50TH ANNIVERSARY

OXFORD

Isolation and detection of DNA–protein crosslinks in mammalian cells

Ignacio Torrecilla [1,†], Annamaria Ruggiano [1,†], Kostantin Kiianitsa [2], Ftoon Aljarbou[1],
Pauline Lascaux[1], Gwendoline Hoslett[1], Wei Song[1], Nancy Maizels[2,3] and Kristijan Ramadan [1,*]

[1]The MRC Weatherall Institute of Molecular Medicine, Department of Oncology, John Radcliffe Hospital, University of Oxford, Oxford, OX3 9DS, UK
[2]Department of Immunology, University of Washington, Seattle, WA 98195-7350, USA
[3]Department of Biochemistry, University of Washington, Seattle, WA 98195-7350, USA

[*]To whom correspondence should be addressed. Tel: +44 1865 222 354; Email: kristijan.ramadan@oncology.ox.ac.uk
[†]The first two authors should be regarded as Joint First Authors.

Abstract

DNA–protein crosslinks (DPCs) are toxic DNA lesions wherein a protein is covalently attached to DNA. If not rapidly repaired, DPCs create obstacles that disturb DNA replication, transcription and DNA damage repair, ultimately leading to genome instability. The persistence of DPCs is associated with premature ageing, cancer and neurodegeneration. In mammalian cells, the repair of DPCs mainly relies on the proteolytic activities of SPRTN and the 26S proteasome, complemented by other enzymes including TDP1/2 and the MRN complex, and many of the activities involved are essential, restricting genetic approaches. For many years, the study of DPC repair in mammalian cells was hindered by the lack of standardised assays, most notably assays that reliably quantified the proteins or proteolytic fragments covalently bound to DNA. Recent interest in the field has spurred the development of several biochemical methods for DPC analysis. Here, we critically analyse the latest techniques for DPC isolation and the benefits and drawbacks of each. We aim to assist researchers in selecting the most suitable isolation method for their experimental requirements and questions, and to facilitate the comparison of results across different laboratories using different approaches.

Graphical abstract
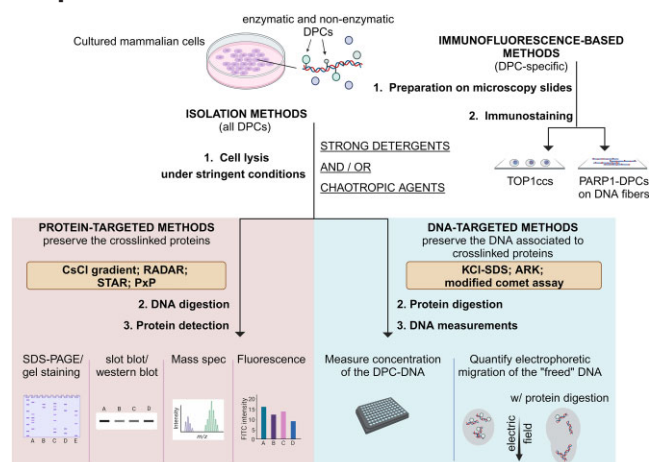
Introduction

DNA–protein crosslinks (DPCs) are common and potentially mutagenic DNA lesions that threaten genome stability if not effectively resolved. DPCs can manifest in different forms, with some being specific and transient physiological intermediates of regular enzymatic reactions. Examples include DPCs generated by topoisomerases that relax or initiate de-catenation of DNA during DNA replication and transcription (1), DPCs formed by DNA polymerase beta at oxida-tive lesions during DNA repair (2), and DPCs generated by HMCES at abasic sites (3). Specific, stable DPCs can arise from abortive topoisomerase reactions in cells treated with topoisomerase poisons (4). In contrast, non-specific DPCs arise from crosslink formation between DNA and a variety of proteins normally in proximity to DNA. Non-specific DPCs are induced by ionising radiation, environmental chemicals and metabolic by-products such as acetaldehyde and formaldehyde. Detailed insights into the nature and the repair
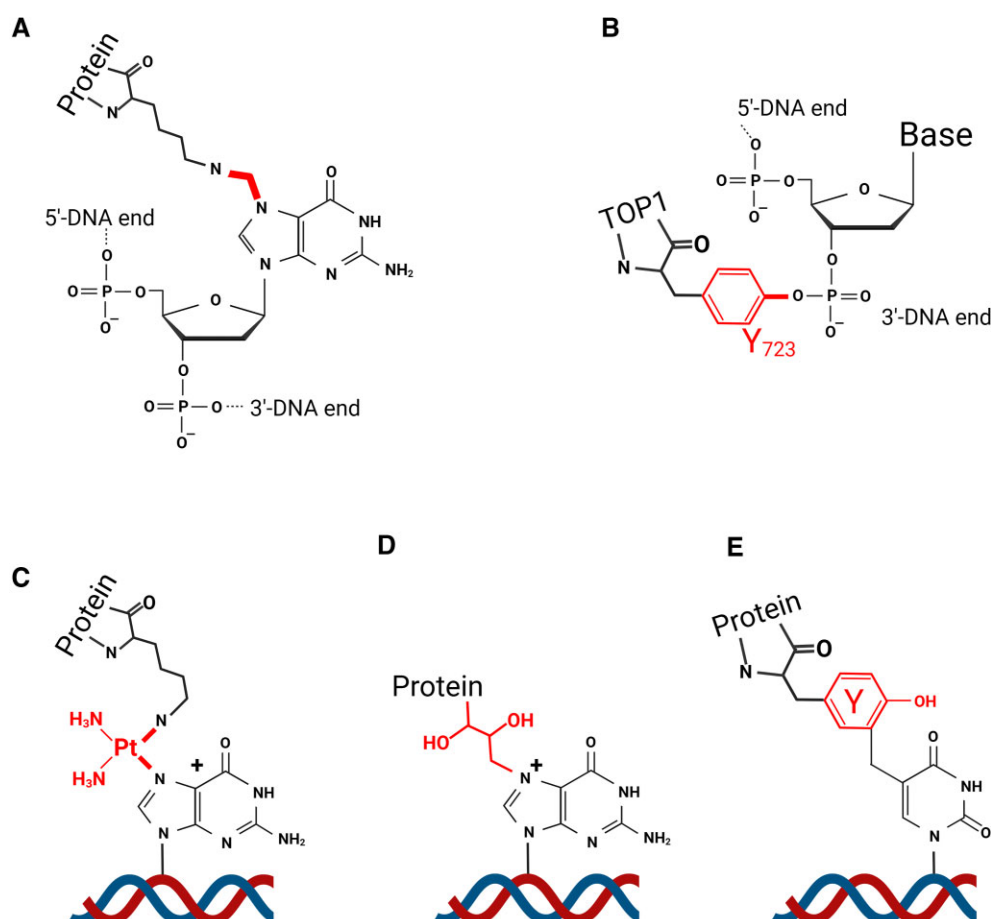
Figure 1. Examples of DNA–protein crosslinks, as detected *in vitro*. DPCs induced by reactions with exogenous crosslinkers (in red) or by endogenous enzyme: (**A**) chemical structure of formaldehyde-induced DPC linking a lysine residue with N7-guanine, highlighting the resulting methylene bridge. (**B**) Structure of covalent linkage of TOP1 Tyr[723] in the TOP1 active site to the 3′-DNA end phosphate group, highlighting the Tyr bridge in red. (**C**) cisplatin-induced DPC linking a lysine residue with N7-guanine, highlighting the cisplatin bridge in red. (**D**) 1,3-Butadiene-induced DPC on N7-guanine. (**E**) Direct crosslink by a free radical between tyrosine and the methyl group in thymine.

mechanisms of specific and non-specific DPCs can be found elsewhere (5–15).

Since DPCs block DNA transcription, replication and repair, numerous potent therapeutic strategies have been developed aimed at exploiting DPC toxicity. Fluoroquinolone antibiotics (ciprofloxacin, levofloxacin), which poison the bacterial type II topoisomerases, are used to treat a wide range of bacterial infections. Topoisomerase poisons are used in cancer therapy to create stable DPCs with TOP2 (doxorubicin, etoposide) or TOP1 (camptothecin and its derivatives topotecan, irinotecan, belotecan, and the antibody-drug conjugate trastuzumab-deruxtecan). PARP inhibitors olaparib and talazoparib, which generate tight, DPC-like complexes between PARP-1/2 and DNA, effectively induce cell death in cancer cells (16,17). Similarly, 5-aza-2′-deoxycytidine (5-aza-dC), an anti-leukemic drug that inhibits DNMT1 DNA methyltransferase, induces the formation of DPCs containing PARP1 at sites of CpG methylation (18). Ionising radiation, the most commonly used cancer treatment, is a potent inducer of DPCs, in addition to causing DNA double-strand breaks (DSB) and single-strand breaks (SSB). Furthermore, platinum-based chemotherapeutics (oxaliplatin, cisplatin) and nitrogen mustard compounds (cyclophosphamide, chlorambucil, melphalan) can induce DNA–protein crosslinks as well as DNA-DNA crosslinks (19,20).

Formation, chemical structure and biological identification of DPCs

DPCs exhibit a wide range of chemical bonds, reflecting the chemical diversity of agents that induce them and their selectivity towards specific amino acid side chains and DNA sites (Figure 1). This chemical variability has been extensively studied *in vitro* (20–26). For example, formaldehyde mediates the formation of transient *Schiff* bases with various amino acid side chains, resulting in methylene-bridge DPCs with various protein side chains (20). Platinum compounds like cisplatin preferentially target basic amino acids like arginine and lysine. Nitrogen mustards and 1,2,3,4-diepoxybutane (DEB) show a preference for cysteines (27). Not surprisingly, the N7 position on guanine is especially susceptible to DPC formation (27,28). Reactive oxygen species (ROS) and reactive nitrogen species (RNS) induce formation of DPCs between lysine and tyrosine side chains in proteins and guanine, cytosine or thymine in DNA (22,29). These diverse chemical structures confer on DPCs distinctive physicochemical properties that affect their stability and thermal lability, an important consideration when isolating and analysing DPCs.

Several research groups have made efforts to characterise the genome-wide adductome resulting from normal metabolic conditions and exposure to DPC-inducing agents
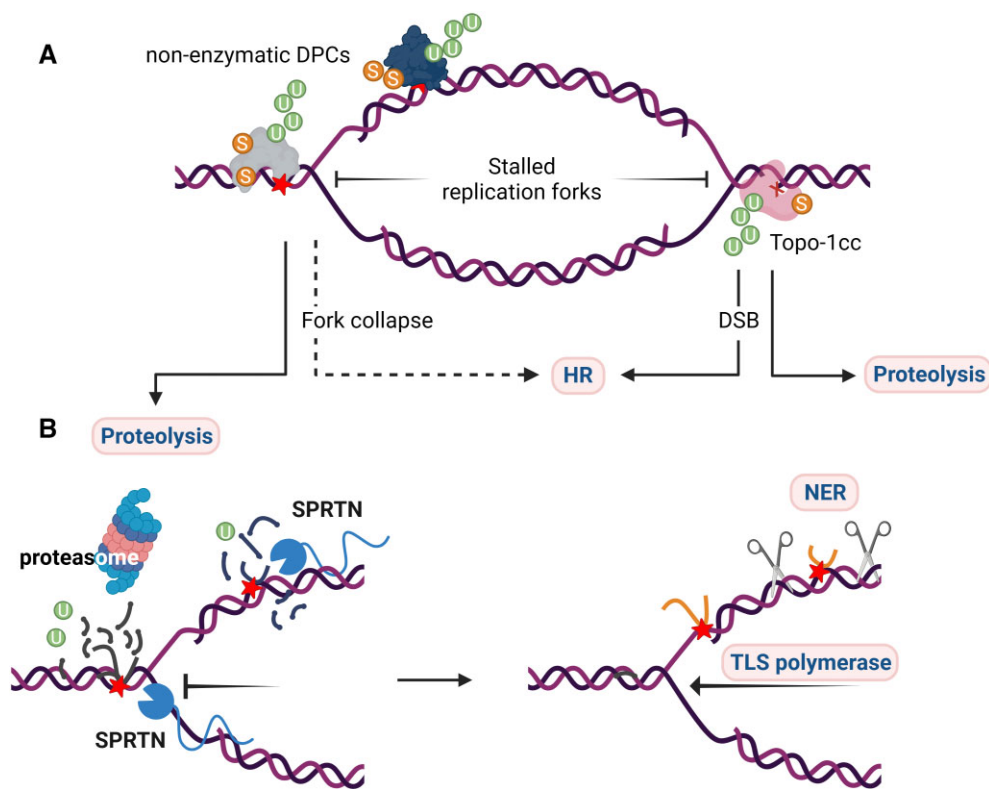
Figure 2. Replication-dependent and post-replicative DPC repair. The three main repair pathways are highlighted in pink boxes. (**A**) DPCs ahead of the replication fork can cause prolonged stalling and fork collapse, which is fixed by homologous recombination. Similarly, replication forks running into a TOP1cc result in single-ended DSBs which requires homologous recombination for repair. Processing of enzymatic and non-enzymatic DPCs by proteolysis is facilitated by ubiquitylation and SUMOylation. (**B**) SPRTN and the proteasome are the main proteolytic activities during and post-replication. The remnant peptide ahead of replication forks can be bypassed by a translesion (TLS) polymerase. Post-replicative remnant peptides are repaired by nucleases of the nucleotide excision repair (NER) pathway, which in mammalian cells can only operate on small DPCs.

