Dissertation zur Erlangung des Doktorgrades der

Fakultät für Chemie und Pharmazie

der Ludwig-Maximilians-Universität München

Characterisation of DNA-protein crosslink repair with the Purification of x-linked Proteins technique

Pedro Weickert Vivancos

aus

Huelva, Spanien

2023

Erklärung

Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. Julian Stingele betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, ...15.05.2024..................................

....

Pedro Weickert Vivancos

Dissertation eingereicht am   20.12.2023

1. Gutachter: Prof. Dr. Julian Stingele

2. Gutachter: Prof. Dr. Karl-Peter Hopfner

Mündliche Prüfung am   10.05.2024

"We have shown that, in vitro, pronase can cleave the [protein-bridged] Interstrand crosslinks induced in DNA by formaldehyde.... If a proteolytic enzyme was to act in the same way in vivo, this would constitute DNA repair." [1]

Table of content

List of publications

This dissertation has been prepared from October 2019 until November 2023 under the supervision of Prof. Dr. Julian Stingele at the Gene Center of the Ludwig-Maximilian Universität in Munich. This cumulative thesis is based on the following two publications and one manuscript.

Publication 1

Pedro Weickert*, Hao-Yi Li*, Maximilian J. Götz, Sophie Dürauer, Denitsa Yaneva, Shubo Zhao, Jacqueline Cordes, Aleida C. Acampora, Ignasi Forne, Axel Imhof & Julian Stingele (2023). SPRTN patient variants cause global-genome DNA-protein crosslink repair defects. ***Nat. Commun***. 14, 352. *equal contribution

Publication 2

Pedro Weickert and Julian Stingele (2022). DNA–Protein Crosslinks and Their Resolution. ***Annu. Rev. Biochem***. 91:157-181.

Manuscript 1

Pedro Weickert*, Sophie Dürauer*, Maximilian J. Götz, Hao-Yi Li and Julian Stingele. Identifying and monitoring covalent DNA-protein crosslinks by PxP. *equal contribution

Summary

DNA-Protein Crosslinks (DPCs) have emerged as an important source of endogenous and exogenous DNA damage. DPCs arise when proteins get covalently trapped on DNA, what can happen during the action of enzymes that naturally establish covalent intermediates with DNA -such as topoisomerases- but also by the action of reactive compounds. One of these compounds, formaldehyde, is an environmental toxin and a metabolite produced during one-carbon metabolism and methanol detoxification. Formaldehyde is extremely reactive and generates protein-protein crosslinks, RNA-protein crosslinks and DPCs. DPCs are toxic because they can block essential chromatin transactions such as replication or transcription. Their toxicity has been exploited in chemotherapy with the use of different drugs that induce these genomic lesions either specifically, camptothecin (CPT) traps TOP1 or etoposide traps TOP2, or unspecifically, with compounds like cisplatin commonly used in breast cancer treatment.

DPC-repair generally involves specialized proteases which mediate the destruction of the protein adduct. In higher eukaryotes, the metalloprotease SPRTN is recruited to the lesion in a process that involves the collision of the replication fork with the adduct. SPRTN is activated by the DNA structure generated by polymerase stalling, cleaves the protein and allows peptide bypass by translesion synthesis (TLS). The study of DPC-repair and the identification of DPCs generated by non-specific crosslinkers has remained elusive due to limited methodology to study these lesions. Previous methods to isolate DPCs, rely on DNA precipitation and are prone to technical biases due to the presence of protein aggregates, giving false positives, or changes in DNA precipitation behaviour due to the variable crosslinked protein nature, giving false negatives. Therefore, I established during my master thesis, and optimized during my PhD, a method for the isolation, identification and monitoring of DPCs, the Purification of x-linked Proteins (PxP). Using PxP in combination with mass spectrometry, I uncovered the identity of DPCs generated by physiologically relevant levels of formaldehyde. Strikingly, formaldehyde-induced DPCs are less complex than anticipated, they mostly consist of crosslinked nucleosomes. Then, we decided to apply PxP for the study of 5-azadC-induced DPCs, which consists mostly of the DNA Methyl Transferase 1 (DNMT1). We observed DNMT1-DPC formation in a dose-dependent manner and could monitor their repair kinetics. We discovered that repair of this crosslinks involved the SUMOylation of the adduct and SUMO-targeted ubiquitylation by the StUbL RNF4, which triggered proteasomal degradation of crosslinked DNMT1. In agreement with biochemical data and previous work, *RNF4* knock-out cells displayed 5-azadC sensitivity.

Interestingly, we observed a cleaved DNMT1-DPC fragment whose appearance was dependent on the same modifications -SUMO and ubiquitin- but not generated by the proteasome. We identified SPRTN as the enzyme responsible for this cleavage and, through structure function analysis, conclude that the UBZ domain is responsible for the recruitment of SPRTN to DNMT1-DPCs. Strikingly, only the loss of the UBZ domain completely phenocopied the *SPRTN-ΔC* allele, which is causative for Ruijs-Aalfs syndrome. Our data support a replication-independent role for the metalloprotease SPRTN during DPC-repair and highlight the importance of the UBZ domain during lesion recognition.

Given the multiple applications of the PxP for the study of DPC biology, we decided to write an article in which we described the method step by step to facilitate its implementation in other labs. We show that the PxP can be used to isolate also other types of DPCs such as those generated by non-competitive inhibitors (e.g etoposide) and that it can successfully isolate relevant physiological DPCs, like those generated by HMCES after short wavelength ultraviolet light (UVC) exposure. In conclusion, we believe that this manuscript will be a great resource for laboratories which do research on the DNA damage field.

Introduction

1. DNA is the storage of genetic information in living organisms

Genes store biological information that must be accurately copied during cell division. This information is encoded in DNA, a molecule made up of two opposing and antiparallel strands of nucleotides. Each nucleotide contains a nitrogen-containing base (A, C, T, or G) and a sugar-phosphate backbone. Hydrogen bonds between complementary base pairs (A-T and C-G) form the double helix structure of DNA. This structure allows for the precise replication and transmission of genetic information from one generation to the next[2]. Additionally, DNA is also transcribed in RNA, which will be further translated into proteins. During this introduction I will discuss essential processes related to the DNA molecule with a special focus on its repair.

1.1 DNA replication

DNA replication is an active process in which a parental double-stranded DNA (dsDNA) is duplicated before cell division. In *E. coli*, the prokaryotic model organism, replication initiates with the binding of a sensor protein, DnaA, to the chromosomal origin of replication C (oriC). The oriC is approximately 245 base-pair (bp) long and rich in AT sequences. Binding of DnaA, triggers separation of the DNA strands, leading to recruitment and loading of the helicase DnaB and other enzymes later (e.g. primase and DNA Polymerase III)[3]. Once primase synthesizes a de novo RNA primer, this is extended by DNA polymerase III, what only occurs in a 5'-3' direction. Replication happens in a bidirectional fashion, with a leading strand needing just one primer and a lagging strand synthesized in short pieces known as Okazaki fragments. During lagging strand replication, single-stranded stretches of DNA are coated by the *E. coli* single strand binding protein (SSB). In *E. coli*, a circular chromosome with a single origin of replication assures the presence of two replication forks, which can replicate the DNA at a remarkable speed of more than 1 kb per second[4]. Termination is determined by the binding of the Tus protein to multiple predetermined sequences in the circular chromosome, blocking the replisome unidirectionally[5]. After termination, both chromosomes are decatenated by topoisomerase IV, allowing the generation of two independent chromosomes[6]. In contrast to prokaryotes, eukaryotes have linear chromosomes with multiple replication origins. The origin recognition complex (ORC, formed by Orc1, Orc4 and Orc5) recognizes replication origins, which in budding yeast are formed by 10 bp and named autonomously replicating sequences (ARS)[7,8]. Replicative helicases can be loaded in origins associated with ORC. Two Mcm2-7 helicase cores in a head to head orientation form the prereplicative complex, what requires additional factors[7,8]. Cdc45 and GINS associate with Mcm2-7 to allow helicase activity, a process which is highly dependent on the cell cycle[9-11]. The Dbf4-

dependent protein kinase (DDK) and the cyclin dependent protein kinase (CDK) regulate replication initiation through phosphorylation, controlling this interaction and further recruiting factors needed for the formation of the replication fork[12-14]. Polymerase epsilon (Polε) replicates the leading strand and polymerase delta (Polδ) the lagging strand[15]. Additionally, polymerase alpha (Polα) and the RPA trimer (formed by the subunits RPA 14, 32 and 70) perform the same function as *E. coli* primase and SSB respectively. During replication termination in eukaryotes, cells face the end replication problem, given that the degradation of the last Okazaki fragment uncovers highly reactive single-stranded DNA (ssDNA) which is prone to recombination and nucleolytic reactions[16]. Thus, cells have evolved a specific structure, telomeres, which protect the end of linear chromosomes[17].

1.2 DNA transcription

DNA transcription is a fundamental process in which the cell, using a DNA template, synthetizes a complementary RNA strand. RNA is a temporary carrier of information that gets translated in a protein by the ribosome[18]. Furthermore, RNA is formed by the ribonucleotides A,G,C and U (instead of T), what changes its chemical characteristic making it more unstable and sensitive to alkaline hydrolysis[19]. Transcription initiates with the binding of an RNA polymerase to a promoter sequence. In prokaryotes, a unique RNA polymerase is regulated by the binding of different sigma factors (σ), which mediates promoter recognition and transcription initiation. A second layer of organization of transcription in bacteria is the organization of genes in arrays named operons. Operons contain a cluster of genes for a particular function and are transcribed as polycistronic messenger RNA, what allows coordinated regulation of the process. Initiation of transcription is directed by binding of the RNAP to its respective promoter aided by its bound σ factor, triggering partial unwinding of the DNA helix and followed by transcription bubble progression. During elongation, the σ factor binding to RNAP is outcompeted by a second protein, NusA, which promotes elongation and dissociates during transcription terminaton. The termination of transcription can then occur in a rho (ρ) -dependent or independent manner. In ρ-independent transcription termination, a specific transcribed sequence generates a hairpin structure which leads to RNAP release and RNA dissociation. In ρ-dependent termination, the ρ helicase loads and translocates along the RNA promoting dissociation of the polymerase and the transcript[17]. In contrast to bacteria, eukaryotic cells have three RNAPs which exert different specific functions. RNAPI transcribes pre-ribosomal RNA (pre-rRNA), RNAPII transcribes messenger RNAs (mRNAs) and non-coding RNAs (ncRNAs), and RNAPIII produces transfer RNAs (tRNAs), the 5S rRNA and ncRNAs. RNAPII is probably the most studied polymerase of them and shares high conservation with its bacterial homolog in structure, function and mechanism. The largest subunit of RNAPII, DNA-directed RNA polymerase II subunit 1 (RPB1) contains a long C-

terminal tail named carboxyl-terminal domain (CTD). RNAPII is recruited to the promoter sequence, recognized by the TATA binding protein (TBP) in a complex with other factors, in a closed state named preinitiation complex[20]. The transcription factor II H (TFIIH) drives partial unwinding of the helix and helps create the transcription "bubble", the CTD is then phosphorylated and the polymerase initiates transcription. During elongation, transcription factors are replaced by elongation factors, which confer processivity to the enzyme. During termination, elongation factors dissociate and the CTD is dephosphorylated and the RNA polymerase can then be recycled[21].

1.3 DNA repair

Faithful DNA replication, and its transcription, are essential processes for the survival of the species. Therefore, preserving genomic sequence is crucial while, at the same time, mutagenesis is indispensable for evolution. It is known that DNA is a reactive molecule and can be modified by endogenous or exogenous insults. Therefore, cells have evolved a sophisticated set of systems that coordinate for the minimization of DNA damage consequences[22]. Disruptions or deregulations in such mechanisms drive cancer and ageing[23]. Furthermore, understanding the role of DNA damage during cancer can lead to the development of novel and more effective targeted therapies[24].

1.3.1 Sources of DNA damage

DNA is subjected to thousands of lesions everyday which range in toxicity and bulkiness[25,26]. These different insults can be categorized in two groups depending on their origin, endogenous or exogenous. Endogenous damage arises during normal physiological processes, single-strand breaks (SSBs) can generate from oxidative species produced during metabolism. One of the major lesions produced by reactive oxygen species (ROS) is 8-oxoguanine, which if, unrepaired, quickly depurinates generating abasic/apurinic apimidinic sites (AP sites)[27]. AP sites are unstable and can undergo β-elimination leading to the formation of a SSB, the most common form of DNA damage[26,28]. Accumulation of SSBs in close proximity or replication of DNA templates containing such breaks leads to the formation of double-strand breaks (DSBs). DSBs are considered the most toxic form of DNA damage and, if left unrepaired, can lead to chromosomal rearrangements and ultimately cell death[29]. DNA adducts are also common lesions, with formaldehyde (together with ROS and reactive nitrogen species, RNS) being a relevant source of this type of base damage (discussed in detail in section 2.1.1)[30]. Notably, mistakes during incorporation of nucleotides or misincorporation of ribonucleotides also constitutes a constant challenge to genomic integrity. Despite having a formidable fidelity rate and 3' to 5' correcting exonuclease activity, the replicative polymerases Polε and Polδ also commit mistakes, incorporating one wrong

nucleotide every 10^6-10^7 or 10^5-10^6 respectively[31]. Moreover, the engagement of translesion synthesis (TLS) polymerases for the bypass of DNA adducts also comes at a cost. TLS polymerases have a low fidelity rate, thousand folds lower than replicative polymerases[32]. Fidelity rates can ultimately be affected by the presence of repetitive DNA regions, generating insertions or deletions in the process[33]. Spontaneous chemical decay of DNA is also a frequent source of damage, being the deamination of C the most common one and happening mostly on ssDNA during replication and transcription[34]. Deamination of C, leads to the generation of genomic U, which pairs with A instead of G, generating a C:G to a T:A transition during replication[35]. This reaction is also enzymatically performed by enzymes of the cytosine deaminase family (APOBECs and AIDs), which are mutagenic drivers in some cancer types[36-38]. Moreover, modification of 5-methyl C, an important epigenetic mark, leads to the formation of T by deamination, also causing transitions to A:T[39]. Another source of DNA damage is caused by topoisomerases activity on aberrant substrates, leading to the formation of DPCs (further discussed in section 2.2.1 and 2.2.2)[40]. Furthermore, RNA:DNA hybrids or R-loops, which can arise from stabilized transcription intermediates, have recently gained interest as a common endogenous source of DNA lesions due to the exposure of long ssDNA fragments, which are additionally recombinogenic and can cause transcription-replication conflicts (TRCs)[41-43]. Notably, the presence of selfish genomic elements in the genome, with the LINE-1 elements massively expanded in the human genome, presents a threat to genomic integrity[44]. LINE-1 mobilization has been linked to an increase in DSBs, senescence and apoptosis in cultured human cells[45,46]. However, the exact mechanism by which LINE-1-mediated DNA damage is induced is not fully understood and might be related to replication-integration conficts[47,48].

Exogenous DNA insults are organized in two main groups which purely depend on damage source: chemical or physical DNA damage inducers. Chemical DNA damage agents, comprise a plethora of compounds which react with DNA, or enzymes which interact with DNA, in an almost infinite different manner[22]. SSBs, can be generated indirectly by strong oxidizing agents such as potassium bromate ($KBrO_3$) that damage DNA bases leading to spontaneous base loss and subsequent AP site hydrolysis[49]. Base damage includes also hydrolysis and alkylation, the latter types being a frequent lesion induced by chemicals. Base alkylation is mechanistically produced by the addition of an alkyl group to highly nucleophilic ring nitrogens or, with less frequency, to the oxygen[50,51]. One example, methyl methanesulfonate (MMS) produces N7-methylguanine and N3-methyladenine, which are excised by specific glycosylases generating AP sites[52]. Alkylating agents, like temozolomide (TMZ), are of great interest due to their common use in chemotherapy[53]. Another type of alkylating agents are bifunctional crosslinkers (e.g. nitrogen mustards, discussed in section

2.1.3), able to induce intra- and inter-strand DNA crosslinks and DNA-protein crosslinks[54,55]. The most important physical agents which can damage DNA are ionizing radiation (IR) and ultraviolet radiation (UV). IR is able to directly generate SSBs with DNA breaks containing a 3'-phosphate or 3'-phosphoglycolate instead of a 3'-hydroxil group[56,57]. Additionally, multiple SSBs in close proximity lead to the generation of DSBs[58]. Furthermore, IR also indirectly damages the DNA by generating highly reactive ROS[59]. UV light is categorized depending on its wavelength in UVA (320–400nm), UVB (290–320 nm) or UVC (190–290 nm)[60]. UVC possesses maximal absorption by DNA and produces more photoproducts than the others[61]. Main lesions produced by UVC include cyclobutane pyrimidine dimers (CPDs) and 6-4 photoproducts (6-4 PPs)[60,62].
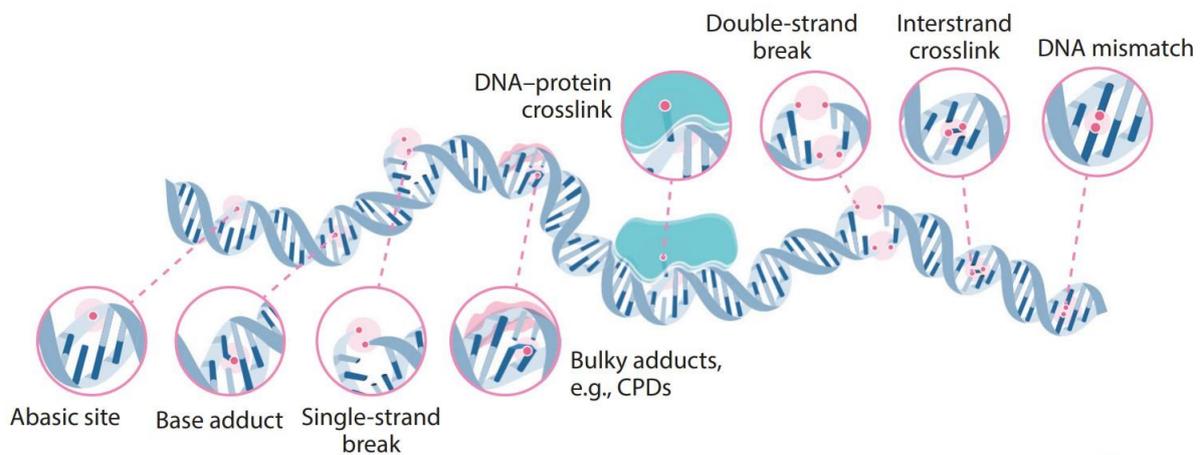

Figure 1. Schematic depiction of common DNA lesions. Adapted from[63].

1.3.2 Post-translational modifications during DNA repair

The activity of some proteins can be modulated by the addition of different chemical groups, sometimes even by entire proteins, which can trigger multiple consequences ranging from conformational changes to targeted protein degradation[21].

1.3.2.1 Ubiquitin

Ubiquitin (Ub) is a highly conserved, and essential, 76 amino acids protein which can be found in a free form or covalently attached to other proteins[64]. Ub is involved in a wide range of processes which include protein degradation, autophagy, endocytosis or DNA repair[65]. Ub is attached to a lysine residue, or the N-terminus of a protein, through its C-terminal glycine[66]. A key feature of Ub resides in its seven lysines, which can get further modified by other Ubs, generating isopeptide Ub-chains with different functions[67]. The most common chains are K48-linked[68], which generally target for proteasomal degradation, and K63-linked, which has more complex functions such as regulating the recruitment of different factors[67,69]. Ub attachment to its substrate follows a three-step process. First, the Ub-E1 enzyme (Ub-activating enzyme)

uses ATP to transfer a Ub protein to a cysteine in its catalytic centre, forming a high energy thioester bond[70]. Second, the Ub is transferred from the E1 to an E2, which is also known as Ub-conjugating enzyme, through a transthiolation reaction[71]. In a final step, a Ub-E3, also known as Ub-ligase, mediates transfer of the Ub to the target protein. Most E3-ligases belong to either the really interesting new gene (RING) or the Homologous to E6AP C-terminus (HECT) family, with the difference being that RING ligases uniquely scaffolds the ligation reaction, while HECT establishes a covalent intermediate with Ub[72,73]. The presence of Ub-binding or Ub-interacting domains (UBD and UIM) provide platforms for interaction with a variety of factors[74].

During the DNA damage response (DDR), ubiquitylation regulates the recruitment and action of many factors by the addition of mono- or diverse poly-Ub moieties. For instance, Ub is essential in the recognition of damaged sites, where modification of the histones H2A, H2B and H2AX destabilizes the nucleosome facilitating break recognition[75]. One of the most studied targets for ubiquitylation is the tumour suppressor protein p53, which is a direct substrate of the E3-ligase MDM2. The constant modification of p53 triggers its proteasomal degradation but upon DNA damage, MDM2 gets inhibited by phosphorylation, what stabilizes p53 allowing to perform its proapoptotic functions[76,77]. Another example of the role of ubiquitylation on DNA damage is the orchestrated recruitment of breast cancer susceptibility gene 1 (BRCA1) to DSBs sites by K63-linked Ub-chains on histones H2A and H2AX[78,79]. Additionally, PCNA modification by Ub determines pathway choice. Its monoubiquitylation (K164) facilitates the recruitment of TLS polymerases, but the extension of this Ub on a K63-linked chain, leads to template switching (TS)[80,81]. However, ubiquitylation is reversible by the action of deubiquitylating enzymes (DUBs). DUBs are proteases, mostly cysteine proteases[82,83], and coordinate all the physiological processes where Ub is involved[84]. In DNA damage, the ubiquitin carboxyl-terminal hydrolase 1 and 7 (USP1 and USP7) are two of the most prominent DUBs and they regulate most of the DNA repair pathways[85-87].

1.3.2.2 SUMO

SUMOylation is a vital post-translational modification catalysed by a limited set of enzymes but regulating numerous different processes. SUMO, as Ub, can be attached as a monomer or polymer to proteins[88]. While model organisms such as *Saccharomyces cerevisiae* or *Xenopus laevis* have a unique SUMO isoform, mammals have up to five[89]. As for Ub, the SUMO machinery is composed of activating-E1, conjugating-E2 and a ligase-E3 enzyme, but with a modest number of members compared to Ub[90]. Moreover, there is a unique SUMO-E2, the SUMO-conjugating enzyme UBC9 (or UBE2I), which can add SUMO in an E3-independent manner[91,92]. Additionally, SUMOylation is also reversible by the action of SUMO proteases, being the sentrin-specific protease (SENP) family the most numerous one[93]. Recognition of

SUMO modifications is mediated by SUMO-interacting motifs (SIMs), which contain a core of hydrophobic and flanking acidic aminoacids[94]. SUMOylation plays a crucial role in preserving genome stability through the regulation of extensive sets of target proteins involved in various DNA repair and genome stability pathways[95]. A good example is DSB-repair, where SUMOylation promotes the recruitment of factors such as BRCA1 or 53BP1 through the action of the SUMO-E3 protein inhibitor of activated STAT1 and 4 (PIAS1 and PIAS4)[96]. Of note, SUMO-targeted ubiquitylation by the STUbLs RNF4 and RNF111 is of great importance for the coordinated degradation of SUMOylated proteins, as well as for the repair of DPCs (extensively discussed in section 2)[97-100].

1.3.2.3 NEDD8

The neural precursor cell expressed down-regulated 8 (NEDD8) is another Ub-like modifier protein which is ligated to proteins in a chemically identical manner as for Ub and SUMO (E1, E2, E3 cascade) and via its C-terminal glycine[101]. In contrast to ubiquitylation, the NEDD8 cascade normally leads to mono-neddylation and the number of substrates of NEDD8 are very limited. The best characterized targets are cullin-RING Ub-E3-ligases (CRLs), which mediate up to 20% of total protein degradation by the proteasome[102]. Modification of CRLs by mononeddylation in their C-terminus, stimulates the recruitment of the E2 and abolishes binding to the cullin-associated NEDD8 dissociated 1 (CAND1), a native cullin inhibitor[103-106]. The role of NEDD8-mediated CRLs activity is well established in various DNA damage repair pathways including nucleotide-excision repair (NER), base excision repair (BER) and DSB-repair[104].

1.3.2.4 Phosphorylation

Protein phosphorylation is a vital cellular regulatory mechanism, as it triggers the activation or deactivation of numerous enzymes and receptors through phosphorylation and dephosphorylation processes[107]. Phosphorylation mechanisms involves the addition of a phosphate group to the polar side chain of different amino acids. This modification generally targets the hydroxyl group of threonines, serines or tyrosines, being phosphoserine the most relevant within them[108,109]. One of the most prominent roles of phosphorylation is the control of cell cycle progression, what is mainly done by the action of CDKs, also involved in transcription and splicing control[110]. In the absence of DNA damage, CDKs (such as CDK2, 4 and 6) phosphorylate retinoblastoma (Rb) allowing transition to S-phase. Rb phosphorylation is maintained during S/G2/M, inactivating the protein and promoting cell cycle progression. Rb is a tumor suppressor that commonly harbors mutations in several cancers[111,112]. In addition, the crosstalk between DNA repair factors and cell cycle progression is performed by the evolutionary conserved phosphatidyl-inositol 3′ kinase-related kinases (PIKKs), comprised of Ataxia-telangiectasia mutated (ATM), ataxia telangiectasia and Rad3-related protein (ATR)

and DNA-dependent protein kinase (DNA-PK), and PIKK-regulated downstream kinases, Checkpoint kinase 1 and 2 (CHK1 and CHK2)[113]. PIKKs play an essential role in the coordination of DNA repair, fork stability and cell cycle control through the phosphorylation of hundreds of substrates[114,115].

1.3.3 Direct, base excision and mismatch repair

Nature has evolved several mechanisms in which specific DNA damages can be directly removed without affecting the DNA backbone. This process is commonly known as direct-repair and comprises the repair of three major lesions: UV-light induced, O6-alkylguanine and N-alkylated base adducts. The two main UV photoproducts, CPDs and 6-4 PPs, can be targeted by the action of photolyases. These enzymes, which are absent in humans, use blue or near-UV light and the cofactor flavin adenine dinucleotide (FAD) in its reduced form to split the base dimer[116-118]. The repair of O6-alkylguanine lesions through direct reversal has gained interest in the recent years due to the use of temozolomide, which mainly induces this lesion, in cancer treatment[52,53]. A unique protein, methylated-DNA-protein-cysteine methyltransferase (MGMT) also known as alkyl-guanine transferase (AGT), acts as a recipient for the alkylation by transferring it from the base to its catalytic cysteine[119]. The S-alkylated MGMT cannot be restored and undergoes fast degradation by the ubiquitin-proteasome system[120]. Of note, activity of MGMT on lesions produced by ethylene dibromide, a known carcinogen, paradoxically increases its toxicity by the formation of a MGMT-DPC[121]. The last type of direct DNA repair is mediated by the AlkB family of demethylases, which using an iron site, activate the oxygen molecule to catalyse the oxidation of aberrant alkyl groups. This reaction generally targets methyl groups on N1-adenine or N3-cytosine, and yields the unmodified base and formaldehyde as a by-product[122-124].

BER is a process for the repair of small lesions which do not distort the DNA structure. There are two main BER sub-pathways which depend on the length of synthesis after lesion removal: short-patch BER, for a single nucleotide, and long-patch, when the synthesis is longer than one nucleotide. BER initiates by cleavage of the damage base by a glycosylase, which results in the formation of an AP site[125], by monofunctional glycosylases, or a SSB, by a bifunctional glycosylase. Monofunctional glycosylases include the uracil-DNA glycosylase (UDG) or N-Methylpurine-DNA glycosylase (MPG), which removes numerous alkyl-base adducts[126]. On the other hand, bifunctional glycosylases include the 8-oxoguanine glycosylase 1 (OGG1), important for the repair of the most prominent oxidative lesion in cells, which directly incises the AP with its lyase activity generating a 3'-α, β-polyunsaturated aldehyde (PUA) and a 5'-phosphate[127]. The AP site generated by monofunctional glycosylases as well as the 3'-PUA generated by bifunctional glycosylases as OGG1, must be removed for proper repair of the lesion. This reaction is catalysed by AP-endonuclease (APE1), which cleaves the DNA-

backbone 5', next to the AP site or the 3'-PUA, and generates a gap with a 3'-hydroxyl and a 5'-deoxyribose phosphate (5'-dRP) flap[128-130]. After APE1 cleavage, polymerase β (Polβ) is recruited to the lesion and performs its double enzymatic function by the addition of a nucleotide to the 3'-end, polymerase activity, and removal of the 5'-dRP, mediated by its lyase activity[131,132]. The ligation of the ends by DNA ligase I or III (LIG1/3), after addition of a phosphate group to the 5' end by polynucleotide kinase/phosphatase (PNKP), constitutes the final step of short-patch BER[126]. Long-patch BER seems to be favoured after the activity of bifunctional glycosylases, where APE1 cleavage of the 3'-PUA leaves a clean 3'-OH and 5'-phosphate flanked by a single nucleotide gap, but also during replication due to the availability of long-patch repair factors or by the presence of 5'-adducts that Polβ cannot remove[126,133]. In this alternative process, PCNA is loaded on the gap with a replicative polymerase (Polε or Polδ, but can use Polβ in non-replicative cells), and after strand synthesis, normally of 2-12 nucleotides, the structure generated is cleaved by Flap structure-specific endonuclease 1 (FEN1), what allows LIG1 ligation and break sealing[126,134]. The exact moment of intervention of the poly-ADP-ribose polymerase 1 (PARP1) and the X-ray repair cross complementing 1 (XRCC1) is a matter of debate. PARP1 is known as a first responder because it binds AP sites and SSBs which are intermediates of BER[135-137]. After lesion recognition, PARP1 undergoes heavy modification (poly-ADP ribosylation or PARylation) of itself and target proteins, what recruits XRCC1 and forms a platform for the further recruitment of more factors implicated in BER[136].

The DNA mismatch repair (MMR) process is essential for maintaining the stability of DNA within cells. It works to rectify errors like mismatched base pairs and small insertions or deletions, which commonly emerge during DNA replication[138,139]. MMR involves four steps: recognition of the lesion, initiation of repair, excision and resynthesis of DNA[140]. Furthermore, MMR also has a preferential bias to act on actively transcribed genes. The two main groups of MMR proteins are the MutS homologs (MSH) and the MutL homologs (MLH)[141-143]. Human MMR machinery comprises eight genes which code for proteins that associate forming heterodimers, what changes its specificity towards lesions[142,143]. For example, MSH2 forms dimers with MSH3 or MSH6 and correct mismatches in the newly replicated DNA strand, by sliding through the DNA until they recognize the distortion caused by the mismatch[144]. The recognition of a mismatch initiates binding of other molecules, generally the heterodimer MLH1-MLH4, having the second endonuclease activity. Endonucleolytic cut by MLH4, generates a SSB that allows exonuclease 1 (EXO1) activity and dissociation of the complex from DNA[140,145,146].

1.3.4 Nucleotide excision repair

NER is the most versatile repair pathway because it can remove a great range of DNA-helix distorting lesions. Some of these lesions include CPDs, 6-4 PPs or DNA bulky adducts[147]. The two main NER sub-pathways only differentiate in the lesion recognition process which, if not involving transcription is named global-genome NER (GG-NER), while if it involves actively-transcribed DNA strands is referred as transcription-coupled NER (TC-NER)[147,148]. Mutations in GG-NER and general NER genes are causative for xeroderma pigmentosum, while mutations in TC-NER related genes exclusively cause Cockaine syndrome[149]. The first element in DNA damage recognition by GG-NER is xeroderma pigmentosum factor C (XPC), which forms a complex with UV excision repair protein RAD23 homolog B (RAD23B) and centrin 2 (CETN2)[150]. This complex can detect helix distortions by a sequential mapping step which involves the XPC damage-independent DNA-binding domain and diffusion along DNA by continuous association and dissociation in a process known as "hopping"[151]. The lower energy of unpaired nucleotides -due to the presence of a lesion- allows XPC to flip-out damage-containing nucleotides and open their conformation[152,153]. XPC is inefficient in the recognition of UV-damages because they do not dramatically distort the helix. Therefore, a second protein, UV-damaged DNA-binding protein (UV-DDB), a heterodimer formed by DDB1 and DDB2 (also known as XPE), introduces a kink next to the lesion, favouring further recognition by XPC[150]. Moreover, DDB1 recruits an E3-ligase, CUL4-RBX1, which ubiquitylates XPC increasing its DNA-binding activity[154]. The binding to the lesion by XPC engages XPB binding, which further recruits the general transcription factor IIH (TFIIH)[150].

In contrast to GG-NER, TC-NER initiates with the stalling of RNAPII on a DNA lesion during transcription elongation. The Cockayne syndrome group B protein (CSB) is a DNA-dependent ATPase which enhances its RNAPII binding upon stalling, driving the recruitment of a second factor, CSA[155]. CSA is a substrate recognition factor of a Ub-E3-ligase complex which, upon recruitment, catalyses the polyubiquitylation of RPB1 on lysine K1268, triggering its degradation and the recruitment of an accessory protein, the UV-sensitive syndrome protein A (UVSSA), which further facilitates the recruitment of TFIIH[156,157].

TFIIH plays a key role in damage verification and unwinds the DNA to expose the lesion, allowing assembly of the repair machinery[148]. The increase of DNA flexibility by TFIIH activity, permits XPD to unwind the DNA and filter the damage by size, allowing for the detection of bulky lesions[147,150]. Lesion filtering process recruits XPA and XPG, which stabilize the opened conformation, expel TFIIH and permit RPA association with the undamaged ssDNA, protecting it from nuclease attack[158,159]. When XPD encounters the damage, XPG displaces XPC and allows recruitment of the pre-incision complex XPF-ERCC1, recruited by XPA[160,161]. In the final step of the process, XPF-ERCC1, positioned on the 5' end, and XPG, positioned on the 3'

end, perform strand incision and release the lesion-containing DNA, which is approximately 30 nucleotides long[161-163]. Downstream processing includes the same factors as for long-patch BER, which extend the excised strand and religate the ends (e.g. PCNA, XRCC1, replicative polymerases, etc.)[164].

1.3.5 Single- and double-strand break repair

As previously mentioned, SSBs are one of the most common lesions in cells, what requires a rapid removal system to avoid their accumulation[25]. The repair of SSBs overlaps with other repair pathways which have SSBs as intermediates such as BER or NER[165]. Therefore, during this section, I will mainly focus on the role of PARP1 and associated factors, important for the repair of SSBs. PARP proteins (PARP1 and PARP2) rapidly detect SSBs in cells, what triggers their catalytic activation[166,167]. Using NAD^+ as a cofactor, these proteins trigger self-PARylation and catalyse the PARylation of other proteins in the vicinity. This activity also controlled by the histone PARylation factor 1 (HPF1), which facilitates the modification of serine residues on histones near the break[168,169]. Although the role is currently unclear, new studies suggest that also DNA and RNA can be directly PARylated[170]. PAR-modification has a short half-life and is rapidly removed by the poly-ADP-ribose glycohydrolase (PARG), while mono-ADP-ribosylation, mostly catalysed by PARP3, is removed by mono-ADP-ribose hydrolases[171]. Modification of the environment around the break by PARPs constitutes an extremely dynamic signalling, which likely controls chromatin accessibility to allow the recruitment of downstream factors[172-174]. In addition to the effect of PARylation on DNA opening[174], the modification also recruits chromatin remodellers such as the chromodomain-helicase-DNA-binding protein 1-like (CHD1L, also known as ALC1)[175,176] or the facilitates chromatic transaction complex (FACT). A second major role for PAR on chromatin is the direct recruitment of XRCC1, which forms complexes with almost all the necessary enzymes for the repair of SSBs[177]. In contrast to the role of direct recruitment of XRCC1 to SSBs, the need for PARylation during BER seems less obvious, given that BER intermediates can be repaired by Polβ which directly allows ligation by XRCC1/LIG3[178-180]. The recruitment of XRCC1 starts a negative feedback loop which "disengages" PARP1 from DNA and acts as an "antitrapping" factor, what ends the process and allows the completion of SSB-repair, either long- or short-patch[181].

DSBs are considered the most toxic form of DNA damage and the DSB repair system is tightly regulated to counter the effect of such deleterious insults[182]. DSBs are predominantly repaired through non-homologous end joining (NHEJ), an error prone pathway which can religate virtually any kind of DSB. An alternative during the repair of DSBs is homologous recombination (HR), a higher fidelity pathway which uses a second allele copy as a template for repair[183]. The cell cycle phase, the availability of certain factors and the presence of a homologous template determines pathway choice.

1.3.5.1 Non-homologous end joining

NHEJ can be itself divided depending on whether there is end resection by nucleases or not. Therefore, three main NHEJ branches can be found in cells: classical NHEJ (cNHEJ) with no resection, alternative NHEJ (altNHEJ) and single-strand annealing (SSA), these two latter ones requiring resection.

cNHEJ initiates with the binding of the heterodimer Ku70-Ku80, which interacts with the DNA ends and nucleates the recruitment of factors such as DNA-PK and LIG4, including the scaffolding factors XRCC4, PAXX and the XRCC4-like factor XLF[184-186]. DNA-bound Ku is an activator of DNA-PK, which triggers phosphorylation of many targets, including itself, and also serine S139 in histone H2A.X for phosphorylation (γH2AX), done by ATM or ATR in alternative DSB repair pathways, which opens the chromatin surrounding the break[187-190]. The recruitment of these factors generates a synaptic complex which aligns the two DNA ends, with its core consisting of XRCC4, XPF and PAXX, allowing ligation by LIG4[186,191]. Although LIG4 can tolerate certain base damages and mismatches[192], many DNA ends need reshaping for ligation, what is performed by enzymes such as PNPK, polymerases such as Polμ and Polλ, or nucleases such as Artemis[193].

End-resection mediates the start of two alternative forms of NHEJ, altNHEJ (also known as microhomology-mediated MMEJ) and SSA[194]. End-resection is determined by the presence of the MRN complex and the phosphorylation of C-terminal-binding protein-interacting protein (CtIP), activated in the S-G2 phase of the cell cycle by cyclin dependent kinases[195,196]. The MRN complex, which requires activity stimulation by CtIP, is composed of three core units: Meiotic Recombination 11 (MRE11), RAD50 and Nijmegen breakage syndrome 1 (NBS1)[196,197]. The MRN complex stimulates the activity of ATM at the end of the DSB, and γH2AX formation will recruit the mediator of DNA damage checkpoint 1 (MDC1), an adaptor protein which will further allow RNF8, an E3-ligase, to catalyse K63-polyubiquitylation on histone H2A[187]. These K63-chains are further amplified by RNF168[78,198]. K63-Ub-Linked H2A, recruits the BRCA1 complex, outcompeting the binding by P53-binding protein 1 (53BP1) that inhibits resection by binding the dimethylated H4K20, the monoubiquitylated H2AK15 and γH2AX, which are important for the repair of DSBs in heterochromatin. The competition between 53BP1 and BRCA1 has been subjected to extensive study given that *53BP1* loss can rescue repair by HR in *BRCA1* deficient cells[199]. An additional layer of control on resection is the inhibition of 53BP1 and RNF8 by phosphorylation by checkpoint kinases, what controls their activity throughout the cell cycle[200,201]. During altNHEJ, Polθ, a specialized polymerase with helicase function, displaces RPA from the resected ssDNA via its helicase domain[202,203]. In addition, PARP1 performs an essential function in the synapsis of the ssDNA strands, which is mediated by microhomology – of two to twenty nucleotides – and can be extended by

Polθ[204]. The flap generated during the bridging by PARP1 and the extension of the microhomology by Polθ can be cleaved by APE2, which not only possesses flap activity, but also can nucleolytically remove different 3'-end blocks[205]. In a final step, ends are ligated by XRCC1-LIG3 complex, recruited by PARP1 to the lesion site[206]. In contrast to altNHEJ, SSA requires larger homology, causing large deletions in repetitive sequences[207]. Annealing and homology search is mediated by RAD52, while the nucleolytic incision of the flaps is mediated by the ERCC1/XPF complex[208-210]. The exact polymerase directly implicated in SSA is still not known but it probably involves Polθ, Polδ or Polζ[211,212].

1.3.5.2 Homologous recombination

After resection, the presence of a sister chromatid will provide HR preference[213]. Although considered high-fidelity, when compared with NHEJ, HR can also have deleterious consequences such as loss of heterozygosity if the homology template is the homologous chromosome[214]. Despite having multiple subpathways, the initiation of HR is always the loading of the recombinase RAD51 by BRCA2 on the resected DNA. During this process, RPA is displaced progressively and the RAD51 filament formation allows the search for sequence homology[215]. The partner and localizer of BRCA2 (PALB2), as well as the BRCA1 and BRCA1-associated RING domain protein 1 (BARD1), aid in the formation of a D-loop structure by stimulating RAD51-ssDNA filaments ATPase activity[216,217]. The free 3'-end of the invading strand can be extended by Polδ using the invaded strand as a template[211]. Of note, the pathway can subdivide in this step depending on active or passive involvement of the second DNA strand in the process. If the invading strand is extended while the non-invading is not, the invading strand reanneals after polymerase activity, serving as a template for the extension of the non-invading strand, a process known as synthesis-dependent strand annealing (SDSA)[218]. In contrast, both DNA strands can get extended if the non-invading strand anneals with the displaced strand of the homologous sequence, a process known as classical HR, and which is prominent in meiotic recombination[218]. Classical HR involves the formation of a double Holliday junction which can be dissolved, this means with no cross overs, by the activity of the Bloom helicase (BLM) and the BTR (BLM, TOP3α, RMI1 and RMI2 complex[219]. Alternatively, crossover can occur by the action of structure specific nucleases or resolvases such as the essential meiotic endonuclease 1 homolog 1, 2 (EME1 and EME2) and methansulfonate, UV sensitive 81 (MUS81)[219]. A final mechanism, known as break-induced replication (BIR), can also occur after strand invasion. BIR is a highly mutagenic process and happens primarily during replication, where after strand invasion the invading strand is extended until the end of the chromosome, even for hundreds of kbps, unless stopped by a blockage[220,221]. In addition, the exposure of such long ssDNA tracts, subjects them to extensive DNA damage, what can drive mutagenicity, and severe chromosomal rearrangements[222]. BIR has acquired high

relevance during recent years because it provides cancer cells, which lack telomerase, with an alternative pathway to extend their telomeres, a process known as alternative lengthening of telomeres (ALT)[223].

1.3.6 DNA interstrand crosslink-repair, the Fanconi anemia pathway

Fanconi anaemia (FA) is a rare disorder which is caused by mutations with partial or total loss of function in, at least, one of the 22 FA genes described up to date[224]. The cause for the disease is probably the incapability of these cells to repair interstrand DNA crosslinks (ICLs), what also causes that FA patients present higher cancer rate[225,226]. The repair of ICLs by FA generally requires S-phase and is triggered by convergence of replication forks at the ICL[227]. ICLs can be recognized after fork stalling, triggering binding of the FA anchoring complex, which contains the translocase FANCM. FANCM allows most replication forks to bypass the ICL so it can be repaired postreplicatively[228]. A second option, is excision of the ICL by the NEIL3 glycosylase. Not all ICLs can be targeted by NEIL3 and its cleavage on the N-glycosyl bonds releases an AP site which is bypassed by TLS[229]. Upon binding of FANCM, in cooperation with FAAP24 and MHF, the FA core complex, which contains the E2-conjugating enzyme FANCT (or UBE2T) and the E3-ligase FANCL, is recruited to the damage[230]. Recruitment of the core complex mediates monoubiquitylation of the FANCD2 and FANCI heterodimer, which is enhanced by DNA binding, by these enzymes[231,232]. In addition, loss of any of the factors in the anchoring complex does not completely avoid FANCI and FANCD2 modification, suggesting alternative pathways for the recruitment of the core complex[233]. Functional consequences of the monoubiquitylation of the ID2 complex (FANCD2 and FANCI), reversible by the action of USP1, include the recruitment of structure-specific nucleases by FANCP (SLX4), which mediate the unhooking of the ICL (e.g. XPF-ERCC1, MUS81-EME1 or SLX1), and TLS polymerases to bypass the ICL-bound dinucleotide[234-240]. The final unhooking step drives the generation of DSBs, which are mostly repaired by HR factors, many of which are also FA proteins (e.g. BRCA1 is FANCS)[240,241].

1.3.7 R-loops, TRCs and their resolution

R-loops have crucial physiological roles in cells such as chromatin organization, DNA replication and repair. However, R-loops have gained interest as potential sources of DNA damage and genomic instability, highlighting the relevance of their regulation[242]. One of the main drivers of R-loop toxicity is mediated by replication-hybrid collision, which can be subdivided in head-on, when the CMG helicase encounters the RNA polymerase, or co-directional, being less toxic due to the CMG capacity of unwinding also RNA:DNA hybrids[243]. Head-on collisions can stall the replication fork, leading to ATR activation and fork reversal, which also requires MUS81 activity[244]. Notably, RNase H1 and H2, are also important for R-loop resolution because they have endoribonuclease activity at specific DNA:RNA hybrid

structures[245,246]. Additionally, a large group of helicases have recently been involved in R-loop suppression, and some examples include senataxin, aquarius and members of the DEAD-box and DEAH-box helicase families[247-250].

1.4 DNA damage tolerance

The majority of DNA damage is removed efficiently but some lesions persist until DNA replication, driving drastic consequences for the cell such as fork stalling and collapse or chromosomal breakage[251]. DNA-damage tolerance pathways (DDTs) cope with this problem, allowing replication of the damaged template and reducing genomic instability. DDTs are conserved in almost all organisms, highlighting their evolutionary importance[251,252].

1.4.1 Translesion synthesis

Despite tolerating certain base lesion damages, the replicative polymerases Polε and Polδ stall when they encounter bulky lesions. In this context, specialized polymerases are recruited for direct synthesis prior to recoupling[253]. Main TLS polymerases in humans belong to the Y-family and are polymerase η (Polη), ι (Polι), κ (Polκ) and the reversionless 1 (REV1), including a B-family member, polymerase ζ (Polζ)[86,254]. TLS polymerases make loose and non-specific contacts with the DNA template, what allows them to accommodate a wide range of DNA distortions and bulky lesions. The use of TLS polymerases also comes at a cost, due to their low fidelity, what explains why their recruitment is tightly regulated[255]. Different TLS polymerases seem to have different efficiencies towards diverse damages. Polη is efficiently replicating across CPDs while it is uncapable of doing so with 6-4 PPs[256-258]. Furthermore, Polι, which arose as a RAD30 (yeast Polη homolog) gene duplication, has partially overlapping functions with Polη with respect to UV-damage but has additional functions[259-262]. Polι also possesses lyase activity and interacts with XRCC1, what allows it to substitute *in vitro* Polβ function, although *in vivo* data are still lacking. Polκ is the most conserved Y-family polymerase with homologs found in almost all domains of life but curiously not in yeast[263]. Despite being blocked by dinucleotide lesions, Polκ can bypass a large variety of monoadducts such as N2-dG alkylation[264,265]. Additionally, Polκ can also extend a C inserted by Polι in certain lesions[266], although the extension process seems to be primarily performed by Polζ[267]. Of note, the role of Polκ is not limited to lesion bypass, given that Polκ also performs gap-filling during NER in cooperation with Polδ[268]. REV1 is a C polymerase with an unusual activity[269,270]. Instead of pairing with G, it acts as a template-independent transferase by swinging the DNA, pairing the G with a loop in its own structure and coordinating the C with one of its residues so it can be placed in the right position[271,272]. This allows REV1 to bypass numerous G-adducts and AP sites[269-271,273]. In contrast to the previously described Y-polymerases, Polζ belongs to the B-family, closely related to the replicative polymerases α,δ and ε[274]. Its core is composed of

REV3, which harbours the catalytic activity, two REV7 subunits, which stimulates REV3 activity also linking Polζ to REV1[275-277]. In addition, as Polδ, REV3 also associates with Pol31 and Pol32, suggesting common features in the recruitment and regulation of both polymerases[278]. The role of Polζ is thought to be the extension of the first nucleotide placed opposite the damage by Y-polymerases, a process which is error-prone but still more accurate than replication by Y-polymerases, what might explain its involvement in the repair of almost all damage-induced mutagenesis[255,279].

Polymerase switching is tightly controlled by Polδ, which outcompetes TLS polymerases to limit mutagenesis[253]. TLS recruitment is triggered by persistent RPA filaments on ssDNA, generated during fork stalling, what recruits the RAD6-RAD18 E2-E3 complex, catalyzing PCNA monoubiquitylation and Polδ to TLS-polymerase switch[280-282]. The recruitment via Ub is mediated by specific Ub-binding domains in each protein such as Ub-binding zinc-fingers (Polκ and Polη) or Ub-binding motifs (Polι and REV1), which work in addition to their PIP-boxes (REV1 has no PIP-box but interacts through its BRCT domain)[283]. REV1 also binds the REV7 subunit of Polζ and the other Y-polymerases through PIP-like motifs, aiding in their recruitment in a process that seems to be lesion specific[284-287].

1.4.2 Repriming

The uncoupling of the replication fork can also be rescued by repriming. Repriming consists of the synthesis of a new primer downstream of the lesion, allowing replication on the leading strand to continue[287]. Although Polα could potentially fulfil this role, it has proven to be extremely inefficient in the synthesis of primers on the leading strand due to its strong inhibition by RPA[288]. Although during RPA exhaustion Polα could potentially perform this function, in higher eukaryotes, repriming is mediated by another primase termed PrimPol. PrimPol is directly recruited by RPA to the ssDNA generated by uncoupling, where it mediates not just priming but also extension downstream of the lesion, due to its double activity as primase and polymerase[289]. While in some cells lines, such as avian DT40 cells, PrimPol loss causes damage sensitivity, growth arrest and fork slowing[290], in human cells, its deletion does not result in major consequences, although the additional loss of Polη or Polζ does have a synergistic effect towards damage sensitivity[291,292]. PrimPol seems to be relevant for providing tolerance towards damages which are not bypassable by TLS, such as chain-terminating nucleosides or ICLs[292,293]. This might suggest that if the efficiency of TLS damage-bypass is lower than repriming, the second would take place[80,287]. After repriming, Polδ-interacting protein 2 (PolDIP2) enhances primer extension activity by PrimPol to then facilitate the switch to Polδ, what promotes recoupling[253,294,295].

1.4.3 Fork reversal

Fork reversal generates a "chicken foot" structure at active replication forks in response to DNA damage. The reverse fork is protected from degradation and permits a controlled stalling of replication. Different outcomes can follow the formation of a reversed fork[296]. The nascent leading strand can be extended using the lagging strand as template[297]. A second function of fork reversal might be the use of the lagging strand as substrate for the repair of the leading strand on a dsDNA context[297]. Fork reversal is initiated by monoubiquitylation-mediated polyubiquitylation of PCNA by helicase-like transcription factor (HLTF), which has helicase and E3-ligase activities, or SNF2 histone linker PHD RING helicase (SHPRH), a second E3-ligase[298,299]. The presence of RPA and polyubiquitylated PCNA recruits two translocases: the SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A-like protein 1 (SMARCAL1), and zinc finger RANBP2-type containing 3 (ZRANB3)[300-302]. The translocases will reshape the fork in a mechanism not yet fully understood and will generate the characteristic "chicken foot" structure, which is directly observable under the transmission electron microscopy in human cells treated with genotoxic agents[303]. As for TS, reverse fork generation requires RAD51 that also aids in protecting the reverse forks from degradation[304,305]. After fork reversal and extension of the leading strand, the RecQ-Like Type 1 (RECQ1) can catalyse reversed-branch migration and fork restart, or alternatively, the fork can suffer a controlled resection by the Werner Syndrome helicase in a complex with the DNA replication ATP-dependent helicase/nuclease DNA2 (DNA2), promoting repair of the lesion and restart[306,307]

2 DNA-protein crosslinks

DPCs are highly toxic DNA lesions which arise when proteins get covalently trapped on DNA. Their toxicity relies on their ability to block essential chromatin processes such as replication and transcription. In the following section, I will describe in detail sources and repair of these pervasive lesions and the consequences of defective DPC-repair.

2.1 Non-enzymatic DNA protein crosslinks

Non-enzymatic DPCs are generated by reactive metabolites and chemotherapeutics which can crosslink virtually any DNA-binding protein. In the following subsection, I summarize some of the main DPC-inducers which have pathophysiological and clinical relevance.

2.1.1 Reactive aldehydes

Reactive aldehydes are common byproducts of endogenous metabolism. Acetaldehyde, for example, drastically increases its cellular levels after ethanol consumption. Ethanol is rapidly oxidized by alcohol dehydrogenases, which use NAD^+ as a cofactor, and the acetaldehyde generated is detoxified by the acetaldehyde dehydrogenase 2 (ALDH2)[308]. Mutations in ALDH2 are causative for Asian flush syndrome and higher cancer predisposition[309]. Acetaldehyde induces various forms of DNA damage, mainly DNA adducts such as N2-ethylidene-2′-deoxyguanosine, N2-ethyl-2′-deoxyguanosine, N2-propano-2′-deoxyguanosine, and N2-etheno-2′-deoxyguanosine[309]. N2-propano-2′-deoxyguanosine is a crosslink precursor, and can react with the N2-amine of a second guanine generating an ICL[310]. Additionally, proteins in the vicinity can get crosslinked to N2-ethyl-2′-deoxyguanosine generating DPCs[311].

Formaldehyde is one of the most relevant, reactive and poisonous aldehydes and has been exploited for decades for the preservation of biological samples. It also constitutes one of the most relevant environmental toxins and can be found in polluted air, water or tobacco smoke[312]. Endogenous formaldehyde is generated during one-carbon metabolism, where serine is cleaved to glycine and formaldehyde by serine hydroxymethyl transferase[313]. Lipid oxidation also contributes to elevated cell formaldehyde during malondialdehyde metabolism[314] and the degradation of adrenaline, which produces methylamine further deaminated into formaldehyde[315]. Furthermore, formaldehyde is produced in high concentrations within chromatin during histone[316] and DNA demethylation[317], generating high concentration waves during transcriptional reprogramming[318], but also in the cytoplasm upon RNA demethylation[319]. Moreover, methanol, either from endogenous or exogenous sources, is rapidly converted in cells to formaldehyde by alcohol dehydrogenases[320]. The first tier of

protection against formaldehyde is constituted by two enzymes, alcohol dehydrogenase 5 (ADH5) and ALDH2. ADH5 conjugates formaldehyde with glutathione generating S-formylglutathione, excreted in the urine. ALDH2 oxidizes formaldehyde to formic acid using NAD^+ as a cofactor, generating reductive power in the process[321]. An additional enzymatic system for the degradation of formaldehyde includes the cytochrome P450 monooxygenase superfamily, which conjugates atomic oxygen to the substrate, generating formic acid. Atomic oxygen is obtained by reducing one of the atoms in molecular oxygen with NADPH as a cofactor[322]. The second tier of protection is constituted by a broad range of DNA-repair mechanisms which mainly involve the FA pathway[310,318] but also DPC-specific proteases such as SPRTN[323] or TC-NER factors as CSB[324]. Dependency on certain repair pathways to repair lesions induced by formaldehyde, can be explained by the type of damage produced by it. Formaldehyde forms, due to its electrophilic nature, covalent adducts with amino- and thiol-groups[325]. The formation of Schiff base intermediates with DNA, RNA and proteins drives the generation of mono-adducts, nucleic acid crosslinks and protein-protein crosslinks[55,311,326]. Interestingly, the reactivity of formaldehyde with proteins and DNA was already described several decades ago, where the authors also hypothesize the existence of a protease-dependent pathway for the repair of DPCs[1]. The efficient generation of DPCs by formaldehyde is widely used for different molecular biology techniques, which seek to capture interactions between proteins and nucleic acids, such as chromatin immunoprecipitation sequencing (ChIP-Seq)[327] or RNA interactome capture method (RIC)[328,329]. One very important feature of these lesions is that despite being induced really efficiently, they are reversible through spontaneous hydrolysis, what limits their toxicity and facilitates their application in the listed methods[330].

Of note, AP sites are one of the most relevant endogenous aldehydes. AP sites are able to crosslink to cysteines and lysines. The formation of such crosslinks (e.g. between histones and DNA) has already been described in the literature, posing AP sites as one of the major drivers of aldehyde-mediated DPC-formation in cells due to their abundance[311,331,332].
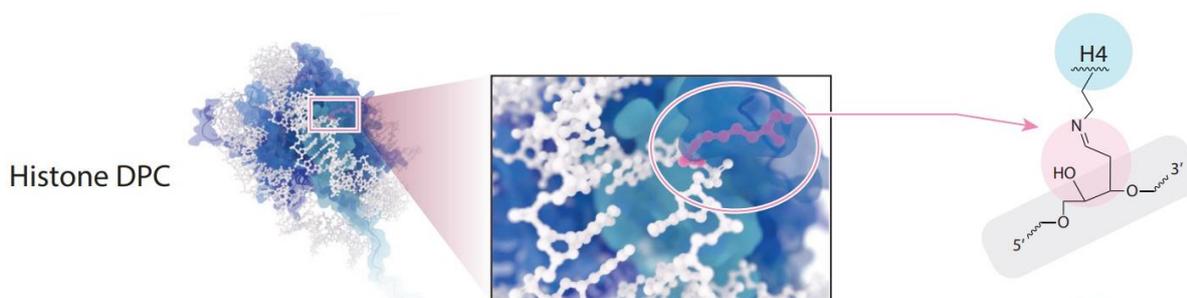

Figure 2. Schematic depiction of the chemical structure of a histone H4 crosslinked to an AP site. Grey colour represents DNA backbone, light blue the protein crosslinked and pink the chemical bond between the protein and DNA. Adapted from[63]

2.1.2 Platinum-derived compounds

Platinum compounds such as cisplatin (or cis-diamminedichloroplatinum (II)) are frequently used in chemotherapy as anticancer drugs. The major toxicity-derived mechanism of these drugs is the formation of DNA-DNA crosslinks (mainly 1,2-GG and 1,2-AG) but they can also generate DPCs[333]. DPC-formation by platin compounds involves the bridging of the N7-guanine in DNA with cysteines, arginines and lysines in protein side chains[333,334]. Of note, cisplatin is more effective in the generation of these lesions than the other commonly used drug, and its isomer, transplatin (trans-dichlorodiammineplatinum (II))[55,333].

2.1.3 Nitrogen mustards

Nitrogen mustards belong to one of the oldest groups of chemicals used in chemotherapy. One example, mechlorethamine can alkylate nucleophilic sites in proteins and DNA, what facilitates their bridging and DPC-formation[335]. The N7 substituent of guanine and cysteines in proteins side chains have been shown to be responsible for the main DPC-generation capacity of the compound[336,337]. Despite being able to form DPCs, nitrogen mustards' toxicity seems to be mainly mediated by DNA-DNA crosslink formation[55,338].

2.1.4 Radiation

In addition to the generation of base damage and strand breaks, IR and UV have also been described as DPC-generating agents[339,340]. Direct base damage by IR and UV, as well as water-generated reactive species, form unstable DNA and protein radicals which can result in covalent interactions between them generating DPCs. Interestingly, DPC-formation by IR is enhanced in the absence of oxygen, known as the reverse oxygen effect, highlighting the importance of these lesions during radiotherapy in solid tumours[341-345].

2.2 Enzymatic DNA-protein crosslinks

Some enzymes establish covalent intermediates with DNA during their physiological activity. The stabilization of these intermediates either by specific inhibitors, or by DNA distortions, generates pathological covalent complexes, which are classified as enzymatic DPCs.

2.2.1 Topoisomerase I

Topoisomerases are present in all domains of life and resolve different topological problems caused by genome length, folding and physical processes affecting its structure (e.g. replication, transcription or chromatin remodelling)[346]. Topoisomerases cleave and religate the DNA backbone establishing covalent intermediates through their active site tyrosine and a phosphate at the end of the DNA molecule. Topoisomerase I (TOP1) cleaves DNA introducing a SSB, remains attached to the 3'-end and relieves DNA supercoils by controlled rotation around the non-cleaved strand[40]. The reason why topoisomerases cleavage complexes

(TOP1 and TOP2ccs) are transient, is the nucleophilic behaviour of the deoxyribose hydroxyl group towards the tyrosyl-DNA bond. In humans, TOP1 is an essential gene, and its molecular function is particularly important during transcription, removing negative and positive supercoiling generated by the transcription bubble[347,348], as well as during replication, where it also removes positive supercoiling, being particularly important in ribosomal gene arrays[349,350]. TOP1cc stabilization has been historically used in chemotherapy after the discovery of a compound, CPT, which intercalates in the enzyme-DNA interface. CPT is an alkaloid firstly isolated from a rapidly growing tree from China, *Camptotheca acuminata Decne*[351]. Interestingly, this plant's TOP1 encodes a mutation, N722S, closed to TOP1 active site (Y723) that makes it resistant to its self-produced alkaloid[352]. This mutation was also found in CEM leukaemia cells after CPT treatment, constituting a resistance mechanism of tumours to TOP1 poisons[353]. New generation TOP1 poisons, such as topotecan and irinotecan, are currently used in chemotherapy due to their higher stability and improved bioavailability, and are derivatives of CPT[354]. In addition to CPT-induced trapping, TOP1-DPCs can be formed by the action of the enzyme on faulty DNA-templates. The misincorporation of ribonucleotides in DNA[355,356], mismatches, SSBs and DSBs, AP sites or base oxidative damages are examples of nearby lesions that can disturb TOP1 activity resulting in DPC formation[357,358]. Notably, topoisomerases DPCs are so frequent, that cells have evolved specific enzymes to cope with them (discussed in section 2.4.1).

Figure 3. Schematic depiction of the chemical structure of a TOP1cc. Grey colour represents DNA backbone, light blue the protein crosslinked and pink the chemical bond between the protein and DNA. Adapted from[63]

2.2.2 Topoisomerase II

TOP2 as TOP1, establishes covalent intermediates with DNA to relieve torsional stress[40]. In contrast to TOP1, TOP2 is a homodimer and after DNA cleavage remains bound to the 5'-end, which generates a DSB with 5'-base overhang[359]. TOP2 has two main isoforms in humans, TOP2α and TOP2β, and while the α isoform is required during mitosis[360], the β isoform activity seems to be more important for transcriptional regulation[40]. Additionally, while TOP2α is important for the resolution of highly positive supercoiled DNA, TOP2β is also able to resolve negative supercoiling[361], being both isoforms involved in the resolution of structures

such as DNA catenates or knots[346,362,363]. Formation of TOP2-DPCs is caused by similar DNA damages than those which trap TOP1 (e.g. AP sites, mismatches or base damage)[364-366]. Moreover, DPC formation is also forced by chemotherapeutics, such as etoposide, doxorubicin or mitoxantrone, that stabilize the covalent enzyme-DNA intermediate. These drugs act in a similar fashion as CPT for the generation of TOP1-DPCs[367]. In addition, it remains to be tested whether TOP2β-DPCs are the source of the observed TOP2β-dependent DSB formation during transcriptional activation of the androgen and oestrogen receptor[368,369]. Understanding the origin of these lesions will be crucial to address hormone-driven carcinogenesis in people with DNA-repair pathways mutations (e.g. BRCA in breast and ovarian cancer)[370,371]. Of note, TOP2-DPCs are not uniquely formed by chemotherapeutics, the antibiotic ciprofloxacin targets DNA gyrase, the bacterial type II topoisomerase, in a mechanism which involves the stabilization of the enzyme-DNA covalent intermediate[372]. In addition, the same strategy is used for gateway cloning, in which the negative selection marker is *ccdb*, which codes for a bacterial toxin that also generates gyrase-DPCs[373].
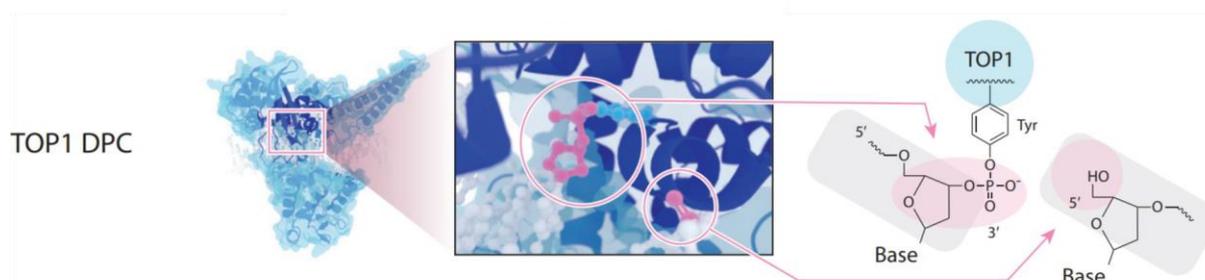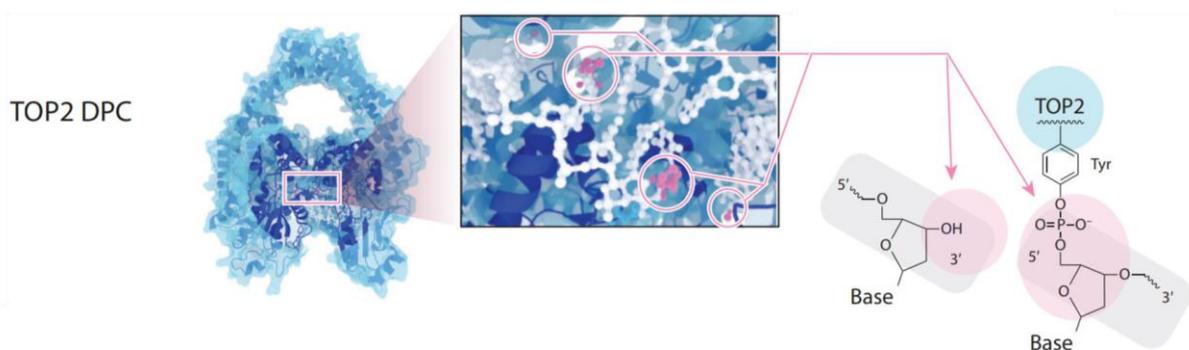

Figure 4. Schematic depiction of the chemical structure of a TOP2cc. Grey colour represents DNA backbone, light blue the protein crosslinked and pink the chemical bond between the protein and DNA. Adapted from[63]

2.2.3 Polymerase Beta

AP sites are unstable and their oxidation can generate 2-deoxyribonolactone (dL)[374]. Common dL inducing agents include hydrogen peroxide, UV light[375], ionizing radiation[376] or anticancer drugs such as neocarzinostatin[377] or tirapazamine[378]. This lesion becomes a suicidal substrate for Polβ during short-patch BER what results in the generation of a DPC[379]. Crosslinking of Polβ to dL relies on its lyase activity given that a mutation in the catalytic lysine completely abolishes DPC formation. Interestingly, processing of the dL lesion by APE1 is a requirement for Polβ-DPC formation[380].

2.2.4 DNMT1

DNA methyltransferases (DNMTs) mediate DNA methylation, a modification essential for the regulation of chromatin structure in mammals[381]. DNA methylation mostly takes place on palindromic CpG sites (regions with CG repeats), where cytosines are modified on their

C5[382,383]. While DNMT3A and DNMT3B are responsible for *de novo* methylation, DNMT1 displays high affinity for hemimethylated dsDNA, being responsible for the maintenance of methylation patters after DNA replication[384,385]. The interplay between DNMTs and the ten-eleven translocation (TET) methyl cytosine dioxygenases, regulates chromatin methylation and confers dynamism to the system[382]. DNMT1 has a highly conserved catalytic domain which transfers a methyl group from S-adenosylmethionin to its target cytosine[384,386]. During the methylation reaction, DNMT1 establishes a covalent enzyme-DNA complex through its catalytic cysteine with the C6 position of the aromatic ring, which is resolved by transfer of the methyl group to the C5, triggering release of the enzyme and yielding the modified C5 base[384,386]. During the methylation process, the target cytosine is flipped out, what presents the active site residues for the reaction start[387]. In addition, DNMT1 activity is tightly regulated by ubiquitin like with PHD and ring finger domains 1 (UHRF1), an E3-ligase which mediates histone H3 ubiquitylation, stimulating DNMT1 methylation[388,389]. A second crucial feature of DNMT1 is its PCNA-binding domains, that allows recruitment of DNMT1 to replication forks for its replication-coupled methylation activity[390,391]. The stabilization of DNMT1 covalent intermediate with DNA has been exploited for decades in the use of myelodysplastic syndromes[392,393]. An analogue of deoxycytidine, 5-aza-2-deoxycytidine (5-azadC), also referred to as decitabine, is incorporated in DNA during S-phase. When DNMT1 tries to methylate 5-azadC, the β-elimination in the final step cannot occur due to the presence of the nitrogen at position 5. This leads to an irreversible trapping of DNMT1 as a DPC[392]. Treatment with 5-azadC presents to major consequences: first, DNMT1-DPCs are lesions that the cell must repair[394]. Second, the high-efficiency trapping of DNMT1 leads to a fast depletion of the enzyme, triggering a dramatic reduction in global methylation and allowing the re-expression of tumour suppressor genes, normally silenced in cancer cells[395].

Figure 5. Schematic depiction of the chemical structure of a DNMT1-DPC formed by a suicidal reaction between the catalytic cysteine of DNMT1 and the pseudosubstrate 5-azadC. Grey colour represents DNA backbone, light blue the protein crosslinked and pink the chemical bond between the protein and DNA. Adapted from[63]

2.2.5 PARP1

PARP inhibitors (PARPis) have revolutionized cancer treatment because they have greater specificity than previously used chemotherapeutics and radiotherapy[396]. The involvement of PARP1 in different repair pathways (SSB-, DSB- repair or DDT), and its role as back up in all

of them, makes it a perfect target for inhibition in cancers where one of the factors in alternative pathways is mutated or dysfunctional[172]. One of the most well documented examples is cancer with *BRCA1* or *BRCA2* mutations, where inhibition of PARP1 specifically drives cancer cell death due to their synthetic lethal interaction[397]. All approved PARPis are pseudosubstrates for PARP proteins and bind competitively to the catalytic site of the enzyme, avoiding NAD$^+$ binding, and trapping the protein on DNA[398]. In addition to inhibiting the PARPs activity, trapping of PARP1 has been proposed as one of the main mechanisms contributing to PARPis toxicity[398]. The binding affinity of PARP1 for DNA is so exceptionally high that trapped PARP1 behaves as a "pseudo" DPC, which is extracted by the segregase valosin containing protein (VCP or p97) in a process coordinated by the SUMO/Ub system (described in section 2.4)[399]. In addition, PARP1 can also crosslink to AP sites, but also to some of the AP-derived lesions (e.g. 3'-PUA), during BER. PARP1-DPC formation can be enhanced after PARPi treatment, due to the increase in residence time of PARP1 on these reactive lesions[400,401].

2.3 Physiological DNA protein crosslinks

Physiological DPCs are those generated by enzymes whose crosslinking is essential for their function. Some of these, play a key role during processes such as AP sites quenching or meiotic recombination. The following section summarizes the most important, or best characterized, examples of physiological DPCs.

2.3.1 HMCES

5-hydroxymethylcytosine (5hmC) binding, ES cell-specific (HMCES) was recently described as an AP site shield that protects ssDNA during replication[402]. HMCES, and its homolog in *E. coli* YedK, possess an N-terminal cysteine -the methionine is removed co-translationally- which reacts with the open conformation of AP sites generating a stable thiazolidine DPC[403,404]. Loss of HMCES in human cells, sensitizes to DNA damage agents which generate AP sites such as $KBrO_3$ or the alkylating agent MMS. HMCES contains an evolutionary conserved SOS response associated peptidase (SRAP) domain which has its catalytic activity and a C-terminal disordered region, which harbours a PIP box[402]. Resolution of HMCES-DPCs, *in vitro* and cells, seem to involve different strategies. On the one hand, HMCES-DPCs can get ubiquitylated and degraded by the proteasome[402], what could allow TLS bypass of the remnant peptide. A second proteolysis-mediated strategy involves the helicase FANCJ, which possesses unfoldase activity, allowing DNA-protein interface exposure and cleavage by SPRTN[405]. On the other hand, HMCES-DPCs can revert on dsDNA, meaning that FANCJ-mediated TLS bypass of the intact adduct, or alternative processes such as TS, might serve as platforms for HMCES-DPC reversal. Reversal, which is evolutionary conserved, relies on a single glutamate residue (Glu107 in YedK and Glu127 in HMCES)[406,407].
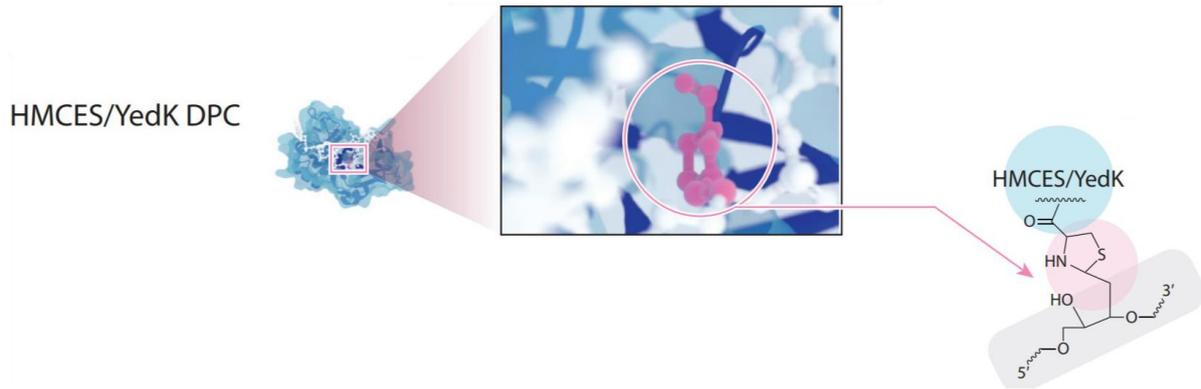
Figure 6. Schematic depiction of the chemical structure of a HMCES or YedK DPC, between its catalytic cysteine which forms a thiazolidine intermediate with the open conformation of an AP site. Grey colour represents DNA backbone, light blue the protein crosslinked and pink the chemical bond between the protein and DNA. Adapted from[63]

2.3.2 UdgX

The Uracil DNA glycosilase X (UdgX) from *Mycobacterium smegmatis* was initially described as a protein capable to tightly bind uracil in a complex resistant to detergent, reducing conditions and heat[408]. It was later observed that this tight binding was due to the formation of a covalent complex mediated by the suicide inactivation of the protein. The process starts with the attack of a histidine located in the protruding loop at the entrance of the uracil-binding site, which forms a stable intermediate with the C1' position of the deoxyribose at the AP site of the ssDNA containing uracils[409-411]. UdgX might play an important role during replication, where uracils generated due to cytosine deamination, can be turned into AP sites by uracil glycosilases. The presence of AP sites on ssDNA by uracil excision, can be abolished by UdgX, which generates a covalent linkage that could avoid strand breakage during replication[408]. Additionally, UdgX has also gained interest as a molecular tool for the detection of uracils in genomic DNA[412-414], but also for the direct measure of cytosine deaminases' activity in live cells[415].

2.3.3 SPO11

The deliberate generation of DSB during meiosis is essential for genetic recombination. The introduction of these lesions is tightly regulated and mediated by SPO11[416]. Evolved from a type IIB topoisomerase, SPO11 generates 5' overhangs[417] by cleaving opposite strands of the DNA backbone with its two-active site tyrosines[418]. Unlike other topoisomerases, SPO11 cannot perform strand passage and resealing of the ends after catalysing break formation, what generates the accumulation of SPO11-DPCs during meiotic recombination[419]. SPO11-DPC removal is mediated by the MRN[420] complex (MRX in yeast[421-423]), which initiates processing of the DSB with its single-strand endonuclease and 3'-5' exonuclease activity. Release of SPO11-DPCs and resection of the DNA ends are key to promote synapsis, recombination and meiotic progression[424].

2.3.4 EBNA1

The Epstein-Bar virus nuclear antigen 1 (EBNA1) is the only protein required for its DNA replication and episomal maintenance during virus latent infection[425,426]. It binds to repetitive DNA elements, in the viral origin of replication, and this binding is essential for its molecular function[425]. EBNA1 generates crosslinks with a similar chemical identity than those formed by topoisomerases. EBNA1 is a tyrosine recombinase and forms the DNA adduct through a single-strand cleavage which involves the formation of a phosphodiester intermediate with the DNA backbone[425,427,428]. EBNA1-DPCs are formed in a cell cycle dependent fashion and are enriched in S/G2 phase. Mutations in the active tyrosine undergo the loss of recombination intermediates which are important for the viral genome replication[429].

2.3.5. Viral terminal proteins

Terminal proteins (TP) are found in different groups of DNA viruses and form a DPC through a phosphodiester bond involving the hydroxyl radical of its active serine and the 5'-end nucleotide of the viral genome[430,431]. One of the most studied TP is the 55-kDa (TP-55) of adenovirus. This protein, which is translated as a precursor (precursor terminal protein, pTP), is cleaved after viral packaging by a viral protease to generate its mature form[432,433]. pTP primes viral DNA replication after crosslinking to a single nucleotide, in a process that involves a "jumping back" mechanism[434]. Remarkably, this mechanism is not only present in eukaryotic viruses, but also in some bacteriophages as the *Bacillus subtilis* bacteriophage Φ29 and *E. coli* PRD1[435-437].

2.4 DNA protein crosslink repair

DPCs can be targeted by multiple repair pathways which will rely on their composition and structure. In addition, active chromatin transactions trigger DPC-repair, being replication-coupled DPC-repair the best understood process. In the following section I summarize the current knowledge on DPC-repair with a special focus on the proteolytic degradation of the protein adduct.

