Dissertation zur Erlangung des Doktorgrades der

Fakultät für Chemie und Pharmazie

der Ludwig-Maximilians-Universität München

Mechanisms of DNA-protein crosslink formation, resolution and repair

Sophie Dürauer

aus

St. Pölten, Österreich

2025

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, 11.11.2025 ....

....

Sophie Dürauer

Dissertation eingereicht am 20.08.2025

1. Gutachter: Prof. Dr. Julian Stingele
2. Gutachter: Prof. Dr. Karl Duderstadt

Mündliche Prüfung am 30.10.2025

“In the top-ranked model, the hydrophobic Ile44-patch of ubiquitin was predicted to interact with a hydrophobic interface at the back of the SprT domain, hereafter referred to as ubiquitin-binding interface at the SprT domain (USD).” [1]

Internally also referred to as “ubiquitin-binding interface discovered by Sophie Dürauer”.

Table of content

List of publications

This dissertation has been prepared from September 2021 until August 2025 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 three publications.

Publication 1

Maximilian Donsbach*, **Sophie Dürauer***, Florian Grünert, Kha T Nguyen, Richa Nigam, Denitsa Yaneva, Pedro Weickert, Rachel Bezalel-Buch, Daniel R Semlow and Julian Stingele.
“A non-proteolytic release mechanism for HMCES-DNA-protein crosslink.”
EMBO J. 2023. 10.15252/embj.2022113360; *equal contribution

Publication 2

Pedro Weickert*, **Sophie Dürauer***, Maximilian J. Götz, Hao-Yi Li and Julian Stingele.
“Electro-elution-based purification of covalent DNA–protein cross-links”
Nat. Protoc. 2024. 10.1038/s41596-024-01004-z; *equal contribution

Publication 3

Sophie Dürauer, Hyun-Seo Kang, Christian Wiebeler, Yuka Machida, Dina S Schnapka, Denitsa Yaneva, Christian Renz, Maximilian J Götz, Pedro Weickert, Abigail C Major, Aldwin S Rahmanto, Sophie M Gutenthaler-Tietze, Lena J Daumann, Petra Beli, Helle D Ulrich, Michael Sattler, Yuichi J Machida, Nadine Schwierz and Julian Stingele.
“Allosteric activation of the SPRTN protease by ubiquitin maintains genome stability.”
Nat. Commun. 2025. 10.1038/s41467-025-61224-z

Summary

DNA integrity is continuously challenged by endogenous and exogenous sources of damage, leading to various forms of DNA lesions. Among the most complex lesions are DNA-protein crosslinks (DPCs), where proteins become covalently attached to DNA. DPCs can form when enzymes that usually form transient complexes with DNA get trapped or through exposure to reactive metabolites and chemotherapeutic agents. In the last years, intensive research has shed light on how cells control DPC formation and resolution.

HMCES is a highly conserved protein that forms physiological DPCs by crosslinking to abasic (AP) sites in single-stranded DNA (ssDNA) during replication. These HMCES-DPCs shield AP sites from spontaneous strand scission and cleavage by endonucleases like APE1, preventing formation of toxic DNA double-strand breaks. Proteolytic degradation via the proteasome or the DPC-specific protease SPRTN was long thought to be the main resolution pathway for these DPCs. This study discovered a non-proteolytic release mechanism, where the stability and release of HMCES-DPCs depend on the local DNA context. They remain stable in ssDNA or at junctions between ssDNA and double-strand DNA (dsDNA) but are released quickly in dsDNA mediated by a conserved glutamate residue within HMCES' active site and HMCES' affinity towards the present DNA-structure. Release of HMCES allows downstream repair enzymes to access the lesion. Therefore, the non-proteolytic release mechanism of HMCES-DPCs limits DPC formation to necessary situations, supporting genome integrity.

To analyse DPCs and their repair, the Purification of x-linked Proteins (PxP) method was developed. Briefly, for PxP cells are treated with DPC-inducing agents, embedded in low-melt agarose, lysed under denaturing conditions and subjected to electro-elution to remove soluble proteins. Crosslinked proteins remain with the DNA in the agarose plug and can be analysed using standard biochemical techniques or mass spectrometry-based proteomics. Compared to other techniques, the PxP offers higher specificity as it does not rely on precipitation. Usage of the PxP is demonstrated with DPCs induced by non-competitive inhibitors like etoposide or 5-azadC, as well as for following the formation and resolution of HMCES-DPCs in mammalian cells. The protocol can be performed by any trained molecular biologist using standard laboratory equipment and is adaptable to various sample types.

DPC repair often involves proteolytic degradation of the protein adducts, in humans this step is mainly performed by the proteasome and the metalloprotease SPRTN. Its activity must be tightly regulated to avoid unwanted cleavage of chromatin-interacting proteins. While DNA-specificity and monoubiquitylation were already identified as regulatory mechanisms, they do not fully explain SPRTN's selectivity and activity towards DPCs. This study investigated whether DPC ubiquitylation, a known and common modification of DPCs during repair, affects

SPRTN. Indeed, *in vitro* reconstitution of this process strongly activated SPRTN independent of known ubiquitin-binding domains. NMR and molecular dynamics simulations revealed a novel ubiquitin-binding interface at the back of SPRTN's protease domain mediating this activation. Ubiquitin-binding at this site induces and stabilises an open and active conformation of the enzyme. Disrupting the interaction by replacing key residues within this interface prohibited allosteric activation, leading to genomic instability and cell cycle defects in cells expressing Ruijs-Aalfs syndrome SPRTN patient variants. Notably, this ubiquitin-mediated activation also occurs in a replication-independent context, when DPCs get SUMOylated and subsequently ubiquitylated.

In summary, this study enhances understanding of DPC biology by revealing regulatory mechanisms governing HMCES-DPC formation and resolution as well as SPRTN activation via a novel ubiquitin-binding interface and by introducing the PxP as a reliable and sensitive method for DPC detection.

1 Introduction

1.1 DNA damage and repair

DNA encodes the genetic information of all living organisms and is constantly challenged by various types of lesions (Figure 1), which may result in mutations within the DNA sequence[2]. Although mutations had a beneficial effect during evolution, driving genomic diversity[3,4], they are also linked to specific human diseases, cancer and ageing[5]. Therefore, cells have evolved complex DNA repair systems, targeting various types of lesions to ensure genome integrity[3].

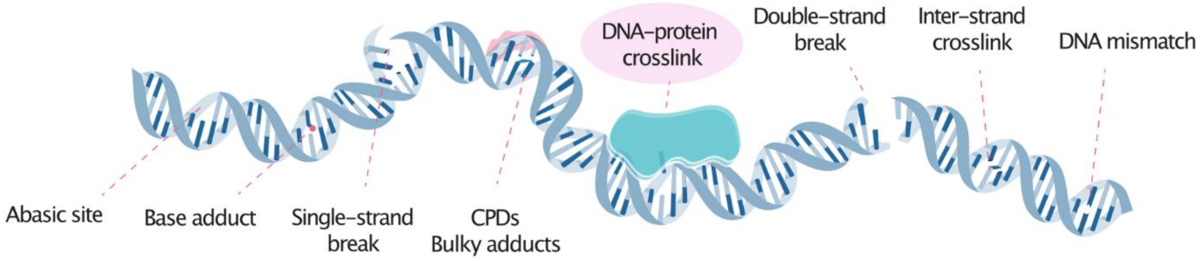


Figure 1. Schematic overview of common DNA lesions. Depicted lesions include abasic sites, base adducts, single-strand breaks, bulky adducts (e.g. cyclobutene pyrimidine dimer – CPDs), DNA-protein crosslinks, double-strand breaks, inter-strand crosslinks and DNA mismatches. Figure adapted from Weickert and Stingele[6].

1.1.1 DNA damage

Sources for DNA damage can be classified by origin into two main classes, endogenous and exogenous. Endogenous sources include metabolic processes in the cell with their byproducts and replication errors. Exogenous sources of DNA damage are further categorised into chemical and physical sources, encompassing chemicals, radiation and other environmental factors. DNA lesions triggered by both sources vary in toxicity and bulkiness[2,3].

1.1.1.1 Errors during replication

During replication, high-fidelity replicative DNA polymerases (Pol), Polδ and Polε, duplicate the DNA, generating a copy, which will be passed on to daughter cells during cell division[7]. Polδ and Polε are high-fidelity polymerases, due to their proofreading function, with 3'-5' exonuclease activity, but still they occasionally incorporate incorrect nucleotides[8,9]. These errors, if uncorrected, are passed on to daughter cells and enter the next replication cycle. Other polymerases, like the translesion synthesis (TLS) polymerases (reversionless 1 (REV1), Polζ, Polη, Polκ and Polι) have lower fidelity and lack proofreading ability. While this allows them to bypass bulky DNA lesions, such as protein adducts at stalled replication forks, it also increases the likelihood of nucleotide misincorporations, causing mutations[10]. Additionally, repetitive DNA sequences can disrupt replication accuracy and result in sequence insertions or deletions[11].

1.1.1.2 Damage to DNA bases

Constant exposure to diverse arrays of environmental chemicals as well as internally generated free radicals and electrophiles, trigger chemical modifications of the DNA, including the formation of DNA adducts, which interfere with DNA replication. The main processes causing these modifications are hydrolysis, alkylation and oxidation[2,3].

Hydrolytic deamination and abasic sites

During hydrolytic DNA deamination the exocyclic amino group of DNA bases is removed, converting adenine (A), guanine (G), cytosine (C) and 5-methly cytosine (5mC) to hypoxanthine, xanthine, uracil (U) and thymine (T) respectively, affecting their natural base pairing[3]. Inherited by daughter cells these base changes lead to genomic mutations. Deamination of C and 5mC occur most frequently and are catalysed by cytosine deaminases (APOBECs and AIDs)[12–14]. Deamination occurs more often in single-stranded DNA (ssDNA) than in double-stranded DNA (dsDNA), but in both cases can lead to the formation of abasic sites[2]. Abasic sites refer to apurinic/apyrimidinic sites (AP sites), positions in DNA lacking a nucleobase. AP sites form spontaneously via hydrolysis of the N-glycosidic bond between the nucleobase and the deoxyribose[3,15], as a result of DNA oxidation (8-oxo-7,8-dihydroguanine, 8-oxoG)[16], or as intermediates during base excision repair (BER), generated by DNA glycosylases[17]. AP sites are highly unstable and can undergo β-elimination leading to single-strand breaks (SSBs)[4,15].

Base adducts

Chemical modifications of DNA bases can result in adducts varying in size and severity. Small adducts arise primarily from alkylation and oxidation. DNA alkylation involves the transfer of an alkyl group to a nitrogen base of the DNA, common sites include O-6-methylguanine (O6-MeG), N-3-methyladenine (N3-MeA) and N-7-methylguanine (N7-MeG). These adducts may form spontaneously or triggered by endogenous (e.g. nitrosamines) or exogenous alkylating agents[18,19]. Alkylators are present in high abundance in the environment, found in air, water, food and pollutants like tobacco smoke or fuel emissions[20–22]. Common experimental agents include methyl methane sulfonate (MMS) and N-methyl-N’-nitro-N-nitrosoguanidine (MNNG), which mainly form N7-MeG and N3-MeA[23]. Temozolomide acts similarly, forming mainly O6-MeG and is clinically used in chemotherapy[24]. During DNA oxidation, an oxygen atom is added to a nucleobase. A prominent example is 8-oxoG, generated by endogenous reactive oxygen species (ROS) during cell respiration[16]. Exogenous sources for DNA oxidation, can either directly act as oxidative agents or act indirectly by generating ROS. Examples are potassium bromate ($KBrO_3$)[25] and polycyclic aromatic hydrocarbons (PAHs)[26], but also ionising radiation (IR) can generate ROS[27]. While small base adducts do not drastically alter DNA structure, they can trigger AP site formation, mutations or stalling of DNA synthesis[28,29]. Of note, PAHs can also trigger the formation of bulky base adducts[26].

Bulky base adducts

Prominent types of bulky DNA adducts are cyclobutene pyrimidine dimers (CPDs) and 6-4 photoproducts (6-4 PPs), both involving covalent bonds between pyrimidine bases on the same DNA strand and induced by ultraviolet (UV) radiation[30–33]. UV light is categorized by emission wavelength into UVA (320-400 nm), UVB (290-320 nm) and UVC (190-290 nm). DNA absorbs light at 260 nm, falling into the UVC spectra, making UVC the most potent inducer of UV-specific lesions[3,30].

DNA crosslinking damage

Moreover, reactive aldehydes like formaldehyde, acetaldehyde and malondialdehyde, along with ROS and reactive nitrogen species (RNS), can act as “crosslinkers”[3,34], forming DNA adducts, interstrand crosslinks (ICLs), covalent linkages between two nucleobases from opposing DNA strands, and DNA-protein crosslinks (DPCs), covalent attachment of proteins to DNA[35,36]. All of these lesions interfere with DNA metabolism as they block replication and transcription[34,37]. Bifunctional crosslinkers, such as nitrogen mustards, platinum compounds and mitomycin C are used in chemotherapy taking advantage of their potential to induce ICLs and DPCs[38,39].

1.1.1.3 DNA strand breaks

SSBs are the most common DNA lesion and typically form during replication and transcription[4,15], but are physiologically induced by topoisomerase 1 (TOP1) activity to reduce DNA supercoiling[40]. SSB accumulation poses a risk for double-strand break (DSB) formation, highly toxic lesions that can cause major loss of genetic information, cell-cycle stalling, checkpoint activation and ultimately lead to cell death[41]. IR, including α-, β-, γ-radiation and X-rays, produced for example by medical devices[42], can directly induce SSBs containing 3’-phosphates or 3’-phosphoglycolates instead of the 3’-hydroxyl group[43,44]. The SSBs are caused by the high energy of IR, which produces ions that disrupt covalent bonds in biological molecules[45].

1.1.1.4 Replication-transcription conflicts and mobile genetic elements

Furthermore, RNA:DNA hybrids (R-loops), stabilised transcription intermediates, have emerged as DNA lesions linked to replication-transcription conflicts[46]. Mobile genetic elements such as LINE-1 elements also threaten genome stability[47], with their activation being associated with DSBs, senescence and apoptosis in cultured human cells[48,49]. Though, the exact mechanism remains unclear, replication-integration conflicts may be involved[50].

1.1.2 DNA damage response

Due to the constant exposure to endogenous and exogenous DNA-damaging agents, cells have evolved specialised DNA damage response (DDR) pathways to resolve and limit DNA lesions and thereby preserve genome stability[3].

One of these mechanisms is direct repair, which reverses base modifications without altering the DNA backbone or requiring new DNA synthesis[51]. Direct repair mainly targets UV-light induced DNA lesions, including CPDs and 6-4 PPs, O6-alkylguanine and N-alkylated base adducts. Specialised enzymes, called photolyases, repair CPDs and 6-4 PPs in bacteria and yeast, but are absent in humans[52]. In humans, O6-alkylguanine and most N-alkylated bases are repaired by enzymes of the methylguanine DNA methyltransferase (MGMT) and anaplastic lymphoma kinase (Alk) family[53,54], MGMT enzymes catalyse demethylation of the lesion by transferring the alkyl group to its own catalytic cysteine. This S-alkylated MGMT is then quickly degraded by the proteasome[55]. The Alk family of demethylases uses an iron-dependent mechanism to oxidise aberrant alkyl groups, restoring the original base and releasing the methyl group as formaldehyde[54,56,57].

In addition, single damaged DNA bases can be repaired by BER, which removes the incorrect base and exchanges it with the correct one[51]. BER operates via two sub-pathways: short-patch BER, replacing a single nucleotide, and long-patch BER, where after repair of the damaged base a stretch of 2-12 nucleotides is newly synthesised. BER starts with the removal of the damaged base by DNA glycosylases. Monofunctional glycosylases, like uracil-DNA glycosylase (UDG) excise the base and leave behind an AP site[58,59]. In contrast, bifunctional glycosylases, such as 8-oxoguanine glycosylase 1 (OGG1) also cleave the DNA backbone, leaving behind a 3'-polyunsaturated aldehyde (3’-PUA) and a 5'-phosphate[60]. AP endonuclease 1 (APE1) processes these sites (AP site or 3’-PUA) generating a 3’-hydroxyl and a 5’-deoxyribose phosphate (5’-dRP) flap[61–63]. Next, Polβ is recruited to the damage site, fills the gap and removes the 5’-dRP[64,65]. In short-patch BER, DNA ligase I or III seal the nick, often following phosphorylation of the 5’-end by polynucleotide kinase/phosphatase (PNKP). Long-patch BER, initiated by bifunctional DNA glycosylases or during replication when Polβ cannot remove the 5’-adduct[59,66], involves loading of proliferating cell nuclear antigen (PCNA) and synthesis by replicative polymerases (Polδ or Polε), though Polβ may act in non-dividing cells. A 2-12 nucleotide stretch is synthesised, replacing DNA adjacent to the lesion and creating a flap structure, which is cleaved by Flap structure-specific endonuclease 1 (FEN1), followed by DNA ligase I-mediated ligation of the remaining nick[59,67]. There is an ongoing debate on the precise timing of poly (ADP-ribose) polymerase 1 (PARP1) and X-ray repair cross-complementing protein 1 (XRCC1) in BER[68–70]. PARP1 is considered an early responder, binding to BER intermediates (SSBs and AP sites), triggering PARylation of itself and surrounding proteins. This facilitates recruitment of other repair factors, including XRCC1, which may serve as a scaffold coordinating repair[69].

SSB repair pathways overlap with other DDR pathways, as SSBs often arise as intermediates in BER and nucleotide excision repair (NER)[71]. PARP1 and PARP2 recognise SSBs and upon binding catalyse PARylation of themselves and nearby proteins in a nicotinamide adenine dinucleotide (NAD^+) consuming reaction[72,73]. PARylation is regulated by histone PARylation factor 1 (HPF1)[74,75] and can be reversed by poly-ADP-ribose glycohydrolase (PARG)[76]. PARylation modifies chromatin structure allowing recruitment of repair factors and chromatin remodelers[77–79]. Moreover, PARylated chromatin directly recruits XRCC1, which serves as a scaffold for other SSB repair enzymes[80]. Interestingly, in BER, XRCC1 can be recruited independently of PARylation via Polβ post-synthesis[81,82]. XRCC1 promotes repair completion, following short- or long-patch BER, by disengaging PARP1 from the DNA[83].

DNA mismatch repair (MMR) corrects replication errors, such as mismatched base pairing and small insertions or deletions, preferentially in actively transcribed genes. It proceeds through four main steps: lesion recognition, repair initiation, excision and DNA re-synthesis[84]. The system relies on MutS and MutL homologs and in humans eight genes encode MMR proteins which form different heterodimers conferring lesion specificity[85,86].

NER removes bulky, helix-distorting lesions such as CPDs, 6-4 PPs and DNA adducts. Two sub-pathways are distinguished, global-genome NER (GG-NER), which scans the entire genome, and transcription-coupled NER (TC-NER), which targets lesions blocking transcription[87,88]. Mutations in GG-NER genes or core NER genes cause xeroderma pigmentosum, which is mainly characterised by extreme sensitivity to UV radiation. In TC-NER genes mutations lead to Cockayne syndrome, correlating with developmental delays, neurological dysfunction and premature ageing[89]. For GG-NER, xeroderma pigmentosum factor C (XPC) in a complex with UV excision repair protein RAD23 homolog B (RAD23B) and centrin 2 (CETN2)[90], initiates repair by sensing helix distortions and flipping out damaged bases[91,92]. Since UV-induced lesions cause only minimal distortions, which XPC cannot detect, UV-damaged DNA-binding protein (UV-DDB) heterodimer (DDB1 and DDB2) kinks DNA adjacent to the lesion, enabling recognition by XPC. Moreover, DDB1 recruits E3-ligases for XPC ubiquitylation (see section 1.4.4), increasing its DNA-binding. DNA-bound XPC is recognised by XPB and recruits transcription factor IIH (TFIIH)[90]. In TC-NER, stalled RNA Pol II recruits Cockayne syndrome group B protein (CSB), which subsequently promotes CSA binding[93]. CSA is part of a E3-complex that mediates ubiquitylation of RPB1 (RNA Pol II subunit), leading to its degradation and recruitment of the UV-sensitive syndrome protein A (UVSSA) and TFIIH[94,95]. TFIIH unwinds the DNA, allowing XPD to scan for bulky lesions[87,90].

XPA, XPG and replication protein A (RPA) stabilise and protect the open DNA complex during this process[96,97]. Once XPD encounters lesions, XPC is displaced by XPG, initiating recruitment of XPF- excision repair cross-complementing group 1 (ERCC1) by XPA[98,99]. The endonucleases XPG and XPF-ERCC1 then excise approximately 30 nucleotides containing the lesion[99–101]. Subsequent DNA re-synthesis and ligation follow long-patch BER[102].

DSBs are repaired by three pathways: non-homologous end joining (NHEJ), microhomology-mediated end joining (MMEJ) and homologous recombination (HR) with pathway choice depending on cell cycle phase and template availability[103].

NHEJ is further categorised into classical NHEJ (cNHEJ) and single-strand annealing (SSA). cNHEJ involves Ku70/Ku80 binding to DNA ends, recruiting DNA-dependent protein kinase (DNA-PK), DNA-ligase 4 (LIG4) and scaffolding factors (XRCC4, PAXX and XRCC4-like factor XLF)[104–106]. DNA-PK activation leads to autophosphorylation and phosphorylation of downstream targets, like histone H2AX (γH2AX). γH2AX is phosphorylated by ataxia telangiectasia mutated (ATM) or ataxia telangiectasia and Rad3 related (ATR) kinase, promoting chromatin relaxation around the break[107–109]. Recruited factors form a synaptic complex, aligning the two DNA ends and allowing ligation by LIG4[106,110]. Many DNA ends require prior end processing, performed by PNPK, polymerases (Polμ and Polλ) and nucleases (e.g. Artemis)[111]. In contrast, SSA requires prior end resection, initiated by the MRN complex, formed by meiotic recombination 11 (MRE11), RAD50 and Nijmegen breakage syndrome 1 (NBS1), and *C*-terminal-binding protein-interacting protein (CtIP)[112,113]. CtIP gets activated via phosphorylation during S-G2 phase by cyclin-dependent kinases[114], but resection requires further modifications, including ubiquitylation (see section 1.4.4). Compared to MMEJ, SSA relies on longer homologous sequences for repair, which often cause deletions in repetitive sequences[115]. Here, RAD52 mediates strand annealing, while ERCC1/XPF remove flaps generated during DNA synthesis[116–118]. The polymerases involved in SSA remain unclear[119].

MMEJ also begins with MRN-mediated end resection but requires only short microhomology sequences (2-20 nucleotides). Polθ displaces RPA from ssDNA via its helicase function and performs the initial extension before Polδ takes over DNA synthesis[120]. APE2 processes generated flaps and XRCC1-LIG3 ligates the ends[121,122].

Compared to NHEJ, HR is a high-fidelity pathway, which uses the sister chromatid as a template[123]. After end resection by MRN and CtIP, RPA bound to ssDNA is gradually replaced by RAD51, which is loaded onto DNA by breast cancer type (BRCA) 2. RAD51 performs homology search and together with partner and localiser of BRCA2 (PALB2), BRCA1 and BRCA1-associated RING domain protein 1 (BARD1) initiates strand invasion, forming a displacement loop (D-loop) structure[124–126]. The invaded strand is then used as a template for extension of the 3'-end of the invading strand by Polδ[119]. Depending on the involvement of the second DNA strand, HR can proceed in three different sub-pathways. First, synthesis-dependent strand annealing (SDSA), where only the invading strand is extended before it dissociates and reanneals to serve as a template for the invaded strand[127]. Second, classic HR, mainly acting during meiotic recombination, where both strands are extended by the formation of a double Holliday junction[127]. Holliday junctions can be resolved without crossover by the Bloom helicase (BLM) and the BTR complex (BLM, TOP3α, RMI1 and RMI2) or with crossover by essential meiotic endonuclease 1 homolog (EME) 1 and 2 and methansulfonate, UV sensitive 81 (MUS81)[128]. Finally, break-induced replication (BIR), where the invading strand is extended until the end of the chromosome, if not interrupted[129,130]. Although BIR plays a key role during replication and is involved in alternative telomere lengthening (ALT) in telomerase-deficient cancer cells[131], it is highly mutagenic and prone to causing chromosomal rearrangements[132].

Fanconi anaemia (FA) is a rare genetic disorder caused by mutations in any of the 22 known FA gens[133], leading to impaired repair of ICLs and increasing risk for cancer[133,134]. During S-phase, replication forks stalled at ICLs trigger recruitment of the FA anchoring complex, including Fanconi anaemia complementation group (FANC) M, FA core complex-associated protein 24 (FAAP24) and FANCM-associated histone fold protein (MHF) 1 and 2, followed by the FA core complex, containing the ubiquitin E2-conjugating enzyme FANCT and the ubiquitin E3-ligase FANCL[135,136]. These enzymes ubiquitylate the FANCD2-FANCI (ID2) complex, which recruits structure-specific nucleases for ICL unhooking (e.g. XPF-ERCC1, SLX1) via FANCP, and TLS polymerases to bypass the lesion. The resulting DSBs are repaired via HR, involving additional FA proteins (e.g. BRCA1 also called FANCS)[135,137–139]. Alternatively, ICLs can be excised by the glycosylase Nei endonuclease VIII-like 3 (NEIL3), generating an AP site, that is similarly bypassed by TLS polymerases[140].

1.1.3 DNA damage tolerance

Although DNA repair pathways aim to resolve lesions as quickly and efficiently as possible, some lesions persist into DNA replication, risking fork stalling or collapse and chromosome breakage. To prevent such situations, cells activate DNA damage tolerance (DDT) pathways that allow replication to continue on damaged templates[141].

One of these pathways is repriming, where synthesis resumes downstream of the lesion. In metazoans, this is carried out by PrimPol[142,143], a specialised primase recruited by RPA to ssDNA generated during fork uncoupling. PrimPol can both prime and extend DNA past the lesion[144], a function especially important when TLS cannot bypass the lesions, for instance at

ICLs[145]. Following repriming Polδ-interacting protein 2 (PolDIP2) enhances PrimPol's primer extension activity to facilitate the switch to Polδ, resuming physiological replication[146,147].

When the replicative polymerases Polδ and Polε stall at bulky DNA lesions, TLS polymerases are recruited[148]. Since interactions of TLS polymerases with DNA are loose and non-specific, they can accommodate and bypass distorted DNA and bulky lesions. However, their low fidelity comes at a cost, increasing the risk of mutations, which is why their recruitment and activity is tightly regulated[149]. In humans, the main TLS polymerases include the Y-family members Polη, Polι, Polκ and REV1 and Polζ, which belongs to the B-family[150]. Each TLS polymerase has distinct lesion preferences. Polη targets CPDs, but not 6-4 PPs[151,152], while sharing some functions with Polι in UV-induced damage repair[153–155]. Polκ bypasses monoadducts like N2-dG alkylation and can assist by extending bases inserted by other TLS polymerases[156,157]. REV1, uniquely inserts cytosine independently of the DNA template, using a loop structure within its active site, allowing bypass of numerous G-adducts and AP sites[158–160]. Polζ is composed of REV3 (catalytic subunit), two REV7 subunits, and the accessory subunits Pol31 and Pol32. It primarily extends nucleotides inserted by Y-polymerases[161,162]. Though error-prone, its fidelity is higher than that of Y-family polymerases[149]. TLS is initiated by stalled replication forks and persistent RPA-coated ssDNA, which recruits the RAD6-RAD18 E2-E3 complex, leading to PCNA monoubiquitylation[30,163,164]. This modification recruits TLS polymerases through their ubiquitin-binding and PCNA-interacting protein (PIP) motifs[165]. REV1 also acts as a scaffold, binding the REV7 subunit of Polζ and Y-family polymerases via PIP-like motifs, facilitating lesion-specific TLS polymerase recruitment. Polymerase switching is tightly regulated by Polδ, which reclaims replication forks after lesion bypass to minimise mutagenesis[143,166,167].

An alternative, error-free DDT pathway is template switching (TS), that restarts stalled replication forks by using another region of the parental strand, a sister chromatid or even the newly synthesised strand as a template[142]. TS is regulated by small ubiquitin-related modifier (SUMO)-ylation and ubiquitylation of PCNA. SUMOylation mediated by protein inhibitor of activated STAT (PIAS) 1 and PIAS4 prevents TS via recruitment of SHI related sequence 2 (SRS2)[168]. RAD6/RAD18-dependent monoubiquitylation of PCNA triggers TLS, whereas extension of the monoubiquitin to K63-linked polyubiquitin by UBC13/MMS2 and RAD5 initiates TS[169–171]. Resolution of the paired nascent strands resembles HR termination[142,172].

Fork reversal forms a “chicken foot” structure at sites of DNA damage, stabilising and protecting stalled forks from degradation[173]. This allows the nascent leading strand to be extended using the lagging strand as a template or to be repaired in a dsDNA context, using the lagging strand as substrate[174]. Fork reversal is triggered by polyubiquitylation of PCNA via helicase-like transcription factor (HLTF)[175] or SNF2 histone linker PHD RING helicase (SHPRH)[176], which recruit two translocases SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A-like protein 1 (SMARCAL1)[177] and zinc finger RANBP2-type containing 3 (ZRANB3)[178] to remodel the fork. RAD51 is essential during these processes for both reversed fork formation and protection[179]. After reversal and extension of the leading strand, RecQ-Like Type 1 (RECQ1) promotes fork restart by branch migration, or the fork undergoes controlled resection by the Werner-Syndrome (WRN) helicase and DNA replication ATP-dependent helicase/nuclease DNA2 (DNA2) to enable lesion repair and replication resumption[180,181].

In conclusion, cells employ a variety of specialised, tightly regulated repair pathways to address different types of DNA lesions, many of which have been the focus of extensive research over the years[2–4]. Among these lesions, DPCs stand out as particularly complex, with scientific attention on them rising in the last decade[6]. Since DPCs are the main topic of this study, they are described separately with a special focus on their repair pathways.

1.2 DNA-Protein Crosslinks

DPCs are covalent crosslinks of proteins to DNA. DPCs disrupt essential chromatin related processes like replication and transcription, making them highly toxic DNA lesions. They are very diverse, varying by the crosslinked protein, the DNA structure involved and the chemical nature of the covalent crosslink. Based on the crosslinked protein, they are categorized into non-enzymatic and enzymatic DPCs[6].

1.2.1 Non-enzymatic DPCs

Non-enzymatic DPCs form when reactive metabolites or chemotherapeutic agents covalently link nearby proteins to DNA, which often occurs alongside other lesions like monoadducts and DNA-DNA crosslinks (e.g. ICLs). Reactive metabolites form during physiological cellular processes and include reactive aldehydes, like acetaldehyde formed during ethanol detoxification[182,183] or formaldehyde, present in micromolar concentrations in mammalian serum and produced in direct vicinity to the DNA, during histone demethylation[184–187]. Formaldehyde reacts with amino and thiol groups[188] to form Schiff-base intermediates via dehydration, followed by formation of the final crosslink (Figure 2a)[38,58,189]. Due to its crosslinking efficiency, formaldehyde is widely used to study DPC formation in cells or in molecular biology techniques for the detection of DNA-protein interactions (e.g. chromatin immunoprecipitation sequencing, ChIP-Seq)[190]. It is also a major environmental toxin, present in air, water and tobacco smoke[191]. Interestingly, AP sites also generate aldehydes, as they constantly cycle between a closed ring and an open-ring aldehyde state. This open-ring aldehyde is very reactive and can crosslink to lysine or cysteine residues of nearby proteins, for example histones (Figure 2b)[58,192,193].

Furthermore, ROS and DNA metabolism intermediates, like 5-formylcytosine (5fC) produced by ten-eleven translocation (TET) enzymes during DNA methylation, can lead to DPCs[38,194,195], as well as many compounds used in chemotherapy. Platinum-based chemotherapeutics (e.g. cisplatin, oxaliplatin and carboplatin) primarily form DNA-DNA crosslinks, but can also bridge the N7-G in DNA to protein side chains of cysteine, arginine and lysine (Figure 2c)[196,197]. Similarly, nitrogen mustards such as mechlorethamine, alkylate nucleophilic sites on DNA and proteins, initiating covalent linkages[198], though their main cytotoxic effect also arises from DNA-DNA crosslink formation[38,199]. Finally, both IR and UV radiation promote DPC formation, alongside causing base damage and DNA strand breaks[200,201]. Radiation generates DNA and protein radicals, which subsequently can form covalent bonds. Interestingly, DPC formation of IR is enriched in hypoxic conditions, a phenomenon known as reverse oxygen effect and relevant for the treatment of solid tumours[202–206].

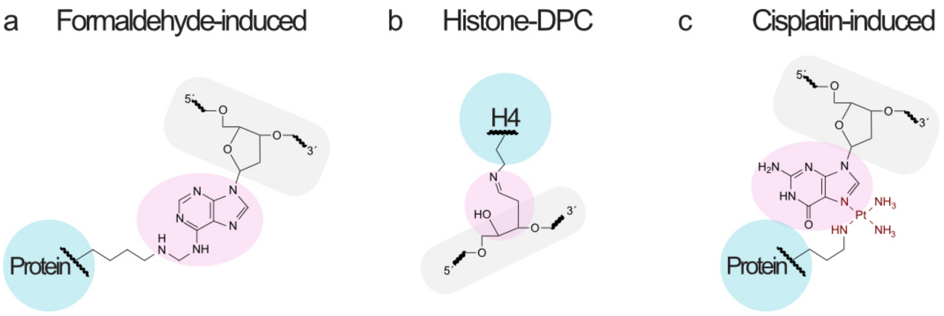


Figure 2. Chemical structures of common non-enzymatic DPCs. (a) DNA-protein crosslink (DPC) induced by formaldehyde. (b) DPC formed between histone H4 and an apurinic/apyrimidinic (AP) site. (c) DPC induced by cisplatin, coloured in red. DNA backbone is shown in grey, crosslinked protein in blue and the chemical bond in pink. Figure adapted from Weickert and Stingele[6] and Kühbacher and Duxin[207].

1.2.2 Enzymatic DPCs

Enzymatic DPCs form when DNA-processing enzymes, which normally create transient complexes with DNA, become trapped due to structural distortions in the DNA or chemical inhibitors[6]. One of the most studied examples here is TOP1, an evolutionary conserved enzyme that relaxes DNA supercoils during replication, transcription and chromatin remodelling by inducing a SSB[208]. TOP1 forms a transient covalent bond between its active site tyrosine and the 3’-end of DNA (Figure 3a), referred to as the TOP1 cleavage complex (TOP1cc). The TOP1cc creates a SSB to release torsional stress by allowing rotation of the DNA strand, followed by re-ligation of the SSB and TOP1 release[40,209]. Under physiological conditions, TOP1ccs are short-lived, but inhibitors like camptothecin (CPT) can stabilise the complex by intercalating at the TOP1-DNA interface[210]. Interestingly, the CPT-producing plant

Camptotheca acuminata expresses a CPT-resistant TOP1 mutant (N722S, next to the active site tyrosine Y723)[211], which was also found in drug-resistant leukaemia cells upon CPT selection[212]. Nowadays, chemical derivatives of CPT (topotecan, irinotecan) are used in the clinic, due to improved stability and bioavailability[213]. TOP1ccs can also arise naturally, as nearby DNA lesions like AP sites or damaged bases can interfere with completion of the enzymatic reaction[214,215].

TOP2, like TOP1, forms covalent intermediates with DNA to release topological stress[40,209]. In contrast to TOP1, TOP2 acts as a homodimer and introduces DSBs, by cleaving both DNA strands, while remaining bound to the 5'-ends and generating a 5'-overhang (Figure 3b). Like TOP1ccs, TOP2ccs can be trapped at DNA lesions, such as AP sites, mismatches and modified bases. They can also be stabilised by chemotherapeutics (etoposide, doxorubicin and mitoxantrone), mimicking the mechanism of action of CPT for TOP1[216]. Additionally, antibiotics like ciprofloxacin exploit the toxicity of TOP2ccs, targeting the bacterial TOP2 homolog (DNA gyrase)[217]. This principle is also used in molecular biology techniques, like gateway cloning, where bacterial gyrase is trapped in a DPC-like complex via a *CcdB*-encoded toxin[218].

DNA methyltransferase 1 (DNMT1) maintains DNA methylation patterns after replication. DNA methylation is essential for chromatin structure and primarily targets CpG sites where cytosines are modified on position C5[219–222]. DNMT1 catalyses methylation by forming a transient covalent crosslink via its catalytic cysteine, transferring a methyl group from S-adenosylmethionine to the C5 of cytosine and completing the reaction via β-elimination for enzyme release[221,223]. The transient DNMT1-DNA complex can be trapped by chemotherapeutics, like 5-aza-2'-deoxycytidine (5-azadC/decitabine), a cytidine analogue incorporated into DNA during replication. DNMT1 attempts to methylate 5-azadC, but the nitrogen at position 5 prevents β-elimination, resulting in a stable DNMT1-DPC (Figure 3c)[224,225]. Besides DNMT1-DPCs formation, which require repair, this process depletes DNMT1, causing global DNA hypomethylation and reactivation of silenced tumour suppressor genes[226].

Moreover, Polβ can get trapped and form a DPC when it encounters 2-deoxyribonolactone (dL)[227], a cytotoxic, reactive derivative of AP sites[228], generated by hydrogen peroxide, UV-light[229], IR[230] and certain chemotherapeutics[231,232]. dL acts as a suicidal substrate for Polβ during short-patch BER[233]. PARP can form pseudo-DPCs when trapped on DNA by PARP inhibitors (PARPi), which block NAD^+-binding sites at PARP's active site and are widely used in anti-cancer therapy[234,235]. Moreover, PARP1 can crosslink to AP sites or related lesions such as 3’-PUA during BER. PARPi treatment exacerbates this by prolonging PARP1 residence time on damaged DNA[236,237].

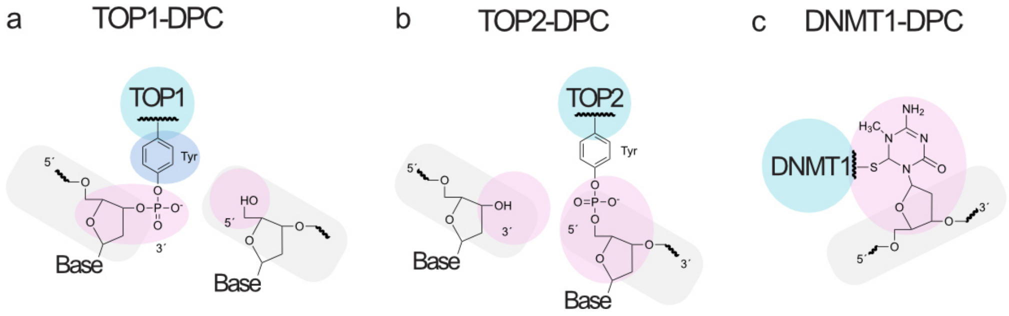


Figure 3. Chemical structures of common enzymatic DPCs. (a) Topoisomerase (TOP) 1 cleavage complex, with a covalent phosphotyrosyl bond between the 3’-end of a single-strand DNA break and the active site tyrosine. (b) TOP2 cleavage complex, with covalent phosphotyrosyl bonds between the 5’-end of a double-strand DNA break and the active site tyrosine of each TOP2 subunit. (c) DNA methyltransferase 1 (DNMT1) covalently trapped by 5-aza-2’-deoxycytidine (5-azadC). DNA backbone is shown in grey, crosslinked protein in blue and the chemical bond in pink. Figure adapted from Weickert and Stingele[6].

1.2.3 Physiological DPCs

Although DPCs pose a major threat to genome stability, requiring rapid resolution, some enzymes deliberately form covalent DNA adducts for functional purposes[6]. SPO11, a TOP2-like enzyme, generates DSBs during meiotic recombination and forms a covalent bond with the 5’-ends of the DSB[238,239]. Unlike TOP1 and TOP2, SPO11 remains attached until the lesion is repaired as it cannot religate the DNA[240]. The Epstein-Barr virus nuclear antigen (EBNA1) forms covalent DPCs at the *oriP* episomal sequence via its active site tyrosine, which is important for viral episome maintenance. How EBNA1-DPCs are released afterwards, remains unknown[6,241]. Moreover, viral terminal proteins (TPs), such as TP-55 in adenovirus, form DPCs by linking their active site serine to the 5’-end of the viral genome[242,243] initiating viral replication in the case of TP-55[244]. Similar mechanisms are found in bacteriophages (*Bacillus subtilis* bacteriophage Φ29 and *E. coli* PRD1)[245–247]. DPCs can also form to shield labile DNA structures. In *Mycobacterium smegmatis* uracil-DNA glycosylase X (UdgX) excises uracil in DNA and forms a DPC with the resulting AP site, preventing strand breaks. Uracil in DNA is mainly formed during replication via cytosine deamination[248–250]. UdgX is a specific example, however, mechanisms like this for AP site shielding are evolutionary conserved[6]. In humans, 5-hydroxymethylcytosine (5-hmC) binding ESC-specific (HMCES) crosslinks to AP sites in ssDNA during replication[251].

1.2.3.1 HMCES

HMCES was initially identified in a mass spectrometry (MS) screen for the epigenetic nucleobase 5-hmC and proposed to act as an epigenetic reader[252], while others proposed a 5-hmC dependent nuclease activity[253]. However, no 5-hmC-related function could ever be confirmed[251]. Interestingly, a bacterial ortholog (YedK) exists, despite the absence of 5-hmC in bacteria, suggesting an alternative role[254]. Indeed, studies in the last years revealed that HMCES protects AP sites in ssDNA at stalled replication forks by forming a DPC via its catalytic SOS response-associated peptidase (SRAP) domain. Of note, human HMCES also carries a PIP-box in the unstructured *C*-terminal tail, through which it binds PCNA and travels with the replication fork[251]. The SRAP domain is highly conserved and carries the three catalytic residues, Cys2, Glu127 and His210 (Figure 4a)[255]. The crosslink forms at HMCES' *N*-terminal cysteine, Cys2 (methionine is proteolytically removed during translation) and the AP site through a thiazolidine ring, which is stabilised by Glu127 and His210 (Figure 4b-c). This mechanism starts with interactions of HMCES' *N*-terminal amino group with the open-ring aldehyde form of the AP site, followed by the formation of a Schiff-base intermediate with the sulfhydryl group of HMCES' Cys2 and subsequent conversion into a stable thiazolidine ring (Figure 4b)[256–258].

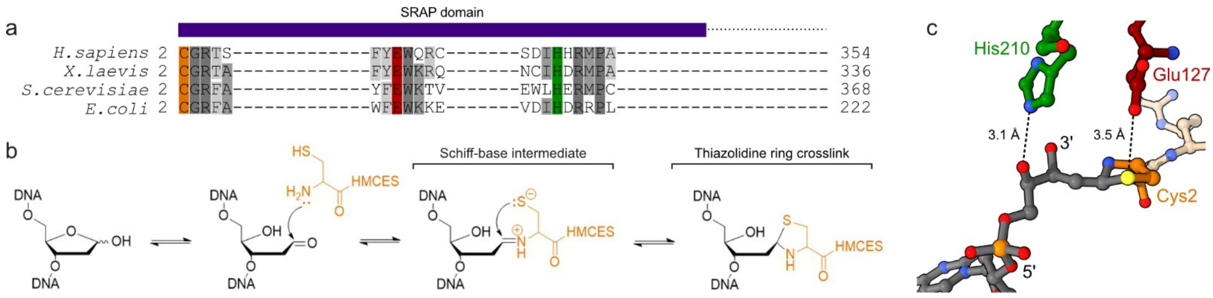


Figure 4. HMCES-DPC formation via its conserved SRAP domain. (a) Multiple sequence alignment of the SOS response-associated peptidase (SRAP) domain, highlighting active site residues in *H. sapiens, X. laevis, S. cerevisiae* and *E. coli* (Cys2 = orange, Glu127 = red, His210 = green). (b) Proposed reaction mechanism for the crosslink formation of SRAP with apurinic/apyrimidinic (AP) sites. (c) Crystal structure of SRAP's active site crosslinked to an AP site (PDB: 6oe7). Active site residues are labelled as in (a). DNA is coloured in grey. Interatomic distances (Å) are noted. Figure adapted from Donsbach *et al.*[259].

HMCES-DPCs are particularly important when cells face increased AP site formation, which can be caused by genotoxic agents (e.g. IR, UV, $KBrO_3$ or MMS)[260], upon overexpression of APOBEC3A[261,262] or by AID-induced hypermutations[263]. HMCES preferentially binds to 3'-ssDNA-dsDNA junctions typically found at stalled replication forks[257].

Beyond replication, HMCES also functions in DSB repair. By relying on its DNA-binding ability HMCES binds the ends of DSBs, where it promotes and mediates MMEJ[264]. Considering DDT pathways acting at stalled replication forks, HMCES-DPCs correlated with accumulation of the

TLS polymerase Polζ[251,261]. Moreover, in *X. laevis* egg extracts, HMCES crosslinks to AP sites, generated by the DNA glycosylase NEIL3 during ICL unhooking. Experiments in *X. laevis* egg extracts further indicated that, the Cdc45-MCM-GINS (CMG) helicase can bypass the AP site, before HMCES crosslinks[265], but polymerases stall at the HMCES-DPC. Resolution relies on proteolytic degradation of crosslinked HMCES by the SprT-like *N*-terminal domain (SPRTN) protease[265], requiring prior unfolding of the protein adduct by the FANCJ helicase[266]. TLS polymerases then synthesise over the remaining peptide, preferentially inserting a deoxyguanosine[265]. While these results could be reproduced *in vitro* with recombinant human proteins[266], HMCES-DPCs were primarily degraded by the proteasome in mammalian cells[251]. Notably, HMCES-DPCs disappeared over time upon SPRTN depletion and proteasomal inhibition in *X. laevis* egg extracts[265], suggesting an additional mechanism for HMCES-DPC resolution. Crosslink reversal has been recently proposed as an option[267], though how this reconciles with AP site protection in ssDNA remains unclear.

1.2.4 Methods for DPC detection

Growing interest in DPCs and their impact on genome stability has driven the development of reliable methods for their detection and identification. These approaches are generally categorised into protein-targeted and DNA-targeted methods, based on the DPC component used for detection[268].

1.2.4.1 Protein-targeted DPC detection methods

Protein-targeted methods isolate DPC-containing DNA from soluble or simply DNA-bound proteins, to subsequently remove the DNA and allow identification of the crosslinked proteins. The rapid approach to DNA adducts recovery (RADAR) and its derivatives, as well as caesium chloride (CsCl) density gradients belong to this group[268].

The RADAR uses a specialised “RADAR” buffer, high in detergents and chaotropic agents to precipitate DNA along with crosslinked proteins. Ethanol is added to facilitate DNA precipitation, followed by centrifugation steps and ethanol washes to remove salt and contaminants. The resulting DNA pellet is resuspended in an alkaline buffer, typically sodium hydroxide, and can be quantified and analysed using standard techniques such as immunoblotting (Figure 5a)[269]. The superior method for true DNA-protein crosslink recovery (STAR) adds an additional lysis step before DNA precipitation to the RADAR protocol (Figure 5b). This extra step is expected to reduce non-specific protein background and improve distinction between genuine DPCs and copurified contaminants[270].

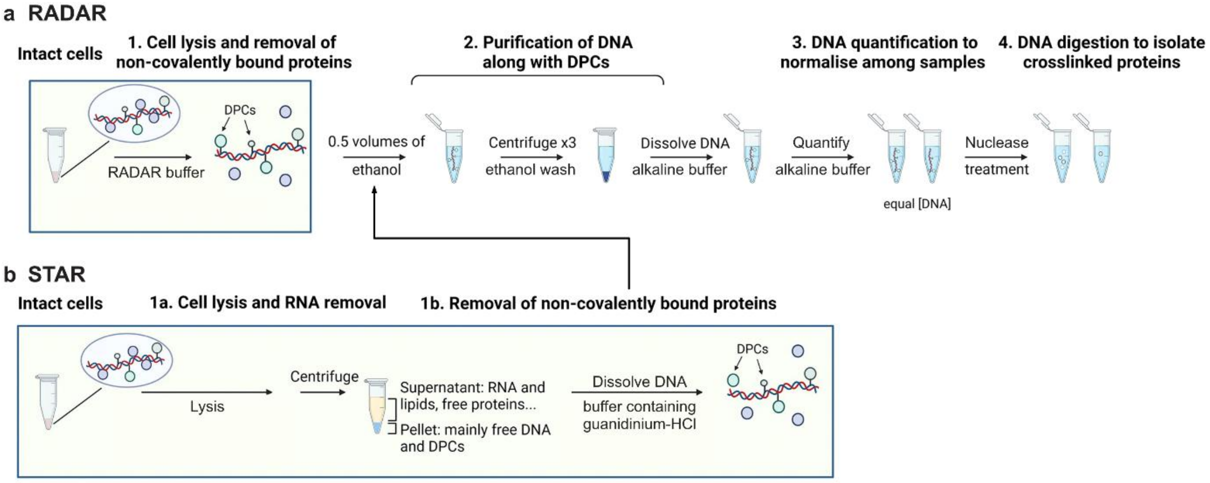


Figure 5. Schematic depiction of RADAR and STAR. (a-b) Depicted are the main steps of the rapid approach to DNA adducts recovery (RADAR) (a) and its derivative the superior method for true DNA-protein crosslink recovery (STAR) (b), used for the purification and analysis of DNA-protein crosslinks (DPCs). Figure adapted from Torrecilla *et al.*[268].

CsCl gradients offer an alternative for DPC purification that avoids precipitation, but instead uses ultracentrifugation to separate DPC-bound DNA from soluble proteins by density (Figure 6)[271]. However, this method is highly time consuming, as centrifugation steps take around 16 hours, DNA retrieval is prone to cross-contamination and sample throughput is limited by the capacity of ultracentrifuges[268].

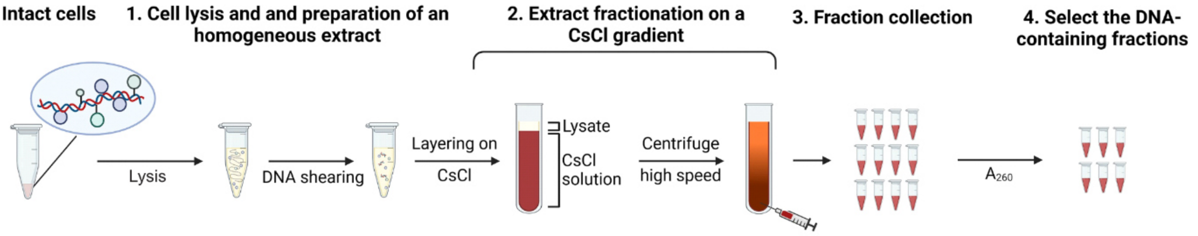


Figure 6. Schematic depiction of CsCl gradients. Depicted are the main steps of caesium chloride (CsCl) density gradients used for the purification and analysis of DNA-protein crosslinks (DPCs). Figure adapted from Torrecilla *et al.*[268].

1.2.4.2 DNA-targeted DPC detection methods

DNA-targeting methods estimate DPC levels by separating DPC-containing DNA from non-DPC-containing DNA. These techniques are primarily quantitative and cannot identify the crosslinked protein but can help localise DPCs within the genome. Common examples rely on protein precipitation, like the potassium chloride and sodium dodecyl sulphate coprecipitation assay (KCl-SDS) and advanced recovery of K-SDS precipitates (ARK), or use electrophoretic mobility as the comet assay[268].

The KCl-SDS assay involves lysing cells in an SDS-containing buffer, shearing the DNA (e.g. via sonication) and precipitating proteins with KCl at low temperatures. After high-speed centrifugation, soluble DNA remains in the supernatant while DNA containing DPCs coprecipitates with proteins in the pellet. The protein pellet is washed and again resolved in KCl buffer at high temperature. This washing step is repeated to eliminate non-specifically precipitated DNA. Finally, proteins are digested (e.g. with proteinase K) and the remaining DNA is quantified. The ratio of precipitated to soluble DNA reflects DPC levels (Figure 7a)[272]. Though KCl-SDS does not allow identification of crosslinked proteins, it is a fast and easy scalable method for quantitative analysis of DPCs. The ARK assay combines features of the KCl-SDS and the RADAR, using a chaotropic buffer for cell lysis followed by ethanol precipitation and KCl-SDS treatment. After precipitation the DNA pellet is resuspended in a KCl-SDS buffer, allowing isolation of DNA-protein complexes from the total DNA, which is further purified with acetone washes (Figure 7b). Compared to the RADAR, the ARK assay improves stringency and detection specificity[273].

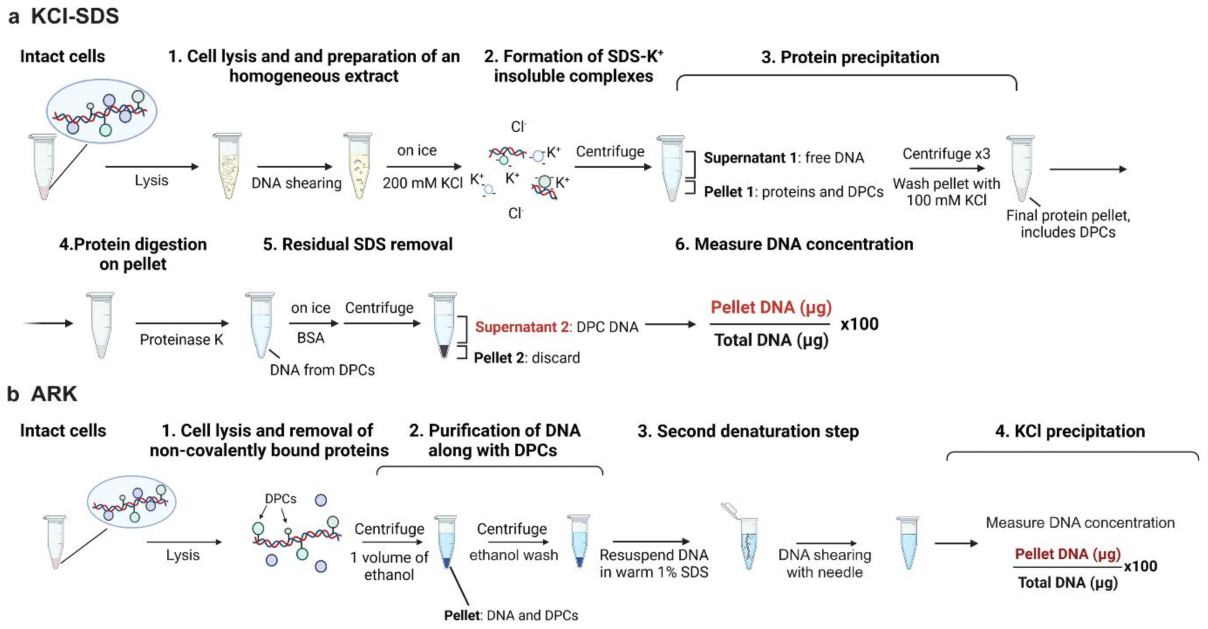


Figure 7. Schematic depiction of KCl-SDS and ARK. (a-b) Depicted are the main steps of the potassium chloride and sodium dodecyl sulphate coprecipitation (KCl-SDS) assay (a) and the advanced recovery of K-SDS precipitates (ARK) (b) used for the purification and analysis of DNA-protein crosslinks (DPCs). Figure adapted from Torrecilla *et al.*[268].

The comet assay is a single-cell electrophoresis method, usually used to assess DNA damage, mainly DNA strand breaks. Cells are embedded in agarose on a microscope slide and exposed to an electric field, which separates intact DNA from fragmented DNA, which will migrate in the electric field and create a comet tail[274]. Under alkaline conditions, the assay detects various types of lesions ranging from SSB and DSB over alkali-labile sites and DNA-

DNA crosslinks to DPCs[268,275]. To differentiate DPCs from other lesions, especially DNA-DNA crosslinks, proteinase K is added to digest crosslinked proteins and release DNA fragments allowing them to migrate further[276,277]. The difference in DNA fluorescence (the tail moment) between proteinase K treated and untreated samples allows quantification of DPC levels (Figure 8). However, compared to other methods, this method lacks sensitivity and does not allow for direct identification of the crosslinked protein[276–278].

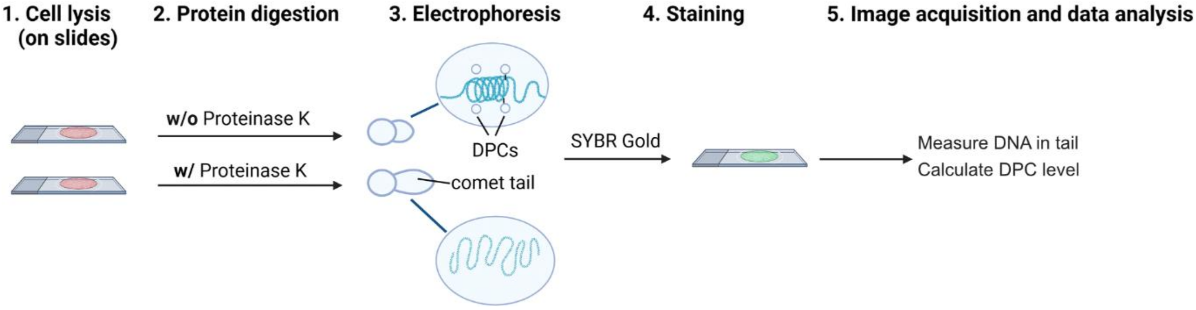


Figure 8. Schematic depiction of Comet assay. Depicted are the main steps of the comet assay used for the purification and analysis of DNA-protein crosslinks (DPCs). Figure adapted from Torrecilla *et al.*[268].

In addition, techniques for the detection of specific DPCs have emerged. For example, a monoclonal antibody against TOP1ccs distinguishes it from free TOP1. It can be used for flow cytometry and immunofluorescence without the need for prior DPC purification[279]. Moreover, PARP1-DPCs can be visualised on DNA fibres using a specific antibody[236]. A recent study described the protein-crosslinked DNA extraction (XDNAX) for the isolation of photo-crosslinked DNA-protein complexes. Here, non-crosslinked proteins were removed with a modified TRIZOL extraction, RNase digestion and silica column washes[280].

1.2.5 DPC Repair

Since DPCs consist of three different components – DNA, protein and the crosslink itself – each of them can be targeted and initiate repair pathways. However, DPC resolution typically requires a combination of pathways, acting consecutively. While repair varies by the type of DPC, proteolytic degradation of the crosslinked protein has emerged as a key step in many pathways[6,36].

1.2.5.1 The DPC Protease SPRTN

The first protease identified to specifically target DPCs was weak suppressor of smt3 (Wss1) in yeast[281]. Wss1-deficient cells exhibit sensitivity towards DPC-inducing agents (e.g. CPT) with strong effects on genome stability and cellular fitness[6,282,283] and they are hypersensitive to formaldehyde, implicating a role for Wss1 in the repair of Top1-DPCs and for DPCs induced by non-specific crosslinkers. Deletion of *Wss1* and tyrosyl-DNA phosphodiesterase (Tdp) 1, which acts in the repair of Top1-DPCs as it hydrolyses the covalent bond between Top1 and

DNA[284], led to growth defects, which could be rescued by *Top1* depletion, further supporting Wss1 acting in Top1-DPC repair[281]. Wss1's protease domain is located at the *N*-terminus and followed by a DNA-binding site and protein-protein interaction motifs including a site for interaction with the segregase Cdc48 (ortholog of mammalian p97) and SUMO interacting motifs (SIMs)[285]. Wss1 and Cdc48 interactions are mediated by Cdc48's cofactor Ubx5, which regulates Wss1-dependent DPC repair. Ubx5 accumulates at induced site-specific DPCs after Wss1 depletion, prohibiting alternative repair, however, DPC tolerance can be rescued to wild-type levels by deletion of *Ubx5*, suggesting a pathological role for Ubx5 in the absence of Wss1[286]. Wss1-like enzymes are found in bacteria, yeast and plants, playing crucial roles in genome maintenance. In metazoans, related proteases are present, belonging to the SPRTN-family. Despite early divergence, SPRTN-like and Wss1-like enzymes are both present in prokaryotes, they perform similar functions and have some structural domains in common[285].

Human SPRTN is a 55-kDa protein composed of a *N*-terminal metalloprotease domain, followed by a zinc-binding domain (ZBD) and a basic region (BR), that bind ssDNA and dsDNA, respectively[287,288]. The protease domain and the ZBD together form the conserved SprT domain. Its unstructured *C*-terminal tail carries motifs for interactions with the segregase p97 (SHP-box)[289,290], PCNA (PIP-box) and a ubiquitin-binding zinc finger (UBZ) [291,292]. Another motif interacting with ubiquitin (MIU) has been predicted in front of the protease domain (Figure 9a)[293]. While p97 involvement seems likely for SPRTN-dependent DPC resolution, where p97 potentially unfolds DPCs to enable SPRTN cleavage, the precise role of p97 in DPC repair is still uncertain[294], as is PCNA's role in SPRTN recruitment[291,292]. Notably, SPRTN-p97 and SPRTN-PCNA interactions seem at least partially dispensable as SPRTN variants lacking both the SHP- and PIP-box are still viable[36], in contrast to full *SPRTN* knockouts[295].

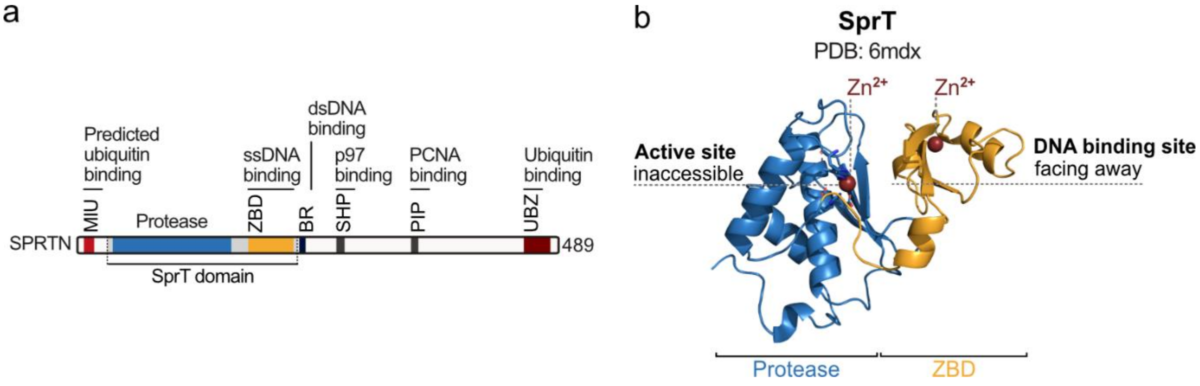


Figure 9. Structure and domains of the metalloprotease SPRTN. (a) Schematic depiction of SPRTN's domain structure, including the predicted motif interacting with ubiquitin (MIU), protease domain, zinc-binding domain (ZBD), basic region (BR), SHP-box for p97-binding, PCNA-binding domain (PIP) and the ubiquitin-binding zinc finger (UBZ). The conserved SprT domain, formed by the protease domain and the ZBD is highlighted. (b) Crystal structure of the SprT domain (aa28-214) (PDB: 6mdx). Protease domain is coloured in blue, ZBD in orange and Zn^{2+} ions in red. Figure adapted from Dürauer *et al.*[1].

Due to SPRTN's flexible *C*-terminal tail, attempts to obtain experimental structures of the human full-length protein were unsuccessful. However, in 2019, a crystal structure of SPRTN's SprT domain (aa28-214) was solved[296]. This structure indicated a Zn^{2+} ion coordinated within in the active site, similar to Wss1, using a well conserved HEXXH motif, mediated by H111, E112, H115 and H130 (Figure 9b)[295,296]. The glutamate side chain residue at position E112, marks the catalytic active site, mediating substrate cleavage by polarization of a water molecule. Mutation of E112 (E112Q) renders the enzyme catalytically inactive[295,296]. Moreover, this study discovered the adjacent ZBD[296], connected to the protease domain via a flexible linker and binding a second Zn^{2+} ion (Figure 9b). In the crystal structure, the ZBD folds back towards the protease domain, presenting a "closed" conformation and prohibiting access to SPRTN's active site[296]. Since the ZBD binds ssDNA, a conformational change upon DNA-binding, allowing access to the active site may occur[295,296].

Loss of SPRTN causes genome instability, with severity increasing with the complexity of the organism. For instance, SPRTN-deficient worms are highly sensitive to DPC-inducing agents but still viable[290,291,295]. In flies, loss of the SPRTN homolog maternal haploid (Mh) results in female sterility. Female flies can still produce eggs, though they do not hatch following fertilisation, because paternal chromosomes fail to condense during early embryonic mitosis, resulting in a loss of paternal DNA and inviable haploid embryos. This suggests that maternal SPRTN is essential for DPC repair in paternal DNA after fertilisation[297]. In mammals, SPRTN is essential, with complete loss of *Sprtn* in mice causing early embryonic lethality. Conditional knockout in mouse embryonic fibroblasts (MEF) leads to rapid proliferation arrest and severe genome instability, marked by micronuclei formation, chromatin bridges and γH2AX and RAD51 foci[298].

In humans, hypomorphic germline SPRTN mutations cause Ruijs-Aalfs syndrome (RJALS), characterised by premature ageing and early-onset liver hepatocellular carcinomas (Figure 10a)[293,299]. Three patients are known, with one carrying a homozygous 1-bp deletion at the beginning of exon 5, resulting in a frame shift and a premature stop codon, which deletes the *C*-terminal tail of SPRTN (K241+X8; $SPRTN^{\Delta C}$) (Figure 10b). The two other patients are compound heterozygotes for a similar $SPRTN^{\Delta C}$ allele caused by a 4-bp deletion at the end of exon 4 (K239+X7). In addition, these patients carry a second allele with a missense mutation close to SPRTN's catalytic active site (Y117C) (Figure 10b)[6,293]. $SPRTN^{\Delta C}$ lacks all features included in the *C*-terminal tail (SHP-box, PIP-box and UBZ) and mislocalises as the *C*-terminal nuclear localisation signal (NLS) is lost. Mouse models of hypomorphic SPRTN, with reduced expression of wild-type SPRTN, mimic RJALS phenotypes with premature ageing (Figure 10c) and high susceptibility for liver tumours[298,300]. In human cells, SPRTN deficiency compromises DPC repair and increases sensitivity to crosslinking agents[295,301,302].

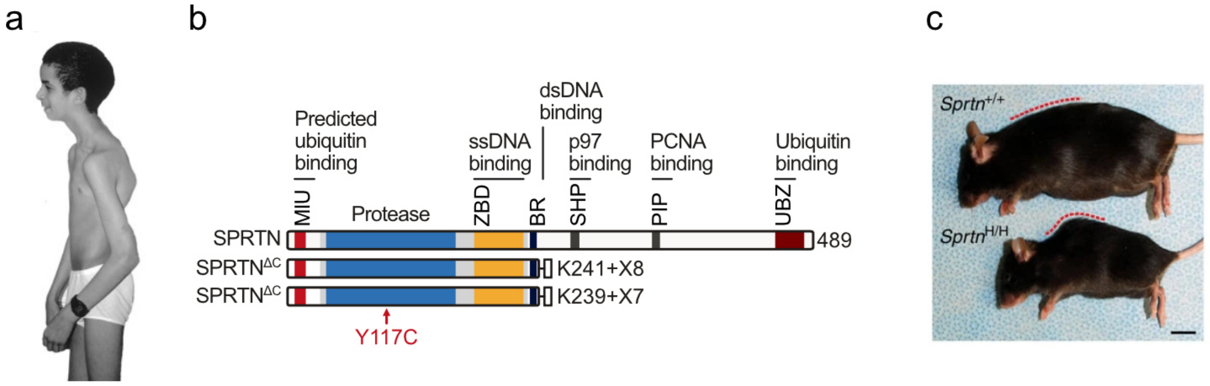


Figure 10. Ruijs-Aalfs syndrome SPRTN variants and phenotypes. (a) 15-year-old boy with Ruijs-Aalfs syndrome (RJALS), characterised by premature ageing and corresponding features. Figure adapted from Ruijs *et al.*[299] (b) Schematic depiction of SPRTN domain structure, RJALS patient SPRTN variants are labelled, with mutations (Y117C), premature stop codons (K241, K239) and additional eight/seven amino acids at the truncated *C*-terminal tails. Figure adapted from Dürauer *et al.*[1] (c) 12-month-old *Sprtn*$^{+/+}$ and SPRTN deficient (*Sprtn*$^{H/H}$) mouse, showing progeroid phenotypes, like lordokyphosis labelled by red dotted lines. Scale bar = 1 cm. Figure adapted from Maskey *et al.*[298].

1.2.5.2 The proteasome

The 26S proteasome is the primary proteolytic machine in cells. It belongs to the AAA+ family and features a barrel-like structure that uses adenosine triphosphate (ATP) hydrolysis to translocate unfolded proteins through axial channels to its proteolytic chamber for degradation[303,304]. While it typically recognises substrates tagged with polyubiquitin, recent studies showed it can also degrade proteins independent of ubiquitylation, relying on degron sequences or shuttling factors, like ubiquilins[305]. Its role in DPC repair was confirmed in replication-coupled DPC repair in *X. laevis* egg extracts, mediated by specific ubiquitin E3-ligases (see section 1.4.5)[306].

1.2.5.3 Other DPC proteases

Besides SPRTN and the proteasome, several other proteases have been identified to be involved in DPC repair[307]. In yeast, DNA damage inducible 1 (Ddi1) acts alongside Wss1. Ddi1 depletion in Wss1-deficient yeast cells enhances sensitivity towards Top1-DPCs and formaldehyde-induced DPCs[308,309]. Like the proteasome, Ddi1 degrades polyubiquitylated substrates, via a retroviral-like protease domain[310]. In humans two Ddi1 homologs (DDI1 and DDI2) were identified, acting similar to yeast Ddi1 and involved in degradation of replication termination factor 2 (RTF2), preventing defective fork restarts[311]. Their direct role in DPC repair remains inconclusive, though DDI2 proteolytically activates the transcription factor nuclear respiratory factor 1 (NRF1) upon proteasomal inhibition, promoting expression of proteasomal subunits[312]. Notably DDI1/DDI2 activity appears linked to polyubiquitylation, suggesting a potential shuttling function where they guide DPCs to the proteasome[313]. Acidic repeat-containing protein (ACRC), also called germ cell nuclear antigen (GCNA), is another DPC protease, primarily active in germ cells and during meiosis. It bears a SprT-protease domain and an intrinsically disordered acidic region. ACRC/GCNA deficiency leads to sterility and genome instability in various species, which is linked to defective SPO11-DPC repair during meiosis[314]. Combined loss with SPRTN causes hypersensitivity to DPC-inducing agents like etoposide and hydroxyurea, associated with accumulation of TOP2-DPCs[315]. Interestingly, in rodents ACRC/GCNA lacks the protease domain, yet depletion still causes sterility, suggesting non-proteolytic roles in DPC repair[316]. Finally, trypsin-like serine proteases of the FAM111 family, have emerged as DPC-proteases. In addition to the *C*-terminal protease domain and two ubiquitin-like domains, FAM111A possesses a PIP-box and a ssDNA-binding domain[317]. FAM111A localises to replication forks and resolves protein adducts, mainly TOP1-DPCs. FAM111A loss slows down replication forks and sensitises cells to TOP1 poisons and PARPi, suggesting a role in replication fork integrity[318]. Hyperactive FAM111A variants are cytotoxic[319] and cause Kenny-Caffey syndrome type 2, characterised by impaired skeletal development and hypoparathyroidism[320]. A proteolytically hyperactive variant of FAM111B causes hereditary fibrosing poikiloderma, associated with pulmonary fibrosis, tendon contractures and myopathy[319]. Both FAM111A and FAM111B have also been implicated as antiviral restriction factors, where the link to DPC repair remains unclear[321,322].

1.2.5.4 Repair by direct crosslink hydrolysis

Since TOP1ccs and TOP2ccs are frequent DNA lesions, specialised enzymes TDP1 and 2, have evolved to resolve them. TDP1 and TDP2 act by hydrolysing the phosphodiester bond between trapped TOP1/TOP2 and DNA[6].

TDP1 primarily targets 3’-DNA-ends, where it mainly processes TOP1ccs, but it can also process other lesions, like AP sites or 3’-phosphoglycolates[284]. It leaves a 3’-phosphate, which is further dephosphorylated by PNKP. PNKP also phosphorylates the 5’-end to allow canonical SSB repair[323]. Due to the bulkiness of TOP1ccs, TDP1 often requires prior proteolysis to access the TOP1-DNA interface[324]. Proteolysis, performed by the proteasome, includes PIAS4 mediated SUMOylation of the TOP1ccs and subsequent ubiquitylation by the SUMO-targeted ubiquitin E3-ligase (STUbL) ring finger protein 4 (RNF4)[325,326]. Structural studies have shown that TDP1 locally melts the DNA during 3’-end processing[284]. TDP1 loss causes hypersensitivity to CPT and knockouts in yeast are synthetically lethal in combination with DSB-repair genes[327] and Wss1[295]. Germline mutations in *TDP1* are linked to spinocerebellar ataxia with axonal neuropathy (SCAN1), a neurodegenerative disorder caused by the gain of function mutation H493R, which disrupts TDP1’s active site and traps it on DNA[328,329]. Post-translation modifications (PTMs), such as phosphorylation by DDR kinases and PARylation by PARP1 enhance TDP1 stability and interactions with XRCC1[330] and PARP1[331,332].

In contrast, TDP2 processes 5'-linked TOP2ccs without the need of prior proteolysis[333,334]. Though proteasomal degradation can assist the process, as it does for TDP1[325,326]. More important is SUMOylation of the TOP2cc by the E3-ligase ZATT, which reshapes the TOP2cc, allowing TDP2 to access the DNA-protein interface and hydrolyse the bond. The result is a clean DSB with a 5'-phosphate, which is further processed by NHEJ[335,336]. TDP2 recruitment is aided by SUMO2/3 and ubiquitin, which bind to TDP2's *N*-terminal ubiquitin associated (UBA) domain[337]. TDP2 loss causes hypersensitivity to TOP2 poisons, like etoposide[338] and *TDP2* mutations lead to a premature stop codon, resulting in enzymatic loss of function and are linked to spinocerebellar ataxia, autosomal recessive 23 (SCAR23)[336].

1.2.5.5 Repair by nucleolytic cleavage

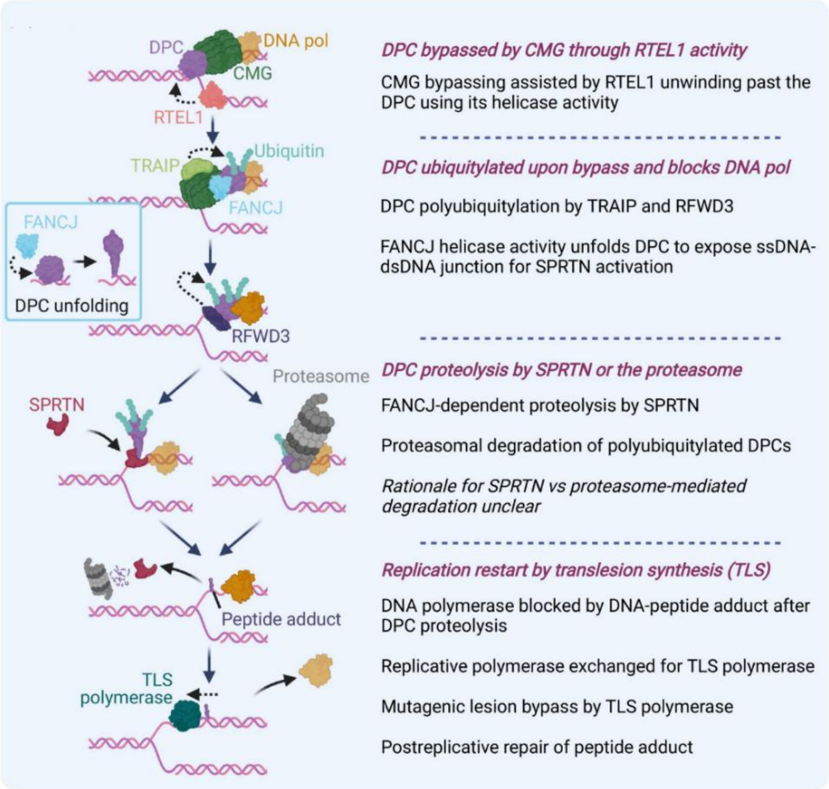
DPCs can be processed nucleolytically by canonical DNA repair enzymes, like the MRN complex. The MRN complex processes DSBs by cleaving DNA ends using endo- and exonuclease activity. DPCs resolved by MRN include the physiological SPO11-DPCs and drug-induced TOP2ccs[240]. Its endonuclease activity is activated by protein adducts at DNA ends, resulting in a DNA clipping 15-25 bps away from the lesion. Notably, while MRN has no specificity for the nature of the block, its activity is strictly limited to DNA ends[339–341]. Structural studies of the MR complex in bacteria (prokaryotes lack NSB1) indicated a heterotetrameric structure, formed by two molecules of both RAD50 and MRE11. The two coiled-coils domains of RAD50 form a ring around the DNA, scanning for DSBs. Once DNA ends are detected, a conformational change is initiated, where the coiled-coils lock the complex next to the DPC via ring-to-rod transition, triggering endonucleolytic cleavage by MRE11[342]. Recently, this structure and screening process was confirmed for the human MRN complex[343]. However, DPCs that bridge DNA ends, like the TOP2ccs, likely require partial proteolysis upstream, potentially mediated by ZATT-dependent SUMOylation and ubiquitylation[325,342]. Additionally, NER enzymes were reported to remove small DPCs (≤ 10 kDa)[344–347], suggesting that bulkier DPCs might always require prior proteolytic processing for repair via nucleolytic cleavage[6].

1.2.5.6 Replication-coupled DPC repair

Studies using *X. laevis* egg extract provided an elucidating view into the sequence of molecular events, that occur when replication forks encounter DPCs (Figure 11). During replication the CMG helicase performs DNA unwinding by translocating on the leading strand. CMG can bypass protein adducts on the lagging strand, via a process potentially involving partial opening of a lateral channel in the CMG complex[306]. However, CMG initially stalls at DPCs on the leading strand, requiring a second helicase, regulator of telomere elongation 1 (RTEL1), to load on the lagging strand and unwind DNA downstream of the DPC. This generates a stretch of ssDNA, which allows CMG to also bypass this DPC[348]. Nonetheless, DNA polymerases cannot bypass DPCs, leading to uncoupling of helicase and polymerase.

Uncoupling triggers ubiquitylation of the DPC by two distinct E3-ligases (see section 1.4.5) and promotes proteasomal degradation of the protein adduct (Figure 11)[6,306,349–351].

In parallel or as an alternative mechanism, the DPC protease SPRTN is recruited to the ssDNA-dsDNA junction created by stalled replication forks[285,295]. SPRTN-related DPC-cleavage requires prior unfolding of the crosslinked protein, performed by the FANCJ helicase[266]. FANCJ loads on the ssDNA downstream of the DPC travelling 5'-3'. By translocating into the DPC, FANCJ unfolds the protein adduct and exposes the underlying ssDNA-dsDNA junction, enabling efficient DPC cleavage by SPRTN[266]. Interestingly, SPRTN-mediated DPC-cleavage did not require DPC ubiquitylation in *X. laevis* egg extract[306]. Once the protein adduct has been degraded by SPRTN or the proteasome, TLS polymerases synthesise across the remaining peptide and initiate downstream repair (Figure 11). This comes with a risk of mutagenesis due to the low fidelity of TLS polymerases[352]. How the remaining peptide adduct is removed remains unclear[351].