(18,28,30–37). Typically, these studies involve a two-step strategy wherein DNA is initially isolated from lysed cells by a method that preserves DPCs, the DPCs are digested with proteases and the resulting peptides are subjected to the qualitative and/or quantitative compositional analysis by mass spectrometry (discussed below). The recent advancement of accurate DPC isolation methods, coupled with the considerable development of mass spectrometry-based techniques for biological samples, has provided initial insights into the DPC adductomes and has enabled the identification of proteins susceptible to crosslinking.

Repair of DPCs

We are now gaining a better understanding of the molecular mechanisms involved in DPC resolution. Two proteolytic repair pathways have emerged in recent years as essential mechanisms for coping with DPCs (Figure 2). During the initial steps of repair, the bulk of DPCs is cleaved by SPRTN (aka DVC1), the metazoan functional homolog of the yeast protease Wss1 (32,38–41). A second pathway relies on the 26S proteasome and is partly redundant with SPRTN for DPC proteolysis repair (33,38). Proteolysis by either SPRTN or the 26S proteasome leaves a remnant peptide attached to DNA, which can be bypassed by translesion polymerases during replication, and later excised by endonucleases or phosphodiesterases (*e.g.* TDP1/TDP2) (reviewed in (5,6,42). Canonical DNA repair pathways like nucleotide excision repair (NER) and homologous recombination (43–45) can also repair DPC, independently of or in concert with proteolytic repair (Figure 2).

Humans affected with Ruijls-Aalfs Syndrome (RJALS), also known as SPRTN (SPARTAN) syndrome, demonstrate that SPRTN is essential in humans (46,47). In this monogenic disease, biallelic hypomorphic mutations in the *SPRTN* gene result in the accumulation of DPCs and accelerated ageing and cancer (48). DPC proteolysis by SPRTN is crucial to the progression of DNA replication forks (32,49,50) and likely for translesion DNA synthesis (51–55). The role of SPRTN extends to replication-independent processes, though it remains restricted to the S- and G2- phases of the cell cycle (33,52). Functions and regulation of SPRTN have been reviewed elsewhere (5,56). Other human proteases (FAM111A, FAM111B, GCNA, DDI1, DDI2) can also repair DPCs but their roles are less defined (8,57,58) (reviewed in (59)).

DPC recognition and processing are aided by post-translational modifications on DPCs themselves (ubiquitin, SUMO and PAR) (38,57,59–63), DNA structures surrounding DPCs (64), physical and functional interactions with the p97 ATPase and with FANCJ (60,65,66), as well as by post-translational modifications on SPRTN (phosphorylation by CHK1, deubiquitylation by USP7, USP11 and VCIP135, and acetylation) (49,67–69). Regulation of DPC proteolysis repair by post-translational modifications has been reviewed in (70).

Table 1. Summary of DPCs isolation and detection methods, their advantages and limitations. References are not exhaustive

Method group	Technique	References	Advantages	Challenges
Protein-Targeted methods	CsCl gradient	(76–84)	Well-established technique.	Laborious and time-intensive. Requires specific instruments and reagents.
	RADAR	(3,18,32,63,65,68, 85,86,96,99)	Simple execution. Sensitive and specific.	Potential contaminations with non-crosslinked proteins and RNA-protein crosslinks.
	STAR	(31)	More stringent isolation. Removal of RNA. Enrichment of proteolytic fragments.	High protein content. Unclear improvement over RADAR.
	PxP	(33)	Able to isolate large adducts. Highly confident protein identification	Potential false negatives. Lack of scalability.
Detection of DPCs isolated by protein-targeted methods	Immuno-detection	(3,18,31–33,59,65, 68,86,95,96)	Well-established	Requires prior DPC identification. Requires good primary antibodies.
	Radio / Fluorescence labelling	(44,84,104–107)	Sensitive. Accurate quantification.	Unable to detect proteins of interest. Pre-labelling may result in potential cross-reactivity.
	Mass spectrometry	(28,31–34,37)	Sensitive and quantitative. Powerful and versatile.	Requires specialised instruments.
DNA-Targeted methods	KCl-SDS precipitation	(32,50,57,69,121)	Reproducible. Small number of steps.	Precipitates free proteins: high background and limits to measure total DPCs
	ARK	(68,122,123)	Low background. Highly sensitive.	Does not isolate particular proteins.
	Comet assay	(125,126,131,132,135)	Fluorescence microscopy images. Alternative to biochemical isolation methods.	Low sensitivity. Not possible to detect a protein of interest.
Immunodetection with specific DPC antibodies	TOP1cc	(65,67,138,139,142)	Reproducible. Sensitive and specific to TOP1cc.	Potential weak staining in IF.
	PARP1 on DNA fibres	(152,155,156,158)	Low number of cells needed. Large number of DNA fibres Robust data.	Complex quantification.

For full reference records, see the main manuscript.

Intention of this review

Appreciation of the importance and variety of DPCs has grown in recent years. Key to systematic investigation of DPC formation and resolution is a method that recovers covalently-bound proteins while eliminating loosely-bound proteins, and facilitates quantitative analysis of the isolated DPCs. The goal of this review is to highlight the technical advances that enable rapid, systematic and cost-effective analysis of DPCs. We hope that this will enable researchers to identify the most suitable DPC isolation method for their needs and to facilitate the comparison of results across different laboratories using different approaches.

By detecting either bound protein or bound DNA, DPC preparation methods can be classified into protein-targeted methods (discussed in **Section 1**) and DNA-targeted methods (discussed in **Section 2**). Additionally, we discuss the use of antibodies to detect specific DPCs (**Section 3**). We evaluate the advantages, limitations and applications of the most frequently used methods for isolating and/or detecting DPCs in mammalian cells (Table 1), based on published literature as well as our own practical experience. Additionally, we delve into the identification of the proteins implicated in DPCs.

Section 1. Protein-targeted methods for DPC preparation

Protein-targeted methods separate DNA from free proteins and analyse the proteins associated with the DNA. Early techniques used DNA-zol® (71–73) or phenol:chloroform/SDS (36), but more sophisticated approaches have since been developed, including caesium chloride density gradient, RADAR, STAR and PxP. In addition to describing these methods, we detail a diverse array of detection methods that aim to provide a qualitative, quantitative and sensitive assessment of the DPC content in a sample.

Caesium chloride density gradient separation

One of the earliest and most versatile techniques for DPC isolation is their purification through a caesium chloride (CsCl) density gradient (74) (Figure 3). CsCl is a dense salt that creates a linear gradient during ultracentrifugation, with decreasing density from bottom to top. Cellular material of varying molecular weight but similar density aligns in the CsCl gradient upon centrifugation (75). The covalent attachment of proteins to DNA modifies its DNA density, leading to a shift in its sedimentation properties.

Briefly, cells are lysed using a high detergent buffer (*e.g.* 1% Sarkosyl) to obtain chromatin, although sequential lysis steps with milder detergents are sometimes used to enhance DNA purity, such as 0.5% NP40 followed by nuclei/chromatin centrifugation and washes with 0.5% Triton X-100 (76). DNA is sheared by passage through a syringe needle. The prepared chromatin is applied to a CsCl gradient and centrifuged until equilibrium (*e.g.* 500,000 g for 4 hours, but this depends on the size of the sample and the tube's volume). DNA-bound proteins migrate into the gradient, while free proteins remain at the top. The bottom of the tube is poked and fractions are collected. DNA-containing fractions, whether native DNA or DPCs, are commonly selected based on A_{260} measurement and further processed and analysed conforming to the appropriate DPC detection method.
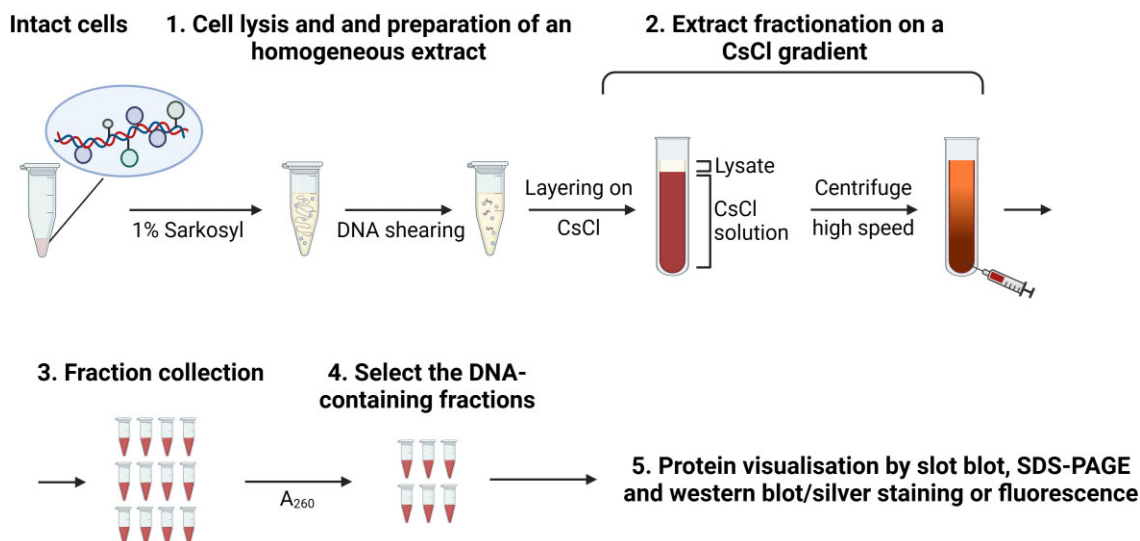
Figure 3. Caesium chloride gradient method. Schematic of CsCl gradient method according to (76).

A CsCl-based method was adapted to study DNA-bound topoisomerases and renamed ICE (In-vivo Complex of Enzyme bioassay)/ICT (In-vivo Complex Topoisomerase) assay. Originally employed to assess TOP1 association with ribosomal DNA (77), it was also applied to analyse the trapping effect of TOP1/TOP2 poisons and UV exposure (78–80). Topoisomerases covalent complexes (TOPccs) are conveniently detected *via* slot blotting (2–20 μg of DNA) using a specific primary antibody (80,81). Similarly, the CsCl gradient fractions can be blotted with antibodies specific to other crosslinked proteins of interest, such as DNMT1 crosslinks induced by 5-aza-dC (82) or cytokeratins induced by the anti-tumour drug aminoflavone (83). Besides individual proteins, the CsCl isolates can be analysed for total DPCs. Isolated proteins might be post-labelled with fluorescein isothiocyanate (FITC) (see below) for subsequent analysis by fluorescence readings or western blotting against FITC. This method proved quantitative and sensitive enough to track DPC increases and repair in response to physiological doses of aldehydes (*i.e.* ≤ 220 μM) (84).

Advantages
CsCl gradient ultracentrifugation is a well-established method for fractionating nucleic acids, with a long history and an extensive literature.

Limitations
This technique has fallen out of fashion because it requires costly, legacy equipment that is rarely available in contemporary laboratories. It is also time-consuming and labour-intensive, yields preparations of variable purity and has limited scalability.

