Biochemical studies on the role of the XPF-ERCC1 endonuclease in repairing damaged DNA replication forks

A thesis submitted to the University of Oxford for the degree of

Doctor of Philosophy

Ummi Abdullah

Wolfson College
Weatherall Institute of Molecular Medicine
University of Oxford

Trinity Term 2015

Abstract

Biochemical studies on the role of the XPF-ERCC1 endonuclease in repairing damaged DNA replication forks

A thesis submitted to the University of Oxford for the degree of Doctor of Philosophy

The human XPF (ERCC4) and ERCC1 proteins form a heterodimeric endonuclease that plays a critical role in maintaining genome stability. Mutations in the XPF gene cause several heritable disorders including Fanconi anemia (FA). This highlights the importance of DNA repair processes controlled by XPF-ERCC1, including the repair at replication forks stalled by DNA interstrand crosslinks (ICLs), for human health. Unrepaired ICLs, formed by by-products of cellular metabolism, are believed to be responsible for the genomic instability, cellular attrition and cancer predisposition seen in FA patients. Additionally, ICL-inducing agents are extensively utilised in anti-cancer therapy.

Previous reports suggest that ICL processing is triggered when the leading strand of a replication fork collides with an ICL, where XPF-ERCC1 catalyses fork incisions to initiate a process termed ‘ICL unhooking’. However, the mechanism(s) of fork processing by XPF-ERCC1 remain poorly explored. Using biochemical reconstitution, I show that ICL unhooking, the initiating steps of replication-coupled ICL repair, is driven by the collaborative efforts of XPF-ERCC1; the replicative single-stranded DNA binding protein RPA; and the 5’ to 3’ exonuclease SNM1A.

XPF-ERCC1 incises ICL-containing fork structures within the duplex DNA region. However, the presence of a model nascent leading strand, mimicking the effects of replication arrest by ICL, eliminates this activity. Strikingly, addition of RPA restores XPF-ERCC1 activity on such structures. SNM1A is able to load onto XPF-ERCC1-RPA induced incisions and digest past the ICL to unhook the ICL from duplex DNA.

I postulate that during replication-coupled ICL repair, the arrest of nascent leading strands by ICLs produces a substrate that is inhibitory to XPF-ERCC1. This inhibition can be overcome through the marked stimulation of XPF-ERCC1 by RPA. XPF-ERCC1-RPA induced-incision enables SNM1A to digest past the ICL, which unhooks the ICL from the duplex DNA, enabling subsequent repair process to occur.

Acknowledgement

First and foremost, I would like to thank my supervisor Prof. Dr. Peter McHugh for taking me on as a student. Thank you for giving me the opportunity to work on such an exciting and thought-provoking project. I also would like to thank my funding body, the Malaysia King's Scholarship, for paving the road for me to obtain my DPhil.

To every member, past and present, of the DNA Damage & Repair Group: Anna Olsen, Lonnie Swift, Mika Abu, Blanka Collis, Sovan Sarkar, Tom Ward, Anderson Wang, Lianne Castle, Sook Yee Lee, Hannah Baddock, Erkin Erdal, and Ghadeer Al-Muhaini - thank you for being my friends during my time in Oxford and may our friendship remains for many more years to come.

To my parents Dr. Abdullah Yusoff and Dr. Che Rabiaah Mohamed; my siblings Ira, Wani, Afiq, Alia and Aleef; my brother-in-law Mohd Syafie; my nephew Irsyad; my nieces Sofea and Hanan; thank you for being 'here' with me – in my phone and in my laptop. Your presence, love and support made my life as a DPhil student much more fun and meaningful. Last but no means least, I would like to thank my roommate who is also my personal shrink and cheerleader, Mohd Fathuallah Mohd Nawi, who also happens to be my husband. I am forever grateful to have you by my side during this memorable journey.

My fondest memory during my DPhil was after winning my first ever award as a DPhil student – second place for poster presentation at the '2015 Young Life Scientists' Symposium for DNA Damage in Physiology and Medicine' in France – I told Peter that during the symposium, I was really determined to win because it was my only chance to ever win something during my DPhil. He shrugged his shoulders, smiled and said 'Well, I think this is the first of many'. Little did he know how profound his words were on me that a few months later I went on to win the Best Student Presenter for Molecular Oncology, and then the Ita Askonas Medal as the Best Student Presenter for the Weatherall Institute of Molecular Medicine (WIMM), University of Oxford. Thank you Peter, because of you, I B.E.L.I.E.V.E!

Table of content

List of Abbreviations

ssDNA	Single-stranded DNA
dsDNA	Double-stranded DNA
FA	Fanconi anemia
XP	Xeroderma pigmentosum
CS	Cockayne's syndrome
ROS	Reactive oxygen species
ICL	Interstrand crosslink
DDR	DNA damage response
ATM	Ataxia-telangiectasia mutated
ATR	ATM and Rad-3 related
MRN	Mre11-Rad50-Nbs1
ATRIP	ATR interacting partner
AP	Apurinic/apyrimidic sites
SSB	Single-strand break
DSB	Double-strand break
CPD	Cyclobutane pyrimidine dimer
6-4PP	6-4 pyrimidone photoproducts
UV	Ultraviolet
IR	Irradiation
PIKK	phosphatidylinositol 3-kinases (PI3K)-related protein kinase
MDC1	mediator of DNA damage checkpoint 1
CHK2	Checkpoint kinase 2
CHK1	Checkpoint kinase 1
CDC25A	Cell cycle division 25A
BRCA1	Breast cancer susceptibility gene 1

53BP1	p53 binding protein 1
CDK2	Cyclin-dependent kinase 2
RPA	Replication protein A
XPF	Xeroderma pigmentosum complementation group F
ERCC1	Excision repair cross-complementation group 1
9-1-1	Rad9-Hus1-Rad1
RFC	Replication factor C
TOPBP1	Topoisomerase-binding protein-1
MGMT	O^6-methylguanine-DNA-methyltransferase
AGT	O^6-alkyguanine-DNA alkyltransferase
5'-dRP	5'-deoxyribose-5'-phosphate
NER	Nucleotide excision repair
BER	Base excision repair
MMR	Mismatch repair
hHRad23B	Human homolog of Rad23B
UV-DDB	UV-damaged DNA binding protein
TFIIH	Transcription factor II H
CtIP	C-terminal binding protein (CtBP)-interacting protein
DNA-PKcs	DNA-dependent protein kinase catalytic subunit
BLM	Bloom syndrome
HJ	Holliday Junction
SDSA	Synthesis-dependent-strand-annealing
SSA	Single-strand annealing
$(HhH)_2$	Tandem helix-hairpin-helix domains
OFS	Celebro-oculo-facio-skeletal

1 Introduction

DNA is the exclusive repository of genetic information in all living organisms therefore assaults that challenge its integrity and stability are detrimental to the cells' viability, which may affect the wellbeing of the organisms. Cells have evolved a multitude of DNA repair mechanisms to combat different types of damage to the cellular genome (reviewed in (Iyama and Wilson 2013)). Deficiency in DNA repair results in an array of genetic disorders including Fanconi anemia (FA), Xeroderma pigmentosum (XP) and Cockayne's syndrome (CS). Due to the lethality of DNA damage to cells, many DNA damaging agents are utilised in anti-cancer regimens to selectively kill cancer cells (reviewed in (Deans and West 2011)). However, cells' ability to repair DNA damage causes resistance to the anti-cancer treatments and recurrence of cancer. Therefore, insights on how DNA damage is processed in the cells is not only crucial for our understanding of a variety of DNA repair disorders, it could also improve cancer prevention and treatment strategies.

In this chapter, the major types of DNA damage and their respective repair mechanisms in mammalian cells are introduced. Then, a cytotoxic DNA lesion called DNA interstrand crosslinks (ICLs) and its repair mechanism is described. Subsequently, the subject of this study, the heterodimeric endonuclease protein complex XPF-ERCC1 implicated in incisions of ICLs during ICL repair is discussed in detail. The aims of this study are laid out in detail at the end of this chapter.

1.1 Sources of DNA Damage

1.1.1 Endogenous DNA Damage

DNA is constantly subjected to a myriad of endogenous assaults from its spontaneous chemical reaction in the aqueous cellular environment and its reaction with by-products of other cellular metabolisms. It is estimated that a single human cell faces $10^4 – 10^6$ DNA damage per day, with approximately 10^4 DNA damage are oxidized bases and single-strand breaks (SSB) (Lindahl 1993). Therefore, an adult human requires approximately $10^{16} – 10^{18}$ repair events per day (Hoeijmakers 2009).

A common source of endogenous DNA damage is the spontaneous hydrolysis of *N*-glycosidic bond between the DNA base and the deoxyribose duplex that occurs 18,000 times per day which causes the loss of the DNA base, resulting in abasic or apurinic/apyrimidic (AP) sites (Loeb, et al. 1986). Additionally, DNA bases containing exocyclic amino groups are inclined to spontaneous hydrolysis that occurs 100-500 times per day resulting in deamination, especially of cytosine to uracil (Lindahl 1993). DNA is also subjected to oxidative damage by reactive oxygen species (ROS), molecules with unpaired electrons therefore are highly reactive, generated from normal cellular processes such as aerobic metabolism. ROS can induce the formation of a variety of DNA damage such as SSB or double-strand breaks (DSBs); DNA-protein crosslinks and base modification; and alkylating species resulting in intra- and interstrand crosslinks (ICLs) (Beckman and Ames 1997; Burrows and Muller 1998). Physiological DNA processes also contribute to endogenous DNA damage such as DNA mismatches, insertions and deletions. The endogenous source of ICLs is discussed further in section 1.4.1 of this chapter.

1.1.2 Exogenous DNA Damage

A major source of exogenous or environmental DNA damage is exposure to ultraviolet (UV) light from the sun. UV light induces bulky DNA photoadducts, which are covalent bonds formed between adjacent pyrimidine bases resulting in cyclobutane pyrimidine dimers (CPD) and 6-4 pyrimidone photoproducts (6-4PP) (Ravanat, et al. 2001).

Another major source of exogenous DNA damage is exposure to ionizing radiation (IR) from natural (e.g., cosmic and gamma radiation) and artificial sources (e.g., medical treatments such as X-rays and radiotherapy). IR can directly ionize the DNA by direct absorption of the radiation energy by DNA or absorption by molecules in a cell (i.e., water) forming ROS that can induce a variety of DNA lesions. Two or more DNA lesions within one or two helical turns of DNA by a single radiation track is termed 'DNA clustered damage sites' (reviewed in (Eccles, et al. 2011). Repair of non-DSBs clustered damage sites can result in additional formation of DSBs, which is one of the most harmful DNA lesions that can result in loss of large amount of genetic information.

Another source of exogenous DNA damage is anti-cancer chemotherapeutic agents which selectively kill cancer cells by inducing DNA damage. The most widely-utilised chemotherapeutic agents are bifunctional alkylating agents that produce ICLs. Formation of ICLs from anti-cancer therapy is discussed further in section 1.4.2 of this chapter.

1.2 Cellular responses to DNA damage

Cellular response to DNA damage encompasses the physical reversal or excision of DNA lesions by specific repair pathways and the activation of signal transduction cascade collectively termed ‘DNA damage response’ (DDR) that would lead to a number of cellular fates including: cell cycle arrest to allow time for repair before replication restart and cell division; cellular senescence or programmed cell death termed ‘apoptosis’ if the lesions are irreparable; and activation of appropriate DNA repair mechanisms. **Figure 1. 1** summarises cellular responses to DNA damage.

1.2.1 DNA damage response (DDR)

The central components in DDR signal transduction cascades are the phosphatidylinositol 3-kinases (PI3K)-related protein kinases (PIKKs) family of proteins which are serine/threonine-protein kinases that regulate signal transduction through phosphorylation (Lempiainen and Halazonetis 2009). Two PIKKs are the major players in DDR: ATM (ataxia-telangiectasia mutated) and ATR (ATM and Rad-3 related), activated by DSBs and a broad spectrum of DNA damage, respectively, to promote cell cycle arrest and DNA repair (Cimprich and Cortez 2008). ATM and ATR have an overlapping but non-redundant roles in DDR. **Figure 1. 2** provides a schematic representation of the ATM and ATR signal transduction cascades.

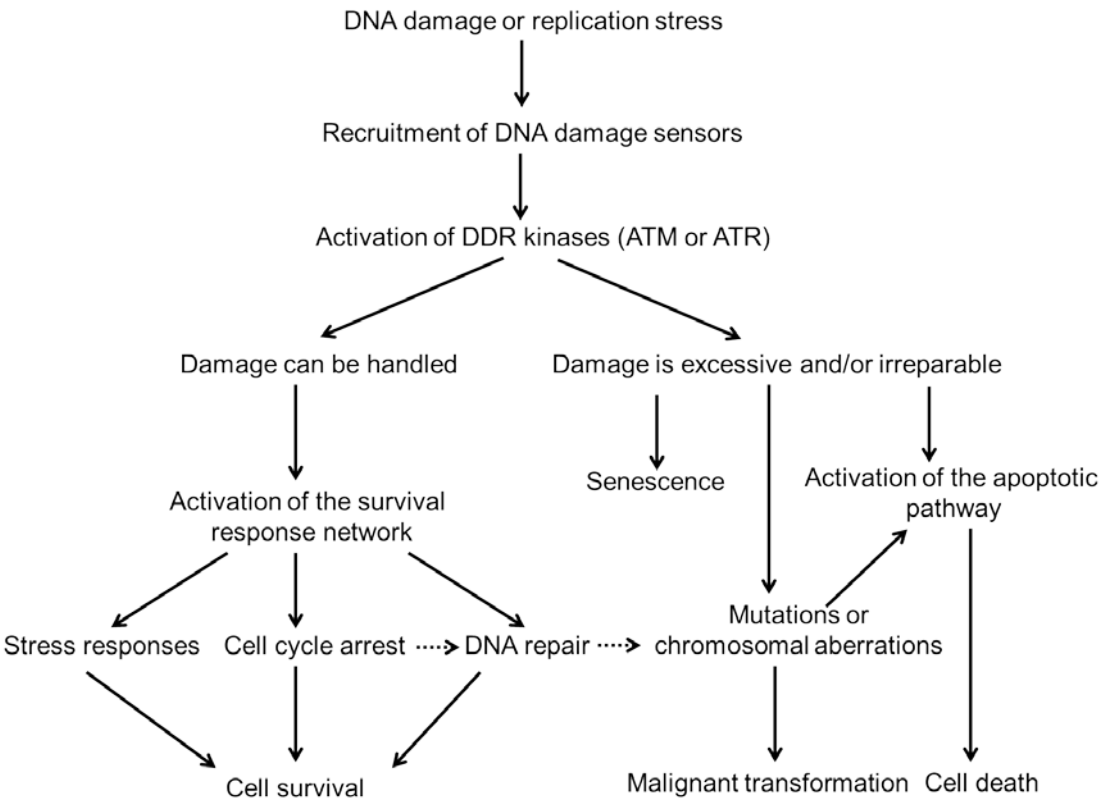


Figure 1. 1: A simplified representation of cellular responses to DNA damage.

Following detection of DNA damage, DNA damage sensors are recruited to damage sites followed by the activation of DNA damage response (DDR) kinases ATM or ATR. The basic response to DNA damage is the activation of the survival response network which includes stress responses (i.e., delayed origin firing) and cell cycle arrest to allow time for DNA repair. Mutations or chromosomal aberrations can occur if the damage is incorrectly repaired or left unrepaired when the damage is irreparable. Alternatively, the cells can undergo cellular senescence or the apoptotic pathway which leads to cell death. Adapted from (Shiloh 2003).

1.2.1.1 ATM activation

Formation of DSBs induces the recruitment of the MRN (Mre11-Rad50-Nbs1) complex to the DSBs (Stracker and Petrini 2011), which in turn recruits ATM to the DSBs and activates the kinase activity of ATM (Lamarche, et al. 2010). Activated ATM phosphorylates a network of target substrates (Matsuoka, et al. 2007) including BRCA1 (breast cancer susceptibility gene 1), 53BP1 (p53 binding protein 1) and histone variant H2AX at serine-139 (γH2AX formation) (Burma, et al. 2001). γH2AX facilitates the recruitment of MDC1 (mediator of DNA damage checkpoint 1) to the DSBs that amplify the DDR signalling by additional recruitment of ATM-MRN complexes. Activated ATM also phosphorylates CHK2 (checkpoint kinase 2) which in turn phosphorylates CDC25A (cell cycle division 25A) phosphatase, targeting it for degradation by proteolysis. Loss of CDC25A phosphatase prevents the removal of the inhibitory phosphorylations on CDK2 (cyclin-dependent kinase 2), preventing its binding to cyclin E and B resulting in cell cycle arrest from G1-to-S or G2-to-M respectively (Lobrich and Jeggo 2007). Alternatively, ATM can phosphorylate p53 which leads to the transcription of numerous p53 target genes including CDK inhibitor p21, which leads to G1 – S arrest to promote replication arrest to resolve the DSBs, or apoptosis. DSB repair is further discusses in section 1.2.3.5 of this chapter.

1.2.1.2 ATR activation

ATR is activated in response to a broad spectrum of DNA damage including DSBs and DNA lesions that impede replication. ATR and ATRIP (ATR interacting partner) are recruited to ssDNA, an intermediate of a variety of DNA damage, which is rapidly coated by the single-strand DNA binding protein RPA (replication protein A) suggesting that ssDNA is the structure that activates the kinase activity of ATR (Zou

and Elledge 2003). ATR is also activated by DSBs following end-resection of DSB ends that generates ssDNA (Jazayeri, et al. 2006). In addition to ATR-ATRIP, the 9-1-1 (Rad9-Hus1-Rad1) complex is also recruited to RPA-coated ssDNA which is loaded onto ssDNA by the clamp-loader RAD17-RFC (RAD17-replication factor C). This leads to the recruitment of an ATR activator TOPBP1 (topoisomerase-binding protein-1) (Cimprich and Cortez 2008) and Rhino (Cotta-Ramusino, et al. 2011). In the presence of Claspin, TOPBP1 activates ATR which in turn phosphorylate CHK1 (checkpoint kinase 1). Activated CHK1 induces cell cycle arrest by phosphorylating CDC25A phosphatase, which targets it for degradation. Loss of CDC25A phosphatase prevents the removal of the inhibitory phosphorylations on CDK2 and preventing its binding to cyclin E, resulting in arrest in S phase. ATR could also stabilise and promote replication fork restart. ATR also phosphorylates DNA repair factors including FANCD2-FANC1 complex, BRCA1, WRN and BLM to promote DNA repair (Andreassen, et al. 2004).

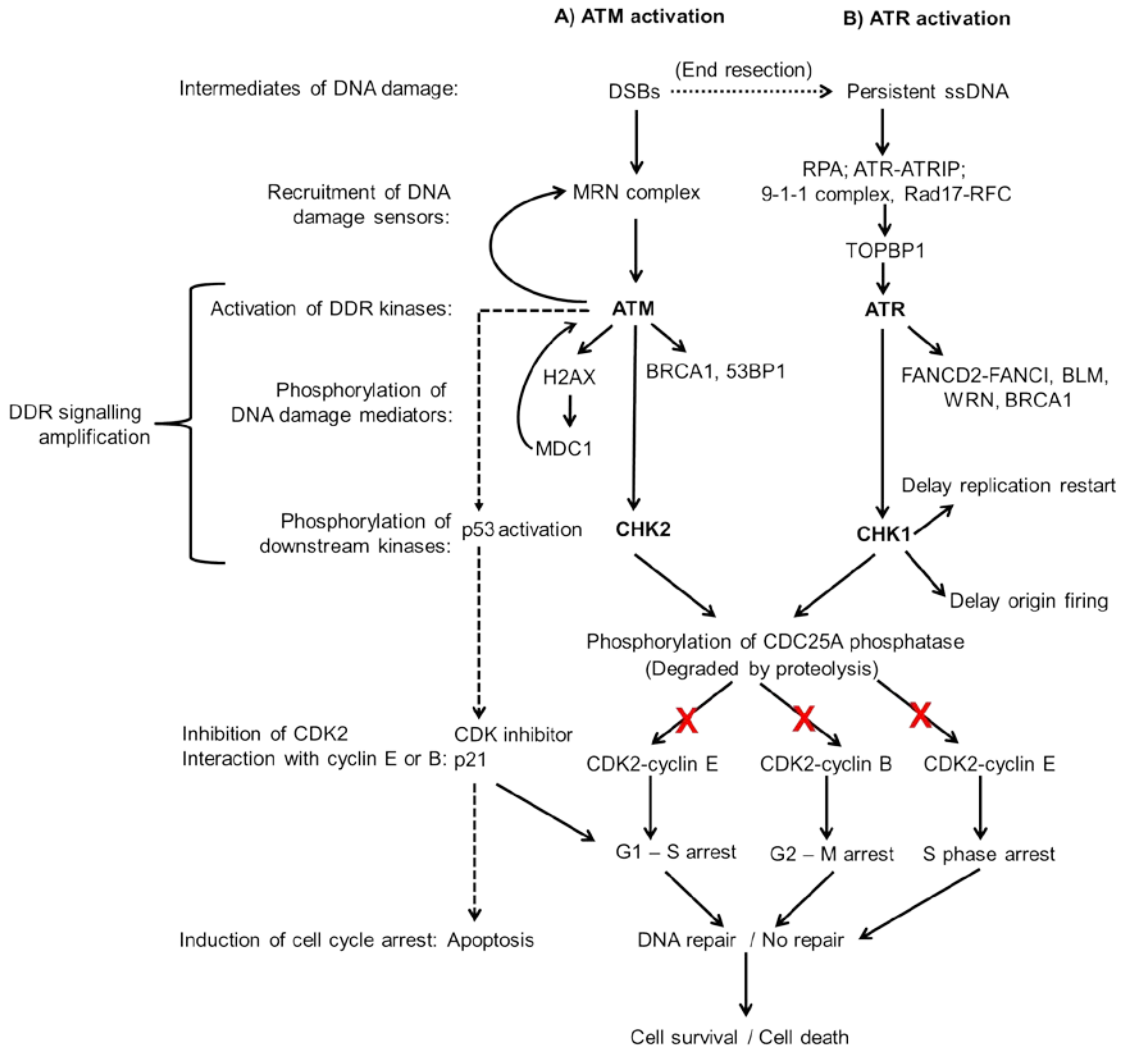


Figure 1. 2 A simplified representation of the ATM and ATR signal transduction cascades.

A) ATM activation: Induction of DSBs induces the recruitment of the MRN complex and the activation of ATM. Activated ATM phosphorylates an array of target substrates including BRCA1, 53BP1 and H2AX (γH2AX formation). γH2AX facilitates the recruitment MDC1 that amplify the DDR signalling. ATM also phosphorylates CHK2 which in turn phosphorylate CDC25A phosphatase, targeting it for degradation. Loss of CDC25A phosphatase prevents the removal of the inhibitory phosphorylations on CDK2, inhibiting its binding to cyclin E and B resulting in cell cycle arrest from G1-to-S or G2-to-M respectively. Alternatively, ATM can phosphorylate p53 which leads to

the transcription of numerous p53 target genes including CDK inhibitor p21 resulting in G1-S phase arrest or apoptosis. B) ATR activation: The presence of persistent ssDNA from stalled replication forks or DSB processing induces the recruitment of RPA, ATR-ATRIP, 9-1-1 complex and RAD17-RFC. An ATR activator TOPBP1 is then recruited to activate ATR. ATR in turn phosphorylates a variety of target substrates including DNA repair factors such as FANCD2-FANCI complex, WRN and BLM and BRCA1. CHK1 is also activated by ATR which in turn phosphorylate CDC25A phosphatase, targeting it for degradation. Loss of CDC25A phosphatase prevents the removal of the inhibitory phosphorylations on CDK2 and inhibiting its binding to cyclin E, resulting in arrest in S phase. DNA repair could occur during cell cycle arrest by ATM and ATR activation which could lead to cell survival. However, if the DNA lesions are not repaired, cell death could occur. Red crosses represent inhibition of downstream process. Adapted from (Cimprich and Cortez 2008; Shiloh 2003).

1.2.2 DNA repair pathways

1.2.2.1 Reversal of DNA damage

Reversal of DNA damage is a 'direct repair' process without the incision of the DNA backbone. Three major mechanisms of DNA damage reversal are: 1) reversal of UV light-induced photolesions by photolyases; 2) reversal of O-alkylated DNA damage by O^6-alkyguanine-DNA alkyltransferase (AGTs); and 3) reversal of N-alkylated base adducts by AlkB family of dioxygenases.

UV light-induced DNA damage CPDs and 6-4PPs are reversed by CPD photolyases and [6-4] photolyases, respectively, which use visible light and fully reduced flavin cofactor FADH(-) to monomerise the intradimer bonds of CPDs and 6-4PPs (Brettel and Byrdin 2010). Methylation of O^6 atom on guanines (O^6-meG) causes mispairing of the O^6-meG with thymine instead of cytosine. O^6-meG are reversed by a 'suicide protein' O^6-methylguanine-DNA-methyltransferase (MGMT) which transfers the methyl group on the O^6 atom to a cysteine residue at its active site therefore restoring guanine to its undamaged state. This in turn inactivates MGMT, which is then ubiquitinated and degraded by proteasome (Iyama and Wilson 2013). N-alkylated base adducts are reversed by human homologues of *E.coli* AlkB family of dioxygenases using an iron (II) site to activate the dioxygen molecule for oxidation of the alkyl group attached to the N^1 atom of adenine and N^3 atom of cytosine, which restore the base to its undamaged state (Yi, et al. 2010).

1.2.2.2 Base excision repair (BER)

Base excision repair (BER) is the main pathway for the repair of chemically modified or damaged DNA bases resulting from deamination (e.g: cytosine to uracil), alkylation (e.g: 3-methyladenine) or oxidation (e.g: 8-oxoguanine) (reviewed in (Krokan and Bjoras 2013)). The BER pathway is activated when the damaged bases are recognised and excised by DNA glycosylases which hydrolyse the *N*-glycosidic bond between the base and deoxyribose, creating an apurinic/apyrimidinic (AP) site. The BER sub-pathways is summarised in **Figure 1. 3**.

If the damaged base is recognised by a monofunctional DNA glycosylase such as MPG (N-methylpurine DNA glycosylase), an AP endonuclease APE1 is recruited to hydrolyse the phosphodiester bond 5' to the AP site resulting in 3'-OH and 5'-dRP (5'-deoxyribose-5'-phosphate) (Wilson and Barsky 2001). A bifunctional DNA glycosylase such as OGG1 (8-oxoguanine DNA glycosylase) exhibit an intrinsic 3' AP lyase activity in addition to its DNA glycosylase activity, therefore it is able to incise 3' of the AP site resulting in either 3'-PUA (3'-phosphate α,β-unsaturated aldehyde) and 5'-phosphate or 3'-phosphate and 5'-phosphate. To enable subsequent repair synthesis and ligation, 3'-PUA is removed by AP endonuclease APE1 while 3'-phosphate is removed by polynucleotide kinase (PNK) (Wiederhold, et al. 2004). BER typically proceeds via the 'short-patch' (SP) sub-pathway which utilises DNA polymerase β to incorporate a single nucleotide to replace the missing nucleotide and the nick is ligated by DNA ligase IIIα-XRCC1 complex (Sobol, et al. 1996).

An alternative BER sub-pathway is the ‘long-patch’ (LP) strand displacement synthesis sub-pathway which typically occurs during S phase or when the level of ATP is low causing low ligation efficiency. During LP, the replicative DNA polymerase δ or ε and associated factors PCNA and RCF synthesizes 2-13 oligonucleotide (nt) past the AP site (Gary, et al. 1999). This results in a 5’-flap structure which is incised by the 5’-flap endonuclease Fen-1, and the nick is sealed by DNA ligase (Levin, et al. 2004).

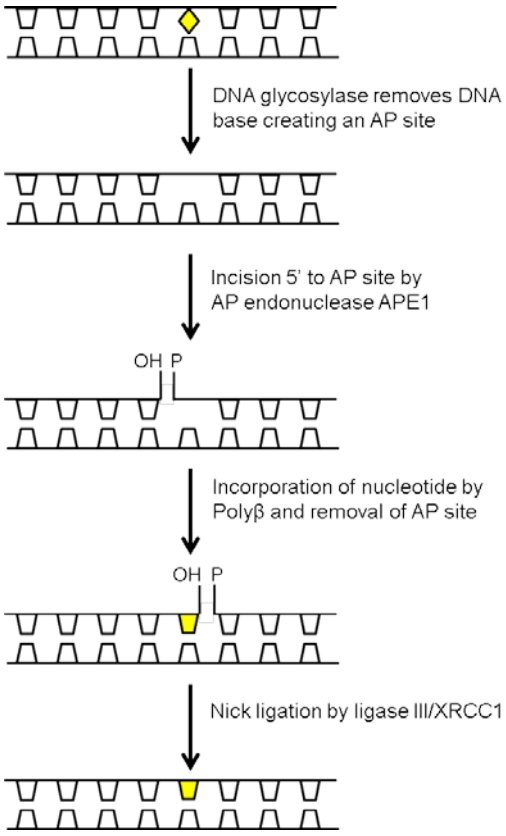


Figure 1. 3: A schematic model of the predominant BER pathway in mammalian cells.

DNA glycosylase removes the damaged base by hydrolysing the *N*-glycosidic bond between the base and sugar phosphate backbone, creating an apurinic/apyrimidinic (AP) site. The AP site is recognised by an AP endonuclease APE1 which hydrolyse the phosphodiester bond 5' to the AP site. Polymerase β then incorporates a single nucleotide and hydrolyses the phosphodiester bond 3' to the AP site by its AP lyase activity resulting in the removal of the AP site from the DNA. The nick is then sealed by DNA ligase III which interact with polymerase β through XRCC1. Adapted from (Scharer 2003). Alternative pathways in BER are described in the text.

1.2.2.3 Mismatch repair (MMR)

DNA mismatch repair pathway is the main method for eliminating base-base mismatches and nucleotide deletions/insertions introduced mainly from replication errors by DNA polymerases, but can also arise from other biological processes (Jiricny 2006). MMR is crucial because it reduces replication error by 100 orders of magnitude (reviewed in (Scharer 2003)). In mammalian cells, the MMR protein machinery is directed to the error-containing DNA strand by discriminating between the parental strand and the newly synthesized daughter strand. Lagging strand is recognised by the transient strand discontinuities of the Okazaki fragments but the recognition of the leading strand is still poorly understood. The strand discontinuities of the Okazaki fragments are marked by the presence of PCNA or RFC at the 3' or 5' end of the break, respectively, which is recognised by the MMR factors during repair (Dzantiev, et al. 2004). **Figure 1. 4** provides a schematic model of the MMR pathway in mammalian cells.

The MMR pathway is initiated by the recruitment of MutS heterodimer to the mismatch base in the daughter strand (Drotschmann, et al. 2001). In human, there are two isoforms of MutS heterodimer: MutSα heterodimer (consisting of MSH2 and MSH6) which is the most abundant MutS heterodimer is recruited to single mismatches or short (1-2 nt) insertion/deletion loops (IDL) (Drummond, et al. 1995; Palombo, et al. 1995); and the MutSβ heterodimer (consisting of MSH3 and MSH6) is recruited to longer (2 nt or longer) IDLs (Genschel, et al. 1998). The binding of MutS proteins to the base mismatches triggers the recruitment of the MutL heterodimers (Li and Modrich 1995). There are three isoforms of MutL heterodimer: MutLα (consisting of MLH1 and PMS2)

is the major MutL himolog that participates in MMR and is recruited if the pathway is initiated by MutSα; and MutLβ (consisting of MLH1 and MLH3) and MutLγ (consisting of MLH1 and PMS1). The MutSα-MutLα complex undergoes an ATP-driven conformational change to a 'sliding clamp' that enables its translocation away from mismatch (Gradia, et al. 1999). MutSα-MutLα complex that moves 5' upstream of the mismatch will encounter RFC bound at the 5'-end of the strand break. RFC is replaced by Exo1 exonuclease which, guided by the MutSα-MutLα complex, degrades the strand in a 5' to 3' direction past the mismatch base, generating a ssDNA gap that is stabilised by RPA (Genschel, et al. 1998). Upon removal of the mismatch, Exo1 endonuclease activity is inhibited by the MutSα-MutLα complex and RPA, displacing it from the DNA. Polymerase δ is then recruited to fill in the ssDNA gap and DNA ligase I sealed the nick to complete the repair process.

Alternatively, when the MutSα-MutLα complex moves downstream of the mismatch or when the mismatch is on a leading strand, the latent endonuclease activity of MutLα is activated upon binding with MutSα, which promote the incision 5' of the mismatch. The nicks generated by MutLα enable the loading of Exo1 which digest the strand in the 5' to 3' direction past the mismatch (Kadyrov, et al. 2006).

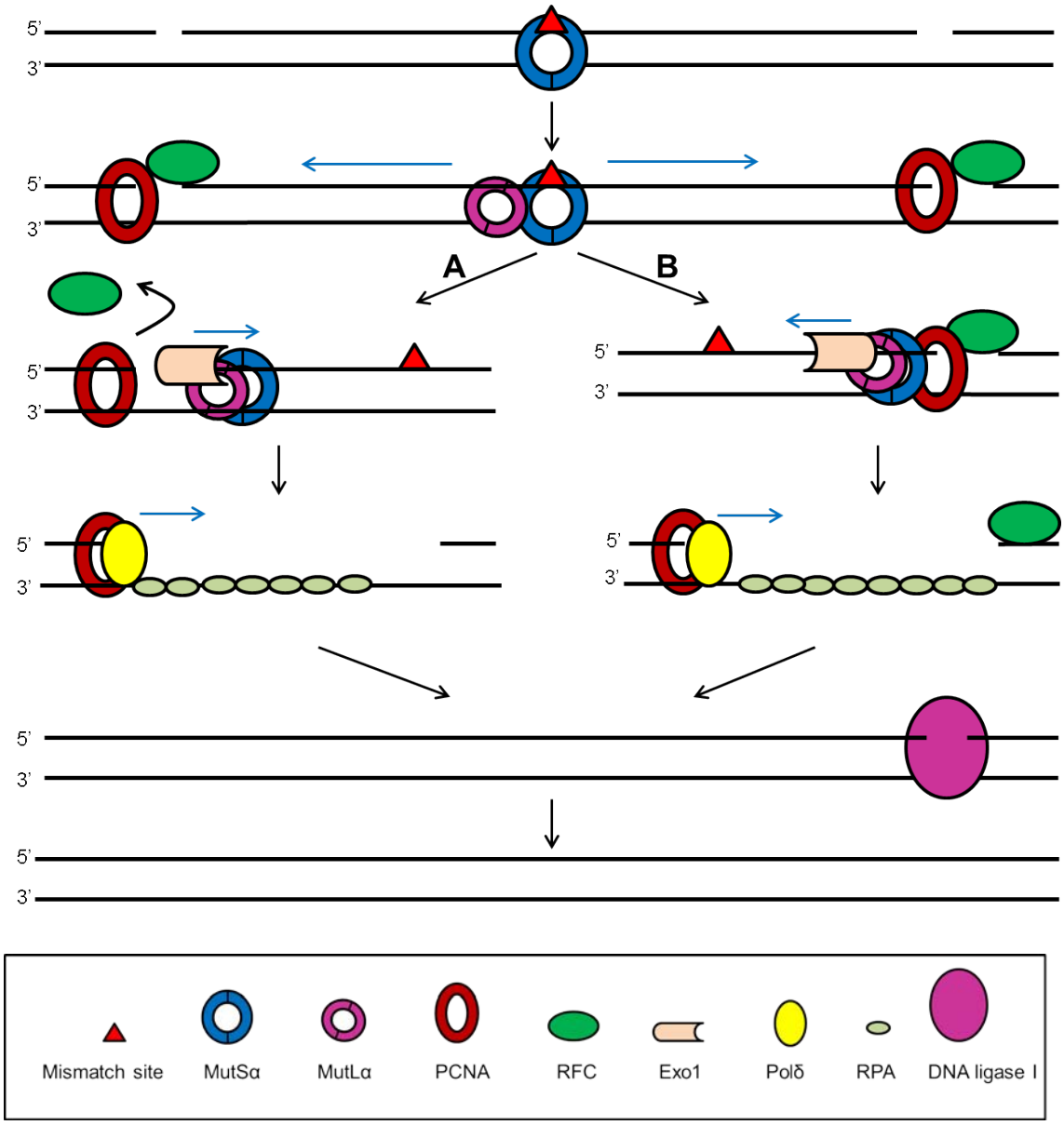


Figure 1. 4 A simplified representation of the MMR pathway in mammalian cells.

The erroneous base (red triangle) on the newly synthesized daughter strand (as seen by the discontinuities of the Okazaki fragments) is detected by MutSα. MutLα is then recruited and the MutSα-MutLα complex undergoes an ATP-driven conformational change to a 'sliding clamp' that enables them translocate away from the mismatch. A) MutSα-MutLα that move upstream of the mismatch encounters RFC bound at the 5'-end of the strand break. RFC is replaced by Exo1 exonuclease which, guided by the

MutSα-MutLα complex, degrades the strand in a 5' to 3' direction past the mismatch base, generating a ssDNA gap that is stabilised by RPA. Upon removal of the mismatch, Exo1 endonuclease activity is inhibited by the MutSα-MutLα complex and RPA, displacing it from the DNA. Polymerase δ is then recruited to fill in the ssDNA gap and DNA ligase I sealed the nick to complete the repair process. B) Alternatively, when the MutSα-MutLα complex moves downstream of the mismatch or when the mismatch is on a leading strand, the latent endonuclease activity of MutLα is activated upon binding with MutSα, which promote the incision 5' of the mismatch. The nicks generated by MutLα enable the loading of Exo1 which digest the strand in the 5' to 3' direction past the mismatch. Adapted from (Jiricny 2006).

1.2.2.4 Nucleotide excision repair (NER)

Nucleotide excision repair (NER) is the main pathway in mammals that resolves helix-distorting bulky DNA lesions such as cyclobutane pyrimidine dimers (CPDs) and 6,4-photoproducts (6-4PPs) formed by UV light; environmental mutagens and chemotherapeutic agents such as cisplatin (Gillet and Scharer 2006). The efficiency of repair by NER was proposed to correlate with the degree of helical distortion caused by the lesion (Gunz, et al. 1996). The global genome NER (GG-NER) sub-pathway mediates repair throughout the genome (Gillet and Scharer 2006) while the transcription-coupled NER (TC-NER) is specific for lesions on actively transcribed strand of DNA that blocks the progression of RNA polymerase II (RNAPII) (Hanawalt and Spivak 2008). The sub-pathways differ at the damage recognition stage, but share the same downstream processes of repair. **Figure 1. 5** provides a schematic model of the NER sub-pathways in mammalian cells.

In GG-NER, damage recognition is initiated by the recruitment of XPC and hHRad23B (human homolog of Rad23B) to the undamaged strands opposite the lesion but less helix distorting lesion such as CPD utilises UV-DDB1/2 (UV-damaged DNA binding proteins 1 and 2) (Sugasawa, et al. 2001). The ability of the damage recognition machinery to recognise a variety of the DNA lesion is due to its ability to detect the destabilisation of the DNA helix and bulkiness of the lesion and not the chemical properties of the lesion (Hess, et al. 1997). Following damage recognition, transcription factor II H (TFIIH) complex which consists of ten subunits is recruited via direct interaction with XPC-hHRad23B (Riedl, et al. 2003). TFIIH subunit XPB promote unwinding of the DNA by its 3' to 5' helicase activity (Fan, et al. 2006). The local denaturation of the DNA generates a 'bubble' structure, which comprises of two dsDNA-ssDNA junctions and regions of ssDNA. The formation of the bubble structure enable TFIIH subunit XPD that track along the DNA until it locks onto the lesion and verifies the chemical modification and bulkiness of the lesion (Mathieu, et al. 2010; Mathieu, et al. 2013). This promotes the recruitment of the pre-incision complex which comprises of XPA, RPA and XPG; and the departure of XPC-hHRad23B (Riedl, et al. 2003). XPA plays a centre-stage role because its binding to 5' side of the bubble structure is suggested for interaction with XPF-ERCC1 and RPA for correct incisions of the DNA lesion (Krasikova, et al. 2010). RPA is suggested to bind to the nondamaged strand and facilitate correct positioning of the XPF-ERCC1 and XPG endonucleases (de Laat, et al. 1998b). Once XPF-ERCC1 is recruited to the bubble complex via interaction with XPA, it incises 5' to the lesion on the damaged strand (Fagbemi, et al. 2011). Repair synthesis is initiated by DNA polymerase δ and DNA pol κ or DNA pol ε and associated factors PCNA and RFC from the 3'-OH generated by incision by XPF-ERCC1 (Lehmann 2011; Staresincic, et al. 2009). This is followed by incision 3' to the

lesion by XPG which releases the lesion-containing oligonucleotide with TFIIH bound to it (Kemp, et al. 2012). The nick is sealed by DNA ligase IIIα/XRCC1 or DNA ligase I for completion of repair (Moser, et al. 2007). The role of XPF-ERCC1 in NER is further described in **section 1.11.1** of this chapter.

In TC-NER, stalling of RNAPII at the DNA lesion within the gene being transcribed serve as a signal for the recruitment of CSB (ERCC6), CSA (ERCC8) and XAB2 (Hanawalt and Spivak 2008). Chromatin remodelling activity of CSB occurs upstream of the stalled RNAPII to enable RNAPII to back tract from the lesion, allowing downstream NER factors to be recruited to the damaged site. The CSA-containing ubiquitin ligase complex is recruited to the damaged site via a CSB-dependent manner and is implicated in ubiquitination and degradation of CSB, which terminate TC-NER and restore transcription (Groisman, et al. 2006).

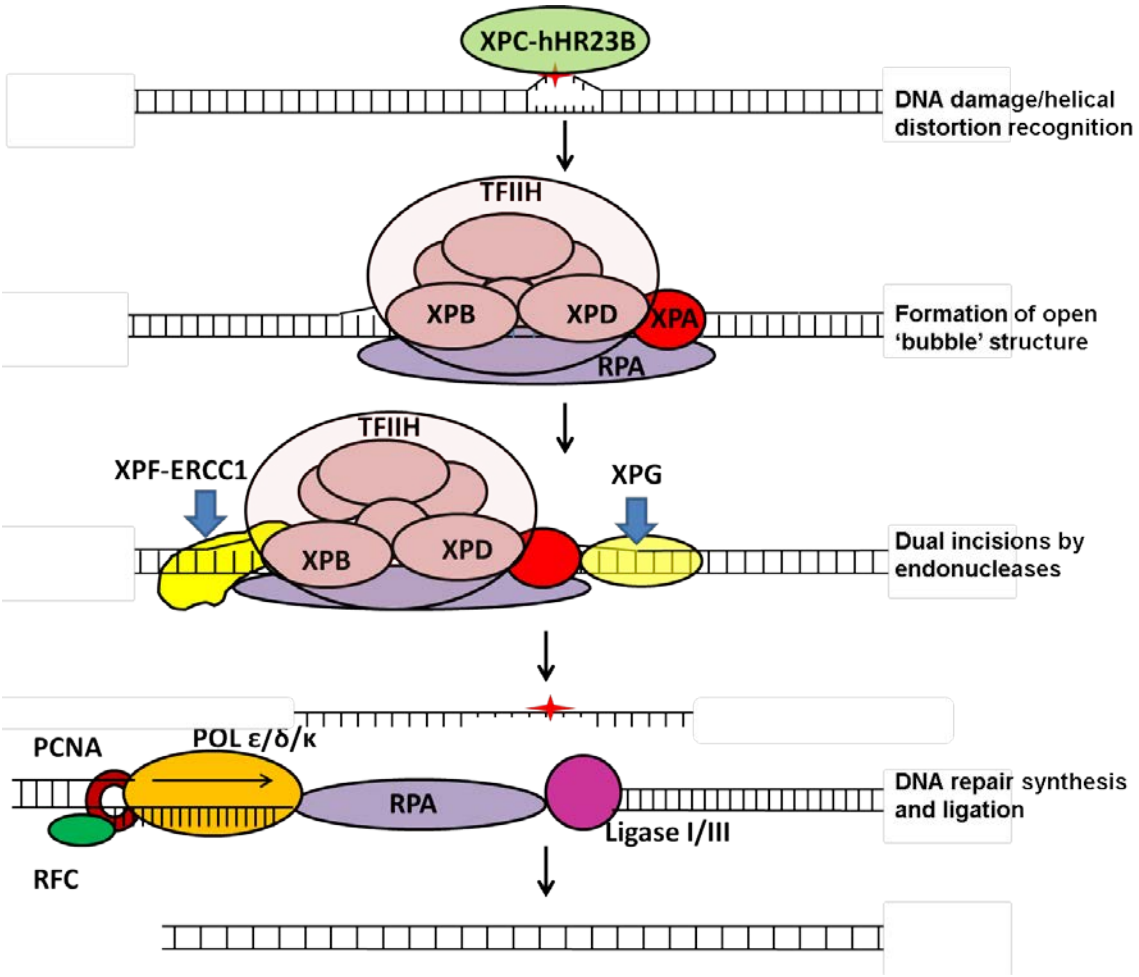


Figure 1. 5 A schematic representation of GG-NER pathway in mammalian cells.

In the presence of a DNA lesion, XPC and hHR23B are recruited to the undamaged complementary strands. The 3' to 5' helicase XPB and the 5' to 3' helicase XPD, subunits of transcription factor II H (TFIIH) unwind the DNA generating a 'bubble' structure. XPA and RPA are then recruited to the 'bubble' structure and facilitate the correct positioning of XPG and XPF-ERCC1 which incises 3' and 5' to the lesion, respectively. The incisions released a 30-mer oligonucleotide containing the lesion. The ssDNA gap is then filled by a DNA polymerase possibly Pol δ,κ or ε and re-ligated for completion of repair. Adapted from (Friedberg 2001).

1.2.2.5 Double-strand break (DSB) repair

DNA double-strand breaks (DSBs) are amongst the most biological hazardous types of DNA damage because it can result in chromosomal aberrations that can affect many genes and loss of large amount of genetic information (Bohgaki, et al. 2010; Jackson and Bartek 2009). DSBs are repaired through two independent pathways which differ in the requirement of homologous DNA template and fidelity of repair: non-homologous end joining (NHEJ) which involves direct ligation of the broken ends for repair and thus error-prone because it causes a small deletion of the DNA sequence (Lieber 2010); and homologous recombination (HR) which utilises homologous DNA of the undamaged sister chromatid as a template for repair and is largely error-free (Li and Heyer 2008). NHEJ is the predominant pathway in mammalian cells operating at all phases of the cell cycle, while HR is restricted to the late-S and G2 phases.

1.2.2.5.1 Non-homologous end joining (NHEJ)

NHEJ which involves direct ligation of the broken ends for repair is activated by the recruitment of the Ku70/Ku80 heterodimer (Ku), which adopts a symmetrical ring-shaped structure that encircles the exposed termini of the DSB (Walker, et al. 2001). The presence of Ku at DSB ends juxtaposed the two ends and prevents aberrant end-resection and suppresses its repair by other pathways (Kass and Jasin 2010). Ku serves as a scaffold for DNA-PKcs (DNA-dependent protein kinase catalytic subunit) which upon recruitment causes the translocation of the Ku heterodimer inward on the dsDNA strand allowing DNA-PKcs to contact DNA termini (Yoo and Dynan 1999). Binding of DNA-PKcs to DSB ends promotes phosphorylation of target substrates including autophosphorylation of DNA-PKcs *in trans* across the DSBs (Gottlieb and Jackson

1993). It also induces synapsis or tethering of the DNA ends which facilitates the recruitment of downstream NHEJ factors. XLF stimulates DNA ligase IV complex (consisting of ligase IV and XRCC4) to seal the DNA ends (Chen, et al. 2000).