DPC bypassed by CMG through RTEL1 activity

CMG bypassing assisted by RTEL1 unwinding past the DPC using its helicase activity

DPC ubiquitylated upon bypass and blocks DNA pol

DPC polyubiquitylation by TRAIP and RFWD3

FANCJ helicase activity unfolds DPC to expose ssDNA-dsDNA junction for SPRTN activation

DPC proteolysis by SPRTN or the proteasome

FANCJ-dependent proteolysis by SPRTN

Proteasomal degradation of polyubiquitylated DPCs

Rationale for SPRTN vs proteasome-mediated degradation unclear

Replication restart by translesion synthesis (TLS)

DNA polymerase blocked by DNA-peptide adduct after DPC proteolysis

Replicative polymerase exchanged for TLS polymerase

Mutagenic lesion bypass by TLS polymerase

Postreplicative repair of peptide adduct

Figure 11. Replication-coupled DPC repair. During replication the CMG helicase bypasses DNA-protein crosslinks (DPCs), with help of another helicase RTEL1, however replicative DNA polymerases (DNA pol) get stalled. The DPC then gets polyubiquitylated by the ubiquitin E3-ligases TRAIP and RFWD3, promoting proteasomal degradation of the protein adduct. Alternatively, DPCs can be degraded independently of the proteasome, beginning with FANCJ-mediated DPC unfolding. This recruits the DPC-protease SPRTN, which gets activated by single-stranded (ss)-double-stranded (ds) DNA junction, present at stalled forks. Both SPRTN and the proteasome leave behind a DNA-bound peptide, which can be bypassed by translesion synthesis (TLS) polymerases. Figure adapted from Carnie *et al.*[351].

1.2.5.7 Global-genome DPC repair

Replication-independent repair of DPCs was first described for TOP1ccs and TOP2ccs, which form cell cycle phase independent and are proteolytically processed by TDP1 and TDP2 (see section 1.2.5.4)[6]. Involvement of SPRTN and the proteasome in global-genome DPC repair was described for DNMT1-DPCs induced by the antineoplastic drug 5-azadC[36]. Here, repair is initiated by polySUMOylation of the DPC, mediated by DNA-resident SUMO-E3 ligases of the PIAS family[291]. SUMOylation triggers further ubiquitylation of the DPC by STUbLs (see section 1.4.5)[351]. Similar to replication-coupled DPC repair, DPC ubiquitylation leads to proteolysis of the protein adduct by the proteasome and SPRTN (Figure 12)[325,353,354]. The remaining peptide adduct requires further processing. NER has been reported to remove these adducts, although the exact underlying mechanism remains elusive[347,355,356].

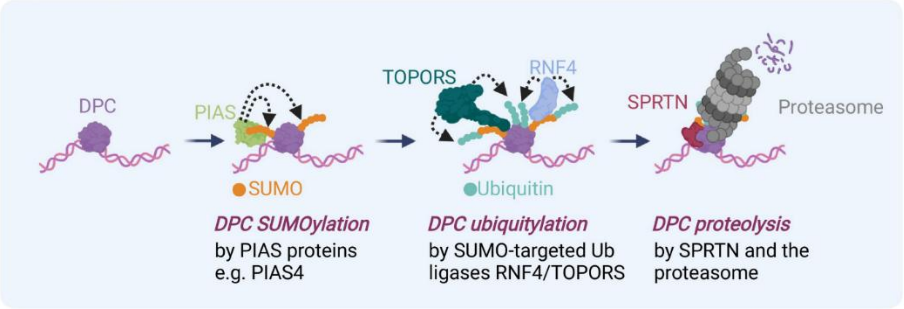


Figure 12. Global-genome DPC repair. In global-genome DNA-protein crosslink (DPC)-repair, the DPC first gets SUMOylated by SUMO E3-ligases (e.g. PIAS4). The SUMO-chains are recognised by the SUMO-targeted ubiquitin E3-ligases RNF4 and TOPORS, leading to polyubiquitylation of the DPC, promoting proteolytic degradation of the protein adduct by SPRTN or the proteasome. Figure adapted from Carnie *et al.*[351].

1.2.5.8 Transcription-coupled DPC repair

Recent studies revealed DPC repair pathways acting during transcription. A process distinct from classic TC-NER, however the initiating events, including CSB recognition of stalled RNA Pol II and CSA recruitment are shared (Figure 13)[357–359]. In TC-NER, CSB is stabilised by UVSSA, which associates with TFIIH and the XPD subunit being responsible for lesion verification and ubiquitin-dependent RNA Pol II removal (see section 1.4.5). This allows subsequent repair steps by XPA, XPG and XPF-ERCC1[94,95,360,361]. Transcription-coupled DPC repair in contrast, does not require RNA Pol II ubiquitylation and UVSSA seems to play a minor role, as only mild dependencies for the repair of formaldehyde-induced DPC and transcription recovery were described[351,358,359]. Whether TFIIH is required remains unclear, as the bulky nature of DPCs may prevent its loading on DNA. TFIIH is recruited to stalled RNA Pol II upon formaldehyde treatment, which is CSB-dependent but independent of UVSSA[351,358,359]. Moreover, XPA, a core licensing factor in TC-NER is dispensable in transcription-coupled DPC repair[357–359]. XPA loss does not impair transcription recovery or sensitise cells to DNMT1-

recovery or sensitise cells to DNMT1-DPCs[357]. This indicates that transcription-coupled DPC repair may use a distinct licensing mechanism to distinguish DPC-stalled RNA Pol II from smaller lesions, which can undergo conventional NER[351].

Unlike NER, transcription-coupled DPC repair does not require excision of the lesion-containing DNA stretch, since formaldehyde-induced DPCs do not trigger oligonucleotide excision and cells lacking XPF-ERCC1 or XPG show no delay in transcription recovery[357,359].

The fact that CSA recruitment by CSB is crucial for transcription-coupled DPC repair but ubiquitylation of RPB1 is not essential, lead to the hypothesis that CSA might instead directly polyubiquitylate the crosslinked protein[357–359]. Polyubiquitylation marks the protein for degradation by the proteasome, which requires prior activity of the unfoldase p97 (Figure 13)[358,359]. The fate of the remaining peptide crosslinked to DNA remains unknown. After lesion removal, transcription resumes via proteasomal degradation of the damage-induced inhibitory transcription factor, activating transcription factor 3 (ATF3)[357,362,363].

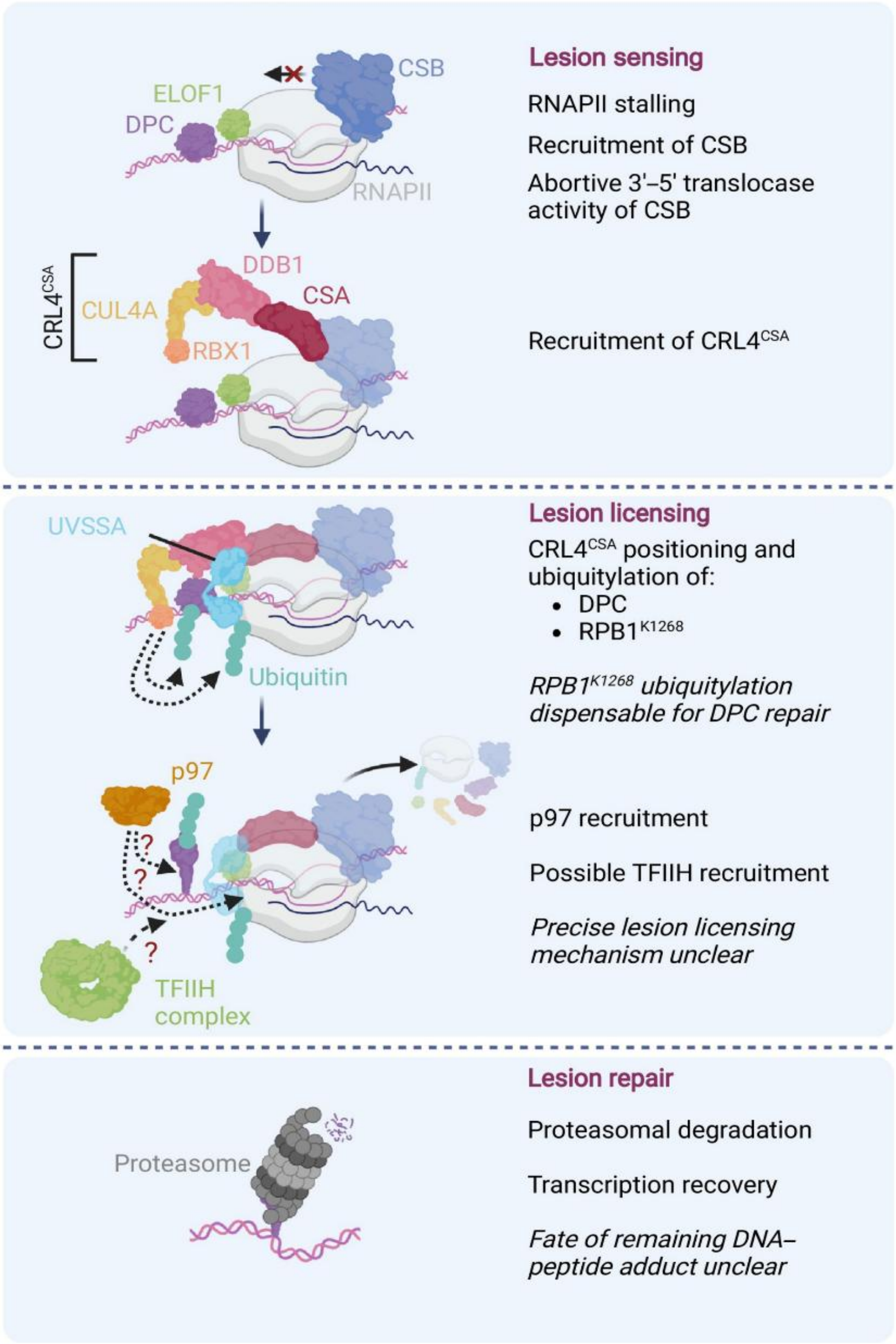


Figure 13. Transcription-coupled DPC repair. Transcription-coupled DNA-protein crosslink (DPC)-repair is initiated when RNA polymerase II (RNAPII) stalls at a DPC, prompting recognition by CSB, which then recruits the CRL4^{CSA} complex. This leads to polyubiquitylation of RNAPII's RPB1 subunit at K1268, a process likely assisted by ELOF1 and UVSSA, though this step is not essential for DPC repair. The critical function of CRL4^{CSA} is instead believed to lie in polyubiquitylation of the DPC itself, enabling its degradation by the proteasome, with the help of the ubiquitin-dependent segregase p97, which unfolds the DPC. However, the precise role of p97 and a potential contribution of the TFIIH complex remain unresolved. Figure adapted from Carnie *et al.*[351].

1.3 Regulation of SPRTN activity

A protease acting in direct proximity to DNA, like SPRTN requires tight regulation not only to avoid unintended degradation of chromatin-bound proteins, but also since SPRTN lacks substrate specificity. Over the past years some regulatory mechanisms underlying SPRTN's protease activity have been revealed[6].

SPRTN was initially identified as a regulator for TLS. Early studies showed that SPRTN loss led to prolonged retention of Polη on chromatin following DNA damage[289,290]. Via its PIP-box and its UBZ, SPRTN localises to UV-induced DNA damage sites, where it potentially recognises ubiquitylated PCNA[364]. Moreover, SPRTN interacts with the segregase p97, via its SHP domain. This was originally thought to promote Polη removal from chromatin[289,290]. However, later findings suggest SPRTN may instead facilitate Polη recruitment and promote TLS initiation via interaction with RAD18[364]. Thus, its precise role in TLS remains uncertain. Another study reported a joint function of SPRTN, p97 and Testis-expressed protein 264 (TEX264) in the repair of TOP1ccs. Here, TEX264 acts as a sensor recruiting p97 with SPRTN to the TOP1ccs. p97 then unfolds crosslinked TOP1, enabling SPRTN-mediated cleavage. Since TEX264 is not involved in most DPC repair pathways, this mechanism appears specific for TOP1cc[292].

Further investigations of SPRTN activity revealed three independent key molecular switches for SPRTN regulation (Figure 14)[6]. First, DNA-dependency, SPRTN and Wss1 are both inactive by themselves but get activated upon DNA binding[281,295,301]. DNA likely serves both as a scaffold, recruiting proteases to their substrates and acts as a trigger for conformational changes to expose their active site[295,296]. SPRTN activation requires DNA-interactions with its both DNA-binding domains (ZBD and BR). Since ZBD and BR recognise ssDNA and dsDNA, respectively, SPRTN activation only occurs at locations containing both structures, such as stalled replication forks or DNA bubbles. However, even very limited amounts of ssDNA are efficient for SPRTN activation. For example, dsDNA ends, or DNA nicks and gaps can activate SPRTN (Figure 14a). Once activated, SPRTN can cleave protein adducts near the activating DNA structure[287]. In addition to substrate cleavage, DNA-binding also triggers SPRTN autocleavage *in trans*, serving as a self-inactivation off switch that removes it from damage sites (Figure 14b)[295,301]. This mechanism is further regulated by ubiquitin[365]. In cells, SPRTN exists in two states, an unmodified and a constitutively monoubiquitylated one, with its UBZ shielding the modification[289,290,364,365]. Monoubiquitylation on the one hand primes SPRTN for polyubiquitylation followed by proteasomal degradation, while on the other hand it enhances autocleavage *in trans* (Figure 14b)[365]. Upon DPC recognition, SPRTN is deubiquitylated by ubiquitin-specific protein 7 (USP7), reducing autocleavage and proteasomal degradation, which increases SPRTN stability and availability for DPC repair (Figure 14c)[365]. Besides USP7, two other deubiquitylating (DUB) enzymes were suggested to be involved in SPRTN deubiquitylation, USP11 and VCPIP1[366,367].

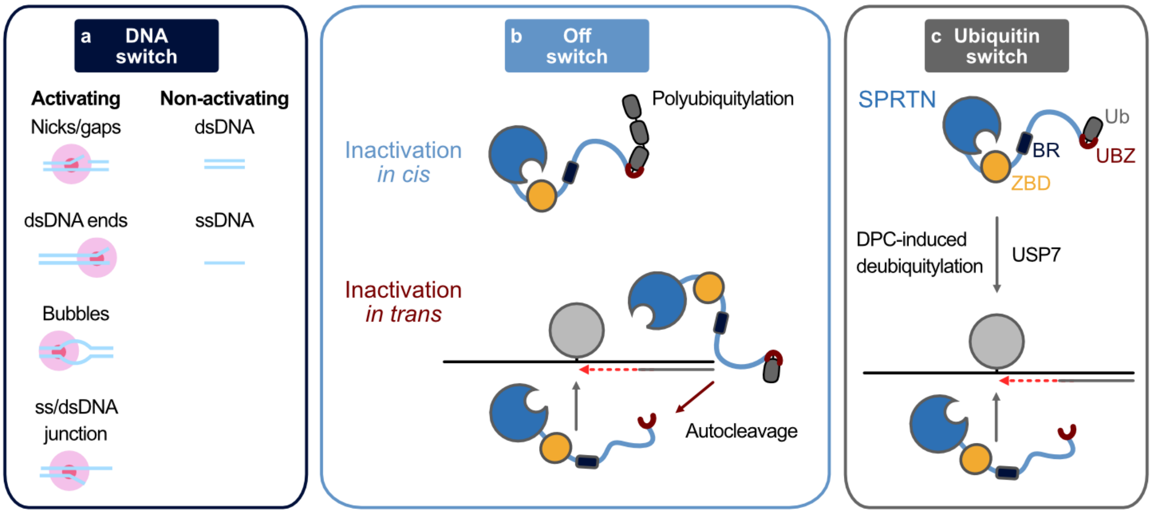


Figure 14. Regulatory mechanisms for SPRTN activity. Three molecular switches regulate SPRTN's proteolytic activity (a) DNA switch, SPRTN only gets activated by binding to DNA structures featuring ssDNA and dsDNA, recognised by the Zinc-binding domain (ZBD) and the basic region (BR), respectively. (b) Off switch, in cells SPRTN is monoubiquitylated and the ubiquitin (Ub) molecule is protected by its ubiquitin-binding zinc finger (UBZ). This modification negatively regulates SPRTN *in cis* by triggering polyubiquitylation and subsequent degradation by the proteasome and *in trans* by inducing autocleavage. (c) Presence of DNA-protein crosslinks (DPCs) triggers deubiquitylation of SPRTN by USP7, which stabilises the enzyme and promotes DPC cleavage. Figure adapted from Weickert and Stingele[6].

Further regulatory effects of ubiquitin were suggested, wherein ubiquitin may regulate SPRTN recruitment, as DPCs get ubiquitylated during repair[306,325,348,353,354]. However, SPRTN still cleaved DPCs in *X. laevis* egg extract when ubiquitylation was blocked[306], but ubiquitylated DPCs accumulate after SPRTN depletion, suggesting that SPRTN preferentially targets them[368]. Furthermore, SPRTN's UBZ has been proposed to mediate interactions with ubiquitylated DPCs, as UBZ-deficient mutants show impaired DPC cleavage in mammalian cells and *X. laevis* egg extracts[306,353]. Yet, SPRTN$^{\Delta C}$ which lacks the entire *C*-terminal tail, including the UBZ (Figure 10b), is viable[293] and rescues phenotypes of inducible *SPRTN* knockouts in MEFs[300]. Thus, the exact role of DPC ubiquitylation and SPRTN's UBZ in the regulation of SPRTN's activity remains unclear.

1.4 Ubiquitin and its role in DNA repair

Ubiquitin (Ub) is a small 76-residue (8.6 kDa) protein highly conserved across eukaryotes. It adopts a so-called β-grasp fold, consisting of five anti-parallel β-sheets, holding a single α-helix[369–371]. During ubiquitylation, the *C*-terminal glycine residue of ubiquitin gets covalently attached to lysine residues of its target protein via an isopeptide bond (see section 1.4.1). Ubiquitylation is the major PTM in eukaryotic cells[372]. Unlike other PTMs, such as phosphorylation, methylation, acetylation and glycosylation, an entire protein is attached to the substrate[373]. Ubiquitylation was originally described as a signal for proteasomal degradation[372]. However, by now ubiquitin could be linked to various cellular processes, including protein quality control, apoptosis, cell cycle progression, endocytosis, DNA repair pathways, autophagy, cellular signalling pathways, trafficking, transcription and immune response[374–378].

Notably, ubiquitin itself can also be modified, expanding its regulatory potential. It can be further ubiquitylated (polyubiquitylated) through each of its seven lysine residues and the *N*-terminus (see section 1.4.2)[374,379], certain residues can be phosphorylated[380,381] and six of the seven lysine residues can be acetylated[382]. Moreover, deamidation, ADP-ribosylation[383] and modifications with other ubiquitin-like proteins (UBLs) have been reported[384,385].

Despite low sequence similarity, UBLs share the characteristic β-grasp fold with ubiquitin[386,387]. UBLs are distinguished depending on whether they are conjugated to substrates or not[388,389]. Type I UBLs possess a *C*-terminal glycine motif and are conjugated to substrates in a similar enzymatic cascade as ubiquitin (see section 1.4.1). This group includes SUMO, neural precursor cell expressed developmentally downregulated protein 8 (NEDD8), interferon-stimulated gene 15 (ISG15), HLA-F adjacent transcript 10 (FAT10), autophagy-related gene (ATG) 8 and ATG12, ubiquitin-related modifier (URM) 1 and ubiquitin-fold modifier (UFM) 1. As ubiquitin, they are involved in various cellular processes, such as protein degradation (NEDD8, FAT10), transcription regulation and DNA repair (SUMO, NEDD8), immune response (FAT10, ISG15 – antiviral defence), autophagy (ATG8, ATG12), tRNA modification (URM1) and stress response (SUMO, UFM1 – endoplasmic reticulum stress response)[389,390]. Type II UBLs are not conjugated and mostly found in multi-domain proteins. While not being directly related to the conjugation process, some are found in certain E1- and E3-enzymes[391,392]. Of note, the protein FUBI is released after proteolytic cleavage of the FAU protein and possesses a *C*-terminal type I-like diGly motif, but to date there is no evidence for conjugation[393].

1.4.1 Ubiquitylation

Ubiquitylation is a dynamic, reversible process that attaches Ub-molecules to target proteins in a thioester cascade regulated by three classes of enzymes, Ub-activating enzymes (E1), Ub-conjugating enzymes (E2) and Ub-ligases (E3)[372]. The process begins with an ATP-dependent thioester bond formation between ubiquitin's *C*-terminal Gly76 and the cysteine group of the E1, followed by ubiquitin transfer to the active site cysteine residue of the E2-enzyme and further transmission to the substrate mediated by E3-ligases. Next, an isopeptide bond between a lysine residue of the substrate and ubiquitin's Gly76 is formed (Figure 15)[374,394].

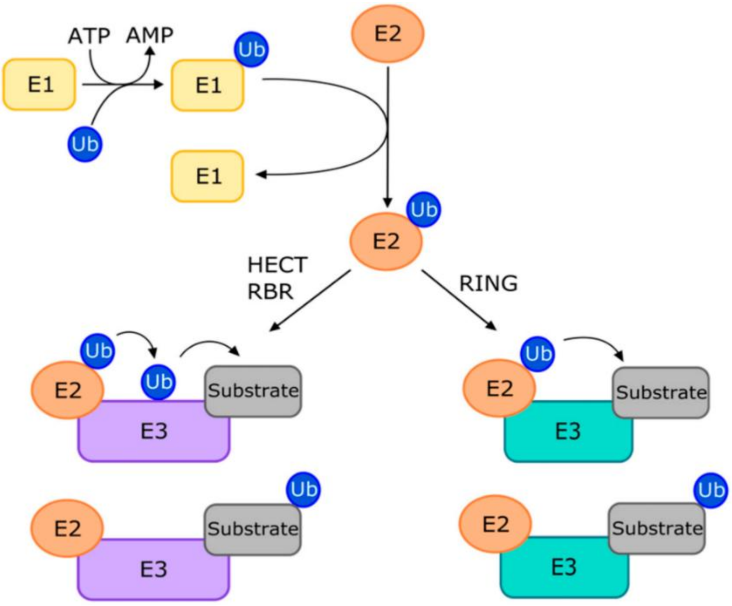


Figure 15. Enzymatic ubiquitylation cascade. Ubiquitylation starts with the activating E1-enzyme, which binds ubiquitin (Ub) in an ATP-consuming reaction. Next Ub is transferred to the conjugating E2-enzyme. Ub is further transferred to the substrate, with the help of E3-ligases. HECT and RBR E3-ligases use a two-step approach, where they first covalently bind Ub themselves. RING E3-ligases bring the Ub-carrying E2 and the substrate together and catalyse a direct transfer of Ub. Figure adapted from LaPlante and Zhang[395].

Humans possess two E1-enzymes (UBA1 and UBA6), approximately 40 E2-enzymes[396,397] and over 600 E3-ligases[372,374,394] for ubiquitylation. Three classes of E3-ligases are distinguished, based on the ubiquitin transfer mechanism. First, really interesting new gene (RING) E3-ligases, which act as scaffolds to enable direct ubiquitin transfer from the E2-enzymes to the substrate (Figure 15). RING E3-ligases are the most abundant group and recruit E2-enzymes either via a zinc-binding domain or a U-box[374,398,399]. Second, homologous to the E6AP carboxyl terminus (HECT) E3-ligases, like E1- and E2-enzymes, they first form a thioester intermediate bond with ubiquitin before transferring it to the substrate (Figure 15). The bond is formed via the catalytic cysteine at the *C*-terminus[374,399,400]. Finally, RING-between-RING (RER) E3-ligases, which combine features of the two other families. They possess two RING domains, one to interact with E2-enzymes (RING1), similar to a standard RING-domain, and a second one (RING2) bearing a catalytically active cysteine for the formation of intermediate linkages with ubiquitin[374,399,401]. Beyond canonical ubiquitylation at lysine residues ubiquitin can form a hydroxyester bond with serine, threonine and tyrosine residues as well as a peptide bond with *N*-termini of target proteins[402].

As mentioned before, ubiquitylation is reversible, with the main enzymes responsible for removal of Ub-moieties being DUBs. DUBs are proteases that fine-tune Ub-signalling by regulating and modifying chain architecture. They can recognise mono- or polyUb-chains and remove Ub-moieties by hydrolysing the isopeptide bond between ubiquitin's Gly76 and the substrate's lysine residue[374,403]. Seven DUB families are distinguished based on their catalytic domain. Six are cysteine proteases: USP, Ub *C*-terminal hydrolases (UCH), Machado-Joseph disease proteases (MJD), ovarian tumour domain proteases (OTUs), motif interacting with Ub containing novel DUB family (MINDYs) and zinc finger with UFM1-specific peptidase domain protein proteases (ZUFSPs). The seventh family is the metalloprotease JAB1/MPN/MOV34 (JAMMs) family. More than one DUB can act in parallel on the same substrate, allowing precise modulation and ubiquitylation dynamics[374,404,405].

1.4.2 Ubiquitin linkage types

Ubiquitylation varies not only in the number of Ub-molecules attached but also in the linkages formed. A single ubiquitin attached to substrates is termed monoubiquitylation[406], while multiple Ub-moieties on distinct lysine residues of the same substrate are called multi-monoubiquitylation[374,407]. These modifications often regulate protein localisation, activity or interactions, influencing processes like endocytosis, protein trafficking and chromatin remodelling[408–412]. Polyubiquitin chains can form on each of ubiquitin's seven lysine residues or at the *N*-terminal methionine (M1)[413,414]. Polyubiquitin chains are named after the lysine residue they are built on (K6-, K11-, K27-, K29-, K33-, K48- and K63-chains) and M1-linked for polyubiquitin chains formed on the *N*-terminal methionine (Figure 16)[379,383,415]. If all Ub-molecules within a polyubiquitin chain are linked via the same Lys, one is referring to homotypic chains (Figure 16c), whereas heterotypic chains combine different Ub-linkages in one chain (Figure 16d) and heterologous chains can be formed together with other PTMs (Figure 16e)[374,385].

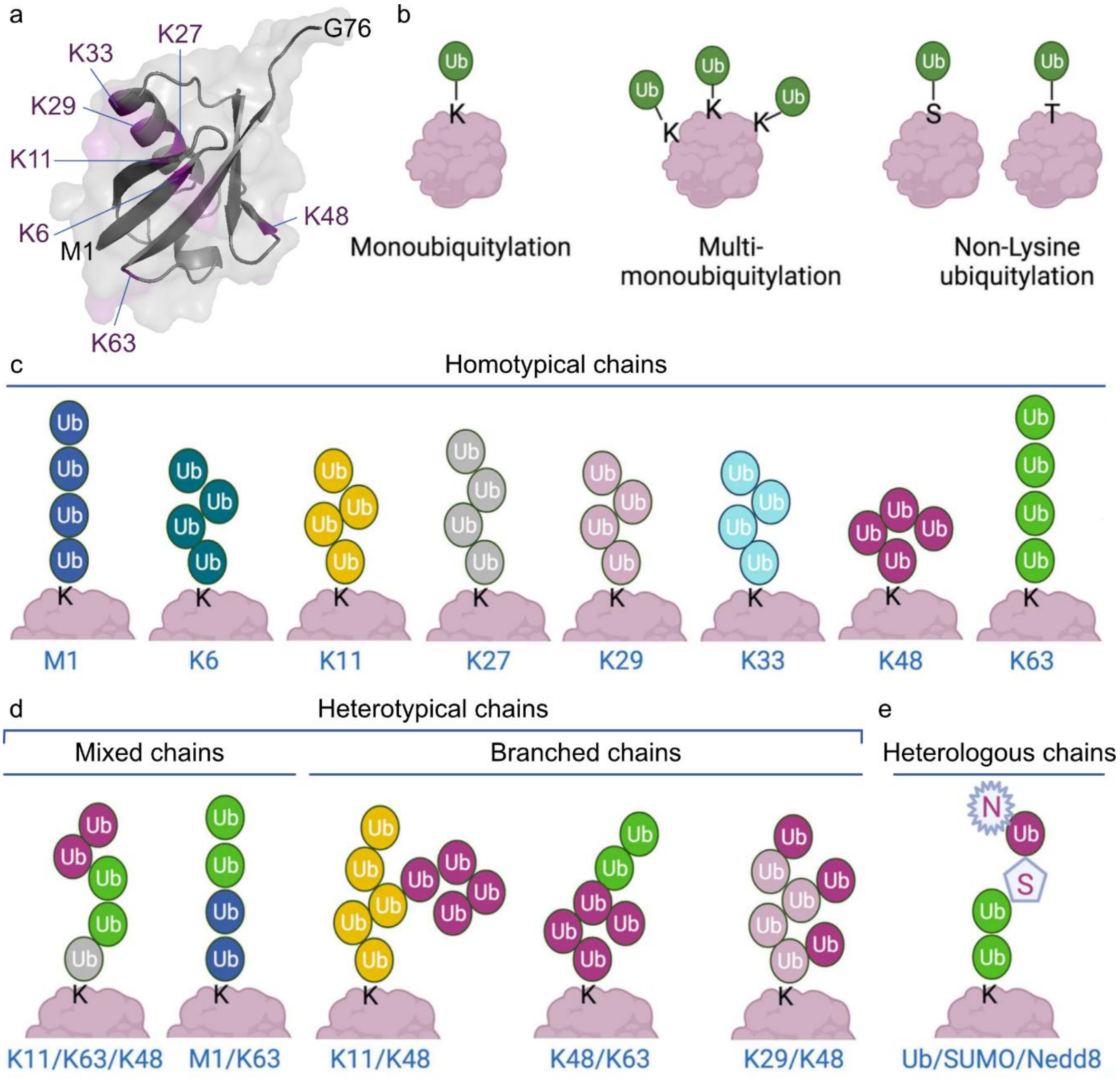


Figure 16. Ubiquitin chain diversity. (a) Structure of ubiquitin (Ub) (PDB: 1ubq), highlighting the seven lysine residues (K6, K11, K27, K29, K33, K48, K63), the *N*-terminal methionine (M1) and the *C*-terminal glycine (G76). (b) For monoubiquitylation, ubiquitin's G76 gets attached to a lysine residue in the target protein. Multi-monoubiquitylation occurs when multiple Ub-molecules are attached independently to different lysine residues on the same target. In rare cases ubiquitylation also occurs on serine or threonine residues. (c-d) Polyubiquitylation involves the sequential addition of Ub-molecules, where each subsequent Ub-molecule is linked to a lysine residue of the preceding one. PolyUb-chains can form on each lysine residue and on the M1 residue, forming homotypic (c) or heterotypical (mixed or branched) chains (d). (e) Combined with other post-translational modifications, like small ubiquitin-related modifier (SUMO) or neural precursor cell expressed developmentally downregulated protein 8 (NEDD), heterologous chains can be formed. Cartoon representatives of different chain architectures are shown. Figure adapted from Gonzalez-Santamarta *et al.*[374].

K48-linked polyubiquitin is the most abundant type and primarily signals for proteasomal degradation[372]. Crystal structures of K48-diUb, revealed a closed conformation, burying ubiquitin's Ile44 hydrophobic patch (Figure 17a-b). However, as the isopeptide linkage is flexible and interactions between the two Ub-molecules are weak, a dynamic equilibrium between the closed and an open conformation exists, allowing access to the important interaction site at the Ile44-patch[416–419]. Such conformational changes can for example be triggered by pH changes[420].

K63-Ub-chains, the second type to be characterised, are mainly involved in non-proteolytic processes. These processes include cell signalling, with the nuclear factor kappa B (NF-κB) pathway, DDR and DDT pathways (see sections 1.4.4-1.4.5), inflammation, autophagy and protein trafficking[374,421–425]. Their extended conformation resembles M1-linked chain (Figure 17c-d)[426–428]. K63 is located very close to the *N*-terminus, which together with the conformational fold of K63-chains led to the assumption that linear (M1-linked) chains could substitute for K63-polyubiquitin[419,428]. However, the *N*-terminal linkage lacks flexibility, differs in the local molecular environment, and is recognised by specific ubiquitin-binding domains (UBDs)[429].

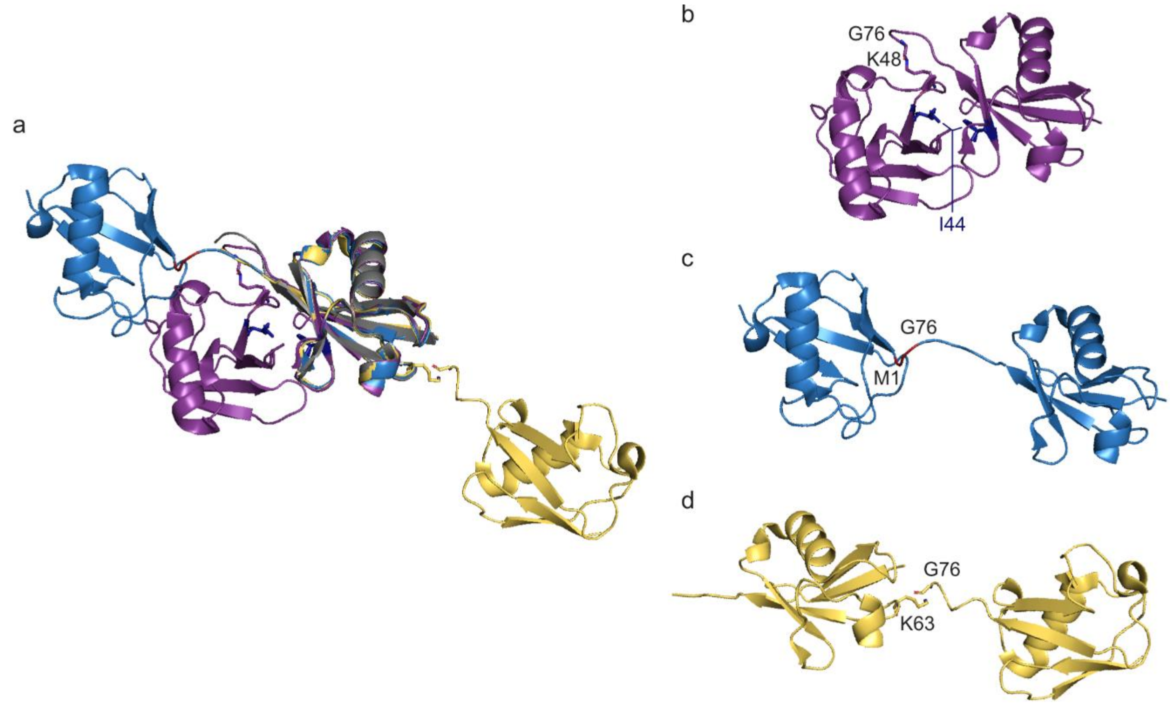


Figure 17. Structural comparison of ubiquitin linkages. (a-d) Structural comparison of monoubiquitin (PDB: 1ubq) in grey, K48-linked diubiquitin (PDB: 3m3j) in purple (b), linear diubiquitin (PDB: 2w9n) in blue (c) and K63-linked diubiquitin (PDB: 2jf5) in yellow (d), highlighting amino acid residues involved in the linkage and the hydrophobic isoleucine (I44) in blue for K48-linked diubiquitin (b). Superimposed structures aligned to monoubiquitin are shown in (a).

Other ubiquitin linkages are less well characterised. M1-linked Ub-chains participate in protein quality control, NF-κB signalling and cell death[430–433]. K11-polyubiquitin functions in cell cycle control, interferon (IFN) signalling and DDR pathways[434–438]. Their structural similarity to K48-chains, also suggested a role in proteasomal degradation. However, other studies stated that homotypic K11-chains do not bind the proteasome sufficiently to induce degradation[374,439]. K6-chains[125,377,440], K27-polyubiquitin and K29-chains, are involved in HR[441]. Moreover, K6-chains were reported in mitophagy[374,436,442] and RNA-protein crosslinks (RPCs) resolution[443,444], while K27-, K29- and K33-linkages contribute to inflammation, autophagy, cell signalling and protein trafficking[374,436]. Structurally, K11- and K27-linked Ub-chains resemble the compact conformation of K48-chains, while K29- and K33-linked chains do not show this conformation. K6-chains also have a closed conformation, however different from the others as proximity to the Ile44-patch may interfere with binding[419,445].

Another layer of complexity is added through heterotypic and branched chains, where different Ub-linkages coexist within the same chain or ubiquitin is modified on multiple lysine residues (Figure 16d)[374,385]. Though still not fully understood, branched chains are estimated to make up 5-20% of all polyubiquitin in the cell[446]. Notable examples include heterotypic K11/K48-chains involved in proteasomal degradation and found especially on cell cycle substrates[447]. K29/K48-branched chains were reported in ER-associated protein degradation (ERAD)[448]. K48/K63-branched and K63/M1-mixed chains were described in NF-κB signalling[449–451] and K48/K63/K11-mixed chains are implicated in Cyclin B1 regulation[452]. Additionally, Ub-chains can be further modified by other PTMs, such as phosphorylation, SUMOylation or NEDDylation, altering chain architecture and function (Figure 16e)[374,453,454].

1.4.3 Ubiquitin interactions

Ubiquitin is primarily recognised via its hydrophobic Ile44-patch, which includes Ile44, Leu8, Val70 and His68 (Figure 18a) and serves as the main interaction site for UBDs and the proteasome[414,455,456]. A second hydrophobic region, the Ile36-patch, involving Ile36, Leu71 and Leu73 (Figure 18a)[414,455], mediates interactions with Ub-chains and is targeted by HECT E3-ligases[457], DUBs[458] and UBDs[459]. The Phe4 patch, formed by Phe4, Gln2 and Thr14 (Figure 18a), is essential for yeast cell division and potentially acts in protein trafficking[460]. Structural differences in this region help DUBs distinguish ubiquitin from its closest homolog NEDD8[461]. Additionally, higher eukaryotes possess a TEK-box within ubiquitin, critical for mitotic degradation. The TEK-box includes Thr12, Thr14, Glu34, Lys6 and Lys11 (Figure 18a)[414,435].

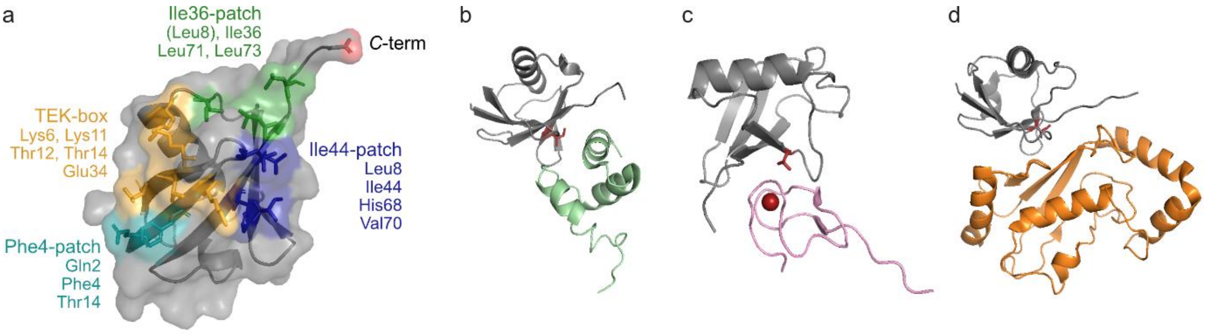


Figure 18. Ubiquitin interaction patches and common ubiquitin-binding domains. (a) Structure of monoubiquitin (PDB: 1ubq) in grey, highlighting its main interaction patches, with the Ile44-patch in blue, the Ile36-patch in green, the Phe4-patch in cyan and the TEK-box in orange. Figure adapted from Komander and Rape[414]. (b) Structure of the ubiquitin-associated (UBA) domain (green) of protein linking IAP with cytoskeleton 1 (PLIC1) bound to ubiquitin (grey) mediated by Ile44, coloured in red (PDB: 2jy6). (c) Structure of the zinc-finger (pink), with the Zn^{2+} ion coloured in red, of nuclear protein localization 4 (Npl4) bound to ubiquitin (grey) mediated by Ile44, coloured in red (PDB: 1q5w). (d) Structure of the ubiquitin conjugating E2-enzyme UBE2D3 in orange bound to ubiquitin (grey) mediated by Ile44, coloured in red (PDB: 2fuh). Figure adapted from Dikic *et al.*[455].

1.4.3.1 Ubiquitin-binding domains

UBDs are structural motifs in proteins that recognise ubiquitin. These domains vary in structure, binding specificity and recognition mechanisms and new ones are constantly identified. The main ones are described here. UBDs most commonly fold into α-helices, zinc fingers (ZnFs), plekstrin homology (PH) folds, ubiquitin-conjugation (UBC)-related domains, Src homology 3 (SH3) structures or WD40 β-propellers[455,462].

Many UBDs rely on α-helical structures that interact mainly with ubiquitin's hydrophobic Ile44-patch. Common single α-helix UBDs include ubiquitin interacting motif (UIM), MIU, inverted ubiquitin interacting motif (IUIM), UIM- and MIU-related (UMI) and double-sided UIM (DUIM)[462]. UIMs use a specific motif (LeuXXAlaLeu) to bind ubiquitin[463,464], whereas MIUs use the same motif in the opposite direction (LeuAlaXXLeu)[465]. DUIMs can even bind two Ub-molecules simultaneously[466]. Multi-helix domains like coupling of ubiquitin conjugation to endoplasmic reticulum (CUE), golgi-localized, gamma-ear-containing, ARF-binding protein 3 (GGA) and target of myb1 (TOM1) (GAT) and UBA domains bundle three helices and often show certain Ub-chain type preferences[462]. UBAs (Figure 18b) vary in selectivity, with some preferring K48-chains[467,468]. CUE domains often bind mono-Ub but also recognise K48-chains[469,470]. GAT domains can engage two Ub-molecules or one Ub-molecule and an additional binding partner[462,471,472]. Ubiquitin recognising proteins, often combine multiple UBDs via flexible linkers to fine-tune affinity and chain type specificity[462]. The proteasomal subunit S5a, for example, contains two UIMs that bind a wide range of linkage types (K48, K63, K6, K11, K29)[473,474]. Rap80, involved in DNA repair, combines multiple UIMs to form a continuous helix which specifically binds to K63-linked chains[421,475,476]. Furthermore, oligomerisation enhances linkage specificity[477].

ZnF domains are the second most abundant UBD fold. They are stabilised by coordinated Zn^{2+} ions and interact with various regions within ubiquitin. Examples are UBZ and nuclear protein localization 4 (Npl4) ZnF (NZF) (Figure 18c), which target ubiquitin's hydrophobic Ile44-patch[462]. NZF are mainly described to recognise K63-linked Ub-chains due to steric constrains, while also possessing a conserved Thr-Phe motif interacting with mono-Ub[421,478–480]. Other ZnFs such as RABEX5 target specific residues within ubiquitin, in this case Asp58 via hydrogen bonds[481,482]. ZnFs are also used in DUBs to recognise the *C*-terminal Gly75-Gly76 tail of ubiquitin and enable cleavage[459,462].

PH domains are typically composed of a 7-stranded β-sheet and a *C*-terminal α-helix and mainly bind phosphoinositides (PIs) within membranes and act in cellular signalling pathways[462]. However, two PH-like domains also bind ubiquitin. Gram-like ubiquitin-binding in Eap45 (GLUE) binds ubiquitin's Ile44-patch via a binding site opposite the PI-binding region[483–485]. Plekstrin-like receptor for ubiquitin (PRU) forms hydrogen bonds with ubiquitin's His68 and forms high-affinity interactions with Ile44[486,487].

UBC domains bear the catalytic Cys of Ub-conjugating E2-enzymes, which mediates thioester bond formation during ubiquitylation (see section 1.4.1). They consist of ~150 residues, which form four stranded α-helices and a 4-stranded antiparallel β-sheet (Figure 18d)[462,488]. While all E2-enzymes covalently bind ubiquitin, UbcH5c was found to also form non-covalent bonds, which play an important role in BRCA1-polyubiquitylation[489]. Ubiquitin-conjugating enzyme E2 variant (UEV) domains resemble UBC folds but lack the catalytically active Cys. They act for instance in human immunodeficiency virus-1 (HIV-1) budding and multi-vesicular body (MVB) sorting [462,490].

SH3 domains adopt a β-barrel fold and typically mediate protein-protein interactions by binding proline-rich motifs (PxxP)[491]. Some SH3 domains also bind ubiquitin, for example Sla1 in yeast and its mammalian homolog CIN85. For Sla1, ubiquitin's hydrophobic Ile44-patch is predicted to compete with PxxP motifs for the first SH3 motif[492,493]. Contrary to CIN85 where all three SH3 domains are thought to bind ubiquitin[462,494].

WD40 β-propellers present for example in SKP1, CUL1, F-box protein (SCF) E3-ligase and the Cdc48/p97 adaptor Doa/Ufd3, recognise ubiquitin's hydrophobic Ile44-patch via loop structures[462,495].

1.4.4 Ubiquitylation in DNA damage response

With their broad and diverse functionality, ubiquitin and ubiquitylation are also essential regulators of DNA damage repair and tolerance pathways.

In NER, many proteins are polyubiquitylated, regulating their stability and protein-protein interactions[496,497]. In GG-NER, DDB2 and XPC are polyubiquitylated by the DDB-CUL4A-RBX1 E3-ligase complex (CRL4A^{DDB2}). DDB2-ubiquitylation leads to its dissociation from the lesion and subsequent proteasomal degradation, while ubiquitylation of XPC enhances its DNA-binding[498]. However, regulation of DDB2 is complex, as a later study suggested stabilisation at the damage site, when XPC prohibited ubiquitylation of DDB2, after UVC-irradiation. Persistent DDB2 ubiquitylation initiated and ensured efficient repair in these scenarios[499]. Both DDB2 and XPC are extracted from chromatin by the p97 segregase complex and delayed removal reduces repair efficiency and increases genotoxicity[500]. DDB2 is likely modified with K48-linked chains, since it gets marked for proteasomal degradation[497], whereas the precise nature of XPC ubiquitylation remains unclear[501]. XPC is also subject to SUMOylation, and these SUMO2-chains are further modified with K63-linked Ub-chains by UBC13-MMS2 (E2) and RNF111 (E3), promoting its release from DNA and facilitating downstream factor recruitment[502,503]. In TC-NER, stalled RNA Pol II recruits CSA, CSB and UVSSA[95,504]. CSA forms a DDB1-CUL4-RBX1 E3-ligase complex (CRL4^{CSA})[505], like DDB2, which ubiquitylates CSB, marking it for p97-mediated extraction and proteasomal degradation[497]. Persistent CSB increases RNA synthesis recovery but impairs cell survival post UV-irradiation, highlighting the need for timely clearance[506]. RNA Pol II itself is also ubiquitylated and degraded, a process with debated roles in repair pathway regulation[507]. Recent studies indicate that modification of RPB1 (Pol II subunit) at K1268 by CRL4^{CSA}, is crucial for the recruitment of TFIIH and UVSSA ubiquitylation, preventing Cockayne syndrome phenotypes *in vivo*[94,508].

During DSB repair, histone ubiquitylation is a pivotal step, that alters chromatin structures and guides pathway choice between HR or NHEJ[509]. The serine/threonine-protein kinase ATM activates DSB repair by phosphorylating H2AX[109], allowing mediator of DNA damage checkpoint protein 1 (MDC1) to bind and upon phosphorylation by ATM recruit the E3-ligases RNF8 and RNF168[510–513]. RNF8 ubiquitylates histone H1 with K63-linked chains, enabling RNF168 to ubiquitylate H2A at K13 and K15. These modifications recruit BRCA1 and 53BP1, which compete to direct repair towards HR and NHEJ, respectively[514–516]. RNF168 also catalyses H2AK15 monoubiquitylation and H4K20 dimethylation, modifications recognised by 53BP1, promoting NHEJ via complex formation (53BP-RIF1-REV7, 53BP1-PTIP-Artemis) and blocking of BRCA1[517,518]. In addition, APC/C^{Cdh1}, an E3-ligase formed by anaphase-promoting complex/cyclosome (APC/C) and its cofactor Cdh1, limits HR by marking CtIP for degradation post-mitosis and in late S/G2 phase[519]. BRCA1 can initiate HR via RAP80, which binds K63-linked Ub-chains on chromatin, formed by RNF8[34,440,520] or through BARD1, which directly recognises monoubiquitylated H2AK15 via its BRCT-domain-associated ubiquitin-dependent recruitment motif (BUDR), leading to BRCA1 recruitment[521,522]. In S/G2 phase, BRCA1 promotes 53BP1 dephosphorylation via PP4C/PP4R2, evicting it from DNA to allow end resection and subsequent HR steps[523]. BRCA1 also facilitates CtIP localisation to chromatin through ubiquitylation[524]. Additionally, BRCA1/BARD1 further ubiquitylate H2A at K125, K127 and K129, recruiting SMARCAD1, which removes 53BP1-binding marks from histones. This leads to repositioning of 53BP1 and completes resection[525–527].

In the FA pathway, monoubiquitylation of FANCD2 at K561 and FANCI at K523 is central for retention on chromatin and initiation of ICL repair. The modification is catalysed by UBA1 (E1), FANCT/UBE2T (E2) and FANCL (E3)[528,529] and anchors the ID2 complex at the lesion site[137]. Interestingly, most FA patients carry mutations that disrupt FANCD2 monoubiquitylation. In cells, lack of ID2 ubiquitylation leads to hypersensitivity towards ICL-inducing agents[530,531]. Structural studies showed that FANCD2 alone forms an inactive homodimer, while the FANCD2/FANCI-heterodimer binds DNA and monoubiquitylation locks the complex around DNA to initiate repair[532,533]. FANCD2 ubiquitylation enhances DNA-binding, while FANCI modification prevents ID2 deubiquitylation by USP1-UAF1, prolonging its DNA association[534,535]. Modification of each subunit promotes modification of the other subunit, coordinating ID2's DNA retention and repair initiation[536].

In DDT pathways PCNA ubiquitylation is important for pathway choice[537]. Upon replication stress, PCNA is monoubiquitylated at K164 by UBA1 (E1), RAD6 (E2) and RAD18 (E3)[170], a process facilitated by RAD6-RAD18 recruitment to stalled forks via RPA[538]. Notably, also other E3-ligases (such as RNF4 and CRL4^{Cdt2}) can monoubiquitylate PCNA, though less efficiently[539–541]. Monoubiquitylated PCNA shifts binding from high-fidelity polymerases (Polδ and Polε) to low-fidelity TLS polymerases (Polη, Polι, Polκ, Polλ, Polζ and REV1), enabling lesion bypass (see section 1.1.3). TLS polymerases recognise PCNA through various domains, including UBMs, PIP-boxes, BRCT domains and/or polymerase-associated domain (PAD)[165]. The Ub-moiety at PCNA's K164 can be extended to a K63-linked Ub-chain by Mms2-Ubc13-Rad5 (in yeast), promoting error-free TS and fork reversal[542]. Mms2-Ubc13 are conserved in mammals and two orthologs of Rad5 are expressed: HLTF and SHPRH. While HLTF and SHPRH were reported to be involved in PCNA polyubiquitylation[176,543], MEFs lacking both factors showed no sensitivity towards DNA-damaging agents[544]. Polyubiquitylated PCNA is recognised by ZRANB3 via its NZF-motif, its PIP-box and AlkB2 PCNA-interaction motif (APIM)[178,545]. ZRANB3 can, on the one hand, promote fork reversal with its annealing helicase activity, slowing down replication forks and promoting replication restart. On the other hand, ZRANB3 has strand-specific endonuclease activity with which it can cleave intermediately formed D-loops to limit aberrant recombination[165,178,546].

1.4.5 Ubiquitylation in DPC repair

During replication-coupled DPC repair, uncoupling of the helicase and polymerase triggers DPC ubiquitylation. Two distinct E3-ligases get recruited by this uncoupling. First, TRAF interacting protein (TRAIP), travelling with the CMG helicase, ubiquitylates DPCs as CMG bypasses them[306]. Secondly, ring finger and WD repeat domain 3 (RFWD3), gets activated by RPA-bound ssDNA formed near stalled replication forks. RFWD3 further extends Ub-chains on DPCs, marking them for proteasomal degradation (Figure 11)[349,350].

In global-genome DPC repair, DPC SUMOylation, catalysed by SAE1/UBA2 (E1), UBC9 (E2)[547] and PIAS4 (E3) precedes ubiquitylation[291,353,354]. SUMO-chains on DPCs are recognised by two STUbLs, RNF4 and TOP1 binding arginine/serine rich protein (TOPORS), acting to a certain extent redundantly (Figure 12). While loss of both E3-ligases has detrimental effects on cell fitness, lack of RNF4 alone only compromises degradation of DNMT1-DPCs[325,353,354,548,549]. E3-ligase choice may depend on chromatin context, although persistent DPCs will eventually be ubiquitylated by either enzyme[548]. As in replication-coupled DPC repair, ubiquitylation triggers proteolytic degradation by the proteasome or SPRTN[325,353,354].

Like in TC-NER, during transcription-coupled DPC repair RPB1 gets ubiquitylated at K1268 by CRL4^{CSA} (Figure 13)[94,508]. However, in contrast to TC-NER, this modification is not essential for DPC repair, as cells lacking RPB1 ubiquitylation enzymes or expressing a non-ubiquitylatable RBP1^{K1268R} mutant, show only mild repair defects. Alternatively, CRL4^{CSA} might directly polyubiquitylate the crosslinked protein in this pathway, marking it for proteasomal degradation (Figure 13) [357–359].

2 Aim of this study

DNA-protein crosslinks (DPCs) are highly toxic DNA lesions that pose a severe threat to genome integrity, particularly during replication. DPCs can arise from endogenous metabolic processes or can be induced by chemotherapeutic agents. Understanding how DPCs form and how they are resolved is essential for understanding cellular repair mechanisms and improving therapeutic approaches relying on DPC induction. This study explores how cells regulate DPC formation and resolution, with special emphasis on DPC detection and DPC repair pathways. Therefore, I focused on physiological DPCs formed by HMCES, techniques used for the detection and analysis of DPCs and the main human DPC protease SPRTN.

First, I investigated the role of the conserved protein HMCES, which forms covalent crosslinks with abasic sites in ssDNA to prevent strand breaks during replication. It is known that HMCES-DPCs are targeted by proteases during repair pathways. In this study, I focused on an additional resolution mechanism, proposing that HMCES-DPCs are reversal. I used biochemical reconstitution experiments to investigate this release mechanism and understand how it balances crosslink release with the need to protect abasic sites.

Secondly, I addressed a constant challenge in DPC research, by developing an in-detail protocol for a new technique called purification of x-linked proteins (PxP), enabling reliable isolation and analysis of DPCs from mammalian cells.

Finally, I investigated DPC repair by the metalloprotease SPRTN, focusing on how its activity is regulated and restricted to cleavage of crosslinked proteins, particularly in relation to ubiquitin. By combining biochemical and structural analysis, I investigated effects of DPC ubiquitylation on SPRTN activity and involved domains. Furthermore, I tested the relevance of these interactions in context of Ruijs-Aalfs syndrome.

Overall, this study aims to improve the understanding of both non-proteolytic and proteolytic DPC resolution and to introduce a robust new technique for DPC identification, supporting future research and therapeutic developments.

3 Publications

3.1 A non-proteolytic release mechanism for HMCES-DNA-protein crosslinks

Contribution report

This publication confirms and explores an auto-release mechanism for HMCES-DPCs, depending on DNA context. *In vitro* experiments shown in the publication were done by Maximilian Donsbach and me, with help from Denitsa Yaneva and Florian Grünert. Together with Pedro Weickert, I generated HMCES overexpression cell lines. *X. laevis* experiments were performed by Kha T. Nguyen and Richa Nigam. Writing and editing of the manuscript as well as figure preparation were done by Maximilian Donsbach, Julian Stingele and me with input from all authors.

Summary

The conserved protein HMCES crosslinks to abasic (AP) sites in ssDNA to prevent strand scission and the formation of toxic dsDNA breaks during replication. Here, we report a non-proteolytic release mechanism for HMCES-DNA-protein crosslinks (DPCs), which is regulated by DNA context. In ssDNA and at ssDNA-dsDNA junctions, HMCES-DPCs are stable, which efficiently protects AP sites against spontaneous incisions or cleavage by APE1 endonuclease. In contrast, HMCES-DPCs are released in dsDNA, allowing APE1 to initiate downstream repair. Mechanistically, we show that release is governed by two components. First, a conserved glutamate residue, within HMCES' active site, catalyses reversal of the crosslink. Second, affinity to the underlying DNA structure determines whether HMCES re-crosslinks or dissociates. Our study reveals that the protective role of HMCES-DPCs involves their controlled release upon bypass by replication forks, which restricts DPC formation to a necessary minimum.

Article

SOURCE DATA
TRANSPARENT PROCESS
OPEN ACCESS

THE EMBO JOURNAL

A non-proteolytic release mechanism for HMCES-DNA-protein crosslinks

Maximilian Donsbach[1,2,†], Sophie Dürauer[1,2,†], Florian Grünert[1,2], Kha T Nguyen[3], Richa Nigam[3], Denitsa Yaneva[1,2], Pedro Weickert[1,2], Rachel Bezalel-Buch[4], Daniel R Semlow[3] & Julian Stingele[1,2,*]

Abstract

The conserved protein HMCES crosslinks to abasic (AP) sites in ssDNA to prevent strand scission and the formation of toxic dsDNA breaks during replication. Here, we report a non-proteolytic release mechanism for HMCES-DNA-protein crosslinks (DPCs), which is regulated by DNA context. In ssDNA and at ssDNA-dsDNA junctions, HMCES-DPCs are stable, which efficiently protects AP sites against spontaneous incisions or cleavage by APE1 endonuclease. In contrast, HMCES-DPCs are released in dsDNA, allowing APE1 to initiate downstream repair. Mechanistically, we show that release is governed by two components. First, a conserved glutamate residue, within HMCES' active site, catalyses reversal of the crosslink. Second, affinity to the underlying DNA structure determines whether HMCES re-crosslinks or dissociates. Our study reveals that the protective role of HMCES-DPCs involves their controlled release upon bypass by replication forks, which restricts DPC formation to a necessary minimum.

Keywords abasic (AP) sites; DNA repair; DNA replication; DNA-protein crosslinks; HMCES
Subject Categories DNA Replication, Recombination & Repair; Post-translational Modifications & Proteolysis
DOI 10.15252/embj.2022113360 | Received 21 December 2022 | Revised 17 July 2023 | Accepted 19 July 2023 | Published online 31 July 2023
The EMBO Journal (2023) 42: e113360

Introduction

Covalent crosslinks between proteins and DNA (DNA-protein crosslinks, DPCs) are dangerous lesions caused by a variety of endogenous and exogenous sources, including widely used chemotherapeutic agents (Tretyakova *et al*, 2015; Stingele *et al*, 2017). DPCs are toxic because they interfere with DNA replication (Duxin *et al*, 2014). Therefore, cells possess conserved repair mechanisms that target DPCs in replication-dependent and -independent manners (Weickert & Stingele, 2022). DPC repair involves the proteolytic destruction of the protein adduct by DPC-specific proteases of the SPRTN/Wss1 family or by proteasomal degradation (Stingele *et al*, 2014, 2016; Vaz *et al*, 2016; Larsen *et al*, 2019). Failure to degrade DPCs has drastic consequences; complete loss of SPRTN is lethal in mammalian cells, while partial loss-of-function results in premature aging and predisposition to liver cancer (Lessel *et al*, 2014; Maskey *et al*, 2014, 2017). Despite the severe phenotypes associated with the absence of SPRTN alone, several additional proteases appear to target DPCs (Borgermann *et al*, 2019; Bhargava *et al*, 2020; Dokshin *et al*, 2020; Kojima *et al*, 2020; Serbyn *et al*, 2020). The diversity of repair mechanisms underlines the threat posed by DPCs. However, some DPCs have important physiological roles. The human protein HMCES forms crosslinks with abasic (AP) sites to protect genome integrity (Mohni *et al*, 2019).

AP sites are frequent endogenous DNA lesions, which arise spontaneously or enzymatically during base excision repair and active DNA demethylation (Thompson & Cortez, 2020). AP sites exist in equilibrium between a closed-ring furanose and an open-ring aldehyde form. The latter is prone to undergo spontaneous β-elimination, resulting in strand scission and DNA single-strand break (SSB) formation, which can also arise enzymatically upon AP site cleavage by AP endonucleases and lyases (Krokan & Bjoras, 2013; Amidon & Eichman, 2020). If such SSBs form in double-stranded DNA (dsDNA), they are swiftly repaired by the cellular SSB repair machinery (Abbotts & Wilson, 2017). In contrast, incision of AP sites in ssDNA, for example, at the replication fork, will result in the formation of toxic DSBs (Mehta *et al*, 2020; Semlow *et al*, 2022). To prevent such a catastrophic scenario, the conserved catalytic SOS response-associated peptidase (SRAP) domain of HMCES (Fig 1A) associates with replication forks to crosslink to AP sites in ssDNA (Mohni *et al*, 2019). Crosslinking occurs between the N-terminal cysteine residue of the SRAP domain (methionine is proteolytically removed) and an AP site, resulting in the formation of a thiazolidine

---

1 Department of Biochemistry, Ludwig-Maximilians-University Munich, Munich, Germany
2 Gene Center, Ludwig-Maximilians-University Munich, Munich, Germany
3 Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA, USA
4 Department of Biological Chemistry and Molecular Biophysics, Washington University School of Medical, Saint Louis, MO, USA
*Corresponding author. Tel: +49 89 2180 71101; E-mail: stingele@genzentrum.lmu.de
†These authors contributed equally to this work

ring, which prohibits strand scission (Fig 1B and C; Halabelian *et al*, 2019; Thompson *et al*, 2019; Wang *et al*, 2019). DPC formation has been suggested to be initiated by the N-terminal amino group attacking the AP sites' open-ring aldehyde form (Halabelian *et al*, 2019; Thompson *et al*, 2019; Wang *et al*, 2019). The resulting Schiff-base intermediate is then converted into a thiazolidine ring upon reaction with the sulfhydryl group of Cys2 (Fig 1B).

The protective function of HMCES-DPCs is particularly important, when cells face substantial amounts of AP sites, for example, upon exposure to genotoxic agents (Srivastava *et al*, 2020), overexpression of the cytosine deaminase APOBEC3A (Mehta *et al*, 2020; Biayna *et al*, 2021), or AID-induced somatic hypermutation (Wu *et al*, 2022). In addition, HMCES-DPCs were shown in *Xenopus* egg extracts to arise as intermediates of replication-coupled DNA interstrand crosslink (ICL) repair (Semlow *et al*, 2022). Unhooking of an AP site-induced ICL (AP-ICL) by a DNA glycosylase yields an AP site, to which HMCES crosslinks. In egg extracts, HMCES-DPCs appear to be mainly degraded by the SPRTN protease (Semlow *et al*, 2022), which requires unfolding of the protein adduct by the FANCJ helicase (Yaneva *et al*, 2023). While human SPRTN cleaves HMCES-DPCs *in vitro* if FANCJ is present (Yaneva *et al*, 2023), it is unclear to what extent SPRTN is required for repair in mammalian cells, where proteasomal HMCES-DPC degradation has been reported (Mohni *et al*, 2019). Notably, HMCES-DPCs are slowly lost over time in egg extracts when SPRTN is depleted and proteasomal activity inhibited (Semlow *et al*, 2022), suggesting that additional mechanisms may resolve HMCES-DPCs. Recent work indicated that SRAP-DPCs can undergo reversal in principle (Paulin *et al*, 2022), but it remained unclear how reversal can be reconciled with the need to protect AP sites in ssDNA.

Here, we use *in vitro* reconstitution to dissect the principles of a non-proteolytic release mechanism for HMCES-DPCs. We demonstrate that DPC release is determined by DNA context and occurs in two steps. First, a conserved glutamate residue located in HMCES' active site catalyses the reversal of the thiazolidine crosslink. Second, HMCES either re-crosslinks, if affinity to the underlying DNA structure is high, or releases the AP site, if affinity is low. As a consequence, HMCES efficiently protects AP sites in ssDNA and at ssDNA-dsDNA junctions but releases them once the DPC is bypassed by the replication machinery and transferred into dsDNA.

Results

HMCES-DNA-protein crosslinks are reversible

Once HMCES-DPCs form, they appear stable over several days at room temperature *in vitro* (Thompson *et al*, 2019). To test whether HMCES remains irreversibly attached during incubation or constantly cycles between a crosslinked and a non-crosslinked state, we designed an assay to assess the reversibility of HMCES-DPCs (Fig 1D, schematic). First, we generated AP sites by incubating a Cy5-labelled 30mer DNA oligonucleotide containing a deoxyuridine (dU) at position 15 with uracil-DNA glycosylase (UDG). DNA containing dT instead of dU served as a control. DPCs were then generated by addition of recombinant full-length HMCES (HMCES^{FL}) or the catalytic SRAP domain (HMCES^{SRAP}). Next, reactions were exposed to a short heat treatment (5 min, 60°C), which inactivates free HMCES while not affecting crosslinked HMCES (Yaneva *et al*, 2023). Finally, a 6-FAM-labelled AP site-containing DNA oligonucleotide was added to all reactions to test whether HMCES can be released from the Cy5-oligonucleotide and re-crosslink to the 6-FAM-oligonucleotide. Indeed, we observed formation of DPCs between 6-FAM-labelled DNA and HMCES^{FL} and HMCES^{SRAP} (Figs 1E, lanes 6 and 7, F, and EV1A), suggesting that some DPCs between HMCES and the Cy5-oligonucleotide reverted which in turn allowed re-crosslinking to the 6-FAM-oligonucleotide. 6-FAM-DPCs did not form if a Cy5-dT-oligonucleotide was used (Fig 1E, lanes 2 and 3), indicating that inactivation of free HMCES was efficient, or if HMCES' catalytic cysteine was replaced by serine (HMCES^{SRAP}-C2S) (Fig 1E, lane 8).

Next, we asked whether HMCES-DPC reversal occurs spontaneously or whether it is an enzymatic process. The active site of HMCES features, in addition to the catalytic cysteine at position 2, two highly conserved amino acid residues, Glu127 and His210 (Fig 1A and C). Structural data suggest that both residues stabilize the transient Schiff-base intermediate during DPC formation (Halabelian *et al*, 2019; Thompson *et al*, 2019; Wang *et al*, 2019). Nonetheless, substitution of the corresponding glutamate residue in the prokaryotic HMCES orthologue YedK results in only reduced DPC formation (Thompson *et al*, 2019; Wang *et al*, 2019), while the effect of substituting the histidine remains controversial with reports of decreased and increased DPC formation (Thompson *et al*, 2019; Wang *et al*, 2019). Consistently, we observed that human HMCES^{SRAP} with substitution of Glu127 (E127A) or His210 (H210A) were able to form DPCs with Cy5-labelled AP site-containing DNA in our assay (Figs 1G, lanes 7 and 8, Cy5 scan, H, EV1B and C). However, HMCES^{SRAP}-E127A and -H210A variants did not form DPCs with the subsequently added 6-FAM-oligonucleotide (Fig 1G, lanes 7 and 8, 6-FAM scan, and H). These results suggest that stabilization of the Schiff-base intermediate by Glu127 and His210 is not essential for DPC formation *per se* but may rather be important to reverse thiazolidine ring formation, perhaps explaining the strict conservation of both residues during evolution. In agreement with a recent study (Paulin *et al*, 2022), we conclude that HMCES-DPCs are reversible, that released HMCES retains the ability to re-crosslink, and that release is an enzymatic process requiring conserved active-site residues.

Release of HMCES-DPCs is determined by DNA context

The fact that HMCES-DPCs are reversible raises the question of whether the release is regulated. AP sites must be protected in ssDNA to prohibit strand breakage, but HMCES-DPC formation may be less favourable in dsDNA, where it would prohibit initiation of AP site repair by AP endonucleases. In line with the need to stabilize AP sites in ssDNA, DPC formation by HMCES^{SRAP}-WT occurs efficiently in ssDNA and at ssDNA-dsDNA junctions with a 5′-flap (Figs 2A and B) (Mohni *et al*, 2019; Thompson *et al*, 2019). In contrast, DPC formation does not occur in dsDNA (Fig 2A and B) (Mohni *et al*, 2019). It has been speculated that this is due to HMCES not being able to accommodate AP sites in its active site if dsDNA is present on the 3′-site of the lesion (Thompson *et al*, 2019). Interestingly, however, HMCES-DPC formation occurred efficiently, when the DNA strand across the AP site contained a nick or a 4-nucleotide gap (Fig 2C and D). This indicates that HMCES does not necessarily require long stretches of ssDNA to form a DPC, but rather relies on the bendability of the substrate DNA induced by a nick or gap.

Substitution of Glu127 and His210 had no effect on the specificity of DPC formation, but HMCES^{SRAP}-E127A crosslinked slower (Fig 2A and B), which may be related to a recently proposed role for Glu127 in AP site ring opening (Paulin *et al*, 2022). To understand whether DNA context also influences DPC release, we first generated DPCs between HMCES^{SRAP} and an AP site in ssDNA before annealing complementary reverse oligonucleotides to generate either a ssDNA-dsDNA junction or fully dsDNA (Fig 3A, schematic). Strikingly, HMCES^{SRAP}-DPCs were stable in ssDNA or ssDNA-dsDNA junctions but reversed in dsDNA (Fig 3A and B). In contrast, HMCES^{SRAP} variants H210A and E127A were partially (H210A) or entirely (E127A) defective for reversal in dsDNA (Fig 3A

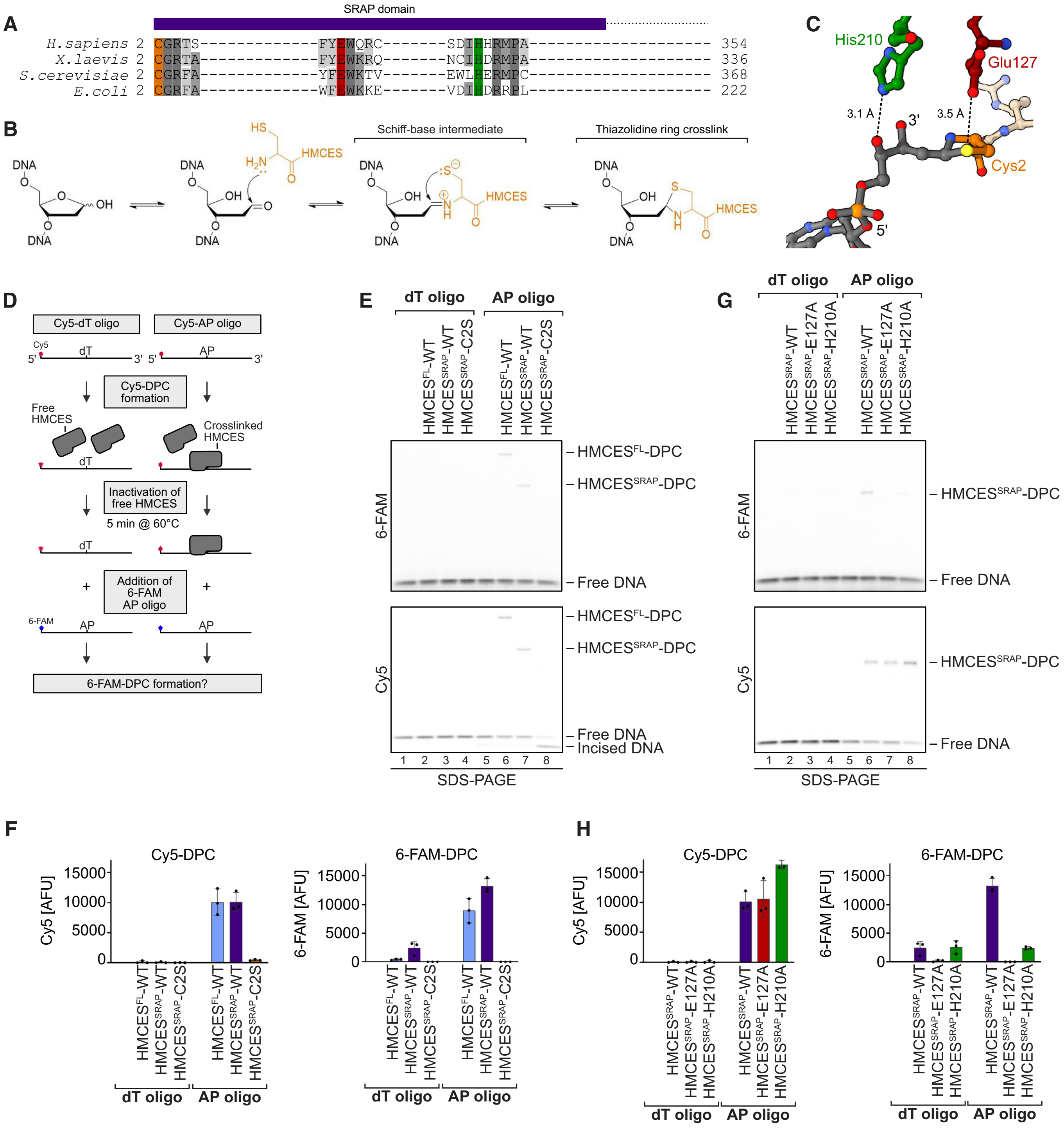


Figure 1.

◀

Figure 1. HMCES-DNA-protein crosslinks are reversible.

A SRAP domain sequence alignment highlighting key active site residues in *H. sapiens*, *X. laevis*, *S. cerevisiae* and *E. coli* HMCES homologues (Cys2 = orange, Glu127 = red and His210 = green).

B Proposed reaction mechanism of SRAP domain crosslinking to an AP site.

C Crystal structure of HMCES' active site crosslinked to an AP site. PDB: 6OE7 (Halabelian *et al*, 2019). DNA is shown in grey. Active site residues are coloured as in (A). Interatomic distances (Å) are labelled.

D Schematic of the assay shown in (E) and (G). HMCES^{FL} and HMCES^{SRAP} (WT or active site variants) were incubated for 1 h at 37°C with a Cy5-labelled 30mer oligonucleotide containing either a dT or an AP site at position 15. Afterwards, non-crosslinked HMCES was inactivated by heat denaturation at 60°C for 5 min. A second 6-FAM-labelled 30mer oligonucleotide containing an AP site was added and formation of 6-FAM DPCs was assessed after an additional incubation for 120 min.

E HMCES^{FL}- and HMCES^{SRAP}-WT and HMCES^{SRAP}-C2S-DPC formation with Cy5- and 6-FAM-oligonucleotides was analysed using denaturing SDS–PAGE. Incised DNA is caused by spontaneous hydrolysis of the AP site.

F Quantification of DPC formation assays shown in (E), left panel: DPC formation to Cy5 oligonucleotide, right panel: DPC formation to 6-FAM oligonucleotide.

G DPC formation of HMCES^{SRAP}-WT and variants (E127A or H210A) with Cy5- and 6-FAM-oligonucleotides was analysed using denaturing SDS–PAGE. Incised DNA is caused by spontaneous hydrolysis of the AP site.

H Quantification of DPC formation assays shown in (G), left panel: DPC formation to Cy5 oligonucleotide, right panel: DPC formation to 6-FAM oligonucleotide.

Data information: Bar graphs in (F) and (H) show the mean of three independent experiments ± SD. Two WT data points are common between (F) and (H). Source data are available online for this figure.

and B). HMCES^{SRAP}-E127A was previously reported to display increased DNA binding (Wang *et al*, 2019), which we confirmed using fluorescence polarization (Fig EV2A) and electrophoretic mobility shift assays (Fig EV2B). To test whether the inability of this variant to reverse is related to increased DNA binding, we combined E127A with a substitution of Arg98 (R98E), which is located within the HMCES^{SRAP}-ssDNA interface (Halabelian *et al*, 2019). DPC formation and release were not affected by the R98E substitution (Figs 2A and B, 3A and B, and EV1B and C), despite severely reduced DNA-binding activity (Fig EV2A and B) (Mohni *et al*, 2019). In combination with E127A, substitution of Arg98 decreased DNA binding below WT levels (Fig EV2A and B) but did not restore the ability to revert the crosslink (Fig EV2C and D). Thus, the reversal defect of HMCES^{SRAP}-E127A-DPCs is unrelated to increased DNA-binding affinity. We conclude that DPC release is not only an active process requiring Glu127 (and partially His210) but is also determined by DNA context. DPC release displays opposite specificity to DPC formation, which correlates with the biological need to protect AP sites in ssDNA but not in dsDNA.

Release of HMCES-DPCs is determined by binding affinity to the underlying DNA

Next, we wanted to understand how DNA context controls the release of HMCES^{SRAP}-DPCs. Our results so far could be explained by a model in which all HMCES-DPCs constantly revert independent of DNA context and that specificity is only determined by HMCES' ability to reform the crosslink after release, which does not occur in dsDNA. However, it remained unclear how HMCES could efficiently protect AP sites, if it would constantly dissociate from the lesion. To gain more detailed insights into DPC reversal in different DNA structures, we first generated HMCES^{SRAP}-DPCs in ssDNA and at ssDNA-dsDNA junctions (DPCs in dsDNA released too quickly to be assessed by this assay). We then added HMCES^{FL} in 10-fold excess to outcompete HMCES^{SRAP} upon release of the AP site (Fig 4A and B, schematic). This set-up allowed us to evaluate release of HMCES^{SRAP}-DPCs by monitoring the appearance of HMCES^{FL}-DPCs over time. Notably, HMCES^{SRAP}-DPCs were released over time in ssDNA (Fig 4A, lanes 3–8) but were much more stable at ssDNA-dsDNA junctions (Fig 4B, lanes 3–8). DPCs formed by the E127A variant were not released in either setting (Fig 4A, lanes 15–20 and B, lanes 15–20). We wondered whether the enhanced release of WT-DPCs from ssDNA was related to the previously reported preferential binding of the SRAP domain to ssDNA-dsDNA junctions compared to ssDNA (Thompson *et al*, 2019). Accordingly, we hypothesized that the active site of HMCES may constantly cycle between a crosslinked and a non-crosslinked state independent of DNA context, but that actual dissociation from the underlying DNA substrate would in addition be determined by binding affinity. To test this idea, we asked whether the reduced DNA-binding affinity of HMCES^{SRAP}-R98E would affect reversal. Indeed, HMCES^{SRAP}-R98E-DPCs reversed much more rapidly in ssDNA and at ssDNA-dsDNA junctions than WT-DPCs (Fig 4A, lanes 9–14 and B, lanes 9–14). Taken together, these results suggest that HMCES-DPC release is governed by two major components. First, the principal capacity of Glu127 to catalyse reversal ensures cycling of the active site between a crosslinked and a non-crosslinked state. Second, the binding strength to the underlying DNA structure then determines whether HMCES re-crosslinks or dissociates while in the non-crosslinked state, which occurs if affinity is low (e.g., within dsDNA or in the context of R98E-DPCs).