2.4.1 Direct hydrolysis: TDP1 and TDP2

TOP1- and TOP2-DPCs are so frequent that cells have evolved a specific set of enzymes which can directly repair these lesions. Tyrosyl-DNA phosphodiesterase 1 and 2 (TDP1 and TDP2) are key enzymes mediating TOP1 and TOP2 crosslink reversal, respectively. TDP1 processes 3'- DNA-ends not only containing TOP1ccs but also certain end-blocks such as 3' phosphoglycolate or AP sites[438]. During the repair of TOP1ccs, TDP1 hydrolyses the phosphodiester bond between TOP1 and the DNA, releasing the enzyme and leaves a phosphate in the 3'-end, which is a further target of downstream factors such as PNKP, DNA polymerases and ligases[235]. The bulkiness of TOP1ccs blocks access of TDP1 to the interface,

thereby needing proteolytic pre-processing of the lesion for its resolution[439]. The degradation of the protein adduct can be done by the proteasome in a process that involves SUMOylation by the E3-ligase PIAS4 and ubiquitylation by RNF4[97,440].
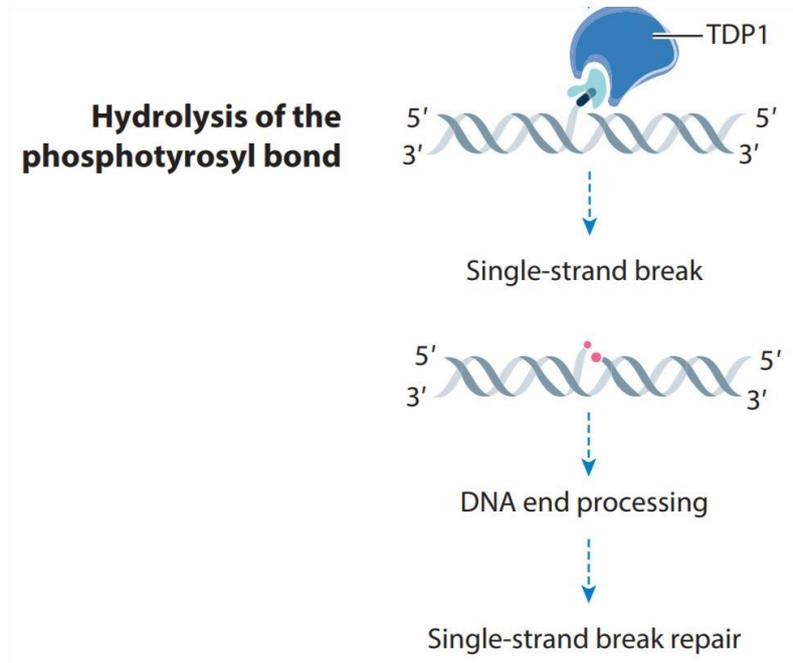
Figure 7. Schematic depiction of the activity of TDP1 on a TOP1cc peptide after partial degradation by specialized proteases including downstream processing of the lesion. Light blue colour represents the remnant peptide after TOP1cc partial degradation and the black colour represents the presence of a covalent bond. Adapted from[63]
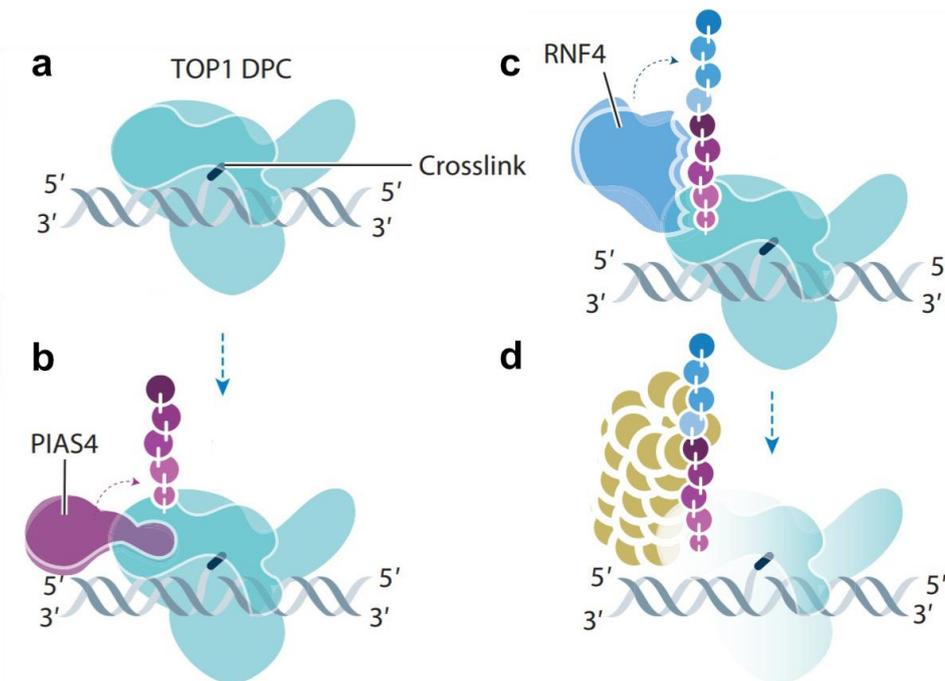
Figure 8. Schematic depiction of the processing of a TOP1cc/DPC by the proteasome in a SUMO targeted ubiquitylation fashion. a. Represents formation of the TOP1-DPC. b. Represents the detection by the SUMO E3-

Structural studies of TDP1 revealed that it is able to separate DNA strands within the lesion site, suggesting that this local melting is necessary for 3'-end processing[438]. TDP1 was first discovered in yeast[441] and its loss hypersensitizes to CPT. *Tdp1* knock-out yeast also display synthetic lethality with genes involved in DSB-repair[442] and with the metalloprotease Wss1[323]. In humans, germline mutations in *TDP1* are causative for spinocerebellar ataxia with axonal neuropathy (SCAN1)[443,444]. Interestingly, this syndrome is not caused by a total loss of function in TDP1, but rather a gain of function because a substitution (H493R), disrupts the active site and stabilizes the normally transient intermediate between TDP1 and the DNA, generating a TDP1-DPC[443]. Cells carrying homozygous alleles *TDP1*[H493R] are hypersensitive to TOP1-trapping agents. Despite being a gain-of-function mutation, the phenotype just manifests in a recessive trait, given that wild-type TDP1 can still repair TOP1ccs as well as TDP1[H493R]-crosslinks [445]. In human cells, the recruitment of TDP1 to the lesion can be enhanced by a combination of PTMs. TDP1 is phosphorylated by DDR kinases at S81, what seems to increase its stability and association with XRCC1 in response to CPT-induced DSBs[446]. Moreover, PARylation has recently gained interest as one of the predominant PTMs that controls repair of TOP1-DPCs by modulating TDP1 recruitment. A direct protein-protein interaction between PARP1 C-terminal domain and TDP1 seems to be enhanced after PARylation, what also aids in the removal of TOP1-DPCs by direct PARylation of TOP1, facilitating its religation in the presence of CPT[447,448]. In addition, *PARP1* knock-out cells display lower TDP1 activity, while a combination treatment of PARPis and CPT kills cells in a synergistic manner[447].

In contrast to TDP1, TDP2 catalyses the release of TOP2 adducts from the 5'-DNA-end[449], but proteolytic pre-processing of TOP2ccs[450] is not essential for their resolution. In addition to PIAS4, RNF4 and proteasomal repair, TOP2ccs can undergo SUMOylation by a different SUMO E3-ligase, ZATT. SUMOylation by ZATT on TOP2ccs seems to reshape the crosslink, facilitating accessibility to the protein-DNA interface by TDP2 active site[451]. Additionally, recruitment of TDP2 to TOP2-DPCs is enhanced by SUMO2/3 binding[451] and the presence of ubiquitin chains, which can interact with the TDP2 UBA domain in its N-terminus[452]. In agreement with its molecular function, *TDP2* knock-out cells display hypersensitivity to TOP2 poisons such as etoposide[453]. Mutations in *TDP2* are causative for spinocerebellar ataxia, autosomal recessive 23 (SCAR23) due to an enzyme loss of function by the presence of a premature stop codon in the coding sequence[454].
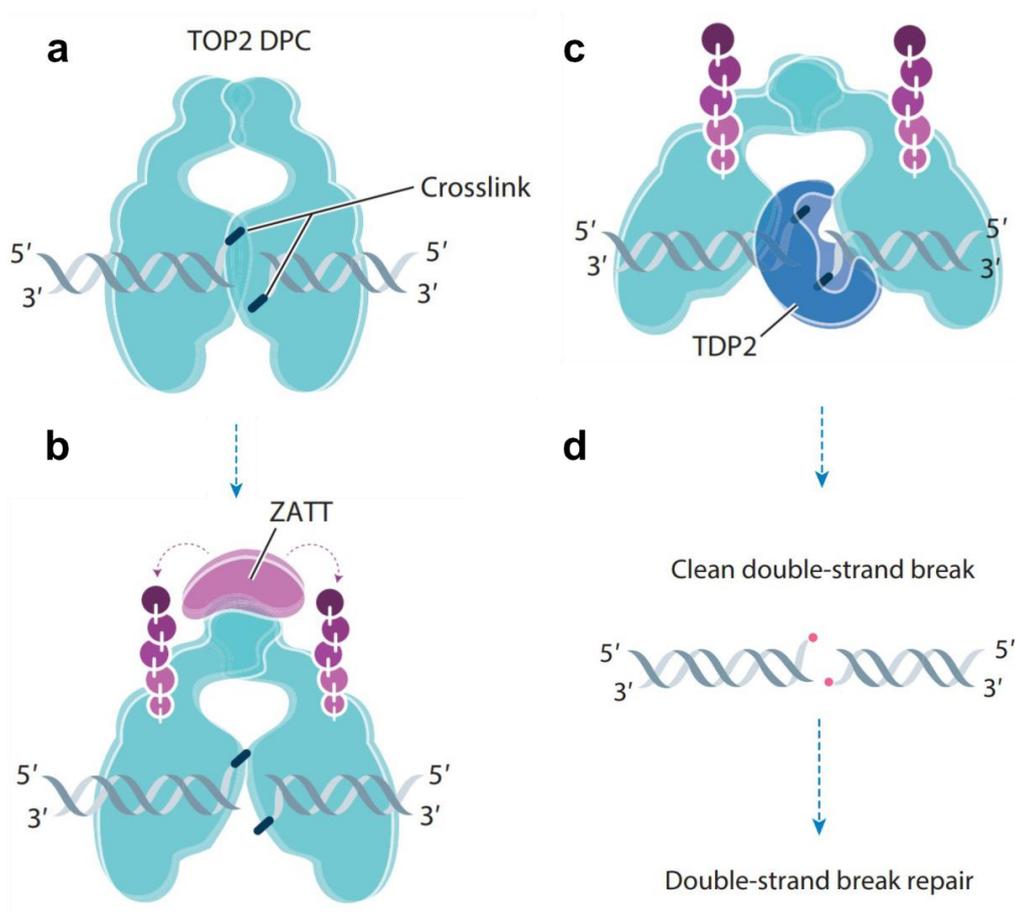
Figure 9. Schematic depiction of the processing of a TOP2cc/DPC mediated by ZATT SUMOylation, crosslink reshaping and TDP2 activity, which leaves a DSB which is processed downstream. a. Represents TOP2-DPC formation. b. Recognition of the TOP2-DPC by ZATT, what triggers SUMOylation of the adduct and reshapes its structure. c. The change in structure after SUMOylation grants accessibility of TDP2 to the DNA-protein interface. d. The DSB generated is processed downstream by canonical DSB-repair pathways. Adapted from[63]

Of note, TDP2 mediates the removal of the viral protein genome-linked (VPg), which covalently caps the 5' end of some positive RNA viruses (e.g. picornaviruses), a process needed for the viral cycle, making TDP2 an interesting antiviral target[455].

2.4.2 Nucleolytic repair, the MRN complex

Two main nucleolytic pathways, MRN mediated resection and NER, can tackle the accumulation of DPC, but they are restricted to very specific contexts. The MRN complex requires the adduct to be located at the end of a DSB. The role of the MRN complex is prominent in the resolution of physiological SPO11-DPCs, but also the pathological or drug-induced TOP2-DPCs[456]. The sensing and processing of the DSB-end DNA adduct has been characterized with fine detail in the *E. coli* homolog of the MRN, the MR complex (*E. coli* lacks NBS1). The Mre11 and Rad50 assemble and form heterotetramers and get activated by a conformational change induced by ATP[457]. The presence of a block at the end of the DSB does not impede MR endonucleolytic activity, allowing Mre11 to cut DNA 15 to 25 bp away from the block[458]. MR binding to DNA triggers the formation of a narrow clamp around it by

Rad50 coiled coils, what permits the binding of Mre11 to the DNA end. The coiled coils will slide over the DNA strand and map the DNA adduct, what can be not just a DPC but virtually any DNA structure[459]. The detection of the adduct induces a conformational transition to a "rod", the nucleolytic-active state[459]. Given that the connection of the DNA ends by internal blocks cannot be processed by MR(N), it is expected that TOP2-DPCs would require partial proteolytic degradation, probably aided by the ZATT SUMO E3-ligase and SUMO-targeted ubiquitylation[97,459].
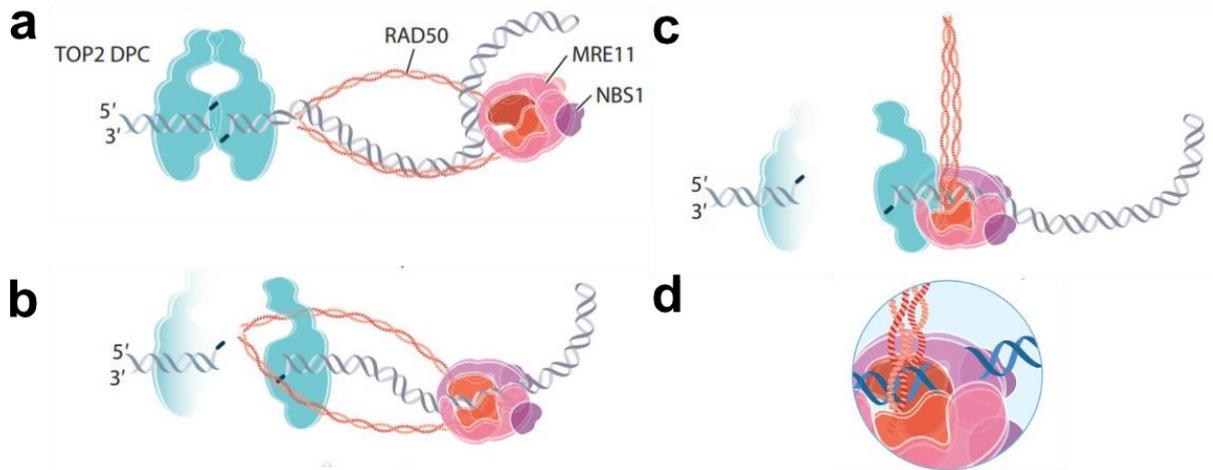
Figure 10. Schematic depiction of the processing of a TOP2cc/DPC mediated by nucleolytic processing by the MRN complex. a. Mapping of the DNA helix by the MRN complex. b. Detection of the DSB blocking adduct, in this context a TOP2cc/DPC. c and d. Transition of the RAD50 filaments to a "rod", what allows endonucleolytic cleavage adjacent to the adduct. Adapted from[63]

In addition to the processing of DSB-Protein adducts, the NER endonucleases can participate in the removal of small DPCs, under 12-14 kDa, and DNA-peptide crosslinks in *E. coli*. Interestingly, NER seems not to be important for the tolerance towards bulkier lesions, while HR proves essential in this matter, because the loss of recA extremely sensitizes the cells to formaldehyde and azacytidine[460]. Experiments *in vitro* further support the importance of NER endonucleases in DPCs and DNA-peptide crosslinks removal, being the XPG nuclease relevant in this process[461].

2.4.3 Specialised DNA-protein crosslink proteases

As previously stated, a protease-dependent repair pathway for DPCs induced by formaldehyde was already proposed almost five decades ago[1]. It took almost 40 years to find the first dedicated protease for the repair of DPCs in yeast, Wss1[462].

2.4.3.1 Wss1 and SPRTN

Wss1 (weak suppressor of smt3) is a zinc metalloprotease which was firstly identified in a yeast screen as suppressor for chromosomal aberrations observed upon Smt3p loss[463]. Wss1 proteases are widely distributed in different life kingdoms such as plants, fungi or bacteria[464].

Wss1 is directly implicated in the repair of Top1-DPCs because yeast lacking Wss1 and Tdp1 have a severely compromised cell fitness. In support of these data, the deletion of Top1, which is not essential in *S. cerevisiae,* restores growth to wild-type levels. In agreement, Δ*wss1*Δ*tdp1* yeast are also extremely sensitive to CPT and Wss1 can directly process these lesions by cleaving the Top1-DNA crosslink and itself in a DNA dependent manner. The loss of Wss1 also sensitized yeast to formaldehyde, highlighting that Wss1 is not only implicated in the repair of Top1-DPCs but also those formed by unspecific crosslinkers[462]. Wss1 tightly associates with the segregase Cdc48 (p97 in humans) a feature which is evolutionary conserved in its metazoan homolog SPRTN. It is remarkable however, that this targeting, which is mainly mediated by the Cdc48 cofactor Ubx5, sensitizes the cells to DPC-induction agents in the absence of Wss1, suggesting that commitment to Wss1-dependent proteolytic pathway is a one-way road[465]. An in-depth analysis of the phylogenesis of DPC-proteases revealed that despite belonging to evolutionary distant families, they share highly related domains and protein interaction motifs[464]. Within their differences, the lack of Ub-binding domains in Wss1 present in SPRTN and the presence of SIMs in Wss1 not present in SPRTN is probably the most notable one[464].

SPRTN (SprT domain at the N-terminus) was firstly described as a TLS regulator. It was observed in cells that loss of SPRTN increased the retention time of Polη on chromatin after DNA damage induction. In addition, this process is regulated by multiple factors. Firstly, SPRTN requires PCNA interaction, mediated by its PIP-box, and with the segregase p97, in a process that was thought to mediate Polη extraction from chromatin. Secondly, all this process was coordinated in a ubiquitin-dependent manner, and required the ubiquitin-binding zinc finger (UBZ) domain at SPRTN's C-terminus, which unbiasedly binds different Ub-chains[466]. Strikingly, a second study in human cells showed exactly the opposite, proposing that SPRTN promotes Polη recruitment and aids during TLS initiation by physically interacting with RAD18[467]. After demonstrating that SPRTN performs the same function as Wss1 in metazoans, many studies focused on its structure and activation mechanism[323,468]. SPRTN contains a zinc metalloprotease (SprT) domain on its N-terminus followed by two DNA interacting motifs, the zinc-binding domain (ZBD) and the basic region (BR), which interact with ss- and ds-DNA structures respectively, and a C-terminal tail, containing the p97 binding SHP motif, the PCNA interacting PIP-box and the UBZ[63]. The DNA-dependent activation of SPRTN has been extensively described using *in vitro* DPC-models. SPRTN displays a DNA-structure dependent activation which presumably limits potential off-target effects. SPRTN utilizes both DNA interfaces, the BR and the ZBD, to bind to different DNA structural features such as bubbles, junctions, gaps or DSBs[469]. The simultaneous engagement of both motifs in

DNA binding might allow conformational changes and protease activity by opening of the active site and efficient substrate processing[469,470].
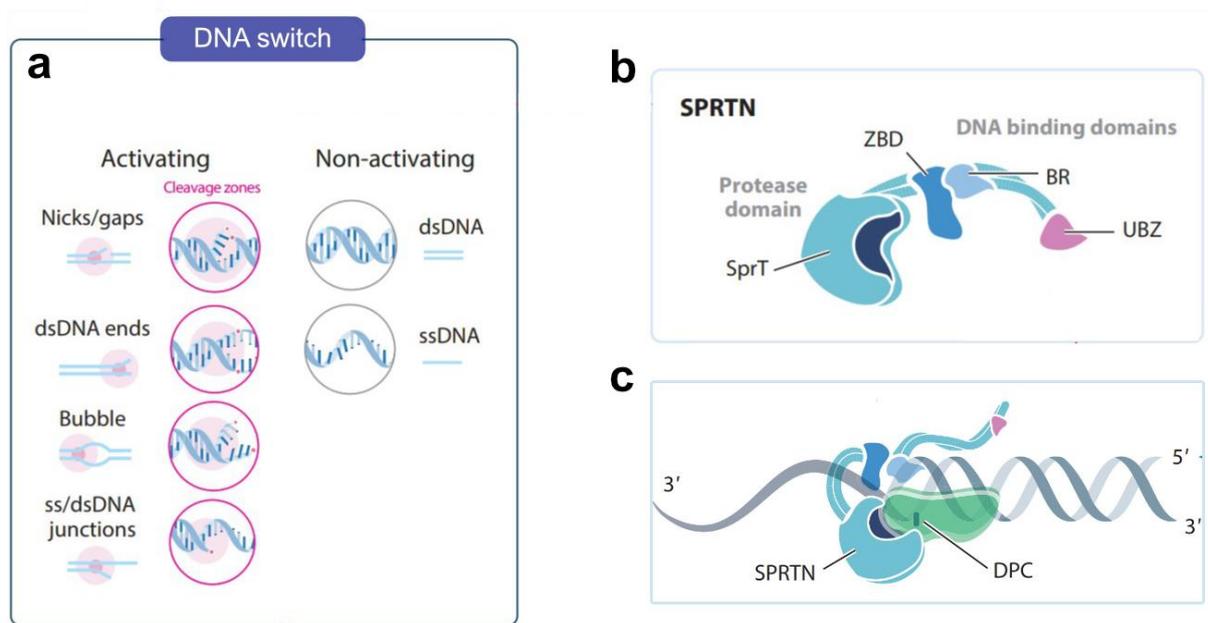
Figure 11. Schematic depiction of DNA-substrate dependent SPRTN's activity during DPC-cleavage. a. DNA structures which can activate SPRTN's proteolysis. b. SPRTN domain distribution, SprT protease domain, Zinc Binding Domain (ZBD), Basic Region (BR) and Ubiquitin-Binding Zinc finger (UBZ). c. SPRTN structural changes during DPC-cleavage upon DNA binding. Adapted from[63].

A second layer on SPRTN's activity control is mediated by ubiquitylation. SPRTN coexists in cells in two states, unmodified or monoubiquitylated[323,467,468]. The modification by monoUb is shielded by the enzyme's own UBZ given that mutations in this domain completely shifts the equilibrium towards unmodified[467]. Modification of the protein has two main functions. First, it primes the protein for polyubiquitylation and proteasomal degradation, what allows the maintenance of constant levels of SPRTN and is probably implicated in the protein cell cycle dependent regulation[471]. Second, ubiquitylated-SPRTN also increases its autocatalytic activity *in trans*, allowing effective inactivation of the monoubiquitylated protein upon DNA binding[471]. USP7 is the most relevant DUB implicated in SPRTN deubiquitylation and confers tolerance towards DPC-inducing agents[471], but the DUBs VCPIP and USP11 have also been implicated in this function[472,473].
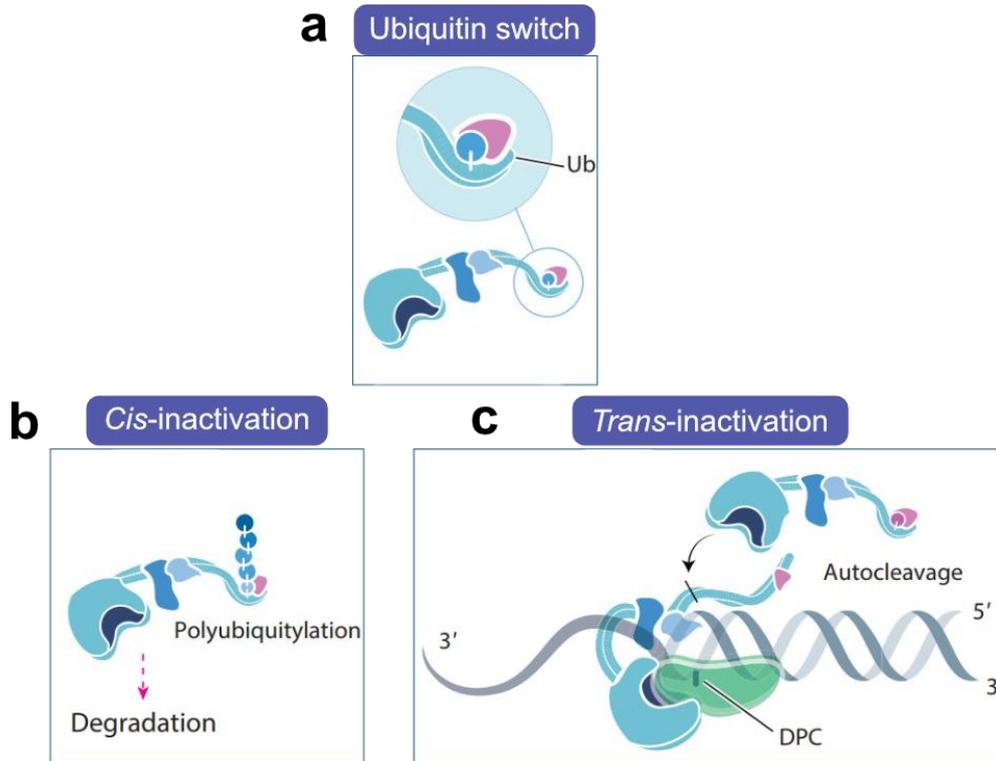
Figure 12. Schematic depiction of SPRTN's regulation by ubiquitylation. a. SPRTN is constitutively monoubiquitylated and coexists in an approximate 50% equilibrium between modified and unmodified. Monoubiquitylation is shielded by the UBZ domain. b. The presence of monoubiquitin on SPRTN, can trigger polyubiquitylation and proteasomal degradation, what is known as cis-inactivation. c. Trans-inactivation occurs when a monoubiquitylated SPRTN cleaves a second SPRTN on chromatin, probably due to a more opened conformation of the protease domain which enhances its activity. Adapted from[63].

SPRTN loss has deleterious consequences which scale up with evolution. *C. elegans* tolerate the loss of DVC-1 (SPRTN in worms) but they become hypersensitive to DPC-inducing agents[323]. In *D. melanogaster* SPRTN is named maternal haploid (MH) because the loss of the protein in the female fly does not allow progeny generation even if the male is *mh*-sufficient. After fertilization, numerous chromosomal missegregation events lead to the loss of paternal DNA, what results in inviable eggs which just contain maternal DNA. This could be explained in a scenario in which paternal DNA bears large amounts of DPCs that must be repaired before the first zygotic division[474]. In mammals, SPRTN is essential and its loss is embryonic lethal. In addition, *Sprtn* conditional knock-out mouse embryonic fibroblasts display cell cycle arrest after few divisions and accumulate DNA damage markers. In humans, mutations in SPRTN are causative for Ruijs-Aalfs syndrome (RJALS), which is characterized by progeroid features, genomic instability and early onset hepatocellular carcinoma[475]. There are currently three characterized RJALS patients, two of them have deletions at the end of SPRTN's exon 4 or beginning of exon 5 which lead to a C-terminal truncated protein referred as SPRTN-ΔC. SPRTN-ΔC lacks all the features on SPRTN's C-terminus, the SHP, the PIP and the UBZ, and is additionally mislocalized due to the loss of a C-terminal NLS. The third patient is heterozygous and bears, in addition to the *SPRTN-ΔC* allele, a mutation in close proximity to

the enzyme catalytic site (SPRTN^{Y117C}) which presumably results in compromised enzymatic function[323,468,476,477].
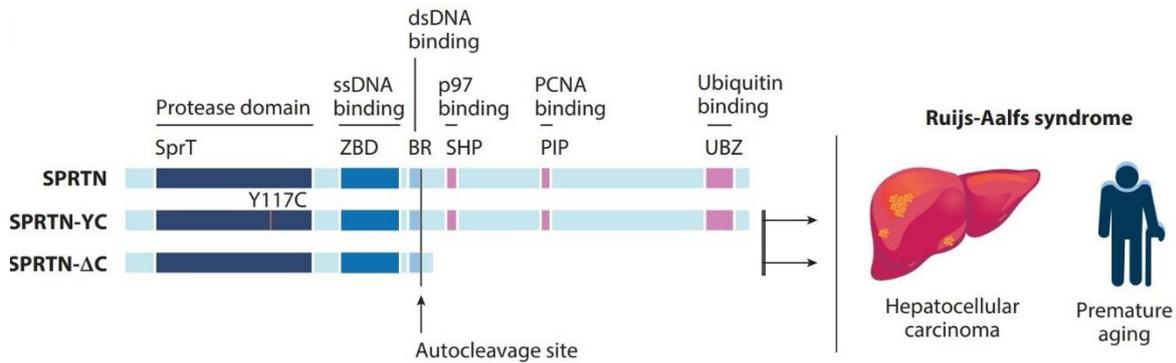
Figure 13. Schematic depiction of the domain distribution on SPRTN and Ruijs-Aalfs syndrome patient variants. SPRTN-YC contains a mutation, Y117C, in close proximity to the catalytic glutamate (E112), what probably disturbs protease activity. SPRTN-ΔC is cause by a premature stop codon in SPRTN's exon 5, what causes the mistranslation of the protein leading to a truncated form which lacks SHP, PIP box and UBZ. Ruijs-Aalfs syndrome is characterized by early onset hepatocellular carcinoma and premature aging. Adapted from[63].

2.4.3.2 The proteasome

The 26S proteasome is the most relevant protease in cells and performs protein degradation not just in the cytosol but also in the nucleus. The proteasome is a threonine protease which belongs to the ATPases associated with various cellular activities (AAA+) family. The structure is similar to a "barrel" and it uses ATP-driven force to pull polypeptides through its axial cavities into its internal degradation chamber, what excludes tightly folded substrates and large polypeptides. The process is coordinated by the use of Ub-chains as a degradation mark, but recent studies support the existence of Ub-independent proteasomal degradation, mediated by specific degron sequences or shuttling factors such as Ubiquilins[74,478-480]. The specific role of the proteasome in DPC-repair was already proposed before but recently confirmed in *Xenopus* extract systems, requiring the action of specific E3-ligases (further discussed in 2.4.4)[380,481].

2.4.3.3 Additional DPC proteases

Additional proteases have been recently implicated in DPC-repair. The DNA damage inducible 1 (Ddi1) is a yeast aspartyl protease whose loss displays synthetic lethality with the loss of Tdp1 and Wss1. Ddi1 acts on Top1-, Flp recombinase- DPCs and stalled RNPII promoting their degradation in a manner dependent on its protease activity. In addition, Ddi1 mitigates the triple loss of Ubx5, Wss1 and Tdp1, further supporting its redundant role in DPC-repair in *S. cerevisiae*[465,482]. Ddi1 has two homologs in humans, DDI1 and DDI2, which have been implicated in the degradation of the replication termination factor 2 (RTF2), avoiding fork restart defects[483]. The implication of these proteases in DPC-repair is less clear and *in vitro* reconstitution experiments will be needed to assess a direct role in the process. Moreover, the major role of DDI2 seems to be the cleavage of the transcription factor NRF1, important for

the transcription of genes related to proteasomal regulation, including its own subunits[484]. Furthermore, the targeting of these proteases requires polyubiquitylation, what might imply that they aid as proteasome shuttle factors during DPC-repair[485].

The germ-cell nuclear antigen or acidic repair containing protein (GCNA/ACRC) is a second protein recently implicated in DPC-repair. GCNA has a SprT protease domain and an acidic intrinsically disordered region (IDR). The loss of *Gcna* in flies leads to genome instability leading to embryonic lethality defects, while in worms these effects are observed at later stages in combination with a mortal germline phenotype, where lifespan and fertility are reduced over generations[486]. Additionally, mutant worms display hypersensitivity to hydroxyurea and etoposide in a synergistic manner with DVC1, suggesting partial redundancy between both pathways[487]. *Gcna* knock-out in both cases displayed elevated levels of TOP2-DPCs, which might be accumulating due to defective repair by GCNA[486]. Furthermore, the presence of four tandem SIMs on its N-terminus targets GCNA to polySUMOylated DPCs in cells[99]. Notably, the role of GCNA in DPC-repair seems to be restricted to fertilization and embryonic development and, in addition, rodents lost the Sprt-protease domain but still recapitulate the phenotypes of *Gcna* loss in other species, what challenges the role of GCNA as a DPC-repair protease[488].

Lastly, the serine proteases FAM111 trypsin-like peptidase A and B (FAM111A and FAM111B) have recently being implicated in DNA-repair as DPC-repair proteases. Autocleavage of FAM111 proteases has been observed *in vitro* and in cells when the proteins are overexpressed[489,490]. From these additional proteases, FAM111A is probably the best characterized in DPC-repair [491]. FAM111A contains a PIP-box in its N-terminus which is followed by two ubiquitin-like domains (UBLs), the second contains its ssDNA binding activity, and a C-terminal protease domain. FAM111B resembles exactly the same domain composition but lacks the PIP-box, suggesting that it arose in evolution after gene duplication[491]. FAM111A localizes at replication forks and confers tolerance towards TOP1-poisons and PARPi and its loss slows replication fork progression. Therefore, it is plausible that this protease cleaves general proteinic adducts encountered during replication, being TOP1-DPCs probably the most relevant ones[490]. Mutations in the *FAM111A* gene are causative for Kenny-Caffey Syndrome type 2. Intriguingly, most of the patients carry gain-of-function mutations which lead to protease hyperactivation and probably cell toxicity[492,493]. Mutations on *FAM111B* cause hereditary fibrosing poikiloderma with tendon contractures, myopathy, and pulmonary fibrosis. As for FAM111A, mutations increase proteolytic activity of the enzyme, but with different phenotypical features[489]. Additionally, high FAM111B expression is also a poor prognosis in breast cancer, due to its importance for cell proliferation and migration[494]. Remarkably, both proteases are restriction factors for different viruses,

acting FAM111A as a barrier for simian virus 40 replication and FAM111B highly upregulated after adenoviral infection in human cells, but whether there is any relation between viral defence and DPC-repair is not yet understood[495].

2.4.4 Replication-coupled DNA-protein crosslink repair

Studies in frog egg extracts have been key to understand the order of events when a replication fork encounters a DPC. The leading strand replicative helicase CMG will slow down its progression after clashing against the protein adduct, still, in a yet unknown mechanism, it is able to bypass the lesion in a process which might involve partial opening of a side channel of CMG[481]. The bypass of the adduct by CMG requires a second helicase, the regulator of telomere elongation 1 (RTEL1), normally involved in the resolution of DNA secondary structures that would otherwise block replication. RTEL1 unwinds the DNA downstream of the DPC on the lagging strand, allowing CMG bypass[496].
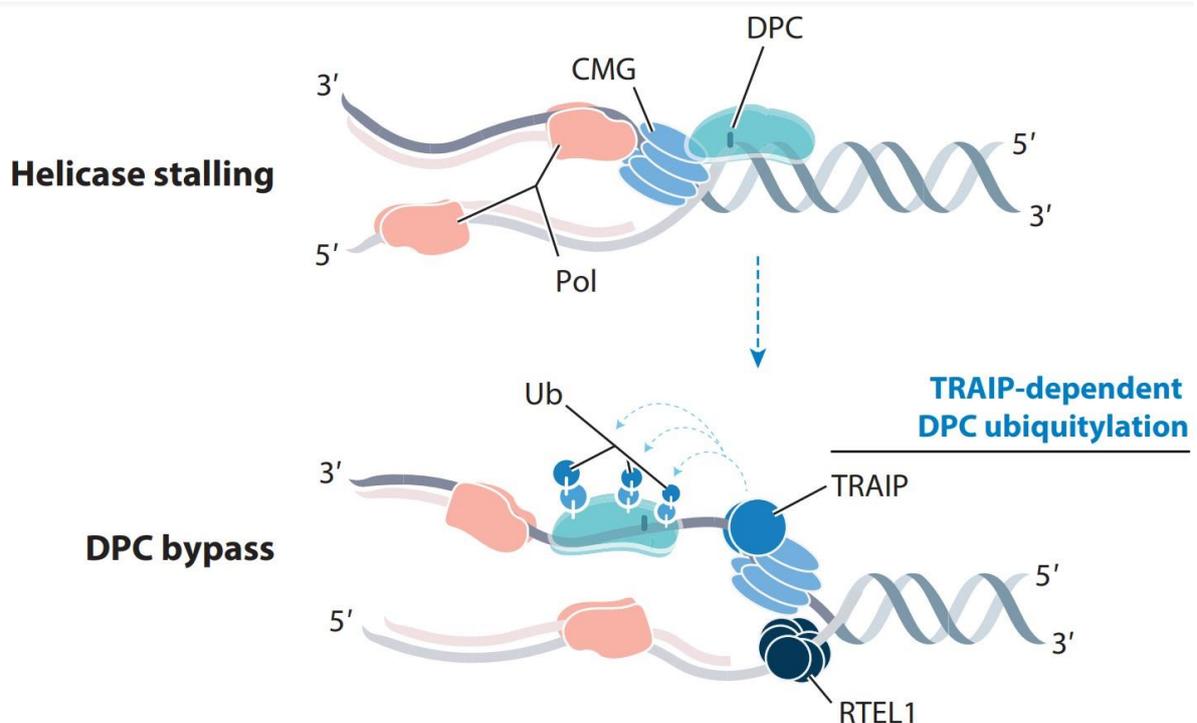
Figure 14. Schematic depiction of the bypass of a DPC during replication by the CMG. First, CMG stalls in front of the lesion, what recruits a second helicase RTEL1, which unwinds the lagging strand and permits bypass of the DPC by CMG. The process leads to ubiquitylation of the adduct by TRAIP. Adapted from[63].

Despite being bypassed by CMG, the DPC will block DNA polymerase progression, generating CMG-polymerase uncoupling and triggering two specific mechanisms: first, the TRAF interacting protein (TRAIP), a master Ub E3-ligase which travels with CMG, will ubiquitylate the DPC during bypass[481]. Second, the ring finger and WD repeat domain 3 (RFWD3), a second E3-ligase recently linked to FA and which coordinates HR and TLS, gets activated by the presence of ssDNA adjacent to the DPC and further extend the Ub-chains on the

protein[497,498]. The presence of Ub on the DPC will recruit the proteasome, which will degrade the protein up to a peptide[481]. Additionally, or optionally, SPRTN will be recruited to the specific ss-dsDNA junction generated by the stalled polymerase, cleaving the adduct closed to the DNA interface[323,469]. The recruitment of TLS polymerases allows synthesis through the remnant peptide, resuming replication but at the cost of potential nucleotide misincorporation on the lesion[499].
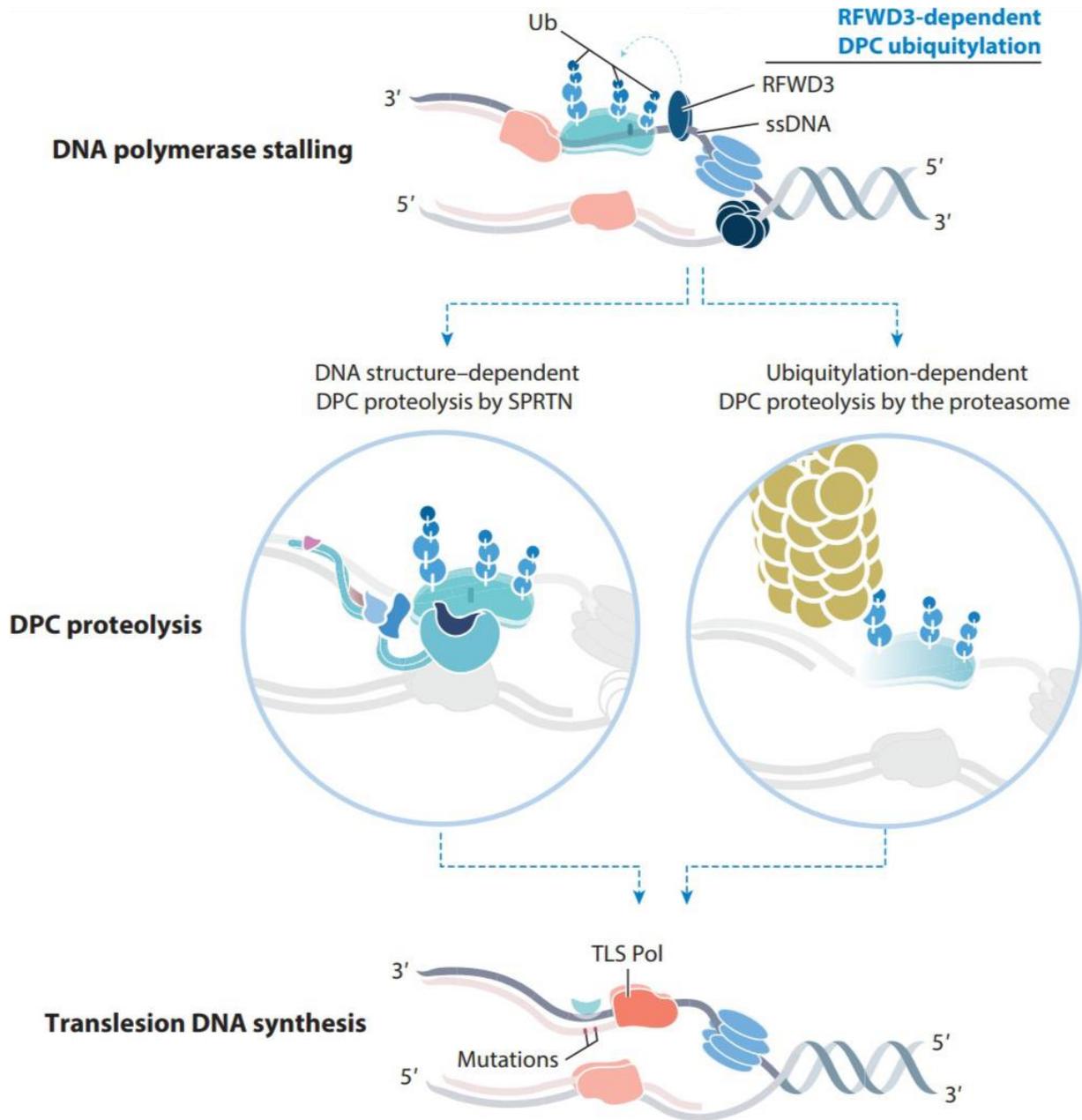
Figure 15. Schematic depiction of the downstream processing of a DPC after CMG bypass. First, DNA polymerase stalling will lead to uncoupling with the CMG helicase, generating ssDNA stretches in front of the DPC, what will recruit the ssDNA-dependent ubiquitin E3-ligase RFWD3, which will increase the ubiquitylation marks on the protein adduct. Second, the ss-dsDNA junction generated by polymerase stalling will recruit the DNA-dependent metalloprotease SPRTN, which will cleave the crosslinked protein in closed proximity to the DNA interface. Additionally, or alternatively, long ubiquitin chains will recruit the proteasome, which can also degrade the DPC. The remnant peptide can then be bypassed by specialized TLS polymerases, allowing replication to continue. Adapted from[63].

Interestingly, a new helicase has emerged as a DPC-resolution cofactor. The helicase FANCJ, is an intriguing protein with helicase and translocase activity that displays "unfoldase" activity *in vitro*, what exposes the crosslinked region of the protein and enhances SPRTN cleavage[405]. It would be tempting to speculate that the role of FANCJ could be extremely important for tightly folded DPCs, or those without exposed lysines, given that SPRTN does not require DPC-ubiquitylation for its activity[481]. In addition to FANCJ, a second factor has emerged as a potential new DPC-resolution protein. Recent data suggest that the testis-expressed protein 264 (TEX264) is also relevant for SPRTN activity. Briefly, TEX264, which is located in the nuclear periphery, promotes the resolution of TOP1-DPCs encountered during replication by associating with the segregase p97, which unfolds crosslinked TOP1 allowing SPRTN cleavage[500].

2.4.5 Global genome or replication independent DNA-protein crosslink repair

Replication-independent repair of DPCs has just been recently described. TOP1 and TOP2 poisons can induce DPCs regardless of the cell cycle status so it is conceivable that cells have evolved additional mechanisms, apart from those related to replication-coupled DPC-repair, allowing them to also repair these lesions. TDP1 and TDP2 require prior proteolytic processing of the adduct by the proteasome, what is achieved by SUMO-targeted ubiquitylation. Although SUMOylation by PIAS4 seems highly redundant, given that *PIAS4* knock-out cell lines are just mildly sensitive to DPC-inducing agents, the loss of RNF4 is more dramatic to cell survival.

2.4.5.1 RNF4 is the main StUbL in human cells and targets DNA-protein crosslinks

RNF4 is a relatively small protein of 190 amino acids with four well characterized SIMs within its N-terminal domain[501]. The SIMs mediate the binding to SUMO chains and allow the ubiquitylation of the substrate[502,503]. RNF4 is a homodimer, in yeast a heterodimer formed by Slx5/8[504], which controls the stability of SUMOylated proteins in cells, being the SUMO cascade components one of the main targets of its activity[505]. The role of RNF4 in DNA damage has been intensively studied. RNF4 is a key regulator of DSB-repair and mediates timely removal of different factors to allow the proper repair of the break. The function of RNF4 in DPC-removal has been characterized in studies using model DPC-substrates. RNF4 targets TOP1-, TOP2- and the 5-azadC-induced DNMT1-DPCs to proteasomal degradation[100]. Thus, *RNF4* knock-out cells display sensitivity towards DPC-inducing agents such as CPT and 5-azadC[97,100]. Interestingly, RNF4 loss increases resistance to etoposide, suggesting that proteolytic degradation of TOP2-DPCs is deleterious for the cell and that the ZNF451-TDP2 or the MRN mediated repair are preferable options[506]. Of note, RNF4 is also implicated in the degradation of promyelocytic leukemia protein (PML) upon arsenic treatment, a therapy used in acute promyelocytic leukemia, where PML is fused to the retinoic receptor alpha due to a chromosomal translocation[503,507-509]. It is then tempting to speculate that RNF4 degradation of

chromatin-bound proteins, or nuclear aggregates, is a general evolutionary repurposed mechanism that mediates eviction by p97 -or proteasomal degradation when eviction is not possible- and not just a DPC-specific process. This is supported by the same pathway acting on "trapped" PARP1, where there is not protein-DNA covalent linkage involved[399].

2.5 Techniques for the identification, isolation and study of DNA-protein crosslink

Numerous different methods have been established for the isolation and identification of DPCs.

2.5.1 RADAR and derivates

The rapid approach to DNA adducts recovery (RADAR) is based on a denaturing precipitation of DNA in the presence of high-concentration of detergent and chaotropic agents with the "RADAR" buffer. The precipitation of DNA is achieved by addition of pure ethanol, centrifugations and sequential ethanol washing steps to remove salt and contaminants. The precipitated DNA can be then resuspended in an alkaline buffer, generally sodium hydroxide, quantified and resolved by classical protein detection techniques[510]. A modification of the RADAR, the Superior method for True DNA-protein crosslinks Recovery (STAR), has recently been published as a more stringent RADAR, allowing for a better distinction between true crosslinked proteins and copurifying contaminants. The STAR method incorporates an extra lysis step before the addition of the previously described "RADAR" buffer, what seems to reduce background proteins after DNA precipitation[511].
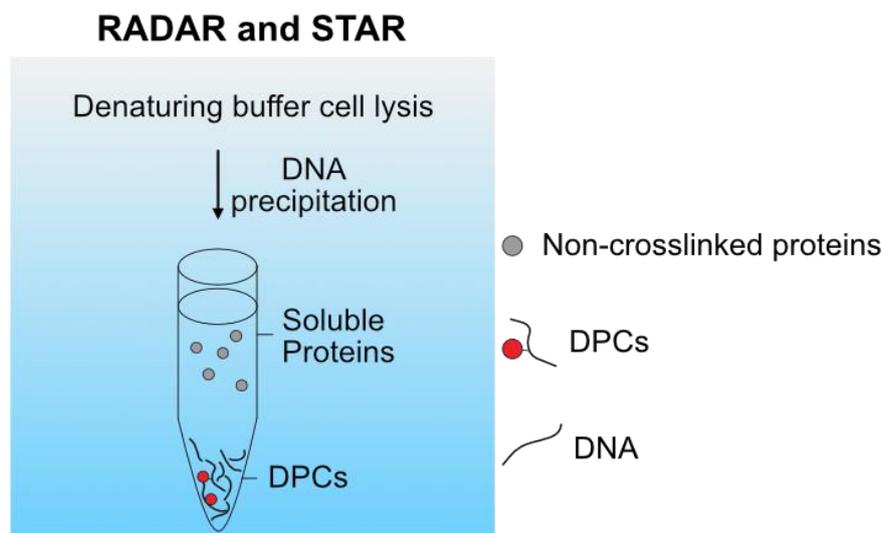
Figure 16. Schematic depiction of the RADAR, or its derivate STAR, assay for the isolation of DPCs. In brief, cells are lysed in a denaturing buffer and DNA is precipitated with high ethanol concentration. The DNA pellet is washed several times to reduce copurifying contaminants and, in a final step, DNA is quantified in different samples and digested, releasing the proteins crosslinked to it.

2.5.2 KCl-SDS and derivates

The KCl-SDS precipitation technique is based on the formation of precipitates of proteins denatured with SDS in the presence of potassium chloride. In brief, cells are lysed in a denaturing buffer that contains SDS. After lysis, they are transferred to tubes and the DNA is mechanically sheered (e.g. by sonication or using a needle). The proteins in the sample are then precipitated by the addition of KCL, which at low temperature will form obvious precipitates that are centrifuged at high speed. The supernatant, which contains soluble DNA is taken for quantification and the precipitate is resuspended again in KCl buffer and resolubilized at high temperature. The process is repeated several times to remove any non-specifically precipitated DNA. In a last step the proteins in the precipitate are digested (e.g. with proteinase K) and the DNA in the sample is quantified. The ratio between soluble DNA and precipitated DNA can be used as a proxy for DPC-abundance. The KCl-SDS is therefore a purely quantitative method which does not uncover the identity of the protein, but it is a good approach for assessing DPC-formation by a compound and it is fast and easily scalable. A method derived from the KCl-SDS assay, the advance recovery of K-SDS precipitates (ARK), uses a combination of RADAR and KCl-SDS. Firstly, cells are processes as for the RADAR and lysed in a chaotropic buffer. After ethanol addition, the DNA is precipitated and resuspended in KCl-SDS buffer, what allows purification of the total protein-DNA complexes from the total DNA previously obtained. This method could be considered a more stringent RADAR assay given that after the KCl precipitation the proteins can also be recovered with an acetone wash and analysed by slot blot or similar[512].
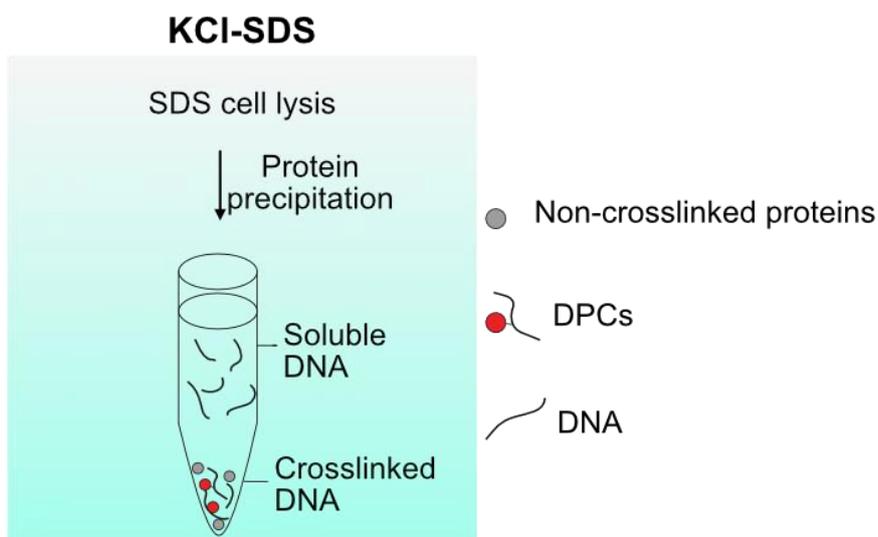
Figure 17. Schematic depiction of the KCl-SDS assay for the detection of DPCs. In brief, cells are lysed in an SDS-containing buffer and proteins are precipitated by the addition of a KCl containing buffer. The protein pellet is washed several times and, in a final step, proteins are digested and coprecipitated DNA is quantified in the sample as a proxy for DPC-formation.

2.5.3 Caesium chloride isolation

Caesium chloride purification of DPCs is probably the most different technique from the ones previously shown, since it does not include any precipitation step that could clearly bias the outcome. In this method, the separating principle is density gradient ultracentrifugation. Cell suspensions are carefully loaded on the top of a layered denaturing solution with increasing concentrations of caesium chloride. During centrifugation in an ultracentrifuge (125.000g), the DNA -which can contain DPCs or not- is separated from the soluble proteins. However, this process is extremely time consuming, the centrifugation step lasts for 16 hours, and once obtained, the sample can be easily cross-contaminated during DNA-retrieval from the tube. Additionally, most of the ultracentrifuges can accommodate a limited number of samples, what also limits the method's throughput[513].

2.6 Model systems for the study of DNA-protein crosslink repair

In addition to the previously mentioned enzymatic DPCs, several models have been established for the study of DPC-repair *in vitro* and *in vivo*. The Flp-recombinase has been engineered to generate aberrant enzymatic reactions which undergo its irreversible crosslink to DNA. This is achieved by a single mutation in a histidine important for its religation step. The Flp-H305L is chemically identical to a TOP1-DPC and it requires the same factors for its repair in yeast (Tdp1, Wss1 and Ddi1)[514]. The advantage of using the abortive Flp, is that it specifically crosslinks in the *FRT* locus, and therefore, the location of the crosslink is known. The system has been coupled to quantitative PCR to monitor the repair of the Flp-DPC at single locus resolution[482]. The study of in vitro replication in *Xenopus* eggs extract is an extremely powerful tool since it allows the immunodepletion of any protein and the monitoring of its effect in the system, regardless of whether this factor is essential in the process. DPC-repair has been extensively studied in these extracts using the repair of a crosslinked HpaII DNA-methyl transferase. The specific crosslinking of this enzyme can be done by using a fluorinated pseudosubstrate (5-fluoro-2'-deoxycytidine) with a mechanism of action similar to that of 5-azadC on DNMT1[515]. This system has shed light on the role of the proteasome in replication-coupled DPC-repair, the E3-ligases implicated in the process, the TLS polymerases or the role of RTEL1 and FANCJ in DPC bypass and repair respectively[405,481,496,499]. Furthermore, the generation of oligo-coupled DPCs has been essential to understand SPRTN DNA-dependent activation. The use of protein-G as a substrate uncovered the DNA-structures responsible for the opening of the enzyme conformation which allow DPC-repair[469]. Moreover, the use of HMCES catalytic domain, SRAP, in the same system also characterized *in vitro* the first DNA-helicase mediating protein unfolding, FANCJ[405].

Introduction

Aim of this study

DPCs are toxic DNA lesions which threat genomic integrity and cell survival. Furthermore, DPCs are a common form of DNA damage generated by chemotherapeutics and environmental toxins. Uncovering the identity of the proteins which crosslink to DNA in different conditions and how are they repair is therefore key to increase our understanding on these pervasive lesions.

The aim of this study is to apply a new method for the detection, identification and monitoring of DPCs, the Purification of x-linked Proteins (PxP), to further understand how such lesions are proteolytically repaired. We used the PxP to uncover the identity of formaldehyde-induced DPCs, which were less complex than expected, since it mostly consisted of crosslinked nucleosomes. Formaldehyde-induced DPCs have a major drawback because they spontaneously revert in cells after damage induction, which complicates the analysis of their repair. Therefore, we focused on a different type of lesion, DNMT1-DPCs.

DNMT1-DPCs are clinically relevant because they are induced by 5-azadeoxycitidine (5-azadC), commonly used to treat myelodysplastic syndrome. Using the PxP, we could monitor the formation and repair of these lesions and observed that their repair is SUMO, ubiquitin and proteasome dependent. Furthermore, we observed cleavage by a second protease, the DPC-repair factor SPRTN. The second aim of this thesis was to characterize which are the SPRTN domains important for the recruitment to and cleavage of DNMT1-DPCs. In addition, we tested whether *SPRTN-ΔC* cells, which mimic the Ruijs-Aalfs syndrome disease variant, could also process such lesions.

We discovered that the UBZ, a domain located in SPRTN's C-terminus, was essential for the recruitment of the protein to these lesions. Moreover, cells engineered to express a SPRTN variant lacking the UBZ, could mimic the phenotype of *SPRTN-ΔC*, concluding that this is the most important feature lost in Ruijs-Aalfs syndrome patients.

Aim of this study

Contribution report

Publication 1

This publication presents a new methodology for the Purification of x-linked Proteins (PxP). After successfully establishing the PxP, we could observe the formation of camptothecin-induced TOP1-DPCs in a dose-dependent manner. We coupled PxP with mass spectrometry to identify the "crosslinkome" of HeLa cells exposed to formaldehyde. We identified histones as the main DPCs formed by formaldehyde. In combination with 5-azadC treatment, we monitored the repair of DNMT1-DPCs in cells with PxP. We discovered that, RNF4 is a master SUMO-dependent ubiquitin E3-ligase that coordinates DNMT1-DPC destruction by the proteasome. Moreover, we uncovered that the metalloprotease SPRTN can also repair these lesions in a replication-independent manner. SPRTN mutants which are causative for disease, lose the ability to repair DNMT1-DPCs. We performed structure-function analysis to map the domain responsible for replication-independent SPRTN function and we found that the loss of the UBZ domain completely phenocopies SPRTN disease variants (C-terminal truncation) phenotypes. Together with Hao-Yi Li, I performed the PxP experiments, cell viability assays, data display and analysis, and FACS and mass spectrometry sample preparation. Mass spectrometry data analysis was performed by Maximilian J. Götz and *in vitro* reconstitution experiments were done by Sophie Dürauer.

Publication 2

This publication is a review on the current knowledge on DPC-repair. The review consists of seven sections. In the first section, we summarized with a general introduction all types of DNA damage and their consequences. The second section comprises all the current understanding on DPC-formation and the identity of different crosslinked proteins. The third section focuses on the proteolytic destruction of the DPC protein adduct by specialized proteases and the regulation of their activity. The fourth section describes replication-coupled DPC-repair which is currently well characterized. The fifth and the sixth section explain alternative DPC-repair pathways by direct hydrolysis or by nucleolytic cleavage, respectively. The last section contains the concluding remarks which also describe current knowledge gaps and future research directions. The review was written by Julian Stingele and I.

Manuscript 1

This manuscript presents a step by step protocol of the PxP, summarizing all its potential applications and comparing it with other methods. The protocol contains experiments with

different crosslinking agents (e. g. etoposide-induced TOP2ccs or formaldehyde). The detailed description of every step would allow any experienced molecular biologist to perform the technique. The experiments in the protocol were performed by Sophie Dürauer and me. The photos and the videos were taken by Hao-Yi Li and Sophie Dürauer. Maximilian J. Götz wrote a mass spectrometry analysis section including an analysis example. The writing, editing and preparation of figures for the manuscript were done by all the authors.

Discussion

1 The PxP allows the unbiased extraction of DPCs from cells

The contribution of DPCs to genomic instability has just recently gained interest with the discovery of SPRTN as a specialized protease targeting these lesions. The methodology established for the detection and monitoring of the protein component of the DPC in cells, has proven previously successful for the study of enzymatic DPCs such as TOP1 or TOP2 but presents major drawbacks. The RADAR assay, and its derivates, rely on DPC-containing DNA behaving as DNA during the precipitation but does not consider that the presence of large adducts or even DNA regions rich in crosslinked proteins could prevent the precipitation of the corresponding DNA fragment. In addition, protein aggregates might also copurify during RADAR given that most of the treatments which induce non-enzymatic DPCs also induce pleiotropic damage such as protein aggregation (e.g. formaldehyde). The PxP is able to detect the presence of such lesions in a completely unbiased manner. With the PxP, we can detect the presence of CPT- and etoposide-induced topoisomerase DPCs in a dose-dependent manner. These DPCs also show obvious modifications when they are resolved by SDS-PAGE and detected by immunoblotting, suggesting that the PxP is also able to isolate posttranslational modifications on DPCs.

With the aim of uncovering the "crosslinkome" after sublethal treatments of cells with formaldehyde, we coupled the PxP with label-free quantitative mass spectrometry. Formaldehyde is a common endotoxin produced in cells during one-carbon metabolism, but also in high local concentrations in chromatin during histone demethylation[321]. Moreover, epigenetic reprogramming has already been linked to DNA damage, given that hematopoietic stem cells lacking Fanconi anaemia genes display lower differentiation rates[318]. Lesions caused by formaldehyde range from RNA-protein crosslinks to ICLs or DPCs[516]. Our work adds a new layer of complexity to the nature of formaldehyde induced DNA damage by identifying histones as the main DPCs formed. Although the formation of nucleosomal-DPCs might be expected, provided the numerous basic residues in closed contact with DNA and their abundance, their complexity should not be underestimated. Nucleosomes are wrapped on large DNA sequences and crosslink to these DNA could easily block replication or transcription. On the other hand, DPCs on ssDNA, can be easily bypassed by replication-coupled DPC-repair or detected and destroyed by global-genome (GG-DPC-) repair. In contrast, detection of histone-DPCs by the SUMO-system is unlikely, given that a crosslinked nucleosome would display a similar conformation to a non-crosslinked one. I would then predict that resolution of nucleosomal-DPCs would only rely on active-chromatin processes, such as replication, transcription or even chromatin (de)-compaction. Of note, it is interesting

to observe the presence of several DNA-repair proteins as formaldehyde-induced DPCs. The presence of APEX1, which cleaves AP sites, or PARP1, which tightly binds DNA-ends, could represent a bias of formaldehyde damage to open-chromatin areas where DNA transactions, mainly repair, is taking place[130,398]. In addition, the absence of non-DNA binders, highlights the stringency of the method. The comparison between non-treated samples and its nuclease control, also identified a new unknown endogenous DPC. The mitochondrial single-strand DNA binding protein SSBP1. Of note, mitochondrial DNA is continuously subjected to the attack of ROS and RNS generated during respiration, what could be causative for SSBP1-DPC formation. In addition, this also proves that the PxP is able to isolate mitochondrial genomes and therefore, it would be also applicable for the study of DPC-repair in this organelle.

2 DNMT1-DPCs are repaired in a replication-independent manner

5-azadC-induced DNMT1-DPCs are generated after the replication fork has incorporated the pseudosubstrate in the DNA[392]. Given the fast repair of these lesions within one to two hours, they constitute a great model for the study of replication-independent DPC-repair[100]. Furthermore, no effect on the repair of DNMT1-DPCs is observed if the cells are cotreated with aphidicolin, a DNA-polymerase inhibitor, proving that replisome collision with the lesion is not required for its repair.

The rapid SUMOylation wave observed after 5-azadC treatment had been previously linked to PIAS4[97,100]. However, we could not confirm those findings and just observed minor reductions in DNMT1-DPC SUMOylation and no effect on repair upon depletion of PIAS4. It is known that several other SUMO E3-ligases are implicated in the DNA damage response. For example, PIAS1 is known to collaborate with PIAS4 in DSB-repair, while ZNF451 specifically targets TOP2-DPCs for their further repair by the TDP2 enzyme[96,451]. Thus, it is likely that high overlap exists between all these enzymes and that loss of any of these factors can be largely compensated by others. Additionally, *PIAS4* knock-outs only display mild sensitivity against CPT, highlighting that TOP1-DPC SUMOylation by PIAS4 is either not as relevant as expected for the repair of the lesion, or that certain back-up enzyme acts when PIAS4 is not present[506]. On the other hand, and in agreement with previous reports, the depletion of RNF4 displayed drastic consequences towards DNMT1-DPC-repair[100]. In support of these data, the inhibition of the SUMO-E1 or the Ub-E1 activating enzymes displayed the same repair defect after two hours.

SUMO1 or SUMO2/3?

DPC SUMOylation by different E3s?
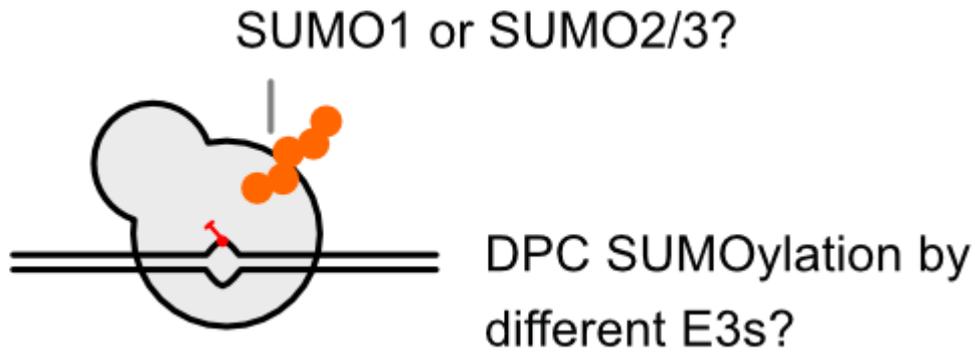
Figure 18. DNMT1-DPCs are formed upon addition of 5-azadC, what triggers its fast SUMOylation by SUMO by different SUMO E3-ligases. Adapted from [517].

Given that an RNF4 RING domain mutant protein could not rescue sensitivity to 5-azadC, we concluded that ubiquitylation by RNF4 is necessary for resolution of DNMT1-DPCs. RNF4 has four tandem SIM domains which presumably bind the SUMO-chains on crosslinked DNMT1, what triggers RNF4 ubiquitylation of the adduct. Still, some controversy surrounds the SUMO isoform linked to DPC-repair. While some reports point towards SUMO-2/3 hybrid chains, others state that SUMO-1 is the main isoform linked to these pervasive lesions after formaldehyde damage. It would be interesting to test whether different DPC-substrates, chromatin processes encountering the lesion or cell cycle phases would bias the SUMO isoform used towards SUMO-1 or SUMO-2/3 chains[518][100]. Strikingly, we also observed that DNMT1-DPCs could also be degraded, albeit much slower, in the absence of RNF4 in a SUMO and Ub-dependent manner. These data suggest the activity of a second StUbL in DPC-repair. Up to date, just one more StUbL in humans has been characterized, RNF111[98]. Whether this is the enzyme which can also target DNMT1-DPCs and trigger its destruction still needs to be determined. Moreover, it would be expected that the loss of this second StUbL would be synthetic lethal with *RNF4* knock-out and that its transient depletion in this background would hypersensitize cells to 5-azadC.

Ub chain

RNF4

RNF111

Another STUbL?
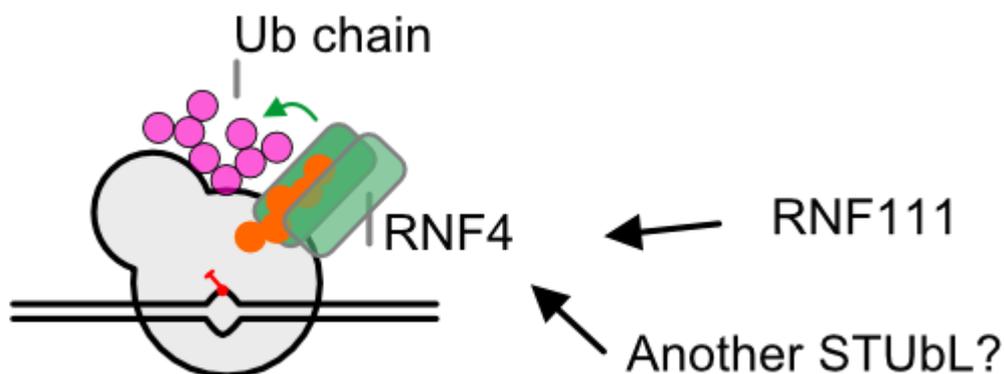
Figure 19. RNF4 is rapidly recruited to polySUMOylated DNMT1-DPCs, what triggers its activation and the polyubiquitylation of the adduct. The loss of RNF4 did not completely abolished DNMT1-DPC-repair, what suggests that there is a second STUbL which might also be implicated. Adapted from [517].

The modification of DNMT1-DPCs by Ub, targeted its fast degradation by the proteasome in a p97-dependent manner. In yeast, the interplay between Wss1 and Cdc48 has been well characterized and it seems to require Ubx5 as a cofactor[465]. In human cells, our data are the first proof of this factor acting independent of SPRTN on these lesions[500]. Despite having proved its role in the extraction of cytotoxic PARP1 from chromatin, and the importance of p97-mediated unfolding of DPCs *in vitro* to facilitate SPRTN cleavage, the exact contribution of p97 to DPC-repair was still unknown[519]. We therefore hypothesize that p97 unfolding is directly upstream of proteasomal degradation and, for DNMT1-DPCs, its activity is essential for repair. Further studies should focus on understanding the role of different cofactors during p97 recruitment. In addition, the type of Ub-chains mediating DPC-unfolding could also be studied by performing coimmunoprecipitations of DNMT1-DPCs. Finally, characterization *in vitro* of p97 DPC-unfolding with different substrates could help to untangle the contribution of each unfoldase to DPC-repair[405,519].

3 SPRTN displays replication-independent DPC-repair

We have uncovered an unexpected role for SPRTN in the repair of replication-independent DPCs. We named this pathway, which comprises SUMOylation, ubiquitylation and recruitment of the proteasome and SPRTN, GG-DPC-repair. We validated the absence of replisome collision against the DPC by monitoring SPRTN autocleavage in cells treated with CPT, formaldehyde and etoposide. SPRTN's autocleavage was not affected in any of the treatments when cells were pretreated with aphidicolin, suggesting that SPRTN's replication-independent role is not a DNMT1-DPC specific feature, but is rather extensible to other DPCs.

The relative contribution of SPRTN during the repair of DNMT1-DPCs is not clear. After the cleavage of the protein, which is almost 200 kDa in size, we observed a faster migrating species which runs ~ 60-80 kDa lower. Whether SPRTN cleavage clearly aids in the repair of this lesion is unlikely. On the one hand, repair of the DPC after longer chases in the presence of proteasomal or p97 inhibitor only seems to accumulate the cleaved product, suggesting further degradation of the fragment by the proteasome. This result could also be misinterpreted given that both treatments, proteasomal and p97 inhibition, deplete the total Ub-pool in the cell, what could interfere with the ubiquitylation of DNMT1, consequently impairing the SPRTN-mediated repair of the cleaved product. In addition, cleavage by SPRTN generates a new N-terminus, which might be targeted for degradation through the N-end rule pathway by specialized E3-ligases[520]. On the other hand, SPRTN might perform sequential cleavage on the adduct facilitating its repair, what could be validated by monitoring repair of a C-terminal tagged DNMT1. Furthermore, if SPRTN is complementing the activity of the proteasome, it would be plausible that an increase in DNMT1-DPC formation by its

overexpression in cells, might require SPRTN for their total removal. In contrast, it might also be possible that SPRTN processes only a subset of lesions, which might be in a specific genomic location (e.g. hetero or euchromatin). Conducting ChIP-seq experiments in cells treated with 5-azadC, either with proteasomal inhibitor or SPRTN-depleted, and performing the immunoprecipitation of DNMT1-DPCs either on their N- or C-terminus would provide an answer to this question. If the chromatin location determines the repair by one protease or the other, proteasomal inhibition would enrich the N-terminal pulldown for a certain chromatin state, while this bias would not be observed for the C-terminal pulldown. The depletion of SPRTN should provide exactly the opposite result, meaning that it is chromatin context what determines repair by either protease.

A second question arising from SPRTN's cleavage relates to the substrate. SPRTN's activity is tightly regulated by DNA-structures and the presence of a distortion within the helix, ssDNA-dsDNA junction, bubbles or DNA-ends, trigger its activation. The presence of unpaired bases was required for DNMT1-DPC cleavage because we could not rescue this cleavage with a ZBD mutant, which cannot bind ssDNA[469]. DNMT1-DPCs form within dsDNA but present a really specific characteristic provided by the methylase reaction. During methylation, the target cytosine is "flipped out" of the DNA. In addition, the unpaired guanine on the other strand is displaced due to the DNMT1's DNA intercalation, what itself triggers the flipping out of a second nucleotide from the DNA strand [521,522]. Given that this state is stabilized by 5-azadC, it would be possible that this distortion within the helix would be enough to allow SPRTN's activation and cleavage[469]. Furthermore, the presence of DNMT1-DPCs in "clusters" due to sequential crosslinking in GC-repeats, might generate local melting on the DNA-template due to the presence of several flipped bases in closed proximity[521,522].

Interestingly, SPRTN's expression is tightly regulated in a cell cycle manner, as demonstrated by its depletion after palbociclib treatment. Palbociclib is a newly developed drug which stalls cells in G1 by inhibiting CDK4/6 without displaying any toxicity[523]. SPRTN was almost non-detectable in cells treated with this drug, what raises the possibility that GG-DPC-repair is mostly active in S/G2/M phases but not in G1. To prove this hypothesis, experiments in which the repair of TOP1- and TOP2-DPCs in palbociclib treated cells could be conducted. This experiment would be interesting for two reasons: first, the treatment with CPT and etoposide in non-replicating cells, will mostly bias the presence of TOP1 and TOP2 crosslinks to transcription bubbles, what could allow the discovery of factors implicated in tolerance towards these lesions during transcription. Second, the degradation of these DPCs in a SUMO and Ub-dependent manner would suggest that the RNF4-proteasome/SPRTN axis is functional throughout the whole cell cycle. The function of SPRTN in specialized cell cycle phases raises the possibility of an alternative pathway for the repair of DPCs during transcription. In this

scenario, SPRTN is unlikely to be involved. First, flavopiridol treatment before 5-azadC incorporation, which completely inhibited transcription, did not affect DNMT1-DPC-repair. Second, transcription obstacles are dangerous throughout all the cell cycle and not just in S/G2/M phases, what suggests that other mechanisms are involved in the tolerance of these obstacles during RNA polymerase stalling. If such a mechanism would exist, it would probably involve the transcription-coupled NER protein CSB. CSB is required for the repair of formaldehyde-induced damage and Cockaine syndrome patients, who carry mutations in this gene, display a set of symptoms which cannot be reconciled with a role of CSB just repairing UVC-induced lesions[324]. This raises the possibility that if CSB could act as a sensor for RNA polymerase stalling, what DPCs would cause, and recruit CSA, which could then drive the polyubiquitylation of the adduct with its consequent degradation by the proteasome. The degradation of the protein could then recruit downstream factors of the pathway and mediate the nucleolytic removal of the peptide, allowing transcription to resume. In this scenario, CSA would also probably mediate the ubiquitylation of the RNA polymerase, what could also trigger its eviction by p97 and destruction by the proteasome[156,324,524].
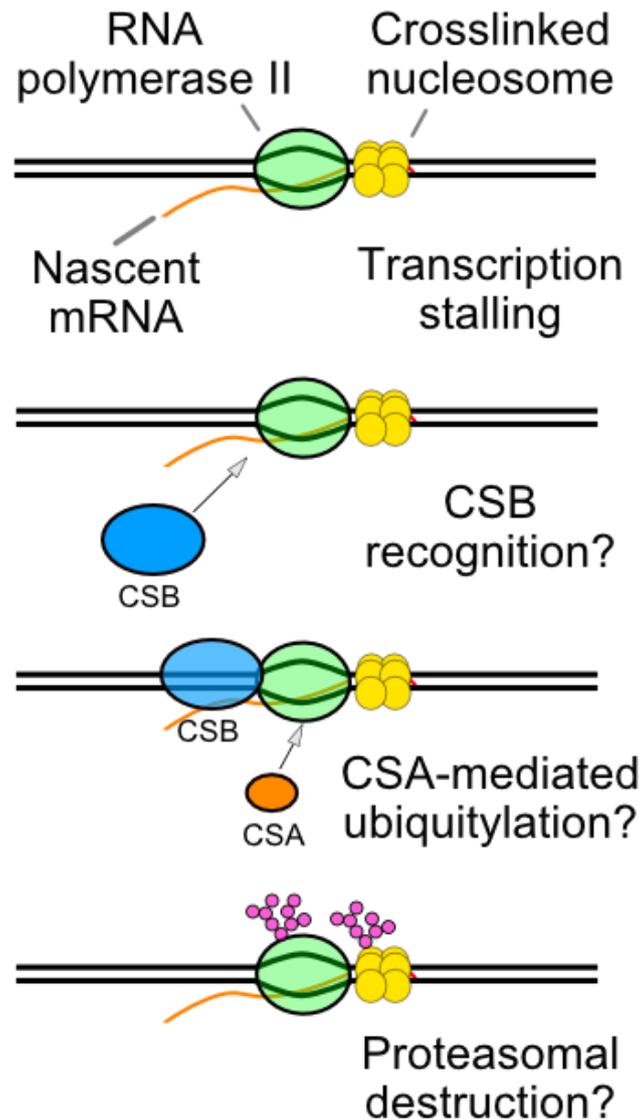
Figure 20. Hypothetical schematic depiction of transcription-coupled DPC-repair. The detection mechanism of a DPC would be the stalling of an RNA-polymerase, what would recruit CSB, leading to CSA mediated ubiquitylation of the DPC but also of the RNA polymerase itself.

We generated cell lines carrying the Ruijs-Aalfs *SPRTN* patient's genetic variant *SPRTN-ΔC*. SPRTN-ΔC is a hypomorph and displays reduced protease activity *in vitro*. In addition, SPRTN contains an NLS on its C-terminus which is lost in the patient's variant. Therefore, it is also mislocalized, what can be detected by immunofluorescence, mostly in the cytoplasm, because it is probably continuously exported from the nucleus. Strikingly however, the patient variants SPRTN-ΔC lacked the ability to cleave DNMT1-DPCs not because of its mislocalization, but because of the loss of an extra feature on its C-terminus[323,476]. Although we could also observe residual activity by SPRTN-ΔC on DNMT1-DPCs upon proteasome and p97 inhibition. The fact that SPRTN-ΔC can still perform protease activity on this DPC, albeit much worse, might explain why *SPRTN-ΔC* cells are still viable regardless of *SPRTN*'s essentiality. Finding the alternative mechanism for which SPRTN-ΔC mediates DNMT1-DPC-repair is still an open question that must be addressed. In agreement with these data, Ruijs-Aalfs patients display a

set of defects on non-replicative tissues such as cataracts or fat tissue dysfunction, which are not explainable by a purely replicative role of SPRTN[476].

