The structure and mechanism of the EXD2 nuclease in DNA repair

Karolina Maria Chabowska

The Institute of Cancer Research
University of London

This thesis is submitted for the degree of Doctor of Philosophy.

January 2023

Declaration

The work described in this thesis was carried out at the Institute of Cancer Research, London, under the supervision of Professor Wojciech Niedzwiedz and Professor Alessandro Vannini.

I, Karolina Maria Chabowska, declare that the work present in this thesis is my own and where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Date: January 2023

Signature

Chabowska

Acknowledgments

First, I would like to thank Prof. Wojciech Niedzwiedz and Prof. Alessandro Vannini for giving me the opportunity to work on this project. Your knowledge and enthusiasm are extremely encouraging and inspired me to pursue a career in research. I am especially grateful to Wojciech for his guidance, support, and lots of scientific discussions.

My big thank you to Dr Jadwiga Nieminuszczy and Dr Erin Cutts who were always helpful and gave me loads of invaluable advice. You taught me not only the wet lab skills, but also how to become a better scientist.

Next, I would like to thank Dr Fabienne Beuron from Electron Microscopy facility, who performed negative staining data collection and analysis and helped with the structural characterization of EXD2. Thank you also to Dr Guiller Abascal Palacios and Dr Jerome Gouge for the help with structural analysis.

I wish to thank all other members of Wojciech's and Alessandro's teams: Dr Peter Martin, Dr Joanna Krwawicz, Dr Alexandra Kanellou, Camelia Mocanu, Dr Ronan Broderick, Dr Thangavelu Kaliyappan, Caroline Clarke, Malgorzata Dylewska, Dr Lucy Curnow, Dr Ewan Ramsay, Dr Helen King, Dr Rosie Shen and Dr Charlotte Smith. You created a friendly atmosphere and provided an excellent humour, making my PhD an enjoyable experience. My big thank you to Gosia, Alex and Peter, for the support especially during the COVID-times lab shift pattern, and to Ronan for help during the writing.

I would like to thank my amazing parents, Dorota and Mirek, for always believing in me and all the support throughout the years. Together with Michal and Daria, you have always encouraged me to dream big and I am extremely grateful for that.

Lastly, thank you to my partner Piotrek for loads of smiles and hugs. The completion of this thesis would have been much tougher without your support.

Abstract

DNA double strand breaks (DSBs) are one of the most toxic DNA lesions. They can jeopardize stability of the genome and lead to cancer development. There are two main pathways of DSB repair: homologous recombination (HR) and non-homologous end joining (NHEJ). EXD2 is a recently identified 3'-5' exonuclease, which promotes HR by facilitating DNA end resection. It has been established that the MRN complex (MRE11, RAD50, NBS1), as well as EXD2, are required for efficient HR. However, the exact contribution of EXD2 in HR, and other DNA repair pathways, is unknown.

To elucidate the molecular mechanism of DNA end resection and the role of EXD2 in this process, I optimized purification of full-length hEXD2 expressed in insect cells and performed biochemical characterization of EXD2 activity *in vitro*. EXD2 exhibits 3'-5' exonuclease activity on multiple DNA structures, including 3' phosphorylated DNA, mismatched nucleotides, hairpins, telomeric G-quadruplexes, gapped DNA and structures mimicking active and reversed replication forks. This data suggests that EXD2 may function to 'clean' modified DSBs to enable their efficient repair. It can also process non-canonical DNA structures, facilitating DNA replication. Moreover, RPA inhibits EXD2 activity on ssDNA ends and decreases digestion efficiency on substrate with an accessible dsDNA end.

Interestingly, this thesis reports the first observation of EXD2's weak endonuclease activity. Generation of a novel (H442A, H463A) mutant highlights the importance of HNH domain, which is present in many endonucleases. A preliminary 3D structure of full-length EXD2, obtained by Dr Fabienne Beuron, shows that both exonuclease and HNH-domains are located close to a tunnel, which may be an entry site for DNA substrates.

The results of this PhD will help to better understand mechanism of EXD2 in DNA repair and may pave the way for developing new therapeutic approaches to treat cancer.

Table of contents

List of tables

List of figures

List of abbreviations

2D/3D	Two/three dimensional
32P	Phosphorus-32 radioactive isotope
Å	Angstrom
A_{280}	Absorbance at 280 nm
aa	Amino acids
ALT	Alternative lengthening of telomeres pathway
Alt-EJ	Alternative end-joining
AP	Apurinic/apyrimidinic site
APS	Ammonium persulfate
ATM	Ataxia telangiectasia mutated kinase
ATP	Adenosine triphosphate
ATR	Ataxia telangiectasia Rad3-related kinase
ATRIP	ATR-interacting protein
AU	Absorbance units
BER	Base excision repair
BIR	Break-induced replication
BLM	Bloom syndrome RecQ-like helicase
BME	2-mercaptoethanol
bp	Base pair
BRCA 1/2	Breast cancer associated protein (1/2)
BSA	Bovine Serum Albumin
CDK	Cyclin dependent kinase
CHK 1/2	Checkpoint kinase (1/2)
Cntr	Control
CO	Crossover
Co	Cobalt
CtIP	CtBP-interacting protein
Da	Dalton
DDR	DNA damage response
dHJ	Double Holliday junction
DNA	Deoxyribonucleic acid
DNA2	DNA replication helicase/nuclease 2
DNA-PKcs	DNA-dependent protein kinase catalytic subunit
dsDNA	Double stranded DNA
DSB	Double strand break

DSE	Double strand end
E	Extinction coefficient
EM	Electron microscopy
EMSA	Electrophoretic mobility shift assay
EXD2	Exonuclease 3’-5’ domain-containing protein 2
EXO1	Exonuclease 1
FAM	Fluorescein
FANC	Fanconi anemia pathway
g	Relative centrifugal force
G1/2	Gap 1/2 stage during cell cycle
G4	G-quadruplexes
GF	Gel filtration
H2AX	Histone 2AX
HR	Homologous recombination
ICL	Interstrand DNA crosslink
IR	Ionizing radiation
Kb	Kilobase
Lig	Ligase
M	Molar concentration
Mg	Magnesium
ml	Millilitre
mM	Milimolar
MMR	Mismatch repair
Mn	Manganese
MRE11	Meiotic recombination 11
MRN	MRE11 - RAD50 - NBS1 complex
MUT	Mutant
Mw	Molecular weight
MWCO	Molecular weight cutoff
μl	Microlitre
μM	Micromolar
NBS1	Nibirin
NCO	Non-crossover
NER	Nucleotide excision repair
NHEJ	Non-homologous end joining
Ni	Nickel

NLS	Nuclear localization signal
nm	Nanometre
NMM	N-methyl mesoporphyrin IX
nt	Nucleotides
PARP	Poly(ADP-ribose) polymerase
PCR	Polymerase Chain Reaction
PMSF	Phenylmethylsulfonyl fluoride
PNK	Polynucleotide kinase
Pol	Polymerase
PS	Phosphorothioate modifications
RAD50	DNA repair protein RAD50
RCF	Relative centrifugal force
RFU	Relative fluorescence units
R-loops	RNA/DNA hybrids
RNA	Ribonucleic acid
RNAP	RNA polymerase
RPA	Replication protein A
S	DNA synthesis stage during cell cycle
SDSA	Synthesis-dependent strand annealing
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SSA	Single-strand annealing
SSB	Single-strand break
TEMED	Tetramethylethylenediamine
TM	Transmembrane domain
TOP 1/2	Topoisomerase 1/2
TOPccs	Topoisomerase cleavage complexes
UV	Ultraviolet
WRN	Werner syndrome protein
WT	Wild type

1. Introduction

1.1 DNA damage

1.1.1 Sources and types of DNA damage

Deoxyribonucleic acid (DNA) is the fundamental unit of inheritance, used as a template for replication and transcription. Its high reactivity and susceptibility for chemical modifications may lead to mutations which can jeopardize stability of the genome. Every day, each human cell has to cope with ~ 70 000 lesions (Tubbs & Nussenzweig, 2017), which can be triggered by exogenous and endogenous factors.

The most known sources of exogenous DNA damage are listed in (Table 1), (Chatterjee & Walker, 2017). The less known factors are food preservatives (N-nitrosamines, sodium benzoate, potassium benzoate, potassium sorbate), food additives (citric acid, phosphoric acid, brilliant blue, sunset yellow) and environmental stress (extreme heat or cold, hypoxia), (Chatterjee & Walker, 2017).

Table 1. Sources of exogenous DNA damage, with types of DNA damage they generate. Intra-strand crosslinks - crosslinks of DNA bases within the same DNA strand. Inter-strand crosslinks - crosslinks of DNA bases on opposite DNA strands. DNA adducts - chemical modifications of DNA.

Exogenous sources of DNA damage	Types of DNA damage
Ionizing radiation (IR): Environment (soil, rocks, radon) Medical devices (X-rays machines)	Single strand breaks (SSB), Double strand breaks (DSB), Base lesions
Ultraviolet radiation (UV): Sunlight, Laboratory devices	Covalent linkages between two adjacent pyrimidines, DNA - protein crosslinks, DNA adducts, DNA strand breakages
Alkylating agents: Cigarette smoke, Industrial processing, Chemotherapeutics	DNA adducts, intra- and inter- strand DNA crosslinks, DNA - protein crosslinks
Aromatic Amines: Cigarette smoke, Fuel, Industrial dyes, Pesticides	DNA adducts
Polycyclic Aromatic Hydrocarbon: Cigarette smoke, Fossil fuels, Vehicle exhaust, Charred food	DNA adducts

DNA damage can be also caused by endogenous factors, during normal cellular metabolism. Although the accuracy of DNA replication is high, DNA polymerases may generate errors, including single-base pair substitutions and single-base deletions (Kunkel, 2009). Repetitive sequences may cause replication slippage leading to frameshift mutations (Viguera, Canceill, & Ehrlich, 2001). The action of topoisomerases (TOP), which solve topological problems of DNA during replication and transcription (J. C. Wang, 2002), can also lead to DNA damage. To relieve torsional stress, topoisomerases introduce transient cleavages of DNA (Pommier, Sun, Huang, & Nitiss, 2016). When DNA lesions are present and topoisomerases become trapped at DNA ends, more DNA damage can be created, including DSBs (Pommier et al., 2016). DSBs are one of the most toxic DNA lesions (Mehta & Haber, 2014) and they are discussed in more detail in subchapter (1.2).

Another endogenous source of DNA damage is spontaneous base deamination (Chatterjee & Walker, 2017). In this process cytosine (C) can become uracil (U), leading to further mutations during replication. Base lesions can also be created through spontaneous methylation of DNA by S-adenosylmethionine (SAM), or through DNA oxidation by reactive oxygen species (ROS). Spontaneous hydrolysis of N-glycosyl bonds between nitrogenous bases and the sugar phosphate backbone can create abasic sites, which can be converted into single-strand breaks (SSBs), (Chatterjee & Walker, 2017).

1.1.2 The role of DNA damage in carcinogenesis and cancer treatment

Increased rates of DNA damage and mutations leads to genome instability. Loss of DNA repair genes increases the number of mutations further, leading to cellular transformation and cancer (Tubbs & Nussenzweig, 2017).

In order for normal human cells to become tumorigenic, several adaptations must occur, termed the hallmarks of cancer (Hanahan & Weinberg, 2000). These include the ability to sustain proliferative signalling, evade growth suppressors, avoid immune destruction, enable replicative immortality, activate invasion and metastasis, induce or access the vascular system, resist cell death and deregulate cellular metabolism (Hanahan, 2022). Currently, the core hallmarks also comprise two enabling characteristics: tumour-promoting inflammation and genome instability and mutation (Hanahan, 2022).

Elevated mutations in DNA repair genes are associated with numerous cancers. Identification of these mutations is extremely important, as it allows tailoring treatments to supress DNA damage responses of cancer cells. Mutations in certain DNA repair genes may increase the dependence on other genes/pathways for survival, making DNA damage repair a major clinical target (R. Huang & Zhou, 2021). One example of this is the use of poly(ADP-ribose) polymerase inhibitors (PARPi) for treatment of patients with *BRCA*-mutated cancers (Lord & Ashworth, 2017). BRCA1, BRCA2 and PARP are proteins involved in different DNA repair pathways (BRCA1/2 in homologous recombination, PARP in base excision repair). Inhibition of PARP in BRCA1/BRCA2-mutated cancer cells, disables efficient DNA repair and results in chromosomal instability and apoptosis (Farmer et al., 2005). Mutations in DNA repair pathway genes can lead to hereditary cancers (Romero-Laorden & Castro, 2017). One of the most known examples are *BRCA1*/*BRCA2* mutations, which affect the DSB repair pathway, increasing the risk of hereditary breast and ovarian cancer. Mutations in checkpoint kinase 2 (*CHEK2*) and ataxia telangiectasia mutated (*ATM*), proteins engaged in DNA damage response, also increase the risk of hereditary breast cancer.

As one of the hallmarks of cancer is the ability to sustain proliferating signalling and replicative immortality, many cancer treatments target DNA replication (Helleday, Petermann, Lundin, Hodgson, & Sharma, 2008). Replication inhibitors, like aphidicolin (inhibiting DNA polymerases) and hydroxyurea (inhibiting synthesis of dNTPs), affect replication fork progression and lead to fork collapse and DSB formation. Alkylating agents, which either modify single DNA bases (monofunctional alkylating agents) or generate DNA-protein crosslinks, intra- and inter- strand DNA crosslinks (bifunctional alkylating agents), are also used as anticancer drugs (cisplatin, mitomycin C). The chemical modifications they introduce affect DNA replication and may lead to other toxic lesions like DSBs (bifunctional alkylators). Another type of treatment that affects DNA replication are topoisomerase inhibitors (camptothecins, etoposide), which trap topoisomerases on DNA, creating SSBs and DSBs. Other therapies, like ionizing radiation and bleomycin, introduce DSBs independently of replication (Helleday et al., 2008).

1.1.3 DNA damage response

DNA damage is recognized by DNA damage sensors that identify the lesion, trigger cell signalling and recruit other proteins to the site (R. X. Huang & Zhou, 2020). As there are multiple types of DNA damage, different DNA damage response (DDR) mechanisms can be engaged (Table 2), (Jackson & Bartek, 2009). In mismatch repair (MMR), MutSα or MutSβ recognize and bind mismatched bases, then recruit other components of the complex (MutLα, proliferating cell nuclear antigen - PCNA, replication factor C – RFC).

Their assembly initiates incision of ssDNA near the mismatch, which is further extended by Exonuclease 1 (Exo1), filled by DNA Polymerase δ and ligated by Ligase 1 (Pećina-Šlaus, Kafka, Salamon, & Bukovac, 2020). In base excision repair (BER), DNA glycosylase recognize and remove damaged bases leaving an abasic site (Krokan & Bjørås, 2013). This site is further processed *via* different BER sub-pathways comprising incision, end processing, gap filling and ligation steps. Nucleotide excision repair (NER), repairs various lesions distorting the DNA helix, including bulky base adducts. It consists of the global genomic NER (GG-NER) sub-pathway, which acts on the overall genome, and transcription-coupled NER, which repairs lesions blocking RNA polymerases (Kusakabe et al., 2019). The Fanconi anemia pathway (FANC) repairs inter-strand DNA crosslinks (ICLs), (Ceccaldi, Sarangi, & D'Andrea, 2016). FANC proteins detect ICLs, which are repaired by cooperation of multiple proteins from different DNA damage pathways. DDR mechanisms engaged in DSBs repair, including non-homologous end joining and homologous recombination, are described in more detail in (1.3.2) and (1.4) chapters. Signalling cascades in the response to DSBs are described in (1.2.4).

Table 2. Chosen DNA damage response (DDR) mechanisms with main DNA lesions they act upon, based on (Jackson & Bartek, 2009).

DDR mechanism	DNA damage
Mismatch repair (MMR)	DNA mismatches, insertions, deletions
Base excision repair (BER)	Base lesions
Nucleotide excision repair (NER)	DNA helix distorting lesions
Fanconi anaemia pathway (FANC)	Inter-strand DNA crosslinks
Homologous recombination (HR)	DSBs, inter-strand DNA crosslinks
Non-homologous end joining (NHEJ)	DSBs

The cell cycle, which can be divided into 4 phases (G1, S, G2, M), is regulated by multiple proteins, including cyclins and cyclin dependent kinases (CDKs), (Nurse, 2002), (R. X. Huang & Zhou, 2020). Preserving of genome stability is ensured by cell cycle checkpoints, which protect cells from entering the next phase until favourable conditions exist. The checkpoints are divided into DNA structure checkpoints (DSCs) and spindle assembly checkpoints (SACs), which regulate mitotic entry. Depending on the extent of the damage, DDR either mediates arrest of the cell cycle at checkpoints and promotes DNA repair, or leads to senescence and apoptosis (Zannini, Delia, & Buscemi, 2014). In the presence of DNA damage, arrest at S-phase inhibits DNA synthesis, whereas at G2/M it prevents entry into mitosis.

1.2 Double strand DNA breaks

1.2.1 Sources of double strand DNA breaks

Disruption of both DNA strands results in double strand break (DSB) formation (Kass, Moynahan, & Jasin, 2016). It has been estimated that ~ 1 % of single-strand lesions in normal human cells is converted into ~ 50 DSBs per cell per cell cycle (Vilenchik & Knudson, 2003).

One of the main exogenous sources of DSBs formation is ionizing radiation (IR). It can arise from compounds naturally occurring in the environment (radon) or as a result of medical treatments (radiotherapy), (Jackson & Bartek, 2009). IR-induced DSBs usually arise when clustered SSBs are formed in the opposite strands in the close proximity (<10 bp), (Vítor, Huertas, Legube, & de Almeida, 2020). Action of radiomimetic drugs like bleomycin, and infections with cancer-associated bacteria *Helicobacter pylori* (Toller et al., 2011) are among other exogenous sources of DSBs.

As introduced in subchapter (1.1.1), topoisomerase cleavage complexes (TOPccs), trapped by anticancer drugs or pre-existing DNA lesions, may also lead to DSBs formation (Pommier, Nussenzweig, Takeda, & Austin, 2022). When TOP1cc cleaves one DNA strand close to a nick on the second strand (Figure 1 A) or when cleavage of both strands by TOP2 becomes irreversible, DSBs may be formed. Collision of trapped TOP1cc with replication forks is another source of DSBs (Figure 1 B).

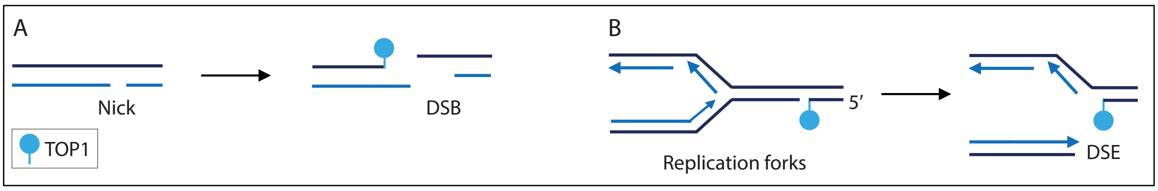


Figure 1. Trapped topoisomerase cleavage complexes can lead to DSBs, (based on Pommier et al., 2016). A) TOP1cc cleavage opposite to a nick generates DSB. B) Collision of trapped TOP1cc with a replication fork generates a double stranded end (DSE).

The majority of endogenously derived DSBs are generated during DNA replication (Cannan & Pederson, 2016). Replication forks may stall not only because of trapped TOPcss, but also when they encounter DNA secondary structures, different types of DNA lesions (inter-strand crosslinks, abasic sites, oxidized bases) and SSBs which may be formed either by damaging agents or as a repair intermediate. This polymerase stalling may further result in DSBs (described in detail in subchapter 1.5).

DSB formation can also occur due to transcription (Ui, Chiba, & Yasui, 2020). During transcription elongation by RNA Polymerase II (RNAPII), DNA lesions may inhibit forward translocation of the enzyme (Lans, Hoeijmakers, Vermeulen, & Marteijn, 2019). Lesion-blocked transcription may lead to the generation of RNA/DNA hybrids (R-loops). The unpaired DNA strand in R-loops is more prone to damage than dsDNA. R-loops may also be precursors of DSBs (Wimberly et al., 2013). Research performed in *E. coli* shows that R-loops may prime replication forks that collapse at DNA template nicks, leading to the generation of DSEs (Figure 2).

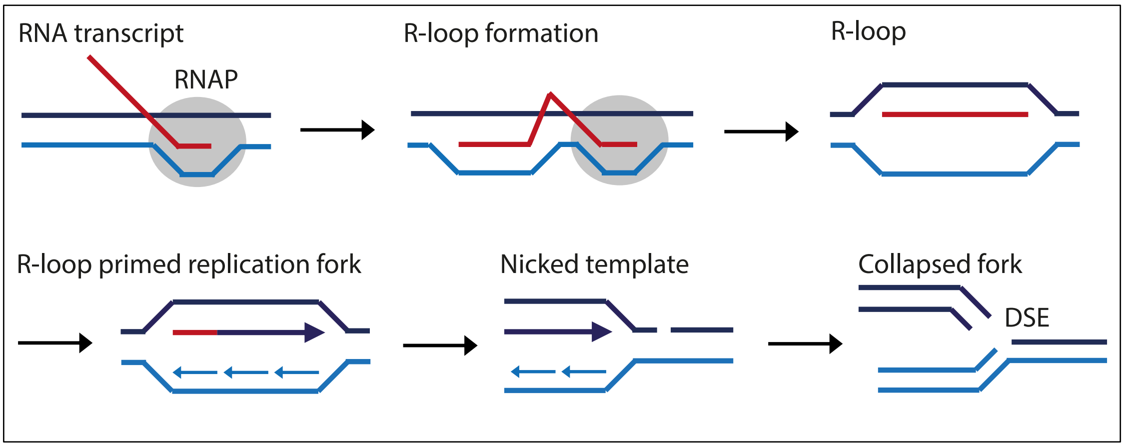


Figure 2. Transcription promoted formation of DSEs, (based on Wimberly et al., 2013). R-loops form when RNA (red) is incorporated into DNA (dark blue, light blue) behind the site of transcription. After RNA polymerase (RNAP) is removed, an R-loop can form a replication fork. When this replication fork encounters a nick in the DNA template, it may collapse, generating a double-stranded end (DSE).

Some DSBs are generated as a result of processing of other DNA lesions (Cannan & Pederson, 2016). Oxidized bases and apurinic/apyrimidinic sites (AP) are repaired in base excision repair pathway (BER), which forms gapped DNA intermediates. If the lesions are located close to each other and two BER reactions are taking place at the same time, SSBs can be created on both DNA strands leading to formation of DSBs (Figure 3).

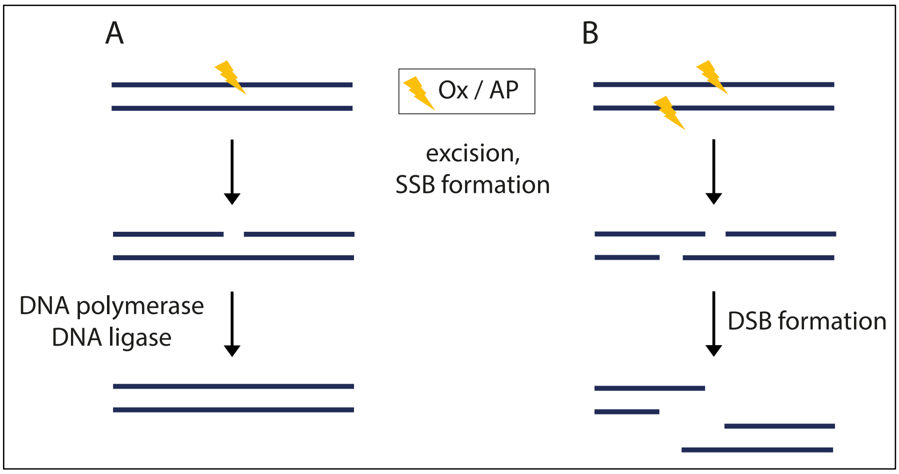


Figure 3. Base excision repair (BER), (based on Cannan & Pederson, 2016). Oxidized bases (Ox) are excised by DNA glycosylases, apurinic/apyrimidinic sites (AP) are excised by AP endonuclease. A) Single base gaps are repaired by DNA polymerase β and DNA ligase III α. B) Excision of closely opposed lesions leads to closely opposed gaps, which can be spontaneously converted to DSBs before repair is completed.

1.2.2 Types of DNA double strand breaks

DSBs are structurally diverse and can be either chemically ‘clean’ or ‘dirty’, hence they can be repaired through different pathways and with different kinetics (Cannan & Pederson, 2016). Clean ends have one 5’ phosphate end and one 3’ hydroxyl end. They can be blunt-ended or possess 3’ or 5’ overhangs, which can create secondary structures affecting DNA end resection (Cejka, 2015). Clean ends can also be created by endonucleases (Mehta & Haber, 2014).

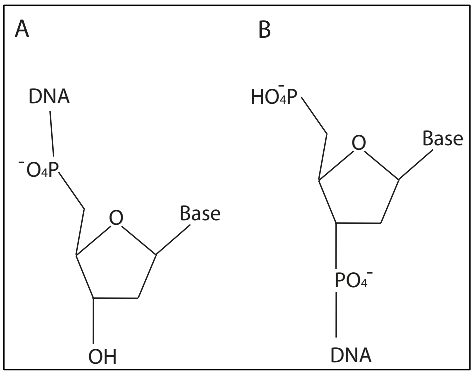


Figure 4. Schematic representation of clean DNA ends with A) 3’ hydroxyl end and B) 5’ phosphate end.

Dirty ends possess chemical alterations and can be almost impossible to repair directly (Cannan & Pederson, 2016), therefore they require additional processing enabling re-joining. DSBs induced by free radicals, which may originate from IR, may generate nucleoside 5’-aldehydes on the 5’ end, and phosphoglycolate (PG), phosphoglycoaldehyde, formyl phosphate and 3’-keto-2’deoxynucleotide on the 3’ end (Figure 5), (Povirk, 2012). They are mostly unstable (except for PG) and may break down spontaneously generating 3’-and 5’- phosphates. IR-induced DSBs can be also accompanied by other clustered lesions including base damage and base loss (Schipler & Iliakis, 2013).

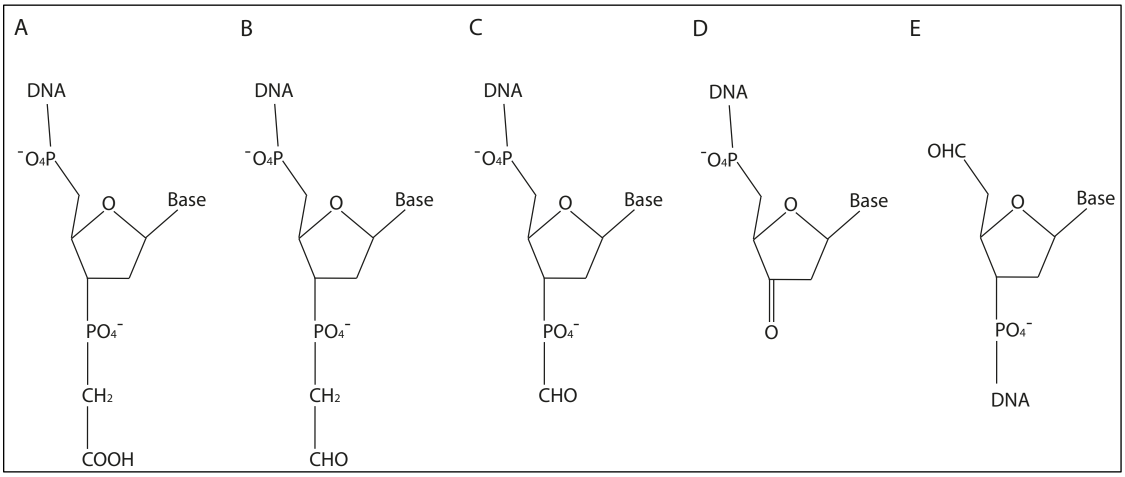


Figure 5. Schematic representation of ‘dirty’ DSBs, (Povirk, 2012). A) 3’-phosphoglycolate. B) 3’-phosphoglycoaldehyde. C) 3’-formyl phosphate. D) 3’-keto-2’-deoxynucleotide. E) 5’-aldehyde.

Pathological, irreversible complexes of human topoisomerases also create ‘dirty’ ends as they form DNA-protein (topoisomerase) crosslinks (TOP-DPCs), (Pommier et al., 2022). Irreversible DSBs created by TOP2 can have a 3’- hydroxyl end and 5’- overhang (4 bases), (Povirk, 2012).

1.2.3 Consequences of DNA double strand breaks

Introduction of DSBs can be either beneficial or pathological (F. A. Khan & Ali, 2017). DSBs that are created as a part of physiological processes, allow diversification of immune responses through V(D)J recombination (Soulas-Sprauel et al., 2007). Recombination of variable (V), diversity (D) and joining (J) gene segments allows generation of unique antigen receptors enabling B and T lymphocytes to respond to foreign pathogens. DSBs allow also genomic diversification of germ cells (Lam & Keeney, 2015). During meiosis I, when homologous chromosomes are segregated, programmed formation and repair of DSBs allows recombination of genetic material (Lam & Keeney, 2015). Controlled topoisomerase-mediated DSBs allows manipulation of DNA topology facilitating DNA replication, transcription and regulation of gene expression (F. A. Khan & Ali, 2017).

Pathological DSBs are believed to be the most harmful type of DNA damage, which can jeopardize stability of the genome (R. Huang & Zhou, 2021). They may cause rearrangements of large DNA fragments, known as chromosomal translocations, leading to cell death or carcinogenesis (Gómez-Herreros, 2019). These rearrangements occur when ends of two distinct DSBs are joined (Hu et al., 2016). The frequency of translocations depends on DSB generation rate and their re-joining speed, as well as on the frequency of close physical contact between two sites of the genome. If two DSBs are closely linked in the same chromosome, intra-chromosomal deletions may occur (Hu et al., 2016). Clustered DSBs are associated with mutational process called chromothripsis, leading to complex chromosomal rearrangements in somatic and germ cells (Nazaryan-Petersen, Bjerregaard, Nielsen, Tommerup, & Tümer, 2020). In this process even hundreds of chromosome fragments can be joined together in random order, leading to cancer and developmental disorders (Kass et al., 2016). It is estimated that up to 5% of tumours shows signs of chromothripsis.

To preserve genome integrity, cells detect and signal the presence of damage using DNA-damage response pathways (DDR), which promote DNA repair (Jackson & Bartek, 2009). Defects in DSB DDR pathways leads to the development of numerous diseases with common phenotypes of immunodeficiency, radiosensitivity and predisposition to cancer (McKinnon & Caldecott, 2007).

1.2.4 DNA damage response in the context of double strand breaks

There are several DNA damage sensors engaged in the response to DSBs, including γH2AX, mediator of DNA damage checkpoint 1 (MDC1), p53-binding protein 1 (53BP1), MRN (Meiotic recombination 11, ATP-binding cassette-ATPase, Nijmegen breakage syndrome / Nibirin) and BRCA1 (R. X. Huang & Zhou, 2020), (Zannini et al., 2014), (Jackson & Bartek, 2009). They can localize to sites of DSBs within seconds or minutes, recruit other proteins to the damaged sites and regulate each other activities. Phosphorylation (S139) of H2AX variant of the core histone is an early response to DNA damage (Burma, Chen, Murphy, Kurimasa, & Chen, 2001). Hence, cellular visualisation of γH2AX foci is often used to validate the presence of lesions. γH2AX recruits also other proteins engaged in DDR: MDC1, 53BP1, BRCA1 (Stucki et al., 2005), (Kleiner, Verma, Molloy, Chait, & Kapoor, 2015), (Krum, Dalugdugan, Miranda-Carboni, & Lane, 2010). Through interaction with γH2AX, MDC1 is present at the site of the damage, and it controls recruitment of other DNA repair factors (Stucki et al., 2005). Depending on the activated sensors, different DSBs repair pathways can be promoted (discussed further in 1.3.1).

One of the most important sensors of DSBs is MRN complex, which recruits activated kinases ataxia telangiectasia mutated (ATM) to DSBs (Uziel et al., 2003). ATM is a main regulator of cellular response to DSBs (R. X. Huang & Zhou, 2020), (Zannini et al., 2014). Together with ataxia telangiectasia Rad3- related (ATR), it starts a cascade of responses and leads to phosphorylation of different proteins at different stages of the cell cycle. One of the best-known ATM/ATR targets are CHK1 and CHK2, which can influence CDK activities (R. X. Huang & Zhou, 2020). Inhibition of CDKs can cause cell cycle arrest, giving more time for the repair before replication and/or mitosis. In the meantime, ATM/ATR stimulate repair by recruiting DNA-repair proteins and promoting their activity. One of the examples of checkpoint regulation is phosphorylation of BRCA1, which regulates ATM/ATR to phosphorylate the transcription factor p53 (R. X. Huang & Zhou, 2020). Then, phosphorylated p53 induces p21, which inhibits CDK and leads to cell cycle arrest (R. X. Huang & Zhou, 2020).

1.3 Double strand DNA breaks repair

1.3.1 Double strand DNA break repair pathway choice

Cells possess at least four mechanisms of DSB repair: canonical non-homologous end-joining (NHEJ), homologous recombination (HR), alternative end-joining (alt-EJ) and single-strand annealing (SSA), (Ceccaldi, Rondinelli, & D'Andrea, 2016). The major difference between NHEJ and other pathways is the requirement for extensive DNA end resection, which is prevented in NHEJ (Chang, Pannunzio, Adachi, & Lieber, 2017). End resection provides stretches of single stranded DNA that can be annealed in an HR, alt-EJ or SSA - dependent manner (Jasin & Rothstein, 2013). HR requires over 100 base pair (bp) of homology, SSA over 20 bp, whereas Alt-EJ requires around 2 – 20 bp (Chang et al., 2017).

It is believed that resection of DNA ends may regulate DSB repair pathway choice (Ceccaldi, Rondinelli, et al., 2016). Proteins protecting DNA ends can prevent DSBs from extensive resection, thereby favouring NHEJ. Apart from Ku70 - Ku80 which is a crucial component of the NHEJ pathway, the most known DNA protection factors are TP53-binding protein (53BP1), RAP1-interacting factor 1 homologue (RIF1), the Shieldin complex (SHLD1, SHLD2, SHLD3, REV7) and the CST-Polα complex (CTC1, STN1, TEN1), (B. Zhao, Rothenberg, Ramsden, & Lieber, 2020). 53BP1 is a DNA damage response protein, interacting with RIF1 and Shieldin, which may directly block resection. CST-Polα, which functions downstream of 53BP1-RIF1-Shieldin, mediates DNA synthesis protecting DNA from extensive digestion (Mirman et al., 2018).

The Ku70 and Ku80 proteins form a heterodimer which is very abundant in the cell and is recruited with fast kinetics and high affinity towards DSBs (Reginato & Cejka, 2020). It was proposed that binding of Ku is the first step in repair of DSBs by NHEJ. When Ku-dependent NHEJ is not possible, for example when DNA ends are chemically altered and difficult to ligate, MRN and C-terminal binding protein interacting protein (CtIP) can cleave DNA past Ku-blocked ends (Cejka & Symington, 2021). This cleavage shifts the repair towards HR (Reginato & Cejka, 2020). Removal of Ku can also be mediated by its DNA-PKcs dependent phosphorylation, which decreases its DNA binding affinity, or by ubiquitination dependent degradation (Deshpande et al., 2020). The ATPase p97 regulates DSB repair pathway choice as it removes ubiquitinated Ku80 from the breaks, promoting the HR (van den Boom et al., 2016). p97 is an important component of the ubiquitin-proteosome degradation mechanism. Its inhibition blocks the disassembly of MRN from DNA damage sites and results in excessive DNA degradation by MRE11

(Kilgas et al., 2021). Increased generation of ssDNA prevents repair *via* HR and promotes repair through the more error-prone SSA pathway (Kilgas et al., 2021).

NHEJ can occur during all phases of the cell cycle (G1, S, G2, M) but the cell cycle affects the choice between NHEJ and the resection dependent pathways (Ceccaldi, Rondinelli, et al., 2016). In G1, 53BP1 and Rif1 localize to DSBs promoting NHEJ (Di Virgilio et al., 2013), (Ceccaldi, Rondinelli, et al., 2016). Activity of 53BP1 is regulated through phosphorylation by Ataxia Telangiectasia Mutated (ATM), (Feng et al., 2015). When cells enter S phase (DNA synthesis), expression of HR genes increases (Scully, Panday, Elango, & Willis, 2019). To prevent resection in G1, when a sister chromatid is not available as a template for repair by HR, cells regulate resection by cyclin dependent kinases (CDKs). During S and G2, CDKs stimulate DNA end resection *via* phosphorylation of multiple proteins, including CtIP (Huertas & Jackson, 2009). When NBS1 (an MRN complex component), senses CtIP phosphorylation, it mediates MRN endonuclease-dependent DNA cleavage, which initiates short range resection in the HR pathway (Anand, Ranjha, Cannavo, & Cejka, 2016). Phosphorylation of BRCA1 by ATM also promotes resection dependent pathway and interplay between 53BP1 and BRCA1 modulates NHEJ – HR pathway choice (Feng et al., 2015), (Scully et al., 2019).

The structure of a DSB may also affect the choice of repair. Simple DSBs are more likely to be repaired by NHEJ, while extensively damaged/modified DSBs which may require end resection by HR and SSA (B. Zhao et al., 2020). The length of DNA ends may also influence pathway choice, as long ssDNAs are weakly bound to by the Ku heterodimer (Scully et al., 2019).

1.3.2 Non-homologous end joining

Non-homologous end joining (NHEJ) is the major DSBs repair pathway of dividing and non-dividing cells and occurs with fast kinetics (B. Zhao et al., 2020), (Shibata et al., 2014). As the name implies, NHEJ relies on re-ligation of DNA ends with little or no homology (0 – 4 bp), (Mimitou & Symington, 2011).

The three main steps of NHEJ are synapsis (positioning diffused DNA ends into close proximity), end processing and ligation (B. Zhao et al., 2020). This is achieved by the coordinated action of belt proteins (Ku70/Ku80), kinase/phosphatase component (PNK), nucleases (Artemis:DNA-PKcs), polymerases (Pol μ, Pol λ), and ligase machinery (XLF:XRCC4:PAXX:DNA ligase IV), (Lieber, 2010), (Ochi et al., 2015).

The key component is the DNA-dependent protein kinase (DNA-PK), (Smith & Jackson, 1999), which consists of Ku heterodimer and DNA-dependent protein kinase catalytic subunit (DNA-PKcs), (Khanna & Jackson, 2001). Ku heterodimer (Ku70 and Ku80) exhibits high affinity towards DNA ends (Dynan & Yoo, 1998) protecting them from extensive resection. It promotes binding of DNA-dependent protein kinase (DNA-PK) and stimulates its activity (Gottlieb & Jackson, 1993). Ku interacts with DNA polymerases (Pol μ and Pol λ), (Bebenek, Pedersen, & Kunkel, 2014), DNA ligase IV – X-ray repair cross-complementing protein 4 (XRCC4) complex (Costantini, Woodbine, Andreoli, Jeggo, & Vindigni, 2007) and (XRCC4:XLF) paralog PAXX (Ochi et al., 2015).

In NHEJ, DNA ends can be either directly ligated or processed before joining (Figure 6), (Chang et al., 2017). If it is possible, blunt ends are preferentially ligated without processing. The type of the processing and recruited proteins depends on the structure of the broken DNA.

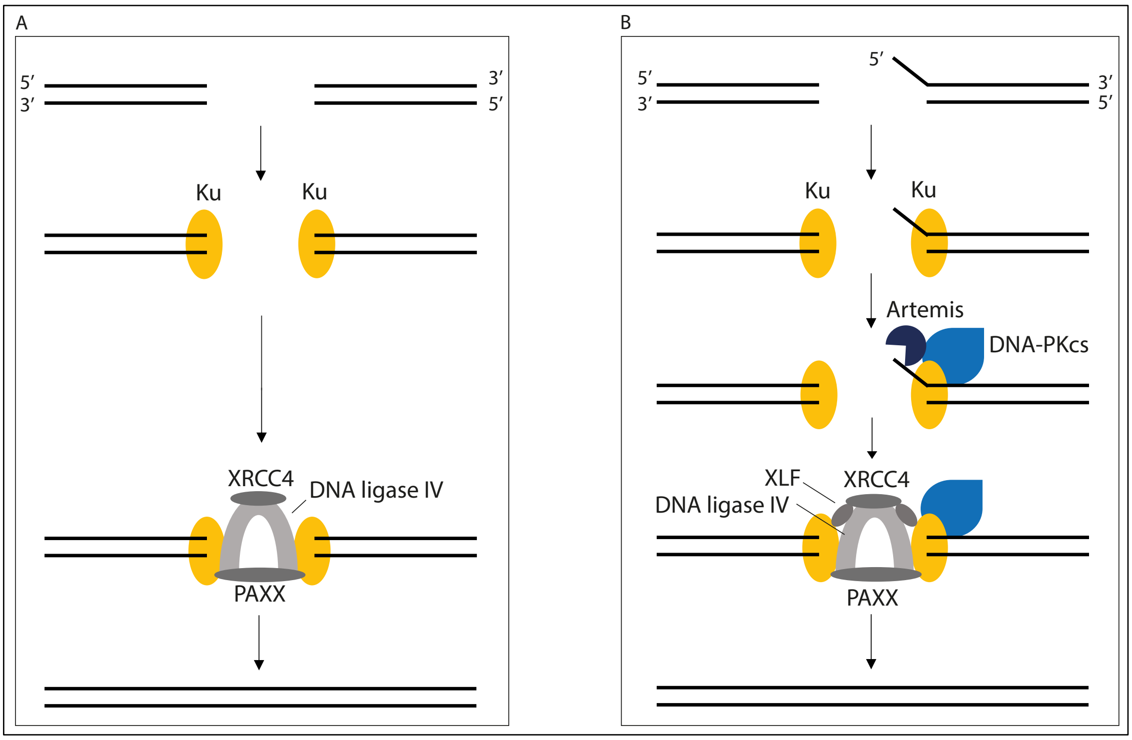


Figure 6. Nonhomologous end-joining, (based on Chang et al., 2017). A) Blunt ends ligation by XRCC4-DNA ligase IV is stimulated by PAXX. B) DSBs with incompatible 5' ends are processed by Artemis – DNA-PKcs. Ligation is stimulated by PAXX and XLF.

If the ends lack phosphate, PNK can phosphorylate the 5' end. If the phosphorylation is undesirable, it can remove the 3' phosphate, facilitating further ligation (Bernstein et al., 2005). When limited DNA resection is required, Artemis is engaged together with DNA-PKcs (Chang et al., 2017). Artemis is believed to be the main NHEJ nuclease, with endonuclease and 5'- exonuclease activities (B. Zhao et al., 2020). Its endonuclease activity is facilitated by DNA-PKcs autophosphorylation (Goodarzi et al., 2006). While digesting 3’ overhangs and hairpins, it usually generates a short 3’ overhang (4 nt), (S. Li et al., 2014). When processing 5’ overhangs, it cleaves mostly at the ssDNA-dsDNA junction leaving the ends blunt. Together with DNA-PKcs it removes 3’-PGs from IR induced DSBs (Figure 7), (Chang et al., 2017). Alternatively, tyrosyl DNA phosphodiesterase 1 (TDP1) can remove PG, before the action of Artemis. DNA ends can be further processed by polymerases, which generate short regions of homology (Figure 7 B). Processing of the broken DNA ends facilitates ligation by DNA ligase IV – XRCC4 complex, supported by XRCC4-like factor (XLF) and/or by a paralogue of XRCC4 and XLF (PAXX), (B. Zhao et al., 2020), (Ma et al., 2004), (Grawunder, Zimmer, Fugmann, Schwarz, & Lieber, 1998), (Ahnesorg, Smith, & Jackson, 2006), (Ochi et al., 2015). Each DNA strand can be ligated independently.

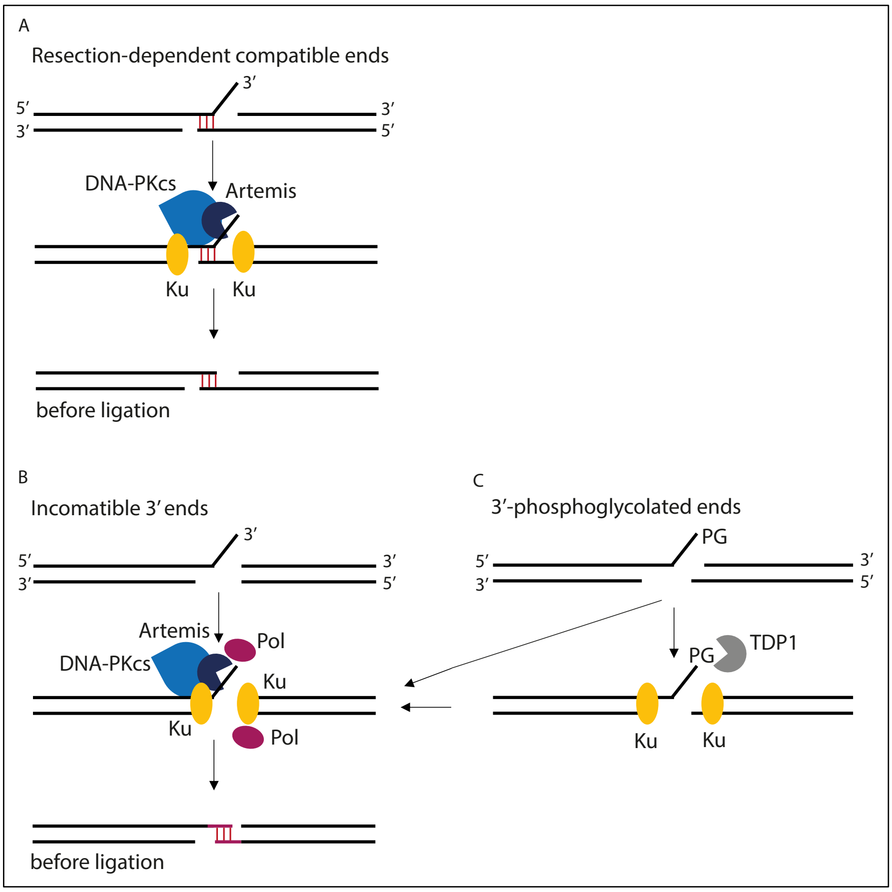


Figure 7. Processing of 3’ ends during NHEJ, (based on Chang et al., 2017). A) DSB with short homologous regions and non-homologous 3’ overhang is processed by Artemis – DNA-PKcs, which cleaves the overhang before ligation. B) A DSB with incompatible 3’ ends is processed by Artemis – DNA-PKcs, while DNA polymerases add nucleotides to generate short homologous regions (red). C) A DSB with 3’ PG can be processed by TDP1 or by Artemis – DNA-PKcs.

Repair by NHEJ is considered to be mutagenic, as NHEJ restores integrity of the DNA but not necessary its original sequence (B. Zhao et al., 2020). It can lead to insertions, deletions and translocations (Chapman, Taylor, & Boulton, 2012). However, its fast kinetics plays an important role in preserving genomic integrity (Ceccaldi, Rondinelli, et al., 2016).

1.3.3 Selected nucleases involved in non-homologous end joining

Artemis nuclease

According to (Yosaatmadja et al., 2021), Artemis (SNM1C/DCLRE1C) is a key NHEJ endonuclease, with only minor 5’-3’ exonuclease activity. Its mutation results in severe immunodeficiency (RS-SCID). Artemis digests different types of DNA structures, including 3’- and 5’- overhangs, hairpins, and DNA with ss-ds junctions (Chang & Lieber, 2016). Its preferred substrates contain ss/dsDNA boundaries and Artemis cleaves at the ssDNA side of the boundary (Chang, Watanabe, & Lieber, 2015). It can cleave hairpin loops, 5’ flaps, split arms and structures with gaps, bubbles, and heterologous loops (Chang et al., 2015) and exhibits 5’ exonuclease activity on ssDNA (S. Li et al., 2014). Its endonuclease activity on different types of dsDNA structures is observed either when in complex with DNA-PKcs or without DNA-PKcs (Chang et al., 2015). According to (Chang et al., 2015), to cut blunt ended dsDNA, Artemis needs both DNA-PKcs and Ku. When only Artemis and DNA-PKcs is present, digestion of blunt ended dsDNA depends on the sequence. Artemis repairs DSBs generated by neocarzinostatin (NCS) and bleomycin. These types of breaks can contain 5’ phosphates, 3’ phosphates, 3’ phosphoglycolates and may be blunt ended or have short overhangs (Chang & Lieber, 2016).

Nuclease activity assays using the catalytic core of the protein (1-361 residues) also show preference towards substrates with ssDNA regions (Yosaatmadja et al., 2021). Here, the catalytic core exhibited higher activity on ssDNA and substrates with 3’ or 5’ overhangs, than on blunt ended dsDNA. Artemis digested the template strand of fork-like structures more efficiently when a “nascent strand” was not present (Yosaatmadja et al., 2021). In experiments performed by (Yosaatmadja et al., 2021), Artemis did not require the presence of other proteins (DNA-PKcs, Ku) for its endonuclease activity on these substrates.

Werner syndrome protein (WRN)

Werner syndrome protein (WRN) is an ATPase with (3’-5’) exonuclease and (3’-5’) helicase activities (Von Kobbe, Thomäs, Czyzewski, Pavletich, & Bohr, 2003). Its mutation results in Werner syndrome (WS), characterized by premature aging and predisposition to cancer. WRN interacts with different types of DNA structures, including

Holliday junctions, split arms, and D-loops (Von Kobbe et al., 2003). It digests dsDNA with 5' overhangs in the presence of either Mn^{2+} or Mg^{2+} (Perry et al., 2006). Its exonuclease activity is higher with Mn^{2+} than with Mg^{2+}, whereas presence of both ions together is indistinguishable from Mn^{2+} alone. WRN digests *in vitro* substrates with telomeric 3' overhangs, not requiring ATPase or helicase activities (B. Li, Reddy, & Comai, 2009). This digestion is limited to the ssDNA region within telomeric substrates and does not depend on the length of this region. Interestingly, non-telomeric dsDNA with 3' overhangs is not digested by WRN (B. Li et al., 2009).

WRN interacts with Ku70/80 (Perry et al., 2006) and XRCC4 – DNA ligase IV (Kusumoto et al., 2008), which stimulates its exonuclease activity. It may process certain types of DNA ends to enable NHEJ. WRN regulates DSB pathway choice between classical NHEJ and alt-EJ (Shamanna et al., 2016). In alt-EJ, MRN and CtIP resect the broken ends. WRN presence at DSBs inhibits recruitment of MRN and CtIP, protecting the 5' ends from digestion. WRN helicase and nuclease activities promote NHEJ. WRN also takes part in HR, discussed further in (1.4). However, WRN depletion results in mild inhibition of HR suggesting it plays a minor role in this pathway. Chromosomes of WS patients display deletions, translocations, and telomere fusions (Shamanna et al., 2016). WRN interacts with components of Shelterin complex, taking part in telomere protection.

1.3.4 Alternative end joining and single-strand annealing

Alternative end joining

Alternative end-joining (alt-EJ), also called microhomology mediated end-joining (MMEJ), is DSB repair pathway which requires around 2 – 20 bp of microhomology (Pannunzio, Watanabe, & Lieber, 2018). In the absence of NHEJ, alt-EJ acts as a backup mechanism (Shamanna et al., 2016).

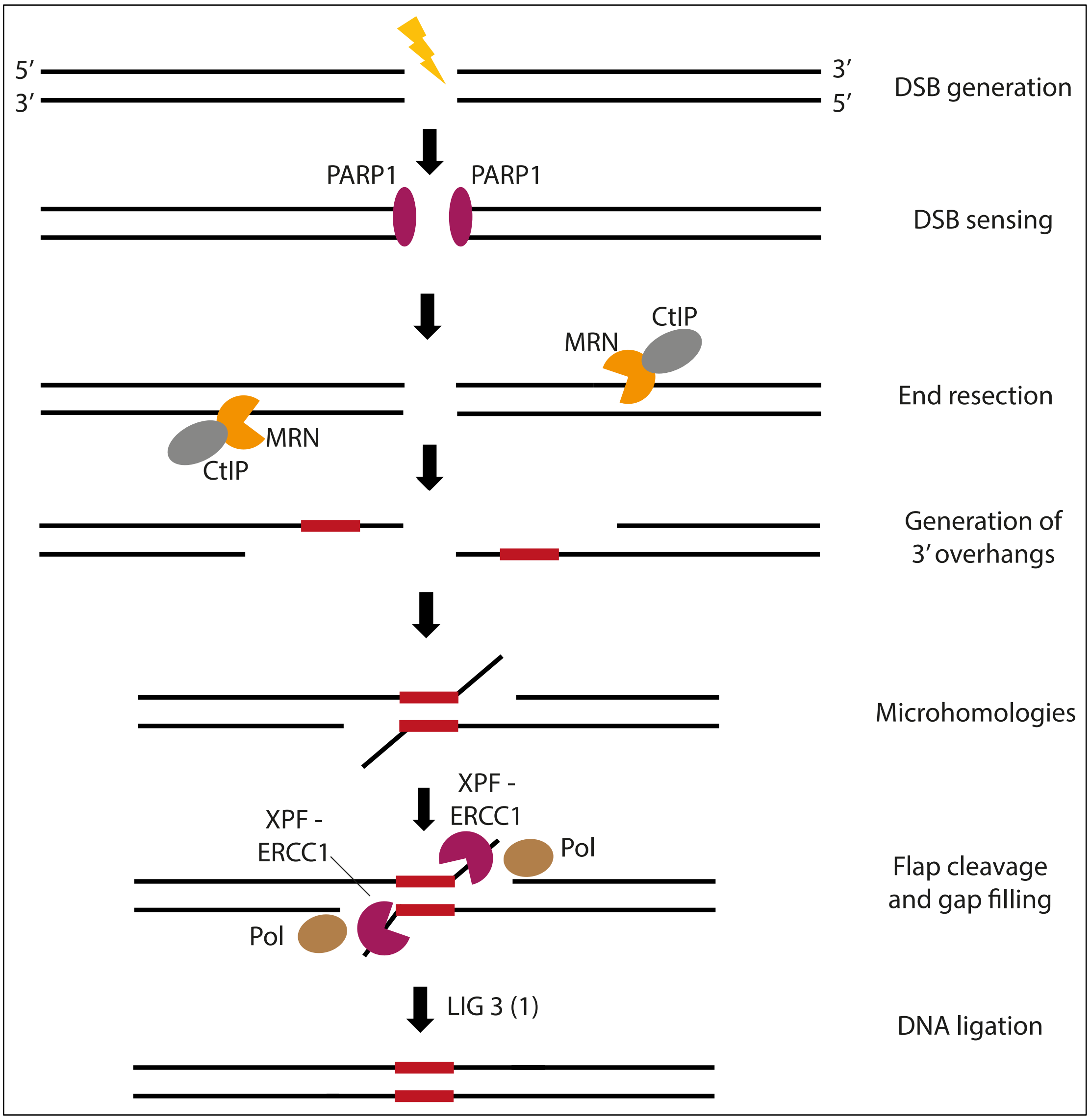


Figure 8. Mechanism of alternative end joining. After DSB recognition by PARP1, MRN with CtIP generates 3' overhangs exposing microhomologies. Heterologous flaps are removed by XPF-ERCC1, DNA gaps are filled by Pol θ, and DNA ends are ligated by LIG3 or LIG1.

Poly ADP-ribose polymerase 1 (PARP1) recognizes DSBs (Sfeir & Symington, 2015) and recruits MRN to perform resection (Figure 8), (Sallmyr & Tomkinson, 2018). MRN, stimulated by CtIP, generates 3' single stranded overhangs (15 – 100 nt) exposing short complementary regions (Chang et al., 2017). RPA, which may bind these overhangs, is displaced by the helicase activity of DNA Polymerase θ (Pol θ), which also stabilizes strand annealing (Mateos-Gomez et al., 2017). Complementary regions anneal together leaving heterologous tails, which are cleaved by XPF-ERCC1 (Seol, Shim, & Lee, 2018). Pol θ fills in the gaps and the ends are ligated by DNA ligase III (LIG3) and DNA ligase I (LIG1).

Single-strand annealing

Repair of DSBs through single-strand annealing (SSA) starts with DNA resection by MRN and CtIP, followed by more extensive resection by the EXO1 nuclease, Bloom syndrome RecQ-like helicase (BLM) or DNA replication helicase/nuclease 2 (DNA2), (Chang et al., 2017). Generated 3' ssDNA overhangs reveal homologous sequences that are annealed in a process mediated by RAD52 (Bhargava, Onyango, & Stark, 2016). The nonhomologous 3' ssDNA ends are cleaved by XPF-ERC1 complex, and the intermediate is processed by polymerase to enable ligation. Type of DNA polymerase and ligase present is SSA is still not fully understood.

The main difference between these two pathways is that SSA requires longer stretch of DNA homology (> 20 bp) and more extensive resection than alt-EJ (Bhargava et al., 2016). Mediators of DNA end synapsis also differ, as for alt-EJ it is PARP and Pol θ, whereas for SSA it is RAD52. Both alt-EJ and SSA pathways are mutagenic as they result in DNA deletions and chromosomal translocations (Sallmyr & Tomkinson, 2018).

1.4 Homologous recombination

1.4.1 Main steps of homologous recombination pathway

Homologous recombination (HR) is a key DSB repair pathway during the S and G2 phases of the cell cycle, when an intact DNA strand (sister chromatid) is available as a template for repair (Moynahan & Jasin, 2010). The pathway starts with the resection of DNA breaks and leads to the formation of single stranded DNA overhangs (Figure 9). In eukaryotes DSB processing is usually a two-step mechanism, starting with short-range, followed by long-range resection (Cejka, 2015). Short-range resection is limited to close proximity of the DNA ends, is slow and capable of digestion of non-canonical DNA structures, whereas long-range resection is fast but less efficient in processing of DNA ends with introduced protein blocks (Cejka & Symington, 2021).

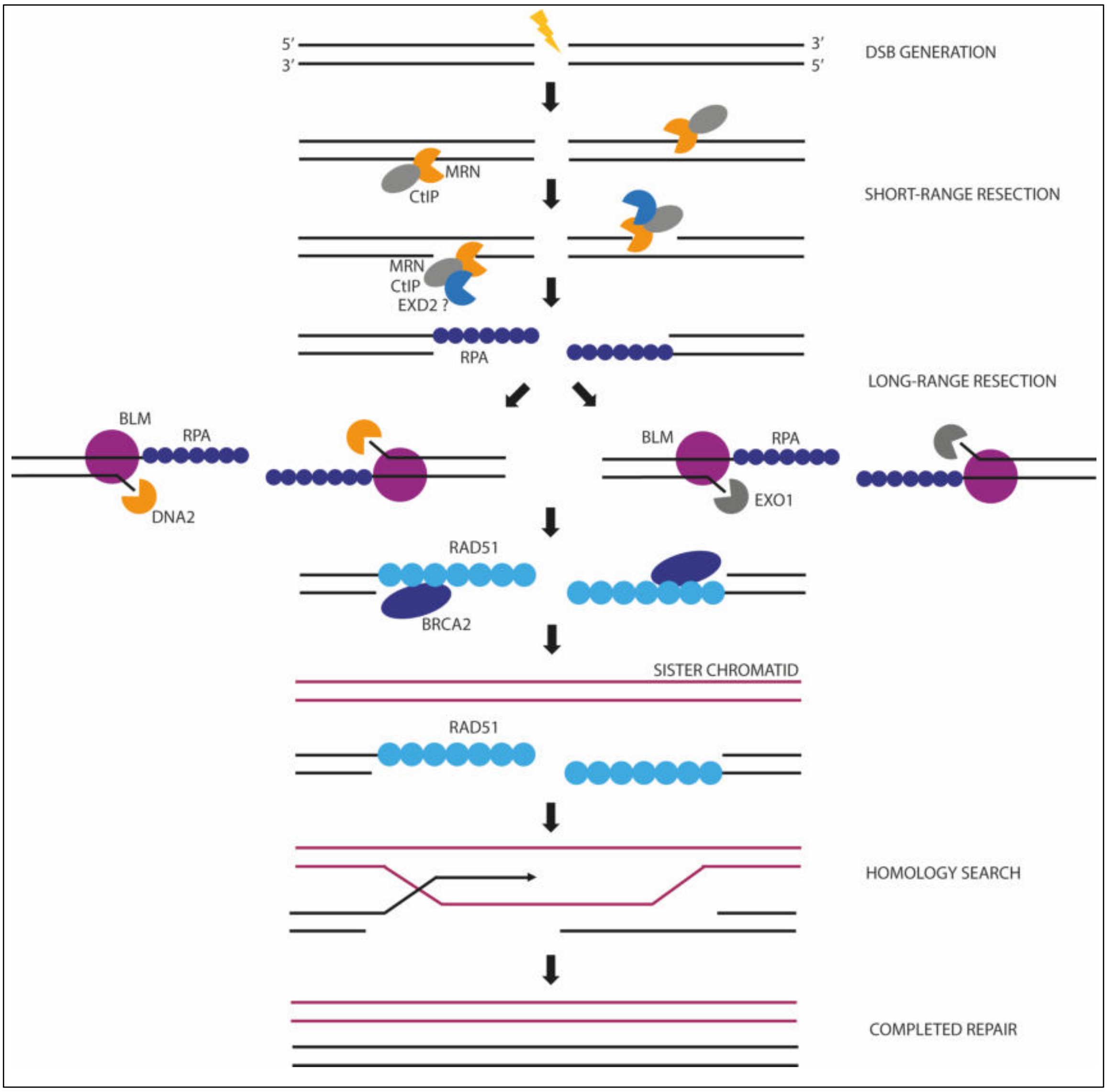


Figure 9. Main steps of homologous recombination. Short range resection by MRN, CtIP and possibly EXD2, is followed by one of two long range resection pathways. Generated long ssDNA overhangs are coated by RAD51, which mediates homology search and strand invasion.

Short-range resection is initiated by the recruitment of the MRN complex (MRE11, RAD50, NBS1) to DSBs (Syed & Tainer, 2018). It is proposed that MRN, stimulated by CtIP, begins resection by endonucleolytic nicking performed several hundred base pairs from the DSB. Then, according to (Syed & Tainer, 2018), MRE11 in association with CtIP starts exonucleolytic degradation in the 3’-5’ direction, towards the break. Recent data by our lab (Broderick et al., 2016) propose that another nuclease - EXD2, is also involved in the initial processing of broken DNA ends, acting in concert with MRE11 to promote short range resection (Figure 9).

After initial short-range resection, one of the two possible long-range resection pathways follow (Nimonkar et al., 2011). The first is mediated by the Bloom helicase (BLM), DNA2 helicase/nuclease, MRN and RPA. Here, MRN increases the affinity of BLM helicase for

DNA ends. The unwinding of dsDNA by BLM in an ATP-dependent manner is then required for DNA resection by DNA2 (Nimonkar et al., 2011). Here, the presence of RPA has been shown to modulate the direction of resection, as when the RPA is present, DNA2 resects in the 5'-3' direction (Cejka et al., 2010).

The second mechanism consists of Exonuclease 1 (EXO1), BLM, MRN and RPA (Nimonkar et al., 2011). Here both BLM and MRN stimulate EXO1 binding to DNA, and MRN additionally increases its processivity allowing EXO1 to resect DNA in the 5'-3' direction. In both long-range resection pathways, thousands of nucleotides are resected from DNA ends leading to the generation of 3' ssDNA overhangs. Here, ssDNA is bound by RPA that melts its secondary structures promoting loading of downstream factors such as RAD51 (Sung, Krejci, Van Komen, & Sehorn, 2003), (Kass et al., 2016). RPA controls the accessibility of RAD51 recombinase to ssDNA by occluding its binding. Therefore, mediator proteins such as BRCA2 are needed to replace RPA with RAD51 (Lisby & Rothstein, 2015), (Von Nicolai, Ehlén, Martin, Zhang, & Carreira, 2016). The RAD51 nucleoprotein filament invades homologous DNA and forms a DNA joint molecule, termed a displacement (D)-loop (Raynard, Niu, & Sung, 2008), (Spírek et al., 2018). DNA synthesis is primed within the D-loop (Heyer, Ehmsen, & Liu, 2010) by DNA polymerases (Sebesta et al., 2013). Pol δ, together with proliferative cell nuclear antigen (PCNA), can extend D-loops up to 2 kb. Translesional polymerases (TLS), including Pol η and Pol κ, generate shorter extension tracks (150 nt). As Pol η extends substrates independently of PCNA, it is possible that PCNA regulates choice of polymerase, affecting the length of D-loop extension (Sebesta et al., 2013).

After D-loop formation, break repair proceeds through one of the HR subpathways: synthesis-dependent strand annealing (SDSA) or via generation of the double Holliday junction (dHJ), (Heyer et al., 2010). In the SDSA pathway, the newly synthetized strand is annealed with the resected second end of the DSB, resulting in a non-crossover product. In the second sub- pathway, a double Holliday junction is formed through recapturing of the second DNA end, followed by either topological dissolution or nucleolytic resolution (Bell & Kowalczykowski, 2016). In the dissolution process the dHJ is resolved by the action of BLM helicase and TOP3, resulting in a non-crossover product. In the resolution mechanism, the dHJ is cut by endonucleases (SLX-MUS, GEN1) creating either non-crossover or crossover products.

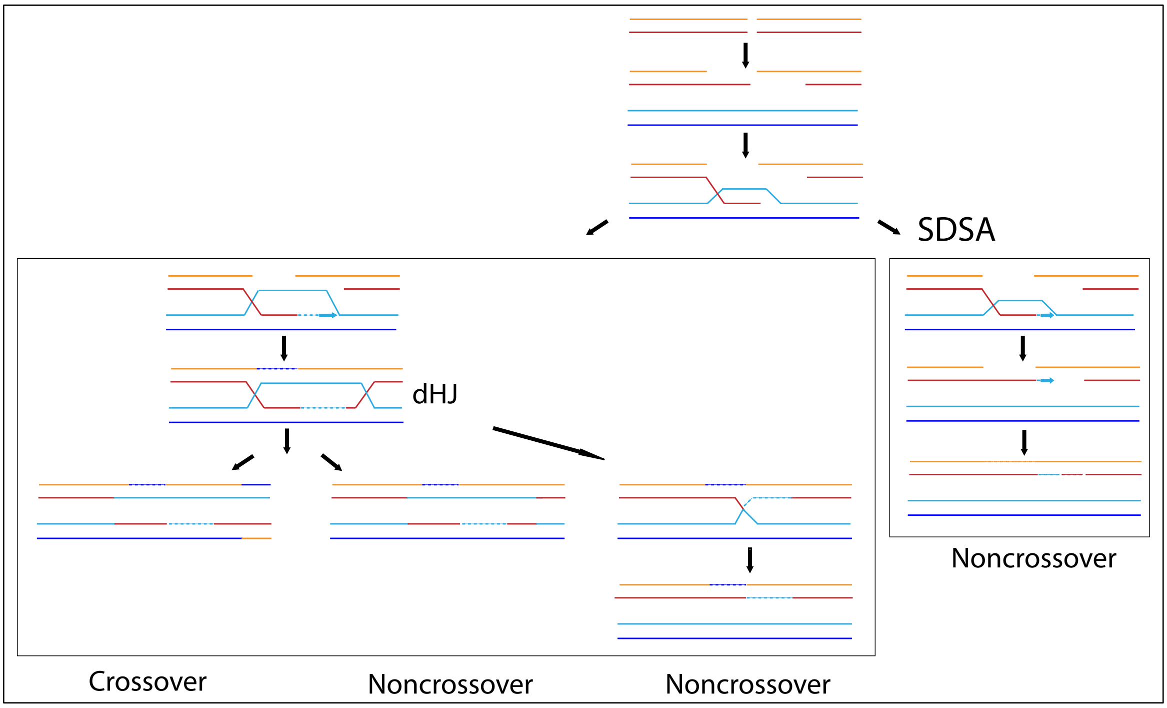


Figure 10. Homologous recombination sub-pathways, (based on Heyer et al., 2010). SDSA – synthesis - dependent strand annealing; dHJ – double Holliday junction; broken lines – newly synthetized DNA.

The activity of proteins taking part in the HR pathway is described in more detail in subchapters (1.4.2), (1.4.3), (1.4.4), and (1.4.5).

1.4.2 The role of MRN and CtIP in homologous recombination

MRN complex (MRE11, RAD50, NBS1) is critical for recognition and processing of DSBs (Deshpande, Lee, Arora, & Paull, 2016). It consists of a nuclease (MRE11), an ATPase (RAD50) and a regulatory docking protein (NBS1), which contains a nuclear localization signal (NLS), (Reginato & Cejka, 2020). MRE11 directly interacts with RAD50, NBS1 and DNA. NBS1 regulates MRN interactions through binding of multiple proteins, including ATM and CtIP. Signalling the presence of DSBs to ATM leads to phosphorylation of hundreds of proteins in DNA damage response pathways. Activated ATM phosphorylates MDC1, which amplifies DDR signalling by recruiting more MRN to DSBs (Syed & Tainer, 2018). CtIP is phosphorylated by ATM and ATR in a DNA damaged-induced manner and is phosphorylated by CDK in a cell cycle-dependent fashion (Makharashvili et al., 2014).

Human MRE11 is manganese dependent (3’-5’) exonuclease, which efficiently digests dsDNA with blunt or 3’ recessed ends (its activity on ssDNA is much lower, whereas there is nearly no activity on 3’ overhangs), (Paull & Gellert, 1998). Its exonuclease activity is higher in complex with RAD50 (Paull & Gellert, 1998). MRE11 asymmetrically opens DNA hairpin loops, suggesting it has endonuclease activity (Paull & Gellert, 1998). In complex with both RAD50 and NBS1, it efficiently cleaves fully paired hairpins (Paull & Gellert, 1999). This activity is observed only in the presence of NBS1, which also stabilizes MR-DNA complexes and enables partial unwinding of dsDNA by the MRN complex (which is further increased by the addition of ATP), (Paull & Gellert, 1999). Interestingly, the presence of ATP changes the digestion pattern of fully paired hairpin substrates. Further *in vitro* analysis shows that human MRN has sequential endo- and exo- nucleolytic activities on DNA ends containing 5' protein adducts (Deshpande et al., 2016). Here, MRN can target both DNA strands close to these adducts. NBS1 is required for endonucleolytic cleavage close to the site of protein attachment, and it also promotes the exonuclease activity of MRN near the blocked end. However, on clean DNA ends NBS1 inhibits exonucleolytic digestion. According to (Deshpande et al., 2016), MRN endonuclease activity at blocked DNA ends requires ATP and Mn^{2+}, and is stimulated by phosphorylated CtIP and discontinuities in the DNA structure (nicks and gaps). However, experiments performed by (Anand et al., 2016) suggest that MRN cannot perform endonucleolytic cleavage on its own, requiring phosphorylated CtIP for this activity. The human MRN-pCtIP complex preferentially cleaves the 5' strand near the protein adduct, initiating DSBs resection in HR (Anand et al., 2016). Hydrolysis of ATP by RAD50, as well as the presence of both Mg^{2+} and Mn^{2+}, are required for MRE11's endonuclease activity, but not for manganese-dependent exonucleolytic processing. According to (Anand et al., 2016), pCtIP does not have nuclease activity on its own, which contradicts findings by Wang and colleagues (H. Wang et al., 2014). NBS1 senses phosphorylation of CtIP and then directly interacts with MRE11 promoting cleavage by MR (Anand et al., 2019). This mechanism restricts MR resection to S-G2 phases of the cell cycle, when CtIP is extensively phosphorylated. In the absence of NBS1, MR exhibits lower nuclease activity, which is promoted by CtIP (CtIP phosphorylation not required), (Anand et al., 2019). Possibly, this limited resection is sufficient in the G1 phase. Interaction between MR alone and CtIP is increased when CtIP is not phosphorylated (Anand et al., 2019). Although NBS1 promotes cleavage of non-nicked DNA by MR with pCtIP, NBS1 is dispensable for cleavage opposite to a strand discontinuity (Anand et al., 2019). Interestingly, under the same conditions (using a substrate with a pre-existing nick in the opposite strand), both MR and MRN could perform endonucleolytic cleavage, even when pCtIP was not present (Anand et al., 2019).

The inability to detect CtIP endonuclease activity described by (Anand et al., 2016) is in contradiction to (Makharashvili et al., 2014), who showed that CtIP acts as a 5' flap endonuclease on branched DNA structures. According to (Makharashvili et al., 2014), CtIP requires phosphorylation for its catalytic activity and its action is independent of MRN. It exhibits activity in the presence of Mg^{2+}, Co^{2+}, Ni^{2+}, with the preference towards Mn^{2+} (Makharashvili et al., 2014). It cuts within the ssDNA tails of hairpins with both 5' and 3' flaps, not cutting the hairpin itself. It cannot cut substrate resembling a hairpin with a 5' ssDNA overhang (Makharashvili et al., 2014). It may act to remove DNA adducts at DSBs as its mutation results in deficiency in processing IR-induced breaks and TOPO adducts in human cells (Makharashvili et al., 2014).

In vitro experiments showed that MRN can remove Ku adducts from DSBs by cleaving at around ~30 nt from the DNA end (Myler et al., 2017). When both CtIP and DNA-PKcs are present, MRN-CtIP cleaves DNA around 45 nt away from the DNA-PK blocked ends (Ku with DNA-PKcs), (Deshpande et al., 2020). Hydrolysis of ATP by RAD50, as well as the presence of pCtIP, but not its catalytic activity, are required for the efficient digestion of this substrate. The presence of DNA-PKcs in the reaction increases cleavage efficiency by ~ 50-fold. When the reaction was deprived from mix of Mg^{2+} and Mn^{2+}, and only Mg^{2+} was present, the cleavage was observed only when both Ku and DNA-PKcs were included (Deshpande et al., 2020).

MRE11 in processing of DNA hairpin structures

As the majority of *in vitro* experiments on DNA secondary structures were performed only on yeast Mre11, not human, the activity of MRX (the yeast MRN homologue) and Sea2 (yeast CtIP homologue) are also mentioned in this chapter. Research on *Saccharomyces cerevisiae* indicates that the Mre11 complex attacks hairpins with loops shorter than 9 nt (Ait Saada et al., 2021)(Makharashvili et al., 2014). However, upon RPA depletion, spacers longer than 8 nt can become targets for the Mre11 complex (it is unlikely that RPA binds ssDNA loops shorter than 8 nt), (Ait Saada et al., 2021). Biochemical assays performed on a fully paired hairpin, as well as on a hairpin with a 30 nt loop, show that yeast Mre11 can cleave at the distal end of the hairpin loop (relative to the 3' end) and at the junction between dsDNA and 3' ssDNA (Trujillo & Sung, 2001). This yeast Mre11 endonuclease activity is enhanced by Rad50, in the presence of ATP (Trujillo & Sung, 2001). *In vitro* analysis performed by (Lengsfeld, Rattray, Bhaskara, Ghirlando, & Paull, 2007) describes "inefficient but detectable" cutting at the hairpin tip by MR and MRX. According to their research, Sea2 cleaves ssDNA close to the hairpin structure and it cooperates with MRX in digestion of hairpin substrates *in vitro* (Lengsfeld et al., 2007). However, (Cannavo & Cejka, 2014) claim that Sae2 does not exhibit nuclease activity and does not promote MRX nuclease activity on hairpins. Experiments performed using human proteins show that hairpins with different lengths of mismatched nucleotides on the tip were opened by MRE11 at numerous positions near the centre. Fully paired hairpins were cut with over 10-fold lower efficiency (Paull & Gellert, 1998). As the part of MRN complex, MRE11 cuts the fully paired hairpins over 60 times more efficiently than MRE11 alone, and over 25 times more than the MR subcomplex (Paull & Gellert, 1999). Interestingly, the MN subcomplex is more active on this kind of substrate than MR (Paull & Gellert, 1999).

Types of hairpins and importance of their formation is described in more detail in the subchapter (5.3.1).

1.4.3 The role of DNA2 and EXO1 in homologous recombination

The second stage of DSBs resection is long range resection performed by either DNA2 or EXO1 (1.4.1), (Nimonkar et al., 2011). This step is important for DNA repair as sequences located 2 - 3 kb away from the break are preferred for homology search (Nimonkar et al., 2011). Depletion of both DNA2 and EXO1 nucleases together, results in impaired formation of RPA foci indicating defective generation of ssDNA (Gravel, Chapman, Magill, & Jackson, 2008).

DNA replication ATP-dependent helicase/nuclease DNA2

In the BLM-DNA2-RPA-MRN pathway, the BLM helicase and DNA2 nuclease activities are essential (Nimonkar et al., 2011). As DNA2 exhibits bipolar nuclease activity (and weak 5’-3’ helicase activity), its 5’-3’ polarity is enforced by RPA, which is also required for DNA unwinding by BLM. MRN stimulates this resection by recruiting BLM to DNA ends. CtIP also enhances resection as it stimulates BLM helicase activity and upregulates DNA2-dependent endonucleolytic cleavage (Daley et al., 2017). However, it does not change the incision site. In the presence of RPA, stimulation of DNA2 by CtIP is lower as RPA and CtIP may compete for the access to the substrate. In human cells, BLM promotes the resection as a part of BLM-TOPOIIIα-RMI1-RMI2 (BTRR) complex (Sturzenegger et al., 2014). BTRR catalyses dissolution of double Holliday junction in the later steps of HR.

BLM is not the only helicase from the RecQ family that can cooperate with DNA2 to mediate long range resection (Sturzenegger et al., 2014). WRN (3’-5’) helicase/exonuclease, together with DNA2 and RPA, can also take part in digestion of 5’ recessed DNA ends. The efficiency of this digestion *in vitro* is even higher when BLM is replaced with WRN. Both BLM (Karow, Chakraverty, & Hickson, 1997) and WRN (Brosh, Waheed, & Sommers, 2002) helicases are loaded on DNA with 3’ ssDNA tails and unwind it in the 3’-5’ direction. In the WRN-DNA2-RPA pathway, resection is dependent on the endonuclease activity of DNA2 and helicase activity of WRN as was validated by generation of separation of function mutants (Sturzenegger et al., 2014).

DNA2 acts as an endonuclease that efficiently cleaves ssDNA (from both ends) and ssDNA within a 5’ flap (Zheng, Meng, Campbell, & Shen, 2020). It localizes to the telomeres of mammalian cells, where it may facilitate resolution of telomeric G-quadruplexes (G4) and the progression of DNA replication. DNA2 cleaves telomeric G4s *in vitro* (Lin et al., 2013). Experiments using different types of substrates show that DNA2 can specifically cleave G4s within the DNA bubble or within the single-stranded 5’ flap, but not a substrate with a G4 opposite to the gapped strand. This cleavage depends on DNA2’s nuclease activity, while its helicase activity is not required (Lin et al., 2013).

Exonuclease 1

In the EXO1-BLM-RPA-MRN pathway, EXO1 can inefficiently digest DNA (5’-3’) on its own. BLM increases its affinity to DNA ends, whereas MRN and RPA stimulate EXO1-dependent resection (Nimonkar et al., 2011). The helicase activity of BLM is not required for EXO1 stimulation, which is mediated by the interaction between these two proteins (Nimonkar, Özsoy, Genschel, Modrich, & Kowalczykowski, 2008). RPA interacts with BLM and stimulates DNA unwinding.

EXO1 is a (5’-3’) exonuclease, which exhibits also 5’ flap endonuclease and RNase H activities (B. I. Lee & Wilson, 1999). Its exonuclease activity is higher on blunt ended dsDNA and dsDNA with a 3’ overhang, than on ssDNA or dsDNA with a 5’ overhang (B. I. Lee & Wilson, 1999). When the length of the 3’ overhang was increased from 4 nt to ~ 100 nt, the yeast Exo1 activity was increased further (Y. Li, Shen, & Niu, 2019). Processing of ssDNA is sequence-specific, as poly(dT)oligonucleotides were not digested (Y. Li et al., 2019). Formation of hairpins promote binding of yeast Exo1 to ssDNA (Y. Li et al., 2019). Here, hairpins with a stem of 4 bp (or longer) can activate digestion of ssDNA harbouring poly(dT) sequences. Exo1 nuclease activity on ssDNA is strongly inhibited by RPA (Y. Li et al., 2019). Exo1 also takes part in mismatch repair, where it can remove mispaired nucleotides (Goellner, Putnam, & Kolodner, 2015).

1.4.4 The role of RPA and RAD51 in homologous recombination

RPA is a single stranded DNA binding complex, consisting of three subunits: RPA70, RPA32 and RPA14 (Yates et al., 2018). RPA70 possesses four DNA binding domains

(DBD-A, B, C, F), whereas RPA32 and RPA14 possess single DBD each (DBD-D and DBD-E). RPA binding to ssDNA is performed mainly by DBD-A, -B, -C and -D. RPA can associate with ssDNA either in a low affinity mode, binding 8 - 12 nt through DBD-A and -B, or in a high affinity mode binding 28 nt through four major DBDs (Yates et al., 2018). Due to its high affinity towards ssDNA, RPA is often used as a marker of DNA end resection (Chen, Lisby, & Symington, 2013). *In vivo* it can associate with ssDNA of many kilobases (Yates et al., 2018), generated during HR. As described in 1.4.3, in HR RPA promotes DNA unwinding and long range resection in the 5' - 3' direction (Chen et al., 2013), (Figure 9). RPA also prevents the formation of DNA hairpins, which can be generated due to the folding of long stretches of ssDNA.

In response to DNA damage, RPA bound to ssDNA can be phosphorylated by ATR, ATM, CDK and DNA-PKcs (Soniat, Myler, Kuo, Paull, & Finkelstein, 2019). This can be used as a readout of DSB resection. Phosphorylation of RPA (pRPA) inhibits DNA resection both in cells and *in vitro* (Soniat et al., 2019). It also leads to abrogation or changes in the interaction between RPA and BLM, affecting BLM activity (Soniat et al., 2019).

During HR, two homologous DNA strands align in a process called synapsis and HR can be divided into presynaptic (formation of Rad51-nucleoprotein filaments) and postsynaptic (extension of paired nucleotides) phases (Eggler, Inman, & Cox, 2002). As many experiments were performed on yeast Rad51, lowercase letters are used to enable distinction. RAD51 forms nucleoprotein filament on ssDNA and promotes extensive pairing of long homologous DNA stretches (Sigurdsson, Van Komen, Petukhova, & Sung, 2002). Rad51 - ssDNA nucleoprotein filament, also called the presynaptic filament, has dsDNA binding site and can hold two DNA molecules in close proximity (Sung et al., 2003). RPA is important in Rad51-mediated DNA strand exchange in both presynaptic (Baumann & West, 1997) and the postsynaptic manner (Eggler et al., 2002). First, its role in the removal of secondary DNA structures facilitates formation of complete Rad51 filaments (Baumann & West, 1997). Then, in the postsynaptic phase, RPA stabilizes DNA pairing (Eggler et al., 2002). However, RPA also has a negative effect on recombination as its competes with Rad51 for DNA binding and this leads to inhibition of the presynaptic filament assembly (Sung et al., 2003). This inhibition is overcome by the recombination-mediator BRCA2, which binds both RAD51 and ssDNA. The RAD51 nucleoprotein filament is stabilized by BRCA2 (Esashi, Galkin, Yu, Egelman, & West, 2007). Direct interaction between RAD51 and BRCA2 regulates dynamics of the filament and stimulates HR.

1.4.5 The role of BRCA proteins in homologous recombination

BRACA1 forms dimer with BARD1, which is essential for their stability *in vivo* (Prakash, Zhang, Feng, & Jasin, 2015). As a part of phosphorylation cascades, BRACA1-BARD1 take part in activation of checkpoints at different stages of the cell cycle (G1/S, S-phase, G2/M), (Roy, Chun, & Powell, 2012). BRACA1's role in HR involves two steps: i) DNA end resection and ii) loading of RAD51 on ssDNA (Prakash et al., 2015). After DNA damage, BRCA1 colocalizes with MRN and interacts directly with CtIP. BRCA1 promotes resection by blocking 53BP1 from damaged DNA (Chapman, Sossick, Boulton, & Jackson, 2012). In the absence of BRCA1, 53BP1 localizes to DSBs supressing end resection and HR.

Mutations of either BRCA1 or BRCA2 genes account for majority of familial breast and ovarian cancers and increase risk of developing other malignancies (Powell & Kachnic, 2003). As described in (1.4.4), BRCA2 stimulates RAD51-dependent DNA strand exchange during HR (Holloman, 2011). It mediates loading of RAD51 on RPA coated ssDNA and stabilizes RAD51 nucleoprotein filaments. BRCA1 promotes recruitment of BRCA2 to DNA damage sites (Prakash et al., 2015), as both proteins interact *via* a bridging protein (PALB2/FANCN), (F. Zhang et al., 2009). Disruption of BRCA1 impairs PALB2, BRCA2 and RAD51 focus formation *in vivo*, whereas disruption of BRCA2 reduces only RAD51 foci (Prakash et al., 2015). In PALB2-deficient Fanconi anemia cells, interaction between BRCA1 and BRCA2 is abolished (F. Zhang et al., 2009).

1.4.6 Resolution of Holliday junctions (HJ)

During HR, strand exchange leads to the formation of a four-way DNA intermediate called Holliday junction (HJs), which usually comprise covalently linked sister chromatids (Punatar, Martin, Wyatt, Chan, & West, 2017). To ensure proper chromosome segregation, HJs can be dissolved by BLM-TOP3α-RMI1-RMI2 (BTR complex) or resolved by structure specific nucleases, called HJ resolvases (Figure 11). There are two pathways of resolution, which are mediated either by the SLX-MUS complex or GEN1 (Punatar et al., 2017). Depletion of proteins engaged in either of these pathways results in defects in chromosome segregation and proliferation (Wyatt, Sarbajna, Matos, & West, 2013). BTR-mediated dissolution removes most of the HJs during normal cell growth and results in non-crossover products (Wyatt et al., 2013). Increased sister chromatid exchanges (SCEs) observed after DNA damage indicate that BTR cannot dissolve high numbers of HJs (Wyatt et al., 2013). However, endonuclease-mediated resolution mechanisms can generate both crossover (CO) and non-crossover products (NCO), (Wechsler, Newman, & West, 2011).

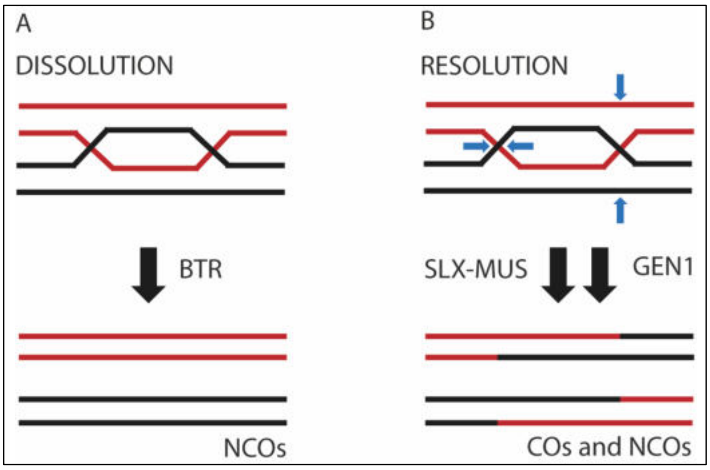


Figure 11. A) Topological dissolution and B) nucleolytic resolution of dHJs, (based on Wyatt et al., 2013). Topological dissolution creates non-crossover products (NCOs), whereas nucleolytic resolution generates both cross-over (COs) and non-crossover products. Blue arrows indicate position of the cleavage.

In dHJ dissolution, two HJs are branch migrated towards each other until they fuse forming hemicatenated intermediate (Bizard & Hickson, 2014). Decatenation of this intermediate by the BTR complex restores DNA structure.

HJ resolution can be performed by the SLX1-SLX4-MUS81-EME1 complex (SLX-MUS), stimulated by XPF-ERCC1 (Wyatt et al., 2013), (Wyatt, Laister, Martin, Arrowsmith, & West, 2017). MUS81-EME1 is a 3’ flap endonuclease, which cannot efficiently cleave HJs (Wyatt et al., 2013). MUS81-EME1 phosphorylation stimulates its interaction with SLX1-SLX4 endonuclease, which introduce the first nick at the junction (Wyatt et al., 2013), (Punatar et al., 2017). As MUS81-EME1 has a preference towards nicked HJs, it performs the second cleavage. SLX-MUS association is increased at the G2/M stage of the cell cycle (Wyatt et al., 2013). *In vitro* experiments showed that SLX1-SLX4 cleaves replication forks, 5’ and 3’ flaps, nicked HJs, and intact mobile and immobile HJs (Wyatt et al., 2013). From this group of substrates, only replication forks, 3’-flaps and nicked HJs were efficiently cleaved by MUS81-EME1 (Wyatt et al., 2013). SLX4 interacts also with XPF-ERCC1, which is nucleotide excision repair endonuclease (Wyatt et al., 2017). SLX-MUS together with XPF-ERCC1 form the SMX tri-nuclease, which cleaves a variety of branched DNA structures (replication forks, 5’ and 3’ flaps, nicked HJ, intact HJ), (Wyatt et al., 2017). The SLX4 scaffold coordinates the action of SLX1, MUS81 and EME1 in HJ resolution and this reaction is stimulated by XPF-ERCC1. SMX is more active than the three separate nucleases (Wyatt et al., 2017). SLX4 interacts directly also with Mutsβ, which is a component of DNA mismatch repair pathway (Young et al., 2020).

Mutsβ binds HJs and stimulates their resolution by the SMX complex. Loss of Mutsβ activity results in increased numbers of HR-associated ultrafine bridges (HR-UFBs) in mitosis (Young et al., 2020). Anaphase HR-UFBs are normally cleaved by GEN1 or SMX (Young et al., 2020).

Experiments performed *in vitro* shown that GEN1, a 5' flap endonuclease, efficiently cleaves both single and double HJs (Punatar et al., 2017). It has a weak sequence preference for cleavage between two G residues in a T-rich region. GEN1 performs symmetrical incisions across the junction (Punatar et al., 2017).

1.5 DNA replication

1.5.1 Main steps of DNA replication

DNA is replicated by the a multi-protein complex called the replisome, which has DNA helicase, primase and DNA polymerase activities (Berti, Cortez, & Lopes, 2020). DNA is synthetised in a 5' – 3' direction in a continuous manner on the leading strand, in the same direction as the moving replication fork. Discontinuous synthesis of Okazaki fragments (150 – 200 nt) on the lagging strand is performed in the opposite direction to the replication fork movement (B. Liu, Hu, Wang, & Kong, 2017).

Active CMG helicase, which consists of CDC45, MCM2-7 and GINS, unwinds the DNA (Burgers & Kunkel, 2017). Synthesis on both leading and lagging strand is initiated by DNA polymerase α (Pol α) – primase, which generates short RNA-DNA primers (Zhou, Lujan, Burkholder, Garbacz, & Kunkel, 2019). The majority of synthesis on the leading strand is performed by DNA polymerase ε (Pol ε), while DNA polymerase δ (Pol δ) performs most of the lagging strand synthesis (Zhou et al., 2019). Polymerase δ (Pol δ) participates also in initiation and termination of leading strand replication (Zhou et al., 2019). After DNA synthesis, RNA primers need to be removed to enable joining of Okazaki fragments by DNA ligase I (Lig I). RNA – DNA primers can be displaced by Pol δ to generate flap structures, which are cleaved by Fen1 flap endonuclease and Dna2 (B. Liu et al., 2017). The flap cleavage pathway can be either short, when Fen1 directly removes the primers, or long when Dna2 cleaves long RPA – coated flaps subsequently cleaved by Fen1. Alternatively, primers can be digested by RNAse H2 and Exo1. After the primers are removed, DNA gaps are filled by Pol δ and sealed by Lig I.

1.5.2 The cellular response to DNA replication stress

Although DNA replication is generally highly accurate, stability of replication forks is often challenged by DNA lesions and secondary structures (Berti et al., 2020). When DNA lesions stall the leading strand polymerase, helicase-polymerase uncoupling of the replisome may occur (Berti et al., 2020). Termination of leading strand synthesis, while the replisome is still moving and the lagging strand is still being synthetised, results in accumulation of ssDNA, which becomes coated by RPA. ssDNA activates ATR, which directs multiple paths of the replication stress response (Saldivar, Cortez, & Cimprich, 2017). Uncoupled stalled replication forks may be restarted through processes of fork repriming, fork reversal or fork incision (Figure 12), (Berti et al., 2020).

Fork repriming starts with synthesis of short RNA fragments, which are required to restart DNA synthesis (Berti et al., 2020). hPrimPol1 is a human DNA primase-polymerase, which is recruited to stalled replication forks in an RPA dependent manner (Wan et al., 2013). Repriming by PrimPol1 is followed by post replicative gap filling in a template switching process or by translesion synthesis polymerases (TLS), (Berti et al., 2020). In the template switching process, when the parental strand is damaged, the newly synthetised strand is used as a template for DNA synthesis.

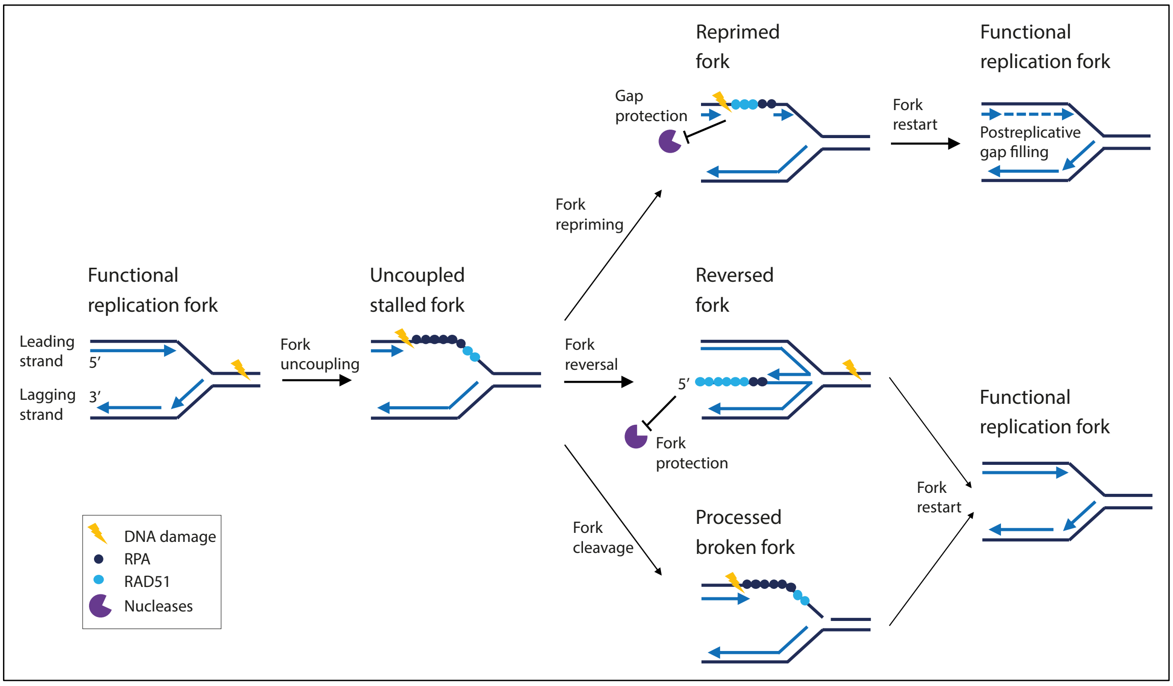


Figure 12. Mechanisms of replication fork plasticity, (based on Berti et al., 2020). Upon replication stress, an active fork stalls. The stalled fork then undergoes fork repriming, fork reversal or fork cleavage.

Another response to replication stress is the formation of reversed replication forks (Zellweger et al., 2015). Replication fork reversal is a transformation of a typical replication fork into a four-way junction, also called a regressed fork or chicken-foot structure (Neelsen & Lopes, 2015). The main steps of the process are unwinding of newly synthetised DNA strands, reannealing of the parental strands and annealing of the nascent strands (Figure 13), (Neelsen & Lopes, 2015). The new, fourth arm is called the regressed arm.

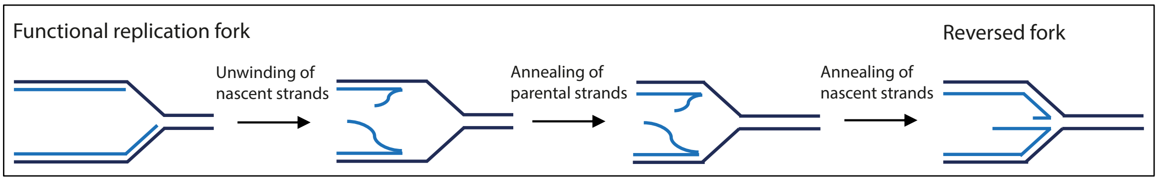


Figure 13. Replication fork reversal, (based on Neelsen & Lopes, 2015).

Fork reversal prevents further DNA synthesis across the damaged template and preserves genome stability allowing more time for DNA repair (Neelsen & Lopes, 2015). However, when the four-way junction is cleaved in an unscheduled way, fork reversal may lead to increased instability (Neelsen & Lopes, 2015).

Replication fork reversal is regulated by replication protein A (Bhat & Cortez, 2018). RPA can diffuse along ssDNA and destabilize secondary structures like hairpins (Nguyen et al., 2014). RPA-coated ssDNA binds to ATR-interacting protein (ATRIP), (Zou & Elledge, 2003), allowing activation of ATR pathways. RPA interacts directly with SWI/SNF-related matrix-associated actin dependent regulator of chromatin subfamily A-like protein 1 (SMARCAL1) and regulates its activity on replication forks (Bétous et al., 2013). When bound to the lagging template strand of an unchallenged fork, it inhibits fork reversal by SMARCAL1 (Figure 14). When bound to the leading template strand of the stalled fork, it stimulates SMARCAL1-mediated fork reversal (Bhat & Cortez, 2018). During replication fork reversal, as parental strands reanneal, RPA has to be removed from the parental ssDNA (Bhat & Cortez, 2018)

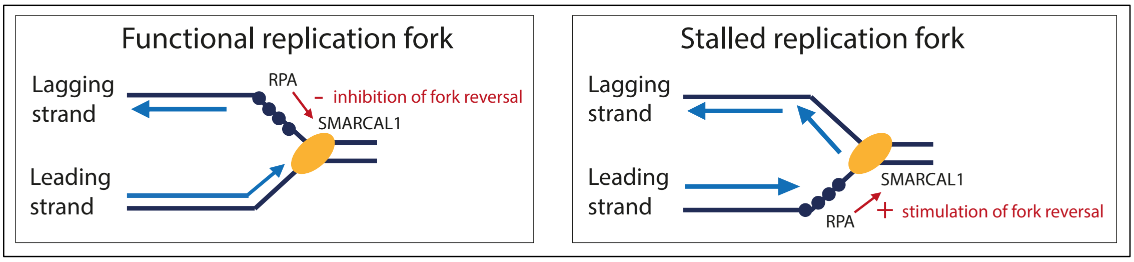


Figure 14. RPA binding can either stimulate or inhibit SMARCAL1-dependent fork reversal, (based on Bhat & Cortez, 2018).

RPA coated ssDNA is also required to recruit the Bloom syndrome helicase (BLM) to sites of DNA replication stress (Shorrocks et al., 2021). BLM, as a part of TOP3A -RMI1-RMI2 complex (BTR) promotes restart of stalled replication forks, though the mechanism of how this is achieved is not well understood. It has been suggested that it may remodel the fork promoting its regression or it can unwind aberrant DNA structures, including G-quadruplexes (Shorrocks et al., 2021).

Another protein that is required for fork reversal is RAD51 (Zellweger et al., 2015), a recombinase and DNA dependent ATPase that binds both ss and dsDNA (Bhat & Cortez, 2018). RAD51 mediated fork reversal is independent of BRCA2 (Mijic et al., 2017). However, to limit resection of reversed forks, BRCA2 loads RAD51 on the regressed arms (Mijic et al., 2017). RAD51 nucleofilaments protect DNA from extensive nucleolytic degradation by MRE11 (Figure 12), (Kolinjivadi, Sannino, De Antoni, et al., 2017). In the absence of BRCA proteins, unprotected reversed replication forks are resected by MRE11 and EXO1 nucleases (Lemaçon et al., 2017). CtIP initiates the activity of MRE11 on unprotected regressed arms, and DNA degradation is further extended by EXO1 (Lemaçon et al., 2017). In the absence of BRCA2, the partially resected reversed fork, may be cleaved by MUS81 (Lemaçon et al., 2017).

The reversed fork can be restored to a typical, three-way structure in a fork restart process (Neelsen & Lopes, 2015). Restart of replication forks, reversed by the action of TOP1 inhibitors, is promoted by human RECQ1 helicase (Figure 15), (Berti et al., 2013). The activity of RECQ1 is regulated by poly(ADP-ribose) polymerase 1 (PARP1), which promotes accumulation of regressed forks limiting their premature restart (Berti et al., 2013). Inhibition of PARP1, after TOP1 inhibition, results in increased DSB formation (Berti et al., 2013).

The second mechanism of reversed fork restart involves the activities of the WRN ATPase and DNA2 nuclease (Figure 15), (Thangavel et al., 2015). Here, WRN opens the reversed arm which are then resected by DNA2. The ssDNA overhang of the regressed arm may promote formation of HJs lading to fork restart (Thangavel et al., 2015). The binding of RECQ1 to reversed forks may limit DNA2 activity, protecting replication forks from an extensive degradation (Thangavel et al., 2015). CtIP, MRE11 and EXO1 are not involved in the DNA2-dependent reversed fork processing pathway (Thangavel et al., 2015). *In vitro* experiments showed that DNA2 degrades four-way junctions (a substrate mimicking a reversed replication fork) more efficiently than dsDNA (Thangavel et al., 2015). Here, the presence of RPA stimulated DNA2 activity *in vitro*. DNA2 promotes genome stability maintenance, as it not only processes stalled replication forks, but also counteracts replication stress by resolving secondary DNA structures, including G-quadruplexes (Lin et al., 2013).

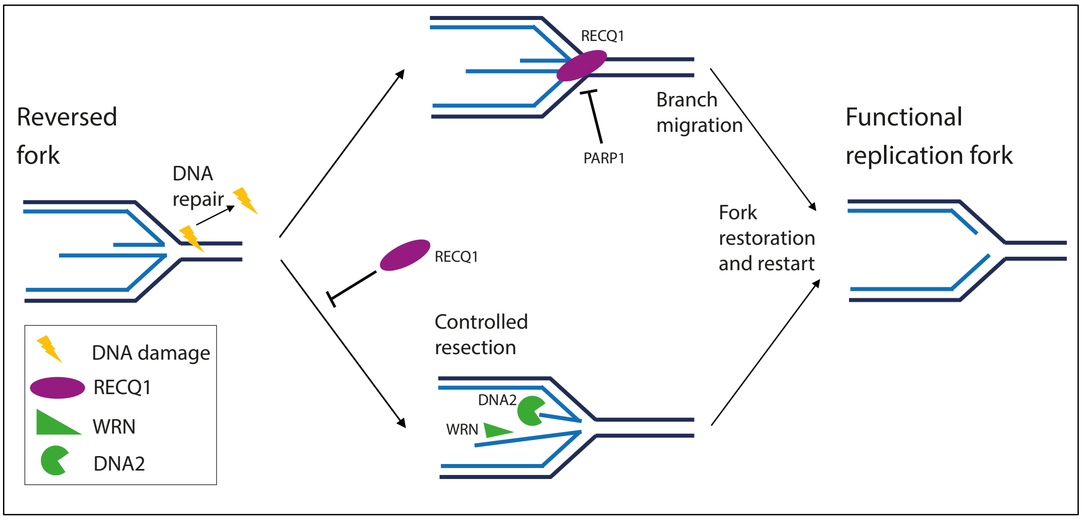


Figure 15. Restoration of the reversed forks, based on (Berti et al., 2020) and (Neelsen & Lopes, 2015). Restart of the fork is promoted by a RECQ1- mediated mechanism or by controlled resection by the WRN and DNA2 nucleases.