Release of HMCES-DPCs restricts crosslink formation to physiologically relevant situations

Next, we wanted to understand how DPC release relates to HMCES' ability to block APE1 endonuclease from incising AP sites. APE1 efficiently cleaves AP sites at ssDNA-dsDNA junctions and in dsDNA but shows little activity in ssDNA (Fig EV3A; Wilson *et al*, 1995). Therefore, we generated HMCES^{SRAP}-WT, -R98E and -E127A-DPCs at ssDNA-dsDNA junctions and in dsDNA and incubated them with APE1. As reported previously (Mohni *et al*, 2019), WT-DPCs shielded AP sites from APE1 incision at ssDNA-dsDNA junctions (Fig 5A, lanes 8–10). In contrast, R98E-DPCs failed to protect against APE1 (Fig 5A, lanes 14–16), suggesting that the increased release of this variant compromises its ability to protect AP sites against APE1 incision. In dsDNA, both WT- and R98E-DPCs did not prevent AP site cleavage (Fig 5B, lanes 8–10 and 14–16, respectively), while

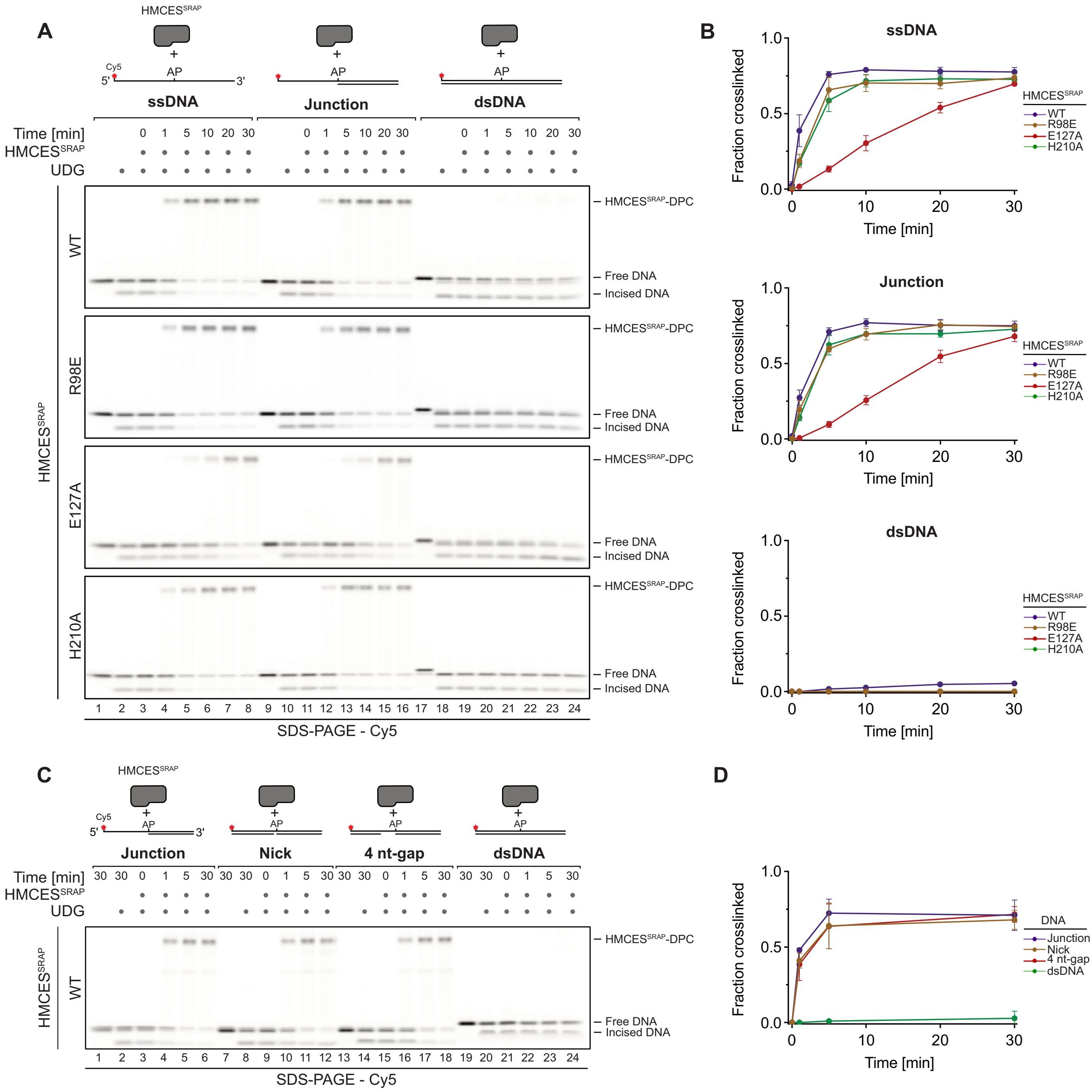


Figure 2. Formation of HMCES-DPCs is determined by DNA context.

A Kinetics of DPC formation by HMCES^{SRAP} (WT, R98E, E127A or H210A variants) to ssDNA, junction DNA and dsDNA. Corresponding reverse oligonucleotides were annealed to ssDNA to create DNA junction and dsDNA prior to adding HMCES^{SRAP}. To ssDNA, a non-complementary oligonucleotide was added as control. HMCES^{SRAP}-WT and variants were incubated with different DNA structures for the indicated amount of time at 37°C prior to separation by denaturing SDS–PAGE.

B Quantification of DPC formation assays shown in (A)

C Kinetics of DPC formation by HMCES^{SRAP}-WT to junction DNA and dsDNA containing a nick or a 4 nt-gap. Corresponding reverse oligonucleotides were annealed to create junction DNA and dsDNA containing a nick or a 4 nt-gap prior to adding HMCES^{SRAP}-WT. HMCES^{SRAP}-WT was incubated with the indicated DNA structures for the indicated amount of time at 37°C prior to separation by denaturing SDS–PAGE.

D Quantification of DPC formation assays shown in (C).

Data information: Data in (B) and (D) represent the mean of three independent experiments ± SD.
Source data are available online for this figure.

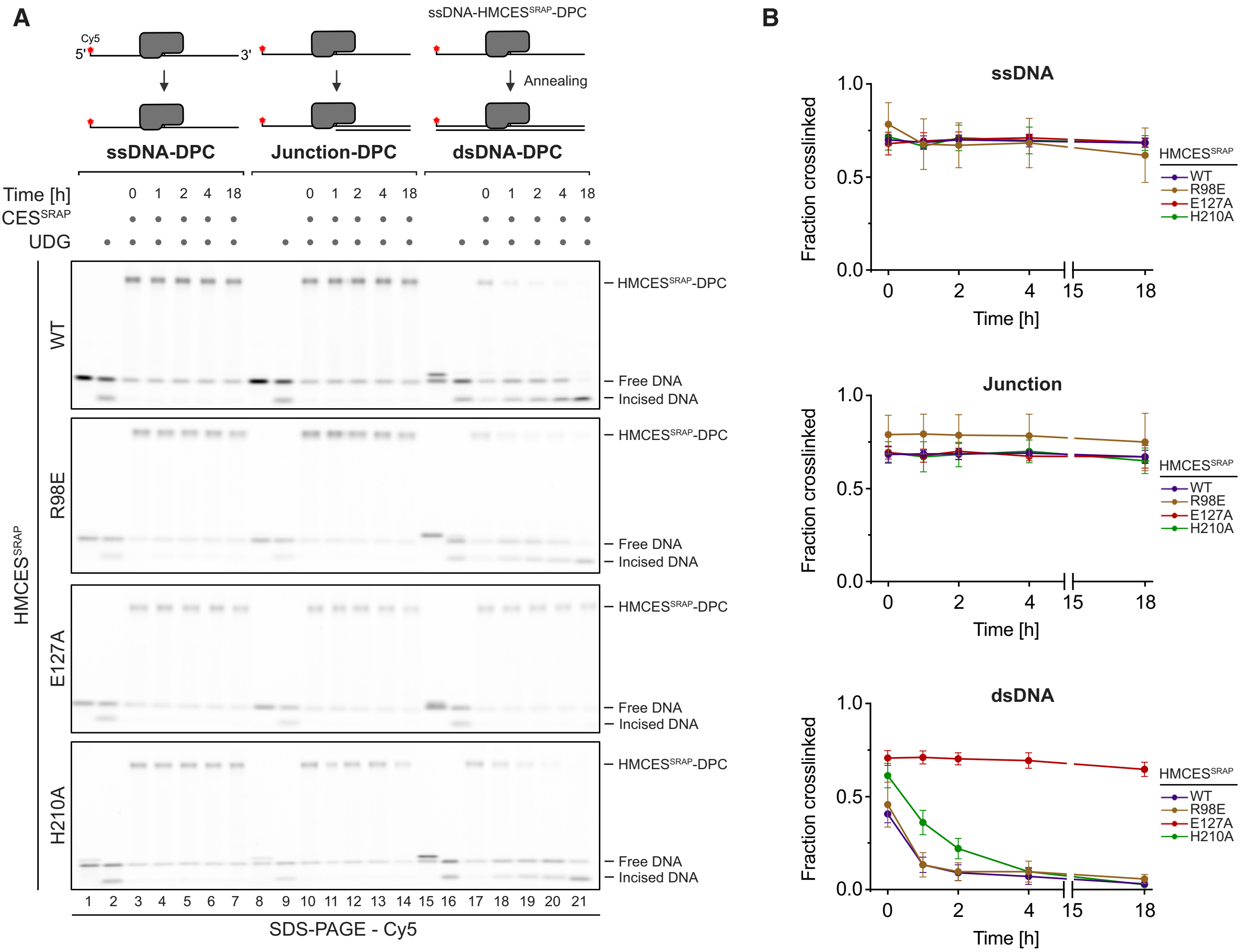


Figure 3. Release of HMCES-DPCs is determined by DNA context.

A DPC reversal kinetics of indicated variants in ssDNA, DNA junction and dsDNA. DPCs were pre-formed in ssDNA before corresponding reverse oligonucleotides were annealed (for ssDNA reactions, a non-complementary oligonucleotide was added). DPC reversal was then monitored after incubation for the indicated amount of time at 37°C using denaturing SDS–PAGE.

B Quantification of DPC reversal assays using HMCES^{SRAP}-WT and variants shown in (A).

Data information: Data in (B) represent the mean of three independent experiments ± SD.
Source data are available online for this figure.

E127A-DPCs fully blocked incision (Fig 5A and B, lanes 20–22). Therefore, failure to undergo auto-release would lead to the inhibition of AP site repair in dsDNA.

Importantly, we observed comparable results (with slight variations) when using the prokaryotic HMCES-orthologue YedK or *Xenopus laevis* HMCES (xl-HMCES) (Fig EV3B), indicating that the capacity of SRAP domain DPCs to auto-release is conserved across different species. YedK-DPCs were stable at ssDNA-dsDNA junctions, prohibiting cleavage of the AP site by APE1 (Fig EV3C, lanes 7–8). In dsDNA, APE1 was able to cleave the AP site, indicating release of the DPC (Fig EV3D, lanes 7–8). AP site cleavage in dsDNA was not observed upon replacing Glu105 (corresponding to Glu127 in HMCES, Fig 1A) with alanine (Fig EV3D, lanes 11–12). Of note, in contrast to human HMCES^{SRAP}, release of the DPC was barely detectable in the absence of APE1 (Fig EV3D, lanes 5–6). While the protection of AP sites at ssDNA-dsDNA junctions against APE1 cleavage by xl-HMCES-DPCs was less strong than observed for the human or prokaryotic protein (Fig EV3C, lanes 15–16), substitution of Glu129 (corresponding to Glu127 in HMCES, Fig 1A) entirely blocked reversal at ssDNA-dsDNA junctions and in dsDNA (Fig EV3C, lanes 19–20, and D, lanes 19–20, respectively).

Collectively, these data show that HMCES-DPC release must be finely balanced to (i) ensure protection of AP sites at ssDNA-dsDNA junctions against potentially catastrophic APE1 incisions (which is compromised upon hyper-release in the HMCES^{SRAP}-R98E variant) and to (ii) allow deprotection of AP site in dsDNA so that APE1 can initiate repair (which is compromised upon hypo-reversal in the HMCES^{SRAP}-E127A variant).

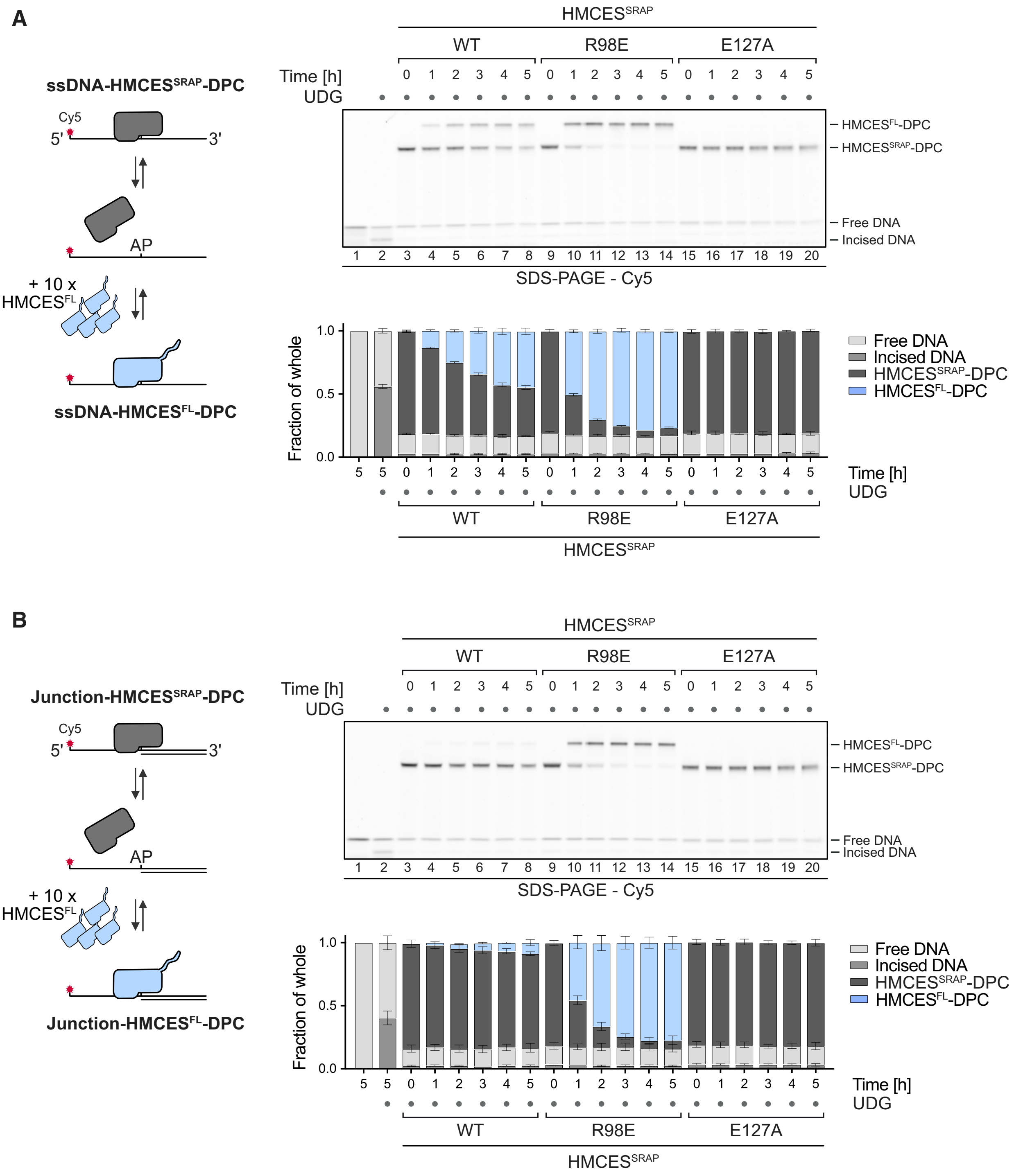


Figure 4. Release of HMCES-DPCs is determined by binding affinity to the underlying DNA.

A, B Competition assay between HMCES^{FL} and indicated HMCES^{SRAP} variants. HMCES^{SRAP}-DPCs in ssDNA (A) or at ssDNA-dsDNA junctions (B) were pre-formed and then incubated with 10-fold excess of HMCES^{FL} for the indicated amount of time at 37°C prior to separation by denaturing SDS gel (upper panels).

Data information: Quantification of competition assay: Bar graphs show the mean of three independent experiments ± SD (lower panels).
Source data are available online for this figure.

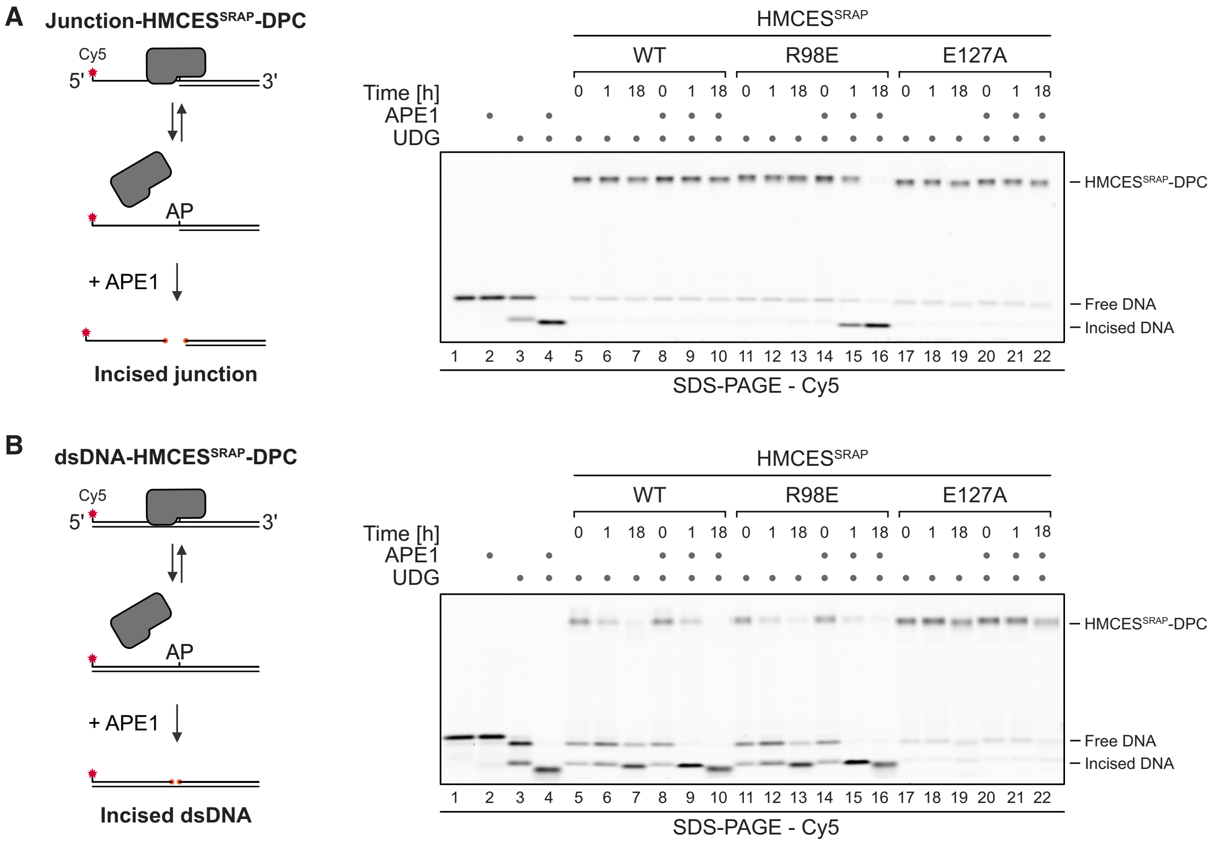


Figure 5. Auto-release of HMCES-DPCs restricts crosslink formation to physiologically relevant situations.

A, B APE1 incision of an AP site protected by the indicated HMCES^{SRAP}-DPC variants at ssDNA-dsDNA junctions (A) or within dsDNA (B). Free dU-containing DNA was incubated alone or in the presence of UDG and HMCES^{SRAP} for 1 h at 37°C. Next, corresponding reverse oligonucleotides were annealed to generate an ssDNA-dsDNA junction (A) or dsDNA (B), and reactions were incubated alone or with APE1 for the indicated amount of time at 37°C prior to separation by denaturing SDS–PAGE.

Source data are available online for this figure.

SPRTN-dependent proteolysis is the dominant mechanism for HMCES-DPC removal in *Xenopus* egg extracts

Next, we sought to test whether crosslink reversal contributes to HMCES removal in a more physiological system. We monitored HMCES-DPC stability during replication-coupled repair of a plasmid containing a site-specific AP-ICL (pICL-*lacO*^{AP}) in *Xenopus* egg extracts supplemented with WT recombinant 3xFlag-tagged *Xenopus laevis* HMCES protein (rHMCES-3xFlag) or E129A-mutated rHMCES-3xFlag. As shown in Fig EV3C and D, the WT protein undergoes efficient DPC reversal from dsDNA *in vitro* while the E129A-mutated protein does not. Similar to endogenous HMCES present in egg extract, both the WT and the E129A-mutated protein were barely detectable on chromatin isolated from extract containing SPRTN but accumulated on chromatin isolated from SPRTN-depleted extract (Fig EV4A–F). We therefore conclude that, relative to reversal, SPRTN-dependent proteolysis represents the primary mechanism for HMCES-DPC removal during ICL repair in egg extracts. However, technical challenges prevented us from determining whether reversal contributes to HMCES-DPC resolution when proteolysis is blocked (Fig EV4A–K, see figure legend for discussion).

Translesion synthesis across HMCES-DPCs triggers their release

The fact that HMCES-DPCs form specifically in ssDNA contexts (Fig 2A and B) (Mohni *et al*, 2019) leads to the question as to how HMCES-DPCs are transferred to dsDNA, where release could occur. Translesion synthesis (TLS) polymerases can extend nascent strands across intact HMCES-DPCs, as has been observed in *Xenopus* egg extracts and *in vitro* (Yaneva *et al*, 2023). Thus, we tested whether TLS across an HMCES-DPC triggers reversal. We placed HMCES^{SRAP}-WT or -E127A-DPCs in template DNA downstream of a primer (Fig 6A) and added TLS polymerase Pol ζ-Rev1 and the helicase FANCJ, which promotes TLS across intact DPCs through unfolding of the crosslinked protein adduct (Yaneva *et al*, 2023). These assays were analysed using UREA–PAGE, which allows separation of the template strand and the extended primer. As a control, we annealed a complementary 45mer reverse oligonucleotide, mimicking full extension. Indeed, extension of the primer by Pol ζ-Rev1 appeared to trigger release of HMCES-DPCs, as evidenced by a loss of WT-DPCs but not of E127A-DPCs (Fig 6B, compared lanes 4 and 5 and lanes 7 and 8, and Fig 6C); the assessment of DPC release was complicated by a fraction of the Cy5 signal remaining in the gel

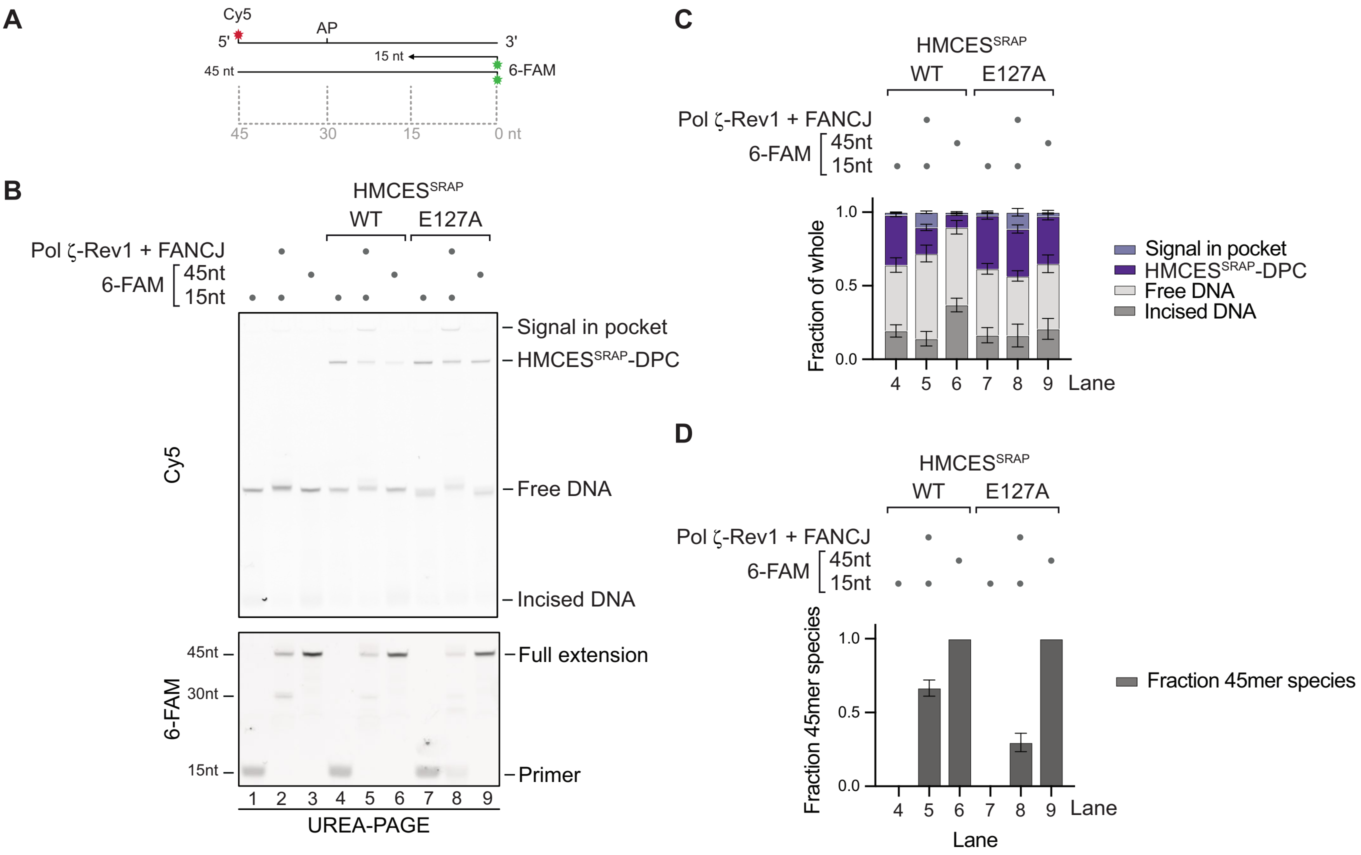


Figure 6. Translesion synthesis across an HMCES-DPC can trigger reversal *in vitro*.

A, B Primer extension assay using Pol ζ-Rev1. Fluorescently labelled primer-template substrates containing an AP site at the indicated position were incubated alone or in the presence of HMCES^{SRAP}-WT or -E127A, recombinant human FANCJ and Pol ζ-Rev1 as indicated for 2 h at 37°C prior to separation by denaturing UREA–PAGE. (A) Model of oligonucleotides. (B) Cy5 scan and 6-FAM scan of denaturing UREA–PAGE.

C Quantification of Cy5 signals shown in (B).

D Quantification of 6-FAM signals shown in (B).

Data information: Bar graphs in (C) and (D) show the mean of four independent experiments ± SD for lanes 4 to 7 and the mean of three independent experiments ± SD for lanes 8 and 9. One replicate was excluded for these conditions for technical reasons.

Source data are available online for this figure.

pocket in the presence of Pol ζ-Rev1-FANCJ (Fig 6B, lane 5 and 8). DPC release was more pronounced when the complementary 45mer oligonucleotide was annealed to the template (Fig 6B and C, lane 6), which is in line with the fact that TLS did not extend all primers across the DPC (Fig 6B, 6-FAM scan and Fig 6D, lane 5). Of note, we observed that extension of the primer was less efficient in templates containing a HMCES^{SRAP}-E127A-DPC (Fig 6B and D, compare lanes 5 and 8). Thus, we cannot exclude that reversal of some DPCs occurs prior to TLS-dependent extension. These results suggest that DPC reversal can be triggered by physiological processes that transfer HMCES-DPCs from ssDNA into dsDNA.

Discussion

In this study, we found that HMCES-DPCs do not necessarily require proteolytic repair because they feature a built-in release mechanism. Our data suggest a model in which auto-release of HMCES-DPCs occurs in two distinct steps (Fig 7). First, the conserved Glu127 residue (with a minor contribution of His210) catalyses the reversal of the crosslink between HMCES' active site cysteine and the AP site, as also observed in other recent work (Paulin *et al*, 2022). Second, the cysteine either re-crosslinks or HMCES dissociates from DNA resulting in release of the AP site. The decision between these two options appears to be determined by binding strength to the underlying DNA. HMCES binds tightly to ssDNA-dsDNA junctions, which favours re-crosslinking over release. In contrast, HMCES binds poorly to dsDNA, resulting in release. Thus, this model explains how HMCES can protect AP sites at ssDNA-dsDNA junctions against incisions by AP endonucleases, while promoting AP site cleavage within dsDNA.

An important question arises regarding the mechanisms that transfer HMCES-DPCs from ssDNA to dsDNA to create the conditions for DPC release. Two settings seem plausible: one, TLS polymerases extend nascent strands across intact HMCES-DPCs with the help of FANCJ (Fig 6; Yaneva *et al*, 2023); two, nascent strands may

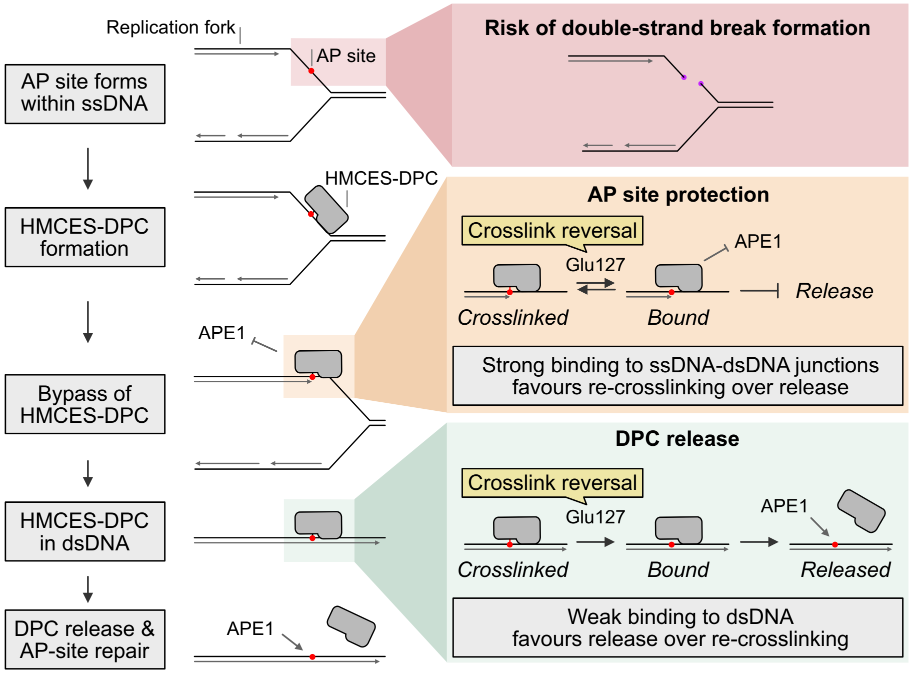


Figure 7. Model of AP site protection by coordinated formation and release of HMCES-DPCs.

AP sites within ssDNA at the replication fork are dangerous because they can lead to the formation of toxic DNA double-strand breaks. The conserved protein HMCES covalently crosslinks to AP sites in ssDNA to prevent strand scission.

HMCES-DPCs efficiently protect AP sites in ssDNA and at ssDNA-dsDNA junctions against spontaneous or enzymatic incisions. While the active site constantly cycles between a crosslinked and a non-crosslinked state (catalysed by Glu127), HMCES affinity to the underlying DNA is high which favours re-crosslinking over dissociation. In dsDNA, affinity is low which favours release of HMCES over re-crosslinking and thereby enables APE1 endonuclease to initiate AP site repair in dsDNA.

be extended upon template switching, which would avoid the need for TLS and thus ensure error-free repair (Mohni *et al*, 2019; Mehta *et al*, 2020; Srivastava *et al*, 2020). Notably, in either case, extension of the nascent strand past the protein adduct would prevent DPC proteolysis by SPRTN, which requires the presence of an ssDNA-dsDNA junction in close proximity to the protein adduct to become activated (Larsen *et al*, 2019; Reinking *et al*, 2020). The relative importance of release, which has the added benefit of recycling the enzyme, versus proteolytic repair of HMCES-DPCs, is an interesting future question. In *Xenopus* egg extracts, SPRTN-dependent proteolysis appeared to be the dominant mechanism for the resolution of -HMCES-DPCs forming during replication-coupled ICL repair (Fig EV4). However, template switching does not occur under these conditions (Semlow *et al*, 2022), but is the preferred mechanism to fill in DNA gaps during replication in mammalian cells (Tirman *et al*, 2021).

In mammalian cells, HMCES$^{\mathrm{E127A}}$ has been reported to complement the sensitivity to AP site-inducing drugs caused by an *HMCES* knock-out (Srivastava *et al*, 2020) and to protect AP sites during somatic hypermutation (Wu *et al*, 2022), which is in line with this variant's ability to form DPCs. We did not observe toxicity in human cells upon overexpression of HMCES$^{\mathrm{E127A}}$ (Fig EV5A–C), which may indicate that proteolytic repair fully compensates for the lack of DPC release. Both proteases targeting HMCES-DPCs, SPRTN and the proteasome are essential for cell viability, which prohibits the analysis of HMCES$^{\mathrm{E127A}}$ toxicity in the absence of DPC proteolysis in cells. Thus, the effects of defective auto-release remain unclear. However, human cells expressing HMCES$^{\mathrm{R98E}}$ are sensitive to ionizing radiation (Mohni *et al*, 2019), which together with our data indicate that increased DPC release compromises HMCES' ability to protect cells against AP sites in ssDNA. The precise regulation of the HMCES-DPC auto-release mechanism identified in this study emphasizes the need to restrict DPC formation to an absolute minimum.

Materials and Methods

Protein expression and purification

HMCES^{SRAP}

An open reading frame containing human HMCES$^{\mathrm{SRAP}}$ (amino acids 1–270) was codon optimized for bacterial expression and cloned in frame with a C-terminal His6-tag in a pNIC plasmid. HMCES$^{\mathrm{SRAP}}$-variants (-C2S, -R98E, -E127A, -H210A, -R98E/E127A) were generated by introducing point mutations using the Q5 site-directed mutagenesis kit (New England BioLabs), following manufacturer's instructions. Mutations were confirmed by Sanger sequencing. BL21 (DE3) *Escherichia coli* cells were transformed with the corresponding plasmids for protein expression. Cells were grown in terrific broth (TB) medium at 37°C to OD$_{600}$ 0.7. Protein expression was induced by addition of 0.5 mM isopropyl- β-D-thiogalactoside (IPTG) for 4 h. Cells were harvested, snap-frozen in liquid nitrogen and pellets were stored at −80°C. For protein purification, cells were resuspended in buffer A (50 mM HEPES/KOH pH 7.8, 500 mM KCl, 5 mM $MgCl_2$, 30 mM imidazole, 10% glycerol, 0.1% IGEPAL, 0.04 mg/ml Pefabloc SC, cOmplete EDTA-free protease inhibitor cocktail tablets and 1 mM tris(2-carboxyethyl) phosphine hydrochloride (TCEP)) and lysed by sonication. All subsequent steps were carried out at 4°C. Cell lysates were incubated with benzonase nuclease (45 U/ml lysate) for 30 min before cell debris was removed by centrifugation at 18,000 *g* for 30 min. Cleared and filtered supernatants were applied to 3 ml Ni-NTA Agarose (QIAGEN) equilibrated in buffer B (20 mM HEPES/KOH pH 7.8, 500 mM KCl, 5 mM $MgCl_2$, 30 mM imidazole, 10% glycerol and 1 mM TCEP). Next, beads were washed with 15 column volumes (CV) of buffer B before protein was eluted in 2 CV of buffer C (20 mM HEPES/KOH pH 7.8, 500 mM KCl, 5 mM $MgCl_2$, 300 mM imidazole, 10% glycerol and 1 mM TCEP). The elution was concentrated to 2 ml using a 10 kDa molecular weight cut-off Amicon ultra centrifugal filter prior to loading on a HiLoad® 16/600 Superdex® 200 pg column equilibrated in buffer D (20 mM HEPES/KOH pH 7.8, 150 mM KCl, 5 mM $MgCl_2$, 10% glycerol and 1 mM TCEP) for size exclusion chromatography. Eluted protein fractions were collected and concentrated with a 10 kDa molecular weight cut-off Amicon Ultra centrifugal filter. Concentrated protein was aliquoted, snap-frozen and stored at −80°C. Removal of the N-terminal methionine was confirmed by mass spectrometry.

YedK

An open reading frame containing bacterial YedK was cloned in frame with a C-terminal His6-tag in a pNIC plasmid. YedK-variant (−E105A) was generated by introducing a point mutation using the Q5 site-directed mutagenesis kit (New England BioLabs), following manufacturer's instructions. Mutation was confirmed by Sanger sequencing. Purification was as described above for $HMCES^{SRAP}$.

$HMCES^{FL}$

For full-length HMCES ($HMCES^{FL}$), the open reading frame was codon optimized and cloned in a pNIC plasmid in frame with a C-terminal TwinStrep-ZB-tag. Recombinant HMCES protein was expressed and purified using a protocol for purification of SPRTN (Reinking *et al*, 2020), with small modifications to some buffers. Cell pellets were resuspended in buffer A (20 mM HEPES/KOH pH 7.5, 500 mM KCl, 5 mM $MgCl_2$, 30 mM imidazole, 10% glycerol, 0.1% IGEPAL, 0.04 mg/ml Pefabloc SC, cOmplete EDTA-free protease inhibitor cocktail tablets and 1 mM TCEP). For washing steps, buffer B (20 mM HEPES/KOH pH 7.5, 500 mM KCl, 5 mM $MgCl_2$ and 1 mM TCEP) was used. Protein was eluted from Strep-Tactin®XT Superflow® high-capacity cartridges with buffer B containing 50 mM Biotin and from HiTrap Heparin HP affinity columns in buffer C (20 mM HEPES/KOH pH 7.5, 1 M KCl, 5 mM $MgCl_2$ and 1 mM TCEP). For size exclusion chromatography and storage of the protein, buffer D (20 mM HEPES/KOH pH 7.8, 150 mM KCl, 5 mM $MgCl_2$, 10% glycerol and 1 mM TCEP) was used. Removal of the N-terminal methionine was confirmed by mass spectrometry.

FANCJ

Human FANCJ-WT followed by a C-terminal TEV-cleavage site and TwinStrep-ZB tag was expressed and purified from Hi5 cells as described previously (Yaneva *et al*, 2023) with minor modifications. Briefly, cells were lysed in 200 ml of lysis buffer (50 mM Tris–HCl pH 8, 500 mM KCl, 0.1% Triton X-100, 10 mM $MgCl_2$, smDNAse nuclease, 0.04 mg/ml Pefabloc SC, cOmplete EDTA-free protease inhibitor cocktail tablets and 1 mM TCEP) with a Dounce homogenizer (25×). The cleared lysate was loaded on a 5 ml Strep-Tactin XT 4Flow cartridge. The column was washed with five column volumes (CV) of wash buffer (50 mM Tris–HCl pH 8, 150 mM KCl, 1 mM TCEP) and proteins were eluted with strep elution buffer (50 mM Tris–HCl pH 8, 150 mM KCl, 50 mM Biotin and 1 mM TCEP). Fractions were pooled and loaded on a 1 ml HiTrap Heparin HP affinity column equilibrated in wash buffer, and eluted in heparin elution buffer (50 mM Tris–HCl pH 8, 1 M KCl and 1 mM TCEP). Fractions were pooled, diluted down to 500 mM KCl and the Z-basic-TwinStrep tag was removed over night with the addition of His-TEV protease. Next, the protein sample was loaded on Superdex 200 Increase 10/300 GL column equilibrated in equilibration buffer (50 mM Tris–HCl pH 8, 200 mM KCl, 10% glycerol and 1 mM TCEP). Eluted proteins were concentrated with 10 kDa cut-off Amicon ultra centrifugal filters before snap-freezing in liquid nitrogen and storage at -80°C.

Pol ζ-Rev1

Pol ζ and Rev1 were purified as described previously (Kochenova *et al*, 2017).

Biotinylated LacI

Biotinylated LacI was purified as described previously (Dewar *et al*, 2015). Briefly, pET11a[LacR-Avi] and pBirAcm (Avidity) vectors were transformed into T7 express-competent cells. LacI and biotin ligase expression was induced with 1 mM IPTG in Luria–Bertani (LB) medium supplemented with 50 μM biotin for 2 h at 37°C. Cells were harvested, snap frozen and stored at −80°C. Cell pellets were lysed in lysis buffer (50 mM Tris–HCl pH 7.5, 5 mM EDTA, 100 mM NaCl, 1 mM DTT, 10% sucrose, cOmplete protease inhibitors, 0.2 mg/ml lysozyme and 0.1% Brij 58) for 30 min at room temperature (RT). Lysates were centrifuged at 21,300 *g* for 1 h at 4°C. Pellets containing chromatin-bound LacI were then suspended in 50 mM Tris–HCl pH 7.5, 5 mM EDTA, 1 M NaCl, 30 mM IPTG, 1 mM DTT and LacI was released from DNA by sonication followed by addition of polymin P to 0.03–0.06% (w/v) at 4°C. Biotinylated LacI was precipitated with 37% ammonium sulphate, pelleted by centrifugation and then suspended in 50 mM Tris–HCl pH 7.5, 1 mM EDTA, 2.6 M NaCl, 1 mM DTT and cOmplete protease inhibitors. Biotinylated LacI was then bound to SoftLINK avidin, washed with 50 mM Tris–HCl pH 7.5, 1 mM EDTA, 2.6 M NaCl, 1 mM DTT, cOmplete protease inhibitors and eluted with 50 mM Tris–HCl pH 7.5, 100 mM NaCl, 1 mM EDTA, 5 mM biotin and 1 mM DTT. Pooled fractions containing biotinylated LacI were buffer exchanged into 50 mM Tris–HCl pH 7.5, 150 mM NaCl, 1 mM EDTA and 1 mM DTT using an Amicon ultra-0.5 ml 3 kDa molecular weight cut-off filter unit. Biotinylated LacI was aliquoted, snap frozen, and stored at −80°C.

USP2-cc

To purify USP2-catalytic core (USP2-cc), $pH_{10}E$ USP-cc plasmid was transformed into BL21 (DE3) *E. coli* cells. Expression was induced with 0.5 mM IPTG in LB medium for 16 h at 18°C. Cells were pelleted and lysed in lysis buffer E (20 mM Tris pH 8.0, 300 mM NaCl, 10 mM imidazole, 10% glycerol, 1% Triton X-100, 1 mg/ml lysozyme, cOmplete EDTA-free protease inhibitor cocktail tablets, 5 mM 2-β-mercaptoethanol (BME) and 10 U/ml benzonase (Sigma, 70746-3)). Cell lysates were cleared by centrifugation at 18,000 *g* for 20 min. His-tagged Usp2-cc was bound to Ni-NTA Agarose (QIAGEN) equilibrated in buffer F (20 mM Tris pH 8.0, 300 mM NaCl, 10 mM imidazole, 10% glycerol, 0.05% Triton X-100 and 5 mM BME) for 4 h at 4°C, washed three times with buffer F and then eluted with buffer G (20 mM Tris pH 8.0, 300 mM NaCl, 300 mM imidazole, 10% glycerol, 0.05% Triton X-100 and 1 mM TCEP). Eluted protein was dialyzed in dialysis buffer H (20 mM Tris pH 8.0, 150 mM NaCl, 10% glycerol and 1 mM TCEP). Protein was then aliquoted, snap frozen and stored at −80°C.

xl-HMCES

An open reading frame containing *Xenopus laevis* HMCES (amino acid 2–336) was cloned in frame with an N-terminal His10-ubiquitin (Ub) to generate pHUE-*xl*.HMCES plasmid (Catanzariti *et al*, 2004). pHUE-*xl*.HMCES(E129A) plasmid was generated by inverse PCR using primer pairs (5'-CAG GAC GGT GAA AAA CAA CCG TAC-3′/ 5'-GCG TTT CCA TGC ATA GAA CCC GTC C-3′). All constructs were confirmed by Sanger sequencing and transformed into ArcticExpress (DE3)-competent cells for protein expression. Cells were grown in LB medium at 37°C to OD_{600} 0.6. Protein expression was induced by addition of 0.5 mM IPTG for 24 h. Cells were harvested, washed once with PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4 and 1.8 mM KH_2PO_4), snap-frozen and stored at −80°C. For protein purification, cells were resuspended in buffer E, incubated on ice for 30 min and briefly sonicated. Cell debris was removed by centrifugation at 18,000 *g* for 20 min. Cleared supernatants were applied to Ni-NTA Agarose (QIAGEN) equilibrated in buffer F. Next, beads were washed thrice with 15 CV of buffer F and the protein was eluted with 4 CV of buffer G. The eluted protein was dialyzed against buffer H. The His-tagged proteins were incubated overnight at 4°C with His10-USP2-cc (molar ratio 1/100) to cleave the His10-Ub tag from the N-terminus of HMCES. The cleavage reaction mixtures were incubated with 1 ml prewashed Ni-NTA agarose to remove His10-Ub, His10-USP2-cc, and uncleaved His10-Ub-HMCES. HMCES in the flowthrough was further purified by anion exchange chromatography using mono Q50/5 GL column (Cytiva). Samples were eluted over a gradient of 150 to 100 mM NaCl. Fractions containing proteins were pooled and concentrated using Amicon Ultra-15 centrifugal filter unit with 10 kDa molecular weight cut-off. Protein was aliquoted, snap-frozen and stored at −80°C.

xl-HMCES-3xFlag

To purify xl-HMCES-3xFlag and xl-HMCES(E129A)-3xFlag, a DNA sequence encoding 3xFlag was inserted downstream of xl-HMCES and xl-HMCES(E129A) in pHUE backbone plasmid to generate pHUE-xl.HMCES-3xFlag and pHUE-xl.HMCES (E129A)-3xFlag, respectively. Correct sequences were confirmed by Sanger sequencing followed by transformation of plasmids into ArcticExpress (DE3)-competent cells. Xl-HMCES-3xFlag and xl-HMCES(E129A)-3xFlag proteins were expressed and purified as described above for xl-HMCES and xl-HMCES(E129A).

Generation of HMCES-DPCs

Crosslinking reactions with different HMCES variants were carried out in 10 μl reactions containing 8.02 μl reaction buffer (20 mM HEPES/KOH pH 7.5, 50 mM KCl, 10 mM $MgCl_2$, 2 mM TCEP and 0.1 mg/ml BSA), 0.5 μl $HMCES^{FL}/HMCES^{SRAP}$ (prediluted to 40 μM in purification buffer D), 1 μl Cy5-labelled forward oligonucleotide (prediluted to 10 μM in DPC dilution buffer—50 mM HEPES/KOH pH 7.5, 100 mM KCl, 10% glycerol and 0.4 mg/ml BSA) and 0.48 μl UDG (New England BioLabs), adding up to final concentrations of 2 μM $HMCES^{FL}/HMCES^{SRAP}$, 1 μM DNA and 0.1 U/μl UDG (New England Biolabs). Reactions were incubated for 1 h at 37°C. Crosslinking reactions with YedK and xl-HMCES (Fig EV3C and D) were carried out identically as described above, except that the 10 μl reactions contained 1.5 μl YedK/xl-HMCES (prediluted to 13.2 μM in purification buffer D) and 7.02 μl reaction buffer. As a standard and if not stated otherwise, a 30mer oligonucleotide containing a central dU (5′-Cy5-CCC AAA AAA AAA AAdU AAA AAA AAA AAA CCC-3′) was used for crosslinking. For generation of different DNA structures, 1 μl of corresponding reverse oligonucleotides (diluted to 12 μM in nuclease-free H_2O) were added to the crosslinking reaction and incubated for 2 min at 37°C before the temperature was decreased by 1°C/min until 20°C was reached to allow annealing of the reverse oligo. For ssDNA samples, a non-complementary reverse oligo was added (5′-AAA CCC CCC CCC CCA CCC CCC CCC AAA-3′); for ssDNA-dsDNA junction samples, a 15mer reverse oligo was added (5′-GGG TTT TTT TTT TTT-3′); and for dsDNA samples, a 30mer reverse oligo was added (5′-GGG TTT TTT TTT TTT ATT TTT TTT TTT GGG-3′). Reverse oligonucleotides were annealed prior to crosslinking for experiments shown in Fig 2A and C. For experiments shown in Figs 3A, 4, 5, EV2C, EV3C and D, and EV4I, reverse oligonucleotides were annealed after crosslinking.

HMCES-DPC formation assays

For the experiments shown in Fig EV1B, the indicated HMCES variants were prediluted to 64, 32, 16, 8, 4, 2, 1 and 0.5 μM in purification buffer D prior to crosslinking. 0.5 μl of the predilutions were added to the crosslinking reactions as described above resulting in final HMCES concentrations of 3.2, 1.6, 0.8, 0.4, 0.2, 0.1, 0.05 and 0.025 μM. Otherwise, crosslinking reactions were carried out as described above. Reactions were stopped by addition of 5.5 μl LDS sample buffer and boiling for 1 min at 95°C. Samples were resolved on 4–12% SDS–PAGE gels. Gels were photographed using a BioRad Chemidoc MP system using appropriate filter settings for Cy5 fluorescence. Crosslinking was quantified using ImageJ by measuring the relative fraction of Cy5 signal in the DPC band.

For the experiments shown in Fig 2A and C, crosslinking reactions were set up as described above. Incubation and annealing of reverse oligonucleotides were performed in the absence of HMCES to generate desired DNA structures (ssDNA, ssDNA-dsDNA junction and dsDNA) before incubation with HMCES. For experiments in Fig 2C, a different 30mer oligonucleotide containing a central dU (5′-Cy5-CCC CCG GAA AAA AAdU AAA AAA AAG GCC CCC-3′) was used and annealed with a 15mer reverse oligonucleotide (5′-Fam-GGG GGC CTT TTT TTT-3′) for ssDNA-dsDNA junction, two 15mer reverse oligonucleotide (5′-Fam-GGG GGC CTT TTT TTT-3′ and 5′-TTT TTT TTC CGG GGG-3′) for dsDNA containing a nick, a 15mer reverse oligonucleotide and a 10mer reverse oligonucleotide (5′-Fam-GGG GGC CTT TTT TTT-3′ and 5′-T TTC CGG GGG-3′) for dsDNA containing a 4 nt-gap and a 30mer reverse oligonucleotide (5′-GGG GGC CTT TTT TTT TTT TTT TTC CGG

GGG-3′) for dsDNA. Following annealing, 0.5 µl HMCES (prediluted to 40 µM in purification buffer D) was added to the reactions. Reactions were incubated for 0, 1, 5, 10, 20, or 30 min at 37°C before being stopped by addition of 5.5 µl LDS sample buffer and boiling for 1 min at 95°C. Samples were frozen in liquid nitrogen and stored at −80°C. Before resolving samples on 4–12% SDS–PAGE gels, samples were boiled again at 95°C for 30 s. Gels were photographed using a BioRad Chemidoc MP system using appropriate filter settings for Cy5 fluorescence. Quantification was performed using ImageJ by measuring the relative fraction of Cy5 signal in the DPC band.

HMCES-DPC release assays

The experiments shown in Fig 1E and G indicated that HMCES variants were crosslinked to a 30mer oligonucleotide containing a central dU or dT (5′-Cy5-CCC AAA AAA AAA AAdU/dT AAA AAA AAA AAA CCC-3′) as described above. In parallel, crosslinking reactions containing a 6-FAM-labelled 30mer oligonucleotide also containing a central dU (5′-6-FAM-CCC AAA AAA AAA AAdU AAA AAA AAA AAA CCC-3′) with 0.5 µl purification buffer D instead of protein were prepared and incubated at 37°C for 1 h as well. To inactive non-crosslinked HMCES, reactions containing the Cy5-oligonucleotide were incubated for 5 min at 60°C. In the following step, reactions containing the Cy5-oligonucleotide and HMCES were mixed 1:1 with reactions containing the 6-FAM-labelled oligonucleotide and incubated for 2 h at 37°C. To stop reactions, 11 µl of LDS sample buffer was added and reactions were boiled for 1 min at 95°C before analysis on 4–12% SDS–PAGE gels. Gels were photographed using a BioRad Chemidoc MP system using appropriate filter settings for Cy5 and 6-FAM fluorescence. 6-FAM- and Cy5-DPC formation was quantified using ImageJ.

For experiments shown in Figs 3A, EV2C and EV4G, indicated HMCES variants were crosslinked to an ssDNA oligonucleotide, as described above. Afterwards, corresponding reverse oligonucleotides were annealed as described above. Following annealing, 1 µl of the crosslinking reaction was added to 9 µl of master mix, resulting in a final buffer composition of 17.1 mM HEPES, 85.6 mM KCl, 3.1% glycerol, 5.5 mM TCEP, 2 mM $MgCl_2$ and 0.1 mg/ml BSA. Reactions were either stopped directly after annealing (0 h) or after 1, 2, 4, or 18 h incubation at 37°C by addition of 5.5 µl LDS sample buffer. For experiments shown in Fig EV4G, corresponding samples were incubated at 20°C after annealing. Reactions were boiled for 1 min at 95°C before being frozen in liquid nitrogen and stored at −80°C. Samples were boiled at 95°C for 30 s before being resolved on 4–12% SDS–PAGE gels. Gels were photographed using a BioRad Chemidoc MP system using appropriate filter settings for Cy5 fluorescence. Quantification was performed using ImageJ by measuring the relative fraction of Cy5 signal in the DPC band.

For the experiments shown in Fig 4, indicated $HMCES^{SRAP}$ variants were crosslinked as described above. Afterwards, corresponding reverse oligonucleotides were annealed to create an ssDNA-dsDNA junction, while a non-complimentary oligonucleotide was added in ssDNA conditions. $HMCES^{FL}$ was prediluted to 20 µM in competition buffer (150 mM KCl, 50 mM HEPES and 10% glycerol). The final assay was carried out in a reaction volume of 10 µl with 1 µl of the crosslinking reaction and 1 µl of prediluted $HMCES^{FL}$, in a final buffer composition of 17.1 mM HEPES, 85.6 mM KCl, 3.1% glycerol, 5.5 mM TCEP, 2 mM $MgCl_2$ and 0.1 mg/ml BSA. Reactions were incubated for 0, 1, 2, 3, 4 or 5 h at 37°C before being stopped by addition of 5.5 µl LDS sample buffer. The reactions were boiled for 1 min at 95°C before being frozen in liquid nitrogen and stored at −80°C. After thawing, samples were boiled at 95°C for 30 s before being resolved on 4–12% SDS–PAGE gels. Gels were photographed using a BioRad Chemidoc MP system, using appropriate filter settings for Cy5 fluorescence. Quantification was done using ImageJ, by measuring the relative Cy5 signals of $HMCES^{FL}$-DPCs and $HMCES^{SRAP}$-DPCs.

APE1 incision assays

The experiments shown in Fig 5 indicated $HMCES^{SRAP}$ variants were crosslinked to a 30mer oligonucleotide containing a central dU as described above. Reverse oligonucleotides were annealed to generate an ssDNA-dsDNA junction or dsDNA. After annealing, 1 µl of the $HMCES^{SRAP}$-DNA crosslinking reaction and 0.5 µl of APE1 (New England BioLabs) were added to 8.5 µl final reaction buffer, bringing final concentrations to 17.1 mM HEPES, 85.6 mM KCl, 3.1% glycerol, 5.5 mM TCEP, 2 mM $MgCl_2$ and 0.1 mg/ml BSA. For experiments shown in Fig EV3C and D, reactions were prepared similarly except that 1.5 µl prediluted YedK-WT or -E105A and xl-HMCES-WT or -E129A were added with the crosslinking reaction with 7.02 µl reaction buffer, as described before. To samples not containing APE1, 0.5 µl APE1 buffer was added (10 mM Tris–HCl, 50 mM NaCl, 1 mM DTT, 0.05 mM EDTA, 200 µg/ml BSA and 50% glycerol, pH 8). Reactions were either stopped directly (0 h) or after 1 or 18 h of incubation at 37°C by the addition of 5.5 µl LDS sample buffer. Reactions were boiled for 1 min at 95°C before being frozen in liquid nitrogen and stored at −80°C. Samples were boiled again at 95°C for 30 s after thawing before being resolved on 4–12% SDS–PAGE gels. Gels were photographed using a BioRad Chemidoc MP system using appropriate filter settings for Cy5 fluorescence.

DNA-binding assays

Electrophoretic mobility shift assays

$HMCES^{SRAP}$-WT and variants were prediluted to 40, 10 and 2.5 µM in purification buffer D. Binding reactions were carried out in 10 µl with 0.5 µl of $HMCES^{SRAP}$ dilutions, 1 µl of 1 µM Cy5-labelled 30mer dT-oligonucleotide and 8.5 µl reaction buffer (20 mM HEPES/KOH pH 7.5, 50 mM KCl, 10 mM $MgCl_2$, 2 mM TCEP and 0.1 mg/ml BSA). Reactions were incubated for 20 min on ice before addition of 4 µl 6× Orange G-loading dye. Samples were then resolved at 4°C on 6% native PAGE gels using 0.5× TBE as running buffer. Gels were photographed using a BioRad Chemidoc MP system using appropriate filter settings for Cy5 fluorescence.

Fluorescence polarization

$HMCES^{SRAP}$-WT and variants were prediluted to 200 µM, 40 µM, 8 µM, 1.6 µM, 0.32 µM, 1.28 nM and 0.256 nM in purification buffer D. Binding was carried out in 50 µl final volume with 5 µl of $HMCES^{SRAP}$ predilutions, 5 µl of 250 nM Cy5-labelled 30mer dT-oligonucleotide and 40 µl of reaction buffer (20 mM HEPES/KOH pH 7.5, 50 mM KCl, 10 mM $MgCl_2$, 2 mM TCEP and 0.1 mg/

ml BSA). Binding reactions were incubated for 20 min on ice before 10 µl of the reactions were pipetted into a 384-well microplate (Greiner Bio-One). Fluorescence polarization was measured using a Tecan Spark multimode microplate reader using appropriate filter settings for Cy5 fluorescence.

Primer extension assay

Primer extension assays were used to analyse HMCES^{SRAP} reversibility following bypass by TLS polymerases Pol ζ-Rev1. HMCES^{SRAP}-WT-DPCs and -E127A-DPCs were crosslinked as described above to a 45mer forward Cy5-labelled oligonucleotide containing a dU at position 30 (5′-CY5-ACC AGT GCC TTG CT[U] GGA CAT CTT TGC CCA CCT GCA GGT TCA CC-3′). To generate an AP site, 0.1 U/µl UDG (New England Biolabs) was added to the crosslinking reaction. After 1 h incubation at 37°C, either a 15mer 6-FAM-labelled primer (5′-6-FAM-GGG TGA ACC TGC AGG-3′) or a corresponding 45mer 6-FAM labelled oligonucleotide (5′-6-FAM-GGG TGA ACC TGC AGG TGG GCA AAG ATG TCC AAG CAA GGC ACT GGT-3′) to generate dsDNA was annealed as described above. FANCJ-dependent primer extension with Pol ζ-Rev1 was performed as described previously (Yaneva *et al*, 2023). For the final reaction, 1 µl of the HMCES^{SRAP}-DPC reaction, 100 nM FANCJ, 25 nM Pol ζ and 40 nM Rev1 were mixed in a final volume of 10 µl and the following conditions: 17.1 mM HEPES/KOH pH 7.5, 5.5 mM Tris–HCl pH 8.0, 70.6 mM KCl, 2.5 mM NaCl, 6.6% glycerol, 5 mM TCEP, 2 mM ATP, 0.2 mM dNTPs, 2.9 mM $MgCl_2$ and 0.1 mg/ml BSA. Reactions were incubated for 2 h at 37°C and stopped by the addition of 10 µl UREA-loading buffer (8 M UREA, 15% Ficoll). The reactions were then boiled for 10 min at 95°C and resolved on denaturing 12% UREA–PAGE gels (12% acrylamide, 8 M UREA and 1×TBE) at 60°C in 1× TBE running buffer. Gels were photographed using a BioRad Chemidoc MP system using appropriate filter settings for Cy5 and 6-FAM fluorescence. Quantification was performed using ImageJ, by measuring the relative Cy5 signals (Fig 5C) or the relative 6-FAM signals of the 45mer species (Fig 5D).

Generation of cell lines

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) foetal bovine serum (FBS). For doxycycline-inducible overexpression of HMCES variants, the coding sequence of HMCES was amplified from cDNA using Q5 Master Mix (M0544, NEB) before being shuttled into p221 plasmid using BP clonase (11789100, Thermo Fisher). Next, the E127A mutation was introduced with Q5® Site-Directed Mutagenesis Kit Protocol (E0554, NEB) and both sequences were subcloned into -pcDNA5/FRT/TO-mVenus-3xFlag-Gateway (Addgene, #40999) using LR clonase (11791020, Thermo Fisher) before generation of stable cell lines using the T-REx Flp-In system (Thermo Fisher) according to manufacturer's instructions. Briefly, HeLa T-REx Flp-In cells were grown to 50% confluency in six-well plates prior to transfection of pOG44 (1.8 µg) and the respective pcDNA5-FRT/TO plasmids (0.2 µg, containing HMCES-WT-mVenus-3xFlag, HMCES-E127A-mVenus-3xFlag or mVenus-3xFlag (the gateway recombination cassette was deleted) using Lipofectamine 2000 (Invitrogen). Sixteen hours after transfection, cells were selected in 150 µg/ml hygromycin B (Fisher Scientific)-containing DMEM media for 10 days.

Cell viability assay

To analyse cell viability of HeLa T-REx Flp-In cells overexpressing HMCES-WT-mVenus-3xFlag or HMCES-E127A-mVenus-3xFlag, cells were counted and seeded in 12-well plates (10,000 cells per well) with DMEM −/+ 1 µg/ml doxycycline in technical triplicates and incubated for 4 days. HeLa T-REx Flp-In cells overexpressing mVenus-3xFlag were included as a control. After this, cell viability was assessed by AlamarBlue assay (Sigma, R7017-1G, 0.04% in PBS assay concentration). One well of each condition was harvested and analysed by western blotting using anti-HMCES antibody (Santa Cruz, #sc-514238), anti-Flag-M2 antibody (Sigma, #F3165) and anti-Vinculin antibody (Santa Cruz, #sc-73614). Plates were afterwards stained with 0.5% crystal violet and scanned.

Preparation of oligonucleotide duplexes with AP-ICL

To generate the AP-ICL-containing oligonucleotide duplex, the complementary 5′-phosphorylated oligonucleotides (AP-ICL top: 5′-GCA CCT TCC GCT CdUT CTT TC-3′ and AP-ICL bottom: CCC TGA AAG AAG AGC GGA AG) heated for 5 min at 95°C in 30 mM HEPES-KOH pH 7.0, 100 mM NaCl and annealed by cooling at 1°C per min to 18°C. The annealed duplex was treated with uracil glycosylase (NEB) in 1 × UDG buffer for 2 h at 37°C followed by phenol/chloroform extraction and ethanol precipitation. The oligo duplex was suspended in 50 mM HEPES-KOH pH 7.0, 100 mM NaCl and then incubated at 37°C for 5 days to allow for crosslink formation. Crosslinked DNA duplex was separated on a 20% polyacrylamide, 1 × Tris-borate-EDTA (TBE) and 8 M urea gel, and the crosslinked product was excised from the gel and eluted into TE (pH 8.0) buffer. Eluted DNA was concentrated by adding 4.5 times volume of 1-butanol, extracted with phenol:choloroform:isoamyl alcohol (25:24:1; pH 8.0) and precipitated with ethanol. The AP-ICL DNA oligo was then suspended in 10 mM Tris–HCl (pH 8.5) and stored at −80°C.

Preparation of crosslink DNA construct pICL-*lacO*^{AP}

pICL-*lacO*^{AP} was prepared as described previously (Semlow *et al*, 2016). Briefly, the backbone plasmid (with 48 *lacO* repeats) was incubated with BbsI in NEBuffer 2.1 for 24 h at 37°C followed by phenol/chloroform extraction. The digested plasmid was further purified using a HiLoad 16/60 Superdex 200 column, which was equilibrated in TE pH 8.0 buffer. Fractions containing the linearized plasmid were pooled, precipitated in ethanol and dissolved in 10 mM Tris–HCl pH 8.5. The abasic site interstrand crosslink (ICL^{AP})-containing duplexes were ligated into the linearized plasmid backbone using T4 DNA ligase (NEB). The ligated plasmid was dialysed into TE pH 8.0 buffer and concentrated using an Amicon Ultra-15 10 kDa molecular weight cut-off filter unit. The covalently closed circular plasmids were further extracted using the CsCl ethidium bromide method. Ethidium bromide was then removed from DNA by mixing with equal volume of saturated isobutanol. The purified pICL-*lacO*^{AP} was then dialyzed into TE pH 8.0, concentrated, snap frozen and stored at −80°C.

Xenopus egg extracts

Unfertilized eggs were collected from female *Xenopus laevis* frogs (Xenopus1, Cat# 4280; age > 2 years). Sperm chromatin was prepared from male *Xenopus laevis* frogs (Xenopus1, Cat# 4290, age > 1 year). All animal work was performed at Caltech and by the IACUC (Protocol IA20-1797, approved 28 May 2020). The institution has an approved Animal Welfare Assurance (no. D16-00266) from the NIH Office of Laboratory Animal Welfare. The high-speed supernatant (HSS) and nucleoplasmic extracts (NPE) were prepared from Xenopus laevis eggs as described previously (Semlow *et al*, 2022). Briefly, six adult female X. laevis frogs were induced to produce eggs by injection with 500 IU hCG. Eggs were collected and dejellied in 1 l of 2.2% (w/v) cysteine, pH 7.7. Dejellied eggs were then washed with 2 l of 0.5× Marc's modified Ringer's solution (2.5 mM HEPES-KOH, pH 7.8, 50 mM NaCl, 1 mM KCl, 0.25 mM $MgSO_4$, 1.25 mM $CaCl_2$ and 0.05 mM EDTA) followed by 1 l of egg lysis buffer (ELB; 10 mM HEPES-KOH, pH 7.7, 50 mM KCl, 2.5 mM $MgCl_2$ and 250 mM sucrose) supplemented with 1 mM DTT and 50 μg/ml cycloheximide. Eggs were then packed and crushed in the presence of 5 μg/ml aprotinin, 5 μg/ml leupeptin and 2.5 μg/ml cytochalasin B by centrifugation at 20,000 *g* for 20 min at 4°C. The soluble extract layer (the low-speed supernatant (LSS)) was collected and supplemented with 50 μg/ml cycloheximide, 1 mM DTT, 10 μg/ml aprotinin, 10 μg/l leupeptin and 5 μg/ml cytochalasin B. LSS was centrifuged in thin-walled ultracentrifuge tubes at 260,000 *g* for 90 min at 2°C in a TLS 55 rotor. Supernatant (HSS) was collected, aliquoted, snap-frozen in liquid nitrogen and stored at −80°C. To prepare NPE, LSS was prepared from eggs collected from 20 female X. laevis frogs as described above. LSS was then supplemented with 50 μg/ml cycloheximide, 1 mM DTT, 10 μg/ml aprotinin, 10 μg/ml leupeptin, 5 μg/ml cytochalasin B and 3.3 μg/ml nocodazole, and centrifuged at 20,000 *g* for 15 min at 4°C. After removing lipids, the clarified cytoplasmic fraction was collected and supplemented with ATP-regenerating mix (2 mM ATP, 20 mM phosphocreatine and 5 μg/ml phosphokinase) and 4,400 demembranated X. laevis sperm chromatin/μl to initiate nuclei formation. After ~90 min incubation at RT, reaction mixture was centrifuged for 3 min at 18,000 *g* at 4°C. The nuclei layer was collected from the top of the tubes and centrifuged at 260,000 *g* for 30 min at 2°C. The lipid layer was removed and the NPE fraction was collected, aliquoted, snap frozen in liquid nitrogen and stored at − 80°C.

Immunodepletions

Immunodepletions of SPRTN and REV1 were performed as described (Semlow *et al*, 2022). Briefly, protein A Sepharose fast flow beads were washed in 1× PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4 and 1.8 mM KH_2PO_4) and then incubated with an appropriate volume of antibodies overnight at 4°C. For SPRTN depletion, three volumes of polyclonal SPRTN anti-serum (Pocono Rabbit Farm and Laboratory rabbit 31053) were used for each volume of beads. For REV1 depletion, one volume of polyclonal REV1 C-terminus anti-serum (Pocono Rabbit Farm and Laboratory rabbit 1010) or one volume of polyclonal REV1 N-terminus anti-serum (Pocono Rabbit Farm and Laboratory rabbit 714) was used for each volume of beads. Equivalent volumes of rabbit pre-immune serum were incubated with beads for mock depletions. The beads were then washed twice with PBS, once with ELB, twice with ELB supplemented with 500 mM NaCl and thrice with ELB. HSS and NPE were immunodepleted by three rounds of incubation at 4°C for 60 min with protein A Sepharose-bound antibodies (five volume extract per volume of beads). In the case of REV1 depletion, two rounds of depletion were performed using the N-terminal antibody and one round was performed using the C-terminal antibody. Extracts were centrifuged for 30 s at 622 *g* in an S-24-11-AT rotor using an Eppendorf 5430R centrifuge and the supernatants were collected.

Replication reactions

For replication reactions, licensing was conducted by incubating HSS with 15 ng/μl pICL-*lacO*^{AP} plasmid in the presence of 3 μg/ml nocodazole, 20 mM phosphocreatine, 2 mM ATP and 5 μg/ml creatine phosphokinase with or without 0.1 μM 3,000 Ci/mmol [α-^{32}P] dCTP for 30 min at RT. Replication was then initiated by mixing two volumes of NPE mix (50% (v/v) NPE, 20 mM phosphocreatine, 2 mM ATP, 5 μg/ml creatine phosphokinase and 13.5 mM DTT in ELB) with one volume of licensing mix. Replication reactions were additionally supplemented with 0.2 or 2 μM Xl-HMCES-3xFlag, as indicated. ^{32}P-radiolabeled reactions were quenched by adding 1 μl of replication reaction to 6 μl of replication stop buffer (8 mM EDTA, 0.13% phosphoric acid, 10% ficoll, 5% SDS, 0.2% bromophenol blue and 80 mM Tris–HCl, pH 8.0) at the indicated time points followed by digestion with proteinase K (2.5 mg/ml) for 60 min at 37°C. Replication products were resolved on 0.8% agarose gels and visualized by phosphor imaging using a GE Healthcare Typhoon FLA 9500 (FujiFilm FLA 9500 user interface v.1.1). Images were analysed using Image Lab v.6.4.0 (Bio-Rad).

Plasmid pulldowns and immunoblotting

Plasmid pulldowns were performed as described previously (Semlow *et al*, 2022). Briefly, streptavidin-coupled magnetic Dynabeads (10 μl per pull down) were washed thrice with bead wash buffer 1 (50 mM Tris–HCl pH 7.5, 150 mM NaCl, 1 mM EDTA pH 8.0 and 0.02% Tween-20). Dynabeads were then incubated with biotinylated LacI (0.4 pmol per 1 μl of beads) at RT for 60 min. The beads were then washed four times with pulldown buffer (20 mM Tris pH 7.5, 150 mM NaCl, 2 mM EDTA pH 8 and 0.5% IGEPAL-CA630). Eight microlitres replication reaction was quenched into 400 μl of pulldown buffer and stored on ice. Samples were then incubated with 10 μl of LacI-coated streptavidin Dynabeads at 4°C for 30 min on a rotating wheel. The beads were washed thrice with wash buffer 2 (20 mM Tris–HCl [pH 7.5], 1.5 M NaCl, 2 mM EDTA and 0.5% IGEPAL-CA630) and then twice with Benzonase equilibration buffer (20 mM Tris–HCl [pH 7.5], 20 mM NaCl, 2 mM $MgCl_2$ and 0.02% Tween-20). Beads were then suspended in 7.5 μl of benzonase buffer (20 mM Tris–HCl [pH 8.0], 20 mM NaCl, 2 mM $MgCl_2$ and 0.02% Tween-20) containing 12 U benzonase and incubated for 1 h at 37°C. 7.5 μl of 2 × Laemmli loading buffer was added and the samples were incubated at 95°C for 5 min. The supernatants were collected and resolved on a 10% Criterion TGX precast midi protein gel (Bio-Rad), and transferred to polyvinyl difluoride membranes (Thermo Scientific). Membranes were blocked with 5% dried milk in PBST for 60 min at room temperature, rinsed several times with PBST and incubated with primary antibodies, as indicated, overnight at 4°C. Endogenous and rHMCES-3xFlag proteins were detected by immunoblotting with affinity-purified HMCES SRAP domain antiserum (1:5,000 dilution in PBST; Pocono Rabbit Farm and Laboratory rabbit 38,389). SPRTN was detected using SPRTN antiserum (Pocono Rabbit Farm and Laboratory rabbit 31053; 1:5,000 dilution in PBST). REV1 was detected using REV1 C-terminus antiserum (Pocono Rabbit Farm and Laboratory rabbit 1010; 1:5,000 dilution in PBST). Membranes were washed thrice with PBST for 10 min at room temperature, incubated with goat anti-rabbit secondary antibody peroxidase conjugate (1:20,000 dilution in PBST; Jackson ImmunoResearch Laboratories Inc., 111-035-003) for 30 min at room temperature, and then washed thrice with PBST for 10 min at room temperature. Blots were imaged with chemiluminescence using SuperSignal West Dura Extended Duration Substrate (Thermo Scientific) using a Bio-Rad ChemiDoc Imaging System (user interface v.2.4) and analysed using Image Lab v.6.4.0 (Bio-Rad).

Quantification and statistical analysis

Statistical details of each experiment (including the exact value of n, what n represents and precision measures) can be found in the figure legends.

Resource availability

Materials availability

All plasmids and cell lines are available on request from the corresponding author.

Data availability

Original gel images of all main figures are provided as source data. Original gel images of extended view figures will be shared by the corresponding author upon request. This study did not generate original code. Any additional information required to reanalyse the data reported in this paper is available from the corresponding author upon request.

Expanded View for this article is available online.

Acknowledgements

We thank T. Fröhlich for mass spectrometry analyses of recombinant HMCES protein and T. Mackens-Kiani for help with preparing Fig 1C. Research in the lab of DRS is supported by NIH grant no. GM129422 and an award from the Shurl and Kay Curci Foundation. Research in the lab of JS is supported by the European Research Council under the European Union's Horizon 2020 research and innovation programme (grant agreement number 801750), by the Alfried Krupp Prize for Young University Teachers awarded by the Alfried Krupp von Bohlen und Halbach Foundation, the European Molecular Biology Organization (YIP4644), the Vallee Foundation and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) (Project ID 213249687—SFB 1064 and Project-ID 393547839—SFB 1361). NIH grant GM118129 to Peter M. Burgers supported RB-B. Open Access funding enabled and organized by Projekt DEAL. Open Access funding enabled and organized by Projekt DEAL.

Author contributions

Maximilian Donsbach: Conceptualization; investigation; writing – original draft; writing – review and editing. **Sophie Dürauer:** Conceptualization; investigation; writing – original draft; writing – review and editing. **Florian Grünert:** Investigation; writing – review and editing. **Kha T Nguyen:** Investigation; writing – review and editing. **Richa Nigam:** Investigation; writing – review and editing. **Denitsa Yaneva:** Investigation; writing – review and editing. **Pedro Weickert:** Investigation; writing – review and editing. **Rachel Bezalel-Buch:** Resources. **Daniel R Semlow:** Supervision; funding acquisition; investigation; writing – review and editing. **Julian Stingele:** Conceptualization; supervision; funding acquisition; writing – original draft; project administration; writing – review and editing.

Disclosure and competing interests statement

The authors declare that they have no conflict of interest.