Figure 1. 6 provides a simplified schematic of NHEJ pathway in mammalian cells.

The mechanism described above is utilised for aligned, compatible DNA termini that possess a 5' phosphate and a 3' hydroxyl. However, more complex DNA ends require additional factors to process the DNA ends: the exonuclease Artemis is required for opening of DNA hairpins by trimming the 3' and 5' ssDNA overhangs to reveal complementary nucleotide stretches upon stimulation by DNA-PK (Ma, et al. 2002); the bifunctional polynucleotide kinase/phosphatase (PNKP) is utilised to phosphorylate 5'-OH termini and dephosphorylate 3'-phosphate termini upon interaction with XRCC4, therefore providing the correct chemical groups required for DNA ligation (Chappell, et al. 2002). The complex DNA ends are usually produced by irradiation or ROS, or the hairpin coding ends produced during rearrangement of the V(D)J immunoglobulin genes. NHEJ is prone to errors because gaps in DNA are likely to arise during end alignment and processing: DNA ends can be rejoined in the incorrect orientation or misjoining could occur when more DSBs are in close proximity; and DNA sequence is likely to be altered if additional processing is required at the DNA ends. A sub-pathway of NHEJ is termed 'microhomology' which utilises larger stretches of sequence homology than required for NHEJ but is not active unless core NHEJ factors are deficient.

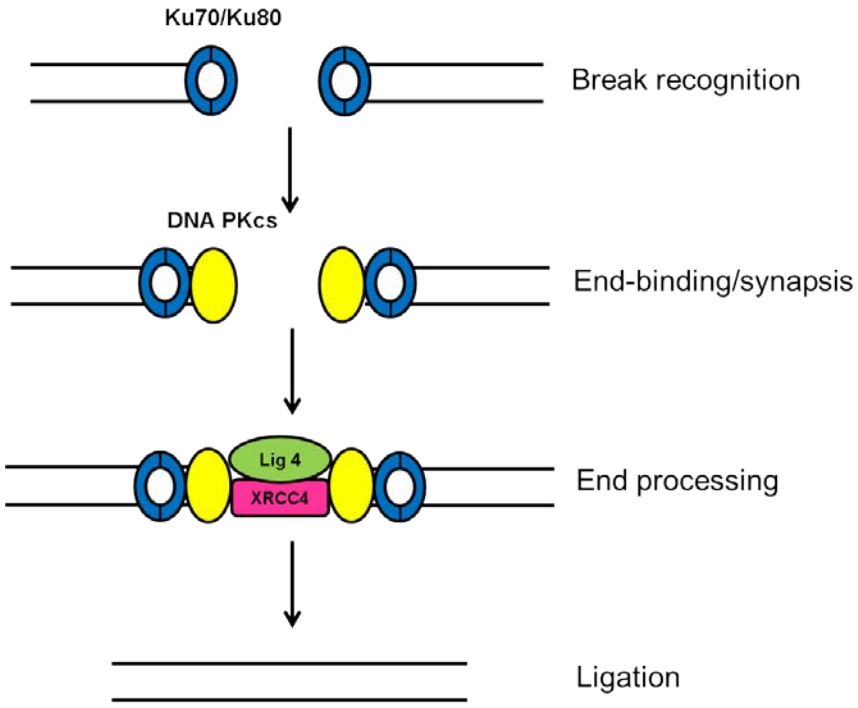


Figure 1. 6 Simplified schematic of NHEJ pathway in mammalian cells.

Ku70/Ku80 heterodimer (Ku) is recruited and encircles the exposed termini of the DSB. Upon recruitment of DNA-PKcs, Ku heterodimer is translocated inward on the dsDNA strand allowing DNA-PKcs to contact DNA termini. Binding of DNA-PKcs to DSB ends promotes phosphorylation of target substrates as well as autophosphorylation of DNA-PKcs *in trans* across the DSBs. The synapsis or tethering of the DNA ends by DNA-PKcs facilitates the recruitment of downstream NHEJ factors. XLF stimulates the activity of DNA ligase IV complex (consisting of ligase IV and XRCC4) is then recruited to DSBs to seal the DNA ends.

1.2.2.5.2 Homologous recombination (HR)

During HR, DSB ends are detected by the MRN complex (Stracker and Petrini 2011) which in turn recruits ATM to the DSBs and activates the kinase activity of ATM (Lamarche, et al. 2010), as discussed in **section 1.2.1.1** of this chapter. MRN associates with CtIP (C-terminal binding protein (CtBP)-interacting protein) which initiates the 5' to 3' end resection of DSB termini by CtIP and the exonuclease activity of Mre11, generating 3' ssDNA overhangs required for strand exchange (Sartori, et al. 2007). Additional resection is carried out by Exo1 (human exonuclease 1) in cooperation with BLM (Bloom Syndrome) helicase (Nimonkar, et al. 2008). The resulting 3'-ssDNA overhangs are rapidly coated by RPA which activates the kinase activity of ATR by its interaction with ATRIP (Zou and Elledge 2003), as discussed in **section 1.2.1.2** of this chapter. Subsequently, RPA is displaced by Rad51 in a Rad52-dependent manner (McIlwraith and West 2008). Brca2 stimulates the assembly of Rad51 nucleoprotein filaments on the 3'-ssDNA overhang (Jensen, et al. 2010) which invades into the intact homologous sister chromatid forming a Holliday Junction (HJ) structure. Repair proceeds via either one of the independent sub-pathways: HJ resolution by structure-selective endonucleases**;** HJ dissolution via helicase-mediated branch migration; synthesis-dependent-strand-annealing (SDSA); and single-strand annealing (SSA). **Figure 1. 7** provide a schematic representation of the HR sub-pathways in mammalian cells.

During the resolution and dissolution sub-pathways of HR, whereby following strand invasion into the intact sister chromatid by Rad51-coated ssDNA, the 3′-end of the invading strand serves as a primer for DNA synthesis and strand extension possibly by

DNA polymerase η (McIlwraith, et al. 2005) forming a displacement loop (D-loop) structure. The second DSB end can be captured by the D-loop, forming a Double Holliday Junction (DHJ) structure that can be processed that can undergo 'dissolution' whereby the DHJ is brought together by branch migration by the helicase activity of the BLM/hTOP3α/hRMI1 complex resulting in non-crossover products (Wu and Hickson 2006). The DHJ can also undergo 'resolution' by several structure-selective endonucleases including GEN1 SLX1-SLX4 and MUS81-EME1 resulting in non-crossover products if the DHJ is incised in the same orientation (Svendsen and Harper 2010). However, changes to the template strand (crossover) occur if the DHJs are incised in different orientations, which mainly occur in meiotic cells. SDSA sub-pathway occurs without the formation of DHJ whereby the newly synthesized strand can be displaced from the D-loop by DNA helicase(s) and anneals with its complementary strand (Ferguson and Holloman 1996). Further synthesis occurs and repair is completed without changes to the template DNA (non-crossover).

The SSA sub-pathway which is extensively investigated in yeast takes place when two repeat sequences are present on both sides of a DSB (Haber 2006). The repeat sequences can anneal with each other forming 3'-flaps which is then cleaved by Rad1-Rad10 complex, a yeast homolog of human XPF-ERCC1, resulting in sequence deletions (Fishman-Lobell and Haber 1992). The role of Rad1-Rad10/XPF-ERCC1 in DSB repair is further described in **section 1.11.2** of this chapter.

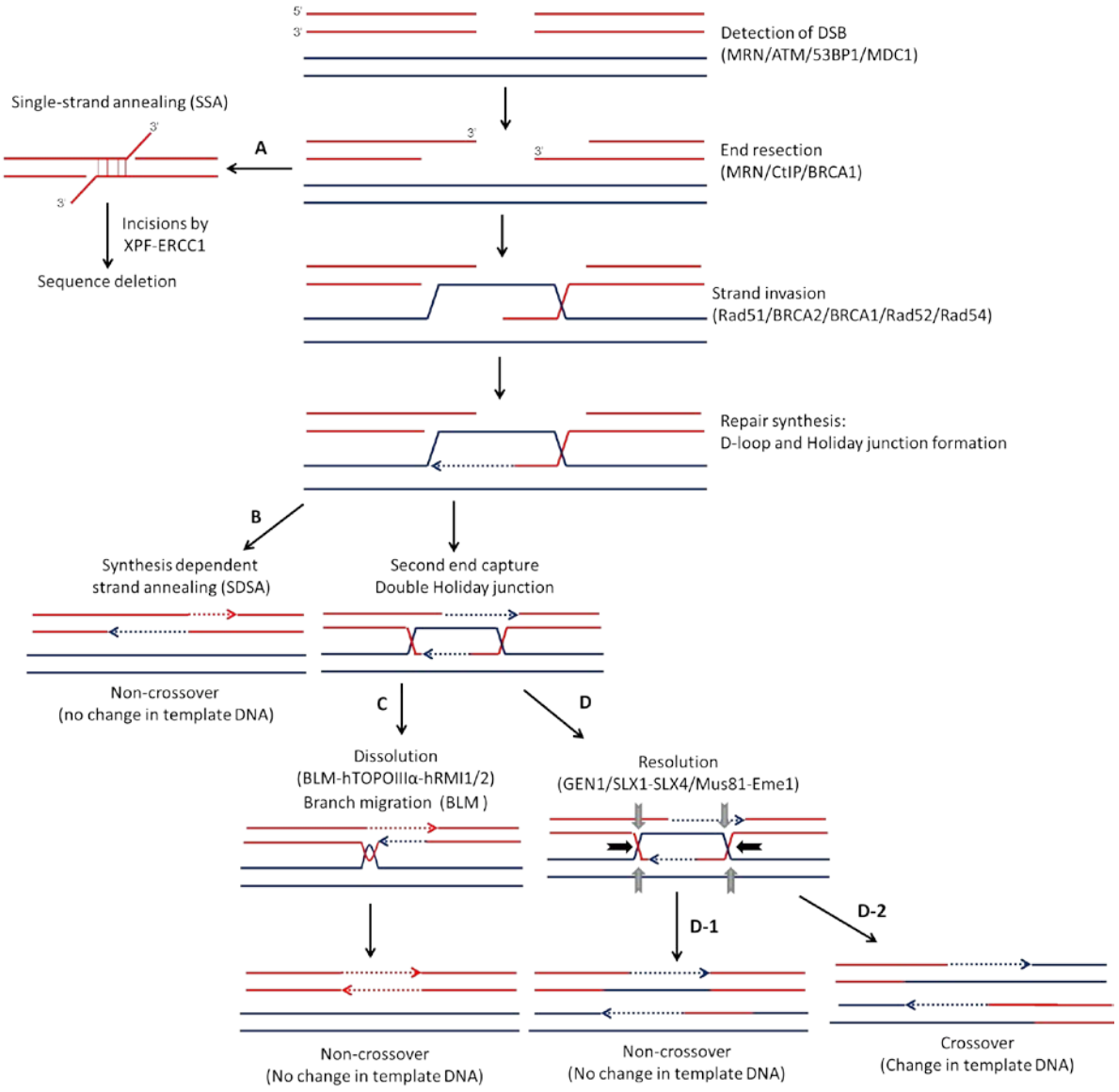


Figure 1. 7 Schematic representation of HR pathway in mammalian cells.

DSB ends are detected by the MRN complex, 53BP1 and MDC1. 5' to 3' end-resection of DSB termini is carried out by CtIP and the exonuclease activity of Mre11, generating 3' ssDNA overhangs which is rapidly coated by RPA. Rad51 displaces RPA by forming nucleoprotein filaments on the ssDNA. Rad51-dependent strand invasion into the homologous sister chromatid resulted a Holliday Junction (HJ) structure. The 3′ end of the invading strand serves as a primer for DNA synthesis and strand extension using the

sister chromatid as a template, forming a displacement loop (D-loop) structure. A) During single-strand annealing (SSA), two repeat sequences that are present on both sides of a DSB can anneal with each other forming 3'-flaps, which is cleaved by the 3'-flap endonuclease XPF-ERCC1, resulting in sequence deletions. B) During synthesis-dependent-strand-annealing (SDSA), the nascent strand is displaced from the D-loop and anneals with the complementary strand. Further synthesis occurs and repair is completed without changes to the template strand (non-crossover). Alternatively, the second DSB end is captured by the D-loop, forming Double Holliday Junctions (DHJs) structure that can be processed in different ways: C) The DHJ can undergo 'dissolution' where the DHJs are brought together by branch migration by the helicase activity of the BLM/hTOPOIIIα/hRMI1/2 resulting non-crossover products. D) The DHJs can undergo 'resolution' by several structure-selective endonuclease including GEN1, SLX1-SLX4 and MUS81-EME1. D-1) if the DHJs are incised in the same orientation (grey or black arrows, non-crossover products are generated. D-2) If the DHJs are incised in different orientations (grey and black arrows) the resulting recombinant molecule will have changes to the template DNA (crossover). Adapted from (Moynahan and Jasin 2010; Yata and Esashi 2009).

1.3 Formation of DNA interstrand cross-links (ICLs)

DNA interstrand cross-links (ICLs) are covalent adducts formed by bi-functional alkylating agents, molecules with two electrophilic groups, that reacted with nucleophilic centers of two DNA bases on complementary strands, the most reactive being the N7 atom on guanine and N3 atom on adenine (Clauson, et al. 2013). ICLs are physical blocks to DNA strand separation, therefore inhibits essential cellular processes such as replication and transcription. A single ICL in a bacterial genome and approximately twenty ICLs in a mammalian genome are lethal when the repair of ICLs in these cells is deficient (Lawley and Phillips 1996).

1.3.1 Endogenous source of ICLs

The major potential source of endogenous ICLs is reactive aldehydes, the by-products of natural cellular processes such histone demethylation (producing formaldehyde), alcohol metabolism (producing acetaldehyde) and lipid peroxidation (producing malondialdehyde and acrolein) (reviewed in (Clauson, et al. 2013)). **Figure 1. 8** represent the structure of acrolein and acrolein-induced ICL. Endogenous ICLs are believed to be one of the underlying causes of Fanconi Anemia (FA), a rare genetic disorder due to ICL repair deficiency that result in cancer predisposition, bone marrow failure and developmental defects. Recent studies underscore the crucial requirement of FA factors in suppressing the toxic effect of reactive aldehydes when aldehyde-catabolysing enzymes are impaired (Langevin, et al. 2011; Oberbeck, et al. 2014; Rosado, et al. 2011).

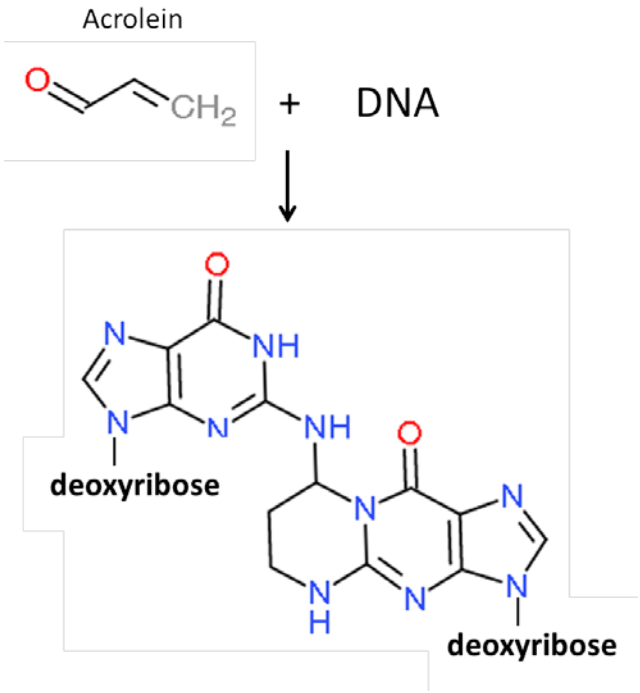


Figure 1. 8: Structure of acrolein and acrolein-induced ICL between guanines on complementary DNA strands.

Acrolein is a by-product of lipid peroxidation and a potential source of endogenous bifunctional alkylating chemical that forms ICLs when reacted with the DNA. Adapted from (Gates 2009).

1.3.2 ICL-inducing agents in anti-cancer therapy

Given the extreme cytotoxicity of ICLs, ICL-inducing agents were one of the earliest forms of anti-cancer chemotherapeutic agents and remain one of the most widely utilised to this day (reviewed in (Deans and West 2011)). Each clinically-active ICL-inducing agents such as nitrogen mustard (HN2), cisplatin and mitomycin C (MMC) differ according to sequence selectivity, location of ICL formation within the DNA, percentage of ICLs formed compared to other type of lesions and degree of DNA structural perturbation. These differences may results in a distinct cellular response to each ICL-inducing agent. It was previously reported that incisions of ICLs that significantly distort the DNA helix is greater than non-distorting ICLs (Smeaton, et al. 2008).

Most experimental bifunctional alkylating agents react with guanine residues in complementary DNA strands to from ICLs. Nitrogen mustard forms ICLs within the major groove of the DNA where as MMC produces ICLs within the minor groove (Noll, et al. 2006; Norman, et al. 1990). ICLs cause DNA helical distortion due to the difference in the length of the ICLs and the distance between the crosslinked nucleotides. ICLs formed by nitrogen mustard is 5.1 angstrom (Å) in length, which is shorter than the minimal distance of 6.8 Å between the guanines that form the ICLs in B-form DNA, causing a structural perturbation of the DNA (Rajski and Williams 1998). A static bend of approximately 14^{o} in the DNA was reported for ICLs produced by mechlorethamine, a derivative of nitrogen mustard (Rink and Hopkins 1995), whereas ICLs produced by cisplatin were reported to cause 45^{o} bend and 80% unwinding of short synthetic DNA oligonucleotides (Malinge, et al. 1994; Sip, et al. 1992).

In addition to ICLs, other types of DNA lesions such as monoadducts and intrastrand crosslinks are also produced by ICL-inducing agents: MMC, cisplatin and nitrogen mustard produces 5-10% ICLs. However it has been determined that ICLs are the most lethal lesions (Gargiulo, et al. 1995) and the relative toxicity of each ICL-inducing agent corresponds with its ability to form ICLs (O'Connor and Kohn 1990; Palom, et al. 2002). **Table 1. 1** provides important information on ICL-inducing agents utilised in anti-cancer therapy.

1.4 ICL repair in mammalian cells

In mammalian cells, ICLs are repaired predominantly in S-phase, when cells are undergoing replication. Replication-dependent ICL repair was first suggested when regardless of the stage of the cell cycle ICLs induced by photo-activated psoralen are introduced in human fibroblasts, cells arrest in late S/G2 phase (Akkari, et al. 2000). This implies that ICLs are exclusively recognised during S-phase, and trigger cell cycle arrest in S-phase to allow time for repair. Furthermore, DNA double-strand breaks (DSBs), an intermediate of ICL repair, are detectable only after passage through S-phase in dividing yeast or rodent cells, and not in stationary cells when the cells are treated with an ICL-inducing agent (De Silva, et al. 2000; McHugh, et al. 2000).

ICL repair outside of S-phase is suggested to occur mainly in non-dividing, terminally differentiated cells. While ICL repair in G1 has been extensively reported in *E.coli* (Berardini, et al. 1997) and yeast (McHugh and Sarkar 2006; McHugh, et al. 2000), replication-independent ICL repair in mammalian cells was first demonstrated recently, highlighting the requirement for NER and TLS factors (Muniandy, et al. 2009).

ICL-inducing agent	Target sequence	Example of agents	Cancer treated with agents
Nitrogen mustard	5'-GNC	Chlorambucil	Chronic lymphocytic leukemia (CLL)
		Cyclophosphamide	Lymphoma
		Melphalan	Multiple myeloma, melanoma, ovarian
		Ifosfamide	Non-small cell lung carcinoma (NSCLC)
Platinum compounds	5'-GC	Cisplatin	Testicular, ovarian, NSCLC
		Carboplatin	Ovarian
		Oxaliplatin	Colorectal
		Satraplatin	Prostate, breast
Mitomycin C	5'-CG	-	Esophageal, bladder
Psoralen	5'-TA	Furocoumarins from plants and fungi	Cutaneous T-cell lymphoma

Table 1. 1: Common ICL-inducing agents utilised in the clinic for anti-cancer therapy. Adapted from (Clauson, et al. 2013).

1.5 Replication-dependent ICL repair pathway

Physical constraints imposed by ICLs onto the DNA demands the coordination of multiple repair pathways for the successful removal of ICLs. In replication-dependent ICL repair, four major classes of genes have been identified using genetic analysis as key players in ICL repair: 1) Fanconi Anemia (FA) factors which are defective in FA patients; 2) structure-selective endonucleases which recognise and incise specific DNA structures; 3) translesion synthesis (TLS) polymerases and 4) DNA recombinases required during homologous recombination. Some of these pathways have already been described in **section 1.2.2** of this chapter in the context of repair of other types of lesions. This section describes the landmark stages of replication-dependent ICL repair which highlights how distinct repair pathways are coordinated for the successful removal of ICLs.

1.5.1 Fanconi Anemia (FA) pathway in ICL repair

Replication-dependent ICL repair is governed by the Fanconi Anemia (FA) pathway, which constitutes at least 17 FA factors (Wang and Smogorzewska 2015). A protein is assigned as an FA complementation group (FANCA to FANCS) when mutations in the protein result in the clinical features of FA, a rare autosomal recessive disorder characterised by bone marrow failure, cancer predisposition and developmental defects. The FA factors and other FA-associated proteins (factors that have not been shown to be mutated in FA patients but interact with the known FA factors) work in concert at various stages of ICL repair. **Figure 1. 9** illustrates the interaction of FA factors and FA-associated proteins during replication-dependent ICL repair.

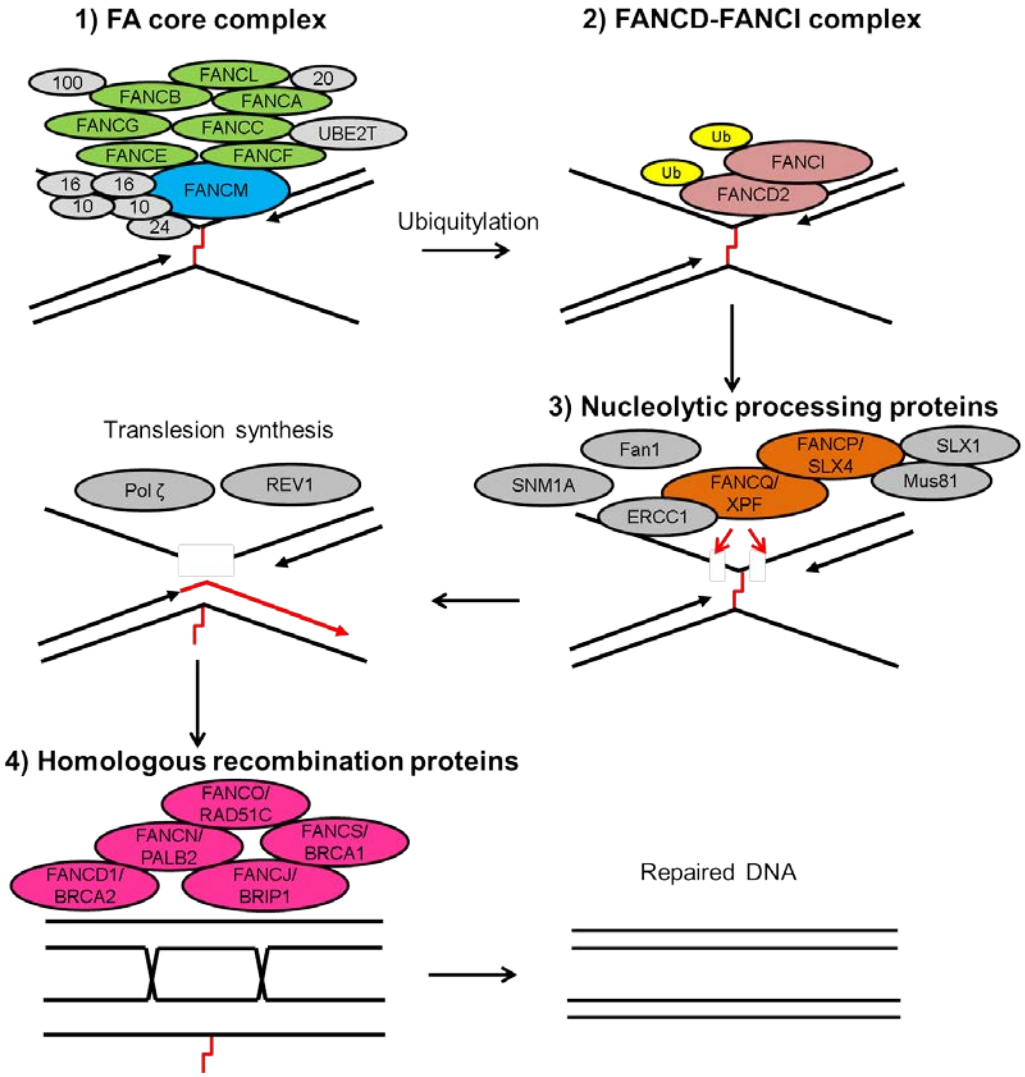


Figure 1. 9 A schematic representation of the interaction of Fanconi Anemia (FA) factors and FA-associated factors during replication-dependent ICL repair.

In the converging replication forks model, collision of the replication forks with an ICL triggers the recruitment FANCM-FAAP24 heterodimer, FAAP16/MHF1 and FAAP10/MHF2. Subsequently, the FA core complex (FANCA, FANCB, FANCC, FANCE, FANCF, FANCG, FANCL, FAAP100 and FAAP20) is assembled on the chromatin. The ubiquitin E3 ligase FANCL facilitated by UBE2T monoubiquitylates FANCD2-FANC1 complex (FANCD2-I), signalling the activation of ICL repair. Structure-selective nuclease(s), possibly XPF/FANCQ-ERCC1, SLX1, Mus81 bound to

the scaffold protein FANCP/SLX4; SNM1A and Fan1 are recruited to incise the ICL from the DNA termed ‘ICL unhooking’. TLS polymerases Rev1 and Polζ are recruited to extend the nascent leading strand past the ICL. The broken DNA strand is repaired by homologous recombination factors including FANCJ/BRIP1, FANCS/BRCA1, FANCN/PALB2, FANCD1/BRCA2, AND FANCO/RAD51 to enable replication fork restart. FA factors are depicted in colors; FA-associated factors are in grey. Adapted from (Wang and Smogorzewska 2015).

1.5.2 ICL recognition and repair activation

ICLs are recognised when the replication forks collide with the ICLs during S-phase, halting replication forks progression. The Fanconi Anemia (FA) factors are the central component in ICL recognition and repair activation. FANCM and FA-associated factor FAAP24 (Fanconi anemia associated protein 24 kDa) forms a heterodimer and are the first to be recruited to ICL sites. FAAP24 is reported to stabilise FANCM positioning onto the chromatin (Ciccia, et al. 2007). Histone-like proteins FAAP16/MHF1 and FAAP10/MHF2 are also recruited to stimulate FANCM activity (Yan, et al. 2010). FANCM possesses a functional helicase domain and FANCM-dependent translocation resulted in the recruitment of the single-stranded DNA (ssDNA)-binding replication protein A (RPA) to the ssDNA region within the vicinity of the ICL (Huang, et al. 2010; Vare, et al. 2012). Binding of RPA to ssDNA activates ATR signalling cascade which promotes phosphorylation of multiple targets including FANCD2-I complex and the MRN complex and activation of DNA damage response (DDR), as described in **section 1.2.1.2** of this chapter.

Subsequently, the FA core complex which comprises of FANCA, FANCB, FANCC, FANCE, FANCF, FANCG, FANCL, FAAP100 and FAAP20 is recruited to the site of ICL. The assembly of the FA core complex is required for monoubiquitylation of FANCD2-FANCI (FANCD2-I) complex on lysine-561 and lysine-523, respectively, by the ubiquitin E3 ligase FANCL facilitated by the ubiquitin-conjugating E2 enzyme UBE2T (Cole, et al. 2010; Garner and Smogorzewska 2011). FANCD2-I monoubiquitylation signify the activation of the pathway because failure in FANCD2-I

monoubiquitylation results in defect in subsequent events in ICL repair (Knipscheer, et al. 2009).

It is suggested that FANCD2-I monoubiquitylation is required to recruit downstream repair factors via their ubiquitin-binding zinc finger (UBZ) domains. Fan1 (Fanconi-associated nuclease 1) which contain a UBZ domain was recently shown to physically interact with monoubiquitylated FANCD2 (Kratz, et al. 2010; MacKay, et al. 2010; Smogorzewska, et al. 2010). SLX4, a scaffold protein is also suggested to be recruited to monoubiquitylated FANCD2 via its UBZ domain, bringing with it the structure-selective endonucleases implicated in ICL repair that are bound to it which are SLX1, XPF-ERCC1 and Mus81-EME1 (Yamamoto, et al. 2011).

1.5.3 ICL incisions by structure-specific nucleases

The second step in ICL repair is dual incisions flanking the ICL termed 'ICL unhooking' which result in double-strand breaks (DSBs) formation. Several structure-selective nucleases implicated in ICL unhooking are XPF-ERCC1, MUS81-EME1, and SLX1 which bind to the scaffold protein SLX4; FAN1; SNM1A and SNM1B. Deficiency in any of the nucleases leads to cellular hypersensitivity to ICL-inducing agents (discussed in (Hodskinson, et al. 2014)) but the exact endo/exonuclease(s) required for ICL incisions remain to be solved. Additionally, the DNA structure requirement that triggers ICL unhooking and whether the leading strand template or the lagging strand template is chosen for incisions also remain conjectural, and will be discussed in **section 1.7** of this chapter.

The 3'-flap endonuclease Mus81-Eme1 has been implicated in the first incision of ICL because its depletion suppresses DSB formation, an intermediate of ICL repair (Hanada, et al. 2007; Hanada, et al. 2006). However, our lab has demonstrated that cell cycle progression of Mus81-depleted cells was not impaired following ICL-induction, which fails to support the notion that Mus81 is an essential prerequisite for the initial incisions and subsequent downstream processing (Wang, et al. 2011). Additionally, Mus81-Eme1 was reported to be involved in the resolution of homologous recombination (HR) intermediates in vivo, and is able to cleave synthetic HR interemediates; placing its role downstream of ICL unhooking (Chen, et al. 2001; Sarbajna, et al. 2014).

The 5'-flap endonuclease Fan1 was shown to physically interact with monoubiqutylated FANCD2 suggesting that it could be recruited for ICL incisions (Kratz, et al. 2010; MacKay, et al. 2010; Smogorzewska, et al. 2010). Additionally, recombinant Fan1 is able to incise ICL placed in a synthetic DNA substrate through flanking incisions, suggesting its role in ICL unhooking (Wang, et al. 2014). However, deficiency in Fan1 does not result in FA phenotypes in humans and it is not epistatic with other FA factors in DT40 chicken cells (Trujillo, et al. 2012; Zhou, et al. 2012). Instead, Fan1 deficiency result in karyomegalic interstitial nephritis (KIN) (Zhou, et al. 2012)

Recent study that measures replication of plasmids containing a single site-specific ICL in *Xenopus leavis* egg extract have demonstrated that ICL repair is abolished when XPF is immunodepleted, while immunodepletion of Mus81 and FAN1 did not significantly affect ICL repair (Klein Douwel, et al. 2014). Additionally, recombinant XPF-ERCC1 alone was reported to unhook an ICL located in a synthetic DNA substrate (Kuraoka, et al. 2000), or when stimulated by the presence of SLX4 (Hodskinson, et al. 2014).

Furthermore, XPF-ERCC1-depleted cells accumulate DSBs, an intermediate of ICL repair suggesting that the first incision can efficiently occur in the absence of XPF-ERCC1 but XPF-ERCC1 is required to resolve the DSBs possibly by creating the second incision to complete ICL unhooking. (Niedernhofer, et al. 2004; Wang, et al. 2011) Collectively, these findings underscore the role of XPF-ERCC1 in ICL unhooking, possibly by creating one incision or both incisions flanking the ICLs.

Previous work in our lab has shown that SNM1A-depleted human cells are hypersensitive to ICL-inducing agents and accumulate DSBs, similar to XPF-ERCC1-depleted cells. Purified SNM1A is able to load onto a nick by XPF-ERCC1 and digest past an ICL placed in a synthetic duplex DNA substrate (Wang, et al. 2011). Therefore, there are two distinct scenarios that could explain how SNM1A collaborates with XPF-ERCC1 during ICL repair: 1) SNM1A may digest the DNA strand after the initial incision by an endonuclease, possibly XPF-ERCC1, to complete ICL unhooking. 2) after ICL unhooking by endonuclease(s), SNM1A digests the oligonucleotides tethered to the ICL to enable translesion synthesis and subsequent downstream ICL repair steps.

1.5.4 Translesion synthesis (TLS) past the ICL

After ICL unhooking, nascent DNA strand extension past the ICL requires a TLS polymerase because replicative polymerases are incapable of extending past the ICL. Pol ζ (consisting of REV3 and REV7 subunits) and REV1 are key TLS polymerases implicated in ICL repair because cells defective in either one of these factors are extremely sensitive to ICL-inducing agents (Gan, et al. 2008; Nojima, et al. 2005). REV1 possesses a deoxycytidyl transferase activity which may be involved in inserting a nucleotide opposite the lesion (Minko, et al. 2008). Immunodepletion of Rev7 resulted

in defect in repair of site specific ICL contained in plasmids in replication assays in *Xenopus laevis* egg extract (Raschle, et al. 2008). Other TLS polymerases including Pol η, Pol κ and Pol ν are suggested to play a minor or redundant role during TLS in ICL repair.

1.5.5 Homologous recombination (HR) completes ICL repair

ICL repair is suggested to be completed via homologous recombination (HR) because cells depleted of key factors in HR namely Rad51, BRCA1 and BRCA2 are hypersensitive to ICL-inducing agents (Bhattacharyya, et al. 2000; Ito, et al. 2005). Rad51 which facilitate strand invasion into the template DNA of the sister chromatid was also reported to be recruited to ICL site prior to ICL incisions, independent of ICL repair activation by FANCD2-FANCI (Long, et al. 2011). This suggests that the FA pathway and HR are coordinated together for ICL repair. Additionally, mutations in five HR factors results in FA, and these factors have been assigned as FA complementation groups highlighting the importance of HR in ICL repair: FANCJ/BRIP1, FANCS/BRCA1, FANCN/PALB2, FANCD1/BRCA2, AND FANCO/RAD51. Upon completion of ICL repair, dissociation of the FA core complex enables the recruitment of the deubiquitylating enzyme USP1 (ubiquitin specific peptidase 1), which is responsible for removing the monoubiquitin from FANCD2 and FANCI to terminate ICL repair (Kim, et al. 2009).

1.6 Models for replication-dependent ICL repair

Several models for replication-dependent ICL repair have been proposed which takes into account 1) the DNA structure requirement that trigger ICL unhooking and 2) whether the leading strand template or the lagging strand template is incised. This section describes the major models for replication-dependent ICL repair. **Figure 1. 10** provide a schematic representation of the possible mechanism of replication-dependent ICL repair.

1.6.1 Dual forks convergence onto an ICL model

Recently, elegent replication assays of a single site-specific ICL located in 6 kb plasmids in *Xenopus laevis* egg extract has provided molecular detail of repair intermediates during ICL repair (Klein Douwel, et al. 2014; Raschle, et al. 2008; Zhang, et al. 2015). In this system, the convergence of dual replication forks, which creates an X-shaped structure, is an absolute trigger for ICL unhooking. When the replication forks encounter an ICL, both nascent leading strands initially stalls -20 nucleotides (nt) from the ICL, and one nascent leading strand gradually progresses to 1 nt from the ICL. The arrival of the leading strand stimulates dual incisions flanking the ICL on the opposite template strand, which unhooks the ICL from the DNA duplex. This enables translesion synthesis to occur and subsequent repair using homologous recombination which results in replication fork restart.

1.6.2 Single fork collision with an ICL model

The earliest model for replication-coupled ICL repair is the trigger of ICL repair by a single replication fork collision with an ICL (Niedernhofer, et al. 2005; Wang 2007). Even though this ‘classic’ model lack molecular detail, there is strong reason to

postulate that a single replication fork is sufficient to trigger ICL repair because replication origins are spaced 100 kilobase (kb) apart and forks move at a rate of 1.5 kb per minute (Duderstadt, et al. 2014). A single fork is more likely to strike an ICL that formed at random locations within the DNA than fork convergence. In this model, it is not clear whether the leading strand or the lagging strand template is chosen for incisions. However, it is possible either one of the template strand is chosen for incisions. Attempt to recapitulate single fork trigger of ICL repair in *Xenopus laevis* replication assay by delaying the arrival of the second fork to the ICL demonstrates that a single replication fork fails to trigger ICL repair. The leading strand fails to approach the ICL, and downstream repair event is not detected. The study demonstrates that delay in the arrival of the second fork, as would occur in vivo is still efficient to trigger ICL repair (Zhang and Walter 2014). However, an updated model for single replication fork collision with an ICL has been devised based on the dual-replication fork. In the updated model, incision occurs on the lagging strand template which translesion synthesis to occur using the existing replication apparatus.

Our lab has recently proposed several context-dependent replication-coupled ICL repair models: ICL detection by a single replication fork occurs in early S-phase or by two converging replication forks most likely occurring during late S-phase (Wang, et al. 2011). In addition, the repair pathways will be influenced by whether the leading-strand template or the lagging-strand template is targeted for the initial incisions (Wang, et al. 2011).

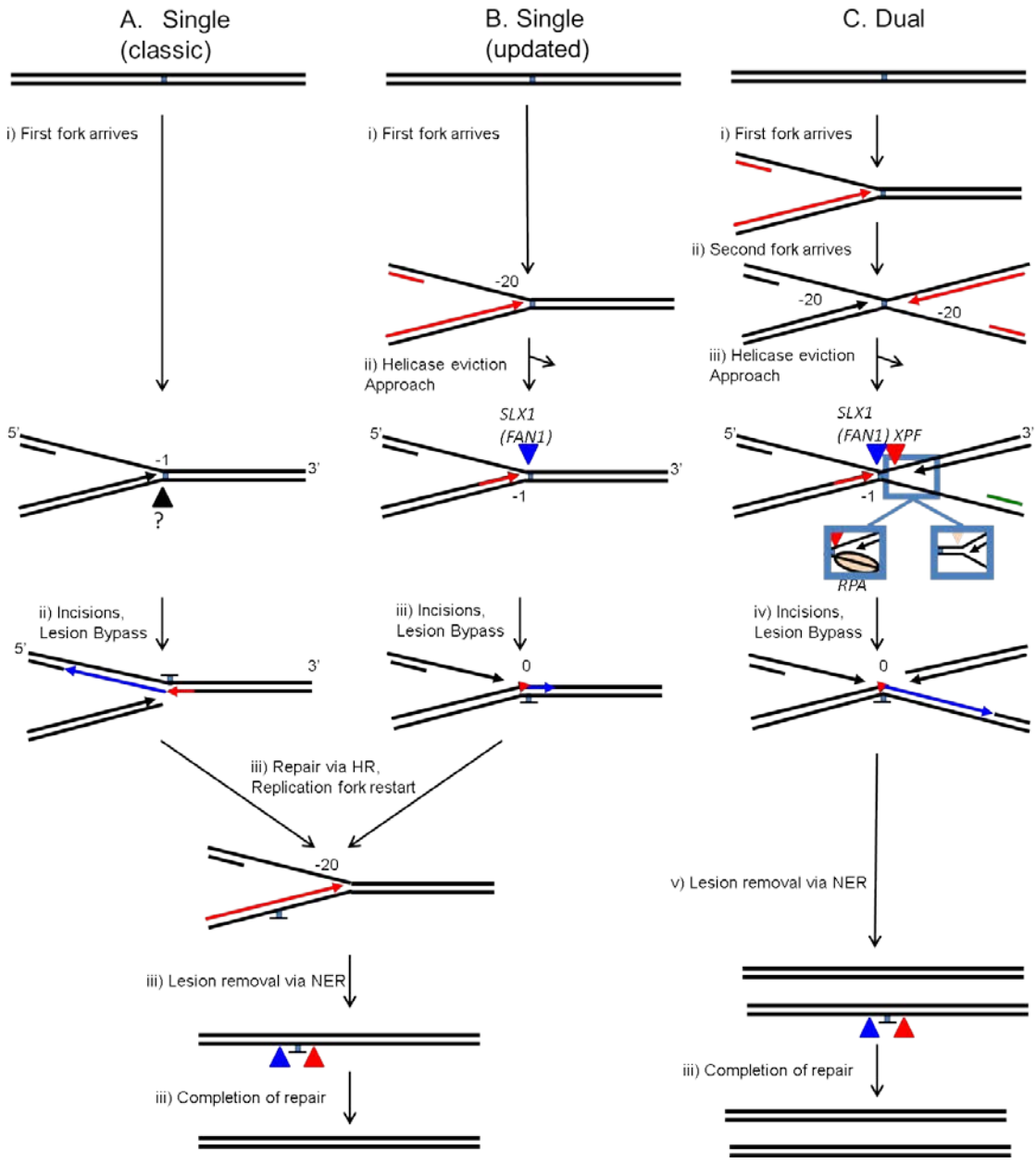


Figure 1. 10 Possible mechanisms for replication-coupled ICL repair.

Three possible mechanisms are depicted which are A) 'classic' single replication fork collision with an ICL which depicts the possibility of incision on the leading strand template; B) 'updated' single replication fork collision with an ICL which depicts incision on the lagging strand template; and C) dual replication collision onto an ICL. Coloured arrows represent ICL incisions by endonucleases implicated in ICL incisions. Adapted from (Zhang and Walter 2014).

1.7 Replication-independent ICL repair pathway

Replication-independent ICL repair was first observed from successful repair of site-specific ICL located in reporter plasmid that do not have replication origins necessary for replication-dependent repair and pointed toward requirement for TLS polymerases and NER factors for successful repair (Enoiu, et al. 2012; Wang, et al. 2001; Zheng, et al. 2003). Based on this evidence, it is suggested that replication-independent repair involves two rounds of ICL unhooking via NER, however several *in vitro* studies reported that NER proteins did not incise cisplatin ICL (Zamble, et al. 1996) or create dual incision both 5' to psoralen or alkyl ICL (Smeaton, et al. 2008). **Figure 1.11** represents the current model for replication-independent ICL repair pathway.

The global genome NER (GG-NER) sub-pathway mediates repair throughout the genome (Gillet and Scharer 2006) while the transcription-coupled NER (TC-NER) is specific for repair on actively transcribed strand of DNA that blocks the progression of RNA polymerase II (RNAPII) (Hanawalt and Spivak 2008). The sub-pathways differ at the damage recognition stage, but share the same downstream processes of repair. **Figure 1.11** provides a schematic model of the NER sub-pathways in mammalian cells. In GG-NER, damage recognition is initiated by the recruitment of XPC and hHRad23B (human homolog of Rad23B) (Sugasawa, et al. 2001). Once XPF-ERCC1 is recruited to the lesion, it incises 5' to the lesion on the damaged strand (Fagbemi, et al. 2011). Repair synthesis is initiated by DNA polymerase δ and DNA pol κ or DNA pol ε and associated factors PCNA and RFC from the 3'-OH generated by incision by XPF-ERCC1 (Lehmann 2011; Staresincic, et al. 2009). The nick is sealed by DNA ligase

IIIα/XRCC1 or DNA ligase I for completion of repair (Moser, et al. 2007). The role of XPF-ERCC1 in NER is further described in **section 1.11.1** of this chapter.

In transcription-coupled NER (TC-NER), stalling of RNAPII at the DNA lesion within the gene being transcribed serve as a signal for the recruitment of CSB (ERCC6), CSA (ERCC8) and XAB2 (Hanawalt and Spivak 2008). Chromatin remodelling activity of CSB occurs upstream of the stalled RNAPII to enable RNAPII to back tract from the lesion, allowing downstream NER factors to be recruited to the damaged site. The CSA-containing ubiquitin ligase complex is recruited to the damaged site via a CSB-dependent manner and is implicated in ubiquitination and degradation of CSB, which terminate TC-NER and restore transcription (Groisman, et al. 2006).

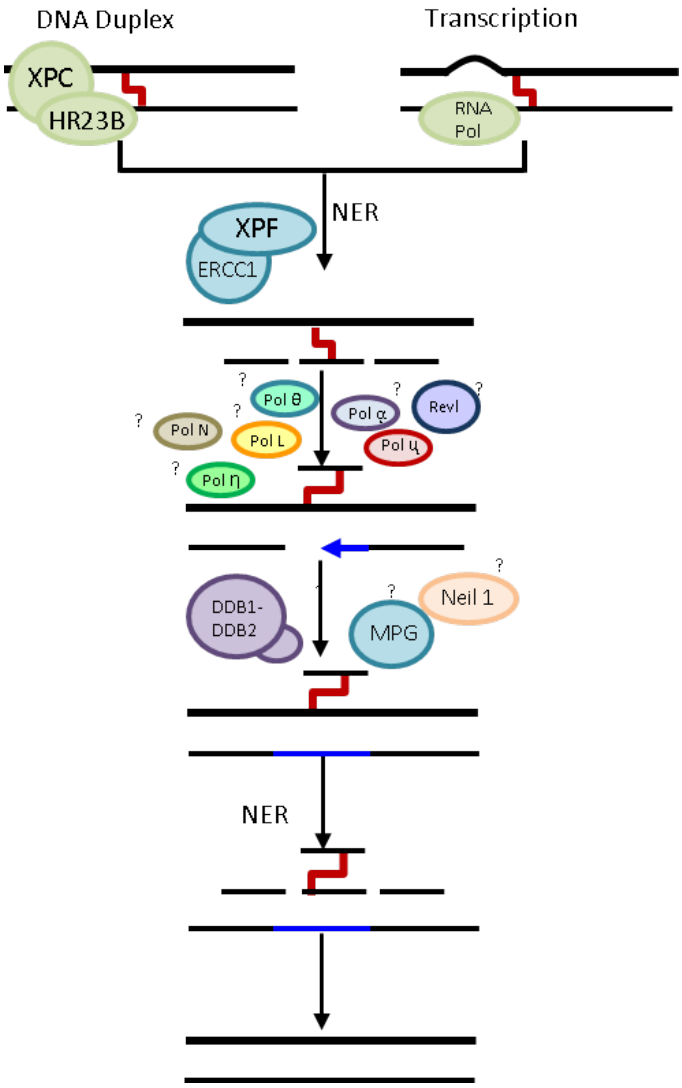


Figure 1. 11 Replication-independent ICL repair pathways

Replication-independent ICL repair pathways occur via A) global genome NER (GG-NER) and transcription-coupled NER (TC-NER). Adapted from (Muniandy, et al. 2010).

