BEYOND THE FANCONI ANEMIA PATHWAY:

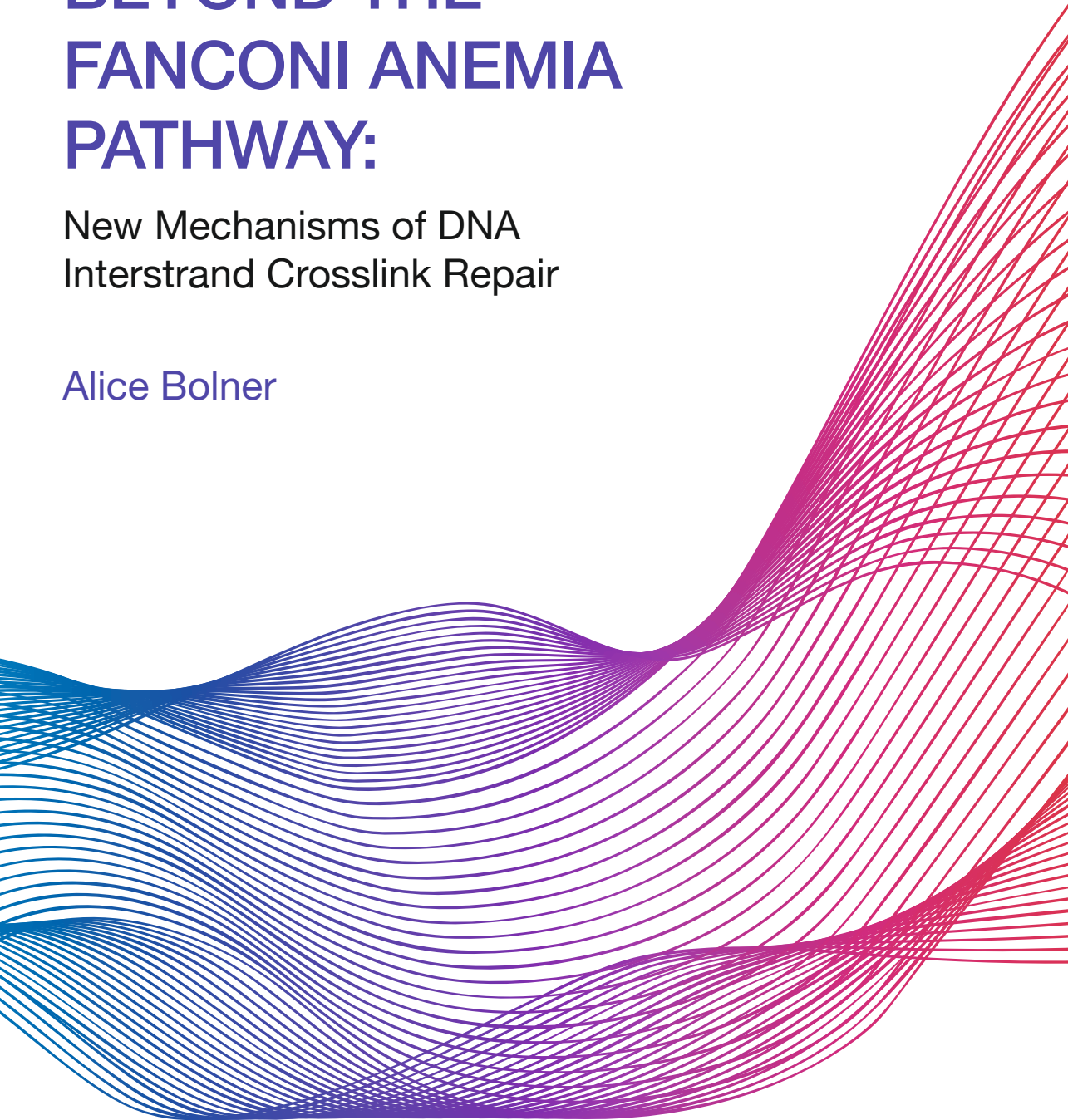## New Mechanisms of DNA Interstrand Crosslink Repair

Alice Bolner

Beyond the Fanconi anemia pathway: new mechanisms of DNA interstrand crosslink repair

Alice Bolner

Beyond the Fanconi anemia pathway: new mechanisms of DNA interstrand crosslink repair

Verder kijken dan de Fanconi anemie route: nieuwe DNA interstrand crosslink reparatie mechanismen

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

donderdag 9 december 2021 des middags te 12.15 uur

door

Alice Bolner

geboren op 9 november 1990
te Trento, Italië

Promotor:
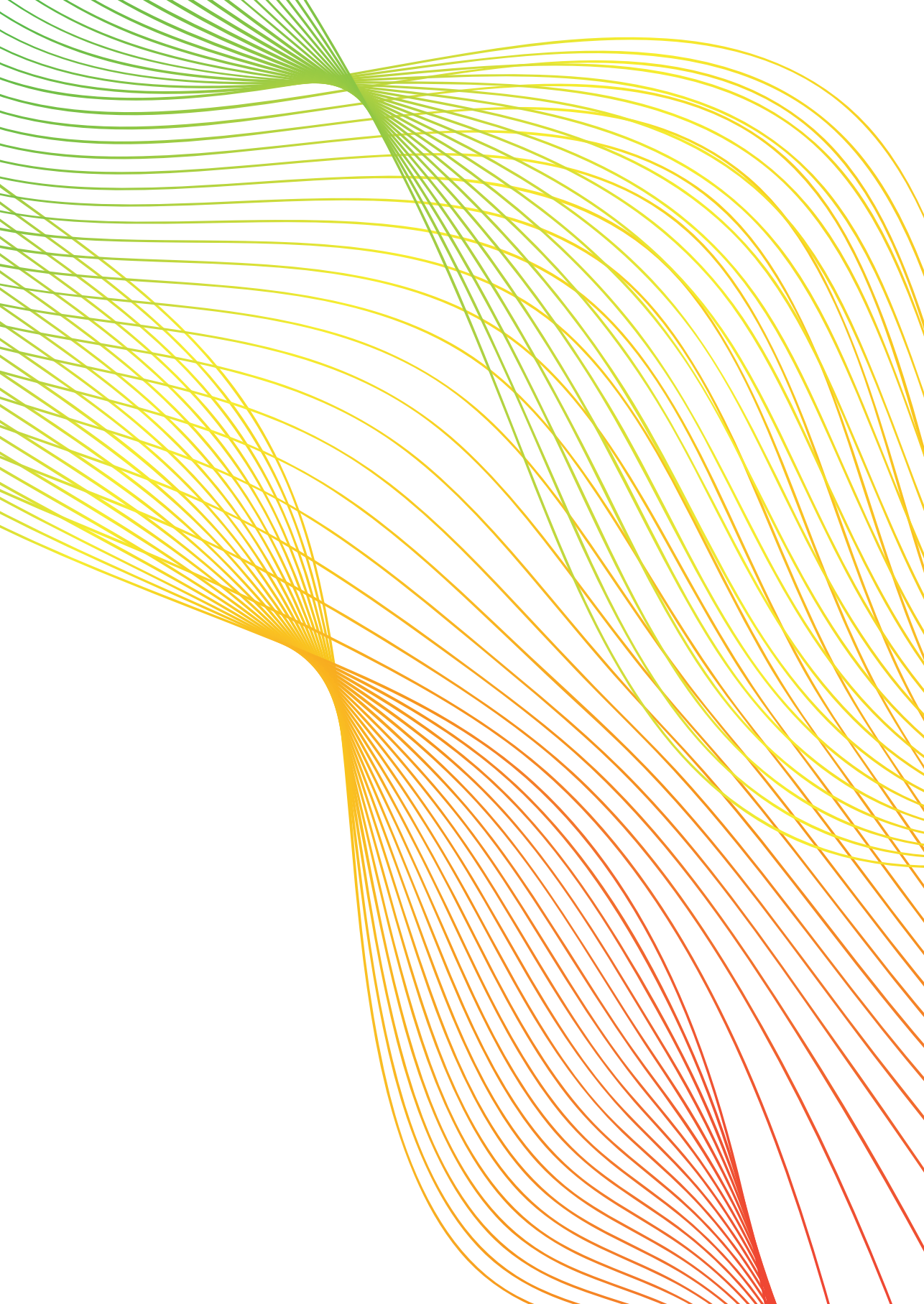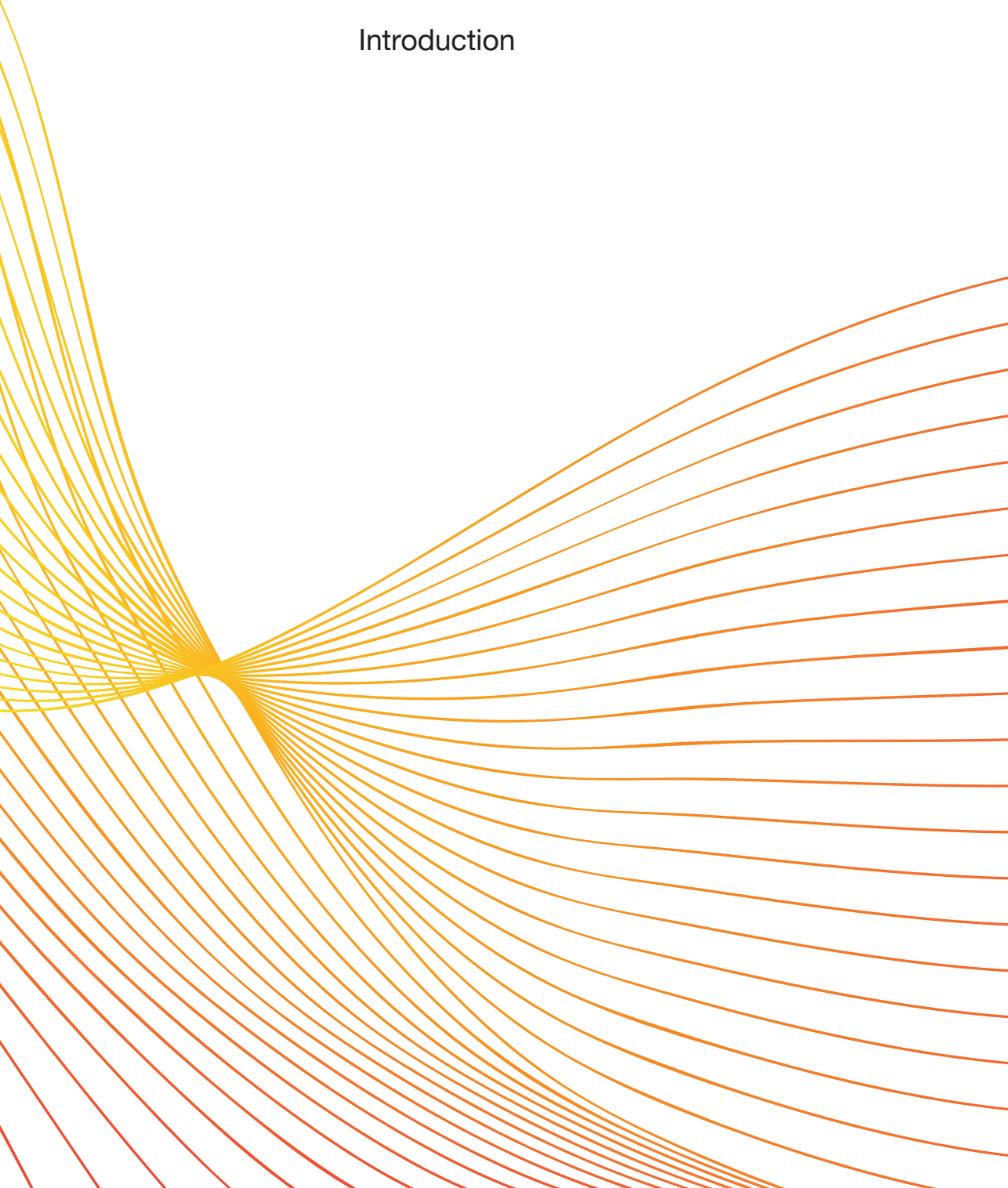Prof. dr. G.J.P.L. Kops

Copromotor:
Dr. P. Knipscheer

Table of contents

CHAPTER 1

Introduction

Table of Contents

1 (margin tab)

1.   Genome maintenance: the DDR

The maintenance of genetic sequence information is vital for living organisms and the perpetuation of life. However, DNA is highly reactive and susceptible to damage by a wide range of both exogenous and endogenous agents. DNA damage is defined as an abnormal chemical modification of DNA, which is not compatible with faithful preservation and transmission of the genetic sequence. If left unrepaired, DNA damage can lead to mutagenesis or cell death, and ultimately it can affect the organism by causing diseases such as cancer and neurodegenerative disorders.

There are different types of DNA damage, such as strand breaks, base modifications, abasic sites, and many more. To safeguard DNA integrity, cells are equipped with a sophisticated network of DNA repair and damage signaling pathways, which is referred to as the DNA damage response (DDR). In the DDR, upon detection of DNA damage, the cell activates specialized pathways that arrest the cell cycle to allow time for DNA repair[1]. In the following sections, I will briefly introduce the signaling events that take place during the DDR and the DNA damage repair pathways that ensure the removal of DNA lesions. This will be followed by a more thorough description of DNA interstrand crosslink repair, as it is the focus of this thesis.

1.1 ATM/ATR signaling

DNA damage induces a complex signaling cascade that depends on two key kinases: ATM (Ataxia-Telangiectasia Mutated) and ATR (Ataxia-Telangiectasia and Rad3-related)[1]. These factors coordinate the recruitment and assembly of the repair machinery at the site of DNA damage and promote cell cycle arrest. ATM and ATR are respectively activated by DNA double-stranded breaks (DSBs) and RPA (Replication Protein A)-coated single-stranded DNA (ssDNA). ATM/ATR signaling inhibits CDK (cyclin-dependent kinase) activity through phosphorylation of CHK1 and CHK2 (checkpoint kinases 1 and 2), and this results in a cell cycle arrest in the G1/S, intra-S, or G2/M cell cycle phases[1,2]. The blockage of cell cycle progression is necessary to create a time window to complete DNA repair prior to DNA replication or cell division. At the same time, ATM/ATR signaling induces transcription of DNA repair genes and activates DNA-repair proteins by post-translational modifications (PTMs)[3]. Once the DNA lesion is removed, DDR signaling is inactivated and the cell cycle is resumed. If there is too much damage that cannot be repaired, ATM/ATR signaling can promote cellular senescence or apoptosis[1].

1.2 Chromatin structure in the DDR

Genomic DNA is packaged in a condensed nucleoprotein structure called chromatin. To facilitate access of the repair machinery to damaged DNA, the chromatin structure is altered at sites of DNA damage[4]. A well-studied example of a DNA damaged-induced chromatin modification is the phosphorylation of serine-139 of

the histone H2A variant H2AX at the site of DSBs. This phosphorylation is mediated by ATM, ATR, and DNA-PK (DNA-dependent protein kinase) and it is necessary for the accumulation of repair proteins at the site of the DSB[3] and the activation of checkpoint signaling. Highlighting the importance of H2AX phosphorylation in the DDR, mice deficient in H2AX have cancer predisposition and increased genomic instability[5].

1.3 DNA repair

No single DNA repair pathway can deal with all the different types of DNA lesions that can arise in the genome, therefore cells have evolved a variety of repair pathways that are highly conserved among eukaryotes (Figure 1). In the following sections, I will delineate the major pathways of DNA repair and DNA damage tolerance. I will then provide a detailed description of the mechanisms of DNA interstrand crosslink repair, as this is the topic of this thesis.
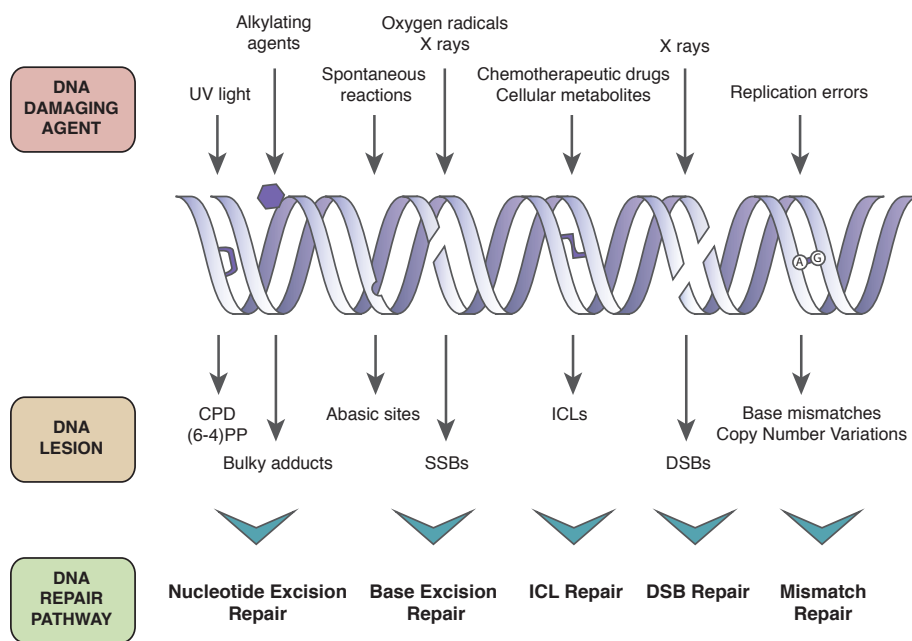
Figure 1: Genome maintenance pathways
The genome is challenged constantly by a wide range of endogenous and exogenous DNA damaging agents. Many types of DNA lesions exist, such as abasic sites, bulky adducts, DNA strand breaks, base mismatches, and interstrand crosslinks. To avoid genome instability, five main DNA repair pathways exist: nucleotide excision repair, base excision repair, double strand break repair, mismatch repair, and DNA interstrand crosslink repair.

Base excision repair

Base excision repair (BER) is a highly conserved pathway involved in the repair of small, non-distorting DNA lesions caused by endogenous processes such as oxidation, deamination, depurination, and alkylation of DNA bases. Moreover, BER is involved in the repair of single-strand breaks (SSBs) (Figure 1)[6]. BER acts by removal of the damaged base and cleavage of the DNA backbone at the site of repair, resulting in a DNA nick. This is then repaired by DNA synthesis and ligation (Figure 2).
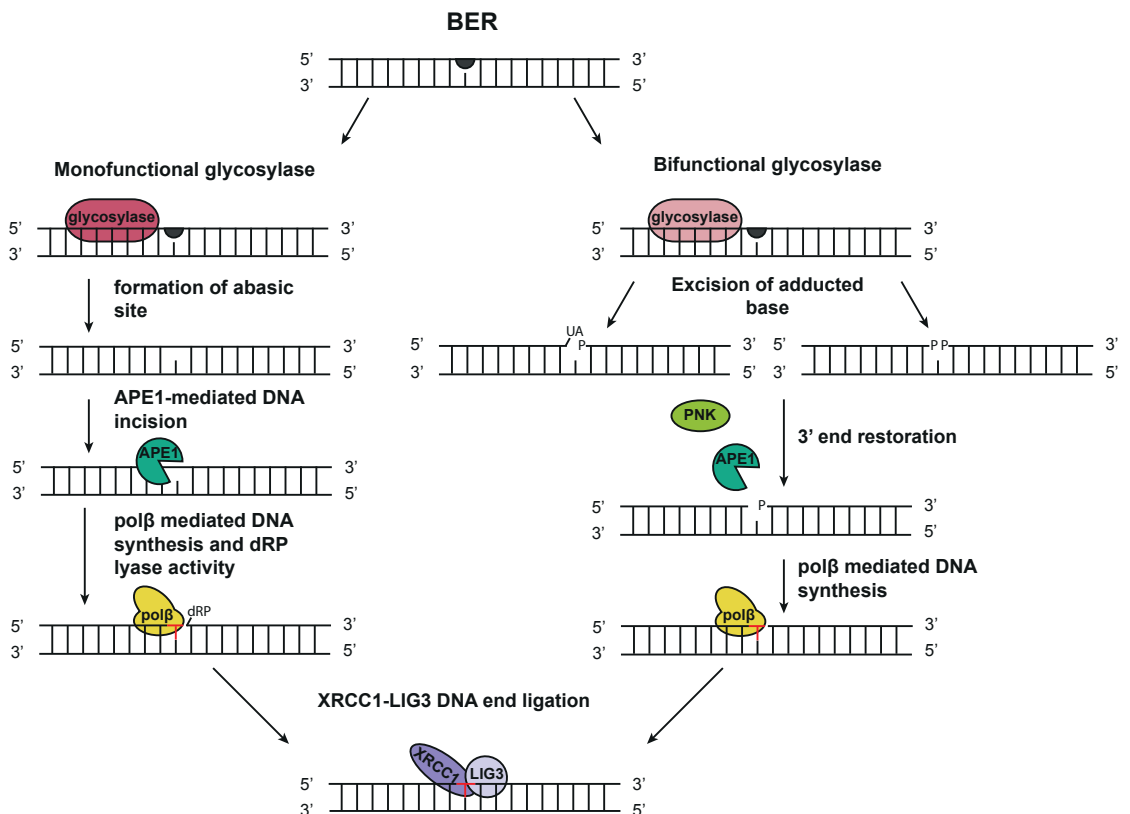
Figure 2: Base excision repair (BER)
BER removes small lesions on DNA bases. Monofunctional glycosylases recognize and excise the damaged DNA base by acting on the *N*-glycosyl bond. This results in the formation of an abasic site, which is removed by APE1-mediated DNA incisions, followed by DNA synthesis and ligation. Bifunctional glycosylases can both recognize the damaged base and incise the DNA, creating a single strand DNA break with phosphate groups (P) or unsaturated aldehydes (UA) at the DNA ends. The OH-group at the 3' end is then restored to allow completion of repair through DNA synthesis and ligation.

The first step in BER is the recognition of DNA damage by DNA glycosylases. In humans, there are 11 glycosylases that together can recognize many different DNA base adducts. Glycosylases can be divided into two groups: monofunctional and bifunctional glycosylases. Monofunctional glycosylases cleave the *N*-glycosyl bond between the DNA base and its sugar, resulting in an abasic site[6-8](Figure 2, left panel). APE1 (apurinic/apyrimidinic endonuclease 1) recognizes this abasic site and cleaves the DNA backbone, resulting in a nick with a 5' sugar and 3' OH-group[6,9]. This is then recognized by Polβ (DNA polymerase β) which removes the sugar through its lyase activity and inserts the missing nucleotide[10,11]. Finally, repair is completed through DNA ligation by a protein complex composed of XRCC1 (X-ray cross complementation protein 1) and Lig3 (DNA ligase 3α)[6,7] (Figure 2).

Bifunctional glycosylases have the ability to both excise the damaged base and cleave the DNA backbone, forming a nick with either a 5' OH-group and a 3' unsaturated aldehyde or a phosphate[8] (Figure 2, right panel). Either APE1 or PNK (polynucleotide kinase) restores the 3' end to an OH-group, allowing repair synthesis by Polβ. Repair is finalized by ligation, as described above. These processes are called short patch BER since the DNA gap can be sealed by a single nucleotide. If the sugar is not efficiently removed from the 5' end of the DNA nick, the mechanism can be redirected to long patch BER. In this mechanism, a polymerase (either a replicative polymerase or Polβ) displaces the damaged strand creating a 5' DNA flap. The flap is then removed by FEN1 (Flap endonuclease 1) and the DNA nick is sealed by Lig1 (Ligase 1)[6-8].

Nucleotide Excision Repair

Nucleotide excision repair (NER) is a versatile DNA repair pathway that removes a wide range of more bulky DNA lesions such as cyclobutane–pyrimidine dimers (CPDs), 6–4 pyrimidine–pyrimidone photoproducts (6–4PPs), cyclopurines, and intrastrand crosslinks derived from alkylating agents (Figure 1)[12]. In contrast to BER, which only excises the damaged base, NER acts by removing a stretch of DNA including the damaged nucleotide through nucleolytic incisions in the DNA backbone. This generates a ssDNA tract which is then filled by DNA synthesis (Figure 3).
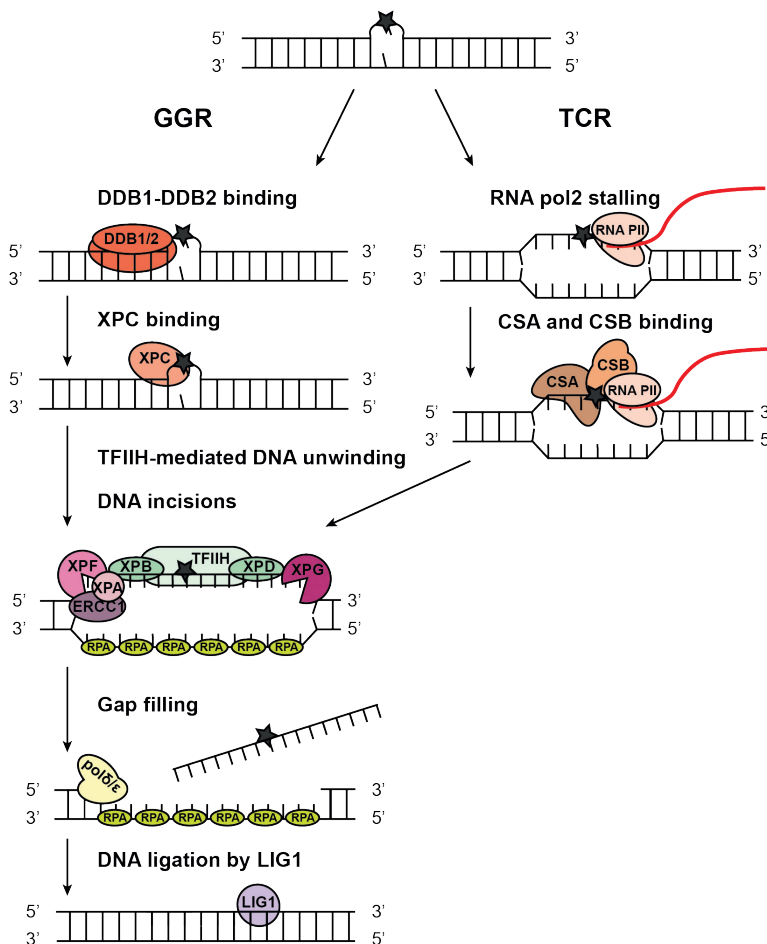
Figure 3: Nucleotide excision repair (NER)
NER acts on bulky DNA adducts that cannot be removed by BER. In **global genome repair** (GGR) the distortion of the helix caused by the DNA lesion is sensed by DDB1-DDB2 and XPC. In **transcription-coupled repair** (TCR), the DNA adduct is recognized through the stalling of the RNA polymerase, which promotes CSA and CSB binding. The two pathways converge with the recruitment of the helicase TFIIH, XPD, XPB which recruit the nucleases XPA, XPG and XPF-ERCC1. The nucleases incise the DNA around the damaged lesion, allowing the eviction of a piece of ssDNA containing the damage. This creates a gap that is then filled by DNA synthesis.

NER has two branches: global genome NER (GG-NER), which is active genome wide, and transcription coupled NER (TC-NER), which only acts in transcriptionally active regions of the genome (Figure 3)[12,13]. In GG-NER the DNA damage is recognized by a complex consisting of XPC (Xeroderma pigmentosum group C), RAD23B, and CETN2 (Centrin 2)[14,15]. This complex recognizes the destabilization of the DNA duplex rather than the DNA lesion itself and is, therefore, able to detect a wide range of DNA adducts. Lesions that do not significantly destabilize the DNA

duplex are first recognized by the DDB1-DDB2 complex (DNA Damage Binding protein 1 and 2), and subsequently by XPC[13]. TC-NER is initiated by the stalling of RNA Pol II (RNA polymerase II) at the DNA lesion, which results in the recruitment of CSA and CSB (Cockayne syndrome protein A and B) (Figure 3)[12,16]. These two proteins promote the assembly of the TC-NER machinery, which includes UVSSA (UV-stimulated scaffold protein A), USP7 (Ubiquitin carboxyl-terminal hydrolase 7), and XAB2[16]. These proteins are thought to induce backtracking of RNA polymerase II, rendering the lesion accessible for repair.

After lesion recognition, the two branches of the pathway converge when the TFIIH (general transcription and DNA repair factor IIH) complex is recruited to the DNA lesion. TFIIH contains the DNA helicases XPB and XPD, which have opposite polarities and can open the DNA duplex around the lesion. The helicase activity of the TFIIH complex together with XPA (Xeroderma pigmentosum group A) and RPA are necessary to verify the lesion[17,18]. The open structure of the DNA, together with the TFIIH complex, XPA and RPA, provide the structure required for recruiting the structure-specific endonucleases XPG (Xeroderma pigmentosum group G) and XPF-ERCC1 (Xeroderma pigmentosum group F and excision repair cross-complementation group protein 1). The first nucleolytic incision is carried out by the XPF-ERCC1 complex, which cleaves 5' to the damage, followed by an XPG-mediated incision at 3' side[19]. These incisions allow the eviction of a small piece of single stranded DNA containing the damaged nucleotide creating a DNA gap of 22-30 nucleotides. This gap is filled by DNA polymerase δ (Polδ), ε (Polε) or κ (Polκ). Finally, the remaining nick is ligated by Lig1 (DNA ligase 1)[12,20].

Mismatch Repair

DNA mismatches are defined as non-Watson-Crick base pairs and mostly occur during DNA replication when a polymerase incorporates the wrong nucleotide[21]. A mismatch, if not promptly repaired, is converted to a mutation during DNA replication in the S-phase of the following cell cycle, and can therefore affect the physiology of the cell. Lack of efficient mismatch repair (MMR) is associated with cancer predisposition[22], underlining the importance of this DNA repair mechanism for genome maintenance.

The first step in MMR is damage recognition by MutSα or MutSβ. The MutSα complex (which is composed of MSH2 and MSH6) can recognize base mismatches and small deletions/insertions, while MutSβ recognizes larger deletions/insertions[23,24]. After mismatch recognition, the MutLα complex, which is composed of MLH1 and the PMS2 endonuclease, is recruited to the damage and incises the nascent DNA strand upon activation by PCNA (proliferating cell nuclear antigen)[23,25]. Because a mismatch contains undamaged bases, it is essential to recognize the newly synthesized DNA strand as the strand that needs to be repaired, to avoid mutations. The newly synthesized DNA strand is thought to be recognized because of its discontinuities in mechanisms that involve the gaps between

1

Okazaki fragments, or by nicks created during the excision of ribonucleotides[26,27]. Alternatively, strand discrimination is achieved through the interaction between MutLα and PCNA[24-26,28]. Then, EXO1 (exonuclease 1) is loaded and degrades DNA in a 5′ to 3′ direction from the nick, excising the mismatch[29]. An alternative mechanism to excise the mismatch is through the strand-displacement activity of Polδ or Polε[29,30] or by their 3' to 5' exonuclease activity [31]. The last stage of MMR is the ligation of the nicks by Lig1[21].

Double Strand Break repair
DNA double strand breaks (DSBs) are extremely toxic DNA lesions that can result in chromosomal rearrangements, insertions, and deletions. DSBs can derive either from exposure to exogenous sources such as ionizing radiation, or they can form during cellular processes, such as immunoglobulin class switching and meiosis[32]. Aberrant DNA topoisomerase reactions, which can occur spontaneously or upon exposure to specific chemotherapeutic agents, can also cause DSBs[33]. Finally, DSBs can form in conditions of prolonged DNA replication stress that cause replication fork collapse[34-36]. DSBs can be repaired by multiple mechanisms and the choice of repair mechanism depends on the phase of the cell cycle and DNA resection at the DSB ends.

Non-homologous end joining
Non-homologous end joining (NHEJ) is responsible for the repair of the majority of DSBs outside the S and G2 phases of the cell cycle. The double strand break is first recognized and bound by the KU complex, an ATP-dependent DNA helicase composed of two subunits, KU70 and KU80[37] (Figure 4, left panel). KU recruits the DNA-dependent protein kinase catalytic subunit (DNA-PKcs), which phosphorylates itself, the KU complex, the nuclease Artemis, and XLF (XRCC4-like factor). In the case of blunt DNA ends, the ends are ligated by a complex composed of XRCC4 (X-ray repair cross-complementing protein 4), Lig4 (DNA ligase 4), and XLF. However, in the majority of cases the ends of the DSB are not compatible for direct ligation and have to be processed. This is done by nucleases such as the MRN complex (MRE11/RAD50/NBS1) and the structure specific nuclease Artemis[38,39]. The DSB ends can also be processed either by addition of nucleotides at the level of the break by dTd (terminal deoxynucleotidyl transferase), Polβ, Polλ (DNA polymerase λ), or Polμ (DNA polymerase μ). These remodeling processes allow ligation of the DSB ends and DNA repair, however processing of the DSB ends can result in deletions[38,40,41].
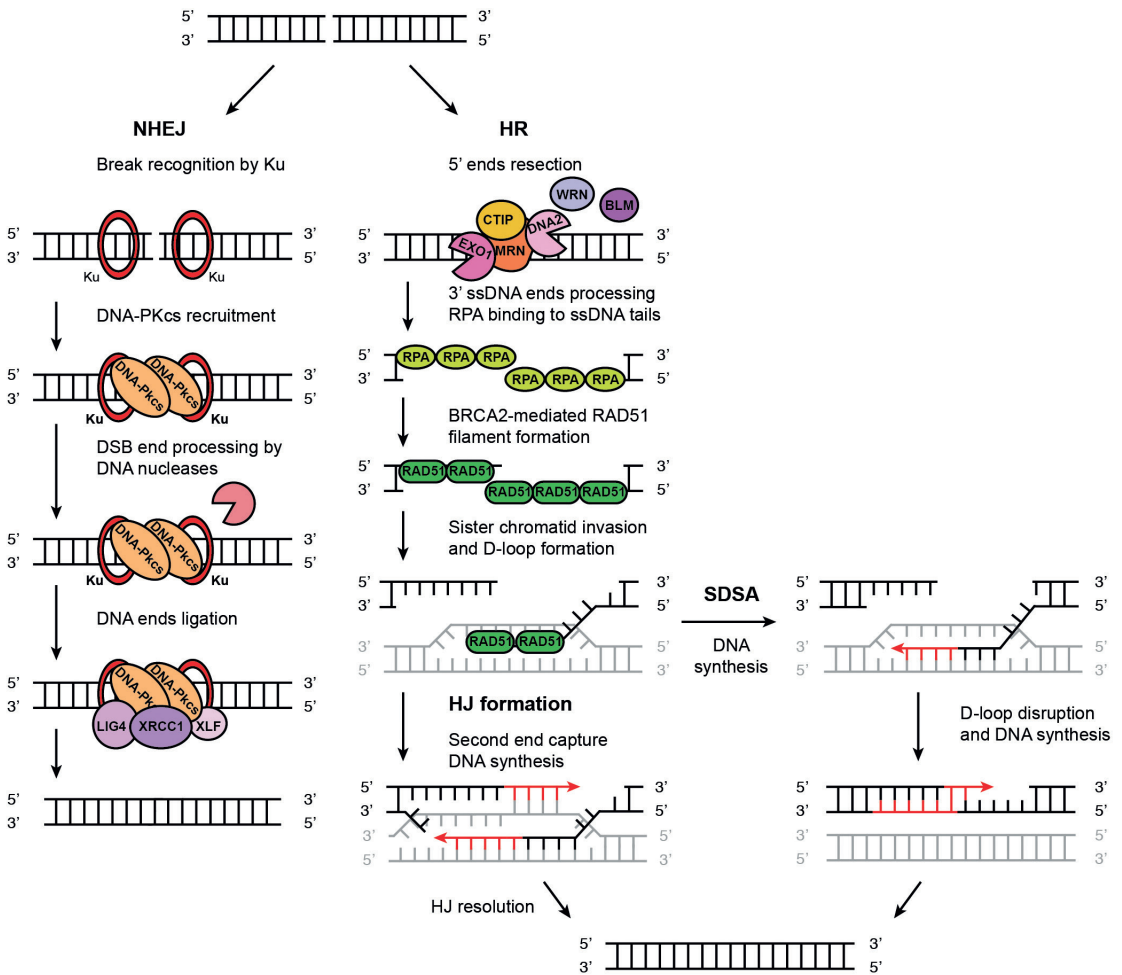
Figure 4: Repair mechanisms of DNA double strand breaks

Double strand breaks (DSBs) are repaired mainly through non-homologous end joining (NHEJ) or homologous recombination (HR). In **NHEJ** the two DSB ends are recognized by the KU proteins, which promotes direct ligation of the DNA ends with minimal to no resection. During **HR** the DSB ends are first resected by the MRN/CTIP complex and then by action of EXO1, WRN, DNA2, and BLM. The resected ssDNA ends are first bound by RPA and then by RAD51. The DNA-RAD51 filaments invade the sister chromatid in search of a region of homology, forming a D-loop. The homologous sequence is used as a template for repairing the DSB, and the D-loop is resolved either through Holliday junction (HJ) resolution or by synthesis dependent strand annealing (SDSA).

Homologous recombination

Homologous recombination (HR) repairs DSBs using the homologous DNA sequence of the sister chromatid as a template. Because of the requirement of the sister chromatid, HR can occur only during S and G2 phases of the cell cycle[42]. HR

is initiated by the processing of the DNA ends to create a ssDNA overhang required for strand invasion in the sister chromatid. This is initiated by the MRN complex and CTIP (CtBP-interacting protein), which incise and resect the DNA to create a short ssDNA overhang[42,43,44] (Figure 4, right panel). EXO1, DNA2, BLM (Bloom syndrome protein), and WRN (Werner syndrome ATP-dependent helicase) act together to further resect the DNA to generate a long 3' ssDNA overhang[45-48] that is bound by the ssDNA binding protein RPA. BRCA2 (Breast cancer type 2 susceptibility protein) promotes the replacement of RPA with ssDNA binding protein RAD51. The resulting RAD51 nucleoprotein filament[49] probes DNA in the sister chromatid for homology, and forms a stable synaptic complex at the homologous sequence. This is followed by invasion of the dsDNA of the sister chromatid by the nucleoprotein filament, creating a D-loop (Figure 4)[44,50]. DNA polymerases can extend the 3' end of the DSB within the D-loop, which is followed by either the disruption of the D-loop and re-annealing of the displaced strand to the other DSB end (Figure 4, synthesis-dependent strand annealing, SDSA), or formation of a double Holliday junction by capture of the second DSB end within the D-loop (Figure 4, Holliday junction -HJ- formation). Holliday junctions are then resolved by the action of helicases or nucleases[51].

Alternative end-joining pathways

Apart from NHEJ and HR, two other pathways can repair DSBs: microhomology-mediated end joining (MMEJ) and single-strand annealing (SSA). These two pathways, like HR, involve resection of the DSB ends, but they do not make use of the sister chromatid to complete repair[38,52].

MMEJ relies on the presence of regions of complementarity ranging from 2 to 20 nucleotides on either side of the DSB. Similar to HR, the DSB ends are processed by the MRN complex and CTIP[53]. The resected ends are then annealed at the microhomology regions through a mechanism that is still poorly understood. After annealing, the non-homologous 3' tails are removed and Polθ (DNA polymerase θ) fills the ssDNA gaps[54].

SSA is the pathway of choice when resection of the DSB ends exposes regions of homology that are longer than 25 nucleotides. The DNA repair protein RAD52 promotes annealing of the homologous sequences, generating a DNA duplex with two non-complimentary 3' tails, which are later removed by the endonuclease complex XPF-ERCC1[55,56]. Repair is finalized through gap filling and ligation[57]. Both MMEJ and SSA can generate genomic rearrangements, insertions, and deletions, and therefore contribute to genomic instability.

Direct reversal

A small set of DNA lesions can be repaired through mechanisms that revert the damage without incision of the DNA backbone or *N*-glycosyl bond. These DNA repair processes involve the restoration of the normal DNA base directly after

DNA damage removal, and they are therefore named 'direct reversal' (or 'direct repair') mechanisms. The main advantage of this repair modality is the avoidance of potentially mutagenic steps.

To date, three mechanisms of direct damage reversal have been identified[58]. The first mechanism is the reversal of UV-induced photolesions. It requires the action of specialized enzymes named photolyases, that use light and the reduced cofactor FADH- (flavin adenine dinucleotide) to revert the damage[58,59]. This DNA damage reversion mechanism is only active in plants and bacteria, not in humans[60]. The second mechanism of direct damage reversal acts on alkyl adducts present in position $O6$ of guanine and thymine. These adducts are detected and repaired by the 6-meG alkyltransferase MGMT, which promotes the transfer of the alkyl group from the DNA to MGMT itself. Once alkylated, the protein is degraded, therefore this mechanism of repair is referred to as 'suicidal'[58,61]. Lastly, N-alkylation DNA damage can be reverted by the AlkB family of proteins. In mammals there are nine members of the AlkB family, however only ALKBH2 and ALKBH3 act in DNA damage reversal through an oxidative dealkylation mechanism[62]. These proteins can act on a variety of DNA adducts such as N1-methyladenine and N3-methylcytosine[58,63] and, in the case of the bacterial enzyme AlkB, on alkyl groups present in position N2 of guanine[63].

1.4 DNA damage tolerance pathways

Replicative DNA polymerases stall at the level of DNA lesions during replication. If the stalling persists, the replication fork can collapse, leading to DSB induction and genomic instability. To avoid this, cells developed a second line of response to DNA damage called DNA damage tolerance, which promotes replication past DNA lesions[64]. In the following paragraphs, I will cover the two mechanisms of DNA damage tolerance: translesion synthesis and template switching.

Translesion synthesis

Translesion synthesis (TLS) involves the replacement of a stalled replicative polymerase with a specialized TLS polymerase, which can accommodate damaged bases and bulky adducts in its active site and can bypass them. In mammals, there are several polymerases involved in TLS: Polη (DNA polymerase η), Polι (DNA polymerase ι), Polκ (DNA polymerase κ) and REV1 all belong to the Y-family of polymerases, while Polζ (DNA polymerase ζ) belongs to the B-family. Polθ (DNA polymerase θ) and Polν (DNA polymerase ν) are part of the A-family of polymerases and display some TLS activity, although it is not their main function[65].

As mentioned above, lesion bypass starts with the replacement of the replicative polymerase with one that supports TLS. Such polymerase switching often depends on monoubiquitination of PCNA that results in the recruitment of a Y-family polymerase[65,66]. The specialized polymerase inserts a short stretch of nucleotides starting from the lesion which is then extended by a second TLS

polymerase. This extension step is necessary for lesion bypass, as it prevents the excision of the newly inserted nucleotides by the proofreading activity of the replicative polymerases. In the final step of TLS, the TLS polymerase is replaced by the replicative DNA polymerase[67]. Since often TLS polymerases lack proofreading activity and have non-restrictive active sites, they are error prone and can result in DNA mutations[65].

Template switching

The stalled replication fork can also use the sister chromatid as a template to bypass a DNA lesion using a process called template switching (TS). Two mechanisms of TS have been reported: the first is lesion skipping, and the second involves a structural rearrangement of the replication fork called fork reversal[68,69].

Lesion skipping occurs when, as a consequence of DNA replication fork stalling, the replicative CMG (CDC45-MCM-GINS) helicase is uncoupled from the rest of the replisome[68]. In this context, a DNA primase can generate a primer downstream of the lesion, which is used to continue DNA replication and thereby skip the DNA lesion. The resulting ssDNA gap is then filled by using the sister chromatid as a template[70-72]. Cellular studies identified PRIMPOL (DNA-directed primase/polymerase) as one of the possible primases acting in lesion skipping of UV-induced DNA lesions[73].

Replication fork reversal is another mechanism of template switching in response to replication stress. When a replication fork is stalled at a DNA lesion, the two parental DNA strands can reanneal together, causing the nascent DNA strand to regress into a four-way junction[68,69]. If the DNA lesion did not block DNA synthesis in the lagging strand, the lagging strand can be used to extend the leading strand and bypass the DNA lesion. Otherwise, fork reversal acts as a mechanism to promote fork stability during DNA repair. Studies identified multiple proteins involved in fork reversal: the fork stabilizers RAD51, BRCA1, and BRCA2 (breast cancer susceptibility protein 1 and 2), and the DNA translocases SMARCAL1 (SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A-like protein 1) and ZRANB3 (Zinc finger Ran-binding domain-containing protein 3)[69,72,74,75]. While more studies are highlighting a role for template switching pathways in genome maintenance, the mechanism of these processes is still poorly understood.

2. Interstrand Crosslink Repair

DNA-interstrand crosslinks (ICLs) are DNA lesions that covalently bind the two strands of the double helix[76]. Because ICLs block strand separation, which is required for DNA replication and transcription, they are among the most toxic DNA lesions. ICL-inducing agents are effective in chemotherapy because they preferentially affect rapidly-dividing cells, such as cancer cells[77].

2.1 ICL inducing agents

Exogenous sources of ICLs
Many drugs used for chemotherapy induce ICLs, such as cisplatin and its derivatives, mitomycin C (MMC), deoxybutane (DEB), psoralens, and nitrogen mustards. These agents cause a multitude of DNA adducts, and ICLs represent just a small percentage of the lesions formed[78]. Yet, ICLs are considered the main cause of toxicity of these drugs.

Although all ICLs are formed by crosslinking the two strands of the DNA, the chemical and structural properties of the various ICLs differ widely. First, the sequence context in which an ICL can form depends on the chemistry of the ICL-inducing agent. For example, cisplatin and MMC induce ICLs in a 5' CG context, while psoralens act in a 5' TA sequence[78]. Secondly, the structure of the ICL and the resulting distortion of the DNA double helix differs depending on the ICL-inducing agent (Figure 5). Cisplatin causes a major distortion of the double helix, forcing the bases opposite the crosslinked Gs to be extrahelical. On the other hand, MMC, psoralens, and nitrogen mustards cause minor distortions in the DNA structure[78].

Endogenous sources of ICLs
Studying ICLs formed endogenously in cells is challenging due to the scarcity of these lesions and the lack of detection methods. However, both *in vivo* and *in vitro* studies identified aldehydes as a possible endogenous source of ICLs. Aldehydes are a highly reactive class of endogenous metabolites that can form a variety of DNA adducts, including ICLs[79] (Figure 5). Acetaldehyde is a product of ethanol oxidation and the metabolism of carbohydrates, and can form ICLs between guanines[80,81]. Formaldehyde is an aldehyde that is highly concentrated in human plasma and is formed during various cellular processes such as dealkylation of methylated DNA and histone demethylation [81-83]. Other ICL-inducing aldehydes are generated through lipid peroxidation, such as acrolein, malondialdehyde, crotonaldehyde, and 4-hydroxynonenal[84].

In **chapter 2** we show that endogenous acetaldehyde-induced ICLs are repaired by two distinct mechanisms.
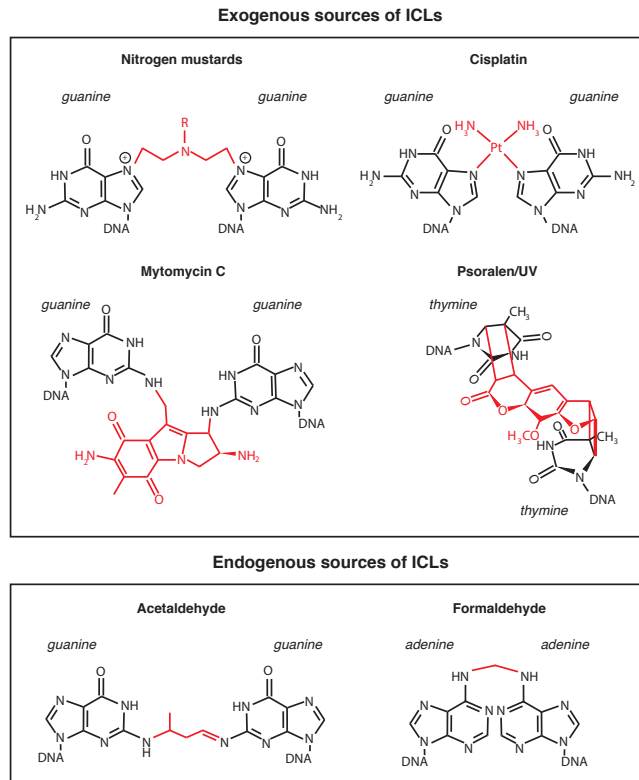
Figure 5: Distinct agents generate structurally diverse interstrand crosslinks
DNA interstrand crosslinks (ICLs) can derive from both endogenous and exogenous sources. Chemical structures of different ICLs that are derived from various agents. The crosslinks are depicted in red, the crosslinked nucleobases in black.

2.2 The Fanconi anemia pathway
The major pathway of ICL repair is the Fanconi anemia (FA) pathway, which comprises DNA repair factors that are also involved in nucleotide excision repair, translesion synthesis, and homologous recombination. The disorder was first identified by the pediatrician Guido Fanconi in 1927, who described three brothers who presented developmental malformations and developed lethal anemia[85]. Cells derived from FA patients are highly sensitive to ICL inducing agents and present genomic instability, supporting the hypothesis that the FA pathway is involved in the repair of ICLs[86].

Fanconi anemia
FA is an autosomal recessive disease that affects one in 150.000 individuals and is caused by mutations in one of the 22 FA genes (named *FANCA, -B, -C, -E, -F,*

-G, -L, -M, -D1, -D2, -I, -J, -N, -O, -P, -Q, -R, -S, -T, -U, -V, -W). The phenotype of the disease differs widely depending on the causative mutation. During their life, FA patients often develop platelet deficiency during childhood, which progresses to pancytopenia and finally leads to bone marrow failure. Many patients also suffer from congenital malformations at birth and are highly predisposed to develop hematological cancers and head and neck cancers later in life[87-89]. At the cellular level, the phenotype is characterized by chromosomal instability (e.g. broken and radial chromosomes upon exposure to ICL inducing agents)[90], stalling in the G2/M phase of the cell cycle, and increased sensitivity to DNA crosslinking agents[86,87]. At the moment, hematopoietic stem cell transplantation is the most effective treatment for bone marrow failure in FA[91]. However, the treatment of cancers associated with FA is still challenging. This is partly due to the fact that many chemotherapeutic drugs induce ICLs and they are therefore extremely toxic to FA patients.

The endogenous source of crosslinks that contribute to the FA phenotype has been elusive for many years, however, a strong genetic link between endogenous aldehydes and FA was recently uncovered[92-96]. Patel and colleagues investigated the relationship between the metabolism of acetaldehyde and FA in mouse models by co-deleting the *Aldh2 (*aldehyde dehydrogenase 2) and *Fancd2* genes[92,94]. Since the enzyme ALDH2 converts acetaldehyde to acetic acid, a lack of ALDH2 induces accumulation of acetaldehyde in cells. Interestingly, single deletion of either *Aldh2* or *Fancd2* in mice did not result in the characteristic hematological defect observed in FA patients. Possibly, this is due to differences between mouse and human metabolism. However, *Aldh2/Fancd2* double knockout mice displayed developmental malformations and were predisposed to developing fatal acute lymphoblastic leukaemia[92,94]. Challenging these mice with ethanol, an acetaldehyde precursor, led to the depletion of the hematopoietic stem cell pool[92] and consequently bone marrow failure[94]. Therefore, a lack of a functional FA pathway combined with a defective acetaldehyde clearance mechanism in mice leads to the FA phenotype. These findings led to the investigation of the relationship between acetaldehyde metabolism and the FA phenotype in a cohort of Japanese FA patients with a dominant negative mutation in *ALDH2*[97]. In agreement with the results from the mice studies, mutations in *ALDH2* increased the severity of the FA phenotype[97]. Similar results were observed in mice lacking FANCD2 and ADH5 (alcohol dehydrogenase 5), an enzyme involved in the clearance of formaldehyde from the cell[96]. These mice developed bone marrow failure and rapid depletion of hematopoietic stem cells. A recent study found that in mice formaldehyde can also be a substrate for ALDH2[98]. Loss of both ALDH2 and ADH5 in mice resulted in accumulation of formaldehyde-induced DNA damage and hematopoietic failure and leukemia[98]. Importantly, the authors identified an inherited bone marrow failure and pre-leukemic syndrome in children with mutations in both *ALDH2* and *ADH5*, indicating that endogenous formaldehyde clearance is essential for hematopoiesis and for limiting genomic instability[98]. These studies raised the possibility of using

aldehyde-targeting drugs for the treatment of Fanconi anemia, such as the aldehyde scavenger metformin, which is currently being tested in a clinical trial[99]. Another novel potential approach for the treatment of FA is the small molecule Alda-1, which activates ALDH2 and leads to the reduction of endogenous acetaldehyde (see also **chapter 5**)[100,101].

Collectively, there is a strong genetic link between the metabolism of aldehydes and Fanconi anemia, indicating a role for endogenous aldehydes in the development of the disease. However, knocking out genes such as *Aldh2* and *Adh5* leads to the accumulation of multiple reactive aldehydes that can induce different types of DNA lesions. Therefore, it is impossible to link FA specifically to acetaldehyde or formaldehyde-induced ICLs using only genetic studies. To show a direct interaction between FA proteins and an endogenous ICL, we recently investigated the repair mechanism of acetaldehyde-induced ICLs using *Xenopus* egg extracts (see **chapter 2**). We demonstrated that the FA pathway repairs acetaldehyde-induced ICLs but also identified a novel repair mechanism that acts on these crosslinks. This novel pathway is further discussed in **chapter 2**. In the following section, I will provide a detailed overview of the mechanism of ICL repair by the FA pathway.

ICL detection and initiation of repair

The first step in ICL repair is the recognition of the DNA lesion. This was proposed to be dependent on the protein FANCM, in complex with FAAP24 (FA-associated polypeptide 24kDa), FAAP16 (FA-associated polypeptide 16kDa), and FAAP10 (FA-associated polypeptide 10kDa)[102-104]. However, the role of FANCM in the activation of the FA pathway was later disputed, since cells lacking FANCM still displayed partial activation of the FA pathway[105,106]. In line with this, *FANCM* is no longer considered a *bona fide* FA gene, since the ICL repair defect found in the *FANCM* patient cell line was later attributed to a second mutation in *FANCA*[107-109]. Therefore, FANCM is believed to be involved in ICL repair but dispensable for the activation of the FA pathway.

These observations raise the possibility that multiple mechanisms exist for sensing the ICL and initiating the FA pathway. In line with this, recent evidence suggests that UHRF1 and UHRF2 (ubiquitin-like with PHD and RING finger domain 1 and 2) are also involved in the recognition of an ICL[110,111]. UHRF1 can bind to DNA containing ICLs, and its deficiency increases the cellular sensitivity to ICL-inducing agents. Cellular studies found that the role of UHRF1 in ICL repair is independent of the FA pathway[110,112]. Given the ability of UHRF1 to bind to the nucleases XPF-ERCC1 and MUS81, cellular studies proposed a model in which UHRF1 recruits the nucleases required for ICL processing in a mechanism independent of the FA pathway[110]. These findings are however in disagreement with a recent report which shows that deletion of either UHRF1 or its binding partner UHRF2 affects recruitment and ubiquitination of FANCD2[111]. Further studies are therefore necessary to establish the mechanism of ICL detection, taking into account that the choice of

ICL-inducing agent determines the chemistry and structure of the crosslink, and possibly affects its detection.

FANCI-FANCD2 monoubiquitination
Core complex structure

The key activating step in the FA pathway is considered to be the monoubiquitination of the FANCI-FANCD2 (ID2) heterodimer by the FA core complex[113,114]. The FA core complex is a multi-subunit ubiquitin ligase composed of 7 FA proteins (FANCA, -B, -C, -E, -F, -G, and -L) and two FA-associated proteins (FAAP20 and FAAP100). The FANCL subunit, which contains the E3-ubiquitin ligase activity within the complex, recruits UBE2T (FANCT), which is the ubiquitin conjugating E2 enzyme[115,116]. Currently, it is assumed that the only target of the FA core complex is the FANCI-FANCD2 heterodimer, which, once monoubiquitinated, promotes ICL unhooking[114,117] (Figure 6). The FA core complex consists of three subcomplexes: FANC-A-G-FAAP20 (AG20 subcomplex), FANC-B-L-FAAP 100 (BL100 subcomplex), and FANC-C-E-F (CEF subcomplex)[118-120]. The structure of the FA core complex was recently solved by cryo-EM[120,121] revealing an asymmetric composition around a dimer of BL100 subcomplexes, which acts as a scaffold for the other FA core complex components.

Mechanism of FANCD2 ubiquitination

Both biochemical reconstitution experiments and structural studies showed that the ID2 heterodimer binds to the CEF subcomplex. This is followed by monoubiquitination of FANCD2 first, and then FANCI, by the combined action of one of the two FANCL subunits and UBE2T (E2 ubiquitin-conjugating enzyme, also FANCT)[118,121]. While *in vitro* FANCD2 could be monoubiquitinated using only recombinant FANCL, UBE2T, E1 enzyme[122], the addition of DNA and FANCI to the reaction robustly stimulated the reaction[123,124]. The importance of DNA binding for FANCD2 monoubiquitination was also shown in cellular studies, in which mutation in the DNA binding domain of *FANCD2* sensitized cells to MMC treatment[125].

For years the mechanistic details of FANCI-FANCD2 monoubiquitination were unknown. One open question was how the core complex could access the ID2 monoubiquitination sites, that are sequestered within the FANCI-FANCD2 interface. Recent structural studies provided evidence that upon binding to DNA, the ID2 complex undergoes structural rearrangements that expose its monoubiquitination sites and therefore facilitate monoubiquitination[121,126]. The structure of the FA core-UBE2T complex bound to both ID2 and DNA showed that these rearrangements juxtapose the active site of UBE2T with the site of FANCD2 monoubiquitination, consistent with the data reporting that FANCD2 is ubiquitinated before FANCI[121]. Once monoubiquitinated, the ID2 complex undergoes structural changes that lock the complex around DNA[121].

Regulation of ID2 monoubiquitination

As mentioned above, *in vitro* studies indicated that FANCI is dispensable for FANCD2 monoubiquitination[122]. However, both *in vitro* and cellular studies showed that ATR-mediated phosphorylation of FANCI stimulates ubiquitination of the ID2 heterodimer by increasing its retention onto DNA and by preventing its de-ubiquitination[124,127,128]. FANCI could also be required for the localization of FANCD2 to the DNA, as purified *chicken* FANCD2 in the absence of FANCI forms a homodimer in a closed conformation that does not allow DNA binding[126]. Even though it is still unclear whether a FANCD2 homodimer can form *in vivo*, the formation of the ID2 heterodimer could be a mechanism within the FA pathway to regulate FANCD2 binding to DNA. Recently, another mechanism for regulating ID2 recruitment to DNA involving the phosphorylation of FANCD2 was proposed[129]. The authors found that CK2 (casein kinase 2) phosphorylates a six-residue cluster on FANCD2, resulting in impaired recruitment of the ID2 complex to DNA and a subsequent defect in ID2 monoubiquitination[129]. Phosphorylation of FANCD2 could therefore prevent undesired recruitment of the ID2 complex to DNA. However, both the role and the mechanistic details of FANCD2 phosphorylation and de-phosphorylation during ICL repair still remain to be established.

Structural consequences of ID2 monoubiquitination

A recent study found that monoubiquitinated ID2 complex bound to fluorescently labeled linear DNA could not bind to competitor DNA, while in the absence of ubiquitination the ID2 complex could exchange DNA[126]. These results indicate that monoubiquitination of the ID2 complex locks the complex onto DNA[126]. This was confirmed by the Cryo-EM structures of monoubiquitinated ID2, which revealed that monoubiquitinated ID2 adopts a closed conformation around the DNA, forming a DNA clamp[126,130,131]. A recent study suggested that the ID2-Ub clamp could slide away from the ICL on DNA, acting as a sliding clamp[131]. This was based on an experiment showing that over time monoubiquitinated ID2 dissociates from linear dsDNA but not from circular dsDNA. The authors proposed that this dissociation from linear DNA is due to the complex sliding off the DNA ends, instead of adopting an open conformation and disengaging from DNA. These findings are in contradiction with the experiment from Alcon and colleagues, which showed that monoubiquitinated ID2 could not exchange efficiently on linear DNA[126]. Recently, a Cryo-EM study found that monoubiquitinated ID2 complex forms large filament-like arrays on plasmid DNA[130]. The authors proposed that this function may serve for the stabilization of the stalled replication forks or chromatin remodeling. Taken together, these studies indicate that monoubiquitination of the ID2 complex induces a structural rearrangement that stabilizes the complex as a clamp around the DNA, but how this promotes subsequent ICL repair steps will need to be further examined.

De-ubiquitination of the ID2 complex

To complete repair, the ID2 complex must also be de-ubiquitinated. Genetic studies in chicken cells indicate that depletion of USP1-UAF1 (ubiquitin carboxyl-terminal hydrolase 1 - USP1 associated factor 1), the complex that de-ubiquitinates FANCD2, induces sensitivity to ICL-inducing agents[132]. In line with this, mice deficient in USP1 display an FA-like phenotype[133,134]. Since monoubiquitination locks the ID2 complex onto DNA[126], de-ubiquitination could be necessary for the release of the complex from DNA and for ID2 recycling[128]. However, biochemical studies have found that USP1-mediated ID2 de-ubiquitination is possible only when the ID2 complex is disengaged from DNA[118]. Therefore, more studies using biochemical reconstitution experiments are needed to elucidate the mechanism of ID2 de-ubiquitination and how this promotes ICL repair.

DNA replication fork convergence, fork remodeling, and CMG unloading

FA-mediated ICL repair requires the convergence of two replication forks at the lesion site[135]. Because the replicative helicase CMG cannot pass through the ICL, the two replication forks stall approximately 20-40 bp before the lesion, consistent with the CMG footprint on DNA[136]. Upon fork convergence, the CMG is unloaded from DNA by the p97 segregase. When fork convergence is inhibited by preventing one of the replication forks from reaching the ICL, the CMG is retained on the DNA and ICL repair is abrogated[135]. Studies in *Xenopus* egg extracts suggested that the mechanism of unloading of the CMG dependent on BRCA1, since the CMG was retained on DNA in extracts depleted of BRCA1[137]. However, a recent study showed that depletion of BRCA1 from extracts co-depletes TRAIP (TRAF-interacting protein)[138], which is the actual mediator of CMG unloading. TRAIP is an E3 ubiquitin ligase that ubiquitinates the CMG helicase, which is subsequently unloaded from DNA by p97[139].

After CMG unloading, one of the two replication forks approaches the lesion, stalling one nucleotide before the ICL ('-1 position')[137,139] (Figure 6). The approach is assumed to be carried out by replicative polymerases[140] and is an essential step during FA-mediated ICL repair[137]. In *Xenopus* egg extracts, inhibition of approach by aphidicolin (a chemical that blocks progression of the B-type polymerases Polδ, Polε, Polα and Polζ) resulted in impairment of ICL unhooking and therefore inhibition of ICL repair[135,137]. Importantly, aphidicolin treatment did not prevent ID2 monoubiquitination, and it is currently unclear how DNA synthesis up to the ICL promotes repair. Fork approach may create a DNA structure that favors ICL unhooking (see 'nucleolytic incisions' below). A recent study using *Xenopus* egg extracts in combination with electron microscopy demonstrated that CMG unloading and approach to the ICL is required for at least a subset of replication forks to undergo fork reversal[141]. The authors proposed that this rearrangement of the replication fork may create the DNA template for unhooking incisions[141], however direct evidence that fork reversal is necessary for ICL repair is still missing.

