Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

Substrate-specificity of the DNA-protein crosslink repair protease SPRTN

Hannah Kristina Reinking
aus
Düsseldorf, Deutschland

2021

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, 27.09.2021

_____

Hannah Reinking

Dissertation eingereicht am 01.10.2021

1. Gutachter: Prof. Dr. Julian Stingele
2. Gutachter: Prof. Dr. Johannes Stigler

Mündliche Prüfung am 18.02.2022

This thesis has been prepared from December 2017 to September 2021 in the laboratory of Professor Dr. Julian Stingele at the Gene Center of the Ludwig-Maximilians-University München.

This is a cumulative thesis based on the following publications:

Hannah K. Reinking, Hyun-Seo Kang, Maximilian J. Götz, Hao-Yi Li, Anja Kieser, Shubo Zhao, Aleida C. Acampora, Pedro Weickert, Evelyn Fessler, Lucas T. Jae, Michael Sattler, Julian Stingele, DNA Structure-Specific Cleavage of DNA-Protein Crosslinks by the SPRTN Protease, *Molecular Cell,* Volume 80, Issue 1, 1 October 2020, Pages 102-113, https://doi.org/10.1016/j.molcel.2020.08.003

Shubo Zhao, Anja Kieser, Hao-Yi Li, Hannah K. Reinking, Pedro Weickert, Simon Euteneuer, Denitsa Yaneva, Aleida C. Acampora, Maximilian J. Götz, Regina Feederle, Julian Stingele, A ubiquitin switch controls autocatalytic inactivation of the DNA-protein crosslink repair protease SPRTN, *Nucleic Acids Research*, Volume 49, Issue 2, 25 January 2021, Pages 902-915, https://doi.org/10.1093/nar/gkaa1224

Hannah K. Reinking, Kay Hofmann, Julian Stingele, Function and evolution of the DNA-protein crosslink proteases Wss1 and SPRTN, *DNA Repair*, Volume 88, 6 February 2020, Pages 102822, https://doi.org/10.1016/j.dnarep.2020.102822 (Review article)

Hannah K. Reinking, Julian Stingele, Protein-oligonucleotide conjugates as model substrates for DNA-protein crosslink repair proteases, *STAR Protocols,* Volume 2, Issue 2, 18 June 2021, Pages 100591, https://doi.org/10.1016/j.xpro.2021.100591 (Experimental protocol)

Contents

Abstract

The integrity of DNA is constantly challenged by a huge variety of damaging agents. If not repaired, damaged DNA can lead to severe diseases and cell death. Therefore, cells employ specific pathways, which target and repair those lesions in order to maintain genome stability. Most of these pathways have been in the focus of intense research efforts, which resulted in a deep understanding of the underlying mechanisms. However, the repair of DNA-protein crosslinks (DPCs) has only received attention during the last years. These DNA lesions occur when chromatin proteins or enzymes acting on DNA become permanently trapped. DPCs are highly toxic, because they can interfere with replication and transcription and therefore cells need to repair them.

A recently discovered repair mechanism involves the proteolytic degradation of the protein component by the DPC-protease SPRTN. SPRTN is a mammalian protease, which is mainly expressed in S-phase and is involved in the repair of DPCs *in vivo* and *in vitro.* In cells, SPRTN promotes resistance towards DPC-inducing agents such as formaldehyde and camptothecin. Its important function is underlined by the fact that a complete *SPRTN* knockout in cells is not viable. In addition, mutations within the *SPRTN* gene are associated with a disease called Ruijs-Aalfs-Syndrome (RJALS), which features premature ageing and hepatocellular carcinoma. *In vitro* analyses showed that SPRTN is a highly promiscuous DNA-dependent protease, which is able to degrade DNA-binding proteins irrespectively of their size, shape or amino acid sequence. This broad substrate specificity enables the enzyme to target a diverse set of DPCs. On the other hand however, it carries the risk of uncontrolled proteolysis of essential chromatin proteins. Despite the major discoveries that have been made during recent years, crucial questions about SPRTN´s regulation remain unanswered. First, the exact mechanism ensuring that SPRTN´s activity is restricted to toxic DPCs, while leaving important DNA-binding proteins untouched, remains elusive. Second, SPRTN is constitutively monoubiquitylated and actively deubiquitylated upon induction of DPCs. However, whether this modification affects SPRTN´s activity directly is not known.

The first part of this thesis deals with the biochemical characterization of SPRTN´s DNA-dependent activation. SPRTN is revealed to bear a DNA-structure specific activity. The enzyme only becomes active if a DPC is positioned at disruptions within double-stranded DNA (dsDNA). This structure-specificity restricts proteolysis to a very narrow window, thereby

protecting chromatin proteins in the vicinity. By using NMR spectroscopy, it is shown that this specificity is achieved by the direct engagement of SPRTN´s two DNA-binding domains: the zinc-binding domain (ZBD) is interacting with unpaired nucleotides and the basic region (BR) binds to the negatively-charged phosphate backbone of dsDNA. Upon simultaneous engagement of these two domains, SPRTN becomes active and cleaves its substrates.

The second part of the thesis addresses the effect of SPRTN´s monoubiquitylation on its activity. Previous studies suggested that the ubiquitylation-state coordinates chromatin access. However, recent data show that monoubiquitylation negatively regulates the SPRTN pool in cells. This happens either by priming the enzyme for degradation by the proteasome or by increasing its autocatalytic cleavage activity *in trans*. Here, the direct effect of monoubiquitylation on recombinant SPRTN proteins is tested using *in vitro* auto- and substrate cleavage assays. It is shown that monoubiquitylation increases autocatalytic cleavage activity of the enzyme towards other SPRTN molecules. However, cleavage of DNA-binding or crosslinked proteins is only mildly affected. These data point towards a role of monoubiquitylation in decreasing the cellular SPRTN concentration. Upon DPC induction, the modification is removed and SPRTN´s half-life extended to process toxic DPCs.

Together, both studies contribute to the understanding of the regulatory mechanisms underlying SPRTN´s activation. Since many chemotherapeutics are known to generate DPCs, understanding the pathway that repairs those lesions may enable and guide the development of combinational anti-cancer treatments targeting DPC repair.

1 Introduction

1.1 DNA Damage and Repair

Genetic information of cellular organisms is stored in deoxyribonucleic acid (DNA) (Travers and Muskhelishvili, 2015). Every day, the integrity of DNA is challenged by various agents, which can induce sequential or structural alterations (Chatterjee and Walker, 2017). Any kind of alteration within the genetic code can have a severe impact on the organism´s properties. While some mutations may be beneficial during evolution, others are harmful and can lead to ageing and diseases such as cancer (Chatterjee and Walker, 2017; Schumacher et al., 2021). DNA damages can be of diverse nature ranging from mismatches, single base or nucleotide alterations, DNA adducts or even breaks within the DNA strands and can be repaired by specific DNA repair pathways (Figure 1) (Chatterjee and Walker, 2017).

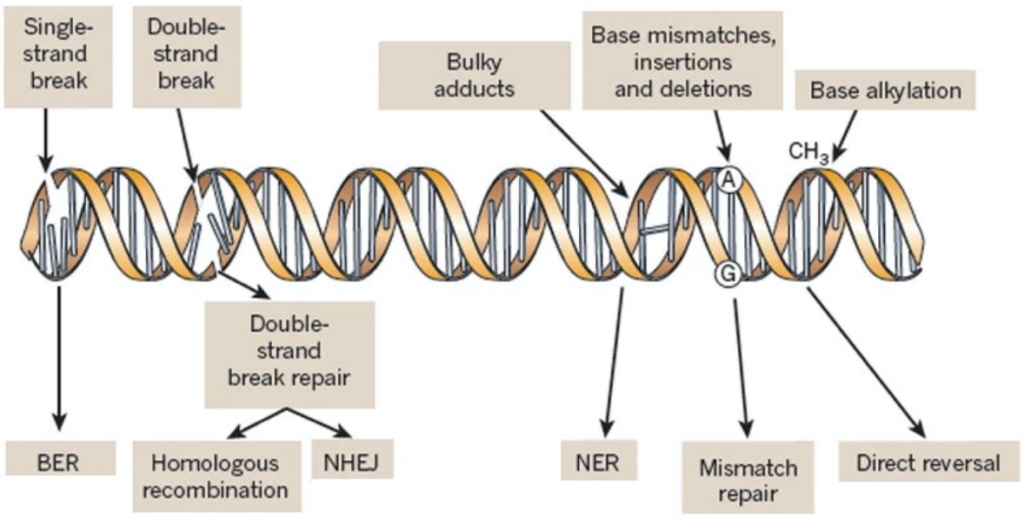


Figure 1: DNA damage and repair pathways. Endogenous and exogenous agents can induce various DNA damages such as mismatches, base modifications, bulky adducts or even single-strand or double-strand breaks. These lesions are targeted by different repair pathways such as mismatch repair, nucleotide excision repair (NER) or base excision repair (BER). Double-strand breaks are repaired by homologous recombination or non-homologous end joining (NHEJ). (Figure adapted from (Lord and Ashworth, 2012))

1.1.1 Endogenous Sources of DNA Damage

DNA changes or lesions can be generated through endogenous processes such as replication. Normally, during replication an identical DNA copy is generated with the help of highly accurate polymerases (Loeb and Monnat, 2008). These polymerases contain a proof-reading activity, which helps to pass the correct genetic information to daughter cells. However, it is still possible that some misincorporated nucleotides escape proof-reading and are retained in the newly synthesized strand. These alterations involve substitutions of nucleotides, deletions but also insertions and alter the DNA sequence (Kunkel, 2004; 2011). Once a nucleotide is falsely incorporated, the incorrect information will be passed on to daughter cells during the next replication cycle, leading to mutations within the genome.

Furthermore, any chemical alteration of the DNA is a potential threat to genome integrity. These include base deamination, abasic sites and DNA methylation (Chan et al., 2013; Lindahl, 1993; Rydberg and Lindahl, 1982; Yonekura et al., 2009). During base deamination an amino group is removed from the DNA base, which can happen spontaneously (Kow, 2002). A prominent example is the deamination from cytosine or 5-methyl-cytosine to generate uracil or thymine, respectively (Lindahl, 1993). These base changes will influence the base pairing properties during the next round of replication (e.g. a C:G will result in U:A), which will alter the sequence of the genetic code. In addition, DNA methylation can also lead to mispairing and to the incorporation of incorrect nucleotides during replication (Rydberg and Lindahl, 1982; Sedgwick, 2004). Moreover, abasic sites are continuously generated and lack purine or pyrimidine bases (apurinic/apyrimidinic site; AP site) (Lindahl, 1993). They are either formed by spontaneous hydrolysis of the glycosidic bond between the nitrogen base and the sugar phosphate backbone or can be generated as an intermediate reaction product during base excision repair (BER) by glycosylases (Krokan and Bjoras, 2013; Lindahl, 1993).

Reactive oxygen species (ROS) are other frequent sources of endogenous DNA damage (Cooke et al., 2003). They are generated during metabolic processes or occur through radiolysis of water when exposed to ionizing radiation (IR) (Cooke et al., 2003; Srinivas et al., 2019). The most common ROS are hydroxyl radicals (OH•), which can attack multiple points within the DNA and can lead to chemical alterations. A prominent example is the generation of 8-oxo-guanine (Fujimoto et al., 2007). This modified base pairs with adenine instead of cytosine and therefore carries the risk of introducing point mutations. In addition,

ROS can induce strand breaks, which are highly toxic and belong to the most detrimental DNA damages (Srinivas et al., 2019).

1.1.2 Exogenous Sources of DNA Damage

In addition to endogenously produced lesions, DNA can face various toxic exogenous agents, e.g. IR, ultraviolet (UV) radiation or alkylating agents (Chatterjee and Walker, 2017). IR includes α, β, γ radiation as well as x-rays, which are, for instance, produced by medical devices (Desouky et al., 2015). IR can either directly damage DNA or indirectly by the generation of OH• radicals during the radiolysis of water. As described above, OH• radicals are also endogenously produced, which explains the overlap in DNA damage induced by ROS such as 8-oxo-guanine, base lesions and strand breaks (Chatterjee and Walker, 2017; Vignard et al., 2013). DNA can further be damaged by direct absorptions of UV light or by energy transfer of nearby molecules (Chatterjee and Walker, 2017). UV light is found in the atmosphere and consists of UVA (320-400 nm), UVB (290-320 nm) and UVC (290 nm) (Chatterjee and Walker, 2017; Davies, 1995). Since DNA absorbs light at 260 nm, UV light can lead to the formation of photochemical alterations, which induce conformational changes within the DNA (Yu and Lee, 2017).

Finally, alkylating agents are found in many pollutants like tobacco smoke and fuel combustion products (Fu et al., 2012). These compounds have the ability of transferring alkyl carbon groups onto DNA, thereby changing its chemical properties. In particular, they can target DNA bases by reacting with the ring nitrogens and extracyclic oxygens in order to generate covalent adducts (Fu et al., 2012). These adducts range from the addition of simple methyl groups up to the formation of complex structures and depend on the reactive sites within the alkylating agents itself (Drablos et al., 2004; Fu et al., 2012; Shrivastav et al., 2010). Due to their highly cytotoxic effect, some chemotherapeutic treatments include alkylating agents in order to kill cancer cells efficiently (Fu et al., 2012).

The above-mentioned exogenous agents are just a small subset of sources that cause DNA damage. Many other toxic agents are also able to induce DNA lesions and constantly challenge the integrity of DNA.

1.1.3 Canonical DNA Repair Pathways

In order to maintain genome stability, cells have developed specialized pathways that repair, tolerate or control the accumulation of DNA lesions (Figure 1) (Chatterjee and Walker, 2017). Germline mutations within genes that belong to DNA repair enzymes are associated with premature ageing and several diseases, indicating their importance for human health (Jackson and Bartek, 2009). Therefore, different repair pathways exist that target specific DNA lesions, which are briefly discussed in the following paragraph.

First, the incorporation of intact but non-complementary nucleotides can be sensed and repaired by the mismatch repair (MMR) pathway, which acts in a post-replicative manner and contributes to the high fidelity of replication (Jiricny, 2013; Kunkel, 2009). This pathway is initiated by MutSα and MutSβ, which bind the mismatch and recruit other proteins to the lesion including MutLα, the proliferating-cell-nuclear-antigen (PCNA) and RFC (Pecina-Slaus et al., 2020). The assembly of proteins culminates in an endonucleolytic incision thereby leading to a single-strand break. An exonuclease subsequently removes the misincorporated nucleotide and the replicative machinery is able to fill and ligate the remaining gap.

The repair of damaged DNA bases can be accomplished using different approaches. First, some enzymes are able to directly reverse the base modification. As an example, alkyltransferases can directly remove base alkylation by transferring the alkyl adduct to their active site cysteines (Tubbs et al., 2007). Alternatively, damaged DNA bases can be targeted by the base excision repair (BER) pathway, which recognizes damages like base oxidation, deamination, alkylation or abasic sites (Krokan and Bjoras, 2013). Upon damage recognition and base excision, an AP site remains within the DNA. After backbone cleavage by an AP endonuclease and processing of the DNA ends, a polymerase can fill the resulting nucleotide gap, which can be further ligated to repair the DNA damage.

Furthermore, bulkier DNA lesions can be recognized and repaired by the nucleotide excision repair (NER) pathway (Nouspikel, 2009). NER can be differentiated into the global genome NER (GG-NER) and the transcription-coupled NER (TC-NER) pathway, which differ in their initiating steps (Scharer, 2013; Vermeulen and Fousteri, 2013). In GG-NER distortions or sterically aberrations within the DNA strands are recognized by XPC-RAD23B (Scharer, 2013; Sugasawa et al., 1998; Sugasawa et al., 2001). TC-NER is initiated by a stalling event of the RNA polymerase II during transcription and involves the specific recruitment of the

proteins CSA and CSB, which lead to the assembly of other components of the NER machinery (Vermeulen and Fousteri, 2013). In GG-NER and TC-NER the incision events are performed by a common pathway. The transcription factor TFIIH opens the DNA with the help of the helicase subunit activities of XPB and XPD and creates a bubble around the lesion (Evans et al., 1997; Scharer, 2013). Next, a preincision complex by recruitment of XPA, RPA and XPG is formed at the lesion site (Volker et al., 2001). With the recruitment of ERCC1-XPF the dual incision event is triggered (Fagbemi et al., 2011). The two endonucleases ERCC1-XPF and XPG cleave the DNA on both lesion sides. ERCC1-XPF generates a free OH-end, which can be used by the replication machinery whereas XPG generates a 5´phosphate end for the ligation after strand synthesis. The excised DNA fragment is released from the DNA while bound to TFIIH and the gap of approx. 25-30 nucleotides is closed and ligated by the replication machinery (Ogi et al., 2010; Scharer, 2013; Vermeulen and Fousteri, 2013).

Interstrand crosslinks (ICLs) occur when the two complementary DNA strands become covalently attached to each other (Rageul and Kim, 2020). This type of DNA damage can be induced by exogenous sources such as cisplatin and blocks DNA replication. A disease, which is related to this DNA damage, is called Fanconi anaemia (FA) (Kottemann and Smogorzewska, 2013). Mutations within FA genes lead to variable phenotypes including bone-marrow failure, acute myeloid leukaemia (AML) and solid tumors (Auerbach, 2009). The repair of ICLs by the FA pathway consists of an interplay of multiple DNA repair pathways such as nucleotide excision repair, homologous recombination and translesion synthesis (Ceccaldi et al., 2016).

The most detrimental DNA lesions are breaks within both DNA strands and are referred to as DNA double-strand breaks (DSB). Unrepaired DSBs can lead to genomic rearrangements, loss of genetic information and even to cell death (Aparicio et al., 2014). Two major pathways of DSB repair exists: non-homologous end joining (NHEJ) and homologous recombination (HR) (Chang et al., 2017; Scully et al., 2019). The choice between HR and NHEJ depends on the cell cycle state. HR is active in S- or G2-phase because it needs a homologous DNA sequence and uses the sister chromatid as a template (Scully et al., 2019). HR is initiated by binding of the MRN complex (MRE11-RAD50-NBS1) and CtIP as well as exonuclease EXO1 and DNA2-Bloom syndrome protein (BLM) heterodimer (Sartori et al., 2007; Scully et al., 2019; Symington, 2014). 3´single-stranded overhangs are generated, a process called DNA end resection. These overhangs are covered

with replication protein A (RPA), which is further displaced with the help of breast cancer type 2 protein (BRCA2) in order to allow loading of the RAD51 recombinase (Scully et al., 2019). These filaments then scan the genome for a homologous sequence. After strand invasion of the 3´end, the homologous sequence is used as a template for nascent strand synthesis (Renkawitz et al., 2014; Renkawitz et al., 2013).

Compared to HR, NHEJ does not require the existence of a homologous sequence and rather ligates the broken ends together (Scully et al., 2019). This process carries a high risk of introducing mutations such as nucleotide deletions or insertions. Initiation of NHEJ is carried out by the Ku complex, consisting of Ku70 and Ku80, which recognize the DSB (Lieber, 2010). This further leads to the recruitment of DNA-dependent protein kinase (DNA-PKs) and Artemis, which process the ends for ligation in order to repair the DSB (Goodarzi et al., 2006; Scully et al., 2019).

The mechanisms described above all aim to repair the DNA lesion. However, cells also developed DNA damage tolerance mechanisms that help them to survive in the presence of DNA damage. Translesion synthesis (TLS) is an example of such a mechanism (Sale, 2013). It enables replication across damaged nucleotides by employing specialized polymerases. Even though this type of DNA synthesis can lead to an increased incorporation of incorrect nucleotides, it can protect cells from even more severe DNA damages such as strand breaks.

1.2 DNA-Protein Crosslink Repair Proteases

1.2.1 DNA-Protein Crosslinks

DNA-protein crosslinks (DPCs) are highly toxic DNA lesions that interfere with processes on DNA such as replication and transcription and therefore endanger genome integrity (Duxin et al., 2014; Fu et al., 2011; Ji et al., 2019; Nakano et al., 2013; Nakano et al., 2012). DPCs can be formed by the covalent entrapment of proteins that act directly on DNA (enzymatic DPCs) or are in close proximity to it (non-enzymatic DPCs) (Stingele et al., 2015). A major challenge in the repair of DPCs originates from their large diversity: DPCs can vary in the nature of the crosslinked protein, the crosslink between DNA and protein itself as well as the DNA structure that is involved in the crosslinking reaction (Figure 2) (Nakano et al., 2017). An additional challenge is that DPCs form frequently, because they are induced by a large variety of endogenous or exogenous sources (Barker et al., 2005a; Stingele et al., 2017). Different crosslinking agents include aldehydes, UV light, IR, different metals and also chemotherapeutics (Barker et al., 2005a; Barker et al., 2005b; Ide et al., 2018).

DPCs can be formed within dsDNA and are induced by a variety of compounds such as reactive aldehydes. Aldehydes are produced endogenously and especially formaldehyde is present at high concentration in human plasma (Luo et al., 2001). Formaldehyde is even generated in the direct vicinity of DNA during metabolic processes such as histone demethylation (Shi et al., 2004; Tsukada et al., 2006). Here it has the ability to crosslink proteins via the amino acid residues lysine, cysteine, histidine or tryptophan to the exocyclic amine of DNA bases by forming a methylene bridge (Figure 2) (Lu et al., 2010). Due to its high crosslinking efficiency formaldehyde is often used to study sensitivity towards DPCs in cells. Acetaldehyde is another reactive aldehyde and is produced endogenously as a byproduct of ethanol metabolism and was shown to crosslink proteins to DNA (Behrens et al., 1988; Kurtz and Lloyd, 2003; Sako et al., 2003). However, even though aldehydes are very efficient in inducing DPCs, they also lead to other lesions such as protein-protein crosslinks, small DNA adducts or ICLs, which should be considered when analysing the cytotoxic effect of DPCs induced by these compounds.

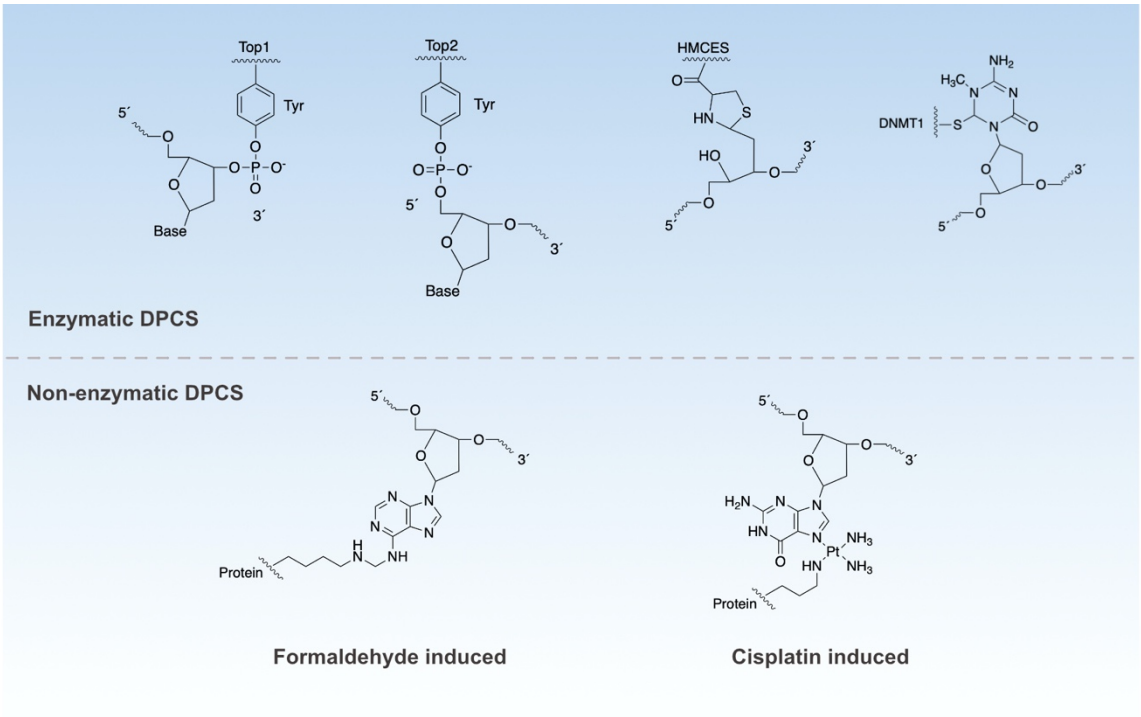


Figure 2: Chemical structures of different DPCs. Enzymatic DPCs are formed when proteins that act on DNA become permanently trapped as in the case of TOP1, TOP2, HMCES and DNMT1. Non enzymatic DPCs involve a variety of proteins that are in the close vicinity to DNA and trapping is induced by reactive compounds such as formaldehyde or cisplatin. DPCs are very heterogenous, as the crosslink can differ with respect to the amino acid within the protein as well as to the DNA position that is targeted. (Figure modified from (Kuhbacher and Duxin, 2020))

Some chemotherapeutics have the ability to crosslink proteins to DNA. Platinum compounds like cisplatin, carboplatin and oxaliplatin are commonly used cancer chemotherapeutics (Groehler et al., 2017). Indeed, cisplatin was the first platinum anticancer drug that was approved by the FDA and is used to treat bladder, head, lung or ovarian cancer (Dasari and Tchounwou, 2014; Kelland, 2007). Even though platinum-based compounds are mainly known to induce ICLs, they also contribute to the formation of DPCs (Kasparkova et al., 2006). Cisplatin is hereby targeting cysteine, arginine and lysine side chain residues and crosslinks them to purine bases (Figure 2) (Groehler et al., 2017; Kuhbacher and Duxin, 2020). Other chemotherapeutic agents with crosslinking abilities are nitrogen mustards. These chemotherapeutics are used in the treatment of lymphoma, leukemia, breast and brain

cancer (Emadi et al., 2009). DPCs induced by these compounds are formed by a linkage between the guanine base and a cysteine residue of the protein (Loeber et al., 2008).

Another commonly used chemotherapeutic drug is the cytosine analog 5′-aza-2-deoxycytidine (5-aza-dC). This compound traps DNA methyltransferase 1 (DNMT1) on DNA (Kuhbacher and Duxin, 2020). Upon treating cells with 5-aza-dC, the analog is incorporated into DNA during replication. Normally DNMT1 transfers methyl groups to the newly synthesized DNA strand (Gujar et al., 2019). However, the enzyme fails to complete the methylation reaction of the pseudo-substrate and remains covalently crosslinked through a cysteine side chain (Figure 2) (Schermelleh et al., 2005; Wu and Santi, 1987). In addition, entrapment of DNMT1 causes global hypomethylation, which leads to the re-expression of silenced tumor suppressor genes (Robert et al., 2003).

Furthermore, other exogenous sources like UV light and IR can generate crosslinks between proteins and DNA. Here, DPC generation by IR depends on the cellular oxygen level (Nakano et al., 2017). Under hypoxic conditions DPC formation is prominent compared to other DNA lesions, such as base damage, single-strand or double-strand breaks (Meyn et al., 1987; Miyagi et al., 1997; Murray et al., 1988; Nakano et al., 2015) The above mentioned endogenous and exogenous crosslinking agents are able to crosslink proteins to dsDNA very efficiently. However, theoretically they are also able to crosslink proteins to ssDNA using the same underlying mechanisms.

Another example of a crosslink formed in dsDNA is the entrapment of topoisomerase 1 (TOP1). Throughout replication and transcription TOP1 acts on DNA to release torsional stress (Champoux and Dulbecco, 1972). During the catalytic reaction, a covalent bond between the active centre tyrosine and the sugar phosphate backbone of DNA is formed. In a normal reaction cycle, the DNA is religated and TOP1 released from DNA (Pommier et al., 2014). However, nearby DNA damage can lead to distortion of the DNA and can prohibit realignment and religation of the strands, thereby trapping the enzyme on the 3′end of a single-strand break (SSB) (also called TOP1 cleavage complex or TOP1cc) (Figure 2) (Pourquier et al., 1997). Certain chemotherapeutics such as camptothecin and topotecan act as intercalating inhibitors and can prevent religation intentionally (Pommier, 2006; Pommier and Marchand, 2011). The resulting TOP1ccs are specifically toxic in highly proliferating cells, therefore these so called TOP1 poisons are widely used to treat ovarian, cervical and lung cancers.

Moreover, proteins, which are involved in the repair of oxidized abasic sites, can become trapped to DNA as in the case of polymerase ß (Polß) (DeMott et al., 2002; Quinones et al., 2015; Sung and Park, 2005). Polß crosslinks are localized at a SSB but compared to TOP1ccs are formed between a lysine residue and the 5´-end of the DNA (DeMott et al., 2002). Other proteins, like poly(ADP-ribose) polymerase 1 (PARP1) can become crosslinked to either the 3´- or 5´-end of a SSB while binding to abasic sites (Ide et al., 2018; Khodyreva et al., 2010; Kuhbacher and Duxin, 2020; Prasad et al., 2014; Prasad et al., 2019; Prasad et al., 2020).

DPCs also occur at DSBs. A prominent example is the entrapment of topoisomerase 2 (TOP2) forming a covalent cleavage complex (TOP2cc, Figure 2) (Pommier et al., 2016). TOP2 is a homodimer and releases torsional stress by cleaving both strands and forming a covalent linkage with both 5´-ends of the DNA. The DSB is normally religated, but - as in the case of TOP1- specific intercalating agents such as etoposide and doxorubicin can interfere with this process and trap TOP2 on DNA. A TOP2-like crosslink to the 5´end of DNA can also be formed by the topoisomerase-like enzyme SPO11 (Hartsuiker et al., 2009; Mimitou et al., 2017; Neale et al., 2005). SPO11 initiates meiotic recombination by inducing DSBs and covalently attaching to the 5´ DNA ends (Mimitou et al., 2017). Normally, these DPCs are removed by endonucleolytic cleavage, which depends on Mre11. Here, it was shown that defective Mre11-dependent initiation of resection can lead to the formation of SPO11 cleavage complexes (SPO11ccs) (Paiano et al., 2020).

Even though most DPCs pose a serious threat to genome integrity, some DPCs exist that play an important role in preventing more detrimental DNA lesions. 5-Hydroxymethylcytosine Binding, ES Cell Specific (HMCES) is a protein, which intentionally crosslinks to abasic sites to prevent the formation of a DSB (Figure 2) (Mehta et al., 2020; Mohni et al., 2019). Abasic sites are formed spontaneously or as an intermediate of the BER pathway. Here, different glycosylases sense and repair damaged bases and can generate an AP site. This AP site can be further processed by an AP endonuclease such as APE1 (Marenstein et al., 2004). APE1 incises the DNA backbone within the phosphodiester bond 5´ to the AP site, generating a SSB. APE1 cleaves the DNA backbone mainly in dsDNA, however, AP sites on ssDNA can also be targeted (Marenstein et al., 2004). If an AP site within ssDNA is targeted by such an incision event, a DSB would be generated. By crosslinking to an abasic site in ssDNA via a thiazolidine linkage, HMCES is shielding this site and protects it from further processing, thereby avoiding the generation of DSBs (Mehta

et al., 2020; Mohni et al., 2019; Thompson et al., 2019). The HMCES crosslink is afterwards resolved by proteasomal degradation (Mohni et al., 2019).
As a conclusion, DPCs are highly diverse DNA lesions with regard to the crosslinked protein as well as the DNA structure that is involved. Due to their heterogenous nature their repair is extremely challenging and specialized pathways with overlapping functions are required to maintain genome stability.

1.2.2 The Metalloproteases SPRTN and Wss1

A dedicated DPC repair mechanism was first described with the discovery of the yeast metalloprotease Wss1 (weak suppressor of smt3) (Stingele et al., 2014). Wss1 was initially characterized as being a suppressor of a hypomorphic small ubiquitin-like modifier (SUMO) variant, but its exact molecular function had remained elusive (Biggins et al., 2001). A genome-wide screen in yeast discovered that the simultaneous loss of Wss1 and Tdp1 leads to the accumulation of TOP1ccs (Stingele et al., 2014). The TOP1ccs further result in cell cycle arrest and DNA damage checkpoint activation. The synthetic lethality of *Δtdp1 Δwss1* led to the conclusion that Wss1 plays a major role in repairing DPCs.
Further studies in different model organisms discovered other proteases with a similar role in DPC repair. In *Xenopus laevis (X.laevis)* egg extracts DPCs are repaired in a replication-dependent manner by an (at the time) unknown protease (Duxin et al., 2014). Bioinformatic analysis identified proteases in metazoans, that are similar to Wss1 and revealed the metalloprotease SPRTN as being the closest candidate in mammalian cells (Stingele et al., 2015). Subsequently, several studies addressed the role of SPRTN as a DPC protease in higher eukaryotes (Maskey et al., 2017; Morocz et al., 2017; Stingele et al., 2016; Vaz et al., 2016).
On a protein-level, Wss1 and SPRTN show many similarities. Even though the sequence identity is rather low and Wss1 is smaller than SPRTN, the domain organization and motives resemble each other (Figure 3) (Reinking et al., 2020a; Stingele et al., 2015). Both proteins contain an N-terminal metal-binding protease domain, which is crucial for their proteolytic activity. DNA-binding domains are located C-terminally to the protease domain. Additionally, the C-terminal tails contain regions for protein-protein interactions for the chaperone-like enzymes Cdc48/p97 (Wss1/SPRTN), SUMO/ubiquitin (Wss1/SPRTN) or the replication clamp protein PCNA (SPRTN).

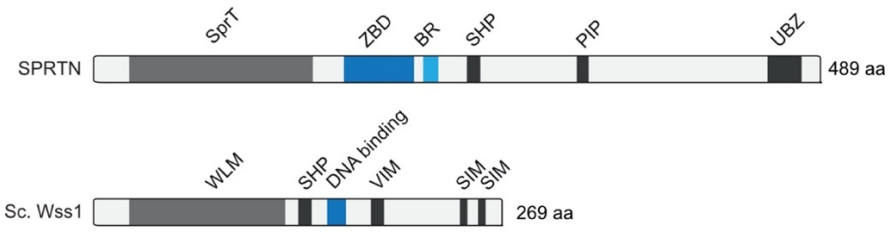


Figure 3: Domain organization of SPRTN and *S.cerevisiae* Wss1. Protease domains are depicted in grey, DNA-binding domains in blue and protein-protein interacting motives in black. SPRTN contains two different DNA-binding domains: the zinc-binding domain (ZBD) and the basic region (BR). SPRTN and Wss1 interact with the chaperon-like enzymes Cdc48/p97 (SHP/VIM), with SUMO or ubiquitin (SIM/UBZ) or with PCNA (PIP).

Wss1 and SPRTN execute very similar functions *in vivo*. Both proteases are essential for genome stability and their loss has severe consequences in different organisms. In yeast, loss of Wss1 leads to increased recombination and chromosomal rearrangements (Stingele et al., 2014). Plants missing Wss1 show strong developmental defects (Enderle et al., 2019a; Enderle et al., 2019b). Mice carrying a hypomorphic *Sprtn* mutation suffer from premature ageing and show predisposition to developing cancer (Maskey et al., 2017). Mammalian *SPRTN*-knockout cells are not viable. Loss of SPRTN results in cell death within a few cell cycles, underlining the essential role in DPC repair (Stingele et al., 2016). *SPRTN* mutations within the human genome are rare and lead to a disease called Ruijs-Aalfs-Syndrome (RJALS) (Lessel et al., 2014; Ruijs et al., 2003). RJALS mutations were found in three patients, which suffered from segmental progeria, genomic instability and early onset hepatocellular carcinomas (HCC) (Figure 4) (Lessel et al., 2014). Sanger sequencing analysis revealed biallelic mutations within the *SPRTN* gene for all three patients (Figure 4b). The first patient showed a 1 bp deletion, which resulted in a frameshift leading to a premature stop codon and truncating the entire C-terminal tail with its protein-protein interacting domains (Lys241 + X8; SPRTN-ΔC). The two other patients were compound heterozygous for a missense mutation leading to the replacement of tyrosine 117 to cysteine (Y117C), which is close to the active site residues within the protease domain. The second allele carried a 4 bp deletion introducing a premature stop codon (K239 + X7). The effect of the mutations found in RJALS was also analysed in human cells (Lessel et al., 2014). Depletion of SPRTN in U2OS cells results in DNA-replication defects, which can be restored by expressing the WT protein. By expression of the SPRTN-ΔC mutant, fork progression can be retained but with a much lesser extend compared to the wild type (WT).

The replacement of Y117C is completely unable to restore progression, highlighting the importance of the protease domain in SPRTN function.

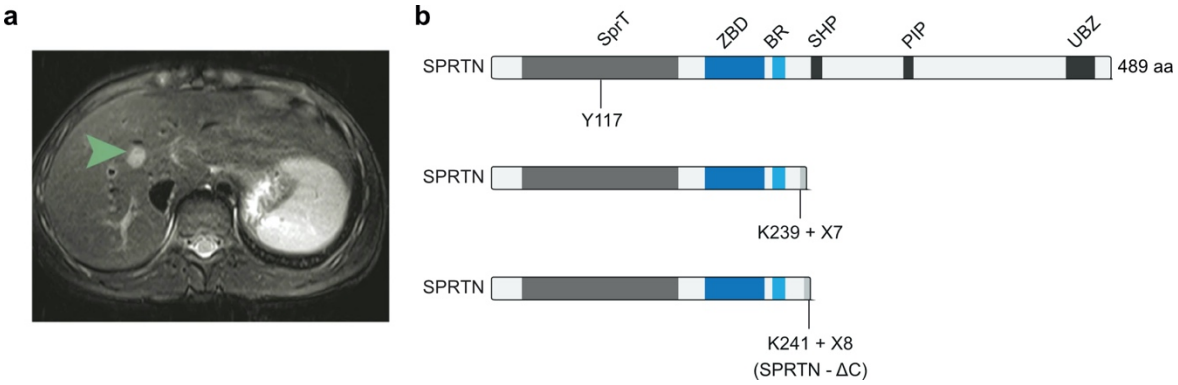


Figure 4: *SPRTN* mutations lead to a disease called Ruijs-Aalfs-Syndrome (RJALS). a) Axial view of magnetic resonance imaging of the liver of a patient suffering from hepatocellular carcinoma. (Figure adapted from (Lessel et al., 2014)) b) Schematic overview of the domain organization of SPRTN. The SprT domain is depicted in grey, the DNA-binding regions ZBD and BR are shown in blue and the protein-protein interacting motives SHP, PIP and UBZ are shown in black. Three biallelic mutations were found in patients suffering from RJALS. The first set of patients contained a missense mutation at Y117C and a 4 bp deletion leading to a preliminary stop codon at K239 + X7 amino acids. The second patient mutation contained a 1 bp deletion leading to the premature stop codon at K241 + X8 amino acids. The frameshift mutations result in the truncation of the C-terminal tail. The mutation K241 + X8 is often studied in the context with SPRTN activity and is referred to as SPRTN–ΔC variant.

In vitro experiments with Wss1 and SPRTN indicated similar proteolytic activities towards their substrates. Both proteases are activated in the presence of DNA and exhibit two distinct proteolytic activities (Stingele et al., 2016; Stingele et al., 2014; Vaz et al., 2016). First, once bound to DNA, they undergo autocleavage *in trans*, meaning one enzyme is cleaving a second. Second, they are able to cleave various DNA-binding proteins, but not non DNA-binding proteins. Since the presence of DNA was found to be absolutely essential for these activities, it was proposed that DNA might act as a scaffold bringing the protease in close proximity to its substrates (Stingele et al., 2016). Additionally, it was speculated that DNA might induce conformational changes within the protease, thereby leading to its activation (Stingele et al., 2016).

Importantly, with respect to substrate identities, both DPC proteases do not show any specificity towards substrate sequence or structure. Wss1 is able to degrade Top1, H1 or Hmg1 and also SPRTN is able to digest the DNA-binding proteins H1, H2A, H2B, H3 and

Hgm1 efficiently (Stingele et al., 2016; Stingele et al., 2014). Non DNA-binding proteins like GFP or BSA are not digested by neither protease. This promiscuity towards substrates allows these DPC proteases to be very efficient, but also carries the risk of unwanted proteolysis of important non-crosslinked DNA-binding proteins. Several studies discussed the role of how different types of DNA might control this proteolytic activity (e.g. ssDNA vs. dsDNA) but the results were conflicting and no common conclusion could be made (Stingele et al., 2016; Toth et al., 2017; Vaz et al., 2016).

In order to gain structural insights into the DNA-dependent activation mechanism, attempts have been made to obtain crystal structures of SPRTN/Wss1 family members (Stingele et al., 2016). However, studies addressing the full length proteins including the C-terminal tail were unsuccessful and proteins were insoluble when expressed in *E.coli*. Furthermore, the C-terminal flexible tail interfered with crystallization (Stingele et al., 2016; Yang et al., 2017). Finally, the protease domain of Wss1b from *Schizosaccharomyces pombe (S.pombe)* was solved (5JIG) (Stingele et al., 2016). It shows a compact fold with 4 α-helices and four-stranded antiparallel β-sheets. In the active site a Ni^{2+}-ion is found as a replacement of the catalytic Zn^{2+} likely due to the His-tag purification. This Ni^{2+} is coordinated by three histidine residues, two water molecules and one oxygen molecule (Figure 5), a typical arrangement for metalloproteases containing a HEXXH motive. A glutamate side chain residue (human numbering E112) is found to polarize a water molecule for the nucleophilic attack of the substrate. By mutating this glutamate to glutamine (E112Q) the entire enzyme is inactivated but still retains its structure (PDB: 5LN5). Interestingly, the catalytic centre containing the metal-binding site is highly solvent exposed and no obvious substrate binding pocket is visible. This open access to the active site could explain the high promiscuity of Wss1 and SPRTN towards their various DNA-binding substrates as discussed above.

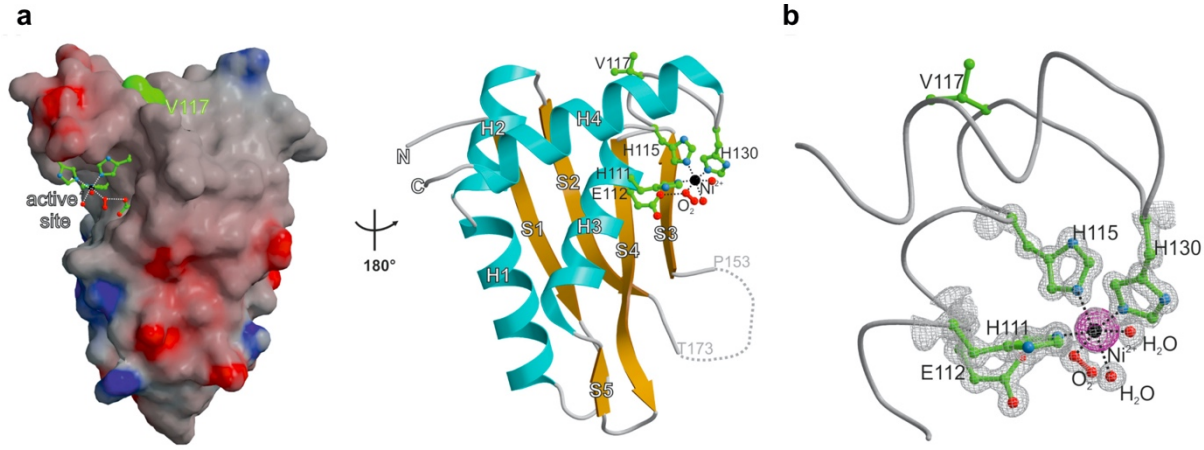


Figure 5: Crystal structure of Wss1 protease domain. a) Structure of the Wss1b protease domain from *S.pombe* in surface (left) and cartoon (right) presentation. Amino acid numbering corresponds to the human sequence. Active site residues coordinating the Ni^{2+} ion are presented as sticks. The Ni^{2+} ion probably displaced the Zn^{2+} ion during His-tag purification. b) Detailed view of the active site residues H111, E112, H115, and H130 coordinating the Ni^{2+} and polarizing a water molecule for the nucleophilic attack. (Figure from (Stingele et al., 2016))

Shortly after the structure of the Wss1 protease domain from *S.pombe* was solved, the protease structure from *Saccharomyces cerevisiae (S.cerevisiae)* was determined as well (Yang et al., 2017). Here again, first crystallization attempts addressed the full-length protein but were unsuccessful and lead to rapid degradation, allowing only crystallization of the protease domain. Both protease structures, from *S.pombe* and *S.cerevisiae*, are highly similar with the most conserved region in the catalytic core comprising the HEXXH motive for metal binding and activation. Importantly, none of the two structures contain the DNA-binding domains, hence no conclusion could be made on how Wss1 is activated by DNA.
In 2019 the protease structure belonging to human SPRTN was finally solved (Li et al., 2019). This structure includes the protease domain with a Zn^{2+} ion coordinated similarly to Wss1 (Figure 6). Additionally, it reveals the presence of a second Zn^{2+} binding domain (ZBD) between the protease domain and a basic region (BR), which was previously characterized in DNA-binding studies (Stingele et al., 2016). The ZBD is connected to the protease domain by a flexible loop and it is speculated that it shields the active centre thereby restricting substrate access (Li et al., 2019). Furthermore, a di-nucleotide (dC-dC) is found to intercalate into the ZBD and to interact with its aromatic site chains, forming a DNA-base binding pocket. This newly discovered DNA-binding site was proposed to engage with ssDNA and thereby controlling access to the active centre. It was assumed that only flexible

substrate proteins - not globular proteins - could gain access and that substrates of SPRTN might require unfolding before their proteolysis.

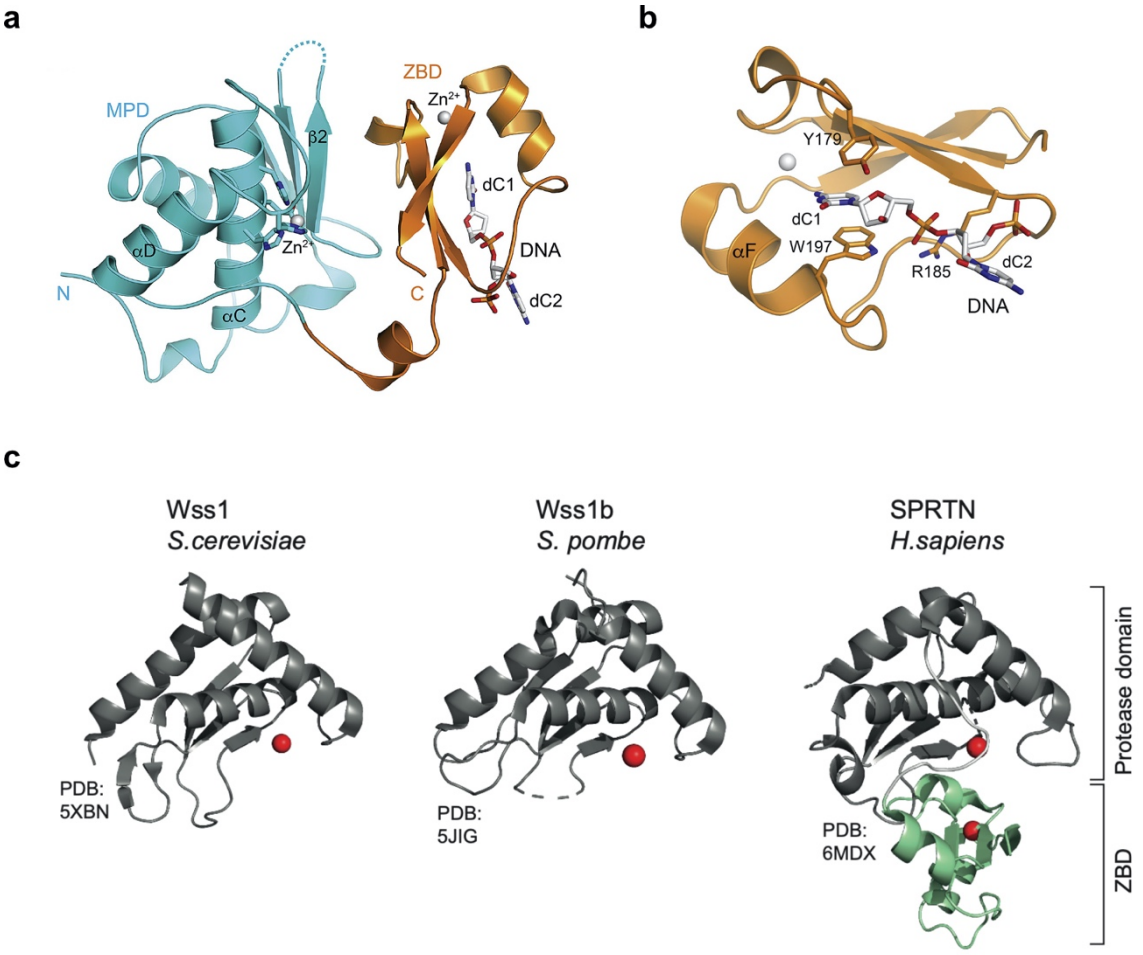


Figure 6: Protein structures from Wss1 and SPRTN protease domains. a) Crystal structure of the human SPRTN metalloprotease domain (MPD, blue) and zinc-binding domain (ZBD, orange). Both domains are coordinating a Zn^{2+} ion. The ZBD crystallized with a 5 nt oligo dC leading to the assumption that this domain contributes to ssDNA binding. (Figure adapted from (Li et al., 2019)) b) Close-up view of the ZBD in complex with a nucleotide. The DNA is stabilized by aromatic side chains forming a base stacking pocket as well as electrostatic interactions of positively charged amino acids with the negative phosphate backbone. (Figure adapted from (Li et al., 2019)) c) Cartoon representation of Wss1 from yeast and human SPRTN. All proteins contain the metalloprotease domain with the HEXXH active site residues and have a similar overall fold. Human SPRTN additionally contains the ZBD, which is missing in Wss1 proteins. (Figure from (Reinking et al., 2020a))

By comparing Wss1 and SPRTN protease structures it becomes clear that their overall architectures are very similar (Figure 6c). However, the ZBD that was found in SPRTN is missing in Wss1 and the catalytic core of Wss1 is completely solvent exposed. Since the ZBD is speculated to restrict proteolysis of substrate proteins by controlling access to the active centre, another regulatory mechanism likely regulates Wss1. Since the only structural information include the protease domains or the ZBD, it is not entirely understood how the C-terminal tails contribute to the proteases´ function.

The C-terminal tail of Wss1 contains domains for the interaction with the segregase Cdc48 (SHP and VIM motives) as well as SUMO-interacting motives (SIMs) (Stingele et al., 2014). Cdc48 is a chaperon-like enzyme, which interacts with ubiquitylated or SUMOylated proteins to segregate them from higher complexes (Bergink et al., 2013; Bodnar and Rapoport, 2017; Twomey et al., 2019) Whereas SUMO binding is not strictly required for *in vivo* and *in vitro* activity of Wss1, the binding to Cdc48 is essential *in vivo* for cell growth and DPC resistance (Stingele et al., 2014). SPRTN´s C-terminal tail contains domains for various protein-protein interactions: a SHP box for ATPase p97 interaction (homolog to Cdc48), a PIP-box for the interaction with PCNA and a ubiquitin-binding zinc finger (UBZ) domain (Stingele et al., 2016). Various studies assigned different roles to these domains for SPRTN´s function, but their exact role and how they contribute to DPC-resistance remains mostly unclear:

Initial studies discovered a role of SPRTN in regulation of TLS based on its C-terminal protein binding motives (Centore et al., 2012; Ghosal et al., 2012; Juhasz et al., 2012; Machida et al., 2012; Morocz et al., 2017). It was shown that SPRTN plays an important role in cellular resistance to UV radiation by interacting with PCNA. SPRTN was found to localize to sites of UV-induced DNA damage and to recognize ubiquitylated PCNA, which depends on its PIP-box and UBZ domain (Centore et al., 2012). Additionally SPRTN was shown to interact with p97 via its SHP box and to recruit p97 to sites of DNA damage (Davis et al., 2012; Mosbech et al., 2012). In this context it was assumed that this might facilitate removal of TLS polymerases from PCNA to prevent mutagenesis.

A recent study proposed a joint role of SPRTN, p97 and TEX264 in facilitating the repair of TOP1ccs (Fielden et al., 2020). TEX264 is thought to act as an adaptor recruiting the p97-SPRTN complex to TOP1ccs and thereby promoting their cleavage. p97 may unfold TOP1 to allow degradation by SPRTN. However, because TEX264 is not involved in the repair of a majority of DPCs, this pathway seems to be specific for TOP1ccs.

On the contrary to the above-mentioned observations, other evidence suggests that the C-terminal tail is largely dispensable for SPRTN activity. First of all, patients who suffer from RJALS, which contain truncations in the C-terminus (SPRTN-ΔC), are viable whereas a complete *SPRTN* knockout in cells is lethal (Lessel et al., 2014; Stingele et al., 2016). Compared to the WT protein, SPRTN-ΔC is overexpressed in cells and in addition mislocalized (Lessel et al., 2014; Stingele et al., 2016). However, even though the domains for protein-protein interactions are missing, SPRTN remains functional. It partially complements formaldehyde-induced sensitivity upon SPRTN depletion and is able to be recruited to chromatin (Stingele et al., 2016).

Another study using MEF cells confirmed the hypothesis of the C-terminal tail being dispensable for SPRTN activity. The authors analysed the cell cycle state in SPRTN-depleted cells and determined the contribution of the C-terminal domains in cell cycle progression (Maskey et al., 2014). The total loss of SPRTN leads to cell cycle arrest and cells accumulate in S/G2 phase. This cell cycle defect can be rescued by expressing the WT protein, but not by expressing the catalytically inactive EQ variant. However, mutants lacking the interaction with p97, PCNA or ubiquitin are able to rescue the cell cycle defect. Again, it seems that the domains within the C-terminal tail are dispensable for SPRTN´s essential function.

Nevertheless, recent evidence suggests a crucial role of the UBZ domain in DPC removal during replication (Larsen et al., 2019). A study in *X.laevis* egg extracts tested the importance of the protein-interaction domains for replication of a DPC-containing plasmid. SPRTN was depleted from the extract and recombinant SPRTN including mutations within the domains was readded to the extract (SHP*, PIP* and UBZ*) (Larsen et al., 2019). While extract with a total depletion of SPRTN lost the ability to degrade DPCs, the addition of the mutants SPRTN-SHP* and SPRTN-PIP* could rescue this effect. This leads to the assumption that p97 and PCNA interactions are dispensable for SPRTN function in the frog egg extract. The re-addition of SPRTN-UBZ* however is not able to rescue depletion of SPRTN WT. A mutant in which the C-terminal UBZ domain was deleted confirms this effect and is not able to restore DPC proteolysis either. These results show that DPC proteolysis in *X.laevis* depends on SPRTN´s UBZ domain, but not on interaction with p97 or PCNA. How the UBZ domain contributes to DPC repair remains, however, unresolved.

1.2.3 Other DPC Proteases

In addition to Wss1/SPRTN, other proteases have been identified that are involved in DPC repair (Figure 7). The protease *germ cell nuclear antigen* (GCNA) or *acid-repeat containing protein* (ACRC) is structurally similar to SPRTN (Carmell et al., 2016). GCNA contains a SprT-like metalloprotease domain and an intrinsically disordered region (IDR) with SUMO interacting motives and is primarily expressed in germ cells (Bhargava et al., 2020; Carmell et al., 2016). Before the discovery of GCNA as a DPC repair protease, an antibody against GCNA was used to label and distinguish reproductive from somatic cells in mice and rats for two decades (Carmell et al., 2016). However, almost nothing was known about this protein and only recently it was shown that GCNA might play a role in the repair of certain DPCs.

In mice and worms GCNA interacts with TOP2 and it was proposed that GCNA promotes resolution of TOP2 DPCs in germlines and early embryos (Dokshin et al., 2020). In *C.elegans* GCNA and TOP2 colocalize during mitosis and *gcna-* mutants are sensitive to TOP2 poisons. In leptotene spermatocytes mutations within GCNA result in chromatin abnormalities and display premature chromatin condensation. Interestingly, compared to the SPRTN *C.elegans* homolog Dvc-1*,* which is primarily expressed during S-phase, GCNA expression reaches its peak in G2/M-phase. When analysing protein localization in worms, the two proteins exhibit complementary localization dynamics with Dvc-1 being enriched when GCNA is absent (Dokshin et al., 2020). These observations could indicate that GCNA and Dvc-1 function in parallel pathways to promote genome integrity. Due to the expression and localization differences, it can be speculated that SPRTN acts mainly in S-phase during DNA replication, whereas GCNA functions during mitosis. This would mean that DPC repair can differ within the cell cycle phase (S- vs. M-phase) as well as cell type (germline vs. somatic cells).

Another class of proteases was recently reported to contribute to DPC repair (Serbyn et al., 2020; Svoboda et al., 2019). This class contains a retroviral aspartic protease domain (Figure 7). Its members, the DNA-damage inducible proteins (DDI1/DDI2 in humans and Ddi1 in yeast) were reported to be ubiquitin shuttling factors, which shuttle polyubiquitylated proteins to the proteasome for degradation (Nowicka et al., 2015). In this context, it was shown that DDI1 and DDI2 are important for the recovery of stalled replication forks upon induction of replication stress. By interacting with the proteasome and ubiquitylated

Replication Termination Factor 2 (RTF2) they remove this replisome component from stalled forks (Kottemann et al., 2018). RTF2 removal promotes the restart of the replication fork and subsequent cell cycle progression. However, the removal was independent of the proteolytic activity of DDIs itself, but rather depended on their role as proteasome shuttles (Kottemann et al., 2018).

Recent evidence revealed an additional role of yeast Ddi1 in DPC repair, which depends on its proteolytic activity. Ddi1 provides resistance to HU sensitivity in yeast independent of its role in proteasomal shuttling (Svoboda et al., 2019). A contemporaneous study also conducted in yeast revealed Ddi1 as a new DPC protease and its participation in TOP1-like DPC removal. This fully required the catalytic activity of the protease domain *in vivo* (Serbyn et al., 2020). The authors showed that yeast Wss1 and Ddi1 have redundant function in the repair of TOP1cc-like DPCs and that Ddi1 function acts independent of the 26S proteasome. However, in neither of the two studies *in vitro* protease activity of Ddi1 could be shown. This left its direct role as a protease, which might act exclusively in DPC repair, unanswered.

Finally, *Family with sequence similarity 111 member A* (FAM111A) was proposed to act as a DPC protease as well and to colocalize with PCNA at replication forks (Alabert et al., 2014; Kojima et al., 2020). This serine protease was initially shown to play a role in restricting viral replication and the role in DPC repair was only recently discovered (Fine et al., 2012; Kojima et al., 2020). *FAM111A* knockout cells are specifically sensitive to TOP1 poisons but not towards other crosslinking agents such as etoposides, 5-aza-dC, formaldehyde or cisplatin, indicating a specialized role in the repair of those crosslinks (Kojima et al., 2020). In addition, *FAM111A* knockout cells are hypersensitive to PARP inhibitors causing PARP1 trapping. For its activity *in vivo* the protease domain, DNA- and PCNA-binding domains are required (Figure 7). FAM111A, like SPRTN, undergoes autocleavage *in vivo* and *in vitro* (Hoffmann et al., 2020). Interestingly, compared to SPRTN, FAM111A also demonstrated a role in releasing proteins that are just trapped on the DNA (e.g. PARP1) (Kojima et al., 2020).

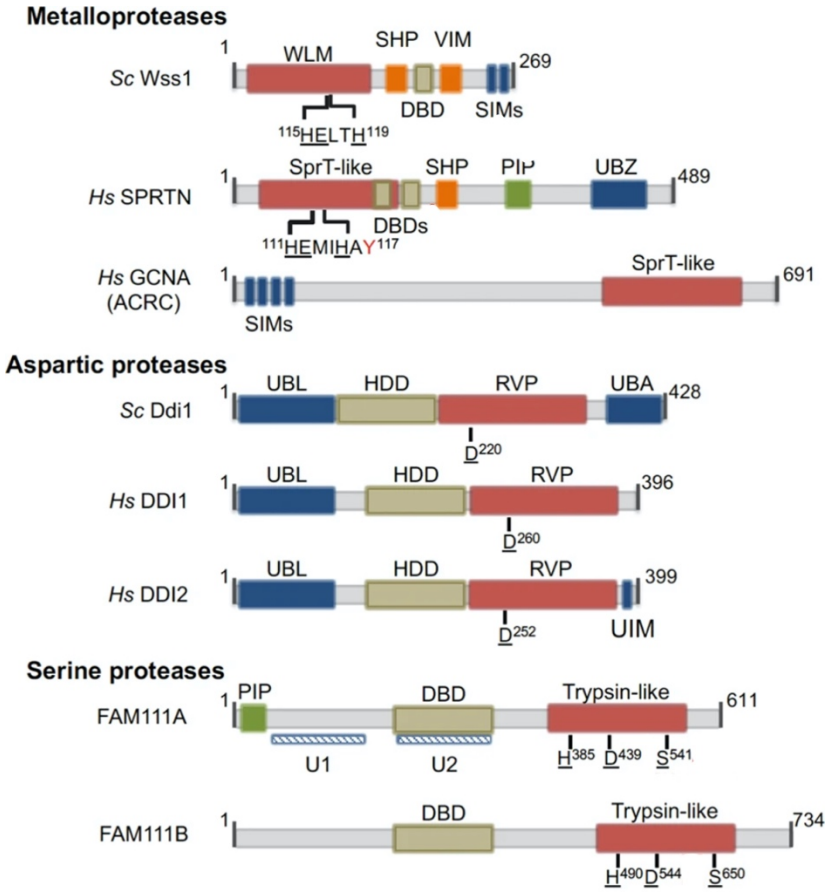


Figure 7: Domain organization of different DPC proteases. In addition to the metalloproteases Wss1 and SPRTN, other enzymes were identified, which were shown to play a role in the repair of certain DPCs. Protease domains are colored in red, p97/Cdc48 interacting motives in orange (SHP/VIM), PCNA interacting motives in green (PIP), DNA-binding sites in ochre and SUMO or ubiquitin/proteasome binding sites in blue. Catalytic residues of the protease domains are underlined. (Figure adapted from (Ruggiano and Ramadan, 2021))

In summary, the recent discoveries of additionally DPC proteases underline the complexity of DPCs and their repair. By acting in different cell cycle phases, tissues or by targeting different types of substrates a large variety of DPCs can be degraded. However, it is not entirely clear how these proteases are regulated and how they interplay.