References

Abbotts R, Wilson DM 3rd (2017) Coordination of DNA single strand break repair. *Free Radic Biol Med* 107: 228–244

Amidon KM, Eichman BF (2020) Structural biology of DNA abasic site protection by SRAP proteins. *DNA Repair (Amst)* 94: 102903

Bhargava V, Goldstein CD, Russell L, Xu L, Ahmed M, Li W, Casey A, Servage K, Kollipara R, Picciarelli Z *et al* (2020) GCNA preserves genome integrity and fertility across species. *Dev Cell* 52: 38–52

Biayna J, Garcia-Cao I, Alvarez MM, Salvadores M, Espinosa-Carrasco J, McCullough M, Supek F, Stracker TH (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

Borgermann N, Ackermann L, Schwertman P, Hendriks IA, Thijssen K, Liu JC, Lans H, Nielsen ML, Mailand N (2019) SUMOylation promotes protective responses to DNA-protein crosslinks. *EMBO J* 38: e101496

Catanzariti AM, Soboleva TA, Jans DA, Board PG, Baker RT (2004) An efficient system for high-level expression and easy purification of authentic recombinant proteins. *Protein Sci* 13: 1331–1339

Dewar JM, Budzowska M, Walter JC (2015) The mechanism of DNA replication termination in vertebrates. *Nature* 525: 345–350

Dokshin GA, Davis GM, Sawle AD, Eldridge MD, Nicholls PK, Gourley TE, Romer KA, Molesworth LW, Tatnell HR, Ozturk AR *et al* (2020) GCNA interacts with Spartan and Topoisomerase II to regulate genome stability. *Dev Cell* 52: 53–68

Duxin JP, Dewar JM, Yardimci H, Walter JC (2014) Repair of a DNA-protein crosslink by replication-coupled proteolysis. *Cell* 159: 346–357

Halabelian L, Ravichandran M, Li Y, Zeng H, Rao A, Aravind L, Arrowsmith CH (2019) Structural basis of HMCES interactions with abasic DNA and multivalent substrate recognition. *Nat Struct Mol Biol* 26: 607–612

Kochenova OV, Bezalel-Buch R, Tran P, Makarova AV, Chabes A, Burgers PM, Shcherbakova PV (2017) Yeast DNA polymerase zeta maintains consistent activity and mutagenicity across a wide range of physiological dNTP concentrations. *Nucleic Acids Res* 45: 1200–1218

Kojima Y, Machida Y, Palani S, Caulfield TR, Radisky ES, Kaufmann SH, Machida YJ (2020) FAM111A protects replication forks from protein obstacles via its trypsin-like domain. *Nat Commun* 11: 1318

Krokan HE, Bjoras M (2013) Base excision repair. *Cold Spring Harb Perspect Biol* 5: a012583

Larsen NB, Gao AO, Sparks JL, Gallina I, Wu RA, Mann M, Raschle M, Walter JC, Duxin JP (2019) Replication-coupled DNA-protein crosslink repair by

SPRTN and the proteasome in *Xenopus* egg extracts. *Mol Cell* 73: 574 – 588

Lessel D, Vaz B, Halder S, Lockhart PJ, Marinovic-Terzic I, Lopez-Mosqueda J, Philipp M, Sim JC, Smith KR, Oehler J *et al* (2014) Mutations in SPRTN cause early onset hepatocellular carcinoma, genomic instability and progeroid features. *Nat Genet* 46: 1239 – 1244

Maskey RS, Kim MS, Baker DJ, Childs B, Malureanu LA, Jeganathan KB, Machida Y, van Deursen JM, Machida YJ (2014) Spartan deficiency causes genomic instability and progeroid phenotypes. *Nat Commun* 5: 5744

Maskey RS, Flatten KS, Sieben CJ, Peterson KL, Baker DJ, Nam HJ, Kim MS, Smyrk TC, Kojima Y, Machida Y *et al* (2017) Spartan deficiency causes accumulation of Topoisomerase 1 cleavage complexes and tumorigenesis. *Nucleic Acids Res* 45: 4564 – 4576

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

Mohni KN, Wessel SR, Zhao R, Wojciechowski AC, Luzwick JW, Layden H, Eichman BF, Thompson PS, Mehta KPM, Cortez D (2019) HMCES maintains genome integrity by shielding abasic sites in single-strand DNA. *Cell* 176: 144 – 153

Paulin KA, Cortez D, Eichman BF (2022) 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

Price NE, Johnson KM, Wang J, Fekry MI, Wang Y, Gates KS (2014) Interstrand DNA-DNA cross-link formation between adenine residues and abasic sites in duplex DNA. *J Am Chem Soc* 136: 3483 – 3490

Reinking HK, Kang HS, Gotz MJ, Li HY, Kieser A, Zhao S, Acampora AC, Weickert P, Fessler E, Jae LT *et al* (2020) DNA structure-specific cleavage of DNA-protein crosslinks by the SPRTN protease. *Mol Cell* 80: 102 – 113

Semlow DR, Zhang J, Budzowska M, Drohat AC, Walter JC (2016) Replication-dependent unhooking of DNA interstrand cross-links by the NEIL3 glycosylase. *Cell* 167: 498 – 511

Semlow DR, MacKrell VA, Walter JC (2022) The HMCES DNA-protein cross-link functions as an intermediate in DNA interstrand cross-link repair. *Nat Struct Mol Biol* 29: 451 – 462

Serbyn N, Noireterre A, Bagdiul I, Plank M, Michel AH, Loewith R, Kornmann B, Stutz F (2020) The aspartic protease Ddi1 contributes to DNA-protein crosslink repair in yeast. *Mol Cell* 77: 1066 – 1079

Srivastava M, Su D, Zhang H, Chen Z, Tang M, Nie L, Chen J (2020) HMCES safeguards replication from oxidative stress and ensures error-free repair. *EMBO Rep* 21: e49123

Stingele J, Schwarz MS, Bloemeke N, Wolf PG, Jentsch S (2014) A DNA-dependent protease involved in DNA-protein crosslink repair. *Cell* 158: 327 – 338

Stingele J, Bellelli R, Alte F, Hewitt G, Sarek G, Maslen SL, Tsutakawa SE, Borg A, Kjaer S, Tainer JA *et al* (2016) Mechanism and regulation of DNA-protein crosslink repair by the DNA-dependent metalloprotease SPRTN. *Mol Cell* 64: 688 – 703

Stingele J, Bellelli R, Boulton SJ (2017) Mechanisms of DNA-protein crosslink repair. *Nat Rev Mol Cell Biol* 18: 563 – 573

Thompson PS, Cortez D (2020) New insights into abasic site repair and tolerance. *DNA Repair (Amst)* 90: 102866

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 – 618

Tirman S, Quinet A, Wood M, Meroni A, Cybulla E, Jackson J, Pegoraro S, Simoneau A, Zou L, Vindigni A (2021) Temporally distinct post-replicative repair mechanisms fill PRIMPOL-dependent ssDNA gaps in human cells. *Mol Cell* 81: 4026 – 4040

Tretyakova NY, Groehler A 4th, Ji S (2015) DNA-protein cross-links: formation, structural identities, and biological outcomes. *Acc Chem Res* 48: 1631 – 1644

Vaz B, Popovic M, Newman JA, Fielden J, Aitkenhead H, Halder S, Singh AN, Vendrell I, Fischer R, Torrecilla I *et al* (2016) Metalloprotease SPRTN/DVC1 orchestrates replication-coupled DNA-protein crosslink repair. *Mol Cell* 64: 704 – 719

Wang N, Bao H, Chen L, Liu Y, Li Y, Wu B, Huang H (2019) Molecular basis of abasic site sensing in single-stranded DNA by the SRAP domain of *E. coli* yedK. *Nucleic Acids Res* 47: 10388 – 10399

Weickert P, Stingele J (2022) DNA-protein crosslinks and their resolution. *Annu Rev Biochem* 91: 157 – 181

Wilson DM 3rd, Takeshita M, Grollman AP, Demple B (1995) Incision activity of human apurinic endonuclease (Ape) at abasic site analogs in DNA. *J Biol Chem* 270: 16002 – 16007

Wu L, Shukla V, Yadavalli AD, Dinesh RK, Xu D, Rao A, Schatz DG (2022) HMCES protects immunoglobulin genes specifically from deletions during somatic hypermutation. *Genes Dev* 36: 433 – 450

Yaneva D, Sparks JL, Donsbach M, Zhao S, Weickert P, Bezalel-Buch R, Stingele J, Walter JC (2023) The FANCJ helicase unfolds DNA-protein crosslinks to promote their repair. *Mol Cell* 83: 43 – 56

Expanded View Figures

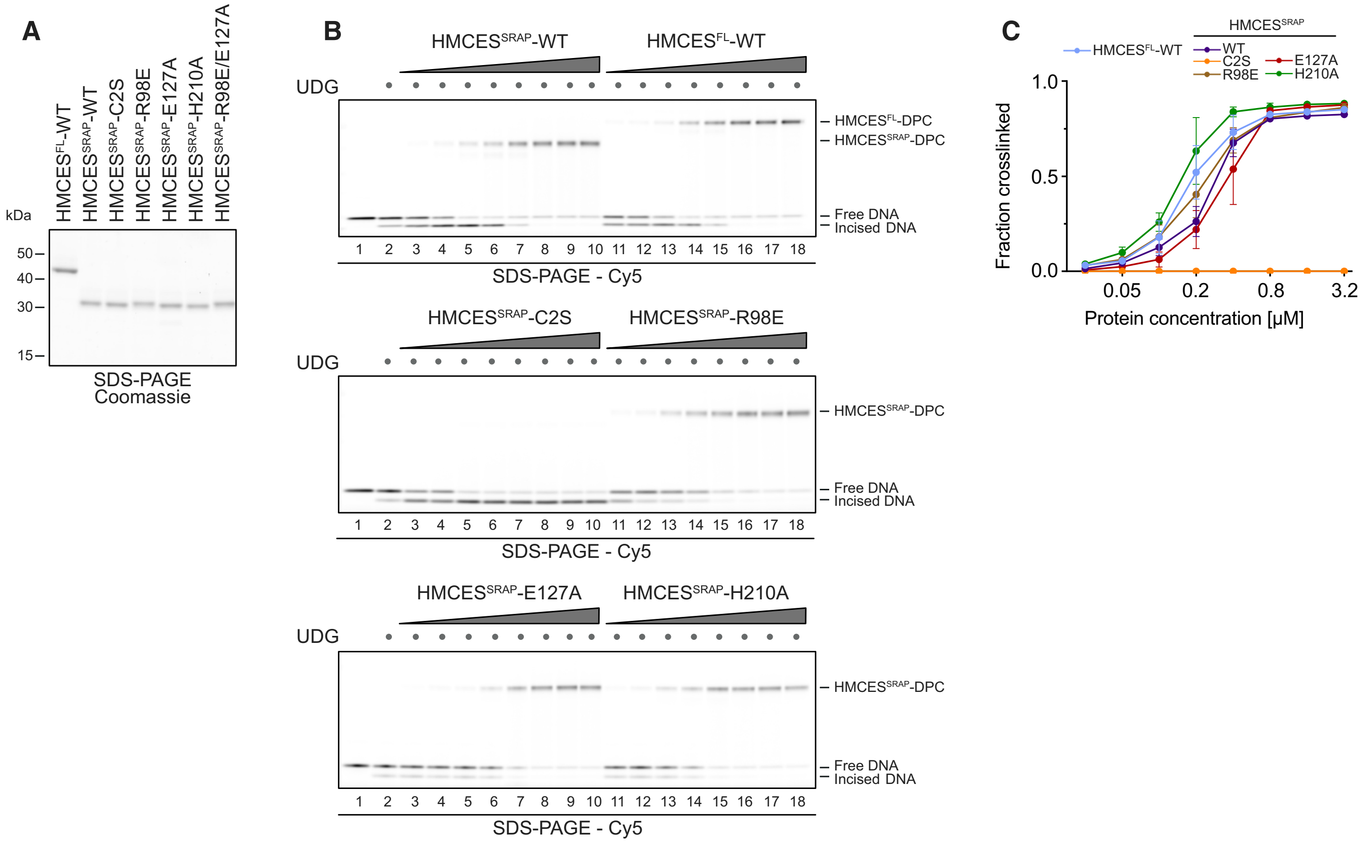


Figure EV1. HMCES-DNA-protein crosslinks are reversible. Related to Fig 1.

- A Coomassie-stained SDS–PAGE gel showing recombinant purified human HMCES^{FL}-WT, HMCES^{SRAP}-WT, HMCES^{SRAP}-C2S, HMCES^{SRAP}-R98E, HMCES^{SRAP}-E127A, HMCES^{SRAP}-H210A and HMCES^{SRAP}-R98E/E127A proteins used in this study.
- B DPC formation of HMCES^{FL} and HMCES^{SRAP} (WT or indicated variants). dU-containing DNA (0.1 μM) was incubated alone or with UDG and increasing concentrations of HMCES (0.025, 0.05, 0.1, 0.2, 0.4, 0.8, 1.6 and 3.2 μM), as indicated for 1 h at 37°C prior to analysis by denaturing SDS–PAGE.
- C Quantification of HMCES-DPC formation assays shown in (B).

Data information: Data in (C) represent the mean of three individual experiments ± SD.

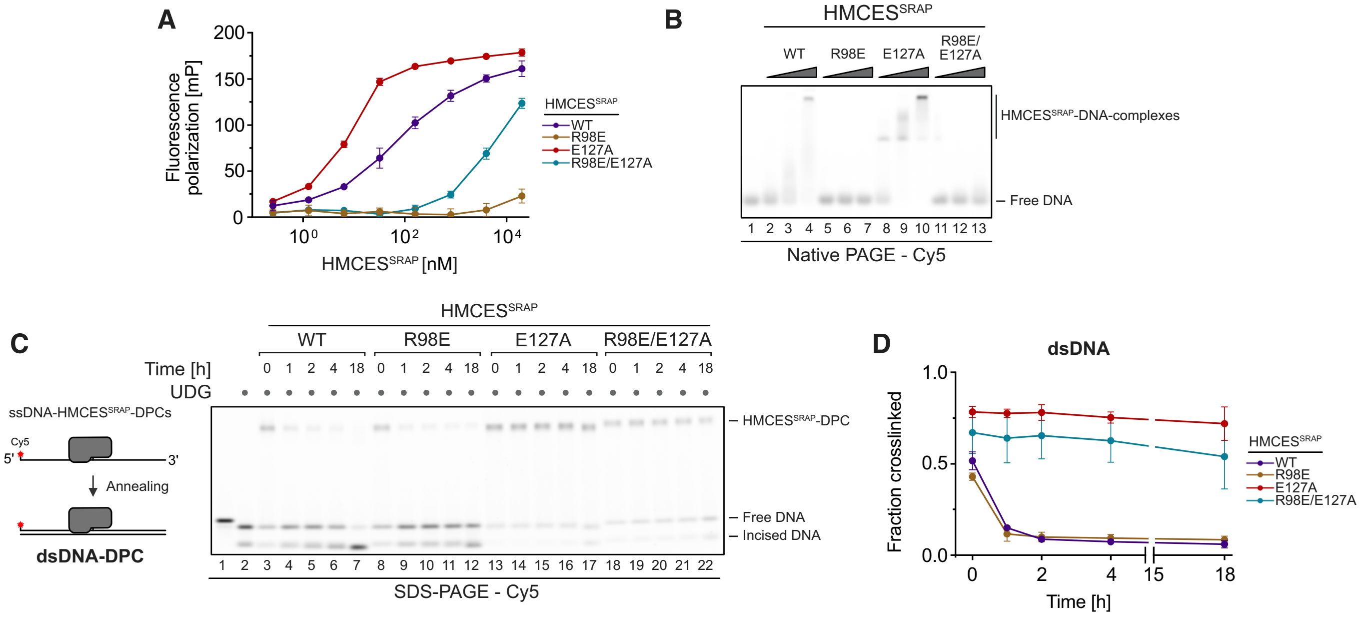


Figure EV2. Release of HMCES-DPCs is determined by DNA context. Related to Fig 3.

A Fluorescence polarization measurements of Cy5-labelled ssDNA (25 nM) incubated with increasing concentrations of HMCES^{SRAP}-WT, HMCES^{SRAP}-R98E, HMCES^{SRAP}-E127A or HMCES^{SRAP}-R98E/E127A for 20 min on ice prior to measuring fluorescence polarization.

B Non-covalent DNA binding of indicated HMCES^{SRAP} variants was assessed by electrophoretic mobility shift assay. A Cy5-labelled 30-mer ssDNA (0.1 μM) was incubated with HMCES^{SRAP} (0, 0.125, 0.5 or 2 μM) for 20 min at 4°C prior to analysis by native PAGE.

C DPC reversal kinetics of indicated HMCES^{SRAP} variants in dsDNA. A corresponding reverse oligonucleotide was annealed to HMCES^{SRAP}-DPCs, before incubation for the indicated amount of time at 37°C prior to separation by denaturing SDS–PAGE.

D Quantification of DPC reversal assays shown in (C).

Data information: Data in (A) and (D) represent the mean of three independent experiments ± SD.

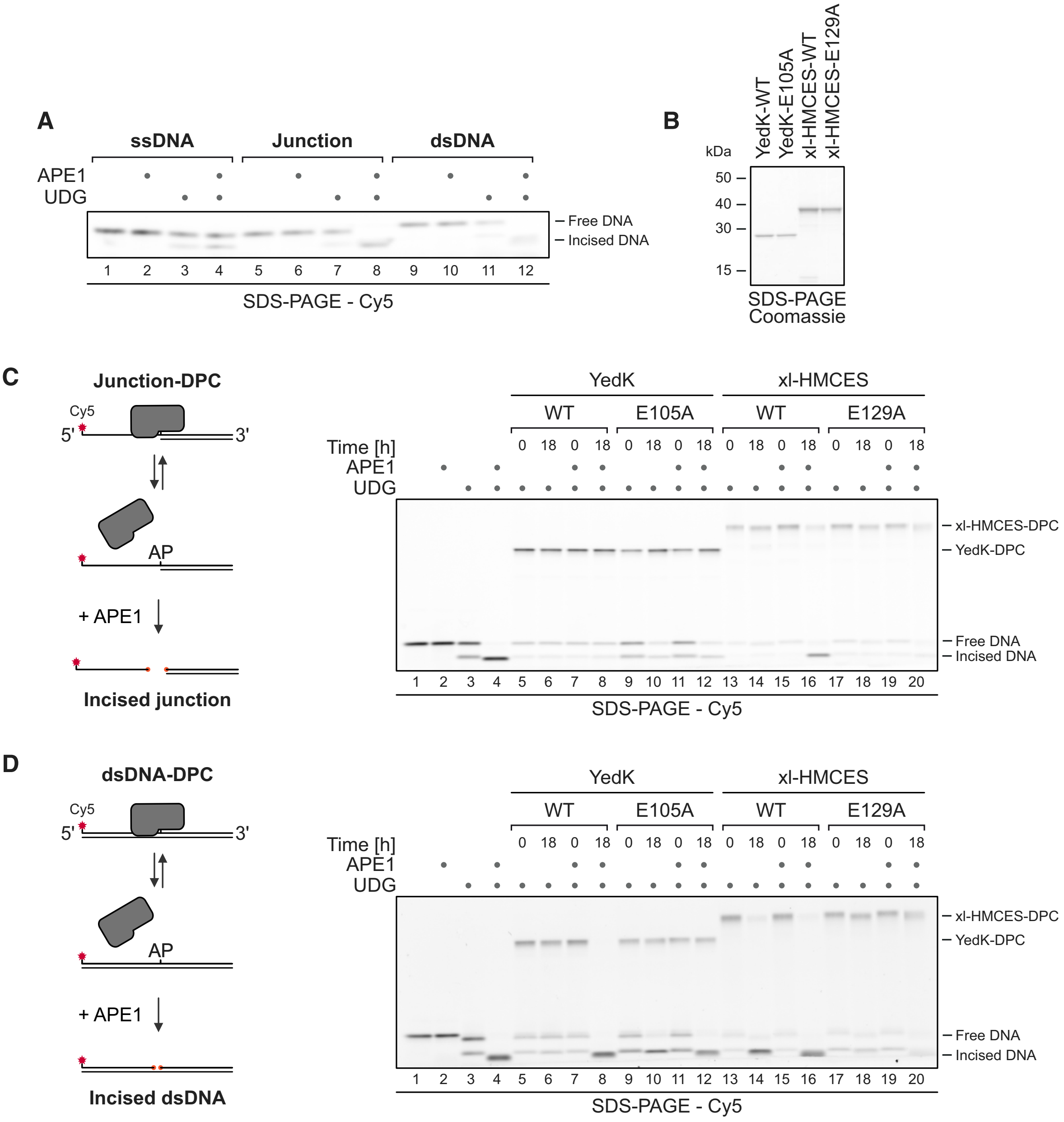


Figure EV3. Auto-release of HMCES-DPCs restricts crosslink formation to physiologically relevant situations. Related to Fig 5.

A APE1 incision of an AP site in ssDNA, DNA junction and dsDNA. A Cy5-labelled 30-mer ssDNA was incubated alone or with UDG for 1 h at 37°C. Corresponding reverse oligonucleotides for DNA junction or dsDNA were annealed (for ssDNA, a non-complementary oligonucleotide was added). Next, samples were incubated with APE1 for 18 h at 37°C before separation by denaturing SDS–PAGE.

B Coomassie-stained SDS–PAGE gel showing recombinant purified bacterial YedK-WT, YedK-E105A, xl-HMCES-WT and xl-HMCES-E129A proteins used in this study.

C, D APE1 incision of an AP site protected by the indicated YedK-WT-DPC or YedK-E105A-DPC and xl-HMCES-WT-DPC or xl-HMCES-E129A-DPC at ssDNA-dsDNA junctions (C) or within dsDNA (D). Free dU-containing DNA was incubated alone or in the presence of UDG and YedK/xl-HMCES for 1 h at 37°C. Next, corresponding reverse oligonucleotides were annealed to generate an ssDNA-dsDNA junction (C) or dsDNA (D), and reactions were incubated alone or with APE1 for the indicated amount of time at 37°C prior to separation by denaturing SDS–PAGE.

Figure EV4. Resolution of HMCES-DPCs during replication-coupled ICL repair. Related to Fig 5.

A Schematic depiction of the NEIL3-dependent repair of an AP-ICL, a lesion that forms when an AP site reacts with a nucleobase of the opposing DNA strand forming a covalent crosslink (Price *et al*, 2014). In *Xenopus* egg extracts, such crosslinks are primarily unhooked by the NEIL3 glycosylase (Semlow *et al*, 2016), which yields an AP site leading to formation of an HMCES-DPC.

B–F In the absence of SPRTN, the intact HMCES-DPC is presumably bypassed by TLS and transferred into dsDNA. To test whether this triggers autorelease, we analysed the stability of DPCs formed by wild-type and E129A-mutated *Xenopus laevis* rHMCES-3xFlag proteins during ICL repair in egg extract. pICL-*lacO*^{AP} was replicated in mock- or SPRTN-depleted extracts (B) supplemented with WT or E129A rHMCES-3xFlag. At the time points indicated, plasmid was recovered under stringent conditions, the DNA was digested and released proteins were separated by SDS–PAGE. HMCES-DPCs were detected using an antibody raised against the SRAP domain that permits simultaneous monitoring of endogenous HMCES protein and the recombinant 3xFlag-tagged HMCES (which migrates slower during SDS–PAGE due to the 3xFlag). In this experimental setup, the endogenous protein serves as a control for the effects of SPRTN depletion and autorelease. Like the endogenous HMCES (C), both WT (D) and E129A-mutated rHMCES-3xFlag (E) were stabilized by SPRTN depletion, implying that proteolysis is the dominant mechanism for removing HMCES-DPC under these conditions. However, it was challenging to assess the relative behaviour of tagged WT and mutant protein because DPCs formed by the wild-type recombinant protein (like those formed by the endogenous protein) are resolved slowly in SPRTN-depleted extract (on the timescale of hours, somewhat slower than the timescale for observed for reversal *in vitro*). Additionally, the E129A-mutated recombinant flag-tagged protein crosslinked less efficiently than endogenous HMCES, making it difficult to detect even when present in large excess (F). We are, therefore, unable to determine from these data whether HMCES-DPC reversal occurs during ICL repair in egg extract under the conditions tested. Orange dots denote non-specific bands or bands corresponding to contaminating IgG.

G One explanation for the discrepancy in the degree of reversal observed between the *in vitro* reconstitution and egg extract systems could be that the *in vitro* reactions were all performed at 37°C, while replication in egg extracts must be performed at 20°C. Therefore, we assessed reversal of HMCES^{SRAP}-WT or -E127A-DPCs in dsDNA at the indicated temperatures for the indicated amount of time before analysis by denaturing SDS–PAGE. Indeed, autocatalytic reversal was significantly delayed at 20°C.

H Quantification of DPC reversal assays using HMCES^{SRAP}-WT and -E127A shown in (G).

I The extracts used in the replication reactions shown in (J and K) were immunoblotted for SPRTN, Rev1 and HMCES.

J As an alternative additional strategy to determine whether reversal contributes to HMCES-DPC resolution, we tested whether REV1 depletion results in stabilization of HMCES-DPCs, reasoning that blocking TLS (and transfer of the DPC into dsDNA) may inhibit reversal. pICL-*lacO*^{AP} was replicated in mock-, REV1-, SPRTN- or REV1- and SPRTN-depleted egg extracts, as indicated. At the indicated time points, plasmid was recovered under stringent conditions, the DNA was digested and released proteins were separated by SDS–PAGE. HMCES was detected by blotting. As expected, depletion of SPRTN alone resulted in a strong stabilization of HMCES-DPCs. Depletion of REV1 alone did not stabilize HMCES-DPCs, consistent with our data indicating that SPRTN represents the dominant mechanism for HMCES-DPC resolution in egg extract. Surprisingly, when combined with SPRTN depletion, REV1 depletion partially suppressed the accumulation of HMCES-DPCs. Superficially, this result is contrary to our expectations based on data presented in Fig 6. However, we interpret the result to indicate when the HMCES-DPC is maintained at an ssDNA/dsDNA junction due to inefficient TLS, residual SPRTN or another protease can eventually degrade the HMCES-DPC. Therefore, while these data do not provide evidence for HMCES-DPC reversal during ICL repair in egg extract, they do reinforce the need for alternative removal mechanisms for HMCES-DPCs present in dsDNA that are refractory to proteolysis.

K In parallel with the reactions shown in (J), pICL-*lacO*^{AP} was replicated in the indicated egg extracts supplemented with [α-^{32}P]dCTP. Replication intermediates were separated on a native agarose gel and visualized by autoradiography. SC, supercoiled. OC, open circular. Consistent with a TLS defect upon Rev1 depletion, we observed accumulation of gapped, circular plasmids in replication gels, implying that the HMCES-DPC is maintained at an ssDNA-dsDNA junction.

Data information: Data in (H) represent the mean of three independent experiments ± SD.

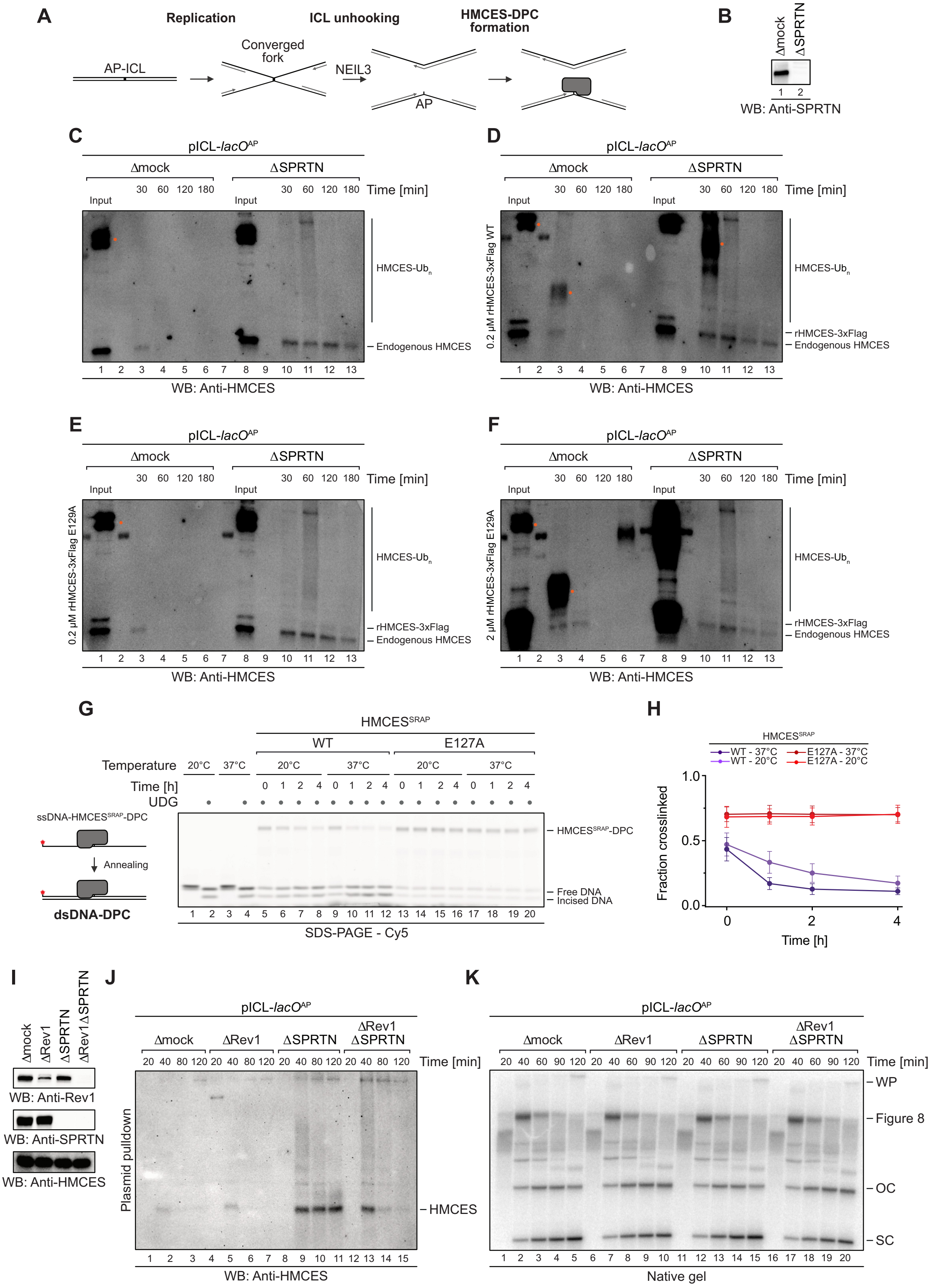

A
Replication
ICL unhooking
HMCES-DPC formation
AP-ICL
Converged fork
NEIL3
AP
B
Δmock
ΔSPRTN
1 2
WB: Anti-SPRTN
C
pICL-lacO^AP
Δmock
ΔSPRTN
Input 30 60 120 180 Time [min]
HMCES-Ub_n
Endogenous HMCES
1 2 3 4 5 6 7 8 9 10 11 12 13
WB: Anti-HMCES
D
0.2 µM rHMCES-3xFlag WT
rHMCES-3xFlag
Endogenous HMCES
E
0.2 µM rHMCES-3xFlag E129A
F
2 µM rHMCES-3xFlag E129A
G
HMCES^SRAP
WT
E127A
Temperature 20°C 37°C
Time [h] 0 1 2 4
UDG
ssDNA-HMCES^SRAP-DPC
Annealing
dsDNA-DPC
HMCES^SRAP-DPC
Free DNA
Incised DNA
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
SDS-PAGE - Cy5
H
HMCES^SRAP
WT - 37°C
WT - 20°C
E127A - 37°C
E127A - 20°C
Fraction crosslinked
1.0
0.5
0.0
0 2 4
Time [h]
I
Δmock
ΔRev1
ΔSPRTN
ΔRev1ΔSPRTN
WB: Anti-Rev1
WB: Anti-SPRTN
WB: Anti-HMCES
J
pICL-lacO^AP
Δmock
ΔRev1
ΔSPRTN
ΔRev1 ΔSPRTN
20 40 80 120 Time [min]
Plasmid pulldown
HMCES
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
WB: Anti-HMCES
K
20 40 60 90 120 Time [min]
WP
Figure 8
OC
SC
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Native gel

Figure EV4.

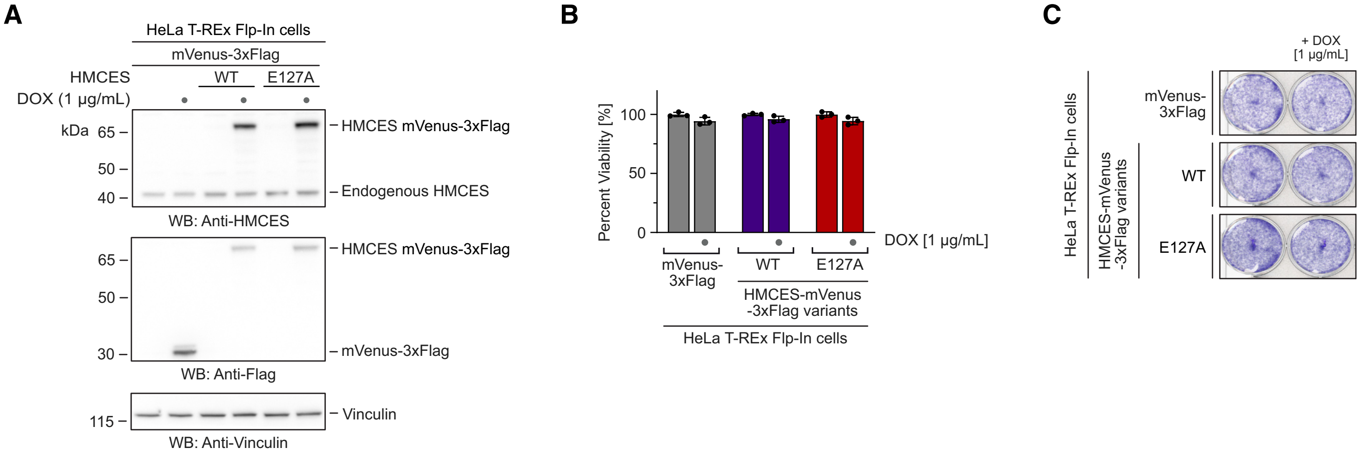


Figure EV5. Cell viability is not affected upon overexpression of reversal-defective HMCES variants. Related to Fig 7.

A–C HeLa T-REx Flp-In cells expressing the indicated doxycycline-inducible HMCES variants with a C-terminal mVenus-3xFlag-tag were grown in the presence of 1 μg/ml doxycycline, as indicated. Expression levels were analysed by Western blotting (A). Cell viability was determined using AlamarBlue cell viability assay (B), or crystal violet staining (C).

Data information: Data in (B) represent the mean ± SD of three technical replicates normalized to the mean of untreated control (mVenus-DOX).

3.2 Electro-elution-based purification of covalent DNA–protein cross-links

Contribution report

This publication provides a comprehensive, step by step protocol for the purification of x-linked proteins (PxP). The experiments shown were carried out by Pedro Weickert and myself, while photos and videos were taken by Hao-Yi Li and me. Maximilian J. Götz contributed the mass spectrometry analysis section, including a practical example for data analysis. Manuscript writing, editing and figure preparation were done by Pedro Weickert, Julian Stingele and me with input from all authors.

Summary

Covalent DNA-protein cross-links (DPCs) are pervasive DNA lesions that challenge genome stability and can be induced by metabolic or chemotherapeutic cross-linking agents including reactive aldehydes, topoisomerase poisons and DNMT1 inhibitors. The purification of x-linked proteins (PxP), where DNA cross-linked proteins are separated from soluble proteins via electro-elution, can be used to identify DPCs. Here we describe a versatile and sensitive strategy for PxP. Mammalian cells are collected following exposure to a DPC-inducing agent, embedded in low-melt agarose plugs and lysed under denaturing conditions. Following lysis, the soluble proteins are extracted from the agarose plug by electro-elution, while genomic DNA and cross-linked proteins are retained in the plug. The cross-linked proteins can then be analysed by standard analytical techniques such as sodium dodecyl-sulphate polyacrylamide gel electrophoresis 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 overcomes the limitations of alternative methods to analyse DPCs that rely on precipitation as the separating principle and can be performed by users trained in molecular or cell biology within 2–3 days. The protocol has been optimised to study DPC induction and repair in mammalian cells but may also be adapted to other sample types including bacteria, yeast and tissue samples.

nature protocols

https://doi.org/10.1038/s41596-024-01004-z

Protocol

Check for updates

Electro-elution-based purification of covalent DNA–protein cross-links

Pedro Weickert[1,2,3], **Sophie Dürauer**[1,2,3], **Maximilian J. Götz**[1,2], **Hao-Yi Li**[1,2] & **Julian Stingele**[1,2]

Abstract

Covalent DNA–protein cross-links (DPCs) are pervasive DNA lesions that challenge genome stability and can be induced by metabolic or chemotherapeutic cross-linking agents including reactive aldehydes, topoisomerase poisons and DNMT1 inhibitors. The purification of x-linked proteins (PxP), where DNA–cross-linked proteins are separated from soluble proteins via electro-elution, can be used to identify DPCs. Here we describe a versatile and sensitive strategy for PxP. Mammalian cells are collected following exposure to a DPC-inducing agent, embedded in low-melt agarose plugs and lysed under denaturing conditions. Following lysis, the soluble proteins are extracted from the agarose plug by electro-elution, while genomic DNA and cross-linked proteins are retained in the plug. The cross-linked proteins can then be analyzed by standard analytical techniques such as sodium dodecyl-sulfate–polyacrylamide gel electrophoresis 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 overcomes the limitations of alternative methods to analyze DPCs that rely on precipitation as the separating principle and can be performed by users trained in molecular or cell biology within 2–3 d. 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.

Key points

- The assay separates proteins in agarose-embedded cell lysates based on their propensity to move through an electric field. Proteins that are covalently cross-linked to DNA are not eluted from the agarose plugs, which can then be analyzed using mass spectrometry or antibody-based staining techniques.
- Isolation of cross-linked proteins by electro-elution provides an alternative over purification methods that rely on the coprecipitation of DNA and proteins.

Key references

Weickert, P. et al. *Nat. Commun.* **14**, 352 (2023): https://doi.org/10.1038/s41467-023-35988-1

Carnie, C. J. et al. *Nat. Cell Biol.* **26**, 797–810 (2024): https://doi.org/10.1038/s41556-024-01391-1

---

[1]Department of Biochemistry, Ludwig-Maximilians-Universität München, Munich, Germany. [2]Gene Center, Ludwig-Maximilians-Universität München, Munich, Germany. [3]These authors contributed equally: Pedro Weickert, Sophie Dürauer. ✉e-mail: stingele@genzentrum.lmu.de

Introduction

DNA–protein cross-links (DPCs) are highly toxic DNA lesions that have emerged as important sources of genome instability[1]. DPCs can be classified as nonenzymatic or enzymatic[2]. Nonenzymatic DPC formation is induced by bifunctional cross-linking 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 analog), 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 cross-links to abasic (AP) sites within single-stranded DNA (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 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–double-stranded DNA 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 (refs. 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 cross-linked DNA or proteins. In the KCl– sodium dodecyl-sulfate (SDS) assay[36], proteins are precipitated from denaturing lysates and quantification of coprecipitating cross-linked DNA indicates the extent of DPC formation. In contrast, the rapid approach to DNA adduct recovery (RADAR) assay[37] and several variations[38,39], are based on the precipitation of DNA and the identification of coprecipitating cross-linked proteins.

Development of the PxP protocol

The identification of DPCs by purification of x-linked proteins (PxP) enables the separation of cross-linked proteins from soluble proteins using electro-elution. 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 cross-linked 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 analyze DPC formation in cells. PxP is thus based on the idea that cross-linked proteins cannot be separated from immobilized DNA in an electric field, while other proteins (including DNA-bound ones) 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 cross-linked proteins due to hydrolysis. Additionally, we tested various parameters (e.g., varying number of cells, agarose concentrations) to enable efficient electro-elution of non-cross-linked proteins. Finally, we explored different options to retrieve cross-linked proteins from plugs following electro-elution, including agarose 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].

Overview of the PxP protocol

The protocol is organized in three main stages describing (1) the 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 cross-links chromatin-associated proteins to DNA with the majority of DPCs being formed by histone proteins[28]
    - Induction of DPCs using the topoisomerase II poison etoposide (TOP2–DPCs). Etoposide forms a stable ternary complex at the topoisomerase–DNA interface and stabilizes the otherwise reversable covalent bond between the tyrosines of the TOP2 active site and DNA[7]. 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 postreplicative chromatin[25]
    - Induction of DPCs between HMCES and AP sites using UVC irradiation or CD437 (refs. 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 the cell suspension 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–polyacrylamide gel electrophoresis (PAGE) gel resulting in electro-elution of non-cross-linked 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, the plugs are fixed followed by in-plug tryptic digestion and standard liquid chromatography tandem mass spectrometry (LC–MS/MS) analysis

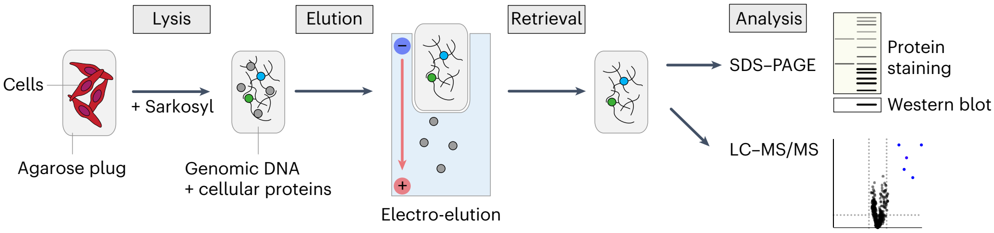


Fig. 1 | A schematic illustration of the PxP workflow. Overview of the PxP method. First, cells are mixed with low-melt agarose and cast in plugs using plug molds. The agarose plugs are then lysed in a denaturing buffer containing sarkosyl to disrupt the cell membrane and denature cellular proteins. 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 cross-linked 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).

Experimental design

This protocol is optimized for the analysis of DPCs in cultured adherent mammalian cells and we have successfully used it 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 the cross-links arise primarily during DNA replication. Additionally, it is critical to choose appropriate drug concentrations for DPC induction. It is important to avoid high drug doses that cause cells to die and detach before they can be collected for PxP. Moreover, the stability of the DPC must be considered when planning the experiment. For example, camptothecin and etoposide are noncompetitive inhibitors of TOP1 and TOP2, respectively. Once the compounds are removed from the culture media, topoisomerase DPCs immediately revert. Therefore, we recommend to place all culture dishes for topoisomerase poisons experiments on ice for 5 min before collecting to minimize cross-link reversal. Another consequence of the immediate reversal of topoisomerase DPCs is that it is not possible to wash out the drug and track the DPC repair process over time. In contrast, the fate of DNMT1-, HMCES- and formaldehyde-induced DPCs can be monitored by a pulse-chase assay where drug-free media is added after DPC induction. In pulse-chase assays, it is possible to include small molecule inhibitors, for example, the proteasome inhibitor MG132, during the chase to assess its effect on DPC repair. When conducting PxP experiments, we recommend using formaldehyde as a positive control; PxP samples obtained from formaldehyde-treated cells should reveal a characteristic histone pattern when analyzed by SDS–PAGE (see 'Anticipated results').

For most types of experiments, one plug 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 per plug is low or if the final samples need to be analyzed by western blotting with several antibodies. We recommend casting several plugs per experimental condition and pool them after electro-elution to increase the final amount of sample. PxP samples can also 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 labeling with amino acids in cell culture-based or tandem mass tag-based quantitation could be used.

No specific expertise is needed to perform the PxP protocol, beyond standard cell biology skills. At first, handling and transferring plugs can be challenging, but it typically only requires some practice. To analyze 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.

Comparison with other methods and limitations

The KCl–SDS assay[36] is a sensitive method to detect DPC formation but cannot be used to study the identity of the cross-linked proteins. The RADAR assay[37] and various derivatives can be used to identify cross-linked proteins but rely on DNA precipitation to isolate coprecipitating DPCs. The reliance on precipitation is a potential risk because the behavior during precipitation is influenced by various features of the cross-linked 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-cross-linked proteins through electro-elution rather than directly purifying cross-linked proteins. In addition, we developed an optional nuclease control, which allows to distinguish between DPCs and copurified contaminants. For the nuclease control, DNA is digested within plugs before electro-elution, which causes cross-linked proteins to elute from the plug while unspecific contaminants remain.

Nonetheless, the PxP assay has limitations. First, DPCs induced by reversible inhibitors, such as camptothecin or etoposide, can be challenging to study, because cells must be collected and embedded in agarose before the denaturing lysis step. Therefore, there is the risk that such DPCs are lost due to reversal before lysis. This is not an issue in KCl–SDS or RADAR assays because the cells are lysed using a denaturing buffer directly in the culture dish. A second limitation of the PxP assay is that the number of cells that can be embedded in one 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.

Materials

Equipment

- Biological safety cabinets (e.g., Thermo Fisher Scientific, Herasafe 2030i, cat. no. 51032330)
- CO_2 incubator (e.g., PHC, cat. no. MCO-230AICUV)
- Water bath (e.g., Memmert, cat. no. WNE10)
- Inverted routine tissue culture microscope with fluorescence module (e.g., Nikon, ECLIPSE Ts2-FL, C-LED505 Epi-FL filter block, EX 496/29, DM 518, EM 543/37)
- Laboratory fume hood
- Cold room between 4 and 10 °C
- Centrifuge for 1.5-ml tubes (e.g., Eppendorf, centrifuge 5424R, cat. no. EP5404000138)
- Automated cell counter (e.g., Thermo Fisher Scientific, Countess II FL Automated, cat. no. AMQF1000)
- Countess cell counting chamber slides (Thermo Fisher Scientific, cat. no. C10228)
- Vacuum pump (e.g., Vacuubrand, cat. no. PC 3004 VARIO)
- Rotating shaker (e.g., Stuart, Rotators SB3, cat. no. 445-2101)
- Thermoblock (e.g., Eppendorf, ThermoMixer C, cat. no. 5382000015)
- Electrophoresis system for SDS–PAGE for Novex WedgeWell gels (e.g., Thermo Fisher Scientific, Mini Gel Tank, cat. no. A25977)
- Tissue culture dishes 6, 10 and 15 cm (Sarstedt, cat. nos. 422-83.3901, 422-83.3902 and 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. nos. 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 (e.g., VWR, cat. no. 216-4334)

(Optional) For SDS–PAGE

- Electrophoresis system for SDS–PAGE (e.g., Thermo Fisher Scientific XCell4 SureLock Midi-Cell Electrophoresis System, cat. no. WR0100)
- Protein blotting cell (e.g., Bio-Rad, Criterion blotter with plate electrodes, cat. no. 1704070)
- Immobilon-P membrane, PVDF, 0.45 µm (Merck Millipore, cat. no. IPVH00010)
- Western blot imaging system (e.g., Bio-Rad, ChemiDoc MP Imaging System)
- Scanner (e.g., Epson, Perfection V850 Pro, cat. no. B11B224401)
- Image analyzing software (e.g., ImageJ software)

(Optional) For UVC treatment

- UVC irradiator (e.g., Analytik Jena, UVP cross-linker CL-1000, cat. no. 849-30101-2)

(Optional) For mass spectrometry data analysis

- 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]

Biological materials

- HAP1 cells (Horizon, cat. no. C631, RRID: CVCL_Y019)
- HeLa T-REx Flp-In cells (originally obtained from The Francis Crick Institute, London, derivative of HeLa, RRID: CVCL_0030)
- HeLa T-REx Flp-In cells expressing doxycycline-inducible HMCES (wild type (WT) or a catalytically compromised C2S variant in which the catalytic cysteine is replaced by serine) with a C-terminal mVenus–3xFlag tag (derivative of HeLa, RRID: CVCL_0030)

Reagents

- Dulbecco's modified Eagle medium (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, 1× (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, cat. no. 11585916001)
- Bolt MOPS SDS running buffer (Invitrogen, cat. no. B0001-02)
- Lithium dodecyl sulfate (LDS) sample buffer, 4× (Thermo Fisher Scientific, cat. no. NP0007)
- Bolt reducing agent, 10× (Thermo Fisher Scientific, cat. no. B0009)
- Coomassie-based protein stain (e.g., GRP, Der Blaue Jonas, cat. no. GRP1)

(Optional) For DPC induction

- 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)

(Optional) For mass spectrometry sample preparation

- 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)

(Optional) For SDS–PAGE

- SDS–PAGE gels (e.g., 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, 20× (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.
- Enhanced chemiluminescence (ECL) substrate for western blot detection (e.g., Thermo Fisher Scientific, SuperSignal West Pico or Femto PLUS chemiluminescent substrate, cat. nos. 34577, 34095)

(Optional) For TCA precipitation

- TCA (Sigma, cat. no. T6399)
  ▲ **CAUTION** TCA can cause damage to eyes and skin. Handle in a fume hood, wearing protective gloves and discard according to local regulations.

(Optional) Primary antibodies used in this protocol

- Histone H3 (Cell Signaling Technology, cat. no. 4499S RRID: AB_10544537) (1:1,000 dilution in TBS-T containing 5% milk)
- Actin (Santa Cruz, cat. no. sc-47778, RRID: AB_626632) (1:1,000 dilution in TBS-T containing 2.5% BSA)
- Vinculin (Santa Cruz, cat. no. sc-73614, RRID: AB_1131294) (1:2,000 dilution in TBS-T containing 2.5% BSA)
- HMCES (Human Altas, cat. no. HPA044968, RRID: AB_2679160) (1:500 dilution in TBS-T containing 2.5% BSA)
- HMCES (Santa Cruz, cat. no. sc-514238, RRID: AB_2813859) (1:500 dilution in TBS-T containing 2.5% BSA)
- Flag-M2 (Sigma, cat. no. F3165, RRID: AB_259529) (1:2,000 dilution in TBS-T containing 2.5% BSA)
- TOP2 (Abcam, cat. no. ab109524, RRID: AB_10859793) (1:1,000 dilution in TBS-T containing 2.5% BSA)
- DNMT1 (Cell Signaling Technology, clone D63A, cat. no. 5032, RRID: AB_10548197) (1:1,000 dilution in TBS-T containing 2.5% BSA)
- GAPDH (Cell Signaling Technology, 14C10, cat. no. 2118, RRID: AB_561053) (1:2,000 dilution in TBS-T containing 2.5% BSA)

Reagent setup

Low-melt agarose solution

Final concentration: 2% low-melt agarose, 1× 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 1× 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. See the 'Troubleshooting' table.

▲ **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.

Lysis buffer (1 ml is required per agarose plug, prepare fresh)

Final concentration: 1× PBS, 2% sarkosyl, 0.5 mM EDTA, 1× 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 10× 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, 25 °C) and the final buffer at 4 °C. The buffer can be stored at 4 °C for at least 8 h.

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.

(Optional) Thymidine medium

Final concentration: 2 mM thymidine in standard culture medium, 10% FBS, for IMDM + 1% PSG. For 500 ml of thymidine medium, weigh 250 mg thymidine powder and transfer it to a 500 ml bottle of culture medium (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 medium by filtering through a 0.22 µm PES vacuum filter in a sterile tissue culture hood. 55 ml FBS is then added to reach a final concentration of 10%. For IMDM, 5 ml PSG is added as well at a final concentration of 1%. Thymidine medium 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 medium for a maximum of 1 month at 4 °C.

(Optional) Etoposide stock solution

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 two or three freeze–thaw cycles.

▲ **CAUTION** Etoposide is a toxic compound. Handle in a fume hood, wearing protective gloves and discard according to local regulations.

(Optional) Doxycycline stock solution

Final concentration: 50 mg/ml doxycycline in 1× PBS. Dissolve 1 g doxycycline hyclate in 20 ml 1× 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 1× PBS. This will yield a 1 mg/ml (1,000×) 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.

(Optional) 5-azadC stock solution

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 two or three freeze–thaw cycles.

▲ **CAUTION** 5-azadC is a toxic compound. Handle in a fume hood, wearing protective gloves and discard according to local regulations.

(Optional) CD437 stock solution

Final concentration: 10 mM CD437 in DMSO. Carefully resuspend 10 mg CD437 in 2.51 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 two or three freeze–thaw cycles.

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.

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.

Ammonium bicarbonate buffer (optional for mass spectrometry sample preparation)

Final concentration: 100 mM ammonium bicarbonate in deionized water; to be prepared fresh. 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: 1×, 10% methanol. For 2 l, add 50 ml commercial 20× 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: 1×. For 500 ml, mix 25 ml commercial 20× 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 10× 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 10× can be stored at RT for at least 2 months.

For 1 l of 1x TBS-T, dilute 100 ml 10× 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.

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 d (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.

TCA
Final concentration: 70% (wt/vol). 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

Stage 1: DPC induction

● **TIMING** 2–3 d

1. Induce covalent DPCs using a DPC-inducing agent of choice or with one of the four example procedures described in this protocol. Perform each experiment at least three times and use appropriate negative and positive controls for the chosen DPC-inducing agent (Table 1 and Boxes 1–4).

Table 1 | Agents used for DPC induction in this protocol

A	Formaldehyde	Induces DPCs by unspecific cross-linking of chromatin proteins	Box 1
B	Etoposide	Induces formation of TOP2–DPCs	Box 2
C	5-azadC (decitabine)	Induces formation of postreplicative DNMT1–DPCs	Box 3
D	UVC irradiation/POLα inhibition	Induce replication-coupled formation of HMCES–DPCs	Box 4

Stage 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 80 °C 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 45 °C.
5. Resuspend cell pellets collected in Stage 1 in ice-cold 1× PBS to a concentration of 25,000 cells/µl. In the examples described in Stage 1, you should have obtained ~2 million cells per condition, which would require resuspension with 80 µl of 1× PBS.

   ▲ **CRITICAL STEP** Resuspend cells thoroughly by pipetting up and down with a 20–200 µl pipette. 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 each aliquot to clean 1.5-ml plastic tubes containing 25 µl 4× LDS sample buffer, 10 µl 10× reducing agent and 55 µl water.
7. Boil samples, which will be used as input controls, in a thermoblock at 99 °C for 15–20 min.
8. 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.
9. Place a plug mold and a vortex next to the 45 °C thermoblock.

BOX 1

DPC induction using formaldehyde

● **TIMING** 2 d, handling 2–3 h

Procedure

1. Day 1: in the evening, prewarm medium (IMDM + 10% FBS + 1% PSG) in a water bath at 37 °C.
2. Collect HAP1 cells cultured in IMDM medium 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 1× PBS followed by trypsinization (trypLE, 1 ml). Once cells have detached, inactivate trypsin with 9 ml medium and transfer the cell suspension to a 15-ml conical tube.
3. 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.
4. Seed 1.5 million cells per condition in fresh 6-cm dishes, adding at least 3 ml medium. The example described here requires three 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.
5. Day 2: in the morning, prewarm medium (IMDM + 10% FBS + 1% PSG) in a water bath at 37 °C.
6. 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 Stage 2 of the main procedure (Step 5).
7. Dilute formaldehyde to 1 M with 1× PBS before adding it to the medium. This is done by mixing 100 µl 16% formaldehyde solution (5.3 M) with 430 µl 1× PBS (final concentration 1 M). After mixing, add 1 µl 1 M formaldehyde solution per ml medium (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 medium for a final concentration of 2 mM). For treating a 6-cm dish, a minimum of 2.5 ml medium is required.
8. Aspirate the medium from both dishes and add formaldehyde-containing medium to one dish. Add fresh medium without formaldehyde to the other dish as a nontreated control. Incubate for 1 h at 37 °C.
9. After incubation, place dishes on ice. Remove the medium, wash once with 3 ml 1× PBS, add 1 ml ice-cold 1× PBS and quickly scrape cells.
10. 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.

BOX 2

Etoposide treatment to induce TOP2–DPCs

● **TIMING 2 d, handling 2–3 h**

Procedure

1. Day 1: in the evening, prewarm medium (IMDM + 10% FBS + 1% PSG) in a water bath at 37 °C.
2. Collect HAP1 cells cultured in IMDM medium 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 1× PBS followed by trypsinization (trypLE, 1 ml). Once cells have detached, inactivate trypsin with 9 ml medium and transfer the cell suspension to a 15-ml conical tube.
3. 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.
4. Seed 1.5 million cells per condition in fresh 6-cm dishes, adding at least 3 ml medium. The example described here requires four dishes (one nontreated control, two etoposide concentrations and one extra dish for determining the number of cells).
5. Day 2: in the morning, prewarm medium (IMDM + 10% FBS + 1% PSG) in a water bath at 37 °C.
6. 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 Stage 2 of the main procedure (Step 5).
7. Add 5 or 10 µl etoposide stock solution (50 mM) to 10 ml medium to reach a final concentration of 25 µM and 50 µM, respectively. 2.5 ml medium is required per dish.
8. Aspirate the medium from all remaining three dishes. Add fresh medium without drugs to one dish as a nontreated control. Add medium containing 25 µM and 50 µM etoposide to the other two dishes. Incubate for 1 h at 37 °C.
9. After incubation, place dishes on ice for 5 min.
   ▲**CRITICAL STEP** Cooling the cells immediately limits reversal of topoisomerase–DPCs during further processing.
10. Remove the medium, add 1 ml ice-cold 1× 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.
11. 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.

10. Each cell suspension is processed individually for Steps 10 and 11. Briefly vortex cells, transfer to the thermoblock set at 45 °C 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 afterward.
    ◆**TROUBLESHOOTING**
11. Pipette the low-melt agarose/cell mixture into the plug mold until it is filled up (~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, and Supplementary Video 1). This step must be swiftly completed to prevent premature solidification of the low-melt agarose.
    ◆**TROUBLESHOOTING**
12. Once all agarose plugs have been cast, place the mold in a fridge at 4 °C 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**
13. 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 and Supplementary Video 2). 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:

BOX 3

5-azadC treatment to induce DNMT1–DPCs and monitor their repair

● **TIMING** 3 d, handling 3–4 h

Procedure

1. Day 1: in the morning, prewarm medium (DMEM + 10% FBS) in a water bath at 37 °C.
2. Collect HeLa T-REx Flp-In cells cultured in DMEM containing 10% FBS. 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 1× PBS followed by trypsinization (trypLE, 1 ml). Once cells have detached, inactivate trypsin with 9 ml medium and transfer the cell suspension to a 15-ml conical tube.
3. 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.
4. Seed 750,000 cells per condition in fresh 6-cm dishes, adding at least 3 ml medium. The example described here requires seven dishes, which includes one additional dish to determine the number of cells.
5. Allow cells to attach for at least 8 h.
6. In the evening, prewarm thymidine medium (DMEM + 10% FBS + 2 mM thymidine) in a water bath at 37 °C.
7. Confirm that cells have attached using a microscope.
8. Initiate the synchronization of cells by a double-thymidine block by aspirating the medium from all dishes and carefully add 2.5 ml thymidine medium to each dish.
9. Incubate cells overnight at 37 °C.
10. Day 2: in the morning, prewarm medium (DMEM + 10% FBS) in a water bath at 37 °C.
11. Remove the thymidine medium from the dishes and wash twice with 3 ml 1× PBS each to release cells from the first thymidine block. Add fresh medium without thymidine and incubate for 8–9 h.
12. In the evening, prewarm thymidine medium (DMEM + 10% FBS + 2 mM thymidine) in a water bath at 37 °C.
13. Repeat steps 8–9 above.
14. Day 3: in the morning, prewarm medium (DMEM + 10% FBS) in a water bath at 37 °C.
15. Remove the thymidine medium from the dishes and wash twice with 3 ml 1× PBS each. Add fresh medium without thymidine to release cells from the second thymidine block.
16. 3 h following release, DNMT1–DPC formation can be induced.
17. 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 Stage 2 of the main procedure (Step 5).
18. Aspirate medium from the six remaining dishes. Add a minimum of 2.5 ml fresh medium containing 10 µM 5-azadC (add 4 µl 5-azadC stock solution (50 mM) to 20 ml prewarmed medium without thymidine) to four dishes. Add medium without 5-azadC to the remaining two nontreated control dishes. Incubate for 30 min at 37 °C.
19. Following the 30 min incubation, place one nontreated control dish and one 5-azadC-treated dish on ice (timepoint 0 h).
20. Proceed by rapidly washing the remaining four dishes twice with 3 ml 1× PBS each and add fresh 5-azadC-free medium. Start timing the recovery time from the moment you replace the medium. Let cells recover for the desired time (1, 2 and 3 h in the example described here) at 37 °C. The remaining nontreated control dish stays unchanged at 37 °C (timepoint 3 h).
21. Remove the medium from the dishes placed on ice in step 19 above. Wash once with 3 ml 1× PBS, add 1 ml ice-cold 1× 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.
22. Repeat step 21 at every desired recovery timepoint (here 1, 2 and 3 h after removal of 5-azadC-containing medium).

    ■ **PAUSE POINT** Cell pellets can be stored at −80 °C for up to 1 month.

1.5-ml tube for one plug, 2-ml tubes for two plugs, 5-ml tubes for up to five plugs. Large numbers of plugs (more than ten) 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.

14. 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 color at this step (Fig. 2c).
15. Place the tubes on a rotating wheel at 4 °C. Rotate for 4 h at 25 rpm to lyse cells within the agarose plugs.
16. 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 ten plugs. Do not unpack the gel yet.
17. 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.

BOX 4

UVC irradiation or POLα inhibition to induce HMCES–DPCs

● **TIMING 3 d, handling 3-4 h**

Procedure

1. Day 1: in the morning, prewarm medium (DMEM + 10% FBS) in a water bath at 37 °C.
2. Collect 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 medium 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 1× PBS followed by trypsinization (trypLE, 1 ml). Once cells have detached, inactivate trypsin with 9 ml medium and transfer the cell suspension to a 15-ml conical tube.
3. 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.
4. Seed 750,000 cells per condition and cell line in fresh 6-cm dishes, adding at least 3 ml medium. The example described here requires three dishes per cell line, which includes one additional dish per cell line to determine the number of cells.
5. Allow cells to attach for at least 8 h.
6. In the evening, prewarm thymidine medium (DMEM + 10% FBS + 2 mM thymidine) in a water bath at 37 °C.
7. Confirm that cells have attached using a microscope.
8. Initiate the synchronization of cells by double-thymidine block by aspirating the medium from all dishes and carefully add 2.5 ml thymidine medium to each dish.
9. Incubate cells overnight at 37 °C.
10. Day 2: in the morning, prewarm medium (DMEM + 10% FBS) in a water bath at 37 °C.
11. Remove the thymidine medium from the dishes and wash twice with 3 ml 1× PBS each to release cells from the first thymidine block. Add fresh medium without thymidine and incubate for 8–9 h.
12. In the evening, prewarm thymidine medium (DMEM + 10% FBS + 2 mM thymidine) in a water bath at 37 °C.
13. Repeat steps 8–9 above, but additionally add doxycycline to the thymidine medium 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 medium in a 50-ml conical tube).
14. Day 3: in the morning, prewarm medium (DMEM + 10% FBS) in a water bath at 37 °C.
15. Confirm the successful induction of HMCES expression using a fluorescence microscope with appropriate filters (excitation 496 nm, emission 543 nm) for HMCES's mVenus-tag.
16. Remove the thymidine medium from the dishes and wash twice with 3 ml 1× PBS each. Add fresh medium without thymidine but containing doxycycline (1 µg/ml) to release cells from the second thymidine block.
17. 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 1× PBS, before adding 2 ml fresh 1× PBS. Irradiate one dish per cell line in a UVC irradiator with a dose of 50 J/m^2. Following irradiation, remove the 1× PBS and add fresh prewarmed doxycycline-free medium. Also change the medium of the nontreated control dishes and let cells recover for 6 h at 37 °C
    - For POLα inhibition, aspirate medium from all dishes. Add a minimum of 2.5 ml fresh medium containing 5 µM CD437 (add 7.5 µl CD437 stock solution (10 mM) to 15 ml prewarmed medium without thymidine) to one dish per cell line. Add medium without CD437 to the remaining nontreated control dishes. Incubate for 1 h at 37 °C.
18. 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 Stage 2 of the main procedure (Step 5).
19. Place the remaining dishes on ice. Remove the medium, wash once with 3 ml 1× PBS, add 1 ml ice-cold 1× 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 −80 °C for up to 1 month.

▲ **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**

18. Optional nuclease control can be performed at this point (see 'Nuclease control', Steps 19–25). Alternatively, continue with Step 26.

(Optional) Nuclease control

● **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.

19. 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

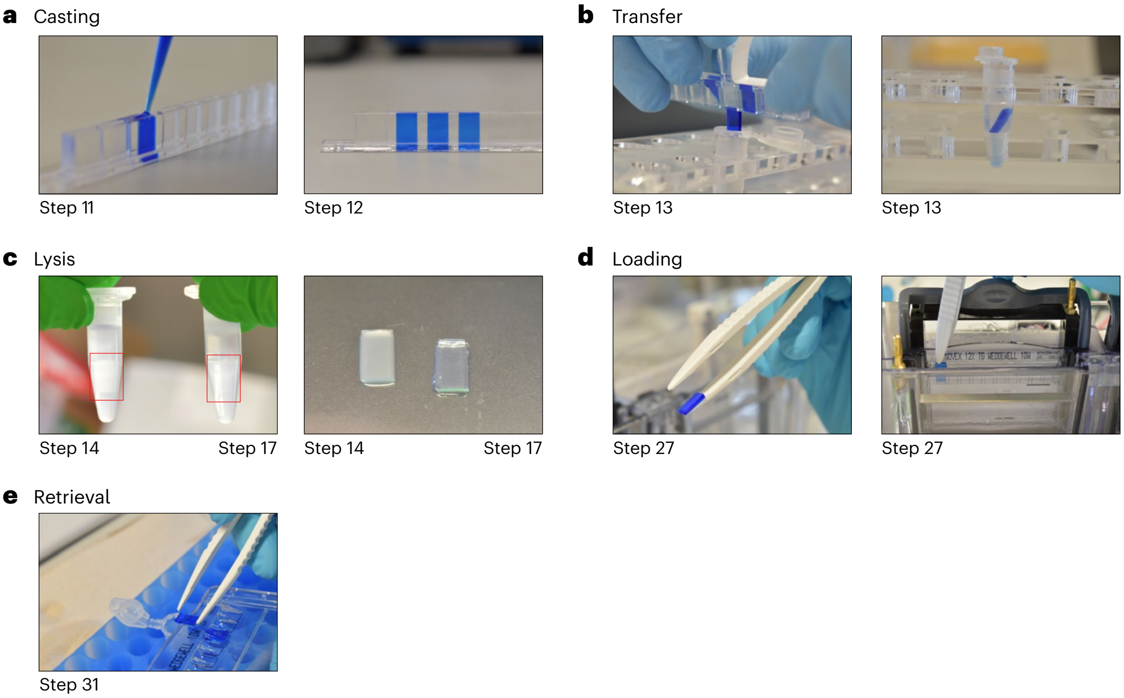


Fig. 2 | Critical steps of the PxP protocol. Photographs presenting correct handling of plugs during the most critical steps of the PxP protocol. For better visualization, agarose plugs were stained with blue bromophenol dye in **a**,**b**,**d** and **e**. The corresponding steps of the protocol are 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 color 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.

opaque after the first wash, due to some coprecipitation of $MgCl_2$ and sarkosyl, but this will not affect the DNA digestion.

20. Label the tubes with the plugs that will be digested with nuclease. We recommend digesting half of the plugs per condition.
21. 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.
22. Incubate all plugs at 37 °C for 30 min with vigorous shaking (500–600 rpm).
23. Place tubes on ice for 5 min.
24. Aspirate the buffer from all tubes and keep on ice.
25. Proceed with Step 26.

Electro-elution

26. 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 16. Once the running buffer has reached all wells, remove ~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 before loading. If the buffer level is too high, plugs will float away during loading.

    ◆ **TROUBLESHOOTING**
27. Use plastic tweezers to load the plugs into the wells of the Novex WedgeWell SDS–PAGE gel (Fig. 2d and Supplementary Video 3).

▲ **CRITICAL STEP** 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**

28. Carefully refill 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.

29. Connect the power supply and start the electro-elution at constant amperage (20 mA per gel) for 60 min. Initial voltage should be 36–45 V.

◆ **TROUBLESHOOTING**

30. 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 (Supplementary Video 4).

▲ **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**

31. Transfer electro-eluted plugs to fresh 1.5-ml plastic tubes using plastic tweezers (Fig. 2e and Supplementary Video 4). At this step, a maximum of ten plugs corresponding to the same condition can be pooled together (see 'Experimental design').
32. 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 homogeneous electro-elution of noncross-linked proteins from the plugs (Fig. 3).

Stage 3: DPC detection and identification

● **TIMING** 2 h–2 d

33. DPCs isolated by PxP can be characterized 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 (Steps 34–66) or B (Steps 67–82), respectively.

Option A: SDS–PAGE followed by western blot or fluorescent staining

● **TIMING** 2 d, handling 1–3 h

34. Place plugs for 2–3 min at 99 °C until molten.
35. Centrifuge at 10,000*g* for 30 s at RT.
36. Molten plugs (~80 µl) can be mixed directly with 40 µl 4× LDS sample buffer and 10 µl 10× reducing agent (proceed directly to Step 48). Alternatively, samples can be further concentrated by TCA precipitation (Steps 37–47).
37. For TCA precipitation, add 80 µl wash buffer containing 200 U benzonase per ml to each tube.

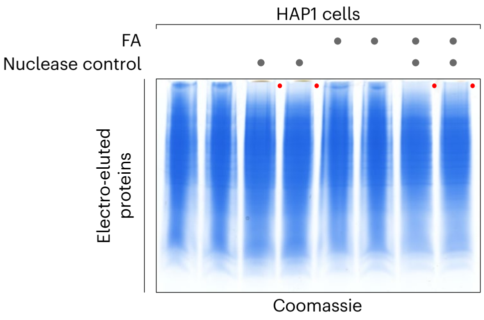


Fig. 3 | Electro-elution of agarose plugs. DPCs were isolated by PxP from HAP1 cells treated with 1 mM formaldehyde (FA) for 1 h, including a nuclease control step. The gel used for electro-elution was stained with a Coomassie-based protein stain, showing eluted noncross-linked proteins. The red dots highlight signal depletion in nuclease control samples.

38. Vortex shortly and incubate at 37 °C for 30 min with vigorous shaking (500–600 rpm).
39. Add 1 U agarase per plug and vortex briefly.
40. Incubate at 42 °C for additional 30 min before placing samples on ice. If the agarase digestion was successful, the samples should not solidify anymore.
41. Add 440 µl deionized water and mix by vortexing.
42. Add 100 µl 70% TCA and incubate for 20 min on ice.
43. Centrifuge at high speed (>20,000*g*) for 20 min at 4 °C in a precooled centrifuge.
44. Carefully remove supernatant and add 1 ml acetone (precooled to −20 °C).
45. Centrifuge again at high speed (>20,000*g*) for 20 min at 4 °C. A white pellet should be visible at this step.
46. After removing the supernatant, place tubes with open lids in a fume hood to evaporate any remaining acetone, normally 10–20 min are enough.
47. Resuspend each pellet in 50 µl 1× LDS sample buffer with 1× reducing agent (12.5 µl 4× LDS sample buffer, 5 µl 10x reducing agent and 32.5 µl deionized water).
48. Boil at 99 °C for 20 min to revert bonds between DNA and cross-linked proteins.
    ◆ **TROUBLESHOOTING**
49. Samples containing agarose must be kept warm before 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 at 99 °C for 2 min.
50. Resolve samples using standard SDS–PAGE gel electrophoresis. We suggest using ultrapure water for the preparation of running buffers. Given that formaldehyde mainly cross-links 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 51 for silver staining, with Step 52 for fluorescent staining or with Step 59 for analysis by western blotting.
51. 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.
52. 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).
53. Place on a shaker with gentle agitation (20–25 rpm) for 30 min at RT.
54. Remove fixation solution and add sufficient SYPRO ruby protein stain to cover the gel. Wrap the box with aluminum foil and incubate overnight with gentle agitation at RT.
55. The next morning, remove the SYPRO ruby protein gel stain and cover the gel with destaining solution.
56. Incubate with gentle agitation for 15 min at RT.
57. Repeat Steps 55–56.
58. 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.
59. To detect cross-linking 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 1× NuPAGE transfer buffer for 50 min at 100 V for small proteins and 1 h and 10 min for larger proteins.
60. Following transfer, block the membrane for 1 h in 5% milk in TBS-T.
61. 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.
62. Incubate overnight in the cold room in primary antibody diluted in TBS-T containing 2.5% BSA or 5% milk (see 'Reagents').

63. The next day, wash the membrane three times for 10 min with TBS-T.
64. Add the corresponding secondary antibody diluted in 5% milk in TBS-T and incubate for 1 h.
65. Wash the membrane at least three times for 10 min with TBS-T.
66. Incubate the membrane with an ECL substrate and acquire the chemiluminescent signal using a western blot imaging system. Avoid overexposure.

Option B: DPC identification by mass spectrometry

● **TIMING** variable, handling 2–6 h

67. 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.
68. Aspirate the fixation solution.
69. 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.
70. Remove buffer and repeat Step 69.
71. Aspirate the buffer.

    ■ **PAUSE POINT** Samples can be stored at 4 °C overnight.
72. 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.
73. Submit the cut agarose plugs to a mass spectrometry facility, which can process them with standard protocols for in-gel tryptic digestion[49].
74. 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).
75. Import the results into RStudio.
76. Remove reverse identified proteins and contaminants.
77. Log2 transform the intensities.
78. 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.
79. Normalize the intensities between the biological replicates. We use the R package preprocessCore to perform quantile normalization.
80. 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 nonrandomly (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.
81. 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 false discovery rate (FDR) can be calculated from either the *t*-statistic or the *P* value returned by limma.
82. Visualize data in R using dedicated graphics packages such as ggplot2. Alternatively, the results can be exported as a text file and further analyzed 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 table

Step	Problem	Possible reason	Solution
Reagent setup	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 1× PBS if needed and keep concentration at 2%
10	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, resulting in premature solidification of the agarose. Alternatively, cell suspension and low-melt agarose solution were not mixed thoroughly, or cell pellets were not well resuspended	Allow enough time for warming up cell suspension before mixing with low-melt agarose solution. Mix cell suspension by pipetting just before placing it in the thermoblock
11	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
12	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 was not long enough. Alternatively, agarose stock has the wrong concentration	Make sure that all liquid was removed from cell pellets before freezing. Increase incubation in the fridge. If the problem persists, prepare a new agarose stock solution
17	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
26–27	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 before loading. Make sure that the buffer level is below the wells during loading
29	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)
30	Plugs do not stay in position when opening the gel	There was remaining buffer left before opening the gel	Use Supplementary Video 4 as a reference and try to remove as much buffer as possible before opening the gel
48	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

Stage 1: DPC induction: 2–3 d (depending on the chosen treatment)
Step 1

Stage 2: DPC isolation by PxP: 6–7 h
Steps 2–18, Plug casting and lysis
Steps 19–25 (optional), Nuclease control: 45 min, handling 15 min
Steps 26–32, Electro-elution: 1–2 h, handling 30 min
Step 33, A segue to Options A and B of Stage 3

Stage 3: DPC detection and identification: 2 h–2 d (depending on the chosen analysis)
Option A, Steps 34–66, Western blot or fluorescent staining: 2 d, handling 1–3 h
Option B, Steps 67–82, Analysis by mass spectrometry: variable, handling 2–6 h

Boxes
Box 1, DPC induction using formaldehyde: 2 d, handling 2–3 h
Box 2, Etoposide treatment to induce TOP2–DPCs: 2 d, handling 2–3 h
Box 3, 5-azadC treatment to induce DNMT1–DPCs and monitor their repair: 3 d, handling 3–4 h
Box 4, UVC irradiation or POLα-inhibition to induce HMCES–DPCs: 3 d, handling 3–4 h

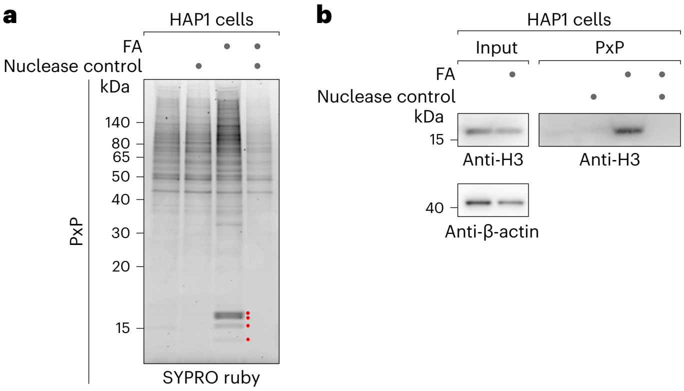


Fig. 4 | DPC induction using formaldehyde. a,**b**, DPCs were isolated by PxP from HAP1 cells treated with 1 mM formaldehyde (FA) for 1 h, including a nuclease control step. PxP samples were resolved in a 12% SDS–PAGE gel and stained with SYPRO ruby protein stain (**a**). In parallel, PxP and input samples were analyzed by western blotting using the indicated antibodies (**b**).

Anticipated results

Electro-elution of agarose plugs

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 32, 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).

Analysis of DPCs by SDS–PAGE followed by western blot or fluorescent staining

DPC induction using formaldehyde

PxP samples obtained from formaldehyde-treated cells should display distinct bands at ~15 kDa when analyzed by SDS–PAGE followed by fluorescent staining with SYPRO ruby protein stain (Fig. 4a, red dots). These bands correspond to cross-linked histone proteins and are sensitive to the nuclease control treatment before 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 noncross-linked proteins as evident from the accumulation of β-actin in

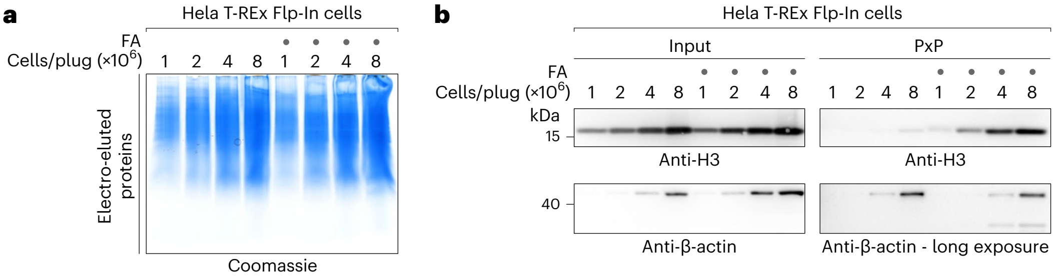


Fig. 5 | Effect of the number of cells per plug on specific and unspecific signals in PxP experiments. a,**b**, DPCs were isolated by PxP from HeLa T-REx Flp-In cells treated with 2 mM formaldehyde (FA) for 1 h with increasing numbers of cells embedded per plug. The gel used for electro-elution was stained with a Coomassie-based protein stain, showing eluted non-cross-linked proteins (**a**). PxP samples were resolved in a 4–12% SDS–PAGE gel and were analyzed by western blotting using the indicated antibodies (**b**).

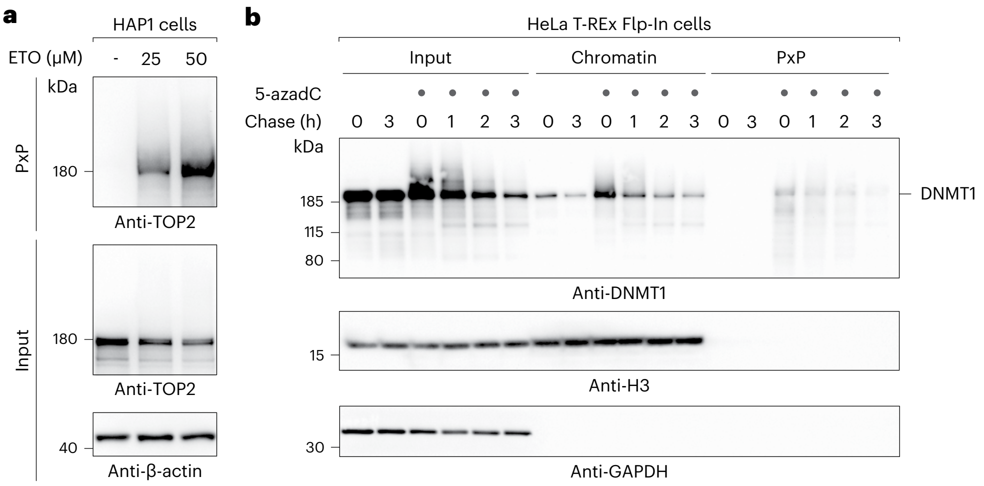


Fig. 6 | Induction of TOP2– and DNMT1–DPCs. a, DPCs were isolated by PxP from HAP1 cells treated with the indicated concentrations of etoposide (ETO) for 1 h. PxP and input samples were analyzed by western blotting using the indicated antibodies. **b**, Analysis of DNMT1–DPC repair by chromatin fractionation and PxP–WB in HeLa T-REx Flp-In. Cells 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 collected after 5-azadC exposure or after an additional incubation in drug-free media for 1, 2 or 3 h, as indicated. Samples were analyzed by western blotting using the indicated antibodies.

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

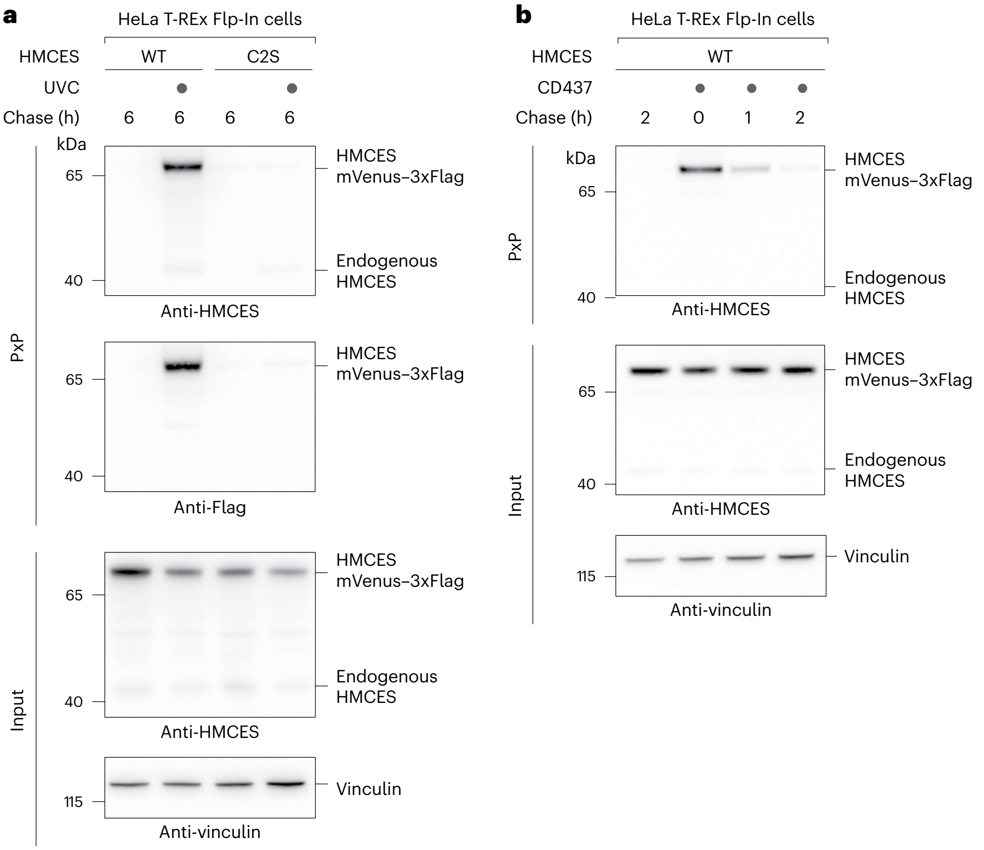


Fig. 7 | Induction of HMCES–DPCs. a, HeLa T-REx Flp-In cells expressing doxycycline-inducible HMCES–mVenus–3xFlag variants (WT or a catalytically compromised C2S variant) were synchronized via a double-thymidine block and released for 2 h into early/mid S-phase, before irradiation with UVC (50 J/m^2). 6 h after irradiation, DPCs were isolated by PxP. PxP samples were analyzed 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 collected or after the indicated chase (1 or 2 h) in drug-free media. Cells were subjected to PxP and samples were analyzed by western blotting using the indicated antibodies.