FANCONI ANEMIA PATHWAY

GLYCOSYLASE PATHWAY
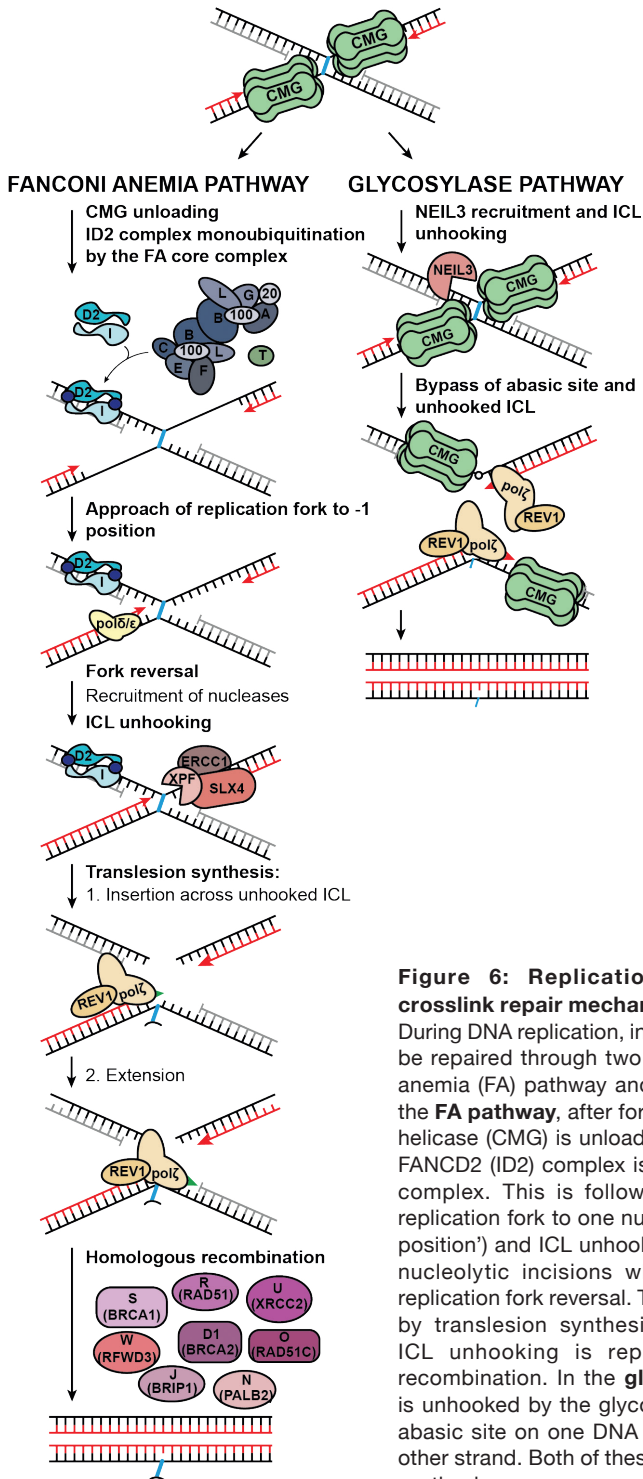
Figure 6: Replication-dependent interstrand crosslink repair mechanisms
During DNA replication, interstrand crosslinks (ICLs) can be repaired through two major pathways: the Fanconi anemia (FA) pathway and the glycosylase pathway. In the **FA pathway**, after fork convergence, the replicative helicase (CMG) is unloaded from DNA and the FANCI-FANCD2 (ID2) complex is ubiquitinated by the FA core complex. This is followed by the approach of one replication fork to one nucleotide upstream the ICL ('-1 position') and ICL unhooking by XPF-ERCC1-mediated nucleolytic incisions which might be mediated by replication fork reversal. The unhooked ICL is bypassed by translesion synthesis, and the DSB formed by ICL unhooking is repaired through homologous recombination. In the **glycosylase pathway**, the ICL is unhooked by the glycosylase NEIL3, resulting in an abasic site on one DNA strand and an adduct on the other strand. Both of these are bypassed by translesion synthesis.

Nucleolytic incisions

Nucleolytic incisions that unhook the ICL from one of the two strands represent a crucial step in ICL repair by the FA pathway. Monoubiquitination of FANCD2 was shown to promote these unhooking incisions[114] (Figure 6). By replicating a plasmid containing an ICL in *Xenopus* egg extracts, Knipscheer and colleagues demonstrated that in the absence of monoubiquitinated FANCD2 unhooking incisions were abrogated, and the replication fork stalled one nucleotide before the ICL ('-1 position')[114]. However, how FANCD2 promotes ICL unhooking is still a subject of debate.

Two models currently exist for ICL unhooking: 1) It can involve two endonucleolytic incisions on either side of the crosslink (Figure 6), which can be performed by the same or different endonucleases, or 2) it can be mediated by one endonucleolytic incision 5' of the ICL followed by 5' to 3' digestion of the DNA past the ICL by an exonuclease. Because FANCD2 and FANCI lack nuclease domains, they are unlikely to be directly involved in endonucleolytic incisions. Work in our laboratory has shown how the structure-specific endonuclease complex XPF-ERCC1 and the scaffold protein SLX4 are required for ICL unhooking during ICL repair[117]. In the XPF-ERCC1 complex, XPF is the subunit with nucleolytic activity while ERCC1 provides substrate specificity[142]. The XPF-ERCC1 complex preferably cleaves DNA structures with a 3' flap such as DNA bubble structures, stem-loop structures, and splayed arms. In all these structures, the XPF-ERCC1 complex preferably incises dsDNA near the junction between double-stranded and single-stranded DNA[143-145]. Lack of functional XPF-ERCC1 complex can sensitize cells to ICL inducing agents[146]. Accordingly, mutations in XPF (FANCQ) are associated with Fanconi anemia[147]. In *Xenopus* egg extracts, immuno-depletion of XPF-ERCC1 results in complete abrogation of ICL unhooking[117]. This indicates that XPF-ERCC1 is either responsible for both nucleolytic incisions, or that the complex is necessary for the first incision in ICL unhooking which is required for the second incision to take place[148,149].

Recruitment of XPF-ERCC1 to the ICL is mediated by the scaffold protein SLX4, which is also a FA protein, FANCP[150]. Accordingly, cells lacking SLX4 display sensitivity to ICL inducing agents[151]. Apart from the XPF-ERCC1 complex, SLX4 binds two other structure-specific endonucleases: SLX1 (SLX1 structure-specific endonuclease subunit) and MUS81-EME1 (Crossover junction endonucleases MUS81-EME1). It is therefore possible that SLX4 could recruit multiple nucleases during ICL unhooking and position two nucleases at either side of the lesion[152]. Previous work in our laboratory has excluded a requirement for SLX1 and MUS81 during ICL repair[117,153]. FAN1 (Fanconi associated nuclease 1), a nuclease that has 5' to 3' exonuclease/endonuclease activity[154], was also proposed to be involved in ICL unhooking. However, in *Xenopus* egg extracts depletion of FAN1 has no effect on ICL repair[117]. Consistently, patients with a mutation in FAN1 develop karyomegalic interstitial nephritis and not Fanconi anemia[155]. An essential role for FAN1 in the FA

pathway can therefore be excluded. The nucleases SNM1A and SNM1B (Sensitive to Nitrogen Mustards 1A and 1B) have also been associated with ICL repair since cells lacking these proteins are sensitive to ICL inducing agents[152,156]. It was proposed that XPF-ERCC1 could make the first nucleolytic incision at the 5' side of the ICL which would allow the loading of SNM1A and SNM1B on DNA. The nucleases could then use their exonuclease activity to digest the ssDNA past the ICL in a 5' to 3' direction, thereby unhooking the crosslink[157]. Nevertheless, more studies are necessary to establish the role of these proteins in ICL unhooking in the context of active DNA replication.

Taken together, the mechanistic details of ICL unhooking are still poorly understood. Future studies are required for the identification of the nucleases involved in ICL unhooking, and the mechanism by which FANCD2 monoubiquitination promotes this important step in ICL repair.

ICL-adduct bypass

ICL unhooking results in the formation of an adducted nucleotide on one of the two strands (Figure 6). Since replicative polymerases cannot synthesize DNA past this adduct, translesion synthesis polymerases are required. Translesion synthesis occurs in two steps: first, a nucleotide is inserted across from the adducted nucleotide ('insertion step'); secondly, another TLS polymerase extends the newly synthesized strand past the adduct ('extension step') (also see 'translesion synthesis' paragraph). This process prevents fork stalling at the cost of replication fidelity.

Cells deficient for either Polζ or REV1 display increased sensitivity to ICL-inducing agents, indicating a role for these TLS polymerases in the FA pathway[158]. REV1 is a dCMP transferase that was shown to function at abasic sites and modified guanines[159]. Given its ability to interact with many other TLS polymerases (such as Polη, Polκ, Polι, and Polζ) as well as with PCNA, it is proposed that REV1 may be a scaffold protein that facilitates polymerase switching[160,161]. Polζ belongs to the B-family of polymerases and is composed of two subunits, REV7 and REV3, and the accessory factors PolD2 and PolD3. In agreement with the cellular data, depletion of either REV7 or REV1 impairs bypass of the unhooked cisplatin ICL in *Xenopus* egg extracts[140,162]. Specifically, depletion of any of the two TLS polymerases affects the extension step in TLS[140,162]. The two TLS polymerases are recruited during cisplatin-ICL repair through the FA core complex, by a mechanism that does not depend on FANCD2[140,163].

It is currently unclear which polymerase acts in the mutagenic insertion step of cisplatin-ICL repair[140]. Interestingly, depletion of REV7 in *Xenopus* egg extracts does not affect the repair of nitrogen mustards-ICL. This finding raises the possibility that the chemical structure of the ICL determines which TLS polymerases are recruited to the crosslink, impacting mutagenesis[162,164]. In agreement with this, in **chapter 2** we described a role for REV1-Polζ in the insertion step of acetaldehyde-induced

ICL repair, while in cisplatin-induced ICL repair these polymerases are involved in extension.

Homologous recombination and adduct removal

In addition to an adducted strand, ICL unhooking incisions also create a double strand break on the other strand, which is repaired through homologous recombination in the FA pathway (see Figure 6 and 'Double Strand Break repair'). Many FANC proteins are also involved in canonical homologous recombination, such as FANCD1 (BRCA2), FANCJ (BRIP1), FANCN (PALB2), FANCO (RAD51C), FANCR (RAD51), FANCS (BRCA1), FANCU (XRCC2), FANCW (RFWD3)[165-168]. However, it is thought these proteins have specific roles in FA pathway-dependent HR. This was observed studying a dominant-negative mutation in RAD51 which disrupts ICL repair but not general HR, indicating a difference in the mechanism for the two pathways[169]. The protein HSF2BP provides another example of how the mechanism of canonical and FA-dependent HR differ. HSF2BP is a BRCA2-interacting protein that is involved in HR during meiosis. Ectopic expression of HSF2BP sensitized cells to ICL-inducing agents but not to ionizing radiation[170]. Biochemical experiments showed that this protein removes BRCA2 from the ICL site, abolishing HR[170].

The second product of unhooking incisions is the crosslinked adduct on one of the two DNA strands (Figure 6), which is first bypassed by TLS and then must be removed from DNA. This adduct is believed to be a substrate for canonical NER[171], although the precise mechanism of adduct removal is still unknown. Possibly, the adduct is removed in the following cell cycle, which could explain why adduct removal is inefficient in *Xenopus* egg extracts[162].

Additional functions of FA proteins

Various studies found additional roles of some FA proteins outside of the FA pathway of ICL repair. The ID2 heterodimer acts also in the protection of stalled replication forks by preventing fork degradation and avoiding inappropriate firing of dormant replication origins[172-176]. It is still unclear whether FANCD2 ubiquitination is necessary for its roles in fork protection. The FANC proteins BRCA1, BRCA2, and RAD51 have also been implicated in the protection of stalled replication forks[172,173,177]. Moreover, FANCI promotes dormant origin firing in response to low levels of hydroxyurea, in a mechanism that is independent of the FA-core complex[174]. In agreement with a role for FA proteins in the protection of stalled replication forks, FA patient cells exhibit high levels of chromosomal breaks at fragile sites[178]. The FA proteins FANCD2, -A, -L, -M, and BRCA2 also protect cells from the accumulation of DNA:RNA hybrids (R-loops) that form during transcription and can induce genomic instability[179,180]. Moreover, experiments in mice lacking FANCD2 have suggested an additional role for the protein in mitochondria homeostasis[181]. These additional roles of FA proteins in genome maintenance mechanisms could contribute to the high cancer incidence

in FA patients. Further investigation of these mechanisms could contribute to the understanding of the complex phenotype of the disease Fanconi anemia.

2.3 Alternative mechanisms of ICL repair

For many years the FA pathway was believed to be the only repair route acting on ICLs. However, recently additional pathways have been identified that repair. In the following paragraphs, I will provide an overview of the currently known alternative mechanisms of ICL repair.

The glycosylase pathway

Recent work demonstrated that a subset of ICLs can be unhooked by a glycosylase, through cleavage of the N-glycosyl bond between the sugar and the base of a nucleotide[182]. This was first observed by replicating a plasmid containing a psoralen-induced ICL in *Xenopus* egg extracts[182]. The authors observed that, in the absence of the FA pathway, the extracts could still repair the crosslink via the glycosylase route. This new repair modality is faster than the FA pathway and it is dependent on the glycosylase NEIL3. NEIL3 unhooks the ICL by N-glycosyl bond cleavage, resulting in the formation of an abasic site on one DNA strand and an adducted base on the other (Figure 6). Both of these are subsequently bypassed by TLS. NEIL3 also acts on ICLs formed at abasic sites, which are sites where a nucleotide is missing its base. Abasic sites are often found in the genome and they can crosslink with the opposite DNA strand, therefore they are a possible endogenous source of ICLs[182]. Remarkably, when the glycosylase pathway was repressed by depletion of NEIL3, these ICLs were repaired through the FA pathway. This finding has two important implications: first, it indicates that multiple repair pathways can act on the same crosslink; second, pathway switching can occur if one repair route fails.

In agreement with the observations in *Xenopus* egg extracts, loss of NEIL3 in both mouse embryonic fibroblasts (MEF), human cervical carcinoma (HeLa) and human osteosarcoma (U2OS) cells induces sensitivity to psoralen-induced ICLs[183]. Interestingly, *Neil3* knock-out mice do not present any severe phenotype, possibly indicating no major issues in the maintenance of the genome[184]. This result may reflect the low abundance of endogenous ICLs that can be repaired by the glycosylase NEIL3. Alternatively, loss of the glycosylase does not induce genomic instability because the FA pathway can readily take over to repair ICLs.

Ubiquitination of the replicative CMG helicase plays a key role in the mechanism of pathway switching between the NEIL3 and the FA routes[138]. When two replication forks converge at the ICL, the ubiquitin E3 ligase TRAIP ubiquitinates the CMG helicase. Short ubiquitin chains on the CMGs recruit NEIL3 and therefore promote repair through the glycosylase pathway. However, if NEIL3 cannot process the ICL or if NEIL3 is absent, continued TRAIP action creates long ubiquitin chains on the CMG. This results in the p97-mediated unloading of the CMG and promotes repair through the FA pathway.

While the FA pathway is able to act on a variety of ICLs, the glycosylase route avoids the creation of a DSB which decreases the likelihood of inducing large chromosomal rearrangements. Future studies should investigate the relative usage and mutational consequences of repair through the FA and the NEIL3 pathway.

ICL traverse

The notion that ICLs are absolute blocks to DNA replication has been recently challenged by a series of studies[185-187]. The authors developed a fluorescence-based, single-molecule assay that allowed the visualization of both a psoralen-induced ICL and the newly synthesized DNA strand. With this method, they demonstrated that DNA could be synthesized downstream of the ICL without ICL unhooking, indicating that a replication fork could traverse the crosslink. Contrary to the FA and glycosylase pathways, which both require replication fork convergence, fork traverse occurs upon collision of a single replication fork with an ICL. Replication can then quickly resume downstream of the ICL. During fork traverse, FANCD2 and ATR promote the interaction between FANCM and the CMG helicase[186]. This interaction is followed by a structural rearrangement of the CMG and the traverse of the ICL involving the translocase activity of FANCM[186]. While this study did not address further repair steps of the ICL, traverse would lead to the generation of an X-shape structure reminiscent of two converging replications forks. Therefore, subsequent repair steps could be similar to those described for the FA or glycosylase pathways (Figure 6).

The exact mechanism of ICL traverse and the frequency of this process in cells is still unclear and requires further investigation.

Replication-independent ICL repair

ICLs are mainly repaired during the S-phase of the cell cycle[188]. However, replication-independent ICL repair (RIR) has also been described, and it could function in slow or non-diving cells, such as neurons[189-192]. RIR has been studied using various model systems, such as mammalian cells in G1 phase and cell extracts[189-193]. The detection of ICLs outside of S phase can occur through different mechanisms. During transcription, ICLs can be detected by RNA polymerase II stalling (transcription-initiated RIR) and TC-NER proteins[189-191]. In the absence of polymerases colliding with the ICL, the lesion must be sensed by genome maintenance proteins with the ability to recognize structural changes in the double helix, such as NER factors. There are conflicting reports regarding a role for the NER factor XPC in RIR, possibly because the different studies used distinct ICL-inducing agents [192,194]. Following detection, the ICL is thought to be unhooked by the NER proteins XPF and XPG[190,191,195]. Understanding how NER proteins contribute to RIR is further complicated because these factors process intrastrand crosslinks, which are also induced by crosslinking drugs[196].

Recent work in *Xenopus* egg extracts found that MMR proteins can also repair ICLs outside of the S-phase[193]. In this RIR pathway, the MutSα complex detects the ICLs, and MutLα and EXO1 act together in unhooking[193].

Even though multiple mechanisms that can act on an ICL outside of S-phase were described, RIR is still poorly characterized. The distortion caused by the ICL may be an important feature in RIR[193], as it could influence the mechanism of ICL detection and consequently the repair pathway choice. Future studies should take into account the structure and chemistry of ICLs to characterize the molecular mechanism of RIR.

3. Chromatin remodeling during DNA damage repair

In the nucleus, DNA is packed in a nucleoprotein structure called chromatin. The basic unit of chromatin is the nucleosome, which is composed of DNA wrapped around two copies of the four core histones: H2A, H2B, H3 and H4 (Figure 7). The nucleosome is formed by deposition of two H3/H4 dimers onto DNA, followed by incorporation of two H2A/H2B dimers[197]. A mature nucleosome contains a histone octamer that is wrapped by approximately 150 bp of DNA[198,199].
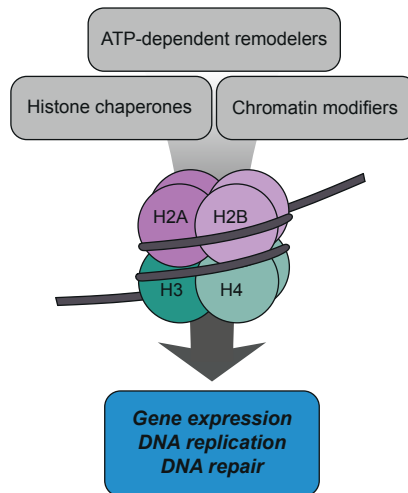
Figure 7: Nucleosome dynamics impact DNA processing.
The fundamental unit of chromatin is the nucleosome, a nucleoprotein structure formed by DNA and histones. In the nucleosome DNA is wrapped around a histone octamer, formed by two H2A-H2B dimers and an H3-H4 tetramer. To allow the binding of regulatory factors to DNA, the structure of chromatin must be altered. This is achieved by the action of multiple factors such as histone chaperones, ATP-dependent remodelers, and chromatin modifiers.

Dynamic changes in the organization of chromatin are essential for all DNA-related processes, as they allow regulatory factors to acces the genetic material. These chromatin changes are referred to as epigenetic modifications, and they provide another layer of regulation for processes like transcription, DNA replication and repair.

The modulation of chromatin conformation can occur through multiple mechanisms. Firstly, small chemical groups (such as phosphate, acetyl, or methyl groups) or small peptides (such as SUMO or ubiquitin) can be added to histone residues. Such chemical modifications affect the interaction between histones and DNA or between multiple nucleosomes, leading to either the opening or tightening of chromatin[200]. Secondly, chromatin structure can be altered by replacing canonical histones by histone variants[201], which can, in turn, affect protein accessibility, histone post-translational modifications (PTMs), and chromatin folding. Lastly, the chromatin landscape can also be actively altered by ATP-dependent remodelers that either evict or slide nucleosomes, or alter the interaction between nucleosomes and DNA[202]. Changes in the activity of chromatin regulators can be deleterious and lead to diseases as cancer, highlighting the importance of chromatin structure for cellular function[203].

In the context of DNA repair, nucleosomes must be rearranged to allow access of repair factors to DNA. Experiments with cells treated with UV light and hydroxyurea demonstrated that the regions of chromatin undergoing active DNA repair were more sensitive to micrococcal DNA nuclease (MNase) digestion. Since this nuclease can digest only non-nucleosomal DNA, the authors concluded that the repair of DNA lesions is associated with a relaxation of chromatin[204]. Based on these findings, the "*a*ccess, *r*epair, *r*estore" model (ARR) was formulated. In the ARR-model, the first steps of DNA repair include the reorganization of nucleosomes to allow repair factors to access the DNA. Upon completion of DNA repair, the canonical nucleosomal organization is restored[205]. These changes allow the recruitment of DNA repair proteins and DNA damage signaling factors, and therefore regulation of the DNA damage response.

The role of chromatin remodeling in DNA repair has been studied for many years, however, there is still a lack of mechanistic information on how histone modifications recruit proteins involved in DNA damage repair and signaling. In **chapter 3** we address how nucleosomes are rearranged during ICL repair. We show that nucleosomes are initially depleted from a region surrounding the crosslink, and they are later placed back in a repair-coupled process.

Currently, virtually nothing is known about the mechanisms of nucleosomal dynamics during ICL repair. Interestingly, the FA protein FANCD2 was reported to have nucleosome-assembly activity *in vitro*[206]. This action is stimulated by FANCI, even though FANCI itself lacks nucleosome assembly activity. Moreover, a mutation in FANCD2 that causes Fanconi anemia (*human* FANCD2[R302W])[207], was defective in histone chaperone activity. Since this mutant could be monoubiquitinated, the

authors proposed that the ICL repair defect was not due to the canonical function of FANCD2 in the FA pathway. Instead, they suggested that the histone chaperone activity of FANCD2 may play a pivotal role in the repair of ICLs. In **chapter 4** we studied the histone chaperone function of *Xenopus* FANCD2 and we observe that the protein can interact with histones but does not act as a histone chaperone during ICL repair.

Other studies have implicated a role for chromatin remodeling in ICL repair. One study reported an interaction between FANCD2 and the chromatin remodeler TIP60 in yeast and human cells[208]. TIP60 is a histone acetyltransferase involved in DSB repair and it can acetylate various DNA repair proteins, including ATM and p53. Knockdown of *TIP60* in fibroblasts sensitized the cells to the ICL-inducing agent MMC, indicating a potential role in ICL repair. The ATPase INO80 was also associated with ICL repair, as INO80-deficient cells are sensitive to MMC treatment[209]. The authors also found that INO80 directly interacted with FANCM, suggesting a possible relationship between INO80 and the FA pathway[209]. However, there is currently no direct evidence that these remodelers affect the chromatin during ICL repair, therefore more investigations are necessary to identify which chromatin remodelers act in the FA pathway, how they interact with FA proteins, and how they affect ICL repair. In **chapter 3** we address this open question using a mass spectrometry-based approach to identify chromatin remodeling factors that localize to the ICL during repair.

4. Study of ICL repair using *Xenopus* egg extracts

Extracts of the eggs of the African clawed frog *Xenopus laevis* are a powerful system for the mechanistic understanding of vertebrate DNA replication and repair (reviewed in [210,211]). These extracts are recovered from *Xenopus* eggs through a series of centrifugation steps and contain the full egg proteome including the proteins required for DNA replication and repair. For the work presented in this thesis, two extracts have been used: the high-speed supernatant (HSS) and the nucleoplasmic extract (NPE). High-speed supernatant *per se* does not support DNA replication, but it can perform replication-independent DNA repair, given its high concentration of DNA repair factors. HSS also allows the binding of histones to DNA and subsequent nucleosome formation, creating a physiological environment for DNA repair[211-213]. Incubating plasmid DNA in HSS promotes the loading of the pre-replication complex (composed of the ORC, CDC6, CDT1, and MCM2-7 subunits) onto the DNA. In cells, this happens in the G1-phase of the cell cycle. To activate replication, NPE is added. NPE is a highly concentrated extract of nuclear proteins and supplies high concentrations of CDK2 (Cyclin-dependent kinase 2), Cyclin E, and other replication factors, allowing initiation of one single round of DNA replication[214].

The *Xenopus* egg extract system offers many advantages for studying ICL repair; firstly, plasmids containing a single chemically defined site-specific ICL can be used[162]. This allows the study of the mechanism of ICL repair in the absence of other lesions, such as intrastrand crosslinks or base adducts, which are often formed upon exposure to ICL-inducing agents and complicate the analysis of cellular experiments. Secondly, since DNA replication and repair occur relatively synchronously on every plasmid, it is possible to isolate and study the repair intermediates formed at various times. Lastly, the absence of membranes allows efficient immunodepletion of DNA repair factors or supplementation of extracts with recombinant proteins or drugs.

In the experimental chapters of this thesis, I will describe novel insights into the mechanisms of ICL repair using the *Xenopus* egg extract system.

5. Thesis outline

DNA interstrand crosslinks are highly toxic DNA lesions that covalently bind the two strands of the DNA double-helix and thereby inhibit cellular processes as DNA replication and DNA repair. For years it was believed that the only replication-dependent ICL repair route was the FA pathway, but recent studies found that the NEIL3 glycosylase can unhook crosslinks induced by Psoralen/UV or that arise at abasic sites[182]. This was the first study to report a mechanism of ICL unhooking that does not involve incisions of the DNA backbone. In **chapter 2** we investigated the mechanism of repair of an endogenous acetaldehyde-induced ICL using the *Xenopus* egg extract system. We found that the lesion can be repaired by the FA pathway, establishing acetaldehyde as a possible source of ICLs that cause Fanconi anemia. Unexpectedly, we found a second mechanism of acetaldehyde-ICL repair, that does not involve FA proteins nor the glycosylase NEIL3. This novel repair route is coupled to DNA replication and involves ICL unhooking through incisions within the crosslink. We therefore provide evidence for a new mechanism of ICL repair that acts on endogenous crosslinks.

Studies on the mechanism of NER and DSB repair showed that nucleosomes are rearranged during repair to promote the recruitment of repair factors and to support DNA damage signaling[215-217]. **In chapter 3** we investigated whether nucleosomes are dynamically rearranged in the context of ICL repair. We found that at the earliest stages of ICL repair nucleosomes are displaced from the region surrounding the ICL. After ICL unhooking, nucleosomes are placed back with a mechanism that is coupled to repair. This study represents the first evidence for nucleosomal rearrangements during ICL repair and provides the unique opportunity to study how this is regulated and how the various stages of repair are coupled to nucleosomal dynamics.

Since previous studies reported FANCD2 acting as a histone chaperone during ICL repair, in **chapter 4** we investigated the nucleosome assembly/disassembly properties of FANCD2. We found no evidence for a direct role of *xl*FANCD2 in nucleosome dynamics in a reconstitution assay, nor during ICL repair, leaving open the question of how nucleosomes are remodeled during ICL repair. In **chapter 5** I discuss the findings of the previous chapters and provide future directions.

1

References

1. Elledge, B.-B.S.Z.S.J. The DNA damage response: putting checkpoints in perspective. *Nature* **408**(2000).

2. Shiloh, Y. ATM and related protein kinases: safeguarding genome integrity. *Nat Rev Cancer* **3**, 155-68 (2003).

3. Huen, M.S. & Chen, J. The DNA damage response pathways: at the crossroad of protein modifications. *Cell Res* **18**, 8-16 (2008).

4. Misteli, T. & Soutoglou, E. The emerging role of nuclear architecture in DNA repair and genome maintenance. *Nat Rev Mol Cell Biol* **10**, 243-54 (2009).

5. Celeste, A. et al. Genomic instability in mice lacking histone H2AX. *Science* **296**, 922-7 (2002).

6. Wallace, S.S. Base excision repair: a critical player in many games. *DNA Repair (Amst)* **19**, 14-26 (2014).

7. Krokan, H.E. & Bjoras, M. Base excision repair. *Cold Spring Harb Perspect Biol* **5**, a012583 (2013).

8. Svilar, D. Goellner, E. Almeida, K & Sobol, R.W. Base Excision Repair and Lesion-Dependent Subpathways for Repair of Oxidative DNA Damage. *Antioxidants & Redox Signaling* **14**(2011).

9. Doetsch, P.V & Cunningham, R.P. The enzymology of apurinic/apyrimidinic endonucleases. *Mutation Research* **236**, 173-201 (1990).

10. Matsumoto, Y. & Kim, K. Excision of Deoxyribose Phosphate Residues by DNA Polymerase B During DNA Repair. *Science* **269**, 699-702 (1995).

11. Prasad, R. Beard, W. Strauss, P. & Wilson, S.H. Human DNA Polymerase B Deoxyribose Phosphate Lyase. *The Journal of Biological Chemistry***273**, 15263–15270 (1998).

12. Marteijn, J.A., Lans, H., Vermeulen, W. & Hoeijmakers, J.H. Understanding nucleotide excision repair and its roles in cancer and ageing. *Nat Rev Mol Cell Biol* **15**, 465-81 (2014).

13. Spivak, G. Nucleotide excision repair in humans. *DNA Repair (Amst)* **36**, 13-18 (2015).

14. Sugasawa, K. et al. HHR23B, a Human Rad23 Homolog, Stimulates XPC Protein in Nucleotide Excision Repair In Vitro. *Molecular and Cellular Biology* **16**, 4852-4861 (1996).

15. Nishi, R. et al. Centrin 2 stimulates nucleotide excision repair by interacting with xeroderma pigmentosum group C protein. *Mol Cell Biol* **25**, 5664-74 (2005).

16. Vermeulen, W. & Fousteri, M. Mammalian transcription-coupled excision repair. *Cold Spring Harb Perspect Biol* **5**, a012625 (2013).

17. Sugasawa, K., Akagi, J., Nishi, R., Iwai, S. & Hanaoka, F. Two-step recognition of DNA damage for mammalian nucleotide excision repair: Directional binding of the XPC complex and DNA strand scanning. *Mol Cell* **36**, 642-53 (2009).

18. Li, C.L. et al. Tripartite DNA Lesion Recognition and Verification by XPC, TFIIH, and XPA in Nucleotide Excision Repair. *Mol Cell* **59**, 1025-34 (2015).

19. Staresincic, L. et al. Coordination of dual incision and repair synthesis in human nucleotide excision repair. *EMBO J* **28**, 1111-20 (2009).

20. Araujo, S. et al. Nucleotide excision repair of DNA with recombinant human proteins: definition of the minimal set of f   actors, active forms of TFIIH, and modulation by CAK. *Genes Dev* **14** 349–359 (2000).

21. Kunkel, T.A. & Erie, D.A. Eukaryotic Mismatch Repair in Relation to DNA Replication. *Annu Rev Genet* **49**, 291-313 (2015).

22. Peltomäki, P. DNA mismatch repair and cancer. *Mutation Research/Reviews in Mutation Research* **488**(2001).

23. Harrington, J.M. & Kolodner, R.D. Saccharomyces cerevisiae Msh2-Msh3 acts in repair of base-base mispairs. *Mol Cell Biol* **27**, 6546-54 (2007).

24. Kunz, C., Saito, Y. & Schar, P. DNA Repair in mammalian cells: Mismatched repair: variations on a theme. *Cell Mol Life Sci* **66**, 1021-38 (2009).

25. Kadyrov, F.A., Dzantiev, L., Constantin, N. & Modrich, P. Endonucleolytic Function of MutLa in Human Mismatch Repair. *Cell* **126**(2006).

26. Pavlov, Y.I., Mian, I.M. & Kunkel, T.A. Evidence for Preferential Mismatch Repair of Lagging Strand DNA Replication Errors in Yeast. *Current Biology* **13**, 744-748 (2003).

27. Ghodgaonkar, M.M. et al. Ribonucleotides misincorporated into DNA act as strand-discrimination signals in eukaryotic mismatch repair. *Mol Cell* **50**, 323-32 (2013).

28. Kadyrov, F. et al. Saccharomyces cerevisiae MutLα is a mismatch repair endonuclease. *J Biol Chem.* **282**, 37181–37190 (2007).

29. Goellner, E.M., Putnam, C.D. & Kolodner, R.D. Exonuclease 1-dependent and independent mismatch repair. *DNA Repair (Amst)* **32**, 24-32 (2015).

30. Farid A. Kadyrov et al. A Possible Mechanism for Exonuclease 1-Independent Eukaryotic Mismatch Repair. *Proceedings of the National Academy of Sciences* **106**, 8495-8500 (2009).

31. Tran, H.T., Gordenin, D.A. & Resnick, M.A. The 3-5' Exonucleases of DNA Polymerases δ and ε and the 5-3' Exonuclease Exo1 Have Major Roles in Postreplication Mutation Avoidance in Saccharomyces cerevisiae. *Molecular and Cellular Biology* **19**(1999).

32. Khanna, K.K. & Jackson, S.P. DNA double-strand breaks: signaling, repair and the cancer connection. *Nature Genetics* **27**(2001).

33. Canela, A. et al. Genome Organization Drives Chromosome Fragility. *Cell* **170**, 507-521 e18 (2017).

34. Zeman, M.K. & Cimprich, K.A. Causes and consequences of replication stress. *Nature Cell Biology* **16**(2014).

35. Hanada, K. et al. The structure-specific endonuclease Mus81 contributes to replication restart by generating double-strand DNA breaks. *Nat Struct Mol Biol* **14**, 1096-104 (2007).

36. Petermann, E., Orta, M.L., Issaeva, N., Schultz, N. & Helleday, T. Hydroxyurea-stalled replication forks become progressively inactivated and require two different RAD51-mediated pathways for restart and repair. *Mol Cell* **37**, 492-502 (2010).

37. Ramsden, A. D. & Gellert, M. Ku protein stimulates DNA end joining by mammalian DNA ligases: a direct role for Ku in repair of DNA double-strand breaks. *The EMBO Journal* **17**, 609–614 (1998).

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

39. Yunmei Ma, Ulrich Pannicke, Klaus Schwarz & Lieber, M.R. Hairpin Opening and Overhang Processing by an Artemis/DNA-Dependent Protein Kinase Complex in Nonhomologous End Joining and V(D)J Recombination. *Cell, Vol.* **108**, 781–794 (2002).

40. Bertocci, B., De Smet, A., Weill, J.C. & Reynaud, C.A. Nonoverlapping functions of DNA polymerases mu, lambda, and terminal deoxynucleotidyltransferase during immunoglobulin V(D)J recombination in vivo. *Immunity* **25**, 31-41 (2006).

41. Gu, J. et al. XRCC4:DNA ligase IV can ligate incompatible DNA ends and can ligate across gaps. *The EMBO Journal* **26**, 1010–1023 (2007).

42. Sung, P. & Klein, H. Mechanism of homologous recombination: mediators and helicases take on regulatory functions. *Nat Rev Mol Cell Biol* **7**, 739-50 (2006).

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

44. Wyman, C. & Kanaar, R. DNA double-strand break repair: all's well that ends well. *Annu Rev Genet* **40**, 363-83 (2006).

45. Mimitou, E.P. & Symington, L.S. DNA end resection--unraveling the tail. *DNA Repair (Amst)* **10**, 344-8 (2011).

46. Serge Gravel, J. Ross Chapman, Christine Magill & Jackson, S.P. DNA helicases Sgs1 and BLM promote DNA double-strand break resection. *Genes & Development* **22**, 2767–2772 (2008).

47. Mimitou, E.P. & Symington, L.S. Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing. *Nature* **455**, 770-4 (2008).

48. Liao, S., Toczylowski, T. & Yan, H. Identification of the Xenopus DNA2 protein as a major nuclease for the 5'->3' strand-specific processing of DNA ends. *Nucleic Acids Res* **36**, 6091-100 (2008).

49. Liu, J., Doty, T., Gibson, B. & Heyer, W.D. Human BRCA2 protein promotes RAD51 filament formation on RPA-covered single-stranded DNA. *Nat Struct Mol Biol* **17**, 1260-2 (2010).

50. Renkawitz, J., Lademann, C.A. & Jentsch, S. Mechanisms and principles of homology search during recombination. *Nat Rev Mol Cell Biol* **15**, 369-83 (2014).

51. Sarbajna, S. & West, S.C. Holliday junction processing enzymes as guardians of genome stability. *Trends Biochem Sci* **39**, 409-19 (2014).

52. Mladenov, E. & Iliakis, G. Induction and repair of DNA double strand breaks: the increasing spectrum of non-homologous end joining pathways. *Mutat Res* **711**, 61-72 (2011).

53. Truong, L.N. et al. Microhomology-mediated End Joining and Homologous Recombination share the initial end resection step to repair DNA double-strand breaks in mammalian cells. *Proc Natl Acad Sci U S A* **110**, 7720-5 (2013).

54. Kent, T., Chandramouly, G., McDevitt, S.M., Ozdemir, A.Y. & Pomerantz, R.T. Mechanism of microhomology-mediated end-joining promoted by human DNA polymerase theta. *Nat Struct Mol Biol* **22**, 230-7 (2015).

55. Rothenberg, E., Grimme, J.M., Spies, M. & Ha, T. Human Rad52-mediated homology search and annealing occurs by continuous interactions between overlapping nucleoprotein complexes. *Proc Natl Acad Sci U S A* **105**, 20274-9 (2008).

56. Motycka, T.A., Bessho, T., Post, S.M., Sung, P. & Tomkinson, A.E. Physical and functional interaction between the XPF/ERCC1 endonuclease and hRad52. *J Biol Chem* **279**, 13634-9 (2004).

57. Bhargava, R., Onyango, D.O. & Stark, J.M. Regulation of Single-Strand Annealing and its Role in Genome Maintenance. *Trends Genet* **32**, 566-575 (2016).

58. Yi, C. & He, C. DNA Repair by Reversal of DNA Damage. *Cold Spring Harb Perspect Biol.* **5**, a012575. (2013).

59. Liu, Z., Wang, L. & Zhong, D. Dynamics and mechanisms of DNA repair by photolyase. *Phys Chem Chem Phys* **17**, 11933-49 (2015).

60. Li, W.F., Kim, S. & Sancar, A. Evidence for lack of DNA photo reactivating enzyme in humans. *Proc Natl Acad Sci U S A* **90**, 4389-4393 (1993).

61. Daniels, D.S. et al. DNA binding and nucleotide flipping by the human DNA repair protein AGT. *Nat Struct Mol Biol* **11**, 714-20 (2004).

62. Duncan, T. et al. Reversal of DNA alkylation damage by two human dioxygenases. *Proc Natl Acad Sci U S A* **99**, 16660-5 (2002).

63. Fedeles, B.I., Singh, V., Delaney, J.C., Li, D. & Essigmann, J.M. The AlkB Family of Fe(II)/alpha-Ketoglutarate-dependent Dioxygenases: Repairing Nucleic Acid Alkylation Damage and Beyond. *J Biol Chem* **290**, 20734-42 (2015).

64. Chang, D.J. & Cimprich, K.A. DNA damage tolerance: when it's OK to make mistakes. *Nat Chem Biol* **5**, 82-90 (2009).

65. Gao, W.Y.a.Y. Translesion and Repair DNA Polymerases: Diverse Structure and Mechanism. *Annual Review of Biochemistry* **87**, 239-261 (2018).

66. Bienko, M. et al. Ubiquitin-Binding Domains in Y-Family Polymerases Regulate Translesion Synthesis. *Science* **310**, 1821-1824 (2005).

67. Shachar, S. et al. Two-polymerase mechanisms dictate error-free and error-prone translesion DNA synthesis in mammals. *EMBO J* **28**, 383-393 (2009).

68. Branzei, D. & Szakal, B. DNA damage tolerance by recombination: Molecular pathways and DNA structures. *DNA Repair* **44**, 68-75 (2016).

69. Marians, K.J. Lesion Bypass and the Reactivation of Stalled Replication Forks. *Annual Review of Biochemistry* **87**, 217-238 (2018).

70. Lopes, M., Foiani, M. & Sogo, J.M. Multiple mechanisms control chromosome integrity after replication fork uncoupling and restart at irreparable UV lesions. *Mol Cell* **21**, 15-27 (2006).

71. Yeeles, J.T.P. & Marians, K.J. The Escherichia coli Replisome Is Inherently DNA Damage Tolerant. *Science* **334**, 235-238 (2011).

72. Fumasoni, M., Zwicky, K., Vanoli, F., Lopes, M. & Branzei, D. Error-free DNA damage tolerance and sister chromatid proximity during DNA replication rely on the Polalpha/Primase/Ctf4 Complex. *Mol Cell* **57**, 812-823 (2015).

73. Garcia-Gomez, S. et al. PrimPol, an archaic primase/polymerase operating in human cells. *Mol Cell* **52**, 541-53 (2013).

74. Vujanovic, M. et al. Replication Fork Slowing and Reversal upon DNA Damage Require PCNA Polyubiquitination and ZRANB3 DNA Translocase Activity. *Mol Cell* **67**, 882-890 e5 (2017).

75. Taglialatela, A. et al. Restoration of Replication Fork Stability in BRCA1- and BRCA2-Deficient Cells by Inactivation of SNF2-Family Fork Remodelers. *Mol Cell* **68**, 414-430 e8 (2017).

76. Noll, D.M., Mason, T.M. & Miller, P.S. Formation and repair of interstrand cross-links in DNA. *Chem Rev* **106**, 277-301 (2006).

77. Deans, A.J. & West, S.C. DNA interstrand crosslink repair and cancer. *Nat Rev Cancer* **11**, 467-80 (2011).

78. Guainazzi, A. & Scharer, O.D. Using synthetic DNA interstrand crosslinks to elucidate repair pathways and identify new therapeutic targets for cancer chemotherapy. *Cell Mol Life Sci* **67**, 3683-97 (2010).

1

79. Stone, M.P. et al. Interstrand DNA cross-links induced by alpha,beta-unsaturated aldehydes derived from lipid peroxidation and environmental sources. *Acc Chem Res* **41**, 793-804 (2008).

80. Wang, M. et al. Identification of DNA Adducts of Acetaldehyde. *Chem Res Toxicol* **13**, 1149-1157 (2000).

81. Loenarz, C. & Schofield, C.J. Expanding chemical biology of 2-oxoglutarate oxygenases. *Nature Chemical Biology* **4**(2008).

82. Yu, F. Chaw, L. Crane, E. Lange, P & Shapiro, R. Isolation and Identification of Cross-Links from Formaldehyde-Treated Nucleic Acids. *Biochemistry* **19**, 5525-5531 (1980).

83. Mosammaparast, N. & Shi, Y. Reversal of histone methylation: biochemical and molecular mechanisms of histone demethylases. *Annu Rev Biochem* **79**, 155-79 (2010).

84. Esterbauer, H., Schaur, R.J. & Zollner, H. Chemistry and Biochemistry of 4-Hydroxynonenal, malonaldehyde and related aldehydes. *Free Radical Biology & Medicine* **11**, 81-128 (1991).

85. Lobitz, S. & Velleuer, E. Guido Fanconi (1892–1979): a jack of all trades. *Nature Reviews Cancer* **6**(2006).

86. Sasaki, M.S. & Tonomura, A. A High Susceptibility of Fanconi's Anemia to Chromosome Breakage by DNA Cross-linking Agents. *Cancer Research* **33**, 1829-1836.

87. Risitano, A.M. et al. Twenty years of the Italian Fanconi Anemia Registry: where we stand and what remains to be learned. *Haematologica* **101**, 319-27 (2016).

88. Glanz, A. & Fraser, F.C. Spectrum of anomalies in Fanconi anaemia. *Journal of Medical Genetics* **19**, 412-416 (1982).

89. Alter, B.P. Cancer in Fanconi anemia, 1927-2001. *Cancer* **97**, 425-40 (2003).

90. Hanlon Newell, A.E. et al. Loss of homologous recombination or non-homologous end-joining leads to radial formation following DNA interstrand crosslink damage. *Cytogenet Genome Res* **121**, 174-80 (2008).

91. Rosenberg, P.S., Alter, B.P., Socie, G. & Gluckman, E. Secular trends in outcomes for Fanconi anemia patients who receive transplants: implications for future studies. *Biol Blood Marrow Transplant* **11**, 672-9 (2005).

92. Garaycoechea, J.I. et al. Genotoxic consequences of endogenous aldehydes on mouse haematopoietic stem cell function. *Nature* **489**, 571-5 (2012).

93. Garaycoechea J.I. et al. Alcohol and endogenous aldehydes damage chromosomes and mutate stem cells. *Nature* **553**, 171-177 (2018).

94. Langevin, F., Crossan, G.P., Rosado, I.V., Arends, M.J. & Patel, K.J. Fancd2 counteracts the toxic effects of naturally produced aldehydes in mice. *Nature* **475**, 53-8 (2011).

95. Oberbeck, N. et al. Maternal aldehyde elimination during pregnancy preserves the fetal genome. *Mol Cell* **55**, 807-17 (2014).

96. Pontel, L.B. et al. Endogenous Formaldehyde Is a Hematopoietic Stem Cell Genotoxin and Metabolic Carcinogen. *Mol Cell* **60**, 177-88 (2015).

97. Hira, A. et al. Variant ALDH2 is associated with accelerated progression of bone marrow failure in Japanese Fanconi anemia patients. *Blood* **122**(2013).

98. Dingler, F.A. et al. Two Aldehyde Clearance Systems Are Essential to Prevent Lethal Formaldehyde Accumulation in Mice and Humans. *Mol Cell* **80**, 996-1012 e9 (2020).

99. Zhang, Q.S. et al. Metformin improves defective hematopoiesis and delays tumor formation in Fanconi anemia mice. *Blood* **128**, 2774-2784 (2016).

100. Perez-Miller, S. et al. Alda-1 is an agonist and chemical chaperone for the common human aldehyde dehydrogenase 2 variant. *Nat Struct Mol Biol* **17**, 159-64 (2010).

101. Van Wassenhove, L.D., Mochly-Rosen, D. & Weinberg, K.I. Aldehyde dehydrogenase 2 in aplastic anemia, Fanconi anemia and hematopoietic stem cells. *Mol Genet Metab* **119**, 28-36 (2016).

102. Huang, M. et al. Human MutS and FANCM complexes function as redundant DNA damage sensors in the Fanconi Anemia pathway. *DNA Repair (Amst)* **10**, 1203-12 (2011).

103. Singh, T.R. et al. MHF1-MHF2, a histone-fold-containing protein complex, participates in the Fanconi anemia pathway via FANCM. *Mol Cell* **37**, 879-86 (2010).

104. Ciccia, A. et al. Identification of FAAP24, a Fanconi anemia core complex protein that interacts with FANCM. *Mol Cell* **25**, 331-43 (2007).

105. Yan, Z. et al. A histone-fold complex and FANCM form a conserved DNA-remodeling complex to maintain genome stability. *Mol Cell* **37**, 865-78 (2010).

106. Yang, Z. et al. A role for the base excision repair enzyme NEIL3 in replication-dependent repair of interstrand DNA cross-links derived from psoralen and abasic sites. *DNA Repair (Amst)* **52**, 1-11 (2017).

107. Bogliolo, M. et al. Biallelic truncating FANCM mutations cause early-onset cancer but not Fanconi anemia. *Genet Med* **20**, 458-463 (2018).

108. Catucci, I. et al. Individuals with FANCM biallelic mutations do not develop Fanconi anemia, but show risk for breast cancer, chemotherapy toxicity and may display chromosome fragility. *Genet Med* **20**, 452-457 (2018).

109. Singh, T.R. et al. Impaired FANCD2 monoubiquitination and hypersensitivity to camptothecin uniquely characterize Fanconi anemia complementation group M. *Blood* **114**, 174-80 (2009).

110. Tian, Y. et al. UHRF1 contributes to DNA damage repair as a lesion recognition factor and nuclease scaffold. *Cell Rep* **10**, 1957-66 (2015).

111. Motnenko, A. et al. Identification of UHRF2 as a novel DNA interstrand crosslink sensor protein. *PLoS Genet* **14**, e1007643 (2018).

112. Lopez-Martinez, D., Liang, C.C. & Cohn, M.A. Cellular response to DNA interstrand crosslinks: the Fanconi anemia pathway. *Cell Mol Life Sci* **73**, 3097-114 (2016).

113. Ceccaldi, R., Sarangi, P. & D'Andrea, A.D. The Fanconi anaemia pathway: new players and new functions. *Nat Rev Mol Cell Biol* **17**, 337-49 (2016).

114. Knipscheer, P. et al. The Fanconi Anemia Pathway Promotes Replication-Dependent DNA Interstrand Cross-Link Repair. *Science* **326**, 1689-1701 (2009).

115. Cole, A.R., Lewis, L.P. & Walden, H. The structure of the catalytic subunit FANCL of the Fanconi anemia core complex. *Nat Struct Mol Biol* **17**, 294-8 (2010).

116. Hodson, C., Purkiss, A., Miles, J.A. & Walden, H. Structure of the human FANCL RING-Ube2T complex reveals determinants of cognate E3-E2 selection. *Structure* **22**, 337-44 (2014).

117. Klein Douwel, D. et al. XPF-ERCC1 acts in Unhooking DNA interstrand crosslinks in cooperation with FANCD2 and FANCP/SLX4. *Mol Cell* **54**, 460-71 (2014).

118. van Twest, S. et al. Mechanism of Ubiquitination and Deubiquitination in the Fanconi Anemia Pathway. *Mol Cell* **65**, 247-259 (2017).

119. Medhurst, A.L. et al. Evidence for subcomplexes in the Fanconi anemia pathway. *Blood* **108**, 2072-80 (2006).

120. Shakeel, S. et al. Structure of the Fanconi anaemia monoubiquitin ligase complex. *Nature* **575**, 234-237 (2019).

121. Wang, S., Wang, R., Peralta, C., Yaseen, A. & Pavletich, N. Structure of the FA core ubiquitin ligase closing the ID clamp on DNA. *Nat Struct Mol Biol.* **28**(2021).

122. Alpi, A.F., Pace, P.E., Babu, M.M. & Patel, K.J. Mechanistic insight into site-restricted monoubiquitination of FANCD2 by Ube2t, FANCL, and FANCI. *Mol Cell* **32**, 767-77 (2008).

123. Rajendra, E. et al. The genetic and biochemical basis of FANCD2 monoubiquitination. *Mol Cell* **54**, 858-69 (2014).

124. Sato, K., Toda, K., Ishiai, M., Takata, M. & Kurumizaka, H. DNA robustly stimulates FANCD2 monoubiquitylation in the complex with FANCI. *Nucleic Acids Res* **40**, 4553-61 (2012).

125. Niraj, J. et al. The identification of FANCD2 DNA binding domains reveals nuclear localization sequences. *Nucleic Acids Res* **45**, 8341-8357 (2017).

126. Alcon, P. et al. FANCD2-FANCI is a clamp stabilized on DNA by monoubiquitination of FANCD2 during DNA repair. *Nat Struct Mol Biol* **27**, 240-248 (2020).

127. Smogorzewska, A. et al. Identification of the FANCI protein, a monoubiquitinated FANCD2 paralog required for DNA repair. *Cell* **129**, 289-301 (2007).

128. Ishiai, M. et al. FANCI phosphorylation functions as a molecular switch to turn on the Fanconi anemia pathway. *Nat Struct Mol Biol* **15**, 1138-46 (2008).

129. Lopez-Martinez, D. et al. Phosphorylation of FANCD2 Inhibits the FANCD2/FANCI Complex and Suppresses the Fanconi Anemia Pathway in the Absence of DNA Damage. *Cell Rep* **27**, 2990-3005 e5 (2019).

130. Tan, W. et al. Monoubiquitination by the human Fanconi anemia core complex clamps FANCI:FANCD2 on DNA in filamentous arrays. *Elife* **9**(2020).

131. Wang, R., Wang, S., Dhar, A., Peralta, C. & Pavletich, N.P. DNA clamp function of the monoubiquitinated Fanconi anaemia ID complex. *Nature* **580**, 278-282 (2020).

132. Oestergaard, V.H. et al. Deubiquitination of FANCD2 is required for DNA crosslink repair. *Mol Cell* **28**, 798-809 (2007).

133. Kim, J.M. et al. Inactivation of murine Usp1 results in genomic instability and a Fanconi anemia phenotype. *Dev Cell* **16**, 314-20 (2009).

134. Parmar, K. et al. Hematopoietic stem cell defects in mice with deficiency of Fancd2 or Usp1. *Stem Cells* **28**, 1186-95 (2010).

135. Zhang, J. et al. DNA interstrand cross-link repair requires replication-fork convergence. *Nat Struct Mol Biol* **22**, 242-7 (2015).

136. Fu, Y.V. et al. Selective bypass of a lagging strand roadblock by the eukaryotic replicative DNA helicase. *Cell* **146**, 931-41 (2011).

137. Long, D.T., Joukov, V., Budzowska, M. & Walter, J.C. BRCA1 promotes unloading of the CMG helicase from a stalled DNA replication fork. *Mol Cell* **56**, 174-85 (2014).

138. Wu, R.A. et al. TRAIP is a master regulator of DNA interstrand crosslink repair. *Nature* **567**, 267-272 (2019).

139. Fullbright, G., Rycenga, H.B., Gruber, J.D. & Long, D.T. p97 Promotes a Conserved Mechanism of Helicase Unloading during DNA Cross-Link Repair. *Mol Cell Biol* **36**, 2983-2994 (2016).

140. Budzowska, M., Graham, T.G., Sobeck, A., Waga, S. & Walter, J.C. Regulation of the Rev1-pol zeta complex during bypass of a DNA interstrand cross-link. *EMBO J* **34**, 1971-85 (2015).

141. Amunugama, R. et al. Replication Fork Reversal during DNA Interstrand Crosslink Repair Requires CMG Unloading. *Cell Rep* **23**, 3419-3428 (2018).

142. Ciccia, A., McDonald, N. & West, S.C. Structural and functional relationships of the XPF/MUS81 family of proteins. *Annu Rev Biochem* **77**, 259-87 (2008).

143. Tsodikov, O.V., Enzlin, J.H., Scharer, O.D. & Ellenberger, T. Crystal structure and DNA binding functions of ERCC1, a subunit of the DNA structure-specific endonuclease XPF-ERCC1. *Proc Natl Acad Sci U S A* **102**, 11236-41 (2005).

144. de Laat, W. Appeldoorn, E. Jaspers, N & Hoeijmakers, J.H.J. DNA Structural Elements Required for ERCC1-XPF Endonuclease Activity. *The Journal of Biological Chemistry* **273**, 7835–7842 (1998).

145. Bowles, M. et al. Fluorescence-based incision assay for human XPF-ERCC1 activity identifies important elements of DNA junction recognition. *Nucleic Acids Res* **40**, e101 (2012).

146. Niedernhofer, L.J. et al. The structure-specific endonuclease Ercc1-Xpf is required to resolve DNA interstrand cross-link-induced double-strand breaks. *Mol Cell Biol* **24**, 5776-87 (2004).

147. Bogliolo, M. et al. Mutations in ERCC4, Encoding the DNA-Repair Endonuclease XPF, Cause Fanconi Anemia. *The American Journal of Human Genetics* **92**, 800–806 (2013).

148. Fisher, L.A., Bessho, M. & Bessho, T. Processing of a psoralen DNA interstrand cross-link by XPF-ERCC1 complex in vitro. *J Biol Chem* **283**, 1275-81 (2008).

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

150. Kim, Y. et al. Mutations of the SLX4 gene in Fanconi anemia. *Nat Genet* **43**, 142-6 (2011).

151. Fekairi, S. et al. Human SLX4 is a Holliday junction resolvase subunit that binds multiple DNA repair/recombination endonucleases. *Cell* **138**, 78-89 (2009).

152. Zhang, J. & Walter, J.C. Mechanism and regulation of incisions during DNA interstrand cross-link repair. *DNA Repair (Amst)* **19**, 135-42 (2014).

153. Hoogenboom, W.S., Boonen, R. & Knipscheer, P. The role of SLX4 and its associated nucleases in DNA interstrand crosslink repair. *Nucleic Acids Res* **47**, 2377-2388 (2019).

154. Lachaud, C. et al. Ubiquitinated Fancd2 recruits Fan1 to stalled replication forks to prevent genome instability. *Science* **351**, 846-9 (2016).

155. Zhou, W. et al. FAN1 mutations cause karyomegalic interstitial nephritis, linking chronic kidney failure to defective DNA damage repair. *Nat Genet* **44**, 910-5 (2012).

156. Ishiai, M. et al. DNA cross-link repair protein SNM1A interacts with PIAS1 in nuclear focus formation. *Mol Cell Biol* **24**, 10733-41 (2004).

157. Abdullah, U.B. et al. RPA activates the XPF-ERCC1 endonuclease to initiate processing of DNA interstrand crosslinks. *EMBO J* **36**, 2047-2060 (2017).

158. Niedzwiedz, W. et al. The Fanconi anaemia gene FANCC promotes homologous recombination and error-prone DNA repair. *Mol Cell* **15**, 607-20 (2004).

159. Kim, N., Mudrak, S.V. & Jinks-Robertson, S. The dCMP transferase activity of yeast Rev1 is biologically relevant during the bypass of endogenously generated AP sites. *DNA Repair (Amst)* **10**, 1262-71 (2011).

160. Guo, C. et al. Mouse Rev1 protein interacts with multiple DNA polymerases involved in translesion DNA synthesis. *The EMBO Journal* **22**, 6621-6630 (2003).

161. Ohashi, E. et al. Interaction of hREV1 with three human Y-family DNA polymerases. *Genes Cells* **6**, 523-31 (2004).

162. Räschle, M. et al. Mechanism of replication-coupled DNA interstrand crosslink repair. *Cell* **134**, 969-80 (2008).

1

163. Kim, H., Yang, K., Dejsuphong, D. & D'Andrea, A.D. Regulation of Rev1 by the Fanconi anemia core complex. *Nat Struct Mol Biol* **19**, 164-70 (2012).

164. Roy, U., Mukherjee, S., Sharma, A., Frank, E.G. & Scharer, O.D. The structure and duplex context of DNA interstrand crosslinks affects the activity of DNA polymerase eta. *Nucleic Acids Res* **44**, 7281-91 (2016).

165. Howlett, N. et al. Biallelic Inactivation of BRCA2 in Fanconi Anemia. *Science* **297**, 606-609 (2002).

166. Long, D.T. Räschle, M. Joukov, V. & Walter, J.C. Mechanism of RAD51-dependent DNA interstrand cross-link repair. *Science* **333**, 84-7 (2011).

167. Inano, S. et al. RFWD3-Mediated Ubiquitination Promotes Timely Removal of Both RPA and RAD51 from DNA Damage Sites to Facilitate Homologous Recombination. *Mol Cell* **66**, 622-634 e8 (2017).

168. Niraj, J., Farkkila, A. & D'Andrea, A.D. The Fanconi Anemia Pathway in Cancer. *Annu Rev Cancer Biol* **3**, 457-478 (2019).

169. Wang, A.T. et al. A Dominant Mutation in Human RAD51 Reveals Its Function in DNA Interstrand Crosslink Repair Independent of Homologous Recombination. *Mol Cell* **59**, 478-90 (2015).

170. Sato, K. et al. HSF2BP negatively regulates homologous recombination in DNA interstrand crosslink repair. *Nucleic Acids Res* **48**, 2442-2456 (2020).

171. Mouw, K.W. & D'Andrea, A.D. Crosstalk between the nucleotide excision repair and Fanconi anemia/BRCA pathways. *DNA Repair (Amst)* **19**, 130-4 (2014).

172. Lossaint, G. et al. FANCD2 binds MCM proteins and controls replisome function upon activation of s phase checkpoint signaling. *Mol Cell* **51**, 678-90 (2013).

173. Schlacher, K., Wu, H. & Jasin, M. A distinct replication fork protection pathway connects Fanconi anemia tumor suppressors to RAD51-BRCA1/2. *Cancer Cell* **22**, 106-16 (2012).

174. Chen, Y.H. et al. ATR-mediated phosphorylation of FANCI regulates dormant origin firing in response to replication stress. *Mol Cell* **58**, 323-38 (2015).

175. Chaudhury, I., Sareen, A., Raghunandan, M. & Sobeck, A. FANCD2 regulates BLM complex functions independently of FANCI to promote replication fork recovery. *Nucleic Acids Res* **41**, 6444-59 (2013).

176. Chaudhury, I., Stroik, D.R. & Sobeck, A. FANCD2-controlled chromatin access of the Fanconi-associated nuclease FAN1 is crucial for the recovery of stalled replication forks. *Mol Cell Biol* **34**, 3939-54 (2014).

177. Rickman, K.A. et al. Distinct roles of BRCA2 in replication fork protection in response to hydroxyurea and DNA interstrand cross-links. *Genes Dev* (2020).

178. Howlett, N.G., Taniguchi, T., Durkin, S.G., D'Andrea, A.D. & Glover, T.W. The Fanconi anemia pathway is required for the DNA replication stress response and for the regulation of common fragile site stability. *Hum Mol Genet* **14**, 693-701 (2005).

179. Garcia-Rubio, M.L. et al. The Fanconi Anemia Pathway Protects Genome Integrity from R-loops. *PLoS Genet* **11**, e1005674 (2015).

180. Schwab, R.A. et al. The Fanconi Anemia Pathway Maintains Genome Stability by Coordinating Replication and Transcription. *Mol Cell* **60**, 351-61 (2015).

181. Zhang, T. et al. Fancd2 in vivo interaction network reveals a non-canonical role in mitochondrial function. *Sci Rep* **7**, 45626 (2017).

182. Semlow, D.R., Zhang, J., Budzowska, M., Drohat, A.C. & Walter, J.C. Replication-Dependent Unhooking of DNA Interstrand Cross-Links by the NEIL3 Glycosylase. *Cell* **167**, 498-511 e14 (2016).

183. Li, N. et al. Cooperation of the NEIL3 and Fanconi anemia/BRCA pathways in interstrand crosslink repair. *Nucleic Acids Res* **48**, 3014-3028 (2020).

184. Rolseth, V. et al. No cancer predisposition or increased spontaneous mutation frequencies in NEIL DNA glycosylases-deficient mice. *Sci Rep* **7**, 4384 (2017).

185. Huang, J. et al. The DNA translocase FANCM/MHF promotes replication traverse of DNA interstrand crosslinks. *Mol Cell* **52**, 434-46 (2013).

186. Huang, J. et al. Remodeling of Interstrand Crosslink Proximal Replisomes Is Dependent on ATR, FANCM, and FANCD2. *Cell Rep* **27**, 1794-1808 e5 (2019).

187. Mutreja, K. et al. ATR-Mediated Global Fork Slowing and Reversal Assist Fork Traverse and Prevent Chromosomal Breakage at DNA Interstrand Cross-Links. *Cell Rep* **24**, 2629-2642 e5 (2018).

188. Legerski, R.J. Repair of DNA interstrand cross-links during S phase of the mammalian cell cycle. *Environ Mol Mutagen* **51**, 540-51 (2010).

189. Hlavin, E.M., Smeaton, M.B. & Miller, P.S. Initiation of DNA interstrand cross-link repair in mammalian cells. *Environ Mol Mutagen* **51**, 604-24 (2010).

190. Muniandy, P.A., Liu, J., Majumdar, A., Liu, S.T. & Seidman, M.M. DNA interstrand crosslink repair in mammalian cells: step by step. *Crit Rev Biochem Mol Biol* **45**, 23-49 (2010).

191. Sarkar, S., Davies, A.A., Ulrich, H.D. & McHugh, P.J. DNA interstrand crosslink repair during G1 involves nucleotide excision repair and DNA polymeraseζ. *The EMBO Journal* **25**, 1285-1294 (2006).

192. Zheng, H. et al. Nucleotide excision repair- and polymerase eta-mediated error-prone removal of mitomycin C interstrand cross-links. *Mol Cell Biol* **23**, 754-61 (2003).

193. Kato, N. et al. Sensing and Processing of DNA Interstrand Crosslinks by the Mismatch Repair Pathway. *Cell Rep* **21**, 1375-1385 (2017).

194. Enoiu, M., Jiricny, J. & Scharer, O.D. Repair of cisplatin-induced DNA interstrand crosslinks by a replication-independent pathway involving transcription-coupled repair and translesion synthesis. *Nucleic Acids Res* **40**, 8953-64 (2012).

195. Wang, X. et al. Involvement of nucleotide excision repair in a recombination-independent and error-prone pathway of DNA interstrand cross-link repair. *Mol Cell Biol* **21**, 713-20 (2001).

196. Zamble, D.B., Mu, D., Reardon, J.T., Sancar, A. & Lippard, S.J. Repair of Cisplatin-DNA Adducts by the Mammalian Excision Nuclease. *Biochemistry* **35**, 10004-10013 (1996).

197. Wilhelm, F.X., Wilhelm, M.L., Erard, M. & Daune, M.P. Reconstitution of chromatin: assembly of the nucleosome. *Nucleic Acids Research* **5**(1978).

198. Kornberg, R.D. Chromatin Structure: A Repeating Unit of Histones and DNA. *Science* **184**, 868-871 (1974).

199. Zhou, K., Gaullier, G. & Luger, K. Nucleosome structure and dynamics are coming of age. *Nat Struct Mol Biol* **26**, 3-13 (2019).

200. Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693-705 (2007).

201. Talbert, P.B. & Henikoff, S. Environmental responses mediated by histone variants. *Trends Cell Biol* **24**, 642-50 (2014).