1.3 Canonical Repair Pathways Involved in DPC Repair

Since DPCs are very diverse, not only specialized proteases but also other canonical repair pathways were found to contribute to DPC repair. Here, the repair pathway choice often depends on the type of crosslink that is targeted. The following paragraph will discuss the major roles of canonical repair pathways in DPC repair.

1.3.1 Tyrosyl-DNA Phosphodiesterase 1 and 2

Since TOP1ccs and TOP2ccs are highly toxic DNA lesions that can result in DSBs, a lot of research has been put into understanding their repair. Repair of both DPCs requires an interplay of multiple DNA repair components and involves specialized enzymes, which target the phosphotyrosyl bond between the protein and DNA (Pommier et al., 2014). These specialized enzymes are called Tyrosyl-DNA phosphodiesterase 1 and 2 (TDP1 and TDP2) and target TOP1ccs and TOP2ccs respectively (Figure 8) (Pommier et al., 2014; Stingele et al., 2017).

TOP1 is crosslinked via a tyrosine side chain, which forms a phosphotyrosyl bond with the 3´end of the DNA as described above (Interthal and Champoux, 2011). TDP1 is responsible for the hydrolysis of this bond but needs the help of other proteins that preprocess the DPC beforehand and close the gap after hydrolysis (Debéthune et al., 2002). TDP1 is not able to act on full-length TOP1 and its catalytic activity was shown to be restricted to smaller TOPI-derived peptides up to 108 amino acids whereas bigger fragments up to 333 amino acids are not cleaved efficiently (Interthal and Champoux, 2011). Several studies indicated an important role of ubiquitin-dependent proteasomal degradation, which allows subsequent access of DNA repair enzymes including TDP1 (Desai et al., 1997; Interthal and Champoux, 2011; Lin et al., 2008). After protein trimming, TDP1 hydrolyses the linkage, which results in a 3´phosphate and a 5´OH end (Yang et al., 1996). To allow successful religation of the break, polynucleotide kinase phosphatase (PNKP), an enzymatic member of the single-strand break repair (SSBR) complex, needs to further process these ends (Debéthune et al., 2002; Plo et al., 2003). PNKP exhibits a 5´-DNA kinase and 3´-phosphatase activity and generates a 3´OH and a 5´phosphate DNA end, which can then be sealed by the single-stand break repair machinery (Caldecott, 2008).

TDP1 stability and recruitment is regulated by posttranslational modifications like phosphorylation, SUMOylation and PARylation. Phosphorylation of serine 81 by ATM and DNA-PK kinases increases the stability and promotes the recruitment of TDP1 to damaged DNA sites (Chiang et al., 2010; Das et al., 2009). Furthermore this phosphorylation promotes interaction with other proteins of the SSBR pathway (e.g. XRCC1 and Ligase IIIα) (Chiang et al., 2010). SUMOylation at lysine 111 additionally promotes its accumulation to sites of DNA damage (Hudson et al., 2012). Moreover, PARP1 directly interacts with the N-terminal domain of TDP1 and also PARylation stabilizes and recruits TDP1 to the DNA damage (Das et al., 2014).

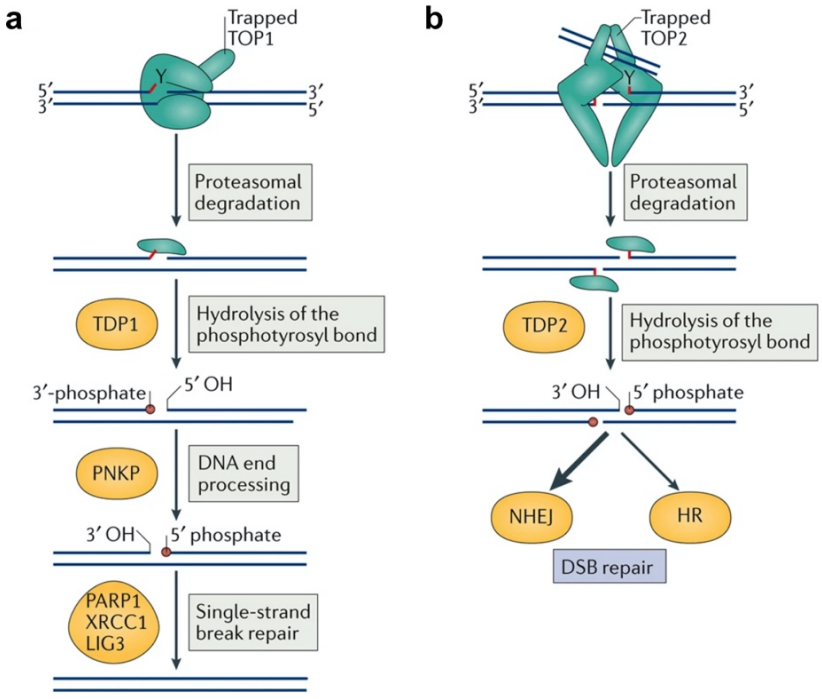


Figure 8: Repair of TOP1 and TOP2 cleavage complexes by the tyrosyl-DNA phosphodiesterase 1 and 2. a) TOP1 is crosslinked to a 3´end of a SSB. The protein component of TOP1 is proteolytically degraded by the proteasome, which enables TDP1 to access and hydrolyze the crosslink. After hydrolysis a phosphate group is attached to the 3´DNA end whereas the 5´end contains a free OH group. Polynucleotide kinase phosphatase (PNKP) contains a 5´kinase and 3´phosphatase activity and processes the DNA ends for the subsequent repair by the SSB repair machinery. PARP1, XRCC1 and LIG3 can complete the repair by strand ligation. b) The homodimer TOP2 cleaves both strands of the DNA. Entrapment results in a DSB with TOP2 attached to the 5´DNA ends by its active site tyrosine. TDP2 requires the crosslinked protein to be degraded, which might be performed by the proteasome. The accessible crosslink can then be hydrolyzed by TDP2 generating 3´OH and 5´phosphate DNA ends, which can be repaired by NHEJ or HR. (Figure adapted from (Stingele et al., 2017))

TOP2ccs can be repaired by the enzymatic activity of TDP2 that hydrolyses the linkage between the tyrosine side chain and the 5´end of DNA (Pommier et al., 2014; Pommier et al., 2016). TOP2 is a homodimer and cleaves both DNA strands, which generates a DSB. TDP2 activity generates 5´phosphate and 3´OH ends, which can be directly religated by the NHEJ machinery (Gomez-Herreros et al., 2013). As in the case of TDP1, the activity of TDP2 seems to require the prior degradation or denaturation of the protein component (Gao et al., 2014). Several studies indicated that the proteasome is the responsible protease exposing the crosslink for repair (Mao et al., 2001; Zhang et al., 2006). However, the molecular basis for regulation and coordination of TOP2cc repair by the proteasome is still not fully understood.

Other evidence for a proteasomal-independent TOP2cc repair mechanism by TDP2 involves the SUMO ligase ZATT (ZNF451) (Schellenberg et al., 2017). TOP2 is modified with SUMO during mitosis and after treatment with topoisomerase inhibitors (Agostinho et al., 2008; Azuma et al., 2003; Ryu et al., 2010). The SUMO ligase ZATT was recently shown to directly bind to TOP2 and to be recruited to chromatin after TOP2 poisoning (Schellenberg et al., 2017). TDP2 interacts with SUMOylated TOP2, which depends on ZATT. This further leads to TDP2 recruitment to chromatin where its catalytic activity is required to cleave intact SUMOylated TOP2.

TDP2 is not only binding to SUMOylated but also to polyubiquitinated proteins (Schellenberg et al., 2020). TDP2 was shown to bind multiple forms of ubiquitin via its N-terminal domain and this binding is important for its ability to repair TOP2-mediated DNA damages (Rao et al., 2016). It was hypothesized that TDP2 is recruited to TOP2ccs through interaction with ubiquitinated proteins but the role of this ubiquitin signalling pathways remains poorly defined. Recent evidence suggests that TDP2 interaction with polyubiquitin serves a regulatory function and influences TDP2 hydrolase activity (Schellenberg et al., 2020). The authors speculated that binding to ubiquitin induces a conformational change, which stabilizes the enzyme in an active conformation with increased activity and further promotes repair of TOP2 generated DNA damage.

Recently, the impact of SUMO and ubiquitin modifications on the repair of TOP1 and TOP2 DPCs was determined in human and yeast cells (Sun et al., 2020). The authors discovered a novel repair mechanism, which removes both, TOP1- and TOP2-DPCs, independently of

DNA replication, transcription or DDR. This pathway involves sequential SUMOylation and ubiquitylation, which primes the DPCs for proteasomal degradation.

In a first step, TOP1- and TOP2-DPCs undergo rapid SUMOylation catalysed by the SUMO E3 ligase PIAS4 (protein inhibitor of activated STAT protein 4) (Figure 9) (Sun et al., 2020). The enzyme is first attaching SUMO-2/3 polymeric chains to the DPCs followed by SUMO-1 capping. This reaction depends on the enzyme´s N-terminal DNA-binding SAP (SAF-A/B, Acinus and PIAS) domain, which recruits the protein to DNA. Following SUMOylation, the SUMO-targeting ubiquitin ligase (STUbL) RNF4 (RING finger protein 4) ubiquitylates the SUMOylated DPC. RNF4 itself is suggested to be recruited to the DPC via its SUMO-interacting motive (SIM). It was further proposed that key proteasome components can then interact with the DPC. PSMD14 (proteasome non-adenosine triphosphatase regulatory subunit 14) deubiquitylates the DPC before passing it to a chymotrypsin-like proteolytic subunit of the 20S core particle PSMB5 (proteasome subunit beta type-5) for degradation. However, even though this SUMO-Ub-proteasome pathway was shown to be involved in TOP1 and 2 crosslink removal, its inactivation did not completely abolish DPC repair either. It can therefore be assumed that other repair pathways exist that target TOP-DPCs in parallel.

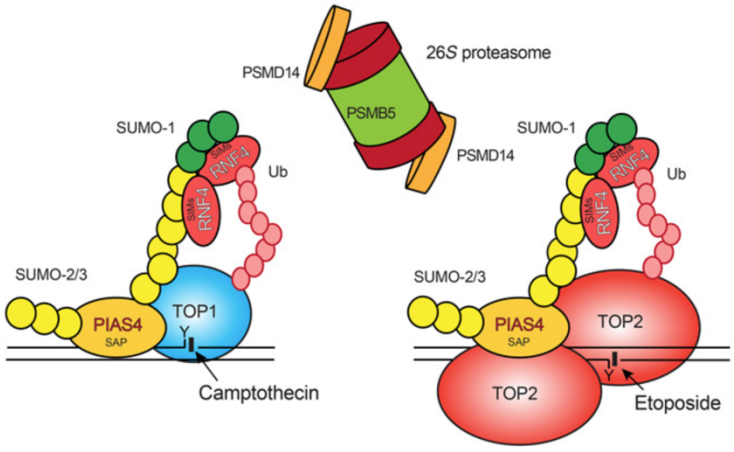


Figure 9: SUMO- and ubiquitin-dependent removal of TOP1- and TOP2-DPCs. The DPCs are sequentially modified with SUMO-2/3 and SUMO-1 by the SUMO ligase PIAS4, which depends on its DNA-binding SAP domain. Following SUMOylation, the SUMO-targeting ubiquitin ligase (STUbL) RNF4 is recruited via its SIM motive and ubiquitylates the DPCs. Polyubiquitylation primes the substrates for proteasomal degradation. (Figure from (Sun et al., 2020))

1.3.2 The MRN complex in DPC repair

DPCs at DSBs have to be removed in order to allow repair of the break by NHEJ or HR. These DPCs can be targeted by the MRN complex (MRX in yeast), which consists of Mre11, Rad50 and Nbs1 (Xrs2) (Rupnik et al., 2008). Normally, the MRN complex contributes to HR-dependent break repair by generating 3´ssDNA overhangs and by recruiting other nucleases to sites of DNA damage (Symington, 2014). However, multiple studies revealed a role of the complex in preprocessing DNA ends containing protein adducts to generate "clean" DSB ends, which can be ligated (Aparicio et al., 2016; Hoa et al., 2016; Lee et al., 2012). Examples of these DPCs include TOP2ccs as well as trapped SPO11 adducts (Hoa et al., 2016; Neale et al., 2005). SPO11 is required for DSB formation during meiotic recombination and exhibits a topoisomerase-like activity, forming a covalent adduct with the 5´DNA end (Mimitou et al., 2017). *S.cerevisiae* strains deficient in Rad50 were shown to accumulate SPO11 DPCs (Keeney et al., 1997). In addition, the MRX complex in yeast was shown to act on the DNA component and to release SPO11 bound oligonucleotides (Neale et al., 2005). After generation of 3´ssDNA overhangs, these DSBs could be repaired by HR.

Moreover, it was shown that the MRN complex can use the nuclease activity of Mre11 to remove TOP2ccs (Hoa et al., 2016). The authors analysed sensitivity towards TOP2 poisons in Mre11-deficient chicken DT40 and human lymphoblast cells. These cells are sensitive towards TOP2 poison and display increased chromosomal DSBs. Even in the absence of exogenous poison, TOP2ccs accumulate upon Mre11 deficiency. This effect is reversed by the overexpression of TDP2, indicating an important role of the MRN complex in repairing these lesions.

In vitro reconstitution of the MRN complex activity for DNA ends containing protein adducts was performed by several laboratories (Cannavo and Cejka, 2014; Deshpande et al., 2016; Saathoff et al., 2018). Deshpande *et al.* showed that the human MRN complex specifically processes dsDNA ends, which are blocked by protein adducts (Deshpande et al., 2016). During end processing, MRN exhibits multiple activities and catalyses sequential endo- and exonucleolytic cleavage reactions (Figure 10):

In a first step, the DNA is endonucleolytically cleaved in close proximity to the 5´protein adduct. This cleavage creates a nick on the 5´ or 3´strand in close proximity to the adduct. In a second step, the DNA is digested in a 3´to 5´ direction by the exonuclease activity of

MRN. This digestion is directed away from the nick and either towards the blocked or away from the blocked end. Third, a last MRN endonucleolytic cleavage event is performed on the opposite strand of the initial nick. This results in clipping-off of the protein adduct and the generation of a clean DSB end, which can be further processed by HR or NHEJ. This specific activity of MRN is tightly controlled by Nbs1, ATP and CtIP. Nbs1 is a key regulator in this process, which stimulates endo- and exonucleolytic cleavage of these "dirty" ends. In contrast, Nbs1 blocks resection of clean, open ends and inhibits their 3´ to 5´exonucleolytic cleavage. Additionally, phosphorylated CtIP stimulates endonucleolytic activity of MRN at blocked ends, thereby contributing to their repair (Aparicio et al., 2016; Deshpande et al., 2016).

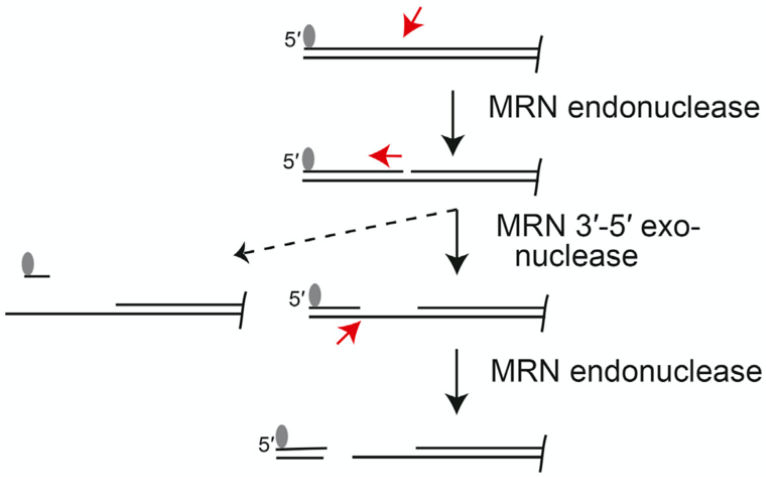


Figure 10: Nuclease-dependent removal of a protein adduct attached to the 5´end of a DSB. In the presence of a protein-adduct on the DSB, MRN performs sequential cleavage events in order to generate DNA ends that can be used in HR. First, an endonucleolytic activity cleaves the DNA in proximity to the 5´adduct generating a DNA nick. Second, exonucleolytic activity resects the DNA in a 3´ to 5´direction. A last endonucleolytic cut releases the DNA containing the protein adduct. A "clean" DSB end is generated, which can be further utilized by downstream repair machineries. MRN activities during this process strictly depend on Nbs1, ATP and the presence of a protein adduct at the 5´end. (Figure from (Deshpande et al., 2016))

The complex consisting of the nuclease Mre11 and ATPase Rad50 is evolutionary highly conserved and processes abnormal terminal DNA structures, hairpins and DSB among different species (Stracker and Petrini, 2011). Compared to the eukaryotic complex, which contains Nbs1 as a third subunit, the bacterial homolog SbcCD only consists of the ATPase

(SbcC) and nuclease (SbcD) (Connelly and Leach, 1996; Stracker and Petrini, 2011). This complex was also shown to contain a 3´- 5´exonuclease activity and to cleave the DNA near a protein-blocked DNA end as well as 5´of the loop of a hairpin structure (Connelly et al., 2003; Connelly and de Leau, 1999). In addition, SbcCD contains an instrinsic endonuclease activity and can cut both DNA strands, which depends on ATP hydrolysis (Lim et al., 2015). The ATP-dependent nuclease activities of the SbcCD complex were further studied in detail by Saathoff *et al.* (Saathoff et al., 2018). The authors proposed the following model for SbcCD´s endo- and exonuclease activities (Figure 11): if a protein blocks the DNA end, the endonucleolytic activity of SbcCD is stimulated. The DNA is locally melted and a bubble is generated, which allows DNA incision. This process requires repeated ATP hydrolysis. SbcCD is then cleaving the DNA on opposing strands by hydrolysing the phosphodiester bond either at the 3´or 5´side and therefore generating chemically distinct DNA ends. After the cleavage event, phosphorylated 5´ and 3´ends remain. For the exonuclease activity at DNA ends it is proposed that the SbcCD complex binds to the DNA in a reversed orientation compared to the orientation to a protein block and hydroxylated 3´and 5´termini are generated (Saathoff et al., 2018).

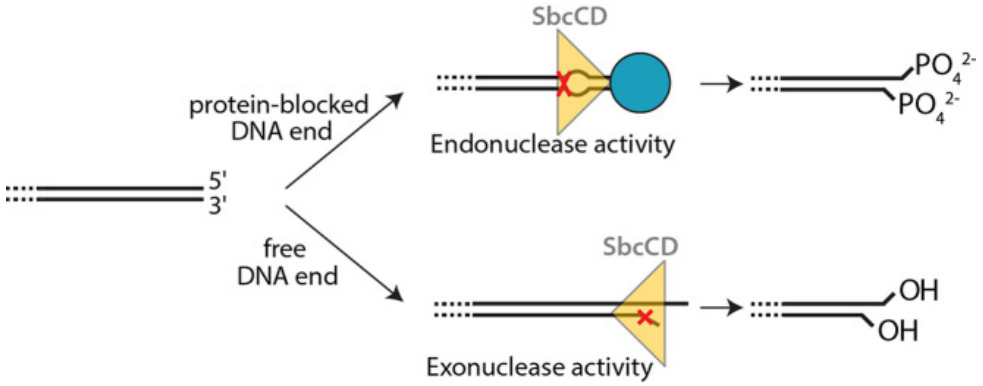


Figure 11: Endo- and exonuclease activity of the SbcCD complex. The endonuclease activity of SbcCD complex (shown in yellow) is stimulated by a protein adduct at the DNA end (blue sphere). After DNA melting and bubble formation, the DNA is cleaved on both strands, leaving phosphorylated 5´and 3´ends. For the exonuclease activity the complex is oriented in the opposite direction and hydroxylated ends are generated. (Figure from (Saathoff et al., 2018))

However, structural insights in how the complex senses and processes diverse DNA ends was still missing. The recent discovery of the cryo-electron microscopy (cryo-EM) structure of the SbcCD complex shed light into this mystery (Kashammer et al., 2019). The complex

was shown in its resting state bound to ATPγS and in a recognition and cutting state bound to a DNA break. In the resting state the nuclease of Mre11 is blocked by ATP-Rad50. In the presence of DNA, the nucleotide binding domain (NBD) together with the coiled coil domain (CC) domain of Rad50 form a clamp around the dsDNA. Mre11 can then bind the DNA end and can form a DNA cutting channel for its nuclease activity. The structural and biochemical contribution of this study reveal how the complex can recognize different DNA termini in order to allow subsequent repair through the DSB repair machineries.

1.3.3 Nucleotide Excision Repair in DPC Repair

NER is another repair pathway that plays a role in the repair of specific DPCs in different organisms. The role of NER enzymes in DPC repair was analysed in a bacterial system using *in vivo* and *in vitro* approaches (Nakano et al., 2007). The UvrABC complex is involved in NER in *E. coli* and cleaves the DNA close to the damaged nucleotide. Repair-deficient *uvrA* mutants were shown to be sensitive towards formaldehyde. Further *in vitro* analysis investigated the role of the protein size on UvrABC nuclease activity. The excision of the DPC in a 60mer oligonucleotide is restricted to proteins smaller than 16 kDa, which probably depends on steric hindrance of the UvrABC to load onto the DNA. This size-dependent repair was also overserved *in vivo.* The repair of DPCs was analysed over time and was revealed to be restricted to proteins smaller than 14 kDa. The combination of *in vivo* and *in vitro* results in bacteria suggest that NER is involved in the repair of smaller DPCs, while the authors proposed that larger DPCs are processed by HR.
Studies in mammalian cells confirmed that NER is restricted to smaller proteins or peptide adducts (Baker et al., 2007; Nakano et al., 2009; Reardon et al., 2006). The same research group that performed NER analysis in bacteria assessed DPC repair by NER in mammalian cells (Nakano et al., 2009). The authors showed that NER is slightly differently involved in DPC repair in mammalian cells compared to *E. coli.* Compared to NER in bacteria, NER in mammalian cells could not repair chromosomal DPCs *in vivo* and could only process very small formaldehyde-induced peptide adducts less than 7 kDa. Moreover, nuclease activity was tested against a 150mer oligonucleotide crosslinked to a variety of peptides and proteins of different sizes. Incision for proteins for up to 11 kDa is visible. Both, the bacterial and the mammalian system show the highest efficiency towards very small substrates of 1.6 kDa. For mammalian cells, the authors proposed that HR constitutes the major mechanism to

conquer DPC lesions and that NER is not an alternative pathway for the removal of larger DPCs. However, it is an important pathway for the repair of small crosslinked protein adducts induced by formaldehyde.

Another research group investigated the role of NER in mammalian cell-free extract for the repair of a crosslinked DNA methyltransferase to an oligonucleotide (Baker et al., 2007). After protein digestion of the full-length methyltransferase by a protease, NER seems to participate in DPC removal (Baker et al., 2007). The authors proposed a role for the proteasome to degrade the protein component of the DPC, making it a suitable substrate for NER. However, this data stands in conflict with results obtained by Nakano *et al.,* which contradict the involvement of the proteasome (Nakano et al., 2009).

A more recent approach monitored the repair of different DPC plasmid substrates in mammalian cells using a PCR-based assay (Chesner and Campbell, 2018). The results showed that NER is involved in the repair of DPCs greater than 10 kDa and even up to 38 kDa. However, smaller peptide adducts (4 kDa) are repaired twice as efficiently as larger adducts. It was hypothesized that this decreased activity towards larger substrates originates from steric hindrance of repair proteins to recognize and bind the lesion. Here again, the authors speculated that the protein component of the DPC is degraded prior to removal by NER. Still, the question which proteases are involved in this process remains under debate.

1.3.4 Fanconi Anaemia Pathway in DPC Repair

The Fanconi anaemia (FA) pathway is well known to be involved in the repair of ICLs, where two DNA strand are covalently attached (Ceccaldi et al., 2016). Cells that are deficient in this repair pathway are sensitive towards reactive aldehydes (Ridpath et al., 2007). However, acetaldehyde and formaldehyde are both capable of introducing DNA-DNA crosslinks as well as DNA-protein crosslinks (Duxin and Walter, 2015). Therefore, it can be speculated, that the sensitivity towards aldehydes in FA-deficient cells additionally comes from unsuccessful DPC repair.

On one side, evidence supporting a role in DPC repair was obtained by the observation that FA proteins seem to be important for the repair of trapped DNMT1 by the inhibitor 5-aza-dC (Orta et al., 2013). Unrepaired DNMT lesions result in replication fork collapse and DSBs, which trigger the HR pathway. Cells depleted of FANCG, a protein participating in the FA pathway, are sensitive towards 5-aza-dC and unable to activate HR-mediated repair

of the DNA lesion. This results in increased replication fork collapse and accumulation of DSBs. Even though this evidence suggests a contribution of the FA pathway in DPC repair, other data support a model in which FA and DPC repair are distinct DNA repair mechanisms:

First, the depletion of FANCI-FANCD2 in *X.laevis* egg extract effects only replication-coupled ICL repair but not the repair of DPCs (Duxin et al., 2014). Second, impairment of the FA pathway by depleting FANCD2 in synchronized worm larvae did not indicate increased sensitivity towards formaldehyde (Stingele et al., 2016). The lack of sensitivity towards formaldehyde in these worms could be explained by the fact that synchronized L1 larvae arrest in G1/S transition and do not undergo replication. Since ICL repair is replication-dependent, the depletion did not have any major negative effect, not even on ICL repair. Moreover, by comparing cisplatin-sensitivity of SPRTN-, FANCD2- or a double worm mutant it was proposed that FANCD2 and SPRTN act in independent repair pathways and cannot complement each other (Stingele et al., 2016). Besides, MEF cells lacking FANCD2 are able to repair formaldehyde-induced DPCs whereas cells lacking SPRTN cannot. Additionally, depletion of FANCD2 did not result in accumulation of DPCs compared to control cells. Last, siRNA-mediated depletion of FANCD2 in HeLa cells results in increased sensitivity towards formaldehyde but not to camptothecin (Vaz et al., 2016). Thus, it seems that FA can participate in the repair of ICLs but is not repairing the specific DPCs induced by camptothecin.

In summary, in order to fully resolve the conflicting results on the contribution of the FA pathway on DPC repair, it still requires further investigations.

1.3.5 Replication-Coupled DPC Repair

Even though some evidence exists that DPCs can be repaired by NER, HR or the FA pathway, the direct contributions of these pathways are not entirely clear and conflicting data consists about the different mechanisms as discussed above.

Some studies revealed the existence of a replication-coupled DPC repair pathway (Duxin et al., 2014). This pathway involves the proteolytic degradation of the DPC´s protein component followed by replication across the remaining peptide adduct by employing TLS. Interestingly, the repair occurs without the formation of DSBs and might therefore be a preferred choice to HR, which carries the risk of chromosomal rearrangements.

The authors employed an *in vitro* approach using *X.laevis* egg extract in which replication of a DPC-containing plasmid can be recapitulated (Walter et al., 1998). The plasmid contains a fluorinated cytosine (5′fluoro-2′-deoxycytosine) incorporated at a recognition site (CCGG) for the DNA methyltransferase HpaII (45 kDa, Figure 12). Recognition and binding of HpaII to this site results in crosslink formation between the protein and the DNA (Chen et al., 1991). By introducing the DNA modification to either the leading or lagging strand, chemically-defined and site-specific DPCs can be generated and replication and repair of these plasmids can be monitored (Figure 12).

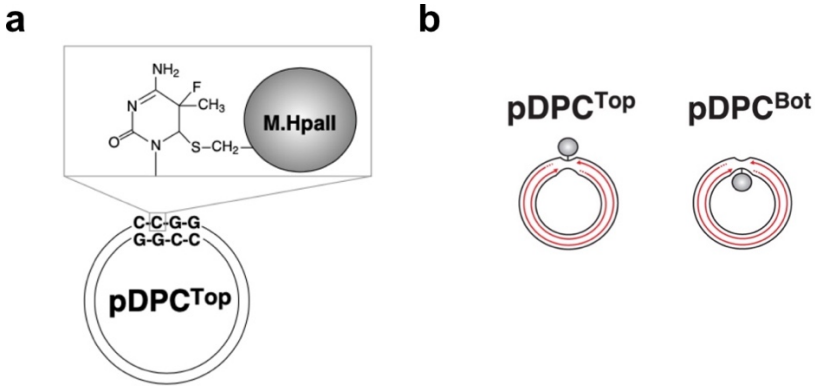


Figure 12: DPC-containing plasmid used for replication analysis in *X.laevis* egg extract. a) By introducing 5′fluoro-2′-deoxycytosine into the recognition site "CCGG" of the DNA methyltransferase HpaII (M.HpaII), the enzyme becomes covalently attached to the plasmid. b) Plasmids containing M.HpaII crosslinks in the top or bottom strand. By introducing the crosslink into the top or bottom strand of the plasmid, replication of a leading or lagging strand DPC can be analysed in the extract. (Figure adapted from (Duxin et al., 2014))

The initial study in 2014 proposed that replication-coupled DPC repair is a multistep mechanism (Duxin et al., 2014): When the eukaryotic helicase CMG (complex of CDC45, MCM2-7 and GINS) travels along the leading strand in 3′ to 5′ direction, it can encounter a DPC. This encounter was first described to lead to CMG stalling and only after proteolytic degradation of the protein component by an unknown protease, CMG was able to bypass the adduct. After subsequent polymerase approach and stalling at the remaining adduct, replication could further continue by employing TLS.

It was speculated that a short peptide adduct remains on the parental DNA strands after replication, which might be further processed by other DNA repair pathways like NER. In this initial study, inhibition of the proteasome did not inhibit DPC repair but the depletion

of ubiquitin did. It was therefore speculated that degradation was proteasome-independent and other proteases (e.g. SPRTN) might participate in the proteolytic degradation of the DPC to allow replication. However, the origin of this protease remained unknown at that time.

Two following studies shed more light into the replication-dependent DPC repair (Larsen et al., 2019; Sparks et al., 2019). The first study readdressed the question of CMG fate after encountering a DPC (Sparks et al., 2019). It was shown, contrary to the initial claims, that the CMG helicase can bypass the intact crosslink before its degradation and that this bypass is essential for efficient DPC proteolysis (Figure 13). CMG bypass of the leading-strand DPC requires the presence of ssDNA downstream of the crosslink. It was shown that an accessory helicase called regulator of telomere elongation helicase 1 (RTEL1, a 5´to 3´helicase), which probably travels with the replisome, is generating this ssDNA stretch. In addition, when the replisome encounters a DPC, this DPC is ubiquitylated by the E3 ligase TRAIP, which also contributes to CMG bypass (Larsen et al., 2019). However, how this ubiquitylation affects this process is not completely understood yet. CMG bypass does not require its unloading and it can be speculated that either the DPC or the CMG undergo conformational changes (Sparks et al., 2019). After bypass, the helicase slows down probably due to uncoupling of the replication machinery. The DPC remains in ssDNA, which leads to further ubiquitylation likely by the recently discovered E3 ubiquitin ligase RFWD3 (Gallina et al., 2021) (Figure 13). RFWD3 was shown to ubiquitylate different proteins when crosslinked to ssDNA and is essential for TLS across DNA-peptide adducts in frog extract. After CMG bypass, the DPC is degraded and TLS synthesis is able to continue replication across the lesion.

The second study analysed the proteases, which are responsible for DPC degradation and how their activity influences DPC repair (Larsen et al., 2019). Using the same cell-free system, the authors discovered that SPRTN and the proteasome both participate in replication-coupled degradation of the DPC but in two distinct and independent pathways (Figure 13). Both proteases accumulate on the DPC plasmid in a replication-dependent manner and the depletion of each of the proteases does not impair recruitment of the other. Depletion of SPRTN or inhibition of the proteasome by an MGM262 inhibitor alone results in only mild delays in DPC replication but does not prevent DPC degradation completely. When both proteases are inhibited however, the DPC is stabilized. This data shows that these

two proteases act independently from each other and that DPC repair is impaired in their simultaneous absence.

The authors further suggested two distinct recruitment and regulatory mechanisms for these proteases: proteasomal degradation requires CMG bypass, which transfers the DPC to ssDNA and further leads to it is ubiquitylation. In order to study the importance of DPC ubiquitylation, the lysine residues were chemically methylated, which prohibits ubiquitylation of HpaII. Pull-down experiments showed that this inhibition abolished proteasome recruitment to the DPC whereas the recruitment of SPRTN was only mildly affected. Additionally, degradation of the methylated crosslink was only observed in the presence of SPRTN but not the proteasome, further indicating a necessity of DPC ubiquitylation for the proteasome pathway. TRAIP was discovered as an E3 ligase that ubiquitylates the DPC upon replisome encounter as described above. However, after CMG bypass and generation of ssDNA around the DPC, the DPC is ubiquitylated probably by the E3 ligase RFWD3 independent of TRAIP (Gallina et al., 2021; Larsen et al., 2019). Therefore, it was proposed that TRAIP is mainly required for the bypass of the CMG helicase as described previously.

SPRTN is also binding to ubiquitin and it was shown that SPRTN activity strictly depends on its UBZ domain in the extract (Larsen et al., 2019). However, ubiquitylation of the DPC is not essential for SPRTN activity and SPRTN can also degrade the methylated DPC. Therefore, the authors speculated that SPRTN rather employs its UBZ domain for the recruitment to the DPC by interacting with other ubiquitylated proteins. Concerning the DPC degradation itself, it was shown that SPRTN is regulated through another mechanism. As for the proteasome, SPRTN needs the CMG helicase to bypass the intact DPC. Yet, it additionally requires nascent strand extension of the polymerase to within a few nucleotides of the DPC. Polymerase stalling at the DPC triggers SPRTN-mediated DPC degradation, which in turn allows replication by TLS polymerases across the lesion.

The presence of two different replication-coupled proteases indicates that DPC repair is a complex process and requires multiple components acting in parallel. By targeting different DPCs or responding to different signals, both proteases contribute to the repair of DPCs and thereby helping cells to cope with this DNA damage.

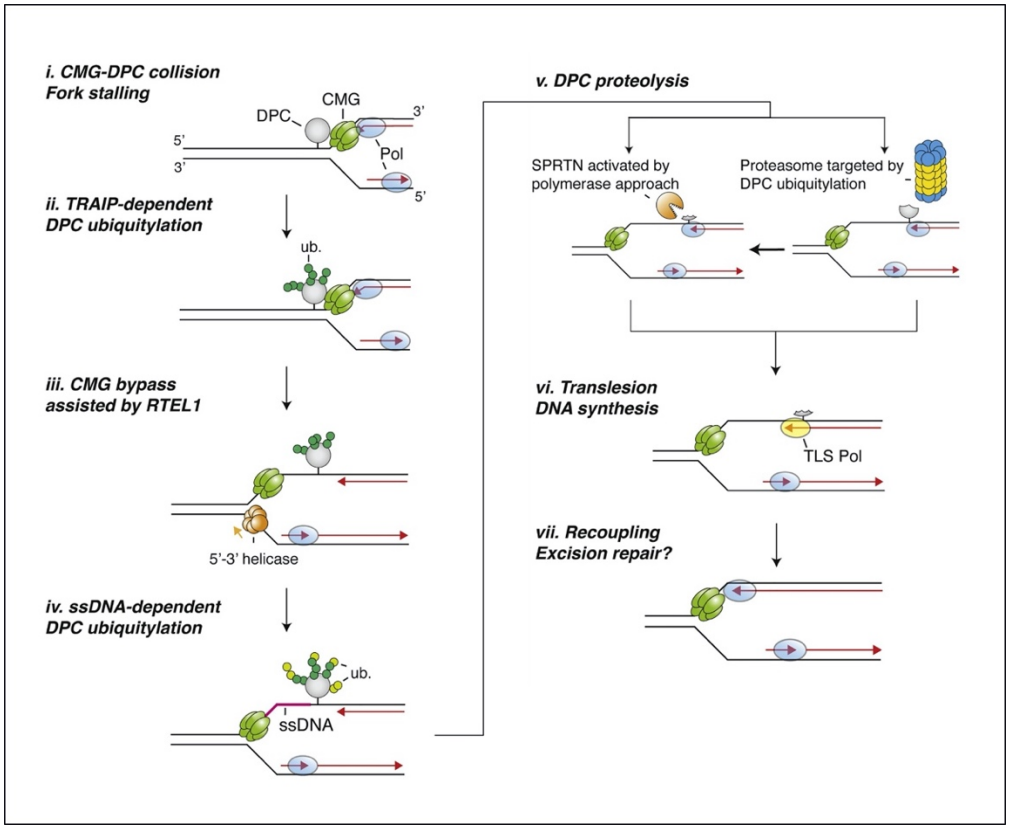


Figure 13: Replication-coupled DPC repair in *X.laevis* egg extract. When the helicase CMG encounters a DPC on the leading strand, the DPC becomes ubiquitylated by the E3 ligase TRAIP. This ubiquitylation promotes CMG bypass. Moreover, an accessory helicase RTEL1 generates ssDNA downstream of the DPC and enables the CMG to bypass the intact crosslink, which leads to uncoupling from the replication machinery and slows down the helicase. The DPC remains in ssDNA and becomes further ubiquitylated probably by the E3 ligase RFWD3. After CMG bypass, the protein component of the DPC is degraded by SPRTN or the proteasome. The proteasome is degrading ubiquitylated DPCs whereas SPRTN requires the polymerase to extend the nascent strand to within a few nucleotides apart from the DPC. After DPC proteolysis, the remaining peptide adduct can be bypassed by TLS. (Figure adapted from (Kuhbacher and Duxin, 2020))

1.4 Aim of this Study

DNA-protein crosslinks (DPCs) are highly toxic DNA lesions, which interfere with important processes such as replication and transcription. The protease SPRTN was recently discovered to play a major role in the repair of DPCs. Its essential function is demonstrated by the fact that a complete knockout of SPRTN in cells is not viable. Many studies have addressed SPRTN´s contribution in DPC repair but several key questions remain unanswered.

SPRTN only becomes active when bound to DNA and can only degrade other DNA-binding proteins. However, once activated, it is highly promiscuous towards its substrates and can target proteins independent of their structure, sequence or size. This feature is beneficial for degrading diverse DPCs, but it also carries the risk of unwanted proteolysis of surrounding chromatin proteins. This thesis addresses how SPRTN´s activity is restricted to the toxic DPC while leaving other DNA-binding proteins unharmed. Therefore, biochemical approaches are employed to characterize SPRTN´s specific activation. In detail, novel model DPC-substrates are generated, which allow *in vitro* analysis of SPRTN´s activity towards different substrates. Additionally, NMR analysis of the two DNA-binding domains is used in order to understand their differential contribution to the DNA-dependent activation.

Furthermore, this thesis attempts to reveal how the activity of SPRTN is regulated by monoubiquitylation. SPRTN is constantly mono- and deubiquitylated in a dynamic process but the effect of this modification on SPRTN´s activity is not entirely understood. In order to gain insights in how monoubiquitin effects SPRTN activity, recombinant SPRTN proteins are characterized in biochemical *in vitro* assays. The results contribute to a deeper understanding how the cellular SPRTN pool is regulated.

2 Publications

2.1 DNA Structure-Specific Cleavage of DNA-Protein Crosslinks by the SPRTN Protease

Hannah K. Reinking, Hyun-Seo Kang, Maximilian J. Götz, Hao-Yi Li, Anja Kieser, Shubo Zhao, Aleida C. Acampora, Pedro Weickert, Evelyn Fessler, Lucas T. Jae, Michael Sattler, Julian Stingele, DNA Structure-Specific Cleavage of DNA-Protein Crosslinks by the SPRTN Protease, *Molecular Cell,* Volume 80, Issue 1, 1 October 2020, Pages 102-113.e6., https://doi.org/10.1016/j.molcel.2020.08.003

Summary

This publication reveals a DNA-structure specific activity of the DPC protease SPRTN. In order to analyse SPRTN´s activity *in vitro*, novel model DPC-substrates were generated by conjugating protein G to defined positions within 30mer oligonucleotides. By employing different biochemical assays, we found that SPRTN activity is tightly coupled to the recognition of certain DNA structures. The protease only becomes activated in close proximity to disruptions within dsDNA: at a hairpin loop, ss- to dsDNA junctions or at dsDNA ends. By applying NMR analysis to SPRTN´s two DNA-binding domains, we showed that these two domains are differently involved in this activation mechanism. The zinc-binding domain (ZBD) engages with unpaired nucleotides within these disruptions whereas the basic region (BR) interacts with the negative phosphate backbone of the dsDNA duplex. Mutations within either of those domains diminish activity, indicating their importance for enzyme function. The observed DNA-structure specific activation probes SPRTN activity towards biological relevant scenarios and prohibits unwanted degradation of chromatin-bound proteins.

Author contribution

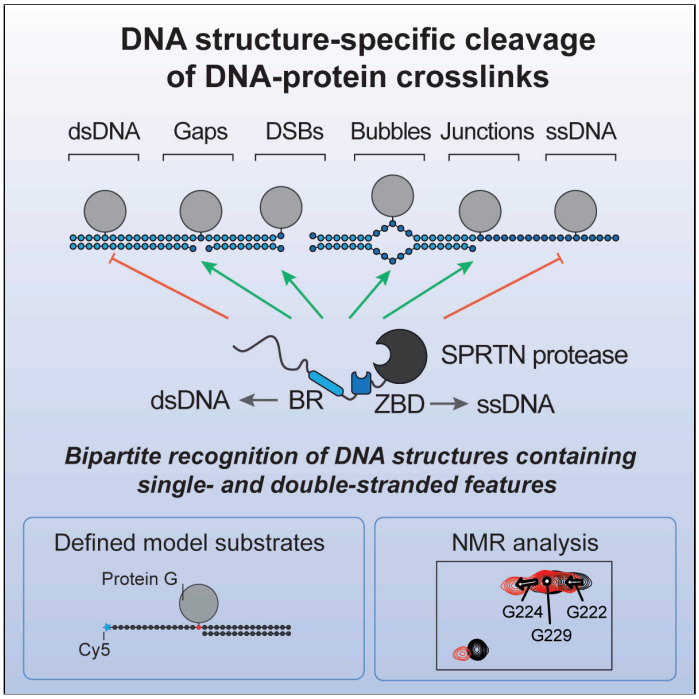
I established the expression and purification strategies for recombinant SPRTN and purified all variants used in this study. In addition, I generated defined model DPC-substrates, which I further employed in biochemical cleavage and binding assays. Moreover, I produced all labelled proteins used for NMR analysis by Hyun-Seo Kang from the Sattler group at the Technical University of Munich. I was involved in discussing and interpreting the NMR results with Julian Stingele, Hyun-Seo Kang and Michael Sattler. I wrote the manuscript together with Julian Stingele.

Molecular Cell

Article

DNA Structure-Specific Cleavage of DNA-Protein Crosslinks by the SPRTN Protease

Graphical Abstract



Highlights

- DNA-protein crosslink cleavage by SPRTN is coupled to recognition of DNA context
- DNA-protein crosslinks are only cleaved in proximity to activating DNA structures
- Two distinct interfaces recognize DNA with single- and double-stranded features
- Activation of SPRTN depends on simultaneous engagement of both DNA binding interfaces

Authors

Hannah K. Reinking, Hyun-Seo Kang, Maximilian J. Götz, ..., Lucas T. Jae, Michael Sattler, Julian Stingele

Correspondence

stingele@genzentrum.lmu.de

In Brief

Reinking et al. show that the protease SPRTN degrades DNA-protein crosslinks in a DNA structure-specific manner, which restricts cleavage to biologically relevant scenarios. NMR analysis reveals that specificity is achieved by a bipartite strategy relying on two DNA binding interfaces that recognize single- and double-stranded features within the substrate.

Reinking et al., 2020, Molecular Cell *80*, 102–113
October 1, 2020
https://doi.org/10.1016/j.molcel.2020.08.003

CellPress

CellPress
OPEN ACCESS

Molecular Cell

Article

DNA Structure-Specific Cleavage of DNA-Protein Crosslinks by the SPRTN Protease

Hannah K. Reinking,[1,2] Hyun-Seo Kang,[3,4] Maximilian J. Götz,[1,2] Hao-Yi Li,[1,2] Anja Kieser,[1,2] Shubo Zhao,[1,2] Aleida C. Acampora,[1,2] Pedro Weickert,[1,2] Evelyn Fessler,[1,2] Lucas T. Jae,[1,2] Michael Sattler,[3,4] and Julian Stingele[1,2,5,*]

[1]Department of Biochemistry, Ludwig Maximilians University, 81377 Munich, Germany
[2]Gene Center, Ludwig Maximilians University, 81377 Munich, Germany
[3]Center for Integrated Protein Science Munich at the Department of Chemistry, Technical University of Munich, 85747 Garching, Germany
[4]Institute of Structural Biology, Helmholtz Zentrum München, 85764 Neuherberg, Germany
[5]Lead Contact
*Correspondence: stingele@genzentrum.lmu.de
https://doi.org/10.1016/j.molcel.2020.08.003

SUMMARY

Repair of covalent DNA-protein crosslinks (DPCs) by DNA-dependent proteases has emerged as an essential genome maintenance mechanism required for cellular viability and tumor suppression. However, how proteolysis is restricted to the crosslinked protein while leaving surrounding chromatin proteins unharmed has remained unknown. Using defined DPC model substrates, we show that the DPC protease SPRTN displays strict DNA structure-specific activity. Strikingly, SPRTN cleaves DPCs at or in direct proximity to disruptions within double-stranded DNA. In contrast, proteins crosslinked to intact double- or single-stranded DNA are not cleaved by SPRTN. NMR spectroscopy data suggest that specificity is not merely affinity-driven but achieved through a flexible bipartite strategy based on two DNA binding interfaces recognizing distinct structural features. This couples DNA context to activation of the enzyme, tightly confining SPRTN's action to biologically relevant scenarios.

INTRODUCTION

Genome stability is constantly challenged by various types of DNA damage (Lindahl, 1993). Efficient detection and repair of DNA lesions is crucially important to prevent premature aging and cancer development (Jackson and Bartek, 2009). A particular type of lesion, covalent DNA-protein crosslinks (DPCs), has recently become the focus of intense research efforts. DPCs are induced by various reactive metabolites and chemotherapeutic agents and can also be caused by entrapment of enzymatic reaction intermediates (Barker et al., 2005; Stingele et al., 2017). DPCs are highly toxic because they block chromatin transactions such as transcription and replication (Duxin et al., 2014; Fu et al., 2011; Nakano et al., 2012, 2013). DPCs pose an exceptional challenge for repair because they are very diverse in nature with respect to the identity of the crosslinked protein and depending on the DNA context in which they occur (Nakano et al., 2017). DPCs form within double-stranded DNA (dsDNA) (e.g., those induced by formaldehyde or acetaldehyde), at DNA nicks (trapped topoisomerase 1 [TOP1]), DNA gaps (polymerase β adducts), or at dsDNA ends/breaks (SPO11 adducts, trapped topoisomerase 2 [TOP2]) (Chen et al., 2013; Lu et al., 2010; Neale et al., 2005; Quiñones et al., 2015).

DPCs can be repaired through degradation of the protein component by proteases of the Wss1/SPRTN family, which is essential for maintaining genome stability, cellular viability, tumor suppression, and prevention of premature aging (Lessel et al., 2014; Lopez-Mosqueda et al., 2016; Maskey et al., 2014; Mórocz et al., 2017; Reinking et al., 2020; Stingele et al., 2014, 2016; Vaz et al., 2016). These proteases tackle the complexity of DPCs with an open and, thus, unselective active site, which allows them to degrade virtually any protein irrespective of amino acid sequence (Stingele et al., 2016; Vaz et al., 2016). This, however, creates the need to prohibit unwanted cleavage of non-crosslinked cellular proteins. Accordingly, the human DPC protease SPRTN appears to be highly regulated. Mono-ubiquitinated SPRTN is excluded from chromatin, with the presence of DPCs triggering deubiquitylation and concurrent relocalization to chromatin (Stingele et al., 2016). Moreover, SPRTN's protease activity depends entirely on the presence of DNA. SPRTN is inactive *in vitro* when incubated on its own but becomes strongly activated upon DNA binding (Stingele et al., 2016; Vaz et al., 2016). DNA is thought to act as a scaffold bringing substrate and enzyme together, triggering non-specific degradation of DNA-bound proteins (non-DNA-binding proteins are not targeted by SPRTN even in the presence of DNA). If true *in vivo*, then recruiting SPRTN to DNA would carry enormous risks because all nearby chromatin proteins would potentially be subjected to uncontrolled degradation. However, insights obtained using a model system of replication-coupled DPC repair (using frog egg extracts) indicate that proteolytic action is exquisitely controlled; SPRTN cleaves plasmid-borne DPCs only when the

Molecular Cell
Article

CellPress
OPEN ACCESS

replisome has passed over the lesion and when the daughter strand has been extended on the DPC, whereas replisome and chromatin factors remain untouched (Duxin et al., 2014; Larsen et al., 2019; Sparks et al., 2019). How this specificity is achieved and whether it requires sophisticated regulation is unknown.

Here we identify an entirely unexpected DNA structure specificity of SPRTN by analyzing its activity for the first time using defined model DNA-protein conjugates. Moreover, NMR experiments suggest that SPRTN achieves such high precision using a unique bipartite strategy: two distinct DNA-binding interfaces reliably read out structural features and DNA context and couple it to activation of the enzyme. This regulatory mechanism results in tight spatial restriction of SPRTN's activity, which allows degradation of crosslinked proteins in a controlled and safe manner.

RESULTS

SPRTN Cleaves DPCs at dsDNA Ends

To understand how SPRTN's activity is influenced by different types of DNA, we initially focused on an intriguing conundrum. SPRTN has been reported to be efficiently activated by DNA oligonucleotides, whether they were single- or double-stranded (Lopez-Mosqueda et al., 2016; Vaz et al., 2016). In contrast, others observed a striking difference using long circular DNA for activation. Single-stranded DNA (ssDNA) circles were found to activate SPRTN much more strongly than dsDNA circles (Stingele et al., 2016). Remarkably, these seemingly contradictory results hold true when conducted in the same experiment. ssDNA circles (ΦX174 phage DNA, 5.4 kb) induce SPRTN activity much more efficiently than dsDNA circles, as judged by autocleavage and cleavage of histone H1 (Figures 1A and S1A–S1C). However, 60-mer single- and double-stranded oligonucleotides activate SPRTN very similarly, although generally less than ssDNA circles. The specific inability of long circular dsDNA to activate SPRTN becomes even more obvious under more stringent high-salt assay conditions (150 mM KCl). Denaturation of dsDNA circles (ΦX174 phage DNA or pMAX-GFP plasmids) to ssDNA by heating and snap-cooling on ice restores their activation potential (Figures 1B and S1D–S1G). We conclude that it is indeed the double-strandedness that prohibits SPRTN activation by dsDNA circles. To test whether the reason for the differential activation of SPRTN by dsDNA circles and double-stranded oligonucleotides is simply the difference in length, we next tested PCR-generated dsDNA fragments of decreasing size for activation. Strikingly, the shorter the dsDNA fragment, the more strongly it activates SPRTN under high-salt conditions (Figures 1C, S1H, and S1I). Of note, histone H1 cleavage cannot be observed, which indicates that it requires stronger activation of SPRTN or reflects the binding preference of H1 itself. Importantly, when using shorter DNA fragments, the total amount of DNA was kept constant. Thus, the number of dsDNA ends increases when shorter fragments are used, which raises the possibility that SPRTN is activated by dsDNA ends (Figure 1D).

To test whether SPRTN is indeed active at dsDNA ends, we generated defined model DPCs: protein G conjugated in a site-specific manner to Cy5-labeled 30-mer oligonucleotides followed by purification via ion-exchange chromatography (Figure 1E). Drastically reduced enzyme concentrations in the low nanomolar range (100-fold less compared with previous assays) can be used to assess cleavage of these substrates. Wild-type (WT) SPRTN, but not the catalytically inactive E112Q (EQ) variant, efficiently cleaves the protein adduct when crosslinked to the terminal base at the 3′ or 5′ end of a dsDNA oligonucleotide (Figures 1F and 1G). In stark contrast, the adduct is not processed at an internal position despite SPRTN binding to it very similarly, as determined by electrophoretic mobility shift assays (EMSAs) (Figure 1H). This apparent specificity of SPRTN is striking and potentially explains how dsDNA-bound chromatin proteins are protected from cleavage.

SPRTN Cleaves DPCs at Hairpins and ssDNA to dsDNA Junctions

It is unlikely that activation takes place exclusively at dsDNA ends because ssDNA circles activate SPRTN very efficiently. To gain insights into activation of SPRTN by ssDNA, we assessed cleavage of the same model DPCs in their single-stranded versions (Figure 2A). Remarkably, the cleavage preference shifts dramatically. The internal adduct is cleaved most efficiently, the 5′ adduct is still processed but to a lower degree, and the 3′ adduct is barely cleaved at all (Figures 2B and 2C). Again, SPRTN binds similarly to all substrates (Figure 2D). Next we wanted to find out whether cleavage preference is related to secondary structures forming within the ssDNA (the long ssDNA circles that efficiently activate SPRTN contain various hairpin structures). The sequence used for the model DPCs is predicted to form a stable hairpin at assay temperature (25°C), and the cleavage efficiency of the protein G adduct appeared to correlate with the proximity to the hairpin. Thus, we tested the isolated hairpin for activation of SPRTN and observed efficient induction of autocleavage and histone cleavage (Figures 2E–2G and S2A). A mutation predicted to result in collapse of the hairpin strongly reduces activation, whereas the double-stranded versions of both sequences activate indistinguishably. Notably, abolishment of hairpin formation does not only reduce activation but also binding by SPRTN (Figure S2B). Furthermore, strictly ssDNAs (poly(dA) or poly(dT)) do not induce SPRTN autocleavage but do so when annealed to each other (Figures 2H, 2I, and S2C). Finally, we tested cleavage of a model DPC substrate containing strictly ssDNA ($C_3A_{11}XA_{12}C_3$) (Figure 2J) and observed neither cleavage nor efficient binding by SPRTN (Figures 2K–2M). Taken together, these data indicate that formation of secondary structures is required for binding and activation of SPRTN by ssDNA. Next we annealed complementary 15-mer or 30-mer oligonucleotides to the single-stranded model DPC, which restored strong binding by SPRTN (Figures 2J and 2M). However, efficient cleavage of the DPC occurs only at the ss/dsDNA junction (Figures 2K and 2L). We conclude that a short section of paired DNA bases is needed for SPRTN to bind efficiently. Cleavage, however, appears to also require the presence of DPCs at specific DNA structures, either at dsDNA ends, in proximity to the stem loop of a hairpin, or at a ss/dsDNA junction.

SPRTN's Structure-Specific Activity Requires Two Distinct DNA-Binding Interfaces

Having established that SPRTN's protease activity displays strict preferences for certain DNA contexts, we wanted to find out how

CellPress
OPEN ACCESS

Molecular Cell
Article

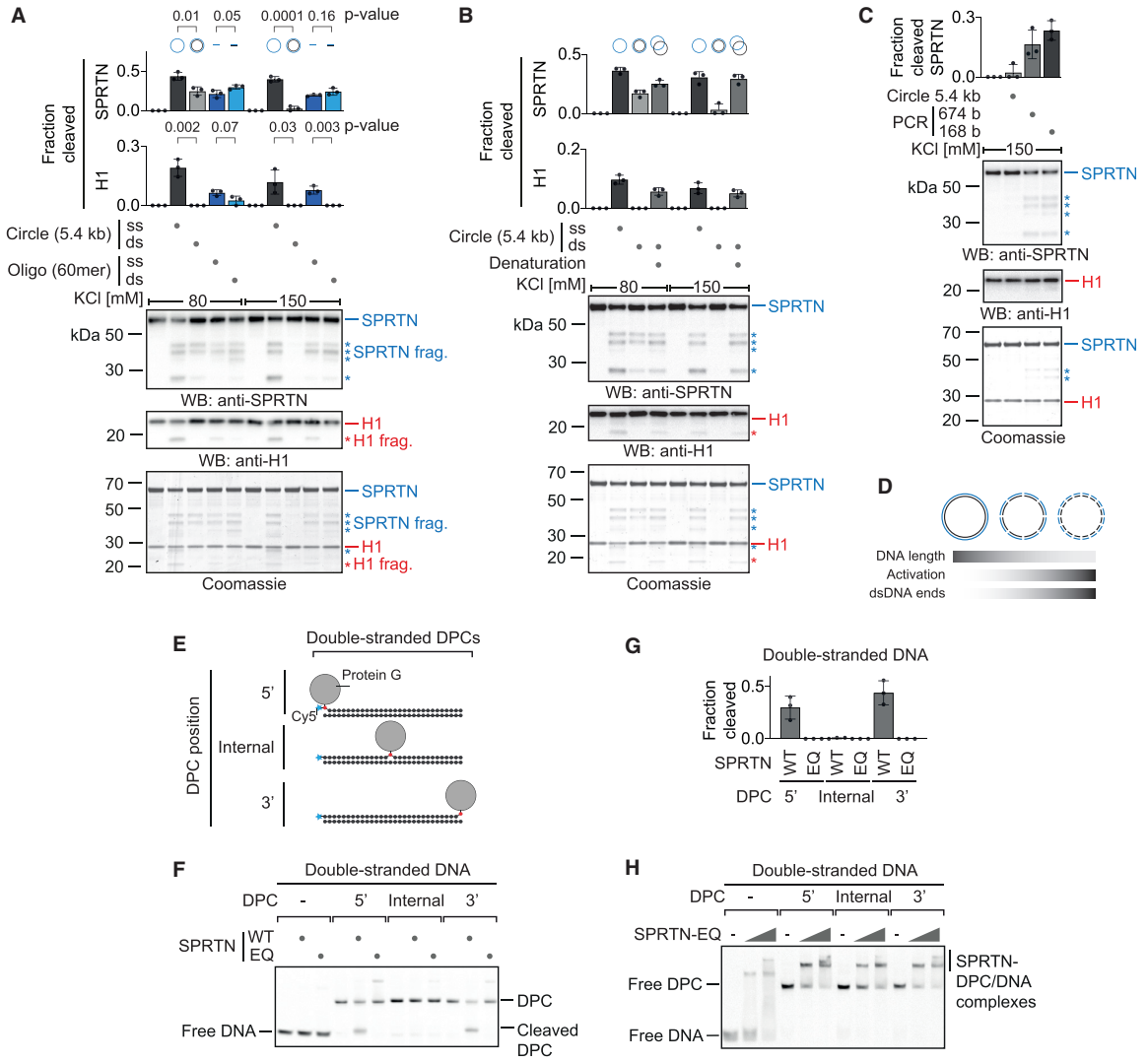


Figure 1. SPRTN Cleaves DPCs at dsDNA Ends

(A) Recombinant SPRTN (500 nM) and histone H1 (500 nM) were incubated alone or in the presence of DNA (5.4 kb circles [ΦX174] or 60-mer oligonucleotides, each single-stranded or double-stranded) for 2 h at 25°C. DNA concentrations were 1 μM for 60-mer oligonucleotides or the corresponding amount of circular DNA (11.4 nM). Reactions were analyzed by SDS-PAGE followed by western blotting and Coomassie staining. Cleaved fragments of SPRTN and H1 are indicated by asterisks. Quantification of western blots results of SPRTN and histone H1 cleavage: values represent the mean ± SD of three independent experiments. The p values were calculated using an unpaired t test.

(B) Reactions and quantification were conducted as in (A) but also included dsDNA (ΦX174) denatured by heating and snap-cooling on ice.

(C) PCR-generated dsDNA fragments were tested for activation of SPRTN as in (A).

(D) Schematic representation of SPRTN's activation by dsDNA and its correlation with DNA length and the number of dsDNA ends.

(E) Schematic of the model DPCs used in (F) and (H). Protein G was conjugated site-specifically to fluorescently labeled 30-mer oligonucleotides prior to annealing complementary reverse oligonucleotides.

(F) Free DNA or the indicated model DPCs (25 nM) were incubated alone or in the presence of recombinant SPRTN (5 nM, WT or the catalytically inactive E112Q [EQ] variant) for 2 h at 25°C prior to separation by native PAGE.

(G) Quantification of the DPC cleavage assay shown in (F). Values represent the mean ± SD of three independent experiments.

(H) EMSAs were used to assess binding of catalytically inactive SPRTN EQ (12.5 and 50 nM) to free dsDNA or the indicated DPCs (25 nM).

See also Figure S1.