1.8 XPF/Mus81 family of DNA repair proteins

XPF/Mus81 family of DNA repair proteins are dimeric structure-selective endonucleases which cleave branched DNA structures that arises during a variety of DNA processes including DNA replication, meiotic recombination and nucleotide excision repair (NER) (reviewed in (Ciccia, et al. 2008)). Its family members can be divided into catalytic and non-catalytic members. Most eukaryotes possess four family members that assemble into two distinct heterodimeric endonucleases: XPF-ERCC1 (Sijbers, et al. 1996) and Mus81-EME1 (Ciccia, et al. 2003); consisting of a catalytic (mentioned first) and a non-catalytic member (mentioned last). Additional family members which exist exclusively in vertebrates are the heterodimer FANCM-FAAP24; both subunits are catalytically inactive but FANCM possesses a functional helicase domain required in DNA translocation, and FAAP24 is reported to target FANCM to DNA (Ciccia, et al. 2007). Humans possess an additional heterodimeric endonuclease Mus81-EME2 (Pepe and West 2014). Archael homologs of the XPF/Mus81 family function as homodimers hence contain a functional nuclease domain in both subunits (Nishino, et al. 2003; Roberts and White 2005). Figure 1.12 provides a schematic representation of the domain arrangement of the XPF/Mus81 family of proteins.

1.9 Domain features of XPF/Mus81 family members

Catalytically active members of the XPF/MUS81 family share a characteristic core consisting of an excision repair cross complementation group 4 (ERCC4) endonuclease domain and a tandem helix-hairpin-helix $(HhH)_2$ domain (Shao and Grishin 2000). The ERCC4 endonuclease domain contains two elements: a helix-β-strand-helix dimerization motif and a subdomain distantly related to the nuclease domain of the type

II restriction endonucleases where the acidic residues within the motif coordinate a divalent metal ion that is required for catalysis (Nishino, et al. 2003; Pingoud, et al. 2005). The $(HhH)_2$ domain is the major determinant for DNA structure selectivity and binding (Komori, et al. 2002; Nishino, et al. 2003; Nishino, et al. 2005). The $(HhH)_2$ domain is also responsible for the dimerization of the XPF/Mus81 family members as a stable heterodimeric complex. Some XPF/Mus81 family members possess a helicase domain, which is functional in FANCM and *P.furiosus* Hef but degenerate in XPF and its orthologs because they lack key residues required for ATP binding and hydrolysis (Sgouros, et al. 1999). The ERCC4-$(HhH)_2$ core is the smallest functional unit required for endonuclease activity and DNA targeting (Nishino, et al. 2005; Tsodikov, et al. 2005). Inactive members retain the same core but their ERCC4 endonuclease domain is catalytically inactive (therefore denoted as the 'central' domain).

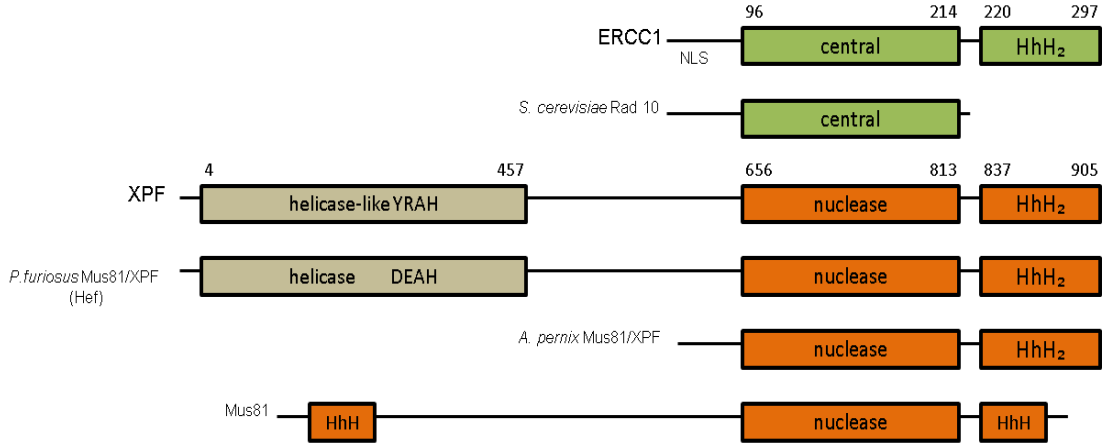


Figure 1.12: A schematic representation of the domain arrangement of the XPF/Mus81 family of proteins.

1.10 Structure of the human XPF-ERCC1

XPF and ERCC1 dimerises through the $(HhH)_2$ domains located at the C-terminal region of both proteins, and their interaction stabilises both proteins *in vivo* (de Laat, et al. 1998c). XPF consists of three domains: an N-terminal helicase-like domain (residue 4-457) is homologous to superfamily II helicases but lack several key residues for DNA unwinding activity. The domain has been suggested to contribute to DNA binding activity, although its function has not been examined in detail. The central nuclease domain (residue 656-813) harbours conserved metal-binding residues as well as functionally important basic residues that presumably interact with DNA substrates (Nishino, et al. 2003). The C-terminal HhH2 domain (residue 837-905) specifically dimerises with ERCC1 and also contributes to DNA binding (Tripsianes, et al. 2005). ERCC1 consists of two domains: A conserved central domain (residue 96-214) predicted to be structurally similar to the nuclease domain of XPF but devoid of the acidic residues characteristic of a nuclease domain. The C-terminal $(HhH)_2$ domain of ERCC1 (residue 220-297) dimerises with XPF and also contribute to DNA binding.

Equilibrium binding assays probing for the binding affinity of each domain of ERCC1 revealed that both the central and the $(HhH)_2$ domains bind ssDNA, with the central domain having 8-fold higher affinity to 5'-ssDNA compared to 3'-ssDNA indicating that the central domain binds ssDNA with defined polarity (Tsodikov, et al. 2005). $(HhH)_2$ motif is a widely distributed DNA-binding motif that binds non-specifically to dsDNA and ssDNA, but with 6-fold higher affinity to ssDNA. Binding assays of the XPF-ERCC1 $(HhH)_2$ complex indicates that the each HhH2 heterodimer binds two

ssDNA oligomers indicating that each HhH2 from XPF and ERCC1 interact with a ssDNA each (Newman, et al. 2005).

1.11 Functions of XPF-ERCC1 in DNA repair

1.11.1 XPF-ERCC1 in NER

Human XPF complemented the repair defect of human XPF-deficient cell extracts and Chinese hamster ovary (CHO) cells groups 4 and 11, confirming that XPF is an NER factor (Sijbers, et al. 1996). XPF-ERCC1 is one of the two structure-selective endonucleases required for the incisions of bulky DNA lesions such as UV photoproducts during NER. The NER pathway is discussed in detail in **section 1.2.2.4** of this chapter. During NER, XPF-ERCC1 incises 5' to the lesion on the damaged strand (Fagbemi, et al. 2011), which generate a 3'-OH that enable repair synthesis by DNA polymerase δ and DNA pol κ or DNA pol ε and associated factors PCNA and RFC (Lehmann 2011; Staresincic, et al. 2009). This is followed by incision 3' to the lesion by XPG which releases the lesion-containing oligonucleotide (Kemp, et al. 2012). Additionally, *in vitro* studies showed that recombinant XPF-ERCC1 is showed to incise a stem-loop substrate, a DNA structure that resembles an open 'bubble' intermediate formed by the lesion, within the duplex DNA 5' to the lesion near the ssDNA/dsDNA junction (Sijbers, et al. 1996).

1.11.2 XPF-ERCC1 DSB repair

Orthologs of XPF-ERCC1 in lower eukaryotes such as Arabidopsis thaliana, Drosophilla and yeast play a vital role in single-strand annealing (SSA) sub-pathway during DSB repair (Fishman-Lobell and Haber 1992; Hefner, et al. 2003). Additional discussion on DSB repair is provided in **section 1.2.2.5** of this chapter. It was first observed in yeast that Rad1-Rad10, homolog XPF-ERCC1 in yeast, was able to remove

long non-homologous 3' overhangs from invading homologous tails during DSB repair (Bardwell, et al. 1994; Fishman-Lobell and Haber 1992), and mutations of Rad1 and Rad10 suppresses HR between sequence repeats (Prado and Aguilera 1995). Human XPF-ERCC1 was also reported to participate in SSA sub-pathway and gene conversion (Adair, et al. 2000; Ahmad, et al. 2008; Al-Minawi, et al. 2008).

1.11.3 XPF-ERCC1 in ICL repair

It was suggested that XPF-ERCC1 may have a role outside of NER because Chinese hamster ovary (CHO) cells defective in XPF or ERCC1 display much higher sensitivity to ICL-inducing agents than cells deficient in other NER factors (De Silva, et al. 2000; Hoy, et al. 1985). XPF is recently characterised as a Fanconi Anemia complementation group denoted as FANCQ, which is a genetic disorder due to deficiency in ICL repair confirming that XPF-ERCC1 is a key player in ICL repair. In the absence of XPF-ERCC1, cells accumulate ICL-induced replication-dependent DSBs (McCabe, et al. 2008; Niedernhofer, et al. 2004; Vare, et al. 2012), suggesting incomplete repair of ICLs without XPF-ERCC1. The role in XPF-ERCC1 is described in detail in **section 1.6.3** of this chapter and also in the introduction section of chapter 3.

1.12 XPF and ERCC1 in DNA repair disorders

Mutations in XPF have been reported in patients with clinical features of several rare autosomal recessive DNA repair disorders: Xeroderma pigmentosum (XP), Cockayne syndrome (CS), Fanconi anemia (FA) and XFE progeroid syndrome; and celebro-oculo-facio-skeletal (OFS) syndrome and XP for mutations in ERCC1 (reviewed in (Gregg, et al. 2011)). It is suggested that the clinical manifestations of each disorder is determined by how the mutations affect protein expression, function and subcellular localization of XPF or ERCC1 (Ahmad, et al. 2010).

Xeroderma pigmentosum (XP) is a disorder caused by deficiency in NER characterized by severe photosensitivity; 1000-fold higher risk of early onset of skin cancer, 10-fold higher risk of other tumors and, in severe cases, neurodegeneration (Kraemer, et al. 1994). Most XPF and ERCC1 patients exhibit mild symptoms of XP which include sun sensitivity, freckling of the skin and basal or squamous cell carcinomas typically during early adulthood. Only a subset of XPF patients also suffers from neurological degeneration, with one patient exhibiting dramatic accelerated ageing (Niedernhofer, et al. 2006).

Fanconi Anemia (FA) is a rare disorder that occurs in 1 in every 350,000 live births with median survival of 20 years (Auerbach 2009). FA is caused by deficiency in any of the 17 FA factors known to date, characterised by cancer predisposition, developmental defects and bone marrow failure. It is inherited in autosomal recessive manner except for FANCB which is X-linked. XPF is recently assigned an FA complementation group FANCQ in light of identification of patients defective in XPF with clinical characterisation of FA (Bogliolo, et al. 2013; Kashiyama, et al. 2013). The causative mutation in a minority of patients with clinical manifestations of FA are yet to be identified, therefore these patients are not assigned to any complementation group (Kottemann and Smogorzewska 2013).

Cockayne syndrome (CS) is a disorder caused by deficiency in transcription-coupled NER characterised by neurodevelopment degeneration and premature ageing. Previously, all CS-affected individuals have been associated with mutations in CSA

(ERCC8) and CSB (ERCC6) CSB, factors involved in TC-NER but recently three patients with mutations in XPF exhibit clinical features of CS (Kashiyama, et al. 2013).

1.13 Objective of this study

The main objective of this study is to examine the role of XPF-ERCC1 in repairing damaged DNA replication forks. Using *in-vitro* biochemical reconstitution assays of model replication fork structures, I attempted to achieve the following aims:

1. **To fully characterise XPF-ERCC1 endonuclease activity on model native and damaged replication fork structures.**

   The most prominent current model for replication-coupled ICL repair proposes that ICL unhooking requires the convergence of two replication forks upon an ICL which creates an X-shaped structure surrounding the ICL; and the gradual extension of a nascent leading strand triggers ICL unhooking. (Klein Douwel, et al. 2014; Raschle, et al. 2008; Zhang, et al. 2015). The authors of the model also argues that a single replication fork fails to trigger ICL unhooking (Klein Douwel, et al. 2014) This is contrary to classical model of replication-coupled ICL repair which proposes that collision of a single replication fork with an ICL is a more likely event to occur in vivo (Niedernhofer, et al. 2005; Wang 2007). To address this discrepancy, I therefore used model native and damaged DNA substrates that represent both model - a single and dual replication forks collision with an ICL.

2. **To determine the contribution of the replicative single-stranded DNA binding protein RPA to XPF-ERCC1 activity on native and damaged replication fork structures.**

It was previously reported that during NER, RPA is recruited to the non-damaged complementary strand opposite a DNA lesion and facilitate the positioning of XPF-ERCC1 and XPG which enable the endonucleases to incise 5’ and 3’ of the lesion, respectively (de Laat, et al. 1998b). However, the contribution of RPA on XPF-ERCC1 activity during ICL repair is not entirely understood even though replication fork stalling by an ICL give rise to ssDNA intermediates surrounding the ICL which may be stabilised by RPA (Raschle, et al. 2008). I therefore asked the question whether RPA modulates XPF-ERCC1 activity by using model native and damaged replication forks that mimic a single or dual replication forks stalling onto an ICL.

3. **To determine the collective activities of XPF-ERCC1 and RPA with the 5’ to 3’ SNM1A exonuclease on native and damaged replication fork structures.**

Biochemical analysis of purified human SNM1A demonstrated that SNM1A loads onto and digests an ICL-containing dsDNA substrate from either blunt ends, or, of importance here, a single XPF-ERCC1-induced incision 5’ to the ICL (Wang, et al. 2011). I address whether SNM1A collaborate with XPF-ERCC1 during ICL repair using model native and damaged replication fork structures.

2 Materials and Method

2.1 Purification of human XPF-ERCC1 from insect cells

Methods to purify the XPF-ERCC1 protein complexes were modified from published protocol (Enzlin and Scharer 2002). Purification of human XPF-ERCC1 was conducted at Research Complex at Harwell with the help of Dr. Joanna McGouran (Department of Chemistry, University of Oxford) and Dr. Denis Ptchelkin (Research Complex at Harwell).

The pFastBac1 vectors with the cDNA encoding the full-length XPF and the full-length ERCC1-His, a kind gift from J.H. Enzlin and O. D. Scharer (University of Zurich, Sweden), were individually transformed into competent DH10Bac *E.coli* cells. Bacmid DNA were isolated and transfected into Sf-21 insect cells to amplify the baculovirus according to the manufacturer's instructions (BAC TO BAC system, Invitrogen). For protein production, 500 mL Hi-5 insect cell culture at the concentration of $2.5x10^7$ cells/mL were co-infected with XPF and ERCC1 viruses. Cells were harvested 65 hours post-infection and centrifuged at 400xg for 10 minutes at room temperature (RT). Cells were gently re-suspended in 200 mL 100 mM NaCl and centrifuged at 400xg for 10 minutes at RT. Cells were re-suspended in 100 mL lysis buffer (50 mM HEPES pH 8.0, 800 mM NaCl, 5 mM TCEP, 10% glycerol, 0.1% NP-40 and one dissolved EDTA-free protease inhibitor cocktail tablet (Roche)) and sonicated for 6 cycles (1 minute sonication, 2 minutes rest) with 9.5 mm probe using Soniprep 150 Plus (MSE). Cell lysate was clarified by centrifugation at 20,000 rpm for 1 hour at 4^oC. The supernatant were incubated with 2.5 mL Ni-NTA agarose beads (Qiagen) overnight on a rotating platform at 4^oC. An aliquot of cells prior to sonication, cell pellet and supernatant after

high-speed centrifugation were loaded in SDS-PAGE gel to determine protein expression level.

The beads were collected by centrifugation at 400xg for 5 minutes at 4^{o}C and re-suspended in 20 mL Ni-buffer (50mM HEPES pH 8.0, 10% glycerol, 5mM TCEP) containing 800 mM NaCl/4 mM imidazole and packed in a column. The column was washed twice via gravity flow with 50 mL Ni-buffer containing 500 mM NaCl/5 mM imidazole and 500 mM NaCl/20 mM imidazole, respectively. XPF-ERCC1 was eluted twice with 12 mL Ni-buffer containing 300 mM NaCl/50 mM imidazole and in 1 mL fractions with 12 mL Ni-buffer containing 300 mM NaCl/250 mM imidazole, respectively. An aliquot of the fractions were loaded onto SDS-PAGE gel to determine protein expression level. The fractions with highest level of XPF and ERCC1 were chosen for purification, and were loaded in a dialysis casette with 3.5K molecular weight cut-off (Thermo Scientific) and dialysed overnight in 2 L dialysis buffer (50 mM Tris pH 8.0, 10% glycerol, 500 mM NaCl, 5 mM DTT). The protein was further purified by gel-filtration using a HiLoad 16/60 Superdex 200 column (Pharmacia) equilibrated with gel filtration buffer (25 mM HEPES pH 8.0, 150 mM NaCl, 10% glycerol and 5 mM TCEP). Fractions containing the XPF–ERCC1 heterodimer, eluting from 13 to 15 mL, were pooled and concentrated. The protein concentration was measured directly at 280 nm and the purified protein was flash-frozen in aliquots and stored at -80^{o}C. Typically, 0.25 – 1 mg/ml of purified complex, with a concentration of 0.1 – 0.6 mg/ml, was obtained.

Full-length human RPA purified from *E.coli* was a kind gift from Fumiko Esashi (Dunn School of Pathology, University of Oxford). Human SNM1A truncated for its first 675

residues (aa 676-1040) was purified from insect cells by a member of our group, Sook Yee Lee. Full length *E.coli* SSB was purchased from Sigma Aldrich.

2.2 Site-directed mutagenesis

Site-directed mutagenesis to introduce point mutation at proline-85 to a serine (P85S) in pFastBac1-XPF(WT) was performed using PCR. PfuUltra High Fidelity DNA polymerase and primers harbouring P85S mutation (forward primer: 5'-GTTGAACACCTCTCTCGCCGTGTAAC-3'; reverse primer: 5'-GTTACACGGCGAGAGAGGTGTTCAAC-3') were included in the PCR reaction mixture. The PCR cycling condition is listed below:

Cycle step	**Temperature (^{0}C)**	**Duration**	**Number of cycles**
Initial denaturation	95	2 minutes	1
Denaturation Annealing Extension	95 61 72	30 seconds 30 seconds 8 minutes	30
Final extension	68 4	10 minutes Hold	1

Table 1: PCR cycling condition used for PCR-based site-directed mutagenesis.

The PCR product was digested using DpnI (NEB) for 1 hour at 37^{0}C, transformed into competent *E.coli* cells and selected under selection overnight by plating onto LB agar plate containing 100 µg/mL ampicillin. Single colonies were further selected overnight under selection. Plasmids from each colony were purified using Miniprep kit (Qiagen). The plasmids were sequenced for the entire length of the XPF cDNA to confirm the existence of the P85S point mutation and no other mutations were detected. Mutant XPF proteins were co-expressed with wild-type ERCC1 in insect cells and purified using the same procedure used for the wild type heterodimer.

2.3 Generation of radiolabelled DNA substrates

For 5'-end radiolabelling, 10 pmoles of DNA oligonucleotide was incubated with 1 unit T4 polynucleotide kinase (T4 PNK) (NEB) in the presence of 6.8 pmoles [γ-^{32}P]dATP for 1 hour at 37^{o}C. For 3'-end radiolabelled, 50 pmoles of DNA oligonucleotide was incubated 1 unit of terminal deoxytransferase (TdT) (NEB) in the presence of 3.3 pmoles [α-^{32}P]dATP for 1 hour at 37^{o}C. Then, the radiolabelling reaction mixture was loaded in a Bio-Gel P-6 spin column (Bio-Rad), centrifuged at 1,000xg for 4 minutes and diluted with water to the concentration of 500 nM. The radiolabelled oligonucleotide was annealed with unlabelled oligonucleotide(s) by boiling at 95^{o}C for 5 minutes and gradually cooled to below 30^{o}C in the presence of annealing buffer (10 mM Tris pH 7.5, 50 mM NaCl, 1 mM EDTA). The final concentration of radiolabelled oligonucleotide is 100 nM. The quality of every batch of radiolabelled DNA substrates generated is analysed using 10% non-denaturing polyacrylamide (PAGE) gel as described in section 3.2 of chapter 3.

Code	Sequence of DNA oligonucleotides
1	5′-ATAAATATTTTTTATTAATAATAGATCACCTTTCTTTCTCTTCTCCCCTT-3′
2	5′-AAGGGGAGAAGAGAAAGAAAGGTGATCTATTATTAATAAAAAATATTTAT-3′
3	5′-TGATCTATTATTAATAAAAAATATTTAT-3′
4	5′-AAGGGGAGAAGAGAAAGAAAGGAGATCT-3′
5	5′-TTCCCCTCCTCTCCTTCCTTCCTGATCTATTATTAATAAAAAATATTTAT-3′
6	5′-AAGGGGAGAAGAGAAAGAAAGG-3′
7	5′-AAGGGGAGAAGAGAAAGAAAG-3′
8	5′-AAGGGGAGAAGAGAAAGAAA-3′
9	5′-AAGGGGAGAAGAGAAAGAA-3′
10	5′-AAGGGGAGAAGAG-3′
11	5′-GGAAGGAAGGAGAGGAGGGGAA-3′

12	5′-CTGCCATGTCACTATGGTTAGCCGT-3′
13	5′-AGTAACGTGACAGCGTTGGCTAAGT-3′
14	5′-(UUCCCCUCCUCUCCUUCCUUCC)TGATCTATTATTAATAAAAAATATTTAT-3′
CL-1	5′-ATAAATATTTTTTATTAATAATAGATCACCTTTCTTTCTCTTCTCCCCTT-3′ 3'P-TATTTATAAAAAATAATTATTATCTAGTCCTTCCTTCCTCTCCTCCCCTT-5'Bio
CL-4	5′-ATAAATATTTTTTATTAATAATAGATCACCTTTCTTTCTCTTCTCCCCTT-3′ 3'P-TATTTATAAAAAATAATTATTATCTAGTCCTTCCTTCCTCTCCTCCCCTT-5'Bio
CL-7	5′-ATAAATATTTTTTATTAATAATAGATCACCTTTCTTTCTCTTCTCCCCTT-3′ 3'P-TATTTATAAAAAATAATTATTATCTAGTCCTTCCTTCCTCTCCTCCCCTT-5'Bio
X	5′-GGTACAGTGATACCAATCGGCATACGGCTAACCATAGTGACATGG-3′ 3'-TTGCACTGTCGCAACCGATTCATACTTAGCCAACGCTGTCACGTT-5′

Table 2: Sequence of DNA oligonucleotides used to generate the DNA substrates for this study. Red lines represent a triazole interstrand crosslink; non-complementary regions are underlined; sequence in brackets stands for RNA; ‘P’stands for 3’-phosphate; ‘Bio’ stands for 5’-biotin.

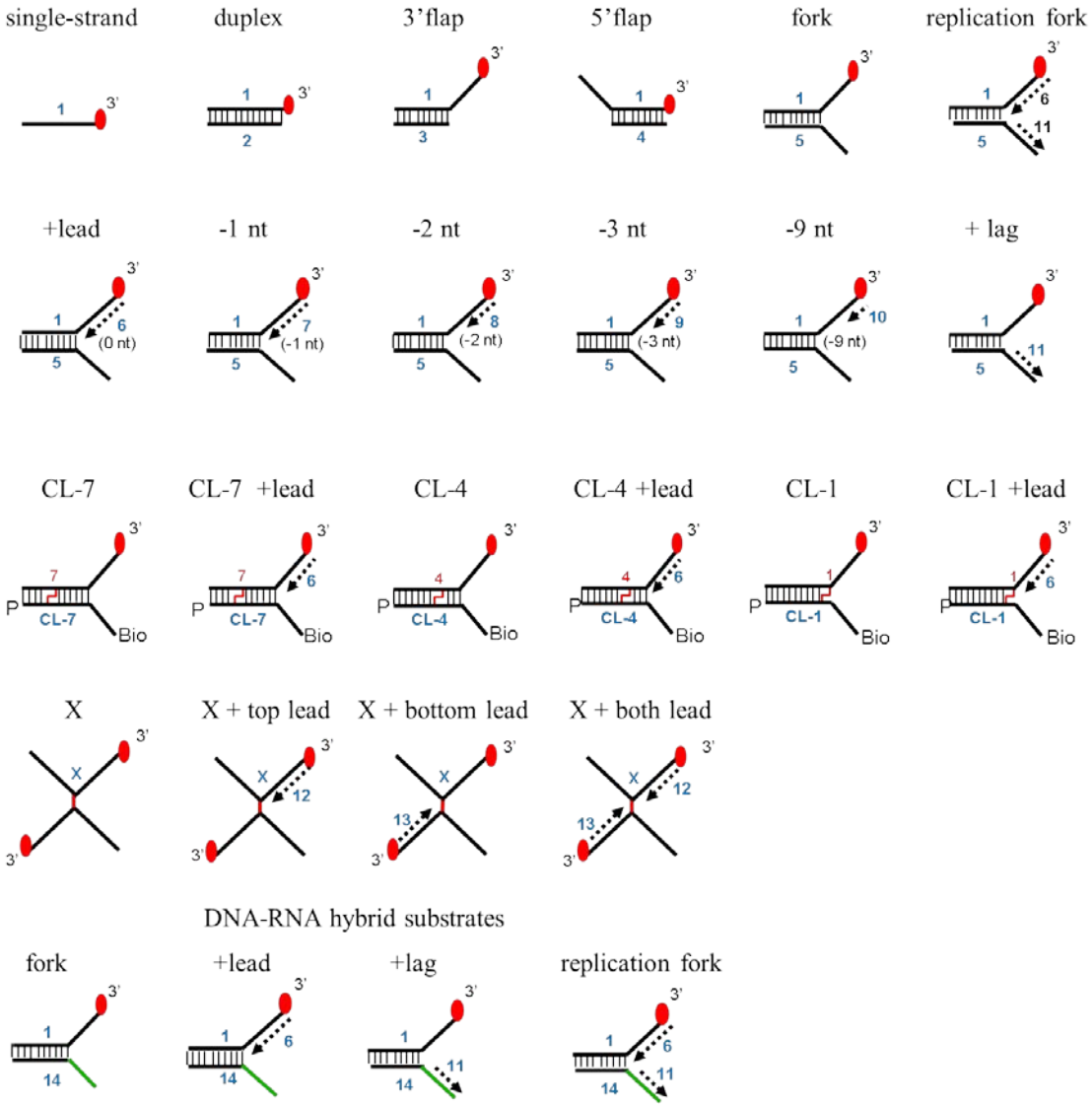


Figure 2. 1: A schematic representation of the DNA substrates utilised in this study. The code for the DNA oligonucleotides (as stated in the table 2) used to assemble the DNA substrates are indicated in blue. 'P' stands for 3'-phosphate; 'Bio' stands for 5'-biotin; green lines represent regions of RNA; red numbers represent the location of the triazole DNA interstrand crosslink with respect to the ss/dsDNA junction.

2.4 Nuclease assay

Unless otherwise stated, nuclease assays were carried out using 10 nM DNA substrate and 40 nM (62 ng) XPF-ERCC1 in nuclease buffer (25 mM HEPES pH 8.0, 40 mM NaCl, 10% glycerol, 0.5 mM ß-mercaptoethanol, 0.1 mg/ml BSA) containing either 0.4 mM $MnCl_2$ or 10 mM $MgCl_2$ in a 10 µL reaction mixture.

To analyse the nuclease activity of XPF-ERCC1, DNA substrates and XPF-ERCC1 proteins were incubated at 30^{o}C for 1 hour in nuclease buffer containing 0.4 mM $MnCl_2$. To analyse the effect of RPA on the nuclease activity of XPF-ERCC1, DNA substrates were pre-incubated with 80 nM RPA (92.8 ng) on ice for 10 minutes followed by the addition of XPF-ERCC1. The reactions were incubated at 30^{o}C for 1 hour in nuclease buffer containing 0.4 mM $MnCl_2$. To analyse the activity of hSNM1A post-incision by XPF-ERCC1, DNA substrates were reacted with XPF-ERCC1 at 30^{o}C for 1 hour in nuclease buffer containing 10 mM $MgCl_2$, followed by the addition of 0.8 nM hSNM1A (1µL of 8 nM stock concentration). The reactions were further incubated at 37^{o}C for increasing time.

Nuclease assays were quenched by adding 3 µL stop solution (95 % formamide/5 % EDTA) and heating at 95^{o}C for 5 minutes. Samples were loaded onto 10 % or 20 % denaturing (7M urea) PAGE gel (19:1 Acylamide/Bis) gels containing 1x TBE and run for 2 hours at 525 V. Gels were fixed in fixing solution (40% methanol/ 20% Acetic Acid/5% glycerol) for 1 hour and dried in a Gel Dryer at 50^{o}C for 4 hours or 80^{o}C for 2 hours. Reaction products were visualised by phosphorimager. Unless otherwise stated, all nuclease assays are representative of experiments performed at least twice.

2.5 Enzyme mobility shift assay (EMSA)

Unless otherwise stated, EMSA were carried out using 10 nM DNA substrate in nuclease buffer (25 mM HEPES pH 8.0, 40 mM NaCl, 10% glycerol, 0.5 mM ß-mercaptoethanol, 0.1 mg/ml BSA) containing 5 mM $CaCl_2$ in a 10 µL reaction mixture.

To analyse RPA-DNA complex formation, DNA substrates were incubated with indicated amounts of RPA on ice for 30 minutes. To analyse RPA-XPF-ERCC1 complex formation on DNA substrates, DNA substrates were pre-incubated with 80 nM RPA on ice for 10 minutes. 40 nM of XPF-ERCC1 was then added and incubation proceeded for a further 30 minutes on ice.

Reactions for EMSA were quenched by adding 3 µL ice cold 50% glycerol, and samples were immediately loaded onto chilled 5% non-denaturing PAGE gel (19:1 Acylamide/Bis) gels containing 0.5x TBE and run for 2 hours at 100 V in the cold room. Gels were fixed in fixing solution (40% methanol/20% Acetic Acid/5% glycerol) for 10 minutes and dried in a Gel Dryer at 50^{o}C for 4 hours or 80^{o}C for 2 hours. Reaction products were visualised by phosphorimager. Unless otherwise stated, all nuclease assays are representative of experiments performed at least twice.

3 Characterisation of XPF-ERCC1 endonuclease activity on model native and damaged replication fork structures

3.1 Introduction

The human XPF-ERCC1 heterodimeric complex is a structure-selective endonuclease which incises branched DNA intermediates that arise during nucleotide excision repair (NER), inter-strand crosslink (ICL) repair and homologous recombination (HR), amongst other DNA processes. Recombinant XPF-ERCC1 complex cleaves branched DNA substrates 2-8 nucleotides (nt) from the ssDNA/dsDNA (fork) junction within the DNA duplex region on the strand with a 3'-ssDNA arm (de Laat, et al. 1998a).

A specialised role for XPF and ERCC1 in ICL repair was implied when XPF or ERCC1 defective Chinese hamster ovaries (CHO) cells were hypersensitive to cross-linking agents compared to cells defective for other NER factors (De Silva, et al. 2000; Hoy, et al. 1985). Efforts to elucidate the role of XPF-ERCC1 in a cell-free replication-coupled repair system have demonstrated ICL repair is abolished when XPF is immuno-depleted, while ablation of other nucleases also previously implicated in ICL repair namely Mus81 and Fan1, did not significantly affect ICL repair (Klein Douwel, et al. 2014). Taken together, these data suggest that XPF-ERCC1 is a key endonuclease required for ICL repair.

The most prominent current model for replication-coupled ICL repair was derived from studies in cell-free replication assays of plasmids containing a site-specific ICL in *Xenopus laevis* egg extracts. Their model proposes that ICL unhooking requires the

convergence of two replication forks upon an ICL, which creates an X-shaped structure surrounding the ICL (Klein Douwel, et al. 2014; Raschle, et al. 2008; Zhang, et al. 2015). In this model, when the replication forks encounter an ICL, both nascent leading strands stall 20-40 nt from the ICL. Eviction of the CMG helicase complexes at both replication forks enables the gradual extension of one nascent leading strand to 1 nt from the ICL which triggers ICL unhooking on the opposite template strand (Long, et al. 2014; Raschle, et al. 2008). The authors argue that the formation of the X-shaped structure surrounding an ICL is the essential requirement for ICL unhooking (Zhang, et al. 2015).

The aim of this chapter is to extend our current understanding on the role of XPF-ERCC1 in replication-coupled ICL repair. The approach I adopted was biochemical reconstitution assays using recombinant human XPF-ERCC1 on synthetic DNA substrates. The recombinant XPF-ERCC1 was first tested against DNA substrates described in the literature to confirm its endonuclease activity. Its activity is then analysed against DNA substrates that mimic the structures that arise at stalled replication forks. Finally, XPF-ERCC1 activity is tested against substrates containing a site-specific ICL that resemble the structures that arise when a single replication fork collides with an ICL and on an X-shaped structure that mimic dual replication fork convergence onto an ICL.

3.2 Purification of human XPF-ERCC1 from insect cells

To enable the biochemical characterisation of the XPF-ERCC1 heterodimer, recombinant XPF and ERCC1 were co-purified from baculovirus-infected Hi5 insect cells using 6xHis tag on ERCC1. I have purified two versions of the complex: XPF(WT)-ERCC1 and XPF(D676A)-ERCC1. The nuclease-defective version of the complex XPF(D676A)-ERCC1 where the metal-binding residue aspartate-676 is mutated to an alanine was purified as a control to confirm that the endonuclease activity observed is attributed specifically to XPF-ERCC1 (Enzlin and Scharer 2002). I have also tested the XPF-ERCC1 that I have purified against the one purified by J. Enzlin (Enzlin and Scharer 2002), both proteins have the same level of activity. The XPF-ERCC1 complex was purified using chromatography on nickel-agarose and gel filtration (See Materials and Methods, Chapter 2 for details).

As published in the literature for gel filtration profiles for XPF-ERCC1 purification, peaks eluted in enzyme buffer (25 mM HEPES pH 8.0, 150 mM NaCl, 10% glycerol and 5 mM TCEP) at ~10 mL corresponds to proteins in an aggregated state in the void volume of the column, peak at ~14 mL corresponds to the XPF-ERCC1 complex and was used in our biochemical assays. The peaks eluting at ~17 mL correspond to monomeric ERCC1 (**Figure 3. 1**a) (Enzlin and Scharer 2002). The fractions for XPF-ERCC1 complex (eluted at 13-15 mL) were pooled and an aliquot was run on SDS-PAGE gel which confirms that the fractions contain highly purified XPF-ERCC1 (**Figure 3. 1**b). I have also confirmed that XPF and ERCC1 were the proteins purified using Western blotting (data not shown).

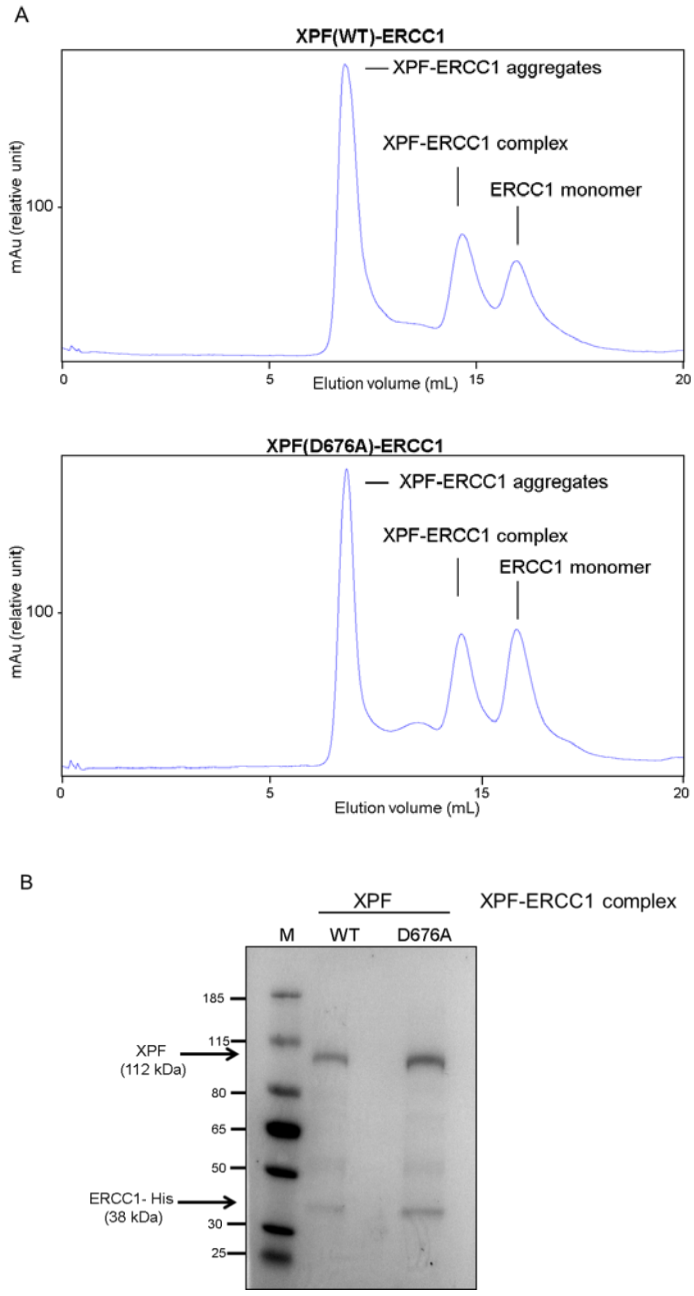


Figure 3. 1: Purification of human XPF-ERCC1.

A) Gel filtration profiles for XPF(WT)-ERCC1 and XPF(D676A)-ERCC1. Peaks at ~10mL elution volume correspond to proteins in an aggregated state in the void volume of the column, the peak at ~14mL to the XPF-ERCC1 complex and the peak at ~17mL to monomeric ERCC1. B) SDS-PAGE gel (4-12%) of the fraction for XPF-ERCC1 complex after gel filtration stained with InstantBlue. The molecular weight of XPF and ERCC1 are 112 and 38 kDA, respectively. M = molecular weight marker.

3.2 Generation of radiolabelled DNA substrates

The endonuclease activity of the recombinant XPF-ERCC1 was analysed against an array of synthetic DNA substrates that mimic the DNA intermediates that may arise during DNA repair and at stalled replication forks. The substrates were generated by first labelling a synthetic 50-mer DNA oligonucleotide (sequence in Materials and Methods, Chapter 2) introducing a (^{32}P) radioactive phosphate at its 5'- or 3'-end. 3'-end labelling was achieved by incubating the oligonucleotide with radioactive [α-^{32}P]dATP in the presence of terminal deoxynucleotidyl transferase (TdT). Under the conditions used, our laboratory has previously demonstrates that TdT facilitates the transfer of a single radioactive [α-^{32}P]dATP to the 3' end of the oligonucleotide. 5'-end labelling was achieved using [γ-^{32}P]dATP in the presence of T4 polynucleotide kinase (T4 PNK), which transfers of radioactive [γ-^{32}P] to the 5'-end of the oligonucleotide. The radiolabelled oligonucleotide is then annealed to unlabelled complementary DNA oligonucleotide(s), which will define the secondary structure of the DNA substrates. The quality of the DNA substrates generated was then determined by analysis on 10% non-denaturing polyacrylamide electrophoresis (PAGE) gels (**Figure 3. 2**). Only one band is predominantly present for each substrate, which indicates that the labelled oligonucleotide is almost quantitatively annealed to the complementary oligonucleotide(s) and the annealed product is pure. Additionally, by comparing the migration of each substrate structure in the gel compared to a 50-mer marker and against other substrates, the different structures generated were presumed correct: the substrates with a simpler secondary structure travel further in the gel compared to more complex substrates.

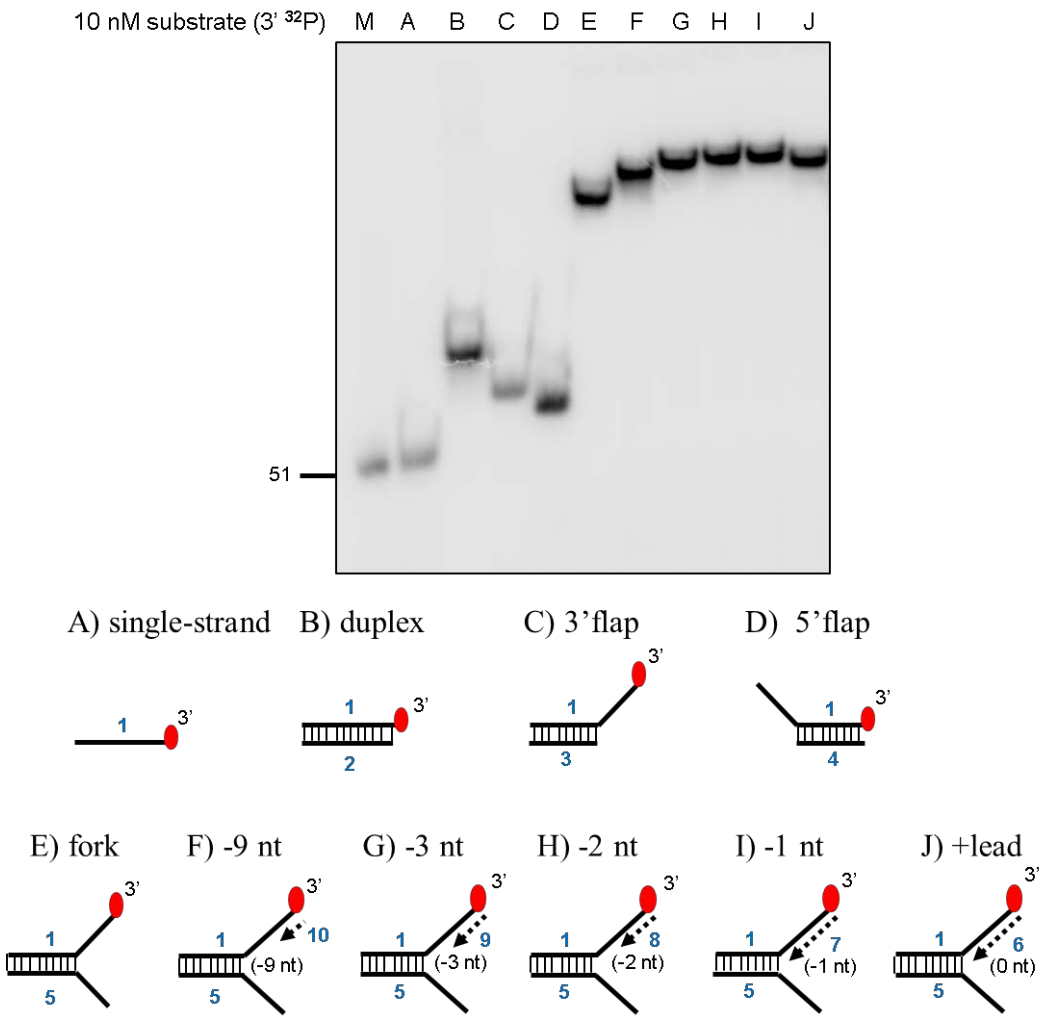


Figure 3. 2: Analysis of the quality of the radiolabelled DNA substrates generated for this study.

A representative analysis of the quality of radiolabelled DNA substrates generated in this study using a 10% non-denaturing PAGE gel. The substrates are radiolabelled using [α-^{32}P]dATP, denoted as a red dot on its 3'-end. Substrate concentrations of 10 nM per reaction are used throughout this study. Every batch of radiolabelled DNA substrates generated are analysed for its quality prior to use in assays. Alphabets correspond to the DNA structure assembled, as shown below the gel. Blue numbers stands for the DNA oligonucleotide utilised to assemble the DNA substrates, as described in the Chapter 2 – Materials and Methods. M = 3'radiolabelled oligonucleotides of the size indicated.

3.3 Analysis of XPF-ERCC1 activity on model replication fork structures

3.3.1 XPF-ERCC1 is a 3'-flap endonuclease

To confirm the structure-selective properties of the recombinant XPF-ERCC1 that I purified, 10 nM of 'simple' DNA substrates labelled on its 3'-end (ssDNA, dsDNA, 3' flap, 5' flap and fork) were incubated with increasing concentrations of the protein complex (0, 20, and 40 nM) in the presence of 0.4 mM manganese (Mn^{2+}) for 1 hour at 30^{o}C. Two divalent cations are required by many DNA metabolising enzymes to hydrolyse the phosphodiester bonds where the divalent cations are jointly coordinated by the DNA substrate and catalytic residue on the enzyme (Yang, et al. 2006). It was previously shown that XPF-ERCC1 activity is optimal in the presence of 0.4 mM manganese (Mn^{2+}) (Enzlin and Scharer 2002). Reactions were stopped with 95 % formamide/5 % EDTA and heating at 95^{o}C for 5 minutes. The samples were analysed in 20 % denaturing PAGE.

As reported in the literature, XPF-ERCC1 selectively incises DNA substrates with a protruding 3'-ssDNA arm: a 3'-flap and the top strand of a 'splayed arms' or 'fork-like' substrate. XPF-ERCC1 activity was undetectable on ssDNA, dsDNA and 5'-flap DNA substrates. The concentration of XPF-ERCC1 that shows maximal cleavage of the DNA fork substrate is four-fold higher than that of the substrate, in line with previous studies. Therefore this concentration was used throughout this study (Enzlin and Scharer 2002; Kuraoka, et al. 2000) (**Figure 3. 3**a). XPF-ERCC1 incises a fork substrate within the dsDNA region at two main locations: 2 nt from the junction which releases a 25-mer product; and 6nt from the junction which releases a 29-mer product (**Figure 3. 3**b). At

the concentration of 40 nM, XPF-ERCC1 activity is approximately 2-fold higher on a fork substrate (~90% substrate incisions) compared to a 3'-flap (~40% substrate incisions) suggesting that the 5'-ssDNA arm on the fork substrate may have a positive effect on the XPF-ERCC1 activity (**Figure 3. 3**c), as previously observed (Tsodikov, et al. 2005).