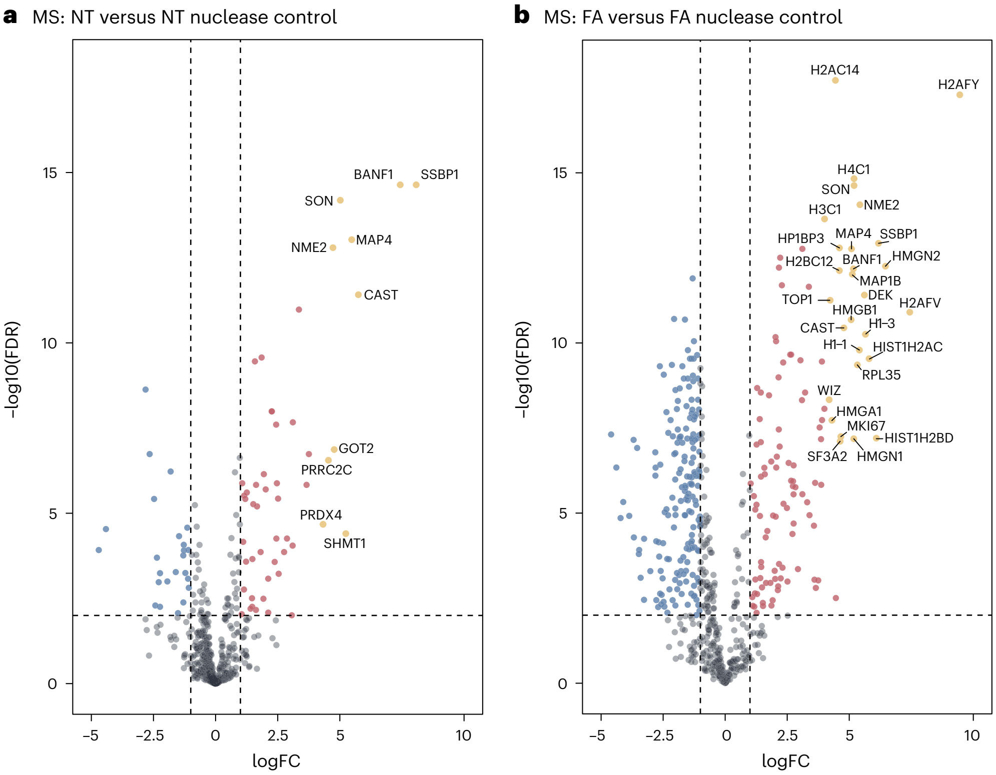


Fig. 8 | Analysis of DPCs by mass spectrometry. a,**b**, Volcano plots depicting log2 fold changes (logFC) plotted against the FDR (−log10) of proteins isolated by PxP from HeLa cells, comparing either nontreated control cells (NT) (**a**) or cells treated with 2 mM formaldehyde (FA) for 1 h (**b**) 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.

even strongly reducing PxP signals (Extended Data Fig. 1a,b). Therefore, we have omitted these steps in this protocol.

Etoposide treatment to induce TOP2–DPCs

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 nontreated control cells. Notably, partial depletion of TOP2 can be observed in input samples, which can serve as a positive control for etoposide treatment.

5-azadC treatment to induce DNMT1–DPCs and monitor their repair

DNMT1–DPCs can be detected in PxP samples obtained from S-phase synchronized cells after a 30 min treatment with 5-azadC by western blotting against endogenous DNMT1. When cells are let to recover in drug-free media for 1, 2 or 3 h 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 nontreated (Fig. 6b).

UVC irradiation/POLα-inhibition to induce HMCES–DPCs

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, cross-linking HMCES–WT can be observed upon treatment of cells with the POLα inhibitor CD437 (Fig. 7b). Once the inhibitor is removed, the cross-links are resolved rapidly (Fig. 7b), probably related to the autocatalytic reversal of HMCES–DPCs[41,51].

Analysis of DPCs by mass spectrometry

To identify endogenous DPCs, compare proteins identified in PxP samples from nontreated cells with the respective nuclease control samples (Fig. 8a and Supplementary Tables 1 and 2). DPCs are expected to be depleted in the nuclease control samples. To identify drug-induced DPCs, compare the treated condition with the nontreated condition and the treated condition to the respective nuclease control (Fig. 8b and Supplementary Tables 1 and 2). Only proteins significantly enriched in a treated condition over the respective nuclease control should be considered DPCs. Typically, we consider as meaningful hits proteins with a log2 fold-change larger than 1 and an FDR smaller than 0.05. This protocol has recently been used to monitor the resolution of formaldehyde-induced DPCs[52].

Reporting summary

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

Data availability

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. Uncropped scans of all blots and gels generated in this study are provided in the Source data file. Data presented in graphs are provided in the Source data file. Source data are provided with this paper.

Received: 24 April 2023; Accepted: 19 February 2024;

References

1. Weickert, P. & Stingele, J. DNA-protein crosslinks and their resolution. *Annu. Rev. Biochem.* **91**, 157–181 (2022).
2. Stingele, J., Habermann, B. & Jentsch, S. DNA–protein crosslink repair: proteases as DNA repair enzymes. *Trends Biochem. Sci.* **40**, 67–71 (2015).
3. Stingele, J., Bellelli, R. & Boulton, S. J. Mechanisms of DNA–protein crosslink repair. *Nat. Rev. Mol. Cell. Biol.* **18**, 563–573 (2017).
4. Wang, M., Dingler, F. A. & Patel, K. J. Genotoxic aldehydes in the hematopoietic system. *Blood* **139**, 2119–2129 (2022).
5. Vijayraghavan, S. & Saini, N. Aldehyde-associated mutagenesis horizontal line current state of knowledge. *Chem. Res. Toxicol.* https://doi.org/10.1021/acs.chemrestox.3c00045 (2023).
6. Pommier, Y. et al. Tyrosyl-DNA-phosphodiesterases (TDP1 and TDP2). *DNA Rep.* **19**, 114–129 (2014).
7. Pommier, Y. & Marchand, C. Interfacial inhibitors: targeting macromolecular complexes. *Nat. Rev. Drug Discov.* **11**, 25–36 (2011).
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).
9. Mohni, K. N. et al. HMCES maintains genome integrity by shielding abasic sites in single-strand DNA. *Cell* **176**, 144–153 e113 (2019).
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).
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).
12. Ruggiano, A. & Ramadan, K. DNA–protein crosslink proteases in genome stability. *Commun. Biol.* **4**, 11 (2021).
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).
14. Serbyn, N. et al. The aspartic protease Ddi1 contributes to DNA–protein crosslink repair in yeast. *Mol. Cell* **77**, 1066–1079 e1069 (2020).
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).
16. Vaz, B. et al. Metalloprotease SPRTN/DVC1 orchestrates replication-coupled DNA–protein crosslink repair. *Mol. Cell* **64**, 704–719 (2016).
17. Lopez-Mosqueda, J. et al. SPRTN is a mammalian DNA-binding metalloprotease that resolves DNA–protein crosslinks. *eLife* https://doi.org/10.7554/eLife.21491 (2016).
18. Kojima, Y. et al. FAM111A protects replication forks from protein obstacles via its trypsin-like domain. *Nat. Commun.* **11**, 1318 (2020).
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).
20. Sparks, J. L. et al. The CMG helicase bypasses DNA–protein cross-links to facilitate their repair. *Cell* **176**, 167–181 e121 (2019).
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).
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).
23. Yaneva, D. et al. The FANCJ helicase unfolds DNA–protein crosslinks to promote their repair. *Mol. Cell* **83**, 43–56 e10 (2023).
24. Gallina, I. et al. The ubiquitin ligase RFWD3 is required for translesion DNA synthesis. *Mol. Cell* **81**, 442–458 e449 (2021).
25. Borgermann, N. et al. SUMOylation promotes protective responses to DNA–protein crosslinks. *EMBO J.* https://doi.org/10.15252/embj.2019101496 (2019).
26. Sun, Y. et al. A conserved SUMO pathway repairs topoisomerase DNA–protein cross-links by engaging ubiquitin-mediated proteasomal degradation. *Sci. Adv*. https://doi.org/10.1126/sciadv.aba6290 (2020).
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).
28. Weickert, P. et al. SPRTN patient variants cause global-genome DNA–protein crosslink repair defects. *Nat. Commun.* **14**, 352 (2023).
29. Maskey, R. S. et al. Spartan deficiency causes genomic instability and progeroid phenotypes. *Nat. Commun.* **5**, 5744 (2014).
30. Lessel, D. et al. Mutations in SPRTN cause early onset hepatocellular carcinoma, genomic instability and progeroid features. *Nat. Genet.* **46**, 1239–1244 (2014).
31. Donsbach, M. et al. A non-proteolytic release mechanism for HMCES-DNA-protein crosslinks. *EMBO J.* **42**, e113360 (2023).
32. Reinking, H. K. & Stingele, J. Protein–oligonucleotide conjugates as model substrates for DNA–protein crosslink repair proteases. *STAR Protoc.* **2**, 100591 (2021).
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).
34. Serbyn, N. et al. SUMO orchestrates multiple alternative DNA–protein crosslink repair pathways. *Cell Rep.* **37**, 110034 (2021).
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).

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).
37. Kiianitsa, K. & Maizels, N. A rapid and sensitive assay for DNA–protein covalent complexes in living cells. *Nucleic Acids Res.* **41**, e104 (2013).
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).
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).
40. Wilhelm, L. et al. SMC condensin entraps chromosomal DNA by an ATP hydrolysis dependent loading mechanism in *Bacillus subtilis*. *eLife* https://doi.org/10.7554/eLife.06659 (2015).
41. Rua-Fernandez, J. et al. Self-reversal facilitates the resolution of HMCES DNA-protein crosslinks in cells. *Cell Rep.* **42**, 113427 (2023).
42. *RStudio: Integrated Development Environment for R* (RStudio, 2023).
43. Bolstad, B. M., Irizarry, R. A., Astrand, M. & Speed, T. P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185–193 (2003).
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).
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).
46. Ritchie, M. E. et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
47. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2016).
48. Strimmer, K. fdrtool: a versatile R package for estimating local and tail area-based false discovery rates. *Bioinformatics* **24**, 1461–1462 (2008).
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).
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).
51. Donsbach, M. et al. A non-proteolytic release mechanism for HMCES–DNA–protein crosslinks. *EMBO J.* **42**, e113360 (2023).
52. Carnie, C. J. et al. Transcription-coupled repair of DNA–protein cross-links depends on CSA and CSB. *Nat. Cell Biol.* **26**, 797–810 (2024).

Acknowledgements

P.W. and S.D. are supported by the International Max-Planck Research School for Molecules of Life. Research in the laboratory of J.S. is funded by European Research Council (ERC Starting Grant 801750 DNAProteinCrosslinks), the Alfried Krupp Prize for Young University Teachers awarded by the Alfried-Krupp von Bohlen und Halbach-Stiftung, European Molecular Biology Organization (YIP4644), a Vallee Foundation Scholarship and Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) (project ID 213249687 – SFB 1064, project-ID 393547839 – SFB 1361).

Author contributions

P.W., S.D. and H.-Y.L. provided anticipated results and designed figures. M.J.G. analyzed 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.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41596-024-01004-z.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41596-024-01004-z.

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

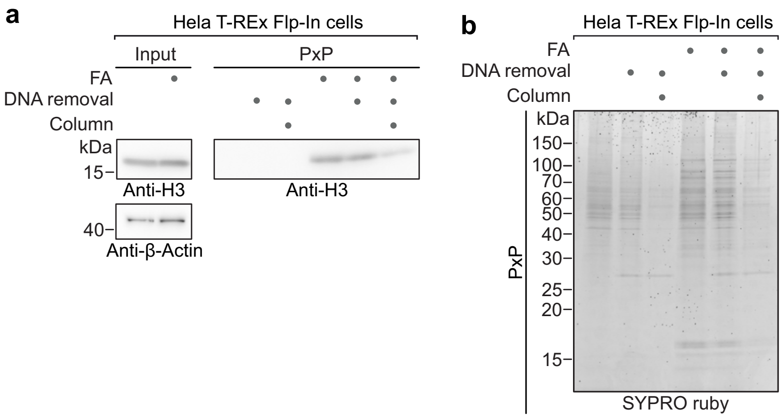
Peer review information *Nature Protocols* thanks Fuyi Wang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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

Related links

Key references using this protocol

Weickert, P. et al. *Nat. Commun.* **14**, 352 (2023): https://doi.org/10.1038/s41467-023-35988-1
Carnie, C. J. et al. *Nat. Cell Biol.* **26**, 797–810 (2024): https://doi.org/10.1038/s41556-024-01391-1



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 cross-linked 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 analyzed by western blotting using the indicated antibodies (a) or resolved in a 12% SDS-PAGE gel and stained with SYPRO ruby protein stain (b).

3.3 Allosteric activation of the SPRTN protease by ubiquitin maintains genome stability

Contribution report

This publication examines the regulatory mechanisms that restrict the activity of the DNA-dependent protease SPRTN to DPC degradation, focusing on the role of ubiquitin in this process. I performed *in vitro* experiments shown in the manuscript with help from Denitsa Yaneva and Dina S. Schnapka. Together with Pedro Weickert, I generated SPRTN overexpression cell lines and performed PxP experiments. Cellular SPRTN autocleavage assays were done by Pedro Weickert. Dina S. Schnapka and Maximilian J. Götz established the system for SUMO-targeted DPC-ubiquitylation, and Hyun-Seo Kang performed NMR measurements. Christian Wiebeler and Nadine Schwierz planned and carried out all molecular dynamics simulations. Yuichi J. Machida and Yuka Machida generated *Sprtn*$^{F/-}$ MEF cell lines and performed genome stability experiments. Christian Renz generated ubiquitylated HMCES^{SRAP} and Aldwin S. Rahmanto performed mass spectrometry analysis of ubiquitylated DPCs. Sophie M. Guthenthaler-Tietze completed ICP-OES measurements of recombinant SPRTN. Figures were prepared by myself, and the manuscript was written by Julian Stingele and me with input from all authors.

Summary

The DNA-dependent protease SPRTN maintains genome stability by degrading toxic DNA-protein crosslinks (DPCs). To understand how SPRTN's promiscuous protease activity is confined to cleavage of crosslinked proteins, we reconstitute the repair of DPCs including their modification with SUMO and ubiquitin chains *in vitro*. We discover that DPC ubiquitylation strongly activates SPRTN independently of SPRTN's known ubiquitin-binding domains. Using protein structure prediction, MD simulations and NMR spectroscopy we reveal that ubiquitin binds to SPRTN's protease domain, promoting an open, active conformation. Replacing key interfacial residues prevents allosteric activation of SPRTN by ubiquitin, leading to genomic instability and cell cycle defects in cells expressing truncated SPRTN variants that cause premature ageing and liver cancer in Ruijs-Aalfs syndrome patients. Collectively, our results reveal a ubiquitin-dependent regulatory mechanism that ensures SPRTN activity is deployed precisely when and where it is needed.

nature communications

Article

https://doi.org/10.1038/s41467-025-61224-z

Allosteric activation of the SPRTN protease by ubiquitin maintains genome stability

Received: 23 December 2024

Accepted: 12 June 2025

Sophie Dürauer[1,2], Hyun-Seo Kang[3,4], Christian Wiebeler[5], Yuka Machida[6], Dina S. Schnapka[1,2], Denitsa Yaneva[1,2], Christian Renz[7], Maximilian J. Götz[1,2], Pedro Weickert[1,2], Abigail C. Major[5], Aldwin S. Rahmanto[7,8], Sophie M. Gutenthaler-Tietze[9,10], Lena J. Daumann[9], Petra Beli[7,8], Helle D. Ulrich[7], Michael Sattler[3,4], Yuichi J. Machida[6], Nadine Schwierz[5] & Julian Stingele[1,2]

The DNA-dependent protease SPRTN maintains genome stability by degrading toxic DNA-protein crosslinks (DPCs). To understand how SPRTN's promiscuous protease activity is confined to cleavage of crosslinked proteins, we reconstitute the repair of DPCs including their modification with SUMO and ubiquitin chains in vitro. We discover that DPC ubiquitylation strongly activates SPRTN independently of SPRTN's known ubiquitin-binding domains. Using protein structure prediction, MD simulations and NMR spectroscopy we reveal that ubiquitin binds to SPRTN's protease domain, promoting an open, active conformation. Replacing key interfacial residues prevents allosteric activation of SPRTN by ubiquitin, leading to genomic instability and cell cycle defects in cells expressing truncated SPRTN variants that cause premature aging and liver cancer in Ruijs-Aalfs syndrome patients. Collectively, our results reveal a ubiquitin-dependent regulatory mechanism that ensures SPRTN activity is deployed precisely when and where it is needed.

Cells invest in extensive repair mechanisms to ensure fidelity of the genetic information stored in their DNA. Defective DNA repair results in mutagenesis and genome instability, major hallmarks of cancer, aging and aging-related diseases[1,2]. Cellular DNA repair activities are organized by sophisticated networks of post-translational modifications[3,4]. Regulatory ubiquitylation events are critical to recruit DNA repair factors in highly controlled manners. Monoubiquitylation of PCNA promotes DNA damage tolerance by recruiting translesion synthesis (TLS) polymerases[5], while monoubiquitylation of the FANCD2/FANCI heterodimer traps the complex on DNA, initiating DNA repair by the Fanconi anemia pathway[6].

Tight regulation is especially important for DNA repair enzymes that are potentially toxic. The SPRTN protease employs a promiscuous activity to degrade covalent DNA-protein crosslinks (DPCs), but it has remained enigmatic how the enzyme achieves specificity for crosslinked proteins and how the unwanted cleavage of chromatin proteins is prevented. DPCs arise upon stabilization of covalent intermediates between DNA-processing enzymes and their substrates[7]. Additionally, various endogenous and environmental reactive agents crosslink proteins to DNA[8,9]. DPCs are toxic because they block DNA replication and transcription[10–13]. The collision of the replication machinery with crosslinked proteins initiates repair by SPRTN[14,15], which can

---

[1]Gene Center, Ludwig-Maximilians-Universität München, Munich, Germany. [2]Department of Biochemistry, Ludwig-Maximilians-Universität München, Munich, Germany. [3]Institute of Structural Biology, Molecular Targets and Therapeutics Center, Helmholtz Munich, Neuherberg, Germany. [4]Bavarian NMR Center and Department of Bioscience, TUM School of Natural Sciences, Technical University of Munich, Garching, Germany. [5]Institute of Physics, University of Augsburg, Augsburg, Germany. [6]Developmental Therapeutics Branch, Center for Cancer Research, National Cancer Institute, Bethesda, MD, USA. [7]Institute of Molecular Biology gGmbH, Mainz, Germany. [8]Institute of Developmental Biology and Neurobiology (IDN), Johannes Gutenberg-Universität Mainz, Mainz, Germany. [9]Chair of Bioinorganic Chemistry, Heinrich-Heine Universität Düsseldorf, Düsseldorf, Germany. [10]Department of Chemistry, Ludwig-Maximilians-Universität München, Munich, Germany. e-mail: stingele@genzentrum.lmu.de

additionally be triggered by global-genome mechanisms[9]. The repair of DPCs by SPRTN is essential for viability. Its loss is lethal in human cell lines[16] and leads to dramatic genome instability and early embryonic lethality in mice[17].

SPRTN features a metalloprotease domain at the *N*-terminus, which, together with the single-stranded DNA (ssDNA) -binding zinc-binding domain (ZBD), forms the conserved SprT domain (Fig. 1a)[18,19]. The SprT domain is followed by a basic region (BR) that interacts with double-stranded DNA (dsDNA)[20]. ZBD and BR couple SPRTN activity to the recognition of ssDNA-dsDNA junctions[21], that arise when DNA polymerases stall at DPCs during replication[14]. However, the recognition of DNA junctions cannot explain how specificity is achieved during DPC repair, given that these structures are common throughout the genome, for example on the lagging strand during DNA replication. In addition to its DNA-binding domains, SPRTN bears interaction motifs for binding to the segregase p97 (SHP box) and PCNA (PIP box)[22–25] but neither is required for SPRTN's DPC repair function[9,14,17]. Furthermore, SPRTN carries a *C*-terminal ubiquitin-binding zinc finger (UBZ), promoting SPRTN ubiquitylation and thereby its inactivation[26]. A motif interacting with ubiquitin (MIU) has been predicted at SPRTN's *N*-terminus but has not been experimentally confirmed[27]. The presence of ubiquitin-binding

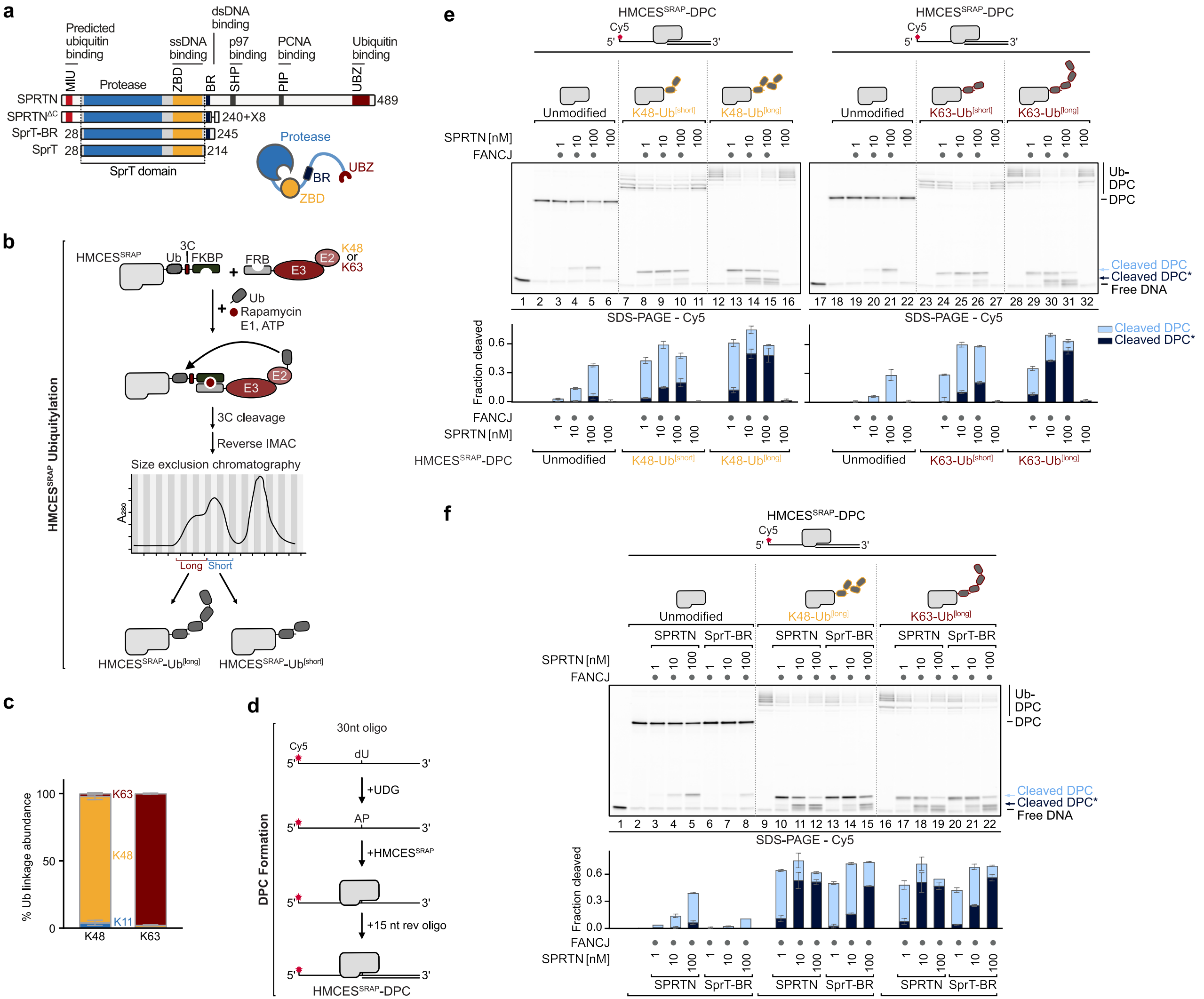


Fig. 1 | Ubiquitylation of DPCs promotes their cleavage by SPRTN. a Schematic of SPRTN's domain structure and truncated variants, featuring motif interacting with ubiquitin (MIU), protease domain, zinc-binding domain (ZBD), basic region (BR), SHP box for p97-binding, PCNA-interacting motif (PIP) and ubiquitin-binding zinc finger (UBZ). SPRTN$^{\Delta C}$ is caused by a frameshift mutation resulting in a variant composed of SPRTN's *N*-terminal 240 residues followed by eight additional amino acids (X8). **b** Schematic of HMCES^{SRAP} ubiquitylation to generate DPCs shown in **e**, **f**, Fig. 4 and Supplementary Fig. 5b and 6b. HMCES^{SRAP}-Ub(G76V)-3C-FKBP was incubated with FRB-E3 + E2 (K48 or K63) in the presence of ubiquitin, rapamycin, ubiquitin-E1 and ATP for 2 h (K63) or 6.5 h (K48) at 30 °C. After cleavage of the FKBP-tag via 3C-protease, ubiquitylated HMCES^{SRAP} was purified by reverse immobilized metal affinity chromatography (IMAC) and size-exclusion chromatography (SEC). **c** Mass spectrometry analysis of ubiquitin linkages formed by ubiquitylation of HMCES^{SRAP} as shown in (**b**). Bar chart shows the mean ± SD of three biological replicates. **d** Schematic of the generation of HMCES^{SRAP}-DPCs. HMCES^{SRAP} was incubated for 30 min at 37 °C with a Cy5-labeled 30nt oligonucleotide containing a dU at position 15 and UDG. After crosslinking a complementary 15nt reverse oligonucleotide was annealed to form a ssDNA-dsDNA junction. **e** Indicated HMCES^{SRAP}-DPCs (10 nM) were incubated alone or in the presence of FANCJ (100 nM) and indicated concentrations of SPRTN (1-100 nM) for 1 h at 30 °C. Quantification: bar graphs represent the mean ± SD of three independent experiments. All samples derive from the same experiment and gels were processed in parallel. Values for cleavage of unmodified HMCES^{SRAP}-DPC are the same as in Supplementary Fig. 1b. Source data are provided as a Source Data file. **f** Indicated HMCES^{SRAP}-DPCs (10 nM) were incubated alone or in the presence of FANCJ (100 nM) and indicated concentrations of SPRTN or SprT-BR (1-100 nM) for 1 h at 30 °C. Quantification: bar graphs represent the mean ± SD of three independent experiments. All samples derive from the same experiment and gels were processed in parallel. Source data are provided as a Source Data file.

domains indicates a critical role of ubiquitin in regulating SPRTN-mediated DPC repair.

Indeed, DPCs are ubiquitylated during replication by the ubiquitin-E3s TRAIP and RFWD3[14,15,28], while SUMOylation precedes ubiquitylation of the protein adduct by the SUMO-targeted ubiquitin-E3s RNF4 and TOPORS during global-genome repair[9,29–32]. DPC ubiquitylation can promote proteasomal degradation of crosslinked proteins[9,14,15,29,30], but it has remained controversial whether it is important for SPRTN-mediated repair. Cleavage of a model DPC by SPRTN in frog egg extracts occurs even if the protein adduct has been treated with formaldehyde to prevent ubiquitylation[14]. Nonetheless, ubiquitylated DPCs accumulate upon SPRTN depletion[33], indicating that they are substrates of the protease. Furthermore, SPRTN's UBZ domain supports efficient DPC cleavage in frog egg extracts and cells[9,14], which has led to the speculation that the UBZ may help to recruit SPRTN to ubiquitylated DPCs. Surprisingly however, the UBZ domain is not essential for SPRTN function. Patients with Ruijs-Aalfs syndrome (RJALS) express truncated versions of SPRTN that lack the *C*-terminal part of the enzyme including the UBZ (SPRTN$^{\Delta C}$, Fig. 1a)[27]. RJALS patients suffer from premature aging and liver cancer[27], phenotypes that are recapitulated in mice with reduced SPRTN function[17]. Yet, truncated SPRTN patient variants are clearly compatible with life, in contrast to full loss of SPRTN. Indeed, the severe growth defects associated with SPRTN loss in conditional mouse knock-out cells are rescued by expression of a truncated SPRTN variant[34]. It has remained enigmatic how SPRTN patient variants target DPCs in the absence of the UBZ and, more generally, whether and how SPRTN activity is regulated by DPC ubiquitylation.

Here, we investigate the role of ubiquitin in SPRTN activation by biochemical reconstitution of DPC ubiquitylation, molecular dynamics (MD) simulations, NMR experiments and cellular assays. We find that DPC ubiquitylation activates SPRTN more than one hundred-fold. Activation occurs independently of SPRTN's UBZ domain but involves a ubiquitin-binding interface at the back of its protease domain. This interface is required in cells expressing truncated RJALS patient variants to maintain genome stability and cellular fitness. Collectively, our results reveal a regulatory mechanism that confines SPRTN's protease activity by linking its activation to DPC modification. Moreover, given that ubiquitin-dependent activation is retained in truncated SPRTN variants, our data explain how residual SPRTN function is maintained in RJALS patients.

Results

Ubiquitylation of DNA-protein crosslinks promotes their cleavage by SPRTN

To directly test whether DPC ubiquitylation regulates SPRTN, we reconstituted DPC ubiquitylation in vitro. To modify DPCs with ubiquitin chains of defined linkages, we employed synthetic engineered ubiquitin-E3s (streamlined versions of the previously described Ubiquiton system[35]), enabling us to modify the catalytic SRAP domain of HMCES (HMCES^{SRAP}) with K48- or K63-linked ubiquitin chains prior to DPC formation with an oligonucleotide containing an abasic (AP) site. HMCES actively crosslinks to AP sites within ssDNA to prevent AP site scission during DNA replication[36]. First, we fused a *C*-terminal tag containing a mono-ubiquitin moiety and a FK506-binding protein (FKBP) domain to HMCES^{SRAP}. We then incubated this substrate with ubiquitin, an engineered ubiquitin-E3 carrying an FKBP-rapamycin-binding (FRB) domain, ubiquitin-E1, ubiquitin-E2, ATP and rapamycin (Fig. 1b). Rapamycin induces proximity between the substrate and the E3, promoting modification of the ubiquitin moiety fused to HMCES^{SRAP} with either K48- or K63-linked polyubiquitin chains (depending on the identity of the E2/E3 enzymes used in the assay). Following cleavage of the 3C-site between ubiquitin and FKBP, HMCES^{SRAP} modified with short or long ubiquitin chains was purified over several steps (Fig. 1b and Supplementary Fig. 1a, for all recombinant proteins used in this study). Mass spectrometry (MS) analysis confirmed the specific formation of K48- and K63-linked polyubiquitin chains on HMCES^{SRAP} (Fig. 1c). DPCs were then generated by incubating unmodified or ubiquitylated HMCES^{SRAP} with an AP site-containing fluorescently-labeled ssDNA-dsDNA junction (Fig. 1d)[37,38].

Next, we incubated the DPCs with SPRTN and the helicase FANCJ, which is required for SPRTN activity in these assays. FANCJ loads on the ssDNA portion of the substrate and translocates into the crosslinked protein, resulting in unfolding of the protein adduct, which in turn enables SPRTN to cleave the DPC[37]. SPRTN cleaved ubiquitylated DPCs more efficiently than unmodified protein adducts, with long chains activating stronger than shorter ones, independently of linkage type (Fig. 1e, lanes 7-16 (K48) and lanes 23-32 (K63)). The ubiquitin-dependent activation of SPRTN was substantial with the extent of cleavage of ubiquitylated DPCs by 1 nM of SPRTN being comparable to the cleavage of unmodified DPCs by 100 nM of SPRTN (Fig. 1e, compare lanes 5 and 13 (K48) and lanes 21 and 29 (K63)). Remarkably, in addition to the fragment produced upon cleavage of unmodified DPCs (Fig. 1e, Cleaved DPC), smaller cleavage products (Fig. 1e, Cleaved DPC*) appeared upon cleavage of ubiquitylated DPCs. Of note, smaller cleavage products were also detected upon addition of free K48- or K63-linked tetra-ubiquitin chains, although to a lesser extent (Supplementary Fig. 1b, cleaved DPC*, lanes 7-9 (K48) and lanes 17-19 (K63)).

To test whether SPRTN's known ubiquitin-binding domains are mediating the stimulating effect of DPC ubiquitylation, we utilized a minimal active SPRTN variant (SprT-BR, aa28-245), that lacks both, MIU and UBZ (Fig. 1a). While the truncated SprT-BR variant showed reduced cleavage of unmodified DPCs compared to the wild-type (WT) enzyme (Fig. 1f, compare lanes 3-5 with lanes 6-8), DPC ubiquitylation strongly boosted its activity (Fig. 1f, compare lanes 10-12 with lanes 13-15 (K48) and lanes 17-19 with lanes 20-22 (K63)). The stimulating effect of DPC ubiquitylation on truncated SprT-BR suggested to us that this region likely contains an additional ubiquitin-binding site that mediates the effect of ubiquitin on SPRTN activation.

Ubiquitin promotes an open SPRTN conformation

To explore this possibility, we used ColabFold[39] to predict complexes between SprT-BR and ubiquitin. In the top-ranked model, the hydrophobic Ile44-patch of ubiquitin was predicted to interact with a hydrophobic interface at the back of the SprT domain (Supplementary Fig. 2a-b), hereafter referred to as ubiquitin-binding interface at the SprT domain (USD). Interestingly, in all models, the SprT domain was predicted to adopt an open conformation with a highly accessible active site facing the DNA binding site of the ZBD. A similar conformation was also predicted in the absence of ubiquitin, in stark contrast to the published crystal structure of the SprT domain (PDB:6mdx[19]) that shows a closed conformation with the ZBD restricting access to the active site (Fig. 2a–c).

To explore whether the predicted open SprT conformation is in equilibrium with the closed conformation and whether ubiquitin binding may affect SprT conformation, we conducted all-atoms MD simulations. We used either the crystal structure or ColabFold-based predictions of the SprT domain, alone or in combination with ubiquitin, as starting points (Fig. 2d–f and Supplementary Fig. 2c). The compact conformation observed in the crystal structure remained largely unchanged over the entire 400 ns timeframe in three independent simulations (Fig. 2d and Supplementary Movie 1). To reveal the predominant conformations within all simulations, we employed RMSD-based clustering (Fig. 2g-i), revealing a single cluster with a closed conformation (Fig. 2g). In contrast, simulations of the ColabFold-predicted SprT structure displayed larger conformational changes during the simulations (Fig. 2e). We observed collapses to a compact conformation with a smaller radius of gyration (Fig. 2e, red arrow). Collapses were followed by rapid reopening of the structure

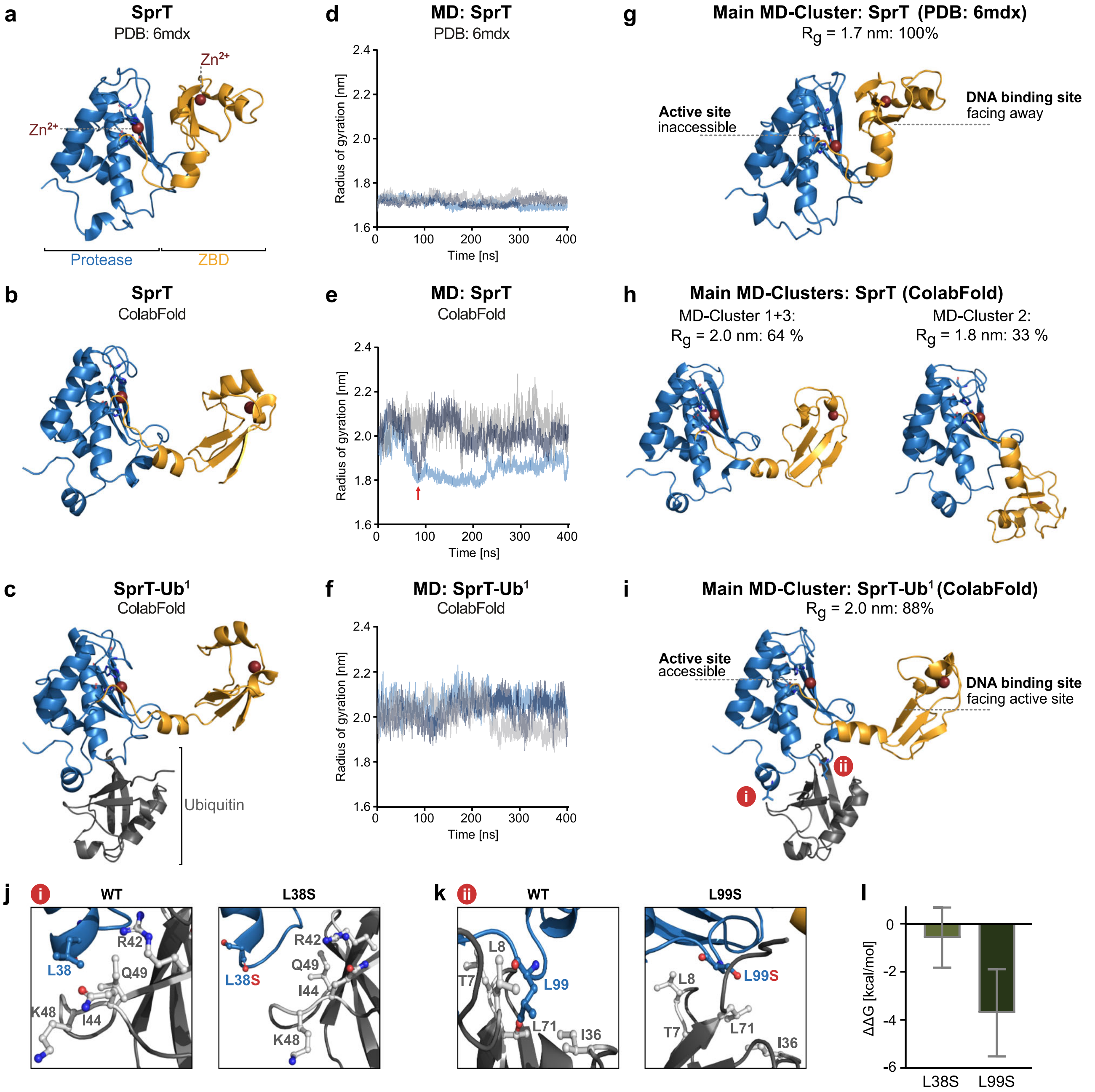


Fig. 2 | Ubiquitin promotes an open SPRTN conformation. a–c Experimental structure of SPRTN's SprT domain (SPRTN$^{aa28-214}$), PDB: 6mdx (**a**), ColabFold predicted structure of SprT (**b**) and ColabFold predicted structure of a SprT-ubiquitin (Ub^1) complex (**c**). Protease domain is colored in blue, zinc-binding domain (ZBD) in orange and the Ub^1 in grey. Zn^{2+} ions are colored in red. **d–f** Radius of gyration (Rg) of the indicated structures over 400 ns of molecular dynamics (MD) simulation. Each curve represents an independent MD trajectory (*n* = 3). Source data are provided as a Source Data file. **g–i** Main MD-clusters of the indicated structures during MD simulation for 400 ns, generated from three independent trajectories. For SprT (ColabFold predicted) two of three main MD-clusters are depicted. Rg correlating frequencies among all performed simulations are labeled above the structures. **j**, **k** Zoom-in to regions **i** and **ii** of the SprT-Ub^1 complex (**i**), showing amino acids of ubiquitin (in grey) surrounding residue Leu38 (**j**) or L99 (**k**) of SPRTN (in blue) in the wild-type (WT) protein (left) and upon L38S or L99S replacement, respectively (right). **l** SprT-Ub^1 binding energy difference (ΔΔG) between SprT-L38S or -L99S and WT protein obtained from alanine scanning. Bar graphs show the mean ± SD of 301 snapshots from PBSA calculations for the central structure of the largest cluster. Source data are provided as a Source Data file.

(Fig. 2e, dark blue trace) or retention of the compact conformation (Fig. 2e, light blue trace, and Supplementary Movie 2). Clustering revealed two clusters with an open conformation (Fig. 2h, left) and one cluster with a closed conformation (Fig. 2h, right). Strikingly, the presence of ubiquitin prevented transitions of the SprT domain to the closed conformation (Fig. 2f and Supplementary Movie 3) and simulations predominantly remained in an open conformation (Fig. 2i). Moreover, ubiquitin binding to the USD interface of the SprT remained stable across all three independent simulations (Fig. 2f). These data indicated to us that ubiquitin binding at the SprT domain may promote SPRTN activation by stabilizing an open conformation of the enzyme with an accessible active site.

Next, we wanted to determine amino acid residues within the USD interface that are important for ubiquitin-binding. In the predicted SprT-ubiquitin complex, Leu38 and Leu99 of SPRTN appeared to mediate the interaction via hydrophobic interactions involving multiple amino acids within ubiquitin's hydrophobic Ile44- and Ile36-patch, respectively (Fig. 2i-k and Supplementary Fig. 2d-e). Both residues, Leu38 and Leu99, are highly conserved throughout evolution (Supplementary Fig. 2f). To assess the effect of replacing either leucine residue with a hydrophilic serine (L38S, L99S), we conducted free energy end-point calculations using MMPBSA in conjunction with alanine scanning (see Methods for details), which enabled us to quantify the effect of each leucine-to-serine replacement to the overall binding affinity of the SprT-ubiquitin complex. We calculated a decrease in binding affinity of around 0.6 kcal/mol for the L38S replacement and a more substantial decrease of 3.74 kcal/mol for L99S (Fig. 2l). This effect is explained by replacement of Leu38 or Leu99 resulting in the loss of hydrophobic contacts to ubiquitin's Ile44- and Ile36-patch, respectively (Fig. 2i–k and Supplementary Fig. 2d, e).

Taken together, our MD simulations results suggest a model wherein ubiquitin binding to the USD promotes SPRTN activity by stabilizing an open conformation with an accessible active site.

DNA- and ubiquitin-binding affect SPRTN's conformation synergistically

To experimentally test whether ubiquitin binds to the USD interface and whether ubiquitin binding affects SPRTN's interaction with DNA, we used NMR spectroscopy. Heteronuclear single quantum coherence (HSQC) spectra of SprT-BR showed well-dispersed peaks (Supplementary Fig. 3a, b). Comparisons with a ZBD-BR construct enabled us to transfer many chemical shifts based on our previous analysis of the ZBD-BR construct[21] (Supplementary Fig. 3b, c, see Figure legend for details). In particular, we could unambiguously assign Trp ε1 ^{1}H,^{15}N resonances to the ZBD (Fig. 3, zoom-ins, orange labels) and protease domain (Fig. 3, zoom-ins, blue labels). Next, we compared NMR spectra of SprT-BR and SprT-BR-L99S, which superimposed very well (Supplementary Fig. 3d), except for those resonances in vicinity to the mutation site, indicating that structural integrity is not affected upon replacement of Leu99. Upon adding ubiquitin in five-fold excess, we observed some changes in the protease domain of SprT-BR spectra (Fig. 3a, blue boxes). In the L99S variant, the effects of ubiquitin addition were reduced, implying that they correspond to ubiquitin binding to SPRTN's USD interface (Fig. 3b, blue boxes). While the ubiquitin-induced effects were subtle and mostly affected resonances corresponding to the protease domain, we also observed line-broadening for signals corresponding to ZBD (Supplementary Fig. 3e, note Ile212). While Trp ε1 resonances were only marginally affected by the addition of ubiquitin (Fig. 3a, b, zoom-ins), the addition of an activating DNA structure in two-fold excess led to major spectral changes in ZBD-BR regions (Fig. 3c). DNA-induced line-broadening was comparable between WT and L99S constructs (Fig. 3d), demonstrating that alteration of the USD does not affect DNA binding. Strikingly, upon combined addition of both DNA and ubiquitin, severe line-broadening was observed in SprT-BR that was more pronounced than the individual effects of ubiquitin or DNA binding (Fig. 3e, red boxes), suggesting that the simultaneous binding of DNA and ubiquitin has synergistic effects on SPRTN's conformation. These effects were virtually absent in the L99S variant (Fig. 3f, red boxes). Consistently, addition of ubiquitin with a mutated Ile44-patch had little effect (Supplementary Fig. 4a, b).

Collectively, our NMR data indicate that ubiquitin amplifies the effects of DNA binding on SPRTN conformation allosterically by binding to the USD interface at the back of the protease domain. Interestingly, ubiquitin had only small effects on its own, implying that DNA binding occurs first and promotes ubiquitin binding at the USD.

Ubiquitin stimulates DPC cleavage by binding to SPRTN's USD interface

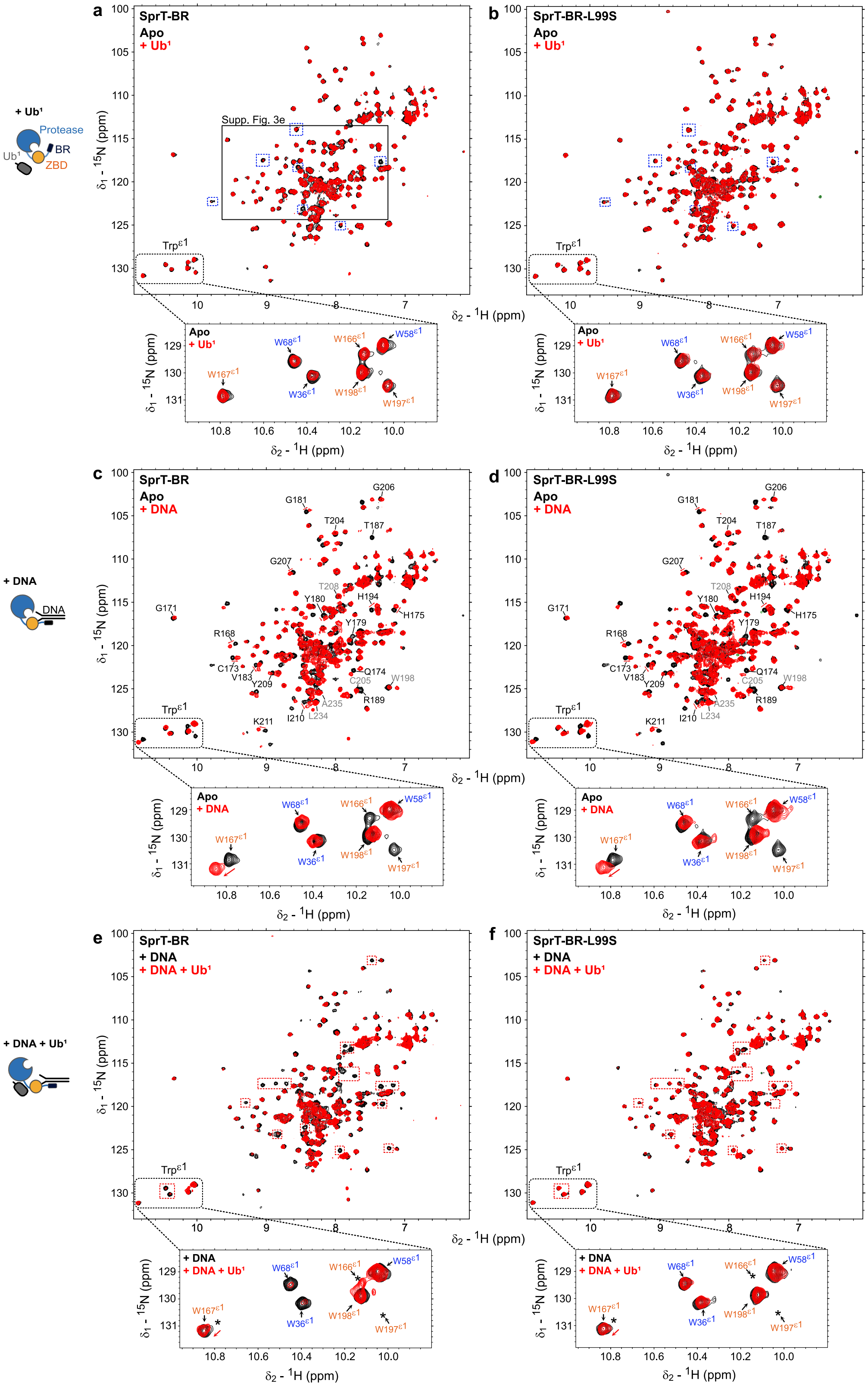
To test whether DPC ubiquitylation stimulates SPRTN activity through binding to the USD interface, we produced full-length SPRTN with an L38S or L99S substitution. Both variants showed cleavage of unmodified HMCES^{SRAP}-DPCs to the same degree as the WT protein (Fig. 4a, compare lanes 3-5, with 6-8 (L38S) and 9-11 (L99S)). While DPC ubiquitylation increased overall activity also in USD mutant variants, the formation of smaller additional cleavage fragments (Cleaved DPC*) observed upon cleavage of ubiquitylated DPCs with the WT protease was reduced (L38S) or almost absent (L99S) (Fig. 4b, c, compare lanes 3-5 with lanes 6-8 (L38S) and lanes 9-11 (L99S)). Combination of the L38S and L99S substitution had no additional effects over the single L99S mutation (Supplementary Fig. 5a, b, compare lanes 6-8 (L99S) with lanes 9-11 (L38S + L99S)). These results suggest that DPC ubiquitylation promotes DPC cleavage through two distinct mechanisms. First, DPC ubiquitylation boosts overall cleavage by SPRTN independent of the USD interface (see Discussion). Second, DPC ubiquitylation allosterically activates SPRTN by binding to the USD interface, enabling the protease to cleave crosslinked proteins more efficiently.

SUMO-targeted DPC ubiquitylation activates SPRTN in vitro and in cells

Encouraged by the strong effects observed using the synthetic DPC ubiquitylation system, we wanted to reconstitute SUMO-targeted DPC ubiquitylation using the enzymes that modify crosslinked proteins in cells. Therefore, we generated DPCs using full-length HMCES protein (HMCES^{FL}); we used HMCES^{FL} because it contains a canonical SUMOylation site in its *C*-terminal tail that is absent in HMCES^{SRAP} constructs. HMCES^{FL}-DPCs were incubated with the SUMOylation machinery, consisting of SUMO-E1, SUMO-E2, SUMO-E3 PIAS4, SUMO2 and ATP (Fig. 5a, b). Successful SUMOylation of the crosslinked protein was indicated by slower migrating HMCES^{FL}-DPC species that were absent in reactions lacking SUMO-E1 (Fig. 5b, compare lanes 3 and 4). For the subsequent ubiquitylation, SUMOylated DPCs were incubated with ubiquitin, ubiquitin-E1, ubiquitin-E2 UBE2D3 and the SUMO-targeted ubiquitin-E3 RNF4 (Fig. 5a, b). Ubiquitylation of SUMOylated DPCs was evident as further upshifts in gel migration and was confirmed by western blot (Fig. 5b, lane 7). We used MS to determine the identity of the ubiquitylated lysine residues and the involved ubiquitin linkages. We identified K48-, K63- and K11-linked ubiquitin chains on SUMOylated DPCs (Fig. 5c), as has been observed in cells[32]. Ubiquitin chains formed on various HMCES lysine residues and on three distinct SUMO2 lysine residues (Fig. 5d). Ubiquitylation was lost in the absence of ubiquitin-E1 or in the absence of SUMOylation (Fig. 5b, lanes 5 and 6 respectively), demonstrating bona fide SUMO-targeted DPC ubiquitylation.

Next, we incubated modified DPCs with SPRTN and FANCJ. Consistent with our results with the synthetic system, we observed enhanced cleavage of the ubiquitylated protein adduct by SPRTN, compared to unmodified DPCs and SUMOylated DPCs (Fig. 5e, compare lanes 3 and 5 with lane 7). Again, additional cleavage products appeared upon DPC ubiquitylation (Fig. 5e, Cleaved DPC*), which were reduced in variants with an altered USD interface (Fig. 5f, compare lanes 3-5 with lanes 6-8 (L38S) and lanes 9-11 (L99S)).

To test whether SUMO-targeted DPC ubiquitylation activates SPRTN also in cells, we monitored the cleavage of DNA methyltransferase 1 (DNMT1)-DPCs induced with 5-azadC[40]. DNMT1-DPCs are swiftly SUMOylated[41], triggering their ubiquitylation by RNF4[9,29,30] and TOPORS[31,32] and, subsequently, cleavage by SPRTN. While *SPRTN*$^{\Delta C}$ cells are viable, they fail to efficiently cleave 5-azadC-induced DNMT1-DPCs[9]. Therefore, we complemented HeLa T-REx Flp-In cells expressing patient-mimicking *SPRTN*$^{\Delta C}$ alleles from the endogenous locus with doxycycline-inducible full length SPRTN variants (WT, E112Q, L38S and L99S) and assessed cleavage of DNMT1-DPCs by the purification of x-linked proteins (PxP) assay (refs. 9,42, Fig. 5g and Methods). DNMT1-DPCs formed in all cell lines upon 5-azadC treatment (Fig. 5g). Following a 2-h chase in drug-free media, a specific cleavage band formed in *SPRTN*$^{\Delta C}$ cells expressing SPRTN-WT but not in cells



expressing catalytically inactive SPRTN-E112Q (Fig. 5g, red dots), as observed previously[9] (DPCs are still resolved in these cells because they are additionally targeted by proteasomal degradation[9,29]). SPRTN-dependent DNMT1-DPC cleavage was strongly reduced in cells expressing SPRTN-L38S or SPRTN-L99S (Fig. 5g, red dots), indicating that SUMO-targeted ubiquitylation promotes DPC cleavage in cells by activating SPRTN at the USD interface.

To corroborate this observation, we additionally assessed 5-azadC-induced SPRTN autocleavage (a marker of SPRTN activation) in the absence of DPC ubiquitylation. To abrogate ubiquitylation of DNMT1-DPCs, we depleted RNF4 using siRNA in HAP1 *TOPORS* knock-out cells. Simultaneous depletion of RNF4 and TOPORS resulted in a complete loss of SPRTN autocleavage (Supplementary Fig. 6a), confirming that DPC ubiquitylation is critical for efficient SPRTN activation in cells.

Fig. 3 | DNA- and ubiquitin-binding affect SPRTN's conformation synergistically. **a**–**f** Comparison of NMR spectra, highlighting Trp ε1 amide signals in ^{1}H,^{15}N-HSQC experiments of SprT-BR and SprT-BR-L99S. Trp ε1 region is labeled and boxed (bottom). Resonance assignments corresponding to the Trp ε1's in the zinc-binding domain (ZBD) are shown in orange and those in the protease domain in blue. Broadened or shifted signals upon dsDNA addition are shown as asterisk. **a**, **b** SprT-BR (**a**) and SprT-BR-L99S (**b**) alone (= Apo) (black), with mono-ubiquitin (Ub^1) (5x molar excess) (red). Minor changes are boxed in blue to highlight the spectral differences between SprT-BR and SprT-BR-L99S upon adding Ub^1. Zoom-in region in Supplementary Fig. 3e is marked with a black box (**b**). **c**, **d** SprT-BR (**c**) and SprT-BR-L99S (**d**) alone (= Apo), with dsDNA (2x molar excess) (red). Some of the ZBD resonances affected by dsDNA are labeled in black while the unchanged are labeled in grey. **e**, **f** Superimpositions of SprT-BR (**e**) and SprT-BR-L99S (**f**) in the presence of dsDNA (2x molar excess) (black) and of both dsDNA (2x molar excess) and Ub^1 (5x molar excess) (red). Additional resonance changes upon adding Ub^1 to the dsDNA-bound SprT-BR are shown with red boxes.

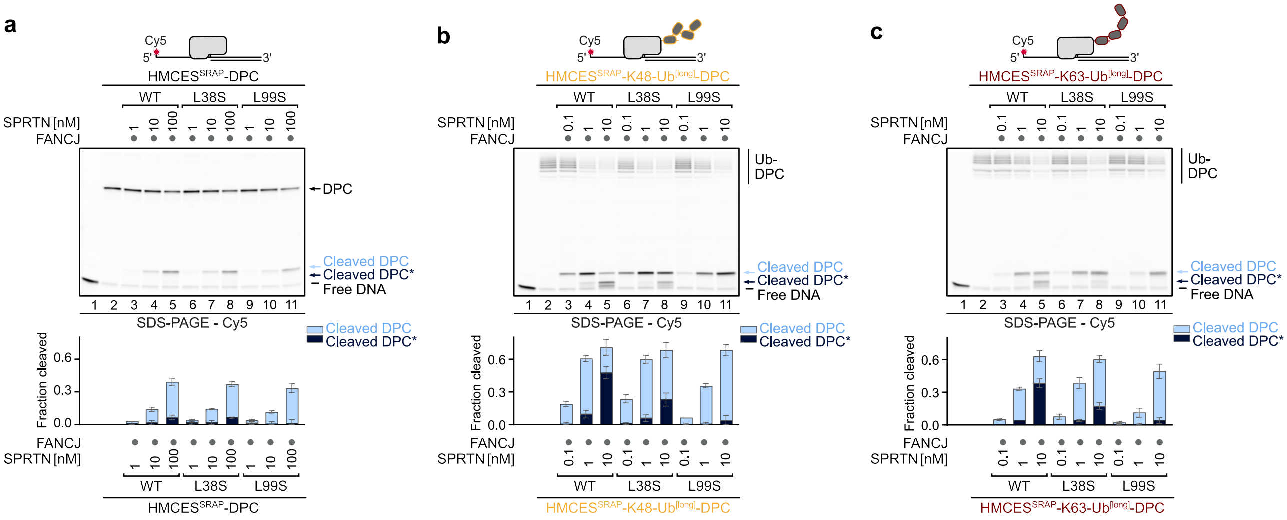


Fig. 4 | The ubiquitin-dependent activation of SPRTN is mediated by the USD. **a**–**c** Indicated HMCES^{SRAP}-DPCs (10 nM) were incubated alone or in the presence of FANCJ (100 nM) and indicated concentrations (0.1–100 nM) and variants of SPRTN (WT, L38S, L99S) for 1 h at 30 °C. Quantification: bar graphs represent the mean ± SD of three independent experiments. Source data are provided as a Source Data file.

Given that DNMT1-DPC repair in cells is compromised upon replacement of critical USD residues and upon loss of SPRTN's *C*-terminal tail in RJALS SPRTN$^{\Delta C}$ patient variants[9], we wanted to examine potential synergistic effects of both alterations using our reconstituted system. We compared cleavage of DPCs modified by SUMO-targeted ubiquitylation by SPRTN^{FL} and SPRTN$^{\Delta C}$ with intact or mutated USD interfaces. While SPRTN$^{\Delta C}$ displayed only slightly reduced DPC cleavage compared to the WT enzyme (Fig. 5h, compare lanes 3-5 with lanes 9-11), the extent of cleavage by SPRTN$^{\Delta C}$ was strongly reduced upon additional replacement of Leu99 by serine (Fig. 5h, compare lanes 9-11 and lanes 18-20). The synthetic cleavage defect of SPRTN$^{\Delta C}$-L99S was only partially explained by loss of the UBZ domain, given that SPRTN$^{\Delta UBZ}$-L99S variant displayed a less pronounced phenotype (Fig. 5h, lanes 15–17). Notably, the defect of SPRTN$^{\Delta C}$ was specific to DPCs modified by SUMO-targeted ubiquitylation. DPCs modified using the synthetic ubiquitylation system were cleaved comparably well by SPRTN$^{\Delta C}$ and the WT enzyme, while a USD mutant variant (L99S) displayed clear defects (Supplementary Fig. 6b and Discussion).

Taken together, our results suggest that SUMO-targeted DPC ubiquitylation allosterically activates SPRTN at the USD interface to promote DPC repair. Our in vitro data further imply that the ubiquitin-dependent activation of SPRTN is specifically important to support the residual cleavage of RJALS SPRTN$^{\Delta C}$ patient variants towards DPCs modified by SUMO-targeted ubiquitylation.

Ubiquitin-dependent activation of SPRTN maintains genome stability in Ruijs-Aalfs syndrome

Next, we wanted to determine whether the ubiquitin-dependent activation of SPRTN at the USD interface is important to maintain the residual function of SPRTN$^{\Delta C}$ patient variants in cells. To this end, we complemented conditional *Sprtn*$^{F/-}$ *CreER*^{T2} knock-out mouse embryonic fibroblasts (MEFs) with either an empty vector (EV) or different human SPRTN variants (FL and ΔC) tagged with a *C*-terminal Strep-tag (Supplementary Fig. 7a, b). Of note, SPRTN$^{\Delta C}$ variants expressed at much higher levels than the WT enzyme (Supplementary Fig. 7a, b), as previously observed in RJALS patients[27]. Loss of endogenous *Sprtn* was induced by 4-hydroxytamoxifen (4-OHT), with the solvent MeOH serving as control (Supplementary Fig. 7c, d), and resulted in diverse phenotypes including growth arrest (Fig. 6a, b), formation of micronuclei and chromatin bridges (Fig. 6c–e), as wells as arrest in the G2/M phase of the cell cycle (Supplementary Fig. 7e–h), as described previously[17]. All phenotypes were rescued by expression of human WT SPRTN but not by catalytically inactive SPRTN-E112Q (Fig. 6a and d). Also, expression of SPRTN$^{\Delta C}$ complemented all phenotypes induced by *Sprtn* knock-out (Fig. 6b and e). While the replacement of USD residues Leu38 or Leu99 had no effect on the ability of full-length SPRTN to complement cell fitness and cell cycle defects upon loss of mouse *Sprtn* (Fig. 6a and Supplementary Fig. 7e), loss of Leu99 resulted in intermediate growth defects and G2/M arrest in SPRTN$^{\Delta C}$ (Fig. 6b and Supplementary Fig. 7f). These defects were accompanied by severe signs of genome instability, observed as micronuclei and chromatin bridges in cells expressing SPRTN$^{\Delta C}$-L99S (Fig. 6c and e).

Collectively, these experiments demonstrate that SPRTN's USD interface, and thus the allosteric activation of SPRTN by ubiquitin, is critical to maintain fitness and genome stability in cells expressing truncated RJALS patient variants.

Discussion

Over the last decade, DPC repair has emerged as a conserved cellular process that is essential for maintaining genome stability[7]. Since the

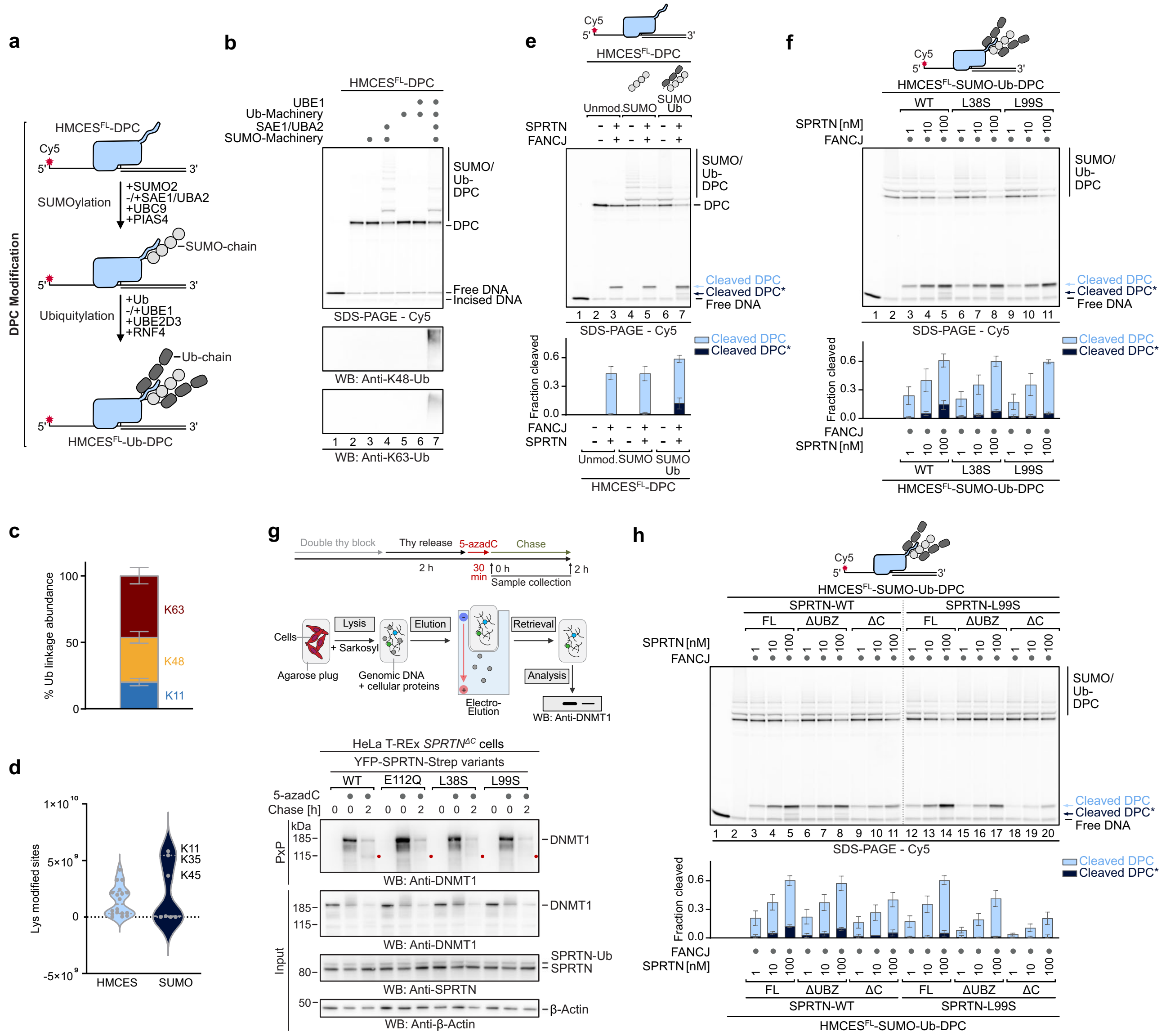


Fig. 5 | SUMO-targeted DPC ubiquitylation activates SPRTN. **a** Schematic of SUMO-targeted ubiquitylation of HMCES^{FL}-DPCs used in b-f and h. HMCES^{FL}-DPCs were incubated alone or in the presence of SUMO2, UBC9 and PIAS4, with or without SAE1/UBA2 for 30 min at 37 °C. Next unmodified or SUMOylated HMCES^{FL}-DPCs were incubated alone or in the presence of ubiquitin (Ub), RNF4, UBE2D3, with or without UBE1 for 30 min at 37 °C. **b** SUMO-targeted ubiquitylated HMCES^{FL}-DPCs generated as described in (**a**), separated by denaturing SDS-PAGE and immunoblotting. Source data are provided as a Source Data file. **c** Mass spectrometry analysis of ubiquitin linkages formed by SUMO-targeted ubiquitylation of HMCES^{FL}-DPCs. Bar chart shows the mean ± SD of four biological replicates. **d** Mass spectrometry analysis of lysine residues within HMCES or SUMO modified upon SUMO-targeted ubiquitylation. Violin blots show the mean ± SD of four biological replicates. **e** Indicated HMCES^{FL}-DPCs (10 nM) were incubated alone or in the presence of FANCJ (100 nM) and SPRTN (100 nM) for 1 h at 30 °C. Quantifications: bar graphs represent the mean ± SD of three independent experiments. Source data are provided as a Source Data file. **f** Indicated HMCES^{FL}-DPCs (10 nM) were incubated alone or in the presence of FANCJ (100 nM) and indicated concentrations (1-100 nM) and variants of SPRTN (WT, L38S, L99S) for 1 h at 30 °C. Quantifications: bar graphs represent the mean ± SD of three independent experiments. All samples derive from the same experiment and gels were processed in parallel. Source data are provided as a Source Data file. **g** HeLa-TREx *SPRTN*$^{\Delta C}$ Flp-In cells complemented with indicated YFP-SPRTN^{FL}-Strep-tag variants were treated as depicted (top) with 5-azadC (10 μM) and harvested at indicated time points. DNMT1-DPCs were isolated using PxP (middle, see Methods) and analyzed by immunoblotting (bottom). Shown is a representative of three independent experiments. Source data are provided as a Source Data file. **h** Indicated HMCES^{FL}-DPCs (10 nM) were incubated alone or in the presence of FANCJ (100 nM) and indicated concentrations (1–100 nM) and variants of SPRTN (FL-WT/L99S, ΔUBZ-WT/L99S, ΔC-WT/L99S) for 1 h at 30 °C. Quantifications: bar graphs represent the mean ± SD of three independent experiments. All samples derive from the same experiment and gels were processed in parallel. Source data are provided as a Source Data file.

identification of dedicated DPC proteases in yeast and humans[10,20,43–48], it has remained enigmatic how specificity for crosslinked protein adducts is achieved. The DPC protease SPRTN features a bipartite DNA binding module, consisting of ZBD and BR, which provides a first layer of specificity by restricting activity to the cleavage of DPCs near ssDNA-dsDNA junctions and other structures with single- and double-stranded features[14,19,21]. However, because such structures occur frequently across the genome, SPRTN's DNA structure-specific activity alone is insufficient to explain how the protease achieves specificity.

Our study reveals that SPRTN activation is controlled by the ubiquitylation of the crosslinked protein by ubiquitin-E3 ligases. By reconstituting DPC ubiquitylation in vitro, we observed that this modification stimulates SPRTN activity by up to two orders of magnitude, regardless of ubiquitin chain linkage type. Our results indicate

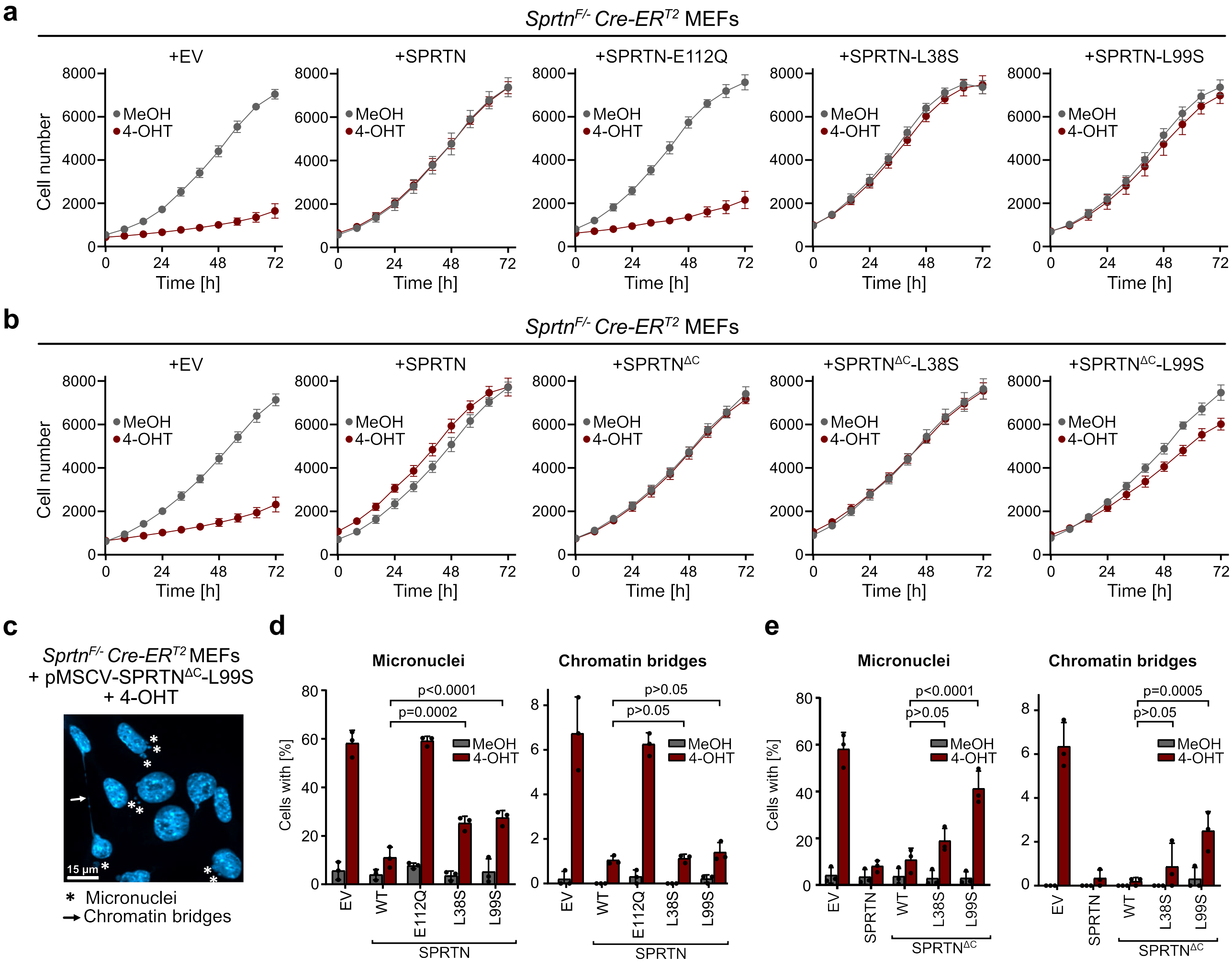


Fig. 6 | Ubiquitin-dependent activation of SPRTN maintains genome stability in Ruijs-Aalfs syndrome. a, **b** Proliferation of *Sprtn*$^{F/-}$ Cre-ER^{T2} mouse embryonic fibroblasts (MEFs) complemented with indicated SPRTN variants or empty vector (EV, pMSCV) treated with methanol (MeOH) or (Z)–4-hydroxytamoxifen (4-OHT) (2 µM) for 48 h. After seeding, cell numbers were counted at indicated time points. Values are the mean ± SD of eight technical replicates. Shown is a representative of three independent experiments. Source data are provided as a Source Data file. **c** Image showing micronuclei (asteriks) and chromatin bridges (arrow) in *Sprtn*$^{F/-}$ Cre-ER^{T2} MEFs + pMSCV-SPRTN$^{\Delta C}$-L99S treated with 4-OHT (2 µM) for 48 h. DNA was visualized by DAPI staining. Scale bar corresponds to 15 µm. **d**, **e** Quantification of micronuclei and chromatin bridges formation in *Sprtn*$^{F/-}$ Cre-ER^{T2} MEFs complemented with indicated SPRTN variants or EV (pMSCV) treated with MeOH or 4-OHT (2 µM) for 48 h. DNA was visualized by DAPI staining. Bar graphs show the mean ± SD of three independent experiments. The p values were calculated using a two-way ANOVA with Dunnett's multiple comparison test. P values: **d** Micronuclei (left): SPRTN-WT vs. SPRTN-L38S = 0.0002; SPRTN-WT vs. SPRTN-L99S < 0.0001. Chromatin bridges (right): SPRTN-WT vs. SPRTN-L38S = 0.9992; SPRTN-WT vs. SPRTN-L99S = 0.8634. **e** Micronuclei (left): SPRTN$^{\Delta C}$-WT vs. SPRTN$^{\Delta C}$-L38S = 0.1411; SPRTN$^{\Delta C}$-WT vs. SPRTN$^{\Delta C}$-L99S < 0.0001. Chromatin bridges (right): SPRTN$^{\Delta C}$-WT vs. SPRTN$^{\Delta C}$-L38S = 0.4745; SPRTN$^{\Delta C}$-WT vs. SPRTN$^{\Delta C}$-L99S = 0.0005. Source data are provided as a Source Data file.

two distinct mechanisms for the activation of SPRTN. On the one hand, ubiquitin activates SPRTN by binding to the USD interface at the back of the protease domain. As a result, the enzyme processes DPCs to a much greater extent, which may be crucial for enabling TLS polymerases to efficiently bypass the remaining peptide adduct during replication-coupled DPC repair[10]. Our MD simulations and NMR data further suggest that the USD-ubiquitin interaction stabilizes a DNA-binding induced open conformation of the enzyme, exposing its active site. In addition, DPC ubiquitylation stimulates overall DPC cleavage independent of the USD interface. The underlying mechanisms remain unclear and are an exciting topic for future research.

These insights help explain why cells ubiquitylate DPCs[9,10,12–15,28–32] and why ubiquitylated DPCs accumulate in cells following SPRTN depletion[33]. Of note, the observation that SPRTN-dependent cleavage can occur without DPC ubiquitylation in frog egg extracts[14,15] is not necessarily inconsistent with our findings. It is plausible that the DPC cleavage observed in egg extract in the absence of ubiquitylation originates from SPRTN's basal, ubiquitin-independent activity, which is also evident in our assays.

Consistently, while amino acid substitutions within the USD interface substantially reduced cleavage of ubiquitylated DPCs in vitro and of DNMT1-DPCs in cells, they did not completely abolish SPRTN function. SPRTN with a replacement of the USD residue Leu99, which consistently showed stronger effects compared to replacing Leu38, suppressed almost all phenotypes caused by the loss of *Sprtn* in MEFs. The same is true for the RJALS SPRTN$^{\Delta C}$ patient variant. Thus, only a minimal amount of SPRTN activity appears to be necessary to fulfil its essential role in suppressing genome instability. The critical role of the USD became evident when Leu99 was replaced in SPRTN$^{\Delta C}$, resulting in cell fitness defects and formation of micronuclei and chromatin bridges in mitosis.