Molecular Cell
Article

CellPress
OPEN ACCESS

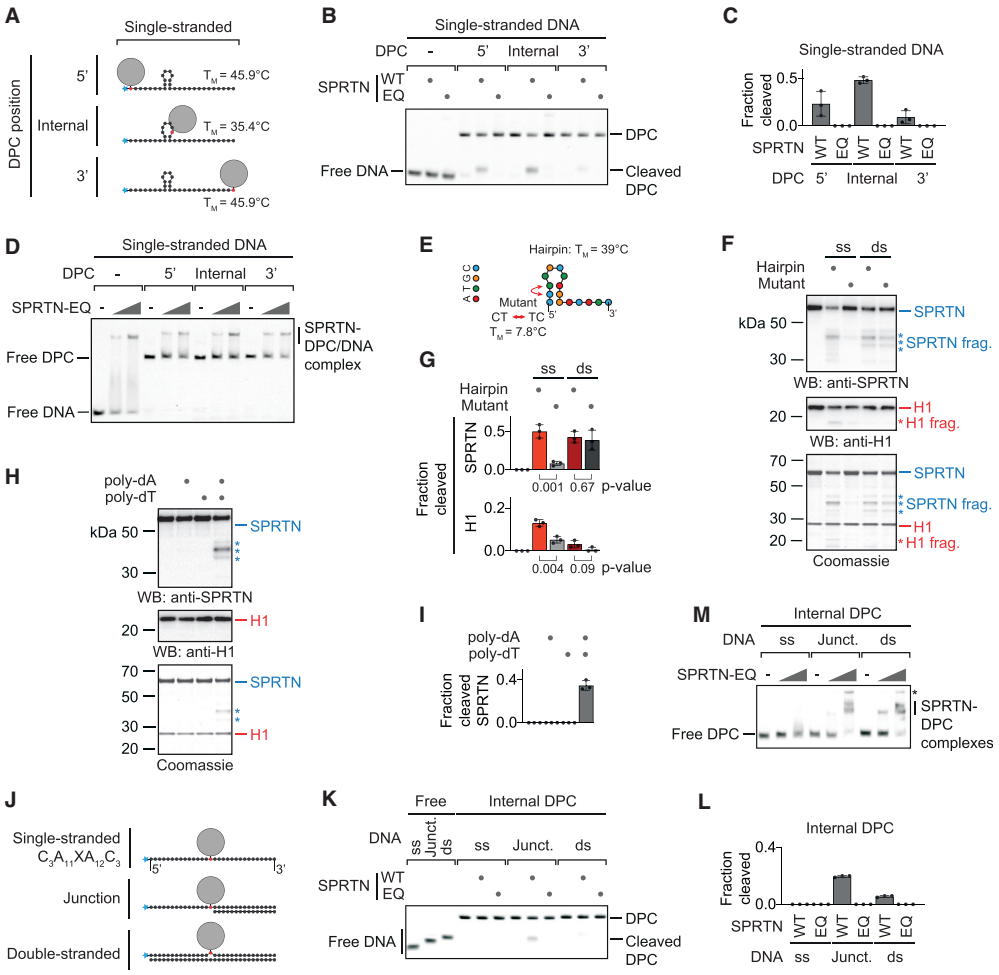


Figure 2. SPRTN Cleaves DPCs at Hairpins and ss/dsDNA Junctions

(A) Schematic of the model DPCs used in (B) and (D). Protein G was conjugated site-specifically to fluorescently labeled 30-mer oligonucleotides. Secondary structures and respective melting temperatures (T_M) were predicted using the mfold webserver.

(B) Free DNA or the indicated model DPCs (25 nM) were incubated alone or in the presence of recombinant SPRTN (5 nM, WT or the catalytically inactive EQ variant) for 2 h at 25°C prior to separation by native PAGE.

(C) Quantification of the DPC cleavage assay shown in (B). Values represent the mean ± SD of three independent experiments.

(D) EMSA assays were used to assess binding of catalytically inactive SPRTN EQ (12.5 and 50 nM) to free ssDNA and the indicated DPCs (25 nM).

(E) Schematic of the 15-mer DNA hairpin and its mutant variant used for activation of SPRTN in (F).

(F and G) Recombinant SPRTN (500 nM) and histone H1 (500 nM) were incubated alone or in the presence of the indicated DNAs (4 μM) for 2 h at 25°C and 80 mM KCl. Reactions were analyzed by SDS-PAGE, followed by western blotting and Coomassie staining. Cleaved fragments of SPRTN and H1 are indicated by asterisks. Quantification of western blots results of SPRTN and histone H1 cleavage: values represent the mean ± SD of three independent experiments. The p values were calculated using an unpaired t test.

(H and I) 15-mer poly(dA) or poly(dT) oligonucleotides (4 μM) were tested for activation of SPRTN. Reactions and quantification were as in (F) and (G).

(J) Schematic of the model DPCs used in (K) and (M). Protein G was conjugated site-specifically to fluorescently labeled 30-mer oligonucleotides prior to annealing complementary reverse oligonucleotides.

(K) The indicated model DPCs (25 nM) were incubated alone or in the presence of recombinant SPRTN (12.5 nM, WT or the catalytically inactive EQ variant) for 2 h at 25°C prior to separation by native PAGE.

(L) Quantification of the DPC cleavage assay shown in (K). Values represent the mean ± SD of three independent experiments.

(M) EMSAs were used to assess binding of catalytically inactive SPRTN EQ (12.5 and 50 nM) to the indicated model DPCs (25 nM). An asterisk indicates non-resolvable high-molecular-weight aggregates.

See also Figure S2.

CellPress
OPEN ACCESS

Molecular Cell
Article

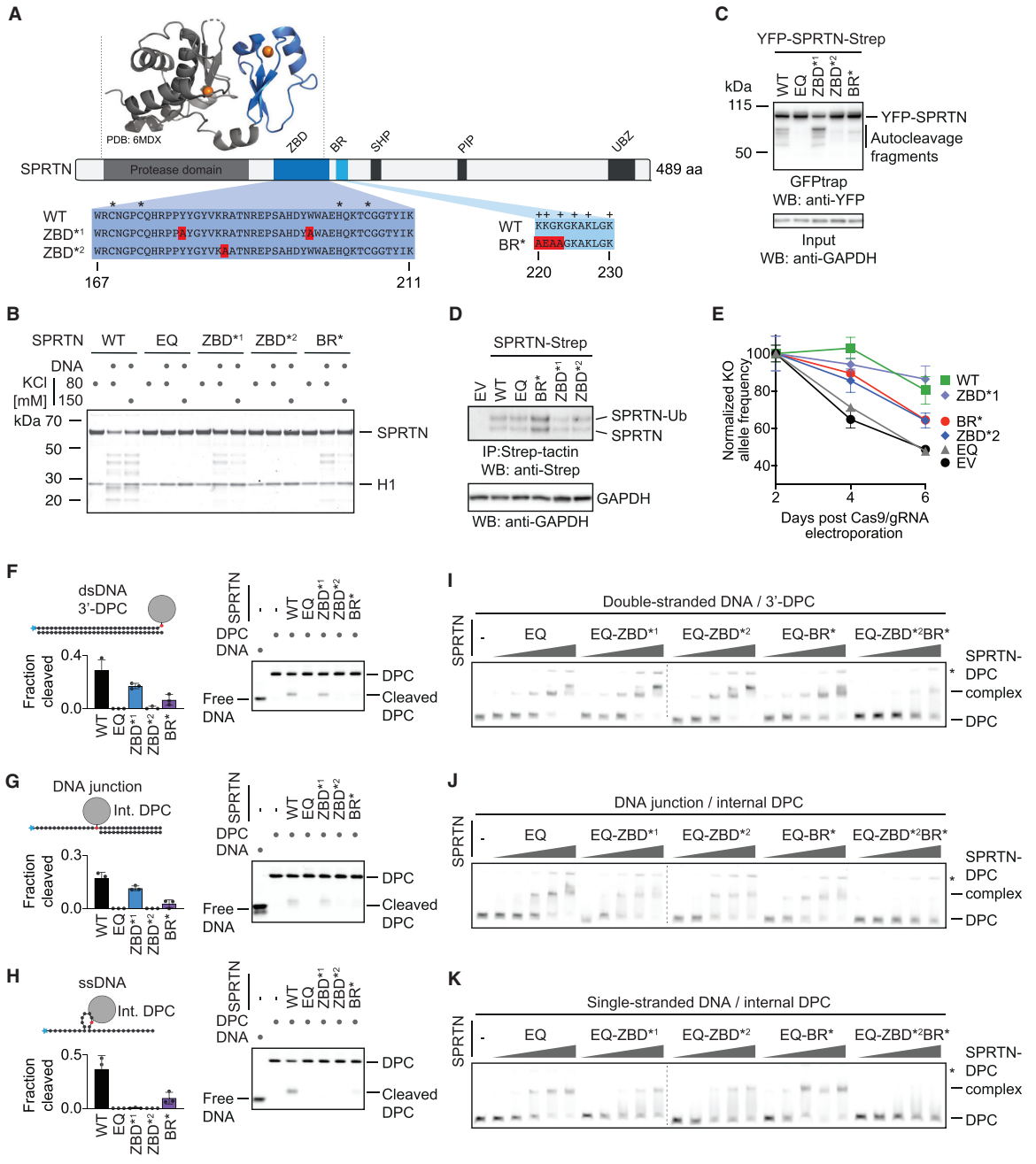


Figure 3. SPRTN's Structure-Specific Activity Requires Two Distinct DNA-Binding Domains

(A) Schematic of SPRTN's domain structure, highlighting the zinc-binding domain (ZBD), the basic DNA-binding region (BR), the SHP box (p97 binding), the PCNA-interacting motif (PIP), and the ubiquitin-binding zinc finger (UBZ). Asterisks indicate the zinc-coordinating residues within the ZBD, and plus signs indicate positively charged amino acids within the BR. The function of the ZBD and BR were tested in this study using the indicated amino acid replacements (ZBD*1, Y179A/W197A; ZBD*2, R185A; BR*, K220A/K221E/G222A/K223A).

(B) Recombinant SPRTN (500 nM, WT or the indicated variants) and histone H1 (500 nM) were incubated alone or in the presence of ssDNA circles (ΦX174 virion) for 2 h at 25°C in the presence of 80 or 150 mM KCl. Reactions were analyzed by SDS-PAGE followed by Coomassie staining.

(legend continued on next page)

Molecular Cell
Article

CellPress
OPEN ACCESS

specificity is achieved. SPRTN is a 55-kDa protein, with the N-terminal part of the enzyme bearing the catalytic metalloprotease domain (Figure 3A). The largely unstructured C-terminal tail contains several protein-protein interaction domains (a ubiquitin-binding zinc finger, a proliferating cell nuclear antigen (PCNA)-interacting protein motif, and a SHP box required for binding to the chaperone-like protein p97) (Centore et al., 2012; Davis et al., 2012; Mosbech et al., 2012; Stingele et al., 2015). Between the tail and protease domain, a basic DNA-binding region (BR) of low complexity was identified that bears several positively charged amino acids (Mórocz et al., 2017; Stingele et al., 2016; Toth et al., 2017). A recent crystal structure of an N-terminal SPRTN fragment revealed an unexpected zinc-binding domain (ZBD) immediately after the protease domain and preceding the BR (PDB: 6MDX; Li et al., 2019). The ZBD was speculated to constitute a ssDNA-binding domain, which is interesting given that we cannot detect efficient binding of SPRTN to substrates containing only ssDNA. Consistent with previous data, we observed reduced autocleavage in SPRTN variants with specific amino acid replacements in the ZBD domain (the ZBD^{*2} [R185A] variant displays a more severe effect than ZBD^{*1} [Y179A_W197A]) (Figure 3B; Li et al., 2019). Similarly, a SPRTN variant with amino acid replacements in the BR domain (BR*; K220A_K221E_G222A_K223A) shows a comparable reduction in activity. Consistent with their crucial role *in vitro*, the more severe ZBD^{*2} variant and the BR* variant display decreased autocleavage when expressed in cells, although recruitment to chromatin after DPC induction by formaldehyde is not affected (Figures 3C, S3A, and S3B). To test whether ZBD and BR contribute to SPRTN's essential function in cells, we expressed cDNAs of the respective SPRTN variants with a retroviral vector in human haploid HAP1 cells (Figure 3D). Next we transfected these cells with recombinant nuclear localization signal (NLS)-Cas9/guide RNA (gRNA) complexes targeting the 5′ and 3′ UTR of the endogenous allele (Figure S3C). The persistence of the resulting SPRTN KO allele was then monitored over time using qPCR. HAP1 cells complemented with WT SPRTN or ZBD^{*1} can tolerate loss of the endogenous SPRTN allele whereas cells transduced with an empty vector (EV) or catalytically inactive SPRTN-EQ cannot (Figure 3E). SPRTN-BR* and ZBD^{*2} display only partial complementation, highlighting the importance of both modules.

To understand how ZBD and BR contribute to SPRTN's activity, we tested the respective SPRTN variants for DPC cleavage and binding. Cleavage of a protein adduct at a dsDNA end, a ss/dsDNA junction, or a hairpin structure is severely reduced in the BR* and ZBD^{*2} variants (Figures 3F–3H). The less stringent ZBD^{*1} mutation mostly affects cleavage of the hairpin DPC. Remarkably, despite being crucial for proteolytic activity, the SPRTN-ZBD* and BR* variants do not show observable defects in substrate binding (Figures 3I–3K and S3D). A severe effect on binding is only observed upon introduction of simultaneous alterations in both DNA binding regions (ZBD^{*2}/BR*). Taken together, these results demonstrate that both DNA binding regions are required for activity and also suggest that recognition of substrates by SPRTN depends on two distinct features recognized by the ZBD and BR, respectively.

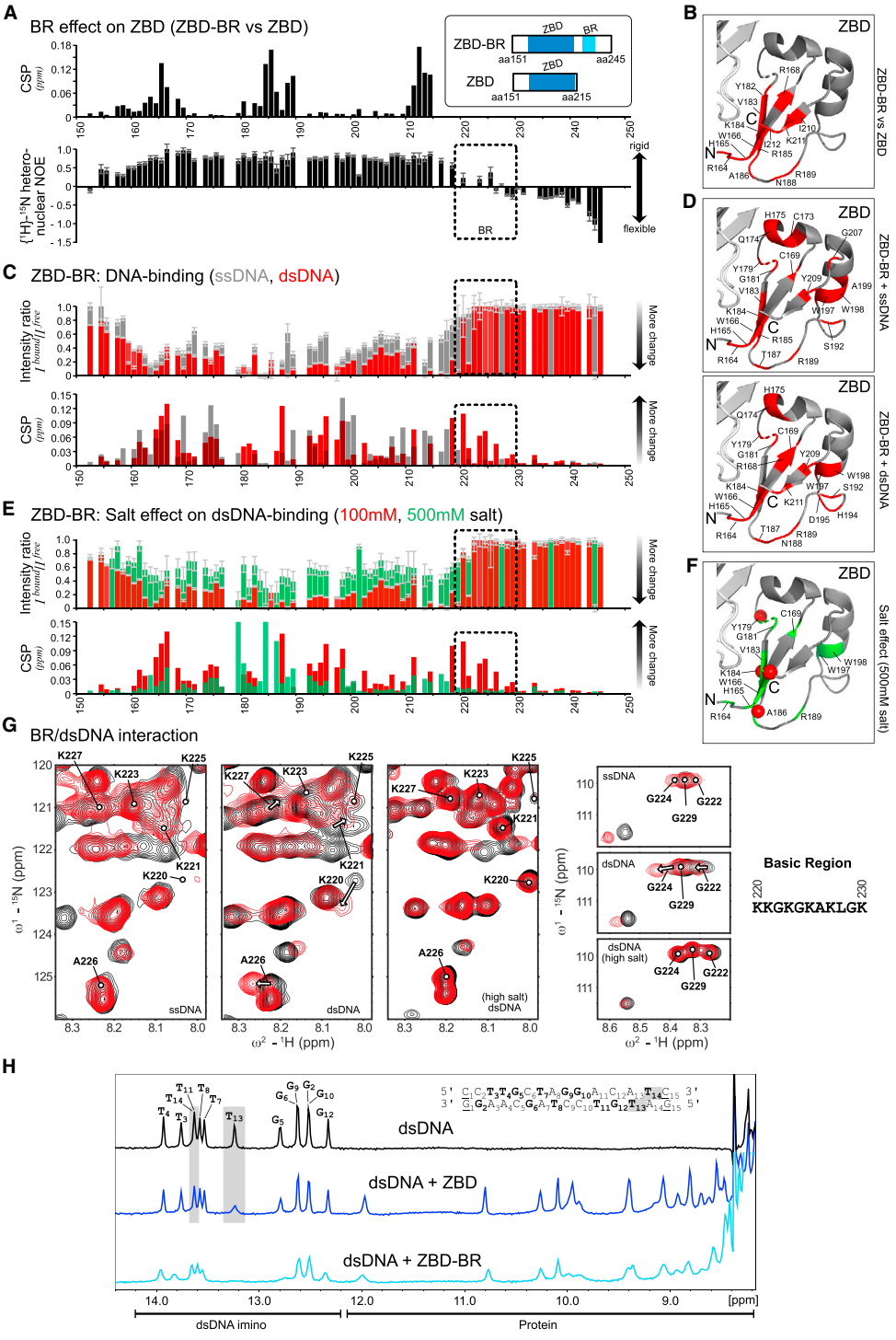
NMR Analysis Reveals Bipartite Recognition of DNA Structures by SPRTN

To probe the structural contributions of ZBD and BR for DNA binding, we analyzed two constructs comprising the entire ZBD-BR module or just the ZBD using NMR. NMR backbone chemical shift assignments enabled analysis of the DNA interactions (Figures 4 and S4). First, when comparing ZBD-BR and ZBD in the absence of DNA, we observed significant chemical shift differences in the β sheet of the ZBD (Figure 4A, top; 4B; and S4A). This suggests transient contacts between the BR and the β sheet of the ZBD. This is further supported by the NMR relaxation experiments, which show that the BR is less flexible on a sub-nanosecond timescale, especially in comparison with the C-terminal end (Figure 4A, bottom). Together, these data suggest a dynamic interaction of the intrinsically disordered BR with the ZBD. Next we monitored chemical shift perturbations (CSPs) for ZBD-BR and ZBD in ^{1}H,^{15}N correlation experiments upon adding 15-mer ssDNA (poly(dA)) or dsDNA (the same sequence as used in Figures 2F and S2B for binding and activation assays). Binding to ssDNA and dsDNA by ZBD-BR and ZBD is readily observed, as evidenced by significant chemical shift changes and line broadening (intensity changes; Figures 4C, S4B, and S4C). Notably, however, large CSPs for the BR region are only observed upon binding dsDNA but not ssDNA, whereas CSPs of the ZBD are observed with ssDNA and dsDNA

(C) SPRTN autocleavage assessed in cells. The indicated YFP-SPRTN-Strep variants were transiently transfected in HeLa Flp-In TRex cells. SPRTN autocleavage fragments were enriched on GFP trap resins, followed by western blotting against the N-terminal YFP tag. Western blotting against GAPDH of cell lysates served as loading control.
(D) HAP1 cell lines complemented by retroviral transduction with cDNAs encoding the indicated C-terminally Strep-tagged SPRTN variants. SPRTN-Strep was enriched on Strep-tactin beads prior to western blotting because of low expression levels. Western blotting against GAPDH of cell lysates served as loading control.
(E) The indicated cell lines were transfected with NLS-Cas9/gRNA complexes targeting the UTRs of the endogenous SPRTN allele. The ratio between the resulting knockout (KO) allele compared with the WT allele was monitored over time using qPCR. A schematic of the genotyping strategy is depicted in Figure S3C. Values represent the mean ± SD of three technical replicates normalized to day 2.
(F–H) The indicated fluorescently labeled model DPCs (25 nM) were incubated alone or in the presence of recombinant SPRTN (WT or the indicated variants) for 2 h at 25°C prior to separation by native PAGE. SPRTN concentrations were 5 nM in (F) and (H) and 12.5 nM in (G). Quantification: values represent the mean ± SD of three independent experiments.
(I–K) EMSAs were used to assess binding of catalytically inactive SPRTN EQ (alone or in combination with the indicated amino acid replacements in the ZBD/BR) to the indicated DPCs (25 nM). SPRTN concentrations were 3.125, 6.25, 12.5, 25, and 50 nM. Asterisks indicate non-resolvable high-molecular-weight aggregates.
See also Figure S3.

CellPress
OPEN ACCESS

Molecular Cell
Article



(legend on next page)

Molecular Cell
Article

CellPress
OPEN ACCESS

(Figures 4C, 4D, S4B, and S4C). This is in line with electrostatic interactions of positively charged side chains within the BR with the negatively charged phosphate backbone of the double-stranded region of the dsDNA ligand. Consistent with this interpretation, the interaction between BR and dsDNA is strongly reduced when titration is performed at higher salt concentrations (500 mM) (Figures 4E–4G and S4B). In contrast, the interaction between ZBD and DNA is much less affected, which is in line with the previous observation that the ZBD binds to DNA bases through stacking of its aromatic residues (Li et al., 2019).

Next we asked which features in the DNA are recognized by SPRTN's ZBD. To this end, we monitored spectral changes of the imino NMR signals in the base pairs of the 15-mer dsDNA upon binding to ZBD-BR or ZBD (Figure 4H). Intriguingly, binding of the isolated ZBD mainly affects NMR signals of base pairs at one end of the dsDNA (i.e., T_{13} and T_{14}). In contrast, when the low-complexity and highly charged BR is present, most imino signals are affected and experience line broadening. This further indicates that the BR contributes binding to the double-stranded part of the oligonucleotide. Thus, we hypothesize that the ZBD interacts specifically with unpaired DNA bases available for interaction at the dsDNA end. This idea is in agreement with the fact that the ZBD interacts with the presumably less stable end of the oligonucleotide (GAT versus CCT). Accordingly, we argue that the common feature recognized by the ZBD is the presence of ssDNA at "frayed" dsDNA ends, ss/dsDNA junctions, or at the ends of a DNA hairpin, whereas the BR enhances binding through non-specific interactions with the double-stranded parts of these structures. If correct, then DPC processing by SPRTN should be enabled by introduction of DNA disruptions that allow local unwinding and, thus, result in the presence of unpaired DNA bases in the vicinity of the DPC.

SPRTN Cleaves DPCs in Close Proximity to Disruptions within dsDNA

To test this hypothesis, we generated model DPCs containing specific disruptions expected to result in local opening of duplex DNA in close proximity to the DPC. First, we disrupted the duplex by a nick, a nick combined with a mismatch (1 bp), or a gap (1 bp) opposite the protein adduct (Figure 5A). Strikingly, this enables cleavage of the DPC depending on SPRTN's ZBD and BR domain (Figures 5A–5C and S5A). Second, we inserted a bubble of increasing size opposite the protein adduct (Figure 5D). Disrupting the 30-mer duplex by a bubble larger than 2 bp enables efficient DPC cleavage by SPRTN (Figures 5D, 5E and S5B). Cleavage of the protein adduct within the bubble again depends on both DNA binding domains (Figure 5F). Having established the requirement for discontinuities within duplex DNA for DPC cleavage by SPRTN, we investigated the spatial interdependency between the activating structure and the position of the protein adduct. To this end, we recessed the DNA strand opposite a 3′ DPC in small steps, moving the putatively activating ss/dsDNA junction farther and farther away from the protein adduct (Figure 5G). Remarkably, an initial increase in cleavage (with a peak around 5 bp between the junction and the adduct) is followed by a sharp decrease when the junction is moved farther away from the adduct (Figures 5G and 5H), whereas binding to the substrates is only mildly affected (Figure 5I). Next we assessed the inverted scenario, in which we brought the activating junction closer to an internal adduct (Figure 5J). In this scenario, cleavage of the protein adduct again depends on close proximity between the junction and the adduct; DPC proteolysis increases sharply at distance smaller than 5 bp (Figures 5J–5L). Again, this effect did not correlate with binding to the substrate (Figure 5L). We conclude that activation of SPRTN happens in a spatially confined manner that restricts substrate cleavage to a very narrow window around specific DNA structures.

Figure 4. NMR Analysis Reveals Bipartite Recognition of DNA Structures by SPRTN

(A) Comparison of NMR data for two SPRTN constructs comprising the ZBD only or ZBD and the BR (ZBD-BR). Top: chemical shift differences of the backbone amide resonances between ZBD and ZBD-BR. Bottom: backbone flexibility of ZBD-BR from {^{1}H}-^{15}N-heteronuclear NOE data. Errors for heteronuclear NOE values were estimated from error propagation of peak height uncertainties based on average noise levels (six randomly chosen positions in each NMR spectra). The dotted area indicates the BR region.

(B) Mapping of chemical shift differences of ZBD in the presence of BR from (A) onto the ZBD structure (PDB: 6MDW). Red color highlights residues with CSPs of more than 0.025 ppm in (A).

(C) Chemical shift perturbations (CSPs) and intensity differences (line broadening) of backbone amides in ZBD-BR upon addition of an equimolar ratio of ssDNA (gray) and dsDNA (red). Errors for intensity ratios upon DNA-binding were estimated from error propagation of peak height uncertainties based on average noise levels (six randomly chosen positions in each NMR spectra). The dotted area indicates the BR region. No boxes are shown for prolines, unassigned, or ambiguous (overlapped) residues. Spectral overlays are show in Figure S4B.

(D) Spectral changes upon DNA binding are mapped onto the ZBD structure (PDB: 6MDW). Changes observed for binding of ZBD-BR to ssDNA (top) or dsDNA (bottom) are shown in red for residues with an intensity ratio of less than 0.15 (85% intensity loss) or CSPs of more than 0.05 ppm.

(E) CSP and intensity changes of ZBD-BR upon addition of an equimolar dsDNA at 100 mM (low salt, red) and 500 mM (high salt, green) salt concentrations. Errors as in (C). Spectral overlays are shown in Figure S4B.

(F) Spectral changes upon dsDNA binding at high salt concentration are mapped onto the ZBD structure (PDB: 6MDW), where the 10 residues with the highest intensity or CSP changes are shown in green. Red spheres indicate changes with an intensity ratio of less than 0.15 (85% intensity loss) or CSPs of more than 0.05 ppm (as in D).

(G) NMR signals (black, free; red, bound) in Figure S4A, highlighting BR residues upon addition of an equimolar ssDNA, dsDNA at 100 mM salt concentration, and dsDNA at 500 mM salt concentration. See Figure S4B for the experimental conditions.

(H) Top: ^{1}H-NMR spectrum of the 15-mer dsDNA. Assignments of the imino resonances of T and G in base pairs in the dsDNA ligand are shown in bold in the sequence. Only 13 signals are observed because of fraying of the terminal base pairs (underlined in the sequence). Center and bottom: ^{1}H-NMR imino spectra of the dsDNA in the presence of an equimolar amount of ZBD or ZBD-BR, respectively. The gray box indicates strongly affected signals (line-broadening) upon addition of the ZBD. NMR spectra were recorded with 100-μM sample concentration in 100 mM potassium chloride, 50 mM HEPES (pH 7.5), 2 mM TCEP at 298 K on a 600-MHz spectrometer.

See also Figure S4.

CellPress
OPEN ACCESS

Molecular Cell
Article

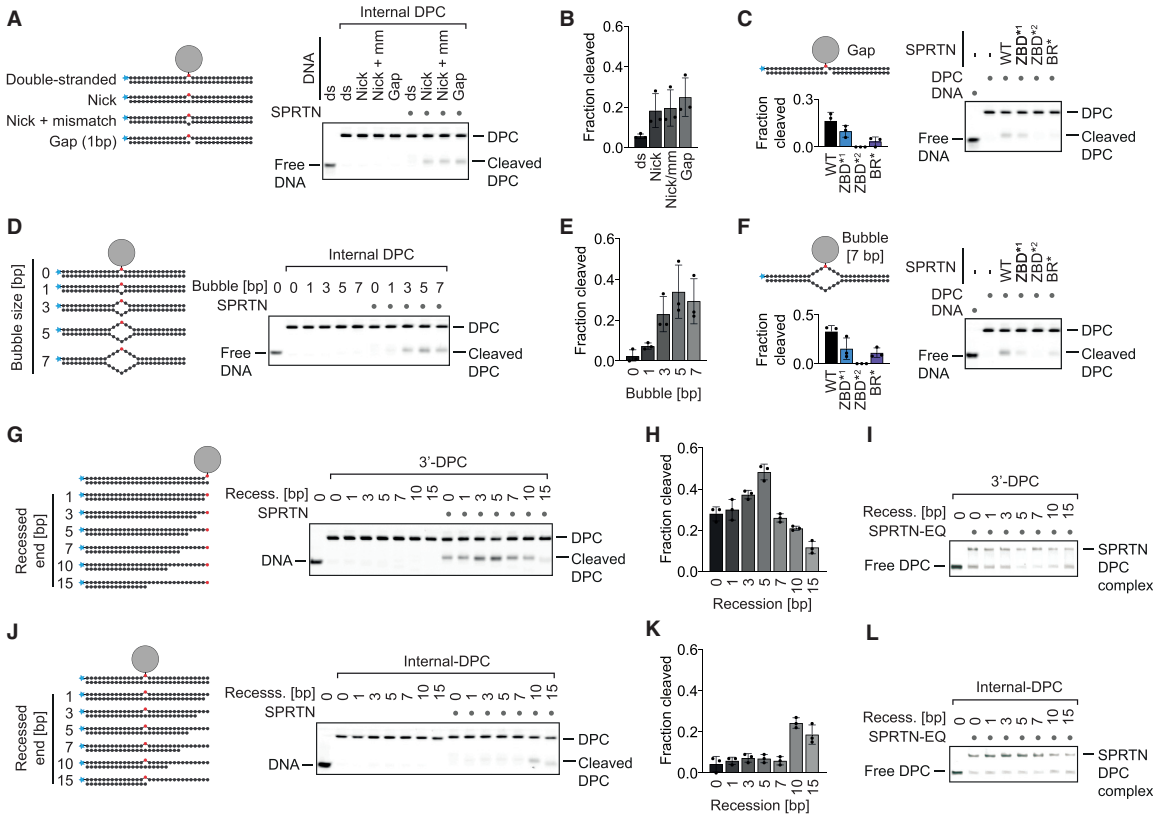


Figure 5. SPRTN Cleaves DPCs in Close Proximity to Disruptions within dsDNA

(A, D, G, and J) Cleavage of model DPCs. Protein G was conjugated site-specifically to fluorescently labeled 30-mer oligonucleotides prior to annealing complementary reverse oligonucleotides to generate the indicated substrates. Model DPCs (25 nM) were incubated alone or in the presence of recombinant SPRTN (WT, 5 nM) for 2 h at 25°C prior to separation by native PAGE.

(B, E, H, and K) Quantifications of DPC cleavage assays shown in (A), (D), (G), and (J). Values represent the mean ± SD of three independent experiments.

(C and F) Both DNA binding domains of SPRTN are required for DPC processing. The indicated fluorescently labeled model DPCs (25 nM) were incubated alone or in the presence of recombinant SPRTN (WT or the indicated variants, 5 nM) for 2 h at 25°C prior to separation by native PAGE. Quantification: values represent the mean ± SD of three independent experiments.

(I and L) SPRTN binds similarly to the model DPCs shown in (G) and (J). EMSA assays were used to assess binding of catalytically inactive SPRTN EQ (25 nM) to the indicated model DPCs (25 nM).

See also Figure S5.

DISCUSSION

Many DNA repair mechanisms (e.g., nucleotide excision repair or the Fanconi anemia pathway) are dispensable for viability unless cells are exposed to high levels of damage (Langevin et al., 2011; Setlow et al., 1969). In contrast, loss of the DPC protease SPRTN is lethal in mammalian cells, indicating constant life-threatening levels of DPCs (Hart et al., 2015; Maskey et al., 2014). Detection and repair of those crosslinks is complicated by several challenges. The diversity of these lesions (type of protein adduct/DNA structure) makes it difficult to evolve sensor proteins with high affinity for DPCs. The exception is enzymes specifically involved in repairing only certain protein adducts, such as TDP1 and TDP2, which target TOP1 and TOP2 adducts, respectively (Cortes Ledesma et al., 2009; Pouliot et al., 1999). Moreover, the DPC repair machinery must reliably distinguish covalent adducts from mere DNA-bound proteins (which are present in very large excess). Here we discovered that such specificity is achieved by recognition of DNA context, which is directly coupled to DPC cleavage. Importantly, several types of frequent DPCs form specifically at those structures, which trigger SPRTN activation. First, SPRTN protects cells against the toxicity of drugs (e.g., etoposide) inducing entrapment of TOP2 and appears to also be important for processing covalent SPO11 adducts during meiosis (Dokshin et al., 2020; Lopez-Mosqueda et al., 2016; Vaz et al., 2016). In both scenarios, TOP2 and SPO11 form covalent adducts with the 5′ ends of a dsDNA end. Second, SPRTN repairs covalent TOP1 adducts (induced by compounds such as camptothecin), which occur at DNA nicks (Maskey et al., 2017; Pommier, 2006). Third,

Molecular Cell
Article

CellPress
OPEN ACCESS

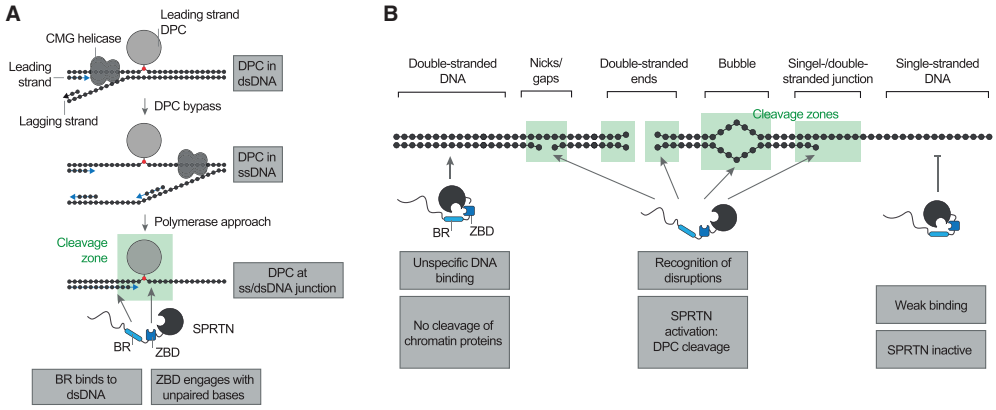


Figure 6. Model of SPRTN's DNA Structure-Specific Protease Activity

(A) Model of replication-coupled transfer of DPCs from dsDNA into a ss/dsDNA junction. The ss/dsDNA junction bears both features required for SPRTN activation: dsDNA, which is recognized by the BR, and unpaired DNA bases, which engage the ZBD.

(B) Schematic overview of the DNA structures activating SPRTN. DNA nicks, gaps, ends, bubbles, and junctions contain both features required for SPRTN activation: dsDNA and unpaired DNA bases.

polymerase ß can become covalently trapped at DNA gaps during base excision repair (SPRTN's role in repairing those adducts has not yet been assessed) (Quiñones et al., 2015). In all of these cases, the DPC already encompasses a DNA structure, which allows activation of SPRTN. The situation is different for non-specific DPCs induced by reactive metabolites, such as formaldehyde or acetaldehyde, which are expected to form within intact dsDNA. These lesions require pre-processing to make them amenable to cleavage by SPRTN. Recent data obtained using frog egg extracts indicate that this happens in a replication-dependent manner (Larsen et al., 2019; Sparks et al., 2019). A leading-strand DPC initially stalls progression of the replicative helicase, but the crosslink is eventually bypassed (presumably depending on a second helicase, RTEL1, unwinding the stalled fork) (Figure 6A). This transfers the protein adduct into ssDNA. However, proteolysis of the DPC only occurs when the DNA polymerase extends the newly synthesized strand to the lesion, creating a ss/dsDNA junction at the DPC, a DNA structure allowing activation of SPRTN. Thus, the structure-specific activity of SPRTN enables controlled repair of various DPCs and allows its coupling to processes such as replication.

SPRTN achieves precision through a flexible, bipartite strategy based on two distinct DNA binding interfaces. SPRTN binds efficiently to DPCs within dsDNA (Figure 1H). However, binding alone is not sufficient to induce substrate cleavage. This may explain why chromatin proteins are not subjected to random cleavage by SPRTN *in vivo*. Induction of activity requires simultaneous engagement of ZBD and BR with DNA, which is only possible when the DNA has single- and double-stranded character. Our NMR analysis shows that the BR mediates sequence-independent electrostatically driven interactions with the negatively charged phosphate backbone of the dsDNA. In contrast, the ZBD binds to ssDNA—either to unpaired DNA bases at ss/dsDNA junctions and bubbles or unpaired bases formed by unwinding/breathing of the terminal base pairs at DNA nicks or dsDNA ends (Figure 6B). The exact molecular nature of the resulting activation remains to be determined, but previous results suggest that it involves conformational changes within SPRTN (Stingele et al., 2016). In agreement, the ZBD appears to constrain access to SPRTN's active site and would likely need to move aside for efficient substrate processing (Li et al., 2019). Taken together, the principles discovered here shift the current paradigm that DPC proteases are non-specific enzymes. On the contrary, our data demonstrate that SPRTN is a precise tool whose activation is spatially restricted, only allowing DPC cleavage in a very narrow window around the activating DNA structure. Furthermore, our results raise interesting questions regarding recruitment of SPRTN to sites of DPC formation in cells. SPRTN appears to have no specific affinity for its target structures. For example, it is activated similarly by a short DNA hairpin and 15-mer duplex DNA despite binding more strongly to dsDNA (Figures 2F and S2B). Thus, we favor a model in which SPRTN is initially recruited via protein-protein interactions and not through DNA binding. In agreement, it has been proposed that recruitment of SPRTN to chromatin upon formaldehyde exposure requires a ubiquitylation signal (Borgermann et al., 2019). Moreover, SPRTN recruitment to TOP1 DPCs depends on direct interaction between the protease and the adaptor protein TEX264 (Fielden et al., 2020). Hence, initial recruitment appears to be highly context-dependent. When recruited, SPRTN can utilize its non-specific DNA binding ability to scan the DNA in the vicinity for the presence of activating structures, which then trigger local activation of the protease and concurrent cleavage of protein adducts.

Our data raise the intriguing additional possibility that DPCs can be made "degradable" by DNA nicking or by creating a DNA bubble, which would be sufficient to allow activation of SPRTN and cleavage of the protein adduct. In this context, it is tempting to speculate that bubble-generating processes, such as transcription, might enable activation of SPRTN. In line with

CellPress
OPEN ACCESS

Molecular Cell
Article

this idea, genetic evidence obtained in flies and worms suggest that SPRTN does not act exclusively in a replication-dependent manner (Delabaere et al., 2014; Stingele et al., 2016). Finally, recent revelations of additional cellular proteases acting on DPCs raise the exciting possibility that specific proteases target DPCs in specific DNA contexts, analogous to cleavage of diverse DNA structures by various structure-specific endonucleases (Borgermann et al., 2019; Dehé and Gaillard, 2017; Kojima et al., 2020; Serbyn et al., 2020; Svoboda et al., 2019). To understand the increasing complexity of DPC repair, it will be paramount to understand the *in vitro* specificity of these enzymes, which appear to have distinct but also partially overlapping functions *in vivo*. Given that these enzymes protect cells against various chemotherapeutic agents, they constitute promising novel drug targets to serve as adjuvants for anti-cancer therapies.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell Lines
- METHOD DETAILS
  - Purification of Recombinant SPRTN
  - DNAs for Activation Assays
  - Protein-Oligonucleotide Conjugation
  - SPRTN Autocleavage/Histone H1 Cleavage Assays
  - Model DNA-Protein Crosslink Cleavage Assays
  - DNA Binding Assays
  - DNA-Protein Crosslink Binding Assays
  - Cellular Autocleavage Assay
  - Strep-Tactin Pull-down
  - Cas9/gRNA RNP Transfection and qPCR Analysis
  - Chromatin Fractionation
  - Immunofluorescence Staining
  - NMR Spectroscopy
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.molcel.2020.08.003.

ACKNOWLEDGMENTS

We thank K. Ramadan for providing the pNIC-ZB SPRTN plasmid; D. Yaneva for help with protein purification; and S. Panier, G. Hewitt, R. Bellelli, and P. Wolf for discussions and comments on the manuscript. H.-Y.L. is supported by the Peter and Traudl Engelhorn Foundation, S.Z. by the LMU – China Scholarship Council Program, and A.C.A. and P.W. by the International Max-Planck Research School for Molecular Life Sciences. J.S. is supported by the European Research Council (ERC Starting Grant 801750 DNAProteinCrosslinks), by the Alfried Krupp Prize for Young University Teachers awarded by the Alfried-Krupp von Bohlen und Halbach-Stiftung, and the Deutsche Forschungsgemeinschaft (CRC1064). L.T.J. is supported by the European Research Council (ERC Starting Grant SOLID). J.S., L.T.J., and M.S. acknowledge support from the Center for Integrated Protein Science Munich (CIPSM).

AUTHOR CONTRIBUTIONS

Conceptualization, H.K.R. and J.S.; Investigation, H.K.R., H.-S.K., M.J.G., A.K., H.-Y.L., S.Z., A.C.A., P.W., E.F., L.T.J., and J.S.; Writing – Original Draft, J.S.; Writing – Review & Editing, H.K.R., H.-S.K., M.S., and J.S.; Funding Acquisition, M.S. and J.S.; Supervision, M.S. and J.S.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: April 16, 2020
Revised: July 3, 2020
Accepted: August 4, 2020
Published: August 26, 2020

REFERENCES

Barker, S., Weinfeld, M., and Murray, D. (2005). DNA-protein crosslinks: their induction, repair, and biological consequences. Mutat. Res. *589*, 111–135.

Bellelli, R., Castellone, M.D., Guida, T., Limongello, R., Dathan, N.A., Merolla, F., Cirafici, A.M., Affuso, A., Masai, H., Costanzo, V., et al. (2014). NCOA4 transcriptional coactivator inhibits activation of DNA replication origins. Mol. Cell *55*, 123–137.

Borgermann, N., Ackermann, L., Schwertman, P., Hendriks, I.A., Thijssen, K., Liu, J.C., Lans, H., Nielsen, M.L., and Mailand, N. (2019). SUMOylation promotes protective responses to DNA-protein crosslinks. EMBO J. *38*, e101496.

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

Chen, S.H., Chan, N.-L., and Hsieh, T.S. (2013). New mechanistic and functional insights into DNA topoisomerases. Annu. Rev. Biochem. *82*, 139–170.

Cortes Ledesma, F., El Khamisy, S.F., Zuma, M.C., Osborn, K., and Caldecott, K.W. (2009). A human 5′-tyrosyl DNA phosphodiesterase that repairs topoisomerase-mediated DNA damage. Nature *461*, 674–678.

Davis, E.J., Lachaud, C., Appleton, P., Macartney, T.J., Näthke, I., and Rouse, J. (2012). DVC1 (C1orf124) recruits the p97 protein segregase to sites of DNA damage. Nat. Struct. Mol. Biol. *19*, 1093–1100.

Dehé, P.-M., and Gaillard, P.H.L. (2017). Control of structure-specific endonucleases to maintain genome stability. Nat. Rev. Mol. Cell Biol. *18*, 315–330.

Delabaere, L., Orsi, G.A., Sapey-Triomphe, L., Horard, B., Couble, P., and Loppin, B. (2014). The Spartan ortholog maternal haploid is required for paternal chromosome integrity in the Drosophila zygote. Curr. Biol. *24*, 2281–2287.

Dokshin, G.A., Davis, G.M., Sawle, A.D., Eldridge, M.D., Nicholls, P.K., Gourley, T.E., Romer, K.A., Molesworth, L.W., Tatnell, H.R., Ozturk, A.R., et al. (2020). GCNA Interacts with Spartan and Topoisomerase II to Regulate Genome Stability. Dev. Cell *52*, 53–68.e6.

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

Fielden, J., Wiseman, K., Torrecilla, I., Li, S., Hume, S., Chiang, S.-C., Ruggiano, A., Narayan Singh, A., Freire, R., Hassanieh, S., et al. (2020). TEX264 coordinates p97- and SPRTN-mediated resolution of topoisomerase 1-DNA adducts. Nat. Commun. *11*, 1274.

Fu, Y.V., Yardimci, H., Long, D.T., Ho, T.V., Guainazzi, A., Bermudez, V.P., Hurwitz, J., van Oijen, A., Schärer, O.D., and Walter, J.C. (2011). Selective bypass of a lagging strand roadblock by the eukaryotic replicative DNA helicase. Cell *146*, 931–941.

Hart, T., Chandrashekhar, M., Aregger, M., Steinhart, Z., Brown, K.R., MacLeod, G., Mis, M., Zimmermann, M., Fradet-Turcotte, A., Sun, S., et al.

Molecular Cell
Article

(2015). High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. Cell *163*, 1515–1526.

Jackson, S.P., and Bartek, J. (2009). The DNA-damage response in human biology and disease. Nature *461*, 1071–1078.

Jae, L.T., Raaben, M., Herbert, A.S., Kuehne, A.I., Wirchnianski, A.S., Soh, T.K., Stubbs, S.H., Janssen, H., Damme, M., Saftig, P., et al. (2014). Virus entry. Lassa virus entry requires a trigger-induced receptor switch. Science *344*, 1506–1510.

Kojima, Y., Machida, Y., Palani, S., Caulfield, T.R., Radisky, E.S., Kaufmann, S.H., and Machida, Y.J. (2020). FAM111A protects replication forks from protein obstacles via its trypsin-like domain. Nat. Commun. *11*, 1318.

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

Larsen, N.B., Gao, A.O., Sparks, J.L., Gallina, I., Wu, R.A., Mann, M., Räschle, M., Walter, J.C., and Duxin, J.P. (2019). Replication-Coupled DNA-Protein Crosslink Repair by SPRTN and the Proteasome in Xenopus Egg Extracts. Mol. Cell *73*, 574–588.e7.

Lee, W., Tonelli, M., and Markley, J.L. (2015). NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. Bioinformatics *31*, 1325–1327.

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

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

Lindahl, T. (1993). Instability and decay of the primary structure of DNA. Nature *362*, 709–715.

Lopez-Mosqueda, J., Maddi, K., Prgomet, S., Kalayil, S., Marinovic-Terzic, I., Terzic, J., and Dikic, I. (2016). SPRTN is a mammalian DNA-binding metalloprotease that resolves DNA-protein crosslinks. eLife *5*, e21491.

Lu, K., Ye, W., Zhou, L., Collins, L.B., Chen, X., Gold, A., Ball, L.M., and Swenberg, J.A. (2010). Structural characterization of formaldehyde-induced cross-links between amino acids and deoxynucleosides and their oligomers. J. Am. Chem. Soc. *132*, 3388–3399.

Maskey, R.S., Kim, M.S., Baker, D.J., Childs, B., Malureanu, L.A., Jeganathan, K.B., Machida, Y., van Deursen, J.M., and Machida, Y.J. (2014). Spartan deficiency causes genomic instability and progeroid phenotypes. Nat. Commun. *5*, 5744.

Maskey, R.S., Flatten, K.S., Sieben, C.J., Peterson, K.L., Baker, D.J., Nam, H.-J., Kim, M.S., Smyrk, T.C., Kojima, Y., Machida, Y., et al. (2017). Spartan deficiency causes accumulation of Topoisomerase 1 cleavage complexes and tumorigenesis. Nucleic Acids Res. *45*, 4564–4576.

Mórocz, M., Zsigmond, E., Tóth, R., Enyedi, M.Z., Pintér, L., and Haracska, L. (2017). DNA-dependent protease activity of human Spartan facilitates replication of DNA-protein crosslink-containing DNA. Nucleic Acids Res. *45*, 3172–3188.

Mosbech, A., Gibbs-Seymour, I., Kagias, K., Thorslund, T., Beli, P., Povlsen, L., Nielsen, S.V., Smedegaard, S., Sedgwick, G., Lukas, C., et al. (2012). DVC1 (C1orf124) is a DNA damage-targeting p97 adaptor that promotes ubiquitin-dependent responses to replication blocks. Nat. Struct. Mol. Biol. *19*, 1084–1092.

Mulder, F.A., Schipper, D., Bott, R., and Boelens, R. (1999). Altered flexibility in the substrate-binding site of related native and engineered high-alkaline Bacillus subtilisins. J. Mol. Biol. *292*, 111–123.

Nakano, T., Ouchi, R., Kawazoe, J., Pack, S.P., Makino, K., and Ide, H. (2012). T7 RNA polymerases backed up by covalently trapped proteins catalyze highly error prone transcription. J. Biol. Chem. *287*, 6562–6572.

Nakano, T., Miyamoto-Matsubara, M., Shoulkamy, M.I., Salem, A.M., Pack, S.P., Ishimi, Y., and Ide, H. (2013). Translocation and stability of replicative DNA helicases upon encountering DNA-protein cross-links. J. Biol. Chem. *288*, 4649–4658.

Nakano, T., Xu, X., Salem, A.M.H., Shoulkamy, M.I., and Ide, H. (2017). Radiation-induced DNA-protein cross-links: Mechanisms and biological significance. Free Radic. Biol. Med. *107*, 136–145.

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

Pommier, Y. (2006). Topoisomerase I inhibitors: camptothecins and beyond. Nat. Rev. Cancer *6*, 789–802.

Pouliot, J.J., Yao, K.C., Robertson, C.A., and Nash, H.A. (1999). Yeast gene for a Tyr-DNA phosphodiesterase that repairs topoisomerase I complexes. Science *286*, 552–555.

Quiñones, J.L., Thapar, U., Yu, K., Fang, Q., Sobol, R.W., and Demple, B. (2015). Enzyme mechanism-based, oxidative DNA-protein cross-links formed with DNA polymerase β in vivo. Proc. Natl. Acad. Sci. USA *112*, 8602–8607.

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

Sattler, M., Schleucher, J., and Griesinger, C. (1999). Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Prog. Nucl. Magn. Reson. Spectrosc. *34*, 93–158.

Serbyn, N., Noireterre, A., Bagdiul, I., Plank, M., Michel, A.H., Loewith, R., Kornmann, B., and Stutz, F. (2020). The Aspartic Protease Ddi1 Contributes to DNA-Protein Crosslink Repair in Yeast. Mol. Cell *77*, 1066–1079.e9.

Setlow, R.B., Regan, J.D., German, J., and Carrier, W.L. (1969). Evidence that xeroderma pigmentosum cells do not perform the first step in the repair of ultraviolet damage to their DNA. Proc. Natl. Acad. Sci. USA *64*, 1035–1041.

Sparks, J.L., Chistol, G., Gao, A.O., Räschle, M., Larsen, N.B., Mann, M., Duxin, J.P., and Walter, J.C. (2019). The CMG Helicase Bypasses DNA-Protein Cross-Links to Facilitate Their Repair. Cell *176*, 167–181.e21.

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

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

Stingele, J., Bellelli, R., Alte, F., Hewitt, G., Sarek, G., Maslen, S.L., Tsutakawa, S.E., Borg, A., Kjær, S., Tainer, J.A., 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., and Boulton, S.J. (2017). Mechanisms of DNA-protein crosslink repair. Nat. Rev. Mol. Cell Biol. *18*, 563–573.

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

Toth, A., Hegedus, L., Juhasz, S., Haracska, L., and Burkovics, P. (2017). The DNA-binding box of human SPARTAN contributes to the targeting of Polη to DNA damage sites. DNA Repair (Amst.) *49*, 33–42.

Vaz, B., Popovic, M., Newman, J.A., Fielden, J., Aitkenhead, H., Halder, S., Singh, A.N., Vendrell, I., Fischer, R., Torrecilla, I., et al. (2016). Metalloprotease SPRTN/DVC1 Orchestrates Replication-Coupled DNA-Protein Crosslink Repair. Mol. Cell *64*, 704–719.

CellPress
OPEN ACCESS

Molecular Cell
Article

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-Strep-tag II antibody	Abcam	Cat#ab76949; RRID:AB_1524455
Anti-Histone H3 antibody	Abcam	Cat#ab10799; RRID:AB_470239
Anti-GFP from mouse IgG1κ (used for YFP detection)	Sigma	Cat#11814460001; RRID:AB_390913
GFP antibody rabbit polyclonal (used for YFP detection)	Chromotek	Cat#PABG1; RRID:AB_2749857
GAPDH (14C10) Rabbit mAb	Cell Signaling	Cat#2118; RRID:AB_561053
Anit-H1.10 antibody	Abcam	Cat#ab11079; RRID:AB_2295032
Anti-SPRTN mAB (6F2)	Stingele lab	Clone6F2
Goat Anti-Rat Immunoglobulins/HRP	Sigma	Cat#A9037; RRID:AB_258429
Goat Anti-Mouse Immunoglobulins/HRP	Dako	Cat#P0447; RRID:AB_2617137
Swine Anti-Rabbit Immunoglobulins/HRP	Dako	Cat#P0399; RRID:AB_2617141
Goat anti-Mouse IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 488	Thermo Scientific	Cat#A-11001; RRID:AB_2534069
Bacterial and Virus Strains		
BL21(DE3)	Thermo Scientific	Cat#C600003
Chemicals, Peptides, and Recombinant Proteins		
16% Formaldehyde (w/v), Methanol-free	Thermo Scientific	Cat#28906
Instant*Blue*	Sigma	Cat#ISB1L
Doxycycline Hyclate	Sigma	Cat#D9891
ProLong Gold Antifade Reagent	Thermo Fisher	Cat#P10144
DAPI Solution	Thermo Fisher	Cat#62248
4x NuPAGE LDS sample buffer	Thermo Scientific	Cat#NP0007
Phusion HF enzyme	NEB	Cat#M0530
UltraPure BSA	Thermo Scientific	Cat#AM2616
Histone H1° Human	NEB	Cat#M2501S
Protein G	BioVision	Cat#6510
Lipofectamine 2000	Thermo Scientific	Cat#11668030
IGEPAL	Sigma	Cat#I8896
Biotin	IBA Lifesciences	Cat#2-1016-005
Pefabloc SC	Merck	Cat#11585916001
TCEP	ROTH	Cat#HN95.2
cOmplete EDTA-free protease inhibitor cocktail	Merck	Cat#4693132001
Critical Commercial Assays		
proFIRE Amine Coupling Kit	Dynamic Biosensors	Cat#PF-NH2-1
NucleoSpin® Gel and PCR Clean-up	MACHEREY-NAGEL	Cat#740609
Quant-iT PicoGreen dsDNA Assay Kit	Thermo Scientific	Cat#P11496
GeneJET Genomic DNA purification kit	Thermo Scientific	Cat#K0722
SsoAdvanced Universal SYBR Green Supermix	Bio-Rad	Cat#1725271

(*Continued on next page*)

Molecular Cell
Article

CellPress
OPEN ACCESS

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental Models: Cell Lines		
Human HeLa Flp-In-T-REx	The Francis Crick Institute Cell Services	N/A
Human HAP1	Thijn Brummelkamp, NKI Amsterdam	N/A
Oligonucleotides		
Oligonucleotide sequences used in this study are provided in Table S1	N/A	N/A
Recombinant DNA		
Ds ΦX174 phage DNA RFI	NEB	Cat#N3021S
Ss ΦX174 phage DNA virion	NEB	Cat#N3023S
pMAX-GFP	LONZA	Cat#VDC-1040
pNIC-STREP-ZB-SPRTN-WT	This study	N/A
pNIC-STREP-ZB-SPRTN-Equation (E112Q)	This study	N/A
pNIC-STREP-ZB-SPRTN-ZBD^{1*} (Y179A_W197A)	This study	N/A
pNIC-STREP-ZB-SPRTN-ZBD^{2*} (R185A)	This study	N/A
pNIC-STREP-ZB-SPRTN-BR* (K220A_K221E_G222A_K223A)	This study	N/A
pNIC-STREP-ZB-SPRTN-EQ-ZBD^{1*} (E112Q_Y179A_W197A)	This study	N/A
pNIC-STREP-ZB-SPRTN-EQ-ZBD^{2*} (E112Q_R185A)	This study	N/A
pNIC-STREP-ZB-SPRTN-EQ-BR* (E112Q_K220A_K221E_G222A_K223A)	This study	N/A
pNIC-STREP-ZB-SPRTN-EQ-ZBD^{2*}-BR* (E112Q_R185A_K220A_K221E_G222A_K223A)	This study	N/A
pNIC-STREP-ZB-SPRTN-ZBD-BR (aa151-245)	This study	N/A
pNIC-STREP-ZB-SPRTN-ZBD (aa151-215)	This study	N/A
pcDNA5-FRT/TO-YFP-SPRTN-WT-Strep	Stingele et al., 2016	N/A
pcDNA5-FRT/TO-YFP-SPRTN-Equation (E112Q)-Strep	Stingele et al., 2016	N/A
pcDNA5-FRT/TO-YFP-SPRTN-ZBD^{1*} (Y179A_W197A)-Strep	This study	N/A
pcDNA5-FRT/TO-YFP-SPRTN-ZBD^{2*} (R185A)-Strep	This study	N/A
pcDNA5-FRT/TO-YFP-SPRTN-BR* (K220A_K221E_G222A_K223A)-Strep	This study	N/A
pOG44	Thermo Scientific	Cat#V600520
pBABE-puro	Addgene	Cat#1764
pBABE-puro-SPRTN-WT-Strep	This study	N/A
pBABE- puro-SPRTN-Equation (E112Q)-Strep	This study	N/A
pBABE- puro-SPRTN-ZBD^{1*} (Y179A_W197A)-Strep	This study	N/A
pBABE- puro-SPRTN-ZBD^{2*} (R185A)-Strep	This study	N/A
Software and Algorithms		
Prism 7	GraphPad Software	https://www.graphpad.com/
ImageJ	NIH	https://imagej.net/Fiji/Downloads

(*Continued on next page*)

CellPress
OPEN ACCESS

Molecular Cell
Article

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Adobe Photoshop CC2018	Adobe	https://www.adobe.com/es/products/photoshop.html
Other		
HiTrap Heparin HP affinity columns	GE Healthcare	Cat#17040701
PD-10 Desalting columns	GE Healthcare	Cat#17085101
Strep-Tactin®XT Superflow® high capacity cartridges	IBA Lifesciences	Cat#2-4026-001
HiLoad 16/600 Superdex 200 pg column	GE Healthcare	Cat#GE28-9893-35
10 kDa cutoff Amicon Ultra centrifugal filters	Merck	Cat#UFC801096
GFP-Trap Magnetic Agarose	Chromotek	Cat#gtma-10

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Julian Stingele (stingele@genzentrum.lmu.de).

Materials Availability

All plasmids are available on request.

Data and Code Availability

This study did not generate code or reposited datasets.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell Lines

Human HeLa Flp-In-T-Rex (female) cells were obtained from and authenticated by Francis Crick Institute Cell Services. HeLa Flp-In-T-Rex cells expressing YFP-SPRTN-Twin-Strep-tag variants were generated using the Flp-In-T-REx system (Thermo Fisher) using pOG44 and the respective pcDNA5-FRT/TO plasmids according to manufacturer's instructions and grown in Dulbecco's modified Eagle Medium (DMEM) supplemented with 10% (v/v) fetal bovine serum (FBS). Protein expression was induced by overnight incubation with doxycycline (final concentration 1 μg/mL). Human HAP1 (male) cells (generated and kindly provided by Thijn Brummelkamp, NKI Amsterdam) stably expressing SPRTN variants were generated by transduction as described preciously (Jae et al., 2014). In brief, HEK293T cells were transfected with pBabe-puro (Addgene #1764) empty vector or containing the coding sequence for SPRTN variants together with pAdvantage (Clontech) and the standard retroviral packaging plasmids VSV-g and Gag-pol. 48h after transfection, viral supernatant was collected and HAP1 cells were transduced with the 0.45 μm filtrate in the presence of 8 μg/mL protamine sulfate (Sigma). After 24 h transduced HAP1 cells were selected with 1 μg/mL puromycin (Invivogen).

METHOD DETAILS

Purification of Recombinant SPRTN

The sequence of full-length human *SPRTN* in the pNIC-ZB-SPRTN plasmid (Vaz et al., 2016) was replaced with a version codon-optimized for bacterial expression and the His-tag was replaced by a Twin-Strep-tag. For protein expression plasmids were transformed into BL21(DE3) *Escherichia coli* cells and grown at 37°C in Terrific broth (TB) medium until they reached OD 0.7. Protein expression was induced by addition of 0.5 mM IPTG over night at 18°C. Next, cells were harvested, resuspended in buffer A (50 mM HEPES/KOH pH 7.2, 500 mM KCl, 1 mM $MgCl_2$, 10% Glycerol, 0.1% IGEPAL, 0.04 mg/mL Pefabloc SC, cOmplete EDTA-free protease inhibitor cocktail tablets, 1 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP), pH 7.2) and lysed by sonication. All steps were carried out at 4°C. Cell lysate was incubated with benzonase (45 U/ mL lysate) for 30 min on ice prior to the removal of cell debris by centrifugation at 18000 g for 30 min. Cleared supernatant was applied to Strep-Tactin®XT Superflow® high capacity cartridges, 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, pH 7.2). Proteins were eluted in 6 CV buffer C (50 mM HEPES/KOH pH 7.2, 500 mM KCl, 10% Glycerol, 1 mM TCEP and 50 mM Biotin, pH 7.2). Eluted proteins were further applied to HiTrap Heparin HP affinity columns 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, pH 7.2). Eluted fractions containing recombinant SPRTN protein were desalted against buffer B using PD-10 desalting columns. The affinity tag was cleaved off over night at 4°C by the addition of

Molecular Cell
Article

CellPress
OPEN ACCESS

His-tagged TEV protease with 1:10 mass ratio. Cleaved recombinant SPRTN protein was further purified by size exclusion chromatography using a HiLoad 16/600 Superdex 200 pg column equilibrated in buffer E (50 mM HEPES/KOH pH 7.2, 500 mM KCl, 10% Glycerol, 0.5 mM TCEP, pH 7.2). Eluted proteins were concentrated with 10 kDa cutoff Amicon Ultra centrifugal filters before snap-freezing in liquid nitrogen and storing at −80°C. Proteins used for NMR analysis were expressed in ^{15}N or ^{13}C-/^{15}N-containing media and purified as described above including minor changes. After cleavage of the affinity tag the samples were applied again on Strep-Tactin®XT Superflow® high capacity cartridges. The flow through was collected and further purified by size exclusion chromatography.