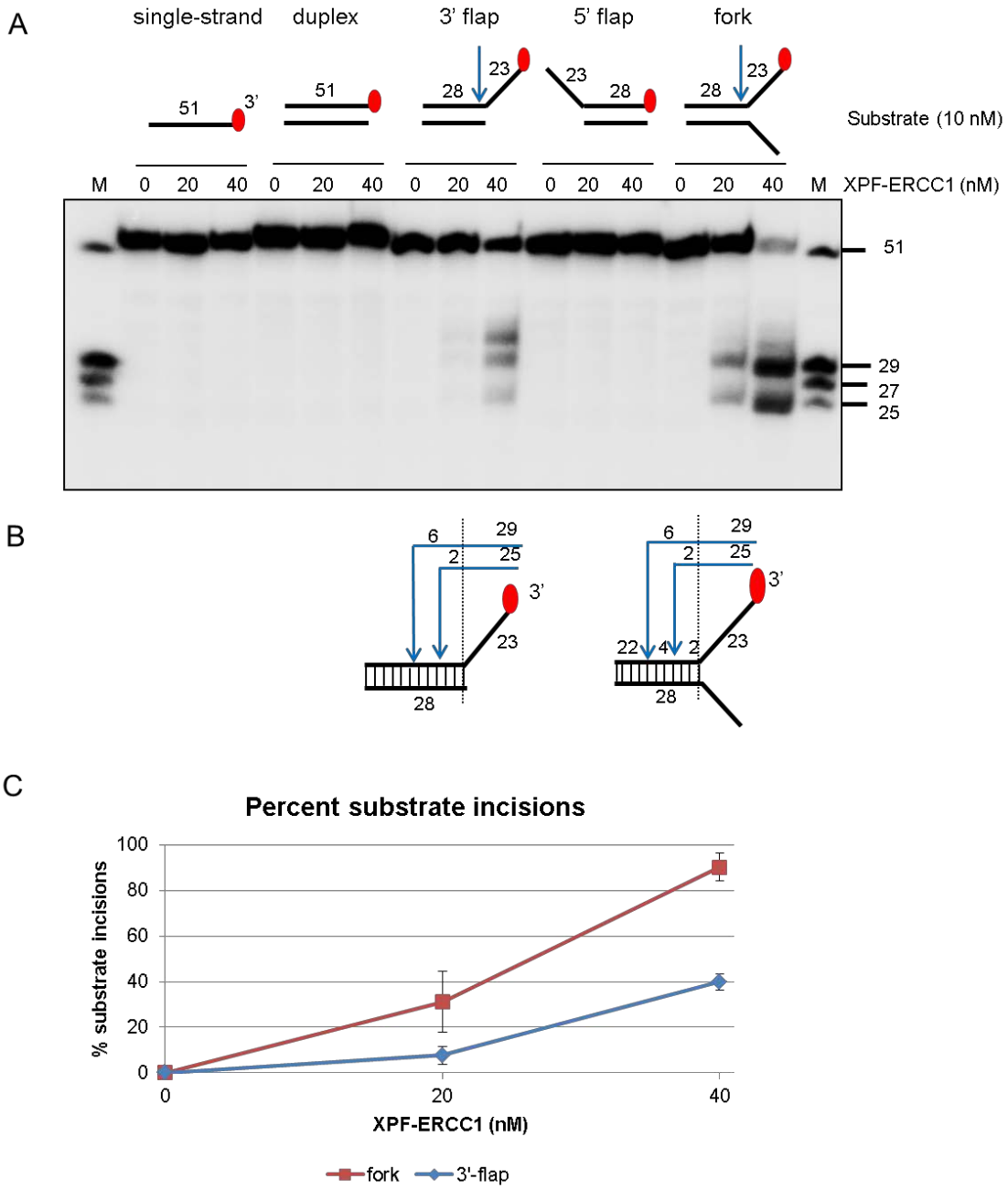


Figure 3. 3: XPF-ERCC1 is a 3’ flap endonuclease.

A) A nuclease assay using 10 nM of 3’-radiolabelled ‘simple’ DNA substrates (ssDNA, dsDNA, 3’-flap, 5’-flap and fork) against increasing concentration of XPF-ERCC1 (0, 20, 40 nM) for 1 hour at 30^{o}C, in the presence of 0.4 mM manganese (Mn^{2+}). XPF-ERCC1 only incises substrates with protruding 3’ssDNA arm: a 3’ flap and a model fork substrate. B) A schematic representation of the locations of XPF-ERCC1 incisions: 2 nt (giving a 25-mer product) and 6 nt (29-mer product) from the ssDNA/dsDNA junction, within the duplex region on the strand with a protruding 3’ssDNA arm. M = radiolabelled oligonucleotides of the sizes indicated. C) Quantification of percentage of XPF-ERCC1 incisions on 3’-flap and fork substrates. All data points represent an average of three independent experiments with SEM as error bars.

3.3.2 XPF-ERCC1 incises a model fork structure at two main positions

By conducting a time-course assays using a fork substrate, I determined that the two incision products formed almost concurrently as seen by their presence at equivalent intensity throughout the assay. This suggests that XPF-ERCC1 incises an individual fork substrate at either 2 nt (giving a 25-mer product) or 6 nt (giving a 29-mer product) from the junction (**Figure 3. 4**). Therefore it is possible that XPF-ERCC1 does not incise both sites on the same molecule of substrate because the first incision at either location abolishes the substrate recognised by XPF-ERCC1 to carry out the second incision.

3.3.3 Mutation of a metal-binding residue in XPF-ERCC1 eliminates endonuclease activity

The metal-binding sites of an enzyme can be determined using affinity cleavage technique which is the replacement of the metal cofactor (manganese, Mn^{2+} or magnesium, Mg^{2+}) with iron (Fe^{2+}) (Zaychikov, et al. 1996). Divalent iron has the ability to reduce molecular oxygen to form superoxide and hydroxyl radicals which cause cleavage of peptide bonds in regions of the metal-binding site. By sequencing the peptide fragments obtained, the location of the metal-binding site, usually aspartate or glutamate residues can be determined (Lykke-Andersen, et al. 1997; Zaychikov, et al. 1996). A metal-binding residue of XPF-ERCC1 was determined to be aspartate-676 which when mutated to alanine (D676A) has been reported to render the complex inactive (Enzlin and Scharer 2002).

To confirm that the endonuclease activity described elsewhere in this chapter were attributable to XPF-ERCC1, I also purified the nuclease-defective mutant, XPF(D676A)-ERCC1. Substrate incisions were not observed using the mutant complex, further confirming that the endonuclease activity observed using the wild-type complex is attributed to XPF-ERCC1 (**Figure 3. 5**).

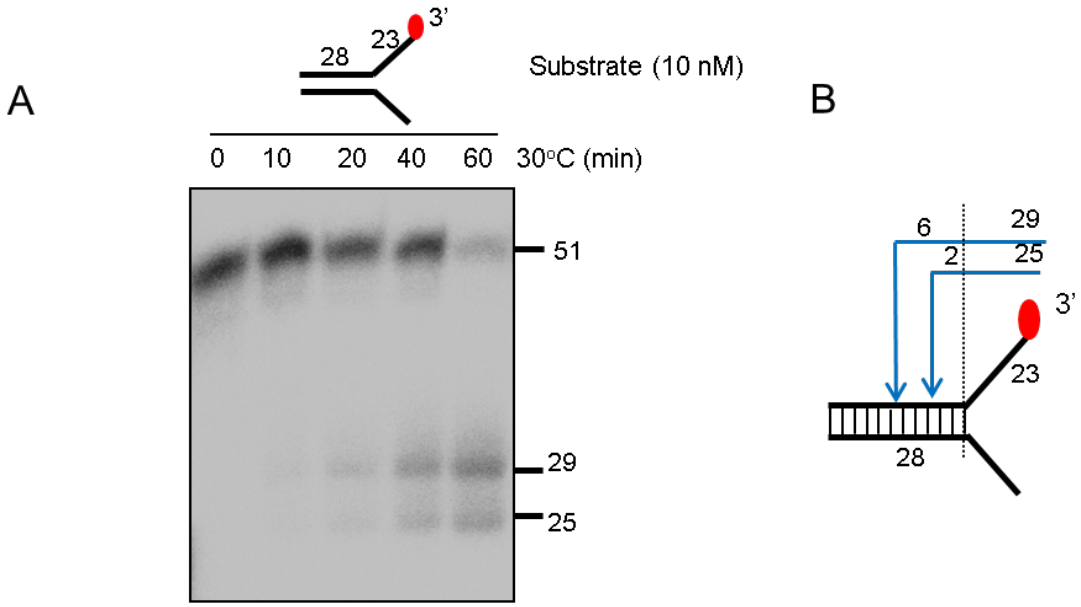


Figure 3. 4: Two XPF-ERCC1 incision products formed at almost equivalent intensity in a time-course assay.

A) A nuclease assay using 10nM 3'-radiolabelled fork substrate incubated with 40 nM XPF-ERCC1 for increasing time (0, 10, 20, 40, 60 min) at 30^{o}C. B) A schematic representation of XPF-ERCC1 incisions: two products corresponding to incisions at 2 nt (giving a 25-mer) and 6 nt (giving a 29-mer) from the junction. M = radiolabelled oligonucleotides of the sizes indicated.

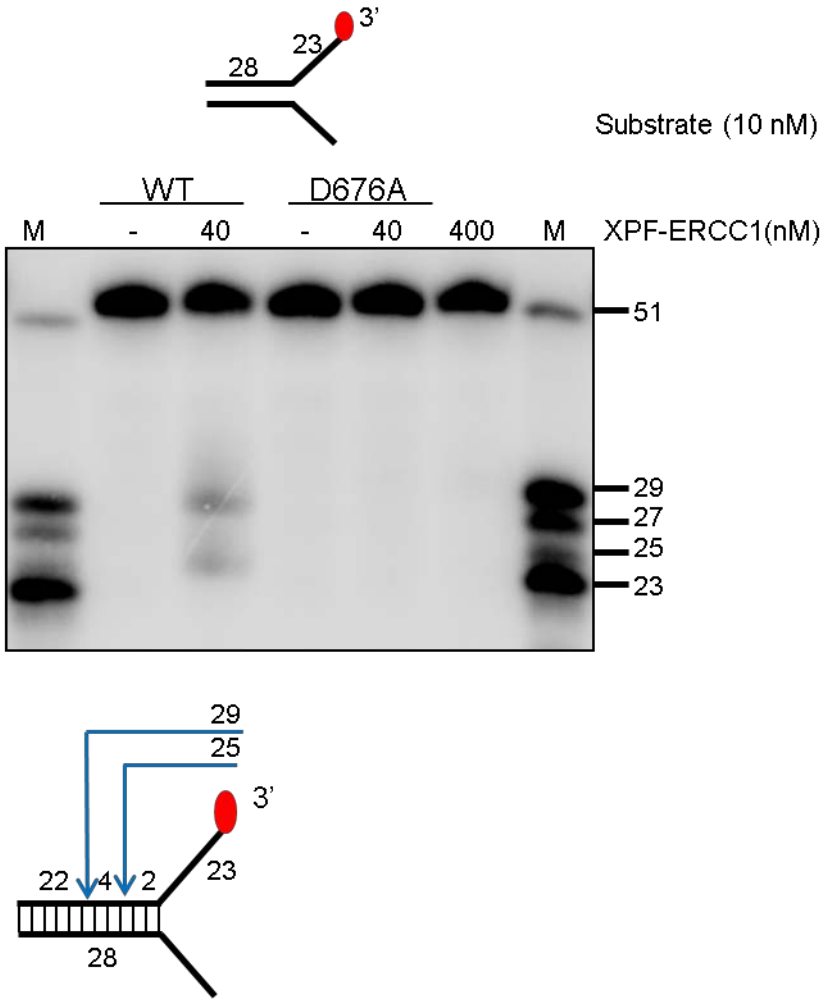


Figure 3. 5: Mutation at a metal-binding residue in XPF renders it devoid of any nuclease activity.

Activity of wild-type (WT) and nuclease-defective (D676A) XPF-ERCC1 against a model 3'-radiolabelled fork substrate. Incision products are not detectable with the nuclease-defective D676A mutant. M = radiolabelled oligonucleotides of the sizes indicated.

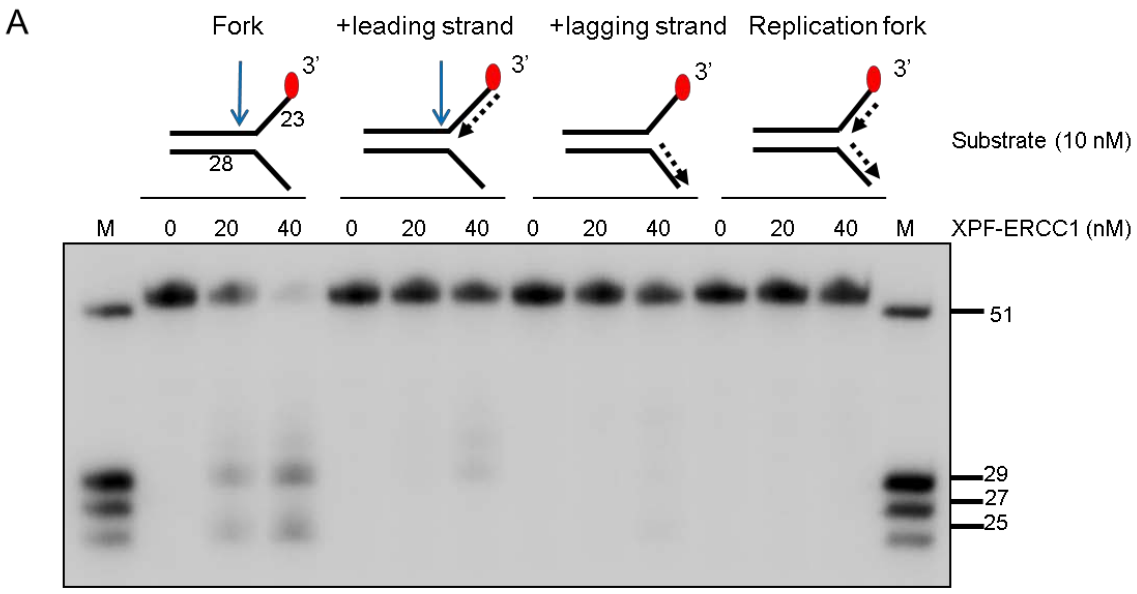
3.3.4 XPF-ERCC1 activity is altered by the presence of a model nascent leading and/or lagging strands

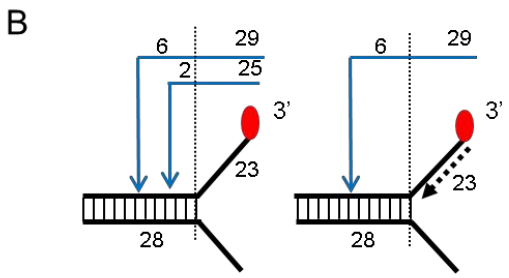
After confirming the structure-selective property of the recombinant XPF-ERCC1 complex, I wished to analyse its activity on DNA substrates that mimic structures that may arise at stalled replication forks. Current models for ICL repair suggest that the arrival of a nascent leading strand to the fork junction triggers ICL incision (Klein Douwel, et al. 2014; Raschle, et al. 2008; Zhang, et al. 2015). To further investigate the effect of leading strand and/or lagging strands, the fork substrate described were annealed with oligonucleotide(s) complementary to the 3'-ssDNA and/or 5'-ssDNA arms to mimic the presence of a nascent leading and/or lagging strands, respectively. The array of fork substrates generated is denoted as: fork, fork with a model nascent leading strand (+leading strand), fork with a model nascent lagging strand (+lagging strand), and fork with both model nascent leading and lagging strands (replication fork) (**Figure 3. 6**a).

Intriguingly, on a fork substrate with a model nascent leading strand, XPF-ERCC1 activity is significantly reduced compared to its activity on a simple fork substrate. At the concentration of 40 nM, the presence of a leading strand on a fork substrate reduces XPF-ERCC1 incision to only ~15%, compared to ~90% incision on a fork substrate. XPF-ERCC1 is 6-fold more efficient at incising a fork substrate compared to a fork with a leading strand (**Figure 3. 6**c). As described earlier, XPF-ERCC1 incises a fork substrate at two main locations: 2 nt from the junction which releases a 25-mer product; and 6 nt from the junction which releases a 29-mer product. However, on a fork with a

leading strand, only the incision 6 nt from the junction is detectable (29-mer product), with loss of incision 2 nt from the junction (25-mer product) (**Figure 3. 6**b). This indicates that the presence of a leading strand on the fork structure has both an inhibitory effect on XPF-ERCC1 activity as well as shifting and focusing its incision to one site 5' further into the DNA duplex, 6 nt from the junction. Notably, XPF-ERCC1 activity is completely abrogated on forks with a model nascent lagging strand (+lagging strand) or a fork with both leading and lagging strands (replication fork), further confirming the stringent structure-selective property of the endonuclease protein complex.

To further confirm the location of XPF-ERCC1 incisions, the same array of fork substrates were radiolabelled on its 5'-end, where the label denoted as a green dot in the figures. 5'-radiolabelling of the substrate enables the detection of additional products produced by potential additional incisions closer to the 5'-end of the substrate. The XPF-ERCC1 incision profile on 5'-radiolabelled fork substrates corresponds to the incision sites previously described for 3'-radiolabelled substrates: the incision 2 nt and 6 nt from the junction releases a 26-mer and a 22-mer products, respectively, on a 5'-radiolabelled substrates. The incision at 2 nt from junction (26-mer product), is also absent on 5'radiolabelled fork with a leading strand. Additional minor incisions of XPF-ERCC1 further into the 5' end were also detectable: 9 nt (giving a 19-mer product) and 11 nt (giving a 17-mer product) from the junction (**Figure 3. 7**), however they will not be discussed in detail in this study since they represent very minor site of incision.


A
Fork
+leading strand
+lagging strand
Replication fork
3'
23
28
Substrate (10 nM)
M 0 20 40 0 20 40 0 20 40 0 20 40 M
XPF-ERCC1 (nM)
51
29
27
25


B
6
29
2
25
3'
23
28

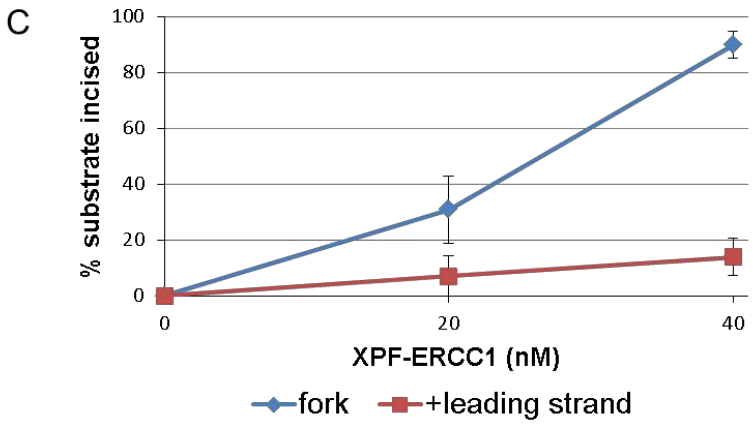

C
% substrate incised
100
80
60
40
20
0
0
20
40
XPF-ERCC1 (nM)
fork
+leading strand

Figure 3. 6: The presence of a model nascent leading strand and/or lagging strands on a fork substrate alters XPF-ERCC1 activity.

A) XPF-ERCC1 titration (0, 20, 40 nM) against an array of 3'-radiolabelled fork substrates (fork; fork with a model nascent leading strand (+leading strand); fork with a model nascent lagging strand (+lagging strand); and a fork with both nascent leading and lagging strand (replication fork) B) A schematic representation of the locations of XPF-ERCC1 incisions on a fork substrate and a fork with a leading strand. XPF-ERCC1 incises a fork substrate at two locations: 2 nt (giving a 25-mer product) and 6 nt (29-mer product) from the fork junction, but incises a fork with a leading strand at one location: 6 nt (29-mer product) from the fork junction. M = radiolabelled oligonucleotides of the sizes indicated. C) Quantification of percentage of XPF-ERCC1 incisions on a fork substrate and fork with a leading strand. XPF-ERCC1 activity is 6-fold higher on a fork substrate compared to a fork with a leading strand. All data points represent an average of three independent experiments with SEM as error bars.

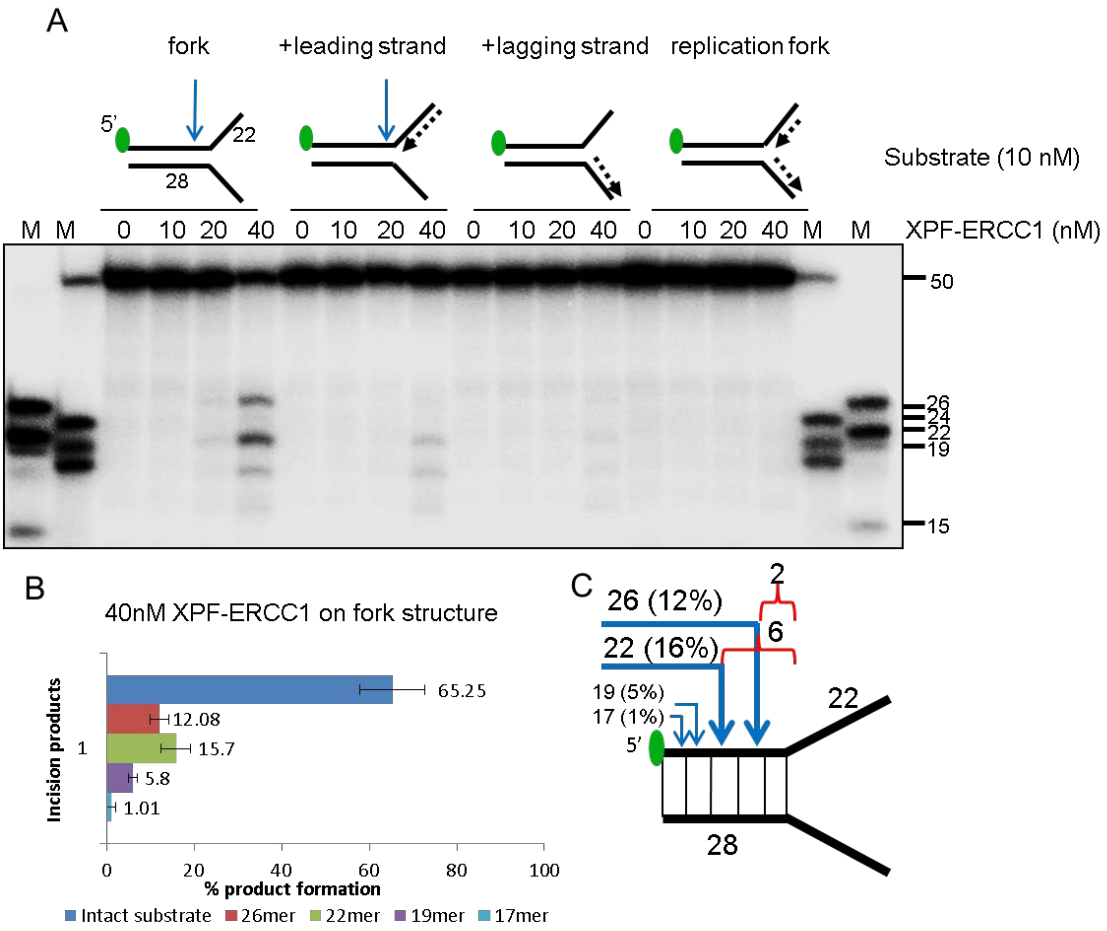


Figure 3. 7: XPF-ERCC1 activity on 5'-radiolabelled array of fork substrates.

XPF-ERCC1 titration (0, 20, 40 nM) against an array of 5'-radiolabelled fork substrates. 5'-radiolabelling of the substrates detects any additional minor incisions by XPF-ERCC1 5' further into the duplex region. B) Quantification of percentage of 40 nM XPF-ERCC1 incisions at each position on a fork substrate. All data points represent an average of three independent experiments with SEM as error bars. C) A schematic representation of the locations and products of XPF-ERCC1 incisions on a fork substrate. The percentage corresponds to graph in C. M = radiolabelled oligonucleotides of the sizes indicated.

3.3.5 XPF-ERCC1 does not incise DNA substrates with 5'-flap

I then determined whether XPF-ERCC1 also incises the 'bottom strand' which has a 5'-ssDNA arm of the same array of fork substrates, used previously where this strand was unlabelled as shown in **Figure 3. 6** and **Figure 3. 7**. In this assay, the same array of fork substrates was 5'-radiolabelled either on the 'top' or the 'bottom' strand which enable simultaneous analysis of XPF-ERCC1 on either strand. XPF-ERCC1 activity was not detectable on the 'bottom strand' of any of substrates in the array, further confirming that XPF-ERCC1 is a 3'-flap endonuclease (**Figure 3. 8**).

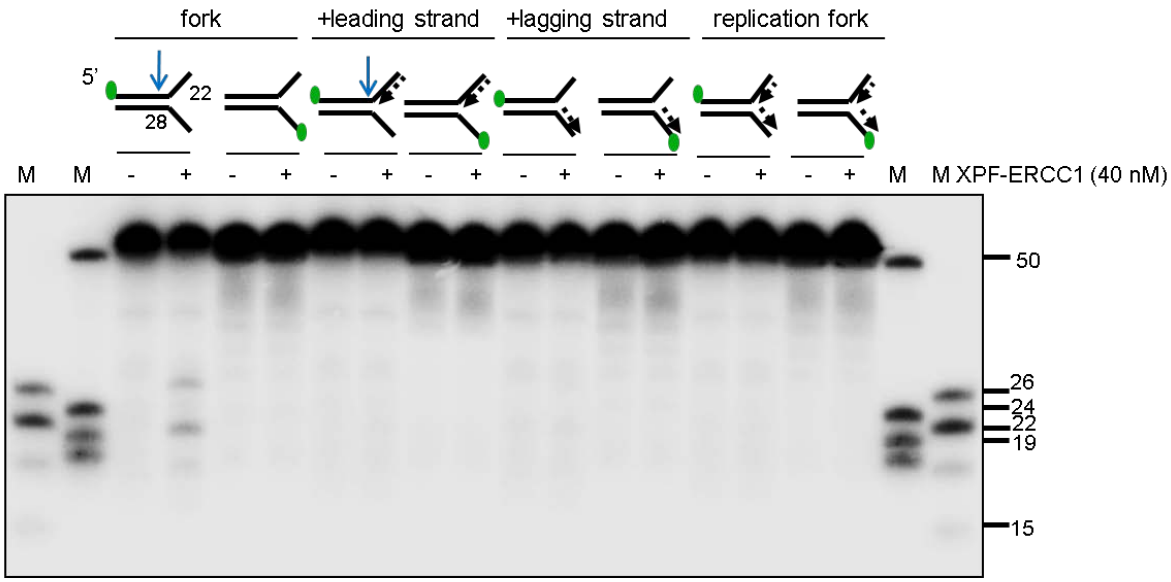


Figure 3. 8: XPF-ERCC1 activity is absent on the 'bottom strand' with 5'-ssDNA arm of fork substrates.

40 nM XPF-ERCC1 against an array of fork substrates 5'-radiolabelled either on the 'top' or the 'bottom' strand. XPF-ERCC1 activity is not detectable on the bottom strand. M = radiolabelled oligonucleotides of the sizes indicated.

3.3.6 Leading strand progression to the fork junction inhibits XPF-ERCC1 activity

Following on from the discovery that the presence of a model nascent leading strand inhibits XPF-ERCC1 activity on fork structures, I wished to explore further the effect of leading strand progression to the fork on XPF-ERCC1 activity, because the current model for ICL repair illustrates that when the replication forks collides with an ICL, the nascent leading strands initially stall 20 nt from the fork junction. One nascent leading gradually progresses to 1 nt to the ICL, which triggers dual incisions flanking the ICL (Klein Douwel, et al. 2014; Raschle, et al. 2008; Zhang, et al. 2015).

To emulate the dynamic nature of replication fork progression *in vitro*, I generated a fork substrate annealed to increasing length of model nascent leading strand denoted as -10 nt, -4 nt, -3 nt, -2 nt and -1 nt from the junction. The fork substrate without a leading strand possesses 22 nt ssDNA arms, which represent the first initial stalling of the replication fork, as described in the literature. Each substrate in this array represents a snapshot of the gradual extension of the leading strand to the fork junction *in vivo*. Intriguingly, XPF-ERCC1 activity gradually decreases as the leading strand extends closer to the junction (**Figure 3. 9**a). At the concentration of 40 nM, XPF-ERCC1 activity is highest on a fork substrate and gradually decreases as the leading strand extends closer to the junction Furthermore, XPF-ERCC1 incision is also gradually shifted and focused to a single position 6 nt from the fork junction, with loss of the incision closest to the junction, starting from a fork substrate with a leading strand -4 to -1 nt from the junction (**Figure 3. 9**b). This data further confirms that the presence of a leading strand has a major inhibitory effect on XPF-ERCC1 activity.

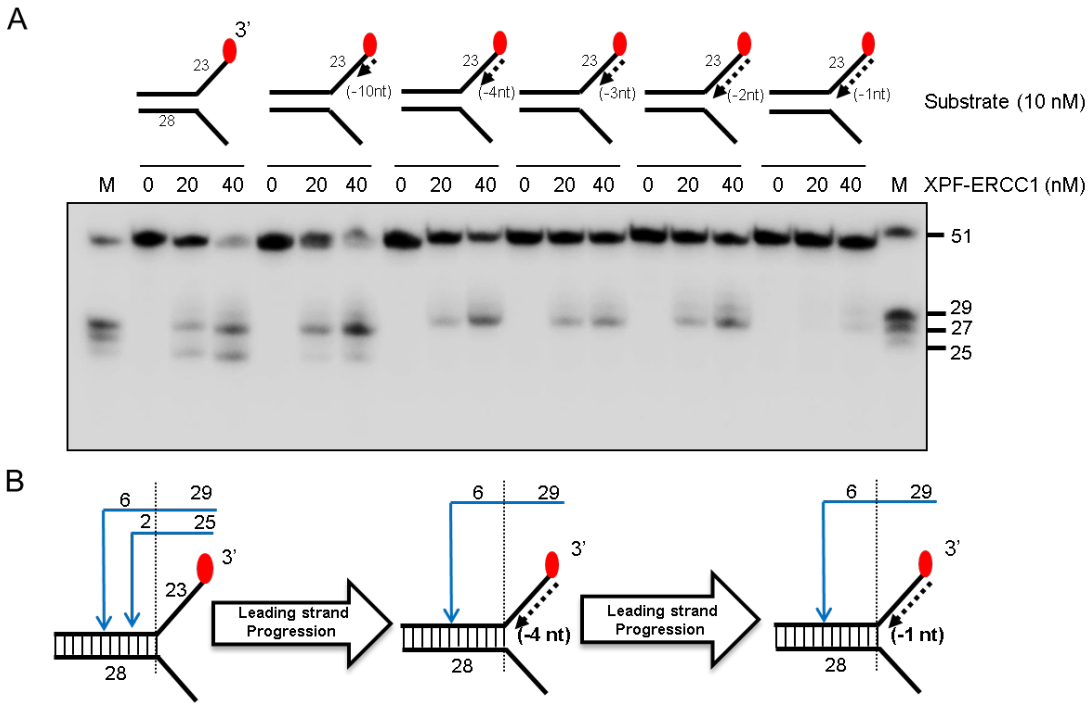


Figure 3. 9: XPF-ERCC1 activity gradual decreases when the leading strand extends closer to the fork junction.

A) XPF-ERCC1 titration (0, 20, 40 nM) against 3'-radiolabelled fork substrates with increasing length of a model nascent leading strand -10 nt, -4 nt, -3 nt, -2 nt and -1 nt from the junction. XPF-ERCC1 activity gradually decreases as the leading strand extends closer to the fork junction. B) A schematic representation of XPF-ERCC1 incisions on a fork substrate with increasing length of a leading strand. XPF-ERCC1 incision is shifted and focused to one single site further into the duplex, 6 nt from the junction (giving a 29-mer product) as the leading strand extends from -4 to -1 nt from the junction with loss of incision 2 nt from the junction. M = radiolabelled oligonucleotides of the sizes indicated.

3.4 XPF-ERCC1 activity on substrates that model a single replication fork collision with an ICL

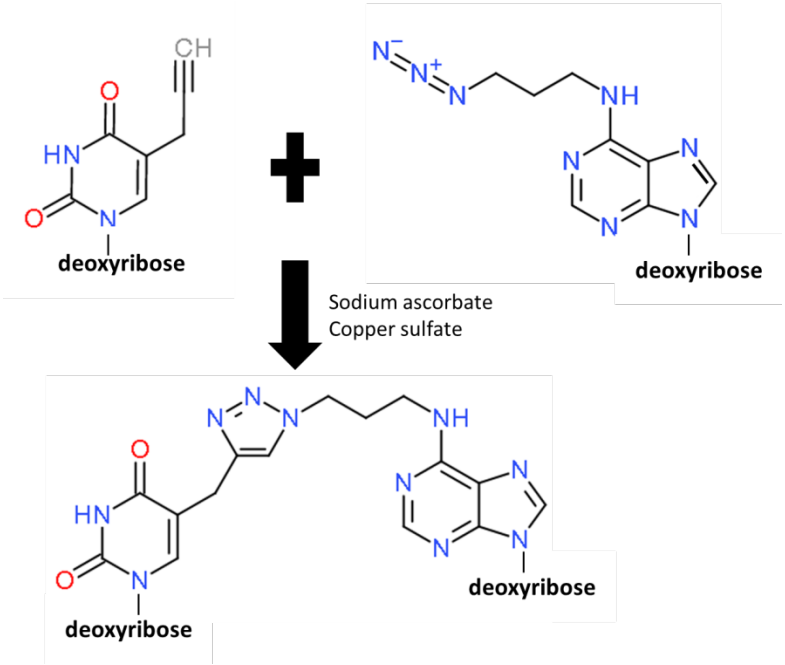
Based on the findings on model native undamaged fork substrates, I designed ICL-containing substrates that model a single replication fork collision with an ICL. The substrates contain a single site-specific triazole ICL, placed at three different locations. The ICL-containing substrates were synthesized by Dr. Joanna McGouran in Prof. Tom Brown's lab (Deparment of Chemistry, University of Oxford). Triazole ICL have previously been used in *in vitro* biochemical studies of Fan1 in ICL processing (Wang, et al. 2014). The location of the ICL is chosen based on the locations of XPF-ERCC1 incisions targeted on the model undamaged fork substrate. On substrate CL-7, XPF-ERCC1 predicted incisions are 3' to the ICL; on substrate CL-4, XPF-ERCC1 predicted incisions are 5' and 3' flanking the ICL; and on substrate CL-1, the ICL is located at the junction and XPF-ERCC1 predicted incisions are 5' to the ICL.

The ICL is introduced in the fork substrate by modifying the 'top strand' on the nucleotide where the ICL will be located with an azide group (location of ICL 5' from the fork junction: 7^{th} nt on CL-7; 4^{th} nt on CL-4 and 1^{st} nt on CL-1) and an alkyne group on the bottom strand (location of ICL 5' from the fork junction: 8^{th} nt on CL-7; 4^{th} nt on Y3 and 1^{st} nt on CL-1). The top and bottom strands were reacted in a 'click reaction', which drives the formation of a triazole ICL from the alkyne and the azide group. The bottom strand of the substrates was blocked with a phosphate group on its 3'-end and a biotin on its 5'-end to ensure that only the top strand is radiolabelled. The structure of the triazole DNA interstrand crosslink is represented in **Figure 3. 10**a. Based on the

incision products on a native undamaged fork previously established, the predicted location of XPF-ERCC1 incisions and approximate mass of products for 3'-radiolabelled (blue arrows) or 5'-radiolabelled (maroon brackets) ICL-containing fork substrates are indicated in **Figure 3. 10**b.

When the ICL-containing fork substrates are 3'-radiolabelled, incisions 5' to the ICL will generate higher molecular weight (MW) products on denaturing gels because the bottom strand will be tethered to the products via the ICL; and incisions 3' to the ICL will generate the same low MW products as on a native undamaged fork substrate (29-mer and 25-mer, corresponding to incisions 6 nt and 2 nt 5' from the junction, respectively) since they will be released upon denaturation. Conversely, when the ICL-containing fork substrates are 5'-radiolabelled, incisions 5' to the ICL will generate low MW weight products as for a native undamaged fork substrates (22-mer and 26-mer, corresponding to incisions 6 nt and 2 nt 5' from the junction, respectively) but incisions 3' to the ICL will generate higher MW products because the bottom strand will be tethered to the products via the ICL.

A


CH
O
HN
O
N
deoxyribose
N⁻
N⁺
N
NH
N
N
N
deoxyribose
Sodium ascorbate
Copper sulfate
N
N
N
NH
O
HN
O
N
deoxyribose
N
N
N
N
deoxyribose

B

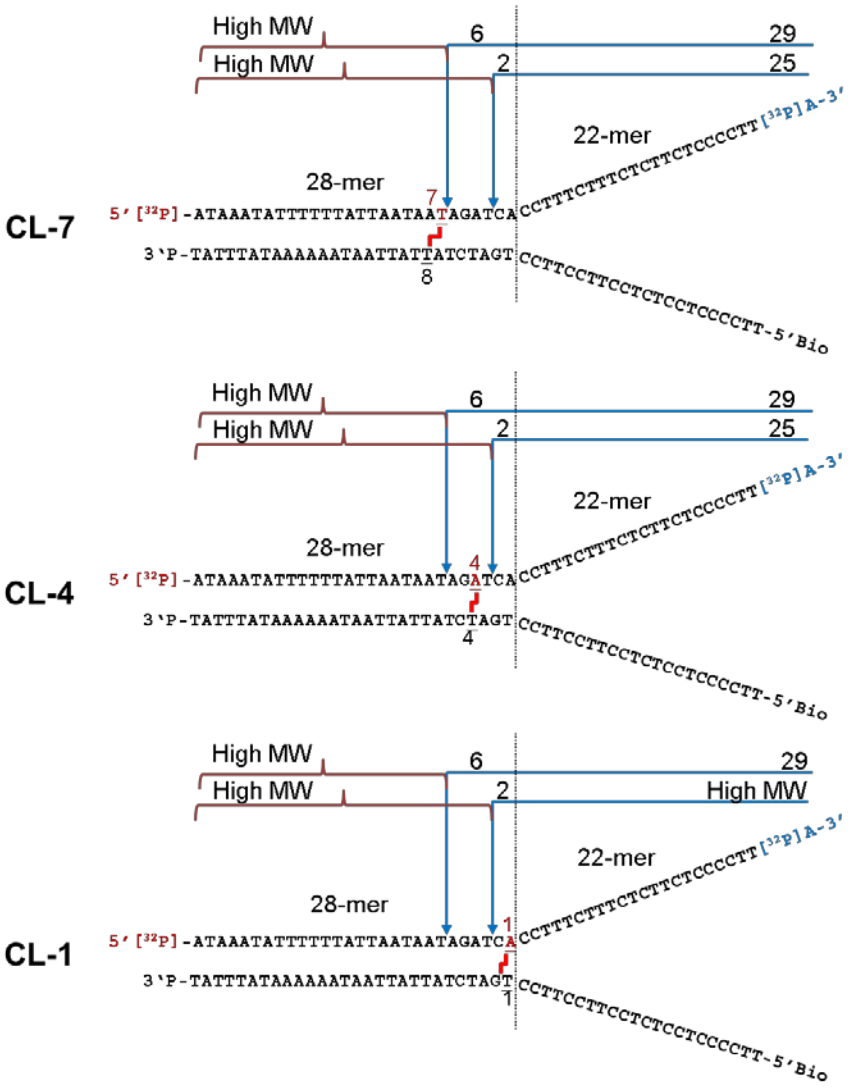

CL-7
High MW
High MW
6
2
29
25
22-mer
28-mer
7
5′ [32P] -ATAAATATTTTTTATTAATAATAGATCA
3′P-TATTTATAAAAAATAATTATTATCTAGT
8
CCTTTCTTTCTCTTCTCCCCTT [32P] A-3′
CCTTCCTTCCTCTCCTCCCCTT-5′Bio
CL-4
High MW
High MW
6
2
29
25
22-mer
28-mer
4
5′ [32P] -ATAAATATTTTTTATTAATAATAGATCA
3′P-TATTTATAAAAAATAATTATTATCTAGT
4
CCTTTCTTTCTCTTCTCCCCTT [32P] A-3′
CCTTCCTTCCTCTCCTCCCCTT-5′Bio
CL-1
High MW
High MW
6
2
29
High MW
22-mer
28-mer
1
5′ [32P] -ATAAATATTTTTTATTAATAATAGATCA
3′P-TATTTATAAAAAATAATTATTATCTAGT
1
CCTTTCTTTCTCTTCTCCCCTT [32P] A-3′
CCTTCCTTCCTCTCCTCCCCTT-5′Bio

Figure 3. 10: Schematic representation of ICL-containing fork substrates that model a single replication fork collision with an ICL.

A)Structure of the triazole DNA interstrand crosslink used in this study. B) The location of the single site-specific triazole ICL is chosen based on XPF-ERCC1 incisions targeted on the model undamaged fork substrate. On substrate CL-7, XPF-ERCC1 predicted incisions are 3' to the ICL; on substrate CL-4, XPF-ERCC1 predicted incisions are 5' and 3' flanking the ICL; and on substrate CL-1, the ICL is located at the junction and XPF-ERCC1 predicted incisions are 5' to the ICL. Predicted incision products on 3'-radiolabelled (blue arrows) or 5'-radiolabelled (maroon brackets) substrates relative to the ICL are also indicated. 'P' stands for phosphate; 'Bio' stands for biotin. Length of model nascent leading strand annealed to the crosslinked substrate is 22 bases.

3.4.1 XPF-ERCC1 activity is inhibited by a model nascent leading strands on ICL-containing fork substrates

First, I analysed XPF-ERCC1 activity on 5'-radiolabelled ICL-containing fork substrates in the absence and presence of a model nascent leading strand (**Figure 3. 11**). For CL-7 (predicted incisions are 3' to the ICL), the expected high MW products were detectable corresponding to predicted incisions 3' to the ICL. For CL-4 (predicted incisions 5' and 3' flanking the ICL), the predicted products are a high MW product (corresponding to incision 5' to the ICL) and a 22mer, corresponding to incision 3' to the ICL). However, only 22-mer product was detectable indicating that incision 3' to the ICL, which is the incision 2 nt from the junction did not occur. For CL-1 (predicted incisions are 5' to the ICL), predicted products are 22-mer and 26-mer corresponding to

incisions 5’ to the ICL. However, only the 22-mer product was detectable, with loss of the 26-mer product indicating that the incision 2nt from the junction did not occur. Intriguingly, the presence of a model nascent leading strand on the ICL-containing fork substrates also inhibits XPF-ERCC1 activity, irrespective of the location of the ICL. The additional low MW products observed corresponding to minor incisions further into the duplex as also seen on 5’-radiolabelled native non-damaged fork substrates were already described in **Figure 3. 7**.

Taken together, this data indicates that 1) XPF-ERCC1 activity is altered in the presence of an ICL. Specifically, XPF-ERCC1 incision closest to the junction in CL-4 and CL-1 is inhibited. 2) A model nascent leading strand is also indicating that, a model nascent leading strand is also inhibitory for XPF-ERCC1 activity on ICL-containing substrates.

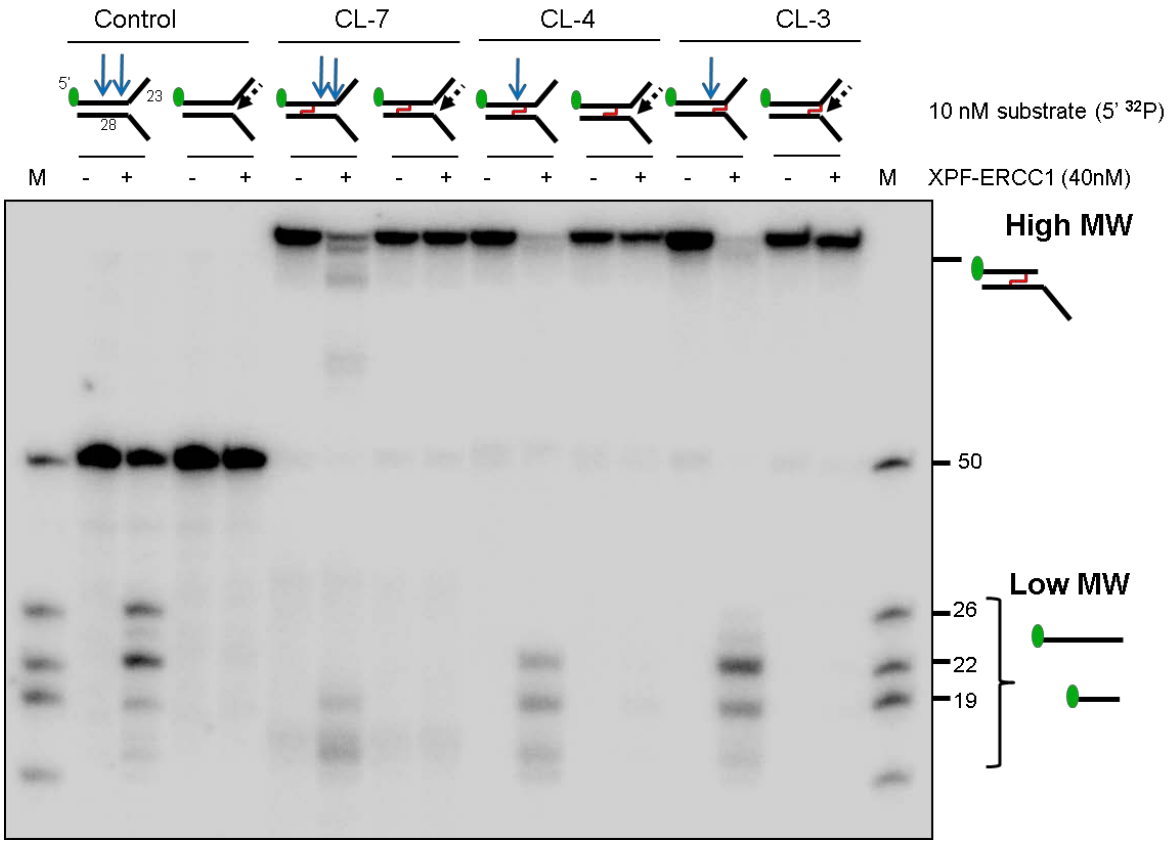


Figure 3. 11 XPF-ERCC1 activity is inhibited when a model nascent leading strands is present on ICL-containing fork substrates.

40 nM XPF-ERCC1 is reacted with 10 nM 5'radiolabelled ICL-containing substrates (CL-7, CL-4, and CL-1) in the presence of absence of a model nascent leading strand. On CL-7, XPF-ERCC1 predicted incisions are 3' to the ICL; on substrate CL-4, XPF-ERCC1 predicted incisions are 5' and 3' flanking the ICL; and on substrate CL-1, the ICL is located at the junction and XPF-ERCC1 predicted incisions are 5' to the ICL. M = radiolabelled oligonucleotides of the sizes indicated.

3.4.2 XPF-ERCC1 incision profile is altered by the presence of an ICL

To confirm the position of ICL incisions, the XPF-ERCC1 was reacted with 3'-radiolabelled ICL-containing substrates. The reaction samples were first analysed in a 20% denaturing PAGE gel. For CL-7 (predicted incisions are 3' to the ICL), 29-mer to 27-mer products were detectable corresponding to predicted incisions both 3' to the ICL. For CL-4 (predicted incisions 5' and 3' flanking the ICL), the predicted high MW corresponding to incision 5' to the ICL was not resolved in the gel; and the predicted 25-mer product corresponding to incision 3' to the ICL was not detectable. For CL-1 (predicted incisions are 5' to the ICL) both predicted incisions would result in high MW weight products which were not resolved by the gel (**Figure 3. 12**).The reaction samples were then analysed on a lower percentage (10%) denaturing PAGE gel (**Figure 3. 13**), which enable a greater resolution of high MW products from the intact substrates. On all of the substrates, high MW weight products were observed indicating incisions 5' to the ICL occur. However, low MW weight products (a smear of bands in low percenage gels) are only detectable on CL-7, as seen on 20% denaturing gel.