202. Tyagi, M., Imam, N., Verma, K. & Patel, A.K. Chromatin remodelers: We are the drivers!! *Nucleus* **7**, 388-404 (2016).

203. Dawson, M.A. & Kouzarides, T. Cancer epigenetics: from mechanism to therapy. *Cell* **150**, 12-27 (2012).

204. Cleaver, J.E. Nucleosome structure controls rates of excision repair in DNA of human cells. *Nature* **270**(1977).

205. Gontijo, A.M., Green, C.M. & Almouzni, G. Repairing DNA damage in chromatin. *Biochimie* **85**, 1133-47 (2003).

206. Sato, K. et al. Histone chaperone activity of Fanconi anemia proteins, FANCD2 and FANCI, is required for DNA crosslink repair. *EMBO J* **31**, 3524-36 (2012).

207. Timmers, C. et al. Positional Cloning of a Novel Fanconi Anemia Gene, FANCD2. *Molecular Cell* **7**, 241–248 (2001).

208. Hejna, J. et al. Tip60 is required for DNA interstrand cross-link repair in the Fanconi anemia pathway. *J Biol Chem* **283**, 9844-51 (2008).

209. Andreev, V. et al. Mammalian INO80 chromatin remodeler cooperates with FANCM to mediate DNA interstrand crosslink-induced checkpoint activation and repair. *DNA Repair (Amst)* **74**, 38-50 (2019).

210. Hoogenboom, W.S., Klein Douwel, D. & Knipscheer, P. Xenopus egg extract: A powerful tool to study genome maintenance mechanisms. *Dev Biol* **428**, 300-309 (2017).

211. Blow, J.J. & Laskey, R.A. Xenopus cell-free extracts and their contribution to the study of DNA replication and other complex biological processes. *Int J Dev Biol* **60**, 201-207 (2016).

212. Almouzni, G. Assembly of spaced chromatin involvement of ATP and DNA topoisomerase activity. *The EMBO Journal* **7**, 4355-4365 (1988).

213. Blow, J.J., Dilworth, S.M., Dingwall, C., Mills, A.D. & Laskey, R.A. Chromosome replication in cell-free systems from Xenopus eggs. *Phil. Trans. R. Soc. London* **317**, 483-494 (1987).

214. Lebofsky, R., Takahashi, T. & Walter, J.C. DNA Replication in Nucleus-Free Xenopus Egg Extracts. *Methods in Molecular Biology* **521**(2009).

215. Seeber, A., Hauer, M. & Gasser, S.M. Nucleosome remodelers in double-strand break repair. *Curr Opin Genet Dev* **23**, 174-84 (2013).

216. van Attikum, H. & Gasser, S.M. Crosstalk between histone modifications during the DNA damage response. *Trends Cell Biol* **19**, 207-17 (2009).

217. Stadler, J. & Richly, H. Regulation of DNA Repair Mechanisms: How the Chromatin Environment Regulates the DNA Damage Response. *Int J Mol Sci* **18**(2017).
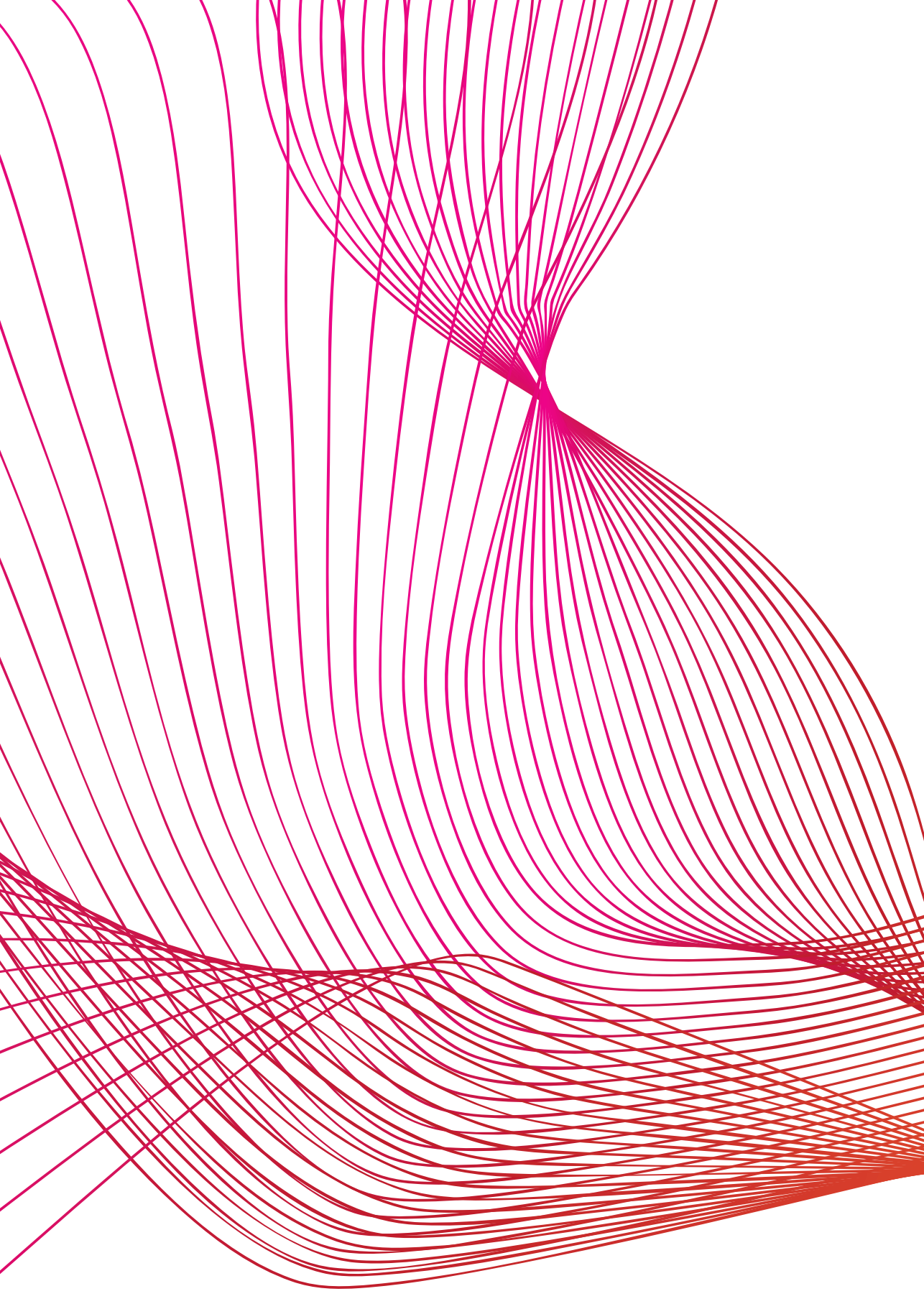
CHAPTER 2

Alcohol derived crosslinks are repaired by two distinct mechanisms

Michael R. Hodskinson[1,8], Alice Bolner[2,8], Koichi Sato[2], Ashley N. Kamimae-Lanning[1], Koos Rooijers[2], Merlijn Witte[2], Mohan Mahesh[1,5], Jan Silhan[1,6], Maya Petek[1,7], David M. Williams[3], Jop Kind[2], Jason W. Chin[1], Ketan J. Patel[1,4] & Puck Knipscheer[2]

[1]*MRC Laboratory of Molecular Biology, Cambridge, UK*

[2]*Oncode Institute, Hubrecht Institute-KNAW and University Medical Center Utrecht, Utrecht, The Netherlands*

[3]*Department of Chemistry, The University of Sheffield, Sheffield, UK*

[4]*MRC Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford, UK*

[5]*Present address: Department of Chemistry, Imperial College London, London, UK*

[6]*Present address: Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, Prague, Czech Republic*

[7]*Present address: Department of Biochemistry, University of Cambridge, Cambridge, UK*

[8]*These authors contributed equally: Michael R. Hodskinson, Alice Bolner*
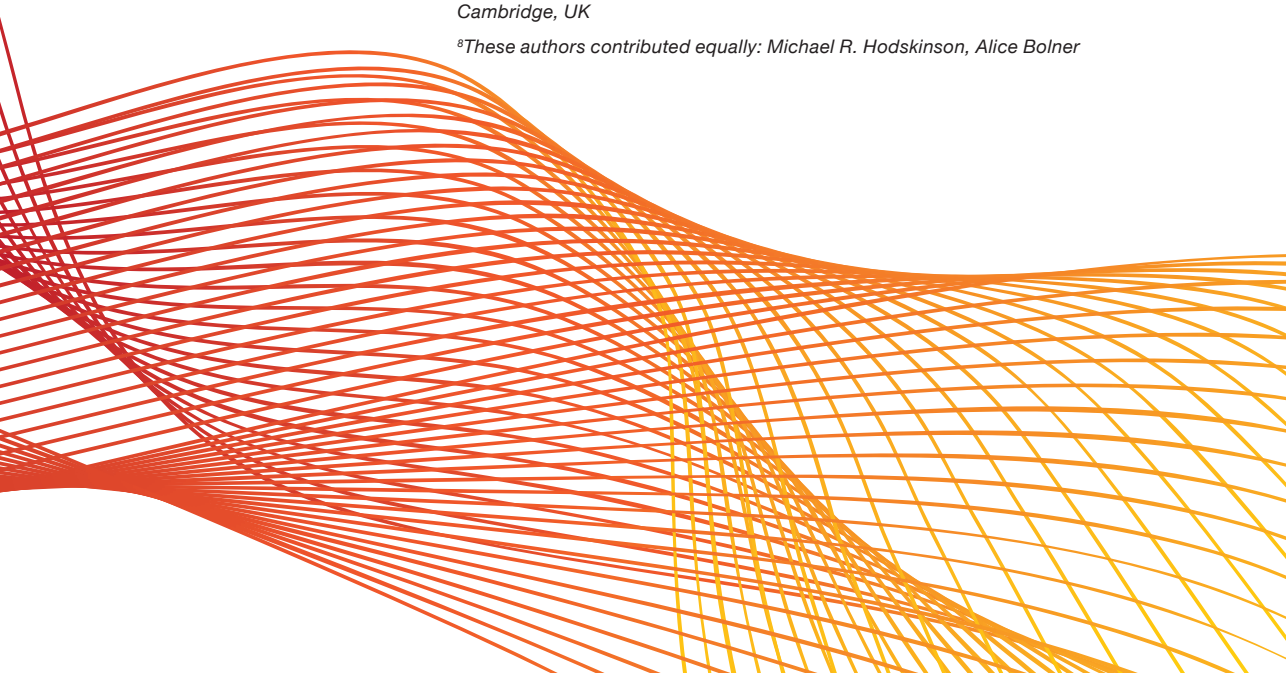
Acetaldehyde is a highly reactive, DNA-damaging metabolite that is produced upon alcohol consumption[1]. Impaired detoxification of acetaldehyde is common in the Asian population, and is associated with alcohol-related cancer[1,2]. Cells are protected against acetaldehyde-induced damage by DNA crosslink repair, which when impaired causes Fanconi anaemia (FA), a disease resulting in failure to produce blood cells and a predisposition to cancer[3,4]. The combined inactivation of acetaldehyde detoxification and the FA pathway induces mutation, accelerates malignancies and causes the rapid attrition of blood stem cells[5-7]. However, the nature of the DNA damage induced by acetaldehyde and how this is repaired remains a key question. Here we generate acetaldehyde-induced DNA interstrand crosslinks and determine their repair mechanism in *Xenopus* egg extracts. We find that two replication-coupled pathways repair these lesions. The first is the FA pathway, which operates using excision— analogous to the mechanism used to repair the interstrand crosslinks caused by the chemotherapeutic agent cisplatin. However, the repair of acetaldehyde-induced crosslinks results in increased mutation frequency and an altered mutational spectrum compared with the repair of cisplatin-induced crosslinks. The second repair mechanism requires replication fork convergence, but does not involve DNA incisions—instead the acetaldehyde crosslink itself is broken. The Y-family DNA polymerase REV1 completes repair of the crosslink, culminating in a distinct mutational spectrum. These results define the repair pathways of DNA interstrand crosslinks caused by an endogenous and alcohol-derived metabolite, and identify an excision-independent mechanism.

To study the repair of alcohol-induced DNA damage, we generated an acetaldehyde-crosslinked DNA substrate. Acetaldehyde reacts with guanine to create a crosslink precursor, N^2-propanoguanine (PdG)[8] (Fig. 1a). In a 5′-CpG sequence, PdG reacts with the N^2-amine of guanine on the opposite strand to create an acetaldehyde interstrand crosslink (AA-ICL); this crosslink exists in equilibrium between three states[9]. We synthesized a site-specific native AA-ICL, denoted AA_{NAT}-ICL, within an oligonucleotide duplex (Extended Data Fig. 1a, b, d, Supplementary Fig. 1). A control reaction of PdG with deoxyinosine—which lacks an N^2-amine—did not crosslink, thus confirming the site-specificity of the AA_{NAT}-ICL (Extended Data Fig. 1c). AA_{NAT}-ICLs were stable at physiological pH and temperature, with less than 10% reversal after 72 h at 37 °C (Extended Data Fig. 1e). However, increased temperature (55 °C) or exposure to acidic conditions reversed AA_{NAT}-ICLs, which is consistent with Schiff-base hydrolysis and protonation of the deoxyguanine (dG) N^2-amine[10] (Extended Data Fig. 1e, f). DNA crosslink repair is conserved among vertebrates and has been comprehensively studied in *Xenopus* egg extracts[11]. To examine the repair of AA_{NAT}-ICLs using this system, the oligonucleotide was ligated into a plasmid (pICL-AA_{NAT}). We also generated plasmids containing a cisplatin interstrand crosslink (pICL-Pt) or PdG (pPdG), and unmodified control plasmids (pCon) (Extended Data Fig. 1g, h). Crosslinked vectors were stable in non-replicating *Xenopus* egg extracts (Extended Data Fig. 1i).

Acetaldehyde crosslinks are repaired by two routes

Cisplatin ICLs are repaired by the replication-dependent FA pathway, which involves unhooking of the ICL by endonucleases, translesion synthesis (TLS) to bypass the adduct, and homologous recombination to resolve double-strand breaks[12-14] (Fig. 1b, Extended Data Fig. 2a). Replication of pICL-Pt in *Xenopus* egg extracts generates a temporal pattern of repair intermediates, starting with converged forks ('figure 8 'structure) and followed by low-mobility products that include homologous recombination intermediates (termed replication/repair intermediates (RRI)) and resolved nicked and supercoiled products[14] (Fig. 1b, c). Because the structures of cisplatin crosslinks and acetaldehyde crosslinks differ substantially (Extended Data Fig. 1j), we asked whether they were repaired by similar mechanisms. We replicated pICL-AA_{NAT} and pICL-Pt in *Xenopus* egg extracts, along with non-crosslinked controls, and separated the products by electrophoresis (Fig. 1c). Nicked and supercoiled products accumulated rapidly in experiments with pCon and pPdG, indicating that there was little or no impediment to replication. Experiments with pICL-AA_{NAT} resulted in RRI products similar to those seen with pICL-Pt; however, an earlier accumulation of nicked and supercoiled products compared with pICL- Pt suggested that some pICL- AA_{NAT} were repaired quickly.
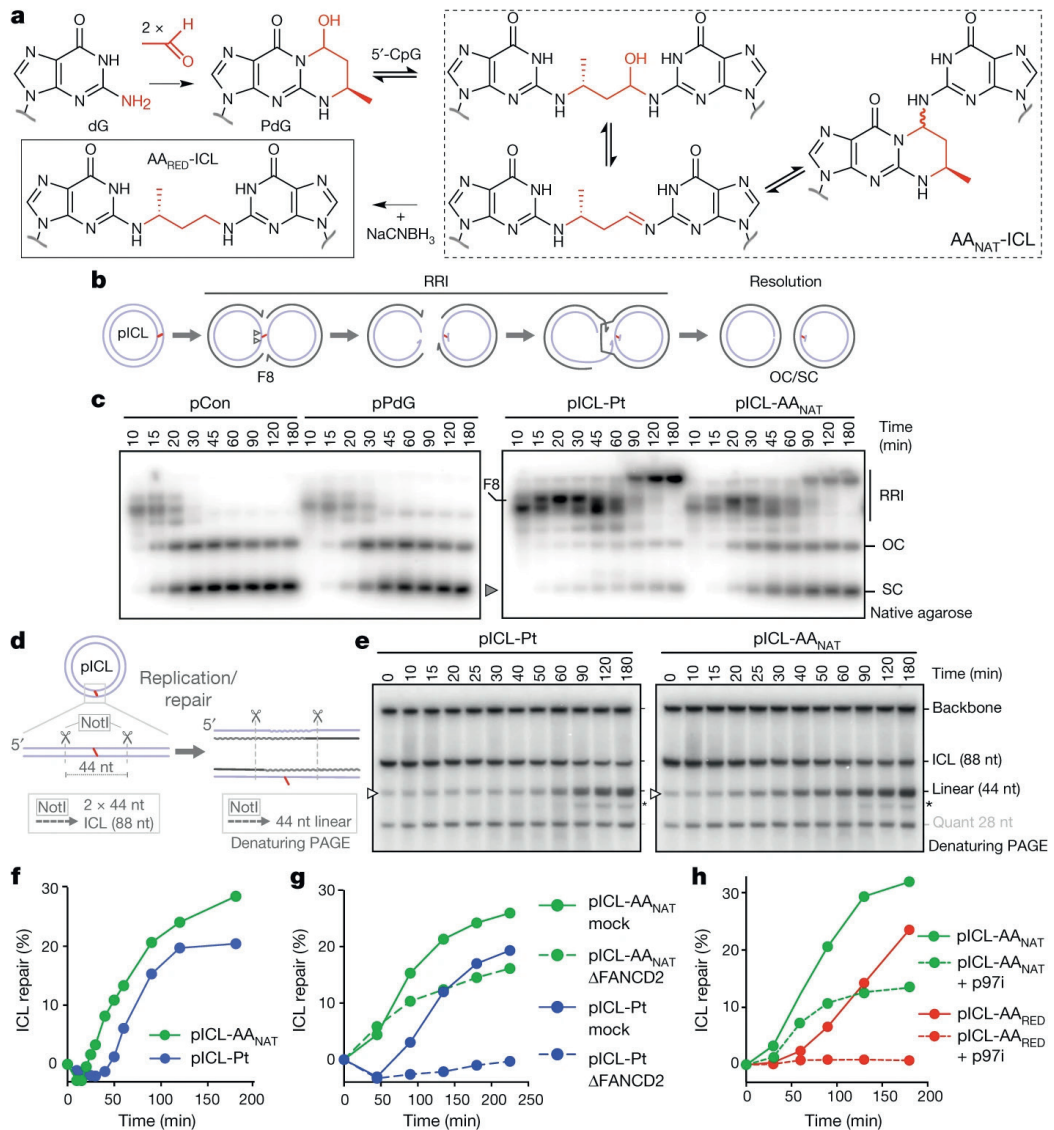
Fig. 1: AA-ICL repair by Fanconi-dependent and Fanconi-independent mechanisms.
a, Reaction scheme of the formation of AA_{NAT}-ICL. Two acetaldehyde molecules react with deoxyguanine (dG) to generate PdG; this reacts with a 5′-CpG guanine on the opposite strand. The AA_{NAT}-ICL exists in an equilibrium between three forms, shown in the dashed box. The crosslink can be chemically reduced with sodium cyanoborohydride to form ICL-AA_{RED} (solid box). **b**, Replication intermediates generated during ICL repair. **c**, Plasmids were replicated in *Xenopus* egg extracts, and the reaction products were resolved by electrophoresis and visualized by autoradiography. Figure 8 structures (F8), later RRIs, open circle (OC) and supercoiled (SC) products (grey arrow) are indicated. Six independent experiments were performed. **d**, Scheme for the NotI ICL repair assay. Wavy lines indicate the part of the strand that is synthesized during repair. **e**, Plasmids were replicated in *Xenopus* egg extracts, repair intermediates were isolated,

digested by NotI, and resolved by PAGE. Accumulation of the 44-nt product (white arrow) indicates replication/ repair. The asterisk marks a product that is probably generated from end- joining activity in some extracts. Ten independent experiments were performed. **f**, Quantification of ICL repair based on the gels in **e**, as described in the Supplementary Methods. Ten independent experiments were performed. **g**, Quantification of repair in mock-depleted or FANCD2-depleted (ΔFANCD2) extract, based on the gel in Extended Data Fig. 2e. Three independent experiments were performed. **h**, Quantification of repair in the presence or in the absence of p97i, based on the gel in Extended Data Fig. 3c. Three independent experiments were performed.

We developed an assay, termed the 'NotI assay', to determine whether the nicked and supercoiled products that we observed were indeed AA_{NAT}-ICL repair products. Repair intermediates were digested using NotI, labelled at the 3′ end and resolved by denaturing PAGE (Fig. 1d). Before DNA replication ($t = 0$), this resulted in fragments of 88 nucleotides (nt) (2×44 nt, crosslinked), 44 nt (low-level background of non-crosslinked plasmids) and the unresolved vector backbone. For pICL- AA_{NAT} and pICL-Pt, the proportion of the 44-nt fragment increased over time, confirming ICL repair (Fig. 1e). We quantified the repair products and found that they accumulated in greater quantities for pICL- AA_{NAT} compared with pICL-Pt (around 30% compared with around 20%, respectively, at 180 min) and, notably, pICL- AA_{NAT} displayed a faster rate of repair (11% compared with 1% at 50 min) (Fig. 1f, Extended Data Fig. 1k, l). These data suggest that a proportion of pICL- AA_{NAT} was processed in a similar manner to pICL-Pt, but that pICL- AA_{NAT} was additionally repaired by a second, faster mechanism.

Because cisplatin ICLs are repaired by the FA pathway (Extended Data Fig. 2a), we examined whether the same pathway also repaired AA_{NAT}- ICLs. In *Xenopus* egg extracts, both pICL-Pt and pICL-AA_{NAT} stimulated monoubiquitination of FANCD2— the activation step of the FA pathway (Extended Data Fig. 2b). FANCD2 depletion causes a defect in the unhooking of cisplatin ICLs, which results in the persistence of nascent leading strands one nucleotide before the ICL[12] (–1 position, Extended Data Fig. 2a). To visualize this, we replicated plasmids in mock and FANCD2-depleted extracts, digested the intermediates, and separated the products on a sequencing gel (Extended Data Fig. 2c, d). For both pICL-Pt and pICL-AA_{NAT}, FANCD2 depletion resulted in persisting –1 products (Extended Data Fig. 2d) and fewer extension products, indicating that the repair of both pICL-Pt and pICL-AA_{NAT} involves the FA pathway. We then questioned whether the second route of AA_{NAT}- ICL repair required the FA pathway. In NotI assays, FANCD2-depleted extracts did not support pICL–Pt repair, but pICL-AA_{NAT} was still partially repaired (Fig. 1g, Extended Data Fig. 2e–h). FA-dependent pICL-Pt repair requires unloading of the CMG helicase by the p97 segregase[15]. Consistent with published results, a p97 segregase inhibitor (referred to as p97i) halted pICL-Pt repair (Extended Data Fig. 2i–m). By contrast, p97i blocked only a proportion of AA_{NAT}-ICL repair, while the faster route of repair was unaffected by this treatment (Extended Data Fig. 2i–m). Together, these results indicate that pICL-AA_{NAT} repair proceeds through both an FA-dependent and an FA-independent mechanism.

Chemical reduction of AA_{NAT}-ICL by sodium cyanoborohydride results in a single, stable form of the acetaldehyde crosslink (AA_{RED}-ICL), as confirmed by high-performance liquid chromatography (HPLC) analysis (not shown), resistance to hydrolysis, and matrix-assisted laser desorption/ionization (MALDI) mass spectrometry (Fig. 1a, Extended Data Figs. 1f, 3a). Replication of a plasmid containing an AA_{RED}-ICL (pICL-AA_{RED}) in *Xenopus* egg extracts yielded RRIs that resembled those of the FA pathway, and there was very little accumulation of nicked and supercoiled products (Extended Data Fig. 3b). The NotI assay revealed that repair of pICL-AA_{RED} was slower than that of pICL-AA_{NAT}, and it was abolished by p97i (Fig. 1h, Extended Data Fig. 3c–e). AA_{RED}-ICL is therefore repaired exclusively by the FA pathway, indicating that the alternative repair route is restricted to the native AA-ICL.

Fork convergence is required for AA-ICL repair

To further characterize the second, faster mechanism of AA_{NAT}-ICL repair, we first asked whether it was replication-dependent. The addition of recombinant geminin, which inhibits DNA replication, blocked all repair of both pICL-Pt and pICL-AA_{NAT} (Fig. 2a, Extended Data Fig. 4a–c). Two replication forks must converge in order for pICL-Pt repair to initiate, promoting the ubiquitination and unloading of CMG helicase. Although CMG unloading is not required for the second route of AA_{NAT}-ICL repair, we questioned whether fork convergence was necessary. We generated an AA_{NAT}-ICL plasmid containing a *lacO* array that, when bound by Lac repressor (LacR), blocks the replication fork[16,17]. Because the AA_{NAT}-ICL is non-symmetrical we generated two versions of this plasmid, with the ICL in either orientation with respect to the rightward fork (pICL-AA-*lacO* and pICL-AAreverse-*lacO*, respectively, Fig. 2b). We replicated these plasmids and separated the digested repair inter mediates on a sequencing gel (Fig. 2b, c). In the absence of LacR, both leftward and rightward leading strands arrived at the –20 position, approached the –1 position, and were extended past the lesion over time. In the presence of LacR, arrival of the leftward fork was inhibited (as indicated by the absence of –1 and –20 products), showing that the LacR block was functional. The rightward fork persisted at the –20 position, which suggests a failure in repair progression. Moreover, the formation of extension products was impeded upon incubation with LacR, and this was similar for both orientations of AA_{NAT}-ICL (Fig. 2c) and pICL-Pt-*lacO* (Extended Data Fig. 4d). This indicates that the unhooking of AA_{NAT}-ICL requires replication fork convergence. Consistent with this, in the presence of LacR, repair of pICL-AA-*lacO* and pICL-AAreverse-*lacO* was greatly reduced (Extended Data Fig. 4e–h). These data show that the collision of one replication fork with an AA_{NAT}-ICL is insufficient for repair.
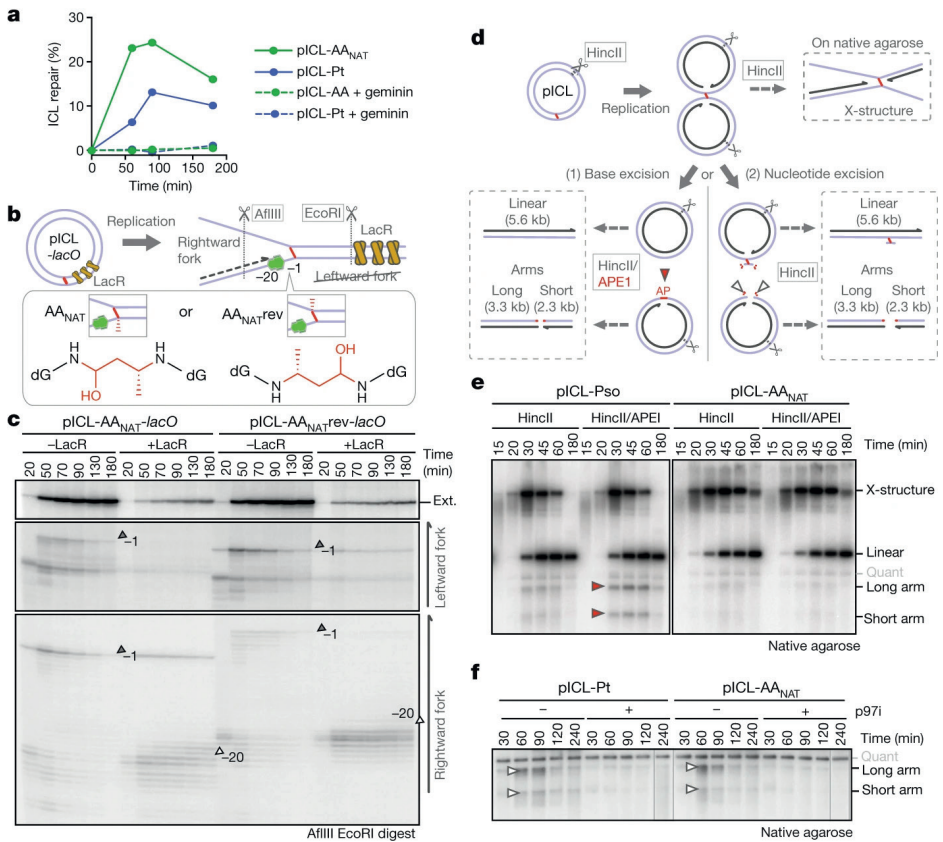
Fig. 2: The alternative AA-ICL repair route requires DNA replication and fork convergence but not DNA excision.

a, Quantification of repair in the presence or in the absence of geminin, based on the gel in Extended Data Fig. 4a. Three independent experiments were performed. **b**, Schematic representation of AA-ICL *lacO* plasmids, showing replication intermediate and digestion sites. **c**, Plasmids were replicated in *Xenopus* egg extracts in the presence or in the absence of LacR. Repair intermediates were digested and separated on a sequencing gel. Grey arrows, −1 product; white arrows, −20 product. Two independent experiments were performed. **d**, Scheme of the assays used to detect base excision (left) and nucleotide excision (right) pathways in **e** and **f**, respectively. DNA fragments formed after HincII/APE1 digestion (left box) or HincII digestion (top and right box). **e**, Plasmids were replicated in *Xenopus* egg extracts with p97i. Repair intermediates were digested and separated by electrophoresis. Red arrows indicate the arms from APE1 incisions. The quantification of these species is shown in Extended Data Fig. 5e. Four independent experiments were performed. **f**, Plasmids were replicated in *Xenopus* egg extracts with or without p97i. Repair intermediates were digested and separated by electrophoresis. White arrows indicate the arms from backbone incisions. The quantification of these species is shown in Extended Data Fig. 5h. Six independent experiments were performed.

Although it is plausible that FA-independent AA$_{NAT}$-ICL unhooking is a consequence of two forks colliding, we think that the mechanism is more likely to be enzymatic, as is the case for all known ICL repair pathways.

New repair route does not involve DNA excision

Whereas cisplatin ICLs are unhooked by nucleolytic incisions, psoralen and abasic site ICLs are preferentially cleaved at a glycosidic bond by NEIL3 glycosylase[12,15,18]. To investigate whether NEIL3 has a role in AA_{NAT}-ICL repair, we complemented NEIL3-depleted extract with recombinant wild-type or catalytically inactive NEIL3 (Extended Data Fig. 5a). Unlike those of the psoralen-ICL plasmid (pICL-Pso)[15], pICL-AA_{NAT} repair intermediates were unaffected by the lack of NEIL3 activity (Extended Data Fig. 5b). Furthermore, human HAP1 cells in which *NEIL3* was knocked out (Δ*NEIL3*) were not hypersensitive to acetaldehyde, and NEIL3 deficiency did not further sensitize a FANCL-deficient strain (Extended Data Fig. 5c, d). It is plausible that another glycosylase unhooks the AA_{NAT}-ICL, so we examined the accumulation of abasic sites—the product of glycosylase cleavage. Recombinant APE1 cleaves abasic sites resulting in the generation of arm fragments upon linearization of replication intermediates. Such arms were generated during pICL-Pso repair, but they did not form during AA_{NAT}-ICL repair[15] (route 1, Fig. 2d, e, Extended Data Fig. 5e–g). This indicates that no abasic site is formed and that AA_{NAT}-ICL repair does not cut the *N*-glycosyl bond.

Next, we tested whether nucleotide excision products are formed when an AA_{NAT}-ICL is processed in *Xenopus* egg extract. Backbone incisions should generate arm fragments when repair intermediates are linearized, as shown for repair of pICL-Pt (route 2, Fig. 2d, f). These arms were also formed during pICL-AA_{NAT} repair, which is consistent with FA pathway activity. Notably, when p97i was added—blocking the FA repair route—no incision products formed, indicating that the second ICL-AA_{NAT} repair pathway does not create backbone incisions (Fig. 2f, Extended Data Fig. 5h–j). To further confirm this, we examined the formation of adducts. In the FA pathway, the incisions of one strand result in an unhooked adduct on the other (Extended Data Fig. 5k). Therefore, late repair intermediates of pICL-Pt contain adducts on the top or the bottom strand, because either can be incised. Repair of pICL-AA_{NAT} also generated adducts, but they were largely restricted to the top stand—it is possible that adducts on the bottom strand either were present in amounts below the detection limit or were processed in the extract (Extended Data Fig. 5k, right). However, after the addition of p97i no adducts were detected, indicating that they depend on the FA pathway. In conclusion, the second, faster repair route for AA_{NAT}-ICLs does not involve a DNA excision step. This route must therefore operate by cutting within the crosslink itself.

Role of REV1 in acetaldehyde crosslink repair

Such a repair pathway would create an adduct on one or both strands, and should require TLS for nucleotide insertion opposite the adduct and extension beyond it. The TLS factor REV1 is critical for pICL-Pt repair, and so we tested whether it also operates in AA_{NAT}-ICL repair. Plasmids were replicated in mock and in REV1-depleted extracts, and digested intermediates were analyzed on a sequencing gel (Fig. 3a–c). For pICL-Pt, REV1 depletion led to the accumulation of insertion products (0 products) and reduced amounts of extension products[19] (Fig. 3c). By contrast, for pICL-AA_{NAT}, REV1 depletion led to the accumulation of –1 products, which is indicative of REV1-mediated insertion opposite the adduct (Fig. 3c, Extended Data Fig. 6a). Notably, this was also the case for pICL-AA_{RED} repair, indicating that there are different TLS mechanisms within the FA pathway. Furthermore, REV1 depletion caused more extensive leading-strand stalling at the rightward than at the leftward fork for AA_{NAT}-ICL (Fig. 3c). This suggests that unhooking by the second pathway may create an adduct on the bottom strand, which is bypassed by REV1. To test this, we inhibited the FA pathway using p97i and examined lesion bypass in the second pathway of AA_{NAT}-ICL repair. As expected, repair of pICL-Pt in the presence of p97i resulted in persistent stalling at the –20 to –40 position, due to defective CMG unloading[15] (Extended Data Fig. 6a). Similar stalled products were observed for AA_{NAT}-ICL repair, indicating that the FA pathway repair was inhibited (Extended Data Fig. 6a). However, we observed an accumulation of rightward –1 products, indicating that REV1 depletion prevents lesion bypass by the second pathway. By contrast, the leftward fork was extended without hindrance as no stalled products accumulated and a considerable amount of extension products were formed (Extended Data Fig. 6a). Depletion of REV7 (the regulatory subunit of Pol ζ, which is encoded by *FANCV*, also known as *MAD2L2*), had a very similar effect on AA-ICL repair to that of REV1 depletion (Extended Data Fig. 6d). It also caused an insertion defect in the FA pathway during AA_{RED}-ICL repair, as well as persistent stalling—especially of the rightward fork—in AA_{NAT}-ICL repair. As Pol ζ is primarily known for its extension activity[20], this defect could be due to co-depletion of the REV1–Pol ζ complex[19] (Extended Data Fig. 6b, c). In summary, the second pathway of AA_{NAT}-ICL repair generates an adduct on the bottom strand that requires REV1 and Pol ζ for bypass. The top strand, however, is readily extended without these TLS factors.
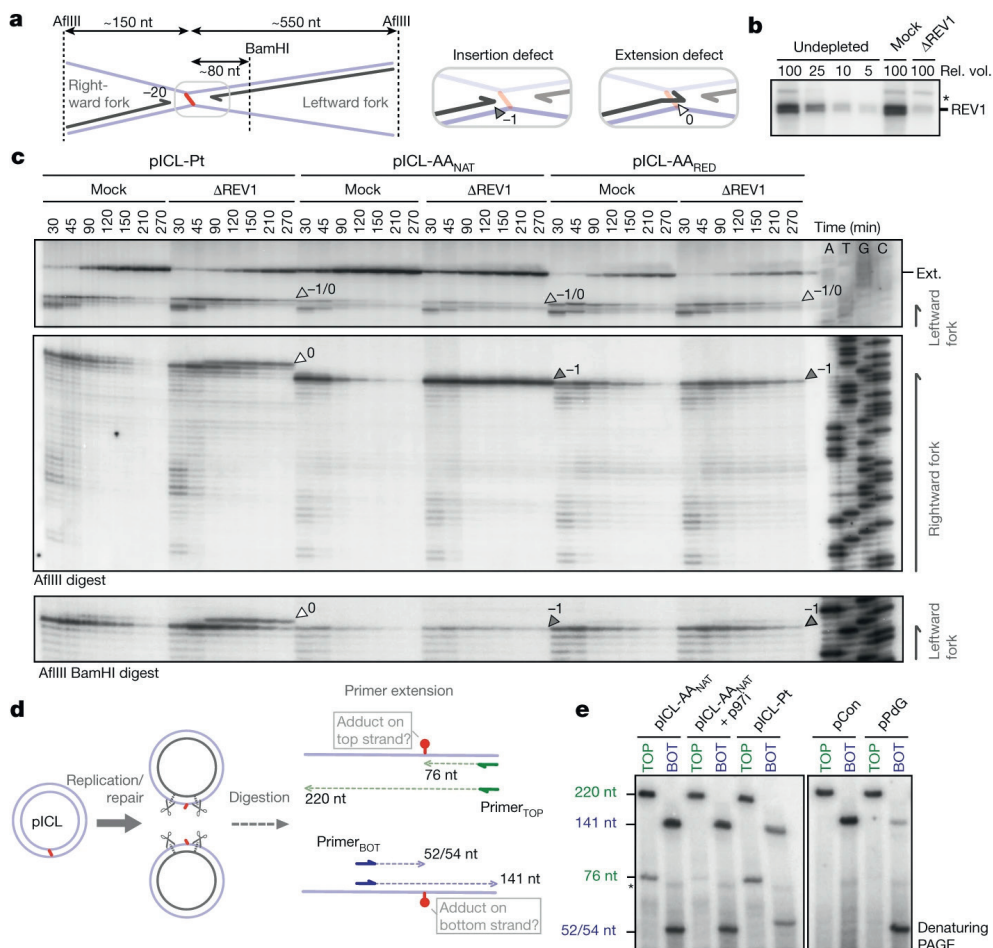
Fig. 3: AA-ICL repair is mediated by REV1.
a, Scheme for the formation of products detected on the sequencing gel in c. **b**, Western blot analysing REV1 in the mock-depleted and the REV1-depleted (ΔREV1) extract, compared with a titration of undepleted extract. The asterisk indicates a background band. Nine independent experiments were performed. **c**, Plasmids were replicated in mock-depleted or REV1-depleted extracts and repair intermediates were digested and separated on a sequencing gel alongside a sequencing ladder. White arrows, 0 products; grey arrows, −1 products; open arrows, 0/−1 products. Three independent experiments were performed. **d**, Scheme of the primer extension assay. Late repair products were digested and used as template for primer extension from labelled TOP or BOT (bottom) primers. **e**, Primer extension products were separated by denaturing PAGE. Four independent experiments were performed.

To further examine adduct formation, we isolated late repair products and subjected them to primer extension reactions at the ICL region using a high-fidelity polymerase (Fig. 3d). As expected, pICL-Pt repair products—which contain adducts on either parental strand—induced the formation of stalled extension products on

both strands (Fig. 3e). For AA$_{NAT}$-ICL, stalling also occurred on both strands but was more extensive on the bottom strand. Moreover, treatment with p97i—that is, blocking the FA pathway—almost entirely eliminated stalling on the top strand (Fig. 3e). These results suggest that the second repair pathway regenerates dG on the top strand, but creates a dG adduct on the bottom strand.

Acetaldehyde crosslink repair is mutagenic

Finally, we examined the fidelity of AA$_{NAT}$-ICL repair. To achieve this, we replicated plasmids in *Xenopus* egg extracts, recovered late repair products, and subjected them to high-throughput sequencing (Extended Data Table 1, Fig. 4a, b). Consistent with a previous report[19], we confirmed that pICL-Pt is repaired with a low error rate—less than 2% of the products carried a mutation at sites corresponding to crosslinked guanines (Fig. 4b, Extended Data Fig. 7a). The most common substitution was G>T transversion. Repaired pICL-AA$_{NAT}$ products show two differences: first, around 10% carry mutations at the crosslinked sites; and second, the mutational spectrum differs, with C>G, C>A, G>C and G>T transversions all observed (Fig. 4b, Extended Data Fig. 7a–c). The frequency of consecutive mutations (at least 2 nt) around the ICL was around 100-fold lower than that of single mutations, in agreement with a report suggesting that ICLs do not drive tandem mutations caused by acetaldehyde[21]. AA$_{NAT}$-ICL repair products generated in the presence of p97i—blocking the FA route—lost most of their mutations at position G8. This indicates that FA-dependent bypass of the top strand adduct is the predominant source of mutation at G8. Under these conditions, mutations are almost entirely restricted to C7 and the mutational spectrum is almost identical to that obtained from pPdG repair (Fig. 4b, Extended Data Fig. 7b). These data strongly suggest that the second route of AA$_{NAT}$-ICL repair reverses this crosslink to yield a monoadduct that is similar or identical to the original propanoguanine. Consistent with this model, we found that the insertion step of TLS past a PdG adduct is mediated by REV1 (Extended Data Fig. 6e).
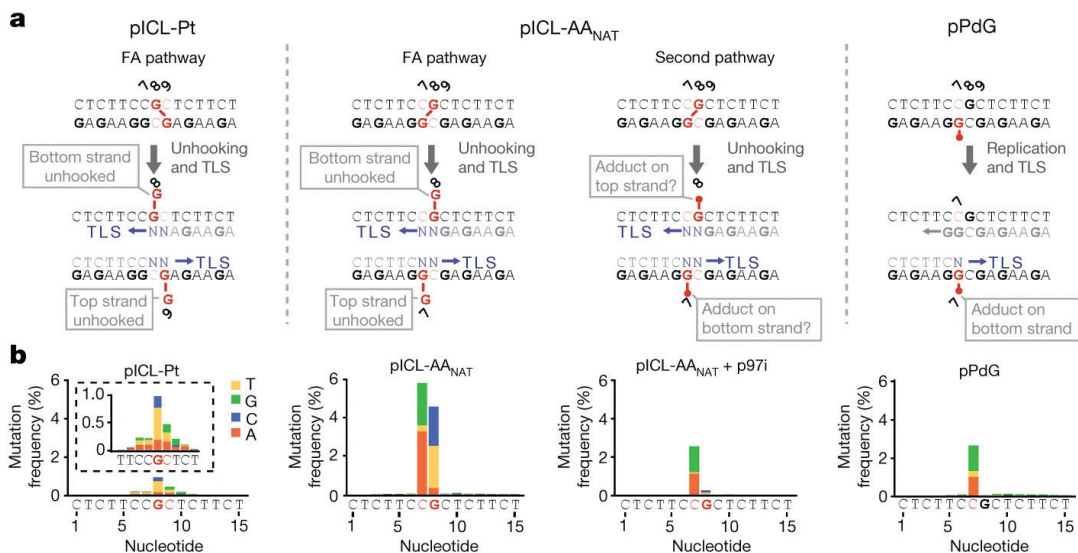
Fig. 4: AA-ICL repair causes point mutations.
a, Scheme showing the positions of DNA lesions and potential adducts (shown in red) generated during repair. Mutations are generated during translesion synthesis (blue Ns and arrow). **b**, Distribution and frequency of nucleotide misincorporation in a 15-bp region flanking the lesions. Strand specificity is lost because sample preparation involves PCR amplification; as such, only the top sequence is indicated below the graphs. Mutation frequency is corrected for mutations found in pCon. Two independent experiments were performed.

Conclusion

In summary, we have determined the repair mechanisms of an important class of endogenous DNA crosslinks—those caused by acetaldehyde. The central role of the FA repair pathway in removing such crosslinks agrees with the strong genetic evidence underpinning two- tier protection against this aldehyde[3,6,7]. However, we also uncovered a new pathway of DNA crosslink repair that operates without excision repair. This mechanism requires replication fork convergence and is unique in that it unhooks the ICL by cutting within the crosslink itself. Repair of AA$_{NAT}$-ICLs by both pathways is error-prone and requires the TLS polymerases REV1 and Pol ζ. However, this new repair modality has an obvious advantage in that it avoids the creation of DNA strand breaks or abasic sites, both of which can promote large-scale genome instability.

2

Author contributions
K.J.P. and P.K. initiated and supervised the study. M.R.H. co-ordinated key aspects of the project. K.S., A.N.K.-L. and K.R. contributed equally. M.R.H., M.M., J.S., M.P., D.M.W. and J.W.C. designed the strategy and synthesized the AA-ICL and PdG adducts. A.N.K.-L. performed cell toxicity experiments. A.B., M.W. and P.K. designed biochemical assays in *Xenopus* egg extracts. A.B., K.S. and M.W. conducted *Xenopus* extract assays. K.R. and J.K. performed bioinformatic analysis. K.J.P. and P.K. wrote the manuscript and M.R.H. designed and created the figures.

Extended Data Figures
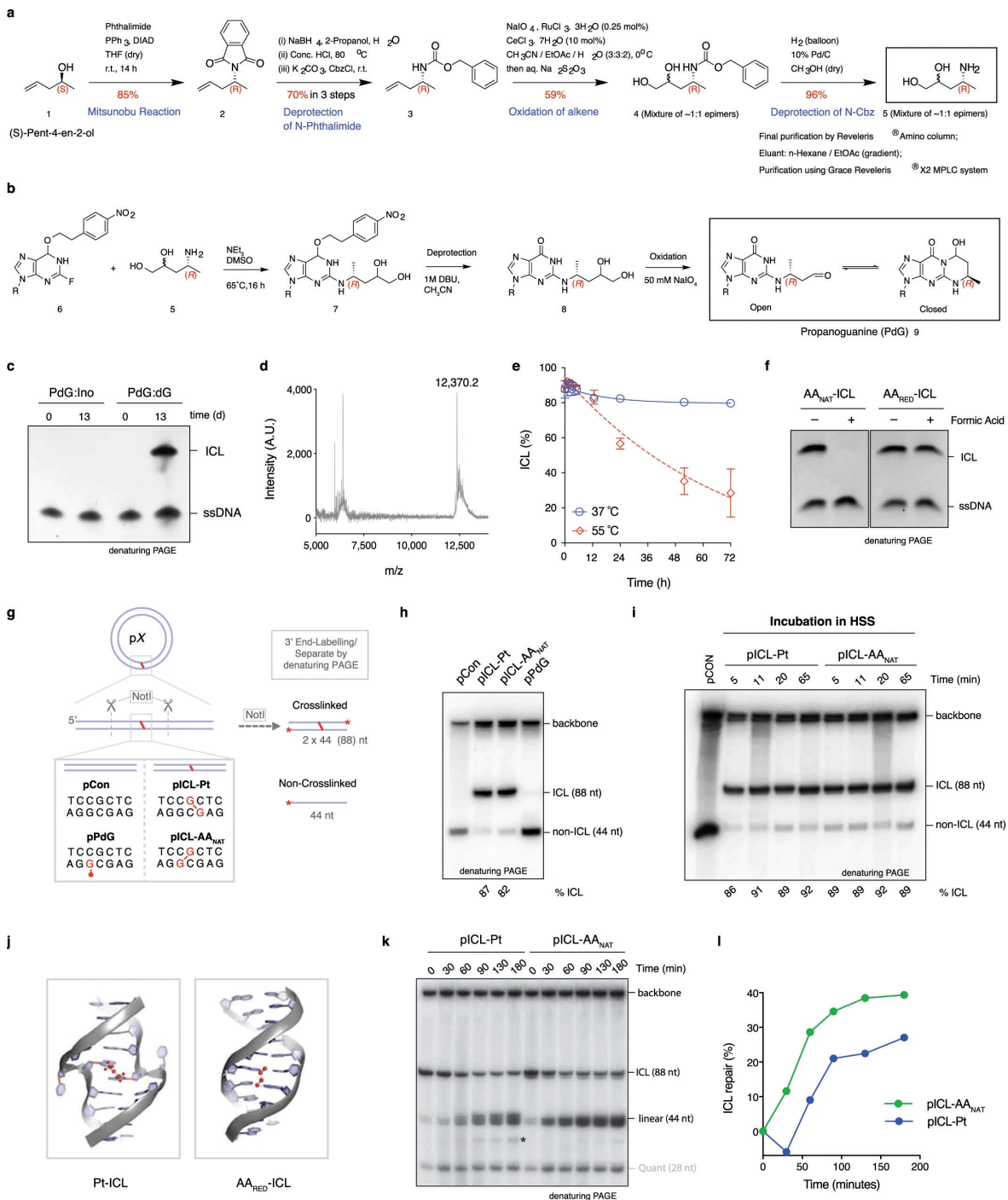
Extended Data Fig. 1: Acetaldehyde ICLs are stable and are repaired in *Xenopus* egg extracts.
a, Scheme for the synthesis of the precursor, 4-(R)-aminopentane-1,2-diol. **b**, Site-specific synthesis of a PdG adduct in a DNA oligonucleotide. **c**, Denaturing PAGE showing crosslink formation between dG and PdG, but not between dG and inosine (ino), which lacks an N2 amine. Two independent experiments were performed. **d**, Confirmation of AA-ICL formation by MALDI MS. The peak at m/z 12,370.2 represents the imine or pyrimidopurinone form. Two further peaks at m/z 5,979.41 and 6,409.74 equate to masses for the two parent oligonucleotides, consistent with the mass of the carbinolamine form after dissociation back to PdG and dG under the desorption/ionization conditions. Three independent experiments were performed. **e**, Stability of AA$_{NAT}$-ICL as a function of temperature and time, as determined by radiolabelling ([α-32P]dCTP) and resolution by denaturing PAGE. Error bars represent s.e.m. from three independent experiments. **f**, AA$_{NAT}$-ICL is susceptible to hydrolysis in aqueous acid, whereas AA$_{RED}$-ICL is stable. Pre-purification crosslink reactions were incubated with or without formic acid and products were resolved by denaturing PAGE. Three independent experiments were performed. **g**, Scheme depicting the type and position of the DNA lesions used in this study. Duplex DNA with or without the indicated lesions was annealed into a backbone vector to generate circular plasmids with or without damage. **h**, To determine the percentage of crosslinks, the ICL-containing plasmids were digested with NotI, labelled at the 3′-end by end filling with [α-32P]dCTP, and separated by denaturing PAGE. Crosslinked DNA (88 nt) shows slower mobility compared with non-crosslinked DNA (44 nt). The percentage of crosslinks was calculated by comparing the 88-nt product with the 44-nt products. Two independent experiments were performed. **i**, AA$_{NAT}$-ICL and Pt-ICL are stable in *Xenopus* egg extracts. Plasmids were incubated in a high-speed supernatant (HSS) extract. DNA was extracted and analysed as described in **h**. Three independent experiments were performed. **j**, Solution structures of a cisplatin ICL and a reduced form of an acetaldehyde ICL (PDB: 1DDP22 and 2HMD23, cartoon representation generated in PyMOL). **k,** The indicated plasmids were replicated in *Xenopus* egg extracts and repair intermediates were digested with NotI, labelled at the 3′-end, and resolved by denaturing PAGE. The increase in intensity of the 44-nt band over time indicates ongoing replication and repair. A higher mobility band, probably generated from end-joining activity in some extracts, is indicated by an asterisk. This gel is the independent experimental duplicate of that in Fig. 1e. **l**, Quantification of repair based on the intensity of the 44-nt product on the gel in **k**, as described in the Supplementary Methods. This graph is the independent experimental duplicate of that in Fig. 1f. Additional replicates of these experiments are presented in Fig. 2a, Extended Data Figs. 2k–m, 4b.
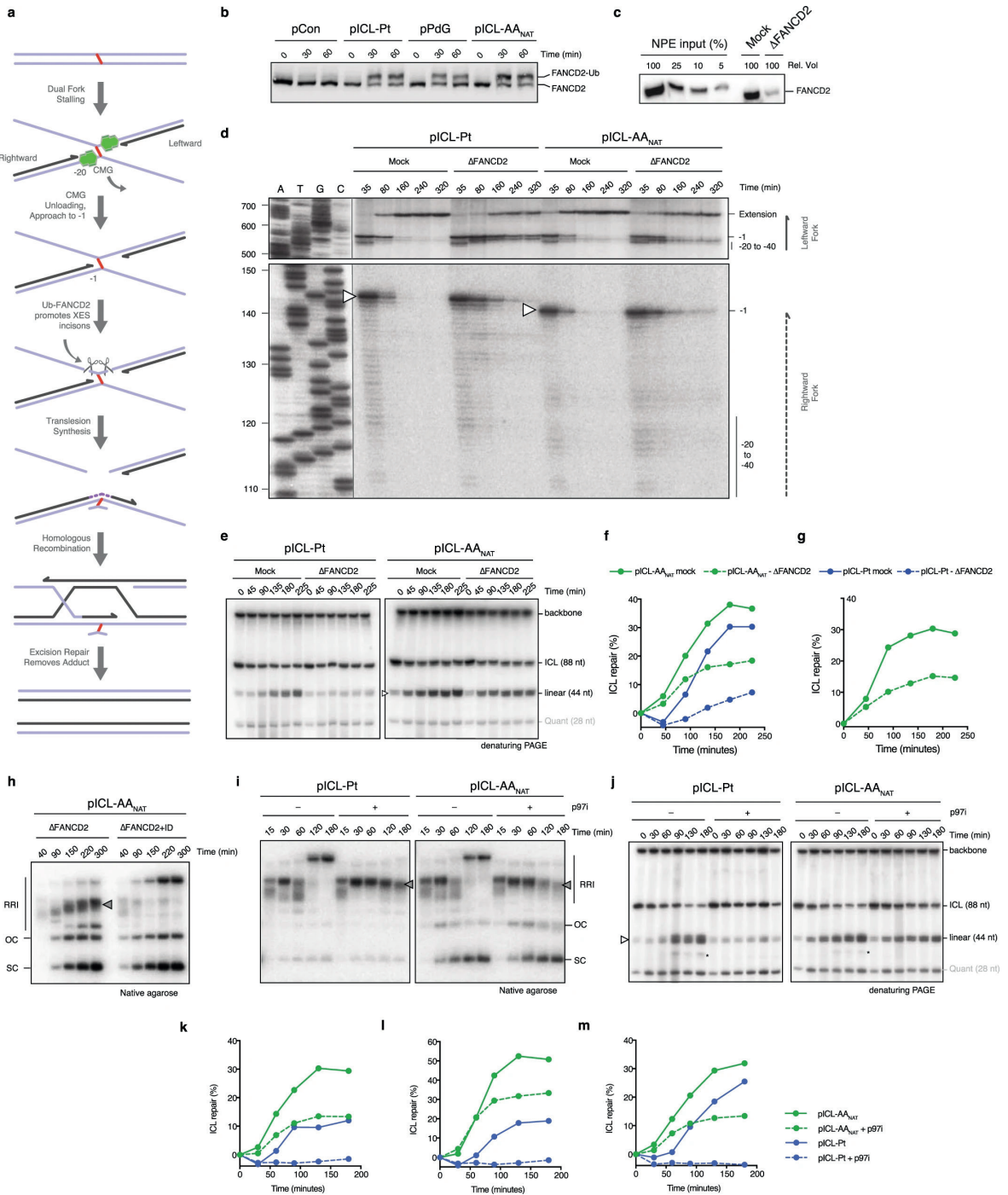
Extended Data Fig. 2: Acetaldehyde ICLs are repaired by Fanconi-dependent and Fanconi-independent mechanisms.

a, Model for ICL repair by the FA pathway. Upon convergence of two replication forks at the crosslink, the CMG helicase is unloaded from the DNA to enable the approach of one replication fork to the −1 position. Ubiquitylation of FANCD2 promotes the recruitment of the XPF-ERCC1-SLX4 (XES) complex to the ICL, which enables nucleolytic incisions that unhook the crosslink. This step could be preceded by fork reversal of one of the stalled replication forks[24]. Incisions generate a broken strand and a strand with an adduct; the latter is bypassed by TLS whereas the broken strand is repaired by homologous recombination. In mammalian cells, it has been shown that a single fork can pass over the ICL without unhooking[25]; this 'traverse' gives rise to a structure that resembles the one generated after fork convergence and CMG unloading and could follow the same steps subsequently. **b**, The indicated plasmids were replicated in *Xenopus* egg extracts and reaction samples were analysed by western blot with FANCD2 antibody. Two independent experiments were performed. **c**, Western blot of FANCD2, showing a titration of *Xenopus* egg extracts compared to mock and FANCD2-depleted extracts. Two independent experiments were performed. **d**, The indicated plasmids were replicated in mock or in FANCD2-depleted extracts in the presence of [α-32P]dCTP. Repair products were digested by AflIII, separated on a sequencing gel alongside a ladder derived from extension primer S, and visualized by autoradiography. The white arrow denotes the −1 product, which is 2 nt larger in pICL-Pt owing to the position of the ICL. Three independent experiments were performed. **e**, The indicated plasmids were replicated in mock or FANCD2-depleted extracts and repair intermediates were digested with NotI, labelled at the 3′-end, and resolved by denaturing PAGE. Quantification of repair based on the intensity of the 44-nt product is shown in Fig. 1g. **f**, The independent experimental duplicate of Fig. 1g. **g**, The independent experimental triplicate of Fig. 1g, but using only pICL-AA$_{NAT}$. **h**, Plasmid pICL-AA$_{NAT}$ was replicated in FANCD2-depleted extract, or FANCD2-depleted extract supplemented with a recombinant FANCI–FANCD2 complex (ID). Reaction samples were resolved by native agarose gel and visualized by autoradiography. RRIs, open circle (OC) and supercoiled (SC) products are indicated. The stalled repair product (grey arrow) is indicated. Two independent experiments were performed. **i**, The indicated plasmids were replicated in *Xenopus* egg extracts in the presence or in the absence of p97i and the intermediates were resolved by native agarose gel electrophoresis. The stalled repair products (grey arrow) are indicated. Seven independent experiments were performed. **j**, The indicated plasmids were replicated in *Xenopus* egg extracts in the presence or in the absence of p97i, and repair intermediates were digested with NotI, labelled at the 3′-end, and resolved by denaturing PAGE. The increase in intensity of the 44-nt band (white arrow) over time indicates ongoing replication and repair. A higher mobility band, probably generated from end-joining activity in some extracts, is indicated with an asterisk. **k**, Quantification of repair based on the intensity of the 44-nt product on the gel shown in j, as indicated in the Supplementary Methods. **l**, Quantified independent experimental duplicate of j. **m**, Quantified independent experimental triplicate of j.
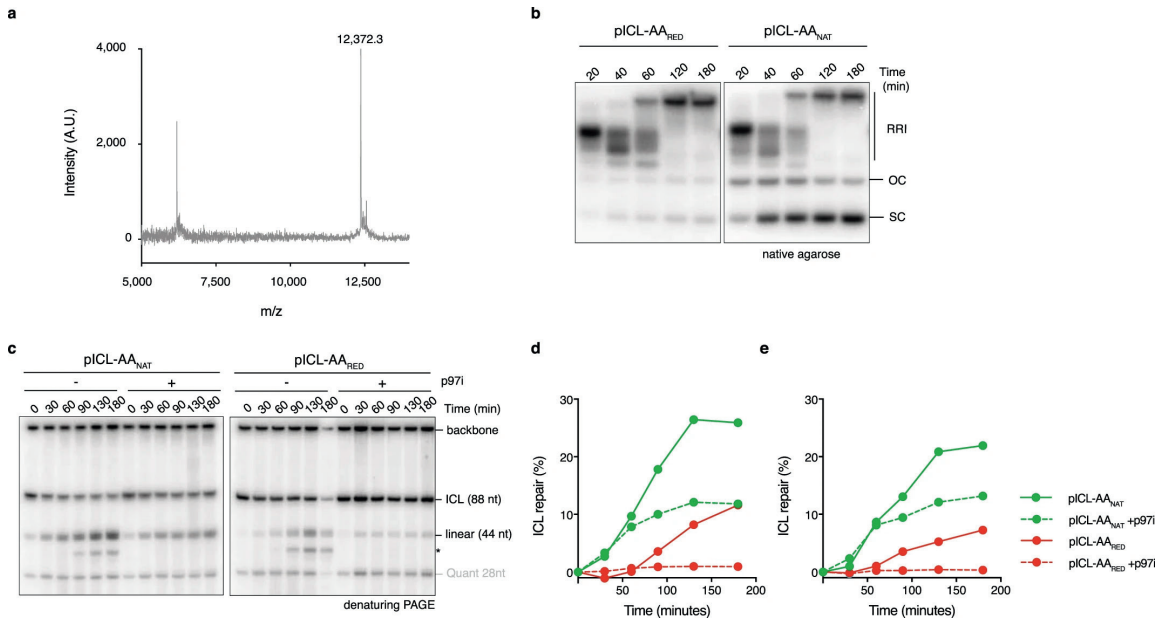
Extended Data Fig. 3: Reduced acetaldehyde ICLs are repaired by the Fanconi pathway.
a, MALDI MS data confirms the identity and the stability of the reduced ICL. Three independent experiments were performed. **b**, Plasmids were replicated in extract and products were resolved on a native agarose gel. Three independent experiments were performed. **c**, The indicated plasmids were replicated in *Xenopus* egg extracts in the presence or in the absence of p97i and the repair intermediates were analysed using the NotI digestion assay. The increase in intensity of the 44-nt band indicates ongoing replication and repair. A band with higher mobility, probably generated from end-joining activity in some extracts, is indicated with an asterisk. Quantification of repair based on this gel is shown in Fig. 1h. Three independent experiments were performed. **d**, The independent experimental duplicate of Fig. 1h. **e**, The independent experimental triplicate of Fig. 1h.