DNAs for Activation Assays

Oligonucleotides were used as follows: 60-mer ssDNA = oJS_63, 60-mer ds DNA = oJS_63 + oJS_64, 15-mer hairpin = oJS_106, 15-mer hairpin mutant = oJS_119, 15-mer hairpin dsDNA = oJS_106 + oJS_107, 15-mer hairpin mutant dsDNA = oJS_119 + oJS_120 (sequences are provided in Table S1). Single-stranded DNAs were incubated for 10 min at 95°C before snap-cooling on ice. Double-stranded DNAs were annealed in a PCR machine (5 min incubation at 95°C followed by a decrease in temperature of 2°C/min until 10°C was reached). A standard PCR protocol using Phusion HF enzyme was used to generate PCR fragments with double-stranded ΦX174 (RF I) DNA as template and the following primer combinations: oJS_31 + oJS_30, oJS_122 + oJS_30, oJS_35 + oJS_34, oJS_123 + oJS_34. PCR fragments were gel purified (NucleoSpin Gel and PCR Clean-up) before used in activation assays. Denaturation of double-stranded DNA circles (ΦX174 (RF I) or pMAX-GFP) was induced by incubation at 95°C for 10 min followed by immediate snap-cooling on ice. Successful denaturation was confirmed using PicoGreen a fluorescent dye specific for double-stranded DNA.

Protein-Oligonucleotide Conjugation

Protein G was crosslinked to oligonucleotides X1, X15, X30 and $C_3A_{11}XA_{12}C_3$, which contained a 5′-Cy5 label and a 3′ phosphate group. An Amino-C6-dT was incorporated at the intended crosslinking position and its terminal primary amine group was further processed to yield a reduced thiol (SH-C9-dT) (Ella Biotech GmbH). Conjugation was carried out with 3 nmol oligonucleotide and 50 μL of 5 mg/mL Protein G using the proFIRE Amine Coupling Kit. During the coupling reaction, the terminal thiol group of SH-C9-dT was further functionalized to an NHS-ester, which can react with a primary amine group of proteins. Crosslinked oligonucleotides (conjugates) were purified by ion exchange chromatography using a proFIRE device (Dynamic Biosensors) according to the manufacturer's instructions. Next, the conjugates were desalted against storage buffer (50 mM HEPES/KOH pH 7.2, 100 mM KCl and 10% Glycerol, pH 7.2) and then snap-frozen in liquid nitrogen and stored at −80°C. Conjugate concentration was determined by measuring Cy5 absorbance with a SpectraMax Paradigm Multi-Mode Detection platform (Molecular Devices). The conjugates were used to generate model DPCs by annealing complementary reverse oligonucleotides (see scheme in Table S2 for details). Annealing was carried out directly prior to cleavage reactions or EMSAs. Conjugates were annealed with complementary reverse oligonucleotides by mixing them at a ratio of 1:1.2 (conjugate:oligonucleotide) in a reaction buffer of 25 mM HEPES/KOH pH 7.2, 50 mM KCl, 5% Glycerol, and 0.2 mg/mL BSA. Annealing was accomplished by incubating the reaction for 20 minutes at 37°C for X1, X15 or X30 conjugates. $C_3A_{11}XA_{12}C_3$ conjugates were annealed by incubating the reaction for 2 min at 37°C followed by a decrease in temperature of 1°C/min to 25°C.

SPRTN Autocleavage/Histone H1 Cleavage Assays

Reactions were performed at 25°C in 20 μl containing 500 nM SPRTN, 500 nM histone H1 and DNA (amount was kept constant in all assays and corresponded to 1 μM of a 60-mer oligonucleotide). The reaction buffer comprised 50 mM HEPES/KOH pH 7.2, 2.9% glycerol and either 80 or 150 mM KCl. Reactions were stopped by addition of 4 x LDS sample buffer supplemented with β-mercaptoethanol and boiling at 95°C for 10 min, resolved on 4%–12% Bis-Tris gradient gels using MOPS buffer and stained with InstantBlue or analyzed by western blotting using anti-SPRTN and anti-H1 antibodies and HRP-conjugated anti-rat IgG or anti-mouse IgG, respectively, as secondary antibodies. The intensity of western blots and scanned gels was adjusted globally using Adobe Photoshop. Cleavage reactions were quantified by dividing the amount of cleaved protein by the total amount of protein (cleaved and uncleaved) as determined by analysis of western blot results using ImageJ.

Model DNA-Protein Crosslink Cleavage Assays

Cleavage of model DPCs by SPRTN was performed in a reaction volume of 10 μl containing 5 nM SPRTN (or as indicated in the figure legend) and 25 nM DPC in a final reaction buffer of 17.5 mM HEPES/KOH pH 7.2, 80 mM KCl, 3.5% Glycerol, 5 mM TCEP and 0.1 mg/mL BSA. Reactions were incubated for 2 h at 25°C. 2 μL of 6x Orange G loading dye was added and cleaved DPC fragments were resolved on 20% TBE gels using 1X TBE as running buffer at 4°C. Gels were photographed using a BioRad Chemidoc MP system using filter settings for Cy5 fluorescence. The intensity of scanned gels was adjusted globally using ImageJ, which was also used to quantify cleavage by dividing the amount of cleaved conjugate by the total amount of conjugate (cleaved and uncleaved) and subtraction of background signal (determined from lanes without SPRTN).

DNA Binding Assays

Electrophoretic mobility shift assays (EMSAs) were used to analyze DNA binding of recombinant proteins. Assay composition was exactly as in SPRTN autocleavage assays with varying amounts of catalytically inactive SPRTN-E112Q. Binding reactions were

CellPress
OPEN ACCESS

Molecular Cell
Article

incubated for 20 min on ice prior to separation on 6% native PAGE gels with 0.5x TBE as running buffer at 4°C. Gels were photographed using a BioRad Chemidoc MP system using filter settings for Cy5 fluorescence. The intensity of the scanned images was adjusted globally using ImageJ.

DNA-Protein Crosslink Binding Assays

Electrophoretic mobility shift assays (EMSAs) were used to analyze binding of catalytically inactive SPRTN-E112Q variants to diverse model DPCs. Therefore 25 nM model DPC was incubated with varying concentrations of recombinant SPRTN proteins for 15 minutes on ice. The total reaction volume was kept to 10 μL with a final reaction buffer of 17.5 mM HEPES/KOH pH 7.2, 80 mM KCl, 3.5% Glycerol, 5 mM TCEP and 0.1 mg/mL BSA. SPRTN-bound DPCs were separated on 6% native PAGE gels in 0.5x TBE running buffer at 4°C. Gels were photographed using a BioRad Chemidoc MP system using filter settings for Cy5 fluorescence. The intensity of the scanned images was adjusted globally using ImageJ.

Cellular Autocleavage Assay

pcDNA5-FRT/TO plasmids encoding YFP-SPRTN-Strep variants (3 μg) were transiently transfected using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Protein expression was induced by overnight (16h) incubation with doxycycline (final concentration 1 μg/mL). SPRTN autocleavage was induced by treating with 200 μM Formaldehyde for 2 hours. Cells lysed on ice in 500 μl lysis buffer (50 mM HEPES pH 7.5, 1 M NaCl, 10% glycerol, 1% IGEPAL CA-630, 20 mM iodoacetamide, 0.04 mg/ml PefaBloc SC and cOmplete EDTA- free protease inhibitor cocktail tablets (1 tablet/50 ml)). After addition of benzonase (4U/ml), lysates were incubated for 30 min on ice. Lysates were cleared by centrifugation at 4°C and applied to 15 μL of GFP-Trap Magnetic Agarose (Chromotek) according to manufacturer's instructions. Finally, samples were resuspended in 40 μl 1X LDS-sample buffer, subjected analysis by SDS-PAGE and western blotting with anti-GFP antibody (PABG1, Chromotek) and peroxidase-conjugated anti-rabbit IgG as secondary antibody. Input samples were analyzed by SDS-PAGE and western blotting with anti-GAPDH antibody (Cell Signaling) and peroxidase-conjugated anti-rabbit IgG as secondary antibody.

Strep-Tactin Pull-down

Cells were lysed on ice in lysis buffer (10 mM Tris/HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 0.5% NP-40, 2 mM $MgCl_2$, 20 mM iodoacetamide, 0.04 mg/ml PefaBloc SC and cOmplete EDTA-free protease inhibitor cocktail tablets (1 tablet/50 ml)). After addition of benzonase (4U/ml), lysates were incubated for 30 min on ice. Lysates were cleared by centrifugation at 4°C and incubated with Strep-Tactin®XT Superflow® beads for 4h. Beads were washed three times with lysis buffer before resuspension in 30 μl 2x LDS-sample buffer. Finally, samples were subjected to analysis by SDS-PAGE and western blotting with Anti-Strep-tag II antibody (Abcam).

Cas9/gRNA RNP Transfection and qPCR Analysis

Human HAP1 cells expressing cDNA encoding C-terminally Strep-tagged SPRTN variants cells were electroporated with NLS-Cas9/gRNA RNPs using a 4D-Nucleofector (Lonza). In brief, crRNA1 and crRNA2 are incubated with tracRNA (95°C, 5 minutes), respectively, to generate gRNAs. gRNAs were mixed with NLS-Cas9 and incubated for 10 minutes at RT to generate RNPs. $1x10^6$ cells were resuspended in 20 μl Nucleofection Solution (Lonza, SE. Cell line 4D-Nucleofector X Kit). Suspended cells were then mixed with RNPs and electroporated (program EN-138). Cells were plated and samples collected every 48 hours after electroporation for genomic DNA extraction (GeneJET Genomic DNA purification kit, Thermo Scientific). The relative amount of KO and WT allele was monitored for each cell line at each time point by qPCR analysis. Each 10 μl reaction contained 20 ng genomic DNA, 0.4 μl forward and reverse primer (10 μM) and 5 μl SYBR Green Supermix (Bio-Rad). PCR reaction was performed in technical triplicates using primers amplifying either WT or KO allele. For analysis, Cq^{WT} was subtracted from Cq^{KO} to obtain ΔCq. $2^{-(\Delta Cq)}$ was calculated for each time point and normalized to the day 2 value ($2^{-(\Delta\Delta Cq)}$).

Chromatin Fractionation

Chromatin fractionation experiments were performed as described before (Bellelli et al., 2014). In brief, cells in the mid-exponential phase of growth were collected by scraping in ice-cold 1x phosphate-buffered saline (PBS). Cells were then equally split and either directly resuspended in 1x LDS buffer or incubated for 10 min in ice-cold CSK buffer (10 mM PIPES, 100 mM NaCl, 1.5 mM MgCl2, 5 mM EDTA, 300 mM sucrose and 0.5% Triton X-100, protease inhibitors and phosphatase inhibitors). Chromatin-bound proteins were isolated by low speed centrifugation (3,000 rpm, 3 min at 4°C). Finally, samples were subjected to analysis by SDS-PAGE and western blotting with Anti-Strep-tag II (Abcam) and anti-histone H3 (Sigma) antibodies.

Immunofluorescence Staining

For indirect immunofluorescence, cells were pre-extracted in CSK buffer containing 0.5% Triton X-100 (10 min on ice) and/or fixed in 4% paraformaldehyde, permeabilized with PGBT buffer (PBS, 0.2% fish skin gelatin, 0.5% BSA, 0.5% Triton X-100) (45 min at room temperature) and then incubated with anti-GFP antibody (Sigma) overnight at 4°C. Coverslips were then washed 3 times for 5 min in PGBT buffer and incubated with Alexa Fluor 488 goat anti-mouse antibody (Thermo Scientifc) and DAPI counterstaining (0.5 μg/ml) for 1h at room temperature. Coverslips were mounted in Prolong Gold Antifade Mountant (Thermo Fisher). Pictures were acquired with a ZEISS LSM710 confocal microscope.

Molecular Cell
Article

NMR Spectroscopy

NMR samples (non-labeled or uniformly $^{15}N^{13}C$-/^{15}N-labeled for SPRTN-ZBD/ZBD, non-labeled for dsDNA) were prepared at protein concentrations of 100 – 350 μM in three buffer conditions (100 mM KCl, 50 mM HEPES pH 7.5, 2 mM TCEP; 20 mM NaCl, 50 mM sodium phosphate pH 6.5, 2 mM TCEP; 500 mM KCl, 50 mM HEPES pH 7.5, 2 mM TCEP) with 10% D_2O added as lock signal. NMR experiments were recorded at 278 K and 298 K on 900-, 800-, 600-MHz Bruker Avance NMR spectrometers, equipped with cryogenic or room-temperature triple resonance gradient probes. NMR spectra were processed by TOPSPIN3.5 (Bruker), then analyzed using NMRFAM-SPARKY (Lee et al., 2015). Backbone resonance assignments of both SPRTN-ZBD and SPRTN-ZBD-BR were obtained from a uniformly ^{15}N,^{13}C-labeled protein employing standard triple resonance experiments HNCA, HNCACO, HNCACB and CBCA(CO)NH (Sattler et al., 1999). ^{1}H-^{15}N Heteronuclear NOE experiments were recorded on a 600-MHz spectrometer at 298 K with an interleaved manner with and without proton saturation. Imino resonances were obtained through 2D 1H-1H NOESY with mixing time of 150 - 200 msec at 278 K and 298 K on 600- and 900-MHz spectrometers. CSP values were calculated based on the following, $\Delta\delta_{HN,N} = \sqrt{\Delta\delta_{HN}^2 + (\Delta\delta_N/R_{scale})^2}$, where 6.5 was applied to the chemical shift change of ^{15}N as R_{scale} factor, as suggested previously (Mulder et al., 1999).

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses (unpaired t test) were performed using Prism 7 (GraphPad Software). Statistical details of each experiment (including the exact value of n, what n represents and precision measures) can be found in the figure legends.

Molecular Cell, Volume *80*

Supplemental Information

DNA Structure-Specific Cleavage of DNA-Protein Crosslinks by the SPRTN Protease

Hannah K. Reinking, Hyun-Seo Kang, Maximilian J. Götz, Hao-Yi Li, Anja Kieser, Shubo Zhao, Aleida C. Acampora, Pedro Weickert, Evelyn Fessler, Lucas T. Jae, Michael Sattler, and Julian Stingele

Figure S1 (related to Figure 1)

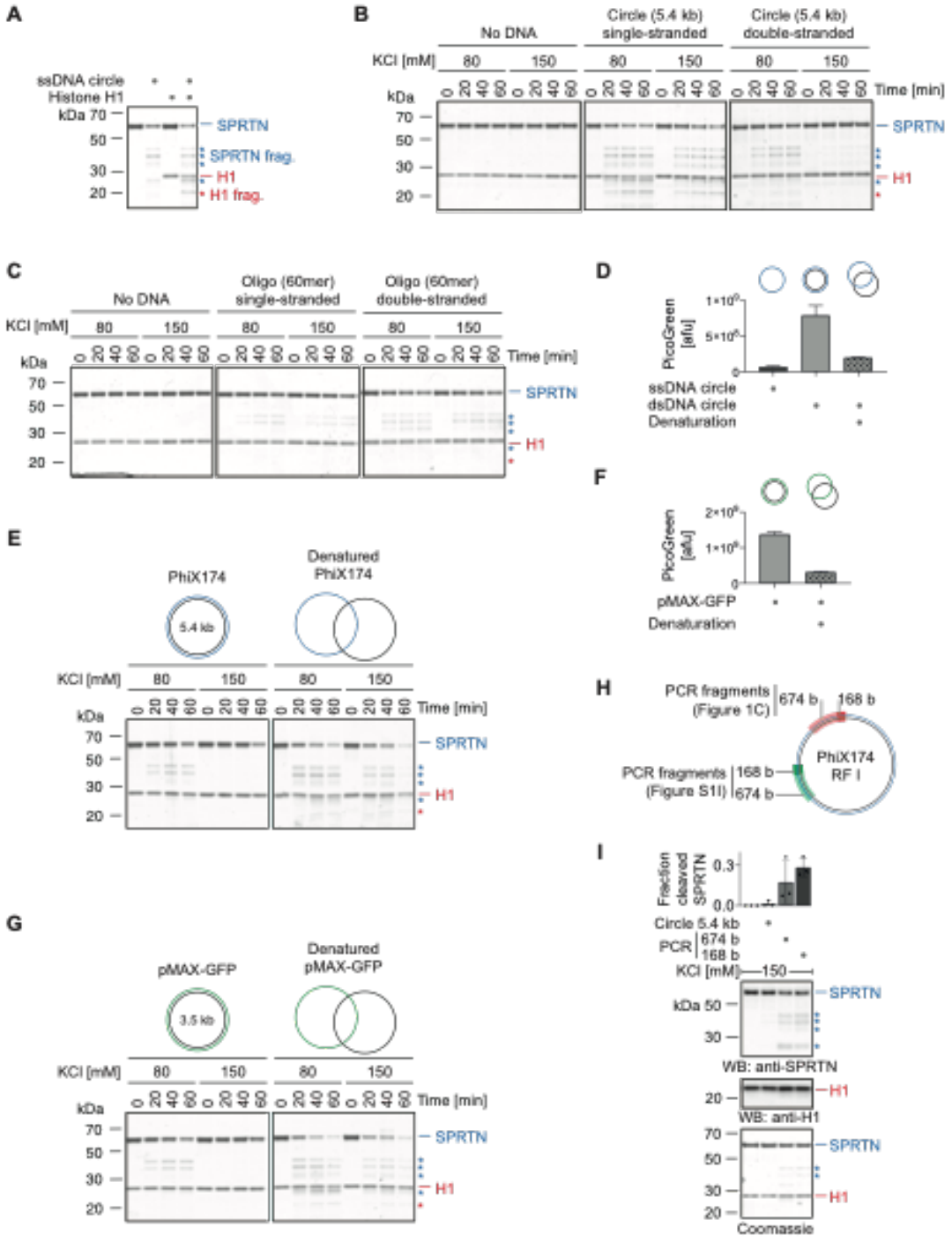


Figure S1. DNA length determines activation of SPRTN by double-stranded DNA, Related to Figure 1.

(A) SPRTN protease assay in the absence and presence of histone H1 to reveal H1 proteolytic fragments. Recombinant SPRTN (500 nM) - with and without histone H1 (500 nM) - was incubated alone or in the presence of ssDNA circles (ΦX174 virion) for 2 h at 25°C. Reactions were analysed by SDS-PAGE followed by Coomassie staining. Proteolytic fragments of SPRTN and H1 are indicated by asterisks.

(B-C) Kinetics of enzymatic reactions shown in Figure 1A. Recombinant SPRTN (500 nM) and histone H1 (500 nM) were incubated alone or in the presence of DNA (5.4 kb circles (B) or 60mer oligonucleotides (C), each either single- or double-stranded for the indicated amount of time at 25°C. DNA concentrations were 1 µM for 60mer oligonucleotides or the corresponding amount of circular DNA (11.4 nM). Reactions were analysed by SDS-PAGE followed by Coomassie staining. Cleaved fragments of SPRTN and H1 are indicated by asterisks.

(D) Denaturation of circular dsDNA (ΦX174 RFI) monitored by PicoGreen fluorescence. dsDNA was melted at 95°C prior to snap-cooling on ice. Denaturation was assessed using PicoGreen, a fluorescent dye specific for dsDNA.

(E) Kinetics of enzymatic reactions shown in Figure 1B. SPRTN (500 nM) and histone H1 (500 nM) were incubated in the presence of double- or denatured dsDNA (ΦX174 RFI) for the indicated amount of time at 25°C. Reactions were analysed by SDS-PAGE followed by Coomassie staining. Cleaved fragments of SPRTN and H1 are indicated by asterisks.

(F) Denaturation of circular dsDNA (pMAX-GFP) monitored by PicoGreen fluorescence. dsDNA was melted at 95°C prior to snap-cooling on ice. Denaturation was assessed using PicoGreen, a fluorescent dye specific for dsDNA.

(G) SPRTN (500 nM) and histone H1 (500 nM) were incubated in the presence of ds- or denatured dsDNA (pMAXGFP plasmid) for the indicated amount of time at 25°C. Reactions were analysed by SDS-PAGE followed by Coomassie staining. Cleaved fragments of SPRTN and H1 are indicated by asterisks.

(H) Schematic of the ΦX174 RFI regions amplified by PCR for use in Figure 1C and **I**.

(I) Fragment length determines activation of SPRTN by PCR-generated double-stranded DNA (as Figure 1C but using different DNA fragments). SPRTN (500 nM) and histone H1 (500 nM) were incubated in the presence of the indicated types of DNA for 2 h at 25°C. Cleaved fragments of SPRTN and H1 are indicated by asterisks. Quantification of Western blot results of SPRTN and histone H1 cleavage: values represent the mean ± SD of three independent experiments.

Figure S2 (related to Figure 2)

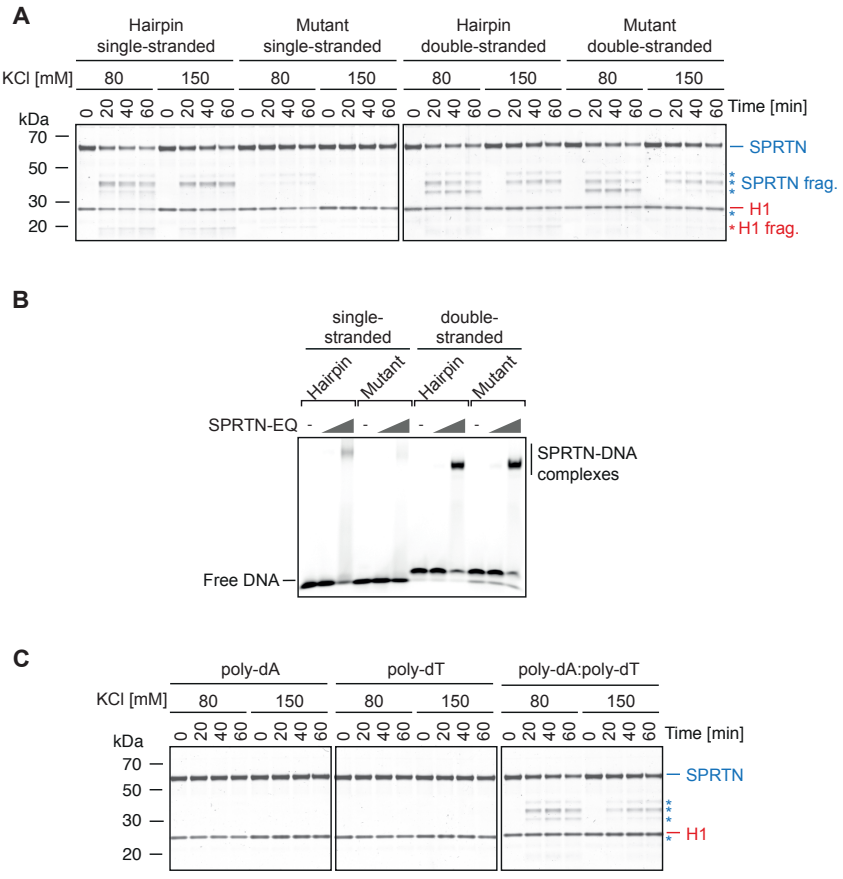


Figure S2. SPRTN is not activated by strictly single-stranded DNA, Related to Figure 2.

(A) Kinetics of enzymatic reactions shown in Figure 2F. Recombinant SPRTN (500 nM) and histone H1 (500 nM) were incubated alone or in the presence of the indicated DNAs (4 μM) for the indicated amount of time at 25°C. Reactions were analysed by SDS-PAGE followed by Coomassie staining. Cleaved fragments of SPRTN and H1 are indicated by asterisks.

(B) SPRTN binds to a DNA hairpin. EMSA assays were used to asses binding of catalytically inactive SPRTN E112Q (500 nM and 4 μM) to the indicated fluorescently-labelled oligonucleotides (4 μM).

(C) Kinetics of enzymatic reactions shown in Figure 2H. Reactions were incubated for the indicated amount of time at 25°C. Reactions were analysed by SDS-PAGE followed by Coomassie staining. Cleaved fragments of SPRTN and H1 are indicated by asterisks.

Figure S3 (related to Figure 3)

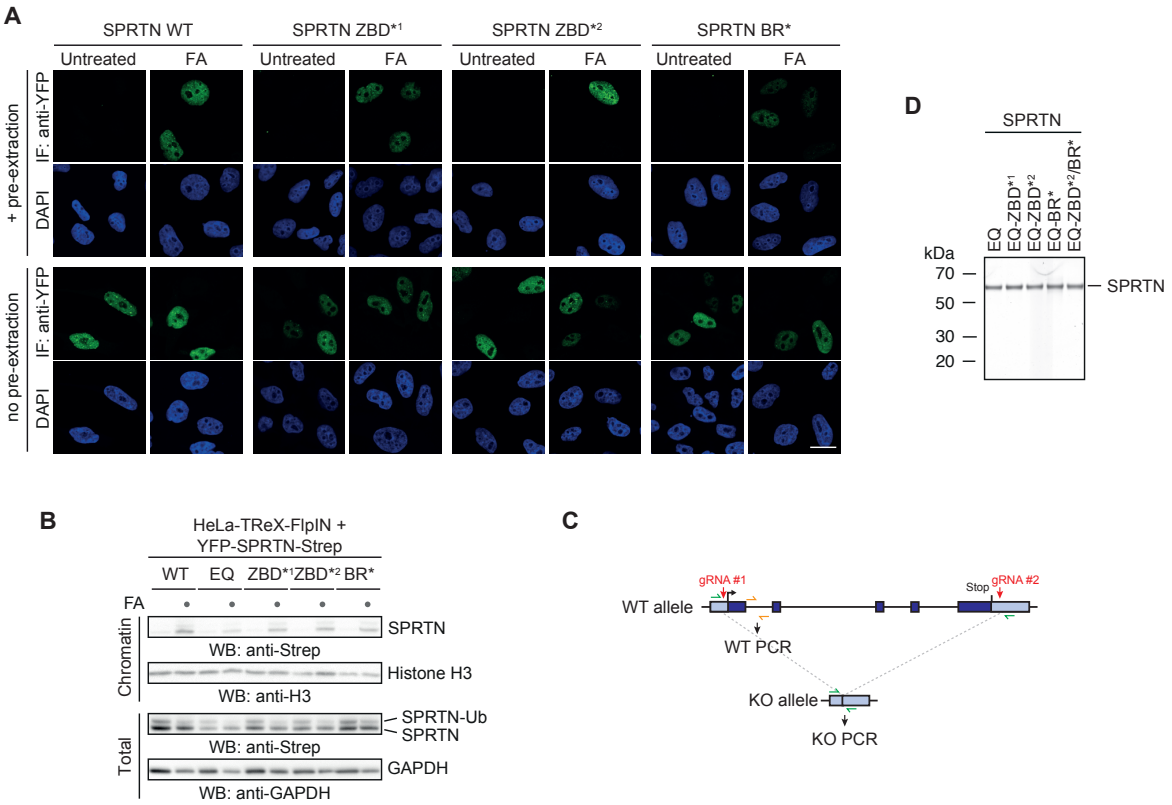


Figure S3. The DNA-dependent induction of SPRTN's protease activity requires two distinct DNA binding domains, Related to Figure 3.

(A) Recruitment of SPRTN to chromatin upon DPC induction is unaltered in ZBD and BR mutant variants. Doxycycline-inducible YFP-SPRTN-Strep HeLa Flp-In TRex cells expressing the indicated SPRTN variants were treated with 500 μM formaldehyde (FA) for 2 h prior to immunofluorescence-staining with or without pre-extraction prior to fixation. Scale bar represents 20 μm.

(B) Recruitment of SPRTN to chromatin upon DPC induction is unaltered in ZBD and BR mutant variants. Doxycycline-inducible YFP-SPRTN-Strep HeLa Flp-In TRex cells expressing the indicated SPRTN variants were treated with 500 μM formaldehyde (FA) for 2 h prior to chromatin fractionation and Western blotting against Strep-tag, histone H3 and GAPDH.

(C) Schematic depiction of the knock-out and genotyping strategy used in Figure 3E. NLS-Cas9/gRNA complexes targeting the UTRs of the endogenous SPRTN allele were transfected in HAP1 cells complemented with empty vector or different SPRTN variants. The abundance of the resulting SPRTN KO allele was then monitored over time using qPCR with the indicated primers.

(D) Coomassie-stained SDS-PAGE showing purified SPRTN-E112Q variants used in electrophoretic mobility shift assays (Figure 3I-K).

Figure S4 (related to Figure 4)

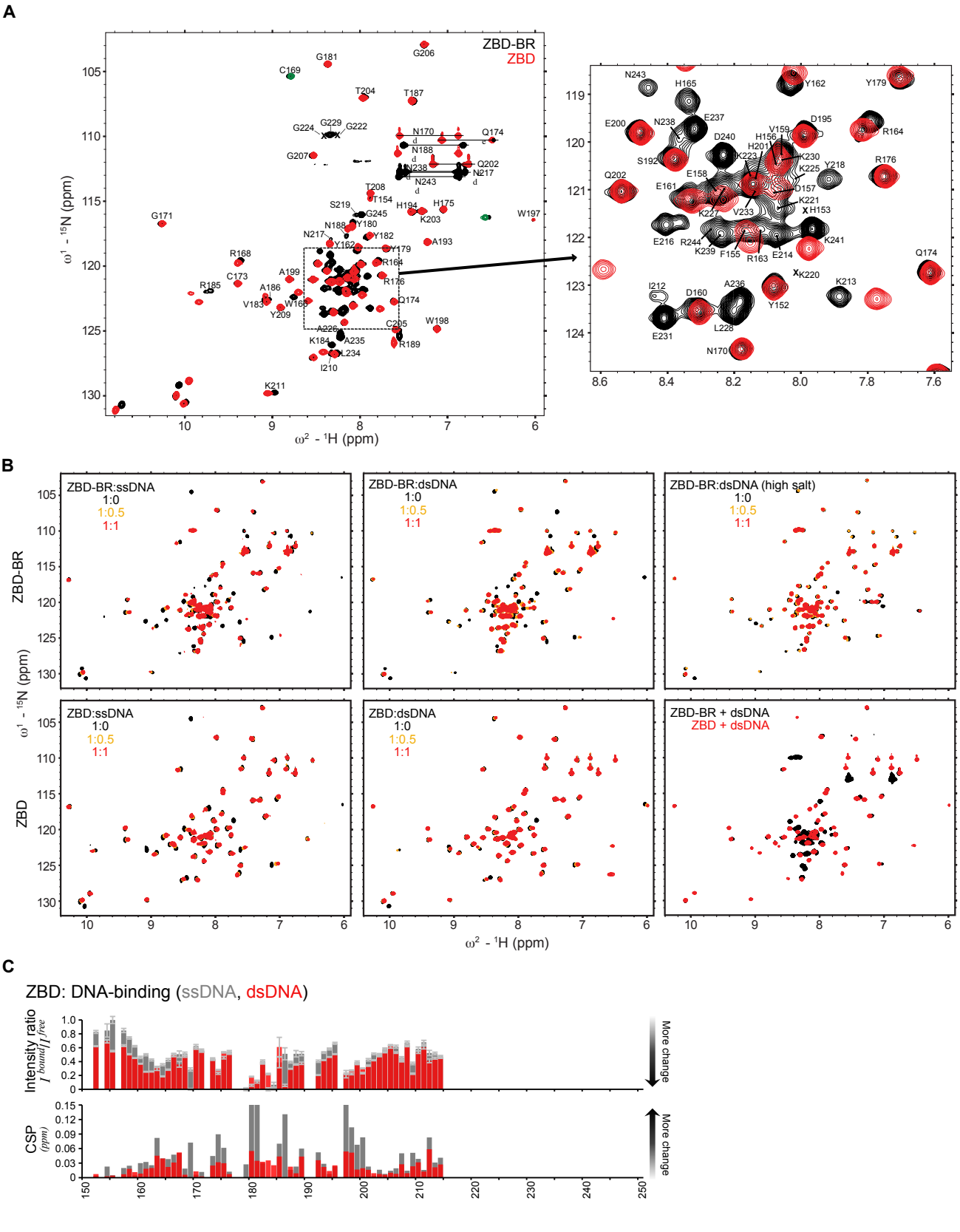


Figure S4. NMR analysis reveals a bipartite recognition of DNA structures by SPRTN, Related to Figure 4.

(A) Chemical shift assignments of amide signals in a ^{1}H, ^{15}N correlation spectrum (HSQC) of ZBD-BR (black). The spectrum of ZBD only (red) is superimposed. Both samples were measured at 100 μM concentration in 20 mM sodium chloride, 50 mM sodium phosphate pH 6.5, 2 mM TCEP at 298 K on a 600 MHz spectrometer.

(B) Superimposed HSQC of ZBD-BR and ZBD, and ZBD-BR at high salt at three different molar ratios of protein to ssDNA (left) or dsDNA (middle) (1:0, black; 1:0.5, orange; 1:1, red). ZBD-BR titrations with dsDNA at high salt concentration is shown on right, top. Superimposed spectra of ZBD-BR (black) and ZBD (red) with equimolar dsDNA are shown on right, bottom. All the samples (protein concentration 100 μM) were measured in 100 mM (500 mM for high salt) potassium chloride, 50 mM HEPES pH 7.5, 2 mM TCEP at 298 K at 600 MHz (ssDNA, dsDNA at high salt) or 900 MHz ^{1}H Larmor frequency (dsDNA).

(C) Chemical shift perturbations (CSPs) and intensity differences (line-broadening) of backbone amides in ZBD upon addition of an equimolar ratio of ssDNA (grey) and dsDNA (red). Errors for intensity ratios upon DNA-binding were estimated from error propagation of peak height uncertainties based on average noise levels (six randomly chosen positions in each NMR spectra). No boxes are shown for proline, unassigned, ambiguous (overlapped) residues in both plots.

Figure S5 (related to Figure 5)

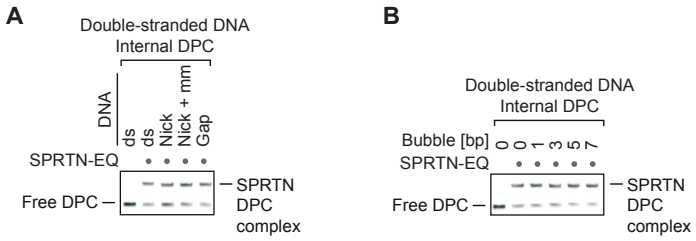


Figure S5. Cleavage by SPRTN requires disruptions within duplex DNA in close proximity to the DNA-protein crosslink, Related to Figure 5.

(A-B) SPRTN binds similarly to the model DPCs used in Figure 5A and 5D. Binding of catalytically inactive SPRTN E112Q (EQ, 25 nM) to the indicated model DPCs (25 nM) was assessed using EMSA assays.

2.2 A ubiquitin switch controls autocatalytic inactivation of the DNA–protein crosslink repair protease SPRTN

Shubo Zhao, Anja Kieser, Hao-Yi Li, Hannah K. Reinking, Pedro Weickert, Simon Euteneuer, Denit isa Yaneva, Aleida C. Acampora, Maximilian J. Götz, Regina Feederle, Julian Stingele, A ubiquitin switch controls autocatalytic inactivation of the DNA-protein crosslink repair protease SPRTN, *Nucleic Acids Research*, Volume 49, Issue 2, 25 January 2021, Pages 902-915, https://doi.org/10.1093/nar/gkaa1224

Summary

This publication investigates the role of SPRTN´s monoubiquitylation. A fraction of SPRTN is constantly monoubiquitylated. Upon treatment of cells with formaldehyde - a highly efficient crosslinking agent - SPRTN becomes deubiquitylated. So far, the purpose of this deubiquitylation step was not entirely understood and the enzyme performing this reaction unknown. Previous studies proposed that SPRTN´s ubiquitylation state effects its recruitment to chromatin. Monoubiquitylated SPRTN was thought to be excluded from DNA and only upon DPC induction and deubiquitylation, the enzyme gains access to the DNA lesion. However, by employing cell-based assays in combination with biochemical *in vitro* analyses, we showed that monoubiquitylation serves another function and negatively regulates SPRTN activity. First, it promotes polyubiquitylation and thereby making SPRTN a target for proteasomal degradation. Second, it enhances autocatalytic cleavage activity *in trans,* which further leads to a decrease in the cellular SPRTN pool. Additionally, the enzyme, which is responsible for removing ubiquitin upon DPC-induction was identified. USP7 was shown to interact with ubiquitylated SPRTN and to remove the modification, thereby prolonging its half-life and enabling it to remove toxic DPCs.

Author contribution

I performed the *in vitro* characterization of SPRTN´s ubiquitin-dependent activity. Therefore, I expressed and purified different SPRTN variants and tested them in

autocatalytic cleavage as well as substrate cleavage assays (entire Figure 6). I was involved in the discussion of the data with Shubo Zhao and Julian Stingele.

902–915 *Nucleic Acids Research, 2021, Vol. 49, No. 2* *Published online 22 December 2020*
doi: 10.1093/nar/gkaa1224

A ubiquitin switch controls autocatalytic inactivation of the DNA–protein crosslink repair protease SPRTN

Shubo Zhao[1,2], Anja Kieser[1,2], Hao-Yi Li[1,2], Hannah K. Reinking[1,2], Pedro Weickert[1,2], Simon Euteneuer[1,2], Denitsa Yaneva[1,2], Aleida C. Acampora[1,2], Maximilian J. Götz[1,2], Regina Feederle[3] and Julian Stingele[1,2,*]

[1]Department of Biochemistry, Ludwig-Maximilians-University, 81377 Munich, Germany, [2]Gene Center, Ludwig-Maximilians-University, 81377 Munich, Germany and [3]Institute for Diabetes and Obesity, Monoclonal Antibody Core Facility, Helmholtz Zentrum München, 85764 Neuherberg, Germany

Received April 29, 2020; Revised December 02, 2020; Editorial Decision December 03, 2020; Accepted December 04, 2020

ABSTRACT

Repair of covalent DNA–protein crosslinks (DPCs) by the metalloprotease SPRTN prevents genome instability, premature aging and carcinogenesis. SPRTN is specifically activated by DNA structures containing single- and double-stranded features, but degrades the protein components of DPCs promiscuously and independent of amino acid sequence. This lack of specificity is useful to target diverse protein adducts, however, it requires tight control in return, in order to prohibit uncontrolled proteolysis of chromatin proteins. Here, we discover the components and principles of a ubiquitin switch, which negatively regulates SPRTN. We demonstrate that monoubiquitylation is induced in an E3 ligase-independent manner and, in contrast to previous assumptions, does not control chromatin access of the enzyme. Data obtained in cells and *in vitro* reveal that monoubiquitylation induces inactivation of the enzyme by triggering autocatalytic cleavage *in trans* while also priming SPRTN for proteasomal degradation *in cis*. Finally, we show that the deubiquitylating enzyme USP7 antagonizes this negative control of SPRTN in the presence of DPCs.

INTRODUCTION

Covalent DNA–protein crosslinks (DPCs) are particularly dangerous DNA lesions, which interfere with all basic chromatin transactions including transcription and replication (1–3). Endogenous DPCs are not only caused by toxic metabolites such as reactive aldehydes but also by entrapment of covalent reaction intermediates of enzymes such as topoisomerases (4). Moreover, crosslinking can be induced by exogenous agents such as ionizing radiation as well as by various widely-used chemotherapeutics (5,6). The protein component of DPCs is targeted for repair by proteases of the Wss1/SPRTN family (7–14). The human protease SPRTN is essential for viability in mammalian cells, which highlights the scale of the threat posed by endogenous DPCs (15). Moreover, germline mutations in *SPRTN*, which result in the deletion of the enzyme's C-terminal tail, are causative for Ruijs-Aalfs syndrome (RJALS) (16,17). RJALS is characterized by premature aging and early-onset hepatocellular carcinomas with similar phenotypes being observed in hypomorphic *Sprtn* mutant mice (16–18). SPRTN is a DNA-dependent metalloprotease, which is activated by DNA structures containing single- (ss) and double-stranded (ds) features, such as ss-/dsDNA junctions or frayed dsDNA ends (19). However, SPRTN's proteolytic activity is highly promiscuous (8,9,11). The lack of preference for certain amino acid sequences is required to target diverse DPCs, but it demands tight control in return. A substantial fraction of SPRTN (30-50%) is constitutively monoubiquitylated (20–22). The modification is strongly reduced in SPRTN variants with amino acid replacements in the enzyme's C-terminal ubiquitin-binding zinc finger (UBZ) (20–22). Attempts to identify the site of monoubiquitylation by mass spectrometry revealed four potentially-modified lysine (K) residues, but SPRTN variants with collective lysine-to-arginine (KR) substitutions retained the modification (8). It has been proposed that monoubiquitylation regulates chromatin access of the enzyme because the recruitment of SPRTN to chromatin upon DPC induction is accompanied by rapid deubiquitylation (8). However, testing this model directly, has not been possible because the involved deubiquitylating enzyme (DUB) is unknown. Accordingly, the mechanistic principles of SPRTN's regulation by monoubiquitylation remain unclear.

Here, we identify the DUB USP7 as the factor responsible for deubiquitylating SPRTN when cells are challenged by DPCs. Moreover, we reveal that monoubiquitylation induces direct inactivation of SPRTN rather than regulating chromatin recruitment of the enzyme. The modifica-

*To whom correspondence should be addressed. Tel: +49 89 2180 71080; Email: stingele@genzentrum.lmu.de

tion triggers autocleavage of the protease while also enhancing proteasomal degradation by priming polyubiquitylation. Finally, *in vitro* experiments suggest that the constitutive monoubiquitylation occurs in a highly promiscuous E3 ligase-independent manner. Taken together, we unravel the principles and components of a regulatory switch, which allows safe operation of a potentially dangerous enzymatic activity in human cells.

MATERIALS AND METHODS

Antibodies and inhibitors

Anti-Strep (ab76949) and anti-Histone H3 (ab10799) antibodies were purchased from Abcam; anti-Tubulin (T6074), anti-Flag (F3165) and anti-Rat IgG (A9037) antibodies were purchased from Sigma; anti-GAPDH (2118) antibody was purchased from Cell Signaling; anti-USP7 (sc-137008) and anti-Histone H1 (sc-377468) were purchased from Santa Cruz; Goat Anti-Mouse Immunoglobulins/HRP (P0447), Swine Anti-Rabbit Immunoglobulins/HRP (P0399) antibodies were purchased from Dako and anti-GFP (PABG1, used for detection of YFP) was purchased from Chromotek. Rat monoclonal anti-human SPRTN antibody (6F2) was generated by immunization of Wistar rats with purified GST-tagged SPRTN-ΔC, which was expressed in insect cells as described previously (8). Hybridoma supernatants were screened by ELISA for binding to purified untagged SPRTN protein and further validated by western blot analysis on HeLa cell lysates as well as recombinant protein. Clone SPRT 6F2 (IgG2a) was subcloned twice by limited dilution to obtain a stable monoclonal hybridoma cell line. For inhibition of the ubiquitin-activating enzyme E1 MLN7243 (TAK-243) was purchased from Chemietek and used at 1 μM final concentration (23). For inhibition of proteasomal degradation MG132 was purchased from Sigma (M7449) and used at 5 μM final concentration. For inhibition of protein synthesis cycloheximide was purchased from Sigma (C4859) and used at 100 μg/ml final concentration.

Cell lines

HCT116 wild-type (WT), HCT116 *USP7* KO and HAP1 WT, *VCPIP1* KO, *USP11* KO cells were purchased from Horizon Discovery. HeLa T-REx Flp-In, 293 T-REx Flp-In and DLD1 cells were provided by Cell Services, The Francis Crick Institute. HeLa T-REx Flp-In cells stably expressing YFP-SPRTN-Strep-tag variants were generated using the Flp-In system (Thermo Fisher) according to manufacturer's instructions.

Transient transfection

For transient transfections cells were grown to 70–90% confluency in 12-well plates. Plasmids (1 μg plasmid) and Lipofectamine 2000 reagent (Invitrogen, 1 μl/μg plasmid) were diluted in 50 μl Opti-MEM Medium each. Plasmid and Lipofectamine 2000 dilutions were mixed following a 5 min incubation. After an additional 15 min incubation, the transfection mix was added to the cells.

siRNA transfection

Cells were grown to 20–30% confluency in 6-well plates. 3 μl siRNA (20 μM, ON TARGETplus SMARTpool, Horizon, USP7 (L-006097-00-0005), USP11 (L-006063-00-0005), VCPIP1 (L-019137-00-0005), Control (D-001810-10-05)) and 3 μl Lipofectamine RNAiMAX Transfection Reagent (Invitrogen) were each diluted in 100 μl Opti-MEM Medium. siRNA and Lipofectamine RNAiMAX Transfection Reagent dilutions were mixed following a 5 min incubation. After an additional 15 min incubation, the transfection mix was added to the cells. After 48 h, cells were reseeded into 60 mm dishes, followed by chromatin fractionation the following day.

Generation of USP7 knock-out cells

USP7 gRNA1 (GGTCTGTCTGGATAAAAGCG) and gRNA2 (GAGTGATGGACACAACACCG) were inserted into Lenti-multi-CRISPR plasmid (Addgene #85402; RRID: Addgene_85402) as described previously (24). The resulting plasmid was then transiently transfected into HAP1 or DLD1 cells using Lipofectamine 2000 (Invitrogen) to generate USP7 knock-out cells. One day after transfection, cells were selected by Puromycin for 48 h. Selected cells were then reseeded in 96-well plates (0.5 cell/well) to generate single clones. Single clones were then screened using western blotting with anti-USP7 antibody.

Purification of partially ubiquitylated YFP-SPRTN-EQ-Strep

293 T-REx Flp-In cells expressing YFP-SPRTN-EQ-Strep were grown in two 245 × 245 mm dishes to 50% confluency before overnight induction of protein expression by addition of 1 μg/ml doxycycline. Cells were harvested by scraping, washed twice in PBS and stored at −80°C. For purification, cells were thawed and resuspended in 10 ml lysis buffer (50 mM HEPES/NaOH pH 7.2, 1 M NaCl, 10% glycerol, 1% IGEPAL CA-630, cOmplete EDTA free protease inhibitors, 0.04 mg/ml PefaBloc SC and 20 mM iodoacetamide). Following sonication and Benzonase (4 U/ml) digestion for 30 min at 4°C, lysates were cleared by centrifugation (23 500 g, 45 min, 4°C). 60 μl MagStrep type3 XT beads (5% (v/v) suspension) were incubated with the supernatant for 1 h at 4°C prior to three wash steps with wash buffer (50 mM HEPES/NaOH pH 7.2, 0.5 M NaCl, 10% glycerol). Finally, purified YFP-SPRTN-EQ-Strep was eluted in 3 × 80 μl elution buffer (50 mM HEPES/NaOH pH 7.2, 0.5 M NaCl, 10% glycerol, 50 mM biotin).

DUB screen

71 cDNAs encoding human DUBs (hORFeome v8.1 Deubiquitinating Enzymes collection + seven additional ORFeome clones: CloneIds: 100011387, 100010734, 100002718, 100066416, 100070362, 100068239) were sub-cloned into pDEST17 using Gateway LR Clonase II (Invitrogen). Plasmids were then transformed into *Escherichia coli* BL21 (DE3) cells for protein expression in 50 ml cultures. Cell pellets were resuspended in BugBuster reagent (Merck Millipore, 5 ml/g). Benzonase (25 U/ml)

and DTT (5 mM) were added prior to an incubation of 20 min at room temperature. Lysates were then cleared by centrifugation (16 000 g, 20 min, 4°C). Lysates were mixed in pools of three prior to assessing their ability of deubiquitylating YFP-SPRTN-EQ-Strep. To this end, purified partially ubiquitylated YFP-SPRTN-EQ-Strep was incubated with lysate pools for 30 min at 25°C. Lysates of non-transformed BL21 cells served as negative control, the unspecific deubiquitylating activity of the catalytic domain of USP2 ($USP2^{cd}$, BostonBiochem, E-504) as positive control. Deubiquitylation was then assessed using SDS-PAGE followed by western blotting using anti-Strep antibody.

DUB activity assays

Candidate DUBs were partially purified using a standard Ni-NTA-purification strategy and then tested for their activity either using a ubiquitin–rhodamine cleavage assay kit (BostonBiochem) following the manufacture's instructions or by incubation with partially ubiquitylated YFP-SPRTN-EQ-Strep for 30 min at 25°C followed by SDS-PAGE and western blotting using anti-Strep antibody.

Expression and purification of recombinant proteins

SPRTN codon-optimized for bacterial expression (using GeneOptimizer) was expressed from a pNIC-Strep-ZB-SPRTN plasmid as previously described (19). SPRTN-Ub^{LF} was generated using Gibson assembly cloning. Flag-SPRTN was generated using insertional mutagenesis. For protein expression plasmids were transformed into BL21(DE3) *E. coli* cells and grown at 37°C in Terrific broth (TB) medium until they reached OD 0.7. Protein expression was induced by addition of 0.5 mM IPTG over night at 18°C. Next, cells were harvested, resuspended in buffer A (50 mM HEPES/KOH pH 7.2, 500 mM KCl, 1 mM $MgCl_2$, 10% glycerol, 0.1% IGEPAL, 0.04 mg/ml Pefabloc SC, cOmplete EDTA-free protease inhibitors, 1 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP), pH 7.2) and lysed by sonication. All steps were carried out at 4°C. Cell lysate was incubated with benzonase (45 U/ml lysate) for 30 min on ice prior to the removal of cell debris by centrifugation at 18 000 g for 30 min. Cleared supernatant was applied to Strep-Tactin®XT Superflow® high capacity cartridges, 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, pH 7.2). Proteins were eluted in 6 CV buffer C (50 mM HEPES/KOH pH 7.2, 500 mM KCl, 10% glycerol, 1 mM TCEP and 50 mM biotin, pH 7.2). Eluted proteins were further applied to HiTrap Heparin HP affinity columns 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, pH 7.2). Eluted fractions containing recombinant SPRTN protein were desalted against buffer B using PD-10 desalting columns. The affinity tag was cleaved off over night at 4°C by the addition of His-tagged TEV protease with 1:10 mass ratio. Cleaved recombinant SPRTN protein was further purified by size exclusion chromatography using a HiLoad 16/600 Superdex 200 pg column equilibrated in buffer E (50 mM HEPES/KOH pH 7.2, 500 mM KCl, 10% Glycerol, 0.5 mM TCEP, pH 7.2). Eluted proteins were concentrated with 10 kDa cutoff Amicon Ultra centrifugal filters before snap-freezing in liquid nitrogen and storing at −80°C.

UBE2D3 was sub-cloned into pDEST17 using Gateway LR Clonase II (Invitrogen). For protein expression plasmids were transformed into BL21(DE3) *E. coli* cells and grown at 37°C in Terrific broth (TB) medium until they reached OD 0.7. Protein expression was induced by addition of 0.5 mM IPTG for 3 h at 37°C. Next, cells were harvested, resuspended in buffer A (50 mM NaH_2PO_4 pH 8, 150 mM NaCl, 10 mM Imidazol and 0.5 mM TCEP), with the addition of 0.04 mg/ml Pefabloc SC and cOmplete EDTA-free protease inhibitors. The cells were then lysed by sonication. All steps were carried out at 4°C. Cell lysate was incubated with benzonase (45 U/ml lysate) for 30 min at 4°C prior to the removal of cell debris by centrifugation at 16 000 g for 30 min. Cleared supernatant was applied twice to Ni-NTA beads equilibrated in buffer A, washed with 5 CV of buffer A and 7 CV of buffer B (50 mM NaH_2PO_4 pH 8, 500 mM NaCl, 20 mM Imidazol, 0.5 mM TCEP). Proteins were eluted in 5 CV buffer C (50 mM NaH_2PO_4 pH 8, 500 mM NaCl, 250 mM Imidazol, 0.5 mM TCEP). Eluted fractions containing recombinant His-UBE2D3 protein were desalted against buffer D (20 mM Tris pH 7.5, 150 mM NaCl, 10% glycerol, 2 mM TCEP) using PD-10 desalting columns. The recombinant protein was further purified by size exclusion chromatography using a HiLoad 16/600 Superdex 200 pg column equilibrated in buffer D. Eluted proteins were concentrated with 3 kDa cutoff Amicon Ultra centrifugal filters before snap-freezing in liquid nitrogen and storing at −80°C.

In vitro ubiquitylation assay

The E2 screen was conducted using the E2 screening kit (UBPBio) according to the manufacturer's instructions. In brief, human E2 ubiquitin conjugating enzymes (2 μM) were incubated together with catalytically inactive SPRTN-E112Q (EQ) (2 μM), E1 ubiquitin activating enzyme (100 nM), ubiquitin (R&D Systems, U-100H, 50 μM) or no lysines N-Terminal Biotin ubiquitin (R&D Systems, UB-NOK-050, 50 μM), DNA (11.1 nM ΦX174 virion) and ATP (2 mM) for 1.5 h at 30°C. All other *in vitro* ubiquitylation assays contained the indicated SPRTN variants, E1 ubiquitin activating enzyme (300 nM), ubiquitin (50 μM), ATP (2 mM) and purified His-tagged UBE2D3 (concentrations as indicated in figure legends) and were incubated for 1.5 h at 30°C. The catalytic domain of USP2 ($USP2^{cd}$) was purchased from BostonBiochem and was included as indicated. Ubiquitylation reactions were stopped by addition of 4× LDS sample buffer supplemented with β-mercaptoethanol and boiling at 95°C for 10 min and then subjected to SDS-PAGE followed by staining with InstantBlue Coomassie protein stain. Contrast of scanned images was adjusted globally using Adobe Photoshop software

In vitro protease assays

Reactions were performed in 20 μl containing the indicated SPRTN variants (500 nM), histone H1 (500 nM, NEB) as

indicated and either ΦX174 Virion ssDNA or RFI dsDNA (11.1 nM, NEB). The reaction buffer comprised 19.5 mM HEPES/KOH pH 7.2, 2.9% glycerol, 5 mM TCEP and either 80 mM or 150 mM KCl. Reactions were incubated at 25°C for 1 h and stopped by the addition of 4x LDS sample buffer supplemented with β-mercaptoethanol. Samples were boiled at 95°C for 10 min, resolved on 4–12% Bis–Tris NuPAGE gradient gels and stained using InstantBlue Coomassie protein stain or analysed by western blotting using anti-SPRTN, anti-Flag and anti-H1 antibodies. The intensity of western blots and scanned gels was adjusted globally using Adobe Photoshop. Cleavage reactions were quantified by dividing the amount of cleaved protein by the total amount of protein (cleaved and uncleaved) as determined by analysis of western blot results using ImageJ.

Protein-oligonucleotide conjugate cleavage assay

Protein G was conjugated to the 5′-terminal, 3′-terminal or an internal base of a 30mer oligonucleotide (5′-Cy5-AC CAGTGCCTTGCTAGGACATCTTTGCCCA-3′) as described previously (19). Double-stranded DPCs were generated by annealing a complementary reverse oligonucleotide. Annealing was carried out immediately prior to cleavage reactions by mixing conjugates and reverse oligonucleotides at a ratio of 1:1.2 in annealing buffer (25 mM HEPES/KOH pH 7.2, 50 mM KCl, 5% Glycerol, and 0.2 mg/ml BSA) followed by an incubation for 2 min at 37°C and a subsequent decrease in temperature of 1°C/min until 25°C were reached. Cleavage reactions with model DPCs were performed in a reaction volume of 10 μl containing 6.25 nM SPRTN and 25 nM DPC in a final reaction buffer of 17.5 mM HEPES/KOH pH 7.2, 80 mM KCl, 3.5% glycerol, 5 mM TCEP and 0.1 mg/ml BSA. Reactions were incubated for 2 h at 25°C. 2 μl of 6× Orange G loading dye was added before reactions were resolved on 20% TBE gels using 1× TBE as running buffer at 4°C. Gels were photographed using a BioRad Chemidoc MP system using filter settings for Cy5 fluorescence. The intensity of scanned gels was adjusted globally using ImageJ, which was also used to quantify cleavage by dividing the amount of cleaved conjugate by the total amount of conjugate (cleaved and uncleaved) and subtraction of background signal (determined from lanes without SPRTN).

Co-immuno-precipitation

To test binding between USP7/VCPIP1/USP11 and SPRTN, HeLa-T-REx Flp-In cells stably expressing YFP-SPRTN-Strep variants were seeded in 60 mm tissue culture plates, grown to 50% confluency and then transiently transfected with pcDNA5-FRT/TO plasmids encoding Flag/Flag-USP7/VCPIP1/USP11 variants using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Sixteen hours after transfection and concurrent induction of protein expression by doxycycline (1 μg/ml), cells were washed once with ice-cold PBS and harvested by scraping in lysis buffer (20 mM Tris/HCl pH 7.5, 137 mM NaCl, 1% IGEPAL CA-630, 2 mM EDTA, 2 mM $MgCl_2$, 4 U/ml Benzonase, cOmplete EDTA free protease inhibitors, 0.04 mg/ml PefaBloc SC and 20 mM iodoacetamide). Lysates were incubated for 30 min on ice, before centrifugation at 16 000 g for 30 min. Supernatants were then used for immuno-precipitation using 5 μl magnetic anti-Flag M2 beads (Sigma) at 4°C for 1 h. The beads were then washed three times with wash buffer (10 mM Tris/HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA and 0.5% IGEPAL CA-630). Finally, samples were resuspended in 100 μl 1× LDS sample buffer, analysed by SDS-PAGE and western blotting with anti-Flag, anti-GAPDH and anti-Strep antibodies.

Cellular autocleavage assay

pcDNA5-FRT/TO plasmids encoding YFP-SPRTN-Strep or YFP-SPRTN-Ub^{LF} variants (1 μg) and Flag/Flag-$USP7^{WT/C223S}$ (3 μg) were transiently transfected using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Protein expression was induced by overnight (16 h) incubation with doxycycline (final concentration 1 μg/ml). Cells were lysed on ice in 1 ml lysis buffer (50 mM HEPES pH 7.5, 1 M NaCl, 10% glycerol, 1% IGEPAL CA-630, 2 mM $MgCl_2$, 20 mM iodoacetamide, 0.04 mg/ml PefaBloc SC and cOmplete EDTA-free protease inhibitors). After addition of benzonase (4 U/ml), lysates were incubated for 30 min on ice. Lysates were cleared by centrifugation (16 000 g, 30 min) at 4°C and applied to 15 μl GFP-trap Magnetic Agarose (Chromotek) and processed according to manufacturer's instructions. Finally, samples were resuspended in 65 μl 1× LDS sample buffer, subjected to analysis by SDS-PAGE and western blotting with anti-GFP antibody (PABG1, Chromotek).

Chromatin fractionation

Cells in the mid-exponential phase of growth were collected by scraping in ice-cold 1X PBS. Cells were then equally split and either directly resuspended in 1X LDS buffer or incubated for 10 min in ice-cold CSK buffer (10 mM PIPES, 100 mM NaCl, 1.5 mM $MgCl_2$, 5 mM EDTA, 300 mM sucrose and 0.5% Triton X-100, 0.04 mg/ml PefaBloc SC and cOmplete EDTA free protease inhibitors). Chromatin-bound proteins were then isolated by low speed centrifugation (3000 rpm, 5 min at 4°C).

Cycloheximide chase

Cells were seeded in 12-well tissue culture plates, grown to 80% confluency and then treated with 5 μM MG132. After 2 h cells were treated with 100 μg/ml cycloheximide (Sigma) for the indicated amount of time. Finally, cells were lysed in 150 μl 1X LDS sample buffer, followed by SDS-PAGE and western blotting with the indicated antibodies.

Formaldehyde sensitivity assay

Long term treatment: 10^4 cells were seeded per well in 12-well plates and treated with the indicated formaldehyde concentration the next day. After 72 h, medium was replaced with alamarBlue cell viability reagent (36 μg/ml resazurin in PBS) and plates kept for an additional 1 h incubation at 37°C. Cell viability was then assessed by measuring fluorescence (560 nm excitation/590 nm emission). Short term

treatment: 5×10^4 cells were seeded per well in 12-well plates and treated with 1 mM formaldehyde concentration the next day for 2 h. After 48 h, medium was replaced with alamarBlue cell viability reagent (36 μg/ml resazurin in PBS) and plates kept for an additional 1 h incubation at 37°C. Cell viability was then assessed by measuring fluorescence (560 nm excitation/590 nm emission).

Detection of formaldehyde-induced DNA–protein crosslinks

DPCs were induced by treating HAP1 WT or *USP7* KO cells with 75 μM formaldehyde for 2 h. DPCs were measured using a KCl/SDS precipitation assay as described before (25). Briefly, cells were washed twice with PBS and lysed in 400 μl denaturing lysis buffer (2% SDS, 20 mM Tris/HCl, pH 7.5), frozen in liquid nitrogen and stored at −80°C until further processing. Lysates were thawed at 55°C for 5 min with 1200 rpm shaking, followed by pipetting samples up and down 30 times. Cellular protein was then precipitated by adding 400 μl precipitation buffer (200 mM KCl, 20 mM Tris pH 7.5) and incubation on ice for 5 min. The precipitated protein was separated by full speed centrifugation at 4°C for 5 min. Next, 400 μl supernatant was saved and used for soluble DNA measurement. The pellet was resuspended in 400 μl precipitation buffer and resolved by shaking at 55°C for 5 min followed by cooling down on ice for 5 min and full speed centrifugation at 4°C for 5 min. After repeating the wash procedure 3 times, protein precipitate was resuspended in 400 μl Proteinase K buffer (200 mM KCl, 20 mM Tris pH 7.5. Proteinase K 0.2 mg/ml) and incubated at 55°C for 45 min. Finally, 10 μl BSA (50 mg/ml) was added to the solution followed by cooling down on ice for 5 min followed full speed centrifugation at 4°C for 5 min. Next, supernatant containing crosslinked DNA was collected. Total DNA and crosslinked DNA were treated with 0.2 mg/ml RNase A for 30 min at 37°C. DNA concentrations were determined using Qubit™ dsDNA HS Assay Kit (Thermo Fisher) according to the manufacturer's instructions. The amount of DPCs was calculated as the ratio between crosslinked DNA and total DNA (crosslinked plus soluble DNA).