We mapped the critical domain needed for this activity to the very C-terminal UBZ and validated it with rescue experiments and with the generation of *SPRTN-ΔUBZ* cells. These cells completely phenocopied SPRTN-ΔC, suggesting that this is indeed the most important feature lost in Ruijs-Aalfs syndrome. SPRTN's UBZ domain is important for the shielding of the protein monoubiquitylation, which is also important for its regulation. DPC-induction, for example by formaldehyde, triggers SPRTN's deubiquitylation by the DUB USP7. Two other DUBs have also been implicated in SPRTN's, VCPIP and USP11, and the understanding of the contribution from each DUB to SPRTN's deubiquitylation is still missing. Interestingly, the deubiquitylation of SPRTN does not seem to specifically trigger SPRTN's recruitment to chromatin but rather reduce its autocleavage[471-473]. This raises the possibility that the loss of monoubiquitin would release the UBZ and allow its binding to the ubiquitin chains on the DPC. SPRTN's UBZ can bind both K48- and K63-chains, but K48 seem to be preferentially conjugated to SUMOylated DPCs[100,466]. Surprisingly, we could not observe any contribution of SPRTN's PIP or SHP boxes, what we proved by fully rescuing DNMT1-DPC cleavage with a variant lacking both domains but still located in the nucleus, SPRTN-Δ241-400. Moreover, the insertion of a single mutation in this construct, the SPRTN-Δ241-400-UBZ*, which lacks a functional UBZ domain, almost completely abolished SPRTN's activity on the lesion. SPRTN's PIP box is important for its recruitment to laser and UVC irradiation damage sites and hydroxyurea-induced foci. The type of stress induced by those treatments, suggests that the PIP box is relevant for SPRTN's repair where replication stress, for all treatments, or damage-induced synthesis, by UVC and laser microirradiation, takes place[466]. Therefore, the PIP box seems uniquely relevant in a scenario in which the lesion is encountered during the replication of a DNA template, which requires PCNA[323,466]. The SHP motif loss did not influence DNMT1-DPC cleavage either. These data supports the fact that inhibition of p97 during DNMT1-DPC-repair did not decrease SPRTN cleavage but rather increased it. Furthermore, the role of the SHP motif in SPRTN's activity seems to be strictly restricted to a replication-coupled activity, where TEX264 might also be implicated driving relocation of TOP1ccs to the nuclear membrane and facilitating their repair[500]. In addition, the role of the SHP domain was also studied in conditional *SPRTN* knock-out mouse embryonic fibroblasts, which display genomic instability, senescence and undergo cell death after few cell divisions upon tamoxifen treatment. The *SPRTN-SHP* mutant could just partially rescue replication fork progression in DNA fiber assays after UVC-induced DNA damage[476].

The genetic relation between *SPRTN-ΔC* and *RNF4* is complex and the phenotype severity is cell line dependent. While in HeLa T-REx *SPRTN-ΔC* cells we observed compromised cell

growth upon RNF4 knock-down, in U2OS T-REx *SPRTN-ΔC* cells we observed a striking synthetic lethality. This phenotype was surprising because RNF4 is completely upstream of SPRTN's activity during GG-DPC-repair, what we validated with autocleavage assays in wild-type and *RNF4* knock-outs cells after DPC induction. In agreement with this epistasis, the depletion of RNF4 in HeLa T-REx *SPRTN-ΔC* cells did not enhance 5-azadC sensitivity. In contrast, the depletion of SPRTN in *RNF4* knock-out cells by knock-down showed the opposite phenotype, while in both experiments cells displayed a compromised cell growth. SPRTN is involved in replication-coupled DPC-repair in addition to GG-DPC-repair[323,468]. Replication-coupled DPC-repair is completely independent of RNF4, what explains the growth defects and lethality observed when both pathways are affected. Given that DNMT1-DPC-repair does not seem to highly rely on SPRTN activity, *SPRTN-ΔC* cells are not sensitive to 5-azadC because they have a fully functional proteasomal-RNF4. In agreement, *RNF4* knock-out cells display 5-azadC sensitivity because in this scenario SPRTN's and proteasomal GG-DPC-repair are fully compromised. The loss of SPRTN in *RNF4* knock-out cells did enhance sensitivity to 5-azadC, what could be explained by a double reduction in GG-DPC-repair, due to the loss of RNF4, and replication-coupled DPC-repair, due to the loss of SPRTN. It would be possible that DNMT1-DPCs not removed by RNF4 and the proteasome could be encountered in the next S-phase, what would require SPRTN. It would be tempting to speculate that loss of SPRTN in *RNF4* knock-out cells would also cause the accumulation of endogenous DPCs, whose identity could be uncovered with the PxP.

4 Downstream lesion processing

Cleavage by SPRTN or the proteasome leaves a peptide that must be removed for full lesion resolution. Therefore, the enzyme required for peptide removal should have nuclease activity. The loss of these nucleases would confer sensitivity towards DPC-inducing agents but the contribution of each of these factors to repair might be context dependent. One example is the ERCC1-XPF complex, which participate in the repair of TOP1ccs probably by cutting the 3' end on the DNA strand and releasing the peptide after protease activity on the protein adduct[525,526]. Furthermore, the loss of both factors sensitizes cells to formaldehyde, highlighting a potential implication of this complex in DPC-repair [506]. A second nuclease which also enhances sensitivity to CPT is MUS81, which forms complex with EME1 or EME2. MUS81 has been implicated in DNA structure dependent cleavage during fork convergence at a TOP1-DPC, what would generate a DSB and allow HR[527]. Thus, MUS81 might just be involved in DPC-removal during replication-dependent repair, where the DNA structure would allow its activity, which requires a 3' flap, a replication fork or a nicked holliday junction [528]. Additionally, another nuclease FEN1, has recently been involved in DPC-repair in a ADP-ribosylation dependent manner[529]. In addition to the previously mentioned nucleases, APE2

has recently gained interest due to its ability to remove 3' blocked ends, what includes those generated by TOP1cc peptides[530]. The size of the peptide as well as the factors required for its removal could be potentially studied by PxP. Moreover, an antibody raised against the DNA-protein interfaced of the crosslinked DNMT1, could be of great used since it would allow the study of the peptide repair by immunofluorescence as well as the identification of factors implicated in the removal through Fluorescence-Activated Cell Sorting (FACS)-based genetic screens.

The peptide generated can be alternatively bypassed during replication by TLS polymerases [499]. In frog eggs extracts bypass of the adduct is purely Polζ dependent but *in vitro*, almost any TLS polymerase can bypass small peptides[531]. Furthermore, TLS also requires the E3-ligase RFWD3, which also targets the DPC for destruction by mediating its polyubiquitylation during replication-coupled DPC-repair[497]. Therefore, RFWD3 might be the factor coordinating both TLS and DPC-repair pathways. RFWD3 as a master regulator between both pathways, might also explain the associated role of SPRTN with TLS, in which it was proposed to limit mutagenicity by TLS-polymerases. Alternatively, bypass of an intact DPC requires its partial unfolding by the helicase FANCJ and its cooperation with polζ-Rev1[405]. Given that bypass of an intact protein adduct is possible with the aid of FANCJ, it would be tempting to speculate that in certain conditions, for example with limited Ub supply or no SPRTN nor proteasomal availability, DPCs could be tolerated during replication and later targeted by GG-DPC-repair. It would be interesting to test whether this is the reason why several HR factors have been implicated in DPC-tolerance[460,532]. Finally, generating a non-Ub modifiable DPC, for example by mutating all the surface lysines on DNMT1, would allow us to understand whether this is the reason why certain DPCs are preferably "tolerated" instead of targeted.

5 Future directions and open questions

In addition to DNMT1, new DPC-models should be established to study their repair in different cell cycle phases and their effect on chromatin transactions. One model for the study of replication-coupled DPC-repair could be HMCES, which can be easily detected upon UVC exposure in synchronized cells. However, its reversal mechanism makes difficult it's the study of its repair by DPC-proteases[406,407]. An alternative to HMCES could be the *mycobacterium smegmatis* UdgX, which crosslinks to uracils within ssDNA. A recent paper showed that UdgX in coexpression with APOBEC3A, which deaminates ssDNA cytosines to uracils, is able to form clear foci in human cells upon replisome stalling. The authors induced foci formation by treating cells with cisplatin, but any compound slowing fork progression and generating ssDNA stretches could be potentially used[415].
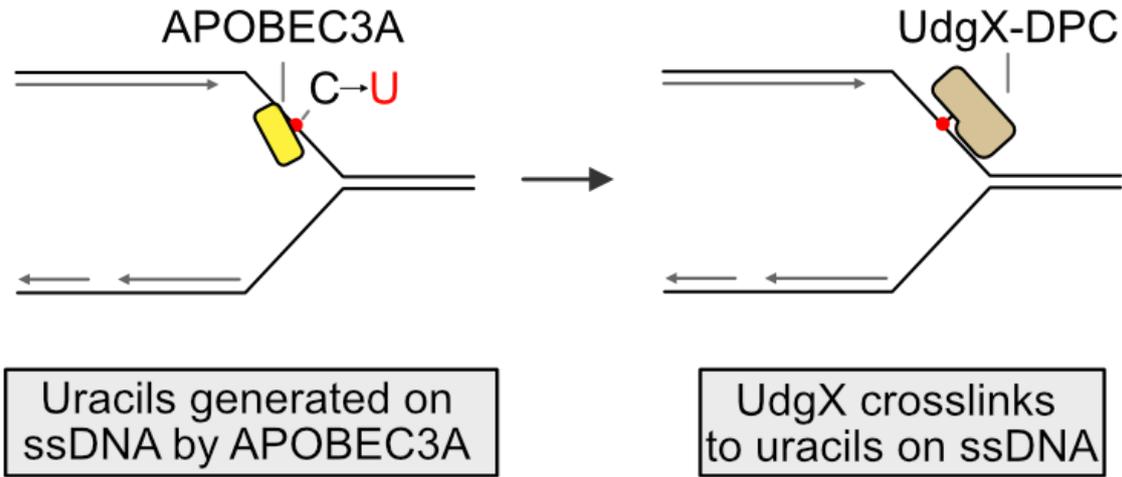
Figure 21. Schematic depiction of a model system for the study of replication-coupled DPC-repair. The coexpression of the cytidine deaminase APOBEC3A and the mycobacterium smegmatis UdgX, would allow for the generation of UdgX crosslinks on ssDNA. The process could be triggered by the CMG-polymerase uncoupling.

An alternative to study DPC-formation independent of cell cycle could be the generation of *TDP1* knock-out cell lines complemented with SCAN1 *TDP1* mutants. The formation of TDP1-DPCs could then be induced by simply adding a TOP1 poison, CPT or topotecan for example, what would trigger the formation of TDP1-DPCs due to faulty religation during the TOP1 peptide release[533]. The advantage of this model system would be that cells could be stalled in G1 and the formation and repair of these lesions could be easily monitored in this cell phase. The main disadvantage would be that the formation of TDP1-DPCs would be linked to the previous formation of TOP1-DPCs, which are themselves extremely toxic for the cell.
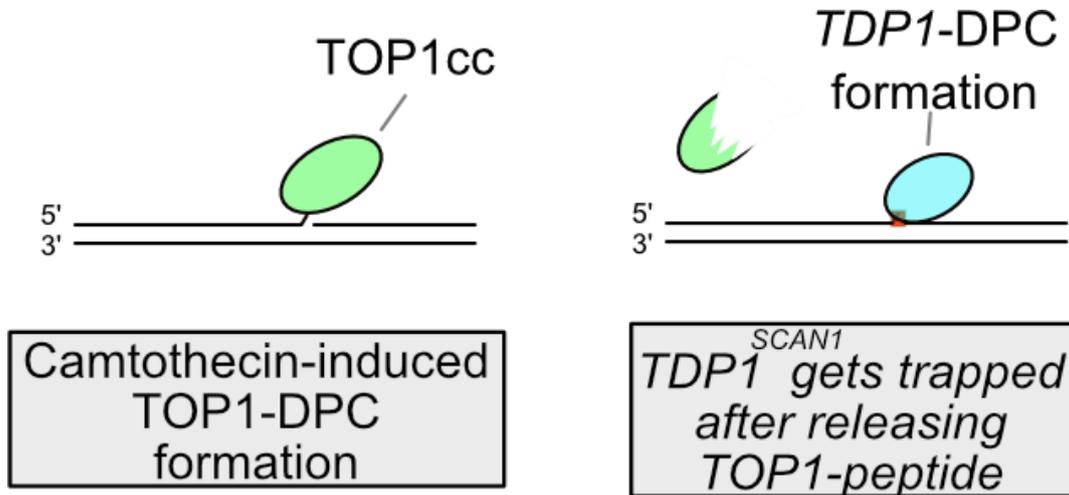
Figure 22. Schematic depiction of a model system for the study of cell cycle-independent DPC-repair. In brief, a TDP1 knock-out cell line complemented with a TDP1^{SCAN1} mutant is treated with CPT, what generates TOP1ccs/DPCs. These lesions are attacked by the proteasome in a RNF4-dependent manner, the remnant peptide is hydrolysed by TDP1, but the SCAN1 variant cannot get released from DNA and becomes irreversibly trapped.

One type of enzymatic DPC that would circumvent the disadvantages of the two previous ones would be MGMT. MGMT repairs the promutagenic O6-alkyl-guanine lesions by a direct transfer of the aberrant alkyl group to an acceptor cysteine in its catalytic center[534,535]. This

reaction is irreversible and alkyl-MGMT is rapidly degraded by the proteasome upon reaction completion. MGMT reacts with the 1,2-dibromoethane, historically used as a gasoline additive and in fumigants, to form a toxic episulfonium ion intermediate. This intermediate reacts with DNA, mainly guanines, generating MGMT adducts[536,537]. Thus, higher expression of MGMT is associated with increased genotoxicity by 1,2-dibromoethane[538]. Therefore, MGMT-DPCs would be an interesting model for the study of the repair of such lesions. They would be easy to induce, by treatment of cells with 1,2-dibromoethane, the identity of the crosslinked protein is known, and in addition, their induction would be cell cycle independent.

The use of reactive compounds which induce DPCs, within other lesions, could also be a subject of study with the PxP. Cisplatin DPCs, would not present the same instability problem as formaldehyde-induced DPCs, given that they cannot be subjected to reversal by hydrolysis and they must be actively repaired[330]. It is highly probable that cisplatin generates a "crosslinkome" relatively similar to the one observed upon formaldehyde exposure. The main disadvantage of using cisplatin would be the generation of ICLs, which are the main toxicity mechanism of action of platin compounds[333,334]. Finally, it is of great clinical relevance to study radiation-induced DPCs. Radiotherapy is generally used in cancer treatment and it is extremely efficient in killing cells due to its capacity to generate SSBs and DSBs. Interestingly however, radiotherapy generally relies on the activation of oxygen closed to DNA, what generates ROS that can attack the different nucleotides. It was already observed that lower concentrations of oxygen surrounding the lesion, what can happen in the hypoxic tumour core, generates DPCs instead[58,341,342]. Uncovering the identity of these proteins and identifying tumour-specific markers which can affect their repair could improve treatment with radiotherapeutic agents. In addition, studying whether their repair is mediated by the proteasome, SPRTN or both, if not implicating new factors, could lead to the development of new DNA-repair targeted inhibitors.

$[Pt(NH_3)_2Cl_2]$
Cisplatin

$[O_2] \downarrow$
Ionizing radiation

Any DNA-interacting
protein
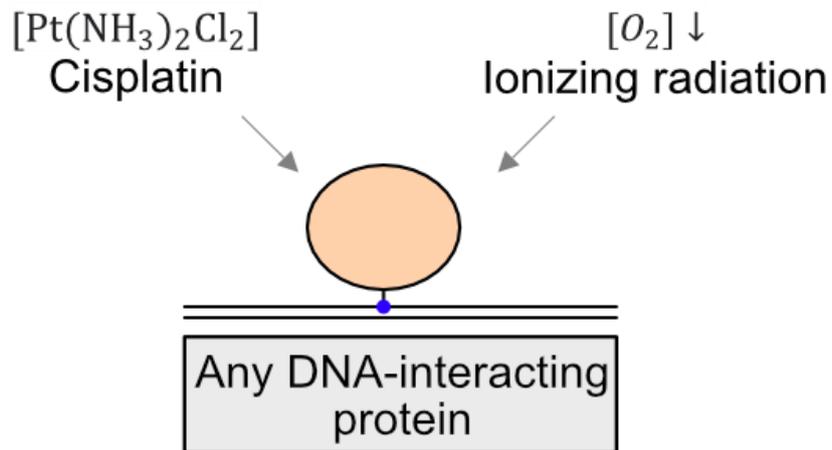
Figure 23. Schematic depiction of a model system to study non-enzymatic DPC-repair. Uncovering the identity of the crosslinked proteins by ionizing radiation -in hypoxic conditions- and cisplatin could be of great clinical relevance.

Future applications of the PxP, which could include the sequencing of DNA fragments crosslinked to specific proteins, would require sample recovery optimization from the plug. Furthermore, the modification of crosslinked proteins could also be easily studied by coupling the PxP with diGLY proteomics, for Ub, or phosphoproteomics. However, the PxP presents two major drawbacks. First, the scalability is limited and even though several plugs can be casted per condition, the number of cells that can be casted in a plug is restricted and would require lysis and electroelution optimization. Second, time sensitive treatments with non-competitive inhibitors, such as CPT or etoposide, might lead to underestimation of DPC-formation due to slow plug-casting step.

Finally, understanding what determines pathway choice during GG-DPC-repair, will be of great interest to understand the role of SPRTN during DPC-repair. Whether SPRTN aids the proteasome in GG-DPC-repair, but can act independently during replication-coupled DPC-repair, or if they always act together. Uncovering these questions will shed light on the essentiality of SPRTN in mammals and its role in cancer, aging and disease.
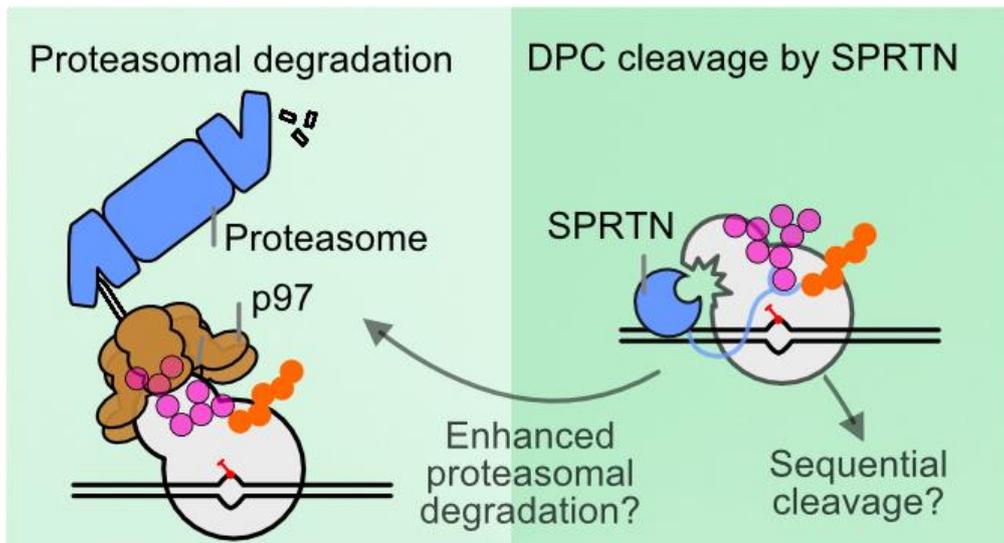
Figure 24. Proteolysis of a DNMT1-DPC by the proteasome and SPRTN. The destruction of the protein adduct by the proteasome requires its polyubiquitylation and unfolding by p97. On the other hand, SPRTN cleavage just requires ubiquitin. Whether SPRTN cleavage enhances proteasomal activity on the adduct by engaging N-end rule pathway E3-ligases or whether SPRTN is able to completely degrade crosslinked DNMT1 by engaging in sequential cleavage is still and opened question. Modified from [517].

List of abbreviations

5-azadC	5-aza-2-deoxycytidine
6-4 PPs	6-4 photoproducts
A	Adenine
AP site	Apurinic/apyrimidinic site
APE	AP site endonuclease
APOBEC	Apolipoprotein B mRNA editing enzyme
BER	Base-excision repair
BIR	Break-induced replication
Bp	Base pair
BR	Basic region
C	Cytosine
CDKs	Cyclin-dependent kinases
Cisplatin	Cis-diamminedichloroplatinum (II)
CMG	Replicative helicase composed of Cdc45 MCM2-7 and GINS
CPD	Cyclobutene pyrimidine dimers
CPT	Camptothecin
CS	Cockaine syndrome
DDR	DNA damage response
DNMT1	DNA methyltransferase 1
DPC	DNA-Protein Crosslink
DSB	Double-Strand break
dsDNA	Double-stranded DNA
DUB	Deubiquitylating enzyme
EBNA1	Epstein-Bar virus nuclear antigen 1
FAD	Flavin adenine dinucleotide
FA	Fanconi anaemia

List of abbreviations

FAN	Fanconi anaemia complementation group
G	Guanine
GG-DPC	Global genome DPC-repair
GG-NER	Global genome nucleotide excision repair
HMCES	5-hydroxymethylcytosine (5hmC) binding, ES cell-specific
ICL	Interstrand DNA crosslinks
IR	Ionizing radiation
KBrO$_3$	Potassium bromate
KCl-SDS	Potassium chloride and sodium dodecyl sulfate coprecipitation assay
MMS	Methyl methanesulfonate
MRN	Complex formed by MRE11, RAD50 and NBS1
NAD+	Nicotinamide adenine dinucleotide
PAR	Poly-ADP-ribosylation
PARP	Poly-ADP-ribose polymerase
PARPi	PARP inhibitors
PCNA	Proliferating cell nuclear antigen
PIAS	Protein inhibitor of activated STAT
PIP	Proliferating cell nuclear antigen-interacting protein box
Pol	Polymerase
PxP	Purification of x-linked Proteins
RADAR	Rapid approach to DNA adducts recovery
RING	Really interesting new gene
RNAP	RNA polymerase
RNS	Reactive nitrogen species
ROS	Reactive oxygen species
RPA	Replication protein A
SENP	Sentrin-specific protease
SHP	p97 interacting motif

List of abbreviations

SIM	SUMO-interacting motif
SPRTN	SprT-like N-Terminal Domain
SRAP	SOS response associated peptidase
SSB	Single-strand break
ssDNA	Single-stranded DNA
StUbL	SUMO-targeted Ubiquitin Ligase
SUMO	Small Ubiquitin-like MOdifier
T	Thymine
TC-NER	Transcription coupled nucleotide excision repair
TDP	Tyrosyl-DNA phosphodiesterase
TLS	translesion synthesis
TMZ	Temozolomide
TLS	Translesion synthesis
TOP	Topoisomerase
TOPcc	Topoisomerase cleavage complex
TP	Terminal protein
TS	Template switching
U	Uracil
Ub	Ubiquitin
UBZ	Ubiquitin-binding Zinc finger
UdgX	Uracil DNA glycosilase X
UV	Ultraviolet
Wss1	Weak suppressor of smt3
XP	Xenoderma pigmentosum
ZBD	Zinc-binding domain

List of abbreviations

Acknowledgements

Bua, this wasn't easy, but it could have also been much worse. During these last years (4.25 since I started my PhD, but almost 6 since I first stepped in the lab), my life has been a rollercoaster of emotions which ranged from "we are almost there" to "just two more weeks and we submit". There are many people that I have to thank for being part of my life during this time, and of course, making it much better. (Prepare for this, these acknowledgements are long).

I would like to start thanking my super boss, Herr Professor Doktor Julian Stingele, for being a constant source of support, help and guidance. Even though most of the times in science the road is uphill, you are always there to push us forward and you should know that we all acknowledge and appreciate what a great supervisor you are. You gave me the opportunity of joining the lab and I will always be thankful for it. Now that I have written such a magnificent thesis and, hopefully, I will defend soon, there is just one thing that constantly comes to my mind "Auch ein blindes Huhn findet mal ein Korn" (inside joke of course). I would say that the only thing that makes this lab more amazing than having you as a supervisor is clearly the people in it, so I will of course dedicate some words to them.

By order of arrival I would like to start thanking Anja, the best technician even, the most reliable person in the world and the one who makes things happen. You cannot imagine how much have we missed you around these two years and how happy we are that you are finally back. Additionally, you should keep the pink hair for two reasons: first, it just fits you and I cannot imagine you without it, and second, it makes it really easy to locate you in a group of people (great when we were going to the Klinikum Mensa). Second in line is, of course, Doktorin Reinking. Hannah, it was really great to have someone to fight with, laugh with, and sometimes, loot the downstairs' Mensa for free food after seminars. It was really fun to have you around and I am really happy that everything has worked out so well for you. In my memory will always remain your love for EMSAs and how exhausted you were after lab meetings. Ricky (A.K.A Dr. Li), I absolutely loved working with you. It was amazing having you on my side and you cannot imagine how much have I learnt from you. My experience would have been really different (and for sure much worse) if I would not have had you by my side. You are super patient, talented and easy to work with (even though I know I am a bit, just a bit, difficult sometimes). I am already expecting the invitation to give a talk in Taiwan when you become a group leader. I will always remember the beers after work and, most important, your legacy when you ask me something "sounds really interesting" (proceeds to do whatever he wants considering 0 of what you just told him). Max, it was absolutely fantastic to have you

in the lab. You always try to help, support and understand what others are doing and to give feedback in their projects. This effort sometimes is unappreciated, but it is what makes you a great scientist. I am sure I will see an AG Götz at some point somewhere around here, but don't take it as a must. Remember that academia sometimes is pretty unfair and that you can always use your data analysis skills in a company to make real money. I guess I can add you here Niels. It is just a pity that we got to know each other so late but I am also sure that if we would have worked in the same lab the world would have collapsed. You are really fun and you laugh at my jokes, seriously, what else can I ask for? P.S. Hurry up with the thesis so we can defend together.

Wait, is the next person who I think it is? Is it the absolutely best, awesomest, impresivest, intelligentest, talentedest (really long etc.) postdoc (exPhD) in the lab? Of course, it is Doktorin Zhao. Shubo, there is not a single day in which I do not get impressed by you. You are extremely talented, hardworking, smart and a great scientist. But, more important than this, you are a wonderful person and a great friend who is always there to lend a hand. It is extremely funny to sit next to you so I can annoy you from time to time (on my defence I am almost like your HiWi). I will miss you a lot when you go back to China but I do promise to visit soon (and this will go online so I should not lie). Of courrrrrse, the next person is Denitsa, what can I say about you that I have not already told you? The Äkta queen and a super doer. I could give you the most difficult protein on earth and the next week it would be purified on my bench (but always on ice). If I know one thing, is that I will never compete against you because you would obviously win (and probably shout at me). You are not just an amazing scientist but also a wonderful friend and I love having you by my side. Jacqueline, the emotional compass of the lab, the one who is always there to give you a hug when you are having a bad day, or tell you the latest gossip. It has been really fun to work with you and soon you will join me in the goal line (wait, is there a goal line?). Honestly, if I would not have you there I don't know who would laugh at my extremely, extremely, extremely funny (sometimes slightly mean) jokes. You are also a superscientist and I am sure that you will also write an awesome thesis (almost as good as mine, you know modesty is not my thing). Sophie, yes, there are also few words for you (I know you were already getting nervous and shaking your arms up and down). You are just truly amazing and I really love working with you. You are extremely smart, technically impressive and extremely funny when angry. I envy to the extreme your absurdly well-organized data/labbook/schedule (with millions of colours that I cannot even differentiate). In addition to this brutal set of professional skills, you and Denitsa have been a constant emotional support because you are both just lovely (and extremely fun to work with). MaxD, you have a great creative mind but also an expert on driving me insane sometimes (it is also cool that we can shout at each other but the next day is all fine). I hereby acknowledge that

we are friends (you can use this as a proof when I mock you). P.S. As a defence present you can get me a ticket to watch one football match involving any of the many clubs you support (I hope Sevilla FC is now between them).

One of the things I liked the most of doing a PhD has been the opportunity to supervise students, who helped me to constantly learn and improve. Chimeg, you are an extremely talented scientist and you will success wherever (and whoever) you work (with). Despite not having been able to help you so much, I am more than happy to see that things have turned out so well for you and I am sure that, after such a stressful master thesis, your PhD will be easy peasy (it also helps that you are really motivated and curious, which are the most important scientific quality). Sara, you were my first student and it was extremely fun to work with you. You are an achiever and a science machine and I am super happy every time I see you at the MPI. I guess that after working with the infamous SSBP1, purifying proteins is not a challenge for you anymore.

Erik (or Doktor Van de Logt), after reading such inspiring acknowledgements from you, I cannot really find the words to thank you (of course it is sarcastic, he wrote 4 lines where I was not even included). Let me first start by wishing you and Anja (and the soon +1) the best of the lucks in the future. It has been great to have you by my side these years and that you always found the time to have some beers (or cook something nice).

I also would like to thank our neighbours from the Jae lab but specially the person who started at the same time and is submitting at the same time. Eva, you are one of the strongest people I ever met and you deserve your PhD more than anyone. You are truly amazing and, even though you had not much luck during these years, you are an awesome scientist. It is a true shame that academia misses out on scientists like you. I would like to thank the R4 community but especially the people I got to meet because of it. An especial mention to Thomas, who will also get his thesis done also soon, and Fillip, who is super fun to hang out with (and who will keep the millenary quiz tradition in the IMPRS). I would like to thank my TAC members, Christof and Axel for the advice and the support during my PhD. In addition, I would like to thank Chris for the very fruitful TOPORS collaboration.

Thanks to the IMPRS I have gotten to know wonderful people who are exactly in the same situation. I would like to start thanking the coordinators of the school Hans-Joerg, Marta and Ingrid, who made it all possible. Mr. Karl (or future doctor Olsen), I still do not know how will you fit all the massive amounts of titles you hold in your business card. You are a super scientist and whether you end up doing a postdoc or not, you will achieve whatever you set your mind to. P.S. Still waiting for the kimchi invitation. Georg (almost doctor Jocher), you will do a stellar postdoc in Harvard. You have everything a scientist need to succeed and I sure

that you will. You are also a super fun person; a great friend and I will always remember our iconic ice cream picture in Horumersiel (let's see if Marlene ever invites us again). Dennis (future Doktor Feigenbutz) my champions league mate and who always joins for the plans, even for the spontaneous ones (what makes him a one in a kind within Germany). Special mention to my Schwille mates, but with special emphasis on three of them. Sveto, you are exceptionally bighearted, fun and generous and I always look forward to our monthly hangouts were we (just) eat junk food and watch JJK. Henry, you are definitely my favourite neighbour (even though you are a natural Spain-born enemy, 50% British 50% French, and also the only one under 80 years old in the neighbourhood). I love chatting with you and your sense of humour is almost as good as mine. Anastasija, thanks for helping me to troubleshoot my life problems and mainly for helping me to laugh at them (even more).

Adrián (o doctor Merino Salomón), creo que tú te mereces un capítulo aparte. No te puedes imaginar cómo de importante has sido durante estos años, cuánto me has ayudado y lo afortunado que soy de que seas mi amigo. Solo decirte, que eso no va a cambiar me vaya a Boston, Londres o las Filipinas. Indudáblemente echaré especialmente de menos las tardes de Naruto y shishita durante el covid, los domingos de fabada, siesta y película y los viajes en coche. Más te vale visitarme pronto. Cristina, you are a natural-born scientist who will achieve anything she sets her mind to. You are sharp, smart, hardworking and talented and you are an extremely interesting person I love expending time with. Your idea of setting up this "tradition" of having breakfast together once a week is really great and I will really miss it when I go away. Special thanks for bringing Adrián and I back home when required and for taking care of us.

During these years I have also met wonderful people out of the PhD world and some of them became friends that I will keep for the rest of my life. A special thanks goes to Danila and Juliano who are just wonderful and supportive friends. Aljosha, you have been one of my greatest discoveries. I still remember the first day in the neurobiology lecture and thinking "this guy seems normal, gonna sit next to him" (I was wrong). I know I will always have you by my side and I am looking forward to our trip to (insert destination here).

Por ultimo, pero no por ello menos importante, quiero agradecer a mis padres y mis hermanas, y a Curro, por ser un ejemplo constante de trabajo duro y una fuente inagotable de apoyo. Gracias también a mis padrinos por encontrar siempre un huequecito para llamarme. Mención especial a mi hermana María, que se convertirá pronto en la segunda doctora Weickert de la familia.

References

1.   Wilkins, R.J. & Macleod, H.D. Formaldehyde induced DNA-protein crosslinks in Escherichia Coli. *Mutat Res* **36**, 11-6 (1976).

2.   Watson, J.D. & Crick, F.H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737-8 (1953).

3.   Wolanski, M., Donczew, R., Zawilak-Pawlik, A. & Zakrzewska-Czerwinska, J. oriC-encoded instructions for the initiation of bacterial chromosome replication. *Front Microbiol* **5**, 735 (2014).

4.   Marians, K.J. Prokaryotic DNA replication. *Annu Rev Biochem* **61**, 673-719 (1992).

5.   Neylon, C., Kralicek, A.V., Hill, T.M. & Dixon, N.E. Replication termination in Escherichia coli: structure and antihelicase activity of the Tus-Ter complex. *Microbiol Mol Biol Rev* **69**, 501-26 (2005).

6.   Duggin, I.G., Wake, R.G., Bell, S.D. & Hill, T.M. The replication fork trap and termination of chromosome replication. *Mol Microbiol* **70**, 1323-33 (2008).

7.   Costa, A., Hood, I.V. & Berger, J.M. Mechanisms for initiating cellular DNA replication. *Annu Rev Biochem* **82**, 25-54 (2013).

8.   Bell, S.P. & Kaguni, J.M. Helicase loading at chromosomal origins of replication. *Cold Spring Harb Perspect Biol* **5**(2013).

9.   Bell, S.D. & Botchan, M.R. The minichromosome maintenance replicative helicase. *Cold Spring Harb Perspect Biol* **5**, a012807 (2013).

10.   Costa, A. et al. The structural basis for MCM2-7 helicase activation by GINS and Cdc45. *Nat Struct Mol Biol* **18**, 471-7 (2011).

11.   Ilves, I., Petojevic, T., Pesavento, J.J. & Botchan, M.R. Activation of the MCM2-7 helicase by association with Cdc45 and GINS proteins. *Mol Cell* **37**, 247-58 (2010).

12.   Tanaka, S. et al. CDK-dependent phosphorylation of Sld2 and Sld3 initiates DNA replication in budding yeast. *Nature* **445**, 328-32 (2007).

13.   Yabuuchi, H. et al. Ordered assembly of Sld3, GINS and Cdc45 is distinctly regulated by DDK and CDK for activation of replication origins. *EMBO J* **25**, 4663-74 (2006).

14.   Zegerman, P. & Diffley, J.F. Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast. *Nature* **445**, 281-5 (2007).

15.   Johansson, E. & Dixon, N. Replicative DNA polymerases. *Cold Spring Harb Perspect Biol* **5**(2013).

16.   Ishikawa, F. & Naito, T. Why do we have linear chromosomes? A matter of Adam and Eve. *Mutat Res* **434**, 99-107 (1999).

References

17. Berg, J.M., Tymoczko, J.L., Gatto, G.J. & Stryer, L. *Biochemistry*, xlii, 1096, 45, 60, 45 pages (W.H. Freeman/Macmillan Learning, New York, 2019).

18. Brenner, S., Jacob, F. & Meselson, M. An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature* **190**, 576-581 (1961).

19. Smellie, R.M. Biochemistry of deoxyribonucleic acid and ribonucleic acid replication. *Br Med Bull* **21**, 195-202 (1965).

20. Nogales, E., Louder, R.K. & He, Y. Structural Insights into the Eukaryotic Transcription Initiation Machinery. *Annu Rev Biophys* **46**, 59-83 (2017).

21. Nelson, D.L., Cox, M.M. & Hoskins, A.A. *Lehninger principles of biochemistry*, 1 volume (various pagings) (Macmillan Learning, Austin, 2021).

22. Chatterjee, N. & Walker, G.C. Mechanisms of DNA damage, repair, and mutagenesis. *Environ Mol Mutagen* **58**, 235-263 (2017).

23. Clarke, T.L. & Mostoslavsky, R. DNA repair as a shared hallmark in cancer and ageing. *Mol Oncol* **16**, 3352-3379 (2022).

24. Huang, R. & Zhou, P.K. DNA damage repair: historical perspectives, mechanistic pathways and clinical translation for targeted cancer therapy. *Signal Transduct Target Ther* **6**, 254 (2021).

25. Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362**, 709-15 (1993).

26. Lindahl, T. & Barnes, D.E. Repair of endogenous DNA damage. *Cold Spring Harb Symp Quant Biol* **65**, 127-33 (2000).

27. Cheng, K.C., Cahill, D.S., Kasai, H., Nishimura, S. & Loeb, L.A. 8-Hydroxyguanine, an abundant form of oxidative DNA damage, causes G----T and A----C substitutions. *J Biol Chem* **267**, 166-72 (1992).

28. Tubbs, A. & Nussenzweig, A. Endogenous DNA Damage as a Source of Genomic Instability in Cancer. *Cell* **168**, 644-656 (2017).

29. Friedberg, E.C. A brief history of the DNA repair field. *Cell Res* **18**, 3-7 (2008).

30. Hwa Yun, B., Guo, J., Bellamri, M. & Turesky, R.J. DNA adducts: Formation, biological effects, and new biospecimens for mass spectrometric measurements in humans. *Mass Spectrom Rev* **39**, 55-82 (2020).

31. Kunkel, T.A. Balancing eukaryotic replication asymmetry with replication fidelity. *Curr Opin Chem Biol* **15**, 620-6 (2011).

32. Loeb, L.A. & Monnat, R.J., Jr. DNA polymerases and human disease. *Nat Rev Genet* **9**, 594-604 (2008).

33. Li, K., Luo, H., Huang, L., Luo, H. & Zhu, X. Microsatellite instability: a review of what the oncologist should know. *Cancer Cell Int* **20**, 16 (2020).

References

34. Yonekura, S., Nakamura, N., Yonei, S. & Zhang-Akiyama, Q.M. Generation, biological consequences and repair mechanisms of cytosine deamination in DNA. *J Radiat Res* **50**, 19-26 (2009).

35. De Bont, R. & van Larebeke, N. Endogenous DNA damage in humans: a review of quantitative data. *Mutagenesis* **19**, 169-85 (2004).

36. Swanton, C., McGranahan, N., Starrett, G.J. & Harris, R.S. APOBEC Enzymes: Mutagenic Fuel for Cancer Evolution and Heterogeneity. *Cancer Discov* **5**, 704-12 (2015).

37. Langenbucher, A. et al. An extended APOBEC3A mutation signature in cancer. *Nat Commun* **12**, 1602 (2021).

38. Petljak, M. et al. Mechanisms of APOBEC3 mutagenesis in human cancer cells. *Nature* **607**, 799-807 (2022).

39. Cooper, D.N. & Youssoufian, H. The CpG dinucleotide and human genetic disease. *Hum Genet* **78**, 151-5 (1988).

40. Pommier, Y., Nussenzweig, A., Takeda, S. & Austin, C. Human topoisomerases and their roles in genome stability and organization. *Nat Rev Mol Cell Biol* **23**, 407-427 (2022).

41. Barlow, J.H. et al. Identification of early replicating fragile sites that contribute to genome instability. *Cell* **152**, 620-32 (2013).

42. Crossley, M.P., Bocek, M. & Cimprich, K.A. R-Loops as Cellular Regulators and Genomic Threats. *Mol Cell* **73**, 398-411 (2019).

43. Stork, C.T. et al. Co-transcriptional R-loops are the main cause of estrogen-induced DNA damage. *Elife* **5**(2016).

44. Mir, A.A., Philippe, C. & Cristofari, G. euL1db: the European database of L1HS retrotransposon insertions in humans. *Nucleic Acids Res* **43**, D43-7 (2015).

45. Belgnaoui, S.M., Gosden, R.G., Semmes, O.J. & Haoudi, A. Human LINE-1 retrotransposon induces DNA damage and apoptosis in cancer cells. *Cancer Cell Int* **6**, 13 (2006).

46. Gasior, S.L., Wakeman, T.P., Xu, B. & Deininger, P.L. The human LINE-1 retrotransposon creates DNA double-strand breaks. *J Mol Biol* **357**, 1383-93 (2006).

47. Flasch, D.A. et al. Genome-wide de novo L1 Retrotransposition Connects Endonuclease Activity with Replication. *Cell* **177**, 837-851 e28 (2019).

48. Ardeljan, D. et al. Cell fitness screens reveal a conflict between LINE-1 retrotransposition and DNA replication. *Nat Struct Mol Biol* **27**, 168-178 (2020).

49. Ballmaier, D. & Epe, B. Oxidative DNA damage induced by potassium bromate under cell-free conditions and in mammalian cells. *Carcinogenesis* **16**, 335-42 (1995).

References

50. Singer, B. & Kusmierek, J.T. Chemical mutagenesis. *Annu Rev Biochem* **51**, 655-93 (1982).

51. Singer, B. O-alkyl pyrimidines in mutagenesis and carcinogenesis: occurrence and significance. *Cancer Res* **46**, 4879-85 (1986).

52. Wyatt, M.D. & Pittman, D.L. Methylating agents and DNA repair responses: Methylated bases and sources of strand breaks. *Chem Res Toxicol* **19**, 1580-94 (2006).

53. Jalali, R. et al. Encouraging experience of concomitant Temozolomide with radiotherapy followed by adjuvant Temozolomide in newly diagnosed glioblastoma multiforme: single institution experience. *Br J Neurosurg* **21**, 583-7 (2007).

54. Lawley, P.D. Effects of some chemical mutagens and carcinogens on nucleic acids. *Prog Nucleic Acid Res Mol Biol* **5**, 89-131 (1966).

55. Tretyakova, N.Y., Groehler, A.t. & Ji, S. DNA-Protein Cross-Links: Formation, Structural Identities, and Biological Outcomes. *Acc Chem Res* **48**, 1631-44 (2015).

56. Henner, W.D., Grunberg, S.M. & Haseltine, W.A. Sites and structure of gamma radiation-induced DNA strand breaks. *J Biol Chem* **257**, 11750-4 (1982).

57. Henner, W.D., Rodriguez, L.O., Hecht, S.M. & Haseltine, W.A. gamma Ray induced deoxyribonucleic acid strand breaks. 3' Glycolate termini. *J Biol Chem* **258**, 711-3 (1983).

58. Iliakis, G. The role of DNA double strand breaks in ionizing radiation-induced killing of eukaryotic cells. *Bioessays* **13**, 641-8 (1991).

59. Vignard, J., Mirey, G. & Salles, B. Ionizing-radiation induced DNA double-strand breaks: a direct and indirect lighting up. *Radiother Oncol* **108**, 362-9 (2013).

60. Davies, R.J. Royal Irish Academy Medal Lecture. Ultraviolet radiation damage in DNA. *Biochem Soc Trans* **23**, 407-18 (1995).

61. Rastogi, R.P., Richa, Kumar, A., Tyagi, M.B. & Sinha, R.P. Molecular mechanisms of ultraviolet radiation-induced DNA damage and repair. *J Nucleic Acids* **2010**, 592980 (2010).

62. Mitchell, D.L. & Nairn, R.S. The biology of the (6-4) photoproduct. *Photochem Photobiol* **49**, 805-19 (1989).

63. Weickert, P. & Stingele, J. DNA-Protein Crosslinks and Their Resolution. *Annu Rev Biochem* **91**, 157-181 (2022).

64. Goldknopf, I.L., French, M.F., Musso, R. & Busch, H. Presence of protein A24 in rat liver nucleosomes. *Proc Natl Acad Sci U S A* **74**, 5492-5 (1977).

65. Hershko, A. & Ciechanover, A. The ubiquitin system. *Annu Rev Biochem* **67**, 425-79 (1998).

References

66.   Pickart, C.M. Mechanisms underlying ubiquitination. *Annu Rev Biochem* **70**, 503-33 (2001).

67.   Kulathu, Y. & Komander, D. Atypical ubiquitylation - the unexplored world of polyubiquitin beyond Lys48 and Lys63 linkages. *Nat Rev Mol Cell Biol* **13**, 508-23 (2012).

68.   Thrower, J.S., Hoffman, L., Rechsteiner, M. & Pickart, C.M. Recognition of the polyubiquitin proteolytic signal. *EMBO J* **19**, 94-102 (2000).

69.   Chen, Z.J. & Sun, L.J. Nonproteolytic functions of ubiquitin in cell signaling. *Mol Cell* **33**, 275-86 (2009).

70.   Schulman, B.A. & Harper, J.W. Ubiquitin-like protein activation by E1 enzymes: the apex for downstream signalling pathways. *Nat Rev Mol Cell Biol* **10**, 319-31 (2009).

71.   Ye, Y. & Rape, M. Building ubiquitin chains: E2 enzymes at work. *Nat Rev Mol Cell Biol* **10**, 755-64 (2009).

72.   Rotin, D. & Kumar, S. Physiological functions of the HECT family of ubiquitin ligases. *Nat Rev Mol Cell Biol* **10**, 398-409 (2009).

73.   Deshaies, R.J. & Joazeiro, C.A. RING domain E3 ubiquitin ligases. *Annu Rev Biochem* **78**, 399-434 (2009).

74.   Komander, D. & Rape, M. The ubiquitin code. *Annu Rev Biochem* **81**, 203-29 (2012).

75.   Biswas, M., Voltz, K., Smith, J.C. & Langowski, J. Role of histone tails in structural stability of the nucleosome. *PLoS Comput Biol* **7**, e1002279 (2011).

76.   Mantovani, F., Collavin, L. & Del Sal, G. Mutant p53 as a guardian of the cancer cell. *Cell Death Differ* **26**, 199-212 (2019).

77.   Momand, J., Zambetti, G.P., Olson, D.C., George, D. & Levine, A.J. The mdm-2 oncogene product forms a complex with the p53 protein and inhibits p53-mediated transactivation. *Cell* **69**, 1237-45 (1992).

78.   Doil, C. et al. RNF168 binds and amplifies ubiquitin conjugates on damaged chromosomes to allow accumulation of repair proteins. *Cell* **136**, 435-46 (2009).

79.   Ramadan, K. & Meerang, M. Degradation-linked ubiquitin signal and proteasome are integral components of DNA double strand break repair: New perspectives for anti-cancer therapy. *FEBS Lett* **585**, 2868-75 (2011).

80.   Daigaku, Y., Davies, A.A. & Ulrich, H.D. Ubiquitin-dependent DNA damage bypass is separable from genome replication. *Nature* **465**, 951-5 (2010).

81.   Hoege, C., Pfander, B., Moldovan, G.L., Pyrowolakis, G. & Jentsch, S. RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. *Nature* **419**, 135-41 (2002).

References

82.   Kwasna, D. et al. Discovery and Characterization of ZUFSP/ZUP1, a Distinct Deubiquitinase Class Important for Genome Stability. *Mol Cell* **70**, 150-164 e6 (2018).

83.   Coleman, K.E. & Huang, T.T. In a Class of Its Own: A New Family of Deubiquitinases Promotes Genome Stability. *Mol Cell* **70**, 1-3 (2018).

84.   Reyes-Turcu, F.E., Ventii, K.H. & Wilkinson, K.D. Regulation and cellular roles of ubiquitin-specific deubiquitinating enzymes. *Annu Rev Biochem* **78**, 363-97 (2009).

85.   Kee, Y. & Huang, T.T. Role of Deubiquitinating Enzymes in DNA Repair. *Mol Cell Biol* **36**, 524-44 (2016).

86.   Sale, J.E., Lehmann, A.R. & Woodgate, R. Y-family DNA polymerases and their role in tolerance of cellular DNA damage. *Nat Rev Mol Cell Biol* **13**, 141-52 (2012).

87.   Huang, T.T. et al. Regulation of monoubiquitinated PCNA by DUB autocleavage. *Nat Cell Biol* **8**, 339-47 (2006).

88.   Wang, L. et al. SUMO2 is essential while SUMO3 is dispensable for mouse embryonic development. *EMBO Rep* **15**, 878-85 (2014).

89.   Liang, Y.C. et al. SUMO5, a Novel Poly-SUMO Isoform, Regulates PML Nuclear Bodies. *Sci Rep* **6**, 26509 (2016).

90.   Nacerddine, K. et al. The SUMO pathway is essential for nuclear integrity and chromosome segregation in mice. *Dev Cell* **9**, 769-79 (2005).

91.   Gareau, J.R. & Lima, C.D. The SUMO pathway: emerging mechanisms that shape specificity, conjugation and recognition. *Nat Rev Mol Cell Biol* **11**, 861-71 (2010).

92.   Cappadocia, L. & Lima, C.D. Ubiquitin-like Protein Conjugation: Structures, Chemistry, and Mechanism. *Chem Rev* **118**, 889-918 (2018).

93.   Kunz, K., Piller, T. & Muller, S. SUMO-specific proteases and isopeptidases of the SENP family at a glance. *J Cell Sci* **131**(2018).

94.   Hecker, C.M., Rabiller, M., Haglund, K., Bayer, P. & Dikic, I. Specification of SUMO1- and SUMO2-interacting motifs. *J Biol Chem* **281**, 16117-27 (2006).

95.   Schwertman, P., Bekker-Jensen, S. & Mailand, N. Regulation of DNA double-strand break repair by ubiquitin and ubiquitin-like modifiers. *Nat Rev Mol Cell Biol* **17**, 379-94 (2016).

96.   Galanty, Y. et al. Mammalian SUMO E3-ligases PIAS1 and PIAS4 promote responses to DNA double-strand breaks. *Nature* **462**, 935-9 (2009).

97.   Sun, Y. et al. A conserved SUMO pathway repairs topoisomerase DNA-protein cross-links by engaging ubiquitin-mediated proteasomal degradation. *Sci Adv* **6**(2020).

98.   Poulsen, S.L. et al. RNF111/Arkadia is a SUMO-targeted ubiquitin ligase that facilitates the DNA damage response. *J Cell Biol* **201**, 797-807 (2013).

References

99.   Borgermann, N. et al. SUMOylation promotes protective responses to DNA-protein crosslinks. *EMBO J* **38**(2019).

100.   Liu, J.C.Y. et al. Mechanism and function of DNA replication-independent DNA-protein crosslink repair via the SUMO-RNF4 pathway. *EMBO J* **40**, e107413 (2021).

101.   Rabut, G. & Peter, M. Function and regulation of protein neddylation. 'Protein modifications: beyond the usual suspects' review series. *EMBO Rep* **9**, 969-76 (2008).

102.   Petroski, M.D. & Deshaies, R.J. Function and regulation of cullin-RING ubiquitin ligases. *Nat Rev Mol Cell Biol* **6**, 9-20 (2005).

103.   Bosu, D.R. et al. C. elegans CAND-1 regulates cullin neddylation, cell proliferation and morphogenesis in specific tissues. *Dev Biol* **346**, 113-26 (2010).

104.   Brown, J.S. & Jackson, S.P. Ubiquitylation, neddylation and the DNA damage response. *Open Biol* **5**, 150018 (2015).

105.   Brown, J.S. et al. Neddylation promotes ubiquitylation and release of Ku from DNA-damage sites. *Cell Rep* **11**, 704-14 (2015).

106.   Baek, K., Scott, D.C. & Schulman, B.A. NEDD8 and ubiquitin ligation by cullin-RING E3 ligases. *Curr Opin Struct Biol* **67**, 101-109 (2021).

107.   Li, X., Wilmanns, M., Thornton, J. & Kohn, M. Elucidating human phosphatase-substrate networks. *Sci Signal* **6**, rs10 (2013).

108.   Schwartz, P.A. & Murray, B.W. Protein kinase biochemistry and drug discovery. *Bioorg Chem* **39**, 192-210 (2011).

109.   Nishi, H., Shaytan, A. & Panchenko, A.R. Physicochemical mechanisms of protein regulation by phosphorylation. *Front Genet* **5**, 270 (2014).

110.   Pavletich, N.P. Mechanisms of cyclin-dependent kinase regulation: structures of Cdks, their cyclin activators, and Cip and INK4 inhibitors. *J Mol Biol* **287**, 821-8 (1999).

111.   Lundberg, A.S. & Weinberg, R.A. Functional inactivation of the retinoblastoma protein requires sequential modification by at least two distinct cyclin-cdk complexes. *Mol Cell Biol* **18**, 753-61 (1998).

112.   Lukasik, P., Zaluski, M. & Gutowska, I. Cyclin-Dependent Kinases (CDK) and Their Role in Diseases Development-Review. *Int J Mol Sci* **22**(2021).

113.   Lanz, M.C., Dibitetto, D. & Smolka, M.B. DNA damage kinase signaling: checkpoint and repair at 30 years. *EMBO J* **38**, e101801 (2019).

114.   Smolka, M.B., Albuquerque, C.P., Chen, S.H. & Zhou, H. Proteome-wide identification of in vivo targets of DNA damage checkpoint kinases. *Proc Natl Acad Sci U S A* **104**, 10364-9 (2007).

References

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

116. Sancar, A. Structure and function of DNA photolyase and cryptochrome blue-light photoreceptors. *Chem Rev* **103**, 2203-37 (2003).

117. Sancar, A. Photolyase and cryptochrome blue-light photoreceptors. *Adv Protein Chem* **69**, 73-100 (2004).

118. Muller, M. & Carell, T. Structural biology of DNA photolyases and cryptochromes. *Curr Opin Struct Biol* **19**, 277-85 (2009).

119. Lindahl, T., Demple, B. & Robins, P. Suicide inactivation of the E. coli O6-methylguanine-DNA methyltransferase. *EMBO J* **1**, 1359-63 (1982).

120. Srivenugopal, K.S., Yuan, X.H., Friedman, H.S. & Ali-Osman, F. Ubiquitination-dependent proteolysis of O6-methylguanine-DNA methyltransferase in human and murine tumor cells following inactivation with O6-benzylguanine or 1,3-bis(2-chloroethyl)-1-nitrosourea. *Biochemistry* **35**, 1328-34 (1996).

121. Chowdhury, G., Cho, S.H., Pegg, A.E. & Guengerich, F.P. Detection and characterization of 1,2-dibromoethane-derived DNA crosslinks formed with O(6)-alkylguanine-DNA alkyltransferase. *Angew Chem Int Ed Engl* **52**, 12879-82 (2013).

122. Kurowski, M.A., Bhagwat, A.S., Papaj, G. & Bujnicki, J.M. Phylogenomic identification of five new human homologs of the DNA repair enzyme AlkB. *BMC Genomics* **4**, 48 (2003).

123. Sedgwick, B. Repairing DNA-methylation damage. *Nat Rev Mol Cell Biol* **5**, 148-57 (2004).

124. Yi, C., Yang, C.G. & He, C. A non-heme iron-mediated chemical demethylation in DNA and RNA. *Acc Chem Res* **42**, 519-29 (2009).

125. Nakamura, J. & Swenberg, J.A. Endogenous apurinic/apyrimidinic sites in genomic DNA of mammalian tissues. *Cancer Res* **59**, 2522-6 (1999).

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

127. Wang, K., Maayah, M., Sweasy, J.B. & Alnajjar, K.S. The role of cysteines in the structure and function of OGG1. *J Biol Chem* **296**, 100093 (2021).

128. Maher, R.L. & Bloom, L.B. Pre-steady-state kinetic characterization of the AP endonuclease activity of human AP endonuclease 1. *J Biol Chem* **282**, 30577-85 (2007).

129. Schermerhorn, K.M. & Delaney, S. Transient-state kinetics of apurinic/apyrimidinic (AP) endonuclease 1 acting on an authentic AP site and commonly used substrate analogs: the effect of diverse metal ions and base mismatches. *Biochemistry* **52**, 7669-77 (2013).

References

130. Freudenthal, B.D., Beard, W.A., Cuneo, M.J., Dyrkheeva, N.S. & Wilson, S.H. Capturing snapshots of APE1 processing DNA damage. *Nat Struct Mol Biol* **22**, 924-31 (2015).

131. Beard, W.A. & Wilson, S.H. Structure and mechanism of DNA polymerase beta. *Biochemistry* **53**, 2768-80 (2014).

132. Liu, Y. et al. DNA polymerase beta and flap endonuclease 1 enzymatic specificities sustain DNA synthesis for long patch base excision repair. *J Biol Chem* **280**, 3665-74 (2005).

133. Svilar, D., Goellner, E.M., Almeida, K.H. & Sobol, R.W. Base excision repair and lesion-dependent subpathways for repair of oxidative DNA damage. *Antioxid Redox Signal* **14**, 2491-507 (2011).

134. Beard, W.A., Horton, J.K., Prasad, R. & Wilson, S.H. Eukaryotic Base Excision Repair: New Approaches Shine Light on Mechanism. *Annu Rev Biochem* **88**, 137-162 (2019).

135. Hassa, P.O. & Hottiger, M.O. The diverse biological roles of mammalian PARPS, a small but powerful family of poly-ADP-ribose polymerases. *Front Biosci* **13**, 3046-82 (2008).

136. Cuneo, M.J. & London, R.E. Oxidation state of the XRCC1 N-terminal domain regulates DNA polymerase beta binding affinity. *Proc Natl Acad Sci U S A* **107**, 6805-10 (2010).

137. D'Amours, D., Desnoyers, S., D'Silva, I. & Poirier, G.G. Poly(ADP-ribosyl)ation reactions in the regulation of nuclear functions. *Biochem J* **342 (Pt 2)**, 249-68 (1999).

138. Su, S.S. & Modrich, P. Escherichia coli mutS-encoded protein binds to mismatched DNA base pairs. *Proc Natl Acad Sci U S A* **83**, 5057-61 (1986).

139. Modrich, P. Mechanisms in E. coli and Human Mismatch Repair (Nobel Lecture). *Angew Chem Int Ed Engl* **55**, 8490-501 (2016).

140. Liu, D., Keijzers, G. & Rasmussen, L.J. DNA mismatch repair and its many roles in eukaryotic cells. *Mutat Res Rev Mutat Res* **773**, 174-187 (2017).

141. Lipkin, S.M. et al. MLH3: a DNA mismatch repair gene associated with mammalian microsatellite instability. *Nat Genet* **24**, 27-35 (2000).

142. Clark, N., Wu, X. & Her, C. MutS Homologues hMSH4 and hMSH5: Genetic Variations, Functions, and Implications in Human Diseases. *Curr Genomics* **14**, 81-90 (2013).

143. Amaral-Silva, G.K. et al. Mismatch repair system proteins in oral benign and malignant lesions. *J Oral Pathol Med* **46**, 241-245 (2017).

144. Brown, M.W. et al. Dynamic DNA binding licenses a repair factor to bypass roadblocks in search of DNA lesions. *Nat Commun* **7**, 10607 (2016).

References

145. Fishel, R. Mismatch repair. *J Biol Chem* **290**, 26395-403 (2015).

146. Jeon, Y. et al. Dynamic control of strand excision during human DNA mismatch repair. *Proc Natl Acad Sci U S A* **113**, 3281-6 (2016).

147. Kusakabe, M. et al. Mechanism and regulation of DNA damage recognition in nucleotide excision repair. *Genes Environ* **41**, 2 (2019).

148. Gillet, L.C. & Scharer, O.D. Molecular mechanisms of mammalian global genome nucleotide excision repair. *Chem Rev* **106**, 253-76 (2006).

149. Rapin, I., Lindenbaum, Y., Dickson, D.W., Kraemer, K.H. & Robbins, J.H. Cockayne syndrome and xeroderma pigmentosum. *Neurology* **55**, 1442-9 (2000).

150. Mu, H., Geacintov, N.E., Broyde, S., Yeo, J.E. & Scharer, O.D. Molecular basis for damage recognition and verification by XPC-RAD23B and TFIIH in nucleotide excision repair. *DNA Repair (Amst)* **71**, 33-42 (2018).

151. Cheon, N.Y., Kim, H.S., Yeo, J.E., Scharer, O.D. & Lee, J.Y. Single-molecule visualization reveals the damage search mechanism for the human NER protein XPC-RAD23B. *Nucleic Acids Res* **47**, 8337-8347 (2019).

152. Paul, D. et al. Tethering-facilitated DNA 'opening' and complementary roles of beta-hairpin motifs in the Rad4/XPC DNA damage sensor protein. *Nucleic Acids Res* **48**, 12348-12364 (2020).

153. Buterin, T., Meyer, C., Giese, B. & Naegeli, H. DNA quality control by conformational readout on the undamaged strand of the double helix. *Chem Biol* **12**, 913-22 (2005).

154. Sugasawa, K. et al. UV-induced ubiquitylation of XPC protein mediated by UV-DDB-ubiquitin ligase complex. *Cell* **121**, 387-400 (2005).

155. Hanawalt, P.C. & Spivak, G. Transcription-coupled DNA repair: two decades of progress and surprises. *Nat Rev Mol Cell Biol* **9**, 958-70 (2008).

156. Nakazawa, Y. et al. Ubiquitination of DNA Damage-Stalled RNAPII Promotes Transcription-Coupled Repair. *Cell* **180**, 1228-1244 e24 (2020).

157. van der Weegen, Y. et al. The cooperative action of CSB, CSA, and UVSSA target TFIIH to DNA damage-stalled RNA polymerase II. *Nat Commun* **11**, 2104 (2020).

158. Krasikova, Y.S., Rechkunova, N.I., Maltseva, E.A., Petruseva, I.O. & Lavrik, O.I. Localization of xeroderma pigmentosum group A protein and replication protein A on damaged DNA in nucleotide excision repair. *Nucleic Acids Res* **38**, 8083-94 (2010).

159. Overmeer, R.M. et al. Replication protein A safeguards genome integrity by controlling NER incision events. *J Cell Biol* **192**, 401-15 (2011).

160. Kokic, G. et al. Structural basis of TFIIH activation for nucleotide excision repair. *Nat Commun* **10**, 2885 (2019).

161. Tsodikov, O.V. et al. Structural basis for the recruitment of ERCC1-XPF to nucleotide excision repair complexes by XPA. *EMBO J* **26**, 4768-76 (2007).

<div align="center">References</div>

162. Fagbemi, A.F., Orelli, B. & Scharer, O.D. Regulation of endonuclease activity in human nucleotide excision repair. *DNA Repair (Amst)* **10**, 722-9 (2011).

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

164. Krasikova, Y., Rechkunova, N. & Lavrik, O. Nucleotide Excision Repair: From Molecular Defects to Neurological Abnormalities. *Int J Mol Sci* **22**(2021).

165. Caldecott, K.W. DNA single-strand break repair and human genetic disease. *Trends Cell Biol* **32**, 733-745 (2022).

166. Azarm, K. & Smith, S. Nuclear PARPs and genome integrity. *Genes Dev* **34**, 285-301 (2020).

167. Langelier, M.F., Eisemann, T., Riccio, A.A. & Pascal, J.M. PARP family enzymes: regulation and catalysis of the poly(ADP-ribose) posttranslational modification. *Curr Opin Struct Biol* **53**, 187-198 (2018).

168. Gibbs-Seymour, I., Fontana, P., Rack, J.G.M. & Ahel, I. HPF1/C4orf27 Is a PARP-1-Interacting Protein that Regulates PARP-1 ADP-Ribosylation Activity. *Mol Cell* **62**, 432-442 (2016).

169. Bonfiglio, J.J. et al. Serine ADP-Ribosylation Depends on HPF1. *Mol Cell* **65**, 932-940 e6 (2017).

170. Groslambert, J., Prokhorova, E. & Ahel, I. ADP-ribosylation of DNA and RNA. *DNA Repair (Amst)* **105**, 103144 (2021).

171. Rack, J.G.M., Palazzo, L. & Ahel, I. (ADP-ribosyl)hydrolases: structure, function, and biology. *Genes Dev* **34**, 263-284 (2020).

172. Ray Chaudhuri, A. & Nussenzweig, A. The multifaceted roles of PARP1 in DNA repair and chromatin remodelling. *Nat Rev Mol Cell Biol* **18**, 610-621 (2017).

173. Kim, M.Y., Mauro, S., Gevry, N., Lis, J.T. & Kraus, W.L. NAD+-dependent modulation of chromatin structure and transcription by nucleosome binding properties of PARP-1. *Cell* **119**, 803-14 (2004).

174. Tulin, A. & Spradling, A. Chromatin loosening by poly(ADP)-ribose polymerase (PARP) at Drosophila puff loci. *Science* **299**, 560-2 (2003).

175. Ahel, D. et al. Poly(ADP-ribose)-dependent regulation of DNA repair by the chromatin remodeling enzyme ALC1. *Science* **325**, 1240-3 (2009).

176. Gao, Y. et al. SSRP1 Cooperates with PARP and XRCC1 to Facilitate Single-Strand DNA Break Repair by Chromatin Priming. *Cancer Res* **77**, 2674-2685 (2017).

177. Caldecott, K.W. XRCC1 protein; Form and function. *DNA Repair (Amst)* **81**, 102664 (2019).

References

178.   Bennett, R.A., Wilson, D.M., 3rd, Wong, D. & Demple, B. Interaction of human apurinic endonuclease and DNA polymerase beta in the base excision repair pathway. *Proc Natl Acad Sci U S A* **94**, 7166-9 (1997).

179.   Mol, C.D., Izumi, T., Mitra, S. & Tainer, J.A. DNA-bound structures and mutants reveal abasic DNA binding by APE1 and DNA repair coordination [corrected]. *Nature* **403**, 451-6 (2000).

180.   Kubota, Y. et al. Reconstitution of DNA base excision-repair with purified human proteins: interaction between DNA polymerase beta and the XRCC1 protein. *EMBO J* **15**, 6662-70 (1996).

181.   Demin, A.A. et al. XRCC1 prevents toxic PARP1 trapping during DNA base excision repair. *Mol Cell* **81**, 3018-3030 e5 (2021).

182.   Vilenchik, M.M. & Knudson, A.G. Endogenous DNA double-strand breaks: production, fidelity of repair, and induction of cancer. *Proc Natl Acad Sci U S A* **100**, 12871-6 (2003).

183.   Gillyard, T. & Davis, J. DNA double-strand break repair in cancer: A path to achieving precision medicine. *Int Rev Cell Mol Biol* **364**, 111-137 (2021).

184.   Britton, S., Coates, J. & Jackson, S.P. A new method for high-resolution imaging of Ku foci to decipher mechanisms of DNA double-strand break repair. *J Cell Biol* **202**, 579-95 (2013).

185.   Gottlieb, T.M. & Jackson, S.P. The DNA-dependent protein kinase: requirement for DNA ends and association with Ku antigen. *Cell* **72**, 131-42 (1993).

186.   Ochi, T. et al. DNA repair. PAXX, a paralog of XRCC4 and XLF, interacts with Ku to promote DNA double-strand break repair. *Science* **347**, 185-188 (2015).

187.   Lee, J.H. & Paull, T.T. Direct activation of the ATM protein kinase by the Mre11/Rad50/Nbs1 complex. *Science* **304**, 93-6 (2004).

188.   Scully, R. & Xie, A. Double strand break repair functions of histone H2AX. *Mutat Res* **750**, 5-14 (2013).

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

190.   Blackford, A.N. & Jackson, S.P. ATM, ATR, and DNA-PK: The Trinity at the Heart of the DNA Damage Response. *Mol Cell* **66**, 801-817 (2017).

191.   Ahnesorg, P., Smith, P. & Jackson, S.P. XLF interacts with the XRCC4-DNA ligase IV complex to promote DNA nonhomologous end-joining. *Cell* **124**, 301-13 (2006).

192.   Tsai, C.J., Kim, S.A. & Chu, G. Cernunnos/XLF promotes the ligation of mismatched and noncohesive DNA ends. *Proc Natl Acad Sci U S A* **104**, 7851-6 (2007).

193.   Stinson, B.M. & Loparo, J.J. Repair of DNA Double-Strand Breaks by the Nonhomologous End Joining Pathway. *Annu Rev Biochem* **90**, 137-164 (2021).

References

194. Scully, R., Panday, A., Elango, R. & Willis, N.A. DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nat Rev Mol Cell Biol* **20**, 698-714 (2019).

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

196. Huertas, P. & Jackson, S.P. Human CtIP mediates cell cycle control of DNA end resection and double strand break repair. *J Biol Chem* **284**, 9558-65 (2009).

197. Syed, A. & Tainer, J.A. The MRE11-RAD50-NBS1 Complex Conducts the Orchestration of Damage Signaling and Outcomes to Stress in DNA Replication and Repair. *Annu Rev Biochem* **87**, 263-294 (2018).

198. Stewart, G.S. et al. The RIDDLE syndrome protein mediates a ubiquitin-dependent signaling cascade at sites of DNA damage. *Cell* **136**, 420-34 (2009).

199. Bunting, S.F. et al. 53BP1 inhibits homologous recombination in Brca1-deficient cells by blocking resection of DNA breaks. *Cell* **141**, 243-54 (2010).

200. Lee, D.H. et al. Dephosphorylation enables the recruitment of 53BP1 to double-strand DNA breaks. *Mol Cell* **54**, 512-25 (2014).

201. Orthwein, A. et al. Mitosis inhibits DNA double-strand break repair to guard against telomere fusions. *Science* **344**, 189-93 (2014).

202. Yu, A.M. & McVey, M. Synthesis-dependent microhomology-mediated end joining accounts for multiple types of repair junctions. *Nucleic Acids Res* **38**, 5706-17 (2010).

203. Mateos-Gomez, P.A. et al. The helicase domain of Poltheta counteracts RPA to promote alt-NHEJ. *Nat Struct Mol Biol* **24**, 1116-1123 (2017).

204. Sallmyr, A. & Tomkinson, A.E. Repair of DNA double-strand breaks by mammalian alternative end-joining pathways. *J Biol Chem* **293**, 10536-10546 (2018).

205. Fleury, H. et al. The APE2 nuclease is essential for DNA double-strand break repair by microhomology-mediated end joining. *Mol Cell* **83**, 1429-1445 e8 (2023).

206. Audebert, M., Salles, B. & Calsou, P. Involvement of poly(ADP-ribose) polymerase-1 and XRCC1/DNA ligase III in an alternative route for DNA double-strand breaks rejoining. *J Biol Chem* **279**, 55117-26 (2004).

207. Blasiak, J. Single-Strand Annealing in Cancer. *Int J Mol Sci* **22**(2021).

208. 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).

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

References

210. Sugiyama, T., New, J.H. & Kowalczykowski, S.C. DNA annealing by RAD52 protein is stimulated by specific interaction with the complex of replication protein A and single-stranded DNA. *Proc Natl Acad Sci U S A* **95**, 6049-54 (1998).

211. McVey, M., Khodaverdian, V.Y., Meyer, D., Cerqueira, P.G. & Heyer, W.D. Eukaryotic DNA Polymerases in Homologous Recombination. *Annu Rev Genet* **50**, 393-421 (2016).

212. Hicks, W.M., Kim, M. & Haber, J.E. Increased mutagenesis and unique mutation signature associated with mitotic gene conversion. *Science* **329**, 82-5 (2010).

213. Johnson, R.D. & Jasin, M. Sister chromatid gene conversion is a prominent double-strand break repair pathway in mammalian cells. *EMBO J* **19**, 3398-407 (2000).

214. Moynahan, M.E. & Jasin, M. Loss of heterozygosity induced by a chromosomal double-strand break. *Proc Natl Acad Sci U S A* **94**, 8988-93 (1997).

215. Elbakry, A. & Lobrich, M. Homologous Recombination Subpathways: A Tangle to Resolve. *Front Genet* **12**, 723847 (2021).

216. Zhang, F. et al. PALB2 links BRCA1 and BRCA2 in the DNA-damage response. *Curr Biol* **19**, 524-9 (2009).

217. Zhao, W. et al. BRCA1-BARD1 promotes RAD51-mediated homologous DNA pairing. *Nature* **550**, 360-365 (2017).

218. Paques, F. & Haber, J.E. Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae. *Microbiol Mol Biol Rev* **63**, 349-404 (1999).

219. Bizard, A.H. & Hickson, I.D. The dissolution of double Holliday junctions. *Cold Spring Harb Perspect Biol* **6**, a016477 (2014).

220. Nguyen, M.O., Jalan, M., Morrow, C.A., Osman, F. & Whitby, M.C. Recombination occurs within minutes of replication blockage by RTS1 producing restarted forks that are prone to collapse. *Elife* **4**, e04539 (2015).

221. Anand, R.P., Lovett, S.T. & Haber, J.E. Break-induced DNA replication. *Cold Spring Harb Perspect Biol* **5**, a010397 (2013).

222. Elango, R. et al. Break-induced replication promotes formation of lethal joint molecules dissolved by Srs2. *Nat Commun* **8**, 1790 (2017).

223. Dilley, R.L. et al. Break-induced telomere synthesis underlies alternative telomere maintenance. *Nature* **539**, 54-58 (2016).

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

225. Nepal, M., Che, R., Zhang, J., Ma, C. & Fei, P. Fanconi Anemia Signaling and Cancer. *Trends Cancer* **3**, 840-856 (2017).

226. Bogliolo, M. & Surralles, J. Fanconi anemia: a model disease for studies on human genetics and advanced therapeutics. *Curr Opin Genet Dev* **33**, 32-40 (2015).

227. Walden, H. & Deans, A.J. The Fanconi anemia DNA repair pathway: structural and functional insights into a complex disorder. *Annu Rev Biophys* **43**, 257-78 (2014).

228. Rohleder, F. et al. FANCM interacts with PCNA to promote replication traverse of DNA interstrand crosslinks. *Nucleic Acids Res* **44**, 3219-32 (2016).

229. 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).

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

231. Rickman, K.A. et al. Deficiency of UBE2T, the E2 Ubiquitin Ligase Necessary for FANCD2 and FANCI Ubiquitination, Causes FA-T Subtype of Fanconi Anemia. *Cell Rep* **12**, 35-41 (2015).

232. Meetei, A.R. et al. A novel ubiquitin ligase is deficient in Fanconi anemia. *Nat Genet* **35**, 165-70 (2003).

233. Liang, C.C. et al. UHRF1 is a sensor for DNA interstrand crosslinks and recruits FANCD2 to initiate the Fanconi anemia pathway. *Cell Rep* **10**, 1947-56 (2015).

234. Knipscheer, P. et al. The Fanconi anemia pathway promotes replication-dependent DNA interstrand cross-link repair. *Science* **326**, 1698-701 (2009).

235. El-Khamisy, S.F. To live or to die: a matter of processing damaged DNA termini in neurons. *EMBO Mol Med* **3**, 78-88 (2011).

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

237. Stoepker, C. et al. SLX4, a coordinator of structure-specific endonucleases, is mutated in a new Fanconi anemia subtype. *Nat Genet* **43**, 138-41 (2011).

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

239. Cohn, M.A. et al. A UAF1-containing multisubunit protein complex regulates the Fanconi anemia pathway. *Mol Cell* **28**, 786-97 (2007).

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

241. 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).

242. Petermann, E., Lan, L. & Zou, L. Sources, resolution and physiological relevance of R-loops and RNA-DNA hybrids. *Nat Rev Mol Cell Biol* **23**, 521-540 (2022).

References

243. Hamperl, S., Bocek, M.J., Saldivar, J.C., Swigut, T. & Cimprich, K.A. Transcription-Replication Conflict Orientation Modulates R-Loop Levels and Activates Distinct DNA Damage Responses. *Cell* **170**, 774-786 e19 (2017).

244. Matos, D.A. et al. ATR Protects the Genome against R Loops through a MUS81-Triggered Feedback Loop. *Mol Cell* **77**, 514-527 e4 (2020).

245. Bubeck, D. et al. PCNA directs type 2 RNase H activity on DNA replication and repair substrates. *Nucleic Acids Res* **39**, 3652-66 (2011).

246. Lockhart, A. et al. RNase H1 and H2 Are Differentially Regulated to Process RNA-DNA Hybrids. *Cell Rep* **29**, 2890-2900 e5 (2019).

247. Skourti-Stathaki, K., Proudfoot, N.J. & Gromak, N. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol Cell* **42**, 794-805 (2011).

248. Sollier, J. et al. Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Mol Cell* **56**, 777-85 (2014).

249. Song, C., Hotz-Wagenblatt, A., Voit, R. & Grummt, I. SIRT7 and the DEAD-box helicase DDX21 cooperate to resolve genomic R loops and safeguard genome stability. *Genes Dev* **31**, 1370-1381 (2017).

250. Cristini, A., Groh, M., Kristiansen, M.S. & Gromak, N. RNA/DNA Hybrid Interactome Identifies DXH9 as a Molecular Player in Transcriptional Termination and R-Loop-Associated DNA Damage. *Cell Rep* **23**, 1891-1905 (2018).

251. Ulrich, H.D. Conservation of DNA damage tolerance pathways from yeast to humans. *Biochem Soc Trans* **35**, 1334-7 (2007).

252. Kunz, B.A., Anderson, H.J., Osmond, M.J. & Vonarx, E.J. Components of nucleotide excision repair and DNA damage tolerance in Arabidopsis thaliana. *Environ Mol Mutagen* **45**, 115-27 (2005).

253. Guilliam, T.A. & Yeeles, J.T.P. Reconstitution of translesion synthesis reveals a mechanism of eukaryotic DNA replication restart. *Nat Struct Mol Biol* **27**, 450-460 (2020).

254. Gan, G.N., Wittschieben, J.P., Wittschieben, B.O. & Wood, R.D. DNA polymerase zeta (pol zeta) in higher eukaryotes. *Cell Res* **18**, 174-83 (2008).

255. McCulloch, S.D. & Kunkel, T.A. The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Res* **18**, 148-61 (2008).

256. McDonald, J.P., Levine, A.S. & Woodgate, R. The Saccharomyces cerevisiae RAD30 gene, a homologue of Escherichia coli dinB and umuC, is DNA damage inducible and functions in a novel error-free postreplication repair mechanism. *Genetics* **147**, 1557-68 (1997).

References

257. Johnson, R.E., Prakash, S. & Prakash, L. Efficient bypass of a thymine-thymine dimer by yeast DNA polymerase, Poleta. *Science* **283**, 1001-4 (1999).

258. Biertumpfel, C. et al. Structure and mechanism of human DNA polymerase eta. *Nature* **465**, 1044-8 (2010).

259. McDonald, J.P. et al. Novel human and mouse homologs of Saccharomyces cerevisiae DNA polymerase eta. *Genomics* **60**, 20-30 (1999).

260. Tissier, A. et al. Misinsertion and bypass of thymine-thymine dimers by human DNA polymerase iota. *EMBO J* **19**, 5259-66 (2000).

261. Bebenek, K. et al. 5'-Deoxyribose phosphate lyase activity of human DNA polymerase iota in vitro. *Science* **291**, 2156-9 (2001).

262. Petta, T.B. et al. Human DNA polymerase iota protects cells against oxidative stress. *EMBO J* **27**, 2883-95 (2008).

263. Ohmori, H. et al. The Y-family of DNA polymerases. *Mol Cell* **8**, 7-8 (2001).

264. Ohashi, E. et al. Fidelity and processivity of DNA synthesis by DNA polymerase kappa, the product of the human DINB1 gene. *J Biol Chem* **275**, 39678-84 (2000).

265. Yuan, B., Cao, H., Jiang, Y., Hong, H. & Wang, Y. Efficient and accurate bypass of N2-(1-carboxyethyl)-2'-deoxyguanosine by DinB DNA polymerase in vitro and in vivo. *Proc Natl Acad Sci U S A* **105**, 8679-84 (2008).