Another response to replication stress is the incision of stalled fork junction and generation of broken replication fork, a single-ended DSB (seDSB), (Berti et al., 2020). According to (Di Marco et al., 2017), MUS81-EME1 endonuclease cooperates with RECQ5 to cleave late replication intermediates at early mitosis. As RAD51 nucleoprotein filaments at stalled replication forks are stabilized by BRCA and Fanconi anemia (FA) proteins, DNA is protected from degradation (Figure 16), (Di Marco et al., 2017). RECQ5 removes RAD51 from stalled replication forks, enabling MUS81-EME1 access to the DNA and stimulating their crossover junction endonuclease activity (Di Marco et al., 2017). Cleavage of late replication intermediates by MUS81-EME1 during early mitosis promotes DNA-repair synthesis at common fragile sites (CFSs) and proper chromosome segregation (Di Marco et al., 2017).

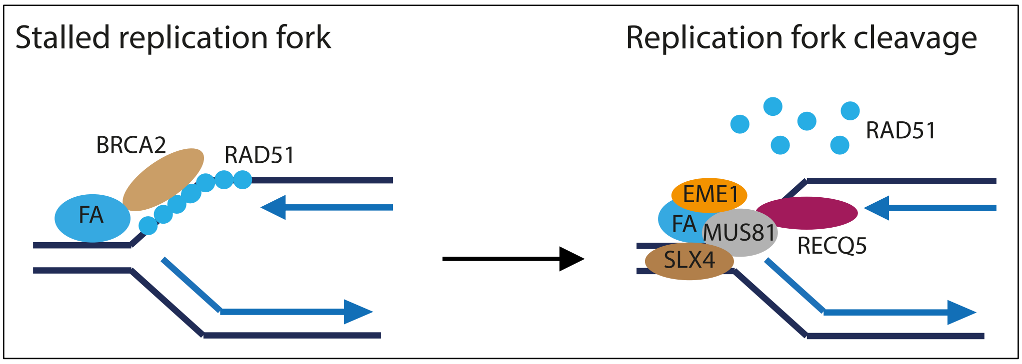


Figure 16. Cleavage of replication forks, (based on Di Marco et al., 2017). RAD51, stabilized by BRCA2 and FA, protects DNA from degradation. RECQ5 removes RAD51 from stalled replication forks enabling fork cleavage by MUS81-EME1.

According to (Pepe & West, 2014), MUS81-EME2 endonuclease cleaves stalled replication forks enabling their restart. This MUS81-EME2 activity is restricted to S-phase, whereas MUS81-EME1 functions later in the cell cycle to cleave recombination intermediates (Pepe & West, 2014).

Broken replication forks can restart through a break induced replication (BIR) mechanism, which depends on action of the PIF1 helicase and DNA polymerase delta 3 accessory subunit (POLD3), (Berti et al., 2020). RAD52 facilitates this repair (Sotiriou et al., 2016). BIR starts with resection of one-ended DSB in the 5'-3' direction (Figure 17), (Kockler, Osia, Lee, Musmaker, & Malkova, 2021). This 3' overhang invades homologous template creating a D - loop. DNA synthesis progresses through migrating bubble which may extend till the end of the chromosome (L. Liu & Malkova, 2022). BIR results in conservative inheritance of newly synthetised DNA leading to increased genetic instability (L. Liu & Malkova, 2022). It has been estimated that DNA synthesis through BIR generates between 100 - 1000 times more mutations than S – phase replication (L. Liu & Malkova, 2022).

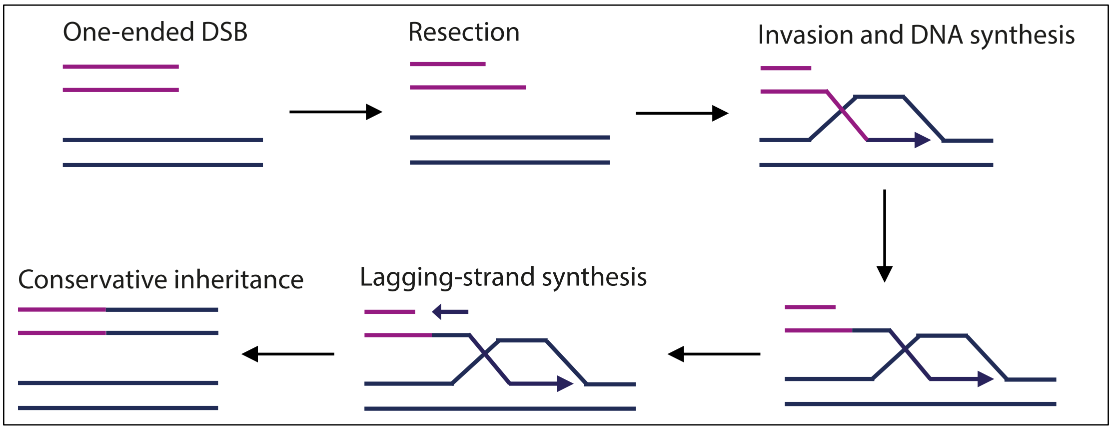


Figure 17. Mechanism of break induced replication pathway (BIR), (based on Kockler et al., 2021).

Another way to restore functionality of the broken fork is RAD52 mediated reannealing of the broken strand and its re-ligation by DNA ligase 4 (LIG4) – XRCC4 (Berti et al., 2020), (Chappidi et al., 2020). This cleavage - re-ligation process allows restart of the fork only during transcription - replication conflicts (TRCs) and enables semiconservative DNA synthesis.

1.5.3 DNA replication and telomeres

Telomeres are repetitive DNA sequences associated with the Shelterin protein complex (TRF1, TRF2, RAP1, TIN2, TPP1, POT1) at the end of eukaryotic chromosomes, which protect telomeric DNA against degradation and fusion (Gilson & Géli, 2007). In humans, telomeres are composed of G-rich repeats of $(TTAGGG)_n$ and 3' ssDNA overhang (Gilson & Géli, 2007). To prevent telomeres from being recognized as DSBs requiring repair, the telomeric ssDNA overhang invades telomeric dsDNA leading to T-loop formation, which is bound to and protected by the Shelterin complex (Shay & Wright, 2019).

Replication of telomeres is challenging as forks often stall within the telomeric repeats (Lin et al., 2013). If the replication is terminated prematurely, it may result in rapid loss of telomeric sequences (Lin et al., 2013). To increase efficiency of the replication, multiple factors are engaged. Shelterin complex recruits several helicases (including BLM, WRN, RTEL1) to recognize and unwind telomeric G-quadruplexes (Lu & Pickett, 2022). RECQL4, together with Shelterin proteins, resolves telomeric D-loops to enable replication (Ghosh et al., 2012). Both BRCA1 and BRCA2 facilitate replication of telomeric regions abundant in G-quadruplexes (G4), (Zimmer et al., 2016). BRCA2 binding to telomeric G4s enables RAD51-mediated restart of G4-stalled forks (Junyeop Lee et al., 2022). By interaction with telomeric G4s, BRCA2 protects telomeres from MRE11-mediated degradation during replication (Junyeop Lee et al., 2022).

In normal cells, telomeres shorten with every cell division due to incomplete DNA synthesis (the so-called "end replication problem"), (Shay & Wright, 2019). After a critical number of cell divisions, the presence of very short, uncapped telomeres results in DNA damage signalling, permanent arrest of the cell cycle (cellular senescence) or apoptosis (Gilson & Géli, 2007). To overcome this, cancer cells activate telomere maintenance pathways. They can be maintained by expression of telomerase - reverse transcriptase that extends the 3' ends of chromosomes in 85 – 90 % of all human cancers, or by alternative lengthening of telomeres pathway (ALT) present in 10 – 15 % of cancers (Shay & Wright, 2019).

The ALT pathway is initiated from DNA breaks at the telomere, likely formed from collapsed replication forks (J. M. Zhang & Zou, 2020). ALT is carried out in ALT - associated promyelocytic leukemia (PML) bodies (APBs). APBs are nuclear structures, which require BLM for their formation (J. M. Zhang & Zou, 2020). They contain telomeric DNA and proteins engaged in DNA replication and repair, creating favourable environment for DNA recombination and synthesis. Liquid – liquid phase separation (LLPS) leads to clustering of telomeres, driving ALT (Figure 18), (Min, Wright, & Shay, 2019). In APBs, BIR is induced by one ended DSBs and is mediated through RAD52-dependent (BIR1) or independent (BIR2) mechanisms (J. M. Zhang, Yadav, Ouyang, Lan, & Zou, 2019). RAD52 can anneal DNA and stimulates telomeric D-loop formation, in the presence and absence of RPA (J. M. Zhang et al., 2019). RAD51 cannot promote D-loop formation on telomeric DNA when RPA is present (J. M. Zhang et al., 2019).

Contrary to the RAD52 dependent pathway, the RAD52-independent pathway is responsible for the formation of extrachromosomal single stranded C-rich telomeric DNA circles (C-circles). Their level is increased in the absence of RAD51 and MRE11 (J. M. Zhang et al., 2019). As C-circles are present in ALT cells expressing RAD52, it suggests that both BIR pathways can act simultaneously (J. M. Zhang et al., 2019). The conservative ALT DNA synthesis is mediated by POL3 and POL4 (J. M. Zhang & Zou, 2020).

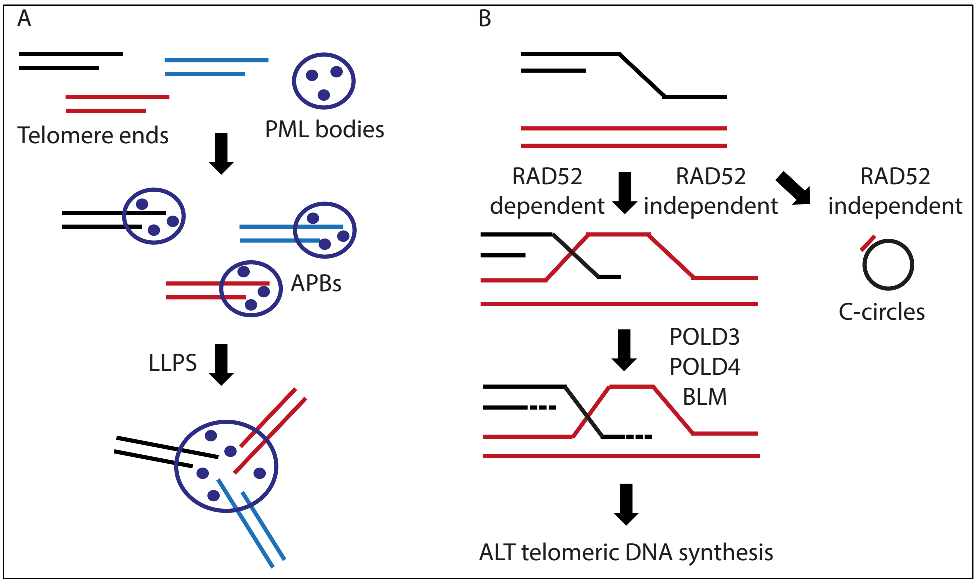


Figure 18. Alternative lengthening of telomeres, (based on J. M. Zhang & Zou, 2020). A) Clustering of telomeres drives ALT. B) RAD52 dependent and independent BIR pathways. Conservative BIR synthesis is dependent on POLD3/POD4 and promoted by BLM. RAD52-independent BIR leads to conservative synthesis and is associated with C-circle formation.

1.6 Exonuclease 3’-5’ domain-containing protein 2 (EXD2)

1.6.1 EXD2 as a component of the homologous recombination pathway

Exonucleases are enzymes which catalyse excision of nucleoside monophosphates from either the 3’- or 5’- nucleic acid end (Shevelev & Hübscher, 2002). They play important roles in DNA metabolism, including replication and repair.

Exonuclease 3’-5’ domain-containing protein 2 (EXD2, EXDL2) is a recently identified nuclease which promotes HR by facilitating DNA end resection (Broderick et al., 2016). Here, the authors demonstrated that EXD2 interacts with DSB repair factors, including CtIP, MRE11 and BRCA1, and it is recruited to damaged DNA with fast kinetics. It is required for the response to DSBs inducing agents, as its depletion sensitizes cells to IR, camptothecin (CPT, inhibitor of TOPO1) and phleomycin (Broderick et al., 2016). EXD2’s nuclease activity promotes DNA end resection, the generation of ssDNA and is required for efficient RPA loading in DSB repair (Broderick et al., 2016). Deficient RPA loading results in reduced formation of RAD51 foci. Depletion of EXD2 affects short range resection, but not initial sensing of the break, as DNA damage responses (phosphorylation of histone H2AX and CHK2) are intact (Broderick et al., 2016). EXD2 depletion leads to increased chromosomal aberrations and sensitization of cells to the PARP inhibitor Olaparib (Broderick et al., 2016).

1.6.2 EXD2 as a part of homologous recombination resection machinery

The HR pathway is initiated by DSB resection, performed by the MRN complex and CtIP (Stracker & Petrini, 2011). Analysis of RPA and RAD51 focus formation shows that combined depletion of EXD2 and MRE11, results in comparable resection inhibition as for cells depleted in MRE11 alone (Broderick et al., 2016). Depletion of EXD2 alone leads to a weaker resection defect than depletion of MRE11 alone, suggesting that EXD2 acts downstream in the resection pathway (Broderick et al., 2016). MRE11 may create multiple nicks on the resected strand (Garcia, Phelps, Gray, & Neale, 2011) that possibly can act as an entry sites for EXD2, as *in vitro* experiments confirm high exonuclease activity of EXD2 on a nicked and a gapped dsDNA (Broderick et al., 2016). Importantly, inhibition of MRE11 endonuclease activity leads to a robust resection defect in cells, whereas inhibition of its exonuclease activity results in a milder phenotype (Broderick et al., 2016), (Shibata et al., 2014). Depletion of EXD2 alone has a stronger resection defect than inhibition of MRE11 exonuclease activity alone, while EXD2 depletion in the presence of a MRE11 endonuclease inhibitor did not decrease the resection further than inhibitor alone (Broderick et al., 2016). All of this is in line with the suggestion that EXD2 acts downstream of MRE11 endonuclease activity. *In vitro* studies confirm that MRE11 and EXD2 functionally interact, as their combination resulted in higher ssDNA degradation than observed for MRN alone (Broderick et al., 2016). However, it is unclear why two nucleases with the same polarity would be required for the initial processing of the DNA DSBs and whether they synergise or antagonise each other depending on the cell type and/or DNA DSBs contexts.

1.6.3 The role of EXD2 in other DNA damage repair pathways

Experiments performed on IR treated cells in G1 show that EXD2-depleted cells have a milder phenotype (% of unrepaired DSBs) than cells deficient in LIG4 and MRE11, contributing to around 39 % of NHEJ events during 6 h repair time (69 % for MRE11), (Akagawa et al., 2020). Hence, EXD2 may take part in DSBs repair in preparation for NHEJ.

A model described by (Shibata et al., 2017) suggests that in G1, EXD2 takes part in resection dependent canonical NHEJ. Here, silencing of EXD2 (siEXD2) reduced the number of end joining events in the reporter assay and decreased amount of pRPA foci. EXD2 depletion partially rescued the repair defect of Artemis-deficient cells (number of γH2AX foci after silencing of EXD2 decreased), (Shibata et al., 2017). Therefore, they proposed that EXD2 together with MRE11 exonuclease activity and EXO1, perform the resection of DSBs before the endonuclease activity of Artemis completes the process.

1.6.4 The role of EXD2 in DNA replication

The replication of genetic material relies on replication forks, whose action is challenged when the DNA template is damaged or is difficult to replicate (Kolinjivadi, Sannino, de Antoni, et al., 2017). EXD2 is one of the key factors of the replication stress response and it is recruited to stalled replication forks with fast kinetics (Nieminuszczy et al., 2019). EXD2 prevents excessive fork reversal, protecting the stalled forks from over resection (Nieminuszczy et al., 2019). Its deficiency results in unscheduled fork reversal, degradation, and collapse, meaning complete loss of replication ability (Figure 19), (Nieminuszczy et al., 2019). Silencing of SMARCAL1, in cells lacking EXD2, limits forks reversal and promotes fork restart (Nieminuszczy et al., 2019). Incubation of EXD2 deficient cells with the PARP inhibitor Olaparib, counteracts RECQ1's inhibition by parylation, leading to efficient fork restart (Nieminuszczy et al., 2019). Reactions with hEXD2 purified from bacteria, show that EXD2 digests fork-like structures *in vitro*, suggesting that this resection promotes fork restart (Nieminuszczy et al., 2019).

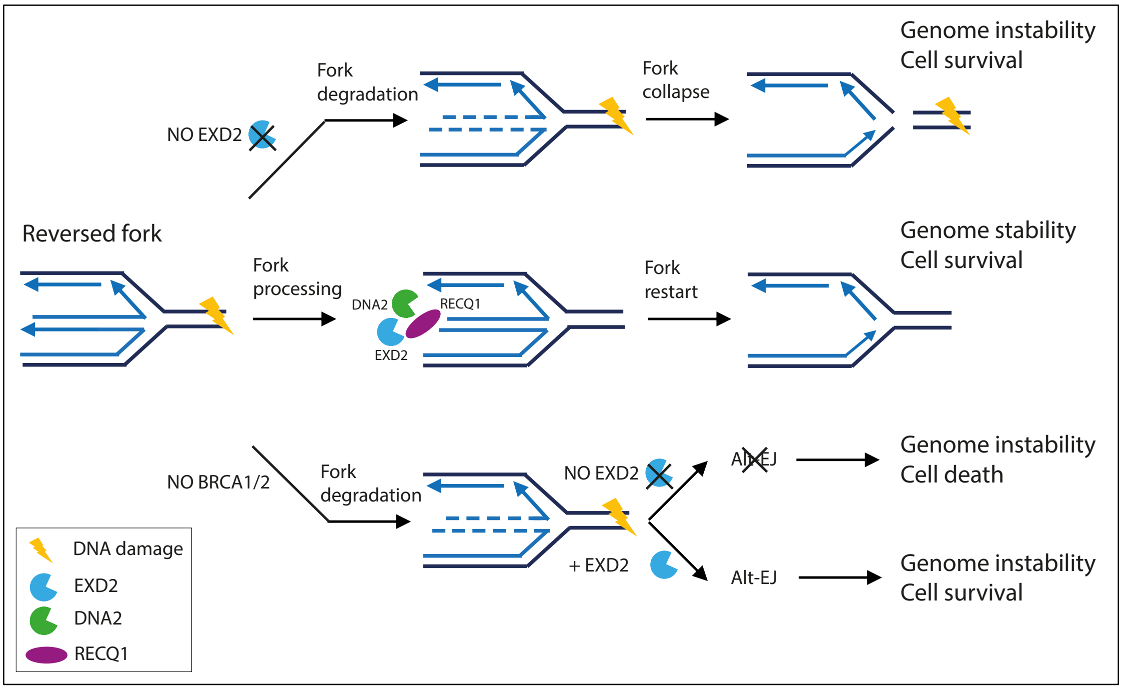


Figure 19. EXD2 promotes restart of reversed replication forks, (based on Nieminuszczy et al., 2019). Combined depletion of EXD2 and BRCA1/2 leads to genome instability due to the absence of functional HR and Alt-EJ pathways.

EXD2 deficiency in cells, results in hypersensitivity to replication inhibitors and in genome instability (Nieminuszczy et al., 2019). Formation of anaphase bridges suggests that DNA is under-replicated in EXD2-deficient cells and loss of EXD2 is synthetic lethal with BRCA1/2, proteins which also perform fork protection functions (Nieminuszczy et al., 2019). In the deficiency of BRCA1/2, chromosomal breaks are repaired via alternative end-joining pathway (Alt-EJ). Depletion of EXD2 leads to decreased amount of chromosome end-to-end fusion, suggesting a defect in Alt-EJ pathway (Nieminuszczy et al., 2019). When both EXD2 and BRCA1/2 are depleted, replication forks cannot be rescued by HR or alt-EJ, leading to cell death (Nieminuszczy et al., 2019)

DNA replication forks can also stall at DNA crosslinks. Depletion of EXD2 results in sensitivity to mitomycin C (MMC), which indicates that EXD2 may also have a role in the repair of inter-strand DNA crosslinks (ICL), (Smogorzewska et al., 2010).

1.6.5 The role of EXD2 in DNA transcription

DNA transcription is challenged by genotoxic stress which leads to arrest of gene expression followed by DNA repair and recovery of mRNA synthesis (RRS), (Sandoz et al., 2022). Upon UV irradiation, EXD2 transiently interacts with RNA Polymerase II (RNAPII), promoting degradation of mRNA synthesised during genotoxic stress (Sandoz et al., 2022). EXD2 is crucial for RRS as it efficiently digests the nascent mRNA allowing restart of transcription (Sandoz et al., 2022). When RNAPII backtracks in front of damage, the 3’ RNA end is no longer located within the RNAPII active site; preventing restart of transcription. EXD2 may assist persistently stalled RNAPII in 3’-5’ mRNA degradation, resetting the 3’ mRNA end so that RNAPII can reactivate (Sandoz et al., 2022). It is not clear why the 3’-5’ nuclease activities of both proteins: EXD2 and RNAPII, would be required to resect mRNA upon the genotoxic stress. Recombinant human EXD2 expressed in insect cells, used for mRNA digestion *in vitro* (Sandoz et al., 2022), was purified by me.

1.6.6 The role of EXD2 at mitochondria

According to (Silva et al., 2018), EXD2 is a mitochondrial ribonuclease preventing inappropriate association of mRNAs with mitoribosome which facilitates mitochondrial translation. It is required for metabolic homeostasis and its depletion leads to alterations in metabolites from glycolysis and tricarboxylic acid (TCA) cycle, decreased ATP production, impaired respiration, and increased number of reactive oxygen species (ROS), (Silva et al., 2018). EXD2 is targeted to mitochondria through mitochondrial targeting sequence (MTS) and its disruption (deletion of 1-61 residues) results in a nucleocytoplasmic signal (Silva et al., 2018). Interestingly, the function and intra-mitochondrial localization of EXD2 proposed by Silva and colleagues (Silva et al., 2018) is contradicted by data obtained by Park and colleagues (Park et al., 2019), who propose that EXD2 is anchored to the mitochondrial outer membrane (OMM) with a cytosolic C-terminus. The OMM localization of EXD2 has been also proposed by (Hensen, Moretton, Van Esveld, Farge, & Spelbrink, 2018) and (Yoo & Rhee, 2020). Experiments performed by (Qin, Myers, Carey, Carr, & Ting, 2021) shown that EXD2 has higher RNA-binding activity in the OMM than in the nucleus.

1.6.7 EXD2 as a target for cancer treatment

As described in subchapter (1.1.2), DNA damage repair and DNA replication are major clinical targets in cancer treatment (R. Huang & Zhou, 2021). Interestingly, EXD2 takes part in both, DSB repair and replication stress response (Broderick et al., 2016), (Nieminuszczy et al., 2019). As cancer cells heavily rely on HR, EXD2 is an interesting target for cancer treatment (Nieminuszczy, Broderick, & Niedzwiedz, 2016). Mutations in certain DNA repair pathways lead to increased dependence on other pathways for survival. Cells defective in HR factors, including EXD2, are sensitive to PARP inhibitors (Nieminuszczy et al., 2016), (McCabe et al., 2006). As EXD2 and BRCA1/2 are synthetic lethal, targeting EXD2 could be used as a treatment of cancer patients bearing BRCA mutations (Nieminuszczy et al., 2019). Given that EXD2 depletion sensitizes cells to common anticancer agents, including CPT or IR, (Broderick et al., 2016), inhibition of EXD2 may be used to enhance the toxicity of anticancer therapies. Therefore, development of small molecule inhibitors targeting EXD2 specifically may become a promising strategy in cancer treatment.

Another reason to target EXD2 is its role in mitochondrial metabolism (Stracker, 2018). The TCA cycle, which is altered in EXD2 depleted cells, generates ATP, and enables the growth of many cancers (Silva et al., 2018). EXD2 depletion results in a defect in mitochondrial translation, which is interesting as inhibitors targeting mitochondrial translation may affect cancer initiating cell population (Stracker, 2018), (Silva et al., 2018).

1.6.8 EXD2 in human diseases

Patients with interstitial deletions in chromosome band 14q24.1q24.3, which encodes EXD2, have congenital heart defects, brachydactyly and mild intellectual disability (Oehl-Jaschkowitz et al., 2014). However, the exact contribution of EXD2 to these phenotypes is not known, as described deletion affects also other 18 neighbouring genes. Patients with chronic thromboembolic pulmonary hypertension (CTEPH) and pulmonary arterial hypertension (PAH) have increased level of autoantibodies against EXD2 in sera (Naito et al., 2019). Therefore, these antibodies could be used as a disease biomarker.

1.6.9 EXD2 in other model organisms

D. melanogaster EXD2 orthologue CG6744 (dEXD2) was shown to influence metabolism and development of *Drosophila* (Silva et al., 2018). Its depletion resulted in premature germline stem cells attrition, delays in development and lifespan extension (Silva et al., 2018). Here, dEXD2 expression was shown to be higher in the ovary than in the brain or gut, and the authors claim that EXD2-mutants have reduced fertility. This loss of fecundity may affect lifespan, which Silva and colleagues show was dramatically increased in dEXD2-deficient females. dEXD2 deficient flies were characterized by impaired respiration and reduced mitochondrial translation (Silva et al., 2018). Therefore, the observed physiological changes were attributed to increased ROS and disrupted mitochondrial homeostasis. In a separate study by Chennuri and colleagues, mutation of EXD2 in *Drosophila* (dEXD2^{c05871}) resulted in repair deficiency of inter-strand DNA crosslinks (ICL), introduced by diepoxybutane (DEB) and mitomycin C (MMC), (Chennuri, Cox, & Saunders, 2018). As ICLs may block replication and transcription of

DNA, dEXD2^{c05871} had increased genome instability (Chennuri et al., 2018). This study performed by Chennuri and colleagues does not describe *Drosophila* phenotype, which was observed by Silva and colleagues.

An orthologue of EXD2 was also identified in the *Xenopus* genome (Cuykendall & Houston, 2010), where it may play a role in germ cell development.

1.6.10 Biochemical characterization of EXD2 activity *in vitro*

EXD2 stimulates DSB resection and its 3’-5’ exonuclease activity has been confirmed *in vitro* (Broderick et al., 2016). Nuclease assays performed with full-length EXD2 purified from bacteria show 3’-5’ exonuclease activity on ssDNA (50 nt) and almost no activity towards blunt end dsDNA (50 bp), (Broderick et al., 2016). The truncated (76 - 564 residues), highly soluble form of EXD2 containing the exonuclease domain behaved indistinguishably from the full-length protein in several *in vitro* resection assays (Broderick et al., 2016).

The same type of EXD2 truncation (aa 76 - 564), purified from *E.coli* by (Park et al., 2019), was shown to have substrate discrimination activity depending on the coordinated metal cation (tested on single stranded nucleic acids substrates). In the presence of Mn^{2+} ions, EXD2 exhibited exonuclease activity towards both DNA and RNA substrates (Park et al., 2019). In the presence of other divalent metal ions (Mg^{2+}, Zn^{2+}, Co^{2+}), EXD2 exhibited no exonuclease activity towards DNA. However, RNA digestion activity was stimulated by both Mn^{2+} and Mg^{2+} ions, with the stronger effect for Mn^{2+} (Park et al., 2019). Interestingly, an EXD2 variant (aa 76 – 284) lacking the C-loop that contacts the active site, had DNA nuclease activity in presence of both Mg^{2+} and Mn^{2+} (Park et al., 2019). A shorter variant (aa 76 – 260) without the whole C-segment, had no nuclease activity on this substrate (Park et al., 2019).

EXD2 deprived of its N-terminal domain (aa 61 - 621) was also purified from *E. coli* by (Silva et al., 2018). It exhibited nuclease activity on both ssDNA and dsDNA (40 bp). It also efficiently digested ssRNA and dsRNA (40 bp) and both strands of DNA-RNA hybrids (40 bp). Similarly to (Park et al., 2019), in the presence of Mg^{2+} ions, EXD2 exhibited activity only on ssRNA and RNA-DNA hybrids. Electrophoretic mobility shift assays (EMSA) revealed a binding preference towards ssRNA, and no clear supershift with ssDNA or DNA-RNA substrates (Silva et al., 2018). Another truncated EXD2 (aa 61-621), purified from bacteria by (Jia et al., 2022), demonstrated stable binding to both dsRNA and dsDNA. It exhibited nuclease activity on RNA-DNA hybrids, with a cleavage preference towards RNA. However, contrary to (Silva et al., 2018), it barely cleaved dsDNA.

tEXD2 (aa 76 - 564) and full-length EXD2 purified by (Broderick et al., 2016) were able to initiate resection from a nicked and gapped (1 nt) dsDNA. These substrates resembled substrates generated during the initial stage of DSBs resection, *via* MRE11's endonuclease activity. Nuclease assays performed by (Silva et al., 2018) also showed EXD2 activity on substrates (35 bp) with an introduced nick or 1 nucleotide gap.

Additionally, tEXD2 (aa 76 - 564) exhibited weak activity towards dsDNA with resected 3' ends (5' overhang, 50/60 bp), (Broderick et al., 2016). Assays performed by (Silva et al., 2018) also showed EXD2 activity on dsDNA with a recessed 3' end (16/35 nt) and reduced EXD2 activity on ssDNA with a 3' phosphate.

1.6.11 The mechanism of substrate cleavage by EXD2 *in vitro*

According to (Jia et al., 2022), EXD2 unwinds RNA-DNA hybrids before their cleavage. The proposed mechanism involves cyclic unwinding of 4 bp, followed by their "one at a time" excision. As DNA in the hybrid is cleaved less efficiently than RNA, it enters a rewinding and stalling stage after several unwinding-excision cycles (Jia et al., 2022).

1.6.12 EXD2 protein structure

The structure of full-length human EXD2 is not known, as only its exonuclease domain (aa 76-295) has been crystalized (Park et al., 2019). Data presented by (Park et al., 2019) shows that the 3'-5' exonuclease domain consists of a canonical DEDDy domain followed by a C-segment (Figure 20).

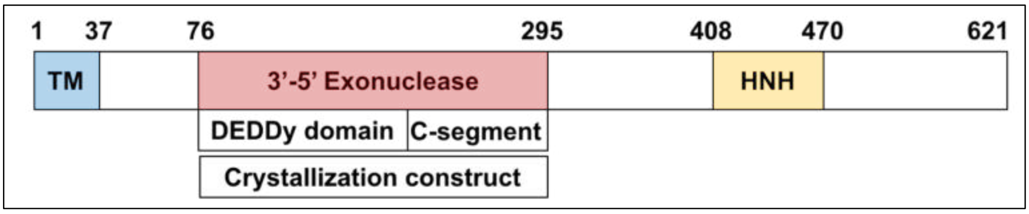


Figure 20. Domain structure of human EXD2, (based on Park et al., 2019). TM – transmembrane domain. DEDDy domain – catalytic core residues taking part in metal ions coordination (Asp, Glu, Asp, Asp, Tyr). HNH – endonuclease-like domain.

DEDDy is a conserved amino acid motif present in DnaQ-like nucleases, which performs 3'-5' exonuclease digestion of DNA or RNA (Yang, 2011). Contrary to other exonucleases from the DEDD superfamily, the active site of EXD2 is unusually organized as a domain-swapped dimer, resulting in a chimeric active site (Park et al., 2019). The composite active site suggests that EXD2 functions as an obligate dimer, whose dimerization interface is based on a large surface of hydrophobic interactions (Figure 21).

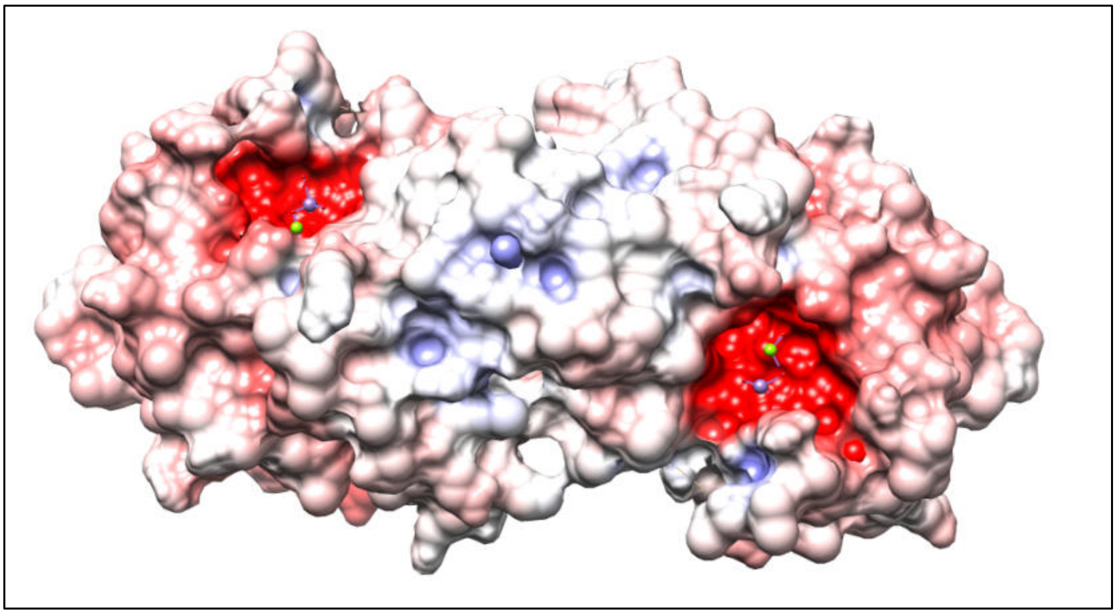


Figure 21. Electrostatic charges on the protein surface of EXD2 exonuclease domain dimer, (based on structure from Park et al., 2019). Picture obtained from UCSF Chimera program, by Dr Jerome Gouge. Negative charge – red, positive charge – blue, uncharged – white, small green sphere – Magnesium ion, small purple sphere – Manganese ion.

EXD2 shares high sequence similarity with Werner syndrome protein (WRN) exonuclease domain and like other nucleases from DEDDy family it contains two metal ions at the active site (Park et al., 2019). WRN is 3’-5’ helicase and exonuclease, which takes part in DNA repair (Perry et al., 2006). Its mutation results in Werner syndrome, which is characterized by greatly increased predisposition to cancer (Perry et al., 2006).

According to (Park et al., 2019) the catalytic core residues taking part in metal ions coordination in EXD2 are: Asp108, Glu110, Asp171, Tyr 242 and Asp246 (hence DEDDy). Alignment between EXD2 and WRN sequences (Figure 90) indicates two key amino acids (Asp108 and Glu110) within the active site of EXD2 (equivalent to Asp82 and Glu84 in WRN), (Broderick et al., 2016). These catalytic core residues coordinate a manganese ion in the active site of EXD2 (Figure 22), (Park et al., 2019). Mutation of the Asp82 residue in WRN (B. Li et al., 2009), as well as Asp108 and Glu110 mutations in EXD2 (Broderick et al., 2016) result in a lack of exonuclease activity.

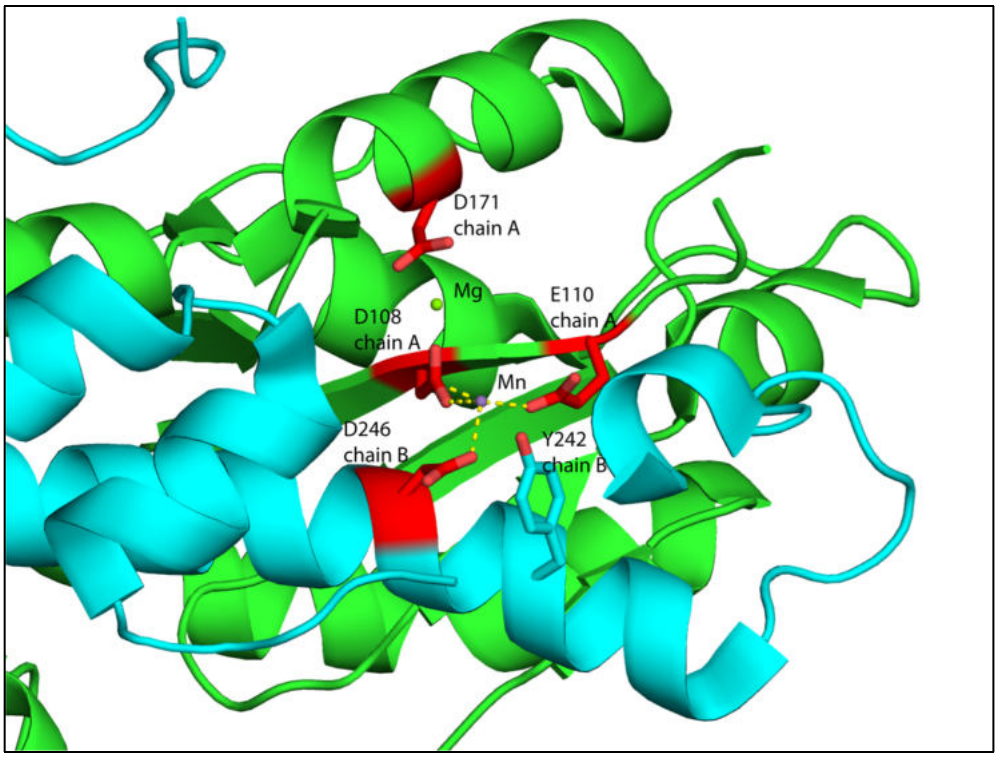


Figure 22. Composite active site of EXD2, (based on structure from Park et al., 2019). Picture from PyMOL program, by Dr Jerome Gouge.

Apart from its exonuclease domain, EXD2 is predicted to have a HNH endonuclease-like domain characterized by tandem CXXC motifs (Silva et al., 2018), which will be further described in the chapters (5.5) and (6.1).

2. Aims of the project

DSB repair is performed *via* the coordinated action of multiple proteins, including the recently identified EXD2 nuclease. It has been shown that EXD2 promotes HR and takes part in the replication stress response. However, the exact role of EXD2 in these pathways has not been fully characterized.

The main aim of my PhD project is to elucidate the molecular mechanism of DNA end resection by EXD2. To achieve this, I focused on:

1. Purification of full-length human EXD2.
2. Identification of the nucleic acid substrates EXD2 is able to process, including damaged and non-canonical DNA structures.
3. Examining if human EXD2 protein exhibits any other biochemical activities.
4. Examining if EXD2 activity can be regulated by RPA.
5. Structural characterization of full-length human EXD2.
6. Purification of the human MRN complex to reconstitute the resection reaction *in vitro* in the future.

3. Materials and methods

3.1 Molecular biology techniques

3.1.1 DNA constructs

DNA constructs, used for expression of single proteins, are listed in (Table 3). Generation of plasmids, bacmids and baculoviruses is further described in subchapter (3.1).

Table 3. DNA constructs used for expression of single proteins. Plasmids were synthetised by Genscript or generated using molecular biology techniques. Some bacmids and baculoviruses were generated by Dr J. Krwawicz, as indicated. Those marked with '+' were generated by me.

Construct (in pLIB)	Plasmid	Bacmid	Baculovirus
10xHis-TEV-**hEXD2**-3C-2xStrep (WT)	+	+	+
10xHis-TEV-**hEXD2**-3C-2xStrep (D108A E110A)	+	+	+
10xHis-TEV-**truncEXD2**-3C-2xStrep, without transmembrane domain	+	+	+
10xHis-TEV-**hEXD2** WT	GenScript	J. Krwawicz	+
10xHis-TEV-**hEXD2** (D108A E110A)	GenScript	J. Krwawicz	+
hMRE11-TEV-10xHis (WT)	GenScript	+	+
hRAD50-TEV-2xStrep (WT)	GenScript	+	+
hNBS1-TEV-2xStrep (WT)	GenScript	+	+
hNBS1 (WT)	+	+	+

DNA constructs, used for expression of protein complexes, are listed in Table 4.

Table 4. DNA constructs used for expression of protein complexes. Some bacmids and baculoviruses were generated by Dr J. Krwawicz, as indicated in the table. Those marked with '+' were generated by me.

Construct (in pBIG1a)	Subunits	Bacmid	Baculovirus
MR His Strep WT	hMRE11-TEV-10xHis (WT) hRAD50-TEV-2xStrep (WT)	+	+
MR His Strep MUT	hMRE11-TEV-10xHis (H129L D130V) hRAD50-TEV-2xStrep (WT)	+	+
MRN His 2Strep WT	hMRE11-TEV-10xHis (WT) hRAD50-TEV-2xStrep (WT) hNBS1-TEV-2xStrep (WT)	J. Krwawicz	
MRN His 2Strep MUT	hMRE11-TEV-10xHis (H129L D130V) hRAD50-TEV-2xStrep (WT) hNBS1-TEV-2xStrep (WT)	J. Krwawicz	
MRN HIS single Strep WT	hMRE11-TEV-10xHis (WT) hRAD50-TEV-2xStrep (WT) hNBS1 (WT)	+	+
MRN HIS single Strep MUT	hMRE11-TEV-10xHis (H129L D130V) hRAD50-TEV-2xStrep (WT) hNBS1 (WT)	+	+

3.1.2 Bacterial transformation

Plasmid DNA (2 μl) was mixed with *Escherichia coli* DH5α competent cells (20 μl) and incubated on ice for 30 min. Heat shock reaction was carried out for 45 seconds at 42°C and recovered on ice for 10 min. Cells were incubated with Luria-Bertani (LB) medium (200 μl), shaking for ~1 h at 37°C. The bacteria were plated on L-agar plate supplemented with designated antibiotic.

3.1.3 Plasmid isolation

Bacteria plates were incubated at 37°C overnight. The desired colonies were transferred to 3 ml of LB supplemented with designated antibiotics. Bacteria cultures were incubated overnight, shaking (210 rpm) at 37°C. Plasmids were isolated from centrifuged bacteria pellet using GeneJet Plasmid MiniPrep Kit (Thermo Scientific). Purified plasmids were stored at - 20°C.

3.1.4 Glycerol stocks of bacteria

Bacteria cultures (1.5 ml) were centrifuged (~ 3000 g), and the supernatant discarded. Pellets were resuspended in 1 ml of 50 % glycerol in LB. After being frozen on dry ice, samples were transferred to - 80°C.

3.1.5 biGBac system

Constructs described in Table 4 were generated using biGBac protocol based on (Cutts & Vannini, 2018). To assemble pBIG1, each gene of interest (GOI) needs to have overlapping regions introduced by Polymerase Chain Reaction (PCR). Different primers were used for different genes. Each primer had a region homologous to the pLIB vector and a region that introduced overlap required for the assembly. After the successful PCR (3.1.6), product was purified from agarose gel (3.1.7) and used for the Gibson assembly reaction (3.1.8).

3.1.6 Modification of DNA using Polymerase Chain Reaction (PCR)

MR: hMRE11-TEV-10xHis (WT / H129L D130V), hRAD50-TEV-2xStrep (WT)

To allow MR assembly, overlapping regions were introduced to genes encoding MRE11 and RAD50 (Table 5). The reactions (50 μl) were composed of Q5 buffer (1X), dNTPs (0.2 mM), Q5 polymerase (0.5 μl), template (100 ng) and primers (0.5 μM each), (Table 5). Sequences of primers are listed in (Table 20).

Table 5. Introduction of overlapping regions required for MR assembly. PCR conditions.

Reaction components	Reaction conditions
Template: hMRE11-TEV-10xHis (WT) hMRE11-TEV-10xHis (H129L D130V) hRAD50-TEV-2xStrep (WT)	• Denaturation: 98°C – 30 s • Reaction, 35 cycles: 1) 98°C – 10 s 2) 65°C – 20 s 3) 72°C – 2 min 30 s • Final extension: 72°C – 4 min
Forward/Reverse primer: For MRE11: Cas I F, Cas I R For RAD50: Cas II F, Cas omega R	

After the PCR, 1 μl of DpnI was added to the sample and incubated at 37°C for 1 hour.

NBS1

To obtain construct encoding hNBS1 without any tags, hNBS1-TEV-2xStrep (WT) was used as a template for PCR reaction. The 25 μl reaction was composed of Marathon buffer (1X), dNTPs (0.3 mM), Marathon polymerase (0.5 μl), template (60 ng) and primers (0.2 μM each), (Table 6). Using the primer overhang, a stop codon was introduced. Sequences of primers are listed in (Table 20).

Table 6. PCR reaction to generate hNBS1 without any tags.

Reaction components	Reaction conditions
Template: hNBS1-TEV-2xStrep (WT)	• Denaturation 95°C – 60 s • Reaction, 35 cycles: 1) 95°C – 20 s 2) 59°C – 20 s 3) 68°C – 7 min 20 s • Final extension: 68°C – 5 min
Forward/Reverse primer: 1FtaglessNbs1,1RtaglessNbs1	

MRN: hMRE11-TEV-10xHis (WT / H129L D130V), hRAD50-TEV-2xStrep (WT), hNBS1 (tagless)

To allow MRN assembly, overlapping regions were introduced to genes encoding MRE11, RAD50 and NBS1 (Table 7). The reactions (50 μl) were composed of Q5 buffer (1X), dNTPs (0.2 mM), Q5 polymerase (0.5 μl), template (100 ng for MRE11, 100 ng for RAD50, 150 ng for NBS1) and primers (0.5 μM each). Sequences of primers are listed in (Table 20).

Table 7. Introduction of overlapping regions required for MRN assembly. PCR conditions.

Reaction components	Reaction conditions
Template: hMRE11-TEV-10xHis (WT), hMRE11-TEV-10xHis (H129L D130V), hRAD50-TEV-2xStrep (WT), hNBS1 (WT)	• Denaturation: 98°C – 30 s • Reaction, 35 cycles: 1) 98°C – 10 s 2) 65°C – 20 s 3) 72°C – 2 min 30 s • Final extension: 72°C – 4 min
Forward/Reverse primer: For MRE11: Cas I F, Cas I R For RAD50: Cas II F, Cas II R For NBS1: Cas III F, Cas omega R	

EXD2: 10xHis-TEV-hEXD2-3C-2xStrep (WT / D108A E110A)

To introduce 2xStrep on the C-terminus sequence of 10xHis-TEV-hEXD2, two PCR reactions were performed. pLIB plasmid encoding hRAD50-TEV-2xStrep was used as a template for first PCR reaction. The reaction product contained pLIB backbone with 2xStrep sequence, but without the hRAD50 gene. For the second PCR reaction, pLIB encoding 10xHis-TEV-hEXD2 was used as a template. The designed primer introduced a 3C cleavage site on the C-terminal end of EXD2. The reactions were composed of Q5 buffer (1X), dNTPs (0.2 mM), Q5 polymerase (0.5 μl), template (100 ng; either hRAD50-TEV-2xStrep or 10xHis-TEV-hEXD2) and primers (0.5 μM each; for RAD50: fstrep3C.1 and RtevEXD2.1; for EXD2: FhisEXD2.2 and RhisEXD2.2). Sequences of primers are listed in (Table 20).

EXD2 without transmembrane domain: 10xHis-TEV-truncEXD2-3C-2xStrep

To generate a construct encoding human EXD2 without its transmembrane domain (without aa 2-37), two PCR reactions were performed. In the first PCR reaction primers 1FEXTRANS and 2REXTRANS were used, in second PCR reaction primers 3FEXTRANS and 4FEXTRANS were used. The reactions were composed of Q5 buffer (1X), dNTPs (0.2 mM), Q5 polymerase (0.5 μl), template (89 ng) and primers (0.5 μM each), (Table 8). Sequences of primers are listed in (Table 20).

Table 8. PCR reaction to generate10xHis-TEV-truncEXD2-3C-2xStrep

Reaction components	Reaction conditions
Template: 10xHis-TEV-hEXD2-3C-2xStrep	• Denaturation: 98°C – 30 s • Reaction, 35 cycles: 1) 98°C – 10 s 2) 66°C – 20 s 3) 72°C – 2 min • Final extension: 72°C – 2 min
Forward/Reverse primer: 1) 1FEXTRANS and 2REXTRANS 2) 3FEXTRANS and 4REXTRANS	

3.1.7 DNA purification from agarose gels

Loading dye (6x) was added to the samples before resolution on agarose gels. Depending on DNA length, different amount of agarose was dissolved in Tris-Borate-EDTA (TBE) or Tris-acetate-EDTA (TAE) buffer. To visualize the DNA, SYBR Green (Thermo Fisher), SYBR Gold (Thermo Fisher) or GelRed (Biotium) were added either directly to the samples or to dissolved agarose. GeneRuler 1 kb DNA (Thermo Scientific) or GeneRuler 50 bp DNA (Thermo Scientific) were used as ladders. The results of the run were visualised using UV or blue light systems. The desired bands were cut out from the gel by scalpel and purified using the Zymoclean Gel DNA Recovery Kit (Zymoresearch). Plasmids were eluted in DNA elution buffer or in water.

3.1.8 Gibson assembly

After modification of DNA (3.1.6), PCR products were used for Gibson assembly reactions. Before the assembly, the pBIG1a vector was digested overnight with SwaI (New England Biolabs, NEB). After the digestion, enzyme was inactivated by incubation at 65°C for 20 min.

Assembly of a human MR expression construct

hMRE11-TEV-10xHis (WT / H129L D130V) and hRAD50-TEV-2xStrep (WT)

To assemble three fragments of DNA, reaction components were mixed on ice and incubated in 50°C for 45 min (in 1X Gibson Assembly Master Mix; reaction volume 20 μl), (Table 9).

Table 9. Gibson assembly of hMRE11-TEV-10xHis (WT / H129L D130V) and hRAD50-TEV-2xStrep (WT).

Reaction components	Amount
pBIG1a	63 ng
MRE11 or MRE11 (H129L D130V)	129 ng 179 ng
RAD50 (WT)	82.5 ng

The mixture was used for *E. coli* DH5α transformation (3.1.2). After performing colony PCR (3.1.9), purified plasmids (MR WT 1, MR MUT 3) were sequenced by Integrated DNA Technologies (IDT) and used for bacmid generation (3.1.10). Primers used for the sequencing reaction (1-14MRN) are listed in (Table 20).

Assembly of a human NBS1 expression construct

To assemble a construct to express hNBS1 (WT) without any tags, two ends of one PCR product (3.1.6) were joined together, reaction components were mixed on ice and incubated at 50°C for 30 min (in 1X Gibson Assembly Master Mix; reaction volume 20 μl), (Table 10).

Table 10. Gibson assembly of hNBS1 (WT) without any tags.

Reaction components	Amount
NBS1 PCR product	300 ng

Results of the assembly were validated by enzymatic digest of purified plasmids (BamHI and HindIII from NEB). Primers used for the sequencing reaction (NbsF1-F6seq) are listed in (Table 20).

Assembly of a human MRN expression construct

hMRE11-TEV-10xHis (WT / H129L D130V), hRAD50-TEV-2xStrep (WT) and hNBS1 (tagless)

To assemble four fragments of DNA, reaction components were mixed on ice and incubated at 50°C for 45 min (in 1X Gibson Assembly Master Mix; reaction volume 20 μl), (Table 11).

Table 11. Gibson assembly of hMRE11-TEV-10xHis (WT / H129L D130V), hRAD50-TEV-2xStrep (WT) and tagless hNBS1.

Reaction components	Amount
pBIG1a	58.8 ng
MRE11 (33 ng/μl) or MRE11 (H129L D130V)	115.5 ng 125 ng
RAD50	80 ng
NBS1	120 ng

After performing colony PCR (3.1.9), purified plasmids were sequenced using primers listed in (Table 20).

Assembly of a human EXD2 expression construct

10xHis-TEV-hEXD2-3C-2xStrep (WT / D108A E110A)

To assemble two fragments of DNA, reaction components were mixed on ice and incubated at 50°C for 20 min (in 1X Gibson Assembly Master Mix; reaction volume 20 μl), (Table 12).

Table 12. Gibson assembly of 10xHis-TEV-hEXD2-3C-2xStrep (WT / D108A E110A).

Reaction components	Amount
Product of pLIB hRAD50-TEV 2xStrep PCR	88 ng
Product of pLIB 10xHis-TEV-hEXD2 PCR	144 ng

Results of the assembly were validated by PCR of purified plasmids. Primers used for the validation reaction (HisEXD2strepassemb1F, HisEXD2strepassemb1R) and the sequencing reaction (1F-5F_HisEXD2Strep) are listed in (Table 20). Plasmids labelled as hisEXD2strep WT colony 1 (178 ng/μl) and hisEXD2strep mut colony 5 (151 ng/μl), were used for generation of the bacmids.

Assembly of a human EXD2 without transmembrane domain expression construct

10xHis-TEV-truncEXD2-3C-2xStrep WT, (without aa 2-37)

Reaction components of 10xHis-TEV-truncEXD2-3C-2xStrep WT assembly (two DNA fragments) are listed in (Table 13). The assembly was performed in the presence of Gibson Assembly Master Mix (1X), the total volume of reaction was 26 μl. For this construct, digestion with 1 μl of DpnI (37°C, 1h) was performed after the reaction.

Table 13. Gibson assembly of 10xHis-TEV-truncEXD2-3C-2xStrep WT.

Reaction components	Amount
Product of PCR reaction with 1FEXTRANS and 2REXTRANS primers	90 ng
Product of PCR reaction with 3FEXTRANS and 4REXTRANS primers	68.6 ng

Primers (1-6_ SEQTREXD2) used for the sequencing of purified plasmid (EXD2wtTrunc colony 2, 121 ng/μl) are listed in (Table 20).

3.1.9 Colony PCR

To validate results of Gibson assembly reactions, colony PCR was performed. For chosen constructs, several colonies were replated. Pipette tips used for the replating (separate tip for each colony) were put into 80 μl of water. The water samples were boiled for 10 – 15 min and 4 μl was used as a template for colony PCR reaction (Table 14). The reactions were composed of Marathon buffer (1x), dNTPs (0.3 mM), Marathon polymerase (0.5 μl), template (boiled bacteria) and primers (0.5 μM each). Primers (S1A, S1B, S1C, S1D) used for the validation of both MR and MRN assembly, are listed in (Table 20).

Table 14. Colony PCR of bacteria expressing MR (hMRE11-TEV-10xHis and hRAD50-TEV-2xStrep) and MRN (hMRE11-TEV-10xHis, hRAD50-TEV-2xStrep and hNBS1).

Reaction conditions
• Denaturation: 95°C – 3 min
• Reaction, 30 - 35 cycles:
1) 95°C – 15 s
2) 58°C – 15 s
3) 68°C – 45 s
• Final extension: 68°C – 1 min

Colonies which generated PCR products of the correct size upon 2% agarose TBE gel analysis were picked from the plates and amplified overnight (in 3 ml of LB supplemented with Ampicillin and Spectinomycin). The assembly of MR and MRN was additionally checked by digestion of the plasmids with SwaI restriction enzyme (NEB).

3.1.10 Bacmid preparation

Constructs for protein expression were transposed into *E. coli* DH10 EMBacY*,* which carries a baculovirus shuttle vector (bacmid), enabling the construct to be transposed *via* the Tn7 attachment sites. DH10 EMBacY contains tetracycline-resistant helper plasmid encoding the transposase (Rowland, Wu, Liu, & Davies, 2020). Transposition results in interference of the LacZ gene, enabling selection of colonies through blue-white screening. Yellow fluorescent protein (YFP) reporter gene enables detection of the virus after transfection of insect cells with the bacmid.

Transposition Transformation

Plasmid (2 μl) was mixed with *E. coli* DH10 EMBacY cells (75 μl) and incubated on ice for 30 min. Heat shock reaction was carried out for 45 seconds at 42°C and recovered on ice for 5 min. Super Optimal broth with Catabolite repression (SOC), Super Optimal Broth (SOB) or LB medium (900 μl) was added to the bacteria and incubated for 4-6 h, shaking at 37°C. Sample (~ 1/10 volume) was plated on the bacmid transposition plate [L Agar (1L), Kanamycin (50 mg), Tetracycline (10 mg), Gentamicin Sulfate (7 mg), Bluo-gal (100 mg), IPTG (40 mg)] and incubated 24 h – 36 h at 37°C. White colonies were restreaked and used for preparation of overnight cultures (4 ml LB with antibiotics). If needed, glycerol stocks were prepared.

Bacmid isolation from bacteria pellets

Bacteria cells were centrifuged at 14000 g for 1 min (or slower and longer). Supernatant was removed and the pellet frozen at -20°C (if necessary). Fresh or thawed pellet was resuspended in 0.3 ml resuspension buffer (Thermo Scientific, Plasmid Miniprep kit), and transferred to 2 ml tube. Subsequently, 0.3 ml of lysis buffer (Thermo Scientific, Plasmid Miniprep kit), followed by 0.3 ml sodium acetate (3M, pH 5.5), was added. The sample was placed on ice for 5 – 10 min and centrifuged at 14000 g for 10 min. The supernatant was transferred to the tube containing 0.8 ml isopropanol. Transferring any white precipitate was avoided. The sample was mixed by inversion, placed on ice for 5 – 10 min and centrifuged at 14000 g for 10 min at RT. The supernatant was removed, and 0.5 ml of 70% ethanol was added. The sample was centrifuged at 14000 g for 5 min at RT. The supernatant was removed and the pellet air dried at RT (5 – 10 min or longer). The purified DNA was dissolved in 50 μl of sterile water and stored at -20°C.

3.2 Protein Expression

3.2.1 Protein expression in insect cells

Transfection

Transfection was carried out in Sf9 cells, with cell counting and viability calculated by Trypan blue staining (1:1). Cells were diluted to $5*10^5$ cells/ml in media without any antibiotics (Insect-Xpress medium). 2 ml of cells per well were plated on a 6-well plate. After cells adhered (Microscope Nikon Eclipse TS100), the old media supplemented with antibiotics (mix of penicillin 50 units/ml + streptomycin 50 μg/ml) was replaced with antibiotic free media. Solution A (20 μl of bacmid DNA with 200 μl of media) and B (20 μl of Cellfectin II surfactant Gibco with 200 μl of media) were prepared. After mixing A and B together, 200 μl of this mixed solution was added to each well. Plates were then placed at 27°C in a humid box. Supernatant containing virus was harvested after 3 days (P1) and stored in 2% FBS at 4°C.

P2 generation

Virus was amplified by adding P1 virus stock to 25 mL of Sf9 cells ($5*10^5$ cells/ml). Cells were incubated shaking at 27°C, in a medium supplemented with penicillin and streptomycin. Virus was harvested by spinning down cells after 2-4 days (P2), once cells expressed the fluorescence reporter (YFP) and cells viability began to drop (Luna Dual Fluorescence Cell Counter). P2 (supernatant) was stored with 2% FBS at 4°C and used for further cell infection and protein expression.

Test expression

Different amounts of P2 (100 μl – 1 ml) were added to 50 ml of Hi5 cells ($5*10^5$ cells/ml). After infection, viability and fluorescence of cells were checked every day (Luna Dual Fluorescence Cell Counter) and aliquots were frozen.

3.3 Protein purification

3.3.1 Protein purification - overview

The general protocol for recombinant proteins purification had three main steps. Two types of affinity chromatography, allowing highly specific binding of recombinant proteins to the resin, followed by size exclusion (SEC). The last step separated proteins based on their size, allowing removal of contaminants and aggregates. All the steps were performed on chromatography columns controlled by AKTA Pure System (GE Healthcare), equipped with fixed wavelength monitor (280 nm). All buffers were filtered and degassed before the use. Quality of purified proteins was controlled after each purification step using SDS-PAGE. Protocols described in the Methods section are final and optimised. The purification optimization procedure is described in the Results section.

3.3.2 Purification of full-length EXD2 WT and MUT (D108A E110A)

10xHis-TEV-hEXD2-3C-2xStrep WT and MUT (D108A E110A)

Frozen pellet, from 1 L of infected Hi5 insect cells, was resuspended in 200 ml of lysis buffer (50 mM Tris pH 7.5, 400 mM NaCl, 10 % glycerol, 2 mM $MgCl_2$, 1 % Triton X-100, 1 mM PMSF, 2 mM BME, 2 SIGMAFAST protease inhibitor cocktail tablets from Sigma-Aldrich). Cells were lysed on ice using Dounce homogenizer, followed by pulse sonication (Amplitude 20 %; Pulse: 5 s on, 10 s off; Time: 2 x 5 min) and centrifugation (37000 g, 40 min, 4°C). Filtered supernatant was loaded twice on 5 ml StrepTrap HP column (Cytiva). For column equilibration and wash, lysis buffer without protease inhibitor cocktail tablets was used. Proteins were eluted with Strep elution buffer (10 mM desthiobiotin, 50 mM Tris pH 7.5, 400 mM NaCl, 10 % glycerol, 2 mM $MgCl_2$, 0.5 % Triton X-100, 1 mM PMSF, 2 mM BME, final pH 7.55). After addition of imidazole buffer (final imidazole concentration = 10 mM), sample was loaded on 5 ml HisTrap HP column (Cytiva). For His-Tag affinity chromatography two buffers were prepared: buffer A without imidazole (50 mM Tris pH 7.5, 500 mM NaCl, 10 % glycerol, 0.5 % Triton X-100, 2 mM BME) and imidazole buffer B (A with 500 mM imidazole). During the chromatography, buffers were mixed with desired proportions (2% B for column equilibration, 6 % B for wash, 6 – 80 % for elution). Eluted fractions were concentrated to 5 ml, using a Sartorius

Vivaspin 20 Centrifugal Concentrator (MWCO 30 kDa). Concentrated sample was spun for 3 min at 12000 x g and loaded on a HiLoad 16/60 Superdex 200 column. Protein was eluted in gel filtration buffer (50 mM Tris pH 7.5, 300 mM NaCl, 10% glycerol, 0.05% Triton X-100 and 2 mM BME), concentrated and flash frozen in liquid nitrogen. Small aliquots (900 μl), designated for a long-term storage in 30 % glycerol, were supplemented with 60 % glycerol (600 μl) before freezing. All the proteins were stored at - 80°C.

3.3.3 Purification of truncated EXD2

10xHis-TEV-hEXD2-3C-2xStrep WT, without the transmembrane domain (2-37 aa)

The purification protocol was the same as for the full-length EXD2 (3.3.2). However, because of increased solubility of the truncated protein, no detergent (Triton X-100) was used.

3.3.4 Purification of full-length EXD2, novel mutant (H442A, H463A)

10xHis-TEV-hEXD2-3C-2xStrep, novel mutant (H442A, H463A)

Newly designed EXD2 mutant (H442A, H463A) was basically purified according to the full-length 10xHis-TEV-hEXD2-3C-2xStrep WT and (D108A E110A) mutant purification protocol. However, due to very low yield, no size exclusion step was performed. After Strep-Tag chromatography (with re-loading of flow through on the column) and His-Tag affinity chromatography, protein was concentrated using a Sartorius Vivaspin 20 Centrifugal Concentrator. During protein concentration, HisTrap elution buffer was exchanged with the gel filtration buffer (50 mM Tris pH 7.5, 300 mM NaCl, 10% glycerol, 0.05% Triton X-100 and 2 mM BME). During concentration, 20 ml of gel filtration buffer was added to 5 ml of the protein. When the volume decreased to 2 ml, another 18 ml of gel filtration buffer was added. Sample was concentrated to ~ 1.5 ml and flash frozen in the liquid nitrogen as above.

3.3.5 Cleavage of Strep-Tag from purified 10xHis-TEV-hEXD2-3C-2xStrep

Small aliquots of frozen 10xHis-TEV-hEXD2-3C-2xStrep: WT, (D108A E110A) mutant (purified as in 3.3.2) and (H442A, H463A) mutant (purified as in 3.3.4) were cleaved with 3C protease. Protease GST-3C (in 50 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM TCEP, 40 % glycerol) was a kind gift from Dr Marios Koliopoulos. Around 21.7 μg of each protein (high glycerol aliquots for WT) was incubated with 1.2 μl of GST-3C (2 mg/ml) mixing at 4°C. After washing StrepTactin Superflow (IBA) with water and gel filtration buffer, resin was incubated with cleaved proteins for ~ 2 h, mixing at 4°C. After 10 min centrifugation, the supernatant was absorbed, and efficiency of the cleavage validated via SDS-PAGE. Proteins without C-terminal Strep-Tag were flash frozen in a liquid nitrogen and used only with biotinylated DNA substrates (Figure 71), (Figure 72), (Figure 73).

3.3.6 Purification of MRN complex

hMRE11-TEV-10xHis (WT / H129L D130V), hRAD50-TEV-2xStrep, hNBS1- TEV-2xStrep)

Pellet from 1 L of infected Hi5 cells was resuspended in 200 ml of the lysis buffer (50 mM Tris pH 7.5, 300 mM NaCl, 10% glycerol, 2 mM $MgCl_2$, 1 mM PMSF, 2 mM BME, SIGMAFAST protease inhibitor cocktail tablet), supplemented with 10 µl benzonase nuclease (≥250 U/µl, Milipore). Cells were lysed using Dounce homogenizer, sonicated (Amplitude 20 %; Pulse: 5 s on, 10 s off; Time: 2 x 5 min), mixed at 4°C (1 h) and centrifugated (37000 g, 40 min, 4°C). The filtrated supernatant was loaded twice on a 5 ml StrepTrap HP column (Cytiva) equilibrated with Strep buffer (50 mM Tris pH 7.5, 300 mM NaCl, 10% glycerol, 2 mM $MgCl_2$, 1 mM PMSF, 2 mM BME). Proteins were eluted with Strep buffer supplemented with 10 mM desthiobiotin and loaded on 1 ml Hi Trap Heparin HP column (Cytiva). For heparin affinity chromatography two buffers were prepared: A without any salt, and B with 1 M NaCl (50 mM Tris, 10 % glycerol, 1 mM PMSF, 2 mM BME). Increasing concentration of buffer B was used for the elution. Chosen fractions were concentrated using Sartorius Vivaspin 6 Centrifugal Concentrator (MWCO 100 kDa) and loaded on Superose 6 10/300 GL column. Proteins were eluted in gel filtration buffer (50 mM Tris, 150 mM NaCl, 10 % glycerol, 2 mM BME, final pH 8), concentrated and flash frozen in a liquid nitrogen as above.

3.3.7 Purification of MR WT

hMRE11-TEV-10xHis, hRAD50-TEV-2xStrep

Frozen pellet, from 1 L of infected Hi5 insect cells, was resuspended in 200 ml of the MR lysis buffer (50 mM Tris pH 7.5, 300 mM NaCl, 10 % glycerol, 2 mM $MgCl_2$, 1 mM PMSF, 2 mM BME, 2 SIGMAFAST protease inhibitor cocktail tablets). Cells were lysed with Dounce homogenizer, pulse sonicated (Amplitude 20 %; Pulse: 5 s on, 10 s off; Time: 2 x 5 min) and centrifugated (37000 g, 50 min, 4°C). Filtered supernatant was loaded on 5 ml StrepTrap HP column (Cytiva). Lysis buffer without protease inhibitor cocktail tablets was used for column equilibration and wash. Elution buffer was supplemented with 10 mM desthiobiotin. Chosen fractions were loaded on 5 ml HiTrap Q HP column (Cytiva), after adjusting NaCl concentration to 100 mM. For the ion exchange chromatography two buffers were prepared: buffer A without salt (25 mM Tris, 0 M NaCl, 10 % glycerol, 2 mM BME, final pH = 8) and buffer B (buffer A with 1 M NaCl). Different ratios of buffer A and B were used (100 mM NaCl for column equilibration, gradient of buffer B for elution). Eluted fractions were concentrated using Sartorius Vivaspin 6 Centrifugal Concentrator (MWCO 100 kDa). The final NaCl concentration was ~ 400 mM. Protein was flash frozen in a liquid nitrogen.

3.4 Biochemical characterization of proteins

3.4.1 SDS-PAGE

Proteins were denatured by boiling at 100°C (5 – 7 min) with addition of NuPage LDS sample buffer (Invitrogen) and Bolt reducing agent (Novex). Samples were separated using NuPage Bis-Tris 4-12% gels (Invitrogen), run in MOPS buffer. Gels were stained with InstantBlue Coomassie Stain (Abcam). If the protein concentration was too low to observe the results, gels were silver stained according to Pierce Silver Stain protocol (Thermo Scientific).

3.4.2 Western blot analysis

Wet transfer was performed in 300 mA for 1h – 1h15 min. Sponge, 2 filter papers, gel, nitrocellulose membrane, additional 2 filter papers and another sponge, were placed on each other to allow the transfer. Transfer buffer was supplemented with 20% methanol (10x transfer buffer composition: 61 g Tris, 288 g Glycine, 2 L volume). Membrane was stained with Ponceau (0.2% Ponceau in 5% Acetic Acid) to assess quality of the transfer. Depending on the used antibody, different blocking conditions were chosen (Table 15). After incubation with primary antibody, membrane was extensively washed with TBST (TBS with 0.1% Tween 20), incubated with secondary antibody for an hour and again extensively washed in TBST. Secondary antibodies conjugated with HRP were treated with Immobilon Western HRP substrate peroxide and luminol solutions (MERCK) and visualised using film.

Table 15. Antibodies used for detection of purified EXD2, Western blot analysis.

Antibody		Blocking conditions
Anti-EXD2	Rabbit, ATLAS ab HPA005848	5% milk TBST
Anti-Strep	Strep Mab Classic HRP, IBA	3% BSA PBST
Anti-His Tag	Mouse, RnDsystems	5% milk TBST
Anti-mouse HRP	Dako polyclonal goat	5% milk TBST
Anti-rabbit HRP	Dako polyclonal goat	5% milk TBST

3.4.3 Oligonucleotide 5’ end labelling

To generate 5’ labelled DNA substrates, oligonucleotides (Integrated DNA Technologies) were labelled using γ-32P Adenosine 5'-triphosphate (ATP) and T4 polynucleotide kinase (NEB). Reaction components (10 μl of 1 μM oligo, 2.5 μl PNK buffer, 1 μl T4 PNK, 2 - 4 μl γ-32P ATP, 7.5 - 9.5 μl water) were incubated at 37°C for around an hour. To purify labelled DNA from unincorporated labelled nucleotides, Ilustra MicroSpin G-50 columns (Cytiva) were used. After the purification, the final volume increased to 30 – 33 μl. Labelled nucleotides were kept at 4°C.

3.4.4 Annealing of dsDNA substrates

To obtain dsDNA, 5’ end labelled substrates were annealed with unlabelled nucleotides. Reaction components (5 μl of 32P labelled oligo, 3 μl of 1 μM unlabelled oligo, annealing buffer, final volume 25 μl) were incubated at 95 - 100°C, for 5 - 7 min and gradually cooled to room temperature. Annealing reactions were performed in 10 mM Tris pH 7.5, 100 mM NaCl and 0.1 mM EDTA. Annealed substrates were kept at 4°C. DNA sequences that were used are listed in (Table 21).

To allow proper folding, ssDNA hairpins (YM164, StableCru_PC217) were also boiled and gradually cooled to the room temperature.

3.4.5 Blocking of DNA substrates with streptavidin

To block DNA ends, a 10 x molar excess of streptavidin was added to biotinylated DNA. Biotinylated substrates were prepared by Integrated DNA Technologies.

3.4.6 Preparation of G-quadruplex substrates

After 5’ end labelling (3.4.3) substrates were diluted in 100 mM KCl, 10 mM Tris-HCl, pH 7.5 and boiled for 5 min. After gradual cooling to room temperature, G-quadruplex containing substrates were stored at 4°C. Prior to nuclease activity reactions, they were additionally diluted in 100 mM KCl and 10 mM Tris-HCl.

3.4.7 Reaction of G-quadruplex substrates with NMM

N-methyl mesoporphyrin IX (NMM) was incubated with G-quadruplex substrates in RT for ~ 20 min (NMM final concentration = 1 μM, DNA final concentration = 0.25 μM). Reactions were carried out in 10 mM Tris pH 7.5 and 100 mM KCl. Fluorescence intensity was measured using SpectraMax iD5 plate reader (Excitation: 450 nm, Emission: 610 nm).

3.4.8 Nuclease activity assays with 32-P labelled DNA substrates and EXD2

In vitro assays with purified EXD2 and 32-P 5' labelled DNA were carried in a buffer containing: 25 mM Tris-HCl pH 7.5, 0.25 mg/ml BSA, 5 mM $MnCl_2$ and 0.5 mM DTT. The final reaction volume was 15 μl, where 3 μl was purified EXD2 diluted in gel filtration buffer (50 mM Tris pH 7.5, 300 mM NaCl, 10 % glycerol, 0.05 % Triton X-100 and 2 mM BME) and 2 μl was DNA substrate diluted in an annealing buffer (10 mM Tris pH 7.5, 100 mM NaCl and 0.1 mM EDTA). If not stated otherwise, the final concentration of labelled DNA in the reaction was 1 – 1.1 nM. The amount of EXD2 used for the reactions varied depending on the purification prep. Most of the experiments were performed with EXD2 WT (10xHis-TEV-hEXD2-3C-2xStrep) frozen at ~ 0.1 mg/ml (diluted 49 times before addition to the reaction) and EXD2 mutant (D108A E110) frozen at ~ 0.067 mg/ml (diluted 32.6 times). For these assays, final protein concentration was ~ 5.3 nM (M_w = 76,47 kDa; EXD2 WT stock concentration = 1307 nM). If another purification preparation was used in the reaction (proteins purified according to the same protocol), the final protein concentration was 1.6 nM. If not stated otherwise, reactions were carried at 37°C and stopped by addition of 15 μl (1:1) of stop solution (91 % formamide, 20.4 mM EDTA, loading dye), (denaturating loading dye composition, 5 ml: TBE 10x – 1ml, Ficol – 600 mg, urea – 2.1 g, bromophenol blue, xylene cyanol). Samples were boiled for 5 min and loaded directly on TBE-urea gels or stored at 4°C and boiled second time before loading.

3.4.9 Denaturating TBE-Urea gels

To analyse resection of DNA substrates, TBE-Urea gel electrophoresis was carried out. Gels were prepared in advance, wells extensively washed and pre-ran in TBE buffer. Samples were resolved at 1500 Volts and 15 – 20 Watts (for one/two gels) for 1 h 45 min – 2 h. If necessary, some gels were fixed (50 % methanol, 10 % acetic acid, 5 % glycerol) and dried. Results were analysed using Typhoon Fla9500.

Table 16. Composition of a denaturating TBE-Urea gel.

Denaturating 15 % acrylamide gel (25 ml)	
Urea pellets (Sigma-Aldrich)	10.5 g
TBE buffer (10 x)	2.5 ml
Acrylamide: bis-acrylamide (19:1), 40 % solution (Fisher Scientific)	9.4 ml
Water	5 ml
APS (1.6 %)	1.1 ml
TEMED	16.5 μl

3.4.10 Native acrylamide gels

To analyse the quality of oligonucleotides used for the reactions, DNA substrates were resolved in 12 % native gels (Table 17), at 100 V and 2 W, for 1 h. Gels were pre-ran in TBE buffer. Native loading dye (Ficoll) was added to the samples.

Table 17. Composition of a native gel.

Native 12 % acrylamide gel (10 ml)	
TBE buffer (10 x)	1 ml
Acrylamide: bis-acrylamide (19:1), 40 % solution (Fisher Scientific)	3 ml
Water	5.6 ml
APS (1.6 %)	440 μl
TEMED	6.6 μl

3.4.11 EMSA

To analyse binding or RPA protein to DNA substrates, EMSA was performed. Composition of the reaction buffer was the same as in EXD2 nuclease activity assays (3.4.8). Final reaction volume was 15 μl, where 2 μl was RPA purified by Dr Joanna Krwawicz. Protein was diluted in RPA purification buffer (50 mM Tris-HCl, pH 8, 100 mM NaCl, 5 % glycerol, 1 mM DTT). RPA was incubated with 32-P 5’ labelled DNA on ice. After 30 min, native loading dye (Ficoll) was added, and samples were loaded directly on the pre-ran gel. Native 6 % acrylamide gels were run on ice, in precooled tanks and cold TBE buffer, at 100 V and 2 W, for ~ 1h 20 min.

3.4.12 Nuclease activity assays with fluorescently labelled DNA

To anneal substrates composed from three DNA strands, with both 6-FAM Fluorescein and a Quencher, reaction components (6 μl of 10 μM FAM oligo, 10 μl of 10 μM Q oligo, 6 μl of 10 μM oligo, annealing buffer, final volume 50 μl) were boiled for 5 min and cooled gradually in a PCR machine. Annealed substrates were kept at 4°C.

Composition of the reaction buffer was the same as in (3.4.8), the final volume was 50 μl. Final protein concentration was ~ 131 nM (5 μl of undiluted purified EXD2), final concentration of fluorescently labelled DNA strand was 24 nM (2 μl of 2 x diluted). Reactions were prepared on ice, in black 96 well plates. Changes in fluorescence intensity were measured by SpectraMax iD5 plate reader (detection kinetics mode, Excitation 480 nm, Emission 520 nm). In the negative control, purified EXD2 was replaced with gel filtration buffer.

3.4.13 Nuclease activity assays on plasmid DNA

Circular ssDNA (ΦX174) was incubated with purified EXD2 at 37°C in a plasmid reaction buffer (20 mM HEPES KOH pH 7.5, 50 mM KCl, 100 μg/ml BSA, 5 % glycerol, 0.05 % Triton X-100, 0.5 mM DTT) supplemented with 5 mM Mn^{2+} / 5 mM Mg^{2+} / 1 mM ATP. Final concentration of ΦX174 substrate (NEB, stock: 1000 μg/ml, M_w = 1.70 x 10^6 Da) was 10 μg/ml. Reactions (10 μl) were stopped with 2.5 μl of a buffer containing 50 mM EDTA and 2 % SDS. After addition of a loading dye, samples were resolved in a 1.2 % TAE agarose gel stained with SYBR Gold (Invitrogen). If not stated otherwise, final EXD2 WT and MUT (D108A E110A) concentration was ~ 46 nM. The only exception is the 2 – 8 h time course where concentration of the proteins was ~ 110 nM (Figure 57 A).

3.5 Structural analysis

3.5.1 Preparation of EXD2 for structural studies

Full-length EXD2 purified according to (3.3.2) was frozen in a gel filtration buffer (50 mM Tris pH 7.5, 300 mM NaCl, 10 % glycerol, 2 mM BME, 0.05 % Triton X-100). To use purified protein in the negative stain analysis, buffer was exchanged using a Zeba Spin Desalting column (5 ml). This column was centrifuged (1000 g, 2 min) to remove storage solution and equilibrated three times in a structural buffer (25 mM Tris pH 7.5, 100 mM NaCl, 5 % glycerol, 0.5 mM BME, no Triton X-100). Between each equilibration step, the column was centrifuged (1000 g, 3 min). Purified protein (90 μl) was loaded on the column, centrifuged (1000 g, 2 min) and collected as the flow through (150 μl).

3.5.2 Negative staining of EXD2 (Dr Fabienne Beuron)

Quantifoil grids R1.2/1.3 (SPT Labtech) supporting a continuous carbon layer were glow-discharged in a PELCO easiGlow unit for 1 minute. EXD2 was adsorbed on the grid for 1 min, blotted, washed twice with water, washed with 2% uranyl acetate, incubated on a drop of 2% uranyl acetate for 30 s, blotted and left to dry.

3.5.3 Data collection of negatively stained EXD2 (Dr Fabienne Beuron)

The negatively stained grids were examined using a Tecnai F20 transmission electron microscope (ThermoFisher Scientific) operating at an accelerating voltage of 200 kV and equipped with a CMOS 4kx4k F516 camera (TVIPS, Germany). Micrographs were acquired at a nominal magnification of x50,000 corresponding to a pixel size of 1.732 Å using the EM-Tools software (TVIPS, Germany).

219 micrographs were selected for processing in Relion (Scheres, 2012). 374 particles were manually picked, then aligned to generate 2D class averages used for template-based autopicking in Relion 3. A total of 18,876 particles were autopicked and extracted with a box size of 80 pixels. Particles were then subjected to 6 iterative rounds of 2D classification, and particles which did not classify well in each round were discarded. The remaining 9,416 particles were subjected to a final round of 2D classification, and representative class averages were selected to generate an *ab initio* 3D reconstruction. This map was then used as a reference volume for a round of 3D classification of the particles in 3 groups. The 3D reconstruction shown here was calculated from 5807 particles.

For truncated EXD2 (without aa 2-37) 137 micrographs were collected, and 1486 particles picked manually and submitted to 2D classification in Relion.

3.5.4 Alpha-fold model fitting of EXD2 (Dr Fabienne Beuron)

The exonuclease domain crystal structure (PDB 6K1A) was fitted into the 3D volume using ‘fit in map’ in Chimera (Pettersen et al., 2004). The exonuclease domain of the AlphaFold model was then aligned to the fitted exonuclease domain in Pymol. This positions part of the predicted structure away from the density. The endonuclease domain (residues 344-621) was fitted separately as a single rigid body in Chimera as described in the Results section.

Results

4. Protein purification

4.1 Purification of EXD2

4.1.1 Purification of EXD2 - overview

To date the majority of analyses conducted on EXD2 *in vitro*, both biochemical (Broderick et al., 2016), (Silva et al., 2018) and structural (Park et al., 2019), have been performed using truncated versions of the protein. Truncated EXD2 has improved solubility and purification yield compared to the full-length. To fully characterize EXD2, I needed to generate the full-length EXD2 construct expressed with higher efficiency. So far only protocols describing EXD2 purification from bacteria have been established (Broderick et al., 2016), (Silva et al., 2018), (Park et al., 2019).

I decided to use insect cells as the expression system as they exhibit more accurate protein folding and post-translational modifications that are not performed in bacteria (Fernandez, J.M. & Hoeffler, J.P., eds., 1999). Protein folding and the presence of post-translational modifications may affect activity and structure of purified hEXD2, hence it is best to use an expression system as closely related to human cells as possible. Additionally, the expression of full-length hEXD2 in insect cells may lead to a higher purification yield than the one obtained from bacteria.

To define how the active EXD2 regulates DNA resection, I needed the inactive negative control. As the (D108A, E110A) EXD2 exhibits exonuclease-dead phenotype (Broderick et al., 2016), I purified both WT and (D108A E110A) proteins.

4.1.2 Optimization of full-length EXD2 purification protocol

Expression optimisation began with the generation of baculovirus encoding 10xHis-TEV-hEXD2, used for protein expression in insect cells. The molecular weight of the recombinant protein monomer was 72.6 kDa (the untagged hEXD2 is ~ 70.4 kDa). Results of test expression showed low solubility of full-length EXD2 (Figure 23), therefore buffers used in the following purifications were supplemented with non-ionic detergent (Triton X-100). It penetrates phospholipid bilayers (Johnson M., 2013) increasing solubility of membrane proteins like EXD2. Non-ionic detergents break protein-lipid associations, but not protein-protein interactions, allowing solubilization of membrane proteins in their native and active form.

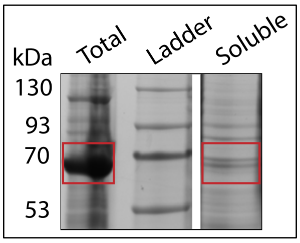


Figure 23. Purification of 10xHis-TEV-hEXD2 mutant (D108A E110A) without any detergent, Coomassie InstantBlue stained gel. Total – lysate before centrifugation; Soluble – soluble lysate after centrifugation. Red boxes indicate decrease of protein solubility after centrifugation.

The trial purification of 10xHis-TEV-hEXD2 using HisPur Ni-NTA resin did not result in high purity (Figure 24 A), therefore I introduced a cleavable 3C-2xStrep tag on the C-terminus. The molecular weight of the 10xHis-TEV-hEXD2-3C-2xStrep monomer was 76.5 kDa. The outcome of this 10xHis-TEV-hEXD2-3C-2xStrep protein purification using Strep-Tactin Superflow resin, showed less contaminants in the sample (Figure 24 B).

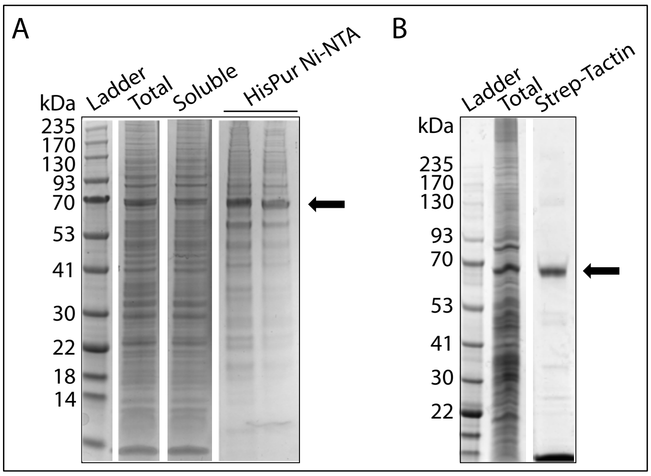


Figure 24. Trial purification of 10xHis-TEV-hEXD2 WT (A) and 10xHis-TEV-hEXD2-3C-2xStrep WT (B), Coomassie InstantBlue stained gels are shown. Arrows indicate presence of EXD2. Both lysis buffers were supplemented with 1% Triton X-100. Total – lysate before centrifugation; Soluble – soluble lysate after centrifugation. A) Purification using HisPur Ni-NTA resin: HisPur Ni-NTA – elution with buffer containing 250 mM imidazole. B) Purification using Strep-Tactin Superflow high-capacity resin: Strep-Tactin – elution by boiling the beads.

For large scale protein purification, I used both types of affinity chromatography (Strep-Tag affinity followed by His-Tag affinity) performed on an AKTA protein purification system. As high concentration of detergent in the final purification step may affect activity of the protein, I decided to decrease amount of Triton X-100 throughout the purification process. 1% Triton X-100 was present in the lysis, decreased to 0.5% during Strep-Tag chromatography. Strep-Tag elution with buffer supplemented with 0.1% Triton X-100, resulted in decrease in purification yield (Figure 25 B). Therefore, in the final protocol, the 0.5% Triton X-100 was present in the sample until loading the protein on the gel filtration column.

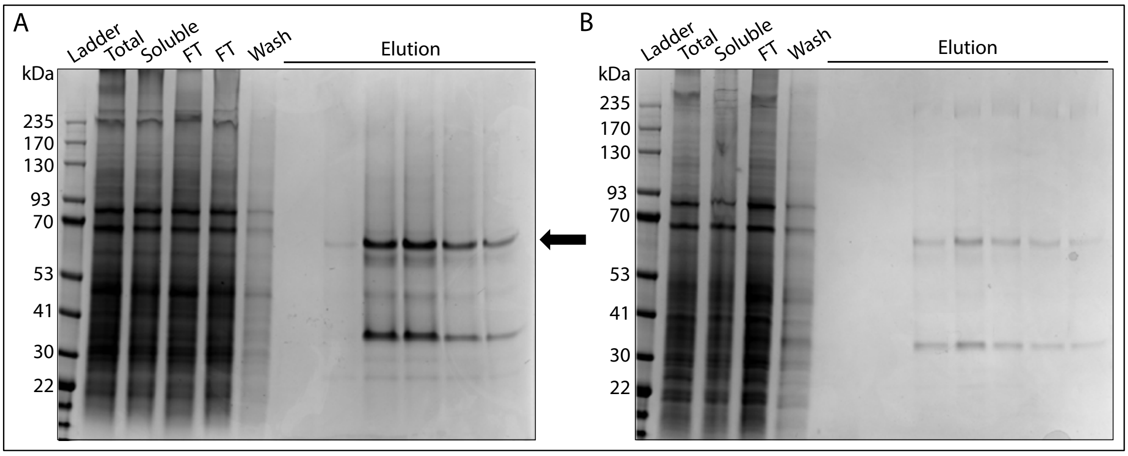


Figure 25. Triton X-100 affects 10xHis-TEV-hEXD2-3C-2xStrep WT purification yield. Purification results from two different pellets of Hi5 cells, Coomassie InstantBlue stained gels. Total – lysate before centrifugation; Soluble – soluble lysate after centrifugation; FT – flow through StrepTrap HP column; Wash – wash of StrepTrap HP column; Elution – elution from StrepTrap HP column. A) 1% Triton X-100 used during the lysis; 0.5% Triton X-100 used in wash and elution steps. 14 μl of elution loaded on the gel. B) 1% Triton X-100 used during the lysis; 0.1% Triton X-100 used in wash and elution steps. 13 μl of elution loaded on the gel.

To confirm the presence of EXD2 after one of purification trials, I used anti-EXD2 antibody (Figure 26 A). Western blotting with anti-His Tag and anti-Strep antibodies indicate presence of C-terminal EXD2 degradation products (Figure 26 B and C).

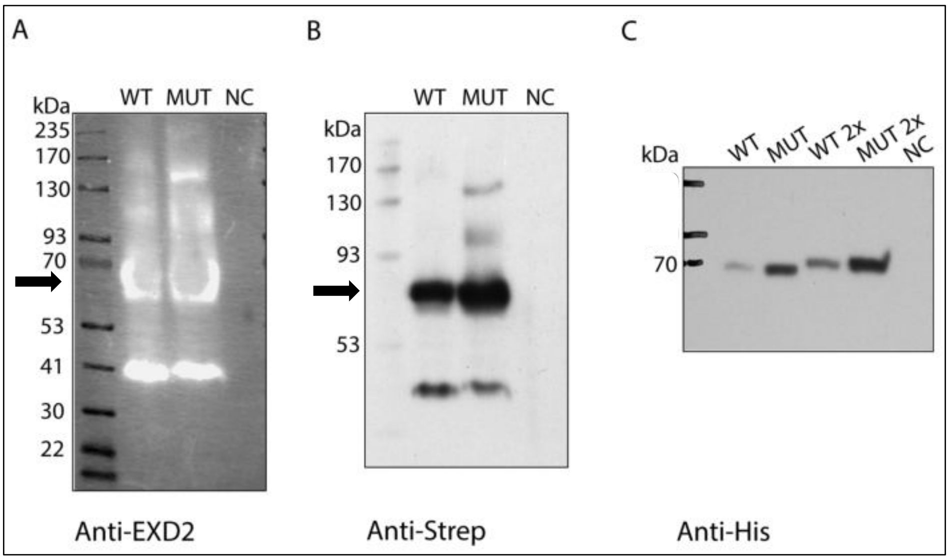


Figure 26. Chemiluminescence analysis of purified 10xHis-TEV-hEXD2-3C-2xStrep, WT and MUT (D108A E110A). NC – negative control - soluble lysate of uninfected insect cells. Results of incubation with (A) anti-EXD2 antibody, (B) anti-Strep antibody and (C) anti-His Tag antibody. Full-length EXD2 (indicated by arrows) and C-terminal degradation products are present in the sample.

During first trials of EXD2 purification, I used benzonase – a recombinant endonuclease that digests DNA and RNA to nucleotides of 2-5 base pairs (Grøntved et al., 2012). It is commonly used to decrease DNA contamination during protein purification. However, as it may affect results of EXD2 nuclease activity assays, I decided not to use it in final purification protocol.

4.1.3 Purification of full-length EXD2 – final protocol

10xHis-TEV-hEXD2-3C-2xStrep WT and MUT (D108A E110A)

Hi5 insect cells were cultured in 27°C for 3 to 4 days after the viral infection. Harvested cell pellets were washed with PBS prior freezing in -20°C. After defrosting, pellet from 1 litre of cells was resuspended in 200 ml of the EXD2 lysis buffer. Cells were lysed using Dounce homogenizer, pulse sonicated and centrifugated. The supernatant volume was divided into two parts, each loaded and eluted from StrepTrap HP column (5 ml) separately to improve purification yield. Peak fractions eluted with desthiobiotin (Figure 27 A, "Strep I, Strep II") were combined and loaded on a HisTrap HP (5 ml) column. The molecular weight of the 10xHis-TEV-hEXD2-3C-2xStrep is 76.5 kDa. However, purified protein was observed on the gel just below the 70 kDa weight standard band. The electrophoresis molecular weight standard (Blue Wide Range Protein Ladder, Cleaver Scientific), used in all gels in the thesis, has corresponding molecular weight when resolved in a Bis-Tris 10 % acrylamide gel in MOPS buffer. As all purified proteins were resolved in Bis-Tris 4-12 % acrylamide gels in MOPS buffer, the M_w of the ladder bands may slightly differ in these conditions (exact M_w for these conditions was not provided by the producer).

As buffers were supplemented with 0.5 % Triton X-100, which creates high absorbance background at 280 nm, the elution profiles from StrepTrap HP and HisTrap HP columns were not included in (Figure 27). EXD2 was eluted from the HisTrap HP column using buffer with increasing imidazole concentration. Chosen fractions (eluted with 198 mM – 400 mM imidazole) were combined and concentrated using Sartorius Vivaspin 20 Centrifugal Concentrator (MWCO 30 kDa). Concentrated protein was loaded on HiLoad 16/60 Superdex 200 column. During the size exclusion, concentration of Triton X-100 was decreased to 0.05%. The first peak in the gel filtration elution profile (Figure 27 C), indicated by the blue lines in the elution plot (Figure 27 C) and in the Coomassie stained gel (Figure 27 B), corresponded to protein aggregates. The second peak contained not only purified EXD2, but possibly a high amounts of Triton X-100, which increased absorbance values at 280 nm (Figure 27 C). The first part of this peak, indicated by red lines in the elution plot (Figure 27 C) and in the Coomassie stained gel (Figure 27 B), were combined, concentrated and flash frozen in liquid nitrogen, in buffer containing 50 mM Tris pH 7.5, 300 mM NaCl, 10% glycerol, 0.05% Triton X-100 and 2 mM BME. Based on the M_w, from chromatographic separation for the standard proteins on HiLoad 16/60 Superdex 200 columns (according to GE Healthcare), frozen fractions contained both EXD2 monomers (76.5 kDa) and dimers. The second part of the peak, indicated by the green lines in the elution profile (Figure 27 C) and in the Coomassie stained gel (Figure

27 B), most likely contained high concentrations of Triton X-100 (the absorbance value is high, whereas intensity of the band is much weaker). Detailed purification protocol is described in the Methods section (3.3.2). Results of 10xHis-TEV-hEXD2-3C-2xStrep WT purification, which was then used for majority of the biochemical assays, are shown in (Figure 27). The yield of shown purification was around 0.38 mg (0.1 mg/ml; 3.8 ml).

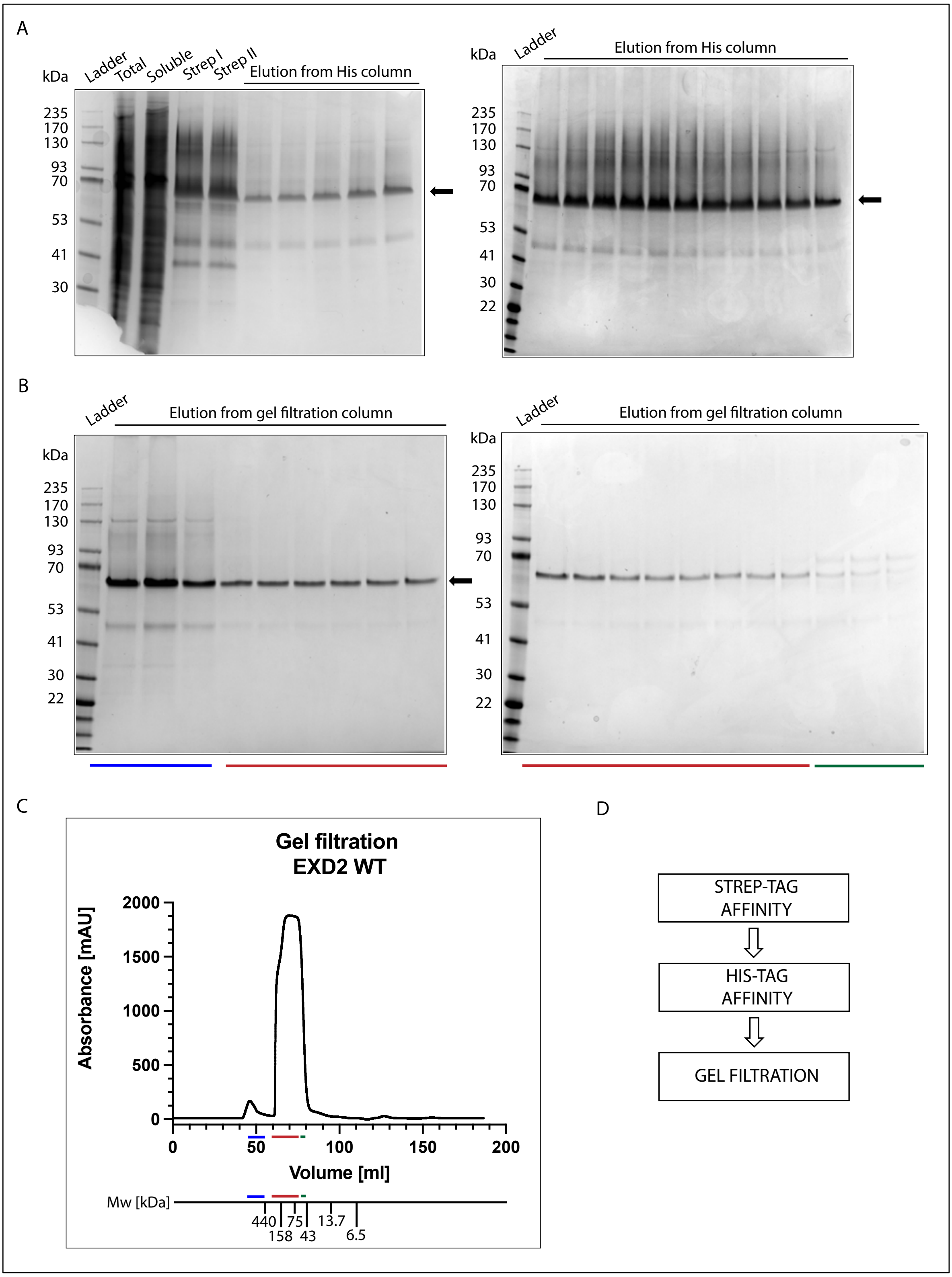


Figure 27. Purification of WT 10xHis-TEV-hEXD2-3C-2xStrep from 1L of Hi5 cells. A) and B) Coomassie InstantBlue stained gels after each purification step. Total – lysate before centrifugation; Soluble – soluble lysate (centrifugation supernatant) loaded on the StrepTrap HP column twice (as the first and the second half); Strep I – first half of the lysate eluted from StrepTrap HP column; Strep II – second half of the lysate eluted from StrepTrap HP column; Elution from His column - Fractions eluted from HisTrap HP column with increasing imidazole concentration. Arrows indicate presence of EXD2. C) Elution profile from HiLoad 16/60 Superdex 200 column - the final purification step. Blue, red and green lines correspond to the eluted fractions on Coomassie stained gels (B) and to the molecular weight standard below the plot. M_w standard is based on a chromatographic separation

for standard proteins on HiLoad 16/60 Superdex 200 columns according to GE Healthcare. D) Main steps of the purification.

Results of 10xHis-TEV-hEXD2-3C-2xStrep MUT (D108A E110A) purification, which was further used as a negative control for majority of the biochemical assays, are shown on (Figure 28). The purification was performed according to the same protocol as for the 10xHis-TEV-hEXD2-3C-2xStrep WT. Even though the same Triton X-100 concentration was used, here the loss of protein after centrifugation is more prominent than in the purification of EXD2 WT. The amount of EXD2 MUT (D108A E110A) was much lower in the supernatant (Figure 28 A, comparison between ‘Total’ and ‘Soluble’ lines in the Coomassie stained gel), which lead to a decreased amount of the protein in the subsequent steps. The shape of the elution profile from the final purification step (Figure 28 B) highly resembles the result obtained for EXD2 WT purification (Figure 27 C). The same fractions, indicated by red lines in the elution profile (Figure 28 B) and in the Coomassie stained gel (Figure 28 A), were concentrated and frozen. Based on the M_w marker, frozen fractions should contain monomers (76.5 kDa) and dimers of EXD2 MUT (D108A E110A). The yield of mutant purification was around 0.2 mg (0.067 mg/ml; 2.9 ml), which is almost two times lower than the one obtained for EXD2 WT. The high absorbance intensity of the main peak (Figure 28 B) is most likely caused by the Triton X-100 present in the sample.

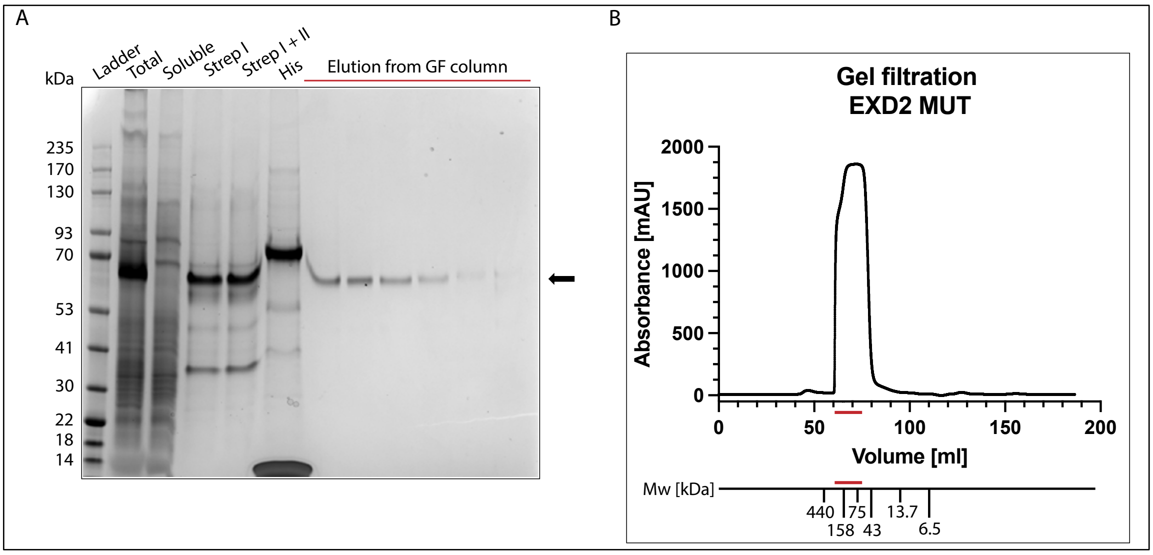


Figure 28. Purification of 10xHis-TEV-hEXD2-3C-2xStrep mutant (D108A E110A) from 1L of Hi5 cells. A) Coomassie InstantBlue stained gel showing results after each purification step. Total – lysate before centrifugation; Soluble – soluble lysate (centrifugation supernatant) loaded on the StrepTrap HP column twice (as the first and the second half); Strep I – first half of the lysate eluted from StrepTrap HP column; Strep I + II – combined elution from StrepTrap HP column. His – combined and concentrated fractions after elution from HisTrap HP column. Elution from GF column – fractions eluted from HiLoad 16/60 Superdex 200, corresponding to the (B) elution profile. Arrow indicates presence of EXD2. B) Elution profile from HiLoad 16/60 Superdex 200 column - the final purification step. Red line on the plot corresponds to the eluted fractions shown on the gel (A) and to the

molecular weight standard below the plot. M_w standard is based on a chromatographic separation for standard proteins according to GE Healthcare.

To assess the concentration of purified proteins, I used Coomassie InstantBlue staining and compared EXD2 bands intensity with BSA protein standard (Figure 29). The absorbance at 280 nm could not be used to determine proteins concentration, as Triton X-100 was present in the purification buffer. Concentration of frozen protein is ~ 0.1 mg/ml for 10xHis-TEV-hEXD2-3C-2xStrep WT, and ~ 0.067 mg/ml for (D108A E110A) mutant.