2
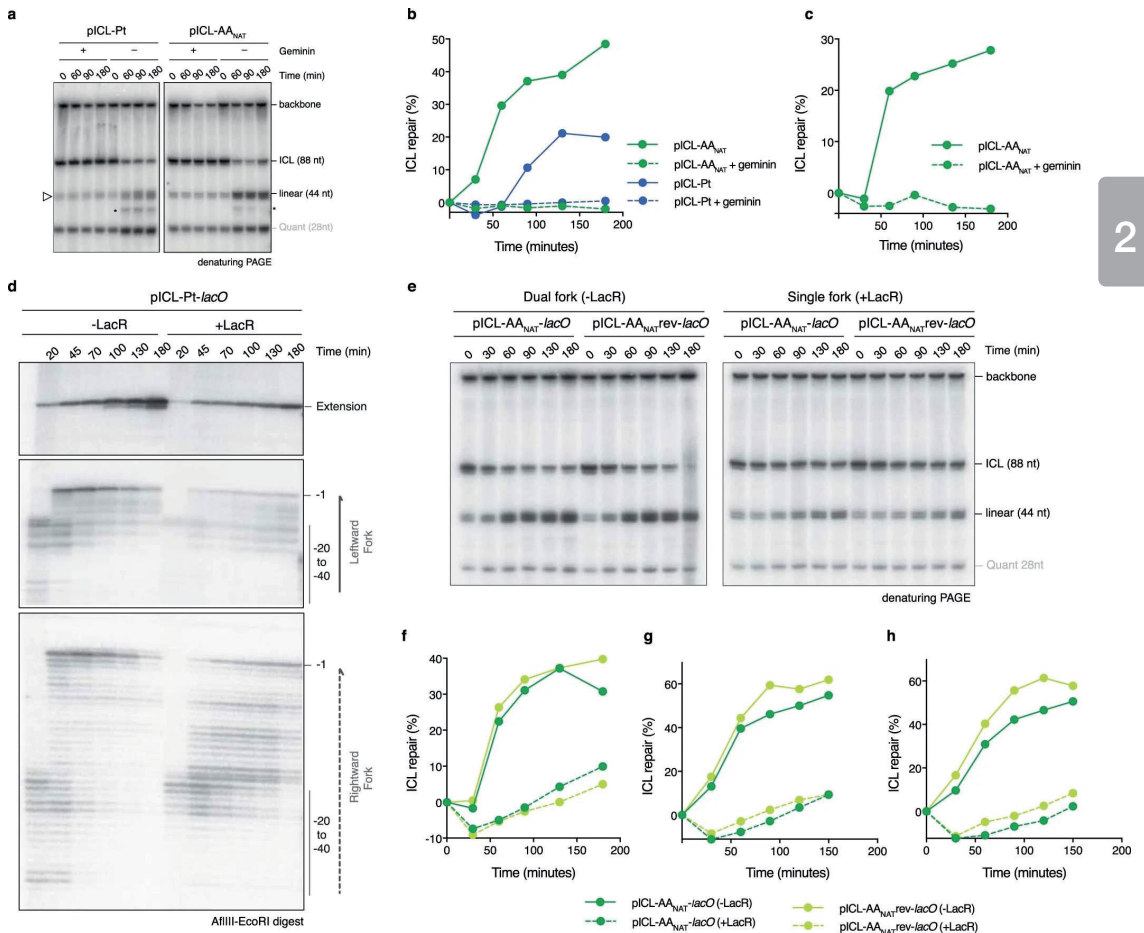
Extended Data Fig. 4: Acetaldehyde ICL repair requires DNA replication and replication fork convergence.
a, The indicated plasmids were replicated in *Xenopus* egg extracts in the presence or in the absence of geminin. Repair intermediates were digested with NotI, labelled at the 3′-end, and resolved by denaturing PAGE. Quantification of repair based on the intensity of the 44-nt product is shown in Fig. 2a. **b**, The independent experimental duplicate of Fig. 2a. **c**, The independent experimental triplicate of Fig. 2a but using only pICL-AA$_{NAT}$. **d**, pICL-Pt-*lacO* was replicated in *Xenopus* egg extracts containing [α-32P]dCTP in the presence or in the absence of LacR. The repair intermediates were digested with AflIII and EcoRI, separated on a sequencing gel and visualized by autoradiography. Two independent experiments were performed. **e**, The indicated plasmids were replicated in extracts in the presence or in the absence of LacR. Repair products were digested with NotI, labelled at the 3′-end, and resolved by denaturing PAGE. Three independent experiments were performed. **f**, Quantification of repair based on the intensity of the 44-nt product on the gel shown in e, as described in the Supplementary Methods. Three independent experiments were performed. **g**, The independent experimental duplicate of f. **h**, The independent experimental triplicate of f.
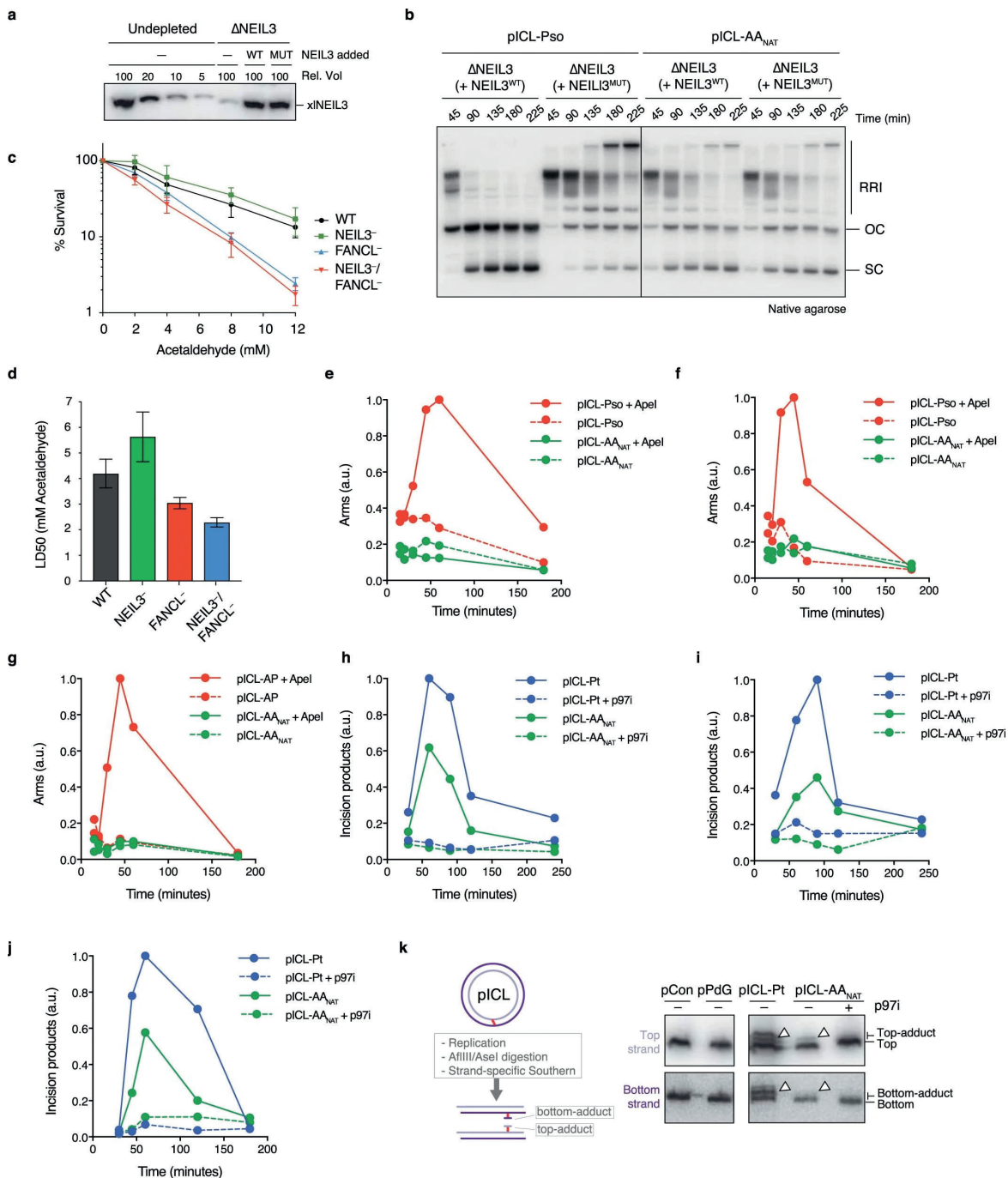
Extended Data Fig. 5: The alternative route of AA-ICL repair does not involve DNA excision.
a, Western blot showing a titration of *Xenopus* egg extracts compared to the NEIL3-depleted (ΔNEIL3) extract and NEIL3-depleted extract supplemented with recombinant wild-type NEIL3 (WT) or catalytically inactive NEIL3 (MUT). Three independent experiments were performed. **b**, The indicated plasmids were replicated in NEIL3-depleted extract containing [α-32P]dCTP, supplemented with wild-type (WT) or catalytically inactive (MUT) NEIL3. Replication intermediates were resolved by native agarose gel electrophoresis and visualized by autoradiography. Three independent experiments were performed. **c**, Clonogenic survival of wild-type, NEIL3-, FANCL- or NEIL3/FANCL-deficient human HAP1 cells after a 2-h exposure to acetaldehyde. Three independent experiments were performed. Data are mean ±s.e.m. **d**, The median lethal dose (LD50) of acetaldehyde for the survival of wild-type and deficient HAP1 cells was calculated by regression analysis of the curves presented in c. Data are mean ±s.e.m. Three independent experiments were performed. **e**, Quantification of the arm fragments resulting from APE1 treatment (AP sites) from the gel in Fig. 2e. **f**, Quantification of the APE1 arms as in e, from an independent duplicate experiment without the addition of p97i. **g**, Quantification of the APE1 arms as in e, from an independent triplicate experiment without the addition of p97i. As a positive control we used a plasmid containing an abasic-site-induced interstrand crosslink (pICL-AP) that is also repaired via the glycosylase NEIL3. **h**, Quantification of the HincII arm fragments from the gel in Fig. 2f. **i**, Quantification of the HincII arms as in h, from an independent duplicate experiment. **j**, Quantification of the HincII arms as in h, from an independent triplicate experiment. **k**, Schematic representation of the formation of DNA adducts by unhooking incisions during ICL repair (left). These adducts are not removed during ICL repair in *Xenopus* egg extracts[14] and can therefore be visualized. Plasmids were replicated in *Xenopus* egg extracts in the presence or in the absence of p97i. Late reaction samples were digested with AflIII and AseI and separated on a sequencing gel. Adducts on either the top or the bottom strand (white arrowheads) were detected by strand-specific Southern blotting (right). Three independent experiments were performed.

a

AflIII digest

AflIII - BamHI digest

b

REV7

c

REV7

REV1

d

Ext.

AflIII - BamHI digest

Rightward Fork

AflIII digest

Leftward Fork

AflIII - BamHI digest

e

Ext.

Rightward Fork

Leftward Fork

AFLIII - BamHI digest
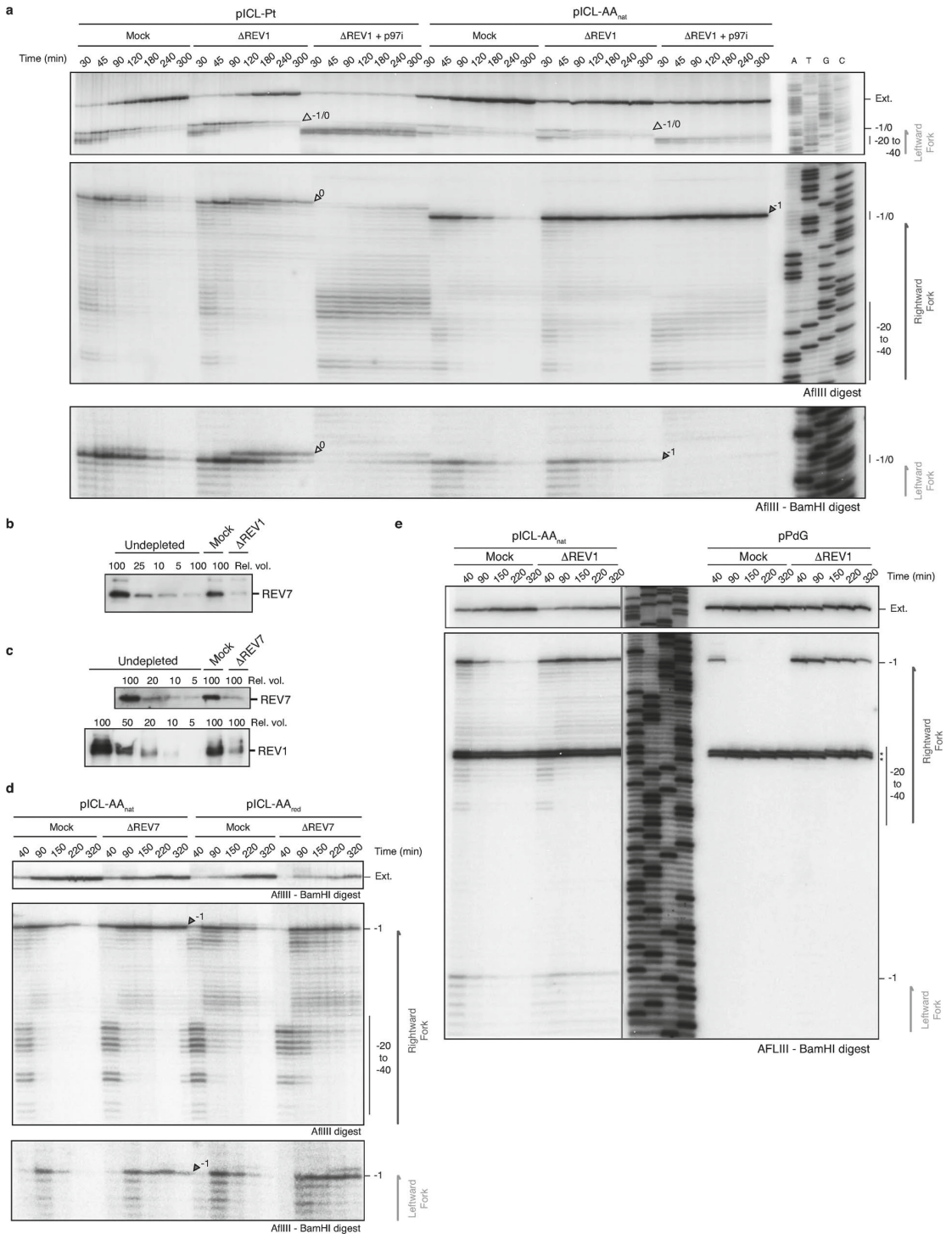
Extended Data Fig. 6: Both routes of AA-ICL repair are mediated by REV1 and REV7.
a, The indicated plasmids were replicated in mock or REV1-depleted extracts, in the presence or in the absence of p97i. Reaction intermediates were digested by either AflIII or AflIII and BamHI, separated on a sequencing gel alongside a ladder derived from extension of primer S, and visualized by autoradiography. White arrows denote 0 products, dark grey arrows indicate −1 products, and light grey arrows indicate 0/−1 products (not separated). Two independent experiments were performed. **b**, Western blot detection of REV7 in REV1- or mock-depleted *Xenopus* egg extracts compared to a titration of undepleted extract. Two independent experiments were performed. **c**, Western blot detection of REV7 and REV1 in REV7-depleted (ΔREV7) or mock-depleted *Xenopus* egg extracts compared to a titration of undepleted extract. Two independent experiments were performed. **d**, The indicated plasmids were replicated in mock- or REV7-depleted extracts and reaction intermediates were digested by either AflIII or AflIII and BamHI, separated on a sequencing gel and visualized by autoradiography. Grey arrows indicate −1 products. Two independent experiments were performed. **e**, Indicated plasmids were replicated in mock- or REV1-depleted extracts and reaction intermediates were digested by AflIII and BamHI, separated on a sequencing gel alongside a ladder derived from extension of primer S, and visualized by autoradiography. The asterisk indicates a 121-nt background fragment caused by a second BamHI restriction site in the leftward fork. Two independent experiments were performed.

2
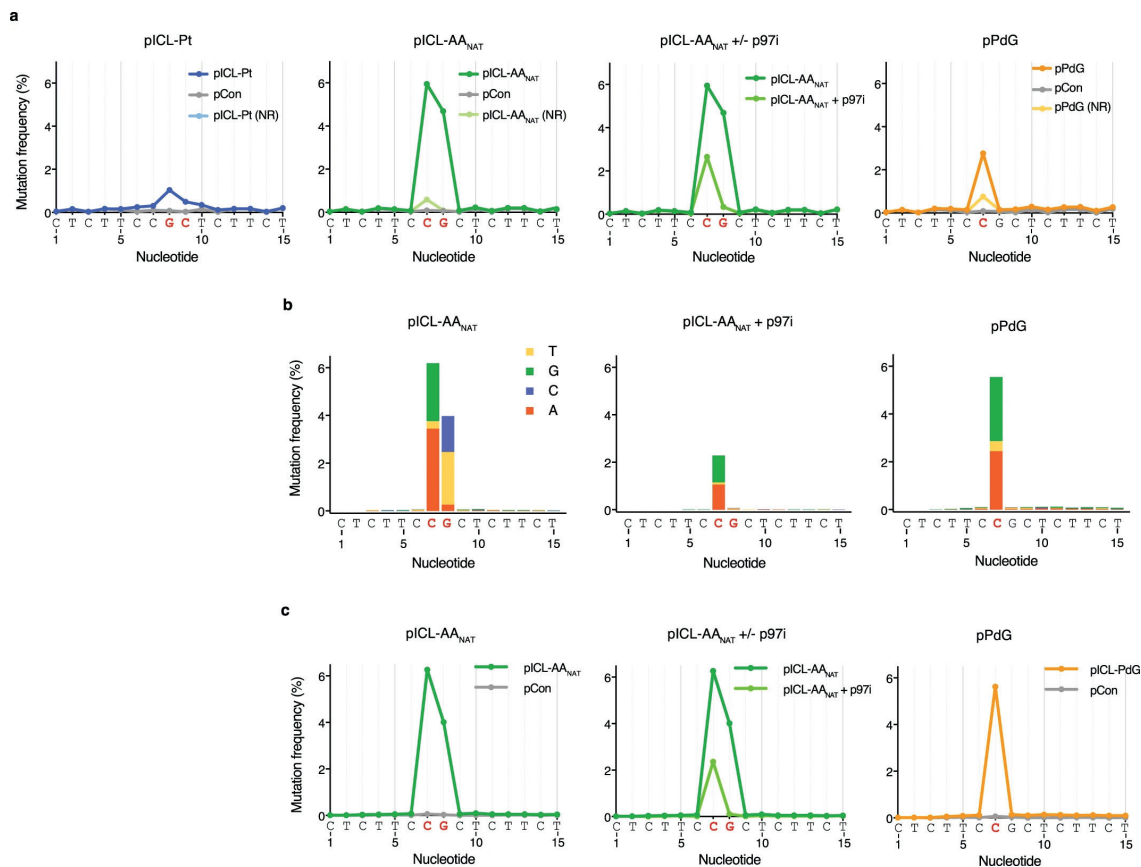
Extended Data Fig. 7: Mutagenic outcome of acetaldehyde crosslink repair.
a, Frequency of nucleotide misincorporation in a 15-bp region flanking the lesions present in the indicated plasmids. The mutation frequencies for the same plasmid that has not been replicated in *Xenopus* egg extracts (NR) and for the control vector (pCon) are plotted together. Strand specificity is lost because the sample preparation involves PCR amplification; as such, only the top sequence is indicated below the graphs. See Fig. 4. **b**, Distribution and frequency of nucleotide misincorporations in a 15-bp region flanking the lesions present in the indicated plasmids. Independent duplicate sequencing experiment of Fig. 4b. The heights of the bars represent the mutation frequency minus the baseline mutations found in pCon. **c**, Frequency of nucleotide misincorporations in a 15-bp region flanking the lesions present in the indicated plasmids (data from the same sequencing experiment as in b). The mutation frequency for pCon is also plotted.

Extended Data Table 1: Sequence and read numbers for high-throughput sequencing experiments.

a

Amplified amplicon	
Experiment 1	AGAACCAATGCATGCGGCCGCGAAGACAGCCCTCTTCCGCTCTT CTTTCGTGCGCGGCCGCGATCCGCTGCATTAATGAAT
Experiment 2	CTCGAGCGGAAGTGCAGAACCAATGCATGCGGCCGCGAAGACA GCCCTCTTCCGCTCTTCTTTCGTGCGCG GCCGCGATCCGCTGCA TTAATGAATCGGCCAACGC GCGGGGAGAGGCGGTTTGCGTATT

b

Sample	Total reads	Pair-matched reads	Reads with indels	Reads with substitutions	Perfect match reads
pICL-Pt	8886804	7928127	2649263	249138	5029726
pICL-AA	14852974	13541980	2396503	1458825	9686652
pPdG	10122605	9331173	69747	565567	8695859
pCon	10449164	9644715	6466	240998	9397251
pICL-AA + p97i	32164058	29539112	491796	1600097	27447219
pICL-Pt-NR	2571879	2385136	102534	58870	2223732
pICL-AA-NR	7316392	6729595	170216	200250	6359129
pPdG-NR	8857493	8137089	52886	270870	7813333

c

Sample	Total reads	Pair-matched reads	Reads with indels	Reads with substitutions	Perfect match reads
pICL-AA	13877258	4566626	388256	538491	3639879
pPdG	15334790	6649408	32855	585454	6031099
pCon	15403988	6714207	7772	171840	6534595
pICL-AA + p97i	15455175	5803423	75922	284148	5443353

a, Amplicon sequence for sequencing experiments 1 and 2. **b**, Total and specific read numbers for sequencing experiment 1 (see Fig. 4). **c**, Total and specific read numbers for sequencing experiment 2 (see Extended Data Fig. 7).

References

1. Brooks, P. J., Enoch, M. A., Goldman, D., Li, T. K. & Yokoyama, A. The alcohol flushing response: an unrecognized risk factor for esophageal cancer from alcohol consumption. *PLoS Med.* **6**, e1000050 (2009).

2. Lai, C. L. et al. Dominance of the inactive Asian variant over activity and protein contents of mitochondrial aldehyde dehydrogenase 2 in human liver. *Alcohol. Clin. Exp. Res.* **38**, 44–50 (2014).

3. Langevin, F., Crossan, G. P., Rosado, I. V., Arends, M. J. & Patel, K. J. Fancd2 counteracts the toxic effects of naturally produced aldehydes in mice. *Nature* **475**, 53–58 (2011).

4. Garaycoechea, J. I. & Patel, K. J. Why does the bone marrow fail in Fanconi anemia? *Blood* **123**, 26–34 (2014).

5   Hira, A. et al. Variant ALDH2 is associated with accelerated progression of bone marrow failure in Japanese Fanconi anemia patients. *Blood* **122**, 3206–3209 (2013).

6. Garaycoechea, J. I. et al. Genotoxic consequences of endogenous aldehydes on mouse haematopoietic stem cell function. *Nature* **489**, 571–575 (2012).

7. Garaycoechea, J. I. et al. Alcohol and endogenous aldehydes damage chromosomes and mutate stem cells. *Nature* **553**, 171–177 (2018).

8. Wang, M. et al. Identification of DNA adducts of acetaldehyde. *Chem. Res. Toxicol.* **13**, 1149–1157 (2000).

9. Cho, Y. J. et al. Stereospecific formation of interstrand carbinolamine DNA cross-links by crotonaldehyde- and acetaldehyde-derived α-CH3-γ-OH-1,N2-propano-2′-deoxyguanosine adducts in the 5′-CpG-3′ sequence. *Chem. Res. Toxicol.* **19**, 195–208 (2006).

10. Loudon, G. M. Organic Chemistry 874–878 (*Oxford Univ. Press*, 2002).

11. Duxin, J. P. & Walter, J. C. What is the DNA repair defect underlying Fanconi anemia? *Curr. Opin. Cell Biol.* **37**, 49–60 (2015).

12. Knipscheer, P. et al. The Fanconi anemia pathway promotes replication-dependent DNA interstrand cross-link repair. *Science* **326**, 1698–1701 (2009).

13. Long, D. T., Räschle, M., Joukov, V. & Walter, J. C. Mechanism of RAD51-dependent DNA interstrand cross-link repair. *Science* **333**, 84–87 (2011).

14. Räschle, M. et al. Mechanism of replication-coupled DNA interstrand crosslink repair. *Cell* **134**, 969–980 (2008).

15. Semlow, D. R., Zhang, J., Budzowska, M., Drohat, A. C. & Walter, J. C. Replication- dependent unhooking of DNA interstrand cross-links by the NEIL3 glycosylase. *Cell* **167**, 498–511 (2016).

16. Duxin, J. P., Dewar, J. M., Yardimci, H. & Walter, J. C. Repair of a DNA-protein crosslink by replication-coupled proteolysis. *Cell* **159**, 346–357 (2014).

17. Zhang, J. et al. DNA interstrand cross-link repair requires replication-fork convergence. *Nat. Struct. Mol. Biol.* **22**, 242–247 (2015).

18. Klein Douwel, D. et al. XPF-ERCC1 acts in unhooking DNA interstrand crosslinks in cooperation with FANCD2 and FANCP/SLX4. *Mol. Cell* **54**, 460–471 (2014).

19. Budzowska, M., Graham, T. G. W., Sobeck, A., Waga, S. & Walter, J. C. Regulation of the Rev1–Pol ζ complex during bypass of a DNA interstrand cross-link. *EMBO J.* **34**, 1971–1985 (2015).

20. Johnson, R. E., Washington, M. T., Haracska, L., Prakash, S. & Prakash, L. Eukaryotic polymerases ɩ and ζ act sequentially to bypass DNA lesions. *Nature* **406**, 1015–1019 (2000).

21. Matsuda, T., Kawanishi, M., Yagi, T., Matsui, S. & Takebe, H. Specific tandem GG to TT base substitutions induced by acetaldehyde are due to intra-strand crosslinks between adjacent guanine bases. *Nucleic Acids Res.* **26**, 1769–1774 (1998).

2

Supplemental note 1: Methods

Synthesis of site-specific crosslinks

Reagents

Chemicals were supplied by Sigma (unless otherwise stated), except the reagents for HPLC and (4R)-4-aminopentane-1,2-diol synthesis, supplied by Link Technologies Ltd and Fluorochem Ltd. Culture medium was supplied by Gibco (Thermo Scientific).

Synthesis of the Cross Linker Aminodiol, 5

We first approached the synthesis of the cross linker aminodiol **5**, following a protocol described by Cho *et al.*[9], who reported a 6-step synthesis of ^{13}C-labelled aminodiol **5** starting from (R)-2-amino-1-propanol and used $K^{13}CN$ in one of the key step. However, in our hands, this 6-step protocol appeared to be cumbersome for scale up due to few non-reproducible, low-yielding steps and also involved handling highly toxic KCN and OsO_4 reagents in couple of key steps. Hence, we designed an alternative route for the synthesis of aminodiol **5** starting from commercially available (S)-pent-4-en-2-ol **1** (SCHEME 1). Each and every step in this synthesis of aminodiol **5** were optimised to afford good to excellent yields that were reproducible during scale-up. Additionally, in this modified synthesis of **5**, highly toxic reagents like KCN and OsO_4 were not used.

SCHEME 1: Synthesis of the Cross Linker Aminodiol 5
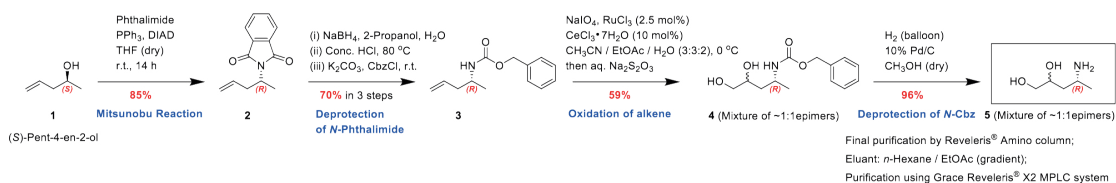
Abbreviations: DIAD = diisopropyl azodicarboxylate; Ph = phenyl; THF = tetrahydrofuran; r.t. = room temperature; h = hour(s); conc. = concentrated; °C = degree(s) Celsius; CBz = benzyloxycarbonyl; DMAP = 4-dimethylaminopyridine; aq.= aqueous

The (4R)-4-aminopentane-1,2-diol **5** was synthesised in 6 steps starting from commercially available (S)-pent-4-en-2-ol **1** with an overall yield of 34%. The synthesis commenced by stereospecific conversion of the enantiomerically pure alcohol **1** to the corresponding *N*-phthalimide derivative **2** by reaction with phthalimide *via* a Mitsunobu reaction[26,27] using PPh_3 and DIAD reagents in anhydrous THF as a solvent in 85% yield. The *N*-phthalimide **2** was deprotected[26] to

the corresponding (R)-amine under reductive conditions using $NaBH_4$ in 2-propanol and subsequent treatment with conc. HCl at 80 °C. The free amine was protected[26] with Cbz-Cl using K_2CO_3 as a base to afford the corresponding N-Cbz protected amine **3** in an overall yield of 70% in three steps. The alkene of **3** was subsequently oxidised[28] using $NaIO_4$, $RuCl_3$ (0.25 mol%) and $CeCl_3 \cdot 7H_2O$ (10 mol%) as a Lewis acid in a mixture of CH_3CN / EtOAc / H_2O (3:3:2) at 0 °C to obtain the corresponding diol **4** in 59% yield as a ~1:1 mixture of inseparable epimers. Deprotection of the N-Cbz group was achieved by hydrogenolysis of the diol **4** using H_2 gas and 10% Pd/C in CH_3OH to obtain the desired amino-1,2-diol **5** as a ~1:1 mixture of epimers in 96% yield after subsequent purification step. The product **5** was purified by Grace Reveleris® X2 system using n-Hexane/EtOH (gradient, using ELS detector) on Reveleris® amino cartridge and isolated as a clear colourless oil. This was stored in a freezer at -20 °C for several months without any decomposition. (R)-2-(Pent-4-en-2-yl)isoindoline-1,3-dione,[26] **2** and benzyl N-[(2R)-pent-4-en-2-yl]carbamate,[26] **3** are known compounds from the literature,[26] while benzyl [(2R)-4,5-dihydroxypentan-2-yl]carbamate, **4** and (4R)-4-aminopentane-1,2-diol, **5** are novel compounds reported here.

(R)-2-(Pent-4-en-2-yl)isoindoline-1,3-dione, 2

Phthalimide
PPh_3, DIAD
THF (dry)
r.t., 14 h

85%

Mitsunobu Reaction

1

(S)-Pent-4-en-2-ol

2

Chemical Formula: $C_{13}H_{13}NO$
Exact Mass: 215.09463
Molecular Weight: 215.2520
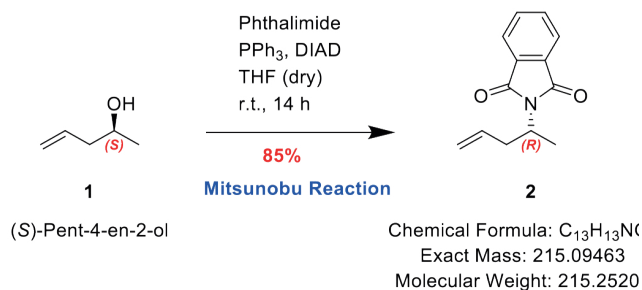
(R)-2-(Pent-4-en-2-yl)isoindoline-1,3-dione, **2** was synthesised following a slightly modified protocol for Mitsunobu reaction described in the literature[26,27]. An oven-dried 500 mL 3-necked round-bottomed flask with a magnetic stirring bar was dried *in vacuo* using a heat gun and purged with argon. This procedure was repeated 3 times and cooled to r.t. under an argon atmosphere. The flask was then charged with (S)-pent-4-en-2-ol **1** (4.307 g, 50.0 mmol, 1.0 eq.), phthalimide (8.092 g, 55.0 mmol, 1.1 eq.) and PPh_3 (14.430g, 55.0 mmol, 1.1 eq.) and dissolved in dry THF (200 mL) under an argon atmosphere. The contents were cooled to 0 °C and stirred vigorously. DIAD (12.638 g, 62.5 mmol, 1.25 eq.) was added dropwise to the mixture for a period of 10 min at 0 °C under an argon atmosphere. The contents were

warmed to r.t. for 30 min and then stirred for a further period of 2 h at r.t. (total time = 2.5 h). The contents were then evaporated to dryness and dissolved in EtOAc (400 mL) and sequentially washed with aqueous NaOH solution (0.5 M, 3 × 150 mL), aqueous HCl (1 M, 150 mL) and saturated brine solution (3 × 400 mL). The organic phase was then separated, dried over anhydrous Na_2SO_4, filtered and evaporated to dryness. The crude material was then subjected to flash chromatography on SiO_2 [gradient; eluant: EtOAc/n-hexane = 2:98 to 5:95] to obtain the desired (R)-2-(pent-4-en-2-yl)isoindoline-1,3-dione,[2] **2** as a colourless oil (9.190 g, 85%); $[\alpha]_D^{20}$ = -19.0° (c 1.0, CHCl$_3$); R$_f$ = 0.48 (SiO$_2$ plate, EtOAc/n-hexane = 1:9); IR (neat) v$_{max}$ 3078, 2977, 2938, 1772, 1701, 1642, 1467, 1393, 1376, 1359, 1332, 1248, 1173, 1136, 1081, 1058, 1015, 994, 916, 895, 872, 793, 717 cm^{-1}; ^{1}H NMR (400.13 MHz, CDCl$_3$) δ 1.49 (d, J = 6.9 Hz, 3H), 2.51 (ddd, J = 14.1, 7.8, 6.8 Hz, 1H), 2.80 (dddd, J = 14.1, 8.4, 7.8, 1.3 Hz, 1H), 4.43 (ddq, J = 8.4, 7.8, 6.9 Hz, 1H), 4.95 (app. d, J = 10.1 Hz, 1H), 5.05 (dd, J = 16.8, 1.3 Hz, 1H), 5.71 (dddd, J = 16.8, 10.1, 7.8, 6.8 Hz, 1H), 7.68 (dd, J = 5.4, 3.0 Hz, 1H), 7.80 (dd, J = 5.4, 3.0 Hz, 1H); ^{13}C NMR (100.61 MHz, CDCl$_3$) δ 18.5 (CH$_3$), 38.3 (CH$_2$), 47.1 (CH), 117.8 (CH$_2$), 123.2 (CH), 132.1 (C), 133.9 (CH), 134.9 (CH), 168.6 (C); MS (ESI+) m/z (rel intensity) 238 [(M+Na)$^+$, 100%], 216 [(M+H)$^+$, 62], 205 (19), 201 (4); HRMS (ESI+) m/z calc'd for C$_{13}$H$_{14}$NO$_2$ [M+H]$^+$: 216.1019, found 216.1016 (Δ = -1.32 ppm). The structure of the product **2** was further characterised by 2D [1H, 1H] COSY, 2D [1H, 13C] HSQC, 2D [1H, 13C] HMBC NMR spectroscopic experiments.

Benzyl N-[(2R)-pent-4-en-2-yl]carbamate, 3
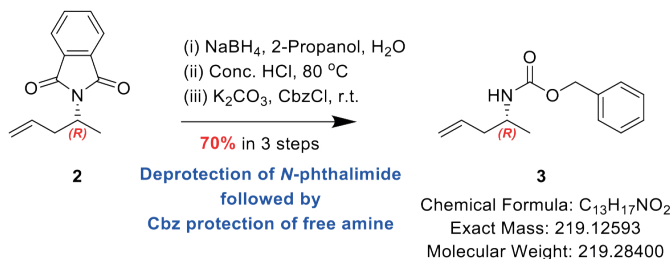
Benzyl N-[(2R)-pent-4-en-2-yl]carbamate, **3** was synthesised in 3 steps starting from the phthalimide **2** following a slightly modified protocol described in the literature by Gebauer *et al*[26]. (R)-2-(Pent-4-en-2-yl)isoindoline-1,3-dione, **2** (4.951 g, 23.0 mmol, 1.0 eq.) was dissolved in isopropanol (170 mL) and H$_2$O (30 mL). NaBH$_4$ (8.701 g, 230.0 mmol, 10.0 eq.) was added portion wise (8 × 1.088 g) to the solution for every 5 min (5 × 8 = 40 min total) and the contents were stirred at r.t. overnight for 16 h. A white precipitate appeared in the reaction mixture and the reaction mixture was quenched by addition of conc. HCl (15.0 mL), stirred at r.t. for 10 min, then heated to 80 °C for 2 h and then cooled to r.t. A thick white

precipitate was observed and the mixture was then neutralised with K_2CO_3 to pH 8.0, after which the mixture was cooled to 0 °C and benzyl chloroformate (5.885 g, 34.5 mmol, 1.5 eq.) was added dropwise for a period of 15 min. The mixture was stirred at 0 °C and DMAP (0.030 mg, 0.246 mmol, 0.0107 eq.) was added to it. The contents were then stirred at r.t. under an argon atmosphere and the progress of the reaction was monitored by TLC analysis (SiO_2 plate, EtOAc/n-hexane = 3:17). After 2.5 h, the reaction was judged to be complete and the reaction mixture was then quenched by addition of saturated brine solution (200 mL), extracted with Et_2O (3 × 200 mL). The aqueous phase was further extracted with EtOAc (50 mL). The combine organic extracts were washed with saturated brine solution (3 × 500 mL), dried over anhydrous Na_2SO_4, evaporated to dryness using a rotatory evaporator to obtain a colourless oil. This was purified by flash chromatography on SiO_2 [gradient; eluant: EtOAc/n-hexane = 1:19 → 1:9 → 3:17 → 1:4] to obtain the desired product benzyl N-[(2R)-pent-4-en-2-yl]carbamate,[26] **3** as a colourless oil (3.518 g, 70%) which crystallised upon freezing on long storage at -20 °C in freezer; $[\alpha]_D^{20}$ = +13.1° (c 1.0, $CHCl_3$); R_f = 0.48 (SiO_2 plate, EtOAc/n-hexane = 1:9); ^{1}H NMR (400.13 MHz, $CDCl_3$) d 1.16 (d, J = 6.5 Hz, 3H), 2.22 (m, 2H), 3.81 (tq, J = 6.5, 6.5 Hz, 1H), 4.64 (broad s, 1H), 5.02-5.19 (2 × m, 4H), 5.77 (ddt, J = 16.9, 10.2, 7.2, 1H), 7.28-7.45 (broad m, 5H); ^{13}C NMR (100.61 MHz, $CDCl_3$) d 20.6 (CH_3), 41.2 (CH_2), 46.6 (CH), 66.6 (CH_2), 118.1 (CH_2), 128.2 (2 × CH), 128.6 (CH), 134.3 (CH), 136.8 (C), 155.8 (C); IR (CH_2Cl_2) v_{max} 3324, 3068, 3033, 2972, 1693, 1642, 1530, 1454, 1405, 1376, 1334, 1280, 1250, 1221, 1100, 1058, 1028, 994, 915, 776 cm^{-1}; MS (ESI+) m/z (rel intensity) 242 [(M+Na)$^+$, 40%], 220 [(M+H)$^+$, 100], 176 (3), 159 (12), 117 (2), 88 (5); HRMS (ESI+) m/z calc'd for $C_{13}H_{18}NO_2$ [M+H]$^+$: 220.1332, found 220.1327 (Δ = -2.25 ppm). The structure of the product **3** was further characterised by 2D [1H, 1H] COSY, 2D [1H, 13C] HSQC, 2D [1H, 13C] HMBC NMR spectroscopic experiments.
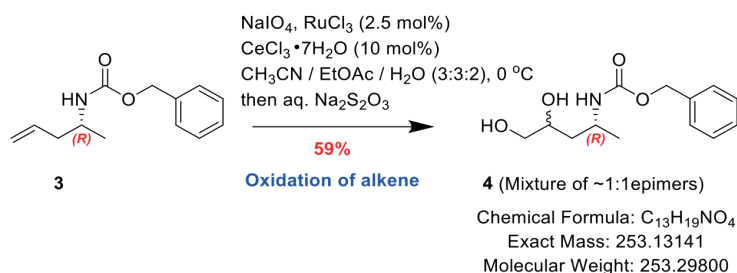
Benzyl [(2R)-4,5-dihydroxypentan-2-yl]carbamate, 4

NaIO$_4$, RuCl$_3$ (2.5 mol%)
CeCl$_3$·7H$_2$O (10 mol%)
CH$_3$CN / EtOAc / H$_2$O (3:3:2), 0 °C
then aq. Na$_2$S$_2$O$_3$

59%

Oxidation of alkene

3

4 (Mixture of ~1:1epimers)
Chemical Formula: C$_{13}$H$_{19}$NO$_4$
Exact Mass: 253.13141
Molecular Weight: 253.29800

Benzyl [(2R)-4,5-dihydroxypentan-2-yl]carbamate, **4** was prepared by ruthenium tetraoxide (generated in situ using RuCl$_3$/NaIO$_4$/CeCl$_3$·7H$_2$O) mediated oxidation of benzyl N-[(2R)-pent-4-en-2-yl]carbamate, **3** following a general procedure described

in the literature by Tiwari[28] *et al*. A mixture of $NaIO_4$ (1.749 g, 8.175 mmol, 1.5 eq.) and $CeCl_3 \cdot 7H_2O$ (0.203 g, 0.545 mmol, 0.1 eq.) in H_2O (11 mL) was stirred at r.t. for 10 min in a 100 mL round-bottomed flask fitted with a magnetic stirring bar. The off-white mixture was cooled to 0 °C and stirred for 4-5 min, followed by sequential addition of EtOAc (16 mL), CH_3CN (33 mL) and $RuCl_3$ (0.0283 g, 0.1363 mmol, 0.025 eq.) at 0 °C. The hetereogeneous mixture was then stirred for 2 min and a solution of the benzyl *N*-[(2*R*)-pent-4-en-2-yl]carbamate, **3** (1.195 g, 5.45 mmol, 1.0 eq.) in EtOAc (16 mL) was added in one shot to this mixture. The contents were stirred at 0 °C and progress of the reaction was monitored by TLC analysis (SiO_2; Eluant: EtOAc/*n*-hexane = 3:17 and also using Eluant: EtOAc = 100%). After 22 min, the reaction mixture turned more purple brown from black and was stopped by diluting the mixture using saturated $Na_2S_2O_3$ solution (200 mL) on a separating funnel and addition of additional EtOAc (150 mL). The mixture was shaken vigorously in the separating funnel and the aqueous layer was further extracted with EtOAc (2 × 75 mL). The combined organic layers were washed sequentially with saturated aqueous $Na_2S_2O_3$ solution (200 mL), saturated aqueous $NaHCO_3$ solution (2 × 200 mL) and finally with saturated brine solution (200 mL). The organic layer was separated, dried over anhydrous Na_2SO_4 and evaporated to dryness to obtain a colourless clear oil. The crude material was subjected to flash chromatography on SiO_2 [gradient; Eluant: EtOAc/*n*-hexane = 3:1 to EtOAc = 100%) to obtain the desired product, *benzyl [(2R)-4,5-dihydroxypentan-2-yl]carbamate*, **4** as a colourless oil (0.817 g, 59%), which was found to be ~1:1 inseparable mixture of epimers and rotamers as observed in ^1H and ^{13}C NMR spectroscopic analysis; R_f = 0.35 and 0.41 (for 2 epimers); ^1H NMR (400.13 MHz, $CDCl_3$) d (Mixture of epimers) 1.18 and 1.20 (2 × d, *J* = 7.0, 6.8 Hz, 3H), 1.24-1.38 and 1.40-1.76 (2 × m, 2H), 2.24 and 2.82 (2 × broad s, 1H), 3.08 and 3.40 (2 × broad s, 1H), 3.42-3.65 (2 × m, 2H), 3.70-3.80 (2 × m, 1H), 3.81-3.90 and 3.91-4.06 (2 × m, 1H), 4.23 and 4.95 (2 × app. d, *J* = 2.6 and 8.3 Hz, 1H), 5.07 and 5.10 (2 × app. d, *J* = 11.8 and 12.1 Hz, 2H), 7.27-7.44 (broad m, 5H); ^{13}C NMR (100.61 MHz, $CDCl_3$) d (Mixture of epimers) 21.6 (2 × CH_3), 40.3 (CH_2), 41.2 (CH_2), 44.1 (CH), 45.4 (CH), 66.6 (CH_2), 66.8 (CH_2), 67.2 (2 × CH_2), 68.8 (CH), 70.3 (CH), 128.2 (CH), 128.2 (2 × CH), 128.4 (CH), 128.6 (CH), 128.7 (CH), 136.3 (C), 136.6 (C), 156.3 (C), 157.3 (C); IR (CH_2Cl_2) v_{max} 3401, 3325, 2970, 2933, 1690, 1516, 1455, 1411, 1377, 1339, 1275, 1248, 1088, 1055, 1020, 871, 843, 764, 757 cm^{-1}; MS (ESI+) *m/z* (rel intensity) 276 [(M+Na)+, 100], 254 [(M+H)+, 12], 210 (14); HRMS (ESI+) *m/z* calc'd for $C_{13}H_{19}NO_4{}^{23}Na$ [M+Na]+ : 276.1206, found 276.1197 (Δ = -3.51 ppm) and *m/z* calc'd for $C_{13}H_{20}NO_4$ [M+H]+ : 254.1387, found 254.1384 (Δ = -1.28 ppm). The structure of the diol **4** was further characterised by 2D [1H, 1H] COSY, 2D [1H, 13C] HSQC, 2D [1H, 13C] HMBC NMR spectroscopic experiments.

(4R)-4-Aminopentane-1,2-diol, 5

4 (Mixture of ~1:1epimers)   **Deprotection of N-Cbz**   5 (Mixture of ~1:1epimers)
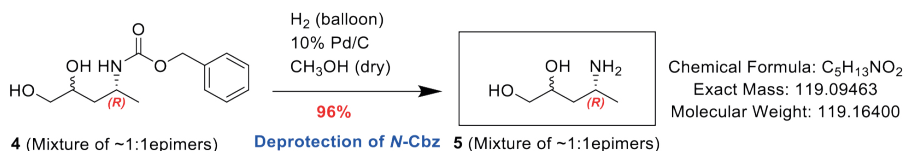
Final purification by Reveleris® Amino column;
Eluant: *n*-Hexane / EtOAc (gradient);
Purification using Grace Reveleris® X2 MPLC system

An oven-dried 100 ml round-bottomed flask with a magnetic stirring bar was fitted with a Quickfit® 3-way stopcock adapter and dried *in vacuo* using a heat gun and purged with nitrogen gas. This procedure was repeated thrice. The flask was cooled to r.t. and was charged with palladium on activated carbon (10% Pd, 0.500 g) under a nitrogen atmosphere and a solution of benzyl [(2*R*)-4,5-dihydroxypentan-2-yl] carbamate, **4** (0.778 g, 3.0715 mmol) dissolved in CH_3OH (25 mL) was added to it. The atmosphere inside the flask was evacuated using a diaphragm pump and purged with hydrogen gas (H_2 balloon) using the 3-way stopcock and this procedure was repeated thrice. The heterogenous mixture was then stirred under a hydrogen atmosphere (H_2 balloon) at r.t. and the progress of the reaction was monitored for the disappearance of the starting material **4** by TLC analysis (SiO_2 plate; Eluant:100% EtOAc) and LC-MS analysis (C18 reverse phase column, H_2O-CH_3CN as mobile phase, gradient). After 6 h, the reaction was judged to be complete and reaction mixture was filtered through Celite® 503 and washed with copious amount of CH_3OH (3 × 25 mL) and EtOAc (2 × 25 mL). The combined filtrate was evaporated to dryness to obtain *(4R)-4-aminopentane-1,2-diol*, **5** as a clear colourless oil (0.366 g, quant. yield) which was observed to be a ~1:1 mixture of epimers as observed from its ¹H and ¹³C NMR spectroscopic analysis. This was further purified and desalted on a Reveleris® Amino column using a Grace Reveleris® X2 MPLC system (gradient; using ELS detector; eluant: 100% *n*-hexane to *n*-hexane / EtOAc = 1:1; the product was isolated between *n*-hexane / EtOAc = 4:1 → 7:3) to afford *(4R)-4-aminopentane-1,2-diol*, **5** as a clear colour oil (0.353 g, 96% yield after purification). The aminodiol product **5** was stored in a freezer at -20 °C; ¹H NMR (400.13 MHz, $CDCl_3$) d (Mixture of epimers) 1.12 and 1.14 (2 × d, *J* = 3.0 and 3.1 Hz, 3H), 1.23-1.40 and 1.44-1.65 (2 × m, 2H), 3.00-3.11 and 3.18-3.24 (2 × m, 1H), 3.19-3.35 (broad s, 4H), 3.36-3.48 and 3.49-3.58 (2 × m, 2H), 3.80-3.88 and 3.89-3.96 (2 × m, 1H); ¹³C NMR (100.61 MHz, $CDCl_3$) d (Mixture of epimers) 24.2 (CH_3), 27.0 (CH_3), 40.4 (CH_2), 40.7 (CH_2), 44.4 (CH), 47.3 (CH), 66.8 (CH_2), 66.9 (CH_2), 69.8 (CH), 72.5 (CH); IR (CH_2Cl_2) v_{max} 3352, 3284, 2947, 2924, 2904, 2869, 1608, 1457, 1380, 1349, 1337, 1266, 1153, 1096, 1074, 1031, 988, 935, 853, 824, 732, 702 cm⁻¹; MS (ESI+) *m/z* (rel intensity) 161 [{(M+CH_3CN)+H}⁺, 7], 120 [(M+H)⁺, 100], 102 (3); HRMS (ESI +) *m/z*

calc'd for $C_5H_{14}NO_2$ [M+H]$^+$: 120.1019, found 120.1019 (Δ = -0.05 ppm). The structure of the aminodiol **5** was further characterised by 2D [1H, 1H] COSY, 2D [1H, 13C] HSQC, 2D [1H, 13C] HMBC NMR spectroscopic experiments.

Synthesis of *R*-methyl deoxypropanoguanine containing deoxyoligonucleotides

Although acetaldehyde forms both *R*- and *S*- stereoisomers of deoxypropanoguanine, the *R*- form is more stereochemically favourable for crosslink formation in a 5´-CpG-3´ on which we based this study[29]. A custom oligonucleotide, 5'-[phosphoryl]-GCACGAAAGAAGAGC-2FdI-GAAG was synthesized by Eurogentec at 1 μmol scale, using 5'-Dimethoxytrityl-2-fluoro-*O6*-p-nitrophenylethyl-2'-deoxyInosine,3'-[(2-cyanoethyl)-(*N,N*-diisopropyl)]-phosphoramidite and shipped on support. Support (20 mg, ca. 0.5 μmol oligo) containing 2FdIno, **6** was incubated with 12 mg (4*R*)-4-aminopentane-1,2-diol **5** in 280 μl DMSO with 140 μl TEA and incubated with agitation at rt overnight to give compound **7**. The support was washed three times with DMSO (200 μl) and three times with CH_3CN (400 μl), followed by deprotection of the *O6*-p-nitrophenylethyl group with 1 M DBU dissolved in CH_3CN (300 μl) at rt for 1 h, to give compound **8**. The support was washed three times with CH_3CN, dried with N_2 and treated with aq. 28% NH_4OH at 55 °C for 6 h to remove remaining protecting groups and elute the oligonucleotide from the support. The oligonucleotide was freeze-dried (Scanvac) and reactants separated by HPLC (Thermo Accela) on a AdvanceBio Oligonucleotide, 4.6 x 150 mm, 2.7 μm (Agilent) in 15 mM TEA/ 400 mM HFIP, (pH 7.0) at 0.5 ml min^{-1} at 60°C, over 30 min, effecting a 15 – 27.5% MeOH gradient; **8** eluted at 19.2 min and **6** at 20.5 min. Fractions containing **8** were lyophilized and reacted with 50 mM $NaIO_4$ for 1 h at rt, and the product desalted over an Oligo R3 column (Thermo Scientific). The identity of **9** was confirmed by MALDI and/or LC-MS (calc'd mass: 6408.18, found 6409.0 (± ≤ 0.02% error).

R-methyl deoxypropanoguanine crosslinking and purification

Unmodified oligonucleotides were supplied HPLC purified and verified by LC-MS (Integrated DNA Technologies, BVBA). Either deoxyguanine containing (5'-[phosphoryl]-CTTTCTTCTC-dG-CCTTCTCCC) or deoxyinosine (5'-[phosphoryl]-CTTTCTTCTC-dI-CCTTCTCCC) oligonucleotides were mixed at 1:1 molar ratio with the PdG containing oligo (5'-[phosphoryl]-GCACGAAAGAAGAGC-PdG-GAAG) in PBS, annealed (85°C for 5 min, ramped to 25 °C, at - 0.1°C s^{-1}) and incubated at 37°C for 7-14 d to allow crosslink formation. The reaction progression was analysed by denaturing PAGE stained with Sybr Gold (Thermo Scientific). Final PdG-dG crosslink yields were approx. 50%. Crosslinks were purified by denaturing PAGE and/or HPLC. For PAGE purification, bands were stained briefly with Sybr Gold, visualized by trans-illumination at 450 nm (Clare Chemical Research) and excised. DNA was electroeluted from the gel (Elutrap, Whatman), concentrated by ethanol precipitation and quantified by Nanodrop (260 nm). HPLC purification of the crosslink followed

the same conditions detailed above, crosslinked oligonucleotides eluted between 25.8 – 26.2 min, and were confirmed by denaturing PAGE analysis. Fractions containing crosslinked DNA were evaporated to dryness and reconstituted in 1 x PBS.

Crosslink Stability Assay
Crosslinked DNA was back-filled with exo-deficient Klenow (New England Biolabs) to incorporate a single ^{32}P-CTP radiolabel. Labelled oligos were purified by R3 Oligo resin and quantified by scintillation counting (Perkin Elmer). Labelled oligos were incubated in 1 x PBS and after the indicated temperature and time combinations samples were removed, mixed with formamide loading buffer and resolved by formamide-urea denaturing PAGE. Dried gels were exposed on phosphorimager 24-48 h and visualized using an Amersham Typhoon Biomolecular Imager (GE Healthcare). Crosslink stability was calculated from band densitometry using Fiji software[30] and the results were plotted using Prism 7 (Graphpad Software) from three independent experiments.

Preparation of plasmid substrates

The plasmids containing site-specific interstrand crosslinks were prepared as described[31,17]. Briefly, duplexes containing ICLs derived from cisplatin, psoralen or acetaldehyde were ligated into a vector linearized with Bbs1. After ligation, the plasmid was purified using a cesium chloride gradient. To ligate the psoralen duplex, the pSVRLuc backbone was modified by site-directed mutagenesis to substitute an adenine to guanine present on one of the Bbs1 5 'overhangs.

The plasmids containing the *lacO* sequence were prepared by ligating the crosslinked oligos duplexes into pSVR*lacO* vector (gift from Johannes Walter)[17]. To make pICL-AAreverse- *lacO*, the pSVR*lacO* was first digested with Not1 and a new insert containing Bbs1 overhangs in reversed position was inserted. This plasmid was used to insert the crosslinked duplexes. To make pPdG, ligation of a duplex containing a PdG into a backbone linearized by Bbs1 was followed by cesium gradient purification, as described above.

Xenopus repair assay

All animal procedures and experiments were performed in accordance with national animal welfare laws and were reviewed by the Animal Ethics Committee of the Royal Netherlands Academy of Arts and Sciences (KNAW). All animal experiments were conducted under a project license granted by the Central Committee Animal Experimentation (CCD) of the Dutch government and approved

by the Hubrecht Institute Animal Welfare Body (IvD), with project license number AVD80100201711044. Sample sizes were chosen based on previous experience, randomization and blinding are not relevant to this study.

Xenopus egg Extracts and DNA Replication

Xenopus laevis female frogs (aged >2 years) were purchased from Nasco and used as a source of eggs. Preparation of *Xenopus* egg extracts and DNA replication were performed as previously described[32,33]. For DNA replication, plasmids were incubated in a high-speed supernatant extract (HSS) at a final concentration of 7.5 ng µl^{-1} for 20 min at room temperature to license the DNA. Two volumes of nucleoplasmic extract (NPE) were added to start DNA replication. For nascent strand labeling, HSS was supplemented with ^{32}P-α-dCTPs. Where indicated, an unrelated non-damaged control plasmid (pQuant) was added at concentrations up to 0.8 ng ul^{-1}, to be used as internal control for quantifications. To block CMG unloading, p97 inhibitor (NMS-873, Sigma)[34] was added to NPE at a final concentration of up to 40 µM. To inhibit DNA replication recombinant Geminin was added to HSS at 300 nM. For the reactions with the pICL-*lacO* plasmids, the plasmids were incubated prior to the start of replication with purified biotinylated LacR protein at a final concentration of 17 µM for 45 minutes at room temperature[35]. For CMG uncoupling experiments, aphidicolin (Sigma) was added to the replication mix 5 minutes after the start of replication, at a final concentration of 12 µM.

For analysis of undigested DNA replication products, replication reactions were stopped by adding five volumes of replication stop solution I (Stop I: 80 mM Tris pH 8, 5% SDS, 0.13% phosphoric acid, 10% Ficoll, 8 mM EDTA, 0.1% bromophenol blue). The samples were treated with proteinase K (1.5 µg µl^{-1}) for 40 min at 37°C and resolved by 0.8% native agarose gel electrophoresis. The gel was dried and visualized by autoradiography.

For extraction of DNA from replication reactions, aliquots of the reaction were stopped with 10 volumes of replication stop solution II (Stop II: 50 mM Tris pH 7.5, 0.5% SDS, 10 mM EDTA pH 8). Samples were then treated with RNase (0.13 µg µl^{-1} final concentration) for 30 min at 37°C, followed by Proteinase K (0.5 µg µl^{-1}) treatment for 1 h at 37 °C or overnight at room temperature. DNA was phenol/chloroform extracted, ethanol precipitated with glycogen (0.3 µg µl^{-1}), and resuspended in 10 mM Tris pH 7.5 in a volume equal to the reaction sample taken.

Antibodies and Immunodepletions

Antibodies against *xl*FANCD2, *xl*REV1, *xl*REV7 were previously described[12,14,19]. The *xl*NEIL3 antibody was raised against a C-terminal peptide of *Xenopus laevis* NEIL3 and affinity purified (New England peptide). FANCD2 and REV7 depletions were carried out as described[12,19]. To rescue the FANCD2 depletion 260 nM of recombinant FANCI-FANCD2 was added to NPE. For depletion of NEIL3, one volume of protein A Sepharose Fast Flow beads (GE Healthcare) was incubated

with 2 volumes of *xl*NEIL3 affinity purified antibody (1 mg ml^{-1}) overnight at 4°C. For depletion from NPE, three rounds were performed in which 1 volume of antibody-bound PAS beads was incubated with 4.5 volumes of extract for 20 min at room temperature. HSS was depleted using the same volumes but for 2 rounds. Recombinant NEIL3 was added to NPE at 150 nM final concentration. For REV1 depletion, 1 volume of Dynabeads™ Protein A beads (Thermo Fisher Scientific) was pre-incubated with 0.5 volumes of antibody for 30 minutes at room temperature. To deplete NPE, 1 volume of antibody-bound beads was incubate with 1.5 volume of extract for 30 minutes at RT for two rounds. HSS was depleted with the same volume of beads in one round.

Protein purification
Recombinant *xl*FANCI-FANCD2 complex was prepared as previously described[12]. Biotinylated LacR protein was prepared as described[35].

NotI repair assay
Plasmids were replicated in *Xenopus* egg extract in the absence of ^{32}P-α-dCTPs and intermediates were purified as described above. Replication reactions were supplemented with a modified pQuant plasmid (pQuant 28nt) that produces a 28 nt fragment upon NotI digestion, which was used as internal control for quantification. Repair intermediates were digested with NotI and 3' labeled by fill-in of the 5'-overhangs with Sequenase DNA polymerase (USB) in the presence of ^{32}P-α-dCTPs and non-labeled dGTP. One volume of denaturing PAGE Gel Loading Buffer II (Invitrogen™) was added, samples were denatured by incubation at 98 °C for 3 minutes, and the DNA fragments were separated by 20% Urea PAGE, products were visualized by autoradiography and quantified using ImageQuant software (GE Healthcare). NotI digested and labeled products that still contain an ICL are 88 nt in size while non- crosslinked molecules are 44 nt. The temporal reduction of the 88 nt product is caused by ongoing ICL unhooking but also by replication fork stalling because this results in a DNA structure that is partially single stranded and therefore not digested by NotI.

The temporal increase of the uncrosslinked 44 nt product was used as a readout for ongoing ICL repair because these are the products of TLS or HR. First, the intensity of pQuant 28 nt fragment was used to normalize the intensity of the 44 nt products, to make these products independent of amplification by replication/ repair and to correct for extraction/loading variations. Then, the normalized linear 44 nt products were expressed as percentage of the total of 88 and 44 nt fragments at time point zero, which represents the percentage of the input DNA that is accumulating over time in the 44 nt band. Finally, the percentage of non-crosslinked background was subtracted to yield the percentage of molecules that have undergone TLS or HR during repair. Of note, 2 or 3 bands can be seen at the height of the 44 nt fragment in some gels. First, the sequence difference between

the bottom and top strand causes a slight difference in mobility. Second, upon ICL repair by the FA pathway the unhooked adduct will cause a shift in size (see also Extended Data Figure 5k).

Nascent strand analysis

Nascent strand analysis was performed as previously described[14]. In brief, DNA repair products were digested either with AflII, with AflIII or EcoRI, or with AflIII and BamHI (New England Biolabs). After addition of one volume of denaturing PAGE Gel Loading Buffer II (Invitrogen™), the samples were separated on a 7% polyacrylamide sequencing gel, and visualized by autoradiography. The sequencing gel ladder was produced using the Thermo Sequenase Cycle Sequencing Kit (USB) and primer S (CATGTTTTACTAGCCAGATTTTTCCTCCTCTCCTG).

APE1 glycosylase assay

Plasmids were replicated in the presence of pQuant. After extraction of the DNA repair intermediates, samples were digested with HincII (New England Biolabs) only and with HincII in combination with APE1 (New England Biolabs). The digested products were separated on a 0.8% native agarose gel and visualized by autoradiography. Quantification was done using Image Quant (GE Healthcare). The arm fragments were first normalized against pQuant and the highest value was set to 1.

HincII incision assay

ICL-containing plasmids were replicated in the presence of pQuant. Repair intermediates were extracted and digested with HincII. Fragments were separated on a 0.8% native agarose gel and visualized by autoradiography. Quantification was performed using ImageQuant software (GE Healthcare). The intensity of the HincII arm fragments was first normalized to pQuant, the highest value in the experiment was set to 1 and the data was plotted against time.

Strand-specific Southern blot assay

Strand-specific Southern blot was performed as described previously[14,17]. Briefly, purified replication intermediates were digested with AflIII and AseI, 0.5 volume of denaturing PAGE Gel Loading Buffer II (Invitrogen™) was added, and the DNA fragments were separated on a 7% denaturing polyacrylamide sequencing gel. The DNA was transferred to a Hybond-N+ membrane (Amersham) overnight at 4°C. The membrane was rinsed once with 2X SSC buffer and UV-irradiated to crosslink the DNA to the membrane. The membrane was pre-hybridized with 12 ml of Ultrahyb buffer (Ambion) for 3 h at 42°C and probed overnight at 42°C. The membrane was then washed three times at 42°C in pre-warmed washing buffer (0.5X SCC, 0.25% SDS), dried and exposed to a phosphorimager screen. To generate the probes pCon was amplified by PCR using the primers XL-1 (CCTGCTGTCCATTCCTTATTCC)

and XL-3 (GCATTGGTTCTGCACTTCCGC). The amplicon was purified and used as template for a primer extension reaction in the presence of ^{32}P-α-dCTPs to generate the probes[14] . Primer S was used for the probe that detects the bottom strand and primer XL-3 for the top strand.

Primer extension assay

Primer extension assay was performed as previously described[19]. In brief, late ICL repair products were isolated and purified as described above. Unreplicated pControl and pPdG were used as controls. Samples were digested with AflIII and BamHI, and the DNA fragments were used as template for primer extension. Primers (Forward: CTCGAGCGGAAGTGCAGAAC, Reverse: AATACGCAAACCGCCTCTCC) were first labelled radioactively with PNK at the 5' end with ^{32}P-γ-ATP and were subsequently annealed to DNA in a thermocycler. Each reaction contained only one primer (either forward or reverse). The primers were then extended using the Phusion High-Fidelity DNA polymerase (NEB) for one round. The resulting DNA fragments were first concentrated by ethanol precipitation (see above) and then separated on a 20% denaturing PAGE gel and visualized by autoradiography.

Sequencing analysis of DNA repair products

Sequencing of the DNA repair products was performed using a similar method as described previously[19]. Briefly, plasmids were replicated in *Xenopus* egg extract and late replication/repair intermediates (240 min for the crosslinked plasmids, 120 min for pCon and pPdG) were isolated. The DNA was amplified by PCR using the Phusion High-Fidelity DNA Polymerase (New England Biolabs) for 15 cycles using primers with unique barcodes. For experiment 1 we used forward primer: NNNNNN-barcode-AGAACCAATGCATGCGGC and reverse primer: NNNNNNATTCATTAATGCAGCGGATCG, where N represents random nucleotides and the barcode is composed of 3 nucleotides. For experiment 2 we used forward primer: NNNNNN-barcode-CTCGAGCGGAAGTGCAGAAC and reverse primer: NNNNNN-barcode-AATACGCAAACCGCCTCTCC, the barcode was composed of 6 nucleotides. This generated a product of 96 nt for experiment 1 presented in the main figures, and 154 nt for experiment 2 in the supplement, with the lesion site roughly in the middle (Extended Data Item 8a). In addition to the replicated DNA samples, we also used 75 ng unreplicated pICL-Pt, pICL-AA and pPdG as template in independent PCR reactions. For each sample two independent PCRs were performed to create technical replicates. The products were separated by TBE PAGE and visualized by Sybr Gold (Thermo Fisher Scientific) staining. For experiment 1 we excised the 96 nt fragment including products up to ~10 nt smaller, for experiment 2 we excised the 154 nt band. Gel slices were broken into small pieces and incubated in elution buffer (10 mM Tris pH 8.0, 1 mM EDTA pH 8.0, 300

mM NaCl) overnight at room temperature under agitation. The DNA was recovered using a Spin-X column (0.45 µM, Sigma) and further purified using paramagnetic beads (CleanPCR kit, CleanNA). The DNA was ethanol-precipitated in the presence of 0.15 mg ml^{-1} glycogen and 0.3 M sodium acetate and resuspended in 10 mM Tris pH 8.0. Illumina TruSeq adapters were ligated to the DNA according to the manufacturer's instructions (Illumina), and the DNA was sequenced using paired end 75 bp NextSeq sequencing for dataset 1, and paired end 150 bp NextSeq sequencing for dataset 2 (Utrecht sequencing facility).