266. Washington, M.T. et al. Efficient and error-free replication past a minor-groove DNA adduct by the sequential action of human DNA polymerases iota and kappa. *Mol Cell Biol* **24**, 5687-93 (2004).

267. Shachar, S. et al. Two-polymerase mechanisms dictate error-free and error-prone translesion DNA synthesis in mammals. *EMBO J* **28**, 383-93 (2009).

268. Ogi, T. & Lehmann, A.R. The Y-family DNA polymerase kappa (pol kappa) functions in mammalian nucleotide-excision repair. *Nat Cell Biol* **8**, 640-2 (2006).

269. Zhang, Y. et al. Response of human REV1 to different DNA damage: preferential dCMP insertion opposite the lesion. *Nucleic Acids Res* **30**, 1630-8 (2002).

270. Nelson, J.R., Lawrence, C.W. & Hinkle, D.C. Deoxycytidyl transferase activity of yeast REV1 protein. *Nature* **382**, 729-31 (1996).

271. Nair, D.T., Johnson, R.E., Prakash, L., Prakash, S. & Aggarwal, A.K. Protein-template-directed synthesis across an acrolein-derived DNA adduct by yeast Rev1 DNA polymerase. *Structure* **16**, 239-45 (2008).

272. Nair, D.T., Johnson, R.E., Prakash, L., Prakash, S. & Aggarwal, A.K. Rev1 employs a novel mechanism of DNA synthesis using a protein template. *Science* **309**, 2219-22 (2005).

References

273. Otsuka, C., Loakes, D. & Negishi, K. The role of deoxycytidyl transferase activity of yeast Rev1 protein in the bypass of abasic sites. *Nucleic Acids Res Suppl*, 87-8 (2002).

274. Sale, J.E. Translesion DNA synthesis and mutagenesis in eukaryotes. *Cold Spring Harb Perspect Biol* **5**, a012708 (2013).

275. Nelson, J.R., Lawrence, C.W. & Hinkle, D.C. Thymine-thymine dimer bypass by yeast DNA polymerase zeta. *Science* **272**, 1646-9 (1996).

276. Acharya, N. et al. Complex formation of yeast Rev1 and Rev7 proteins: a novel role for the polymerase-associated domain. *Mol Cell Biol* **25**, 9734-40 (2005).

277. Malik, R. et al. Structure and mechanism of B-family DNA polymerase zeta specialized for translesion DNA synthesis. *Nat Struct Mol Biol* **27**, 913-924 (2020).

278. Acharya, N., Johnson, R.E., Pages, V., Prakash, L. & Prakash, S. Yeast Rev1 protein promotes complex formation of DNA polymerase zeta with Pol32 subunit of DNA polymerase delta. *Proc Natl Acad Sci U S A* **106**, 9631-6 (2009).

279. Lawrence, C.W. Cellular roles of DNA polymerase zeta and Rev1 protein. *DNA Repair (Amst)* **1**, 425-35 (2002).

280. Miyase, S. et al. Differential regulation of Rad18 through Rad6-dependent mono- and polyubiquitination. *J Biol Chem* **280**, 515-24 (2005).

281. Watanabe, K. et al. Rad18 guides poleta to replication stalling sites through physical interaction and PCNA monoubiquitination. *EMBO J* **23**, 3886-96 (2004).

282. Davies, A.A., Huttner, D., Daigaku, Y., Chen, S. & Ulrich, H.D. Activation of ubiquitin-dependent DNA damage bypass is mediated by replication protein a. *Mol Cell* **29**, 625-36 (2008).

283. Leung, W., Baxley, R.M., Moldovan, G.L. & Bielinsky, A.K. Mechanisms of DNA Damage Tolerance: Post-Translational Regulation of PCNA. *Genes (Basel)* **10**(2018).

284. Boehm, E.M. & Washington, M.T. R.I.P. to the PIP: PCNA-binding motif no longer considered specific: PIP motifs and other related sequences are not distinct entities and can bind multiple proteins involved in genome maintenance. *Bioessays* **38**, 1117-1122 (2016).

285. Boehm, E.M., Spies, M. & Washington, M.T. PCNA tool belts and polymerase bridges form during translesion synthesis. *Nucleic Acids Res* **44**, 8250-60 (2016).

286. Murakumo, Y. et al. Interactions in the error-prone postreplication repair proteins hREV1, hREV3, and hREV7. *J Biol Chem* **276**, 35644-51 (2001).

287. Guilliam, T.A. Mechanisms for Maintaining Eukaryotic Replisome Progression in the Presence of DNA Damage. *Front Mol Biosci* **8**, 712971 (2021).

288. Taylor, M.R.G. & Yeeles, J.T.P. Dynamics of Replication Fork Progression Following Helicase-Polymerase Uncoupling in Eukaryotes. *J Mol Biol* **431**, 2040-2049 (2019).

289. Bianchi, J. et al. PrimPol bypasses UV photoproducts during eukaryotic chromosomal DNA replication. *Mol Cell* **52**, 566-73 (2013).

290. Bailey, L.J., Bianchi, J., Hegarat, N., Hochegger, H. & Doherty, A.J. PrimPol-deficient cells exhibit a pronounced G2 checkpoint response following UV damage. *Cell Cycle* **15**, 908-18 (2016).

291. Bailey, L.J., Bianchi, J. & Doherty, A.J. PrimPol is required for the maintenance of efficient nuclear and mitochondrial DNA replication in human cells. *Nucleic Acids Res* **47**, 4026-4038 (2019).

292. Kobayashi, K. et al. Repriming by PrimPol is critical for DNA replication restart downstream of lesions and chain-terminating nucleosides. *Cell Cycle* **15**, 1997-2008 (2016).

293. Quinet, A. et al. PRIMPOL-Mediated Adaptive Response Suppresses Replication Fork Reversal in BRCA-Deficient Cells. *Mol Cell* **77**, 461-474 e9 (2020).

294. Maga, G. et al. DNA polymerase delta-interacting protein 2 is a processivity factor for DNA polymerase lambda during 8-oxo-7,8-dihydroguanine bypass. *Proc Natl Acad Sci U S A* **110**, 18850-5 (2013).

295. Guilliam, T.A., Bailey, L.J., Brissett, N.C. & Doherty, A.J. PolDIP2 interacts with human PrimPol and enhances its DNA polymerase activities. *Nucleic Acids Res* **44**, 3317-29 (2016).

296. Neelsen, K.J. & Lopes, M. Replication fork reversal in eukaryotes: from dead end to dynamic response. *Nat Rev Mol Cell Biol* **16**, 207-20 (2015).

297. Berti, M. & Vindigni, A. Replication stress: getting back on track. *Nat Struct Mol Biol* **23**, 103-9 (2016).

298. Unk, I. et al. Human SHPRH is a ubiquitin ligase for Mms2-Ubc13-dependent polyubiquitylation of proliferating cell nuclear antigen. *Proc Natl Acad Sci U S A* **103**, 18107-12 (2006).

299. Blastyak, A., Hajdu, I., Unk, I. & Haracska, L. Role of double-stranded DNA translocase activity of human HLTF in replication of damaged DNA. *Mol Cell Biol* **30**, 684-93 (2010).

300. Ciccia, A. et al. Polyubiquitinated PCNA recruits the ZRANB3 translocase to maintain genomic integrity after replication stress. *Mol Cell* **47**, 396-409 (2012).

301. Betous, R. et al. Substrate-selective repair and restart of replication forks by DNA translocases. *Cell Rep* **3**, 1958-69 (2013).

References

302. Kolinjivadi, A.M. et al. Smarcal1-Mediated Fork Reversal Triggers Mre11-Dependent Degradation of Nascent DNA in the Absence of Brca2 and Stable Rad51 Nucleofilaments. *Mol Cell* **67**, 867-881 e7 (2017).

303. Zellweger, R. et al. Rad51-mediated replication fork reversal is a global response to genotoxic treatments in human cells. *J Cell Biol* **208**, 563-79 (2015).

304. Schlacher, K. et al. Double-strand break repair-independent role for BRCA2 in blocking stalled replication fork degradation by MRE11. *Cell* **145**, 529-42 (2011).

305. Berti, M., Cortez, D. & Lopes, M. The plasticity of DNA replication forks in response to clinically relevant genotoxic stress. *Nat Rev Mol Cell Biol* **21**, 633-651 (2020).

306. 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).

307. Thangavel, S. et al. DNA2 drives processing and restart of reversed replication forks in human cells. *J Cell Biol* **208**, 545-62 (2015).

308. 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).

309. Mizumoto, A. et al. Molecular Mechanisms of Acetaldehyde-Mediated Carcinogenesis in Squamous Epithelium. *Int J Mol Sci* **18**(2017).

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

311. Nakamura, J. & Nakamura, M. DNA-protein crosslink formation by endogenous aldehydes and AP sites. *DNA Repair (Amst)* **88**, 102806 (2020).

312. Guerin, M.R., Higgins, C.E. & Griest, W.H. The analysis of the particulate and vapour phases of tobacco smoke. *IARC Sci Publ*, 115-39 (1987).

313. Zheng, Y. et al. Mitochondrial One-Carbon Pathway Supports Cytosolic Folate Integrity in Cancer Cells. *Cell* **175**, 1546-1560 e17 (2018).

314. Nakamura, J. et al. Evidence that endogenous formaldehyde produces immunogenic and atherogenic adduct epitopes. *Sci Rep* **7**, 10787 (2017).

315. Yu, P.H., Wright, S., Fan, E.H., Lun, Z.R. & Gubisne-Harberle, D. Physiological and pathological implications of semicarbazide-sensitive amine oxidase. *Biochim Biophys Acta* **1647**, 193-9 (2003).

316. Gupta, S. et al. Histone methylation regulates memory formation. *J Neurosci* **30**, 3589-99 (2010).

317. Okano, M., Bell, D.W., Haber, D.A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247-57 (1999).

318. Shen, X. et al. A Surge of DNA Damage Links Transcriptional Reprogramming and Hematopoietic Deficit in Fanconi Anemia. *Mol Cell* **80**, 1013-1024 e6 (2020).

References

319. Li, T. et al. Formaldehyde and De/Methylation in Age-Related Cognitive Impairment. *Genes (Basel)* **12**(2021).

320. MacAllister, S.L., Choi, J., Dedina, L. & O'Brien, P.J. Metabolic mechanisms of methanol/formaldehyde in isolated rat hepatocytes: carbonyl-metabolizing enzymes versus oxidative stress. *Chem Biol Interact* **191**, 308-14 (2011).

321. 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).

322. Anzenbacher, P. & Anzenbacherova, E. Cytochromes P450 and metabolism of xenobiotics. *Cell Mol Life Sci* **58**, 737-47 (2001).

323. Stingele, J. et al. Mechanism and Regulation of DNA-Protein Crosslink Repair by the DNA-Dependent Metalloprotease SPRTN. *Mol Cell* **64**, 688-703 (2016).

324. Mulderrig, L. et al. Aldehyde-driven transcriptional stress triggers an anorexic DNA damage response. *Nature* **600**, 158-163 (2021).

325. McGhee, J.D. & von Hippel, P.H. Formaldehyde as a probe of DNA structure. r. Mechanism of the initial reaction of Formaldehyde with DNA. *Biochemistry* **16**, 3276-93 (1977).

326. Grafstrom, R.C., Fornace, A.J., Jr., Autrup, H., Lechner, J.F. & Harris, C.C. Formaldehyde damage to DNA and inhibition of DNA repair in human bronchial cells. *Science* **220**, 216-8 (1983).

327. Johnson, D.S., Mortazavi, A., Myers, R.M. & Wold, B. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**, 1497-502 (2007).

328. Castello, A. et al. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**, 1393-406 (2012).

329. Castello, A. et al. System-wide identification of RNA-binding proteins by interactome capture. *Nat Protoc* **8**, 491-500 (2013).

330. Quievryn, G. & Zhitkovich, A. Loss of DNA-protein crosslinks from formaldehyde-exposed cells occurs through spontaneous hydrolysis and an active repair process linked to proteosome function. *Carcinogenesis* **21**, 1573-80 (2000).

331. Chan, W. et al. Quantification of a Novel DNA-Protein Cross-Link Product Formed by Reacting Apurinic/Apyrimidinic Sites in DNA with Cysteine Residues in Protein by Liquid Chromatography-Tandem Mass Spectrometry Coupled with the Stable Isotope-Dilution Method. *Anal Chem* **91**, 4987-4994 (2019).

332. Admiraal, S.J. & O'Brien, P.J. Reactivity and Cross-Linking of 5'-Terminal Abasic Sites within DNA. *Chem Res Toxicol* **30**, 1317-1326 (2017).

333. Jung, Y. & Lippard, S.J. Direct cellular responses to platinum-induced DNA damage. *Chem Rev* **107**, 1387-407 (2007).

References

334. Huang, H., Woo, J., Alley, S.C. & Hopkins, P.B. DNA-DNA interstrand cross-linking by cis-diamminedichloroplatinum(II): N7(dG)-to-N7(dG) cross-linking at 5'-d(GC) in synthetic oligonucleotides. *Bioorg Med Chem* **3**, 659-69 (1995).

335. Hemminki, K. & Kallama, S. Reactions of nitrogen mustards with DNA. *IARC Sci Publ*, 55-70 (1986).

336. Loeber, R.L. et al. Proteomic analysis of DNA-protein cross-linking by antitumor nitrogen mustards. *Chem Res Toxicol* **22**, 1151-62 (2009).

337. Michaelson-Richie, E.D. et al. Mechlorethamine-induced DNA-protein cross-linking in human fibrosarcoma (HT1080) cells. *J Proteome Res* **10**, 2785-96 (2011).

338. Rajski, S.R. & Williams, R.M. DNA Cross-Linking Agents as Antitumor Drugs. *Chem Rev* **98**, 2723-2796 (1998).

339. Peak, M.J. & Peak, J.G. DNA-to-protein crosslinks and backbone breaks caused by far- and near-ultraviolet, and visible light radiations in mammalian cells. *Basic Life Sci* **38**, 193-202 (1986).

340. Strniste, G.F., Hardin, J.M. & Rall, S.C. Isolation and characterization of stable protein-DNA adducts induced in chromatin by ultraviolet light. *Adv Exp Med Biol* **86A**, 571-94 (1977).

341. Nakano, T. et al. Induction of DNA-protein cross-links by ionizing radiation and their elimination from the genome. *Mutat Res* **771**, 45-50 (2015).

342. Shoulkamy, M.I. et al. Detection of DNA-protein crosslinks (DPCs) by novel direct fluorescence labeling methods: distinct stabilities of aldehyde and radiation-induced DPCs. *Nucleic Acids Res* **40**, e143 (2012).

343. Meyn, R.E., vanAnkeren, S.C. & Jenkins, W.T. The induction of DNA-protein crosslinks in hypoxic cells and their possible contribution to cell lethality. *Radiat Res* **109**, 419-29 (1987).

344. Cadet, J., Anselmino, C., Douki, T. & Voituriez, L. Photochemistry of nucleic acids in cells. *J Photochem Photobiol B* **15**, 277-98 (1992).

345. Cadet, J. & Douki, T. Formation of UV-induced DNA damage contributing to skin cancer development. *Photochem Photobiol Sci* **17**, 1816-1841 (2018).

346. Pommier, Y., Sun, Y., Huang, S.N. & Nitiss, J.L. Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. *Nat Rev Mol Cell Biol* **17**, 703-721 (2016).

347. Liu, L.F. & Wang, J.C. Supercoiling of the DNA template during transcription. *Proc Natl Acad Sci U S A* **84**, 7024-7 (1987).

348. Baranello, L. et al. RNA Polymerase II Regulates Topoisomerase 1 Activity to Favor Efficient Transcription. *Cell* **165**, 357-71 (2016).

References

349. Le, T.T. et al. Synergistic Coordination of Chromatin Torsional Mechanics and Topoisomerase Activity. *Cell* **179**, 619-631 e15 (2019).

350. Keszthelyi, A., Minchell, N.E. & Baxter, J. The Causes and Consequences of Topological Stress during DNA Replication. *Genes (Basel)* **7**(2016).

351. Liu, L.F. et al. Mechanism of action of camptothecin. *Ann N Y Acad Sci* **922**, 1-10 (2000).

352. Sirikantaramas, S., Yamazaki, M. & Saito, K. Mutations in topoisomerase I as a self-resistance mechanism coevolved with the production of the anticancer alkaloid camptothecin in plants. *Proc Natl Acad Sci U S A* **105**, 6782-6 (2008).

353. Fujimori, A., Harker, W.G., Kohlhagen, G., Hoki, Y. & Pommier, Y. Mutation at the catalytic site of topoisomerase I in CEM/C2, a human leukemia cell line resistant to camptothecin. *Cancer Res* **55**, 1339-46 (1995).

354. Wang, X., Zhuang, Y., Wang, Y., Jiang, M. & Yao, L. The recent developments of camptothecin and its derivatives as potential anti-tumor agents. *Eur J Med Chem* **260**, 115710 (2023).

355. Williams, J.S., Lujan, S.A. & Kunkel, T.A. Processing ribonucleotides incorporated during eukaryotic DNA replication. *Nat Rev Mol Cell Biol* **17**, 350-63 (2016).

356. Reijns, M.A. et al. Enzymatic removal of ribonucleotides from DNA is essential for mammalian genome integrity and development. *Cell* **149**, 1008-22 (2012).

357. Pourquier, P. et al. Effects of uracil incorporation, DNA mismatches, and abasic sites on cleavage and religation activities of mammalian topoisomerase I. *J Biol Chem* **272**, 7792-6 (1997).

358. Saha, L.K. et al. Topoisomerase I-driven repair of UV-induced damage in NER-deficient cells. *Proc Natl Acad Sci U S A* **117**, 14412-14420 (2020).

359. Nitiss, J.L. DNA topoisomerase II and its growing repertoire of biological functions. *Nat Rev Cancer* **9**, 327-37 (2009).

360. Lee, J.H. & Berger, J.M. Cell Cycle-Dependent Control and Roles of DNA Topoisomerase II. *Genes (Basel)* **10**(2019).

361. McClendon, A.K., Rodriguez, A.C. & Osheroff, N. Human topoisomerase IIalpha rapidly relaxes positively supercoiled DNA: implications for enzyme action ahead of replication forks. *J Biol Chem* **280**, 39337-45 (2005).

362. Valdes, A., Segura, J., Dyson, S., Martinez-Garcia, B. & Roca, J. DNA knots occur in intracellular chromatin. *Nucleic Acids Res* **46**, 650-660 (2018).

363. Griffith, J.D. & Nash, H.A. Genetic rearrangement of DNA induces knots with a unique topology: implications for the mechanism of synapsis and crossing-over. *Proc Natl Acad Sci U S A* **82**, 3124-8 (1985).

References

364. Vann, K.R., Oviatt, A.A. & Osheroff, N. Topoisomerase II Poisons: Converting Essential Enzymes into Molecular Scissors. *Biochemistry* **60**, 1630-1641 (2021).

365. Cline, S.D., Jones, W.R., Stone, M.P. & Osheroff, N. DNA abasic lesions in a different light: solution structure of an endogenous topoisomerase II poison. *Biochemistry* **38**, 15500-7 (1999).

366. Sabourin, M. & Osheroff, N. Sensitivity of human type II topoisomerases to DNA damage: stimulation of enzyme-mediated DNA cleavage by abasic, oxidized and alkylated lesions. *Nucleic Acids Res* **28**, 1947-54 (2000).

367. Nitiss, J.L. Targeting DNA topoisomerase II in cancer chemotherapy. *Nat Rev Cancer* **9**, 338-50 (2009).

368. Haffner, M.C., De Marzo, A.M., Meeker, A.K., Nelson, W.G. & Yegnasubramanian, S. Transcription-induced DNA double strand breaks: both oncogenic force and potential therapeutic target? *Clin Cancer Res* **17**, 3858-64 (2011).

369. Haffner, M.C. et al. Androgen-induced TOP2B-mediated double-strand breaks and prostate cancer gene rearrangements. *Nat Genet* **42**, 668-75 (2010).

370. Morimoto, S. et al. Type II DNA Topoisomerases Cause Spontaneous Double-Strand Breaks in Genomic DNA. *Genes (Basel)* **10**(2019).

371. Shi, Q. et al. Estradiol increases risk of topoisomerase IIbeta-mediated DNA strand breaks to initiate Xp11.2 translocation renal cell carcinoma. *Cell Commun Signal* **19**, 114 (2021).

372. Collin, F., Karkare, S. & Maxwell, A. Exploiting bacterial DNA gyrase as a drug target: current state and perspectives. *Appl Microbiol Biotechnol* **92**, 479-97 (2011).

373. Bernard, P. & Couturier, M. Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. *J Mol Biol* **226**, 735-45 (1992).

374. Rhaese, H.J. & Freese, E. Chemical analysis of DNA alterations. I. Base liberation and backbone breakage of DNA and oligodeoxyadenylic acid induced by hydrogen peroxide and hydroxylamine. *Biochim Biophys Acta* **155**, 476-90 (1968).

375. Urata, H. & Akagi, M. Photo-induced formation of the 2-deoxyribonolactone-containing nucleotide for d(ApCpA); effects of neighboring bases and modification of deoxycytidine. *Nucleic Acids Res* **19**, 1773-8 (1991).

376. Chan, W. et al. Quantification of the 2-deoxyribonolactone and nucleoside 5'-aldehyde products of 2-deoxyribose oxidation in DNA and cells by isotope-dilution gas chromatography mass spectrometry: differential effects of gamma-radiation and Fe2+-EDTA. *J Am Chem Soc* **132**, 6145-53 (2010).

377. Kappen, L.S. & Goldberg, I.H. Identification of 2-deoxyribonolactone at the site of neocarzinostatin-induced cytosine release in the sequence d(AGC). *Biochemistry* **28**, 1027-32 (1989).

378. Daniels, J.S., Gates, K.S., Tronche, C. & Greenberg, M.M. Direct evidence for bimodal DNA damage induced by tirapazamine. *Chem Res Toxicol* **11**, 1254-7 (1998).

379. DeMott, M.S. et al. Covalent trapping of human DNA polymerase beta by the oxidative DNA lesion 2-deoxyribonolactone. *J Biol Chem* **277**, 7637-40 (2002).

380. Quinones, J.L. et al. Enzyme mechanism-based, oxidative DNA-protein cross-links formed with DNA polymerase beta in vivo. *Proc Natl Acad Sci U S A* **112**, 8602-7 (2015).

381. Globisch, D. et al. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS One* **5**, e15367 (2010).

382. Jeltsch, A. & Jurkowska, R.Z. New concepts in DNA methylation. *Trends Biochem Sci* **39**, 310-8 (2014).

383. Chatterjee, A. & Eccles, M.R. DNA methylation and epigenomics: new technologies and emerging concepts. *Genome Biol* **16**, 103 (2015).

384. Jurkowska, R.Z., Jurkowski, T.P. & Jeltsch, A. Structure and function of mammalian DNA methyltransferases. *Chembiochem* **12**, 206-22 (2011).

385. Jeltsch, A. & Jurkowska, R.Z. Allosteric control of mammalian DNA methyltransferases - a new regulatory paradigm. *Nucleic Acids Res* **44**, 8556-8575 (2016).

386. Jeltsch, A. Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases. *Chembiochem* **3**, 274-93 (2002).

387. Bashtrykov, P. et al. Specificity of Dnmt1 for methylation of hemimethylated CpG sites resides in its catalytic domain. *Chem Biol* **19**, 572-8 (2012).

388. Qin, W. et al. DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination. *Cell Res* **25**, 911-29 (2015).

389. Nishiyama, A. et al. Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication. *Nature* **502**, 249-53 (2013).

390. Chuang, L.S. et al. Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. *Science* **277**, 1996-2000 (1997).

391. Gowher, H. & Jeltsch, A. Mammalian DNA methyltransferases: new discoveries and open questions. *Biochem Soc Trans* **46**, 1191-1202 (2018).

392. Christman, J.K. 5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. *Oncogene* **21**, 5483-95 (2002).

393. Fenaux, P. et al. Efficacy of azacitidine compared with that of conventional care regimens in the treatment of higher-risk myelodysplastic syndromes: a randomised, open-label, phase III study. *Lancet Oncol* **10**, 223-32 (2009).

References

394. Maslov, A.Y. et al. 5-aza-2'-deoxycytidine-induced genome rearrangements are mediated by DNMT1. *Oncogene* **31**, 5172-9 (2012).

395. Daskalakis, M. et al. Demethylation of a hypermethylated P15/INK4B gene in patients with myelodysplastic syndrome by 5-Aza-2'-deoxycytidine (decitabine) treatment. *Blood* **100**, 2957-64 (2002).

396. Rose, M., Burgess, J.T., O'Byrne, K., Richard, D.J. & Bolderson, E. PARP Inhibitors: Clinical Relevance, Mechanisms of Action and Tumor Resistance. *Front Cell Dev Biol* **8**, 564601 (2020).

397. Hodgson, D.R. et al. Candidate biomarkers of PARP inhibitor sensitivity in ovarian cancer beyond the BRCA genes. *Br J Cancer* **119**, 1401-1409 (2018).

398. Murai, J. et al. Trapping of PARP1 and PARP2 by Clinical PARP Inhibitors. *Cancer Res* **72**, 5588-99 (2012).

399. Krastev, D.B. et al. The ubiquitin-dependent ATPase p97 removes cytotoxic trapped PARP1 from chromatin. *Nat Cell Biol* **24**, 62-73 (2022).

400. Prasad, R. et al. Suicidal cross-linking of PARP-1 to AP site intermediates in cells undergoing base excision repair. *Nucleic Acids Res* **42**, 6337-51 (2014).

401. Prasad, R., Horton, J.K. & Wilson, S.H. Requirements for PARP-1 covalent crosslinking to DNA (PARP-1 DPC). *DNA Repair (Amst)* **90**, 102850 (2020).

402. Mohni, K.N. et al. HMCES Maintains Genome Integrity by Shielding Abasic Sites in Single-Strand DNA. *Cell* **176**, 144-153 e13 (2019).

403. Thompson, P.S., Amidon, K.M., Mohni, K.N., Cortez, D. & Eichman, B.F. Protection of abasic sites during DNA replication by a stable thiazolidine protein-DNA cross-link. *Nat Struct Mol Biol* **26**, 613-618 (2019).

404. Halabelian, L. et al. Structural basis of HMCES interactions with abasic DNA and multivalent substrate recognition. *Nat Struct Mol Biol* **26**, 607-612 (2019).

405. Yaneva, D. et al. The FANCJ helicase unfolds DNA-protein crosslinks to promote their repair. *Mol Cell* **83**, 43-56 e10 (2023).

406. Donsbach, M. et al. A non-proteolytic release mechanism for HMCES-DNA-protein crosslinks. *EMBO J*, e113360 (2023).

407. Paulin, K.A., Cortez, D. & Eichman, B.F. The SOS response-associated peptidase (SRAP) domain of YedK catalyzes ring opening of abasic sites and reversal of its DNA-protein cross-link. *J Biol Chem* **298**, 102307 (2022).

408. Sang, P.B., Srinath, T., Patil, A.G., Woo, E.J. & Varshney, U. A unique uracil-DNA binding protein of the uracil DNA glycosylase superfamily. *Nucleic Acids Res* **43**, 8452-63 (2015).

409. Ahn, W.C. et al. Covalent binding of uracil DNA glycosylase UdgX to abasic DNA upon uracil excision. *Nat Chem Biol* **15**, 607-614 (2019).

410. Tu, J., Chen, R., Yang, Y., Cao, W. & Xie, W. Suicide inactivation of the uracil DNA glycosylase UdgX by covalent complex formation. *Nat Chem Biol* **15**, 615-622 (2019).

411. Aroli, S., Woo, E.J., Gopal, B. & Varshney, U. Mutational and structural analyses of UdgX: insights into the active site pocket architecture and its evolution. *Nucleic Acids Res* **51**, 6554-6565 (2023).

412. Datta, M. et al. Development of mCherry tagged UdgX as a highly sensitive molecular probe for specific detection of uracils in DNA. *Biochem Biophys Res Commun* **518**, 38-43 (2019).

413. Ji, T.T. et al. Enzymatic Cleavage-Mediated Extension Stalling Enables Accurate Recognition and Quantification of Locus-Specific Uracil Modification in DNA. *Anal Chem* **95**, 8384-8392 (2023).

414. Jiang, L. et al. UdgX-Mediated Uracil Sequencing at Single-Nucleotide Resolution. *J Am Chem Soc* **144**, 1323-1331 (2022).

415. Stewart, J.A., Schauer, G. & Bhagwat, A.S. Visualization of uracils created by APOBEC3A using UdgX shows colocalization with RPA at stalled replication forks. *Nucleic Acids Res* **48**, e118 (2020).

416. Sun, H., Treco, D., Schultes, N.P. & Szostak, J.W. Double-strand breaks at an initiation site for meiotic gene conversion. *Nature* **338**, 87-90 (1989).

417. Buhler, C., Lebbink, J.H., Bocs, C., Ladenstein, R. & Forterre, P. DNA topoisomerase VI generates ATP-dependent double-strand breaks with two-nucleotide overhangs. *J Biol Chem* **276**, 37215-22 (2001).

418. Corbett, K.D., Benedetti, P. & Berger, J.M. Holoenzyme assembly and ATP-mediated conformational dynamics of topoisomerase VI. *Nat Struct Mol Biol* **14**, 611-9 (2007).

419. Liu, J., Wu, T.C. & Lichten, M. The location and structure of double-strand DNA breaks induced during yeast meiosis: evidence for a covalently linked DNA-protein intermediate. *EMBO J* **14**, 4599-608 (1995).

420. Zhang, B., Tang, Z., Li, L. & Lu, L.Y. NBS1 is required for SPO11-linked DNA double-strand break repair in male meiosis. *Cell Death Differ* **27**, 2176-2190 (2020).

421. Alani, E., Padmore, R. & Kleckner, N. Analysis of wild-type and rad50 mutants of yeast suggests an intimate relationship between meiotic chromosome synapsis and recombination. *Cell* **61**, 419-36 (1990).

422. Ivanov, E.L., Korolev, V.G. & Fabre, F. XRS2, a DNA repair gene of Saccharomyces cerevisiae, is needed for meiotic recombination. *Genetics* **132**, 651-64 (1992).

423. Nairz, K. & Klein, F. mre11S--a yeast mutation that blocks double-strand-break processing and permits nonhomologous synapsis in meiosis. *Genes Dev* **11**, 2272-90 (1997).

References

424. Yadav, V.K. & Claeys Bouuaert, C. Mechanism and Control of Meiotic DNA Double-Strand Break Formation in S. cerevisiae. *Front Cell Dev Biol* **9**, 642737 (2021).

425. Frappier, L. Ebna1. *Curr Top Microbiol Immunol* **391**, 3-34 (2015).

426. Lindner, S.E. & Sugden, B. The plasmid replicon of Epstein-Barr virus: mechanistic insights into efficient, licensed, extrachromosomal replication in human cells. *Plasmid* **58**, 1-12 (2007).

427. Castillo, F., Benmohamed, A. & Szatmari, G. Xer Site Specific Recombination: Double and Single Recombinase Systems. *Front Microbiol* **8**, 453 (2017).

428. Crozat, E., Fournes, F., Cornet, F., Hallet, B. & Rousseau, P. Resolution of Multimeric Forms of Circular Plasmids and Chromosomes. *Microbiol Spectr* **2**(2014).

429. Dheekollu, J. et al. Cell-cycle-dependent EBNA1-DNA crosslinking promotes replication termination at oriP and viral episome maintenance. *Cell* **184**, 643-654 e13 (2021).

430. Pronk, R. & van der Vliet, P.C. The adenovirus terminal protein influences binding of replication proteins and changes the origin structure. *Nucleic Acids Res* **21**, 2293-300 (1993).

431. Robinson, A.J., Younghusband, H.B. & Bellett, A.J. A circula DNA-protein complex from adenoviruses. *Virology* **56**, 54-69 (1973).

432. Zhao, L.J. & Padmanabhan, R. Nuclear transport of adenovirus DNA polymerase is facilitated by interaction with preterminal protein. *Cell* **55**, 1005-15 (1988).

433. Benevento, M. et al. Adenovirus composition, proteolysis, and disassembly studied by in-depth qualitative and quantitative proteomics. *J Biol Chem* **289**, 11421-11430 (2014).

434. King, A.J. & van der Vliet, P.C. A precursor terminal protein-trinucleotide intermediate during initiation of adenovirus DNA replication: regeneration of molecular ends in vitro by a jumping back mechanism. *EMBO J* **13**, 5786-92 (1994).

435. Salas, M. Protein-priming of DNA replication. *Annu Rev Biochem* **60**, 39-71 (1991).

436. Hoeben, R.C. & Uil, T.G. Adenovirus DNA replication. *Cold Spring Harb Perspect Biol* **5**, a013003 (2013).

437. Munoz-Espin, D., Holguera, I., Ballesteros-Plaza, D., Carballido-Lopez, R. & Salas, M. Viral terminal protein directs early organization of phage DNA replication at the bacterial nucleoid. *Proc Natl Acad Sci U S A* **107**, 16548-53 (2010).

438. Flett, F.J. et al. Structural basis for DNA 3'-end processing by human tyrosyl-DNA phosphodiesterase 1. *Nat Commun* **9**, 24 (2018).

439. Interthal, H. & Champoux, J.J. Effects of DNA and protein size on substrate cleavage by human tyrosyl-DNA phosphodiesterase 1. *Biochem J* **436**, 559-66 (2011).

References

440. Lin, C.P., Ban, Y., Lyu, Y.L., Desai, S.D. & Liu, L.F. A ubiquitin-proteasome pathway for the repair of topoisomerase I-DNA covalent complexes. *J Biol Chem* **283**, 21074-83 (2008).

441. Pouliot, J.J., Yao, K.C., Robertson, C.A. & Nash, H.A. Yeast gene for a Tyr-DNA phosphodiesterase that repairs topoisomerase I complexes. *Science* **286**, 552-5 (1999).

442. Liu, C., Pouliot, J.J. & Nash, H.A. Repair of topoisomerase I covalent complexes in the absence of the tyrosyl-DNA phosphodiesterase Tdp1. *Proc Natl Acad Sci U S A* **99**, 14970-5 (2002).

443. Takashima, H. et al. Mutation of TDP1, encoding a topoisomerase I-dependent DNA damage repair enzyme, in spinocerebellar ataxia with axonal neuropathy. *Nat Genet* **32**, 267-72 (2002).

444. El-Khamisy, S.F. et al. Defective DNA single-strand break repair in spinocerebellar ataxia with axonal neuropathy-1. *Nature* **434**, 108-13 (2005).

445. Hirano, R. et al. Spinocerebellar ataxia with axonal neuropathy: consequence of a Tdp1 recessive neomorphic mutation? *EMBO J* **26**, 4732-43 (2007).

446. Das, B.B. et al. Optimal function of the DNA repair enzyme TDP1 requires its phosphorylation by ATM and/or DNA-PK. *EMBO J* **28**, 3667-80 (2009).

447. Das, B.B. et al. PARP1-TDP1 coupling for the repair of topoisomerase I-induced DNA damage. *Nucleic Acids Res* **42**, 4435-49 (2014).

448. Das, S.K. et al. Poly(ADP-ribose) polymers regulate DNA topoisomerase I (Top1) nuclear dynamics and camptothecin sensitivity in living cells. *Nucleic Acids Res* **44**, 8363-75 (2016).

449. Cortes Ledesma, F., El Khamisy, S.F., Zuma, M.C., Osborn, K. & Caldecott, K.W. A human 5'-tyrosyl DNA phosphodiesterase that repairs topoisomerase-mediated DNA damage. *Nature* **461**, 674-8 (2009).

450. Gao, R. et al. Proteolytic degradation of topoisomerase II (Top2) enables the processing of Top2.DNA and Top2.RNA covalent complexes by tyrosyl-DNA-phosphodiesterase 2 (TDP2). *J Biol Chem* **289**, 17960-9 (2014).

451. Schellenberg, M.J. et al. ZATT (ZNF451)-mediated resolution of topoisomerase 2 DNA-protein cross-links. *Science* **357**, 1412-1416 (2017).

452. Rao, T. et al. Novel TDP2-ubiquitin interactions and their importance for the repair of topoisomerase II-mediated DNA damage. *Nucleic Acids Res* **44**, 10201-10215 (2016).

453. Zeng, Z., Cortes-Ledesma, F., El Khamisy, S.F. & Caldecott, K.W. TDP2/TTRAP is the major 5'-tyrosyl DNA phosphodiesterase activity in vertebrate cells and is critical

for cellular resistance to topoisomerase II-induced DNA damage. *J Biol Chem* **286**, 403-9 (2011).

454. Gomez-Herreros, F. et al. TDP2 protects transcription from abortive topoisomerase activity and is required for normal neural function. *Nat Genet* **46**, 516-21 (2014).

455. Virgen-Slane, R. et al. An RNA virus hijacks an incognito function of a DNA repair enzyme. *Proc Natl Acad Sci U S A* **109**, 14634-9 (2012).

456. Neale, M.J., Pan, J. & Keeney, S. Endonucleolytic processing of covalent protein-linked DNA double-strand breaks. *Nature* **436**, 1053-7 (2005).

457. Hopfner, K.P. et al. Structural biochemistry and interaction architecture of the DNA double-strand break repair Mre11 nuclease and Rad50-ATPase. *Cell* **105**, 473-85 (2001).

458. Connelly, J.C., de Leau, E.S. & Leach, D.R. Nucleolytic processing of a protein-bound DNA end by the E. coli SbcCD (MR) complex. *DNA Repair (Amst)* **2**, 795-807 (2003).

459. Kashammer, L. et al. Mechanism of DNA End Sensing and Processing by the Mre11-Rad50 Complex. *Mol Cell* **76**, 382-394 e6 (2019).

460. Nakano, T. et al. Nucleotide excision repair and homologous recombination systems commit differentially to the repair of DNA-protein crosslinks. *Mol Cell* **28**, 147-58 (2007).

461. Baker, D.J. et al. Nucleotide excision repair eliminates unique DNA-protein cross-links from mammalian cells. *J Biol Chem* **282**, 22592-604 (2007).

462. Stingele, J., Schwarz, M.S., Bloemeke, N., Wolf, P.G. & Jentsch, S. A DNA-dependent protease involved in DNA-protein crosslink repair. *Cell* **158**, 327-338 (2014).

463. Biggins, S., Bhalla, N., Chang, A., Smith, D.L. & Murray, A.W. Genes involved in sister chromatid separation and segregation in the budding yeast Saccharomyces cerevisiae. *Genetics* **159**, 453-70 (2001).

464. Reinking, H.K., Hofmann, K. & Stingele, J. Function and evolution of the DNA-protein crosslink proteases Wss1 and SPRTN. *DNA Repair (Amst)* **88**, 102822 (2020).

465. Noireterre, A., Serbyn, N., Bagdiul, I. & Stutz, F. Ubx5-Cdc48 assists the protease Wss1 at DNA-protein crosslink sites in yeast. *EMBO J* **42**, e113609 (2023).

466. Davis, E.J. et al. DVC1 (C1orf124) recruits the p97 protein segregase to sites of DNA damage. *Nat Struct Mol Biol* **19**, 1093-100 (2012).

467. Centore, R.C., Yazinski, S.A., Tse, A. & Zou, L. Spartan/C1orf124, a reader of PCNA ubiquitylation and a regulator of UV-induced DNA damage response. *Mol Cell* **46**, 625-35 (2012).

References

468.   Vaz, B. et al. Metalloprotease SPRTN/DVC1 Orchestrates Replication-Coupled DNA-Protein Crosslink Repair. *Mol Cell* **64**, 704-719 (2016).

469.   Reinking, H.K. et al. DNA Structure-Specific Cleavage of DNA-Protein Crosslinks by the SPRTN Protease. *Mol Cell* **80**, 102-113 e6 (2020).

470.   Li, F., Raczynska, J.E., Chen, Z. & Yu, H. Structural Insight into DNA-Dependent Activation of Human Metalloprotease Spartan. *Cell Rep* **26**, 3336-3346 e4 (2019).

471.   Zhao, S. et al. A ubiquitin switch controls autocatalytic inactivation of the DNA-protein crosslink repair protease SPRTN. *Nucleic Acids Res* **49**, 902-915 (2021).

472.   Huang, J. et al. Tandem Deubiquitination and Acetylation of SPRTN Promotes DNA-Protein Crosslink Repair and Protects against Aging. *Mol Cell* **79**, 824-835 e5 (2020).

473.   Perry, M. et al. USP11 mediates repair of DNA-protein cross-links by deubiquitinating SPRTN metalloprotease. *J Biol Chem* **296**, 100396 (2021).

474.   Delabaere, L. et al. The Spartan ortholog maternal haploid is required for paternal chromosome integrity in the Drosophila zygote. *Curr Biol* **24**, 2281-7 (2014).

475.   Lessel, D. et al. Mutations in SPRTN cause early onset hepatocellular carcinoma, genomic instability and progeroid features. *Nat Genet* **46**, 1239-44 (2014).

476.   Maskey, R.S. et al. Spartan deficiency causes genomic instability and progeroid phenotypes. *Nat Commun* **5**, 5744 (2014).

477.   Ruijs, M.W. et al. Atypical progeroid syndrome: an unknown helicase gene defect? *Am J Med Genet A* **116A**, 295-9 (2003).

478.   Makaros, Y. et al. Ubiquitin-independent proteasomal degradation driven by C-degron pathways. *Mol Cell* **83**, 1921-1935 e7 (2023).

479.   Bard, J.A.M. et al. Structure and Function of the 26S Proteasome. *Annu Rev Biochem* **87**, 697-724 (2018).

480.   Finley, D. Recognition and processing of ubiquitin-protein conjugates by the proteasome. *Annu Rev Biochem* **78**, 477-513 (2009).

481.   Larsen, N.B. et al. Replication-Coupled DNA-Protein Crosslink Repair by SPRTN and the Proteasome in Xenopus Egg Extracts. *Mol Cell* **73**, 574-588 e7 (2019).

482.   Serbyn, N. et al. The Aspartic Protease Ddi1 Contributes to DNA-Protein Crosslink Repair in Yeast. *Mol Cell* **77**, 1066-1079 e9 (2020).

483.   Kottemann, M.C., Conti, B.A., Lach, F.P. & Smogorzewska, A. Removal of RTF2 from Stalled Replisomes Promotes Maintenance of Genome Integrity. *Mol Cell* **69**, 24-35 e5 (2018).

484.   Radhakrishnan, S.K. et al. Transcription factor Nrf1 mediates the proteasome recovery pathway after proteasome inhibition in mammalian cells. *Mol Cell* **38**, 17-28 (2010).

References

485. Dirac-Svejstrup, A.B. et al. DDI2 Is a Ubiquitin-Directed Endoprotease Responsible for Cleavage of Transcription Factor NRF1. *Mol Cell* **79**, 332-341 e7 (2020).

486. Bhargava, V. et al. GCNA Preserves Genome Integrity and Fertility Across Species. *Dev Cell* **52**, 38-52 e10 (2020).

487. Dokshin, G.A. et al. GCNA Interacts with Spartan and Topoisomerase II to Regulate Genome Stability. *Dev Cell* **52**, 53-68 e6 (2020).

488. Carmell, M.A. et al. A widely employed germ cell marker is an ancient disordered protein with reproductive functions in diverse eukaryotes. *Elife* **5**(2016).

489. Hoffmann, S. et al. FAM111 protease activity undermines cellular fitness and is amplified by gain-of-function mutations in human disease. *EMBO Rep* **21**, e50662 (2020).

490. Kojima, Y. et al. FAM111A protects replication forks from protein obstacles via its trypsin-like domain. *Nat Commun* **11**, 1318 (2020).

491. Welter, A.L. & Machida, Y.J. Functions and evolution of FAM111 serine proteases. *Front Mol Biosci* **9**, 1081166 (2022).

492. Kenny, F.M. & Linarelli, L. Dwarfism and cortical thickening of tubular bones. Transient hypocalcemia in a mother and son. *Am J Dis Child* **111**, 201-7 (1966).

493. Caffey, J. Congenital stenosis of medullary spaces in tubular bones and calvaria in two proportionate dwarfs--mother and son; coupled with transitory hypocalcemic tetany. *Am J Roentgenol Radium Ther Nucl Med* **100**, 1-11 (1967).

494. Li, W., Hu, S., Han, Z. & Jiang, X. YY1-Induced Transcriptional Activation of FAM111B Contributes to the Malignancy of Breast Cancer. *Clin Breast Cancer* **22**, e417-e425 (2022).

495. Ip, W.H. et al. Differential Regulation of Cellular FAM111B by Human Adenovirus C Type 5 E1 Oncogenes. *Viruses* **13**(2021).

496. Sparks, J.L. et al. The CMG Helicase Bypasses DNA-Protein Cross-Links to Facilitate Their Repair. *Cell* **176**, 167-181 e21 (2019).

497. Gallina, I. et al. The ubiquitin ligase RFWD3 is required for translesion DNA synthesis. *Mol Cell* **81**, 442-458 e9 (2021).

498. 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* **78**, 192 (2020).

499. Duxin, J.P., Dewar, J.M., Yardimci, H. & Walter, J.C. Repair of a DNA-protein crosslink by replication-coupled proteolysis. *Cell* **159**, 346-57 (2014).

500. Fielden, J. et al. TEX264 coordinates p97- and SPRTN-mediated resolution of topoisomerase 1-DNA adducts. *Nat Commun* **11**, 1274 (2020).

References

501. Moilanen, A.M. et al. Identification of a novel RING finger protein as a coregulator in steroid receptor-mediated gene transcription. *Mol Cell Biol* **18**, 5128-39 (1998).

502. Sun, H., Leverson, J.D. & Hunter, T. Conserved function of RNF4 family proteins in eukaryotes: targeting a ubiquitin ligase to SUMOylated proteins. *EMBO J* **26**, 4102-12 (2007).

503. Tatham, M.H. et al. RNF4 is a poly-SUMO-specific E3 ubiquitin ligase required for arsenic-induced PML degradation. *Nat Cell Biol* **10**, 538-46 (2008).

504. Chang, Y.C., Oram, M.K. & Bielinsky, A.K. SUMO-Targeted Ubiquitin Ligases and Their Functions in Maintaining Genome Stability. *Int J Mol Sci* **22**(2021).

505. Kumar, R., Gonzalez-Prieto, R., Xiao, Z., Verlaan-de Vries, M. & Vertegaal, A.C.O. The STUbL RNF4 regulates protein group SUMOylation by targeting the SUMO conjugation machinery. *Nat Commun* **8**, 1809 (2017).

506. Olivieri, M. et al. A Genetic Map of the Response to DNA Damage in Human Cells. *Cell* **182**, 481-496 e21 (2020).

507. Geoffroy, M.C., Jaffray, E.G., Walker, K.J. & Hay, R.T. Arsenic-induced SUMO-dependent recruitment of RNF4 into PML nuclear bodies. *Mol Biol Cell* **21**, 4227-39 (2010).

508. de The, H., Chomienne, C., Lanotte, M., Degos, L. & Dejean, A. The t(15;17) translocation of acute promyelocytic leukaemia fuses the retinoic acid receptor alpha gene to a novel transcribed locus. *Nature* **347**, 558-61 (1990).

509. de The, H. et al. The PML-RAR alpha fusion mRNA generated by the t(15;17) translocation in acute promyelocytic leukemia encodes a functionally altered RAR. *Cell* **66**, 675-84 (1991).

510. Kiianitsa, K. & Maizels, N. A rapid and sensitive assay for DNA-protein covalent complexes in living cells. *Nucleic Acids Res* **41**, e104 (2013).

511. Glumac, M. et al. SPRTN-dependent DPC degradation precedes repair of damaged DNA: a proof of concept revealed by the STAR assay. *Nucleic Acids Res* **51**, e35 (2023).

512. Hu, Q. et al. The ARK Assay Is a Sensitive and Versatile Method for the Global Detection of DNA-Protein Crosslinks. *Cell Rep* **30**, 1235-1245 e4 (2020).

513. Subramanian, D., Furbee, C.S. & Muller, M.T. ICE bioassay. Isolating in vivo complexes of enzyme to DNA. *Methods Mol Biol* **95**, 137-47 (2001).

514. Nielsen, I. et al. A Flp-nick system to study repair of a single protein-bound nick in vivo. *Nat Methods* **6**, 753-7 (2009).

515. Chen, L. et al. Direct identification of the active-site nucleophile in a DNA (cytosine-5)-methyltransferase. *Biochemistry* **30**, 11018-25 (1991).

References

516. Zhao, S. et al. RNF14-dependent atypical ubiquitylation promotes translation-coupled resolution of RNA-protein crosslinks. *Mol Cell* (2023).

517. Weickert, P. et al. SPRTN patient variants cause global-genome DNA-protein crosslink repair defects. *Nat Commun* **14**, 352 (2023).

518. Ruggiano, A. et al. The protease SPRTN and SUMOylation coordinate DNA-protein crosslink repair to prevent genome instability. *Cell Rep* **37**, 110080 (2021).

519. Kroning, A., van den Boom, J., Kracht, M., Kueck, A.F. & Meyer, H. Ubiquitin-directed AAA+ ATPase p97/VCP unfolds stable proteins crosslinked to DNA for proteolysis by SPRTN. *J Biol Chem* **298**, 101976 (2022).

520. Sherpa, D., Chrustowicz, J. & Schulman, B.A. How the ends signal the end: Regulation by E3 ubiquitin ligases recognizing protein termini. *Mol Cell* **82**, 1424-1438 (2022).

521. Adam, S. et al. DNA sequence-dependent activity and base flipping mechanisms of DNMT1 regulate genome-wide DNA methylation. *Nat Commun* **11**, 3723 (2020).

522. Ren, W., Gao, L. & Song, J. Structural Basis of DNMT1 and DNMT3A-Mediated DNA Methylation. *Genes (Basel)* **9**(2018).

523. Mangini, N.S., Wesolowski, R., Ramaswamy, B., Lustberg, M.B. & Berger, M.J. Palbociclib: A Novel Cyclin-Dependent Kinase Inhibitor for Hormone Receptor-Positive Advanced Breast Cancer. *Ann Pharmacother* **49**, 1252-60 (2015).

524. Verma, R., Oania, R., Fang, R., Smith, G.T. & Deshaies, R.J. Cdc48/p97 mediates UV-dependent turnover of RNA Pol II. *Mol Cell* **41**, 82-92 (2011).

525. Vance, J.R. & Wilson, T.E. Yeast Tdp1 and Rad1-Rad10 function as redundant pathways for repairing Top1 replicative damage. *Proc Natl Acad Sci U S A* **99**, 13669-74 (2002).

526. Takahata, C. et al. Repair synthesis step involving ERCC1-XPF participates in DNA repair of the Top1-DNA damage complex. *Carcinogenesis* **36**, 841-51 (2015).

527. Zhang, H. et al. TDP1-independent pathways in the process and repair of TOP1-induced DNA damage. *Nat Commun* **13**, 4240 (2022).

528. Pepe, A. & West, S.C. Substrate specificity of the MUS81-EME2 structure selective endonuclease. *Nucleic Acids Res* **42**, 3833-45 (2014).

529. Sun, Y. et al. Flap endonuclease 1 repairs DNA-protein crosslinks via ADP-ribosylation. *bioRxiv*, 2023.10.19.563118 (2023).

530. Alvarez-Quilon, A. et al. Endogenous DNA 3' Blocks Are Vulnerabilities for BRCA1 and BRCA2 Deficiency and Are Reversed by the APE2 Nuclease. *Mol Cell* **78**, 1152-1165 e8 (2020).

References

531. Ghodke, P.P. & Guengerich, F.P. DNA polymerases eta and kappa bypass N(2)-guanine-O(6)-alkylguanine DNA alkyltransferase cross-linked DNA-peptides. *J Biol Chem* **297**, 101124 (2021).

532. Nakano, T. et al. Homologous recombination but not nucleotide excision repair plays a pivotal role in tolerance of DNA-protein cross-links in mammalian cells. *J Biol Chem* **284**, 27065-76 (2009).

533. Interthal, H. et al. SCAN1 mutant Tdp1 accumulates the enzyme--DNA intermediate and causes camptothecin hypersensitivity. *EMBO J* **24**, 2224-33 (2005).

534. Pegg, A.E. Multifaceted roles of alkyltransferase and related proteins in DNA repair, DNA damage, resistance to chemotherapy, and research tools. *Chem Res Toxicol* **24**, 618-39 (2011).

535. Guengerich, F.P., Fang, Q., Liu, L., Hachey, D.L. & Pegg, A.E. O6-alkylguanine-DNA alkyltransferase: low pKa and high reactivity of cysteine 145. *Biochemistry* **42**, 10965-70 (2003).

536. Cmarik, J.L. et al. Mutation spectrum and sequence alkylation selectivity resulting from modification of bacteriophage M13mp18 DNA with S-(2-chloroethyl)glutathione. Evidence for a role of S-(2-N7-guanyl)ethyl)glutathione as a mutagenic lesion formed from ethylene dibromide. *J Biol Chem* **267**, 6672-9 (1992).

537. Loeber, R., Rajesh, M., Fang, Q., Pegg, A.E. & Tretyakova, N. Cross-linking of the human DNA repair protein O6-alkylguanine DNA alkyltransferase to DNA in the presence of 1,2,3,4-diepoxybutane. *Chem Res Toxicol* **19**, 645-54 (2006).

538. Liu, L., Pegg, A.E., Williams, K.M. & Guengerich, F.P. Paradoxical enhancement of the toxicity of 1,2-dibromoethane by O6-alkylguanine-DNA alkyltransferase. *J Biol Chem* **277**, 37920-8 (2002).

Publications

Article

SPRTN patient variants cause global-genome DNA-protein crosslink repair defects

Check for updates

Pedro Weickert[1,2,4], Hao-Yi Li[1,2,4], Maximilian J. Götz[1,2], Sophie Dürauer[1,2], Denitsa Yaneva[1,2], Shubo Zhao[1,2], Jacqueline Cordes [1,2], Aleida C. Acampora[1,2], Ignasi Forne [3], Axel Imhof [3] & Julian Stingele [1,2]

DNA-protein crosslinks (DPCs) are pervasive DNA lesions that are induced by reactive metabolites and various chemotherapeutic agents. Here, we develop a technique for the Purification of x-linked Proteins (PxP), which allows identification and tracking of diverse DPCs in mammalian cells. Using PxP, we investigate DPC repair in cells genetically-engineered to express variants of the SPRTN protease that cause premature ageing and early-onset liver cancer in Ruijs-Aalfs syndrome patients. We find an unexpected role for SPRTN in global-genome DPC repair, that does not rely on replication-coupled detection of the lesion. Mechanistically, we demonstrate that replication-independent DPC cleavage by SPRTN requires SUMO-targeted ubiquitylation of the protein adduct and occurs in addition to proteasomal DPC degradation. Defective ubiquitin binding of SPRTN patient variants compromises global-genome DPC repair and causes synthetic lethality in combination with a reduction in proteasomal DPC repair capacity.

Unrepaired DNA damage causes ageing and cancer formation[1,2]. Therefore, cells employ DNA repair pathways, which operate not only in a transcription-[3] or replication-coupled[4] manner, but also involve global mechanisms that scan the entire genome for lesions[5]. Covalent DNA-protein crosslinks (DPCs) are a particular pervasive type of DNA damage and are targeted by multiple repair enzymes[6]. DNA-protein crosslinking arises from enzymatic and non-enzymatic sources[7]. Non-enzymatic DPC formation is induced by bifunctional chemical crosslinkers such as platinum-based chemotherapeutics or formaldehyde, which is even produced within chromatin during histone demethylation and is present at micromolar concentrations in mammalian blood[8]. Enzymatic DPCs are caused by entrapment of normally transient covalent enzyme-DNA reaction intermediates and are induced by various chemotherapeutic agents including topoisomerase poisons and the antineoplastic drug 5-aza-2′-deoxycytidine (5-azadC)[9]. 5-azadC is incorporated into DNA during replication, where it acts as pseudo-substrate for DNA methyltransferase 1 (DNMT1) leading to formation of a covalent complex between the modified base and DNMT1's active site cysteine[10,11]. However, upon

methylating 5-azadC, DNMT1 fails to release from DNA, thereby forming a stable DPC.

DPC repair involves the proteolytic degradation of the protein adduct by metalloproteases of the Wss1/SPRTN family[12–17]. While loss of SPRTN is lethal in mammalian cells, hypomorphic variants cause Ruijs-Aalfs syndrome, which is characterized by premature ageing and early-onset hepatocellular carcinoma[16,18,19]. Ruijs-Aalfs syndrome is primarily caused by frame-shift mutations resulting in expression of C-terminally truncated SPRTN-ΔC variants, which lack nuclear localisation signals and various protein-protein interaction motifs[18]. Data obtained in frog egg extracts demonstrated that DPC cleavage by SPRTN can be initiated by a replication fork colliding with a DPC[20–22]. While the replicative helicase is able to bypass the protein adduct, DNA polymerases fail to synthesize across the DPC[20,21]. SPRTN recognizes the resulting single-/double-stranded DNA junction using a bipartite DNA-binding module, which triggers local activation of the enzyme and concurrent DPC cleavage[23]. In egg extracts, DPCs are additionally targeted by replication-coupled proteasomal degradation[20]. Recent reports suggest that the proteasome also targets DPCs outside of

[1]Department of Biochemistry, Ludwig-Maximilians-University, 81377 Munich, Germany. [2]Gene Center, Ludwig-Maximilians-University, 81377 Munich, Germany. [3]Protein Analysis Unit (ZfP), BioMedical Center (BMC), Ludwig-Maximilians-University, 82152 Martinsried, Germany. [4]These authors contributed equally: Pedro Weickert, Hao-Yi Li. e-mail: stingele@genzentrum.lmu.de

replication, which relies on initial SUMOylation of the protein adduct and subsequent ubiquitylation by the SUMO-targeted ubiquitin ligase RNF4[24,25]. In contrast, it is currently believed that SPRTN acts exclusively at the replication fork, where it relies on ubiquitin signals for recruitment[20,26]. No consensus has emerged regarding the role of SUMO modifications for SPRTN-dependent DPC repair. SUMOylation has been suggested to block alternative repair pathways to promote SPRTN-dependent repair[26,27]; while SUMOylation was found to be dispensable for SPRTN function in another study[28]. At any rate, the existence of at least two proteolytic systems to degrade DPCs - SPRTN and the proteasome - indicates significant evolutionary pressure to cope with these insults in order to preserve genome integrity. However, the relationship between proteasome- and SPRTN-dependent repair as well as their relative contribution towards DPC cleavage in mammalian cells remain unknown.

Exploring DPC repair in mammalian cells in mechanistic detail has remained challenging not only due to the essential function of the SPRTN protease, but also due to limitations of the currently available techniques for the study of DPCs. DPC formation in mammalian cells can be assessed by separating DPCs from non-crosslinked proteins using ultra-centrifugation of caesium chloride gradients[29,30]. However, this approach is laborious, low throughput, and requires substantial amounts of material. Most other DPC assays are based on precipitation as the separating principle. In the KCl-SDS assay (and its derivative ARK)[31,32], proteins are precipitated from denaturing lysates and co-precipitating DNA is quantified as a proxy for the amount of DPCs. The RADAR (rapid approach to DNA adduct recovery) assay employs the opposite principle; DNA is precipitated from lysates and co-precipitating proteins are analysed using slot-blotting or silver-staining[33]. The reliance on precipitation is a major drawback of these assays. DPCs are diverse in nature, which will affect their behaviour during precipitation. While smaller protein adducts may efficiently co-precipitate with DNA, larger adducts may even prevent DNA from precipitating.

Here, we present a method for the Purification of x-linked Proteins (PxP) that overcomes these limitations. PxP is based on electro-elution of non-covalently attached proteins from DNA embedded in agarose plugs and can be combined with label-free quantitative mass spectrometry to determine the identity of unknown DPCs. In addition, we developed genetically-engineered hypomorphic *SPRTN* mutant cell lines expressing patient-mimicking variants enabling not only structure-function analysis of SPRTN in cells, but also the genetic exploration of relationships between different DPC repair factors. Using these tools, we describe an unexpected role for SPRTN in replication-independent DPC repair. We find that this global-genome DPC cleavage by SPRTN requires SUMO-targeted ubiquitylation of the DPC, occurs independent of proteasomal degradation, and is defective in cells expressing Ruijs-Aalfs syndrome-associated SPRTN variants. As a consequence, reduction of proteasomal DPC degradation causes synthetic defects in *SPRTN* mutant cell lines. Finally, structure-function analysis of SPRTN demonstrates that the loss of a ubiquitin-binding domain in patient variants is responsible for defective global-genome DPC repair.

Results

A strategy for the purification of crosslinked proteins
The technique described here was inspired by chromosome entrapment experiments that had been designed to investigate interactions between prokaryotic condensin and DNA in *Bacillus subtilis*[34]. In these experiments, bacterial chromosomes were immobilized in low-melt agarose plugs to assess topological interactions with covalently-closed condensin rings. We hypothesized that a similar principle could be utilized to monitor and identify DPCs in mammalian cells. Based on this idea, we designed an assay for the Purification of x-linked Proteins (PxP) (Fig. 1a). In brief, mammalian cells are harvested and embedded in low-melt agarose plugs. Next, plugs are transferred to denaturing lysis

buffer containing 2% sarkosyl. Upon completion of cell lysis, plugs are transferred to wells of an SDS-PAGE gel and subjected to electro-elution. During electrophoresis, cellular proteins exit the plug, while DNA (due to its high molecular weight) and crosslinked proteins remain inside. Plugs are then retrieved, melted, and DNA is digested with a nuclease to release the crosslinked proteins. Finally, crosslinked proteins are analysed using SDS-PAGE coupled with western blotting or silver staining. To test our experimental strategy, we first analysed camptothecin (CPT)-induced TOP1-DPCs which formed in a dose-dependent manner with no background signal detectable in untreated cells (Fig. 1b and Supplementary Fig. 1a). Having established that the PxP procedure allows detection of specific DPCs, we next asked whether it can also reveal the identity of non-enzymatic DPCs induced by chemical crosslinkers. To answer this question, we first introduced a control that allows the distinction between co-purifying contaminants and DPCs. Cells of each experimental condition were cast into two plugs. One plug was digested with a nuclease prior to electro-elution, while the second plug was incubated in buffer only. DPCs are expected to elute from the plug upon DNA digestion, while co-purifying contaminants are not (Fig. 1a). We subjected cells to a 1-h formaldehyde pulse, performed PxP extraction, and analysed samples on silver-stained SDS-PAGE gels. Distinct formaldehyde-induced bands could be detected, which were sensitive to nuclease treatment prior to electro-elution, suggesting that treatment with formaldehyde results in crosslinking of specific proteins (Fig. 1c and Supplementary Fig. 1b). To reveal the identity of formaldehyde-induced DPCs, we combined PxP with label-free quantitative proteomics. Plugs were retrieved after electro-elution, fixed, and subjected to in-plug tryptic digestion and detection by LC-MS/MS. Thirty-five proteins were significantly enriched in PxP plugs after formaldehyde exposure (Fig. 1d, e). The most abundant formaldehyde-induced DPCs were formed by core histones (Fig. 1d, e). Histone crosslinking was confirmed by western blotting and could be observed at low formaldehyde concentrations, which did not affect long-term viability (Fig. 1f, g and Supplementary Fig. 1c). We conclude that PxP enables identification of unknown DPCs. In addition, we find that the challenge to preserve genome integrity after formaldehyde exposure is less complex than previously anticipated as formaldehyde-induced DPCs mainly consist of crosslinked nucleosomes.

Replication-independent repair of 5-azadC-induced DNMT1-DPCs monitored by PxP
5-azadC-induced DNMT1-DPCs form post-replicatively and are, thus, an ideal model lesion to study replication-independent DPC repair[25]. Therefore, we tested whether PxP can be used to track the fate of DNMT1-DPCs. We synchronized cells using a double thymidine block, released them into early/mid S-phase, and subjected them to a 30-minute pulse of increasing 5-azadC concentrations (Fig. 2a). Using PxP followed by western blotting, we observed dose-dependent formation of DNMT1-DPCs (Fig. 2b), which were sensitive to nuclease treatment prior to electro-elution (Supplementary Fig. 2a, b). To monitor repair of DNMT1-DPCs, we harvested cells either immediately after 5-azadC treatment or following a chase in drug-free media for 2 h (Fig. 2a). The bulk of DNMT1-DPCs was repaired during the chase, which, in agreement with a previous report[25], was blocked by pre-treating cells with proteasome inhibitor MG132 (Fig. 2c), by depleting the sole SUMO E2 conjugating enzyme UBC9 (Supplementary Fig. 2c), and by chemical inhibition of SUMO-E1 or ubiquitin-E1 activating enzymes (Fig. 2d, e). Because pre-treatment with ubiquitin-E1 inhibitor interfered with DNMT1-DPC formation (Supplementary Fig. 2d), it was added together with 5-azadC (see scheme in Fig. 2a). Moreover, chemical inhibition of the ATPase p97, which is required for proteasomal degradation of many chromatin proteins[35], blocked bulk DNMT1-DPC repair (Supplementary Fig. 2e). Interestingly, we noted the appearance of a faster migrating DNMT1-DPC species in PxP and input samples 2 h after 5-azadC exposure (Fig. 2c–e and Supplementary Fig. 2c–e, orange

dots). While this species increased upon proteasome or p97 inhibition (Fig. 2c and Supplementary Fig. 2e, orange dots), it was absent after blocking of either SUMOylation or ubiquitylation (Fig. 2d, e and Supplementary Fig. 2c, orange dots). Taken together, this indicated to us that DNMT1-DPCs are proteolytically cleaved in a SUMO- and ubiquitin-dependent manner by an alternative DPC protease, which occurs in parallel to the previously reported proteasomal degradation.

DPC SUMOylation and subsequent ubiquitylation have been proposed to rely on the SUMO E3 ligase PIAS4 and the SUMO-targeted ubiquitin ligase RNF4[24,25]. However, we did not observe a reduction in DPC degradation upon siRNA-mediated depletion or knock-out of *PIAS4* (Supplementary Fig. 2f, g), perhaps indicating redundancy with another SUMO-E3 ligase. In contrast, knock-out of *RNF4* resulted in

clear reduction of bulk degradation and reduced formation of the putative DNMT1-DPC cleavage fragment (Fig. 2f). Consistently and in line with a previous report[25], we found *RNF4* knock-out (KO) cells to be sensitive to 5-azadC exposure (Supplementary Fig. 3a, b). Of note, while RNF4 depletion clearly delayed repair, we observed residual degradation and appearance of the cleaved DNMT1 fragment after a prolonged chase period of up to 6 h (Fig. 2g). Residual repair in *RNF4* KO cells was blocked by chemical inhibition of SUMO- or ubiquitin E1-activating enzymes (Fig. 2g and Supplementary Fig. 3c), suggesting that a second SUMO-targeted ubiquitin ligase activity acts as a, albeit less efficient, back-up to RNF4. We conclude that in addition to the proteasome a second proteolytic activity acts downstream of SUMO-targeted ubiquitylation during global-genome DPC repair.
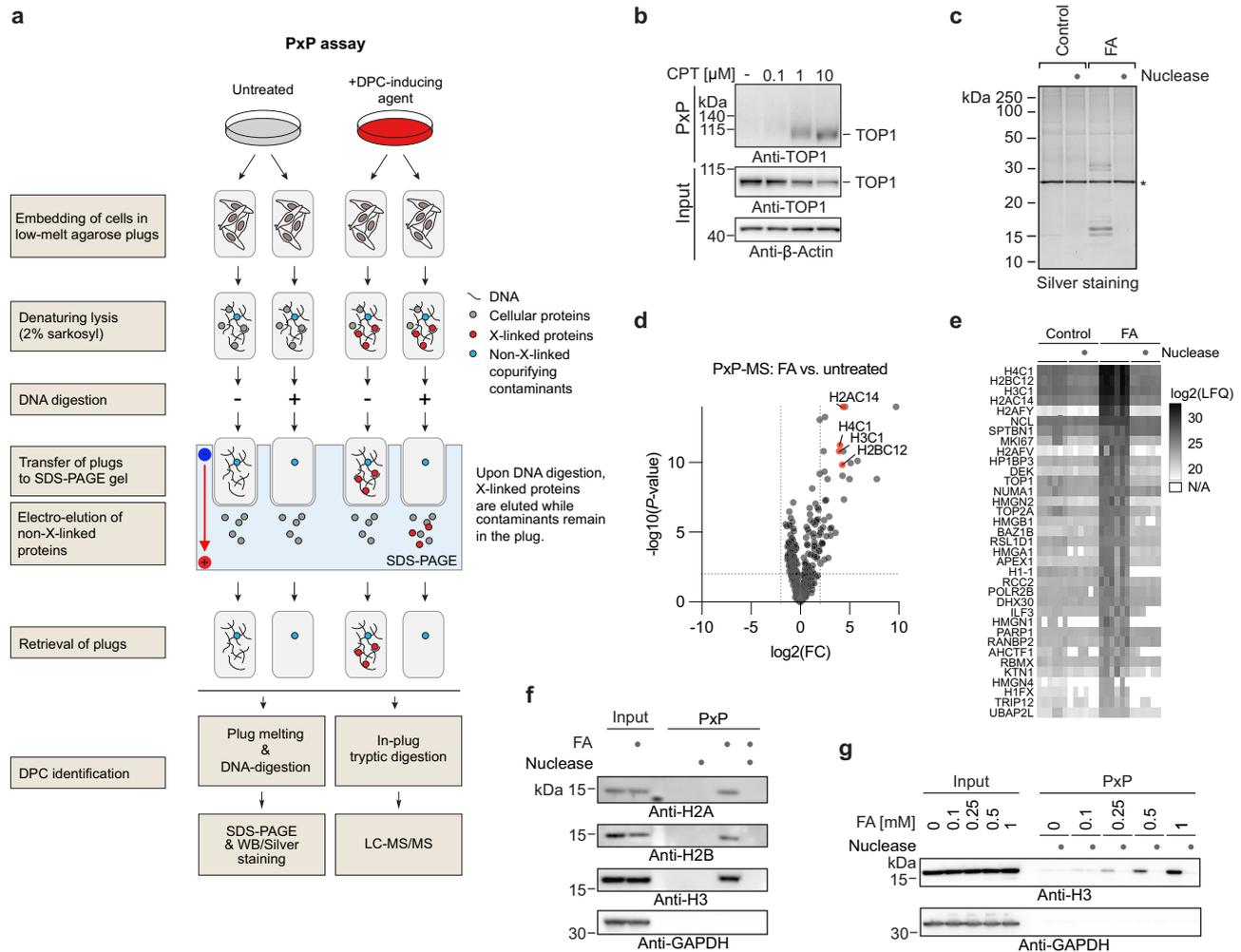
Fig. 1 | A strategy for the purification of crosslinked proteins. a Schematic depiction of the Purification of x-linked Proteins (PxP) assay. Cells are harvested and embedded in low-melt agarose plugs. Plugs are transferred to denaturing lysis buffer. Upon completion of lysis, DNA is optionally digested using a nuclease. Next, plugs are transferred to an SDS-PAGE gel and subjected to electro-elution. For DPC detection, plugs are melted following electro-elution, digested with nuclease and analysed using SDS-PAGE followed by western blotting or silver-staining. Alternatively, plugs are fixed and subjected to in-plug tryptic digestion for quantitative proteomics. **b** Camptothecin (CPT)-induced TOP1-DPC formation assessed by PxP. HeLa T-REx Flp-In cells were treated for 30 min with the indicated doses of CPT prior to isolation of DPCs using PxP and analysis by western blotting. **c** Untreated or formaldehyde (FA)-treated (2 mM, 1 h) HeLa cells were processed as depicted in (**a**) and analysed by SDS-PAGE and silver staining. Asterisk indicates Benzonase nuclease used to digest all samples prior to running the final SDS-PAGE. **d** Mass spectrometry analysis of PxP samples comparing untreated and FA-treated (2 mM,

1 h) HeLa cells. Six plugs per condition were subjected to in-plug tryptic digestion followed by label-free quantitative mass spectrometry. Volcano plot depicting fold change (FC, log2) between conditions plotted against FDR-adjusted *P*-value (-log10). See also Supplementary Data 1, Supplementary Data 2. **e** Heatmap showing normalized intensities of six replicates of statistically significant FA-induced DPCs (FDR-adjusted $P < 0.01$, FC > 2) identified in (**d**) ranked by average intensity upon FA-treatment. See also Supplementary Data 1, Supplementary Data 2. **f** PxP analysis of FA-induced histone crosslinks. Cells were treated for 1 h with 2 mM FA and subjected to PxP extraction including a nuclease digestion as indicated and analysed by western blotting. The experiment was repeated twice and similar results were obtained. **g** PxP analysis of histone H3 crosslinks induced by increasing concentrations of FA. Cells were treated for 1 h with the indicated doses of FA and subjected to PxP analysis including a nuclease digestion as indicated and analysed by western blotting. The experiment was repeated three times and similar results were obtained. Source data are provided as a Source Data file.
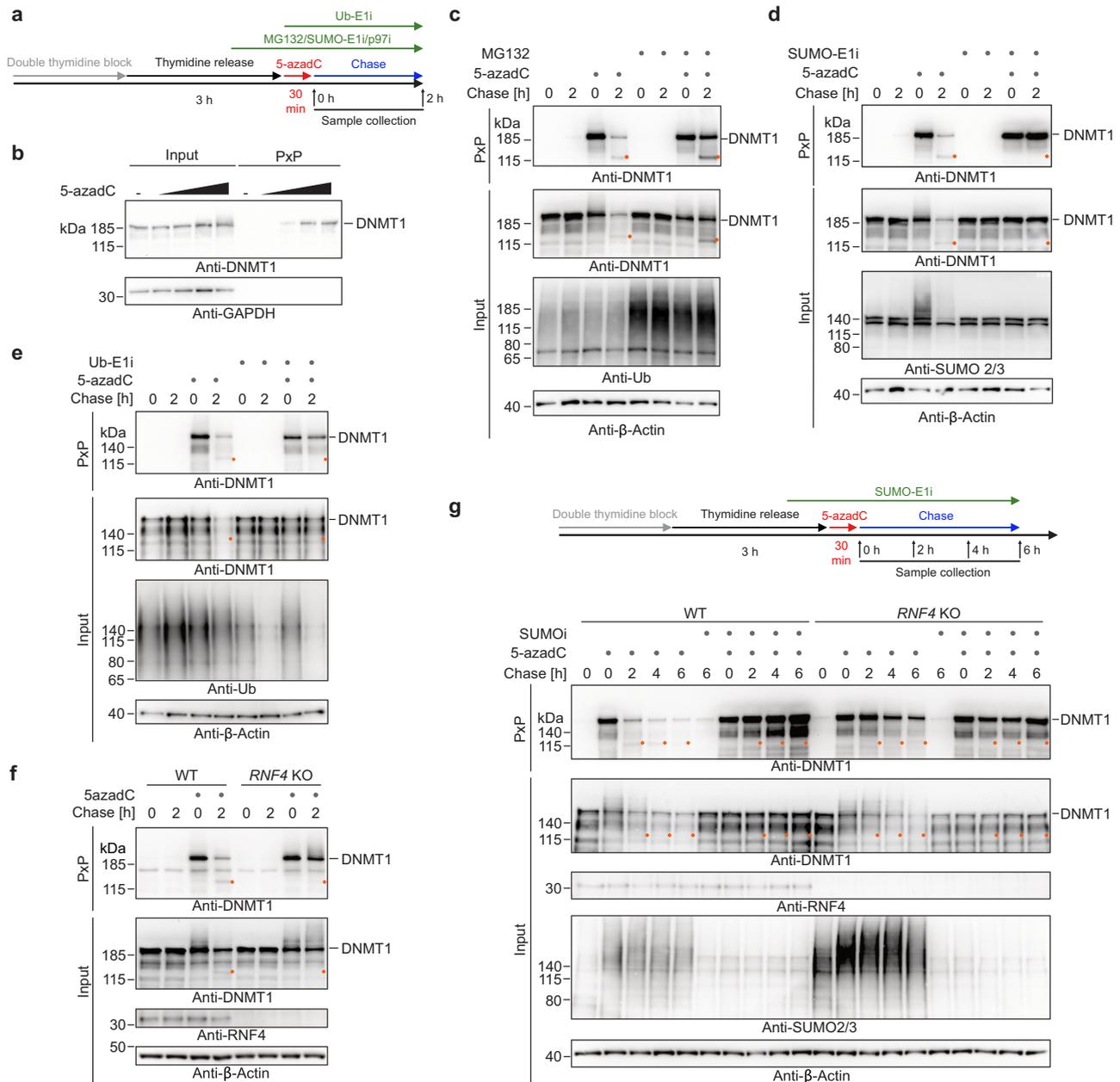
Fig. 2 | Global-genome repair of 5-azadC-induced DNMT1-DPCs monitored by PxP. a Schematic depiction of the experimental workflow used to monitor the repair of 5-azadC-induced DNMT1-DPCs. Cells were synchronized via a double thymidine block and released into early/mid S-phase for 3 h prior to induction of DNMT1-DPCs by a 30-min pulse of 5-azadC. Samples were collected either immediately after 5-azadC exposure or following a chase in drug-free media. Proteasome inhibitor (MG132, 5 μM), p97 inhibitor (p97i CB-5083, 5 μM) and SUMOylation inhibitor (SUMO-E1i ML-792, 5 μM) were added 1 h prior to induction of DPCs and kept during the chase with 5-azadC-free medium. Ubiquitylation inhibitor (Ub-E1i TAK-243, 1 μM) was added together with 5-azadC. **b** 5-azadC-induced DNMT1-DPC formation assessed by PxP. HeLa T-REx Flp-In cells were treated as depicted in (**a**) with the

indicated doses of 5-azadC for 30 min prior to immediate isolation of DPCs using PxP and western blotting analysis. **c**–**f** 5-azadC-induced DNMT1-DPC formation and repair upon proteasome inhibition (**c**), inhibition of SUMOylation (**d**), inhibition of ubiquitylation (**e**), or knock-out of *RNF4* (**f**) assessed by PxP. HeLa T-REx Flp-In cells were treated as depicted in (**a**) prior to extraction of DPCs using PxP, and analysis of samples by western blotting using the indicated antibodies. **g** HeLa WT and *RNF4* knock-out (KO) cells were treated and analysed as depicted including an optional treatment with SUMOylation inhibitor (SUMO-E1i ML-792, 5 μM), prior to extraction of DPCs using PxP and analysis of samples by western blotting using the indicated antibodies. Experiments in (**c**–**g**) were repeated three times and similar results were obtained. Source data are provided as a Source Data file.