Complementation of *USP7* KO cells

DLD1 WT or *USP7* KO cells were seeded in six-well plates, grown to 50% confluency before transient transfection with pcDNA5-FRT/TO plasmids encoding YFP/YFP-USP7$^{WT/C223S}$ (2 μg) using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. 16 h after transfection, cells were reseeded into 12-well plates (5×10^4 cells per well) followed by treatment with 1 mM formaldehyde for 2 h the next day. After 48 h cell viability was determined using alamarBlue cell viability assay.

RESULTS

The monoubiquitylation of SPRTN's C-terminal tail is promiscuous

In order to understand the regulation of SPRTN by monoubiquitylation, we first attempted to map the modified lysine residue(s) using protein truncations. Ubiquitylation of full length SPRTN (N-terminally YFP-, C-terminally Strep-tagged) is readily observed upon transient transfection and is sensitive to inhibition of the ubiquitin-activating enzyme (E1) (Figure 1A). The truncated SPRTN variant found in RJALS patients (SPRTN-ΔC) is not ubiquitylated, while the isolated C-terminal tail (SPRTN-ΔN) is modified (Figure 1A). The modification can be further mapped to a small 7 kDa fragment, which contains the UBZ domain and five lysine residues (SPRTN-Δ425). Surprisingly, substitution of all five lysines (5KR) does not alter the level of modification (Figure 1A). The N-terminal YFP-tag and linker contain various lysines and we suspected that these residues might undergo modification as well. Indeed, deletion of the YFP-tag leads to a severe reduction in ubiquitylation of the SPRTN-Δ425-5KR fragment (Figure 1B). However, a slightly extended variant (SPRTN-Δ400) with the same lysine replacements (SPRTN-Δ400-5KR) remains ubiquitylated, unless all additional lysines are replaced as well (SPRTN-Δ400-9KR) (Figure 1C). Full-length SPRTN with the 9KR replacement is unstably expressed, but appears to remain monoubiquitylated (Figure 1D). Notably, a SPRTN-Δ400-8KR fragment with only one remaining lysine residue displays multiple modifications, which indicates that the monoubiquitylation can be further modified (Supplementary Figure S1A). We conclude that the monoubiquitylation of SPRTN can target various lysine residues (even those of the YFP-tag) and can be extended to a ubiquitin chain.

E3-independent monoubiquitylation of SPRTN

Monoubiquitylation of proteins bearing ubiquitin-binding domains is frequently observed and has been proposed to occur in an E3-independent manner (26–28). Thus, we tested a collection of human recombinant ubiquitin-conjugating enzymes (E2s) for their ability to ubiquitylate catalytically-inactive SPRTN *in vitro*. Strikingly, ten out of twenty-nine E2s induce SPRTN monoubiquitylation in the absence of an E3 ubiquitin ligase (Figure 1E and Supplementary Figure S1B). *In vitro* ubiquitylation by the E2 UBE2D3 even triggers multi-monoubiquitylation of SPRTN, as indicated by multiple modifications with a ubiquitin variant containing no lysines and a biotinylated N-terminus (Supplementary Figure S1C). Notably, a SPRTN variant with an altered UBZ domain (UBZ*, D473A) undergoes modification *in vitro*, but its modification is sensitive to the addition of an unspecific deubiquitylation activity (USP2 catalytic domain) (Figure 1F). This could indicate that SPRTN-UBZ* variants lack monoubiquitylation in cells because a functional UBZ domain is important to shield the modification from cellular DUB activities. The fact that monoubiquitylation of SPRTN occurs in an E3-independent manner (although an involvement of E3 ligases in cells cannot be excluded) and the high level of modification in basal conditions argues that the generation of ubiquitylated SPRTN (SPRTN-Ub) is a constitutive process. In turn, this infers that cellular control of the modification must occur through deubiquitylation.

Nucleic Acids Research, 2021, Vol. 49, No. 2 **907**

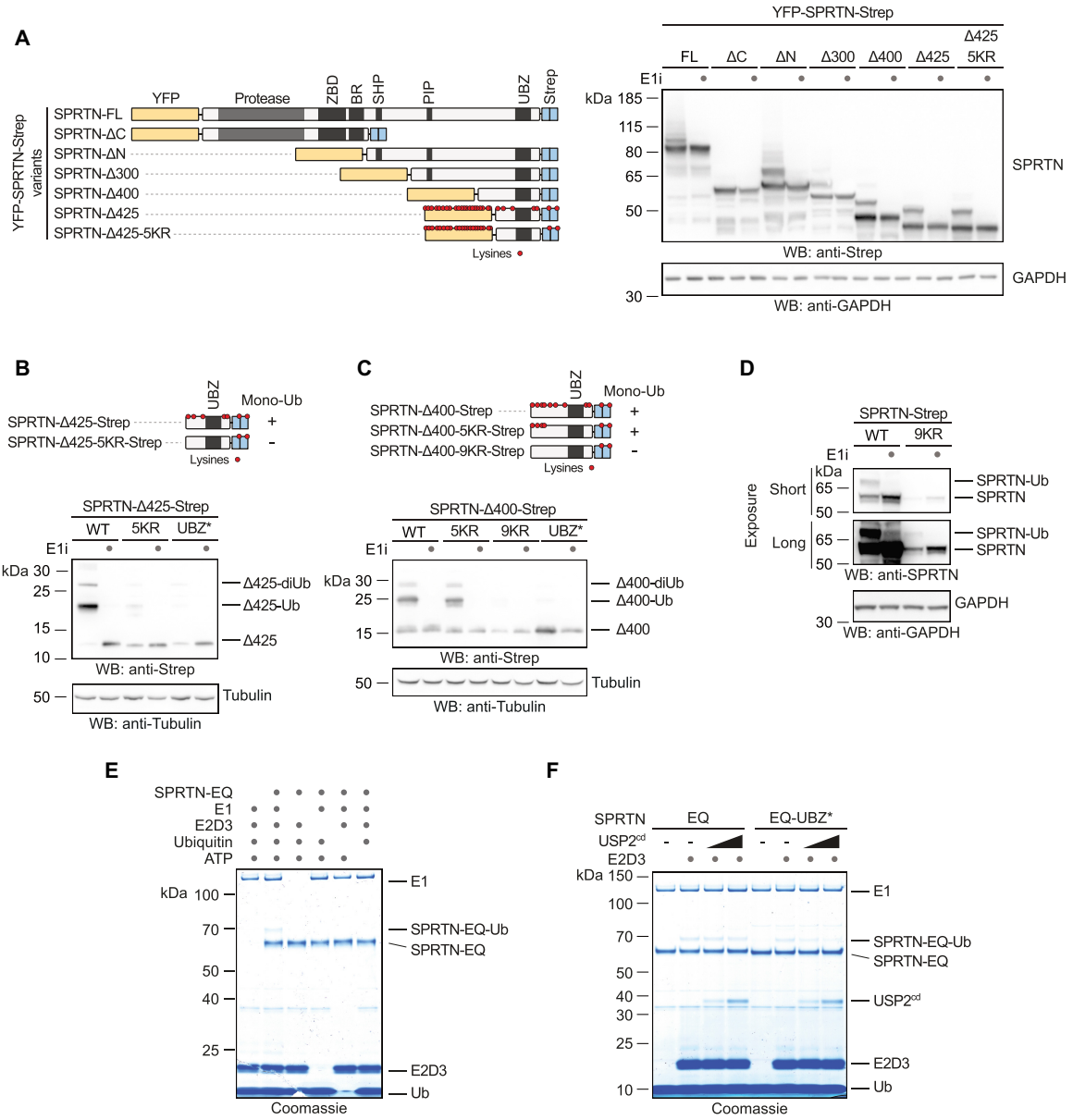


Figure 1. Promiscuous E3-independent monoubiquitylation of SPRTN's C-terminal tail. (**A–D**) Monoubiquitylation status of truncated SPRTN variants. Plasmids encoding tagged full-length (FL) SPRTN or truncations (carrying the indicated lysine to arginine (KR) substitutions or the UBZ* variant, D473A) were transiently transfected in HeLa T-REx Flp-In cells. Expression of SPRTN was induced by addition of doxycycline for 6 h prior to cell lysis (including a co-treatment with a ubiquitin-activating enzyme E1 inhibitor (E1i) as indicated) and analysed by SDS-PAGE and western blotting. (**E**) *In vitro* ubiquitylation assays containing SPRTN-EQ (410 nM), UBE2D3 (4 μM), E1 ubiquitin activating enzyme (300 nM), ubiquitin (50 μM) and ATP (2 mM) as indicated were incubated for 1.5 h at 30°C. Reactions were stopped by addition of LDS sample buffer and subjected to SDS-PAGE followed by staining with InstantBlue Coomassie protein stain. (**F**) *In vitro* ubiquitylation assays containing SPRTN-EQ or SPRTN-EQ-UBZ* (410 nM), E1 ubiquitin activating enzyme (300 nM), ubiquitin (50 μM), ATP (2 mM), UBE2D3 as indicated (8 μM) and increasing amounts of the catalytic domain of USP2 ($USP2^{cd}$) (0, 250 or 500 nM) were incubated for 1.5 h at 30°C. Reactions were stopped by addition of LDS sample buffer and subjected to SDS-PAGE followed by staining with InstantBlue Coomassie protein stain.

An *in vitro* screen reveals that USP7 targets ubiquitylated SPRTN

To identify the DUB responsible for deubiquitylating SPRTN, we designed an *in vitro* screen (Figure 2A). We subcloned an arrayed cDNA library containing sequences of seventy-one human DUBs into bacterial expression plasmids. Upon expression in *E. coli*, lysates were prepared and pooled in twenty-four sets of three. Deubiquitylation activity was assessed by incubating each pool with partially monoubiquitylated SPRTN-EQ purified from human cells. Addition of five out of twenty-four pools triggered SPRTN deubiquitylation (Supplementary Figure S2A). Each positive pool contained one lysate able to deubiquitylate SPRTN (Figure 2B). The respective plasmids were re-isolated and determined to encode four different DUBs: USP4, USP7, USP15 and USP42. All four candidates were re-expressed, partially purified and successfully re-tested for their ability to deubiquitylate SPRTN (Supplementary Figure S2B-C). To test whether these DUBs possess the ability to act on SPRTN-Ub in cells, we monitored SPRTN monoubiquitylation in cells upon overexpression of the respective candidates. Overexpression of USP7 but not of the other candidates leads to a loss of endogenous and exogenously-expressed SPRTN-Ub and a concurrent increase in unmodified SPRTN (Figure 3A-B). Of note, whether USP42 can deubiquitylate SPRTN in cells remains unclear, given that it was not expressed at significant levels. Importantly, overexpression of a catalytically inactive variant of the DUB ($USP7^{CS}$, C223S) does not trigger deubiquitylation (Figure 3C-D). Consistently, USP7 binds to SPRTN in co-immunoprecipitation experiments (Figure 3E). Notably, catalytically inactive USP7 ($USP7^{CS}$) binds preferentially to ubiquitylated SPRTN. Interestingly, monoubiquitylated species of SPRTN-UBZ* and SPRTN-ΔC co-immunoprecipitate with $USP7^{CS}$, although this modification is not detectable in input samples (Figure 3E). This observation is in agreement with our finding that SPRTN-UBZ* can be monoubiquitylated *in vitro*, but then fails to protect the modification (Figure 1F). USP7 bears an N-terminal TRAF domain, which precedes the catalytic domain (CD) and five C-terminal ubiquitin-like domains (UBLs). Deletion of the CD or the UBLs abrogates preferential binding of USP7 to SPRTN-Ub indicating that these domains are important to provide specificity for modified SPRTN, but are not essential for the interaction per se (Figure 3F). A USP7 variant lacking the TRAF domain (USP7-ΔTRAF) is deficient in SPRTN binding. However, interpretation of this result is complicated by the fact that this truncation is expressed at low levels, which may indicate more general defects (Figure 3F). We conclude that USP7 interacts specifically with SPRTN-Ub and has the ability to deubiquitylate the protease *in vitro* and in cells.

USP7 deubiquitylates SPRTN upon DPC induction

Next, we tested whether USP7 is the DUB responsible for regulating SPRTN's chromatin access by deubiquitylation upon DPC induction. Thus, we treated HCT116 WT and *USP7* knock-out (KO) cells for 2 h with formaldehyde before assessing recruitment of SPRTN by chromatin fractionation. Indeed, endogenous SPRTN fails to be deubiquitylated in the absence of USP7 upon formaldehyde exposure (Figure 4A). Unexpectedly however, the lack of deubiquitylation does not result in impaired recruitment. In *USP7* KO cells, also SPRTN-Ub is found on chromatin. These results indicate that deubiquitylation occurs downstream or in parallel to recruitment and is not preceding SPRTN's relocalization to chromatin. To understand the contribution of USP7-mediated deubiquitylation to DPC repair, we generated DLD1 and HAP1 *USP7* KO cells because sensitivity of HCT116 *USP7* KO cells is difficult to assess due to their strong growth defect. HAP1 and DLD1 *USP7* KO cells show defective SPRTN deubiquitylation and hypersensitivity towards formaldehyde exposure (Figure 4B–E). Importantly, the formaldehyde sensitivity of DLD1 *USP7* KO cells is complemented by transient transfection of $USP7^{WT}$ but not of $USP7^{CS}$ (Figure 4F and Supplementary Figure S3A). Furthermore, HAP1 *USP7* KO cells accumulate higher levels of DPCs following a 2-h exposure to formaldehyde as determined using a KCl-SDS precipitation assays (Figure 4G).

Of note, while this study was under consideration, it was proposed that SPRTN is deubiquitylated by VCPIP1 (29). In addition, a recent preprint argued that USP11 is responsible for SPRTN deubiquitylation (30). Given our identification of USP7, we compared the contribution of all three enzymes to SPRTN deubiquitylation. Neither VCPIP1 nor USP11 induce SPRTN deubiquitylation when overexpressed, while USP7 does (Supplementary Figure S3B). VCPIP1 and USP11 interact weakly with SPRTN in co-immunoprecipitating experiments, but show no (VCPIP1) or only weak (USP11) preference for SPRTN-Ub (Supplementary Figure S3B). siRNA-mediated depletion of VCPIP1 or USP11 has no effect on SPRTN deubiquitylation in DLD1 cells, while depletion of USP7 does (Supplementary Figure S3C). Moreover, we obtained HAP1 *USP11* and *VCPIP1* KO cells which show sensitivity towards formaldehyde but no defects in SPRTN deubiquitylation (Supplementary Figure S3D-E).

To conclude, under the conditions tested here, USP7 but not VCPIP1 or USP11 has a prominent role in deubiquitylating SPRTN in cells. Finally, the formaldehyde sensitivity and DPC accumulation observed in *USP7* KO cells argue that deubiquitylation, although not involved in SPRTN's chromatin recruitment, must have an important function in DPC repair. Therefore, we further explored the effects of monoubiquiytation on the SPRTN protease.

Monoubiquitylation promotes degradation and autocleavage of SPRTN

Monoubiquitylation can lead to proteasomal degradation by priming polyubiquitylation (31,32). Thus, we assessed whether monoubiquitylation destabilizes SPRTN using cycloheximide-chase experiments. Indeed, endogenous SPRTN-Ub has a shorter half-life than non-ubiquitylated SPRTN, with degradation being blocked by proteasome inhibition (Figure 5A and Supplementary Figure S4A). Degradation is not affected by loss of USP7, which indicates that deubiquitylation is not involved in SPRTN protein stability under basal conditions (Supplementary Figure S4A). Proteasomal inhibition leads to accumulation of polyubiq-

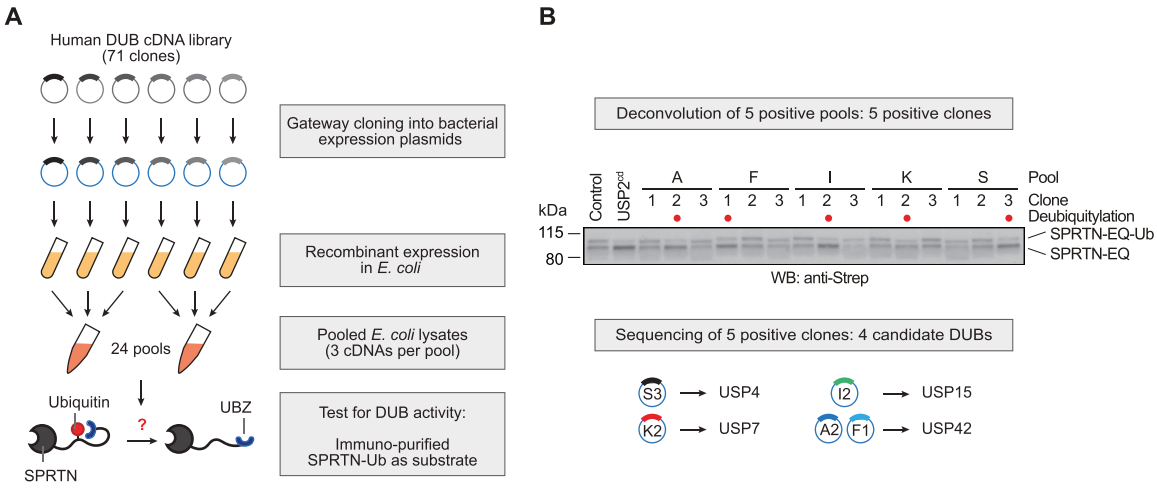


Figure 2. An *in vitro* screen identifies DUBs targeting ubiquitylated SPRTN. (**A**) Schematic depiction of the screening strategy employed to test seventy-one human deubiquitylating enzymes (DUBs) for their ability to deubiquitylate SPRTN. (**B**) Deconvolution of positive pools identified in Supplementary Figure S2A reveals four candidate DUBs. Lysates prepared from fifteen clones present in the five positive pools were incubated for 30 min at 25°C together with purified partially-monoubiquitylated YFP-SPRTN-EQ-Strep. Reactions were stopped by addition of LDS sample buffer and analysed by SDS-PAGE and western blotting using anti-Strep antibody. Lysates of BL21 cells served as negative control, the unspecific deubiquitylating activity of the catalytic domain of USP2 ($USP2^{cd}$) as positive control.

uitylated SPRTN species, which are strongly reduced in the SPRTN-UBZ* variant, which provides further support for monoubiquitylation inducing degradation by priming polyubiquitylation (Figure 5B and Supplementary Figure S4B). Furthermore, a linear fusion of SPRTN with ubiquitin (SPRTN-Ub^{LF}, omitting the two C-terminal glycines), which has been previously suggested to mimic monoubiquitylation, destabilizes the entire SPRTN pool (Figure 5C) (21).

While conducting these experiments, we noted that cells expressing SPRTN-Ub^{LF} display various protein fragments, which are recognized by an antibody specific for the N-terminal YFP-tag (Figure 5D). These fragments become even more obvious when enriched on GFP-trap resins and correspond to the previously reported autocatalytic fragments seen in cells, which express WT SPRTN but are absent in cells expressing catalytically inactive SPRTN-EQ (Figure 5D) (8,9). These results raise the possibility that monoubiquitylation of SPRTN triggers enhanced autocleavage of the enzyme. In agreement, autocleavage of the non-ubiquitylated SPRTN-UBZ* variant is barely detectable, unless linearly-fused to ubiquitin (SPRTN-UBZ*-Ub^{LF}) (Figure 5D). Furthermore, deubiquitylation of SPRTN induced by overexpression of $USP7^{WT}$, but not of catalytically inactive $USP7^{CS}$, leads to reduced formation of autocatalytic SPRTN fragments (Figure 5E). Autocleavage of endogenous SPRTN is induced by formaldehyde exposure and is more prominent at lower concentrations while deubiquitylation is observable at higher concentrations (Figure 5F) (8). Remarkably, if monoubiquitylation of endogenous SPRTN is blocked by pre-treating cells with ubiquitin E1 inhibitor, autocleavage is strongly reduced (Figure 5F). Interestingly, autocleavage of endogenous SPRTN also increases in cells, which have been treated with proteasome inhibitors (Figure 5G and Supplementary Figure S4C and D). This observation provides support for a model in which monoubiquitylated SPRTN is either degraded by the proteasome or undergoes autocleavage. In agreement, the short half-life of SPRTN-Ub is independent of the enzyme's catalytic activity and a SPRTN truncation ($SPRTN^{aa1-227}$) corresponding to the shortest autocleaved fragment is not particularly unstable (Supplementary Figure S4E-F). Taken together, proteasomal degradation and autocleavage appear to be independent outcomes induced by monoubiquitylation.

Next, we tested whether enhancement of SPRTN autocleavage by monoubiquitylation stems from a direct effect on the enzyme's activity. To this end, we produced recombinant SPRTN-Ub^{LF} and compared its autocleavage and substrate cleavage activity to WT SPRTN. Indeed, SPRTN-Ub^{LF} displays markedly increased DNA-dependent autocatalytic cleavage *in vitro* (Figure 6A). The effect is particularly strong in the presence of dsDNA and in high salt conditions. In contrast, cleavage of histone H1 or that of model DPC substrates (fluorescently-labelled protein G-oligonucleotide conjugates) is not significantly increased (Figure 6A and B). SPRTN autocleavage occurs *in trans* with one SPRTN molecule cleaving a second (8,9). A catalytically inactive Flag-tagged SPRTN-EQ variant is cleaved more efficiently by SPRTN-Ub^{LF} than by WT SPRTN. This suggests that modification of the SPRTN molecule cleaving *in trans* is sufficient to enhance autocleavage (Figure 6C). These *in vitro* data demonstrate that enhanced autocleavage of SPRTN-Ub in cells is caused by a direct effect of the modification on SPRTN activity. Thus, we conclude that monoubiquitylation negatively controls the SPRTN pool not only by inducing proteasomal degradation *in cis* but also by triggering the inactivation of other SPRTN molecules through *in trans* autocleavage.

910 *Nucleic Acids Research, 2021, Vol. 49, No. 2*

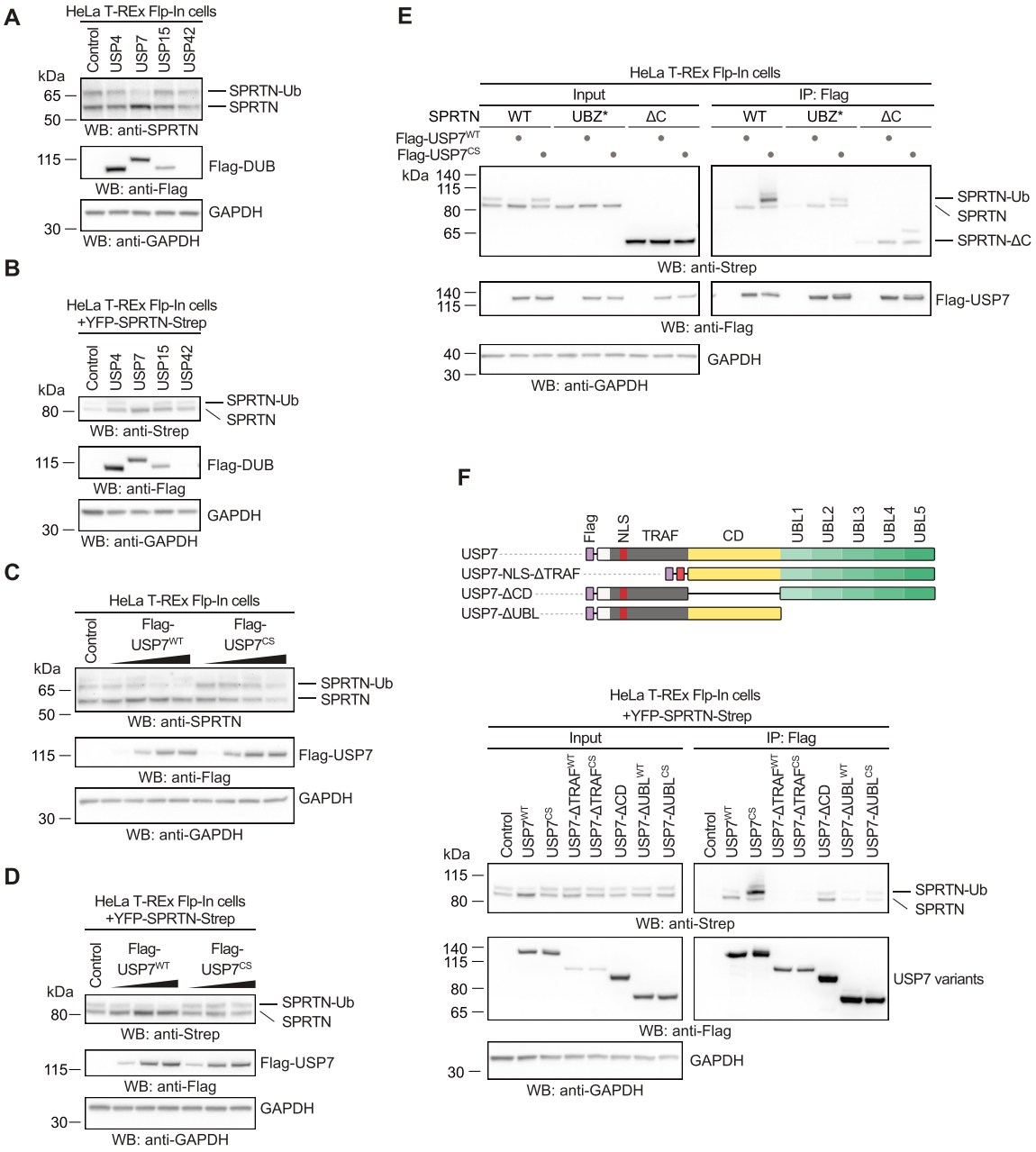


Figure 3. USP7 interacts with and targets ubiquitylated SPRTN in human cells. (**A**) Analysis of DUB overexpression-induced deubiquitylation of endogenous SPRTN in HeLa T-REx Flp-In cells. Indicated N-terminally Flag-tagged DUBs were transiently expressed for one day before cells were lysed and analysed by western blotting. (**B**) Analysis of DUB overexpression-induced deubiquitylation of doxycycline-inducible YFP-SPRTN-Strep stably expressed in HeLa T-REx Flp-In cells. Indicated N-terminally Flag-tagged DUBs were transiently expressed for one day before cells were lysed and analysed by western blotting. (**C**) Increasing amounts of N-terminally Flag-tagged USP7 (or the catalytically inactive CS variant) were transiently expressed in HeLa T-REx Flp-In cells for one day before cells were lysed and analysed by western blotting. (**D**) Increasing amounts of N-terminally Flag-tagged USP7 (or the catalytically inactive CS variant) were transiently expressed in HeLa T-REx Flp-In cells stably expressing doxycycline-inducible YFP-SPRTN-Strep for one day before cells were lysed and analysed by western blotting. (**E**) Plasmids encoding Flag-tagged full-length USP7 (WT or the catalytically inactive CS variant) were transiently transfected in HeLa T-REx Flp-In cells stably expressing the indicated doxycycline-inducible YFP-SPRTN-Strep variants. Binding was analysed by co-immunoprecipitation using anti-Flag beads followed by western blotting. (**F**) Schematic depiction of USP7's domain structure and protein truncations used for co-immunoprecipitation analysis with SPRTN (upper panel). Plasmids encoding Flag-tagged full-length USP7 (WT or the catalytically inactive CS variant) or the respective truncations were transiently transfected in HeLa T-REx Flp-In cells stably expressing doxycycline-inducible YFP-SPRTN-Strep. Binding was analysed by co-immunoprecipitation using anti-Flag beads followed by western blotting (lower panel).

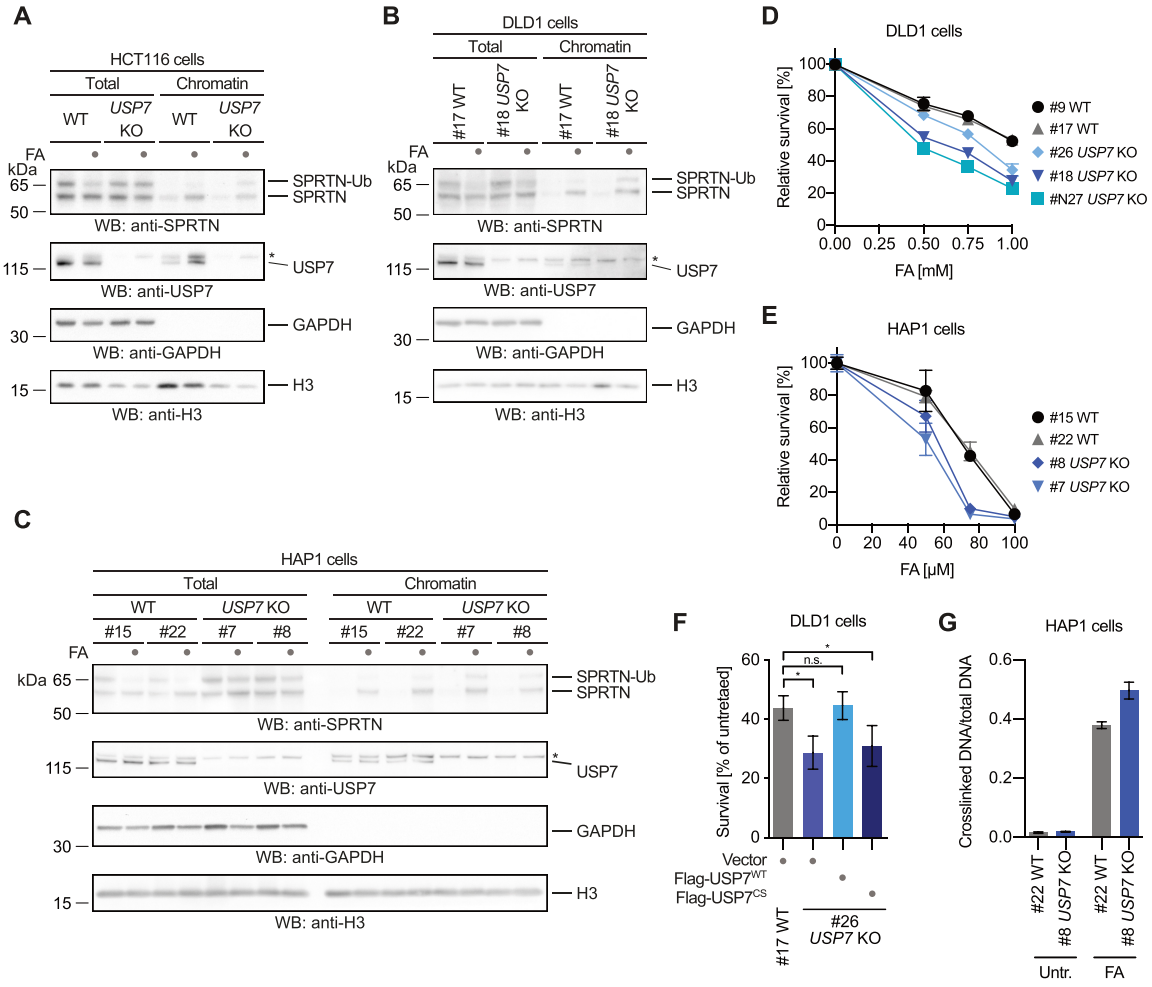


Figure 4. USP7 deubiquitylates SPRTN upon DPC induction. (**A**) HCT116 WT or *USP7* knock-out (KO) cells were treated with 2 mM formaldehyde (FA) for 2 h. Cells were either lysed directly in LDS sample buffer (total) or subjected to chromatin fractionation. Samples were then analyzed by SDS-PAGE followed by western blotting. Asterisks indicates a cross-reactive band. (**B**) Clonal DLD1 *USP7* KO cells and matched WT control cells were treated with 2 mM formaldehyde (FA) for 3 h. Cells were either lysed directly in LDS sample buffer (total) or subjected to chromatin fractionation. Samples were then analysed by SDS-PAGE followed by western blotting. Asterisks indicates a cross-reactive band. (**C**) Clonal HAP1 *USP7* KO cells and matched WT control cells were treated with 2 mM formaldehyde (FA) for 2 h. Cells were either lysed directly in LDS sample buffer (total) or subjected to chromatin fractionation. Samples were then analysed by SDS-PAGE followed by western blotting. Asterisks indicates a cross-reactive band. (**D**) Clonal DLD1 *USP7* KO cells and matched WT control cells were treated with the indicated formaldehyde concentrations for 2 h. After 48 h cell viability was determined using the alamarBlue cell viability assay. Values represent the mean ± SD of three technical replicates normalized to the mean of untreated controls of each cell line (**E**) Clonal HAP1 *USP7* KO cells and matched WT control cells were treated with the indicated formaldehyde concentrations for 72 h. Cell viability was then determined using the alamarBlue cell viability assay. Values represent the mean ± SD of three technical replicates normalized to the mean of untreated controls of each cell line. (**F**) YFP-tagged full-length USP7 (WT or the catalytically inactive CS variant) or the empty vector were transiently transfected in DLD1 *USP7* KO cells and matched WT control cells. Cells were treated with 1 mM formaldehyde for 2 h. After 48 h cell viability was determined using the alamarBlue cell viability assay. Values represent the mean ± SEM of four independent biological replicates normalized to the mean of untreated controls of each cell line. Significance was determined using a paired *t*-test (**P*-value < 0.05). (**G**) Cellular DPCs were quantified in clonal HAP1 *USP7* KO cells and matched WT control cells treated with 75 μM formaldehyde for 2 h using a KCl/SDS precipitation assay. DPCs were measured as the ratio of crosslinked DNA compared to total DNA. Values represent the mean ± SD of three technical replicates.

912 Nucleic Acids Research, 2021, Vol. 49, No. 2

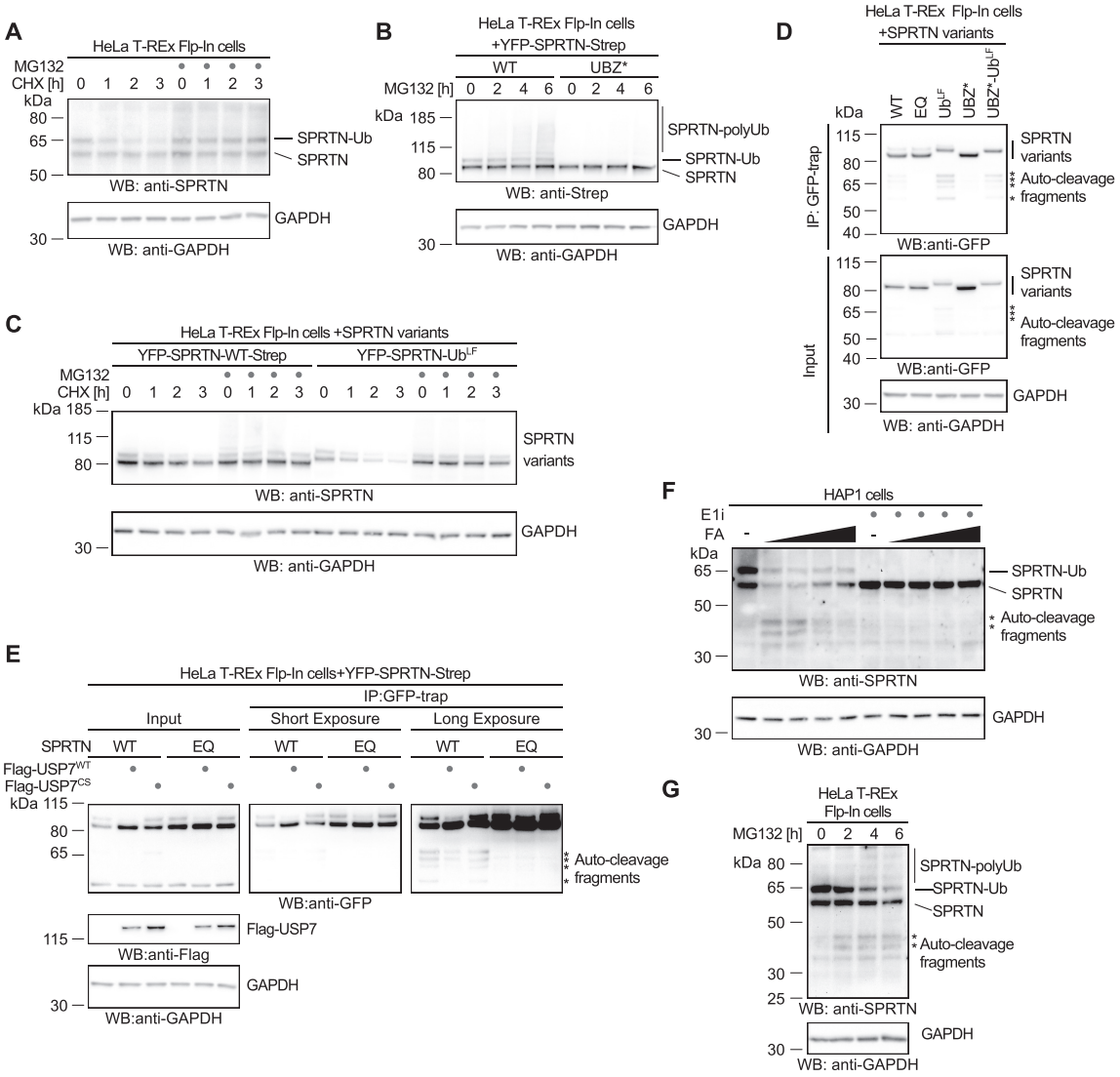


Figure 5. Monoubiquitylation promotes SPRTN degradation and autocleavage. (**A**) Stability of endogenous SPRTN was determined with a cycloheximide-chase experiment in HeLa-T-REx Flp-In cells. Cells were incubated with cycloheximide for the indicated amount of time (with or without a 2-h pre-treatment with the proteasome inhibitor MG132) prior to cell lysis and analysis by western blotting. (**B**) Polyubiquitylation of stably expressed doxycycline-inducible YFP-SPRTN-Strep or of YFP-SPRTN-UBZ*-Strep was determined in HeLa-T-REx Flp-In cells upon treatment with proteasome inhibitor MG132 for the indicated amount of time prior to cell lysis and analysis by western blotting. (**C**) Stability of stably expressed doxycycline-inducible YFP-SPRTN-Strep or a linear SPRTN-Ubiquitin fusion (YFP-SPRTN-Ub^{LF}) was determined in HeLa-T-REx Flp-In cells using a cycloheximide-chase experiment. Cells were incubated in the presence of cycloheximide for the indicated amount of time (with or without a 2-h pre-treatment with the proteasome inhibitor MG132) prior to cell lysis and analysis by western blotting. (**D**) Indicated YFP-SPRTN-Strep or linear SPRTN-Ubiquitin fusion (YFP-SPRTN-Ub^{LF}) variants were transiently transfected in HeLa-T-REx Flp-In cells. SPRTN autocleavage fragments were enriched on GFP-trap resins, followed by western blotting against the N-terminal YFP-tag. Western blotting of cell lysates against GAPDH serves as loading control. Asterisks indicate autocleavage fragments. (**E**) Indicated YFP-SPRTN-Strep variants were transiently transfected in HeLa-T-REx Flp-In cells in combination with Flag-tagged full-length USP7 (WT or the catalytically inactive CS variant) or the empty vector. SPRTN autocleavage fragments were enriched on GFP-trap resins, followed by western blotting against the N-terminal YFP-tag. Western blotting against GAPDH of cell lysates serves as loading control. Asterisks indicate autocleavage fragments. (**F**) HAP1 cells were treated with increasing amounts of formaldehyde (FA, 0.25, 0.5, 1 and 2 mM) for 2 h (either with or without a 2-h pre-treatment with ubiquitin-activating enzyme E1 inhibitor as indicated) prior to cell lysis and analysis by western blotting. Asterisks indicate autocleavage fragments. (**G**) HeLa-T-REx Flp-In cells were treated with proteasome inhibitor MG132 for the indicated amount of time prior to cell lysis and analysis by western blotting. Asterisks indicate autocleavage fragments.

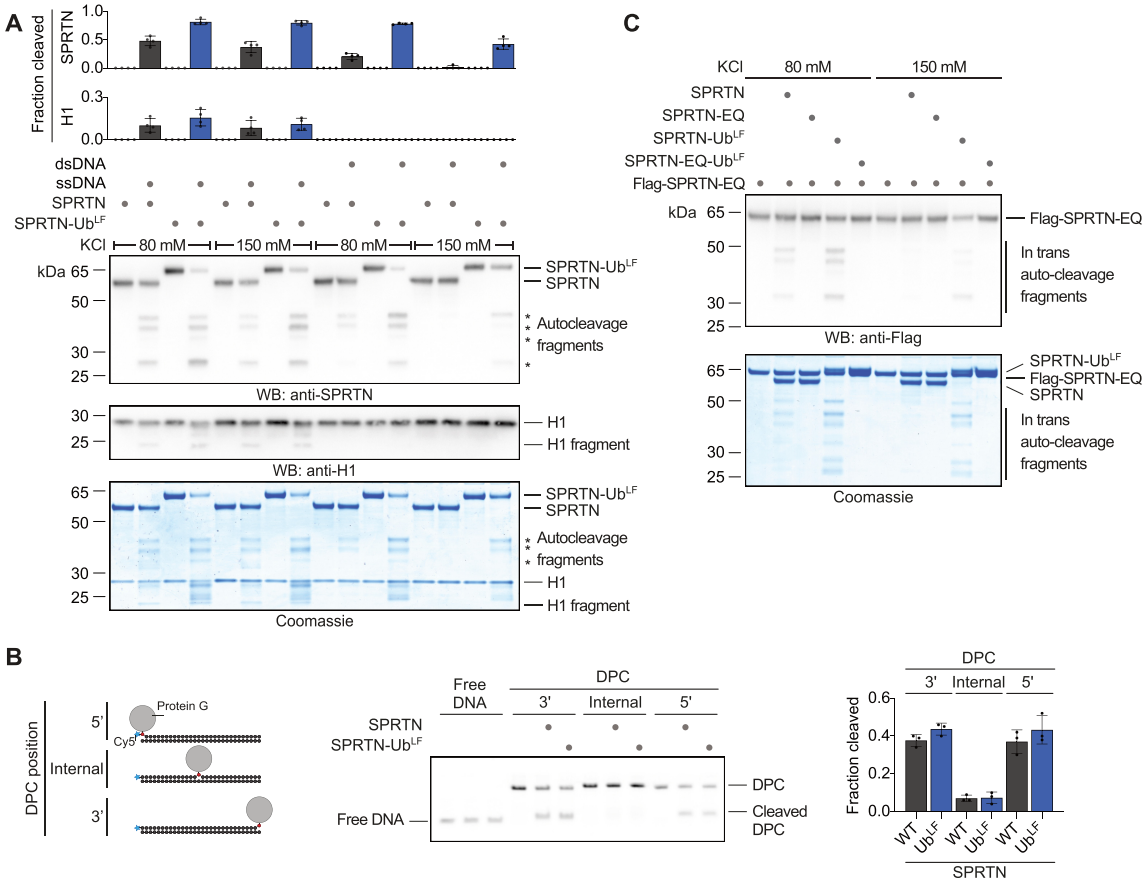


Figure 6. Monoubiquitylation promotes SPRTN autocleavage *in trans*. (**A**) Recombinant SPRTN or a linear SPRTN-Ubiquitin fusion (SPRTN-Ub^{LF}) (500 nM) were incubated with histone H1 alone or in the presence of either single- (ss) Virion or double-stranded (ds) RFI ΦX174 DNA (11.1 nM) for 60 min at 25°C. Salt concentrations were as indicated. Reactions were analysed by SDS-PAGE followed by western blotting and staining with InstantBlue Coomassie protein stain. Quantification of western blots of results of SPRTN and histone H1 cleavage: values represent the mean ± SD of four independent experiments. (**B**) Indicated model protein G-oligonucleotide conjugates (25nM) were incubated alone or in the presence of recombinant SPRTN (6.25 nM, WT or a linear SPRTN-Ubiquitin fusion (SPRTN-Ub^{LF})) for 2 h at 25°C prior to separation by native PAGE. Right panel, quantification of DPC cleavage: values represent the mean ± SD of three independent experiments. (**C**) Recombinant catalytically inactive Flag-SPRTN-EQ (500 nM) was incubated alone or in combination with active SPRTN (500 nM, WT or a linear SPRTN-Ubiquitin fusion (SPRTN-Ub^{LF})) in the presence of DNA (ΦX174 RFI dsDNA, 11.1 nM) for 60 min at 25°C. Salt concentrations were as indicated. Reactions were subjected to SDS-PAGE followed by staining with InstantBlue Coomassie protein stain and western blotting.

DISCUSSION

The regulatory ubiquitin switch revealed here is distinct from other prominent types of monoubiquitylation events occurring during genome maintenance. FANCD2 and FANCI are ubiquitylated by the Fanconi anemia core complex in a site-specific manner upon recruitment to chromatin during the repair of inter-strand crosslinks (33). PCNA is site-specifically monoubiquitylated as a response to stalled DNA synthesis, which fosters recruitment of translesion synthesis polymerases (34–37). In contrast, SPRTN monoubiquitylation is not triggered by DPC induction but instead appears to be a constitutive process. Our data demonstrate that the modification can have two distinct outcomes, both of which lead to inactivation of the enzyme (Figure 7). Firstly, monoubiquitylation primes SPRTN *in cis* for proteasomal degradation by fostering polyubiquitylation. Secondly, it further reduces the amount of active SPRTN by promoting autocleavage *in trans*. Importantly, SPRTN autocleavage requires the presence of DNA *in vitro*, which infers that monoubiquitylation-triggered autocleavage in cells is specifically affecting DNA-bound SPRTN molecules. If cells face DPC induction, USP7-mediated deubiquitylation appears to be important to stall this negative regulation in order to prolong the half-life of active DNA-bound SPRTN. Formaldehyde exposure triggers wide-spread ubiquitylation events in cells (38). It is thus attractive to speculate that in the presence of DPCs SPRTN's UBZ domain engages with specific ubiquitylation signals. In turn, this would expose the monoubiquitylation and thereby allow USP7 to remove the modification. Al-

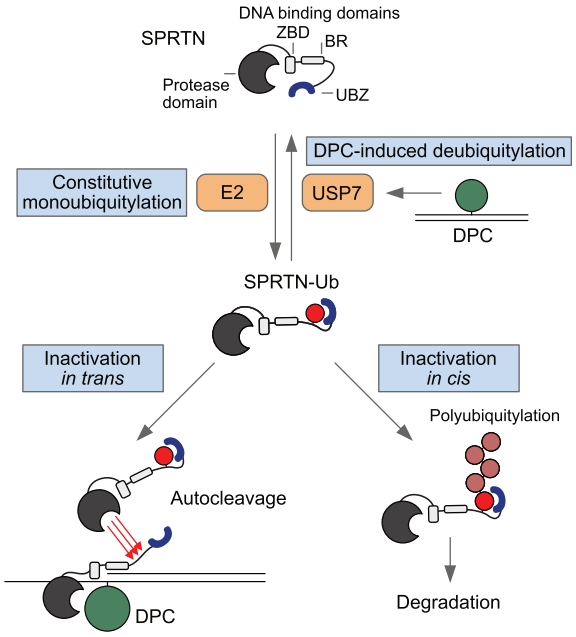


Figure 7. Regulation of SPRTN by monoubiquitylation and USP7. Proposed model for the regulation of SPRTN by monoubiquitylation and USP7-mediated deubiquitylation. SPRTN is subjected to constitutive promiscuous monoubiquitylation of its C-terminal tail. The modification is shielded by SPRTN's ubiquitin binding zinc-finger (UBZ). Monoubiquitylation affects SPRTN twofold. It primes SPRTN *in cis* for proteasomal degradation by inducing polyubiquitylation while also triggering inactivation by fostering autocleavage of other SPRTN molecules *in trans*. USP7 relieves this inhibition by deubiquitylating SPRTN upon induction of DNA–protein crosslinks (DPCs).

though the regulation described here is important to control SPRTN protein levels, it does not participate in the recruitment of SPRTN to chromatin. This finding is further supported by the fact that the RJALS syndrome variant SPRTN-ΔC retains a large degree of function despite lacking the UBZ and not being monoubiquitylated. In this context, it will be interesting to investigate whether the phenotypes observed in RJALS are caused, at least in part, by the loss of the negative regulation of SPRTN and not only by a reduction in DPC repair capacity. Notably, recruitment of SPRTN to UV-induced lesions (but not DPCs) has been shown to depend on the UBZ domain potentially indicating that this domain serves dual purposes (20–22). However, how SPRTN is recruited to DPCs remains controversial with evidence pointing towards ubiquitylation or SUMOylation signals (39,40). Understanding how the presence of crosslinks is signalled to DPC repair enzymes will be critical to decipher decision making during DPC repair. The recent identification of several additional proteases targeting DPCs implies that DPC repair pathway choice is a complex cellular process (13,39,41–45). At any rate, the intricate negative regulation described here highlights not only the complexity of DPC repair but also the importance of controlling SPRTN's potentially toxic proteolytic activity.

DATA AVAILABILITY

Data are available from the corresponding author upon reasonable request.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank S. Panier, G. Hewitt and R. Bellelli for discussions and comments on the manuscript.

FUNDING

LMU – China Scholarship Council Program [to S.Z.]; Peter and Traudl Engelhorn Foundation [to H.L.]; International Max-Planck Research School for Molecular Life Sciences [to A.C.A., P.W.]; European Research Council [ERC Starting Grant 801750 DNAProteinCrosslinks to J.S.]; Alfried Krupp Prize for Young University Teachers awarded by the Alfried Krupp von Bohlen and Halbach-Stiftung [to J.S.]; Center for Integrated Protein Sciences Munich (CIPSM) [to J.S.]. Funding for open access charge: ERC [Starting Grant 801750 DNAProteinCrosslinks].
Conflict of interest statement. None declared.

REFERENCES

1. Nakano,T., Miyamoto-Matsubara,M., Shoulkamy,M.I., Salem,A.M., Pack,S.P., Ishimi,Y. and Ide,H. (2013) Translocation and stability of replicative DNA helicases upon encountering DNA–protein cross-links. *J. Biol. Chem.*, **288**, 4649–4658.
2. Nakano,T., Ouchi,R., Kawazoe,J., Pack,S.P., Makino,K. and Ide,H. (2012) T7 RNA polymerases backed up by covalently trapped proteins catalyze highly error prone transcription. *J. Biol. Chem.*, **287**, 6562–6572.
3. Duxin,J.P., Dewar,J.M., Yardimci,H. and Walter,J.C. (2014) Repair of a DNA-protein crosslink by replication-coupled proteolysis. *Cell*, **159**, 346–357.
4. Barker,S., Weinfeld,M. and Murray,D. (2005) DNA–protein crosslinks: their induction, repair, and biological consequences. *Mutat. Res.*, **589**, 111–135.
5. Stingele,J., Bellelli,R. and Boulton,S.J. (2017) Mechanisms of DNA–protein crosslink repair. *Nat. Rev. Mol. Cell Biol.*, **18**, 563–573.
6. Nakano,T., Xu,X., Salem,A.M.H., Shoulkamy,M.I. and Ide,H. (2017) Radiation-induced DNA–protein cross-links: mechanisms and biological significance. *Free Radic. Biol. Med.*, **107**, 136–145.
7. Stingele,J., Schwarz,M.S., Bloemeke,N., Wolf,P.G. and Jentsch,S. (2014) A DNA-dependent protease involved in DNA-protein crosslink repair. *Cell*, **158**, 327–338.
8. Stingele,J., Bellelli,R., Alte,F., Hewitt,G., Sarek,G., Maslen,S.L., Tsutakawa,S.E., Borg,A., Kjaer,S., Tainer,J.A. *et al.* (2016) Mechanism and regulation of DNA-Protein crosslink repair by the DNA-dependent metalloprotease SPRTN. *Mol. Cell*, **64**, 688–703.
9. Vaz,B., Popovic,M., Newman,J.A., Fielden,J., Aitkenhead,H., Halder,S., Singh,A.N., Vendrell,I., Fischer,R., Torrecilla,I. *et al.* (2016) Metalloprotease SPRTN/DVC1 orchestrates replication-coupled DNA-protein crosslink repair. *Mol. Cell*, **64**, 704–719.
10. Lopez-Mosqueda,J., Maddi,K., Prgomet,S., Kalayil,S., Marinovic-Terzic,I., Terzic,J. and Dikic,I. (2016) SPRTN is a mammalian DNA-binding metalloprotease that resolves DNA–protein crosslinks. *Elife*, **5**, e21491.
11. Mórocz,M., Zsigmond,E., Tóth,R., Enyedi,M.Z., Pintér,L. and Haracska,L. (2017) DNA-dependent protease activity of human Spartan facilitates replication of DNA–protein crosslink-containing DNA. *Nucleic Acids Res.*, **45**, 3172–3188.

12. Reinking,H.K., Hofmann,K. and Stingele,J. (2020) Function and evolution of the DNA–protein crosslink proteases Wss1 and SPRTN. *DNA Repair (Amst.)*, **88**, 102822–102822.
13. Larsen,N.B., Gao,A.O., Sparks,J.L., Gallina,I., Wu,R.A., Mann,M., Raschle,M., Walter,J.C. and Duxin,J.P. (2019) Replication-coupled DNA-protein crosslink repair by SPRTN and the proteasome in xenopus egg extracts. *Mol. Cell*, **73**, 574–588.
14. Sparks,J.L., Chistol,G., Gao,A.O., Raschle,M., Larsen,N.B., Mann,M., Duxin,J.P. and Walter,J.C. (2019) The CMG helicase bypasses DNA-protein cross-links to facilitate their repair. *Cell*, **176**, 167–181.
15. Maskey,R.S., Kim,M.S., Baker,D.J., Childs,B., Malureanu,L.A., Jeganathan,K.B., Machida,Y., van Deursen,J.M. and Machida,Y.J. (2014) Spartan deficiency causes genomic instability and progeroid phenotypes. *Nat. Commun.*, **5**, 5744.
16. Lessel,D., Vaz,B., Halder,S., Lockhart,P.J., Marinovic-Terzic,I., Lopez-Mosqueda,J., Philipp,M., Sim,J.C., Smith,K.R., Oehler,J. *et al.* (2014) Mutations in SPRTN cause early onset hepatocellular carcinoma, genomic instability and progeroid features. *Nat. Genet.*, **46**, 1239–1244.
17. Ruijs,M.W.G., van Andel,R.N.J., Oshima,J., Madan,K., Nieuwint,A.W.M. and Aalfs,C.M. (2003) Atypical progeroid syndrome: an unknown helicase gene defect? *Am. J. Med. Genet. A*, **116A**, 295–299.
18. Maskey,R.S., Flatten,K.S., Sieben,C.J., Peterson,K.L., Baker,D.J., Nam,H.-J., Kim,M.S., Smyrk,T.C., Kojima,Y., Machida,Y. *et al.* (2017) Spartan deficiency causes accumulation of Topoisomerase 1 cleavage complexes and tumorigenesis. *Nucleic Acids Res.*, **45**, 4564–4576.
19. Reinking,H.K., Kang,H.S., Götz,M.J., Li,H.Y., Kieser,A., Zhao,S., Acampora,A.C., Weickert,P., Fessler,E., Jae,L.T. *et al.* (2020) DNA structure-specific cleavage of DNA-protein crosslinks by the SPRTN protease. *Mol. Cell*, **80**, 102–113.
20. Centore,R.C., Yazinski,S.A., Tse,A. and Zou,L. (2012) Spartan/C1orf124, a reader of PCNA ubiquitylation and a regulator of UV-induced DNA damage response. *Mol. Cell*, **46**, 625–635.
21. Mosbech,A., Gibbs-Seymour,I., Kagias,K., Thorslund,T., Beli,P., Povlsen,L., Nielsen,S.V., Smedegaard,S., Sedgwick,G., Lukas,C. *et al.* (2012) DVC1 (C1orf124) is a DNA damage-targeting p97 adaptor that promotes ubiquitin-dependent responses to replication blocks. *Nat. Struct. Mol. Biol.*, **19**, 1084–1092.
22. Davis,E.J., Lachaud,C., Appleton,P., Macartney,T.J., Nathke,I. and Rouse,J. (2012) DVC1 (C1orf124) recruits the p97 protein segregase to sites of DNA damage. *Nat. Struct. Mol. Biol.*, **19**, 1093–1100.
23. Hyer,M.L., Milhollen,M.A., Ciavarri,J., Fleming,P., Traore,T., Sappal,D., Huck,J., Shi,J., Gavin,J., Brownell,J. *et al.* (2018) A small-molecule inhibitor of the ubiquitin activating enzyme for cancer treatment. *Nat. Med.*, **24**, 186–193.
24. Cao,J., Wu,L., Zhang,S.M., Lu,M., Cheung,W.K., Cai,W., Gale,M., Xu,Q. and Yan,Q. (2016) An easy and efficient inducible CRISPR/Cas9 platform with improved specificity for multiple gene targeting. *Nucleic Acids Res.*, **44**, e149.
25. Zhitkovich,A. and Costa,M. (1992) A simple, sensitive assay to detect DNA–protein crosslinks in intact cells and in vivo. *Carcinogenesis*, **13**, 1485–1489.
26. Hoeller,D., Crosetto,N., Blagoev,B., Raiborg,C., Tikkanen,R., Wagner,S., Kowanetz,K., Breitling,R., Mann,M., Stenmark,H. *et al.* (2006) Regulation of ubiquitin-binding proteins by monoubiquitination. *Nat. Cell Biol.*, **8**, 163–169.
27. Woelk,T., Oldrini,B., Maspero,E., Confalonieri,S., Cavallaro,E., Di Fiore,P.P. and Polo,S. (2006) Molecular mechanisms of coupled monoubiquitination. *Nat. Cell Biol.*, **8**, 1246–1254.
28. Hoeller,D., Hecker,C.M., Wagner,S., Rogov,V., Dotsch,V. and Dikic,I. (2007) E3-independent monoubiquitination of ubiquitin-binding proteins. *Mol. Cell*, **26**, 891–898.
29. Huang,J., Zhou,Q., Gao,M., Nowsheen,S., Zhao,F., Kim,W., Zhu,Q., Kojima,Y., Yin,P., Zhang,Y. *et al.* (2020) Tandem deubiquitination and acetylation of sprtn promotes dna-protein crosslink repair and protects against aging. *Mol. Cell*, **79**, 824–835.
30. Perry,M., Kollala,S.S., Biegert,M., Su,G., Kodavati,M., Mallard,H., Kreiling,N., Holbrook,A. and Ghosal,G. (2020) USP11 deubiquitinates monoubiquitinated SPRTN to repair DNA–protein crosslinks. bioRxiv doi: https://doi.org/10.1101/2020.06.30.180471, 01 July 2020, preprint: not peer reviewed.
31. Koegl,M., Hoppe,T., Schlenker,S., Ulrich,H.D., Mayer,T.U. and Jentsch,S. (1999) A novel ubiquitination factor, E4, is involved in multiubiquitin chain assembly. *Cell*, **96**, 635–644.
32. Braten,O., Livneh,I., Ziv,T., Admon,A., Kehat,I., Caspi,L.H., Gonen,H., Bercovich,B., Godzik,A., Jahandideh,S. *et al.* (2016) Numerous proteins with unique characteristics are degraded by the 26S proteasome following monoubiquitination. *PNAS*, **113**, E4639–E4647.
33. Ceccaldi,R., Sarangi,P. and D'Andrea,A.D. (2016) The Fanconi anaemia pathway: new players and new functions. *Nat. Rev. Mol. Cell Biol.*, **17**, 337–349.
34. Moldovan,G.-L., Pfander,B. and Jentsch,S. (2007) PCNA, the maestro of the replication fork. *Cell*, **129**, 665–679.
35. Bienko,M., Green,C.M., Crosetto,N., Rudolf,F., Zapart,G., Coull,B., Kannouche,P., Wider,G., Peter,M., Lehmann,A.R. *et al.* (2005) Ubiquitin-binding domains in Y-family polymerases regulate translesion synthesis. *Science*, **310**, 1821–1824.
36. Stelter,P. and Ulrich,H.D. (2003) Control of spontaneous and damage-induced mutagenesis by SUMO and ubiquitin conjugation. *Nature*, **425**, 188–191.
37. Plosky,B.S., Vidal,A.E., Fernandez de Henestrosa,A.R., McLenigan,M.P., McDonald,J.P., Mead,S. and Woodgate,R. (2006) Controlling the subcellular localization of DNA polymerases iota and eta via interactions with ubiquitin. *EMBO J.*, **25**, 2847–2855.
38. Ortega-Atienza,S., Rubis,B., McCarthy,C. and Zhitkovich,A. (2016) Formaldehyde is a potent proteotoxic stressor causing rapid heat shock transcription factor 1 activation and lys48-linked polyubiquitination of proteins. *Am. J. Pathol.*, **186**, 2857–2868.
39. Borgermann,N., Ackermann,L., Schwertman,P., Hendriks,I.A., Thijssen,K., Liu,J.C., Lans,H., Nielsen,M.L. and Mailand,N. (2019) SUMOylation promotes protective responses to DNA–protein crosslinks. *EMBO J.*, **38**, e101496.
40. Fielden,J., Wiseman,K., Torrecilla,I., Li,S., Hume,S., Chiang,S.-C., Ruggiano,A., Narayan Singh,A., Freire,R., Hassanieh,S. *et al.* (2020) TEX264 coordinates p97- and SPRTN-mediated resolution of topoisomerase 1-DNA adducts. *Nat. Commun.*, **11**, 1274.
41. Bhargava,V., Goldstein,C.D., 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.
42. Dokshin,G.A., Davis,G.M., Sawle,A.D., Eldridge,M.D., Nicholls,P.K., Gourley,T.E., Romer,K.A., Molesworth,L.W., Tatnell,H.R., Ozturk,A.R. *et al.* (2020) GCNA interacts with spartan and topoisomerase II to regulate genome stability. *Dev. Cell*, **52**, 53–68.
43. Serbyn,N., Noireterre,A., Bagdiul,I., Plank,M., Michel,A.H., Loewith,R., Kornmann,B. and Stutz,F. (2019) The aspartic protease ddi1 contributes to dna-protein crosslink repair in yeast. *Mol. Cell*, **77**, 1066–1079.
44. Svoboda,M., Konvalinka,J., Trempe,J.F. and Grantz Saskova,K. (2019) The yeast proteases Ddi1 and Wss1 are both involved in the DNA replication stress response. *DNA Repair (Amst.)*, **80**, 45–51.
45. Kojima,Y., Machida,Y., Palani,S., Caulfield,T.R., Radisky,E.S., Kaufmann,S.H. and Machida,Y.J. (2020) FAM111A protects replication forks from protein obstacles via its trypsin-like domain. *Nat. Commun.*, **11**, 1318.