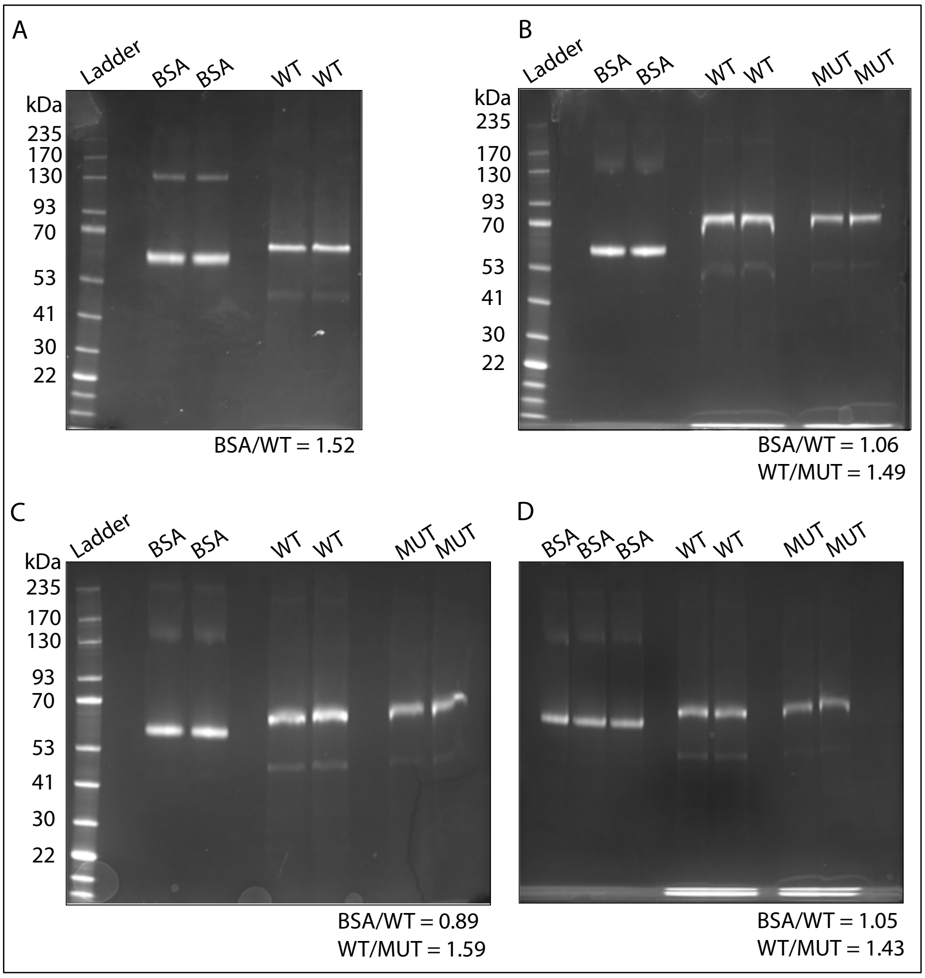


Figure 29. Assessing concentration of purified 10xHis-TEV-hEXD2-3C-2xStrep WT and mutant (D108A E110A), Coomassie InstantBlue stained gels. Negative images used for protein quantification. A)B)C)D) Experimental repeats performed on the same batch of protein. Concentration of BSA is 0.1 mg/ml. Bands intensity ratio between BSA and WT EXD2 is shown for each repeat. The mean BSA/WT ratio is 1 – 1.1; the mean WT/MUT ratio is 1.5.

The quality of purified EXD2, both WT and (D108A E110A), was assessed by liquid chromatography mass spectrometry (LC-MS/MS) by Dr Lu Yu (Table 22), (Table 23). The proteins were digested by trypsin and the resulting mixture of peptides analysed by LC-MS/MS. The data was processed and searched against Uniprot protein databases and the in-house contaminate database. As the result of tryptic digest, 79 peptide sequences were detected for EXD2 WT sample and 67 for EXD2 MUT (D108A E110A) sample. The coverage values of 89 % (for WT) and 86 % (for MUT) indicate that high percentage of the EXD2 sequence was covered by identified peptides. The main contaminants were heat shock proteins and tubulins, which should not affect activity of EXD2.

4.1.4 Purification of truncated EXD2

10xHis-TEV-hEXD2-3C-2xStrep, without the transmembrane domain (without 2-37 residues)

EXD2 contains an N-terminal transmembrane domain (Park et al., 2019) and the primary sequence of EXD2 (Figure 94) shows many N-terminal hydrophobic and small nonpolar amino acids. Therefore, to increase EXD2's solubility for use in structural studies, I generated truncated 10xHis-TEV-hEXD2-3C-2xStrep construct without N-terminal transmembrane domain (aa 2-37). The lack of this transmembrane domain improved protein solubility and increased the purification yield (Figure 30). Hence, Triton X-100 was not required during the purification and the protein was used in the structural analysis (Figure 103).

The purification procedure for the truncated EXD2 was the same as for the full-length protein (3.3.2), but no detergent was used. Lack of the detergent is important in the context of structural studies, as Triton X-100 creates undesirable background affecting electron microscopy analysis (discussed in more detail in 6.5).

Results of 10xHis-TEV-hEXD2-3C-2xStrep purification from 1L of infected Hi5 cells are shown in (Figure 30). Proteins present in the StrepTrap HP elution peak (Figure 30 B) were subsequently purified using a HisTrap HP column. Fractions indicated by the red lines in the elution profile (Figure 30 C) and in the Coomassie stained gel (Figure 30 E) were combined and loaded on the gel filtration column. The first peak, indicated by the blue lines in the elution plot (Figure 30 D) and in the Coomassie stained gel (Figure 30 D), most likely contains protein aggregates (M_w over 440 kDa). Fractions from the second peak, indicated by the red lines (Figure 30 D, G), were concentrated and flash frozen in a buffer containing 50 mM Tris pH 7.5, 300 mM NaCl, 10% glycerol and 2 mM BME. According to the M_w standard they contained mostly protein monomers. It is possible that the lack of Triton X-100 in the sample affects the shape of the elution profile, hence the difference between gel filtration plots of the full-length (Figure 27) and the truncated protein (Figure 30). The yield of purification of truncated EXD2 was around 0.94 mg (0.323 mg/ml; 2.9 ml; A_{280}=0.309; 0.382; 0.358; 0.373; M_w=72.5 kDa, E=79840 $M^{-1}cm^{-1}$). It was 2.5 times higher than for full-length WT EXD2.

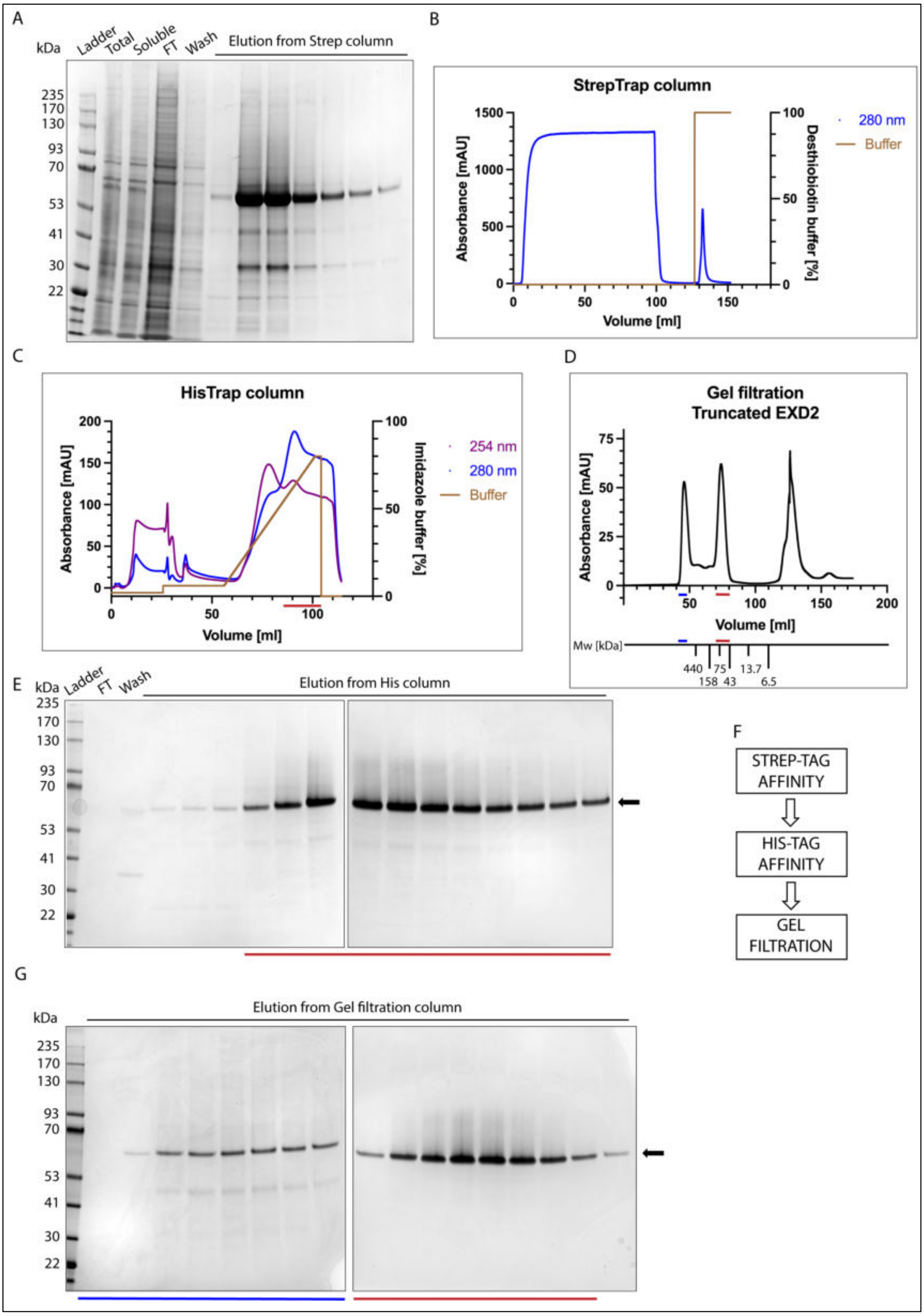


Figure 30. Purification of truncated 10xHis-TEV-hEXD2-3C-2xStrep WT (without 2-37 residues) from 1L of Hi5 cells. A)E)G) InstantBlue Coomassie stained gels showing results after each purification step. Arrows indicate presence of EXD2. A) Total – lysate before centrifugation; Soluble – soluble lysate (centrifugation supernatant) loaded on StrepTrap HP; FT – flow through the StrepTrap HP column; Wash – wash of the StrepTrap HP column. Elution from StrepTrap HP column. B) Elution profile from StrepTrap HP. E) FT – flow through the HisTrap HP column; Wash – wash of the HisTrap HP column; Elution from HisTrap HP column – red line from the Coomassie stained gel corresponds

to fractions indicated by the red line in (C) elution profile from HisTrap HP column. C) Absorbance at 254 nm indicates high concentration of DNA in the sample. Brown line shows concentration (%) of 0.5 M imidazole in the buffer. G) Fractions eluted from HiLoad 16/60 Superdex 200 column – blue and red lines correspond to the fractions indicated by these colours in (D) elution profile from HiLoad 16/60 Superdex 200. D) M_w standard is based on a chromatographic separation for standard proteins according to GE Healthcare. F) Main purification steps.

5. Biochemical characterization of full-length EXD2

5.1 Basic characterization of EXD2

5.1.1 EXD2 characterization - overview

EXD2 is a 3'-5' exonuclease (Broderick et al., 2016). *In vitro* nuclease activity assays with full-length EXD2 purified from bacteria showed exonuclease activity on ssDNA (50 nt) and very weak activity on blunt ended dsDNA (50 bp), (Broderick et al., 2016). The truncated EXD2 (aa 76 – 564; tEXD2), still containing the exonuclease domain, behaved indistinguishably from the full-length protein in several *in vitro* resection assays (both proteins were purified from bacteria), (Broderick et al., 2016). Additionally, tEXD2 had weak activity towards dsDNA with a resected 3' end (50/60 bp), (Broderick et al., 2016). Truncated EXD2 (aa 61 – 621), expressed in *E. coli,* was also purified by (Silva et al., 2018). It exhibited nuclease activity on ssDNA, dsDNA (40 bp) and dsDNA with a recessed 3' end (16/35 nt) (Silva et al., 2018).

In vitro nuclease activity assays performed by (Broderick et al., 2016) revealed that manganese ions are required for EXD2 activity on DNA. Experiments conducted by (Silva et al., 2018) showed that replacing manganese (Mn) with magnesium (Mg) affects EXD2 activity in a different way depending on the substrate. On ssDNA and dsDNA the activity is completely blocked, on RNA-DNA hybrids it is reduced, whereas on ssRNA it is nearly the same as in the presence of Mn^{2+} (Silva et al., 2018). Data presented by (Park et al., 2019) confirmed that EXD2 exhibits nuclease activity on ssDNA, only in the presence of Mn^{2+}. Replacing Mn^{2+} with Mg^{2+}, Zn^{2+} or Ca^{2+} completely blocked the activity (Park et al., 2019).

To fully characterise hEXD2 *in vitro*, I performed nuclease activity assays with full-length wild type and exonuclease deficient mutant EXD2. Proteins were expressed in insect cells and purified as described in (3.3.2).

5.1.2 EXD2 nuclease activity in the presence of different metal ions

I started with determining EXD2 ion dependency. To this end I analysed EXD2 nuclease activity on 50 bp dsDNA (BS40* + UA1) in the presence of different metal ions. I tested (0.01 – 0.1 – 1 – 5 – 10 mM) Mn^{2+}, Mg^{2+}, Zn^{2+} and Ca^{2+}. As expected, EXD2 is a manganese dependent nuclease (Figure 31). It is in line with the fact that catalytic core residues coordinate Mn^{2+} in the exonuclease active site crystalized by (Park et al., 2019). The presence of Mg^{2+} resulted in extremely weak EXD2 activity at 1 or 5 mM but not 10 mM concentration, whereas Zn^{2+} and Ca^{2+} did not promote any digestion (Figure 31). EXD2 digestion efficiency has been quantified by measurements of a substrate band intensity in the presence of different concentrations of Mn^{2+} and Mg^{2+} (Figure 31 B). This quantification method has some limitations as it may not detect subtle kinetic differences, it is prone to human made errors (pipetting) and it can be affected by the quality of self-made gels (sharpness of the bands). A more accurate method of digestion quantification, which is based on automatic measurements of changes in fluorescence intensity by SpectraMax iD5 plate reader (for substrates containing both a fluorophore and a quencher), is described in (5.8.2). However, it requires higher amounts of the protein for each reaction and is more costly due to the substrates price.

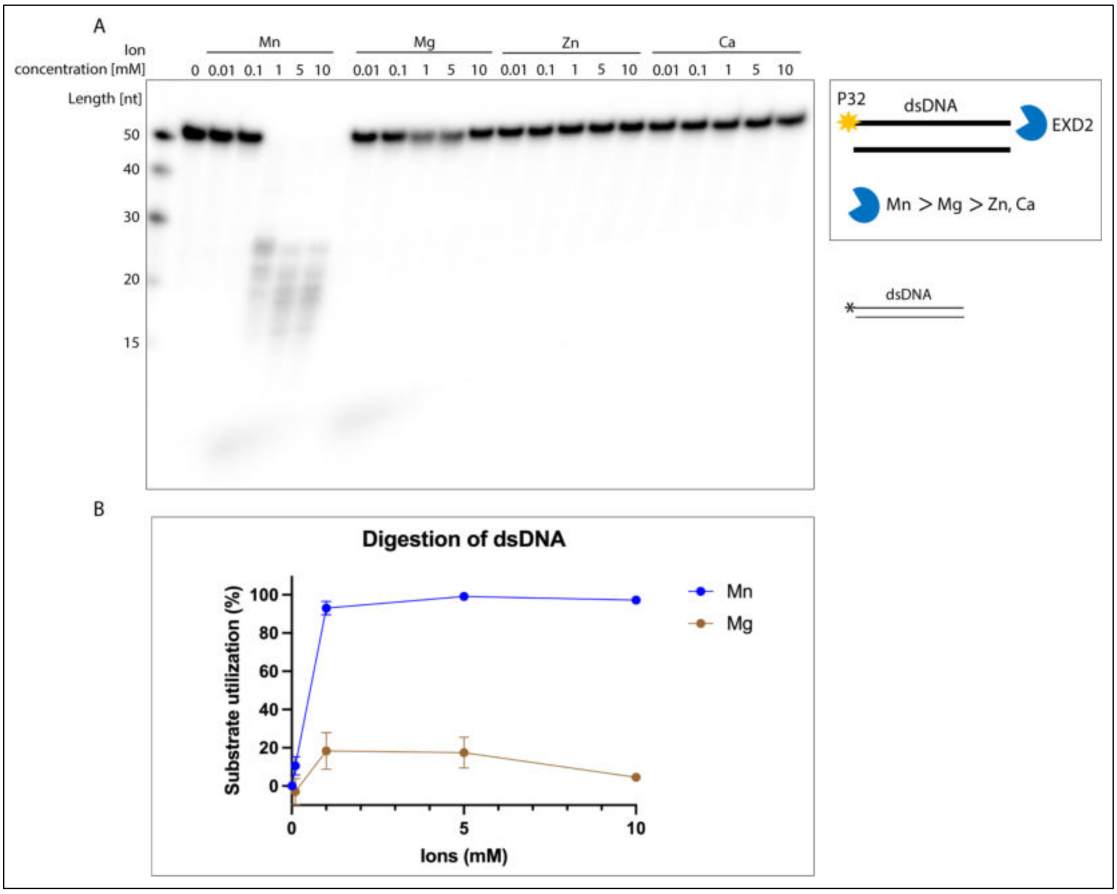


Figure 31. EXD2 WT is manganese dependent exonuclease. Mg^{2+} stimulates extremely weak EXD2 activity, whereas Zn^{2+} and Ca^{2+} did not promote any activity. A) 5'-radiolabelled 50 bp dsDNA substrate (~ 1 nM) was incubated with EXD2 WT (5.3 nM) for 30 min at 37°C. Samples were resolved in TBE-urea polyacrylamide gel and visualised by phosphor imaging. Representative image of one of three repeats. B) Quantification of experiments such as in panel (A) for reactions supplemented with either Mn^{2+}or Mg^{2+} ; n = 2; error bars; SEM. If no error bar is present, it means that it is shorter than the size of the symbol.

5.1.3 EXD2 nuclease activity on a single stranded DNA

To test activity of purified proteins (10xHis-TEV-hEXD2-3C-2xStrep, both WT and D108A E110A mutant) on ssDNA, I used the same 5'-radiolabelled ssDNA substrate (50 nt, BS40) as (Broderick et al., 2016) used to establish that EXD2 is a 3'-5' exonuclease.

Digestion of the labelled substrate by EXD2 WT, contrary to (D108A E110A) mutant, resulted in oligo truncation and its faster migration through the acrylamide gel (Figure 32). The length of the 5’-radiolabelled products gradually decreases over time indicating 3’-5’ exonuclease activity of the protein.

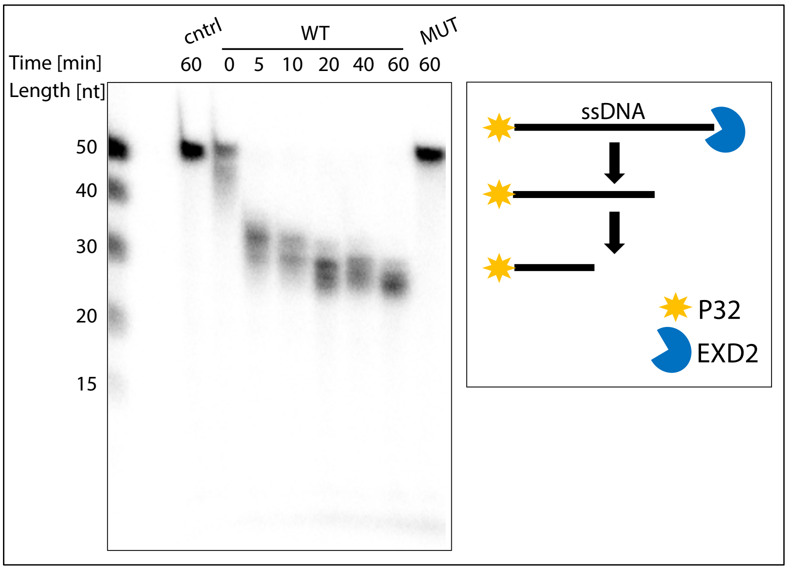


Figure 32. Purified WT EXD2, contrary to (D108A E110A) mutant, displays 3’-5’ exonuclease activity in vitro on ssDNA. 5’-radiolabelled 50 nt DNA substrate (~ 1 nM) was incubated with EXD2 WT (1.6 nM) or EXD2 (D108A E110) MUT (1.6 nM) for indicated amounts of time at 37℃. Samples were resolved in TBE-urea polyacrylamide gel and visualised by phosphor imaging. The reaction buffer contained 5 mM $MnCl_2$. Representative image of one of three repeats.

To analyse EXD2’s preferences towards DNA length, I used 4 substrates of the same 3’ end sequence but different lengths (20, 40, 50 and 73 nt). In the first 5 min of reaction, the substrates were shortened by up to 4, 18, 22 and 50 nucleotides respectively (Figure 33). It suggests that in this length range, the longer substrate was preferable. However, longer substrates have higher likelihood of self-duplex and hairpin formation, which may affect digestion efficiency. The shortest digestion product was around 15 nt long, with 20 nt being enough to initiate digestion. It is possible that EXD2 requires ~ 15 nt to bind ssDNA, hence it cannot digest the substrates any further.

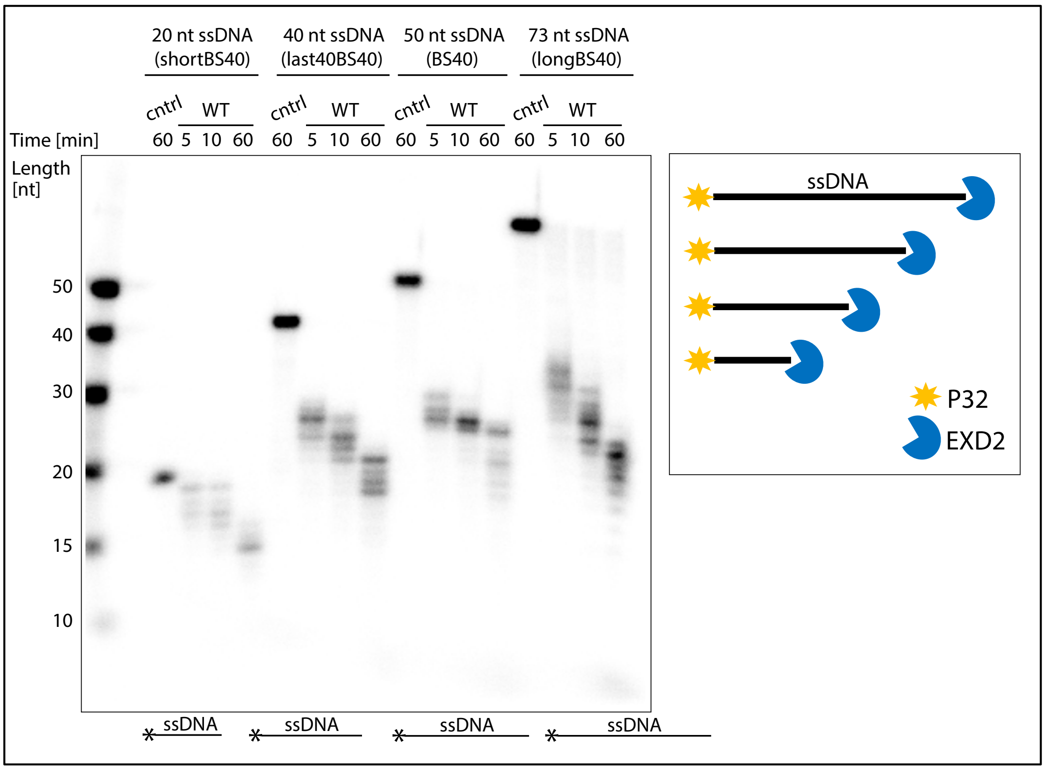


Figure 33. WT EXD2 efficiently digests single stranded DNA of different length (20 – 73 nt) in vitro. 5'-radiolabelled DNA substrates (~ 1 nM) were incubated with EXD2 WT (5.3 nM) for indicated amounts of time at 37℃. Samples were resolved in TBE-urea polyacrylamide gel and visualised by phosphor imaging. The reaction buffer contained 5 mM $MnCl_2$. Representative image of one of two/three repeats.

5.1.4 EXD2 nuclease activity on double stranded DNA

To characterise EXD2's exonuclease activity on dsDNA I used three types of substrates: blunt ended (50 bp; BS40*+UA1), with a 5' overhang (50/60 nt; BS40*+UA1_60) and with a 3' overhang (60/50 nt; BS40_10* +UA1), (Figure 34). The 50 nt strand had the same sequence as 50 nt ssDNA used in (Figure 32). I labelled one strand of each substrate on the 5' terminus. Both, blunt ended and 5' overhang substrates are resected with high efficiency with a similar digestion pattern (Figure 34, A), with the shortest digestion product being around 15 nt. The substrate with a 3' overhang is also shortened by ~35 nt, leaving a product of ~25 nt long (Figure 34, B).

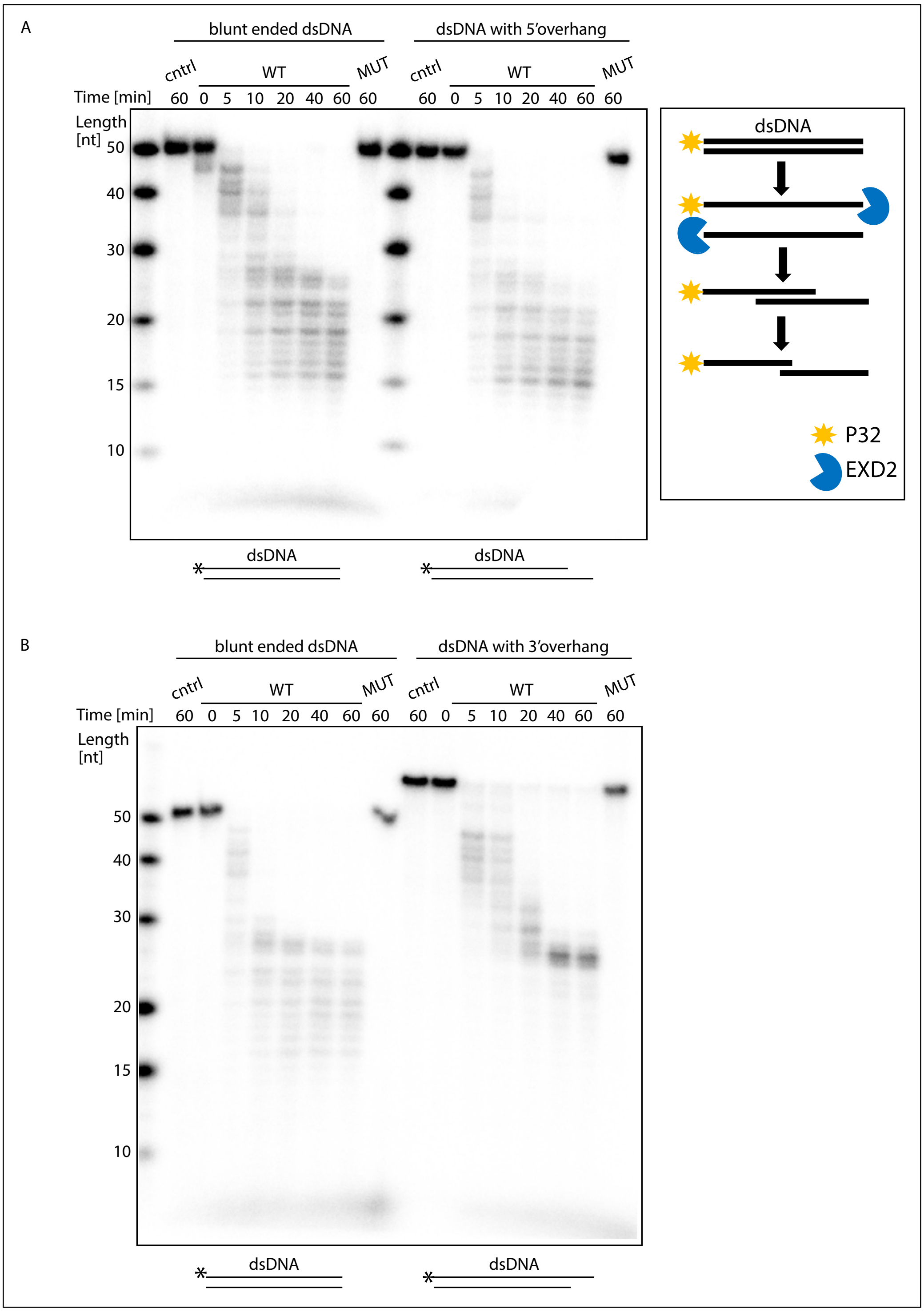


Figure 34. Purified WT EXD2, contrary to (D108A E110A) mutant, displays 3’-5’ exonuclease activity in vitro on dsDNA with blunt ends, B) 3’ overhang and A) 5’ overhang. A) 5’-radiolabelled DNA substrates (50 bp; 50/60 nt; ~ 1 nM) were incubated with EXD2 WT (1.6 nM) or EXD2 (D108A E110) MUT (1.6 nM) for indicated amounts of time at 37°C. B) 5’-radiolabelled DNA substrates (50 bp; 60/50 nt; ~ 1 nM) were incubated with EXD2 WT (5.3 nM) or EXD2 (D108A E110) MUT (5.3 nM) for indicated amounts of time at 37°C. Samples were resolved in TBE-urea polyacrylamide gel and visualised by phosphor imaging. The reaction buffer contained 5 mM $MnCl_2$. Representative image of one of three repeats.

To analyse EXD2’s preferences towards dsDNA of different lengths (20, 40, 50 and 73 nt), I used substrates of the same sequence and length as in (Figure 33), but double stranded (Figure 35). In the first 5 min of the reaction, the dsDNA substrates were shortened by up to 4, 23, 31 and 43 nucleotides respectively (Figure 35). Similarly, as for ssDNA, the longer dsDNA the higher the digestion efficiency in this length range. The shortest product of dsDNA digestion was around 12 nt long.

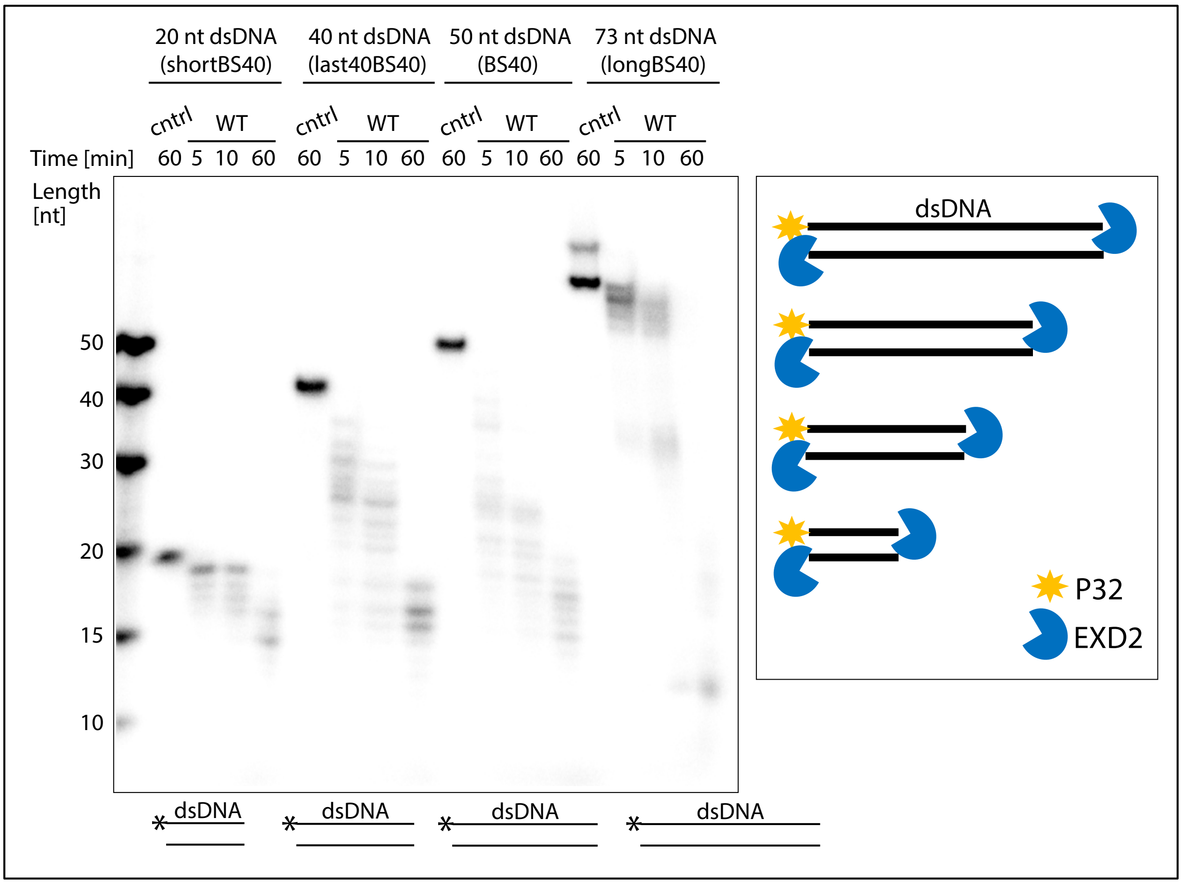


Figure 35. WT EXD2 efficiently digests double stranded DNA of different length (20 – 73 nt) in vitro. 5'-radiolabelled DNA substrates (~ 1 nM) were incubated with EXD2 WT (5.3 nM) or EXD2 (D108A E110) MUT (5.3 nM) for indicated amounts of time at 37°C. Samples were resolved in TBE-urea polyacrylamide gel and visualised by phosphor imaging. The reaction buffer contained 5 mM $MnCl_2$. Representative image of one of three repeats.

To examine what is EXD2's preferable substrate, I performed reactions on ssDNA and dsDNA using different concentrations of the protein (Figure 36). As the reaction kinetics are very fast (no substrate band is present at the smallest protein concentration), I decided to compare the increase of the digestion products ≤ 20 nt. The signal at the bottom of the gel (< 2 nt) is probably a background, hence it was not included in the quantification. Comparison of increase of the products ≤ 20 nt, indicated that EXD2 digested dsDNA faster than the ssDNA of the same sequence. This is contrary to the experiments performed with EXD2 purified from bacteria (Broderick et al., 2016), which exhibits preference towards ssDNA. It is highly possible that the observed difference is caused by using a different protein expression system.

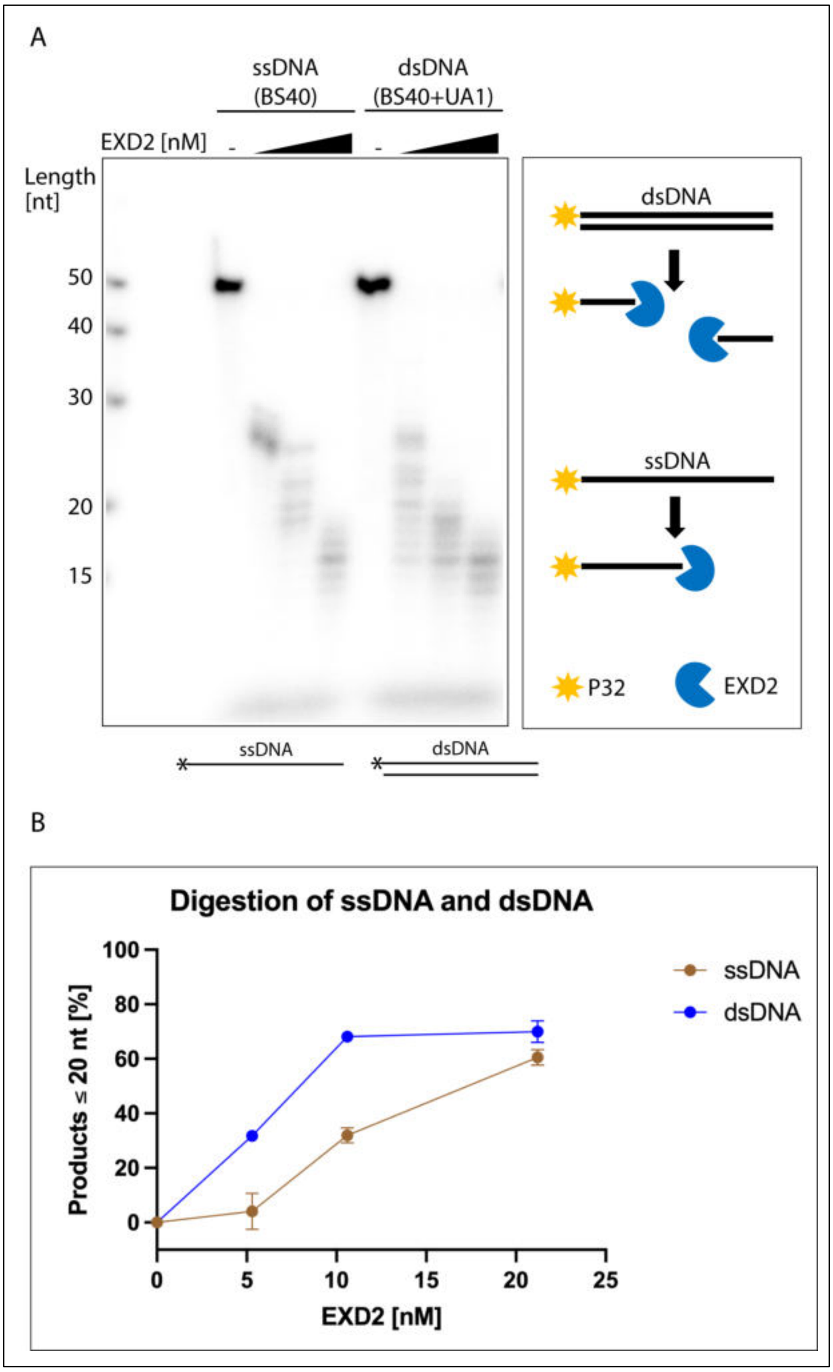


Figure 36. EXD2 WT digests dsDNA faster than the ssDNA of the same sequence. A) 5’-radiolabelled DNA substrates (~ 1 nM) were incubated with EXD2 WT (0; 5.3; 10.6; 21.2 nM,) for 1 h at 37℃. Samples were resolved in TBE-urea polyacrylamide gel and visualised by phosphor imaging. The reaction buffer contained 5 mM $MnCl_2$. Representative image of one of two repeats. B) Quantification of experiments such as in panel (A); n = 2 ; error bars; SEM. If no error bar is present, it means that it is shorter than the size of the symbol.

Conclusions

EXD2 WT, contrary to the nuclease-dead (D108A E110A) mutant, efficiently digests ssDNA and dsDNA, including blunt ended substrates and substrates with a 3’ or 5’ overhang. Digestion patterns of substrates of different length but the same 3’ end sequence, suggests that EXD2 has higher preferences towards longer DNA substrates, with 20 nt substrates being long enough to initiate digestion. Analysis of the length of products generated by EXD2 nuclease activity on ssDNA and dsDNA of the same sequence, indicates higher digestion preferences towards ds than ssDNA. The shortest product of digestion was ~12 nt long, which may suggest a minimal length of substrate required for processing by EXD2, or may indicate the DNA binding footprint of EXD2.

5.2 EXD2 nuclease activity on damaged DNA

To identify if EXD2 can take part in the repair of modified DNA strands that may be in the proximity of DSBs, I performed *in vitro* nuclease activity assays on the following substrates: (i) phosphorylated DNA, (ii) DNA with an abasic lesion and (iii) a substrate with mismatched nucleotides. I also tested EXD2's ability to process nicked and gapped sequences, which may be formed by MRN at the initial stage of end resection.

5.2.1 Phosphorylated DNA - overview

DNA phosphodiester bonds between 3'-O of one nucleoside and 5'-O of another, can be hydrolysed to either a 3'- or 5' phosphate (Mikkola, Lönnberg, & Lönnberg, 2018). DNA ends with a 3'- phosphate can be generated when tyrosyl-DNA-phosphodiesterase I (Tdp1) removes topoisomerase I (TopI) trapped by inhibitors (Zhelkovsky & McReynolds, 2014). Cancer treatment agents, like bleomycin and ionizing radiation, may also lead to DNA damage with a 3'- phosphate end. Importantly, some DNA repair reactions, including ligation, cannot act on a 3'- phosphorylated DNA, as they need a free terminal 3'-hydroxyl group (Zhelkovsky & McReynolds, 2014).

5.2.2 EXD2 nuclease activity on phosphorylated DNA

To test EXD2's ability to digest 3'- phosphorylated substrates, I used ssDNA (50 nt) phosphorylated on the 3' terminus, with the same nucleotide sequence as the unmodified substrate (Figure 37). Both substrates were efficiently digested indicating that EXD2 can process 3'- phosphorylated DNA ends.

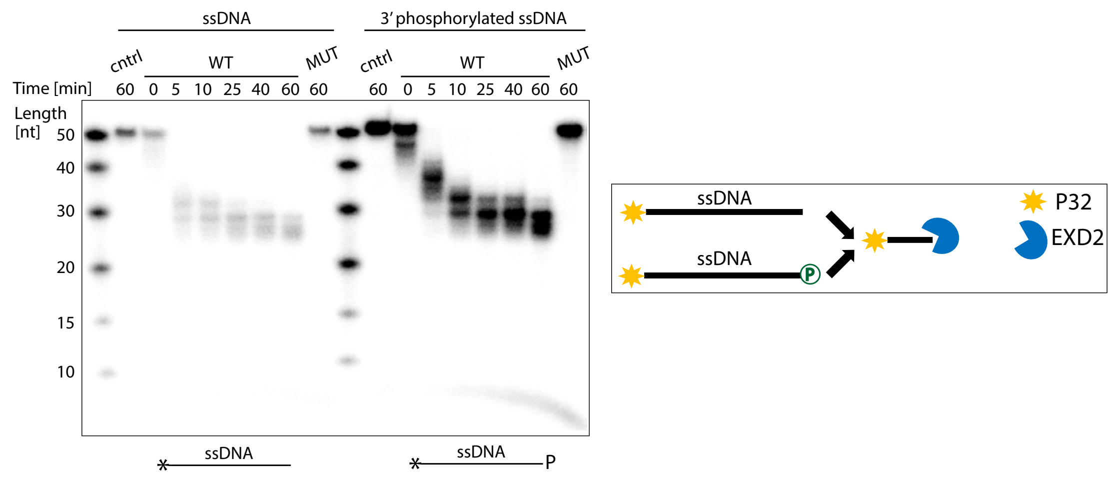


Figure 37. EXD2 WT, contrary to (D108A E110A) mutant, exhibits 3’-5’ exonuclease activity on 3’ phosphorylated substrate. 5’-radiolabelled DNA substrates (~ 1 nM) were incubated with EXD2 WT (1.6 nM) or EXD2 (D108A E110) MUT (1.6 nM) for indicated amounts of time at 37°C. Samples were resolved in TBE-urea polyacrylamide gel and visualised by phosphor imaging. The reaction buffer contained 5 mM MnCl$_2$. Representative image of one of three repeats.

5.2.3 DNA lesions - overview

Abasic DNA lesions (AP) are created when the N-glycosylic bond between nitrogenous base and deoxyribose breaks, either spontaneously or by the action of specialised DNA glycosylases (P. S. Thompson & Cortez, 2020). APs can be generated by alkylating agents used in cancer treatment and can be the source of genomic instability (P. S. Thompson & Cortez, 2020). As a consequence of the AP formation, other lesions, including interstrand crosslinks, may be formed (P. S. Thompson & Cortez, 2020).

Another DNA defect is a DNA mismatch (MM), which is formed when two non-complementary bases are aligned together (Rossetti et al., 2015). It can arise as a result of ionizing radiation, mutagenic chemicals or replication (Rossetti et al., 2015).

5.2.4 EXD2 nuclease activity on abasic DNA lesions and on mismatched DNA

To test if EXD2 can process abasic DNA, I used a dsDNA substrate with an abasic site located before the last 4 nucleotides of the upper strand (depicted in Figure 38). EXD2 WT processed the 3’ end of this substrate until it reached the abasic site, which completely blocked further digestion. DNA of the same sequence, but without an abasic lesion, was efficiently digested (Figure 38).

EXD2 digested dsDNA containing a mismatched nucleotide (G replacing T) and a fully complementary substrate with high efficiency (Figure 38). The digestion kinetics were assessed by quantification of products shorter than 25 nt, indicating digestion past the mismatch site.

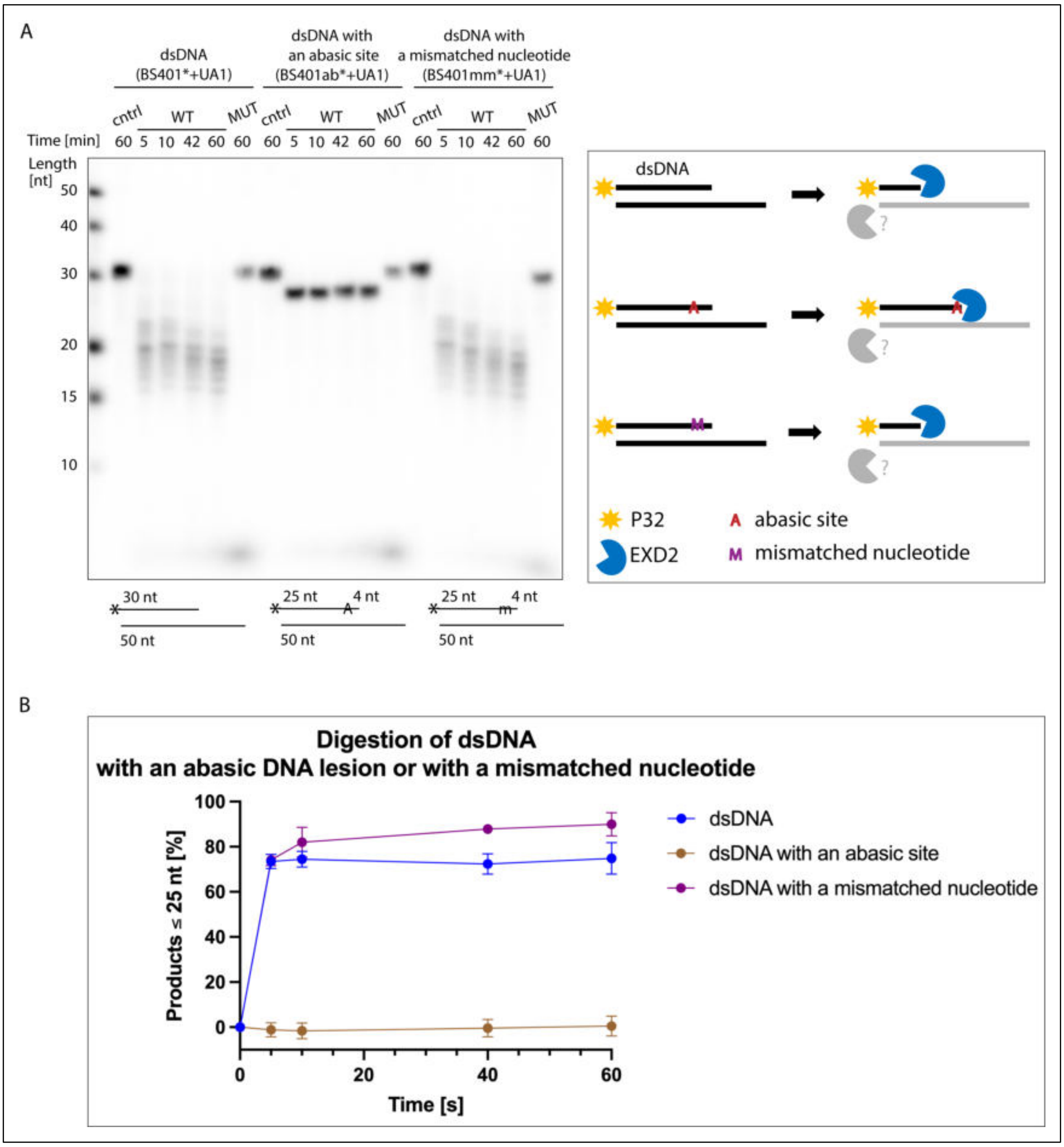


Figure 38. EXD2 WT, contrary to (D108A E110A) mutant, exhibits 3'-5' exonuclease activity on mismatched DNA but not on abasic DNA lesions. A) 5'-radiolabelled DNA substrates (~ 1 nM) were incubated with EXD2 WT (5.3 nM) or EXD2 (D108A E110) MUT (5.3 nM) for indicated amounts of time at 37℃. Samples were resolved in TBE-urea polyacrylamide gel and visualised by phosphor imaging. The reaction buffer contained 5 mM $MnCl_2$. Representative image of one of three repeats. B) Quantification of experiments such as in panel A; n = 2; error bars; SEM. If no error bar is present, it means that it is shorter than the size of the symbol.

5.2.5 Nicked and gapped DNA substrates - overview

Experiments performed by (Broderick et al., 2016) showed that truncated EXD2 (Lys76-Val564, expressed in bacteria) efficiently digests nicked dsDNA and dsDNA with a 1 nucleotide gap. This kind of substrate can be generated by MRN-dependent endonucleolytic cleavage in HR and may represent an entry point for EXD2, which then proceeds with resection (Broderick et al., 2016). As EXD2 purified by (Broderick et al., 2016) did not exhibit nuclease activity on blunt ended dsDNA, the nicked and gapped substrate they used did not require introduction of blocks at the ends of the DNA strands.

5.2.6 EXD2 nuclease activity on nicked and gapped DNA substrates

In order to test the full-length human EXD2's ability to process nicked or gapped DNA substrates, I started with repetition of the above experiments using the same sequences of nicked and gapped blunt ended DNA (Figure 39 A, Figure 39 B) as (Broderick et al., 2016). Both substrates were efficiently digested by WT EXD2. Modification of the substrates, to introduce a 5' overhang, did not affect the digestion pattern (Figure 39 A, Figure 39 B). To test, if EXD2 can start digestion from a nick or from a 1 nt gap, I introduced 5 phosphorothioate functional groups (PS) on the 3' end of the substrate sequence. This type of DNA modification is commonly used to block the activity of exonucleases (Deshpande et al., 2016). Both substrates were efficiently digested, meaning that EXD2 can start processing from a nick or from1 nt gap (Figure 39 C). Introduction of an additional block on 3' end of the unlabelled strand, changed the digestion pattern (Figure 39 D). Analysis of the impact of PS functional groups on EXD2's digestion pattern is described in more detail in chapter (5.6).

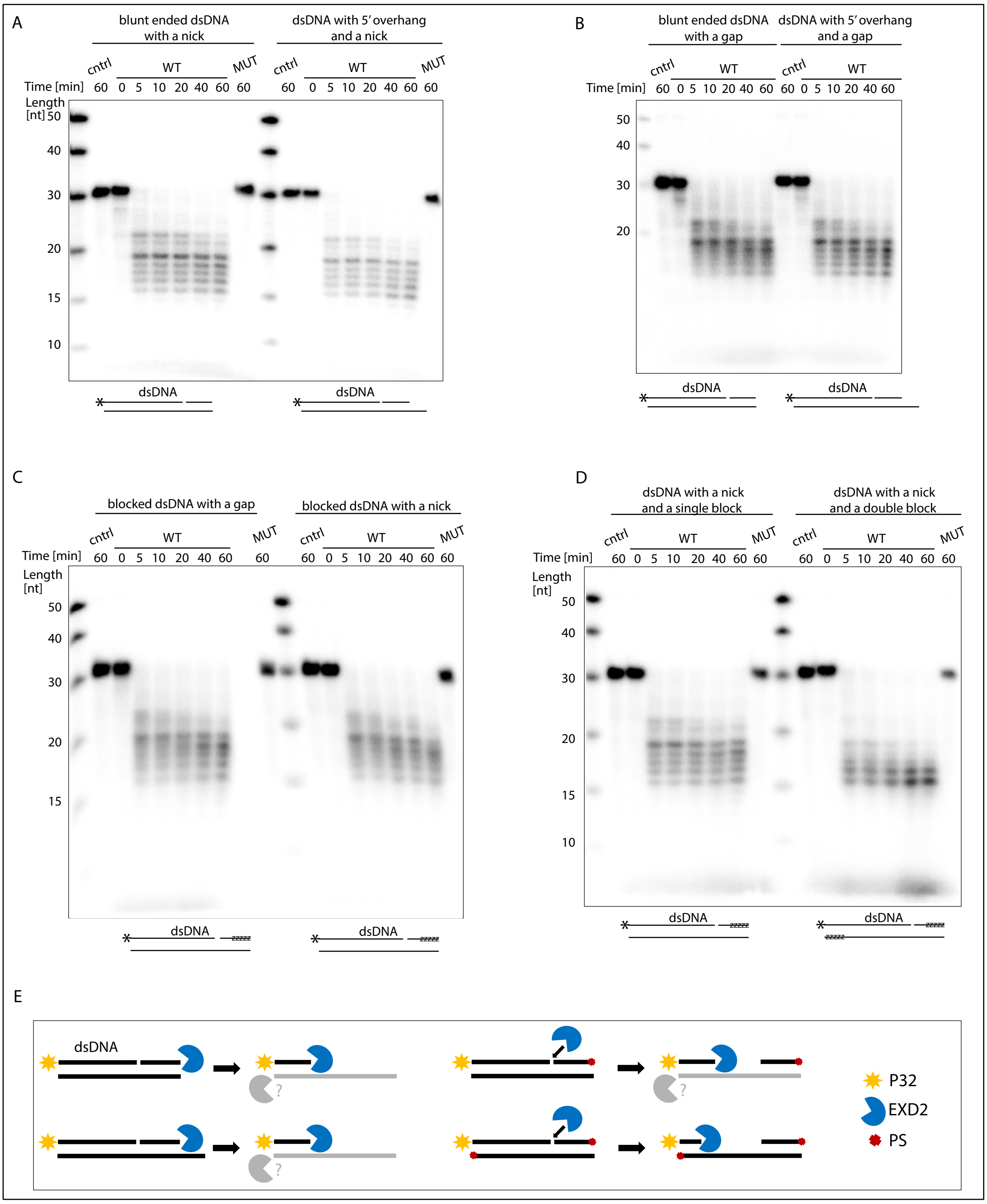


Figure 39. EXD2 WT, contrary to (D108A E110A) mutant, exhibits exonuclease activity on A) dsDNA with a nick, and B) dsDNA with a 1 nt gap. C) EXD2 WT can start processing DNA from a nick and from a 1 nt gap. D) Introduction of an additional block, on 3' end of the unlabelled strand changed the digestion pattern. 5'-radiolabelled DNA substrates (~ 1 nM) were incubated with EXD2 WT (5.3 nM) or EXD2 (D108A E110) MUT (5.3 nM) for indicated amounts of time at 37℃. Samples were resolved in TBE-urea polyacrylamide gel and visualised by phosphor imaging. The reaction buffer contained 5 mM MnCl$_2$ Representative images of several repeats. E) Schematic representation of WT EXD2 activity. As the digestion pattern of the unlabelled strand was not analysed, it is indicated with the grey colour. The PS at the end of the strand block exonucleolytic digestion.

To test if size of DNA gap can affect EXD2 activity, I used dsDNA with 1-, 3-, 5-, and 10-nucleotide gaps (Figure 40). The labelled strand had the same length for all the substrates, while the length of DNA on the 3' side from the gap differed depending on the gap size. Five PS modifications on the 3' end blocked nuclease-dependent processing of these DNA ends. All substrates were efficiently digested, meaning that EXD2 can start processing from gaps of different sizes.

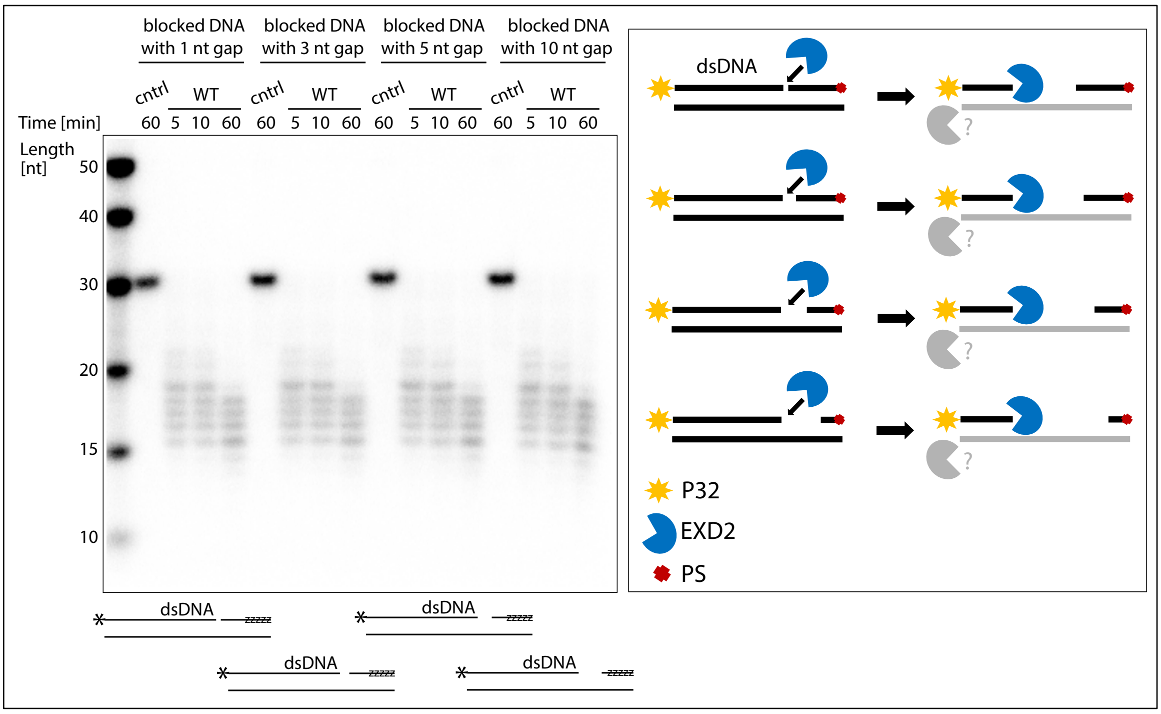


Figure 40. EXD2 WT can start processing DNA from 1-, 3-, 5- and 10- nucleotide gap. 5'-radiolabelled DNA substrates (~ 1 nM) were incubated with EXD2 WT (5.3 nM) for indicated amounts of time at 37℃. Samples were resolved in TBE-urea polyacrylamide gel and visualised by phosphor imaging. The reaction buffer contained 5 mM $MnCl_2$. Representative image of one of three repeats.

Conclusion

Full-length human EXD2 has the ability to digest different types of damaged DNA including, 3' phosphorylated ssDNA (Figure 37) and mismatched DNA (Figure 38). It digests dsDNA with a nick or with a gap inside the sequence (Figure 39 A and B) and can process DNA starting from a nick or from a gap (Figure 39 C). The size of a gap within 1 – 10 nt range, does not affect EXD2's activity (Figure 40). It suggests that EXD2 promotes efficient resection towards DSB and may play a role in removal of damaged DNA during this process. The only tested substrate in this chapter that could not be digested by EXD2 was dsDNA with an abasic site (Figure 38).

5.3 EXD2 activity on non-canonical DNA structures

DNA can adopt conformations that differ from canonical right-handed double helical DNA (B-DNA), (J. Zhao, Bacolla, Wang, & Vasquez, 2010), (G. Wang & Vasquez, 2014), (Bochman, Paeschke, & Zakian, 2012). Non-B DNA forming sequences are associated with genetic instability as they often colocalize with hotspots of DNA damage (G. Wang & Vasquez, 2014). To test EXD2's activity on these non-B DNA structures, I performed nuclease activity assays on hairpins, cruciforms and G-quadruplexes.

5.3.1 DNA hairpins - overview

Formation of non-canonical DNA structures can be promoted by DNA replication, transcription, and repair (J. Zhao et al., 2010). Generation of long regions of single stranded DNA can facilitate folding of secondary structures (J. Zhao et al., 2010). The structure of non-B DNA may affect accessibility of DNA damage and repair factors (G. Wang & Vasquez, 2014). Bases in single stranded DNA are more exposed than in canonical B-DNA, where they are located between sugar-phosphate backbones, hence in ssDNA they may be more prone to the damage (G. Wang & Vasquez, 2014).

Secondary DNA structures can be formed by palindromic sequences (two closely spaced or unspaced inverted repeats), (Miklenić & Svetec, 2021). In ssDNA the inverted repeats can pair with one another creating a hairpin. Each hairpin has a stem, formed by paired nucleotides, and a symmetric centre in the shape of a single stranded loop (J. Zhao et al., 2010). When two hairpins are created in the same position of both DNA strands, a cruciform structure is formed (G. Wang & Vasquez, 2014). Cruciform resolution leads to formation of double strand breaks with two hairpin-capped ends (Casper, Greenwell, Tang, & Petes, 2009). Research in yeast showed that processing of these ends is dependent on Mre11 and Sae2 which cleave the hairpin tips facilitating HR. (Casper et al., 2009). Hairpins can also be created during replication of the lagging DNA strand. Processing of this type of hairpin may be independent of Mre11 and Sae2 (Casper et al., 2009). According to (Mirkin, 2006), repetitive sequences may fold into hairpin-like structures in lagging strand templates, leading to contractions, or in nascent lagging strands, resulting in expansions (Figure 41).

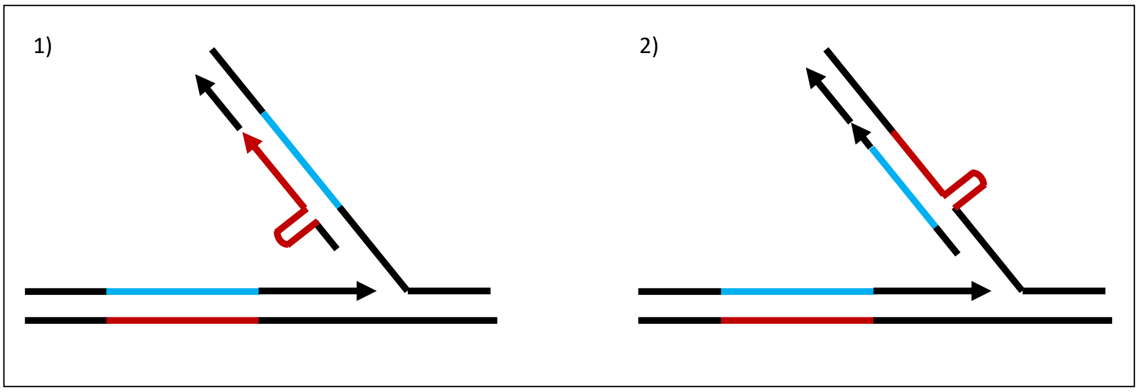


Figure 41. Generation of hairpin-like structures in nascent lagging strand 1), and in lagging strand template 2). Stands prone to structure formation are coloured in red, the complementary strands are blue. Adapted from (Mirkin, 2006).

5.3.2 EXD2 nuclease activity on DNA hairpins

DNA substrates labelled as 3A and H40 have already been used to examine the *in vitro* activity of Exonuclease 1 (Exo1), (Y. Li et al., 2019). Substrate 3A consist of ssDNA (29T) followed by (AAATTT) on the 3’ terminus (Figure 42). The sequence (AAATTT) at the 3’ end is very short, hence it is unlikely that it folds into a hairpin. The H40 hairpin has 6 bp long stem, with one G-A mismatch, located between 8 and 23 nt from the 5’ end. EXD2 WT, contrary to the exonuclease dead mutant, digested 3A substrate with high efficiency (Figure 42). However, a substrate containing a hairpin located in the middle of the sequence, with a longer stem and GC pairing at the base, was not efficiently digested by EXD2 (H40, Figure 42). The length of the digestion product suggests that EXD2 stops resection of the H40 substrate before reaching the base of the hairpin stem. The putative formation of the H40 hairpin was based on a prediction, not on experimental evidence.

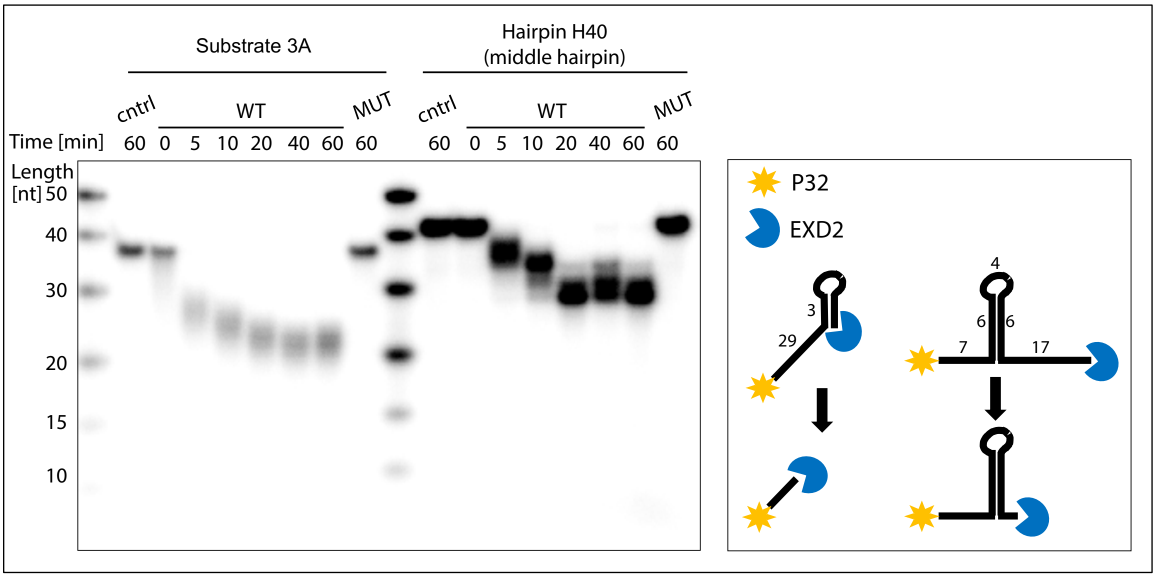


Figure 42. EXD2 WT, contrary to (D108A E110A) mutant, digests 3A substrate [(29 T) followed by (AAATTT)]. The H40 hairpin was not efficiently processed by EXD2. 5’-radiolabelled DNA substrates (3A; H40; ~ 1 nM) were incubated with EXD2 WT (1.6 nM) or EXD2 (D108A E110) MUT (1.6 nM) for indicated amounts of time at 37°C. Samples were resolved in TBE-urea polyacrylamide gel and visualised by phosphor imaging. The reaction buffer contained 5 mM $MnCl_2$. Formation of the hairpin was based on a prediction, not on experimental evidence. Representative image of one of three repeats.

To establish if EXD2 can exhibit hairpin opening activity, I used the YM164 substrate which was efficiently digested by Artemis nuclease (S. Li et al., 2014). The YM164 was predicted to form 20 bp long, fully paired stem and 6 nt overhang on the 5’ terminus. Here, the digestion pattern suggests that EXD2 process the stem exonucleolytically, reaching the hairpin tip, but probably cannot process any further (Figure 43). The observed activity is very different from the digestion pattern generated by Artemis, which opened YM164 hairpin endonucleolytically (S. Li et al., 2014). The putative formation of the YM164 hairpin was based on a prediction, not on experimental evidence.

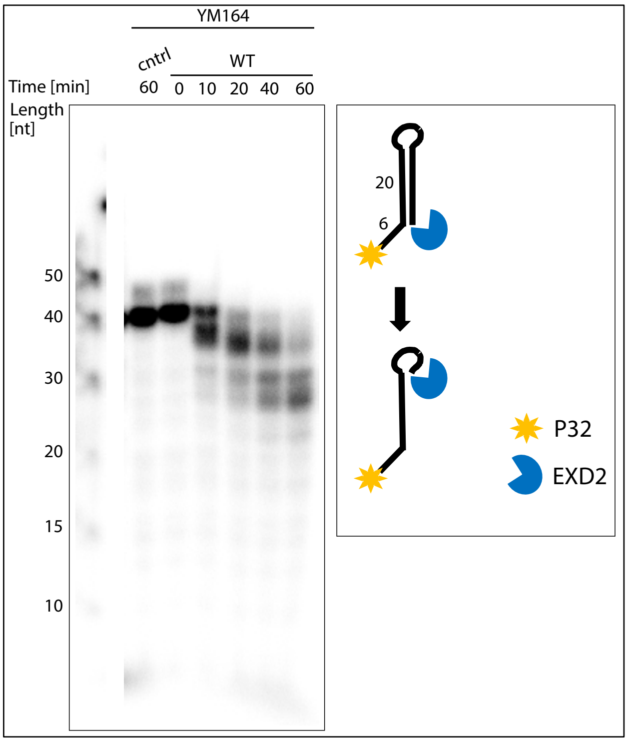


Figure 43. EXD2 WT resects YM164 hairpin stems but cannot digest past the hairpin tip. 5'-radiolabelled DNA substrate (YM164; ~ 1 nM) was incubated with EXD2 WT (5.3 nM) for indicated amounts of time at 37℃. Samples were resolved in TBE-urea polyacrylamide gel and visualised by phosphor imaging. The reaction buffer contained 5 mM $MnCl_2$. Formation of the hairpin was based on a prediction, not on experimental evidence. Representative image of one of three repeats.

To further assess what kind of hairpins can be digested by EXD2, I designed a DNA hairpin by modifying the PC217 substrate from (Reginato, Cannavo, & Cejka, 2017). Here, the 70 nt long DNA which encodes "AGGG**CA**CGCCCT" sequence was modified into "AGGGCGCCCT" to create a hairpin with a complementary 5 bp stem ('hairpin_PC217', 68 nt). According to MacVector predictions, it is difficult to assess what conformation it adapts, as both PC217 and PC217_hairpin may form multiple secondary structures (Figure 44 A). The modification I introduced may lead to the formation of a hairpin with a 5 bp stem or a much longer, complex structure with partially uncomplimentary nucleotides. As EXD2 process dsDNA more efficiently than ssDNA, I also created a hairpin of the same sequence but extruded from dsDNA. Therefore, I annealed the PC217_hairpin sequence (68 nt) with 60 nt long DNA (Figure 44 B); referred to as 'double stranded hairpin'. There is no experimental evidence as to what the conformation of generated substrates are.

Figure 44. A) Secondary structures that may be formed by PC217 and hairpin_PC217 sequences, according to the MacVector prediction software. The blue colour indicates which parts of sequences may form DNA hairpins. The difference between nucleotide sequences of PC217 and hairpin_PC217 is indicated by the underline. B) Structure of a hairpin extruded from the double stranded DNA (dsPC217_hairpin).

In the hairpin_PC217 substrate, an AT pair at the stem base should be followed by 4 fully paired GC base pairs. The stem structure of a hairpin extruded from dsDNA may be different as AT at the stem base may complement with TA from the second strand, leaving only 4 GC base pairs out (Figure 44 B). Analysis of EXD2 activity showed that the hairpin extruded from dsDNA was efficiently digested (Figure 45). Quantification of products shorter than 40 nt indicates that EXD2 resected this structure past the hairpin tip. The digestion pattern of the ssDNA substrate, which may fold into the hairpin, varied (Figure 45 A, B). Therefore, digestion quantification was performed for each repeat separately (Figure 45 C, D). On a ssDNA substrate, EXD2 WT either processed part of the hairpin stem or was capable of digestion of the whole hairpin (Figure 45). It is possible that the variability between the repeats is caused by differences in substrate folding.

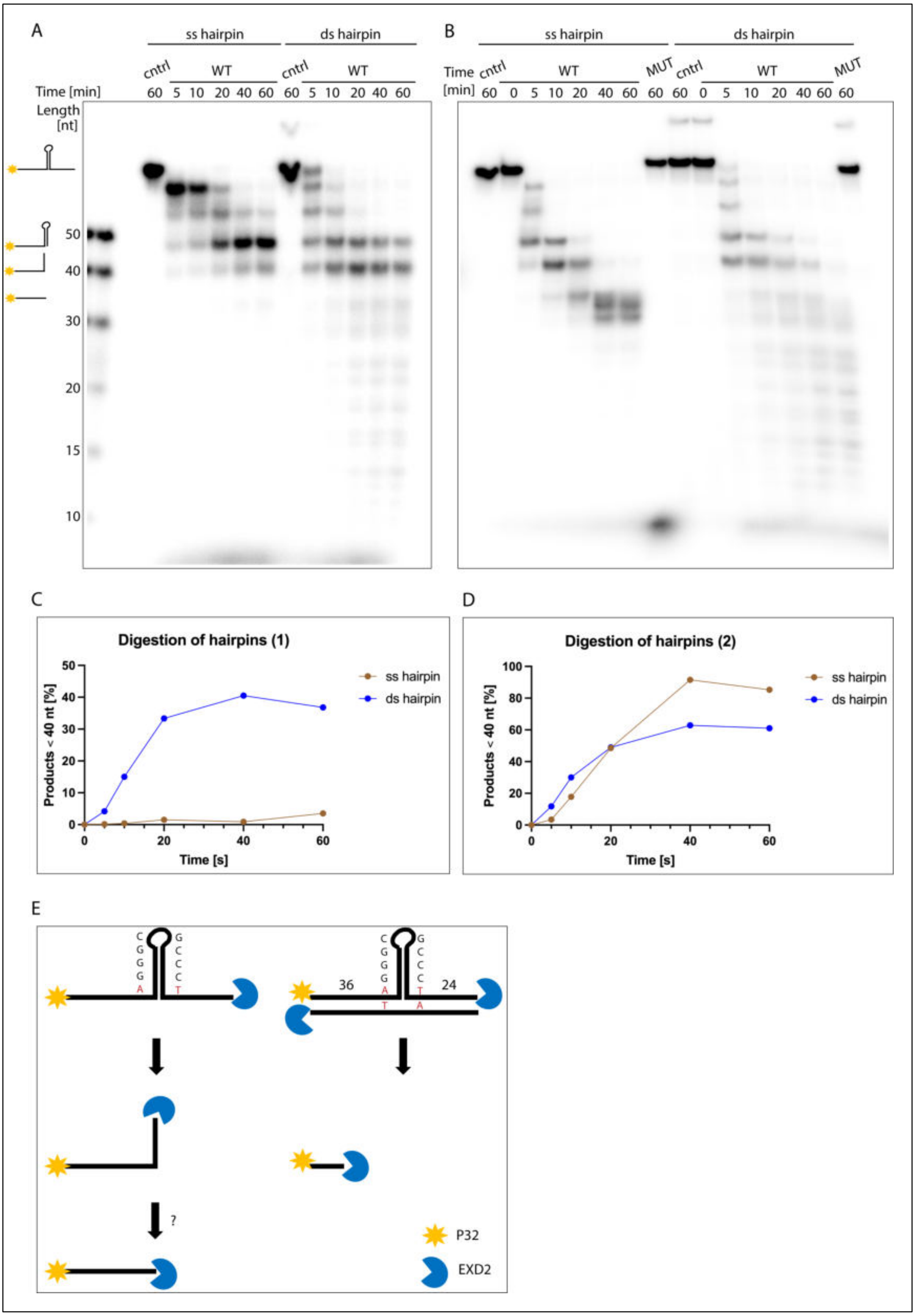


Figure 45. EXD2 WT, contrary to (D108A E110A) mutant, digests hairpin extruded from dsDNA. Activity of EXD2 WT on a ssDNA hairpin (hairpin_PC217) varied between the repeats (it is difficult to conclude if EXD2 can digest past the hairpin tip). A) B) 5'-radiolabelled DNA substrates (~ 1 nM) were incubated with EXD2 WT (5.3 nM) or EXD2 (D108A E110) MUT (5.3 nM) for indicated amounts of time at ~ 37℃. Samples were resolved in TBE-urea polyacrylamide gel and visualised by phosphor imaging. The reaction buffer contained 5 mM MnCl$_2$. Formation of the hairpins is based on a

prediction, not on experimental evidence. Representative images of two from three repeats. C) Quantification of experiments in panel A; D) Quantification of experiments in panel B; E) Schematic representation of substrate digestion.

Digestion pattern of both hairpins showed two distinct products around 40 and 45 nt long. To check if this digestion pattern may be the result of two endonucleolytic cleavages at the stem base, I introduced PS modifications at the tip of PC217_ hairpin substrate (AGGG*C*GCCCT). The PS should block the exonucleolytic activity of EXD2. If the hairpin is cut out endonucleolytically at both sides of the stem base, the PS modifications at the hairpin tip should not inhibit digestion. The digestion pattern showed that EXD2 cannot process through the PS block and stops at the hairpin tip (Figure 46). It indicates that EXD2 processes the ds hairpins exonucleolytically. It is possible that two distinct bands (~40 and 45 nt) observed in ss and ds hairpin digestion pattern (Figure 45) may correspond not to the cleavage at the both sides of the stem base, but to the cleavage at the hairpin tip and the cleavage on the 3’ end of the stem base. However, most likely they are the result of exonucleolytic processing.

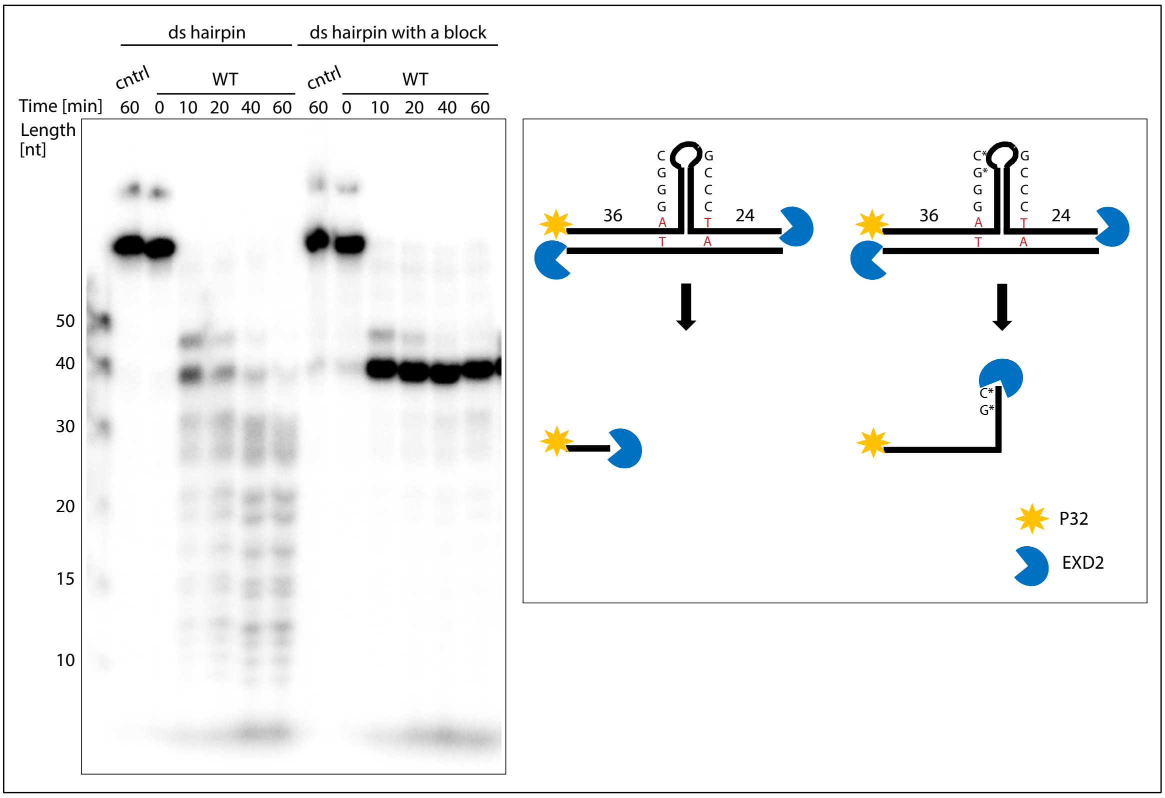


Figure 46. EXD2 WT digests dsPC217_hairpin with high efficiency. Introduction of the PS modifications at the hairpin tip inhibits hairpin digestion. 5’-radiolabelled DNA substrates (~ 1 nM) were incubated with EXD2 WT (5.3 nM) for indicated amounts of time at 37℃. Samples were resolved in TBE-urea polyacrylamide gel and visualised by phosphor imaging. The reaction buffer contained 5 mM $MnCl_2$. Formation of the hairpins is based on a prediction, not on experimental evidence. Representative image of one of two repeats.

The quality of hairpins used in the reactions was checked on native gels (Figure 47). The migration pattern for single stranded hairpin_PC217 varied as two bands of different intensity were observed. They may be a result of self-dimer formation, which is highly possible based on the PC217 sequence and the fact that two bands were observed also for ssDNA PC217 sequence.

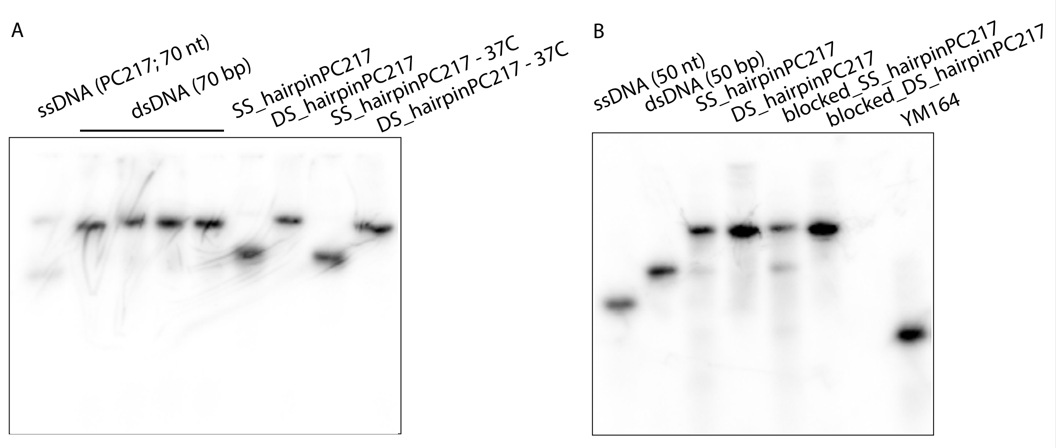


Figure 47. Native gel analysis of DNA substrates used for the nuclease activity assays.

5.3.3 DNA cruciform structures - overview

Cruciform structures can be formed as a result of negative supercoiling of DNA containing inverted repeats of 6 or more nucleotides, which are present in many promotor regions (Brázda, Laister, Jagelská, & Arrowsmith, 2011). Cruciform structures may be formed during DNA replication and transcription, and many proteins from these pathways have cruciform binding properties (Brázda et al., 2011). Cruciform structures may also have a role in epigenetics, as they protect DNA from methylation (Brázda et al., 2011).

The main parts of the cruciform structure are the stem, loop, and a branch point (Brázda et al., 2011). If the inverted repeats are located directly next to each other, the size of a loop is minimal. Cruciform structures are not stable on naked linear DNA, because of branch migration (Brázda et al., 2011). They can be stabilized in a circular plasmid DNA *in vitro* (Brázda et al., 2011). Cleavage of cruciform structures can be performed by topoisomerase II (G. E. Lee, Kim, & Chung, 1998). As long cruciforms can also be recognized by junction resolving enzymes, they can lead to DNA breaks and genome instability (Brázda et al., 2011). There are two hypotheses of cruciform structure cleavage: i) centre break and ii) resolution (Inagaki et al., 2013). In the centre break mechanism each hairpin is nicked near the tip (Cunningham, Coté, Cam-Ozdemir, & Lewis, 2003). The resolution mechanism can be performed by GEN1, which performs diagonal cleavage of the cruciform (Inagaki et al., 2013). It leads to the formation of hairpin ends which can be opened by Artemis. This sequential cleavage of cruciform structures leads to chromosomal translocations (Inagaki et al., 2013).

5.3.4 EXD2's activity on cruciform-like structures

As EXD2 efficiently digested hairpins extruded from dsDNA (Figure 44 B), I used the same sequence (hairpin_PC217) to generate a structure mimicking a mobile cruciform. Hence DNA strands of the linear substrate were fully complementary to each other (Figure 48 A), the generated substrate was not stable and it could configure as a bubble, a fully annealed dsDNA or a mobile cruciform. There is no experimental evidence indicating what structure this substrate formed. The length of the stem was 5 bp, which is less than 6 bp generally required for the cruciform formation. However, according to (Miura, Ogake, Yoneyama, Kikuchi, & Ohyama, 2019), there are cruciform structures with the short stems of 5 bp.

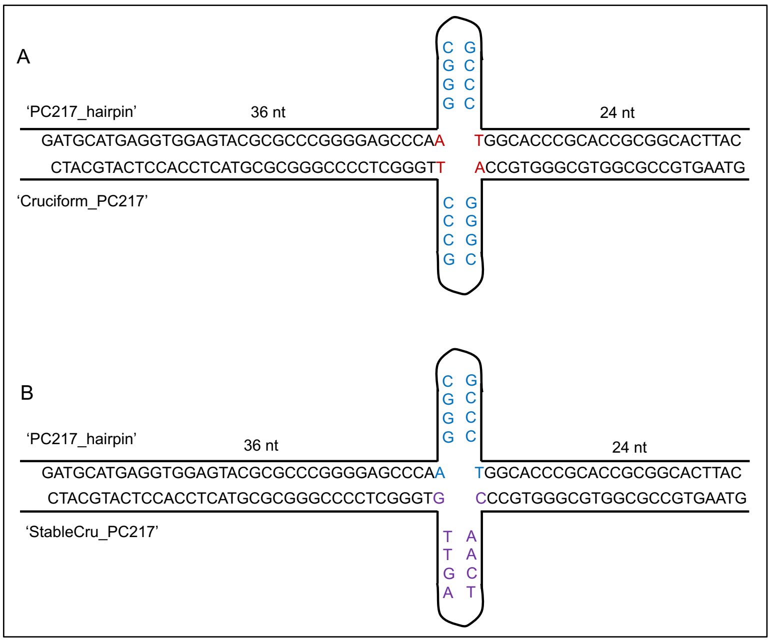


Figure 48. Structure of A) a fully complementary mobile cruciform and B) a more stable non-complementary cruciform.

EXD2 digested both strands of this mobile cruciform-like structure and one of the strands was resected more efficiently than the other (Figure 49). Nuclease activity assays were carried out at around 37°C, whereas the Tm of the single strand was 78.7°C (according to IDT Tm calculator).

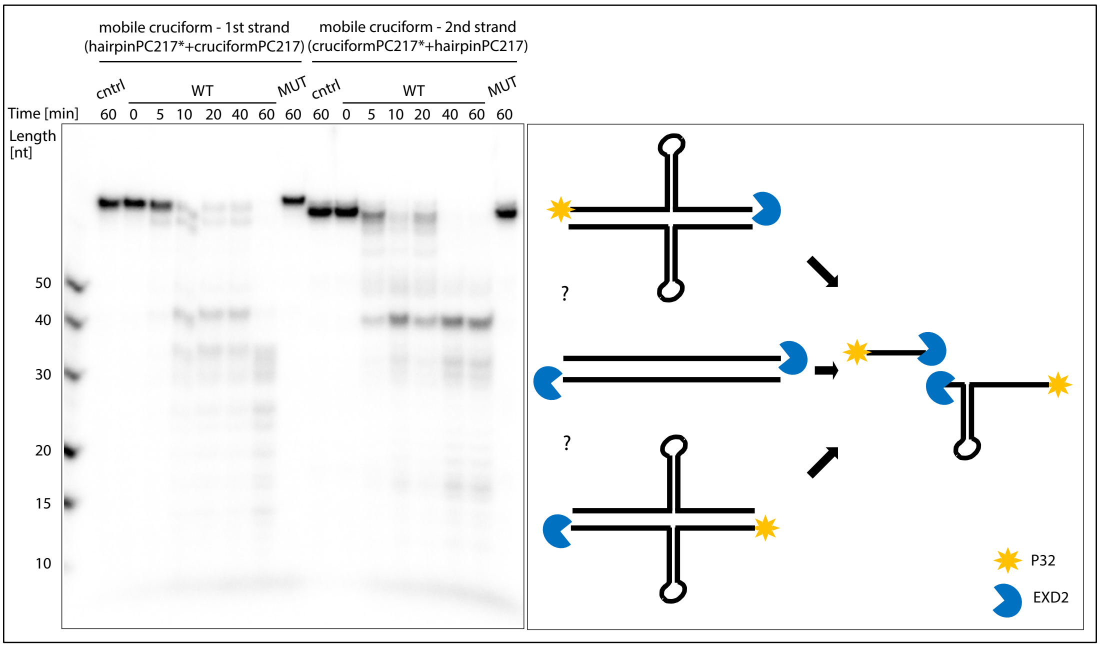


Figure 49. EXD2 WT, contrary to (D108A E110A) mutant, digests structure mimicking a mobile cruciform, which due to its instability can fully anneal into dsDNA. 5'-radiolabelled DNA substrates (~ 1 nM) were incubated with EXD2 WT (5.3 nM) or EXD2 (D108A E110) MUT (5.3 nM) for indicated amounts of time at ~ 37 °C. Samples were resolved in TBE-urea polyacrylamide gel and visualised by phosphor imaging. The reaction buffer contained 5 mM $MnCl_2$. There is no experimental evidence regarding the conformation the substrate adopted. Representative image of one of three repeats.

To increase the stability of these cruciform substrate, I modified the sequence so that the DNA strands were no longer complementary between each other (Figure 48 B). Digestion reactions were then conducted with increasing concentration of EXD2 WT, which fully digested cruciform-like structure and ssDNA of the same sequence, which should fold into a hairpin ('StableCru_PC217'), (Figure 50). Digestion efficiency was quantified by examining the products shorter than 25 nt, indicating resection past the hairpin tip. Nuclease-dead EXD2 mutant was not capable of performing any digestion, even at the highest protein concentrations.

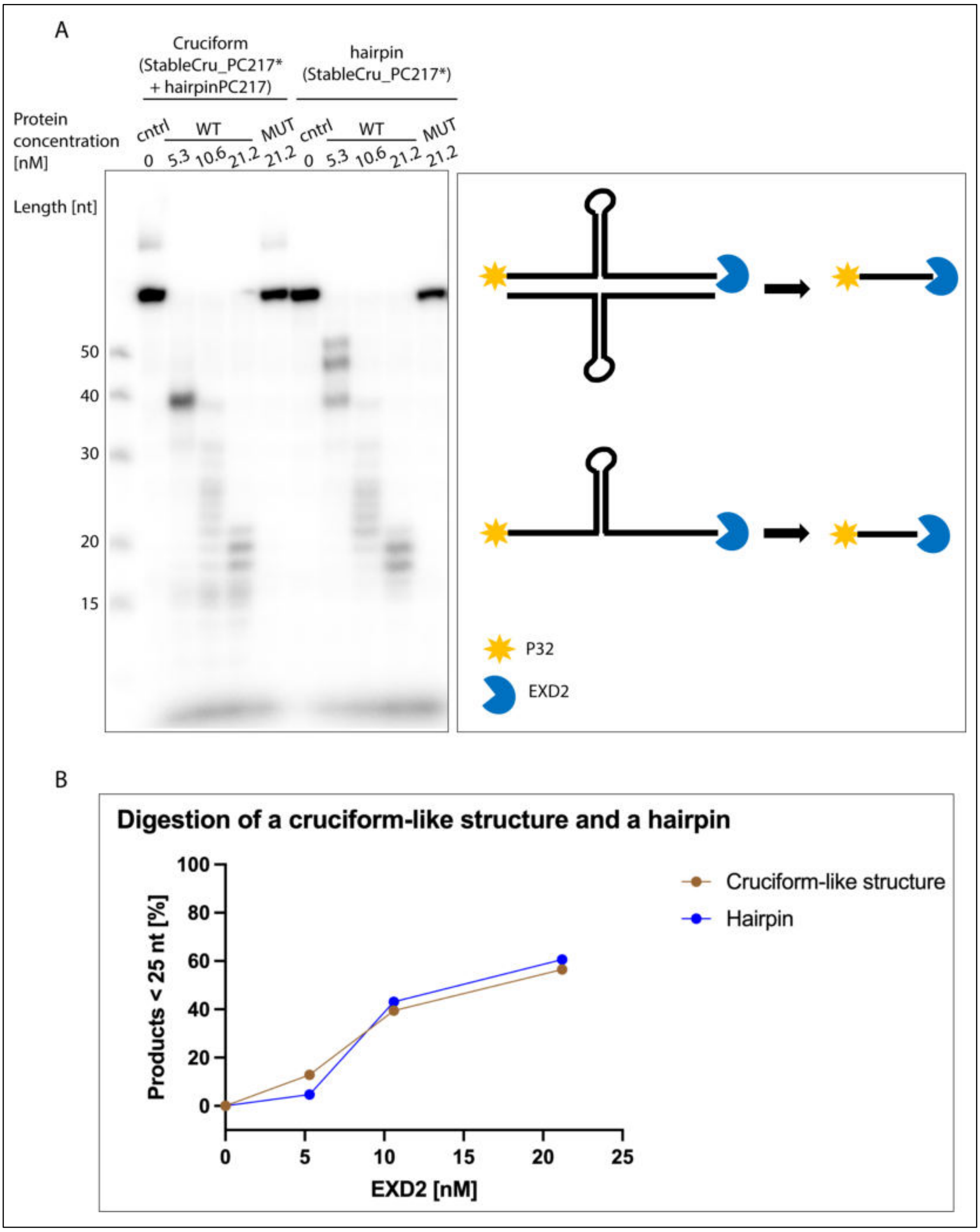


Figure 50. EXD2 WT, contrary to (D108A E110A) mutant, digests structures mimicking cruciforms with non-complementary regions between the two strands, and digests ssDNA hairpins. Formation of these structures is based on a predication, not on experimental evidence. A) 5'-radiolabelled DNA substrates (~ 1 nM) were incubated with indicated concentrations of EXD2 WT or EXD2 (D108A E110) MUT for 1 h at 37℃. Samples were resolved in TBE-urea polyacrylamide gel and visualised by phosphor imaging. The reaction buffer contained 5 mM $MnCl_2$. B) Quantification of experiments such as in panel A; n = 3; error bars; SEM. If no error bar is present, it means that it is shorter than the size of the symbol. Representative image of one of three repeats.

Conclusions

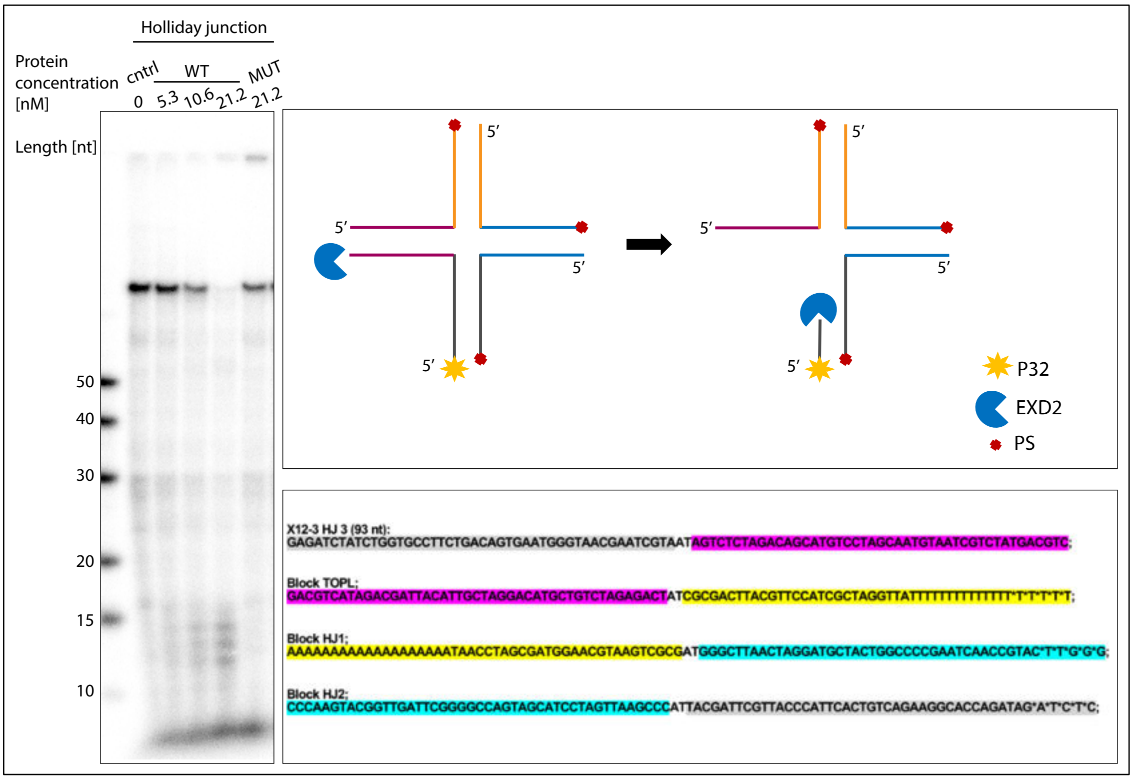
At high protein concentration EXD2 WT efficiently digested dsDNA with a non-complementary sequence, which could either fold into the short cruciform or mimic a DNA bubble. At high protein concentrations, EXD2 WT also efficiently resected ssDNA hairpins. Formation of these secondary structures is based on a prediction and was not confirmed experimentally. As cruciform structures are thermodynamically unstable in their linearized form, substrates based on a circular plasmid DNA should also be tested in the future. Folding of hairpin structures can be experimentally confirmed by fluorescence resonance energy transfer (FRET) analysis of fluorescently labelled substrates.

5.3.5 Holliday junctions – overview

A holliday junction (HJ) is a four-way stranded DNA intermediate linking two dsDNA molecules (Song, Hu, Yin, Wang, & Yin, 2022). It is formed during HR and regression of stalled replication forks. Similarly to cruciform structures, it has a stem and a branch point (Brázda et al., 2011). HJs have two main configurations: X-stacked and open-planar, which can be regulated depending on the salt concentration (open-planar is dominant in low salt conditions), (Song et al., 2022). Unprocessed HJs lead to replication damage, and the formation of ultrafine anaphase bridges, mis-segregation of chromosomes and even cell death (Chan, Fugger, & West, 2018), (Song et al., 2022). Therefore, HJs-targeting ligands could be used in an anticancer therapy. The mechanisms of HJ resolution are discussed in (1.4.6).

5.3.6 EXD2's activity on Holliday junctions

To test EXD2's activity on HJs, I used substrates published by (Thangavel et al., 2015). To mimic continuity of the HJs and block multiple entry points for exonuclease activity, I introduced PS modifications on three 3' ends of the substrate (Figure 51). EXD2 WT efficiently digested the labelled strand past the branch point, until the strand was shorter than 15 nt. This digestion pattern is in line with EXD2 activity on other DNA structures with introduced PS modifications (effect of PS blocks on EXD2 activity is discussed in more detail in 5.6).



X12-3 HJ 3 (93 nt):
GAGATCTATCTGGTGCCTTCTGACAGTGAATGGGTAACGAATCGTAATAGTCTCTAGACAGCATGTCCTAGCAATGTAATCGTCTATGACGTC;

Block TOPL;
GACGTCATAGACGATTACATTGCTAGGACATGCTGTCTAGAGACTATCGCGACTTACGTTCCATCGCTAGGTTATTTTTTTTTTTTTT*T*T*T*T;

Block HJ1;
AAAAAAAAAAAAAAAAAAAATAACCTAGCGATGGAACGTAAGTCGCGATGGGCTTAACTAGGATGCTACTGGCCCCGAATCAACCGTAC*T*T*G*G*G;

Block HJ2;
CCCAAGTACGGTTGATTCGGGGCCAGTAGCATCCTAGTTAAGCCCATTACGATTCGTTACCCATTCACTGTCAGAAGGCACCAGATAG*A*T*C*T*C;

Figure 51. EXD2 WT, contrary to (D108A E110A) mutant, exonucleolytically digests substrates mimicking Holliday junctions. To block EXD2 activity, all 3' ends apart from the labelled strand, were blocked with PS modifications. Sequences used for generation of the substrate are highlighted with different colours, depending on the strand sequence they anneal to. 5'-radiolabelled DNA substrate (~ 1 nM) was incubated with EXD2 WT (0; 5.3; 10.6; 21.2 nM) or EXD2 (D108A E110) MUT (21.2 nM) for 1 h at 37℃. Samples were resolved in TBE-urea polyacrylamide gel and visualised by phosphor imaging. The reaction buffer contained 5 mM MnCl$_2$. Representative image of one of three repeats.

Conclusions

EXD2 WT efficiently digests Holliday junctions past its branch point. It has been shown that unprocessed HJs result in replication damage (Song et al., 2022), hence it is possible that EXD2 is acting with other nucleases to mediate HJ resolution and enable efficient replication. As HJ substrates mimic reversed replication forks, the controlled resection may be in line with the role of EXD2 in replication fork protection (Nieminuszczy et al., 2019).

5.3.7 G-quadruplexes - overview

Guanine rich sequences of DNA can fold into four-stranded structures, known as G-quadruplexes (G4s), (Hänsel-Hertsch, Di Antonio, & Balasubramanian, 2017). They are formed from G tetrads, which are planar structures coordinated by a central cation, stacked on top of one another. Depending on number of DNA strands engaged, G4s can be classified as intra- or inter- molecular (Liano, Monti, Chowdhury, Raguseo, & Di Antonio, 2022). Intra-molecular G4s are formed within one DNA strand, from 4 or more repeats of three Gs. Their biological importance is characterized in more detail than the role of inter-molecular G4s, which are formed between multiple DNA strands.

G4 forming sequences are especially abundant in promoters and telomeres (Valton & Prioleau, 2016). Interestingly, they are more often present in regulatory genes and oncogenes, than in tumour suppressor genes (Rhodes & Lipps, 2015). G4s may both positively and negatively regulate transcription. On the one hand they can act as blocks for polymerase processivity, but on the other they may mediate transcription by stabilizing R-loops (Robinson, Raguseo, Nuccio, Liano, & Di Antonio, 2021). It is possible that G4s promote long-range DNA interactions, for example between proteins bound to promoter and enhancer regions (Robinson et al., 2021). Telomeric G4s are substrates for human telomerase, which colocalizes with them *in vivo,* and has ability to partially unwind them before telomeric DNA extension (Moye et al., 2015). In dsDNA, the probability of forming G4s increases during replication, transcription, and DNA repair, when DNA is transiently single-stranded (Rhodes & Lipps, 2015). The presence of G4s may affect DSB formation, as genomic regions with predicted G4 motifs are enriched with DSBs (Hänsel-Hertsch et al., 2017). G4s can block DNA polymerases, leading to stalled replication forks and genome instability (Valton & Prioleau, 2016).

5.3.8 EXD2 activity on G-quadruplexes

To test if EXD2 can process G4s, I used a substrate encoding a modified G4 forming sequence located in the MYC promotor region. MYC oncogene is overexpressed in up to 70 % of all human cancers, hence its targeting is a promising therapeutic strategy (W. Wang et al., 2020). This substrate contained the MycPu22 sequence (TGA**GGG**T**GGG**TA**GGG**T**GGG**TAA), with introduced overhangs (total length of the oligonucleotide - 70 nt, ‘extendMyc’), which was a kind gift from Sophie Williams (Genome Replication Lab at the ICR). According to (W. Wang et al., 2020), Pu22 G4 structure is thermodynamically stable (melting temperature > 85°C) and it adopts a conformation as shown in (Figure 52) from (W. Wang et al., 2020).