The synthetic effect observed between the combined loss of SPRTN's *C*-terminal tail and a functional USD interface, is only partially explained by the loss of the UBZ domain. While the UBZ is required for

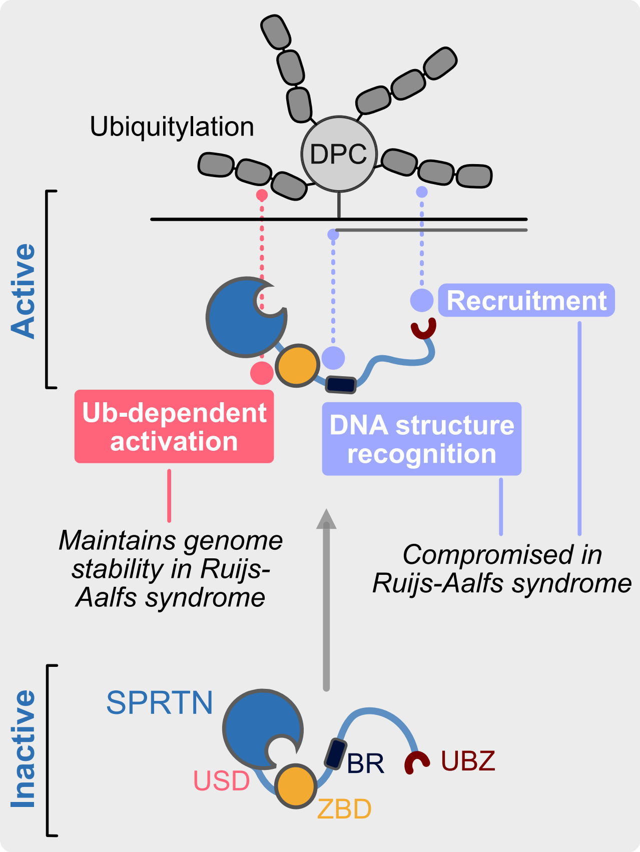


Fig. 7 | 'Triple-lock' model for SPRTN activation. The ubiquitin-binding zinc finger (UBZ) recruits SPRTN to ubiquitylated DPCs. Binding of both DNA-binding domains, zinc-binding domain (ZBD) and the basic region (BR) to activating DNA structures induces an open conformation. This open conformation is stabilized by ubiquitin binding to the ubiquitin binding interface at the SprT domain (USD). Recruitment and DNA structure recognition are compromised in Ruijs-Aalfs syndrome patients, which therefore fully rely on the Ub-dependent activation via the USD to maintain genome stability.

efficient DPC cleavage in cells[9] and egg extracts[14], it showed weaker defects in processing ubiquitylated DPCs in combination with the L99S substitution than the corresponding SPRTN$^{\Delta C}$-L99S variant. In addition to lacking the UBZ domain, SPRTN$^{\Delta C}$ also exhibits reduced DNA binding compared to the FL protein[20,44], which may contribute to its reliance on the USD for full functionality. Based on these considerations, we propose a partially speculative 'triple-lock' model in which SPRTN activity is controlled by at least three mechanisms (Fig. 7). First, the UBZ supports SPRTN function by recruiting it to ubiquitylated DPCs, as previously suggested[9,14]. This recruitment function is likely more important in the crowded environment of a cell than in our in vitro experiments, explaining why the loss of the UBZ had no or only weak phenotypes in our assays. Second, the binding of an activating DNA structure induces an open conformation of SPRTN. Third, this open, active conformation is further stabilized by binding of ubiquitylated DPCs to SPRTN's USD interface, facilitating rapid and complete proteolysis of the crosslinked protein adduct.

This model offers a potential explanation for why SPRTN$^{\Delta C}$ displayed defects in processing DPCs modified by SUMO-targeted ubiquitylation but not of DPCs modified using the synthetic ubiquitylation system. In the synthetic system, the DPC is modified exclusively at the *C*-terminal ubiquitin tag[35]. In contrast, our MS analysis revealed that SUMO-targeted ubiquitylation affects multiple lysine residues within the DPC and the SUMO chain. Some of these ubiquitylation events may hinder SPRTN function by interfering with efficient DNA binding. Thus, SPRTN's full DNA binding capacity is likely required in this context. In cells, potential steric hindrance of DNA access by SPRTN due to DPC ubiquitylation may be overcome by p97-dependent unfolding of the crosslinked protein[49].

The ability of SPRTN to be activated by both K48- and K63-linked ubiquitin chains raises an hypothesis as to why SPRTN is essential, despite acting redundantly with the proteasome in most experimental systems investigated so far[9,14,15]. Given that the proteasome mainly targets substrates modified with K48-linked ubiquitin[50], key endogenous substrates of SPRTN may be modified by K63-linked ubiquitin and are consequently not amenable to proteasomal degradation.

In conclusion, we have demonstrated that ubiquitin does not simply recruit SPRTN but allosterically activates the enzyme, which is essential for maintaining genome stability in RJALS patients. This sophisticated mechanism likely evolved to constrain the potentially toxic activity of SPRTN and represents a unique ubiquitin-dependent regulatory principle in DNA repair.

Methods

Mammalian cell lines

HeLa TREx Flp-In *SPRTN*$^{\Delta C}$ cells (ref. 9) stably expressing YFP-SPRTN-Strep-tag 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 Scientific). Protein expression was induced by overnight incubation with doxycycline hyclate (DOX) (D9891, Sigma) (final concentration 1 µg/mL). HeLa TREx Flp-In *SPRTN*$^{\Delta C}$ cells were grown in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% (v/v) fetal bovine serum (FBS).

HAP1 wild-type (WT) cells (C631, Horizon Discovery) and HAP1 *TOPORS* knock-out (KO) cells (HZGHC008005c006, Horizon Discovery) were grown in Iscove's Modified Dulbecco's Medium (IMDM) supplemented with 10% (v/v) FBS and 1% (v/v) Penicillin-Streptomycin-Glutamine (PSG).

Sprtn$^{F/-}$ mouse embryonic fibroblasts (MEFs) (H7) immortalized with SV40 large T and transduced with a retroviral vector expressing Cre-ER^{T2} (ref. 17) were cultured in DMEM supplemented with 10% (v/v) FBS. *Sprtn* KO was induced by treating 4×10^5 cells with methanol (MeOH) (vehicle control) or 2 µM (Z)−4-hydroxytamoxifen (4-OHT) (Ht7904, Sigma) for 48 h. Conversion of the floxed *Sprtn* allele **f** to the KO allele (-) was verified by PCR using WT- (5'-GTGCTGGGATCTGCACCTAT-3') and KO-specific (5'-CCATCAGGGACGTTTTCTTG-3') forward primers and a common reverse primer (5'-TGCACAGCTGTAAACCCTTG-3'). PCR conditions were 35 cycles of 94 °C for 30 s, 60 °C for 30 s, and 72 °C for 1 min. PCR products are 527 bp and 278 bp for the floxed and the KO alleles, respectively. For exogenous expression of human SPRTN in MEFs, cells were infected with retroviral vectors produced in HEK293T/17 (CRL-11268, ATCC) by co-transfecting pMSCV.hyg-SPRTN-Strep with gag-pol and VSV-G packaging plasmids. Infected cells were selected with Hygromycin B (200 µg/mL) (10687010, Thermo Scientific) for 8 days.

To confirm protein expression, cells were lysed in NP-40 lysis buffer (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 0.1% NP-40 (IGEPAL) (I8896, Sigma), 10% glycerol, 5 mM EDTA (BP118-500, Fisher BioReagents), 50 mM NaF, 1 mM Na_3VO_4) supplemented with protease inhibitor cocktail (P8849, Sigma). Cell lysates containing 30 µg protein were resolved on SDS-PAGE gels (4-12% Bis-Tris NuPAGE, Thermo Scientific) using MOPS buffer. Resolved proteins were subsequently immunoblotted using anti-SPRTN antibody (1:500) (6F2) (ref. 26) and anti-β-actin antibody (1:1000) (Sc47778, Santa Cruz Biotechnology).

Protein expression and purification

SPRTN. Amino acid replacements and deletions for SPRTN variants were generated using the Q5-site-directed mutagenesis kit (E0554S, NEB). Recombinant SPRTN (Full-length and ΔUBZ – WT or in combination with L38S, L99S amino acid replacements) protein was expressed in BL21 (DE3) *E. coli* cells (C600003, Thermo Scientific) and purified as previously described with slight modifications[21].

BL21 (DE3) *E. coli* cells were grown at 37 °C in Terrific broth (TB) medium (prepared with tap water) until they reached OD 0.7. Protein expression was induced by addition of 0.5 mM Isopropyl-β-D-thiogalactoside (IPTG) (I6758, Sigma) overnight at 18 °C. The next day, cells were harvested, snap-frozen in liquid nitrogen and stored at −80 °C. All subsequent steps were carried out at 4 °C. For protein purification, cell pellets were resuspended in buffer A (50 mM HEPES/KOH pH 7.2, 500 mM KCl, 1 mM $MgCl_2$, 10% Glycerol, 0.1% IGEPAL (I8896, Sigma), 0.04 mg/mL Pefabloc SC (76307, Sigma), cOmplete EDTA-free protease inhibitor cocktail tablets (4693132001, Roche), 1 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP)) (HN95.3, Roth) and lysed by sonication. Cell lysate was incubated with smDNAse (45 U/mL lysate) (MPI for Biochemistry) for 30 min on a roller prior to removal of cell debris by centrifugation at 18,000 × *g* for 30 min. Cleared supernatant was filtered using syringe filters (PVDF, 0.22 μm) and applied to Strep-Tactin®XT 4Flow® high-capacity cartridges (2-5028-001, IBA Lifesciences), washed with 3 column volumes (CV) of buffer A and 4 CV of buffer B (50 mM HEPES/KOH pH 7.2, 500 mM KCl, 10% Glycerol, 1 mM TCEP (HN95.3, Roth)). Proteins were eluted in 6 CV buffer C (50 mM HEPES/KOH pH 7.2, 500 mM KCl, 10% Glycerol, 1 mM TCEP (HN95.3, Roth) and 50 mM Biotin). Eluted proteins were further applied to HiTrap Heparin HP affinity columns (17040701, Cytiva) and washed with 3 CV buffer B before eluting in buffer D (50 mM HEPES/KOH pH 7.2, 1 M KCl, 10% Glycerol, 1 mM TCEP (HN95.3, Roth)). Eluted fractions containing recombinant SPRTN protein were desalted against buffer B using PD-10 desalting columns (17085101, Cytiva). The affinity tag was cleaved off at 4 °C overnight by addition of His-tagged TEV protease (ref. 37) with 1:10 mass ratio. Cleaved recombinant SPRTN protein was further purified by size exclusion chromatography (SEC) using a HiLoad 16/600 Superdex 200 pg column (28989335, Cytiva) equilibrated in buffer B (50 mM HEPES/KOH pH 7.2, 500 mM KCl, 10% Glycerol, 1 mM TCEP (HN95.3, Roth)). Eluted proteins were concentrated with 10 kDa cut-off Amicon Ultra centrifugal filters (UFC801096, Merck) before aliquoting, snap-freezing in liquid nitrogen and storage at −80 °C.

Following SPRTN purification, metalation of the protein was examined by Inductively Coupled Plasma Optical Emission Spectrometry (ICP-OES) (see Supplementary Table 1), which confirmed correct metalation with three Zn^{2+} ions per full-length SPRTN molecule.

For truncated SPRTN variants smaller than 30 kDa including $SPRTN^{ΔC}$ (WT or L99S), SprT-BR (WT, L99S, W36G and W58G), ZBD-BR and ZBD, a Strep-tagged TEV protease (ref. 9) was used. Prior to SEC, Strep-tagged TEV protease, residual uncleaved protein and the cleaved Tag were removed by a Strep-Tactin®XT 4Flow® high capacity cartridges (2-5028-001, IBA Lifesciences)[9].

For NMR experiments, SprT-BR (L99S, W36G and W58G), ZBD, and ZBD-BR were expressed in ^{15}N- or ^{13}C-/^{15}N-containing media. Here, cells were grown to OD 0.4, before the temperature of the incubator was lowered to 18 °C and $MnCl_2$ was added to a final concentration of 1.5 mM. Once OD 0.7 was reached protein expression was induced with 0.5 mM IPTG (I6758, Sigma) and performed overnight at 18 °C. For SEC, buffer E (50 mM HEPES/KOH pH 7.2, 500 mM KCl, 1% Glycerol, 2 mM TCEP (HN95.3, Roth), pH 7.2) was used.

Mono-Ubiquitin. For purification of mono-ubiquitin (Ub^1) a plasmid encoding Ub^1 with a *N*-terminal His6-Tag was provided by Brenda Schulman (MPI for Biochemistry, Martinsried, Germany). Ub^1-I44A was generated by introducing point mutations using the Q5-site-directed mutagenesis kit (E0554S, NEB). Protein was expressed in Rosetta *E. Coli* cells (70-954-3, Sigma), grown at 37 °C in TB (prepared with tap water) to OD 0.7. Protein expression was induced with 0.5 mM IPTG (I6758, Sigma) overnight at 18 °C. Cells were harvested the next day and directly resuspended in buffer A (50 mM Tris-HCl pH 7.5, 250 mM NaCl) (20 mL/ L culture), snap-frozen in liquid nitrogen and stored at −80 °C. All subsequent steps were carried out at 4 °C. For protein purification, cell lysates were thawed and Pefabloc SC (0.04 mg/mL) (76307, Sigma), $MgCl_2$ (1 mM) and smDNAse (45 U/mL lysate) (MPI for Biochemistry) were added. Cells were lysed by sonication and incubated for 30 min on a roller prior to removal of cell debris by centrifugation at 50,000 × *g* for 30 min. Clarified lysate was filtered using syringe filters (PVDF, 0.22 μm) and mixed with Ni-NTA Agarose (30250, Qiagen) equilibrated in buffer A and incubated for 1 h on a roller to allow binding. The beads were transferred to a gravity flow column, washed with 15 CV of buffer A and protein was eluted in fractions of 1 CV each with buffer B (50 mM Tris-HCl pH 7.5, 250 mM NaCl, 300 mM imidazole (3899.1, Roth). Fractions were checked via SDS-PAGE and Coomassie-based staining for presence of Ub^1. Ub^1-containing fractions were pooled and after addition of GST-tagged 3C-protease (0.5 mg/L culture) (MPI for Biochemistry), dialyzed against buffer A overnight. Cleaved protein was passed through Ni-NTA Agarose (30250, Qiagen) the next day for removal of uncleaved protein and His6-tag. The flow-through was collected, concentrated to 1 mL and loaded on a Superdex 200 Increase 10/300 GL column (28990944, Cytiva) equilibrated in buffer C (50 mM HEPES/KOH pH 7.2, 500 mM KCl, 1% Glycerol, 2 mM TCEP (HN95.3, Roth)). Eluted protein was concentrated with 10 kDa cut-off Amicon ultra centrifugal filters (UFC801096, Merck) before aliquoting, snap-freezing in liquid nitrogen and storage at −80 °C.

FANCJ. Recombinant FANCJ protein was expressed in High Five™ cells (B85502, Thermo Scientific) and purified as previously described[38].

$HMCES^{SRAP}$. Recombinant $HMCES^{SRAP}$, protein was expressed in BL21 (DE3) *E. coli* cells (C600003, Thermo Scientific) and purified as previously described[38], using TB (prepared with tap water). For synthetic ubiquitylation of $HMCES^{SRAP}$, a sequence encoding for Ub^1(G76V) followed by an FKBP-domain, including linkers and a 3C-protease cleavage site was codon optimized for bacterial expression and inserted at the *C*-terminal end of $HMCES^{SRAP}$, in front of the His6-tag, in the pNIC_$HMCES^{SRAP}$ plasmid. Purification followed protocols described for $HMCES^{SRAP}$ and the final protein was further processed as described below.

$HMCES^{FL}$. Recombinant $HMCES^{FL}$ protein was expressed in BL21 (DE3) *E. coli* cells (C600003, Thermo Scientific) and purified as previously described[38], analogously to recombinant SPRTN using TB (prepared with tap water).

UBC9. For purification of recombinant UBC9, the open reading frame was codon optimized and cloned in a pBAD plasmid carrying a *N*-terminal His6-tag. Protein was expressed in BL21(DE3) *E. coli* cells (C600003, Thermo Scientific) and grown in TB (prepared with tap water) at 37 °C to OD 0.7 before induction of protein expression with 0.1% L-arabinose (A3256, Sigma) at 18 °C overnight. Cells were harvested the next day, snap-frozen in liquid nitrogen and stored at −80 °C. All subsequent steps were performed at 4 °C. For protein purification, cell pellets were thawed, resuspended in buffer A (20 mM HEPES/KOH pH 7.0, 2 mM $Mg(OAc)_2$, 300 mM KOAc, 10% glycerol, 30 mM imidazole (3899.1, Roth), 0.1% IGEPAL (I8896, Sigma), 1 mM TCEP (HN95.3, Roth), cOmplete protease inhibitor (4693132001, Roche), 0.04 mg/mL Pefabloc SC (76307, Sigma), 1 mg/mL lysozyme (8259.3, Roth), 45 U/mL smDNAse (MPI for Biochemistry) and incubated on a roller for 30 min. The lysate was sonicated for 15 min prior to cell debris removal by centrifugation at 18,000 x *g* for 40 min. Clarified lysate was filtered using syringe filters (PVDF, 0.22 μm) and incubated with Ni-NTA agarose (30250, Qiagen) on a roller for 1 h at

4 °C. The beads were transferred to a gravity flow column and washed with 15 CV buffer B (20 mM HEPES/KOH pH 7.0, 2 mM $Mg(OAc)_2$, 300 mM KOAc, 10% glycerol, 30 mM imidazole (3899.1, Roth)) before elution in 2 CV buffer C (20 mM HEPES/KOH pH 7.0, 2 mM $Mg(OAc)_2$, 300 mM KOAc, 10% glycerol, 300 mM imidazole (3899.1, Roth)). The His6-tag was cleaved by the addition of His-tagged TEV protease (1 mg/L culture) (ref. 37) and dialyzed overnight against buffer D (20 mM HEPES/KOH pH 7.0, 2 mM $Mg(OAc)_2$, 300 mM KOAc). The next day, cleaved protein was passed through Ni-NTA agarose (30250, Qiagen) to remove His-tagged TEV protease, residual uncleaved protein and His6-Tag. Flow-through was concentrated to 1 mL and loaded on a Superdex 200 Increase 10/300 GL column (28990944, Cytiva) equilibrated in buffer E (20 mM HEPES/KOH pH 7.0, 2 mM $Mg(OAc)_2$, 300 mM KOAc, 10% glycerol). Eluted protein was concentrated with 10 kDa cut-off Amicon ultra centrifugal filters (UFC801096, Merck) before aliquoting, snap-freezing in liquid nitrogen, and storage at −80 °C.

PIAS4. The open reading frame of PIAS4 was codon optimized and cloned in a pNIC plasmid in frame with a *N*-terminal TwinStrep-ZB-tag. Recombinant PIAS4 protein was expressed in BL21 (DE3) *E. coli* cells (C600003, Thermo Scientific), grown in TB (prepared with tap water) at 37 °C to OD 0.7 before induction with 1 mM IPTG (I6758, Sigma) and expression at 18 °C overnight. Protein purification was done analogously to SPRTN.

UBE2D3. For purification of UBE2D3, the open reading frame was codon optimized and cloned into a pDEST17 plasmid carrying an *N*-terminal His6-tag. UBE2D3 was expressed in BL21(DE3) *E. coli* cells (C600003, Thermo Scientific), grown in TB media (prepared with tap water) to an OD of 0.7 at 37 °C. Expression was induced by the addition of 0.5 mM IPTG (I6758, Sigma) for 3 h at 37 °C. Cells were harvested, snap frozen in liquid nitrogen and stored at −80 °C. All subsequent steps were performed at 4 °C. For protein purification, cell pellets were thawed and resuspended in 50 mL buffer A (50 mM Na_2HPO_4/NaH_2PO_4 pH 8.0, 150 mM NaCl, 10 mM imidazole (3899.1, Roth), 1 mM TCEP (HN95.3, Roth), cOmplete protease inhibitor (4693132001, Roche), 0.04 mg/mL Pefabloc SC (76307, Sigma)) and lysed by sonication. DNA was digested by the addition of smDNAse (45 U/mL lysate) (MPI for Biochemistry) for 30 min on a roller, followed by centrifugation at 18,000 x *g* for 30 min to remove cell debris. Clarified lysate was filtered using syringe filters (PVDF, 0.22 μm) and incubated with Ni-NTA agarose (30250, Qiagen) on a roller for 1 h at 4 °C. The beads were washed with 20 mL buffer B (50 mM Na_2HPO_4/NaH_2PO_4 pH 8.0, 500 mM NaCl, 20 mM imidazole (3899.1, Roth), 1 mM TCEP (HN95.3, Roth)) and eluted in 5 mL buffer C (50 mM Na_2HPO_4/NaH_2PO_4 pH 8.0, 500 mM NaCl, 250 mM imidazole, 1 mM TCEP (HN95.3, Roth)). The eluted protein was dialyzed against buffer D (20 mM Tris/HCl pH 7.5, 150 mM NaCl, 10% glycerol, 0.5 mM TCEP (HN95.3, Roth)) overnight followed by SEC on a Superdex 200 Increase 10/300 GL column (28990944, Cytiva) equilibrated in buffer D. Eluted protein was concentrated using 10 kDa cut-off Amicon ultra centrifugal filters (UFC801096, Merck) before aliquoting, snap-freezing in liquid nitrogen and storage at −80 °C.

RNF4. For purification of recombinant RNF4, the open reading frame was codon optimized and cloned in a pNIC plasmid in frame with a *N*-terminal TwinStrep-ZB-tag. RNF4 protein was expressed in BL21 (DE3) *E. coli* cells (C600003, Thermo Scientific), grown in TB (prepared with tap water) and purified analogously to SPRTN. For SEC, buffer E (50 mM HEPES/KOH pH 7.2, 150 mM KCl, 10% glycerol, 1 mM TCEP (HN95.3, Roth)) was used.

SUMO2. For purification of recombinant SUMO2, the open reading frame was codon optimized and cloned in a pBAD plasmid carrying a *N*-terminal His6-tag. SUMO2 was expressed in BL21 (DE3) *E. coli* cells (HN95.3, Roth) and grown in TB (prepared with tap water) at 37 °C to OD 0.7 before induction with 0.02% L-arabinose (A3256, Sigma) and expression at 18 °C overnight. Cells were harvested the next day, snap-frozen in liquid nitrogen and stored at −80 °C. All subsequent steps were performed at 4 °C. For protein purification, cell pellets were thawed, resuspended in buffer A (50 mM Na_2HPO_4/NaH_2PO_4 pH 7.5, 500 mM NaCl, 10% glycerol, 30 mM imidazole (3899.1, Roth), 0.2% Triton-X-100 (T8787, Sigma), 1 mM TCEP (HN95.3, Roth), cOmplete protease inhibitor (4693132001, Roche), 0.04 mg/mL Pefabloc SC (76307, Sigma), 1 mg/mL lysozyme (8259.3, Roth), 45 U/mL smDNAse (MPI for Biochemistry)) and incubated on a roller for 30 min. Cell lysate was sonicated for 15 min before removal of cell debris by centrifugation at 18,000 x *g* for 30 min. Clarified lysate was filtered using syringe filters (PVDF, 0.22 μm) and incubated with Ni-NTA agarose (30250, Qiagen) on a roller for 1 h at 4 °C. The beads were transferred to a gravity flow column and washed with 15 CV buffer B (50 mM Na_2HPO_4/NaH_2PO_4 pH 7.5, 500 mM NaCl, 10% glycerol, 30 mM imidazole (3899.1, Roth)) before elution in 2 CV buffer C (50 mM Na_2HPO_4/NaH_2PO_4 pH 7.5, 500 mM NaCl, 10% glycerol, 300 mM imidazole (3899.1, Roth)). The His-tag was cleaved by the addition of His-tagged TEV protease (1 mg/L culture) (ref. 37). The protein was dialyzed against buffer D (20 mM HEPES/KOH pH 7.5, 100 mM KCl) overnight. The next day, cleaved protein was passed through Ni-NTA agarose (30250, Qiagen) to remove His-tagged TEV protease, residual uncleaved protein and the His-Tag. Flow-through was concentrated to 1 mL and loaded on a Superdex 200 Increase 10/300 GL column (28990944, Cytiva) equilibrated in buffer E (20 mM HEPES/KOH pH 7.5, 100 mM KCl, 10% glycerol, 1 mM TCEP (HN95.3, Roth)). Eluted protein was concentrated with 3 kDa cut-off Amicon ultra centrifugal filters (UFC8003, Merck) before aliquoting, snap-freezing in liquid nitrogen and storage at −80 °C.

In vitro HMCES-DPC generation

DPCs were generated between $HMCES^{SRAP}$, $HMCES^{SRAP}$-K48-$Ub^{[short]/[long]}$, $HMCES^{SRAP}$-K63-$Ub^{[short]/[long]}$ or $HMCES^{FL}$ and a 30nt Cy5-labeled forward oligonucleotide (5′-Cy5-CCCAAAAAAAAAAAdUAAAAAAAAAAAACCC-3′), as previously described[37,38]. For $HMCES^{FL}$-DPCs final concentrations differed from published protocols: $HMCES^{FL}$ (13 μM), UDG (0.1 U/μL) (M0280L, NEB), DNA (1.25 μM). For all crosslinking reactions, incubation was shortened to 30 min at 37 °C. To form ssDNA-dsDNA junctions 1 μL complementary 15nt reverse oligonucleotide (5′-GGGTTTTTTTTTTTT-3′) (12 μM in nuclease-free H_2O) was annealed to all crosslinking reactions.

$HMCES^{SRAP}$ Ubiquitylation using synthetic ubiquitin E3 ligases

A simplified Ubiquiton system[35], based on fusions of a complete ubiquitin instead of split-ubiquitin as a starting point, was used. In brief, $HMCES^{SRAP}$-Ub(G76V)−3C-FKBP-His6 was K48-poly-ubiquitylated in a reaction containing substrate (10 μM), ubiquitin (30 μM) (U6253, Merck), Ub^1-K48R (10 μM) (IMB gGmbH), His-Uba1 (50 nM) (refs. 51,52), Ubc7-His (E2) (4 μM) (refs. 51,52), His-FRB-E3[48] (10 μM) (IMB gGmbH), ATP (1 mM) (R0441, Thermo Scientific) and rapamycin (50 μM) (SEL-S1039, Biozol) in ubiquitylation buffer (40 mM HEPES/NaOH pH 7.4, 50 mM NaCl, 8 mM $Mg(OAc)_2$) for 6.5 h at 30 °C. K48-modified $HMCES^{SRAP}$ was separated from other reaction components by cleaving the dimerization tag using His-3C-protease (IMB gGmbH) at 4 °C overnight, reverse immobilized metal affinity chromatography (IMAC) and SEC (20 mM HEPES/KOH pH 7.8, 150 mM KCl, 2 mM $MgCl_2$, 1 mM TCEP (HN95.3, Roth), 10% glycerol).

$HMCES^{SRAP}$-Ub(G76V)−3C-FKBP-His6 was K63-poly-ubiquitylated in a reaction containing substrate (10 μM), ubiquitin (30 μM) (U6253, Merck), Ub^1-K63R (10 μM) (IMB gGmbH), His-Uba1 (50 nM) (refs. 51,52), His-Ubc13·Mms2 (E2) (2 μM) (refs. 51,52), His-FRB-L20-E3[63] (10 μM) (IMP gGmbH), ATP (1 mM) (R0441, Thermo Scientific) and rapamycin

(50 μM) (SEL-S1039, Biozol) in ubiquitylation buffer (40 mM HEPES/NaOH pH 7.4, 50 mM NaCl, 8 mM $Mg(OAc)_2$) for 2 h at 30 °C and purified as described above.

Ubiquitin mutants, His-Uba1, Ubc7-His, His-Ubc13 and Mms2 were purified as previously described[51,52]. His-FRB-E3[48] and His-FRB-L20-E3[63] were produced in *E. coli* and purified by IMAC followed by SEC (20 mM HEPES/NaOH pH 7.4, 150 mM NaCl, 10% glycerol, 1 mM DTT (D0632, Merck).

In vitro SUMOylation and ubiquitylation of HMCES-DPCs

SUMOylation reactions were performed in 20 μL for 30 min at 37 °C, containing HMCES-DPC (125 nM), SUMO2 (1.250 μM), SAE1/UBA2 (100 nM) (NKM-ATGP3363, Hölzel), UBC9 (200 nM) and PIAS4 (125 nM). The reaction buffer comprised 20 mM HEPES/KOH pH 7.5, 110 mM KOAc, 5.32 mM $MgCl_2$, 2 mM $Mg(OAc)_2$, 0.05% TWEEN20 (P7949, Sigma), 0.2 mg/ml BSA (AM2616, Thermo Scientific), 1 mM TCEP (HN95.3, Roth), 2.5 mM ATP (R0441, Thermo Scientific). If no further reactions were carried out 5 μL reaction buffer were added and DPCs were either used in DPC cleavage assays or directly mixed with 4x LDS sample buffer (NP0007, Thermo Scientific) supplemented with 5% β-mercaptoethanol (β-ME) (M3148, Sigma), followed by boiling for 1 min at 95 °C prior to SDS-PAGE analysis. For subsequent ubiquitylation, 5 μL ubiquitin master mix were added, and reactions were incubated for 30 min at 37 °C. The ubiquitin master mix contained mono-ubiquitin (1 μM), UBE1 (100 nM) (182UB101, Lifesensors), RNF4 (200 nM) and UBE2D3 (200 nM). DPCs were either used in DPC cleavage assays or directly mixed with 4x LDS sample buffer (NP0007, Thermo Scientific) supplemented with 5% β-ME (M3148, Sigma), followed by boiling for 1 min at 95 °C prior to SDS-PAGE analysis. Samples were resolved on SDS-PAGE gels (12% Bis-Tris BOLT, Thermo Scientific) using MOPS buffer. Gels were scanned using a BioRad Chemidoc MP system with appropriate filter settings for Cy5 fluorescence. Gels were subsequently analyzed by immunoblotting using anti-K48-Ub (D9D5) (1:500) (8081S, Cell Signaling) and anti-K63-Ub (D7A11) (1:500) (5621S, Cell Signaling) antibodies.

For analysis of SUMOylated and ubiquitylated $HMCES^{FL}$-DPC by mass spectrometry (MS), reactions were scaled up to 50 μL, ubiquitin concentration was increased (5 μM) and incubation time for ubiquitylation was extended (1 h at 37 °C). Reactions were stopped by addition of 4x LDS sample buffer (NP0007, Thermo Scientific) supplemented with 5% β-ME (M3148, Sigma). Samples were stored at −20 °C until MS analysis.

DPC cleavage assay

DPC cleavage by SPRTN was assessed in 10 μL reactions at 30 °C for 1 h, containing SPRTN (WT or variants, as indicated – concentrations ranging from 0.1–100 nM), DPC or free DNA (10 nM) with or without FANCJ (100 nM) and with or without free K48-linked tetra-ubiquitin (Ub^4) (SI4804, Lifesensors) or K63-Ub^4 (SI6304, Lifesensors) (400 nM, referring to concentrations of individual ubiquitin moieties). The reaction buffer comprised 17.1 mM HEPES/KOH pH 7.5, 85.6 mM KCl, 3.1% glycerol, 5 mM TCEP (HN95.3, Roth), 2.1 mM $MgCl_2$, 0.12 mg/ml BSA (AM2616, Thermo Scientific) and 1 mM ATP (R0441, Thermo Scientific). Reactions were stopped with 4x LDS sample buffer (NP0007, Thermo Scientific) supplemented with 5% β-ME (M3148, Sigma) and boiling for 1 min at 95 °C. Samples were resolved on SDS-PAGE gels (12% Bis-Tris BOLT, Thermo Scientific) using MOPS buffer. Gels were scanned using a BioRad Chemidoc MP system with appropriate filter settings for Cy5 fluorescence. DPC cleavage was quantified using ImageJ (v1.54 f), by dividing the amount of cleaved DPCs by the total amount of DPC (cleaved + uncleaved). For Cleaved DPC*, the sum of cleavage fragment signals and the corresponding signal for free DNA was calculated minus free DNA signals inferred from control DPC reactions.

Cellular SPRTN autocleavage assays

For cellular SPRTN autocleavage assays, 1×10^6 cells were seeded in 6-well plates. 24 h later 4 μL siRNA (20 μM) and 20 μL Lipofectamine RNAiMAX Transfection Reagent (13778075, Thermo Scientific) were each diluted in 200 μL Opti-MEM serum-free 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. 24 h after transfection, each well was split in 4 wells of a 12-well plate. The next morning, cells were treated with 5-azadC (10 μM) (A3656, Sigma) for 2, 4 or 8 h or left untreated for each transfected siRNA. At desired time points, cells were directly lysed in 1x LDS (NP0007, Thermo Scientific) and boiled for 20 min at 95 °C. Samples were resolved on SDS-PAGE gels (4-12% Bis-Tris NuPAGE, Thermo Scientific) using MOPS buffer and subsequently immunoblotted using anti-DNMT1 antibody (1:1000) (#5032, Cell Signaling), anti-SPRTN antibody (1:500) (6F2) (ref. 26), anti-RNF4 antibody (1:500) (AF7964, R&D systems) and anti-Vinculin antibody (1:1000) (sc-73614, Santa Cruz Biotechnology). The following siRNAs (Horizon Discovery) were used: siCTRL (Control pool, D-001810-10-20), siRNF4 (SMARTpool,L-006557-00-0005).

Purification of x-linked proteins (PxP)

For PxP experiments, 7.5×10^5 cells were seeded in 6-cm dishes, and thymidine-containing media (2 mM) (T9250, Sigma) was added after 8 h. After approximately 16 h, thymidine media was removed, and cells were washed twice with PBS and released into normal media for 9 h, before thymidine media was re-added and expression of YFP-SPRTN-Strep-tag variants was induced with DOX (1 μg/mL) (D9891, Sigma). After another 16 h in thymidine media, cells were washed twice with PBS and released into normal media for 2 h before adding fresh media containing 5-azadC (10 μM) (A3656, Sigma). After a 30 min incubation, 5-azadC containing media was removed, cells were washed twice with PBS and recovery was allowed for 2 h. Cells were scraped on ice at indicated time points and cell pellets were stored at −80 °C. PxP to isolate DNMT1-DPCs was performed as previously described[9,42]. In brief, 10 μL of each cell suspension were directly lysed in 1x LDS sample buffer (NP0007, Thermo Scientific) to serve as input samples before plug casting. 1.5×10^6 cells were embedded into low-melt agarose (1613111, Bio-Rad) plugs, extracted by PxP[9,42] and prepared for analysis by SDS-PAGE at the end of the protocol. Samples were resolved on SDS-PAGE gels (4-12% Bis-Tris NuPAGE, Thermo Scientific) using MOPS buffer and subsequently immunoblotted using anti-DNMT1 antibody (1:1000) (#5032, Cell Signaling), anti-SPRTN antibody (1:500) (6F2) (ref. 26), and anti-β-actin antibody (1:1000) (Sc47778, Santa Cruz Biotechnology).

Cell viability

For cell proliferation assays, 1,000 cells were seeded per well (n = 8) in a 96-well plate, and cell numbers were recorded every 8 h for 3 days using Cytation 5 (BioTek) equipped with a 4x objective and the Gen5 software (ver. 3.14).

Flow cytometry

Cells were labeled with EdU (10 μM) for 45 min. EdU staining was performed with the Click-iT EdU Alexa Fluor 488 Flow Cytometry Assay Kit (C10425, Thermo Scientific) following the manufacturer's protocol. Cells were next stained with 4′,6-diamidino-2-phenylindole (DAPI) (4 μg/mL) (62248, Thermo Scientific) and analyzed using the BD LSRFortessa cell analyzer (BD Biosciences) with the FACSDiva™ software (ver. 6.2). Figures were generated using FlowJo (ver. 10.10).

Microscopy

Cells grown on a cover glass were washed once with PBS, fixed with paraformaldehyde (4%) (P6148, Sigma) in PBS for 10 min, and permeabilized with 0.2% Triton X-100 (T8787, Sigma) in PBS for 10 min.

After washing with PBS, cells were stained with DAPI (1 µg/mL) (62248, Thermo Scientific) in PBS for 10 min, and the cover glass was mounted with ProLong Glass Antifade Mountant (P36980, Thermo Scientific). Images were captured using a Zeiss Axio Observer 7 equipped with Apotome 3 and the Axiocam 820 camera. At least 300 DAPI-stained nuclei were scored manually for the presence of micronuclei or chromatin bridges by an observer blinded to sample identities. Statistical analysis was performed by two-way ANOVA with Dunnett's multiple comparison test in GraphPad Prism (ver. 10.3.0).

Identification of ubiquitin-linkages by quantitative mass spectrometry analysis

For SUMO-targeted ubiquitylated HMCES^{FL}-DPCs (4 biological replicates per condition), reactions were terminated by boiling samples at 70 °C. Synthetically ubiquitylated HMCES^{SRAP} (3 biological replicates per condition) was directly used for mass spectrometry measurements. For quantitative mass spectrometry analysis, samples were subsequently cleaned-up using the paramagnetic-based SP3 technology as described previously[53]. Briefly, 100 µg of freshly pre-equilibrated SP3 beads (45152105050250, GE Healthcare), were added to 20 µL of samples. Purification of total proteins from in vitro reactions was next completed through three-rounds of 80% (v/v) ethanol-solvation of the SP3-sample mixture at room temperature. The resulting purified proteins were then subjected to trypsin digestion (1 µg) in 50 mM ammonium bicarbonate pH 8.0 for 16 h at 37 °C. Digested peptides were acidified using trifluoroacetic acid and desalted on reverse-phase C18 StageTips for MS analysis. Eluted samples were analyzed on a quadrupole Orbitrap mass spectrometer Exploris 480 (Thermo Scientific) equipped with a UHPLC EASY-nLC 1200 system (Thermo Scientific). Samples were loaded onto a C18 reversed-phase column (55 cm length, 75 mm inner diameter, packed in-house with ReproSil-Pur 120 C18-AQ 1.9-µm beads) (r119.aq, Dr. Maisch GmbH) and eluted with a gradient from 2.4 to 32% Acetonitrile.

The mass spectrometer was operated in data-dependent mode, automatically switching between MS and MS2 acquisition. Survey full scan MS spectra (m/z 300–1,650; resolution: 60,000; target value: 3×10^6; maximum injection time: 28 ms) were acquired in the Orbitrap. The 15 most intense precursor ions were sequentially isolated, fragmented by higher energy C-trap dissociation (HCD) and scanned in the Orbitrap mass analyzer (normalized collision energy: 30%; resolution: 15,000; target value: 1×10^5; maximum injection time: 40 ms; isolation window: 1.4 m/z (LFQ run)). Precursor ions with unassigned charge states, as well as with charge states of +1 or higher than +6, were excluded from fragmentation. Precursor ions already selected for fragmentation were dynamically excluded for 25 s.

Peptide identification: Raw data files were analyzed using MaxQuant (development version 1.5.2.8). Parent ion and MS2 spectra were searched against a database containing 98,566 human protein sequences obtained from UniProtKB (April 2018 release) using the Andromeda search engine. Spectra were searched with a mass tolerance of 6 ppm in MS mode, 20 ppm in HCD MS2 mode and strict trypsin specificity, allowing up to three miscleavages. Protein N-terminal acetylation and methionine oxidation were searched as variable modifications. The dataset was filtered based on posterior error probability (PEP) to arrive at a false discovery rate (FDR) of less than 1% estimated using a target-decoy approach.

ICP-OES measurements

SPRTN samples in storage buffer and expression media (TB) as control were digested using an *Anton Paar* Multiwave 5000 microwave. For this, 160 µL of each sample (corresponding to 0.95 mg protein) were placed into PTFE digestion vessels. To this, 1 mL HNO_3 (69%) (450041 M, VWR) was added. The used digestion program was: 5 min ramp up to 180 °C, then 10 min at 200 °C and 15 min at 220 °C. After digestion, samples were allowed to cool to room temperature before being diluted to final volumes (10 mL) with ultrapure water (type 1, 18.2 MΩ·cm at 25 °C) for Inductively Coupled Plasma Optical Emission Spectrometry (ICP-OES) analysis. Blank samples were treated analogously.

ICP-OES was performed on a Varian Vista RL instrument operating in radial mode to determine the concentrations of Co, Fe, Mn and Zn. Calibration standards were prepared in HNO_3 (2%) by diluting a certified multi-element ICP standard (1.09492, Merck) containing the elements of interest to obtain a 4-point linear calibration curve ranging from 0 µg/mL to 4 µg/mL. ICP-OES operating conditions were set as follows: plasma power at 1.25 kW, plasma gas at 13.5 L/min, nebulizer pressure at 170 kPa, auxiliary gas flow rate at 1.5 L/min with three replicates per measurement cycle, which were automatically averaged. The following emission lines were selected for Co at 230.786, 231.160, 237.863 and 258,033 nm, Fe at 234.350, 238.204, 258.588 and 259.940 nm, Mn at 257.610, 259.372 and 293.931 nm and Zn at 202.548, 206.200 and 213.857 nm. Quality control was ensured by analyzing blanks within the sequence and a certified reference material alongside the samples, with recoveries within acceptable limits. See Supplementary Table 1 for results.

Protein structure predictions

To prepare the crystal structure (PDB: 6mdx) for MD simulations, Swiss Model[54] was employed to model the two missing residues, to adjust the modified amino acids to their natural counterpart and to remove the ligands from crystallization. From the structures generated with Swiss Model, we took the one that was the closest to the crystal structure with an RMSD of 0.082 Å. Structures for SprT-BR (SPRTN$^{aa28-245}$) and SprT-BR-ubiquitin were predicted using ColabFold[39,55,56] and AlphaFold2[57] using alphafold2_ptm. Figures were generated using PyMOL (ver. 3.0.3).

Molecular dynamics simulations

Starting structures for SprT and SprT-Ubiquitin were generated as described above, in both cases the top-ranked model (Rank_1) was selected for further analysis. Predictions for SPRTN variants (L38S, L99S) were generated analogously. In case of SprT-ubiquitin complexes, the interface predicted for the WT enzyme was also used for SPRTN variants. Starting with these predicted structures, two Zn^{2+} ions were added based on their binding sites in the crystal structure. Subsequently, hydrogen atoms were added to these structures as well as to the model of the crystal structure employing the H++ server, which determines protonation states based on continuum electrostatics[58]. Specifically, H++ employs the Poisson–Boltzmann equation to estimate the pKa of the ionizable residues in a macromolecule and assigns the protonation states accordingly[59]. This way, the model for SprT and for the crystal structure consisted of the same atoms and of amino acids in the same protonation states. Pdb-files of the SprT domain (WT and variants L38S, L99S) are provided as supplementary data (Supplementary Data 1-3). Figures and movies were generated using PyMOL (ver. 3.0.3).

For setting-up the system for the simulations, we employed AmberTools20[60]. The structures were placed in rectangular simulation boxes with a minimum distance of 15 Å between the solute and the boundaries of the box. The boxes were filled with water and NaCl was added to neutralize the system and to achieve a physiological concentration of around 150 mM NaCl leading to system sizes of about 72.324 to 95.881 atoms for SprT and the SprT-Ub^1 complex, respectively.

The force field parameters for the proteins were taken from the Amber ff19SB force field[61] and the OPC water model was used[62]. After conversion of the topology and coordinate files to gromacs with parmed from AmberTools20, the parameters for NaCl were replaced by the ones from ref. 63 and for Zn^{2+} by the parameters from ref. 64.

MD simulations were performed using the Gromacs simulation package[65], version 2024. Initially, the energy of the systems was minimized using the steepest descent algorithm. Subsequently, the systems were equilibrated for 1 ns, first in the NVT and then in the NPT ensemble. For the production run, we performed 400 ns long simulations employing the velocity-rescaling thermostat with a stochastic term and a time constant of 0.1 ps and isotropic Parrinello-Rahman pressure coupling with a time constant of 5.0 ps. Each production was repeated three times with random velocities. We used clustering analysis with Gromacs for the production runs based on RMSD to group similar conformations, allowing us to identify the dominant structural states and to calculate the radius of gyration (Rg) of the clusters. In this analysis, 100 ns were discarded for equilibration. For the clustering, we employed the Gromacs bulit-in tool gmx cluster using the Daura clustering algorithm with an RMSD cut-off of 0.5 nm for all atoms of the protein. The algorithm identifies neighbors for each structure within the specified cut-off, selects the structure with the most neighbors as the first cluster center, and groups it with its neighbors. These are then removed from the pool, and the process repeats until all structures are clustered[66]. The numbers of clusters for the simulated systems and the fraction of structures in the three most populated clusters are summarized in Supplementary Table 2.

To calculate the binding affinity, additional simulations were performed. First 1 ns of NVT and NPT simulations with stronger restraints of 2.1 kcal/mol Å^2. Then production runs with weak restraints of 0.1 kcal/mol. We used representative structures from the largest cluster of the unrestrained simulations as starting structures. The productions were performed for 4 ns neglecting the first ns for equilibration and structures were sampled every 10 ps. Each simulation was repeated three times. The end-point free energy calculations were performed using the MMPBSA program from the Amber package[67] using the gmx_MMPBSA tool[68]. Water molecules and ions were removed and the trajectories were re-evaluated using the mbondi3 radii and parameters from ref. 69 denoted as igb=8 in Amber. To account for hydrophobic solvation, we used a surface area-dependent tension model with surface tension coefficient $\gamma = 0.005$ kcal/mol Å^2. No conformational changes were considered. From the alanine scanning procedure, we obtained the contribution of the two mutations to the binding free energy. The energy contribution of the mutation was calculated by cutting all atoms after the Cβ-atom of this residue. This procedure was performed for the simulations of the WT and the simulations of the mutated complexes. The difference yielded the binding energy contributions of the L38S and the L99S mutations.

NMR spectroscopy

All NMR samples (uniformly labeled ^{13}C,^{15}N-ZBD-BR, ^{15}N-SprT-BR, ^{15}N-SprT-BR-L99S, ^{15}N-SprT-BR-W36G, ^{15}N-SprT-BR-W58G, ^{15}N-ZBD) were prepared at concentrations of 100 μM and 250 μM in 20 mM HEPES/KOH pH 7.2, 150 mM KCl with 10% D_2O, as lock signal. All NMR experiments were recorded at 308 K on a 600 MHz Bruker Avance NMR spectrometer, equipped with a cryogenic triple-resonance gradient probe. NMR spectra were processed using NMRPIPE[70] or TOPSPIN3.7 (Bruker) and analyzed using NMRFAM-SPARKY[71]. Using the previous backbone resonance assignments for ZBD-BR from ref. 21, aromatic resonances were further assigned using 2D CBHD, CBHE, aromatic ^{1}H,^{13}C-HSQC and 3D ^{15}N/^{13}C-edited NOESY experiments. Trp ε1 resonances for the protease domain were assigned by mutation (W36G or W58G), while W68 resonance was assigned by exclusion. For 2D ^{1}H,^{15}N-HSQC comparisons, 100 μM SPRTN was mixed with 500 μM Ub^1 or Ub^1-I44A (5x molar excess) and / or 200 μM dsDNA (2x molar excess) (fwd: 5′-CCTTGCTAGGACATC-3′ + rev: 5′-GATGTCCTAGCAAGG-3′, annealed to dsDNA) accordingly. Chemical shift perturbations (CSP) values were calculated based as $\Delta\delta_{HN,N} = \sqrt{\Delta\delta_{HN}{}^2 + \left(\frac{\Delta\delta_N}{R_{scale}}\right)^2}$, where $R_{\text{scale}} = 6.5$ was applied as suggested previously[72].

Reporting summary

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

Data availability

Mass spectrometry data reported in this manuscript have been deposited to the ProteomeXchange Consortium (www.proteomexchange.org) via the Proteomics Identification Database (PRIDE) partner repository with the dataset identifier PXD063921. All remaining data supporting the findings of this study are available from the corresponding author upon request. Source data are provided with this paper.

Code availability

All conformers from the MD simulation trajectories are available from the corresponding author upon request.

References

1. Tubbs, A. & Nussenzweig, A. Endogenous DNA damage as a source of genomic instability in cancer. *Cell* **168**, 644–656 (2017).
2. 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).
3. Dantuma, N. P. & van Attikum, H. Spatiotemporal regulation of posttranslational modifications in the DNA damage response. *EMBO J.* **35**, 6–23 (2016).
4. 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).
5. Bienko, M. et al. Ubiquitin-binding domains in Y-family polymerases regulate translesion synthesis. *Science* **310**, 1821–1824 (2005).
6. Alcón, P. et al. FANCD2–FANCI is a clamp stabilized on DNA by monoubiquitination of FANCD2 during DNA repair. *Nat. Struct. Mol. Biol.* **27**, 240–248 (2020).
7. Weickert, P. & Stingele, J. DNA-Protein Crosslinks and Their Resolution. *Annu Rev. Biochem.* **91**, 157–181 (2022).
8. Tretyakova, N. Y., Groehler, A. & Ji, S. DNA-Protein Cross-Links: Formation, Structural Identities, and Biological Outcomes. *Acc. Chem. Res* **48**, 1631–1644 (2015).
9. Weickert, P. et al. SPRTN patient variants cause global-genome DNA-protein crosslink repair defects. *Nat. Commun.* **14**, 352 (2023).
10. 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).
11. Carnie, C. J. et al. Transcription-coupled repair of DNA-protein cross-links depends on CSA and CSB. *Nat. Cell Biol.* **26**, 797–810 (2024).
12. van Sluis, M. et al. Transcription-coupled DNA-protein crosslink repair by CSB and CRL4CSA-mediated degradation. *Nat. Cell Biol.* **26**, 770–783 (2024).
13. Oka, Y., Nakazawa, Y., Shimada, M. & Ogi, T. Endogenous aldehyde-induced DNA-protein crosslinks are resolved by transcription-coupled repair. *Nat. Cell Biol.* **26**, 784–796 (2024).
14. 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).
15. Sparks, J. L. et al. The CMG Helicase Bypasses DNA-Protein Cross-Links to Facilitate Their Repair. *Cell* **176**, 167–181.e21 (2019).
16. Tsherniak, A. et al. Defining a Cancer Dependency Map. *Cell* **170**, 564–576.e16 (2017).
17. Maskey, R. S. et al. Spartan deficiency causes genomic instability and progeroid phenotypes. *Nat. Commun.* **5**, 5744 (2014).
18. 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).

19. 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).
20. Stingele, J. et al. Mechanism and Regulation of DNA-Protein Crosslink Repair by the DNA-Dependent Metalloprotease SPRTN. *Mol. Cell* **64**, 688–703 (2016).
21. Reinking, H. K. et al. DNA Structure-Specific Cleavage of DNA-Protein Crosslinks by the SPRTN Protease. *Mol. Cell* **80**, 102–113.e6 (2020).
22. 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).
23. 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).
24. 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).
25. Kim, M. S. et al. Regulation of error-prone translesion synthesis by Spartan/C1orf124. *Nucleic Acids Res*. **41**, 1661–1668 (2013).
26. 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).
27. Lessel, D. et al. Mutations in SPRTN cause early onset hepatocellular carcinoma, genomic instability and progeroid features. *Nat. Genet* **46**, 1239–1244 (2014).
28. Gallina, I. et al. The ubiquitin ligase RFWD3 is required for translesion DNA synthesis. *Mol. Cell* **81**, 442–458.e9 (2021).
29. 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).
30. 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).
31. Carnie, C. J. et al. Decitabine cytotoxicity is promoted by dCMP deaminase DCTD and mitigated by SUMO-dependent E3 ligase TOPORS. *EMBO J*. **43**, 2397–2423 (2024).
32. Liu, J. C. Y. et al. Concerted SUMO-targeted ubiquitin ligase activities of TOPORS and RNF4 are essential for stress management and cell proliferation. *Nat. Struct. Mol. Biol*. **31**, 1355–1367 (2024).
33. Ruggiano, A. et al. The protease SPRTN and SUMOylation coordinate DNA-protein crosslink repair to prevent genome instability. *Cell Rep*. **37**, 110080 (2021).
34. Maskey, R. S. et al. Spartan deficiency causes accumulation of Topoisomerase 1 cleavage complexes and tumorigenesis. *Nucleic Acids Res*. **45**, 4564–4576 (2017).
35. Renz, C. et al. Ubiquiton-An inducible, linkage-specific polyubiquitylation tool. *Mol. Cell* **84**, 386–400.e11 (2024).
36. Mehta, K. P. M., Lovejoy, C. A., Zhao, R., Heintzman, D. R. & Cortez, D. HMCES Maintains Replication Fork Progression and Prevents Double-Strand Breaks in Response to APOBEC Deamination and Abasic Site Formation. *Cell Rep*. **31**, 107705 (2020).
37. Yaneva, D. et al. The FANCJ helicase unfolds DNA-protein crosslinks to promote their repair. *Mol. Cell* **83**, 43–56.e10 (2023).
38. Donsbach, M. et al. A non-proteolytic release mechanism for HMCES-DNA-protein crosslinks. *The EMBO Journal*, e113360 (2023).
39. Mirdita, M. et al. ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
40. 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).
41. Borgermann, N. et al. SUMOylation promotes protective responses to DNA-protein crosslinks. *EMBO J*. **38**, e101496 (2019).
42. Weickert, P., Dürauer, S., Götz, M. J., Li, H.-Y. & Stingele, J. Electro-elution-based purification of covalent DNA-protein cross-links. *Nat. Protoc*. **19**, 2891–2914 (2024).
43. 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).
44. Vaz, B. et al. Metalloprotease SPRTN/DVC1 Orchestrates Replication-Coupled DNA-Protein Crosslink Repair. *Mol. Cell* **64**, 704–719 (2016).
45. Lopez-Mosqueda, J. et al. SPRTN is a mammalian DNA-binding metalloprotease that resolves DNA-protein crosslinks. *eLife* **5**, e21491 (2016).
46. Kojima, Y. et al. FAM111A protects replication forks from protein obstacles via its trypsin-like domain. *Nat. Commun*. **11**, 1318 (2020).
47. Dokshin, G. A. et al. GCNA Interacts with Spartan and Topoisomerase II to Regulate Genome Stability. *Dev. Cell* **52**, 53–68.e6 (2020).
48. Serbyn, N. et al. The Aspartic Protease Ddi1 Contributes to DNA-Protein Crosslink Repair in Yeast. *Mol. Cell* **77**, 1066–1079.e9 (2020).
49. Kröning, 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).
50. Xu, P. et al. Quantitative proteomics reveals the function of unconventional ubiquitin chains in proteasomal degradation. *Cell* **137**, 133–145 (2009).
51. Carvalho, A. F. et al. High-yield expression in Escherichia coli and purification of mouse ubiquitin-activating enzyme E1. *Mol. Biotechnol*. **51**, 254–261 (2012).
52. Renz, C. et al. Ubc13-Mms2 cooperates with a family of RING E3 proteins in budding yeast membrane protein sorting. *J. Cell Sci*. **133**, jcs244566 (2020).
53. Hughes, C. S. et al. Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nat. Protoc*. **14**, 68–85 (2019).
54. Waterhouse, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*. **46**, W296–W303 (2018).
55. Mitchell, A. L. et al. MGnify: the microbiome analysis resource in 2020. *Nucleic Acids Res*. **48**, D570–D578 (2020).
56. Eastman, P. et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput Biol*. **13**, e1005659 (2017).
57. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
58. Gordon, J. C. et al. H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res*. **33**, W368–W371 (2005).
59. Anandakrishnan, R., Aguilar, B. & Onufriev, A. V. H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res*. **40**, W537–W541 (2012).
60. Case, D. A. et al. AmberTools. *J. Chem. Inf. Model*. **63**, 6183–6191 (2023).
61. Tian, C. et al. ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theory Comput* **16**, 528–552 (2020).
62. Izadi, S., Anandakrishnan, R. & Onufriev, A. V. Building Water Models: A Different Approach. *J. Phys. Chem. Lett*. **5**, 3863–3871 (2014).
63. Mamatkulov, S. & Schwierz, N. Force fields for monovalent and divalent metal cations in TIP3P water based on thermodynamic and kinetic properties. *J. Chem. Phys*. **148**, 074504 (2018).
64. Li, P., Roberts, B. P., Chakravorty, D. K. & Merz, K. M. Rational Design of Particle Mesh Ewald Compatible Lennard-Jones Parameters for +2 Metal Cations in Explicit Solvent. *J. Chem. Theory Comput* **9**, 2733–2748 (2013).
65. Van Der Spoel, D. et al. GROMACS: fast, flexible, and free. *J. Comput Chem*. **26**, 1701–1718 (2005).

66. Daura, X. et al. Peptide Folding: When Simulation Meets Experiment. *Angew. Chem. Int. Ed.* **38**, 236–240 (1999).
67. Miller, B. R. I. et al. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **8**, 3314–3321 (2012).
68. Valdés-Tresanco, M. S., Valdés-Tresanco, M. E., Valiente, P. A. & Moreno, E. gmx_MMPBSA: A New Tool to Perform End-State Free Energy Calculations with GROMACS. *J. Chem. Theory Comput.* **17**, 6281–6291 (2021).
69. Nguyen, H., Roe, D. R. & Simmerling, C. Improved Generalized Born Solvent Model Parameters for Protein Simulations. *J. Chem. Theory Comput.* **9**, 2020–2034 (2013).
70. Delaglio, F. et al. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
71. Lee, W., Tonelli, M. & Markley, J. L. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325–1327 (2015).
72. Mulder, F. A., Schipper, D., Bott, R. & Boelens, R. Altered flexibility in the substrate-binding site of related native and engineered high-alkaline Bacillus subtilisins. *J. Mol. Biol.* **292**, 111–123 (1999).

Acknowledgements

We thank B. Schulman for providing a plasmid for ubiquitin. We gratefully acknowledge Jiaxuan Chen from the IMB Proteomics Core Facility in Mainz for help with mass spectrometry experiments. S.D. and P.W. are supported by the International Max-Planck Research School for Molecules of Life. We are grateful to Sam Asami and Gerd Gemmecker for help with NMR measurements at the Bavarian NMR center. We thank Dr. Shar-yin N. Huang at the National Cancer Institute for her technical support. Research in the lab of J.S. is funded by European Research Council (ERC StG 801750 DNAProteinCrosslinks, ERC CoG 101124695 DECONSTRUCT), the Alfried-Krupp von Bohlen und Halbach-Stiftung, European Molecular Biology Organization (YIP4644), a Vallee Foundation Scholarship, and Deutsche Forschungsgemeinschaft (Project ID 213249687 - SFB 1064). H.D.U. acknowledges funding by the European Research Council (ERC AdG 101140447). We acknowledge funding by the Deutsche Forschungsgemeinschaft (J.S., H.D.U., and P.B.: Project-ID 393547839 – SFB 1361; M.S. and L.J.D: Project-ID 325871075 – SFB1309). The authors acknowledge the scientific support and HPC resources provided by the Erlangen National High Performance Computing Center (NHR@FAU) of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) under the NHR project b119ee and the resources on the LiCCA HPC cluster of the University of Augsburg, co-funded by the Deutsche Forschungsgemeinschaft under Project-ID 499211671).

Author contributions

Conceptualization: S.D. and J.S. Investigation: S.D., D.S.S., D.Y., P.W., M.J.G., Y.M., Y.J.M., and C.R. NMR: H.S.K. MD-simulations: C.W., A.C.M., and N.S. Mass spectrometry: A.S.R. ICO-OES: S.M.G.T. Writing – Original draft: S.D. and J.S. Writing – Review & Editing: S.D. and J.S. with input from all authors. Funding Acquisition and Supervision: J.S., L.J.D., P.B., H.D.U., M.S., Y.J.M., and N.S.

Funding

Open Access funding enabled and organized by Projekt DEAL.

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/s41467-025-61224-z.

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

Peer review information *Nature Communications* thanks Javier O. Sanlley Hernandez and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

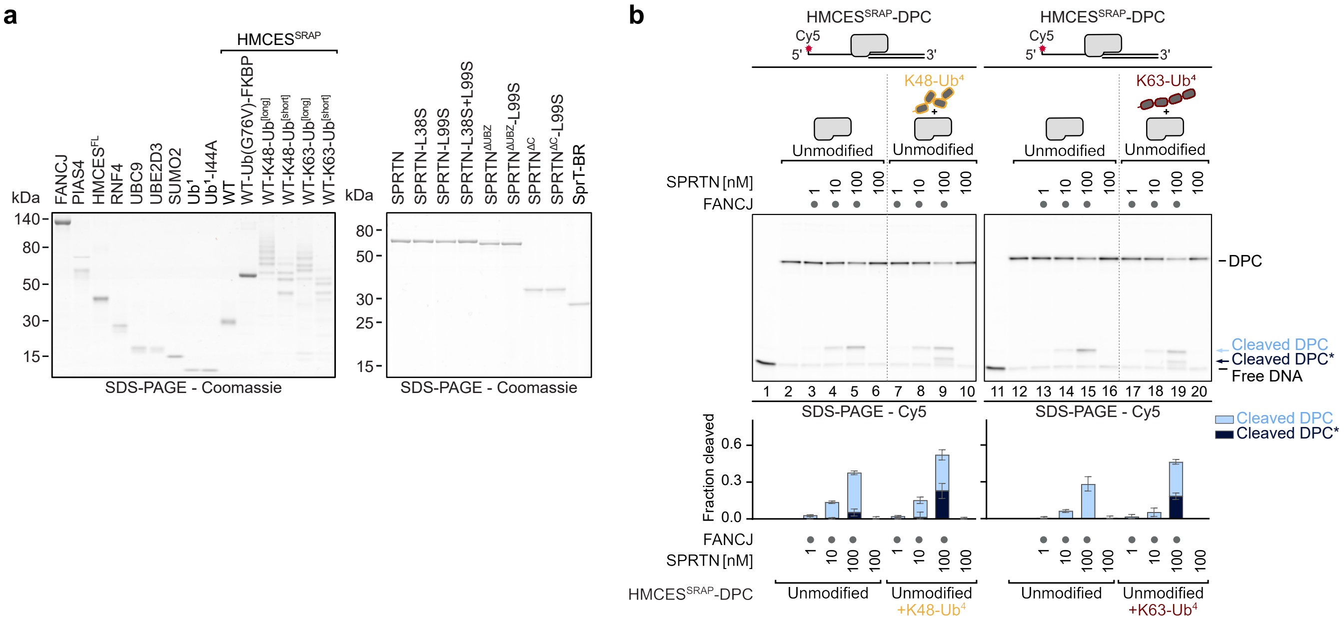
Supplementary Information

Allosteric activation of the SPRTN protease by ubiquitin maintains genome stability

Sophie Dürauer[1,2], Hyun-Seo Kang[3,4], Christian Wiebeler[5], Yuka Machida[6], Dina S Schnapka[1,2], Denitsa Yaneva[1,2], Christian Renz[7], Maximilian J Götz[1,2], Pedro Weickert[1,2], Abigail C Major[5], Aldwin S Rahmanto[7,8], Sophie M Gutenthaler-Tietze[9,10], Lena J Daumann[9], Petra Beli[7,8], Helle D Ulrich[7], Michael Sattler[3,4], Yuichi J Machida[6], Nadine Schwierz[5] and Julian Stingele[1,2,*]

[1] Gene Center, Ludwig-Maximilians-Universität München, Munich, Germany
[2] Department of Biochemistry, Ludwig-Maximilians-Universität München, Munich, Germany
[3] Institute of Structural Biology, Molecular Targets and Therapeutics Center, Helmholtz Munich, Neuherberg, Germany
[4] Bavarian NMR Center and Department of Bioscience, TUM School of Natural Sciences, Technical University of Munich, Garching, Germany
[5] Institute of Physics, University of Augsburg, Augsburg, Germany
[6] Developmental Therapeutics Branch, Center for Cancer Research, National Cancer Institute, Bethesda, MD, USA
[7] Institute of Molecular Biology gGmbH, Mainz, Germany
[8] Institute of Developmental Biology and Neurobiology (IDN), Johannes Gutenberg-Universität Mainz, Germany
[9] Chair of Bioinorganic Chemistry, Heinrich-Heine Universität Düsseldorf, Düsseldorf, Germany
[10] Department of Chemistry, Ludwig-Maximilians-Universität München, Munich, Germany
* Correspondence: stingele@genzentrum.lmu.de

SUPPLEMENTARY FIGURE 1

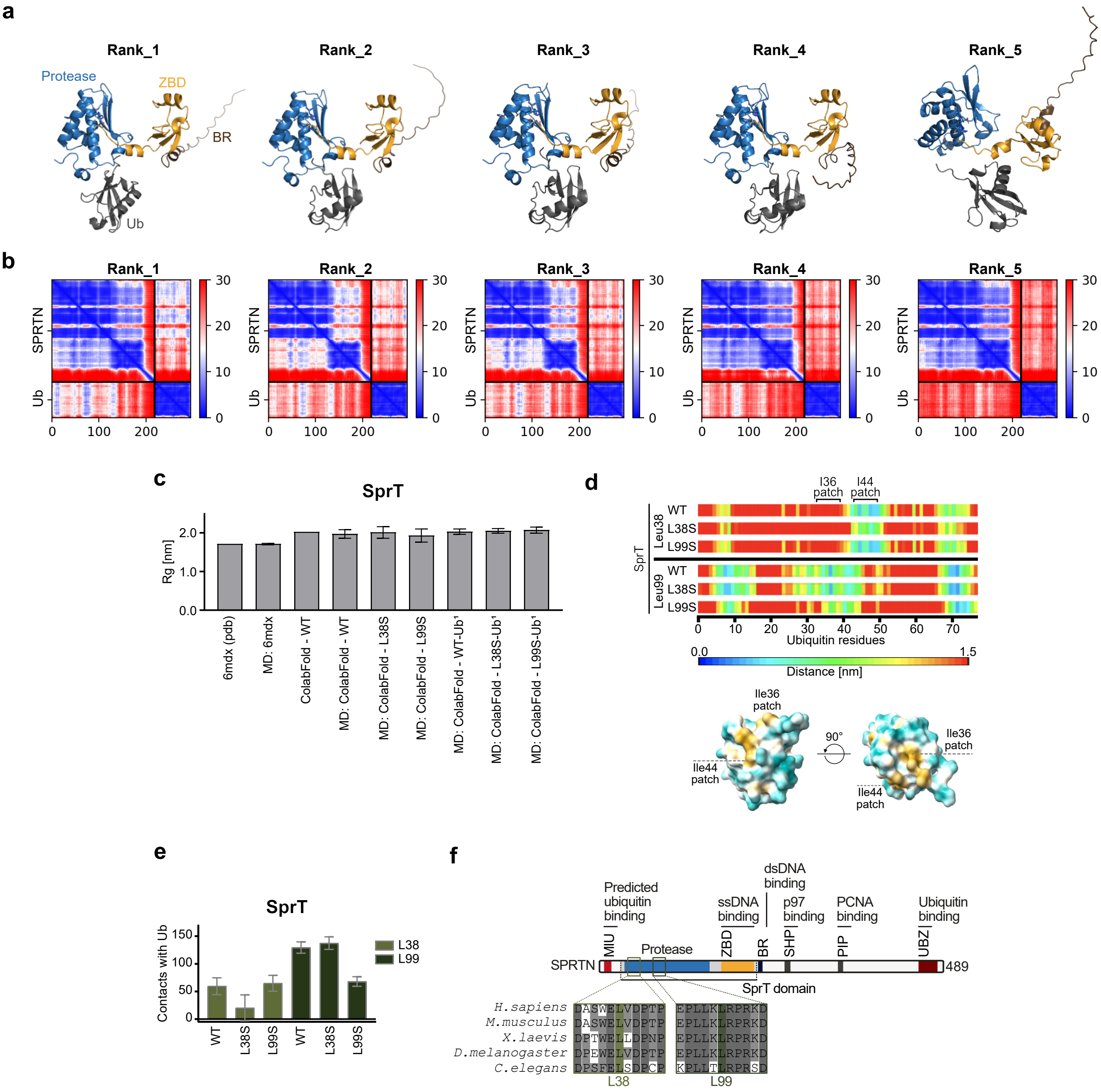


Supplementary Fig. 1. Ubiquitin boosts DPC cleavage by SPRTN

(a) Coomassie stained SDS-PAGE gel, showing equimolar amounts of purified recombinant human FANCJ, PIAS4, HMCES$^{\text{FL}}$, RNF4, UBC9, UBE2D3, SUMO2, mono-ubiquitin (Ub^{1}) and Ub^{1}-I44A, HMCES$^{\text{SRAP}}$, HMCES$^{\text{SRAP}}$-Ub(G76V)-FKBP, HMCES$^{\text{SRAP}}$-K48-Ub$^{[\text{long}]}$, HMCES$^{\text{SRAP}}$-K48-Ub$^{[\text{short}]}$, HMCES$^{\text{SRAP}}$-K63-Ub$^{[\text{long}]}$ and HMCES$^{\text{SRAP}}$-K63-Ub$^{[\text{short}]}$, SPRTN, SPRTN-L38S, SPRTN-L99S, SPRTN-L38S+L99S, SPRTN$^{\Delta\text{UBZ}}$, SPRTN$^{\Delta\text{UBZ}}$-L99S, SPRTN$^{\Delta\text{C}}$, SPRTN$^{\Delta\text{C}}$-L99S, and SprT-BR used for *in vitro* assays. Source data are provided as a Source Data file.

(b) HMCES$^{\text{SRAP}}$-DPCs (10 nM) were incubated alone or in the presence of FANCJ (100 nM), K48-tetra-ubiquitin (Ub^{4}) or K63-Ub^{4} (400 nM, referring to the concentration of individual ubiquitin moieties) and indicated concentrations of SPRTN (1-100 nM) for 1 h at 30°C. Quantification: bar graphs represent the mean ± SD of three independent experiments. All samples derive from the same experiment and gels were processed in parallel. Values for cleavage of unmodified HMCES$^{\text{SRAP}}$-DPC are the same as in Fig. 1e. Source data are provided as a Source Data file.

SUPPLEMENTARY FIGURE 2



Supplementary Fig. 2. A novel ubiquitin binding interface at the SprT domain

(a) Structures of SprT-BR-ubiquitin complex predicted by ColabFold using AlphaFold2_ptm. All five models (Rank_1-5) are shown (left to right), highlighting SPRTN's protease domain (blue), Zinc-binding domain (ZBD) (orange) and basic region (BR) (brown). Ubiquitin (Ub^1) is shown in grey.

(b) Predicted aligned error (PAE) blots of predicted SprT-BR-Ub^1 complexes shown in (a).

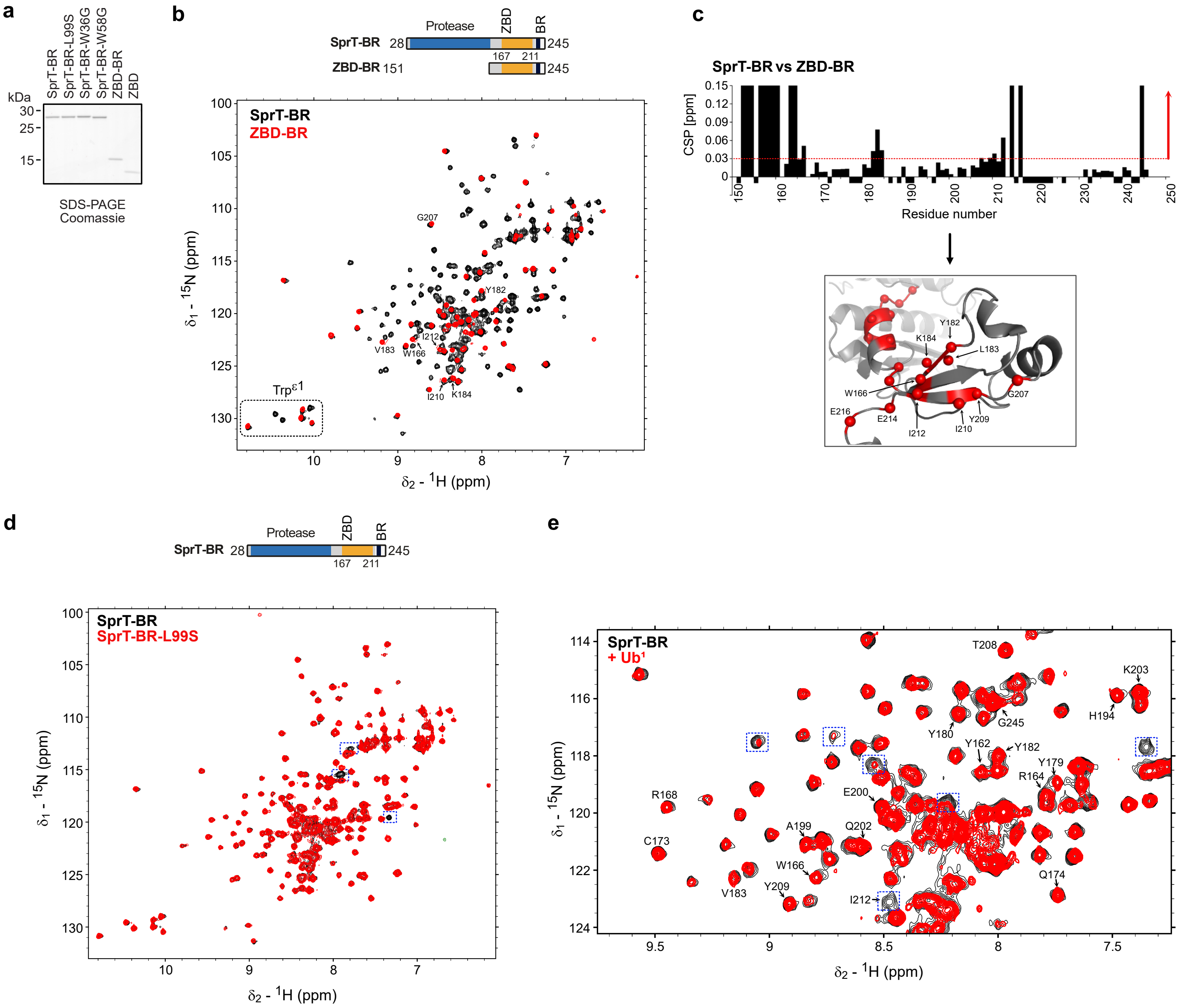
(c) Bar charts showing radius of gyration (Rg) before and after molecular dynamics (MD) simulations for SprT (PDB: 6mdx and ColabFold predicted) and SprT-Ub^1 (ColabFold predicted). In case of MD simulations, the mean ± SD of snapshots from three independent 400 ns MD trajectories is given and the first 100 ns of each trajectory were discarded for equilibration. Source data are provided as a Source Data file.

(d) Heat map indicating minimum distances of all amino acids in ubiquitin to residues at positions 38 (upper part) and 99 (lower part) of SPRTN, including wild-type (WT) conditions (L38, L99) and mutated states (L38S, L99S) (top). Structure of ubiquitin colored by hydrophobicity (bottom). Ile36 and Ile44 patches are highlighted.

(e) Average number of contacts between SprT residues 38 and 99 and ubiquitin, defined as interatomic distance <0.6 nm between side chain or backbone. Contacts for multiple atoms of the SprT residue with an atom of ubiquitin are only counted once. For the WT protein: forms approximately 60 (residue 38) and approximately 130 (residue 99) contacts. In the L38S variant contacts are reduced to approximately 20 at residue 38 and largely unchanged at residue 99. In L99S variant contacts at residue 99 decrease to around 68, while interactions at residue 38 remain similar to the WT. Source data are provided as a Source Data file.

(f) Schematic of SPRTN's domain structure with sequence alignment highlighting key residues in the SprT domain in *H. sapiens*, *M. musculus*, *X. laevis*, *D. melanogaster* and *C. elegans* SPRTN homologues (L38 = light green, L99 = dark green).

SUPPLEMENTARY FIGURE 3



Supplementary Fig. 3. NMR analysis of SPRTN's SprT domain

(a) Coomassie stained SDS-PAGE gel, showing equimolar amounts of purified recombinant human SprT-BR, SprT-BR- L99S, SprT-BR-W36G, SprT-BR-W58G, ZBD-BR, and ZBD used for NMR measurements.

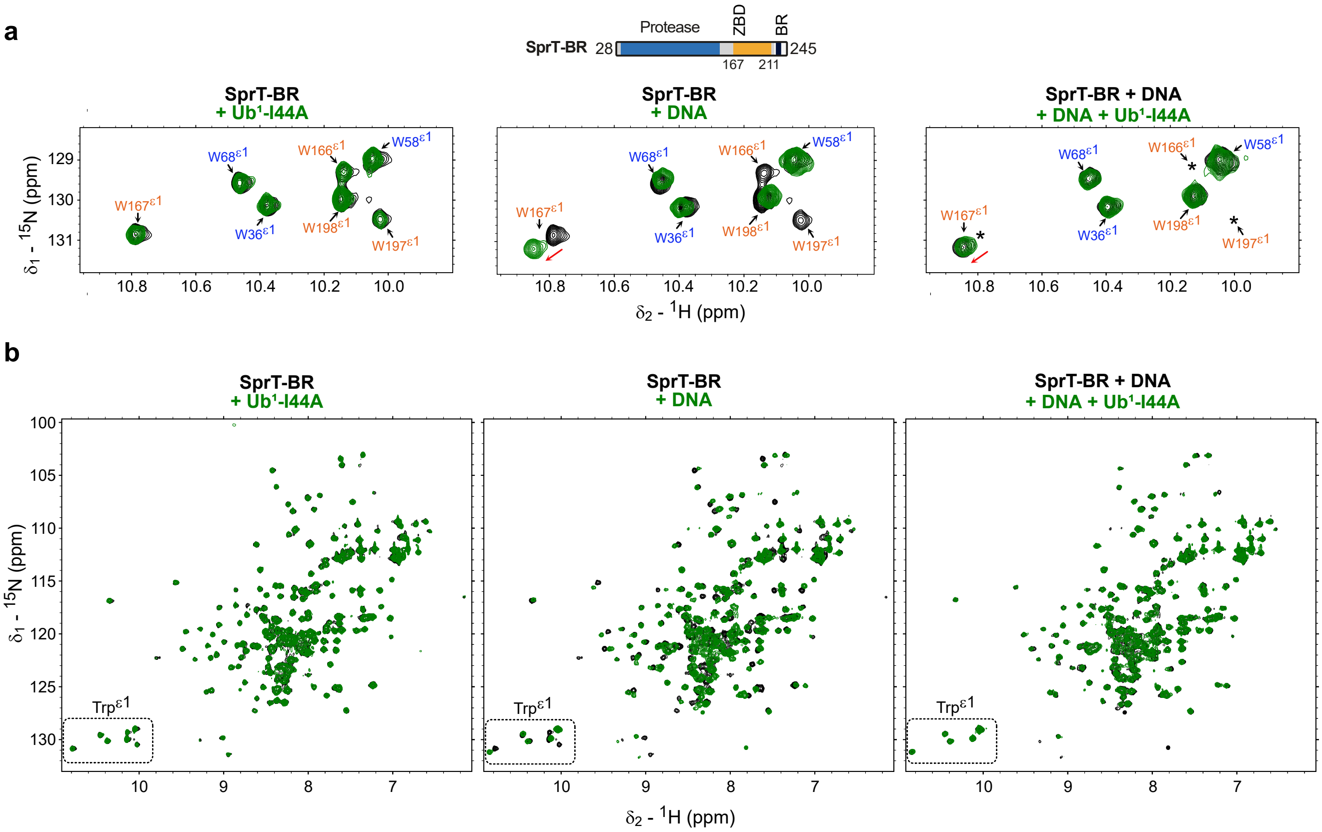
(b) Comparison of NMR spectra of SprT-BR (aa28-245) (black) and ZBD-BR (aa151-245) (red). Some representative peaks with Chemical Shift Perturbation (CSP) > 0.3 ppm are annotated on the spectrum. The crowded middle region corresponds to the unstructured linkers and the *C*-terminal BR region.

(c) CSP of SprT-BR against ZBD-BR. Residues with CSP > 0.3 ppm (red arrow) are highlighted in the structure as red spheres with labels for some representative residues (bottom). Negative values indicate proline or unassigned residues. Large changes, which could not be traced are given a CSP value of 1.5 ppm. CSP differences between SprT-BR and ZBD-BR highlight some differences in the linker between protease and ZBD (aa151-160) and on the β-sheet fold of the ZBD. Nonetheless, we were able to transfer many chemical shifts from our previous analysis of the ZBD-BR construct[21]. The large CSPs of residues aa151-160, presumably reflect interaction of this region with the protease domain. Due to the non-optimal sample stability of SprT-BR, we could not assign the individual resonances of the protease domain. However, by exclusion and introducing mutations of the tryptophanes in the protease domain (W36G and W58G), the resonances belonging to the protease domain could readily be distinguished from those in the ZBD and BR (see Fig. 3 for the Trp ε1 resonances).

(d) Comparison of NMR spectra of SprT-BR (black) and SprT-BR-L99S (red). Spectral differences are highlighted with blue boxes.