RADAR (rapid approach to DNA adduct recovery)

RADAR (rapid approach to DNA adduct recovery; Figure 4A) is currently one of the most widely used methods for the biochemical isolation of DPCs (85). It is rapid, sensitive and requires very little starting material (10^4 cells, which yields the equivalent of 60 ng of DNA), making it suitable for high throughput analyses and ELISA immunodetection (86).

A typical RADAR extraction buffer for mammalian cells, made without proprietary reagents, contains the chaotropic agent guanidinium isothiocyanate (GTC, 6M) (85) and the detergents Triton X-100 (4%) and Sarkosyl (1%), along with 10 mM Tris-HCl (pH 6.5) and 20 mM EDTA (85). The distinctive combination of GTC with detergents creates a solution highly unfavourable to non-specific protein-DNA interactions. After extraction, nucleic acids are immediately alcohol-precipitated, the supernatant (which contains free protein) carefully removed and the pellet washed with 75% ethanol and re-precipitated. The washed pellet, which contains DNA and DPC, is resuspended in 8 mM NaOH, quantified with Quant-iT™ Picogreen™ to enable sample normalisation, and digested with benzonase to destroy nucleic acids (both DNA and RNA). The resulting protein isolate (*i.e.* the crosslinked proteins) can be analysed by various methods, including ELISA, slot-blot and Western blotting to detect specific proteins, or by mass spectroscopy for an unbiased view of the adduct repertoire. The entire adductome can be examined by silver or Flamingo staining, as specified below. Negative controls have confirmed the specificity of crosslink detection (3,87).

RADAR has been widely applied to characterise chemotherapeutically relevant DPC-inducing agents, to measure drug efficacy and to clarify the mechanism of action of anti-tumour drugs. First applied to the detection of DNMT1 crosslinks induced by the anti-tumour drug 5-aza-dC (85), RADAR has been employed to analyse the formation of TOP1 and TOP2 adducts by camptothecin (CPT) and etoposide (32,59,65,85,88,89); to compare the efficacy of conventional topoisomerase trapping agents with new topoisomerase inhibitors, such as the second generation of indenoisoquinolines, which have better pharmacokinetics and lower toxicity than CPT (90); to establish that exatecan induces TOP1cc more effectively than either CPT or topotecan (91); and to demonstrate that the compound CX-5461, originally described as an RNA polymerase I inhibitor, caused the formation of TOP2B crosslinks (92). RADAR was also used to demonstrate that lipid peroxidation by-products accumulating in CPT-treated cells are themselves capable of inducing DPC. Among these, 4-hydroxy-2-nonenal stabilises TOP1ccs with efficacy
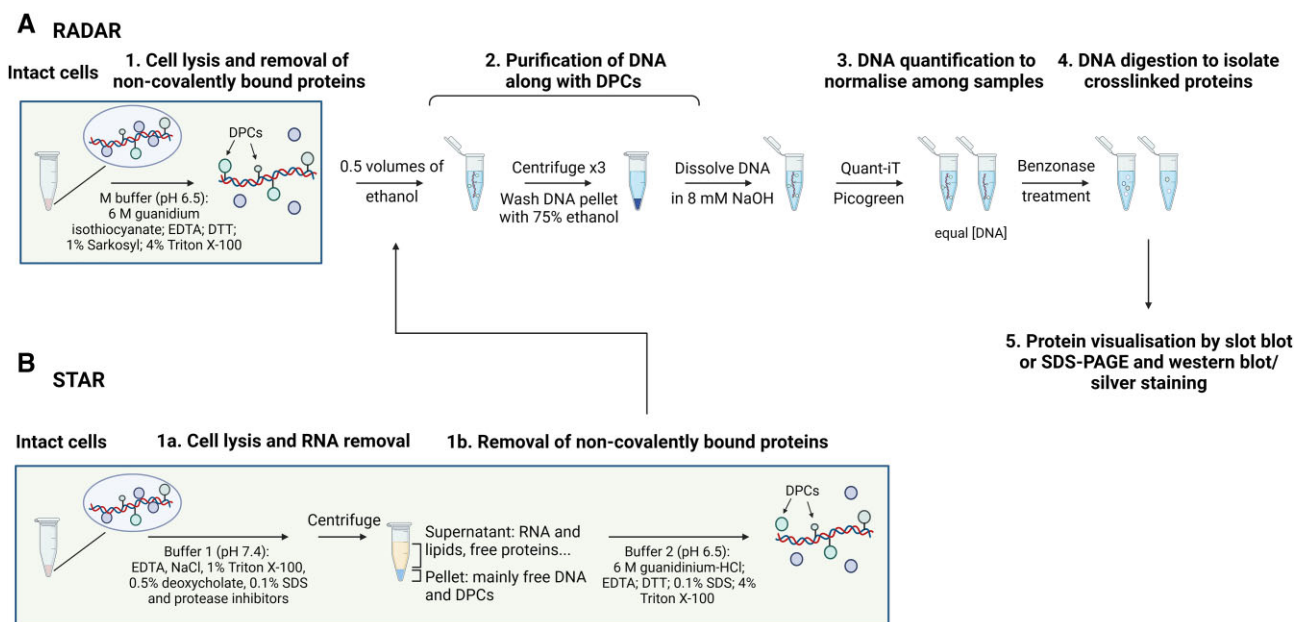
Figure 4. RADAR and STAR methods. (**A**) Schematic of RADAR method according to (85). (**B**) Schematic of STAR according to (31).

comparable to CPT, thus showing that CPT causes the covalent complex by both direct and indirect mechanisms (93).

RADAR, through its ability to co-precipitate DNA and RNA, has enabled the demonstration that TOP3B crosslinks to RNA in addition to DNA, and is the only topoisomerase known to be capable of doing so (94). RADAR has also been instrumental in characterising DPCs formed by DNA polymerase beta in response to oxidative stress (2), by histones H3/H4 in response to formaldehyde treatment (32), and by the Epstein-Barr virus EBNA1 protein (87).

The analysis of specific crosslinked substrates alongside total DPC isolates has shown a role in DPC repair for the protease SPRTN (32), the ATPase p97 and its cofactor TEX264 (65), the proteasome (3), the DNA break repair factor XRCC1 (95), the de-ubiquitylating enzyme (DUB) USP11 (68), and the exonuclease EXO1 (96) in human cells; and for the protease GCNA in *Drosophila* and zebrafish (97). Furthermore, post-translational modifications such as ubiquitylation and SUMOylation can be detected on Western blots of DPCs (59) or specific adducts (63,98), establishing a role for the ubiquitin/SUMO system in DPC repair.

Finally, RADAR isolates have been subjected to downstream mass spectrometry and Illumina sequencing (18,32,97,99). Mass spectrometry analysis of total DPCs from SPRTN-depleted cells and GCNA knock-out Drosophila embryos have defined the most abundant crosslinked proteins in these systems (32,97). In addition, it defined the relatively limited repertoire of proteins and protein families that constitute the 'adductome' in human cells (37).

Illumina sequencing was used to sequence the regions where PARP1 covalent adducts form as a consequence of DNMT1 activity (Adduct-seq, (18)) and where TOP1ccs preferentially form (TOP1 CAD-seq, (99)).

Advantages

RADAR is easy to execute and relies on non-proprietary reagents and equipment commonly found in contemporary laboratories. It can be performed on human cells, tissues,

Drosophila and zebrafish embryos, and Mycobacteria, which are notoriously difficult to lyse (85,86,97,100). RADAR is sensitive, specific, versatile and suitable for downstream applications such as mass spectrometry (32,37,97) and Illumina sequencing, provided that the DNA is resuspended in a buffer compatible with library preparation and that chromatin-immunoprecipitation (ChIP)-grade antibodies are available for the protein of interest (18,99).

Limitations

Insufficient removal of the non-crosslinked proteins can result in a high background. In order to minimise it, the user should pay particular attention to the following steps: (i) the cells must be lysed in the appropriate volume of buffer (advised no less than 1 ml per 2×10^6 cells); (ii) following the initial ethanol precipitation, the DNA pellet must be washed carefully and extensively.

Useful modification

Both DNA and RNA are recovered following alcohol precipitation of cell lysates. RNA may be removed by RNase A treatment of purified nucleic acids (100 μg/ml, 30 min at 4°C) (85,86) (42,62,94).

STAR (superior method for true DNA–protein crosslinks recovery)

The Superior method for True DNA–protein crosslinks Recovery (STAR; Figure 4B) (31) is an adaptation of the RADAR method aimed at reducing RNA-associated proteins from DPC preparations. Cells are lysed in a buffer containing detergents (1% Triton X-100, 0.1% SDS, 0.5% deoxycholate, similar to RIPA buffer) but lacking GTC. Crude nuclear pellets are recovered by high speed precipitation and centrifugation and then resuspended in a buffer containing guanidinium-HCl. Subsequently, DNA precipitation and sample treatments are carried out as per the RADAR protocol.

STAR appears sound by several criteria. It efficiently detected formaldehyde-dependent crosslinking of histone H3,

an abundant DPC, in line with previous analyses (32,101). It also established that crosslinking of the RNA-dependent helicases DDX5 and DDX9 increases with increasing formaldehyde concentrations. In addition, when STAR was employed to assess DPC repair kinetics over 4 hours after treatment with 400 or 800 μM formaldehyde, the kinetics of DPC removal and the role of SPRTN in clearing DPCs within this time frame were consistent with previous studies using RADAR (32,102).

The DPC preparation by STAR was separated by SDS-PAGE and the gel was sliced into higher and lower molecular weight components (>60 kDa or <60 kDa). Proteins from each slice were eluted and analysed by mass spectrometry. The faster-migrating fraction contained peptides of proteins larger than 60 kDa in molecular weight (such as TOP1), which are presumably the result of cleavage/digestion of full-size DPCs, in line with previous reports of proteolytic repair (10,59).

Proteins isolated by the STAR method were analysed by semi-quantitative mass spectrometry and, while a repertoire of over 300 proteins was found, the majority of DPC pools contained only a few highly abundant proteins from the core histone and hnRNP families and some filamentous structural proteins. Approximately two hundred other proteins, including ubiquitin and SPRTN itself, were found to comprise a minor proportion of the DPC pool. After treatment with formaldehyde, proteins involved in RNA metabolism showed the highest increase (31).

Advantages

The bulk of RNA and soluble proteins are efficiently removed with the preliminary lysis step and centrifugation, which effectively only precipitates DNA and DPCs. The result is a DPC preparation devoid of cytoplasmic RNA and contamination from non-crosslinked proteins.

Limitations

Preliminary cell fractionation in a buffer lacking chaotropic salts may be accompanied by proteolysis and nucleolytic digestion that destroys or modifies DPCs.

Other considerations

As recovery of RNA is indeed intrinsic to the nucleic acid isolation, several other protocols have implemented RNase treatment to remove RNA (31,42,62,85,94). A somewhat similar step to eliminate cytoplasmic RNA was also previously implemented (18) using a proprietary detergent (MPER, ThermoFisher) to disrupt cell membranes, followed by high-speed pelleting of the nuclear material and denaturation in RADAR buffer. In this respect, the preliminary lysis for RNA removal in the STAR protocol provides limited advance.

There is no convincing evidence that the STAR protocol achieved higher levels of sample purity compared to RADAR (31). The claim appears to have been based on a comparison between a STAR preparation and a particularly contaminated RADAR preparation. For example, while DPCs isolated by RADAR have previously been shown to increase in a time- and dose-dependent manner after formaldehyde treatment, with a detectable increase at 150 μM (59), DPC accumulation in the RADAR preparation from Glumac *et al.* (31) cannot be detected even in cells treated with 400 μM formaldehyde. In addition, a comparison between the RADAR isolates and total extracts shows barely any difference in the level of cytosolic and nuclear contaminating proteins; on the other hand, histone H3, an abundant protein prone to crosslinking (32),

remains undetectable in the DPC preparation. Hence, based on these contradictions with previously published reports, the benefit of cell lysis in detergent before RADAR extraction remains unclear.

PxP: purification of crosslinked proteins

As an alternative to precipitation, DNA can be extracted from cells and immobilised on a solid matrix such as agarose, from which non-crosslinked proteins are removed by electroelution (Figure 5). This approach to 'Purification of x-linked proteins' (PxP) (33) represents an adaptation of the chromosome entrapment experiments originally designed to assess interactions between prokaryotic condensin rings and DNA (103).