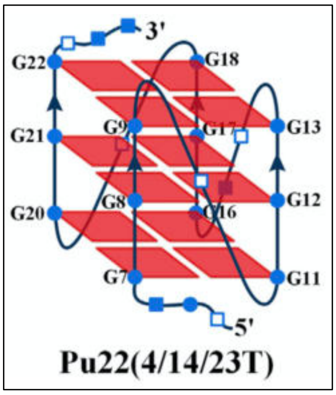


Figure 52. Pu22 structure from (W. Wang et al., 2020). Pu22 is modified Pu27 sequence, with G to T mutations at G4, G14, G23.

Before the reaction, this substrate was boiled and gradually cooled in the presence of 100 mM KCl (in 10 mM Tris-HCl, pH 7.5), as G4 stabilization is higher in the presence of K^+ than Na^+ (Hänsel-Hertsch et al., 2017). However, as the nuclease activity assays were performed in EXD2 reaction buffer, the final concentration of KCl was decreased 7.5 times. This massive reduction of KCl concentration could significantly decrease the stability of G4s during reactions with EXD2. As high temperature could also affect the stability of G4s, nuclease activity assays were performed at both RT and 37°C (Figure 53). EXD2 WT initiated digestion of the overhang but could not process the sequence that may form G4s. As expected, EXD2 (D108A E110A) mutant was not able to perform any digestion.

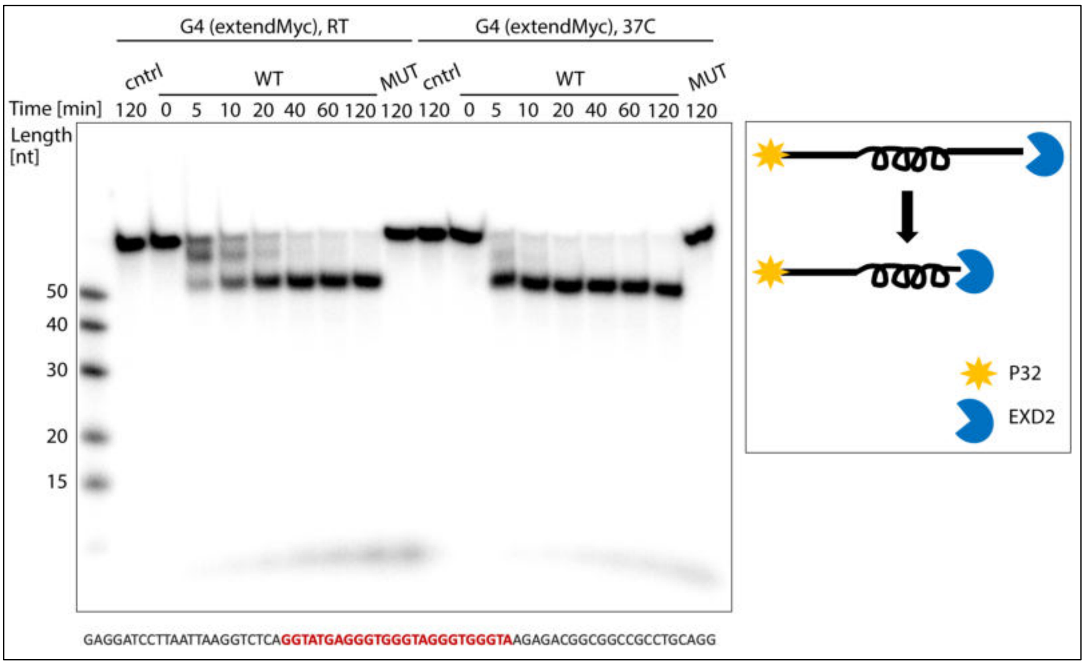


Figure 53. Both EXD2 WT and (D108A E110A) mutant, cannot digest sequence that may form G-quadruplexes. Nucleotide composition of ‘extendMyc’ sequence is shown beneath the gel, with red letters indicating G-rich sequence. 5’-radiolabelled DNA substrate (extendMyc; ~ 1 nM) was incubated with EXD2 WT (5.3 nM) or EXD2 (D108A E110) MUT (5.3 nM) for indicated amounts of time at RT or 37°C. Reactions were performed in RT to increase stability of the substrate. The reaction buffer contained 5 mM $MnCl_2$ and ~ 13 mM KCl. Low KCl concentration could decrease stability of G4s during the reaction. Samples were resolved in TBE-urea polyacrylamide gel and visualised by phosphor imaging. Representative image of one of three repeats.

Next, I tested if EXD2 can process telomeric G4 forming sequences. Telomeric sequences in humans are formed from $(TTAGGG)_n$ repeats (Moye et al., 2015). They vary in length, ranging from 5 to 10 kb in human blood leukocytes, to 14 kb in human sperm (Pickett, Henson, Au, Neumann, & Reddel, 2011). According to (Petraccone et al., 2011), $(TTAGGG)_n$, where n = 4, 8, 12, spontaneously folds into stable G4s structures. The melting temperature of $(TTAGGG)_4$ is around 62°C (Petraccone et al., 2011).

I used a (TTAGGG)$_4$ substrate with the same overhangs as those introduced to the MycPu22 sequence (with 1 additional nt on the 5’ end, for the total length of 70 nt; ‘extendTelom’). This and all the other G4 forming sequences were prepared the same way as the first substrate (gradual cooling after boiling, the same buffer composition). As the nuclease activity assays were performed in the EXD2 reaction buffer, the final KCl concentration in the sample was decreased from 100 mM to ~ 13 mM, which could significantly reduce the stability of G4s. The aim of conducting reactions at RT was to improve the stability of G4s. In both temperature conditions, EXD2 WT performed resection of a sequence which may form telomeric G4s (Figure 54).

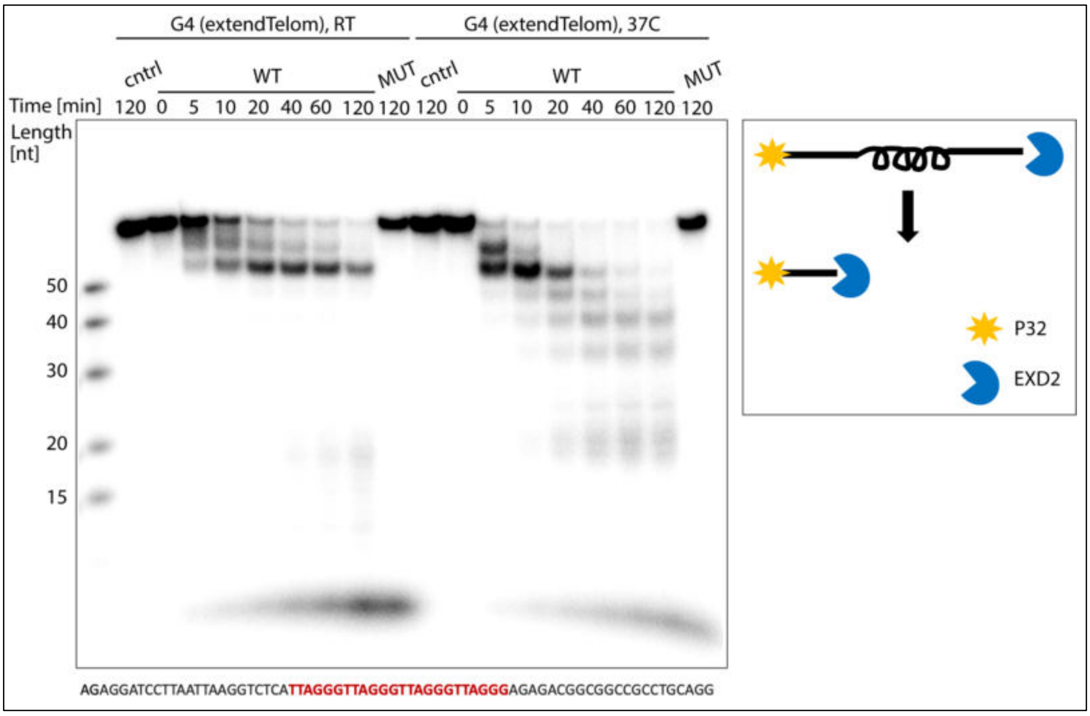


Figure 54. EXD2 WT, contrary to (D108A E110A) mutant, digests sequence that may form telomeric G-quadruplexes. Nucleotide composition of ‘extendTelom’ sequence is shown beneath the gel, with red letters indicating position of possible G-quadruplex formation. 5’-radiolabelled DNA substrate (extendTelom; ~ 1 nM) was incubated with EXD2 WT (5.3 nM) or EXD2 (D108A E110) MUT (5.3 nM) for indicated amounts of time at RT or 37 ℃. Reactions were performed in RT to increase stability of the substrate. The reaction buffer contained 5 mM MnCl$_2$ and ~ 13 mM KCl. Low KCl concentration could decrease stability of G4s during the reaction. Samples were resolved in TBE-urea polyacrylamide gel and visualised by phosphor imaging. Representative image of one of three repeats.

To validate if the tested substrates form G4s, I used N-methyl mesoporphyrin IX (NMM). NMM has higher affinity for G4s than for dsDNA and it emits fluorescence upon G4 binding (Endoh et al., 2016). After incubation of oligonucleotides with NMM (in the presence of 10 mM Tris pH 7.5 and 100 mM KCl), I measured its fluorescence intensity (3.4.7). As a control, I used a solution without any oligonucleotide. I also tested modified ‘extendMyc’ and ‘extendTelom’ substrates, where the middle G from each GGG sequence was replaced with A to prevent G4s formation. NMM fluorescence intensity values were higher in the presence of substrates which should form G-quadruplexes

(Table 18). However, the difference between values obtained for telomeric G4 sequence (extendTelom) and the values for substrate not forming G4s (Mutated extendTelom) is small. This may indicate that G4s are not efficiently formed by this sequence.

Table 18. Fluorescence of NMM in the presence of different DNA substrates. Values after subtraction of the control.

Tested substrate	450 nm / 610 nm
'extendMyc' (G4)	**39524**
'Mutated extendMyc' (non-G4)	2376
'extendTelom' (G4)	**4765**
'Mutated extendTelom' (non-G4)	2447

I tested EXD2 activity on mutated 'extendMyc' and 'extendTelom' substrates, where the "middle" G from each GGG sequence was replaced with A to prevent G4 formation. 'Mutated extendMyc' sequence was efficiently digested by EXD2 WT, indicating that 'extendMyc' sequence forms G4s which block EXD2 activity (Figure 55). The digestion pattern of sequences forming telomeric G4s (extendTelom) differed from the non-G4 forming sequence (mutated extendTelom), (Figure 55). Here, when the "middle" G from each (TTAGGG) repeat was replaced with A to prevent G4 formation, EXD2 digested the substrate to shorter products (the 50 nt long product disappeared as indicated by an arrow).

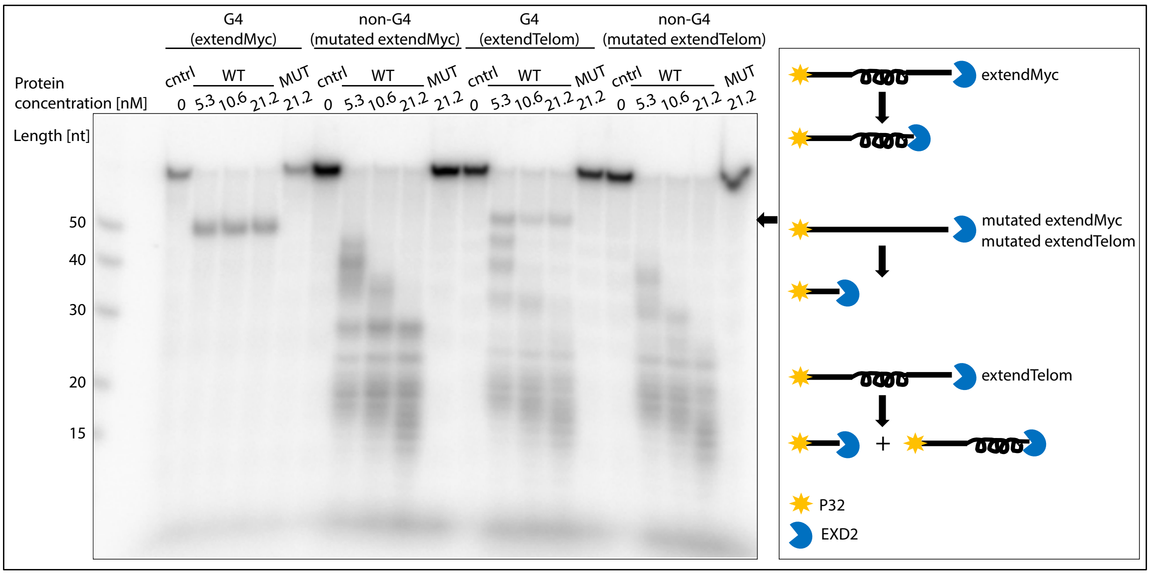


Figure 55. EXD2 WT, contrary to the (D108A E110A) mutant, digests sequence that may form telomeric G4s (extendTelom) and sequences that do not form G4s (mutated extendTelom, mutated extendMyc). EXD2 cannot digest G4s in MycPu22 (part of the extendMyc sequence). The non-G4 forming sequences are digested to shorter products than telomeric G4s (the 50 nt long product disappeared as indicated by an arrow). As the final KCl concentration in the reaction is low (13 mM), it is possible that G4s were not stable. 5'-radiolabelled DNA substrates (~ 1 nM) were incubated with EXD2 WT (0; 5.3; 10.6; 21.2 nM) or EXD2 (D108A E110) MUT (21.2 nM) for 1 h at 37 ℃. The reaction buffer contained 5 mM $MnCl_2$. Samples were resolved in TBE-urea polyacrylamide gel and visualised by phosphor imaging. Representative image of one of two repeats.

Recent data by (Junyeop Lee et al., 2022) showed that MRE11 resects telomeric G4s. One of the substrates they used for their *in vitro* reactions were ‘TelG5’ and ‘TelG5-G3’. The TelG5 sequence consists of $(TTAGGG)_4$ with GG and TTAG flanking at the 5’- and 3’- ends respectively. TelG5-G3 is based on the same sequence but one of the middle Gs is replaced with T (GGT TA**G GG**T TA**G GG**T TA**G TG**T TA**G GG**T TAG). According to (Junyeop Lee et al., 2022), TelG5 can form G4, whereas TelG5-G3 mimics G3 intermediate structures.

Using NMM, I validated if TelG5 and TelG5-G3 forms G4s, and I tested EXD2 nuclease activity on both DNA substrates. Observed fluorescence values of NMM were higher in the presence of ‘TelG5’. However, as the fluorescence differences are minimal (Cntr = 10112; TelG5-G3 = 11555; TelG5 = 14334), it is possible that G4s were not formed. EXD2 partially digested both substrates (Figure 56), with slightly higher activity on TelG5-G3 than on TelG5. The digestion pattern highly resembled resection performed by MRE11 shown by (Junyeop Lee et al., 2022). Unfortunately, results presented by (Junyeop Lee et al., 2022) do not include a nucleotide ladder so it is impossible to analyse extend of the digestion and compare it with EXD2 activity.

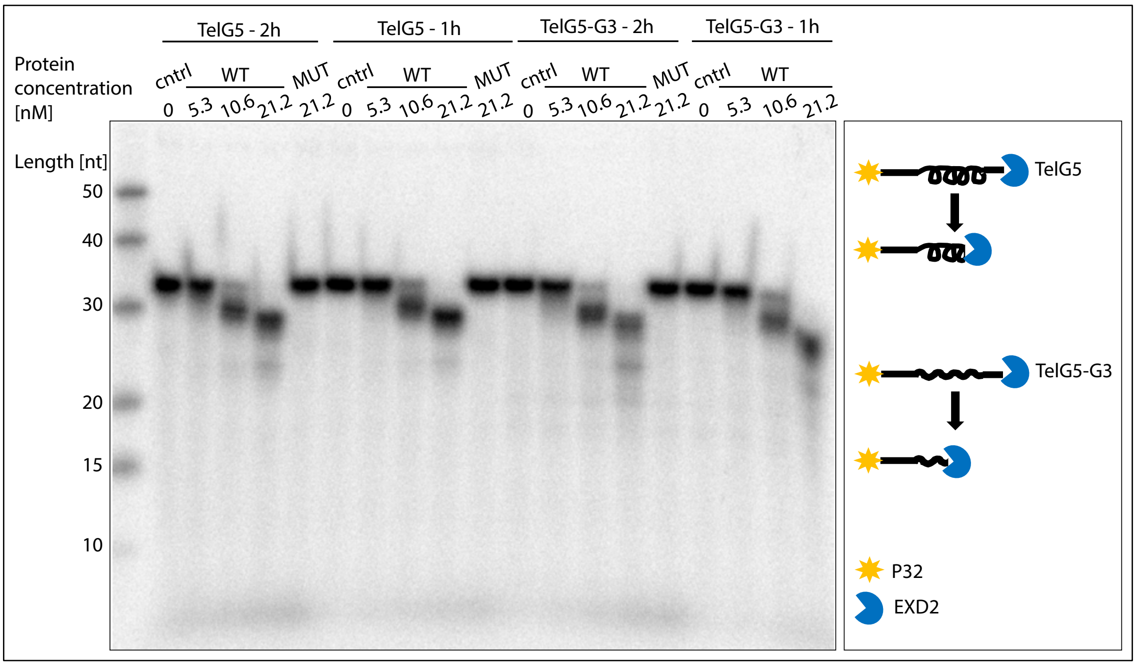


Figure 56. EXD2 WT, contrary to the (D108A E110A) mutant, partially digests sequences of TelG5 (which may form a telomeric G4) and TelG5-G3 (telomeric G3 intermediate). 5’-radiolabelled DNA substrates (~ 1 nM) were incubated with EXD2 WT (0; 5.3; 10.6; 21.2 nM) or EXD2 (D108A E110) MUT (21.2 nM) for either 1 or 2 h, at 37℃. The reaction buffer contained 5 mM $MnCl_2$ and ~ 13 mM KCl. Low KCl concentration could decrease stability of G4s during the reaction. Samples were resolved in TBE-urea polyacrylamide gel and visualised by phosphor imaging. Representative image of one of three repeats.

Conclusions

EXD2's activity on substrates which may form G-quadruplexes differs depending on the tested sequence. EXD2 digests sequences predicted to form telomeric G4s, but its activity is completely inhibited by G4s in the MycPu22 sequence. To further validate these results, it is essential to repeat the reactions in the wide range of KCl concentrations, at they affect stability of G4s. As the tested substrates are predicted to form either weak G4s (telomeric sequences) or very strong G4s (MycPu22 sequence), it would be beneficial to test EXD2 activity on other G4s forming sequences.

5.4 EXD2 nuclease activity on closed ssDNA

EXD2 contains a predicted HNH-like domain (Park et al., 2019), which is present in some endonucleases (discussed in more detail in 6.1). Therefore, to test if EXD2 can exhibit endonuclease activity, I used a ΦX174 substrate, of 5386 bases in length. It is a single stranded viral DNA commercially isolated from bacteriophage ΦX174. According to the producer (NEB), more than 85 % of the molecules are circular. This substrate requires initial endonuclease cleavage to undergo further resection, as was shown with the MRN complex (Broderick et al., 2016).

EXD2 was incubated with ΦX174 between 2 and 8 h (Figure 57A). Results of the digestion were visualized using SYBR Gold nucleic acid stain. During the 2 h reaction, very little substrate was digested. However, 8 h incubation resulted in nearly complete disappearance of the substrate suggesting that EXD2 may have some endonuclease activity. Substrate incubation with exonucleolytically dead mutant (D108A E110A) resulted in appearance of a band that can corresponds to either cleaved linearized ΦX174 or DNA contamination present in the protein purification prep. As the intensity of the substrate band do not decrease, it is possible that SYBR Gold stains the DNA copurified with the protein. The reactions were repeated, but this time EXD2 was incubated with ΦX174 overnight (Figure 57 B). As the substrate was almost fully digested, it confirms that EXD2 WT may nick closed ssDNA and then process it exonucleolytically. The nicking activity is not very efficient as EXD2 requires hours to perform the digestion. The presence of ATP did not significantly affect the resection.

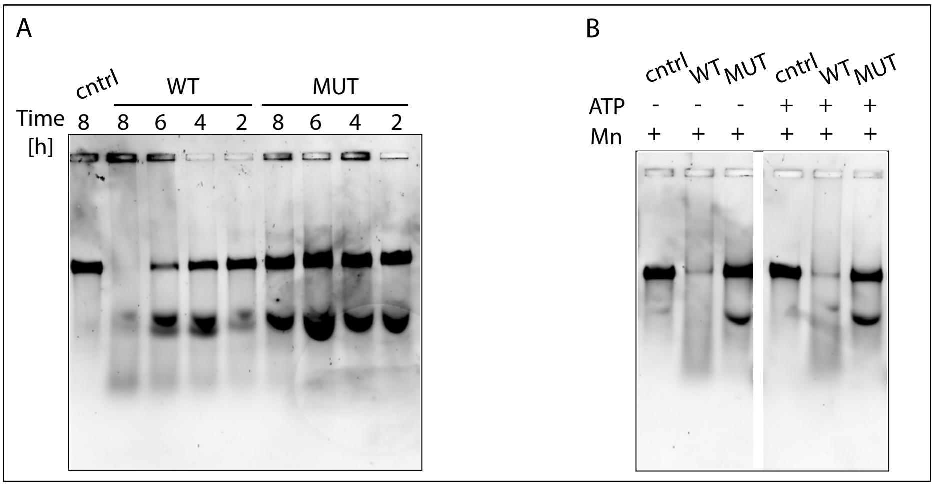


Figure 57. Activity of EXD2 WT and (D108A E110A) mutant on ΦX174 substrate. Circular ssDNA (ΦX174; 10 µg/ml) was incubated with EXD WT and (D108A E110A) MUT at 37°C. Samples were resolved in a 1.2 % TAE agarose gel stained with SYBR Gold. A) Reactions performed in the presence of 5 mM Mn^{2+} for 2 – 8 h, protein concentration: ~ 110 nM. B) Reactions performed overnight in the presence of 5 Mn^{2+} and + / - 1 mM ATP, protein concentration: ~ 46 nM.

Differences in the amount of digested substrate between control, EXD2 WT and MUT, indicate that WT EXD2 can perform the cleavage but the tested conditions may not be favourable. Therefore, I tested EXD2 endonuclease activity also in the presence of Mg^{2+}. When Mn^{2+} was replaced with Mg^{2+}, EXD2 WT exhibited extremely weak activity on ΦX174 substrate, even after overnight reaction (Figure 58 A). Supplementation with ATP did not affect the activity. The presence of both Mg^{2+} and Mn^{2+} (Figure 58B), resulted in higher digestion efficiency, but lower than for Mn^{2+} alone.

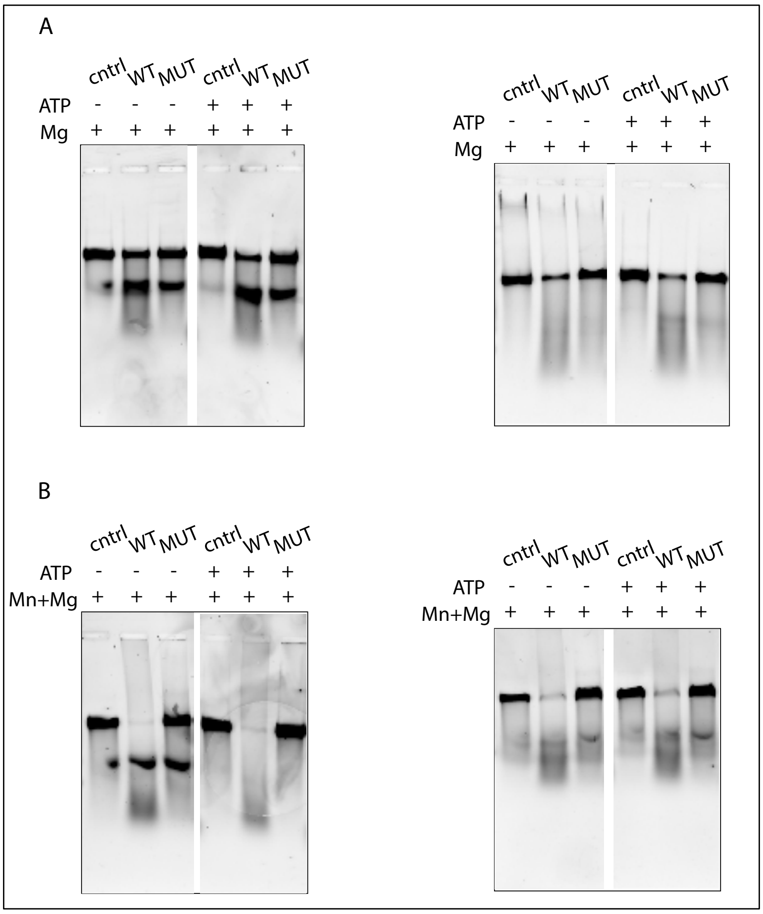


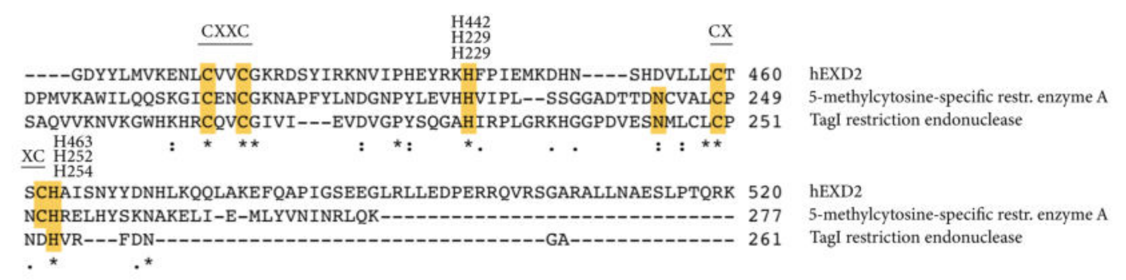
Figure 58. Activity of EXD2 WT and (D108A E110A) mutant on ΦX174 substrate (10 μg/ml). Proteins (~ 46 nM) were incubated with plasmid overnight at 37℃. Samples were resolved in a 1.2 % TAE agarose gel stained with SYBR Gold A) Reactions in the presence of 5 mM Mg^{2+} and +/- 1 mM ATP. Two repeats. B) Reactions in the presence of 5 mM Mn^{2+}, 5 mM Mg^{2+} and +/- 1 mM ATP. Two repeats.

Based on these experiments, it is difficult to conclude if the exonucleolytic-dead EXD2 mutant (D108A E110A) can nick closed ssDNA. EXD2 WT cleaves a ΦX174 substrate but its endonuclease activity may require a cofactor, which is not present in the reaction.

5.5 Mutation of EXD2's HNH domain

5.5.1 Design of the mutations in the HNH domain

To further validate if EXD2 may act as an endonuclease, I performed a sequence homology search between the HNH endonuclease-like domain of EXD2 and other proteins. The HHpred server indicated similarities between EXD2 and several endonucleases, including *Escherichia coli* 5-methylcytosine-specific restriction enzyme A (EcoKMcrA) and *Thermocrispum agreste* restriction endonuclease (TagI). Both of the proteins belong to the HNH endonuclease family, named after crucial His-Asn-His residues. The first histidine from the HNH motif is considered to act as the nucleophile activator (H229 in EcoKMcrA; H229 in TagI), whereas the second histidine is involved in the coordination of the active site metal ion (H252 in EcoKMcrA; H254 in TagI), (Czapinska et al., 2018), (Kisiala, Copelas, Czapinska, Xu, & Bochtler, 2018). The H229A and H252A EcoKMcrA mutants, as well as H229A and H254L TagI mutants are catalytically inactive (Czapinska et al., 2018), (Kisiala et al., 2018). Sequence alignment between hEXD2, EcoKMcrA and TagI shows strict conservation of these histidine residues, highlighting their importance (Figure 59).



```
               CXXC                 H442                 CX
                                    H229
                                    H229
----GDYYLMVKENLCVVCGKRDSYIRKNVIPHEYRKHFPIEMKDHN----SHDVLLLCT 460  hEXD2
DPMVKAWILQQSKGICENCGKNAPFYLNDGNPYLEVHHVIPL--SSGGADTTDNCVALCP 249  5-methylcytosine-specific restr. enzyme A
SAQVVKNVKGWHKHRCQVCGIVI---EVDVGPYSQGAHIRPLGRKHGGPDVESNMLCLCP 251  TagI restriction endonuclease
            :  *  **        :  *:    *.     . .      : :  **
   H463
XC H252
   H254
SCHAISNYYDNHLKQQLAKEFQAPIGSEEGLRLLEDPERRQVRSGARALLNAESLPTQRK 520  hEXD2
NCHRELHYSKNAKELI-E-MLYVNINRLQK------------------------------ 277  5-methylcytosine-specific restr. enzyme A
NDHVR---FDN---------------------------------GA-------------- 261  TagI restriction endonuclease
. *      .*
```

Figure 59. Amino acid alignment of human EXD2 with two HNH endonucleases: EcoKMcrA and TagI. The conserved residues are indicated by stars *. https://www.ebi.ac.uk/Tools/msa/clustalo/

In addition to the HNH motif, other highly conserved residues between EXD2, EcoKMcrA and TagI are the tandem CXXC residues (for TagI it is one CXXC and one CXXD), (Figure 59). The cysteines from CXXC motifs are ligands for structural Zn^{2+} in HNH endonucleases. Their mutations (C210A; C213A) drastically reduce TagI nuclease activity (Kisiala et al., 2018). Mutations of corresponding cysteines in EXD2 (420CXXC to SXXS and 460CXXC to SXXS) are described by (Silva et al., 2018). The EXD2-C1/C2 mutant exhibits lower nuclease activity on ssRNA, both strands of RNA-DNA hybrid, ssDNA and dsDNA, yet it is not nucleolytic dead (Silva et al., 2018). Hence, mutations of histidine residues (H442 and H463) seemed more likely to affect potential EXD2 endonuclease activity, than mutations in CXXC motif.

To predict the role of H442 and H463 EXD2 residues, I decided to model EXD2 sequence (aa 411 – 471) on the structural template of EcoKMcrA. Together with Dr Guiller Abascal Palacios, we used the Phyre2 One 2 One threading program. According to this structural prediction, H463 may be involved in the coordination of the active site metal ion in EXD2, as it structurally overlaps with the position of H252 in EcoKMcrA. The H442 EXD2 residue is not shown in the (Figure 60), as it is predicted to be localized much further away.

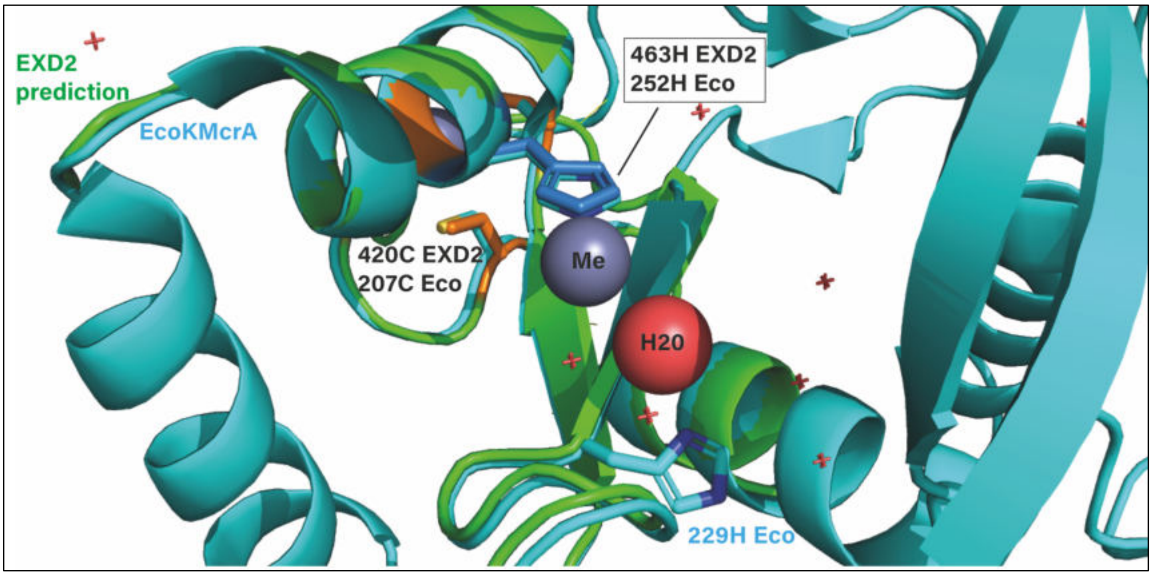
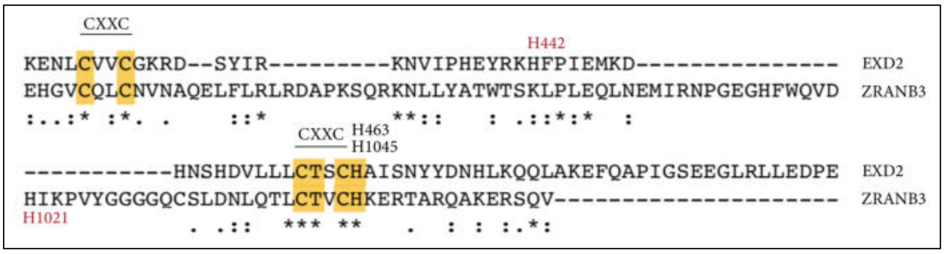


Figure 60. Result of Phyre2 One 2 One threading with Dr Guiller Abascal Palacios. EXD2 model – green, EcoKMcrA template (6ghc from PDB) - blue. Picture from PyMOL program. Red crosses indicate presence of water particles. According to this prediction, H463 of hEXD2 may be involved in the coordination of the active site metal ion.

Another HNH nuclease I decided to choose for EXD2 sequence alignment was the Zinc-finger, RAN-binding domain containing 3 (ZRANB3) protein (not indicated by HHpred server). ZRANB3 is an eukaryotic ATP-dependent endonuclease which is recruited to DNA breaks and stressed replication forks (Sebesta, Cooper, Ariza, Carnie, & Ahel, 2017).



```
    CXXC
                                                H442
KENLCVVCGKRD--SYIR---------KNVIPHEYRKHFPIEMKD---------------  EXD2
EHGVCQLCNVNAQELFLRLRDAPKSQRKNLLYATWTSKLPLEQLNEMIRNPGEGHFWQVD  ZRANB3
:..:* :*. .     ::*         **::    : .::*:*   :
                     CXXC H463
                          H1045
-----------HNSHDVLLLCTSCHAISNYYDNHLKQQLAKEFQAPIGSEEGLRLLEDPE  EXD2
HIKPVYGGGGQCSLDNLQTLCTVCHKERTARQAKERSQV---------------------  ZRANB3
H1021
              . .::  *** **    .  : : :.*:
```

Figure 61. Amino acid alignment of human EXD2 with ZRANB3. Conserved residues are indicated by stars *. https://www.ebi.ac.uk/Tools/msa/clustalo/

The double CXXC motif, as well as the following histidine (H463 for EXD2; H1045 for ZRANB3) are conserved between ZRANB3 and EXD2 (Figure 61). However, the EXD2 H442 residue (conserved between hEXD2, EcoKMcrA and TagI) does not seem to be conserved in ZRANB3 (Figure 61). It may be the result of program limitations, as it is possible to align H442 (EXD2) with H1021 (ZRANB3). The H1021A ZRANB3 mutant, as well as H1045A mutant, exhibit no endonuclease activity (Sebesta et al., 2017).

The importance of the histidine residue from the CXXCH motif is consistent between the EcoKMcrA, TagI and ZRANB3 proteins. It corresponds to the H463 EXD2 residue, which may have a catalytic role in the nucleolytic cleavage. The H442 EXD2 residue is also highly conserved between endonucleases (EcoKMcrA, TagI and possibly ZRANB3).

According to COSMIC (Catalogue of somatic mutations in cancer) the H442 EXD2 mutation is present in one sample of large intestine carcinoma, with a pathogenic prediction score (according to Functional Analysis through Hidden Markov Models - FATHMM). Hence, to validate any potential endonuclease activity of EXD2, I designed a mutant bearing both H442A and H463A mutations. This mutated plasmid was prepared by GenScript company.

5.5.2 Purification of full-length EXD2 (H442A and H463A) mutant

10xHis-TEV-hEXD2-3C-2xStrep, endonuclease mutant (H442A and H463A)

EXD2 endonuclease mutant (H442A, H463A) was purified according to the full-length 10xHis-TEV-hEXD2-3C-2xStrep WT and (D108A E110A) mutant purification protocols. However, due to a very low yield, only Strep-Tag and His-Tag affinity chromatography were performed (no gel filtration), (Figure 62). Purification of the (H442A, H463A) mutant was conducted only once, and was not assessed by mass spectrometry analysis.

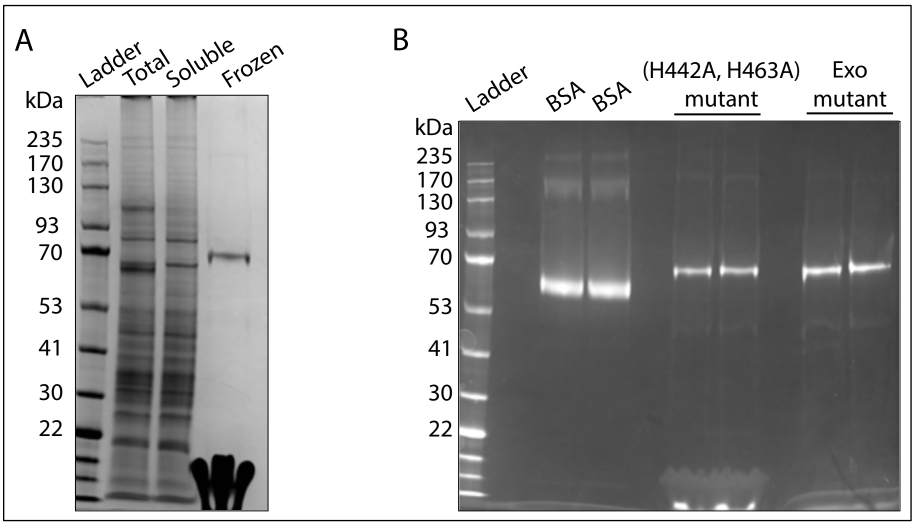


Figure 62. A) Purification of full-length 10xHis-TEV-hEXD2-3C-2xStrep (H442A, H463A) mutant, Coomassie stained gel. Total – lysate before centrifugation; Soluble – soluble lysate after centrifugation; Frozen – final result of the purification. B) Assessing concentration of the purified 10xHis-TEV-hEXD2-3C-2xStrep (H442A, H463A) mutant by SDS-PAGE and Coomassie staining. This is negative image used for protein quantification. Concentration of BSA is 0.1 mg/ml. Band intensity ratio between EXD2 (H442A, H463A) mutant and EXD2 exonuclease mutant (D108A E110A) is 0.81.

After His-Tag affinity chromatography, buffer exchange was performed on a Sartorius Vivaspin Centrifugal Concentrator, while concentrating the protein. A shadow on the bottom of the Coomassie stained gel corresponds to the presence of the detergent (Figure 62). This indicates that concentration of the Triton X-100 (0.5 % in the HisTrap elution buffer) was not efficiently decreased (desired concentration was 0.05 %, as in the gel filtration buffer). The yield of the purification, from 1L of insect cells, was around 83 μg (0.055 mg/ml; 1.5 ml).

5.5.3 Activity of EXD2 (H442A and H463A) mutant

10xHis-TEV-hEXD2-3C-2xStrep (H442A and H463A) mutant

To validate activity of purified EXD2 (H442A and H463A) mutant, I used ssDNA (70 nt), dsDNA of the same sequence and dsDNA with PS modifications introduced on the strand opposite to the labelled one (Figure 63). As discussed in more detail in the next chapter (5.6), the presence of the PS block opposite to the labelled strand stimulates EXD2 WT activity. The EXD2 (H442A and H463A) mutant has very weak activity on ssDNA and no activity on both types of dsDNA. Even when the concentration of the mutant was 10 times higher than the WT, (H442A and H463A) did not digest dsDNA. This is interesting, as this activity is different to WT EXD2, which prefers digestion of dsDNA over ssDNA.

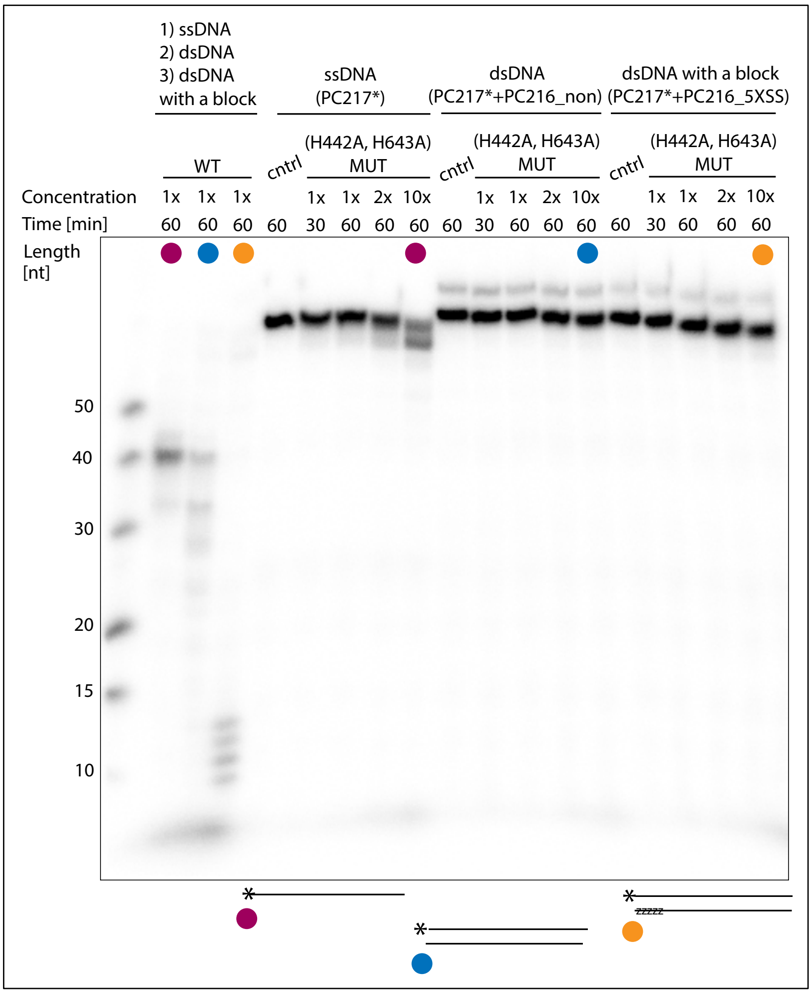


Figure 63. EXD2 (H442A and H463A) mutant exhibits very weak activity on ssDNA and no activity on dsDNA. This activity is contrary to EXD2 WT, which prefers digestion of dsDNA over ssDNA. To facilitate comparison between EXD2 WT and (H442A and H463A) mutant, different types of substrates (~ 1 nM) are marked with different colours. Reactions were carried at 37℃ for either 30 min or 1 h, in the presence of 5 mM $MnCl_2$. Samples were resolved in TBE-urea polyacrylamide gel and visualised by phosphor imaging. Representative image of one of two/three repeats.

Activity of EXD2 (H442A and H463A) mutant on biotin blocked DNA is shown in the next chapter (5.6.4).

Conclusions

The HNH domain is important for EXD2 activity. Mutation of (H442A and H463A) dramatically reduced the nuclease activity on ssDNA and completely abolished this activity on dsDNA. To further validate these results, the EXD2 mutant (H442A and H463A) purification should be repeated. It is unlikely that the residual higher concentration of Triton X-100 (as it was exchanged on Vivaspin, not on the gel filtration column) is the reason of dramatically reduced activity of the mutant. EXD2 WT frozen in a high concertation of Triton (0.5 %), exhibited activity on both ssDNA and dsDNA (data not shown, tested after one of the trial purifications of EXD2 WT).

5.6 EXD2 nuclease activity on blocked DNA ends

As described in the previous chapter, EXD2 may exhibit weak endonuclease activity. Therefore, I tested EXD2's nuclease activity on DNA substrates with introduced different types of blocks. As DSBs in cells are usually 'dirty', I used blocks to mimic additional damage or protein adducts at the DNA ends.

5.6.1 Phosphorothioate modifications - overview

The phosphorothioate (PS) modification is formed when non-bridging oxygen in the oligonucleotide phosphate backbone is replaced by sulphur (Monia, Johnston, Sasmor, & Cummins, 1996). It is one of the most common synthetic modifications created to inhibit degradation of oligonucleotides by nucleases. The PS functional groups have been used to prevent exonucleolytic degradation *in vitro* by MRN (Deshpande et al., 2016) and MRX (Cannavo, Reginato, & Cejka, 2019). *In vivo*, DNA phosphorothioate modifications are widespread in many bacteria, including the human microbiome (Sun et al., 2020).

5.6.2 The impact of phosphorothioate modifications on EXD2-dependent nucleolytic processing

I used oligonucleotides with introduced PS modifications to mimic small blocks either at DNA ends or inside the sequence. First, to examine if PS inhibit EXD2 exonuclease activity, I used a 70 nt ssDNA substrate with five PS functional groups near the 3' terminus, followed by 10 nt (PC217_middle), (Figure 64 A). The sequence of the non-blocked substrate was published by (Cannavo et al., 2019) as PC217. Here, the modified substrate is digested exonucleolytically until EXD2 reaches position of the PS block which inhibit further digestion. The dsDNA, with PS modifications introduced 10 nt away from the 3' end of the labelled strand, is degraded the same way (Figure 64 B).

Introduction of additional PS functional groups, to block degradation of the unlabelled strand, does not affect the digestion of the blocked upper strand (Figure 64 B).

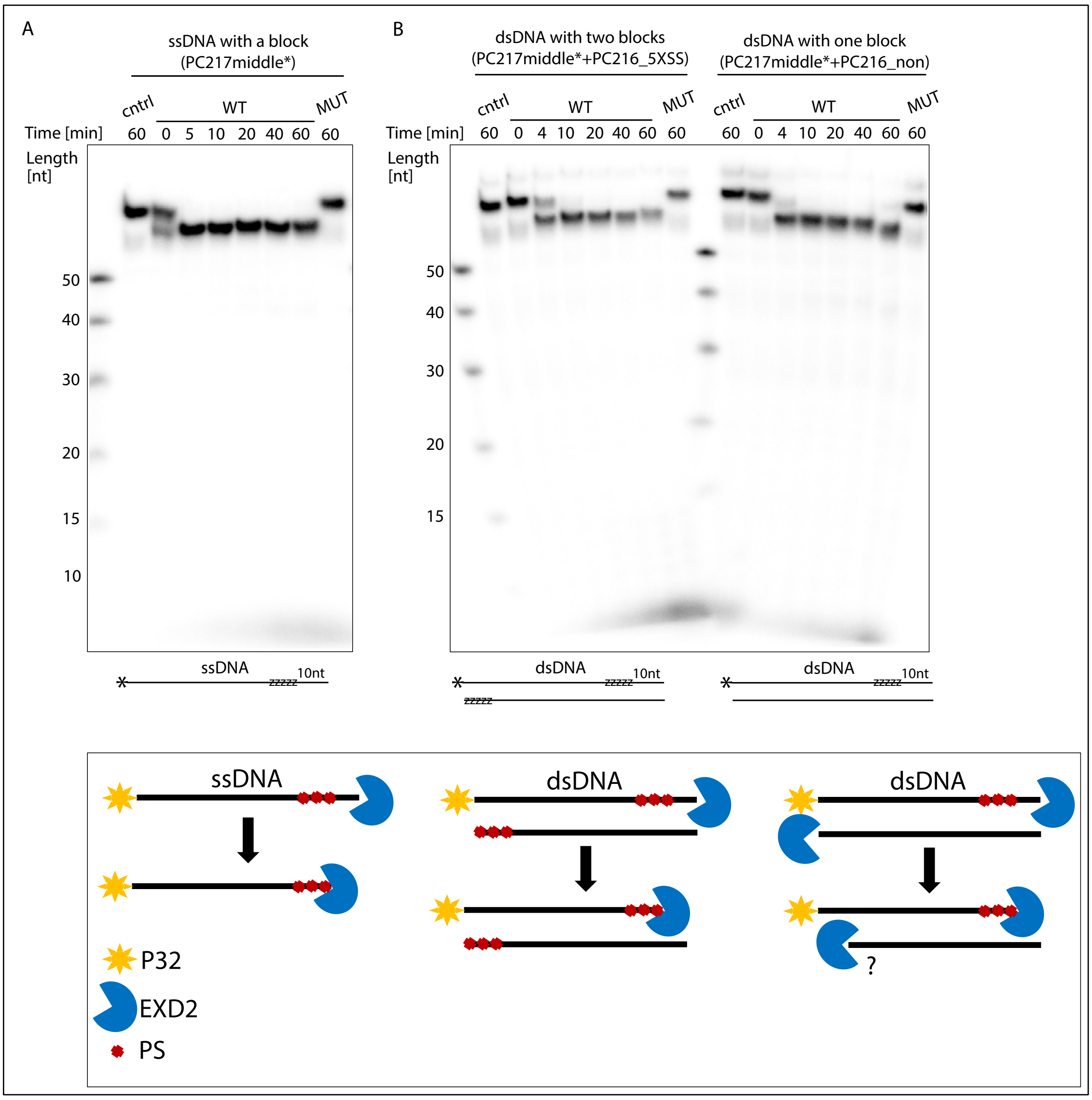


Figure 64. Purified WT EXD2, contrary to the (D108A E110A) mutant, displays 3'-5' exonuclease activity on nonblocked DNA end. PS modifications inhibit further digestion. 5'-radiolabelled DNA substrates (~ 1 nM) were incubated with EXD2 WT (5.3 nM) or EXD2 (D108A E110) MUT (5.3 nM) for the indicated amounts of time at 37℃. Samples were resolved in TBE-urea polyacrylamide gel and visualised by phosphor imaging. The reaction buffer contained 5 mM $MnCl_2$. Representative image of one of three repeats.

To further elucidate the impact of the PS block on EXD2's digestion pattern I used dsDNA with PS modifications introduced on the opposite strand to the labelled one (Figure 65). I compared it with the digestion of ssDNA of the same sequence, but without the PS. The ssDNA was exonucleolytically digested with a typical 'ladder like' resection pattern (Figure 65). The same 70 bp dsDNA sequence with five PS modifications introduced on the 3' end of the opposite strand, was either exonucleolytically digested with very fast kinetics, or resected exonucleolytically up to around 10 nt and then cut endonucleolytically close to the 5' end of the labelled strand (Figure 65). The lack of visible middle-length products may suggest that a PS block stimulates EXD2's endonuclease activity. dsDNA of the same sequence, but without the blocks (PC217*+PC216_non, Figure 67 A), was digested with the same pattern as ssDNA (PC217*). Hence, the difference in EXD2's activity is block dependent.

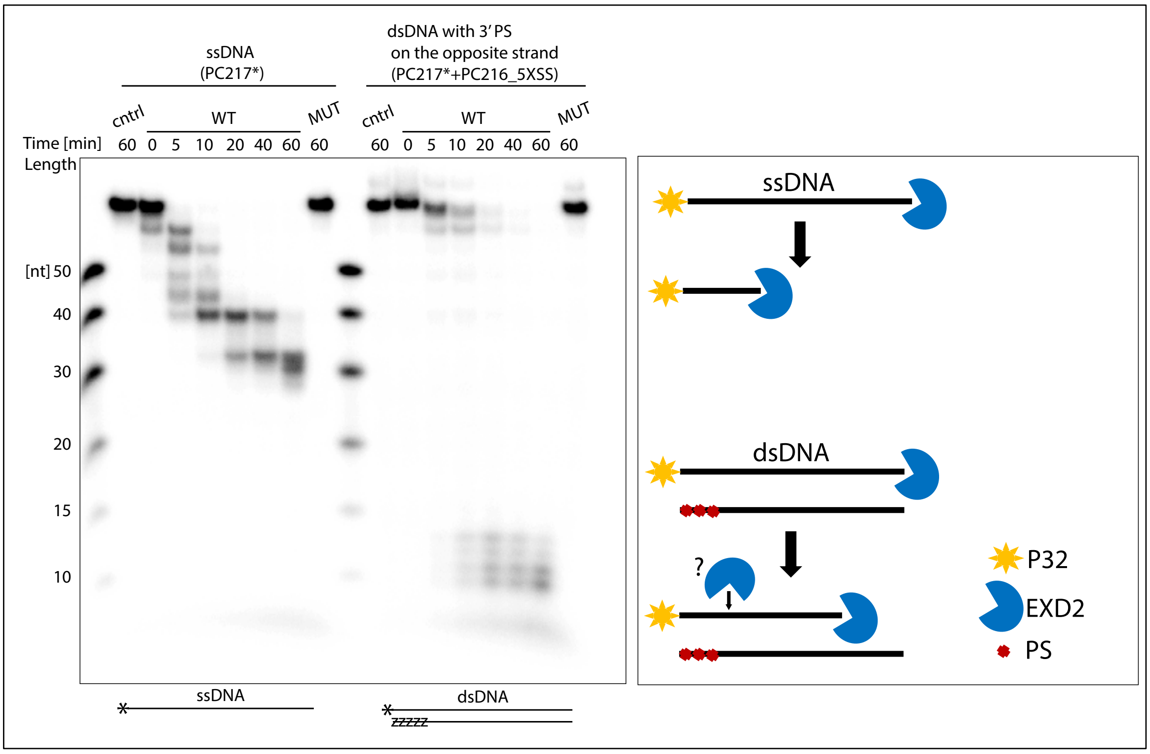


Figure 65. The presence of PS modifications can alter the activity of EXD2. EXD2 WT, contrary to the (D108A E110A) mutant, exhibits 3'-5' exonuclease activity on ssDNA. Introduction of PS modifications on the opposite strand, changes the digestion pattern of the labelled strand. 5'-radiolabelled DNA substrates (~ 1 nM) were incubated with EXD2 WT (1.6 nM) or EXD2 (D108A E110) MUT (1.6 nM) for the indicated amounts of time at 37°C. Samples were resolved in TBE-urea polyacrylamide gel and visualised by phosphor imaging. The reaction buffer contained 5 mM $MnCl_2$. Representative image of one of three repeats.

To be sure that the change of digestion pattern is not sequence dependent, I tested EXD2's activity on a different substrate. Here, a 50 bp dsDNA with two PS modifications on the 3' end of the opposite strand to the labelled one, had a different digestion pattern than a non-blocked 50 bp dsDNA substrate of the same sequence (Figure 66). The change of the digestion pattern in (Figure 66) is analogous to the change observed in (Figure 65). Here, the length of the shortest products is similar. However, resection of the blocked substrate from (Figure 66) results in some middle-length products observed after 5 min of reaction. The observed digestion kinetics of the PS blocked 50 bp substrate from (Figure 66) may be slower than the PS blocked 70 bp substrate from (Figure 65), which may account for the presence of some middle-length products. This is line with chapter (5.1) which shows that longer substrates are preferable for degradation by EXD2. However, it is still possible that the PS block stimulates the potential endonuclease activity of EXD2 as the digestion pattern changes in a similar way in both cases and not all middle-length products are visible.

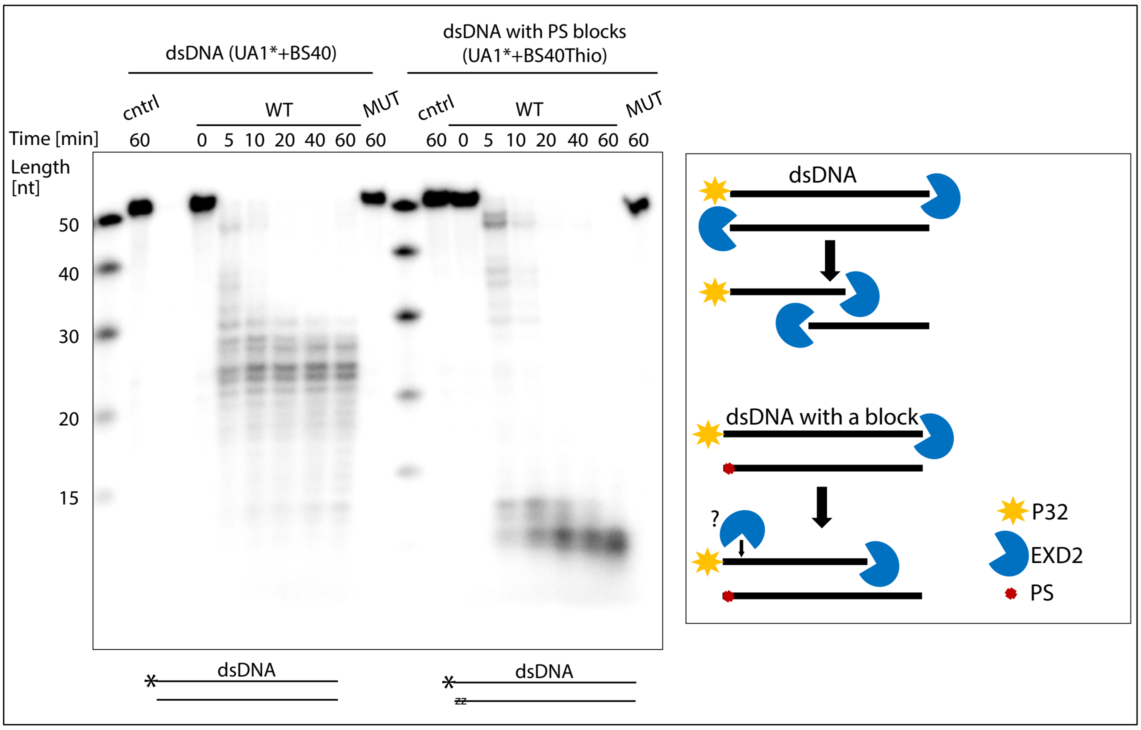


Figure 66. EXD2 WT, contrary to the (D108A E110A) mutant, exhibits 3'-5' exonuclease activity on dsDNA. Introduction of a PS modifications on the opposite strand, changes the digestion pattern of the labelled strand. 5'-radiolabelled DNA substrates (~ 1 nM) were incubated with EXD2 WT (5.3 nM) or EXD2 (D108A E110) MUT (5.3 nM) for indicated amounts of time at 37℃. Samples were resolved in TBE-urea polyacrylamide gel and visualised by phosphor imaging. The reaction buffer contained 5 mM $MnCl_2$. Representative image of one of four repeats (two performed on UA1*+BS40Thio and two performed on UA1_60*+BS40Thio).

To check if the change of the digestion pattern depends on the position of the PS block, I used a 70 bp dsDNA: 1) without any block (PC216_non), 2) with one strand containing five PS functional groups 25 nt from the 3' end (PC216Xafter40); 3) with one strand containing five PS functional groups 16 nt from the 3' end (PC216near3) and 4) with one strand containing five PS functional groups on the 3' end (PC216_5XSS), opposite to the P32 label (Figure 67 A). When five PS functional groups are positioned 25 nt from the 3' end, the majority of the digestion product is ~ 32 nt long. When the substrate has blocks 16 nt from the end, the main product is ~22 nt long. This may be because EXD2 stalls on the sequence opposite to the block, hence the position of the PS functional groups inside the sequence affects the digestion pattern of the opposite strand. The labelled strand of all three substrates (DNA without any block, with blocks 25 nt and 16 nt from the 3' end) is digested in a 'ladder like' pattern, with products ranging from 70 till 12 nt in length. This suggests that any possible immobilisation of EXD2 is only temporary, as products of different lengths are being generated. Interestingly, the digestion pattern of the substrate with PS blocks on the 3' end does not show products of intermediate length (in line with Figure 65). This suggests that if there is endonucleolytic activity taking place, it is stimulated only when the block is present on the very end of the opposite strand. Labelling of the opposite strand to the one labelled in (Figure 67A) shows that if there are no PS blocks, both strands of the dsDNA are efficiently digested (Figure 67B). When 5 PS blocks are introduced 10 nt away from the 3’ end of the unlabelled strand, the main digestion product is 15 nt long (Figure 67B). This is in line with the observation that the length of the main product depends on the position of the block on the opposite strand.

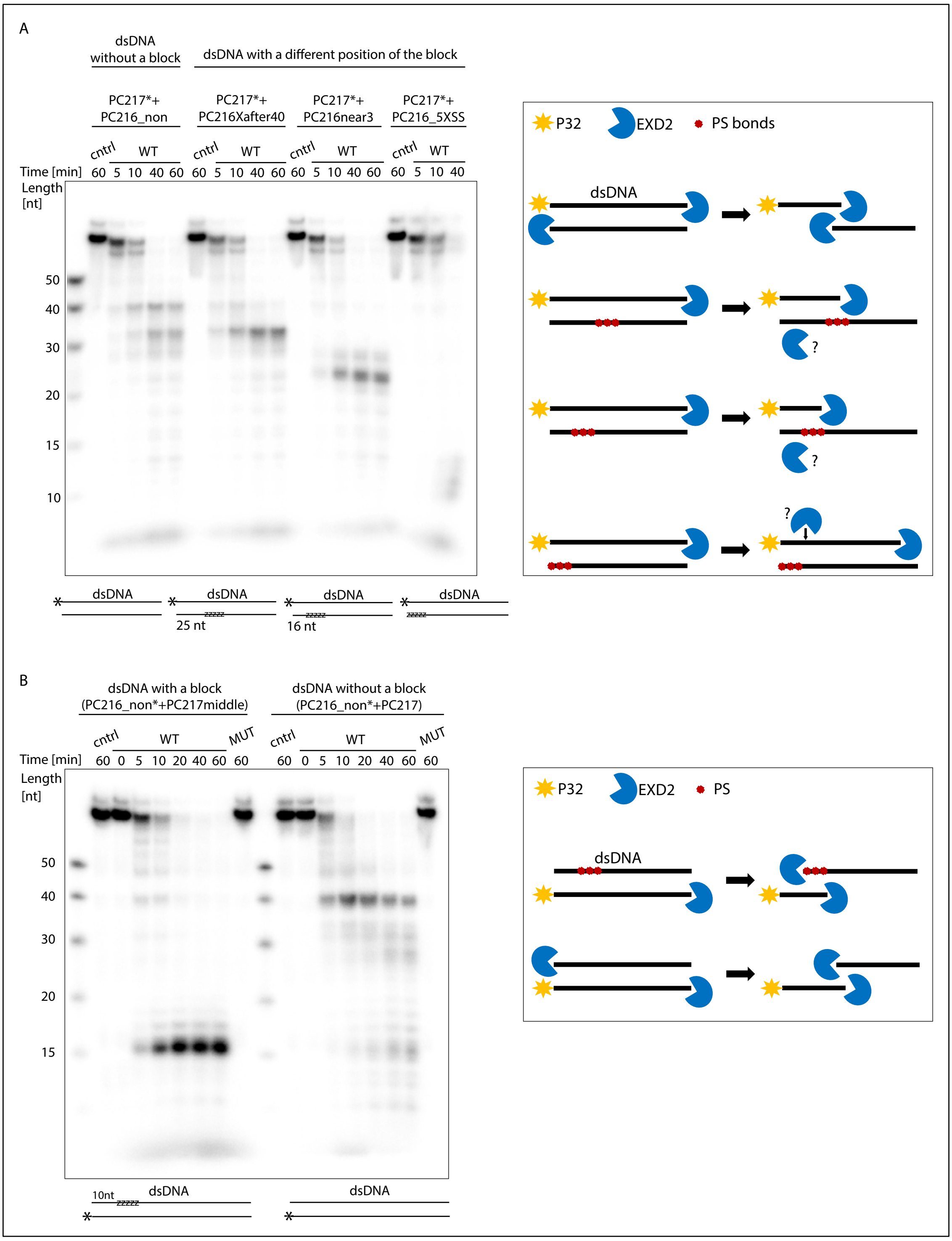


Figure 67. Position of the PS modifications on the opposite strand changes EXD2's digestion pattern of the labelled strand. 5'-radiolabelled DNA substrates (~ 1 nM) were incubated with EXD2 WT (5.3 nM) or EXD2 (D108A E110) MUT (5.3 nM) for the indicated amounts of time at 37℃. Samples were resolved in TBE-urea polyacrylamide gel and visualised by phosphor imaging. The reaction buffer contained 5 mM MnCl$_2$. Representative image of one of three repeats.

To check the digestion pattern of the blocked strand I used the same dsDNA substrate as in (Figure 65) and in (Figure 67 A), (PC217+PC216_5XSS), but this time I labelled the strand that was blocked (PC216_5XSS). The five PS modifications on the 3' end completely inhibited exonucleolytic digestion of the blocked strand of dsDNA (Figure 68).

Introduction of five PS modifications near the 3’ end of the second strand, to have both strands with PS blocks, had little effect on the digestion (Figure 68). The appearance of a very weak signal, 15 nt long (lanes 6-7), may indicate inefficient endonucleolytic cleavage opposite to the block.

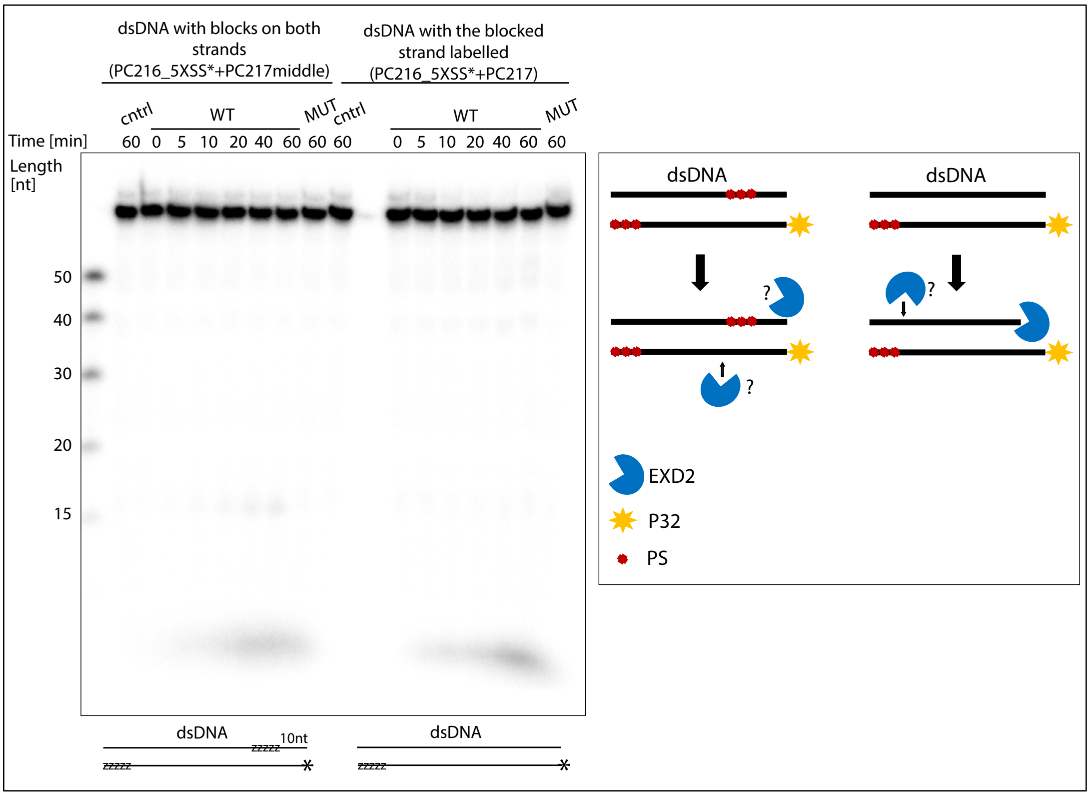


Figure 68. ***PS blocks introduced on 3’ end of dsDNA inhibit exonucleolytic digestion of the strand by EXD2. Presence of PS blocks on both DNA strands does not promote efficient digestion but may result in a very weak endonuclease activity. 5’-radiolabelled DNA substrates (~ 1 nM) were incubated with EXD2 WT (5.3 nM) or EXD2 (D108A E110) MUT (5.3 nM) for the indicated amounts of time at 37°C. Samples were resolved in TBE-urea polyacrylamide gel and visualised by phosphor imaging. The reaction buffer contained 5 mM MnCl$_2$. Representative image of one of three repeats.***

Next, I analysed what is the digestion pattern when blocks on the opposite strands are positioned closer to each other. I used dsDNA with one strand PS blocked after 19 nt from 5’ end (PC216middle) and one strand PS blocked after 10 nt from 3’ end (PC217middle). I labelled either first or second strand on the 5’ terminus (Figure 69). The bottom strand was exonucleolytically digested until EXD2 reached position of the five PS functional groups. Hence, the main product is ~25 nt long. The upper strand was digested exonucleolytically until reaching the block, with no clear sign of endonucleolytic activity.

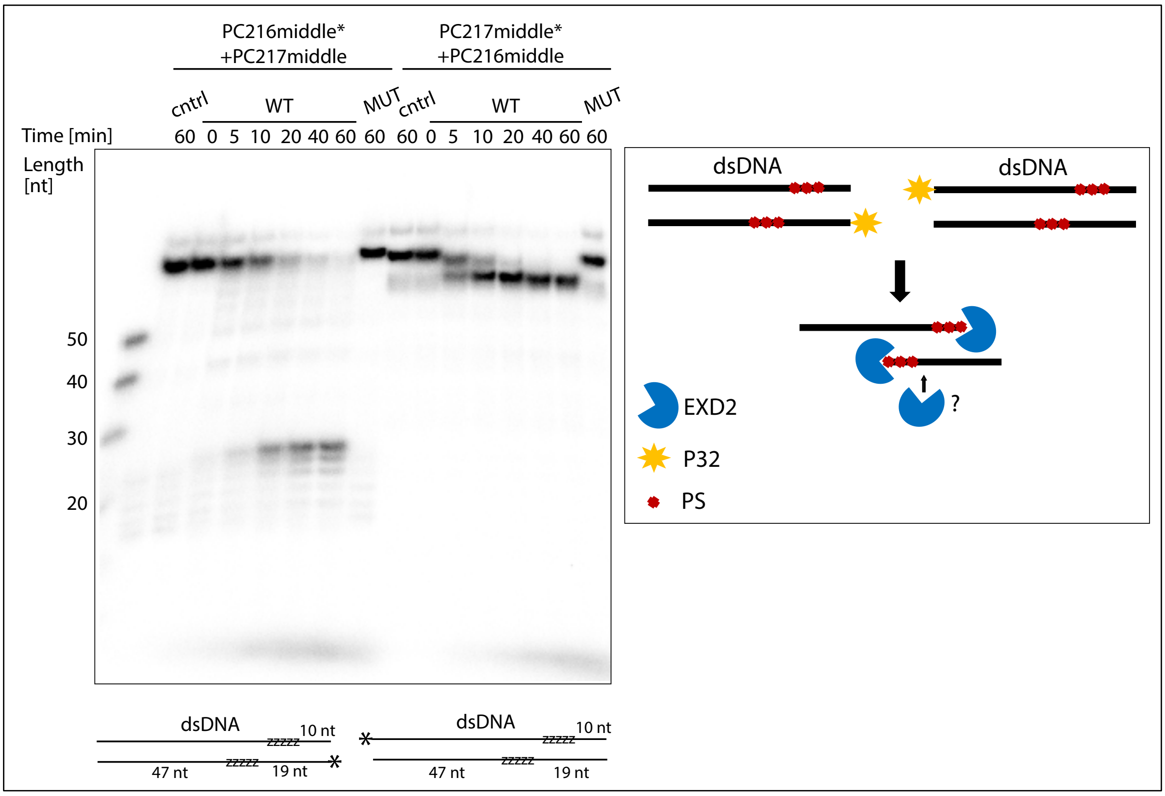


Figure 69. PS blocks inhibit exonucleolytic digestion of DNA by EXD2. It is unlikely that blocks on two strands close to each other may promote nicking activity. 5'-radiolabelled DNA substrates (~ 1 nM) were incubated with EXD2 WT (5.3 nM) or EXD2 (D108A E110) MUT (5.3 nM) for the indicated amounts of time at 37°C. Samples were resolved in TBE-urea polyacrylamide gel and visualised by phosphor imaging. The reaction buffer contained 5 mM $MnCl_2$. Representative image of one of two repeats.

To understand how PS blocks affect EXD2's nucleolytic activity, I used a substrate with two sets of PS blocks on the opposite strand to the labelled one (Figure 70). The digestion pattern of the labelled strand was the same for substrate having one set of five PS blocks as for substrate with two sets of PS blocks (Figure 70). It is possible that one EXD2 molecule may get stuck on the 3' end of the blocked strand occluding movement of the second EXD2 molecule digesting the labelled strand from the opposite direction. This would mean that the digestion pattern of the labelled strand is the result of a rapid exonucleolytic activity which is stopped when one EXD2 molecule bumps into the second one. It may explain why there is no difference in the digestion pattern between substrates containing one and two sets of blocks (Figure 70). However, if EXD2 stalls on the block and occludes the movement of the protein on the opposite strand, we should not see digestion products shorter than the length between the 3' end and the PS blocks, which is what we observe in (Figure 67). The labelled strand of the substrate with PS blocks introduced on the opposite strand 25 nt from 3' end (Figure 67) is digested in a ladder-like pattern, with products ranging from 70 till 12 nt. Therefore, in my opinion it is more likely that the digestion pattern in (Figure 70) may be the result of weak EXD2 endonuclease activity, which is stimulated by a PS block on the 3' end.

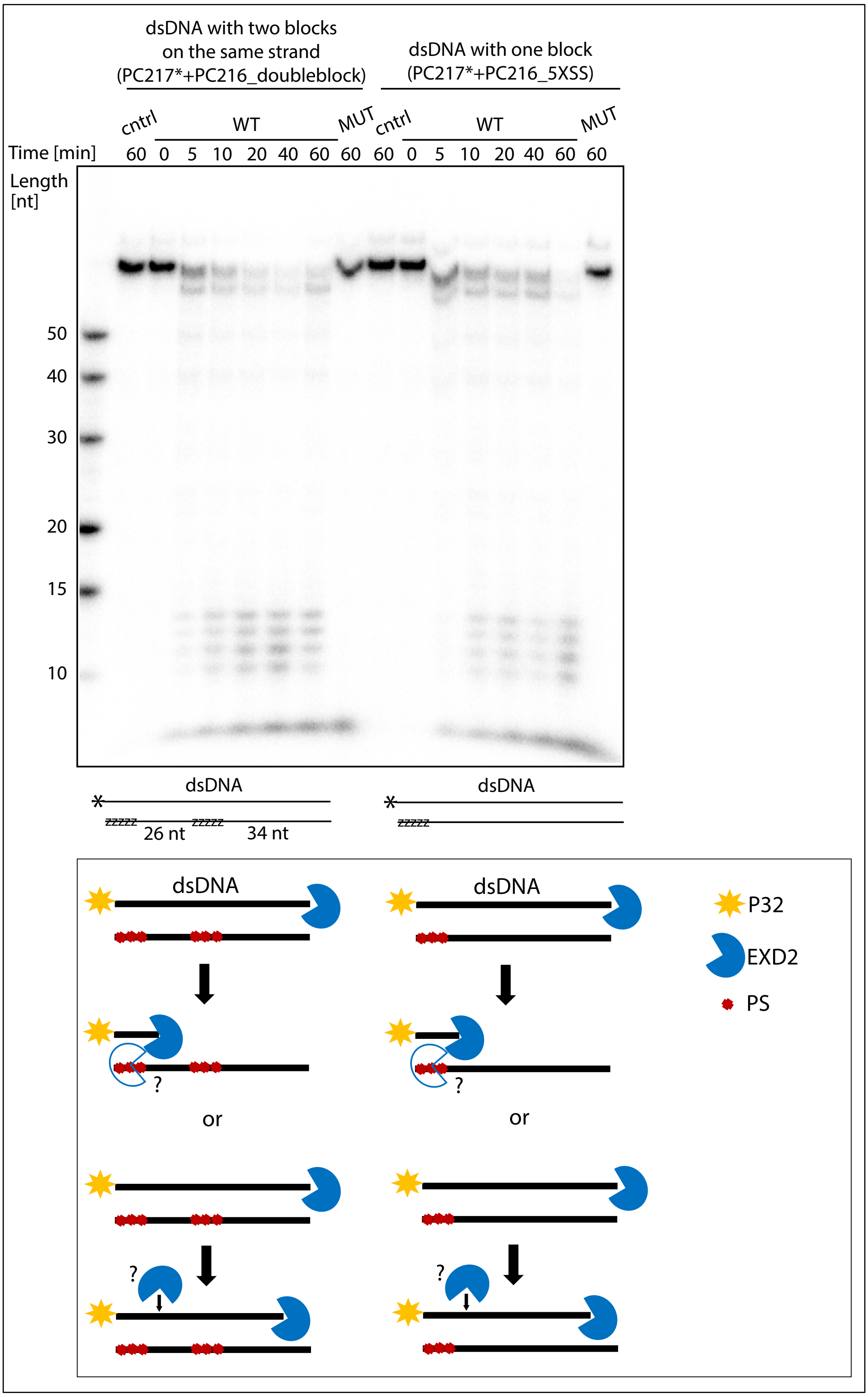


Figure 70. EXD2's digestion pattern of the labelled strand is the same for substrates having one set of PS blocks on the opposite strand, as for substrates with two sets of PS blocks. 5'-radiolabelled DNA substrates (~ 1 nM) were incubated with EXD2 WT (5.3 nM) or EXD2 (D108A E110) MUT (5.3 nM) for the indicated amounts of time at 37℃. Samples were resolved in TBE-urea polyacrylamide gel and visualised by phosphor imaging. The reaction buffer contained 5 mM $MnCl_2$. Representative image of one of two/three repeats.

5.6.3 Biotin blocks - overview

Biotin is a small and stable molecule, which can be introduced at the end of oligonucleotide sequences. It interacts with the streptavidin protein that can bind up to four biotin molecules (Pierce Protein Methods, ThermoFisher Scientific). Incubation of biotin-labelled DNA with streptavidin is often used to mimic DNA-protein adducts *in vitro* (Deshpande et al., 2016).

5.6.4 EXD2 nuclease activity on DNA with biotin-blocked ends

To test EXD2's activity on protein-blocked DNA ends, I used 3' biotinylated oligonucleotides incubated with streptavidin. Before running the reactions, EXD2 (10xHis-TEV-hEXD2-3C-2xStrep) was cleaved with 3C protease and incubated with Strep-Tactin Sepharose. This procedure removes the Strep-tag from the protein and reduces the high binding affinity of EXD2 to streptavidin. The final concentration of the cleaved protein in the nuclease activity assays in this subchapter should be around 24 nM. However, it is difficult to assess the exact concentration, due to some loss at the Strep-Tactin resin after the Strep-Tag cleavage.

I performed a nuclease activity assay using the 3' biotin-streptavidin blocked ssDNA substrate (Figure 71). As expected, the biotin-streptavidin complex completely inhibited EXD2's exonuclease activity. Lack of any digestion products indicates that EXD2 exhibits no endonuclease activity on this type of substrate. To test if any potential endonuclease activity could be stimulated by additional factors, I supplemented the reactions with Mn^{2+} alone, with both Mn^{2+} and ATP, and with all Mn^{2+}, ATP and Mg^{2+} (Figure 71). None of these conditions resulted in a digestion pattern suggesting an endonuclease activity. Addition of ATP and Mg^{2+} decreased EXD2's exonuclease activity on non-blocked ssDNA substrates.

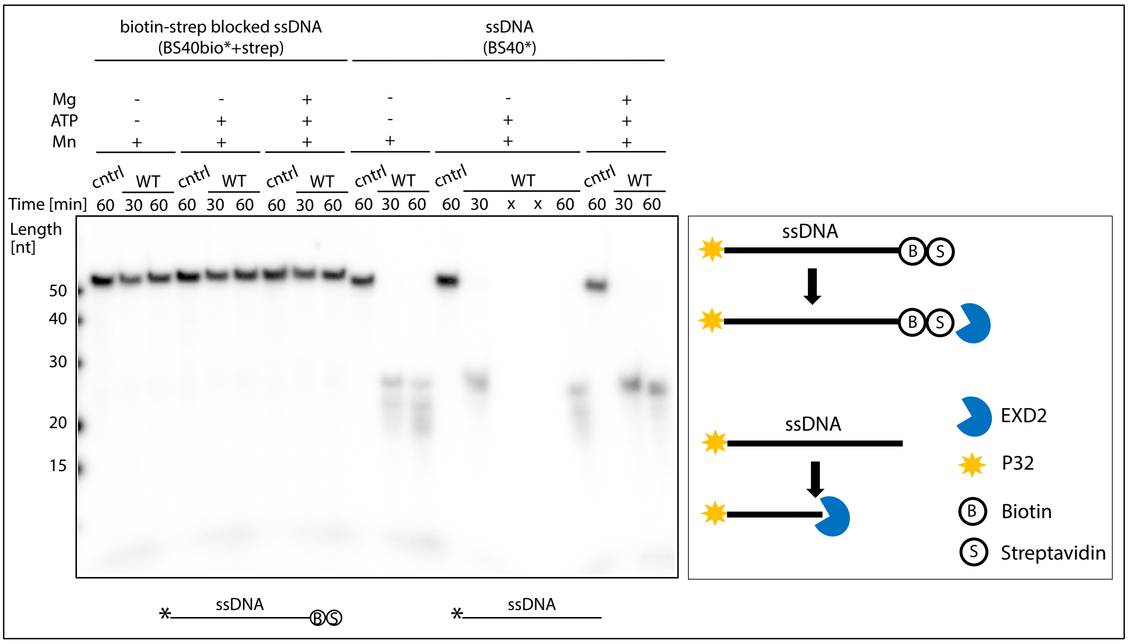


Figure 71. Addition of a biotin-streptavidin complex blocks EXD2's 3'-5' exonucleolytic activity. EXD2 exhibits no endonuclease activity on ssDNA with a biotin-streptavidin blocked end. 5'-radiolabelled DNA substrates (~ 1 nM) were incubated with EXD2 WT (after Strep-Tag cleavage) for the indicated amounts of time at 37℃. To block DNA ends, a 10 x molar excess of streptavidin was added to biotinylated DNA. Samples were resolved in TBE-urea polyacrylamide gel and visualised by phosphor imaging. The reaction buffer was supplemented with 5 mM Mn$^{2+,}$ 2 mM Mg^{2+} or 1 mM ATP, as indicated. Representative image of one of three repeats.

Next, I tested EXD2's nuclease activity on dsDNA with biotin-streptavidin blocks introduced on the 3' end of either one or both strands (Figure 72). When both DNA strands were blocked, there was no exonucleolytic activity and no clear endonuclease activity. However, the amount of the substrate decreased, and a weak signal appeared at the bottom of the gel. The result is reproducible, and it may suggest that EXD2 has a very weak endonucleolytic activity when the 3' ends of both DNA strands are blocked. When only the unlabelled strand was biotin-streptavidin blocked on the 3' end, the substrate was digested in a similar way to the PS blocked substrate from (Figure 65). The main digestion product was ~15 nt long. Addition of ATP and Mg^{2+}, which inhibited resection, did not result in the formation of middle-length products (Figure 72). If the digestion pattern we see on blocked substrates is the result of fast digestion kinetics, then it would be likely that inhibition of resection by ATP and Mg^{2+} should lead to formation of additional, middle-length products. At this point it is difficult to conclude how such a digestion pattern is generated.

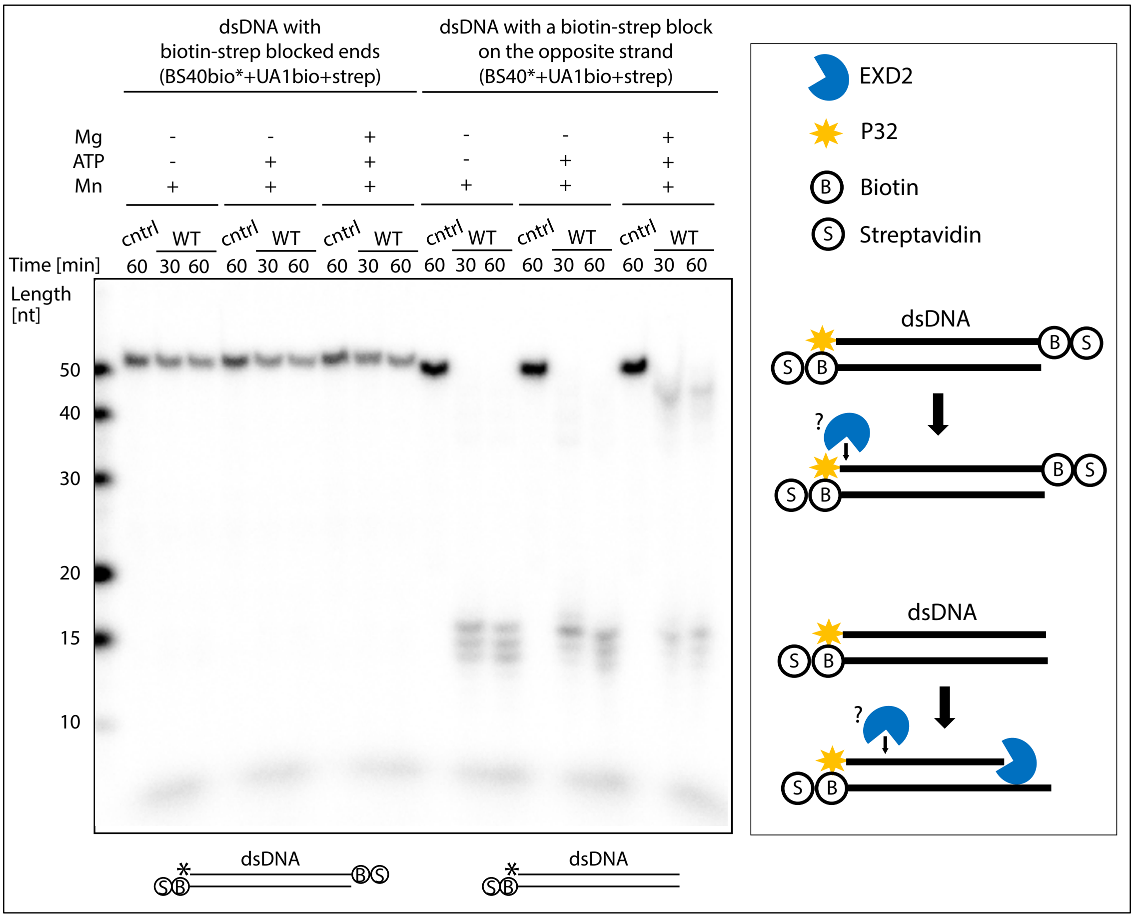


Figure 72. Blocking both strands of dsDNA with biotin-streptavidin complexes inhibits EXD2's exonucleolytic activity. It is possible that EXD2 exhibits some weak endonuclease activity but there is no clear evidence. 5'-radiolabelled DNA substrates (~ 1 nM) were incubated with EXD2 WT (after Strep-Tag cleavage) for the indicated amounts of time at 37°C. To block DNA ends, a 10 x molar excess of streptavidin was added to biotinylated DNA. Samples were resolved in TBE-urea polyacrylamide gel and visualised by phosphor imaging. The reaction buffer was supplemented with 5 mM Mn^{2+}, 2 mM Mg^{2+} or 1 mM ATP, as indicated. Representative image of one of three repeats.

To check how biotin on its own, without addition of streptavidin, affects EXD2's digestion of DNA substrates, I used a 3' biotinylated ssDNA and compared it with a non-biotinylated strand (Figure 73). I tested the activity of EXD2 WT, EXD2 exonuclease mutant (D108A E110A) and EXD2 (H442A, H463A) mutant (designed to disrupt potential endonuclease activity). All the proteins (10xHis-TEV-hEXD2-3C-2xStrep) were cleaved with 3C protease and incubated with Strep-Tactin Sepharose before the reaction. All the reactions were carried for 60 min and performed using different range of protein concentrations with two concentrations exceeding amount of the proteins typically used for these reactions. Even though the reactions were carried out with high amounts of these proteins, there was no endonucleolytic activity on ssDNA with a 3' biotinylated end. The biotin on its own, without streptavidin addition, completely blocked EXD2's exonuclease activity. The nonbiotinylated substrate was efficiently digested by WT EXD2. The EXD2 endonucleolytic mutant had some exonucleolytic activity, in line with the results showed in (5.5.3). The presence of an excess of EXD2 exonucleolytic mutant in this reaction, resulted in the generation of signal at the bottom of the gel, which may indicate some endonucleolytic activity near the 5' terminus.

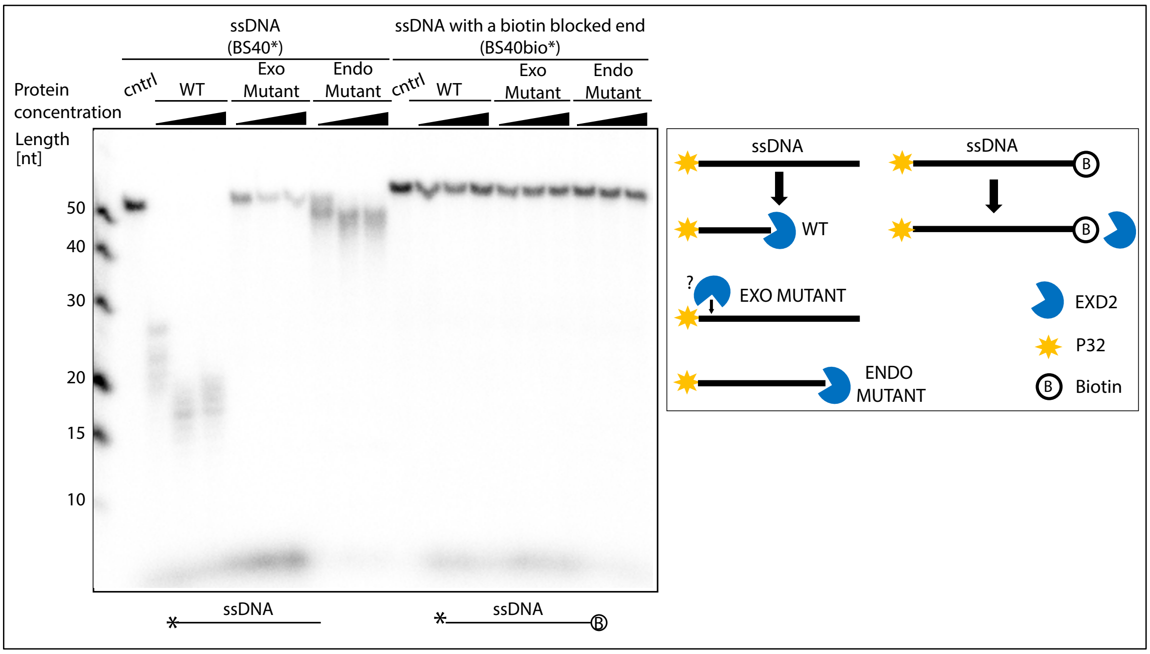


Figure 73. Biotin on the 3' end of a ssDNA substrate blocks EXD2's nuclease activity. Non-blocked ssDNA may be cut endonucleolytically by an excess of EXD2 exonuclease mutant (D108A E110A). 5'-radiolabelled DNA substrates (~ 1 nM) were incubated with EXD2 WT, EXD2 (D108A E110) MUT or EXD2 (H442A, H463A) MUT (proteins after Strep-Tag cleavage) for 1 h at 37℃. Concentrations of the proteins were ~ 12; 24; 30 nM. However, it is difficult to assess their exact concentration due to some loss at the Strep-Tacin resin after the Strep-Tag cleavage. Samples were resolved in TBE-urea polyacrylamide gel and visualised by phosphor imaging. The reaction buffer was supplemented with 5 mM Mn^{2+}. Representative image of one of two/three repeats

Conclusions

The phosphorothioate modifications (Figure 64), biotin (Figure 73), and biotin-streptavidin complex (Figure 71) inhibit EXD2's exonuclease activity. The presence of a block changes the digestion pattern of the opposite strand of dsDNA. This observed change is not sequence dependent (Figure 65 and Figure 66), but depends on the position of the block (Figure 67). There is no clear conclusion how the strand opposite to the block is digested. It is possible that one EXD2 molecule gets trapped on the 3' end of the blocked strand, occluding movement of the second EXD2 molecule digesting the labelled strand (in line with Figure 70). There is also a chance that EXD2 starts processing the unblocked strand exonucleolytically and then cleaves it endonucleolytically opposite to the block, hence no middle-length products are observed (Figure 64 and Figure 72). It is also not clear if the blocked strand of dsDNA can be endonucleolytically digested, as there are some indications of EXD2's nicking activity (Figure 72 and Figure 68).

5.7 The effect of metal ions and ATP on EXD2's nuclease activity

5.7.1 Role of magnesium – overview

Role of magnesium in stability of the genome

Magnesium is a key element required for genomic stability (Hartwig, 2001). It ensures DNA replication fidelity and acts as a cofactor in DNA damage repair pathways (Hartwig, 2001). Cellular concentration of Mg is between 5 and 30 mM, while the free Mg^{2+} content is 0.4 – 0.6 mM (Hartwig, 2001). In cells most Mg^{2+} is chelated by ATP, hence the level of free Mg^{2+} changes upon hydrolysis of chelating ATP (Ohyama, 2019). The amount of free intracellular Mg^{2+} increases during apoptosis (Hartwig, 2001). Mg stabilizes DNA conformation and affects chromatin structure, stimulating formation of heterochromatin and nucleosome self-assembly (Ohyama, 2019). The amount of free intracellular Mg^{2+} increases at metaphase, concurrent with chromosome condensation (Ohyama, 2019).

Regulation of nuclease activity by different metal ions

Experiments performed on MRN showed that MRE11's exonuclease vs endonuclease activity can be regulated by the presence of other proteins (RAD50, NBS1, phosphorylated CtIP), the type of substrate used and the presence of metal ions (Pinto, Anand, & Cejka, 2018). MRE11 is manganese-dependent exonuclease, however, for MRE11's endonuclease activity the reaction needs to be supplemented with magnesium (manganese only is not sufficient) and ATP (Pinto et al., 2018).

5.7.2 EXD2 nuclease activity is regulated by the ratio between manganese and magnesium

EXD2 is manganese-dependent nuclease, showing very weak activity in the presence of Mg^{2+} alone (Figure 31). Therefore, I decided to test if a mixture of Mn^{2+} and Mg^{2+} will increase EXD2's digestion efficiency. Surprisingly, addition of 5 mM Mg^{2+} inhibited manganese-dependent resection by EXD2 (Figure 74 A). Titration of $MgCl_2$, into samples supplemented with 5 mM $MnCl_2$, shows that inhibition is Mg concentration-dependent (Figure 74 B). Here, the highest resection efficiency was observed in the presence of 5 mM Mn^{2+} alone. A mixture of (5 mM Mn^{2+} + 1 mM Mg^{2+}) and (5 mM Mn^{2+} + 2 mM Mg^{2+}) decreased EXD2's activity, whereas (5 mM Mn^{2+} + 5 mM Mg^{2+}) inhibited resection almost completely.

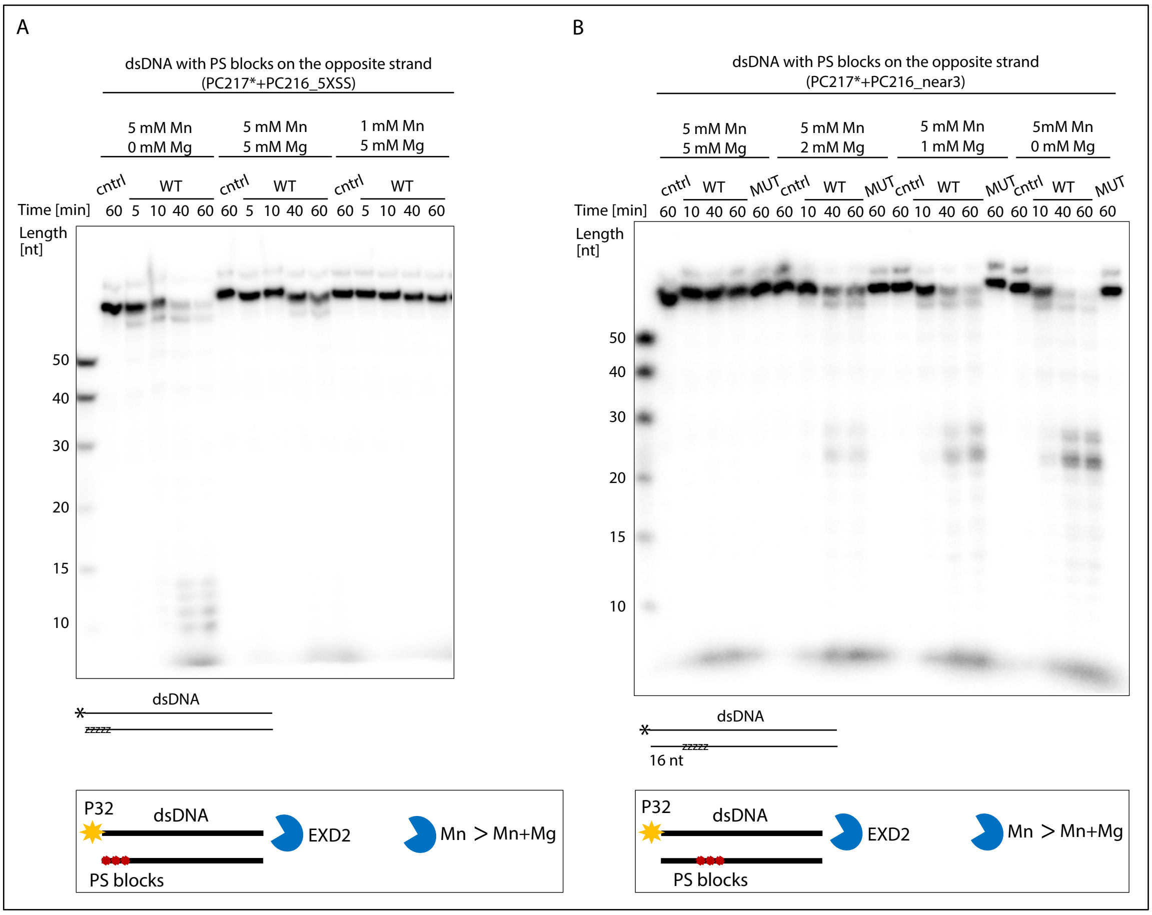


Figure 74. EXD2 WT, contrary to the (D108A E110A) mutant, digests substrates in a manganese-dependent manner. A) Addition of Mg^{2+} and Mn^{2+} together inhibits digestion by EXD2. B) Increasing concentrations of Mg^{2+} leads to inhibition of manganese-dependent resection. A)B) 5'-radiolabelled DNA substrates (~ 1 nM) were incubated with EXD2 WT (5.3 nM) or EXD2 (D108A E110) MUT (5.3 nM) for the indicated amounts of time at 37℃. Samples were resolved in TBE-urea polyacrylamide gel and visualised by phosphor imaging. The reaction buffer contained indicated amounts of Mn^{2+} and Mg^{2+}. Representative images of one of three repeats.

5.7.3 Investigating the effect of ATP on EXD2 nuclease activity

As adenosine triphosphate (ATP) alters the activity of the MRN complex (Paull & Gellert, 1999), I decided to test how it affects EXD2's nucleolytic digestion of DNA. Addition of 1 mM ATP to a reaction supplemented with 5 mM Mn^{2+}, resulted in decreased digestion of dsDNA (Figure 75). This may be caused by the ability of ATP to chelate Mn (M. M. T. Khan & Martell, 1962), leaving lower amount of free Mn^{2+} required for EXD2-dependent resection. Addition of both metal ions: Mn^{2+} and Mg^{2+}, together with ATP, massively blocked EXD2's activity on a dsDNA substrate (Figure 75). This result is in line with ATP's ability to chelate both Mg^{2+} and Mn^{2+} (M. M. T. Khan & Martell, 1962).

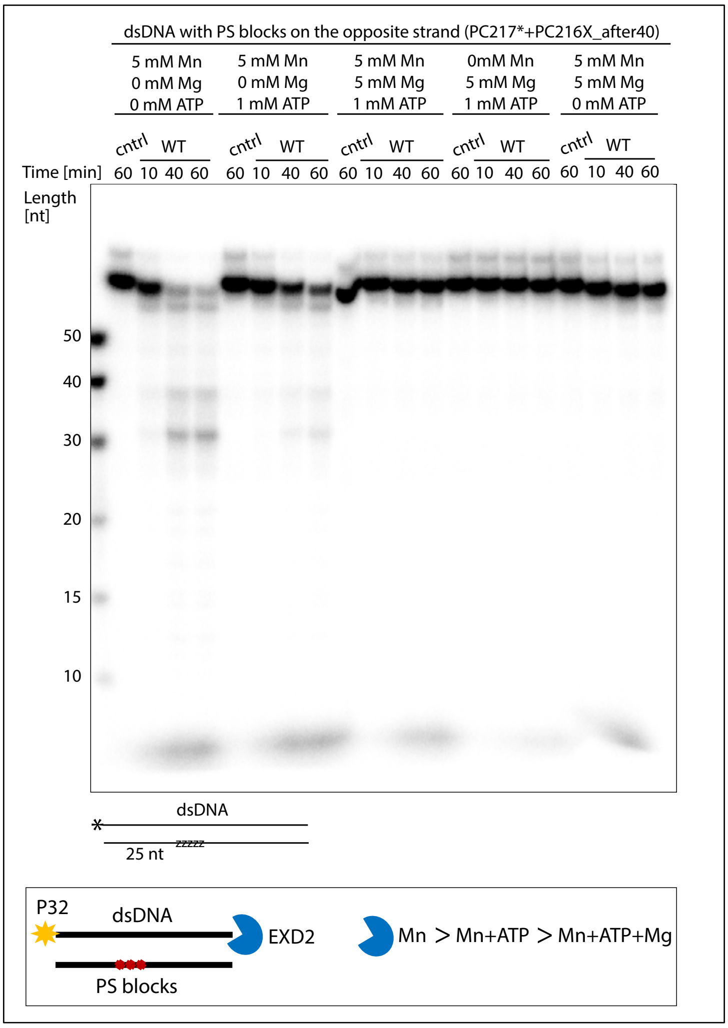


Figure 75. ATP inhibits digestion by EXD2. Supplementation with both ATP and Mg^{2+} almost completely blocks manganese-dependent resection. 5'-radiolabelled DNA substrate (~ 1 nM) was incubated with EXD2 WT (5.3 nM) or EXD2 (D108A E110) MUT (5.3 nM) for the indicated amounts of time at 37℃. Samples were resolved in TBE-urea polyacrylamide gel and visualised by phosphor imaging. The reaction buffer contained indicated amounts of Mn^{2+}, $Mg^{2+}$and ATP. Representative image of one of two repeats.

The effect of ATP, Mn^{2+} and Mg^{2+} is also examined in chapter (5.6), where different types of substrates were tested.

Conclusions

EXD2 is a manganese-dependent nuclease, exhibiting very weak activity in the presence of magnesium alone (Figure 31). Supplementation of reactions with both Mn^{2+} and Mg^{2+}, decreases resection efficiency in a Mg^{2+} dependent manner (Figure 74). Addition of ATP also inhibits the digestion, possibly due to its ability to chelate metal ions (Figure 75).

5.8 Examining EXD2's nuclease activity in the context of DNA replication

Truncated hEXD2 (K76-V564), purified from bacteria, digests fork-like structures *in vitro* (Nieminuszczy et al., 2019). To validate the activity of full-length protein expressed in insect cells, I started with a repetition of these nuclease activity assays (5.8.1) using the same fork-like substrates as (Nieminuszczy et al., 2019).

5.8.1 EXD2 activity on structures resembling DNA replication forks

To test EXD2's activity in the context of replication, I used DNA substrates resembling replication forks with one nascent strand (either leading or lagging), split arms and full forks with two nascent strands. In all substrates, only one template strand was 5'-radiolabelled. Therefore, here I analyse digestion of only this strand.