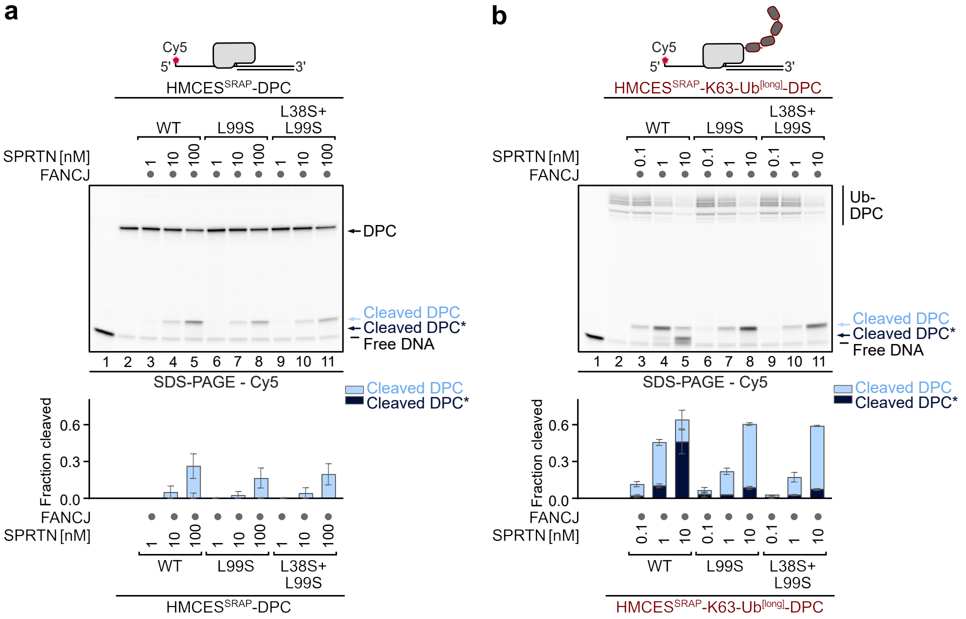
(e) Comparison of NMR spectra of SprT-BR alone (black) and with mono-ubiquitin (Ub^1) (5x molar excess) (red). Some resonances corresponding to the zinc-binding domain (ZBD) are labeled for the dispersed region. Unlabeled resonances in the dispersed region generally correspond to the protease domain. The broadened signals in the presence of Ub^1 are highlighted with blue boxes. Full spectrum shown in Fig. 3a.

SUPPLEMENTARY FIGURE 4



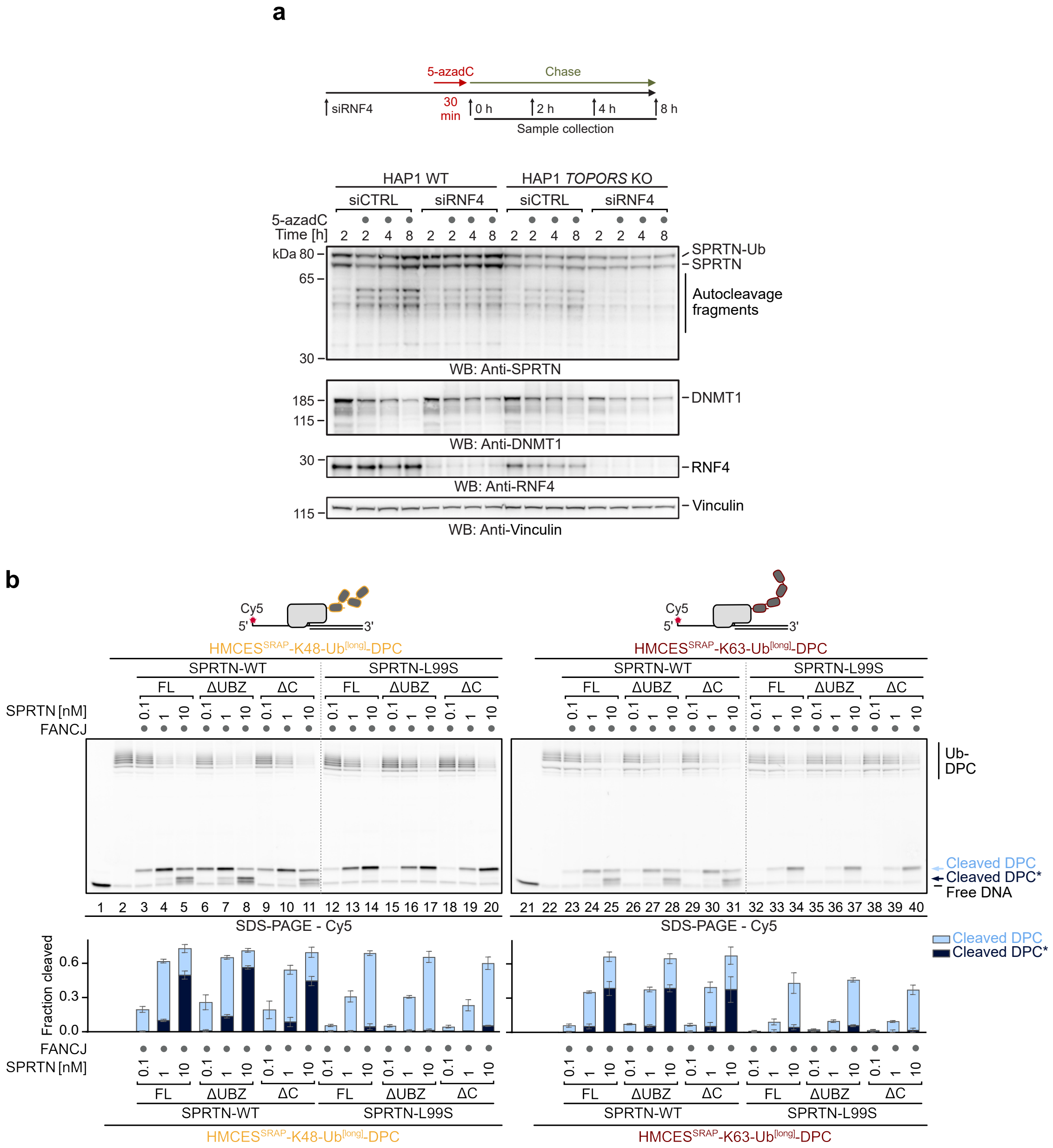
Supplementary Fig. 4. Ubiquitin binding at the USD depends on ubiquitin's Ile44 patch
(a) Comparison of NMR spectra (Trp ε1 amide signals in ^{1}H,^{15}N-HSQC experiments) of SprT-BR alone (black) (=Apo), with mono-ubiquitin (Ub^1)-I44A (5x molar excess) (green) (left) or with dsDNA (2x molar excess) (green) (middle). Right panel shows superimpositions of SprT-BR in the presence of dsDNA (2x molar excess) (black) and of both dsDNA (2x molar excess) and Ub^1-I44A (5x molar excess) (green). Resonances corresponding to the Trp ε1's in the ZBD are labeled in orange or shown as asterisk when broadened. Trp ε1's in the protease domain are labeled in blue. Full spectra are shown in (b).
(b) Full NMR spectra of Trp ε1 amide signals shown in (a). Trp ε1 region is boxed.

SUPPLEMENTARY FIGURE 5



Supplementary Fig. 5. The USD promotes cleavage of ubiquitylated DPC by SPRTN. (a-b) Indicated HMCES^{SRAP}-DPCs (10 nM) were incubated alone or in the presence of FANCJ (100 nM) and indicated concentrations (0.1-100 nM) and variants of SPRTN (WT, L99S, L38S+L99S) for 1 h at 30°C. Quantification: bar graphs represent the mean ± SD of three independent experiments. Source data are provided as a Source Data file.

SUPPLEMENTARY FIGURE 6

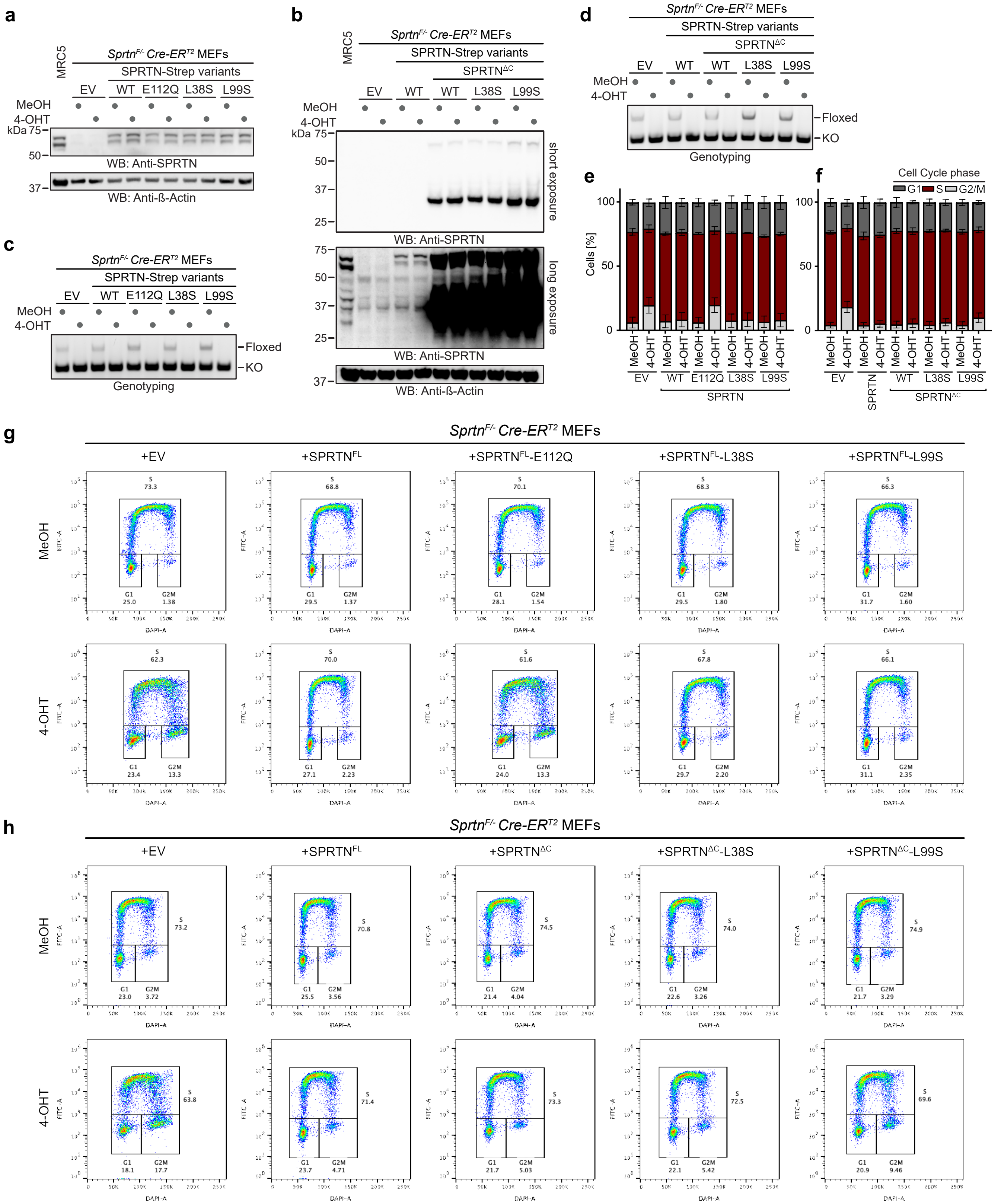


Supplementary Fig. 6. The USD is especially important in hypomorphic SPRTN variants

(a) HAP1 wild-type (WT) or HAP1 *TOPORS* knock-out (KO) cells transfected with indicated siRNAs were treated with 5-azadC (10 μM) and harvested as depicted (top). Whole cell lysates were analyzed by immunoblotting (bottom). Source data are provided as a Source Data file.

(b) Indicated HMCES^{SRAP}-Ub$^{[long]}$-DPCs (10 nM) were incubated alone or in the presence of recombinant FANCJ (100 nM) and indicated concentrations (0.1-10 nM) and variants of SPRTN (FL-WT/L99S, ΔUBZ-WT/L99S, ΔC-WT/L99S) for 1 h at 30°C. Quantification: bar graphs represent the mean ± SD of three independent experiments. All samples derive from the same experiment and gels were processed in parallel. Source data are provided as a Source Data file.

SUPPLEMENTARY FIGURE 7



Supplementary Fig. 7. Ubiquitin-dependent activation of SPRTN maintains genome stability in Ruijs-Aalfs syndrome.

(a-b) Expression of indicated SPRTN variants or empty vector (EV) (pMSCV) in *Sprtn*$^{F/-}$ Cre-ER^{T2} mouse embryonic fibroblasts (MEFs) treated with methanol (MeOH) or (Z)-4-hydroxytamoxifen (4-OHT) (2 μM) for 48 h. Whole cell lysates were anlayzed by immunoblotting. Source data are provided as a Source Data file.

(c-d) PCR-based genotyping of Sprtn alleles in *Sprtn*$^{F/-}$ Cre-ER^{T2} MEFs, complemented with indicated SPRTN variants or EV (pMSCV), treated with MeOH or 4-OHT (2 μM) for 48 h. Source data are provided as a Source Data file.

(e-h) Cell cycle profiling of *Sprtn*$^{F/-}$ Cre-ER^{T2} MEFs, complemented with indicated SPRTN variants or EV (pMSCV), treated with MeOH or 4-OHT (2 μM) for 48 h. Cells were labeled with EdU for 45 min and analyzed by flow cytometry. Bar charts represent the mean ± SD of three independent experiments (e-f). Flow charts show a representative of these three experiments (g-h). Source data are provided as a Source Data file.

Supplementary Table 1

Sample	MW (kDa)	Input		Resuspended in (mL)	Con		Measurement (mean of all measured wavelengths)							
		μL	mg		mg/mL	μM	Co		Fe		Mn		Zn	
							μg/mL	μM	μg/mL	μM	μg/mL	μM	μg/mL	μM
TB	-	160	-	10	-	**-**	0.0035	**-**	0.081	**-**	0.001667	**-**	0.075333	**-**
SPRTN	55.2	160	0.95	10	0.095	**1.72**	0.002	**0.03**	0.03125	**0.56**	<LOD	**<LOD**	0.348	**5.32**

Atomic mass metals	
Fe	55.845
Zn	65.380
Mn	54.938
Co	58.933

LOD = Limit of Detection

Supplementary Table 1. ICP-OES measurements.

Inductively Coupled Plasma Optical Emission Spectrometry (ICP-OES) measurement results for recombinant SPRTN and the used expression media (TB). The file contains: The Molecular weight (kDa), input volume (µL) and mass (mg), volume for resuspension (mL), concentration of sample (mg/mL and µM) and corresponding measured amounts (mg/mL and µM) for Co, Fe, Mn and Zn (mean of all measured wavelengths. Additionally, the atomic mass for Co, Fe, Mn and Zn are given. LOD = Limit of Detection.

Supplementary Table 2

Overview:

Model	$n_{Cluster}$	Fraction Σ Top 3-Cluster
SprT - PDB: 6mdx	1	1.00
SprT - ColabFold	9	0.98
SprT-L38S - ColabFold	19	0.77
SprT-L99S - ColabFold	8	0.95
SprT+Ub[1] - ColabFold	9	0.97
SprT-L38S+Ub[1] - ColabFold	9	0.98
SprT-L99S+Ub[1] - ColabFold	3	1.00

Details:

Model	Cluster	Structures	Prob.	Rg [nm]	SD [nm]	n: 90003
SprT - PDB: 6mdx	1	90003	1.00	1.71	0.02	
SprT - ColabFold	1	37295	0.41	2.02	0.06	
	2	30001	0.33	1.84	0.04	
	3	21072	0.23	2.05	0.06	
	All	90003	1.00	1.97	0.11	
SprT-L38S - ColabFold	1	43149	0.48	2.06	0.08	
	2	18669	0.21	1.78	0.07	
	3	7375	0.08	2.16	0.04	
	All	90003	1.00	2.01	0.15	
SprT-L99S - ColabFold	1	34661	0.39	2.08	0.08	
	2	30916	0.34	1.75	0.05	
	3	19954	0.22	1.92	0.07	
	All	90003	1.00	1.93	0.17	
SprT+Ub1 - ColabFold	1	79430	0.88	2.03	0.07	
	2	5367	0.06	2.05	0.04	
	3	2796	0.03	2.05	0.06	
	All	90003	1.00	2.03	0.07	
SprT-L38S+Ub1 - ColabFold	1	71075	0.79	2.05	0.05	
	2	10972	0.12	2.06	0.03	
	3	6318	0.07	2.02	0.07	
	All	90003	1.00	2.05	0.06	
SprT-L99S+Ub1 - ColabFold	1	76721	0.85	2.09	0.05	
	2	10649	0.12	1.91	0.10	
	3	2633	0.03	2.08	0.05	
	All	90003	1.00	2.07	0.08	

Supplementary Table 2. Clustering overview MD-simulations.

Summary of numbers of clusters from molecular dynamics (MD) simulations of SprT (PDB: 6mdx), SprT (ColabFold), SprT-L38S (ColabFold), SprT-L99S (ColabFold), SprT-Ub[1] (ColabFold), SprT-L38S+Ub[1] (ColabFold), SprT-L99S+Ub[1] (ColabFold). The file contains: number of clusters (n) and the fraction Σ of the top 3 cluster, for each model (Overview); On the second sheet (Detail): for each model, the three largest clusters (SprT – PDB: 6mdx only showed one cluster) are listed along with their probability (Prob.;based on the number of structures), their radius of gyration (Rg; nm) including standard deviation (SD; nm), as well as the radius of gyration when all structures of a model are considered (All).

Log$_2$-transformed normalised intensities of all proteins measured and depicted in Fig. 1d-e. The file contains: The Uniprot identifier (UniprotID), HGNC gene symbol (Gene name), the normalised log$_2$-transformed intensity in each replicate of the respective conditions (FA_Rn: formaldehyde-treated replicate n, FABenzCrtl_Rn: Formaldehyde-treated replicate n nuclease control, Untr_Rn: untreated replicate n and UntrBenzCrtl_Rn: untreated replicate n nuclease control. n indicates the replicate numbers 1-6). NA accounts for non-detected.

4 Discussion

4.1 HMCES-DPCs are reversible

HMCES crosslinks to AP sites in ssDNA, protecting them from incision and thereby preventing formation of SSBs, which if not repaired, can lead to toxic DSBs[251]. While these DPCs are crucial to ensure genome stability, they interfere with replication and transcription, as DNA and RNA polymerases stall in front of the DPC. Proteolytic degradation of HMCES-DPCs by SPRTN or the proteasome has been described in *X. laevis* egg extract. However, they also disappeared in situations where no proteolytic cleavage was possible[251,265]. One study indicated that HMCES^{SRAP}-DPCs can undergo reversal, proposing an additional resolution mechanism[267]. This raises the question of how such a mechanism can ensure protection of AP sites in ssDNA.

HMCES-DPCs formed *in vitro* are stable for multiple days at room temperature, initially suggesting irreversibility[257]. However, our *in vitro* experiments with recombinant human HMCES^{SRAP} and a prior study show that HMCES-DPCs can revert and that released HMCES can re-crosslink[259,267]. The auto-release mechanism depends primarily on the conserved Glu127 residue, with minor contribution of His210. During crosslink formation, these residues catalyse AP site ring opening and stabilise the covalent bond with the AP site[267]. Once released, HMCES either re-crosslinks or dissociates. Mutations at these residues (E127A and H210A) impaired crosslink reversal but had only slight (E127A) or no (H210A) effects on initial crosslinking ability. Similar results were seen for the bacterial ortholog YedK, indicating evolutionary conservation of this release mechanism[267]. The decision to re-crosslink or dissociate is driven by the local DNA structure. HMCES crosslinks to AP sites in ssDNA or at ssDNA-dsDNA junctions but not in dsDNA[251,257]. 3'-dsDNA was long suggested to act as a steric hindrance for crosslinking, with HMCES failing to accommodate the AP within its active site[257]. However, we observed efficient crosslinking at DNA nicks and gaps, implying that HMCES simply requires certain flexibility and bendability of the DNA, rather than long stretches of ssDNA per se. Consistent with this, DPCs in dsDNA reverted faster than those in ssDNA or at ssDNA-dsDNA junctions. Crosslink formation and reversal were related to HMCES' DNA-binding affinity, which is higher for DNA junctions than for ssDNA[257]. HMCES-DPC reversal therefore relies on two main steps. First Glu127, with minor contribution of His210, catalyses release of the crosslink (Figure 19a). Secondly, whether HMCES re-crosslinks or dissociates depends on the local DNA structure and its binding affinity towards it (Figure 19b).

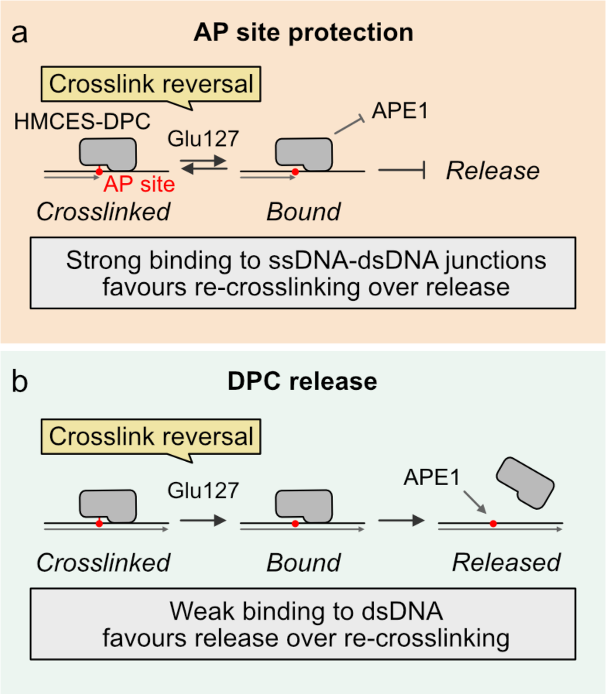


Figure 19. Model of AP site protection by HMCES-DPCs via crosslink reversal. (a-b) At abasic (AP) sites in ssDNA and at ssDNA-dsDNA junctions HMCES efficiently crosslinks, protecting the AP site from spontaneous or enzymatic incision (e.g. by AP endonuclease 1, APE1). Glu127 within HMCES' active site, mediates constant cycling between crosslinked and non-crosslinked states. High affinity to the underlying DNA structure favours re-crosslinking (a). Binding to dsDNA is weak, triggering release of HMCES and initiating AP site repair after cleavage by APE1 (b). Figure adapted from Donsbach *et al.*[259].

During replication, HMCES-DPCs form in ssDNA, raising the question, of how the dsDNA triggering HMCES release is formed. One possible mechanism is bypass by TLS polymerases, which were able to synthesise across HMCES-DPC *in vitro*[266,550]. However, efficient bypass required prior DPC unfolding by FANCJ[266] or proteolytic degradation[550]. Alternatively, annealing of the nascent strand followed by TS may generate dsDNA, enabling error-free repair[251,260,261]. By producing fully dsDNA around the DPC, SPRTN remains inactive, indicating that these pathways preferentially promote HMCES reversal over proteolytic degradation[285,306].

4.1.1 Physiological relevance

So far, the primary repair pathway for HMCES-DPCs remains unclear. While SPRTN mediates replication-coupled ICL repair, via HMCES-DPC cleavage, in *X. laevis* egg extract[265], TS is the main mechanism for DNA gap filling in mammals[551]. In mammalian cells, HMCES-DPCs are resolved even after proteasome inhibition and SPRTN knockdown, highlighting auto-release as an efficient repair pathway[552]. At the same time proteolysis may compensate when release is impaired, as overexpression of HMCES^{E127A} caused no toxicity and rescued sensitivity for AP site inducing agents of *HMCES* knockout cells[260]. Studying HMCES-DPC resolution is challenging as SPRTN and the proteasome are essential for cell viability, making it difficult to completely abolish proteolysis. Interestingly, the hyper-reverting HMCES^{R98E} variant sensitises cells to IR[251], suggesting that tight crosslinking is essential when AP sites accumulate. A major advantage of crosslink reversal is that the enzyme can be recycled and used again, whereas proteolysis degrades the protein and requires re-synthesis.

While in the aforementioned situations, HMCES-DPCs form during replication on the template strand, a recent study in HR-deficient cells showed HMCES also crosslinking to AP sites on the nascent strand[553]. These DPCs formed after 5hmdU misincorporation in the nascent strand, forming toxic BER intermediates that stall replication forks and trigger formation of post-replicative ssDNA gaps. These gaps were formed either by APE1/APE2 incision or by PRIMPOL-mediated repriming. HMCES-DPCs accumulated in HR-deficient *Fancd2*$^{-/-}$ cells, suggesting a role in protecting AP sites besides RAD51. They proposed a model where BER-mediated ssDNA flap exposure is promoted after 5hmdU incorporation in the nascent DNA strand by staggered action of SMUG1 and APE1/APE2. Initial nicks caused by multiple AP sites triggered BER-mediated ssDNA flap exposure, to which HMCES efficiently crosslinked. This blocks BER completion and stalls replication in FA/BRCA-deficient cells[553]. Lack of FANCD2-FANCI nucleofilament formation can lead to defective RAD51 stabilisation, which leaves AP sites unprotected[553]. While HMCES crosslinking would shield these AP sites from incision it also impairs replication fork progression. As here, auto-release is not an option, resolution fully depends on proteolysis by the proteasome or SPRTN, with prior FANCJ-mediated unfolding[251,266]. Moreover, HMCES-DPCs in the nascent strand also induced PRIMPOL-dependent ssDNA gaps and sister chromatid exchange (SCE). As *Fancd2*$^{-/-}$ cells lack RAD51-mediated HR, it would be interesting to investigate whether RAD52 or other paralogs promote strand invasion and faithful SCE completion in these cells. Notably, HMCES expression correlated with survival in HR-deficient tumours, suggesting it could be a biomarker for HR-deficient cancer susceptibility[553].

Beyond crosslinking, HMCES also plays a role in MMEJ repair, mediated by its DNA-binding ability. It binds class switch regions, protects ssDNA overhangs and promotes MMEJ. Mice lacking HMCES show significant defects in class switch recombination[264,554], suggesting a RPA-like function in stabilising ssDNA during repair[555].

4.1.2 Comparison to other physiological DPCs

Unlike most DPCs, which form as parts of pathological processes, HMCES forms stable crosslinks at AP sites in ssDNA to prevent incision. Other enzymes like TOP1 and TOP2[40,209] or DNMT1[219–222] form transient complexes with DNA and are released upon successful DNA processing. Here, stable DPC formation is considered a detrimental consequence of the entrapment of the enzyme, which can be induced for instance by chemical compounds and requires DPC repair pathways for resolution[210,216,224,225]. While Polβ and PARP1 were also described to crosslink to AP sites, no beneficial, protective function was reported so far[227,236,237]. Therefore HMCES' functions and its built-in resolution mechanism seem distinct from most other physiological DPCs, including also SPO11, which requires proteolytic repair[240]. Notably, UdgX acts similar to HMCES, binding and protecting AP sites, but also removes the uracil from DNA before crosslinking[248–250]. No DNA glycosylase activity has been reported for HMCES so far, but it is an interesting question for future investigations.

Given HMCES' protective role in ssDNA, inhibiting DPC formation in combination with drugs generating long stretches of ssDNA, like the Polα inhibitor CD437[556], could enhance therapeutic efficacy by increasing ssDNA vulnerability.

4.2 Advantages, limitations and opportunities for the PxP method

A major challenge in DPC research has long been the reliable visualisation and identification of DPCs. With the purification of x-linked proteins (PxP), we established a straightforward protocol for the identification of crosslinked proteins, accessible to trained molecular biologists using standard lab equipment. Briefly, cells are embedded into agarose plugs and lysed under denaturing conditions after DPC-inducing treatments. Soluble, non-crosslinked proteins are removed via electro-elution after lysis, leaving genomic DNA and crosslinked proteins in the plug. These can then be analysed by standard techniques, such as western blotting and fluorescent staining or used for quantitative MS-based proteomics (Figure 20).

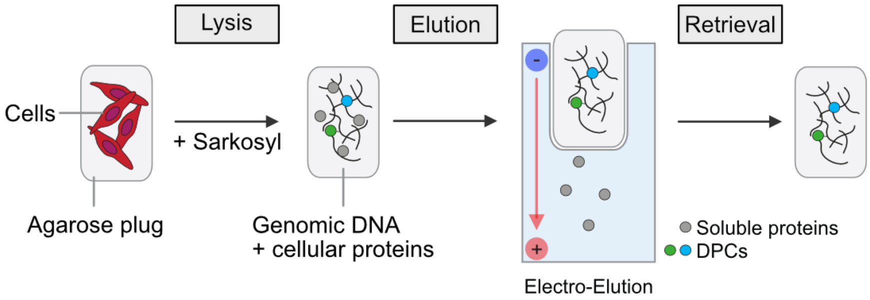


Figure 20. Schematic depiction of the PxP method. Depicted are the main steps of the purification of x-linked proteins (PxP) protocol for the purification and analysis of DNA-protein crosslinks (DPCs). Figure adapted from Weickert *et al.*[557].

Compared to other established methods for DPC analysis, like KCl-SDS[272] or the RADAR [558] and its derivatives, the PxP does not rely on precipitation. This is beneficial as precipitation can be affected by various factors, for instance features of the crosslinked protein or protein aggregation caused by pleiotropic agents like formaldehyde, which then can be mistaken for

DPCs. Moreover, we have included a nuclease control in the PxP protocol, which digests DNA in the plug prior to electro-elution, enabling distinction between true DPCs and non-specific contaminants.

A limitation of the PxP is the number of cells that can be embedded into one plug, as too many cells can impair lysis and increase background signals. One can overcome this issue by casting multiple plugs per condition. Another limitation lies in the detection of reversible DPCs, as this can be challenging since cells must be embedded into agarose before lysis, giving the DPCs time to revert. Despite this, we successfully detected TOP2-DPCs induced by etoposide and another study purified TOP1-DPCs induced by CPT, where even TOP1-PTMs could be detected[353]. Moreover, repair of DNMT1-DPCs after 5-azadC treatment could be monitored with the PxP[353]. After UVC treatment or Polα inhibition we were able to detect HMCES-DPCs, consistent with results obtained using RADAR[552]. Importantly, no DPCs were observed for the HMCES^{C2S} variant (mutation at the crosslink forming Cys2), and after drug removal HMCES-DPCs disappeared, most likely related to HMCES' auto-release function[259,552].

While HMCES-DPCs can be efficiently purified with the PxP, less stable AP site crosslinks, including Schiff-base intermediates, may be more challenging. 5fC forms Schiff-base intermediates with histones, which can be chemically stabilised using $NaCNBH_3$ after nuclei extraction to allow direct access to chromatin[559]. These Schiff-base intermediates are also formed by proteins linked to AP sites in DNA repair pathways[16,59,64], but may be too labile to survive PxP purification. *In vitro* studies showed that $NaCNBH_3$ can stabilise the Schiff-base intermediate between an AP site and YedK (bacterial homolog of HMCES) active site variants (C2A, C2S), enabling detection of stable DPCs[257]. Incorporation of a $NaCNBH_3$ reduction step in the PxP protocol, however, is difficult as it requires access to nuclei which is usually achieved by adding $NaCNBH_3$ during cell lysis. In the PxP, cells are embedded in agarose before cell lysis, making it challenging to enable access for $NaCNBH_3$ to nuclei and Schiff-base intermediates. Studying Schiff-base intermediates would be especially interesting following treatment with the Polα inhibitor CD437, which leads to ssDNA accumulation, RPA depletion and increased AP site formation[556]. HMCES-DPCs form after CD437 treatment and disappear after drug removal, with only mild contribution of the proteasome or SPRTN, suggesting a dominant role for HMCES' auto-release mechanism here[552]. We also observed SPRTN activation in human cells post CD437-treatment, evident by SPRTN autocleavage fragments. This corresponds with SPRTN targeting DNA structures like stalled replications forks with ssDNA-dsDNA junctions. However, SPRTN did not seem to target HMCES-DPCs[552]. It remains elusive whether other DPCs form under these conditions and whether they are SPRTN substrates. Stabilisation of Schiff-base intermediates followed by PxP could help identify these potential other DPCs.

The PxP has also been successfully coupled with unbiased MS-based proteomics studies, using formaldehyde to induce DPCs, yielding insights into cellular responses to formaldehyde-induced DNA damage[353,357]. Similar approaches could be applied to other DPC-inducing compounds, which would be especially interesting for unspecific crosslinkers like methylating agents (MMS or MNNG), oxidising agents (H_2O_2 or KBrO3), other aldehydes (acetaldehyde or malondialdehyde) or radiation (UV or IR). A specific crosslinker could serve as a control in these experiments, for instance, etoposide, which induces TOP2-DPCs[216]. Such a characterisation of DPCs and compounds would be a valuable resource for the field and may uncover novel, physiologically or therapeutically relevant DPCs.

4.3 A ubiquitin-binding interface at the SprT domain

Interactions of SPRTN and ubiquitin have primarily been linked to SPRTN's *C*-terminal UBZ domain, which facilitates SPRTN recruitment to ubiquitylated DPCs during global-genome DPC repair[353] and regulates protein levels via monoubiquitylation in cells[365]. However, the UBZ is not essential, as RJALS patients' SPRTN variants lacking the UBZ are still viable, albeit with reduced activity[293]. There is no evidence to support ubiquitin binding at the predicted *N*-terminal MIU[293], leaving the question open whether UBZ-deficient SPRTN variants still bind ubiquitin.

We addressed this by modelling interactions between ubiquitin and SPRTN's catalytic core, the SprT domain. Interestingly, the model predicted ubiquitin's hydrophobic Ile44-patch binding a previously undescribed site at the back of the SprT domain (Figure 21a), now referred to as ubiquitin-binding interface at the SprT domain (USD). Moreover, unlike the published crystal structure[296] (Figure 21b), the predicted structure adopts an open conformation in which the catalytic site and the DNA-binding site of the ZBD are both accessible facing each other. This conformation persisted with or without ubiquitin in molecular dynamics simulations, but the SprT-Ub complex showed greater structural stability. Using NMR, we could confirm ubiquitin binding at the USD, with detectable spectral shifts upon addition of DNA and ubiquitin. Furthermore, we identified key amino acid residues mediating ubiquitin interactions at the USD. Leu38 interacts with ubiquitin's Ile44-patch and Leu99 with the Ile36-patch, both via hydrophobic interactions. Substitution of these leucine residues with hydrophilic serine (L38S, L99S) disrupted ubiquitin binding. Both residues, and the entire SprT domain are evolutionarily highly conserved. Notably, Leu38 lies within an α-helix (Figure 21a), a fold commonly found in UIMs, MIUs and UBAs and it also interacts with ubiquitin's Ile44-patch[462]. However, no classic UIM or MIU motifs are found around Leu38. No common UBD structures are predicted around Leu99.

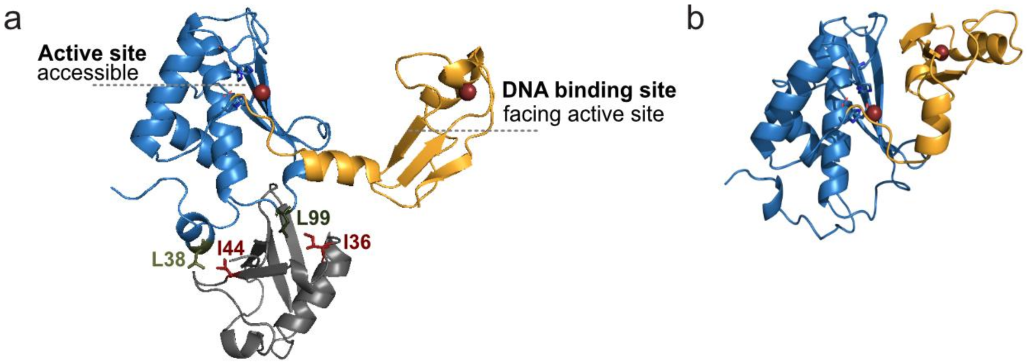


Figure 21. ColabFold model and crystal structure of SprT. (a-b) ColabFold predicted structure of SprT (aa28-214) in complex with ubiquitin (a) and published crystal structure of SprT (PDB: 6mdx) (b). Protease domain is coloured in blue and Zinc-binding domain (ZDB) in orange. Ubiquitin is shown in grey and Zn^{2+} ions in red. Important amino acid residues mediating ubiquitin-binding to SprT are labelled and coloured in green (L38 and L99). Their corresponding interaction patches in ubiquitin are highlighted in red (I44 and I36). Figure adapted from Dürauer *et al.*[1].

Even though human SPRTN (SprT domain), bacterial SprT and yeast Wss1 are homologs, they display some structural differences (Figure 22). Only human SPRTN has a characterised ZBD (Figure 22a), but a similar fold exists in bacterial SprT, although lacking experimental evidence for DNA binding (Figure 22b). Since in the case of SPRTN, DNA and ubiquitin binding trigger and stabilise a conformational change, it would be interesting to test whether bacterial SprT also has structural flexibility and whether potential conformational changes are triggered by DNA binding. Notably, most bacteria do not express ubiquitin, therefore the process must either fully rely on DNA binding or include other not yet known factors. For this model, testing and identifying potential DNA-binding domains in bacterial SprT will be essential.

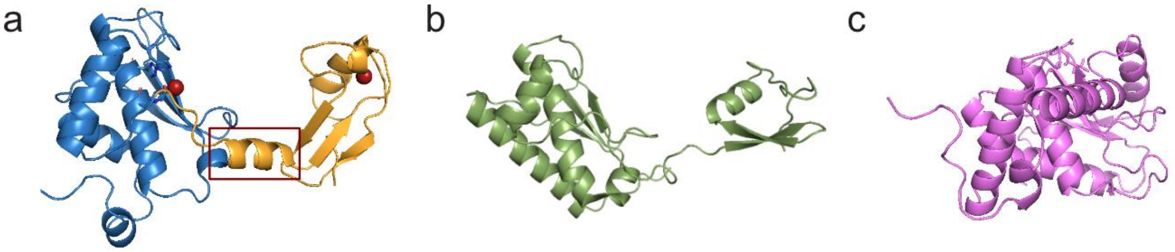


Figure 22. AlphaFold models of SPRTN, SprT and Wss1. (a) ColabFold predicted structure of human SPRTN (SprT domain, aa28-214). Protease domain is coloured in blue, Zinc-binding domain (ZBD) in orange and Zn^{2+} ions in red. The acidic linker between the protease domain and the ZBD is boxed in red. Figure adapted from Dürauer *et al.*[1]. (b) AlphaFold predicted structure of SPRTN's bacterial homolog SprT from *E. coli* (AlphaFoldDB: P39902). (c) AlphaFold predicted structure of SPRTN's yeast homolog Wss1 from *S. cerevisiae* (AlphaFoldDB: P38838). Largely unstructured domains were removed for easier visualisation.

A notable difference lies in the linker between the two domains. In human SPRTN, this linker connects the protease domain and the ZBD and is predicted to form a highly negatively charged and acidic α-helix consisting almost entirely of aspartate and glutamate residues (DEVDE) (Figure 22a, red box). Such an α-helix structure is absent in bacterial SprT. Considering that open and closed conformation differ mainly by ZBD positioning, this acidic linker may mediate these structural transitions. Interestingly, a point mutation (in SPRTN) within this region (D157K) increases autocleavage activity of the enzyme[296], suggesting that the linker contributes to SPRTN regulation, potentially acting as an on and off switch via conformational changes. Mutations that disrupt the linker's charge may lock SPRTN in an open, active state. While high proteolytic activity is beneficial for DPC cleavage, uncontrolled protease activity poses a great risk for cells especially near chromatin. Further NMR analysis of modified linker variants could assess conformational stability. Also, by monitoring spectral changes upon DNA and ubiquitin addition, where one would expect less changes than for the wild-type protein if the variant already has the open conformation. Preliminary experiments would be needed to determine whether a single point mutation, multiple substitutions or potentially exchanging all aspartate and glutamate residues, further enhance activity. However, extensive changes may also affect protein folding and stability and should be tested beforehand. Evaluating the activity of hyperactive mutants in cells will most likely require depletion or at least activity reduction of endogenous SPRTN, which can be achieved by using *SPRTN*$^{\Delta C}$ cells[353] or by siRNA-mediated knockdown. If the linker plays a regulatory role, its disruption could lead to constitutive SPRTN activation which could impair cell viability. Nonetheless, as there are other regulatory layers for SPRTN activity, like its DNA specificity, toxicity might be limited to conditions involving DPCs.

In contrast, Wss1 lacks a ZBD or a similar domain (Figure 22c) and seems unlikely to undergo large conformational changes upon activation. Despite this, Wss1 remains inactive in the absence of DNA[281]. Ubiquitin stimulation has not been tested so far, but given its structure, Wss1 likely does not require conformational changes, suggesting this mechanism to be specific for the SPRTN protease family. However, since Wss1 contains SIM motifs[285], SUMO may function similarly to ubiquitin in this context by activating the protease.

4.4 Allosteric activation of SPRTN by DPC ubiquitylation

Ubiquitin binding at the USD strongly enhanced SPRTN's DPC cleavage activity, with ubiquitylated DPCs being cleaved up to one hundred-fold more efficiently than unmodified. This suggests that DPC ubiquitylation[306,325,349,352–354,358,359,548] serves not only as a signal for proteasomal degradation but also directly activates SPRTN, revealing a novel regulatory mechanism for its DNA-dependent protease activity. Notably, ubiquitylated DPCs accumulate upon SPRTN depletion[368], supporting an essential non-redundant role of SPRTN alongside the proteasome. We and others observed no ubiquitin linkage specificity for SPRTN activation or binding[560], only ubiquitin chain length made a difference in our experiments. Therefore, SPRTN could act on DPCs modified with other types of ubiquitin chains than K48-linked chains, which are the main target of the proteasome[561]. By targeting these substrates, SPRTN would have a critical function in DPC repair distinct from the proteasome[306,348,353].

The proteasome is the main protein degradation machinery in cells, recognising K48-linked ubiquitylated targets[561], which were also found on DPCs in cells in combination with K63- and K11-linked Ub-chains[549]. SPRTN was able to cleave similarly modified DPCs *in vitro*, indicating similar activity in cells, acting in parallel with the proteasome. Besides the proteasome, DUBs are another example of proteases activated by ubiquitin. DUBs can be divided into different classes, one of them is the JAMM metalloprotease family[404,405]. Similar to SPRTN, they coordinate a Zn^{2+} ion within their active site which they use together with a nucleophilic water to hydrolyse the isopeptide linkage formed by ubiquitin[562]. However, DUBs solely cut Ub-chains, often in a target- and linkage-specific way[562,563]. SPRTN, on the other hand, cleaves diverse crosslinked protein adducts and could potentially also act on Ub-chains at the DPC. Other DPC proteases show varying degrees of ubiquitin dependency. Yeast Ddi1 cleaves only highly ubiquitylated substrates (chains of more than eight Ub-molecules) and polyubiquitylated DPCs accumulated with compromised Ddi1 activity[310]. However, the human homologs DDI1/DDI2 were described to shuttle DPCs to the proteasome rather than cleaving the protein adducts themselves[313]. The cleavage of polyubiquitylated DPCs performed by Ddi1 in yeast, therefore, could be taken over by SPRTN in humans. ACRC/GCNA, mainly expressed in germ cells, lacks UBDs but instead contains three SIMs, through which it recognises SUMOylated DPCs[291]. It may act upstream of SUMOylation-dependent ubiquitylation of DPCs[325,353,354,548,549] or target non-ubiquitylated DPCs. FAM111A and FAM111B both possess ubiquitin-like domains but ubiquitin binding has not been confirmed[317]. FAM111A's DPC-cleavage activity appears PCNA- and ssDNA-dependent[318]. Therefore, activation of SPRTN by DPC-ubiquitylation represents a unique mechanism among human DPC proteases.

Moreover, the allosteric activation of SPRTN by ubiquitin differs from other typical ubiquitin functions in DDR and DDT, where ubiquitin mainly recruits repair factors to lesion sites or marks proteins for degradation (see sections 1.4.4 and 1.4.5). Although SPRTN's UBZ domain may mediate recruitment of the enzyme[306,353], our *in vitro* experiments showed that even minimal SPRTN constructs with an intact USD can cleave ubiquitylated DPCs, indicating that the UBZ is not essential for DPC cleavage. Yet, cells with truncated SPRTN variants lacking the UBZ, showed reduced cleavage of 5-azadC-induced DNMT1-DPCs[353], a discrepancy further discussed in section 4.5.

Interestingly, besides enhancing general DPC cleavage by SPRTN, we observed additional smaller cleavage fragments running very close to the free DNA for ubiquitylated DPCs. Formation of these fragments was dependent on ubiquitin binding to the USD, as they were almost absent for the USD variant $\mathrm{SPRTN}^{\mathrm{L99S}}$. It would be interesting to further investigate these cleavage fragments and identify precise cleavage sites. To minimise the remaining peptide adduct makes sense, as it facilitates downstream repair mechanisms, for instance TLS polymerase bypass[352]. Of note, the final fate of the remaining peptide remains unclear. Mapping SPRTN cleavage sites could improve understanding of these processes and potentially indicate certain substrate-specificity. Our preliminary MS-approaches were complicated by excess free non-crosslinked $\mathrm{HMCES}^{\mathrm{SRAP}}$ and DNA attached to the fragments, both interfering with MS measurements. Further optimisation would be needed, such as removing DNA at the end of the reaction using nucleases or increasing protein concentration to enable visualisation of cleavage products on SDS-PAGE gels with protein-based staining (e.g. Coomassie). In this way, different fragments and free protein could be better separated by cutting out the corresponding bands from the gel after electrophoresis. However, size differences between free and crosslinked $\mathrm{HMCES}^{\mathrm{SRAP}}$ are minor, making it hard to distinguish them. Alternatively, the DPC could be purified using size exclusion chromatography (SEC) to separate it from free DNA and protein, before usage in cleavage assays. Coupling SEC to MS, could also improve peptide separation, an approach which could be especially useful for analysis of SPRTN autocleavage fragments. We initially used GluC for MS analysis of semi-specific peptides, as HMCES-DPC cleavage products of SPRTN and trypsin (standardly used for protein digestion before MS) appeared very similar on SDS-PAGE gels. However, direct comparison of SPRTN and trypsin cleavage without additional sample digestion may be a better approach. Moreover, this set-up could answer the longstanding question of whether SPRTN shares cleavage sites with trypsin. Alternatively, Cosenza-Contreras *et al.* developed a method for MS to identify neo-*N*-termini arising from endogenous proteolytic activity[564]. They specifically modify *N*-termini before MS measurements to distinguish *N*-termini formed by the protease of interest from cleavage sites formed during experimental proteolytic digestion, for example by trypsin[564]. In our case this could help to distinguish SPRTN-generated fragments from digestion products. Mapping SPRTN cleavage sites could improve understanding of the protease on multiple levels. First, it could help identify other substrates for SPRTN by searching for similar motifs in other DPCs. Secondly, for the ubiquitylated DPCs it could determine whether SPRTN cuts progressively or in distinct steps, which then could be verified by synthesising or purifying identified cleavage products and using them in SPRTN cleavage assays or by mutating cleavage sites in HMCES.

Even though the USD-variant SPRTN^{L99S} lacked additional DPC cleavage products for ubiquitylated DPCs, general DPC cleavage activity was still increased, suggesting that other factors in the reaction could also be affected. FANCJ is known to unfold HMCES-DPCs, enabling SPRTN cleavage *in vitro*[266], which we used in our experiments. We investigated whether DPC ubiquitylation alters FANCJ activity and whether unmodified and ubiquitylated DPCs are unfolded in the same way. Using a HMCES^{SRAP} variant (R98E) that can be thermally unfolded, enabling SPRTN cleavage without FANCJ[266] and adding free tetraUb to reactions, we could confirm that formation of additional cleavage fragments does not per se require FANCJ but relies on a functional USD. Nevertheless, limited proteolysis experiments with trypsin following FANCJ-mediated DPC unfolding showed different cleavage patterns for unmodified and ubiquitylated DPCs, implying that DPC ubiquitylation indeed affects FANCJ activity. This could explain why USD variants lacked additional cleavage fragments but still showed enhanced general DPC cleavage. Although FANCJ has no known UBDs, DPC ubiquitylation may alter how it processes DPCs. One hypothesis is that FANCJ stalls at the DPC and partially displaces the protein adduct to expose the underlying DNA, allowing SPRTN to bind. Larger DPCs, like highly ubiquitylated ones, could cause more FANCJ stalling and thereby create more accessible SPRTN substrates and promote cleavage. However, SUMOylated DPCs, despite having a similar size to HMCES^{SRAP}-Ub$^{[short]}$-DPCs, did not increase DPC cleavage, suggesting a ubiquitin-specific mechanism. To test whether FANCJ stalls at the DPC or goes through it, a HaeIII cleavage site could be included in the DNA for crosslinking, as previously described[266]. If FANCJ goes through the DPCs and the HaeIII cleavage site is downstream of the DPC, it would separate the two DNA strands, prohibiting HaeIII cleavage. If it stalls in front of the DPCs, DNA strands remain annealed forming a functional HaeIII cleavage site.

Notably, AlphaFold-based protein structure predictions identify the Ub-E3-ligase RFWD3, involved in DPC ubiquitylation[349,350], as a likely FANCJ interactor[565]. Since both proteins localise to stalled replication forks[266,349,350], a functional interaction is plausible where either of the two proteins could recruit the other to the fork, leading to DPC ubiquitylation by RFWD3 and DPC unfolding by FANCJ, but further experiments, involving *in vitro* reconstitution, are needed, to confirm this predicted interaction.

Importantly, SPRTN also cleaved DPCs when DPC ubiquitylation was prohibited in *X. laevis* egg extract[306,348]. Although this seems contradictory to our findings at first, DPC cleavage in these cases likely reflects SPRTN's basal, Ub-independent activity, which we also observed in our experiments.

4.5 The USD's role in Ruijs-Aalfs syndrome SPRTN variants

Amino acid substitutions within SPRTN's USD interface reduced cleavage of ubiquitylated DPCs *in vitro* and of DNMT1-DPCs in cells, though it did not fully abolish SPRTN activity. The functional importance of the USD became evident in the hypomorphic SPRTN$^{\Delta C}$ RJALS patient variant. Full-length SPRTN with a L99S substitution rescued all phenotypes caused by *Sprtn* loss in conditional knockout MEFs, similar to wild-type SPRTN$^{\Delta C}$. Introducing the same mutation in SPRTN$^{\Delta C}$ however, severely impaired cell fitness and genome stability, indicated by increased micronuclei and chromatin bridge formation, hallmarks for missegregated chromosomes[566] and chromosome entanglement[567], respectively. Phenotypes previously observed in *Sprtn* conditional knockout MEFs[298]. These findings suggest that minimal SPRTN activity is sufficient to fulfil SPRTN's role in genome stability maintenance. Notably, SPRTN$^{\Delta C}$ and SPRTN$^{\Delta C\text{-}L99S}$ had no defect in DPC cleavage of synthetically ubiquitylated DPCs. SPRTN$^{\Delta C\text{-}L99S}$ failed to produce additional, smaller cleavage fragments, similar to full-length SPRTN^{L99S}. In contrast, cleavage of SUMO-targeted ubiquitylated HMCES-DPCs was reduced with SPRTN$^{\Delta C}$ and further impaired by replacing Leu99. Interestingly, this phenotype cannot be completely explained by loss of the *C*-terminal tail and the UBZ, as SPRTN$^{\Delta UBZ}$ cleaved SUMO-targeted ubiquitylated DPCs comparable to full-length SPRTN. Notably, SPRTN$^{\Delta C}$ also experiences decreased DNA binding[295,301], which may explain its decreased activity on heavily modified DPCs, where access to the underlying DNA is restricted. MS analysis of SUMO-targeted and synthetically ubiquitylated DPCs revealed clear differences in the modification patterns. In the synthetic system DPCs were solely ubiquitylated on the *C*-terminally fused Ub-moiety, whereas SUMO-targeted ubiquitylation modified several lysine residues within both HMCES and SUMO. This likely increases steric hinderance, requiring full DNA-binding capacity of full-length SPRTN for efficient cleavage.

Although SPRTN's UBZ is dispensable *in vitro*, it is essential for SPRTN function in cells[353] and in *X. laevis* egg extract[306], likely facilitating DPC recognition and SPRTN recruitment to the damage site[353]. Different from our *in vitro* set-up this might be more important in the crowded environment of a cell. In cells, the UBZ also ensures SPRTN monoubiquitylation, which regulates protein levels[365]. This regulatory function explains higher expressions of SPRTN$^{\Delta C}$ variants compared to full-length SPRTN in overexpressing cell lines[293]. Nevertheless, our pull-down assays indicated that ubiquitin binding is primarily mediated by the UBZ, whereas NMR analysis confirmed ubiquitin interactions at the USD. Of note, due to the size of full-length SPRTN, it cannot be used in NMR analysis. However, Ub-interactions at the USD appeared weak with low affinity. The interactions are likely transient and may not be stable enough to be detected in pull-down assays. Unlike human SPRTN, its *Drosophila* ortholog, Mh, contains two *C*-terminal UBZs[297], but ubiquitin interactions have not been experimentally confirmed. Interestingly, Mh interferes with Top2-mediated processing of satellite DNA in cross-species experiments of *D. melanogaster* and *D. simulans*[568], suggesting that Mh cleaves Top2ccs. Since human TOP2ccs are SUMOylated and ubiquitylated[335,337], it would be interesting to investigate Top2 modifications in *Drosophila* and test whether Mh is stimulated by ubiquitin *in vitro* with recombinant proteins.

In conclusion, while SPRTN's UBZ is crucial for its regulation and recruitment in cells SPRTN$^{\Delta C}$ variants remain viable[293], yet depend heavily on a functional USD to preserve DPC processing and genome stability.

4.6 Regulation of SPRTN activity by DPC ubiquitylation

Combining our findings, we propose a partially speculative "triple lock" model (Figure 23) for how DPC ubiquitylation regulates SPRTN's DNA-dependent protease activity. In this model, SPRTN gets activated through three mechanisms. First, the UBZ recruits SPRTN to the ubiquitylated DPC, as previously suggested[306,353]. Second, binding to an activating ssDNA-dsDNA junction induces a conformational shift to an open, active state, allowing access to SPRTN's active site. Third, this open and active conformation is further stabilised by ubiquitin binding to SPRTN's USD, enhancing proteolysis of crosslinked proteins.

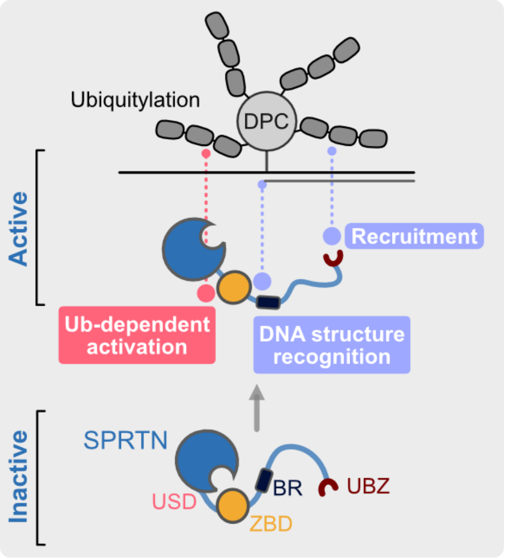


Figure 23. "Triple-lock" model for SPRTN regulation and activation. SPRTN's ubiquitin-binding zinc-finger (UBZ) binds to ubiquitylated DNA-protein crosslinks (DPCs) and recruits SPRTN. DNA-binding of its two DNA-binding domains, the zinc-binding domain (ZBD) to ssDNA and the basic region (BR) to dsDNA, triggers a conformation change towards an open and active state. Ubiquitin (Ub) binding to the ubiquitin-binding interface at the SprT domain (USD) stabilises this open conformation, ensuring efficient DPC cleavage. Figure adapted from Dürauer *et al.*[1].

A remaining question is how DPCs would be unfolded during replication-independent repair to allow cleavage by SPRTN. DNMT1-DPC cleavage by SPRTN in a global-genome context[353] could be linked to the USD, yet FANCJ cannot act here due to the absence of ssDNA needed for its loading[266]. Moreover, helicase activity outside of DNA transactions like replication or transcription threatens genome integrity.

5-azadC-induced DNMT1-DPCs are ubiquitylated by RNF4 and TOPORS and rapidly targeted by p97, leading to proteasomal degradation[325,353,354,548,549]. In yeast, Wss1 interacts with the p97 homolog Cdc48 via Ubx5[286], similar mechanisms may exist in human cells, as SPRTN bears a p97 interaction motif (SHP-box). While p97 can extract chromatin-bound PARP1 and enhance DPC cleavage by SPRTN *in vitro*, its exact role in mammalian DPC repair remains unclear[569]. For DNMT1-DPC repair p97 was proposed to act upstream of the proteasome, where it could potentially also facilitate SPRTN activity[353].

Considering the dsDNA context in the scenario of 5-azadC-induced DNMT1-DPCs, it is questionable how a ssDNA-dsDNA junction needed for SPRTN activation would form. One possibility is that when p97 threads the protein adduct through its central pore it exerts a pulling force on the DNA, locally separating the two DNA strands. This could form a bubble, sufficient for SPRTN activation[287]. This is supported by evidence of p97 translocating small DNA oligonucleotides through its catalytic centre *in vitro*[569], though this may not be representative for chromosomal DNA. In this study, ubiquitylated Eos was used as model DPC, and p97 unfolded it using Ufd1-Npl4, as Ub-adaptors[569], however different adaptors may be involved for other DPCs. Future studies should aim to identify p97 adaptors and cofactors involved in DPC repair, for example, via co-immunoprecipitation following DPC induction in cells (e.g. DNMT1-DPCs induced by 5-azadC or HMCES-DPCs induced by CD437). Alternatively, recombinant p97 and Ufd1-Npl4 could be used with synthetically ubiquitylated DPCs *in vitro*, which allows optimisation and variation of Ub-chain types and length. By this, ideal conditions for p97 could be identified, while coupling the reaction with SPRTN DPC cleavage. Comparative studies between different unfoldases (e.g. p97 vs. FANCJ) and resulting SPRTN cleavage products could also illuminate how cells decide between proteasomal and SPRTN-mediated DPC degradation.

In summary, our findings reveal an additional layer for SPRTN regulation and provide tools and experimental set-ups for further investigation of DPC repair mechanisms, including the coordination between p97, the proteasome and SPRTN in proteolytic processing of DPCs.

Acknowledgements

Time really does fly, it's hard to believe it's been four years since I started my PhD, and ten since I finished school and started university. If someone had told me back then that I'd end up doing a PhD in Biochemistry I would've laughed and yet here I am, grateful I landed exactly where I am. These past four years have been intense, emotional and full of growth. To give some numbers, during this time I filled 5 hand-written lab books, ran 728 experiments, cloned 177 plasmids, 144 of them SPRTN-related and since I joined the lab we purified 381 proteins, with one-third being SPRTN-preps. There was laughter, joy, frustration, and even a few tears (and I rarely cry) but still I'd do it all over again. Doing my PhD in the Stingele lab was one of the best decisions of my life and I am deeply thankful to everyone who supported me throughout this journey.

First, I want to thank my supervisor, Julian. Thank you for believing in me from day one, despite my lack of experience with protein purification. You gave me the freedom to grow, offered guidance when needed and always knew what to say and do when things got difficult or in moments of doubt (even when I thought Ramadan is scoping me). Your energy, optimism and creativity make this lab an incredible place and I've learned a lot from you. Thank you for creating such an inspiring lab environment and for giving me the opportunity to be part of it.

I'm deeply grateful to my wonderful lab colleagues for their support, friendship and all the fun we had together. Here are a few words for each of you. Pedro, my forever favourite Spanish guy. You are a brilliant and crazy fast-thinking scientist and truly the king of cloning. It was a joy working with you, I've learnt so much from you and I'm very proud of our shared paper. Also, outside the lab I'm happy I can call you my friend, you really are the life of the party and it's always fun with you. I am sure you will do great whatever path you pick, and I hope to see you soon again. Denitsa, my bench and desk partner, we never had a project together but for sure I wouldn't have a single paper without you. You helped me get started in the lab, introduced me to the in vitro world and were ready to jump in and pipette like crazy when things got busy. You really are the in vitro queen, and I feel honoured I got to learn from you and truly admire your dedication and persistence. You also became a really good friend of mine and I cherish our coffee and smoke breaks where we could talk about everything. Max D. (I know you are really curious to read what I have to say about you), despite our occasional disagreements, I'm happy I got to work with you and proud of what we achieved together. You might be a bit chaotic, but your creativity always amazes me, and I am happy we became friends along the way. Maybe you can't drink as much as me, but we certainly share our love for Austria. Jackie, you really are the emotional glue of the lab, always having encouraging and kind words for everybody. I loved sharing an office and a lab bay with you and truly admire efficient working style. Shubo, your work ethics are unmatchable, and your positive energy and always friendly mindset is simply contagious. Although I really think you should work less, it was nice to not be alone in the lab on weekends. I've really enjoyed our running club and I'm certain you will thrive as a PI with your own lab. Max G., I have no other word for you then genius, your knowledge is astonishing and I'm very thankful for all our scientific discussions and your help with mass spec data analysis. Anja, the best technician in the whole world. Without your organisation the lab would fall apart. You were always cheerful and ready to help, thank you for everything. Ricky, thank you for your endless patience, even when you had to explain the Incucyte to me for the fifth times, sometimes even via video call. You are a great scientist, and I'm glad I got to work with you. Carla, you impress me so much with both your talent in wet lab and bioinformatics, all while pursuing so many cool hobbies outside the lab and being a super nice person to have around. Hannah, you have already left the lab when I joined but thank you for laying the foundation for my SPRTN work and for always being there to answer questions and stopping by the lab to show me things. To the newer lab members, Fabio, Alan and Shuhan, you were great additions to the team, and it was a pleasure working with you. I look forward to seeing where your projects lead. I also want to mention the extended Roche part of the lab, Niels and Kerstin. Even though we didn't see each other often, I always had fun with you at retreats and parties. You've successfully transitioned to industry, and I'll try to follow your lead now. Supervising students was one of the most rewarding parts of my PhD and I want to thank Naira, Melisa and Dina, for their contributions to the SPRTN-Ubiquitin project. Dina, we spent the most time together and now you're even a PhD student in the lab. I know it was not always easy, but I am very proud and impressed by how organised and structured you approach your own project now. I'm already excited to read your paper on p97 and SPRTN.

I also want to mention and thank our neighbours, the Jae lab. Thank you for sharing autoclaved water and tips when needed and just good company. Especially Eva, the impressive first PhD to graduate in this lab. You are truly amazing and an inspiration with your endurance and resilience. I miss our office chats and coffee breaks but I'm glad you like your new job in industry and to have you as a friend.

Throughout all my projects I had amazing collaboration partners, I am deeply thankful for their contributions, which greatly enriched our publications. Hanso, thank you for the many NMR measurements you did and your patience with explaining the data to me and even helping me when I was struggling with Pymol. I will always think of you when I pass by the Lotto shop at Marienplatz, our dealing spot for 15N media and SPRTN preps. Christian and Nadine, I still don't really understand how the molecular dynamics simulations work, but they were an incredible addition to our paper. Thank you for your contributions and your quick reactions after every discussion. Also, thanks for giving me the chance to once see my SPRTN in 3D. Chris, it's been three years since we first met and talked about a collaboration at the ubiquitin conference in Croatia and since then you've prepared and sent us many proteins. Thanks for an amazing collaboration and our nice talks over a couple of beers at the SFB retreats. Sophie, teaming up with another Sophie from Austria was just bound for success. Thank you for helping me get the best SPRTN ever. Gregor, I started working with you in my first year and kept on coming back to you for help with SAXS, testing new methods or IT stuff. Thank you for all your support and for being part of my thesis committee.

I also want to thank my TAC members, Charo Robles and Karl Duderstadt for their guidance and support throughout this journey and for now both being part of my thesis committee. Thank you also to Lena Daumann and the LMU mentoring program for female junior researchers. I'm also grateful for the R4 community with the Nedialkova, Hampel and Duderstadt Lab for the scientific discussions and fun we had at get-togethers, conferences and the Oktoberfest.

Thank you also to the IMPRS-LS, my graduate school for all the opportunities and friendships it brought. Especially, Adrian, Anja, Filip, Leona, Yusuf, Henry, Maria, Sveto and Filiz. I always had lots of fun with you, and it was great to have friends around in the same situation who understood the struggles of doing a PhD.

Outside of science, I'm thankful for friends who reminded me, that there actually is a life outside of the lab. Thank you, Melissa, Annabelle, Martin, Matthias and Jelica. Georg, you are like the brother I never had. Living together for three years, we had great talks, fun adventures and crazy nights out partying. I'm very much looking forward to seeing you soon back in Germany or when I come visit you in Boston. Marlene, we became friends from the first day we met and throughout this whole journey you were always there for me, in the highs and lows, whether they were personal or lab related. I really appreciate all our shared adventures and look forward to the next ones, also in your new home Mainz. Laura, you have been by my side from the very beginning, back in Vienna when I was applying for PhD positions. With your positive energy you always lifted me up and I feel very lucky to have you as a friend and that you also moved to Munich, so we could spend so much time together.

To my friends back in Austria, thank you for keeping in touch, visiting me in Munich and always welcoming me when I was back home. Your support means a lot to me!

Ein großes Danke geht an meine Familie, meine Eltern, meine Schwestern Hannah, Jasmin und ihren Mann Manuel. Die letzten vier Jahre waren nicht leicht und es tut mir leid, dass ich nicht immer vor Ort bei euch sein konnte. Trotz allem habt ihr mich immer bedingungslos unterstützt und an mich geglaubt. Schon nach meinem ersten Gespräch mit Julian habt ihr mich bestärkt den Schritt nach München zu wagen und das, obwohl die Welt der Wissenschaft noch immer etwas befremdlich auf euch wirkt. Ich weiß ich kann immer auf euch zählen und dafür möchte ich einfach nur DANKE sagen. Ich möchte meine Doktorarbeit auch meiner Oma widmen, die diesen Moment leider nicht mehr miterleben kann. Sie hat mich über mein ganzes Studium über unterstützt und war immer stolz auf mich. Liebe Oma, ich bin jetzt Frau Doktor, ganz ohne heiraten.

Und das Beste kommt zum Schluss, Lukas. In den letzten vier Jahren warst du meine größte Stütze. Oft genug warst du auch der Einzige, der sich Sorgen um meine physische und psychische Gesundheit gemacht hat und mich bremste, wenn ich es wieder mal übertrieb. Nach zwei Jahren Fernbeziehung und viel Zeit im Zug, bist du für mich nach München gezogen. Ich weiß das war nicht leicht für dich und es bedeutet mir sehr viel. Dass wir uns ausgerechnet an dem Tag kennengelernt haben, an dem ich das Angebot für die Stelle in München bekam, bleibt für mich ein Zeichen. Denn du hast dich trotzdem auf mich eingelassen und mir sind zwei der besten Dinge in meinem Leben am gleichen Tag passiert. Danke für Alles Lukas!

Now, there is only one thing left to say…
Danke, es war mir ein Volksfest!

List of abbreviations

3'-PUA	3'-polyunsaturated aldehyde
5-azadC	5-aza-2'-deoxycytidine
5'dRP	5’-deoxyribose phosphate
5fC	5-formylcytosine
5-hmC	5-hydroxymethylcytosine
5mC	5-methly cytosine
6-4 PPs	6-4 photoproducts
8-oxoG	8-oxo-7,8-dihydroguanine
A	Adenine
ACRC	Acidic repeat-containing protein
ALT	Alternative telomere lengthening
ALK	Anaplastic lymphoma kinase
AP site	Apurinic/apyrimidinic site
APC/C	Anaphase-promoting complex/cyclosome
APE1	AP endonuclease 1
APIM	AlkB2 PCNA-interaction motif
ARK	Advance recovery of K-SDS precipitates
ATF3	Activating transcription factor 3
ATG	Autophagy-related gene
ATM	Ataxia telangiectasia mutated
ATP	Adenosine triphosphate
ATR	Ataxia telangiectasia and Rad3 related
BER	Base excision repair
BIR	Break-induced replication
BLM	Bloom helicase
BR	Basic region
BRCA1	Breast cancer type
BTR complex	BLM, TOP3α, RMI1 and RMI2 complex
BUDR	BRCT-domain-associated ubiquitin-dependent recruitment motif
C	Cytosine
CETN2	Centrin 2
ChIP-Seq	Chromatin immunoprecipitation sequencing
CMG	Cdc45-MCM-GINS

cNHEJ	Classical NHEJ
CPD	Cyclobutene pyrimidine dimer
CPD	Cyclobutene pyrimidine dimers
CPT	Camptothecin
CRL4A^{DDB2}	DDB-CUL4A-RBX1 E3-ligase complex
CSB	Cockayne syndrome group B protein
CsCl	Caesium chloride
CtIP	C-terminal-binding protein-interacting protein
CUE	Coupling of ubiquitin conjugation to endoplasmic reticulum
Ddi1	DNA damage inducible 1
DDR	DNA damage response
DDT	DNA damage tolerance
dL	2-deoxyribonolactone
D-loop	Displacement loop
DNA2	DNA replication ATP-dependent helicase/nuclease DNA2
DNA-PK	DNA-dependent protein kinase
DNMT1	DNA methyltransferase 1
DPC	DNA-protein crosslink
dsDNA	Double-stranded DNA
DUB	Deubiquitylating enzyme
DUIM	Double-sided UIM
EBNA1	Epstein-Barr virus nuclear antigen
EME	Essential meiotic endonuclease 1 homolog
ERAD	Endoplasmic reticulum-associated degradation
ERCC1	Excision repair cross-complementing group 1
FA	Fanconi anaemia
FAAP24	FA core complex associated protein 24
FANC	Fanconi anaemia complementation group
FAT10	HLA-F adjacent transcript 10
FEN1	Flap structure-specific endonuclease 1
G	Guanine
GAT	GGA and TOM
GCNA	Germ cell nuclear antigen
GGA	Golgi-localized, Gamma-ear-containing, ARF-binding protein 3
GG-NER	Global-genome NER

GLUE	Gram-like ubiquitin-binding in Eap45
HECT	Homologous to the E6AP carboxyl terminus
HIV	Human immunodeficiency virus-1
HLTF	Helicase-like transcription factor
HMCES	5-hydroxymethylcytosine binding ESC-specific
HPF1	Histone PARylation factor 1
HR	Homologous recombination
ICL	Interstrand crosslink
ID2	FANCD2-FANCI complex
IFN	Interferon
IR	Ionizing radiation
ISG15	Interferon-stimulated gene 15
IUIM	Inverted ubiquitin interacting motif
$KBrO_3$	Potassium bromate
KCl-SDS	Potassium chloride and sodium dodecyl sulphate coprecipitation assay
LIG4	DNA-ligase 4
MDC1	Mediator of DNA damage checkpoint protein 1
MEF	Mouse embryonic fibroblasts
MGMT	Methylguanine DNA methyltransferase
Mh	Maternal haploid
MHF	FANCM-associated histone fold protein
MMEJ	Microhomology-mediated end joining
MMR	Mismatch repair
MMS	Methyl methane sulfonate
MNNG	*N*-methyl-*N*'-nitro-*N*-nitrosoguanidine
MRE11	Meiotic recombination 11
MRN complex	Complex formed by MRE11, RAD50 and NSB1
MS	Mass spectrometry
MVB	Multi-vesicular body
NAD+	Nicotinamide adenine dinucleotide
NBS1	Nijmegen-breakage syndrome protein 1
NEDD8	Neural precursor cell expressed developmentally downregulated protein 8
NEIL3	Nei endonuclease VIII-like 3
NER	Nucleotide excision repair
NF-κB	Nuclear factor kappa B

NHEJ	Non-homologous end joining
NLS	Nuclear localization signal
Npl4	Nuclear protein localization 4
NRF1	Nuclear respiratory factor 1
NZF	Npl4 ZnF
OGG1	8-oxoguanine glycosylase 1
PAD	Polymerase-associated domain
PARG	Poly-ADP-ribose glycohydrolase
PARP1	Poly (ADP-ribose) polymerase 1
PARPi	PARP inhibitor
PCNA	Proliferating cell nuclear antigen
PH	Plekstrin homology
PI	Phosphoinositide
PIAS	Protein inhibitor of activated STAT
PIP	PCNA-interacting protein
PNKP	Polynucleotide kinase/phosphatase
Pol	Polymerase
PolDIP2	Polδ-interacting protein 2
PRU	Plekstrin-like receptor for ubiquitin
PTM	Post-translational modification
PxP	Purification of x-linked proteins
PxxP	Proline-rich motifs
RAD23B	UV excision repair protein RAD23 homolog B
RADAR	Rapid approach to DNA adducts recovery
RBR	RING-between-RING
RECQ1	RecQ-Like Type 1
REV1	Reversionless 1
RFWD3	Ring finger and WD repeat domain 3
RING	Really interesting new gene
RJALS	Ruijs-Aalfs syndrome
RNF4	Ring finger protein 4
RNS	Reactive nitrogen species
ROS	Reactive oxygen species
RPC	RNA-protein crosslink
RPA	Replication protein A

RTEL1	Regulator of telomere elongation 1
SCAN1	Spinocerebellar ataxia with axonal neuropathy
SCAR23	Spinocerebellar ataxia, autosomal recessive 23
SCE	Sister chromatid exchange
SCF	SKP1, CUL1, F-box protein
SDSA	Synthesis-dependent strand annealing
SEC	Size exclusion chromatography
SH3	Src homology 3
SHPRH	SNF2 histone linker PHD RING helicase
SIM	SUMO interacting motif
SMARCAL1	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A-like protein 1
SPRTN	SprT-like *N*-terminal domain
SRAP	SOS response-associated peptidase
SRS2	SHI related sequence 2
SSA	Single-strand annealing
ssDNA	Single-stranded DNA
STAR	Superior method for True DNA-protein crosslink Recovery
STUbL	SUMO-targeted ubiquitin E3-ligase
SUMO	Small ubiquitin-related modifier
T	Thymine
TC-NER	Transcription-coupled NER
TDP	Tyrosyl-DNA phosphodiesterase
TET	Ten-eleven translocation
TEX264	Testis-expressed protein 264
TFIIH	Transcription factor IIH
TLS	Translesion synthesis
TOM1	Target of Myb1
TOP	Topoisomerase
TOP1ccs	TOP1 cleavage complex
TOPORS	TOP1 binding arginine/serine rich protein
TP	Terminal protein
TRAIP	TRAF interacting protein
TS	Template switching
U	Uracil

Ub	Ubiquitin
UBA	Ubiquitin associated
UBC	Ubiquitin-conjugation
UBD	Ubiquitin-binding domain
UBL	Ubiquitin-like proteins
UBZ	Ubiquitin-binding ZnF
UDG	Uracil-DNA glycosylase
UdgX	Uracil-DNA glycosylase X
UEV	Ubiquitin-conjugating enzyme E2 variant
UFM	Ubiquitin-fold modifier
UIM	Ubiquitin interacting motif
UMI	UIM- and MIU-related
URM	Ubiquitin-related modifier
USD	Ubiquitin-binding interface at the SprT domain
USP7	Ubiquitin-specific protein 7
UV	Ultraviolet radiation
UV-DDB	UV-damaged DNA-binding protein
UVSSA	UV-sensitive syndrome protein A
WRN	Werner-Syndrome
Wss1	Weak suppressor of smt3
XPC	Xeroderma pigmentosum factor C
XRCC1	X-Ray repair cross complementing 1
XRCC1	X-ray repair cross-complementing protein 1
ZBD	Zinc-binding domain
ZnF	Zinc finger
ZRANB3	Zinc finger RANBP2-type containing 3

References

1. Dürauer, S. *et al.* Allosteric activation of the SPRTN protease by ubiquitin maintains genome stability. *Nat Commun* **16**, 5422 (2025).
2. Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362**, 709–715 (1993).
3. Chatterjee, N. & Walker, G. C. Mechanisms of DNA damage, repair, and mutagenesis. *Environmental and Molecular Mutagenesis* **58**, 235–263 (2017).
4. Tubbs, A. & Nussenzweig, A. Endogenous DNA Damage as a Source of Genomic Instability in Cancer. *Cell* **168**, 644–656 (2017).
5. 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).
6. Weickert, P. & Stingele, J. DNA-Protein Crosslinks and Their Resolution. *Annu Rev Biochem* **91**, 157–181 (2022).
7. Loeb, L. A. & Monnat, R. J. DNA polymerases and human disease. *Nat Rev Genet* **9**, 594–604 (2008).
8. Kunkel, T. A. Evolving Views of DNA Replication (In)Fidelity. *Cold Spring Harb Symp Quant Biol* **74**, 91–101 (2009).
9. Kunkel, T. A. Balancing eukaryotic replication asymmetry with replication fidelity. *Current Opinion in Chemical Biology* **15**, 620–626 (2011).
10. Goodman, M. F. & Woodgate, R. Translesion DNA polymerases. *Cold Spring Harb Perspect Biol* **5**, a010363 (2013).
11. 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).
12. Swanton, C., McGranahan, N., Starrett, G. J. & Harris, R. S. APOBEC Enzymes: Mutagenic Fuel for Cancer Evolution and Heterogeneity. *Cancer Discov* **5**, 704–712 (2015).
13. Langenbucher, A. *et al.* An extended APOBEC3A mutation signature in cancer. *Nat Commun* **12**, 1602 (2021).
14. Petljak, M. *et al.* Mechanisms of APOBEC3 mutagenesis in human cancer cells. *Nature* **607**, 799–807 (2022).
15. Lindahl, T. & Barnes, D. E. Repair of endogenous DNA damage. *Cold Spring Harb Symp Quant Biol* **65**, 127–133 (2000).
16. 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**, 166–172 (1992).

17. Wilson, D. M. & Barsky, D. The major human abasic endonuclease: formation, consequences and repair of abasic lesions in DNA. *Mutation Research/DNA Repair* **485**, 283–307 (2001).
18. Gates, K. S. An overview of chemical processes that damage cellular DNA: spontaneous hydrolysis, alkylation, and reactions with radicals. *Chem Res Toxicol* **22**, 1747–1760 (2009).
19. Grady, W. M. & Ulrich, C. M. DNA alkylation and DNA methylation: cooperating mechanisms driving the formation of colorectal adenomas and adenocarcinomas? *Gut* **56**, 318–320 (2007).
20. Ballschmiter, K. Pattern and sources of naturally produced organohalogens in the marine environment: biogenic formation of organohalogens. *Chemosphere* **52**, 313–324 (2003).
21. Hecht, S. S. DNA adduct formation from tobacco-specific *N*-nitrosamines. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **424**, 127–142 (1999).
22. Hamilton, J. T. G., McRoberts, W. C., Keppler, F., Kalin, R. M. & Harper, D. B. Chloride Methylation by Plant Pectin: An Efficient Environmentally Significant Process. *Science* **301**, 206–209 (2003).
23. Wyatt, M. D. & Pittman, D. L. Methylating agents and DNA repair responses: Methylated bases and sources of strand breaks. *Chem Res Toxicol* **19**, 1580–1594 (2006).
24. Newlands, E. S., Stevens, M. F., Wedge, S. R., Wheelhouse, R. T. & Brock, C. Temozolomide: a review of its discovery, chemical properties, pre-clinical development and clinical trials. *Cancer Treat Rev* **23**, 35–61 (1997).
25. Ballmaier, D. & Epe, B. Oxidative DNA damage induced by potassium bromate under cell-free conditions and in mammalian cells. *Carcinogenesis* **16**, 335–342 (1995).
26. Geacintov, N. E. Is intercalation a critical factor in the covalent binding of mutagenic and tumorigenic polycyclic aromatic diol epoxides to DNA? *Carcinogenesis* **7**, 759–766 (1986).
27. Vignard, J., Mirey, G. & Salles, B. Ionizing-radiation induced DNA double-strand breaks: A direct and indirect lighting up. *Radiotherapy and Oncology* **108**, 362–369 (2013).
28. Srinivas, U. S., Tan, B. W. Q., Vellayappan, B. A. & Jeyasekharan, A. D. ROS and the DNA damage response in cancer. *Redox Biol* **25**, 101084 (2019).
29. Shrivastav, N., Li, D. & Essigmann, J. M. Chemical biology of mutagenesis and DNA repair: cellular responses to DNA alkylation. *Carcinogenesis* **31**, 59–70 (2010).
30. Davies, R. J. Royal Irish Academy Medal Lecture. Ultraviolet radiation damage in DNA. *Biochem Soc Trans* **23**, 407–418 (1995).