SPRTN cleaves DNMT1-DPCs

The DPC-specific metalloprotease SPRTN is currently believed to act exclusively at the replication fork[14,17,20,36]. Surprisingly however, siRNA-mediated depletion of SPRTN completely abolished the appearance of the faster-migrating DNMT1-DPC species even upon proteasome inhibition, while neither bulk degradation nor DPC formation were affected (Fig. 3a, orange dots). The appearance of the DNMT1-DPC fragment was restored by expression of a siRNA-resistant version of

SPRTN-WT but not by catalytically-inactive SPRTN-E112Q (EQ) (Fig. 3b, orange dots). These data suggest that the observed DNMT1-DPC fragment is a product of SPRTN-dependent proteolysis. DNMT1-DPCs form in the wake of DNA synthesis, therefore it seemed unlikely that the cleaved DPC is a consequence of SPRTN's established role in replication-coupled DPC repair. Indeed, inhibition of DNA synthesis by aphidicolin following induction of DNMT1-DPCs had no effect on SPRTN-dependent DPC cleavage or bulk repair (Fig. 3c, orange dots,
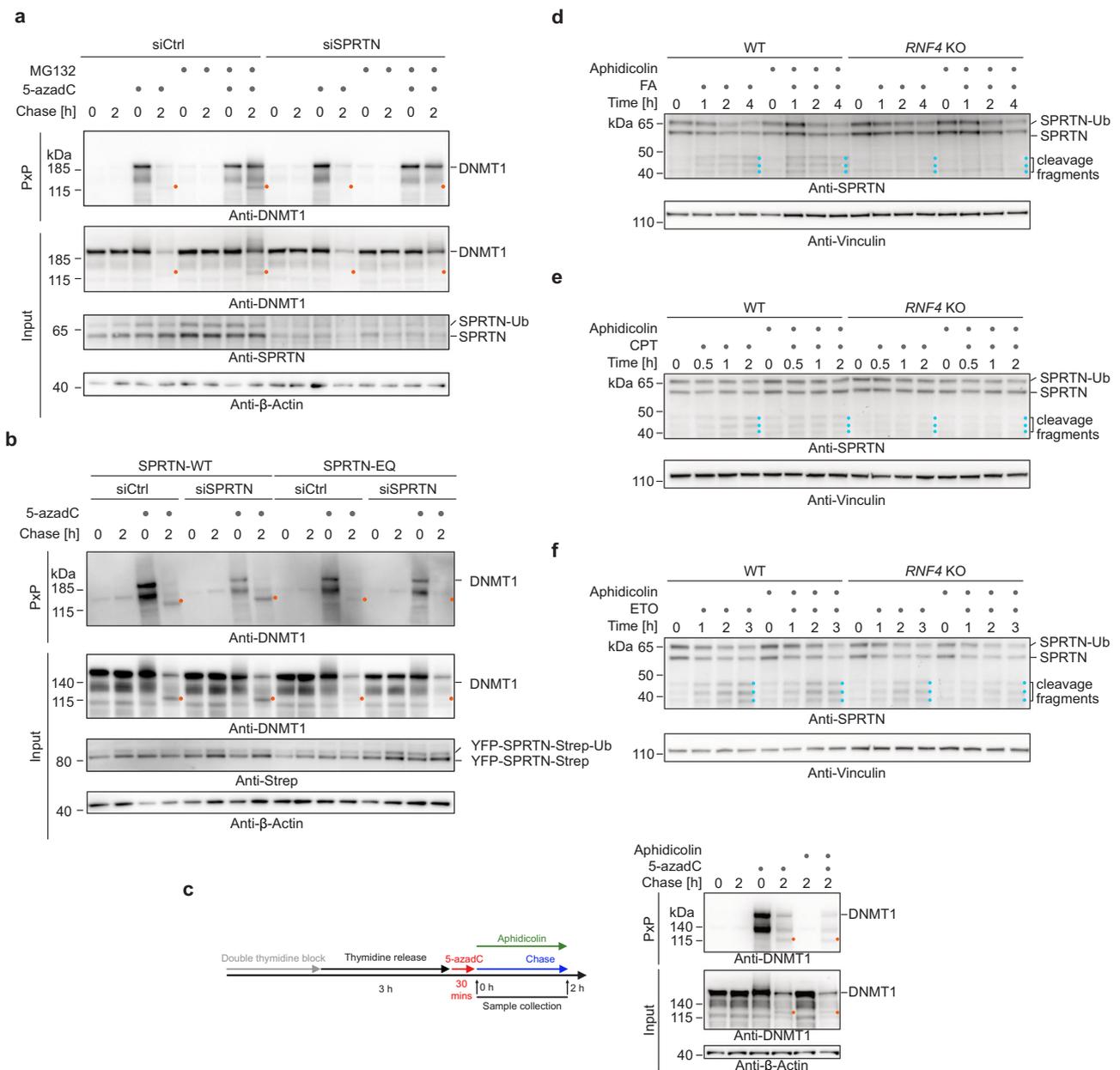
Fig. 3 | The metalloprotease SPRTN cleaves DNMT1-DPCs during global-genome repair. a HeLa T-REx Flp-In cells transfected with the indicated siRNAs were treated as depicted in Fig. 2a. DNMT1-DPCs were isolated using PxP and analysed by western blotting using the indicated antibodies. **b** HeLa T-REx Flp-In cells stably expressing siRNA-resistant SPRTN variants (wildtype (WT) or catalytically inactive E112Q (EQ)) were transfected with the indicated siRNAs and treated as in Fig. 2a. DPCs were isolated by PxP and analysed by western blotting using the indicated antibodies. **c** 5-azadC-induced DNMT1-DPC repair upon inhibition of DNA synthesis assessed by PxP. HeLa T-REx Flp-In cells were treated as depicted including an optional addition of aphidicolin (3 μM) during the chase (left). DNMT1-DPCs were isolated using PxP and analysed by western blotting using the indicated antibodies (right). **d**–**f** HeLa WT or *RNF4* KO cells were treated with formaldehyde (FA, 250 μM) (**d**), camptothecin (CPT, 500 nM) (**e**) or etoposide (ETO, 50 μM) (**f**), including a 2-h pre-treatment with aphidicolin, as indicated, before whole cell lysates were analysed by western blotting using the indicated antibodies. Source data are provided as a Source Data file.

and Supplementary Fig. 4a). Moreover, cleavage was not affected by knock-out of the adaptor protein TEX264 (Supplementary Fig. 4b, orange dots), which was shown previously to be involved in replication-coupled repair of DPCs by SPRTN[36]. We also excluded an involvement of transcription, because inhibition of RNA synthesis using the CDK9-inhibitor flavopiridol showed no effect on DPC cleavage or repair (Supplementary Fig. 4c, d, orange dots).

To test whether SPRTN also responds to other types of DPCs in a replication-independent manner, we monitored autocleavage of the protease, an indicator of SPRTN activation[13,14,37]. We treated cells with formaldehyde (thereby inducing histone-DPCs), CPT (TOP1-DPCs), or etoposide (ETO, causing TOP2-DPCs) and monitored accumulation of

SPRTN autocleavage fragments over time. Formaldehyde- and CPT-induced autocleavage was strongly reduced in *RNF4* KO cells, suggesting that SPRTN activation by TOP1- and histone-DPCs occurs similar to what we observed upon post-replicative induction of DNMT1-DPCs (Fig. 3d, e, blue dots). Interestingly, etoposide-induced SPRTN autocleavage occurred largely independent of RNF4 and was partially reduced by aphidicolin in *RNF4* KO cells (Fig. 3f, blue dots), indicating that TOP2-DPCs are sensed and signalled differently. In the case of CPT and FA however, SPRTN autocleavage was completely unaffected by inhibition of DNA synthesis using aphidicolin (Fig. 3d, e, blue dots, and Supplementary Fig. 4e). This was in contrast to SPRTN's role at replication forks, which relies on DNA polymerases extending

nascent strands up to the protein adduct[20]. We thus conclude that SPRTN responds to various DPCs in a global-genome manner that does not rely on the replication machinery to detect the lesion. Next, we asked whether global-genome DPC repair by SPRTN is also active outside the S/G2-phase (when SPRTN expression levels are high[38]). We arrested cells using the CDK4/CDK6 inhibitor palbociclib in early G1 phase (Supplementary Fig. 4f), which was accompanied by a strong reduction in SPRTN protein levels (Supplementary Fig. 4g–i). Low levels of SPRTN expression made the assessment of autocleavage impossible, but also indicated that it is unlikely that the protease is important in G1 phase. Collectively, these data demonstrate that SPRTN targets DPCs during global-genome repair downstream of SUMO-targeted ubiquitylation and that this mechanism, while being replication-independent, primarily operates in the S/G2 phase of the cell cycle.
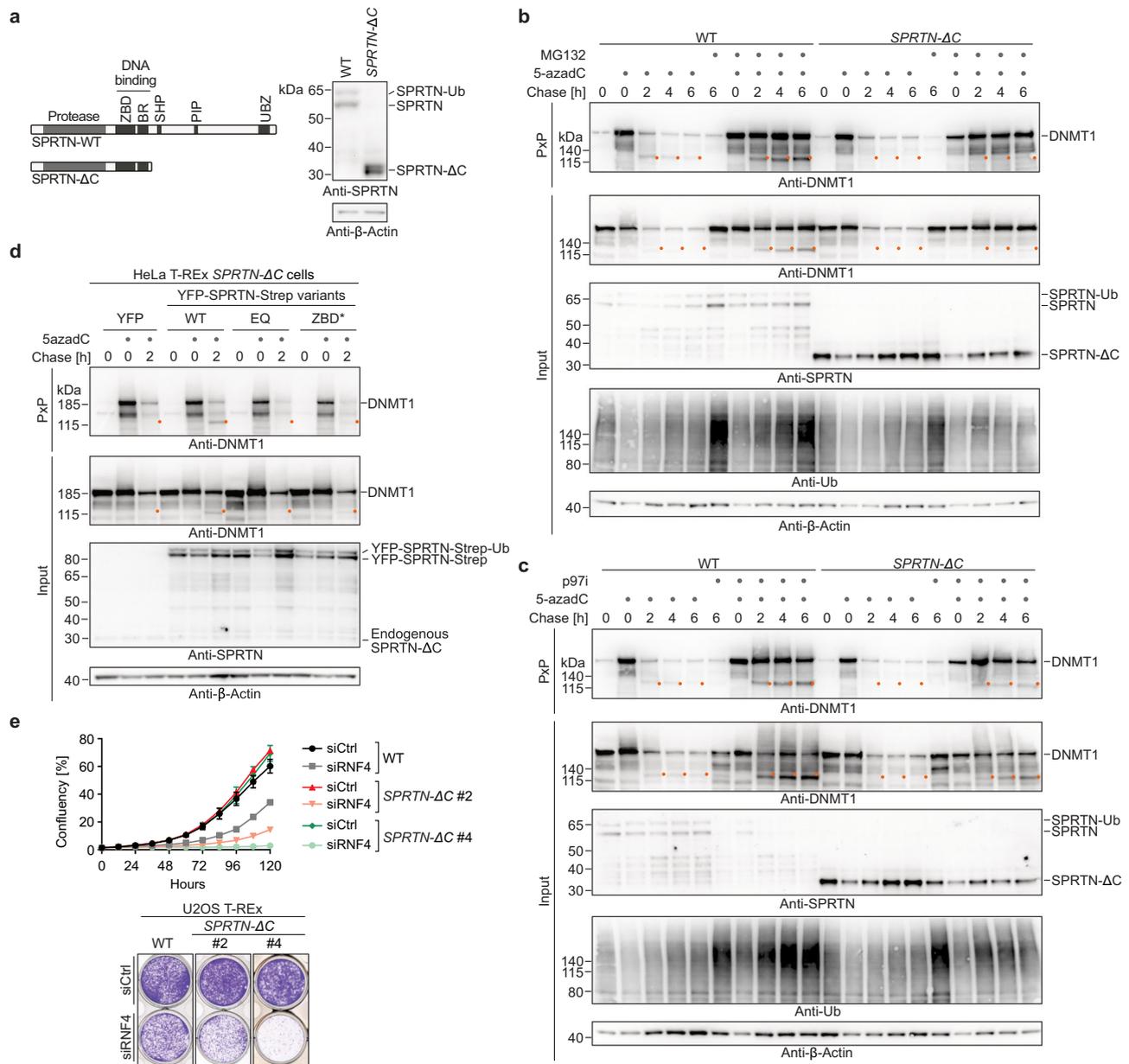
SPRTN patient variants affect replication-independent DNA-protein crosslink repair

Ruijs-Aalfs syndrome is caused by partial loss-of-function *SPRTN* mutations and is characterized by progeroid features and early onset hepatocellular carcinomas[18,39]. Intriguingly, several aspects of the disease are difficult to reconcile with a purely replicative function of SPRTN. Patients and mice bearing hypomorphic *SPRTN* mutations display signs of failed tissue homeostasis in the largely quiescent liver and in postmitotic lens epithelial cells[18,19,39]. In contrast, the highly proliferative haematopoietic system, which is in addition challenged by high endogenous formaldehyde concentrations[40], seems not to be affected. Therefore, we asked whether replication-independent cleavage of DPCs may be affected by patient variants. To investigate this question, we engineered cells to express patient-mimicking variants. We edited the endogenous *SPRTN* locus in HeLa T-REx Flp-In cells using two gRNAs resulting in the deletion of the entire coding region of exon 5 (Supplementary Fig. 5a). The resulting mutant cells express a SPRTN-ΔC variant, which is highly reminiscent of the truncated SPRTN variants observed in Ruijs-Aalfs syndrome patients (Fig. 4a and Supplementary Fig. 5a). While *SPRTN-ΔC* cells were viable, they failed to efficiently cleave DNMT1-DPCs; the DPC cleavage band observed during a 6-h chase in WT cells was hardly detectable in mutant cells (Fig. 4b, c). Residual amounts of DPC cleavage fragments were only observed upon inhibition of proteasomal degradation or p97 activity (Fig. 4b, c). Re-expression of SPRTN-WT, but not of SPRTN-EQ rescued the cleavage of DNMT1-DPCs in *SPRTN-ΔC* cells (Fig. 4d). A catalytically-compromised DNA-binding mutant SPRTN-ZBD* (R185A)[23,41] displayed strongly reduced activity (Fig. 4d). Despite being unable to efficiently cleave DNMT1-DPCs, *SPRTN-ΔC* cells were not sensitive to exposure of 5-azadC (Supplementary Fig. 5b), likely due to redundant DPC degradation by the proteasome. In line with SPRTN acting downstream of RNF4, 5-azadC sensitivity caused by depletion of RNF4 was comparable in *SPRTN-ΔC* and in wild-type HeLa T-REx Flp-In cells (Supplementary Fig. 5b). In addition, we noted that RNF4 depletion resulted in mild synthetic growth defects in *SPRTN-ΔC* cells (Supplementary Fig. 5c), indicating a complex relationship between both factors (see Discussion). We also observed that siRNA-mediated depletion of SPRTN resulted in growth defects in *RNF4* KO cells and led to increased 5-azadC sensitivity (Supplementary Fig. 5d, e). To corroborate these results, we generated U2OS T-REx Flp-In *SPRTN-ΔC* cells by generating frameshift mutations using a single gRNA, which targets the beginning of exon 5 (Supplementary Fig. 5f). In U2OS *SPRTN-ΔC* cells, DNMT1-DPC cleavage was defective (Supplementary Fig. 5g), and depletion of RNF4 caused synthetic lethality/viability defects (Fig. 4e and Supplementary Fig. 5h). Taken together, these results show that SPRTN patient variants compromise replication-independent DPC repair. Moreover, our data indicate that cells can tolerate such reduced repair capacity in principle, but only if proteasomal DPC repair is fully functional.

Compromised ubiquitin binding is the main defect of SPRTN patient variants

Next, we asked why patient variants fail to efficiently cleave DNMT1-DPCs. It has previously been speculated that the major defect of SPRTN-ΔC variants is their mislocalisation to the cytosol due to loss of a C-terminal nuclear localisation signal (NLS)[15,18]. Therefore, we complemented HeLa T-REx Flp-In *SPRTN-ΔC* cells with YFP-tagged SPRTN-ΔC constructs either carrying an additional N-terminal NLS or not (Fig. 5a, top). As expected, SPRTN-ΔC was mislocalised to the cytosol, while NLS-SPRTN-ΔC was found preferentially in the nucleus (Fig. 5a, bottom). Nevertheless, NLS-SPRTN-ΔC was not able to fully restore DNMT1-DPC cleavage in *SPRTN-ΔC* cells (Fig. 5b). Both ΔC variants showed only a slight increase in DPC cleavage, despite being heavily overexpressed and present at much higher levels than SPRTN-WT, which efficiently rescued cleavage (Fig. 5b). Taken together, these results suggest that mislocalisation is not the sole defect of SPRTN-ΔC variants and that the C-terminal part of SPRTN contains an additional critical feature required for replication-independent DPC cleavage. In addition to ensuring nuclear localisation, SPRTN's C-terminal tail contains three protein-protein interaction motifs: a SHP-box (SHP) mediating binding to p97[38,42], a PIP-box (PIP) for interacting with PCNA[43], and a ubiquitin-binding zinc finger (UBZ)[43]. In order to identify the critical domain for DPC cleavage, we complemented *SPRTN-ΔC* cells with SPRTN variants bearing replacements of key amino acids in all three motifs. Expression of SPRTN-WT or PIP*- and SHP*-mutant variants restored DPC cleavage, while SPRTN variants with a defective UBZ domain (D473A - UBZ*) appeared to display reduced cleavage (Supplementary Fig. 6a). To further corroborate that PIP- and SHP-box are dispensable, we complemented *SPRTN-ΔC* cells with a SPRTN variant lacking the entire region between SPRTN's DNA binding domains and the C-terminal NLS and UBZ domain (Fig. 5c, top). Despite lacking both, PIP- and SHP-box, this variant (SPRTN-Δ241-400) fully supported DNMT1-DPC cleavage, unless its UBZ domain was defective as well (SPRTN-Δ241-400-UBZ*) (Fig. 5c, bottom). We conclude that PCNA and p97 binding domains are not required for SPRTN's function in replication-independent DPC repair, while ubiquitin binding appears to be crucial. In addition to recruiting SPRTN to sites of DNA damage, the UBZ domain is also required for stabilising monoubiquitylation of SPRTN (Supplementary Fig. 6a)[38,42,43], which in turn regulates SPRTN autocleavage[37]. To exclude that the loss of monoubiquitylation is causative for the DNMT1-DPC cleavage defects of SPRTN-UBZ*, we tested a linear fusion of ubiquitin to SPRTN-UBZ* (SPRTN-UBZ*-Ub), which we showed previously to restore the regulation of SPRTN autocleavage[37]. However, we observed that this variant remained unable to cleave DNMT1-DPCs (Supplementary Fig. 6b). To further exclude that the reduction of DPC cleavage by SPRTN-UBZ* is a consequence of reduced catalytic activity, we assessed the activity of the recombinant enzyme in vitro, using cleavage of a DPC model substrate (Protein G-oligonucleotide conjugates[23,44]) and autocleavage as a readout. While SPRTN-ΔC showed slightly reduced substrate cleavage and autocleavage, SPRTN-UBZ*'s activity was indistinguishable from the WT enzyme (Supplementary Fig. 6c, d).

Next, we wanted to extend our observations to endogenously expressed SPRTN. We edited the endogenous locus using a gRNA that targets the coding sequence of SPRTN's C-terminal UBZ domain to generate *SPRTN-ΔUBZ* variants. We obtained one *SPRTN-ΔUBZ* clone (#3) with homozygous deletions resulting in premature stop codons. As a consequence, key residues of the UBZ domain are lost (Supplementary Fig. 7a). DNMT1-DPC cleavage was virtually absent in *SPRTN-ΔUBZ* clone #3, confirming that the UBZ domain is critically required for SPRTN's global-genome repair function (Fig. 5d). Sequencing analysis of a second clone (#10) revealed that one allele contained an in-frame deletion resulting in the loss of key UBZ features (#10 Allele 2, Supplementary Fig. 7a). The second allele of clone #10 was identified to bear a premature stop codon, however only downstream of all

Fig. 4 | SPRTN patient variants affect global-genome DNA-protein crosslink repair. a Domain structure of SPRTN wildtype and the C-terminally truncated SPRTN-ΔC variant observed in Ruijs-Aalfs syndrome patients indicating SPRTN's two DNA binding domains (zinc-binding domain, ZBD, and basic region, BR) and interaction motifs/domains for binding to p97 (SHP), PCNA (PIP) and ubiquitin (UBZ) (left). Western blot analysis of HeLa T-REx Flp-In cells genetically-engineered to express patient-like SPRTN-ΔC variants (right). **b, c** 5-azadC-induced DNMT1-DPC repair assessed by PxP in WT and *SPRTN-ΔC* cells. HeLa T-REx Flp-In cells were treated as depicted in Fig. 2g including a pre-treatment with proteasome inhibitor MG132 (**b**) or inhibition of p97 (p97i) (**c**), as indicated. Upon isolation of DPCs by PxP, samples were analysed using western blotting using the indicated antibodies. **d** HeLa T-REx Flp-In *SPRTN-ΔC* cells were complemented with YFP, YFP-SPRTN-Strep variants as indicated (wildtype (WT), E112Q (EQ), R185A (ZBD*)) and treated as shown in Fig. 2a, before DPCs were isolated using PxP and analysed by western blotting using the indicated antibodies. **e** U2OS T-REx Flp-In WT and *SPRTN-ΔC* cells were transfected with the indicated siRNAs. Cell confluency was monitored over 5-days using IncuCyte live cell imaging, (top, values represent the mean ± SD of 3 technical replicates), before cells were stained with crystal violet (bottom). Source data are provided as a Source Data file.

important UBZ residues (#10 Allele 1, Supplementary Fig. 7a). Clone #10 displayed residual SPRTN monoubiquitylation (Fig. 5d), which is consistent with residual UBZ function. In agreement with clone #10 retaining residual levels of ubiquitin binding, we observed minor degrees of DNMT1-DPC cleavage (Fig. 5d). A key role for the UBZ domain was further indicated by CPT-, formaldehyde-, and ETO-induced autocleavage being virtually absent in *SPRTN-ΔUBZ* #3 cells (Supplementary Fig. 7b–d). As observed in *SPRTN-ΔC* cells, RNF4 depletion caused growth defects in HeLa T-REx *SPRTN-ΔUBZ* cells (Supplementary Fig. 7e, f), while resulting in synthetic lethality in U2OS

T-REx *SPRTN-ΔUBZ* cells (Fig. 5e, Supplementary Fig. 7g. We conclude that *SPRTN-ΔUBZ* fully phenocopies the effect of Ruijs-Aalfs syndrome patient variants, suggesting that loss of ubiquitin-binding is the key defect of SPRTN-ΔC.

Discussion

DNA lesions are diverse in nature and are studied using a broad variety of lesion-specific techniques. DPCs have only recently emerged as important endogenous lesions, and the available toolkit to investigate these adducts is therefore limited. We developed a DPC
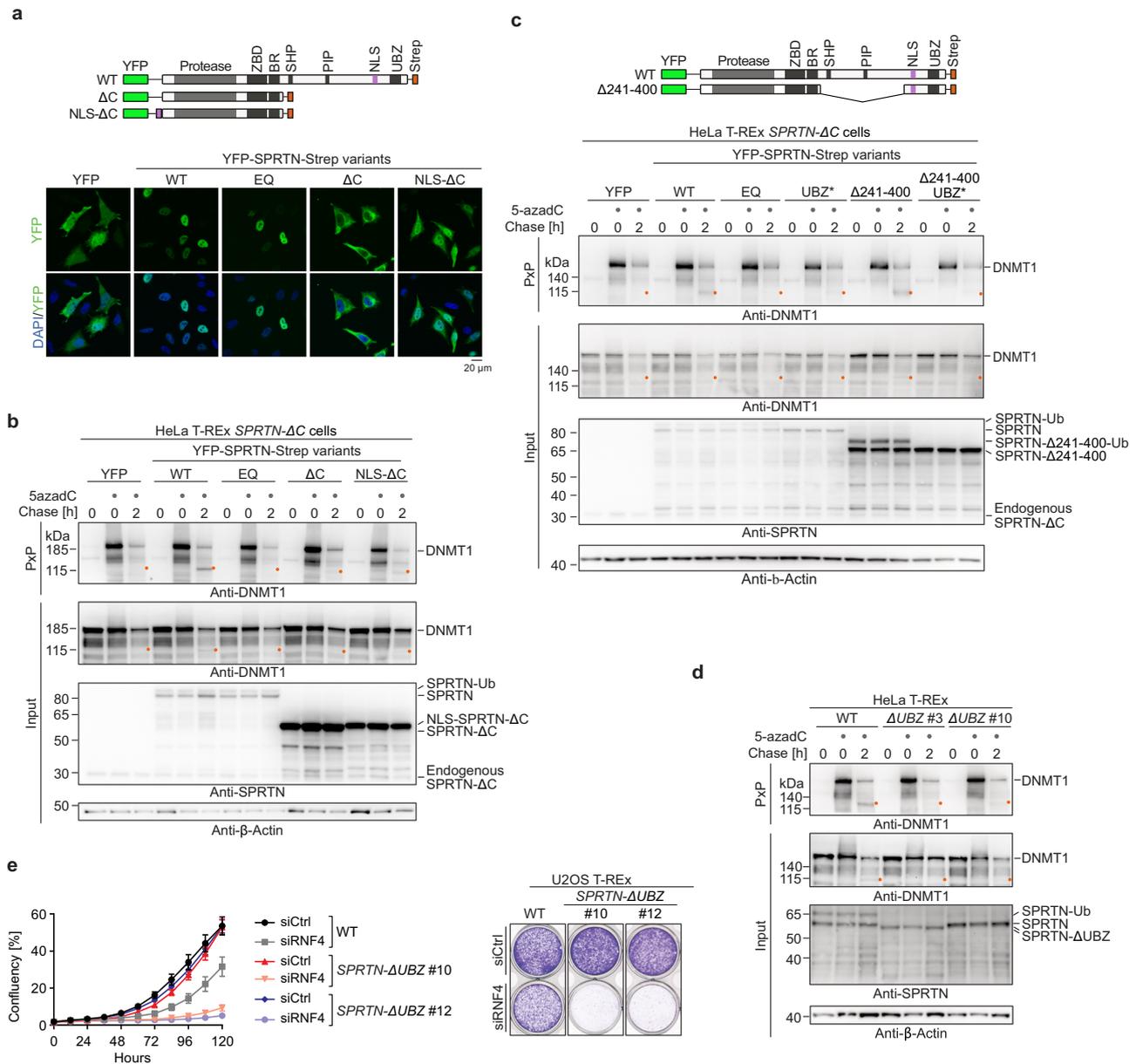
Fig. 5 | Compromised ubiquitin binding is the main defect of SPRTN patient variants. a Schematic depiction of YFP-SPRTN-Strep variants used to complement HeLa T-REx Flp-In *SPRTN-ΔC* cells (top). The localisation of each variant (wildtype (WT), E112Q (EQ), ΔC, and ΔC including an orthogonal N-terminal nuclear localisation signal (NLS-ΔC)) was determined using immunofluorescence staining with anti-GFP antibodies recognizing YFP. Please note that different exposure times are shown for different variants due to different expression levels, see western blot analysis in (**b**). **b** HeLa T-REx Flp-In *SPRTN-ΔC* cells complemented with the indicated SPRTN variants were treated as depicted in Fig. 2a, DPCs were isolated using PxP, and samples were analysed by western blotting using the indicated antibodies. **c** HeLa T-REx Flp-In *SPRTN-ΔC* cells complemented with indicated SPRTN variants (full length or an internally-truncated version lacking amino acids 241–400

(Δ241–400) either WT or in combination with an amino acid replacement within the UBZ domain (D473A, UBZ*), were treated as depicted in Fig. 2a. DPCs were isolated using PxP and analysed by western blotting using the indicated antibodies. **d** Two clonal HeLa T-REx Flp-In *SPRTN-ΔUBZ* cell lines (see Supplementary Fig. 7a) were treated as depicted in Fig. 2a, before isolation of DPCs and analysis by western blotting using the indicated antibodies. **e** U2OS T-REx Flp-In WT and *SPRTN-ΔUBZ* cells were transfected with the indicated siRNAs. Cell confluency was monitored over 5-days using IncuCyte live cell imaging, (left, values represent the mean ± SD of 3 technical replicates), before cells were stained with crystal violet (right). Experiments in (**a**) and (**c**) were performed three times and experiments in (**b**) and (**d**) twice with similar results. Source data are provided as a Source Data file.

extraction method that is compatible with various downstream readouts and is able to detect and identify DPCs in various experimental scenarios. We have combined PxP with quantitative proteomics to reveal that formaldehyde induces less complex DPCs than anticipated. Since formaldehyde is a major source of endogenous DNA damage[8,40], our data indicate that nucleosomal histone-DNA crosslinks are frequent genotoxic challenges faced by mammalian cells.

By studying the repair of 5-azadC-induced DNMT1-DPCs with PxP, we discovered an unexpected role of the SPRTN metalloprotease in replication-independent DPC repair (Fig. 6). Intriguingly, replication-independent DPC cleavage by SPRTN relies on the same initial signals as proteasomal degradation[24,25], namely SUMO-targeted ubiquitylation by RNF4. DPC detection by the SUMO system appears to occur in a global-genome manner that does not rely on transcription or replication machineries to detect the lesion. Despite being replication-
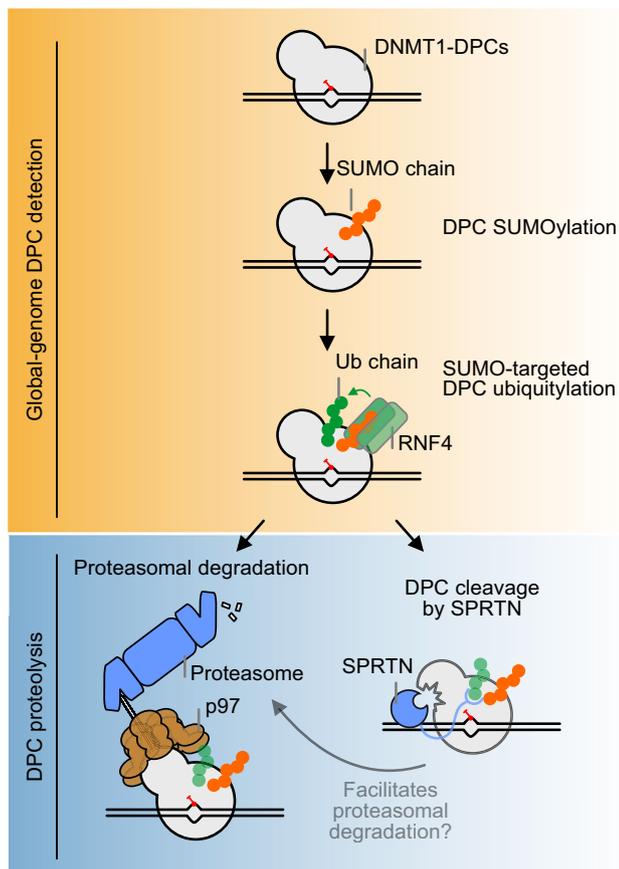
independent, SPRTN-mediated global-genome repair seems restricted to the S/G2 phase, due to low SPRTN expression in G1 cells. The fact that SPRTN-dependent cleavage increases upon proteasome inhibition, suggests that SPRTN acts independently of proteasomal degradation. In addition, we observed that inhibition of the ATPase p97 inhibits proteasomal DPC degradation, while increasing the abundance of the SPRTN-dependent DPC cleavage fragment. p97 has the ability to unfold substrate proteins by threading them through its central pore, which often results in their degradation[35]. Therefore, we propose that SUMO-targeted ubiquitylation results in (a) p97-dependent extraction and subsequent proteasomal degradation or (b) SPRTN-dependent cleavage, perhaps if extraction is inefficient. However, the fate of the DPC fragment produced by SPRTN cleavage remains unclear. It is possible that the cleavage fragment accumulates upon proteasomal inhibition (while it appears only transiently, if proteasome is active), because proteasome and SPRTN are two independent parallel mechanisms that target DNMT1-DPCs. Alternatively, the accumulation of the DNMT1-DPC cleavage fragment upon proteasomal inhibition may indicate that it is itself a substrate for proteasomal degradation. In this hypothetic model, SPRTN cleavage may facilitate proteasomal degradation of DPCs by generating a novel N-terminus, which could trigger additional DPC ubiquitylation by N-end rule E3 ubiquitin ligases. We favour the second scenario because it seems unlikely that the generation of the 115 kDa DNMT-DPC fragment is in itself sufficient for repair.

The fact that SPRTN patient variants displayed compromised global-genome DPC repair may explain the defects observed in non-

Fig. 6 | Model of global-genome DNMT1-DPCs repair. Repair of 5-azadC-induced DNMT1-DPCs is initiated by SUMOylation, followed by subsequent ubiquitylation by the SUMO-targeted ubiquitin ligase RNF4. Modified DNMT1-DPCs are either targeted by p97- and proteasome-dependent degradation or cleaved by SPRTN. The DPC fragment generated by SPRTN cleavage may be subjected to further degradation by p97 and the proteasome.

replicative tissues of Ruijs-Aalfs syndrome patients. PxP in combination with patient-mimicking *SPRTN-ΔC* cells enabled us to conduct a detailed structure-function analysis of critical features within SPRTN for global-genome DPC cleavage. Notably, DPC cleavage was strongly reduced upon mutation of SPRTN's single-stranded DNA binding domain, the ZBD. This observation suggests that DNMT1-DPCs contain a DNA structure with single-stranded or unpaired DNA features, which were shown in vitro to be required for SPRTN activity (2-3 unpaired bases suffice for activation[23]). Interestingly, within the DNMT1-DPC, the 5-azadC base is flipped-out of the DNA duplex into the enzyme's active site, which destabilize the DNA helix locally and result in additional flipped-out bases[45,46]. Structural data indicate that these bases would be accessible for other proteins[46], which would allow SPRTN binding and, thus, activation. Alternatively, DNMT1-DPCs may require pre-processing by a yet to be identified helicase or nuclease activity prior to SPRTN cleavage. The p97-binding motif of SPRTN had no influence on DPC cleavage, which is in line with our observation that p97 activity is only required for proteasomal degradation. In agreement with SPRTN's role in DNMT1-DPC repair being replication-independent, binding to the replication clamp PCNA by SPRTN was also not necessary for activity. In addition to losing p97 and PCNA binding, SPRTN-ΔC variants have three further defects; mislocalisation, reduced DPC cleavage activity, and loss of ubiquitin binding. Importantly however, loss of the UBZ domain alone was sufficient to recapitulate the phenotypes of SPRTN-ΔC suggesting that ubiquitin binding is the critical feature lacking in Ruijs-Aalfs syndrome-associated SPRTN variants. One function of the UBZ domain is the establishment of SPRTN monoubiquitylation[38], which regulates SPRTN autocleavage[37]. However, a linear ubiquitin fusion, which restores autocleavage[37], did not restore DPC cleavage. Therefore, we favour the interpretation that the primary defect of SPRTN-ΔC and SPRTN-ΔUBZ variants is the inability to recognize RNF4-catalyzed DPC ubiquitylation. In support, RNF4 modifies DPCs primarily with K48-linked ubiquitin chains[25], which matches the ability of SPRTN's UBZ to interact with such chains[42].

Our findings that efficient SPRTN-dependent DNMT1-DPC cleavage and formaldehyde- and CPT-induced SPRTN-autocleavage require the presence of RNF4 suggest an epistatic interaction between the two enzymes. Interestingly, however, the genetic relationship between both DPC repair factors is more complex. While SPRTN acts downstream of RNF4 during global-genome repair (this study), it also functions during replication-coupled repair[14,20], which is independent of RNF4[25]. Moreover, in both scenarios SPRTN functions in addition to proteasomal DPC degradation (which is RNF4-dependent outside of replication[24,25]). In agreement, *RNF4* KO cells were 5-azadC sensitive (because SPRTN and proteasomal repair are affected), while *SPRTN-ΔC* cells were not (because proteasomal global-genome repair is still available). In contrast, siRNA-mediated depletion of SPRTN results in increased 5-azadC sensitivity in *RNF4* KO cells, which may reflect a synthetic defect between a reduction in both global-genome DPC repair branches (SPRTN and proteasome) combined with a reduction of SPRTN's replication-coupled repair function. Interestingly, in scenarios of impaired SPRTN function (*SPRTN-ΔC, SPRTN-ΔUBZ*), additional depletion of RNF4 resulted in synthetic viability defects, which are presumably caused by unrepaired endogenous DPCs. Some DPCs may rely more on SPRTN-dependent cleavage than proteasomal degradation, perhaps explaining the differences between 5-azadC sensitivity and viability. Notably, in the absence of RNF4 or upon SPRTN mutation, global-genome cleavage of DPCs was strongly reduced but did still occur. Therefore, the synthetic phenotype may also be caused by simultaneous partial loss-of-function at two critical points of the same pathway.

To conclude, DPCs are not only repaired by replication-coupled mechanisms but are also efficiently targeted by SUMO-dependent global-genome DPC repair mechanisms, that are replication- and transcription-independent. What determines pathway choice during

DPC repair and whether it is linked to genomic context is an exciting open question. Furthermore, whether transcription-coupled DPC repair occurs as well remains to be determined. We anticipate that the PxP methodology will be instrumental to address these key questions on DPC repair.

Methods

Cell lines

HeLa, U2OS T-REx Flp-In and HeLa T-REx Flp-In cells were provided by Cell Services, The Francis Crick Institute, and grown in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% (v/v) fetal bovine serum (FBS). HeLa T-REx Flp-In cells stably expressing siRNA-resistant YFP-SPRTN-Strep-tag and HeLa T-REx Flp-In *SPRTN-ΔC* cells expressing SPRTN variants were generated using the Flp-In system (pOG44, V600520, Thermo Scientific) according to manufacturer's instructions and selected in Hygromycin B (150 μg/ml) (10687010, Thermo Fisher). Protein expression was induced by overnight incubation with doxycycline (D9891, Sigma) (final concentration 1 μg/ml).

Generation of cell lines

Genome-edited cell lines were generated by transfection of pX330 plasmids (#82580, Addgene) encoding the following gRNA sequences: HeLa *RNF4* KO cells (gRNA_RNF4#1 GCTACTCAGAGAAAGCGTCG); U2OS T-REx Flp-In *PIAS4* KO cells (gRNA_PIAS4#1 AGCACGGGGTA GTCAATAT); U2OS T-REx Flp-In *SPRTN-ΔC* cells (gRNA_SPRTN-ΔC#3 ACTAAAAGGGATTACTAGCT); HeLa T-REx Flp-In and U2OS T-REx Flp-In *SPRTN-ΔUBZ* cells (gRNA_SPRTN-ΔUBZ#1 CACTTGGACTGGTGC CTTGA). HeLa T-REx Flp-In *SPRTN-ΔC* cells were generated by co-transfection of two pX330 plasmids containing two different gRNAs (gRNA_SPRTN-ΔC#1 TTGGCAGATAAACCCAACAG and gRNA_SPRTN-ΔC#2 ATTAACCAGAACTTCCTGAC). 16 h after transfection of plasmids using Lipofectamine 2000 (11668030, Thermo Scientific), cells were selected in puromycin-containing (1 μg/ml) media for 48 (HeLa T-REx Flp-In or HeLa cells) or 72 (U2OS T-REx Flp-In cells) hours. Next, cells were seeded in 96 well plates in a concentration of 0.75 cells per well. Single colonies were transferred once confluency was reached and editing efficiency was confirmed by western blotting and Sanger sequencing. Polyclonal HeLa T-REx Flp-In *TEX264* KO cells were generated using two different gRNAs (gRNA_TEX264#1 ATAAGTGCCGA GTGTGCCGT and gRNA_TEX264#2 CTGTGTGCCTATCCTCGGC) with a gRNA targeting the safe-harbour-site *AAVS1* (gRNA_AAVS1#1 GTCC CTAGTGGCCCCACTGT) as control. Editing efficiency of polyclonal pools was confirmed by western blotting following selection and cells were directly used for experiments without selecting single clones. All cell lines generated in this study are available from the corresponding author upon request.

Genotyping of single clones

Genomic DNA of single clones was extracted by lysing cells in 5 mM Tris-HCl pH 8 at 99 °C for 2 min, followed by addition of proteinase K (0.1 mg/ml, 25530049, Invitrogen). Samples were then incubated at 55 °C for 5 h, before proteinase K was heat-inactivated for 45 min at 85 °C. 10 ng of genomic DNA was used as template to amplify the edited region while adding overhangs homologous to the pDONR221 vector (see Supplementary data 3 for primer sequences used for each genotype) using Platinum II Hot-Start Green PCR Master Mix (14001012, Thermo Fischer). Next, PCR products were gel-purified (REF 740611, MACHEREY-NAGEL) and cloned by TEDA-based cloning[47] into a pDONR221 backbone amplified with Q5® Hot Start High-Fidelity 2X Master Mix (M0494S, NEB). Plasmid DNA was isolated from at least five single colonies and analysed by Sanger sequencing.

siRNA transfection

For PxP experiments, cells were transfected in 60 mm dishes. 5 μl siRNA (20 μM) and 12.5 μl Lipofectamine RNAiMAX Transfection Reagent (13778075, Thermo Scientific) were each diluted in 400 μl Opti-MEM Medium. Following a 5 min incubation, siRNA and Lipofectamine RNAiMAX Transfection Reagent dilutions were mixed. After an additional 15 min, the transfection mix was added to cells. After 16 h, cells were reseeded into 60 mm dishes, followed by synchronization using a double thymidine block and PxP extraction 72 h after transfection as described below. For viability and 5-azadC sensitivity assays, siRNA transfections were performed in 6-well plates. 3 μl of siRNA were mixed with 100 μl of Opti-MEM Medium and incubated for 5 min. 7.5 μl of Lipofectamine RNAiMAX were mixed with 100 μl of Opti-MEM Medium and incubated for 5 min. Next, both solutions were mixed, incubated for additional 15 min and added to the well containing 800 μl of media. The following siRNAs (Horizon Discovery) were used: siCTRL (Control pool, D-001810-10-20), siRNF4 (SMARTpool, L-006557-00-0005), siUBC9 (SMARTpool, L-004910-00-0005), siSPRTN#1 (CAAGGAACCAGAGAAUUA) and siPIAS4 (SMARTpool, L-006445-00-0005).

Purification of x-linked proteins (PxP)

DPCs were induced by addition of methanol-free formaldehyde (28906, Fisher Scientific) or camptothecin (CPT, 208925, Sigma) (concentrations indicated in figure legends) to asynchronous cells. For induction of 5-azadC (A3656, Sigma) induced DNMT1-crosslinks, cells were synchronized using a double thymidine block. In brief, cells were seeded in the morning and thymidine-containing media (2 mM, T9250, Sigma) was added after 8 h. The next day, cells were released in thymidine-free medium for 9 h, prior to readdition of thymidine and overnight incubation. Then, cells were released in thymidine-free medium and treated with 5-azadC (10 μM), MG132 (5 μM) (M7449, Sigma), SUMO-E1 inhibitor ML-792 (5 μM) (Axon Medchem, 3109), Ub-E1 inhibitor TAK-243 (1 μM) (AOB87172, Chemietek), aphidicolin (3 μM) (A4487, Sigma) p97i CB-5083 (5 μM) (HY-12861-10mg, Hölzel) or flavopiridol (10 μM) (F3055, Sigma) as indicated in figures.

For PxP, formaldehyde-treated cells were harvested at the respective time points by trypsinisation and counted, while CPT- and 5-azadC-treated cells were scraped in ice-cold PBS (an additional plate per condition was trypsinised and counted to determine the number of cells per plate). For CPT treatments, plates were prechilled on ice for 5 min before scraping to minimize TOP1cc reversal. Next, cells were washed and resuspended in PBS at 2×10^4 cells/μl (cells were optionally pelleted, frozen and stored at −80 °C at this point, apart from CPT-treated cells, which were processed immediately). 10 μl of the cell suspension were directly lysed in 1x NuPAGE LDS sample buffer (NP0007, Thermo Scientific) to serve as input samples. The remaining cell suspension was pre-warmed for 45 s at 45 °C prior to mixing with an equal volume of low melt agarose (2% in PBS, 1613111, Bio-Rad) and immediately cast into plug molds (#1703713, Bio-Rad) with a total volume of ca. 90 μl. Plugs were placed at 4 °C for 5 min, prior to transfer into 1 ml ice-cold lysis buffer (1 x PBS, 0.5 mM EDTA, 2% sarkosyl, cOmplete EDTA-free protease inhibitor cocktail (4693132001, Merck), 0.04 mg/ml Pefabloc SC (11585916001, Merck). Lysis was carried out on a rotating wheel at 4 °C for 4 h. Following lysis, DNA was optionally digested by nuclease. To this end, plugs were transferred to washing buffer (50 mM Tris/HCl pH 8, 0.5 mM MgCl₂, 0.01% sarkosyl). After 10 min, buffer was replaced by fresh washing buffer only or washing buffer containing benzonase nuclease (0.2 U/μl, 70746, Merck Millipore), followed by incubation in a thermoshaker (500 rpm, 37 °C) for 1 h. For electro-elution, plugs were transferred to the wells of 10-well SDS-PAGE gels (12%, 1.5 mm Novex WedgeWell or BOLT gels, ThermoFisher). Electrophoresis was carried out in 300 ml MOPS buffer at 20 mA per gel for 60 min in a Mini Gel Tank (ThermoFisher). Following electro-elution, plugs were retrieved and transferred to tubes containing 1 ml washing buffer, while the gel was stained using InstantBlue (ISB1L, Sigma) to confirm successful extraction of non-crosslinked cellular proteins. Plugs were incubated on a rotating wheel

at 4 °C for 10 min. Plugs of the same conditions were pooled at this stage of the purification (typically two plugs were cast per condition for CPT-, formaldehyde- and 5-azadC-induced DPCs). The supernatant was aspirated and plugs were melted at 99 °C for 5 min, followed by addition of 20 µl washing buffer containing 50 units benzonase nuclease per plug and incubation at 37 °C for 30 min. Samples were then frozen at −80 °C. For analysis by western blotting, NuPAGE LDS sample buffer was added and samples were subjected to western blotting using the indicated antibodies. For silver staining, frozen samples were thawed and centrifuged in a table-top centrifuge at top speed at 4 °C. Supernatant was then passed through 0.45 µM SpinX centrifuge tube filters (CLS8162, Merck) to remove residual agarose. NuPAGE LDS sample buffer was added and samples analysed using SDS-PAGE in a Bolt 12 % 1.5 mm 10-well gel followed by silver staining (SilverQuest Silver Staining Kit, LC6070, ThermoFisher). For analysis by mass spectrometry, plugs were washed twice following electro-elution in washing buffer and fixed in 40% ethanol/10% acetic acid on a rotating wheel at 4 °C for 1 h. Finally, plugs were washed twice in 100 mM ammonium bicarbonate.

Western blotting

Samples were boiled in NuPAGE LDS sample buffer (NP0007, Thermo Scientific) containing NuPAGE Sample Reducing Agent (NP0009, Thermo Scientific), before SDS-PAGE using NuPAGE 4–12% 20 well gels (a 4–12% 12- well Bolt gel was used for Fig. 2b). Following electrophoresis, proteins were transferred on 0.45 µm PVDF membranes (IPVH00010, Merck) using a wet transfer system (#1704070, Bio-Rad) for 70 min at 100 V. Membranes were blocked in 5 % milk in TBS-T for 1 h before addition of primary antibody: Anti-DNMT1 (D63A6) antibody (1:1000) (#5032, Cell Signaling), Anti-Actin antibody (1:1000) (Sc-47778, Santa Cruz Biotechnology), Anti-SUMO2/3 antibody (1:2000) (ab3742, Abcam), Anti-TOP1 antibody (1:1000) (ab109374, Abcam), Anti-Ub antibody (1:1000) (Sc-8017, Santa Cruz Biotechnology), Anti-SPRTN antibody (1:500) (6F2[37]), Anti-RNF4 antibody (1:500) (AF7964, R&D systems), Anti-GAPDH (14C10) antibody (1:2000) (2118, Cell Signaling), Anti-Histone H2A antibody (1:1000) (07-146, Merck), Anti-Histone H2B antibody (1:1000) (10799, Cell Signaling), Anti-Histone H3 antibody (1:1000) (4499 S, Cell Signaling), Anti-PIAS4 antibody (1:500) (SC-166744, Santa Cruz Biotechnology), Anti-Flag (1:2000) (F1804, Sigma-Aldrich) Anti-TEX264 (1:500) (sc-100944, Santa Cruz Biotechnology), Anti-Vinculin (1:1000) (sc-73614, Santa Cruz Biotechnology). Following incubation with primary antibody overnight, membranes were washed with TBS-T and incubated for 1 h with corresponding secondary antibodies (Goat Anti-Mouse Immunoglobulins/HRP, P0447, Dako; Swine Anti-Rabbit Immunoglobulins/HRP, P0399, Dako; Goat Anti-Rat Immunoglobulins/HRP, A9037, Sigma; Rabbit Anti-Goat Immunoglobulins/HRP, A8919, Sigma). To help visualize bands, brigthness and contrast of blots were globally adjusted using ImageLab (Bio-Rad) version 5.2. Uncropped scans of all blots are provided in the Source Data file.

Identification of DNA-protein crosslinks by quantitative proteomics

Agarose plugs were reduced, alkylated and digested with trypsin. The resulting peptides were purified using StageTips and resuspended in 15 µl of 0.1% formic acid solution. For LC-MS/MS purposes, desalted peptides were injected in an Ultimate 3000 RSLCnano system (Thermo) and separated in a 15-cm analytical column (75 µm ID home-packed with ReproSil-Pur C18-AQ 2.4 µm from Dr. Maisch) with a 50-min gradient from 5 to 60% acetonitrile in 0.1% formic acid. The effluent from the HPLC was directly electrosprayed into an LTQ-Orbitrap mass spectrometer XL (Thermo) operated in data dependent mode to automatically switch between full scan MS and MS/MS acquisition. Typical parameters were as follows: survey full scan MS spectra (from m/z 250–1600) were acquired in the Orbitrap with

resolution R = 60,000 at m/z 400 (AGC target of 5×10^5). The three most intense peptide ions with charge states between 2 and 4 were sequentially isolated to a target value of 10,000 and fragmented in the linear ion trap by collision induced dissociation (CID). All fragment ion spectra were recorded in the LTQ part of the instrument. For all measurements with the Orbitrap detector, 3 lock-mass ions from ambient air were used for internal calibration. Typical MS conditions were: spray voltage, 1.5 kV; no sheath and auxiliary gas flow; heated capillary temperature, 200 °C; normalized CID energy 35%; activation q = 0.25; activation time = 30 ms. MaxQuant 1.6.6.0 was used to identify proteins and quantify by iBAQ with the following parameters: Database, Uniprot_UP000005604_Hsapiens_20191107; MS tol, 10ppm; MS/MS tol, 0.5 Da; Peptide FDR, 0.1; Protein FDR, 0.01 Min. peptide Length, 7; Variable modifications, Oxidation (M); Fixed modifications, Carbamidomethyl (C); Peptides for protein quantitation, razor and unique; Min. peptides, 1; Minute. ratio count, 2. To identify significantly enriched proteins, MaxQuant output data were further processes in R. LFQ intensity values were log2 transformed. Missing values were imputed based on a probabilistic dropout function using the proDA R-package setting the untreated benzonase condition as a reference level (Ahlmann-Eltze and Anders, 2020). Proteins that were not identified in at least 3 replicates of either non-benzonase treated condition were removed, if they were simultaneously not detected in more than 12 out of 24 samples. Differential abundance of proteins was calculated using a Wald-test with Benjamini Hochberg FDR correction. Identified proteins were considered significantly enriched if their $\log_2$ fold enrichment was greater than 2 and FDR adjusted p-value smaller than 0.01.

Plasmids and site-directed mutagenesis

pCMV6-RNF4-DDK-Myc was purchased from Origene (#RC207273). pIRES-AcFL was a gift from the Boulton lab. pcDNA5-FRT/TO-YFP-SPRTN-WT-Strep, pcDNA5-FRT/TO-YFP-SPRTN-EQ (E112Q)-Strep pcDNA5-FRT/TO-YFP-SPRTN-ΔC-Strep, pcDNA5-FRT/TO-YFP-SPRTN-PIP*(Y331A/F332A)Strep, pcDNA5-FRT/TO-YFP-SPRTN-SHP*(F253A/L260A)-Strep, pcDNA5-FRT/TO-YFP-SPRTN-ZBD*(R185A)-Strep, pcDNA5-FRT/TO-YFP-SPRTN-UBZ*(D473A)-Strep, pcDNA5-FRT/TO-YFP-SPRTN-UBZ*(D473A)-Ub and pNIC-STREP-ZB-SPRTN-WT have been described previously[13,23,37]. pcDNA5-FRT/TO-YFP-SPRTN-Δ241-400-Strep, pcDNA5-FRT/TO-YFP-SPRTN-Δ241-400-UBZ*-(D473A)-Strep, pcDNA5-FRT/TO-YFP-NLS-SPRTN-ΔC-Strep, pCMV6-RNF4-CS1(C132A/C135A)-DDK-Myc, pNIC-STREP-ZB-SPRTN-UBZ*(D473A), and pNIC-STREP-ZB-SPRTN-ΔC were generated by Q5 site-directed mutagenesis (#E0554S, NEB) according to manufacturer's instructions. siRNA-resistant variants of pcDNA5-FRT/TO-YFP-SPRTN-WT-Strep and pcDNA5-FRT/TO-YFP-SPRTN-EQ (E112Q)-Strep were generated by Q5 site-directed mutagenesis using primers Oshubo-141 (CGAAAACTATTCAAAAAAAGGCAAAGGAAAG) and Oshubo-142 (GGCTCTTTTATTTTTATGTAAGTGCCTCC) introducing silent mutations in the region targeted by siSPRTN#1. All plasmids generated in this study are available from the corresponding author upon request.

Cell viability and drug sensitivity

To measure formaldehyde sensitivity, 500 cells were seeded per well in triplicates in 6-well plates. The next day, cells were treated with the indicated doses formaldehyde for 1 h followed by two washes with PBS. After 7 days, cells were stained with crystal violet.

To measure 5-azadC sensitivity of HeLa WT and *RNF4* KO cells, 5×10^3 cells were seeded in technical quadruplicates in 24-well plates. 5-azadC was added at the indicated concentration 16 h after seeding. After 96 h, cell viability was measured by AlamarBlue assay (Resazurin, R7017, Sigma). To determine complementation of 5-azadC sensitivity, HeLa WT and *RNF4* KO cells were transfected with pIRES-AcFL (expressing GFP-Flag), pCMV6-RNF4-DDK-Myc, or pCMV6-RNF4-CS1-DDK-Myc (CS1, C132A/C135A variant as in[48]) plasmids. 3 µg of plasmid

were mixed with 100 µl of Opti-MEM Medium and incubated for 5 min. 3 µl of Lipofectamine 2000 were mixed with 100 µl of Opti-MEM Medium and incubated for 5 min. After incubation, both solutions were mixed and incubated for additional 15 min. The solution was then added to the cells in 800 µl of media. 5×10^3 transfected cells were then re-seeded in technical quadruplicates in 24-well plates. 5-azadC was added at the indicated concentration 16 h after seeding. After 96 h, cell viability was measured by AlamarBlue assay.

For viability and 5-azadC-sensitivity measurements of HeLa T-Rex Flp-In (WT, *SPRTN-ΔC, AAVS1 #1, SPRTN-ΔUBZ #3 and SPRTN-ΔUBZ #10*), HeLa (WT and *RNF4* KO) cells, and U2OS T-Rex Flp-In (WT, *SPRTN-ΔC* #2, *SPRTN-ΔC* #4, *SPRTN-ΔUBZ* #10 and *SPRTN-ΔUBZ* #12), 2×10^5 cells were seeded in 6-well plates, followed by transfection with siRNAs, as indicated in figure legends. To measure cell viability, transfected cells were re-seeded the following day (5×10^3 cells for HeLa T-Rex Flp-In and HeLa cells; 1×10^4 cells for U2OS T-Rex Flp-In cells) in technical triplicates in 12-well plate. Cell confluency was monitored and analysed using a IncuCyte S3 live cell imaging system every 12 h for 5 days. Following imaging, cells were stained with crystal violet. To measure 5-azadC sensitivity, 5×10^3 cells were re-seeded the day after transfection in technical quadruplicates in 24-well plates. 5-azadC was added at the indicated concentration 16 h after seeding. After 96 h, cell viability was measured by AlamarBlue assay.

DNA and RNA synthesis measurements

1×10^6 cells were seeded in 6-cm dishes. The next day, cells were pre-treated with flavopiridol (10 µM, 1 h), aphidicolin (3 µM, 2 h) or palbociclib (5 µM (Sigma, PZ0383), 48 h), as indicated in figure legends. To measure RNA synthesis, 400 µM 5-ethynyl-uridine (EU, Jena Bioscience, CLK-N002-10) was then added for 30 min. To measure DNA synthesis, 100 µM 5-ethynyl-2′-deoxyuridine (EdU, Jena Bioscience, CLK-N001-100) was added for 30 min. Next, cells were washed twice with PBS, harvested by trypsinization and stained with eFluor780 viability dye (Thermo, 65-0865-14) for 30 min at 4 °C. Next, pellets were washed with PBS containing 1% BSA and fixed with 4% formaldehyde for 15 min. Following another wash, cells were permeabilized by 0.25% Triton-X in PBS, washed, and incubated with Click-it mix (Tris, 39.5 mM, pH8; Alexa Fluor 488-Azide (Thermo, B40953), 0.06 mM; CuSO4, 4 mM; Ascorbic Acid, 40 µg/ml; DAPI 1 µg/ml) for 30 min. After a final wash step, cells were resuspended in PBS containing 1% BSA and analyzed on a BD LSRFortessa (BD Bioscience) equipped with 355/405/488/561/640 nm lasers with a minimum count of 10,000 events. The mean fluorescence intensity (MFI) in FITC channel was measured excluding dead cells (eFluor780 stained) and the results were analyzed using FlowJo™ v10.8.1 Software (BD Life Sciences). In brief, cell debris and aggregates were excluded by SSC-A/FSC-A and single cells were gated by SSC-H/SSC-A. Live cells were identified by SSC-A/APC-Cy7-A and the fraction of EdU/EU positive cells was identified based on FITC-A fluorescence intensity (Supplementary Fig. 8).

Immunofluorescence staining

Doxycycline (1 µg/ml, D9891, Sigma) was added overnight to HeLa-TREx Flp-in cells to induce expression of SPRTN variants. The next day, cells were fixed in 4% formaldehyde (28906, Thermo Scientific) followed by permeabilizing and blocking with PBGT buffer (1X PBS, 0.2% fish skin gelatin, 0.5% BSA, 0.5% Triton X-100) (45 min at room temperature) and then incubated with anti-GFP antibody (Chromotek, PABG1) for 1 h at room temperature. Coverslips were washed 3 times for 5 min with PBGT buffer and incubated with Alexa Fluor 488 goat anti-mouse secondary antibody (A-11001, Thermo Scientific) and DAPI (0.5 µg/ml, 62248, Thermo Fisher) for 1 h at room temperature. Coverslips were mounted in Prolong Gold Antifade Mountant (P10144, Thermo Fisher) and images were acquired using a ZEISS LSM710 confocal microscope and software ZEN 2009 (Carl

Zeiss) version 5.5.0.443. Image processing was done using ImageJ (v1.53t).

Recombinant protein purification

Recombinant SPRTN (WT, EQ (E112Q), UBZ* (D473A)) protein was expressed in *E. coli BL21* and purified as previously described[23]. The protocol was slightly modified for the purification of SPRTN-ΔC, for which the N-terminal Strep-Zb-tag was removed using a Strep-tagged TEV protease. Tag and TEV protease were removed by applying the sample to Strep-Tactin®XT Superflow® high capacity cartridges, before the collected flow-through was further purified by size exclusion chromatography.

In vitro SPRTN autocleavage

Reactions were performed for 2 h at 25 °C in 20 µL containing 2 µM recombinant SPRTN and 11.14 nM circular single-stranded DNA (ΦX174 Virion DNA, #N3023, NEB). The reaction buffer comprised 19.5 mM HEPES/KOH pH 7.2, 2.9% glycerol, 80 mM KCl, and 4.95 mM TCEP. Reactions were stopped by 4 x LDS sample buffer supplemented with 5% β-mercaptoethanol and boiling at 95 °C for 10 min, and resolved on SDS-PAGE gels (4–12% Bis-Tris) using MOPS buffer and stained with SYPRO Ruby (#S12000, Thermo Fisher Scientific) following manufacturer's instructions. Gels were photographed using a BioRad Chemidoc MP system and cleavage was quantified using ImageJ (v1.53t). The fraction of cleaved recombinant SPRTN was calculated by dividing the amount of remaining full-length protein in the presence of DNA by the amount of full-length protein in the absence of DNA.

SPRTN autocleavage in cells

1×10^5 cells per well were seeded in 12-well plates in the evening. The next morning cells were pre-treated or not with aphidicolin (3 µM) for 2 h and then in combination with formaldehyde (250 µM), CPT (500 nM) or etoposide (50 µM, Sigma, 341205). In the indicated timepoint, cells were washed with PBS 1x and resuspended in 1x LDS. The samples were then boiled and resolved in 4–12% 20-well gels. For palbociclib treatment, cells were seeded in 10 cm dishes in the presence of palbociclib (5 µM), reseeded in palbociclib-containing medium after 30 h in 12-well plates, before autocleavage was induced as described above.

Protein G-oligonucleotide conjugation

Protein G-oligonucleotide conjugates were generated as previously described[23,44]. In brief, Protein G (#6510, BioVision) was conjugated to oligonucleotide X15 (5′−6-FAM-ACC AGT GCC TTG CT[SH-C9-dT] GGA CAT CTT TGC CCA-3′) (Ella BioTech GmbH), which contained a 6-FAM label at the 5′-end and a phosphate group at the 3′-end. Conjugation was performed using the proFIRE Amine Coupling Kit (Dynamic Biosensors) and the conjugate purified with the proFIRE device (Dynamic Biosensors) through ion exchange chromatography. The conjugate concentration was determined by measuring 6-FAM fluorescence in a Tecan Spark plate reader using a NanoQuant plate. For Protein G-oligonucleotide conjugate cleavage assays, conjugates were annealed to a 2x excess of complementary reverse oligonucleotide oDY_72 (5′-TGGGCAAAGATGTCC-3′) forming a single-/double-stranded DNA junction.

Protein G-oligonucleotide conjugate cleavage assay

Cleavage of Protein G-oligonucleotide conjugates by SPRTN was performed in a reaction containing increasing concentrations of SPRTN (2, 10 and 50 nM) and 10 nM conjugate (or free DNA as control) in a final reaction buffer of 17.5 mM HEPES/KOH pH 7.2, 85 mM KCl, 3.5% glycerol, 5.5 mM TCEP and 0.12 mg/mL BSA. Reactions were incubated for 2 h at 25 °C. Unstained urea loading dye (15% Ficoll®, 8 M Urea) was added and reactions were resolved on 8 M Urea, 15% Acrylamide, 1x TBE gels using 1x TBE as running buffer. Gels were photographed using

a BioRad Chemidoc MP system and cleavage was quantified using ImageJ (v1.53t) by calculating the ratio between cleaved and total conjugate.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Mass spectrometry data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD026654[49].

Database Uniprot_UP000005604_Hsapiens_20191107 was used to identify proteins. Source data are provided with this paper. All other data that support this study are available from the corresponding author upon request. Source data are provided with this paper.

References

1. Schumacher, B., Pothof, J., Vijg, J. & Hoeijmakers, J. H. J. The central role of DNA damage in the ageing process. *Nature* **592**, 695–703 (2021).
2. Jackson, S. P. & Bartek, J. The DNA-damage response in human biology and disease. *Nature* **461**, 1071–1078 (2009).
3. van den Heuvel, D., van der Weegen, Y., Boer, D. E. C., Ogi, T. & Luijsterburg, M. S. Transcription-coupled DNA repair: from mechanism to human disorder. *Trends Cell Biol.* **31**, 359–371 (2021).
4. Cortez, D. Replication-coupled DNA repair. *Mol. Cell* **74**, 866–876 (2019).
5. 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**, 465–481 (2014).
6. Weickert, P. & Stingele, J. DNA-protein crosslinks and their resolution. *Annu. Rev. Biochem.* **91**, 157–181 (2022).
7. Stingele, J., Habermann, B. & Jentsch, S. DNA-protein crosslink repair: proteases as DNA repair enzymes. *Trends Biochem Sci.* **40**, 67–71 (2015).
8. Morellato, A. E., Umansky, C. & Pontel, L. B. The toxic side of one-carbon metabolism and epigenetics. *Redox Biol.* **40**, 101850 (2021).
9. Stingele, J., Bellelli, R. & Boulton, S. J. Mechanisms of DNA-protein crosslink repair. *Nat. Rev. Mol. Cell Biol.* **18**, 563–573 (2017).
10. Christman, J. K. 5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. *Oncogene* **21**, 5483–5495 (2002).
11. Maslov, A. Y. et al. 5-aza-2'-deoxycytidine-induced genome rearrangements are mediated by DNMT1. *Oncogene* **31**, 5172–5179 (2012).
12. Stingele, J., Schwarz, M. S., Bloemeke, N., Wolf, P. G. & Jentsch, S. A DNA-dependent protease involved in DNA-protein crosslink repair. *Cell* **158**, 327–338 (2014).
13. Stingele, J. et al. Mechanism and regulation of DNA-protein crosslink repair by the DNA-dependent metalloprotease SPRTN. *Mol. Cell* **64**, 688–703 (2016).
14. Vaz, B. et al. Metalloprotease SPRTN/DVC1 orchestrates replication-coupled DNA-protein crosslink repair. *Mol. Cell* **64**, 704–719 (2016).
15. Lopez-Mosqueda, J. et al. SPRTN is a mammalian DNA-binding metalloprotease that resolves DNA-protein crosslinks. *Elife* **5**, e21491 (2016).
16. Maskey, R. S. et al. Spartan deficiency causes accumulation of Topoisomerase 1 cleavage complexes and tumorigenesis. *Nucleic Acids Res.* **45**, 4564–4576 (2017).
17. Morocz, M. et al. DNA-dependent protease activity of human Spartan facilitates replication of DNA-protein crosslink-containing DNA. *Nucleic Acids Res.* **45**, 3172–3188 (2017).
18. Lessel, D. et al. Mutations in SPRTN cause early onset hepatocellular carcinoma, genomic instability and progeroid features. *Nat. Genet* **46**, 1239–1244 (2014).
19. Maskey, R. S. et al. Spartan deficiency causes genomic instability and progeroid phenotypes. *Nat. Commun.* **5**, 5744 (2014).
20. Larsen, N. B. et al. Replication-coupled DNA-protein crosslink repair by SPRTN and the proteasome in xenopus egg extracts. *Mol. Cell* **73**, 574–588.e7 (2019).
21. Sparks, J. L. et al. The CMG helicase bypasses DNA-protein crosslinks to facilitate their repair. *Cell* **176**, 167–181.e21 (2019).
22. 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).
23. Reinking, H. K. et al. DNA structure-specific cleavage of DNA-protein crosslinks by the SPRTN protease. *Mol. Cell* **80**, 102–113 e6 (2020).
24. Sun, Y. et al. A conserved SUMO pathway repairs topoisomerase DNA-protein cross-links by engaging ubiquitin-mediated proteasomal degradation. *Sci. Adv.* **6**, eaba6290 (2020).
25. Liu, J. C. Y. et al. Mechanism and function of DNA replication-independent DNA-protein crosslink repair via the SUMO-RNF4 pathway. *EMBO J.* **40**, e107413 (2021).
26. Ruggiano, A. et al. The protease SPRTN and SUMOylation coordinate DNA-protein crosslink repair to prevent genome instability. *Cell Rep.* **37**, 110080 (2021).
27. Serbyn, N. et al. SUMO orchestrates multiple alternative DNA-protein crosslink repair pathways. *Cell Rep.* **37**, 110034 (2021).
28. Borgermann, N. et al. SUMOylation promotes protective responses to DNA-protein crosslinks. *EMBO J.* **38**, e101496 (2019).
29. Gueranger, Q., Kia, A., Frith, D. & Karran, P. Crosslinking of DNA repair and replication proteins to DNA in cells treated with 6-thioguanine and UVA. *Nucleic Acids Res.* **39**, 5057–5066 (2011).
30. Subramanian, D., Furbee, C. S. & Muller, M. T. ICE bioassay. Isolating in vivo complexes of enzyme to DNA. *Methods Mol. Biol.* **95**, 137–147 (2001).
31. Zhitkovich, A. & Costa, M. A simple, sensitive assay to detect DNA-protein crosslinks in intact cells and in vivo. *Carcinogenesis* **13**, 1485–1489 (1992).
32. Hu, Q. et al. The ARK assay is a sensitive and versatile method for the global detection of DNA-protein crosslinks. *Cell Rep.* **30**, 1235–1245 e4 (2020).
33. Kiianitsa, K. & Maizels, N. A rapid and sensitive assay for DNA-protein covalent complexes in living cells. *Nucleic Acids Res.* **41**, e104 (2013).
34. Wilhelm, L. et al. SMC condensin entraps chromosomal DNA by an ATP hydrolysis dependent loading mechanism in Bacillus subtilis. *Elife* **4**, e06659 (2015).
35. van den Boom, J. & Meyer, H. VCP/p97-mediated unfolding as a principle in protein homeostasis and signaling. *Mol. Cell* **69**, 182–194 (2018).
36. Fielden, J. et al. TEX264 coordinates p97- and SPRTN-mediated resolution of topoisomerase 1-DNA adducts. *Nat. Commun.* **11**, 1274 (2020).
37. Zhao, S. et al. A ubiquitin switch controls autocatalytic inactivation of the DNA-protein crosslink repair protease SPRTN. *Nucleic Acids Res.* **49**, 902–915 (2021).
38. Mosbech, A. et al. DVC1 (C1orf124) is a DNA damage-targeting p97 adaptor that promotes ubiquitin-dependent responses to replication blocks. *Nat. Struct. Mol. Biol.* **19**, 1084–1092 (2012).
39. Ruijs, M. W. et al. Atypical progeroid syndrome: an unknown helicase gene defect? *Am. J. Med Genet A* **116A**, 295–299 (2003).

40. 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).

41. Li, F., Raczynska, J. E., Chen, Z. & Yu, H. Structural insight into DNA-dependent activation of human metalloprotease spartan. *Cell Rep.* **26**, 3336–3346.e4 (2019).

42. Davis, E. J. et al. DVC1 (C1orf124) recruits the p97 protein segregase to sites of DNA damage. *Nat. Struct. Mol. Biol.* **19**, 1093–1100 (2012).

43. Centore, R. C., Yazinski, S. A., Tse, A. & Zou, L. Spartan/C1orf124, a reader of PCNA ubiquitylation and a regulator of UV-induced DNA damage response. *Mol. Cell* **46**, 625–635 (2012).

44. Reinking, H. K. & Stingele, J. Protein-oligonucleotide conjugates as model substrates for DNA-protein crosslink repair proteases. *STAR Protoc.* **2**, 100591 (2021).

45. Adam, S. et al. DNA sequence-dependent activity and base flipping mechanisms of DNMT1 regulate genome-wide DNA methylation. *Nat. Commun.* **11**, 3723 (2020).

46. Ren, W., Gao, L. & Song, J. Structural basis of DNMT1 and DNMT3A-mediated DNA methylation. *Genes (Basel)* **9**, 620 (2018).

47. Xia, Y. et al. T5 exonuclease-dependent assembly offers a low-cost method for efficient cloning and site-directed mutagenesis. *Nucleic Acids Res.* **47**, e15 (2019).

48. Tatham, M. H. et al. RNF4 is a poly-SUMO-specific E3 ubiquitin ligase required for arsenic-induced PML degradation. *Nat. Cell Biol.* **10**, 538–546 (2008).

49. Perez-Riverol, Y. et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450 (2019).

Acknowledgements

Author contributions

Conceptualization: P.W., H.L., and J.S. Investigation: P.W., H.L., M.J.G., D.Y., S.Z., S.D., A.C.A., I.F., A.I., J.C., and J.S. Writing – Original Draft: J.S. and P.W. Writing – Review & Editing: H.L., P.W., and J.S. Funding Acquisition and Supervision: J.S.

Funding

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-023-35988-1.

Correspondence and requests for materials should be addressed to Julian Stingele.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary Information

SPRTN patient variants cause global-genome DNA-protein crosslink repair defects

Pedro Weickert[1,2,4], Hao-Yi Li[1,2,4], Maximilian J. Götz[1,2], Sophie Dürauer[1,2], Denitsa Yaneva[1,2], Shubo Zhao[1,2], Jacqueline Cordes[1,2], Aleida C. Acampora[1,2], Ignasi Forne[3], Axel Imhof[3], and Julian Stingele[1,2,*]

[1]Department of Biochemistry, Ludwig-Maximilians-University, 81377 Munich, Germany.

[2]Gene Center, Ludwig-Maximilians-University, 81377 Munich, Germany.

[3]Protein Analysis Unit (ZfP), BioMedical Center (BMC), Ludwig-Maximilians-University, 82152 Martinsried, Germany.