Full-length hEXD2 WT, contrary to the exonuclease-dead mutant (D108A E110A), digests fork like structures with one nascent strand (Figure 76). Digestion of the template strand is slightly faster in substrates with a nascent lagging strand than those with a nascent leading strand. The difference is very subtle, visible only in the first 5 min of the reaction, but is in line with *in vitro* nuclease activity assays performed by (Nieminuszczy et al., 2019). Faster digestion of template ends without a complementary strand, may seem surprising as EXD2 has higher digestion preferences towards dsDNA than ssDNA (5.1).

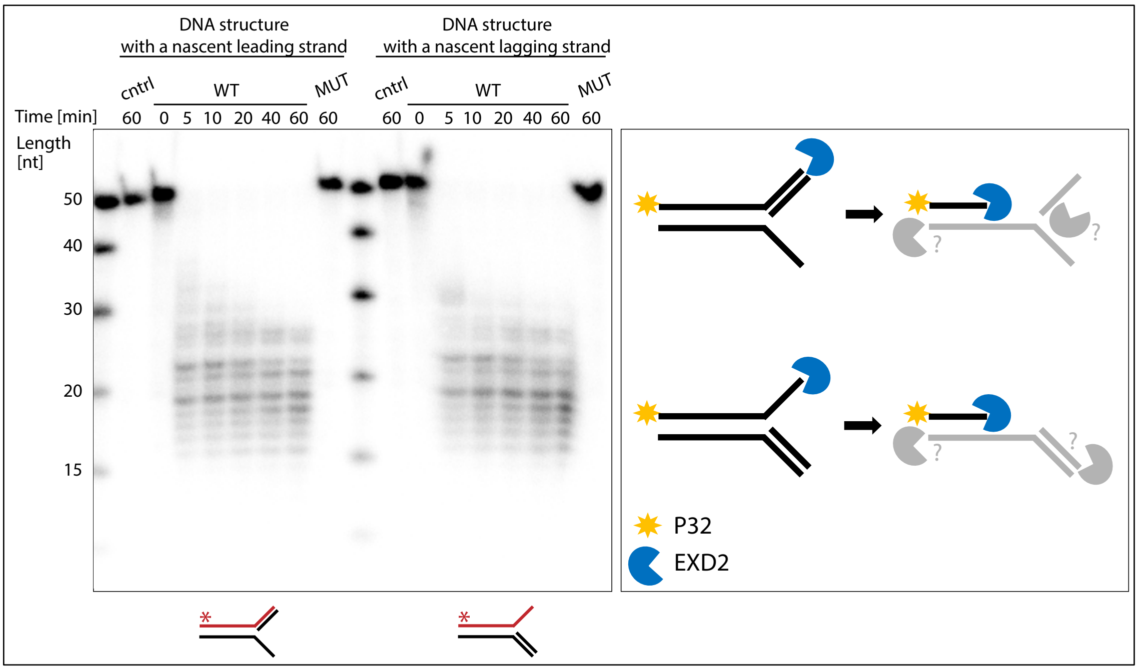


Figure 76. Purified WT EXD2, contrary to the (D108A E110A) mutant, displays 3'-5' exonuclease activity on DNA structures resembling replication forks with one nascent strand. 5'-radiolabelled

DNA substrates (~ 1 nM) were incubated with EXD2 WT (5.3 nM) or EXD2 (D108A E110) MUT (5.3 nM) for the indicated amounts of time at 37℃. Samples were resolved in TBE-urea polyacrylamide gel and visualised by phosphor imaging. The reaction buffer contained 5 mM MnCl$_2$. Representative image of one of three repeats. As digestion of only one template strand was analysed, all other (unlabelled) strands are shown in grey in the graphical representation.

DNA split arms and double stranded replication forks are also efficiently digested by WT EXD2 (Figure 77). Split arms seem to be a more accessible substrate than template strands of a ds replication fork, as they are digested faster.

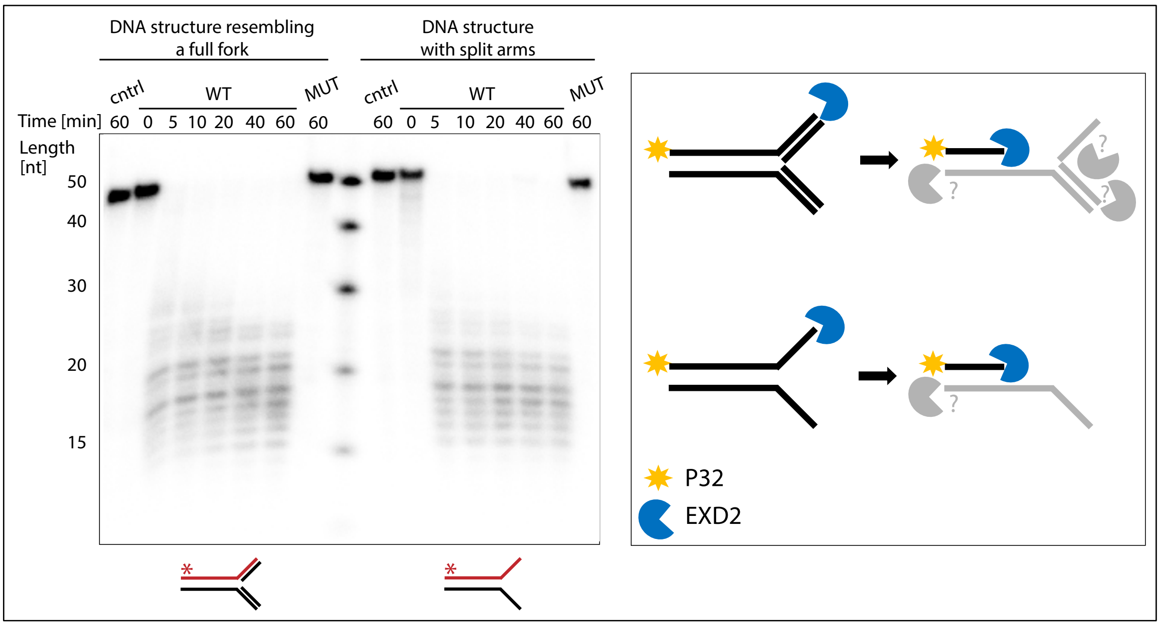


Figure 77. Purified WT EXD2, contrary to the (D108A E110A) mutant, displays 3'-5' exonuclease activity on DNA structures resembling double stranded replication forks and split arms. 5'-radiolabelled DNA substrates (~ 1 nM) were incubated with EXD2 WT (5.3 nM) or EXD2 (D108A E110) MUT (5.3 nM) for the indicated amounts of time at 37℃. Samples were resolved in TBE-urea polyacrylamide gel and visualised by phosphor imaging. The reaction buffer contained 5 mM MnCl$_2$. Representative image of one of three repeats.

Substrates from subchapter (5.8.1) were used by (Nieminuszczy et al., 2019). However, due to the lack of any blocks in the sequence, it is difficult to predict the specific sequence of events that allow for the whole structure to be digested (indicated by question marks in Figure 76 and Figure 77). To address this, I performed reaction rate assays on fork-like structures (5.8.2) and analysed how the introduction of blocks affects digestion of these fork-like structures (5.8.3).

5.8.2 The reaction rate of EXD2 on DNA structures resembling replication forks

To calculate the reaction rate of digestion I used DNA substrates resembling replication forks with nascent leading strand. The 3' end of the template leading strand was fluorescently labelled with 6-FAM, whereas the 5' end of the nascent leading strand was integrated with a quencher. The quencher masks the fluorescence of 6-FAM, unless it is released from the substrate due to digestion of the strand by EXD2 (Figure 78C). The reaction was performed with 5 mM Mn^{2+}. The fluorescence intensity, measured in relative fluorescence units (RFU), corresponds with the amount of product that is created. The speed of the reaction, also called the reaction rate, is described as the increase of the product concentration per unit of time (Figure 78A). As the amount of the substrate decreases over time, the rate of the reaction also slows down (Figure 78B). According to (Figure 78B), the speed of the reaction is the highest between 2.5 min and 5 min, with the highest points at 302 s, 230 s, 182 s and 149 s (values for four repeats respectively).

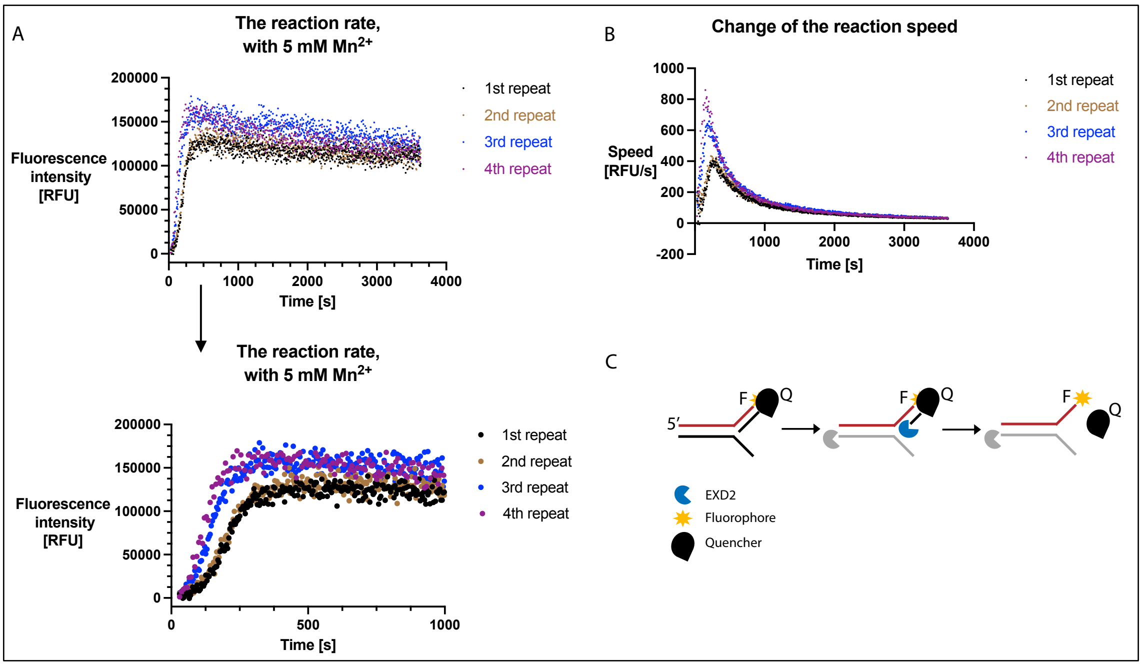


Figure 78. EXD2 WT digests DNA fork-like structure with nascent leading strand. EXD2 WT (~131 nM) was incubated with DNA (containing both 6-FAM Fluorescein and a Quencher) for 1h (3600 s) at RT, in the presence of 5 mM Mn^{2+}. The final concentration of fluorescently labelled DNA strand was 24 nM. Changes in fluorescence intensity were measured by SpectraMax iD5 plate reader. A) The reaction rate is described as increase in fluorescence intensity (RFU) per unit of time (s). The 4th repeat was performed on protein kept overnight in 4℃ after thawing, showing high EXD2's stability. B) The speed of the reaction (RFU/s) changes over time. C) Proposed mechanism of substrate digestion by EXD2.

As the RFU increases linearly at the beginning of the reaction, I used Prism 9 *simple linear regression* tool to calculate the slope. I chose time range in which plots looked most linear (first 300 s for 1st and 2nd repeat; first 250 s for 3rd and 4th repeat). Best-fit values of the slopes are shown in (Figure 79) and they describe initial velocity (Vo) of the reaction. The Vo is almost the same for 1st and 2nd repeat but it is around 1.6 - 1.7x slower than for 3rd and 4th repeat. It may be due to high sensitivity of the assay and subtle differences of the conditions, like using different protein aliquots from the same purification batch.

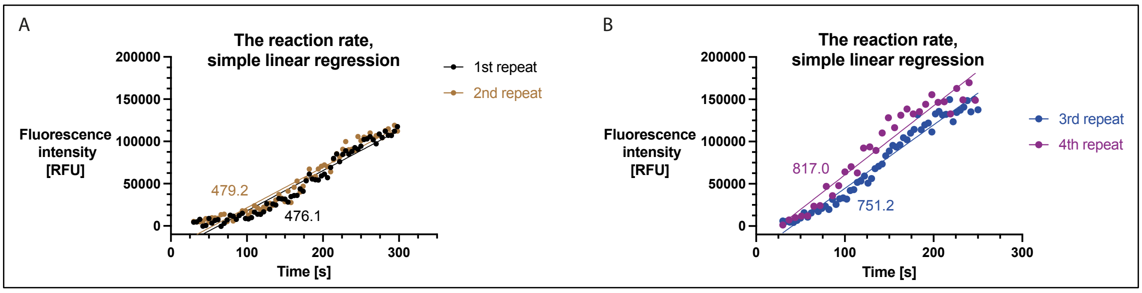


Figure 79. The reaction rate of EXD2 WT on DNA fork-like structure with nascent leading strand based on data points from (Figure 78). Plots include values of simple linear regression. Calculations were performed by the Prism 9 program. A) 1st and 2nd repeat. B) 3rd and 4th repeat.

To investigate how the reaction rate changes upon different concentrations of manganese ions, I performed reactions with the same DNA fork-like structure, but with 0 mM, 0.1 mM, 1 mM, 5 mM and 10 mM Mn^{2+} (Figure 80). The steepest increase in RFU was observed with 10 mM Mn^{2+}, indicating the highest initial velocity of EXD2 in these conditions. Only one repeat with 5 mM Mn^{2+} was performed parallel to the reaction with 10 mM Mn^{2+}, and it had nearly the same reaction rate (Figure 80 A). Decreasing the Mn^{2+} concentration to 0.1 mM delayed the digestion. A complete lack of Mn^{2+} in the reaction, totally inhibited EXD2's nuclease activity.

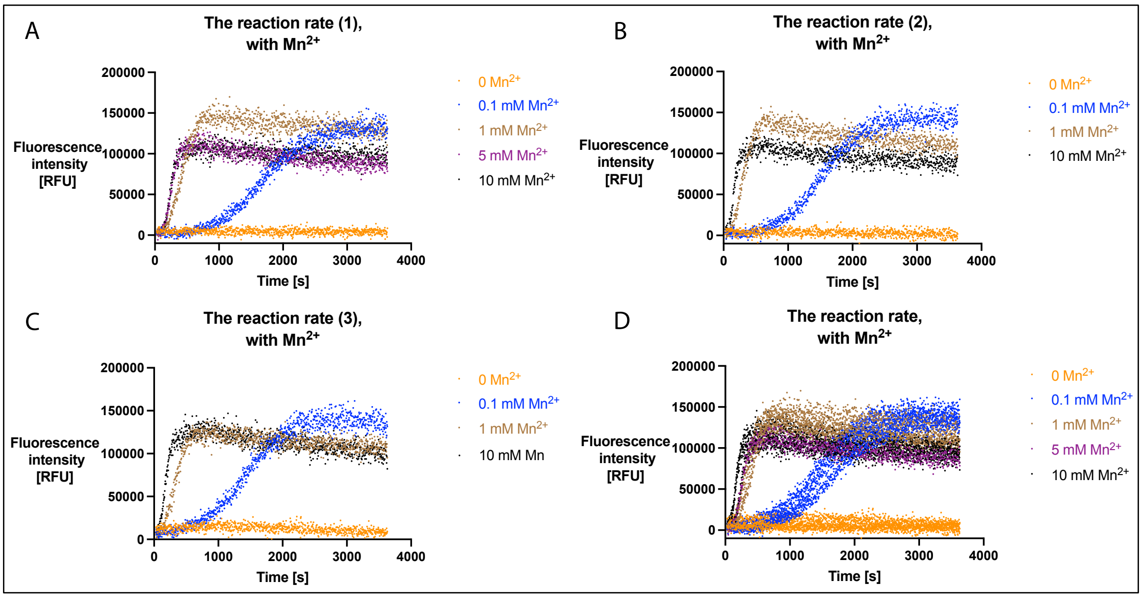


Figure 80. The reaction rate of digestion of substrates mimicking replication forks by EXD2 WT depends on Mn^{2+} concentration. Assays performed on DNA fork-like structure with nascent leading strand. EXD2 WT (~131 nM) was incubated with DNA (containing both 6-FAM Fluorescein and a Quencher) for 1h (3600 s) at RT. The final concentration of fluorescently labelled DNA strand was 24 nM. Changes in fluorescence intensity were measured by SpectraMax iD5 plate reader. The concentration of Mn^{2+} in the reactions varied as indicated. A) B) and C) show reaction rates of separate repeats. D) shows combined reaction rates of all three repeats.

To check how the reaction rate changes upon addition of magnesium, I replaced Mn^{2+} with either 1 mM or 5 mM magnesium. Nuclease activity assays on dsDNA in subchapter (5.1.2) show hardly any EXD2 activity upon Mg^{2+} addition. Here, the reaction rate assays with DNA fork-like structures, show a lack of EXD2 activity in the presence of 1 mM Mg^{2+}

(Figure 81B). Addition of 5 mM Mg^{2+} results in delayed and very inefficient digestion (Figure 81A). As the 4th repeat of 5 mM Mn^{2+} reaction from (Figure 78), was performed in parallel to one of the repeats with 5 mM Mg^{2+}, their reaction rates are plotted together (Figure 81C). Fluorescence intensity after 60 min of the reaction with 5 mM Mg^{2+} is nearly the same, as in the 2nd minute of the reaction with 5 mM Mn^{2+}. A single experiment with the EXD2 exonuclease-dead mutant (D108A E110A), shows a lack of any activity with either 5 mM Mg^{2+} or 5 mM Mn^{2+}.

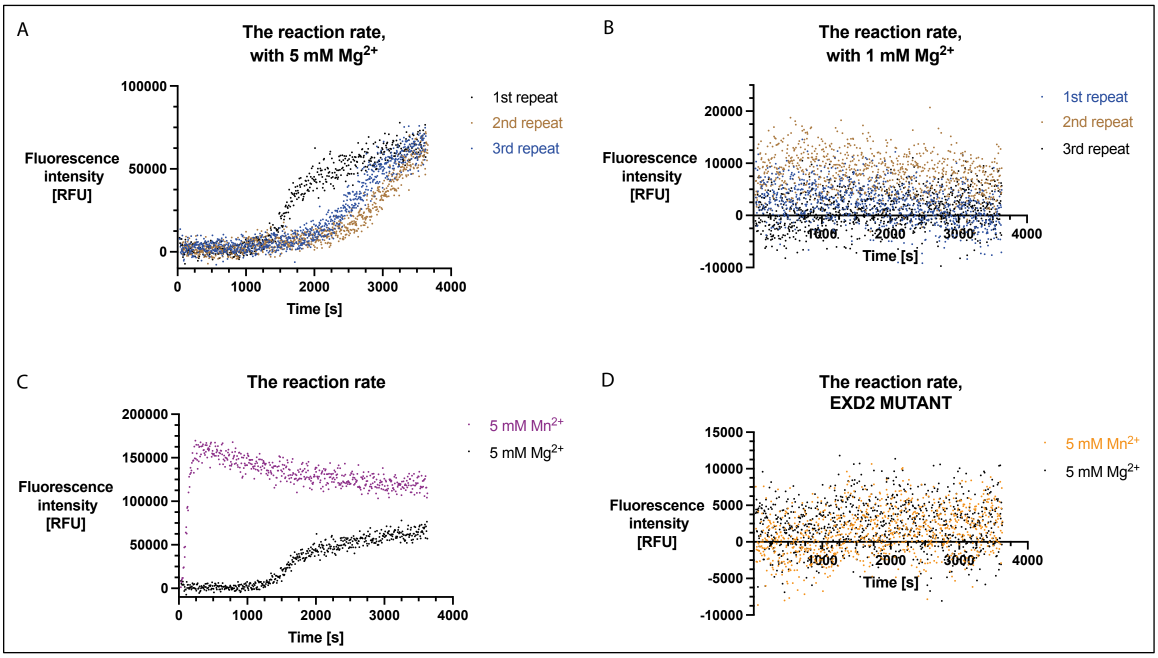


Figure 81. The reaction rate of EXD2 WT depends on the type and concentration of ions used in the reaction. Assays were performed on DNA fork-like structure with nascent leading strand. EXD2 WT or (D108A E110A) mutant (~131 nM) were incubated with DNA (containing both 6-FAM Fluorescein and a Quencher) for 1h (3600 s) at RT. The final concentration of fluorescently labelled DNA strand was 24 nM. Changes in fluorescence intensity were measured by SpectraMax iD5 plate reader. The concentrations of Mn^{2+} and Mg^{2+} in the reactions varied, as indicated. A) Reaction rate in the presence of 5 mM Mg^{2+}. B) Reaction rate in the presence of 1 mM Mg^{2+}. C) Comparison of the reaction rates with different types of ions. D) Reaction rate for EXD2 (D108A E110A) mutant.

5.8.3 Investigating the mechanism of fork-like structure digestion by EXD2

To understand how DNA replication forks with nascent leading strand are digested, I decided to introduce phosphorothioate (PS) modifications at different positions of the fork sequence. To mimic the continuity of the bottom template strand, I placed PS functional groups at the 3’ end (Figure 82). This should block EXD2-dependent 3’-5’ digestion from this end. Additionally, I introduced PS blocks at the junction or on either double stranded or single stranded part of the fork. I performed reaction rate assays on all substrates, with and without the blocks, in the presence of 5 mM Mn^{2+}. Plots of fluorescence intensity versus time, show nearly the same reaction rate for all used substrates (Figure 82). It suggests that the presence of a block on the 3’ end of the bottom template strand, as well as the presence of PS blocks inside the sequence, does not affect digestion of the nascent strand by EXD2. The nascent strand is accessible for EXD2 when the bottom template strand is present.

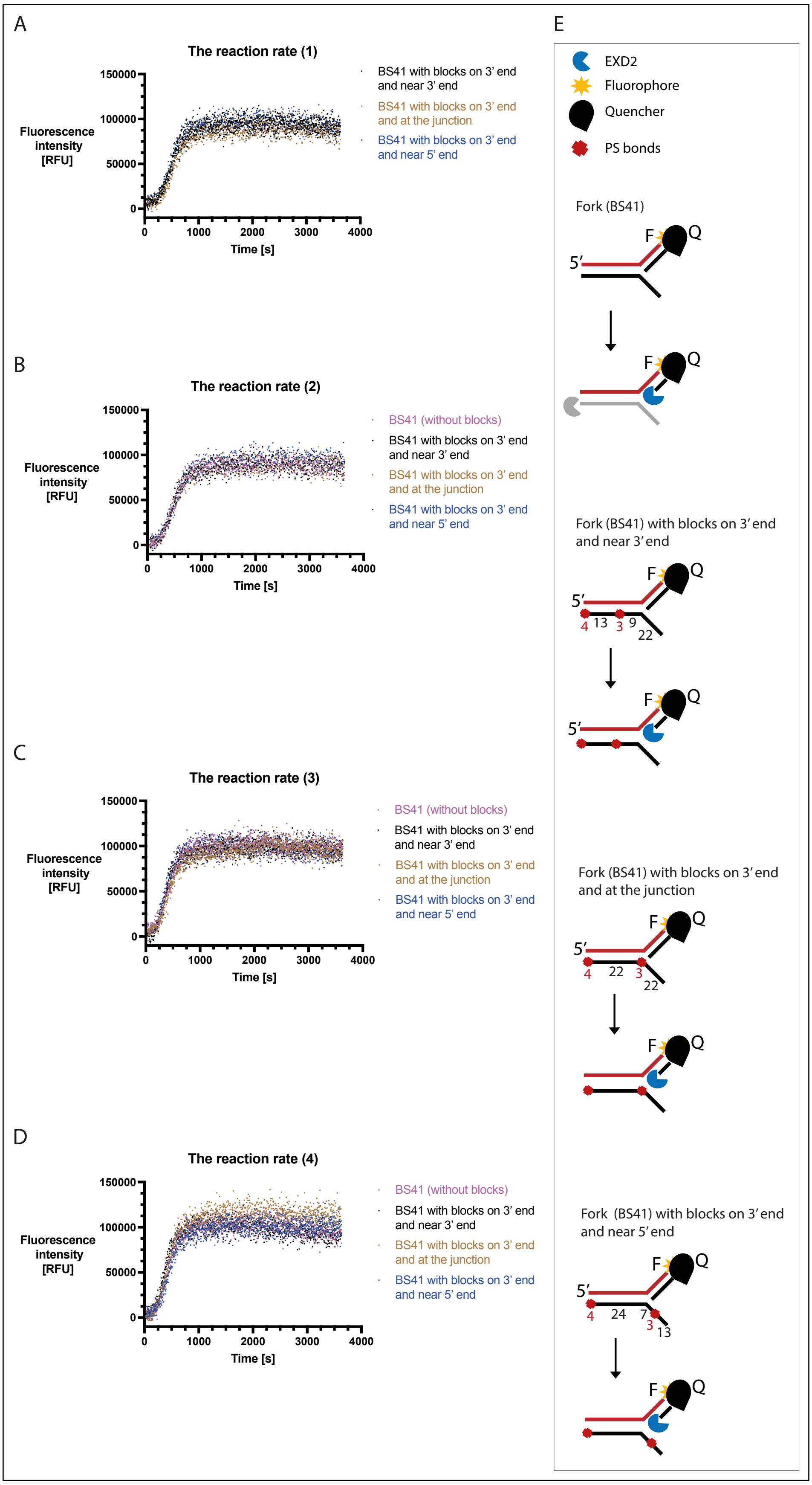


Figure 82. The reaction rate of EXD2 WT's digestion of a fork-like substrate on the nascent leading strand is not affected by a 3' block on a bottom template strand. EXD2 WT (~131 nM) was incubated with DNA substrates (containing both 6-FAM Fluorescein and a Quencher) for 1h (3600 s) at RT, in the presence of 5 mM Mn^{2+}. The final concentration of fluorescently labelled DNA strand was 24 nM.

Changes in fluorescence intensity were measured by SpectraMax iD5 plate reader. A) B) C) D) show reaction rates of separate repeats on different types of substrates, with different positions of PS modifications. E) Model of substrates used in the reactions. Black values represent number of nucleotides between the blocks, red values represent number of the PS functional groups.

To check how the presence of blocks affect the reaction rate when the 3’ end of the bottom template strand is free, I introduced PS modifications close to the junction or at the 5’ end (Figure 83). Introduction of these PS functional groups inside the bottom template strand does not affect the digestion of the nascent strand even when 3’ end of the template strand is accessible for EXD2 (Figure 83). Combined results from (Figure 82) and (Figure 83) suggest that the nascent strand is digested first, being the preferable substrate in this structure.

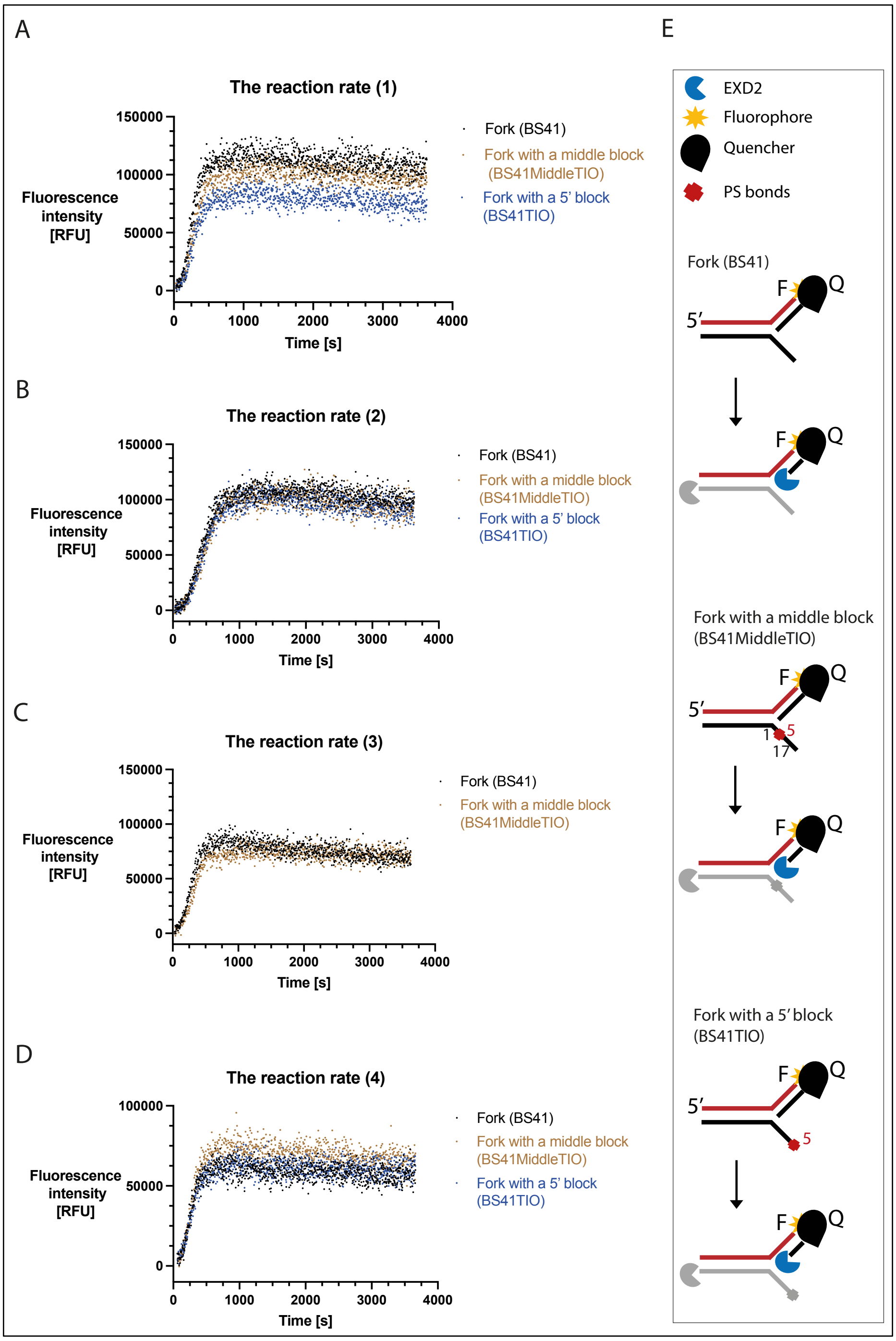


Figure 83. Reaction rate of EXD2 WT's digestion of a fork-like substrate on the nascent leading strand is not affected by the position of PS blocks on the bottom template strand. EXD2 WT (~131 nM) was incubated with DNA (containing both 6-FAM Fluorescein and a Quencher) for 1h (3600 s) at RT, in the presence of 5 mM Mn^{2+}. The final concentration of fluorescently labelled DNA strand was 24 nM. Changes in fluorescence intensity were measured by SpectraMax iD5 plate reader. A) B) C) D) show reaction rates of separate repeats on different types of substrates, with different positions of PS modifications. E) Model of substrates used in the reactions. Black values represent the number of nucleotides between the blocks, red values represent the number of the PS functional groups. Only one negative control was used for different substrates shown in this figure.

To confirm that the fluorescence observed during these reactions is not caused by the digestion of the fluorophore (6-FAM) from the 3' end, I designed a substrate with PS functional groups adjacent to the 6-FAM label (Figure 84). The reaction rate plots show no difference between digestion of this substrate and the substrate with no PS on the labelled strand. This suggests that the template strand with a 3’ 6-FAM label is not digested by EXD2. Hence, the observed fluorescence is most likely caused by the digestion of a nascent strand with a quencher.

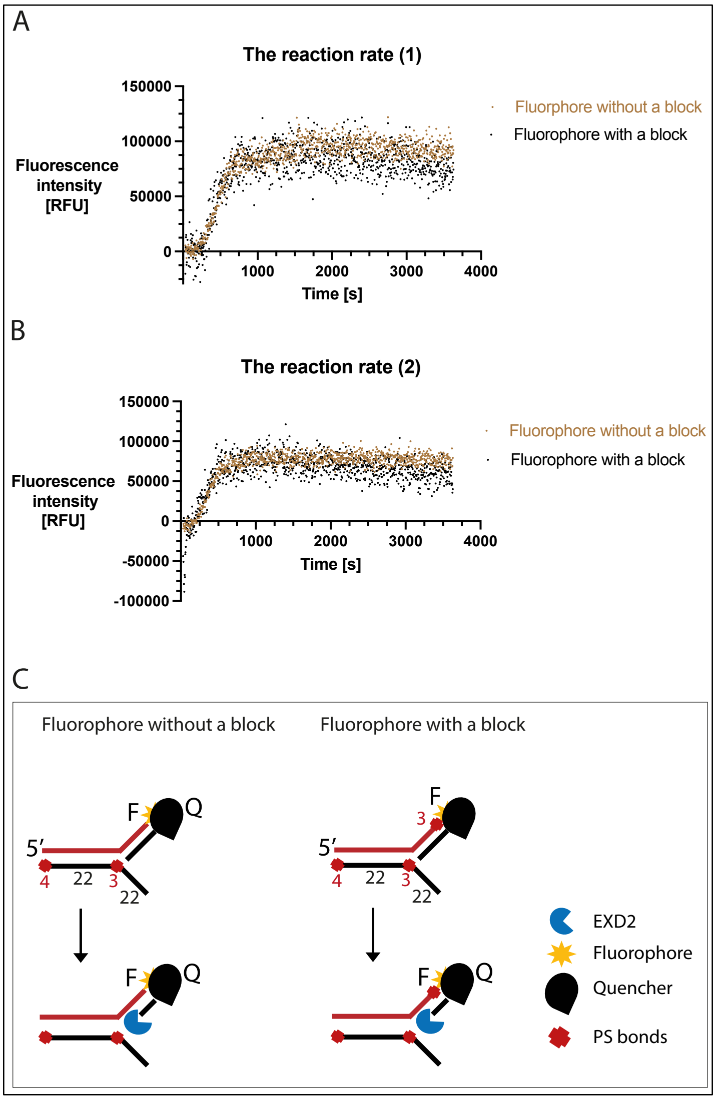


Figure 84. A fluorophore (6-FAM) located on the 3’ end of the template strand is not exonucleolytically digested by WT EXD2. EXD2 WT (~131 nM) was incubated with DNA (containing both 6-FAM Fluorescein and a Quencher) for 1h (3600 s) at RT, in the presence of 5 mM Mn^{2+}. The final concentration of fluorescently labelled DNA strand was 24 nM. Changes in fluorescence intensity were measured by SpectraMax iD5 plate reader. A) and B) show reaction rates of separate repeats on substrates with and without PS blocks adjacent to the 6-FAM label. C) Model of substrates used in these reactions. Black values represent numbers of nucleotides between the blocks, red values represent numbers of PS functional groups.

5.8.4 Estimation of the kinetics of EXD2-dependent nucleolytic processing of DNA.

The canonical way to understand kinetics of an enzyme is to use the Michaelis-Menten (MM) equation (Choi, Rempala, & Kim, 2017). This describes the dependence of a reaction rate on the substrate concentration. K_M, which is Michelin-Menten constant, equals concentration of the substrate, when the reaction rate is half of the maximum velocity (V_{max}).

To calculate EXD2 WT kinetics parameters I ran reactions for 30 min, at different concentrations of a substrate (a fork structure with blocks on the 3' end and near the 3' end, as shown in Figure 82). The 3' end of the template leading strand was labelled with 6-FAM, whereas the 5' end of the nascent leading strand was labelled with a quencher. To visualise the reaction rate, I plotted the increase of the fluorescence over time for each substrate concentration separately (Figure 85 A). Velocity (speed), described as increase of the fluorescence in time (RFU/s), was also plotted versus time (Figure 85 B). According to this representation, the speed of digestion of 48 nM substrate was the highest at ~ 471 s (~ 8 min). Therefore, I averaged 3 values of reaction velocity taken around this time (462 s, 471 s, 480 s) for each amount of the substrate, and I plotted them versus substrate concentrations (9.6; 12; 16; 24; 32; 48 nM), (Figure 85 D). Using Prism 9 software, I determined the K_M and V_{max} for these reactions. The program generated nonlinear fit with best fit values of V_{max} = 748.2 [RFU/s] and K_M = 60.29 [nM]. V_{max} is the maximum velocity, which according to the software is extrapolated to very high substrate concentrations and is almost always higher than any velocity measured in the experiment. As the highest digestion velocity of 32 nM substrate was obtained at ~ 417 s (~ 7 min), I decided to perform the same analysis for the values taken at times (408 s, 417 s, 426 s), (Figure 85 C). At ~ 7 min the best fit values were V_{max} = 577.4 [RFU/s] and K_M = 37.77 [nM], which differ a lot comparing to the values obtained at ~ 8 min. The values obtained at the end of the reaction at ~ 30 min (1794 s, 1803 s, 1812 s) were: V_{max} = 538.6 [RFU/s] and K_M = 179.9 [nM], (Figure 85 E). Reaction velocity plots for the three analysed times (7 min, 8 min, 30 min) were combined on (Figure 85 D).

The huge variety between V_{max} values is caused by insufficient amount of analysed data points. In the MM equation, reaction velocity vs substrate concentration plot should have a rectangular hyperbola shape (Walsh, 2012). Based on the shape of the plots from (Figure 85 F), all analysed concentrations (9.6 – 48 nM) are in the beginning range of EXD2-dependent reaction velocity. To increase the accuracy of this analysis, higher concentrations of substrates should be tested. Unfortunately, as the highest possible amount of the fluorescently labelled substrate was used during the analysis (48 nM), it was not possible to increase the amount of the tested substrate under these conditions (50 μl of total reaction volume). Hence, to estimate V_{max} and K_m values of EXD2, another protocol should be optimized. Duo to the method limitations, fluorescence analysis is not recommended (the highest RFU readout I managed to obtain before saturation error was ~ 435 000).

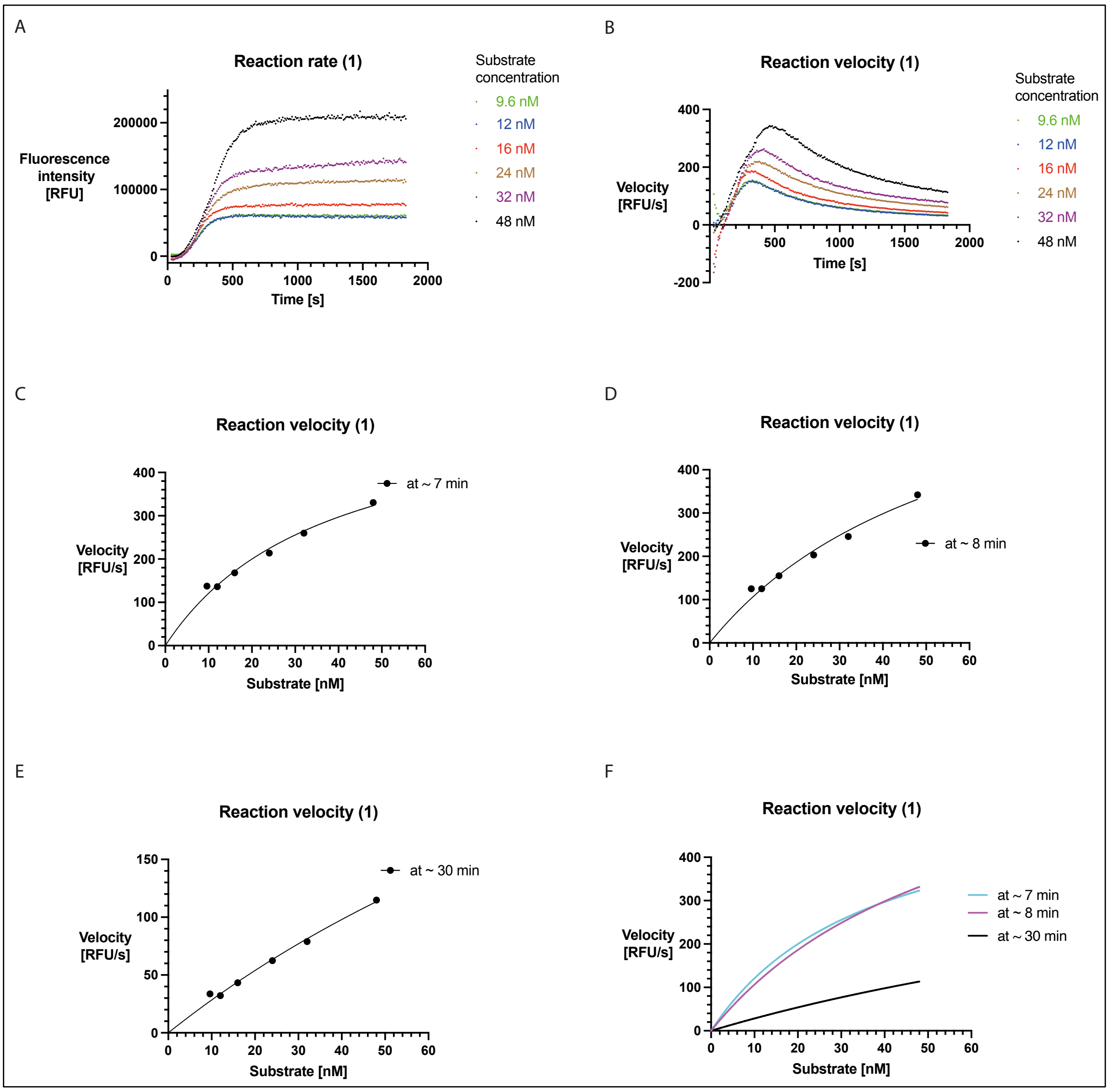


Figure 85. Kinetics parameters of fork-like substrate digestion by EXD2 WT (~131 nM). Reactions were performed with different concentrations of DNA substrate (containing both 6-FAM Fluorescein and a Quencher), for 30 min at RT. Changes in fluorescence intensity were measured by SpectraMax iD5 plate reader. A) Reaction rates (RFU vs time) of EXD2 on different amounts of substrate. B) Reaction velocity (RFU/s) vs time on different amounts of substrate. C) Reaction velocity vs substrate concentration at 7 min. D) Reaction velocity vs substrate concentration at 8 min. E) Reaction velocity vs substrate concentration at 30 min. E) Combined reaction velocity plots, used for generation of K_M and V_{max} values. Number (1) in all shown plots indicates that all data were generated from the same analysis.

5.9 Investigating the effect of RPA on EXD2 activity

5.9.1 RPA binding to DNA substrates

The presence of RPA impacts activity of many nucleases (discussed in more detail in 1.4.4). However, it has not been tested if the presence of RPA may directly affect EXD2's ability to carry out exonucleolytic digestion of DNA substrates. To test this, I performed an *in vitro* analysis, using RPA purified from *E. coli* by Dr Joanna Krwawicz. The protein was kept in - 80°C in a storage buffer (50 mM Tris-HCl, pH 8, 100 mM NaCl, 1 mM DTT, 5 % glycerol) and its concentration (7.9 μM) was assessed spectrophotometrically (M_w = 111000 Da, E = 88085 $M^{-1}cm^{-1}$, A_{280} = 0.7).

To validate the DNA binding activity of this purified RPA, I performed electrophoretic mobility shift assays (EMSA). Here, I used 3 different DNA structures: split arms, dsDNA with a 3' overhang and a fork-like structure with one nascent strand and PS blocks (Figure 86). The final RPA concentration in these reactions were between 0 and 390 nM, whereas final concentrations of labelled DNA strands were ~ 1 nM. RPA bound to all tested DNA substrates in the reaction buffer used for nuclease activity assays (25 mM Tris-HCl pH 7.5, 0.25 mg/ml BSA, 5 mM $MnCl_2$ and 0.5 mM DTT), (Figure 86). As the strongest DNA mobility shift was observed at 390 nM RPA (Figure 86 A), I decided to use this concentration in the nuclease activity assays with EXD2.

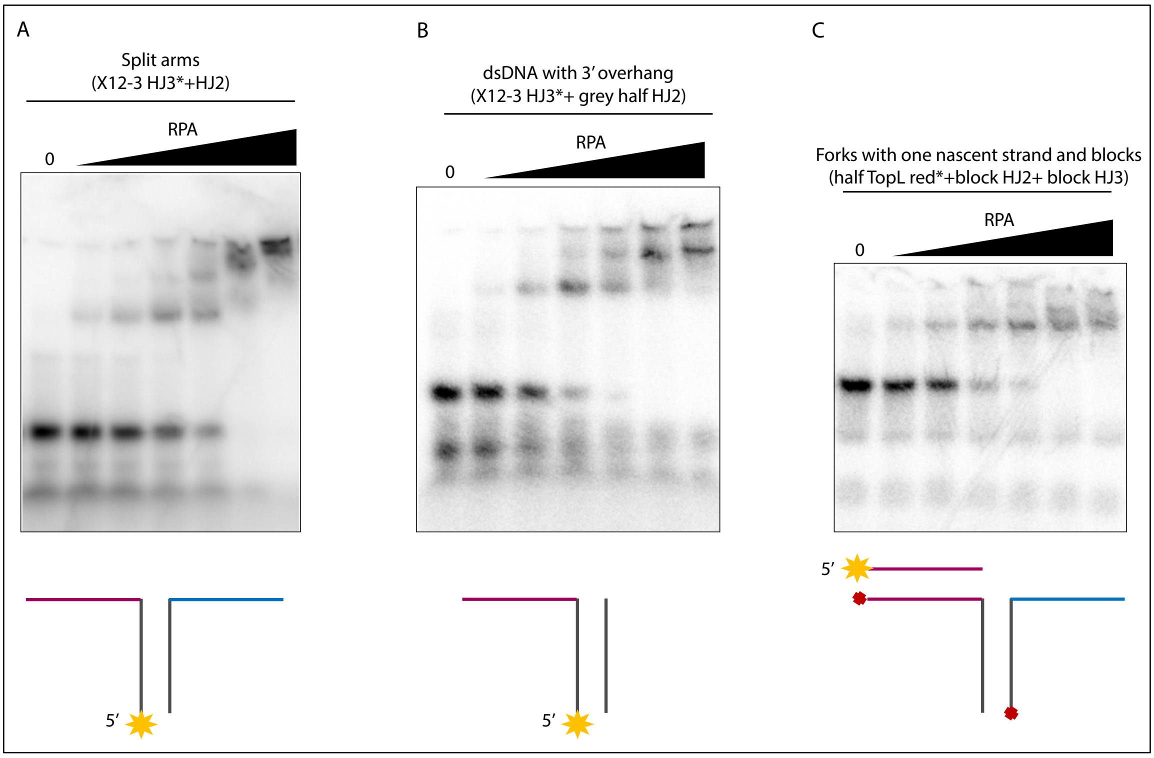


Figure 86. EMSA assay with purified RPA and different types of DNA substrates (length of each ssDNA strand is 45 nt). Star indicates labelled strand; red circle indicates position of the PS blocks. RPA (0, 12; 24; 49; 98; 195; 390 nM) was incubated with 5'-radiolabelled DNA substrates (~ 1 nM) for ~ 30 min at 4°C. Samples were resolved in native 6 % acrylamide gels and visualised by phosphor imaging. A) Representative image of one of three repeats. B) and C) Representative images of one of two repeats for each.

5.9.2 Examining EXD2 nuclease activity in the presence of RPA

To test how RPA affects EXD2 activity *in vitro*, I performed nuclease activity assays in the presence and absence of RPA. As RPA is a ssDNA binding complex with a high affinity towards ssDNA stretches of 28 nt or above (Yates et al., 2018), I used DNA split arms with 45 nt long ssDNA as a substrate (total length of the substrate was 93 nt). To mimic the continuity of dsDNA, I also tested a second substrate where PS modifications were introduced on the double stranded 3' end. In the absence of RPA, EXD2 efficiently digested both types of DNA split arms (Figure 87). The digestion pattern differed between the substrates, due to the presence of the PS blocks on the opposite strand to the labelled one. Incubation of substrates with RPA for 20 min prior to the EXD2 addition, resulted in a lack of EXD2 activity. RPA coated ssDNA ends were almost fully protected from nucleolytic degradation. Only at the highest EXD2 concentration there was an indication of digestion of few first nucleotides. As a control, samples without RPA were supplemented with a buffer of the same composition as the one used for the RPA storage. To test if EXD2 can compete with RPA for access to DNA, I also performed reactions where DNA was added to the solution where both proteins were already present (Figure 87). As the DNA digestion was blocked, it indicates that RPA has higher affinity to ssDNA ends than EXD2.

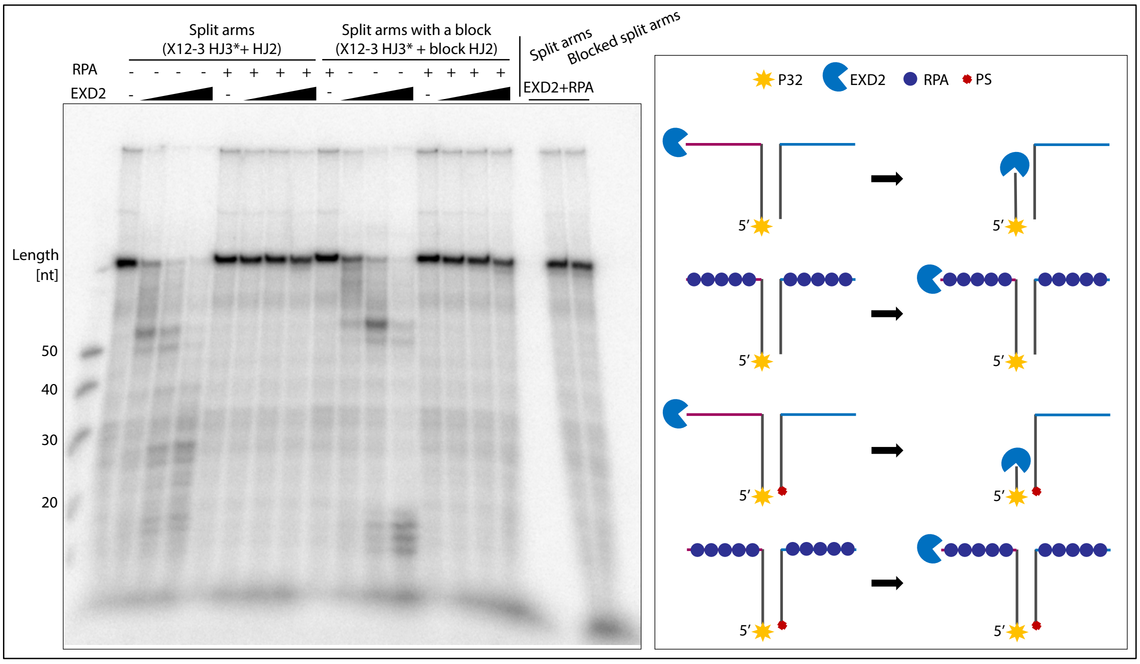


Figure 87. RPA blocks the activity of EXD2 WT on DNA substrates with split arms. 5'-radiolabelled DNA substrates (~ 1 nM) were incubated with RPA (390 nM) for 20 min. Next, EXD2 WT (0; 2.6; 5.3; 10.6 nM) was added for the reaction (1h; 37℃). Samples were resolved in TBE-urea polyacrylamide gel and visualised by phosphor imaging. The reaction buffer contained 5 mM $MnCl_2$. Label (EXD2+RPA) indicates that in last two reactions EXD2 (10.6 nM) and RPA were added to the substrate simultaneously. Representative image of one of three repeats.

As RPA coats 3’ overhangs generated during resection of DSBs in HR, I tested how it affects EXD2 activity on dsDNA with a 3’ overhang (45 nt from the total length of 93 nt). Similarly, the presence of RPA blocked EXD2 digestion on dsDNA with a 3’ overhang (Figure 88). Here, the tested EXD2 concentrations were higher (0; 5.3; 10.6; 21.2 nM), than with the split arms substrates (0; 2.6; 5.3; 10.6 nM). At the highest concentrations, EXD2 was able to perform digestion of few first nucleotides of RPA coated DNA (Figure 88). The same result was observed for a reaction where both RPA and EXD2 (21.2 nM) were added to the substrate simultaneously (last line from the Figure 88).

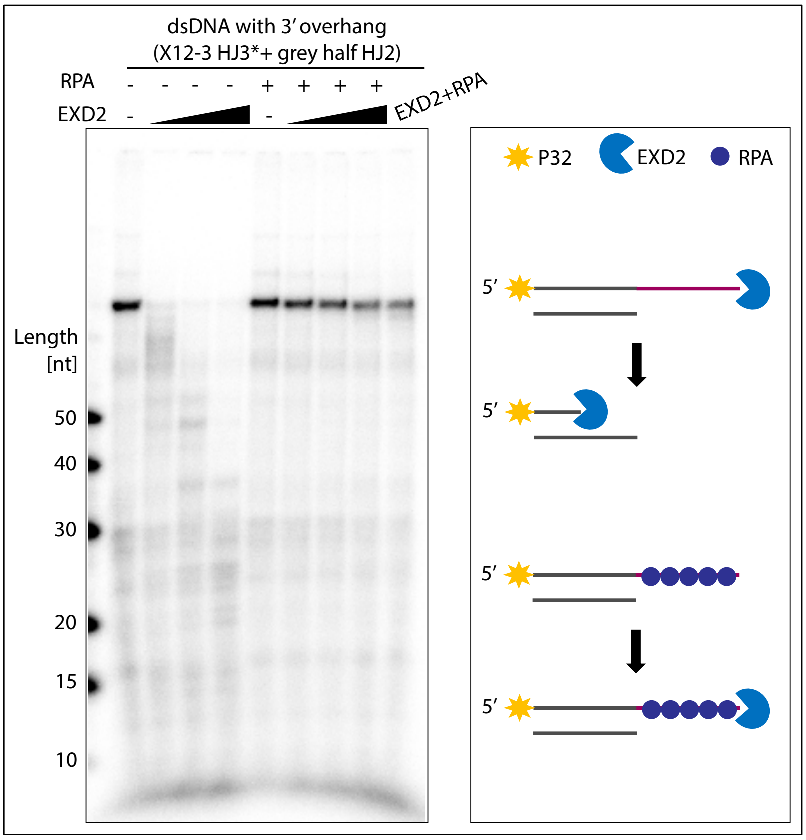


Figure 88. RPA blocks EXD2 WT activity on dsDNA with a 3’ overhang. 5’-radiolabelled DNA substrate (~ 1 nM) was incubated with RPA (390 nM) for 20 min. Next, EXD2 WT (0; 5.3; 10.6; 21.2 nM) was added for the reaction (1h; 37℃). Samples were resolved in TBE-urea polyacrylamide gel and visualised by phosphor imaging. The reaction buffer contained 5 mM $MnCl_2$. Label (EXD2+RPA) indicates that in the last reaction EXD2 (21.2 nM) and RPA were added to the substrate simultaneously. Representative image of one of three repeats.

RPA binds ssDNA over 1000 times more efficiently than dsDNA (T. Liu & Huang, 2016). To examine if RPA may affect EXD2 activity on dsDNA ends, I used a substrate resembling a replication fork with a one nascent strand. To mimic continuity of the strands, PS modifications were introduced on the 3’ ends of the template strands. After preincubation of DNA with RPA for 20 min, I started the reaction by EXD2 addition. Interestingly, RPA decreased EXD2 activity on the annealed strand, but did not block it completely (Figure 89). In the presence of RPA, the digestion pattern resembled a smear suggesting resection of the substrate in a slow gradual manner. In the absence of RPA, most of the DNA was digested to strands shorter than 15 nt, without the products of intermediate length.

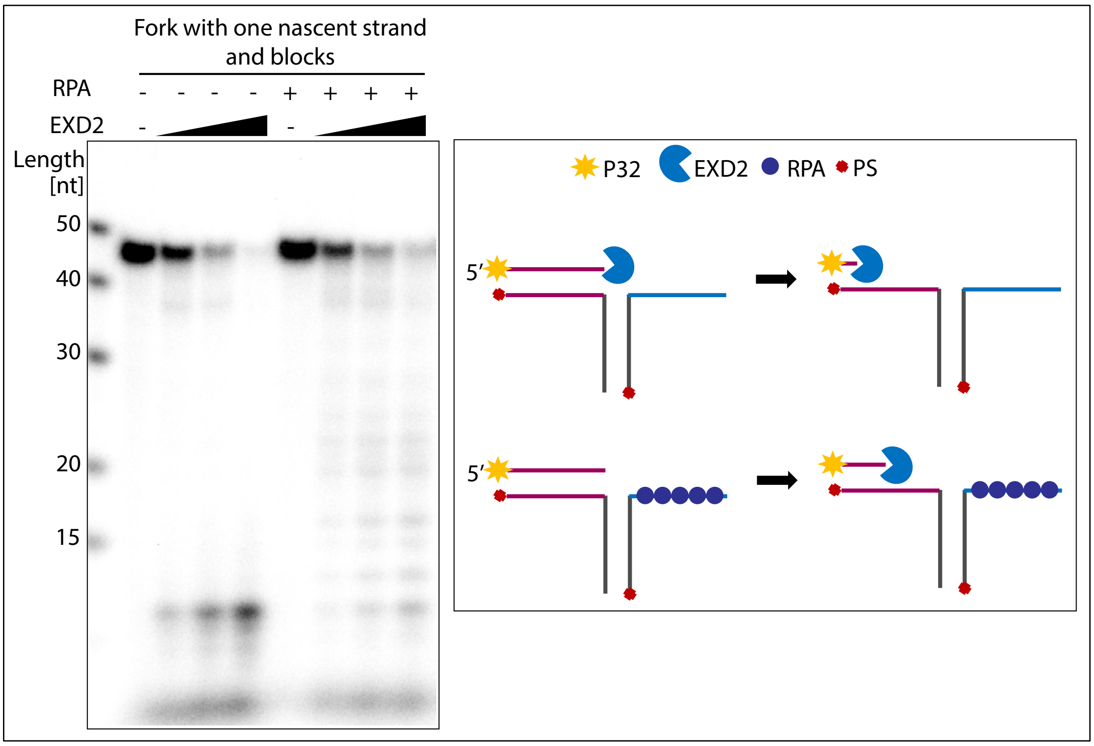


Figure 89. RPA partially decreases EXD2 WT activity on the nascent strand of the fork-like structure. ***5'-radiolabelled DNA substrate (~ 1 nM) was incubated with RPA (390 nM) for 20 min. Next, EXD2 WT (0; 5.3; 10.6; 21.2 nM) was added for the reaction (1h; 37°C). Samples were resolved in TBE-urea polyacrylamide gel and visualised by phosphor imaging. The reaction buffer contained 5 mM $MnCl_2$. Representative image of one of three repeats.***

It is unclear what causes the reduction in EXD2 activity on substrates resembling replication forks with one nascent strand in the presence of RPA. The unannealed template strand, which is most likely coated by RPA, is not accessible for EXD2 exonucleolytic digestion even in the absence of RPA, due to the PS block on the 3' end. While EXD2 digests the nascent strand, the longer regions of the template strand become single stranded. It is possible that even though these regions are too short for high affinity binding of RPA, RPA migrates towards these regions physically occluding smooth movement of EXD2 on double stranded parts of the substrate.

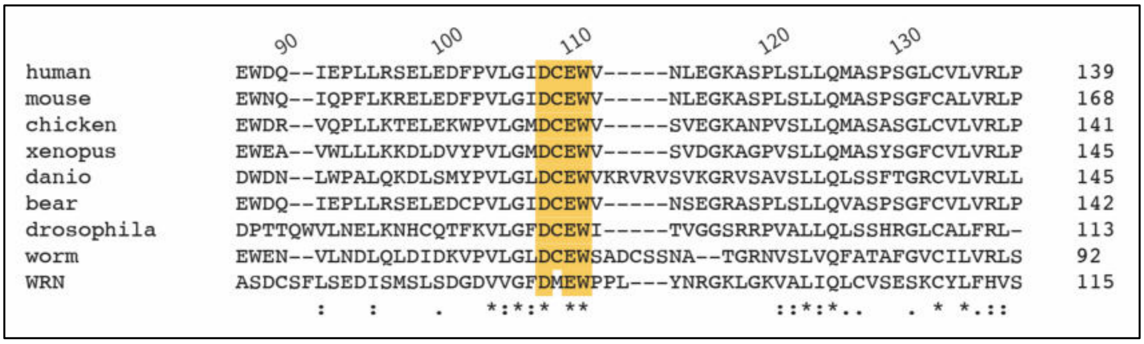
Conclusions

RPA blocks EXD2 activity on ssDNA of different types of substrates (split arms, dsDNA with a 3' overhang). This inhibition is caused by RPA binding to ssDNA with a higher affinity than EXD2. The presence of RPA decreases EXD2 activity on double stranded DNA ends but does not block the digestion completely. It is difficult to conclude what is the cause of this reduced digestion.

6. Structural characterization of EXD2

6.1 Primary structure of EXD2 – bioinformatic analysis

To analyse EXD2 sequence conservation throughout evolution, I used the Clustal Omega server. I aligned amino acid sequences from eight organisms: *Homo sapiens, Mus musculus, Gallus gallus, Xenopus tropicalis, Danio rerio, Ursus maritimus, Drosophila melanogaster* and *Trichinella spiralis* (Figure 90). EXD2 displays some sequence homology between species, which is especially high among vertebrates. Importantly, the catalytic core residues within exonuclease active site (Asp108 and Glu110 in human) are highly conserved among all analysed species. EXD2 shares some sequence similarity with Werner syndrome protein (WRN) exonuclease domain (Figure 90, highlighted), and the two key amino acids within the active site are conserved (Asp82 and Glu84 in WRN). Mutation of the Asp82 residue in WRN (B. Li et al., 2009), as well as Asp108 and Glu110 mutations in EXD2 (Broderick et al., 2016), result in lack of exonuclease activity.

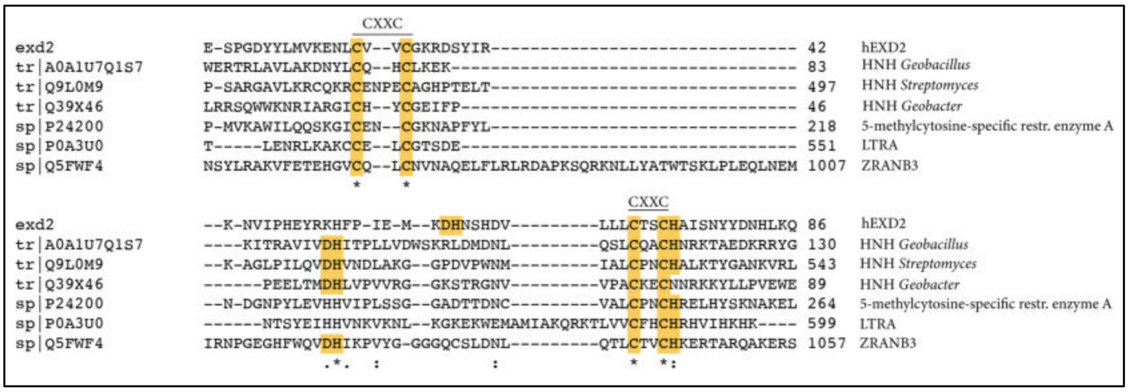


```
                   90          100        110            120       130
human           EWDQ--IEPLLRSELEDFPVLGIDCEWV-----NLEGKASPLSLLQMASPSGLCVLVRLP    139
mouse           EWNQ--IQPFLKRELEDFPVLGIDCEWV-----NLEGKASPLSLLQMASPSGFCALVRLP    168
chicken         EWDR--VQPLLKTELEKWPVLGMDCEWV-----SVEGKANPVSLLQMASASGLCVLVRLP    141
xenopus         EWEA--VWLLLKKDLDVYPVLGMDCEWV-----SVDGKAGPVSLLQMASYSGFCVLVRLP    145
danio           DWDN--LWPALQKDLSMYPVLGLDCEWVKRVRVSVKGRVSAVSLLQLSSFTGRCVLVRLL    145
bear            EWDQ--IEPLLRSELEDCPVLGIDCEWV-----NSEGRASPLSLLQVASPSGFCVLVRLP    142
drosophila      DPTTQWVLNELKNHCQTFKVLGFDCEWI-----TVGGSRRPVALLQLSSHRGLCALFRL-    113
worm            EWEN--VLNDLQLDIDKVPVLGLDCEWSADCSSNA--TGRNVSLVQFATAFGVCILVRLS    92
WRN             ASDCSFLSEDISMSLSDGDVVGFDMEWPPL---YNRGKLGKVALIQLCVSESKCYLFHVS    115
                      :   :    .   *:*:* **              ::*:*..   . * *.::
```

Figure 90. Amino acid alignment of human EXD2 with WRN, and with EXD2 sequences present among other species. Conserved residues are indicated by stars. Asp108 and Glu110 (human EXD2) are highly conserved among analysed species. https://www.ebi.ac.uk/Tools/msa/clustalo/*

As EXD2 exonuclease domain (aa 76-295) has already been crystalized and characterized by (Park et al., 2019), I decided to focus on uncrystallized domains. EXD2 has a HNH endonuclease-like domain characterized by tandem CXXC motifs (Silva et al., 2018). It had been previously reported that EXD2 HNH-like domain lacks the conserved DH catalytic residues required for activity in other HNH endonucleases (Silva et al., 2018). To find remote homology between EXD2 HNH-like domain and other proteins I used HHpred server. HHpred found similarities between the C-terminal part of EXD2 and several proteins, including: 5-methylcytosine-specific restriction enzyme A (sp|P24200), Group II intron-encoded protein LTRA (sp|P0A3U0), HNH endonuclease from *Geobacillus virus E2* (tr|A0A1U7Q1S7), HNHc domain-containing protein from *Streptomyces coelicolor* (tr|Q9L0M9) and HNH endonuclease family protein from *Geobacter metallireducens* (tr|Q39X46). Subsequently I aligned all these proteins, together with DNA annealing helicase and endonuclease ZRANB3 (sp|Q5FWF4) and

EXD2 sequences, using Clustal Omega server (Figure 91, labelled as Uniprot numbers in brackets). The analysis shows DH catalytic residues between the tandem CXXC motif, yet their position in EXD2 sequence is shifted. Importantly, 5-methylcytosine-specific restriction enzyme A and Group II intron-encoded protein LTRA in the alignment, do not encode conserved DH residues, yet they exhibit endonuclease activity. The aspartic residue is not as highly conserved as the histidine residue, which is present in all aligned proteins, including EXD2. This indicates that although EXD2 does not encode DH residues in the classically conserved position, it may exhibit endonuclease activity.



```
                               CXXC
exd2            E-SPGDYYLMVKENLCV--VCGKRDSYIR------------------------------- 42   hEXD2
tr|A0A1U7Q1S7   WERTRLAVLAKDNYLCQ--HCLKEK----------------------------------- 83   HNH Geobacillus
tr|Q9L0M9       P-SARGAVLKRCQKRCENPECAGHPTELT------------------------------- 497  HNH Streptomyces
tr|Q39X46       LRRSQWWKNRIARGICH--YCGEIFP---------------------------------- 46   HNH Geobacter
sp|P24200       P-MVKAWILQQSKGICEN--CGKNAPFYL------------------------------- 218  5-methylcytosine-specific restr. enzyme A
sp|P0A3U0       T-----LENRLKAKCCE--LCGTSDE---------------------------------- 551  LTRA
sp|Q5FWF4       NSYLRAKVFETEHGVCQ--LCNVNAQELFLRLRDAPKSQRKNLLYATWTSKLPLEQLNEM 1007 ZRANB3
                               *    *

                                                           CXXC
exd2            --K-NVIPHEYRKHFP-IE-M--KDHNSHDV---------LLLCTSCHAISNYYDNHLKQ 86   hEXD2
tr|A0A1U7Q1S7   ----KITRAVIVDHITPLLVDWSKRLDMDNL---------QSLCQACHNRKTAEDKRRYG 130  HNH Geobacillus
tr|Q9L0M9       --K-AGLPILQVDHVNDLAKG--GPDVPWNM---------IALCPNCHALKTYGANKVRL 543  HNH Streptomyces
tr|Q39X46       ------PEELTMDHLVPVVRG--GKSTRGNV---------VPACKECNNRKKYLLPVEWE 89   HNH Geobacter
sp|P24200       --N-DGNPYLEVHHVIPLSSG--GADTTDNC---------VALCPNCHRELHYSKNAKEL 264  5-methylcytosine-specific restr. enzyme A
sp|P0A3U0       ------NTSYEIHHVNKVKNL--KGKEKWEMAMIAKQRKTLVVCFHCHRHVIHKHK---- 599  LTRA
sp|Q5FWF4       IRNPGEGHFWQVDHIKPVYG-GGGQCSLDNL---------QTLCTVCHKERTARQAKERS 1057 ZRANB3
                            .*.  :           :             *  *:
```

Figure 91. Alignment of EXD2 HNH-like domain sequence with other proteins. Conserved residues are indicated by stars*. The tandem CXXC motif is highly conserved among analysed proteins. https://www.ebi.ac.uk/Tools/msa/clustalo/

Importantly, the tandem CXXC motif in EXD2 HNH-like domain sequence is highly conserved in EXD2 from various organisms, indicating its important function throughout evolution (Figure 92).

```
worm            KNNLGDVVSQLSDEHWKLFSCAEEDPPFVVQLRFQPKVVAS--KESYYIKEKENQCVVCG    349
drosophila      NQNLGTHISE---------------EPFTVRLNFEPAGRAVGDVGRFYQTIKKNQCVVCG    371
danio           EKGIGELVSE---------------DPFIVRLLFEPSGRPDS-EKDYYLTAKENLCVVCG    418
xenopus         DKGIGDLVST---------------IPFVVKLRFEPSGRPES-NADYYLTVKENLCVVCG    415
chicken         DKGIGDLVST---------------DPFVVKLRFEPSGRPES-DVDYYLMVKENLCVVCG    424
mouse           DKGIGELVSK---------------EPFVVRLQFEPAGRPES-PGDYYLMVKENLCVVCG    453
human           DKGIGELVSE---------------EPFVVKLRFEPAGRPES-PGDYYLMVKENLCVVCG    424
bear            DKGIGELVSE---------------EPFVVKLQFEPAGRPES-PGDYYLMIKENLCVVCG    427
                .:.:*  :*                 ** *:* *:*          :*   *:* *****

worm            RLDLLMRKRIIPHLYRRLFPEELKSFRSHDILLLCFDCHRKAEYYDTLFQRSLAEQCDAP    409
drosophila      DRDAYIRKNVVPREYRKHFPLVMKSHTSDDVLLLCPTCHQLSNISDLRVRSKLAVQCEAP    431
danio           KVDSYIRKNIVPHEYRRHFPSEMKDHNSHDILLLCTACHAASNVHDGALKAELAEEHAAP    478
xenopus         KRESYIRKNIVPHEYRRHFPVQMKDHNSHDVLLLCTSCHAVSNYYDSNLKQKLAEEFNAP    475
chicken         KRESYIRKNIVPHEYRRHFPIQMKDHNSHDVLLLCTSCHAISNYYDNHLKQQLAEEFGAP    484
mouse           KTDTYIRKNIIPHEYRKHFPIEMKDHNSHDVLLLCTSCHAISNYYDNHLKQQLAKEFQAP    513
human           KRDSYIRKNVIPHEYRKHFPIEMKDHNSHDVLLLCTSCHAISNYYDNHLKQQLAKEFQAP    484
bear            KKDSYIRKNVIPHEYRKHFPIEMKDHNSHDVLLLCTSCHAISNYYDNHLKQQLAREFQAP    487
                  :  :**.::*: **: **  :*.. *.*:****  **  ::  *  .: .** :  **
```

Figure 92. Amino acid alignment of human EXD2 with EXD2 sequences present among other species. Conserved residues are indicated by stars*. The tandem CXXC motif in the HNH-like domain is highly conserved among analysed species. https://www.ebi.ac.uk/Tools/msa/clustalo/

6.2 EXD2 secondary structure prediction – bioinformatic analysis

To predict secondary structure of full-length EXD2, I used protein structure prediction servers SABLE and PSIPRED. The results of the PSIPRED analysis (Figure 93) slightly differ from the SABLE analysis (Supplementary Figure), especially in the N-terminal region, where PSIPRED suggests a helical structure, whereas SABLE predicts helices together with beta-strands.

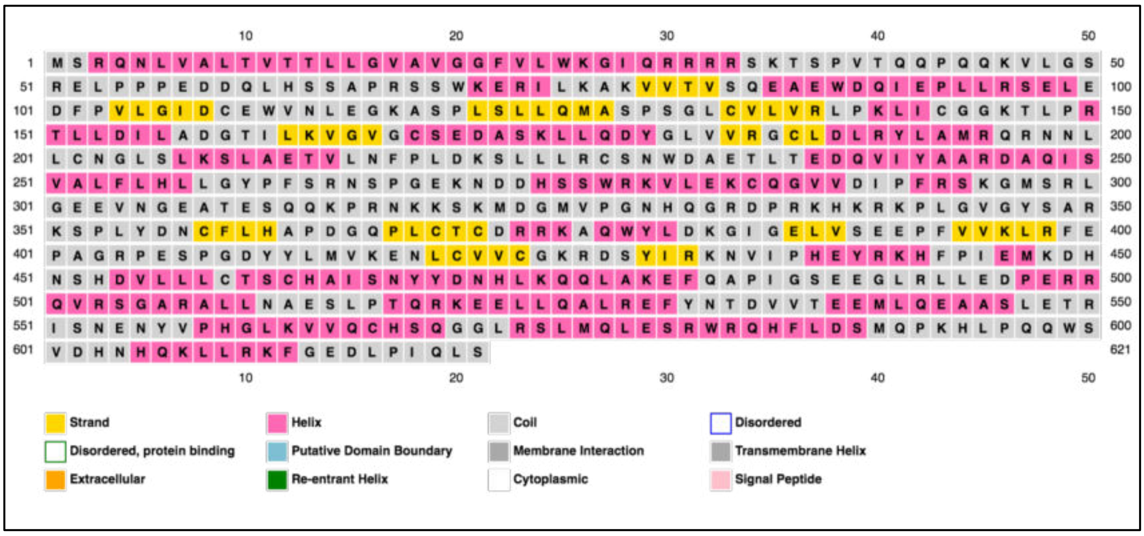


Figure 93. Secondary structure prediction of full-length EXD2; http://bioinf.cs.ucl.ac.uk/psipred/

According to the previous hydropathy analysis (Park et al., 2019), EXD2 may contain a single transmembrane domain at the N-terminus (aa 1-37). The presence of polar amino acids followed by hydrophobic and nonpolar residues on the N-terminus may support this hypothesis (Figure 94). Likewise, TMHMM server predicts that a transmembrane helix should be located between amino acids Val7 and Trp25 (Figure 95A), whereas TMpred (Figure 95B) predicts three strong transmembrane helices (aa 7-25, 118-135 and 240-262, respectively).

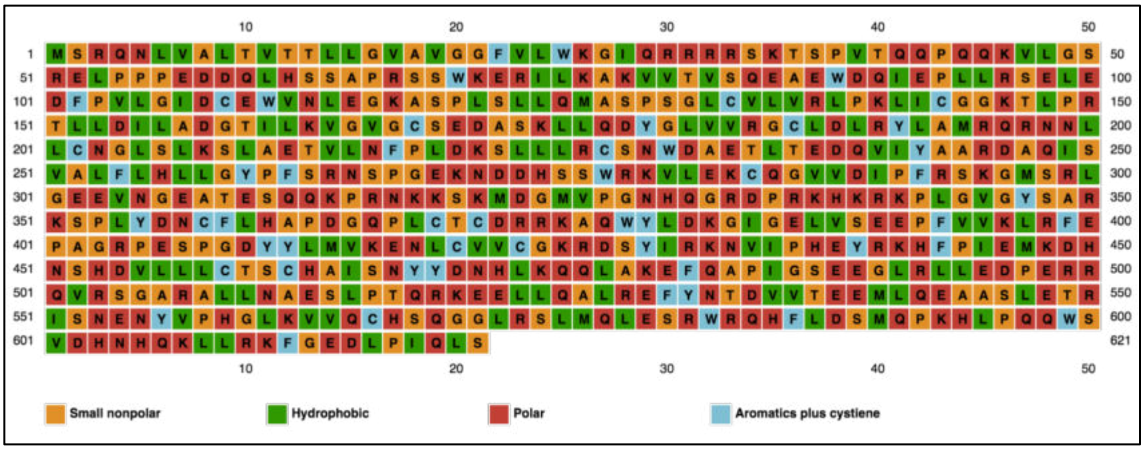


Figure 94. Full-length EXD2 amino acid sequence; http://bioinf.cs.ucl.ac.uk/psipred/

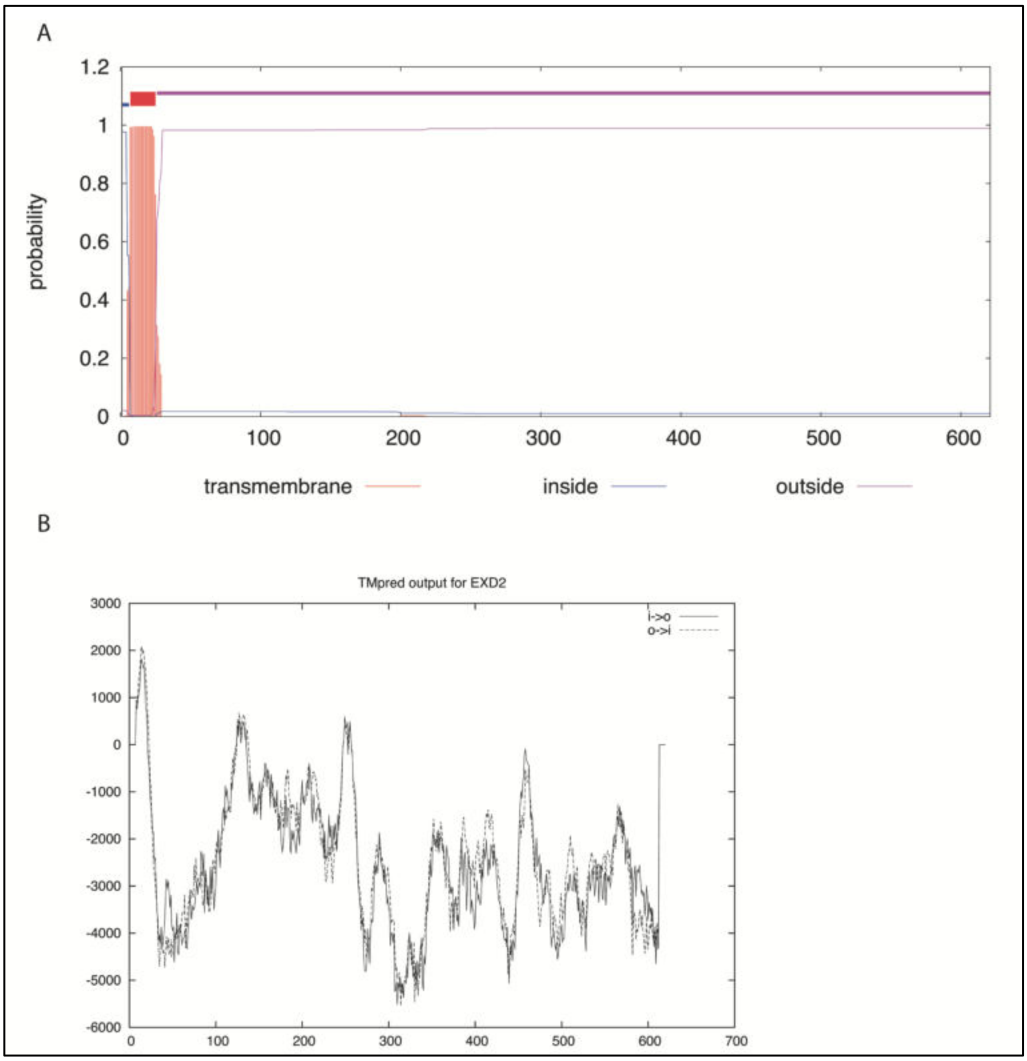


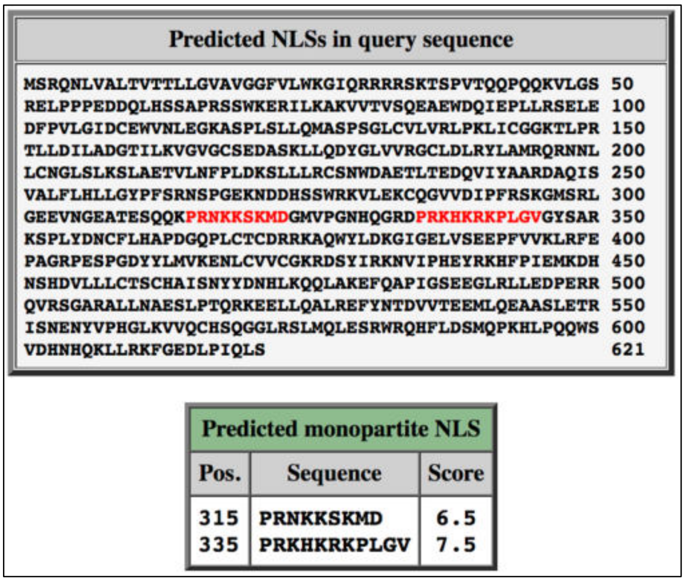
Figure 95. A) Prediction of transmembrane helices in EXD2; http://www.cbs.dtu.dk/services/TMHMM/ B) Prediction of transmembrane helices in EXD2; https://embnet.vital-it.ch/software/TMPRED_form.html.

It is highly possible that a transmembrane domain is indeed present in the N-terminal region of EXD2, nevertheless the results should be used with extreme caution as the predictions differ between the servers.

6.3 Prediction of EXD2 cellular location – bioinformatic analysis

Silva and colleagues (Silva et al., 2018) proposed that EXD2 N-terminus contains a putative mitochondrial targeting sequence (MTS, aa 1-61), which indicates mitochondrial localisation of the protein. However, according to Park and colleagues (Park et al., 2019), the N-terminal transmembrane domain anchors EXD2 to mitochondrial outer membrane, whereas the C-terminus adopts a cytosolic location. Since EXD2 probably has the transmembrane domain, it may be geometrically difficult to associate and form the active dimer. It is not known if the dimer would be formed through homotypic (within intra-mitochondrial membrane) or heterotypic interaction (through inter-membrane).

I decided to focus on nuclear location of EXD2, as it was confirmed from *in vivo* experiments using GFP- and FLAG-tagged EXD2 (Nieminuszczy et al., 2019). GFP-EXD2 exhibits both nuclear and cytosolic localization of the protein, with EXD2 recruited to laser induced DNA damage (Nieminuszczy et al., 2019). Data by Sandoz and colleagues also show that upon DNA damage EXD2 is recruited to chromatin (Sandoz et al., 2022). To predict potential nuclear localization sequence (NLS) of EXD2, I used NLS Mapper software. The obtained results confirm nuclear presence of the protein (Figure 96). However, how EXD2 regulates its mitochondrial – nuclear location has not been established so far.



Predicted NLSs in query sequence

```
MSRQNLVALTVTTLLGVAVGGFVLWKGIQRRRRSKTSPVTQQPQQKVLGS  50
RELPPPEDDQLHSSAPRSSWKERILKAKVVTVSQEAEWDQIEPLLRSELE  100
DFPVLGIDCEWVNLEGKASPLSLLQMASPSGLCVLVRLPKLICGGKTLPR  150
TLLDILADGTILKVGVGCSEDASKLLQDYGLVVRGCLDLRYLAMRQRNNL  200
LCNGLSLKSLAETVLNFPLDKSLLLRCSNWDAETLTEDQVIYAARDAQIS  250
VALFLHLLGYPFSRNSPGEKNDDHSSWRKVLEKCQGVVDIPFRSKGMSRL  300
GEEVNGEATESQQKPRNKKSKMDGMVPGNHQGRDPRKHKRKPLGVGYSAR  350
KSPLYDNCFLHAPDGQPLCTCDRRKAQWYLDKGIGELVSEEPFVVKLRFE  400
PAGRPESPGDYYLMVKENLCVVCGKRDSYIRKNVIPHEYRKHFPIEMKDH  450
NSHDVLLLCTSCHAISNYYDNHLKQQLAKEFQAPIGSEEGLRLLEDPERR  500
QVRSGARALLNAESLPTQRKEELLQALREFYNTDVVTEEMLQEAASLETR  550
ISNENYVPHGLKVVQCHSQGGLRSLMQLESRWRQHFLDSMQPKHLPQQWS  600
VDHNHQKLLRKFGEDLPIQLS                               621
```

Predicted monopartite NLS

Pos.	Sequence	Score
315	PRNKKSKMD	6.5
335	PRKHKRKPLGV	7.5

Figure 96. EXD2 contains a putative nuclear localization sequence. Result from cNLS mapper.

6.4 EXD2 structure prediction by AlphaFold

AlphaFold is an AI system created by DeepMind and EMBL's European Bioinformatics Institute (EMBL-EBI) to predict protein's 3D structure from amino acid sequences (Jumper et al., 2021). The model confidence is based on per-residue confidence score (pLDDT) that ranges between 0 and 100. Protein secondary structures, including alpha-helices and beta-sheets, show spatial conformations of amino acid chains in a cartoon / ribbon visual representation (Kocincová et al., 2017). The spatial positioning of secondary structures is then observed as a 3D view.

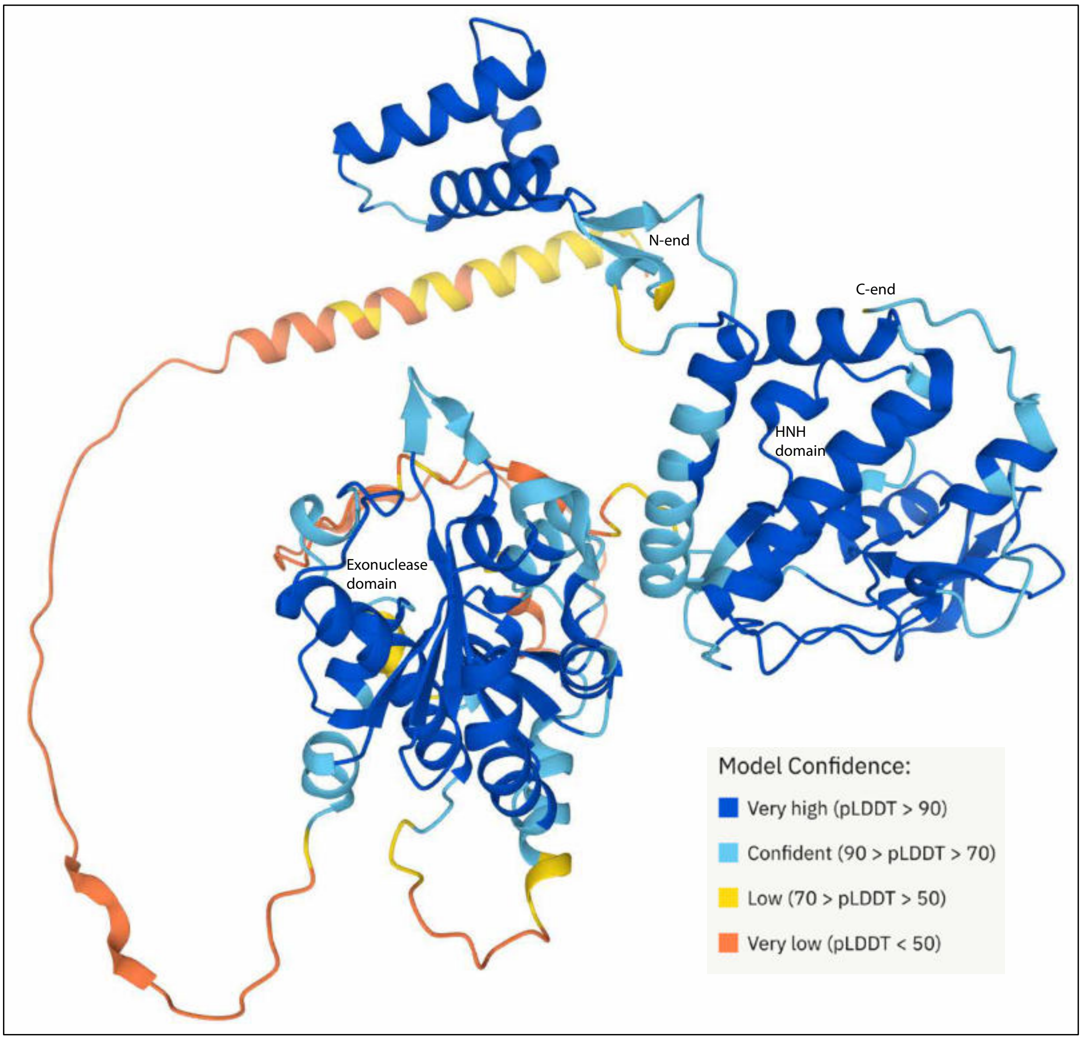


Figure 97. Human EXD2 model – based on AlphaFold prediction. Model confidence differs between the regions (high for the exonuclease domain and HNH domain; low for the N-terminal part of EXD2 and the linker region between the domains). https://alphafold.ebi.ac.uk/entry/Q9NVH0

The human EXD2 AlphaFold model (Figure 97) shows high confidence (blue) in the structure of exonuclease region, which has already been crystalized (Park et al., 2019), and in the predicted structure of the HNH domain. The lowest model confidence is represented by the orange and yellow colours at the N-terminal part of EXD2. A long protein linker connects the N-terminal transmembrane domain with the rest of the protein. It has high flexibility and can rotate at different angles, hence the model confidence is low. The exonuclease and HNH domains are connected with each other by another linker region of a low model confidence. It suggests that these two domains can be positioned at different orientations relative to each other.

6.5 Preliminary 3D structure of purified EXD2

6.5.1 Overview

A full-length EXD2 structure has not been established so far, as only its exonuclease domain (aa 76-295) has been crystalized (Park et al., 2019). To provide structural insights into the full-length protein, purified EXD2 was analysed using negative staining electron microscopy (EM). Preparation of grids, as well as negative staining data collection and analysis was performed by Dr Fabienne Beuron.

Observed specimens must be dehydrated, as the electron path of the microscope is kept under high vacuum (R. F. Thompson, Walker, Siebert, Muench, & Ranson, 2016). One of the most common methods to prepare the sample is a negative staining (Orlova & Saibil, 2011). Here, a droplet of purified protein is placed on the EM carbon support film and then embedded in a heavy metal salt solution, usually uranyl acetate. Excess of the stain is blotted, and the sample is dried before the imaging. Uranyl acetate negatively stains the protein, creating a high contrast (sample is visualised by the lack of stain). This gives information about size, shape, and homogeneity of the sample. This is a fast method of sample preparation, although it has some limitations. The main restriction is the protein size, as particles below 100 – 200 kDa are difficult to visualize. Some molecules are fragile and can disintegrate during staining and drying. Negative staining may also be incompatible with some buffers and can create staining artefacts.

The EM image is created by scattered and unscattered electrons (Orlova & Saibil, 2011). Absorption of part of the beam generates an amplitude contrast, whereas scattering of electrons at different angles creates a phase contrast. Proteins in the solution are present at different orientations and can be observed as two-dimensional projections (Zanotti, 2016). Observation of a large number of homogenous particles of different angular orientations allows reconstitution of a 3D electron density (R. F. Thompson et al., 2016). To image thousands of identical particles, high sample homogeneity is required. Heterogeneity may result from conformation changes and protein composition (degradation, aggregation). Low resolution data from the negative staining (~ > 20 Å) are often used as a starting point to generate high resolution images using cryo-electron microscopy (cryo-EM). Recent advances in cryo-electron microscopy allow for the determination of protein structures at resolution as high as ~ 2 Å (Zanotti, 2016).

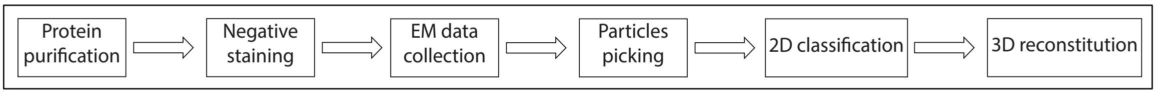


Figure 98. The main steps of protein structure generation.

6.5.2 Negative staining of purified EXD2

Purified full-length EXD2 (10xHis-TEV-hEXD2-3C-2xStrep) was frozen in a gel filtration buffer (50 mM Tris pH 7.5, 300 mM NaCl, 10 % glycerol, 2 mM BME, 0.05 % Triton X-100), which obstructed protein observation under the microscope due to the small size of the particles. Therefore, I performed buffer exchange (25 mM Tris pH 7.5, 100 mM NaCl, 5 % glycerol, 0.5 mM BME) using Zeba spin desalting column. Quality of the purified sample was validated using SDS-PAGE electrophoresis (Figure 99).

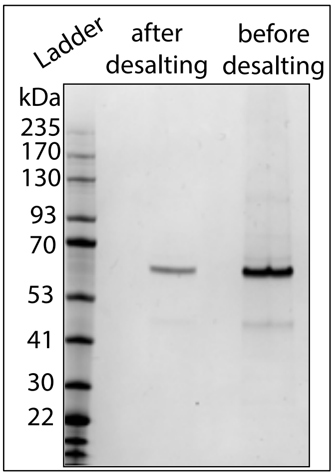


Figure 99. Coomassie stained gel of purified full-length EXD2. After desalting on a Zeba spin column, EXD2 was used for negative stain EM.

EM data from negatively stained protein was collected by Dr Fabienne Beuron, who performed 2D classifications (Figure 100). The preliminary 3D model of purified EXD2 was computed in Relion. According to (Park et al., 2019), EXD2's exonuclease domain forms a dimer through 2-fold rotational symmetry therefore C2 symmetry was applied during the processing. Crystalized EXD2 exonuclease domain (6K1A, Protein Databank) was docked into the negative stain density model using Chimera (Figure 100D). To confirm that the observed particles are not artefacts, gel filtration buffer was used as a negative control.

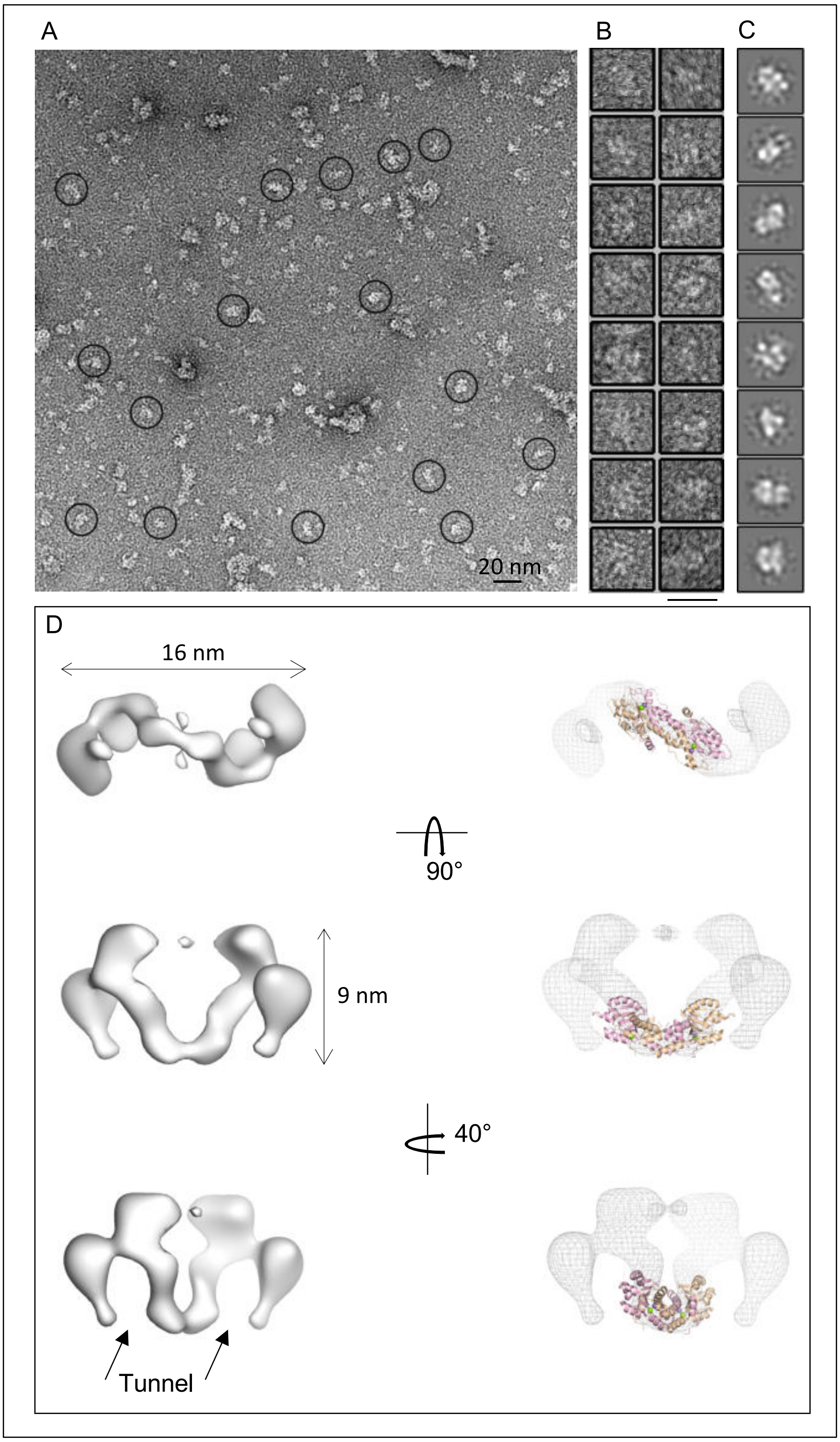


Figure 100. Preliminary 3D structure of purified full-length EXD2 generated by Dr Fabienne Beuron. A) Micrograph of negatively stained EXD2. B) Particles chosen for further analysis. C) 2D classification of chosen particles. D) Preliminary 3D structure of full-length EXD2. 2-fold symmetry applied. Negative stain density map (grey) together with crystalized structure of EXD2 exonuclease domain (ribbon representation, 6K1A from Protein Data Bank). Green and purple spheres indicate active site metal ions.

The preliminary 3D structure (30 Å resolution) shows that the interaction between EXD2 monomers is mostly through the exonuclease domain. The overall shape of the dimer suggests that the exonuclease active site is close to the tunnel, which could be an entry site for DNA substrates. Active site metal ions (indicated by green and purple spheres) are exposed and located in the vicinity of potential substrates. Heterogeneity of negatively stained sample shows that the protein is unstable under these conditions. The selected particles are consistent with a dimer form and smaller molecules can be either monomers or the results of degradation. As presented by (Park et al., 2019), EXD2 exonuclease is a domain-swapped dimer with chimeric active site. The molecular weight of human dimer EXD2 is around 140 kDa, which is on the border line of EM visualisation. Recombinant full-length protein was visualised under detergent-free conditions, which could result in its instability and changes in observed composition (monomers, dimers, aggregates). Importantly, a crystal structure of the exonuclease domain fits the EM density, confirming that the collected data do not visualise artefacts. To obtain data of higher resolution and to use cryo-electron microscopy (cryo-EM), the quality of the sample had to be improved.

To increase homogeneity of the protein and to reduce residual background, I used size exclusion chromatography. The frozen EXD2 (50 mM Tris pH 7.5, 300 mM NaCl, 10 % glycerol, 2 mM BME, 0.05 % Triton X-100) was reloaded on the gel filtration column (HiLoad 16/60 Superdex 200), to separate aggregates and degraded protein. Although the gel filtration buffer was free from the detergent (50 mM Tris pH 7.5, 300 mM NaCl, 10 % glycerol, 2 mM BME), high absorbance values observed on the gel filtration plot suggest that residual Triton X-100 from the frozen sample eluted together with the EXD2 (Figure 101 A). The elution profile (Figure 101 A) suggests that majority of the eluted proteins were monomers (76.5 kDa) and dimers. The negative stain data shown that sample was more consistent with shape and size, but the undesirable background was higher than after Zeba spin desalting column. Therefore, I performed a new purification, where Triton X-100 was used only during the lysis and first affinity step (StrepTrap column). The second affinity step (HisTrap column), as well as the gel filtration (50 mM Tris pH 7.5, 300 mM NaCl, 10 % glycerol, 2 mM BME), were performed without any detergent. Although lack of Triton X-100 eliminated undesirable background from the negative staining, it resulted in massive protein aggregation (Figure 101 B). The shift of the gel filtration peak suggests that majority of EXD2 eluted as the aggregates with the molecular weight of over 440 kDa.

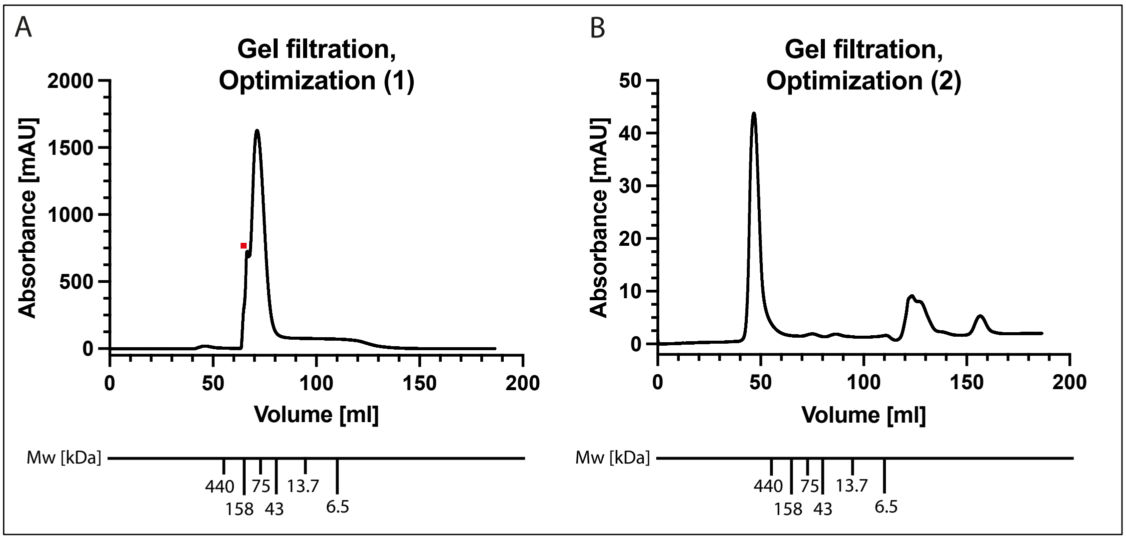


Figure 101. Preparation of EXD2 for the negative staining. A) EXD2 frozen in the presence of 0.05 % Triton X-100 was reloaded on the gel filtration column and eluted with detergent free buffer. Negatively stained protein, indicated by the red square, was homogenous in shape and size. However, undesirable background prevented further analysis. B) EXD2 purification performed without any detergent in the last two steps (His Tag affinity and GF). Lack of Triton X-100 resulted in aggregation of the protein (elution at 50 ml). To use purified protein for other purposes, protease 3C was added between two affinity steps (in B purification only, but this should not affect the observed aggregation). The molecular weight is based on a chromatographic separation for standard proteins on HiLoad 16/60 Superdex 200 columns according to GE Healthcare.

To decrease aggregation of the eluted protein (Figure 101 B), NP40 was added to the sample. It is non-ionic detergent, used for protein solubilization, which should not create undesirable background. Addition of 0.1 % NP40 slightly decreased aggregation after gel filtration, yet the protein was still too aggregated (Figure 102) to improve resolution of preliminary 3D structure obtained so far.