Collectively, this data from 5'-radiolabelling and 3'-radiolabelling of the ICL-containing fork substrates signify that an ICL located near the fork junction may alter XPF-ERCC1 incision profile compared to its incision on a native undamaged fork substrate. The incision closest to the junction (2 nt from the junction) did not occur, as seen for substrate CL-4 and CL-1 but not for CL-7 where the ICL is further into the duplex. Additionally, a model nascent leading strand also has an inhibitory effect on XPF-ERCC1 activity.

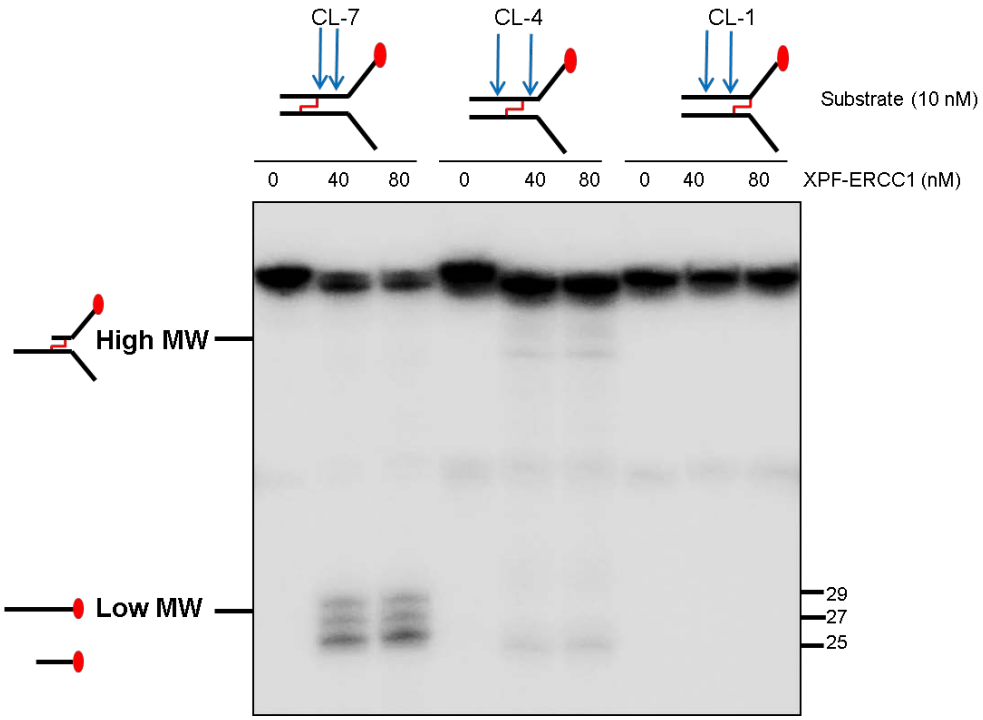


Figure 3. 12 XPF-ERCC1 incisions is altered by the presence of an ICL

40 nM XPF-ERCC1 is reacted with 10 nM 3'radiolabelled ICL-containing substrates (CL-7, CL-4, and CL-1), and analysed in 20% denaturing gel. On CL-7, XPF-ERCC1 predicted incisions are 3' to the ICL; on substrate CL-4, XPF-ERCC1 predicted incisions are 5' and 3' flanking the ICL; and on substrate CL-1, the ICL is located at the junction and XPF-ERCC1 predicted incisions are 5' to the ICL. M = radiolabelled oligonucleotides of the sizes indicated. M = radiolabelled oligonucleotides of the sizes indicated.

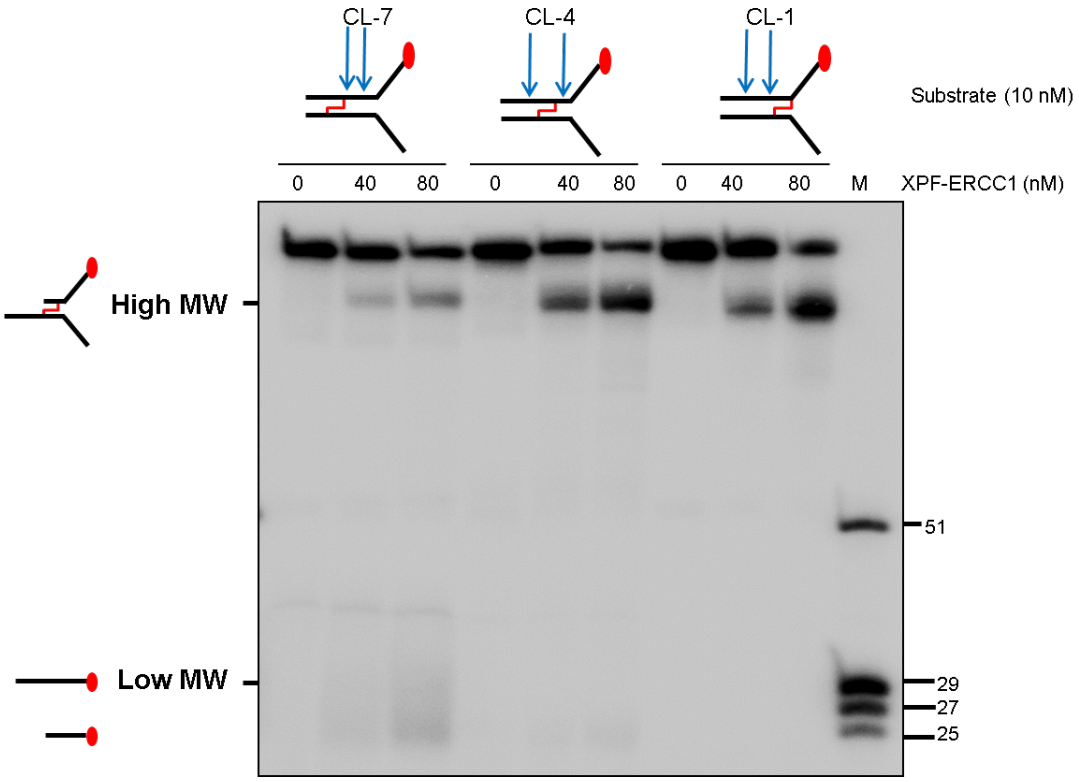


Figure 3. 13 XPF-ERCC1 incisions is altered by the presence of an ICL.

40 nM XPF-ERCC1 is reacted with 10 nM 3'radiolabelled ICL-containing substrates (CL-7, CL-4, and CL-1), and analysed in 10% denaturing gel. On CL-7, XPF-ERCC1 predicted incisions are 3' to the ICL; on substrate CL-4, XPF-ERCC1 predicted incisions are 5' and 3' flanking the ICL; and on substrate CL-1, the ICL is located at the junction and XPF-ERCC1 predicted incisions are 5' to the ICL. M = radiolabelled oligonucleotides of the sizes indicated. M = radiolabelled oligonucleotides of the sizes indicated.

3.5 XPF-ERCC1 activity on substrates that model replication fork convergence onto an ICL

The replication fork convergence model proposed that an X-shaped DNA structure is an essential trigger for ICL incisions and unhooking, most likely by XPF-ERCC1 (Klein Douwel, et al. 2014; Raschle, et al. 2008; Zhang, et al. 2015). To test this hypothesis in a simple *in vitro* system, I therefore designed an X-shaped DNA substrate where an ICL is surrounded by four ssDNA arms. The ssDNA arms on the ICL are 23-mer in length. I also annealed oligonucleotide(s) complementary to the two 3'-ssDNA arms to model nascent leading strands striking the ICL from either the left or the right of the ICL, and also with both leading strands converging onto the ICL. I have determined that the model nascent leading strand are annealed to the X-shaped substrates using 10% non-denaturing gel as previously described in Section 3.2 (data not shown). CL-1 substrate which resembles a single replication fork collision with an ICL was used as a positive control for assays using these X-shaped substrates.

3.5.1 X-shaped structures is not a substrate for XPF-ERCC1.

Intriguingly, XPF-ERCC1 activity is not detectable on any of the X-shaped substrates, in the presence of absence of model nascent leading strands (**Figure 3. 14**). On CL-1 substrate however, XPF-ERCC1 incision is detectable but its activity is reduced in the presence of a model nascent leading strand, as previously described in **Figure 3. 11**. This data indicates that an X-shaped structure is not a preferred substrate of XPF-ERCC1 which suggests that the endonuclease requires regions of dsDNA for it to exert its activity.

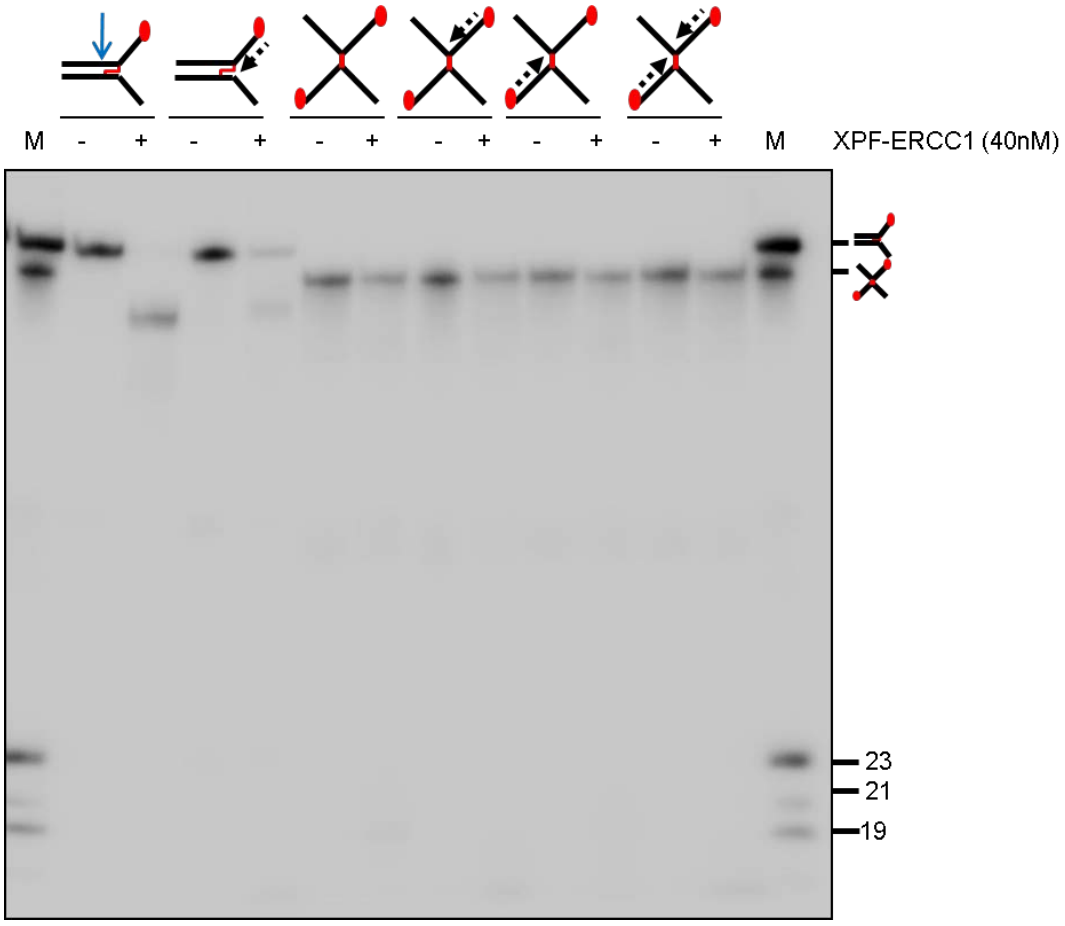


Figure 3. 14 X-shaped structures is not a substrate for XPF-ERCC1.

An array of 3'-radiolabelled X-shaped substrates (X-shaped; X-shaped with rightward model nascent leading strand; X-shaped with leftward model nascent leading strand; and X-shaped with both model nascent leading strands) were reacted with 40 nM XPF-ERCC1 and analysed in 10% denaturing gel. CL-1 substrate which contain an ICL at the fork junction was used a positive control and as a model for single replication fork collision with an ICL. M = radiolabelled oligonucleotide of indicated structures and sizes.

3.6 Discussion

XPF-ERCC1 deficiency causes high sensitivity to ICL-inducing agents *in vivo* suggesting a role for the endonuclease in ICL repair (De Silva, et al. 2000). Experimental evidence from cell-free replication-coupled repair system in *Xenopus laevis* egg extract illustrates that depletion of XPF, not other nucleases previously implicated in ICL repair namely Mus81 and FAN1, abrogates ICL repair, further confirming the importance of the endonuclease in ICL repair (Klein Douwel, et al. 2014).

Models derived from studies in the cell-free replication system propose that replication fork convergence onto an ICL, producing an X-shaped DNA structure is an essential requirement to trigger ICL incision, possibly by XPF-ERCC1 (Klein Douwel, et al. 2014; Raschle, et al. 2008; Zhang, et al. 2015). Additionally, the progression of a nascent leading strand to an ICL initiate incisions on the opposite template strand. While this discovery are the foundation of the most prominent model for replication-coupled ICL repair mechanism, questions arise as to whether replication fork convergence onto an ICL is also an absolute requirement for ICL repair in mammalian cells. This is because it is also possible that only one replication fork will encounter the ICLs (reviewed in (Zhang and Walter 2014)). The approach that we have taken to address this issue is *in vitro* reconstitution assay of replication fork incisions using human recombinant XPF-ERCC1 against synthetic DNA substrates that mimic the possible structures that arise when a single replication fork encounters and ICL or when two replication forks converges onto an ICL.

In agreement with previous studies, I showed that XPF-ERCC1 incises a model simple fork substrate at two locations within the duplex regions, 2 nt and 6 nt from the fork junction. XPF-ERCC1 activity is higher on a fork substrate compared to a 3'-flap (**Figure 3. 3**). Intriguingly, when a model nascent leading strand is present on the fork substrate, XPF-ERCC1 activity is significantly reduced. Furthermore, its incision is shifted and focused to a single site 5' from the fork junction, with loss of incision 2 nt 5' from the fork junction. However, the presence of a lagging strand or both leading and lagging strands completely abrogates XPF-ERCC1 activity (**Figure 3. 6**).

My observations are in agreement with a recent model for XPF-ERCC1 binding to a fork substrate, generated from NMR titration assay and structural studies of XPF-ERCC1 C-terminal domains heterodimer (Das, et al. 2012). The model suggests that XPF-ERCC1 recognises branched DNA substrates by the binding of each helix-hairpin-helix (HhH_2) domains of XPF and ERCC1 to each ssDNA arm; and the binding of the central domain of ERCC1 to the DNA duplex near the fork junction on the strand with protruding 5'ssDNA arm (Tsodikov, et al. 2005). The (HhH_2) domain on XPF specifically binds to the 3'-ssDNA arm, therefore positions the catalytic residue on the nuclease domain at the fork junction on the strand with the 3'-ssDNA arm for cleavage (Das, et al. 2012).

Therefore, my observation that XPF-ERCC1 activity is higher on a fork substrate compared to a 3'-flap substrate might be because ERCC1requires the dsDNA region and the 5'-ssDNA arm of a substrate. Binding of the complex to a fork substrate is possibly more stable compared to a 3'-flap substrate which does not possess a 5'-ssDNA arm, hence the increased level of activity on a fork substrate. The presence of a

leading strand on a fork substrate significantly reduced the activity of the complex possibly by occluding the 3'-ssDNA arm from binding of the (HhH_2) domain on XPF therefore its active residue could not be properly positioned onto the substrate. However, it is interesting that when a lagging strand which occludes the 5'-ssDNA arm is present, XPF-ERCC1 is completely abolished. Most possibly ERCC1 requires 5'-ssDNA arm for binding and the binding increases XPF-ERCC1 incisions, so eventhough a leading strand occludes the binding of XPF and interferes with the positioning of the active site on the incision sites, ERCC1 is still able to bind to the substrate. Therefore the protein complex activity is still functional in the presence of a leading strand on a fork substrate. It is only when the binding of ERCC1 to the substrate is inhibited as in the presence of a lagging strand that its activity is completely abrogated.

While the finding regarding the inhibitory effect of a lagging strand on a fork substrate provides a substantial insight into the structure-selectivity of XPF-ERCC1, a lagging strand is believed to stall 70-140 nt away from the replication fork during collision with an ICL *in vivo* (Klein Douwel, et al. 2014; Raschle, et al. 2008; Zhang, et al. 2015). Therefore I focused on pursuing further the effect of a leading strand because the gradual progression of a leading to an ICL is believed to be the essential trigger for ICL unhooking by XPF-ERCC1. By extending increasing length of a leading strand closer to the fork junction, it further confirms that XPF-ERCC1 is gradually reduced in the presence of a leading strand (**Figure 3. 9**).

In-vitro reconstitution assays showed that XPF-ERCC1 was able to sequentially create dual incisions flanking an ICL, termed 'ICL unhooking', when the ICL is placed near

the junction of a fork substrate (Kuraoka, et al. 2000). However, additional studies illustrate that XPF-ERCC1 can unhook an ICL, but the nuclease scaffold protein SLX4 is required to drive the second incision, which indicates that the second incision is not dependent on the presence of the 3'ssDNA arm (Hodskinson, et al. 2014). On a fork structure, the two XPF-ERCC1 incision products of XPF-ERCC1 are observed at equivalent intensity throughout a one-hour time course indicating that XPF-ERCC1 incises every DNA molecule at either one of the location, 6 nt or 2 nt from the junction (**Figure 3. 4**).

When an ICL is placed on the fork structure which models a single fork collision with the ICL, my data however indicates that the presence of an ICL near the fork junction inhibits XPF-ERCC1 incision closest to the junction (**Figure 3. 11**, **Figure 3. 12** and **Figure 3. 13**). Other investigators claimed that ICL unhooking by XPF-ERCC1 occurs very rapidly in their assays therefore the initial incision products were gel-purified and subjected with reaction to XPF-ERCC1 for the second time (Hodskinson, et al. 2014; Kuraoka, et al. 2000). In our assays, even after an hour of incubation XPF-ERCC1, the second cleavage 3' to the ICL is not observed on fork substrates containing an ICL close to the fork junction. This discrepancy may be due to the difference in helical distortion caused by one ICL compared to another because it is know that different ICL causes different distortion and bending on the DNA (Noll, et al. 2006). Furthermore, as observed on a model undamaged fork substrate, the presence of a leading strand inhibits XPF-ERCC1 activity on all ICL-containing substrates further confirming our finding that the presence of a leading strand inhibits XPF-ERCC1 activity (**Figure 3. 11**).

My analysis of the X-shaped DNA substrate that mimics replication forks convergence onto an ICL indicates that the structure is not a substrate for XPF-ERCC1 (**Figure 3. 14**) which is in line with crystallographic data of the protein complex that suggests ERCC1 binding to dsDNA region (Tsodikov, et al. 2005). The presence of a leading strand at either side of the X-structure or on both sides of structure also does not restore XPF-ERCC1 activity on the structure, contradictory to the replication fork convergence model which proposed that a nascent leading strand striking an ICL triggers ICL incisions. Therefore it is possible that for such structure, other factors are required to collaborate with XPF-ERCC1 to induce incision or it is also possible that the ssDNA arms surrounding the ICL may re-annealed prior to ICL incision by XPF-ERCC1, forming regions of dsDNA.

In this chapter, I determined that the presence of a leading strand on a fork substrate inhibits XPF-ERCC1 activity. I also determined that XPF-ERCC1 alone is not able to create 5’ and 3’ incisions flanking an ICL to unhook it from the duplex DNA. Furthermore, an X-shaped structure which mimics dual forks collision upon an ICL is not a preferred substrate of XPF-ERCC1. Taken together, my data suggests that XPF-ERCC1 alone does not have the capability to incise the DNA upon the arrival of a leading strand to the fork junction. This suggests that additional factors may be required to facilitate XPF-ERCC1 incisions or another endonuclease/exonuclease may be required to complete ICL repair. Our finding in this chapter leads to the question as to whether another factor is required to restore XPF-ERCC1 activity when a leading strand is present on a fork structure.

4 The stimulation on XPF-ERCC1 activity by the human replication protein A (RPA)

4.1 Introduction

The human single-stranded DNA (ssDNA)-binding replication protein A (RPA) is a heterotrimer composed of 70 kDa, 32 kDa and 14 kDa subunits (p70, p32, and p14, respectively) (Fairman and Stillman 1988). The involvement of RPA in various DNA metabolising events including DNA replication, recombination and repair are attributed to its two main properties: First, RPA high binding affinity to ssDNA stabilises the ssDNA and removes secondary structures from the ssDNA regions. Second, its interaction with various proteins modulates the proteins' activity during DNA metabolism. No mammalian mutant cells defective in RPA have been isolated, possibly due to its essential role in DNA replication and other key DNA metabolic processes

Stimulation of XPF-ERCC1 and XPG endonuclease activity by RPA was described in studies using DNA substrates that mimic an 'open bubble', a structure that arise during NER (Matsunaga, et al. 1996). During NER, DNA regions surrounding the lesion is unwind to form a bubble and the damaged DNA strand is cleaved by XPG on the 3' side of the lesion (O'Donovan, et al. 1994) and by ERCC1-XPF on the 5' side (Sijbers, et al. 1996). RPA was also reported to stimulate XPF-ERCC1 activity on a 5'-flap substrate but inhibit XPF-ERCC1 activity on a 3'-flap substrate (de Laat, et al. 1998b).

While physical interactions between XPG and RPA have been described, which suggest that RPA is recruited to the undamaged DNA to position XPG to site of damage (He, et al. 1995), direct interaction between XPF and RPA has also been reported (Bessho, et al. 1997). Furthermore, recombinant XPF-ERCC1 mutated at XPF-RPA interaction site revealed that the mutant complex has similar biochemical properties as the wild-type complex, however, the authors did not determine whether the mutation affect XPF-ERCC1 activity in the presence of RPA (Fisher, et al. 2011). RPA was also shown to enhance the binding of XPF-ERCC1 complex to a bubble substrate, which suggest that RPA may also influence targeting XPF-ERCC1 to the sites of DNA damage and repair (Matsunaga, et al. 1996).

The aim of this chapter is to elucidate the role of RPA on XPF-ERCC1 on DNA substrates that resemble the structures that arise at stalled replication fork. Stalling of replication forks by an inter-strand crosslink (ICL) give rise to ssDNA intermediates surrounding the lesion and the arrival of a nascent leading strand to the ICL is believed to trigger dual incisions flanking the ICL, possibly by XPF-ERCC1 (Klein Douwel, et al. 2014; Raschle, et al. 2008; Zhang, et al. 2015). Therefore the question arises as to whether RPA also interacts with XPF-ERCC1 in this DNA process.

My findings in Chapter 3 illustrated that XPF-ERCC1 activity is significantly reduced by the presence of a model leading strand on a DNA fork substrate. In this chapter, I investigated whether RPA modulates XPF-ERCC1 activity on the array of DNA fork substrates used in Chapter 3. I also explored the mechanism of RPA modulation of XPF-ERCC1 by determining RPA binding affinity to these substrates. My attempt to

purify the mutant XPF-ERCC1 complex reported to disrupt XPF interaction with RPA, an effort to determine the physical interaction between the proteins, is also described.

4.2 The effects of RPA on XPF-ERCC1 activity on fork substrates.

4.2.1 RPA stimulates XPF-ERCC1 activity on a fork substrate

To explore the effect of RPA on XPF-ERCC1 activity, a DNA fork substrate was pre-incubated with increasing concentration of recombinant human RPA on ice for 10 minutes prior to reaction with 40 nM XPF-ERCC1 for 1 hour at 30^{O}C. Increasing substrate incision was observed with increasing concentration of RPA, reaching maximal incisions with 80 nM RPA, two-fold the concentration of XPF-ERCC1. Furthermore, the presence of RPA shifted and focuses XPF-ERCC1 incision to a single site further into the duplex (5') away from the fork junction, as seen in the higher level of incision 6 nt from the junction (giving a 29-mer product) and loss of incision 2 nt from the junction (25-mer product) (**Figure 4. 1**a and b).

When the same assay was conducted using single-strand DNA-binding protein (SSB), a homolog of RPA from *E. coli*, XPF-ERCC1 was inhibited by SSB which suggests that the stimulatory effect on XPF-ERCC1 is specific for RPA (**Figure 4. 1**c). Furthermore, inhibition of XPF-ERCC1 by SSB may be attributed to tighter association or non-specific binding of SSB to ssDNA therefore occluding the region of the DNA substrate recognised by XPF-ERCC1. Indeed, in the pilot experiment containing the DNA substrate in the reaction, SSB were not able to migrate from the wells into the gel (data

not shown). Only upon incubation with 10 µg Proteinase K for 30 minutes at 37^{o}C the DNA substrates were able to migrate into the gel, comparable to migration of DNA substrates reacted with RPA. This confirms the tight association of SSB with the DNA substrate likely producing an aggregate which can only be disrupted with a protease. Taken together, this data signifies that the stimulation of XPF-ERCC1 is specific for RPA, which may be attributed to RPA binding to the DNA substrate, and also possibly by the direct interaction between XPF-ERCC1 and RPA.

4.2.2 RPA does not restore the activity of the nuclease-defective XPF-ERCC1

After confirming that RPA stimulates XPF-ERCC1 activity on a fork substrate, I determined whether RPA is able to restore the endonuclease activity of the nuclease-defective mutant, XPF(D676A)-ERCC1. A fork substrate was reacted with wild-type (lane 2 - 4) or D676A (lane 5 – 7) XPF-ERCC1 with and without 80 nM RPA. Endonuclease activity was detected for wild-type XPF-ERCC1 (lane 3), and its activity is stimulated in the presence of RPA (lane 4). Activity was not detected for the D676A mutant, with or without RPA (lane 6, 7). This illustrates that the RPA-stimulated endonuclease activity of the complex is inherent to XPF, and the presence of RPA does not produce change in active site architecture capable of reactivating the catalytic capacity of the mutant XPF or confer nuclease activity onto ERCC1 (**Figure 4. 2**).

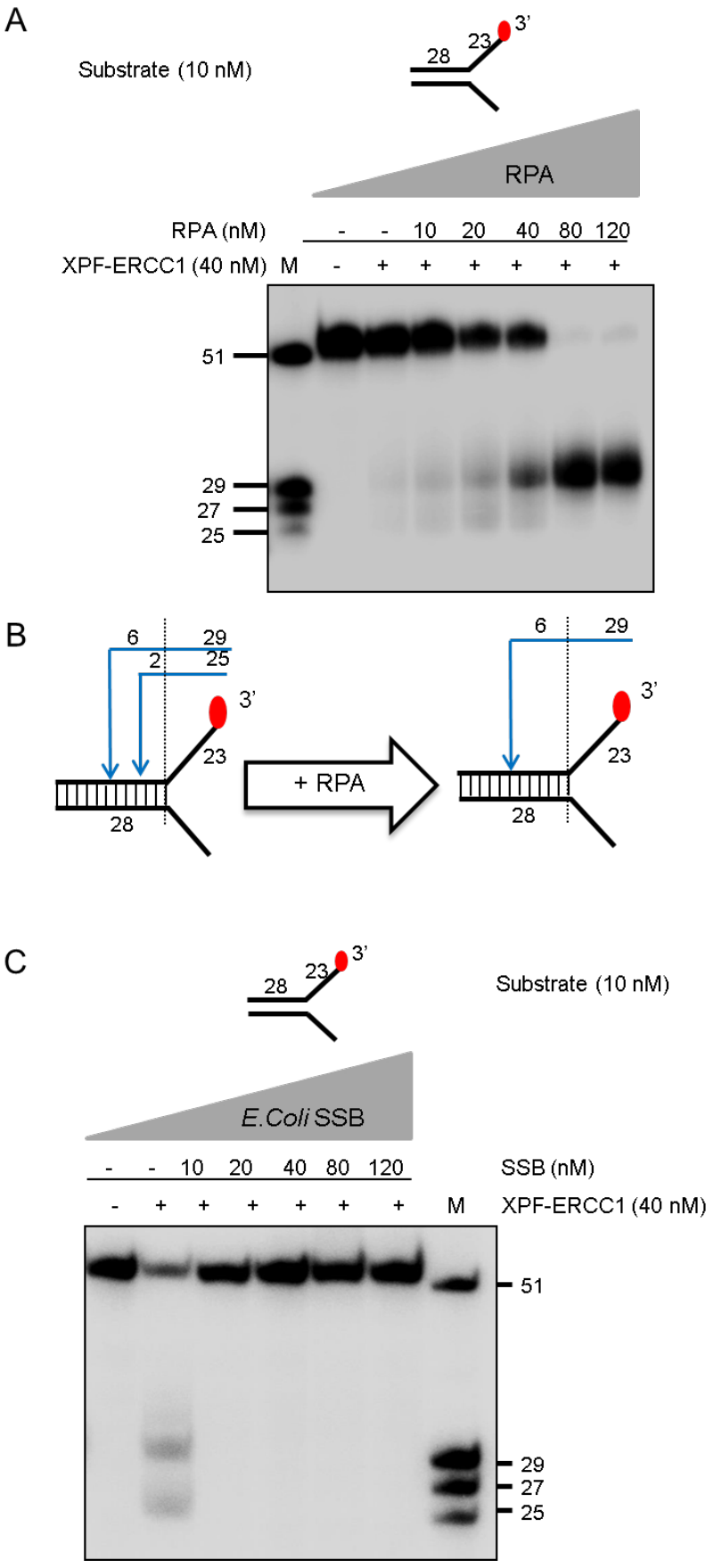

A
Substrate (10 nM)
3'
23
28
RPA
RPA (nM) - - 10 20 40 80 120
XPF-ERCC1 (40 nM) M - + + + + + +
51
29
27
25
B
6
29
2
25
3'
23
28
+ RPA
6
29
3'
23
28
C
3'
23
28
Substrate (10 nM)
E.Coli SSB
- - 10 20 40 80 120
SSB (nM)
- + + + + + + M
XPF-ERCC1 (40 nM)
51
29
27
25

Figure 4. 1: RPA stimulates XPF-ERCC1 activity on a fork substrate.

A)10 nM simple3'-radiolabelled fork substrate was incubated with increasing concentration (0, 10, 20, 40, 80, 120 nM) of human RPA on ice for 10 minutes, and reacted with 40 nM XPF-ERCC1 for 1 hour at 30^{O}C. B) Schematic representation of the shift in XPF-ERCC1 incision sites after incubation with RPA. C) Nuclease reactions as in A were performed using *E. coli* SSB. After reaction with XPF-ERCC1, 1 µL 10 µg/µL Proteinase K was added and the reaction is further incubated for 30 minutes at 37^{O}C. M = radiolabelled oligonucleotides of the sizes indicated.

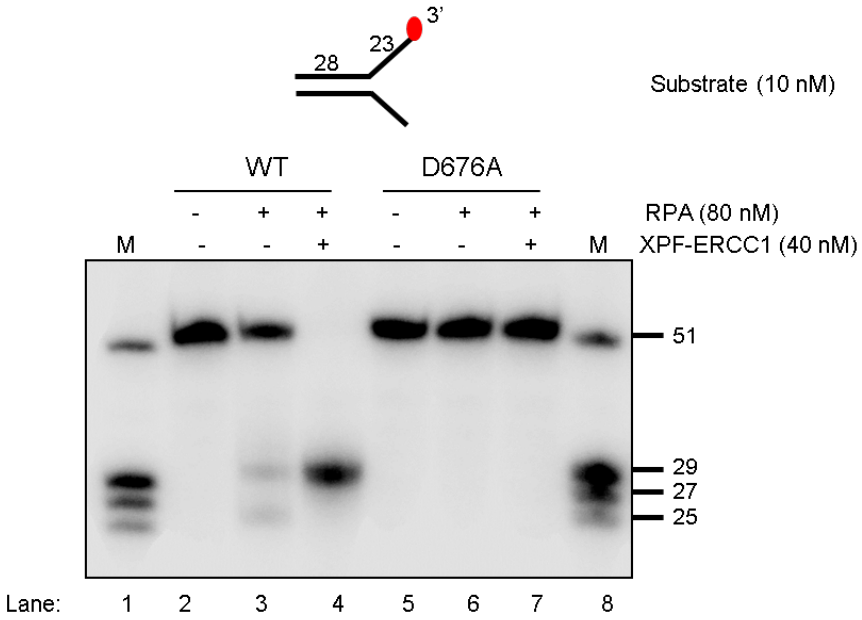


Figure 4. 2: RPA does not restore the activity of the nuclease-defective, XPF(D676A)-ERCC1 mutant.

3'-radiolabelled fork substrate was reacted with wild-type XPF-ERCC1 (lane 2 – 4) or D676A mutant (lane 5 – 7), with or without 80 nM RPA. M = radiolabelled oligonucleotides of the sizes indicated.

4.2.3 RPA restores XPF-ERCC1 activity inhibited by a model nascent leading strand on a fork substrate

In Chapter 3, I determined that when a model nascent leading strand extends closer to a fork junction, XPF-ERCC1 activity gradually decreases with almost complete inhibition when the leading strand is at the fork junction (-1 nt location). I wished to determine whether RPA would be able to restore XPF-ERCC1 activity on the fork substrates with increasing length of nascent model leading strands (fork, -10 nt, -4 nt, -3 nt, -2 nt, -1 nt from the fork junction), previously described (**Figure 3. 9**). The array of substrates was incubated with or without RPA, prior to reaction with XPF-ERCC1. Intriguingly, in the presence of RPA, XPF-ERCC1 activity is gradually restored on these substrates, confirming that RPA is not only able to stimulate XPF-ERCC1 activity on a fork substrate, it is also required to activate XPF-ERCC1 activity inhibited by the presence of a leading strand. RPA restores the incision further into the duplex (5') 6 nt from the fork junction (giving a 29-mer product), but not the incision site 2 nt from the junction (25-mer product) (**Figure 4. 3**a and b).

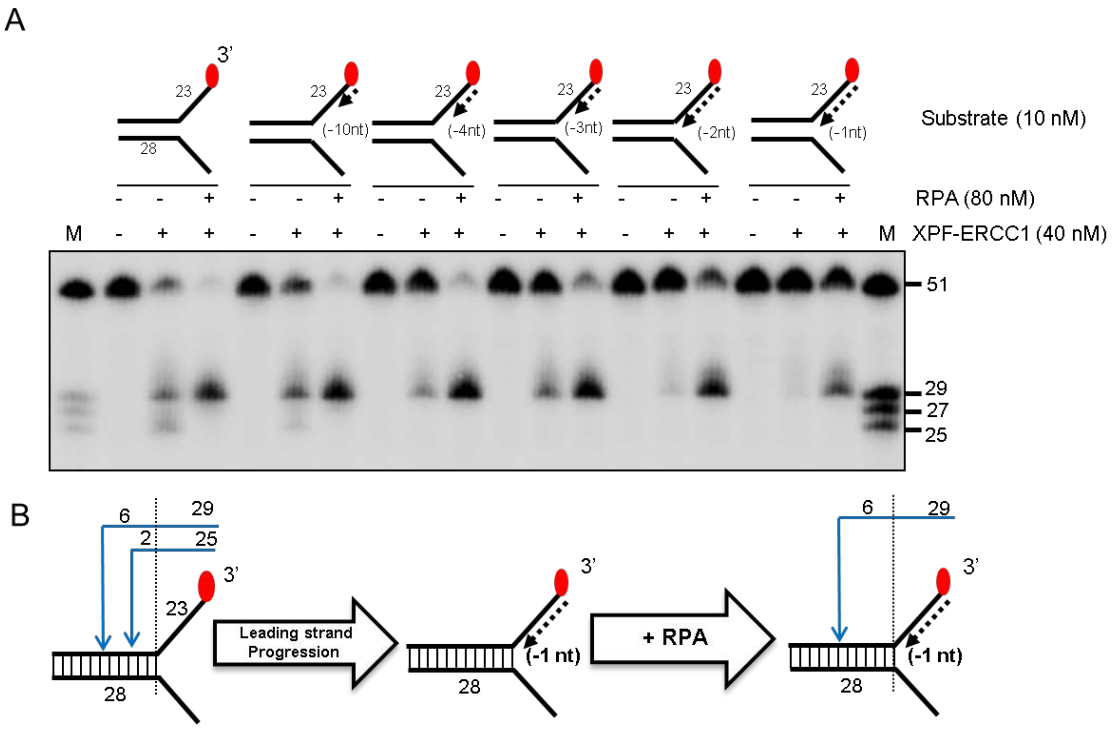


Figure 4. 3: The gradual inhibition of XPF-ERCC1 activity when the leading strand extends closer to the fork junction is overcome by RPA.

A) 3'-radiolabelled fork substrates with increasing length of model nascent leading strand (fork, -10 nt, -4 nt, -3 nt, -2 nt, -1 nt from the fork junction) were incubated with/without 80 nM RPA on ice prior to reaction with 40 nM XPF-ERCC1 for 1 hour at 30^0C. M = radiolabelled oligonucleotides of the sizes indicated. B) Schematic representation which show that XPF-ERCC1 incisions are inhibited by the progression of a model nascent leading strands, which is then restored by the presence of RPA.

4.2.4 RPA does not restore XPF-ERCC1 activity inhibited by a model nascent lagging strand on a fork substrate

After confirming that RPA is able to activate XPF-ERCC1 incisions inhibited by the presence of a model nascent leading strand on a fork substrate, I then asked whether the apparent re-activation of XPF-ERCC1 by RPA is specific for a fork substrate with a leading strand or also for fork substrates with a lagging strand or both leading and lagging strands. The array of fork substrates (simple fork, +leading strand, +lagging strand, replication fork) as previously described in Chapter 3 was reacted with XPF-ERCC1 in the presence of RPA. As shown earlier, RPA stimulated and shifted XPF-ERCC1 incision further into (5') the duplex on a simple fork substrate; and inhibition of XPF-ERCC1 activity by a leading strand on a fork substrate is restored by RPA. Here I determined that RPA fails to restore XPF-ERCC1 activity on fork substrate with a model nascent lagging strand or both model nascent leading and lagging strands (**Figure 4. 4**). This data further illustrates that RPA is required to re-activate XPF-ERCC1 endonuclease activity specifically on a fork structure with a model nascent leading strand. The presence a lagging strand or both leading and lagging strand on a fork substrate however, may occlude the binding of RPA to the substrate therefore preventing restoration of XPF-ERCC1 activity on these substrates. My data also raises the question as to whether RPA has a specific binding preference with respect to the polarity of the ssDNA region for binding.

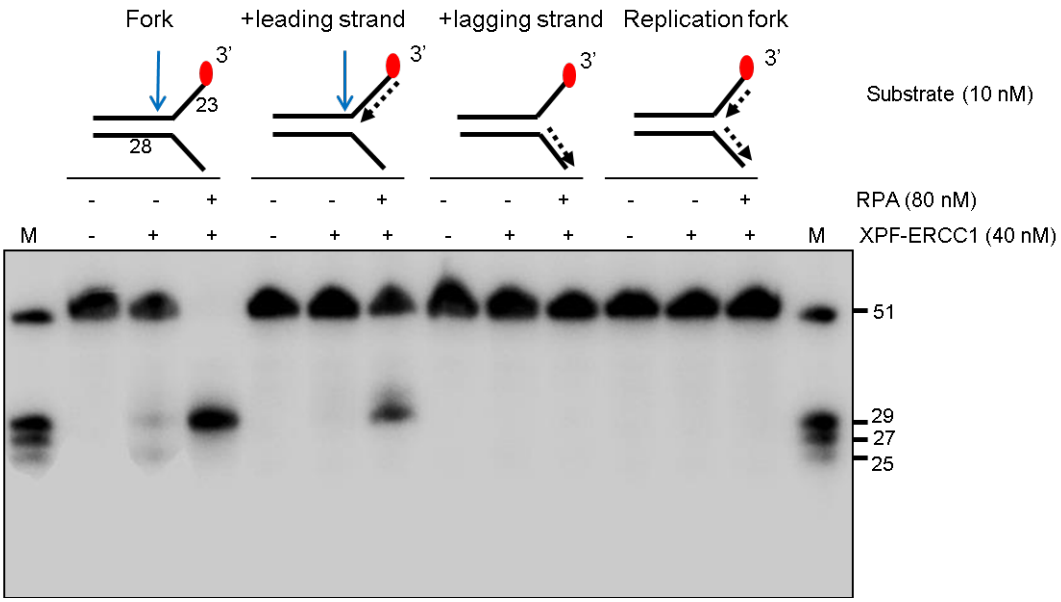


Figure 4. 4: RPA does not restore XPF-ERCC1 activity inhibited by a model nascent lagging strand on a fork substrate.

An array of 3'-radiolabelled fork substrates (fork, +leading strand, +lagging strand, replication fork) was incubated with/without 80 nM RPA on ice prior to reaction with 40 nM XPF-ERCC1 for 1 hour at 30^0C. M = radiolabelled oligonucleotides of indicated sizes.

4.3 RPA binding affinity to DNA fork substrates

In light of my finding that RPA is able to restore XPF-ERCC1 incisions inhibited by a model nascent leading strand on a fork substrate, I explored the mechanism of the stimulation by first determining the binding affinity of RPA to the array of DNA fork substrates used in this study. It has been suggested that RPA binding to DNA involves a particular molecular polarity because RPA is shown to preferentially binds to 3'-flap compared to 5'-flap substrate (de Laat, et al. 1998b). Furthermore, it is has been proposed that the initial hRPA binding occurs on the 5' side of a ssDNA substrate, and then extends in the 3' direction to create a stably bound hRPA (Iftode and Borowiec 2000).

To investigate the binding polarity of RPA to the DNA and how this reflects on its stimulation of XPF-ERCC1, I utilised the enzyme mobility shift assay (EMSA) to answer this question. An EMSA is utilised to determine whether a protein interacts with a radiolabelled DNA substrate observable by the less mobile protein-DNA complex compared to the DNA substrate alone when analysed on non-denaturing polyacrylamide gels (PAGE).

4.3.1 Preferential binding of RPA to 3'-ssDNA arm of fork substrates

To enable the detection of protein-DNA complex in an EMSA, the divalent cation manganese (Mn^{2+}) was replaced with calcium (Ca^{2+}), which is an inhibitor to many DNA metabolising enzymes (Mordasini, et al. 2003). By inhibiting the nuclease activity of XPF-ERCC1, the binding of the protein to the DNA substrate can be properly analysed (de Laat, et al. 1998b). The same array of fork substrates used elsewhere (fork;

+leading strand; +lagging strand; replication fork) was incubated with increasing concentration of RPA. Binding of RPA to a fork substrate is detectable with 80 nM RPA, as seen in the shift of the radiolabelled DNA substrate in the non-denaturing gel (**Figure 4. 5**). This data complements our nuclease assay in **Figure 4. 1** which shows maximal stimulation on XPF-ERCC1 activity is at 80 nM RPA. However, stable RPA binding to the fork substrate with a model nascent leading strand is not detectable even though RPA stimulates XPF-ERCC1 activity on this structure as shown in **Figure 4. 1** and **Figure 4. 4**. Interestingly, RPA binding was observed on the fork substrate with model nascent lagging strand even though XPF-ERCC1 activity is abrogated on this structure and its activity is not restored by RPA. Binding was also not observed on the fork substrate with both leading and lagging strand. This data suggests that RPA has a preferential binding affinity to 3'-ssDNA arms of a fork substrate, which is occluded by the presence of a model nascent leading strand.

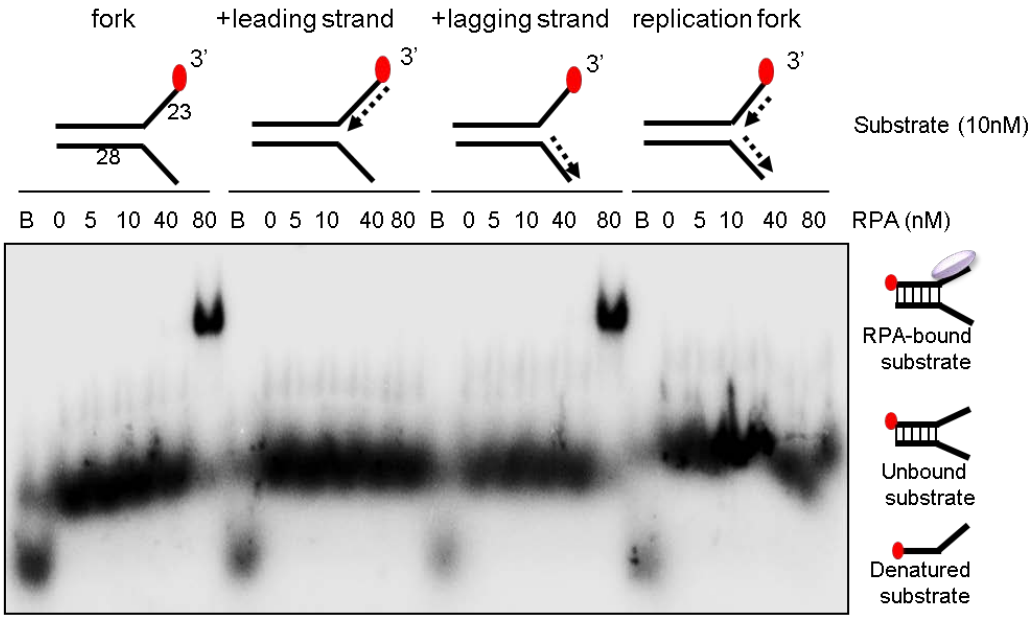


Figure 4. 5: Preferential binding of RPA to 3'-ssDNA arm of a fork substrate.

An array of 3'-radiolabelled fork substrates (fork; +leading strand; +lagging strand; replication fork) were incubated with increasing concentration of RPA (0, 5, 10, 40, 80 nM) in the presence of 5 mM calcium (Ca^{2+}). B = the indicated substrate boiled at $95^{O}C$ for 5 minutes to denature the substrate. Reaction was analysed in a 5 % non-denaturing PAGE gel.

4.3.2 RPA binding to a fork substrate is disrupted by a model nascent leading strand of any length

RPA has been shown to bind to a 8 nt ssDNA (Blackwell, et al. 1996), in line with several studies that suggest 8-10 nt is required as the precursor of the more stable 30 nt binding of RPA (Blackwell and Borowiec 1994). Since RPA binding to a fork substrate is shown to be disrupted by a model nascent leading strand in the previous figure, I then investigated the minimal length of 3'-ssDNA arm of a fork substrate required for RPA binding. Using the array of fork substrates with increasing length of model nascent leading strand (the 3'-ssDNA length not occluded by a model nascent leading strand on these array of substrates are 22 nt, 9 nt, 3 nt, 2 nt, 1 nt and 0 nt corresponding to the fork, -10 nt, -4 nt, -3 nt, -2 nt and -1 nt substrates, respectively), I determined that RPA binding to the fork substrate is disrupted by the presence of a model nascent leading strand of any length (**Figure 4. 6**). This finding further confirms that RPA may have a strong binding affinity for 3'-ssDNA arms of a fork substrate.