[4]These authors contributed equally: Pedro Weickert, Hao-Yi Li


*Correspondence: stingele@genzentrum.lmu.de

Extended Data Fig. 1. Analysis of DNA-protein crosslink formation by Purification of x-linked Proteins (PxP) assay. a InstantBlue-stained SDS-PAGE gel used for electro-elution of camptothecin (CPT)-treated PxP samples shown in Fig. 1b. Two plugs per condition were used. **b** InstantBlue-stained SDS-PAGE gel used for electro-elution of formaldehyde (FA)-treated PxP samples shown in Fig. 1c. Two plugs per condition were used. **c** HeLa cells were treated with the indicated concentrations of FA for 1 hour, followed by growth in drug-free medium for 7 days and staining by crystal violet. Source data are provided as a Source Data file.
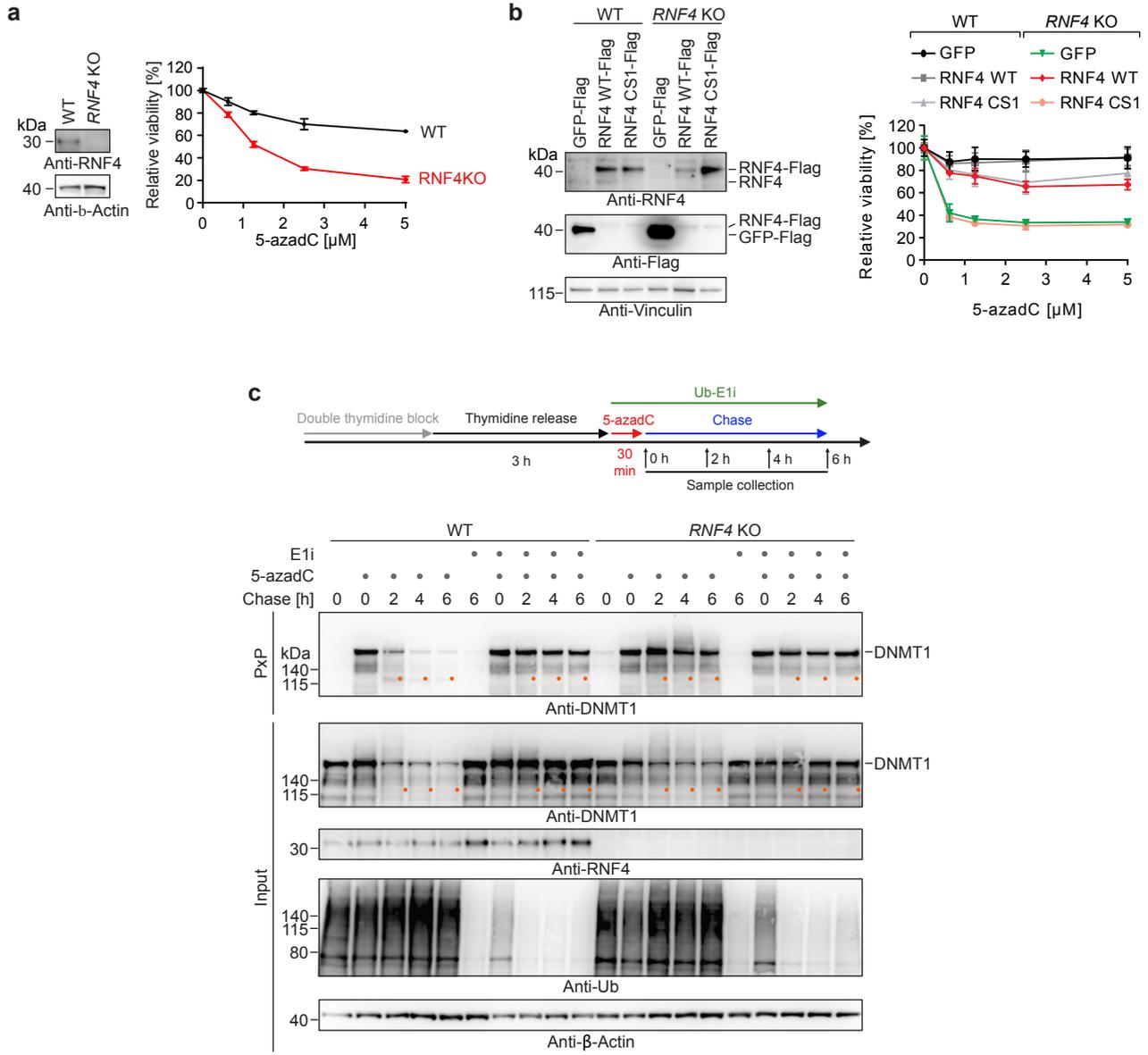
Extended Data Figure 1

a

CPT [µM]   -   0.1   1   10

b

            NT   FA

Nuclease   • •   • •

c

FA [mM]   0   0.1   0.25   0.5   1
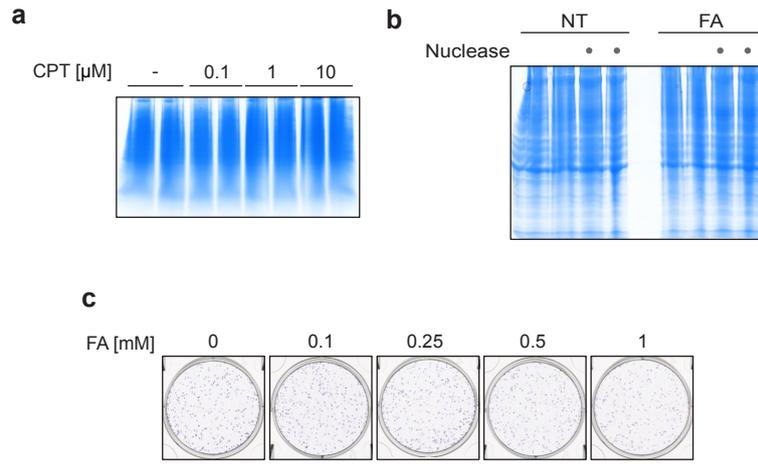
Extended Data Fig. 2. Analysis of 5-azadC-induced DNMT1-DPC repair by PxP assay. a HeLa T-REx Flp-In cells were treated as depicted in Fig. 2a and harvested directly after 5-azadC exposure. DPCs were isolated using PxP including a nuclease treatment prior to electro-elution as indicated, and analysed by western blotting using the indicated antibodies. This control experiment was performed once. **b** InstantBlue-stained SDS-PAGE gel used for electro-elution of 5-azadC-treated PxP samples shown in (**a**). Two plugs per condition were used. **c** HeLa T-REx Flp-In cells transfected with the indicated siRNAs were treated and analysed as depicted in Fig. 2a. DPCs were isolated using PxP and analysed by western blotting using the indicated antibodies. **d** HeLa T-REx Flp-In cells were treated as in Fig. 2e with the difference that ubiquitylation inhibitor (Ub-E1i TAK-243, 1 µM) was added already 1 hour prior to addition of 5-azadC (top). DPCs were isolated using PxP and analysed by western blotting using the indicated antibodies (bottom). **e** 5-azadC-induced DNMT1-DPC formation and repair upon inhibition of p97 (p97i CB-5083, 5µM). HeLa T-REx Flp-In cells were treated as depicted in Fig. 2a prior to extraction of DPCs using PxP and analysis of samples using western blotting using the indicated antibodies. **f** HeLa T-REx Flp-In cells transfected with the indicated siRNAs were treated and analysed as depicted in Fig. 2a. DPCs were isolated using PxP and analysed by western blotting using the indicated antibodies. **g** U2OS T-REx Flp-In WT and *PIAS4* knock-out (KO) cells were treated and analysed as shown in Fig. 2a. DPCs were isolated by PxP and analysed by western blotting using the indicated antibodies. Experiments were repeated three (**c-f**) or two **(g)** times with similar results. Source data are provided as a Source Data file.
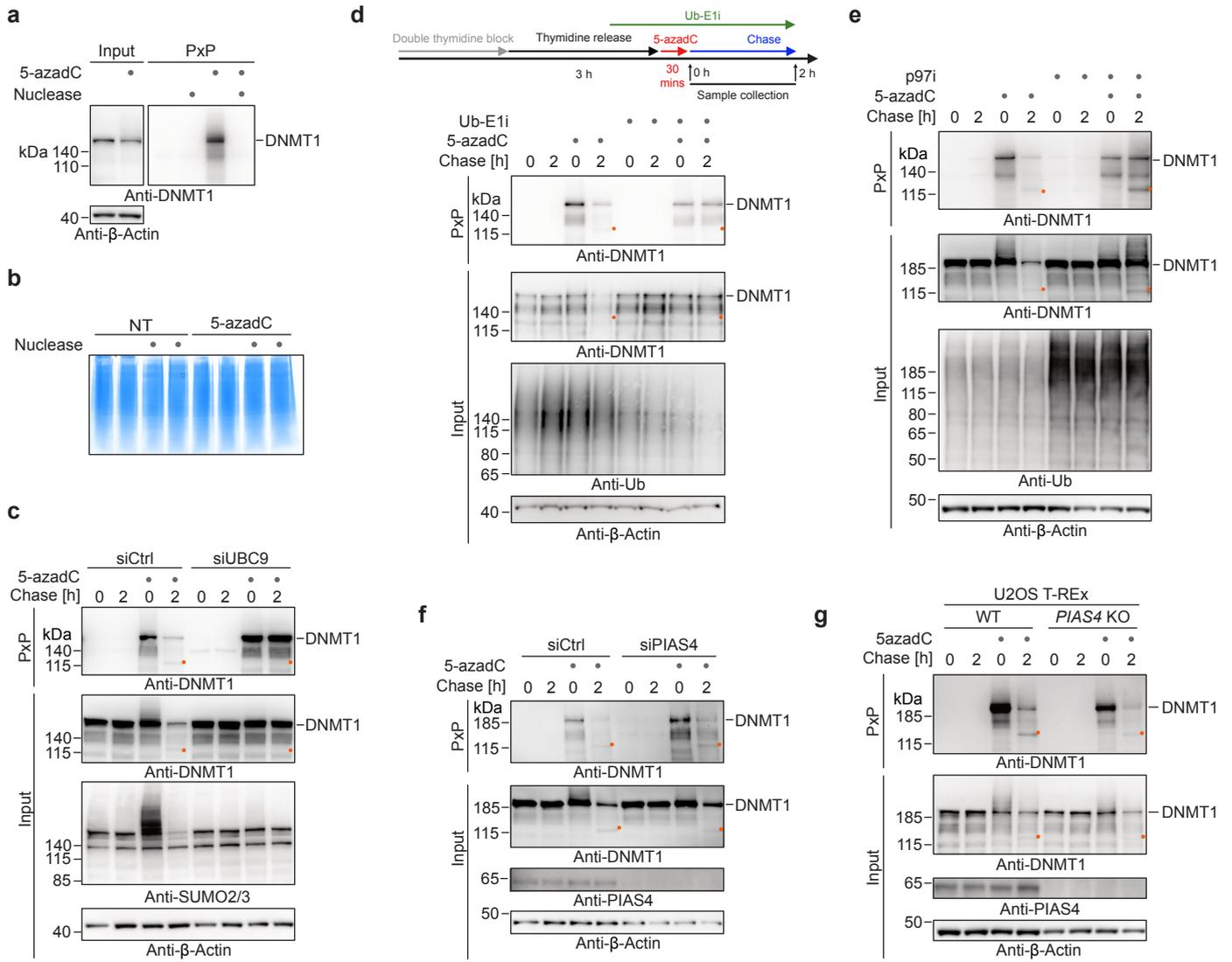
Extended Data Fig. 3. Analysis of 5-azadC-induced DNMT1-DPC repair upon RNF4 loss. a Western blot analysis of HeLa WT and *RNF4* knock-out (KO) cells (left). Cells were treated with increasing concentrations of 5-azadC for 96 hours prior to assessment of viability by AlamarBlue assay. Values represent the mean ± SD of 4 technical replicates normalized to untreated cells (right). **b** Western blot analysis of HeLa WT and *RNF4* KO cells transfected with plasmids encoding AcGFP-Flag or C-terminally Flag-tagged RNF4 (wildtype (WT) or catalytically inactive variant (CS1, C132A/C135A)) (left). Cells were treated with increasing concentrations of 5-azadC for 96 hours prior to assessment of viability by AlamarBlue assay. Values represent the mean ± SD of 4 technical replicates normalized to untreated cells (right). **c** HeLa WT and *RNF4* KO cells were treated and analysed as depicted including an optional treatment with ubiquitylation inhibitor (Ub-E1i TAK-243, 1 μM) (top), prior to extraction of DPCs using PxP and analysis of samples by western blotting using the indicated antibodies (bottom). The experiment was repeated three times with similar results. Source data are provided as a Source Data file.
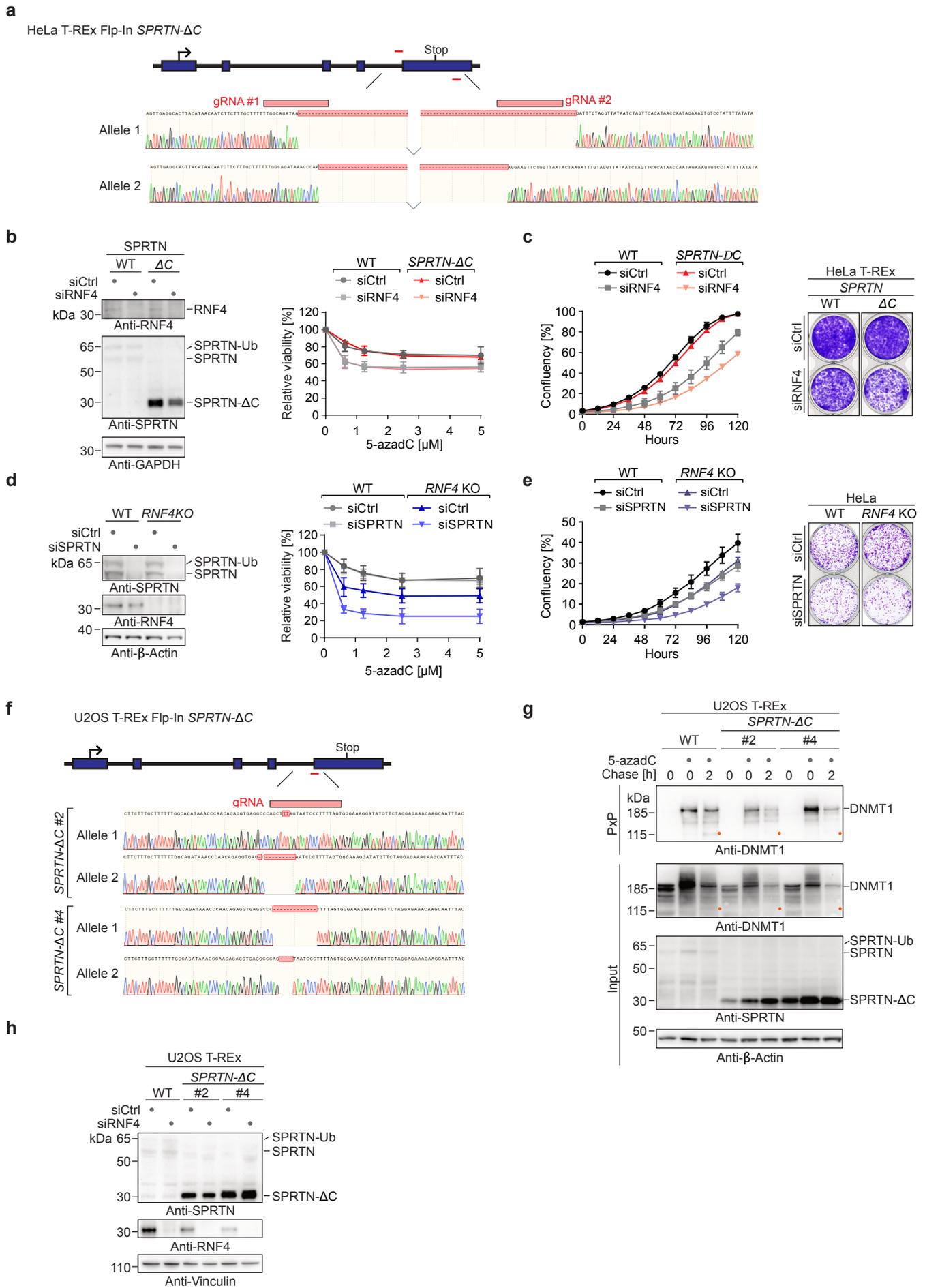
Extended Data Figure 3

a

kDa
30 — Anti-RNF4
40 — Anti-b-Actin

WT | RNF4 KO

b

WT | RNF4 KO

kDa
40 — RNF4-Flag / RNF4

Anti-RNF4

40 — RNF4-Flag / GFP-Flag

Anti-Flag

115 — Anti-Vinculin

WT — GFP / RNF4 WT / RNF4 CS1
RNF4 KO — GFP / RNF4 WT / RNF4 CS1

5-azadC [μM]

c

Double thymidine block — Thymidine release — 5-azadC — Chase — Ub-E1i

3 h — 30 min — 0 h / 2 h / 4 h / 6 h — Sample collection

WT | RNF4 KO

E1i
5-azadC
Chase [h]: 0 0 2 4 6 6 0 2 4 6 0 0 2 4 6 6 0 2 4 6

PxP
kDa
140
115
Anti-DNMT1 — DNMT1

Input
140
115
Anti-DNMT1 — DNMT1

30 — Anti-RNF4

140
115
80
Anti-Ub

40 — Anti-β-Actin

Extended Data Fig. 4. Analysis of DPC-induced SPRTN activation. a HeLa T-REx Flp-In cells were treated with aphidicolin (3 µM) for 2 hours and then incubated with EdU-containing medium for 30 min. EdU incorporation was quantified by flow cytometry. **b** Polyclonal pools of HeLa T-REx Flp-In cells transfected with Cas9 and sgRNAs targeting the safe-harbour site AAVS1 or the TEX264 locus were treated as depicted in Fig. 2a. DPCs were isolated using PxP and analysed by western blotting using the indicated antibodies. The experiment was repeated three times with similar results. **c-d** 5-azadC-induced DNMT1-DPC repair upon inhibition of transcription assessed by PxP. HeLa T-REx Flp-In cells were treated as depicted including an addition of flavopiridol (10 µM), as indicated. EU incorporation (30 min) was quantified by flow cytometry after a 1-hour flavopiridol pre-treatment (**c**). DNMT1-DPCs were isolated using PxP and analysed by western blotting using the indicated antibodies (**d**). The experiment was repeated twice with similar results. **e** HeLa cells were treated with aphidicolin (3 µM) for 2 hours and then incubated with EdU-containing medium for 30 min. EdU incorporation was quantified by flow cytometry. **f** U2OS T-REx Flp-In cells were treated with palbociclib (5 µM) for 48 hours and then incubated with EdU containing medium for 30 min. EdU incorporation was quantified by flow cytometry. **g-i** U2OS T-REx Flp-In cells were treated with palbociclib (5 µM) for 48 hours and then treated with FA (250 µM) (**g**), CPT (500 nM) (**h**) or ETO (50 µM) (**i**) as indicated, before SPRTN autocleavage was assessed by Western blotting. Experiments shown in (**g-i**) were repeated three times with similar results. Source data are provided as a Source Data file.

Extended Data Figure 4

a

HeLa T-REx Flp-In cells

b

c

d

e

HeLa WT cells

f

U2OS T-REx Flp-In cells

g

h

i
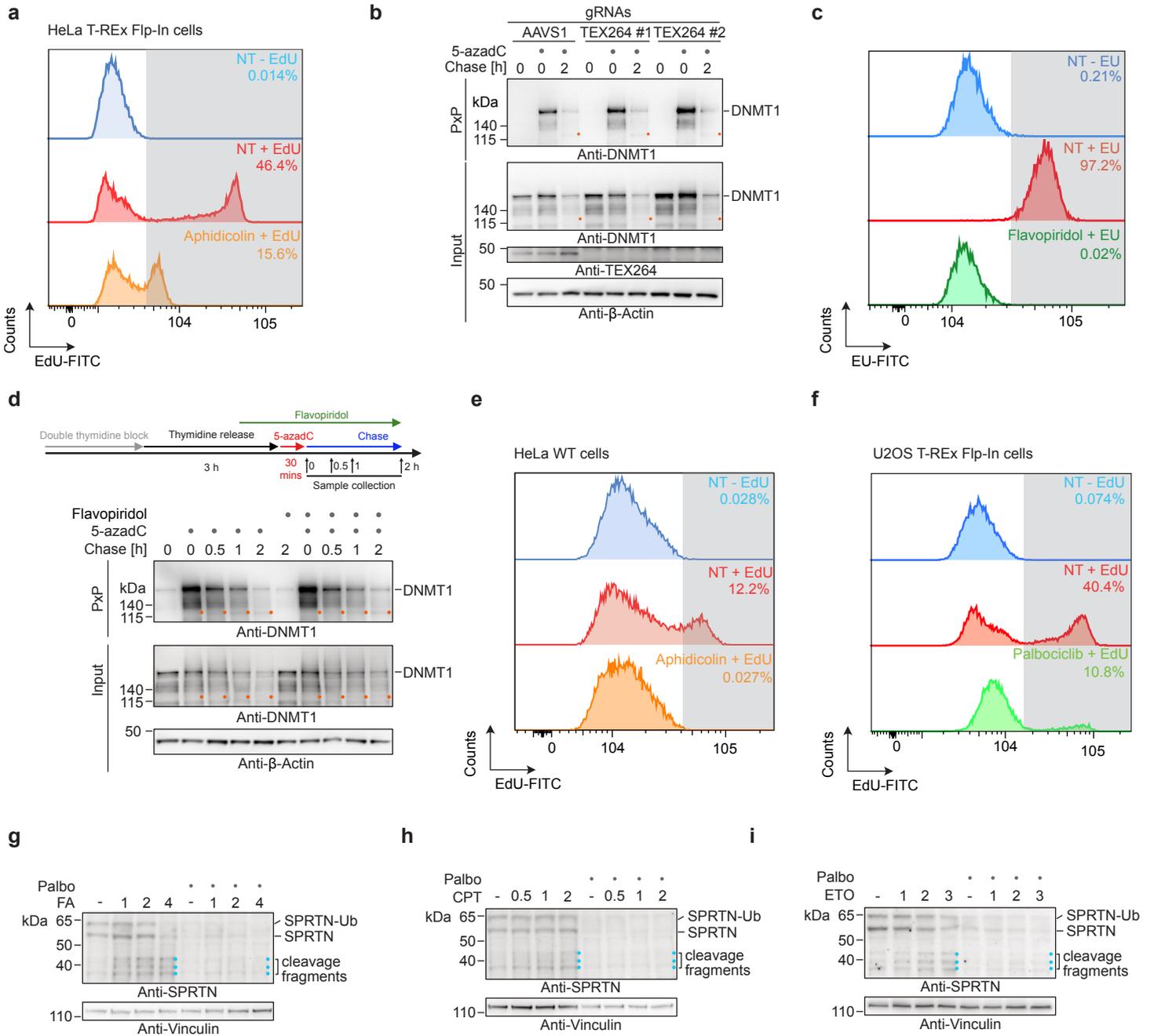
Extended Data Fig. 5. Generation and characterization of *SPRTN-ΔC* cells. a Depiction of the editing strategy for the generation of HeLa T-REx Flp-In *SPRTN-ΔC* cells and Sanger sequencing traces of both edited alleles. **b-e** 5-azadC sensitivity and viability of HeLa T-REx Flp-In WT, *SPRTN-ΔC*, or HeLa WT, *RNF4* knock-out (KO) cells were analysed 72 hours after transfection with the indicated siRNAs. Western blot analysis of HeLa T-REx Flp-In WT, *SPRTN-ΔC* (**b**), or *RNF4* KO cells (**d**) transfected with the indicated siRNAs (left). Transfected cells were exposed to increasing concentrations of 5-azadC for 96 hours, prior to assessment of cell viability using AlamarBlue assay (right). Values represent the mean ± SD of 3 biological replicates normalized to untreated cells. HeLa T-REx Flp-In WT, *SPRTN-ΔC* (**c**), or HeLa WT, *RNF4* KO cells (**e**) were transfected with the indicated siRNAs and cell confluency was monitored over 5 days by IncuCyte live cell imaging every 12 hours (left, values represent the mean ± SD of 3 technical replicates), before cells were stained with crystal violet (right). **f** Depiction of the editing strategy for the generation of U2OS T-REx Flp-In *SPRTN-ΔC* cells and Sanger sequencing traces of edited alleles of two different clones. **g** U2OS T-REx Flp-In WT and *SPRTN-ΔC* cells were treated as depicted in Fig. 2a but with a 1-hour 5-azadC incorporation, before DPCs were isolated by PxP and analysed by western blotting using the indicated antibodies. The experiment was repeated twice with similar results. **h** Western blot analysis of U2OS T-REx Flp-In WT and *SPRTN-ΔC* cells transfected with the indicated siRNAs used for viability assays shown in Fig. 4e. Knock-down efficiency was confirmed twice with similar results. Source data are provided as a Source Data file.
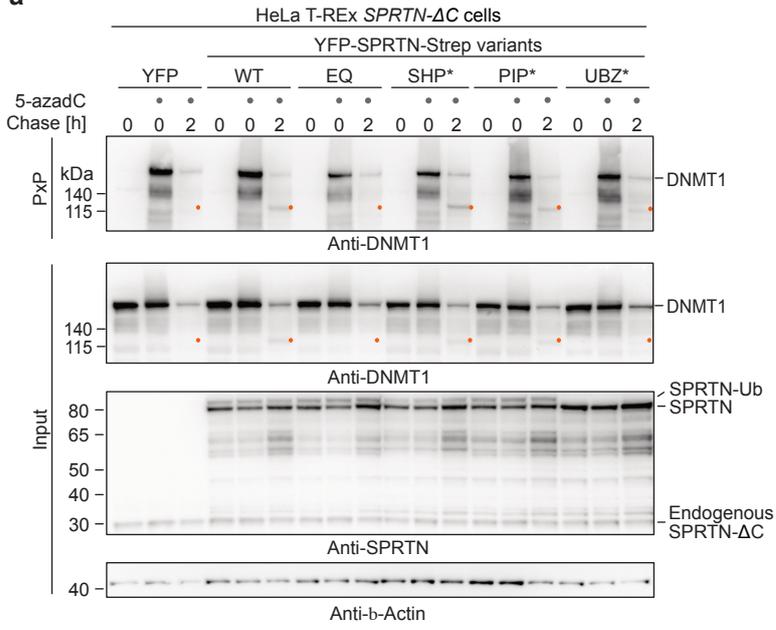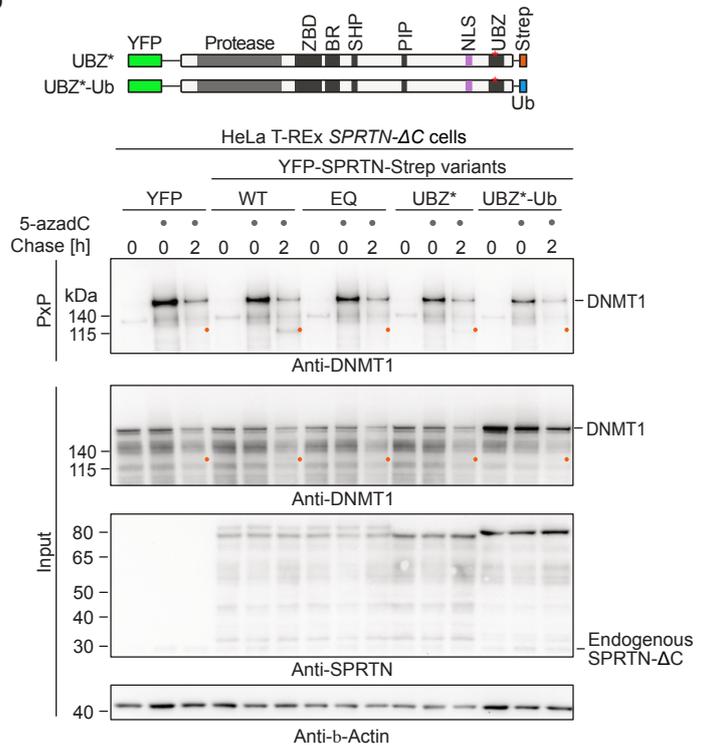
Extended Data Figure 5

Extended Data Fig. 6. Structure-function-analysis of DPC cleavage by SPRTN. a HeLa T-REx Flp-In *SPRTN-ΔC* cells complemented with variants bearing amino acid replacements in SPRTN's interaction motifs/domains for binding to p97 (SHP*), PCNA (PIP*) and ubiquitin (UBZ*) were treated as depicted in Fig. 2a. DPCs were isolated using PxP and analysed by western blotting using the indicated antibodies. The experiment was repeated three times with similar results. **b** HeLa T-REx Flp-In *SPRTN-ΔC* cells were complemented with YFP, YFP-SPRTN-Strep, YFP-SPRTN-EQ-Strep, YFP-SPRTN-UBZ*-Strep and YFP-SPRTN-UBZ*-Ub (ubiquitin C-terminal fusion) and treated as shown in Fig. 2a, before DPCs were isolated using PxP and analysed by western blotting using the indicated antibodies. The experiment was repeated twice with similar results. **c** Protein G covalently conjugated to a fluorescently-labelled single-/double-stranded DNA junction was incubated for 2 hours with increasing concentrations of the indicated recombinant SPRTN variants *in vitro*, followed by analysis using denaturing Urea-PAGE (top). Quantifications show the mean ± SD of 3 independent experiments (bottom). **d** SPRTN autocleavage activity was assessed by incubating the indicated recombinant SPRTN variants in the presence or absence of DNA (ΦX174 Virion) for 2 hours, prior to analysis of cleavage using SYPRO Ruby-stained SDS-PAGE (left). Asterisks indicate SPRTN autocleavage fragments. Quantifications show the mean ± SD of 3 independent experiments (right). Source data are provided as a Source Data file.
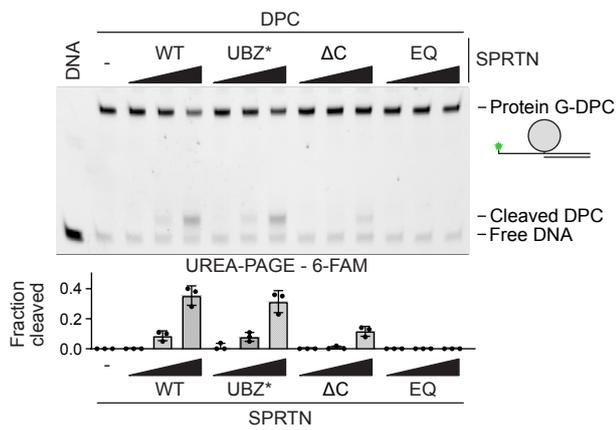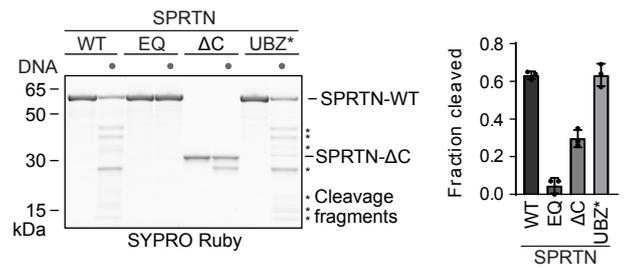
Extended Data Figure 6

a

HeLa T-REx *SPRTN-ΔC* cells

YFP-SPRTN-Strep variants: YFP, WT, EQ, SHP*, PIP*, UBZ*

5-azadC
Chase [h]: 0 0 2 (repeated)

PxP — Anti-DNMT1 (kDa: 140, 115) — DNMT1

Input — Anti-DNMT1 (140, 115) — DNMT1

Input — Anti-SPRTN (80, 65, 50, 40, 30) — SPRTN-Ub, SPRTN, Endogenous SPRTN-ΔC

Anti-b-Actin (40)

b

UBZ*, UBZ*-Ub domain schematic: YFP, Protease, ZBD, BR, SHP, PIP, NLS, UBZ, Strep, Ub

HeLa T-REx *SPRTN-ΔC* cells

YFP-SPRTN-Strep variants: YFP, WT, EQ, UBZ*, UBZ*-Ub

5-azadC
Chase [h]: 0 0 2

PxP — Anti-DNMT1 (kDa: 140, 115) — DNMT1

Input — Anti-DNMT1 (140, 115) — DNMT1

Input — Anti-SPRTN (80, 65, 50, 40, 30) — Endogenous SPRTN-ΔC

Anti-b-Actin (40)

c

DPC

DNA, -, WT, UBZ*, ΔC, EQ — SPRTN

UREA-PAGE - 6-FAM

Protein G-DPC, Cleaved DPC, Free DNA

Fraction cleaved (0.0 – 0.4): -, WT, UBZ*, ΔC, EQ — SPRTN

d

SPRTN: WT, EQ, ΔC, UBZ*

DNA

kDa: 65, 50, 30, 15

SYPRO Ruby

SPRTN-WT, SPRTN-ΔC, Cleavage fragments
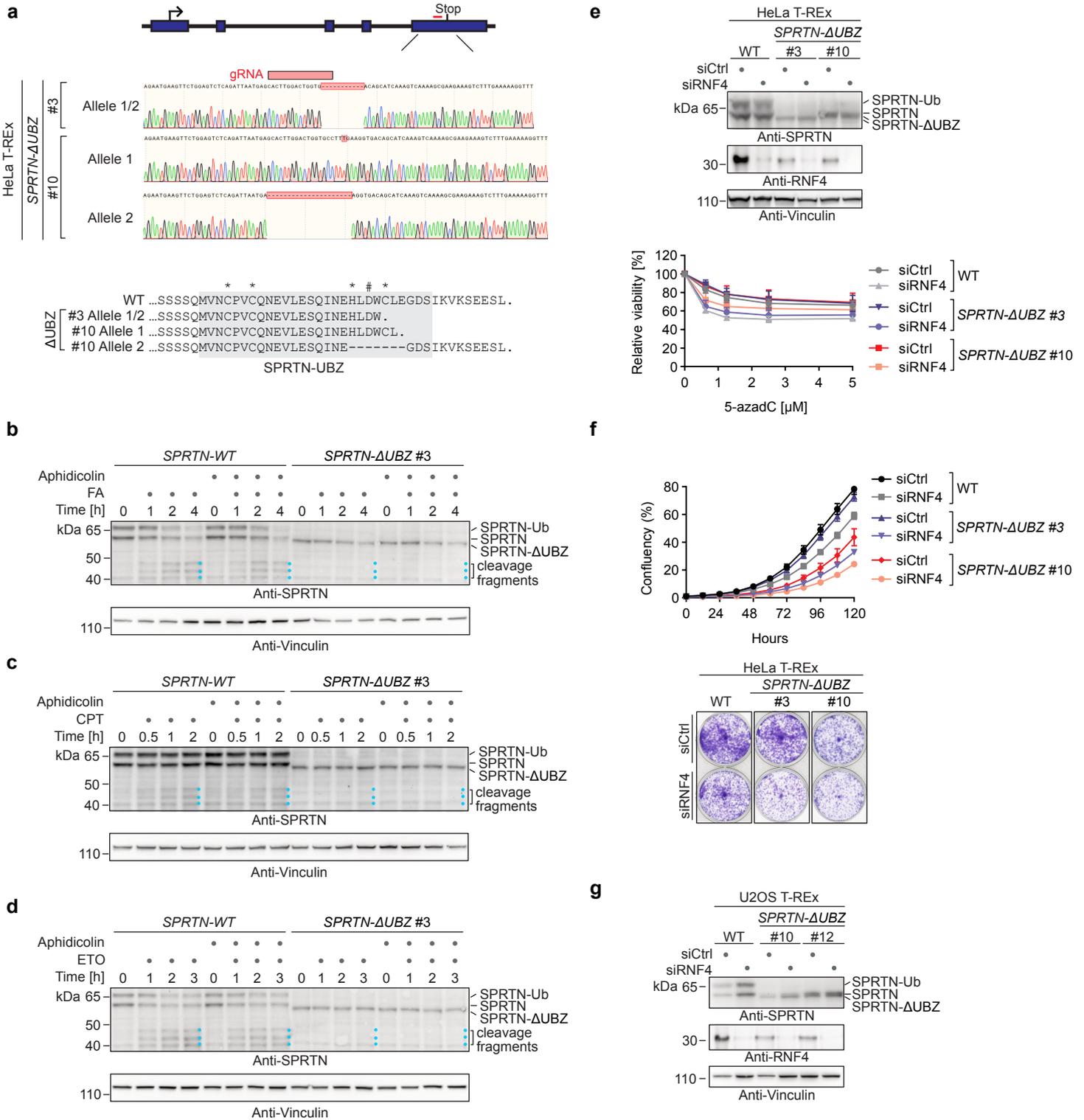
Fraction cleaved (0.0 – 0.8): WT, EQ, ΔC, UBZ* — SPRTN

Extended Data Fig. 7. Generation and characterization of *SPRTN-ΔUBZ* cells a Depiction of the editing strategy for the generation of HeLa T-REx Flp-In *SPRTN-ΔUBZ* cells and sanger sequencing traces of the edited alleles of two different clones (top). Amino acid sequences encoded by edited *SPRTN-ΔUBZ* alleles. The sequence of the SPRTN-UBZ domain is highlighted in grey. Residues coordinating the zinc ion of the UBZ are indicated with asterisks. Aspartate D473, which is essential for ubiquitin binding is indicated with # (bottom). **b-d** HeLa T-REx Flp-In WT or *SPRTN-ΔUBZ* #3 cells were treated with formaldehyde (FA, 250 µM) (**b**), CPT (500 nM) (**c**) or ETO (50 µM) (**d**) as indicated in the absence or presence of aphidicolin including a 2 hours pre-treatment. The whole cell lysates were harvested for analysis of SPRTN autocleavage by western blotting using the indicated antibodies. **(b-d)** were repeated three times with similar results. **e** Western blot analysis of HeLa T-REx Flp-In WT and *SPRTN-ΔUBZ* cells transfected with the indicated siRNAs (top). Transfected cells were exposed to increasing concentrations of 5-azadC for 96 hours, prior to assessment of cell viability using AlamarBlue assay (bottom). Values represent the mean ± SD of 3 biological replicates normalized to untreated cells. **f** HeLa T-REx Flp-In WT and *SPRTN-ΔUBZ* cells were transfected with the indicated siRNAs. Cell confluency was monitored over 5 days by IncuCyte live cell imaging every 12 hours (top, values represent the mean ± SD of 3 technical replicates), before cells were stained with crystal violet (bottom). **g** Western blot analysis of U2OS T-REx Flp-In WT and *SPRTN-ΔUBZ* cells transfected with the indicated siRNAs and used for viability assays shown in Fig. 5e. Knock-down efficiency was confirmed twice with similar results. Source data are provided as a Source Data file.

Extended Data Figure 7

a

HeLa T-REx / SPRTN-ΔUBZ

#3 Allele 1/2

#10 Allele 1

Allele 2

gRNA

```
            *   *           *  #  *
WT            ...SSSSQMVNCPVCQNEVLESQINEHLDWCLEGDSIKVKSEESL.
ΔUBZ  #3 Allele 1/2  ...SSSSQMVNCPVCQNEVLESQINEHLDW.
      #10 Allele 1   ...SSSSQMVNCPVCQNEVLESQINEHLDWCL.
      #10 Allele 2   ...SSSSQMVNCPVCQNEVLESQINE-------GDSIKVKSEESL.
```

SPRTN-UBZ

b *SPRTN-WT* / *SPRTN-ΔUBZ #3*

Aphidicolin, FA, Time [h]

Anti-SPRTN, Anti-Vinculin

SPRTN-Ub, SPRTN, SPRTN-ΔUBZ, cleavage fragments

c *SPRTN-WT* / *SPRTN-ΔUBZ #3*

Aphidicolin, CPT, Time [h]

Anti-SPRTN, Anti-Vinculin

d *SPRTN-WT* / *SPRTN-ΔUBZ #3*

Aphidicolin, ETO, Time [h]

Anti-SPRTN, Anti-Vinculin

e HeLa T-REx / *SPRTN-ΔUBZ*

WT #3 #10

siCtrl, siRNF4

Anti-SPRTN, Anti-RNF4, Anti-Vinculin

SPRTN-Ub, SPRTN, SPRTN-ΔUBZ

Relative viability [%] vs 5-azadC [µM]

- siCtrl / siRNF4] WT
- siCtrl / siRNF4] *SPRTN-ΔUBZ #3*
- siCtrl / siRNF4] *SPRTN-ΔUBZ #10*

f Confluency (%) vs Hours

- siCtrl / siRNF4] WT
- siCtrl / siRNF4] *SPRTN-ΔUBZ #3*
- siCtrl / siRNF4] *SPRTN-ΔUBZ #10*

HeLa T-REx / *SPRTN-ΔUBZ*

WT #3 #10

siCtrl, siRNF4

g U2OS T-REx / *SPRTN-ΔUBZ*

WT #10 #12

siCtrl, siRNF4

Anti-SPRTN, Anti-RNF4, Anti-Vinculin

SPRTN-Ub, SPRTN, SPRTN-ΔUBZ

Extended Data Fig. 8. Gating strategy for flow cytometry analysis. Cell debris and aggregates were excluded by SSC-A/FSC-A and single cells were gated by SSC-H/SSC-A. Live cells were identified by SSC-A/APC-Cy7-A and the fraction of EdU/EU positive cells was identified based on FITC-A fluorescence intensity relative to unlabeled cells.

Extended Data Figure 8

Step 1

Step 2

Step 3

Step 4
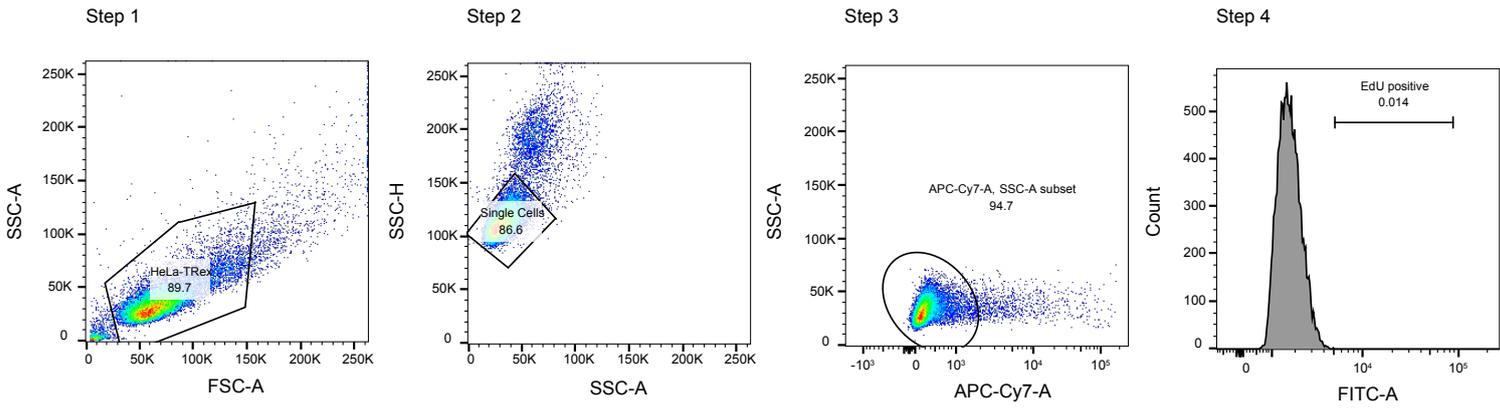
![ANNUAL REVIEWS logo]

DNA–Protein Crosslinks and Their Resolution

Pedro Weickert[1,2] and Julian Stingele[1,2]

[1]Department of Biochemistry, Ludwig Maximilians University, Munich, Germany;
email: stingele@genzentrum.lmu.de

[2]Gene Center, Ludwig Maximilians University, Munich, Germany

![ANNUAL REVIEWS CONNECT]

www.annualreviews.org

- Download figures
- Navigate cited references
- Keyword search
- Explore related articles
- Share via email or social media

Keywords

DNA–protein crosslink, formaldehyde, MRE11, SPRTN, TDP, topoisomerase

Abstract

Covalent DNA–protein crosslinks (DPCs) are pervasive DNA lesions that interfere with essential chromatin processes such as transcription or replication. This review strives to provide an overview of the sources and principles of cellular DPC formation. DPCs are caused by endogenous reactive metabolites and various chemotherapeutic agents. However, in certain conditions DPCs also arise physiologically in cells. We discuss the cellular mechanisms resolving these threats to genomic integrity. Detection and repair of DPCs require not only the action of canonical DNA repair pathways but also the activity of specialized proteolytic enzymes—including proteases of the SPRTN/Wss1 family—to degrade the crosslinked protein. Loss of DPC repair capacity has dramatic consequences, ranging from genome instability in yeast and worms to cancer predisposition and premature aging in mice and humans.

Contents

1. INTRODUCTION

The integrity of the genetic information stored in DNA is constantly challenged by various types of lesions (1). Spontaneous decay, as well as numerous endogenous and exogenous agents, alters the sequence and structure of the DNA molecule (2). The resulting DNA damage is diverse in nature and ranges from abasic sites and mismatched or damaged bases (e.g., by alkylation or oxidation) to single- and double-strand breaks and covalent crosslinks between bases of the same or opposing DNA strands (intrastrand crosslinks or interstrand crosslinks, respectively) (**Figure 1a**). If unrepaired, these lesions can ultimately cause mutagenesis, senescence, cell death, tumorigenesis, and accelerated aging (3, 4). Therefore, cells employ specialized lesion-specific DNA damage response (DDR) pathways to detect the presence of DNA damage, signal to activate the cellular DNA repair machinery, and repair these threats to cellular and organismal integrity. The importance of DNA repair is underlined by germline mutations in DDR genes causing human premature aging and cancer-predisposition syndromes (5). Moreover, DDR genes are frequently subjected to somatic mutation during malignant transformation (6). The simultaneous inactivation of DNA repair and DNA damage signaling mechanisms enables the acquisition of additional mutations while escaping the cellular arrest typically triggered by DNA damage (7). In addition, the loss of DNA repair capacity creates vulnerabilities in cancer cells (8). Radiation therapy and many chemotherapeutics exploit this by inducing cytotoxic DNA damage to kill cancer cells. Moreover, the loss of specific repair pathways can trigger synthetic lethality upon inhibition of alternative repair mechanisms (9). The first clinically approved strategy exploiting such a mechanism is the treatment of patients with tumors defective for homology-directed repair with poly(ADP-ribose) polymerase inhibitors (10–12).

As a consequence, DNA lesions and their repair have received exceptional research attention. However, the significance of one particular type of DNA lesion—covalent DNA–protein crosslinks (DPCs)—has emerged only in the last decade. Nonetheless, it was noted in the early years of DNA repair research that DPC induction contributes significantly to the lethality caused by irradiation (13). Moreover, cellular DPC formation was observed upon exposure of

Abasic site: a position in DNA that lacks a DNA base; an abundant DNA lesion caused by spontaneous hydrolysis or upon enzymatic base excision

Synthetic lethality: synthetic lethality occurs if the combination of two genetic perturbations causes lethality, while individual perturbations do not affect viability
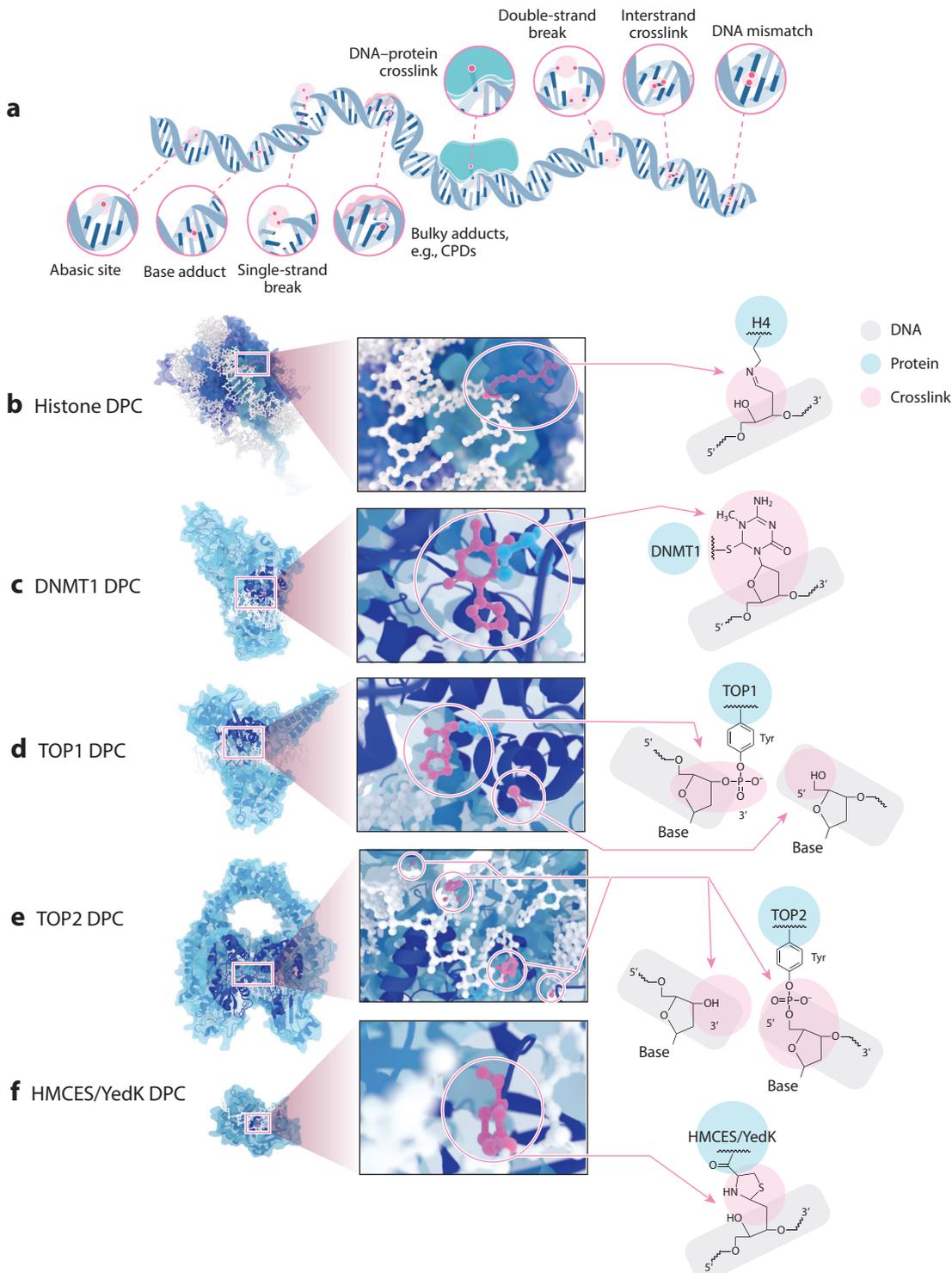
(*Caption appears on following page*)

Figure 1 (*Figure appears on preceding page*)

The diversity of DNA lesions and covalent DPCs. (*a*) Schematic depiction of common DNA lesions. Abasic sites lack a DNA base, while the phosphate backbone remains intact. Small base adducts do not strongly affect the structure of the DNA helix, while bulkier adducts such as UV-light-induced CPDs do. Interstrand crosslinks arise upon crosslinking of DNA bases of opposing DNA strands. DNA mismatches are mainly caused by errors made during DNA replication. (*b–f, left*) Overall structure of diverse DPCs, (*center*) close-up view of the crosslink, and (*right*) chemical structures of the covalent bond. (*b*) Crosslink between a lysine of histone H4 and an abasic site (artistic depiction based on PDB ID 1AOI). (*c*) DNMT1 covalently trapped on methylated 5azadC (PDB ID: 4DA4). (*d*) TOP1 cleavage complex featuring a covalent phosphotyrosyl bond between the active-site tyrosine and the 3′ end of a DNA single-strand break (PDB ID: 1T8I). (*e*) TOP2 cleavage complex featuring covalent phosphotyrosyl bonds between the active-site tyrosine of each TOP2 subunit and the 5′ ends of a DNA double-strand break (PDB ID: 5GWK). (*f*) Covalent HMCES/YedK DPC featuring a covalent thiazolidine linkage between the open-chain conformation of an abasic site and the enzyme's N-terminal cysteine residue (PDB ID: 6NUA). Abbreviations: 5azadC, 5-aza-2′-deoxycytidine; CPD, cyclobutane pyrimidine dimer; DNMT1, DNA methyltransferase 1; DPC, DNA–protein crosslink; HMCES, 5-hydroxymethylcytosine binding ESC-specific; PDB ID, Protein Data Bank identifier; TOP, topoisomerase; UV, ultraviolet.

cells to reactive aldehydes such as formaldehyde or the Penicillin Roquefort toxin (14, 15). The crosslinking ability of formaldehyde was soon employed to capture DNA–protein interactions by chromatin immunoprecipitation–based techniques (16). How cells precisely repair DPCs remained unclear; it was simply assumed that DPCs are targeted by canonical DNA repair pathways (17). However, the first publication describing formaldehyde-induced DPCs had already speculated that proteolysis of the protein adduct might be required for their repair:

> We have shown that, in vitro, pronase can cleave the [protein-bridged] interstrand cross-links induced in DNA by formaldehyde. . . . If a proteolytic enzyme was to act in the same way in vivo, this would constitute DNA repair (15).

Almost 40 years later, the yeast metalloprotease Wss1 was identified to repair DPCs proteolytically (18). Together with the concurrent description of proteolytic DPC repair in *Xenopus laevis* egg extracts this discovery triggered intense interests in DPC formation and repair (19). The subsequent identification of SPRTN (also called Spartan, see the sidebar titled Spartan—What's in a Name?) as the metazoan counterpart of Wss1 and the severe consequences associated with defective DPC repair have brought DPCs into the focus of genome stability research (20–25). In this review, we describe the nature and sources of DPCs, as well as their toxic consequences and the cellular mechanisms resolving these pervasive DNA lesions.

Penicillin Roquefort toxin: a mycotoxin with a reactive aldehyde group causing cytotoxic DPC formation; produced by the blue cheese mold *Penicillium roqueforti*

2. DNA–PROTEIN CROSSLINKS

DPCs are a diverse class of lesions that differ with respect to the identity of the crosslinked protein, the type of DNA structure involved, and the chemical nature of the covalent crosslink

---

SPARTAN—WHAT'S IN A NAME?

The authors have investigated mechanisms of DPC repair for a considerable fraction of their professional lives, and the name of the SPRTN (pronounced Spartan) protease has been an unfailing source of jokes and memes. However, the precise origin of its colorful name has remained enigmatic. SPRTN stands for *SprT*-domain at the N terminus (98). *SprT* is an *E. coli* gene encoding a SPRTN-like protease (87). The *SprT* gene was named by Dr. Ryutaro Utsumi in 1996, but the corresponding work was never published. To settle the etymological question, the authors have contacted Dr. Utsumi, who kindly disclosed that *SprT* stands for stationary phase regulated. His team identified *SprT* in a screen for genes regulating expression of the bolA1 transcriptional regulator in the stationary phase (R. Utsumi, personal communication).

itself (**Figure 1**) (26–28). DPCs can be broadly classified based on the principle of crosslink formation (29). Nonenzymatic DPCs are caused by reactive agents, which nonspecifically crosslink chromatin proteins to DNA. In contrast, enzymatic DPCs involve specific proteins that form covalent intermediates with DNA as part of their evolved enzymatic function. DPC formation occurs when these normally transient intermediates become stabilized, thereby resulting in covalent entrapment of the enzyme. Enzymatic and nonenzymatic DPCs are induced by various endogenous and exogenous agents, including widely used chemotherapeutics (26). DPCs are toxic because their bulkiness efficiently blocks essential chromatin transactions. For instance, they interfere with DNA unwinding and DNA synthesis during replication and block progression of transcribing RNA polymerases (30–34). However, DPC formation is not exclusively a pathological event. Certain enzymatic DPCs form deliberately in cells and have important physiological functions (35, 36).

2.1. Nonenzymatic DNA–Protein Crosslinks

Bifunctional crosslinking agents induce efficient formation of nonenzymatic DPCs, while also causing a variety of other lesions such as monoadducts or DNA–DNA crosslinks. Some crosslinkers arise endogenously as part of normal metabolism (37). The reactive metabolite acetaldehyde is, for instance, produced during detoxification of ethanol (38). Moreover, formaldehyde is a common metabolite present at micromolar concentrations in mammalian serum and is even produced in direct proximity to DNA during histone demethylation (39–42). Reactive oxygen species (ROS) can react with DNA bases or amino acids to generate highly reactive radical species; this results in DPC formation upon reaction with a second macromolecule (28, 43). Furthermore, some intermediates of DNA metabolism bear intrinsic reactivity. 5-Formylcytosine, an oxidized 5-methylcytosine species that occurs as an intermediate during active DNA demethylation by TET (ten-eleven translocation) enzymes, can autocrosslink to proteins (44). Similarly, DNA-binding proteins can react with the reactive open-chain aldehyde conformation of abasic sites to form crosslinks (**Figure 1b**) (45). Apart from these endogenous sources, nonenzymatic DPCs are also caused by a variety of exogenous agents, including transition metals such as chromium, the tobacco-smoke constituent butadiene, and the chemical warfare agent and chemotherapeutic nitrogen mustard (46–48). DPCs are also induced by platinum-based chemotherapeutics (cisplatin, oxaliplatin, and carboplatin), although the principal mechanism of action is commonly attributed to their ability to induce DNA–DNA crosslinks (49). Ionizing radiation (IR) causes crosslinking indirectly through the generation of highly reactive ROS and radical species, which in turn generate a variety of DNA lesions, including DPCs (50). Notably, IR-induced DPC formation occurs preferentially in the absence of molecular oxygen, whereas DNA-break formation prevails in normoxic conditions (51–53). Attempts have been made to determine the identities of nonenzymatic DPCs (28). However, while chromatin proteins are more prone to becoming crosslinked, a systematic and quantitative understanding of what proteins are crosslinked by which agent is still elusive. Thus, it remains unclear to what extent DPCs induced by radiotherapy or platinum-based crosslinkers contribute to their therapeutic success.

2.2. Enzymatic DNA–Protein Crosslinks

That DPC induction can underlie successful chemotherapy is exemplified by several agents exerting their therapeutic effect by crosslinking specific enzymes to DNA (26). The cytidine analogue 5-aza-2′-deoxycytidine (5azadC, often referred to as decitabine) is incorporated into DNA during replication and is used to treat myelodysplastic syndromes (54–56). Once incorporated, DNA methyltransferase 1 (DNMT1) methylates 5azadC, which involves the formation of a covalent

Reactive oxygen species (ROS): a highly reactive by-product of cellular oxygen metabolism, e.g., superoxide and hydroxyl radicals

intermediate. However, the substituted nitrogen at position 5 prohibits the final β-elimination reaction required to release DNMT1 (**Figure 1c**). Entrapment of DNMT1 causes not only the formation of a DPC but also a global decrease in DNA methylation, which contributes to its therapeutic success by leading to reexpression of tumor suppressor genes silenced by pathological hypermethylation (57).

The mode of action of chemotherapeutics resulting in the formation of topoisomerase DPCs is distinct. These so-called enzyme poisons stabilize the physiologically occurring covalent phospho-tyrosyl bonds between DNA termini and the active-site tyrosines of topoisomerases (58). Topoisomerase 1 (TOP1) enzymes relieve torsional stress within the DNA molecule during transcription and replication (59). To this end, TOP1 inserts a single-strand break, which allows rotation of the DNA strand (60). During relaxation, TOP1 remains covalently attached to the 3′ end of the single-strand break via its active-site tyrosine—this state is often referred to as the TOP1 covalent complex (TOP1cc) (**Figure 1d**). The reaction is completed by TOP1 religating the single-strand break and concurrent release of the enzyme. TOP1ccs can be stabilized by the plant alkaloid camptothecin and its chemotherapeutic derivatives irinotecan and topotecan, which intercalate in the TOP1–DNA interface (61). Interestingly, the camptothecin-producing plant *Camptotheca acuminata* encodes a camptothecin-resistant TOP1 variant with an N722S substitution (directly adjacent to the catalytic tyrosine 723) (62). Notably, the identical substitution was observed to appear in CEM leukemia cells after selection with camptothecin (63). Entrapment of TOP1ccs also occurs naturally. Completion of the enzymatic reaction is sensitive to distortions within the DNA molecule caused by damaged bases or abasic sites in the immediate vicinity (64, 65).

Similar to TOP1, TOP2 forms a covalent cleavage complex during DNA processing (60). Each subunit of the homodimeric TOP2 complex induces a single-strand break in the substrate DNA, which results in formation of a double-strand break. TOP2 controls DNA supercoiling by passing a second intact DNA strand through the break while remaining covalently attached to either 5′ end of the double-strand break via the tyrosine from either active site (**Figure 1e**). The covalent TOP2 cleavage complex (TOP2cc) is typically transient but can be stabilized by several enzyme poisons, including the chemotherapeutic agents etoposide, doxorubicin, and mitoxantrone (66). Although cancer therapy by TOP2-DPC-inducing drugs has proven successful, it bears the risk of causing chemotherapy-induced oncogenic translocations (67, 68). The cytotoxicity of TOP2-like DPCs is useful not only for cancer therapy but also as a tool in molecular biology. The popular Gateway cloning system uses the CcdB toxin for counterselection. CcdB is toxic in wild-type *Escherichia coli* cells because it stabilizes the covalent intermediate of the bacterial TOP2-like enzyme DNA gyrase (69). Plasmids encoding CcdB can be expressed only in strains encoding a resistant gyrase variant (gyrA462) (69). However, formation of DPCs by TOP2-like enzymes is not pathological per se; it can also be employed by cells for important physiological functions.

2.3. Physiological DNA–Protein Crosslinks

Covalent reaction intermediates between enzymes and DNA are designed to be highly transient. However, specific DPC-causing enzymes exist that maintain the covalent state for different purposes. The TOP2-like enzyme SPO11 is essential for meiotic recombination and is conserved from yeast to mammals (36, 70). SPO11 induces double-strand breaks during meiosis and remains covalently attached to the 5′ ends of the break until repaired (71). DPC formation by another tyrosine-based enzyme, the Epstein–Barr virus nuclear antigen 1 (EBNA1) protein, is important for viral episome maintenance (72). EBNA1's active-site tyrosine covalently conjugates to the episomal *oriP* sequence, which is needed for resolving a four-way Holliday junction generated during viral replication (72). How EBNA1 DPCs are eventually released and whether release requires the activity of cellular DPC repair enzymes remain untested.

DPCs also arise intentionally to shield otherwise labile DNA lesions during replication. The presence of unstable abasic sites in single-stranded DNA (ssDNA) is dangerous, because they can be converted into single-strand breaks either spontaneously or enzymatically through cleavage by apurinic or apyrimidinic site lyases (73). Break formation in ssDNA during replication causes fork collapse and generation of highly detrimental single-ended double-strand breaks (74). Thus, cells employ a mechanism conserved throughout the tree of life to convert abasic sites into DPCs. 5-Hydroxymethylcytosine binding ESC-specific (HMCES) and its prokaryotic homolog yedK efficiently crosslink to abasic sites within ssDNA (35, 75–77). Crosslinking occurs through the N-terminal cysteine residue of the family-defining SOS response associated peptidase domain, which forms a thiazolidine linkage with the open-chain conformation of abasic sites (**Figure 1***f*) (35, 75–78). HMCES appears to be particularly important for shielding abasic sites generated by base excision repair of uracil-containing DNA originating from cytosine deamination by apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like 3A (APOBEC3A) (79). As a consequence, loss of HMCES is synthetic lethal in combination with the increased activity of APOBEC3A observed in various cancer types (80). Of note, HMCES has also been implicated in the repair of DNA double-strand breaks by the alternative end-joining pathway, which is independent of its ability to form DPCs (81). Despite HMCES crosslinks forming deliberately, it appears that they cannot simply revert; they require repair by cellular DPC repair mechanisms similarly to pathological DPCs (35).

3. DNA–PROTEIN CROSSLINK PROTEOLYSIS

The three components of DPCs—DNA, protein, and crosslink—can be used to initiate repair (26). In fact, successful DPC repair frequently relies on successful collaboration between different repair mechanisms acting consecutively on each of the components. Although, the principle of repair is strongly influenced by the nature of the DPC, proteolytic destruction of the protein component has emerged as a key event during repair of diverse DPCs.

3.1. The DPC Proteases Wss1 and SPRTN

The yeast metalloprotease Wss1 was the first protease identified as being specifically dedicated to the proteolytic repair of DPCs (18). Yeast strains lacking Wss1 are sensitive to induction of enzymatic and nonenzymatic DPCs and display hyperrecombination and dramatic rates of gross chromosomal rearrangements (18, 82, 83).

Wss1-like enzymes are present in bacteria, yeasts, and plants, and their essential role in genome maintenance has been confirmed in several species (84–87). Proteases of the related SPRTN family are present in all metazoans. SPRTN- and Wss1-like enzymes share a common evolutionary origin, but the split occurred early in evolution, with both types of enzymes being present in prokaryotes (87). SPRTN and Wss1 proteases carry out the same role in vivo, have highly similar activities in vitro, and display almost identical domain organization (29). It is therefore conceivable that the last common ancestor of SPRTN and Wss1 was already a DPC protease. Loss of SPRTN proteases causes genome instability in various species, with the severity of the phenotype increasing during evolution. SPRTN-deficient worms are viable but highly sensitive to DPC-inducing agents (20, 88, 89). Flies display a peculiar phenotype upon loss of the SPRTN homolog maternal haploid (mh). *mh*-mutant female flies are sterile even though they produce eggs in normal quantities (90). The eggs fail to hatch, because in the first mitotic divisions following fertilization, paternal chromosomes fail to condense, resulting in dramatic missegregation. Consequently, paternal DNA is lost during the first cell divisions, leading to the development of inviable haploid embryos containing only maternal DNA (90). Although the identity of the crosslink responsible for this

dramatic genome instability remains to be established, one can deduce that paternal fly DNA contains substantial amounts of DPCs, which require maternal SPRTN for repair. In mammals, SPRTN is essential for viability, with $Sprtn^{-/-}$ mouse embryos dying during early embryogenesis (24). Conditional knockout mouse embryonic fibroblasts arrest upon induction of SPRTN depletion within a few cell cycles (24). The arrest is accompanied by pronounced signs of genome instability, including increased micronuclei, chromatin bridges during anaphase, and γH2AX- and RAD51-foci formation (24). In humans, germline hypomorphic mutations in *SPRTN* were identified to be causative for Ruijs-Aalfs syndrome (RJALS) (25). All three known patients suffer from premature aging and early-onset hepatocellular carcinomas (25, 91). One patient is homozygous for a 1-bp deletion at the beginning of exon 5 that results in a frame shift and deletion of the C-terminal tail of SPRTN (this allele is referred to as *SPRTN-ΔC*). The two other patients are compound heterozygous for a highly similar *SPRTN-ΔC* allele caused by a frameshift at the end of exon 4 and a second *SPRTN-YC* mutant allele resulting in an amino acid substitution (Y117C) close to the enzyme's catalytic residues (**Figure 2a**) (25). Hypomorphic *Sprtn*-mutant mice, which express reduced levels of wild-type Sprtn, phenocopy many RJALS patient features including premature aging and predisposition for liver tumors (23, 24). SPRTN depletion in human cells causes DPC repair defects and sensitivity to DPC-inducing agents (20–22, 92). To conclude, Wss1 and SPRTN proteases are essential to protect genome integrity across species. DPC proteases require broad substrate specificity in order to tackle the complexity of DPCs. As a consequence, they require tight control to prohibit unwanted proteolysis of mere chromatin bound proteins.

3.2. Regulation of SPRTN

Human SPRTN is a 55-kDa protein with an N-terminal protease domain followed by a zinc-binding domain (ZBD), a basic region (BR) and a C-terminal tail bearing several protein–protein interaction motifs (**Figure 2a**). ZBD and BR are DNA-binding domains that bind to ssDNA and double-stranded DNA (dsDNA), respectively (93, 94). The unstructured C-terminal tail contains a SHP-box for binding the segregase p97 (88, 95). Binding motifs for p97 (Cdc48 in yeast) are present in most SPRTN and Wss1 homologs, but p97's precise role in DPC repair is unclear (29). It is conceivable that p97 unfolds DPCs to facilitate cleavage by SPRTN. However, p97 binding appears to be at least partially dispensable for SPRTN function, given that the *SPRTN-ΔC* allele supports viability in RJALS patients despite lacking p97-binding motifs (26). Either p97 is needed only for repair of certain nonessential SPRTN substrates or other segregase or unfolding activities compensate for the lack of interaction between SPRTN and p97. The patient variants also lack SPRTN's C-terminal ubiquitin-binding zinc finger (UBZ) and the proliferating cell nuclear antigen (PCNA)–interacting protein (PIP) box required for binding the replication clamp PCNA, whose role in recruiting SPRTN to the lesion is not entirely understood. In general, it remains controversial how SPRTN is recruited to chromatin upon DPC induction, with some evidence pointing toward a ubiquitylation signal and other evidence toward a SUMOylation signal coupled to adaptor proteins (89, 96).

Nevertheless, key principles underlying the regulation of SPRTN have emerged that enable safe operation of this potentially toxic enzymatic activity in human cells. Three interdependent molecular switches control SPRTN's activation and subsequent inactivation (**Figure 2b**). SPRTN's activity is governed by a DNA switch based on recognition of specific DNA structures. SPRTN and Wss1 are inactive on their own but become active in the presence of DNA (18, 20, 21). Activation requires engagement of both of SPRTN's DNA-binding domains, which can occur only at DNA structures bearing single- and double-stranded features such as ss/dsDNA junctions or DNA bubbles (**Figure 2b**) (94). Even residual amounts of ssDNA enable activation; fraying of

Micronuclei: small extranuclear bodies containing DNA; their formation is often caused by errors during mitotic chromosome segregation

Hypomorphic mutation: a mutation resulting in partial loss of gene function or reduced expression of the wild-type protein

Segregase: a chaperone-like enzyme able to extract proteins from their environment
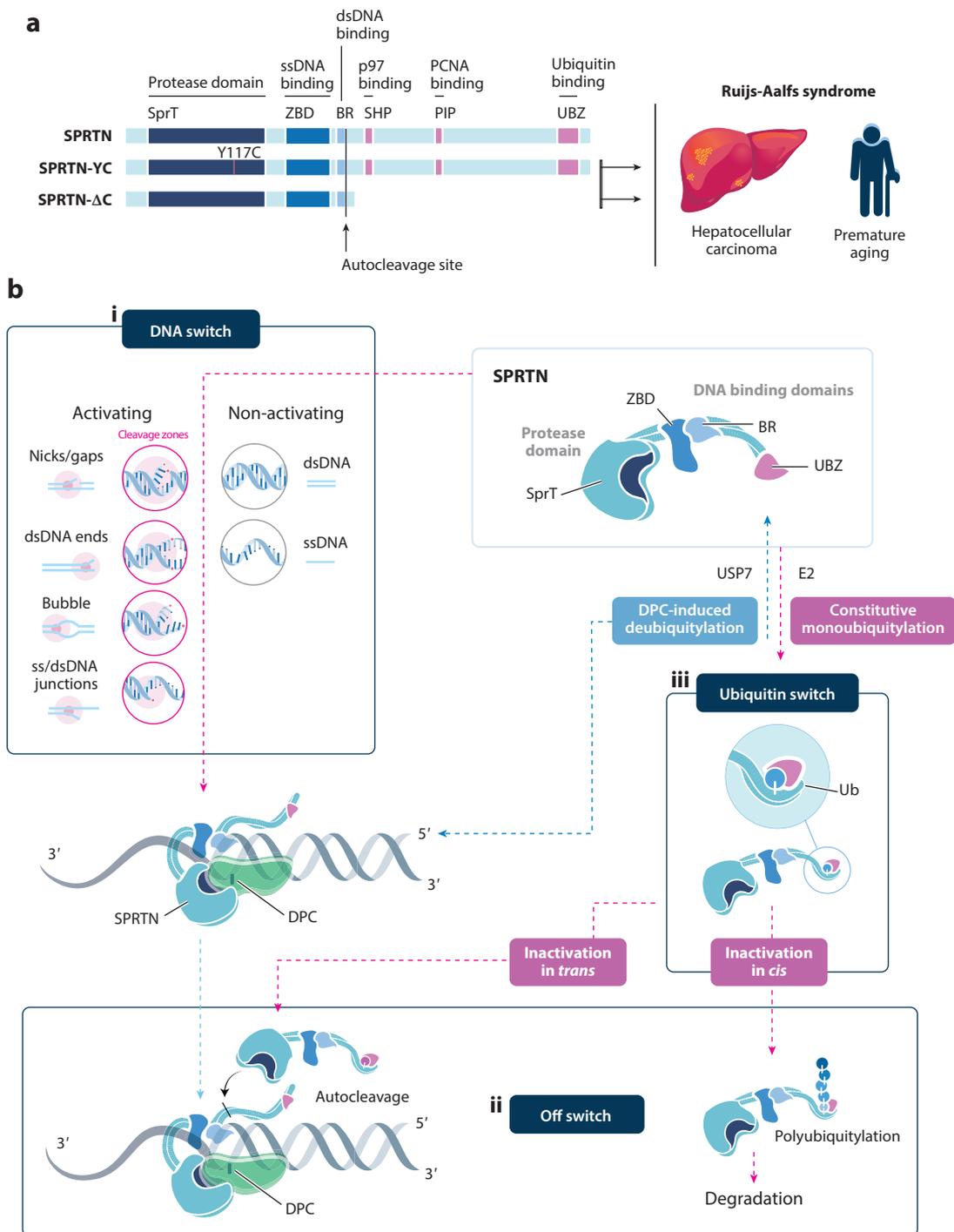
(*Caption appears on following page*)

Figure 2 (*Figure appears on preceding page*)

The DNA protease SPRTN and its regulation. (*a*) Domain organization of human SPRTN and of the two variants causing Ruijs-Aalfs syndrome, which is characterized by premature aging and early-onset hepatocellular carcinomas. SPRTN-YC contains an amino acid substitution (Y117C) close to the enzyme's catalytic residues. SPRTN-ΔC features a deletion of the C-terminal tail of SPRTN due to frame shifts caused by a deletion either at the end of exon 4 or at the beginning of exon 5. (*b*) Three regulatory switches controlling SPRTN activation and subsequent inactivation. (*i*) DNA switch. SPRTN becomes active upon binding DNA structures with ssDNA and dsDNA features. (*ii*) Off switch. Autocatalytic cleavage of SPRTN in *trans* inactivates the enzyme and triggers eviction from chromatin. (*iii*) Ubiquitin switch. This switch negatively controls SPRTN in two ways: SPRTN is constitutively monoubiquitylated, which enhances inactivation in *trans* by promoting autocleavage, and it also causes inactivation in *cis* by priming SPRTN for polyubiquitylation, which in turn leads to proteasomal degradation. Upon DPC induction, USP7 relieves this negative regulation by deubiquitylating SPRTN. Abbreviations: BR, basic region; DPC, DNA–protein crosslink; dsDNA, double-stranded DNA; PIP, proliferating cell nuclear antigen–interacting protein box; ssDNA, single-stranded DNA; Ub, ubiquitin; UBZ, ubiquitin-binding zinc finger; USP7, ubiquitin specific protease 7; ZBD, zinc-binding domain.

dsDNA ends or local unwinding caused by a DNA nick are sufficient for activation of SPRTN. However, DNA that is entirely single or double stranded does not activate SPRTN (94). The nature of the DNA structure–dependent activation of SPRTN is not fully understood, but it has been speculated that the ZBD shields access to the active site and that activation involves a conformational change relieving this inhibition (93). Once activated, SPRTN is able to process protein adducts in close spatial proximity to the activating structure (94). In addition, DNA binding triggers autocatalytic inactivation of SPRTN by inducing autocleavage in *trans* (20, 21). Autocleavage acts as an off switch and results in eviction from cellular sites of DNA damage (20). Notably, the off switch is itself subjected to additional control by a ubiquitin switch, which modulates SPRTN's activity and stability (**Figure 2*b***) (97). SPRTN is constitutively monoubiquitylated to almost 50% in cells, with the modification being shielded by the enzyme's own UBZ domain (88, 95, 97, 98). Monoubiquitylated SPRTN displays increased autocleavage in cells and in vitro, while substrate cleavage is unaffected. Furthermore, the modification inactivates SPRTN by not only triggering autocleavage in *trans* but also priming it for proteasomal degradation (97). Importantly, this negative regulation of SPRTN is turned off in times of need. SPRTN is deubiquitylated by USP7 upon DPC induction, which presumably increases the amount of enzyme available for repair and also increases its lifetime at the damage site by reducing autocleavage. Of note, while USP7 appears to be the major deubiquitylating enzyme acting on SPRTN, USP11 and VCPIP1 have been implicated in SPRTN deubiquitylation as well (97, 99, 100).

3.3. Additional DNA–Protein Crosslink Proteases

Besides SPRTN or Wss1, cells appear to require additional proteases to cope with the amount and complexity of DPCs (101). Acidic repeat–containing protein (ACRC), also called germ cell nuclear antigen, bears a SPRTN/Wss1-like metalloprotease domain and is mainly expressed in germ cells (87, 102). In fact, a monoclonal antibody specific for ACRC is commonly used as a marker for germ cells (102). ACRC deficiency results in sterility and genome instability in various species, which has been attributed to a defect in repairing mitotic SPO11 DPCs (89, 102–104). Notably, while rodents have lost the protease domain of ACRC, knockout mice are sterile, suggesting that the key function of ACRC in meiosis is not proteolytic (102). Furthermore, the trypsin-like serine protease FAM111A has been implicated in digesting protein obstacles that block progression of the replication fork (105). Whether this function is related to its role as a SV40 replication restriction factor remains to be tested (106). Germline mutations in *FAM111A* are causative for Kenny–Caffey syndrome and severe osteocraniostenosis, which are characterized by impaired skeletal development and hypoparathyroidism (107). Interestingly, the dominant patient alleles encode a hyperactive FAM111A variant, which results in cytotoxicity (108, 109).

Hypoparathyroidism: a rare condition characterized by insufficient amounts of parathyroid hormone that leads to low calcium and phosphorus levels in blood

The aspartyl protease Ddi1 was identified as a potential additional DPC protease acting in parallel to Wss1 in yeast (110, 111). Depletion of Ddi1 enhances the sensitivity of Wss1-deficient yeast cells toward Top1 DPCs and formaldehyde-induced DPCs (110). Ddi1 bears a retroviral-like protease domain that preferentially digests polyubiquitylated substrates (112). Of note, human and worm Ddi1 homologs use a similar mechanism to proteolytically activate the polyubiquitylated NFE2L1 transcription factor upon proteasome inhibition, resulting in the expression of genes encoding proteasomal subunits (113–115). Direct biochemical evidence that ACRC, FAM111A, and Ddi1 act proteolytically on DPCs remains sparse. Understanding the in vitro specificities of these enzymes is important for understanding the principles of pathway choice and division of labor in cells. Furthermore, the major cellular proteolytic degradation machine—the proteasome—is involved in destruction of various DPCs in addition to specialized DPC proteases (89, 116–121).

How do DPC proteases know which proteins to degrade? Detecting a DPC is not a trivial task for a cell, given that the lesion likely appears similar to a protein merely interacting with DNA; the actual crosslink is buried in the DNA–protein interface. Therefore, high-affinity sensor proteins, which are commonly used to detect other types of DNA lesions, are less suitable. However, the consequences of DPCs on chromatin processes provide an opportunity for detection.