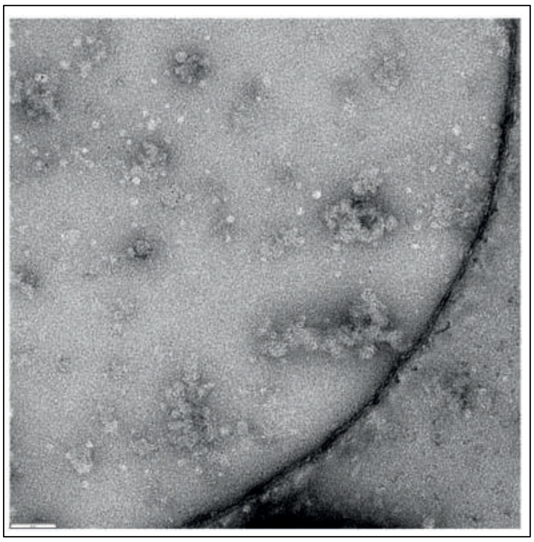


Figure 102. Micrograph of negatively stained EXD2 after addition of 0.1 % NP40, showing high protein aggregation; scale bar of 60 nm. Picture taken by Dr Fabienne Beuron.

All the tested conditions did not improve homogeneity of the sample. Negatively stained protein was either too degraded or too aggregated. Therefore, I decided to generate an EXD2 construct deprived of the transmembrane domain.

Truncated EXD2 (10xHis-TEV-truncEXD2-3C-2xStrep, without aa 2-37) was purified according to the protocol described in (4.1.4). The lack of the transmembrane domain resulted in increased solubility of the protein, hence no detergent was used during the purification (50 mM Tris pH 7.5, 300 mM NaCl, 10 % glycerol, 2 mM BME). As the preliminary results of negatively stained truncated protein were not satisfying, 50 bp dsDNA was added to the solution. The addition of DNA slightly improved the stability of the protein and truncated EXD2 was visualised on the micrograph (Figure 103). Results of 2D classification of the truncated EXD2 were similar to the 2D classes of the full-length protein (Figure 103 B and C). This analysis indicates that the presence of the transmembrane domain does not affect the overall shape of the protein and its ability to dimerize. Due to the low number of homogenous particles, a 3D structure was not generated.

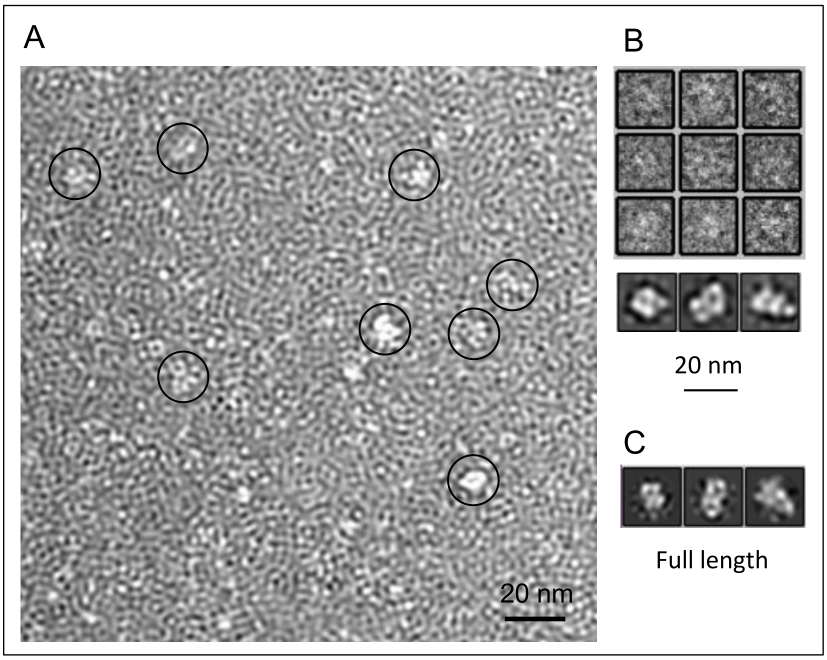


Figure 103. Truncated EXD2 (10xHis-TEV-truncEXD2-3C-2xStrep, without 2-37 aa). A) Negatively stained micrograph in the presence of dsDNA. B) Particles chosen for further analysis and 2D classification of truncated EXD2. C) 2D classification of the full-length EXD2. Dr Fabienne Beuron

6.5.3 Conclusions

To generate a high-resolution 3D structure of either full-length EXD2 or EXD2 without the transmembrane domain, more optimization steps need to be performed. It is essential to increase homogeneity of the sample and reduce undesirable background observed under EM. As EXD2 is a nuclease, the presence of DNA or RNA in the sample may increase the stability of the protein. Addition of chemicals, like Amphipol A8-35 or Lauryl Maltose Neopentyl Glycol, should also improve protein solubilization.

6.6 Comparison of AlphaFold prediction with a preliminary 3D structure of purified human EXD2

AlphaFold structure prediction (6.4) was fitted into a negative stain density map of purified full-length EXD2 (6.5.2). Visualisation was performed by Dr Fabienne Beuron using PyMOL software (Figure 104).

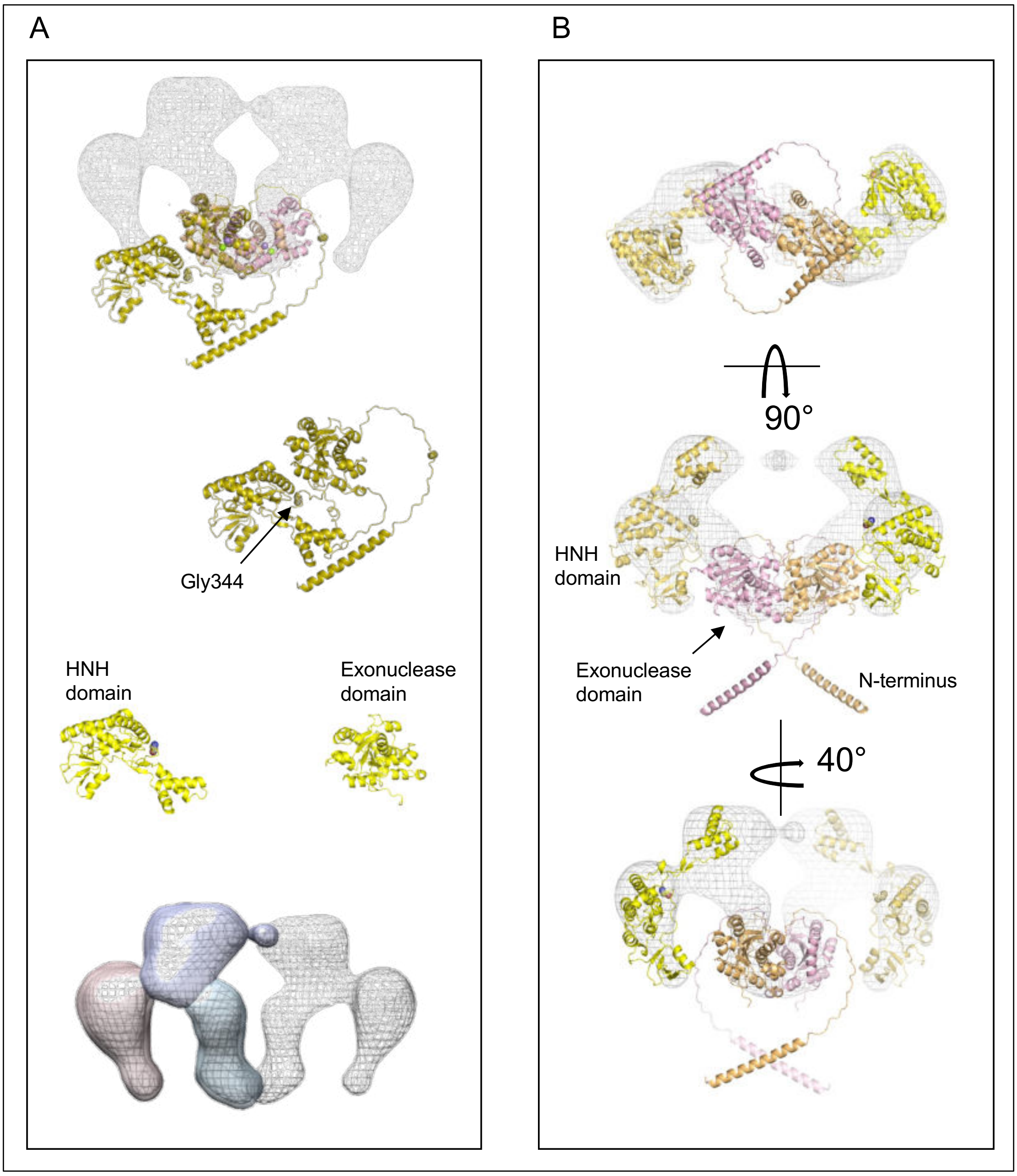


Figure 104. Comparison of AlphaFold prediction (ribbon representation) with a preliminary 3D structure of purified full-length EXD2 (negative stain density map), by Dr Fabienne Beuron. A) The predicted 3D model was cut at Gly344 of the linker region (low prediction confidence of the linker), which allowed to move the HNH domain and the exonuclease domain separately. Gly344 is represented as a red-blue dot in the ribbon structure. B) The cleaved model was fitted into negative stain density map of the full-length EXD2. Two EXD2 monomers were used for the visualisation.

The predicted full-length structure was aligned to the fitted coordinates of the exonuclease domain dimer (Figure 104). Aligning of the exonuclease domain from the

AlphaFold prediction, with the negative stain density map, resulted in positioning half of the predicted structure away from the density. Therefore, the predicted structure was cut at Glycine 344, located in the linker region of a very low confidence - between the exonuclease (aa 76 – 295) and HNH (aa 408 – 470) domains. This allowed the HNH domain to move as a single rigid body and fit more accurately in the negative stain map. As EXD2 acts as a dimer, two monomers were used for the visualization. The AlphaFold structure prediction fits accurately into the negative stain density map. It allows identification of which part of the map represents the N- terminal part of the protein. As the predicted N-terminal transmembrane domain is connected to the core of the protein through a long flexible linker, it is possible that its orientation is different and it is directed upwards, not downwards. To validate its real orientation, higher resolution density map needs to be generated. A higher resolution may enable us to determine if the monomers interact with each other through residues located in the upper part of the density map, not only through the exonuclease domain. Interestingly, AlphaFold prediction automatically positions itself as separate monomers (Figure 104 B), not a dimer with a composite active site as in (Figure 100). It is possible that this is due to the current program limitations. Nevertheless, the positioning difference does not influence the overall shape of the protein. The comparison between the predicted 3D model and experimentally derived data allowed to visualise that both exonuclease and HNH-domains are located close to a channel. This channel maybe the entry site for nucleic acids, which are cleaved by EXD2 nuclease. Similarities between the low-resolution structure and high-resolution AlphaFold ribbon representation suggest the accuracy of experimentally derived structural data.

7. Generation of research tools for reconstitution of DNA resection process

To fully understand the role of EXD2 in DNA end resection, the whole resection reaction should be reconstituted *in vitro*. As the MRN complex, CtIP and RPA are the essential components of short-range resection machinery in HR (Moynahan & Jasin, 2010), they need to be purified and tested together with EXD2. Therefore, I optimized purification protocols for MRN and MR, which can be used in a future project elucidating the molecular mechanisms of DNA end resection.

7.1 MRN complex (MRE11-RAD50-NBS1) purification

7.1.1 Overview

MRN has been the subject of research for past 20 years (Paull, 2018), nevertheless its structure and function are still not fully known. One of the challenges on working with this complex is a low purification yield (J. H. Lee & Paull, 2006), (Myler et al., 2019). Therefore, I have been optimizing an MRN purification protocol for both WT (hMRE11-TEV-10xHis, hRAD50-TEV-2xStrep, hNBS1-TEV-2xStrep) and M(H129L D130V)RN mutant. According to (Anand et al., 2016), MRE11 (H129L D130V) mutation results in the lack of 3'-5' exonuclease and endonuclease activities, hence purification of this protein can be used as a negative control in nuclease assays.

7.1.2 Optimization of MRN purification

To purify MRE11, RAD50 and NBS1 as a complex, Hi5 cells were transfected with one construct encoding all three proteins (hMRE11-TEV-10xHis, hRAD50-TEV-2xStrep, hNBS1-TEV-2xStrep). To optimize the purification protocol, I tested two types of lysis: standard and hypotonic (Table 19), (Figure 105).

Table 19. MRN purification – types of cell lysis tested.

Cell lysis	Procedure	Conditions
Standard	1. Dounce homogenizer 2. Sonication 3. Incubation with benzonase 4. Centrifugation at 37 000 g	Lysis buffer: 50 mM Tris pH 7.5, 300 mM NaCl, 10% glycerol, 2 mM Mg^{2+}, 2 mM Mn^{2+}, 2 mM BME, SIGMAFAST protease inhibitors
Hypotonic	1. Hypotonic buffer 2. Incubation on the roller 3. Hypertonic conditions 4. Incubation on the roller 5. Centrifugation at 800 g 6. Centrifugation at 16 000 g	Hypotonic buffer: 50 mM Tris pH 7.5, 2 mM BME, SIGMAFAST protease inhibitors Hypertonic conditions: + 16.7% (v/v) glycerol, + 305 mM NaCl, + 2 mM Mg^{2+}, + 2 mM Mn^{2+}

Hypotonic lysis was based on a MRN salt extraction protocol by (Anand, Pinto, & Cejka, 2018), with some changes introduced. Centrifugation at 48 000 g, proposed by (Anand et al., 2018), was replaced by two step centrifugation (800 g, followed by 16 000 g). Low speed centrifugation was introduced to pellet the nuclei (Nikos P., 2013) and avoid DNA release. During both purification trials, buffers were supplemented with Mg^{2+} and Mn^{2+}, as both these ions are functionally important for MRN. In the MR complex, MRE11 exhibits Mg^{2+}/Mn^{2+} - dependent nuclease activities (Y. Liu et al., 2016), and RAD50 dimerizes in a Mg^{2+} - dependent manner (Syed & Tainer, 2018). Buffer supplementation with Mg^{2+} should also protect the nuclei from breaking (Nikos P., 2013). However, as Mn^{2+} is not required for the stability of the complex, I decided not to include it in the final protocol.

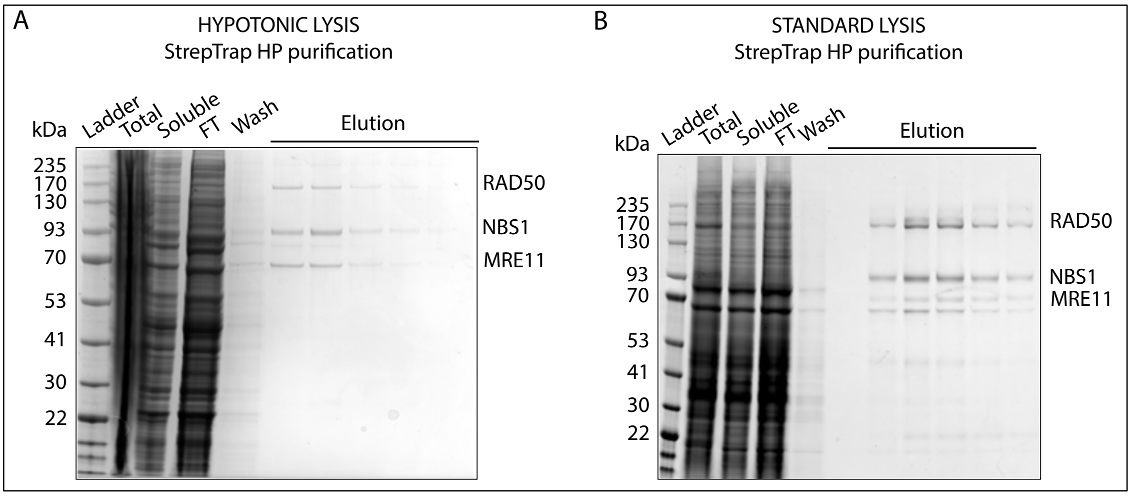


Figure 105. MRN purification from two different pellets of Hi5 cells (240 ml), Coomassie stained gels. Total – lysate before centrifugation; Soluble – soluble lysate after centrifugation; FT – flow through

StrepTrap HP column; Wash – wash of the column; Elution – elution from the column. Total and Soluble fractions were treated with benzonase before loading on the gel. Efficiency of the baculovirus infection varied.

After cell lysis (either standard or hypotonic), the centrifugation supernatant was loaded on the StrepTrap HP column. The Coomassie stained gels (Figure 105) show that fractions eluted with desthiobiotin contained all three subunits of the MRN complex. For further purifications, I chose the standard cell lysis protocol (Table 19). During this step, DNA was degraded by benzonase. Here, the presence of Mg^{2+} ions in the lysis buffer was required for benzonase activity.

To purify MRN as a complex, a gel filtration step was introduced. This should enable separation of the MRN complex from the protein aggregates, subcomplexes and single subunits. Results of this two-step purification (StrepTrap HP followed by Superose 6 10/300 GL) are shown (Figure 106).

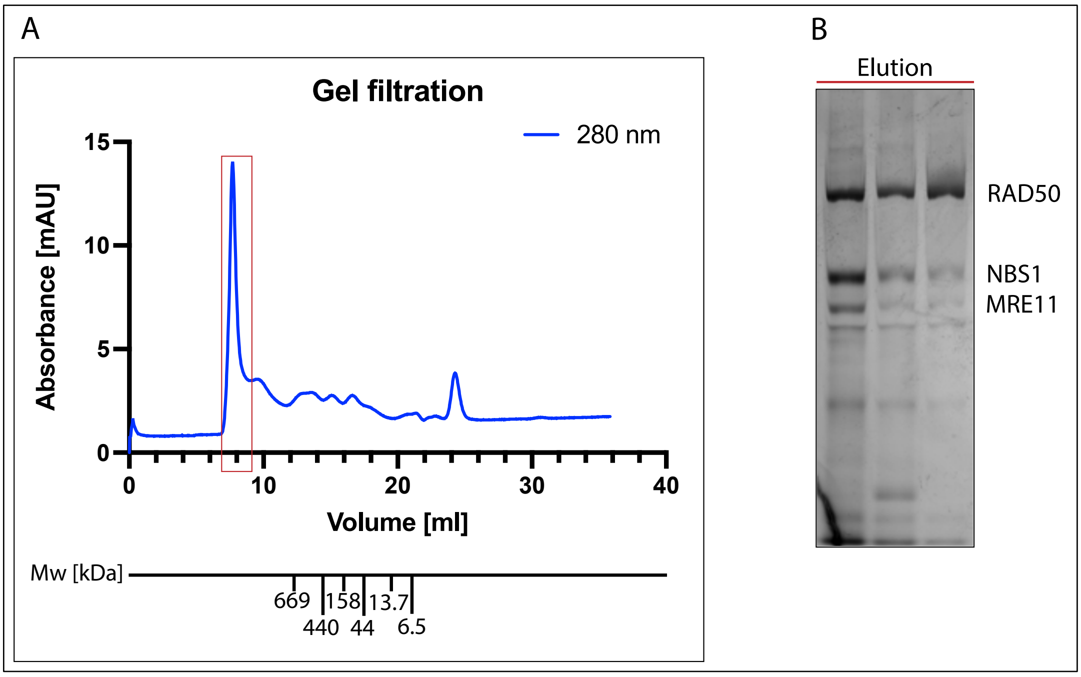


Figure 106. MRN purification performed from 240 ml of Hi5 cells. A) Elution profile from Superose 6 10/300 GL - the final purification step. Fractions indicated by the red box correspond to the eluted fractions shown on the gel (B). The peak contains highly aggregated proteins. B) Silver-stained gel after the final step.

The M_w of the recombinant MRN dimer is around 658.4 kDa. The first peak (Figure 106) corresponds to the proteins eluted in the void volume of the column (in 50 mM Tris pH 7.5, 150 mM NaCl, 10% glycerol, 2 mM Mg^{2+}, 2 mM Mn^{2+}, 2 mM BME), with the M_w of over 669 kDa. Although all MRN subunits were present, they were likely to be protein aggregates.

Observed aggregation could be caused by binding of MRN to the contaminating DNA. Therefore, I introduced a HiTrap Heparin HP column step (Figure 107). Immobilized heparin mimics the structure of nucleic acids, allowing the purification of DNA binding proteins. MRN was eluted from HiTrap Heparin HP with a linear gradient of NaCl (475 mM - first peak; 505 mM - second peak). The purification profile corresponds to (Liao, Guay, Toczylowski, & Yan, 2012) where the majority of MRN was eluted with 400 - 475 mM NaCl. Fractions indicated by the red box in the elution profile (Figure 107 C) and by the red line in the Commassie stained gel (Figure 107 D) were combined and loaded on the gel filtration column. The introduction of HiTrap Heparin HP enabled a better separation of the first two peaks observed in the gel filtration profile (Figure 107 E). Both peaks from gel the filtration profile, indicated either by green or red lines in the plot and in the Coomassie stained gel, contain proteins of M_w of over 669 kDa (based on standard protein resolution according to Cytiva). Although this predicted M_w is higher than the M_w of the recombinant MRN complex dimer (658.4 kDa), it is possible that the first peak (indicated by the green box) contains mostly protein aggregates, whereas the second peak (indicated by the red box) contains the MRN complex.

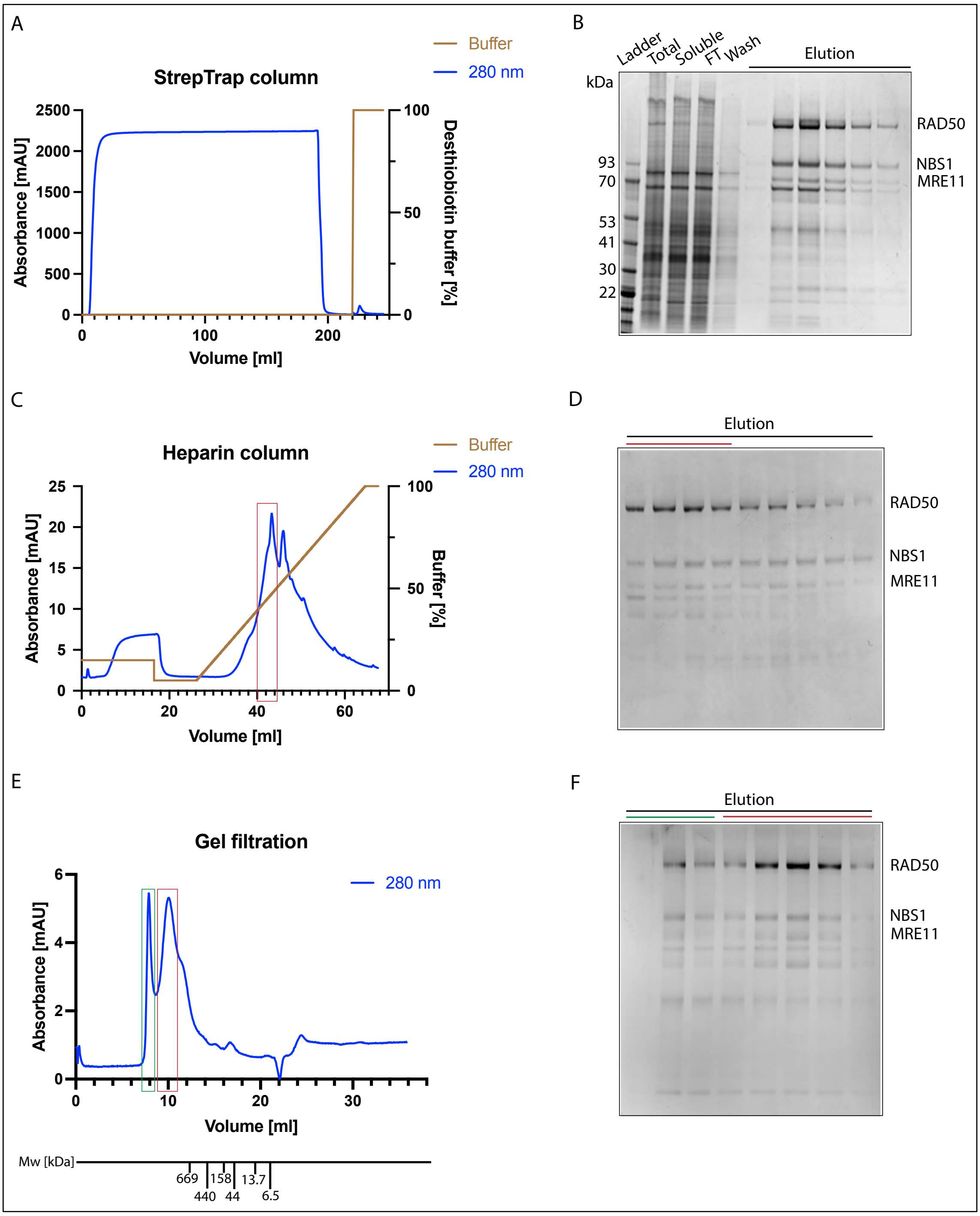


Figure 107. MRN purification from 500 ml of Hi5 cells. Elution profiles from A) StrepTrap HP, C) HiTrap Heparin HP and E) Superose 6 10/300 GL. B) Coomassie staining performed after standard cell lysis. Total – lysate before centrifugation; Soluble – soluble lysate after centrifugation; FT – flow through StrepTrap HP column; Wash – wash of the column; Elution – elution from the column. D) Coomassie staining of fractions eluted from HiTrap Heparin HP column. Red line on the gel corresponds to fractions indicated by the red box in (C) HiTrap Heparin HP elution profile. These fractions were combined and loaded on the GF column. F) Silver staining of fractions eluted from the gel filtration column. Green and red lines on the gel correspond to the fractions indicated by these colours in (E) elution profile from Superose 6 10/300 GL. E) Molecular weight of the proteins is based on standard protein resolution according to Cytiva.

7.1.3 Mass spectrometry analysis of purified MRN

During one of the purification trials, proteins concentrated after the Strep Tag affinity chromatography were resolved in a gel and sent to the mass spectrometry facility in the lab of our collaborator Dr J Choudhary (ICR). Analysis performed by Dr Lu Yu confirmed the presence all MRN subunits (Figure 108). Analysed bands contained also degraded components of the complex, heat shock proteins, tubulins, and other contaminants. A summary of this data is given in (Table 24).

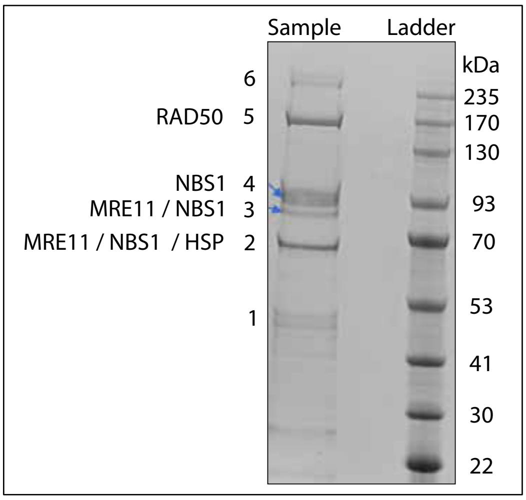


Figure 108. MRN purification. Coomassie staining of proteins concentrated after Strep Tag affinity chromatography. Indicated bands (1-6) were cut from the gel and sent for the mass spectrometry analysis. In the band 2, heat shock proteins (HSP) were present together with MRE11 and NBS.

7.1.4 Purification of MRN

hMRE11-TEV-10xHis, hRAD50-TEV-2xStrep, hNBS1- TEV-2xStrep

To increase the amount of purified MRN complex, I used pellet from 1L of infected Hi5 cells. The main purification steps were two types of affinity chromatography (performed using StrepTrap HP and HiTrap Heparin HP), followed by size exclusion chromatography (Superose 6 10/300 GL column). A detailed purification protocol is described in chapter (3.3.6).

To increase the yield of the purification, the Strep affinity step was performed twice (flow through and wash fractions were reloaded). To improve the purification yield, all fractions indicated by the red box in the HiTrap Heparin HP elution profile (Figure 109 B) were combined and loaded on the gel filtration column. The gel filtration elution profile indicates high amount of the protein aggregates as shown by the green box in the plot (Figure 109 C) and the green line in the Coomassie stained gel (Figure 109 D). Based on the standard protein resolution according to Cytiva, M_w of the fractions indicated by the red box (Figure 109 C) was over 669 kDa. As it was in proximity of M_w of the recombinant MRN complex dimer (658.4 kDa) these fractions were chosen for freezing. Two first fractions indicated by the red box (Figure 109 C) were flash frozen without concentrating (C ~ 0.09 mg / ml, 2 x 500 μl, A_{280nm} = 0.05). The other fractions included in the red box were combined and concentrated before freezing. Protein concentration was calculated based on the Beer-Lambert Law: A = ε*C*l, where A – absorbance, ε - molar absorption coefficient (M^{-1} cm^{-1}), C – molar concentration (M), l – optical path length (cm). For calculations of the MRN dimer concentration, the following values were used: M_w = 658.4 kDa, ε = 365480 M^{-1} cm^{-1}; hMRE11-TEV-10xHis (M_w = 82,8 kDa, ε = 37275 M^{-1} cm^{-1}), hRAD50-TEV-2xStrep (M_w = 157,7 kDa, ε = 72865 M^{-1} cm^{-1}) and hNBS1- TEV-2xStrep (M_w = 88,7 kDa, ε = 72600 M^{-1} cm^{-1}). The highest concentration of the MRN complex fraction eluted from the GF column was C = 0.14 μM; C = 0.09 mg/ml (A_{280nm} = 0.05), with the total yield of over 90 µg. Concentration of the MRN aggregates (Figure 109 C, first peak) was three times higher (A_{280nm} = 0.15).

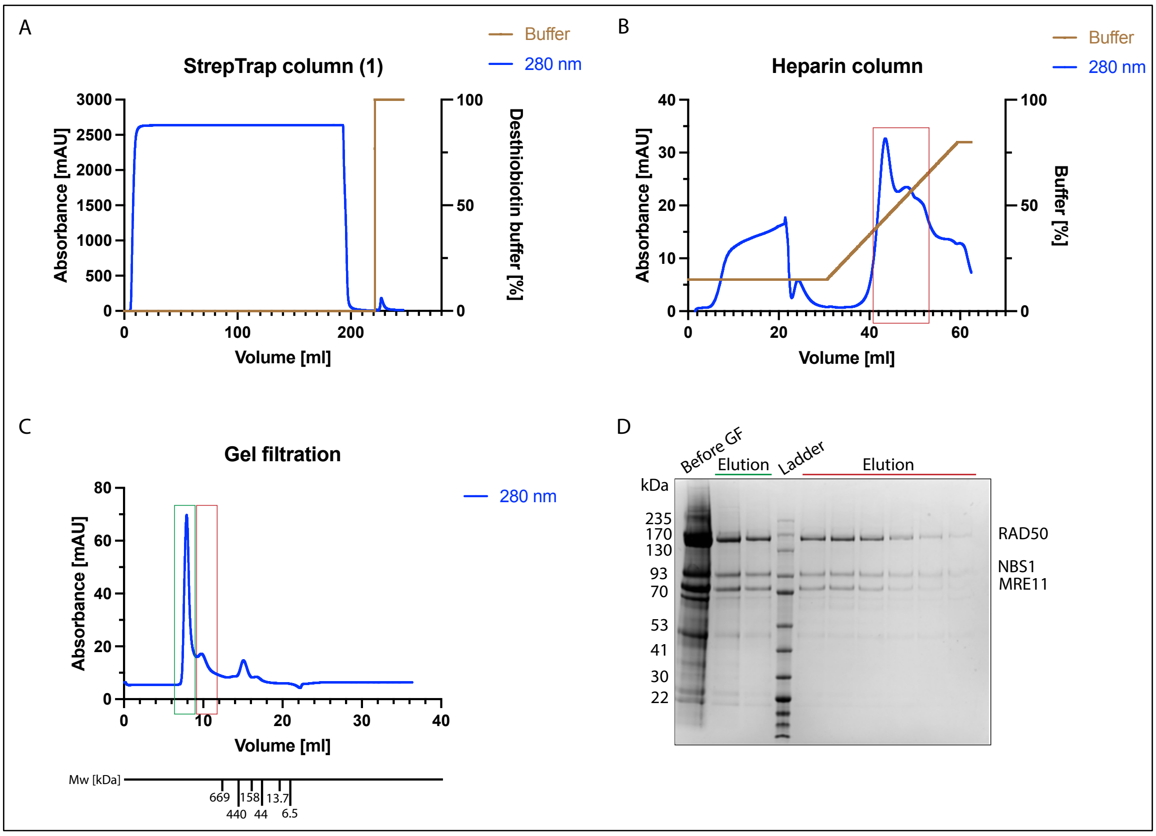


Figure 109. MRN purification from 1 L of Hi5 cells. Elution profiles from A) StrepTrap HP, B) HiTrap Heparin HP and C) Superose 6 10/300 GL. B) Red box indicates fractions eluted from the Heparin column (with ~ 375 nM – 650 nM NaCl) and loaded on the gel filtration column. D) Coomassie staining of fractions eluted from the gel filtration column. Green and red lines correspond to the labelling of the gel filtration elution profile. C) Molecular weight of eluted proteins is based on the standard protein resolution according to Cytiva. Fractions from the second peak (indicated by the red box) were frozen, as they were in the proximity of recombinant MRN complex dimer (658.4 kDa).

7.2 MR subcomplex (MRE11-RAD50) purification

7.2.1 Overview

Purification of MRN resulted in a low yield, with the majority of proteins eluted in the "aggregates" peak (Figure 109). According to (Myler et al., 2019), purification of the MR subcomplex has over six times higher yield than the MRN complex. Therefore, I decided to prepare a separate protocol for MR purification. To use MR in nuclease activity assays, I decided not to use benzonase during the purification.

7.2.2 Optimization of MR purification

hMRE11-TEV-10xHis, hRAD50-TEV-2xStrep

To purify MRE11 and RAD50 as a complex, Hi5 insect cells were transfected with one construct encoding both hMRE11-TEV-10xHis and hRAD50-TEV-2xStrep. As both proteins encode affinity tags, I decided to perform two steps of affinity chromatography. Pellet from 500 ml of infected Hi5 cells was lysed and divided into two batches. The first batch was purified using a StrepTrap affinity column, followed by HisTrap affinity step. The second batch was purified using HisTrap affinity followed by StrepTrap (Figure 110). After each purification step, both MR subunits were present in the eluted fractions. Silver staining of the eluted proteins suggested higher purification yield for the first batch (StrepTrap affinity column followed by HisTrap affinity step).

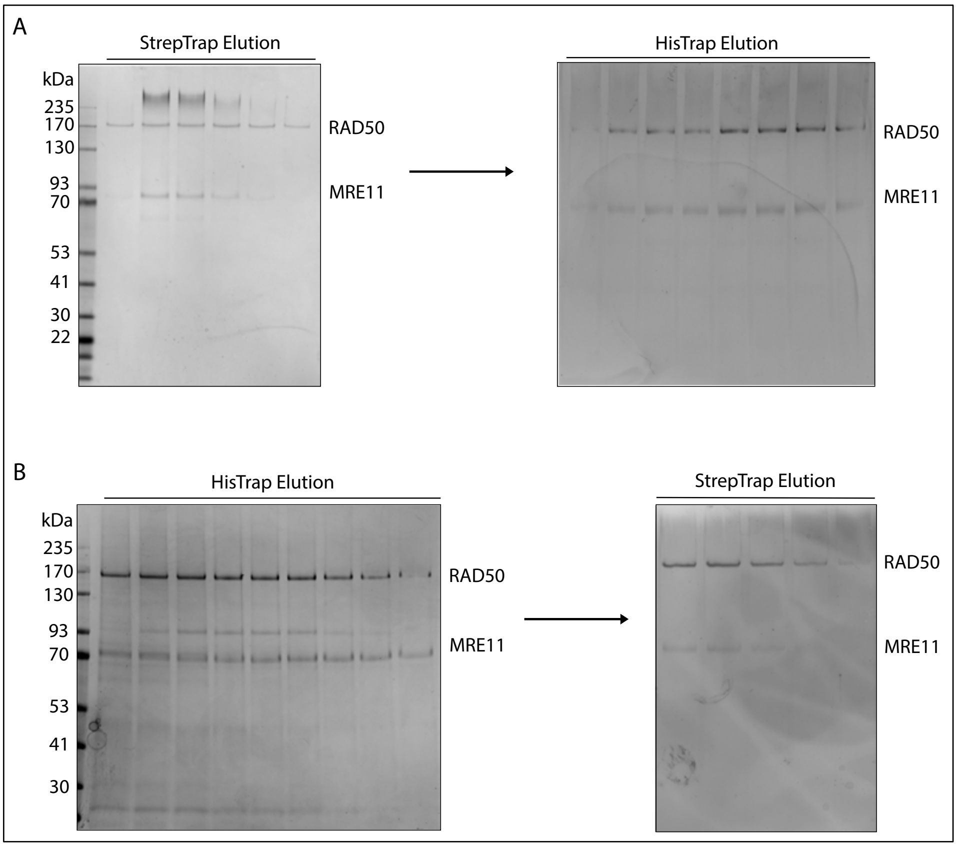


Figure 110. MR purification from 500 ml of Hi5 cells. After lysis, cells were divided into two batches. A) Purification using StrepTrap HP affinity column followed by HisTrap HP. Fractions (1.5 ml) eluted from the StrepTrap column were stained with Coomassie. Fractions (0.5 ml) eluted from the HisTrap column were silver stained. B) Purification using a HisTrap HP affinity column followed by StrepTrap HP. Fractions (1 ml) eluted from HisTrap column were stained with Coomassie. Fractions (0.5 ml) eluted from StrepTrap column were silver stained.

For large-scale protein purification, I used a StrepTrap HP column followed by HisTrap HP (Figure 111). Fractions indicated by the orange box in the HisTrap elution profile (Figure 111 C) corresponds to the fractions indicated by the orange line in the Coomassie stained gel (Figure 111 D). The high absorbance values at 254 nm (Figure 111 A, C) suggest high amount of contaminating DNA.

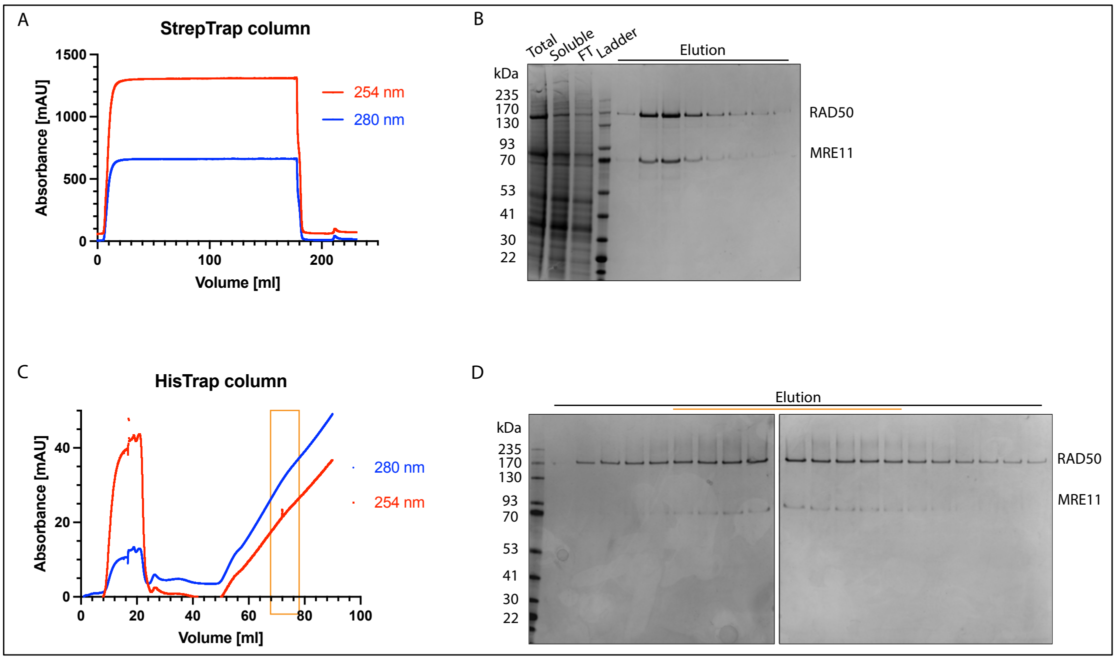


Figure 111. MR purification from 1L of Sf9 cells. Elution profiles from A) StrepTrap HP and C) HisTrap HP columns. High absorbance at 254 nm indicates high amount of contaminating DNA. B) Coomassie stained gel after first steps of purification. Total – lysate before centrifugation; Soluble – soluble lysate after centrifugation; FT – flow through StrepTrap HP column; Elution – elution from the StrepTrap HP column. D) Coomassie stained gel of fractions eluted from HisTrap HP column. Orange line on the gel corresponds to the fractions from the orange box on the (C) chromatogram, which were concentrated and frozen,

Purified MR was concentrated and flash frozen (A_{280nm} = 0.05; C = 0.23 μM; C = 0.110 mg/ml; 350 μl). The purification yield of the MR complex was very low (~38.5 μg), with high DNA contamination (A_{260nm}/A_{280nm} = 1). To increase purity of the proteins, ion exchange chromatography was introduced into the final protocol (7.2.3).

7.2.3 Purification of MR

hMRE11-TEV-10xHis, hRAD50-TEV-2xStrep

The MR complex (WT and MRE11 H129L D130V mutant) was purified from 1 L of Hi5 cells (Figure 112). The main steps were affinity chromatography (StrepTrap HP) followed by ion exchange (HiTrap Q HP). To increase the purification yield, HisTrap chromatography was not included. A detailed purification protocol is described in chapter (3.3.7).

MR complex was eluted from HiTrap Q HP column using increasing salt gradient. Fractions eluted with ~ 400 mM NaCl, indicated by the orange box in the protein purification plot (Figure 112 C) and by the orange line in the Coomassie stained gel (Figure 112 D), were concentrated and frozen. To calculate MR dimer concentration, the following values were used: M_w = 481 kDa, ε = 220280 M^{-1} cm^{-1}; hMRE11-TEV-10xHis (M_w = 82,8 kDa, ε = 37275 M^{-1} cm^{-1}) and hRAD50-TEV-2xStrep (M_w = 157,7 kDa, ε = 72865 M^{-1} cm^{-1}). MR mutant was frozen at the concentration C = 0.5 μM; C = 0.240

mg/ml (A_{280nm} = 0.11; A_{260nm}/A_{280nm} = 0.8; V = 490 μl). The total purification yield was ~ 118 μg. MR WT was frozen at concentration C = 0.73 μM; C = 0.350 mg/ml (A_{280nm} = 0.16; A_{260nm}/A_{280nm} = 0.9; V ~ 700 μl), with a total yield of ~ 245 μg. This is around 2.5 times higher than the purification yield of the MRN complex (Figure 109).

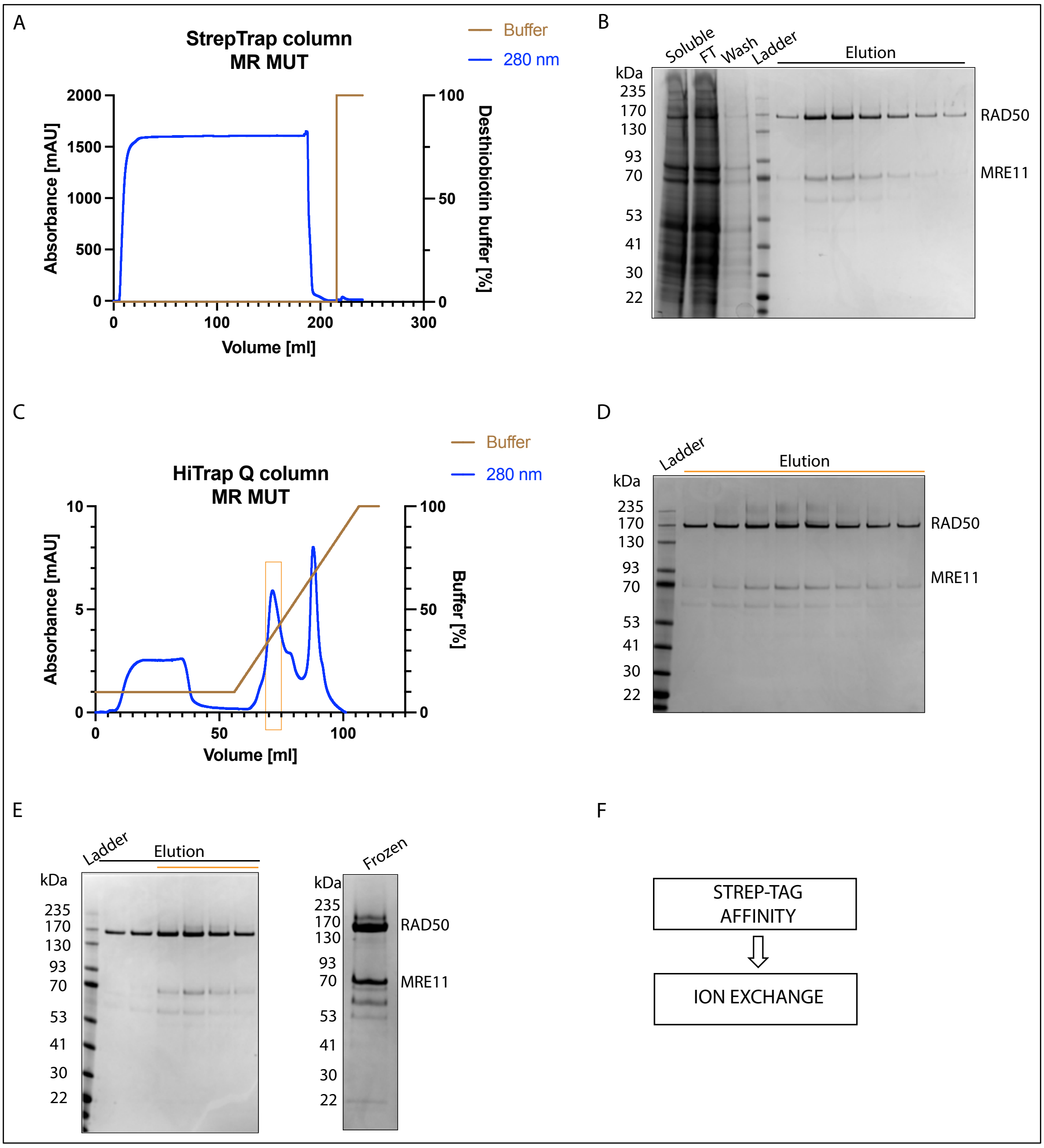


Figure 112. MR purification from 1 L of Hi5 cells. A) – D) Purification of M(H129L D130V)R mutant; E) Final results of MR WT purification. A) Elution profile from StrepTrap HP and C) HiTrap Q HP column. B) Coomassie staining after cell lysis. Soluble – soluble lysate after centrifugation; FT – flow through StrepTrap HP column; Wash – wash of the StrepTrap HP column; Elution – elution from the StrepTrap HP column. D) Coomassie staining of M(H129L D130V)R fractions eluted from HiTrap Q HP column. Concentrated fractions are marked with an orange line, corresponding to the orange box on the (C) chromatogram. E) Coomassie staining of MR WT fractions eluted from HiTrap Q HP column; Frozen – final sample after concentrating. F) Main steps of MR purification.

8. Discussion and future work

8.1 EXD2 purification - discussion

All protocols published so far (Broderick et al., 2016), (Silva et al., 2018), (Park et al., 2019), (Jia et al., 2022) describe purification of hEXD2 from bacteria. Due to its increased solubility and yield, the majority of labs purified a truncated version of the protein: i.e. (aa 61 – 621) for (Jia et al., 2022) and (Silva et al., 2018), (aa 76-564) for (Park et al., 2019). Our lab purified both, truncated (aa 76-564) and the full-length EXD2, which in the case of bacterially expressed protein behaved almost indistinguishably in *in vitro* nuclease activity assays (Broderick et al., 2016).

In this thesis, purification of hEXD2 from insect cells is described for the first time. I have optimized two purification protocols: one for truncated EXD2 (without aa 2-37) which has been tested in structural studies, and one for purification of the full-length protein used for biochemical assays. Yields of both purifications were satisfactory, with around 2.5 times higher yield for the truncated protein. Insect cells were used as an expression system, due to more accurate protein folding and a presence of post-translational modifications that are not performed in bacteria (Fernandez, J.M. & Hoeffler, J.P., eds., 1999). Using this expression system could lead to better quality and higher activity of the protein. This can explain why the full-length EXD2 that I have purified had high nuclease activity on dsDNA, whereas the full-length EXD2 from (Broderick et al., 2016) had very weak activity on dsDNA. Composition of buffers used during protein purification and in nuclease activity assays, may also affect EXD2 activity and can be a potential reason for observed differences.

8.2 EXD2 biochemical and structural characterization – discussion

8.2.1 'Clean' DNA ends

Consistent with results shown by (Broderick et al., 2016) full-length hEXD2 WT is a manganese-dependent 3'-5' exonuclease, whereas the (D108A, E110A) mutant is exonucleolytically inactive (Figure 32). This data is consistent with a crystal structure of the exonuclease domain of EXD2 (Park et al., 2019), where the highly conserved D108 and E110 residues coordinate the active site metal ion.

Interestingly, purified full-length hEXD2 exhibited higher nuclease activity on dsDNA than on ssDNA (Figure 36), which contradicts results generated using bacterially expressed hEXD2. Data from (Broderick et al., 2016) showed almost no activity on blunt ended dsDNA and a weak activity on DNA with a 5' overhang. The contrast between my data and those already published may result from using a different expression system (discussed in 8.1) and suggests the importance of utilising protein expression systems as closely related to human cells as possible.

Purified full-length hEXD2 exhibited high exonuclease activity on blunt ended dsDNA and dsDNA with either a 3’ or 5’ overhang (Figure 34). This is one of the key differences between EXD2 and MRE11, which exhibits extremely low activity on 3’ overhangs (both as MRE11 and MR), (Paull & Gellert, 1998). MRE11/MR efficiently digests dsDNA with blunt ends and ends with 5’ overhang (Paull & Gellert, 1998). This may suggest a different function of these proteins is some of the repair reactions that require resection of substrates with 3’ vs 5’ ssDNA ends i.e. regressed forks processing, Alt-EJ or BIR. To directly compare the activities of MRE11 and EXD2 in the future, reactions should be performed on substrates of the same sequence.

Artemis (together with DNA-PKcs) has endonuclease activity on both 3’ and 5’ overhangs of dsDNA (Chang et al., 2015). At 5’ overhangs it cuts at the ss/ds DNA boundary, but at 3’ overhangs it leaves a 4 nt overhang (Chang et al., 2015). Its activity on blunt ends depends on DNA sequence and may require presence of Ku heterodimer (Chang et al., 2015). A model described by (Shibata et al., 2017) proposed that EXD2 with other nucleases (MRE11, EXO1) performs resection of DSBs before the endonuclease activity of Artemis completes this process. Analysis of IR treated cells in G1 showed that EXD2 depleted cells have a milder phenotype (% of unrepaired DSBs) than cells deficient in MRE11 (Akagawa et al., 2020). It is tempting to speculate that MRE11 prepares the majority of substrates for digestion by Artemis, but not the 3’ overhangs which could be digested by EXD2. dsDNA with a 3’ overhang is not MRE11’s preferred substrate (Paull & Gellert, 1998), whereas EXD2 had similar preferences towards both blunt ended dsDNA and DNA with a 3’/5’ overhang (they are efficiently shortened by around the same length, ~ 35 nt), (Figure 34).

Purified full-length hEXD2 digested both single and double stranded DNA substrates as short as 20 nt (Figure 33). Analysis on 20, 40, 50 and 73 nt/bp long substrates showed that EXD2 had higher resection activity on the longer ones, creating the shortest products of ~ 12 nt (Figure 35). In my hands no successful EMSA analysis with EXD2 was performed (no shift was observed for any of the tested EXD2 concentrations, data not shown) hence I cannot compare binding preferences of EXD2 using DNA substrates of different lengths. Likewise, in EXD2 EMSAs performed by (Silva et al., 2018) no clear shift was observed with DNA, (the clear binding was shown with RNA substrates). DNA binding assays with purified yeast MR using 20, 31, 70, 100, 197 bp substrates, showed a preference towards longer DNA, with the lowest dissociation constant (K_D) for the 197 bp substrate (Kissling et al., 2022). At high protein concentrations, MR binds 20 bp substrate (Kissling et al., 2022).

8.2.2 Mismatched nucleotides and abasic DNA

DNA mismatches can be formed during DNA replication or HR and are repaired by MMR in an EXO1- dependent or independent pathway (Kadyrova et al., 2022). The EXO1 independent repair is promoted by DNA2 and relies on strand displacement DNA synthesis by Pol δ. According to (Kadyrova et al., 2022), EXO1 is the only eukaryotic exonuclease which can excise mismatched nucleotides in MMR. As DNA substrate with mismatched nucleotide was efficiently digested by EXD2 (Figure 38), it would be interesting to establish what is the role of EXD2 in the context of MMR. Hence, the *in vitro* activity of EXD2 and EXO1 on different types of mismatches should be compared in the future.

As EXD2 exhibited no activity on DNA with an abasic site (Figure 38), it is most likely not involved in the BER pathway.

8.2.3 3’ phosphorylated DNA ends

EXD2 efficiently digested 3’ phosphorylated DNA (Figure 37). This is in contrast to bacterially expressed EXD2, which had low activity on phosphorylated ssDNA (Silva et al., 2018). The observed difference may be caused by using different expression system, which may affect activity of the purified protein. DNA 3’ phosphate ends can be generated by IR and they have to be processed to enable DNA ligation (Zhelkovsky & McReynolds, 2014). Neocarzinostatin (NCS) generates DSBs with either a 3’-phosphate or 3’-phosphoglycolate, which *in vivo* are repaired by Artemis (Chang & Lieber, 2016). It would be interesting to test if EXD2 is required for the repair of NCS damage *in vivo*.

In bacteria, Exonuclease III can remove 3’ phosphates from dsDNA (Kuo, McRee, Cunningham, & Tainer, 1993). Apurinic/apyrimidinic endonuclease 1 (APE1), which is involved in the BER pathway, is a homolog of Exo III in mammals (Hadi & Wilson, 2000). APE1 exhibits 3’ phosphodiesterase and 3’ phosphatase activities which enables removal of 3’ phosphates (Thakur et al., 2014). Human polynucleotide kinase (hPNK), is also involved in BER and SSB/DSB repair, acting as DNA 3' phosphatase and 5' kinase enabling ligation of broken ends (Schellenberg & Williams, 2011). According to (Karimi-Busheri, Rasouli-Nia, Allalunis-Turner, & Weinfeld, 2007), human PNK takes part in the NHEJ pathway, but is not involved in HR. Indeed, I could not identify any DNA 3' phosphatase that has been shown to be a part of HR pathway.

A model described by (Syed & Tainer, 2018), proposed that HR resection starts with an endonucleolytically-generated nick several hundred bps from the DSBs with resection carried out towards the break. As IR-induced DSBs can be accompanied by other clustered lesions (Schipler & Iliakis, 2013), it may be possible that some lesions with 3' phosphate ends are present in the sequence between the MRE11 "nicking" site and the DSB. Hence, maybe EXD2 facilitates resection of these 3' phosphate lesions while digesting the strand towards the DSB. It is possible that EXD2 digests 3'-phosphates outside of the HR pathway, for example to enable restart of stalled replication forks. Under some circumstances, when PNK phosphatase is not functional, maybe it could also facilitate DNA ligation in the NHEJ.

8.2.4 EXD2 activity downstream of MRE11

The proposed digestion of 3' phosphates located between the MRE11 "nicking" site and the DSB is in line with another result, showing that EXD2 can initiate digestion from a DNA nick/gap (Figure 39). After endonucleolytic cleavage by MRE11 (MRN-CtIP), nicked DNA can be an entry site for EXD2 which together with MRE11 can perform resection towards the break, as in the model proposed by (Broderick et al., 2016). EXD2 exhibited the same activity on substrates with different size of gaps (1-, 3-, 5-, 10- nt wide), (Figure 40). MR can digests substrates with 5- and 10- nt gaps, but not the one with 15- nt gap (Deshpande et al., 2016).

Recent studies with yeast proteins, showed that MR forms nuclease active higher-order assemblies on DNA (Kissling et al., 2022). One could speculate that, after starting exonuclease digestion, the MRE11 complex may disassemble from this substrate. This may create a wide gap in the nucleotide sequence, which could be an entry site for EXD2. If any DNA damage / non-canonical structures lead to MRE11 disassembly, EXD2 may try to process them starting from this gap. Therefore, it would be interesting to test EXD2 activity on DNA substrates with an even wider gap.

The important question is what kind of substrates, which may block activity of MRE11 or other nucleases, can be digested by EXD2. Therefore, I focused on establishing if EXD2 is capable of digestion of any non-canonical structures.

8.2.5 Hairpins and cruciform-like structures

EXD2 activity on hairpins folded from the ssDNA, varied depending on the substrate sequence. A hairpin containing substrate (H40) with 6 bp long stem and 4 nt loop, which was mostly digested by Exo1 (Y. Li et al., 2019), was not resected by EXD2 (Figure 42). Interestingly, substrate '3A' with a 29T ssDNA followed by (AAATTT) was digested by EXD2 (Figure 42), but not by Exo1 (Y. Li et al., 2019). This shows that regarding digestion of these substrates, EXD2 and Exo1 complement each other. Hairpin YM164, which was cleaved by Artemis (S. Li et al., 2014), was processed by EXD2 only till the hairpin tip (Figure 43). The digestion pattern of this substrate differs between nucleases. This suggests that Artemis opens the hairpin and cleaves at ss/dsDNA junctions (S. Li et al., 2014), whereas EXD2 resected it exonucleolytically. As described, activity on these types of substrates differs between EXD2 and other nucleases. However, no evidence that supports formation of the secondary structures has been reported, hence it is difficult to conclude what is the exact conformation of these substrates in the reactions.

Interestingly, when one hairpin extruded from the dsDNA sequence, it was efficiently digested by EXD2 (Figure 45). These kind of structures can be formed during DNA replication (Mirkin, 2006). As EXD2 localizes to replication forks (Nieminuszczy et al., 2019), it is possible that resection of these structures is one of its role in this process. EXD2 exonucleolytically digested substrates predicted to form cruciform-like structures (Figure 50), which can also be formed during replication (Brázda et al., 2011). Interestingly, BRCA1, which physically and functionally interacts with EXD2 (Broderick et al., 2016), (Nieminuszczy et al., 2019), has cruciform binding properties (Brázda et al., 2011). Future *in vitro* reactions on long stable cruciform structures with EXD2 and BRCA1 can shed a light if BRCA1 enhances EXD2's activity in this process. Data from Inagaki and colleagues show that cruciform resolution can be performed by sequential cleavage, first by GEN1 then by Artemis (Inagaki et al., 2013). It would be interesting to test if GEN1 cleavage can create an entry point for EXD2-dependent nucleolytic processing of cruciform structures. If a product generated by GEN1 can be efficiently cleaved by EXD2, this may provide evidence that these two proteins cooperate to resolve DNA cruciforms.

One of the tested cruciform-like structures had fully complementary strands and was unstable, whereas another had a noncomplementary region forming two hairpins opposite to each other (Figure 48). Hence, these structures (which were both digested by EXD2) may also resemble partially melted dsDNA / small bubbles, not necessarily cruciform, as there was no evidence what is the exact conformation they adapt. They may mimic structures formed during DNA replication and transcription when strands from the DNA duplex are separated. It has been shown that EXD2 has a role in facilitating DNA replication (Nieminuszczy et al., 2019) and transcription (Sandoz et al., 2022). Therefore, it is reasonable to believe that EXD2's activity is regulated *in vivo,* i.e. it cleaves separated duplex strands only when DNA damage needs to be repaired or replication or transcription stress need to be resolved, not when they are a part of undamaged replication / transcription bubbles. DNA substrates with noncomplementary regions (~ 21 nt) within their dsDNA are the preferred substrate for WRN which efficiently degrades them in a 3' – 5' exonucleolytic direction (Machwe, Xiao, Theodore, & Orren, 2002). Therefore, it would be interesting to examine if EXD2 and WRN differ in their digestion preferences towards different bubble-containing substrates.

8.2.6 G-quadruplexes

EXD2 did not digest G-quadruplexes formed by MYC Pu22 sequence (Figure 53). However, sequences which may form telomeric G4s were resected by EXD2 (Figure 55). The digestion efficiency on these structures could be dependent on neighbouring sequences or on the total length of the substrate. A substrate with a total length of 70 nt containing a G4 forming sequence in the middle (extendTelom) was efficiently digested (Figure 54), whereas a 30 nt substrate comprising only a G rich sequence (TelG5) was partially resected (Figure 56). However, it is difficult to conclude if the analysed sequences formed G4s in the reactions. The nuclease activity assays were performed in a buffer required for EXD2 activity, resulting in decreased KCl concentration. This could affect stability of G4s, hence it is important to validate these results in the future, using higher amount of KCl. It would be beneficial to test EXD2 activity also on other sequences, forming more stable G4s than the telomeric one.

According to (Junyeop Lee et al., 2022), MRE11 digests this TelG5 sequence. Unfortunately, as no molecular weight marker is present in this study, it is difficult to analyse the extent of digestion by MRE11 and to compare it with EXD2. It is possible that the shortest product is of the same length for both nucleases. Interestingly, BRCA2 which binds to telomeric G4s, inhibits their digestion by MRE11 (Junyeop Lee et al., 2022). This binding promotes RAD51 mediated restart of the replication fork stalled at the telomeric G4s (Junyeop Lee et al., 2022). It is possible that binding of BRCA2 protects telomeres also from EXD2-mediated degradation, hence it would be interesting to test this experimentally.

EXD2 localizes to telomeres in cells that rely on alternative lengthening of telomeres (ALT) pathway (Broderick et al., 2023). EXD2 deficient ALT cells have increased number of ALT-associated PML bodies and C-circles (both hallmarks of ALT), (Broderick et al., 2023). Hence, it was proposed that EXD2 controls repair pathway choice at ALT telomeres as it promotes HR and supresses RAD52-independent BIR (Broderick et al., 2023). Digestion of sequences which may form telomeric G4s by EXD2 (Figure 54) is in line with (Broderick et al., 2023) as EXD2 may resect these telomeric G4 structures during the initial stages of HR. It is possible, that this processing is performed by coordinated action of both nucleases, MRE11 and EXD2. Hence, it would be interesting to test digestion of telomeric G4s when both enzymes are present in the reaction. Another nuclease that has also been shown to cleave telomeric G4s is DNA2 (Lin et al., 2013). It is tempting to speculate that MRE11 and EXD2 digest telomeric G4 during the short-range resection, whereas DNA2 activity follows during the long-range part of the HR.

As RAP1-interacting factor 1 (Rif1) binds to regions of the genome with predicted G4 motifs (Kanoh et al., 2015), and EXD2 interacts with RIF1 in cells (unpublished data from our lab, Caroline Clarke), it is tempting to speculate that EXD2 has a physiological role in processing of G4 structures not only at telomeres, but at other genomic loci. G4s can be visualised in human cells using structure specific antibodies (BG4, 1H6), (Hänsel-Hertsch et al., 2017), hence it would be interesting to observe if they colocalize with EXD2 at telomeres and other parts of DNA sequences. It can also be tested if amount of G4s increases in EXD2 depleted cells, both in the presence and absence of G4-stabilizing ligands (for example telomestatin). These experiments should validate if EXD2 has a physiological role in the processing of G4 structures.

8.2.7 Homologous recombination

In HR, short range resection is initiated by endonucleolytic cleavage by the MRE11 complex (MRN-CtIP), (Syed & Tainer, 2018). As discussed in (8.2.4), EXD2 can act downstream of MRE11, resecting the DNA from a nick or gap generated by MRE11 proximal to the break site. I propose that in HR, EXD2 engages mainly when DNA damage and non-canonical DNA structures are present between the MRE11-induced nicking sites and the DSB. This is in line with EXD2's ability to process 3' phosphorylated DNA, mismatches and different types of substrates that may form hairpin / cruciform-like structures. EXD2's digestion efficiency is much higher when blocks are present opposite to the digested strand (Figure 66). Hence, modifications and adducts present on DSBs, may stimulate EXD2's activity on the opposite strand.

The 70 bp DNA substrate with PS modifications on the 3' terminus (PC217*+PC216_5XSS), used in (Figure 67), was published by (Cannavo et al., 2019) to characterize MRX activity. They showed 'smear-like' degradation pattern generated by Mre11 alone and a weak exonuclease activity of MRX. Addition of pSae2 stimulated MRX and changed the digestion pattern (Cannavo et al., 2019). Interestingly, Mre11 is required as part of the MRX-pSae2 complex to perform digestion of the same efficiency/pattern as observed for EXD2. Both EXD2 and MRX-pSae2 generate ~ 10 nt long DNA when processing this substrate, with only a small number of middle-length products. This may suggest that on some substrates, EXD2 can rescue inefficient resection caused by lack of pSae2. Unfortunately, these experiments were performed on yeast Mre11 so it is difficult to conclude if the same result would be observed using human proteins. This can be addressed in future by utilising the purification protocols for human MRN that I have developed.

Experiments performed with yeast proteins show that MRX-Sae2 preferentially degrades the 5'- terminated strand of a blunt ended plasmid-length dsDNA, and that this reaction is stimulated by RPA (Cannavo et al., 2019). Unfortunately, I could not find if the same result was obtained for human MRN. Reactions carried out on a closed ssDNA substrate showed that the presence of RPA blocks cleavage performed by hMRN (Anand et al., 2019). Other reactions with hMRN and RPA were performed in the presence of additional nucleases (Nimonkar et al., 2011), hence it is difficult to conclude what is direct effect of RPA on hMRN activity in HR. This can be addressed in future using purified hMRN (developed protocol) and RPA purified by Dr Joanna Krwawicz.

I have also shown that RPA completely blocks EXD2 activity on ssDNA ends (Figure 87). This is in line with the observation that RPA binds ssDNA with very high affinity (Yates et al., 2018). On ssDNA, RPA inhibits the activity of other nucleases, including Exo1 (Y. Li et al., 2019), even though it stimulates EXO1-activity in long range resection during HR (Nimonkar et al., 2011). Therefore, I also tested how RPA effects EXD2 activity on a substrate where the dsDNA end is accessible for resection (a fork-like substrate with one nascent strand), (Figure 89). Here, RPA reduced EXD2 activity (most likely by binding the ssDNA part of the substrate), but EXD2 was still capable of performing digestion. In this experimental setup, RPA's function is opposite to its described role in promoting long-range resection during HR, where RPA enforces 5'-3' polarity of DNA2 and stimulates EXO1 activity (Nimonkar et al., 2011). Therefore, it seems more likely that EXD2 takes part in short-range resection and/or controlled processing of regressed replication forks. More research needs to be performed to validate a potential role of EXD2 in the further steps of HR. It should be tested if EXD2 can process short strand of dsDNA when a very long overhang is coated with RPA. Experiments in human cells, performed by (Broderick et al., 2016), showed that depletion of EXD2 affects both, short- and long- range resection. Knockdown of MRE11 had a similar but stronger phenotype (Broderick et al., 2016) even though MRE11 only stimulates the long-range resection by DNA2 or EXO1 (Nimonkar et al., 2011). Hence, it would be interesting to test if EXD2 can stimulate these long-range resection pathways *in vitro*. It should also be tested if RPA has the same impact on EXD2 and hMRE11 activities.

8.2.8 DNA replication

Results by (Nieminuszczy et al., 2019) showed that EXD2 promotes restart of reversed replication forks. Here, different types of substrates mimicking reversed replication forks were tested for their ability to be degraded by purified EXD2 *in vitro* but digestion when both reversed strands were blunt ended and fully complementary was not observed. I showed that purified full-length hEXD2 resected Holliday junctions with blunt ended complementary strands (Figure 51) in line with its proposed role in replication. The degradation was performed in 3' to 5' direction. The difference between my data and (Nieminuszczy et al., 2019) may result from using a different expression system for EXD2.

It has been shown that the combined activities of DNA2 and WRN lead to resection of reversed replication forks, promoting their restart (Thangavel et al., 2015). The presence of RPA stimulates the activity of DNA2, which carries out resection in the 5'-3' direction (Thangavel et al., 2015). One could speculate that if only one strand needs to be degraded, then controlled resection may be performed either by DNA2-WRN or EXD2, depending on the end accessibility. Perhaps EXD2 and DNA2-WRN compete, and the outcome depends on the presence of RPA which stimulates the activity of DNA2 on 4-

way junctions (Thangavel et al., 2015). Nuclease activity assays on several types of substrates showed that RPA either completely inhibited or partially reduced EXD2 activity (Figure 88), (Figure 89), (not tested on HJ). Possibly, if two regressed arms can be processed, EXD2 and DNA2 could act with different polarities at the same time. In these conditions, no long stretches of ssDNA should be formed so RPA would not inhibit EXD2 activity. To validate these assumptions, *in vitro* nuclease assays should be performed in the presence of both EXD2 and DNA2-WRN.

Nuclease activity assays on split arms, forks with one nascent strand and forks with two nascent strands showed high EXD2 digestion efficiency on all the substrates (Figure 76), (Figure 77). The most prominent difference between the results I have obtained and (Nieminuszczy et al., 2019) was on fork-like structures with two nascent strands, which was not efficiently resected by hEXD2 expressed in bacteria (Nieminuszczy et al., 2019). The presence of a 3' block on the fork template strand did not affect rate of the nascent strand digestion (Figure 82). The nascent strand was accessible for EXD2-dependent digestion when both template strands were present, and it was digested first as the preferred part of the substrate. As discussed in (8.2.5), EXD2 can digests substrates which may form hairpins extruded from one strand of dsDNA and cruciform-like structures. Possibly, the role of EXD2 in DNA replication is not limited to the resection of reversed forks, but may extend to the digestion of non-canonical DNA structures formed in the nascent strand. Importantly, EXD2 could perform nascent strand digestion also in close to physiological conditions when the single stranded template is coated with RPA (Figure 89). Even though the presence of RPA decreased EXD2's resection efficiency in this assay, the nascent strand was gradually digested.

8.2.9 Balance between the ions

EXD2 is manganese-dependent nuclease, and its activity was decreased when magnesium was added to the reaction (Figure 74). Mg can be copurified with a protein, hence its addition to the reaction buffer may result in a too high final concentration. It is also possible that the balance between Mg and Mn regulates EXD2 activity not only *in vitro*. It has been shown that nuclear-encoded EXD2 acts both in mitochondria (Silva et al., 2018) and it in the nucleus (Broderick et al., 2016), (Shibata et al., 2017), (Nieminuszczy et al., 2019), (Sandoz et al., 2022). Most intracellular Mn is present in nuclei, rather than mitochondria (Kalia, Jiang, & Zheng, 2008). According to (Killilea & Killilea, 2022), Mg concentration in mitochondria is 0.4 – 0.7 mM, whereas Mn concentration is 3 – 16 μM. I can speculate that the balance between Mg/Mn ions may allow EXD2 to be active either at specific location or time, but currently there is no data to support this notion.

8.2.10 HNH domain

According to (Silva et al., 2018), EXD2 has a highly conserved HNH endonuclease-like domain ‘lacking catalytic DH residues required for activity in other HNH endonucleases’. Here, I found similarities between EXD2 and HNH endonucleases (EcoKMcrA, TagI), which have crucial H residues that are required for their activity (not DH), (Figure 59). These H residues are strictly conserved between EXD2 and these nucleases, highlighting their importance. Another strictly conserved motif is tandem CXXC, which mutated in EXD2 (CXXC > SXXS) results in lower nuclease activity but not in complete inhibition (Silva et al., 2018). Here, I mutated H residues (H442 and H463) of EXD2 and showed that this purified mutant exhibited extremely low activity on ssDNA and complete lack of activity on dsDNA (Figure 63). It is interesting, as WT EXD2 exhibits higher activity on dsDNA than on ssDNA. As mutations of catalytic residues from the HNH endonuclease domain changes EXD2 activity on ssDNA and completely blocks activity on dsDNA, it suggests that maybe EXD2 is not only an exonuclease, but also an endonuclease. A H442 EXD2 mutation was found in a sample of large intestine carcinoma (COSMIC catalogue). Analysis by (Silva et al., 2018) showed that EXD2’s HNH-like domain is required for efficient mitochondrial translation, and it influences stable binding of EXD2 to ssRNA. Comparison of AlphaFold prediction with preliminary 3D structure of EXD2, by Dr Fabienne Beuron (Figure 104), indicated that the HNH domain may localize close to a tunnel which could be an entry site for DNA substrates.

All these data highlight the importance of EXD2’s HNH domain, even though its exact role in EXD2 activity is not fully understood. Therefore, EXD2 should be analysed not only in the context of its exonuclease domain, but as the full-length protein. It is essential to generate a cell line with EXD2 (H442A, H463A) mutation and analyse if expression of the mutant corrects phenotypes of EXD2 deficient cells (sensitization to DNA damage and impaired formation of RPA foci).

Nuclease activity assays with EXD2 (D108A E110A) exonuclease-dead mutant, showed either a complete lack of activity or some weak signal at the very bottom of the gel that was treated as a background. However, this signal can correspond to products 1 – 2 nt long and could represent weak endonucleolytic cleavage at the very end of the 5’ terminus. An excess of (D108A E110A) exonuclease mutant in the reactions lead to increased amounts of this very short product (Figure 73). It was also observed in the reactions with WT EXD2, but not with the (H442A, H463A) mutant. At this point it is difficult to conclude if this signal is of any endonucleolytic importance, as it can be background related to instability of the substrate. To validate the results, purification of the endonuclease mutant should also be repeated. It would be interesting to see if the mutants can rescue the resection by acting together in the reaction where they are both mixed.

8.2.11 Endonuclease activity

EXD2 WT exhibited endonuclease activity, when incubated with closed circular ssDNA for an extended amount of time (Figure 57). This is in line with the importance of its HNH domain, discussed in (8.2.10). The observed potential endonuclease activity is very weak, as it requires hours, not minutes, to be observed. It is possible that EXD2 requires another protein to stimulate this cleavage. It would be interesting to examine if (phosphorylated) CtiP can be a co-factor of EXD2 endonuclease activity, as it stimulates endonuclease cleavage by MRN complex (Anand et al., 2016) and it interacts with EXD2 (Broderick et al., 2016). This can be carried out in future using purified human MRN, which I have optimised protocols for, and purified CtIP (Anand et al., 2018).

It is possible that EXD2's endonuclease activity is substrate-dependent. In line with this idea, DNA substrates with blocks located opposite to the labelled strand, was digested with a distinct pattern (Figure 67). If the blocks were located at the very end of the substrate, no products of intermediate length were observed, suggesting that these blocks either stimulate EXD2 exonuclease activity or that EXD2 nicks the strand opposite to the block. No distinct cleavage was observed on dsDNA with blocks on both 3' ends (Figure 72). However, the amount of substrate decreased over time indicating possible digestion. A reaction with substrates having PS blocks on both strands also showed what could be interpreted as weak endonuclease activity (Figure 68).

In my opinion EXD2 is an exonuclease with a weak endonuclease activity. The main question is whether the endonuclease activity is physiologically relevant, as it is extremely weak. Therefore, I think it is essential to analyse cells with EXD2 endonuclease mutations (discussed in 8.2.10) and to test potential EXD2 cofactors (such as pCtIP) *in vitro*. As EXD2 exhibits higher activity on RNA than on DNA (Silva et al., 2018), it would be interesting to test endonuclease activity on an RNA substrate. It is possible, that its endonuclease role is physiologically relevant only in the context of RNA.

8.2.12 Cellular localization

It was proposed that EXD2 localizes both in mitochondria (Silva et al., 2018), and in the nucleus (Broderick et al., 2016), (Shibata et al., 2017), (Nieminuszczy et al., 2019), (Sandoz et al., 2022). Here, I showed that EXD2 encodes a putative nuclear localization signal (NLS), (Figure 96), which supports experimentally derived data by (Nieminuszczy et al., 2019).

It is unclear how EXD2 distribution is regulated. DNA2, which acts at both mitochondria and nuclei, controls this localization *via* post-translational modifications (Zheng et al., 2020). hDNA2 does not encode an NLS and its translocation to the nucleus depends on TRAF6-mediated polyubiquitination (Zheng et al., 2020). To test if EXD2 migrates to the nucleus through an NLS-dependent mechanism, it would be required to mutate the predicted NLS sequence in cells (~ 335 position of aa) and observe how it affects EXD2 localization. The MTS (aa 1-61) is located at the N-terminal end of EXD2 (Silva et al., 2018), whereas the predicted NLS is between exonuclease and HNH domains. Perhaps a controlled degradation, causing cleavage of the N-terminus encoding both MTS and a transmembrane domain, could be a mechanism underlying EXD2 distribution in cells.

8.2.13 EXD2 structure

A preliminary 3D structure of full-length EXD2, by Dr Fabienne Beuron, showed that EXD2 monomers interact mostly through their exonuclease domains (Figure 100). This is in line with data by (Park et al., 2019), who shows that EXD2's exonuclease domain has a chimeric active site. Based on our preliminary data and AlphaFold prediction, it seems that the exonuclease and HNH domains are located close to a tunnel, which could be an entry site for nucleic acids (Figure 104). This structure has resolution of ~ 30 Å. Hence, to obtain higher resolution data, the purification and negative staining of EXD2 needs to be further optimized.

8.3 Reconstitution of the resection - discussion

To fully understand the role of EXD2, it is important to purify other proteins and reconstitute the short-range resection pathway *in vitro* (MRN, CtIP, RPA, EXD2). As an initial step towards this goal, I have optimized purification protocols for both the MR and MRN complex.

As the MRN complex has been the object of research since the nineties, several purification protocols have already been published. Some of these articles (Anand et al., 2018), (Joon Lee & Dunphy, 2013), describe purification of the complex without performing a gel filtration step. However, to separate the MRN complex from the aggregates and single subunits, size exclusion chromatography is essential (Figure 109). The MRE11, RAD50 and NBS1 proteins form variety of complexes with different subunit composition and stoichiometry (Van der linden, Sanchez, Kinoshita, Kanaar, & Wyman, 2009). Therefore, it is important to obtain MRN as a complex, as the variable subunit composition may be relevant to MRN function. Introduction of a gel filtration step is described in (Myler et al., 2019), (J. H. Lee & Paull, 2006) and (Van der linden et al., 2009). However, the purification yield of the complex is low. In (Myler et al., 2019) 80 µg of the complex was obtained from 60x15-cm dishes of Sf21 cells, in (J. H. Lee & Paull, 2006) 100 µg of purified protein was obtained from 1.4 L of insect cells. The MRN purification I performed, resulted in over 90 µg of the complex from 1 L of Hi5 cells. Even though it is a low purification yield, it is higher than in the published protocols. It is possible that the increased yield is the result of DNA digestion by benzonase used in the lysis step. However, as benzonase has nuclease activity, this MRN purification protocol should not be performed if proteins are to be used in nuclease activity assays. As contaminating DNA is digested, this protocol is recommended for proteins dedicated for co-immunoprecipitation (co-IP) analysis or structural studies. It should prevent DNA mediated interactions, allowing validation of direct interactions between proteins. MRE11-RAD50, required for nuclease activity assays, can be purified using the MR protocol which does not involve the use of benzonase. The yield of this purification is over 2.5 higher than the MRN, allowing for more efficient protein production.

8.4 Concluding statements

This work for the first time describes purification of the full-length and truncated hEXD2 from insect cells. The protein folding and posttranslational modifications in this expression system are more related to the human cells, than those performed in bacteria. Hence, activity of full-length hEXD2 purified according to this protocol can more closely resemble activity of EXD2 expressed in human cells.

Purified full-length hEXD2 exhibits high nuclease activity on substrates resembling damaged DNA (with mismatched nucleotide or 3’ phosphorylated DNA end) and on substrates, which are likely to form non-canonical DNA structures (hairpins, cruciform-like structures, telomeric G4s). However, folding of these structures should be additionally confirmed. Presence of a block opposite to the digested strand, stimulates EXD2 activity. EXD2 can also start resection from a nick and DNA gaps of different size. Therefore, I propose that EXD2 digests damaged DNA and non-canonical DNA structures during HR short range resection process. Possibly it can also facilitate DNA replication not only by digestion of reversed replication forks but also by digestion of structures inhibiting replication fork progression. Presence of RPA inhibits EXD2 activity on ssDNA, but only partially decreases its activity on dsDNA ends. Hence, EXD2 can perform digestion in close to physiological conditions. It is important to validate if the observed activity is relevant *in vivo*.

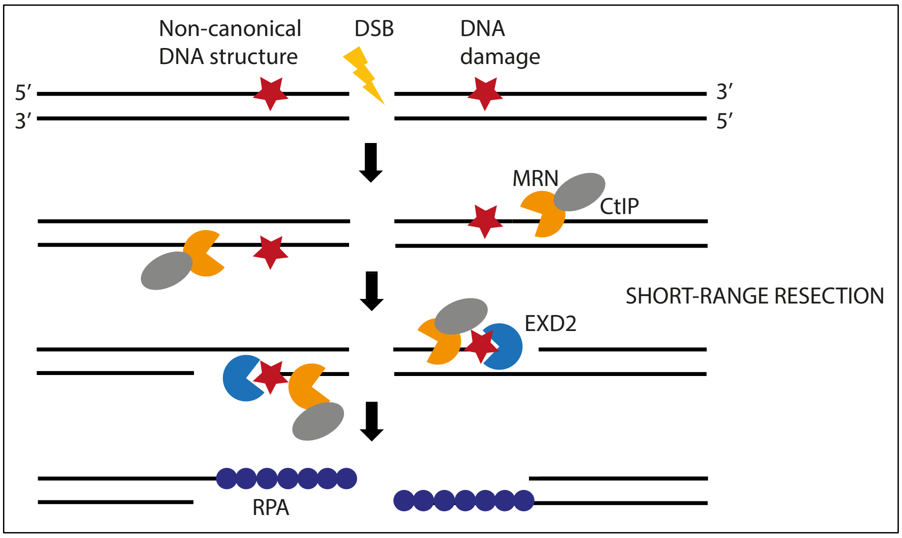


Figure 113. Proposed model of EXD2 activity in HR short-range resection process. After MRN-CtIP initiate resection, EXD2 may exonucleolytically digest damaged DNA and non-canonical DNA structures located in the proximity of DSBs. This may enable generation of ssDNA overhangs, which are coated by RPA.

To summarise, this work shows that full-length hEXD2 purified from insect cells:

- digests both ss and dsDNA, both blunt ended and with overhangs
- digests damaged DNA
- digests non-canonical DNA structures (their folding should be additionally confirmed in the future)
- exhibits distinguished digestion pattern when a block is present on a substrate
- is inhibited by RPA on single stranded ends, but still active when double stranded ends are present
- has weak endonuclease activity
- requires HNH domain for exonuclease activity.

Supplementary figures

Table 20. Primers used in the thesis, as described in the Materials and Methods section. Sequences starting from 'Cas' and 'S1' are from (Cutts & Vannini, 2018).

Cas I F	aacgctctatggtctaaagatttAAATCGACCTACTCCGGAATATTAATAGATC;
Cas I R	aaacgtgcaatagtatccagtttATTTAAATGGTTATGATAGTTATTGCTCAGCG;
Cas II F	aaactggatactattgcacgtttAAATCGACCTACTCCG GAATATTAATAGATC;
Cas omega R	aaccccgattgagatatagatttATTTAAATGGTTATGA TAGTTATTGCTCAGCG;
1F tagless Nbs1	GCGTCGTCGTTAAAAGCTTGTCGAGAAGTACTAGAGGATCA;
1R tagless Nbs1	CGACAAGCTTTTAACGACGACGCTTCAGATACGGGTTGTAAC;
Cas II R	aaacatcaggcatcattaggtttATTTAAATGGTTATGATAGTTATTGCTCAGCG;
Cas III F	aaacctaatgatgcctgatgtttAAATCGACCTACTCCG GAATATTAATAGATC;
Fstrep3C.1	ggaagtgctgttccagggtccgTCATGGAGCCATCCGCAATTTGAAAAAGG;
RtevEXD2.1	atggtgtgccattggatccGCGCCCGATGGTGGGACGGTATGAATAATCC;
FhisEXD2.2	GGATCCAATGGCACACCATCACCATCAC;
RhisEXD2.2	gaccctggaacagcacttccagGCTCAGCTGGATCGGCAGGTCTTC;
1FEXTRANS	TCACTGCATTCTAGTTGTGGTTTGTCCAA;
2REXTRANS	GTTGCTGGGTCACCGGGCCTTGGAAGTACAGATTTTCATGGTGATGGTG;
3FEXTRANS	TTCCAAGGCCCGGTGACCCAGCAACCG;
4REXTRANS	TTGGACAAACCACAACTAGAATGCAGTGA;
1SEQTREXD2	CGAGCATCGTTTGTTCGC;
2SEQTREXD2	GATTATTCATACCGTCCCACC;
3SEQTREXD2	GGTGCTGGGTATTGATTGC;
4SEQTREXD2	ACCCTGACCGAAGATCAAG;
5SEQTREXD2	TGGTGAAGGAAAACCTGTG;
6SEQTREXD2	CCCGTATTAGCAACGAGAAC;
S1A	CGAGGCTAACTGAAACACGG;
S1B	CCCGTAGTGGCTATGGCAG;
S1C	GCGAGATGGTTATCATTTTAATTATCTCC;
S1D	CCTCCCACACCTCCCCCTG;
1MRN for	AGACAGAATAAAACGCACGG;
2MRN for	ACTTCATTCTGCTGGGTGG;
3MRN for	CGTCTGTATCGTATGTTCGTG;
4MRN for	TTTAGCGTGCTGCGTTTC;
5MRN for	AAATGGCGAACGATAGCG;
6MRN for	CAGCCATACCACATTTGTAGAG;
7MRN for	GAAGATGAGCATTCTGGGTG;
8MRN for	CAGGAAGATAGCAACTGGC;
9MRN for	GCTGAACCAGGAAAAGAGC;
10MRN for	ATGCTGACCAAAGACAAGG;
11MRN for	GGATAAACTGAAGAGCACCG;
12MRN for	CGAAAGAACAAGTGAGCCC;
13MRN for	GGAAGAGAAGTATCGTGAGATG;
14MRN for	TTTGAAAAAGGTGGCGGC;

16MRN for	GTTTCTGAAGGCGGTTGAG;
17MRN for	AGAAGTTTCGTATGCTGAGC;
18MRN for	CAGAAGCAAGAAGAGGACG;
2MRN rev	TGCTCAGTTTCTTGTGCTG ;
NbsF1seq	TTTGTTCGCCCAGGACTC;
NbsF2seq	TTGGTCGTAAGAACTGCG;
NbsF3seq	TTTCTGAAGGCGGTTGAG;
NbsF4seq	TCTGAAAACCACCACCCC;
NbsF5seq	CAAAGAGCAACACCTGAG;
NbsF6seq	TGCTGCCGAAGAAACTGC;
HisEXD2strepass emb1F	CCTGGAAACCCGTATTAG;
HisEXD2strepass emb1R	ATGTTTCAGGTTCAGGGGG;
1F HisEXD2Strep	AGGCGAGCATCGTTTGTTC;
2F HisEXD2Strep	GCAGCAAAAAGTTCTGGGTAG;
3F HisEXD2Strep	CTGCTGGGTTATCCGTTTAG;
4F HisEXD2Strep	TGGTGAAGGAAAACCTGTG:
5F HisEXD2Strep	ACTGCTGCGTAAGTTTGGC;

Table 21. DNA sequences used as substrates in nuclease activity assays, as described in the Results section.

BS40	ATAAATATTTTTTATTAATAATAGATCACCTTTCTTTCTCTTCTCCCCTT;
UA1	AAGGGGAGAAGAGAAAGAAAGGTGATCTATTATTAATAAAAAATATTTAT;
shortBS40	TTTCTTTCTCTTCTCCCCTT;
last40BS40	TTTATTAATAATAGATCACCTTTCTTTCTCTTCTCCCCTT;
longBS40	GGCACCCGCACCGCGGCACTTACATAAATATTTTTTATTAATAATAGATCACC TTTCTTTCTCTTCTCCCCTT;
UA1_60	AGTCTGGTAAAAGGGGAGAAGAGAAAGAAAGGTGATCTATTATTAATAAAAAA TATTTAT;
BS40_10	ATAAATATTTTTTATTAATAATAGATCACCTTTCTTTCTCTTCTCCCCTTAGTCT GGTAA;
20 nt dsDNA	shortBS40 + Reverse_shortBS40 Reverse_shortBS40: AAGGGGAGAAGAGAAAGAAA;
40 nt dsDNA	last40BS40 + Reverse_last40BS40 Reverse_last40BS40: AAGGGGAGAAGAGAAAGAAAGGTGATCTATTATTAATAAA;
73 nt dsDNA	longBS40 + ReverselongBS40: ReverselongBS40: AAGGGGAGAAGAGAAAGAAAGGTGATCTATTATTAATAAAAAATATTTATGTA AGTGCCGCGGTGCGGGTGCC;
PhosphoBS40	ATAAATATTTTTTATTAATAATAGATCACCTTTCTTTCTCTTCTCCCCTT/3Phos/;
BS401	ATAAATATTTTTTATTAATAATAGATCACC;
BS401ab	ATAAATATTTTTTATTAATAATAGA/idSp/CACC;
BS401mm	ATAAATATTTTTTATTAATAATAGAGCACC;
Blunt ended dsDNA with a nick	BS401*: ATAAATATTTTTTATTAATAATAGATCACC; BS402: TTTCTTTCTCTTCTCCCCTT; UA1: AAGGGGAGAAGAGAAAGAAAGGTGATCTATTATTAATAAAAAATATTTAT;
dsDNA with 5’ overhang and a nick	BS401* + BS402 + UA1_60
Blunt ended dsDNA with a gap	BS401*: ATAAATATTTTTTATTAATAATAGATCACC; BS403: TTCTTTCTCTTCTCCCCTT; UA1:AAGGGGAGAAGAGAAAGAAAGGTGATCTATTATTAATAAAAAATATTTAT;
dsDNA with 5’ overhang and a gap	BS401* + BS403 + UA1_60
Blocked dsDNA with a gap	BS401* + BS403thio + UA1 BS403thio: TTCTTTCTCTTCTC*C*C*C*T*T;
Blocked dsDNA with a nick	BS401* + BS402thio + UA1 BS402thio: TTTCTTTCTCTTCTC*C*C*C*T*T;

dsDNA with a nick and a double block	BS401* +BS402thio + UA1thio UA1thio: AAGGGGAGAAGAGAAAGAAAGGTGATCTATTATTAATAAAAAATA*T*T*T*A*T;
Blocked DNA with 1 nt gap	BS401* + BS403thio + UA1
Blocked DNA with 3 nt gap	BS401* + BS402thio3 + UA1 BS402thio3: CTTTCTCTTCTC*C*C*C*T*T;
Blocked DNA with 5 nt gap	BS401* + BS402thio5 + UA1 BS402thio5: TTCTCTTCTC*C*C*C*T*T;
Blocked DNA with 10 nt gap	BS401* + BS402thio10 + UA1 BS402thio10: TTCTC*C*C*C*T*T;
Hairpin 3A	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTAAATTT;
Hairpin H40	TTGATAAGAGGTCATTTGAATTCATGGCTTAGAGCTTAAT;
YM164	TTTTTGATTACTACGGTAGTAGCTACGTAGCTACTACCGTAGTAAT;
PC217	GATGCATGAGGTGGAGTACGCGCCCGGGGAGCCCAAGGGCACGCCCTGGCACCCGCACCGCGGCACTTAC;
Hairpin_PC217	GATGCATGAGGTGGAGTACGCGCCCGGGGAGCCCAAGGGCGCCCTGGCACCCGCACCGCGGCACTTAC;
Hairpin_compl216	GTAAGTGCCGCGGTGCGGGTGCCATTGGGCTCCCCGGGCGCGTACTCCACCTCATGCATC;
Ds_hairpin with a block'	topblockhairpinPC217* + Hairpin_compl216 topblockhairpinPC217: GATGCATGAGGTGGAGTACGCGCCCGGGGAGCCCAAGGG*C*GCCCTGGCACCCGCACCGCGGCACTTAC;
Cruciform_PC217	GTAAGTGCCGCGGTGCGGGTGCCAGGGCGCCCTTGGGCTCCCCGGGCGCGTACTCCACCTCATGCATC;
StableCru_PC217	GTAAGTGCCGCGGTGCGGGTGCCCAACTAGTTGTGGGCTCCCCGGGCGCGTACTCCACCTCATGCATC;
Holliday junction	X12-3 HJ3* + Block TOPL + Block HJ1 + Block HJ2 X12-3 HJ3: GAGATCTATCTGGTGCCTTCTGACAGTGAATGGGTAACGAATCGTAATAGTCTCTAGACAGCATGTCCTAGCAATGTAATCGTCTATGACGTC; Block TOPL: GACGTCATAGACGATTACATTGCTAGGACATGCTGTCTAGAGACTATCGCGACTTACGTTCCATCGCTAGGTTATTTTTTTTTTTTTT*T*T*T*T*T; Block HJ1: AAAAAAAAAAAAAAAAAAATAACCTAGCGATGGAACGTAAGTCGCGATGGGCTTAACTAGGATGCTACTGGCCCCGAATCAACCGTAC*T*T*G*G*G; Block HJ2: CCCAAGTACGGTTGATTCGGGGCCAGTAGCATCCTAGTTAAGCCCATTACGATTCGTTACCCATTCACTGTCAGAAGGCACCAGATAG*A*T*C*T*C;
ExtendMyc (extendMycPu22)	GAGGATCCTTAATTAAGGTCTCAGGTATGAGGGTGGGTAGGGTGGGTAAGAGACGGCGGCCGCCTGCAGG;

ExtendTelom	AGAGGATCCTTAATTAAGGTCTCATTAGGGTTAGGGTTAGGGTTAGGGAGAGACGGCGGCCGCCTGCAGG;
Mutated extendMyc	GAGGATCCTTAATTAAGGTCTCAGGTATGAGAGTGAGTAGAGTGAGTAAGAGACGGCGGCCGCCTGCAGG;
Mutated extendTelom (MutTelomeric)	AGAGGATCCTTAATTAAGGTCTCATTAGAGTTAGAGTTAGAGTTAGAGAGAGACGGCGGCCGCCTGCAGG;
TelG5	GGTTAGGGTTAGGGTTAGGGTTAGGGTTAG;
TelG5-G3	GGT TAG GGT TAG GGT TAG TGT TAG GGT TAG;
PC216_non	GTAAGTGCCGCGGTGCGGGTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCGTACTCCACCTCATGCATC;
PC216_5XSS	GTAAGTGCCGCGGTGCGGGTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCGTACTCCACCTCAT*G*C*A*T*C;
PC217middle	GATGCATGAGGTGGAGTACGCGCCCGGGGAGCCCAAGGGCACGCCCTGGCACCCGC*A*C*C*G*CGGCACTTAC;
BS40Thio	ATAAATATTTTTTATTAATAATAGATCACCTTTCTTTCTCTTCTCCCC*T*T;
PC216Xafter40	GTAAGTGCCGCGGTGCGGGTGCCAGGGCGTGCCCTTGGGCT*C*C*C*C*GGGCGCGTACTCCACCTCATGCATC;
PC216near3	GTAAGTGCCGCGGTGCGGGTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCG*C*G*T*A*CTCCACCTCATGCATC;
PC216middle	GTAAGTGCCGCGGTGCGGG*T*G*C*C*AGGGCGTGCCCTTGGGCTCCCCGGGCGCGTACTCCACCTCATGCATC;
PC216_doubleblock	GTAAGTGCCGCGGTGCGGGTGCCAGGGCGTGCCCT*T*G*G*G*CTCCCCGGGCGCGTACTCCACCTCAT*G*C*A*T*C;
BS40bio	ATAAATATTTTTTATTAATAATAGATCACCTTTCTTTCTCTTCTCCCCTT/3Bio/;
UA1bio	AAGGGGAGAAGAGAAAGAAAGGTGATCTATTATTAATAAAAAATATTTAT/3Bio/
DNA structure with a nascent leading strand	BS40* + BS41 + UA2 BS40: ATAAATATTTTTTATTAATAATAGA TCACCTTTCTTTCTCTTCTCCCCTT; BS41: TTCCCCTCCTCTCCTTCCTTCCTGATCTATTATTAATAAAAAATATTTAT; UA2: AAGGGGAGAAGAGAAAGAAAGG;
DNA structure with a nascent lagging strand	BS40* + BS41 + UA6 UA6: GGAAGGAAGGAGAGGAGGGGAA;
DNA structure resembling a full fork	BS40* + BS41 + UA2 + UA6
DNA structure with split arms	BS40* + BS41
Fluorescent fork (BS41)	BS40_FAM + UA2-Q + BS41 BS40_FAM: ATAAATATTTTTTATTAATAATAGATCACCTTTCTTTCTCTTCTCCTCTT/36-FAM/; UA2-Q: /5IABkFQ/AAGAGGAGAAGAGAAAGAAAGG;

Fluorescent fork (BS41) with blocks on 3’ end and near 3’ end	BS40_FAM + UA2-Q + BS41last+double BS41last+double: TTCCCCTCCTCTCCTTCCTTCCTGATCTA TT*A*T*TAATAAAAAATAT*T*T*A*T;
Fluorescent fork (BS41) with blocks on 3’ end and at the junction	BS40_FAM + UA2-Q + BS41last+junction BS41last+junction: TTCCCCTCCTCTCCTTCCTTCC*T*G*ATCTATTATTAATAAAAAATA T*T*T*A*T;
Fluorescent fork (BS41) with blocks on 3’ end and near 5’ end	BS40_FAM + UA2-Q + BS41last+free BS41last+free: TTCCCCTCCTCTC*C*T*TCCTTCCTGATCTATTATTAATAAAAAATA T*T*T*A*T;
Fluorescent fork with a middle block (BS41MiddleTIO)	BS40_FAM + UA2-Q + BS41Mtio BS41Mtio: TTCCCCTCCTCTCCTTC*C*T*T*C*CTGATCTATTATTAATAAAAAATATTTAT;
Fluorescent fork with a 5’ block (BS41TIO)	BS40_FAM + UA2-Q + BS41tio BS41tio: T*T*C*C*C*CTCCTCTCCTTCCTTCCTGATCTATTATTAATAAAAAATATTTAT;
Fluorescent fork where fluorophore is with a block	BS40-blockFAM + UA2-Q + BS41last+junction BS40blockFAM: ATAAATATTTTTTATTAATAATAGATCACCTTTCTTTCTCTTCTCCCC*T*T*/36-FAM/
Split arms	X12-3 HJ3* + HJ2 X12-3 HJ3: GAGATCTATCTGGTGCCTTCTGACAGTGAATGGGTAACGAATCGTAATAGTCTCTAGACAGCATGTCCTAGCAATGTAATCGTCTATGACGTC; X12-3 HJ2: CCCAAGTACGGTTGATTCGGGGCCAGTAGCATCCTAGTTAAGCCCATTACGATTCGTTACCCATTCACTGTCAGAAGGCACCAGATAGATCTC;
dsDNA with 3’ overhang	X12-3 HJ3* + Greyhalf HJ2 X12-3 HJ3: GAGATCTATCTGGTGCCTTCTGACAGTGAATGGGTAACGAATCGTAATAGTCTCTAGACAGCATGTCCTAGCAATGTAATCGTCTATGACGTC; GreyhalfHJ2: ATTACGATTCGTTACCCATTCACTGTCAGAAGGCACCAGATAGATCTC;
Forks with one nascent strand and blocks	halfTOPLred* + Block HJ2 + Block HJ3 halfTOPLred: GACGTCATAGACGATTACATTGCTAGGACATGCTGTCTAGAGACT; Block HJ2: CCCAAGTACGGTTGATTCGGGGCCAGTAGCATCCTAGTTAAGCCCATTACGATTCGTTACCCATTCACTGTCAGAAGGCACCAGATAG*A*T*C*T*C;

	Block HJ3: GAGATCTATCTGGTGCCTTCTGACAGTGAATGGGTAACGAATCGTAATAGTCTCTAGACAGCATGTCCTAGCAATGTAATCGTCTATG*A*C*G*T*C;
Split arms with a block	X12-3 HJ3* + Block HJ2 X12-3 HJ3: GAGATCTATCTGGTGCCTTCTGACAGTGAATGGGTAACGAATCGTAATAGTCTCTAGACAGCATGTCCTAGCAATGTAATCGTCTATGACGTC; Block HJ2: CCCAAGTACGGTTGATTCGGGGCCAGTAGCATCCTAGTTAAGCCCATTACGATTCGTTACCCATTCACTGTCAGAAGGCACCAGATAG*A*T*C*T*C;

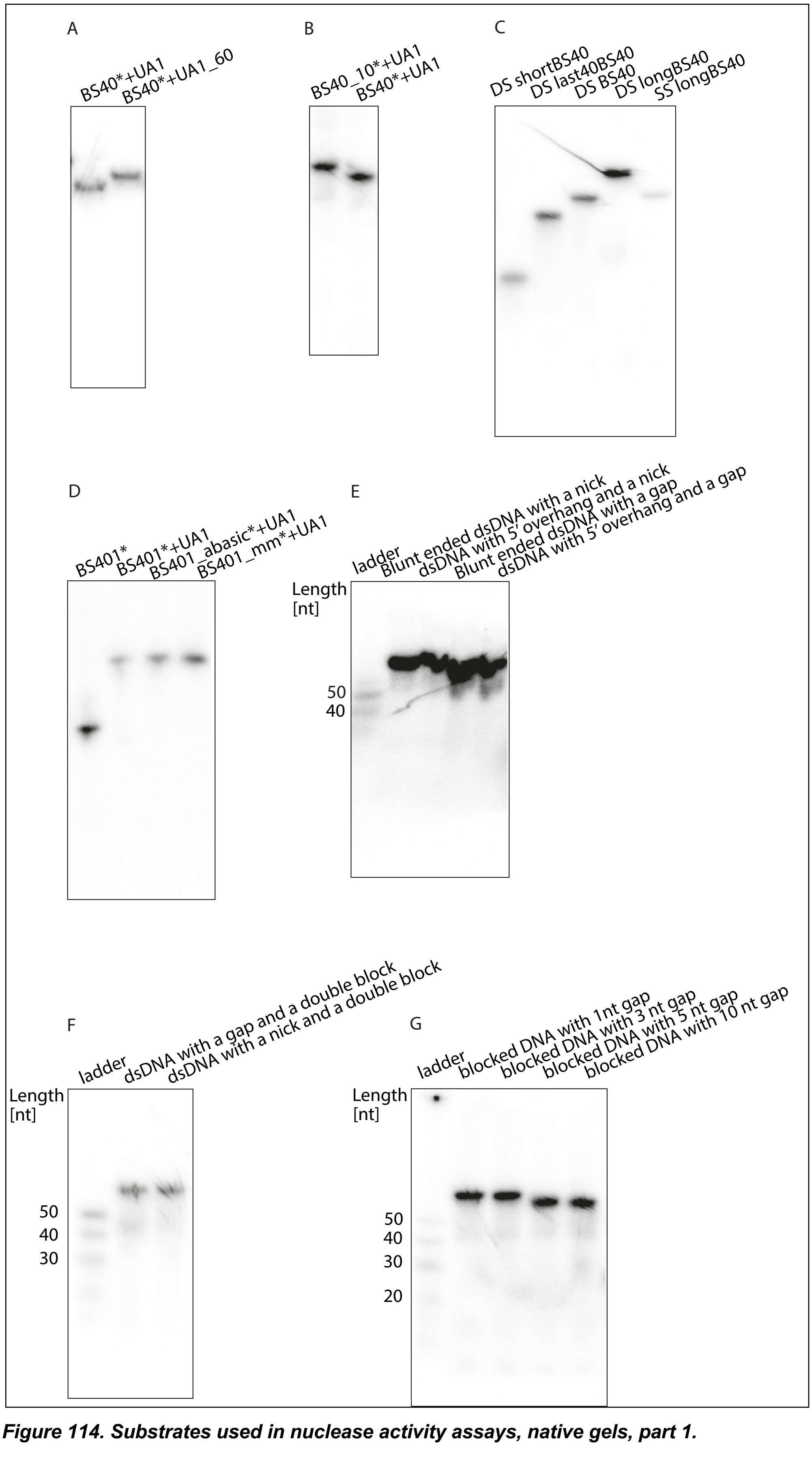


Figure 114. Substrates used in nuclease activity assays, native gels, part 1.