Demultiplexing

Data from experiment 1 was demultiplexed on 3 nt barcodes at the 5'-ends, downstream of a 6 nt random nucleotide segment of the first mate (R1) of read pairs, allowing no mismatches. The 3 nt barcodes were extended on the 3'-end with 9 nt of the forward PCR primer ("AGAACCAAT") to avoid reverse complement reads demultiplexing as forward reads. Reads not demultiplexed as forward reads were then demultiplexed on the second mate (R2) of read pairs, again allowing no mismatches, and again with the 3 nt barcodes which were extended with 9 nt of the forward PCR primer. Data from experiment 2 was demultiplexed on 6 nt barcodes prefixed at the 5'-ends of the first mate (R1) of read pairs, allowing at most 1 mismatch (no indels). Barcodes were designed such that 1 mismatch could be tolerated without introducing ambiguity. For dataset 2, R2 of reads demultiplexed with forward primers and R1 of reads demultiplexed with reverse primers was taken the reverse complement of (using fastx_reverse_complement from FASTX_toolkit v0.0.14). For dataset 1, R2 of demultiplexed reads were trimmed by 6 nt. (using fastx_trimmer from FASTX_toolkit v0.0.14) and then taken the reverse complement of. To ensure R1 and R2 contained the same segment of the amplicon (which is larger than the read lengths in dataset 1) and did not dovetail, R1 was trimmed at the 5'-end using "ATGCGGCCGC" and allowing at most 1 mismatch and no indels, using cutadapt (v1.9.1[36]) and at the 3'-end using "GCGCGGCCGCG", again allowing at most 1 mismatch and no indels. Similarly, R2 reads were trimmed using "GCCGCGAAG" at the 5'-end and "GCATTAA" at the 3'-end. To make the data of dataset 1 comparable to dataset 2, R1 reads of dataset 1 were then prefixed with "CTCGAGCGGAAGTGCAGAACCAATGCATGCGGCCGC" and suffixed with "GCGCGGCCGCGATCCGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCG GTTTGCGTATT", and R2 reads were prefixed with "CTCGAGCGGAAGTGCAGAACCAATGCATGCGGCCGCGAAG" and suffixed with "GCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATT", reconstituting the amplicon of dataset 2.

Alignment to amplicon

Reads were mapped to the amplicon sequence of experiment 2 (Extended Data Item 8a) using blastn (from the BLAST suite, v2.4.0+, with parameters -dust no -parse_

deflines -max_hsps 1 -num_alignments 1 -word_size 7 -evalue 100.0 -outfmt "6 qseqid sstart send qstart qend btop" -strand plus). The blast result generated a table with one line per read, for both the R1 and R2 read mates. From these tables, the Blast trace-back operations (BTOP) string was parsed to extract information about mismatches in the read relative to the amplicon sequence. These tables were further processed using python software. This software filters out read pairs where R1 and R2 do not agree, or do not fully contain the 130 nt amplicon. This software then summarized the BLAST tables into occurrences of each BLAST BTOP, for each sample. Read numbers for total reads, pair-matched reads, and reads with indels and insertions are listed in Extended Data Item 8b and c.

Post-processing of mutation count tables

Technical replicates (such as sequencing lanes and PCR replicates) were merged by summation. Because the reads of dataset 1 and dataset 2 are of different length, dataset 1 was only informative on mutations in the amplicon interval [40, 76) (0-based, half-open interval). Therefore, normalization was done using only mutations that were observed in both datasets (with at least 100 reads). Observed mutation (BTOP) counts were normalized by

$$N_{i,j} = \frac{C_{i,j} + \phi}{\sum_{i \in W} C_{i,j} + \phi |W|}$$

where $C_{i,j}$ is the count of mutation i in sample j, W is the set of mutations that meet the cutoff criterion in both datasets, ϕ is a pseudocount (set to 1 in further analyses) and $N_{i,j}$ is the normalized count of mutation i in sample j.

Cellular Survival Assays

Cell lines

Wild-type and *NEIL3*- HAP1 near-haploid human cells were purchased from Horizon Discovery and cultured at 37 °C and 5% CO2 in IMDM supplemented with 10% fetal calf serum and penicillin/streptomycin. *NEIL3*- cells were confirmed by immunoblotting against NEIL3. For targeting of *FANCL*, WT and *NEIL3*- cells were transfected using Turbofectin 8.0 (Origene) and the following plasmids: pX461 (Cas9 nickase) (Addgene), gRNA vectors (FANCL_left: CCTAATGCAATTCTGCGTGCTGT and FANCL_right: TTTTTCTGGCTCAAGTACCCAGG), and a *FANCL*-Puro targeting construct (Wellcome Trust Sanger Institute). Two days post-transfection, 3.5 µg ml^{-1} puromycin was added, and two days later, cells were plated in 96-well plates with puromycin at a density of 1 x 10^4 cells per well. After 14 days of incubation, individual clones were picked and analyzed for *FANCL* targeting using the SequalPrep Long PCR kit (Applied Biosystems) (using primers FANCL_ LR_FW2,

TGTCTACCCCCTAAGTTCGTTGA; EF1a_R1, GCGATCTCTGGGTTCTACG TTAGTG). Targeted clones were then plated with 100 ng ml^{-1} mitomycin C overnight and analyzed by immunoblotting for FANCD2. *FANCL* knockouts were identified by failure to ubiquitinate FANCD2. All cell lines were confirmed mycoplasma negative using the MycoAlert Mycoplasma Detection Kit (Lonza).

Clonogenic Survival Assay

For colony survival assay, HAP1 cells were prepared at 2 x 10^5 cells ml^{-1}. Cells and acetaldehyde (diluted in culture media) were mixed in 96-well blocks (Greiner Bio-One Masterblock) and foil seals (Bio-Rad Microseal 'F') were applied before culturing cells at 37°C for 2 h. Cells were then serially diluted in PBS using a multi-channel pipette to obtain 1:10 and 1:100 dilutions, and 100 μl of each were plated in duplicate in 24-well plates filled with 1.5 mL of culture media per well. Cells were cultured for 6 days before being stained with crystal violet[37] and colonies were quantified by a GelCount colony counter (Oxford Optronix). Data presented reflects a minimum of three independent experiments, error bars denote S.E.M.

Supplemental note 1 references

26.   Gebauer, J., Rost, D. & Blechert, S. A highly efficient synthesis of (-)-pinidinol. *Heterocycles* **68**, 2129-2132 (2006).

27.   Mahesh, M., Murphy, J.A., LeStrat, F. & Wessel, H.P. Reduction of arenediazonium salts by tetrakis(dimethylamino)ethylene (TDAE): Efficient formation of products derived from aryl radicals. *Beilstein Journal of Organic Chemistry* **5**(2009).

28.   Tiwari, P. & Misra, A.K. An efficient stereoselective dihydroxylation of glycals using a bimetallic system, RuCl3/CeCl3/NaIO4. *Journal of Organic Chemistry* **71**, 2911-2913 (2006).

29.   Lao, Y.B. & Hecht, S.S. Synthesis and properties of an acetaldehyde-derived oligonucleotide interstrand cross-link. *Chemical Research in Toxicology* **18**, 711-721 (2005).

30.   Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9**, 676-682 (2012).

31.   Enoiu, M., Jiricny, J. & Scharer, O.D. Repair of cisplatin-induced DNA interstrand crosslinks by a replication-independent pathway involving transcription-coupled repair and translesion synthesis. *Nucleic Acids Research* **40**, 8953-8964 (2012).

32.   Lebofsky, R., Takahashi, T. & Walter, J.C. DNA replication in nucleus-free Xenopus egg extracts. *Methods Mol Biol* **521**, 229-52 (2009).

33.   Walter, J., Sun, L. & Newport, J. Regulated chromosomal DNA replication in the absence of a nucleus. *Molecular Cell* **1**, 519-529 (1998).

34.   Magnaghi, P. et al. Covalent and allosteric inhibitors of the ATPase VCP/p97 induce cancer cell death. *Nature Chemical Biology* **9**, 548-556 (2013).

35.   Dewar, J.M., Budzowska, M. & Walter, J.C. The mechanism of DNA replication termination in vertebrates. *Nature* **525**, 345-350 (2015).

36.   Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10-12 (2011).

37.   Franken, N.A., Rodermond, H.M., Stap, J., Haveman, J. & van Bree, C. Clonogenic assay of cells in vitro. *Nat Protoc* **1**, 2315-9 (2006).

2

Supplemental note 2

SI Figure 1

NMR spectra of compounds generated for the synthesis of the aminodiol (*5*) crosslinker. **a**. Scheme for the synthesis of aminodiol. **b**. ¹H NMR spectrum and 13C{1H}, ¹³C DEPT- 135 and ¹³C DEPT-90 NMR spectra of (*R*)-2-(Pent-4-en-2-yl)isoindoline-1,3-dione, 2. **c**. ¹H 13 13 NMR spectrum and 13C{1H}, C DEPT-135 and C DEPT-90 NMR spectra of benzyl *N*- 1 13 [(2*R*)-pent-4-en-2-yl] carbamate, 3. **d**. H NMR spectrum and 13C{1H}, C DEPT-135 and ¹³C DEPT-90 NMR spectra of benzyl [(2*R*)-4,5-dihydroxypentan-2-yl]carbamate, 4. **e**. ¹H NMR spectrum and 13C{1H}, ¹³C DEPT-135 and ¹³C DEPT-90 NMR spectra of (4*R*)-4- aminopentane-1,2-diol, 5.

a

b   ## ^{1}H NMR Spectrum

13C{1H}, ^{13}C DEPT -135 and ^{13}C DEPT -90 NMR Spectra

c

^{1}H NMR Spectrum

13C{1H}, ^{13}C DEPT -135 and ^{13}C DEPT -90 NMR Spectra

d

¹H NMR Spectrum

4 (Mixture of epimers)
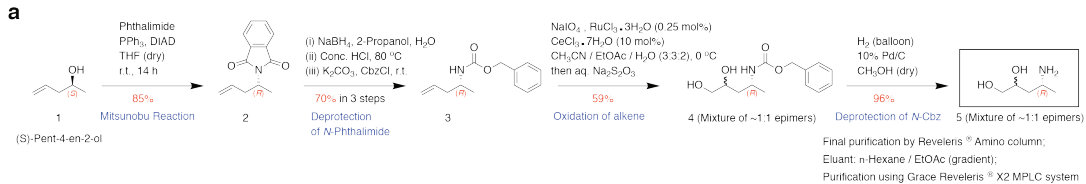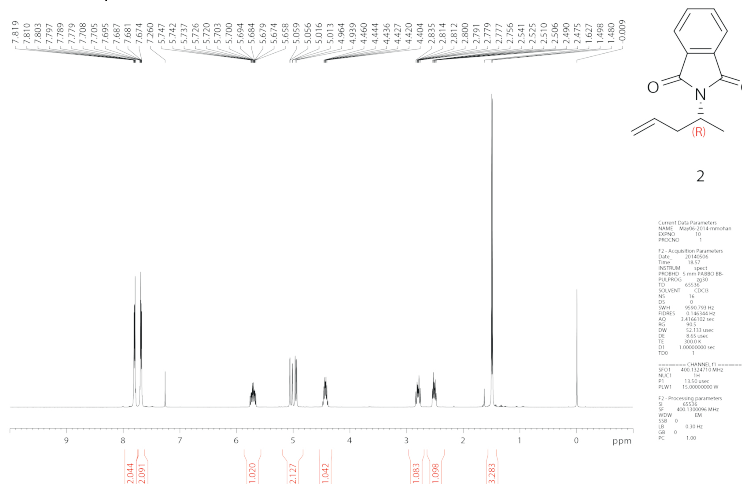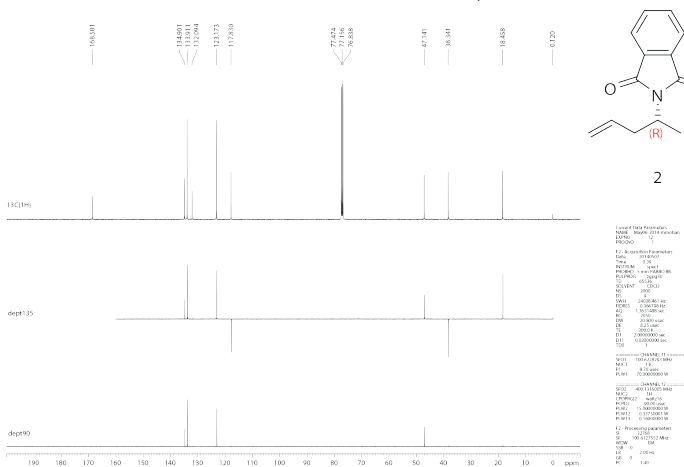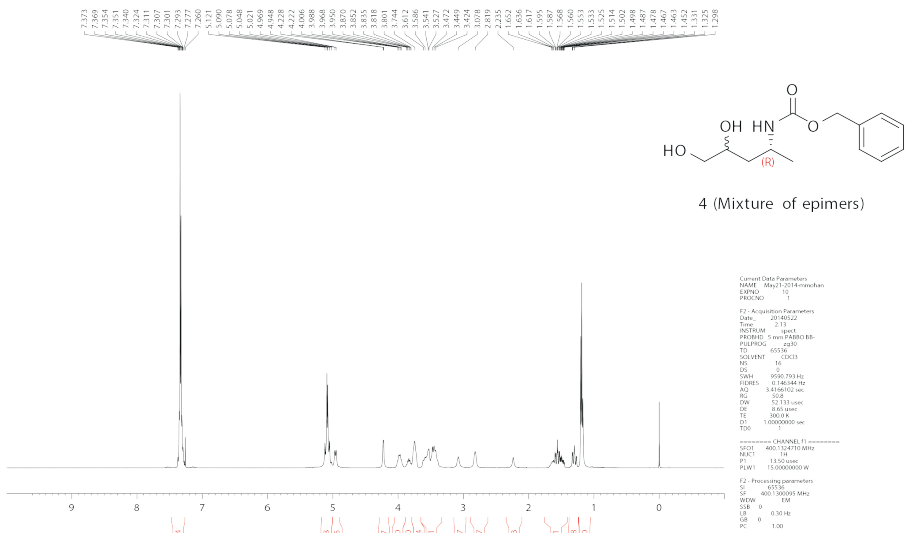
2

13C{1H}, ¹³C DEPT -135 and ¹³C DEPT -90 NMR Spectra !

4 (Mixture of epimers)
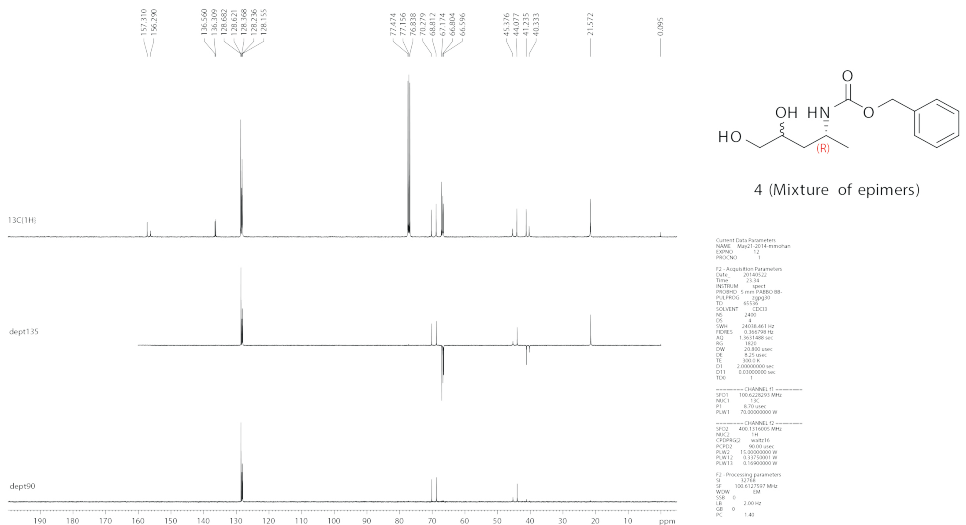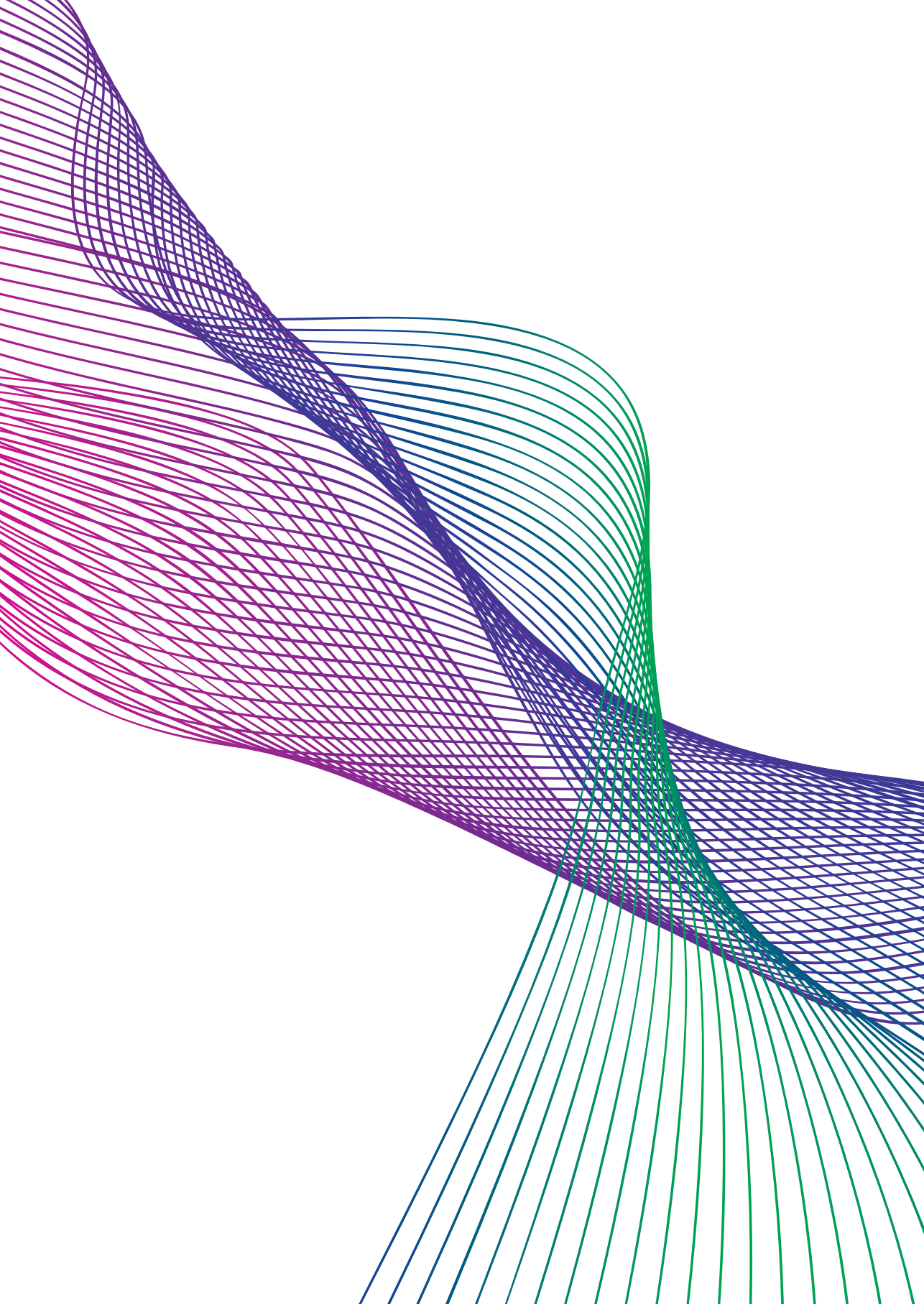
CHAPTER 3

Nucleosome remodeling during DNA interstrand crosslink repair in *Xenopus* egg extracts

Alice Bolner[1], Merijn Kerstens[1], Roxanne van der Sluijs[1], Merlijn Witte[1], Jamie Barnett[1], Harm Post[2], Maarten Altelaar[2] and Puck Knipscheer[1]

[1]*Hubrecht Institute – KNAW, University Medical Center Utrecht & Oncode Institute, the Netherlands*
[2]*Biomolecular Mass Spectrometry and Proteomics, Bijvoet Centre for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, the Netherlands*
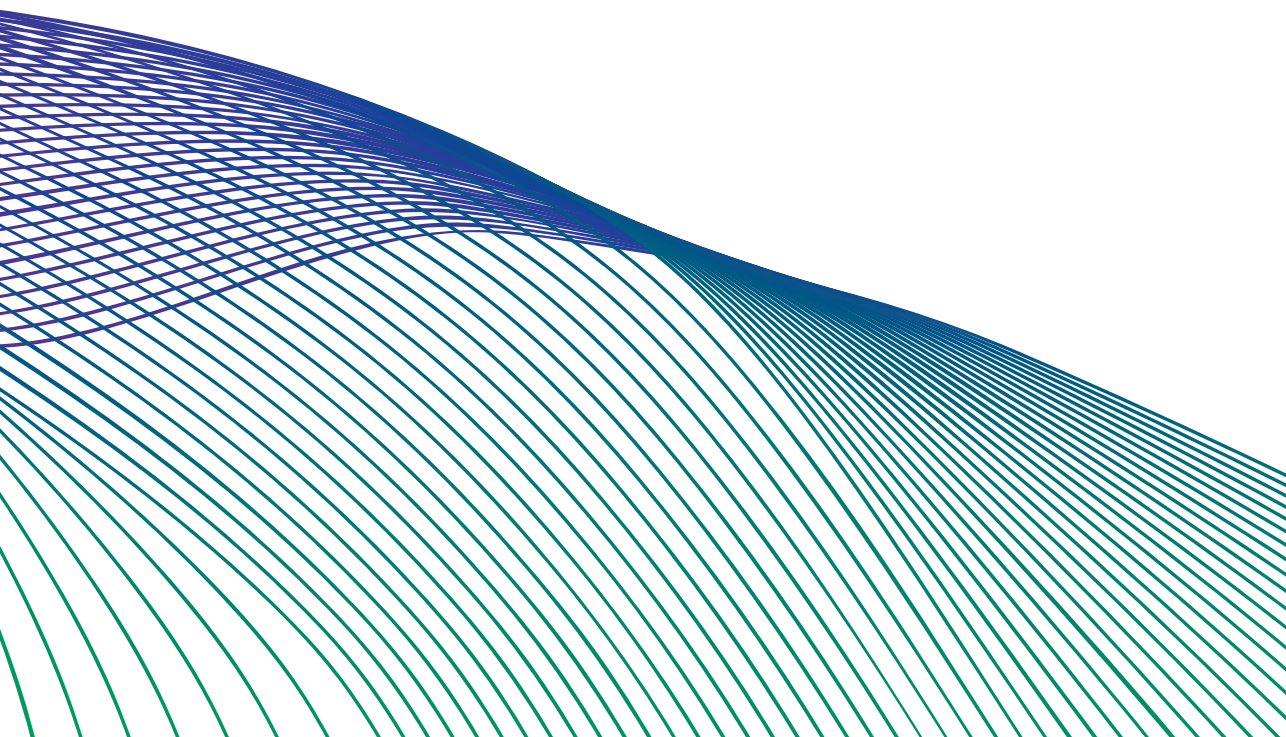
DNA repair is thought to require reorganization of the chromatin environment to allow access of repair proteins and promote DNA damage signaling. However, whether and how this reorganization takes place in the context of DNA interstrand crosslink (ICL) repair is unknown. Here, we describe a method to map nucleosomes during DNA replication-dependent ICL repair in *Xenopus* egg extracts. Using this method, we show that during ICL repair nucleosomes are depleted progressively from the ICL up to 1.5 kb away from it. At the final stages of crosslink repair, nucleosomes are restored both at and around the lesion site. These nucleosomal rearrangements are coupled to repair and take place in two different ICL repair pathways. Using a mass spectrometry-based approach, we identified chromatin remodeling factors that are specifically recruited to the plasmid during ICL repair. Altogether, we present the first demonstration of nucleosome dynamics during ICL repair and provide a unique opportunity to dissect the molecular mechanisms that controls this.

Introduction

The eukaryotic genome is packed into chromatin, a nucleoprotein structure that constitutes the physiological landscape for all DNA processes, including DNA replication and repair. The building block of chromatin is the nucleosome[1,2], a repeating nucleoprotein structure composed of DNA wrapped around a histone octamer formed by a $(H3-H4)_2$ tetramer and two H2A-H2B dimers. Together with the interconnecting DNA, nucleosomes appear as "beads on a string"[1]. Dynamic changes in the organization of chromatin are essential, as they allow access of regulatory factors to the genetic material. The modulation of chromatin conformation can occur through multiple mechanisms. Chromatin structure can be opened or tightened through chemical modification of histones (such as addition of small chemical groups or small peptides)[3], through incorporation of histone variants in nucleosomes[4,5], or by the action of ATP-dependent remodelers that can evict or slide nucleosomes[6].

As chromatin restricts the binding of repair factors to DNA, DNA repair is coupled to nucleosomal rearrangements that facilitate access to the lesion and promote repair[7]. Previous studies showed that regions of chromatin with active DNA repair were more sensitive to nuclease treatment[8-11], providing the basis for the "Access-Repair-Restore" (ARR) model[12]. In this model, DNA damage signaling leads to nucleosomal rearrangements that increase DNA accessibility in mammalian cells. These changes in the chromatin structure are not limited to the lesion site, but can extend several kilobases from the damage[13]. As DNA repair proceeds, nucleosomes are restored and the chromatin structure reverts to a more condensed state. These conformational changes promote DNA repair by allowing repair factors to bind to the lesion site and regulate the DNA damage response[14-16]. An example of how changes in chromatin can regulate DNA damage signaling is the phosphorylation of the histone variant H2A.X at the site of DNA double stranded breaks (DSBs)[17]. This histone variant is phosphorylated by DNA damage checkpoint kinases on Ser139 generating γH2A.X[17], that promotes the recruitment of repair proteins and DNA damage checkpoint signaling factors[18-21]. Although the role of chromatin remodeling in the DNA damage response has been extensively studied, there is still a lack of mechanistic understanding regarding how histone modifications can recruit specific repair factors or remodelers and how this process is regulated. While current data on chromatin remodeling during DNA repair mostly derives from studies on DSB and nucleotide excision repair (NER), these studies indicate that different DNA repair pathways may be associate to distinct mechanisms of chromatin remodeling.

DNA interstrand crosslinks (ICLs) are toxic DNA lesions that covalently bind the two strands of the DNA together and are induced by chemotherapeutic agents or by metabolic by-products in cells (see introduction, chapter 2)[22-24]. The main pathway of ICL repair is the Fanconi anemia pathway, which is mutated in the cancer predisposition syndrome Fanconi anemia (FA). The currently known 22

FA proteins, including endonucleases, recombinases, and translesion synthesis DNA polymerases, act together in a complex process that is coupled to DNA replication[25,26]. Repair in this pathway is initiated upon stalling of two DNA replication forks on either side of the ICL (Figure S1)[27]. This is followed by TRAIP (ubiquitin E3 ligase TRAF-interacting protein)-dependent ubiquitination of the CMG (CDC45, MCM2-7, and GINS) helicase and the subsequent extraction of the CMG from DNA by the p97 ATPase[28,29]. After CMG unloading, one of the leading strands 'approaches' the ICL up to 1 nt from the crosslink, promoting the subsequent steps of repair[30,31] (see Figure S1 and introduction of this thesis). The approached DNA replication fork can undergo fork reversal, a process in which reannealing of the parental strands creates a four-way DNA structure[32]. The crosslink is then unhooked via dual nucleolytic incisions on either side of the ICL[33], in a mechanism dependent on the structure specific endonuclease XPF-ERCC1[34]. The unhooked ICL is bypassed by translesion synthesis (TLS) and the newly synthesized DNA strand is used as template for homology-directed repair of the incised strand[35,36] (Figure S1). In addition to the FA pathway, two other ICL repair pathways have recently been described. The glycosylase pathway that repairs psoralen/UV-induced ICLs and unhooks the crosslink via cleavage of the *N*-glycosyl bond[37] (see introduction of this thesis and Figure S1), and a novel pathway that repairs acetaldehyde-induced ICLs in which the crosslink is unhooked by cutting within the ICL[38] (see chapter 2). While in recent years we gained major insights into the function of the ICL repair factors, there is virtually nothing known about the role and mechanism of chromatin remodeling in ICL repair.

In this study we investigate nucleosome dynamics during ICL repair using the *Xenopus* egg extract system that supports efficient nucleosome assembly during DNA replication and repair[39,40]. To measure nucleosome positioning, we developed an assay based on MNase digestion and quantitative PCR. Using this method, we show that during ICL repair nucleosomes are sequentially removed first from a region close to the crosslink, and then from a region up to 1.5 kb away from the lesion. At the final stages of repair, the nucleosomal landscape is restored. These dynamic rearrangements of nucleosomes occur during both the FA pathway and the glycosylase pathway of ICL repair. Furthermore, we found that blocking DNA replication or repair affects the nucleosomal rearrangements, highlighting the relationship between chromatin dynamics and ICL repair. Finally, we show that multiple chromatin remodeling factors are enriched at the ICL during repair.

Results

Nucleosome assembly on ICL-containing plasmids

Previous studies reported that nucleosomes are efficiently assembled on DNA during replication in *Xenopus* egg extracts[39,40]. To test if nucleosome assembly is inhibited by the presence of an interstrand crosslink we performed a supercoiling assay (Figure 1A). This assay is based on the fact that assembly of nucleosomes on a relaxed plasmid introduces negative superhelical turns in the DNA, resulting in supercoiled DNA (Figure 1A). We used a plasmid containing a site-specific cisplatin ICL (pICL-Pt) and an undamaged control plasmid (pControl) with the same sequence. The plasmid DNA was relaxed by Topoisomerase I treatment and incubated with High-Speed Supernatant (HSS), to promote the loading of the pre-replication complexes onto DNA (Figure S2). After 20 minutes, the nucleoplasmic extract (NPE) was added to promote the start (t=0) of one complete round of replication (for more detailed information see the introduction of this thesis). Replication intermediates were collected at various times and separated by agarose gel electrophoresis. Topoisomerase I treatment induced relaxation of the plasmids, although this was not fully completed as some faster migrating species were still observed (Figure 1B, lane 2). HSS incubation induced inefficient assembly of nucleosomes onto both pICL-Pt and pControl, as observed by the appearance of faster migrating intermediates at the start of replication (Figure 1B, lanes 3 and 9). After addition of NPE, supercoiled species further accumulated in both plasmids leading to a fully supercoiled band in less than 60 minutes (Figure 1B, lanes 4 to 6 and 10 to 12), indicating efficient replication-coupled nucleosome assembly.

To confirm that the replication-coupled supercoiling was caused by nucleosome assembly we performed an MNase digestion assay. MNase digests DNA that is not protected by binding to histones within a nucleosome, resulting in approximately 150 bp DNA fragments that represent the nucleosomal footprints. Replication intermediates of pICL-Pt and pControl were digested with MNase and the products were separated by agarose gel electrophoresis (Figure 1C). After HSS incubation (Figure 1C, lanes 1 and 6) we observed minor accumulation of 150 bp fragment DNA, indicating inefficient replication-independent nucleosome assembly on both plasmids. During replication, the 150bp DNA fragment accumulated in both pControl and pICL-Pt conditions (Figure 1C, lanes 2 to 4 and 7 to 11), indicating efficient replication-coupled nucleosome assembly. These experiments show that *Xenopus* egg extracts promote efficient assembly of nucleosomes on both undamaged and crosslinked plasmids during DNA replication.
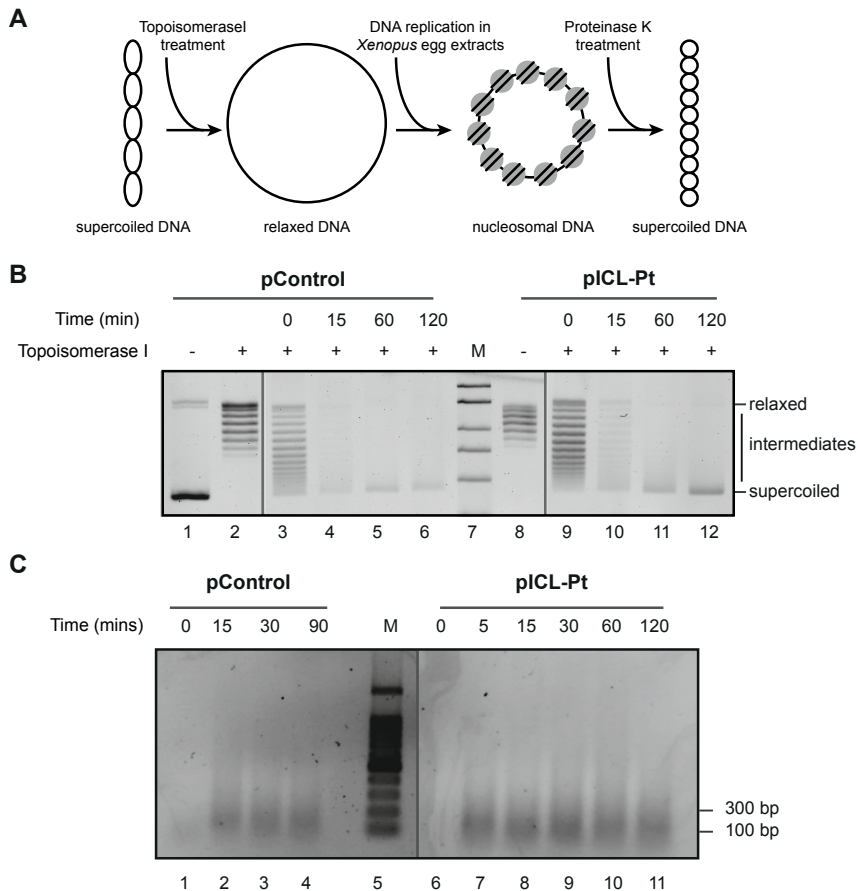
Figure 1: *Xenopus* egg extracts efficiently assemble nucleosomes on crosslinked DNA during DNA replication.
A) Schematic representation of the supercoiling assay to measure nucleosome assembly. Briefly, plasmid DNA is relaxed by Topoisomerase I treatment and incubated in extracts. Assembly of nucleosomes on DNA induces supercoiling of the plasmid, which can be visualized by a shift in gel mobility. **B)** pControl or pICL-Pt were replicated in *Xenopus* egg extracts and samples were taken at various time points. Supercoiling of the plasmid was visualized as a change in mobility upon gel electrophoresis ("M" indicates the DNA marker). All lanes are from the same gel but irrelevant lanes were removed indicated by a vertical line. **C)** Replication samples from B were digested with MNase (see "Methods") and analyzed by gel electrophoresis ("M" indicates the DNA marker). All lanes are from the same gel but irrelevant lanes were removed indicated by a vertical line.

Nucleosomes are removed from the ICL region and subsequently placed back during ICL repair

Since nucleosomal dynamics is an important component in other DNA repair pathways, we aimed to determine whether and how nucleosomes are repositioned during ICL repair. To this end, we developed a nucleosome positioning assay
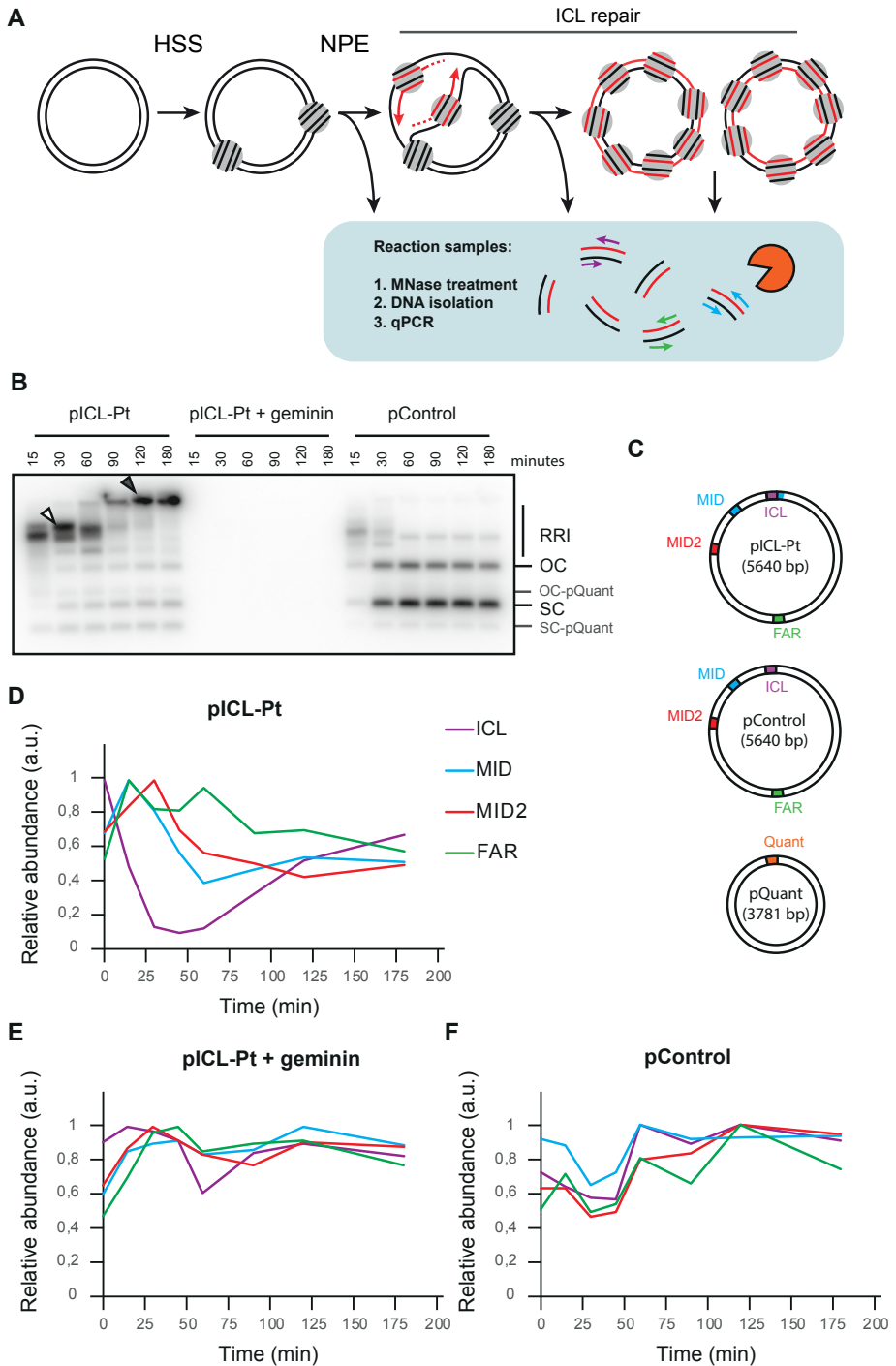
based on MNase digestion and quantitative PCR (Figure 2A and Methods). To map nucleosomes during ICL repair we replicated pICL-Pt in *Xenopus* egg extracts and first analyzed the replication intermediates by agarose gel electrophoresis (Figure 2B and Methods). This resulted in the specific pattern of FA pathway DNA repair intermediates (Replication/repair intermediates, or RRI) starting with the accumulation of "Figure 8" structures (Figure 2B, white arrow, and Figure S2B), followed by the formation of low-mobility DNA structures including homologous recombination intermediates (Figure 2B, grey arrow), and finally the slow accumulation of open circular and supercoiled repaired products (Figure 2B, "OC" and "SC" respectively). Addition of the DNA replication inhibitor geminin completely blocked DNA replication, as shown by the lack of nucleotide incorporation (Figure 2B). The undamaged plasmid pControl was quickly replicated in extracts, resulting in the accumulation of nicked and supercoiled structures within 30 minutes (Figure 2B). To determine the position of nucleosomes during ICL repair, DNA replication/repair intermediates were digested with MNase to obtain the nucleosome protected fraction. The resulting DNA fragments were isolated and mapped onto the plasmid sequence using qPCR with four different primer sets (Figure 2C and Methods): primer set "ICL" (24-132 bp from the crosslink), "MID" (662-775 bp from the crosslink), "MID2" (1500-1607 bp from the crosslink), and "FAR" (3117-3017 bp from the crosslink). The relative abundance of each amplicon was normalized against the internal control pQuant and was plotted against time (see Methods section). At early times during ICL repair the region proximal to the ICL became highly susceptible to MNase treatment (Figure 2D, ICL region and Figure S3). This could be due to replication fork stalling at the ICL, thereby preventing nucleosome assembly at this location. About 20 minutes after the start of replication, when all replication forks reached the ICL and repair started, the DNA from regions further away from the ICL became more susceptible to MNase treatment (Figure 2D, MID and MID2 regions and Figure S3), indicating progressive removal of nucleosomes from these regions during repair. In the FAR region instead, MNase sensitivity remained relatively constant (Figure 2D, FAR and Figure S3), suggesting that nucleosomes were not efficiently depleted from this region. At the time of ICL unhooking, which is quickly followed by TLS, (starting from 60 minutes based on the reduction in "Figure 8" structures, Figure 2B), the ICL, MID and MID2 amplicon abundance increased, reflecting the placement of nucleosomes in these regions. This data shows that during ICL repair nucleosomes are first removed (either by disassembly or sliding) from the site of damage and are later placed back (either by reassembly or sliding), concomitantly with the later stages of repair.

To further confirm this model, pControl and pICL-Pt replication/repair intermediates were digested with MNase, histone H2B was immunoprecipitated, and the H2B-bound DNA fraction was analyzed by qPCR. Immunoprecipitation of H2B did not alter the pattern of amplicon abundance (Figure S4), strongly indicating that the MNase protected fraction represents the nucleosomal footprint on these

plasmids. Furthermore, when pICL-Pt replication was inhibited by addition of geminin (Figure 2B), we did not observe the sequential removal/reassembly of nucleosomes in the ICL, MID and MID2 regions (Figure 2E and Figure S3) indicating that the re-organization of nucleosomes on pICL-Pt depends on DNA replication. Finally, digestion of pControl replication intermediates with MNase did not result in similar variation in MNase sensitivity (Figure 2F and Figure S3), indicating that the nucleosome dynamics observed for pICL-Pt are linked to the repair of the crosslink. Overall, we developed an assay that provides a readout for nucleosomal rearrangements during ICL repair. Using this nucleosome positioning assay, we found that nucleosomes are sequentially removed and placed back in a region surrounding the ICL, providing evidence for nucleosome dynamics during ICL repair.

Figure 2: Nucleosome dynamics at, and around, the ICL during replication-dependent ICL repair.
A) Schematic representation of the nucleosome positioning assay. Briefly, replication/repair samples are treated with MNase and the nucleosome-bound fraction is extracted. The DNA fragments are then analyzed by qPCR with plasmid-specific primers. **B)** Plasmids were replicated in *Xenopus* egg extracts by sequential incubation in HSS and NPE, with or without geminin. Radio-labelled nucleotides (^{32}P-α-dCTP) were added to the replication reactions. Samples were taken at various time points after the start of replication, reaction products were separated by gel electrophoresis, and visualized by autoradiography. Replication/repair intermediates (RRI), open circular (OC), and supercoiled (SC) products are indicated. The white arrow indicates "Figure 8" structures. An unrelated plasmid pQuant was replicated in the same reaction for normalization purposes (OC-pQuant and SC-pQuant) (see methods section). **C)** Schematic representation of the primers used in the nucleosome positioning assay (see methods section for primer sequences). **D)** Replication reactions from A) to which no ^{32}P-α-dCTP was added were digested with MNase and the DNA fragments were isolated and mapped to the plasmid by qPCR using the primer sets in C). For each primer set DNA abundance was normalized against pQuant, the peak value was set to 1, and the values were plotted against reaction time. **E)** as in D), in the presence of geminin **F)** as in D) using pControl. Independent experimental replicates in Figure S3.

A

HSS   NPE   ICL repair

Reaction samples:

1. MNase treatment
2. DNA isolation
3. qPCR

3

B

pICL-Pt   pICL-Pt + geminin   pControl

15 30 60 90 120 180   15 30 60 90 120 180   15 30 60 90 120 180   minutes

RRI
OC
OC-pQuant
SC
SC-pQuant

C

MID
MID2   ICL
pICL-Pt
(5640 bp)
FAR

MID
MID2   ICL
pControl
(5640 bp)
FAR

Quant
pQuant
(3781 bp)

D

pICL-Pt

Relative abundance (a.u.)

—ICL
—MID
—MID2
—FAR

Time (min)

E

pICL-Pt + geminin

Relative abundance (a.u.)

Time (min)

F

pControl

Relative abundance (a.u.)

Time (min)

CMG unloading is required for nucleosome reorganization during ICL repair

To understand the function of this nucleosomal rearrangement, we set out to determine how this process is coupled to specific steps in ICL repair. We inhibited one of the earliest events of the FA pathway, the unloading of the CMG helicase (Figure S1)[29,31,37], which is required for replication fork reversal and ICL unhooking. This step depends on the ubiquitination of the CMG helicase and its subsequent unloading from DNA by the p97 segregase[28]. It can be inhibited by addition of NMS-873, a potent p97 inhibitor (p97i). Consistent with published studies, addition of p97i resulted in persistence of "Figure 8" DNA structures on an agarose gel, that indicates accumulation of repair intermediates and inhibition of repair[28,37] (Figure 3A, white arrow). When mapping nucleosomes in this reaction using the nucleosome positioning assay, we observed a sequential decrease in the nucleosome protected fraction at the ICL and MID regions, similar to the control reaction without p97i (Figure 3B and Figure S5). This indicates that nucleosomes are removed from the regions surrounding the ICL even in the presence of the p97 inhibitor. Nucleosomes were also removed from the MID2 region in the presence of p97i, albeit with slower kinetics compared to the control (Figure 3B and C, and Figure S5). This suggests that CMG unloading is not required for the removal of nucleosomes around the crosslink. However, while the MNase resistant DNA fragments increased at later stages of repair in the control condition (Figure 3B), this was not observed in the presence of p97i (Figure 3C and Figure S5), indicating that nucleosomes could not be placed back when unloading of the CMG helicase was inhibited.

To further confirm these findings, we blocked CMG unloading by disrupting CMG ubiquitination[31] (Figure S1) using ubiquitin vinyl-sulfone (UbVS), an irreversible inhibitor of deubiquitinating enzymes that depletes the free ubiquitin pool[41]. In accordance with the p97 inhibition results, blocking CMG unloading by UbVS treatment resulted in persistent "Figure 8" DNA structures (Figure 3A, black arrow). In the presence of UbVS we observed a sequential increase in MNase sensitivity of the ICL, MID and MID2 regions at early times (Figure 3D), as observed with the p97i (Figure 3C). However, at later stages of repair, the MNase protected DNA fraction did not increase as in the control reaction (Figure 3B), indicating that nucleosomes were not restored around the crosslink. These results suggest that nucleosomes are removed from the ICL, MID and MID2 regions prior to CMG unloading, but restoration of the nucleosome landscape requires continuation of ICL repair, providing evidence for a direct link between ICL repair and nucleosomal dynamics.
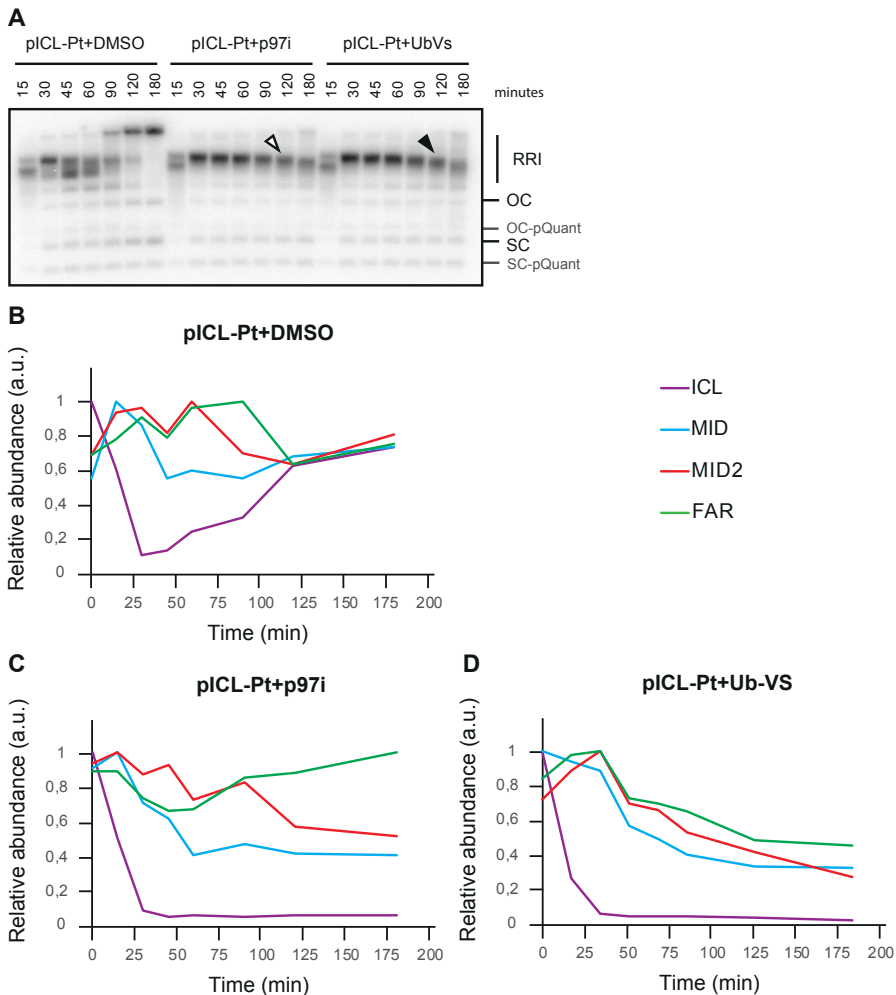
Figure 3: CMG unloading is necessary for restoration of nucleosomes in the FA pathway of ICL repair

A) Plasmids were replicated in *Xenopus* egg extracts in the presence of NMS 873 (p97i), ubiquitin vinyl-sulfone (UbVS), or DMSO. Reactions were split, one part was replicated in the presence of ^{32}P-α-dCTP and the repair products were analyzed by gel electrophoresis and autoradiography. Replication/repair intermediates (RRI), open circular (OC) and supercoiled (SC) products are indicated. An unrelated plasmid pQuant was replicated in the same reaction for normalization purposes. (OC-pQuant and SC-pQuant). The white and black arrows indicate persistent "Figure 8" structures **B-D)** The non-labelled samples generated in A) were digested with MNase and DNA was purified and mapped on pICL-Pt by qPCR. DNA abundance of each amplicon was normalized against pQuant. Experimental replicates in Figure S5A and B.
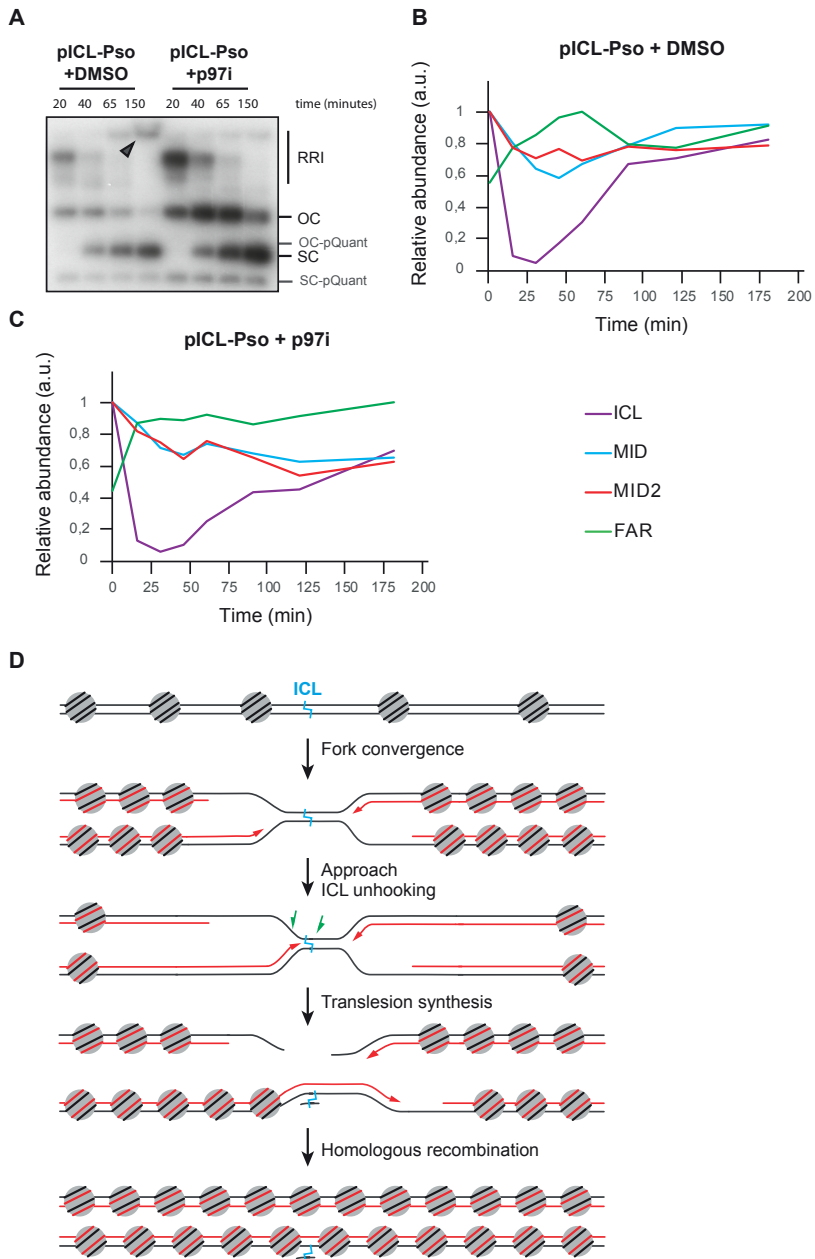
Repair of structurally diverse ICLs involves similar nucleosome dynamics

Structurally different ICLs can be repaired by different pathways[37,38]. Cisplatin and nitrogen mustards-induced ICLs are repaired solely by the FA pathway, while psoralen/UV- or abasic site-induced ICLs can be repaired by both the NEIL3-dependent glycosylase pathway and the FA pathway[37] (Figure S1). In the absence of NEIL3, repair shifts to the FA pathway and vice versa[37]. We wondered whether the chromatin dynamics observed during ICL-Pt repair were specific for the FA pathway or were a more common feature of ICL repair. To investigate this, we replicated a plasmid containing a Psoralen/UV-induced ICL (pICL-Pso) in *Xenopus* egg extract. Consistent with previous studies[37], we found that upon separation of pICL-Pso repair intermediates by agarose gel electrophoresis, the "Figure 8" structures were quickly converted into open circular molecules and supercoiled species (Figure 4A). The FA pathway-specific low-mobility DNA structures (Figure 4A, grey arrow) were relatively scarce, indicating that the majority of the crosslinks were repaired by the faster glycosylase pathway and only a small fraction by the FA pathway. To study exclusively nucleosome dynamics coupled to the glycosylase pathway, we supplemented extracts with p97i, which efficiently blocks the FA pathway without inhibiting the glycosylase route[37]. In this condition, the FA-specific low mobility DNA structures were strongly reduced, indicating that repair occurred through the NEIL3 pathway[37] (Figure 4A). We then made use of the nucleosome positioning assay to study nucleosome dynamics during pICL-Pso repair. In control conditions, where pICL-Pso is repaired by both the glycosylase and the FA pathway, we observed a sequential increase in MNase sensitivity in the ICL, MID and MID2 loci, similar to pICL-Pt (Figure 4B and Figure S5B). As repair progressed, the MNase-resistant DNA fraction increased again in these regions (Figure 4B). This shows that nucleosomes are first removed from the region surrounding the crosslink and later placed back. Upon p97 inhibition, we observed that MNase sensitivity increased in the ICL, MID, and MID2 regions at early stages of repair (Figure 4C), indicating that nucleosomal removal also occurs in the glycosylase pathway. At later stages of repair nucleosomes were reconstituted at the ICL region, albeit slowly compared to the control condition. Furthermore, the repositioning of nucleosomes at the MID regions was reduced compared to the control (Figure 4C). This could be due to an inhibitory effect of p97i on translesion synthesis which would result in incomplete repair (see also Discussion) and is consistent with the increase in open circular structures in the presence of p97i (Figure 4A). Alternatively, p97i may induce some DNA resection and replication fork collapse (see Discussion). In both of these scenarios, a portion of ICLs would not be repaired and therefore the chromatin landscape would not be restored. Nevertheless, since at the ICL site nucleosome restoration occurs still rather efficiently, it is likely that nucleosome reassembly during pICL-Pso repair does not rely on CMG unloading specifically, but rather on the successful completion of ICL repair. In summary, we conclude that the nucleosomal dynamics during both the FA pathway and the glycosylase pathway

are highly similar and consist of the sequential disassembly and reassembly of nucleosomes in the region surrounding the ICL (Figure 4D).
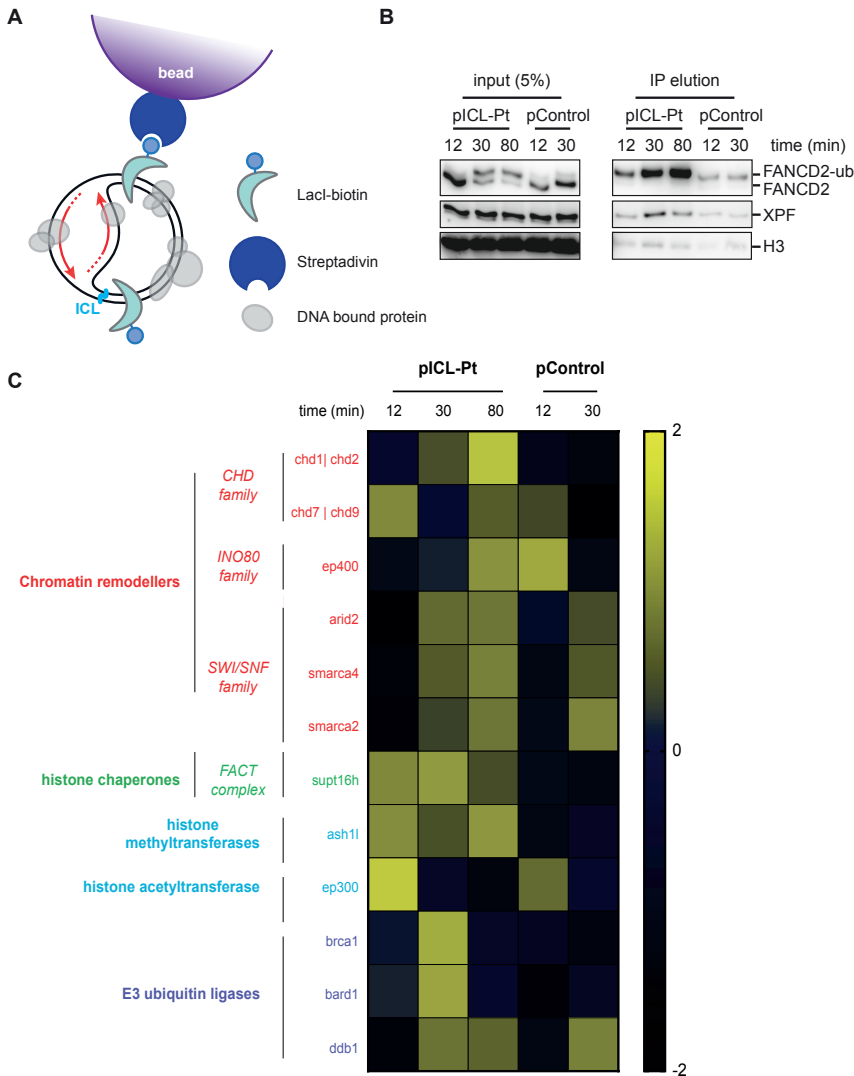
Identification chromatin remodeling factors recruited to ICLs during repair
After determining that ICL repair-coupled nucleosome remodeling takes place in our system, we aimed to identify what chromatin remodeling factors are involved in this process. We used a method in which ICL-containing plasmids are pulled down during repair and plasmid-bound proteins are identified by mass spectrometry. To this end, we replicated pICL-Pt and pControl in *Xenopus* egg extracts and isolated DNA and DNA-bound proteins at various times during repair using streptavidin beads coated with biotinylated LacI, that nonspecifically binds DNA (Figure 5A)[35]. DNA-bound proteins were first analyzed by western blot (Figure 5B). As expected, ubiquitinated FANCD2 and XPF, key proteins in the FA pathway, were enriched on pICL-Pt. In contrast, histone H3 was equally bound to pICL-Pt and pControl (Figure 5B). We next analyzed the DNA-bound proteins by mass spectrometry (see methods section). In addition to factors involved in DNA replication and ICL repair, we found several chromatin remodeling factors enriched on pICL-Pt during repair (Figure 5C). Among these proteins, we found members of ATP-dependent chromatin remodeling complexes (CHD, INO80 and SWI/SNF families), histone chaperones (FACT complex), methyltransferases (ASHL1l), acetyltransferases (ep300) and E3 ubiquitin ligases (BRCA1, BARD1, DDB1) (Figure 5C). Many of these factors have previously been linked to nucleosome dynamics during repair processes, such as DSB repair (CHD1[42], CHD7[43], p400[44,45], SMARCA2 and 4[46], SUPT16H[47], BRCA1/BARD1[48], p300[49]) and NER (ASHL1l[50], DDB1[51], SUPT16H[52]). Interestingly, we observed that histone chaperones, methyltransferases, acetyltransferases and ubiquitin ligases are mainly enriched at the earlier stages of repair (t=12 and 30 minutes) while recruitment of chromatin remodelers peaks later (t=80 minutes). It is therefore possible that the factors recruited at earlier time points may be involved in the initial repair-coupled nucleosome depletion, while the ones that bind the plasmid later may act in nucleosome reconstitution. To study the function of these factors in the various stages of ICL repair, current work in the laboratory is focusing on using specific inhibitors and raising antibodies against these factors.

A

pICL-Pso
+DMSO

pICL-Pso
+p97i

20 40 65 150   20 40 65 150   time (minutes)

RRI

OC
OC-pQuant
SC
SC-pQuant

B

pICL-Pso + DMSO

Relative abundance (a.u.)

Time (min)

C

pICL-Pso + p97i

Relative abundance (a.u.)

Time (min)

— ICL
— MID
— MID2
— FAR

D

ICL

Fork convergence

Approach
ICL unhooking

Translesion synthesis

Homologous recombination

Figure 4: Nucleosome dynamics are similar in different pathways of ICL repair.
A) pICL-Pso was replicated in *Xenopus* egg extracts in the presence or absence of NMS 873 (p97i). Reactions were split, one part was replicated in the presence of ^{32}P-α-dCTP and the repair products were analyzed by gel electrophoresis and autoradiography. Replication/repair intermediates (RRI), open circular (OC) and supercoiled (SC) products are indicated. An unrelated plasmid pQuant (OC-pQuant, SC-pQuant) was replicated in the same reaction for normalization purposes. Grey arrow indicates FA-specific low mobility DNA structures. **B-C)** The non-labelled samples generated in A) were digested with MNase and the DNA fragments were analyzed by qPCR. For each primer set DNA abundance was normalized against pQuant. Samples from Figure S5A were obtained in the same experiment. Experimental replicate in Figure S5B. **D)** Model for nucleosome dynamics during replication-dependent ICL repair by the FA pathway. Upon fork convergence at the ICL, nucleosomes are removed from the lesion site, possibly because of fork stalling. As repair is initiated, nucleosomes are progressively removed from a region approximately 1.5kb away from the crosslink. In the later stages of repair, nucleosomes are reassembled both at the ICL site and the region surrounding it. Similar nucleosomal dynamics are likely occurring during ICL repair through the glycosylase pathway.

3

Figure 5: A mass spectrometry approach to identify chromatin remodeling factors during ICL repair.
A) Schematic representation of the LacI plasmid pulldown. **B)** Indicated plasmids were replicated in *Xenopus* egg extracts and DNA was isolated by streptavidin beads conjugated with LacI-biotin (see methods section). The proteins bound to the plasmid were analyzed by western blotting with the indicated antibodies. **C)** DNA bound proteins were analyzed by mass spectrometry. The heatmap shows the mean of the Z scored $\log_2$ label-free quantitation LFQ intensity from three biochemical replicates.

Discussion

The reorganization of the chromatin environment is thought to be necessary during DNA repair to allow the access of repair factors to DNA and to promote DNA damage signaling[13]. In this study, we used *Xenopus* egg extracts to investigate nucleosome dynamics during ICL repair. We report that, in the earliest stages of ICL repair, nucleosomes are displaced from the ICL region, likely due to the replication forks stalling at the level of the DNA lesion. Nucleosomes are then progressively displaced from the region surrounding the lesion, up to approximately 1.5kb away from it. After ICL unhooking, nucleosomes are restored in a process that is coupled to repair (Figure 4D). Using a mass spectrometry approach, we found several factors involved in nucleosomal dynamics enriched during ICL repair, such as: ATP-dependent chromatin remodeling complexes, histone chaperones, methyltransferases, acetyltransferases and E3 ubiquitin ligases (Figure 5C).