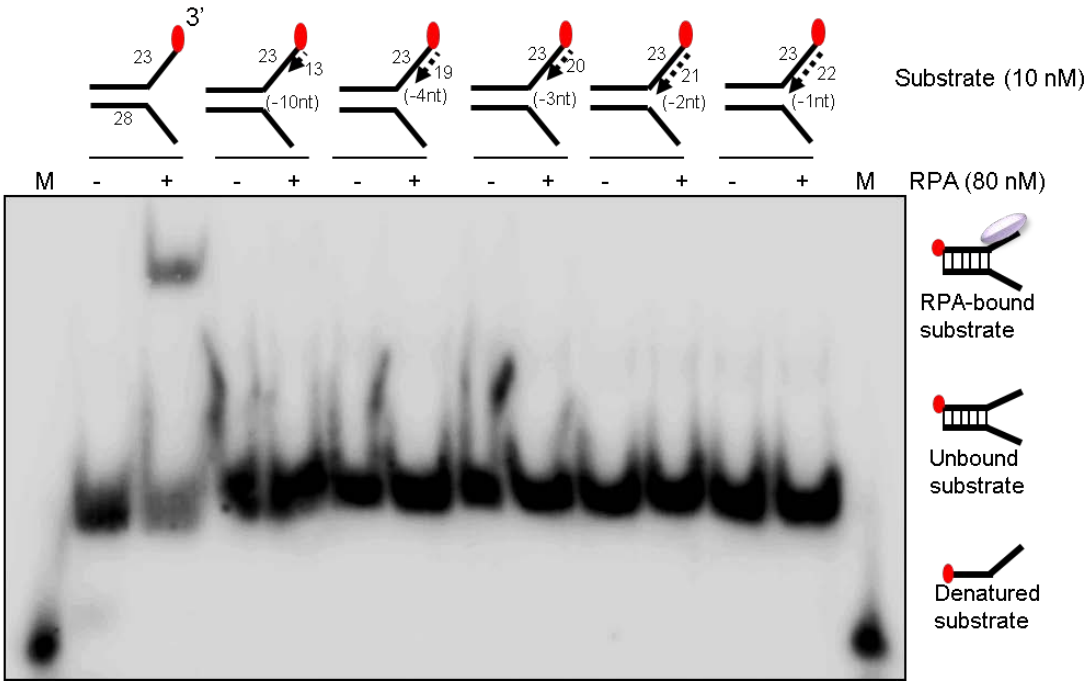


Figure 4. 6: RPA binding to a fork substrate is disrupted by the presence of a model nascent leading strand of any length.

XPF-ERCC1 were incubated with an array of 3'-radiolabelled fork substrates with increasing length of model nascent leading strand (fork, -10 nt, -4 nt, -3 nt, -2 nt and -1 nt with un-occluded 3'-ssDNA lengths of 22 nt, 9 nt, 3 nt, 2 nt, 1 nt and 0 nt, respectively). M = radiolabelled oligonucleotide of the indicated structure. The reaction was analysed in a 5 % non-denaturing PAGE gel.

4.4 Analysis of RPA/XPF-ERCC1 complex formation on DNA fork substrates

Ternary complex formation between a radiolabelled DNA substrate and two proteins can be detected in an EMSA as a slower migrating ternary complex, coined as a 'super-shift', compared to the migration of a DNA-protein complex in a PAGE gel. To determine whether ternary complex is formed between the DNA substrates, XPF-ERCC1 and RPA which will indicate whether the two proteins physical interact with the DNA substrates simultaneously, the array of fork substrates was incubated with either XPF-ERCC1 alone, or RPA alone or RPA and XPF-ERCC1 (**Figure 4. 7**).

XPF-ERCC1 binding is not detectable on any of the fork substrates in the array (lane 2, 6, 10, and 14) possibly because it is transiently recruited to the DNA substrates. RPA binding previously detectable at simple fork and fork with a model nascent lagging strand in **Figure 4. 5** was also observed in this assay (lane 3, 7, 11, 15). For assessment of ternary complex formation, a faint super-shift band above the band shift for RPA-bound substrate is detectable for a simple fork substrate indicating that a small proportion of the DNA substrates are bound by both RPA and XPF-ERCC1 (lane 4). A faint super-shift band is also observed on a fork substrate with a nascent model leading strand (lane 8), however the super-shift band is not detectable on a fork with a model nascent lagging strand (lane 12) and a fork with both model nascent leading and lagging strand (lane 16).

The supershift bands indicating ternary complex formation on a fork substrate and a fork with a model nascent leading strands are in line with my observation that RPA stimulates XPF-ERCC1 activity on these substrates. However, the observed super-shift bands are faint, therefore additional evidence is required to confirm the possible transient interaction of the DNA substrates with the two proteins. Nevertheless, this result provides evidence that RPA and XPF-ERCC1 may physically interact on the DNA fork substrates and together able to efficiently cleave DNA substrates.

The same assay was also conducted using fork substrates with increasing length of model nascent leading strand. Here I determined that while RPA binding is only observed on a fork substrate, and not in the presence of a model nascent leading strand of any length as previously described in **Figure 4. 6**. However, a faint band corresponding to a super-shift is observed on all of the substrates, which indicate that a small proportion of the substrates were bound with both XPF-ERCC1 and RPA simultaneously. This data complements our nuclease assays which illustrate that RPA stimulates XPF-ERCC1 activity on these substrates.

My findings indicate that RPA has a preference for binding to the 3'-ssDNA arm of a fork substrate, which is blocked by the presence of a model nascent leading strand. This data suggests two possibilities: 1) the 5'-ssDNA is required for RPA stimulation of XPF-ERCC1 even though RPA binding to it is not detectable in an EMSA 2) RPA stimulates XPF-ERCC1 activity via direct physical interaction.

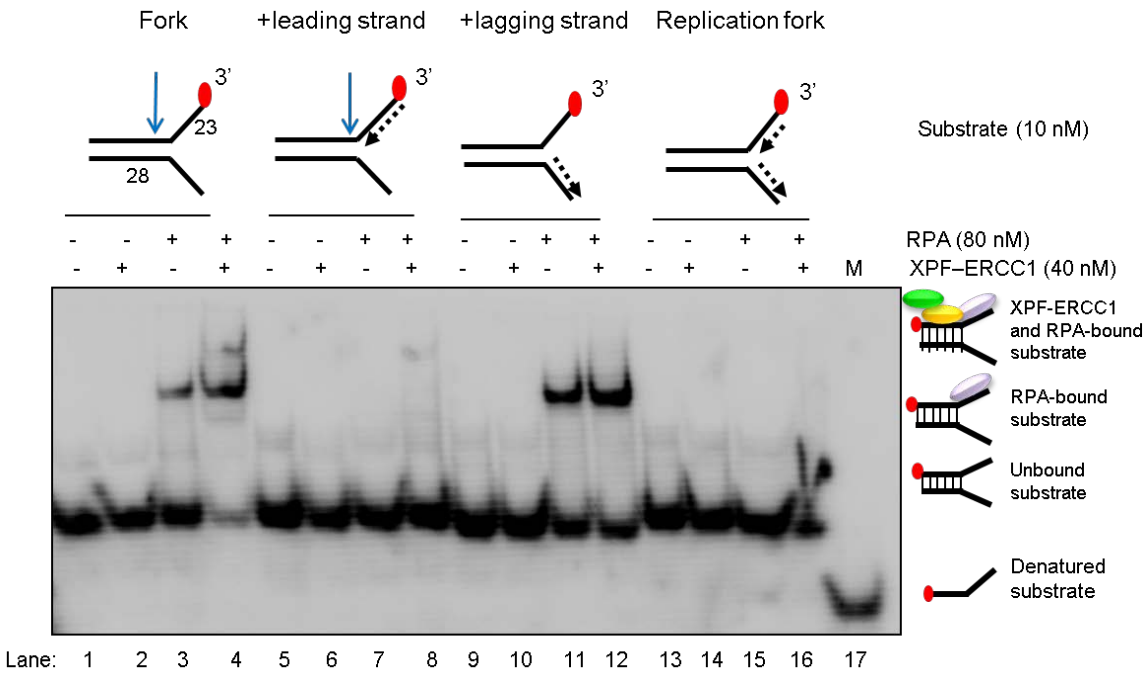


Figure 4. 7: Analysis of ternary complex formation between DNA substrates, RPA and XPF-ERCC1 of an array of fork substrates.

An array of 3'-radiolabelled fork substrates (fork; +leading strand; +lagging strands; replication fork) were incubated with either XPF-ERCC1 alone (lane 2, 6, 10, and 14) or RPA alone (3, 7, 11, 15) or XPF-ERCC1 and RPA (lane 4, 8, 12, 16). M = radiolabelled oligonucleotide of the structure indicated. Reaction was analysed in a 5 % non-denaturing PAGE gel.

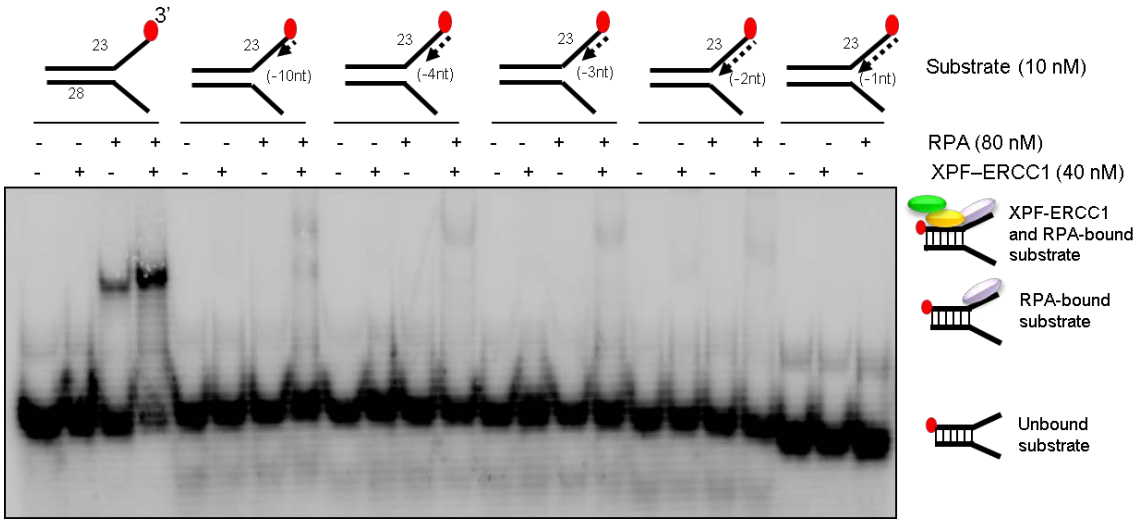


Figure 4. 8: Analysis of ternary complex formation between DNA substrates, XPF-ERCC1, RPA using fork substrates with increasing length of model nascent leading strand.

3'-radiolabelled fork substrates with increasing length of a model nascent leading strand (fork, -10 nt, -4 nt, -3 nt, -2 nt, and -1 nt) were incubated with either XPF-ERCC1 alone, RPA alone, or XPF-ERCC1 and RPA. Reaction was analysed in a 5 % non-denaturing PAGE gel. Note: The reaction for XPF-ERCC1-RPA on the fork with -1 nt model nascent leading strand is not present in the figure.

4.5 Determination of the role of the 5'-ssDNA arm of a fork substrate for RPA stimulation of XPF-ERCC1

Using competition binding assay, it was determined that RPA binding affinity to RNA is 4-orders of magnitude lower than to DNA (Brill and Stillman 1989). Therefore, to test the first possibility namely whether the 5'-ssDNA arm of a fork substrate is required for RPA stimulation of XP-ERCC1 (even though RPA binding to it is not detectable in an EMSA), I replaced the 5'-ssDNA with a 5'ssRNA, denoted as green lines in **Figure 4. 9**. In these experiments, XPF-ERCC1 activity is tested on the DNA and DNA:RNA substrates, with and without 80 nM RPA. On a DNA fork substrate, RPA stimulates and shifted XPF-ERCC1 incision further into the duplex (lane 2 – 4), as previously described in **Figure 4. 4**. On a DNA:RNA hybrid fork substrate however, XPF-ERCC1 activity is reduced, and the presence of RPA abolishes XPF-ERCC1 activity (lane 5 – 7). On a DNA fork with a model nascent leading strand, XPF-ERCC1 activity inhibited by the leading strand is restored by RPA (lane 8 – 10), as described in Figure 4. 4; however XPF-ERCC1 activity is not restored on a DNA:RNA hybrid fork substrate with a model nascent leading strand (lane 11 – 13). Taken together, this data suggests that the presence of a 5'-ssRNA arm itself has an minor inhibitory effect on XPF-ERCC1 activity as seen in reduced XPF-ERCC1 on the DNA:RNA hybrid fork (lane 6). Furthermore, the 5'-ssRNA also inhibit the stimulation by RPA which indicates that RPA requires the 5'ssDNA arm for its stimulation of XPF-ERCC1, eventhough its binding to the 5'-ssDNA arm is not detectable in an EMSA (**Figure 4. 9**).

Additionally, I compared RPA binding to the array of DNA fork substrates (fork; +leading strand; +lagging strand; replication fork) with the same array of fork substrates but the 5'-ssDNA arm is replaced with 5'-ssRNA. RPA binding profile on the DNA:RNA hybrid substrates is the same as on DNA substrates: RPA binding is detectable on fork substrates but not detectable on fork with a model nascent leading strand (**Figure 4. 10**).

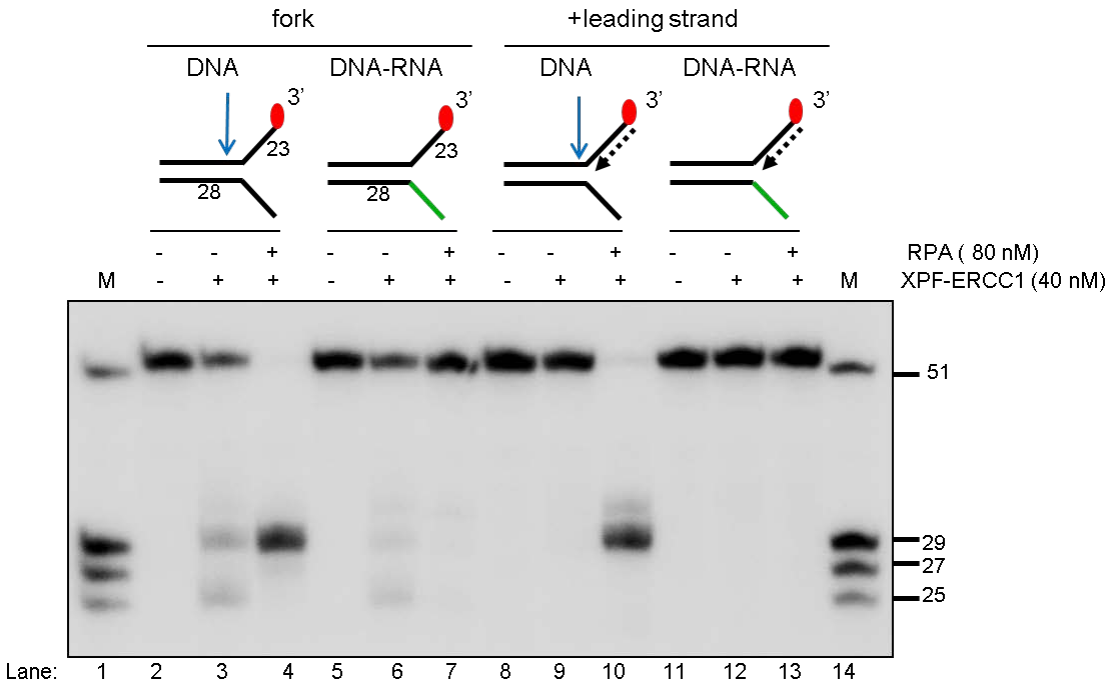


Figure 4. 9: Comparison between XPF-ERCC1 activity on a DNA and a DNA:RNA hybrid fork substrates.

3'-radiolabelled DNA substrates (lane 2 – 4 and 11 – 13) or DNA:RNA hybrid substrates (lane 5 – 7 and 11 – 13) are reacted with XPF-ERCC1 with or without RPA. DNA is represented as black lines and RNA is green lines. M = radiolabelled oligonucleotides of the sizes indicated.

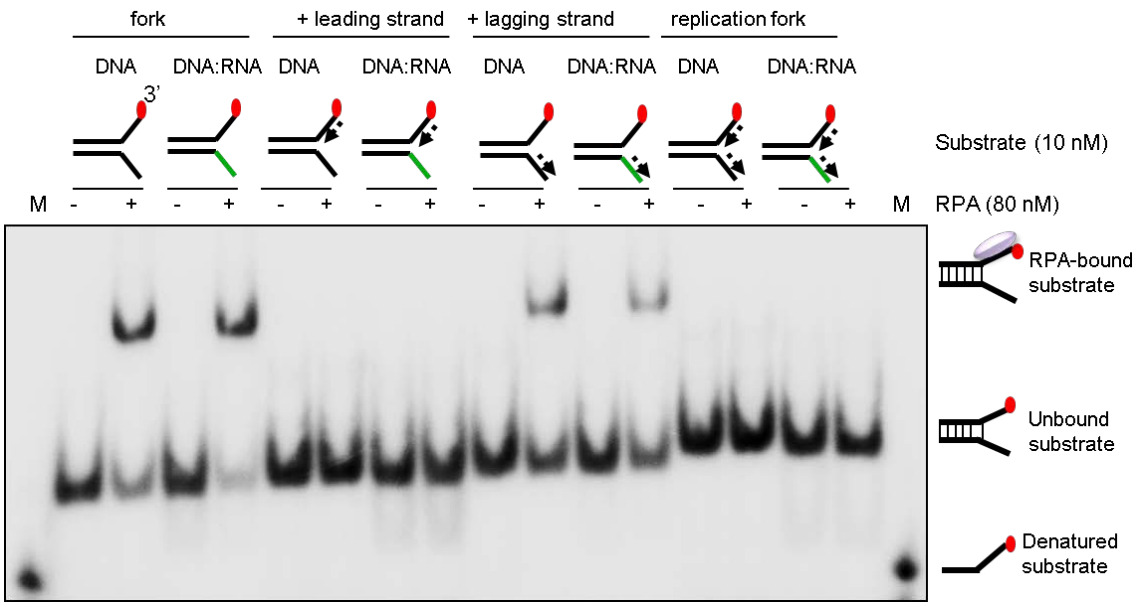


Figure 4. 10: Comparison between RPA binding to a DNA and a DNA:RNA hybrid array of fork substrates.

An array of 3'-radiolabelled fork substrates (fork; +leading strand; +lagging strand; replication fork), either as DNA or DNA:RNA hybrid were incubated with 80 nM RPA. DNA is represented as black lines and RNA is green lines. M = radiolabelled oligonucleotides of the structure indicated.

4.6 Analysis of XPF-ERCC1 interaction with RPA using a XPF-ERCC1 mutated at the site of interaction with RPA

To test the second possibility, which is whether RPA stimulates XPF-ERCC1 activity via direct physical interaction in absence of binding to the 3'-ssDNA arm when a model nascent leading strand is present, I therefore purified a mutant XPF-ERCC1 complex reported to disrupt the interaction of the complex with RPA.

A residue on XPF for RPA interaction has been determined in a reverse yeast two-hybrid assay (Fisher, et al. 2011). In the assay, plasmids containing XPF cDNA were transformed into a mutator *E. coli* strain XL-1 Red which introduces random mutations within the plasmid. The mutated plasmids were then co-transformed with plasmids containing wild-type p70 RPA in a yeast strain, which enable the analysis of growth of colonies in a selective medium. Colonies that failed to grow indicated a lack of interaction between the mutated XPF and p70 RPA.

Out of twelve colonies that failed to grow in the selective medium, they discovered that one colony contain a single amino acid change, proline-85 to serine (P85S). While the authors have purified the XPF(P85S)-ERCC1 complex and determined that the mutant complex has similar biochemical properties as the wild-type complex, they did not determine whether the mutant complex fails to be stimulated by RPA. Therefore, to analyse the effect of the mutation on XPF-ERCC1 activity in the presence of RPA, I purified this mutant protein using the same procedure for the purification of wild-type XPF-ERCC1 used in this study.

The expression of XPF(P85S) and ERCC1 (**Figure 4. 11**a) is comparable to the expression of the wild-type and D676A complex (data not shown). However, unlike the wild-type and D676A mutant complex, where equivalent amount XPF and ERCC1 were eluted from the nickel beads with 250 mM imidazole (data not shown), for the P85S mutant complex, the amount of XPF eluted is significantly less than ERCC1. Furthermore, a distinct peak for XPF(P85S)-ERCC1 complex is not present in the gel filtration profile (**Figure 4. 11**b), unlike for wild-type and D676A mutant where the peak is present at ~14 mL (**Figure 3. 1**a). This suggest that the XPF(P85S) is not stable – it is either aggregated or degraded therefore it is not able to interact with ERCC1 for the elution. The fraction from gel filtration for the predicted XPF(P85S)-ERCC1 complex was pooled and I have determined the fractions contain XPF and ERCC1 using Western blotting (data not shown). I have yet to test the activity of the purified mutant complex in the near future. I wished to modify my purification strategy to increase the yield for the P85S mutant complex in the near future.

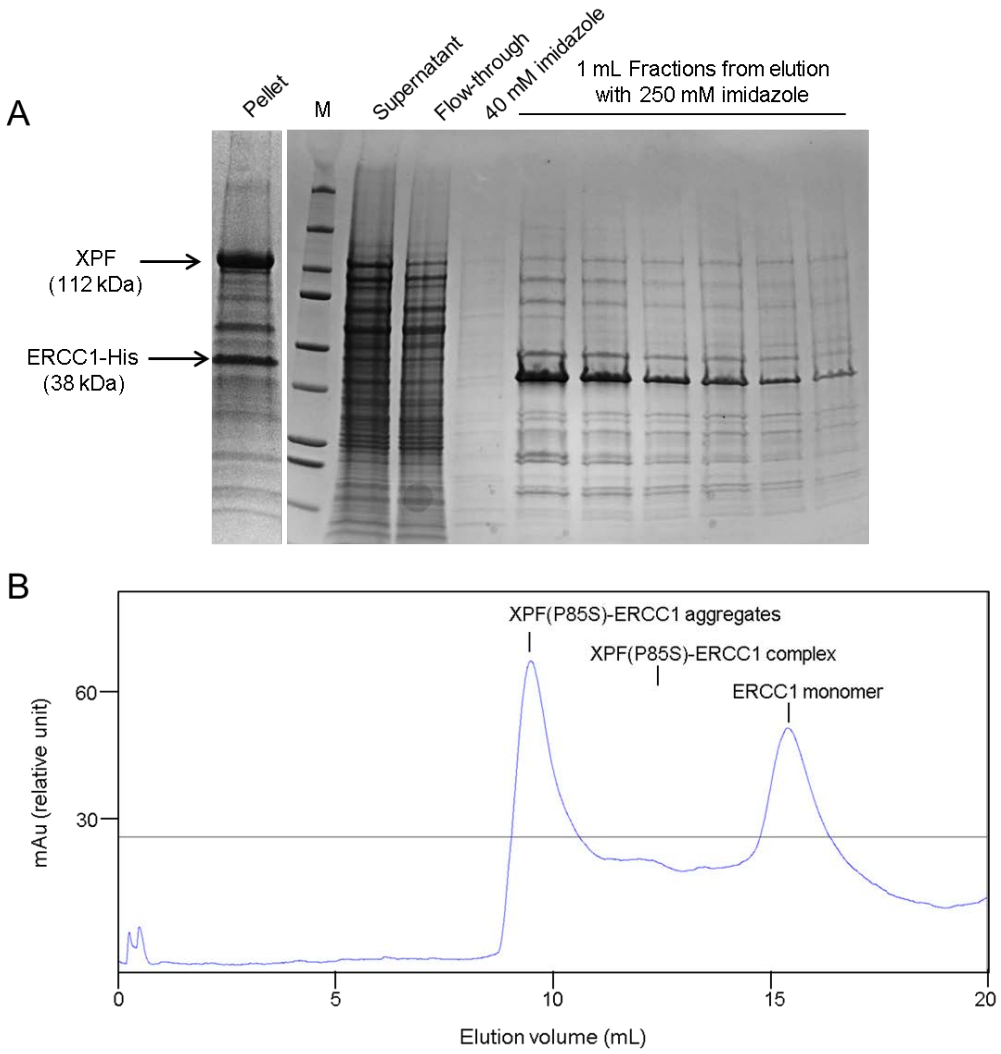


Figure 4. 11: Purification of XPF(P85S)-ERCC1.

A) SDS-PAGE gel (10%) of an aliquot of the fraction as labelled. The molecular weight of XPF and ERCC1 are 112 and 38 kDA, respectively. M = molecular weight marker. B) Gel filtration profiles for XPF(P85S)-ERCC1. The Peak at ~10mL corresponds to proteins in an aggregated state in the void volume of the column, the peak at ~14mL to the XPF-ERCC1 complex and the peak at ~17mL to monomeric ERCC.

4.7 XPF-ERCC1 and RPA activity on substrates that mimic a single replication fork collision with an ICL

I then tested whether RPA is able to stimulate XPF-ERCC1 activity inhibited by a model nascent leading strand on ICL-containing fork substrates that mimic a single replication fork collision with an ICL as previously described in **Figure 3. 11**. The substrates were previously described in Chapter 3, Section 3.4: On substrate CL-7, XPF-ERCC1 predicted incisions are 3' to the ICL; on substrate CL-4, XPF-ERCC1 predicted incisions are 5' and 3' flanking the ICL; and on substrate CL-1, the ICL is located at the junction and XPF-ERCC1 predicted incisions are 5' to the ICL. I have determined that the model nascent leading strand are annealed to the ICL-containing substrates by analysing the substrates in a 10% non-denaturing gel, as previously described in Section 3.2 for analysis of native undamaged substrates (data not shown).

4.7.1 RPA restores XPF-ERCC1 activity inhibited by a model nascent leading strand on ICL-containing fork substrates

The ICL-containing substrates were 5'-radiolabelled and reacted with XPF-ERCC1 in the presence or absence of RPA; and in the presence of absence a model nascent leading strand. XPF-ERCC1 activity inhibited by the model nascent leading strand is restored by RPA (**Figure 4. 12**). This data suggests that when a nascent leading strand of a single replication fork strikes an ICL, XPF-ERCC1 is unable to act on the structure. However, only in the presence of RPA, XPF-ERCC1 is able to incise the ICLs. However, the presence of additional XPF-ERCC1 incision products detectable when the

substrates are 5'radiolabelled complicates interpretation as to whether RPA alters XPF-ERCC1 incisions profile.

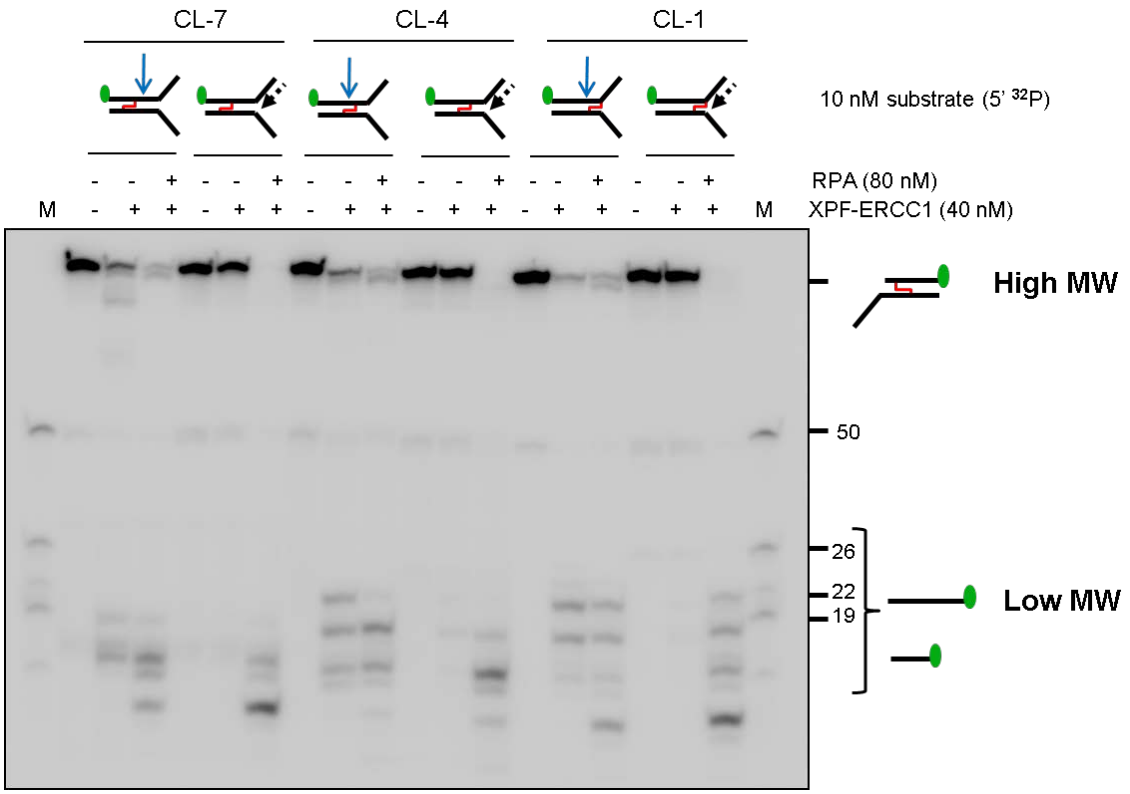


Figure 4. 12: RPA restores XPF-ERCC1 activity inhibited by a model nascent leading strand on ICL-containing fork substrates

40 nM XPF-ERCC1 (in the presence or absence of 80 nM RPA) is reacted with 5'-radiolabelled crosslinked substrates in the presence or absence of a model nascent leading strand. M = radiolabelled oligonucleotides of the sizes indicated.

4.7.2 Incision of ICL by RPA stimulation of XPF-ERCC1 influenced by the location of ICL to the fork junction

To determine whether RPA alters XPF-ERCC1 incisions profile, the ICL-containing substrates were 3'-radioabelled and reacted with XPF-ERCC1 in the presence and absence of RPA; and analysed in 15% non-denaturing PAGE gel (**Figure 4. 13**). RPA restores the low MW products on CL-7 substrate which corresponds to XPF-ERCC1 incisions both 3' to the ICL. On CL-4 (predicted incisions 5' and 3' flanking the ICL) and CL-1 (predicted incisions are both 5' to the ICL). However, low MW weight products corresponding to incisions 3' to the ICL were not detectable. This data suggest that eventhough RPA is able to restore XPF-ERCC1 activity (as seen on substrate CL-7) which is both 3' to the ICL, it is not able to stimulate another incision 3' to the ICL to enable ICL unhooking on substrate CL-4 and CL-1.

High MW products were observed in all of the substrates (in the presence or absence of a model nascent leading strand) when the substrates are 3'-radiolabelled indicative of incision even in the presence of a leading strand. However, no incision products are detectable when a model nascent leading strand is present on 5'-radiolabelled substrates, indicative of inhibition of incisions. This discrepancy could be due to incisions at the 3'-end of the bottom strand, hence the production of the high MW products. However, collectively, my data point towards the inability of XPF-ERCC1 to incise 3' to the ICL even in the presence of absence of RPA.

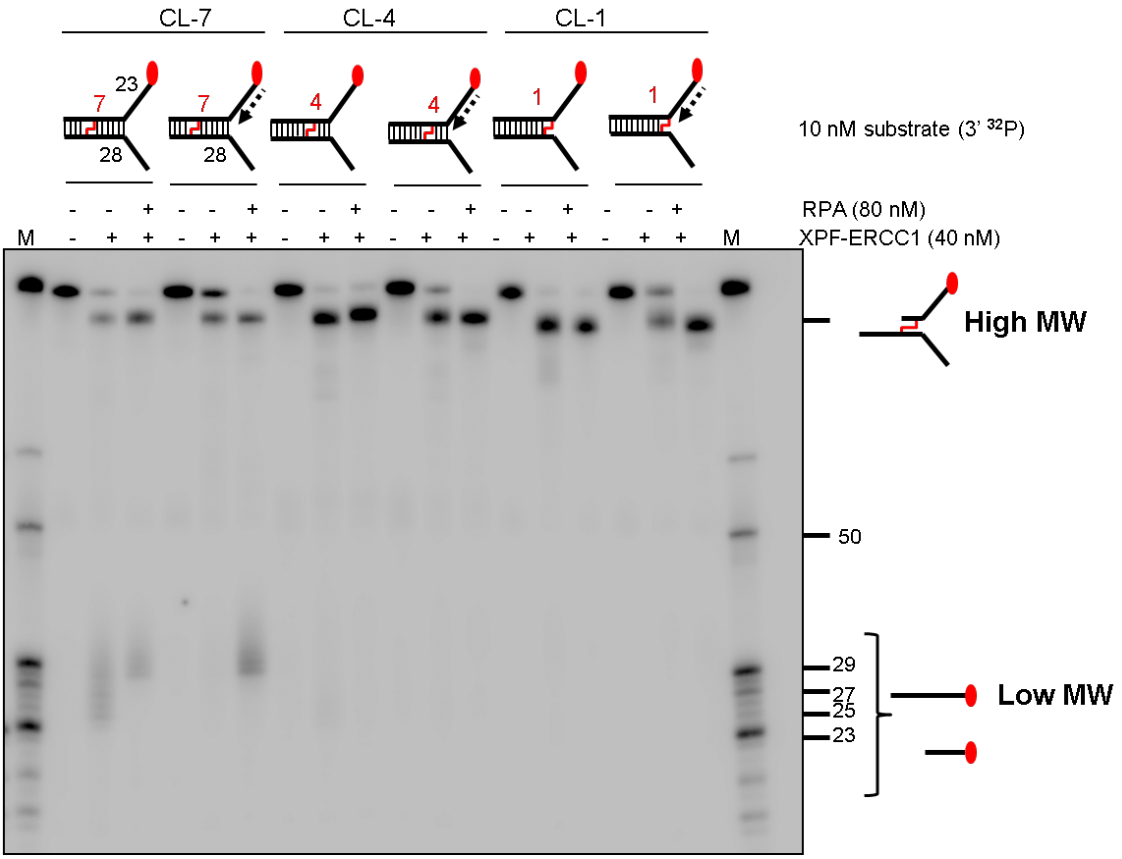


Figure 4. 13 RPA does not stimulate XPF-ERCC1 to incise flanking the ICL.

40 nM XPF-ERCC1 (in the presence or absence of 80 nM RPA) is reacted with 3'-radiolabelled crosslinked substrates in the presence or absence of a model nascent leading strand. M = radiolabelled oligonucleotides of the sizes indicated.

4.8 Substrate that mimic replication forks convergence on a crosslink

4.8.1 RPA is not able to activate XPF-ERCC1 activity on an X-shaped structure

Finally, I tested whether RPA is able to stimulate XPF-ERCC1 activity on X-shaped structures that mimics replication forks convergence onto an ICL. The ssDNA arms on the ICL are 23-mer in length. In **Figure 3. 14**, I have shown that XPF-ERCC1 activity is not detectable on an X-structure indicating that such structure is not a substrate for XPF-ERCC1. Here, I determined that RPA is not able to activate XPF-ERCC1 activity on these structures, whether it is the X-structure alone, or with the presence of a model leftward and/or rightward nascent leading strand. This data further confirms that X-shaped structure is not a preferred substrate of XPF-ERCC1 which suggests that the endonuclease requires regions of dsDNA for it to exert its activity (**Figure 4. 14**). Low MW bands detectable at the bottom of the gel, however it may not correspond to incisions near the ICL because each ssDNA arms surrounding the ICL is 23-mer in length, and the low MW bands are smaller than 23-mer suggesting incision close to the 3'-end.

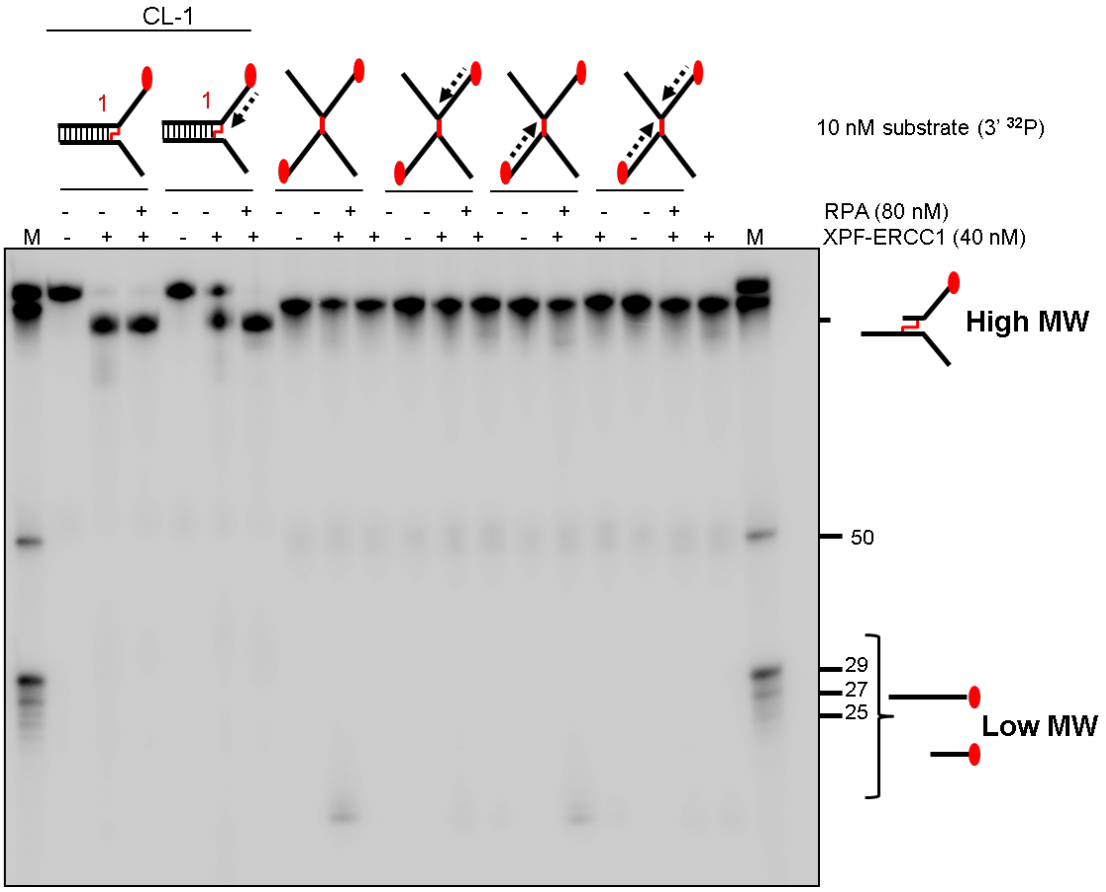


Figure 4. 14 RPA is not able to activate XPF-ERCC1 activity on an X-shaped structure that mimic dual replication forks convergence onto an ICL.

An array of 3'-radioabelled X-shaped substrates (X-shaped; X-shaped with rightward model nascent leading strand; X-shaped with leftward model nascent leading strand; and X-shaped with both model nascent leading strands) were reacted with 40 nM XPF-ERCC1 in the presence or absence of 80 nM RPA. The reaction samples were analysed in 10% denaturing gel. CL-1 substrate which contain an ICL at the fork junction was used a positive control and as a model for single replication fork collision with an ICL. M = radiolabelled oligonucleotide of indicated structures and sizes.

4.9 Discussion

In chapter 3, I have demonstrated that XPF-ERCC1 activity is inhibited by the presence of a model nascent leading strand or lagging strand on a fork substrate. Furthermore, a substrate that resembles a single replication fork collision with an ICL is also inhibited by a model nascent leading strand suggesting that the presence of a leading strand has a major inhibitory effect on XPF-ERCC1 activity. XPF-ERCC1 is however not detectable on an X-structure that mimic dual replication forks convergence onto an ICL. My data adds additional complexity to current understanding regarding ICL unhooking believed to be possibly by XPF-ERCC1 gained from cell-free replication-coupled repair system in *Xenopus laevis* egg extract which posits the arrival of a nascent leading strand to an X-structure that resembles dual replication forks convergence onto an ICL triggers ICL incisions on the opposite template strand (Klein Douwel, et al. 2014; Raschle, et al. 2008; Zhang, et al. 2015). In this chapter, I investigated whether the human ssDNA-binding protein RPA, previously reported to stimulate XPF-ERCC1 activity on a bubble substrate, the structure that may arise during NER, could also stimulates XPF-ERCC1 activity on substrates that mimic replication forks stalling by an ICL.

I determined that RPA stimulates and shifts XPF-ERCC1 incision at the site further into the duplex (6 nt from the junction), with loss of the incision closest to the junction (2 nt from the junction) (**Figure 4. 1**). Intriguingly, RPA specifically restores XPF-ERCC1 activity inhibited by a model nascent leading strand on the fork substrate. However, RPA only restores the incision further into the duplex, with loss of the incision closest to the junction (**Figure 4. 4**). The similar pattern of shifting and focusing of XPF-

ERCC1 incision further into the duplex region either by the presence of RPA on a fork substrate or the presence of a model nascent leading strand on the fork (albeit weakly without RPA) suggests that both the presence of a leading strand or another protein such as RPA causes steric hindrance to the positioning of XPF-ERCC1 active site on the substrate, therefore shifting it to further into the duplex and occluding the incision closest to the junction. However, the XPF-ERCC1 activity inhibited by a model nascent lagging strand on the fork substrate cannot be overcome by RPA (**Figure 4. 4**) possibly because the positioning of the central domain of ERCC1 is severely hindered by the presence of a model nascent lagging strand therefore completely abrogates XPF-ERCC1 activity on these substrates, as previously discussed in the discussion section in Chapter 3, Section 3.6. My data suggests that absence of XPF-ERCC1 activity due to hindrance of ERCC1 binding cannot be rescued by RPA.

As a first step to understanding the mechanism of RPA stimulation of XPF-ERCC1, I analysed the binding affinity of RPA to the DNA fork substrates. RPA has been shown to bind to a minimum 8 nt ssDNA (Blackwell, et al. 1996), suggesting that 8-10 nt is required as the precursor of the more stable 30 nt binding of RPA (Blackwell and Borowiec 1994). The affinity of human RPA for pyrimidines (C, T) is approximately 50-fold higher than its affinity for purines (A, G) however preference is not detectable in ssDNA of arbitrary mixture of pyrimidines and purines (Kim, et al. 1992). For this study, I am using a fork substrate with 22 nt ssDNA arms of an arbitrary mixture of pyrimidines and both arms have the exact same sequence. Therefore, the fork substrate that I am using fits all the criteria required for RPA binding and RPA should not have preference for 3'-ssDNA arm or the 5'-ssDNA arm because both are of the same length and sequence. Here I showed that RPA binds to a fork substrate, however, intriguingly,

RPA binding is not detectable on a fork with a model nascent leading strand even though RPA stimulates XPF-ERCC1 activity on this substrate. Binding of RPA is observed on a fork with a model nascent lagging strand, but not on fork with both model nascent leading and lagging strands (**Figure 4. 5**). Furthermore, RPA binding is inhibited by the presence of any length of leading strand on a fork substrate (**Figure 4. 6**). My data suggests that RPA has a preference for binding to 3'-ssDNA arm of a fork substrate, which is blocked by the presence of a model nascent leading strand. This is in line with current proposal which is that RPA binding to DNA involves a particular molecular polarity because RPA is shown to preferentially binds to 3'-flap compared to 5'-flap substrate (de Laat, et al. 1998b). This data suggests two possibilities: 1) the 5'-ssDNA is required for RPA stimulation of XPF-ERCC1 even though RPA binding to it is not detectable in an EMSA or 2) RPA stimulation of XPF-ERCC1 activity via direct physical interaction with the 3'-ssDNA arm is occluded by the presence of a model nascent leading strand.

To test the first possibility, I changed the 5'-ssDNA arm to a 5'-ssRNA arm for the 'bottom strand' of the fork substrate because RPA has lower binding affinity to RNA (Brill and Stillman 1989). Therefore, if the 5'ssDNA region is not required for RPA to stimulate XPF-ERCC1 activity, the DNA:RNA hybrid fork substrate which contain a 5'ssRNA arm will have the same level of activity as on a DNA fork substrate in the presence of RPA. Intriguingly, XPF-ERCC1 activity is reduced on the DNA:RNA fork substrate and its activity is not restored in the presence of RPA. When a model nascent leading strand is present on the DNA:RNA fork, XPF-ERCC1 activity is also not restored by RPA. This data indicates that XPF-ERCC1 also requires the 5'-ssDNA for activity and RPA also require the 5'-ssDNA arm to stimulate XPF-ERCC1 activity

(**Figure 4. 9**). However, in an EMSA, RPA binding to the DNA:RNA hybrid is similar to that of DNA substrates (**Figure 4. 10**). Taken together my data indicates that both XPF-ERCC1 and RPA requires the 5'-ssDNA arms.

To test the second possibility which is the physical interaction between the proteins, I attempted to observe ternary complex formation between the DNA, RPA and XPF-ERCC1 using an EMSA. DNA-RPA-XPF-ERCC1 ternary complex formation has been observed in an EMSA after the reaction products is crosslinked using gluteraldehyde on 5'-flap substrate (de Laat, et al. 1998b). I wish to determine whether the ternary complex formation can be observed without the use of gluteraldehyde. On a fork substrate, RPA binding is observed. In the presence of XPF-ERCC1, a faint band of super-shift indicating ternary complex formation is observed. However, it is not as strong as it should. On a fork with leading strand, RPA binding is not observed. However, in the presence of XPF-ERCC1, a faint band is observed indicating ternary complex formation. The band is not observed for fork with lagging strand eventhough binding of RPA is observed. However, the faint is very faint so we cannot make a conclusive statement on this (**Figure 4. 7**). I then tested ternary complex formation using fork with increasing length of leading strand. Intriguingly, a minor band is observed on all of the substrates indicating ternary complex formation.

A more conclusive way to investigate the physical interaction between proteins is test their physical interaction directly. First, a study concluded that XPF-ERCC1 complex interacts with RPA via XPF using co-immunoprecipitation (Bessho, et al. 1997). However, the result is not clear and the interaction said is minor. Second, a reverse yeast two-hybrid assay has reported that XPF interacts with the p70 subunit of RPA

(Fisher, et al. 2011). This study indicates that there are many mutations in XPF that interrupt the binding of RPA to XPF, however they further analysed the one has one single amino acid change which is the change from proline-85 to serine (P85S). Their data indicates that the mutant complex has the same level of activity at the wild-type complex This indicates that there are many residues on XPF that interact with RPA.

However, I decided to analyse the P85S mutation further by purifying the mutant complex using the same method as the purification for wild-type XPF-ERCC1. I discovered that even though both XPF and ERCC1 are equally expressed, the same as the expression for wild-type or nuclease-defective complex, however, most of XPF is degraded or aggregated because the band for XPF is smaller compared to ERCC1 after elution with nickel columns, unlike for wild-type and nuclease-defective where the elution of XPF and ERCC1 are equivalent. Gel filtration profile also showed that the expected peak for active XPF-ERCC1 is not present for the P85S. After assessment of the fractions after gel filtration, it shows that the level of ERCC1 is higher than XPF. The author were able to get equal amount of XPF and ERCC1 because they used a tag for both XPF and ERCC1 therefore by first capturing XPF, they were able to remove excess ERCC1. However, they did not publish the gel filtration profile of their purification therefore I was not able to compare my purification with them. I wish to explore the activity of the purified mutant complex in the near future. I am aware that the authors discovered twelve mutations in XPF that fail to interact with RPA, which was not disclosed. Therefore it is also possible that there are additional residues required for XPF interaction with RPA.