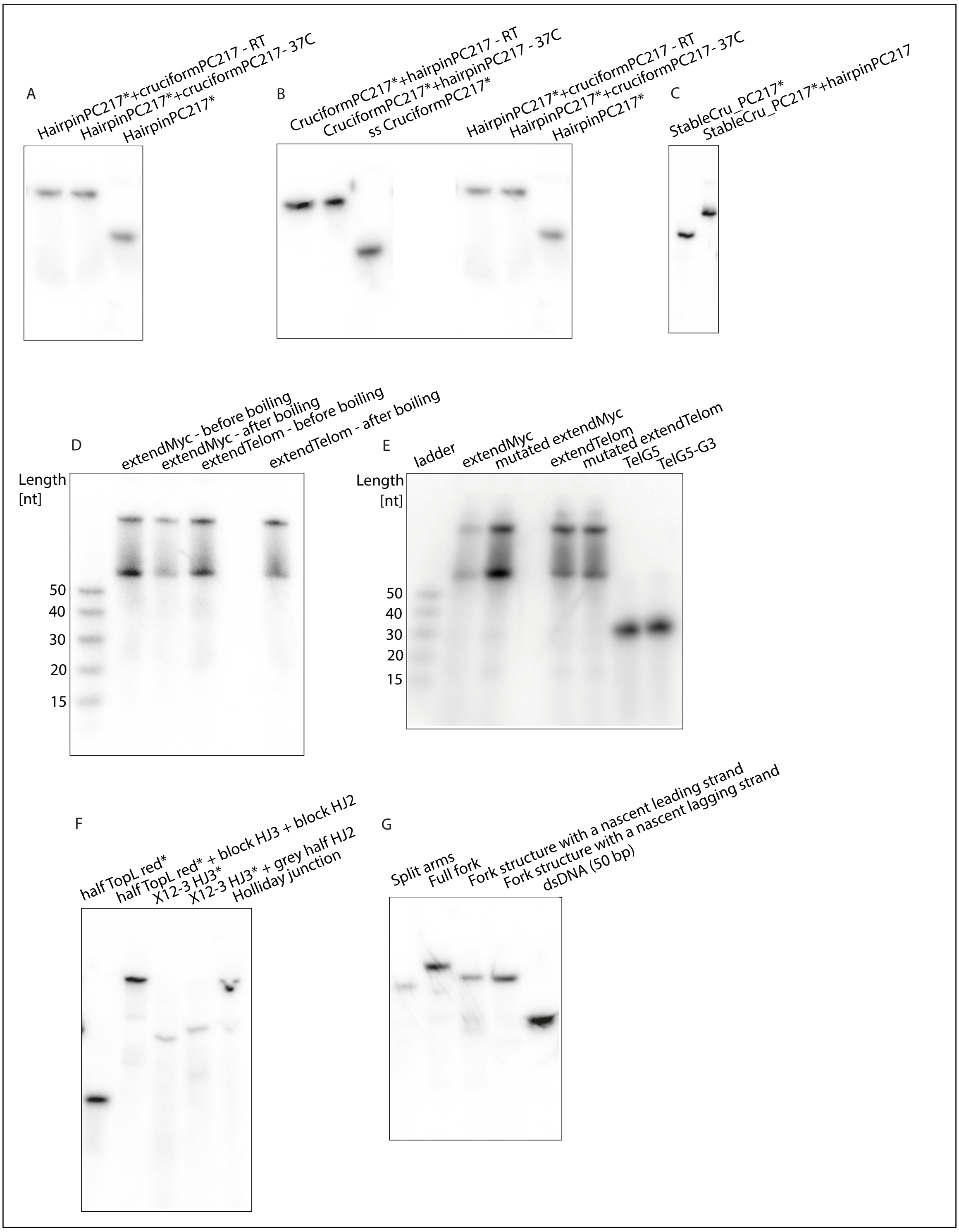


Figure 115. Substrates used in nuclease activity assays, native gels, part 2.

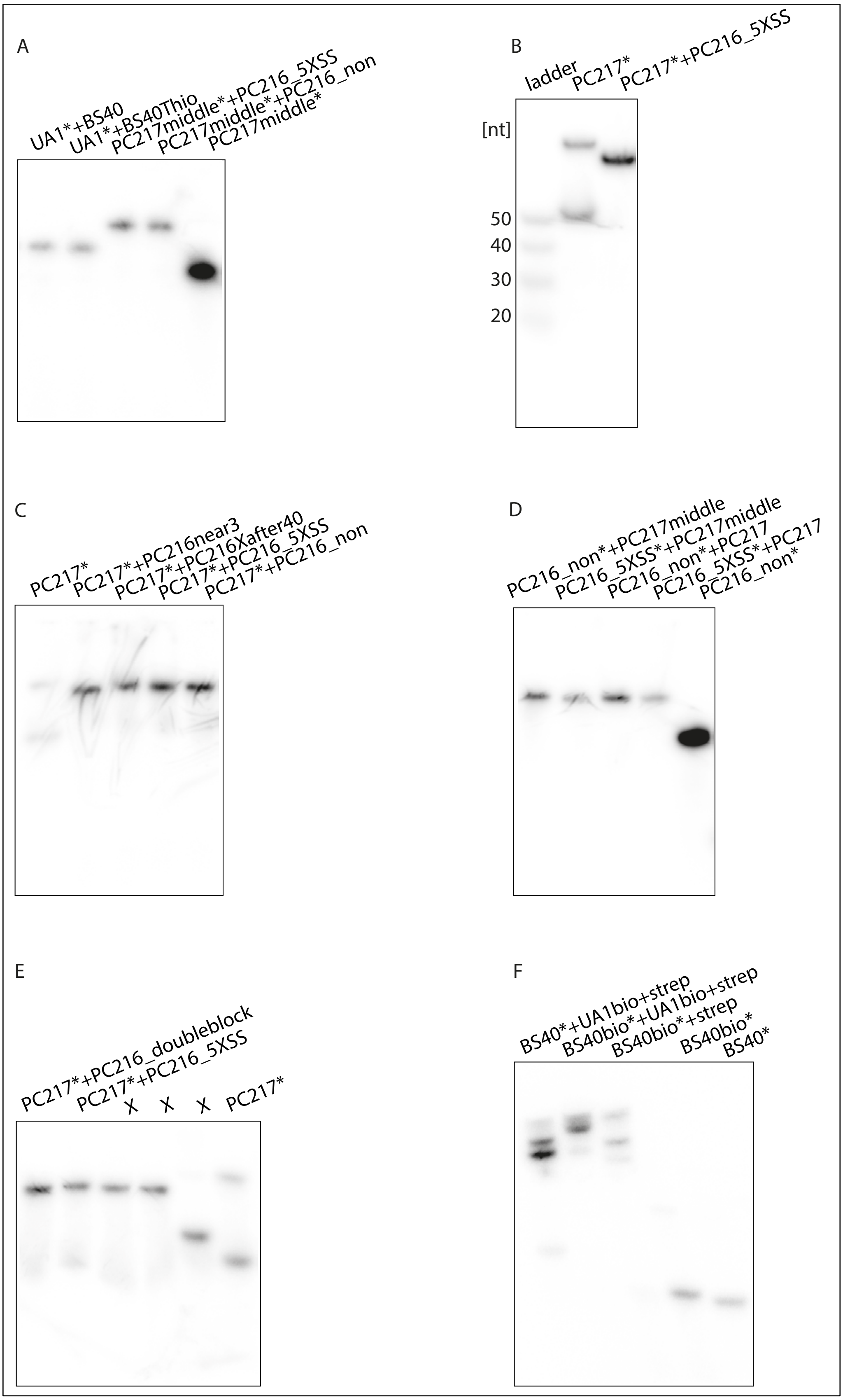


Figure 116. Substrates used in nuclease activity assays, native gels, part 3.

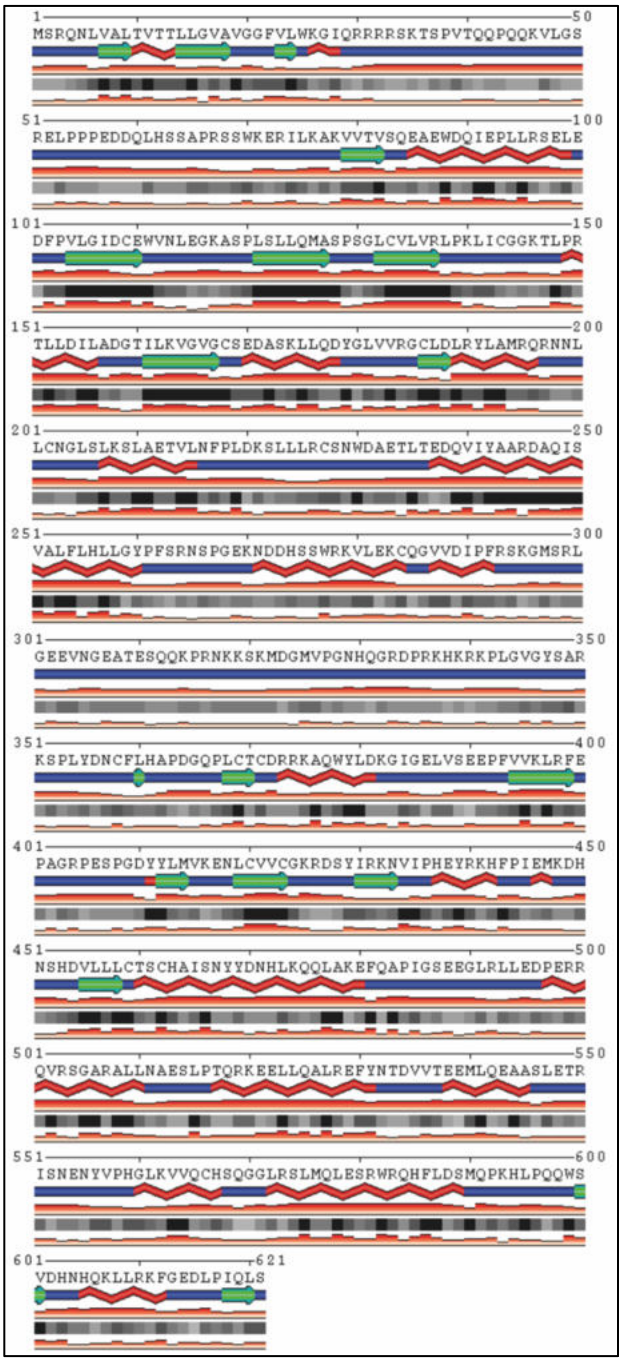


Figure 117. Secondary structure prediction of full-length EXD2; *https://sable.cchmc.org/****. Red - helices, green - B strand, blue - coiled coil.***

Table 22. Mass spectrometry analysis of purified EXD2 WT, by Dr Lu Yu. Table shows 300 from 628 hits.

Accession	Description	Hi5	Sum PEP Score	Coverage [%]	# Peptides	# Unique Peptides	# AAs	MW [kDa]	Score Sequest HT
Q9NVH0	Exonuclease 3'-5' domain-containing protein 2 OS=Homo sapiens OX=9606 GN=EXD2 PE=1 SV=2		1354.828	89	79	79	621	70.3	8067.26
P04350	Tubulin beta-4A chain OS=Homo sapiens OX=9606 GN=TUBB4A PE=1 SV=2		313.498	59	24	0	444	49.6	1227.97
P68371	Tubulin beta-4B chain OS=Homo sapiens OX=9606 GN=TUBB4B PE=1 SV=1		280.456	62	25	0	445	49.8	1145.57
Q9BVA1	Tubulin beta-2B chain OS=Homo sapiens OX=9606 GN=TUBB2B PE=1 SV=1		261.674	59	27	2	445	49.9	1118.23
P07437	Tubulin beta chain OS=Homo sapiens OX=9606 GN=TUBB PE=1 SV=2		265.65	59	24	1	444	49.6	1117.09
Q9BUF5	Tubulin beta-6 chain OS=Homo sapiens OX=9606 GN=TUBB6 PE=1 SV=1		104.828	30	12	1	446	49.8	520.47
Q3ZCM7	Tubulin beta-8 chain OS=Homo sapiens OX=9606 GN=TUBB8 PE=1 SV=2		100.431	23	12	3	444	49.7	505.79
Q71U36	Tubulin alpha-1A chain OS=Homo sapiens OX=9606 GN=TUBA1A PE=1 SV=1		199.961	47	19	1	451	50.1	445.63
Q9NY65	Tubulin alpha-8 chain OS=Homo sapiens OX=9606 GN=TUBA8 PE=1 SV=1		174.263	38	15	1	449	50.1	359.77
P54652	Heat shock-related 70 kDa protein 2 OS=Homo sapiens OX=9606 GN=HSPA2 PE=1 SV=1		106.674	23	13	1	639	70	268.82
P0DMV9	Heat shock 70 kDa protein 1B OS=Homo sapiens OX=9606 GN=HSPA1B PE=1 SV=1		50.831	8	4	1	641	70	68.79
P15924	Desmoplakin OS=Homo sapiens OX=9606 GN=DSP PE=1 SV=3		29.183	3	4	4	2871	331.6	36.73
Q71UI9	Histone H2A.V OS=Homo sapiens OX=9606 GN=H2AZ2 PE=1 SV=3		19.987	35	3	1	128	13.5	31.43
P63267	Actin, gamma-enteric smooth muscle OS=Homo sapiens OX=9606 GN=ACTG2 PE=1 SV=1		8.963	11	3	1	376	41.9	19.5
P00367	Glutamate dehydrogenase 1, mitochondrial OS=Homo sapiens OX=9606 GN=GLUD1 PE=1 SV=2		12.662	9	3	3	558	61.4	24.3
P14923	Junction plakoglobin OS=Homo sapiens OX=9606 GN=JUP PE=1 SV=3		5.51	4	3	3	745	81.7	9.57
Q92547	DNA topoisomerase 2-binding protein 1 OS=Homo sapiens OX=9606 GN=TOPBP1 PE=1 SV=3		9.067	3	2	2	1522	170.6	10.48
A0A7E5VJV7	Tubulin beta chain OS=Trichoplusia ni OX=7111 GN=LOC113494444 PE=3 SV=1	Yes	537	77	39	5	447	50.2	2028.14
A0A7E5VE36	Tubulin beta chain OS=Trichoplusia ni OX=7111 GN=LOC113492972 PE=3 SV=1	Yes	323.426	55	24	1	443	49.5	1325.47
A0A7E5WY23	Heat shock protein 70 OS=Trichoplusia ni OX=7111 GN=LOC113506423 PE=2 SV=1	Yes	552.818	70	43	30	654	71.6	1405.59
A0A7E5VJS6	Tubulin beta chain OS=Trichoplusia ni OX=7111 GN=LOC113494447 PE=3 SV=1	Yes	276.312	59	29	11	457	51.4	881.64
A0A7E5X115	Tubulin beta chain OS=Trichoplusia ni OX=7111 GN=LOC113507665 PE=3 SV=1	Yes	244.155	43	20	1	447	49.9	904.84
A0A7E5X5Y1	Tubulin beta chain OS=Trichoplusia ni OX=7111 GN=LOC113508948 PE=3 SV=1	Yes	308.419	61	25	13	451	50.7	813.14
A0A2H4WW89	Tubulin alpha chain OS=Trichoplusia ni OX=7111 GN=LOC113497986 PE=2 SV=1	Yes	301.388	74	29	5	450	49.9	805.63
A0A7E5WLV3	Tubulin alpha chain OS=Trichoplusia ni OX=7111 GN=LOC113503819 PE=3 SV=1	Yes	253.481	51	21	1	450	49.8	673.26
A0A7E5V8B9	Tubulin beta chain OS=Trichoplusia ni OX=7111 GN=LOC113491657 PE=3 SV=1	Yes	74.974	18	11	1	450	50.7	308.34
A0A7E5WTY1	heat shock 70 kDa protein cognate 5 OS=Trichoplusia ni OX=7111 GN=LOC113505692 PE=3 SV=1	Yes	328.29	57	32	32	686	74.7	450.7
A0A7E5WX33	ATP synthase subunit beta OS=Trichoplusia ni OX=7111 GN=LOC113506412 PE=3 SV=1	Yes	293.773	75	29	24	515	54.9	395.46
A0A7E5VSK5	Heat shock protein 83 OS=Trichoplusia ni OX=7111 GN=LOC113496266 PE=3 SV=1	Yes	237.422	49	31	30	722	83.1	335.68
Q0MUU6	Heat shock cognate 70 protein OS=Trichoplusia ni OX=7111 GN=LOC113494618 PE=3 SV=1	Yes	194.289	60	34	29	659	73.1	287.3
A0A7E5WIQ6	Asparagine--tRNA ligase OS=Trichoplusia ni OX=7111 GN=LOC113503013 PE=3 SV=1	Yes	212.018	38	19	19	554	63	324.44
A0A7E5V8N1	Calcium-transporting ATPase OS=Trichoplusia ni OX=7111 GN=LOC113491725 PE=3 SV=1	Yes	247.643	34	33	33	1026	112.7	326.94
A0A7E5VBE7	ADP/ATP translocase OS=Trichoplusia ni OX=7111 GN=LOC113492335 PE=3 SV=1	Yes	143.36	62	26	26	300	32.7	249.52
A0A7E5WH63	constitutive coactivator of PPAR-gamma-like protein 1 isoform X1 OS=Trichoplusia ni OX=7111 GN=LO	Yes	238.38	24	16	16	962	105.6	276.04
A0A7E5WHV4	Vesicle-fusing ATPase OS=Trichoplusia ni OX=7111 GN=LOC113502529 PE=3 SV=1	Yes	177.295	50	28	28	746	82.5	217.64
A0A7E5WAV9	dnaJ homolog subfamily A member 2-like OS=Trichoplusia ni OX=7111 GN=LOC113501015 PE=3 SV=1	Yes	124.027	50	13	13	403	44.9	216.05
A0A7E5W8U8	dnaJ homolog subfamily A member 1 OS=Trichoplusia ni OX=7111 GN=LOC113500490 PE=3 SV=1	Yes	143.94	47	18	18	404	45.3	194.89
A0A7E5WZM0	ATP synthase subunit alpha OS=Trichoplusia ni OX=7111 GN=LOC113507229 PE=3 SV=1	Yes	116.378	40	19	19	557	59.9	169.75
A0A7E5WI92	Glyoxylate reductase 1 homolog OS=Trichoplusia ni OX=7111 GN=LOC113502935 PE=3 SV=1	Yes	131.543	48	14	14	527	57.3	187.41
A0A7E5WP10	endoplasmic reticulum resident protein 44 isoform X2 OS=Trichoplusia ni OX=7111 GN=LOC113504396	Yes	110.881	38	16	16	505	58.6	149.89
A0A7E5WRU4	Coatomer subunit delta OS=Trichoplusia ni OX=7111 GN=LOC113505136 PE=3 SV=1	Yes	92.419	39	15	15	505	56.1	139.88
A0A7E5WWN3	Coatomer subunit beta OS=Trichoplusia ni OX=7111 GN=LOC113506617 PE=4 SV=1	Yes	117.517	27	15	1	958	103.4	151.87
A0A7E5WW29	Coatomer subunit beta OS=Trichoplusia ni OX=7111 GN=LOC113506395 PE=4 SV=1	Yes	117.533	27	15	1	981	106.4	152.83
A0A7E5WD37	nucleolar protein 58 OS=Trichoplusia ni OX=7111 GN=LOC113501219 PE=3 SV=1	Yes	122.725	41	18	18	535	59.8	154.95
A0A7E5WLA2	heat shock protein 60A-like OS=Trichoplusia ni OX=7111 GN=LOC113503680 PE=3 SV=1	Yes	129.158	48	15	15	572	60.9	174.48
A0A7E5VL01	uncharacterized protein LOC113494730 OS=Trichoplusia ni OX=7111 GN=LOC113494730 PE=4 SV=1	Yes	75.829	29	15	15	582	66.4	114.98
A0A7E5WIZ7	DNA replication licensing factor MCM4 OS=Trichoplusia ni OX=7111 GN=LOC113503128 PE=3 SV=1	Yes	97.95	29	14	14	861	94.9	133.58
A0A7E5VDX7	60S acidic ribosomal protein P0 OS=Trichoplusia ni OX=7111 GN=LOC113492956 PE=3 SV=1	Yes	96.404	66	14	14	316	34	129.44
A0A7E5WZY0	Histone H2A OS=Trichoplusia ni OX=7111 GN=LOC113507348 PE=3 SV=1	Yes	140.562	64	8	6	124	13.3	152.52
A0A7E5VCZ8	UDP-glucose:glycoprotein glucosyltransferase OS=Trichoplusia ni OX=7111 GN=LOC113492729 PE=3	Yes	109.217	22	18	18	1569	177.1	143.4
A0A7E5VCV0	extended synaptotagmin-2-A isoform X1 OS=Trichoplusia ni OX=7111 GN=LOC113492714 PE=4 SV=1	Yes	92.035	25	14	14	821	92.4	135.07
A0A7E5WJL7	60S acidic ribosomal protein P2 OS=Trichoplusia ni OX=7111 GN=LOC113502816 PE=3 SV=1	Yes	107.581	85	15	15	111	11.4	123.82
A0A7E5VU81	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1 OS=Trichoplusia ni OX=7111	Yes	100.094	54	17	17	474	54.1	133.16
A0A7E5WBC0	40S ribosomal protein S14 OS=Trichoplusia ni OX=7111 GN=LOC113500782 PE=3 SV=1	Yes	89.434	72	11	11	151	16.2	128.35
A0A7E5VHB9	Isocitrate dehydrogenase [NAD] subunit, mitochondrial OS=Trichoplusia ni OX=7111 GN=LOC11349378	Yes	79.858	26	10	10	389	42.2	125.46
A0A7E5VVN0	T-complex protein 1 subunit delta OS=Trichoplusia ni OX=7111 GN=LOC113497142 PE=3 SV=1	Yes	117.206	35	10	10	541	56.9	152.64
A0A7E5VW33	engulfment and cell motility protein 1 OS=Trichoplusia ni OX=7111 GN=LOC113497273 PE=4 SV=1	Yes	78.166	24	13	13	743	84.5	126.9
A0A7E5X012	Polypeptide N-acetylgalactosaminyltransferase OS=Trichoplusia ni OX=7111 GN=LOC113507745 PE=3	Yes	106.178	21	11	11	629	71.9	134.76
A0A7E5WMN1	Coatomer subunit beta' OS=Trichoplusia ni OX=7111 GN=LOC113504086 PE=3 SV=1	Yes	92.139	21	13	13	939	105.4	130.12
A0A7E5WY31	40S ribosomal protein S17 OS=Trichoplusia ni OX=7111 GN=LOC113506555 PE=3 SV=1	Yes	68.271	46	4	4	133	15.4	140.64
A0A7E5X1T0	Voltage-dependent anion-selective channel protein 3 OS=Trichoplusia ni OX=7111 GN=LOC113507651	Yes	101.963	58	14	14	282	30.1	123.45
A0A7E5VT10	T-complex protein 1 subunit zeta OS=Trichoplusia ni OX=7111 GN=LOC113496326 PE=3 SV=1	Yes	82.597	33	15	15	531	57.9	104.78
A9XXA7	Elongation factor 1-alpha OS=Trichoplusia ni OX=7111 GN=LOC113501794 PE=2 SV=1	Yes	94.787	38	11	11	463	50.3	109.17
A0A7E5X2B8	Histone H2B OS=Trichoplusia ni OX=7111 GN=LOC113507804 PE=3 SV=1	Yes	26.426	28	4	4	124	13.8	94.9
A0A7E5W2S7	40S ribosomal protein S3a OS=Trichoplusia ni OX=7111 GN=LOC113498677 PE=3 SV=1	Yes	48.129	38	12	12	264	29.9	83.19
A0A7E5VM91	RuvB-like helicase OS=Trichoplusia ni OX=7111 GN=LOC113495045 PE=3 SV=1	Yes	75.526	31	8	8	456	49.7	129.96
A0A7E5X421	importin-7 isoform X2 OS=Trichoplusia ni OX=7111 GN=LOC113508374 PE=4 SV=1	Yes	89.785	17	12	12	1046	120.5	115.51
A0A7E5W030	protein LSM14 homolog A OS=Trichoplusia ni OX=7111 GN=LOC113498215 PE=3 SV=1	Yes	81.693	21	8	8	708	78.6	124.24
A0A7E5WVI5	metal transporter CNNM4-like OS=Trichoplusia ni OX=7111 GN=LOC113505583 PE=3 SV=1	Yes	69.031	23	16	16	928	104.4	89.63
A0A7E5X2F1	probable medium-chain specific acyl-CoA dehydrogenase, mitochondrial isoform X3 OS=Trichoplusia ni	Yes	65.108	32	12	12	422	46	106.59
A0A7E5W0E1	CCT-theta OS=Trichoplusia ni OX=7111 GN=LOC113498305 PE=3 SV=1	Yes	81.262	36	15	15	545	59.2	104.08
A0A7E5VKT5	serine/threonine-protein phosphatase PP2A 65 kDa regulatory subunit OS=Trichoplusia ni OX=7111 GN	Yes	55.465	23	9	9	590	65.2	94.52
A0A7E5WXW5	Aspartate--tRNA ligase, cytoplasmic OS=Trichoplusia ni OX=7111 GN=LOC113507036 PE=3 SV=1	Yes	56.275	38	14	14	536	59.9	96.58
A0A7E5VP72	ATPase family AAA domain-containing protein 3A homolog OS=Trichoplusia ni OX=7111 GN=LOC11349	Yes	60.461	32	15	15	630	70.6	91.73
A0A7E5VLJ2	cAMP-dependent protein kinase type II regulatory subunit OS=Trichoplusia ni OX=7111 GN=LOC11349	Yes	66.115	60	15	15	382	43.4	96.16
A0A7E5WD79	Arginine kinase OS=Trichoplusia ni OX=7111 GN=LOC113501645 PE=3 SV=1	Yes	61.045	32	12	12	379	42.3	89.26
A0A7E5WK34	myelin expression factor 2 isoform X6 OS=Trichoplusia ni OX=7111 GN=LOC113503358 PE=4 SV=1	Yes	56.584	33	12	12	532	57.8	89.51
A0A7E5WL95	regulator of nonsense transcripts 1-like OS=Trichoplusia ni OX=7111 GN=LOC113503603 PE=3 SV=1	Yes	74.669	19	13	13	1038	115.6	94.27
A0A7E5W5Z7	26S proteasome regulatory subunit 7 OS=Trichoplusia ni OX=7111 GN=LOC113499378 PE=3 SV=1	Yes	52.186	34	11	11	434	48.5	94.34
A0A7E5WSZ1	Sidoreflexin OS=Trichoplusia ni OX=7111 GN=LOC113505284 PE=3 SV=1	Yes	69.739	40	10	10	327	36.8	99.22
A0A7E5X1H7	Histone H4 OS=Trichoplusia ni OX=7111 GN=LOC113507807 PE=3 SV=1	Yes	36.658	33	9	9	218	24.3	61.41
A0A7E5X2W4	la-related protein 1-like OS=Trichoplusia ni OX=7111 GN=LOC113508673 PE=4 SV=1	Yes	56.359	15	11	11	1282	141.6	77.93
A0A7E5WJN2	T-complex protein 1 subunit gamma OS=Trichoplusia ni OX=7111 GN=LOC113503333 PE=3 SV=1	Yes	79.755	37	12	12	541	59.1	84.9
A0A7E5WZB6	thioredoxin-related transmembrane protein 1-like OS=Trichoplusia ni OX=7111 GN=LOC113507588 PE=	Yes	28.875	15	4	4	301	34	70.68
A0A7E5WYK0	LOW QUALITY PROTEIN: 26S proteasome regulatory subunit 4-like OS=Trichoplusia ni OX=7111 GN=L	Yes	63.277	34	10	9	439	49.2	85.58
A0A7E5X4I5	NADH-cytochrome b5 reductase OS=Trichoplusia ni OX=7111 GN=LOC113508463 PE=3 SV=1	Yes	44.794	49	13	13	314	35.5	63.24
A0A7E5VGR7	Cytochrome c oxidase subunit 4 OS=Trichoplusia ni OX=7111 GN=LOC113493724 PE=3 SV=1	Yes	30.543	37	5	5	177	20.2	68.93
A0A7E5VSM9	sequestosome-1-like isoform X3 OS=Trichoplusia ni OX=7111 GN=LOC113496361 PE=4 SV=1	Yes	69.95	30	7	2	542	58.6	90.28
A0A7E5WF42	Translocon-associated protein subunit beta OS=Trichoplusia ni OX=7111 GN=LOC113502129 PE=3 SV	Yes	44.516	42	5	5	190	20.9	92.13
A0A7E5W4C9	26S proteasome non-ATPase regulatory subunit 2 OS=Trichoplusia ni OX=7111 GN=LOC113499000 PE	Yes	45.931	20	12	12	928	101.4	71.26
A0A7E5VB26	charged multivesicular body protein 7 isoform X2 OS=Trichoplusia ni OX=7111 GN=LOC113492259 PE=	Yes	62	29	11	10	626	69.6	70.37
A0A7E5VL44	nucleolar protein 56 OS=Trichoplusia ni OX=7111 GN=LOC113494788 PE=3 SV=1	Yes	73.917	33	10	10	459	51.2	83.22
A0A7E5VYD3	ATPase family AAA domain-containing protein 1-like isoform X2 OS=Trichoplusia ni OX=7111 GN=LOC1	Yes	42.156	27	8	8	381	42.9	66.68
A0A7E5WVQ8	ATP synthase subunit beta OS=Trichoplusia ni OX=7111 GN=LOC113505642 PE=3 SV=1	Yes	43.163	13	6	1	652	69.9	74.94
A0A7E5WJD6	actin cytoskeleton-regulatory complex protein PAN1-like isoform X2 OS=Trichoplusia ni OX=7111 GN=L	Yes	47.816	19	10	2	671	72.2	62.51
A0A7E5VN70	DNA replication licensing factor MCM6 OS=Trichoplusia ni OX=7111 GN=LOC113495298 PE=3 SV=1	Yes	58.071	28	12	12	818	91.6	78.57
A0A7E5VHV3	Coatomer subunit gamma OS=Trichoplusia ni OX=7111 GN=LOC113493971 PE=3 SV=1	Yes	71.355	12	7	7	861	94.7	79.58
A0A7E5VGL3	AP complex subunit beta OS=Trichoplusia ni OX=7111 GN=LOC113493578 PE=3 SV=1	Yes	62.732	17	9	9	911	99.8	84.41
A0A7E5WJR6	40S ribosomal protein S3 OS=Trichoplusia ni OX=7111 GN=LOC113502978 PE=3 SV=1	Yes	35.478	37	9	9	243	26.8	59.44

A0A7E5VSK8	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 2 OS=Trichoplusia ni OX=7111	Yes	44.836	24	10	10	621	67.7	60.74
A0A7E5VBV3	DNA damage-binding protein 1 OS=Trichoplusia ni OX=7111 GN=LOC113492460 PE=4 SV=1	Yes	39.08	13	10	10	1136	125.2	65.7
A0A7E5W4P6	Alpha-mannosidase OS=Trichoplusia ni OX=7111 GN=LOC113499077 PE=3 SV=1	Yes	49.402	13	8	8	1156	132.5	75.96
A0A7E5WFR3	Eukaryotic translation initiation factor 3 subunit E OS=Trichoplusia ni OX=7111 GN=LOC113502235 PE	Yes	51.628	23	9	9	445	52	57.09
A0A7E5WP50	Ribosomal protein S6 kinase OS=Trichoplusia ni OX=7111 GN=LOC113504422 PE=3 SV=1	Yes	36.263	19	5	5	452	50.7	64.95
A0A7E5WPX8	Dymeclin OS=Trichoplusia ni OX=7111 GN=LOC113504330 PE=3 SV=1	Yes	49.409	7	4	4	722	80.8	75.2
A0A7E5W6Q4	E3 ubiquitin-protein ligase KCMF1 OS=Trichoplusia ni OX=7111 GN=LOC113499596 PE=4 SV=1	Yes	62.183	27	7	7	373	39.4	71.77
A0A7E5VDW5	60S acidic ribosomal protein P1 OS=Trichoplusia ni OX=7111 GN=LOC113492912 PE=3 SV=1	Yes	56.401	71	6	6	112	11.3	75.96
A0A7E5VGA7	probable 26S proteasome non-ATPase regulatory subunit 3 OS=Trichoplusia ni OX=7111 GN=LOC1134	Yes	49.973	25	10	10	494	56.3	69.53
A0A7E5VSY2	ATP-dependent 6-phosphofructokinase OS=Trichoplusia ni OX=7111 GN=LOC113496310 PE=3 SV=1	Yes	49.597	24	12	12	782	85.7	62.34
A0A7E5VFE9	dnaJ homolog shv OS=Trichoplusia ni OX=7111 GN=LOC113493285 PE=4 SV=1	Yes	50.484	39	8	8	353	39.8	72.3
A0A7E5WNQ1	60S ribosomal protein L23 OS=Trichoplusia ni OX=7111 GN=LOC113504069 PE=3 SV=1	Yes	54.822	57	8	8	140	14.8	65.24
A0A7E5VYG1	Arginyl-tRNA synthetase OS=Trichoplusia ni OX=7111 GN=LOC113497820 PE=3 SV=1	Yes	31.862	20	13	13	683	77.5	45.61
A0A7E5VI16	vesicle-associated membrane protein/synaptobrevin-binding protein isoform X1 OS=Trichoplusia ni OX=	Yes	55.004	35	10	10	278	31	65.34
A0A7E5VVX2	Tubulin gamma chain OS=Trichoplusia ni OX=7111 GN=LOC113497039 PE=3 SV=1	Yes	43.617	29	9	9	456	51.3	69.79
A0A7E5VZ30	mitochondrial import receptor subunit TOM40 homolog 1 OS=Trichoplusia ni OX=7111 GN=LOC1134979	Yes	42.438	47	11	11	315	34.2	58.16
A0A7E5VRS4	H(+)-transporting two-sector ATPase OS=Trichoplusia ni OX=7111 GN=LOC113496116 PE=3 SV=1	Yes	53.441	30	11	11	617	68	61.4
A0A7E5WAY5	Eukaryotic translation initiation factor 3 subunit H OS=Trichoplusia ni OX=7111 GN=LOC113501042 PE	Yes	42.179	29	7	7	337	38.6	67.63
A0A7E5WK01	Eukaryotic translation initiation factor 3 subunit C OS=Trichoplusia ni OX=7111 GN=LOC113503061 PE	Yes	47.545	19	10	10	884	102	59.72
A0A7E5VAX0	LOW QUALITY PROTEIN: uncharacterized protein LOC113492244 OS=Trichoplusia ni OX=7111 GN=LC	Yes	54.361	11	8	8	961	102.9	57.44
A0A7E5VCU3	RNA helicase OS=Trichoplusia ni OX=7111 GN=LOC113492699 PE=3 SV=1	Yes	43.191	22	10	9	734	81.6	54.25
A0A7E5WKS7	protein lingerer-like OS=Trichoplusia ni OX=7111 GN=LOC113503108 PE=4 SV=1	Yes	82.251	13	5	5	947	100.2	78.32
A0A7E5W099	Signal sequence receptor subunit alpha OS=Trichoplusia ni OX=7111 GN=LOC113498050 PE=3 SV=1	Yes	34.697	38	4	4	280	30.9	66.29
A0A7E5VTL8	cullin-associated NEDD8-dissociated protein 1 OS=Trichoplusia ni OX=7111 GN=LOC113496611 PE=3	Yes	31.445	9	7	7	1256	138.3	53.97
A0A7E5VES6	ras GTPase-activating protein-binding protein 2 OS=Trichoplusia ni OX=7111 GN=LOC113493164 PE=4	Yes	30.068	15	6	6	439	47.5	58.31
A0A7E5VWI4	piwi-like protein Siwi OS=Trichoplusia ni OX=7111 GN=LOC113497222 PE=3 SV=1	Yes	27.039	12	8	8	903	101.8	55.65
A0A7E5X1U5	CCT-beta OS=Trichoplusia ni OX=7111 GN=LOC113508363 PE=3 SV=1	Yes	52.096	20	8	8	536	57.8	64.72
A0A7E5WJV7	60S ribosomal protein L17 OS=Trichoplusia ni OX=7111 GN=LOC113503318 PE=3 SV=1	Yes	27.927	25	5	5	187	21.5	41.21
A0A7E5WKN9	uncharacterized protein LOC113503212 isoform X4 OS=Trichoplusia ni OX=7111 GN=LOC113503212 P	Yes	35.391	17	9	1	656	70.5	47.19
A0A7E5WXV7	LOW QUALITY PROTEIN: NADH-ubiquinone oxidoreductase 75 kDa subunit, mitochondrial-like OS=Tric	Yes	45.812	22	8	8	627	68.9	60.98
A0A7E5VDG0	stromal cell-derived factor 2 OS=Trichoplusia ni OX=7111 GN=LOC113492843 PE=4 SV=1	Yes	45.789	52	4	4	225	24.7	84.13
A0A7E5WMQ8	polyubiquitin-A OS=Trichoplusia ni OX=7111 GN=LOC113503581 PE=4 SV=1	Yes	35.179	55	6	6	1293	145.1	57.42
A0A7E5VSV1	sequestosome-1-like isoform X2 OS=Trichoplusia ni OX=7111 GN=LOC113496361 PE=4 SV=1	Yes	51.359	28	6	1	553	60	66.83
A0A7E5WWA9	Heat shock protein 83 OS=Trichoplusia ni OX=7111 GN=LOC113506484 PE=3 SV=1	Yes	48.346	17	8	8	780	88.7	60.44
A0A7E5WBY2	zinc transporter ZIP11-like OS=Trichoplusia ni OX=7111 GN=LOC113501271 PE=4 SV=1	Yes	55.159	12	3	3	353	37.6	75.5
A0A7E5VJJ6	phosphatidylinositide phosphatase SAC1-like OS=Trichoplusia ni OX=7111 GN=LOC113494355 PE=4 S	Yes	38.792	18	7	7	599	67.7	60.88
A0A7E5VU25	ankyrin repeat domain-containing protein 17 isoform X2 OS=Trichoplusia ni OX=7111 GN=LOC11349667	Yes	50.349	7	8	8	2591	275	57.8
A0A7E5VGM4	uncharacterized protein LOC113493678 isoform X1 OS=Trichoplusia ni OX=7111 GN=LOC113493678 P	Yes	44.017	22	7	7	369	40.2	52.93
A0A7E5VSR8	Vesicle-fusing ATPase OS=Trichoplusia ni OX=7111 GN=LOC113496395 PE=3 SV=1	Yes	38.484	20	6	6	440	49.2	45.36
A0A7E5WYW5	transmembrane protein 209 OS=Trichoplusia ni OX=7111 GN=LOC113507403 PE=4 SV=1	Yes	61.27	25	7	7	435	47.1	58.77
A0A7E5VQ68	Y-box factor homolog isoform X2 OS=Trichoplusia ni OX=7111 GN=LOC113495753 PE=4 SV=1	Yes	47.841	41	5	5	272	29.2	57.02
A0A7E5WWK3	Eukaryotic translation initiation factor 3 subunit F OS=Trichoplusia ni OX=7111 GN=LOC113506587 PE	Yes	32.112	26	5	5	286	30.5	45.48
A0A7E5WLD9	zinc finger RNA-binding protein isoform X4 OS=Trichoplusia ni OX=7111 GN=LOC113503285 PE=4 SV=	Yes	40.876	10	4	4	919	98.1	59.64
A0A7E5WYE3	zinc transporter ZIP13 homolog OS=Trichoplusia ni OX=7111 GN=LOC113507215 PE=4 SV=1	Yes	51.042	33	8	8	391	42.3	61.33
A0A7E5VYH3	protein transport protein Sec31A isoform X1 OS=Trichoplusia ni OX=7111 GN=LOC113497802 PE=3 SV	Yes	46.597	7	6	6	1220	131.9	57.09
A0A7E5W6S4	Pyruvate dehydrogenase E1 component subunit alpha OS=Trichoplusia ni OX=7111 GN=LOC11349984	Yes	38.044	29	6	6	401	43.6	55.18
A0A7E5VU29	fragile X mental retardation syndrome-related protein 1 OS=Trichoplusia ni OX=7111 GN=LOC11349672	Yes	50.316	24	8	8	604	67.4	56.13
A0A7E5W0U9	Glutamyl-tRNA synthetase OS=Trichoplusia ni OX=7111 GN=LOC113498448 PE=3 SV=1	Yes	37.869	8	7	7	1635	181.7	57.27
A0A7E5WYR8	40S ribosomal protein S2 OS=Trichoplusia ni OX=7111 GN=LOC113507345 PE=3 SV=1	Yes	30.099	36	8	8	261	28	42.26
A0A7E5VJH5	cytochrome c oxidase subunit 6C-like OS=Trichoplusia ni OX=7111 GN=LOC113494341 PE=3 SV=1	Yes	24.735	51	4	4	77	8.8	44.39
A0A7E5WMG0	Mitochondrial import inner membrane translocase subunit TIM50 OS=Trichoplusia ni OX=7111 GN=LOC	Yes	40.664	24	6	6	386	45.1	60.79
A0A7E5VVS0	mitochondrial import receptor subunit TOM70 OS=Trichoplusia ni OX=7111 GN=LOC113497203 PE=4 S	Yes	36.327	23	9	9	547	61.7	45.07
A0A7E5W6Z0	talin-1 isoform X2 OS=Trichoplusia ni OX=7111 GN=LOC113499982 PE=4 SV=1	Yes	44.048	5	9	9	3585	392.5	53.61
A0A7E5X4Q3	Leucyl-tRNA synthetase OS=Trichoplusia ni OX=7111 GN=LOC113508534 PE=3 SV=1	Yes	56.373	13	8	8	1176	134	56.65
A0A7E5W0C1	Calcium-transporting ATPase OS=Trichoplusia ni OX=7111 GN=LOC113498289 PE=3 SV=1	Yes	51.889	12	9	9	1157	127.3	54.2
A0A7E5V9S2	26S proteasome non-ATPase regulatory subunit 12 OS=Trichoplusia ni OX=7111 GN=LOC113491986 P	Yes	57.82	17	4	4	450	51.5	65.61
A0A7E5VUB5	heterogeneous nuclear ribonucleoprotein R OS=Trichoplusia ni OX=7111 GN=LOC113496737 PE=4 SV=	Yes	43.45	20	9	9	629	69.4	52.99
A0A7E5VDC2	26S proteasome regulatory subunit RPN11 OS=Trichoplusia ni OX=7111 GN=LOC113492787 PE=4 SV=	Yes	30.173	47	8	8	311	34.6	43.61
A0A7E5W2N9	Acetyl-CoA carboxylase OS=Trichoplusia ni OX=7111 GN=LOC113498903 PE=4 SV=1	Yes	35.37	5	8	8	2400	269.7	43.59
A0A7E5VU78	protein hu-li tai shao isoform X5 OS=Trichoplusia ni OX=7111 GN=LOC113496708 PE=3 SV=1	Yes	32.962	22	8	8	701	77.4	45.53
A0A7E5WDP5	atlastin-like isoform X1 OS=Trichoplusia ni OX=7111 GN=LOC113501762 PE=3 SV=1	Yes	33.909	11	7	7	770	86.9	44.88
A0A7E5W7J7	ATP-binding cassette sub-family F member 2 OS=Trichoplusia ni OX=7111 GN=LOC113499837 PE=4 S	Yes	32.475	15	7	7	621	70.7	37.91
A0A7E5X2D5	hypoxia up-regulated protein 1 OS=Trichoplusia ni OX=7111 GN=LOC113508064 PE=3 SV=1	Yes	35.345	12	7	7	913	101.9	41.61
A0A7E5WUT4	uncharacterized protein LOC113505917 OS=Trichoplusia ni OX=7111 GN=LOC113505917 PE=4 SV=1	Yes	40.954	61	4	4	85	9.7	45.09
A0A7E5WZG9	dehydrogenase/reductase SDR family protein 7-like OS=Trichoplusia ni OX=7111 GN=LOC113506976 P	Yes	35.054	31	4	4	312	34.2	60.63
A0A7E5W4V9	regulator of gene activity isoform X1 OS=Trichoplusia ni OX=7111 GN=LOC113499390 PE=3 SV=1	Yes	36.911	29	5	5	420	45.2	51.24
A0A7E5W0P0	3-oxoacyl-[acyl-carrier-protein] reductase OS=Trichoplusia ni OX=7111 GN=LOC113498344 PE=4 SV=1	Yes	39.007	4	6	6	2395	256	45.82
A0A7E5VZG5	protein lifeguard 1 isoform X6 OS=Trichoplusia ni OX=7111 GN=LOC113498048 PE=3 SV=1	Yes	57.119	19	3	3	246	27.2	55.7
A0A7E5WJD3	flotillin-2 isoform X1 OS=Trichoplusia ni OX=7111 GN=LOC113503171 PE=3 SV=1	Yes	28.518	24	8	8	443	48.3	39.95
A0A7E5WNR7	paraplegin OS=Trichoplusia ni OX=7111 GN=LOC113504325 PE=3 SV=1	Yes	26.232	17	8	8	783	87.5	46.69
A0A7E5VJT9	protein 4.1 homolog isoform X2 OS=Trichoplusia ni OX=7111 GN=LOC113494423 PE=4 SV=1	Yes	37.546	5	7	7	3019	327.9	52.92
A0A7E5VFK0	small nuclear ribonucleoprotein-associated protein B OS=Trichoplusia ni OX=7111 GN=LOC113493298	Yes	26.21	32	5	5	188	19.6	40.7
A0A7E5W2K6	Sodium/potassium-transporting ATPase subunit alpha OS=Trichoplusia ni OX=7111 GN=LOC113498884	Yes	40.97	11	7	7	1009	111.5	47.67
A0A7E5W676	protein tumorous imaginal discs, mitochondrial-like isoform X1 OS=Trichoplusia ni OX=7111 GN=LOC11	Yes	35.101	15	6	6	498	55.1	45.51
A0A7E5WZQ1	T-complex protein 1 subunit eta OS=Trichoplusia ni OX=7111 GN=LOC113507248 PE=3 SV=1	Yes	37.87	19	9	9	544	58.9	44.43
A0A7E5WAY7	Protein disulfide-isomerase OS=Trichoplusia ni OX=7111 GN=LOC113501048 PE=3 SV=1	Yes	31.119	19	5	5	494	55.4	40.87
A0A7E5VD94	V-type proton ATPase subunit H OS=Trichoplusia ni OX=7111 GN=LOC113492751 PE=3 SV=1	Yes	25.154	13	6	6	474	55	44.35
A0A7E5WRJ3	26S proteasome regulatory subunit 6A-B OS=Trichoplusia ni OX=7111 GN=LOC113504711 PE=3 SV=1	Yes	21.489	25	7	6	428	47.7	34.91
A0A7E5W6R4	Coiled-coil domain-containing protein 47 OS=Trichoplusia ni OX=7111 GN=LOC113499597 PE=4 SV=1	Yes	38.537	15	4	4	465	52.8	48.37
A0A7E5WZL9	ubiquitin-protein ligase E3A OS=Trichoplusia ni OX=7111 GN=LOC113507639 PE=4 SV=1	Yes	41.691	11	6	6	914	102.9	48.22
A0A7E5VTX1	uncharacterized protein LOC113496684 OS=Trichoplusia ni OX=7111 GN=LOC113496684 PE=4 SV=1	Yes	38.183	31	3	3	162	18.1	45.65
A0A7E5WDY6	cytoplasmic dynein 1 intermediate chain isoform X6 OS=Trichoplusia ni OX=7111 GN=LOC113501446 P	Yes	46.67	20	6	6	671	74.2	59.66
A0A7E5VLC9	26S proteasome regulatory subunit 10B OS=Trichoplusia ni OX=7111 GN=LOC113494826 PE=3 SV=1	Yes	28.951	23	5	5	396	44.7	55.48
A0A7E5VBD8	Phosphate carrier protein, mitochondrial OS=Trichoplusia ni OX=7111 GN=LOC113492332 PE=3 SV=1	Yes	21.173	15	4	4	359	39.3	37.7
A0A7E5VVQ7	congested-like trachea protein OS=Trichoplusia ni OX=7111 GN=LOC113497194 PE=3 SV=1	Yes	23.059	32	7	7	298	32.5	42.33
A0A7E5W8I0	ras-related protein Rab-39B OS=Trichoplusia ni OX=7111 GN=LOC113500058 PE=4 SV=1	Yes	51.636	32	5	5	222	25.1	64.68
A0A7E5VE26	nuclear pore complex protein Nup133 OS=Trichoplusia ni OX=7111 GN=LOC113492995 PE=3 SV=1	Yes	38.082	10	7	7	1134	122.1	39.2
A0A7E5VPM3	eukaryotic translation initiation factor 4 gamma 3 isoform X4 OS=Trichoplusia ni OX=7111 GN=LOC113	Yes	36.562	8	8	8	1742	192.7	42.18
A0A7E5VFM7	RNA-binding protein squid isoform X1 OS=Trichoplusia ni OX=7111 GN=LOC113493348 PE=4 SV=1	Yes	36.313	23	6	6	328	34.8	38
A0A7E5W0W9	Signal recognition particle receptor subunit beta OS=Trichoplusia ni OX=7111 GN=LOC113498464 PE=	Yes	20.67	24	5	5	250	28.7	35.74
A0A7E5WL69	ubiquilin-1 isoform X1 OS=Trichoplusia ni OX=7111 GN=LOC113503585 PE=4 SV=1	Yes	46.721	26	5	5	528	57	64.39
A0A7E5VT50	Polyadenylate-binding protein OS=Trichoplusia ni OX=7111 GN=LOC113496349 PE=3 SV=1	Yes	35.107	18	7	7	611	68.2	40.71
A0A7E5VX36	ATPase family protein 2 homolog OS=Trichoplusia ni OX=7111 GN=LOC113497517 PE=4 SV=1	Yes	31.925	13	6	6	744	83.1	44.19
A0A7E5WDP1	prolactin regulatory element-binding protein OS=Trichoplusia ni OX=7111 GN=LOC113501550 PE=4 SV=	Yes	20.522	22	8	8	424	46.9	30.88
A0A7E5X4J0	Signal transducer and activator of transcription OS=Trichoplusia ni OX=7111 GN=LOC113509063 PE=3	Yes	28.392	11	4	4	774	88.4	51.1
A0A7E5WGL9	uncharacterized protein LOC113502100 OS=Trichoplusia ni OX=7111 GN=LOC113502100 PE=4 SV=1	Yes	32.366	39	2	2	152	17.1	43.42
A0A7E5VTA1	Pyruvate dehydrogenase E1 component subunit beta OS=Trichoplusia ni OX=7111 GN=LOC113496524	Yes	28.048	22	5	5	367	39.7	40.48
A0A7E5VZE7	maternal protein tudor-like isoform X1 OS=Trichoplusia ni OX=7111 GN=LOC113498000 PE=4 SV=1	Yes	33.383	6	7	7	2393	268.5	39.28
A0A7E5X1G5	Eukaryotic translation initiation factor 3 subunit M OS=Trichoplusia ni OX=7111 GN=LOC113508254 PE	Yes	37.203	24	6	6	386	44.1	41.04
A0A7E5WE69	Prohibitin OS=Trichoplusia ni OX=7111 GN=LOC113501490 PE=3 SV=1	Yes	27.061	36	6	6	274	30.1	38.36

A0A7E5X2F8	60S ribosomal protein L38 OS=Trichoplusia ni OX=7111 GN=LOC113508086 PE=3 SV=1	Yes	15.547	50	4	4	70	8.2	25.47
A0A7E5WS51	uncharacterized protein LOC113505205 OS=Trichoplusia ni OX=7111 GN=LOC113505205 PE=4 SV=1	Yes	18.122	26	6	6	394	45.3	34.59
A0A7E5W0Y1	Protein O-GlcNAc transferase OS=Trichoplusia ni OX=7111 GN=LOC113498372 PE=3 SV=1	Yes	38.035	8	4	4	865	96.5	57.85
A0A7E5W568	AFG3-like protein 2 OS=Trichoplusia ni OX=7111 GN=LOC113499196 PE=3 SV=1	Yes	18.745	13	7	7	778	86.2	33.85
A0A7E5VA16	Eukaryotic translation initiation factor 3 subunit D OS=Trichoplusia ni OX=7111 GN=LOC113492047 PE	Yes	32.553	14	3	3	553	62.9	48.01
A0A451FUM5	Actin OS=Trichoplusia ni OX=7111 GN=LOC113507353 PE=2 SV=1	Yes	24.428	20	4	2	376	41.8	35.55
A0A7E5X5F7	RNA helicase OS=Trichoplusia ni OX=7111 GN=LOC113508964 PE=3 SV=1	Yes	30.558	14	6	6	620	69.3	36.26
A0A7E5WGS1	sorting and assembly machinery component 50 homolog A-like OS=Trichoplusia ni OX=7111 GN=LOC11	Yes	32.741	22	6	6	468	49.8	38.2
A0A7E5WDM3	proteasomal ubiquitin receptor ADRM1 isoform X1 OS=Trichoplusia ni OX=7111 GN=LOC113501753 PE	Yes	47.724	23	4	4	432	44.6	55.69
A0A7E5WZK0	60S ribosomal protein L23a OS=Trichoplusia ni OX=7111 GN=LOC113507000 PE=3 SV=1	Yes	36.521	31	5	5	298	31.2	40.37
A0A7E5W287	Staphylococcal nuclease domain-containing protein OS=Trichoplusia ni OX=7111 GN=LOC113498757 P	Yes	40.129	15	7	7	895	99.8	40.8
A0A7E5W8E1	40S ribosomal protein S16 isoform X2 OS=Trichoplusia ni OX=7111 GN=LOC113500369 PE=3 SV=1	Yes	20.062	41	7	7	151	17	26.26
A0A7E5VEF0	E3 ubiquitin-protein ligase RNF13 OS=Trichoplusia ni OX=7111 GN=LOC113493052 PE=4 SV=1	Yes	32.226	21	5	5	400	44.3	40.38
A0A7E5VVV5	Protein YIPF OS=Trichoplusia ni OX=7111 GN=LOC113497220 PE=3 SV=1	Yes	36.866	20	2	2	224	24.9	44.65
A0A7E5VQT9	ataxin-2-like protein OS=Trichoplusia ni OX=7111 GN=LOC113495904 PE=3 SV=1	Yes	34.983	8	5	5	905	94.8	34.87
A0A7E5V881	neurofilament heavy polypeptide isoform X1 OS=Trichoplusia ni OX=7111 GN=LOC113491627 PE=3 SV	Yes	29.309	22	3	3	370	40.8	39.27
A0A7E5WL55	Alkaline phosphatase OS=Trichoplusia ni OX=7111 GN=LOC113503650 PE=3 SV=1	Yes	34.045	18	4	4	529	57.9	40.48
A0A7E5VNH5	40S ribosomal protein S4 OS=Trichoplusia ni OX=7111 GN=LOC113495374 PE=3 SV=1	Yes	22.63	28	7	7	263	29.5	30.18
A0A7E5W5M1	Microtubule-associated protein Jupiter OS=Trichoplusia ni OX=7111 GN=LOC113499617 PE=3 SV=1	Yes	27.733	81	4	4	122	12.4	33.66
A0A7E5WLI9	heterogeneous nuclear ribonucleoprotein 27C isoform X4 OS=Trichoplusia ni OX=7111 GN=LOC113503	Yes	42.534	17	4	4	472	49.3	43.74
A0A7E5VXI9	acyl-CoA Delta(11) desaturase-like OS=Trichoplusia ni OX=7111 GN=LOC113497645 PE=3 SV=1	Yes	23.812	24	5	5	353	40.2	26.46
A0A7E5V9S5	Elongation factor 1-alpha OS=Trichoplusia ni OX=7111 GN=LOC113491975 PE=4 SV=1	Yes	22.229	13	6	6	650	69.9	32.24
A0A7E5WKH2	40S ribosomal protein S15 OS=Trichoplusia ni OX=7111 GN=LOC113503424 PE=3 SV=1	Yes	24.083	46	3	3	147	17	34.7
A0A7E5VZE6	serine palmitoyltransferase 1 OS=Trichoplusia ni OX=7111 GN=LOC113498039 PE=4 SV=1	Yes	21.937	18	4	4	450	50.8	37.9
A0A7E5WQV8	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase 48 kDa subunit OS=Trichoplusia ni OX=	Yes	19.684	22	5	5	435	49	32.28
A0A7E5WWZ6	cathepsin L-like OS=Trichoplusia ni OX=7111 GN=LOC113506377 PE=3 SV=1	Yes	26.684	27	6	6	341	38.1	29.78
A0A7E5VBX4	60S ribosomal protein L5 OS=Trichoplusia ni OX=7111 GN=LOC113492511 PE=3 SV=1	Yes	29.517	24	5	5	299	34.1	32.3
A0A7E5X1F1	heat shock protein 75 kDa, mitochondrial OS=Trichoplusia ni OX=7111 GN=LOC113508241 PE=3 SV=1	Yes	19.055	12	6	5	695	78.8	24.31
A0A7E5VHF4	Clathrin heavy chain OS=Trichoplusia ni OX=7111 GN=LOC113493842 PE=3 SV=1	Yes	15.11	4	4	4	1691	192.2	31.57
A0A7E5VDL2	40S ribosomal protein S8 OS=Trichoplusia ni OX=7111 GN=LOC113492881 PE=3 SV=1	Yes	16.599	35	7	7	208	23.8	26.7
A0A7E5VEY7	Nuclear pore protein OS=Trichoplusia ni OX=7111 GN=LOC113493212 PE=3 SV=1	Yes	33.402	9	5	5	821	92.9	38.38
A0A7E5W1Q7	Replication protein A subunit OS=Trichoplusia ni OX=7111 GN=LOC113498438 PE=3 SV=1	Yes	25.935	19	7	7	603	66.4	33.77
A0A7E5WE99	dnaJ protein homolog 1 OS=Trichoplusia ni OX=7111 GN=LOC113501513 PE=4 SV=1	Yes	16.275	13	4	4	352	39	27.61
A0A7E5W5I6	protein jagunal isoform X1 OS=Trichoplusia ni OX=7111 GN=LOC113499615 PE=3 SV=1	Yes	21.142	14	2	2	200	23.4	34.4
A0A7E5VNS0	Coatomer subunit alpha OS=Trichoplusia ni OX=7111 GN=LOC113495443 PE=4 SV=1	Yes	35.576	8	5	5	1233	137.9	38.4
A0A7E5VTM9	uncharacterized protein LOC113496621 OS=Trichoplusia ni OX=7111 GN=LOC113496621 PE=3 SV=1	Yes	26.403	9	4	4	712	80.3	29.82
A0A7E5WVJ2	syntaxin-7 OS=Trichoplusia ni OX=7111 GN=LOC113505867 PE=3 SV=1	Yes	26.038	31	4	4	262	29.9	38.71
A0A7E5WWV2	probable protein phosphatase 2C 11 isoform X1 OS=Trichoplusia ni OX=7111 GN=LOC113505975 PE=	Yes	29.881	10	3	3	680	72.4	34.99
A0A7E5WE59	cyclin-dependent kinase 10 OS=Trichoplusia ni OX=7111 GN=LOC113501891 PE=3 SV=1	Yes	22.483	17	5	5	404	45.3	29.88
A0A7E5VYY3	60S ribosomal protein L9 OS=Trichoplusia ni OX=7111 GN=LOC113497792 PE=3 SV=1	Yes	15.44	33	4	4	190	21.4	28.68
A0A7E5VSR4	ER membrane protein complex subunit 1 OS=Trichoplusia ni OX=7111 GN=LOC113496338 PE=3 SV=1	Yes	22.546	8	5	5	914	101	36.18
A0A7E5VQC0	failed axon connections OS=Trichoplusia ni OX=7111 GN=LOC113495770 PE=3 SV=1	Yes	28.784	17	4	4	378	42.9	30.93
A0A7E5VK81	protein purity of essence OS=Trichoplusia ni OX=7111 GN=LOC113494523 PE=3 SV=1	Yes	24.492	1	4	4	5261	574.9	35.56
A0A7E5WL52	GPN-loop GTPase 3 OS=Trichoplusia ni OX=7111 GN=LOC113503196 PE=3 SV=1	Yes	22.49	26	4	4	277	31.6	31.81
A0A7E5WY92	tumor suppressor candidate 3 OS=Trichoplusia ni OX=7111 GN=LOC113506805 PE=3 SV=1	Yes	24.389	12	5	5	332	37.8	32.34
A0A7E5WXT0	NADPH--cytochrome P450 reductase OS=Trichoplusia ni OX=7111 GN=LOC113507001 PE=3 SV=1	Yes	22.715	15	5	5	687	77.4	31.62
A0A7E5VSI2	prostatic acid phosphatase isoform X1 OS=Trichoplusia ni OX=7111 GN=LOC113496175 PE=4 SV=1	Yes	19.345	12	4	4	420	48.3	27.72
A0A7E5WMP6	aldehyde dehydrogenase, mitochondrial OS=Trichoplusia ni OX=7111 GN=LOC113503573 PE=3 SV=1	Yes	25.764	12	2	2	509	55.5	38.38
A0A7E5WLY0	mitochondrial 2-oxodicarboxylate carrier OS=Trichoplusia ni OX=7111 GN=LOC113503840 PE=3 SV=1	Yes	11.407	23	4	4	304	33.2	27.9
A0A7E5WM47	Kinesin light chain OS=Trichoplusia ni OX=7111 GN=LOC113503845 PE=3 SV=1	Yes	22.524	15	5	5	532	59.9	28.32
A0A7E5W815	WD and tetratricopeptide repeats protein 1 isoform X2 OS=Trichoplusia ni OX=7111 GN=LOC11350025	Yes	29.147	16	4	4	653	72.8	35.94
A0A7E5VE43	peroxiredoxin 1 isoform X3 OS=Trichoplusia ni OX=7111 GN=LOC113493007 PE=3 SV=1	Yes	23.878	42	6	5	195	21.9	25.6
A0A7E5WVF1	nuclear pore complex protein Nup153 isoform X1 OS=Trichoplusia ni OX=7111 GN=LOC113506133 PE=	Yes	32.061	7	5	5	1366	146	31.48
A0A7E5VBS3	Cleavage and polyadenylation specificity factor subunit 2 OS=Trichoplusia ni OX=7111 GN=LOC113492	Yes	16.433	10	4	4	735	81.8	27.86
A0A7E5WDE9	von Willebrand factor A domain-containing protein 8 OS=Trichoplusia ni OX=7111 GN=LOC113501696 P	Yes	17.315	6	5	5	1395	156.3	24.64
A0A7E5WIN9	Aminoacyl-tRNA hydrolase OS=Trichoplusia ni OX=7111 GN=LOC113502996 PE=4 SV=1	Yes	18.928	32	4	4	161	17.3	21.5
A0A7E5WZ01	dnaJ homolog subfamily C member 11-like OS=Trichoplusia ni OX=7111 GN=LOC113507444 PE=4 SV=	Yes	23.641	15	4	4	576	65.3	33.19
A0A7E5WCK4	Nucleolar GTP-binding protein 2 OS=Trichoplusia ni OX=7111 GN=LOC113501444 PE=3 SV=1	Yes	16.428	9	4	4	929	105.6	28.92
A0A7E5WGW5	1-acyl-sn-glycerol-3-phosphate acyltransferase gamma-like OS=Trichoplusia ni OX=7111 GN=LOC1135	Yes	15.401	19	5	5	393	45.7	17.51
A0A7E5WAH7	serine/threonine-protein kinase 4 isoform X4 OS=Trichoplusia ni OX=7111 GN=LOC113500953 PE=4 S	Yes	29.897	21	5	5	529	58.2	33.96
A0A7E5VI91	Microsomal prostaglandin E synthase 2 OS=Trichoplusia ni OX=7111 GN=LOC113494009 PE=3 SV=1	Yes	12.241	14	4	4	388	44.2	21.09
A0A7E5VB97	Signal sequence receptor subunit delta OS=Trichoplusia ni OX=7111 GN=LOC113492305 PE=3 SV=1	Yes	30.485	44	4	4	167	18.1	30.57
A0A7E5WVW7	uncharacterized protein LOC113506337 isoform X2 OS=Trichoplusia ni OX=7111 GN=LOC113506337 P	Yes	16.047	6	2	2	1172	132.4	32.7
A0A7E5W748	Cytochrome c oxidase subunit OS=Trichoplusia ni OX=7111 GN=LOC113499991 PE=3 SV=1	Yes	24.556	24	2	2	107	12.1	30.95
A0A7E5VZ77	aminoacyl tRNA synthase complex-interacting multifunctional protein 1 OS=Trichoplusia ni OX=7111 G	Yes	23.579	23	5	5	294	32.4	22.88
A0A7E5VJF6	mitochondrial import inner membrane translocase subunit Tim23 OS=Trichoplusia ni OX=7111 GN=LOC	Yes	30.109	34	3	3	194	20.9	36.35
A0A7E5VDN0	DNA replication licensing factor MCM7 OS=Trichoplusia ni OX=7111 GN=LOC113492876 PE=3 SV=1	Yes	17.565	10	5	5	719	81	24.62
A0A7E5VEM5	Cytochrome b-c1 complex subunit Rieske, mitochondrial OS=Trichoplusia ni OX=7111 GN=LOC1134931	Yes	15.849	26	4	4	271	28.9	29.57
A0A7E5WTB8	WD repeat-containing protein 43 OS=Trichoplusia ni OX=7111 GN=LOC113505087 PE=4 SV=1	Yes	15.225	7	4	4	632	69.9	22.92
A0A7E5VMD2	probable RNA-binding protein CG14230 OS=Trichoplusia ni OX=7111 GN=LOC113495090 PE=4 SV=1	Yes	16.005	11	4	4	585	67.4	31.42
A0A7E5X3M5	40S ribosomal protein S10 OS=Trichoplusia ni OX=7111 GN=LOC113508227 PE=3 SV=1	Yes	15.549	30	5	5	158	18	21.97
A0A7E5WGA1	protein transport protein Sec61 subunit alpha OS=Trichoplusia ni OX=7111 GN=LOC113502385 PE=3 S	Yes	31.231	11	4	4	476	52	32.42
A0A7E5WDM1	lamin-C-like OS=Trichoplusia ni OX=7111 GN=LOC113501742 PE=3 SV=1	Yes	20.563	17	6	6	614	70.1	24.67
A0A7E5X1B5	CTP synthase OS=Trichoplusia ni OX=7111 GN=LOC113507725 PE=3 SV=1	Yes	17.307	14	6	6	603	67.2	22.55
A0A7E5VSY1	cytochrome b-c1 complex subunit 2, mitochondrial-like OS=Trichoplusia ni OX=7111 GN=LOC11349641	Yes	27.694	27	5	5	438	46	31.84
A0A7E5WIE4	Threonyl-tRNA synthetase OS=Trichoplusia ni OX=7111 GN=LOC113502506 PE=3 SV=1	Yes	20.426	12	6	6	721	82.7	27.12
A0A7E5VBJ7	Small nuclear ribonucleoprotein Sm D1 OS=Trichoplusia ni OX=7111 GN=LOC113492379 PE=3 SV=1	Yes	19.272	26	2	2	127	13.7	31.98
A0A7E5WGH6	ATP synthase subunit b OS=Trichoplusia ni OX=7111 GN=LOC113502221 PE=3 SV=1	Yes	14.63	14	3	3	243	27.4	23.19
A0A7E5VL92	pyridoxal-dependent decarboxylase domain-containing protein 1 isoform X2 OS=Trichoplusia ni OX=711	Yes	14.805	7	4	4	821	88.9	24.26
A0A7E5WXC4	V-type proton ATPase proteolipid subunit OS=Trichoplusia ni OX=7111 GN=LOC113506850 PE=3 SV=1	Yes	34.289	24	2	2	157	16	38.14
A0A7E5VXD0	uncharacterized protein LOC113497545 OS=Trichoplusia ni OX=7111 GN=LOC113497545 PE=3 SV=1	Yes	18.81	40	3	3	127	13.8	26.12
A0A7E5VGZ6	vesicle-trafficking protein SEC22b OS=Trichoplusia ni OX=7111 GN=LOC113493689 PE=3 SV=1	Yes	19.904	35	5	5	217	24.9	25.81
A0A7E5WA27	LOW QUALITY PROTEIN: protein draper-like OS=Trichoplusia ni OX=7111 GN=LOC113500799 PE=4 S	Yes	14.933	7	4	4	1175	128.7	23.73
A0A7E5WP89	G-protein-signaling modulator 2 OS=Trichoplusia ni OX=7111 GN=LOC113504072 PE=3 SV=1	Yes	19.994	12	4	4	623	68.6	29.34
A0A7E5VNJ8	translocator protein isoform X3 OS=Trichoplusia ni OX=7111 GN=LOC113495402 PE=3 SV=1	Yes	17.461	15	1	1	176	19.5	30.25
A0A7E5VSQ2	ribonucleoside-diphosphate reductase subunit M2 OS=Trichoplusia ni OX=7111 GN=LOC113496347 PE	Yes	19.73	14	2	2	380	44.1	27.64
A0A7E5VTJ8	RNA helicase OS=Trichoplusia ni OX=7111 GN=LOC113496527 PE=4 SV=1	Yes	12.239	9	5	4	827	92.3	22.35
A0A7E5VL31	39S ribosomal protein L12, mitochondrial OS=Trichoplusia ni OX=7111 GN=LOC113494754 PE=3 SV=1	Yes	24.893	29	3	3	181	19.9	32.82
A0A7E5VWF1	multidrug resistance-associated protein 1 OS=Trichoplusia ni OX=7111 GN=LOC113497341 PE=4 SV=1	Yes	19.152	5	4	4	1518	168.9	25.58
A0A7E5WQI8	RNA-binding protein Nova-2-like isoform X5 OS=Trichoplusia ni OX=7111 GN=LOC113504776 PE=4 SV=	Yes	14.809	12	4	3	467	47.8	22.18
A0A7E5V8L2	uncharacterized protein LOC113491711 OS=Trichoplusia ni OX=7111 GN=LOC113491711 PE=4 SV=1	Yes	12.166	10	2	2	433	44.4	24.5
A0A7E5WGS3	protein unc-50 homolog isoform X1 OS=Trichoplusia ni OX=7111 GN=LOC113502143 PE=3 SV=1	Yes	26.272	13	3	3	272	31.8	23.91
A0A7E5X451	Glutaminyl-tRNA synthetase OS=Trichoplusia ni OX=7111 GN=LOC113508384 PE=3 SV=1	Yes	13.601	8	5	5	780	87.6	19.16
A0A7E5WNU3	ATP-dependent zinc metalloprotease YME1 homolog isoform X3 OS=Trichoplusia ni OX=7111 GN=LOC	Yes	22.421	13	5	5	767	84.6	26.67
A0A7E5X472	Inosine-5'-monophosphate dehydrogenase OS=Trichoplusia ni OX=7111 GN=LOC113508532 PE=3 SV=	Yes	17.356	12	3	3	512	55.4	29.8
A0A7E5VP84	nuclear valosin-containing protein-like OS=Trichoplusia ni OX=7111 GN=LOC113495566 PE=4 SV=1	Yes	15.575	8	3	3	866	95.3	29.14
A0A7E5WYK4	gem-associated protein 5-like OS=Trichoplusia ni OX=7111 GN=LOC113507272 PE=4 SV=1	Yes	18.688	6	4	4	1275	141.6	28.11
A0A7E5WCL4	26S proteasome non-ATPase regulatory subunit 7 OS=Trichoplusia ni OX=7111 GN=LOC113501489 PE	Yes	19.385	14	2	2	332	37.6	27.49
A0A7E5WIN3	nicalin-1 OS=Trichoplusia ni OX=7111 GN=LOC113502579 PE=3 SV=1	Yes	15.871	10	4	4	550	60.5	20.8

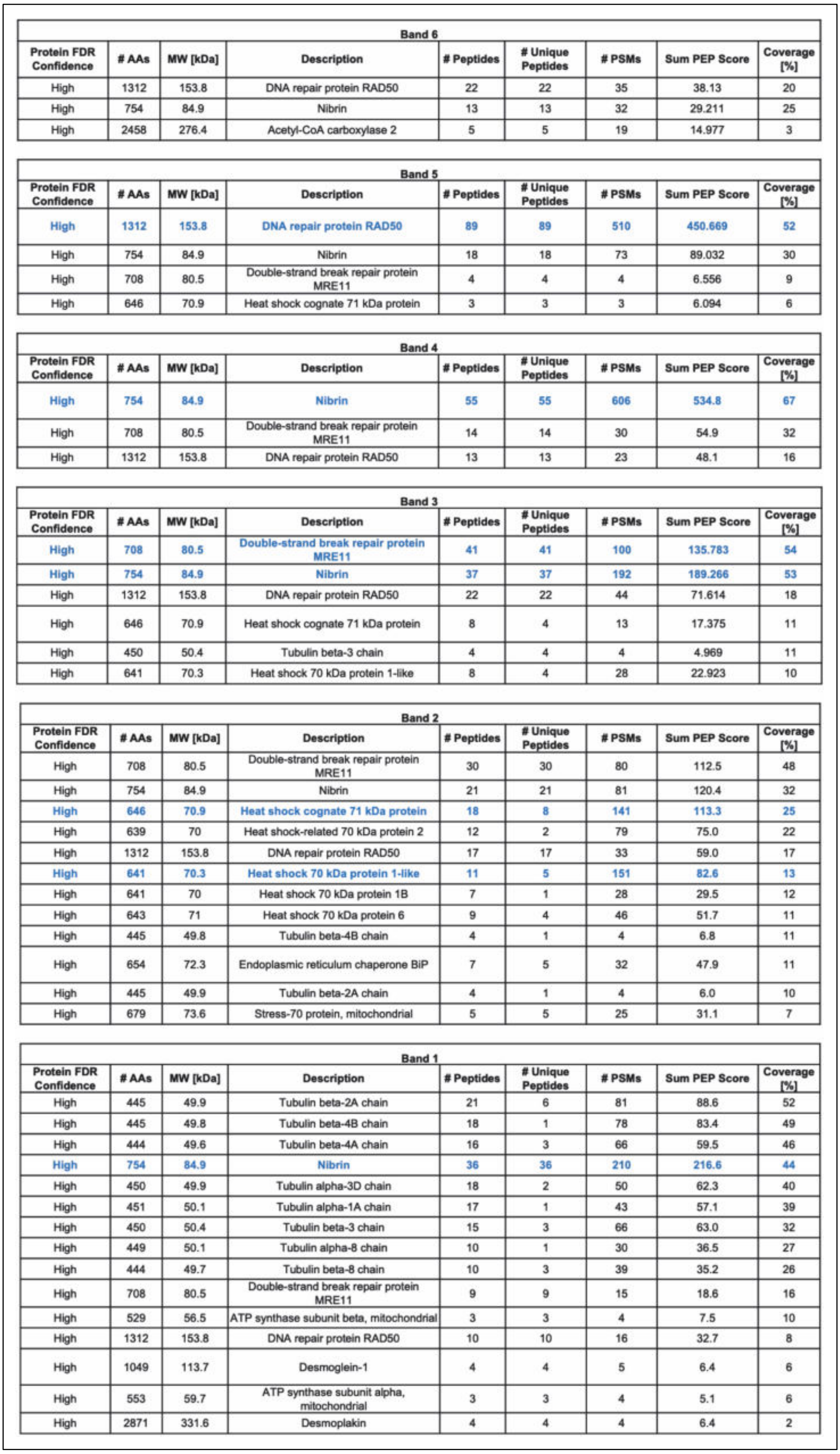
Table 23. Mass spectrometry analysis of purified EXD2 MUT (D108A E110A), by Dr Lu Yu. Table shows 300 from 593 hits.

Accession	Marked as	Description	Sum PEP Score	Coverage [%]	# Peptides	# Unique Peptides	# AAs	MW [kDa]	Score Sequest HT
Q9NVH0		Exonuclease 3'-5' domain-containing protein 2 OS=Homo sapiens OX=9606 GN=EXD2 PE=1 SV=2	969.739	87	68	3	621	70.3	7300.38
Q9NVH0_mutant		Exonuclease 3'-5' domain-containing protein 2 (mutation on D108A E110A) OS=Homo sapiens OX=9606 GN=EXD2_mut PE	999.002	86	67	2	621	70.2	7334.67
P04264		Keratin, type II cytoskeletal 1 OS=Homo sapiens OX=9606 GN=KRT1 PE=1 SV=6	275.677	58	34	31	644	66	664.95
P35908		Keratin, type II cytoskeletal 2 epidermal OS=Homo sapiens OX=9606 GN=KRT2 PE=1 SV=2	328.193	64	34	29	639	65.4	562.01
A0A7E5WY23	CabLooper	Heat shock protein 70 OS=Trichoplusia ni OX=7111 GN=LOC113506423 PE=2 SV=1	315.817	65	31	22	654	71.6	823.59
P35527		SWISS-PROT:P35527 Tax_Id=9606 Gene_Symbol=KRT9 Keratin, type I cytoskeletal 9	383.581	61	29	29	623	62.1	680.94
A0A7E5V8N1	CabLooper	Calcium-transporting ATPase OS=Trichoplusia ni OX=7111 GN=LOC113491725 PE=3 SV=1	241.164	33	29	29	1026	112.7	362.38
A0A2H4WW89	CabLooper	Tubulin alpha chain OS=Trichoplusia ni OX=7111 GN=LOC113497986 PE=2 SV=1	306.119	79	28	6	450	49.9	1004.65
A0A7E5VSK5	CabLooper	Heat shock protein 83 OS=Trichoplusia ni OX=7111 GN=LOC113496266 PE=3 SV=1	192.799	44	28	28	722	83.1	276.02
A0A7E5VK44	CabLooper	Tubulin beta chain OS=Trichoplusia ni OX=7111 GN=LOC113494443 PE=3 SV=1	353.345	77	27	4	448	50.4	1502.85
P13645		SWISS-PROT:P13645 Tax_Id=9606 Gene_Symbol=KRT10 Keratin, type I cytoskeletal 10	377.525	46	27	20	593	59.5	1261.56
A0A7E5WTY1	CabLooper	heat shock 70 kDa protein cognate 5 OS=Trichoplusia ni OX=7111 GN=LOC113505692 PE=3 SV=1	209.194	44	26	26	686	74.7	297.87
A0A7E5WHV4	CabLooper	Vesicle-fusing ATPase OS=Trichoplusia ni OX=7111 GN=LOC113502529 PE=3 SV=1	171.429	51	26	26	746	82.5	207.37
A0A7E5WW29	CabLooper	Coatomer subunit beta OS=Trichoplusia ni OX=7111 GN=LOC113506395 PE=4 SV=1	112.703	37	24	24	981	106.4	167.35
P15924		Desmoplakin OS=Homo sapiens OX=9606 GN=DSP PE=1 SV=3	119.255	12	23	23	2871	331.6	191.01
A0A7E5WLV3	CabLooper	Tubulin alpha chain OS=Trichoplusia ni OX=7111 GN=LOC113503819 PE=3 SV=1	237.871	50	22	1	450	49.8	910.12
A0A7E5W2N9	CabLooper	Acetyl-CoA carboxylase OS=Trichoplusia ni OX=7111 GN=LOC113498903 PE=4 SV=1	116.237	17	22	22	2400	269.7	165.21
Q9BVA1		Tubulin beta-2B chain OS=Homo sapiens OX=9606 GN=TUBB2B PE=1 SV=1	214.36	58	21	1	445	49.9	873.48
P68371		Tubulin beta-4B chain OS=Homo sapiens OX=9606 GN=TUBB4B PE=1 SV=1	242.406	59	21	0	445	49.8	858.76
A0A7E5X5Y1	CabLooper	Tubulin beta chain OS=Trichoplusia ni OX=7111 GN=LOC113508948 PE=3 SV=1	188.186	59	21	11	451	50.7	475.1
A0A7E5VBE7	CabLooper	ADP/ATP translocase OS=Trichoplusia ni OX=7111 GN=LOC113492335 PE=3 SV=1	91.41	61	21	21	300	32.7	258.32
P04350		Tubulin beta-4A chain OS=Homo sapiens OX=9606 GN=TUBB4A PE=1 SV=2	263.457	59	20	0	444	49.6	1012.9
Q0MUU6	CabLooper	Heat shock cognate 70 protein OS=Trichoplusia ni OX=7111 GN=LOC113494618 PE=3 SV=1	133.779	36	20	16	659	73.1	216.93
A0A7E5WX33	CabLooper	ATP synthase subunit beta OS=Trichoplusia ni OX=7111 GN=LOC113506412 PE=3 SV=1	197.041	61	20	16	515	54.9	257.7
A0A7E5VBV3	CabLooper	DNA damage-binding protein 1 OS=Trichoplusia ni OX=7111 GN=LOC113492460 PE=4 SV=1	68.834	23	20	20	1136	125.2	88.12
A0A7E5WD37	CabLooper	nucleolar protein 58 OS=Trichoplusia ni OX=7111 GN=LOC113501219 PE=3 SV=1	99.753	44	19	19	535	59.8	124.56
A0A7E5VJS6	CabLooper	Tubulin beta chain OS=Trichoplusia ni OX=7111 GN=LOC113494447 PE=3 SV=1	182.508	49	18	6	457	51.4	654.78
A0A7E5X115	CabLooper	Tubulin beta chain OS=Trichoplusia ni OX=7111 GN=LOC113507665 PE=3 SV=1	204.357	39	17	0	447	49.9	707.17
A0A7E5VZ30	CabLooper	mitochondrial import receptor subunit TOM40 homolog 1 OS=Trichoplusia ni OX=7111 GN=LOC113497931 PE=3 SV=1	242.35	74	17	17	315	34.2	477.08
P08779		SWISS-PROT:P08779 Tax_Id=9606 Gene_Symbol=KRT16 Keratin, type I cytoskeletal 16	110.623	43	17	10	473	51.2	157.16
P13647		SWISS-PROT:P13647 Tax_Id=9606 Gene_Symbol=KRT5 Keratin, type II cytoskeletal 5	82.745	34	17	12	590	62.3	131.86
A0A7E5WIZ7	CabLooper	DNA replication licensing factor MCM4 OS=Trichoplusia ni OX=7111 GN=LOC113503128 PE=3 SV=1	89.442	29	17	17	861	94.9	104.88
A0A7E5VRS4	CabLooper	H(+)-transporting two-sector ATPase OS=Trichoplusia ni OX=7111 GN=LOC113496116 PE=3 SV=1	70.706	48	17	17	617	68	95.46
A0A7E5W672	CabLooper	Aspartate--tRNA ligase, cytoplasmic OS=Trichoplusia ni OX=7111 GN=LOC113499807 PE=3 SV=1	57.206	40	17	17	536	59.8	74.04
A0A7E5VCV0	CabLooper	extended synaptotagmin-2-A isoform X1 OS=Trichoplusia ni OX=7111 GN=LOC113492714 PE=4 SV=1	61.476	31	17	17	821	92.4	74.46
Q9NY65		Tubulin alpha-8 chain OS=Homo sapiens OX=9606 GN=TUBA8 PE=1 SV=1	187.051	35	16	1	449	50.1	513
P02533		Keratin, type I cytoskeletal 14 OS=Homo sapiens OX=9606 GN=KRT14 PE=1 SV=4	109.14	42	16	9	472	51.5	143.89
A0A7E5WL95	CabLooper	regulator of nonsense transcripts 1-like OS=Trichoplusia ni OX=7111 GN=LOC113503603 PE=3 SV=1	72.047	21	16	16	1038	115.6	103.99
A0A7E5VN70	CabLooper	DNA replication licensing factor MCM6 OS=Trichoplusia ni OX=7111 GN=LOC113495298 PE=3 SV=1	90.637	32	16	16	818	91.6	120.28
P02538		SWISS-PROT:P02538 Tax_Id=9606 Gene_Symbol=KRT6A Keratin, type II cytoskeletal 6A	95.32	33	15	4	564	60	132.42
A0A7E5W8U8	CabLooper	dnaJ homolog subfamily A member 1 OS=Trichoplusia ni OX=7111 GN=LOC113500490 PE=3 SV=1	95.84	44	15	15	404	45.3	129.11
A0A7E5VEY7	CabLooper	Nuclear pore protein OS=Trichoplusia ni OX=7111 GN=LOC113493212 PE=3 SV=1	80.009	25	15	15	821	92.9	95.46
A0A7E5VE26	CabLooper	nuclear pore complex protein Nup133 OS=Trichoplusia ni OX=7111 GN=LOC113492995 PE=3 SV=1	65.621	23	15	15	1134	122.1	77.55
A0A7E5WZL9	CabLooper	ubiquitin-protein ligase E3A OS=Trichoplusia ni OX=7111 GN=LOC113507639 PE=4 SV=1	62.054	25	14	14	914	102.9	81.04
A0A7E5VWF1	CabLooper	multidrug resistance-associated protein 1 OS=Trichoplusia ni OX=7111 GN=LOC113497341 PE=4 SV=1	59.348	17	14	14	1518	168.9	79.96
A0A7E5VDX7	CabLooper	60S acidic ribosomal protein P0 OS=Trichoplusia ni OX=7111 GN=LOC113492956 PE=3 SV=1	136.638	69	13	13	316	34	204.1
P04259		Keratin, type II cytoskeletal 6B OS=Homo sapiens OX=9606 GN=KRT6B PE=1 SV=5	88.657	28	13	2	564	60	135.54
A0A7E5VL01	CabLooper	uncharacterized protein LOC113494730 OS=Trichoplusia ni OX=7111 GN=LOC113494730 PE=4 SV=1	90.98	27	13	13	582	66.4	116.91
A0A7E5WH63	CabLooper	constitutive coactivator of PPAR-gamma-like protein 1 isoform X1 OS=Trichoplusia ni OX=7111 GN=LOC113502610 PE=3	97.565	22	13	13	962	105.6	122.52
A0A7E5WIQ6	CabLooper	Asparagine--tRNA ligase OS=Trichoplusia ni OX=7111 GN=LOC113503013 PE=3 SV=1	90.972	40	13	13	554	63	103.05
A0A7E5W4C9	CabLooper	26S proteasome non-ATPase regulatory subunit 2 OS=Trichoplusia ni OX=7111 GN=LOC113499000 PE=3 SV=1	71.885	25	13	13	928	101.4	94.21
A0A7E5WMN1	CabLooper	Coatomer subunit beta' OS=Trichoplusia ni OX=7111 GN=LOC113504086 PE=3 SV=1	74.007	22	13	13	939	105.4	101.45
A0A7E5WZM0	CabLooper	ATP synthase subunit alpha OS=Trichoplusia ni OX=7111 GN=LOC113507229 PE=3 SV=1	60.204	30	13	13	557	59.9	72.38
A0A7E5VHV3	CabLooper	Coatomer subunit gamma OS=Trichoplusia ni OX=7111 GN=LOC113493971 PE=3 SV=1	90.193	24	12	11	861	94.7	100.47
A0A7E5WK01	CabLooper	Eukaryotic translation initiation factor 3 subunit C OS=Trichoplusia ni OX=7111 GN=LOC113503061 PE=3 SV=1	45.978	23	12	12	884	102	67.08
A0A7E5VSK8	CabLooper	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 2 OS=Trichoplusia ni OX=7111 GN=LOC11349634	46.202	25	12	12	621	67.7	53.55
A0A7E5VSR4	CabLooper	ER membrane protein complex subunit 1 OS=Trichoplusia ni OX=7111 GN=LOC113496338 PE=3 SV=1	51.34	19	12	12	914	101	65.64
A0A7E5VYG1	CabLooper	Arginyl-tRNA synthetase OS=Trichoplusia ni OX=7111 GN=LOC113497820 PE=3 SV=1	32.206	17	12	12	683	77.5	45.1
A0A7E5WAV9	CabLooper	dnaJ homolog subfamily A member 2-like OS=Trichoplusia ni OX=7111 GN=LOC113501015 PE=3 SV=1	84.865	44	11	11	403	44.9	154.18
A0A7E5X2W4	CabLooper	la-related protein 1-like OS=Trichoplusia ni OX=7111 GN=LOC113508673 PE=4 SV=1	68.972	14	11	11	1282	141.6	98.81
A0A7E5WXT0	CabLooper	NADPH--cytochrome P450 reductase OS=Trichoplusia ni OX=7111 GN=LOC113507001 PE=3 SV=1	72.735	28	11	11	687	77.4	105.16
A0A7E5VSY2	CabLooper	ATP-dependent 6-phosphofructokinase OS=Trichoplusia ni OX=7111 GN=LOC113496310 PE=3 SV=1	63.83	20	11	11	782	85.7	89.29
A0A7E5X2F1	CabLooper	probable medium-chain specific acyl-CoA dehydrogenase, mitochondrial isoform X3 OS=Trichoplusia ni OX=7111 GN=LOC	50.977	34	11	11	422	46	87.61
A0A7E5WYR8	CabLooper	40S ribosomal protein S2 OS=Trichoplusia ni OX=7111 GN=LOC113507345 PE=3 SV=1	56.217	45	11	11	261	28	76.43
A0A7E5W6S4	CabLooper	Pyruvate dehydrogenase E1 component subunit alpha OS=Trichoplusia ni OX=7111 GN=LOC113499844 PE=4 SV=1	53.094	40	11	11	401	43.6	75.28
A0A7E5VD94	CabLooper	V-type proton ATPase subunit H OS=Trichoplusia ni OX=7111 GN=LOC113492751 PE=3 SV=1	43.369	26	11	11	474	55	64.39
A0A7E5VKT5	CabLooper	serine/threonine-protein phosphatase PP2A 65 kDa regulatory subunit OS=Trichoplusia ni OX=7111 GN=LOC113494680 P	55.572	26	11	11	590	65.2	63.47
A0A7E5X421	CabLooper	importin-7 isoform X2 OS=Trichoplusia ni OX=7111 GN=LOC113508374 PE=4 SV=1	44.324	20	11	11	1046	120.5	64.21
A0A7E5V9X1	CabLooper	Ubiquitin carboxyl-terminal hydrolase 7 OS=Trichoplusia ni OX=7111 GN=LOC113492018 PE=3 SV=1	37.867	16	11	11	1117	129.3	51.32
A0A7E5VP72	CabLooper	ATPase family AAA domain-containing protein 3A homolog OS=Trichoplusia ni OX=7111 GN=LOC113495539 PE=4 SV=1	36.283	20	11	11	630	70.6	42.93
A0A7E5X451	CabLooper	Glutaminyl-tRNA synthetase OS=Trichoplusia ni OX=7111 GN=LOC113508384 PE=3 SV=1	34.92	18	11	11	780	87.6	43.8
A0A7E5WCE0	CabLooper	cytoplasmic dynein 1 intermediate chain isoform X5 OS=Trichoplusia ni OX=7111 GN=LOC113501446 PE=3 SV=1	41.675	27	11	11	673	74.3	44.64
P54652		Heat shock-related 70 kDa protein 2 OS=Homo sapiens OX=9606 GN=HSPA2 PE=1 SV=1	75.895	17	10	1	639	70	152.07
A0A7E5WHI3	CabLooper	heat shock protein 70 A1-like OS=Trichoplusia ni OX=7111 GN=LOC113502297 PE=3 SV=1	58.997	22	10	4	630	69.6	109.34
A0A7E5WJR6	CabLooper	40S ribosomal protein S3 OS=Trichoplusia ni OX=7111 GN=LOC113502978 PE=3 SV=1	62.343	46	10	8	243	26.8	93.48
A0A7E5WLA2	CabLooper	heat shock protein 60A-like OS=Trichoplusia ni OX=7111 GN=LOC113503680 PE=3 SV=1	80.023	34	10	10	572	60.9	93.88
A0A7E5VUJ5	CabLooper	Eukaryotic translation initiation factor 3 subunit L OS=Trichoplusia ni OX=7111 GN=LOC113496841 PE=3 SV=1	56.693	29	10	10	543	63.4	76.08
A0A7E5X1G5	CabLooper	Eukaryotic translation initiation factor 3 subunit M OS=Trichoplusia ni OX=7111 GN=LOC113508254 PE=3 SV=1	44.307	39	10	10	386	44.1	63.9
A0A7E5WMW3	CabLooper	80 kDa nuclear cap-binding protein OS=Trichoplusia ni OX=7111 GN=LOC113503615 PE=3 SV=1	45.994	19	10	10	799	92.8	62.12
A0A7E5VT10	CabLooper	T-complex protein 1 subunit zeta OS=Trichoplusia ni OX=7111 GN=LOC113496326 PE=3 SV=1	50.835	29	10	10	531	57.9	65.31
A0A7E5VU81	CabLooper	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1 OS=Trichoplusia ni OX=7111 GN=LOC11349678	52.184	28	10	10	474	54.1	64.68
A0A7E5W744	CabLooper	RNA cytidine acetyltransferase OS=Trichoplusia ni OX=7111 GN=LOC113500035 PE=3 SV=1	52.704	20	10	10	986	109.3	62.76
A0A7E5X472	CabLooper	Inosine-5'-monophosphate dehydrogenase OS=Trichoplusia ni OX=7111 GN=LOC113508532 PE=3 SV=1	47.063	27	10	10	512	55.4	62.21
A0A7E5WTB8	CabLooper	WD repeat-containing protein 43 OS=Trichoplusia ni OX=7111 GN=LOC113505087 PE=4 SV=1	35.115	22	10	10	632	69.9	44.11
P14923		Junction plakoglobin OS=Homo sapiens OX=9606 GN=JUP PE=1 SV=3	37.417	17	10	10	745	81.7	45.47
A0A7E5VCU3	CabLooper	RNA helicase OS=Trichoplusia ni OX=7111 GN=LOC113492699 PE=3 SV=1	35.397	18	10	9	734	81.6	46.68
A0A7E5W0Q0	CabLooper	3-oxoacyl-[acyl-carrier-protein] reductase OS=Trichoplusia ni OX=7111 GN=LOC113498344 PE=4 SV=1	36.632	8	10	10	2362	252.7	51.12
A0A7E5VG59	CabLooper	importin subunit beta-1 OS=Trichoplusia ni OX=7111 GN=LOC113493477 PE=3 SV=1	32.658	18	10	10	885	98	39.58
Q9BUF5		Tubulin beta-6 chain OS=Homo sapiens OX=9606 GN=TUBB6 PE=1 SV=1	87.435	28	9	1	446	49.8	378.32
A0A7E5W030	CabLooper	protein LSM14 homolog A OS=Trichoplusia ni OX=7111 GN=LOC113498215 PE=3 SV=1	76.003	22	9	9	708	78.6	108.77
A0A7E5VGA7	CabLooper	probable 26S proteasome non-ATPase regulatory subunit 3 OS=Trichoplusia ni OX=7111 GN=LOC113493512 PE=3 SV=1	64.667	22	9	9	494	56.3	88
A0A7E5VLJ2	CabLooper	cAMP-dependent protein kinase type II regulatory subunit OS=Trichoplusia ni OX=7111 GN=LOC113494793 PE=3 SV=1	53.285	31	9	9	382	43.4	75.15
A0A7E5VVN0	CabLooper	T-complex protein 1 subunit delta OS=Trichoplusia ni OX=7111 GN=LOC113497142 PE=3 SV=1	79.043	33	9	9	541	56.9	96.73
A0A7E5VM91	CabLooper	RuvB-like helicase OS=Trichoplusia ni OX=7111 GN=LOC113495045 PE=3 SV=1	50.298	29	9	9	456	49.7	72.05
A9XXA7	CabLooper	Elongation factor 1-alpha OS=Trichoplusia ni OX=7111 GN=LOC113501794 PE=2 SV=1	44.903	33	9	9	463	50.3	62.32
A0A7E5VNS0	CabLooper	Coatomer subunit alpha OS=Trichoplusia ni OX=7111 GN=LOC113495443 PE=4 SV=1	44.132	13	9	9	1233	137.9	70.32
A0A7E5WVI5	CabLooper	metal transporter CNNM4-like OS=Trichoplusia ni OX=7111 GN=LOC113505583 PE=3 SV=1	39.781	13	9	9	928	104.4	53.9
A0A7E5W0U9	CabLooper	Glutamyl-tRNA synthetase OS=Trichoplusia ni OX=7111 GN=LOC113498448 PE=3 SV=1	40.649	9	9	9	1635	181.7	56.98
A0A7E5VW33	CabLooper	engulfment and cell motility protein 1 OS=Trichoplusia ni OX=7111 GN=LOC113497273 PE=4 SV=1	47.458	20	9	9	743	84.5	58.71