   The observation that nucleosomes are sequentially displaced and re-placed at the lesion site is consistent with previous observations on chromatin dynamics during DSB repair and NER and, more generally, with the ARR model[13]. This model involves transient nucleosomal rearrangements that favor DNA accessibility and repair, however, the extent of this increased accessibility is still unclear. In this study we observed nucleosome removal from a region 1.5 kb from the ICL both for cisplatin and psoralen crosslinks (Figure 2D and Figure 4B), indicating that similar nucleosomal rearrangements are induced by crosslinks of different chemical nature and using different repair mechanisms.

   During ICL repair nucleosomes are first displaced from the ICL site and later from the surrounding DNA (Figure 2D). We speculate that the initial clearance of nucleosomes from the lesion site could be necessary to create of a point of access to DNA for chromatin remodeling factors. Once these chromatin remodelers can access DNA, they could induce the progressive displacement of nucleosomes from the DNA regions surrounding the ICL, creating a large nucleosome-free DNA region. This nucleosome-free DNA could then promote both recruitment of DNA repair factors and allow DNA rearrangements necessary for ICL repair. One of the factors that could require nucleosome-free DNA to act in ICL repair is FANCD2. Previous observations from our lab[34] indicate that FANCD2 first binds DNA at the ICL site and that later multiple FANCD2 molecules are loaded that spread to the area surrounding the ICL. This is in line with recent findings that propose that the ID2 complex is a sliding clamp[53]. Since nucleosomes would constitute an obstacle for the sliding of the ID2 complex, it is possible that nucleosome clearance from the region surrounding the lesion may be necessary for ID2 sliding. However, it is still unknown whether ID2 sliding is functionally relevant for ICL repair. Clearance of nucleosomes could also be required for DNA rearrangements associated to repair, such as fork reversal[32]. This process is suggested to be involved in the promotion

of ICL unhooking[32], however mechanistic details regarding the role of fork reversal in ICL repair are still lacking.

Blocking CMG unloading did not affect removal of nucleosomes, but it prevented nucleosomes from being placed back in the FA pathway, suggesting that completion of repair is necessary for the correct restoration of chromatin (Figure 3C). Given the timing of nucleosome repositioning, we speculate that this process could be coupled to TLS (Figure 2D). This is unexpected since lesion bypass of the strand containing the unhooked adduct by TLS precedes the process of homologous recombination (HR) that restores the incised strand (see Figure S1) and HR is thought to require nucleosome-free DNA[47,54]. To test whether nucleosomal restoration starts before the completion of HR, and may be coupled to TLS, experiments should investigate nucleosome dynamics in extracts depleted of TLS polymerases or HR factors.

We demonstrated that nucleosome remodeling during ICL repair is a replication-dependent process (Figure 2E), indicating that this process may be linked to histone chaperones associated to the replication fork, such as CAF-1 (Chromatin Assembly Factor-1) or FACT (Facilitates Chromatin Transcription). Interestingly, the SUPT16H subunit of the FACT complex was found enriched on pICL-Pt in our mass spec dataset (Figure 5C), indicating a potential role for this complex during repair. FACT is a component of the replisome and it is involved in the deposition of histones H3-H4 during replication[55,56]. Previous studies identified FACT as the factor that promotes deposition of the histone variant H2A.X during repair of UV lesions[52], and it has been associated with nucleosomal dynamics during multiple DNA repair pathways[47,57]. It would be therefore important to determine if this complex has a role during ICL repair and damage signaling.

Our data on nucleosomal dynamics during repair of a psoralen/UV-induced ICL indicates that inhibiting the FA pathway by blocking CMG unloading did not completely restore nucleosomes around the ICL (Figure 4C). This observation is seemingly contradictory, since blocking the FA pathway shifts repair to the glycosylase pathway, which does not require CMG unloading[37]. However, we inhibited CMG unloading by using an inhibitor of the p97 segregase, which could affect additional steps in ICL repair. Treatment with p97i during pICL-Pso repair induced accumulation of open circular structures (Figure 4A), possibly indicating a defect in translesion synthesis, which is one of the final steps in the glycosylase pathway[37] (Figure S1). Delay or partial impairment of translesion synthesis could therefore explain the defect in nucleosome restoration. An alternative explanation could be inefficient pathway switching during pICL-Pso repair upon p97i treatment. While we (Figure 4A) and others[37] have shown that a portion of psoralen/UV-induced-induced ICLs is repaired by the FA pathway, it is not completely clear how efficient the switch to the glycosylase pathway is upon inhibition of the FA pathway. If this is not efficient, ICLs will not be repaired and may induce fork resection or collapse, ultimately affecting nucleosomal restoration.

In summary, we present the first direct evidence for nucleosome dynamics during replication-dependent ICL repair providing multiple advantages for studying the molecular mechanism of this process. Firstly, the use of plasmids containing chemically defined ICLs allows the specific and direct investigation of this process. Secondly, using site-specific ICLs enables us to study the location of nucleosomes with respect to the ICL at mono-nucleosomal resolution during each stage of repair. This is still challenging in cells as there is currently no method to induce a site-specific ICL in the genome. Thirdly, since ICL repair is a synchronous process in *Xenopus* egg extracts, we can link nucleosomal dynamics to specific stages of repair. In the future we aim to identify both the factors involved in this process and their mechanism of action. This will not only provide important new insights in the mechanisms of ICL repair, that can provide opportunities for the improvement of chemotherapy, but may also give important information on this process in other DNA repair pathways.

Materials and Methods

Preparation of plasmids

Plasmids containing site-specific DNA interstrand crosslinks were prepared as previously described[27,58]. In brief, DNA duplexes containing cisplatin or psoralen-induced ICLs were ligated in the pSVRLuc vector linearized with Bbs1. The ligated plasmids were purified using a cesium chloride gradient. For pICL-Pso, the pSVRLuc backbone was first modified to substitute an adenine to guanine on one of the Bbs1 5' overhangs.

Xenopus egg extracts, DNA replication and repair

Xenopus egg extracts were prepared as described previously[59]. To replicate plasmid DNA, plasmids were incubated in HSS at a final concentration of 7.5 ng µl^{-1} for 20 minutes at room temperature to allow formation of pre-replication complexes. The HSS-DNA mix was supplemented with two volumes of NPE to start one round of DNA replication. To visualize the nascent DNA strand, replication was carried out in the presence of ^{32}P-α-dCTP. An unrelated non-damaged control plasmid (pQuant) was added as internal control for quantifications at a final concentration of 0.25 ng µl^{-1}. To inhibit DNA replication HSS was supplemented with recombinant geminin at 300 nM. To block the unloading of the CMG helicase, NPE was supplemented with up to 40 µM p97 inhibitor (NMS-873, Sigma). Ubiquitin signaling was impaired adding ubiquitin vinyl-sulfone (UbVS, Boston Biochem) to NPE at a final concentration of 26 µM. To analyze replication and repair products, replication reactions were stopped by adding five volumes of replication stop solution I (Stop I: 80 mM Tris pH 8, 5% SDS, 0.13% phosphoric acid, 10% Ficoll, 8 mM EDTA, 0.1% bromophenol blue). Following treatment with Proteinase K (1.5 µg µl^{-1} 40 min at

37°C), samples were loaded on a 0.8% native agarose gel. The gel was dried and visualized by autoradiography.

Supercoiling and MNase assay

For the supercoiling assay, plasmid DNA was relaxed using Topoisomerase I (NEB, 10 U Topoisomerase I per μg DNA). DNA was subsequently replicated in *Xenopus* egg extracts as described above at final plasmid concentration of 3.33 ng μl^{-1}. Samples (6 μl replication mix) were taken at various time points and, following RNase A and Proteinase K treatment, DNA was isolated by phenol/chloroform extraction. DNA was analyzed by gel electrophoresis and visualized after SybrGold staining.

For analysis of chromatin assembly by MNase digestion, plasmid DNA was replicated in *Xenopus* egg extracts as described above. At various times samples (12 μl reaction volume, 60 ng DNA) were diluted with 488 μl 1X MNase buffer (50 mM Tris-HCl, 5 mM $CaCl_2$, pH 7.9 at 25°C, NEB) supplemented with 0.1 mg μl^{-1} BSA and digested with 300 U MNase for 10 minutes at 37°C. Digestion was stopped using 120 μl C-buffer (160 mM EDTA, 6.87% SDS). Samples were deproteinated using Proteinase K, DNA was isolated by phenol/chloroform extraction, and samples were analyzed by gel electrophoresis.

Nucleosome positioning assay

To map nucleosomes during ICL repair, 3 μl aliquots of a replication reaction were diluted 10X in MNase buffer (NEB, 10X MNase buffer) supplemented with BSA at 0.1 mg μl^{-1}. Samples were then digested with 600 U MNase for 5 minutes at 37°C. Digestions were stopped by adding 14 volumes of MNase stop buffer (20 mM Tris-HCl pH 7.5, 0.15 M NaCl, 20 mM EDTA, 20 mM EGTA, 0.5% NP-40, 5 μg/mL aprotinin, 5 μg/mL leupeptin, and 2 mM PMSF). Samples of the digested DNA (80 μl) were then treated with RNase A and Proteinase K and purified by phenol/chloroform extraction. DNA was then digested with Bbs1 and analyzed by quantitative PCR using the following primers:
ICL forward: 5'-AGCCAGATTTTTCCTCCTC-3';
ICL reverse: 5'-CATGCATTGGTTCTGCACTT-3';
MID forward: 5'-ACCCTGGGTTCTTTTCCAAC-3';
MID reverse: 5'- CATTTCATCTGGAGCGTCCT-3';
MID2 forward: 5'-GCAACGTGCTGGTTATTGTG-3';
MID2 reverse: 5'-GTGATCATGCGTTTGCGTTG-3';
FAR forward: 5'-AACGCCAATAGGGACTTTCC-3';
FAR reverse: 5'-GGGCGTACTTGGCATATGAT-3';
pQuant forward: 5'-TACAAATGTACGGCCAGCAA-3';
pQuant reverse: 5'-GAGTATGAGGGAAGCGGTGA-3'.
The abundance of ach amplicon was normalized against pQuant amplicon abundance, and for each amplicon the peak value was set to 1.

H2B Immunoprecipitation

Repair samples (3 µl aliquots) were diluted 10X in MNase buffer (NEB, 10X MNase buffer) supplemented with BSA at 0.1 mg µl $^{-1}$. Samples were then digested with 40U MNase for 40 minutes. The reaction was stopped using 14 volumes MNase Stop buffer (see above). To immunoprecipitate H2B, MNase-digested samples were incubated for 1 hour at 4°C with 0.3 volumes (compared to sample volume) of Protein A Sepharose Fast Flow beads (PAS beads, GE Healthcare) pre-washed in ChIP wash buffer (50 mM Tris-HCl pH 7.5, 10 mM EDTA, 0.15 M NaCl). After incubation, beads were spinned down and the supernatant was collected and incubated with 2.4 µg of ChIP-grade α-H2B antibody (Abcam, ab1791) on a rotating wheel overnight at 4°C. The next day 0.3 volumes of pre-washed PAS beads were incubated with the samples for 2 hours at room temperature on a rotating wheel. Beads were washed three times with ChIP wash buffer and DNA was eluted from the beads adding 10 volumes of ChIP elution buffer (50 mM Tris-HCl pH 7.5, 10 mM EDTA, 1% SDS) while shaking in a Thermomixer for 20 minutes at 65°C, at 1200 RPM. The eluted DNA was treated with RNase A and Proteinase K and purified by phenol/chloroform extraction. Samples were then analyzed by quantitative PCR as described above.

LacI plasmid pulldown for mass spectrometry analysis

Magnetic beads coated with streptavidin (Dynabeads M-280, Invitrogen) were washed three times in wash buffer 1 (50 mM Tris pH 7.5, 150 mM NaCl, 1mM EDTA pH 8.0, 0.02% Tween-20). Beads were resuspended in one volume of wash buffer 1 and incubated with biotinylated LacI (2 µM final concentration) for 45 minutes at room temperature, mixing every 10 minutes. Beads were then washed four times with one volume of IP buffer (10 mM Hepes pH 7.7, 50 mM KCl, 2.5 mM MgCl$_2$, 250 mM sucrose, 0.25 mg ml^{-1} BSA, 0.02% Tween-20), resuspended in 6.64 volume of IP buffer, and stored at 4°C. Samples from the replication reaction (150 µl) were mixed with 750 µl beads solution (containing 113 µL biotin-LacI-bound Streptavidin magnetic beads), and the mix was incubated for 30 minutes on water-ice slurry mixing every 10 minutes. Beads were washed three times with IP buffer containing 0.03% Tween-20, and resuspended in 40 µl 1x SDS sample buffer (75 mM Tris-HCl pH6.8, 10% glycerol, 2.5% SDS, 10% Bond-Breaker TCEP Solution (Thermo Fisher), and 0.02% bromophenol blue). Samples were boiled at 95°C for 5 minutes, and analyzed by gel electrophoresis and western blotting.

Western blotting

For western blot analysis of plasmid pull down experiments, antibodies against *xl*FANCD2 (previously described in[60]), Histones H3 (Abcam), and *xl*XPF (previously described in[34]) were used.

Mass spectrometry analysis of plasmid-bound factors

The DNA-bound factors isolated by plasmid pulldown were separated on a 12% Bis-Tris SDS-PAGE gel (Biorad). The gel was run approximately 2-3 cm and it was stained with colloidal coomassie dye G-250 (Gel Code Blue Stain Reagent, Thermo Scientific) and each lane was cut in three pieces. Gel pieces were reduced, alkylated and trypsinized overnight at 37°C. Peptide extraction was done with 100% acetonitrile (ACN) and samples were dried in a vacuum concentrator. Samples were resuspended in 10% (v/v) formic acid for UHPLC-MS/MS. Data acquisition was done using a UHPLC 1290 system coupled to an Orbitrap Q Exactive Biopharma HF mass spectrometer (Thermo Scientific). Samples were first trapped (Dr Maisch Reprosil C18, 3 μm, 2 cm x 100 μm) for 5 minutes at 5 μl min^{-1} in solvent A (0.1 M acetic acid in water). Samples were then separated on an analytical column (Agilent Poroshell EC-C18, 278 μm, 40 cm x 75 μm), at a column flow of 300 nl min^{-1} with a gradient as follows: 13-44% solvent B (0.1 M formic acid in 80% acetonitrile) in 95 minutes, 44-100% in 3 min, 100% solvent B for 1 min and 100-0% in 1 min. The acquisition of full scan MS spectra was performed at m/z 375 – 1600, at a resolution of 60,000 at m/z 400 after accumulation to a target value of 3e6. Up to ten most intense precursor ions were selected for HCD fragmentation, which was done at normalised collision energy of 27% after the accumulation to the target value of 1e4. MS/MS acquisition was at 30.000 resolution. Raw data files were analyzed using the MaxQuant software (version 1.5.0.17) with label-free quantification[61]. False discovery rate (FDR) of 0.01 and minimum peptide length of 7 amino acids were used. A non- redundant *Xenopus* database[62] was used as search engine for the data. Cysteine carbamidomethylation was selected as a fixed modification, while protein N-terminal acetylation and methionine oxidation were selected as variable modifications. For the Andromeda search, trypsin was chosen as enzyme allowing for cleavage N-terminal to proline. A maximum of two missed cleavages was allowed. Mass deviation for fragment ions was 0.5 Dalton and the initial mass deviation of precursor ion was up to 7 ppm. All bioinformatics analysis was carried out with the Perseus software Version 1.6.10.0. For each comparison, the processed data was filtered to contain at least two valid values in at least one of the replicate group (three repeats per condition).

Supplementary Figures
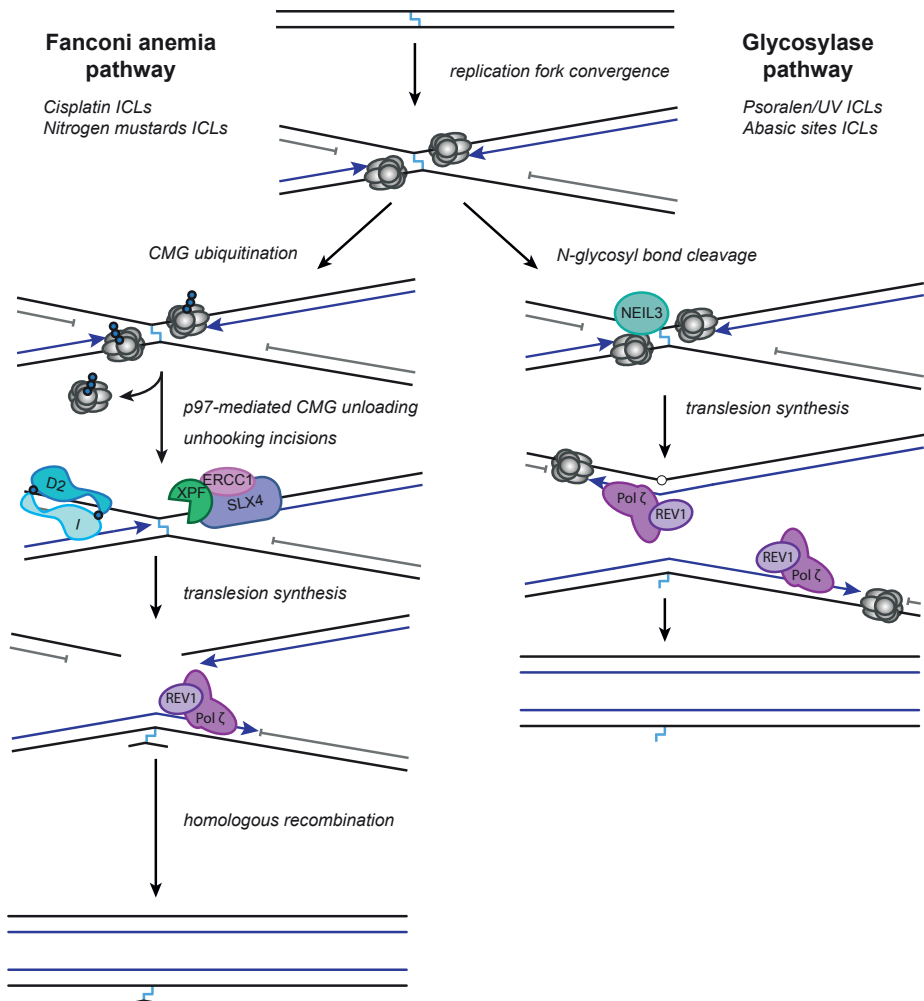
Figure S1: Mechanism of ICL repair by the FA pathway and the glycosylase pathway.
Schematic representation of the FA and the glycosylase pathway. Briefly, upon double fork convergence at the level of the ICL, in the FA pathway the CMG is unloaded by the p97 segregase. This is followed by ubiquitination of FANCI-FANCD2 and recruitment of the nuclease complex XPF-ERCC1-SLX4, that unhooks the ICL. The unhooked adduct is bypassed by translesion synthesis polymerases, generating a template for homology-mediated repair of the double strand break. In the glycosylase pathway, the glycosylase NEIL3 unhooks the ICL, generating an abasic site on one of the two strands (white circle). Both the unhooked adduct and the abasic site are bypasses by translesion synthesis.
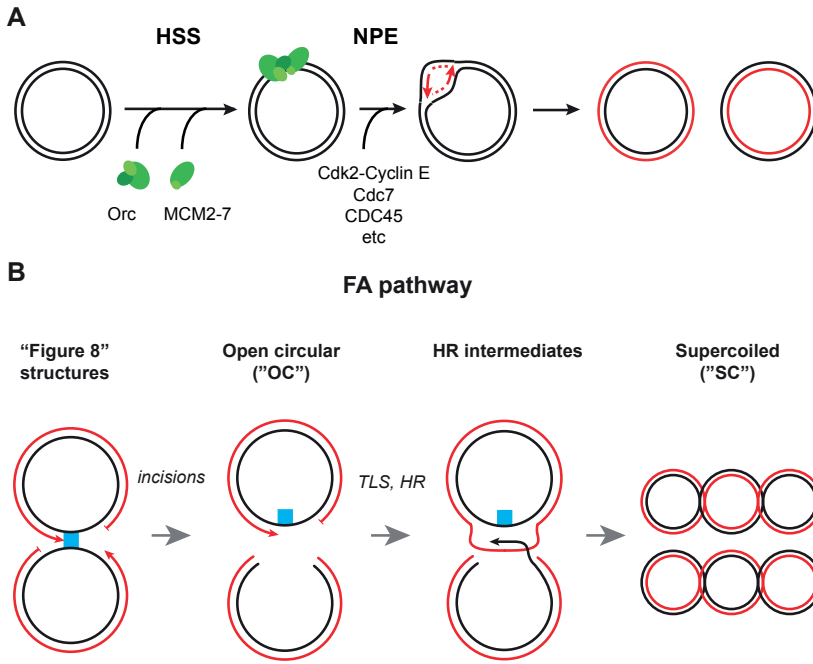
Figure S2: DNA replication and ICL repair in *Xenopus* egg extracts.
A) Schematic representation of DNA replication using the NPE/HSS system. For more information, see the introduction of this thesis. **B)** Schematic representation of the DNA structures formed in the FA pathway; 'TLS', translesion synthesis; 'HR', homologous recombination.
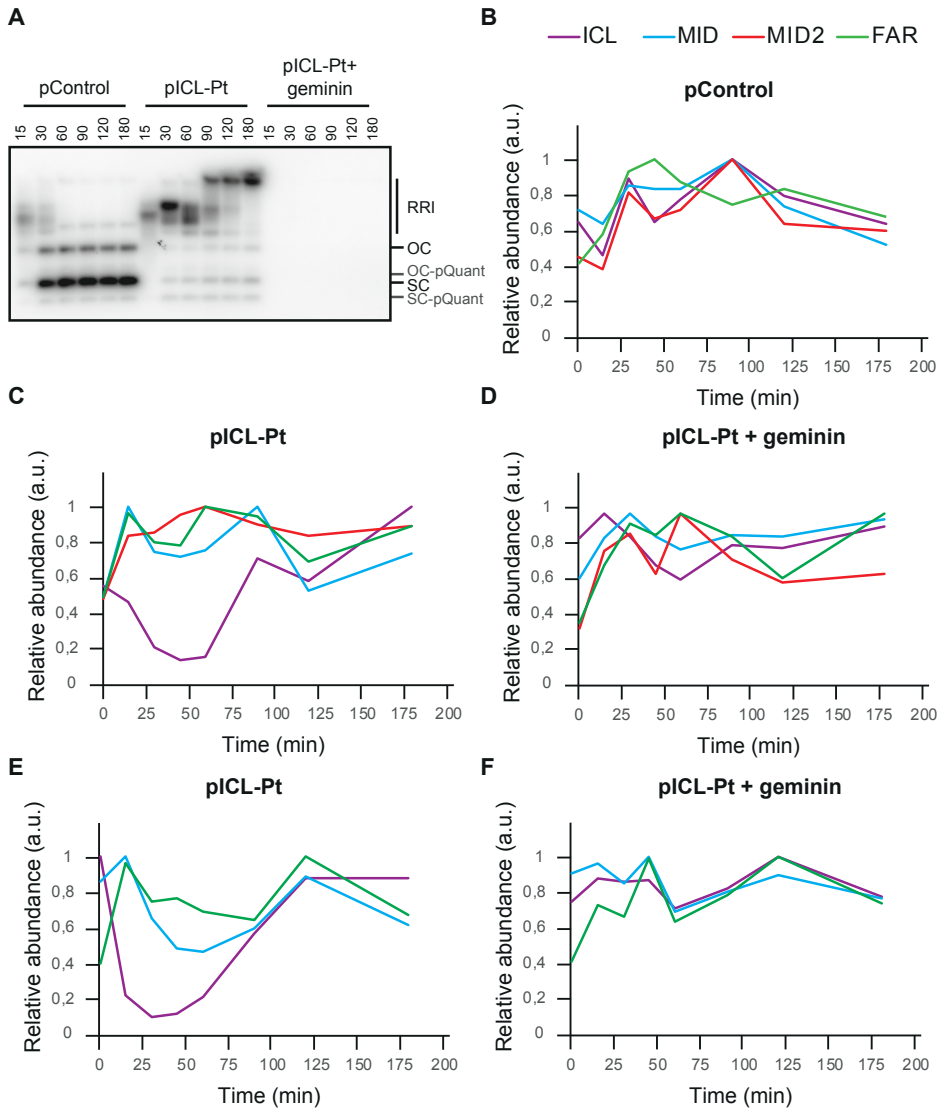
Figure S3: Nucleosome are remodeled during replication-dependent ICL repair.
A) Plasmids were replicated in *Xenopus* egg extracts. Part of the reaction was used to monitor DNA replication by incubation with ^{32}P-α-dCTP. Radioactive reaction samples were resolved by gel electrophoresis and visualized by autoradiography. Geminin was added where indicated to inhibit DNA replication. Replication/repair intermediates (RRI), open circular (OC) and supercoiled (SC) products are indicated. The plasmid pQuant (OC-pQuant and SC-pQuant) was replicated in the same reaction for normalization purposes. **B-D)** The non-radioactive replication/repair samples were digested with MNase and DNA fragments were purified and mapped to the plasmid by qPCR. For each primer set amplicon abundance was normalized against pQuant. Independent experimental replicate of Figure 2C-E. **E)** Independent experimental replicate of C). **F)** Independent experimental replicate of D).
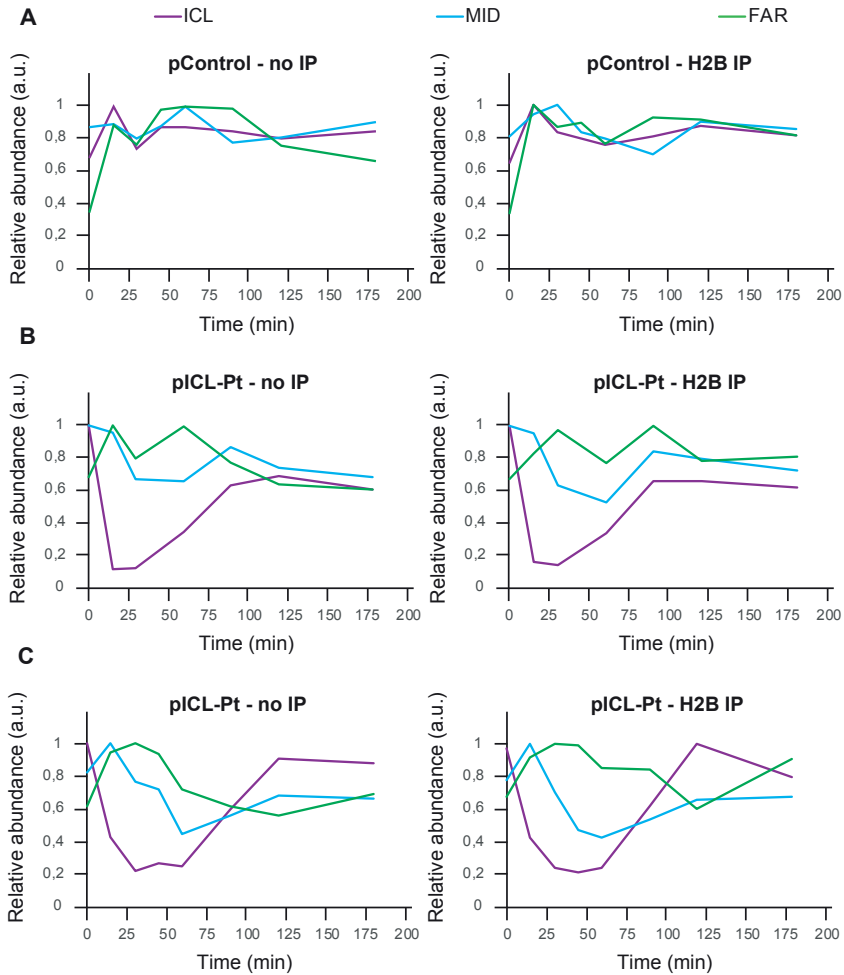
Figure S4: Immunoprecipitation of histone H2B confirms nucleosome dynamics during ICL repair.
A) pControl was replicated in *Xenopus* egg extracts and replication/repair samples were taken at various time points. Samples were treated with MNase and where indicated the DNA fraction bound to histone H2B was isolated. DNA was mapped on pICL-Pt using qPCR (see methods section for quantification). **B)** As in A) but using pICL-Pt. **C)** Independent experimental replicate of B).
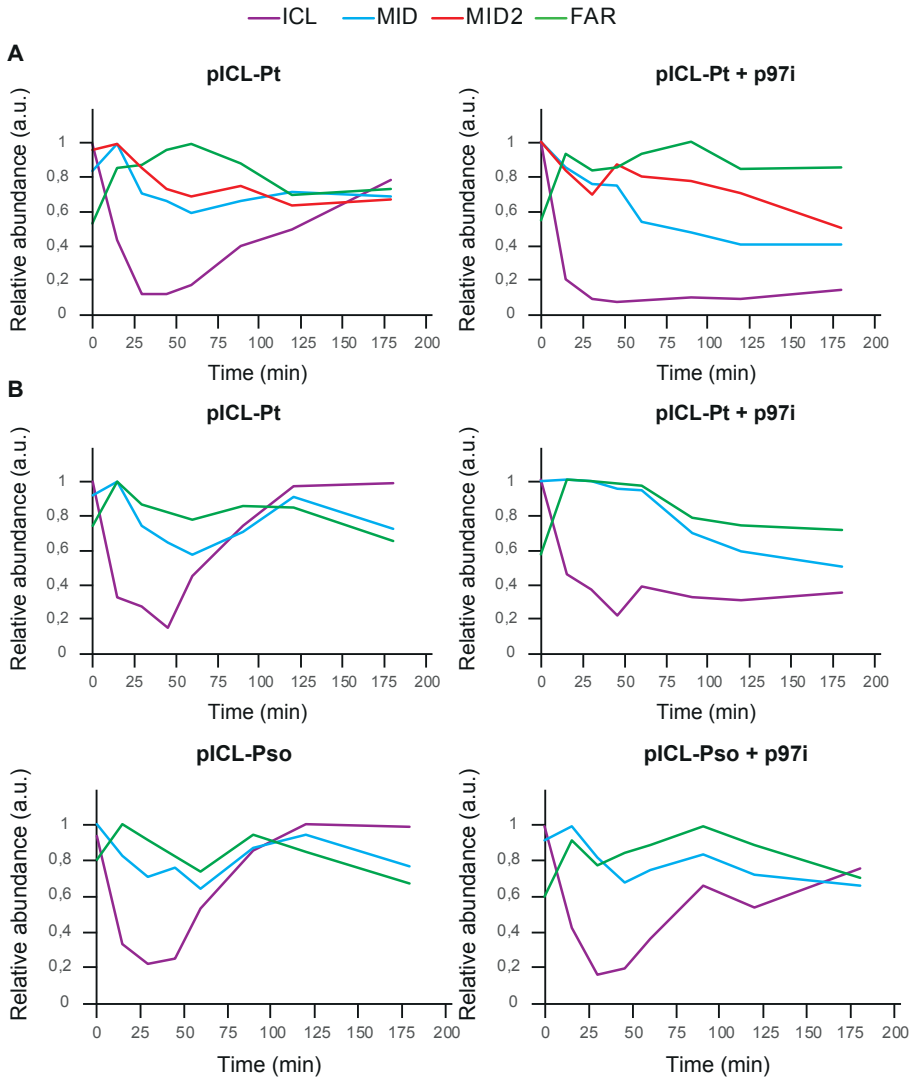
Figure S5: Nucleosome restoration requires ICL repair completion
A) pICL-Pt was replicated in *Xenopus* egg extracts, with the addition of NMS 873 (p97i) where indicated. Samples were digested with MNase and DNA was analyzed by qPCR. Performed in the same experiment as Figure 4B and 4C. Independent experimental replicate of Figure 3B and 3C. **B)** As in A) using either pICL-Pt or pICL-Pso. Independent experimental replicate of A) and Figure 4B and 4C.

References

1. Bustin, M., Goldblatt, D. & Sperling, R. Chromatin Structure Visualization by Immunoelectron Microscopy. *Cell* **7**, 297-304 (1976).

2. Luger, K., Mäder, A., Richmond, R., Sargent, D.F. & Richmond, T.J. Crystal structure of the nucleosome core particle at 2.8Å resolution. *Nature* **389**, 251–260 (1997).

3. Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693-705 (2007).

4. Talbert, P.B. & Henikoff, S. Environmental responses mediated by histone variants. *Trends Cell Biol* **24**, 642-50 (2014).

5. Maze, I., Noh, K.M., Soshnev, A.A. & Allis, C.D. Every amino acid matters: essential contributions of histone variants to mammalian development and disease. *Nat Rev Genet* **15**, 259-71 (2014).

6. Seeber, A., Hauer, M. & Gasser, S.M. Nucleosome remodelers in double-strand break repair. *Curr Opin Genet Dev* **23**, 174-84 (2013).

7. Lemaitre, C. & Soutoglou, E. Double strand break (DSB) repair in heterochromatin and heterochromatin proteins in DSB repair. *DNA Repair (Amst)* **19**, 163-8 (2014).

8. Cleaver, J.E. Nucleosome structure controls rates of excision repair in DNA human cells. *Nature* **270**, 451-453 (1977).

9. Bodell, W.J. & Cleaver, J.E. Transient conformation changes in chromatin during excision repair of ultraviolet damage to DNA. *Nucleic Acids Research* **9**, 203-213 (1981).

10. Smerdon, M.J., Watkins, J.F. & Lieberman, M.W. Effect of histone H1 removal on the distribution of ultraviolet-induced deoxyribonucleic acid repair synthesis within chromatin. *Biochemistry* **21**, 3879-85 (1982).

11. Smerdon, M.J. & Lieberman, M.W. Nucleosome rearrengement in human chromatin during UV-induced DNA repair synthesis. *Proc. Natl. Acad. Sci. USA* **75**, 4238-4241 (1978).

12. Green, C.M. & Almouzni, G. When repair meets chromatin. *EMBO reports* **3**, 28-33 (2002).

13. Polo, S.E. & Almouzni, G. Chromatin dynamics after DNA damage: The legacy of the access-repair-restore model. *DNA Repair (Amst)* **36**, 114-121 (2015).

14. Humpal, S.E., Robinson, D.A. & Krebs, J.E. Marks to stop the clock: histone modifications and checkpoint regulation in the DNA damage response. *Biochemistry and Cell Biology* **87**(2009).

15. van Attikum, H. & Gasser, S.M. Crosstalk between histone modifications during the DNA damage response. *Trends Cell Biol* **19**, 207-17 (2009).

16. Escargueil, A.E., Soares, D.G., Salvador, M., Larsen, A.K. & Henriques, J.A. What histone code for DNA repair? *Mutat Res* **658**, 259-70 (2008).

17. Rogakou, E.P., Pilch, D.R., Orr, A.H., Ivanova, V.S. & Bonner, W.S. DNA Double-stranded Breaks Induce Histone H2AX Phosphorylation on Serine 139. *The Journal of Biological Chemistry* **273**, 5858–5868 (1997).

18. Kobayashi, J., Antoccia, A., Tauchi, H., Matsuura, S. & Komatsu, K. NBS1 and its functional role in the DNA damage response. *DNA Repair (Amst)* **3**, 855-61 (2004).

19. Ward, I.M. & Chen, J. Histone H2AX is phosphorylated in an ATR-dependent manner in response to replicational stress. *J Biol Chem* **276**, 47759-62 (2001).

20. Matsumoto, M. et al. Perturbed gap-filling synthesis in nucleotide excision repair causes histone H2AX phosphorylation in human quiescent cells. *J Cell Sci* **120**, 1104-12 (2007).

21. Burma, S., Chen, B.P., Murphy, M., Kurimasa, A. & Chen, D.J. ATM phosphorylates histone H2AX in response to DNA double-strand breaks. *J Biol Chem* **276**, 42462-7 (2001).

22. Garaycoechea, J.I. et al. Genotoxic consequences of endogenous aldehydes on mouse haematopoietic stem cell function. *Nature* **489**, 571-5 (2012).

23. Duxin, J.P. & Walter, J.C. What is the DNA repair defect underlying Fanconi anemia? *Curr Opin Cell Biol* **37**, 49-60 (2015).

24. Langevin, F., Crossan, G.P., Rosado, I.V., Arends, M.J. & Patel, K.J. Fancd2 counteracts the toxic effects of naturally produced aldehydes in mice. *Nature* **475**, 53-8 (2011).

25. Kim, H. & D'Andrea, A.D. Regulation of DNA cross-link repair by the Fanconi anemia/BRCA pathway. *Genes Dev* **26**, 1393-408 (2012).

26. Ceccaldi, R., Sarangi, P. & D'Andrea, A.D. The Fanconi anemia pathway: new players and new functions. *Nature reviews molecular cell biology* **17**(2016).

27. Zhang, J. et al. DNA interstrand cross-link repair requires replication-fork convergence. *Nat Struct Mol Biol* **22**, 242-7 (2015).

28. Fullbright, G., Rycenga, H.B., Gruber, J.D. & Long, D.T. p97 Promotes a Conserved Mechanism of Helicase Unloading during DNA Cross-Link Repair. *Mol Cell Biol* **36**, 2983-2994 (2016).

29. Wu, R.A. et al. TRAIP is a master regulator of DNA interstrand crosslink repair. *Nature* **567**, 267-272 (2019).

30. Fu, Y.V. et al. Selective bypass of a lagging strand roadblock by the eukaryotic replicative DNA helicase. *Cell* **146**, 931-41 (2011).

31. Long, D.T., Joukov, V., Budzowska, M. & Walter, J.C. BRCA1 promotes unloading of the CMG helicase from a stalled DNA replication fork. *Mol Cell* **56**, 174-85 (2014).

32. Amunugama, R. et al. Replication Fork Reversal during DNA Interstrand Crosslink Repair Requires CMG Unloading. *Cell Rep* **23**, 3419-3428 (2018).

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

34. Klein Douwel, D. et al. XPF-ERCC1 acts in Unhooking DNA interstrand crosslinks in cooperation with FANCD2 and FANCP/SLX4. *Mol Cell* **54**, 460-71 (2014).

35. Budzowska, M., Graham, T.G., Sobeck, A., Waga, S. & Walter, J.C. Regulation of the Rev1-pol zeta complex during bypass of a DNA interstrand cross-link. *EMBO J* **34**, 1971-85 (2015).

36. Long, D.T., Raschle, M., Joukov, V. & Walter, J.C. Mechanism of RAD51-dependent DNA interstrand cross-link repair. *Science* **333**, 84-7 (2011).

37. Semlow, D.R., Zhang, J., Budzowska, M., Drohat, A.C. & Walter, J.C. Replication-Dependent Unhooking of DNA Interstrand Cross-Links by the NEIL3 Glycosylase. *Cell* **167**, 498-511 e14 (2016).

38. Hodskinson, M.R. et al. Alcohol-derived DNA crosslinks are repaired by two distinct mechanisms. *Nature* **579**, 603-608 (2020).

39. Almouzni, G. & Mechali, M. Assembly of spaced chromatin involvement of ATP and DNA topoisomerase activity. *The EMBO Journal* **7**, 4355-4365 (1988).

40. Gaillard, P.L. et al. Chromatin Assembly Coupled to DNA Repair: A New Role for Chromatin Assembly Factor I. *Cell* **86**, 887-896 (1996).

41. Borodovsky, A. et al. A novel active site-directed probe specific for deubiquitylating enzymes reveals proteasome association of USP14. *The EMBO Journal* **20**(2001).

3

42. Zhou, J. et al. Human CHD1 is required for early DNA-damage signaling and is uniquely regulated by its N terminus. *Nucleic Acids Res* **46**, 3891-3905 (2018).

43. Rother, M.B. et al. CHD7 and 53BP1 regulate distinct pathways for the re-ligation of DNA double-strand breaks. *Nat Commun* **11**, 5775 (2020).

44. Courilleau, C. et al. The chromatin remodeler p400 ATPase facilitates Rad51-mediated repair of DNA double-strand breaks. *J Cell Biol* **199**, 1067-81 (2012).

45. Xu, Y. et al. The p400 ATPase regulates nucleosome stability and chromatin ubiquitination during DNA repair. *J Cell Biol* **191**, 31-43 (2010).

46. Rother, M.B. & van Attikum, H. DNA repair goes hip-hop: SMARCA and CHD chromatin remodellers join the break dance. *Philos Trans R Soc Lond B Biol Sci* **372**(2017).

47. Oliveira, D.V. et al. Histone chaperone FACT regulates homologous recombination by chromatin remodeling through interaction with RNF20. *J Cell Sci* **127**, 763-72 (2014).

48. Densham, R.M. et al. Human BRCA1-BARD1 ubiquitin ligase acitivty counteracts chromatin barriers to DNA resection. *Nature Structural & Molecular Biology* **23**(2016).

49. Manickavinayaham, S. et al. E2F1 acetylation directs p300/CBP-mediated histone acetylation at DNA double-strand breaks to facilitate repair. *Nat Commun* **10**, 4951 (2019).

50. Balbo Pogliano, C. et al. ASH1L histone methyltransferase regulates the handoff between damage recognition factors in global-genome nucleotide excision repair. *Nat Commun* **8**, 1333 (2017).

51. Kapetanaki, M.G. et al. The DDB1-CUL4ADDB2 ubiquitin ligase is deficient in xeroderma pigmentosum group E and targets histone H2A at UV-damaged DNA sites. *Proc Natl Acad Sci U S A* **103**, 2588-93 (2006).

52. Piquet, S. et al. The Histone Chaperone FACT Coordinates H2A.X-Dependent Signaling and Repair of DNA Damage. *Mol Cell* **72**, 888-901 e7 (2018).

53. Wang, R., Wang, S., Dhar, A., Peralta, C. & Pavletich, N.P. DNA clamp function of the monoubiquitinated Fanconi anaemia ID complex. *Nature* **580**, 278-282 (2020).

54. Costelloe, T. et al. The yeast Fun30 and human SMARCAD1 chromatin remodellers promote DNA end resection. *Nature* **489**, 581-4 (2012).

55. Kurat, C.F., Yeeles, J.T.P., Patel, H., Early, A. & Diffley, J.F.X. Chromatin Controls DNA Replication Origin Selection, Lagging-Strand Synthesis, and Replication Fork Rates. *Mol Cell* **65**, 117-130 (2017).

56. Yang, J. et al. The Histone Chaperone FACT Contributes to DNA Replication-Coupled Nucleosome Assembly. *Cell Rep* **14**, 1128-1141 (2016).

57. Prendergast, L., Hong, E., Safina, A., Poe, D. & Gurova, K. Histone chaperone FACT is essential to overcome replication stress in mammalian cells. *Oncogene* **39**, 5124-5137 (2020).

58. Enoiu, M., Jiricny, J. & Scharer, O.D. Repair of cisplatin-induced DNA interstrand crosslinks by a replication-independent pathway involving transcription-coupled repair and translesion synthesis. *Nucleic Acids Res* **40**, 8953-64 (2012).

59. Lebofsky, R., Takahashi, T. & Walter, J.C. DNA Replication in Nucleus-Free Xenopus Egg Extracts. in *Methods in Molecular Biology, DNA Replication*, Vol. 521 (2009).

60. Knipscheer, P. et al. The Fanconi Anemia Pathway Promotes Replication-Dependent DNA Interstrand Cross-Link Repair. *Science* **326**, 1698-1701 (2009).

61. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367-72 (2008).

62.   Temu, T., Mann, M., Raschle, M. & Cox, J. Homology-driven assembly of NOn-redundant protEin sequence sets (NOmESS) for mass spectrometry. *Bioinformatics* **32**, 1417-9 (2016).

3
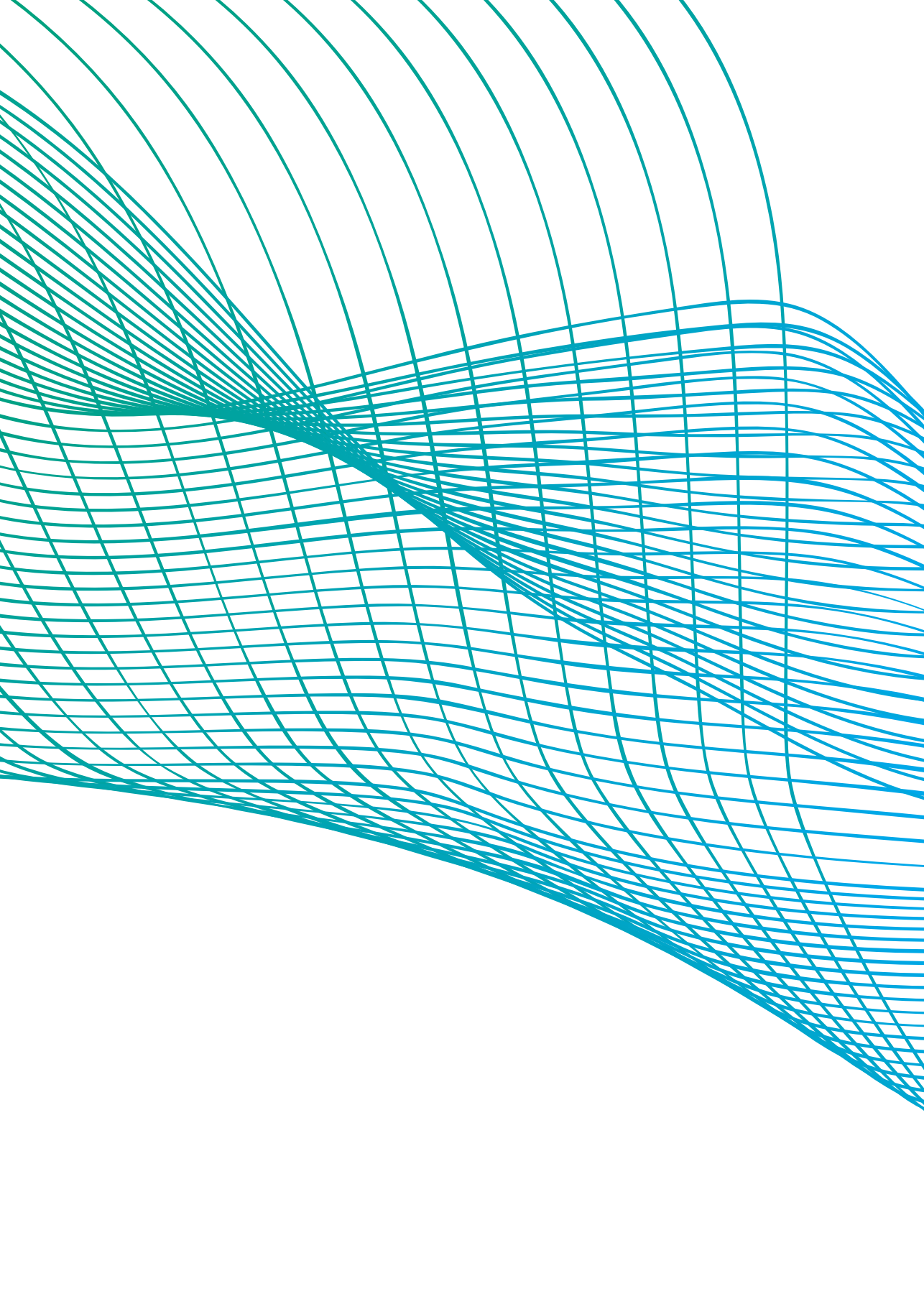
CHAPTER 4

Does FANCD2 act as a histone chaperone during DNA interstrand crosslink repair?

Alice Bolner[1], Wouter Hoogenboom[1], and Puck Knipscheer[1]

[1]Hubrecht Institute – KNAW, University Medical Center Utrecht & Oncode Institute, the Netherlands
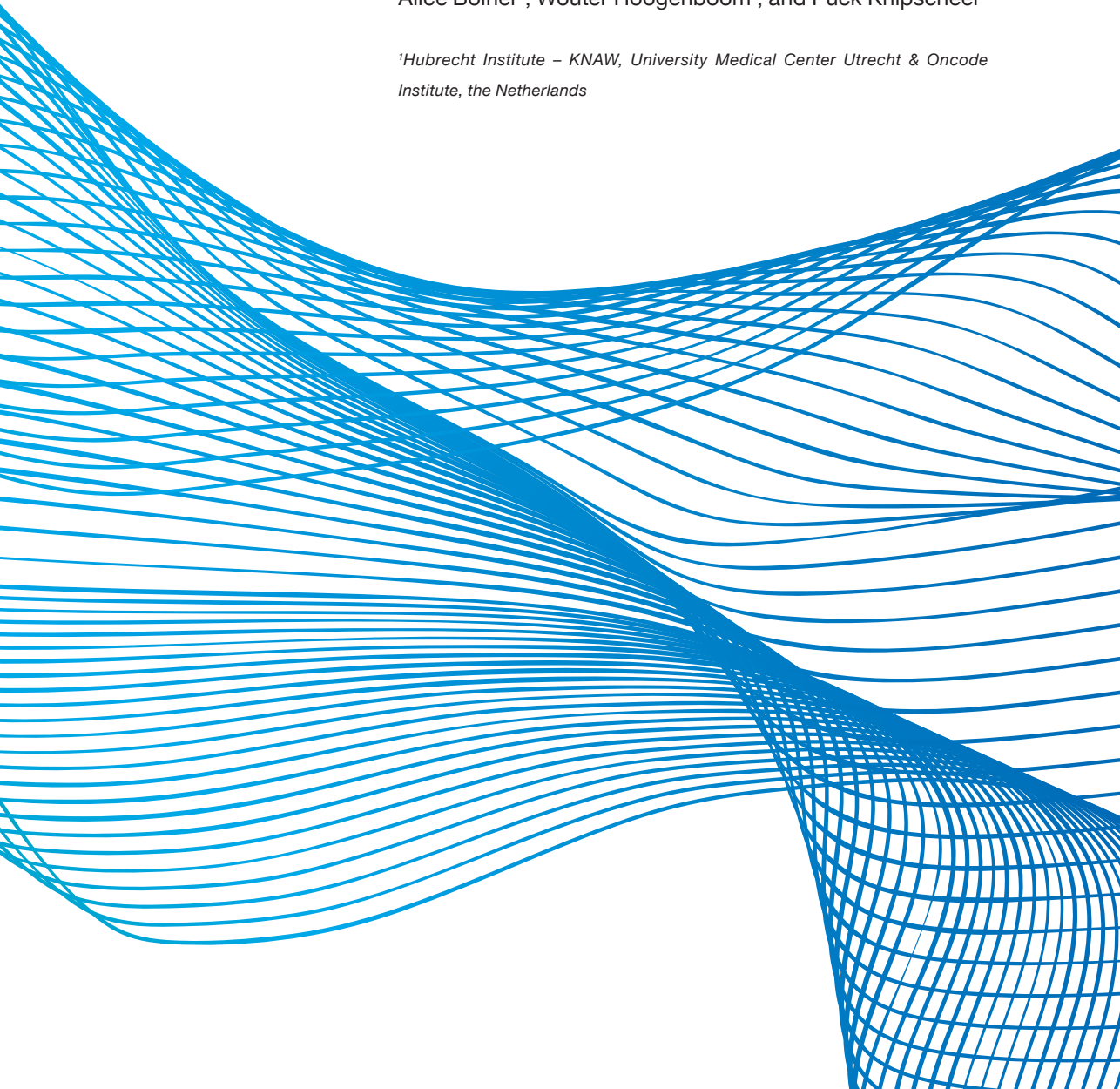
DNA interstrand crosslinks (ICLs) are toxic DNA lesions repaired mainly by the Fanconi anemia (FA) pathway. This repair pathway is coupled to DNA replication and is initiated by the recruitment and monoubiquitination of the FANCI-FANCD2 complex, that promotes ICL-unhooking. In addition to its role in ICL unhooking, FANCD2 was proposed to act as histone chaperone during ICL repair. In the present study, we investigate the role of FANCD2 in chromatin dynamics during ICL repair using assays based on recombinant proteins as well as *Xenopus* egg extracts. We found that recombinant FANCI-FANCD2 complex transiently binds to histones but cannot assemble nucleosomes in a reconstitution assays. Depletion of FANCD2 from *Xenopus* egg extracts did not affect the depletion of nucleosomes from the ICL region at the early stages of repair, however it impaired repair-coupled nucleosomal reconstitution. To understand if this was due to the role of FANCD2 in ICL unhooking or in histone dynamics, we purified FANCD2[R305W], which reportedly lacks histone chaperone activity. In *Xenopus* egg extracts, this mutant was defective in ICL recruitment and ubiquitination, and could therefore not be used to test the histone chaperone function of FANCD2. These observations do not support a direct role for FANCD2 in nucleosome (dis)assembly during ICL repair, however more experiments are necessary to address whether the binding of FANCD2 to histones may have a role in the FA pathway.

Introduction

Fanconi anemia (FA) is a clinically heterogeneous genetic disorder associated with cancer predisposition and bone marrow failure[1]. On the cellular level, FA cells are highly sensitive to compounds that induce extremely toxic DNA lesion: DNA interstrand crosslinks (ICLs)[2-4]. These DNA lesions covalently connect the two strands of the double helix, thereby preventing essential cellular functions such as DNA replication and transcription[2]. FA is caused by mutations in one of the 22 currently known FA genes: *FANC-A, -B, -C, -D1, -D2, -E, -F, -G, -I, -J, -L, -M, -N, -O, -P, -Q, -R, -S, -T, -U, -V* and *-W*, that act together in the FA pathway of ICL repair[3]. A key step in the activation of the FA pathway is the monoubiquitination of the FANCI-FANCD2 (ID2) heterodimer by the FA core complex (FANCA, -B, -C, -E, -F, -G, -L, and the FA core complex associated proteins FAAP20, -24 and -100)[5,6], which leads to the recruitment of DNA repair factors to the stalled replication fork[5-8]. The *Xenopus* egg extract system has been used to define many of the mechanistic details of the FA pathway, including the role of FANCI-FANCD2 ubiquitination during ICL repair[5,6]. By replicating plasmids containing chemically defined ICLs in extracts, we and others have demonstrated that repair is initiated with the convergence of two replication forks at the ICL site (Figure 1A)[6,7] and the ubiquitination of the ID2 complex[5]. This is followed by the unloading of the replicative CMG (CDC45, MCM2-7, GINS) helicase and the approach of one of the leading strands to the ICL[6]. A portion of the approached replication forks can undergo fork reversal, a process that involves the reannealing of the parental strands to form a four-way DNA structure[9]. The ubiquitination of the ID2 heterodimer is a key step in the FA pathway as it promotes the ICL unhooking incisions via the recruitment of the nuclease complex FANCQ (XPF)-ERCC1 and the scaffold protein FANCP (SLX4)[8]. The unhooked crosslink is then bypassed by translesion synthesis (TLS), and the double strand break (DSB) generated at the incised strand is repaired by homology-mediated repair[6].

In addition to its role in ICL repair, other roles for FANCD2 in the DNA damage response have been reported. It localizes to stalled replication forks, where it prevents nucleolytic degradation of the nascent DNA and promotes fork restart upon completion of DNA repair[10-12]. At common fragile sites (CFSs), FANCD2 facilitates the progression of the replication fork, possibly by resolving DNA:RNA hybrids[13]. Additionally, it can regulate dormant origin firing at CFSs, in order to promote completion of DNA replication and avoid mitotic instability[13]. FANCD2 also flanks replication stress-induced ultra-fine DNA bridges, DNA connections that link two sister chromatids in the anaphase of mitosis[14], possibly stabilizing these DNA structures.

The repeating unit of chromatin, the nucleosome, is a nucleoprotein complex formed by a histone octamer (composed of a H3/H4 histone tetramer and two H2A/H2B histone dimers) and 146 bp of DNA. Nucleosomes are dynamic complexes that

are assembled and disassembled during cellular processes such as transcription, and DNA replication and repair. Multiple studies link FANCD2 to chromatin dynamics at the sites of DNA damage[15-17]. FANCD2 was reported to directly bind histone H3/H4 *in vitro* via a C-terminal histone binding domain[15,17]. Mutations in this region compromised cell survival upon cisplatin treatment[15], suggesting that the interaction between FANCD2 and histones may be involved in ICL repair. Moreover, within this histone binding domain of FANCD2, a methyl-lysine-binding domain (MBD) was detected, which allowed FANCD2 to bind methylated H4K20 and form discrete nuclear foci during ICL repair[17]. Disruption of the MBD resulted in decreased affinity of FANCD2 for chromatin and increased sensitivity to ICL inducing agents, suggesting that histone binding may help to recruit FANCD2 to the damage site. In addition, a recent study proposed that *chicken* FANCD2 is directly involved in the assembly of nucleosomes[15]. In support of this idea, *chicken* FANCD2[R305W], the equivalent of a FA-causing mutation, *h*FANCD2[R302W],[18] was shown to lack histone chaperone activity *in vitro*[15]. Cells expressing this mutant were more sensitive to cisplatin treatment compared to wild-type, even though *c*FANCD2[R305W] could still be efficiently ubiquitinated[15]. The authors therefore proposed that FANCD2 may act as a histone chaperone during ICL repair independently of its function in unhooking incisions.

Previous observations in our laboratory indicated that during ICL repair FANCD2 is initially recruited to the lesion site and that in time, multiple FANCD2 molecules are loaded and then spread out to the surrounding DNA region[8]. We hypothesized that this spreading may be linked to the histone chaperone function of FANCD2, and we therefore set out to investigate the role of FANCD2 in nucleosome dynamics during ICL repair. We found that the purified recombinant *Xenopus laevis* FANCI-FANCD2 complex could transiently interact with histones but did not support the assembly of nucleosomes in reconstitution assays. Depletion of FANCD2 from *Xenopus* egg extracts hardly affected the progressive depletion of nucleosomes at early stages of ICL repair, but prevented the restoration of the nucleosomal landscape upon completion of repair. To understand if this defect was caused by inhibition of FANCD2-mediated nucleosome assembly or by a defect in FANCD2-mediated ICL unhooking, we tested a potential separation of function mutant, *xl*FANCD2[R305W]. We observed that the protein had both a recruitment and ubiquitination defect, and could therefore not be used to separate these functions. While we conclude that FANCD2 is likely not involved in nucleosome depletion, we cannot formally exclude that is acts in nucleosome restoration at later stages of repair. Further studies will establish whether the histone binding properties of FANCD2 may play a role in ICL repair.

Results

The *Xenopus laevis* FANCI-FANCD2 complex does not support efficient nucleosome assembly

To study the nucleosome assembly properties of *Xenopus laevis* FANCD2, we purified the recombinant *xl*FANCI-FANCD2 (ID2) complex from insect cells (*xl*FANCD2 alone is not stable, unpublished results) (Figure 1B, see also Methods). We analyzed the nucleosome assembly properties of the recombinant ID2 complex using a DNA supercoiling assay (Figure 1C). In this assay, relaxed circular DNA is incubated with histones and a (potential) histone chaperone, and the induction of negative supercoils in the plasmid is used as readout for nucleosome assembly. As supercoiled DNA migrates faster than relaxed DNA on a native agarose gel, DNA supercoiling can be analyzed by gel electrophoresis. As a positive control we used the truncated form of the histone chaperone CAF-1 (tCAF-1), which is able to assemble nucleosomes *in vitro*[19]. As expected, tCAF-1 induced supercoiling of the plasmid (Figure 1D), indicating efficient nucleosome assembly. In contrast, incubation of ID2 complex with histones and DNA, resulted only in a very low degree of supercoiling (Figure 1D, black arrow), indicating inefficient nucleosome assembly (Figure 1D). Thus, even though ID2 may display some nucleosome assembly properties, *xl*ID2 is less efficient at nucleosome assembly than the *chicken* ID2 complex as previously reported[15].

Although often DNA supercoiling is used as readout for nucleosome assembly, these studies are prone to misinterpretation, since native gel analysis cannot distinguish the wide variety of sub-nucleosomal structures that DNA and histones can form. For this reason, we decided to further examine the nucleosome assembly properties of the *xl*ID2 complex using the quantitative nucleosome assembly assay (NAQ assay), which is based on MNase digestion[20] (Figure 1E).

4

A

Fork convergence
ID2 ubiquitination

CMG unloading

ICL unhooking

TLS

HR

B

ID2

FANCD2
FANCI

C

relaxed DNA

supercoiled
DNA

histone chaperone
+ histones

D

tCAF1   ID2

− + + + + + + + +   *Topoisomerase I*
− − + + + + + + +   *Histones*

Relaxed DNA

Supercoiling
intermediates
*

Supercoiled DNA

F

ID2
ID2 + 207 bp DNA
ID2 + H3/H4 tetramer
ID2 + H2A/H2B tetramer
ID2 + octamer
ID2 + octamer + 207 bp DNA
tCAF1 + octamer + 207 bp DNA
octamer + 207bp DNA
207 bp DNA

nucleosome
hexamer

DNA

DNA

nucleosome
hexamer

histones

1   2   3   4   5   6   7   8   9

E

Nucleosome assembly

histone chaperone
+ histones   DNA   MNase   loading
control

nucleosomal
footprint

visualization of nucleosomes

G

xlID2 + octamer + 207bp DNA
tCAF1 + octamer + 207bp DNA
octamer + 207bp DNA

loading
control
150bp

DNA

1   2   3

Figure 1: FANCD2 interacts with histone but does not assemble nucleosomes in reconstitution assays

A) Schematic representation of the FA pathway of ICL repair. **B)** Recombinant *xl*FANCD2 and *xl*FANCI were co-expressed in insect cells and the ID2 complex was purified using the FLAG tag on *xl*FANCI and gel filtration chromatography. **C)** Schematic representation of the supercoiling assay. In brief, histones and chaperones are incubated with DNA that was relaxed by topoisomerase I treatment. Nucleosome assembly on the plasmid results in introduction of supercoils, that can be visualized by enhanced mobility on an agarose gel after DNA extraction. **D)** A supercoiling assay was performed by incubating either ID2 or tCAF-1 with histones and relaxed plasmid DNA. Samples were analyzed by gel electrophoresis and visualized by Sybr Gold staining. Black arrow indicates supercoiling intermediates formed upon ID2 incubation with histones and DNA. Asterisk indicates DNA present in the ID2 protein preparation. **E)** Schematic representation of the NAQ assay. Histones are first incubated with the histone chaperone, then DNA is added to allow formation of nucleosomes. Nucleosome assembly is analyzed both by PAGE and by MNase digestion followed by PAGE. **F)** Histones were incubated with either ID2 or tCAF-1 and DNA and the reactions were loaded on a PAGE gel. DNA was visualized by Sybr Gold staining (top panel) and histones were visualized by the Alexa Fluor 647 tag (bottom panel). Independent experimental replicate is presented in Figure S1A. **G)** Samples from F were digested with MNase and the nucleosomal footprint was visualized on a PAGE gel stained with Sybr Gold. Independent experimental replicate is presented in Figure S1B.

We incubated the ID2 complex with a 207 bp DNA fragment and fluorescently-labelled histone octamers, and analyzed the products using native polyacrylamide gel electrophoresis (PAGE) and Sybr Gold staining (Figure 1F, top panel and Figure S1A, top panel). In addition, we visualized histones using a fluorescent dye conjugated to an engineered single cysteine (T112C) on histone 2B (H2B) (Figure 1F, bottom panel and Figure S1A, bottom panel). Incubation of tCAF-1 with DNA and histones gave rise to two shifted DNA bands representing the full nucleosome and the hexasome (a known intermediate lacking one of the H2A-H2B dimers[20]), as also confirmed by the fluorescent signal (Figure 1F, lane 7 and Figure S1A). When DNA was incubated only with histone octamer, some background signal was visible but the nucleosome band was not well-defined, indicating it is caused by the aspecific binding of histones to DNA[20](Figure 1F, lane 8 and Figure S1A). In contrast, incubation of the ID2 complex with DNA and histones did not result in a nucleosomal band (Figure 1F lane 6 and Figure S1A). To confirm formation of a full nucleosome, the NAQ assay includes digestion of the products by MNase, followed by isolation of the DNA and separation of the products by native PAGE. This allows visualization of the MNase-protected DNA fragment, corresponding to the nucleosomal footprint (~150 bp). We could observe the formation of the nucleosomal footprint in the reaction products containing tCAF-1 (Figure 1G, lane 2 and Figure S1B), but not in the reaction products containing ID2 (Figure 1G, lane 1 and Figure S1B). In the DNA-octamer only condition, a smear of aspecific bands was detected, confirming non-specific binding of histones to DNA (Figure 1G, lane 3 and Figure S1B). Interestingly, in the presence of the ID2 heterodimer this unspecific signal was not observed (Figure 1F lane 6, Figure 1G, lane 1 and Figure S1A and B), which could indicate that the ID2 complex interacts with histones and/or DNA transiently, affecting their interacting with each other. Overall, these results indicate

that the ID2 complex is not able to assemble nucleosomes in reconstition assays. However, we noticed that incubation of ID2 with histones (and no DNA) resulted in fluorescent molecules entering the native gel as a high molecular weight smear (Figure 1F, lane 3-6, Figure S1A and D), indicating interaction between ID2 and histones. It is therefore possible that the ID2 complex could have some affinity for histones. We wondered if the ID2 heterodimer could interact with preassembled nucleosomes. To test this, we prepared salt assembled nucleosomes (see Methods) on 207 bp or 147 bp DNA fragments. Because the nucleosomal footprint is 147 bp, the 207 bp fragment will leave some free DNA outside of the nucleosome. However, incubation of ID2 with either of these preassembled nucleosomes did not result in a shift of the nucleosomal bands by native PAGE (Figure S1A and C). Moreover, in the presence of ID2 we observed no protection of the DNA flanking the nucleosome assembled on the 207 bp DNA fragment from MNase digestion (Figure S1B), further indicating that while *xl*ID2 has some affinity for free histones, it does not stably bind nucleosomes.