SUPPLEMENTARY DATA

A ubiquitin switch controls autocatalytic inactivation of the DNA-protein crosslink repair protease SPRTN

Shubo Zhao, Anja Kieser, Hao-Yi Li, Hannah K. Reinking, Pedro Weickert, Simon Euteneuer, Denitsa Yaneva, Aleida C. Acampora, Maximilian J. Götz, Regina Feederle and Julian Stingele

Figure S1 (related to Figure 1). E3-independent monoubiquitylation SPRTN *in vitro*.

A. Analysis of monoubiquitylation of truncated SPRTN variants. Plasmids encoding the Strep-tagged SPRTN-Δ400 truncation (carrying the indicated lysine to arginine (KR) substitutions. K407 was not replaced in the SPRTN-Δ400-8KR variant) were transiently transfected in HeLa T-REx Flp-In cells. Expression of SPRTN was induced by addition of doxycycline for 16 hours before cells were lysed and subjected to immunoprecipitation using anti-Strep beads followed by western blotting. Western blotting of cell lysates against Tubulin serves as loading control.

B. Twenty-nine human E2 ubiquitin conjugating enzymes (2 μM) were incubated together with SPRTN-EQ (2 μM), E1 ubiquitin activating enzyme (100 nM), ubiquitin (50 μM), DNA (11.1 nM ΦX174 single-stranded DNA) and ATP (2 mM) for 1.5 hours at 30°C. Reactions were stopped by addition of LDS sample buffer and subjected to SDS-PAGE followed by staining with InstantBlue Coomassie protein stain.

C. *In vitro* ubiquitylation assays containing SPRTN-EQ (410 nM), E1 ubiquitin activating enzyme (300 nM), UBE2D3 (16 μM), ubiquitin WT or no-Lys, N-terminally biotinylated ubiquitin (50 μM) and ATP (2 mM) were incubated for 1.5 hours at 30°C. Reactions were stopped by addition of LDS sample buffer and subjected to SDS-PAGE followed by staining with InstantBlue Coomassie protein stain.

Figure S1 (related to Figure 1)

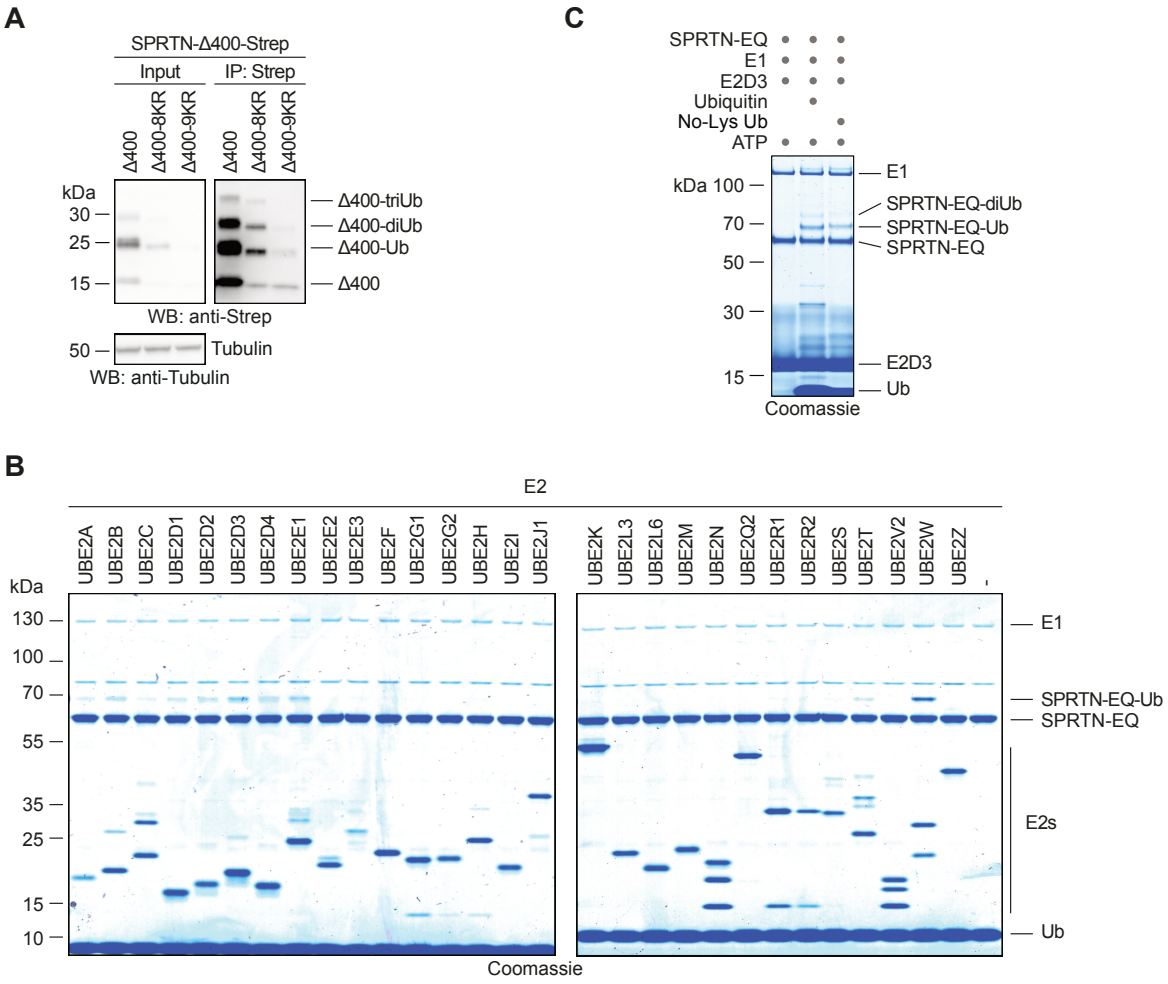


Figure S2 (related to Figure 2). An *in vitro* screen reveals that USP7 targets ubiquitylated SPRTN.

A. Result of the primary screen schematically depicted in Figure 2A. Twenty-four pools (each containing three different lysates of *E. coli* cells expressing different DUB cDNAs) were incubated for 30 min at 25°C together with purified partially-ubiquitylated YFP-SPRTN-EQ-Strep in three experimental sets (A-K, L-N, O-X). Reactions were stopped by addition of LDS sample buffer and analysed by SDS-PAGE and western blotting using anti-Strep antibody. Lysates of BL21 cells served as negative control, the unspecific deubiquitylating activity of the catalytic domain of the USP2 ($USP2^{cd}$) as positive control.

B. Deubiquitylation activity of indicated partially purified DUBs was compared using the commercial Ubiquitin-Rhodamine cleavage assay, which measures the release of a rhodamine fluorophore C-terminally conjugated to ubiquitin with cleavage resulting in increased fluorescence. Left panel, increase in rhodamine fluorescence over time. Right panel, initial velocities of the deubiquitylating reactions. The catalytic domain of the USP2 ($USP2^{cd}$) served as positive control. Values represent the mean ± SD of two technical replicates.

C. Indicated partially purified DUBs were incubated for 30 min at 25°C together with purified partially-ubiquitylated YFP-SPRTN-EQ-Strep. Reactions were stopped by addition of LDS sample buffer and analysed by SDS-PAGE and western blotting using anti-Strep antibody.

Figure S2 (related to Figure 2)

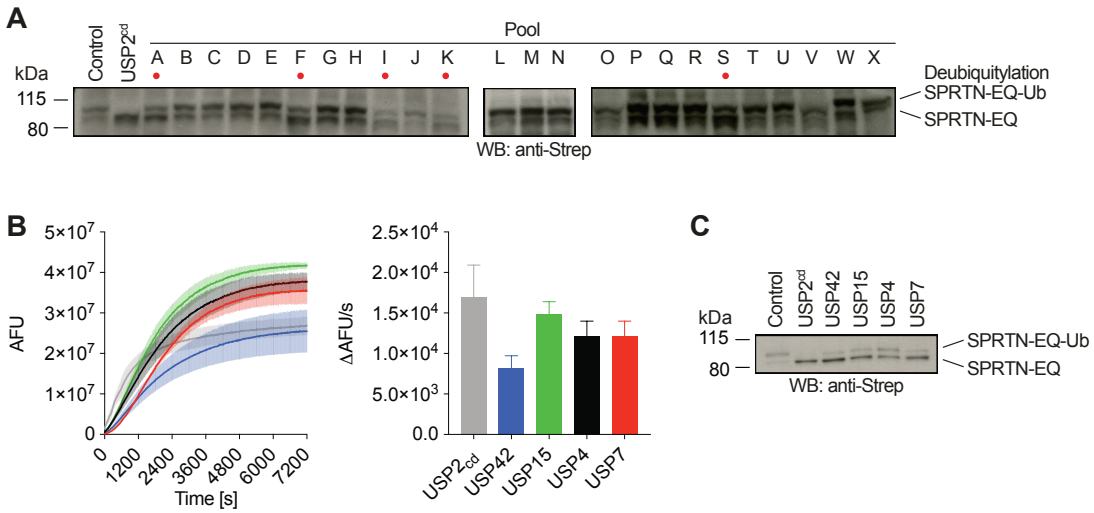


Figure S3 (related to Figure 4). USP7 deubiquitylates SPRTN upon DPC induction.

A. Western blot analysis of DLD1 WT or *USP7* KO cells transiently transfect with empty vector, YFP-tagged USP7 WT or the catalytically inactive USP7 CS variant, as indicated. Asterisks indicates a cross-reactive band.

B. Plasmids encoding Flag-tagged full-length USP7, VCPIP1 or USP11 (WT or catalytically inactive variants) or the empty vector were transiently transfected in HeLa T-REx Flp-In cells stably expressing doxycycline-inducible YFP-SPRTN-Strep. Binding was analysed by co-immunoprecipitation using anti-Flag beads followed by western blotting.

C. DLD1 cells transfected with siRNA pools targeting USP7, VCPIP1 or USP11 were treated for 2 hours with 2 mM formaldehyde (FA) 72 hours after transfection. Cells were lysed in LDS sample buffer and analysed by SDS-PAGE followed by western blotting.

D. HAP1 WT, *USP11* KO and *VCPIP1* KO cells were treated with the indicated formaldehyde concentrations for 72 hours. Cell viability was then determined using the alamarBlue cell viability assay. Values represent the mean ± SD of three technical replicates normalized to the mean of untreated controls of each cell line.

E. HAP1 WT, *USP7* KO, *VCPIP1* KO and *USP11* KO cells were treated with 2 mM formaldehyde (FA) for 3 hours. Cells were either lysed directly in LDS sample buffer (total) or subjected to chromatin fractionation. Samples were then analysed by SDS-PAGE followed by western blotting. Asterisks indicates a cross-reactive band.

Figure S3 (related to Figure 4)

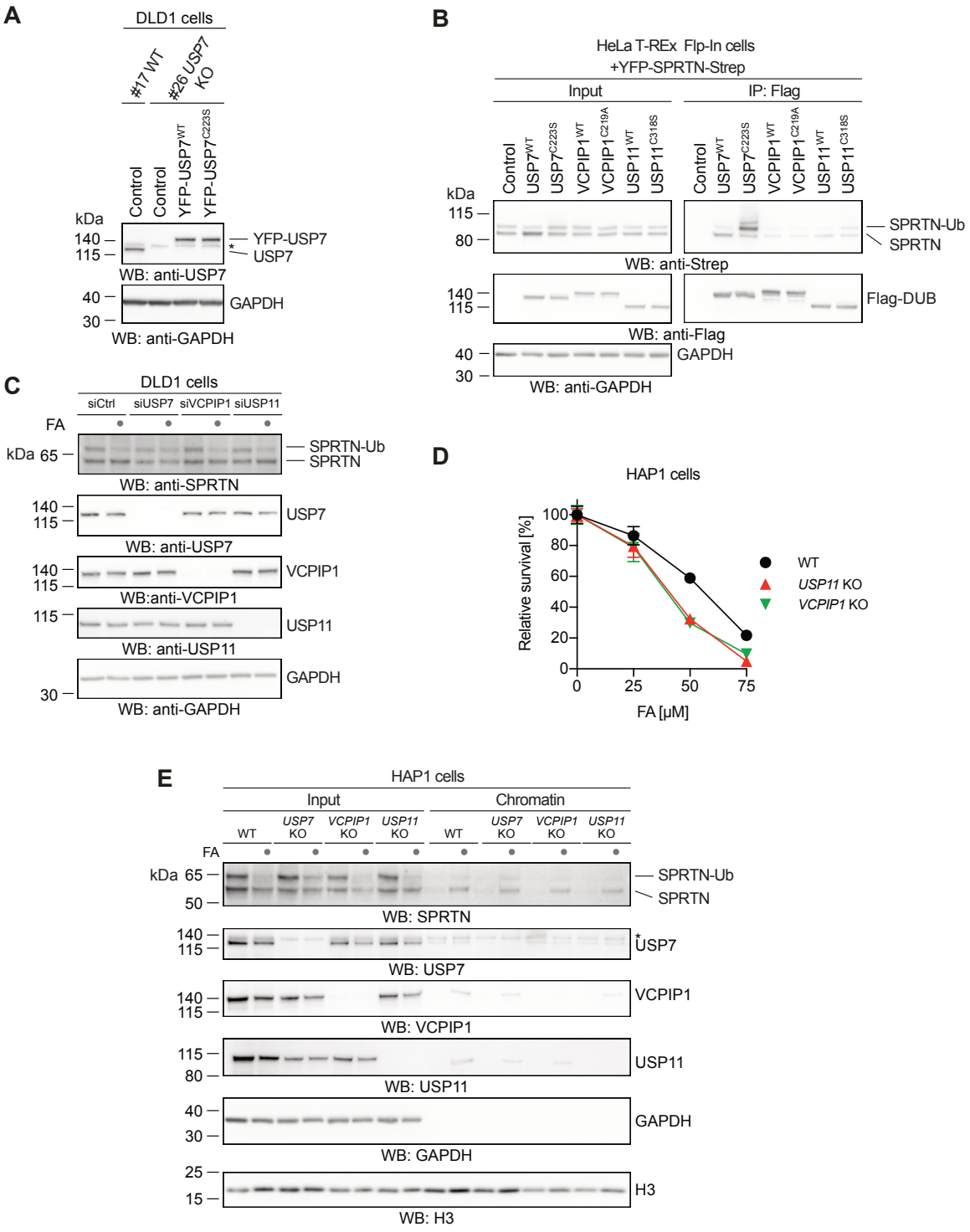


Figure S4 (related to Figure 5). Monoubiquitylation promotes SPRTN degradation and autocleavage.

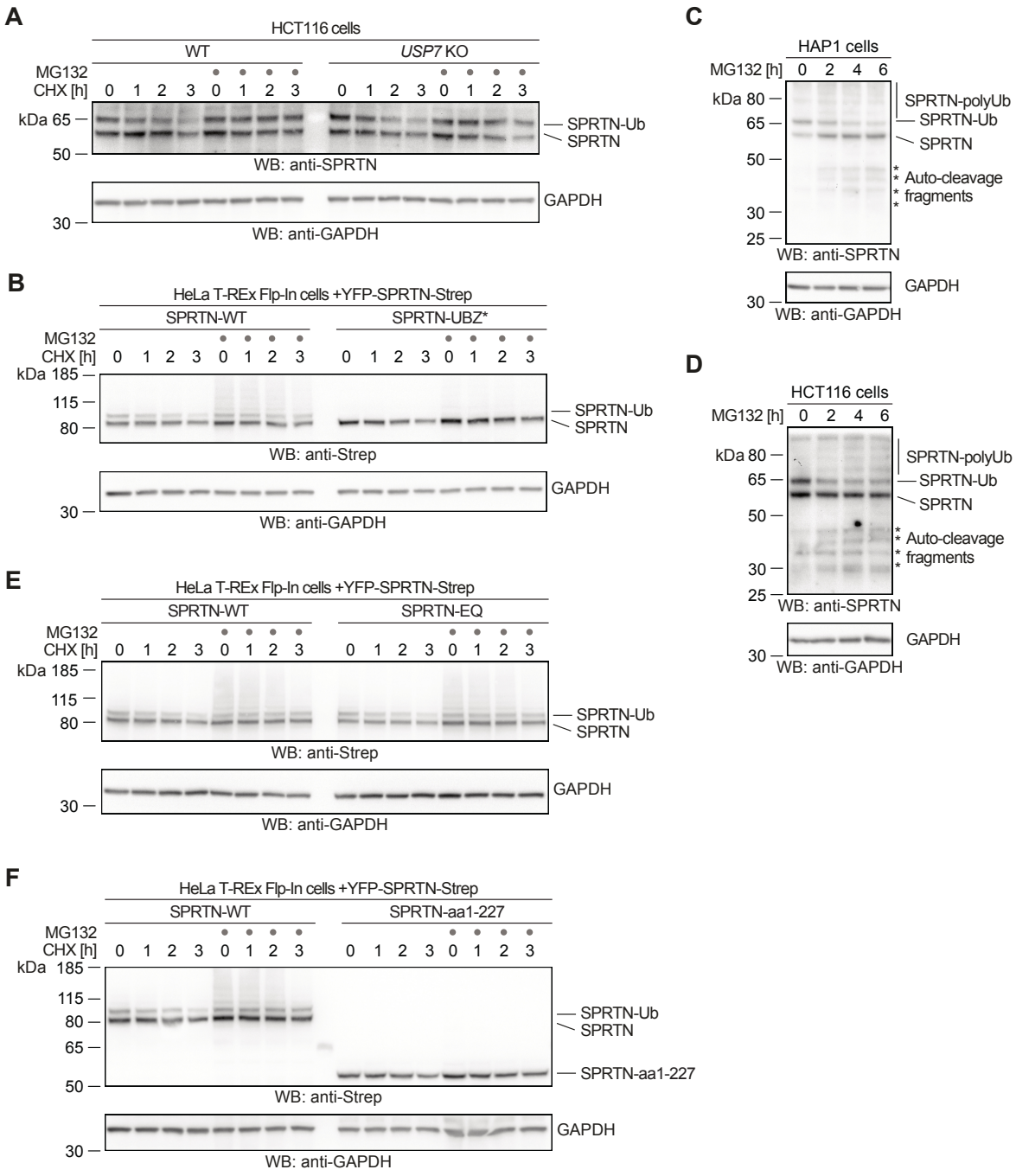
A. Stability of endogenous SPRTN was determined in HCT116 WT or *USP7* KO cells using a cycloheximide-chase experiment. Cells were incubated in the presence of cycloheximide for the indicated amount of time (with or without a 2-hour pre-treatment with the proteasome inhibitor MG132) prior to cell lysis and analysis by western blotting.

B. Stability of stably expressed doxycycline-inducible YFP-SPRTN-Strep and YFP-SPRTN-UBZ*-Strep was determined in HeLa-T-REx Flp-In cells using a cycloheximide-chase experiment. Cells were incubated in the presence of cycloheximide for the indicated amount of time (with or without a 2-hour pre-treatment with the proteasome inhibitor MG132) prior to cell lysis and analysis by western blotting.

C-D. HAP1 or HCT116 cells were treated with proteasome inhibitor MG132 for the indicated amount of time prior to cell lysis and analysis by western blotting. Asterisks indicate autocleavage fragments.

E-F. Stability of stably expressed doxycycline-inducible YFP-SPRTN-Strep and catalytically inactive YFP-SPRTN-EQ-Strep or the truncated YFP-SPRTN-aa1-227-Strep was determined in HeLa-T-REx Flp-In cells using a cycloheximide-chase experiment. Cells were incubated in the presence of cycloheximide for the indicated amount of time (with or without a 2-hour pre-treatment with the proteasome inhibitor MG132) prior to cell lysis and analysis by western blotting.

Figure S4 (related to Figure 5)



2.3 Function and evolution of the DNA-protein crosslink proteases Wss1 and SPRTN

Hannah K. Reinking, Kay Hofmann, Julian Stingele, Function and evolution of the DNA-protein crosslink proteases Wss1 and SPRTN, *DNA Repair*, Volume 88, 6 February 2020, Pages 102822, https://doi.org/10.1016/j.dnarep.2020.102822 (Review article)

Summary

This review article discusses recent findings about the different DNA-protein crosslink proteases. The article first summarizes the main DPC-causing sources and describes DPC repair by canonical DNA repair pathways. In addition, it portrays the initial discovery of DPC proteases and reveals their structural and mechanistical similarities. An in-depth phylogenetic analysis of Wss1/SPRTN proteases is performed, which uncovers overlapping domain architectures and protein-protein interaction motifs with other potential DPC proteases.

Author contribution

I wrote the manuscript together with Kay Hofmann and Julian Stingele.

DNA Repair 88 (2020) 102822

Contents lists available at ScienceDirect

DNA Repair

journal homepage: www.elsevier.com/locate/dnarepair

Review Article

Function and evolution of the DNA-protein crosslink proteases Wss1 and SPRTN

Hannah K. Reinking[a,b], Kay Hofmann[c], Julian Stingele[a,b,*]

[a] Gene Center, Ludwig-Maximilians-University Munich, Munich, Germany
[b] Department of Biochemistry, Ludwig-Maximilians-University Munich, Munich, Germany
[c] Institute for Genetics, University of Cologne, Germany

ARTICLE INFO

Keywords:
DNA-protein crosslink
SPRTN
Wss1
ACRC
GCNA1
DDI1

ABSTRACT

Covalent DNA-protein crosslinks (DPCs) are highly toxic DNA adducts, which interfere with faithful DNA replication. The proteases Wss1 and SPRTN degrade DPCs and have emerged as crucially important DNA repair enzymes. Their protective role has been described in various model systems ranging from yeasts, plants, worms and flies to mice and humans. Loss of DPC proteases results in genome instability, cellular arrest, premature ageing and cancer predisposition. Here we discuss recent insights into the function and molecular mechanism of these enzymes. Furthermore, we present an in-depth phylogenetic analysis of the Wss1/SPRTN protease continuum. Remarkably flexible domain architectures and constantly changing protein-protein interaction motifs indicate ongoing evolutionary dynamics. Finally, we discuss recent data, which suggest that further partially-overlapping proteolytic systems targeting DPCs exist in eukaryotes. These new developments raise interesting questions regarding the division of labour between different DPC proteases and the mechanisms and principles of repair pathway choice.

1. Introduction

DNA integrity is constantly threatened by various types of endogenous DNA damages [1]. Failure to repair these lesions can have severe consequences. Germline mutations in genes encoding crucial DNA repair proteins cause various human premature ageing and cancer predisposition syndromes [2]. Recent insights suggest that a particular class of endogenous lesions - covalent DNA-protein crosslinks (DPCs) - is especially detrimental to genome stability and thus cellular integrity. DPCs are highly toxic because they block chromatin transactions due to their bulkiness. Crosslinked proteins interfere with transcription by impeding RNA polymerase [3,4]. Similarly, DPCs obstruct progression of the replicative helicase and DNA polymerases during replication [5–7]. Moreover, it is conceivable that DPCs interfere with other dynamic chromatin processes such as chromatin remodelling or higher-order chromatin organization through looping by SMC proteins [8,9].

1.1. Sources of DNA-protein crosslinks

DPCs form by covalent trapping of normally transient protein-DNA interactions. Unspecific crosslinking can be induced by reactive endogenous metabolites such as formaldehyde or acetaldehyde [10]. Formaldehyde is even produced in direct vicinity of DNA during histone demethylation [11,12]. Moreover, several intermediates of DNA metabolism have intrinsic crosslinking properties. 5-formylcytosine occurs during TET-protein mediated oxidation of 5-methylcytosine and can form covalent crosslinks with primary amines of histones [13–15]. Moreover, abasic sites, which arise either spontaneously or actively during base excision repair and DNA demethylation crosslink to proteins *in vitro* [16–19].

In addition, DNA repair enzymes, such as polymerase ß can become covalently trapped during base excision repair attempts [20,21]. Similarly, the covalent reaction intermediates of topoisomerases can be stabilized by nearby DNA distortions leading to DPC formation [22,23]. Trapping of topoisomerases 1 (TOP1) and 2 (TOP2) can also be induced by so called enzyme poisons: small molecules, which intercalate within the protein-DNA interface [24,25]. Enzyme poisons targeting TOP1 (irinotecan, topotecan) and TOP2 (etoposide, doxorubicin) are frequently used as chemotherapeutic agents. In general, many chemotherapeutic regimens include agents inducing DNA-protein crosslinking [26]. This includes cisplatin derivatives such as carboplatin and oxaliplatin, which crosslink proteins unspecifically to DNA [27]. In contrast, the cytosine analogue 5-aza-deoxycytidine (5-aza-dC) specifically targets DNA methyltransferase 1 (DNMT1). 5-aza-dC is

* Corresponding author.
E-mail address: stingele@genzentrum.lmu.de (J. Stingele).

https://doi.org/10.1016/j.dnarep.2020.102822
Received 23 December 2019; Received in revised form 3 February 2020; Accepted 5 February 2020
Available online 06 February 2020

incorporated into DNA during replication, where it is recognized by DNMT1 as a pseudo-substrate leading to covalent trapping of the enzyme [28]. Finally, ionizing radiation (IR) can induce DPCs [29]. Importantly, IR-induced DPCs form preferentially under low oxygen conditions; *e.g.* in hypoxic tumour regions during radiotherapy [10,30].

1.2. Canonical DNA repair pathways targeting DNA-protein crosslinks

DPCs are a challenge for cellular repair machineries, because they constitute an extremely diverse class of lesions. Not only with respect to the identity of the crosslinked protein – virtually every chromatin protein can become crosslinked – but also due to the various DNA structures involved. Some repair proteins target only specific DPCs. Tyrosyl-DNA-phosphodiesterase 1 (TDP1) and 2 (TDP2) hydrolyse specifically the covalent bond between DNA and the active site tyrosines of trapped TOP1 and TOP2, respectively [31–33]. In addition, canonical DNA repair pathways such as homologous recombination (HR) and nucleotide excision repair (NER) have been implicated in DPC repair. Bacteria, yeast and human cells lacking crucial NER genes are extremely sensitive to reactive aldehydes [34–36]. Moreover, plasmid-borne DPCs are repaired less efficiently in NER-deficient human cell lines [37]. However, NER activity appears to be restricted to smaller peptides or protein adducts ($< 8-10$ kDa). Furthermore, the nuclease/ATPase complex MRN (consisting of MRE11, RAD50 and NBS1) has been shown to clip off DPCs located at the end of DNA double-strand breaks (DSBs), such as those formed by TOP2 or the meiotic topoisomerase-related enzyme SPO11 [38–40]. It had been assumed that these canonical repair mechanisms would suffice to protect cells from DPCs. The existence of DPC proteases targeting the protein component of DPCs was thus entirely unexpected. The discovery of dedicated DPC repair enzymes has triggered intense research efforts to understand the biology of DPCs and their repair.

2. DPC proteases

2.1. Discovery of DPC proteases

The yeast metalloprotease Wss1 (weak suppressor of smt3) was initially described as a high copy suppressor of a hypomorphic SUMO variant, but its molecular function had remained unclear [41]. A function for Wss1 in DPC repair was unravelled by a genome-wide screen, which revealed that a loss of Wss1 results in synthetic sickness when combined with a loss of Tdp1 [42]. *Δtdp1Δwss1* yeast strains accumulate unrepaired Top1-DPCs, which trigger cell cycle arrest and DNA damage checkpoint activation. Importantly, Wss1 not only protects cells against topoisomerase-induced DPCs but also against non-enzymatic DPCs such as those induced by formaldehyde. Concurrent results obtained in *Xenopus laevis* egg extracts indicated that a similar proteolytic activity exists in higher eukaryotes [43]. Bioinformatic evidence highlighted similarities between Wss1 proteases found in yeasts and plants and the metazoan SPRTN proteins [44,45]. SPRTN was initially described as a regulator of translesion synthesis, but its exact role remained controversial [45–49]. Indeed, several studies revealed SPRTN as the DPC protease acting in higher eukaryotes [50–54]. Interestingly, a SPRTN-like protease, acidic repeat-containing protein (ACRC, also called germ cell nuclear antigen, GCNA) has recently been implicated as an additional DPC protease in metazoans [55–58].

The loss of DPC proteases has severe consequences on genome stability and cellular integrity. In yeast, the loss of Wss1 results in increased recombination and dramatic rates of gross chromosomal rearrangements [42,59]. Plants lacking Wss1 display severe developmental defects [60,61]. In flies, SPRTN deficiency leads to fertility defects in females, which stems from a failure to faithfully replicate and segregate paternal chromosomes during the first mitotic divisions after fertilization [62,63]. This results in a loss of paternal chromosomes and leads to the formation of haploid embryos consisting exclusively of maternal DNA. The gene encoding SPRTN in *Drosophila* is accordingly named *maternal haploid*. In mammals, SPRTN is essential for cellular viability [64,65]. Moreover, even a partial loss of SPRTN function has severe effects in humans. Ruijs-Aalfs syndrome (RJALS) is caused by a hypomorphic SPRTN variant, which lacks the C-terminal tail of the enzyme. RJALS patients display premature ageing and early-onset hepatocellular carcinomas [66,67]. In agreement, cancer predisposition and premature ageing can also be observed in a hypomorphic *Sprtn* mouse model [52]. In contrast to the universal importance of SPRTN, the eukaryotic SPRTN-like protease ACRC/GCNA seems to be especially important in meiotic cells. It has been proposed that ACRC/GCNA is required to process toxic SPO11 adducts during meiotic recombination [55,56]. As a consequence, ACRC/GCNA deficiency leads to fertility defects. Interestingly, ACRC/GCNA was identified as the antigen of a monoclonal antibody, which had been used for decades to specifically stain nuclei of germ cells [58,68].

2.2. Replication-coupled DNA-protein crosslink repair

Experiments analysing the repair of DPC-containing plasmids in *X.laevis* egg extracts revealed the precise order of events and replication-coupled nature of proteolytic DPC processing [43,69,70]. A lagging strand DPC can be immediately bypassed by the replicative CMG helicase, whereas bypass of a leading strand DPC requires the activity of an additional helicase RTEL1 (Fig. 1). DPC bypass by the helicase results in the same scenario independent of the adduct's location: a DPC-stalled DNA polymerase, which then leads to recruitment of SPRTN by an unknown signalling mechanism and degradation of the protein adduct.

However, whether SPRTN acts exclusively in a replication-coupled manner is not entirely clear. Genetic evidence suggests that SPRTN has a function outside of S-phase. First, SPRTN is recruited to chromatin prior to replication in fly embryos [63]. Second, worms lacking SPRTN are sensitive to formaldehyde exposure even when arrested in a non-dividing larval state [50]. It is conceivable that DPCs are sensed outside of S-phase by the stalling of other chromatin processes (*e.g.* transcription). Although, currently little evidence of transcription-coupled DPC repair exists.

2.3. Structure and mechanism of DPC proteases

SPRTN and Wss1 protect cells against formaldehyde-induced DPCs, which potentially involve any chromatin protein. It is remarkable that one enzyme has the ability to target so many different substrates. Notably, analysis of SPRTN degradation products by mass spectrometry did not detect any specific cleavage sites indicating the lack of a preferred amino acid sequence [51]. The crystal structures obtained of the protease domains of Wss1 (*S. pombe* and *S. cerevisiae*) and SPRTN (*H. sapiens*) revealed highly similar compact protease domains (Fig. 2) [50,71,72].

Strikingly, the active sites are extremely open with only few specificity-generating features. This might explain how these enzymes can process such a large variety of substrates. However, such an unspecific protease activity requires tight control in order to prohibit uncontrolled proteolysis. In fact, SPRTN's and Wss1's protease activities are strictly DNA-dependent. Wss1 is inactive when incubated on its own *in vitro*. However, the enzyme becomes strongly activated in the presence of DNA, which allows it to process various DNA-bound substrates [42]. SPRTN displays an identical DNA-induced activity [50,51]. Notably, long single-stranded DNA appears to activate the enzyme much more efficiently than double-stranded DNA. Thus, processing of substrate proteins depends on the presence of single-stranded DNA. The molecular nature of protease activation is not well understood, but likely involves DNA-induced conformational changes [50]. Both, SPRTN and Wss1 bear highly basic unspecific DNA binding domains, which are required for their activity. Interestingly, the recent crystal structure of human SPRTN revealed a second unexpected DNA binding domain

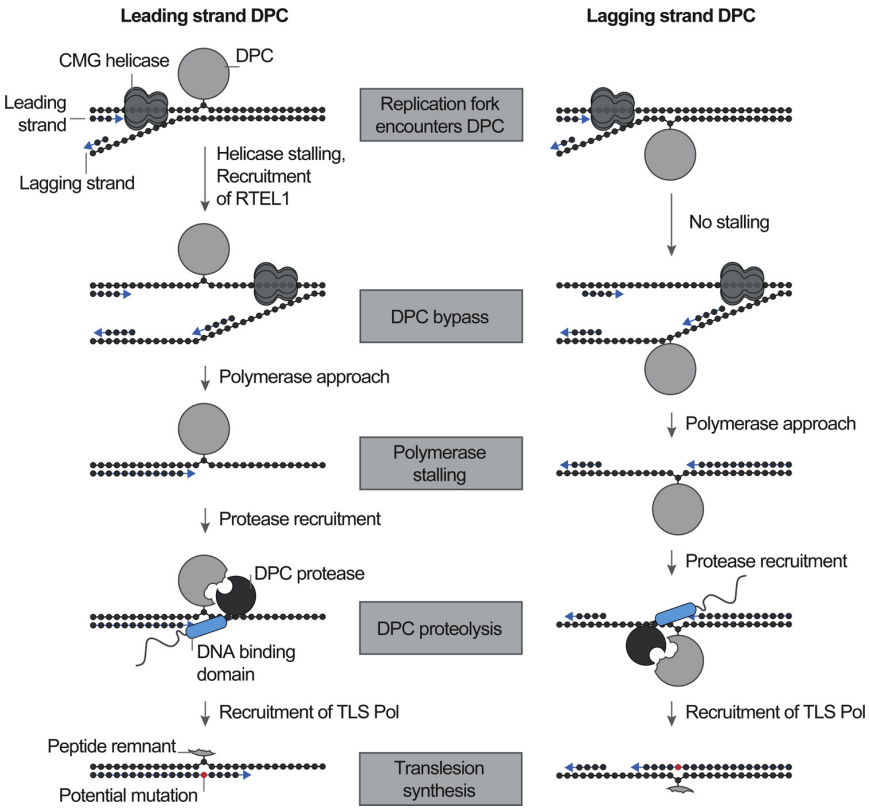


Fig. 1. Replication-coupled DPC repair by DPC proteases. A replisome encountering a DPC will be affected differentially depending on the location of the protein adduct. A leading strand DPC stalls progression of the replicative CMG helicase. Bypass of the lesion by the helicase appears to require unwinding of the fork by a second helicase, RTEL1. This allows the DNA polymerase to approach the adduct. In contrast, a lagging strand DPC does not hinder progression of the replicative helicase, but will stall the lagging strand DNA polymerase. The DPC-stalled DNA polymerase triggers recruitment of the DPC protease by an unknown signalling mechanism. Degradation of the protein component of the DPC allows translesion synthesis (TLS) past the remaining peptide remnant, which can result in the incorporation of mutations.

[72]. A highly unusual zinc-binding domain (ZBD) appears to restrict access to the active site. It has been proposed that the ZBD constitutes a single-strand DNA binding domain. Mutations within the ZBD do affect catalytic activity of the enzyme. However, the exact molecular role of the ZBD remains to be determined. Wss1 does not contain a ZBD, but it has been proposed that access to the active site is controlled through a cysteine switch [73]. Cysteine switches are a common regulatory feature in matrix-metalloproteases, in which a cysteine residue blocks access to the active site by binding directly to the zinc ion [74].

3. Evolution of DPC proteases

In addition to their remarkably similar enzymatic activities and protein structures, Wss1 and SPRTN display highly related domain organizations and protein-protein interaction motifs. Both proteins interact with the segregase p97 (also called VCP or Cdc48) and either ubiquitin or the ubiquitin-like protein SUMO. Despite these similarities no consensus has been reached regarding the evolutionary relationship between Wss1 and SPRTN [44,75]. We thus decided to readdress this question by conducting an in-depth phylogenetic analysis, which revealed an evolutionary continuum of related proteases including SPRTN and Wss1 (Fig. 3A). This spectrum of enzymes contains several related bacterial protease families and four families containing eukaryotic representatives. The WLM/Wss1 family contains three subfamilies (W1, W2 and W3) with mainly fungal representatives. The largely eukaryotic SPRTN family contains three subfamilies (S1, S2 and S3). Human SPRTN and all other metazoan representatives belong to subfamily S1. Moreover, ACRC/GCNA proteases are part of the same continuum of enzymes with fungal (A1) and animal (A2) subfamilies. In addition, we identified a fourth family of eukaryotic proteases

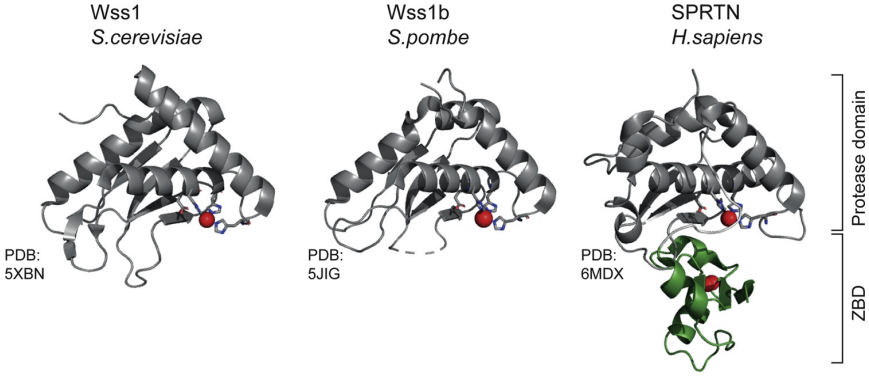


Fig. 2. Crystal structures of DPC protease domains. Structure of the protease domains of Wss1 (S. pombe and S. cerevisiae) and SPRTN (H. sapiens) in cartoon representation. The protease fold is displayed in dark grey, the active site metal ion in red and the zinc-binding domain (ZBD) in green. The catalytic centers are highly solvent exposed and no obvious substrate-binding cleft can be observed. The ZBD domain appears to restrict access to the active site of human SPRTN.

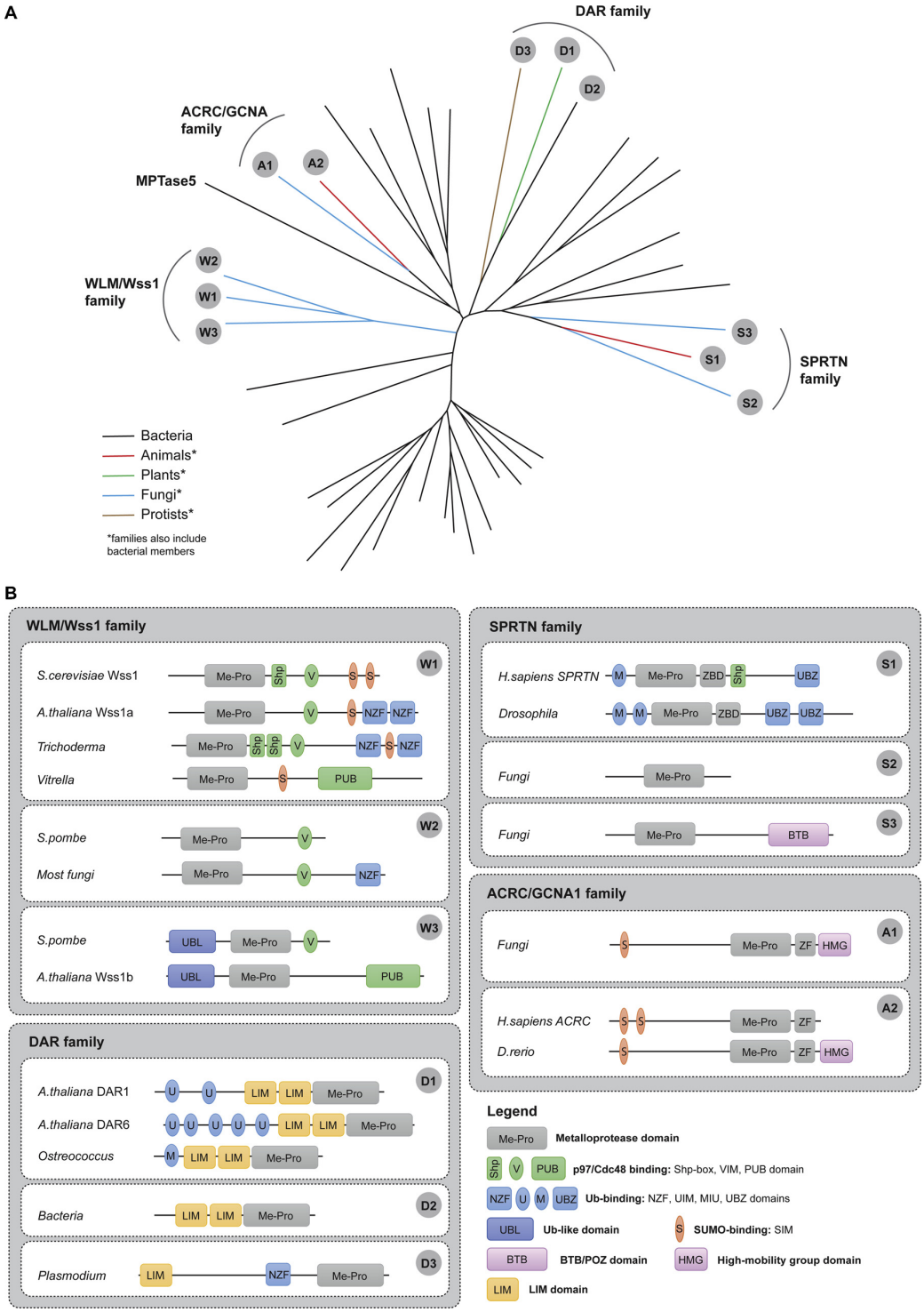


Fig. 3. A) Phylogenetic tree of SPRTN, Wss1 and ACRC proteases. The SPRTN/Wss1 continuum of enzymes was revealed using database searches with profiles and Hidden Markov Models of Wss1, SPRTN and ACRC, which identified roughly 12,000 sequences after removing redundant and highly similar sequences, which were then clustered. Evolutionary relationships were deduced by aligning representative sequences of the metalloprotease domains of each cluster, which were used to generate a phylogenetic tree. Families including exclusively bacterial proteases are labelled black. Families with eukaryotic members are color-coded as indicated. **B) Representative domain/motif architectures of the families with eukaryotic participation.** Bacterial subfamilies are typically restricted to the metalloprotease domain, often followed by a zinc-finger region. Several bacterial families also contain additional domains different from those found in the eukaryotes.

belonging to the Wss1/SPRTN spectrum. This DAR family includes the plant family of DA-1 related ubiquitin receptors (DAR1-6), which we reveal to contain Wss1/SPRTN-like metalloprotease domains [76]. Our analysis also confirms the previously identified presence of SPRTN/Wss1-like metalloprotease domains in MPTase5 bacterial polymorphic toxins [77]. The bacterial representatives of all families are entirely uncharacterized and whether they have similar functions to SPRTN and Wss1 is currently unknown.

Next, we searched each subgroup containing eukaryotic representatives for common domains and protein-protein interaction motifs (Fig. 3B). The diversity of domains even within subfamilies is remarkable. Binding motifs/domains for recognition of ubiquitin and SUMO are a recurring theme and can be found in different combinations in almost all eukaryotic subfamilies. In particular the W1 subfamily (containing *S.cerevisiae* Wss1) includes members with various binding domains for Cdc48, ubiquitin, and/or SUMO – often in variable copy numbers, which indicates dynamic ongoing evolution. Only the S1 subfamily (containing *H.sapiens* SPRTN) contains the aforementioned ZBD. Notably, the zinc-finger domains found downstream of the metalloprotease domain in ACRC/GCNA enzymes is distinct to SPRTN's ZBD. Importantly, the phylogenetic analysis presented here also shows that Wss1 and SPRTN are not direct orthologs. Apparently, they independently acquired p97/Cdc48 and Ub/SUMO binding motifs, which makes them an example for convergent evolution within the same phylogenetic subgroup.

4. Regulation of DPC proteases

The abundance of recognition motifs for ubiquitin or SUMO indicates dynamic regulation of DPC proteases in various organisms. However, the nature of regulation is only superficially understood. SUMO binding is crucial for Wss1's function in DPC repair in yeast [42]. Moreover, DPCs become SUMOylated in human cells, which recruits the ACRC/GCNA protease *via* its SUMO-interacting motifs [57]. In contrast, human SPRTN bears a ubiquitin-binding zinc finger and is monoubiquitylated [45,47]. Monoubiquitylation regulates chromatin access of the enzyme. Ubiquitylated SPRTN is excluded from chromatin. The induction of DPCs triggers deubiquitylation and concurrent relocalization of SPRTN to chromatin [50]. The deubiquitinating enzyme and E3 ligase regulating this ubiquitin-switch remain unidentified.

Furthermore, SPRTN proteins bear a PIP-box for binding the replication clamp PCNA (proliferating cell nuclear antigen). However, the PIP-box is not required for SPRTN's function in replication-coupled DPC repair, but might be related to SPRTN's proposed role in regulating translesion synthesis [64,70]. Interaction with p97 is dispensable for DPC repair by SPRTN in *X.laevis* egg extracts, but is required for Wss1's role in DPC repair in yeast. Nevertheless, binding to the segregase p97/Cdc48 must play an important function because our analysis suggests that this feature was acquired independently in Wss1 and SPRTN proteases. Cdc48/p97 is a chaperone-like protein that has the ability to segregate proteins from their environment [78,79]. Although experimental insights are lacking, it is tempting to speculate that p97/Cdc48 assists DPC cleavage by partially unfolding substrates. Alternatively, the segregase activity may be important for removing SPRTN cleavage products from chromatin.

5. Outlook

The recent identification of additional proteases targeting DPCs was remarkable given the strong phenotypes caused by the loss of Wss1 or SPRTN. Very recent studies provide evidence that the unrelated yeast aspartate protease Ddi1 (DNA-damage inducible 1) acts as backup or alternative to Wss1-dependent DPC repair [80,81]. Moreover, proteasome inhibitors have previously been shown to impair DPC removal, but it remained unclear, whether that was a direct effect [82,83]. Recent experiments now demonstrated that the proteasome acts directly on DPCs [70]. Importantly, proteasomal DPC degradation requires ubiquitylation of the DPC by the E3 ligase TRAIP. In contrast, DPC cleavage by SPRTN is independent of DPC ubiquitylation. These recent developments raise interesting questions regarding DPC repair pathway choice. What determines whether a DPC is processed by Wss1/SPRTN, Ddi1 or the proteasome? These mechanisms are at least partially redundant as implied by synthetic sensitivities in yeast [80,81]. However, SPRTN is essential for viability in mammalian cells indicating that the amount of overlap between the different proteolytic systems is limited. The answer may lie within the diversity of DPCs. Certain protein adducts might be cleavable by only one DPC protease, whereas others might be degraded by several enzymes. In this context it will be paramount, to understand the *in vitro* properties of DPC proteases regarding substrate specificity and preference.

Finally, the abundance of repair mechanisms targeting DPCs demonstrate the scale of the threat posed by DNA-protein crosslinking. In the future, it will thus be important to reveal the origins of cellular DPC formation in order to improve our understanding of endogenous sources of genome instability. In this context, it is also worth noting that cells apparently induce DPCs voluntarily in certain contexts. HMCES (5-hydroxymethylcytosine (5hmC) binding, ESC-specific) auto-crosslinks to abasic sites within single-stranded DNA to ensure DNA integrity during replication [84–86]. Efforts should thus be made to understand whether such programmed DPCs are rather exceptions or perhaps a common theme in chromatin biology.

CRediT authorship contribution statement

Hannah K. Reinking: Conceptualization, Investigation, Writing - review & editing. **Kay Hofmann:** Investigation, Writing - review & editing. **Julian Stingele:** Conceptualization, Investigation, Writing - original draft, Writing - review & editing.

Declaration of Competing Interest

The authors declare that there are no conflicts of interest.

Acknowledgement

We thank P. Weickert for comments on the manuscript and A. Bandera for help with figure preparation. J.S. is supported by the European Research Council (ERC Starting Grant 801750 DNAProteinCrosslinks), by the Alfried Krupp Prize for Young University Teachers awarded by the Alfried-Krupp von Bohlen and Halbach-Stiftung, by the Center for Integrated Protein Sciences Munich (CIPSM) and the Deutsche Forschungsgemeinschaft (CRC1064).

References

[1] T. Lindahl, Instability and decay of the primary structure of DNA, Nature 362 (1993) 709–715.
[2] S.P. Jackson, J. Bartek, The DNA-damage response in human biology and disease, Nature 461 (2009) 1071–1078.
[3] S. Ji, D. Park, K. Kropachev, M. Kolbanovskiy, I. Fu, S. Broyde, M. Essawy, N.E. Geacintov, N.Y. Tretyakova, 5-Formylcytosine-induced DNA–peptide cross-links reduce transcription efficiency, but do not do cause transcription errors in human cells, J. Biol. Chem. (2019).
[4] T. Nakano, R. Ouchi, J. Kawazoe, S.P. Pack, K. Makino, H. Ide, T7 RNA polymerases backed up by covalently trapped proteins catalyze highly error prone transcription, J. Biol. Chem. 287 (2012) 6562–6572.
[5] Y.V. Fu, H. Yardimci, D.T. Long, T.V. Ho, A. Guainazzi, V.P. Bermudez, J. Hurwitz, A. van Oijen, O.D. Scharer, J.C. Walter, Selective bypass of a lagging strand roadblock by the eukaryotic replicative DNA helicase, Cell 146 (2011) 931–941.
[6] T. Nakano, M. Miyamoto-Matsubara, M.I. Shoulkamy, A.M. Salem, S.P. Pack, Y. Ishimi, H. Ide, Translocation and stability of replicative DNA helicases upon encountering DNA-protein cross-links, J. Biol. Chem. 288 (2013) 4649–4658.
[7] H.K. Kuo, J.D. Griffith, K.N. Kreuzer, 5-Azacytidine induced methyltransferase-DNA adducts block DNA replication in vivo, Cancer Res. 67 (2007) 8248–8254.
[8] M.S. van Ruiten, B.D. Rowland, SMC Complexes: Universal DNA Looping Machines with Distinct Regulators, Trends Genet. 34 (2018) 477–487.
[9] C.R. Clapier, J. Iwasa, B.R. Cairns, C.L. Peterson, Mechanisms of action and

regulation of ATP-dependent chromatin-remodelling complexes, Nat. Rev. Mol. Cell Biol. 18 (2017) 407–422.
[10] S. Barker, M. Weinfeld, D. Murray, DNA-protein crosslinks: their induction, repair, and biological consequences, Mutat. Res. 589 (2005) 111–135.
[11] Y. Shi, F. Lan, C. Matson, P. Mulligan, J.R. Whetstine, P.A. Cole, R.A. Casero, Y. Shi, Histone demethylation mediated by the nuclear amine oxidase homolog LSD1, Cell 119 (2004) 941–953.
[12] Y. Tsukada, J. Fang, H. Erdjument-Bromage, M.E. Warren, C.H. Borchers, P. Tempst, Y. Zhang, Histone demethylation by a family of JmjC domain-containing proteins, Nature 439 (2006) 811–816.
[13] S. Ji, I. Fu, S. Naldiga, H. Shao, A.K. Basu, S. Broyde, N.Y. Tretyakova, 5-Formylcytosine mediated DNA-protein cross-links block DNA replication and induce mutations in human cells, Nucleic Acids Res. 46 (2018) 6455–6469.
[14] F. Li, Y. Zhang, J. Bai, M.M. Greenberg, Z. Xi, C. Zhou, 5-formylcytosine yields DNA-Protein cross-links in nucleosome core particles, J. Am. Chem. Soc. 139 (2017) 10617–10620.
[15] S. Ji, H. Shao, Q. Han, C.L. Seiler, N.Y. Tretyakova, Reversible DNA-Protein cross-linking at epigenetic DNA marks, Angew. Chem. Int. Ed. Engl. 56 (2017) 14130–14134.
[16] R.M. Kohli, Y. Zhang, TET enzymes, TDG and the dynamics of DNA demethylation, Nature 502 (2013) 472–479.
[17] J.T. Sczepanski, R.S. Wong, J.N. McKnight, G.D. Bowman, M.M. Greenberg, Rapid DNA–protein cross-linking and strand scission by an abasic site in a nucleosome core particle, Proc. Natl. Acad. Sci. 107 (2010) 22475–22480.
[18] W. Chan, Y.H. Ham, L. Jin, H.W. Chan, Y.L. Wong, C.K. Chan, P.Y. Chung, 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 (2019) 4987–4994.
[19] H.E. Krokan, M. Bjørås, Base excision repair, Cold Spring Harb. Perspect. Biol. 5 (2013) a012583, , https://doi.org/10.1101/cshperspect.a012583.
[20] J.L. Quiñones, U. Thapar, K. Yu, Q. Fang, R.W. Sobol, B. Demple, Enzyme mechanism-based, oxidative DNA–protein cross-links formed with DNA polymerase β in vivo, Proc. Natl. Acad. Sci. 112 (2015) 8602–8607.
[21] J.L. Quinones, B. Demple, When DNA repair goes wrong: BER-generated DNA-protein crosslinks to oxidative lesions, DNA repair 44 (2016) 103–109.
[22] Y. Pommier, S.-y.N. Huang, R. Gao, B.B. Das, J. Murai, C. Marchand, Tyrosyl-DNA-phosphodiesterases (TDP1 and TDP2), DNA Repair 19 (2014) 114–129.
[23] P. Pourquier, L.M. Ueng, G. Kohlhagen, A. Mazumder, M. Gupta, K.W. Kohn, Y. Pommier, Effects of uracil incorporation, DNA mismatches, and abasic sites on cleavage and religation activities of mammalian topoisomerase I, J. Biol. Chem. 272 (1997) 7792–7796.
[24] Y. Pommier, C. Marchand, Interfacial inhibitors: targeting macromolecular complexes, Nature reviews, Drug discovery 11 (2011) 25–36.
[25] J.L. Nitiss, Targeting DNA topoisomerase II in cancer chemotherapy, Nat. Rev. Cancer 9 (2009) 338–350.
[26] J. Stingele, R. Bellelli, S.J. Boulton, Mechanisms of DNA-protein crosslink repair, Nat. Rev. Mol. Cell Biol. 18 (2017) 563–573.
[27] K. Chvalova, V. Brabec, J. Kasparkova, Mechanism of the formation of DNA-protein cross-links by antitumor cisplatin, Nucleic Acids Res. 35 (2007) 1812–1821.
[28] A.Y. Maslov, M. Lee, M. Gundry, S. Gravina, N. Strogonova, C. Tazearslan, A. Bendebury, Y. Suh, J. Vijg, 5-aza-2'-deoxycytidine-induced genome rearrangements are mediated by DNMT1, Oncogene 31 (2012) 5172–5179.
[29] S. Barker, M. Weinfeld, J. Zheng, L. Li, D. Murray, Identification of mammalian proteins cross-linked to DNA by ionizing radiation, J. Biol. Chem. 280 (2005) 33826–33838.
[30] J. Stingele, S. Jentsch, DNA-protein crosslink repair, Nat. Rev. Mol. Cell Biol. 16 (2015) 455–460.
[31] F. Cortes Ledesma, S.F. El Khamisy, M.C. Zuma, K. Osborn, K.W. Caldecott, A human 5'-tyrosyl DNA phosphodiesterase that repairs topoisomerase-mediated DNA damage, Nature 461 (2009) 674–678.
[32] J.J. Pouliot, K.C. Yao, C.A. Robertson, H.A. Nash, Yeast gene for a Tyr-DNA phosphodiesterase that repairs topoisomerase I complexes, Science 286 (1999) 552–555.
[33] S.W. Yang, A.B. Burgin Jr., B.N. Huizenga, C.A. Robertson, K.C. Yao, H.A. Nash, A eukaryotic enzyme that can disjoin dead-end covalent complexes between DNA and type I topoisomerases, Proc. Natl. Acad. Sci. U.S.A. 93 (1996) 11534–11539.
[34] B. de Graaf, A. Clore, A.K. McCullough, Cellular pathways for DNA repair and damage tolerance of formaldehyde-induced DNA-protein crosslinks, DNA repair 8 (2009) 1207–1214.
[35] T. Nakano, S. Morishita, A. Katafuchi, M. Matsubara, Y. Horikawa, H. Terato, A.M. Salem, S. Izumi, S.P. Pack, K. Makino, H. Ide, Nucleotide excision repair and homologous recombination systems commit differentially to the repair of DNA-protein crosslinks, Mol. Cell 28 (2007) 147–158.
[36] T. Nakano, A. Katafuchi, M. Matsubara, H. Terato, T. Tsuboi, T. Masuda, T. Tatsumoto, S.P. Pack, K. Makino, D.L. Croteau, B. Van Houten, K. Iijima, H. Tauchi, H. Ide, 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 (2009) 27065–27076.
[37] L.N. Chesner, C. Campbell, 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 62 (2018) 18–27.
[38] M.J. Neale, J. Pan, S. Keeney, Endonucleolytic processing of covalent protein-linked DNA double-strand breaks, Nature 436 (2005) 1053–1057.
[39] E. Cannavo, P. Cejka, Sae2 promotes dsDNA endonuclease activity within Mre11-Rad50-Xrs2 to resect DNA breaks, Nature 514 (2014) 122–125.
[40] R.A. Deshpande, J.H. Lee, S. Arora, T.T. Paull, Nbs1 converts the human Mre11/Rad50 nuclease complex into an Endo/Exonuclease machine specific for Protein-DNA adducts, Mol. Cell 64 (2016) 593–606.
[41] S. Biggins, N. Bhalla, A. Chang, D.L. Smith, A.W. Murray, Genes involved in sister chromatid separation and segregation in the budding yeast Saccharomyces cerevisiae, Genetics 159 (2001) 453–470.
[42] J. Stingele, M.S. Schwarz, N. Bloemeke, P.G. Wolf, S. Jentsch, A DNA-Dependent Protease Involved in DNA-Protein Crosslink Repair, Cell 158 (2014) 327–338.
[43] J.P. Duxin, J.M. Dewar, H. Yardimci, J.C. Walter, Repair of a DNA-Protein crosslink by replication-coupled proteolysis, Cell 159 (2014) 346–357.
[44] J. Stingele, B. Habermann, S. Jentsch, DNA-protein crosslink repair: proteases as DNA repair enzymes, Trends Biochem. Sci. 40 (2015) 67–71.
[45] A. Mosbech, I. Gibbs-Seymour, K. Kagias, T. Thorslund, P. Beli, L. Povlsen, S.V. Nielsen, S. Smedegaard, G. Sedgwick, C. Lukas, R. Hartmann-Petersen, J. Lukas, C. Choudhary, R. Pocock, S. Bekker-Jensen, N. Mailand, DVC1 (C1orf124) is a DNA damage-targeting p97 adaptor that promotes ubiquitin-dependent responses to replication blocks, Nat. Struct. Mol. Biol. 19 (2012) 1084–1092.
[46] R.C. Centore, S.A. Yazinski, A. Tse, L. Zou, Spartan/C1orf124, a reader of PCNA ubiquitylation and a regulator of UV-induced DNA damage response, Mol. Cell 46 (2012) 625–635.
[47] E.J. Davis, C. Lachaud, P. Appleton, T.J. Macartney, I. Nathke, J. Rouse, DVC1 (C1orf124) recruits the p97 protein segregase to sites of DNA damage, Nat. Struct. Mol. Biol. 19 (2012) 1093–1100.
[48] Y. Machida, M.S. Kim, Y.J. Machida, Spartan/C1orf124 is important to prevent UV-induced mutagenesis, Cell cycle (Georgetown, Tex.) 11 (2012) 3395–3402.
[49] S. Juhasz, D. Balogh, I. Hajdu, P. Burkovics, M.A. Villamil, Z. Zhuang, L. Haracska, Characterization of human Spartan/C1orf124, an ubiquitin-PCNA interacting regulator of DNA damage tolerance, Nucleic Acids Res. 40 (2012) 10795–10808.
[50] J. Stingele, R. Bellelli, F. Alte, G. Hewitt, G. Sarek, S.L. Maslen, S.E. Tsutakawa, A. Borg, S. Kjaer, J.A. Tainer, J.M. Skehel, M. Groll, S.J. Boulton, Mechanism and regulation of DNA-Protein crosslink repair by the DNA-Dependent metalloprotease SPRTN, Mol. Cell 64 (2016) 688–703.
[51] B. Vaz, M. Popovic, J.A. Newman, J. Fielden, H. Aitkenhead, S. Halder, A.N. Singh, I. Vendrell, R. Fischer, I. Torrecilla, N. Drobnitzky, R. Freire, D.J. Amor, P.J. Lockhart, B.M. Kessler, G.W. McKenna, O. Gileadi, K. Ramadan, Metalloprotease SPRTN/DVC1 orchestrates replication-coupled DNA-Protein crosslink repair, Mol. Cell 64 (2016) 704–719.
[52] R.S. Maskey, K.S. Flatten, C.J. Sieben, K.L. Peterson, D.J. Baker, H.-J. Nam, M.S. Kim, T.C. Smyrk, Y. Kojima, Y. Machida, A. Santiago, J.M. van Deursen, S.H. Kaufmann, Y.J. Machida, Spartan deficiency causes accumulation of Topoisomerase 1 cleavage complexes and tumorigenesis, Nucleic Acids Res. 45 (2017) 4564–4576.
[53] M. Mórocz, E. Zsigmond, R. Tóth, M.Z. Enyedi, L. Pintér, L. Haracska, DNA-dependent protease activity of human Spartan facilitates replication of DNA-protein crosslink-containing DNA, Nucleic Acids Res. 45 (2017) 3172–3188.
[54] J. Lopez-Mosqueda, K. Maddi, S. Prgomet, S. Kalayil, I. Marinovic-Terzic, J. Terzic, I. Dikic, SPRTN is a mammalian DNA-binding metalloprotease that resolves DNA-protein crosslinks, eLife 5 (2016) e21491.
[55] G.A. Dokshin, G.M. Davis, A.D. Sawle, M.D. Eldridge, P.K. Nicholls, T.E. Gourley, K.A. Romer, L.W. Molesworth, H.R. Tatnell, A.R. Ozturk, D.G. de Rooij, G.J. Hannon, D.C. Page, C.C. Mello, M.A. Carmell, GCNA interacts with spartan and topoisomerase II to regulate genome stability, Dev. Cell 52 (2020) 53-68.e56.
[56] V. Bhargava, C.D. Goldstein, L. Russell, L. Xu, M. Ahmed, W. Li, A. Casey, K. Servage, R. Kollipara, Z. Picciarelli, R. Kittler, A. Yatsenko, M. Carmell, K. Orth, J.F. Amatruda, J.L. Yanowitz, M. Buszczak, GCNA preserves genome integrity and fertility across species, Dev. Cell 52 (2020) 38–52.e10.
[57] N. Borgermann, L. Ackermann, P. Schwertman, I.A. Hendriks, K. Thijssen, J.C. Liu, H. Lans, M.L. Nielsen, N. Mailand, SUMOylation promotes protective responses to DNA-protein crosslinks, EMBO J. 38 (2019) e101496.
[58] M.A. Carmell, G.A. Dokshin, H. Skaletsky, Y.-C. Hu, J.C. van Wolfswinkel, K.J. Igarashi, D.W. Bellott, M. Nefedov, P.W. Reddien, G.C. Enders, V.N. Uversky, C.C. Mello, D.C. Page, A widely employed germ cell marker is an ancient disordered protein with reproductive functions in diverse eukaryotes, eLife 5 (2016) e19993.
[59] P. Kanellis, M. Gagliardi, J.P. Banath, R.K. Szilard, S. Nakada, S. Galicia, F.D. Sweeney, D.C. Cabelof, P.L. Olive, D. Durocher, A Screen for Suppressors of Gross Chromosomal Rearrangements Identifies a Conserved Role for PLP in Preventing DNA Lesions, PLoS Genet. 3 (2007) e134.
[60] J. Enderle, A. Dorn, N. Beying, O. Trapp, H. Puchta, The protease WSS1A, the endonuclease MUS81, and the phosphodiesterase TDP1 are involved in independent pathways of DNA-protein crosslink repair in plants, Plant Cell 31 (2019) 775–790.
[61] J. Enderle, A. Dorn, H. Puchta, DNA- and DNA-Protein-Crosslink repair in plants, Int. J. Mol. Sci. 20 (2019) 4304.
[62] B. Loppin, F. Berger, P. Couble, Paternal chromosome incorporation into the zygote nucleus is controlled by maternal haploid in Drosophila, Dev. Biol. 231 (2001) 383–396.
[63] L. Delabaere, G.A. Orsi, L. Sapey-Triomphe, B. Horard, P. Couble, B. Loppin, The Spartan ortholog maternal haploid is required for paternal chromosome integrity in the Drosophila zygote, Curr. Biol. 24 (2014) 2281–2287.
[64] R.S. Maskey, M.S. Kim, D.J. Baker, B. Childs, L.A. Malureanu, K.B. Jeganathan, Y. Machida, J.M. van Deursen, Y.J. Machida, Spartan deficiency causes genomic instability and progeroid phenotypes, Nat. Commun. 5 (2014) 5744.
[65] T. Hart, M. Chandrashekhar, M. Aregger, Z. Steinhart, K.R. Brown, G. MacLeod, M. Mis, M. Zimmermann, A. Fradet-Turcotte, S. Sun, P. Mero, P. Dirks, S. Sidhu, F.P. Roth, O.S. Rissland, D. Durocher, S. Angers, J. Moffat, High-resolution CRISPR screens reveal fitness genes and genotype-specific Cancer liabilities, Cell 163 (2015) 1515–1526.
[66] M.W.G. Ruijs, R.N.J. van Andel, J. Oshima, K. Madan, A.W.M. Nieuwint, C.M. Aalfs,