Briefly, 1×10^6 cells per condition are harvested and embedded in 1% low melting agarose plugs and lysed in 2% sarkosyl. The plugs are then subjected to PAGE to elute the mobile cellular proteins from the agarose, while DNA and crosslinked proteins remain inside the plug due to their very high molecular weight. Plugs are washed, the agarose melted and DNA digested with benzonase to release the crosslinked proteins, for downstream analysis by western blot or mass spectrometry. For control purposes, parallel samples are processed similarly, except that DNA digestion is performed on the agarose plug before electro-elution to identify any co-purifying background contaminant proteins.

PxP followed by western blotting has been validated for the detection of TOP1cc, histone H3 and DNMT1, formed in a dose-dependent manner in response to treatment with CPT, formaldehyde and 5-aza-dC, respectively. Interestingly, this experimental strategy served to establish an additional role of SPRTN in post-replicative repair of 5-aza-dC-induced DNMT1 crosslinks. SUMOylation of the protein adduct followed by RNF4-dependent ubiquitylation initiates this replication-independent repair mechanism, generating a modified DNMT1-DPC that may be resolved by either the proteasome or SPRTN (33,61).

PxP has also been used in combination with LC-MS/MS to reveal the identity of DPCs following the exposure of HeLa cells to a high formaldehyde pulse (2 mM, 1 hour). Through label-free quantitative (LFQ) mass spectrometry, 35 proteins were significantly enriched in PxP plugs after formaldehyde exposure. Notably, five core histones (H2AC14, H2AFY, H2BC12, H3C1, H4C1) accounted for the vast majority of the formaldehyde-induced DPCs, indicating that crosslinked nucleosomes are the predominant formaldehyde-induced adducts, as expected if crosslinking is a function of protein proximity to DNA. Interestingly, the complexity of the DPC pool isolated by PxP appeared to be comparatively simpler, both in terms of number and composition, than the DPC pool isolated by alternative methods.

Advantages

PxP provides a high level of confidence in distinguishing DPCs from contaminants, thanks to protein identification in nuclease pre-treated control samples. By comparing the protein profiles between experimental samples and control samples, researchers can confidently discern the presence of DPCs while minimising the risk of false positives or contaminants. The degree of confidence in the results can be adjusted by establishing thresholds for the fold change increase between the experimental and control samples. For instance, by setting a
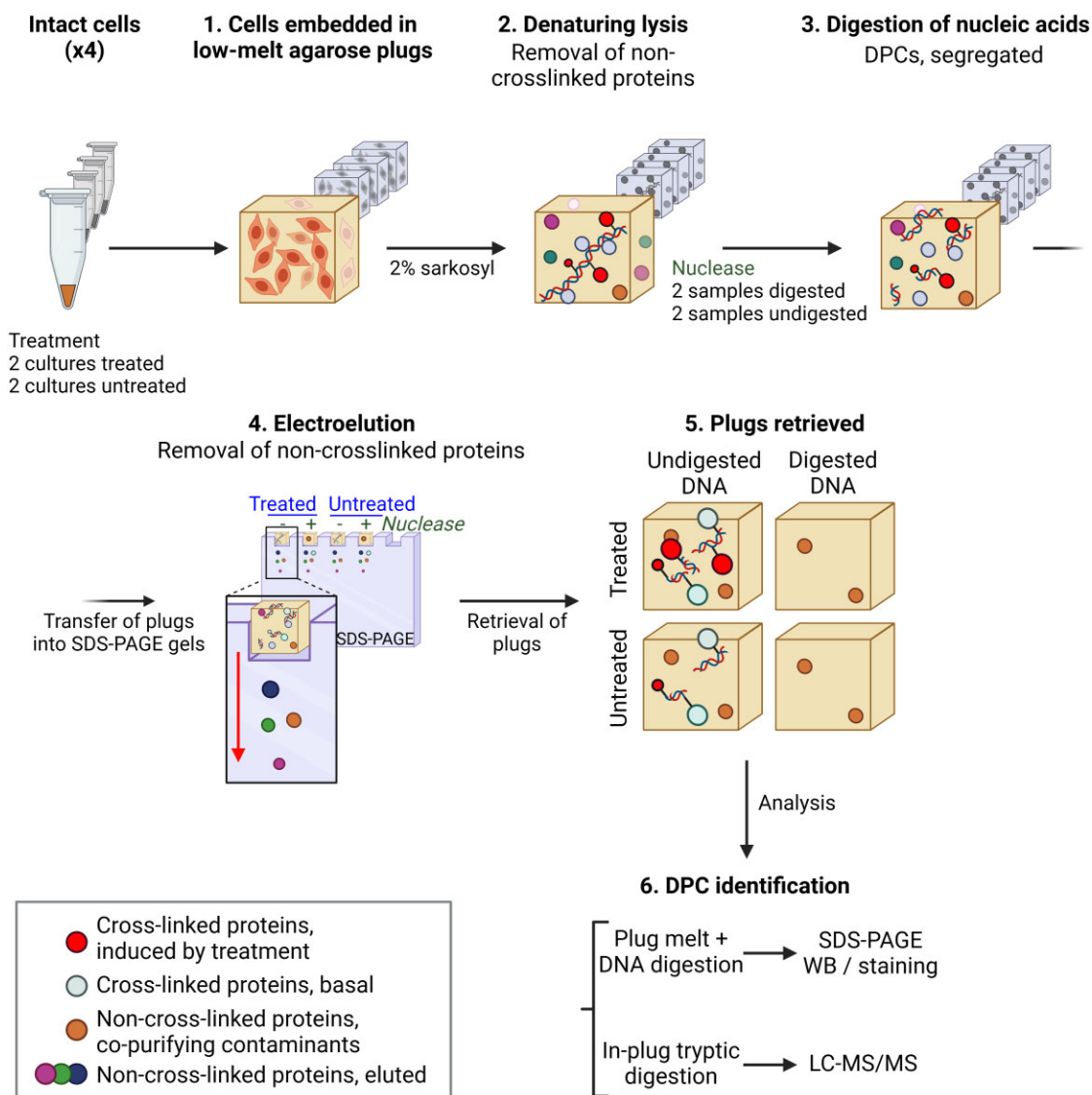
Figure 5. PxP method. Schematic of PxP method according to (33).

threshold of a log$_2$ fold increase of 2, researchers were able to confidently identify 35 proteins as formaldehyde-induced DPCs. This set included seven histones, five members of the HMG family, TOP1, TOP2A and PARP1. Consequently, PxP enhances the accuracy and reliability of DPC studies.

Notably, many proteins integral to large structures like the ribosome, proteasome, collagen, tubulin, actin and even the p97 segregase did not undergo electro-elution effectively. Nevertheless, these proteins could be readily identified as contaminants as levels were comparable in control and experimental samples.

Limitations

Limits on the number of cells that can be embedded in an agarose plug restricts the scalability of the method. Additionally, large protein complexes may not electro-elute from the agarose plugs. Furthermore, while the PxP procedure offers the advantage of high-confidence identification of proteins involved in DPCs, it may also present a challenge when aiming to detect all genuine DPCs. The procedure may potentially lead to false negative results. The number of proteins iden-

tified in the negative control samples pre-treated with benzonase is notably high, and less than 10% of the total identified proteins were considered to be involved in DPC formation after formaldehyde treatment. For example, this study failed to detect certain proteins that were confidently identified as formaldehyde-induced DPCs using RADAR (37). These proteins include histones H1-2, H1-3, H1-4 and H1-5, the transcription factor FACT complex subunit SPT16, the helicase DDX46, the ubiquitin-like protein SUMO2, and some ribosomal proteins. Some of these proteins were also identified as DPCs with high confidence in SPRTN-deficient cells. The discrepancies in results may be attributed to the use of different cell lines, formaldehyde doses and treatment durations. Consequently, the variety of formaldehyde-induced DPCs identified using PxP is narrower in comparison to alternative methodologies like RADAR.

Detection of DPCs prepared by protein-targeted methods

While methods for isolating DPCs are crucial for enriching and purifying DPCs from complex biological samples,
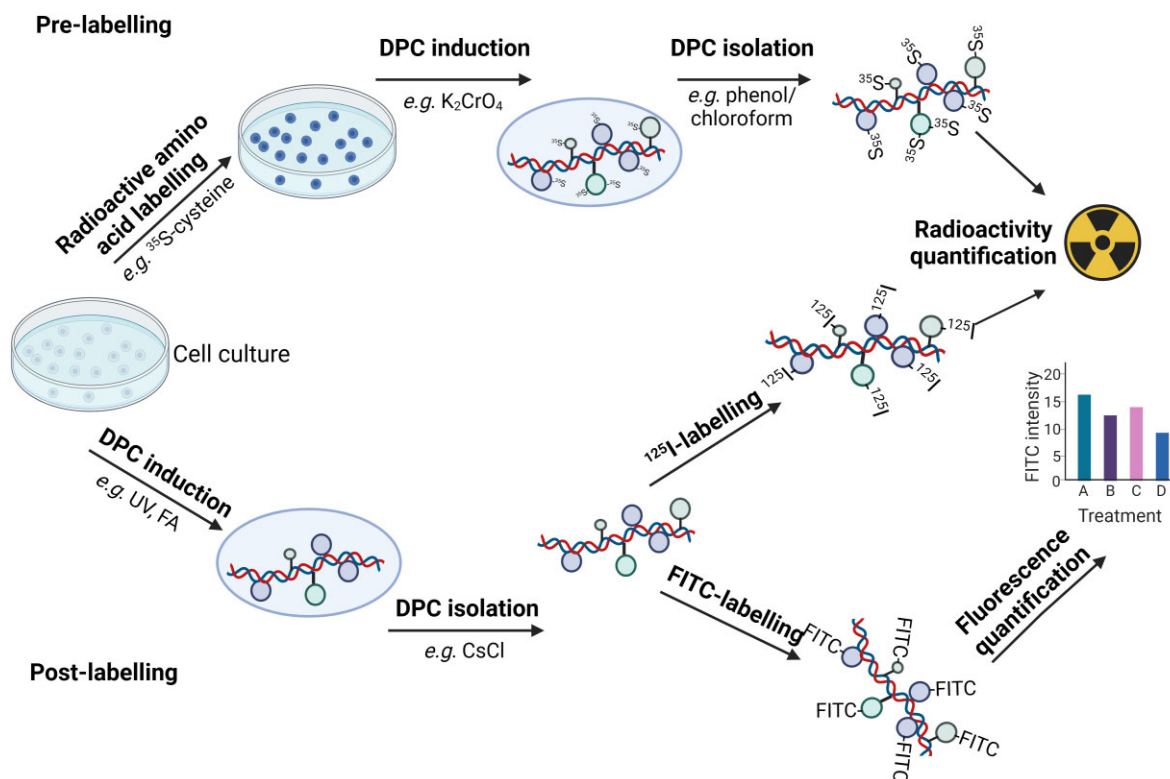
Figure 6. Labelling of DPCs with radioisotopic amino acids and fluorescence chemicals. Upper panel: pre-labelling method before DPC isolation according to (104). Bottom panel: post-labelling method after DPC isolation according to (44,45,84,107).

detection methods aim at identifying and quantitatively assessing the presence of DPCs in a given sample. In recent years, the development of innovative approaches has driven remarkable progress in the field of DPC detection. These advances have expanded the toolbox available for studying DPCs, incorporating principles like immunodetection and mass spectrometry, resulting in significant improvements in sensitivity and accuracy. Here, we explore various methods for detecting DPCs following a protein-targeted DPC preparation. These approaches can be protein-specific, employing immunodetection techniques such as Western blotting and ELISA, or non-specific, utilising radio- or fluorescence labelling. Additionally, we provide an overview of the main techniques relying on mass spectrometry for the characterisation of DPCs.