4. REPLICATION-COUPLED DNA–PROTEIN CROSSLINK REPAIR

The detailed order of events initiated by a replication fork encountering a DPC in the DNA template has been studied by monitoring replication of DPC-containing plasmids in *X. laevis* egg extracts (**Figure 3**) (19, 121–123). The replicative Cdc45–Mcm2–7–GINS (CMG) helicase unwinds the DNA template by translocating on the leading strand (32). A DPC on the leading strand initially blocks progression of the CMG helicase, which triggers ubiquitylation of the protein adduct by the CMG-associated ubiquitin E3 ligase TRAIP (121). Bypass of the DPC requires a second helicase, regulator of telomere elongation helicase 1 (RTEL1) (122). RTEL1 provides genome stability by unwinding challenging DNA secondary structures during replication (124). It is assumed that RTEL1 loads on the lagging strand at the stalled fork to unwind the template strand downstream of the DPC (122). The resulting stretch of ssDNA then allows CMG to bypass the DPC. The precise mechanism is not fully understood but presumably requires opening of a side channel within the CMG helicase. Next, the leading-strand DNA polymerase extends the nascent DNA strand up to the DPC but is unable to simply synthesize across the adduct. Polymerase stalling uncouples DNA synthesis from DNA unwinding by CMG and, thus, results in accumulation of ssDNA downstream of the DPC. Of note, DPCs located on the lagging strand are immediately bypassed by the CMG without requiring additional factors, but they also result in accumulation of ssDNA by stalling progression of the lagging strand DNA polymerase (19, 32). ssDNA is swiftly covered by the ssDNA-binding replication protein A (RPA), which in turn recruits the ubiquitin E3 ligase RFWD3 (123). RFWD3-dependent ubiquitylation of proteins on ssDNA is important for many DNA repair processes including translesion synthesis (TLS) and homologous recombination (123, 125). In addition, interstrand crosslink repair by the Fanconi anemia pathway relies on RFWD3 with germline mutations in *RFWD3* causing Fanconi anemia (126–128). In the context of DPC repair, RFWD3 triggers additional DPC ubiquitylation (123). Resumption of DNA synthesis at the DPC requires debulking of the protein adduct by proteolysis. In *X. laevis* egg extracts this occurs by two redundant processes (**Figure 3**) (121). The approach of the DNA polymerase up to the lesion creates a ss/dsDNA junction in direct proximity to the DPC. This enables engagement of both of SPRTN's DNA-binding domains, triggering local activation of the protease and concurrent DPC proteolysis. Thus, the structure-specific activity of SPRTN enables the coupling of its activity to replication (94). In parallel, DPCs are degraded

Cdc45–Mcm2–7–GINS (CMG) helicase: replicative helicase complex in eukaryotes composed of Cdc45, the hetero-hexameric Mcm2–7 helicase, and the heterotrimeric GINS complex; it is essential for unwinding the template DNA duplex during replication

Translesion synthesis (TLS): a conserved process allowing cells to replicate damaged DNA templates; relies on specialized low-fidelity TLS polymerases

Fanconi anemia: rare human disorder resulting in bone marrow failure and cancer predisposition; characterized by an inability to repair DNA interstrand crosslinks

by the proteasome, which depends on DPC ubiquitylation, while cleavage by SPRTN does not (121). In either case, DPC proteolysis reduces the size of the adduct to a degree that allows specialized TLS polymerases to synthesize past the peptide remnant. The ability of TLS polymerases to accommodate damaged templates in their flexible active sites bears the cost of misincorporating
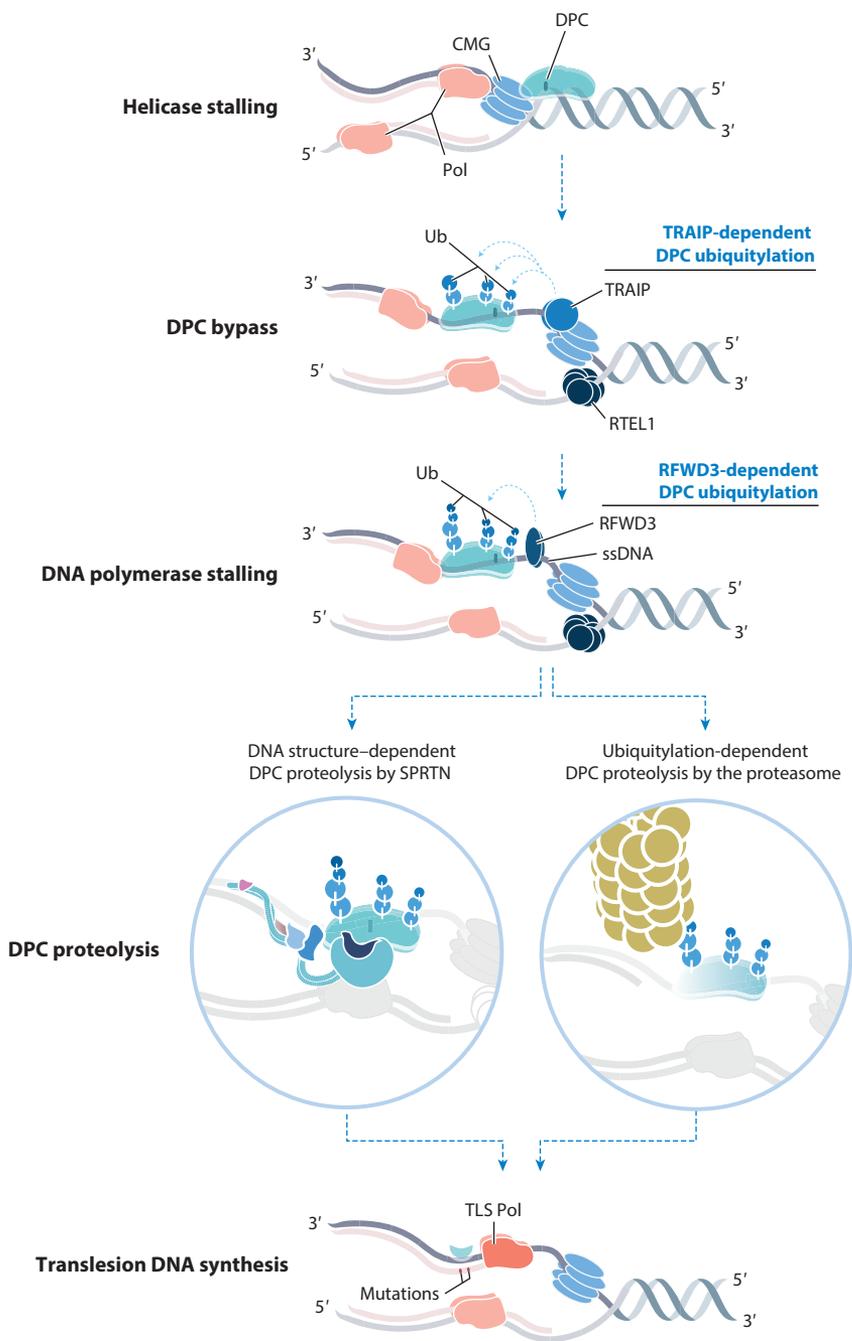
(*Caption appears on following page*)

Figure 3 (*Figure appears on preceding page*)

Replication-coupled DPC repair. A DPC on the leading strand stalls progression of the replicative CMG helicase, which causes ubiquitylation of the protein adduct by the CMG-associated ubiquitin E3 ligase TRAIP. Bypass of the protein adduct requires loading of the RTEL1 helicase on the lagging strand. RTEL1 has the opposite polarity to the CMG helicase, unwinding DNA in the $5'–3'$ direction. RTEL1 activity creates ssDNA downstream of the protein adduct, which enables bypass of CMG. Stalling of the replicative DNA polymerase at the DPC results in accumulation of ssDNA, which in turn recruits the ubiquitin E3 ligase RFWD3 via the ssDNA-binding protein RPA. RFWD3 further ubiquitylates the DPC. Polymerase progression requires DPC proteolysis, either by SPRTN or the proteasome. SPRTN-dependent DPC cleavage requires the presence of a single- to double-stranded DNA junction in direct proximity to the protein adduct but is independent of DPC ubiquitylation. In contrast, proteasomal degradation solely relies on ubiquitylation of the DPC. DNA synthesis across the peptide remnant requires specialized TLS polymerases, which are able to accommodate damaged templates but can cause mutagenesis due to misincorporation of incorrect nucleotides. Abbreviations: CMG, Cdc45–Mcm2–7–GINS; DPC, DNA–protein crosslink; Pol, polymerase; RPA, replication protein A; RTEL1, regulator of telomere elongation helicase 1; ssDNA, single-stranded DNA; TLS, translesion synthesis; Ub, ubiquitin.

incorrect nucleotides (129). Thus, replication-coupled DPC repair can cause mutagenesis (18). How the remaining peptide adducts are removed has not been formally shown, but it is commonly assumed that they are removed by endonucleases of the nucleotide excision repair (NER) pathway. To conclude, replication of DPC-containing templates relies on proteases debulking the lesion. However, DPC proteolysis also allows replication-independent DPC repair enzymes to access the lesion.

5. REPAIR BY CROSSLINK HYDROLYSIS

In general, DPC repair enzymes require broad specificity in order to tackle diverse DPCs. However, topoisomerase DPCs apparently occur with such high frequency in cells that specific repair enzymes evolved. Tyrosyl-DNA phosphodiesterase (TDP1) 1 and 2 directly hydrolyze the covalent bond between the DNA termini and the active-site tyrosines of TOP1 and TOP2, respectively (130). However, both enzymes require upstream processing of the respective topoisomerases in order to gain access to the covalent bond and release the protein adduct.
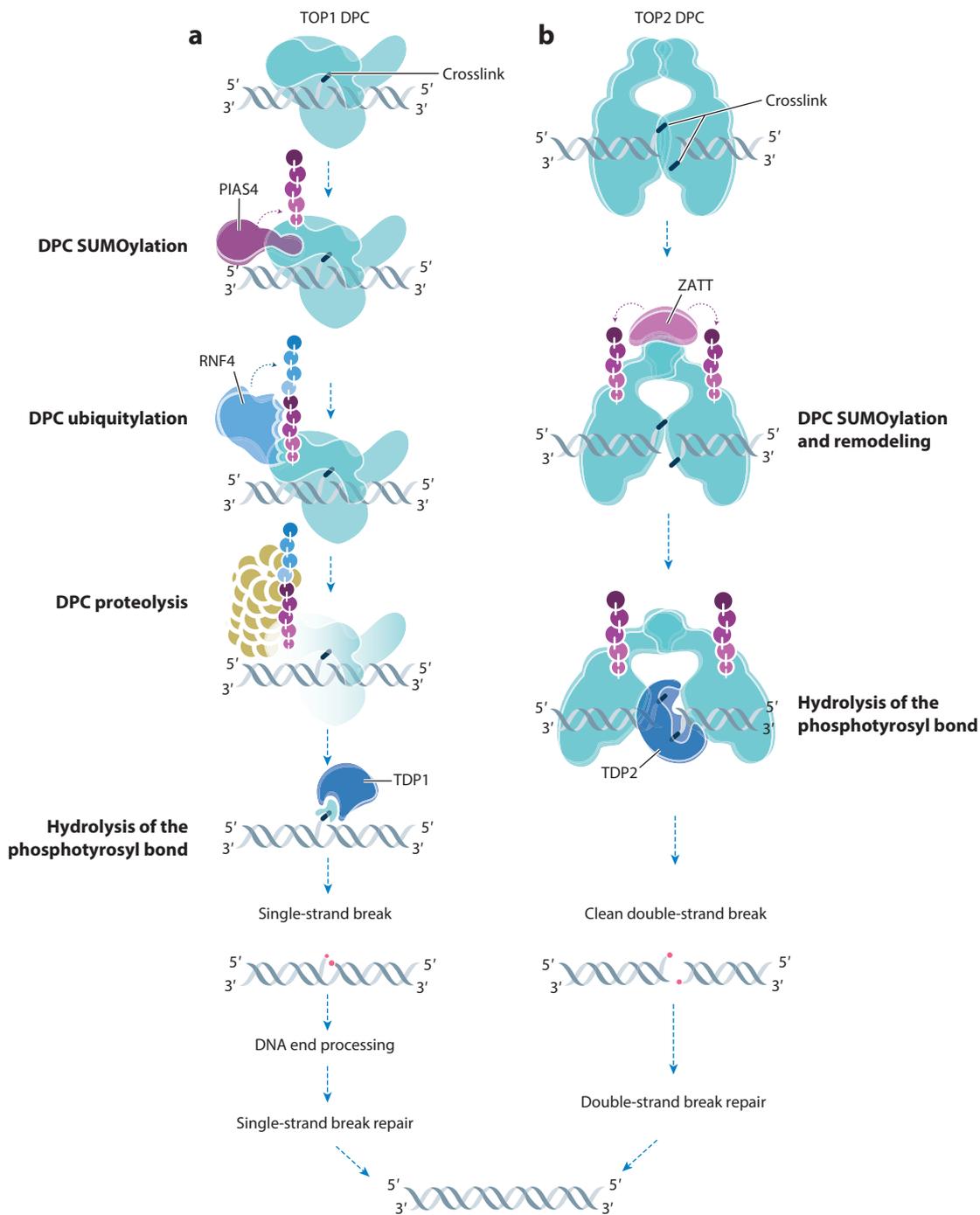
5.1. Tyrosyl-DNA Phosphodiesterase 1

TDP1 was initially discovered in yeast, but homologs are present in all eukaryotes (131, 132). TDP1 deficiency results in sensitivity toward TOP1-DPC-inducing agents and is synthetic lethal with loss of the DPC protease Wss1 in yeast (18, 131). Germline mutations in *TDP1* cause the recessive syndrome spinocerebellar ataxia with axonal neuropathy (SCAN1) (133, 134). Interestingly, SCAN1 is not only a consequence of defective TOP1 DPC repair. The SCAN1 H493R substitution causes TDP1 to become itself covalently trapped on DNA upon release of the covalent TOP1 adduct (135). SCAN1 is recessive because, in heterozygous patients, TDP1[H493R] DPCs are repaired by the remaining wild-type TDP1 (135). The substrate of TDP1—the covalent tyrosyl-DNA phosphodiester bond—is buried in the TOP1–DNA interface. Thus, repair by TDP1 requires proteolytic processing of the adduct by the proteasome (**Figure 4a**) (116, 136). Proteasomal degradation of TOP1 DPCs is initiated by SUMOylation, which in turn triggers polyubiquitylation by SUMO-targeted ubiquitin E3 ligases (STUbLs) (117, 137). Of note, the human STUbL RNF4 has recently been shown to also induce replication-independent proteasomal degradation of DNMT1 DPCs (138). Whether the polyubiquitylation signal is sufficient to trigger recruitment of proteasomes or whether additional factors are involved is currently not known. Upon degradation, TDP1 hydrolyzes the bond between the remaining TOP1 peptide and

Nucleotide excision repair (NER): DNA repair pathway that removes bulky DNA adducts by excising a short single-stranded DNA fragment containing the lesion

a TOP1 DPC

Crosslink

5′
3′

5′
3′

DPC SUMOylation

PIAS4

5′
3′

5′
3′

DPC ubiquitylation

RNF4

5′
3′

5′
3′

DPC proteolysis

5′
3′

5′
3′

**Hydrolysis of the
phosphotyrosyl bond**

TDP1

5′
3′

5′
3′

Single-strand break

5′
3′

5′
3′

DNA end processing

Single-strand break repair

b TOP2 DPC

Crosslink

5′
3′

5′
3′

ZATT

**DPC SUMOylation
and remodeling**

5′
3′

5′
3′

**Hydrolysis of the
phosphotyrosyl bond**

TDP2

5′
3′

5′
3′

Clean double-strand break

5′
3′

5′
3′

Double-strand break repair

5′
3′

5′
3′

(*Caption appears on following page*)

Figure 4 (*Figure appears on preceding page*)

DPC repair by direct hydrolysis. (*a*) Repair of TOP1 DPCs is initiated by polySUMOylation of the adduct by the SUMO E3 ligase PIAS4, which in turn triggers ubiquitylation by the SUMO-targeted ubiquitin E3 ligase RNF4. RNF4-dependent ubiquitylation leads to proteasomal degradation of the TOP1 DPC. Debulking of the DPC enables TDP1 to access and hydrolyze the covalent bond between the active-site tyrosine of TOP1 and the 3′ end of the break. Subsequent DNA end processing by PNKP allows repair of the break by the cellular single-strand break repair machinery. (*b*) Repair of TOP2 DPCs is initiated by SUMOylation of the adduct by the SUMO E3 ligase ZATT. SUMOylation remodels the crosslinked TOP2 adduct in a way that allows TDP2 to access and hydrolyze the covalent bonds between the active-site tyrosines of the TOP2 dimer and the 5′ ends of the double-strand break. Repair is completed by canonical double-strand break mechanisms. Abbreviations: DPC, DNA–protein crosslink; PNKP, polynucleotide phosphatase/kinase; SUMO, small ubiquitin-like modifier; TDP, tyrosyl-DNA phosphodiesterase; TOP, topoisomerase.

the 3′ end of the DNA. Of note, in addition to targeting TOP1 DPCs, TDP1 also repairs several other adducts at 3′ DNA termini (139, 140). TDP1 action generates a 3′-phosphate end, which requires dephosphorylation by bifunctional polynucleotide phosphatase/kinase (PNKP) (141). In addition, PNKP phosphorylates the 5′ end of the break, which allows the canonical single-strand break repair machinery to complete repair (141).

5.2. Tyrosyl-DNA Phosphodiesterase 2

The tyrosyl-DNA phosphodiester bonds of TOP2 DPCs occur at 5′ DNA ends and are hydrolyzed by TDP2 (142, 143). Repair of TOP2 DPCs by TDP2 can be initiated by proteasomal degradation of the protein adduct involving initial SUMOylation and subsequent ubiquitylation as described for TOP1 DPCs (117, 144, 145). However, TDP2 does not rely exclusively on proteolysis to gain access to its substrate. SUMOylation of trapped TOP2 by the SUMO E3 ligase ZATT remodels the adduct in a way that enables TDP2 to access and hydrolyze the covalent crosslink (**Figure 4***b*) (146). The clean double-strand break produced by TDP2 bears phosphorylated 5′ DNA ends, which are directly ligated by the nonhomologous end joining pathway (147).

6. REPAIR BY NUCLEOLYTIC CLEAVAGE

DPCs occurring at specific DNA structures can be repaired by repurposing canonical DNA repair enzymes. The MRN complex plays a crucial role in double-strand break repair and contains the meiotic recombination 11 (MRE11) nuclease, the ATPase RAD50 and the regulatory Nijmegen-breakage syndrome protein 1 (NBS1) subunit (known as Xrs2 in yeast) (148). MRN initiates double-strand break repair by processing the ends of a break using its endo- and exonuclease activities. Using this activity to sense and remove protein adducts at the end of double-strand breaks is a strategy conserved throughout evolution from phages to pro- and eukaryotes (**Figure 5**) (71, 149–151). Protein adducts at DNA ends stimulate the endonuclease activity of MRN (or its respective counterparts in lower organisms) and result in clipping of the adduct (152–154). MRN is insensitive to the nature of the block, which enables it to process diverse DPCs. Yet it seems that yeast MRX is particularly important for repairing TOP2 DPCs as well as SPO11 DPCs during meiosis (71, 155, 156). Similarly, TOP2 DPC repair involves MRN in higher eukaryotes (154, 157, 158). Notably, MRN activity is entirely restricted to DNA ends. Structural analyses of the bacterial MR complex suggest that end sensing occurs when a DNA end slips through the ring formed by the coiled-coil domains of the two RAD50 subunits (**Figure 5**) (159). The ensuing ring-to-rod transition within the coiled-coil domains locks the MR complex adjacent to the protein adduct and triggers endonucleolytic cleavage (159). Importantly, this model suggests that TOP2ccs must be preprocessed in order to expose the double-strand break. It is conceivable that this occurs upon proteasomal degradation of TOP2 DPCs. However, ZATT-dependent remodeling may be sufficient, if it involves breaking the TOP2 dimer.

Figure 5

Nucleolytic repair of DPCs by the MRN complex. The MRN complex consists of two MRE11 nucleases, two RAD50 ATPases, and one or two of the regulatory subunit NBS1. MRN scans the genome for protein adducts at DNA double-strand breaks using its RAD50 subunits. DNA end detection involves the slipping of the DNA end with the protein adduct through the ring formed by the extended Rad50 coiled-coil domains. Importantly, this is possible only if the break is exposed, which could occur upon partial degradation or remodeling of the TOP2 DPCs. End detection triggers a ring-to-rod transition in the coiled-coil domains, which in turn activates the endonuclease activity of MRE11 and results in nucleolytic cleavage of the DPC. Upon clipping off the protein adducts at either end of the break, repair is subsequently completed by canonical double-strand break repair mechanisms. Abbreviations: DPC, DNA–protein crosslink; MRE11, meiotic recombination 11; MRN, MRE11–RAD50–NBS1 complex; NBS1, Nijmegen-breakage syndrome protein 1.
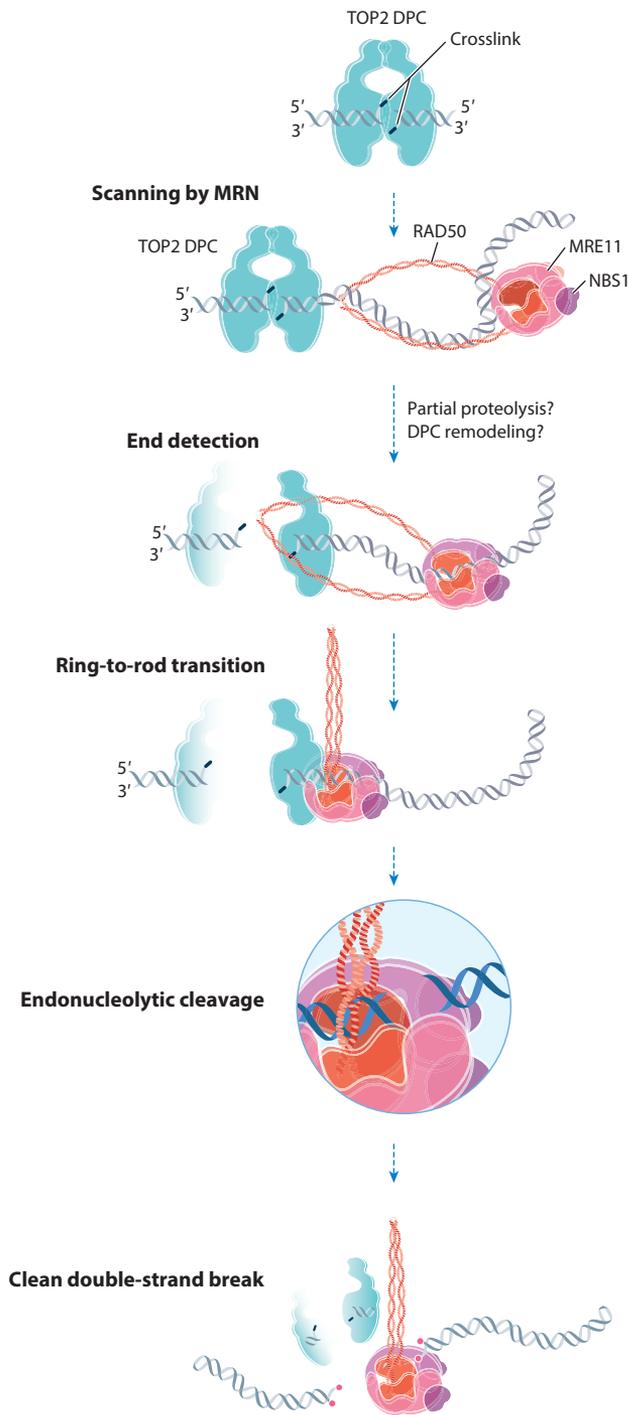
In addition to nucleolytic DPC removal at double-strand breaks, endonucleases of the NER pathway contribute to DPC repair in pro- and eukaryotes (119, 160–162). Importantly, in vitro data indicate that NER can excise small DPCs up to a size of 10 kDa (119, 160, 162). These data suggest that most DPCs require preprocessing by DPC proteases or the proteasome to enable excision of the peptide remnant by NER.

7. CONCLUDING REMARKS

The diversity and complexity of DPCs are reflected in the multitude of options available for repair. However, how the interplay between different DPC repair enzymes is organized and how pathway choice is determined remain relatively unclear. The role of Wss1 in DPC repair emerged because loss of Wss1 was found to be synthetic lethal with Tdp1 deficiency in yeast, suggesting that they operate in parallel repair pathways (18). However, the genetic relationship between SPRTN and TDP1 in mammalian cells is less clear. Some evidence points toward a similar situation as in yeast, while other data suggest synergistic activities of SPRTN and TDP1 (23, 96). In addition, the relationship between DPC proteolysis, TDP2-dependent hydrolysis, and MRN-dependent repair of TOP2 DPCs and SPO11 DPCs is largely unexplored. More-over, it remains unexplained why replication-coupled DPC repair involves at least two proteolytic enzymes—the proteasome and SPRTN. The essential role of SPRTN in mammalian cells suggests that some DPCs rely exclusively on its activity for repair. SPRTN does not require DPC ubiquitylation and may thus be needed if DPC ubiquitylation is not possible (perhaps in the absence of exposed lysines on the protein adduct). At any rate, decision making during DPC repair is likely governed by posttranslational modification of the DPC itself as well as of the surrounding chromatin proteins. The SUMO and ubiquitin systems appear central to the detection of and signaling from DPCs at the replication fork and beyond. SUMOylation appears to be a conserved early event upon DPC induction (89, 163, 164), as is the subsequent generation of a ubiquitin signal by STUbLs (117, 138, 163). However, a consensus regarding the mechanistic principles underlying DPC signaling and pathway choice has still to emerge. We propose that classical concepts of DNA repair pathway choice may not be sufficient to capture the complexity of DPCs. It is conceivable that multiple proteases as well as other DPC repair enzymes act synergistically on the protein. The relative importance of the individual repair enzyme may differ not only with respect to the identity of the protein adduct, the chemical nature of the crosslink, and the DNA structure involved but also with respect to the cellular context.

DPCs are a formidable challenge for cellular repair mechanisms and are central to the mode of action of many widely used chemotherapeutic agents, suggesting that quickly dividing cancer cells are particularly sensitive to DPCs. Thus, targeting DPC repair directly may be a viable strategy for cancer treatment in the future. Importantly, most synthetic lethal strategies exploiting specific genetic vulnerabilities rely on endogenous damage to kill cancer cells. DPCs appear to be particularly pervasive and abundant endogenous lesions. SPRTN is essential in mammalian cells, which is remarkable given that cells have various additional options available for DPC repair. Accordingly, cells must be constantly challenged by life-threatening levels of DPCs. In this context, it is interesting to speculate on why evolution has not selected against the cellular processes causing endogenous DPC formation. Histone and DNA methylation are suspected to cause DPCs, thus, it may be that DPC formation is a price worth paying for sophisticated epigenetic control mechanisms.

FUTURE ISSUES

1. The nature and source of endogenous DPCs remain to be discovered.

2. Methods to determine what proteins are crosslinked where in the genome in any given experimental scenario should be developed.

3. A complete understanding of the molecular mechanism underlying SPRTN's DNA structure–dependent activation is currently lacking.

4. The role of prokaryotic SPRTN and Wss1 homologs should be investigated.

5. The substrate preference and specificity of the ACRC, Ddi1 (DDI1, DDI2), and FAM111A/B proteases needs to be established in vitro.

6. The principles of DPC repair pathway choice remain largely unexplored.

7. Strategies to target DPC repair, and SPRTN in particular, for anticancer therapy need to be developed.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

LITERATURE CITED

1. Lindahl T. 1993. Instability and decay of the primary structure of DNA. *Nature* 362:709–15

2. Chatterjee N, Walker GC. 2017. Mechanisms of DNA damage, repair, and mutagenesis. *Environ. Mol. Mutagen.* 58:235–63

3. Schumacher B, Pothof J, Vijg J, Hoeijmakers JHJ. 2021. The central role of DNA damage in the ageing process. *Nature* 592:695–703

4. Roos WP, Thomas AD, Kaina B. 2016. DNA damage and the balance between survival and death in cancer biology. *Nat. Rev. Cancer* 16:20–33

5. Jackson SP, Bartek J. 2009. The DNA-damage response in human biology and disease. *Nature* 461:1071–78

6. Jeggo PA, Pearl LH, Carr AM. 2016. DNA repair, genome stability and cancer: a historical perspective. *Nat. Rev. Cancer* 16:35–42

7. Curtin NJ. 2012. DNA repair dysregulation from cancer driver to therapeutic target. *Nat. Rev. Cancer* 12:801–17

8. O'Connor MJ. 2015. Targeting the DNA damage response in cancer. *Mol. Cell* 60:547–60

9. O'Neil NJ, Bailey ML, Hieter P. 2017. Synthetic lethality and cancer. *Nat. Rev. Genet.* 18:613–23

10. Lord CJ, Ashworth A. 2017. PARP inhibitors: synthetic lethality in the clinic. *Science* 355:1152–58

11. Bryant HE, Schultz N, Thomas HD, Parker KM, Flower D, et al. 2005. Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature* 434:913–17

12. Farmer H, McCabe N, Lord CJ, Tutt AN, Johnson DA, et al. 2005. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* 434:917–21

13. Smith KC, O'Leary ME. 1967. Photoinduced DNA-protein cross-links and bacterial killing: a correlation at low temperatures. *Science* 155:1024–26

14. Moulé Y, Moreau S, Aujard C. 1980. Induction of cross-links between DNA and protein by PR toxin, a mycotoxin from *Penicillium roqueforti*. *Mutat. Res.* 77:79–89

15. Wilkins RJ, Macleod HD. 1976. Formaldehyde induced DNA–protein crosslinks in *Escherichia Coli*. *Mutat. Res.* 36:11–16

16. Solomon MJ, Varshavsky A. 1985. Formaldehyde-mediated DNA–protein crosslinking: a probe for in vivo chromatin structures. *PNAS* 82:6470–74

17. Barker S, Weinfeld M, Murray D. 2005. DNA–protein crosslinks: their induction, repair, and biological consequences. *Mutat. Res.* 589:111–35

18. **Stingele J, Schwarz MS, Bloemeke N, Wolf PG, Jentsch S. 2014. A DNA-dependent protease involved in DNA-protein crosslink repair. *Cell* 158:327–38**

19. Duxin JP, Dewar JM, Yardimci H, Walter JC. 2014. Repair of a DNA-protein crosslink by replication-coupled proteolysis. *Cell* 159:346–57

20. **Stingele J, Bellelli R, Alte F, Hewitt G, Sarek G, et al. 2016. Mechanism and regulation of DNA-protein crosslink repair by the DNA-dependent metalloprotease SPRTN. *Mol. Cell* 64:688–703**

21. **Vaz B, Popovic M, Newman JA, Fielden J, Aitkenhead H, et al. 2016. Metalloprotease SPRTN/DVC1 orchestrates replication-coupled DNA-protein crosslink repair. *Mol. Cell* 64:704–19**

22. Lopez-Mosqueda J, Maddi K, Prgomet S, Kalayil S, Marinovic-Terzic I, et al. 2016. SPRTN is a mammalian DNA-binding metalloprotease that resolves DNA-protein crosslinks. *eLife* 5:e21491

23. Maskey RS, Flatten KS, Sieben CJ, Peterson KL, Baker DJ, et al. 2017. Spartan deficiency causes accumulation of Topoisomerase 1 cleavage complexes and tumorigenesis. *Nucleic Acids Res.* 45:4564–76

24. Maskey RS, Kim MS, Baker DJ, Childs B, Malureanu LA, et al. 2014. Spartan deficiency causes genomic instability and progeroid phenotypes. *Nat. Commun.* 5:5744

25. Lessel D, Vaz B, Halder S, Lockhart PJ, Marinovic-Terzic I, et al. 2014. Mutations in SPRTN cause early onset hepatocellular carcinoma, genomic instability and progeroid features. *Nat. Genet.* 46:1239–44

26. Stingele J, Bellelli R, Boulton SJ. 2017. Mechanisms of DNA-protein crosslink repair. *Nat. Rev. Mol. Cell Biol.* 18:563–73

27. Nakano T, Xu X, Salem AMH, Shoulkamy MI, Ide H. 2017. Radiation-induced DNA–protein cross-links: mechanisms and biological significance. *Free Radic. Biol. Med.* 107:136–45

28. Tretyakova NY, Groehler AT, Ji S. 2015. DNA–protein cross-links: formation, structural identities, and biological outcomes. *Acc. Chem. Res.* 48:1631–44

29. Stingele J, Habermann B, Jentsch S. 2015. DNA–protein crosslink repair: proteases as DNA repair enzymes. *Trends Biochem. Sci.* 40:67–71

30. Nakano T, Ouchi R, Kawazoe J, Pack SP, Makino K, Ide H. 2012. T7 RNA polymerases backed up by covalently trapped proteins catalyze highly error prone transcription. *J. Biol. Chem.* 287:6562–72

31. Nakano T, Miyamoto-Matsubara M, Shoulkamy MI, Salem AM, Pack SP, et al. 2013. Translocation and stability of replicative DNA helicases upon encountering DNA-protein cross-links. *J. Biol. Chem.* 288:4649–58

32. Fu YV, Yardimci H, Long DT, Ho TV, Guainazzi A, et al. 2011. Selective bypass of a lagging strand roadblock by the eukaryotic replicative DNA helicase. *Cell* 146:931–41

33. Yeo JE, Wickramaratne S, Khatwani S, Wang YC, Vervacke J, et al. 2014. Synthesis of site-specific DNA-protein conjugates and their effects on DNA replication. *ACS Chem. Biol.* 9:1860–68

34. Kuo HK, Griffith JD, Kreuzer KN. 2007. 5-Azacytidine–induced methyltransferase-DNA adducts block DNA replication in vivo. *Cancer Res.* 67:8248–54

18. Identifies Wss1 as the first protease specifically dedicated to DPC repair.

20. Identifies SPRTN as the main DPC protease in higher eukaryotes.

21. Identifies SPRTN as the main DPC protease in higher eukaryotes.

35. Provides evidence
that HMCES protects
abasic sites during
replication by forming a
covalent DPC.

35. **Mohni KN, Wessel SR, Zhao R, Wojciechowski AC, Luzwick JW, et al. 2019. HMCES maintains genome integrity by shielding abasic sites in single-strand DNA.** *Cell* **176:144–53.e13**

36. Keeney S, Giroux CN, Kleckner N. 1997. Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell* 88:375–84

37. O'Brien PJ, Siraki AG, Shangari N. 2005. Aldehyde sources, metabolism, molecular toxicity mechanisms, and possible effects on human health. *Crit. Rev. Toxicol.* 35:609–62

38. Behrens UJ, Hoerner M, Lasker JM, Lieber CS. 1988. Formation of acetaldehyde adducts with ethanol-inducible P450IIE1 in vivo. *Biochem. Biophys. Res. Commun.* 154:584–90

39. Shi Y, Lan F, Matson C, Mulligan P, Whetstine JR, et al. 2004. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* 119:941–53

40. Tsukada Y, Fang J, Erdjument-Bromage H, Warren ME, Borchers CH, et al. 2006. Histone demethylation by a family of JmjC domain-containing proteins. *Nature* 439:811–16

41. Dingler FA, Wang M, Mu A, Millington CL, Oberbeck N, et al. 2020. Two aldehyde clearance systems are essential to prevent lethal formaldehyde accumulation in mice and humans. *Mol. Cell* 80:996–1012.e9

42. Burgos-Barragan G, Wit N, Meiser J, Dingler FA, Pietzke M, et al. 2017. Mammals divert endogenous genotoxic formaldehyde into one-carbon metabolism. *Nature* 548:549–54

43. Groehler A IV, Kren S, Li Q, Robledo-Villafane M, Schmidt J, et al. 2018. Oxidative cross-linking of proteins to DNA following ischemia-reperfusion injury. *Free Radic. Biol. Med.* 120:89–101

44. Ji S, Shao H, Han Q, Seiler CL, Tretyakova NY. 2017. Reversible DNA–protein cross-linking at epigenetic DNA marks. *Angew. Chem. Int. Ed.* 56:14130–34

45. Sczepanski JT, Wong RS, McKnight JN, Bowman GD, Greenberg MM. 2010. Rapid DNA–protein cross-linking and strand scission by an abasic site in a nucleosome core particle. *PNAS* 107:22475–80

46. Michaelson-Richie ED, Loeber RL, Codreanu SG, Ming X, Liebler DC, et al. 2010. DNA–protein cross-linking by 1,2,3,4-diepoxybutane. *J. Proteome Res.* 9:4356–67

47. Macfie A, Hagan E, Zhitkovich A. 2010. Mechanism of DNA–protein cross-linking by chromium. *Chem. Res. Toxicol.* 23:341–47

48. Loeber RL, Michaelson-Richie ED, Codreanu SG, Liebler DC, Campbell CR, Tretyakova NY. 2009. Proteomic analysis of DNA–protein cross-linking by antitumor nitrogen mustards. *Chem. Res. Toxicol.* 22:1151–62

49. Chvalova K, Brabec V, Kasparkova J. 2007. Mechanism of the formation of DNA–protein cross-links by antitumor cisplatin. *Nucleic Acids Res.* 35:1812–21

50. Ward JF. 1994. The complexity of DNA damage: relevance to biological consequences. *Int. J. Radiat Biol.* 66:427–32

51. Zhang H, Koch CJ, Wallen CA, Wheeler KT. 1995. Radiation-induced DNA damage in tumors and normal tissues. III. Oxygen dependence of the formation of strand breaks and DNA–protein crosslinks. *Radiat. Res.* 142:163–68

52. Fornace AJ Jr., Little JB. 1977. DNA crosslinking induced by X-rays and chemical agents. *Biochim. Biophys. Acta Nucleic Acids Protein Synth.* 477:343–55

53. Shoulkamy MI, Nakano T, Ohshima M, Hirayama R, Uzawa A, et al. 2012. Detection of DNA–protein crosslinks (DPCs) by novel direct fluorescence labeling methods: distinct stabilities of aldehyde and radiation-induced DPCs. *Nucleic Acids Res.* 40:e143

54. Christman JK. 2002. 5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. *Oncogene* 21:5483–95

55. Maslov AY, Lee M, Gundry M, Gravina S, Strogonova N, et al. 2012. 5-aza-2'-deoxycytidine-induced genome rearrangements are mediated by DNMT1. *Oncogene* 31:5172–79

56. Fenaux P, Mufti GJ, Hellstrom-Lindberg E, Santini V, Finelli C, et al. 2009. Efficacy of azacitidine compared with that of conventional care regimens in the treatment of higher-risk myelodysplastic syndromes: a randomised, open-label, phase III study. *Lancet Oncol.* 10:223–32

57. Daskalakis M, Nguyen TT, Nguyen C, Guldberg P, Kohler G, et al. 2002. Demethylation of a hypermethylated P15/INK4B gene in patients with myelodysplastic syndrome by 5-Aza-2'-deoxycytidine (decitabine) treatment. *Blood* 100:2957–64

58. Pommier Y, Marchand C. 2011. Interfacial inhibitors: targeting macromolecular complexes. *Nat. Rev. Drug Discov.* 11:25–36

59. Pommier Y, Sun Y, Huang SN, Nitiss JL. 2016. Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. *Nat. Rev. Mol. Cell Biol.* 17:703–21

60. Champoux JJ. 2001. DNA topoisomerases: structure, function, and mechanism. *Annu. Rev. Biochem.* 70:369–413

61. Liu LF, Desai SD, Li TK, Mao Y, Sun M, Sim SP. 2000. Mechanism of action of camptothecin. *Ann. N. Y. Acad. Sci.* 922:1–10

62. Sirikantaramas S, Yamazaki M, Saito K. 2008. Mutations in topoisomerase I as a self-resistance mechanism coevolved with the production of the anticancer alkaloid camptothecin in plants. *PNAS* 105:6782–86

63. Fujimori A, Harker WG, Kohlhagen G, Hoki Y, Pommier Y. 1995. Mutation at the catalytic site of topoisomerase I in CEM/C2, a human leukemia cell line resistant to camptothecin. *Cancer Res.* 55:1339–46

64. Pourquier P, Ueng LM, Kohlhagen G, Mazumder A, Gupta M, et al. 1997. Effects of uracil incorporation, DNA mismatches, and abasic sites on cleavage and religation activities of mammalian topoisomerase I. *J. Biol. Chem.* 272:7792–96

65. Pourquier P, Waltman JL, Urasaki Y, Loktionova NA, Pegg AE, et al. 2001. Topoisomerase I-mediated cytotoxicity of *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine: trapping of topoisomerase I by the O^6-methylguanine. *Cancer Res.* 61:53–58

66. Nitiss JL. 2009. Targeting DNA topoisomerase II in cancer chemotherapy. *Nat. Rev. Cancer* 9:338–50

67. Gothe HJ, Bouwman BAM, Gusmao EG, Piccinno R, Petrosino G, et al. 2019. Spatial chromosome folding and active transcription drive DNA fragility and formation of oncogenic MLL translocations. *Mol. Cell* 75:267–83.e12

68. Canela A, Maman Y, Huang SN, Wutz G, Tang W, et al. 2019. Topoisomerase II-induced chromosome breakage and translocation is determined by chromosome architecture and transcriptional activity. *Mol. Cell* 75:252–66.e8

69. Bernard P, Couturier M. 1992. Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. *J. Mol. Biol.* 226:735–45

70. Bergerat A, de Massy B, Gadelle D, Varoutas PC, Nicolas A, Forterre P. 1997. An atypical topoisomerase II from Archaea with implications for meiotic recombination. *Nature* 386:414–17

71. Neale MJ, Pan J, Keeney S. 2005. Endonucleolytic processing of covalent protein-linked DNA double-strand breaks. *Nature* 436:1053–57

72. Dheekollu J, Wiedmer A, Ayyanathan K, Deakyne JS, Messick TE, Lieberman PM. 2021. Cell-cycle-dependent EBNA1-DNA crosslinking promotes replication termination at *oriP* and viral episome maintenance. *Cell* 184:643–54.e13

73. Beard WA, Horton JK, Prasad R, Wilson SH. 2019. Eukaryotic base excision repair: New approaches shine light on mechanism. *Annu. Rev. Biochem.* 88:137–62

74. Cortez D. 2019. Replication-coupled DNA repair. *Mol. Cell* 74:866–76

75. Thompson PS, Amidon KM, Mohni KN, Cortez D, Eichman BF. 2019. Protection of abasic sites during DNA replication by a stable thiazolidine protein-DNA cross-link. *Nat. Struct. Mol. Biol.* 26:613–18

76. Halabelian L, Ravichandran M, Li Y, Zeng H, Rao A, et al. 2019. Structural basis of HMCES interactions with abasic DNA and multivalent substrate recognition. *Nat. Struct. Mol. Biol.* 26:607–12

77. Wang N, Bao H, Chen L, Liu Y, Li Y, et al. 2019. Molecular basis of abasic site sensing in single-stranded DNA by the SRAP domain of *E. coli* yedK. *Nucleic Acids Res.* 47:10388–99

78. Aravind L, Anand S, Iyer LM. 2013. Novel autoproteolytic and DNA-damage sensing components in the bacterial SOS response and oxidized methylcytosine-induced eukaryotic DNA demethylation systems. *Biol. Direct* 8:20

79. Mehta KPM, Lovejoy CA, Zhao R, Heintzman DR, Cortez D. 2020. HMCES maintains replication fork progression and prevents double-strand breaks in response to APOBEC deamination and abasic site formation. *Cell Rep.* 31:107705

80. Biayna J, Garcia-Cao I, Alvarez MM, Salvadores M, Espinosa-Carrasco J, et al. 2021. Loss of the abasic site sensor HMCES is synthetic lethal with the activity of the APOBEC3A cytosine deaminase in cancer cells. *PLOS Biol.* 19:e3001176

81. Shukla V, Halabelian L, Balagere S, Samaniego-Castruita D, Feldman DE, et al. 2020. HMCES functions in the alternative end-joining pathway of the DNA DSB repair during class switch recombination in B cells. *Mol. Cell* 77:384–94.e4

82. Kanellis P, Gagliardi M, Banath JP, Szilard RK, Nakada S, et al. 2007. A screen for suppressors of gross chromosomal rearrangements identifies a conserved role for PLP in preventing DNA lesions. *PLOS Genet.* 3:e134

83. Balakirev MY, Mullally JE, Favier A, Assard N, Sulpice E, et al. 2015. Wss1 metalloprotease partners with Cdc48/Doa1 in processing genotoxic SUMO conjugates. *eLife* 4:e06763

84. Enderle J, Dorn A, Beying N, Trapp O, Puchta H. 2019. The protease WSS1A, the endonuclease MUS81, and the phosphodiesterase TDP1 are involved in independent pathways of DNA–protein crosslink repair in plants. *Plant Cell* 31:775–90

85. Homchan A, Sukted J, Mongkolsuk S, Jeruzalmi D, Matangkasombut O, Pakotiprapha D. 2020. Wss1 homolog from *Candida albicans* and its role in DNA–protein crosslink tolerance. *Mol. Microbiol.* 114:409–22

86. Noguchi C, Grothusen G, Anandarajan V, Martinez-Lage Garcia M, Terlecky D, et al. 2017. Genetic controls of DNA damage avoidance in response to acetaldehyde in fission yeast. *Cell Cycle* 16:45–58

87. Reinking HK, Hofmann K, Stingele J. 2020. Function and evolution of the DNA-protein crosslink proteases Wss1 and SPRTN. *DNA Repair* 88:102822

88. Mosbech A, Gibbs-Seymour I, Kagias K, Thorslund T, Beli P, et al. 2012. DVC1 (C1orf124) is a DNA damage–targeting p97 adaptor that promotes ubiquitin-dependent responses to replication blocks. *Nat. Struct. Mol. Biol.* 19:1084–92

89. Borgermann N, Ackermann L, Schwertman P, Hendriks IA, Thijssen K, et al. 2019. SUMOylation promotes protective responses to DNA-protein crosslinks. *EMBO J.* 38:e101496

90. Delabaere L, Orsi GA, Sapey-Triomphe L, Horard B, Couble P, Loppin B. 2014. The Spartan ortholog maternal haploid is required for paternal chromosome integrity in the *Drosophila* zygote. *Curr. Biol.* 24:2281–87

91. Ruijs MW, van Andel RN, Oshima J, Madan K, Nieuwint AW, Aalfs CM. 2003. Atypical progeroid syndrome: an unknown helicase gene defect? *Am. J. Med. Genet. A* 116A:295–99

92. Morocz M, Zsigmond E, Toth R, Enyedi MZ, Pinter L, Haracska L. 2017. DNA-dependent protease activity of human Spartan facilitates replication of DNA–protein crosslink-containing DNA. *Nucleic Acids Res.* 45:3172–88

93. Li F, Raczynska JE, Chen Z, Yu H. 2019. Structural insight into DNA-dependent activation of human metalloprotease Spartan. *Cell Rep.* 26:3336–46.e4

94. Reinking HK, Kang HS, Gotz MJ, Li HY, Kieser A, et al. 2020. DNA structure-specific cleavage of DNA-protein crosslinks by the SPRTN protease. *Mol. Cell* 80:102–13.e6

95. Davis EJ, Lachaud C, Appleton P, Macartney TJ, Nathke I, Rouse J. 2012. DVC1 (C1orf124) recruits the p97 protein segregase to sites of DNA damage. *Nat. Struct. Mol. Biol.* 19:1093–100

96. Fielden J, Wiseman K, Torrecilla I, Li S, Hume S, et al. 2020. TEX264 coordinates p97- and SPRTN-mediated resolution of topoisomerase 1-DNA adducts. *Nat. Commun.* 11:1274

97. Zhao S, Kieser A, Li HY, Reinking HK, Weickert P, et al. 2021. A ubiquitin switch controls autocatalytic inactivation of the DNA–protein crosslink repair protease SPRTN. *Nucleic Acids Res.* 49:902–15

98. Centore RC, Yazinski SA, Tse A, Zou L. 2012. Spartan/C1orf124, a reader of PCNA ubiquitylation and a regulator of UV-induced DNA damage response. *Mol. Cell* 46:625–35

99. Perry M, Biegert M, Kollala SS, Mallard H, Su G, et al. 2021. USP11 mediates repair of DNA–protein cross-links by deubiquitinating SPRTN metalloprotease. *J. Biol. Chem.* 296:100396

100. Huang J, Zhou Q, Gao M, Nowsheen S, Zhao F, et al. 2020. Tandem deubiquitination and acetylation of SPRTN promotes DNA-protein crosslink repair and protects against aging. *Mol. Cell* 79:824–35.e5

101. Ruggiano A, Ramadan K. 2021. DNA–protein crosslink proteases in genome stability. *Commun. Biol.* 4:11

102. Carmell MA, Dokshin GA, Skaletsky H, Hu YC, van Wolfswinkel JC, et al. 2016. A widely employed germ cell marker is an ancient disordered protein with reproductive functions in diverse eukaryotes. *eLife* 5:e19993

94. Reveals that SPRTN achieves specificity by cleaving DPCs in a DNA structure–specific manner.

103. Bhargava V, Goldstein CD, Russell L, Xu L, Ahmed M, et al. 2020. GCNA preserves genome integrity and fertility across species. *Dev. Cell* 52:38–52.e10

104. Dokshin GA, Davis GM, Sawle AD, Eldridge MD, Nicholls PK, et al. 2020. GCNA interacts with Spartan and Topoisomerase II to regulate genome stability. *Dev. Cell* 52:53–68.e6

105. Kojima Y, Machida Y, Palani S, Caulfield TR, Radisky ES, et al. 2020. FAM111A protects replication forks from protein obstacles via its trypsin-like domain. *Nat. Commun.* 11:1318

106. Fine DA, Rozenblatt-Rosen O, Padi M, Korkhin A, James RL, et al. 2012. Identification of FAM111A as an SV40 host range restriction and adenovirus helper factor. *PLOS Pathog.* 8:e1002949

107. Unger S, Gorna MW, Le Bechec A, Do Vale-Pereira S, Bedeschi MF, et al. 2013. *FAM111A* mutations result in hypoparathyroidism and impaired skeletal development. *Am. J. Hum. Genet.* 92:990–95

108. Nie M, Oravcova M, Jami-Alahmadi Y, Wohlschlegel JA, Lazzerini-Denchi E, Boddy MN. 2021. FAM111A induces nuclear dysfunction in disease and viral restriction. *EMBO Rep.* 22:e50803

109. Hoffmann S, Pentakota S, Mund A, Haahr P, Coscia F, et al. 2020. FAM111 protease activity undermines cellular fitness and is amplified by gain-of-function mutations in human disease. *EMBO Rep.* 21:e50662

110. Serbyn N, Noireterre A, Bagdiul I, Plank M, Michel AH, et al. 2020. The aspartic protease Ddi1 contributes to DNA-protein crosslink repair in yeast. *Mol. Cell* 77:1066–79.e9

111. Svoboda M, Konvalinka J, Trempe JF, Grantz Saskova K. 2019. The yeast proteases Ddi1 and Wss1 are both involved in the DNA replication stress response. *DNA Repair* 80:45–51

112. Yip MCJ, Bodnar NO, Rapoport TA. 2020. Ddi1 is a ubiquitin-dependent protease. *PNAS* 117:7776–81

113. Dirac-Svejstrup AB, Walker J, Faull P, Encheva V, Akimov V, et al. 2020. DDI2 is a ubiquitin-directed endoprotease responsible for cleavage of transcription factor NRF1. *Mol. Cell* 79:332–41.e7

114. Lehrbach NJ, Ruvkun G. 2016. Proteasome dysfunction triggers activation of SKN-1A/Nrf1 by the aspartic protease DDI-1. *eLife* 5:e17721

115. Koizumi S, Irie T, Hirayama S, Sakurai Y, Yashiroda H, et al. 2016. The aspartyl protease DDI2 activates Nrf1 to compensate for proteasome dysfunction. *eLife* 5:e18357

116. Lin CP, Ban Y, Lyu YL, Desai SD, Liu LF. 2008. A ubiquitin-proteasome pathway for the repair of topoisomerase I-DNA covalent complexes. *J. Biol. Chem.* 283:21074–83

117. Sun Y, Miller Jenkins LM, Su YP, Nitiss KC, Nitiss JL, Pommier Y. 2020. A conserved SUMO pathway repairs topoisomerase DNA-protein cross-links by engaging ubiquitin-mediated proteasomal degradation. *Sci. Adv.* 6:eaba6290

118. Mao Y, Desai SD, Ting CY, Hwang J, Liu LF. 2001. 26 S proteasome-mediated degradation of topoisomerase II cleavable complexes. *J. Biol. Chem.* 276:40652–58

119. Baker DJ, Wuenschell G, Xia L, Termini J, Bates SE, et al. 2007. Nucleotide excision repair eliminates unique DNA-protein cross-links from mammalian cells. *J. Biol. Chem.* 282:22592–604

120. Quievryn G, Zhitkovich A. 2000. Loss of DNA–protein crosslinks from formaldehyde-exposed cells occurs through spontaneous hydrolysis and an active repair process linked to proteosome function. *Carcinogenesis* 21:1573–80

121. Larsen NB, Gao AO, Sparks JL, Gallina I, Wu RA, et al. 2019. Replication-coupled DNA-protein crosslink repair by SPRTN and the proteasome in *Xenopus* egg extracts. *Mol. Cell* 73:574–88.e7

122. Sparks JL, Chistol G, Gao AO, Raschle M, Larsen NB, et al. 2019. The CMG helicase bypasses DNA-protein cross-links to facilitate their repair. *Cell* 176:167–81.e21

123. Gallina I, Hendriks IA, Hoffmann S, Larsen NB, Johansen J, et al. 2021. The ubiquitin ligase RFWD3 is required for translesion DNA synthesis. *Mol. Cell* 81:442–58.e9

124. Vannier JB, Sarek G, Boulton SJ. 2014. RTEL1: functions of a disease-associated helicase. *Trends Cell Biol.* 24:416–25

125. Elia AE, Wang DC, Willis NA, Boardman AP, Hajdu I, et al. 2015. RFWD3-dependent ubiquitination of RPA regulates repair at stalled replication forks. *Mol. Cell* 60:280–93

126. Feeney L, Munoz IM, Lachaud C, Toth R, Appleton PL, et al. 2017. RPA-mediated recruitment of the E3 ligase RFWD3 is vital for interstrand crosslink repair and human health. *Mol. Cell* 66:610–21.e4

127. Inano S, Sato K, Katsuki Y, Kobayashi W, Tanaka H, et al. 2017. RFWD3-mediated ubiquitination promotes timely removal of both RPA and RAD51 from DNA damage sites to facilitate homologous recombination. *Mol. Cell* 66:622–34.e8

121. Shows parallel replication-coupled DPC proteolysis by SPRTN and the proteasome.

122. Describes the mechanism of DPC bypass by the CMG helicase.

128. Knies K, Inano S, Ramirez MJ, Ishiai M, Surralles J, et al. 2017. Biallelic mutations in the ubiquitin ligase RFWD3 cause Fanconi anemia. *J. Clin. Investig.* 127:3013–27

129. Sale JE. 2013. Translesion DNA synthesis and mutagenesis in eukaryotes. *Cold Spring Harb. Perspect. Biol.* 5:a012708

130. Pommier Y, Huang SY, Gao R, Das BB, Murai J, Marchand C. 2014. Tyrosyl-DNA-phosphodiesterases (TDP1 and TDP2). *DNA Repair* 19:114–29

131. Identifies Tdp1 as the enzyme releasing covalent Top1 DPCs in yeast.

131. Pouliot JJ, Yao KC, Robertson CA, Nash HA. 1999. Yeast gene for a Tyr-DNA phosphodiesterase that repairs topoisomerase I complexes. *Science* 286:552–55

132. Identifies Tdp1 as the enzyme releasing covalent Top1 DPCs in yeast.

132. Yang SW, Burgin AB Jr., Huizenga BN, Robertson CA, Yao KC, Nash HA. 1996. A eukaryotic enzyme that can disjoin dead-end covalent complexes between DNA and type I topoisomerases. *PNAS* 93:11534–39

133. Takashima H, Boerkoel CF, John J, Saifi GM, Salih MA, et al. 2002. Mutation of TDP1, encoding a topoisomerase I-dependent DNA damage repair enzyme, in spinocerebellar ataxia with axonal neuropathy. *Nat. Genet.* 32:267–72

134. El-Khamisy SF, Saifi GM, Weinfeld M, Johansson F, Helleday T, et al. 2005. Defective DNA single-strand break repair in spinocerebellar ataxia with axonal neuropathy-1. *Nature* 434:108–13

135. Interthal H, Chen HJ, Kehl-Fie TE, Zotzmann J, Leppard JB, Champoux JJ. 2005. SCAN1 mutant Tdp1 accumulates the enzyme–DNA intermediate and causes camptothecin hypersensitivity. *EMBO J.* 24:2224–33

136. Interthal H, Champoux JJ. 2011. Effects of DNA and protein size on substrate cleavage by human tyrosyl-DNA phosphodiesterase 1. *Biochem. J.* 436:559–66

137. Steinacher R, Osman F, Lorenz A, Bryer C, Whitby MC. 2013. Slx8 removes Pli1-dependent protein-SUMO conjugates including SUMOylated topoisomerase I to promote genome stability. *PLOS ONE* 8:e71960

138. Liu JCY, Kuhbacher U, Larsen NB, Borgermann N, Garvanska DH, et al. 2021. Mechanism and function of DNA replication-independent DNA-protein crosslink repair via the SUMO-RNF4 pathway. *EMBO J.* 40:e107413

139. Inamdar KV, Pouliot JJ, Zhou T, Lees-Miller SP, Rasouli-Nia A, Povirk LF. 2002. Conversion of phosphoglycolate to phosphate termini on 3′ overhangs of DNA double strand breaks by the human tyrosyl-DNA phosphodiesterase hTdp1. *J. Biol. Chem.* 277:27162–68

140. Zhou T, Akopiants K, Mohapatra S, Lin PS, Valerie K, et al. 2009. Tyrosyl-DNA phosphodiesterase and the repair of 3′-phosphoglycolate-terminated DNA double-strand breaks. *DNA Repair* 8:901–11

141. El-Khamisy SF. 2011. To live or to die: a matter of processing damaged DNA termini in neurons. *EMBO Mol. Med.* 3:78–88

142. Identifies TDP2 as a specific DNA repair enzyme for the repair of TOP2 DPCs.

142. Cortes Ledesma F, El Khamisy SF, Zuma MC, Osborn K, Caldecott KW. 2009. A human 5′-tyrosyl DNA phosphodiesterase that repairs topoisomerase-mediated DNA damage. *Nature* 461:674–78

143. Schellenberg MJ, Appel CD, Adhikari S, Robertson PD, Ramsden DA, Williams RS. 2012. Mechanism of repair of 5′-topoisomerase II-DNA adducts by mammalian tyrosyl-DNA phosphodiesterase 2. *Nat. Struct. Mol. Biol.* 19:1363–71

144. Wei Y, Diao LX, Lu S, Wang HT, Suo F, et al. 2017. SUMO-targeted DNA translocase Rrp2 protects the genome from Top2-induced DNA damage. *Mol. Cell* 66:581–96.e6

145. Gao R, Schellenberg MJ, Huang SY, Abdelmalak M, Marchand C, et al. 2014. Proteolytic degradation of topoisomerase II (Top2) enables the processing of Top2·DNA and Top2·RNA covalent complexes by tyrosyl-DNA-phosphodiesterase 2 (TDP2). *J. Biol. Chem.* 289:17960–69

146. Schellenberg MJ, Lieberman JA, Herrero-Ruiz A, Butler LR, Williams JG, et al. 2017. ZATT (ZNF451)-mediated resolution of topoisomerase 2 DNA-protein cross-links. *Science* 357:1412–16

147. Gomez-Herreros F, Romero-Granados R, Zeng Z, Alvarez-Quilon A, Quintero C, et al. 2013. TDP2-dependent non-homologous end-joining protects against topoisomerase II-induced DNA breaks and genome instability in cells and in vivo. *PLOS Genet.* 9:e1003226

148. Paull TT. 2018. 20 years of Mre11 biology: no end in sight. *Mol. Cell* 71:419–27

149. Stohr BA, Kreuzer KN. 2001. Repair of topoisomerase-mediated DNA damage in bacteriophage T4. *Genetics* 158:19–28

150. Connelly JC, de Leau ES, Leach DR. 2003. Nucleolytic processing of a protein-bound DNA end by the *E. coli* SbcCD (MR) complex. *DNA Repair* 2:795–807

151. Lee KC, Padget K, Curtis H, Cowell IG, Moiani D, et al. 2012. MRE11 facilitates the removal of human topoisomerase II complexes from genomic DNA. *Biol. Open* 1:863–73

152. Cannavo E, Cejka P. 2014. Sae2 promotes dsDNA endonuclease activity within Mre11-Rad50-Xrs2 to resect DNA breaks. *Nature* 514:122–25

153. Saathoff JH, Kashammer L, Lammens K, Byrne RT, Hopfner KP. 2018. The bacterial Mre11-Rad50 homolog SbcCD cleaves opposing strands of DNA by two chemically distinct nuclease reactions. *Nucleic Acids Res*. 46:11303–14

154. Deshpande RA, Lee JH, Arora S, Paull TT. 2016. Nbs1 converts the human Mre11/Rad50 nuclease complex into an endo/exonuclease machine specific for protein-DNA adducts. *Mol. Cell* 64:593–606

155. Hartsuiker E, Mizuno K, Molnar M, Kohli J, Ohta K, Carr AM. 2009. Ctp1CtIP and Rad32Mre11 nuclease activity are required for Rec12Spo11 removal, but Rec12Spo11 removal is dispensable for other MRN-dependent meiotic functions. *Mol. Cell. Biol.* 29:1671–81

156. Hartsuiker E, Neale MJ, Carr AM. 2009. Distinct requirements for the Rad32(Mre11) nuclease and Ctp1(CtIP) in the removal of covalently bound topoisomerase I and II from DNA. *Mol. Cell* 33:117–23

157. Aparicio T, Baer R, Gottesman M, Gautier J. 2016. MRN, CtIP, and BRCA1 mediate repair of topoisomerase II–DNA adducts. *J. Cell Biol.* 212:399–408

158. Hoa NN, Shimizu T, Zhou ZW, Wang ZQ, Deshpande RA, et al. 2016. Mre11 is essential for the removal of lethal Topoisomerase 2 covalent cleavage complexes. *Mol. Cell* 64:580–92

159. **Kashammer L, Saathoff JH, Lammens K, Gut F, Bartho J, et al. 2019. Mechanism of DNA end sensing and processing by the Mre11-Rad50 complex. *Mol. Cell* 76:382–94.e6**

160. Nakano T, Morishita S, Katafuchi A, Matsubara M, Horikawa Y, et al. 2007. Nucleotide excision repair and homologous recombination systems commit differentially to the repair of DNA-protein crosslinks. *Mol. Cell* 28:147–58

161. de Graaf B, Clore A, McCullough AK. 2009. Cellular pathways for DNA repair and damage tolerance of formaldehyde-induced DNA-protein crosslinks. *DNA Repair* 8:1207–14

162. Nakano T, Katafuchi A, Matsubara M, Terato H, Tsuboi T, et al. 2009. Homologous recombination but not nucleotide excision repair plays a pivotal role in tolerance of DNA-protein cross-links in mammalian cells. *J. Biol. Chem.* 284:27065–76

163. Serbyn N, Bagdiul I, Noireterre A, Michel AH, Suhandynata RT, et al. 2021. SUMO orchestrates multiple alternative DNA-protein crosslink repair pathways. *Cell Rep.* 37:110034

164. Ruggiano A, Vaz B, Kilgas S, Popović M, Rodriguez-Berriguete G, et al. 2021. The protease SPRTN and SUMOylation coordinate DNA-protein crosslink repair to prevent genome instability. *Cell Rep.* 37:110080

159. Describes the mechanism of DNA end detection by the MR complex.

Contents

Errata

An online log of corrections to *Annual Review of Biochemistry* articles may be found at
http://www.annualreviews.org/errata/biochem

Identifying and monitoring covalent DNA-protein crosslinks by PxP

Pedro Weickert[1,2,#], Sophie Dürauer[1,2,#], Maximilian J. Götz[1,2], Hao-Yi Li[1,2] and Julian Stingele[1,2,*]

[1]Department of Biochemistry, Ludwig-Maximilians-Universität München, 81377 Munich, Germany.

[2]Gene Center, Ludwig-Maximilians-Universität München, 81377 Munich, Germany.

[#]Equal contribution

*Correspondence: stingele@genzentrum.lmu.de

[H1] Abstract

Covalent DNA-protein crosslinks (DPC) are pervasive DNA lesions that challenge genome stability. This protocol describes a versatile and sensitive strategy for the Purification of x-linked Proteins (PxP) to identify and monitor DPCs in mammalian cells. Here, we discuss the use of PxP to track DPCs formed by the enzyme HMCES or induced by metabolic or chemotherapeutic crosslinking agents including reactive aldehydes, topoisomerase poisons and DNMT1 inhibitors. PxP separates crosslinked and soluble proteins by electro-elution, thereby, overcoming limitations of established methods to analyse DPCs which rely on precipitation as the separating principle. For PxP, cells are harvested following exposure to a DPC-inducing agent, embedded in low-melt agarose plugs, and lysed under denaturing conditions. Following lysis, soluble proteins are extracted from the agarose plug by electro-elution, while genomic DNA and crosslinked proteins are retained in the plug. Crosslinked proteins can then be analysed by standard analytical techniques such as SDS-PAGE followed by western blotting or fluorescent staining. Alternatively, quantitative mass spectrometry-based proteomics can be used for the unbiased identification of DPCs. The isolation and analysis of DPCs by PxP can be performed by an experienced molecular or cell biologist within 2-3 days. The protocol has been optimized to study DPC induction and repair in mammalian cells but may also be adapted to other sample types including bacteria, yeast, and tissue samples.

[H1] Introduction

DNA-protein crosslinks (DPCs) are highly toxic DNA lesions that have emerged as important sources of genome instability[1]. DPCs can be classified as non-enzymatic or enzymatic[2]. Non-enzymatic DPC formation is induced by bifunctional crosslinking reagents, including platinum-based chemotherapeutic drugs, nitrogen mustards, and metabolic aldehydes[3]. An abundant endogenous reactive aldehyde is formaldehyde, which is generated during one-carbon metabolism and by various demethylation reactions[4]. Formaldehyde is also a relevant environmental toxin as a common air pollutant and tobacco smoke component[5]. Enzymatic DPCs arise upon stabilization of covalent enzyme-DNA reaction intermediates. Prominent examples are topoisomerases 1 (TOP1) and 2 (TOP2), which establish covalent linkages between their active site tyrosines and 3'- or 5'- DNA ends, respectively[6]. These covalent TOP1- and TOP2-DNA complexes are normally short-lived but can become stabilized by chemotherapeutic topoisomerase poisons (e.g., by camptothecin that targets TOP1 or by etoposide that targets TOP2)[7]. The chemotherapeutic drug decitabine (5-azadC, a deoxycytidine analogue), causes DPC formation by a distinct mechanism. 5-azadC is incorporated into DNA during replication, where it irreversibly entraps DNA Methyltransferase 1 (DNMT1)[8]. In contrast, Embryonic Stem-Cell Specific 5-Hydroxymethylcytosine Binding (HMCES) forms stable DPCs as part of its cellular function[9]. HMCES crosslinks to abasic (AP) sites within ssDNA via its N-terminal cysteine residue, thereby protecting the damaged DNA strand from spontaneous or enzymatic incision[10,11].

DPC repair involves the proteolytic destruction of the protein adduct by proteasomal degradation or by dedicated DPC proteases[12], such as Weak Suppressor

3

of Smt3 1 (Wss1)[13] and DNA Damage Inducible 1 (Ddi1)[14] in yeast and SprT-Like N-Terminal Domain (SPRTN)[15-17] and FAM111 Trypsin Like Peptidase A (FAM111A)[18] in higher eukaryotes. DPC proteolysis can be induced in a replication-coupled or replication-independent, global-genome manner[1]. Replication-coupled repair is initiated by the collision of a replication fork with the DPC[19]. While the replicative CMG helicase (formed by Cdc45, MCM2-7 and GINS) can bypass the DPC, DNA polymerases stall once they approach the protein adduct[20,21]. The resulting ssDNA-dsDNA junction activates DPC cleavage by the SPRTN protease[21,22], which in addition requires unfolding of the protein adduct by the Fanconi Anemia Group J Protein (FANCJ) helicase[23]. In parallel, ubiquitylation of the DPC by replisome-associated E3-ligases triggers degradation by the proteasome[21,24]. The resulting peptide remnant is eventually bypassed by translesion synthesis polymerases[19]. How DPCs are sensed during global-genome repair is not entirely understood but entails the modification of the protein adduct by SUMOylation[25], which recruits the SUMO-targeted ubiquitin E3-ligase RNF4[26,27]. RNF4 ubiquitylates the DPC, leading to its destruction by the proteasome[26,27] and SPRTN[28]. SPRTN is essential for viability in mammalian cells, which highlights the importance of DPC repair[29]. Furthermore, partial loss-of-function of SPRTN leads to premature aging and cancer predisposition in Ruijs-Aalfs syndrome[30], which is caused by premature stop codons that result in the deletion of a critical ubiquitin-binding domain at SPRTN's C-terminus[28].

DPCs and the mechanistic principles of their repair are studied using defined model substrates in reconstitution experiments[22,23,31,32] and in *Xenopus* egg extracts[19-21,23,24,27]. In yeast, the Flp-nick system has been used to generate site-specific DPCs mimicking covalent TOP1 adducts[14,33,34]. While centrifugation of cell lysates in caesium chloride gradients can detect DPC formation in mammalian cells[35], most

assays rely on precipitation to enrich crosslinked DNA or proteins. In the KCl-SDS assay[36], proteins are precipitated from denaturing lysates and quantification of co-precipitating crosslinked DNA indicates the extent of DPC formation. In contrast, the RADAR (Rapid Approach to DNA Adduct Recovery) assay[37] and several variations[38,39], are based on the precipitation of DNA and the identification of co-precipitating crosslinked proteins.

[H2] Development of the PxP protocol

Here, we describe a protocol for the detection and identification of DPCs by Purification of x-linked Proteins (PxP), which separates crosslinked proteins from soluble proteins using electro-elution instead of precipitation. The PxP assay was inspired by experiments investigating the nature of DNA binding by Structural Maintenance of Chromosomes (SMC) proteins in *Bacillus subtillis*[40]. To assess whether the ring structure formed by the bacterial SMC-proteins was topologically binding to chromosomal DNA, the ring interfaces were crosslinked and chromosomal DNA was immobilized in low-melt agarose plugs. In an electric field, bacterial SMC-proteins were retained by the chromosomal DNA in the plug and did not elute, which demonstrated that they were encircling the DNA molecule. We reasoned that a similar principle could be applied to analyse DPC formation in cells. PxP is thus based on the idea that crosslinked proteins cannot be separated from immobilized DNA in an electric field, while merely DNA bound proteins are efficiently eluted. During the development of the PxP protocol, we optimized lysis conditions (e.g., buffer composition and timing) to achieve complete lysis of mammalian cells within agarose plugs, while minimizing incubation times to avoid loss of crosslinked proteins due to hydrolysis. Additionally, we tested various parameters (e.g., varying number of cells, agarose concentrations)

to enable efficient electro-elution of non-crosslinked proteins. Finally, we explored different options to retrieve crosslinked proteins from plugs following electro-elution, including agarase digestion (alone or in combination with a nuclease treatment), thermal melting of the plug, and the use of centrifugal filters to remove remaining agarose.

Using PxP, we recently identified a role for the SPRTN protease in global-genome DPC repair that is compromised by Ruijs-Aalfs syndrome patient mutations[28].

[H2] Overview of the PxP protocol

The protocol is organised in three main steps describing the (1) induction of different types of DPCs in mammalian cells, (2) their isolation by PxP, and (3) their detection and identification.

1. Induction of different types of DPCs:

   - Induction of DPCs by the reactive metabolite formaldehyde. Formaldehyde efficiently crosslinks chromatin proteins to DNA with the majority of DPCs being formed by histone proteins[28].

   - Induction of TOP2-DPCs using the topoisomerase poisons etoposide. Etoposide intercalates in the topoisomerase-DNA interface, thereby stabilizing the naturally occurring covalent complex between the enzyme's active site tyrosines and DNA[7]. Consequentially, these DPCs swiftly revert once the topoisomerase poison is removed.

   - Induction of DNMT1-DPC formation by 5-azadC. Cells synchronized in early S-phase efficiently incorporate 5-azadC in newly synthesized DNA, resulting in the formation of DNMT1-DPCs in post-replicative chromatin[25].

- Induction of DPCs between HMCES and AP sites using UVC irradiation or CD437[9,41], a polymerase alpha (POLα) inhibitor, in cell lines expressing Flag-tagged HMCES variants. UVC irradiation causes AP site formation by directly damaging the DNA. In contrast, treatment with CD437 interferes with lagging-strand synthesis, thereby generating long stretches of ssDNA, which are highly prone to spontaneous AP site formation.

2. DPC purification by PxP (Fig. 1):

- Cells are embedded in agarose plugs by mixing with a low-melt agarose solution and casting into plug molds.

- Cells are lysed within the agarose plugs using a denaturing buffer, which disrupts the cell membrane and denatures all cellular proteins.

- Agarose plugs are inserted into the pockets of an SDS-PAGE gel resulting in electro-elution of non-crosslinked proteins from the agarose plugs during the ensuing electrophoresis. In contrast, DPCs and chromosomal DNA are retained in the plug.

3. Detection and identification of DPCs (Fig. 1):

- For analysis by western blotting, silver or fluorescent staining, plugs are melted at high temperature and mixed with SDS-PAGE sample buffer.

- For mass spectrometry-based identification of DPCs, plugs are fixed followed by in-plug tryptic digestion and standard Liquid Chromatography tandem Mass Spectrometry (LC-MS/MS) analysis.

[H2] Experimental design and required expertise

This protocol is optimized for the analysis of DPCs in cultured adherent mammalian cells and has been successfully used by us in various cell lines, including HeLa, U2OS,

RPE-1 and HAP1 cells. Cells should be carefully handled using sterile techniques to avoid contaminations and checked regularly for the presence of mycoplasma contamination. To ensure reproducible results, cells should not be confluent or starved before seeding for DPC induction.

The precise experimental design is determined by the type of DPCs to be investigated. Formaldehyde and topoisomerase poisons can be used for DPC induction in asynchronous cell populations. However, we recommend synchronizing cells in early/mid S-phase using a double-thymidine block to monitor the repair of 5-azadC-induced DNMT1-DPCs or HMCES-DPCs, because these crosslinks arise primarily in a replication-coupled manner. Additionally, it is critical to choose appropriate drug concentrations for DPC induction. Try to avoid high drug doses that cause cells to detach before harvesting for PxP, because dead cells can result in unspecific accumulation of non-crosslinked proteins in PxP samples. Moreover, the stability of the DPC must be considered when planning the experiment. For example, camptothecin and etoposide are non-competitive inhibitors of TOP1 and TOP2, respectively. Once they are removed, topoisomerase DPCs immediately revert. Therefore, we recommend to preincubate cells treated with topoisomerase poisons on ice for five minutes before harvesting to minimize crosslink reversal. Another consequence of the immediate reversal of topoisomerase DPCs is, that it is not possible to first induce them, wash out the drug and then track their repair over time. In contrast, the fate of DNMT1-, HMCES-, and formaldehyde-induced DPCs can be monitored by chasing with drug-free media following DPC induction. It is possible to include small molecule inhibitors, for example the proteasome inhibitor MG132, during the chase to assess the effect on DPC repair. When conducting first PxP experiments, we recommend using formaldehyde as a positive control; PxP samples obtained from

formaldehyde-treated cells should reveal a characteristic histone pattern when analysed by SDS-PAGE (see anticipated results).

For most types of experiments, one technical replicate per experimental condition is sufficient. However, only a limited number of cells can be embedded per plug, which can be problematic if the amount of DPCs is low or if the final samples need to be analysed by western blotting with several antibodies. In such cases, several plugs can be prepared per each experimental condition and eventually pooled, resulting in larger amounts of sample. Additionally or alternatively, PxP samples can be concentrated by trichloroacetic acid (TCA) precipitation after electro-elution. For identification of DPCs by mass spectrometry, we recommend preparing at least three biological replicates to enable label-free quantitation. Alternatively, Stable Isotope Labelling with Amino acids in Cell culture (SILAC)- or Tandem Mass Tag (TMT)-based quantitation could be used.

No special expertise is needed to perform the PxP protocol. At first, handling and transferring plugs can be challenging, but it typically only requires some practice. To analyse PxP-purified DPCs, basic knowledge of standard techniques such as SDS-PAGE and western blotting are sufficient. Identification of DPCs by mass spectrometry will require access to the expertise and specialized equipment of a proteomics facility.

[H2] Comparison with other methods and limitations of the protocol

The KCl-SDS assay[36] is a sensitive method to detect DPC formation but cannot be used to study the identity of the crosslinked proteins. The RADAR assay[37] and various derivatives can be used to identify crosslinked proteins but rely on DNA precipitation to isolate co-precipitating DPCs. The reliance on precipitation is a potential risk because the behaviour during precipitation is influenced by various features of the

crosslinked protein. In addition, precipitation caused by protein aggregation upon treatment with pleiotropic reactive agents such as formaldehyde may be mistaken for DPC formation. The PxP protocol overcomes these issues by removing non-crosslinked proteins through electro-elution rather than directly purifying crosslinked proteins. In addition, we developed an optional nuclease control which allows to distinguish between DPCs and co-purifying contaminants. For this control, DNA is digested within plugs prior to electro-elution, which causes crosslinked proteins to elute from the plug while unspecific contaminants remain.

Nonetheless, the PxP assay has limitations as well. First, DPCs induced by reversible inhibitors, such as camptothecin or etoposide, can be challenging to study, because cells must be harvested and embedded in agarose before denaturing lysis. Therefore, there is the risk that such DPCs are lost due to reversal prior to lysis. This is not an issue in KCl-SDS or RADAR assays, because there, the cells are lysed using a denaturing buffer directly in the culture dish. A second limitation of the PxP is that the number of cells that can be embedded in a plug is limited. Too many cells per plug result in inefficient lysis and can lead to high background signals in downstream analysis. This limitation can be mitigated by casting multiple agarose plugs per experimental condition.

[H1] Materials

[H2] Equipment

[H3] General equipment

- Biological safety cabinets (for instance, Thermo Fisher Scientific, Herasafe 2030i, cat. no. 51032330)

- CO_2 incubator (for instance, PHC, MCO-230AICUV)

- Water bath (for instance, Memmert, WNE10)

- Inverted routine tissue culture microscope (for instance, Nikon, ECLIPSE Ts2-FL)

- Laboratory fume hood

- Cold room between 4-10°C

- Centrifuge for 1.5 ml tubes (for instance, Eppendorf, centrifuge 5424R, cat. no. EP5404000138)

- Automated cell counter (for instance, Thermo Fisher Scientific, Countess II FL Automated)

- Countess cell counting chamber slides (Thermo Fisher Scientific, cat. no. C10228)

- Vacuum pump (for instance, Vacuubrand, cat. no. PC 3004 VARIO)

- Rotating shaker (for instance, Stuart, Rotators SB3, cat. no. 445-2101)

- Thermoblock (for instance, Eppendorf, ThermoMixer C, cat. no. 5382000015)

- Electrophoresis system for SDS-PAGE for Novex WedgeWell gels (for instance, Thermo Fisher Scientific, Mini-Gel-Tank, cat. no. A25977)

[H3] Consumables

- Tissue culture dishes 6, 10, and 15 cm (Sarstedt, cat. no. 422-83.3901, 422-83.3902, 422-83.3903)

- Cell scrapers (Starlab, cat. no. CC7600-0220)

- CHEF disposable plug molds (Bio-Rad, cat. no. 1703713)

- Plastic tweezers (Samco, cat. no. L760/01)

- SafeSeal tubes 1.5/5 ml (Sarstedt, cat. no. 72.706/72.701)

- Novex WedgeWell 12%, Tris-Glycine, 1.0 mm gel (Invitrogen, cat. no. XP00120BOX)

- 0.22 µm PES vacuum filter (Steritop 45 mm neck size, Millipore express PLUS 0.22 µm PES filter, 250 ml, Merck Millipore, cat. no. S2GPT02RE)

- Staining box for SDS-PAGE gels (for instance, VWR, cat. no. 216-4334)

- Immobilon-P membrane, PVDF, 0.45 µm (Merck Millipore, cat. no. IPVH00010)

[H3] For SDS-PAGE (optional)

- Electrophoresis system for SDS-PAGE (for instance, Thermo Fisher Scientific XCell4 SureLock Midi-Cell Electrophoresis System, cat. no. WR0100)

- Protein blotting cell (for instance, Bio-Rad, Criterion blotter with plate electrodes, cat. no. 1704070)

- Western blot imaging system (for instance, Bio-Rad, ChemiDoc MP Imaging System)

- Scanner (for instance, Epson, Perfection V850 Pro, cat. no. B11B224401)

- Image analysing software (for instance, ImageJ software)

[H3] For UVC treatment (optional)

- UVC irradiator (for instance, Analytik Jena, UVP crosslinker CL-1000, cat. no. 849-30101-2)

[H3] For mass spectrometry data analysis (optional)

- R (version 4.2.2)

- RStudio (version 2023.06.0+421)[42]

- preprocessCore (version 1.64)[43]

- MSnbase (version 2.24.0)[44,45]

- Limma (version 3.54.0)[46]

- ggplot2 (version 3.4.2)[47]

- fdrtool (version 1.2.17)[48]

[H2] Biological materials (used in examples shown here)

- HAP1 cells (Horizon, cat. no. C631)

- HeLa T-REx Flp-In cells

- HeLa T-REx Flp-In cells expressing doxycycline-inducible HMCES (WT or a catalytically compromised C2S variant in which the catalytic cysteine is replaced by serine) with a C-terminal mVenus-3xFlag tag

[H2] Reagents

- Dulbecco modified eagle media (DMEM) (Gibco, cat. no. 41966052)

- Iscove's modified Dulbecco's medium (IMDM) (Gibco, cat. no. 12440061)

- Penicillin-streptomycin-glutamine (PSG) (Gibco, cat. no. 10378016)

- Fetal bovine serum (FBS) (Gibco, cat. no. 10437028)

- Dimethyl sulfoxide (DMSO) (ROTH, cat. no. 4720.4)

- Phosphate-buffered saline (PBS) pH 7.4 (ROTH, cat. no. 1111.1)

- TrypLE express enzyme, 1x (Gibco, cat. no. 12604021)

- Trypan blue solution, 0.4% (Gibco, cat. no. 15250061)

- Low-melt agarose (Bio-Rad, cat. no. 1613111)

- Sarkosyl solution, 20% (Sigma, cat. no. L7414)

- UltraPure ethylenediaminetetraacetic acid (EDTA), 0.5 M, pH 8 (Invitrogen, cat. no. 15575-038)

- UltraPure Tris-HCl, 1 M, pH 8 (Invitrogen, cat. no. 15568-025)

- Magnesium chloride (MgCl$_2$), 1 M (Invitrogen, cat. no. AM9530G)

- Benzonase (Merck Millipore, cat. no. 70746)

- cOmplete, EDTA-free protease inhibitor cocktail mini (Merck Millipore, cat. no. 4693132001)

- Pefabloc SC (Merck Millipore, 11585916001)

- Bolt MOPS SDS running buffer (Invitrogen, cat. no. B0001-02)

- LDS sample buffer, 4x (Thermo Fisher Scientific, cat. no. NP0007)

- Bolt reducing agent, 10x (Thermo Fisher Scientific, cat. no. B0009)

- Coomassie-based protein stain (for instance, GRP, Der Blaue Jonas, cat. no. GRP1)

[H3] For DPC induction described here (optional)

- Thymidine powder (Sigma, cat. no. T9250)

- Formaldehyde, methanol-free, 16% (Pierce, cat. no. 28906)

  !CAUTION Exposure to formaldehyde through inhalation can lead to airway irritation, bronchospasm, and pulmonary edema. Handle in a fume hood, wearing protective gloves and discard according to local regulations.