A0A7E5W7J7	CabLooper	ATP-binding cassette sub-family F member 2 OS=Trichoplusia ni OX=7111 GN=LOC113499837 PE=4 SV=1	40.644	18	9	9	621	70.7	49.3
A0A7E5WWA9	CabLooper	Heat shock protein 83 OS=Trichoplusia ni OX=7111 GN=LOC113506484 PE=3 SV=1	42.858	16	9	9	780	88.7	53.23
A0A7E5X1F1	CabLooper	heat shock protein 75 kDa, mitochondrial OS=Trichoplusia ni OX=7111 GN=LOC113508241 PE=3 SV=1	34.152	19	9	9	695	78.8	51.61
A0A7E5WAY5	CabLooper	Eukaryotic translation initiation factor 3 subunit H OS=Trichoplusia ni OX=7111 GN=LOC113501042 PE=3 SV=1	39.428	36	9	9	337	38.6	48.35
A0A7E5W5Z7	CabLooper	26S proteasome regulatory subunit 7 OS=Trichoplusia ni OX=7111 GN=LOC113499378 PE=3 SV=1	33.499	26	9	9	434	48.5	51.04
A0A7E5WDE9	CabLooper	von Willebrand factor A domain-containing protein 8 OS=Trichoplusia ni OX=7111 GN=LOC113501696 PE=4 SV=1	35.345	9	9	9	1395	156.3	48.81
A0A7E5VL44	CabLooper	nucleolar protein 56 OS=Trichoplusia ni OX=7111 GN=LOC113494788 PE=3 SV=1	50.25	34	9	9	459	51.2	57.72
A0A7E5VWI4	CabLooper	piwi-like protein Siwi OS=Trichoplusia ni OX=7111 GN=LOC113497222 PE=3 SV=1	27.713	12	9	9	903	101.8	38.85
A0A7E5W4P6	CabLooper	Alpha-mannosidase OS=Trichoplusia ni OX=7111 GN=LOC113499077 PE=3 SV=1	37.084	14	9	9	1156	132.5	51.2
A0A7E5WIE9	CabLooper	translation initiation factor eIF-2B subunit gamma OS=Trichoplusia ni OX=7111 GN=LOC113502516 PE=4 SV=1	37.494	31	9	9	459	51.2	49.92
A0A7E5W0E1	CabLooper	CCT-theta OS=Trichoplusia ni OX=7111 GN=LOC113498305 PE=3 SV=1	29.587	19	9	9	545	59.2	40.5
A0A7E5W4M1	CabLooper	cleavage and polyadenylation specificity factor subunit 1 OS=Trichoplusia ni OX=7111 GN=LOC113499249 PE=4 SV=1	36.718	12	9	9	1463	163	43.11
A0A7E5VK22	CabLooper	60 kDa SS-A/Ro ribonucleoprotein homolog OS=Trichoplusia ni OX=7111 GN=LOC113494479 PE=3 SV=1	32.683	20	9	9	634	72	35.22
P02535-1		SWISS-PROT:P02535-1 Tax_Id=10090 Gene_Symbol=Krt10 Isoform 1 of Keratin, type I cytoskeletal 10	43.975	15	8	1	570	57.7	109.47
A0A7E5WQV8	CabLooper	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase 48 kDa subunit OS=Trichoplusia ni OX=7111 GN=LOC1135	41.051	27	8	8	435	49	57.44
A0A7E5WZN5	CabLooper	Voltage-dependent anion-selective channel protein 3 OS=Trichoplusia ni OX=7111 GN=LOC113507651 PE=3 SV=1	44.542	39	8	8	285	30.4	60.81
A0A7E5X4Q3	CabLooper	Leucyl-tRNA synthetase OS=Trichoplusia ni OX=7111 GN=LOC113508534 PE=3 SV=1	42.409	11	8	8	1176	134	62.23
A0A7E5WVJ1	CabLooper	trifunctional enzyme subunit beta, mitochondrial-like OS=Trichoplusia ni OX=7111 GN=LOC113506203 PE=3 SV=1	43.531	26	8	8	474	50.9	59.46
A0A7E5X1U5	CabLooper	CCT-beta OS=Trichoplusia ni OX=7111 GN=LOC113508363 PE=3 SV=1	55.804	27	8	8	536	57.8	64.78
A0A7E5WLH0	CabLooper	ras GTPase-activating protein 1 OS=Trichoplusia ni OX=7111 GN=LOC113503316 PE=4 SV=1	41.728	17	8	8	932	104.6	64.15
A0A7E5W6R4	CabLooper	Coiled-coil domain-containing protein 47 OS=Trichoplusia ni OX=7111 GN=LOC113499597 PE=4 SV=1	55.52	31	8	8	465	52.8	62.74
A0A7E5VH01	CabLooper	Isocitrate dehydrogenase [NAD] subunit, mitochondrial OS=Trichoplusia ni OX=7111 GN=LOC113493780 PE=3 SV=1	43.319	29	8	2	389	42.3	56.18
A0A451FUM5	CabLooper	Actin OS=Trichoplusia ni OX=7111 GN=LOC113507353 PE=2 SV=1	39.159	34	8	1	376	41.8	48.62
Q92547		DNA topoisomerase 2-binding protein 1 OS=Homo sapiens OX=9606 GN=TOPBP1 PE=1 SV=3	42.478	9	8	8	1522	170.6	54.65
A0A7E5VKK1	CabLooper	116 kDa U5 small nuclear ribonucleoprotein component OS=Trichoplusia ni OX=7111 GN=LOC113494621 PE=3 SV=1	35.453	15	8	8	987	110.3	48.74
P60709		Actin, cytoplasmic 1 OS=Homo sapiens OX=9606 GN=ACTB PE=1 SV=1	31.203	34	8	1	375	41.7	40.87
A0A7E5VI16	CabLooper	vesicle-associated membrane protein/synaptobrevin-binding protein isoform X1 OS=Trichoplusia ni OX=7111 GN=LOC1134	27.42	37	8	8	278	31	37.51
A0A7E5WJN2	CabLooper	T-complex protein 1 subunit gamma OS=Trichoplusia ni OX=7111 GN=LOC113503333 PE=3 SV=1	49.485	24	8	8	541	59.1	46.16
A0A7E5VPM3	CabLooper	eukaryotic translation initiation factor 4 gamma 3 isoform X4 OS=Trichoplusia ni OX=7111 GN=LOC113495640 PE=3 SV=1	24.888	6	8	8	1742	192.7	35.71
A0A7E5VJJ6	CabLooper	phosphatidylinositide phosphatase SAC1-like OS=Trichoplusia ni OX=7111 GN=LOC113494355 PE=4 SV=1	32.398	21	8	8	599	67.7	38.79
A0A7E5VZE7	CabLooper	maternal protein tudor-like isoform X1 OS=Trichoplusia ni OX=7111 GN=LOC113498000 PE=4 SV=1	29.489	5	8	8	2393	268.5	38.74
A0A7E5VVX2	CabLooper	Tubulin gamma chain OS=Trichoplusia ni OX=7111 GN=LOC113497039 PE=3 SV=1	27.434	32	8	8	456	51.3	39.9
A0A7E5VCZ8	CabLooper	UDP-glucose:glycoprotein glucosyltransferase OS=Trichoplusia ni OX=7111 GN=LOC113492729 PE=3 SV=1	32.738	9	8	8	1569	177.1	42.02
A0A7E5WD79	CabLooper	Arginine kinase OS=Trichoplusia ni OX=7111 GN=LOC113501645 PE=3 SV=1	28.318	28	8	8	379	42.3	38
A0A7E5WA69	CabLooper	calcium-binding mitochondrial carrier protein Aralar1 isoform X1 OS=Trichoplusia ni OX=7111 GN=LOC113500848 PE=3 SV	23.687	18	8	8	682	76.1	27.7
A0A7E5W2V2	CabLooper	Sodium/potassium-transporting ATPase subunit alpha OS=Trichoplusia ni OX=7111 GN=LOC113498884 PE=3 SV=1	28.28	11	8	8	1036	114.4	34.17
Q9U6Y5		SWISS-PROT:Q9U6Y5 Green fluorescent protein (GFP-Cter-HisTag)	53.931	37	7	7	249	28.1	83
A0A7E5VTM9	CabLooper	uncharacterized protein LOC113496621 OS=Trichoplusia ni OX=7111 GN=LOC113496621 PE=3 SV=1	55.483	21	7	7	712	80.3	68.39
A0A7E5X1H7	CabLooper	Histone H4 OS=Trichoplusia ni OX=7111 GN=LOC113507807 PE=3 SV=1	27.495	23	7	7	218	24.3	42.89
A0A7E5VGL3	CabLooper	AP complex subunit beta OS=Trichoplusia ni OX=7111 GN=LOC113493578 PE=3 SV=1	49.474	14	7	7	911	99.8	72.11
A0A7E5VDN0	CabLooper	DNA replication licensing factor MCM7 OS=Trichoplusia ni OX=7111 GN=LOC113492876 PE=3 SV=1	45.8	16	7	7	719	81	54.93
A0A7E5VHB9	CabLooper	Isocitrate dehydrogenase [NAD] subunit, mitochondrial OS=Trichoplusia ni OX=7111 GN=LOC113493780 PE=3 SV=1	36.594	23	7	1	389	42.2	52.7
A0A7E5VBD8	CabLooper	Phosphate carrier protein, mitochondrial OS=Trichoplusia ni OX=7111 GN=LOC113492332 PE=3 SV=1	21.63	20	7	7	359	39.3	34.76
A0A7E5VSJ9	CabLooper	elongation factor 1-delta isoform X2 OS=Trichoplusia ni OX=7111 GN=LOC113496314 PE=3 SV=1	33.098	38	7	1	242	26.5	40.71
A0A7E5VSF0	CabLooper	probable elongation factor 1-delta isoform X1 OS=Trichoplusia ni OX=7111 GN=LOC113496314 PE=3 SV=1	31.46	31	7	1	271	29.6	39.87
A0A7E5VH18	CabLooper	Lon protease homolog, mitochondrial OS=Trichoplusia ni OX=7111 GN=LOC113493795 PE=3 SV=1	36.204	13	7	7	957	107.5	43.91
A0A7E5VTL8	CabLooper	cullin-associated NEDD8-dissociated protein 1 OS=Trichoplusia ni OX=7111 GN=LOC113496611 PE=3 SV=1	36.898	10	7	7	1256	138.3	42.91
A0A7E5VHF4	CabLooper	Clathrin heavy chain OS=Trichoplusia ni OX=7111 GN=LOC113493842 PE=3 SV=1	36.326	7	7	7	1691	192.2	48.68
A0A7E5X012	CabLooper	Polypeptide N-acetylgalactosaminyltransferase OS=Trichoplusia ni OX=7111 GN=LOC113507745 PE=3 SV=1	29.968	20	7	7	629	71.9	42.62
A0A7E5VV65	CabLooper	CCT-alpha OS=Trichoplusia ni OX=7111 GN=LOC113496858 PE=3 SV=1	30.142	18	7	7	554	59	47.44
A0A7E5W5F8	CabLooper	guanine nucleotide-binding protein subunit beta-like protein OS=Trichoplusia ni OX=7111 GN=LOC113499550 PE=4 SV=1	31.054	33	7	7	319	35.9	44.86
A0A7E5VSE4	CabLooper	elongation factor 1-gamma OS=Trichoplusia ni OX=7111 GN=LOC113496222 PE=4 SV=1	31.885	28	7	7	423	48.1	39.04
A0A7E5W8E1	CabLooper	40S ribosomal protein S16 isoform X2 OS=Trichoplusia ni OX=7111 GN=LOC113500369 PE=3 SV=1	15.057	37	7	7	151	17	25.56
A0A7E5VT50	CabLooper	Polyadenylate-binding protein OS=Trichoplusia ni OX=7111 GN=LOC113496349 PE=3 SV=1	27.247	19	7	7	611	68.2	35.9
A0A7E5VU78	CabLooper	protein hu-li tai shao isoform X5 OS=Trichoplusia ni OX=7111 GN=LOC113496708 PE=3 SV=1	17.596	17	7	7	701	77.4	37.48
A0A7E5W2Y6	CabLooper	neurochondrin homolog OS=Trichoplusia ni OX=7111 GN=LOC113498920 PE=3 SV=1	16.851	11	7	7	737	83	23.6
A0A7E5VL42	CabLooper	guanine nucleotide-binding protein G(O) subunit alpha isoform X1 OS=Trichoplusia ni OX=7111 GN=LOC113494761 PE=4 S	23.877	24	7	7	356	40.6	28.83
A0A7E5WDP5	CabLooper	atlastin-like isoform X1 OS=Trichoplusia ni OX=7111 GN=LOC113501762 PE=3 SV=1	29.408	13	7	7	770	86.9	32.44
A0A7E5WM26	CabLooper	Pyruvate carboxylase OS=Trichoplusia ni OX=7111 GN=LOC113503833 PE=4 SV=1	25.269	11	7	7	1201	131.7	36.76
A0A7E5VXY6	CabLooper	transportin-3 OS=Trichoplusia ni OX=7111 GN=LOC113497709 PE=4 SV=1	19.344	14	7	7	916	102.1	27.41
A0A7E5WQS1	CabLooper	Adipocyte plasma membrane-associated protein OS=Trichoplusia ni OX=7111 GN=LOC113504763 PE=3 SV=1	23.789	18	7	7	668	73.7	29.74
A0A7E5VQ68	CabLooper	Y-box factor homolog isoform X2 OS=Trichoplusia ni OX=7111 GN=LOC113495753 PE=4 SV=1	80.337	37	6	6	272	29.2	233.63
A0A7E5WMQ8	CabLooper	polyubiquitin-A OS=Trichoplusia ni OX=7111 GN=LOC113503581 PE=4 SV=1	23.499	70	6	6	1293	145.1	68.25
A2A4G1		TREMBL:A2A4G1 Tax_Id=10090 Gene_Symbol=Krt15 Keratin complex 1, acidic, gene 15	19.571	11	6	1	456	49.5	55.45
A0A7E5WFR3	CabLooper	Eukaryotic translation initiation factor 3 subunit E OS=Trichoplusia ni OX=7111 GN=LOC113502235 PE=3 SV=1	50.129	18	6	6	445	52	63.72
A0A7E5WWK3	CabLooper	Eukaryotic translation initiation factor 3 subunit F OS=Trichoplusia ni OX=7111 GN=LOC113506587 PE=3 SV=1	35.043	40	6	6	286	30.5	53.67
A0A7E5VBX4	CabLooper	60S ribosomal protein L5 OS=Trichoplusia ni OX=7111 GN=LOC113492511 PE=3 SV=1	41.426	31	6	6	299	34.1	57.8
A0A7E5WK34	CabLooper	myelin expression factor 2 isoform X6 OS=Trichoplusia ni OX=7111 GN=LOC113503358 PE=4 SV=1	40.806	21	6	6	532	57.8	61.82
A0A7E5VTA1	CabLooper	Pyruvate dehydrogenase E1 component subunit beta OS=Trichoplusia ni OX=7111 GN=LOC113496524 PE=4 SV=1	43.944	27	6	6	367	39.7	52.07
A0A7E5WP50	CabLooper	Ribosomal protein S6 kinase OS=Trichoplusia ni OX=7111 GN=LOC113504422 PE=3 SV=1	39.073	22	6	6	452	50.7	47.41
A0A7E5WI92	CabLooper	Glyoxylate reductase 1 homolog OS=Trichoplusia ni OX=7111 GN=LOC113502935 PE=3 SV=1	43.262	22	6	6	527	57.3	52.8
A0A7E5WRU4	CabLooper	Coatomer subunit delta OS=Trichoplusia ni OX=7111 GN=LOC113505136 PE=3 SV=1	25.27	14	6	6	505	56.1	34.62
A0A7E5VSM9	CabLooper	sequestosome-1-like isoform X3 OS=Trichoplusia ni OX=7111 GN=LOC113496361 PE=4 SV=1	44.218	19	6	6	542	58.6	49.78
A0A7E5WGS1	CabLooper	sorting and assembly machinery component 50 homolog A-like OS=Trichoplusia ni OX=7111 GN=LOC113502502 PE=3 SV	40.61	23	6	6	468	49.8	51.21
A0A7E5VBG0	CabLooper	H/ACA ribonucleoprotein complex subunit 4-like OS=Trichoplusia ni OX=7111 GN=LOC113492352 PE=3 SV=1	29.449	16	6	6	514	57.5	39.44
A0A7E5WZQ1	CabLooper	T-complex protein 1 subunit eta OS=Trichoplusia ni OX=7111 GN=LOC113507248 PE=3 SV=1	22.122	15	6	6	544	58.9	33.81
A0A7E5W5U4	CabLooper	UDP-glucuronosyltransferase OS=Trichoplusia ni OX=7111 GN=LOC113499669 PE=3 SV=1	38.778	25	6	6	520	59.2	39.85
A0A7E5WJL7	CabLooper	60S acidic ribosomal protein P2 OS=Trichoplusia ni OX=7111 GN=LOC113502816 PE=3 SV=1	27.387	60	6	6	111	11.4	36.08
A0A7E5VU51	CabLooper	glomulin-like isoform X1 OS=Trichoplusia ni OX=7111 GN=LOC113496685 PE=4 SV=1	23.248	14	6	6	584	66.3	32.28
A0A7E5VFH6	CabLooper	Mitochondrial import inner membrane translocase subunit Tim21 OS=Trichoplusia ni OX=7111 GN=LOC113493317 PE=3 S	19.922	34	6	6	223	25.4	28.06
A0A7E5X2D5	CabLooper	hypoxia up-regulated protein 1 OS=Trichoplusia ni OX=7111 GN=LOC113508064 PE=3 SV=1	24.103	9	6	6	913	101.9	26.7
A0A7E5X256	CabLooper	Dynamin-1-like protein OS=Trichoplusia ni OX=7111 GN=LOC113508433 PE=4 SV=1	24.951	15	6	6	700	78.7	30.93
A0A7E5VJT9	CabLooper	protein 4.1 homolog isoform X2 OS=Trichoplusia ni OX=7111 GN=LOC113494423 PE=4 SV=1	19.413	3	6	6	3019	327.9	26.48
A0A7E5VNH5	CabLooper	40S ribosomal protein S4 OS=Trichoplusia ni OX=7111 GN=LOC113495374 PE=3 SV=1	19.266	25	6	6	263	29.5	22.01
A0A7E5W6Q4	CabLooper	E3 ubiquitin-protein ligase KCMF1 OS=Trichoplusia ni OX=7111 GN=LOC113499596 PE=4 SV=1	34.1	24	6	6	373	39.4	34.65
A0A7E5WC73	CabLooper	nodal modulator 3 OS=Trichoplusia ni OX=7111 GN=LOC113501382 PE=4 SV=1	22.587	10	6	6	1139	121.8	33.68
A0A7E5VE43	CabLooper	peroxiredoxin 1 isoform X3 OS=Trichoplusia ni OX=7111 GN=LOC113493007 PE=3 SV=1	15.504	36	6	6	195	21.9	25.4
A0A7E5WDM1	CabLooper	lamin-C-like OS=Trichoplusia ni OX=7111 GN=LOC113501742 PE=3 SV=1	21.849	12	6	6	614	70.1	26.01
A0A7E5VRX5	CabLooper	speckle targeted PIP5K1A-regulated poly(A) polymerase-like OS=Trichoplusia ni OX=7111 GN=LOC113496117 PE=4 SV=1	28.591	13	6	6	839	92.3	29.92
A0A7E5WNR7	CabLooper	paraplegin OS=Trichoplusia ni OX=7111 GN=LOC113504325 PE=3 SV=1	23.21	14	6	6	783	87.5	32.08
A0A7E5X3M5	CabLooper	40S ribosomal protein S10 OS=Trichoplusia ni OX=7111 GN=LOC113508227 PE=3 SV=1	22.928	30	6	6	158	18	28.08
A0A7E5X4I5	CabLooper	NADH-cytochrome b5 reductase OS=Trichoplusia ni OX=7111 GN=LOC113508463 PE=3 SV=1	15.936	25	6	6	314	35.5	22.63
A0A7E5X2B8	CabLooper	Histone H2B OS=Trichoplusia ni OX=7111 GN=LOC113507804 PE=3 SV=1	45.391	36	5	5	124	13.8	216.94
ENSEMBL:ENSBTAP00000025		(Bos taurus) hypothetical protein	38.165	9	5	1	536	59.3	139.09
A0A7E5WZY0	CabLooper	Histone H2A OS=Trichoplusia ni OX=7111 GN=LOC113507348 PE=3 SV=1	74.157	60	5	3	124	13.3	183.72
Q7Z794		Keratin, type II cytoskeletal 1b OS=Homo sapiens OX=9606 GN=KRT77 PE=1 SV=3	25.295	12	5	4	578	61.9	71.46
P17066		Heat shock 70 kDa protein 6 OS=Homo sapiens OX=9606 GN=HSPA6 PE=1 SV=2	20.47	8	5	1	643	71	53.34
A0A7E5VDG0	CabLooper	stromal cell-derived factor 2 OS=Trichoplusia ni OX=7111 GN=LOC113492843 PE=4 SV=1	50.364	42	5	5	225	24.7	80.24
A0A7E5WSZ1	CabLooper	Sidoreflexin OS=Trichoplusia ni OX=7111 GN=LOC113505284 PE=3 SV=1	54.77	30	5	5	327	36.8	64.81
A0A7E5VVQ7	CabLooper	congested-like trachea protein OS=Trichoplusia ni OX=7111 GN=LOC113497194 PE=3 SV=1	37.395	32	5	5	298	32.5	53.84

A0A7E5VVW4	CabLooper	mitochondrial 2-oxoglutarate/malate carrier protein-like OS=Trichoplusia ni OX=7111 GN=LOC113497226 PE=3 SV=1	46.04	26	5	5	293	31.6	54.61
A0A7E5VSR8	CabLooper	Vesicle-fusing ATPase OS=Trichoplusia ni OX=7111 GN=LOC113496395 PE=3 SV=1	25.341	14	5	5	440	49.2	29.84
A0A7E5WGA1	CabLooper	protein transport protein Sec61 subunit alpha OS=Trichoplusia ni OX=7111 GN=LOC113502385 PE=3 SV=1	31.947	16	5	5	476	52	40.22
A0A7E5WRU3	CabLooper	uncharacterized protein LOC113505025 OS=Trichoplusia ni OX=7111 GN=LOC113505025 PE=4 SV=1	29.459	27	5	5	270	29.9	38.82
A0A7E5VU84	CabLooper	Histone deacetylase OS=Trichoplusia ni OX=7111 GN=LOC113496752 PE=3 SV=1	25.045	22	5	5	486	55.4	43.83
A0A7E5WVQ8	CabLooper	ATP synthase subunit beta OS=Trichoplusia ni OX=7111 GN=LOC113505642 PE=3 SV=1	30.878	11	5	1	652	69.9	42.01
A0A7E5WY31	CabLooper	40S ribosomal protein S17 OS=Trichoplusia ni OX=7111 GN=LOC113506555 PE=3 SV=1	24.098	35	5	5	133	15.4	31.45
A0A7E5WP10	CabLooper	endoplasmic reticulum resident protein 44 isoform X2 OS=Trichoplusia ni OX=7111 GN=LOC113504396 PE=4 SV=1	16.729	14	5	5	505	58.6	27.41
A0A7E5W2S7	CabLooper	40S ribosomal protein S3a OS=Trichoplusia ni OX=7111 GN=LOC113498677 PE=3 SV=1	23.208	23	5	5	264	29.9	28.23
A0A7E5V9S2	CabLooper	26S proteasome non-ATPase regulatory subunit 12 OS=Trichoplusia ni OX=7111 GN=LOC113491986 PE=3 SV=1	31.544	22	5	5	450	51.5	35.09
A0A7E5WE99	CabLooper	dnaJ protein homolog 1 OS=Trichoplusia ni OX=7111 GN=LOC113501513 PE=4 SV=1	18.769	15	5	5	352	39	26.18
A0A7E5VXD0	CabLooper	uncharacterized protein LOC113497545 OS=Trichoplusia ni OX=7111 GN=LOC113497545 PE=3 SV=1	21.542	48	5	5	127	13.8	26.74
P04406		Glyceraldehyde-3-phosphate dehydrogenase OS=Homo sapiens OX=9606 GN=GAPDH PE=1 SV=3	18.62	24	5	5	335	36	27.99
A0A7E5VFM7	CabLooper	RNA-binding protein squid isoform X1 OS=Trichoplusia ni OX=7111 GN=LOC113493348 PE=4 SV=1	27.406	17	5	5	328	34.8	29.63
A0A7E5WZK0	CabLooper	60S ribosomal protein L23a OS=Trichoplusia ni OX=7111 GN=LOC113507000 PE=3 SV=1	23.025	22	5	5	298	31.2	25.96
A0A7E5VQE8	CabLooper	Prohibitin OS=Trichoplusia ni OX=7111 GN=LOC113495703 PE=3 SV=1	25.884	28	5	5	299	33.1	29.08
A0A7E5WNU3	CabLooper	ATP-dependent zinc metalloprotease YME1 homolog isoform X3 OS=Trichoplusia ni OX=7111 GN=LOC113504345 PE=3 S	29.948	15	5	5	767	84.6	39.39
A0A7E5WXV7	CabLooper	LOW QUALITY PROTEIN: NADH-ubiquinone oxidoreductase 75 kDa subunit, mitochondrial-like OS=Trichoplusia ni OX=711	27.133	15	5	5	627	68.9	35.16
A0A7E5W7I0	CabLooper	nucleosome assembly protein 1-like 1 isoform X1 OS=Trichoplusia ni OX=7111 GN=LOC113499831 PE=3 SV=1	14.188	16	5	5	399	45.6	25.03
A0A7E5VEM6	CabLooper	ADP-ribosylation factor 2 OS=Trichoplusia ni OX=7111 GN=LOC113493088 PE=3 SV=1	20.319	42	5	2	180	20.4	31.63
A0A7E5W3F8	CabLooper	Carnitine O-palmitoyltransferase OS=Trichoplusia ni OX=7111 GN=LOC113498996 PE=3 SV=1	19.011	10	5	5	777	89.5	25.53
A0A7E5WZ01	CabLooper	dnaJ homolog subfamily C member 11-like OS=Trichoplusia ni OX=7111 GN=LOC113507444 PE=4 SV=1	13.803	10	5	5	576	65.3	20.62
A0A7E5VAX0	CabLooper	LOW QUALITY PROTEIN: uncharacterized protein LOC113492244 OS=Trichoplusia ni OX=7111 GN=LOC113492244 PE=3	20.903	7	5	5	961	102.9	23.25
A0A7E5V8A8	CabLooper	ran GTPase-activating protein 1 isoform X1 OS=Trichoplusia ni OX=7111 GN=LOC113491648 PE=4 SV=1	25.034	15	5	5	600	63.9	26.97
A0A7E5X5U8	CabLooper	60S ribosomal protein L4 OS=Trichoplusia ni OX=7111 GN=LOC113508986 PE=3 SV=1	19.147	17	5	5	435	48.2	23.43
A0A7E5WWZ6	CabLooper	cathepsin L-like OS=Trichoplusia ni OX=7111 GN=LOC113506377 PE=3 SV=1	17.552	24	5	5	341	38.1	23.32
A0A7E5VE87	CabLooper	Importin subunit alpha OS=Trichoplusia ni OX=7111 GN=LOC113493033 PE=3 SV=1	20.859	16	5	5	515	56.5	25.59
A0A7E5VDL2	CabLooper	40S ribosomal protein S8 OS=Trichoplusia ni OX=7111 GN=LOC113492881 PE=3 SV=1	17.097	28	5	5	208	23.8	22.23
A0A7E5VVL0	CabLooper	serine palmitoyltransferase 2 OS=Trichoplusia ni OX=7111 GN=LOC113497126 PE=3 SV=1	15.675	13	5	5	544	59.3	21.61
A0A7E5WQ48	CabLooper	TBC1 domain family member 9 OS=Trichoplusia ni OX=7111 GN=LOC113504654 PE=4 SV=1	20.475	7	5	5	1137	128.7	24.46
A0A7E5VEH7	CabLooper	Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial OS=Trichoplusia ni OX=7111 GN=LOC11349306	14.797	15	5	5	656	71.5	24
A0A7E5VAS2	CabLooper	myosin heavy chain, non-muscle isoform X1 OS=Trichoplusia ni OX=7111 GN=LOC113492215 PE=3 SV=1	14.5	4	5	5	1970	225.2	24.65
A0A7E5VYD3	CabLooper	ATPase family AAA domain-containing protein 1-like isoform X2 OS=Trichoplusia ni OX=7111 GN=LOC113497783 PE=3 S	16.671	22	5	5	381	42.9	23.39
A0A7E5VB26	CabLooper	charged multivesicular body protein 7 isoform X2 OS=Trichoplusia ni OX=7111 GN=LOC113492259 PE=3 SV=1	23.372	13	5	5	626	69.6	30.24
A0A7E5VCW2	CabLooper	Proton-translocating NAD(P)(+) transhydrogenase OS=Trichoplusia ni OX=7111 GN=LOC113492713 PE=3 SV=1	26	9	5	5	1048	109.1	29.69
A0A7E5W7A8	CabLooper	EH domain-containing protein 3 OS=Trichoplusia ni OX=7111 GN=LOC113499725 PE=4 SV=1	11.545	17	5	5	532	60.9	18.33
A0A7E5VHD5	CabLooper	3'-5' RNA helicase YTHDC2-like OS=Trichoplusia ni OX=7111 GN=LOC113493794 PE=4 SV=1	18.766	8	5	5	1368	152.9	24.53
A0A7E5VUR8	CabLooper	WD repeat-containing protein 55 homolog OS=Trichoplusia ni OX=7111 GN=LOC113496757 PE=4 SV=1	14.198	7	5	5	891	100	16.97
A0A7E5W6Y9	CabLooper	ankyrin repeat and LEM domain-containing protein 2-like OS=Trichoplusia ni OX=7111 GN=LOC113499885 PE=3 SV=1	18.96	10	5	5	874	97	21.32
A0A7E5W1Q7	CabLooper	Replication protein A subunit OS=Trichoplusia ni OX=7111 GN=LOC113498438 PE=3 SV=1	12.135	11	5	5	603	66.4	20.05
A0A7E5VX36	CabLooper	ATPase family protein 2 homolog OS=Trichoplusia ni OX=7111 GN=LOC113497517 PE=4 SV=1	20.328	12	5	5	744	83.1	23.76
A0A7E5X1B5	CabLooper	CTP synthase OS=Trichoplusia ni OX=7111 GN=LOC113507725 PE=3 SV=1	11.094	11	5	5	603	67.2	17.9
A0A7E5VSI2	CabLooper	prostatic acid phosphatase isoform X1 OS=Trichoplusia ni OX=7111 GN=LOC113496175 PE=4 SV=1	15.504	19	5	5	420	48.3	19.7
A0A7E5X5W2	CabLooper	RuvB-like helicase OS=Trichoplusia ni OX=7111 GN=LOC113508933 PE=3 SV=1	12.98	13	5	5	466	51.2	16.57
A0A7E5WVH9	CabLooper	interleukin enhancer-binding factor 2 homolog isoform X1 OS=Trichoplusia ni OX=7111 GN=LOC113505581 PE=4 SV=1	15.569	19	5	5	388	41.9	18.58
P29508		Serpin B3 OS=Homo sapiens OX=9606 GN=SERPINB3 PE=1 SV=2	15.497	18	5	5	390	44.5	20.07
A0A7E5WSR5	CabLooper	DNA polymerase epsilon catalytic subunit OS=Trichoplusia ni OX=7111 GN=LOC113505019 PE=3 SV=1	17.266	4	5	5	2233	254.3	21.98
A0A7E5V897	CabLooper	nuclear pore membrane glycoprotein 210 OS=Trichoplusia ni OX=7111 GN=LOC113491631 PE=4 SV=1	10.632	6	5	5	1714	184.6	24.22
A0A7E5VAW0	CabLooper	Nuclear pore complex protein Nup85 OS=Trichoplusia ni OX=7111 GN=LOC113492237 PE=3 SV=1	15.049	12	5	5	652	72.5	20.51
A0A7E5WV62	CabLooper	sphingomyelin phosphodiesterase 4 isoform X1 OS=Trichoplusia ni OX=7111 GN=LOC113506064 PE=4 SV=1	12.599	7	5	5	777	89.4	16.41
A0A7E5WF42	CabLooper	Translocon-associated protein subunit beta OS=Trichoplusia ni OX=7111 GN=LOC113502129 PE=3 SV=1	53.474	39	4	4	190	20.9	71.27
A0A7E5VDW5	CabLooper	60S acidic ribosomal protein P1 OS=Trichoplusia ni OX=7111 GN=LOC113492912 PE=3 SV=1	49.061	54	4	4	112	11.3	56.54
A0A7E5VSB5	CabLooper	heat shock protein 68-like OS=Trichoplusia ni OX=7111 GN=LOC113496267 PE=3 SV=1	27.4	11	4	1	649	70.8	45.82
A0A7E5WBC0	CabLooper	40S ribosomal protein S14 OS=Trichoplusia ni OX=7111 GN=LOC113500782 PE=3 SV=1	29.882	30	4	4	151	16.2	39.3
A0A7E5W943	CabLooper	cleavage and polyadenylation specificity factor subunit CG7185 OS=Trichoplusia ni OX=7111 GN=LOC113500207 PE=3 S	35.444	10	4	4	745	80.9	40.71
A0A7E5VEF7	CabLooper	TBC1 domain family member 20 OS=Trichoplusia ni OX=7111 GN=LOC113493075 PE=4 SV=1	28.192	22	4	4	427	49	38.65
P06702		Protein S100-A9 OS=Homo sapiens OX=9606 GN=S100A9 PE=1 SV=1	21.491	45	4	4	114	13.2	34.16
A0A7E5WE69	CabLooper	Prohibitin OS=Trichoplusia ni OX=7111 GN=LOC113501490 PE=3 SV=1	29.438	24	4	4	274	30.1	37.53
A0A7E5WCL4	CabLooper	26S proteasome non-ATPase regulatory subunit 7 OS=Trichoplusia ni OX=7111 GN=LOC113501489 PE=3 SV=1	31.439	26	4	4	332	37.6	33.74
A0A7E5WBY2	CabLooper	zinc transporter ZIP11-like OS=Trichoplusia ni OX=7111 GN=LOC113501271 PE=4 SV=1	25.856	21	4	4	353	37.6	34.21
A0A7E5V9S5	CabLooper	Elongation factor 1-alpha OS=Trichoplusia ni OX=7111 GN=LOC113491975 PE=4 SV=1	25.087	10	4	4	650	69.9	27.67
A0A7E5WHE3	CabLooper	glucose-induced degradation protein 8 homolog isoform X2 OS=Trichoplusia ni OX=7111 GN=LOC113502272 PE=4 SV=1	24.412	33	4	4	233	26.3	32.64
A0A7E5W1E2	CabLooper	replication factor C subunit 5 OS=Trichoplusia ni OX=7111 GN=LOC113498508 PE=3 SV=1	26.97	21	4	4	333	37	32.38
A0A7E5VXI9	CabLooper	acyl-CoA Delta(11) desaturase-like OS=Trichoplusia ni OX=7111 GN=LOC113497645 PE=3 SV=1	27.358	16	4	4	353	40.2	32.71
P02769		SWISS-PROT:P02769 (Bos taurus) Bovine serum albumin precursor	13.809	9	4	4	607	69.2	25.85
A0A7E5W145	CabLooper	ras-related protein ORAB-1 OS=Trichoplusia ni OX=7111 GN=LOC113498457 PE=4 SV=1	28.474	30	4	4	203	22.5	33.22
P47929		Galectin-7 OS=Homo sapiens OX=9606 GN=LGALS7 PE=1 SV=2	17.901	39	4	4	136	15.1	22.97
A0A7E5WLI9	CabLooper	heterogeneous nuclear ribonucleoprotein 27C isoform X4 OS=Trichoplusia ni OX=7111 GN=LOC113503328 PE=4 SV=1	29.303	15	4	4	472	49.3	32.48
A0A7E5WL69	CabLooper	ubiquilin-1 isoform X1 OS=Trichoplusia ni OX=7111 GN=LOC113503585 PE=4 SV=1	37.194	19	4	4	528	57	39.25
A0A7E5W1S2	CabLooper	juvenile hormone esterase-like isoform X2 OS=Trichoplusia ni OX=7111 GN=LOC113498447 PE=4 SV=1	25.159	15	4	4	705	78.1	35.26
A0A7E5VDK8	CabLooper	integrator complex subunit 4 OS=Trichoplusia ni OX=7111 GN=LOC113492844 PE=4 SV=1	18.303	9	4	4	969	106	24.77
A0A7E5VEX4	CabLooper	Aspartate aminotransferase OS=Trichoplusia ni OX=7111 GN=LOC113493187 PE=3 SV=1	17.019	14	4	4	429	48	24.19
Q02413		Desmoglein-1 OS=Homo sapiens OX=9606 GN=DSG1 PE=1 SV=2	21.237	8	4	4	1049	113.7	29.24
A0A7E5WBC6	CabLooper	60S ribosomal protein L24 OS=Trichoplusia ni OX=7111 GN=LOC113500785 PE=3 SV=1	13.409	26	4	4	155	17.4	17.79
A0A7E5WXF7	CabLooper	LOW QUALITY PROTEIN: importin-5-like OS=Trichoplusia ni OX=7111 GN=LOC113506199 PE=4 SV=1	24.716	10	4	4	784	87.4	31.15
A0A7E5WGA9	CabLooper	Coatomer subunit gamma OS=Trichoplusia ni OX=7111 GN=LOC113502389 PE=3 SV=1	24.098	8	4	3	863	95.2	29.39
A0A7E5W3B8	CabLooper	ADP-ribosylation factor 1 OS=Trichoplusia ni OX=7111 GN=LOC113498854 PE=3 SV=1	18.781	31	4	1	182	20.7	26.8
A0A7E5VW85	CabLooper	ABC transporter C family member 13 OS=Trichoplusia ni OX=7111 GN=LOC113497169 PE=4 SV=1	18.773	6	4	4	1596	177.3	21.68
A0A7E5WLD3	CabLooper	2-oxoglutarate and iron-dependent oxygenase domain-containing protein 3-like OS=Trichoplusia ni OX=7111 GN=LOC1135	11.155	22	4	4	320	36.7	18.61
A0A7E5VGX8	CabLooper	ATP-dependent Clp protease ATP-binding subunit clpX-like, mitochondrial isoform X4 OS=Trichoplusia ni OX=7111 GN=LO	9.784	8	4	4	579	63.4	16.89
A0A7E5VK81	CabLooper	protein purity of essence OS=Trichoplusia ni OX=7111 GN=LOC113494523 PE=3 SV=1	21.926	2	4	4	5261	574.9	26.3
A0A7E5V9W6	CabLooper	40S ribosomal protein SA OS=Trichoplusia ni OX=7111 GN=LOC113492011 PE=3 SV=1	19.552	24	4	4	307	33.5	22.58
A0A7E5WMG0	CabLooper	Mitochondrial import inner membrane translocase subunit TIM50 OS=Trichoplusia ni OX=7111 GN=LOC113503506 PE=3 S	13.898	18	4	4	386	45.1	21.94
A0A7E5V881	CabLooper	neurofilament heavy polypeptide isoform X1 OS=Trichoplusia ni OX=7111 GN=LOC113491627 PE=3 SV=1	19.663	26	4	4	370	40.8	25.59
A0A7E5W0E2	CabLooper	Calcium-transporting ATPase OS=Trichoplusia ni OX=7111 GN=LOC113498289 PE=3 SV=1	16.015	6	4	4	1129	124.5	21.22
A0A7E5VGM4	CabLooper	uncharacterized protein LOC113493678 isoform X1 OS=Trichoplusia ni OX=7111 GN=LOC113493678 PE=4 SV=1	15.083	19	4	4	369	40.2	23.8
A0A7E5W9P6	CabLooper	CDGSH iron-sulfur domain-containing protein 2 homolog OS=Trichoplusia ni OX=7111 GN=LOC113500730 PE=3 SV=1	14.508	50	4	4	132	14.7	18.71
A0A7E5WP71	CabLooper	Pre-mRNA-processing factor 19 OS=Trichoplusia ni OX=7111 GN=LOC113504409 PE=3 SV=1	23.805	20	4	4	499	54.3	24.05
A0A7E5VVP8	CabLooper	NPC intracellular cholesterol transporter 1 isoform X4 OS=Trichoplusia ni OX=7111 GN=LOC113497190 PE=3 SV=1	10.321	4	4	4	1320	145.5	18.42
A0A7E5X1S9	CabLooper	Dynein light intermediate chain OS=Trichoplusia ni OX=7111 GN=LOC113507896 PE=3 SV=1	17.572	12	4	4	469	51.9	19.47
A0A7E5WQL7	CabLooper	5'-3' exoribonuclease OS=Trichoplusia ni OX=7111 GN=LOC113504793 PE=3 SV=1	13.03	6	4	4	987	112.2	22.46
A0A7E5VAH3	CabLooper	transducin beta-like protein 3 OS=Trichoplusia ni OX=7111 GN=LOC113492152 PE=4 SV=1	21.032	10	4	4	793	87.9	26.32
A0A7E5WH80	CabLooper	ATP-binding cassette sub-family D member 3 OS=Trichoplusia ni OX=7111 GN=LOC113502384 PE=4 SV=1	15.226	10	4	4	669	76.4	17.62
A0A7E5VVS0	CabLooper	mitochondrial import receptor subunit TOM70 OS=Trichoplusia ni OX=7111 GN=LOC113497203 PE=4 SV=1	11.279	11	4	4	547	61.7	15.04
A0A7E5VJN4	CabLooper	KH domain-containing, RNA-binding, signal transduction-associated protein 2-like OS=Trichoplusia ni OX=7111 GN=LOC11	15.314	18	4	4	357	40.6	16.42
A0A7E5X4D0	CabLooper	eukaryotic translation initiation factor 4 gamma 2 OS=Trichoplusia ni OX=7111 GN=LOC113508571 PE=4 SV=1	16.491	10	4	4	730	80	17.49
A0A7E5VGR7	CabLooper	Cytochrome c oxidase subunit 4 OS=Trichoplusia ni OX=7111 GN=LOC113493724 PE=3 SV=1	11.489	34	4	4	177	20.2	13.44
A0A7E5WPH0	CabLooper	GTP-binding protein 1 isoform X1 OS=Trichoplusia ni OX=7111 GN=LOC113504522 PE=4 SV=1	14.211	12	4	4	619	67.9	17.7
A0A7E5VI30	CabLooper	insulin-like growth factor 2 mRNA-binding protein 1 isoform X1 OS=Trichoplusia ni OX=7111 GN=LOC113494044 PE=3 SV=	14.006	11	4	4	536	58.4	16.08
A0A7E5WV74	CabLooper	2-arachidonoylglycerol hydrolase ABHD12 OS=Trichoplusia ni OX=7111 GN=LOC113505513 PE=3 SV=1	12.242	18	4	4	373	42.2	18.86

Table 24. Mass spectrometry analysis by Dr Lu Yu. Identification of proteins from MRN purification, as shown in Coomassie stained gel from the results section.



Band 6

Protein FDR Confidence	# AAs	MW [kDa]	Description	# Peptides	# Unique Peptides	# PSMs	Sum PEP Score	Coverage [%]
High	1312	153.8	DNA repair protein RAD50	22	22	35	38.13	20
High	754	84.9	Nibrin	13	13	32	29.211	25
High	2458	276.4	Acetyl-CoA carboxylase 2	5	5	19	14.977	3

Band 5

Protein FDR Confidence	# AAs	MW [kDa]	Description	# Peptides	# Unique Peptides	# PSMs	Sum PEP Score	Coverage [%]
High	1312	153.8	DNA repair protein RAD50	89	89	510	450.669	52
High	754	84.9	Nibrin	18	18	73	89.032	30
High	708	80.5	Double-strand break repair protein MRE11	4	4	4	6.556	9
High	646	70.9	Heat shock cognate 71 kDa protein	3	3	3	6.094	6

Band 4

Protein FDR Confidence	# AAs	MW [kDa]	Description	# Peptides	# Unique Peptides	# PSMs	Sum PEP Score	Coverage [%]
High	754	84.9	Nibrin	55	55	606	534.8	67
High	708	80.5	Double-strand break repair protein MRE11	14	14	30	54.9	32
High	1312	153.8	DNA repair protein RAD50	13	13	23	48.1	16

Band 3

Protein FDR Confidence	# AAs	MW [kDa]	Description	# Peptides	# Unique Peptides	# PSMs	Sum PEP Score	Coverage [%]
High	708	80.5	Double-strand break repair protein MRE11	41	41	100	135.783	54
High	754	84.9	Nibrin	37	37	192	189.266	53
High	1312	153.8	DNA repair protein RAD50	22	22	44	71.614	18
High	646	70.9	Heat shock cognate 71 kDa protein	8	4	13	17.375	11
High	450	50.4	Tubulin beta-3 chain	4	4	4	4.969	11
High	641	70.3	Heat shock 70 kDa protein 1-like	8	4	28	22.923	10

Band 2

Protein FDR Confidence	# AAs	MW [kDa]	Description	# Peptides	# Unique Peptides	# PSMs	Sum PEP Score	Coverage [%]
High	708	80.5	Double-strand break repair protein MRE11	30	30	80	112.5	48
High	754	84.9	Nibrin	21	21	81	120.4	32
High	646	70.9	Heat shock cognate 71 kDa protein	18	8	141	113.3	25
High	639	70	Heat shock-related 70 kDa protein 2	12	2	79	75.0	22
High	1312	153.8	DNA repair protein RAD50	17	17	33	59.0	17
High	641	70.3	Heat shock 70 kDa protein 1-like	11	5	151	82.6	13
High	641	70	Heat shock 70 kDa protein 1B	7	1	28	29.5	12
High	643	71	Heat shock 70 kDa protein 6	9	4	46	51.7	11
High	445	49.8	Tubulin beta-4B chain	4	1	4	6.8	11
High	654	72.3	Endoplasmic reticulum chaperone BiP	7	5	32	47.9	11
High	445	49.9	Tubulin beta-2A chain	4	1	4	6.0	10
High	679	73.6	Stress-70 protein, mitochondrial	5	5	25	31.1	7

Band 1

Protein FDR Confidence	# AAs	MW [kDa]	Description	# Peptides	# Unique Peptides	# PSMs	Sum PEP Score	Coverage [%]
High	445	49.9	Tubulin beta-2A chain	21	6	81	88.6	52
High	445	49.8	Tubulin beta-4B chain	18	1	78	83.4	49
High	444	49.6	Tubulin beta-4A chain	16	3	66	59.5	46
High	754	84.9	Nibrin	36	36	210	216.6	44
High	450	49.9	Tubulin alpha-3D chain	18	2	50	62.3	40
High	451	50.1	Tubulin alpha-1A chain	17	1	43	57.1	39
High	450	50.4	Tubulin beta-3 chain	15	3	66	63.0	32
High	449	50.1	Tubulin alpha-8 chain	10	1	30	36.5	27
High	444	49.7	Tubulin beta-8 chain	10	3	39	35.2	26
High	708	80.5	Double-strand break repair protein MRE11	9	9	15	18.6	16
High	529	56.5	ATP synthase subunit beta, mitochondrial	3	3	4	7.5	10
High	1312	153.8	DNA repair protein RAD50	10	10	16	32.7	8
High	1049	113.7	Desmoglein-1	4	4	5	6.4	6
High	553	59.7	ATP synthase subunit alpha, mitochondrial	3	3	4	5.1	6
High	2871	331.6	Desmoplakin	4	4	4	6.4	2

Bibliography

Ahnesorg, P., Smith, P., & Jackson, S. P. (2006). XLF interacts with the XRCC4-DNA Ligase IV complex to promote DNA nonhomologous end-joining. *Cell*. https://doi.org/10.1016/j.cell.2005.12.031

Ait Saada, A., Costa, A. B., Sheng, Z., Guo, W., Haber, J. E., & Lobachev, K. S. (2021). Structural parameters of palindromic repeats determine the specificity of nuclease attack of secondary structures. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkab168

Akagawa, R., Trinh, H. T., Saha, L. K., Tsuda, M., Hirota, K., Yamada, S., … Sasanuma, H. (2020). UBC13-Mediated Ubiquitin Signaling Promotes Removal of Blocking Adducts from DNA Double-Strand Breaks. *IScience*. https://doi.org/10.1016/j.isci.2020.101027

Anand, R., Jasrotia, A., Bundschuh, D., Howard, S. M., Ranjha, L., Stucki, M., & Cejka, P. (2019). NBS1 promotes the endonuclease activity of the MRE11-RAD50 complex by sensing CtIP phosphorylation. *The EMBO Journal*. https://doi.org/10.15252/embj.2018101005

Anand, R., Pinto, C., & Cejka, P. (2018). Methods to Study DNA End Resection I: Recombinant Protein Purification. In *Methods in Enzymology*. https://doi.org/10.1016/bs.mie.2017.11.008

Anand, R., Ranjha, L., Cannavo, E., & Cejka, P. (2016). Phosphorylated CtIP Functions as a Co-factor of the MRE11-RAD50-NBS1 Endonuclease in DNA End Resection. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2016.10.017

Baumann, P., & West, S. C. (1997). The human Rad51 protein: Polarity of strand transfer and stimulation by hRP-A. *EMBO Journal*. https://doi.org/10.1093/emboj/16.17.5198

Bebenek, K., Pedersen, L. C., & Kunkel, T. A. (2014). Structure-function studies of DNA polymerase λ. *Biochemistry*. https://doi.org/10.1021/bi4017236

Bell, J. C., & Kowalczykowski, S. C. (2016). Mechanics and Single-Molecule Interrogation of DNA Recombination. *Annual Review of Biochemistry*. https://doi.org/10.1146/annurev-biochem-060614-034352

Bernstein, N. K., Williams, R. S., Rakovszky, M. L., Cui, D., Green, R., Karimi-Busheri, F., … Glover, J. N. M. (2005). The molecular architecture of the mammalian DNA repair enzyme, polynucleotide kinase. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2005.02.012

Berti, M., Chaudhuri, A. R., Thangavel, S., Gomathinayagam, S., Kenig, S., Vujanovic, M., … Vindigni, A. (2013). Human RECQ1 promotes restart of replication forks reversed by DNA topoisomerase I inhibition. *Nature Structural and Molecular Biology*. https://doi.org/10.1038/nsmb.2501

Berti, M., Cortez, D., & Lopes, M. (2020). The plasticity of DNA replication forks in response to clinically relevant genotoxic stress. *Nature Reviews Molecular Cell Biology*. https://doi.org/10.1038/s41580-020-0257-5

Bétous, R., Couch, F. B., Mason, A. C., Eichman, B. F., Manosas, M., & Cortez, D. (2013). Substrate-Selective Repair and Restart of Replication Forks by DNA Translocases. *Cell Reports*. https://doi.org/10.1016/j.celrep.2013.05.002

Bhargava, R., Onyango, D. O., & Stark, J. M. (2016). Regulation of Single-Strand Annealing and its Role in Genome Maintenance. *Trends in Genetics*. https://doi.org/10.1016/j.tig.2016.06.007

Bhat, K. P., & Cortez, D. (2018). RPA and RAD51: Fork reversal, fork protection, and genome stability. *Nature Structural and Molecular Biology*. https://doi.org/10.1038/s41594-018-0075-z

Bizard, A. H., & Hickson, I. D. (2014). The dissolution of double Holliday junctions. *Cold Spring Harbor Perspectives in Biology*. https://doi.org/10.1101/cshperspect.a016477

Bochman, M. L., Paeschke, K., & Zakian, V. A. (2012). DNA secondary structures: Stability and function of G-quadruplex structures. *Nature Reviews Genetics*. https://doi.org/10.1038/nrg3296

Brázda, V., Laister, R. C., Jagelská, E. B., & Arrowsmith, C. (2011). Cruciform structures are a common DNA feature important for regulating biological processes. *BMC Molecular Biology*. https://doi.org/10.1186/1471-2199-12-33

Broderick, R., Nieminuszczy, J., Baddock, H. T., Deshpande, R. A., Gileadi, O., Paull, T. T., … Niedzwiedz, W. (2016). EXD2 promotes homologous recombination by facilitating DNA end resection. *Nature Cell Biology*. https://doi.org/10.1038/ncb3303

Brosh, R. M., Waheed, J., & Sommers, J. A. (2002). Biochemical characterization of the DNA substrate specificity of Werner syndrome helicase. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M111446200

Burgers, P. M. J., & Kunkel, T. A. (2017). Eukaryotic DNA replication fork. *Annual Review of Biochemistry*. https://doi.org/10.1146/annurev-biochem-061516-044709

Burma, S., Chen, B. P., Murphy, M., Kurimasa, A., & Chen, D. J. (2001). ATM Phosphorylates Histone H2AX in Response to DNA Double-strand Breaks. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.C100466200

Cannan, W. J., & Pederson, D. S. (2016). Mechanisms and Consequences of Double-Strand DNA Break Formation in Chromatin. *Journal of Cellular Physiology*.

https://doi.org/10.1002/jcp.25048

Cannavo, E., & Cejka, P. (2014). Sae2 promotes dsDNA endonuclease activity within Mre11-Rad50-Xrs2 to resect DNA breaks. *Nature*. https://doi.org/10.1038/nature13771

Cannavo, E., Reginato, G., & Cejka, P. (2019). Stepwise 5′ DNA end-specific resection of DNA breaks by the Mre11-Rad50-Xrs2 and Sae2 nuclease ensemble. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1820157116

Casper, A. M., Greenwell, P. W., Tang, W., & Petes, T. D. (2009). Chromosome aberrations resulting from double-strand DNA breaks at a naturally occurring yeast fragile site composed of inverted Ty elements are independent of Mre11p and Sae2p. *Genetics*. https://doi.org/10.1534/genetics.109.106385

Ceccaldi, R., Rondinelli, B., & D'Andrea, A. D. (2016). Repair Pathway Choices and Consequences at the Double-Strand Break. *Trends in Cell Biology*. https://doi.org/10.1016/j.tcb.2015.07.009

Ceccaldi, R., Sarangi, P., & D'Andrea, A. D. (2016). The Fanconi anaemia pathway: New players and new functions. *Nature Reviews Molecular Cell Biology*. https://doi.org/10.1038/nrm.2016.48

Cejka, P. (2015). DNA end resection: Nucleases team up with the right partners to initiate homologous recombination. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.R115.675942

Cejka, P., Cannavo, E., Polaczek, P., Masuda-Sasa, T., Pokharel, S., Campbell, J. L., & Kowalczykowski, S. C. (2010). DNA end resection by Dna2-Sgs1-RPA and its stimulation by Top3-Rmi1 and Mre11-Rad50-Xrs2. *Nature*. https://doi.org/10.1038/nature09355

Cejka, P., & Symington, L. S. (2021). DNA End Resection: Mechanism and Control. *Annual Review of Genetics*. https://doi.org/10.1146/annurev-genet-071719-020312

Chan, Y. W., Fugger, K., & West, S. C. (2018). Unresolved recombination intermediates lead to ultra-fine anaphase bridges, chromosome breaks and aberrations. *Nature Cell Biology*. https://doi.org/10.1038/s41556-017-0011-1

Chang, H. H. Y., & Lieber, M. R. (2016). Structure-Specific nuclease activities of Artemis and the Artemis: DNA-PKcs complex. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkw456

Chang, H. H. Y., Pannunzio, N. R., Adachi, N., & Lieber, M. R. (2017). Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nature Reviews Molecular Cell Biology*. https://doi.org/10.1038/nrm.2017.48

Chang, H. H. Y., Watanabe, G., & Lieber, M. R. (2015). Unifying the DNA End-processing

Roles of the Artemis Nuclease. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.m115.680900

Chapman, J. R., Sossick, A. J., Boulton, S. J., & Jackson, S. P. (2012). BRCA1-associated exclusion of 53BP1 from DNA: Damage sites underlies temporal control of DNA repair. *Journal of Cell Science*. https://doi.org/10.1242/jcs.105353

Chapman, J. R., Taylor, M. R. G., & Boulton, S. J. (2012). Playing the End Game: DNA Double-Strand Break Repair Pathway Choice. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2012.07.029

Chappidi, N., Nascakova, Z., Boleslavska, B., Zellweger, R., Isik, E., Andrs, M., … Janscak, P. (2020). Fork Cleavage-Religation Cycle and Active Transcription Mediate Replication Restart after Fork Stalling at Co-transcriptional R-Loops. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2019.10.026

Chatterjee, N., & Walker, G. C. (2017). Mechanisms of DNA damage, repair, and mutagenesis. *Environmental and Molecular Mutagenesis*. https://doi.org/10.1002/em.22087

Chen, H., Lisby, M., & Symington, L. S. (2013). RPA Coordinates DNA End Resection and Prevents Formation of DNA Hairpins. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2013.04.032

Chennuri, P., Cox, L. S., & Saunders, R. D. C. (2018). EXD2 and WRN exonucleases are required for interstrand crosslink repair in Drosophila. *BioRxiv*. https://doi.org/10.1101/284307

Choi, B., Rempala, G. A., & Kim, J. K. (2017). Beyond the Michaelis-Menten equation: Accurate and efficient estimation of enzyme kinetic parameters. *Scientific Reports*. https://doi.org/10.1038/s41598-017-17072-z

Costantini, S., Woodbine, L., Andreoli, L., Jeggo, P. A., & Vindigni, A. (2007). Interaction of the Ku heterodimer with the DNA ligase IV/Xrcc4 complex and its regulation by DNA-PK. *DNA Repair*. https://doi.org/10.1016/j.dnarep.2006.12.007

Cunningham, L. A., Coté, A. G., Cam-Ozdemir, C., & Lewis, S. M. (2003). Rapid, Stabilizing Palindrome Rearrangements in Somatic Cells by the Center-Break Mechanism. *Molecular and Cellular Biology*. https://doi.org/10.1128/mcb.23.23.8740-8750.2003

Cutts, E., & Vannini, A. (2018). *Troubleshooting biGBac: a practical guide*. (June). https://doi.org/10.17504/protocols.io.q3ydypw

Cuykendall, T. N., & Houston, D. W. (2010). Identification of germ plasm-associated transcripts by microarray analysis of xenopus vegetal cortex RNA. *Developmental Dynamics*. https://doi.org/10.1002/dvdy.22304

Czapinska, H., Kowalska, M., Zagorskaite, E., Manakova, E., Slyvka, A., Xu, S. yong, …

Bochtler, M. (2018). Activity and structure of EcoKMcrA. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gky731

Daley, J. M., Jimenez-Sainz, J., Wang, W., Miller, A. S., Xue, X., Nguyen, K. A., … Sung, P. (2017). Enhancement of BLM-DNA2-Mediated Long-Range DNA End Resection by CtIP. *Cell Reports*. https://doi.org/10.1016/j.celrep.2017.09.048

Deshpande, R. A., Lee, J. H., Arora, S., & Paull, T. T. (2016). Nbs1 Converts the Human Mre11/Rad50 Nuclease Complex into an Endo/Exonuclease Machine Specific for Protein-DNA Adducts. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2016.10.010

Deshpande, R. A., Myler, L. R., Soniat, M. M., Makharashvili, N., Lee, L., Lees-Miller, S. P., … Paull, T. T. (2020). DNA-dependent protein kinase promotes DNA end processing by MRN and CtIP. *Science Advances*. https://doi.org/10.1126/sciadv.aay0922

Di Marco, S., Hasanova, Z., Kanagaraj, R., Chappidi, N., Altmannova, V., Menon, S., … Janscak, P. (2017). RECQ5 Helicase Cooperates with MUS81 Endonuclease in Processing Stalled Replication Forks at Common Fragile Sites during Mitosis. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2017.05.006

Di Virgilio, M., Callen, E., Yamane, A., Zhang, W., Jankovic, M., Gitlin, A. D., … Nussenzweig, M. C. (2013). Rif1 prevents resection of DNA breaks and promotes immunoglobulin class switching. *Science*. https://doi.org/10.1126/science.1230624

Dynan, W. S., & Yoo, S. (1998). Interaction of Ku protein and DNA-dependent protein kinase catalytic subunit with nucleic acids. *Nucleic Acids Research*. https://doi.org/10.1093/nar/26.7.1551

Eggler, A. L., Inman, R. B., & Cox, M. M. (2002). The Rad51-dependent pairing of long DNA substrates is stabilized by replication protein A. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M204328200

Endoh, T., Rode, A. B., Takahashi, S., Kataoka, Y., Kuwahara, M., & Sugimoto, N. (2016). Real-Time Monitoring of G-Quadruplex Formation during Transcription. *Analytical Chemistry*. https://doi.org/10.1021/acs.analchem.5b04396

Esashi, F., Galkin, V. E., Yu, X., Egelman, E. H., & West, S. C. (2007). Stabilization of RAD51 nucleoprotein filaments by the C-terminal region of BRCA2. *Nature Structural and Molecular Biology*. https://doi.org/10.1038/nsmb1245

Farmer, H., McCabe, H., Lord, C. J., Tutt, A. H. J., Johnson, D. A., Richardson, T. B., … Ashworth, A. (2005). Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature*. https://doi.org/10.1038/nature03445

Feng, L., Li, N., Li, Y., Wang, J., Gao, M., Wang, W., & Chen, J. (2015). Cell cycle-dependent inhibition of 53BP1 signaling by BRCA1. *Cell Discovery*. https://doi.org/10.1038/celldisc.2015.19

Garcia, V., Phelps, S. E. L., Gray, S., & Neale, M. J. (2011). Bidirectional resection of DNA double-strand breaks by Mre11 and Exo1. *Nature*. https://doi.org/10.1038/nature10515

Ghosh, A. K., Rossi, M. L., Singh, D. K., Dunn, C., Ramamoorthy, M., Croteau, D. L., … Bohr, V. A. (2012). RECQL4, the protein mutated in Rothmund-Thomson syndrome, functions in telomere maintenance. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M111.295063

Gilson, E., & Géli, V. (2007). How telomeres are replicated. *Nature Reviews Molecular Cell Biology*. https://doi.org/10.1038/nrm2259

Goellner, E. M., Putnam, C. D., & Kolodner, R. D. (2015). Exonuclease 1-dependent and independent mismatch repair. *DNA Repair*. https://doi.org/10.1016/j.dnarep.2015.04.010

Gómez-Herreros, F. (2019). DNA Double Strand Breaks and Chromosomal Translocations Induced by DNA Topoisomerase II. *Frontiers in Molecular Biosciences*. https://doi.org/10.3389/fmolb.2019.00141

Goodarzi, A. A., Yu, Y., Riballo, E., Douglas, P., Walker, S. A., Ye, R., … Lees-Miller, S. P. (2006). DNA-PK autophosphorylation facilitates Artemis endonuclease activity. *EMBO Journal*. https://doi.org/10.1038/sj.emboj.7601255

Gottlieb, T. M., & Jackson, S. P. (1993). The DNA-dependent protein kinase: Requirement for DNA ends and association with Ku antigen. *Cell*. https://doi.org/10.1016/0092-8674(93)90057-W

Gravel, S., Chapman, J. R., Magill, C., & Jackson, S. P. (2008). DNA helicases Sgs1 and BLM promote DNA double-strand break resection. *Genes and Development*. https://doi.org/10.1101/gad.503108

Grawunder, U., Zimmer, D., Fugmann, S., Schwarz, K., & Lieber, M. R. (1998). DNA ligase IV is essential for V(D)J recombination and DNA double-strand break repair in human precursor lymphocytes. *Molecular Cell*. https://doi.org/10.1016/S1097-2765(00)80147-1

Grøntved, L., Bandle, R., John, S., Baek, S., Chung, H. J., Liu, Y., … Levens, D. (2012). Rapid genome-scale mapping of chromatin accessibility in tissue. *Epigenetics and Chromatin*. https://doi.org/10.1186/1756-8935-5-10

Hadi, M. Z., & Wilson, D. M. (2000). Second human protein with homology to the escherichia coli abasic endonuclease exonuclease III. *Environmental and Molecular Mutagenesis*. https://doi.org/10.1002/1098-2280(2000)36:4<312::AID-EM7>3.0.CO;2-K

Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. *Cancer Discovery*. https://doi.org/10.1158/2159-8290.CD-21-1059

Hanahan, D., & Weinberg, R. A. (2000). The Hallmarks of Cancer Review Douglas. *Cell*.

Hänsel-Hertsch, R., Di Antonio, M., & Balasubramanian, S. (2017). DNA G-quadruplexes in the human genome: Detection, functions and therapeutic potential. *Nature Reviews Molecular Cell Biology*. https://doi.org/10.1038/nrm.2017.3

Hartwig, A. (2001). Role of magnesium in genomic stability. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*. https://doi.org/10.1016/S0027-5107(01)00074-4

Helleday, T., Petermann, E., Lundin, C., Hodgson, B., & Sharma, R. A. (2008). DNA repair pathways as targets for cancer therapy. *Nature Reviews Cancer*. https://doi.org/10.1038/nrc2342

Hensen, F., Moretton, A., Van Esveld, S., Farge, G., & Spelbrink, J. N. (2018). The mitochondrial outer-membrane location of the EXD2 exonuclease contradicts its direct role in nuclear DNA repair. *Scientific Reports*. https://doi.org/10.1038/s41598-018-23690-y

Heyer, W. D., Ehmsen, K. T., & Liu, J. (2010). Regulation of homologous recombination in eukaryotes. *Annual Review of Genetics*. https://doi.org/10.1146/annurev-genet-051710-150955

Holloman, W. K. (2011). Unraveling the mechanism of BRCA2 in homologous recombination. *Nature Structural and Molecular Biology*. https://doi.org/10.1038/nsmb.2096

Hu, J., Meyers, R. M., Dong, J., Panchakshari, R. A., Alt, F. W., & Frock, R. L. (2016). Detecting DNA double-stranded breaks in mammalian genomes by linear amplification-mediated high-throughput genome-wide translocation sequencing. *Nature Protocols*. https://doi.org/10.1038/nprot.2016.043

Huang, R. X., & Zhou, P. K. (2020). DNA damage response signaling pathways and targets for radiotherapy sensitization in cancer. *Signal Transduction and Targeted Therapy*. https://doi.org/10.1038/s41392-020-0150-x

Huang, R., & Zhou, P. K. (2021). DNA damage repair: historical perspectives, mechanistic pathways and clinical translation for targeted cancer therapy. *Signal Transduction and Targeted Therapy*. https://doi.org/10.1038/s41392-021-00648-7

Huertas, P., & Jackason, S. P. (2009). Human CtIP mediates cell cycle control of DNA end resection and double strand break repair. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M808906200

Inagaki, H., Ohye, T., Kogo, H., Tsutsumi, M., Kato, T., Tong, M., … Kurahashi, H. (2013). Two sequential cleavage reactions on cruciform DNA structures cause palindrome-mediated chromosomal translocations. *Nature Communications*. https://doi.org/10.1038/ncomms2595

Jackson, S. P., & Bartek, J. (2009). The DNA-damage response in human biology and disease. *Nature*. https://doi.org/10.1038/nature08467

Jasin, M., & Rothstein, R. (2013). Repair of strand breaks by homologous recombination. *Cold Spring Harbor Perspectives in Biology*. https://doi.org/10.1101/cshperspect.a012740

Jia, X., Li, Y., Wang, T., Bi, L., Ye, S., Chen, J., … Zhang, X. (2022). *Discrete RNA – DNA hybrid cleavage by the EXD 2 exonuclease pinpoints two rate-limiting steps*.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*. https://doi.org/10.1038/s41586-021-03819-2

Kadyrova, L. Y., Dahal, B. K., Gujar, V., Daley, J. M., Sung, P., & Kadyrov, F. A. (2022). The nuclease activity of DNA2 promotes exonuclease 1-independent mismatch repair. *Journal of Biological Chemistry*, *298*(4), 101831. https://doi.org/10.1016/j.jbc.2022.101831

Kalia, K., Jiang, W., & Zheng, W. (2008). Manganese accumulates primarily in nuclei of cultured brain cells. *NeuroToxicology*. https://doi.org/10.1016/j.neuro.2008.02.012

Kanoh, Y., Matsumoto, S., Fukatsu, R., Kakusho, N., Kono, N., Renard-Guillet, C., … Masai, H. (2015). Rif1 binds to G quadruplexes and suppresses replication over long distances. *Nature Structural and Molecular Biology*. https://doi.org/10.1038/nsmb.3102

Karimi-Busheri, F., Rasouli-Nia, A., Allalunis-Turner, J., & Weinfeld, M. (2007). Human polynucleotide kinase participates in repair of DNA double-strand breaks by nonhomologous end joining but not homologous recombination. *Cancer Research*. https://doi.org/10.1158/0008-5472.CAN-07-0480

Karow, J. K., Chakraverty, R. K., & Hickson, I. D. (1997). The Bloom’s syndrome gene product is a 3’-5’ DNA helicase. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.272.49.30611

Kass, E. M., Moynahan, M. E., & Jasin, M. (2016). When Genome Maintenance Goes Badly Awry. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2016.05.021

Khan, F. A., & Ali, S. O. (2017). Physiological Roles of DNA Double-Strand Breaks. *Journal of Nucleic Acids*. https://doi.org/10.1155/2017/6439169

Khan, M. M. T., & Martell, A. E. (1962). Metal chelates of adenosine triphosphate. *Journal of Physical Chemistry*. https://doi.org/10.1021/j100807a003

Khanna, K. K., & Jackson, S. P. (2001). DNA double-strand breaks: Signaling, repair and the cancer connection. *Nature Genetics*. https://doi.org/10.1038/85798

Kilgas, S., Singh, A. N., Paillas, S., Then, C. K., Torrecilla, I., Nicholson, J., … Ramadan, K. (2021). p97/VCP inhibition causes excessive MRE11-dependent DNA end resection promoting cell killing after ionizing radiation. *Cell Reports*. https://doi.org/10.1016/j.celrep.2021.109153

Killilea, D. W., & Killilea, A. N. (2022). Mineral requirements for mitochondrial function: A connection to redox balance and cellular differentiation. *Free Radical Biology and Medicine*. https://doi.org/10.1016/j.freeradbiomed.2022.02.022

Kisiala, M., Copelas, A., Czapinska, H., Xu, S. Y., & Bochtler, M. (2018). Crystal structure of the modification-dependent SRA-HNH endonuclease TagI. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gky781

Kissling, V. M., Reginato, G., Bianco, E., Kasaciunaite, K., Tilma, J., Cereghetti, G., … Peter, M. (2022). Mre11-Rad50 oligomerization promotes DNA double-strand break repair. *Nature Communications*, *13*(1), 1–16. https://doi.org/10.1038/s41467-022-29841-0

Kleiner, R. E., Verma, P., Molloy, K. R., Chait, B. T., & Kapoor, T. M. (2015). Chemical proteomics reveals a γH2AX-53BP1 interaction in the DNA damage response. *Nature Chemical Biology*, *11*(10), 807–814. https://doi.org/10.1038/nchembio.1908

Kocincová, L., Jarešová, M., Byška, J., Parulek, J., Hauser, H., & Kozlíková, B. (2017). Comparative visualization of protein secondary structures. *BMC Bioinformatics*. https://doi.org/10.1186/s12859-016-1449-z

Kockler, Z. W., Osia, B., Lee, R., Musmaker, K., & Malkova, A. (2021). Repair of DNA Breaks by Break-Induced Replication. *Annual Review of Biochemistry*. https://doi.org/10.1146/annurev-biochem-081420-095551

Kolinjivadi, A. M., Sannino, V., de Antoni, A., Técher, H., Baldi, G., & Costanzo, V. (2017). Moonlighting at replication forks – a new life for homologous recombination proteins BRCA1, BRCA2 and RAD51. *FEBS Letters*. https://doi.org/10.1002/1873-3468.12556

Kolinjivadi, A. M., Sannino, V., De Antoni, A., Zadorozhny, K., Kilkenny, M., Técher, H., … Costanzo, V. (2017). Smarcal1-Mediated Fork Reversal Triggers Mre11-Dependent Degradation of Nascent DNA in the Absence of Brca2 and Stable Rad51 Nucleofilaments. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2017.07.001

Krokan, H. E., & Bjørås, M. (2013). Base excision repair. *Cold Spring Harbor Perspectives in Biology*. https://doi.org/10.1101/cshperspect.a012583

Krum, S. A., Dalugdugan, E. D. L. R., Miranda-Carboni, G. A., & Lane, T. F. (2010). BRCA1 forms a functional complex with γ-H2AX as a late response to genotoxic stress. *Journal of Nucleic Acids*. https://doi.org/10.4061/2010/801594

Kunkel, T. A. (2009). Evolving views of DNA replication (in)fidelity. *Cold Spring Harbor Symposia on Quantitative Biology*. https://doi.org/10.1101/sqb.2009.74.027

Kuo, C. F., McRee, D. E., Cunningham, R. P., & Tainer, J. A. (1993). Purification, crystallization and space group determination of DNA repair enzyme exonuclease III from E. coli. *Journal of Molecular Biology*. https://doi.org/10.1006/jmbi.1993.1021

Kusakabe, M., Onishi, Y., Tada, H., Kurihara, F., Kusao, K., Furukawa, M., … Sugasawa, K. (2019). Mechanism and regulation of DNA damage recognition in nucleotide excision repair. *Genes and Environment*. https://doi.org/10.1186/s41021-019-0119-6

Kusumoto, R., Dawut, L., Marchetti, C., Jae, W. L., Vindigni, A., Ramsden, D., & Bohr, V. A. (2008). Werner protein cooperates with the XRCC4-DNA ligase IV complex in end-processing. *Biochemistry*. https://doi.org/10.1021/bi702325t

Lam, I., & Keeney, S. (2015). Mechanism and regulation of meiotic recombination initiation. *Cold Spring Harbor Perspectives in Biology*. https://doi.org/10.1101/cshperspect.a016634

Lans, H., Hoeijmakers, J. H. J., Vermeulen, W., & Marteijn, J. A. (2019). The DNA damage response to transcription stress. *Nature Reviews Molecular Cell Biology*. https://doi.org/10.1038/s41580-019-0169-4

Lee, B. I., & Wilson, D. M. (1999). The RAD2 domain of human exonuclease 1 exhibits 5’ to 3’ exonuclease and flap structure-specific endonuclease activities. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.274.53.37763

Lee, G. E., Kim, J. H., & Chung, I. K. (1998). Topoisomerase II-mediated DNA cleavage on the cruciform structure formed within the 5′ Upstream region of the human β-globin gene. *Molecules and Cells*.

Lee, J. H., & Paull, T. T. (2006). Purification and Biochemical Characterization of Ataxia-Telangiectasia Mutated and Mre11/Rad50/Nbs1. *Methods in Enzymology*. https://doi.org/10.1016/S0076-6879(06)08033-5

Lee, Joon, & Dunphy, W. G. (2013). The Mre11-Rad50-Nbs1 (MRN) complex has a specific role in the activation of Chk1 in response to stalled replication forks. *Molecular Biology of the Cell*. https://doi.org/10.1091/mbc.E13-01-0025

Lee, Junyeop, Sung, K., Joo, S. Y., Jeong, J. H., Kim, S. K., & Lee, H. (2022). Dynamic interaction of BRCA2 with telomeric G-quadruplexes underlies telomere replication homeostasis. *Nature Communications*, *13*(1). https://doi.org/10.1038/s41467-022-31156-z

Lemaçon, D., Jackson, J., Quinet, A., Brickner, J. R., Li, S., Yazinski, S., … Vindigni, A. (2017). MRE11 and EXO1 nucleases degrade reversed forks and elicit MUS81-dependent fork rescue in BRCA2-deficient cells. *Nature Communications*. https://doi.org/10.1038/s41467-017-01180-5

Lengsfeld, B. M., Rattray, A. J., Bhaskara, V., Ghirlando, R., & Paull, T. T. (2007). Sae2 Is an Endonuclease that Processes Hairpin DNA Cooperatively with the

Mre11/Rad50/Xrs2 Complex. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2007.11.001

Li, B., Reddy, S., & Comai, L. (2009). Sequence-specific processing of telomeric 3' overhangs by the Werner syndrome protein exonuclease activity. *Aging*. https://doi.org/10.18632/aging.100032

Li, S., Chang, H. H., Niewolik, D., Hedrick, M. P., Pinkerton, A. B., Hassig, C. A., … Lieber, M. R. (2014). Evidence that the DNA endonuclease ARTEMIS also has intrinsic 5′-exonuclease activity. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M113.544874

Li, Y., Shen, J., & Niu, H. (2019). DNA duplex recognition activates Exo1 nuclease activity. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.RA119.008549

Liano, D., Monti, L., Chowdhury, S., Raguseo, F., & Di Antonio, M. (2022). Long-range DNA interactions: inter-molecular G-quadruplexes and their potential biological relevance. *Chemical Communications*, 12753–12762. https://doi.org/10.1039/d2cc04872h

Liao, S., Guay, C., Toczylowski, T., & Yan, H. (2012). Analysis of MRE11's function in the 5′→3′ processing of DNA double-strand breaks. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gks044

Lieber, M. R. (2010). The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annual Review of Biochemistry*. https://doi.org/10.1146/annurev.biochem.052308.093131

Lin, W., Sampathi, S., Dai, H., Liu, C., Zhou, M., Hu, J., … Shen, B. (2013). Mammalian DNA2 helicase/nuclease cleaves G-quadruplex DNA and is required for telomere integrity. *EMBO Journal*. https://doi.org/10.1038/emboj.2013.88

Lisby, M., & Rothstein, R. (2015). Cell biology of mitotic recombination. *Cold Spring Harbor Perspectives in Biology*. https://doi.org/10.1101/cshperspect.a016535

Liu, B., Hu, J., Wang, J., & Kong, D. (2017). Direct visualization of RNA-DNA primer removal from okazaki fragments provides support for flap cleavage and exonucleolytic pathways in eukaryotic cells. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M116.758599

Liu, L., & Malkova, A. (2022). Break-induced replication: unraveling each step. *Trends in Genetics*, *38*(7), 752–765. https://doi.org/10.1016/j.tig.2022.03.011

Liu, T., & Huang, J. (2016). Replication protein A and more: Single-stranded DNA-binding proteins in eukaryotic cells. *Acta Biochimica et Biophysica Sinica*. https://doi.org/10.1093/abbs/gmw041

Liu, Y., Sung, S., Kim, Y., Li, F., Gwon, G., Jo, A., … Cho, Y. (2016). ATP -dependent DNA binding, unwinding, and resection by the Mre11/Rad50 complex . *The EMBO*

Journal. https://doi.org/10.15252/embj.201592462

Lord, C. J., & Ashworth, A. (2017). PARP inhibitors: Synthetic lethality in the clinic. *Science*. https://doi.org/10.1126/science.aam7344

Lu, R., & Pickett, H. A. (2022). Telomeric replication stress: The beginning and the end for alternative lengthening of telomeres cancers. *Open Biology*. https://doi.org/10.1098/rsob.220011

Ma, Y., Lu, H., Tippin, B., Goodman, M. F., Shimazaki, N., Koiwai, O., … Lieber, M. R. (2004). A biochemically defined system for mammalian nonhomologous DNA end joining. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2004.11.017

Machwe, A., Xiao, L., Theodore, S., & Orren, D. K. (2002). DNase I footprinting and enhanced exonuclease function of the Bipartite Werner Syndrome protein (WRN) bound to partially melted duplex DNA. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M108880200

Makharashvili, N., Tubbs, A. T., Yang, S. H., Wang, H., Barton, O., Zhou, Y., … Paull, T. T. (2014). Catalytic and Noncatalytic Roles of the CtIP Endonuclease in Double-Strand Break End Resection. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2014.04.011

Mateos-Gomez, P. A., Kent, T., Deng, S. K., Mcdevitt, S., Kashkina, E., Hoang, T. M., … Sfeir, A. (2017). The helicase domain of Polθ counteracts RPA to promote alt-NHEJ. *Nature Structural and Molecular Biology*. https://doi.org/10.1038/nsmb.3494

McCabe, N., Turner, N. C., Lord, C. J., Kluzek, K., Białkowska, A., Swift, S., … Ashworth, A. (2006). Deficiency in the repair of DNA damage by homologous recombination and sensitivity to poly(ADP-ribose) polymerase inhibition. *Cancer Research*. https://doi.org/10.1158/0008-5472.CAN-06-0140

McKinnon, P. J., & Caldecott, K. W. (2007). DNA strand break repair and human genetic disease. *Annual Review of Genomics and Human Genetics*. https://doi.org/10.1146/annurev.genom.7.080505.115648

Mehta, A., & Haber, J. E. (2014). Sources of DNA double-strand breaks and models of recombinational DNA repair. *Cold Spring Harbor Perspectives in Biology*. https://doi.org/10.1101/cshperspect.a016428

Mijic, S., Zellweger, R., Chappidi, N., Berti, M., Jacobs, K., Mutreja, K., … Lopes, M. (2017). Replication fork reversal triggers fork degradation in BRCA2-defective cells. *Nature Communications*. https://doi.org/10.1038/s41467-017-01164-5

Mikkola, S., Lönnberg, T., & Lönnberg, H. (2018). Phosphodiester models for cleavage of nucleic acids. *Beilstein Journal of Organic Chemistry*. https://doi.org/10.3762/bjoc.14.68

Miklenić, M. S., & Svetec, I. K. (2021). Palindromes in dna—a risk for genome stability and implications in cancer. *International Journal of Molecular Sciences*. https://doi.org/10.3390/ijms22062840

Mimitou, E. P., & Symington, L. S. (2011). DNA end resection-Unraveling the tail. *DNA Repair*. https://doi.org/10.1016/j.dnarep.2010.12.004

Min, J., Wright, W. E., & Shay, J. W. (2019). Clustered telomeres in phase-separated nuclear condensates engage mitotic DNA synthesis through BLM and RAD52. *Genes and Development*. https://doi.org/10.1101/gad.324905.119

Mirkin, S. M. (2006). DNA structures, repeat expansions and human hereditary disorders. *Current Opinion in Structural Biology*. https://doi.org/10.1016/j.sbi.2006.05.004

Mirman, Z., Lottersberger, F., Takai, H., Kibe, T., Gong, Y., Takai, K., … de Lange, T. (2018). 53BP1–RIF1–shieldin counteracts DSB resection through CST- and Polα-dependent fill-in. *Nature*. https://doi.org/10.1038/s41586-018-0324-7

Miura, O., Ogake, T., Yoneyama, H., Kikuchi, Y., & Ohyama, T. (2019). A strong structural correlation between short inverted repeat sequences and the polyadenylation signal in yeast and nucleosome exclusion by these inverted repeats. *Current Genetics*. https://doi.org/10.1007/s00294-018-0907-8

Monia, B. P., Johnston, J. F., Sasmor, H., & Cummins, L. L. (1996). Nuclease resistance and antisense activity of modified oligonucleotides targeted to Ha-ras. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.271.24.14533

Moye, A. L., Porter, K. C., Cohen, S. B., Phan, T., Zyner, K. G., Sasaki, N., … Bryan, T. M. (2015). Telomeric G-quadruplexes are a substrate and site of localization for human telomerase. *Nature Communications*. https://doi.org/10.1038/ncomms8643

Moynahan, M. E., & Jasin, M. (2010). Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. *Nature Reviews Molecular Cell Biology*. https://doi.org/10.1038/nrm2851

Myler, L. R., Gallardo, I. F., Soniat, M. M., Deshpande, R. A., Gonzalez, X. B., Kim, Y., … Finkelstein, I. J. (2017). Single-Molecule Imaging Reveals How Mre11-Rad50-Nbs1 Initiates DNA Break Repair. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2017.08.002

Myler, L. R., Soniat, M. M., Zhang, X., Deshpande, R. A., Paull, T. T., & Finkelstein, I. J. (2019). Purification and biophysical characterization of the mre11-Rad50-Nbs1 complex. In *Methods in Molecular Biology*. https://doi.org/10.1007/978-1-4939-9520-2_20

Naito, A., Hiwasa, T., Tanabe, N., Sanada, T. J., Sugiura, T., Shigeta, A., … Tatsumi, K. (2019). Elevated levels of autoantibodies against EXD2 and PHAX in the sera of patients with chronic thromboembolic pulmonary hypertension. *PLoS ONE*.

https://doi.org/10.1371/journal.pone.0211377

Nazaryan-Petersen, L., Bjerregaard, V. A., Nielsen, F. C., Tommerup, N., & Tümer, Z. (2020). Chromothripsis and DNA repair disorders. *Journal of Clinical Medicine*. https://doi.org/10.3390/jcm9030613

Neelsen, K. J., & Lopes, M. (2015). Replication fork reversal in eukaryotes: From dead end to dynamic response. *Nature Reviews Molecular Cell Biology*. https://doi.org/10.1038/nrm3935

Nguyen, B., Sokoloski, J., Galletto, R., Elson, E. L., Wold, M. S., & Lohman, T. M. (2014). Diffusion of human replication protein a along single-stranded DNA. *Journal of Molecular Biology*. https://doi.org/10.1016/j.jmb.2014.07.014

Nieminuszczy, J., Broderick, R., Bellani, M. A., Smethurst, E., Schwab, R. A., Cherdyntseva, V., … Niedzwiedz, W. (2019). EXD2 Protects Stressed Replication Forks and Is Required for Cell Viability in the Absence of BRCA1/2. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2019.05.026

Nieminuszczy, J., Broderick, R., & Niedzwiedz, W. (2016). EXD2 - a new player joins the DSB resection team. *Cell Cycle*. https://doi.org/10.1080/15384101.2016.1161997

Nimonkar, A. V., Genschel, J., Kinoshita, E., Polaczek, P., Campbell, J. L., Wyman, C., … Kowalczykowski, S. C. (2011). BLM-DNA2-RPA-MRN and EXO1-BLM-RPA-MRN constitute two DNA end resection machineries for human DNA break repair. *Genes and Development*. https://doi.org/10.1101/gad.2003811

Nimonkar, A. V., Özsoy, A. Z., Genschel, J., Modrich, P., & Kowalczykowski, S. C. (2008). Human exonuclease 1 and BLM helicase interact to resect DNA and initiate DNA repair. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.0809380105

Nurse, P. (2002). Cyclin dependent kinases and cell cycle control (Nobel lecture). *ChemBioChem*. https://doi.org/10.1002/1439-7633(20020703)3:7<596::AID-CBIC596>3.0.CO;2-U

Ochi, T., Blackford, A. N., Coates, J., Jhujh, S., Mehmood, S., Tamura, N., … Jackson, S. P. (2015). PAXX, a paralog of XRCC4 and XLF, interacts with Ku to promote DNA double-strand break repair. *Science*. https://doi.org/10.1126/science.1261971

Oehl-Jaschkowitz, B., Vanakker, O. M., De Paepe, A., Menten, B., Martin, T., Weber, G., … Tzschach, A. (2014). Deletions in 14q24.1q24.3 are associated with congenital heart defects, brachydactyly, and mild intellectual disability. *American Journal of Medical Genetics, Part A*. https://doi.org/10.1002/ajmg.a.36321

Ohyama, T. (2019). New aspects of magnesium function: A key regulator in nucleosome self-assembly, chromatin folding and phase separation. *International Journal of Molecular Sciences*. https://doi.org/10.3390/ijms20174232

Orlova, E. V., & Saibil, H. R. (2011). Structural analysis of macromolecular assemblies by electron microscopy. *Chemical Reviews*. https://doi.org/10.1021/cr100353t

Pannunzio, N. R., Watanabe, G., & Lieber, M. R. (2018). Nonhomologous DNA end-joining for repair of DNA double-strand breaks. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.TM117.000374

Park, J., Lee, S.-Y., Jeong, H., Kang, M.-G., Van Haute, L., Minczuk, M., … Lee, C. (2019). The structure of human EXD2 reveals a chimeric 3′ to 5′ exonuclease domain that discriminates substrates via metal coordination. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkz454

Paull, T. T. (2018). 20 Years of Mre11 Biology: No End in Sight. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2018.06.033

Paull, T. T., & Gellert, M. (1998). The 3′ to 5′ exonuclease activity of Mre11 facilitates repair of DNA double-strand breaks. *Molecular Cell*. https://doi.org/10.1016/S1097-2765(00)80097-0

Paull, T. T., & Gellert, M. (1999). Nbs1 potentiates ATP-driven DNA unwinding and endonuclease cleavage by the Mre11/Rad50 complex. *Genes and Development*. https://doi.org/10.1101/gad.13.10.1276

Pećina-Šlaus, N., Kafka, A., Salamon, I., & Bukovac, A. (2020). Mismatch Repair Pathway, Genome Stability and Cancer. *Frontiers in Molecular Biosciences*. https://doi.org/10.3389/fmolb.2020.00122

Pepe, A., & West, S. C. (2014). MUS81-EME2 promotes replication fork restart. *Cell Reports*. https://doi.org/10.1016/j.celrep.2014.04.007

Perry, J. J. P., Yannone, S. M., Holden, L. G., Hitomi, C., Asaithamby, A., Han, S., … Tainer, J. A. (2006). WRN exonuclease structure and molecular mechanism imply an editing role in DNA end processing. *Nature Structural and Molecular Biology*. https://doi.org/10.1038/nsmb1088

Petraccone, L., Spink, C., Trent, J. O., Garbett, N. C., Mekmaysy, C. S., Giancola, C., & Chaires, J. B. (2011). Structure and stability of higher-order human telomeric quadruplexes. *Journal of the American Chemical Society*. https://doi.org/10.1021/ja209192a

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*. https://doi.org/10.1002/jcc.20084

Pickett, H. A., Henson, J. D., Au, A. Y. M., Neumann, A. A., & Reddel, R. R. (2011). Normal mammalian cells negatively regulate telomere length by telomere trimming. *Human Molecular Genetics*. https://doi.org/10.1093/hmg/ddr402

Pinto, C., Anand, R., & Cejka, P. (2018). Methods to Study DNA End Resection II: Biochemical Reconstitution Assays. In *Methods in Enzymology*. https://doi.org/10.1016/bs.mie.2017.11.009

Pommier, Y., Nussenzweig, A., Takeda, S., & Austin, C. (2022). Human topoisomerases and their roles in genome stability and organization. *Nature Reviews Molecular Cell Biology*. https://doi.org/10.1038/s41580-022-00452-3

Pommier, Y., Sun, Y., Huang, S. Y. N., & Nitiss, J. L. (2016). Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. *Nature Reviews Molecular Cell Biology*. https://doi.org/10.1038/nrm.2016.111

Povirk, L. F. (2012). Processing of Damaged DNA Ends for Double-Strand Break Repair in Mammalian Cells. *ISRN Molecular Biology*. https://doi.org/10.5402/2012/345805

Powell, S. N., & Kachnic, L. A. (2003). Roles of BRCA1 and BRCA2 in homologous recombination, DNA replication fidelity and the cellular response to ionizing radiation. *Oncogene*. https://doi.org/10.1038/sj.onc.1206678

Prakash, R., Zhang, Y., Feng, W., & Jasin, M. (2015). Homologous recombination and human health: The roles of BRCA1, BRCA2, and associated proteins. *Cold Spring Harbor Perspectives in Biology*. https://doi.org/10.1101/cshperspect.a016600

Punatar, R. S., Martin, M. J., Wyatt, H. D. M., Chan, Y. W., & West, S. C. (2017). Resolution of single and double Holliday junction recombination intermediates by GEN 1. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1619790114

Qin, W., Myers, S. A., Carey, D. K., Carr, S. A., & Ting, A. Y. (2021). Spatiotemporally-resolved mapping of RNA binding proteins via functional proximity labeling reveals a mitochondrial mRNA anchor promoting stress recovery. *Nature Communications*. https://doi.org/10.1038/s41467-021-25259-2

Raynard, S., Niu, H., & Sung, P. (2008). DNA double-strand break processing: The beginning of the end. *Genes and Development*. https://doi.org/10.1101/gad.1742408

Reginato, G., Cannavo, E., & Cejka, P. (2017). Physiological protein blocks direct the Mre11-Rad50-Xrs2 and Sae2 nuclease complex to initiate DNA end resection. *Genes and Development*. https://doi.org/10.1101/gad.308254.117

Reginato, G., & Cejka, P. (2020). The MRE11 complex: A versatile toolkit for the repair of broken DNA. *DNA Repair*. https://doi.org/10.1016/j.dnarep.2020.102869

Rhodes, D., & Lipps, H. J. (2015). Survey and summary G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkv862

Robinson, J., Raguseo, F., Nuccio, S. P., Liano, D., & Di Antonio, M. (2021). DNA G-quadruplex structures: More than simple roadblocks to transcription? *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkab609

Romero-Laorden, N., & Castro, E. (2017). Inherited mutations in DNA repair genes and cancer risk. *Current Problems in Cancer*. https://doi.org/10.1016/j.currproblcancer.2017.02.009

Rossetti, G., Dans, P. D., Gomez-Pinto, I., Ivani, I., Gonzalez, C., & Orozco, M. (2015). The structural impact of DNA mismatches. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkv254

Rowland, R. J., Wu, L., Liu, F., & Davies, G. J. (2020). A baculoviral system for the production of human β-glucocerebrosidase enables atomic resolution analysis. *Acta Crystallographica Section D: Structural Biology*. https://doi.org/10.1107/S205979832000501X

Roy, R., Chun, J., & Powell, S. N. (2012). BRCA1 and BRCA2: Different roles in a common pathway of genome protection. *Nature Reviews Cancer*. https://doi.org/10.1038/nrc3181

Saldivar, J. C., Cortez, D., & Cimprich, K. A. (2017). The essential kinase ATR: Ensuring faithful duplication of a challenging genome. *Nature Reviews Molecular Cell Biology*. https://doi.org/10.1038/nrm.2017.67

Sallmyr, A., & Tomkinson, A. E. (2018). Repair of DNA double-strand breaks by mammalian alternative end-joining pathways. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.TM117.000375

Sandoz, J., Cigrang, M., Catez, P., Donnio, L., Elly, C., Berico, P., … Coin, F. (2022). *Active mRNA degradation by EXD2 nuclease elicits recovery of transcription after genotoxic stress*. https://doi.org/10.1101/2022.07.20.499545

Schellenberg, M. J., & Williams, R. S. (2011). DNA end processing by polynucleotide kinase/phosphatase. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1118214109

Scheres, S. H. W. (2012). RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology*. https://doi.org/10.1016/j.jsb.2012.09.006

Schipler, A., & Iliakis, G. (2013). DNA double-strand-break complexity levels and their possible contributions to the probability for error-prone processing and repair pathway choice. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkt556

Scully, R., Panday, A., Elango, R., & Willis, N. A. (2019). DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nature Reviews Molecular Cell Biology*. https://doi.org/10.1038/s41580-019-0152-0

Sebesta, M., Burkovics, P., Juhasz, S., Zhang, S., Szabo, J. E., Lee, M. Y. W. T., …

Krejci, L. (2013). Role of PCNA and TLS polymerases in D-loop extension during homologous recombination in humans. *DNA Repair*. https://doi.org/10.1016/j.dnarep.2013.05.001

Sebesta, M., Cooper, C. D. O., Ariza, A., Carnie, C. J., & Ahel, D. (2017). Structural insights into the function of ZRANB3 in replication stress response. *Nature Communications*. https://doi.org/10.1038/ncomms15847

Seol, J. H., Shim, E. Y., & Lee, S. E. (2018). Microhomology-mediated end joining: Good, bad and ugly. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*. https://doi.org/10.1016/j.mrfmmm.2017.07.002

Sfeir, A., & Symington, L. S. (2015). Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway? *Trends in Biochemical Sciences*. https://doi.org/10.1016/j.tibs.2015.08.006

Shamanna, R. A., Lu, H., De Freitas, J. K., Tian, J., Croteau, D. L., & Bohr, V. A. (2016). WRN regulates pathway choice between classical and alternative non-homologous end joining. *Nature Communications*. https://doi.org/10.1038/ncomms13785

Shay, J. W., & Wright, W. E. (2019). Telomeres and telomerase: three decades of progress. *Nature Reviews Genetics*. https://doi.org/10.1038/s41576-019-0099-1

Shevelev, I. V., & Hübscher, U. (2002). The 3′-5′ exonucleases. *Nature Reviews Molecular Cell Biology*. https://doi.org/10.1038/nrm804

Shibata, A., Moiani, D., Arvai, A. S., Perry, J., Harding, S. M., Genois, M. M., … Tainer, J. A. (2014). DNA Double-Strand Break Repair Pathway Choice Is Directed by Distinct MRE11 Nuclease Activities. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2013.11.003

Shibata, A., Steinlage, M., Barton, O., Juhász, S., Künzel, J., Spies, J., … Löbrich, M. (2017). DNA Double-Strand Break Resection Occurs during Non-homologous End Joining in G1 but Is Distinct from Resection during Homologous Recombination. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2016.12.016

Shorrocks, A. M. K., Jones, S. E., Tsukada, K., Morrow, C. A., Belblidia, Z., Shen, J., … Blackford, A. N. (2021). The Bloom syndrome complex senses RPA-coated single-stranded DNA to restart stalled replication forks. *Nature Communications*. https://doi.org/10.1038/s41467-020-20818-5

Sigurdsson, S., Van Komen, S., Petukhova, G., & Sung, P. (2002). Homologous DNA pairing by human recombination factors Rad51 and Rad54. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M208004200

Silva, J., Aivio, S., Knobel, P. A., Bailey, L. J., Casali, A., Vinaixa, M., … Stracker, T. H. (2018). EXD2 governs germ stem cell homeostasis and lifespan by promoting mitoribosome integrity and translation. *Nature Cell Biology*.

https://doi.org/10.1038/s41556-017-0016-9

Smith, G. C. M., & Jackson, S. P. (1999). The DNA-dependent protein kinase. *Genes and Development*. https://doi.org/10.1101/gad.13.8.916

Smogorzewska, A., Desetty, R., Saito, T. T., Schlabach, M., Lach, F. P., Sowa, M. E., … Elledge, S. J. (2010). A Genetic Screen Identifies FAN1, a Fanconi Anemia-Associated Nuclease Necessary for DNA Interstrand Crosslink Repair. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2010.06.023

Song, Q., Hu, Y., Yin, A., Wang, H., & Yin, Q. (2022). DNA Holliday Junction: History, Regulation and Bioactivity. *International Journal of Molecular Sciences*, *23*(17). https://doi.org/10.3390/ijms23179730

Soniat, M. M., Myler, L. R., Kuo, H. C., Paull, T. T., & Finkelstein, I. J. (2019). RPA Phosphorylation Inhibits DNA Resection. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2019.05.005

Sotiriou, S. K., Kamileri, I., Lugli, N., Evangelou, K., Da-Ré, C., Huber, F., … Halazonetis, T. D. (2016). Mammalian RAD52 Functions in Break-Induced Replication Repair of Collapsed DNA Replication Forks. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2016.10.038

Soulas-Sprauel, P., Rivera-Munoz, P., Malivert, L., Le Guyader, G., Abramowski, V., Revy, P., & De Villartay, J. P. (2007). V(D)J and immunoglobulin class switch recombinations: A paradigm to study the regulation of DNA end-joining. *Oncogene*. https://doi.org/10.1038/sj.onc.1210875

Spírek, M., Mlčoušková, J., Beláň, O., Gyimesi, M., Harami, G. M., Molnár, E., … Krejci, L. (2018). Human RAD51 rapidly forms intrinsically dynamic nucleoprotein filaments modulated by nucleotide binding state. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gky111

Stracker, T. H. (2018). EXD2: A new regulator of mitochondrial translation and potential target for cancer therapy. *Molecular and Cellular Oncology*. https://doi.org/10.1080/23723556.2018.1445943

Stracker, T. H., & Petrini, J. H. J. (2011). The MRE11 complex: Starting from the ends. *Nature Reviews Molecular Cell Biology*. https://doi.org/10.1038/nrm3047

Stucki, M., Clapperton, J. A., Mohammad, D., Yaffe, M. B., Smerdon, S. J., & Jackson, S. P. (2005). MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. *Cell*. https://doi.org/10.1016/j.cell.2005.09.038

Sturzenegger, A., Burdova, K., Kanagaraj, R., Levikova, M., Pinto, C., Cejka, P., & Janscak, P. (2014). DNA2 cooperates with the WRN and BLM RecQ helicases to mediate long-range DNA end resection in human cells. *Journal of Biological*

Chemistry. https://doi.org/10.1074/jbc.M114.578823

Sun, Y., Kong, L., Wu, G., Cao, B., Pang, X., Deng, Z., … You, D. (2020). DNA phosphorothioate modifications are widely distributed in the human microbiome. *Biomolecules*. https://doi.org/10.3390/biom10081175

Sung, P., Krejci, L., Van Komen, S., & Sehorn, M. G. (2003). Rad51 Recombinase and Recombination Mediators. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.R300027200

Syed, A., & Tainer, J. A. (2018). The MRE11–RAD50–NBS1 Complex Conducts the Orchestration of Damage Signaling and Outcomes to Stress in DNA Replication and Repair. *Annual Review of Biochemistry*. https://doi.org/10.1146/annurev-biochem-062917-012415

Thakur, S., Sarkar, B., Cholia, R. P., Gautam, N., Dhiman, M., & Mantha, A. K. (2014). APE1/Ref-1 as an emerging therapeutic target for various human diseases: Phytochemical modulation of its functions. *Experimental and Molecular Medicine*. https://doi.org/10.1038/emm.2014.42

Thangavel, S., Berti, M., Levikova, M., Pinto, C., Gomathinayagam, S., Vujanovic, M., … Vindigni, A. (2015). DNA2 drives processing and restart of reversed replication forks in human cells. *Journal of Cell Biology*. https://doi.org/10.1083/jcb.201406100

Thompson, P. S., & Cortez, D. (2020). New insights into abasic site repair and tolerance. *DNA Repair*. https://doi.org/10.1016/j.dnarep.2020.102866

Thompson, R. F., Walker, M., Siebert, C. A., Muench, S. P., & Ranson, N. A. (2016). An introduction to sample preparation and imaging by cryo-electron microscopy for structural biology. *Methods*. https://doi.org/10.1016/j.ymeth.2016.02.017

Toller, I. M., Neelsen, K. J., Steger, M., Hartung, M. L., Hottiger, M. O., Stucki, M., … Müller, A. (2011). Carcinogenic bacterial pathogen Helicobacter pylori triggers DNA double-strand breaks and a DNA damage response in its host cells. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1100959108

Trujillo, K. M., & Sung, P. (2001). DNA Structure-specific Nuclease Activities in the Saccharomyces cerevisiae Rad50-Mre11 Complex. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M105482200

Tubbs, A., & Nussenzweig, A. (2017). Endogenous DNA Damage as a Source of Genomic Instability in Cancer. *Cell*. https://doi.org/10.1016/j.cell.2017.01.002

Ui, A., Chiba, N., & Yasui, A. (2020). Relationship among DNA double-strand break (DSB), DSB repair, and transcription prevents genome instability and cancer. *Cancer Science*. https://doi.org/10.1111/cas.14404

Uziel, T., Lerenthal, Y., Moyal, L., Andegeko, Y., Mittelman, L., & Shiloh, Y. (2003).

Requirement of the MRN complex for ATM activation by DNA damage. *EMBO Journal*. https://doi.org/10.1093/emboj/cdg541

Valton, A. L., & Prioleau, M. N. (2016). G-Quadruplexes in DNA Replication: A Problem or a Necessity? *Trends in Genetics*. https://doi.org/10.1016/j.tig.2016.09.004

van den Boom, J., Wolf, M., Weimann, L., Schulze, N., Li, F., Kaschani, F., … Meyer, H. (2016). VCP/p97 Extracts Sterically Trapped Ku70/80 Rings from DNA in Double-Strand Break Repair. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2016.08.037

Van der linden, E., Sanchez, H., Kinoshita, E., Kanaar, R., & Wyman, C. (2009). RAD50 and NBS1 form a stable complex functional in DNA binding and tethering. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkn1072

Viguera, E., Canceill, D., & Ehrlich, S. D. (2001). Replication slippage involves DNA polymerase pausing and dissociation. *EMBO Journal*. https://doi.org/10.1093/emboj/20.10.2587

Vilenchik, M. M., & Knudson, A. G. (2003). Endogenous DNA double-strand breaks: Production, fidelity of repair, and induction of cancer. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.2135498100

Vítor, A. C., Huertas, P., Legube, G., & de Almeida, S. F. (2020). Studying DNA Double-Strand Break Repair: An Ever-Growing Toolbox. *Frontiers in Molecular Biosciences*. https://doi.org/10.3389/fmolb.2020.00024

Von Kobbe, C., Thomäs, N. H., Czyzewski, B. K., Pavletich, N. P., & Bohr, V. A. (2003). Werner Syndrome Protein Contains Three Structure-specific DNA Binding Domains. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M308338200

Von Nicolai, C., Ehlén, Å., Martin, C., Zhang, X., & Carreira, A. (2016). A second DNA binding site in human BRCA2 promotes homologous recombination. *Nature Communications*. https://doi.org/10.1038/ncomms12813

Walsh, R. (2012). Alternative Perspectives of Enzyme Kinetic Modeling. In *Medicinal Chemistry and Drug Design*. https://doi.org/10.5772/36973

Wan, L., Lou, J., Xia, Y., Su, B., Liu, T., Cui, J., … Huang, J. (2013). HPrimpol1/CCDC111 is a human DNA primase-polymerase required for the maintenance of genome integrity. *EMBO Reports*. https://doi.org/10.1038/embor.2013.159

Wang, G., & Vasquez, K. M. (2014). Impact of alternative DNA structures on DNA damage, DNA repair, and genetic instability. *DNA Repair*. https://doi.org/10.1016/j.dnarep.2014.03.017

Wang, H., Li, Y., Truong, L. N., Shi, L. Z., Hwang, P. Y. H., He, J., … Wu, X. (2014). CtIP maintains stability at common fragile sites and inverted repeats by end resection-independent endonuclease activity. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2014.04.012

Wang, J. C. (2002). Cellular roles of DNA topoisomerases: A molecular perspective. *Nature Reviews Molecular Cell Biology*. https://doi.org/10.1038/nrm831

Wang, W., Hu, S., Gu, Y., Yan, Y., Stovall, D. B., Li, D., & Sui, G. (2020). Human MYC G-quadruplex: From discovery to a cancer therapeutic target. *Biochimica et Biophysica Acta - Reviews on Cancer*. https://doi.org/10.1016/j.bbcan.2020.188410

Wechsler, T., Newman, S., & West, S. C. (2011). Aberrant chromosome morphology in human cells defective for Holliday junction resolution. *Nature*. https://doi.org/10.1038/nature09790

Wimberly, H., Shee, C., Thornton, P. C., Sivaramakrishnan, P., Rosenberg, S. M., & Hastings, P. J. (2013). R-loops and nicks initiate DNA breakage and genome instability in non-growing Escherichia coli. *Nature Communications*. https://doi.org/10.1038/ncomms3115

Wyatt, H. D. M., Laister, R. C., Martin, S. R., Arrowsmith, C. H., & West, S. C. (2017). The SMX DNA Repair Tri-nuclease. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2017.01.031

Wyatt, H. D. M., Sarbajna, S., Matos, J., & West, S. C. (2013). Coordinated actions of SLX1-SLX4 and MUS81-EME1 for holliday junction resolution in human cells. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2013.08.035

Yang, W. (2011). Nucleases: Diversity of structure, function and mechanism. *Quarterly Reviews of Biophysics*. https://doi.org/10.1017/S0033583510000181

Yates, L. A., Aramayo, R. J., Pokhrel, N., Caldwell, C. C., Kaplan, J. A., Perera, R. L., … Zhang, X. (2018). A structural and dynamic model for the assembly of Replication Protein A on single-stranded DNA. *Nature Communications*. https://doi.org/10.1038/s41467-018-07883-7

Yoo, C. M., & Rhee, H. W. (2020). APEX, a Master Key to Resolve Membrane Topology in Live Cells. *Biochemistry*. https://doi.org/10.1021/acs.biochem.9b00785

Yosaatmadja, Y., Baddock, H. T., Newman, J. A., Bielinski, M., Gavard, A. E., Mukhopadhyay, S. M. M., … Gileadi, O. (2021). Structural and mechanistic insights into the Artemis endonuclease and strategies for its inhibition. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkab693

Young, S. J., Sebald, M., Shah Punatar, R., Larin, M., Masino, L., Rodrigo-Brenni, M. C., … West, S. C. (2020). MutSβ Stimulates Holliday Junction Resolution by the SMX Complex. *Cell Reports*. https://doi.org/10.1016/j.celrep.2020.108289

Zannini, L., Delia, D., & Buscemi, G. (2014). CHK2 kinase in the DNA damage response and beyond. *Journal of Molecular Cell Biology*. https://doi.org/10.1093/jmcb/mju045

Zanotti, G. (2016). Cryo-EM and X-Ray Crystallography: Complementary or Alternative Techniques? *NanoWorld Journal*. https://doi.org/10.17756/nwj.2016-025

Zellweger, R., Dalcher, D., Mutreja, K., Berti, M., Schmid, J. A., Herrador, R., … Lopes, M. (2015). Rad51-mediated replication fork reversal is a global response to genotoxic treatments in human cells. *Journal of Cell Biology*. https://doi.org/10.1083/jcb.201406099

Zhang, F., Ma, J., Wu, J., Ye, L., Cai, H., Xia, B., & Yu, X. (2009). PALB2 Links BRCA1 and BRCA2 in the DNA-Damage Response. *Current Biology*. https://doi.org/10.1016/j.cub.2009.02.018

Zhang, J. M., Yadav, T., Ouyang, J., Lan, L., & Zou, L. (2019). Alternative Lengthening of Telomeres through Two Distinct Break-Induced Replication Pathways. *Cell Reports*. https://doi.org/10.1016/j.celrep.2018.12.102

Zhang, J. M., & Zou, L. (2020). Alternative lengthening of telomeres: From molecular mechanisms to therapeutic outlooks. *Cell and Bioscience*. https://doi.org/10.1186/s13578-020-00391-6

Zhao, B., Rothenberg, E., Ramsden, D. A., & Lieber, M. R. (2020). The molecular basis and disease relevance of non-homologous DNA end joining. *Nature Reviews Molecular Cell Biology*. https://doi.org/10.1038/s41580-020-00297-8

Zhao, J., Bacolla, A., Wang, G., & Vasquez, K. M. (2010). Non-B DNA structure-induced genetic instability and evolution. *Cellular and Molecular Life Sciences*. https://doi.org/10.1007/s00018-009-0131-2

Zhelkovsky, A. M., & McReynolds, L. A. (2014). Polynucleotide 3′-terminal phosphate modifications by RNA and DNA ligases. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M114.612929

Zheng, L., Meng, Y., Campbell, J. L., & Shen, B. (2020). Multiple roles of DNA2 nuclease/helicase in DNA metabolism, genome stability and human diseases. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkz1101

Zhou, Z. X., Lujan, S. A., Burkholder, A. B., Garbacz, M. A., & Kunkel, T. A. (2019). Roles for DNA polymerase δ in initiating and terminating leading strand DNA replication. *Nature Communications*. https://doi.org/10.1038/s41467-019-11995-z

Zimmer, J., Tacconi, E. M. C., Folio, C., Badie, S., Porru, M., Klare, K., … Tarsounas, M. (2016). Targeting BRCA1 and BRCA2 Deficiencies with G-Quadruplex-Interacting Compounds. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2015.12.004

Zou, L., & Elledge, S. J. (2003). Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. *Science*. https://doi.org/10.1126/science.1083430