DPC detection by SDS-PAGE, western blot and ELISA

Proteins of interest within a DPC preparation isolated through a CsCl gradient, RADAR, STAR and PxP are commonly visualised by Western blot, whereby the DPC isolates are resolved by SDS-PAGE, transferred to a nitrocellulose or PVDF membrane, and probed with specific antibodies (18). The quantity of DNA required per sample depends on the cell line, the abundance of the crosslinked protein, the quality of the antibody used for detection and the desired sensitivity. In our experience with RADAR, samples containing 10–30 μg of DNA (recovered from approximately $2–6 \times 10^6$ human cells) are usually adequate (18,32). Alternatively, the entire DPC preparation resolved by SDS-PAGE can be visualised using silver staining or Flamingo®, where we recommend digesting approximately 80 μg of DNA to achieve good sensitivity.

For high throughput microplate analysis, such as ELISA-based RADAR, samples containing 60 ng of DNA (10^4 human cells) are suitable (86). Here, cells can be cultured in a multi-well plate for practicality. The inclusion of silica gel fines in the lysis buffer (4–16% v/v) facilitates DNA precipitation by centrifugation. The purified DNA is then resuspended in 8 mM NaOH, quantified, diluted to 5–500 ng/ml, and digested with benzonase. Serial dilutions of DPCs are absorbed onto an ELISA plate - taking advantage of their natural affinity for plastic in microtiter plates - and subsequently detected using a primary antibody specific for the protein of interest and an HRP-conjugated secondary antibody (86).

Detection of DPCs by labelling

The protein component of a DPC can be labelled with radioactive isotopes or with fluorescent chemicals to facilitate its detection; this labelling can be done either in cells during cell culture (pre-labelling) or after DPC isolation (post-labelling; Figure 6).

The most common pre-labelling method involves adding a radioactive amino acid, such as ^{35}S-cysteine, to the culture medium. As newly synthesised proteins incorporate the radioactive amino acid, crosslinked proteins will also be labelled and DPC frequencies can be assessed by the total radioactivity in the DPC preparation. This method was employed in Chinese hamster ovary cells (CHO) to quantify DPCs induced by chromium or nickel compounds (10–100 μM K$_2$CrO$_4$ or 0.2–2 mM NiCl$_2$, 20 hours) (104).

Post-labelling methods most commonly employ radioactive iodine (^{125}I) or fluorescein. DPC preparations can incorporate ^{125}I into tyrosine residues at neutral pH in the presence of an oxidant (such as chloramine-T) (105,106). In
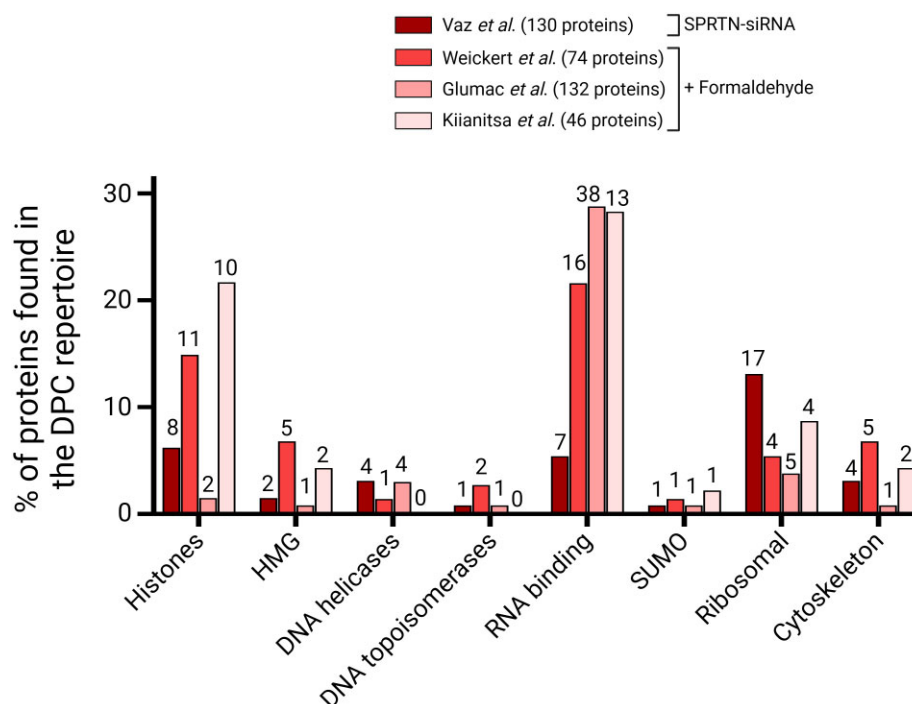
Figure 7. Comparison of protein repertoires identified by mass spectrometry. DPCs were induced by either depleting SPRTN from cells or by exposing cells to formaldehyde. Different DPC isolation methods: RADAR (32,85), PxP (33), STAR (31), cell types, MS instruments and bioinformatic analysis were used in each study. The graph shows the percentage of each protein group in the whole repertoire (bars) and the actual number of proteins found. HMG: high mobility group protein family.

the original study, CHO cells were treated with crosslinking agents (UV, formaldehyde, cisplatin, chromium or nickel compounds), DPCs isolated using phenol/chloroform, and proteins digested with Proteinase K. The residual DNA-bound peptides were post-labelled with ^{125}I and the radioactivity of the sample was measured with a standard γ-counter, along with the 260/280 nm absorbance for normalisation against DNA (107). Other studies used fluorescein isothiocyanate (FITC) post-labelling to quantify the DPCs (purified by CsCl density gradient) from cells exposed to aldehydes and ionising radiation (44,84). The isothiocyanate group from FITC reacts with the amino groups in proteins, therefore labelling the entire DPC preparation. DPCs can be estimated by reading total fluorescence at 520 nm or by SDS-PAGE and blotting with anti-FITC antibodies. FITC post-labelling was used to analyse the involvement of the NER repair pathway in DPC repair (44). FITC labelling revealed a sensible delay in DPC repair in NER-defective cells, where conventional SDS-PAGE/gel staining did not detect any difference. With complementary experiments, the study demonstrated that only a fraction of DPCs, typically those smaller than 8 kDa, is processed by NER (44).

In conclusion, while radio-isotope labelling is not attractive due to the radioactivity involved, fluorescence-based methods are safer and should find greater use among the DPC detection methods.

Advantages

Although relying on conventional DPC biochemical isolation methods, the post-labelling methods are more sensitive, and quantitation can be more precise than other DPC detection systems.

Limitations

In these methods, all crosslinked proteins are labelled, which limits the analysis of a specific DPC. Further, the pre-labelling method does not exclude a direct reaction between radioactive amino acids and DNA, which makes it difficult to measure the real DPCs. Due to the different sensitivities of radioactive amino acids, selecting the appropriate one depends on the context of an experiment.

Analysis of changes in DPC number and composition by mass spectrometry

A collective effort by several research groups to identify proteins involved in DPCs through unbiased mass spectrometry (MS) has established that DPCs are extremely complex and heterogeneous in nature. The complete repertoire of proteins found in DNA–protein conjugates could potentially encompass thousands of different proteins (Figure 7), but it has been consistently observed that the majority of DPCs are formed by only a limited set of abundant proteins that are primarily located in the nucleus and exhibit a strong affinity for DNA. DPCs at baseline levels are found in all cells, underscoring the importance of determining the frequency at which DPCs occur, to understand their relevance to genome stability.

Absolute quantification of DPCs

The first question to characterise DPCs is to determine their occurrence rate in cells. Isotope dilution tandem mass spectrometry assay (35,36,108) is a technique capable of attaining absolute quantification of total DPC numbers in cells and tissues. In brief, the protein constituents from cellular DPCs are broken down into single amino acids and spiked with known amounts of isotopically labelled standard amino acids. Subse-

quent mass spectrometry analysis allows the quantification of the DPC-derived amino acids as a light-to-heavy isotope ratio. While achieving sensitive DPC quantifications, this particular technique does not provide information about the specific proteins involved. Using this technique, the number of the specific DPC dG-Me-Cys endogenously generated in rodent liver tissues was estimated at 15 DPCs per 10^8 dG (or around 450 dG-Me-Cys per cell), above the levels presented in nasal tissue (3.6), bone marrow (2.3) and peripheral blood mononuclear cells (1.3) (108), reflecting the higher exposure of the liver to metabolic aldehydes. As expected, treatments with exogenous crosslinking agents significantly increased the frequency of DPCs. Thus, exposure of rats to airborne formaldehyde at a concentration of 15 ppm for several days increased the number of dG-Me-Cys in nasal tissue from 6.5 to 18.2 per 10^8 dGs. Other studies estimated the number of DPCs at 200 Cys-N7G-EMA adducts per 10^8 dGs and 6 Cys-NOR-N7G adducts per 10^8 nucleotides in cultured human fibrosarcoma HT1080 cells treated with mechlorethamine (50 μM, 3 hours) and nor-nitrogen (100 μM, 3 hours), respectively (35,36). These results indicate the high frequency and formation of DPC conjugates.

Characterisation of naturally occurring DPCs

The detection of naturally occurring DPCs can be greatly enhanced by impairing the activities of SPRTN and/or the proteasome, the two cellular mechanisms that repair them, leading to the accumulation of these crosslinks. In our efforts to identify SPRTN substrates, we isolated DPCs using RADAR from HeLa cells where SPRTN had been depleted through siRNA, as well as from parental HeLa cells (32). Subsequently, we analysed these isolates by LFQ mass spectrometry. We observed 416 different protein constituents of DPCs under basal conditions. Notably, nearly one hundred of these proteins exhibited a significant increase of more than 1.5-fold in the SPRTN-depleted samples, of which one-third had DNA-binding properties, including histones, topoisomerases and DNA helicases. Results were further validated by assays involving the expression of either SPRTN^{WT} or the catalytically inactive mutant SPRTN^{E112A}. Some DPCs that did not show an increase in SPRTN-deficient cells might be resolved through SPRTN-independent mechanisms or their levels might be too low to detect changes using a mass spectrometry approach.

Characterisation of DPCs induced by exogenous agents

Exposure to exogenous chemical crosslinking agents (aldehydes, reactive oxygen species, carcinogen metabolites, anticancer drugs or irradiation) results in significant increases in DPC number and complexity. Since formaldehyde is the main cell metabolite contributing to endogenous DPCs, treatment of cells with this agent may most accurately reproduce the protein range and chemical structures found in physiological conditions.

DPCs induced by formaldehyde

Formaldehyde, the smallest organic chemical with crosslinking properties, is a metabolic by-product derived mainly from histone demethylation, methanol oxidation and methylamine deamination (109,110) and therefore a major source of the endogenous genotoxic challenge faced by cells (111,112). It is also a contaminant from cigarette smoke, automobile exhaust and chemical industries (113).

Kiianitsa *et al*. (37) treated CCRF-CEM lymphoblasts with 0.5, 1 and 2 mM formaldehyde for 1 hour. These doses were chosen as they moderately exceed the physiological concentration of formaldehyde in tissues and organs but are much lower than those used for extensive chromatin crosslinking in ChIP protocols (often 130 mM). Following two rounds of RADAR fractionation, including RNase A treatment, the samples were subjected to LFQ MS analysis. The result of this analysis revealed that neither the MS signal nor the number of crosslinked proteins exhibited an appreciable increase with the formaldehyde dose, confirming that massive chromatin crosslinking did not occur under these conditions. However, over 40 proteins were identified as highly reactive to formaldehyde, and their abundance in DPCs increased in a dose-dependent manner across all replicates. The majority of these DPCs were composed of various histones, HMG family members, and a diverse group of nuclear RNA-binding proteins and splicing factors. The formaldehyde-reactive DPC fraction was also highly enriched with SUMO2 peptides, a post-translational modification that targets DNA-crosslinked proteins for proteolytic repair. Notably, similar experiments in HeLa cells treated with exogenous formaldehyde at low doses (2 mM, 1 hour), followed by DPC isolation using the PxP method (33), yielded a very similar repertoire of formaldehyde-reactive proteins, predominantly composed of core histones and HMG species. They also identified other chromosome-binding proteins such as DEK, BAZ1B and HP1BP3 and the mitochondrial protein SSBP1, backing the suitability of PxP for detecting DPCs of mitochondrial origin.