- Etoposide (Sigma, cat. no. 341205)

  !CAUTION Etoposide is a toxic compound. Handle in a fume hood, wearing protective gloves and discard according to local regulations.

- Doxycycline hyclate (Sigma, cat. no. D9891)

!CAUTION Doxycycline hyclate is a toxic compound. Handle in a fume hood, wearing protective gloves and discard according to local regulations.

- 5-azadC (Sigma, cat. no. A3656)

!CAUTION 5-azadC is a toxic compound. Handle in a fume hood, wearing protective gloves and discard according to local regulations.

- CD437 (Hölzel Diagnostika, cat. no. HY-100532)

[H3] For mass spectrometry sample preparation (optional)

- B Braun cutfix stainless steel scalpel (Thermo Fisher scientific, cat. no. 5518083)

- Acetic acid (ROTH, cat. no. 64-19-7)

- Absolute ethanol (ROTH, cat. no. K928.4)

- Seppro ammonium bicarbonate buffer, 2 M (Sigma, cat. no. S2454)

[H3] For SDS-PAGE (optional)

- SDS-PAGE gels (for instance, Thermo Fisher Scientific, NuPAGE 4-12%, Bis-Tris, 1.0 mm, Midi protein gel, 20-well, cat. no. WG1402BOX or NuPAGE 12 %, Bis-Tris, 1,0 mm, Mini protein gel, 12-well, cat. no. NP0342BOX)

- SilverQuest Staining Kit (Invitrogen, cat. no. LC6070)

- Glacial acetic acid, 100% (ROTH, cat. no. 3738.1)

    !CAUTION Glacial acetic acid can cause damage to eyes and skin. Handle in a fume hood, wearing protective gloves and safety goggles. Store in a safety cabinet and discard according to local regulations.

- Methanol, high-performance technical grade (VWR, cat. no. 20903.368P)

    !CAUTION Methanol is highly flammable and volatile and is toxic upon inhalation or contact. Keep away from ignition sources, wear protective

gloves, and avoid inhalation, swallowing and contact with skin and discard according to local regulations.

- SYPRO ruby protein gel stain (Thermo Fisher Scientific, cat. no. S12000)

- NuPAGE transfer buffer, 20x (Thermo Fisher Scientific, cat. no. NP0006)

- Tris base (ROTH, cat. no. 77-86-1)

- Tween 20 (Sigma, cat. no. P7949)

- Sodium chloride (NaCl) (ROTH, cat. no. 7647-14-5)

- Bovine serum albumin (BSA) (Sigma-Aldrich, cat. no. A9418-500G)

- Powdered milk (ROTH, cat. no. 68514-61-4)

- Sodium azide (ROTH, cat. no. 26628-22-8)

  !CAUTION Sodium azide can cause damage to eyes and skin, is toxic and environmentally hazardous. Handle in a fume hood, wearing protective gloves and safety googles. Store in a safety cabinet and discard according to local regulations.

- ECL substrate for western blot detection (for instance, Thermo Fisher Scientific, SuperSignal West Pico or Femto PLUS chemiluminescent substrate, cat. no. 34577, 34095)

[H3] *For Trichloroacetic acid precipitation (optional)*

- Trichloroacetic acid (TCA), (Sigma, cat. no. T6399)

  !CAUTION Trichloroacetic acid can cause damage to eyes and skin. Handle in a fume hood, wearing protective gloves and discard according to local regulations.

[H3] *Primary antibodies used in this protocol (optional)*

- Histone H3 (Cell signalling, cat. no. 4499S) (1:1000 dilution in TBS-T containing 5% milk)

- Actin (Santa cruz, cat. no. sc-47778) (1:1000 dilution in TBS-T containing 2.5% BSA)

- Vinculin (Santa cruz, cat. no. sc-73614) (1:2000 dilution in TBS-T containing 2.5% BSA)

- HMCES (Human altas, cat. no. HPA044968) (1:500 dilution in TBS-T containing 2.5% BSA)

- HMCES (Santa cruz, cat. no. sc-514238) (1:500 dilution in TBS-T containing 2.5% BSA)

- Flag-M2 (Sigma, cat. no. F3165) (1:2000 dilution in TBS-T containing 2.5% BSA)

- TOP2 (Abcam, cat. no. ab109524) (1:1000 dilution in TBS-T containing 2.5% BSA)

- DNMT1 (Cell signalling, clone D63A, cat. no. #5032) (1:1000 dilution in TBS-T containing 2.5% BSA)

- GAPDH (Cell signalling, 14C10, cat. no. 2118) (1:2000 dilution in TBS-T containing 2.5% BSA)

[H2] Reagent Setup

[H3] Low-melt agarose solution

Final concentration: 2% low-melt agarose, 1x PBS

For 100 ml of low-melt agarose solution, weigh 2 g low-melt agarose and transfer to a 250 ml bottle. Add 100 ml 1x PBS and screw the lid onto the bottle only loosely. Microwave in pulses until the agarose is dissolved. While still hot, the low-melt agarose solution can be aliquoted in 1.5 ml plastic tubes and stored at -20°C. Aliquots are thawed directly before use as described in the experimental procedure below. Aliquots

can be stored for at least 24 months but should not be reused after thawing. TROUBLESHOOTING.

!CAUTION Use adequate equipment when handling hot solutions. Safety goggles, gloves, and a heat-protective holder for handling the bottle. Do not shake the agarose solution violently.

[H3] Lysis buffer (1 ml is required per agarose plug, prepare fresh)

Final concentration: 1x PBS, 2% sarkosyl, 0.5 mM EDTA, 1x cOmplete EDTA-free protease inhibitor cocktail, 0.04 mg/ml Pefabloc SC

For 10 ml of lysis buffer, add to a 15 ml conical tube 1 ml 10x PBS, 1 ml 20% sarkosyl, 10 µl 0.5 M EDTA, 1 mini EDTA-free protease inhibitor cocktail tablet, and 40 µl Pefabloc SC stock solution (10 mg/ml dissolved in deionized water, aliquoted and stored at -20°C). Fill up to 10 ml with deionized water. Rotate at 4°C until the protease inhibitor tablet has dissolved completely, store individual components at room temperature (RT) and the final buffer at 4°C. The buffer can be stored at 4°C for at least 8 h.

[H3] Wash buffer (optional for nuclease control, 1.5 ml per agarose plug, prepare fresh)

Final concentration: 50 mM Tris-HCl pH 8, 0.01% sarkosyl, 1 mM $MgCl_2$

For 20 ml of wash buffer, add to a 50 ml conical tube 1 ml 1 M Tris-HCl pH 8, 10 µl 20% sarkosyl and 20 µl 1 M $MgCl_2$. Fill up to 20 ml with deionized water. Add 200 U benzonase per ml to an aliquot of this buffer (250 µl per plug). Store individual components at RT and the final buffer at 4°C. The buffer can be stored at 4°C for at least 8 h.

[H3] Thymidine media (optional)

Final concentration: 2 mM thymidine in standard culture media, 10% FBS, for IMDM +1% PSG

For 500 ml of thymidine media, weigh 250 mg thymidine powder and transfer it to a 500 ml bottle of culture media (for cell lines used here DMEM or IMDM). Close the bottle, shake vigorously and place in a water bath at 37°C for 30 min with intermittent shaking. Once the thymidine powder has dissolved and no thymidine precipitates are visible, sterilize the media by filtering through a 0.22 µm PES vacuum filter in a sterile tissue culture hood. 55 ml FBS are then added to reach a final concentration of 10%. For IMDM 5 ml PSG are added as well, final concentration 1%. Thymidine media without FBS and PSG can be stored up to 6 months at 4°C. Once FBS and PSG are added, we recommend to store the media for a maximum of 1 month at 4°C.

[H3] Etoposide stock solution (optional)

Final concentration: 50 mM etoposide in DMSO

Carefully resuspend 25 mg etoposide with 850 µl DMSO by pipetting up and down. Aliquot in 1.5 ml tubes and freeze at -80°C. The aliquot in use can be transferred and stored at -20°C. Aliquots are stable for at least 1 year. Before use, thaw at RT, discard aliquots after 2-3 freeze-thaw cycles.

!CAUTION Etoposide is a toxic compound. Handle in a fume hood, wearing protective gloves and discard according to local regulations.

[H3] Doxycycline stock solution (optional)

Final concentration: 50 mg/ml doxycycline in 1x PBS

Dissolve 1 g doxycycline hyclate in 20 ml 1x PBS in a 50 ml conical tube. Aliquot in 1.5 ml tubes and store at -20°C for up to 2 years. The working concentration is 1 µg/ml and therefore, we recommend preparing a second dilution using 20 µl of the 50 mg/ml doxycycline solution and 980 µl 1x PBS. This will yield a 1 mg/ml (1000x) solution which can be stored at -20°C for up to 3 months. Do not reuse thawed diluted (1 mg/ml) aliquots.

!CAUTION Doxycycline hyclate is a toxic compound. Handle in a fume hood, wearing protective gloves and discard according to local regulations.

[H3] 5-azadC stock solution (optional)

Final concentration: 50 mM 5-azadC in DMSO

Carefully resuspend 5 mg 5-azadC in 438 µl DMSO by pipetting up and down. Aliquot in 1.5 ml tubes and store at -20°C. Aliquots are stable for at least 1 year. Before use, thaw at RT, discard aliquots after 2-3 freeze-thaw cycles.

!CAUTION 5-azadC is a toxic compound. Handle in a fume hood, wearing protective gloves and discard according to local regulations.

[H3] CD437 stock solution (optional)

Final concentration: 10 mM CD437 in DMSO

Carefully resuspend 10 mg CD437 in 2.5095 ml DMSO by pipetting up and down. Aliquot in 1.5 ml tubes and store at -80°C for up to 6 months. The aliquot in use can be stored at -20°C for up to 1 month. Before use, thaw at RT, discard aliquots after 2-3 freeze-thaw cycles.

[H3] Fixation solution (optional for mass spectrometry sample preparation or SYPRO ruby staining)

Final concentration: 40% ethanol, 10% glacial acetic acid

For 500 ml, measure 200 ml absolute ethanol in a 500 ml measuring cylinder, add 50 ml glacial acetic acid and fill up to 500 ml with deionized water. Transfer to a 500 ml bottle, close, mix by inversion, and store at RT. The fixation solution is stable for at least 1 year at RT.

!CAUTION Prepare in a fume hood and store in a safety cabinet.

[H3] Destaining solution (optional for SYPRO ruby staining)

Final concentration: 10% methanol, 7% glacial acetic acid

For 500 ml, measure 50 ml methanol in a 500 ml measuring cylinder, add 35 ml glacial acetic acid and fill up to 500 ml with deionized water. Transfer to a 500 ml bottle, close, mix by inversion, and store at RT. The destaining solution is stable for at least 1 year at RT.

!CAUTION Prepare in a fume hood and store in a safety cabinet.

[H3] Ammonium bicarbonate buffer (optional for mass spectrometry sample preparation, prepare fresh)

Final concentration: 100 mM ammonium bicarbonate in deionized water

The commercial buffer stock solution (2 M) can be aliquoted in 15 ml conical tubes and stored at -20°C for at least 2 years. Thaw one aliquot at RT the day of use. For 20 ml 100 mM ammonium bicarbonate buffer, mix 1 ml 2 M Seppro ammonium

bicarbonate buffer with 19 ml deionized water in a 50 ml conical tube. Mix by inversion. We prepare this buffer fresh the day of use and do not store it diluted. Once prepared, store at 4°C for no more than 8 h.

NuPAGE transfer buffer (optional for western blotting)

Final concentration: 1x, 10% methanol

For 2 l, add 50 ml commercial 20x NuPAGE transfer buffer to a 2 l cylinder. Add 1.75 l deionized water, fill up with methanol to reach 2 l, mix by inverting and transfer to a glass bottle. Transfer buffer can be stored for at least 1 month at 4°C.

Bolt MOPS SDS running buffer

Final concentration: 1x

For 500 ml, mix 25 ml commercial 20x Bolt MOPS SDS running buffer with 475 ml deionized water in a 500 ml bottle. Mix by inverting and store at RT. MOPS running buffer can be stored for at least 6 months at RT.

Tris-buffered saline containing 0.1% Tween (TBS-T)

Final concentration: 50 mM Tris, 150 mM NaCl, 0.1% Tween, pH 7.6

For 1 L of 10x TBS solution, weigh 61 g Tris base and 88 g NaCl, transfer to a 1 l beaker and add 750 ml deionized water. Use a magnetic steerer for mixing. Once everything has dissolved, add HCl dropwise to adjust the pH to 7.6 at 25°C. Fill up to 1 l with deionized water, mix by inversion. TBS 10x can be stored at RT for at least 2 months.

For 1 L of 1x TBS-T, dilute 100 ml 10x TBS in 900 ml deionized water, mix by inverting. Cut the end of a 1 ml pipette tip with scissors and add 1 ml Tween 20 to the buffer (flush the tip by pipetting up and down). Use a magnetic steerer to mix until the detergent has fully dissolved, store at RT for no more than 2 weeks.

[H3] TBS-T containing BSA or milk powder (optional for primary antibody dilutions required for western blotting)

Final concentration: 2.5% BSA or 5% milk powder dissolved in TBS-T (optionally 0.1% sodium azide for long-term storage)

For 100 ml, weigh 2.5 g BSA or 5 g milk powder, in a 100 ml bottle and add 100 ml TBS-T. Optionally, add 100 mg sodium azide. Dissolve using a magnetic steerer. The final solution can be stored at 4°C for 2 days (without sodium azide) or for at least 1 year (with sodium azide). Sodium azide should not be added to TBS-T BSA/milk solutions used for blocking or secondary antibody solutions.

!CAUTION Sodium azide can cause damage to eyes and skin, is toxic and environmentally hazardous. Handle in a fume hood, wearing protective gloves and safety goggles. Store in a safety cabinet and discard according to local regulations.

[H3] Trichloroacetic acid (TCA)

Final concentration: 70% (w/v)

Weigh 35 g TCA powder in a 50 ml conical tube, add deionized water and dissolve on a rotating wheel or equivalent at RT. Once dissolved, measure the volume in a tube and add deionized water up to 50 ml. The TCA solution is stable for at least 2 years at 4°C.

Procedure

Part 1. DPC induction. TIMING 2-3 days

1| Induce covalent DNA-protein crosslinks using a DCP-inducing agent of choice or with one of the four example procedures described in this protocol. Perform each experiment at least 3 times and use appropriate negative and positive controls for the chosen DPC-inducing agent.

Table 1. Agents used for DPC induction in this protocol.

A	*Formaldehyde*	Induces DPCs by unspecific crosslinking of chromatin proteins.
B	*Etoposide*	Induces formation of TOP2-DPCs.
C	*5-azadC (decitabine)*	Induces formation of postreplicative DNMT1-DPCs.
D	*UVC irradiation/POLα-inhibition*	Induce replication-coupled formation of HMCES-DPCs.

A. DPC induction using formaldehyde. TIMING 2 days, handling 2-3 h

   I. **Day 1:** In the evening, prewarm media (IMDM + 10% FBS + 1% PSG) in a water bath at 37ºC.

  II. Harvest HAP1 cells cultured in IMDM media containing 10% FBS and 1% PSG. For the example described here one 10-cm dish with a confluency of 70-80% will be sufficient. Wash HAP1 cells with 5 ml 1x PBS followed by trypsinization (trypLE, 1 ml). Once cells are detached, inactivate trypsin with 9 ml media and transfer the cell suspension to a 15 ml conical tube.

 III. Mix an equal volume of cell suspension with trypan blue (for example, 50 µl and 50 µl) and determine the number of viable cells using an automated cell counter.

A 10-cm dish at 70-80% confluency should yield a minimum of 10-15 million cells.

IV. Seed 1.5 million cells per condition in fresh 6-cm dishes, adding at least 3 ml media. The example described here requires 3 dishes, which includes one additional dish for cell counting. If you plan to perform the optional nuclease control seed 3 million cells per dish, because more cells will be required to cast at least two plugs per condition.

V. **Day 2:** In the morning, prewarm media (IMDM + 10% FBS + 1% PSG) in a water bath at 37ºC.

VI. Trypsinize one dish and determine the approximate number of cells per dish using an automated cell counter. The number of cells will be required in Part 2, step 5.

VII. Dilute formaldehyde to 1 M with 1x PBS before adding it to the media. This is done by mixing 100 µl 16% formaldehyde solution (5.3 M) with 430 µl 1x PBS (final concentration 1 M). After mixing, add 1 µl 1 M formaldehyde solution per ml media (for example 5 µl in 5 ml) to reach a final concentration of 1 mM (or add 2 µl 1 M formaldehyde solution per ml media for a final concentration of 2 mM). For treating a 6-cm dish, a minimum of 2.5 ml media is required.

VIII. Aspirate the media from both dishes and add formaldehyde-containing media to one dish. Add fresh media without formaldehyde to the other dish as a non-treated control. Incubate for 1 h at 37°C.

IX. After incubation, place dishes on ice. Remove the media, wash once with 3 ml 1x PBS, add 1 ml ice-cold 1x PBS and quickly scrape cells.

X. Transfer cells to 1.5 ml plastic tubes and centrifuge at 500 g for 5 min at 4°C in a precooled centrifuge. Remove the supernatant. Freeze cell pellets at -80°C. PAUSE POINT. Cell pellets can be stored at -80°C for up to 1 week.

B. *Etoposide treatment to induce TOP2-DPCs. TIMING 2 days, handling 2-3 h.*

   I. **Day 1:** In the evening, prewarm media (IMDM + 10% FBS + 1% PSG) in a water bath at 37°C.

   II. Harvest HAP1 cells cultured in IMDM media containing 10% FBS and 1% PSG. For the example described here one 10-cm dish with a confluency of 70-80% will be sufficient. Wash HAP1 cells with 5 ml 1x PBS followed by trypsinization (trypLE, 1 ml). Once cells are detached, inactivate trypsin with 9 ml media and transfer the cell suspension to a 15 ml conical tube.

   III. Mix an equal volume of cell suspension with trypan blue (for example, 50 µl and 50 µl) and determine the number of viable cells using an automated cell counter. A 10-cm dish at 70-80% confluency should yield a minimum of 10-15 million cells.

   IV. Seed 1.5 million cells per condition in fresh 6-cm dishes, adding at least 3 ml media. The example described here requires 4 dishes (1 non-treated control, 2 etoposide concentrations and 1 extra dish for determining the number of cells).

   V. **Day 2:** In the morning, prewarm media (IMDM + 10% FBS + 1% PSG) in a water bath at 37°C.

   VI. Trypsinize the cells from one dish and count them using an automated cell counter to determine the approximate number of cells per dish. The number of cells will be required in Part 2, step 5.

VII. Add 5 or 10 µl etoposide stock solution (50 mM) to 10 ml media to reach a final concentration of 25 µM and 50 µM, respectively. 2.5 ml media are required per dish.

VIII. Aspirate the media from all remaining three dishes. Add fresh media without drugs to one dish as a non-treated control. Add media containing 25 µM and 50 µM etoposide to the other two dishes. Incubate for 1 h at 37°C.

IX. After incubation, place dishes on ice for 5 min. CRITICAL STEP. Cooling the cells immediately limits reversal of topoisomerase-DPCs during further processing.

X. Remove the media, add 1 ml ice-cold 1x PBS and quickly scrape cells. Transfer cells to 1.5 ml plastic tubes and centrifuge at 500 g for 5 min at 4°C in a precooled centrifuge.

XI. Remove the supernatant, place tubes back on ice and rapidly proceed with sample processing in Part 2. CRITICAL STEP. These samples cannot be frozen or stored.

C. *5-azadC treatment to induce DNMT1-DPCs and monitor their repair. TIMING 3 days, handling 3-4 h*

I. **Day 1:** In the morning, prewarm media (DMEM + 10% FBS for HeLa T-REx Flp-In) in a water bath at 37ºC.

II. Harvest HeLa T-REx Flp-In cells cultured in DMEM +10% FBS respectively. For the example described here one 10-cm dish with a confluency of 70-80% will be sufficient. Wash HeLa T-REx Flp-In cells with 5 ml 1x PBS followed by trypsinization (trypLE, 1 ml). Once cells are detached, inactivate trypsin with 9 ml media and transfer the cell suspension to a 15 ml conical tube.

III. Mix an equal volume of cell suspension with trypan blue (for example, 50 µl and 50 µl) and determine the number of viable cells using an automated cell counter. A 10-cm dish at 70-80% confluency should yield a minimum of 6-8 million cells.

IV. Seed 750,000 cells per condition in fresh 6-cm dishes, adding at least 3 ml media. The example described here requires 7 dishes, which includes one additional dish to determine the number of cells.

V. Allow cells to attach for at least 8 h.

VI. In the evening, prewarm thymidine media (DMEM + 10% FBS + 2 mM thymidine) in a water bath at 37ºC.

VII. Confirm that cells have attached using a microscope.

VIII. Initiate the synchronization of cells by a double-thymidine block by aspirating the media from all dishes and carefully add 2.5 ml thymidine media to each dish.

IX. Incubate cells overnight at 37°C.

X. **Day 2:** In the morning, prewarm media (DMEM + 10% FBS) in a water bath at 37ºC.

XI. Remove the thymidine media from the dishes and wash twice with 3 ml 1x PBS each to release cells from the first thymidine block. Add fresh media without thymidine and incubate for 8-9 h.

XII. In the evening, prewarm thymidine media (DMEM + 10 % FBS + 2 mM thymidine) in a water bath at 37ºC.

XIII. Repeat step VIII-IX.

XIV. **Day 3:** In the morning, prewarm media (DMEM + 10% FBS) in a water bath at 37ºC.

XV. Remove the thymidine media from the dishes and wash twice with 3 ml 1x PBS each. Add fresh media without thymidine to release cells from the second thymidine block.

XVI. 3 h following release, DNMT1-DPC formation can be induced.

XVII. Trypsinize one dish and determine the approximate number of cells per dish using an automated cell counter. The number of cells will be required in Part 2, step 5.

XVIII. Aspirate media from the 6 remaining dishes. Add a minimum of 2.5 ml fresh media containing 10 μM 5-azadC (add 4 μl 5-azadC stock solution (50 mM) to 20 ml prewarmed media without thymidine) to four dishes. Add media without 5-azadC to the remaining two non-treated control dishes. Incubate for 30 min at 37°C.

XIX. Following the 30 min incubation, place one non-treated control dish and one 5-azadC-treated dish on ice (timepoint 0 h).

XX. Proceed by rapidly washing the remaining 4 dishes twice with 3 ml 1x PBS each and add fresh 5-azadC-free media. Start timing the recovery time from the moment you replace the media. Let cells recover for the desired time (1, 2 and 3 h in the example described here) at 37°C. The remaining non-treated control dishes stay unchanged at 37°C (timepoint 3 h).

XXI. Remove the media from the dishes placed on ice in step XIX. Wash once with 3 ml 1x PBS, add 1 ml ice-cold 1x PBS and quickly scrape cells. Transfer cells to 1.5 ml plastic tubes and centrifuge at 500 g for 5 min at 4°C in a precooled centrifuge. Remove the supernatant. Freeze cell pellets at -80°C.

XXII. Repeat step XXI at every desired recovery timepoint (here 1, 2 and 3 h after removal of 5-azadC-containing media). PAUSE POINT. Cell pellets can be stored at -80°C for up to 1 month.

D. UVC irradiation or POLα-inhibition to induce HMCES-DPCs. TIMING 3 days, handling 3-4 h

I. **Day 1:** In the morning, prewarm media (DMEM + 10% FBS) in a water bath at 37ºC.

II. Harvest HeLa T-REx Flp-In cells expressing doxycycline-inducible HMCES-WT and HeLa T-REx Flp-In cells expressing doxycycline-inducible HMCES-C2S cultured in DMEM media containing 10% FBS. For the example described here one 10-cm dish with a confluency of 70-80% is sufficient per cell line. Wash cells with 5 ml 1x PBS followed by trypsinization (trypLE, 1 ml). Once cells are detached, inactivate trypsin with 9 ml media and transfer the cell suspension to a 15 ml conical tube.

III. Mix an equal volume of cell suspension with trypan blue (for example, 50 µl and 50 µl) and determine the number of viable cells using an automated cell counter. A 10-cm dish at 70-80% confluency should yield 6-8 million cells.

IV. Seed 750,000 cells per condition and cell line in fresh 6-cm dishes, adding at least 3 ml media. The example described here requires 3 dishes per cell line, which includes one additional dish per cell line to determine the number of cells.

V. Allow cells to attach for at least 8 h.

VI. In the evening, prewarm thymidine media (DMEM + 10% FBS + 2 mM thymidine) in a water bath at 37ºC.

VII. Confirm that cells have attached using a microscope.

VIII. Initiate the synchronization of cells by double-thymidine block by aspirating the media from all dishes and carefully add 2.5 ml thymidine media to each dish.

IX. Incubate cells overnight at 37°C.

X. **Day 2:** In the morning, prewarm media (DMEM + 10% FBS) in a water bath at 37ºC.

XI. Remove the thymidine media from the dishes and wash twice with 3 ml 1x PBS each to release cells from the first thymidine block. Add fresh media without thymidine and incubate for 8-9 h.

XII. In the evening, prewarm thymidine media (DMEM + 10% FBS + 2 mM thymidine) in a water bath at 37ºC.

XIII. Repeat step VIII-IX, but additionally add doxycycline to the thymidine media to induce HMCES expression (1 µg/ml final doxycycline concentration, add 20 µl of the diluted doxycycline stock solution (1 mg/ml) to 20 ml media in a 50 ml conical tube).

XIV. **Day 3:** In the morning, prewarm media (DMEM + 10% FBS) in a water bath at 37ºC.

XV. Confirm the successful induction of HMCES expression using a fluorescence microscope with appropriate excitation/emission filters for HMCES's mVenus-tag.

XVI. Remove the thymidine media from the dishes and wash twice with 3 ml 1x PBS each. Add fresh media without thymidine but containing doxycycline (1 µg/ml) to release cells from the second thymidine block.

XVII. 2 h after release, HMCES-DPC formation can be induced using UVC irradiation or POLα inhibition.

For UVC irradiation, wash all dishes with 3 ml 1x PBS, before adding 2 ml fresh 1x PBS. Irradiate one dish per cell line in a UVC irradiator with a dose of 50 J/m^2. Following irradiation, remove the 1x PBS and add fresh prewarmed doxycycline-free media. Also change the media of the non-treated control dishes and let cells recover for 6 h at 37 °C.

For POLα inhibition, aspirate media from all dishes. Add a minimum of 2.5 ml fresh media containing 5 μM CD437 (add 7.5 μl CD437 stock solution (10 mM) to 15 ml prewarmed media without thymidine) to one dish per cell line. Add media without CD437 to the remaining non-treated control dishes. Incubate for 1 h at 37°C.

XVIII. 6 h after UVC irradiation or 1 h after addition of POLα inhibitor, trypsinize one untreated control dish per cell line and determine the number of cells using an automated cell counter. The number of cells will be required in Part 2, step 5.

XIX. Place the remaining dishes on ice. Remove the media, wash once with 3 ml 1x PBS, add 1 ml ice-cold 1x PBS and quickly scrape cells. Transfer cells to 1.5 ml plastic tubes and centrifuge at 500 g for 5 min at 4°C in a precooled centrifuge. Remove the supernatant. Freeze cell pellets at -80°C. PAUSE POINT. Cell pellets can be stored at -80ºC for up to 1 month.

Part 2. DPC isolation by PxP. TIMING 6-7 h

2| Prepare lysis buffer and keep at 4°C on a roller to allow protease inhibitor tablet to dissolve.

3| Set a thermoblock to 80ºC and thaw one low-melt agarose aliquot for at least 10 min. Vortex once to ensure that it is melted entirely and well mixed.

4| Set a second thermoblock to 45ºC.

5| Resuspend cell pellets harvested in Part 1 in ice-cold 1x PBS to a concentration of 25,000 cells/µl based on the approximate number of cells determined in Part 1. In the examples described in Part 1, you should have obtained circa 2 million cells per condition, which would require resuspension with 80 µl of 1x PBS. CRITICAL STEP. Resuspend cells thoroughly by pipetting up and down with a 20-200 µl pipet. Make sure that no cell clumps remain and that cells are evenly distributed. Keep cell suspensions on ice all the time.

6| Remove a 10 µl aliquot from each cell suspension and transfer to fresh 1.5 ml plastic tubes containing 25 µl 4x LDS sample buffer, 10 µl 10x reducing agent and 55 µl water. Boil samples, which will be used as input controls, at 99ºC for 15-20 min.

7| Once the low-melt agarose solution at 80°C (step 2) is melted completely, transfer the tube to the 45°C thermoblock and allow equilibration for 1-2 min.

8| Place a plug mold and a vortexer next to the 45°C thermoblock.

9| Each cell suspension is processed individually for steps 9 and 10. Briefly vortex cells, transfer to the thermoblock set at 45ºC, and incubate for 45 s for volumes less than 100 µl and 60 s for larger volumes. CRITICAL STEP. Mix the liquid low-melt agarose solution with the cell suspension in a 1:1 ratio. The final agarose concentration is 1%. The low-melt agarose solution is viscous so pipette slowly to ensure accurate volumes. Mix cells and low-melt agarose solution slowly but thoroughly. Be careful to not introduce bubbles, because it is difficult to remove them afterwards. TROUBLESHOOTING.

10| Pipette the low-melt agarose/cell mixture into the plug mold until it is filled up (approximate volume circa 80-90 µl). CRITICAL STEP. While pipetting the mixture into the mold, it is important to prevent the formation of bubbles. Therefore, pipette the mixture slowly but steadily on the walls of the mold, positioning the pipette tip parallel to the plug mold (Fig. 2a, note that plugs were stained with bromophenol blue for better visualization, Video S1). This step must be swiftly completed to prevent premature solidification of the low-melt agarose. TROUBLESHOOTING.

11| Once all agarose plugs have been cast, place the mold in a fridge at 4ºC and incubate for at least 5 min until the low-melt agarose has completely solidified (Fig. 2a). In the meantime, precool one 1.5 ml plastic tube per agarose plug on ice. Prepare bigger tubes if plugs will be pooled, see next step. TROUBLESHOOTING.

12| To transfer the agarose plugs to the tubes, remove the adhesive sticker on the bottom of the mold. Ensure that it is removed completely, as it tends to rip apart easily. Then, push the agarose plugs out of the mold and into the tubes using the removable tool that is supplied with the mold (Fig. 2b, Video S2). Push from the bottom of the mold (where the sticker was placed). In this step, agarose plugs corresponding to the same condition can be pooled in the same tube. Choose the size of the tube accordingly: 1.5 ml tube for 1 plug, 2 ml tubes for 2 plugs, 5 ml tubes for up to 5 plugs. Large numbers of plugs (> 10) can be pooled in 50 ml conical tubes. We do not recommend using 15 ml conical tubes because it is difficult to retrieve plugs from these tubes.

13| Add 1 ml ice-cold lysis buffer per plug to the tubes. Use at least 25 ml lysis buffer if using 50 ml conical tubes for lysis. The agarose plugs should have a white opaque colour at this step (Fig. 2c).

14| Place the tubes on a rotating wheel at 4°C. Rotate for 4 h at 25 rpm to lyse cells within the agarose plugs.

15| After 3.5 h, start to prepare a running chamber for the electro-elution step. Prepare 300 ml fresh Bolt MOPS SDS running buffer and one Novex WedgeWell SDS-PAGE gel (12%, Tris-Glycine, 1.0 mm) per every 10 plugs. Do not yet unpack the gel.

16| After 4 h of lysis, the agarose plugs should be translucent and almost invisible in the lysis buffer (Fig. 2c). Place tubes on ice and remove as much lysis buffer as possible using a pipette or vacuum pump. CRITICAL STEP. If using a vacuum pump, attach a 10 μl tip and remove lysis buffer slowly. Be careful not to break or damage the agarose plugs at this step. TROUBLESHOOTING.

17| Optional nuclease control can be performed at this point (see section nuclease control, steps 18-24). Alternatively, continue with step 25.

Nuclease control (optional). TIMING 45 min, handling 15 min

CRITICAL. While not essential, we recommend including a nuclease control to confirm that a signal observed in PxP samples is stemming from DPC formation. Note that for this step at least two plugs per condition are required; each plug is placed in a different 1.5 ml tube.

18| Carefully aspirate the lysis buffer from each tube and replace with 1 ml cold wash buffer. Incubate plugs for 10 min at 4°C on a rotating wheel at 25 rpm.

The plugs may turn slightly opaque after the first wash, due to some co-precipitation of MgCl$_2$ and sarkosyl, but this will not affect the DNA digestion.

19| Label the tubes with the plugs that will be digested with nuclease. We recommend digesting half of the plugs per condition.

20| Aspirate wash buffer and add 250 µl wash buffer containing 200 U benzonase per ml to the samples that will be digested. Add wash buffer without benzonase to the remaining plugs.

21| Incubate all plugs at 37ºC for 30 min with vigorous shaking (500-600 rpm).

22| Place tubes on ice for 5 min.

23| Aspirate the buffer from all tubes and keep on ice.

24| Proceed with step 25 of the protocol.

25| Open the Novex WedgeWell SDS-PAGE gel packaging and place the gel in the running chamber. Fill the chamber with Bolt MOPS SDS running buffer prepared in Step 15. Once the running buffer has reached all wells, remove circa 50% of the buffer from the part of the chamber that contains the wells. CRITICAL STEP. The buffer level must be below the wells during loading of the plugs, but the wells must also be wetted prior to loading. If the buffer level is too high, plugs will float away during loading. TROUBLESHOOTING.

26| Use plastic tweezers to load the plugs into the wells of the Novex WedgeWell SDS-PAGE gel (Fig. 2d, Video S3). We recommend to briefly wet the tweezers in Bolt MOPS SDS running buffer before starting. Then try to balance the plug on one lever of the tweezers with half of the plug extending past the lever. Once the plug touches the wet well, let it slide from the tweezer lever into the well.

Gently push the plug with the tip of the tweezers into the well, if necessary. TROUBLESHOOTING.

27| Carefully re-fill the chamber with Bolt MOPS SDS running buffer. CRITICAL STEP. Make sure to not disturb the loaded plugs. If buffer is added too vigorously, some plugs may float out of the wells.

28| Connect the power supply and start the electro-elution at constant amperage (20 mA per gel) for 60 min. Initial voltage should be between 36-45 V. TROUBLESHOOTING.

29| Once electro-elution is completed, carefully remove the gel from the chamber and place it on a paper tissue. Remove residual buffer from the gel by placing each corner on a paper tissue and wait for excess buffer to be absorbed. Make sure that there is as little buffer remaining as possible before opening the gel with a spatula (Video S4). CRITICAL STEP. If buffer is not properly removed, plugs may change position during opening of the gels. It is important to make sure that each plug stays in position to prevent mixing of the different conditions. TROUBLESHOOTING.

30| Transfer electro-eluted plugs to fresh 1.5 ml plastic tubes using plastic tweezers (Fig. 2e, Video S4). At this step, a maximum number of 10 plugs corresponding to the same condition can be pooled together (See Experimental design).

31| If desired, the Novex WedgeWell SDS-PAGE gel used for electro-elution can be transferred to a plastic box for staining with a Coomassie-based protein stain to confirm successful and homogenous electro-elution of non-crosslinked proteins from the plugs (Fig. 3).

Part 3. DPC detection and identification. TIMING 2 h – 2 days

32| DPCs isolated by PxP can be characterised using different approaches. We have included procedures for SDS-PAGE followed by western blotting or fluorescent protein staining and analysis by mass spectrometry. Follow the steps in options A or B, respectively.

A   SDS-PAGE followed by western blot or fluorescent staining. TIMING 2 days, handling 1-3 h

 I. Place plugs for 2-3 min at 99ºC until molten.

 II. Centrifuge at 10,000 g for 30 s at RT.

 III. Molten plugs (ca. 80 µl) can be mixed directly with 40 µl 4x LDS sample buffer and 10 µl 10x reducing agent (proceed directly to Step *A* XV). Alternatively, samples can be further concentrated by TCA precipitation (Step A IV-XIV).

 IV. For TCA precipitation, add 80 µl wash buffer containing 200 U benzonase per ml to each tube.

 V. Vortex shortly and incubate at 37ºC for 30 min with vigorous shaking (500-600 rpm).

 VI. Add 1 U agarase per plug and vortex briefly.

 VII. Incubate at 42ºC for additional 30 min before placing samples on ice. If the agarase digestion was successful, the samples should not solidify anymore.

 VIII. Add 440 µl deionized water and mix by vortexing.

 IX. Add 100 µl 70% TCA and incubate for 20 min on ice.

 X. Centrifuge at high speed (>20,000 g) for 20 min at 4ºC in a precooled centrifuge.

 XI. Carefully remove supernatant and add 1 ml acetone (precooled to -20ºC).

 XII. Centrifuge again at high speed (>20,000 g) for 20 min at 4ºC. A white pellet should be visible at this step.

XIII. After removing the supernatant, place tubes with open lids in a fume hood to evaporate any remaining acetone, normally 10-20 min are enough.

XIV. Resuspend each pellet in 50 µl 1x LDS sample buffer with 1x reducing agent (12.5 µl 4x LDS sample buffer, 5 µl 10x reducing agent and 22.5 µl deionized water).

XV. Boil at 99ºC for 20 min to revert bonds between DNA and crosslinked proteins. TROUBLESHOOTING.

XVI. Samples containing agarose must be kept warm prior to loading on an SDS-PAGE gel, we therefore recommend boiling them just before loading the gel. After boiling they should stay liquid at RT for at least 15 min. If samples solidify, they can be melted again by boiling again at 99ºC for 2 min.

XVII. Resolve samples using standard SDS-PAGE gel electrophoresis. We suggest using ultrapure water for the preparation of running buffers. Given that formaldehyde mainly crosslinks small histone proteins, we recommend resolving formaldehyde-treated samples in 12% SDS-PAGE gels. To visualize larger adducts, DNMT1-DPCs, topoisomerase-DPCs or mVenus-3xFlag-tagged HMCES-DPCs, we recommend 4-12% gradient SDS-PAGE gels. Continue with step XVIII for silver staining, with step XIX for fluorescent staining or with step XXVI for analysis by western blotting.

XVIII. For silver staining, follow the manufacturer's instructions (SilverQuest Staining Kit, Invitrogen, cat. no. LC6070). We have successfully used the basic and the fast staining protocol for the detection of DPCs.

XIX. For fluorescent staining using SYPRO ruby protein gel stain we adapted the manufacturer's protocol. Transfer the SDS-PAGE gel to a plastic box containing

enough fixation solution to cover the gel (for the staining box specified in Equipment, 20 ml are sufficient).

XX. Place on a shaker with gentle agitation (20-25 rpm) for 30 min at RT.

XXI. Remove fixation solution and add sufficient SYPRO ruby protein stain to cover the gel. Wrap the box with aluminium foil and incubate overnight with gentle agitation at RT.

XXII. The next morning, remove the SYPRO ruby protein gel stain and cover the gel with destaining solution.

XXIII. Incubate with gentle agitation for 15 min at RT.

XXIV. Repeat steps XXII-XXIII.

XXV. Wash gel with ultrapure water for 5 min and proceed with imaging. Place the gel in the fluorescent imager and acquire an image using adequate emission/excitation settings. If using the ChemiDoc MP Imaging System, use the UV transilluminator for excitation and the 605/650nm emission filter. Avoid overexposure.

XXVI. To detect crosslinking of specific proteins by western blotting, transfer proteins from the SDS-PAGE gel to a 0.45 μm PVDF membrane that was activated with methanol. When using the Bio-Rad Criterion blotting system transfer in 1x NuPAGE transfer buffer for 50 min at 100 V for small proteins and 1 h and 10 min for larger proteins.

XXVII. Following transfer, block the membrane for 1 h in 5% milk in TBS-T.

XXVIII. If the antibody is diluted in TBS-T containing BSA, rinse the membrane three times with TBS-T to remove excess milk. If the antibody is diluted in TBS-T containing milk, remove the blocking solution, and directly add the antibody solution.

XXIX. Incubate overnight in the cold room in primary antibody diluted in TBS-T containing 2.5% BSA or 5% milk (See Reagents).

XXX. The next day, wash the membrane three times for 10 min with TBS-T.

XXXI. Add the corresponding secondary antibody diluted in 5% milk in TBS-T and incubate for 1 h.

XXXII. Wash the membrane at least three times for 10 min with TBS-T.

XXXIII. Incubate the membrane with an ECL substrate and acquire the chemiluminescent signal using a western blot imaging system. Avoid overexposure.

B   *DPC identification by mass spectrometry. TIMING variable, handling 2-6 h*

I. To prepare agarose plugs for mass spectrometry analysis, add 1 ml fixation solution to each plug and incubate on a rotating wheel for 1 h at 4°C at 25 rpm.

II. Aspirate the fixation solution.

III. Add 1 ml 100 mM ammonium bicarbonate buffer to each tube. Incubate plugs for additional 10 min on a rotating wheel at 4°C at 25 rpm.

IV. Remove buffer and repeat step III.

V. Aspirate the buffer. PAUSE POINT. Samples can be stored at 4°C overnight.

VI. Cut the plugs in smaller pieces (at least two) using clean scalpels on a sterile dish and transfer to fresh tubes. Use different scalpels and dishes for plugs of different conditions.

VII. Submit the cut agarose plugs to a mass spectrometry facility, which can process them with standard protocols for in-gel tryptic digestion[49].

VIII. Proceed with analysis of mass spectrometry data. The mass spectrometry facility will typically provide you with a spreadsheet containing information on

peptide identity and the respective intensities. Alternatively, proteins can be identified from raw mass spectrometry spectra using proteomics software such as MaxQuant. Several freely available programs can be used for statistical analysis of mass spectrometry data. We describe a workflow using R studio and the limma package. However, other programs, which are freely available, can also be considered (e.g. Perseus or DEP R package).

IX. Import the results into RStudio.

X. Remove reverse identified proteins and contaminants.

XI. Log2 transform the intensities.

XII. Remove proteins that were not identified in at least 75% of the replicates of at least one condition. Depending on the number of biological replicates and the desired stringency of the analysis, different requirements can be chosen.

XIII. Normalize the intensities between the biological replicates. We use the R package preprocessCore to perform quantile normalization.

XIV. Potentially remaining missing values have to be imputed at this stage. Here, we use the MinDet imputation method to impute values that are missing due to low protein abundance. Depending on the nature of missing values, different imputation methods should be considered. The R package MSnbase offers several imputation algorithms for randomly and non-randomly (left censored) missing data. It also allows for a hybrid method where data is both missing at random and not at random depending on user defined classification.

XV. Enrichment and statistical significance are calculated using the limma functions lmFit, eBayes and topTable based on a user defined design and contrast matrix that specify conditions and comparisons for all conditions that should be tested. As a default, we define comparisons for all possible combinations of conditions.

To adjust for multiple comparison, the false discovery rate is calculated from the t-statistic using the fdrtool function of the R package fdrtool. The FDR can be calculated from either the t-statistic or the p-value returned by limma.

XVI. Visualize data in R using dedicated graphics packages such as ggplot2. Alternatively, the results can be exported as a text file and further analysed using specialized software such as GraphPad Prism. The results can be visualized as a volcano plot, by plotting the log2 fold-change against the FDR.

Troubleshooting

Troubleshooting advice can be found in Table 2.

Table 2. Troubleshooting

STEP	PROBLEM	POSSIBLE REASON	SOLUTION
Reagent set-up	Clumps in low-melt agarose solution.	Low-melt agarose has not dissolved well.	When preparing the low-melt agarose stock solution, boil it carefully until all clumps are dissolved. Re-add 1x PBS if needed and keep concentration at 2%.
9	Cells are not evenly distributed in agarose plugs.	The cell suspension has not reached the correct temperature before mixing with the low-melt agarose solution,	Allow enough time for warming up cell suspension before mixing with low-melt agarose solution. Mix

		resulting in pre-mature solidification of the agarose. Alternatively, cell suspension and low-melt agarose solution were not mixed thoroughly, or cell pellets were not well resuspended.	cell suspension by pipetting just before placing it in the thermoblock.
10	Plugs contain bubbles.	The agarose/cell mixture was mixed too strong, introducing bubbles, which were transferred to the plug mold. Alternatively, bubbles were formed while filling up the plug mold with the agarose/cell mixture.	Mix low-melt agarose solution with cell suspension carefully and slowly fill up plug molds or use reverse pipetting. Bubbles in the mold can be removed by taking the air out of the bubble with a 10 µl pipette tip.
11	Plugs are too soft to transfer.	The volume of low-melt agarose solution was lower than the volume of cell suspension when mixed or the time for solidification in the fridge	Make sure that all liquid was removed from cell pellets before freezing. Increase incubation in the fridge. If the problem

		was not long enough. Alternatively, agarose stock has the wrong concentration.	persists, prepare a new agarose stock solution.
16	Plugs remain opaque after lysis	The number of cells per plug is too high.	Do not exceed one million cells per plug. If you require more sample for downstream analyses, cast multiple plugs per condition.
25-26	Plugs cannot be loaded in the Novex WedgeWell SDS-PAGE gel wells.	The plugs are not cold enough for loading or the buffer is covering the wells during loading.	Keep plugs on ice for at least 5 min prior to loading. Make sure that the buffer level is below the wells during loading.
28	Voltage is too high or too low.	Wrong amperage or one of the chambers or gels is not running properly.	Make sure every gel is covered by sufficient amounts of buffer. Also check that the chamber lid and gel are assembled and connected correctly. Recheck that the amperage is set to 20

			mA per gel (constant amperage).
29	Plugs do not stay in position when opening the gel.	There was remaining buffer left before opening the gel.	Use video S4 as a reference and try to remove as much buffer as possible before opening the gel.
32A XV	Sample is viscous after boiling.	The sample was not boiled long enough.	Boil for an additional 10-15 min or until the sample is not viscous anymore.

Timing

Part 1. Cell seeding and treatment: 2-3 days (depending on the chosen treatment)

A. *DPC induction using formaldehyde: 2 days, handling 2-3 h*

B. *Etoposide treatment to induce TOP2-DPCs: 2 days, handling 2-3 h*

C. *5-azadC treatment to induce DNMT1-DPCs and monitor their repair: 3 days, handling 3-4 h*

D. *UVC irradiation or POLα-inhibition to induce HMCES-DPCs: 3 days, handling 3-4 h*

Part 2. Sample processing: 6-7 h

Nuclease control (optional): 45 min, handling 15 min

Part 3. Downstream analysis: 2 h - 2 days (depending on the chosen analysis)

A. Western blot or fluorescent staining. TIMING 2 days, handling 1-3 h

B. Analysis by mass spectrometry. TIMING variable, handling 2-6 h

[H1] Anticipated results

[H2] Electro-elution of agarose plugs (Step 31)

The successful electro-elution of agarose plugs prepared from formaldehyde-treated cells can be monitored by staining the gel used for electro-elution with a Coomassie-based protein stain (Step 31, Fig. 3). This gel can also be used to monitor the success of the nuclease control, which should cause a depletion of the Coomassie-stained signal in the gel pocket (Fig. 3, red dots).

[H2] Analysis of DPCs by SDS-PAGE followed by western blot or fluorescent staining (Step 32A)

[H3] DPC induction using formaldehyde (Step 1A)

PxP samples obtained from formaldehyde-treated cells should display distinct bands around the 15 kDa marker when analysed by SDS-PAGE followed by fluorescent staining with SYPRO ruby protein stain (Fig. 4a, red dots). These bands correspond to crosslinked histone proteins and are sensitive to the nuclease control treatment prior to electro-elution, indicating that they are *bona fide* DPCs. Alternatively, silver staining can be used to visualize formaldehyde-induced DPCs, but we find that staining with SYPRO ruby protein stain results in comparable sensitivity while being less laborious. The formation of histone-DPCs can also be visualized by western blotting, using for

example an anti-histone H3 antibody (Fig. 4b), which should reveal a specific signal for histone H3-DPCs in PxP samples from formaldehyde-treated cells, which is absent in nuclease control samples.

If more cells per plug are used than the 1 million recommended in this protocol, increased DPC signals can be observed (Fig. 5a-b). However, increasing the number of cells can also lead to the accumulation of non-crosslinked proteins as evident from the accumulation of actin in PxP samples (Fig. 5b). The precise number of cells at which unspecific accumulation of proteins starts to occur is cell line specific and must thus be experimentally determined. When using 1 million cells per plug, we have not observed unspecific signals in any cell line tested.

In the original version of the PxP procedure[28], PxP samples were digested with benzonase and filtered through nitrocellulose columns to remove DNA and residual agarose, respectively, before analysis by SDS-PAGE. However, both steps turned out to be unnecessary with filtration even strongly reducing PxP signals (Extended Data Fig. 1a-b). Therefore, we have omitted these steps in this protocol.

[H3] Etoposide treatment to induce TOP2-DPCs (Step 1B)

The induction of etoposide-induced TOP2 DPCs can be observed in dose-dependent manner in PxP samples (Fig. 6a). No TOP2 signal is observed in PxP samples obtained from non-treated control cells. Notably, partial depletion of TOP2 can be observed in input samples, which can serve as a positive control for etoposide treatment.

[H3] 5-azadC treatment to induce DNMT1-DPCs and monitor their repair (Step 1C)

DNMT1-DPCs can be detected in PxP samples obtained from S-phase synchronized cells after a 30-minutes treatment with 5-azadC by western blotting against endogenous DNMT1. When cells are let to recover in drug-free media for one, two or three hours after 5-azadC exposure, a gradual decrease of the DNMT1-DPC signal can be observed in PxP samples. The stringency of DPC extraction by PxP is apparent when compared to chromatin fractionation[50], where histone H3, as well as DNMT1 signals can be detected in all conditions, including non-treated (Fig. 6b).

[H3] UVC irradiation/POLα-inhibition to induce HMCES-DPCs (Step 1D)

Six hours after UVC irradiation, HMCES-DPCs can be readily detected in S-phase synchronized cells expressing tagged HMCES-WT by western botting against the tag or HMCES itself (Fig. 7a). In the latter case, DPCs formed by endogenous HMCES are visible as well (Fig. 7a). In contrast, DPC formation is not observed in cells expressing a HMCES variant in which the catalytic cysteine residue has been replaced by serine (C2S) (Fig. 7a). Similarly, crosslinking HMCES-WT can be observed upon treatment of cells with the POLα inhibitor CD437 (Fig. 7b). Once the inhibitor is removed, the crosslinks are resolved rapidly (Fig. 7b), likely related to the autocatalytic reversal of HMCES-DPCs[41,51].

[H2] Analysis of DPCs by mass spectrometry (Step 32B)

To identify endogenous DPCs, compare proteins identified in PxP samples from non-treated cells with the respective nuclease control samples (Fig. 8a). DPCs are expected to be depleted in the nuclease control samples. To identify drug-induced DPCs, compare the treated condition to the non-treated condition and the treated condition to the respective nuclease control (Fig. 8b). Only if a protein is significantly

enriched in a treated condition over the respective nuclease control it should be considered a DPC. Typically, proteins with a log2 fold-change larger than 1 and an FDR smaller than 0.05 can be considered significantly different between conditions.

Figure legends

Fig. 1 Schematic illustration of the PxP workflow.

Overview of the PxP method. First, cells are mixed with low-melt agarose and casted in plugs using plug molds. The agarose plugs are then lysed in a denaturing buffer containing sarkosyl to disrupt the cell membrane and denature cellular proteins. After lysis, agarose plugs are transferred to the wells of an SDS-PAGE gel for electro-elution. During electro-elution, all soluble proteins elute into the SDS-PAGE gel, while genomic DNA and crosslinked proteins are retained in the plug. Finally, plugs are retrieved and DPCs can be identified using standard SDS-PAGE assays (protein staining, western blot) or mass spectrometry-based proteomics (LC-MS/MS).

Fig. 2. Critical steps of the PxP protocol. Photographs presenting correct handling of plugs during the most critical steps of the PxP protocol. For better visualization agarose plugs were stained with blue bromophenol dye in **a-b** and **d-e**. Corresponding steps of the protocol are indicated below each photograph. **a**, Casting of agarose plugs, by pipetting agarose/cell mixture into plug molds (left) and finished plugs in the molds (right). **b**, Transfer of solidified agarose plugs from plug molds to plastic tubes filled with lysis buffer. **c**, Agarose plugs with white opaque colour before lysis (left) and transparent plugs after successful lysis (right). The left photograph shows agarose plugs in plastic tubes with lysis buffer; red rectangles highlight the plugs. In the right

photograph agarose plugs were placed on a dark background for better visualization.

d, Loading of agarose plugs into the wells of a Novex WedgeWell SDS-PAGE gel for electro-elution. Agarose plugs are transferred with tweezers to the Novex WedgeWell SDS-PAGE gel in an electrophoresis chamber. **e**, Retrieval of agarose plugs after electro-elution. The agarose plugs are transferred with tweezers from the opened Novex WedgeWell SDS-PAGE gel to fresh plastic tubes.

Fig. 3. Anticipated results: Electro-elution of agarose plugs. DPCs were isolated by PxP from HAP1 cells treated with 1 mM formaldehyde (FA) for 1 h, including a nuclease control step. The gel used for electro-elution was stained with a Coomassie-based protein stain, showing eluted non-crosslinked proteins. Red dots highlight signal depletion in nuclease control samples.

Fig. 4. Anticipated results: DPC induction using formaldehyde. a-b, DPCs were isolated by PxP from HAP1 cells treated with 1 mM formaldehyde (FA) for 1 h, including a nuclease control step. PxP samples were resolved in a 12% SDS-PAGE gel and stained with SYPRO ruby protein stain (a). In parallel, PxP and input samples were analysed by western blotting using the indicated antibodies (b).

Fig. 5. Anticipated results: Effect of the number of cells per plug on specific and unspecific signals in PxP experiments. a-b, DPCs were isolated by PxP from HeLa T-REx Flp-In cells treated with 2 mM formaldehyde (FA) for 1 h with increasing numbers of cells embedded per plug. The gel used for electro-elution was stained with a Coomassie-based protein stain, showing eluted non-crosslinked proteins (a). PxP

samples were resolved in a 4-12% SDS-PAGE gel and were analysed by western blotting using the indicated antibodies (b).

Fig. 6. Anticipated results: Induction of TOP2- and DNMT1-DPCs. a, DPCs were isolated by PxP from HAP1 cells treated with the indicated concentrations of etoposide (ETO) for 1 h. PxP and input samples were analysed by western blotting using the indicated antibodies. **b**, Analysis of DNMT1-DPC repair by chromatin fractionation and PxP-WB in HeLa T-REx Flp-In cells treated as in **Step 1C**. Cells were synchronized via a double-thymidine block, released for 3 h into early/mid S-phase, and treated with 10 µM 5-azadC for 30 min and directly harvested after 5-azadC exposure or after an additional incubation in drug-free media for 1, 2 or 3 h, as indicated. Samples were analysed by western blotting using the indicated antibodies.

Fig. 7. Anticipated results: Induction of HMCES-DPCs. a, HeLa T-REx Flp-In cells expressing doxycycline-inducible HMCES-mVenus-3xFlag variants (WT or a catalytically compromised C2S variant) were synchronized via a double-thymidine block and released for 2 h into early/mid S-phase, before irradiation with UVC (50 J/m²). 6 h after irradiation, DPCs were isolated by PxP. PxP samples were analysed by western blotting using the indicated antibodies. **b**, HeLa T-REx Flp-In cells expressing doxycycline-inducible HMCES-WT-mVenus-3xFlag were synchronized via a double-thymidine block and released for 2 h into early/mid S-phase, before incubation for 1 h with 5 µM the polymerase α inhibitor CD437. Cells were directly harvested or after the indicated chase (1 or 2 h) in drug-free media. Cells were subjected to PxP and samples were analysed by western blotting using the indicated antibodies.

Fig. 8. Anticipated results: Analysis of DPCs by mass spectrometry.

a-b, Volcano plots depicting log2 fold changes (logFC) plotted against the FDR (-log10) of proteins isolated by PxP from HeLa cells, comparing either non-treated control cells (NT) (a) or cells treated with 2 mM formaldehyde (FA) for 1 h (b) with the respective nuclease control. PxP samples were subjected to in-plug tryptic digestion and label-free quantitative mass spectrometry (MS). Six biological replicates were used per condition. Reanalysis of data from ref[28].

Extended Data Fig. 1. Comparison of different methods for DPC retrieval from agarose plugs following electro-elution. a-b, DPCs were isolated by PxP from HeLa T-REx Flp-In cells treated with 2 mM formaldehyde (FA) for 1 h. To retrieve crosslinked proteins from agarose plugs different methods were used. Plugs were either dissolved by melting only or in combination with enzymatic DNA removal. Additionally, samples were passed through a column as in the original PxP protocol[28], as indicated. Finally, all PxP samples were resolved in a 4-12% SDS-PAGE gel and were analysed by western blotting using the indicated antibodies (a) or resolved in a 12% SDS-PAGE gel and stained with SYPRO ruby protein stain (b).

Table S1. Imputed intensities for MS volcano plots shown Fig. 8. Log2 normalized and imputed intensities of proteins isolated using PxP from non-treated (NT) cells or cells treated with formaldehyde (FA) and corresponding nuclease controls (NuclCtrl). Six biological replicates were used per condition. Reanalysis of data from ref[28].

Table S2. Limma test results for MS volcano plots shown in Fig. 8. Statistical test results of fold changes of proteins isolated by PxP from non-treated (NT) cells or cells

treated with formaldehyde (FA) as returned by limma. Six biological replicates were used per condition. Reanalysis of data from ref[28].

Video S1. Casting of agarose plugs using plug molds. See protocol step 10.

Video S2. Transferring agarose plugs from plug molds to lysis buffer. See protocol step 12.

Video S3. Loading agarose plugs for electro-elution. See protocol step 26.

Video S4. Retrieving agarose plugs after electro-elution. See protocol step 29-30.

Author contributions

P.W., S.D., and H.L. provided anticipated results and designed figures. M.J.G. analysed mass spectrometry data and prepared corresponding figures and text sections. P.W., S.D. and J.S. wrote the original draft. J.S. acquired funding and supervised the work. All authors read and approved the final manuscript.

Data Availability Statement

Mass spectrometry data are available from the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD02665449. Source data are provided with this paper.

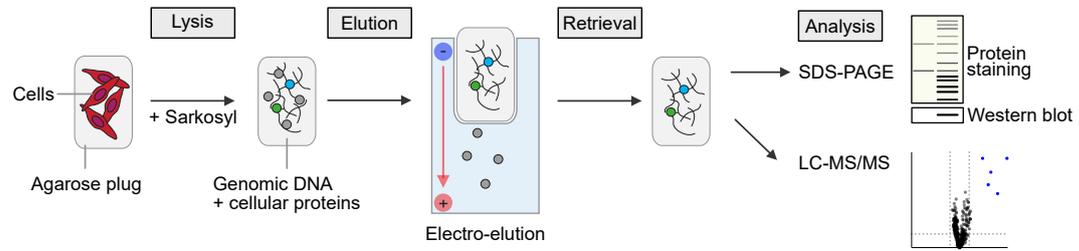**Competing interests**

The authors declare no competing interests.

References

1   Weickert, P. & Stingele, J. DNA-Protein Crosslinks and Their Resolution. *Annu Rev Biochem* **91**, 157-181 (2022). https://doi.org:10.1146/annurev-biochem-032620-105820
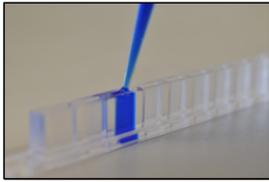
2   Stingele, J., Habermann, B. & Jentsch, S. DNA-protein crosslink repair: proteases as DNA repair enzymes. *Trends Biochem Sci* **40**, 67-71 (2015). https://doi.org:10.1016/j.tibs.2014.10.012

3   Stingele, J., Bellelli, R. & Boulton, S. J. Mechanisms of DNA-protein crosslink repair. *Nat Rev Mol Cell Biol* **18**, 563-573 (2017). https://doi.org:10.1038/nrm.2017.56

4   Wang, M., Dingler, F. A. & Patel, K. J. Genotoxic aldehydes in the hematopoietic system. *Blood* **139**, 2119-2129 (2022). https://doi.org:10.1182/blood.2019004316

5   Vijayraghavan, S. & Saini, N. Aldehyde-Associated Mutagenesis horizontal line Current State of Knowledge. *Chem Res Toxicol* (2023). https://doi.org:10.1021/acs.chemrestox.3c00045

6   Pommier, Y. *et al.* Tyrosyl-DNA-phosphodiesterases (TDP1 and TDP2). *DNA Repair (Amst)* **19**, 114-129 (2014). https://doi.org:10.1016/j.dnarep.2014.03.020

7   Pommier, Y. & Marchand, C. Interfacial inhibitors: targeting macromolecular complexes. *Nat Rev Drug Discov* **11**, 25-36 (2011). https://doi.org:10.1038/nrd3404

8   Christman, J. K. 5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. *Oncogene* **21**, 5483-5495 (2002). https://doi.org:10.1038/sj.onc.1205699

9   Mohni, K. N. *et al.* HMCES Maintains Genome Integrity by Shielding Abasic Sites in Single-Strand DNA. *Cell* **176**, 144-153 e113 (2019). https://doi.org:10.1016/j.cell.2018.10.055

10   Thompson, P. S., Amidon, K. M., Mohni, K. N., Cortez, D. & Eichman, B. F. Protection of abasic sites during DNA replication by a stable thiazolidine protein-DNA cross-link. *Nat Struct Mol Biol* **26**, 613-618 (2019). https://doi.org:10.1038/s41594-019-0255-5

11   Halabelian, L. *et al.* Structural basis of HMCES interactions with abasic DNA and multivalent substrate recognition. *Nat Struct Mol Biol* **26**, 607-612 (2019). https://doi.org:10.1038/s41594-019-0246-6

12   Ruggiano, A. & Ramadan, K. DNA-protein crosslink proteases in genome stability. *Commun Biol* **4**, 11 (2021). https://doi.org:10.1038/s42003-020-01539-3

13   Stingele, J., Schwarz, M. S., Bloemeke, N., Wolf, P. G. & Jentsch, S. A DNA-dependent protease involved in DNA-protein crosslink repair. *Cell* **158**, 327-338 (2014). https://doi.org:10.1016/j.cell.2014.04.053

14   Serbyn, N. *et al.* The Aspartic Protease Ddi1 Contributes to DNA-Protein Crosslink Repair in Yeast. *Mol Cell* **77**, 1066-1079 e1069 (2020). https://doi.org:10.1016/j.molcel.2019.12.007

15   Stingele, J. *et al.* Mechanism and Regulation of DNA-Protein Crosslink Repair by the DNA-Dependent Metalloprotease SPRTN. *Mol Cell* **64**, 688-703 (2016). https://doi.org:10.1016/j.molcel.2016.09.031

16   Vaz, B. *et al.* Metalloprotease SPRTN/DVC1 Orchestrates Replication-Coupled DNA-Protein Crosslink Repair. *Mol Cell* **64**, 704-719 (2016). https://doi.org:10.1016/j.molcel.2016.09.032

17   Lopez-Mosqueda, J. *et al.* SPRTN is a mammalian DNA-binding metalloprotease that resolves DNA-protein crosslinks. *Elife* **5** (2016). https://doi.org:10.7554/eLife.21491

18   Kojima, Y. *et al.* FAM111A protects replication forks from protein obstacles via its trypsin-like domain. *Nat Commun* **11**, 1318 (2020). https://doi.org:10.1038/s41467-020-15170-7

19   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). https://doi.org:10.1016/j.cell.2014.09.024

20   Sparks, J. L. *et al.* The CMG Helicase Bypasses DNA-Protein Cross-Links to Facilitate Their Repair. *Cell* **176**, 167-181 e121 (2019). https://doi.org:10.1016/j.cell.2018.10.053

21   Larsen, N. B. *et al.* Replication-Coupled DNA-Protein Crosslink Repair by SPRTN and the Proteasome in Xenopus Egg Extracts. *Mol Cell* **73**, 574-588 e577 (2019). https://doi.org:10.1016/j.molcel.2018.11.024

22   Reinking, H. K. *et al.* DNA Structure-Specific Cleavage of DNA-Protein Crosslinks by the SPRTN Protease. *Mol Cell* **80**, 102-113 e106 (2020). https://doi.org:10.1016/j.molcel.2020.08.003

23   Yaneva, D. *et al.* The FANCJ helicase unfolds DNA-protein crosslinks to promote their repair. *Mol Cell* **83**, 43-56 e10 (2023). https://doi.org:10.1016/j.molcel.2022.12.005

24   Gallina, I. *et al.* The ubiquitin ligase RFWD3 is required for translesion DNA synthesis. *Mol Cell* **81**, 442-458 e449 (2021). https://doi.org:10.1016/j.molcel.2020.11.029

25   Borgermann, N. *et al.* SUMOylation promotes protective responses to DNA-protein crosslinks. *EMBO J* **38** (2019). https://doi.org:10.15252/embj.2019101496

26   Sun, Y. *et al.* A conserved SUMO pathway repairs topoisomerase DNA-protein cross-links by engaging ubiquitin-mediated proteasomal degradation. *Sci Adv* **6** (2020). https://doi.org:10.1126/sciadv.aba6290

27   Liu, J. C. Y. *et al.* Mechanism and function of DNA replication-independent DNA-protein crosslink repair via the SUMO-RNF4 pathway. *EMBO J* **40**, e107413 (2021). https://doi.org:10.15252/embj.2020107413

28   Weickert, P. *et al.* SPRTN patient variants cause global-genome DNA-protein crosslink repair defects. *Nat Commun* **14**, 352 (2023). https://doi.org:10.1038/s41467-023-35988-1

29   Maskey, R. S. *et al.* Spartan deficiency causes genomic instability and progeroid phenotypes. *Nat Commun* **5**, 5744 (2014). https://doi.org:10.1038/ncomms6744

30   Lessel, D. *et al.* Mutations in SPRTN cause early onset hepatocellular carcinoma, genomic instability and progeroid features. *Nat Genet* **46**, 1239-1244 (2014). https://doi.org:10.1038/ng.3103

31   Donsbach, M. *et al.* An auto-release mechanism for HMCES-DNA-protein crosslinks. *bioRxiv*, 2022.2012.2018.520715 (2022). https://doi.org:10.1101/2022.12.18.520715

32   Reinking, H. K. & Stingele, J. Protein-oligonucleotide conjugates as model substrates for DNA-protein crosslink repair proteases. *STAR Protoc* **2**, 100591 (2021). https://doi.org:10.1016/j.xpro.2021.100591

33   Noireterre, A., Serbyn, N., Bagdiul, I. & Stutz, F. Ubx5-Cdc48 assists the protease Wss1 at DNA-protein crosslink sites in yeast. *EMBO J* **42**, e113609 (2023). https://doi.org:10.15252/embj.2023113609

34   Serbyn, N. *et al.* SUMO orchestrates multiple alternative DNA-protein crosslink repair pathways. *Cell Rep* **37**, 110034 (2021). https://doi.org:10.1016/j.celrep.2021.110034

35   Subramanian, D., Furbee, C. S. & Muller, M. T. ICE bioassay. Isolating in vivo complexes of enzyme to DNA. *Methods Mol Biol* **95**, 137-147 (2001). https://doi.org:10.1385/1-59259-057-8:137

36   Zhitkovich, A. & Costa, M. A simple, sensitive assay to detect DNA-protein crosslinks in intact cells and in vivo. *Carcinogenesis* **13**, 1485-1489 (1992). https://doi.org:10.1093/carcin/13.8.1485

37   Kiianitsa, K. & Maizels, N. A rapid and sensitive assay for DNA-protein covalent complexes in living cells. *Nucleic Acids Res* **41**, e104 (2013). https://doi.org:10.1093/nar/gkt171

38   Glumac, M. *et al.* SPRTN-dependent DPC degradation precedes repair of damaged DNA: a proof of concept revealed by the STAR assay. *Nucleic Acids Res* **51**, e35 (2023). https://doi.org:10.1093/nar/gkad022

39   Hu, Q. *et al.* The ARK Assay Is a Sensitive and Versatile Method for the Global Detection of DNA-Protein Crosslinks. *Cell Rep* **30**, 1235-1245 e1234 (2020). https://doi.org:10.1016/j.celrep.2019.12.067

40   Wilhelm, L. *et al.* SMC condensin entraps chromosomal DNA by an ATP hydrolysis dependent loading mechanism in Bacillus subtilis. *Elife* **4** (2015). https://doi.org:10.7554/eLife.06659

41   Rua-Fernandez, J. *et al.* Self-reversal facilitates the resolution of HMCES-DNA protein crosslinks in cells. *bioRxiv* (2023). https://doi.org:10.1101/2023.06.14.544844

42   RStudio: Integrated Development Environment for R. (2023).

43   preprocessCore: A collection of pre-processing functions_. R package version 1.60.0 (2022).

44   Gatto, L. & Lilley, K. S. MSnbase-an R/Bioconductor package for isobaric tagged mass spectrometry data visualization, processing and quantitation. *Bioinformatics* **28**, 288-289 (2012). https://doi.org:10.1093/bioinformatics/btr645

45   Gatto, L., Gibb, S. & Rainer, J. MSnbase, Efficient and Elegant R-Based Processing and Visualization of Raw Mass Spectrometry Data. *J Proteome Res* **20**, 1063-1069 (2021). https://doi.org:10.1021/acs.jproteome.0c00313

46   Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015). https://doi.org:10.1093/nar/gkv007

47   Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer International Publishing, 2016).

48   Strimmer, K. fdrtool: a versatile R package for estimating local and tail area-based false discovery rates. *Bioinformatics* **24**, 1461-1462 (2008). https://doi.org:10.1093/bioinformatics/btn209

49   Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V. & Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc* **1**, 2856-2860 (2006). https://doi.org:10.1038/nprot.2006.468

50   Zhao, S. *et al.* A ubiquitin switch controls autocatalytic inactivation of the DNA-protein crosslink repair protease SPRTN. *Nucleic Acids Res* **49**, 902-915 (2021). https://doi.org:10.1093/nar/gkaa1224
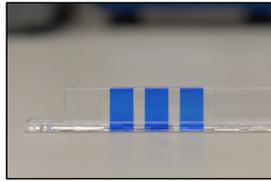
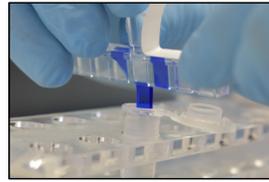51   Donsbach, M. *et al.* A non-proteolytic release mechanism for HMCES-DNA-protein crosslinks. *EMBO J* **42**, e113360 (2023). https://doi.org:10.15252/embj.2022113360

1

Cells — Agarose plug

Lysis → + Sarkosyl → Genomic DNA + cellular proteins

Elution → Electro-elution

Retrieval

Analysis → SDS-PAGE → Protein staining, Western blot

LC-MS/MS

a Casting

Step 10

Step 11

b Transfer

Step 12

Step 12

c Lysis

Step 13   Step 16
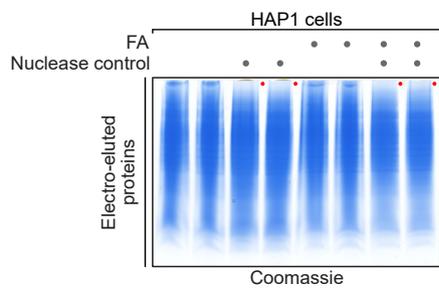

Step 13   Step 16

d Loading
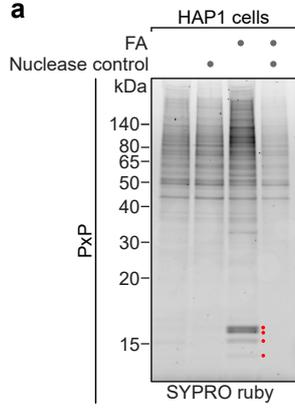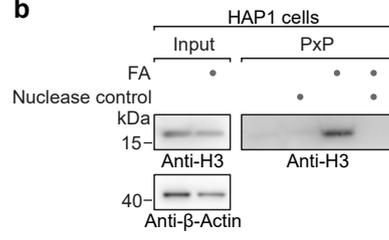

Step 26

Step 26

e Retrieval

Step 30

3

a

HAP1 cells

FA

Nuclease control

kDa

PxP

SYPRO ruby

b

HAP1 cells

| Input | PxP |

FA

Nuclease control

kDa

15–

Anti-H3   Anti-H3

40–

Anti-β-Actin

a

Hela T-REx Flp-In cells

FA

Cells/plug [x10⁶] 1 2 4 8 1 2 4 8

Coomassie

b

Hela T-REx Flp-In cells

Input	PxP

FA

Cells/plug [x10⁶] 1 2 4 8 1 2 4 8   1 2 4 8 1 2 4 8

kDa
15–

Anti-H3   Anti-H3

40–

Anti-β-Actin   Anti-β-Actin - long exposure

a

HAP1 cells

ETO [µM]   -   25   50

kDa

PxP

180-

Anti-TOP2

Input

180-

Anti-TOP2

40-

Anti-β-Actin

b

HeLa T-REx Flp-In cells

	Input	Chromatin	PxP

5-azadC
Chase [h]   0   3   0   1   2   3   0   3   0   1   2   3   0   3   0   1   2   3
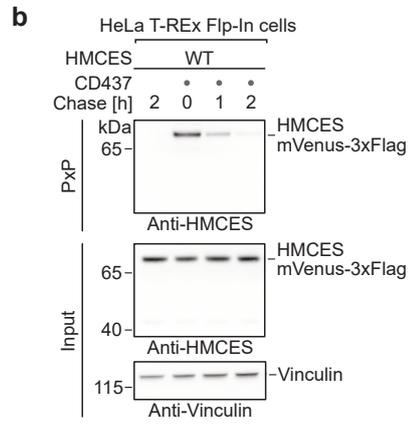
kDa
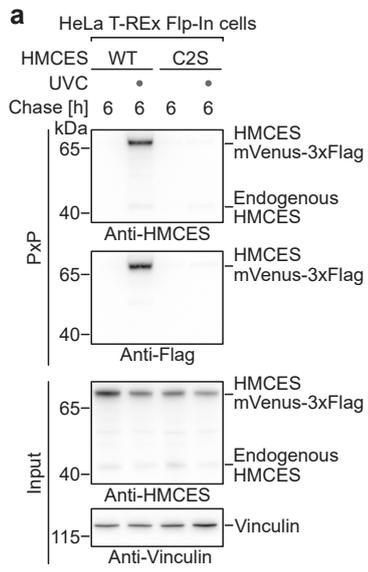
185-
115-
80-

—DNMT1

Anti-DNMT1

15-

Anti-H3

30-

Anti-GAPDH
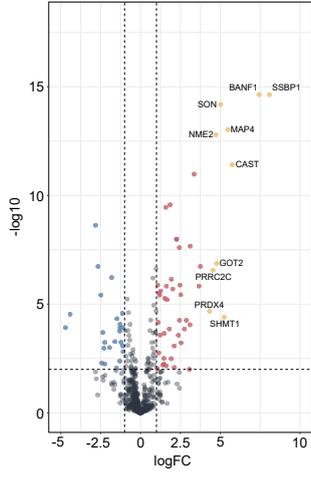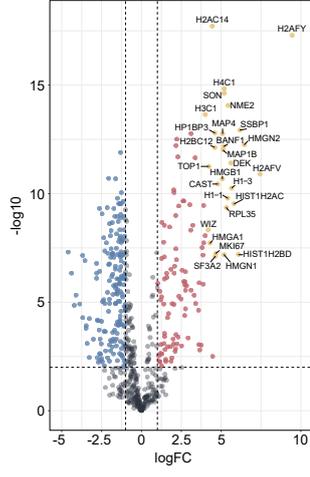
7

a MS: NT vs NT nuclease control

b MS: FA vs FA nuclease control

Extended Data 1

a

Hela T-REx Flp-In cells

	Input	PxP		
FA	•		•	•
DNA removal		• •	•	•
Column		•		•

kDa
15–

Anti-H3   Anti-H3

40–

Anti-β-Actin

b

Hela T-REx Flp-In cells

FA			•	•	•
DNA removal	•	•	•	•	
Column		•			•

PxP

kDa
150–
100–
70–
60–
50–
40–
30–

25–
20–

15–

SYPRO ruby