Another attempt to further investigate the interaction between XPF-ERCC1 and RPA is to confirm its interaction *in vivo*. The authors that determined the XPF mutation that disrupts its interaction with RPA also showed that XPF-defective CHO cells complemented with XPF(P85S) is not able to localise to the nucleus, unlike cells complemented with XPF(WT). They argue that interaction with RPA is required to translocate XPF-ERCC1 to the nucleus.

RPA is also able to restore XPF-ERCC1 activity inhibited by a leading strand on a fork substrate containing an ICL, which mimics a single fork collision in replication-dependent ICL repair (**Figure 4. 12**). However, RPA is not able to reactivate XPF-ERCC1 activity on an X-structure that resembles dual replication forks collision onto an ICL, further confirming that XPF-ERCC1 requires region of dsDNA for its activity (**Figure 4. 14**).

Taken together, my data provide insight on how RPA plays a significant role in modulating XPF-ERCC1 activity during replication-coupled ICL repair.

5 Collective activities of XPF-ERCC1 and RPA with the 5’ to 3’ exonuclease SNM1A

5.1 Introduction

In addition to RPA, XPF-ERCC1 has also been reported to genetically interact with another key factor implicated in ICL repair namely the 5’ to 3’ exonuclease SNM1A and physically with the nuclease scaffold protein SLX4. Previous genetic analysis has shown that SNM1A-depleted and SLX4-depleted human cells are epistatic with XPF-ERCC1-depleted human cells and exhibit the same phenotype of Mus81-dependent accumulation of γH2AX formation, and DSBs likely representing replication fork collapse. Moreover, cells doubly depleted for SNM1A and SLX4 exhibit equivalent hypersensitivity to ICL-inducing agents indicating that SNM1A and SLX4 participate in the same pathway as XPF-ERCC1 (Wang, et al. 2011).

The 5’ to 3’ exonuclease SNM1A is a member of the β-CASP family of proteins which contain an MBL (metallo-β-lactamase) structural domain which is structurally related to the MBL family of prokaryotic antibiotic detoxifying enzymes, and contributes to its ability to hydrolyse phosphodiester bonds (Callebaut, et al. 2002; Sengerova, et al. 2012). Biochemical analysis of purified human SNM1A demonstrated that it loads onto and digests an ICL-containing dsDNA substrate from either blunt ends, or, and of importance here, a single XPF-ERCC1-induced incision 5’ to the ICL (Wang, et al. 2011). Following XPF-ERCC1 incision 5’ from the ICL, SNM1A digestion initially stalls immediately prior to (0 nucleotide) 5’ to the ICL and digestion proceeds past the ICL leaving a single residual nucleotide tethered to the opposing strand, effectively

releasing or ‘unhooking’ the ICL (Wang, et al. 2011). Based on genetic and biochemical analysis of XPF-ERCC1 and SNM1A interactions, there are two scenarios to explain how SNM1A collaborates with XPF-ERCC1 during ICL repair. The first scenario is that after incisions flanking an ICL, possibly by XPF-ERCC1 alone or in conjunction with another endonuclease, an oligonucleotide is left tethered to the ICL. SNM1A is then recruited to digest the tethered oligonucleotide because its removal enables translesion synthesis and subsequent downstream repair processes to occur. This places SNM1A downstream of XPF-ERCC1 in the ICL repair pathway. In the second scenario, SNM1A directly enables ‘ICL unhooking’ where by XPF-ERCC1 initially induces a single incision 5’ to the ICL and the net result of SNM1A digestion past the ICL is to ‘unhook’ the ICL from the DNA substrate. This chapter is dedicated to describe the work in progress for biochemical analysis of XPF-ERCC1 and RPA with the 5’ to 3’ exonuclease SNM1A.

5.2 Biochemical analysis of XPF-ERCC1 and RPA with SNM1A

5.2.1 SM1A requires an incision by XPF-ERCC1 to load onto and digest a DNA fork substrate

In order to explore the combined activity of XPF-ERCC1 and RPA with SNM1A, human SNM1A truncated for its first 675 residues (676-1040) was utilised in assays with purified XPF-ERCC1 and RPA. The recombinant SNM1A was purified from insect cells by a member of our lab, Sook Yee Lee. SNM1A was truncated to increase its solubility during purification and it was determined that the truncation does not alter its 5’ to 3’ exonuclease activity and preference for ssDNA over dsDNA (Sengerova, et

al. 2012; Wang, et al. 2011). The truncated SNM1A will be referred to as simply 'SNM1A' from here on.

Unlike XPF-ERCC1, the divalent cation for optimal SNM1A processivity is magnesium (Mg^{2+}) instead of manganese (Mn^{2+}) (Sengerova, et al. 2012). Therefore, in assays involving XPF-ERCC1 and hSNM1A, 10 mM Mg^{2+} is used as the divalent cation instead of manganese (Mn^{2+}), and this change in divalent should not substantially affect XPF-ERCC1 activity, because XPF-ERCC1 was shown to be only slightly less active in Mg^{2+} (Enzlin and Scharer 2002). The DNA oligonucleotides used to generate the DNA substrates in this study do not contain a 5' phosphate, hence are not digested by the 5' to 3' exonucleolytic activity of SNM1A for their termini. Therefore, any SNM1A digestion observed in this study will only be from XPF-ERCC1-induced incisions which generates a nick in the DNA with 5'-end phosphate required for recognition and loading of SNM1A. The nuclease reaction for XPF-ERCC1 and SNM1A was carried out using 40 nM XPF-ERCC1 for 1 hour at 30^{O}C and further incubated with 0.8 nM hSNM1A for increasing time at 37^{O}C.

On a fork substrate, the reaction with XPF-ERCC1 alone produces 2 incision products 29-mer and 25-mer (lane 3, **Figure 5. 2**), as previously reported in the earlier chapters. When the substrate is further incubated with SNM1A for increasing time, step-wise digestion products were observed (lane 3 to 7), consistent with those previously described (Sengerova, et al. 2012; Wang, et al. 2011) resulting from a 5'-3' exonuclease activity. Within 60 minutes of incubation with SNM1A, digestion products shorter than the 23-mer marker were observed, indicating that SNM1A is able to digest past the fork junction suggesting that its activity is not affected when the DNA substrate transition

from dsDNA to ssDNA. To confirm that the digestion products were from the 5’ to 3’ exonuclease activity of SNM1A after incisions by XPF-ERCC1, the fork substrate was incubated with SNM1A alone for the maximal 60 minutes (lane 8). No incision/digestion products were observed indicating that the initial incisions by XPF-ERCC1 which generated a 5’-end phosphate is required for SNM1A to load and digest the DNA substrate. When a leading strand is present on a fork substrate, XPF-ERCC1 incision is inhibited, as shown in the previous chapters, which in turn prevents SNM1A from loading onto and digesting the DNA substrate (lane 9 to 15). This further confirms that incisions by XPF-ERCC1 are required for DNA substrate digestion by SNM1A.

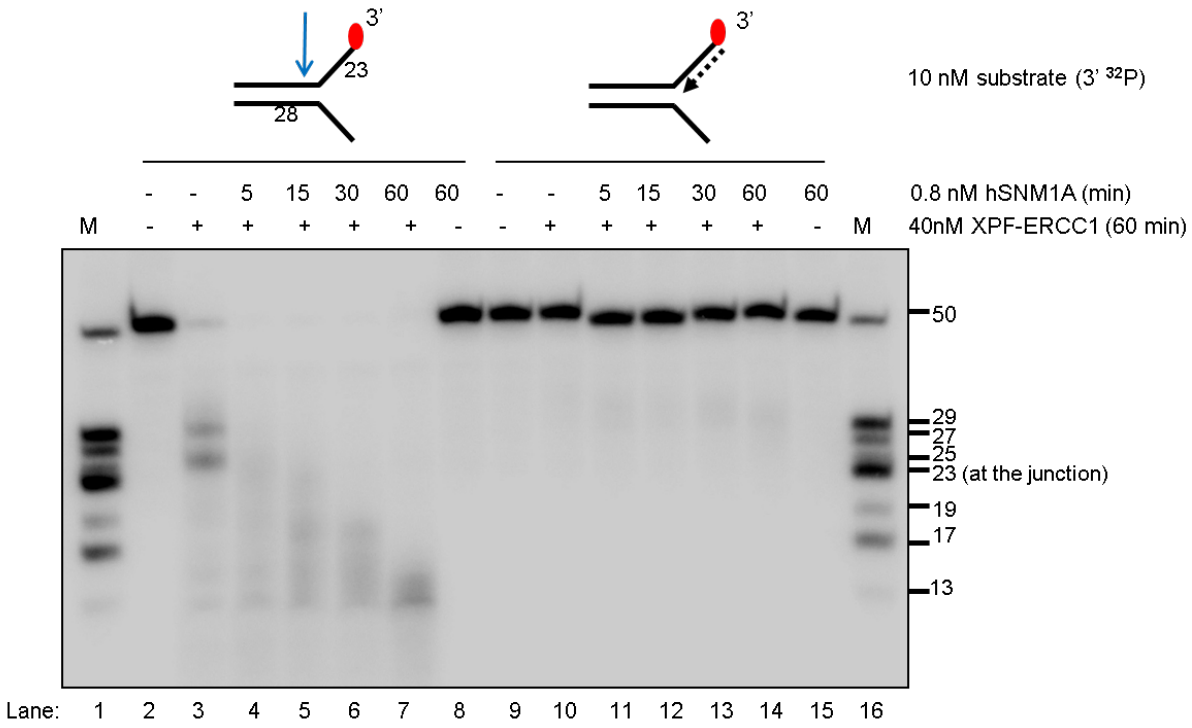


Figure 5. 1 SM1A requires an incision by XPF-ERCC1 to load onto and digest a DNA fork substrate.

10nM 3'-radiolabelled fork substrates (with or without a model nascent leading strand) were incubated with 40 nM XPF-ERCC1 for 1 hour at 30^{o}C, and further incubated with 0.8 nM SNM1A for increasing time (0, 5, 15, 30, 60 minutes) at 37^{o}C. M = radiolabelled oligonucleotides of the sizes indicated.

5.2.2 A leading strand on a fork substrate inhibits SNM1A activity

In the previous chapters, I determined that XPF-ERCC1 activity is inhibited by a nascent leading strand structure on a model fork substrate is restored by RPA. Here, I would like to determine whether, after RPA allows XPF-ERCC1 to overcome the inhibitory effect of a nascent leading strand, SNM1A is able to load onto the substrate from the incisions by produced XPF-ERCC1, and subsequently digest the substrate.

The digestion profile of SNM1A on a fork substrate is similar in the presence or absence of RPA, indicating that SNM1A activity is not affected by RPA (lane 2 to 9, **Figure 5. 3**). As expected, the inhibition of XPF-ERCC1 activity by a leading strand on a fork substrate is restored by RPA (lane 10 to 13), however, SNM1A is not able to digest the substrate even though XPF-ERCC1 incisions will generate the 5'-end phosphates that are required by SNM1A (lane 14 to 17). This data indicates that a leading strand on a fork substrate inhibits SNM1A activity possibly by altering the structure of the DNA therefore prevents SNM1A from loading onto the DNA or that SNM1A is able to load onto the DNA but the leading strand blocks SNM1A from digesting the substrate. It is intriguing that even though XPF-ERCC1 incision is focused to 6 nt 5' from the junction in the presence of RPA, SNM1A was not able to even digest the 6 nt region of the DNA 5' from to the junction.

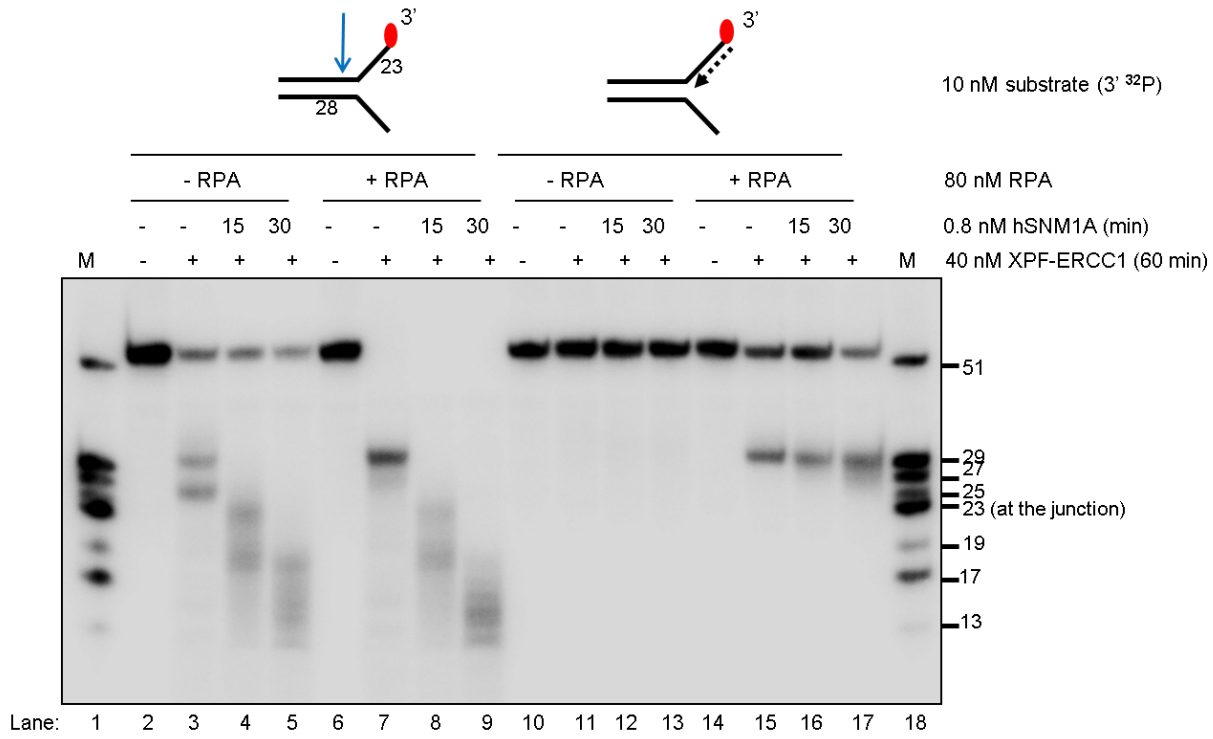


Figure 5. 2 A leading strand on a fork substrate inhibits XPF-ERCC1 activity

10nM 3'-radiolabelled fork substrates (with or without a model nascent leading strand) were incubated with 40 nM XPF-ERCC1 (with or without 80 nM RPA) for 1 hour at 30^{o}C, and further incubated with 0.8 nM SNM1A for increasing time (0, 15, and 30 minutes) at 37^{o}C. M = radiolabelled oligonucleotides of the sizes indicated.

5.2.3 XPF-ERCC1 activity with SNM1A on ICL-containing DNA fork substrates

Previous work in our lab has demonstrated that following XPF-ERCC1 induced incision of duplex DNA containing a single site-specific SJG-136 ICL , SNM1A initially stalls 0 nt 5' to the ICL and digests past the ICL, releasing the ICL from the duplex DNA (Sengerova, et al. 2012; Wang, et al. 2011).

In Chapter 3 (**Figure 3.11** to **3.13**), I demonstrated that XPF-ERCC1 alone is not able to unhook a single-site specific triazole ICL. When an ICL is located 5' upstream of the major XPF-ERCC1 incisions sites targeted on control undamaged fork structures (substrate CL-7), XPF-ERCC1 incises the DNA strand but does not 'unhook' the ICL because both incisions are 3' to the ICL. When an ICL is located in between the two major XPF-ERCC1 incision sites (substrate CL-4) or at the fork junction (substrate CL-1), only high molecular weight products are observed, indicating that incisions 3' of the ICL do not occur. These findings demonstrate that when an XPF-ERCC1 alone is not able to complete 'unhook' of an ICL because it can incise 5' to the ICL, but not 3' to the ICL. **Figure 5. 3** provides a schematic summary XPF-ERCC1 activity on the ICL-containing fork substrates as demonstrated in Chapter three.

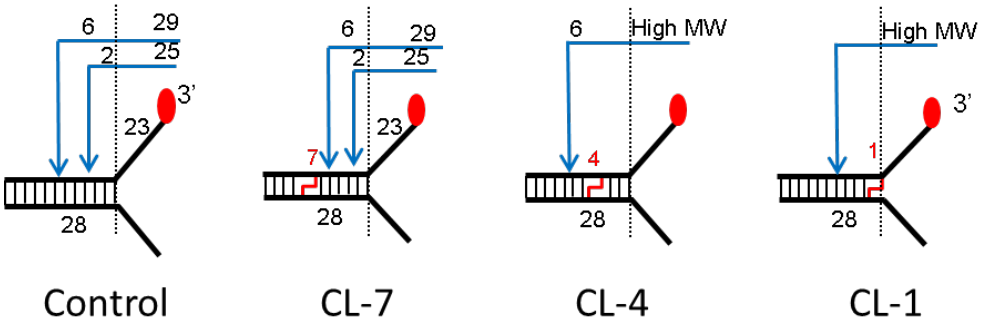


Figure 5. 4 Schematic representation of XPF-ERCC1 incisions on ICL-containing fork substrates.

XPF-ERCC1 incisions as determined in Chapter 3 (Figure 3.11 to figure 3.13) were represented as blue arrows. MW = molecular weight.

I therefore wished to determine whether on the ICL-containing substrates, XPF-ERCC1-induced incision 5' of the ICL enables digestion of the substrate by SNM1A therefore releases or 'unhooks' the ICL from the DNA. Digestion by SNM1A past the ICL from XPF-ERCC1-induced incisions could signify that SNM1A collaborates with XPF-ERCC1 during 'ICL unhooking', a pivotal step in ICL repair process *in vivo*. On the fork substrate that contains an ICL 5' upstream of XPF-ERCC1 incision sites (substrate CL-7), SNM1A is able to digest the substrate past the fork junction, with no sign of stalling (lane 2 to 8, **Figure 5. 6**). However, when the ICL is located in between XPF-ERCC1 incision sites (substrate CL-4), SNM1A digestion products range in between 29-mer to 23-mer suggesting SNM1A stalling 5' to the junction (lane 9 to 15). When the ICL is located at the junction (substrate CL-1), SNM1A digestion products seem be shorter than products from digestion of substrate CL-4, suggesting SNM1A stalling slightly closer (5') to the junction than with CL-4, or just past the junction. However, I am aware that the quality of the figure precludes definitive interpretation of

the result. 1) A product of approximately 50-mer in length is observed from all substrates, which could be a product of XPF-ERCC1 and SNM1A activities, or the denaturation of the crosslinked substrate. The former could be the reason for such product because it is not observed in substrate-only reaction (lane 2, 9 and 16). 2) The marker on the right (lane 23) has shifted up in the gel, therefore does not match the location of the left marker (lane 1) when a straight line is drawn across the gel, making it difficult to interpret the exact sizes of SNM1A digestion products.

However, based on this data, it can be postulated that XPF-ERCC1-induced incision 5' from the ICL might enable SNM1A to digest past the ICL, unhooking the ICL from the DNA duplex.

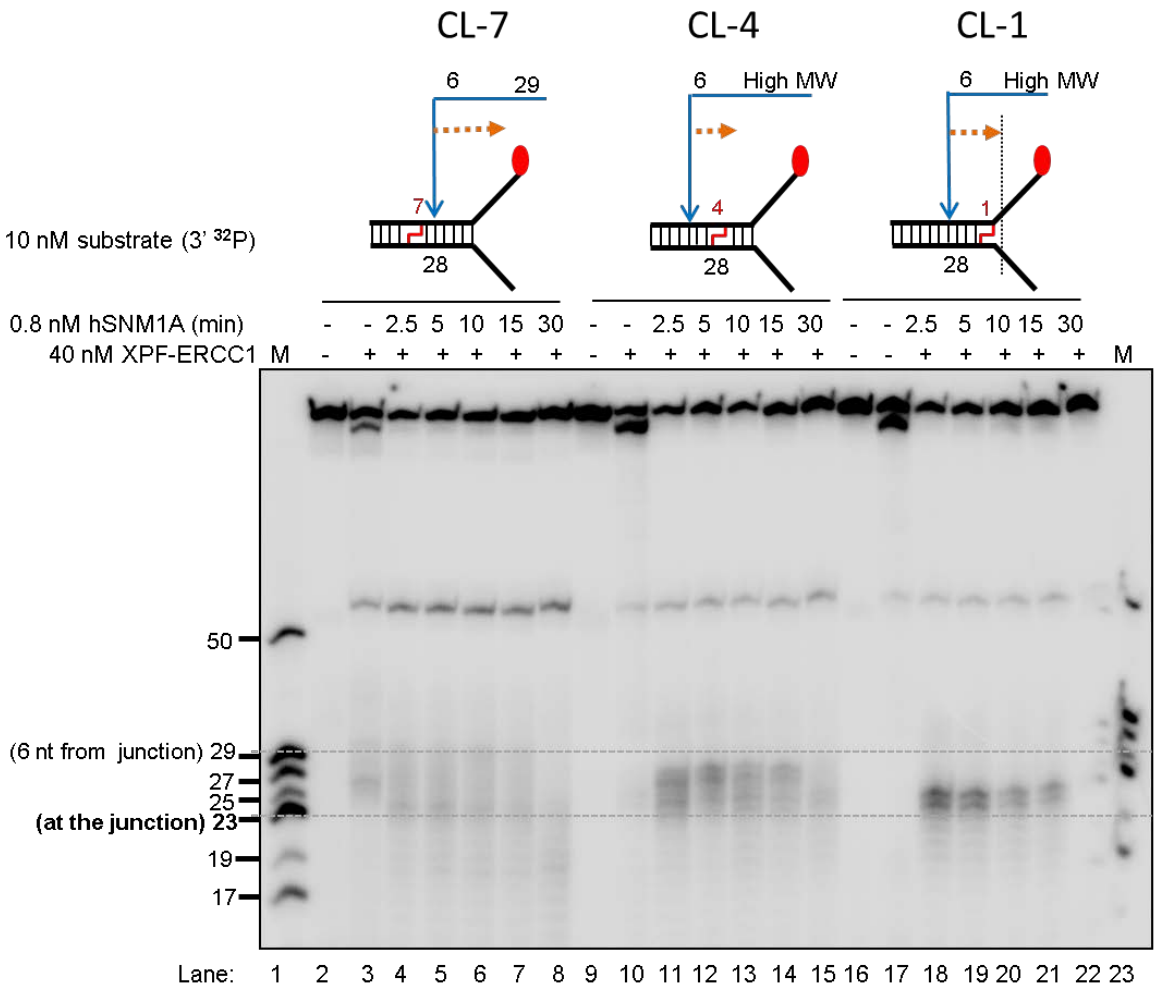


Figure 5. 5 XPF-ERCC1 activity with SNM1A on ICL-containing fork substrates.

10nM 3'-radiolabelled ICL-containing fork substrates (CL-7, CL-4 and CL-1) were incubated with 40 nM XPF-ERCC1 for 1 hour at 30^{o}C, and further incubated with 0.8 nM SNM1A for increasing times (0, 2.5, 5, 10, 15, and 30 minutes) at 30^{o}C. Blue arrows represent XPF-ERCC1 incisions and orange dotted arrows represent SNM1A digestion. M = radiolabelled oligonucleotides of the sizes indicated.

5.2.4 SNM1A is able to digest past an ICL on a fork substrate with a model nascent leading strand when RPA is present

In the previous figure, I determined that SNM1A is able to digest past an ICL located in fork substrates. However, when a replication fork encounters an ICL, ICL incisions are reported to be only be triggered by the progression of a nascent leading strand to 0 nt from ICL (Klein Douwel, et al. 2014; Raschle, et al. 2008; Zhang, et al. 2015). Therefore, I wanted to explore SNM1A activity on ICL-containing fork substrates in the presence of a model nascent leading strand.

In Chapter 4, I demonstrated that XPF-ERCC1 activity inhibited by a model nascent leading strand on ICL-containing fork substrates is restored by RPA. However, RPA does not restore or stimulate XPF-ERCC1 incisions 3' to the ICL when the ICL is located close to the junction (in between the two XPF-ERCC1 incision sites for substrate CL-4 or at the junction for substrate CL-1). Incisions 3' to the ICL are not detectable on these crosslinked substrates even in the presence of RPA. This indicates that even though RPA is able to restore XPF-ERCC1 activity inhibited by a model nascent leading strand, it does not stimulate XPF-ERCC1 to unhook the ICL.

Figure 5. 7 provides a schematic summary of XPF-ERCC1 activity with and without RPA on ICL-containing fork substrates in the presence of a model nascent leading strand, as demonstrated in Chapter four.

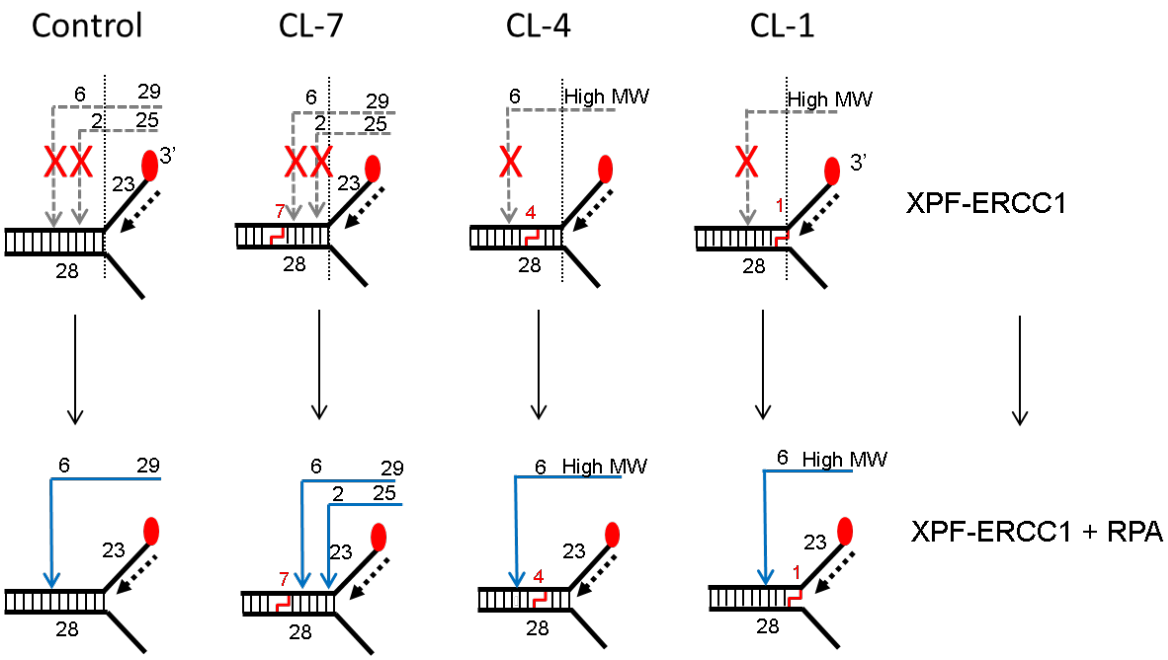


Figure 5. 6 Schematic representation of XPF-ERCC1 incisions on ICL-containing fork substrates in the presence of a model nascent leading strand.

As determined in Chapter four (Figure 4.12 and 4.14), XPF-ERCC1 incisions inhibited by the presence of a model nascent leading strand on ICL-containing fork substrate are restored by RPA. Blue arrows represent XPF-ERCC1 incisions. MW = molecular weight.

Therefore, to enable the analysis of the ICL-containing substrates in the presence and absence a model nascent leading strand, the substrates were reacted with XPF-ERCC1 in the presence of RPA (XPF-ERCC1-RPA), and further incubated with SNM1A for increasing time (**Figure 5. 9**). On a native fork substrate, XPF-ERCC1 activity inhibited by a model nascent leading strand is restored by RPA (lane 2 to 6). However, the model nascent leading strand inhibits SNM1A digestion (lane 7 to 11), as previously observed in Figure 5. 3. On substrate CL-7 (the ICL is located 5' upstream of XPF-ERCC1 at main incision sites for undamaged substrates) both with or without a model nascent leading strand, XPF-ERCC1-RPA incision is localised to 6 nt from the junction releasing a 29-mer product, but SNM1A digestion is not observed (lane 12 to 16; and 19 to 23). The similarity of SNM1A inhibition on substrate CL-7 and on a native fork substrate with a model nascent leading strand possibly indicates that an ICL located further into the duplex causes a similar distortion to the structure of the DNA as when a model nascent leading strand is present on a native fork substrate. Therefore, SNM1A recognises substrate CL-7 as similar to a native fork substrate with a model nascent leading strand.

On substrate CL-4, (the ICL is located in between XPF-ERCC1 main incision sites), both with or without a model nascent leading strand, XPF-ERCC1-RPA incision is localised to 5' of the ICL and SNM1A digested the substrate past the ICL, terminating at approximately 2 nt from the junction (25-mer product) (lane 19 to 23; 24 to 28). On substrate CL-1 (the ICL is located at the junction), XPF-ERCC1-RPA incision is localised to 5' of the ICL and SNM1A digested the substrate past the ICL, terminating at approximately 2 nt after into the junction (21-mer product) (lane 34 to 38; and 39 to 44). This data indicates that an ICL located closer to the junction (4 nt from the junction

for CL-4; and at the junction for CL-1), causes a dissimilar structural distortion to the DNA than an ICL located further into the duplex (7 nt from the junction for CL-7), which enable SNM1A to load onto the DNA and digest past the ICL. However, SNM1A digestion terminates at a different location for CL-4 and CL-1 indicating a structural dissimilarity between the two substrates. **Figure 5. 8** provides a schematic summary of the interplay between XPF-ERCC1-RPA induced incisions and digestion by SNM1A on ICL-containing fork substrates with a model nascent leading strand.

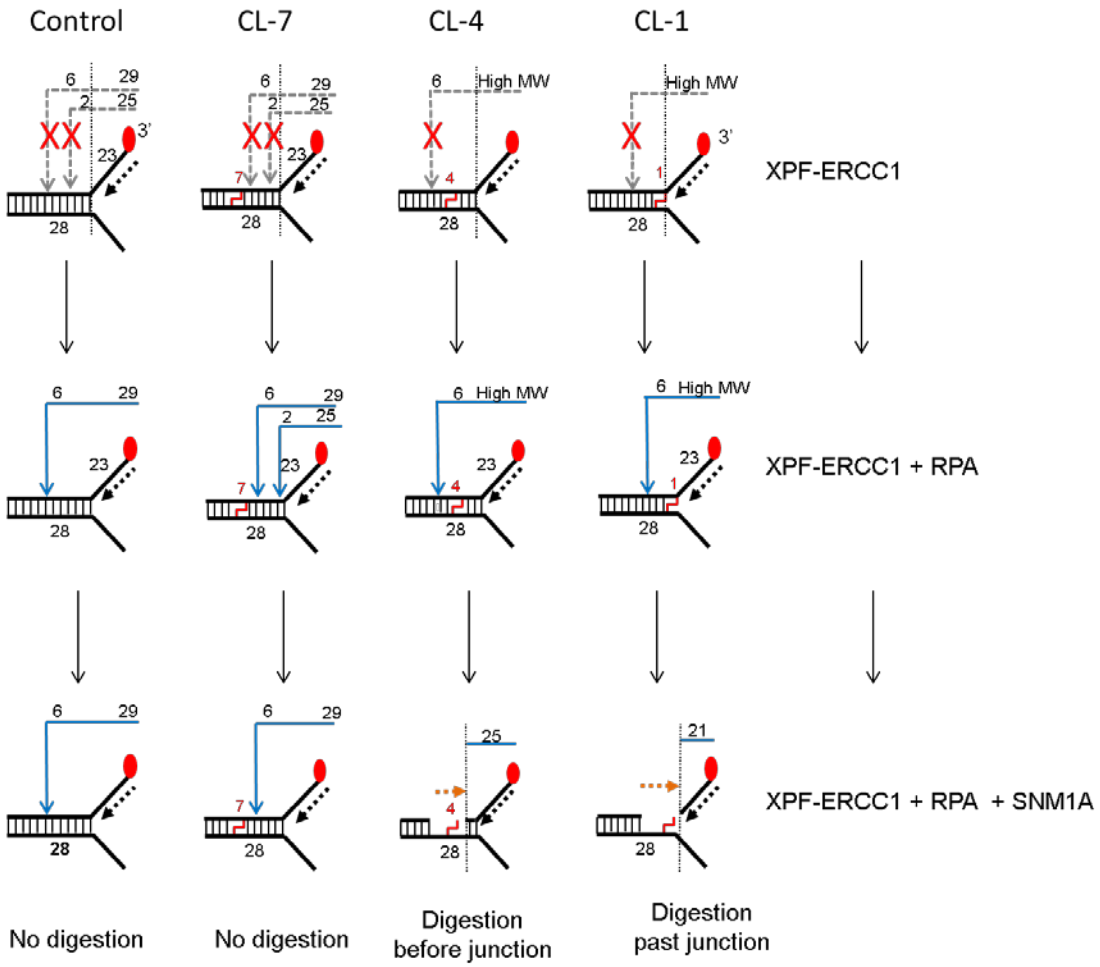


Figure 5. 7 Schematic representation of XPF-ERCC1-RPA incisions and SNM1A digestion on ICL-containing fork substrates in the presence of a model nascent leading strand.

As determined in chapter four (Figure 4.12 and 4.13), XPF-ERCC1 incisions inhibited by the presence of a model nascent leading strand on ICL-containing fork substrates is restored by RPA. XPF-ERCC1-RPA induced incisions enable the loading on SNM1A onto the substrate to digest past the ICL. Blue arrows represent XPF-ERCC1 incisions, orange dotted arrows represent SNM1A digestion. MW = molecular weight.

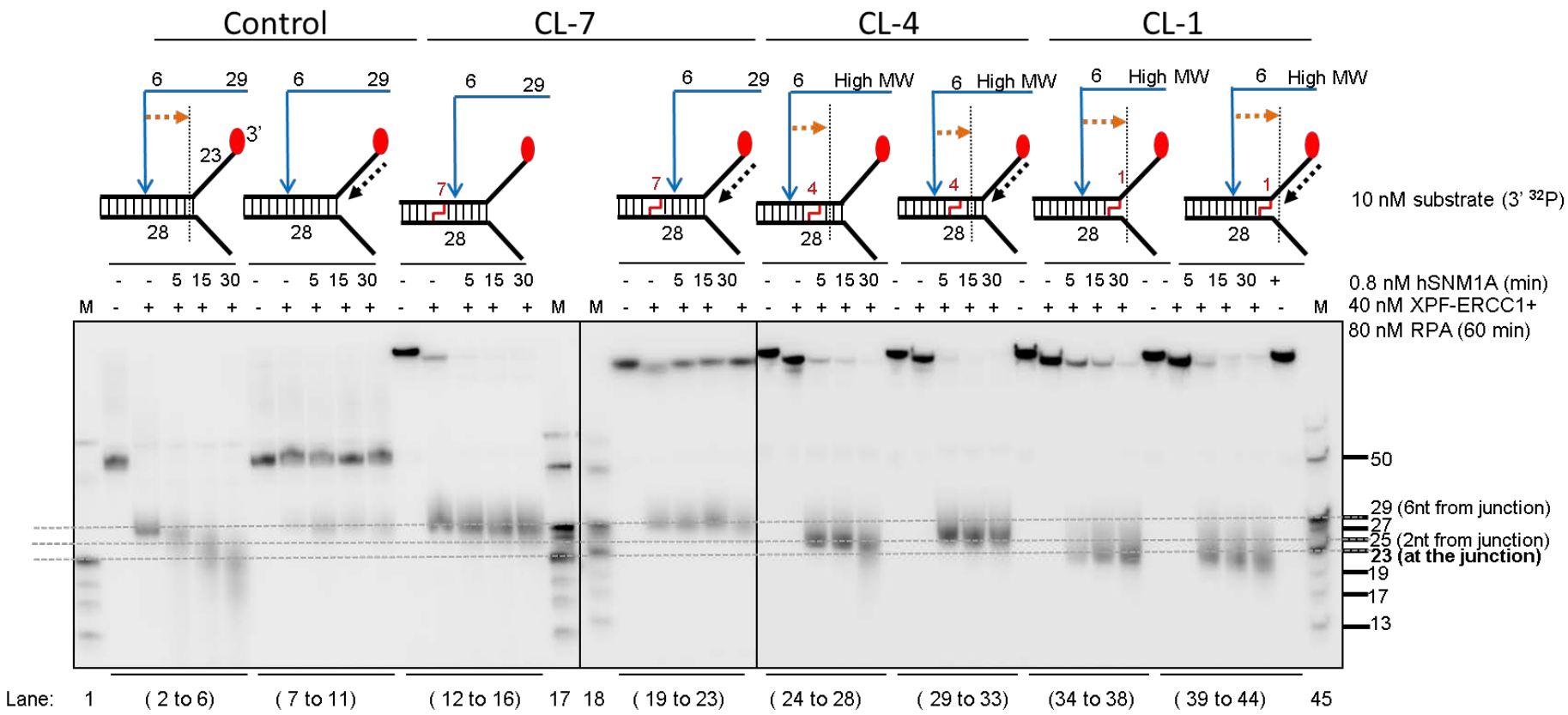

Control
CL-7
CL-4
CL-1
6
29
23
3'
28
7
4
1
High MW
10 nM substrate (3' 32P)
0.8 nM hSNM1A (min)
40 nM XPF-ERCC1+
80 nM RPA (60 min)
M
50
29 (6nt from junction)
27
25 (2nt from junction)
23 (at the junction)
19
17
13
Lane:
1
(2 to 6)
(7 to 11)
(12 to 16)
17
18
(19 to 23)
(24 to 28)
(29 to 33)
(34 to 38)
(39 to 44)
45

Figure 5. 8 SNM1A is able to digest past an ICL on a fork substrate with a model nascent leading strand when RPA is present.

10nM 3'-radiolabelled ICL-containing fork substrates (CL-7, CL-4 and CL-1) with or without a model nascent leading strand were incubated with 40 nM XPF-ERCC1 and 80 nM RPA for 1 hour at 30^{o}C, and further incubated with 0.8 nM SNM1A for increasing time (0, 5, 15, and 30 minutes) at 30^{o}C. Blue arrows represent XPF-ERCC1 incision sites, and orange dotted arrows represent SNM1A digestions. M = radiolabelled oligonucleotides of the sizes indicated.

5.3 Discussion

Previous genetic analysis in our lab has indicated that the 5’ to 3’ exonuclease SNM1A and the scaffold protein SLX4 participate in the same pathways as XPF-ERCC1 during ICL repair (Wang, et al. 2011). Therefore this chapter focuses on the biochemical activities of XPF-ERCC1 and RPA and their possible collaboration with SNM1A or SLX4.

Genetic and biochemical analysis in our lab of XPF-ERCC1 and SNM1A pointed towards two hypotheses to account for how SNM1A collaborates with XPF-ERCC1 during ICL repair. In the first hypothesis, SNM1A is required to digest the gapped intermediate covalently tethered to the incised DNA strand generated after incisions flanking an ICL possibly by XPF-ERCC1 alone or in conjunction with another endonuclease. In the second hypothesis, SNM1A participates in ‘ICL unhooking’ where by XPF-ERCC1 induces a single incision 5’ to the ICL and SNM1A digestion past the ICL is required to ‘unhook’ the ICL from the DNA substrate.

In Chapter 3, I demonstrated that XPF-ERCC1 incises a native fork substrate at two main locations: 2 nt and 6 nt 5’ from the fork junction. However, XPF-ERCC1 alone is not able to unhook a single site-specific ICL. When an ICL is located 5’ upstream of XPF-ERCC1 main incisions sites (substrate CL-7), XPF-ERCC1 incises the DNA strand but does not ‘unhook’ the ICL because both incisions are 3’ to the ICL. When an ICL is located in between the two XPF-ERCC1 incision sites (substrate CL-4) or at the fork junction (substrate CL-1), only high molecular weight products are observed, indicating that incision 3’ of the ICL does not occur. These findings demonstrate that

when an XPF-ERCC1 alone is not able to unhook an ICL because it can incise 5’ to the ICL, but not 3’ to the ICL.

In Chapter 4, I demonstrated that XPF-ERCC1 activity is inhibited by a model nascent leading strand on ICL-containing fork substrates is restored by RPA. However, RPA does not restore or stimulate XPF-ERCC1 incisions 3’ to the ICL when the ICL is located close to the junction (in between the two XPF-ERCC1 incision sites for substrate CL-4 or at the junction for substrate CL-1). Low molecular weight products indicative of incisions 3’ to the ICL are not detectable on these crosslinked substrates even in the presence of RPA. This indicates that even though RPA is able to restore XPF-ERCC1 activity inhibited by a model nascent leading strand, it does not facilitate XPF-ERCC1 ICL unhooking.

In this chapter, I demonstrated that SNM1A is able to digest ICL-containing substrates past the ICL, suggesting that SNM1A could collaborate with XPF-ERCC1 to unhook the ICL. On substrate CL-7 where the ICL is 5’ to (upstream) of XPF-ERCC1 incision sites, XPF-ERCC1 is able to digest past the junction. The key result is its activity on substrate CL-4 and CL-1 where the ICL is located in between XPF-ERCC1 incision sites and at the junction, respectively. SNM1A is digest past the ICL on these substrates indicating that SNM1A is able to collaborate with XPF-ERCC1 to unhook the ICL.

However, the major current model for replication-coupled ICL repair suggests that ICL incisions are to be only be triggered by the progression of a nascent leading strand to 0 nt from ICL (Klein Douwel, et al. 2014; Raschle, et al. 2008; Zhang, et al. 2015). There, I assessed SNM1A activity on a native fork with a model nascent leading strand.

Intriguingly, SNM1A fails to digest the substrate when a model nascent leading strand is present, even though XPF-ERCC1 incisions are restored by RPA. This may suggest that a model nascent leading on a fork substrate alters the structure of the DNA therefore prevents SNM1A from loading onto the DNA or prevent SNM1A is move onto the DNA to digest it.

Therefore, in order to determine whether SNM1A is able to collaborate with XPF-ERCC1 to unhook the ICL when a model nascent leading strand is present on the ICL-containing substrates, I reacted the substrates with XPF-ERCC1 and RPA prior to digestion with SNM1A. XPF-ERCC1-RPA-induced incisions are detectable for all of the substrates indicating that RPA facilitates XPF-ERCC1 incisions inhibited by the leading strand. Like a native fork substrate with a model nascent leading strand, SNM1A digestion is not observed for substrate CL-7 (the ICL is located 5’ upstream of XPF-ERCC1 main incision sites) both with or without a model nascent leading strand possibly due to similarity structural changes caused by a model nascent leading and an ICL further into the duplex, but this remains to be explored. The intriguing observation is for substrate CL-4 and CL-1 where the ICL is located in between XPF-ERCC1 main incision sites or at the junction respectively SNM1A can load from an XPF-ERCC1-RPA-induced incision and digest past the ICL.

Collectively, my data suggests that when XPF-ERCC1 incisions inhibited by a model nascent leading strand are restored by RPA, it enables SNM1A to load onto the DNA from the XPF-ERCC1-RPA-induced incision.

6 Concluding discussion and future work

One of the DNA processes that the XPF-ERCC1 endonuclease participates in is the replication-coupled repair of interstrand crosslinks (ICLs). Failure of XPF-ERCC1 to repair ICLs results in the heritable disorder Fanconi Anemia (FA) characterised by bone marrow failure, predisposition to cancer, and developmental defects. Additionally, ICL-inducing agents are extensively utilised in the clinic. Therefore XPF-ERCC1 is an attractive target to study given its important role in human health.

There are several predominant hypotheses regarding replication-coupled ICL repair that emerged recently from findings in cell-free replication-coupled repair system in X*enopus laevis* egg extracts (Klein Douwel, et al. 2014; Raschle, et al. 2008; Zhang, et al. 2015): 1) ICL repair is triggered by the convergence of dual replication forks onto an ICL and that repair fails to occur when a single replication encounters an ICL. This is a major shift from previous models because a single replication fork will almost always first strike an ICL *in vivo,* whereas replication fork convergence will generally occur some considerable time later; 2) An X-shaped structure corresponding to dual replication forks convergence onto the ICL is an absolute requirement for the initiation of ICL repair; 3) A nascent leading strand that gradually progresses to 1 nt from the ICL triggers dual incisions flanking the ICL on the opposite template strand; 4) Depletion of XPF effectively abolishes ICL repair, but not depletion of other nucleases previously implicated in ICL incisions namely Mus81 and Fan1.

The main objective of this study is to examine the role of XPF-ERCC1 heterodimeric endonucleases in repairing damaged DNA replication forks, taking into account the

current predominant hypotheses regarding its role in replication-coupled ICL repair. I utilised *in-vitro* reconstitution assays to achieve this objective. Specifically, for this study, I characterised XPF-ERCC1 endonuclease activity on model native and damaged replication fork structures; I determined the contribution of the replicative single-stranded DNA binding protein RPA to XPF-ERCC1 activity and I determined the collective activities of XPF-ERCC1 and RPA with the 5' to 3' SNM1A exonuclease. Eventhough *in-vitro* reconstitution assays have its own limitations, for example, the absence of other replication-associated factors or chromatin-bound proteins which play a significant role in modulating replication and hence may affect replication processing by XPF-ERCC1, it is a powerful tool because it allows the analysis of XPF-ERCC1 activity alone in the absence of other nucleases implicated in ICL repair.

In Chapter 3, I characterised XPF-ERCC1 endonuclease activity on simple model native and damaged replication fork structures. Using *in vitro* reconstitution assays, I determined that XPF-ERCC1 incises these simple native fork substrates at two main locations: 2 nt and 6 nt 5' from the fork junction. Based on this finding, I designed model damaged fork structures that contain a single-site specific triazole ICL, placed at three different locations. These structures represent a single replication collision with an ICL. The location of the ICL is chosen based on the locations of XPF-ERCC1 incisions targeted on the model native undamaged fork substrate. On crosslinked substrate CL-7, XPF-ERCC1 predicted incisions are 3' to the ICL; on crosslinked substrate CL-4, XPF-ERCC1 predicted incisions are 5' and 3' flanking the ICL; and on crosslinked substrate CL-1, the ICL is located at the junction and XPF-ERCC1 predicted incisions are 5' to the ICL.

Intriguingly, the position of an ICL affects XPF-ERCC1 incisions profile when compared to its incisions on a model native undamaged fork substrate: By labelling the ICL-containing substrates on their 5'-ends, I determined that an ICL further into the duplex, 5' from the predicted XPF-ERCC1 incisions, does not affect XPF-ERCC1 incisions (CL-7). However, an ICL located in the middle (CL-4) or 3' (CL-1) from the predicted XPF-ERCC1 incisions inhibit its incision that is closest to the junction which is 2 nt from the junction. The inhibition of XPF-ERCC1 activity by an ICL may signify that XPF-ERCC1 alone may be incapable of creating incisions flanking the ICL, even though it was previously demonstrated that XPF-ERCC1 can unhook an ICL (Kuraoka, et al. 2000).