Independently, Glumac *et al*. (31) undertook a qualitative MS approach to identify formaldehyde-induced (400 μM, 15 min) DPCs extracted from HeLa cells using the STAR method. Formaldehyde treatment did not significantly alter the composition of the DPC pool but enriched certain nuclear proteins that were under-represented in untreated cells (ranking within the 99% percentile of MS signal from the bottom). Among the top reactive proteins identified were some involved in ribosome function, RNA maturation, chromatin structure, DNA repair and SUMO2/3, which argues in favour of SUMOylation as a DPC repair signal (59,61,102,114).

The consistency in results across different techniques and cell types confirms the existence of a limited repertoire of proteins that exhibit high reactivity to exogenous formaldehyde at doses only slightly above the physiological level. In addition, the fact that the extent of the DPC damage may be dependent on the formaldehyde dose, as shown for histone H3, may partially explain differences across different studies.

DPCs induced by non-enzymatic anti-tumour and contaminant agents

Chemotherapeutic crosslinkers encompass compounds like nitrogen mustards and platinum-based drugs, which can induce a wide spectrum of DNA–protein and DNA-DNA crosslinks. This field, therefore, has substantial medical significance and deserves further attention. Several studies conducted in HT1080 human fibrosarcoma cells treated with phosphoramide mustard (100 μM) or mechlorethamine (25 μM) (35), where DPCs were isolated by phenol-chloroform/SDS and analysed by qualitative mass spectrometry, identified 134 and 38 proteins, respectively, most of which
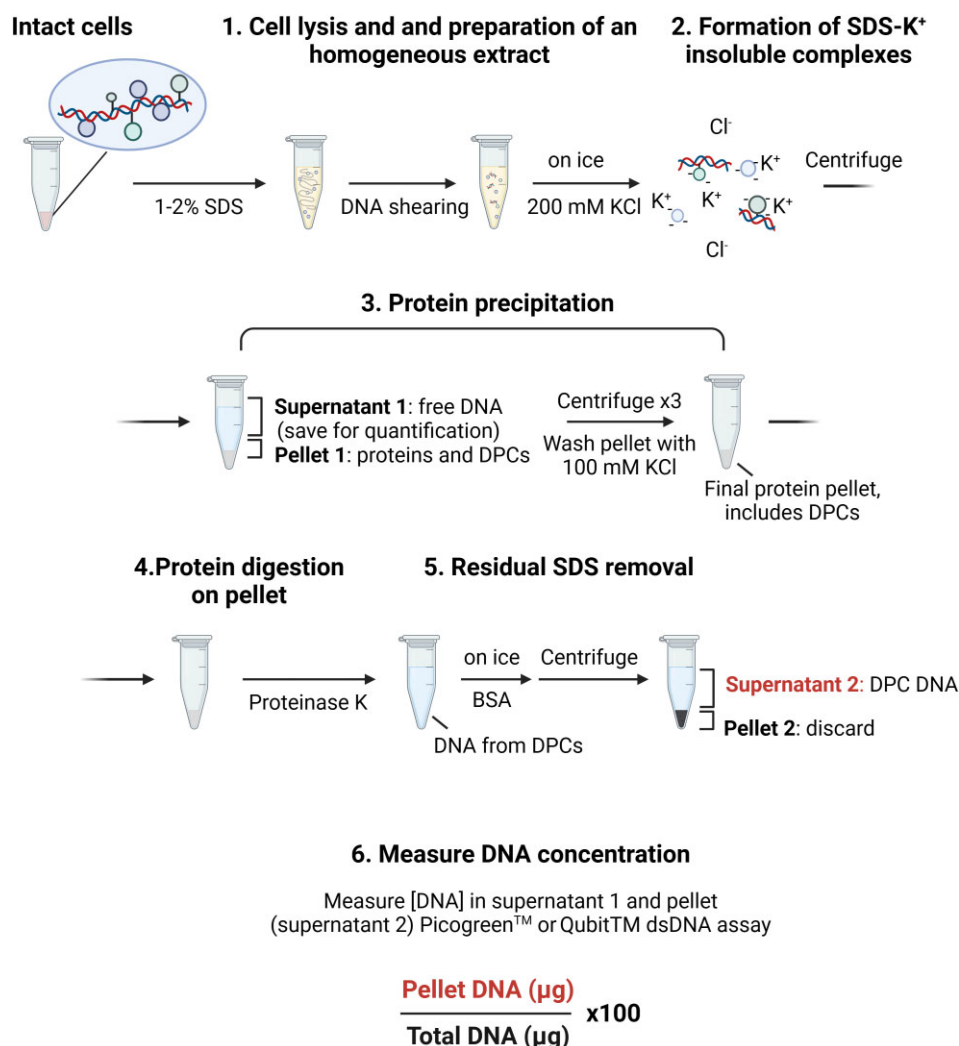
Figure 8. KCl-SDS precipitation. Schematic of KCl-SDS method according to (122).

were nuclear and at least one third with DNA- or RNA-binding motifs. On the other hand, treatment with cisplatin (100 μM) induced the increase of over 250 nuclear proteins, with many belonging to the HMG, histone and elongation factor families (28). Interestingly, several proteins forming DPCs were cytoplasmic, ribosomal and membrane-bound, revealing the wide range of DPC lesions induced by cisplatin. Moreover, the effect of carcinogenic environmental pollutants on DPCs has been exemplified by treating cells with 1,2,3,4-diepoxybutane (DEB, 2 mM) (34,115–117), which crosslinked 152 proteins, most of them with nucleic acid-binding motifs such as histones, HMG proteins, transcription factors, splicing factors and structural proteins.

Overall, these findings indicate that formaldehyde tends to enhance endogenous DPCs while also inducing the formation of DPCs involving proteins less prone to crosslink endogenously. The DPCs formed by nitrogen mustards, cisplatin and DEB do not differ considerably from those induced by formaldehyde. With small variations across studies, the most abundant protein components of DPCs are histones, chromatin modification/ compaction factors, transcription factors, structural proteins and a variety of ribonucleoproteins. The presence of RNA-binding proteins in the DPC pool might be attributed to the fact that many of them also possess an affinity for DNA. While contamination with RNA was not assessed in many cases, such a possibility has been ruled out before by HPLC analysis of the enzymatic digests (35,36).

Section 2. DNA-targeted isolation methods

DNA-targeted methods estimate the amount of DPC through a quantification of DNA. The DNA engaged in DPCs is isolated from the rest either through a co-precipitation of proteins (KCl-SDS precipitation and ARK) or due to distinct electrophoretic mobility (modified comet assay). While these methods may not allow direct visualisation of crosslinked proteins, they may offer the opportunity to study the locations of DPC in the genome.

KCl-SDS precipitation assay

This method is designed to isolate proteins, including DPCs, based on the ability of the ionic detergent SDS to bind and neutralise the cationic sites in proteins, thereby stripping weakly bound proteins from DNA (Figure 8). Subsequent addition of potassium chloride (KCl) forms an insoluble K⁺-SDS complex that is precipitated by centrifugation, leaving free DNA in the supernatant.
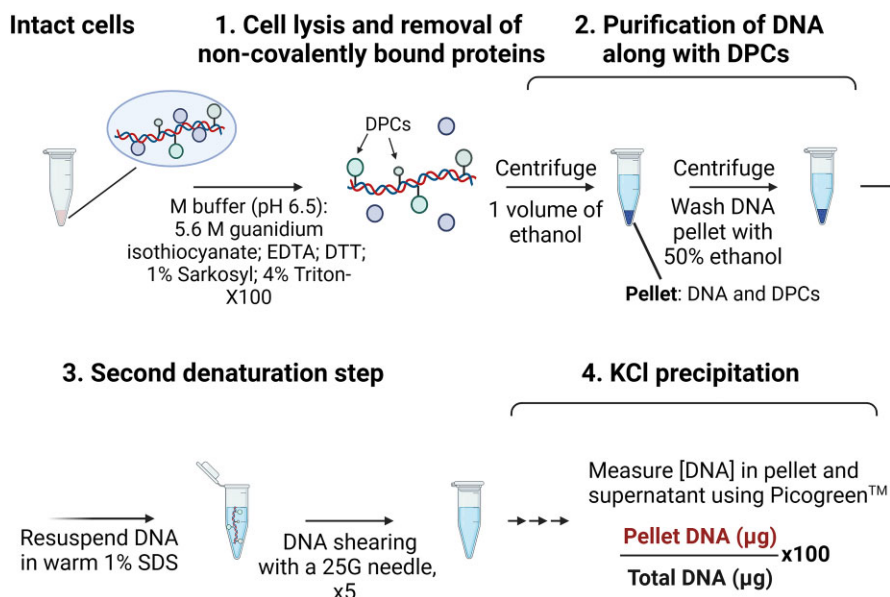
Figure 9. ARK assay. Schematic of the ARK method according to (121).

Initially developed for *in vitro* assays, this method was first applied to isolate avian TOP1 after reaction with ^{32}P-labelled DNA. Counting ^{32}P radioactivity in the precipitate quantifies the extent of TOP1 crosslinking to DNA (118). The ability of this method to isolate covalently attached proteins was substantiated *in vitro* using a mix of plasmid DNA and purified histone H1 or DNA polymerase I proteins (119). The addition of 1% SDS disrupted DNA interaction with these proteins and led to a DNA-free (unlabelled) pellet; conversely, when samples were UV-irradiated before SDS treatment, DNA was detected in the K$^+$-SDS pellets, consistent with the formation of covalent bonds between DNA and histone H1 and DNA polymerase. Presence of the labelled DNA in the pellet was also observed when TOP1 was crosslinked to DNA enzymatically, in the absence of external crosslinking agents (118,119). This method was further applied to determine TOP1 and TOP2 cleavage sites on DNA (120).

The KCl-SDS precipitation method was then adapted for isolating covalently attached proteins from human and murine cells (121) and used to demonstrate defective DPC repair in human cultured cells deficient in the protease SPRTN (32,50) and the DUB USP7 (69). Typically, 1×10^6 cells are lysed with 1–2% SDS. Genomic DNA is sheared by freezing/thawing cycles, sonication and/or robust vortexing. After heating at 65°C, 200 mM KCl is added on ice, forming an insoluble complex. Gentle centrifugation separates the protein pellet (from the supernatant containing free DNA), which is resuspended in 100 mM KCl and washed three times through cycles of heating, cooling and centrifugation before being resuspended in 100 mM KCl for DNA quantification. Proteins are next digested with Proteinase K, and BSA is added to quench the SDS. A final centrifugation step eliminates the SDS, yielding a supernatant containing the DPC-DNA. Concentrations of DPC-DNA (from the pellet) and free DNA (from the initial supernatant) are measured by various assays: Hoechst 33258 (121), Quant-iT™ Picogreen™ (32) or Qubit™ dsDNA HS assay (69) with standard DNA probes. The amount of DPCs is calculated as the ratio between DNA in the pellet and the total DNA (pellet plus supernatant) (named the DPC coefficient). The KCl-SDS method is sensitive enough to detect a dose-dependent increase of DPCs in response to cisplatin or formaldehyde, even at low concentrations of 75 μM formaldehyde (50,57,69,121), but, as the sensitivity depends on DNA fragment size, procedure standardisation is crucial for reproducible results.

Advantages
The KCl-SDS precipitation method is straightforward and can be used to analyse DPCs from various mammalian cell types. It is inexpensive, relies on standard laboratory reagents and no special lab equipment is needed. Therefore, it can be applied to a large number of samples.

Limitations
As this is a DNA-targeted isolation method, it cannot be used for protein visualisation or to check for a protein of interest within the DPC preparation, limiting its applications. Due to the thick and difficult-to-handle K$^+$-SDS precipitate, contamination of the DPC-containing pellet by free DNA is common and generates a high background, so proper resuspension of the pellet during washing is recommended. Strong, non-covalent protein-DNA interactions that resist 1–2% SDS lysis increase the DNA read in the precipitate, contributing to background.

ARK (advanced recovery of K-SDS precipitates) assay

The ARK method (Figure 9) overcomes some limitations of the KCl-SDS assay by combining it with RADAR (122). As in RADAR, lysis with the chaotropic salt GTC in the presence of detergents releases non-covalently associated proteins and the resulting genomic DNA, including DPCs, can then be precipitated with ethanol. Performing lysis at 55°C increases stringency and reduces background. After resuspending the pellet in 1% SDS, a conventional KCl-SDS precipitation protocol is carried out. ARK, however, omits the addition of BSA to avoid interference with fluorometric readout, as DNA

concentration is measured using Picogreen™. Also, as in the KCl-SDS assay, the final estimation of DPCs is determined by a ratio between the amount of protein-associated DNA and the amount of total DNA. Notably, after treating the final suspension with RNAse, no reduction in the protein background was observed, which indicates that guanidinium thiocyanate removes RNA effectively (122).