Atypical progeroid syndrome: An unknown helicase gene defect? Am. J. Med. Genet. A 116A (2003) 295–299.
[67] D. Lessel, B. Vaz, S. Halder, P.J. Lockhart, I. Marinovic-Terzic, J. Lopez-Mosqueda, M. Philipp, J.C. Sim, K.R. Smith, J. Oehler, E. Cabrera, R. Freire, K. Pope, A. Nahid, F. Norris, R.J. Leventer, M.B. Delatycki, G. Barbi, S. von Ameln, J. Hogel, M. Degoricija, R. Fertig, M.D. Burkhalter, K. Hofmann, H. Thiele, J. Altmuller, G. Nurnberg, P. Nurnberg, M. Bahlo, G.M. Martin, C.M. Aalfs, J. Oshima, J. Terzic, D.J. Amor, I. Dikic, K. Ramadan, C. Kubisch, Mutations in SPRTN cause early onset hepatocellular carcinoma, genomic instability and progeroid features, Nat. Genet. 46 (2014) 1239–1244.
[68] G.C. Enders, J.J. May, 2nd, developmentally regulated expression of a mouse germ cell nuclear antigen examined from embryonic day 11 to adult in male and female mice, Dev. Biol. 163 (1994) 331–340.
[69] J.L. Sparks, G. Chistol, A.O. Gao, M. Raschle, N.B. Larsen, M. Mann, J.P. Duxin, J.C. Walter, The CMG helicase bypasses DNA-Protein cross-links to facilitate their repair, Cell 176 (2019) 167-181.e121.
[70] N.B. Larsen, A.O. Gao, J.L. Sparks, I. Gallina, R.A. Wu, M. Mann, M. Raschle, J.C. Walter, J.P. Duxin, Replication-coupled DNA-Protein crosslink repair by SPRTN and the proteasome in Xenopus egg extracts, Mol. Cell 73 (2019) 574–588.e577.
[71] X. Yang, Y. Li, Z. Gao, Z. Li, J. Xu, W. Wang, Y. Dong, Structural analysis of Wss1 protein from saccharomyces cerevisiae, Sci. Rep. 7 (2017) 8270.
[72] F. Li, J.E. Raczynska, Z. Chen, H. Yu, Structural insight into DNA-Dependent activation of human metalloprotease spartan, Cell Rep. 26 (2019) 3336-3346.e3334.
[73] M.Y. Balakirev, J.E. Mullally, A. Favier, N. Assard, E. Sulpice, D.F. Lindsey, A.V. Rulina, X. Gidrol, K.D. Wilkinson, Wss1 metalloprotease partners with Cdc48/Doa1 in processing genotoxic SUMO conjugates, eLife 4 (2015) e06763.
[74] H.E. Van Wart, H. Birkedal-Hansen, The cysteine switch: a principle of regulation of metalloproteinase activity with potential applicability to the entire matrix metalloproteinase gene family, Proc. Natl. Acad. Sci. U.S.A. 87 (1990) 5578–5582.
[75] B. Vaz, M. Popovic, K. Ramadan, DNA-protein crosslink proteolysis repair, Trends Biochem. Sci. 42 (2017) 483–495.
[76] Y. Peng, L. Chen, Y. Lu, Y. Wu, J. Dumenil, Z. Zhu, M.W. Bevan, Y. Li, The ubiquitin receptors DA1, DAR1, and DAR2 redundantly regulate endoreduplication by modulating the stability of TCP14/15 in Arabidopsis, Plant Cell 27 (2015) 649–662.
[77] D. Zhang, R.F. de Souza, V. Anantharaman, L.M. Iyer, L. Aravind, Polymorphic toxin systems: comprehensive characterization of trafficking modes, processing, mechanisms of action, immunity and ecology using comparative genomics, Biol. Direct 7 (2012) 18.
[78] S. Braun, K. Matuschewski, M. Rape, S. Thoms, S. Jentsch, Role of the ubiquitin-selective CDC48(UFD1/NPL4)chaperone (segregase) in ERAD of OLE1 and other substrates, EMBO J. 21 (2002) 615–621.
[79] S. Jentsch, S. Rumpf, Cdc48 (p97): a "molecular gearbox" in the ubiquitin pathway? Trends Biochem. Sci. 32 (2007) 6–11.
[80] M. Svoboda, J. Konvalinka, J.F. Trempe, K. Grantz Saskova, The yeast proteases Ddi1 and Wss1 are both involved in the DNA replication stress response, DNA repair 80 (2019) 45–51.
[81] N. Serbyn, A. Noireterre, I. Bagdiul, M. Plank, A.H. Michel, R. Loewith, B. Kornmann, F. Stutz, The aspartic protease Ddi1 contributes to DNA-Protein crosslink repair in yeast, Mol. Cell (2019) S1097–2765 (1019)30898-30896.
[82] D.J. Baker, G. Wuenschell, L. Xia, J. Termini, S.E. Bates, A.D. Riggs, T.R. O'Connor, Nucleotide excision repair eliminates unique DNA-protein cross-links from mammalian cells, J. Biol. Chem. 282 (2007) 22592–22604.
[83] G. Quievryn, A. Zhitkovich, Loss of DNA-protein crosslinks from formaldehyde-exposed cells occurs through spontaneous hydrolysis and an active repair process linked to proteosome function, Carcinogenesis 21 (2000) 1573–1580.
[84] L. Halabelian, M. Ravichandran, Y. Li, H. Zeng, A. Rao, L. Aravind, C.H. Arrowsmith, Structural basis of HMCES interactions with abasic DNA and multivalent substrate recognition, Nat. Struct. Mol. Biol. 26 (2019) 607–612.
[85] K.N. Mohni, S.R. Wessel, R. Zhao, A.C. Wojciechowski, J.W. Luzwick, H. Layden, B.F. Eichman, P.S. Thompson, K.P.M. Mehta, D. Cortez, HMCES maintains genome integrity by shielding abasic sites in single-strand DNA, Cell 176 (2019) 144–153.e113.
[86] P.S. Thompson, K.M. Amidon, K.N. Mohni, D. Cortez, B.F. Eichman, Protection of abasic sites during DNA replication by a stable thiazolidine protein-DNA cross-link, Nat. Struct. Mol. Biol. 26 (2019) 613–618.

2.4 Protein-oligonucleotide conjugates as model substrates for DNA-protein crosslink repair proteases

Hannah K. Reinking, Julian Stingele, Protein-oligonucleotide conjugates as model substrates for DNA-protein crosslink repair proteases, *STAR Protocols,* Volume 2, Issue 2, 18 June 2021, Pages 100591, https://doi.org/10.1016/j.xpro.2021.100591 (Experimental protocol)

Summary

This protocol paper describes the experimental steps for the generation of protein-oligonucleotide conjugates as *in vitro* substrates for DPC proteases. The protocol employs a detailed description of how these substrates are generated and further used in *in vitro* cleavage and binding assays: In a first step protein G is conjugated to diverse 30mer oligonucleotides over an amine coupling reaction. The resulting conjugates are further isolated from non-conjugated DNA and protein using ion exchange chromatography. In order to assess DPC cleavage by SPRTN, the conjugates are incubated with SPRTN and the resulting products are resolved on native gels. Binding of SPRTN to the different conjugates is analysed using electrophoretic mobility shift assays (EMSAs). The conjugates can be used for the *in vitro* characterization of diverse DPC proteases.

Author contribution

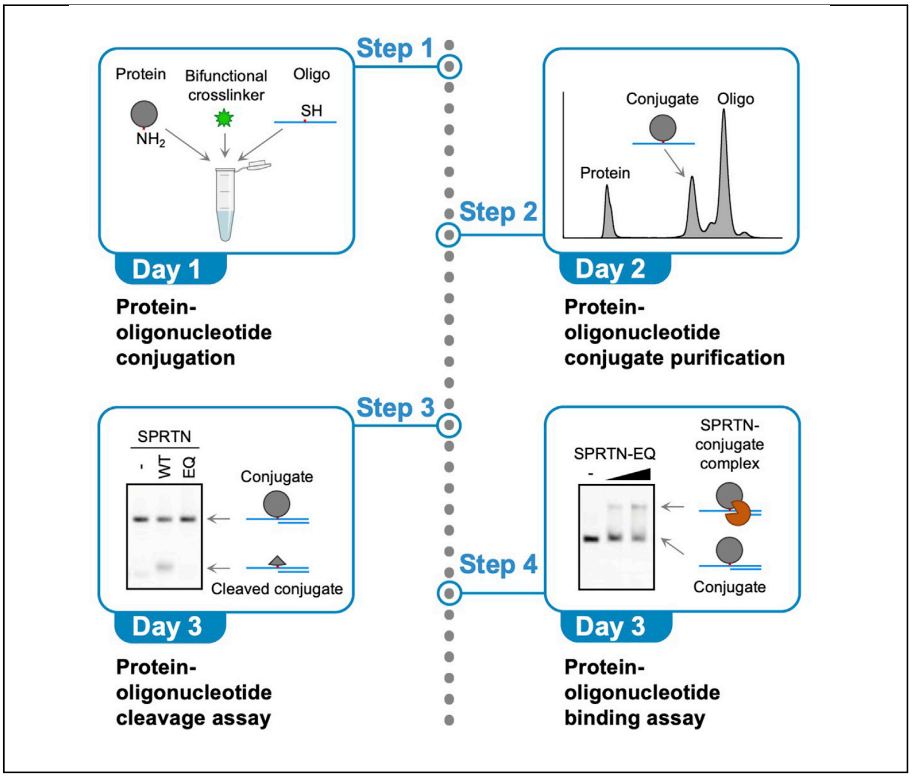
I established the experimental setup for the generation of protein-oligonucleotide conjugates used for SPRTN´s *in vitro* characterization. I performed all described experiments. In addition, I discussed the results with Julian Stingele and wrote the manuscript.

STAR Protocols

CellPress
OPEN ACCESS

Protocol

Protein-oligonucleotide conjugates as model substrates for DNA-protein crosslink repair proteases



Covalent DNA-protein crosslinks (DPCs) have emerged as pervasive sources of genome instability. DPCs are targeted for repair by DNA-dependent proteases of the Wss1/SPRTN family. However, understanding how these enzymes achieve specificity has been hampered by the lack of suitable *in vitro* model substrates. Here, we describe the generation of defined protein-oligonucleotide conjugates as DPC model substrates, which enable the analysis of DPC proteases in activity and binding assays.

Hannah K. Reinking, Julian Stingele

reinking@genzentrum.lmu.de (H.K.R.)
stingele@genzentrum.lmu.de (J.S.)

Highlights

Protocol for the generation of defined protein-oligonucleotide conjugates

Conjugates as model substrates for enzymes processing DNA-protein crosslinks

Conjugates can be used to determine specificity of the SPRTN protease

Reinking & Stingele, STAR Protocols 2, 100591
June 18, 2021
https://doi.org/10.1016/j.xpro.2021.100591

STAR Protocols

CellPress
OPEN ACCESS

Protocol

Protein-oligonucleotide conjugates as model substrates for DNA-protein crosslink repair proteases

Hannah K. Reinking[1,2,3,*] and Julian Stingele[1,2,4,*]

[1]Department of Biochemistry, Ludwig Maximilians University, 81377 Munich, Germany

[2]Gene Center, Ludwig Maximilians University, 81377 Munich, Germany

[3]Technical contact

[4]Lead contact

*Correspondence: reinking@genzentrum.lmu.de (H.K.R.), stingele@genzentrum.lmu.de (J.S.)
https://doi.org/10.1016/j.xpro.2021.100591

SUMMARY

Covalent DNA-protein crosslinks (DPCs) have emerged as pervasive sources of genome instability. DPCs are targeted for repair by DNA-dependent proteases of the Wss1/SPRTN family. However, understanding how these enzymes achieve specificity has been hampered by the lack of suitable *in vitro* model substrates. Here, we describe the generation of defined protein-oligonucleotide conjugates as DPC model substrates, which enable the analysis of DPC proteases in activity and binding assays.

For complete details on the use and execution of this protocol, please refer to Reinking et al. (2020).

BEFORE YOU BEGIN

Preparation of buffers

Timing: 2 h

1. Prepare stock solutions and buffers as described in materials and equipment.

Note: The reducing agent TCEP and BSA should be added freshly on the day of the experiment.

Preparation of proteins

Timing: 2 h

Note: Protein and oligonucleotide are conjugated with the proFIRE Amine Coupling Kit prior to purification of the conjugate using the proFIRE chromatography system. Conjugation is based on amine coupling and is suitable for proteins with a molecular weight larger than 5 kDa. Our protocol is optimized for conjugation of protein G (30 kDa) to a single-stranded 30-mer oligonucleotide. Conjugation of other proteins may require additional optimization steps (troubleshooting 1).

2. Prepare protein stock solutions to be used for conjugation
   a. Resuspend lyophilized proteins in conjugation buffer provided in the proFIRE Amine Coupling Kit to a concentration of 5 mg/mL. Non-lyophilized proteins are concentrated to 15 mg/mL before dilution to 5 mg/mL with conjugation buffer. Alternatively, a buffer exchange using

CellPress
OPEN ACCESS

STAR Protocols
Protocol

centrifugal filters can be performed. However, this may cause protein loss through interaction with the filter membrane.

b. Distribute protein stock solution into 50 μL aliquots in 1.5 mL tubes. Each conjugation reaction requires one aliquot (250 μg protein).

c. Snap freeze aliquots in liquid nitrogen and store at −80°C.

3. Purify DPC proteases

a. Detailed protocols for recombinant expression and purification of the DPC protease SPRTN have been published (Reinking et al., 2020; Stingele et al., 2016; Vaz et al., 2016). The assays described here can also be performed to investigate other proteases implicated in DPC degradation such as Wss1, GCNA, FAM111A or Ddi1 (Bhargava et al., 2020; Dokshin et al., 2020; Kojima et al., 2020; Serbyn et al., 2020; Stingele et al., 2014). Purified proteases should be stored in small aliquots at −80°C.

⚠ CRITICAL: Proteins that are used for DNA conjugation have to be free of protein contaminations. If sample purity is low, the possibility of conjugating the contaminating proteins and, thus, obtaining heterogenous protein-oligonucleotide conjugate mixtures is high.

Ordering modified DNA oligonucleotides for protein conjugation

Timing: 30 min

Note: To conjugate a protein of choice, the oligonucleotide has to contain a modified thymine base at the intended crosslinking position that carries a reduced thiol on a short linker (thio-dT, Figure 1). During conjugation, the reduced thiol reacts with a bifunctional crosslinker provided in the proFIRE Amine Coupling Kit to generate an NHS ester. This NHS ester reacts with a primary amino group of the protein forming a stable bond between the oligonucleotide and the protein. Oligonucleotides between 15 and 150 nucleotides can be used for conjugation. However, according to our experience, 30-mer oligonucleotides produce the best results with respect to yield and purity. Shorter oligonucleotides complicate the separation of conjugated and unconjugated DNA, while conjugations using longer oligonucleotides result in poor yields.

Note: In order to allow the monitoring of conjugate cleavage in gel-based assays, the oligonucleotide has to be modified with a fluorescent label. The label can be attached either to the 3′or 5′end of the oligonucleotide. We have successfully used 5′-Cy5 and 5′-6-FAM labels.

Note: Oligonucleotides with a thio-dT as the 3′-terminal base carry a 3′-phosphate as a remnant of the synthesis strategy. Thus, we purchase all oligonucleotides with a 3′-phosphate to allow the direct comparison of different crosslinking positions.

Figure 1. Modified nucleotide (thio-dT) used for protein conjugation

The nucleotide incorporated at the intended crosslinking position contains a reduced thiol (red highlight) on a linker. During the crosslinking reaction, the thiol is further modified to an NHS ester, which then crosslinks to primary amine groups of proteins.

STAR Protocols
Protocol

CellPress
OPEN ACCESS

Note: Example on how to order an oligonucleotide intended for conjugation of a protein to an internal position:

5′-Cy5-CCCAAAAAAAAAAA-thio-dT-AAAAAAAAAAAACCC-3′-phos

4. Each crosslinking reaction requires 3 nmol of high-quality oligonucleotide (troubleshooting 2). Order HPLC-purified and lyophilized DNA in aliquots of 3 nmol. Order at least three aliquots from a company, which allows ordering of exact amounts per aliquot (e.g., ELLA Biotech GmbH): one for calibrating the anion exchange column; one for protein conjugation and one as backup.
5. Store modified lyophilized DNA at −20°C. Always make sure to protect fluorescently labeled DNA from light.

Alternatives: For the conjugation strategy described here it is important that the oligonucleotide contains an accessible reduced thiol group. However, it is possible to generate protein-oligonucleotide conjugates using other crosslinking strategies. A conjugation kit available from Abcam crosslinks proteins to a 10 to 120-mer single-stranded DNA (ssDNA) or up to 80 bases of double-stranded DNA (dsDNA). However, compared to our strategy, conjugation is only possible at the 3′- and 5′-termini of the oligonucleotide (Oligonucleotide Conjugation Kit, ab218260).

Ordering unmodified DNA oligonucleotides

Timing: 30 min

Note: In order to generate model substrates with different DNA structures, the conjugate has to be annealed to reverse oligonucleotides (Table 1). Those oligonucleotides do not require any modifications and can be ordered from various companies (e.g., Eurofins Genomics, Sigma-Aldrich, or Integrated DNA Technologies).

Note: An oligonucleotide with the same sequence as the modified DNA used for protein conjugation should be ordered as well (forward control oligo, Table 1). This oligonucleotide needs to contain a fluorescent label (Cy5 or 6-FAM) and will be used as a control in protein-oligonucleotide cleavage assays.

Table 1. Sequences of reverse oligonucleotides needed to generate model substrates with different DNA structures and of a non-modified forward oligonucleotide

Name	Resulting DNA structure	Sequence
Reverse oligo 1	Double-stranded/single-stranded DNA junction	5′- GGGTTTTTTTTTTT - 3′
Reverse oligo 2	Double-stranded DNA	5′- GGGTTTTTTTTTTTATTTTTTTTTTGGG - 3′
Forward control oligo	-	5′-Cy5-CCCAAAAAAAAAAATAAAAAAAAAAACCC-3′

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
Protein G	BioVision	Cat#6510
UltraPure™ BSA	Thermo Scientific	Cat#AM2616

(*Continued on next page*)

CellPress
OPEN ACCESS

STAR Protocols
Protocol

Continued		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical commercial assays		
proFIRE Amine Coupling Kit (containing dilution and conjugation buffer)	Dynamic Biosensors	Cat#PF-NH2-1
proFIRE buffer A	Dynamic Biosensors	PF-BU-A-10
proFIRE buffer B	Dynamic Biosensors	PF-BU-B-5
Oligonucleotides		
Internal cross-link 5′-Cy5-CCCAAAAAAAAAAA-thio-dT-AAAAAAAAAAAACCC-3′-phos	Ella Biotech	n/a
Reverse oligo 1 5′- GGGTTTTTTTTTTT - 3′	Sigma	n/a
Reverse oligo 2 5′- GGGTTTTTTTTTTTA TTTTTTTTTTGGG - 3′	Sigma	n/a
Forward control oligo 5′-Cy5-CCCAAAAAAAAAAA TAAAAAAAAAAACCC-3′	Sigma	n/a
Software and algorithms		
ImageJ	NIH	https://imagej.net/Fiji/Downloads
Other		
proFIRE	Dynamic Biosensors	https://www.dynamic-biosensors.com/profire/
proFIRE anion exchange column	Dynamic Biosensors	TB-CC-1-1
Chemidoc XRS+ System	Bio-Rad	1708265
SpectraMax Paradigm Multi-Mode Detection Platform	Molecular Devices	n/a
C1000 Touch Thermal Cycler	Bio-Rad	Cat#1851148
Empty gel cassettes	Thermo Fisher	Cat#NC2010
Mini Gel Tank	Thermo Fisher	Cat#A25977
PCR tubes	Biozym	710970X

MATERIALS AND EQUIPMENT

Stock solutions			
Reagent	Concentration	Amount	Storage
Potassium chloride (KCl)	2.5 M	1 L	20°C, up to 1 year
Potassium chloride (KCl)	0.5 M	10 mL	20°C, up to 1 year
4-(2-Hydroxyethyl)-1-piperazine ethanesulfonic acid (HEPES)/KOH pH 7.2	1 M in H_2O	1 L	4°C, up to 1 year
Tris-(2-carboxyethyl)-phosphine hydrochloride (TCEP) (hazardous)	0.5 M in H_2O	10 mL	−20°C, up to 6 months
TCEP	0.05 M in H_2O	1 mL	−20°C, up to 6 months
10 × TBE	5.5% Boric acid (hazardous), 0.9 M Tris(hydroxymethyl) aminomethane, 0.025 M EDTA (hazardous)	1 L	20°C, up to 1 year
Glycerol	100%	n/a	20°C, up to 2 years
Ammonium persulfate (APS) (hazardous)	10%	10 mL	−20°C, up to 1 year
6 × Orange G loading dye	15% Ficoll type 400, 0.125% Orange G in H_2O	10 mL	−20°C, up to 1 year

STAR Protocols
Protocol

CellPress
OPEN ACCESS

TBE running buffers				
Reagent	Stock concentration	Final concentration	Add in 1 L	Storage
TBE	10 ×	1 ×	100 mL	4°C, up to 6 months
TBE	10 ×	0.5 ×	50 mL	4°C, up to 6 months

Conjugate desalting and storage buffer			
Reagent	Stock concentration	Final concentration	Add in 100 mL
KCl	2.5 M	100 mM	4 mL
HEPES/KOH pH 7.2	1 M	50 mM	5 mL
Glycerol	100%	10%	10 mL
H_2O	-	-	81 mL

Store buffer at 4°C for up to 6 months. After conjugate desalting and measuring DNA concentration, add BSA to a final concentration of 0.4 mg/mL for storage (conjugate storage buffer). The conjugate storage buffer used in cleavage and binding assays has to be prepared fresh on the day of the experiment and can only be used on the same day.

High-salt SPRTN buffer			
Reagent	Stock concentration	Final concentration	Add in 10 mL
KCl	2.5 M	500 mM	2 mL
HEPES/KOH pH 7.2	1 M	50 mM	0.5 mL
Glycerol	100%	10%	1 mL
TCEP	0.5 M	5 mM	0.1 mL
H_2O	-	-	6.4 mL

Store buffer at 4°C for up to 6 months. Add TCEP fresh on the day you perform your assay.

⚠ CRITICAL: Take appropriate precaution when handling the indicated hazardous chemical substances. Consult safety data sheets and standard operating procedures and conduct a risk assessment prior to starting the work. Wear gloves, lab coats and goggles and work under a flow hood, if possible.

STEP-BY-STEP METHOD DETAILS

Conjugation and purification of protein-oligonucleotide conjugates

Timing: approx. 8 h

To generate DPC model substrates, we repurposed a commercially available solution, which was originally developed for the conjugation of protein ligands to DNA-based biosensors. Protein and oligonucleotide are conjugated using the proFIRE Amine Coupling Kit. Subsequently, conjugates are purified using the proFIRE chromatography system. Purification of the conjugate is in principle also possible using other chromatography systems (e.g., ÄKTA (Cytiva) or NGC (Bio-Rad)). However, other systems may require additional optimization to reliably separate proteins and oligonucleotides from conjugates.

Note: First, it is important to determine the elution profile of unconjugated DNA in anion exchange chromatography. The following steps can be performed at 20°C–25°C unless indicated otherwise.

1. Resuspend one aliquot of thio-dT-modified oligonucleotide (3 nmol) in 40 µL dilution buffer (proFIRE Amine Coupling Kit). Resuspend the solution thoroughly by pipetting up-and-down or mild

CellPress
OPEN ACCESS

STAR Protocols
Protocol

vortexing. Centrifuge the tube and place on ice. Always make sure to protect fluorescently labeled DNA from light.

2. Equilibrate the proFIRE anion exchange column first with water and perform a test run using buffer A and B according to the proFIRE manual to ensure that the system is free of leaks and that pump and column pressures are correct.
3. Take 20 µL of the resuspended oligonucleotide and dilute with 140 µL dilution buffer. Keep left-over oligonucleotide (20 µL) on ice, in case the equilibration has to be repeated. Store at −20°C after successful equilibration.
4. Inject 160 µL diluted DNA and run it using default settings that correspond to the DNA length (e.g., 30 nucleotides).
5. Analyze elution profile. Different oligonucleotides interact with varying strength with the anion exchange resin. By increasing the salt concentration during chromatography from 150 mM NaCl (proFIRE buffer A) to 1 M NaCl (proFIRE buffer B), the specific salt concentration at which a particular unconjugated DNA elutes is determined.
6. The protein-oligonucleotide conjugate will elute at lower salt concentrations than unconjugated DNA. Set up a chromatography protocol so that 12 fractions of 0.6 CVs each are collected prior to elution of unconjugated DNA.

Note: Once a calibration for one specific oligonucleotide sequence is performed, the same chromatography program can be applied for other conjugation reactions using the same DNA. Different oligonucleotide lengths or sequences will lead to different elution profiles (Figure 2). If oligonucleotide length or sequence are altered, another calibration has to be performed.

Pause point: It is possible to pause at this point.

Note: During this step the protein is conjugated to the oligonucleotide and the conjugate is separated from unconjugated protein and DNA using anion exchange chromatography. The conjugation reaction requires incubation for at least 12 h at 4°C. We recommend to start the protocol in the afternoon (4 pm) and continue with the purification the next morning (10 am). The oligonucleotide needs to be protected from light during all steps using aluminium foil or using black plastic tubes. The following steps can be performed at 20°C–25°C.

7. Thaw dilution and conjugation buffer (proFIRE Amine Coupling Kit) until they are completely dissolved.
8. Equilibrate two purification spin columns per coupling reaction. Add 400 µL of conjugation buffer and subsequently centrifuge at 1500 × *g* for 1 minute. Discard flow

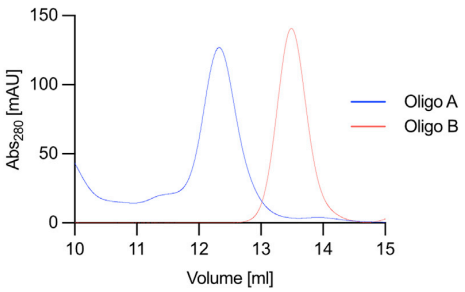


Figure 2. Oligonucleotide size and sequence influence elution properties
Two unconjugated 30-mer oligonucleotides with different sequences (oligo A and B) elute differently in anion exchange chromatography.

STAR Protocols
Protocol

through and add another 400 µL of conjugation buffer. Centrifuge again and discard flow through.

9. Dissolve 3 nmol modified oligonucleotide in 40 µL dilution buffer and vortex mildly until completely dissolved.
10. Dissolve the bifunctional crosslinker (proFIRE Amine Coupling Kit) in 100 µL H_2O and vortex.
11. Add 10 µL of the dissolved crosslinker to the diluted oligonucleotide.
12. Mix reaction by flicking the tube, spin down, and incubate for 5 min at 20°C–25°C.
13. Purify sample by adding the oligonucleotide from step 12 on top of the first purification spin column. Centrifuge at 1500 × *g* for 2 min and collect flow through. Discard spin column after use.
14. Add flow through on top of the second column and centrifuge again at 1500 × *g* for 2 min. Collect flow through in a fresh 1.5 mL tube. The flow through contains the oligonucleotide.
15. Add 50 µL of your aliquoted protein (250 µg) to the oligonucleotide and pipette up and down.
16. Incubate the conjugation between 12–15 h on ice while protected from light.

Optional: Reactions can also be incubated for 1 h at 20°C–25°C which has the advantage of performing crosslinking reaction and conjugate purification in a single day. However, the shorter incubation time may lead to decreased crosslinking efficiency. Therefore, we recommend to incubate the reaction for at least 12 h at 4°C. Incubation for up to 15 h is well tolerated by protein G, but we recommend to test the stability of other proteins before using them in conjugation reactions.

17. On the next day, equilibrate the proFIRE anion exchange column with water and subsequently with buffer B followed by buffer A. Conduct a test run by injecting 160 µL dilution buffer using the chromatography program you generated during calibration.
18. Following the test run, inject at least 160 µL (a) of the conjugation reaction and start the run (b).
    a. The volume of the conjugation reaction is approximately 90 µL. Take the entire reaction into a syringe and fill it up with buffer A to a total volume of 160 µL. Avoid air bubbles inside the syringe to not harm the purification column.
    b. During the run, a salt gradient is applied to the column starting from 150 mM NaCl to 865 mM NaCl in 16 CV (1 CV = 1 mL), followed by 2 CV of 1 M NaCl and 6 CV 150 mM NaCl.
19. Collect fractions containing the protein-oligonucleotide conjugate (0.6 CV/fraction) in 1.5 mL tubes.

⚠ CRITICAL: The purity of the conjugate is more important than its yield. Fractions overlapping with the elution of unconjugated DNA (judged by the absorbance profile, Figure 3) should not be selected as this might interfere with results of conjugate cleavage and binding assays (troubleshooting 3).

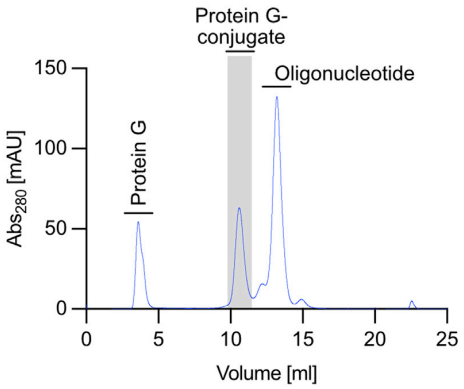


Figure 3. Conjugate purification
Following the crosslinking reaction, the conjugate is separated from unconjugated protein and oligonucleotide using anion exchange chromatography. The conjugate is collected in fractions of 0.6 column volumes (gray-shaded area).

CellPress
OPEN ACCESS

STAR Protocols
Protocol

20. Equilibrate centrifugal filters (3 kDa cut-off buffer exchange columns, proFIRE Amine Coupling Kit) with conjugate desalting buffer.
21. Add 500 µL of the first conjugate fraction on top of the filter and centrifuge at 4°C and 12000 × *g* for 15 min. Discard flow through and repeat until all fractions are collected in one filter. After adding the final 500 µL of conjugate, centrifuge until 100 µL of sample remain on top of the filter.
22. Add 300 µL conjugate desalting buffer and centrifuge at 4°C and 12000 × *g* for 15 min.
23. Repeat previous step three times and discard flow through each time.
24. Centrifuge until 150 µL of sample remains on top of the filter. Transfer the concentrated conjugate to a fresh tube.
25. Prepare a standard curve of unconjugated fluorescently labeled DNA with concentrations ranging from 0.1 to 10 µM and measure Cy5 (649 nm) or 6-FAM (495 nm) absorbance using a SpectraMax Paradigm Multi-Mode Detection platform. Determine Cy5 or 6-FAM absorbance of the purified and concentrated conjugate and calculate the concentration based on the standard.

Optional: Conjugate concentration can also be determined by measuring absorbance at 260 nm using a Nanodrop photometer.

26. After determining the conjugate concentration, add BSA to a final concentration of 0.4 mg/mL. BSA is important to stabilize the conjugate. Without BSA, conjugates tend to aggregate over time (troubleshooting 4).
27. Aliquot sample in 3 µL aliquots in 0.2 mL thin-wall PCR tubes. Snap freeze in liquid nitrogen and store protected from light at −80°C.

Note: Cy5-labelled DNA tends to adhere to plastic tubes. Therefore, we recommend to aliquot and store conjugates in low-binding PCR tubes.

Pause point: Conjugates are stable at −80°C for > 1 year. Ensure that they are protected from light.

Protein-oligonucleotide conjugate cleavage assay

Timing: 8 h

In this step, cleavage of protein-oligonucleotide conjugates by the DPC protease SPRTN is analyzed using native 1 × TBE gels and a fluorescent imaging system. Three model substrates with different DNA structures are generated during the annealing reaction: a ssDPC, a junction-DPC, and a dsDPC (Figure 4). The same DNA structures are also generated without a conjugated protein as controls. Substrates are then tested for cleavage by wild-type (WT) SPRTN with a catalytic inactive E112Q (EQ) variant serving as control.

Note: Ideally, native TBE gels should be casted on the day of the experiment. If necessary, gels can be casted in advance and stored at 4°C after wrapping in tissues soaked with TBE running buffer. Gels should not be stored for more than 3 days. A separating and a stacking part are required for 1 × TBE gels used in conjugate cleavage assays. To achieve reproducible results, we recommend to always cast identically sized separating parts. Therefore, label the intended top position of the separating gel on the gel casting cassettes.

Alternatives: We have good experience using disposable gel casting cassettes (Thermo Fisher). Alternatively, other gel systems can be used to cast TBE gels. 1 × and 0.5 × TBE gels are also commercially available as pre-cast gels. However, native pre-cast gels have the disadvantage of a short shelf-life (< 6 weeks).

STAR Protocols
Protocol

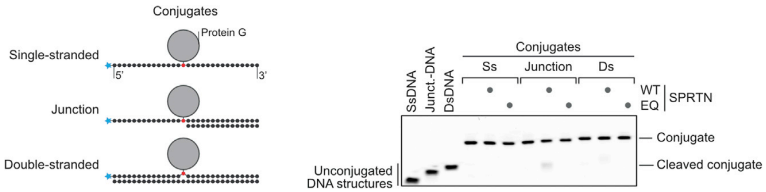


Figure 4. Protein-oligonucleotide conjugate cleavage assay

Schematic depiction of the fluorescently labeled protein G-oligonucleotide conjugates used to probe SPRTN specificity (left panel). Recombinant SPRTN (5 nM, WT or the catalytically inactive EQ variant) is incubated with the indicated model substrates (25 nM) for 2 h at 25°C prior to separation on native TBE gels (right panel).

28. Cast 1 × native TBE gel (8 × 8 cm Mini gel size) for protein-oligonucleotide conjugate cleavage assays according to Table 2 (separating gel: 20% polyacrylamide, stacking gel: 4% polyacrylamide).
    a. Pipette stock solutions for the separating gel into a 15 mL conical tube. Add APS and TEMED last. Close the tube and invert 2–3 times.
    b. Pour the separating gel immediately into the casting cassette until it reaches the intended level. This level should be beneath the comb which will be inserted later on.
    c. Immediately cover the gel with 500 µL isopropanol and let it polymerize for 45 min.
    d. Invert cassette and collect isopropanol with a tissue.
    e. Pipette stock solutions for the stacking gel into a 15 mL conical tube, add APS and TEMED last and invert 2–3 times.
    f. Cast stacking gel and immediately insert comb. Let gel polymerize for 45 min.
29. Place the polymerized gels into a running chamber filled with 1 × TBE running buffer. Buffers should be pre-cooled to 4°C.
30. Flush the wells carefully using a 200 µL pipette.
31. Pre-run gel (preferably at 4°C) at 100 V for 30 min

⚠ CRITICAL: Acrylamide is a toxic chemical. Cast gel in a flow hood and take appropriate caution when handling hazardous substances (wear gloves and goggles). Provide enough time for the gel to polymerize to completion. If the comb is taken out too early, the pockets will not form properly. However, do not polymerize the gel for more than 2 h, because the gels tend to crack, if dried out.

Table 2. Recipe for 1 × TBE gel for protein-oligonucleotide conjugate cleavage assays

Reagent	Stock concentration	Final concentration	
Separating gel			**Add in 8 mL**
Acrylamide/bisacrylamide (29:1) (hazardous)	30%	20%	5.3 mL
TBE	10 ×	1 ×	0.8 mL
H_2O	n/a	n/a	1.9 mL
APS (hazardous)	10%	n/a	50 µL
Tetramethylethylendiamine (TEMED) (hazardous)	n/a	n/a	5 µL
Stacking gel			**Add in 6 mL**
Acrylamide/bisacrylamide (hazardous)	30%	4%	0.8 mL
TBE	10 ×	1 ×	0.6 mL
H_2O	n/a	n/a	4.6 mL
APS (hazardous)	10%	n/a	50 µL
TEMED (hazardous)	n/a	n/a	5 µL

CellPress
OPEN ACCESS

STAR Protocols
Protocol

⚠ CRITICAL: The DPC protease SPRTN tends to precipitate at elevated temperatures. Thus, all running buffers have to be pre-cooled prior to running the gel. Additionally, running gel electrophoreses at 4°C will improve results.

Note: All steps containing Cy5-labelled DNA are carried out in PCR or low-binding tubes. When using standard tubes, Cy5 adheres to the plastic, which may result in differences between samples. Prepare all dilutions on ice and turn off the light to avoid photobleaching. Protect tubes from light with aluminium foil.

Note: Steps 32–36 are also used for the binding assays and will be referred to below.

32. Prepare conjugate storage buffer containing 0.4 mg/mL BSA, 500 mM KCl, 50 mM TCEP and high-salt SPRTN buffer as described in materials and equipment. Add TCEP and BSA to the buffers as indicated.
33. Dilute the protein-oligonucleotide conjugate and the forward control oligo to 100 nM in conjugate storage buffer (Table 3, DNA 1 and 2).

Table 3. Oligonucleotides and conjugate used to generate different model substrates for the protein-oligonucleotide conjugate cleavage assay

DNA	Name	Sequence
1	Forward control oligo	5′-Cy5-CCCAAAAAAAAAAA TAAAAAAAAAAAACCC-3′
2	Protein-oligonucleotide conjugate	5′-Cy5-CCCAAAAAAAAAAA-protein G-dT-AAAAAAAAAAAACCC-3′-phos
3	Reverse oligo 1 (for junction-DPC)	5′- GGGTTTTTTTTTTT - 3′
4	Reverse oligo 2 (for ds-DPC)	5′- GGGTTTTTTTTTTTATTTTTTTTTTGGG - 3′

34. Dilute reverse oligos in H_2O to 120 nM (Table 3, DNA 3 and 4).

Note: The excess of reverse oligo ensures that conjugates anneal to completion.

35. Mix DNAs from Table 3 with a volume ratio of 1:1 in PCR tubes to generate substrates with different DNA structures according to Table 4.

Note: The required amount of annealed substrates depends on the number of conditions to be tested in the assay. 5 µL of annealed substrate is needed for one cleavage/binding reaction. As an example, testing cleavage of SPRTN-WT, SPRTN-EQ and a negative control without protease requires at least 15 µL of annealed substrate (3 × 5 µL). Thus, 10 µL forward DNA are annealed with 10 µL reverse DNA to generate a small excess.

Table 4. Pipetting scheme for the annealing of DNAs for the generation different DNA and DPC structures used in the protein-oligonucleotide conjugate cleavage assay

Substrate	DNA (10 µL)	DNA (10 µL)	Resulting structure
a	1	H_2O	SsDNA
b	1	3	Junction-DNA
c	1	4	DsDNA
d	2	H_2O	SsDPC
e	2	3	Junction-DPC
f	2	4	DsDPC

STAR Protocols
Protocol

CellPress
OPEN ACCESS

Table 5. Parameters for annealing model substrates for the protein-oligonucleotide conjugate cleavage assay in a thermal cycler

Steps	Temperature	Time	Cycles
Initial denaturation	37°C	2 min	1
Annealing	37°C (−1°C/cycle)	1 min	18
Target temperature	20°C	Forever	1

36. Anneal reactions in a thermal cycler using the following settings (Table 5).

Note: The temperature of the initial denaturation step is limited by the stability of the conjugated protein. High temperatures will result in protein denaturation and precipitation. In our experience, denaturing at 37°C is sufficient to allow annealing of most sequences.

37. Prepare master mix (Table 6). Prepare a small excess of the master mix (e.g., a 15 × master mix for 12 cleavage reactions).

Table 6. Master mix required for the protein-oligonucleotide conjugate cleavage assay

Reagent	Stock concentration	Volume for 1 cleavage reaction	Volume for 15 cleavage reactions
KCl	500 mM	0.1 µL	1.5 µL
TCEP	50 mM	0.9 µL	13.5 µL
H_2O	n/a	3 µL	45 µL

38. Distribute 4 µL of master mix into 0.2 mL PCR tubes. Keep on ice (Table 7).
39. Add 5 µL of the substrates a-f (Table 4) to PCR tubes (Table 7).

Table 7. Pipetting scheme for the protein-oligonucleotide conjugate cleavage assay

Reaction	1	2	3	4	5	6	7	8	9	10	11	12
Master mix	4 µL	4 µL	4 µL	4 µL	4 µL	4 µL	4 µL	4 µL	4 µL	4 µL	4 µL	4 µL
Substrate a	5 µL	-	-	-	-	-	-	-	-	-	-	-
Substrate b	-	5 µL	-	-	-	-	-	-	-	-	-	-
Substrate c	-	-	5 µL	-	-	-	-	-	-	-	-	-
Substrate d	-	-	-	5 µL	5 µL	5 µL	-	-	-	-	-	-
Substrate e	-	-	-	-	-	-	5 µL	5 µL	5 µL	-	-	-
Substrate f	-	-	-	-	-	-	-	-	-	5 µL	5 µL	5 µL
Buffer	1 µL	1 µL	1 µL	1 µL	-	-	1 µL	-	-	1 µL	-	-
SPRTN-WT	-	-	-	-	1 µL	-	-	1 µL	-	-	1 µL	-
SPRTN-EQ	-	-	-	-	-	1 µL	-	-	1 µL	-	-	1 µL

40. Flick the tube, spin down, and place on ice. Cover samples with aluminum foil.
41. Thaw the DPC protease (e.g., SPRTN WT and EQ) on ice and prepare the desired stock concentration in high-salt SPRTN buffer.
    a. SPRTN precipitates at elevated temperatures. Thaw enzyme aliquots on ice. Once SPRTN is diluted in high salt SPRTN buffer, add immediately to reactions.
    b. A 50 nM stock concentration of SPRTN will result in a final assay concentration of 5 nM, which should be sufficient for efficient conjugate cleavage. High SPRTN concentrations (> 50 nM assay concentration) will interfere with interpretation of the result due to complex formation between SPRTN and the cleaved conjugate (troubleshooting 5).
42. Add 1 µL of the respective protease or high-salt SPRTN buffer as negative control to the reactions.
    a. The final reaction volume is 10 µL containing 25 nM DPC or DNA in 80 mM KCl, 17.5 mM HEPEs (pH 7.5), 3.5% Glycerol, 0.1 mg/mL BSA and 5 mM TCEP.

CellPress
OPEN ACCESS

STAR Protocols
Protocol

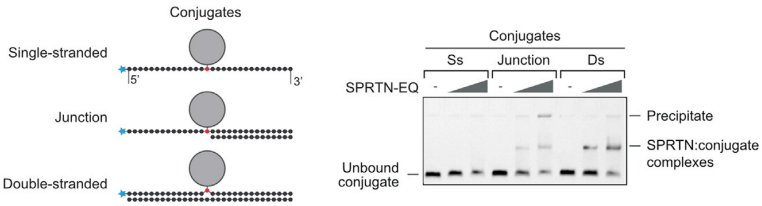


Figure 5. Protein-oligonucleotide conjugate binding assay

Schematic depiction of the fluorescently labeled protein G-oligonucleotide conjugates used to assess SPRTN binding (left panel). Recombinant catalytically inactive SPRTN EQ (12.5 and 50 nM) is incubated with the indicated model substrates (25 nM) prior to separation on native TBE gels (right panel).

43. Incubate reactions at 25°C for 2 h in a thermal cycler.
44. In the meantime, pre-run 1 × TBE gel (ideally at 4°C). Run at 100 V for 30 min in 1 × TBE running buffer.
45. Upon completion of the cleavage reaction, add 2 µL 6 × Orange G loading dye to each reaction and spin down.
46. Load 10 µL of sample per well of the 1 × TBE gel and run at 100 V for 90 min. Cover chamber with aluminum foil.
47. Photograph the gel using a Bio-Rad Chemidoc MP imaging system with filter settings for Cy5 or 6-FAM fluorescence. Avoid overexposure.

Alternatives: A Typhoon FLA scanner (GE Healthcare) or similar fluorescent imaging systems and scanners can be used to image gels.

Protein-oligonucleotide conjugate binding assay

Timing: 6 h

In this step, binding of a catalytically inactive variant of the DPC protease SPRTN (SPRTN-EQ) to different substrates is analyzed using an electrophoretic mobility shift assay (EMSA) with native 0.5 × TBE gels and a fluorescent imaging system. Three different DNA structures are generated during the annealing reaction: a ssDPC, a junction-DPC and a dsDPC (Figure 5). We recommend to titrate the amount of protease by testing a series of concentrations. Therefore at least two different concentrations are used to assess binding to different substrates. High-salt SPRTN buffer is used as negative control.

48. Cast 0.5 × TBE gel (8 × 8 cm Mini gel size) for protein-oligonucleotide conjugate binding assays according to Table 8 (separating gel: 6% polyacrylamide, no stacking gel)
    a. Pipette solutions into a 15 mL conical tube. Add APS and TEMED last. Invert the tube 2–3 times and pour the entire solution directly until the top of the gel.
    b. Insert the comb immediately and let the gel polymerize for 45 min.

Table 8. Recipe for 0.5 × TBE gel for protein-oligonucleotide conjugate binding assays

Reagent	Stock concentration	Final concentration	Add in 10 mL
Separating gel			
Acrylamide/bisacrylamide (29:1) (hazardous)	30%	6%	2 mL
TBE	10 ×	0.5 ×	0.5 mL
H_2O	n/a	n/a	7.5 mL
APS (hazardous)	10%	n/a	50 µL
TEMED (hazardous)	n/a	n/a	5 µL

STAR Protocols
Protocol

49. Place the polymerized gels into a running chamber filled with 0.5 × TBE running buffer. Buffers should be pre-cooled to 4°C.
50. Flush the wells carefully using a 200 µL pipette.
51. Pre-run gel (preferably at 4°C) at 90 V for 30 min
52. Generate substrates for the binding assay by following steps 32 - 36 of the protein-oligonucleotide conjugate cleavage assay.
53. Prepare master mix (Table 9). Prepare a small excess of the master mix (e.g., 12 × master mix for 9 binding reactions).

Table 9. Master mix required for the protein-oligonucleotide conjugate binding assay

Reagent	Stock concentration	Volume for 1 reaction	Volume for 12 reactions
KCl	500 mM	0.1 µL	1.2 µL
TCEP	50 mM	0.9 µL	10.8 µL
H_2O	n/a	3 µL	36 µL

54. Distribute 4 µL of master mix into 0.2 mL PCR tubes. Keep on ice (Table 10).
55. Add 5 µL substrate d-f (Table 4) to PCR tubes according to Table 10.

Table 10. Pipetting scheme for the protein-oligonucleotide conjugate binding assay

Reaction	1	2	3	4	5	6	7	8	9
Mastermix	4 µL	4 µL	4 µL	4 µL	4 µL	4 µL	4 µL	4 µL	4 µL
Substrate d	5 µL	5 µL	5 µL	-	-	-	-	-	-
Substrate e	-	-	-	5 µL	5 µL	5 µL	-	-	-
Substrate f	-	-	-	-	-	-	5 µL	5 µL	5 µL
Buffer	1 µL	-	-	1 µL	-	-	1 µL	-	-
EQ (125 nM)	-	1 µL	-	-	1 µL	-	-	1 µL	-
EQ (500 nM)	-	-	1 µL	-	-	1 µL	-	-	1 µL

56. Flick the tube, spin down, and place on ice. Cover samples with aluminum foil.
57. Thaw the DPC protease (SPRTN-EQ) on ice and prepare the desired stock concentration in high-salt SPRTN buffer.
    a. SPRTN precipitates at elevated temperatures. Thaw enzyme aliquots on ice. Once SPRTN is diluted in high salt SPRTN buffer, add immediately to reactions.
    b. A 125 nM and 500 nM stock concentration of SPRTN will result in a final assay concentration of 12.5 nM and 50 nM.
58. Add 1 µL of protease or high-salt SPRTN buffer as negative control to the reactions according to Table 10.
    a. The final reaction volume is 10 µL containing 25 nM DPC or DNA in 80 mM KCl, 17.5 mM HEPEs (pH 7.5), 3.5% Glycerol, 0.1 mg/mL BSA and 5 mM TCEP.
59. Incubate reactions at 4°C for 15 min on ice.
60. Upon completion of the binding reaction, add 2 µL 6 × Orange G loading dye to each reaction and spin down.
61. Load 10 µL of sample per well of the 0.5 × TBE gel and run at 90 V for 40 min. Cover chamber with aluminum foil.
62. Photograph the gel using a Bio-Rad Chemidoc MP imaging system using filter settings for Cy5 or 6-FAM fluorescence. Avoid overexposure.

Note: To avoid precipitation of SPRTN, run the electrophoreses at 4°C.

EXPECTED OUTCOMES

The success of the conjugation reaction can be estimated by analyzing the anion exchange chromatography elution profile. The profile should show three clearly separated peaks: (1) the non-conjugated

CellPress
OPEN ACCESS

STAR Protocols
Protocol

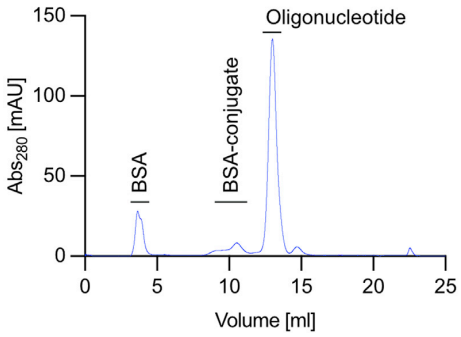


Figure 6. Anion exchange chromatography of a conjugation reaction with BSA and a 30-mer oligonucleotide

The small conjugate peak indicates a poor crosslinking efficiency.

protein eluting first; (2) the protein-oligonucleotide conjugate eluting second; and (3) non-conjugated oligonucleotide eluting last (Figure 3). Some non-conjugated protein and oligonucleotide will always remain after the conjugation reaction. Following the anion exchange chromatography, fractions containing the conjugate are collected and pooled (Figure 3, gray area). Typically, this results in a volume of 1.8 mL with a concentration of around 200–300 nM. Upon concentration and desalting, a final yield of 150 μL with a concentration of 2–3 μM should be obtained. Yield and crosslinking efficiency depend on the protein and oligonucleotide used. In our hands, crosslinking of protein G to a 30-mer oligonucleotide results in high yields and pure conjugates.

Substrate and cleavage fragments are separated using native TBE gels in conjugate cleavage assays. The cleaved conjugate migrates slightly above unconjugated DNA, because it contains the entire oligonucleotide and a peptide remnant (Figure 4). SPRTN displays a strict DNA structure-specific activity with a strong preference for proteins conjugated at or in close proximity to DNA structures containing single- and double-stranded features (Reinking et al., 2020). Thus, SPRTN cleaves a conjugate at an ssDNA/dsDNA junction but not within ssDNA and only poorly within dsDNA (Figure 4).

Binding of a protein to a protein-oligonucleotide conjugate retards migration of the conjugate in native 0.5 × TBE gels (Figure 5). The DPC protease SPRTN-EQ requires a stretch of dsDNA for stable binding (Reinking et al., 2020). Accordingly, SPRTN binds to conjugates at an ssDNA/dsDNA junction and within dsDNA but not within ssDNA. Of note, a signal in the pocket of the gel indicates aggregation and may be observed at higher protein concentrations (Figure 5).

LIMITATIONS

In the protocol described here, oligonucleotides are crosslinked to primary amines within proteins (e.g., to N-termini). Accessibility of these amines will affect the conjugation reaction. Thus, different proteins show distinct conjugation efficiencies (even when of similar molecular weight). Accordingly, the model substrates described here can only partially replicate the complexity of DPCs in cells, which can involve various chromatin proteins. Moreover, this protocol is optimized for analyzing the DPC protease SPRTN. The investigation of other DPC-processing enzymes may thus require additional optimization steps. Finally, conjugate cleavage assays are analyzed using native 1 × TBE gels, which cannot resolve tight DNA-protein interactions. Thus, the presence of other DNA-binding proteins such as RPA may interfere with the assay read-out, due to tight binding to substrate and product of the cleavage reaction.

TROUBLESHOOTING

Problem 1

Poor conjugate yields (Figure 6, note in Preparation of proteins).

STAR Protocols
Protocol

CellPress
OPEN ACCESS

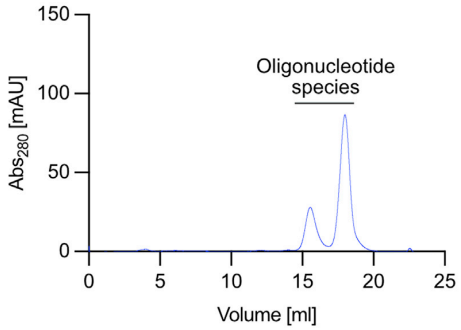


Figure 7. Anion exchange chromatography elution profile of a 30-mer oligonucleotide
Two differentially eluting oligonucleotide species can be observed, which indicates that the integrity of the oligonucleotide is not sufficient to obtain homogenous conjugates.

Potential solution

Some proteins show low conjugation efficiencies, as evident by a small conjugate peak during anion exchange chromatography (Figure 6). Poor accessibility of primary amine groups within the protein may cause low conjugate yields. Introduction of an N-terminal linker can increase accessibility of the protein's N terminus and thereby improve yields. In addition, increased protein concentration and longer incubation times (12 h, 4°C vs 1 h, 20°C–25°C) may help to improve conjugation efficiency.

Problem 2

Multiple oligonucleotide species prior to conjugation (Figure 7, step 4 in Ordering modified DNA oligonucleotides for protein conjugation).

Potential solution

Multiple species observed during ion-exchange chromatography of the unconjugated DNA are indicative of poor oligonucleotide quality (Figure 7). Integrity of the modified oligonucleotide is essential to obtain high-quality substrates. Thus, reorder from a different supplier.

Problem 3

Multiple conjugate-species and contamination with unconjugated DNA (Figure 8, step 19 in Conjugation and purification of protein-oligonucleotide conjugates).

Potential solution

Insufficient protein purity will result in heterogenous conjugates. Thus, determine purity of the protein using an SDS PAGE and Coomassie or Silver staining prior to conjugation. To avoid

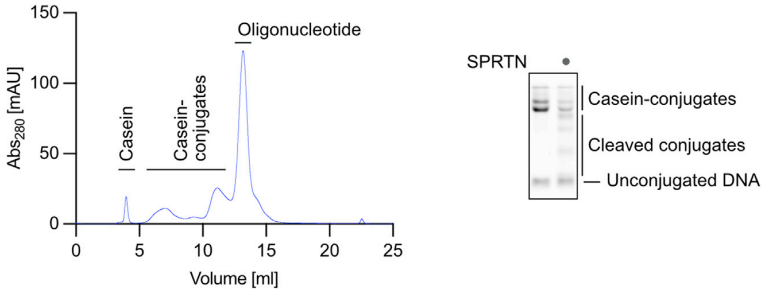


Figure 8. Anion exchange chromatography of a conjugation reaction with casein and a 30-mer oligonucleotide
Multiple conjugate peaks can be observed during the purification (left panel). Cleavage reaction using pooled casein-oligonucleotide conjugates and recombinant SPRTN (right panel). Multiple conjugate bands and contaminating unconjugated oligonucleotide can be detected.

CellPress
OPEN ACCESS

STAR Protocols
Protocol

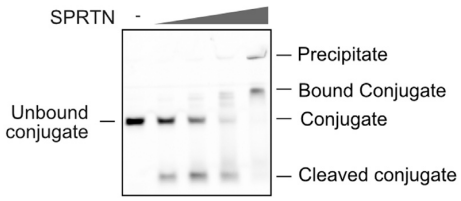


Figure 9. Protein-oligonucleotide conjugate cleavage assay

Increasing concentrations of recombinant SPRTN are incubated with model substrates prior to separation on a native TBE gel. The result is difficult to interpret, because high SPRTN concentrations lead to stable SPRTN:substrate/product complex formation and precipitation.

contamination of the purified conjugate with unconjugated oligonucleotides, collect only the peak of the eluting conjugate.

Problem 4

No or weak fluorescence signal in gel-based assays (step 26 in Conjugation and purification of protein-oligonucleotide conjugates).

Potential solution

Add BSA after conjugate purification and before snap freezing and storage. Additionally, add BSA to the conjugate dilution buffer at the beginning of each assay. Avoid exposure with light and cover samples with aluminum foil.