Figure 2: FANCD2 depletion affects nucleosome positioning during ICL repair.
A) Schematic representation of the ICL repair assay in *Xenopus* egg extracts. In brief, the ICL overlaps with a SAPI restriction enzyme recognition site. Repair is quantified by measuring the reconstitution of the restriction site. **B)** Mock depleted extracts (mock), FANCD2 depleted (ΔFANCD2) extracts and FANCD2 depleted extracts complemented with recombinant ID2 (ΔFANCD2+ID2^{WT}) were used to replicate pICL-Pt in the presence of radio-labelled nucleotides (^{32}P-α-dCTP). Repair efficiency was calculated by quantifying the regeneration of the SAPI restriction site and was plotted. Dotted line represents SAPI fragments from contaminating uncrosslinked plasmid present in varying degrees in different pICL preparations. Independent experimental replicate is presented in Figure S2A. **C)** Schematic representation of the primers used for the nucleosome positioning assay. **D-F)** Replication/repair reactions from B to which no ^{32}P-α-dCTP was added, were digested with MNase. The resulting DNA fragments were purified and mapped to the plasmid by qPCR, using the primers represented in C. The DNA abundance of each primer set was normalized to pQuant and the peak value was set to 1, and the values were plotted against reaction time. Independent experimental replicates are presented in Figure S2B and C.

A

B

C

D Mock

E ΔFANCD2

F ΔFANCD2+ID2^WT
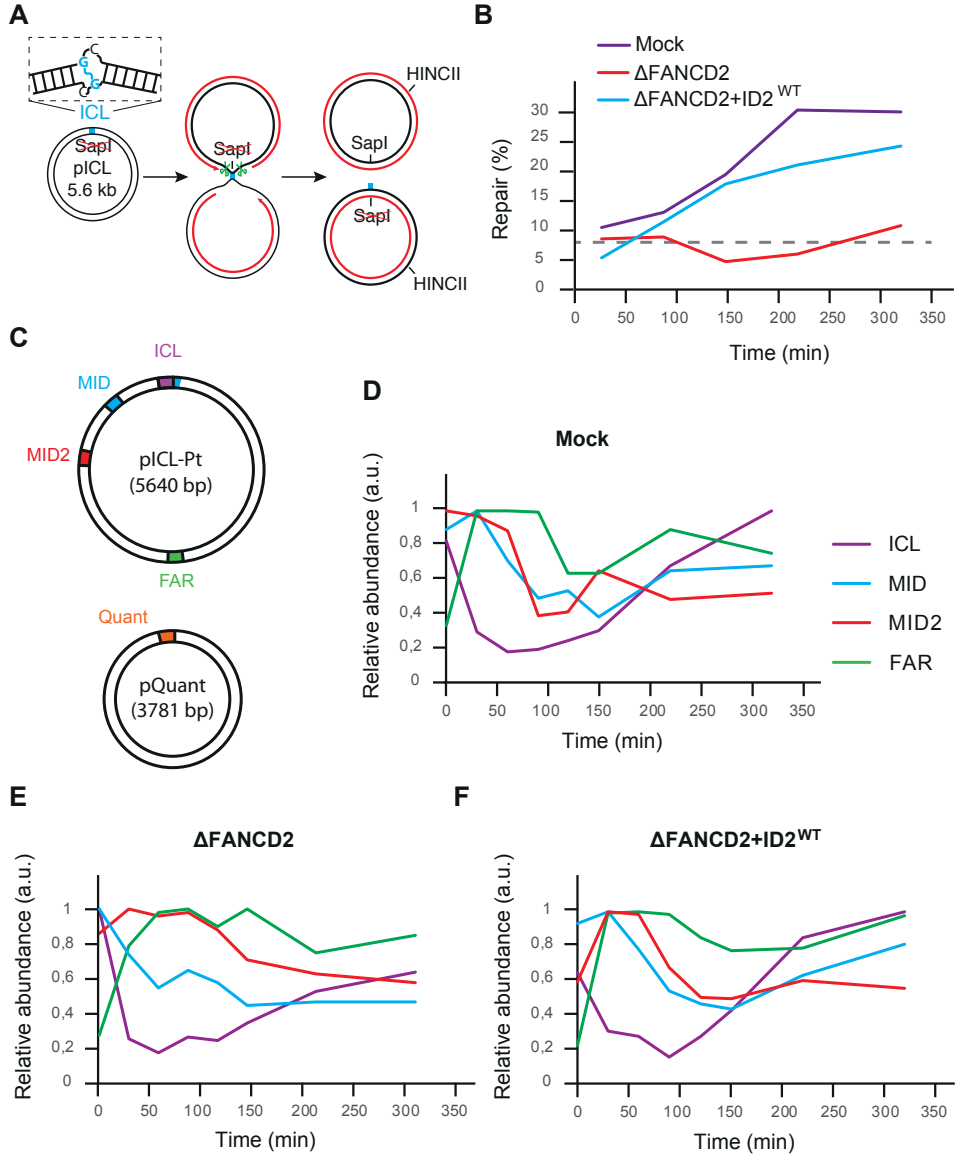
Depletion of FANCD2 affects nucleosomal dynamics during ICL repair in *Xenopus* egg extracts

In the context of ICL repair, the ID2 heterodimer binds to branched DNA structures[21] and interacts with other repair factors. We therefore reasoned that the role of the ID2 complex in chromatin dynamics could depend on one or both of these conditions. To study the potential histone chaperone activity of FANCD2 in the context of ICL repair, we therefore used the *Xenopus* egg extract system, which recapitulates the physiological mechanism of replication-coupled repair of ICLs[6]. To study the effect of FANCD2 depletion on nucleosome assembly during ICL repair in extract, we replicated a 5.6 kb plasmid containing a cisplatin-induced interstrand crosslink (pICL-Pt), which is repaired solely by the FA pathway (Figure 1A)[5]. We used mock-depleted or FANCD2-depleted extracts, and FANCD2-depleted extract supplemented with recombinant ID2 complex. Replication and repair intermediates were isolated, and repair efficiency was determined by measuring the regeneration of a SAPI restriction enzyme recognition site that is blocked by the ICL prior to repair (Figure 2A)[6]. As previously reported[5], depletion of FANCD2 resulted in a defect in ICL repair, that could be rescued by addition of recombinant ID2 (Figure 2B and Figure S2A). To analyze the dynamics of nucleosomes during repair we used the nucleosome positioning assay (also described in chapter 3 of this thesis). Briefly, repair intermediates were isolated and treated with MNase, resulting in DNA fragments protected by nucleosomes. The abundance of protected fragments at various times during repair was then determined by quantitative PCR (qPCR) using a set of primers that map to four regions on the plasmid: "ICL" region (24-132 bp from the ICL), "MID" region (662-775 bp from the crosslink), "MID2" region (1500-1607 bp from the crosslink) and "FAR" region (3117-3017 bp from the crosslink) (Figure 2C). In mock depleted conditions, we observed a decrease of the ICL amplicon relative abundance over time, reaching a minimum at ~50 minutes (Figure 2D "ICL"). This indicates that, at the earliest stages of ICL repair, the ICL region became sensitive to MNase treatment and therefore was depleted of nucleosomes. The amplicon abundance also decreased for the MID and MID2 regions, but this started later and reached a minimum only at ~80 minutes. This indicates that during repair nucleosomes are gradually removed also from the regions further away from the ICL (Figure 2D, "MID" and "MID2"). As repair products started to accumulate after ~150 minutes (Figure 2B), the abundance of the ICL, MID and MID2 amplicons increased, reflecting the restoration of nucleosomes in these regions. This is in agreement with what we previously observed in undepleted extracts (see chapter 3 of this thesis) and it indicates that during ICL repair nucleosomes are dynamically rearranged in the region surrounding the lesion site (Figure S3). In the absence of FANCD2, we observed that nucleosomes were disassembled from the ICL and MID region similarly to mock condition (Figure 2E and Figure S2B) while the removal of nucleosomes from the MID2 region was slightly delayed compared to the control. At later times, MID and MID2 amplicon abundance did not increase in the absence

of FANCD2, indicating that nucleosomes were not placed back in these regions (Figure 2E and Figure S2B). This is likely due to the lack of repair in the absence of FANCD2 and it indicates that ICL repair is required for the restoration of the nucleosomal landscape. Interestingly, at the ICL site, nucleosomes were partially reassembled even in the absence of FANCD2 (Figure 2E and Figure S2B).

We speculate that this could be due to the induction of replication fork reversal during ICL repair, and the subsequent placement of nucleosomes on the reversed nascent strands[9,22] (Figure S3, see also Discussion). Importantly, adding back recombinant ID2 complex resulted in nucleosomal dynamics comparable to the mock control condition (Figure 2F and Figure S2C). We conclude that depletion of FANCD2 has only a minor effect on nucleosomal dynamics at the early stages of ICL repair, however, at later stages of repair, lack of FANCD2 prevents proper nucleosome re-assembly.

*xl*FANCD2[R305W] is not ubiquitinated during ICL repair in *Xenopus* egg extracts

Next, we set out to determine whether the effect of FANCD2 depletion was due to a defect in its potential histone chaperone function, or in its canonical role in promoting unhooking incisions during ICL repair. Previous studies identified a FA-associated FANCD2 mutant (*h*FANCD2[R302W])[18] that could be efficiently ubiquitinated in response to cisplatin treatment, and therefore likely promote ICL unhooking, but lacked histone chaperone activity[15]. This mutant could therefore potentially act as a separation of function mutant. We purified the equivalent *Xenopus laevis* FANCD2 mutant protein (*xl*FANCD2[R305W]) in complex with *xl*FANCI, as well as the wild-type ID2 complex from insect cells (Figure 3A). To assess whether the mutant could rescue the ICL repair defect caused by FANCD2 depletion, we replicated pICL-Pt in FANCD2-depleted extracts (Figure 3C) supplemented with either wild-type (ID2[WT]) or mutant (ID2[R305W]) ID2 complex, and measured ICL repair by SAPI site regeneration. While ID2[WT] rescued the repair defect caused by FANCD2 depletion, the mutant did not (Figure 3B and Figure S4A). However, in contrast with what was previously reported for the chicken mutant protein[15], *xl*FANCD2[R305W] was not efficiently ubiquitinated during ICL repair in our system (Figure 3D and Figure S4B).

To better understand why FANCD2[R305W] could not be ubiquitinated, we examined whether the protein was recruited to the crosslink during repair. For this, we made use of chromatin immunoprecipitation (ChIP, Figure 4A) using αFANCD2 antibody and pICL-Pt specific qPCR primers (Figure 2C)[8]. The ICL-containing plasmid was replicated in FANCD2-depleted extracts supplemented with wild-type or mutant ID2 complex (Figure 4B and Figure S4C). Reaction samples were taken at various times, proteins were crosslinked to DNA, and the plasmid was fragmented by sonication. After FANCD2 immunoprecipitation, the samples were analyzed by qPCR. An unrelated and undamaged plasmid (pQuant) was added to the reaction as a negative control. FANCD2[WT] accumulated first at the ICL region and, with a slight delay, also

at the at the MID2 and FAR regions as previously demonstrated[8] (Figure 4D and Figure S4D), but not at the control plasmid pQuant. In contrast, recruitment of the FANCD2[R305W] mutant was impaired in all regions of the crosslinked plasmid (Figure 4E and S4E). Because DNA binding is required for efficient FANCD2 ubiquitination[21], this defect in recruitment likely explains the lack of xlFANCD2[R305W] ubiquitination and consequently the ICL repair deficiency. The impaired ICL recruitment of this mutant makes it unsuited as a separation of function mutant.
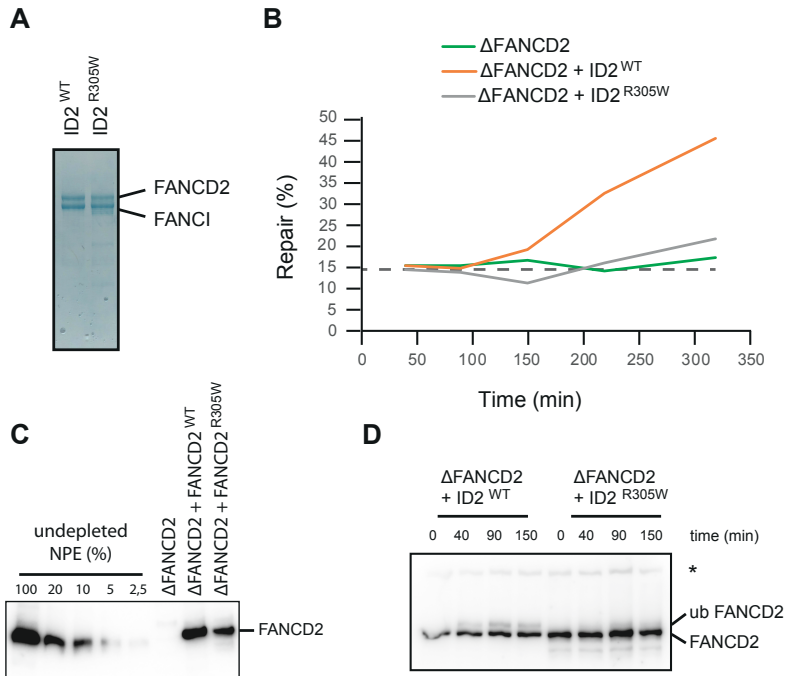
Figure 3: FANCD2[R305W] cannot support ICL due to lack of ubiquitination
A) Recombinant xlFANCD2[WT] or xlFANCD2[R305W] were expressed in complex with xlFANCI in insect cells. The heterodimers were purified using the FLAG tag on xlFANCI and gel filtration chromatography. **B)** Extracts were depleted of FANCD2 (ΔFANCD2) and supplemented with either FANCD2 wild-type in complex with FANCI (ID2[WT]) or FANCD2[R305W] in complex with FANCI (ID2[R305W]). ICL repair was measured as in Figure 2A and plotted. Dotted line represents SAPI fragments from contaminating uncrosslinked plasmid present in varying degrees in different pICL preparations. Independent experimental replicate is presented in Figure S4A. **C)** FANCD2 depleted NPE (ΔFANCD2) and FANCD2 depleted NPE complemented with either wild-type ID2 (ID2[WT]) or FANCD2[R305W] in complex with FANCI (ID2[R305W]) were analyzed by Western blotting using αFANCD2 antibody. A dilution series of undepleted NPE was loaded on the same gel to determine the level of depletion. **D)** Western blot analysis of FANCD2 ubiquitination. Asterisk indicates background band. Independent experimental replicate is presented in Figure S4B.

A

Crosslinked replication sample

Fragmentation of DNA by sonication

Immunoprecipitation of protein of interest and DNA bound to it

Analysis of DNA by qPCR

Antibody against protein of interest

B

C
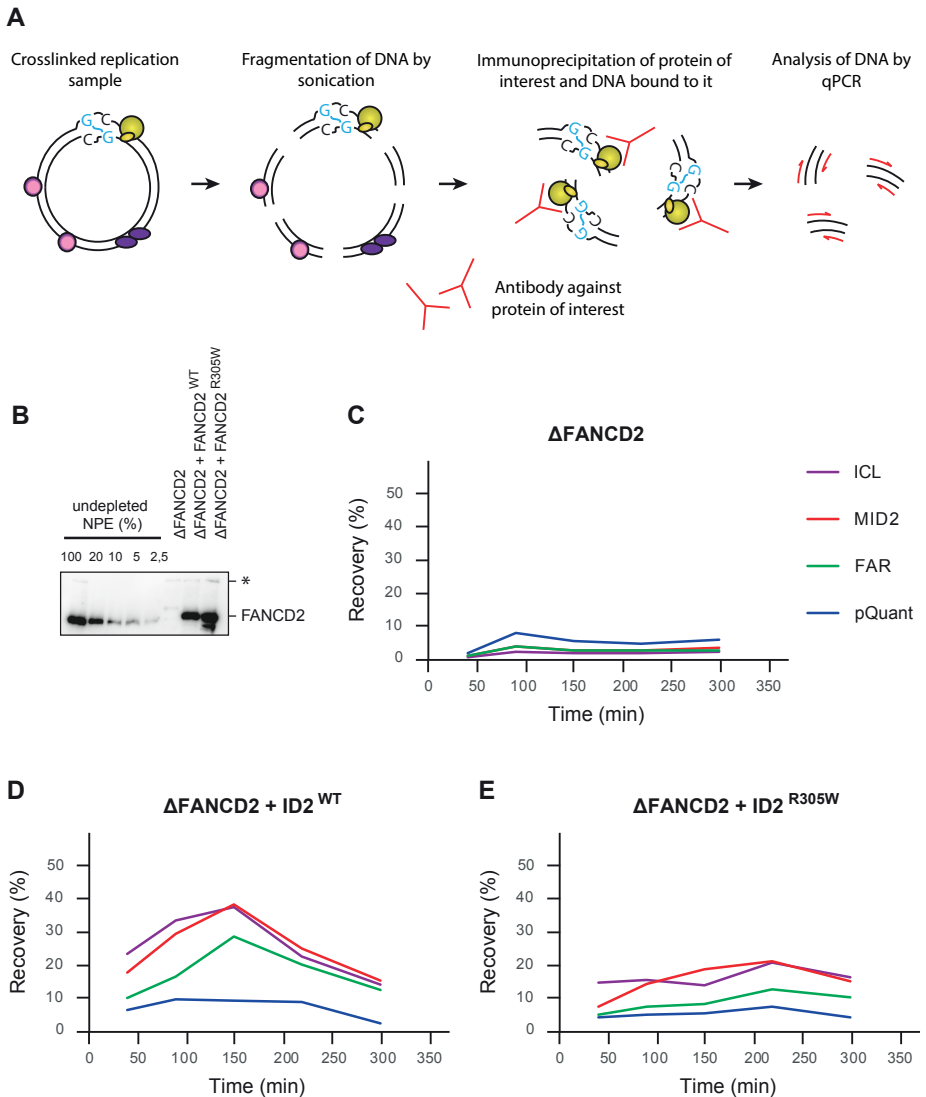
Figure 4: FANCD2^{R305W} is not efficiently recruited to the crosslinked plasmid
A) Schematic representation of the chromatin immunoprecipitation assay (ChIP). **B)** FANCD2 depleted NPE (ΔFANCD2) and FANCD2 depleted NPE complemented with wild-type FANCD2 in complex with FANCI (ID2^{WT}) or FANCD2^{R305W} in complex with FANCI (ID2^{R305W}) were analyzed by Western blot using αFANCD2 antibody. A dilution series of undepleted NPE was loaded on the same blot to determine the level of depletion. **C-E)** FANCD2 depleted extracts (ΔFANCD2), FANCD2 depleted extracts supplemented with wild type ID2 (ΔFANCD2+ ID2^{WT}) and FANCD2 depleted extracts supplemented with FANCD2^{R305W}-FANCI complex (ΔFANCD2+ID2^{R305W}) were used to replicate pICL/pQuant. Reaction samples were isolated and analyzed by FANCD2 ChIP using ICL, MID2, FAR and pQuant primer sets. Independent experimental replicates are presented in Figure S4D and E.

Discussion

The FA pathway has been extensively studied in the past decade, however still virtually nothing is known regarding chromatin dynamics in ICL repair. Both *in vitro* and cellular studies suggested that FANCD2 has histone chaperone functions in ICL repair[15]. Nonetheless, mechanistic information regarding this specific role of FANCD2 is missing. We show that recombinant *xl*ID2 does not assemble nucleosomes and that depletion of FANCD2 during ICL repair in *Xenopus* egg extracts hardly affects initial depletion of nucleosomes around the ICL. Nucleosome repositioning after repair was affected by FANCD2 depletion, however this is likely due to a deficiency in FANCD2-mediated unhooking incisions that are required for repair and would therefore prohibit repair-coupled nucleosome re-assembly. Based on the lack of nucleosome assembly in the reconstitution assays, the low affinity of FANCD2 for histones and the limited effect of FANCD2 depletion on nucleosome removal during ICL repair, we disfavor a model in which FANCD2 directly acts in nucleosome dynamics during ICL repair.

Interestingly, some nucleosomes could be placed back specifically at the ICL locus even in the absence of FANCD2 (Figure 2E). Possibly, under these conditions, a portion of the stalled replication forks undergoes fork reversal (Figure S3, right panel), which is a complex DNA transaction that involves the annealing of the two newly synthesized DNA strands and the re-annealing of the two parental strands, forming a four-way junction. We speculate that nucleosomes could be placed on the reversed fork, resulting in protection of the ICL locus from MNase treatment (Figure S3 right panel). The placement of nucleosomes on a reversed fork has already been observed in previous studies[22] and is believed to promote replication fork restart. Deposition of nucleosomes on the reverse fork could have multiple functions during ICL repair. Firstly, it could limit the access of nucleases to the regressed DNA ends, thereby preventing resection and further fork backtracking. Secondly, post-translational modifications of histones deposited on the reversed fork could recruit factors involved in fork stability and/or restart. Lastly, protection of the regressed fork may promote resection-independent fork restart[23] and prevent unscheduled DNA resection and chromosomal rearrangements.

To formally exclude a role for FANCD2 in nucleosome reconstitution during ICL repair, a separation of function mutant of FANCD2 is required that inhibits this putative role while still promoting ICL unhooking incisions. We purified the FANCD2[R305W] mutant, which was shown to inhibit specifically the histone chaperone function of FANCD2[15], however, in contrast to the previous study, we found that this mutant was not efficiently ubiquitinated nor recruited to the crosslink during repair. Furthermore, in our assays, we found that the *xl*FANCD2[R305W] mutant protein was unstable in *Xenopus* egg extracts (data not shown). We speculate that the defect in the ubiquitination and recruitment of the mutant are due its instability, which is in line with what was previously observed studying cells derived from the FA patient

carrying the mutation (*h*FANCD2[R302W])[18]. The authors studied FANCD2 ubiquitination in the patient-derived fibroblasts and found that the FANCD2 mutant protein was expressed at very low levels. Therefore, both *h*FANCD2[R302W] and its *Xenopus leavis* equivalent seem to be unstable and not support ICL repair. It is possible that the equivalent chicken mutant used in previous studies[15] could be more stable than the human or *Xenopus* equivalents, and therefore could be more functional during ICL repair. Alternatively, since FANCD2[R305W] ubiquitination was previously assessed in cellular assays, it is possible that the overexpression system used may have compensated for the protein instability so that the residual activity of FANCD2[R305W] could promote repair.

Based on our results, we consider unlikely that FANCD2 actively (dis)assembles nucleosomes during ICL repair, however we do not exclude that the interaction between FANCD2 and histones may play a role in ICL repair. Histone binding could be required for the recruitment of FANCD2 to the ICL, as shown previously in cells exposed to mitomycin C[17]. Another study found that FANCD2 binds methylated histone H4 upon induction of replication stress to promote fork stability[16]. It is possible that the histone binding function of FANCD2 acts during ICL repair to stabilize the stalled replication fork and limit DNA resection. To investigate these possibilities, we aim to identify a FANCD2 mutant that can be ubiquitinated and support ICL unhooking but cannot bind to histones. This mutant would be used to test the effect on resection during ICL repair.

Methods

Xenopus egg extracts, DNA replication and repair assays

Xenopus egg extracts were prepared as described previously[24]. The plasmid containing a cisplatin-induced ICL (pICL-Pt) was prepared and replicated in extracts as described in previous studies[6,25]. In brief, the plasmid was first incubated in HSS (high speed supernatant) extract for 20 minutes in the presence of ^{32}P-α-dCTPs at 7.5 ng μl^{-1} final concentration, to allow pre-replication complex assembly. Two volumes of nucleoplasmic egg extract (NPE) were added to HSS, starting a single round of DNA replication. Aliquots of the replication reaction (5 μl sample 2.5 ng μl^{-1} DNA) were taken at various timepoints and stopped with ten volumes of Stop II solution (0.5% SDS, 10 mM ETDA, 50 mM Tris pH 7.5). Samples were treated first with RNase (0.13 μg μl^{-1}) for 30 minutes at 37°C, and then with proteinase K (0.5 μg μl^{-1}) overnight at room temperature. DNA was purified by phenol/chloroform extraction followed by ethanol precipitation, and was resuspended in 5 μl of 10 mM Tris pH 7.5. ICL repair was measured by digesting 1 μl of DNA with HINCII, or HINCII and SAPI, and the digestion products were separated by gel electrophoresis on a 0.8% agarose gel. DNA was visualized by autoradiography and repair efficiency was calculated as reported previously[26].

Antibodies and immunodepletions

FANCD2 antiserum was prepared against residues 1–172 of *xl*FANCD2 and it was tested for specificity by western blotting. To deplete egg extracts of FANCD2, one volume of Dynabeads Protein A (Thermo Fisher) was washed three times with 5 volumes of 100 mM phosphate buffer pH 8, supplemented with 0.25 mg ml^{-1} BSA. Beads were then incubated with 1.4 volumes of affinity purified *xl*FANCD2 antibody on a rolling bank overnight at 4°C. The antibody-coupled beads were then washed four times with five volumes of 100 mM phosphate buffer pH 8 supplemented with 0.25 mg ml^{-1} BSA, and three times with five volumes ELB buffer (10 mM HEPES-KOH pH 7.7, 50 mM KCl, 2.5 mM MgCl$_2$, 250 mM sucrose). One volume of beads was then mixed with 1.4 volumes of extract (HSS or NPE) and was incubated 30 minutes at room temperature on a rolling bank. The extract was harvested and the procedure was repeated one more time for both HSS and NPE. Extracts were then collected and used for DNA replication.

Protein purification

FLAG-tagged *xl*FANCI and STREP-tagged *xl*FANCD2 were cloned into pDONR201 (Life Technologies). The *xl*FANCD2 mutation R305W was introduced in pDONR-FANCD2 using Quickchange site-directed mutagenesis protocol. Baculoviruses were produced using the Baculodirect system, following manufacturers protocol (Life Technologies). Proteins were expressed in Sf21 insect cells by co-infection with FLAG-*xl*FANCI and STREP-*xl*FANCD2 (wild-type or mutant) viruses for 68 hours at 27°C. Cells from 750 ml of culture were collected by centrifugation, resuspended in 30 ml lysis buffer (50 mM Hepes pH8, 100 mM NaCl, 1 mM PMSF, 1 mM EDTA, 1 mM DTT, 1 tablet/10ml Complete Mini-EDTA-free Protease Inhibitor Cocktail (Roche)) and lysed by sonication. The soluble fraction obtained after centrifugation (20000 g 20 minutes 4°C) was incubated with 800 µl anti-FLAG M2 affinity gel (Sigma) that was pre-washed three times with lysis buffer. Incubation was carried at 4°C for 1 hour on a rotating wheel. Beads were collected by centrifugation and washed in lysis buffer three times. The ID2 complex was eluted in lysis buffer containing 100 µg ml^{-1} 3x FLAG peptide (Sigma). ID2^{WT} was then loaded on a HiTrap Heparin HP 1 ml column equilibrated first in MQ water and then in low salt buffer (50 mM Hepes pH 8, 100 mM NaCl, 1 mM PMSF, 1 mM EDTA, 1 mM DTT). A gradient was applied between low salt buffer and high salt buffer (50 mM Hepes pH 8, 1 M NaCl, 1 mM PMSF, 1 mM EDTA, 1 mM DTT) in 20 column volumes (CVs). Elutions 11-13 were pooled, and loaded on a gel filtration column (Superdex 200 hiload 16/600). The column was equilibrated first in MQ water and then in low salt buffer (50 mM Hepes pH8, 100 mM NaCl, 1 mM PMSF, 1 mM EDTA, 1 mM DTT). Sample was loaded and fractions containing the FANCI-FANCD2 heterodimer at 1:1 ratio were collected (55-85 ml). The complex was concentrated using Amicon Ultra centrifugal filter, 100 kDa (Merck Millipore).

The *xl*FANCD2 mutant R305W was purified in complex with *xl*FANCI, following the same procedure as for *xl*FANCD2-WT, omitting the Heparin purification step. Truncated *yeast* CAF-1 (tCAF-1) was purified as previously reported[19]. Denatured *Xenopus laevis* histones (H3^{C110A}, H4, H2A, H2B^{T112C}) were purchased (The Histone Source, Colorado University) and labelled using Alexa Fluor 647. Briefly, either lyophilized histone H3^{C110A} or H2B^{T112C} were resuspended in unfolding buffer (6 M Guanidinium HCl, 20 mM Tris pH 7.5, TCEP 0.2 mM), mixed with the dye in a 1:1 ratio and incubated overnight at 4°C on a rotating wheel. The reaction was quenched in 10 mM DTT for 1 hour at 4°C. To prepare the histone octamer, histones H3, H4 and H2A histones were resuspended in unfolding buffer and mixed with labelled H2B in a 1:1:1:1 ratio. To make H3-H4 tetramers, labelled H3 was mixed with histone H4 in a 1:1 ratio. To make H2A-H2B dimer, labelled H2B was mixed with H2A in a 1:1 ratio. Histones were then dialyzed overnight at 4°C against refolding buffer (2 M NaCl, 10 mM Tris pH 7.5, 1 mM EDTA, 0.2 mM TCEP). Free histones and free dye were removed by gel filtration (Superdex 200 column) and concentrated using 10 kDa Amicon filter (Merck Millipore).

NAQ assay

The NAQ assay was performed ad described previously[20]. Briefly, 200 nM tCAF-1 or *xl*FANCD2 (in complex with *xl*FANCI) were mixed with 100 nM histone octamer in NA buffer (25 mM Tris pH7.5, 150 mM NaCl, 1 mM EDTA, 0.02% Tween-20, 1 mM TCEP) for 15 minutes at room temperature. After the incubation, 150 nM 207 bp oligo DNA was added and the reaction was incubated for 30 minutes at room temperature. Part of the assembly reaction was loaded on a 6% native PAGE and the gel was stained with Sybr Gold staining (Fisher). The gel was imaged with a Typhoon imager (Amersham) using Cy2/488 to visualize DNA and A647 to visualize histones. The remaining of the assembly reaction was digested with 20 U MNase (New England Biolabs) in MNase buffer supplemented with BSA (New England Biolabs). The reaction was incubated 10 minutes at 37°C and then quenched with EDTA (50 mM final concentration). Loading control DNA (50 ng) was added and DNA was purified using the MinElute kit (QIAGEN). The purified DNA was run on a 6% native PAGE gel. The gel was stained by Sybr Gold staining (Fisher) and it was imaged with a Typhoon imager (Amersham) using Cy2/488.

To reconstitute nucleosomes by salt assembly, the histone octamer was mixed with DNA (either 207 bp fragment or 147 bp fragment) in reaction buffer (25 mM Tris pH 7.5, 1 mM EDTA, 0.5 mM TCEP) supplemented with 2 M NaCl, at various DNA:octamer ratios (1:1,4; 1:1,6; 1:1,8). The salt concentration was then gradually reduced to 150 mM NaCl by diluting the assembly reaction in reaction buffer five times, and incubating the reaction 30 minutes at 37°C after each dilution. The assembled nucleosomes were then analyzed on a PAGE gel.

Nucleosome positioning assay

For the nucleosome positioning assay 3 µl aliquots of the replication/repair reactions were diluted ten times in MNase buffer (NEB, 10X MNase buffer) supplemented with BSA 0.1 mg µl^{-1}. This was followed by addition of 600 U MNase (NEB) and incubation for 5 minutes at 37°C. MNase digestion was stopped adding 14 volumes of MNase stop buffer: 20 mM Tris-HCl pH 7.5, 0.15 M NaCl, 20 mM EDTA, 20 mM EGTA, 0.5% NP-40 (Sigma-Aldrich), 5 µg ml^{-1} aprotinin (Sigma-Aldrich), 5 µg ml^{-1} leupeptin (Sigma- Aldrich), 2 mM PMSF (Sigma-Aldrich). Samples were treated with RNase and proteinase K and purified by phenol/chloroform extraction. DNA was digested with BBS1 and analyzed by qPCR using the following primers:
ICL forward: 5'-AGCCAGATTTTTCCTCCTC-3';
ICL reverse: 5'-CATGCATTGGTTCTGCACTT-3';
MID forward: 5'-ACCCTGGGTTCTTTTCCAAC-3';
MID reverse: 5'- CATTTCATCTGGAGCGTCCT-3';
MID2 forward: 5'-GCAACGTGCTGGTTATTGTG-3';
MID2 reverse: 5'-GTGATCATGCGTTTGCGTTG-3';
FAR forward: 5'-AACGCCAATAGGGACTTTCC-3';
FAR reverse: 5'-GGGCGTACTTGGCATATGAT-3';
pQuant forward: 5'-TACAAATGTACGGCCAGCAA-3';
pQuant reverse: 5'-GAGTATGAGGGAAGCGGTGA-3'.
Amplicon abundance was normalized against pQuant, and for each amplicon the peak value was set at 1.

Chromatin immunoprecipitation

Chromatin immunoprecipitations were performed as previously reported[27]. In brief, replication reactions were crosslinked with formaldehyde and DNA was fragmented by sonication. DNA-bound FANCD2 was immunoprecipitated using Protein A Sepharose beads (Abcam) coupled to purified αFANCD2 antibodies. Protein-DNA crosslinks were reversed and DNA was purified and analyzed by quantitative PCR using the indicated primers.

Supercoiling assay

Analysis of nucleosome assembly on a relaxed circular plasmid was adapted from Sato et al[15]. DNA (pSVRLuc) was relaxed by topoisomeraseI (New England Biolabs) treatment for 2 hours at 37°C. The histone octamer (200 nM) was incubated with increasing concentrations of either tCAF-1 or *xl*ID2 (300nM, 600nM, 1µM *xl*FANCD2) at 37°C for 15 minutes in NA buffer (25 mM Tris pH 7.5, 150 mM NaCl, 1 mM EDTA, 0.02% Tween-20, 1 mM TCEP) supplemented with 2mg ml^{-1} BSA. After the incubation, 180 ng relaxed DNA was added to the chaperone/histone mixture and the reaction was incubated for 2 hours at 37°C. This was followed by proteinase K treatment and DNA purification by phenol/chloroform extraction. The extracted DNA was loaded on a 1% agarose gel and the gel was run overnight at 80V at
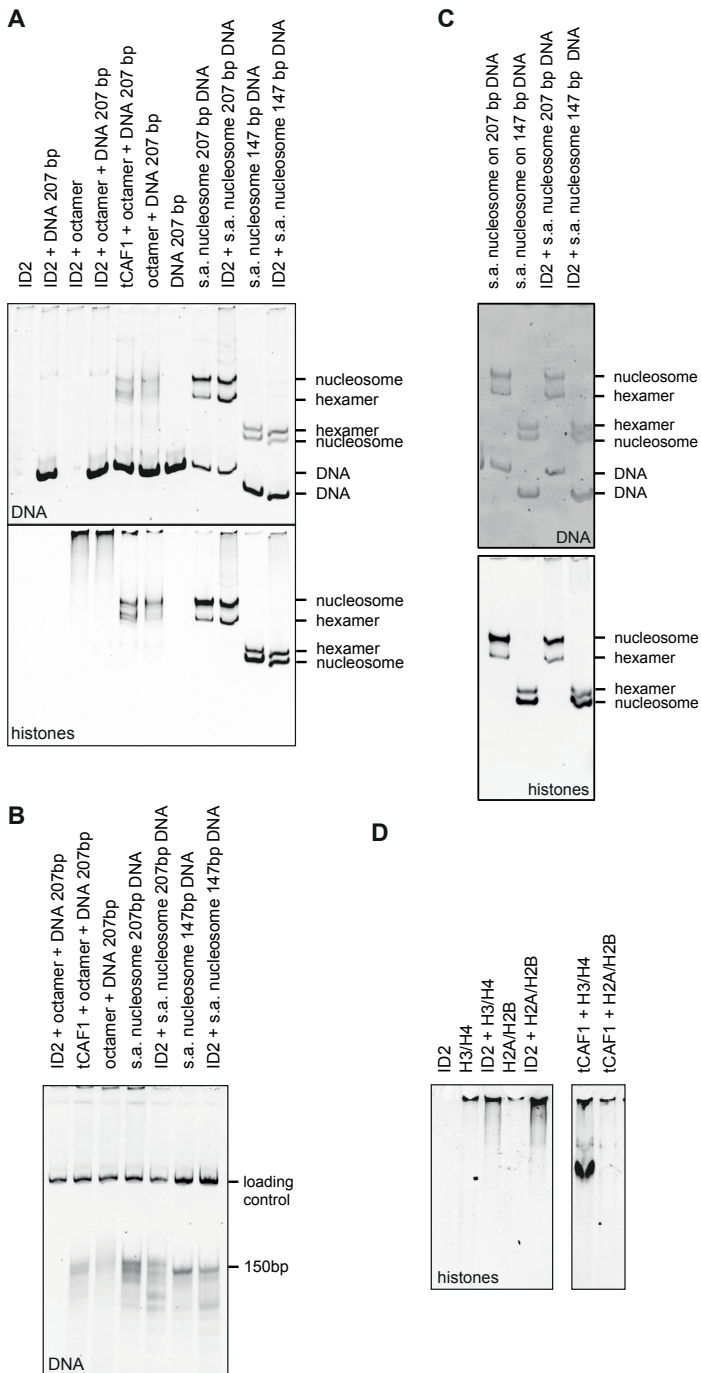
4°C. The gel was stained with Sybr Gold and DNA was visualized using a Typhoon imager (Amersham).

Acknowledgements

4

Supplementary Figures

A

B

C

D

Figure S1: FANCD2 interacts with histones but cannot assemble nor interact with nucleosomes.

A) ID2 complex was incubated with histones and DNA, samples were analyzed on PAGE gel as in Figure 1F. DNA was visualized by Sybr Gold staining and histones were visualized by the Alexa Fluor 647 tag. tCAF-1 was used as positive control. Independent experimental replicate of Figure 1F. 's.a. nucleosome', salt assembled nucleosome. **B)** Samples generated in (A) were digested with MNase and the DNA fragments were analyzed on a PAGE gel stained with Sybr Gold. Independent experimental replicate of Figure 1G. **C)** Nucleosomes were reconstituted by salt assembly and were incubated with recombinant ID2 complex. Samples were analyzed by PAGE gel. DNA was visualized by Sybr Gold staining and histones were visualized through the Alexa Fluor 647 tag. Independent experimental replicate of (A). **D)** Histones and recombinant ID2 were incubated and samples were analysed by PAGE gel. Histones were visualized through the Alexa Fluor 647 tag. tCAF-1 was used as positive control. Independent experimental replicate of A and Figure 1F (lower panel).

4

A

B

ΔFANCD2

C
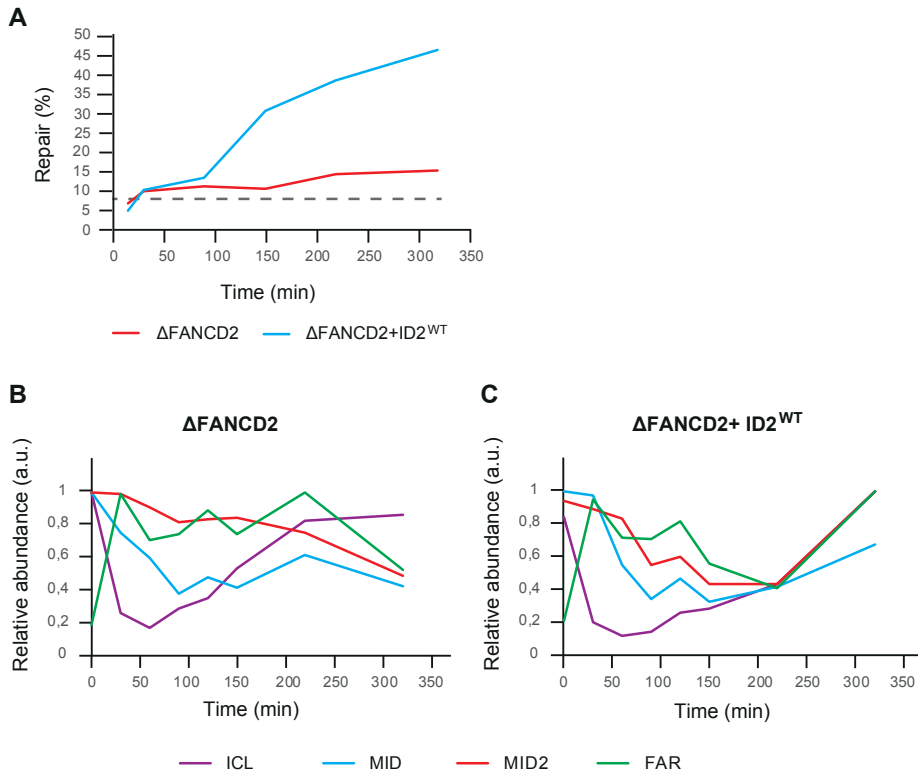
ΔFANCD2+ ID2^{WT}

Figure S2: FANCD2 depletion impairs repair-coupled nucleosome dynamics
A) pICL-Pt was replicated in either extracts depleted of FANCD2 (ΔFANCD2) or extracts depleted of FANCD2 and complemented with recombinant ID2 complex (ΔFANCD2+ID2^{WT}). An unrelated undamaged plasmid, pQuant, was replicated in the same reactions for quantification purposes. The efficiency of repair was calculated (see methods) and plotted over time. Dotted line represents background SAPI regeneration due to an uncrosslinked portion of pICL-Pt. Independent experimental replicate of Figure 2B. **B-C)** Intermediates generated in the repair reactions in A to which no ^{32}P-α-dCTP was added, were digested with MNase. The resulting DNA fragments were isolated and mapped to the plasmid by qPCR, using the primer sets ICL, MID, MID2 and FAR. The abundance of each amplicon was normalised to pQuant and the peak value was set to 1. Values were plotted against reaction time. Independent experimental replicate of Figure 2E and F.
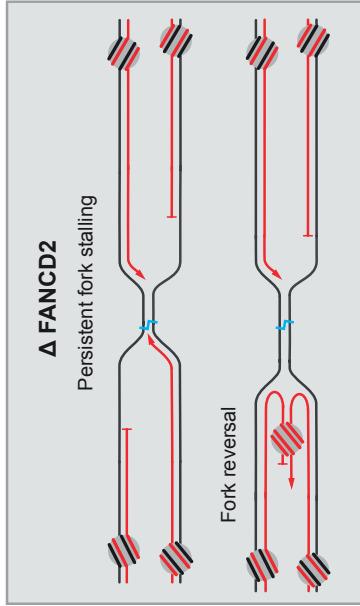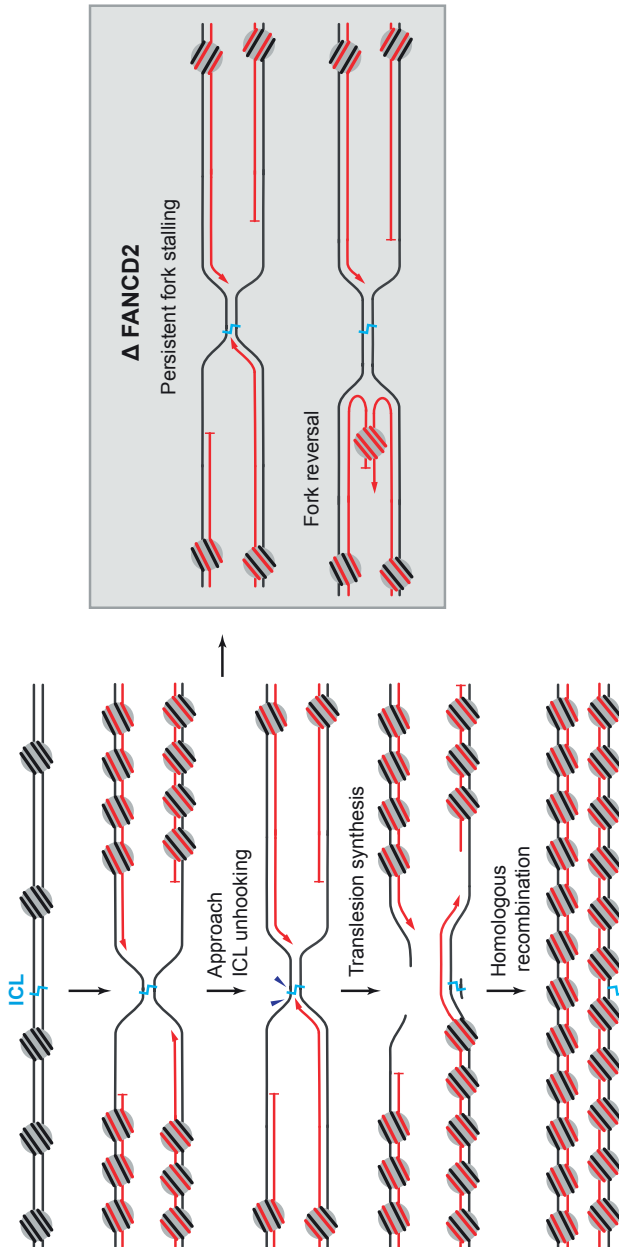
Figure S3: schematic model for nucleosome dynamics during ICL repair

During ICL repair nucleosomes are dynamically removed from the region surrounding the ICL up to approximately 1.5kb from the lesion (see also chapter 3). In the late stages of repair nucleosomes are restored in a mechanism coupled to ICL repair. In the absence of FANCD2 (ΔFANCD2, right panel), replication forks persistently stall at the ICL and some may undergo fork reversal. Nucleosomes may be positioned on the reversed fork to avoid resection and facilitate fork restart.
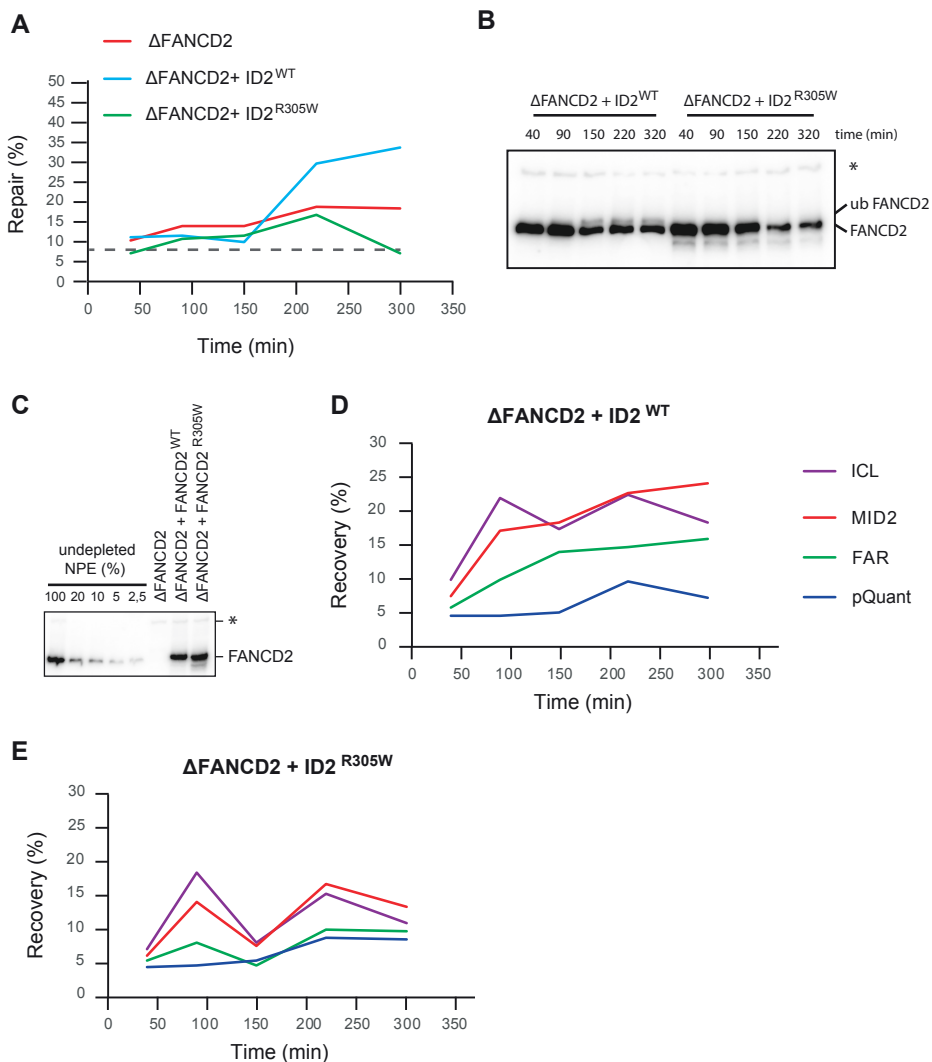
Figure S4: *xl*FANCD2R305W does not support ICL repair and is not efficiently recruited to the crosslink.

A) pICL-Pt and pQuant were replicated in extracts depleted of FANCD2 and supplemented with either ID2WT or IDR305W. Repair was analyzed by quantification of the regeneration of the SAPI restriction site and plotted. Dotted line represents background level of repair due to an uncrosslinked portion of pICL-Pt. Intermediates from this repair reaction were analyzed by ChIP in Figure 4. **B)** Reaction samples were analysed by Western blot using αFANCD2 antibody to investigate FANCD2 ubiquitination during repair. **C)** FANCD2 depleted NPE (ΔFANCD2) and FANCD2 depleted NPE complemented with ID2WT or ID2R305W were analyzed by Western blot using αFANCD2 antibody. A dilution series of undepleted NPE was loaded on the same blot to determine the level of depletion. Relates to D and E. **D-E)** FANCD2 depleted extracts (ΔFANCD2), FANCD2 depleted extracts supplemented with either wild type ID2 (ΔFANCD2+ ID2WT) or FANCD2R305W-FANCI complex (ΔFANCD2+ID2R305W) were used to replicate pICL/pQuant. Reaction samples were isolated and analyzed by FANCD2 Chip using ICL, MID2, FAR and pQuant primer sets. Independent experimental replicate of Figure 4.

References

1. de Winter, J.P. & Joenje, H. The genetic and molecular basis of Fanconi anemia. *Mutat Res* **668**, 11-9 (2009).

2. Nalepa, G. & Clapp, D.W. Fanconi anaemia and cancer: an intricate relationship. *Nat Rev Cancer* **18**, 168-185 (2018).

3. Kottemann, M.C. & Smogorzewska, A. Fanconi anaemia and the repair of Watson and Crick DNA crosslinks. *Nature* **493**, 356-63 (2013).

4. Duxin, J.P. & Walter, J.C. What is the DNA repair defect underlying Fanconi anemia? *Curr Opin Cell Biol* **37**, 49-60 (2015).

5. Knipscheer, P. et al. The Fanconi Anemia Pathway Promotes Replication-Dependent DNA Interstrand Cross-Link Repair. *Science* **326**, 1689-1701 (2009).

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

7. Zhang, J. et al. DNA interstrand cross-link repair requires replication-fork convergence. *Nat Struct Mol Biol* **22**, 242-7 (2015).

8. Klein Douwel, D. et al. XPF-ERCC1 acts in Unhooking DNA interstrand crosslinks in cooperation with FANCD2 and FANCP/SLX4. *Mol Cell* **54**, 460-71 (2014).

9. Amunugama, R. et al. Replication Fork Reversal during DNA Interstrand Crosslink Repair Requires CMG Unloading. *Cell Rep* **23**, 3419-3428 (2018).

10. Lossaint, G. et al. FANCD2 binds MCM proteins and controls replisome function upon activation of s phase checkpoint signaling. *Mol Cell* **51**, 678-90 (2013).

11. Kais, Z. et al. FANCD2 Maintains Fork Stability in BRCA1/2-Deficient Tumors and Promotes Alternative End-Joining DNA Repair. *Cell Rep* **15**, 2488-99 (2016).

12. Schlacher, K., Wu, H. & Jasin, M. A distinct replication fork protection pathway connects Fanconi anemia tumor suppressors to RAD51-BRCA1/2. *Cancer Cell* **22**, 106-16 (2012).

13. Madireddy, A. et al. FANCD2 Facilitates Replication through Common Fragile Sites. *Mol Cell* **64**, 388-404 (2016).

14. Chan, K.L., Palmai-Pallag, T., Ying, S. & Hickson, I.D. Replication stress induces sister-chromatid bridging at fragile site loci in mitosis. *Nat Cell Biol* **11**, 753-60 (2009).

15. Sato, K. et al. Histone chaperone activity of Fanconi anemia proteins, FANCD2 and FANCI, is required for DNA crosslink repair. *EMBO J* **31**, 3524-36 (2012).

16. Higgs, M.R. et al. Histone Methylation by SETD1A Protects Nascent DNA through the Nucleosome Chaperone Activity of FANCD2. *Mol Cell* **71**, 25-41 e6 (2018).

17. Paquin, K.L. et al. FANCD2 binding to H4K20me2 via a methyl-binding domain is essential for efficient DNA crosslink repair. *Mol Cell Biol* (2019).

18. Timmers, C. et al. Positional cloning of a novel Fanconi anemia gene, FANCD2. *Mol Cell* **7**, 241-248 (2001).

19. Mattiroli, F. et al. DNA-mediated association of two histone-bound complexes of yeast Chromatin Assembly Factor-1 (CAF-1) drives tetrasome assembly in the wake of DNA replication. *Elife* **6**(2017).

20. Mattiroli, F., Gu, Y. & Luger, K. Measuring Nucleosome Assembly Activity in vitro with the Nucleosome Assembly and Quantification (NAQ) Assay. *Bio Protoc* **8**(2018).

4

21. Yuan, F., El Hokayem, J., Zhou, W. & Zhang, Y. FANCI protein binds to DNA and interacts with FANCD2 to recognize branched structures. *J Biol Chem* **284**, 24443-52 (2009).

22. Schmid, J.A. et al. Histone Ubiquitination by the DNA Damage Response Is Required for Efficient DNA Replication in Unperturbed S Phase. *Mol Cell* **71**, 897-910 e8 (2018).

23. Berti, M. et al. Human RECQ1 promotes restart of replication forks reversed by DNA topoisomerase I inhibition. *Nat Struct Mol Biol* **20**, 347-54 (2013).

24. Lebofsky, R., Takahashi, T. & Walter, J.C. DNA replication in nucleus-free Xenopus egg extracts. *Methods Mol Biol* **521**, 229-52 (2009).

25. Enoiu, M., Jiricny, J. & Scharer, O.D. Repair of cisplatin-induced DNA interstrand crosslinks by a replication-independent pathway involving transcription-coupled repair and translesion synthesis. *Nucleic Acids Res* **40**, 8953-64 (2012).

26. Knipscheer, P., Raschle, M., Scharer, O.D. & Walter, J.C. Replication-coupled DNA interstrand cross-link repair in Xenopus egg extracts. *Methods Mol Biol* **920**, 221-43 (2012).

27. Pacek, M., Tutter, A.V., Kubota, Y., Takisawa, H. & Walter, J.C. Localization of MCM2-7, Cdc45, and GINS to the site of DNA unwinding during eukaryotic DNA replication. *Mol Cell* **21**, 581-7 (2006).
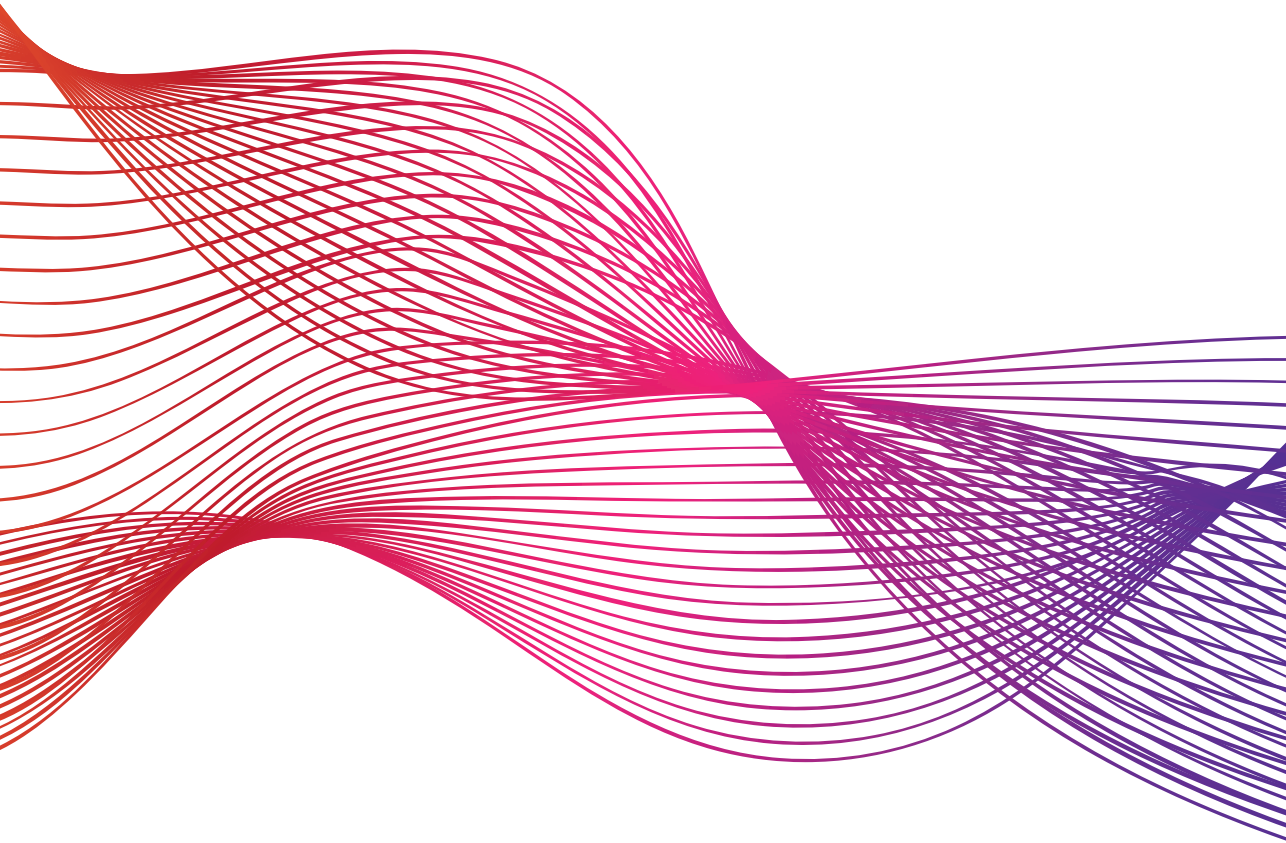
Does FANCD2 act as a histone chaperone during DNA interstrand crosslink repair?
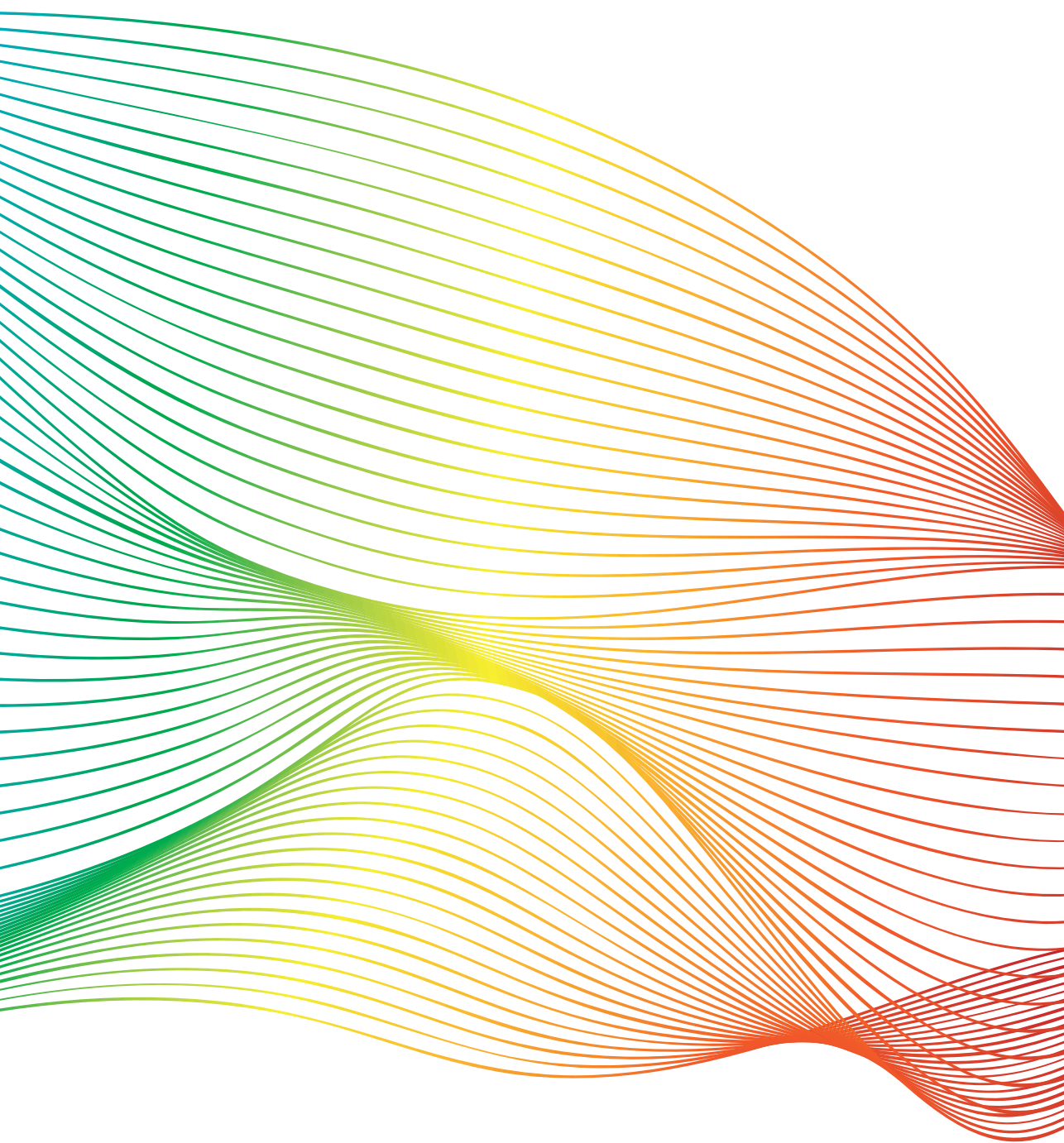
Chapter 5

Discussion

DNA interstrand crosslinks (ICLs) are highly toxic DNA lesions that, by covalently binding the two strands of the DNA double helix, inhibit DNA replication and transcription. ICLs can be induced by chemotherapeutics, such as cisplatin, or by endogenous metabolites, such as reactive aldehydes and nitric oxide[1,2].

Repair of these lesions relies mainly on the DNA replication-coupled Fanconi anemia (FA) pathway.