I also explored the effect of a model nascent leading strand on a fork substrate, which, based on the cell-free replication system, would trigger ICL incisions. Intriguingly, XPF-ERCC1 activity is significantly reduced on native fork substrates or the ICL-containing fork substrates irrespective of position of ICLs on the substrate. This indicates that a model nascent leading strand has a major inhibitory effect on XPF-ERCC1 activity even on ICL-containing substrates and it may be possible that another factor is required to activate XPF-ERCC1 activity when a nascent leading strand strikes an ICL *in vivo*.

I then explored XPF-ERCC1 activity on X-shaped structures that model dual replication fork convergence, where a triazole ICL is flanked by four 22-mer ssDNA arms. XPF-ERCC1 incisions were not detectable, whether on the X-shaped structure alone or in the presence of model leftward and/or rightward nascent leading strands, indicating that an X-shaped structure is not a substrate for XPF-ERCC1. This finding is in agreement with

a recent model for XPF-ERCC1 binding to a fork substrate, generated from NMR titration assay and structural studies of XPF-ERCC1 C-terminal domains heterodimer (Das, et al. 2012; Tripsianes, et al. 2005). The model suggests that ERCC1 bind to dsDNA regions, and implies that a region of dsDNA is required for XPF-ERCC1 activities.

Collectively, my findings in Chapter 3 demonstrate that: 1) XPF-ERCC1 alone possibly incapable of unhooking an ICL on substrates that model a single replication fork collision with an ICL. XPF-ERCC1 may need other factors to activate its activity or it may collaborate with other nucleases to unhook the ICL. 2) X-shaped structures that model dual replication fork convergence onto an ICL is not a substrate for XPF-ERCC1. This suggests that an X-shaped structure may not be structure that was recognised by XPF-ERCC1 during replication forks convergence onto an ICL *in vivo*. Possibly, the DNA within the vicinity of the ICL may re-annealed to form duplex region that is recognised by XPF-ERCC1. Alternatively the annealing properties of certain helicases involved in replication-coupled DNA repair might be able to catalyse localised strand annealing around the ICL to generate a substrate for XPF-ERCC1 (Wu 2012). Therefore it is possible that in the situation when two forks converges onto an ICL, the progression of a nascent leading strand of one replication fork inhibits the nuclease activity of XPF-ERCC1, allowing the re-annealing of the DNA region of the other replication fork. However, another factor may be required to then facilitate XPF-ERCC1 to incise such structure. Based on this hypothesis, I sought to determine whether there are other factors required to activate or stimulate XPF-ERCC1 in the situations described above.

In Chapter 4, I characterised the contribution of the heterotrimeric single-stranded DNA binding protein RPA to XPF-ERCC1 activity on native and damaged replication fork structures. RPA was previously reported to stimulate XPF-ERCC1 and XPG endonucleases during nucleotide excision repair (NER) (Bessho, et al. 1997; Matsunaga, et al. 1996). Therefore, I wished to determine whether RPA could also stimulate XPF-ERCC1 during replication fork structure processing, in the presence and absence of ICLs at the fork. First, I demonstrated that RPA significantly stimulates XPF-ERCC1 activity on a native fork structure and focuses XPF-ERCC1 incision to a single location 6 nt within the duplex region, 5' from the fork junction. Intriguingly, RPA specifically restores XPF-ERCC1 activity inhibited by a model nascent leading strand on a native fork structure.

I therefore attempted to characterise the mechanism of RPA stimulation of XPF-ERCC1 activity. I attempted to address whether RPA stimulates XPF-ERCC1 by its binding or modulation of the DNA substrate, or by its physical interaction with XPF-ERCC1. First, I utilised enzyme mobility shift assays (EMSA) to determine RPA and/or XPF-ERCC1 binding affinity to the native fork substrate. I was not able to detect XPF-ERCC1 binding to the DNA substrates at the concentration used in the nuclease assays. This is possibly because of low or transient binding affinity of XPF-ERCC1 to the fork substrate. RPA binds more strongly the native fork substrate, but its binding was inhibited by a model nascent leading strand on the fork which suggests that the 3'-ssDNA arm is required for RPA binding. However, RPA stimulates XPF-ERCC1 activity a native fork with a model nascent leading strand in the absence of its binding to the DNA, therefore it may suggest that RPA exert its stimulation of XPF-ERCC1 by direct physical interaction with RPA. However, complex formation between XPF-

ERCC1 and RPA on the DNA substrate is observed at very low level, which complicates interpretation. Stimulation of RPA on XPF-ERCC1 is significantly reduced using DNA:RNA hybrid fork substrates with 5'-ssRNA on 'bottom' strand of the fork substrate which indicates that RPA exert its stimulation on XPF-ERCC1 by binding to the 5'-ssDNA arm. However, RPA binds to a fork with a model lagging strand which blocks the 5'-ssDNA arm, but XPF-ERCC1 activity is inhibited on this substrate. Therefore, at this stage, it is still inconclusive as to whether RPA exert its stimulation on XPF-ERCC1 *via* its binding to the DNA substrate or *via* its physical interaction with XPF-ERCC1, or through a combination of both these associations.

To further characterise the mechanism of stimulation of RPA on XPF-ERCC1, I attempted to purify a mutant XPF-ERCC1 reported to abolish its interaction with RPA (Fisher, et al. 2011). The yield for this mutant protein was very low but I have confirmed that both proteins are purified using Western blot. I wish to test its activity in the presence of RPA in the near future. I also would like analyse XPF-ERCC1 activity at the cellular level. I have attempted to create a human XPF-deficient using the CRISPR/Cas9 gene editing technology using U2OS cancer cell line and I currently have a mono-allelic disruption of XPF cell line. Additionally, I am also generating a stable complementation of wild-type or mutant XPF in XPF-deficient Chinese Hamster Ovaries (CHO) cells. I hope to analyse the effect of XPF-ERCC1 interaction with RPA at the cellular level through expression of P85S form. I hope to obtain the stable complement cell lines in the near future.

Additionally, I also analysed whether RPA is able to restore XPF-ERCC1 activity on ICL-containing substrates inhibited by the presence of a model nascent leading strand.

On substrate CL-7, where XPF-ERCC1 predicted incisions are 3’ to the ICL, the presence of RPA restores the appearance of low molecular weight products, indicative of incisions 3’ to the ICL. However, ICL unhooking would not occur because both XPF-ERCC1 incisions are 3’ to the ICL. On substrate CL-4, where XPF-ERCC1 predicted from native fork structure incisions are 5’ and 3’ flanking the ICL, low molecular weight products are not observed even in the presence of RPA, indicating that incision 3’ to the ICL did not take place. On substrate CL-1 where the ICL is located at the junction and XPF-ERCC1 predicted incisions are 5’ to the ICL, low molecular weight products were also not observed, indicating that the incision 3’ to the ICL did not take place.

I then tested whether RPA could stimulate XPF-ERCC1 activity on X-shaped substrates that resemble dual forks convergence onto an ICL. XPF-ERCC1 incisions are not detectable on these substrates even in the presence of RPA indicating that an X-shaped structure is not a substrate for XPF-ERCC1. This extends and confirms the conclusion that a DNA duplex region may be required for XPF-ERCC1 activity.

Collectively my data indicates that even though RPA is able to restore XPF-ERCC1 activity inhibited by a model nascent leading strand, it does not facilitate XPF-ERCC1 ICL unhooking. Based on these findings, I sought to determine whether there are other factors required for ICL unhooking.

In Chapter 5, I analysed the collective activities of XPF-ERCC1 and RPA with the 5’ to 3’ ICL repair exonuclease, human SNM1A. Biochemical analysis of highly purified SNM1A demonstrated that it loads onto and digests an ICL-containing dsDNA substrate

from either blunt ends, or, and of importance here, a single XPF-ERCC1-induced incision 5' to the ICL (Sengerova, et al. 2012; Wang, et al. 2011). Therefore, SNM1A is an attractive candidate for analysis with XPF-ERCC1 and RPA. On ICL-containing fork substrates with a model nascent leading strand, I determined that SNM1A is able to load onto and digest past the ICL from XPF-ERCC1-RPA-induced incisions 5' to the ICL. Collectively, my data suggests that when XPF-ERCC1 incisions inhibited by a model nascent leading strand are restored by RPA, it enables SNM1A to load onto the DNA from the XPF-ERCC1-RPA-induced incision.

Based on the findings in this study and the current model of replication-coupled ICL repair, I postulate that when a single replication fork encounters an ICL, the nascent leading strand initially stalls 20 nt from the ICL. It gradually progresses to 1 nt from the ICL, and its arrival at the ICL triggers an XPF-ERCC1-RPA-induced incision six nucleotides 5' to the junction, in a duplex region. SNM1A loads from these incisions and digest past the ICL, unhooking the ICL from the DNA duplex, leaving a residual single nucleotide moiety, which has been demonstrated as the reaction product using mass-spectrometry to characterise the reaction products of SNM1A activity in previous work (Wang et al., 2011). This enables translesion synthesis to occur and repair of the broken DNA strand via homologous recombination. In the event of dual replication fork convergence onto an ICL, I postulate that both nascent leading strands initially stall 20 nt from the ICL. CMG complexes from both replication forks unload from both leading strands, as previously described (Long, et al. 2014; Zhang, et al. 2015) which enables one nascent leading strand to gradually progresses to 1 nt from the ICL as previously described (Raschle, et al. 2008; Zhang, et al. 2015). I postulate that the region within the vicinity of the other replication fork would re-annealed to form DNA duplex region.

The structure that arises at this stage is inhibitory for XPF-ERCC1, but in the presence of RPA, XPF-ERCC1 is able to incise the structure within the duplex region, 6 nt from the ICL. This XPF-ERCC1-RPA-induced incision enables SNM1A to load onto and digest past the ICL. The net result of XPF-ERCC1-RPA-SNM1A is ICL unhooking, which enable subsequent repair process to occur via translesion synthesis and homologous recombination. **Figure 6. 1** provides a schematic representation of the mechanisms of ICL unhooking that I have described.

In conclusion, this study provides a possible mechanism for ICL unhooking that occur by the collective activities of the endonuclease XPF-ERCC1, the replicative single-strand DNA binding protein RPA and the 5’ to 3’ exonuclease SNM1A.

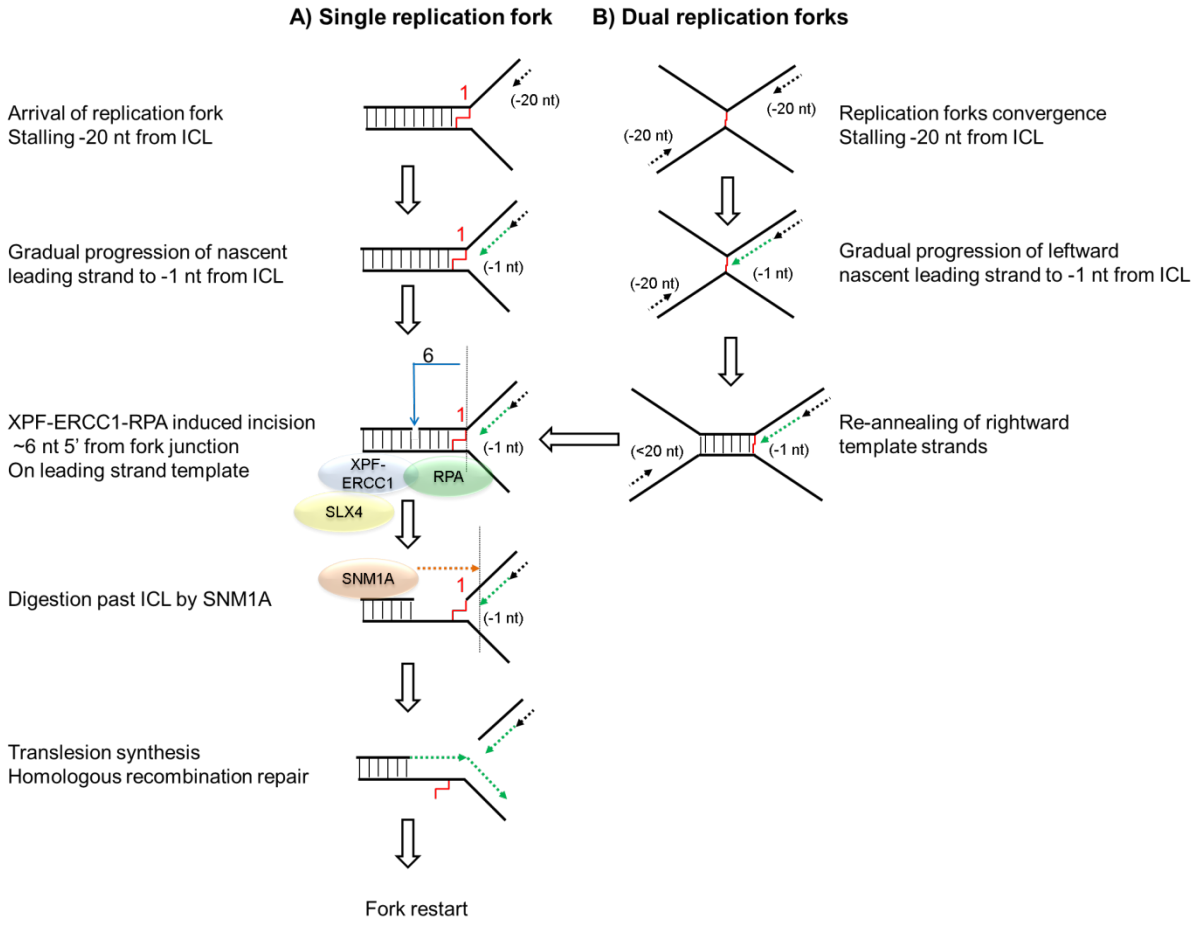


Figure 6. 1 Possible mechanisms of replication-coupled ICL repair based on findings from this study

Possible mechanism of replication-coupled ICL repair based on findings from this study for that described the collective collective activities of the endonuclease XPF-ERCC1, the replicative single-strand DNA binding protein RPA and the 5' to 3' exonuclease SNM1A that result in ICL unhooking. A) single replication fork collision with an ICL B) Dual replication forks convergence onto an ICL.

7 References

Adair, G. M., et al.
2000 Role of ERCC1 in removal of long non-homologous tails during targeted homologous recombination. The EMBO journal 19(20):5552-61.

Ahmad, A., et al.
2010 Mislocalization of XPF-ERCC1 nuclease contributes to reduced DNA repair in XP-F patients. PLoS genetics 6(3):e1000871.

Ahmad, A., et al.
2008 ERCC1-XPF endonuclease facilitates DNA double-strand break repair. Molecular and cellular biology 28(16):5082-92.

Akkari, Y. M., et al.
2000 DNA replication is required To elicit cellular responses to psoralen-induced DNA interstrand cross-links. Molecular and cellular biology 20(21):8283-9.

Al-Minawi, A. Z., N. Saleh-Gohari, and T. Helleday
2008 The ERCC1/XPF endonuclease is required for efficient single-strand annealing and gene conversion in mammalian cells. Nucleic acids research 36(1):1-9.

Andreassen, P. R., A. D. D'Andrea, and T. Taniguchi
2004 ATR couples FANCD2 monoubiquitination to the DNA-damage response. Genes & development 18(16):1958-63.

Auerbach, A. D.
2009 Fanconi anemia and its diagnosis. Mutation research 668(1-2):4-10.

Bardwell, A. J., et al.
1994 Specific cleavage of model recombination and repair intermediates by the yeast Rad1-Rad10 DNA endonuclease. Science 265(5181):2082-5.

Beckman, K. B., and B. N. Ames
1997 Oxidative decay of DNA. The Journal of biological chemistry 272(32):19633-6.

Berardini, M., W. Mackay, and E. L. Loechler
1997 Evidence for a recombination-independent pathway for the repair of DNA interstrand cross-links based on a site-specific study with nitrogen mustard. Biochemistry 36(12):3506-13.

Bessho, T., et al.
1997 Reconstitution of human excision nuclease with recombinant XPF-ERCC1 complex. The Journal of biological chemistry 272(6):3833-7.

Bhattacharyya, A., et al.
2000 The breast cancer susceptibility gene BRCA1 is required for subnuclear assembly of Rad51 and survival following treatment with the DNA cross-linking agent cisplatin. The Journal of biological chemistry 275(31):23899-903.

Blackwell, L. J., and J. A. Borowiec
1994 Human replication protein A binds single-stranded DNA in two distinct complexes. Molecular and cellular biology 14(6):3993-4001.

Blackwell, L. J., J. A. Borowiec, and I. A. Mastrangelo

1996 Single-stranded-DNA binding alters human replication protein A structure and facilitates interaction with DNA-dependent protein kinase. Molecular and cellular biology 16(9):4798-807.

Bogliolo, M., et al.
2013 Mutations in ERCC4, encoding the DNA-repair endonuclease XPF, cause Fanconi anemia. American journal of human genetics 92(5):800-6.

Bohgaki, T., M. Bohgaki, and R. Hakem
2010 DNA double-strand break signaling and human disorders. Genome integrity 1(1):15.

Brettel, K., and M. Byrdin
2010 Reaction mechanisms of DNA photolyase. Current opinion in structural biology 20(6):693-701.

Brill, S. J., and B. Stillman
1989 Yeast replication factor-A functions in the unwinding of the SV40 origin of DNA replication. Nature 342(6245):92-5.

Burma, S., et al.
2001 ATM phosphorylates histone H2AX in response to DNA double-strand breaks. The Journal of biological chemistry 276(45):42462-7.

Burrows, C. J., and J. G. Muller
1998 Oxidative Nucleobase Modifications Leading to Strand Scission. Chemical reviews 98(3):1109-1152.

Callebaut, I., et al.
2002 Metallo-beta-lactamase fold within nucleic acids processing enzymes: the beta-CASP family. Nucleic acids research 30(16):3592-601.

Chappell, C., et al.
2002 Involvement of human polynucleotide kinase in double-strand break repair by non-homologous end joining. The EMBO journal 21(11):2827-32.

Chen, L., et al.
2000 Interactions of the DNA ligase IV-XRCC4 complex with DNA ends and the DNA-dependent protein kinase. The Journal of biological chemistry 275(34):26196-205.

Chen, X. B., et al.
2001 Human Mus81-associated endonuclease cleaves Holliday junctions in vitro. Molecular cell 8(5):1117-27.

Ciccia, A., A. Constantinou, and S. C. West
2003 Identification and characterization of the human mus81-eme1 endonuclease. The Journal of biological chemistry 278(27):25172-8.

Ciccia, A., et al.
2007 Identification of FAAP24, a Fanconi anemia core complex protein that interacts with FANCM. Molecular cell 25(3):331-43.

Ciccia, A., N. McDonald, and S. C. West
2008 Structural and functional relationships of the XPF/MUS81 family of proteins. Annual review of biochemistry 77:259-87.

Cimprich, K. A., and D. Cortez
2008 ATR: an essential regulator of genome integrity. Nature reviews. Molecular cell biology 9(8):616-27.

Clauson, C., O. D. Scharer, and L. Niedernhofer
2013 Advances in understanding the complex mechanisms of DNA interstrand cross-link repair. Cold Spring Harbor perspectives in biology 5(10):a012732.

Cole, A. R., L. P. Lewis, and H. Walden
2010 The structure of the catalytic subunit FANCL of the Fanconi anemia core complex. Nature structural & molecular biology 17(3):294-8.
Cotta-Ramusino, C., et al.
2011 A DNA damage response screen identifies RHINO, a 9-1-1 and TopBP1 interacting protein required for ATR signaling. Science 332(6035):1313-7.
Das, D., et al.
2012 The structure of the XPF-ssDNA complex underscores the distinct roles of the XPF and ERCC1 helix- hairpin-helix domains in ss/ds DNA recognition. Structure 20(4):667-75.
de Laat, W. L., et al.
1998a DNA structural elements required for ERCC1-XPF endonuclease activity. The Journal of biological chemistry 273(14):7835-42.
de Laat, W. L., et al.
1998b DNA-binding polarity of human replication protein A positions nucleases in nucleotide excision repair. Genes & development 12(16):2598-609.
de Laat, W. L., et al.
1998c Mapping of interaction domains between human repair proteins ERCC1 and XPF. Nucleic acids research 26(18):4146-52.
De Silva, I. U., et al.
2000 Defining the roles of nucleotide excision repair and recombination in the repair of DNA interstrand cross-links in mammalian cells. Molecular and cellular biology 20(21):7980-90.
Deans, A. J., and S. C. West
2011 DNA interstrand crosslink repair and cancer. Nature reviews. Cancer 11(7):467-80.
Drotschmann, K., et al.
2001 Asymmetric recognition of DNA local distortion. Structure-based functional studies of eukaryotic Msh2-Msh6. The Journal of biological chemistry 276(49):46225-9.
Drummond, J. T., et al.
1995 Isolation of an hMSH2-p160 heterodimer that restores DNA mismatch repair to tumor cells. Science 268(5219):1909-12.
Duderstadt, K. E., et al.
2014 Replication-fork dynamics. Cold Spring Harbor perspectives in biology 6(1).
Dzantiev, L., et al.
2004 A defined human system that supports bidirectional mismatch-provoked excision. Molecular cell 15(1):31-41.
Eccles, L. J., P. O'Neill, and M. E. Lomax
2011 Delayed repair of radiation induced clustered DNA damage: friend or foe? Mutation research 711(1-2):134-41.
Enoiu, M., J. Jiricny, and O. D. Scharer
2012 Repair of cisplatin-induced DNA interstrand crosslinks by a replication-independent pathway involving transcription-coupled repair and translesion synthesis. Nucleic acids research 40(18):8953-64.
Enzlin, J. H., and O. D. Scharer
2002 The active site of the DNA repair endonuclease XPF-ERCC1 forms a highly conserved nuclease motif. The EMBO journal 21(8):2045-53.

Fagbemi, A. F., B. Orelli, and O. D. Scharer
2011 Regulation of endonuclease activity in human nucleotide excision repair. DNA repair 10(7):722-9.
Fairman, M. P., and B. Stillman
1988 Cellular factors required for multiple stages of SV40 DNA replication in vitro. The EMBO journal 7(4):1211-8.
Fan, L., et al.
2006 Conserved XPB core structure and motifs for DNA unwinding: implications for pathway selection of transcription or excision repair. Molecular cell 22(1):27-37.
Ferguson, D. O., and W. K. Holloman
1996 Recombinational repair of gaps in DNA is asymmetric in Ustilago maydis and can be explained by a migrating D-loop model. Proceedings of the National Academy of Sciences of the United States of America 93(11):5419-24.
Fisher, L. A., et al.
2011 Role of interaction of XPF with RPA in nucleotide excision repair. Journal of molecular biology 413(2):337-46.
Fishman-Lobell, J., and J. E. Haber
1992 Removal of nonhomologous DNA ends in double-strand break recombination: the role of the yeast ultraviolet repair gene RAD1. Science 258(5081):480-4.
Friedberg, E. C.
2001 How nucleotide excision repair protects against cancer. Nature reviews. Cancer 1(1):22-33.
Gan, G. N., et al.
2008 DNA polymerase zeta (pol zeta) in higher eukaryotes. Cell research 18(1):174-83.
Gargiulo, D., et al.
1995 Structural and function modification of DNA by mitomycin C. Mechanism of the DNA sequence specificity of mitomycins. Nucleic acids symposium series (34):169-70.
Garner, E., and A. Smogorzewska
2011 Ubiquitylation and the Fanconi anemia pathway. FEBS letters 585(18):2853-60.
Gary, R., et al.
1999 Proliferating cell nuclear antigen facilitates excision in long-patch base excision repair. The Journal of biological chemistry 274(7):4354-63.
Gates, K. S.
2009 An overview of chemical processes that damage cellular DNA: spontaneous hydrolysis, alkylation, and reactions with radicals. Chemical research in toxicology 22(11):1747-60.
Genschel, J., et al.
1998 Isolation of MutSbeta from human cells and comparison of the mismatch repair specificities of MutSbeta and MutSalpha. The Journal of biological chemistry 273(31):19895-901.
Gillet, L. C., and O. D. Scharer
2006 Molecular mechanisms of mammalian global genome nucleotide excision repair. Chemical reviews 106(2):253-76.
Gottlieb, T. M., and S. P. Jackson

1993 The DNA-dependent protein kinase: requirement for DNA ends and association with Ku antigen. Cell 72(1):131-42.

Gradia, S., et al.
1999 hMSH2-hMSH6 forms a hydrolysis-independent sliding clamp on mismatched DNA. Molecular cell 3(2):255-61.

Gregg, S. Q., A. R. Robinson, and L. J. Niedernhofer
2011 Physiological consequences of defects in ERCC1-XPF DNA repair endonuclease. DNA repair 10(7):781-91.

Groisman, R., et al.
2006 CSA-dependent degradation of CSB by the ubiquitin-proteasome pathway establishes a link between complementation factors of the Cockayne syndrome. Genes & development 20(11):1429-34.

Gunz, D., M. T. Hess, and H. Naegeli
1996 Recognition of DNA adducts by human nucleotide excision repair. Evidence for a thermodynamic probing mechanism. The Journal of biological chemistry 271(41):25089-98.

Haber, J. E.
2006 Transpositions and translocations induced by site-specific double-strand breaks in budding yeast. DNA repair 5(9-10):998-1009.

Hanada, K., et al.
2007 The structure-specific endonuclease Mus81 contributes to replication restart by generating double-strand DNA breaks. Nature structural & molecular biology 14(11):1096-104.

Hanada, K., et al.
2006 The structure-specific endonuclease Mus81-Eme1 promotes conversion of interstrand DNA crosslinks into double-strands breaks. The EMBO journal 25(20):4921-32.

Hanawalt, P. C., and G. Spivak
2008 Transcription-coupled DNA repair: two decades of progress and surprises. Nature reviews. Molecular cell biology 9(12):958-70.

He, Z., et al.
1995 RPA involvement in the damage-recognition and incision steps of nucleotide excision repair. Nature 374(6522):566-9.

Hefner, E., S. B. Preuss, and A. B. Britt
2003 Arabidopsis mutants sensitive to gamma radiation include the homologue of the human repair gene ERCC1. Journal of experimental botany 54(383):669-80.

Hess, M. T., et al.
1997 Bipartite substrate discrimination by human nucleotide excision repair. Proceedings of the National Academy of Sciences of the United States of America 94(13):6664-9.

Hodskinson, M. R., et al.
2014 Mouse SLX4 is a tumor suppressor that stimulates the activity of the nuclease XPF-ERCC1 in DNA crosslink repair. Molecular cell 54(3):472-84.

Hoeijmakers, J. H.
2009 DNA damage, aging, and cancer. The New England journal of medicine 361(15):1475-85.

Hoy, C. A., et al.

1985 Defective DNA cross-link removal in Chinese hamster cell mutants hypersensitive to bifunctional alkylating agents. Cancer research 45(4):1737-43.

Huang, M., et al.
2010 The FANCM/FAAP24 complex is required for the DNA interstrand crosslink-induced checkpoint response. Molecular cell 39(2):259-68.

Iftode, C., and J. A. Borowiec
2000 5' --> 3' molecular polarity of human replication protein A (hRPA) binding to pseudo-origin DNA substrates. Biochemistry 39(39):11970-81.

Ito, M., et al.
2005 Rad51 siRNA delivered by HVJ envelope vector enhances the anti-cancer effect of cisplatin. The journal of gene medicine 7(8):1044-52.

Iyama, T., and D. M. Wilson, 3rd
2013 DNA repair mechanisms in dividing and non-dividing cells. DNA repair 12(8):620-36.

Jackson, S. P., and J. Bartek
2009 The DNA-damage response in human biology and disease. Nature 461(7267):1071-8.

Jazayeri, A., et al.
2006 ATM- and cell cycle-dependent regulation of ATR in response to DNA double-strand breaks. Nature cell biology 8(1):37-45.

Jensen, R. B., A. Carreira, and S. C. Kowalczykowski
2010 Purified human BRCA2 stimulates RAD51-mediated recombination. Nature 467(7316):678-83.

Jiricny, J.
2006 The multifaceted mismatch-repair system. Nature reviews. Molecular cell biology 7(5):335-46.

Kadyrov, F. A., et al.
2006 Endonucleolytic function of MutLalpha in human mismatch repair. Cell 126(2):297-308.

Kashiyama, K., et al.
2013 Malfunction of nuclease ERCC1-XPF results in diverse clinical manifestations and causes Cockayne syndrome, xeroderma pigmentosum, and Fanconi anemia. American journal of human genetics 92(5):807-19.

Kass, E. M., and M. Jasin
2010 Collaboration and competition between DNA double-strand break repair pathways. FEBS letters 584(17):3703-8.

Kemp, M. G., et al.
2012 Mechanism of release and fate of excised oligonucleotides during nucleotide excision repair. The Journal of biological chemistry 287(27):22889-99.

Kim, C., R. O. Snyder, and M. S. Wold
1992 Binding properties of replication protein A from human and yeast cells. Molecular and cellular biology 12(7):3050-9.

Kim, J. M., et al.
2009 Inactivation of murine Usp1 results in genomic instability and a Fanconi anemia phenotype. Developmental cell 16(2):314-20.

Klein Douwel, D., et al.
2014 XPF-ERCC1 acts in Unhooking DNA interstrand crosslinks in cooperation with FANCD2 and FANCP/SLX4. Molecular cell 54(3):460-71.

Knipscheer, P., et al.
2009 The Fanconi anemia pathway promotes replication-dependent DNA interstrand cross-link repair. Science 326(5960):1698-701.
Komori, K., et al.
2002 Novel endonuclease in Archaea cleaving DNA with various branched structure. Genes & genetic systems 77(4):227-41.
Kottemann, M. C., and A. Smogorzewska
2013 Fanconi anaemia and the repair of Watson and Crick DNA crosslinks. Nature 493(7432):356-63.
Kraemer, K. H., et al.
1994 Xeroderma pigmentosum and related disorders: examining the linkage between defective DNA repair and cancer. The Journal of investigative dermatology 103(5 Suppl):96S-101S.
Krasikova, Y. S., et al.
2010 Localization of xeroderma pigmentosum group A protein and replication protein A on damaged DNA in nucleotide excision repair. Nucleic acids research 38(22):8083-94.
Kratz, K., et al.
2010 Deficiency of FANCD2-associated nuclease KIAA1018/FAN1 sensitizes cells to interstrand crosslinking agents. Cell 142(1):77-88.
Krokan, H. E., and M. Bjoras
2013 Base excision repair. Cold Spring Harbor perspectives in biology 5(4):a012583.
Kuraoka, I., et al.
2000 Repair of an interstrand DNA cross-link initiated by ERCC1-XPF repair/recombination nuclease. The Journal of biological chemistry 275(34):26632-6.
Lamarche, B. J., N. I. Orazio, and M. D. Weitzman
2010 The MRN complex in double-strand break repair and telomere maintenance. FEBS letters 584(17):3682-95.
Langevin, F., et al.
2011 Fancd2 counteracts the toxic effects of naturally produced aldehydes in mice. Nature 475(7354):53-8.
Lawley, P. D., and D. H. Phillips
1996 DNA adducts from chemotherapeutic agents. Mutation research 355(1-2):13-40.
Lehmann, A. R.
2011 DNA polymerases and repair synthesis in NER in human cells. DNA repair 10(7):730-3.
Lempiainen, H., and T. D. Halazonetis
2009 Emerging common themes in regulation of PIKKs and PI3Ks. The EMBO journal 28(20):3067-73.
Levin, D. S., et al.
2004 A conserved interaction between the replicative clamp loader and DNA ligase in eukaryotes: implications for Okazaki fragment joining. The Journal of biological chemistry 279(53):55196-201.
Li, G. M., and P. Modrich

1995 Restoration of mismatch repair to nuclear extracts of H6 colorectal tumor cells by a heterodimer of human MutL homologs. Proceedings of the National Academy of Sciences of the United States of America 92(6):1950-4.

Li, X., and W. D. Heyer

2008 Homologous recombination in DNA repair and DNA damage tolerance. Cell research 18(1):99-113.

Lieber, M. R.

2010 The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. Annual review of biochemistry 79:181-211.

Lindahl, T.

1993 Instability and decay of the primary structure of DNA. Nature 362(6422):709-15.

Lobrich, M., and P. A. Jeggo

2007 The impact of a negligent G2/M checkpoint on genomic instability and cancer induction. Nature reviews. Cancer 7(11):861-9.

Loeb, L. A., et al.

1986 Apurinic sites as common intermediates in mutagenesis. Basic life sciences 38:341-7.

Long, D. T., et al.

2014 BRCA1 promotes unloading of the CMG helicase from a stalled DNA replication fork. Molecular cell 56(1):174-85.

Long, D. T., et al.

2011 Mechanism of RAD51-dependent DNA interstrand cross-link repair. Science 333(6038):84-7.

Lykke-Andersen, J., R. A. Garrett, and J. Kjems

1997 Mapping metal ions at the catalytic centres of two intron-encoded endonucleases. The EMBO journal 16(11):3272-81.

Ma, Y., et al.

2002 Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. Cell 108(6):781-94.

MacKay, C., et al.

2010 Identification of KIAA1018/FAN1, a DNA repair nuclease recruited to DNA damage by monoubiquitinated FANCD2. Cell 142(1):65-76.

Malinge, J. M., C. Perez, and M. Leng

1994 Base sequence-independent distorsions induced by interstrand cross-links in cis-diamminedichloroplatinum (II)-modified DNA. Nucleic acids research 22(19):3834-9.

Mathieu, N., N. Kaczmarek, and H. Naegeli

2010 Strand- and site-specific DNA lesion demarcation by the xeroderma pigmentosum group D helicase. Proceedings of the National Academy of Sciences of the United States of America 107(41):17545-50.

Mathieu, N., et al.

2013 DNA quality control by a lesion sensor pocket of the xeroderma pigmentosum group D helicase subunit of TFIIH. Current biology : CB 23(3):204-12.

Matsunaga, T., et al.

1996 Replication protein A confers structure-specific endonuclease activities to the XPF-ERCC1 and XPG subunits of human DNA repair excision nuclease. The Journal of biological chemistry 271(19):11047-50.

Matsuoka, S., et al.
2007 ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. Science 316(5828):1160-6.

McCabe, K. M., et al.
2008 ERCC1 is required for FANCD2 focus formation. Molecular genetics and metabolism 95(1-2):66-73.

McHugh, P. J., and S. Sarkar
2006 DNA interstrand cross-link repair in the cell cycle: a critical role for polymerase zeta in G1 phase. Cell cycle 5(10):1044-7.

McHugh, P. J., W. R. Sones, and J. A. Hartley
2000 Repair of intermediate structures produced at DNA interstrand cross-links in Saccharomyces cerevisiae. Molecular and cellular biology 20(10):3425-33.

McIlwraith, M. J., et al.
2005 Human DNA polymerase eta promotes DNA synthesis from strand invasion intermediates of homologous recombination. Molecular cell 20(5):783-92.

McIlwraith, M. J., and S. C. West
2008 DNA repair synthesis facilitates RAD52-mediated second-end capture during DSB repair. Molecular cell 29(4):510-6.

Minko, I. G., et al.
2008 Replication bypass of the acrolein-mediated deoxyguanine DNA-peptide cross-links by DNA polymerases of the DinB family. Chemical research in toxicology 21(10):1983-90.

Mordasini, T., A. Curioni, and W. Andreoni
2003 Why do divalent metal ions either promote or inhibit enzymatic reactions? The case of BamHI restriction endonuclease from combined quantum-classical simulations. The Journal of biological chemistry 278(7):4381-4.

Moser, J., et al.
2007 Sealing of chromosomal DNA nicks during nucleotide excision repair requires XRCC1 and DNA ligase III alpha in a cell-cycle-specific manner. Molecular cell 27(2):311-23.

Moynahan, M. E., and M. Jasin
2010 Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. Nature reviews. Molecular cell biology 11(3):196-207.

Muniandy, P. A., et al.
2010 DNA interstrand crosslink repair in mammalian cells: step by step. Critical reviews in biochemistry and molecular biology 45(1):23-49.

Muniandy, P. A., et al.
2009 Repair of laser-localized DNA interstrand cross-links in G1 phase mammalian cells. The Journal of biological chemistry 284(41):27908-17.

Newman, M., et al.
2005 Structure of an XPF endonuclease with and without DNA suggests a model for substrate recognition. The EMBO journal 24(5):895-905.

Niedernhofer, L. J., et al.
2006 A new progeroid syndrome reveals that genotoxic stress suppresses the somatotroph axis. Nature 444(7122):1038-43.
Niedernhofer, L. J., A. S. Lalai, and J. H. Hoeijmakers
2005 Fanconi anemia (cross)linked to DNA repair. Cell 123(7):1191-8.
Niedernhofer, L. J., et al.
2004 The structure-specific endonuclease Ercc1-Xpf is required to resolve DNA interstrand cross-link-induced double-strand breaks. Molecular and cellular biology 24(13):5776-87.
Nimonkar, A. V., et al.
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 105(44):16906-11.
Nishino, T., et al.
2003 X-ray and biochemical anatomy of an archaeal XPF/Rad1/Mus81 family nuclease: similarity between its endonuclease domain and restriction enzymes. Structure 11(4):445-57.
—
2005 Structural and functional analyses of an archaeal XPF/Rad1/Mus81 nuclease: asymmetric DNA binding and cleavage mechanisms. Structure 13(8):1183-92.
Nojima, K., et al.
2005 Multiple repair pathways mediate tolerance to chemotherapeutic cross-linking agents in vertebrate cells. Cancer research 65(24):11704-11.
Noll, D. M., T. M. Mason, and P. S. Miller
2006 Formation and repair of interstrand cross-links in DNA. Chemical reviews 106(2):277-301.
Norman, D., et al.
1990 NMR and computational characterization of mitomycin cross-linked to adjacent deoxyguanosines in the minor groove of the d(T-A-C-G-T-A).d(T-A-C-G-T-A) duplex. Biochemistry 29(11):2861-75.
O'Connor, P. M., and K. W. Kohn
1990 Comparative pharmacokinetics of DNA lesion formation and removal following treatment of L1210 cells with nitrogen mustards. Cancer communications 2(12):387-94.
O'Donovan, A., et al.
1994 XPG endonuclease makes the 3' incision in human DNA nucleotide excision repair. Nature 371(6496):432-5.
Oberbeck, N., et al.
2014 Maternal aldehyde elimination during pregnancy preserves the fetal genome. Molecular cell 55(6):807-17.
Palom, Y., et al.
2002 Relative toxicities of DNA cross-links and monoadducts: new insights from studies of decarbamoyl mitomycin C and mitomycin C. Chemical research in toxicology 15(11):1398-406.
Palombo, F., et al.
1995 GTBP, a 160-kilodalton protein essential for mismatch-binding activity in human cells. Science 268(5219):1912-4.
Pepe, A., and S. C. West

2014 Substrate specificity of the MUS81-EME2 structure selective endonuclease. Nucleic acids research 42(6):3833-45.

Pingoud, V., et al.
2005 Identification of base-specific contacts in protein-DNA complexes by photocrosslinking and mass spectrometry: a case study using the restriction endonuclease SsoII. Molecular bioSystems 1(2):135-41.

Prado, F., and A. Aguilera
1995 Role of reciprocal exchange, one-ended invasion crossover and single-strand annealing on inverted and direct repeat recombination in yeast: different requirements for the RAD1, RAD10, and RAD52 genes. Genetics 139(1):109-23.

Rajski, S. R., and R. M. Williams
1998 DNA Cross-Linking Agents as Antitumor Drugs. Chemical reviews 98(8):2723-2796.

Raschle, M., et al.
2008 Mechanism of replication-coupled DNA interstrand crosslink repair. Cell 134(6):969-80.

Ravanat, J. L., T. Douki, and J. Cadet
2001 Direct and indirect effects of UV radiation on DNA and its components. Journal of photochemistry and photobiology. B, Biology 63(1-3):88-102.

Riedl, T., F. Hanaoka, and J. M. Egly
2003 The comings and goings of nucleotide excision repair factors on damaged DNA. The EMBO journal 22(19):5293-303.

Rink, S. M., and P. B. Hopkins
1995 A mechlorethamine-induced DNA interstrand cross-link bends duplex DNA. Biochemistry 34(4):1439-45.

Roberts, J. A., and M. F. White
2005 An archaeal endonuclease displays key properties of both eukaryal XPF-ERCC1 and Mus81. The Journal of biological chemistry 280(7):5924-8.

Rosado, I. V., et al.
2011 Formaldehyde catabolism is essential in cells deficient for the Fanconi anemia DNA-repair pathway. Nature structural & molecular biology 18(12):1432-4.

Sarbajna, S., D. Davies, and S. C. West
2014 Roles of SLX1-SLX4, MUS81-EME1, and GEN1 in avoiding genome instability and mitotic catastrophe. Genes & development 28(10):1124-36.

Sartori, A. A., et al.
2007 Human CtIP promotes DNA end resection. Nature 450(7169):509-14.

Scharer, O. D.
2003 Chemistry and biology of DNA repair. Angewandte Chemie 42(26):2946-74.

Sengerova, B., et al.
2012 Characterization of the human SNM1A and SNM1B/Apollo DNA repair exonucleases. The Journal of biological chemistry 287(31):26254-67.

Sgouros, J., P. H. Gaillard, and R. D. Wood
1999 A relationship betweena DNA-repair/recombination nuclease family and archaeal helicases. Trends in biochemical sciences 24(3):95-7.

Shao, X., and N. V. Grishin

2000 Common fold in helix-hairpin-helix proteins. Nucleic acids research 28(14):2643-50.

Shiloh, Y.
2003 ATM and related protein kinases: safeguarding genome integrity. Nature reviews. Cancer 3(3):155-68.

Sijbers, A. M., et al.
1996 Xeroderma pigmentosum group F caused by a defect in a structure-specific DNA repair endonuclease. Cell 86(5):811-22.

Sip, M., et al.
1992 Distortions induced in DNA by cis-platinum interstrand adducts. Biochemistry 31(9):2508-13.

Smeaton, M. B., et al.
2008 Distortion-dependent unhooking of interstrand cross-links in mammalian cell extracts. Biochemistry 47(37):9920-30.

Smogorzewska, A., et al.
2010 A genetic screen identifies FAN1, a Fanconi anemia-associated nuclease necessary for DNA interstrand crosslink repair. Molecular cell 39(1):36-47.

Sobol, R. W., et al.
1996 Requirement of mammalian DNA polymerase-beta in base-excision repair. Nature 379(6561):183-6.

Staresincic, L., et al.
2009 Coordination of dual incision and repair synthesis in human nucleotide excision repair. The EMBO journal 28(8):1111-20.

Stracker, T. H., and J. H. Petrini
2011 The MRE11 complex: starting from the ends. Nature reviews. Molecular cell biology 12(2):90-103.

Sugasawa, K., et al.
2001 A multistep damage recognition mechanism for global genomic nucleotide excision repair. Genes & development 15(5):507-21.

Svendsen, J. M., and J. W. Harper
2010 GEN1/Yen1 and the SLX4 complex: Solutions to the problem of Holliday junction resolution. Genes & development 24(6):521-36.

Tripsianes, K., et al.
2005 The structure of the human ERCC1/XPF interaction domains reveals a complementary role for the two proteins in nucleotide excision repair. Structure 13(12):1849-58.

Trujillo, J. P., et al.
2012 On the role of FAN1 in Fanconi anemia. Blood 120(1):86-9.

Tsodikov, O. V., et al.
2005 Crystal structure and DNA binding functions of ERCC1, a subunit of the DNA structure-specific endonuclease XPF-ERCC1. Proceedings of the National Academy of Sciences of the United States of America 102(32):11236-41.

Vare, D., et al.
2012 DNA interstrand crosslinks induce a potent replication block followed by formation and repair of double strand breaks in intact mammalian cells. DNA repair 11(12):976-85.

Walker, J. R., R. A. Corpina, and J. Goldberg
2001 Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. Nature 412(6847):607-14.

Wang, A. T., et al.
2011 Human SNM1A and XPF-ERCC1 collaborate to initiate DNA interstrand cross-link repair. Genes & development 25(17):1859-70.
Wang, A. T., and A. Smogorzewska
2015 SnapShot: Fanconi anemia and associated proteins. Cell 160(1-2):354-354 e1.
Wang, R., et al.
2014 DNA repair. Mechanism of DNA interstrand cross-link processing by repair nuclease FAN1. Science 346(6213):1127-30.
Wang, W.
2007 Emergence of a DNA-damage response network consisting of Fanconi anaemia and BRCA proteins. Nature reviews. Genetics 8(10):735-48.
Wang, X., et al.
2001 Involvement of nucleotide excision repair in a recombination-independent and error-prone pathway of DNA interstrand cross-link repair. Molecular and cellular biology 21(3):713-20.
Wiederhold, L., et al.
2004 AP endonuclease-independent DNA base excision repair in human cells. Molecular cell 15(2):209-20.
Wilson, D. M., 3rd, and D. Barsky
2001 The major human abasic endonuclease: formation, consequences and repair of abasic lesions in DNA. Mutation research 485(4):283-307.
Wu, L., and I. D. Hickson
2006 DNA helicases required for homologous recombination and repair of damaged replication forks. Annual review of genetics 40:279-306.
Wu, Y.
2012 Unwinding and rewinding: double faces of helicase? Journal of nucleic acids 2012:140601.
Yamamoto, K. N., et al.
2011 Involvement of SLX4 in interstrand cross-link repair is regulated by the Fanconi anemia pathway. Proceedings of the National Academy of Sciences of the United States of America 108(16):6492-6.
Yan, Z., et al.
2010 A histone-fold complex and FANCM form a conserved DNA-remodeling complex to maintain genome stability. Molecular cell 37(6):865-78.
Yang, W., J. Y. Lee, and M. Nowotny
2006 Making and breaking nucleic acids: two-Mg2+-ion catalysis and substrate specificity. Molecular cell 22(1):5-13.
Yata, K., and F. Esashi
2009 Dual role of CDKs in DNA repair: to be, or not to be. DNA repair 8(1):6-18.
Yi, C., et al.
2010 Iron-catalysed oxidation intermediates captured in a DNA repair dioxygenase. Nature 468(7321):330-3.
Yoo, S., and W. S. Dynan
1999 Geometry of a complex formed by double strand break repair proteins at a single DNA end: recruitment of DNA-PKcs induces inward translocation of Ku protein. Nucleic acids research 27(24):4679-86.
Zamble, D. B., et al.

1996 Repair of cisplatin--DNA adducts by the mammalian excision nuclease. Biochemistry 35(31):10004-13.

Zaychikov, E., et al.
1996 Mapping of catalytic residues in the RNA polymerase active center. Science 273(5271):107-9.

Zhang, J., et al.
2015 DNA interstrand cross-link repair requires replication-fork convergence. Nature structural & molecular biology 22(3):242-7.

Zhang, J., and J. C. Walter
2014 Mechanism and regulation of incisions during DNA interstrand cross-link repair. DNA repair 19:135-42.

Zheng, H., et al.
2003 Nucleotide excision repair- and polymerase eta-mediated error-prone removal of mitomycin C interstrand cross-links. Molecular and cellular biology 23(2):754-61.

Zhou, W., et al.
2012 FAN1 mutations cause karyomegalic interstitial nephritis, linking chronic kidney failure to defective DNA damage repair. Nature genetics 44(8):910-5.

Zou, L., and S. J. Elledge
2003 Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. Science 300(5625):1542-8.