Problem 5

Complex formation between SPRTN and substrate/product in cleavage assays (Figure 9, step 41 in Protein-oligonucleotide conjugate cleavage assay).

Potential solution

High concentrations of SPRTN result in the formation of complexes between SPRTN and cleaved and uncleaved conjugate complexes, which are not resolved in the native TBE gels used to analyze cleavage reactions. This interferes with analyzing cleavage efficiency. Reduce the concentration of SPRTN to substoichiometric amounts. If high SPRTN concentrations are required, use denaturing TBE-UREA gels to analyze cleavage reactions.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Julian Stingele (stingele@genzentrum.lmu.de).

Materials availability

No materials to declare.

Data and code availability

This study did not generate code or reposited datasets.

ACKNOWLEDGMENTS

We thank R. Strasser for advice, D. Yaneva for testing the protocol, and P. Weickert for comments on the protocol. J.S. is supported by the European Research Council (ERC Starting Grant 801750 DNAProteinCrosslinks), by the Alfried Krupp Prize for Young University Teachers awarded by the Alfried-Krupp von Bohlen und Halbach-Stiftung, the European Molecular Biology Organization (YIP4644), and the Deutsche Forschungsgemeinschaft (CRC1064).

STAR Protocols
Protocol

CellPress
OPEN ACCESS

AUTHOR CONTRIBUTIONS

Conceptualization, H.K.R. and J.S.; investigation, H.K.R.; writing – original draft, H.K.R.; writing – review & editing, H.K.R. and J.S.; funding acquisition, J.S.; supervision, J.S..

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Bhargava, V., Goldstein, C.D., 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.e10.

Dokshin, G.A., Davis, G.M., Sawle, A.D., Eldridge, M.D., Nicholls, P.K., Gourley, T.E., Romer, K.A., Molesworth, L.W., Tatnell, H.R., Ozturk, A.R., et al. (2020). GCNA interacts with spartan and topoisomerase II to regulate genome stability. Dev. Cell *52*, 53–68.e56.

Kojima, Y., Machida, Y., Palani, S., Caulfield, T.R., Radisky, E.S., Kaufmann, S.H., and Machida, Y.J. (2020). FAM111A protects replication forks from protein obstacles via its trypsin-like domain. Nat. Commun. *11*, 1318.

Reinking, H.K., Kang, H.S., Gotz, M.J., Li, H.Y., Kieser, A., Zhao, S., Acampora, A.C., Weickert, P., Fessler, E., Jae, L.T., et al. (2020). DNA structure-specific cleavage of DNA-protein crosslinks by the SPRTN protease. Mol. Cell *80*, 102–113.e6.

Serbyn, N., Noireterre, A., Bagdiul, I., Plank, M., Michel, A.H., Loewith, R., Kornmann, B., and Stutz, F. (2020). The aspartic protease Ddi1 contributes to DNA-protein crosslink repair in yeast. Mol. Cell *77*, 1066–1079.e9.

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

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

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

3 Discussion

3.1 DNA Structure-Specific Cleavage of SPRTN

DNA-protein crosslinks are toxic DNA lesions, which are constantly generated within cells through various endogenous and exogenous agents (Stingele et al., 2015). They pose an exceptional challenge to cells because they can disrupt DNA-related processes such as replication and transcription (Duxin et al., 2014; Nakano et al., 2013; Nakano et al., 2012). The discovery of the DPC-specific proteases Wss1 and SPRTN opened an entire new research field dedicated to the repair of DPCs. SPRTN is a mammalian protease, which helps cells to cope with DPC-inducing agents such as formaldehyde or camptothecin (Stingele et al., 2016; Vaz et al., 2016). SPRTN is essential for cell viability with a knockout of the enzyme resulting in cell death (Stingele et al., 2016). Moreover, mutations within the *SPRTN* gene lead to a severe disease associated with premature ageing and cancer (Lessel et al., 2014).

In vitro characterization of the enzyme revealed a strict DNA-dependent activation (Stingele et al., 2016; Vaz et al., 2016). SPRTN only becomes activated once bound to DNA and can then perform two different modes of action. First, it can undergo autocleavage *in trans*, meaning one SPRTN molecule is cleaving another and second, it can degrade any DNA-binding protein. Interestingly, apart from the DNA-dependency, SPRTN is not restricted towards the identity of the substrates and can degrade various proteins (Stingele et al., 2016; Vaz et al., 2016). *In vitro* it degrades various DNA-binding proteins such as H1, H2A, H2B, H3 and Hgm1 irrespective of their size or structure (Stingele et al., 2016). *In vivo* it contributes to the repair of various DPCs including those induced by formaldehyde as well as by chemotherapeutics such as camptothecin. This promiscuity is a two-sided sword: on one side, it can become beneficial when targeting divers DPCs but on the other hand it can be harmful when essential DNA-binding proteins are degraded. In the first part of this thesis, it was shown that SPRTN´s activity is tightly coupled to the presence of certain DNA structures in close vicinity to the crosslinked protein. These structures include hairpins, ss- to dsDNA junctions as well as dsDNA ends. The feature that these structures have in

common is the presence of unpaired nucleotides within dsDNA, or so-called dsDNA disruptions. Since cleavage of proteins is restricted to this very narrow window, DNA-binding proteins, which are not found in the proximity, are protected.

Indeed, certain DPCs contain such DNA structures and SPRTN was shown to participate in their repair. SPRTN contributes to the repair of TOP1-DPCs trapped at 3´DNA nicks (Stingele et al., 2016). Additionally, TOP2 and SPO11 adducts are found at 5´DNA ends and SPRTN protects cells against agents that induce those DPCs. (Dokshin et al., 2020; Lopez-Mosqueda et al., 2016; Vaz et al., 2016). Moreover, it was shown that SPRTN can repair DPCs in a replication dependent manner (Larsen et al., 2019). When the CMG helicase bypasses the intact DPC during replication, the lesion is transferred to ssDNA. Only upon polymerase approach to within a few nucleotides of the DPC, SPRTN is able to degrade the protein component. Here, a ss- to dsDNA junction is generated close to the crosslink, which allows SPRTN activation.

These examples are in line with the DNA-structure dependency discovered in this thesis. However, also other structures were identified during the *in vitro* analysis, which have not been associated with DPC repair by SPRTN so far. It was shown that SPRTN becomes activated once the DPC is placed in a DNA bubble with more than 2 unpaired DNA bases (Reinking et al., 2020b). Transcription is a bubble generating process and some genetic evidence exists, which indicates that SPRTN also acts independent of replication (Stingele et al., 2016). However, if and how SPRTN contributes to transcription-coupled DPC repair is unknown.

On a protein level, SPRTN activation can be explained by the differential engagement of SPRTN´s DNA-binding domains (Reinking et al., 2020b). It was shown that the ZBD interacts with unpaired nucleotides while the BR is engaged with the negative phosphate backbone of the dsDNA duplex. Only upon simultaneous engagement of those two domains at the specific DNA structure, the enzyme becomes activated and cleaves crosslinks located at DNA disruptions. Even though the contribution of those two domains is clear, it still has to be determined how the substrate gains access to the active site. Two different scenarios are conceivable. In the crystal structure of human SPRTN it seems that the ZBD blocks access to the active site (Li et al., 2019). It can be speculated that while engaging with the specific DNA structure, the enzyme undergoes dynamic conformational changes and liberates the access for the protease domain. However, evidence for this hypothesis is still missing. An alternative interpretation is that the enzyme is locked in a stable conformation

and the protein component of the DPC is placed in the right position for degradation. In this scenario, the protein should be unfolded and flexible in order to enter the narrow groove between protease domain and the ZBD. This might require preprocessing of the protein component. SPRTN contains a SHP binding motive for the interaction with the segregase p97, which could unfold the DPC prior to the degradation. However, given the fact that the SHP box is dispensable for viability, the importance of such a mechanism is questionable (Larsen et al., 2019). Moreover, diverse proteins with different structures are cleaved efficiently *in vitro* without the presence of such unfolding machineries, indicating that this hypothesis does not hold true. In order to draw more conclusion about the precise activation mechanism, further studies have to be employed that target the role of the protein component that is crosslinked to the DNA. Therefore, different proteins could be crosslinked to the oligonucleotides and their cleavage by SPRTN assessed in *in vitro* cleavage assays. Additionally, in combination with mass spectrometry certain cleavage sites within the substrate could be determined.

3.2 Monoubiquitylation Effects SPRTN Activity

Another interesting question is how SPRTN is recruited to damaged DNA sites. Affinity towards different DNA structures alone seems unlikely because the enzyme binds non-activating and activating DNA structures similarly *in vitro (*Reinking et al., 2020b*)*. Previous observations proposed a role of a ubiquitin switch controlling chromatin access (Stingele et al., 2016). SPRTN contains a flexible C-terminal tail with protein-protein interacting motives and is constitutively monoubiquitylated (Centore et al., 2012; Davis et al., 2012; Mosbech et al., 2012). Upon treatment with DPC-inducing agents, the modification is removed by a deubiquitylation enzyme (Stingele et al., 2016). It was speculated that this deubiquitylation enables recruitment to chromatin. Interestingly, in the second study of this thesis non-ubiquitylated and monoubiquitylated SPRTN were both found on chromatin, leading to the conclusion that monoubiquitylation is not regulating recruitment. Instead, we could show that monoubiquitylation of SPRTN contributes to its inactivation, thereby controlling its catalytic activity towards DPCs. The first inactivation mechanisms is priming the enzyme for polyubiquitylation and thereby making it a target for proteasomal degradation. The second mechanism is the increased autocatalytic cleavage activity of

monoubiquitylated SPRTN *in trans*, leading to a decrease in the active full-length SPRTN. We propose the following model for the control of SPRTN´s activity: in unchallenged condition SPRTN is constantly monoubiquitylated. This modification is stabilized by SPRTN´s own UBZ domain by shielding it from deubiquitylating enzymes. Since no increased DNA damage is detectable, SPRTN is inactivated by self-proteolysis or proteolysis by the proteasome. Once cells are treated with DPC-inducing agents, the monoubiquitin is removed from the enzyme by the deubiquitylating enzyme USP7, identified during this study. This slows down SPRTN depletion and enables the enzyme to degrade toxic DPCs. Since it was shown that SPRTN is only active in the presence of DNA *in vitro*, the ubiquitylation state would only affect the enzyme when it is already recruited to chromatin. How this recruitment is achieved still not understood but some speculations can be made.

Formaldehyde treatment leads to wide-spread ubiquitylation events in cells (Ortega-Atienza et al., 2016). An interesting assumption would involve SPRTN interaction with ubiquitin signals via its UBZ domain. By engaging with other ubiquitin proteins, the own monoubiquitin is exposed for removal by USP7 and the enzyme is stabilized on DNA. It was shown that SPRTN can degrade non-ubiquitylated proteins, which rules out the fact that the DPC itself is the only interacting partner, which could recruit the enzyme to chromatin (Larsen et al., 2019). However, it is possible that SPRTN engages with other ubiquitylated DNA repair components. Indeed, some studies showed that SPRTN recognized ubiquitylated PCNA after UV-damage and this interaction depended on its PIP-box and UBZ-domain (Centore et al., 2012; Juhasz et al., 2012). Moreover, the interaction with p97 via its SHP box was shown to recruit p97 to sites of DNA damage (Davis et al., 2012; Mosbech et al., 2012). However, opposing to these observations was a recent study in *X.laevis* egg extract, which indicated that PCNA and p97 binding are dispensable for DPC degradation, whereas the UBZ domain is essential (Larsen et al., 2019). Furthermore, SPRTN-ΔC, which is found in patients suffering from RJALS and leaks the C-terminal tail is still partially functional (Maskey et al., 2014; Stingele et al., 2016). Therefore, it still remains an interesting question how DPCs are sensed in the first place and which domains of SPRTN enable its chromatin recruitment.

Another interesting question that arises is how ubiquitin actually influences SPRTN´s activity on a structural level. In this thesis this question was addressed by an *in vitro* analysis of a SPRTN variant, which contains a ubiquitin fusion (SPRTN-Ub^{LF}, omitting the two C-

terminal glycines) (Zhao et al., 2021). This fusion was previously shown to mimic monoubiquitylated SPRTN (Mosbech et al., 2012). SPRTN-WT and SPRTN-Ub^{LF} were tested in autocleavage and substrate cleavage assays using histone H1 and the model DPC-substrates described in the first part of this thesis. SPRTN-WT and -Ub^{LF} were both active in the presence of DNA, indicating that SPRTN does not strictly require the ubiquitin modification for activity. In addition, it was shown that SPRTN-Ub^{LF} still retains its DNA-structure specificity. Interestingly, while SPRTN WT and SPRTN-Ub^{LF} showed similar activity towards its substrates, autocleavage *in trans* seemed to be affected by the modification. SPRTN-Ub^{LF} showed enhanced autocleavage activity when compared to SPRTN-WT. This increased catalytic activity was independent of the ubiquitin-status of the second SPRTN molecule that was degraded and seemed to entirely depend on the active enzyme itself. From this observation it can be concluded that monoubiquitin induces conformational changes within the enzyme resulting in increased activity. On the other side, monoubiquitylation could also influence the affinity of one SPRTN molecule for a second, thereby enhancing autocleavage activity. In line with this, other studies suggested that SPRTN forms a dimer on DNA, which could explain the efficient autocatalytic cleavage activity (Li et al., 2019). However, we could not detect such a dimer formation on DNA during our own investigations. In summary, the question how ubiquitin is changing SPRTN activity is still not resolved and it would be interesting to further understand how this modification affects the structure of the SPRTN molecule.

Taken together, SPRTN is a tightly controlled protease, which protects cells from toxic DPC lesions. Its broad substrate specificity makes it a useful tool to target a large diversity of DPCs. However, the recent discoveries of additional proteases involved in DPC repair raise the possibility that SPRTN is not alone in this fight and that other mechanisms exists that conquer those damages.

4 References

Agostinho, M., Santos, V., Ferreira, F., Costa, R., Cardoso, J., Pinheiro, I., Rino, J., Jaffray, E., Hay, R.T., and Ferreira, J. (2008). Conjugation of human topoisomerase 2 alpha with small ubiquitin-like modifiers 2/3 in response to topoisomerase inhibitors: cell cycle stage and chromosome domain specificity. Cancer Res *68*, 2409-2418.

Alabert, C., Bukowski-Wills, J.C., Lee, S.B., Kustatscher, G., Nakamura, K., de Lima Alves, F., Menard, P., Mejlvang, J., Rappsilber, J., and Groth, A. (2014). Nascent chromatin capture proteomics determines chromatin dynamics during DNA replication and identifies unknown fork components. Nat Cell Biol *16*, 281-291.

Aparicio, T., Baer, R., and Gautier, J. (2014). DNA double-strand break repair pathway choice and cancer. DNA Repair (Amst) *19*, 169-175.

Aparicio, T., Baer, R., Gottesman, M., and Gautier, J. (2016). MRN, CtIP, and BRCA1 mediate repair of topoisomerase II-DNA adducts. J Cell Biol *212*, 399-408.

Auerbach, A.D. (2009). Fanconi anemia and its diagnosis. Mutat Res *668*, 4-10.

Azuma, Y., Arnaoutov, A., and Dasso, M. (2003). SUMO-2/3 regulates topoisomerase II in mitosis. J Cell Biol *163*, 477-487.

Baker, D.J., Wuenschell, G., Xia, L., Termini, J., Bates, S.E., Riggs, A.D., and O'Connor, T.R. (2007). Nucleotide excision repair eliminates unique DNA-protein cross-links from mammalian cells. J Biol Chem *282*, 22592-22604.

Barker, S., Weinfeld, M., and Murray, D. (2005a). DNA-protein crosslinks: their induction, repair, and biological consequences. Mutat Res *589*, 111-135.

Barker, S., Weinfeld, M., Zheng, J., Li, L., and Murray, D. (2005b). Identification of mammalian proteins crosslinked to DNA by ionizing radiation. J Biol Chem *280*, 33826-33838.

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

Bergink, S., Ammon, T., Kern, M., Schermelleh, L., Leonhardt, H., and Jentsch, S. (2013). Role of Cdc48/p97 as a SUMO-targeted segregase curbing Rad51-Rad52 interaction. Nat Cell Biol *15*, 526-532.

Bhargava, V., Goldstein, C.D., 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 e10.

Biggins, S., Bhalla, N., Chang, A., Smith, D.L., and Murray, A.W. (2001). Genes involved in sister chromatid separation and segregation in the budding yeast Saccharomyces cerevisiae. Genetics *159*, 453-470.

Bodnar, N.O., and Rapoport, T.A. (2017). Molecular Mechanism of Substrate Processing by the Cdc48 ATPase Complex. Cell *169*, 722-735 e729.

Caldecott, K.W. (2008). Single-strand break repair and genetic disease. Nat Rev Genet *9*, 619-631.

Cannavo, E., and Cejka, P. (2014). Sae2 promotes dsDNA endonuclease activity within Mre11-Rad50-Xrs2 to resect DNA breaks. Nature *514*, 122-125.

Carmell, M.A., Dokshin, G.A., Skaletsky, H., Hu, Y.C., van Wolfswinkel, J.C., Igarashi, K.J., Bellott, D.W., Nefedov, M., Reddien, P.W., Enders, G.C., et al. (2016). A widely employed germ cell marker is an ancient disordered protein with reproductive functions in diverse eukaryotes. Elife *5*, e19993.

Ceccaldi, R., Sarangi, P., and D'Andrea, A.D. (2016). The Fanconi anaemia pathway: new players and new functions. Nat Rev Mol Cell Biol *17*, 337-349.

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

Champoux, J.J., and Dulbecco, R. (1972). An activity from mammalian cells that untwists superhelical DNA - A possible swivel for DNA replication. Proc Natl Acad Sci U S A *69*, 143-146.

Chan, K., Resnick, M.A., and Gordenin, D.A. (2013). The choice of nucleotide inserted opposite abasic sites formed within chromosomal DNA reveals the polymerase activities participating in translesion DNA synthesis. DNA Repair (Amst) *12*, 878-889.

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

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

Chen, L., MacMillan, A.M., Chang, W., Ezaz-Nikpay, K., Lane, W.S., and Verdine, G.L. (1991). Direct identification of the active-site nucleophile in a DNA (cytosine-5)-methyltransferase. Biochemistry *30*, 11018–11025.

Chesner, L.N., and Campbell, C. (2018). 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.

Chiang, S.C., Carroll, J., and El-Khamisy, S.F. (2010). TDP1 serine 81 promotes interaction with DNA ligase IIIalpha and facilitates cell survival following DNA damage. Cell Cycle *9*, 588-595.

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

Connelly, J.C., and de Leau, E.S.L., D. R. (1999). DNA cleavage and degradation by the SbcCD protein complex from Escherichia coli. Nucleic Acids Res *27*, 1039–1046.

Connelly, J.C., and Leach, D.R. (1996). The sbcC and sbcD genes of Escherichia coli encode a nuclease involved in palindrome inviability and genetic recombination. Genes Cells *1*, 285-291.

Cooke, M.S., Evans, M.D., Dizdaroglu, M., and Lunec, J. (2003). Oxidative DNA damage: mechanisms, mutation, and disease. The FASEB Journal *17*, 1195-1214.

Das, B.B., Antony, S., Gupta, S., Dexheimer, T.S., Redon, C.E., Garfield, S., Shiloh, Y., and Pommier, Y. (2009). Optimal function of the DNA repair enzyme TDP1 requires its phosphorylation by ATM and/or DNA-PK. EMBO J *28*, 3667-3680.

Das, B.B., Huang, S.Y., Murai, J., Rehman, I., Ame, J.C., Sengupta, S., Das, S.K., Majumdar, P., Zhang, H., Biard, D., et al. (2014). PARP1-TDP1 coupling for the repair of topoisomerase I-induced DNA damage. Nucleic Acids Res *42*, 4435-4449.

Dasari, S., and Tchounwou, P.B. (2014). Cisplatin in cancer therapy: molecular mechanisms of action. Eur J Pharmacol *740*, 364-378.

Davies, R.J.H. (1995). Ultraviolet radiation damage in DNA. Biochem Soc Trans *23*, 407–418.

Davis, E.J., Lachaud, C., Appleton, P., Macartney, T.J., Nathke, I., and Rouse, J. (2012). DVC1 (C1orf124) recruits the p97 protein segregase to sites of DNA damage. Nat Struct Mol Biol *19*, 1093-1100.

Debéthune, L., Kohlhagen, G., Grandas, A., and Pommier, Y. (2002). Processing of nucleopeptides mimicking the topoisomerase I–DNA covalent complex by tyrosyl-DNA phosphodiesterase. Nucleic Acids Res *30*, 1198-1204.

DeMott, M.S., Beyret, E., Wong, D., Bales, B.C., Hwang, J.T., Greenberg, M.M., and Demple, B. (2002). Covalent trapping of human DNA polymerase beta by the oxidative DNA lesion 2-deoxyribonolactone. J Biol Chem *277*, 7637-7640.

Desai, S.D., Liu, L.F., Vazquez-Abad, D., and D'Arpa, P. (1997). Ubiquitin-dependent destruction of topoisomerase I is stimulated by the antitumor drug camptothecin. J Biol Chem *272*, 24159-24164.

Deshpande, R.A., Lee, J.H., Arora, S., and Paull, T.T. (2016). Nbs1 Converts the Human Mre11/Rad50 Nuclease Complex into an Endo/Exonuclease Machine Specific for Protein-DNA Adducts. Mol Cell *64*, 593-606.

Desouky, O., Ding, N., and Zhou, G. (2015). Targeted and non-targeted effects of ionizing radiation. Journal of Radiation Research and Applied Sciences *8*, 247-254.

Dokshin, G.A., Davis, G.M., Sawle, A.D., Eldridge, M.D., Nicholls, P.K., Gourley, T.E., Romer, K.A., Molesworth, L.W., Tatnell, H.R., Ozturk, A.R., et al. (2020). GCNA Interacts with Spartan and Topoisomerase II to Regulate Genome Stability. Dev Cell *52*, 53-68 e56.

Drablos, F., Feyzi, E., Aas, P.A., Vaagbo, C.B., Kavli, B., Bratlie, M.S., Pena-Diaz, J., Otterlei, M., Slupphaug, G., and Krokan, H.E. (2004). Alkylation damage in DNA and RNA--repair mechanisms and medical significance. DNA Repair (Amst) *3*, 1389-1407.

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

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

Emadi, A., Jones, R.J., and Brodsky, R.A. (2009). Cyclophosphamide and cancer: golden anniversary. Nat Rev Clin Oncol *6*, 638-647.

Enderle, J., Dorn, A., Beying, N., Trapp, O., and Puchta, H. (2019a). The Protease WSS1A, the Endonuclease MUS81, and the Phosphodiesterase TDP1 Are Involved in Independent Pathways of DNA-protein Crosslink Repair in Plants. Plant Cell *31*, 775-790.

Enderle, J., Dorn, A., and Puchta, H. (2019b). DNA- and DNA-Protein-Crosslink Repair in Plants. Int J Mol Sci *20*, 4304.

Evans, E., Moggs, J.G., Hwang, J.R., Egly, J.M., and Wood, R.D. (1997). Mechanism of open complex and dual incision formation by human nucleotide excision repair factors. EMBO J *16*, 6559-6573.

Fagbemi, A.F., Orelli, B., and Scharer, O.D. (2011). Regulation of endonuclease activity in human nucleotide excision repair. DNA Repair (Amst) *10*, 722-729.

Fielden, J., Wiseman, K., Torrecilla, I., Li, S., Hume, S., Chiang, S.C., Ruggiano, A., Narayan Singh, A., Freire, R., Hassanieh, S., et al. (2020). TEX264 coordinates p97- and SPRTN-mediated resolution of topoisomerase 1-DNA adducts. Nat Commun *11*, 1274.

Fine, D.A., Rozenblatt-Rosen, O., Padi, M., Korkhin, A., James, R.L., Adelmant, G., Yoon, R., Guo, L., Berrios, C., Zhang, Y., et al. (2012). Identification of FAM111A as an SV40 host range restriction and adenovirus helper factor. PLoS Pathog *8*, e1002949.

Fu, D., Calvo, J.A., and Samson, L.D. (2012). Balancing repair and tolerance of DNA damage caused by alkylating agents. Nat Rev Cancer *12*, 104-120.

Fu, Y.V., Yardimci, H., Long, D.T., Ho, T.V., Guainazzi, A., Bermudez, V.P., Hurwitz, J., van Oijen, A., Scharer, O.D., and Walter, J.C. (2011). Selective bypass of a lagging strand roadblock by the eukaryotic replicative DNA helicase. Cell *146*, 931-941.

Fujimoto, H., Pinak, M., Nemoto, T., and Bunta, J. (2007). Structural analysis of base mispairing in DNA containing oxidative guanine lesion. Centr Eur J Phys *5*, 49-61.

Gallina, I., Hendriks, I.A., Hoffmann, S., Larsen, N.B., Johansen, J., Colding-Christensen, C.S., Schubert, L., Selles-Baiget, S., Fabian, Z., Kuhbacher, U., et al. (2021). The ubiquitin ligase RFWD3 is required for translesion DNA synthesis. Mol Cell *81*, 442-458 e449.

Gao, R., Schellenberg, M.J., Huang, S.Y., Abdelmalak, M., Marchand, C., Nitiss, K.C., Nitiss, J.L., Williams, R.S., and Pommier, Y. (2014). Proteolytic degradation of topoisomerase II (Top2) enables the processing of Top2.DNA and Top2.RNA covalent complexes by tyrosyl-DNA-phosphodiesterase 2 (TDP2). J Biol Chem *289*, 17960-17969.

Ghosal, G., Leung, J.W., Nair, B.C., Fong, K.W., and Chen, J. (2012). Proliferating cell nuclear antigen (PCNA)-binding protein C1orf124 is a regulator of translesion synthesis. J Biol Chem *287*, 34225-34233.

Gomez-Herreros, F., Romero-Granados, R., Zeng, Z., Alvarez-Quilon, A., Quintero, C., Ju, L., Umans, L., Vermeire, L., Huylebroeck, D., Caldecott, K.W., and Cortes-Ledesma, F. (2013). TDP2-dependent non-homologous end-joining protects against topoisomerase II-induced DNA breaks and genome instability in cells and in vivo. PLoS Genet *9*, e1003226.

Goodarzi, A.A., Yu, Y., Riballo, E., Douglas, P., Walker, S.A., Ye, R., Harer, C., Marchetti, C., Morrice, N., Jeggo, P.A., and Lees-Miller, S.P. (2006). DNA-PK autophosphorylation facilitates Artemis endonuclease activity. EMBO J *25*, 3880-3889.

Groehler, A.t., Degner, A., and Tretyakova, N.Y. (2017). Mass Spectrometry-Based Tools to Characterize DNA-Protein Cross-Linking by Bis-Electrophiles. Basic Clin Pharmacol Toxicol *121 Suppl 3*, 63-77.

Gujar, H., Weisenberger, D.J., and Liang, G. (2019). The Roles of Human DNA Methyltransferases and Their Isoforms in Shaping the Epigenome. Genes (Basel) *10*, 172.

Hartsuiker, E., Mizuno, K., Molnar, M., Kohli, J., Ohta, K., and Carr, A.M. (2009). Ctp1CtIP and Rad32Mre11 nuclease activity are required for Rec12Spo11 removal, but Rec12Spo11 removal is dispensable for other MRN-dependent meiotic functions. Mol Cell Biol *29*, 1671-1681.

Hoa, N.N., Shimizu, T., Zhou, Z.W., Wang, Z.Q., Deshpande, R.A., Paull, T.T., Akter, S., Tsuda, M., Furuta, R., Tsutsui, K., et al. (2016). Mre11 Is Essential for the Removal of Lethal Topoisomerase 2 Covalent Cleavage Complexes. Mol Cell *64*, 580-592.

Hoffmann, S., Pentakota, S., Mund, A., Haahr, P., Coscia, F., Gallo, M., Mann, M., Taylor, N.M., and Mailand, N. (2020). FAM111 protease activity undermines cellular fitness and is amplified by gain-of-function mutations in human disease. EMBO Rep *21*, e50662.

Hudson, J.J., Chiang, S.C., Wells, O.S., Rookyard, C., and El-Khamisy, S.F. (2012). SUMO modification of the neuroprotective protein TDP1 facilitates chromosomal single-strand break repair. Nat Commun *3*, 733.

Ide, H., Nakano, T., Salem, A.M.H., and Shoulkamy, M.I. (2018). DNA-protein cross-links: Formidable challenges to maintaining genome integrity. DNA Repair (Amst) *71*, 190-197. 10.1016/j.dnarep.2018.08.024.

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

Jackson, S.P., and Bartek, J. (2009). The DNA-damage response in human biology and disease. Nature *461*, 1071-1078.

Ji, S., Park, D., Kropachev, K., Kolbanovskiy, M., Fu, I., Broyde, S., Essawy, M., Geacintov, N.E., and Tretyakova, N.Y. (2019). 5-Formylcytosine-induced DNA-peptide cross-links reduce transcription efficiency, but do not cause transcription errors in human cells. J Biol Chem *294*, 18387-18397.

Jiricny, J. (2013). Postreplicative mismatch repair. Cold Spring Harb Perspect Biol *5*, a012633.

Juhasz, S., Balogh, D., Hajdu, I., Burkovics, P., Villamil, M.A., Zhuang, Z., and Haracska, L. (2012). Characterization of human Spartan/C1orf124, an ubiquitin-PCNA interacting regulator of DNA damage tolerance. Nucleic Acids Res *40*, 10795-10808.

Kashammer, L., Saathoff, J.H., Lammens, K., Gut, F., Bartho, J., Alt, A., Kessler, B., and Hopfner, K.P. (2019). Mechanism of DNA End Sensing and Processing by the Mre11-Rad50 Complex. Mol Cell *76*, 382-394.e386.

Kasparkova, J., Novakova, O., Vrana, O., Intini, F., Natile, G., and Brabec, V. (2006). Molecular aspects of antitumor effects of a new platinum(IV) drug. Mol Pharmacol *70*, 1708-1719.

Keeney, S., Giroux, C.N., and Kleckner, N. (1997). Meiosis-Specific DNA Double-Strand Breaks Are Catalyzed by Spo11, a Member of a Widely Conserved Protein Family. Cell *88*, 375-384.

Kelland, L. (2007). The resurgence of platinum-based cancer chemotherapy. Nat Rev Cancer *7*, 573-584.

Khodyreva, S.N., Prasad, R., Ilina, E.S., Sukhanova, M.V., Kutuzov, M.M., Liu, Y., Hou, E.W., Wilson, S.H., and Lavrik, O.I. (2010). Apurinic/apyrimidinic (AP) site recognition by the 5'-dRP/AP lyase in poly(ADP-ribose) polymerase-1 (PARP-1). Proc Natl Acad Sci U S A *107*, 22090-22095.

Kojima, Y., Machida, Y., Palani, S., Caulfield, T.R., Radisky, E.S., Kaufmann, S.H., and Machida, Y.J. (2020). FAM111A protects replication forks from protein obstacles via its trypsin-like domain. Nat Commun *11*, 1318.

Kottemann, M.C., Conti, B.A., Lach, F.P., and Smogorzewska, A. (2018). Removal of RTF2 from Stalled Replisomes Promotes Maintenance of Genome Integrity. Mol Cell *69*, 24-35.e25.

Kottemann, M.C., and Smogorzewska, A. (2013). Fanconi anaemia and the repair of Watson and Crick DNA crosslinks. Nature *493*, 356-363.

Kow, Y.W. (2002). Repair of deaminated bases in DNA. Free Radic Biol Med *33*, 886-893.

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

Kuhbacher, U., and Duxin, J.P. (2020). How to fix DNA-protein crosslinks. DNA Repair (Amst) *94*, 102924.

Kunkel, T.A. (2004). DNA replication fidelity. J Biol Chem *279*, 16895-16898.

Kunkel, T.A. (2009). Evolving views of DNA replication (in)fidelity. Cold Spring Harb Symp Quant Biol *74*, 91–101.

Kunkel, T.A. (2011). Balancing eukaryotic replication asymmetry with replication fidelity. Curr Opin Chem Biol *15*, 620-626.

Kurtz, A.J., and Lloyd, R.S. (2003). 1,N2-deoxyguanosine adducts of acrolein, crotonaldehyde, and trans-4-hydroxynonenal cross-link to peptides via Schiff base linkage. J Biol Chem *278*, 5970-5976.

Larsen, N.B., Gao, A.O., Sparks, J.L., Gallina, I., Wu, R.A., Mann, M., Raschle, M., Walter, J.C., and Duxin, J.P. (2019). Replication-Coupled DNA-Protein Crosslink Repair by SPRTN and the Proteasome in Xenopus Egg Extracts. Mol Cell *73*, 574-588.e577.

Lee, K.C., Padget, K., Curtis, H., Cowell, I.G., Moiani, D., Sondka, Z., Morris, N.J., Jackson, G.H., Cockell, S.J., Tainer, J.A., and Austin, C.A. (2012). MRE11 facilitates the removal of human topoisomerase II complexes from genomic DNA. Biol Open *1*, 863-873.

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

Li, F., Raczynska, J.E., Chen, Z., and Yu, H. (2019). Structural Insight into DNA-Dependent Activation of Human Metalloprotease Spartan. Cell Rep *26*, 3336-3346.e3334.

Lieber, M.R. (2010). The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. Annu Rev Biochem *79*, 181-211.

Lim, C.T., Lai, P.J., Leach, D.R., Maki, H., and Furukohri, A. (2015). A novel mode of nuclease action is revealed by the bacterial Mre11/Rad50 complex. Nucleic Acids Res *43*, 9804-9816.

Lin, C.P., Ban, Y., Lyu, Y.L., Desai, S.D., and Liu, L.F. (2008). A ubiquitin-proteasome pathway for the repair of topoisomerase I-DNA covalent complexes. J Biol Chem *283*, 21074-21083.

Lindahl, T. (1993). Instability and decay of the primary structure of DNA. Nature *362*, 709-715.

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

Loeber, R., Michaelson, E., Fang, Q., Campbell, C., Pegg, A.E., and Tretyakova, N. (2008). Cross-linking of the DNA repair protein Omicron6-alkylguanine DNA alkyltransferase to DNA in the presence of antitumor nitrogen mustards. Chem Res Toxicol *21*, 787-795.

Lopez-Mosqueda, J., Maddi, K., Prgomet, S., Kalayil, S., Marinovic-Terzic, I., Terzic, J., and Dikic, I. (2016). SPRTN is a mammalian DNA-binding metalloprotease that resolves DNA-protein crosslinks. Elife *5*, e21491.

Lord, C.J., and Ashworth, A. (2012). The DNA damage response and cancer therapy. Nature *481*, 287-294.

Lu, K., Ye, W., Zhou, L., Collins, L.B., Chen, X., Gold, A., Ball, L.M., and Swenberg, J.A. (2010). Structural characterization of formaldehyde-induced cross-links between amino acids and deoxynucleosides and their oligomers. J Am Chem Soc *132*, 3388-3399.

Luo, W., Li, H., Zhang, Y., and Ang, C.Y. (2001). Determination of formaldehyde in blood plasma by high-performance liquid chromatography with fluorescence detection. J Chromatogr B Biomed Sci Appl *753*, 253-257.

Machida, Y., Kim, M.S., and Machida, Y.J. (2012). Spartan/C1orf124 is important to prevent UV-induced mutagenesis. Cell Cycle *11*, 3395-3402.

Mao, Y., Desai, S.D., Ting, C.Y., Hwang, J., and Liu, L.F. (2001). 26 S proteasome-mediated degradation of topoisomerase II cleavable complexes. J Biol Chem *276*, 40652-40658.

Marenstein, D.R., Wilson, D.M., 3rd, and Teebor, G.W. (2004). Human AP endonuclease (APE1) demonstrates endonucleolytic activity against AP sites in single-stranded DNA. DNA Repair (Amst) *3*, 527-533.

Maskey, R.S., Flatten, K.S., Sieben, C.J., Peterson, K.L., Baker, D.J., Nam, H.J., Kim, M.S., Smyrk, T.C., Kojima, Y., Machida, Y., et al. (2017). Spartan deficiency causes accumulation of Topoisomerase 1 cleavage complexes and tumorigenesis. Nucleic Acids Res *45*, 4564-4576.

Maskey, R.S., Kim, M.S., Baker, D.J., Childs, B., Malureanu, L.A., Jeganathan, K.B., Machida, Y., van Deursen, J.M., and Machida, Y.J. (2014). Spartan deficiency causes genomic instability and progeroid phenotypes. Nat Commun *5*, 5744.

Mehta, K.P.M., Lovejoy, C.A., Zhao, R., Heintzman, D.R., and 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.

Meyn, R.E., vanAnkeren, S.C., and Jenkins, W.T. (1987). The induction of DNA-protein crosslinks in hypoxic cells and their possible contribution to cell lethality. Radiat Res *109*, 419-429.

Mimitou, E.P., Yamada, S., and Keeney, S. (2017). A global view of meiotic double-strand break end resection. Science *355*, 40-45.

Miyagi, Y., Zhang, H., and Wheeler, K.T. (1997). Radiation-induced DNA damage in tumors and normal tissues: IV. Influence of proliferation status and cell type on the formation of oxygen-dependent DNA damage in cultured cells. Radiat Res *148*, 29-34.

Mohni, K.N., Wessel, S.R., Zhao, R., Wojciechowski, A.C., Luzwick, J.W., Layden, H., Eichman, B.F., Thompson, P.S., Mehta, K.P.M., and Cortez, D. (2019). HMCES Maintains Genome Integrity by Shielding Abasic Sites in Single-Strand DNA. Cell *176*, 144-153.e113.

Morocz, M., Zsigmond, E., Toth, R., Enyedi, M.Z., Pinter, L., and Haracska, L. (2017). DNA-dependent protease activity of human Spartan facilitates replication of DNA-protein crosslink-containing DNA. Nucleic Acids Res *45*, 3172-3188.

Mosbech, A., Gibbs-Seymour, I., Kagias, K., Thorslund, T., Beli, P., Povlsen, L., Nielsen, S.V., Smedegaard, S., Sedgwick, G., Lukas, C., et al. (2012). DVC1 (C1orf124) is a DNA damage-targeting p97 adaptor that promotes ubiquitin-dependent responses to replication blocks. Nat Struct Mol Biol *19*, 1084-1092.

Murray, D., Meyn, R.E., and Vanankeren, S.C. (1988). Variations in the spectrum of lesions produced in the DNA of cells from mouse tissues after exposure to gamma-rays in air-breathing or in artificially anoxic animals. Int J Radiat Biol Relat Stud Phys Chem Med *53*, 921-933.

Nakano, T., Katafuchi, A., Matsubara, M., Terato, H., Tsuboi, T., Masuda, T., Tatsumoto, T., Pack, S.P., Makino, K., Croteau, D.L., et al. (2009). Homologous recombination but not nucleotide excision repair plays a pivotal role in tolerance of DNA-protein cross-links in mammalian cells. J Biol Chem *284*, 27065-27076.

Nakano, T., Mitsusada, Y., Salem, A.M., Shoulkamy, M.I., Sugimoto, T., Hirayama, R., Uzawa, A., Furusawa, Y., and Ide, H. (2015). Induction of DNA-protein cross-links by ionizing radiation and their elimination from the genome. Mutat Res *771*, 45-50.

Nakano, T., Miyamoto-Matsubara, M., Shoulkamy, M.I., Salem, A.M., Pack, S.P., Ishimi, Y., and Ide, H. (2013). Translocation and stability of replicative DNA helicases upon encountering DNA-protein cross-links. J Biol Chem *288*, 4649-4658.

Nakano, T., Morishita, S., Katafuchi, A., Matsubara, M., Horikawa, Y., Terato, H., Salem, A.M., Izumi, S., Pack, S.P., Makino, K., and Ide, H. (2007). Nucleotide excision repair and homologous recombination systems commit differentially to the repair of DNA-protein crosslinks. Mol Cell *28*, 147-158.

Nakano, T., Ouchi, R., Kawazoe, J., Pack, S.P., Makino, K., and Ide, H. (2012). T7 RNA polymerases backed up by covalently trapped proteins catalyze highly error prone transcription. J Biol Chem *287*, 6562-6572.

Nakano, T., Xu, X., Salem, A.M.H., Shoulkamy, M.I., and Ide, H. (2017). Radiation-induced DNA-protein cross-links: Mechanisms and biological significance. Free Radic Biol Med *107*, 136-145.

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

Nouspikel, T. (2009). DNA repair in mammalian cells : Nucleotide excision repair: variations on versatility. Cell Mol Life Sci *66*, 994-1009.

Nowicka, U., Zhang, D., Walker, O., Krutauz, D., Castaneda, C.A., Chaturvedi, A., Chen, T.Y., Reis, N., Glickman, M.H., and Fushman, D. (2015). DNA-damage-inducible 1 protein (Ddi1) contains an uncharacteristic ubiquitin-like domain that binds ubiquitin. Structure *23*, 542-557.

Ogi, T., Limsirichaikul, S., Overmeer, R.M., Volker, M., Takenaka, K., Cloney, R., Nakazawa, Y., Niimi, A., Miki, Y., Jaspers, N.G., et al. (2010). Three DNA polymerases, recruited by different mechanisms, carry out NER repair synthesis in human cells. Mol Cell *37*, 714-727.

Orta, M.L., Calderon-Montano, J.M., Dominguez, I., Pastor, N., Burgos-Moron, E., Lopez-Lazaro, M., Cortes, F., Mateos, S., and Helleday, T. (2013). 5-Aza-2'-deoxycytidine causes replication lesions that require Fanconi anemia-dependent homologous recombination for repair. Nucleic Acids Res *41*, 5827-5836.

Ortega-Atienza, S., Rubis, B., McCarthy, C., and Zhitkovich, A. (2016). Formaldehyde Is a Potent Proteotoxic Stressor Causing Rapid Heat Shock Transcription Factor 1 Activation and Lys48-Linked Polyubiquitination of Proteins. Am J Pathol *186*, 2857-2868.

Paiano, J., Wu, W., Yamada, S., Sciascia, N., Callen, E., Paola Cotrim, A., Deshpande, R.A., Maman, Y., Day, A., Paull, T.T., and Nussenzweig, A. (2020). ATM and PRDM9 regulate SPO11-bound recombination intermediates during meiosis. Nat Commun *11*, 857.

Pecina-Slaus, N., Kafka, A., Salamon, I., and Bukovac, A. (2020). Mismatch Repair Pathway, Genome Stability and Cancer. Front Mol Biosci *7*, 122.

Plo, I., Liao, Z.Y., Barcelo, J.M.K., G., Caldecott, K.W., and Weinfeld, M.P., Y. (2003). Association of XRCC1 and tyrosyl DNA phosphodiesterase (Tdp1) for the repair of topoisomerase I-mediated DNA lesions. DNA Repair (Amst) *2*, 1087-1100.

Pommier, Y. (2006). Topoisomerase I inhibitors: camptothecins and beyond. Nat Rev Cancer *6*, 789-802.

Pommier, Y., Huang, S.Y., Gao, R., Das, B.B., Murai, J., and Marchand, C. (2014). Tyrosyl-DNA-phosphodiesterases (TDP1 and TDP2). DNA Repair (Amst) *19*, 114-129.

Pommier, Y., and Marchand, C. (2011). Interfacial inhibitors: targeting macromolecular complexes. Nat Rev Drug Discov *11*, 25-36.

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

Pourquier, P., Ueng, L.M., Kohlhagen, G., Mazumder, A., Gupta, M., Kohn, K.W., and Pommier, Y. (1997). Effects of uracil incorporation, DNA mismatches, and abasic sites on cleavage and religation activities of mammalian topoisomerase I. J Biol Chem *272*, 7792-7796.

Prasad, R., Horton, J.K., Chastain, P.D., 2nd, Gassman, N.R., Freudenthal, B.D., Hou, E.W., and Wilson, S.H. (2014). Suicidal cross-linking of PARP-1 to AP site intermediates in cells undergoing base excision repair. Nucleic Acids Res *42*, 6337-6351.

Prasad, R., Horton, J.K., Dai, D.P., and Wilson, S.H. (2019). Repair pathway for PARP-1 DNA-protein crosslinks. DNA Repair (Amst) *73*, 71-77.

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

Quinones, J.L., Thapar, U., Yu, K., Fang, Q., Sobol, R.W., and Demple, B. (2015). Enzyme mechanism-based, oxidative DNA-protein cross-links formed with DNA polymerase beta in vivo. Proc Natl Acad Sci U S A *112*, 8602-8607.

Rageul, J., and Kim, H. (2020). Fanconi anemia and the underlying causes of genomic instability. Environ Mol Mutagen *61*, 693-708.

Rao, T., Gao, R., Takada, S., Al Abo, M., Chen, X., Walters, K.J., Pommier, Y., and Aihara, H. (2016). Novel TDP2-ubiquitin interactions and their importance for the repair of topoisomerase II-mediated DNA damage. Nucleic Acids Res *44*, 10201-10215.

Reardon, J.T., Cheng, Y., and Sancar, A. (2006). Repair of DNA-protein cross-links in mammalian cells. Cell Cycle *5*, 1366-1370.

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

Reinking, H.K., Kang, H.S., Gotz, M.J., Li, H.Y., Kieser, A., Zhao, S., Acampora, A.C., Weickert, P., Fessler, E., Jae, L.T., et al. (2020b). DNA Structure-Specific Cleavage of DNA-Protein Crosslinks by the SPRTN Protease. Mol Cell *80*, 102-113.e106.

Renkawitz, J., Lademann, C.A., and Jentsch, S. (2014). Mechanisms and principles of homology search during recombination. Nat Rev Mol Cell Biol *15*, 369-383. 10.1038/nrm3805.

Renkawitz, J., Lademann, C.A., Kalocsay, M., and Jentsch, S. (2013). Monitoring homology search during DNA double-strand break repair in vivo. Mol Cell *50*, 261-272.

Ridpath, J.R., Nakamura, A., Tano, K., Luke, A.M., Sonoda, E., Arakawa, H., Buerstedde, J.M., Gillespie, D.A., Sale, J.E., Yamazoe, M., et al. (2007). Cells deficient in the FANC/BRCA pathway are hypersensitive to plasma levels of formaldehyde. Cancer Res *67*, 11117-11122.

Robert, M.F., Morin, S., Beaulieu, N., Gauthier, F., Chute, I.C., Barsalou, A., and MacLeod, A.R. (2003). DNMT1 is required to maintain CpG methylation and aberrant gene silencing in human cancer cells. Nat Genet *33*, 61-65.

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

Ruijs, M.W., van Andel, R.N., Oshima, J., Madan, K., Nieuwint, A.W., and Aalfs, C.M. (2003). Atypical progeroid syndrome: an unknown helicase gene defect? Am J Med Genet A *116A*, 295-299.

Rupnik, A., Grenon, M., and Lowndes, N. (2008). The MRN complex. Curr Biol *18*, R455-457.

Rydberg, B., and Lindahl, T. (1982). Nonenzymatic methylation of DNA by the intracellular methyl group donor S-adenosyl-L-methionine is a potentially mutagenic reaction. EMBO J *1*, 211-216.

Ryu, H., Furuta, M., Kirkpatrick, D., Gygi, S.P., and Azuma, Y. (2010). PIASy-dependent SUMOylation regulates DNA topoisomerase IIalpha activity. J Cell Biol *191*, 783-794.

Saathoff, J.H., Kashammer, L., Lammens, K., Byrne, R.T., and Hopfner, K.P. (2018). The bacterial Mre11-Rad50 homolog SbcCD cleaves opposing strands of DNA by two chemically distinct nuclease reactions. Nucleic Acids Res *46*, 11303-11314.

Sako, M., Inagaki, S., Esaka, Y., and Deyashiki, Y. (2003). Histones accelerate the cyclic 1, N 2 -propanoguanine adduct-formation of DNA by the primary metabolite of alcohol and carcinogenic crotonaldehyde. Bioorg Med Chem Lett *13*, 3497-3498.

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

Sartori, A.A., Lukas, C., Coates, J., Mistrik, M., Fu, S., Bartek, J., Baer, R., Lukas, J., and Jackson, S.P. (2007). Human CtIP promotes DNA end resection. Nature *450*, 509-514.

Scharer, O.D. (2013). Nucleotide excision repair in eukaryotes. Cold Spring Harb Perspect Biol *5*, a012609.

Schellenberg, M.J., Appel, C.D., Riccio, A.A., Butler, L.R., Krahn, J.M., Liebermann, J.A., Cortes-Ledesma, F., and Williams, R.S. (2020). Ubiquitin stimulated reversal of topoisomerase 2 DNA-protein crosslinks by TDP2. Nucleic Acids Res *48*, 6310-6325.

Schellenberg, M.J., Lieberman, J.A., Herrero-Ruiz, A., Butler, L.R., Williams, J.G.M.-C., A.M., Mueller, G.A., London, R.E., Cortés-Ledesma, F., and Williams, R.S. (2017). ZATT (ZNF451)–mediated resolution of topoisomerase 2 DNA-protein cross-links. Science *357*, 1412-1416.

Schermelleh, L., Spada, F., Easwaran, H.P., Zolghadr, K., Margot, J.B., Cardoso, M.C., and Leonhardt, H. (2005). Trapped in action: direct visualization of DNA methyltransferase activity in living cells. Nat Methods *2*, 751-756.

Schumacher, B., Pothof, J., Vijg, J., and Hoeijmakers, J.H.J. (2021). The central role of DNA damage in the ageing process. Nature *592*, 695-703.

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

Sedgwick, B. (2004). Repairing DNA-methylation damage. Nat Rev Mol Cell Biol *5*, 148-157.

Serbyn, N., Noireterre, A., Bagdiul, I., Plank, M., Michel, A.H., Loewith, R., Kornmann, B., and Stutz, F. (2020). The Aspartic Protease Ddi1 Contributes to DNA-Protein Crosslink Repair in Yeast. Mol Cell *77*, 1066-1079.e1069.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., and Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell *119*, 941-953.

Shrivastav, N., Li, D., and Essigmann, J.M. (2010). Chemical biology of mutagenesis and DNA repair: cellular responses to DNA alkylation. Carcinogenesis *31*, 59-70.

Sparks, J.L., Chistol, G., Gao, A.O., Raschle, M., Larsen, N.B., Mann, M., Duxin, J.P., and Walter, J.C. (2019). The CMG Helicase Bypasses DNA-Protein Cross-Links to Facilitate Their Repair. Cell *176*, 167-181.e121.

Srinivas, U.S., Tan, B.W.Q., Vellayappan, B.A., and Jeyasekharan, A.D. (2019). ROS and the DNA damage response in cancer. Redox Biol *25*, 101084.

Stingele, J., Bellelli, R., Alte, F., Hewitt, G., Sarek, G., Maslen, S.L., Tsutakawa, S.E., Borg, A., Kjaer, S., Tainer, J.A., 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., and Boulton, S.J. (2017). Mechanisms of DNA-protein crosslink repair. Nat Rev Mol Cell Biol *18*, 563-573.

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

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

Stracker, T.H., and Petrini, J.H. (2011). The MRE11 complex: starting from the ends. Nat Rev Mol Cell Biol *12*, 90-103.

Sugasawa, K., Ng, J.M., Masutani, C., Iwai, S., van der Spek, P.J., Eker, A.P., Hanaoka, F., Bootsma, D., and Hoeijmakers, J.H. (1998). Xeroderma pigmentosum group C protein complex is the initiator of global genome nucleotide excision repair. Mol Cell *2*, 223-232.

Sugasawa, K., Okamoto, T., Shimizu, Y., Masutani, C., Iwai, S., and Hanaoka, F. (2001). A multistep damage recognition mechanism for global genomic nucleotide excision repair. Genes Dev *15*, 507-521.

Sun, Y., Miller Jenkins, L.M., Su, Y.P., Nitiss, K.C., Nitiss, J.L., and Pommier, Y. (2020). A conserved SUMO pathway repairs topoisomerase DNA-protein cross-links by engaging ubiquitin-mediated proteasomal degradation. Sci Adv *6*, eaba6290.

Sung, J.S., and Park, I.K. (2005). Formation of DNA-protein cross-links mediated by C1'-oxidized abasic lesion in mouse embryonic fibroblast cell-free extracts. Integrative Biosciences *9*, 79-85.

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

Symington, L.S. (2014). End resection at double-strand breaks: mechanism and regulation. Cold Spring Harb Perspect Biol *6*, a016436.

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

Toth, A., Hegedus, L., Juhasz, S., Haracska, L., and Burkovics, P. (2017). The DNA-binding box of human SPARTAN contributes to the targeting of Poleta to DNA damage sites. DNA Repair (Amst) *49*, 33-42.

Travers, A., and Muskhelishvili, G. (2015). DNA structure and function. FEBS J *282*, 2279-2295.

Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M.E., Borchers, C.H., Tempst, P., and Zhang, Y. (2006). Histone demethylation by a family of JmjC domain-containing proteins. Nature *439*, 811-816.

Tubbs, J.L., Pegg, A.E., and Tainer, J.A. (2007). DNA binding, nucleotide flipping, and the helix-turn-helix motif in base repair by O6-alkylguanine-DNA alkyltransferase and its implications for cancer chemotherapy. DNA Repair (Amst) *6*, 1100-1115.

Twomey, E.C., Ji, Z., Wales, T.E., Bodnar, N.O., Ficarro, S.B., Marto, J.A., Engen, J.R., and Rapoport, T.A. (2019). Substrate processing by the Cdc48 ATPase complex is initiated by ubiquitin unfolding. Science *365*, eaax1033.

Vaz, B., Popovic, M., Newman, J.A., Fielden, J., Aitkenhead, H., Halder, S., Singh, A.N., Vendrell, I., Fischer, R., Torrecilla, I., et al. (2016). Metalloprotease SPRTN/DVC1 Orchestrates Replication-Coupled DNA-Protein Crosslink Repair. Mol Cell *64*, 704-719.

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

Vignard, J., Mirey, G., and Salles, B. (2013). Ionizing-radiation induced DNA double-strand breaks: a direct and indirect lighting up. Radiother Oncol *108*, 362-369.

Volker, M., Mone, M.J., Karmakar, P., van Hoffen, A., Schul, W., Vermeulen, W., Hoeijmakers, J.H., van Driel, R., van Zeeland, A.A., and Mullenders, L.H. (2001). Sequential assembly of the nucleotide excision repair factors in vivo. Mol Cell *8*, 213-224.

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

Wu, J.C., and Santi, D.V. (1987). Kinetic and catalytic mechanism of HhaI methyltransferase. J Biol Chem *262*, 4778-4786.

Yang, S.W., Burgin, A.B., Huizenga, B.N., Robertson, C.A., Yao, K.C., and Nash, H.A. (1996). A eukaryotic enzyme that can disjoin dead-endcovalent complexes between DNA and type I topoisomerases. Proc Natl Acad Sci U S A *93*, 11534-11539.

Yang, X., Li, Y., Gao, Z., Li, Z., Xu, J., Wang, W., and Dong, Y. (2017). Structural analysis of Wss1 protein from saccharomyces cerevisiae. Sci Rep *7*, 8270.

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

Yu, S.L., and Lee, S.K. (2017). Ultraviolet radiation: DNA damage, repair, and human disorders. Mol Cel Toxicol *13*, 21-28.

Zhang, A., Lyu, Y.L., Lin, C.P., Zhou, N., Azarova, A.M., Wood, L.M., and Liu, L.F. (2006). A protease pathway for the repair of topoisomerase II-DNA covalent complexes. J Biol Chem *281*, 35997-36003.

Zhao, S., Kieser, A., Li, H.Y., Reinking, H.K., Weickert, P., Euteneuer, S., Yaneva, D., Acampora, A.C., Gotz, M.J., Feederle, R., and Stingele, J. (2021). A ubiquitin switch controls autocatalytic inactivation of the DNA-protein crosslink repair protease SPRTN. Nucleic Acids Res *49*, 902-915.

5 List of Abbreviations

5-aza-dC	5-Aza-2′-Deoxycytidine
5hmC	5-Hydroxymethylcytosine
ACRC	Acidic Repeat-Containing Protein
AML	Acute myeloid leukemia
AP	Apurinic/Apyrimidinic
ATM	Ataxia telangiectasia mutated
ATR	Ataxia telangiectasia and Rad3 related
BER	Base Excision Repair
BLM	Bloom syndrome protein
BR	Basic region
BRCA2	Breast Cancer 2
BSA	Bovine serum albumin
CMG	Cdc45-Mcm2-7-GINS
CtIP	C-terminal-binding protein interacting protein
DDI	DNA damage inducible
DNA	Desoxyribonucleic acid
DNA-PKs	DNA-dependent protein kinases
DNMT-1	DNA (cytosine-5)-methyltransferase 1
DPC	DNA protein crosslink
DSB	Double-strand break
dsDNA	Double-stranded DNA
EQ	Inactive SPRTN mutant (E112Q)
FA	Fanconi anaemia
FAM111A	Family with sequence similarity 111 member A
FDA	Food and drug administration
GCNA	Germ cell nuclear antigen
GFP	Green fluorescent protein
GG-NER	global genome nucleotide excision repair
HCC	Hepatocellular carcinomas
HMCES	5-Hydroxymethylcytosine Binding, ES Cell Specific
HR	Homologous recombination
ICL	Interstrand Crosslink

List of Abbreviations

IR	Ionizing radiation
MMR	Mismatch repair
MRN	Mre11, Rad50 and Nbs1
NER	Nucleotide excision repair
NHEJ	Non-homologous end joining
OH	Hydroxyl radical
PARP1	Poly (ADP-Ribose) polymerase 1
PCNA	Proliferating cell nuclear antigen
PIAS4	Protein inhibitor of activated STAT protein 4
PNKP	Polynucleotide kinase 3'-phosphatase
Polβ	Polymerase β
PSMB5	Proteasome subunit beta type-5
PSMD14	proteasome non-adenosine triphosphatase regulatory subunit 14
RJALS	Ruijs-Aalfs-Syndrome
RNF4	RING finger protein 4
ROS	Reactive oxygen species
RPA	Replication protein A
RTEL	Regulator of telomere elongation helicase 1
SAP	SAF-A/B, Acinus and PIAS
SIM	SUMO-interacting motif
SPRTN	SprT-like N-terminal domain
SSB	Single-strand break
ssDNA	Single-stranded DNA
STUbL	SUMO-targeting Ubiquitin ligase
SUMO	Small ubiquitin-like modifier
TC-NER	Transcription-coupled nucleotide excision repair
TDP1	Tyrosyl-DNA-Phosphodiesterase 1
TDP2	Tyrosyl-DNA-Phosphodiesterase 2
TLS	Translesion synthesis
TOP1	Topoisomerase 1
TOP1cc	Topoisomerase 1 cleavage complex
TOP2	Topoisomerase 2

TOP2cc	Topoisomerase 2 cleavage complex
TRAIP	TRAF-interacting protein
UV	Ultraviolet
Wss1	Weak suppressor of Smt3
WT	Wildtype
ZBD	Zinc binding domain

6 Acknowledgements

First of all, I would like to express my deepest gratitude to my supervisor Julian Stingele. I am extremely thankful for the opportunity of performing my PhD in your lab. Not only are you an inspiring scientist full of ideas, enthusiasm, and optimism (even when SPRTN kept precipitating) - you are also a wonderful supervisor, always giving me the feeling of being supported and appreciated. I thank you very much for everything!

I would like to thank my PhD commission Prof. Dr. Johannes Stigler, Prof. Dr. Karl-Peter Hopfner, Prof. Dr. Karl Duderstadt, Prof. Dr. Klaus Förstemann and PD Dr. Dietmar Martin for reading and evaluating my PhD thesis. Moreover, I want to thank the members of my thesis advisory committee Johannes Stigler und Christoph Kurat for their input and discussions on my project.

Michael and Hanso, thank you for this successful collaboration, which resulted in a nice publication. Hanso, I really appreciate how you went through our data with me and explained me the basics of NMR!

Additionally, I would like to express my gratitude to the IRTG graduate school and to Elizabeth Schroeder-Reiter. Thank you for your open ears and support at any time.

I want to specifically thank Anja Kieser (Strauss), the best technician in the world! I do not know, what I would have done without you! Despite of performing amazing experiments to help me, I really enjoyed having such a positive person around. I will truly miss working with you!

Moreover, I want to acknowledge the rest of the Stingele lab members: Denitsa**,** thank you very much for always being there for me and for your support at any level and time. I can just say: I really had a lot of fun with you in your lab. We are the best team! Pedro, thank you for always making the best (and worst) jokes in the right (and wrong) moments. You have supported me in my darkest times (known as “the Christmas party”) and could always make me laugh. The last years would not have been the same without you! Shubo, I am always impressed how you can be so efficient in the lab and at the same time so helpful and nice to everyone. I am happy we could share some very special moments together and I am

sure there will be many more to come! Ricky, no matter which question I had, you were always there to help out. I really much enjoyed working with you and I appreciate how patiently you introduced me to the cell culture! Aleida, thank you very much for all your help at any time! Max, thank you for saving me from many nervous breakdowns (especially when it comes to computer-related problems). I would not be surprised, if I see a Götz-Lab at some fancy research institute one day! Jackie, it is a pity we only got to spend my last 6 months together in the lab. However, it felt like you had been there for ages and I really loved having you “next to me”!

Monika, Sabine and Romy: I really appreciate how you always helped me out, whenever I had any questions or problems!

Finally I want to thank my friends and family for their support and encouragement during the last years: to my father who gave me the freedom to do whatever and wherever I wanted and to my mother who believed in me and cheered me up so many times. A big thank you goes to my two sisters Eli and Vera who I can always count on. Knowing you are there, is the best backup I can have in life.

Last but not least, a big thank you to Eugenio! Somebody gave me the small hint that acknowledgements should not be “too tacky, but still enough”. So here it comes: you are a wonderful person and I could not wish for a better partner (and best friend) to start a new family together. I am truly grateful to have you in my life!