This pathway involves three major repair steps: the unhooking of the ICL through nucleolytic incisions in the DNA backbone, the bypass of the unhooked crosslink by translesion synthesis, and finally the repair of the incised DNA strand by homologous recombination (HR)[3]. Recent studies demonstrated that specific ICLs can also be repaired during DNA replication by FA-independent routes[4,5] (Figure 1, see also **chapter 2**). However, many mechanistic details of these novel pathways are still lacking. Moreover, to fully dissect the mechanism of ICL repair, it is also of interest to examine the context in which this process occurs. While there is increasing evidence that nucleosome dynamics around the lesion are important for DNA repair pathways, the role of the nucleosomal context during ICL repair is still largely unexplored. In this chapter, I will summarize the main findings of this thesis, discuss them in the context of knowledge in the field, and suggest how future experiments may integrate these in the mechanisms of ICL repair.
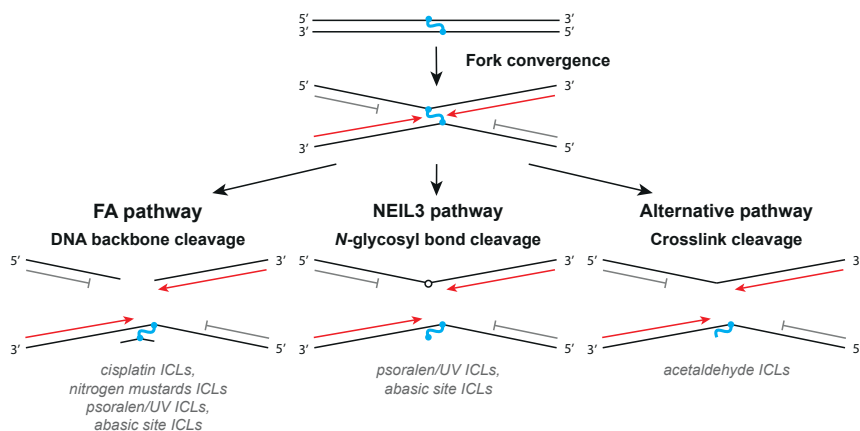
Figure 1: Mechanisms of ICL unhooking.
Schematic representation of ICL unhooking in the various pathways of ICL repair. In the FA pathway, the crosslink is unhooked by dual incisions in the phosphodiester DNA backbone, resulting in a double-strand break in one strand and an ICL-containing adduct in the other strand. The NEIL3 pathway cleaves one of the two glycosyl bonds that form the ICL, resulting in an abasic site (white circle) in one strand and an adduct in the other DNA strand. Acetaldehyde-ICLs are repaired by an alternative route, likely relying on direct reversal. In this mechanism, the cleavage occurs within the ICL and generates a guanine on one strand and a propanoguanine on the other strand.

Multiple routes of repair act on chemically distinct ICLs

Can the structure of an ICL affect the choice of repair pathway?
For many years it was assumed that the only replication-coupled mechanism of ICL repair was the FA pathway. This notion was recently challenged by the finding that both ICLs induced by Psoralen/UV treatment and those forming at abasic sites can be unhooked by the glycosylase NEIL3[4,5] (Figure 1). Moreover, in **chapter 2** we identified yet another pathway of ICL repair acting on acetaldehyde-induced ICLs. In this pathway, the ICL is unhooked through a cut within the crosslink itself, and we speculate that this occurs by a direct reversal mechanism[6] (Figure 1).

Since these alternative routes seem to act on specific ICLs, the chemical structure of the crosslink could determine the repair mechanism. Crosslinks can differ substantially both in terms of the sequence context in which they form and in the distortion that they introduce in the DNA helix. Cisplatin-induced ICLs form between two guanines in a CpG context and cause distortion and bending of the double helix, resulting in the protrusion of the bases next to the crosslinked bases. Psoralen-induced ICLs instead form preferentially between thymines in 5′-TA/AT-3′ sequences, and they induce helix distortion but no bending[7]. The endogenous acetaldehyde-ICLs do not distort the double helix, and form between guanines in a CpG context[8]. All these structural differences may dictate not only how ICLs are recognized and repaired, but also their mutagenic potential. The FA pathway seems to be able to repair a wide range of chemically distinct ICLs, such as the ones induced by nitrogen mustards, MMC, cisplatin, psoralen/UV, abasic sites and acetaldehyde[3-6,9]. How the FA pathway proteins recognize and unhook such a wide range of ICLs is still unclear. It is possible that the FA pathway does not directly recognize the ICL itself, but rather the consequences of the crosslink. These could be the stalling of the two replication forks at the level of the ICL or the presence of stalled replication forks still coupled to the CMG helicase. Alternatively, ICL unhooking in the FA pathway could be carried out by multiple nucleases. To date, only the nuclease XPF has been identified as an FA protein, however, experiments in cells showed that depletion of other nucleases such as MUS81-EME1, FAN1, SLX1, SNM1A and B can also cause sensitivity to ICL inducing agents[10-14]. While previous work from our lab has excluded a role for MUS81, SLX1 and FAN1 in cisplatin ICL repair[15], it is possible that these nucleases may unhook different types of crosslinks. It is therefore important to test the role of these nucleases in the repair of chemically diverse ICLs.

While the FA pathway acts on a wide range of ICLs, this repair mechanism involves the formation of a double strand break (DSB)[3,16], which can result in chromosomal rearrangements and aberrations. The formation of chromosomal breakages and rearrangements upon exposure to ICL-inducing agents is a specific feature of FA cells and is used to diagnose the disease. Limiting genomic instability by avoiding the formation of these DSBs could therefore be a role for the alternative

5

routes of ICL repair. Moreover, in the FA pathway the unhooked ICL is bypassed by translesion synthesis, which is a mutagenic process[17]. As the genetic information in this newly synthesized strand is then used for homology-mediated repair of the DSB, the FA pathway can introduce mutations in both sister chromatids. The alternative route of acetaldehyde-ICL repair results in fewer point mutations compared to the FA pathway, as it introduces mutations only in one of the newly synthesized strands (see **chapter 2**). These considerations suggest that ICL repair by the FA pathway may come at the cost of mutagenicity and genomic instability. Accordingly, cells lacking a functional FA pathway have decreased point mutations compared to wild type cells[18]. Future studies should investigate how both the ICL structure and the repair route affect mutagenicity and genomic stability.

What is the mechanism of the alternative route of acetaldehyde-ICL repair? In **chapter 2** we show that the FA-independent route of acetaldehyde-ICL repair takes place during replication and involves ICL unhooking through a cut within the ICL, possibly forming an undamaged guanine and an adducted guanine[6] (Figure 1). To determine the mechanism of the alternative route of acetaldehyde-ICL repair, further investigation of the ICL unhooking step is necessary. We demonstrated that the crosslink is not broken by the force generated by the collision between one replication fork and the ICL (see **chapter 2**), and even though we cannot formally exclude that the ICL is broken by the force generated by two converged replication forks, we favor a model in which ICL unhooking is enzymatic. Potential candidates for this reaction are the dealkylating enzymes of the AlkB family, which can remove a wide range of alkyl groups from DNA bases. *In vitro* studies showed that *E. coli* AlkB is able to remove alkyl groups attached to the *N2* atom of guanine[19]. Since the *N2* atom of guanine is also involved in the acetaldehyde-ICL, it is possible that the members of the vertebrate ortholog protein family ALKBH could unhook acetaldehyde-induced ICLs through oxidative dealkylation. Current work in the laboratory is focused on investigating a potential role for the ALKBH protein family in acetaldehyde-ICL repair. Moreover, two unbiased approaches to identify the enzyme(s) involved in acetaldehyde-ICL unhooking/repair are currently undertaken. Firstly, during acetaldehyde-ICL repair reactions, ICL-bound proteins are analyzed using plasmid pulldowns coupled to mass spectrometry (see methods section of **chapter 3**). Secondly, wild-type or FA deficient human cells are screened for factors that induce acetaldehyde sensitivity.

A second aspect of the alternative ICL repair route that needs further clarification is how pathway choice is regulated. In psoralen/UV-induced ICLs, the regulation of pathway choice is dependent on the ubiquitination of the CMG helicase by the E3 ligase TRAIP[5]. The glycosylase NEIL3 can bind to short ubiquitin chains on the CMG and initiate the glycosylase route of ICL repair[5]. If NEIL3 cannot process the crosslink, long ubiquitin chains will form on the CMG helicase, promoting CMG unloading and repair by the FA pathway[5]. However, disruption of ubiquitin signaling

during acetaldehyde-ICL repair does not affect unhooking in the alternative pathway (data not shown), indicating that CMG ubiquitination is not involved in pathway choice for this route. To understand how pathway choice is regulated during acetaldehyde-ICL repair, the mechanism of initiation of the alternative route should first be defined. Given the observation that repair is quickly initiated by the alternative pathway, it is likely that the unhooking enzyme travels with the replication fork. However, the arrival of a single replication fork at the lesion site does not initiate ICL repair, indicating that additional steps are necessary for the initiation of the alternative pathway. It is possible that through double fork convergence at the ICL, two copies of the unhooking enzyme are brought to the ICL to initiate ICL unhooking. Alternatively, fork convergence may promote ICL unhooking by inducing activation of the unhooking enzyme through post-translational modifications (PMT), or by changing the DNA structure around the ICL.

During acetaldehyde-ICL repair in *Xenopus* egg extracts, a part of the ICLs is repaired by the alternative route, while the other part is repaired by the FA pathway. This could be caused by multiple factors. Lack of timely initiation of the alternative acetaldehyde-ICL repair route could lead to FA pathway activation, just like for the NEIL3 pathway[5]. A second possibility is that the unhooking enzyme in the alternative pathway may inhibit itself after ICL unhooking, as it has been observed for proteins involved in direct reversal[20]. This auto-inhibitory effect would then limit the concentration of the unhooking enzyme and result in FA pathway activation in conditions where the concentration of ICLs in DNA is too high. To test these hypotheses, it is essential to identify what factor(s) are involved in ICL unhooking in the alternative route. Lastly, the structure of the crosslink itself may drive pathway choice. As acetaldehyde-induced ICLs are present in an equilibrium between three forms (see **chapter 2**), the chemical structure may dictate whether an ICL is repaired by the alternative pathway or by the FA pathway. Currently, efforts are being made in the lab to isolate the three different crosslink forms to test this hypothesis.

Why do patients develop Fanconi anemia if endogenous ICLs can be repaired by multiple routes?

The development of disease in the absence of a functional FA pathway seems in contradiction with the idea that endogenous ICLs can be repaired by multiple pathways. However, there are various explanations for this. First, not all endogenous ICLs could be the right templates for the alternative repair routes, as there is a multitude of sources of endogenous crosslinks[1,21-23]. Factors like the structure of the ICL, the DNA surrounding the lesion, and the local chromatin environment could limit the alternative routes of ICL repair. Secondly, it is possible that in FA patients the disruption of the FA pathway occurs at a stage in the repair process where pathway switching is no longer possible. In line with this, depletion of FANCI-FANCD2 during the repair of psoralen-induced ICLs did not result in complete

switch to the NEIL3 pathway[4], indicating that pathway switching cannot occur at any stage of ICL repair.

How are other aldehyde-induced ICLs repaired?
In addition to acetaldehyde, other reactive aldehydes such as formaldehyde and acrolein can induce crosslinks[24,25] and therefore contribute to the development of FA. However, how these endogenous ICLs are repaired is still unknown. Formaldehyde is an endogenous aldehyde present abundantly in the body and formed during oxidative demethylation of DNA, RNA and proteins[26]. Genetic experiments linked formaldehyde to FA[27], suggesting that formaldehyde-induced ICLs could be repaired by the FA pathway. Importantly, formaldehyde can form crosslinks between structurally dissimilar DNA bases (either two adenines or two guanines), raising the possibility that crosslinks induced by the same compound may be repaired through different pathways[24,28]. Given the abundance of formaldehyde in the body, it would be of major interest to address the mechanism of formaldehyde-ICL repair in the future.

Chromatin remodeling in ICL repair

What is the role of nucleosomal rearrangements during ICL repair?
Chromatin plays an important role in the DNA damage response to various DNA lesions, as DSBs and DNA adducts induced by UV light. In **chapter 3** we studied the nucleosomal rearrangements that occur during ICL repair and we found that nucleosomes are sequentially displaced and replaced in the region surrounding the crosslink. These dynamics may favor the recruitment of repair proteins and promote DNA damage signaling. While our work provides the first direct evidence of nucleosomal rearrangements during ICL repair, the spatial resolution of nucleosomes mapping in our experiments was limited by the use of quantitative PCR. To obtain a more detailed view of nucleosomal dynamics over time, a new method for mapping nucleosomes using high-throughput DNA sequencing is currently being developed in the laboratory. This methodology, coupled to inhibition of ICL repair at various stages, will provide information on how changes in nucleosome localization are associated with the various repair steps.

Is FANCD2 monoubiquitination involved in nucleosomal dynamics in the FA pathway?
Although FANCD2 ubiquitination has been considered to be the key activating step in the FA pathway for decades, the direct function of the ubiquitinated FANCI-FANCD2 complex (ID-ub) is still not clear. Our laboratory previously demonstrated that ubiquitinated FANCD2 promotes ICL unhooking[15], however, current data indicates that this is not mediated by a direct interaction between ubiquitinated

FANCD2 and the nuclease complex SLX4-XPF-ERCC1 (unpublished data). It is therefore still unclear how monoubiquitination of FANCD2 leads to the recruitment of the unhooking nucleases. Both structural and biochemical studies suggest that ubiquitination of the FANCI-FANCD2 complex leads to conformational changes that favor the retention of the complex onto DNA[29-31]. However, how the retention of ID2-ub on the DNA could promote ICL unhooking is still unknown. *In vitro* and cellular studies proposed that FANCD2 may induce nucleosomal rearrangements that could regulate nucleolytic cleavage during ICL unhooking[32]. The authors suggested that the histone chaperone activity of FANCD2 could assemble nucleosomes on one of the two parental strands, thereby limiting DNA incisions to one side only of the ICL[32]. However, in **chapter 4** we tested the nucleosome (dis)assembly function of FANCD2 and we found no clear histone chaperone activity, indicating that FANCD2 likely does not regulate ICL unhooking through nucleosome (dis)assembly. Another possible explanation for the role of monoubiquitinated FANCD2 in the promotion of ICL unhooking is that FANCD2 could bind to an unknown factor, which in turn recruits the nucleases to the ICL. Studies based on mass spectrometry can be used to find which proteins bind both ubiquitinated FANCD2 and SLX4 and therefore identify putative bridging factor(s). Previous work from our lab showed that during ICL repair multiple FANCD2 molecules are loaded on the plasmid at the ICL locus and they spread away from the crosslink[15]. The function of this spreading is still unclear, but it could be involved in the regulation of SLX4-XPF-ERCC1 recruitment.

How to identify chromatin remodeling factors acting in ICL repair?
While studies on nucleotide excision repair and double strand break repair identified multiple chromatin remodeling factors involved in the DNA damage response, little is known about the role of these proteins in ICL repair. To identify factors involved in nucleosomal dynamics during ICL repair in *Xenopus* egg extracts, we used an unbiased approach based on mass spectrometry (**chapter 3).** This method allows the identification of factors that bind ICL-containing plasmids during repair, circumventing the use of ICL-inducing agents that can create many other types of DNA lesions. Moreover, since ICL repair in *Xenopus* egg extracts occurs relatively synchronous, this method can identify proteins involved in specific repair stages. We found various proteins with known roles in nucleosomal dynamics, such as ATP-dependent chromatin remodeling factors, histone chaperones, histone methyltransferases, acetyltransferases and E3 ubiquitin ligases. To understand the role of chromatin dynamics in ICL repair, these factors will be depleted from *Xenopus* egg extracts either by immunodepletion or by using a novel method recently developed in the Knipscheer laboratory. This method is inspired by the Trim away method[33], that relies on the degradation of antibody-bound proteins by the ubiquitin E3 ligase Trim21. Addition of recombinant Trim21, together with the antibody of choice to *Xenopus* egg extracts, results in the efficient depletion of the protein to which the antibody was raised. This method allows depletion of proteins

from extracts after the start of replication, providing the possibility to study factors that act both in DNA replication and repair.

Furthermore, to characterize the changes in chromatin structure during ICL repair, future studies in our laboratory will aim to identify repair-associated histone modifications, such as phosphorylation, methylation, acetylation, ubiquitination and sumoylation. To study how histone PTMs are associated with ICL repair, plasmid pulldowns will be coupled to histone PTM analysis either by western blotting or mass spectrometry. Moreover, since histone modifications play a role in pathway choice for DSB repair[34,35], future experiments will compare histone PTM in the different pathways of ICL repair. These studies will allow us to define mechanistically how changes in chromatin structure support ICL repair and whether chromatin modifications are involved in ICL repair pathway choice.

The understanding of the mechanism of nucleosome dynamics and histone PTMs during ICL repair will provide novel insights into this repair pathway and could help the development of new strategies to improve chemotherapy (see paragraph 'ICL repair and human health').

ICL repair and human health

How can a detailed understanding of ICL repair improve human health?
Most cellular acetaldehyde and formaldehyde is metabolized by two enzymes: ADH5, which removes mainly formaldehyde, and ALDH2, which can clear both formaldehyde and acetaldehyde[36]. Mice deficient in both the metabolism of aldehydes and in the FA pathway, display various symptoms associated with FA, such as bone marrow failure (BMF), leukemia and accumulation of DNA damage in hematopoietic stem cells[1,27,36,37]. In agreement with a role for aldehydes in the pathogenesis of FA, a cohort of Japanese FA patients who carried a mutation in ALDH2 displayed accelerated BMF progression[38]. Importantly, deficiency in ALDH2 is very common in the East Asian population and it results in increased cancer predisposition due to exposure to high levels of aldehydes[39]. Therefore, monitoring exposure to sources of aldehydes (e.g. alcohol) can impact the development and severity of both FA and its related cancers. Since acetaldehyde is both a product of carbohydrates and alcohol metabolism, limiting alcohol consumption can be an effective strategy to reduce the incidence of cancer in the ALDH2-deficient population. However, BMF in most FA patients develops during childhood, and it is therefore not likely to be caused by alcohol intake. Novel approaches for the treatment of FA aim to reduce the endogenous levels of aldehydes by using the aldehyde scavenging drug metformin, which is commonly used for the treatment of type 2 diabetes[40]. Metformin treatment enhanced hematopoiesis in *Fancd2*[-/-] mice and delayed tumor formation in *Fancd2*[-/-] *p53*[-/-] mice[41], indicating that this treatment could counteract the toxic effects of endogenous aldehydes in FA patients. A

second approach that is currently investigated to reduce aldehyde-induced DNA damage is to increase the activity of ALDH2. Studies in rodents identified a small molecule chaperone, Alda-1, which binds ALDH2 and stimulates its dehydrogenase activity[42-44]. *In vitro* experiments demonstrated that Alda-1 increases 2.1 fold the activity of wild-type ALDH2 and up to 11 fold the activity of the ALDH2 variant common in the South East Asian population[42]. When this molecule was tested in *Fancd2*[-/-] mice carrying a dominant negative point mutation in *Aldh2*, it resulted in the restoration of the hematopoietic functions. Alda-1 is therefore a promising candidate for the treatment of FA[43]. The discovery of a second repair pathway acting on endogenous ICLs (see **chapter 2**) also opens up opportunities for therapeutic applications, as the stimulation of the pathway could potentially alleviate some of the symptoms of FA.

While ICL-inducing agents used in chemotherapy cause a wide range of DNA lesions, ICLs are considered the main contributors to cancer cell death[45-47]. Resistance to these therapies is unfortunately common, constituting an obstacle to cancer treatment. Cells acquire resistance to ICL inducing agents by several mechanisms, including by upregulation of repair genes. For example, a non-small cell lung cancer (NSCLC) cell line resistant to cisplatin (A549/DR) was reported to have increased expression of FA genes and increased levels of FANCD2 monoubiquitination compared to its parental cell line[48]. Importantly, targeting FA genes re-sensitized these cells to cisplatin. These findings support the idea that inhibition of FA proteins could be an efficacious strategy to counteract the development of resistance to chemotherapy based on crosslinking agents. Proteins in the FA pathway could also be targeted to induce synthetic lethality. Synthetic lethality is defined as the disruption of two gene products that leads to cell death, and it is an attractive approach to increase the efficacy of chemotherapy. A clinically successful example of synthetic lethality is the use of inhibitors of PARP (Poly(ADP-ribose) polymerase) for the treatment of BRCA1/2-deficient advanced ovarian and breast cancers[49]. Inhibition of PARP results in accumulation of DSBs during replication, that cannot be efficiently resolved in the absence of the HR proteins BRCA1 or BRCA2, resulting in cell death. Despite the successful clinical trials, cancers treated with PARP inhibitors can still acquire resistance to therapy (reviewed in[50]). Detailed knowledge of the mechanisms of ICL repair may therefore help the identification of additional factors that could be targeted to induce synthetic lethality. To do so, specific inhibitors of FA factors should be developed, which could act by disrupting protein interactions, by blocking biochemical activities as strand annealing and exchange, ubiquitin conjugation and ligation, helicase activity or DNA binding. A recent study reported a small molecule inhibitor that prevents FANCL/UBE2T-mediated FANCD2 monoubiquitination, possibly by binding to the RING domain of FANCL[51]. Importantly, when the inhibitor was tested in U2OS cells, it increased sensitivity to carboplatin chemotherapy.

5

Alternatively, novel therapies could aim at factors involved in ICL repair-associated chromatin rearrangements. Inhibitors of histone deacetylases (HDACs) have been tested in combination with cisplatin for the treatment of malignant pleural mesothelioma and lung adenocarcinoma[52]. The combination of HDAC inhibitor and cisplatin resulted in an increase in cell death and reduced the emergence of resistance-phenotype, indicating the inhibition of chromatin remodeling factors is a valid strategy to enhance chemotherapy.

Concluding remarks

The use of *Xenopus* egg extracts in combination with plasmids containing chemically-defined ICLs has recently led to breakthroughs in the understanding of ICL repair. Using this system, I elucidated how endogenous crosslinks induced by acetaldehyde are repaired, and I investigated the nucleosomal rearrangements associated with ICL repair. These projects have contributed to the understanding of key mechanistic details regarding ICL repair and could in the future help improving therapies for cancer and Fanconi anemia.

References

1.   Garaycoechea, J.I. et al. Genotoxic consequences of endogenous aldehydes on mouse haematopoietic stem cell function. *Nature* **489**, 571-5 (2012).

2.   Clauson, C., Scharer, O.D. & Niedernhofer, L. Advances in understanding the complex mechanisms of DNA interstrand cross-link repair. *Cold Spring Harb Perspect Biol* **5**, a012732 (2013).

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

4.   Semlow, D.R., Zhang, J., Budzowska, M., Drohat, A.C. & Walter, J.C. Replication-Dependent Unhooking of DNA Interstrand Cross-Links by the NEIL3 Glycosylase. *Cell* **167**, 498-511 e14 (2016).

5.   Wu, R.A. et al. TRAIP is a master regulator of DNA interstrand crosslink repair. *Nature* **567**, 267-272 (2019).

6.   Hodskinson, M.R. et al. Alcohol-derived DNA crosslinks are repaired by two distinct mechanisms. *Nature* **579**, 603-608 (2020).

7.   Noll, D.M., McGregor Mason, T.M. & Miller, P.S. Formation and Repair of Interstrand Cross-Links in DNA. *Chem. Rev.* **106**, 277-301 (2006).

8.   Cho, Y.J. et al. Stereospecific formation of interstrand carbinolamine DNA cross-links by crotonaldehyde- and acetaldehyde-derived alpha-CH3-gamma-OH-1,N2-propano-2'-deoxyguanosine adducts in the 5'-CpG-3' sequence. *Chem Res Toxicol* **19**, 195-208 (2006).

9.   Knipscheer, P. et al. The Fanconi Anemia Pathway Promotes Replication-Dependent DNA Interstrand Cross-Link Repair. *Science* **326**, 1698-1701 (2009).

10.   Zhou, W. et al. FAN1 mutations cause karyomegalic interstitial nephritis, linking chronic kidney failure to defective DNA damage repair. *Nat Genet* **44**, 910-5 (2012).

11.   Castor, D. et al. Cooperative control of holliday junction resolution and DNA repair by the SLX1 and MUS81-EME1 nucleases. *Mol Cell* **52**, 221-33 (2013).

12.   Demuth, I., Digweed, M. & Concannon, P. Human SNM1B is required for normal cellular response to both DNA interstrand crosslink-inducing agents and ionizing radiation. *Oncogene* **23**, 8611-8 (2004).

13.   Dendouga, N. et al. Disruption of murine Mus81 increases genomic instability and DNA damage sensitivity but does not promote tumorigenesis. *Mol Cell Biol* **25**, 7569-79 (2005).

14.   Ishiai, M. et al. DNA cross-link repair protein SNM1A interacts with PIAS1 in nuclear focus formation. *Mol Cell Biol* **24**, 10733-41 (2004).

15.   Klein Douwel, D. et al. XPF-ERCC1 acts in Unhooking DNA interstrand crosslinks in cooperation with FANCD2 and FANCP/SLX4. *Mol Cell* **54**, 460-71 (2014).

16.   Long, D.T., Raschle, M., Joukov, V. & Walter, J.C. Mechanism of RAD51-dependent DNA interstrand cross-link repair. *Science* **333**, 84-7 (2011).

17.   Budzowska, M., Graham, T.G., Sobeck, A., Waga, S. & Walter, J.C. Regulation of the Rev1-pol zeta complex during bypass of a DNA interstrand cross-link. *EMBO J* **34**, 1971-85 (2015).

18.   Papadopoulo, D., Guillouf, C., Mohrenweiser, H. & Moustacchi, E. Hypomutability in Fanconi Anemia Cells is Associated with Increased Deletion Frequency at the HPRT Locus. *Proc Natl Acad Sci U S A* **87**, 8383-8387 (1990).

5

19. Singh, V. et al. Mechanism of repair of acrolein- and malondialdehyde-derived exocyclic guanine adducts by the alpha-ketoglutarate/Fe(II) dioxygenase AlkB. *Chem Res Toxicol* **27**, 1619-31 (2014).

20. Ryan J. Hansen, S.M.L., Sari J. Paikoff, Anthony E. Pegg, and M. Eileen & Dolan. Role of MGMT in Protecting against Cyclophosphamide-Induced Toxicity in Cells and Animals. *DNA Repair (Amst)* **6**, 1145-1154 (2007).

21. Guainazzi, A. & Scharer, O.D. Using synthetic DNA interstrand crosslinks to elucidate repair pathways and identify new therapeutic targets for cancer chemotherapy. *Cell Mol Life Sci* **67**, 3683-97 (2010).

22. Stone, M.P. et al. Interstrand DNA cross-links induced by alpha,beta-unsaturated aldehydes derived from lipid peroxidation and environmental sources. *Acc Chem Res* **41**, 793-804 (2008).

23. Huang, H. et al. DNA cross-link induced by trans-4-hydroxynonenal. *Environ Mol Mutagen* **51**, 625-34 (2010).

24. Huang, H. & Hopkins, P.B. DNA Interstrand Cross-Linking by Formaldehyde: Nucleotide Sequence Preference and Covalent Structure of the Predominant Cross-Link Formed in Synthetic Oligonucleotides. *J. Am. Chem. Soc.* **115**, 9402-9408 (1993).

25. Kozekov, I.D. et al. DNA Interchain Cross-Links Formed by Acrolein and Crotonaldehyde. *J. Am. Chem. Soc.* **125**, 50-61 (2003).

26. Wu, S.C. & Zhang, Y. Active DNA demethylation: many roads lead to Rome. *Nat Rev Mol Cell Biol* **11**, 607-20 (2010).

27. Pontel, L.B. et al. Endogenous Formaldehyde Is a Hematopoietic Stem Cell Genotoxin and Metabolic Carcinogen. *Mol Cell* **60**, 177-88 (2015).

28. Chaw, Y.M., Crane, L.E., Lange, P. & Shapiro, R. Isolation and Identification of Cross-Links from Formaldehyde-Treated Nucleic Acidst. *Biochemistry* **19**, 5525-5531 (1980).

29. Alcon, P. et al. FANCD2-FANCI is a clamp stabilized on DNA by monoubiquitination of FANCD2 during DNA repair. *Nat Struct Mol Biol* **27**, 240-248 (2020).

30. Wang, R., Wang, S., Dhar, A., Peralta, C. & Pavletich, N.P. DNA clamp function of the monoubiquitinated Fanconi anaemia ID complex. *Nature* **580**, 278-282 (2020).

31. Wang, S., Wang, R., Peralta, C., Yaseen, A. & Pavletich, N. Structure of the FA core ubiquitin ligase closing the ID clamp on DNA. *Nat Struct Mol Biol.* **28**(2021).

32. Sato, K. et al. Histone chaperone activity of Fanconi anemia proteins, FANCD2 and FANCI, is required for DNA crosslink repair. *EMBO J* **31**, 3524-36 (2012).

33. Clift, D., So, C., McEwan, W.A., James, L.C. & Schuh, M. Acute and rapid degradation of endogenous proteins by Trim-Away. *Nat Protoc* **13**, 2149-2175 (2018).

34. Aymard, F. et al. Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks. *Nat Struct Mol Biol* **21**, 366-74 (2014).

35. Escargueil, A.E., Soares, D.G., Salvador, M., Larsen, A.K. & Henriques, J.A. What histone code for DNA repair? *Mutat Res* **658**, 259-70 (2008).

36. Dingler, F.A. et al. Two Aldehyde Clearance Systems Are Essential to Prevent Lethal Formaldehyde Accumulation in Mice and Humans. *Mol Cell* **80**, 996-1012 e9 (2020).

37. Langevin, F., Crossan, G.P., Rosado, I.V., Arends, M.J. & Patel, K.J. Fancd2 counteracts the toxic effects of naturally produced aldehydes in mice. *Nature* **475**, 53-8 (2011).

38. Hira, A. et al. Variant ALDH2 is associated with accelerated progression of bone marrow failure in Japanese Fanconi anemia patients. *Blood* **122**, 3206-9 (2013).

39. Chen, C.H., Ferreira, J.C., Gross, E.R. & Mochly-Rosen, D. Targeting aldehyde dehydrogenase 2: new therapeutic opportunities. *Physiol Rev* **94**, 1-34 (2014).

40. Foretz, M., Guigas, B., Bertrand, L., Pollak, M. & Viollet, B. Metformin: from mechanisms of action to therapies. *Cell Metab* **20**, 953-66 (2014).

41. Zhang, Q.S. et al. Metformin improves defective hematopoiesis and delays tumor formation in Fanconi anemia mice. *Blood* **128**, 2774-2784 (2016).

42. Chen, C.H. et al. Activation of aldehyde dehydrogenase-2 reduces ischemic damage to the heart. *Science* **321**, 1493-5 (2008).

43. Tsai, J. et al. Pharmacologic Activation of Aldehyde Metabolism to Protect Hematopoietic Stem Cells (HSC) in Murine Models of Fanconi Anemia (FA). *Blood* **134**(2019).

44. Perez-Miller, S. et al. Alda-1 is an agonist and chemical chaperone for the common human aldehyde dehydrogenase 2 variant. *Nat Struct Mol Biol* **17**, 159-64 (2010).

45. Osawa, T., Davies, D. & Hartley, J.A. Mechanism of cell death resulting from DNA interstrand cross-linking in mammalian cells. *Cell Death Dis* **2**, e187 (2011).

46. Wynne, P. et al. Enhanced repair of DNA interstrand crosslinking in ovarian cancer cells from patients following treatment with platinum-based chemotherapy. *Br J Cancer* **97**, 927-33 (2007).

47. Spanswick VJ et al. Repair of DNA interstrand crosslinks as a mechanism of clinical resistance to melphalan in multiple myeloma. . *Blood* **100**(2002).

48. Chen, P. et al. The functional status of DNA repair pathways determines the sensitization effect to cisplatin in non-small cell lung cancer cells. *Cellular Oncology* **39**, 511-522 (2016).

49. Farmer, H. et al. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. . *Nature* **14**, 917-921 (2005).

50. Noordermeer, S.M. & van Attikum, H. PARP Inhibitor Resistance: A Tug-of-War in BRCA-Mutated Cells. *Trends Cell Biol* **29**, 820-834 (2019).

51. Cornwell, M.J. et al. Small-Molecule Inhibition of UBE2T/FANCL-Mediated Ubiquitylation in the Fanconi Anemia Pathway. *ACS Chem Biol* **14**, 2148-2154 (2019).

52. Gueugnon, F. et al. New histone deacetylase inhibitors improve cisplatin antitumor properties against thoracic cancer cells. *Oncotarget* **5**, 4504-4515 (2014).

5
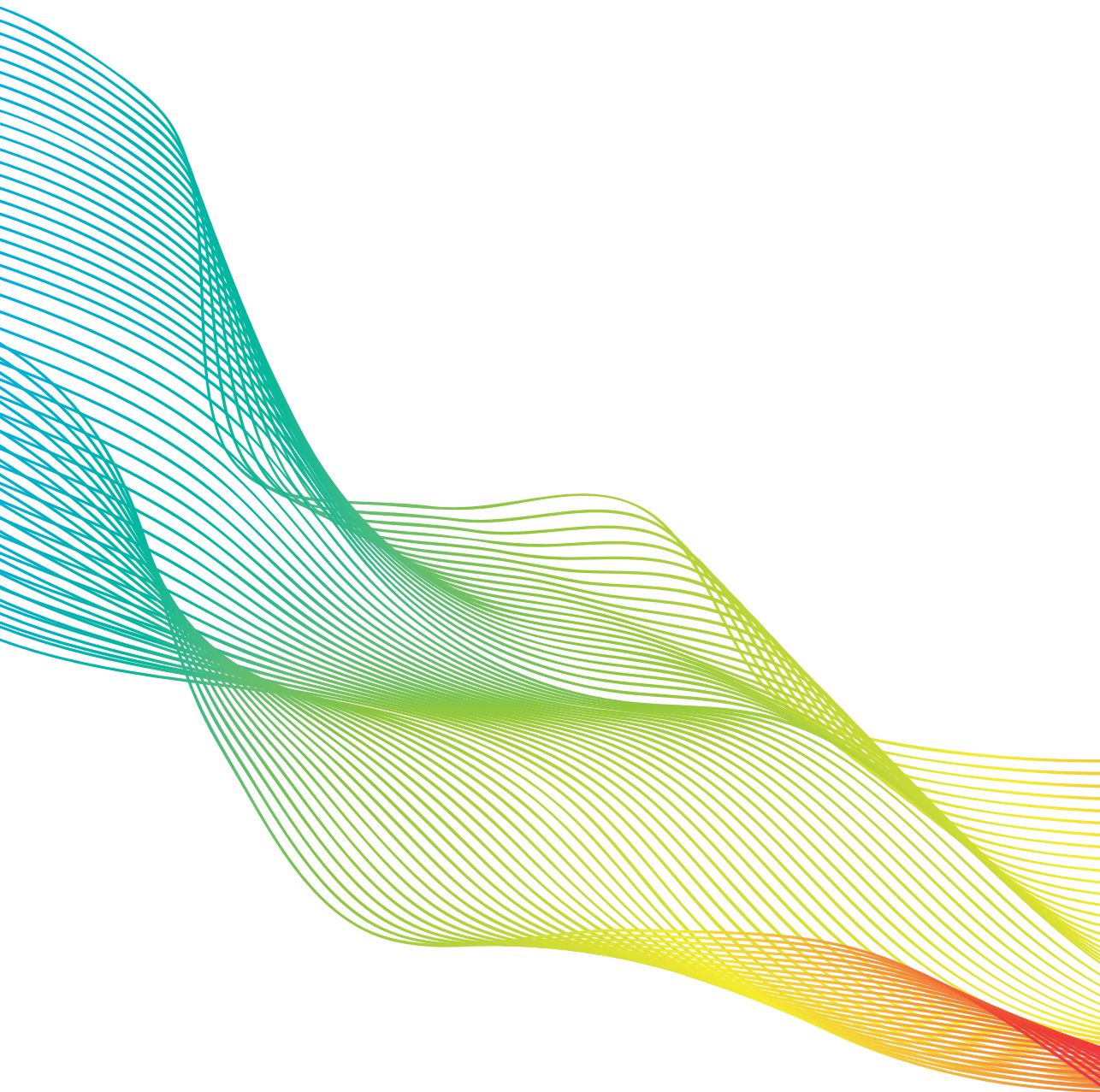
ADDENDUM
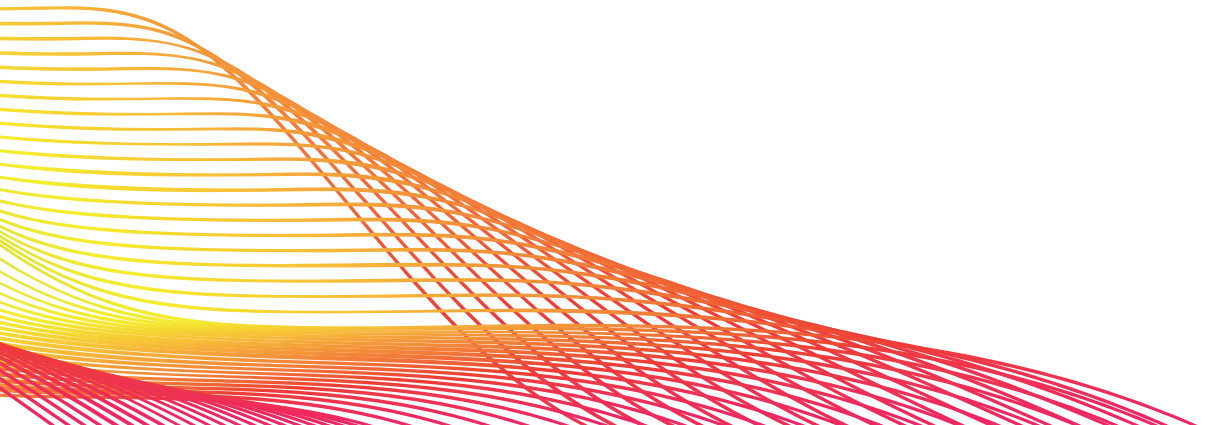
Nederlandstalige samenvatting

De meeste cellen in ons lichaam bevatten, verdeeld over 46 chromosomen, genetische informatie van beide ouders. Elk chromosoom bestaat uit twee lange DNA moleculen die als een dubbele helix om elkaar gewonden zijn. DNA bestaat uit vier bouwstenen, adenine, thymine, cytosine en guanine, de volgorde waarin de bouwstenen voorkomen vormen de code waarin genetische informatie is opgeslagen. De variatie in deze code bepaalt onder andere onder andere lichamelijke kenmerken zoals de kleur van de huid en de ogen. Deze variatie kan echter ook de onderliggende reden zijn voor genetische aandoeningen.

De informatie die opgeslagen ligt in het DNA is essentieel voor cellen omdat het bepaalt wat voor eiwitten er worden gevormd. Eiwitten besturen alle processen in de cel en veranderingen in het DNA, ofwel DNA mutaties, kunnen de productie van deze eiwitten. Dit kan essentiële cel processen verstoren en het is dus van groot belang dat het DNA intact en onveranderd blijft. DNA wordt op vele manieren beschadigd, bijvoorbeeld door zonlicht of sigaretten rook, maar ook door stoffen die het lichaam zelf produceert. Al deze factoren kunnen leiden tot wel 200,000 DNA beschadigingen per cel per dag. Als deze beschadigingen niet correct gerepareerd worden kan dat verschillende gevolgen hebben van DNA mutaties tot zelfs celdood. Gelukkig bevatten onze cellen verschillende mechanismen om DNA schade te repareren en zo zichzelf gezond te houden. In **hoofdstuk 1** beschrijf ik de belangrijkste DNA reparatie mechanismen van de cel.

Dit proefschrift houdt zich bezig met de reparatie mechanismen van één specifiek type DNA schade; DNA interstrand crosslinks, of ICLs. Dit is een bijzonder gevaarlijke vorm van DNA schade omdat het de twee strengen van het DNA op een onnatuurlijke manier stevig aan elkaar koppelt. Om bij de informatie die opgeslagen ligt in het DNA te kunnen, moeten de twee strengen van fysiek van elkaar los gemaakt zodat er ruimte is om de DNA volgorde af te lezen. Een ICL verhinderd deze scheiding van strengen en belemmert daarmee belangrijke processen. Zoals het kopiëren van het DNA, dat moet gebeuren voor elke celdeling, en eiwitproductie. Dit maakt ICLs zo uitzonderlijk schadelijk en wanneer niet op tijd gerepareerd kunnen ze grote fouten in het DNA of zelfs celdood tot gevolg hebben.

ICLs kunnen worden gevormd door verschillende stoffen die in cellen tijdens natuurlijke processen voorkomen. Bijvoorbeeld, Acetaldehyde, een stofje dat tijdelijk vrijkomt tijdens de afbraak van alcohol, kan een ICL induceren. Daarnaast zijn er ook stoffen van buiten ons lichaam die ICLs veroorzaken, bijvoorbeeld chemische stoffen die gebruikt worden tijdens chemotherapie tegen kanker. Deze therapeutische stoffen doden specifiek snel delende cellen zoals kankercellen

omdat deze gevoeliger zijn voor beschadigingen die het kopiëren van het DNA belemmeren.

Het is nog niet duidelijk hoe deze ICLs in cellen precies gerepareerd worden. Veel van de kennis die vandaag de dag hierover hebben, komt van het bestuderen van een zeldzame genetische ziekte: Fanconi anemie. Patiënten met deze ziekte worden geboren met mutaties in de DNA code voor het maken van eiwitten die betrokken zijn bij ICL reparatie. Deze eiwitten worden de Fanconi anemie eiwitten genoemd en werken samen in de "Fanconi anemie route' voor ICL reparatie. De cellen van Fanconi anemie patiënten hebben problemen met het repareren van ICLs. Hierdoor zijn ze zeer gevoelig voor stoffen die dit type DNA schade induceren, zoals bepaalde chemotherapeutica. Het belang van deze DNA reparatie route wordt verder benadrukt door het feit dat Fanconi anemie patiënten vaak aangeboren afwijkingen aan de ledematen vertonen, een extreme vorm van bloedarmoede hebben en ook een sterk toegenomen kans op de ontwikkeling van kanker hebben.

Het is erg ingewikkeld om de details van het ICL reparatieproces te bestuderen om verschillende redenen. Ten eerste, veroorzaken de stoffen die ICLs induceren tegelijk ook veel andere soorten DNA schade die de cellen beïnvloeden. Daarom is het moeilijk te bepalen welk effect van welke schade komt en dus welk effect specifiek toe te wijzen is aan ICLs. Daarnaast weten we dat ICLs voornamelijk gerepareerd worden tijdens het kopiëren van het DNA. DNA replicatie is een zeer ingewikkeld proces dat niet eenvoudig na te bootsen is buiten de cel. Om beide problemen te adresseren hebben we gebruik gemaakt van een uniek modelsysteem gebaseerd op extracten gemaakt van de eitjes van de *Xenopus laevis* kikker. Deze extracten bevatten alle eiwitten die nodig zijn om efficiënt DNA te kopiëren en bepaalde schades te repareren. Wanneer we in dit systeem DNA repliceren met een ICL wordt dit gerepareerd met mechanismen overeenkomstig met die in onze cellen. Door het DNA te modificeren met een ICL op een specifieke plaats kunnen wij zowel de replicatie en reparatie goed bestuderen zonder dat daar een cel voor nodig is.

Hoofdstuk 1 beschrijft in het kort de belangrijkste reparatie mechanismen van DNA schade en wat zij betekenen voor het landschap van het omringende DNA. Een onderliggend begrip van hoe verschillende soorten schades worden gerepareerd helpt ons de zeer complexe mechanismen van ICL-reparatie beter te definiëren. Verder beschrijf ik de huidige kennis van reparatie van ICLs en hoe het *Xenopus* ei-extract systeem hieraan heeft bijgedragen. ICL reparatie is een complex proces bestaand uit meerdere stappen dat begint met de herkenning van de ICL. Dit gebeurt tijdens het kopiëren van het DNA als de onnatuurlijke verbinding van de twee DNA strengen door de ICL het verder aflezen van de DNA volgorde verhinderd. Het stilvallen van het kopieerproces door deze blokkade is een signaal dat DNA

reparatie nodig is. De volgende doorslaggevende stap is het scheiden van de twee verbonden DNA strengen. Dit wordt bewerkstelligt door twee knipjes te maken in een van de twee rondom de de ICL. Nadat de twee verwikkelde strengen van elkaar zijn losgemaakt kan DNA replicatie worden voortgezet en wordt ICL reparatie afgerond.

Hoofdstuk 2 beschrijft het onderzoek naar het reparatie mechanisme van ICLs veroorzaakt door acetaldehyde, een stof die onder andere gevormd wordt tijdens de afbraak van alcohol in cellen. Uit eerder genetisch onderzoek weten we dat cellen twee verschillende verdedigingsmechanismen hebben tegen dit reactieve acetaldehyde. Het eerste mechanisme is gebaseerd op de eliminatie van de stof uit de cel door het enzym genaamd ALDH2. Hierdoor accumuleert er geen acetaldehyde in de cel die ICLs kunnen veroorzaken. Echter, hebben wel 2 miljard mensen wereldwijd een DNA mutatie in het ALDH2 gen dat resulteer in een defect ALDH2 eiwit dat acetaldehyde niet goed kan verwijderen. Deze mensen lopen meer DNA schade veroorzaakt door acetaldehyde op en hebben daardoor een grotere kans om alcohol-gerelateerde kankers te ontwikkelen. Het tweede verdedigingsmechanisme tegen acetaldehyde is de reparatie van deze veroorzaakte DNA schade. Gebruik makende van het *Xenopus* ei extract systeem onderzochten wij hoe acetaldehyde geïnduceerde ICLs worden grepreerd, dit werk is beschreven in **hoofdstuk 2** van dit proefschrift. We ontdekte dat deze natuurlijk voorkomende DNA schade gerepareerd wordt door twee verschillende mechanismen: De eerder genoemde Fanconi anemie route, en een andere tot nog toe nooit eerder beschreven route. Een belangrijk verschil tussen de twee zit in hoe deze mechanismen de ICL losknippen van een van de twee DNA strengen. In de Fanconi anemie route gebeurt dit door een van de DNA streng te knippen naast de plek van de crosslink , terwijl de nieuwe route de crosslink zelf knipt. Dit heeft belangrijke gevolgen voor de rest van het reparatie mechanisme. Onze bevindingen identificeren niet alleen een nieuwe ICL reparatie route maar laten ook zien dat ICLs die door stoffen in de cel gevormd worden op verschillende manieren gerepareerd kunnen worden. Dit werk kan bijdragen aan de ontwikkeling van betere behandelmethoden voor zowel alcohol-gerelateerde kankers als Fanconi anemie.

DNA moleculen in de cel zijn zo'n 2 meter lang en moeten efficiënt opgevouwen worden om in de celkern te passen. Dit wordt voor een deel bewerkstelligd door het binden van het DNA aan specifieke eiwitten, histonen genaamd. Acht histon eiwitten samen vormen een complex waar het DNA strak om gewonden zit. Deze eenheid van een histon-complex met DNA wordt een nucleosoom genoemd. Een groot gedeelte van het DNA is opgevouwen in deze nucleosomen en de structuur van deze meerdere nucleosomen noemen we chromatine dat vervolgens nog verder wordt opgevouwen. Chromatine speelt een belangrijke rol bij het aflezen en kopiëren van DNA. De strakke structuur van chromatine moet bij deze processen tijdelijk

losser worden gemaakt. Dit gebeurt vaak door nucleosomen te verwijderen of te verplaatsen. Wat ook gebeurd tijdens reparatie van DNA schade waarbij speciale eiwitten betrokken zijn, een proces dat we nog niet goed begrijpen.

In hoofdstuk 3 bestuderen we de nucleosomen tijdens de reparatie van ICLs gevormd door cisplatine, een stof gebruikt in chemotherapie. Eerst ontwikkelden we een methode, gebaseerd op de *Xenopus* ei extracten, die ons in staat stelde de positie van nucleosomen te bepalen en zo inzicht te verkrijgen in hoe nucleosomen gehergeorganiseerd worden tijdens ICL reparatie. Hiermee ontdekten we dat in een vroeg stadium van reparatie nucleosomen verdwenen rond de plek van de DNA schade, waarschijnlijk om de DNA reparatie eiwitten toegang te verlenen. In latere stadia van ICL reparatie werden nucleosomen rond the ICL terug geplaats. Om te ontdekken welke eiwitten bij deze herorganisatie van nucleosomen betrokken zijn isoleerde we alle eiwitten die tijdens ICL reparatie aan het DNA binden. Na identificatie van deze eiwitten met behulp van massa spectrometrie bleek dat er verschillende eiwitten, waarvan bekend is dat ze een rol kunnen spelen bij nucleosoom organisatie, teruggevonden werden. Vervolgonderzoek moet uitwijzen of deze eiwitten daadwerkelijk deze rol vervullen tijdens ICL reparatie en hoe ze precies te werk gaan. Als blijkt dat het herorganiseren van nucleosomen inderdaad van belang is voor ICL reparatie zou deze kennis bij kunnen dragen aan de ontwikkeling van betere therapieën tegen kanker.

Een van de Fanconi anemie eiwitten, FANCD2, speelt een belangrijke rol bij de reparatie van ICLs omdat het betrokken is bij het losknippen van de crosslink van een van de twee DNA strengen. Recent onderzoek suggereert dat FANCD2 ook betrokken zou kunnen zijn bij het positioneren van nucleosomen tijdens ICL reparatie. **Hoofdstuk 4** beschrijft het onderzoek naar deze rol van FANCD2 in de herorganisatie van nucleosomen. Gebruik makend van verschillende methoden, zowel het *Xenopus* ei extract systeem als reconstitutie met gezuiverde eiwitten, vonden wij geen bewijs voor de betrokkenheid van FANCD2 bij dit proces. FANCD2 bleek wel te binden aan histon eiwitten en dit zou nog een rol kunnen spelen in het ICL reparatie proces, bijvoorbeeld door het aantrekken van andere eiwitten naar de DNA schade. Vervolgstudies zullen het belang en de functie van deze interactie verder moeten onderzoeken.

In **hoofdstuk 5** bespreek ik mijn onderzoeksresultaten in een bredere context en in het licht van de recente literatuur. Ik beschrijf hoe mijn resultaten bijdragen aan de kennis van ICL reparatie en doe suggesties voor vervolgstappen die de nog openstaande vragen kunnen beantwoorden. Verder bediscussieer ik hoe een beter begrip van het mechanisme van ICL reparatie kan bijdragen aan een gezond leven door een beter begrip van Fanconi anemie en het de mogelijke ontwikkeling van efficiëntere therapieën voor de behandeling van kanker.

A

Addendum

English Summary

Most cells in our body contain the genetic information derived from our parents, divided into 46 chromosomes. Each chromosome consists of DNA, a long molecule in which two strands are wrapped around each other in the shape of a double helix. The genetic information is stored as a "DNA sequence", composed of 4 different units: adenine, thymine, cytosine and guanine. These units are called nucleotides, and they are the building blocks of DNA. Variations in the DNA sequence determine human physical traits such as skin and eye colour. However, sometimes they can also cause genetic diseases.

The information stored in DNA is essential for cells, as it encodes for proteins that carry out and regulate all cellular processes. It is therefore important that the genetic information remains intact, since a change in the DNA sequence, named mutation, could lead to improper cellular functioning. Multiple factors damage DNA, such as UV light, cigarette smoke, as well as cellular metabolites. All these factors lead to around 200,000 DNA lesions per day in a single cell. If left unrepaired, DNA damage can lead to mutations in the DNA sequence or even cell death. Luckily, there are multiple methods to repair DNA lesions and maintain cellular function. In **chapter 1**, I describe the main DNA repair mechanisms available to the cell.

This thesis focuses on one specific type of DNA lesion: the DNA interstrand crosslink (ICL). This is a very toxic DNA damage that sticks together the two DNA strands. In order to access the information encoded in the DNA, the cell must separate the two DNA strands and only then it can "read" the DNA sequence. The DNA strands must also be separated during cell division, a process necessary for the creation of new cells. Before a cell can divide, it needs to create a copy of its DNA sequence that will be passed to the daughter cell after division. This is done through an intricate process called DNA replication that requires separating the two DNA strands, reading the DNA sequence, and finally copying it. Since ICLs glue together the two strands of DNA, they inhibit strand separation and therefore make it impossible to either read or copy the DNA sequence. For this reason, ICLs are extremely toxic and, if left unrepaired, can lead to DNA instability and cell death.

ICLs are induced by a variety of compounds that are formed during natural processes, such as the breaking down of alcohol in the body. Some chemotherapeutics can also induce ICLs, resulting in the death of cancer cells. Since cancer cells divide more often than normal cells, they are more vulnerable to replication-blocking agents like ICLs. Because of this, ICL inducing agents are often used in chemotherapy.

A

How the cell repairs an ICL is still not completely understood. Much of our current knowledge on ICL repair is derived from the study of a rare genetic disorder, Fanconi anemia. Individuals with Fanconi anemia have mutations in the DNA coding for proteins that are necessary for the repair of ICLs. These proteins are called Fanconi anemia proteins and they act in the "Fanconi anemia pathway" of ICL repair. Cells of individuals with Fanconi anemia cannot repair ICLs and are therefore highly sensitive to agents that induce this type of DNA lesions, such as various chemotherapeutic drugs. Highlighting the importance of ICL repair for human health, individuals with Fanconi anemia often display developmental malformations, severe anemia which often results in bone marrow failure, and cancer predisposition.

Studying how ICLs are repaired is very difficult because of technical challenges. Firstly, many of the agents that are used to induce ICLs in cells also create other types of DNA damage. It is therefore very complicated to study specifically how ICLs are repaired when working with cells. However, studying ICL repair outside of a cell is also technically challenging, since this process occurs during DNA replication, the mechanism by which the cell to copies its DNA. DNA replication is a very complex mechanism, which is difficult to reproduce outside of the cell using only purified proteins and DNA. For the work presented in this thesis, we used a unique system based on the eggs from the frog *Xenopus laevis*. These eggs contain all the proteins necessary to copy and repair DNA. Extracts derived from *Xenopus laevis* eggs are able to copy DNA containing a highly defined ICL, and to repair the lesion through a mechanism that is highly conserved between frogs and humans. This system allows researchers to study how ICLs are repaired without using cells.

Chapter 1 describes briefly the most important mechanisms of DNA damage repair, and the associated changes in the surrounding DNA landscape. Knowing how different types of DNA lesions are repaired can help to better define the complex mechanisms of ICL repair. ICL repair is a multi-step process that starts with the detection of the ICL during DNA replication, initiating a signal for DNA repair proteins to start repair. The central step in ICL repair is the separation of the two DNA strands, which is achieved by cutting the DNA surrounding the ICL in one DNA strand. After strand separation, DNA replication can resume and ICL repair is finalized. In **chapter 1** I describe the current state of knowledge regarding the mechanism of ICL repair, and how the *Xenopus* egg extract system has contributed to define this process.

Chapter 2 investigates how naturally occurring ICLs are repaired. As mentioned above, some cellular processes can lead to the formation of reactive compounds that can induce ICLs. This is the case for acetaldehyde, a compound which is formed during the break-down of alcohol and during other cellular processes. Fortunately, cells have two lines of defence against the DNA damage induced

by acetaldehyde. The first line of defence is the removal of acetaldehyde by a specialised protein, the ALDH2 enzyme. This enzyme removes acetaldehyde from the cell before it can damage DNA, thereby preventing the formation of ICLs. However, about 2 billion people worldwide have a mutation in the DNA coding for ALDH2, resulting in a defect in acetaldehyde clearance. Because of this, these individuals are more prone to acetaldehyde-induced DNA damage and have higher risk of developing alcohol-related cancer. The second line of defence is the repair of acetaldehyde-induced DNA damage. In **chapter 2** we investigated the mechanism of repair of acetaldehyde-induced ICLs using the *Xenopus* egg extract system. We found that these naturally occurring ICLs can be repaired in two ways: the Fanconi anemia pathway and a novel faster route. These two mechanisms act differently on the ICL: in the Fanconi anemia pathway the two strands are separated by cutting the DNA surrounding the ICL, while in the novel mechanism the ICL itself is cut. This has important implications for the rest of the repair mechanism. Our finding not only identifies a completely new ICL repair pathway but it also proves that naturally occurring ICLs can be repaired by multiple mechanisms, broadening our understanding of ICL repair. Moreover, new insights in how acetaldehyde-induced DNA damage is repaired could help the development of treatments for alcohol-related cancers and Fanconi anemia.

The approximately 2 meters of DNA in each of our cells must be tightly folded and compacted to fit into the nucleus. This is in part achieved by the binding of DNA to histones, specialized proteins that compact DNA. DNA is wrapped around a complex of eight histones, forming a DNA/histones complex called "the nucleosome". DNA compacted by nucleosomes is called chromatin, which is then further compacted to form the chromosomes. The structure of chromatin plays an essential role in the regulation of important cellular processes such DNA replication and protein production. Since chromatin is so tightly packed, it must be loosened in order to access the information in the DNA sequence. This is achieved by specialised proteins, that can use energy to either loosen or condense chromatin during cellular processes. How this process works, especially during DNA repair, is still poorly understood.

In **chapter 3** we describe the interactions between DNA and histones during repair of an ICL induced by the chemotherapeutic drug cisplatin. We developed a method based on the *Xenopus* egg extract system that allowed us to determine the position of the nucleosomes and thereby study how nucleosomes are (re)organized during ICL repair. We found that at the initial stages of ICL repair, nucleosomes were disassembled around the DNA lesion, likely to allow access of repair proteins to the DNA. Nucleosomes were formed again later, at the final stages of repair, to re-compact DNA. In order to understand what proteins are involved in the disassembly and reassembly of nucleosomes during ICL repair, we used a novel technique

A

that can isolate all the proteins that bind DNA during repair. These proteins were then identified using a method called mass spectrometry. With this method, we found multiple factors that could be involved in the reorganization of nucleosomes during ICL repair. Future research will have to determine which of these proteins are necessary for nucleosome reorganization in order to better define this process. Understanding nucleosome dynamics during ICL repair is not only important for deepening our understanding of this DNA repair mechanism, but it could lead to the development of better cancer therapies.

The protein FANCD2 is an essential component of the Fanconi anemia pathway of ICL repair, as it promotes the unhooking of the ICL from one of the two DNA strands. Recent studies indicated that FANCD2 could have additional functions in ICL repair, possibly acting on nucleosome dynamics. In **chapter 4** we investigated if FANCD2 is involved in the re-organization of nucleosomes during ICL repair. Using various techniques, both based on purified proteins and on the *Xenopus* egg extract system, we found no direct role for FANCD2 in nucleosome dynamics during DNA repair. However, we found that FANCD2 could bind to histones, opening the possibility that this interaction may be important for ICL repair. For example, FANCD2 could bind histones during repair to recruit other DNA repair proteins to the ICL and therefore promote repair. For this reason, further studies are necessary to understand this interaction in the context of ICL repair.

In **chapter 5** I discuss my research in a broader context, considering the current scientific literature. I examine how the experiments described in this thesis contribute to our understanding of ICL repair, and I speculate about what can be the next steps to address some of the remaining open questions. Finally, I analyze how a detailed understanding of ICL repair can promote human health, both for alleviating symptoms of Fanconi anemia and for optimising chemotherapies based on ICL inducing agents.

Curriculum Vitae

Alice Bolner was born on November 9th 1990 in Trento, Italy.

In 2009 she started her Bachelor Degree in Biomolecular Sciences and Technologies at the University of Trento. Under the guidance of Dr. Gabriella Viero, she performed her bachelor thesis on the characterization of polysomes and RNA binding proteins by atomic force microscopy.

In 2013 she moved to the Netherlands to start the Master Degree in Biomedical Sciences at Leiden University. During her Master, she did an internship in the department of Cell and Chemical Biology at Leiden University Medical Centre. Here, she used electron microscopy to study viral replication, under the guidance of Prof. Bram Koster and Dr. Montserrat Bárcena. She did her second internship at the Leiden University Medical Centre, in the Department of Human Genetics, under the guidance of Dr. Harry Vrieling and Dr. Giel Hendriks. For this project she studied the cellular responses to genotoxic stress in embryonic stem cells.

In April 2015 she started her PhD at the Hubrecht Institute in the group of Dr. Puck Knipscheer, investigating the molecular mechanisms of DNA interstrand crosslink repair. The work developed on this subject is presented in this thesis.

A

List of publications

Hodskinson MR*, **Bolner A***, Sato K, Kamimae-Lanning AN, Rooijers K, Witte M, Mahesh M, Silhan J, Petek M, Williams DM, Kind J, Chin JW, Patel KJ, Knipscheer P. Alcohol-derived DNA crosslinks are repaired by two distinct mechanisms. *Nature 579*, *603-608*. (2020).

Lunelli L, Bernabò P, **Bolner A**, Vaghi V, Marchioretto M, Viero G. Peering at Brain Polysomes with Atomic Force Microscopy. *J Vis Exp* **109**. (2016).

*Equal contribution

Acknowledgments

This PhD thesis is the output of the effort and support of several people to whom I am extremely grateful.

First and foremost, I thank **Puck** who encouraged me and helped me through this unpredictable PhD journey. Thank you for your warm support through the years and for your continuous optimism. You are always ready to help no matter how busy you are, thank you for all your support!

I would like to thank the reading committee that has critically assessed and approved the work presented in this thesis: **Edwin Cuppen, Wouter de Laat, Susanne Lens, Francesca Mattiroli, Marcel van Vugt**. In addition, I would like to thank the additional members of the Doctoral Examination committee, **Jeroen Bakkers** and **Harry Vrieling**; thank you all for your time and efforts, I look forward to our discussion on December 9th.

I would like to also express my gratitude to my promotor **Geert Kops**, who guided me through the years and helped me focusing on the important questions. Our discussions were extremely valuable for the completion of this thesis, thank you.

I would also like to thank **Litha Schipper** for her support in this final months and for making sure that everything went smoothly during the preparation of this thesis.

Part of the work presented in this thesis would not have been possible without the contribution of our collaborators from the **Patel lab**, thank you for all your work and the scientific discussions.

My sincere thanks to my paranymphs, **Giulia** and **Roxanne**. Thank you for all the discussion at work and for the drinks afterwards! Especially thank you for your support in this past year, it was a very challenging time but you always managed to make me smile :)

I would like to thank all the past and present members of the **Knipscheer lab** for helping me with my experiments (especially when they were not going so well) and for listening to endless discussions and complaints about ICL repair. Thanks to **Merlijn, Koichi, Jamie, Themis, Lars** and **Alex** as well as the former lab members **Daisy, Anna, Rick, Wouter, Nerea** and **Aiko**, for creating a great atmosphere in the lab. And of course, thanks to the intern students who helped in my projects, **Merijn** and **Jos**.

I am also grateful for all the discussions during lab meetings with the **Mattiroli lab**. Thanks to **Clément, Bruna, Jan, Fabian, Inge**, you are great lab neighbours and I look forward to hearing about your future findings.

A

I would also like to thank **Huub** and the **animal caretakers**, who make it possible for us to work with radioactivity and extracts at the Hubrecht institute.

A special thanks goes to the **Creyghton group**, that for many years made sure we were always on time for lunch! Thanks to **Bas, Caro, Ilia, Sander, Maartje** and **Menno** for the lunch fun and the nice borrels!

Thanks to everyone at the **Hubrecht Institute**, and to the former Hubrechters. I have nice memories of all the fun borrels, Christmas dinners, BBQs and labstapdagens, thanks to everyone who contributed to make the Hubrecht a great place to work in.

I also would like to thank everyone who contributed to **Prout**, especially **Tamar, Vinzenz, Federico, Ilona, Jelmer, Özge, Ellen, Suze, Amaury** and **Romy**, it was a pleasure to work with you and I wish you the best for your PhD projects and for the next steps!

A special thanks to those who managed to make me forget about all the failed experiments and have a good laugh: the **Might Riders** (**Federica, Emilie, Vyom, Jason, Dushyant, Pedro** and **Enikö**), **João, Emma, Ljubisja, Aileen, Chris, Ariana, Anna, Bart, Elena C., Elena K., Michelle, Katrin** and **Matteo**. Thank you all for the fun times!

Ringrazio poi le persone che mi sono a fianco da una vita, nonostante i 1000 km di distanza. Vorrei ringraziare **Lisa** per tutte le chiacchierate e i cocktail annessi, grazie per il sostegno, anche quando è difficile capire che mi passa per la testa. Non vedo l'ora di cantare assieme a squarciagola "let it go" a -15 C in Lapponia!. Grazie a **Veronica**, che con la sua gentilezza e premura mi fa sempre sentire a casa, è stato un piacere assistere al tuo matrimonio ed essere presente per questo tuo nuovo inizio. Aspetto con impazienza la nuova stagione di feste in piscina! E grazie a **Cinzia**, che c'è sempre nonostante i mille impegni, grazie per tutte le serate passate assieme! Riesci sempre a mettermi di buonumore, anche alle 2 di notte quando le chiavi della macchina sono a 50 km di distanza :)

Vorrei infine ringraziare la mia **famiglia**, e specialmente **Simon**, che mi ricorda sempre quando è ora di staccare e andare in vacanza.