31. Mitchell, D. L. & Nairn, R. S. The Biology of the (6–4) Photoproduct. *Photochemistry and Photobiology* **49**, 805–819 (1989).
32. Yu, S.-L. & Lee, S.-K. Ultraviolet radiation: DNA damage, repair, and human disorders. *Mol. Cell. Toxicol.* **13**, 21–28 (2017).
33. 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).
34. 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).
35. Huang, Y. & Li, L. DNA crosslinking damage and cancer - a tale of friend and foe. *Transl Cancer Res* **2**, 144–154 (2013).
36. Stingele, J., Bellelli, R. & Boulton, S. J. Mechanisms of DNA-protein crosslink repair. *Nat Rev Mol Cell Biol* **18**, 563–573 (2017).
37. Enoiu, M., Jiricny, J. & Schärer, O. D. Repair of cisplatin-induced DNA interstrand crosslinks by a replication-independent pathway involving transcription-coupled repair and translesion synthesis. *Nucleic Acids Res* **40**, 8953–8964 (2012).
38. Tretyakova, N. Y., Groehler, A. & Ji, S. DNA-Protein Cross-Links: Formation, Structural Identities, and Biological Outcomes. *Acc Chem Res* **48**, 1631–1644 (2015).
39. Noll, D. M., Mason, T. M. & Miller, P. S. Formation and repair of interstrand cross-links in DNA. *Chem Rev* **106**, 277–301 (2006).
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. Mehta, A. & Haber, J. E. Sources of DNA double-strand breaks and models of recombinational DNA repair. *Cold Spring Harb Perspect Biol* **6**, a016428 (2014).
42. Desouky, O., Ding, N. & Zhou, G. Targeted and non-targeted effects of ionizing radiation. *Journal of Radiation Research and Applied Sciences* **8**, 247–254 (2015).
43. Henner, W. D., Grunberg, S. M. & Haseltine, W. A. Sites and structure of gamma radiation-induced DNA strand breaks. *J Biol Chem* **257**, 11750–11754 (1982).
44. 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**, 711–713 (1983).
45. Borrego-Soto, G., Ortiz-López, R. & Rojas-Martínez, A. Ionizing radiation-induced DNA injury and damage detection in patients with breast cancer. *Genet Mol Biol* **38**, 420–432 (2015).

46. Crossley, M. P., Bocek, M. & Cimprich, K. A. R-Loops as Cellular Regulators and Genomic Threats. *Mol Cell* **73**, 398–411 (2019).
47. Mir, A. A., Philippe, C. & Cristofari, G. euL1db: the European database of L1HS retrotransposon insertions in humans. *Nucleic Acids Res* **43**, D43-47 (2015).
48. 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 International* **6**, 13 (2006).
49. 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**, 1383–1393 (2006).
50. 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).
51. Fu, D., Calvo, J. A. & Samson, L. D. Balancing repair and tolerance of DNA damage caused by alkylating agents. *Nat Rev Cancer* **12**, 104–120 (2012).
52. Sancar, A. Structure and Function of DNA Photolyase and Cryptochrome Blue-Light Photoreceptors. *Chem. Rev.* **103**, 2203–2238 (2003).
53. Ellison, K. S., Dogliotti, E., Connors, T. D., Basu, A. K. & Essigmann, J. M. Site-specific mutagenesis by O6-alkylguanines located in the chromosomes of mammalian cells: influence of the mammalian O6-alkylguanine-DNA alkyltransferase. *Proceedings of the National Academy of Sciences* **86**, 8620–8624 (1989).
54. Sedgwick, B. Repairing DNA-methylation damage. *Nat Rev Mol Cell Biol* **5**, 148–157 (2004).
55. 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**, 1328–1334 (1996).
56. Begley, T. J. & Samson, L. D. AlkB mystery solved: oxidative demethylation of N1-methyladenine and N3-methylcytosine adducts by a direct reversal mechanism. *Trends Biochem Sci* **28**, 2–5 (2003).
57. Yi, C., Yang, C.-G. & He, C. A non-heme iron-mediated chemical demethylation in DNA and RNA. *Acc Chem Res* **42**, 519–529 (2009).
58. Nakamura, J. & Swenberg, J. A. Endogenous apurinic/apyrimidinic sites in genomic DNA of mammalian tissues. *Cancer Res* **59**, 2522–2526 (1999).
59. Krokan, H. E. & Bjørås, M. Base excision repair. *Cold Spring Harb Perspect Biol* **5**, a012583 (2013).
60. 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 (2020).

61. 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**, 924–931 (2015).
62. 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**, 7669–7677 (2013).
63. Maher, R. L. & Bloom, L. B. Pre-steady-state Kinetic Characterization of the AP Endonuclease Activity of Human AP Endonuclease 1 *. *Journal of Biological Chemistry* **282**, 30577–30585 (2007).
64. Beard, W. A. & Wilson, S. H. Structure and Mechanism of DNA Polymerase β. *Biochemistry* **53**, 2768–2780 (2014).
65. Liu, Y. *et al.* DNA Polymerase β and Flap Endonuclease 1 Enzymatic Specificities Sustain DNA Synthesis for Long Patch Base Excision Repair *. *Journal of Biological Chemistry* **280**, 3665–3674 (2005).
66. 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**, 2491–2507 (2011).
67. 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).
68. 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**, 3046–3082 (2008).
69. Cuneo, M. J. & London, R. E. Oxidation state of the XRCC1 N-terminal domain regulates DNA polymerase β binding affinity. *Proceedings of the National Academy of Sciences* **107**, 6805–6810 (2010).
70. 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)**, 249–268 (1999).
71. Caldecott, K. W. DNA single-strand break repair and human genetic disease. *Trends Cell Biol* **32**, 733–745 (2022).
72. 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).
73. Azarm, K. & Smith, S. Nuclear PARPs and genome integrity. *Genes Dev* **34**, 285–301 (2020).
74. Bonfiglio, J. J. *et al.* Serine ADP-Ribosylation Depends on HPF1. *Mol Cell* **65**, 932-940.e6 (2017).

75. 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).
76. Rack, J. G. M., Palazzo, L. & Ahel, I. (ADP-ribosyl)hydrolases: structure, function, and biology. *Genes Dev* **34**, 263–284 (2020).
77. 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).
78. Tulin, A. & Spradling, A. Chromatin loosening by poly(ADP)-ribose polymerase (PARP) at Drosophila puff loci. *Science* **299**, 560–562 (2003).
79. Ahel, D. *et al.* Poly(ADP-ribose)-dependent regulation of DNA repair by the chromatin remodeling enzyme ALC1. *Science* **325**, 1240–1243 (2009).
80. Caldecott, K. W. XRCC1 protein; Form and function. *DNA Repair (Amst)* **81**, 102664 (2019).
81. Bennett, R. A., Wilson, D. M., 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**, 7166–7169 (1997).
82. 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**, 6662–6670 (1996).
83. Demin, A. A. *et al.* XRCC1 prevents toxic PARP1 trapping during DNA base excision repair. *Mol Cell* **81**, 3018-3030.e5 (2021).
84. 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).
85. Amaral-Silva, G. K. do *et al.* Mismatch repair system proteins in oral benign and malignant lesions. *J Oral Pathol Med* **46**, 241–245 (2017).
86. 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).
87. Kusakabe, M. *et al.* Mechanism and regulation of DNA damage recognition in nucleotide excision repair. *Genes and Environment* **41**, 2 (2019).
88. Gillet, L. C. J. & Schärer, O. D. Molecular Mechanisms of Mammalian Global Genome Nucleotide Excision Repair. *Chem. Rev.* **106**, 253–276 (2006).
89. Rapin, I., Lindenbaum, Y., Dickson, D. W., Kraemer, K. H. & Robbins, J. H. Cockayne syndrome and xeroderma pigmentosum. *Neurology* **55**, 1442–1449 (2000).
90. Mu, H., Geacintov, N. E., Broyde, S., Yeo, J.-E. & Schärer, O. D. Molecular basis for damage recognition and verification by XPC-RAD23B and TFIIH in nucleotide excision repair. *DNA Repair* **71**, 33–42 (2018).

91. Paul, D. *et al.* Tethering-facilitated DNA ‘opening’ and complementary roles of β-hairpin motifs in the Rad4/XPC DNA damage sensor protein. *Nucleic Acids Res* **48**, 12348–12364 (2020).
92. Buterin, T., Meyer, C., Giese, B. & Naegeli, H. DNA Quality Control by Conformational Readout on the Undamaged Strand of the Double Helix. *Chemistry & Biology* **12**, 913–922 (2005).
93. Hanawalt, P. C. & Spivak, G. Transcription-coupled DNA repair: two decades of progress and surprises. *Nat Rev Mol Cell Biol* **9**, 958–970 (2008).
94. Nakazawa, Y. *et al.* Ubiquitination of DNA Damage-Stalled RNAPII Promotes Transcription-Coupled Repair. *Cell* **180**, 1228-1244.e24 (2020).
95. 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).
96. 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 Research* **38**, 8083–8094 (2010).
97. Overmeer, R. M. *et al.* Replication protein A safeguards genome integrity by controlling NER incision events. *Journal of Cell Biology* **192**, 401–415 (2011).
98. Kokic, G. *et al.* Structural basis of TFIIH activation for nucleotide excision repair. *Nat Commun* **10**, 2885 (2019).
99. Tsodikov, O. V. *et al.* Structural basis for the recruitment of ERCC1-XPF to nucleotide excision repair complexes by XPA. *EMBO J* **26**, 4768–4776 (2007).
100. Fagbemi, A. F., Orelli, B. & Schärer, O. D. Regulation of endonuclease activity in human nucleotide excision repair. *DNA Repair* **10**, 722–729 (2011).
101. Staresincic, L. *et al.* Coordination of dual incision and repair synthesis in human nucleotide excision repair. *The EMBO Journal* **28**, 1111–1120 (2009).
102. Krasikova, Y., Rechkunova, N. & Lavrik, O. Nucleotide Excision Repair: From Molecular Defects to Neurological Abnormalities. *Int J Mol Sci* **22**, 6220 (2021).
103. 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).
104. 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**, 579–595 (2013).
105. Gottlieb, T. M. & Jackson, S. P. The DNA-dependent protein kinase: requirement for DNA ends and association with Ku antigen. *Cell* **72**, 131–142 (1993).
106. 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).

107. Lee, J.-H. & Paull, T. T. Direct activation of the ATM protein kinase by the Mre11/Rad50/Nbs1 complex. *Science* **304**, 93–96 (2004).

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

109. 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).

110. Ahnesorg, P., Smith, P. & Jackson, S. P. XLF interacts with the XRCC4-DNA ligase IV complex to promote DNA nonhomologous end-joining. *Cell* **124**, 301–313 (2006).

111. 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).

112. Sartori, A. A. *et al.* Human CtIP promotes DNA end resection. *Nature* **450**, 509–514 (2007).

113. 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).

114. Huertas, P. & Jackson, S. P. Human CtIP mediates cell cycle control of DNA end resection and double strand break repair. *J Biol Chem* **284**, 9558–9565 (2009).

115. Blasiak, J. Single-Strand Annealing in Cancer. *Int J Mol Sci* **22**, 2167 (2021).

116. 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**, 13634–13639 (2004).

117. 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).

118. 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**, 6049–6054 (1998).

119. 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).

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

121. 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**, 55117–55126 (2004).

122. 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).

123. Johnson, R. D. & Jasin, M. Sister chromatid gene conversion is a prominent double-strand break repair pathway in mammalian cells. *EMBO J* **19**, 3398–3407 (2000).
124. Elbakry, A. & Löbrich, M. Homologous Recombination Subpathways: A Tangle to Resolve. *Front. Genet.* **12**, (2021).
125. Wu-Baer, F., Lagrazon, K., Yuan, W. & Baer, R. The BRCA1/BARD1 heterodimer assembles polyubiquitin chains through an unconventional linkage involving lysine residue K6 of ubiquitin. *J Biol Chem* **278**, 34743–34746 (2003).
126. Zhang, F. *et al.* PALB2 links BRCA1 and BRCA2 in the DNA-damage response. *Curr Biol* **19**, 524–529 (2009).
127. Pâques, F. & Haber, J. E. Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae. *Microbiol Mol Biol Rev* **63**, 349–404 (1999).
128. Bizard, A. H. & Hickson, I. D. The dissolution of double Holliday junctions. *Cold Spring Harb Perspect Biol* **6**, a016477 (2014).
129. 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).
130. Anand, R. P., Lovett, S. T. & Haber, J. E. Break-induced DNA replication. *Cold Spring Harb Perspect Biol* **5**, a010397 (2013).
131. Dilley, R. L. *et al.* Break-induced telomere synthesis underlies alternative telomere maintenance. *Nature* **539**, 54–58 (2016).
132. Elango, R. *et al.* Break-induced replication promotes formation of lethal joint molecules dissolved by Srs2. *Nat Commun* **8**, 1790 (2017).
133. Nalepa, G. & Clapp, D. W. Fanconi anaemia and cancer: an intricate relationship. *Nat Rev Cancer* **18**, 168–185 (2018).
134. Bogliolo, M. & Surrallés, J. Fanconi anemia: a model disease for studies on human genetics and advanced therapeutics. *Current Opinion in Genetics & Development* **33**, 32–40 (2015).
135. Walden, H. & Deans, A. J. The Fanconi Anemia DNA Repair Pathway: Structural and Functional Insights into a Complex Disorder. *Annual Review of Biophysics* **43**, 257–278 (2014).
136. Rohleder, F. *et al.* FANCM interacts with PCNA to promote replication traverse of DNA interstrand crosslinks. *Nucleic Acids Res* **44**, 3219–3232 (2016).
137. Knipscheer, P. *et al.* The Fanconi anemia pathway promotes replication-dependent DNA interstrand cross-link repair. *Science* **326**, 1698–1701 (2009).
138. Niraj, J., Färkkilä, A. & D'Andrea, A. D. The Fanconi Anemia Pathway in Cancer. *Annual Review of Cancer Biology* **3**, 457–478 (2019).

139. 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**, 174–185 (2014).
140. 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).
141. Ulrich, H. D. Conservation of DNA damage tolerance pathways from yeast to humans. *Biochem Soc Trans* **35**, 1334–1337 (2007).
142. Ashour, M. E. & Mosammaparast, N. Mechanisms of damage tolerance and repair during DNA replication. *Nucleic Acids Res* **49**, 3033–3047 (2021).
143. Guilliam, T. A. Mechanisms for Maintaining Eukaryotic Replisome Progression in the Presence of DNA Damage. *Front. Mol. Biosci.* **8**, (2021).
144. Bianchi, J. *et al.* PrimPol bypasses UV photoproducts during eukaryotic chromosomal DNA replication. *Mol Cell* **52**, 566–573 (2013).
145. Quinet, A. *et al.* PRIMPOL-Mediated Adaptive Response Suppresses Replication Fork Reversal in BRCA-Deficient Cells. *Mol Cell* **77**, 461-474.e9 (2020).
146. 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**, 3317–3329 (2016).
147. Maga, G. *et al.* DNA polymerase δ-interacting protein 2 is a processivity factor for DNA polymerase λ during 8-oxo-7,8-dihydroguanine bypass. *Proceedings of the National Academy of Sciences* **110**, 18850–18855 (2013).
148. 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).
149. McCulloch, S. D. & Kunkel, T. A. The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Res* **18**, 148–161 (2008).
150. 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**, 141–152 (2012).
151. Biertümpfel, C. *et al.* Structure and mechanism of human DNA polymerase eta. *Nature* **465**, 1044–1048 (2010).
152. 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**, 1557–1568 (1997).
153. Tissier, A. *et al.* Misinsertion and bypass of thymine-thymine dimers by human DNA polymerase iota. *EMBO J* **19**, 5259–5266 (2000).

154. Petta, T. B. *et al.* Human DNA polymerase iota protects cells against oxidative stress. *EMBO J* **27**, 2883–2895 (2008).
155. McDonald, J. P. *et al.* Novel human and mouse homologs of Saccharomyces cerevisiae DNA polymerase eta. *Genomics* **60**, 20–30 (1999).
156. 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. *Proceedings of the National Academy of Sciences* **105**, 8679–8684 (2008).
157. 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**, 39678–39684 (2000).
158. 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**, 2219–2222 (2005).
159. 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**, 239–245 (2008).
160. 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* 87–88 (2002) doi:10.1093/nass/2.1.87.
161. Malik, R. *et al.* Structure and mechanism of B-family DNA polymerase ζ specialized for translesion DNA synthesis. *Nat Struct Mol Biol* **27**, 913–924 (2020).
162. Sale, J. E. Translesion DNA synthesis and mutagenesis in eukaryotes. *Cold Spring Harb Perspect Biol* **5**, a012708 (2013).
163. Miyase, S. *et al.* Differential regulation of Rad18 through Rad6-dependent mono- and polyubiquitination. *J Biol Chem* **280**, 515–524 (2005).
164. 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**, 625–636 (2008).
165. Leung, W., Baxley, R. M., Moldovan, G.-L. & Bielinsky, A.-K. Mechanisms of DNA Damage Tolerance: Post-Translational Regulation of PCNA. *Genes (Basel)* **10**, 10 (2018).
166. Boehm, E. M., Spies, M. & Washington, M. T. PCNA tool belts and polymerase bridges form during translesion synthesis. *Nucleic Acids Res* **44**, 8250–8260 (2016).
167. 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).

168. Burkovics, P. *et al.* Srs2 mediates PCNA-SUMO-dependent inhibition of DNA repair synthesis. *EMBO J* **32**, 742–755 (2013).
169. Meister, C., Wong, R. P., Park, Z.-H. & Ulrich, H. D. Reversible association of ubiquitin with PCNA is important for template switching in S. cerevisiae. *DNA Repair (Amst)* **149**, 103842 (2025).
170. 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**, 135–141 (2002).
171. Mohiuddin, M. *et al.* SUMOylation of PCNA by PIAS1 and PIAS4 promotes template switch in the chicken and human B cell lines. *Proc Natl Acad Sci U S A* **115**, 12793–12798 (2018).
172. Giannattasio, M. *et al.* Visualization of recombination-mediated damage bypass by template switching. *Nat Struct Mol Biol* **21**, 884–892 (2014).
173. Neelsen, K. J. & Lopes, M. Replication fork reversal in eukaryotes: from dead end to dynamic response. *Nat Rev Mol Cell Biol* **16**, 207–220 (2015).
174. Berti, M. & Vindigni, A. Replication stress: getting back on track. *Nat Struct Mol Biol* **23**, 103–109 (2016).
175. Blastyák, A., Hajdú, I., Unk, I. & Haracska, L. Role of double-stranded DNA translocase activity of human HLTF in replication of damaged DNA. *Mol Cell Biol* **30**, 684–693 (2010).
176. 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**, 18107–18112 (2006).
177. 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).
178. Ciccia, A. *et al.* Polyubiquitinated PCNA recruits the ZRANB3 translocase to maintain genomic integrity after replication stress. *Mol Cell* **47**, 396–409 (2012).
179. Zellweger, R. *et al.* Rad51-mediated replication fork reversal is a global response to genotoxic treatments in human cells. *J Cell Biol* **208**, 563–579 (2015).
180. Berti, M. *et al.* Human RECQ1 promotes restart of replication forks reversed by DNA topoisomerase I inhibition. *Nat Struct Mol Biol* **20**, 347–354 (2013).
181. Thangavel, S. *et al.* DNA2 drives processing and restart of reversed replication forks in human cells. *J Cell Biol* **208**, 545–562 (2015).
182. O'Brien, P. J., Siraki, A. G. & Shangari, N. Aldehyde sources, metabolism, molecular toxicity mechanisms, and possible effects on human health. *Crit Rev Toxicol* **35**, 609–662 (2005).

183. Behrens, U. J., Hoerner, M., Lasker, J. M. & Lieber, C. S. Formation of acetaldehyde adducts with ethanol-inducible P450IIE1 in vivo. *Biochem Biophys Res Commun* **154**, 584–590 (1988).

184. Shi, Y. *et al.* Histone Demethylation Mediated by the Nuclear Amine Oxidase Homolog LSD1. *Cell* **119**, 941–953 (2004).

185. Tsukada, Y. *et al.* Histone demethylation by a family of JmjC domain-containing proteins. *Nature* **439**, 811–816 (2006).

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

187. Burgos-Barragan, G. *et al.* Mammals divert endogenous genotoxic formaldehyde into one-carbon metabolism. *Nature* **548**, 549–554 (2017).

188. 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**, 3276–3293 (1977).

189. Grafstrom, R. C., Fornace, A. J., Autrup, H., Lechner, J. F. & Harris, C. C. Formaldehyde damage to DNA and inhibition of DNA repair in human bronchial cells. *Science* **220**, 216–218 (1983).

190. Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**, 1497–1502 (2007).

191. Guerin, M. R., Higgins, C. E. & Griest, W. H. The analysis of the particulate and vapour phases of tobacco smoke. *IARC Sci Publ* 115–139 (1987).

192. 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).

193. 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).

194. Groehler, A. *et al.* Oxidative cross-linking of proteins to DNA following ischemia-reperfusion injury. *Free Radic Biol Med* **120**, 89–101 (2018).

195. Ji, S., Shao, H., Han, Q., Seiler, C. L. & Tretyakova, N. Y. Reversible DNA-Protein Cross-Linking at Epigenetic DNA Marks. *Angew Chem Int Ed Engl* **56**, 14130–14134 (2017).

196. 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**, 659–669 (1995).

197. Jung, Y. & Lippard, S. J. Direct Cellular Responses to Platinum-Induced DNA Damage. *Chem. Rev.* **107**, 1387–1407 (2007).
198. Hemminki, K. & Kallama, S. Reactions of nitrogen mustards with DNA. *IARC Sci Publ* 55–70 (1986).
199. Rajski, S. R. & Williams, R. M. DNA Cross-Linking Agents as Antitumor Drugs. *Chem. Rev.* **98**, 2723–2796 (1998).
200. 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).
201. 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**, 571–594 (1977).
202. 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).
203. 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).
204. 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**, 419–429 (1987).
205. Cadet, J., Anselmino, C., Douki, T. & Voituriez, L. Photochemistry of nucleic acids in cells. *J Photochem Photobiol B* **15**, 277–298 (1992).
206. Cadet, J. & Douki, T. Formation of UV-induced DNA damage contributing to skin cancer development. *Photochem Photobiol Sci* **17**, 1816–1841 (2018).
207. Kühbacher, U. & Duxin, J. P. How to fix DNA-protein crosslinks. *DNA Repair (Amst)* **94**, 102924 (2020).
208. Pommier, Y., Sun, Y., Huang, S.-Y. N. & Nitiss, J. L. Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. *Nat Rev Mol Cell Biol* **17**, 703–721 (2016).
209. Champoux, J. J. DNA topoisomerases: structure, function, and mechanism. *Annu Rev Biochem* **70**, 369–413 (2001).
210. Liu, L. F. *et al.* Mechanism of action of camptothecin. *Ann N Y Acad Sci* **922**, 1–10 (2000).
211. 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**, 6782–6786 (2008).

212. 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**, 1339–1346 (1995).
213. 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).
214. 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**, 7792–7796 (1997).
215. Pourquier, P. *et al.* Topoisomerase I-mediated cytotoxicity of N-methyl-N'-nitro-N-nitrosoguanidine: trapping of topoisomerase I by the O6-methylguanine. *Cancer Res* **61**, 53–58 (2001).
216. Nitiss, J. L. DNA topoisomerase II and its growing repertoire of biological functions. *Nat Rev Cancer* **9**, 327–337 (2009).
217. Collin, F., Karkare, S. & Maxwell, A. Exploiting bacterial DNA gyrase as a drug target: current state and perspectives. *Appl Microbiol Biotechnol* **92**, 479–497 (2011).
218. Bernard, P. & Couturier, M. Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. *J Mol Biol* **226**, 735–745 (1992).
219. Jeltsch, A. & Jurkowska, R. Z. New concepts in DNA methylation. *Trends Biochem Sci* **39**, 310–318 (2014).
220. Chatterjee, A. & Eccles, M. R. DNA methylation and epigenomics: new technologies and emerging concepts. *Genome Biology* **16**, 103 (2015).
221. Jurkowska, R. Z., Jurkowski, T. P. & Jeltsch, A. Structure and function of mammalian DNA methyltransferases. *Chembiochem* **12**, 206–222 (2011).
222. Jeltsch, A. & Jurkowska, R. Z. Allosteric control of mammalian DNA methyltransferases - a new regulatory paradigm. *Nucleic Acids Res* **44**, 8556–8575 (2016).
223. Jeltsch, A. Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases. *Chembiochem* **3**, 274–293 (2002).
224. 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).
225. 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**, 223–232 (2009).
226. 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**, 2957–2964 (2002).

227. Quiñones, J. L. *et al.* Enzyme mechanism-based, oxidative DNA–protein cross-links formed with DNA polymerase β in vivo. *Proceedings of the National Academy of Sciences* **112**, 8602–8607 (2015).

228. 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**, 476–490 (1968).

229. 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**, 1773–1778 (1991).

230. 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**, 6145–6153 (2010).

231. 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**, 1027–1032 (1989).

232. Daniels, J. S., Gates, K. S., Tronche, C. & Greenberg, M. M. Direct evidence for bimodal DNA damage induced by tirapazamine. *Chem Res Toxicol* **11**, 1254–1257 (1998).

233. DeMott, M. S. *et al.* Covalent trapping of human DNA polymerase beta by the oxidative DNA lesion 2-deoxyribonolactone. *J Biol Chem* **277**, 7637–7640 (2002).

234. 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).

235. Murai, J. *et al.* Trapping of PARP1 and PARP2 by Clinical PARP Inhibitors. *Cancer Res* **72**, 5588–5599 (2012).

236. Prasad, R. *et al.* Suicidal cross-linking of PARP-1 to AP site intermediates in cells undergoing base excision repair. *Nucleic Acids Res* **42**, 6337–6351 (2014).

237. 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).

238. Bergerat, A. *et al.* An atypical topoisomerase II from Archaea with implications for meiotic recombination. *Nature* **386**, 414–417 (1997).

239. Keeney, S., Giroux, C. N. & Kleckner, N. Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell* **88**, 375–384 (1997).

240. Neale, M. J., Pan, J. & Keeney, S. Endonucleolytic processing of covalent protein-linked DNA double-strand breaks. *Nature* **436**, 1053–1057 (2005).

241. 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).
242. 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**, 2293–2300 (1993).
243. Robinson, A. J., Younghusband, H. B. & Bellett, A. J. D. A circular DNA—Protein complex from adenoviruses. *Virology* **56**, 54–69 (1973).
244. 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**, 5786–5792 (1994).
245. Salas, M. Protein-priming of DNA replication. *Annu Rev Biochem* **60**, 39–71 (1991).
246. Muñoz-Espín, D., Holguera, I., Ballesteros-Plaza, D., Carballido-López, 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**, 16548–16553 (2010).
247. Hoeben, R. C. & Uil, T. G. Adenovirus DNA replication. *Cold Spring Harb Perspect Biol* **5**, a013003 (2013).
248. 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).
249. 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).
250. 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**, 8452–8463 (2015).
251. Mohni, K. N. *et al.* HMCES Maintains Genome Integrity by Shielding Abasic Sites in Single-Strand DNA. *Cell* **176**, 144-153.e13 (2019).
252. Spruijt, C. G. *et al.* Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152**, 1146–1159 (2013).
253. Kweon, S.-M. *et al.* Erasure of Tet-Oxidized 5-Methylcytosine by a SRAP Nuclease. *Cell Rep* **21**, 482–494 (2017).
254. Lee, Y.-J. *et al.* Identification and biosynthesis of thymidine hypermodifications in the genomic DNA of widespread bacterial viruses. *Proc Natl Acad Sci U S A* **115**, E3116–E3125 (2018).
255. Aravind, L., Anand, S. & Iyer, L. M. Novel autoproteolytic and DNA-damage sensing components in the bacterial SOS response and oxidized methylcytosine-induced eukaryotic DNA demethylation systems. *Biol Direct* **8**, 20 (2013).
256. Halabelian, L. *et al.* Structural basis of HMCES interactions with abasic DNA and multivalent substrate recognition. *Nat Struct Mol Biol* **26**, 607–612 (2019).

257. 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).
258. Wang, N. *et al.* Molecular basis of abasic site sensing in single-stranded DNA by the SRAP domain of E. coli yedK. *Nucleic Acids Res* **47**, 10388–10399 (2019).
259. Donsbach, M. *et al.* A non-proteolytic release mechanism for HMCES-DNA-protein crosslinks. *The EMBO Journal* **n/a**, e113360 (2023).
260. Srivastava, M. *et al.* HMCES safeguards replication from oxidative stress and ensures error-free repair. *EMBO Rep* **21**, e49123 (2020).
261. Mehta, K. P. M., Lovejoy, C. A., Zhao, R., Heintzman, D. R. & Cortez, D. HMCES Maintains Replication Fork Progression and Prevents Double-Strand Breaks in Response to APOBEC Deamination and Abasic Site Formation. *Cell Rep* **31**, 107705 (2020).
262. Biayna, J. *et al.* 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 (2021).
263. Wu, L. *et al.* HMCES protects immunoglobulin genes specifically from deletions during somatic hypermutation. *Genes Dev* **36**, 433–450 (2022).
264. Shukla, V. *et al.* HMCES Functions in the Alternative End-Joining Pathway of the DNA DSB Repair during Class Switch Recombination in B Cells. *Mol Cell* **77**, 384-394.e4 (2020).
265. Semlow, D. R., MacKrell, V. A. & Walter, J. C. The HMCES DNA-protein cross-link functions as an intermediate in DNA interstrand cross-link repair. *Nat Struct Mol Biol* **29**, 451–462 (2022).
266. Yaneva, D. *et al.* The FANCJ helicase unfolds DNA-protein crosslinks to promote their repair. *Mol Cell* **83**, 43-56.e10 (2023).
267. 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).
268. Torrecilla, I. *et al.* Isolation and detection of DNA-protein crosslinks in mammalian cells. *Nucleic Acids Res* **52**, 525–547 (2024).
269. Kiianitsa, K. & Maizels, N. Ultrasensitive isolation, identification and quantification of DNA-protein adducts by ELISA-based RADAR assay. *Nucleic Acids Res* **42**, e108 (2014).
270. 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).
271. 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).

272. Zhitkovich, A. & Costa, M. A simple, sensitive assay to detect DNA-protein crosslinks in intact cells and in vivo. *Carcinogenesis* **13**, 1485–1489 (1992).
273. 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).
274. Ostling, O. & Johanson, K. J. Microelectrophoretic study of radiation-induced DNA damages in individual mammalian cells. *Biochem Biophys Res Commun* **123**, 291–298 (1984).
275. Singh, N. P., McCoy, M. T., Tice, R. R. & Schneider, E. L. A simple technique for quantitation of low levels of DNA damage in individual cells. *Exp Cell Res* **175**, 184–191 (1988).
276. Merk, O. & Speit, G. Significance of formaldehyde-induced DNA-protein crosslinks for mutagenesis. *Environ Mol Mutagen* **32**, 260–268 (1998).
277. Merk, O. & Speit, G. Detection of crosslinks with the comet assay in relationship to genotoxicity and cytotoxicity. *Environ Mol Mutagen* **33**, 167–172 (1999).
278. Mórocz, 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).
279. Patel, A. G. *et al.* Immunodetection of human topoisomerase I-DNA covalent complexes. *Nucleic Acids Res* **44**, 2816–2826 (2016).
280. Trendel, J. *et al.* The human proteome with direct physical access to DNA. *Cell* **0**, (2025).
281. 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).
282. Kanellis, P. *et al.* A screen for suppressors of gross chromosomal rearrangements identifies a conserved role for PLP in preventing DNA lesions. *PLoS Genet* **3**, e134 (2007).
283. Balakirev, M. Y. *et al.* Wss1 metalloprotease partners with Cdc48/Doa1 in processing genotoxic SUMO conjugates. *Elife* **4**, e06763 (2015).
284. Flett, F. J. *et al.* Structural basis for DNA 3′-end processing by human tyrosyl-DNA phosphodiesterase 1. *Nat Commun* **9**, 24 (2018).
285. Reinking, H. K., Hofmann, K. & Stingele, J. Function and evolution of the DNA-protein crosslink proteases Wss1 and SPRTN. *DNA Repair* **88**, 102822 (2020).
286. 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).
287. Reinking, H. K. *et al.* DNA Structure-Specific Cleavage of DNA-Protein Crosslinks by the SPRTN Protease. *Mol Cell* **80**, 102-113.e6 (2020).
288. F, L., Je, R., Z, C. & H, Y. Structural Insight into DNA-Dependent Activation of Human Metalloprotease Spartan. *Cell reports* **26**, (2019).

289. 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).

290. 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).

291. Borgermann, N. *et al.* SUMOylation promotes protective responses to DNA-protein crosslinks. *EMBO J* **38**, e101496 (2019).

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

293. Lessel, D. *et al.* Mutations in SPRTN cause early onset hepatocellular carcinoma, genomic instability and progeroid features. *Nat Genet* **46**, 1239–1244 (2014).

294. Stingele, J., Habermann, B. & Jentsch, S. DNA-protein crosslink repair: proteases as DNA repair enzymes. *Trends Biochem Sci* **40**, 67–71 (2015).

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

296. 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).

297. Delabaere, L. *et al.* The Spartan ortholog maternal haploid is required for paternal chromosome integrity in the Drosophila zygote. *Curr Biol* **24**, 2281–2287 (2014).

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

299. Ruijs, M. W. G. *et al.* Atypical progeroid syndrome: an unknown helicase gene defect? *Am J Med Genet A* **116A**, 295–299 (2003).

300. Maskey, R. S. *et al.* Spartan deficiency causes accumulation of Topoisomerase 1 cleavage complexes and tumorigenesis. *Nucleic Acids Res* **45**, 4564–4576 (2017).

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

302. Lopez-Mosqueda, J. *et al.* SPRTN is a mammalian DNA-binding metalloprotease that resolves DNA-protein crosslinks. *eLife* **5**, e21491 (2016).

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

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

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

306. 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).

307. Ruggiano, A. & Ramadan, K. DNA-protein crosslink proteases in genome stability. *Commun Biol* **4**, 11 (2021).

308. Svoboda, M., Konvalinka, J., Trempe, J.-F. & Grantz Saskova, K. The yeast proteases Ddi1 and Wss1 are both involved in the DNA replication stress response. *DNA Repair (Amst)* **80**, 45–51 (2019).

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

310. Yip, M. C. J., Bodnar, N. O. & Rapoport, T. A. Ddi1 is a ubiquitin-dependent protease. *Proc Natl Acad Sci U S A* **117**, 7776–7781 (2020).

311. 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).

312. 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).

313. 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).

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

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

316. Carmell, M. A. *et al.* A widely employed germ cell marker is an ancient disordered protein with reproductive functions in diverse eukaryotes. *Elife* **5**, e19993 (2016).

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

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

319. 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).

320. Kenny, F. M. & Linarelli, L. Dwarfism and cortical thickening of tubular bones. Transient hypocalcemia in a mother and son. *Am J Dis Child* **111**, 201–207 (1966).

321. Ip, W.-H. *et al.* Differential Regulation of Cellular FAM111B by Human Adenovirus C Type 5 E1 Oncogenes. *Viruses* **13**, 1015 (2021).

322. Fine, D. A. *et al.* Identification of FAM111A as an SV40 Host Range Restriction and Adenovirus Helper Factor. *PLOS Pathogens* **8**, e1002949 (2012).

323. 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).

324. Interthal, H. & Champoux, J. J. Effects of DNA and protein size on substrate cleavage by human tyrosyl-DNA phosphodiesterase 1 (TDP1). *Biochem J* **436**, 559–566 (2011).

325. 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).

326. 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**, 21074–21083 (2008).

327. 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**, 14970–14975 (2002).

328. 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**, 267–272 (2002).

329. El-Khamisy, S. F. *et al.* Defective DNA single-strand break repair in spinocerebellar ataxia with axonal neuropathy-1. *Nature* **434**, 108–113 (2005).

330. 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**, 3667–3680 (2009).

331. Das, B. B. *et al.* PARP1-TDP1 coupling for the repair of topoisomerase I-induced DNA damage. *Nucleic Acids Res* **42**, 4435–4449 (2014).

332. 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**, 8363–8375 (2016).

333. 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**, 674–678 (2009).

334. 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**, 17960–17969 (2014).

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

336. Gómez-Herreros, F. *et al.* TDP2 protects transcription from abortive topoisomerase activity and is required for normal neural function. *Nat Genet* **46**, 516–521 (2014).

337. 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).

338. Zeng, Z., Cortés-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**, 403–409 (2011).

339. Jc, C., Es, de L. & Dr, L. Nucleolytic processing of a protein-bound DNA end by the E. coli SbcCD (MR) complex. *DNA repair* **2**, (2003).
340. Hopfner, K. P. *et al.* Structural biochemistry and interaction architecture of the DNA double-strand break repair Mre11 nuclease and Rad50-ATPase. *Cell* **105**, 473–485 (2001).
341. E, C. & P, C. Sae2 promotes dsDNA endonuclease activity within Mre11-Rad50-Xrs2 to resect DNA breaks. *Nature* **514**, (2014).
342. L, K. *et al.* Mechanism of DNA End Sensing and Processing by the Mre11-Rad50 Complex. *Molecular cell* **76**, (2019).
343. Fan, Y. *et al.* Structural basis for DNA double-strand break sensing by human MRE11-RAD50-NBS1 and its TRF2 complex. 2025.03.14.643254 Preprint at https://doi.org/10.1101/2025.03.14.643254 (2025).
344. B, de G., A, C. & Ak, M. Cellular pathways for DNA repair and damage tolerance of formaldehyde-induced DNA-protein crosslinks. *DNA repair* **8**, (2009).
345. 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**, 27065–27076 (2009).
346. Nakano, T. *et al.* Nucleotide excision repair and homologous recombination systems commit differentially to the repair of DNA-protein crosslinks. *Mol Cell* **28**, 147–158 (2007).
347. Baker, D. J. *et al.* Nucleotide excision repair eliminates unique DNA-protein cross-links from mammalian cells. *J Biol Chem* **282**, 22592–22604 (2007).
348. Sparks, J. L. *et al.* The CMG Helicase Bypasses DNA-Protein Cross-Links to Facilitate Their Repair. *Cell* **176**, 167-181.e21 (2019).
349. Gallina, I. *et al.* The ubiquitin ligase RFWD3 is required for translesion DNA synthesis. *Mol Cell* **81**, 442-458.e9 (2021).
350. Inano, S. *et al.* RFWD3-Mediated Ubiquitination Promotes Timely Removal of Both RPA and RAD51 from DNA Damage Sites to Facilitate Homologous Recombination. *Mol Cell* **66**, 622-634.e8 (2017).
351. Carnie, C. J., Jackson, S. P. & Stingele, J. Transcription-coupled repair of DNA–protein crosslinks. *Trends in Cell Biology* **35**, 316–329 (2025).
352. 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).
353. Weickert, P. *et al.* SPRTN patient variants cause global-genome DNA-protein crosslink repair defects. *Nat Commun* **14**, 352 (2023).

354. 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).
355. Reardon, J. T. & Sancar, A. Repair of DNA-polypeptide crosslinks by human excision nuclease. *Proc Natl Acad Sci U S A* **103**, 4056–4061 (2006).
356. Chesner, L. N. & Campbell, C. A quantitative PCR-based assay reveals that nucleotide excision repair plays a predominant role in the removal of DNA-protein crosslinks from plasmids transfected into mammalian cells. *DNA Repair (Amst)* **62**, 18–27 (2018).
357. Carnie, C. J. *et al.* Transcription-coupled repair of DNA-protein cross-links depends on CSA and CSB. *Nat Cell Biol* **26**, 797–810 (2024).
358. van Sluis, M. *et al.* Transcription-coupled DNA-protein crosslink repair by CSB and CRL4CSA-mediated degradation. *Nat Cell Biol* **26**, 770–783 (2024).
359. Oka, Y., Nakazawa, Y., Shimada, M. & Ogi, T. Endogenous aldehyde-induced DNA-protein crosslinks are resolved by transcription-coupled repair. *Nat Cell Biol* **26**, 784–796 (2024).
360. Kokic, G. *et al.* Structural basis for RNA polymerase II ubiquitylation and inactivation in transcription-coupled repair. *Nat Struct Mol Biol* **31**, 536–547 (2024).
361. Carnie, C. J. & Jackson, S. P. The ELOF(1)ant in the room of TCR. *Nat Cell Biol* **23**, 584–586 (2021).
362. van der Weegen, Y. *et al.* ELOF1 is a transcription-coupled DNA repair factor that directs RNA polymerase II ubiquitylation. *Nat Cell Biol* **23**, 595–607 (2021).
363. Epanchintsev, A. *et al.* Cockayne's Syndrome A and B Proteins Regulate Transcription Arrest after Genotoxic Stress by Promoting ATF3 Degradation. *Mol Cell* **68**, 1054-1066.e6 (2017).
364. 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).
365. 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).
366. 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).
367. Perry, M. *et al.* USP11 mediates repair of DNA-protein cross-links by deubiquitinating SPRTN metalloprotease. *J Biol Chem* **296**, 100396 (2021).
368. Ruggiano, A. *et al.* The protease SPRTN and SUMOylation coordinate DNA-protein crosslink repair to prevent genome instability. *Cell Rep* **37**, 110080 (2021).
369. Vijay-Kumar, S., Bugg, C. E. & Cook, W. J. Structure of ubiquitin refined at 1.8 A resolution. *J Mol Biol* **194**, 531–544 (1987).

370. Burroughs, A. M., Iyer, L. M. & Aravind, L. The natural history of ubiquitin and ubiquitin-related domains. *Front Biosci (Landmark Ed)* **17**, 1433–1460 (2012).

371. Overington, J. P. Comparison of three-dimensional structures of homologous proteins. *Current Opinion in Structural Biology* **2**, 394–401 (1992).

372. Hershko, A. & Ciechanover, A. The ubiquitin system. *Annu Rev Biochem* **67**, 425–479 (1998).

373. Sahu, I., Zhu, H., Buhrlage, S. J. & Marto, J. A. Proteomic approaches to study ubiquitinomics. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1866**, 194940 (2023).

374. Gonzalez-Santamarta, M., Bouvier, C., Rodriguez, M. S. & Xolalpa, W. Ubiquitin-chains dynamics and its role regulating crucial cellular processes. *Seminars in Cell & Developmental Biology* **132**, 155–170 (2022).

375. Akutsu, M., Dikic, I. & Bremm, A. Ubiquitin chain diversity at a glance. *Journal of Cell Science* **129**, 875–880 (2016).

376. Chen, R.-H., Chen, Y.-H. & Huang, T.-Y. Ubiquitin-mediated regulation of autophagy. *J Biomed Sci* **26**, 80 (2019).

377. Mendes, M. L., Fougeras, M. R. & Dittmar, G. Analysis of ubiquitin signaling and chain topology cross-talk. *Journal of Proteomics* **215**, 103634 (2020).

378. Pérez Berrocal, D. A., Witting, K. F., Ovaa, H. & Mulder, M. P. C. Hybrid Chains: A Collaboration of Ubiquitin and Ubiquitin-Like Modifiers Introducing Cross-Functionality to the Ubiquitin Code. *Front. Chem.* **7**, (2020).

379. Haglund, K. & Dikic, I. Ubiquitylation and cell signaling. *The EMBO Journal* **24**, 3353–3359 (2005).

380. Swaney, D. L. *et al.* Global analysis of phosphorylation and ubiquitylation cross-talk in protein degradation. *Nat Methods* **10**, 676–682 (2013).

381. Zhou, H. *et al.* Toward a comprehensive characterization of a human cancer cell phosphoproteome. *J Proteome Res* **12**, 260–271 (2013).

382. Choudhary, C. *et al.* Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* **325**, 834–840 (2009).

383. Song, L. & Luo, Z.-Q. Post-translational regulation of ubiquitin signaling. *J Cell Biol* **218**, 1776–1786 (2019).

384. Galisson, F. *et al.* A novel proteomics approach to identify SUMOylated proteins and their modification sites in human cells. *Mol Cell Proteomics* **10**, M110.004796 (2011).

385. Swatek, K. N. & Komander, D. Ubiquitin modifications. *Cell Res* **26**, 399–422 (2016).

386. Hochstrasser, M. Origin and function of ubiquitin-like proteins. *Nature* **458**, 422–429 (2009).

387. Schulman, B. A. & Wade Harper, J. Ubiquitin-like protein activation by E1 enzymes: the apex for downstream signalling pathways. *Nat Rev Mol Cell Biol* **10**, 319–331 (2009).
388. Jentsch, S. & Pyrowolakis, G. Ubiquitin and its kin: how close are the family ties? *Trends in Cell Biology* **10**, 335–342 (2000).
389. Cappadocia, L. & Lima, C. D. Ubiquitin-like Protein Conjugation: Structures, Chemistry, and Mechanism. *Chem. Rev.* **118**, 889–918 (2018).
390. van der Veen, A. G. & Ploegh, H. L. Ubiquitin-like proteins. *Annu Rev Biochem* **81**, 323–357 (2012).
391. Lüders, J., Pyrowolakis, G. & Jentsch, S. The ubiquitin-like protein HUB1 forms SDS-resistant complexes with cellular proteins in the absence of ATP. *EMBO reports* **4**, 1169–1174 (2003).
392. Yashiroda, H. & Tanaka, K. Hub1 is an essential ubiquitin-like protein without functioning as a typical modifier in fission yeast. *Genes Cells* **9**, 1189–1197 (2004).
393. Rossman, T. G., Visalli, M. A. & Komissarova, E. V. fau and its ubiquitin-like domain (FUBI) transforms human osteogenic sarcoma (HOS) cells to anchorage-independence. *Oncogene* **22**, 1817–1821 (2003).
394. Deol, K. K., Lorenz, S. & Strieter, E. R. Enzymatic Logic of Ubiquitin Chain Assembly. *Front. Physiol.* **10**, (2019).
395. LaPlante, G. & Zhang, W. Targeting the Ubiquitin-Proteasome System for Cancer Therapeutics by Small-Molecule Inhibitors. *Cancers (Basel)* **13**, 3079 (2021).
396. Groettrup, M., Pelzer, C., Schmidtke, G. & Hofmann, K. Activating the ubiquitin family: UBA6 challenges the field. *Trends Biochem Sci* **33**, 230–237 (2008).
397. Wu, P.-Y. *et al.* A conserved catalytic residue in the ubiquitin-conjugating enzyme family. *EMBO J* **22**, 5241–5250 (2003).
398. Deshaies, R. J. & Joazeiro, C. A. P. RING domain E3 ubiquitin ligases. *Annu Rev Biochem* **78**, 399–434 (2009).
399. Morreale, F. E. & Walden, H. Types of Ubiquitin Ligases. *Cell* **165**, 248-248.e1 (2016).
400. Weber, J., Polo, S. & Maspero, E. HECT E3 Ligases: A Tale With Multiple Facets. *Front Physiol* **10**, 370 (2019).
401. Walden, H. & Rittinger, K. RBR ligase-mediated ubiquitin transfer: a tale with many twists and turns. *Nat Struct Mol Biol* **25**, 440–445 (2018).
402. McDowell, G. S. & Philpott, A. Non-canonical ubiquitylation: Mechanisms and consequences. *The International Journal of Biochemistry & Cell Biology* **45**, 1833–1842 (2013).
403. Hanpude, P., Bhattacharya, S., Dey, A. K. & Maiti, T. K. Deubiquitinating enzymes in cellular signaling and disease regulation. *IUBMB Life* **67**, 544–555 (2015).

404. Mulder, M. P. C., Witting, K. F. & Ovaa, H. Cracking the Ubiquitin Code: The Ubiquitin Toolbox. *Curr Issues Mol Biol* **37**, 1–20 (2020).

405. Clague, M. J., Urbé, S. & Komander, D. Breaking the chains: deubiquitylating enzyme specificity begets function. *Nat Rev Mol Cell Biol* **20**, 338–352 (2019).

406. Hicke, L. & Dunn, R. Regulation of membrane protein transport by ubiquitin and ubiquitin-binding proteins. *Annu Rev Cell Dev Biol* **19**, 141–172 (2003).

407. Haglund, K., Di Fiore, P. P. & Dikic, I. Distinct monoubiquitin signals in receptor endocytosis. *Trends Biochem Sci* **28**, 598–603 (2003).

408. Hoeller, D. *et al.* Regulation of ubiquitin-binding proteins by monoubiquitination. *Nat Cell Biol* **8**, 163–169 (2006).

409. Wang, X., Jin, C., Tang, Y., Tang, L.-Y. & Zhang, Y. E. Ubiquitination of tumor necrosis factor receptor-associated factor 4 (TRAF4) by Smad ubiquitination regulatory factor 1 (Smurf1) regulates motility of breast epithelial and cancer cells. *J Biol Chem* **288**, 21784–21792 (2013).

410. Su, Y.-T. *et al.* Monoubiquitination of Filamin B Regulates Vascular Endothelial Growth Factor-Mediated Trafficking of Histone Deacetylase 7. *Mol Cell Biol* **33**, 1546–1560 (2013).

411. Sasaki, A. T. *et al.* Ubiquitination of K-Ras enhances activation and facilitates binding to select downstream effectors. *Sci Signal* **4**, ra13 (2011).

412. Pavri, R. *et al.* Histone H2B monoubiquitination functions cooperatively with FACT to regulate elongation by RNA polymerase II. *Cell* **125**, 703–717 (2006).

413. Pickart, C. M. & Fushman, D. Polyubiquitin chains: polymeric protein signals. *Curr Opin Chem Biol* **8**, 610–616 (2004).

414. Komander, D. & Rape, M. The ubiquitin code. *Annu Rev Biochem* **81**, 203–229 (2012).

415. Hicke, L., Schubert, H. L. & Hill, C. P. Ubiquitin-binding domains. *Nat Rev Mol Cell Biol* **6**, 610–621 (2005).

416. Trempe, J. F., Brown, N. R., Noble, M. E. M. & Endicott, J. A. A new crystal form of Lys48-linked diubiquitin. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **66**, 994–998 (2010).

417. Cook, W. J., Jeffrey, L. C., Carson, M., Chen, Z. & Pickart, C. M. Structure of a diubiquitin conjugate and a model for interaction with ubiquitin conjugating enzyme (E2). *Journal of Biological Chemistry* **267**, 16467–16471 (1992).

418. Lu, X., Ebelle, D. L., Matsuo, H. & Walters, K. J. An Extended Conformation for K48 Ubiquitin Chains Revealed by the hRpn2:Rpn13:K48-Diubiquitin Structure. *Structure* **28**, 495-506.e3 (2020).

419. Trempe, J.-F. Reading the ubiquitin postal code. *Curr Opin Struct Biol* **21**, 792–801 (2011).

420. Varadan, R., Walker, O., Pickart, C. & Fushman, D. Structural Properties of Polyubiquitin Chains in Solution. *Journal of Molecular Biology* **324**, 637–647 (2002).
421. Sato, Y. *et al.* Structural basis for specific recognition of Lys 63-linked polyubiquitin chains by tandem UIMs of RAP80. *EMBO J* **28**, 2461–2468 (2009).
422. Tang, Y. *et al.* K63-linked ubiquitination regulates RIPK1 kinase activity to prevent cell death during embryogenesis and inflammation. *Nat Commun* **10**, 4157 (2019).
423. Laplantine, E. *et al.* NEMO specifically recognizes K63-linked poly-ubiquitin chains through a new bipartite ubiquitin-binding domain. *The EMBO Journal* **28**, 2885–2895 (2009).
424. Gallo, L. H. *et al.* Novel Lys63-linked ubiquitination of IKKβ induces STAT3 signaling. *Cell Cycle* **13**, 3964–3976 (2014).
425. Yan, K. *et al.* The role of K63-linked polyubiquitination in cardiac hypertrophy. *Journal of Cellular and Molecular Medicine* **22**, 4558–4567 (2018).
426. Varadan, R. *et al.* Solution Conformation of Lys63-linked Di-ubiquitin Chain Provides Clues to Functional Diversity of Polyubiquitin Signaling*. *Journal of Biological Chemistry* **279**, 7055–7063 (2004).
427. Weeks, S. D., Grasty, K. C., Hernandez-Cuebas, L. & Loll, P. J. Crystal structures of Lys-63-linked tri- and di-ubiquitin reveal a highly extended chain architecture. *Proteins: Structure, Function, and Bioinformatics* **77**, 753–759 (2009).
428. Komander, D. *et al.* Molecular discrimination of structurally equivalent Lys 63-linked and linear polyubiquitin chains. *EMBO reports* **10**, 466–473 (2009).
429. Chen, Z. J. & Sun, L. J. Nonproteolytic Functions of Ubiquitin in Cell Signaling. *Molecular Cell* **33**, 275–286 (2009).
430. Dittmar, G. & Winklhofer, K. F. Linear Ubiquitin Chains: Cellular Functions and Strategies for Detection and Quantification. *Front. Chem.* **7**, (2020).
431. Ikeda, F. *et al.* SHARPIN forms a linear ubiquitin ligase complex regulating NF-κB activity and apoptosis. *Nature* **471**, 637–641 (2011).
432. Gerlach, B. *et al.* Linear ubiquitination prevents inflammation and regulates immune signalling. *Nature* **471**, 591–596 (2011).
433. Tokunaga, F. Linear ubiquitination-mediated NF-κB regulation and its related disorders. *The Journal of Biochemistry* **154**, 313–323 (2013).
434. Matsumoto, M. L. *et al.* K11-Linked Polyubiquitination in Cell Cycle Control Revealed by a K11 Linkage-Specific Antibody. *Molecular Cell* **39**, 477–484 (2010).
435. Jin, L., Williamson, A., Banerjee, S., Philipp, I. & Rape, M. Mechanism of Ubiquitin-Chain Formation by the Human Anaphase-Promoting Complex. *Cell* **133**, 653–665 (2008).
436. Liao, Y., Sumara, I. & Pangou, E. Non-proteolytic ubiquitylation in cellular signaling and human disease. *Commun Biol* **5**, 1–15 (2022).

437. van Huizen, M. & Kikkert, M. The Role of Atypical Ubiquitin Chains in the Regulation of the Antiviral Innate Immune Response. *Front. Cell Dev. Biol.* **7**, (2020).

438. Samant, R. S., Livingston, C. M., Sontag, E. M. & Frydman, J. Distinct proteostasis circuits cooperate in nuclear and cytoplasmic protein quality control. *Nature* **563**, 407–411 (2018).

439. Grice, G. L. *et al.* The Proteasome Distinguishes between Heterotypic and Homotypic Lysine-11-Linked Polyubiquitin Chains. *Cell Reports* **12**, 545–553 (2015).

440. Sobhian, B. *et al.* RAP80 targets BRCA1 to specific ubiquitin structures at DNA damage sites. *Science* **316**, 1198–1202 (2007).

441. Yu, Y. *et al.* K29-Linked Ubiquitin Signaling Regulates Proteotoxic Stress Response and Cell Cycle. *Nat Chem Biol* **17**, 896–905 (2021).

442. Michel, M. A., Swatek, K. N., Hospenthal, M. K. & Komander, D. Ubiquitin Linkage-Specific Affimers Reveal Insights into K6-Linked Ubiquitin Signaling. *Molecular Cell* **68**, 233-246.e5 (2017).

443. Zhao, S. *et al.* RNF14-dependent atypical ubiquitylation promotes translation-coupled resolution of RNA-protein crosslinks. *Mol Cell* **83**, 4290-4303.e9 (2023).

444. Rahmanto, A. S. *et al.* K6-linked ubiquitylation marks formaldehyde-induced RNA-protein crosslinks for resolution. *Molecular Cell* **83**, 4272-4289.e10 (2023).

445. Virdee, S., Ye, Y., Nguyen, D. P., Komander, D. & Chin, J. W. Engineered diubiquitin synthesis reveals Lys29-isopeptide specificity of an OTU deubiquitinase. *Nat Chem Biol* **6**, 750–757 (2010).

446. Swatek, K. N. *et al.* Insights into ubiquitin chain architecture using Ub-clipping. *Nature* **572**, 533–537 (2019).

447. Yau, R. G. *et al.* Assembly and Function of Heterotypic Ubiquitin Chains in Cell-Cycle and Protein Quality Control. *Cell* **171**, 918-933.e20 (2017).

448. Liu, C., Liu, W., Ye, Y. & Li, W. Ufd2p synthesizes branched ubiquitin chains to promote the degradation of substrates modified with atypical chains. *Nat Commun* **8**, 14274 (2017).

449. Emmerich, C. H. *et al.* Activation of the canonical IKK complex by K63/M1-linked hybrid ubiquitin chains. *Proc Natl Acad Sci U S A* **110**, 15247–15252 (2013).

450. Pluska, L. *et al.* The UBA domain of conjugating enzyme Ubc1/Ube2K facilitates assembly of K48/K63-branched ubiquitin chains. *EMBO J* **40**, e106094 (2021).

451. Ohtake, F., Saeki, Y., Ishido, S., Kanno, J. & Tanaka, K. The K48-K63 Branched Ubiquitin Chain Regulates NF-κB Signaling. *Mol Cell* **64**, 251–266 (2016).

452. Kirkpatrick, D. S. *et al.* Quantitative analysis of in vitro ubiquitinated cyclin B1 reveals complex chain topology. *Nat Cell Biol* **8**, 700–710 (2006).

453. Yao, J., Liang, X., Liu, Y. & Zheng, M. Neddylation: A Versatile Pathway Takes on Chronic Liver Diseases. *Front. Med.* **7**, (2020).
454. Han, Z.-J., Feng, Y.-H., Gu, B.-H., Li, Y.-M. & Chen, H. The post-translational modification, SUMOylation, and cancer (Review). *International Journal of Oncology* **52**, 1081–1094 (2018).
455. Dikic, I., Wakatsuki, S. & Walters, K. J. Ubiquitin-binding domains - from structures to functions. *Nat Rev Mol Cell Biol* **10**, 659–671 (2009).
456. Shih, S. C., Sloper-Mould, K. E. & Hicke, L. Monoubiquitin carries a novel internalization signal that is appended to activated receptors. *EMBO J* **19**, 187–198 (2000).
457. Kamadurai, H. B. *et al.* Insights into ubiquitin transfer cascades from a structure of a UbcH5B approximately ubiquitin-HECT(NEDD4L) complex. *Mol Cell* **36**, 1095–1102 (2009).
458. Hu, M. *et al.* Crystal structure of a UBP-family deubiquitinating enzyme in isolation and in complex with ubiquitin aldehyde. *Cell* **111**, 1041–1054 (2002).
459. Reyes-Turcu, F. E. *et al.* The ubiquitin binding domain ZnF UBP recognizes the C-terminal diglycine motif of unanchored ubiquitin. *Cell* **124**, 1197–1208 (2006).
460. Sloper-Mould, K. E., Jemc, J. C., Pickart, C. M. & Hicke, L. Distinct functional surface regions on ubiquitin. *J Biol Chem* **276**, 30483–30489 (2001).
461. Ye, Y. *et al.* Polyubiquitin binding and cross-reactivity in the USP domain deubiquitinase USP21. *EMBO Rep* **12**, 350–357 (2011).
462. Randles, L. & Walters, K. J. Ubiquitin and its binding domains. *Front Biosci (Landmark Ed)* **17**, 2140–2157 (2012).
463. Hofmann, K. & Falquet, L. A ubiquitin-interacting motif conserved in components of the proteasomal and lysosomal protein degradation systems. *Trends Biochem Sci* **26**, 347–350 (2001).
464. Polo, S., Confalonieri, S., Salcini, A. E. & Di Fiore, P. P. EH and UIM: endocytosis and more. *Sci STKE* **2003**, re17 (2003).
465. Pinato, S., Gatti, M., Scandiuzzi, C., Confalonieri, S. & Penengo, L. UMI, a novel RNF168 ubiquitin binding domain involved in the DNA damage signaling pathway. *Mol Cell Biol* **31**, 118–126 (2011).
466. Hirano, S. *et al.* Double-sided ubiquitin binding of Hrs-UIM in endosomal protein sorting. *Nat Struct Mol Biol* **13**, 272–277 (2006).
467. Raasi, S., Orlov, I., Fleming, K. G. & Pickart, C. M. Binding of polyubiquitin chains to ubiquitin-associated (UBA) domains of HHR23A. *J Mol Biol* **341**, 1367–1379 (2004).
468. Raasi, S., Varadan, R., Fushman, D. & Pickart, C. M. Diverse polyubiquitin interaction properties of ubiquitin-associated domains. *Nat Struct Mol Biol* **12**, 708–714 (2005).

469. Shih, S. C. *et al.* A ubiquitin-binding motif required for intramolecular monoubiquitylation, the CUE domain. *EMBO J* **22**, 1273–1281 (2003).

470. Donaldson, K. M., Yin, H., Gekakis, N., Supek, F. & Joazeiro, C. A. P. Ubiquitin signals protein trafficking via interaction with a novel ubiquitin binding domain in the membrane fusion regulator, Vps9p. *Curr Biol* **13**, 258–262 (2003).

471. Prag, G. *et al.* Structural mechanism for ubiquitinated-cargo recognition by the Golgi-localized, gamma-ear-containing, ADP-ribosylation-factor-binding proteins. *Proc Natl Acad Sci U S A* **102**, 2334–2339 (2005).

472. Akutsu, M. *et al.* Structural basis for recognition of ubiquitinated cargo by Tom1-GAT domain. *FEBS Lett* **579**, 5385–5391 (2005).

473. Zhang, N. *et al.* Structure of the s5a:k48-linked diubiquitin complex and its interactions with rpn13. *Mol Cell* **35**, 280–290 (2009).

474. Wang, Q., Young, P. & Walters, K. J. Structure of S5a bound to monoubiquitin provides a model for polyubiquitin recognition. *J Mol Biol* **348**, 727–739 (2005).

475. Sims, J. J. & Cohen, R. E. Linkage-specific avidity defines the lysine 63-linked polyubiquitin-binding preference of rap80. *Mol Cell* **33**, 775–783 (2009).

476. Hofmann, K. Ubiquitin-binding domains and their role in the DNA damage response. *DNA Repair (Amst)* **8**, 544–556 (2009).

477. Yoshikawa, A. *et al.* Crystal structure of the NEMO ubiquitin-binding domain in complex with Lys 63-linked di-ubiquitin. *FEBS Lett* **583**, 3317–3322 (2009).

478. Kulathu, Y., Akutsu, M., Bremm, A., Hofmann, K. & Komander, D. Two-sided ubiquitin binding explains specificity of the TAB2 NZF domain. *Nat Struct Mol Biol* **16**, 1328–1330 (2009).

479. Alam, S. L. *et al.* Ubiquitin interactions of NZF zinc fingers. *EMBO J* **23**, 1411–1421 (2004).

480. Wang, B. *et al.* Structure and ubiquitin interactions of the conserved zinc finger domain of Npl4. *J Biol Chem* **278**, 20225–20234 (2003).

481. Lee, S. *et al.* Structural basis for ubiquitin recognition and autoubiquitination by Rabex-5. *Nat Struct Mol Biol* **13**, 264–271 (2006).

482. Penengo, L. *et al.* Crystal structure of the ubiquitin binding domains of rabex-5 reveals two modes of interaction with ubiquitin. *Cell* **124**, 1183–1195 (2006).

483. Slagsvold, T. *et al.* Eap45 in mammalian ESCRT-II binds ubiquitin via a phosphoinositide-interacting GLUE domain. *J Biol Chem* **280**, 19600–19606 (2005).

484. Alam, S. L. *et al.* Structural basis for ubiquitin recognition by the human ESCRT-II EAP45 GLUE domain. *Nat Struct Mol Biol* **13**, 1029–1030 (2006).

485. Hirano, S. *et al.* Structural basis of ubiquitin recognition by mammalian Eap45 GLUE domain. *Nat Struct Mol Biol* **13**, 1031–1032 (2006).

486. Schreiner, P. *et al.* Ubiquitin docking at the proteasome through a novel pleckstrin-homology domain interaction. *Nature* **453**, 548–552 (2008).
487. Husnjak, K. *et al.* Proteasome subunit Rpn13 is a novel ubiquitin receptor. *Nature* **453**, 481–488 (2008).
488. Pickart, C. M. Mechanisms underlying ubiquitination. *Annu Rev Biochem* **70**, 503–533 (2001).
489. Brzovic, P. S., Lissounov, A., Christensen, D. E., Hoyt, D. W. & Klevit, R. E. A UbcH5/ubiquitin noncovalent complex is required for processive BRCA1-directed ubiquitination. *Mol Cell* **21**, 873–880 (2006).
490. Sundquist, W. I. *et al.* Ubiquitin recognition by the human TSG101 protein. *Mol Cell* **13**, 783–789 (2004).
491. Mayer, B. J. SH3 domains: complexity in moderation. *J Cell Sci* **114**, 1253–1263 (2001).
492. He, Y., Hicke, L. & Radhakrishnan, I. Structural basis for ubiquitin recognition by SH3 domains. *J Mol Biol* **373**, 190–196 (2007).
493. Stamenova, S. D. *et al.* Ubiquitin binds to and regulates a subset of SH3 domains. *Mol Cell* **25**, 273–284 (2007).
494. Bezsonova, I. *et al.* Interactions between the three CIN85 SH3 domains and ubiquitin: implications for CIN85 ubiquitination. *Biochemistry* **47**, 8937–8949 (2008).
495. Pashkova, N. *et al.* WD40 repeat propellers define a ubiquitin-binding domain that regulates turnover of F box proteins. *Mol Cell* **40**, 433–443 (2010).
496. Li, J., Bhat, A. & Xiao, W. Regulation of nucleotide excision repair through ubiquitination. *ABBS* **43**, 919–929 (2011).
497. Rechkunova, N. I., Maltseva, E. A. & Lavrik, O. I. Post-translational Modifications of Nucleotide Excision Repair Proteins and Their Role in the DNA Repair. *Biochemistry (Mosc)* **84**, 1008–1020 (2019).
498. Sugasawa, K. *et al.* UV-Induced Ubiquitylation of XPC Protein Mediated by UV-DDB-Ubiquitin Ligase Complex. *Cell* **121**, 387–400 (2005).
499. Matsumoto, S. *et al.* Functional regulation of the DNA damage-recognition factor DDB2 by ubiquitination and interaction with xeroderma pigmentosum group C protein. *Nucleic Acids Research* **43**, 1700–1713 (2015).
500. Puumalainen, M.-R. *et al.* Chromatin retention of DNA damage sensors DDB2 and XPC through loss of p97 segregase causes genotoxicity. *Nat Commun* **5**, 3695 (2014).
501. Chitale, S. & Richly, H. Timing of DNA lesion recognition: Ubiquitin signaling in the NER pathway. *Cell Cycle* **16**, 163–171 (2017).
502. Van Cuijk, L. *et al.* SUMO and ubiquitin-dependent XPC exchange drives nucleotide excision repair. *Nature Communications* **6**, (2015).

503. Poulsen, S. L. *et al.* RNF111/Arkadia is a SUMO-targeted ubiquitin ligase that facilitates the DNA damage response. *Journal of Cell Biology* **201**, 797–807 (2013).

504. Brückner, F., Hennecke, U., Carell, T. & Cramer, P. CPD damage recognition by transcribing RNA polymerase II. *Science* **315**, 859–862 (2007).

505. Groisman, R. *et al.* The Ubiquitin Ligase Activity in the DDB2 and CSA Complexes Is Differentially Regulated by the COP9 Signalosome in Response to DNA Damage. *Cell* **113**, 357–367 (2003).

506. Groisman, R. *et al.* CSA-dependent degradation of CSB by the ubiquitin–proteasome pathway establishes a link between complementation factors of the Cockayne syndrome. *Genes Dev.* **20**, 1429–1434 (2006).

507. Wilson, M. D., Harreman, M. & Svejstrup, J. Q. Ubiquitylation and degradation of elongating RNA polymerase II: The last resort. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1829**, 151–157 (2013).

508. Tufegdžić Vidaković, A. *et al.* Regulation of the RNAPII Pool Is Integral to the DNA Damage Response. *Cell* **180**, 1245-1261.e21 (2020).

509. Aquila, L. & Atanassov, B. S. Regulation of Histone Ubiquitination in Response to DNA Double Strand Breaks. *Cells* **9**, 1699 (2020).

510. Kolas, N. K. *et al.* Orchestration of the DNA-damage response by the RNF8 ubiquitin ligase. *Science* **318**, 1637–1640 (2007).

511. Stucki, M. *et al.* MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. *Cell* **123**, 1213–1226 (2005).

512. Mailand, N. *et al.* RNF8 ubiquitylates histones at DNA double-strand breaks and promotes assembly of repair proteins. *Cell* **131**, 887–900 (2007).

513. Thorslund, T. *et al.* Histone H1 couples initiation and amplification of ubiquitin signalling after DNA damage. *Nature* **527**, 389–393 (2015).

514. Nowsheen, S. *et al.* L3MBTL2 orchestrates ubiquitin signalling by dictating the sequential recruitment of RNF8 and RNF168 after DNA damage. *Nat Cell Biol* **20**, 455–464 (2018).

515. Mattiroli, F. *et al.* RNF168 ubiquitinates K13-15 on H2A/H2AX to drive DNA damage signaling. *Cell* **150**, 1182–1195 (2012).

516. Bohgaki, M. *et al.* RNF168 ubiquitylates 53BP1 and controls its response to DNA double-strand breaks. *Proc Natl Acad Sci U S A* **110**, 20982–20987 (2013).

517. Botuyan, M. V. *et al.* Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair. *Cell* **127**, 1361–1373 (2006).

518. Fradet-Turcotte, A. *et al.* 53BP1 is a reader of the DNA-damage-induced H2A Lys 15 ubiquitin mark. *Nature* **499**, 50–54 (2013).

519. Lafranchi, L. *et al.* APC/CCdh1 controls CtIP stability during the cell cycle and in response to DNA damage. *The EMBO Journal* **33**, 2860–2879 (2014).

520. Hu, Y. *et al.* RAP80-directed tuning of BRCA1 homologous recombination function at ionizing radiation-induced nuclear foci. *Genes Dev* **25**, 685–700 (2011).

521. Becker, J. R. *et al.* BARD1 reads H2A lysine 15 ubiquitination to direct homologous recombination. *Nature* **596**, 433–437 (2021).

522. Sherker, A. *et al.* Two redundant ubiquitin-dependent pathways of BRCA1 localization to DNA damage sites. *EMBO reports* **22**, e53679 (2021).

523. Isono, M. *et al.* BRCA1 Directs the Repair Pathway to Homologous Recombination by Promoting 53BP1 Dephosphorylation. *Cell Rep* **18**, 520–532 (2017).

524. Yu, X., Fu, S., Lai, M., Baer, R. & Chen, J. BRCA1 ubiquitinates its phosphorylation-dependent binding partner CtIP. *Genes Dev* **20**, 1721–1726 (2006).

525. Densham, R. M. *et al.* Human BRCA1-BARD1 ubiquitin ligase activity counteracts chromatin barriers to DNA resection. *Nature Structural and Molecular Biology* **23**, 647–655 (2016).

526. Kalb, R., Mallery, D. L., Larkin, C., Huang, J. T. J. & Hiom, K. BRCA1 Is a Histone-H2A-Specific Ubiquitin Ligase. *Cell Rep* **8**, 999–1005 (2014).

527. Tong, Z.-B., Ai, H.-S. & Li, J.-B. The Mechanism of Chromatin Remodeler SMARCAD1/Fun30 in Response to DNA Damage. *Front. Cell Dev. Biol.* **8**, (2020).

528. Machida, Y. J. *et al.* UBE2T Is the E2 in the Fanconi Anemia Pathway and Undergoes Negative Autoregulation. *Molecular Cell* **23**, 589–596 (2006).

529. Meetei, A. R. *et al.* A novel ubiquitin ligase is deficient in Fanconi anemia. *Nat Genet* **35**, 165–170 (2003).

530. Shimamura, A. *et al.* A novel diagnostic screen for defects in the Fanconi anemia pathway. *Blood* **100**, 4649–4654 (2002).

531. Tan, W. & Deans, A. J. The ubiquitination machinery of the Fanconi Anemia DNA repair pathway. *Progress in Biophysics and Molecular Biology* **163**, 5–13 (2021).

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

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

534. Rennie, M. L. *et al.* Differential functions of FANCI and FANCD2 ubiquitination stabilize ID2 complex on DNA. *EMBO reports* **21**, e50133 (2020).

535. Tan, W. *et al.* Monoubiquitination by the human Fanconi anemia core complex clamps FANCI:FANCD2 on DNA in filamentous arrays. *eLife* **9**, e54128.

536. Lemonidis, K. *et al.* Structural and biochemical basis of interdependent FANCI-FANCD2 ubiquitination. *The EMBO Journal* **42**, e111898 (2023).

537. Andersen, P. L., Xu, F. & Xiao, W. Eukaryotic DNA damage tolerance and translesion synthesis through covalent modifications of PCNA. *Cell Res* **18**, 162–173 (2008).

538. Hedglin, M., Aitha, M., Pedley, A. & Benkovic, S. J. Replication protein A dynamically regulates monoubiquitination of proliferating cell nuclear antigen. *J Biol Chem* **294**, 5157–5168 (2019).

539. Simpson, L. J. *et al.* RAD18-independent ubiquitination of proliferating-cell nuclear antigen in the avian cell line DT40. *EMBO Rep* **7**, 927–932 (2006).

540. Terai, K., Abbas, T., Jazaeri, A. A. & Dutta, A. CRL4Cdt2 E3 ubiquitin ligase monoubiquitinates PCNA to promote translesion DNA synthesis. *Mol Cell* **37**, 143 (2010).

541. Zhang, S. *et al.* PCNA is ubiquitinated by RNF8. *Cell Cycle* **7**, 3399–3404 (2008).

542. Parker, J. L. & Ulrich, H. D. Mechanistic analysis of PCNA poly-ubiquitylation by the ubiquitin protein ligases Rad18 and Rad5. *EMBO J* **28**, 3657–3666 (2009).

543. Unk, I. *et al.* Human HLTF functions as a ubiquitin ligase for proliferating cell nuclear antigen polyubiquitination. *Proc Natl Acad Sci U S A* **105**, 3768–3773 (2008).

544. Krijger, P. *et al.* HLTF and SHPRH are not essential for PCNA polyubiquitination, survival and somatic hypermutation: Existence of an alternative E3 ligase. *D N A Repair* **10**, 438–444 (2011).

545. Weston, R., Peeters, H. & Ahel, D. ZRANB3 is a structure-specific ATP-dependent endonuclease involved in replication stress response. *Genes Dev* **26**, 1558–1572 (2012).

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

547. Wilkinson, K. A. & Henley, J. M. Mechanisms, regulation and consequences of protein SUMOylation. *Biochemical Journal* **428**, 133–145 (2010).

548. Carnie, C. J. *et al.* Decitabine cytotoxicity is promoted by dCMP deaminase DCTD and mitigated by SUMO-dependent E3 ligase TOPORS. *EMBO J* **43**, 2397–2423 (2024).

549. Liu, J. C. Y. *et al.* Concerted SUMO-targeted ubiquitin ligase activities of TOPORS and RNF4 are essential for stress management and cell proliferation. *Nat Struct Mol Biol* **31**, 1355–1367 (2024).

550. Sugimoto, Y. *et al.* Novel mechanisms for the removal of strong replication-blocking HMCES- and thiazolidine-DNA adducts in humans. *Nucleic Acids Res* **51**, 4959–4981 (2023).

551. Tirman, S. *et al.* Temporally distinct post-replicative repair mechanisms fill PRIMPOL-dependent ssDNA gaps in human cells. *Mol Cell* **81**, 4026-4040.e8 (2021).

552. Rua-Fernandez, J. *et al.* Self-reversal facilitates the resolution of HMCES-DNA protein crosslinks in cells. 2023.06.14.544844 Preprint at https://doi.org/10.1101/2023.06.14.544844 (2023).

553. Peña-Gómez, M. J. *et al.* HMCES corrupts replication fork stability during base excision repair in homologous recombination–deficient cells. *Science Advances* **11**, eads3227 (2025).

554. Masuda, K. *et al.* DNA polymerase theta contributes to the generation of C/G mutations during somatic hypermutation of Ig genes. *Proc Natl Acad Sci U S A* **102**, 13986–13991 (2005).

555. Patterson-Fortin, J. & D'Andrea, A. D. Exploiting the Microhomology-Mediated End-Joining Pathway in Cancer Therapy. *Cancer Res* **80**, 4593–4600 (2020).

556. Ercilla, A. *et al.* Physiological Tolerance to ssDNA Enables Strand Uncoupling during DNA Replication. *Cell Rep* **30**, 2416-2429.e7 (2020).

557. Weickert, P., Dürauer, S., Götz, M. J., Li, H.-Y. & Stingele, J. Electro-elution-based purification of covalent DNA-protein cross-links. *Nat Protoc* **19**, 2891–2914 (2024).

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

559. Raiber, E.-A. *et al.* 5-Formylcytosine organizes nucleosomes and forms Schiff base interactions with histones in mouse embryonic stem cells. *Nature Chem* **10**, 1258–1266 (2018).

560. Song, W. *et al.* The dual ubiquitin binding mode of SPRTN secures rapid spatiotemporal proteolysis of DNA–protein crosslinks. *Nucleic Acids Res* **53**, gkaf638 (2025).

561. Xu, P. *et al.* Quantitative proteomics reveals the function of unconventional ubiquitin chains in proteasomal degradation. *Cell* **137**, 133–145 (2009).

562. Komander, D. Mechanism, specificity and structure of the deubiquitinases. *Subcell Biochem* **54**, 69–87 (2010).

563. Ronau, J. A., Beckmann, J. F. & Hochstrasser, M. Substrate specificity of the ubiquitin and Ubl proteases. *Cell Res* **26**, 441–456 (2016).

564. Cosenza-Contreras, M. *et al.* TermineR: Extracting information on endogenous proteolytic processing from shotgun proteomics data. *Proteomics* **24**, e2300491 (2024).

565. Schmid, E. W. & Walter, J. C. Predictomes, a classifier-curated database of AlphaFold-modeled protein-protein interactions. *Molecular Cell* **85**, 1216-1232.e5 (2025).

566. Krupina, K., Goginashvili, A. & Cleveland, D. W. Causes and consequences of micronuclei. *Curr Opin Cell Biol* **70**, 91–99 (2021).

567. Chan, K. L. & Hickson, I. D. New insights into the formation and resolution of ultra-fine anaphase bridges. *Semin Cell Dev Biol* **22**, 906–912 (2011).

568. Brand, C. L. & Levine, M. T. Cross-species incompatibility between a DNA satellite and the Drosophila Spartan homolog poisons germline genome integrity. *Curr Biol* **32**, 2962-2971.e4 (2022).

569. Kröning, 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).