ARK's high sensitivity was demonstrated in different cellular contexts, including HeLa, HEK293T and SK6 cells, with either *wild-type* or SPRTN, NER or Fanconi anemia-deficient backgrounds (122). Prolonged treatment with physiological doses (≤75 μM) of formaldehyde led to a significant accumulation of DPCs only when SPRTN or FANCI were depleted (∼4-fold and ∼2-fold increase after 12 hours, respectively). However, at higher doses (400 μM, 2 hours), the ARK assay showed a DPC coefficient approximately five times higher than the conventional KCl-SDS precipitation assay.

Furthermore, the ARK assay was also employed to demonstrate the formaldehyde-dependent DPC accumulation in human cultured cells lacking the DUB USP11 (68) and to detect DPCs induced by 100 mM acetaldehyde (∼2.4-fold increase) (123). As well as non-specific DPCs, the ARK assay can also be used to detect enzymatic DPCs generated by treatment with CPT and etoposide. In comparison with RADAR, ARK demonstrated similar sensitivity across various CPT doses (85,122).

Advantages

The protocol is fast and can be completed in 7 hours. The stringent buffer in the ARK assay effectively lowers the background compared to the lysis with SDS alone. Moreover, omitting BSA improves the Picogreen™ read sensitivity compared to the conventional KCl-SDS precipitation.

Limitations

Particular proteins cannot be studied using the ARK assay. However, by removing KCl-SDS from the insoluble ARK DPC precipitate with acetone (124), the protein can be solubilised for gel analysis or mass spectrometry.

Modified comet assay

The comet assay is a single-cell electrophoresis-based method primarily designed for quantifying DNA breaks within individual cells. The procedure involves embedding cells in agarose on a microscope slide and subjecting them to an electric field. In this process, high molecular weight DNA remains immobile, but smaller DNA fragments (the result of DSBs) migrate towards the anode, creating a DNA comet tail (125). In alkaline conditions, the comet tail length is indicative of the overall DNA damage, including DNA DSBs and SSBs, alkali-labile sites, DNA-DNA crosslinks, and DPCs (126–130). To distinguish and specifically estimate DPCs in a sample that contains DNA–DNA crosslinks (131,132), Proteinase K is added to the slides. This enzyme digests proteins (Figure 10), enhancing the migration of DNA fragments damaged by DPCs during electrophoresis. This step is especially important for agents that induce both types of damage, such as formaldehyde.

Briefly, 5×10^4 cells per condition are mixed with 0.5–1% low melting agarose and spread onto two microscopy slides per condition. The slides are soaked in a buffer containing 1% Triton X-100 and 0.5% Sarcosyl (1 hour to O/N). Pro-

teinase K (0.8% w/v) is applied to only one of the two replicate slides. This treatment lyses the cells, unwinds DNA under alkaline conditions and digests crosslinked proteins. The slides are transferred to an electrophoresis tank filled with alkaline buffer and electrophoresis is performed for 25 min at 300 mA (25 V). Slides are allowed to air dry overnight. Then, DNA is stained with a fluorescent dye (*e.g.* SYBR gold) and visualised under a microscope. The DNA fluorescence in the comet tail ('tail moment') is measured using image analysis software (such as the Andor Komet) and expressed as a percentage of the total fluorescence from a single cell. The DPC levels are calculated as the difference between the tail moment with Proteinase K treatment and without it (50,131,132).

The specificity of the modified comet assay for DPCs was confirmed through a comparison of DNA damage caused by mitomycin C or cisplatin treatments, which mainly induce DNA-DNA crosslinks, and damage caused by formaldehyde (132). When Proteinase K was used, it led to an increase in the tail moment for cells treated with formaldehyde but had no such effect on the cells exposed to mitomycin C or cisplatin. Consequently, while the modified assay is less suitable for studying DNA–DNA crosslinkers (132), it has been effectively used to measure DPC levels after exposure to various common environmental contaminants, such as chromate (131), silver and titanium nanoparticles (133), and microwaves (134).

A further protocol modification enables the specific measurement of DPCs formed during the S phase (135). Cells are initially exposed to a short pulse of 5′-bromo-2′-deoxyuridine (BrdU), a thymidine analogue, to label newly synthesised DNA, before being treated with genotoxic agents. BrdU-labelled DNA is stained with an anti-BrdU primary antibody and an Alexa Fluor secondary antibody. The slides are then stained with ethidium bromide and subjected to imaging (135). This approach was employed to demonstrate that cells with a defective SPRTN protein accumulate DPCs during the S phase (50).

Advantages

This is the only fluorescence microscopy-based assay available to analyse non-enzymatic DPCs, alongside trapped in agarose DNA immunostaining (TARDIS) (136). It is a simple alternative to biochemical isolation methods and can complement their results.

Limitations

It is not possible to detect a crosslinked protein of interest using this assay, as with other DNA-targeted methods. Moreover, this assay is less sensitive than other methods, therefore it might be necessary to use a higher concentration of the crosslinking agent to detect changes (50).

Section 3. Detection of chemotherapeutically relevant DPCs

Immunodetection of DPCs formed by TOP1 with the TOP1cc monoclonal antibody

In order to relax supercoiled DNA, TOP1 carries out an isoenergetic reaction in which DNA cleavage is coupled with covalent bond formation between Tyr723 of TOP1 and the DNA substrate. The TOP1-DPC thus formed, also called TOP cleavage complex or TOP1cc, are of therapeutic importance because treatment with TOP1 poisons, such as CPT and related compounds, causes them to persist and promote cytotoxic-
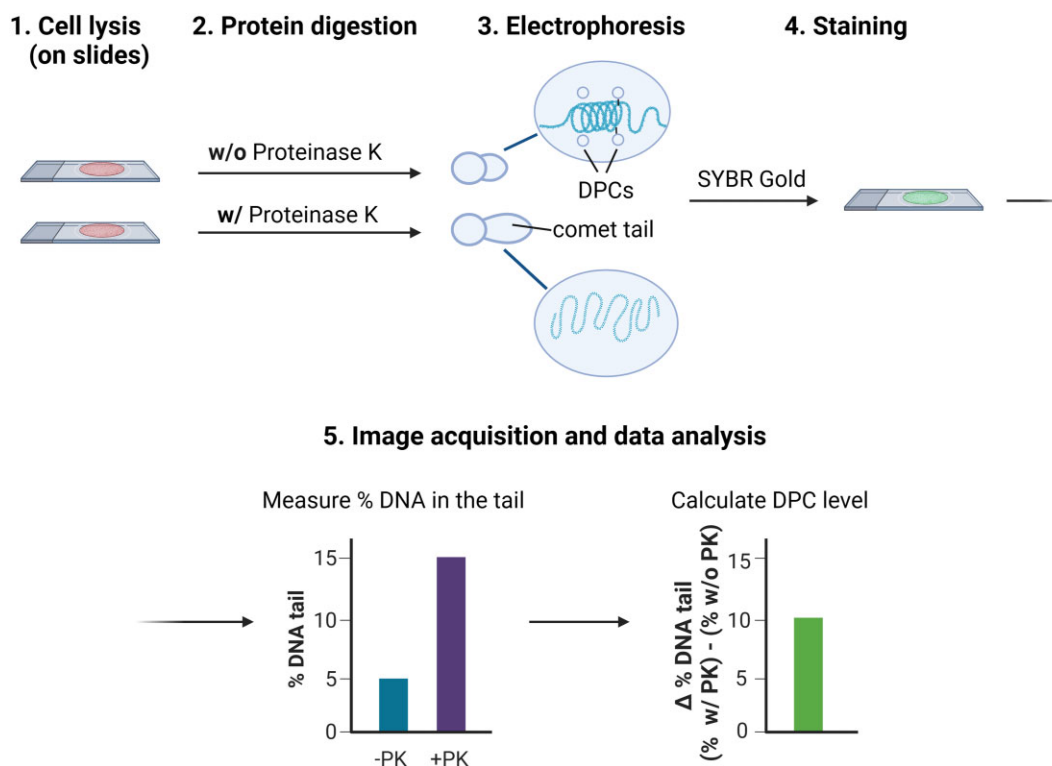
Figure 10. Modified comet assay. Schematic of the modified comet assay according to (131).

ity. They can be quantified by a variety of assays, including alkaline elution (137), ICE (80) and RADAR (85). However, detection was revolutionised by the development of the monoclonal antibody RRID:AB_2756354 which specifically recognises TOP1cc but not free TOP1 or DNA (138).

The TOP1cc antibody enables direct localisation and co-localisation analyses of TOP1cc without the need for cell lysis or DPC purification, and it is suitable for immunoblotting, flow cytometry and immunofluorescence. For immunofluorescence assays (Figure 11), cells are grown on coverslips, then treated, fixed, permeabilised and blocked following standard methods. To render DPCs more accessible, coverslips are further incubated in 1% SDS (5 min) and quenched with 0.1% Triton X-100. The slides are then stained with the TOP1cc antibody (1 hour, and optionally co-stained with additional primary antibodies), washed (0.1% Triton X-100) and treated with Alexa Fluor secondary antibodies. Images are acquired on a confocal microscope and TOP1cc foci or maximal mean fluorescence is quantified (see Methods).

The utilisation of the TOP1cc antibody has advanced our understanding of how TOP1-DPCs form and how they are repaired. We have learned that not only CPT and its analogues trap TOP1 on DNA, but several other genotoxic agents as well, including indenoisoquinolines, actinomycin D, and cisplatin (138). Immunofluorescence analysis has shown that TOP1cc foci colocalise with DNA damage markers, with kinetics suggesting that treatment with TOP1 poisons causes replication forks to stall, then undergo conversion to DSBs due to DNA replication fork collapse (4). Interestingly, TOP1cc foci do not co-localise with DNA damage signalling markers, suggesting that they induce secondary and distant damage on DNA (138). Immunofluorescence studies using the TOP1cc antibody have also confirmed the involvement of several key factors in TOP1cc repair, including tyrosyl DNA-phosphodiesterase 1 (TDP1) (139), the proteases SPRTN (67) and FAM111A (8), the unfoldase p97/VCP and its cofactor TEX264 (65), and confirmed that TOP1ccs are primed for proteasomal degradation by sequential SUMOylation and ubiquitylation (42,140). In addition, using the TOP1cc antibody in combination with CUT&RUN genomic sequencing has enabled the mapping of TOP1 hotspots, which were found to localise mainly in intronic regions (28.9%) and in promoters (31.6%), opening new avenues to characterise fine regulation of protein expression (141).

The TOP1cc antibody also has clinical utility. Its high sensitivity has enabled measurement of TOP1cc repair kinetics following exposure to low CPT doses (65), and has made it possible to predict the sensitivity of acute myeloid leukaemia cells to TOP1 poisons (142). The use of the TOP1cc antibody as a correlating biomarker is currently in a phase II clinical trial (*ClinicalTrials.gov NCT03289910*) (142). Furthermore, T-cells derived from individuals with chronic viral infections (hepatitis C, hepatitis B, HIV) were shown to accumulate TOP1cc foci, a finding with implications for cancer patients with these chronic infections, as the use of CPT analogues can lead to immunodeficiency (143). Additionally, endogenous TOP1ccs accumulate in Huntington's disease cells and in SPRTN hypomorphic mice liver cells, potentially exacerbating genomic instability and liver cancer, respectively (144,145).

Advantages
The TOP1cc monoclonal antibody is specific and does not cross-react with etoposide-stabilised TOP2-DPC (138). It is very sensitive and detects TOP1cc in untreated cells (65,139) and in cells treated at low CPT concentrations: 10–30 nM for immunoblotting, 25 nM for immunofluorescence and 16 nM
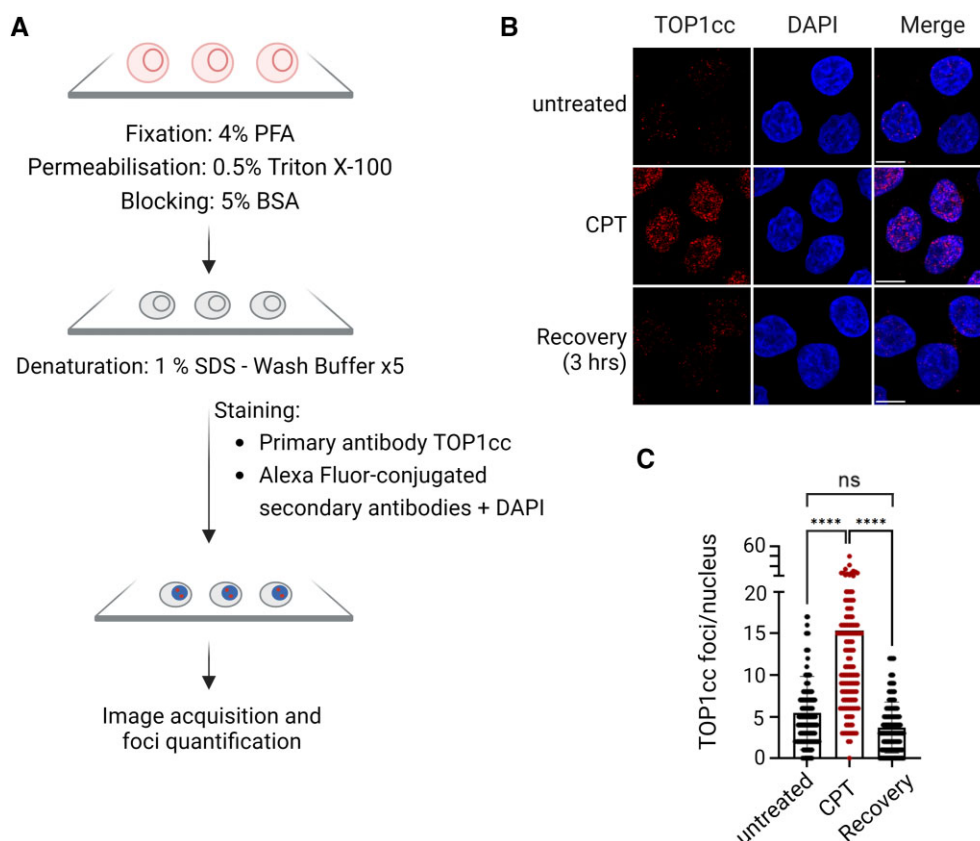
Figure 11. Detection of TOP1cc by immunofluorescence. (**A**) Schematic of the TOP1cc immunofluorescence protocol according to (138). (**B**) TOP1cc immunofluorescence of HeLa cells either untreated, treated with camptothecin (CPT; 50 nM, 1 hour), or 3 hours recovery after treatment. TOP1cc were stained with the primary antibody (Millipore Cat# MABE1084, RRID:AB_2756354) and the Alexa Fluor 555 secondary antibody. Scale bar, 10 μm. (**C**) Count of TOP1cc foci using Cell Profiler™. Significance determined by two-way ANOVA test. PFA: paraformaldehyde. See Supplementary material for raw data.

for flow cytometry (138). The antibody has been used in a variety of cell lines, including A549 (138), RPE-1 (65), myeloid cells (142), rhabdomyosarcoma cells (139), MEFs (145), as well as in astrocytes derived from mouse brain (146). It can be combined with other antibodies without disturbance to co-stained proteins such as γ-H2AX (138). Using cell imaging to visualise TOP1ccs in cells provides spatial resolution that cannot be achieved by biochemical analysis alone.

Limitations

Intense background and non-specific staining may be exhibited by some cell types or staining conditions. More washes or the use of a different secondary antibody may improve results. For example, we noticed that, in HeLa cells, donkey Alexa Fluor 555 secondary antibody (RRID AB_2536180) produced a better signal than goat Alexa Fluor 488 (RRID AB_2534069). When growing loosely attached cells in coverslips lacking a poly-L-lysine coating, the denaturing step of SDS can be problematic.

Visualisation of PARP1-DPC on DNA fibres

Poly(ADP-ribose) polymerase 1 (PARP1) is an enzyme that acts as a sensor of various forms of DNA damage (147,148). It catalyses poly-ADP-ribosylation of itself and other proteins to facilitate the recruitment of downstream repair factors, most notably during single-strand break repair, base excision repair (BER) and Okazaki fragment processing. PARP is a ther-

apeutic target, and PARP inhibitors (olaparib, rucaparib, niraparib, and talazoparib) have been approved for the treatment of prostate, ovarian, breast and pancreatic cancer, targeting a synthetic lethal effect between PARP inhibition and deficiencies in homologous recombination (149,150). PARP inhibitors have been shown to induce the formation of a complex resembling a DPC, although there was some debate about whether these were true covalent adducts or tightly 'trapped' complexes.

PARP1-DPC were initially studied *in vitro* using radiolabelled synthesised DNA substrates, allowing detection of crosslinked PARP1 by phosphor-imaging, and these results were complemented by clear evidence for formation of PARP1-DPC during BER (151–153). PARP1-DPC have also been shown to form at regions rich in CpG dinucleotides in response to damage caused by culture in 5-aza-dC, using the RADAR-based method, Adduct-Seq (18,32,97,99). Questions were raised regarding the role of PARP inhibition in PARP1-DPC formation, particularly whether PARP inhibition enhanced PARP1 crosslinking, and whether PARP inhibition combined with methanesulfonate (MMS) had an effect greater than MMS treatment alone (152–154).

A technique (Figure 12) based on a classical DNA fibre spreading (155,156) incorporates PARP1 immunostaining to enable direct visualisation of PARP1-DPCs on DNA fibres (152). This technique is versatile and can provide valuable insights into how and where PARP1-DPCs are formed. After treatment, cells are harvested by trypsinisation, resus-
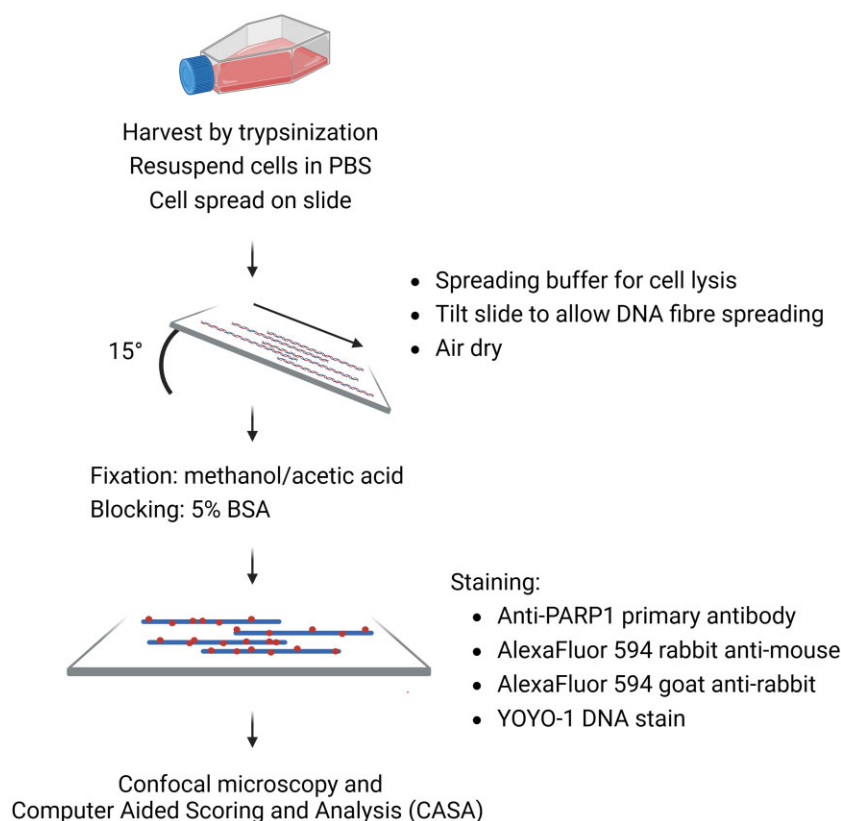
Figure 12. Visualisation of PARP1-DPC on DNA fibres. Schematic of the protocol as described in (152).

pended in ice-cold PBS, and spread onto silane-prep slides. After partial drying by evaporation, a spreading buffer containing 0.5% SDS is added to lyse the cells and gently remove non-covalently bound proteins. Slides are tilted 15° to promote the lysate to move down the slide and DNA fibres spreading. They are then air-dried before, fixed in 3:1 methanol/acetic acid, and stored at –20°C overnight. The immunostaining process begins with washing the slides in PBST buffer (0.1% Tween-20), followed by PBS, and blocking in 5% BSA. Immunostaining is carried out using an anti-PARP1 antibody (*RRID AB_394009*), followed by further washing in PBS and a second blocking step in 5% BSA. Staining is then carried out with secondary antibody (AlexaFluor 594 rabbit anti-mouse) and tertiary antibody (AlexaFluor 594 goat anti-rabbit). YOYO-1 is used to stain DNA before the slides are mounted. DNA fibres are imaged using a confocal microscope, and Computer Aided Scoring and Analysis (CASA) software (157) is used to quantify PARP1 levels per Mb of DNA. Typically, over 600 fibres are analysed per condition (152) thereby generating robust data. While conventional DNA fibre spreading techniques use thymidine analogues like 5-bromo-2′-deoxyuridine (BrdU), 5-chloro-2′-deoxyuridine (CldU) and 5-iodo-2′-deoxyuridine (IdU) to pulse-label replication sites, this requires an HCl DNA denaturation step before immunolabelling which is not necessary for the above protocol. In principle, 5-ethynyl-2′-deoxyuridine (EdU) could be incorporated into DNA and labelled with a photostable AlexaFluor dye via a click reaction. This modification would allow the visualisation of PARP1-DPCs at replication sites with minimal modification to the established protocol.

Advantages

Each sample requires only $2–3 \times 10^3$ cells (158), several orders of magnitude less than what is necessary for biochemical analogous. Samples may be characterised at a single-cell level, with individual fibres being analysed. Visualisation of PARP1-DPCs directly on DNA fibres enables spatial characterisation, for example, to observe signal clustering. The inclusion of other markers enables assessment of PARP1 localisation at DNA damage sites, just as a similar DNA fibre approach using a fluorescently-labelled aldehyde reactive probe was used to visualise abasic sites formed during BER (159).

Limitations

The quantification of PARP1-DPCs using the CASA software relies on the assumption that DNA stretches uniformly, providing a direct conversion between distance and Mb of DNA. However, the described DNA spreading technique is prone to the non-uniform spreading of DNA, and frequent crossing of fibres occurs, making image acquisition and analysis more time-consuming (158). The DNA combing technique, which uses specialised equipment that stretches fibres at a constant rate of 2 kb/μm, offers more accurate DNA tract length determination (158,160,161), but it involves slower preparations and includes a Proteinase K digestion step, making it incompatible with visualising DNA–protein crosslinks. Another limitation of this technique is the absence of essential controls to demonstrate whether the detected PARP1 is covalently linked to DNA. While it has been suggested that histone core proteins are removed from fibres, leaving only covalently bound proteins (141), this assertion has not been conclusively

demonstrated. Nevertheless, the technique remains valid for visualising PARP inhibitors-induced trapped PARP1 (16,162).

Conclusion and further perspectives

The proteins that make up DPCs display a diverse range of sizes and biochemical properties, which significantly influence the outcomes when isolating DPCs using different methods. Although most protein-targeted methods consistently reveal that the majority of DPCs comprise only a few abundant proteins, the observed variations in composition and occurrence in the DPC pools can be attributed to less represented protein components. Such variability poses challenges when comparing and interpreting data obtained from different isolation methods.

The methods described here have their advantages and limitations (Table 1). Researchers must carefully consider these factors when designing experiments and interpreting results, as the method choice would ultimately depend on the specific experimental requirements. For example, when studying the total DPC pool induced by a treatment, both protein-targeted and DNA-targeted based methods are valuable. Protein-targeted methods are particularly useful when studying specific DPCs or their post-translational modifications, where various subsequent approaches facilitate targeted detection. Protein-targeted methods also enable downstream applications such as mass spectrometry or sequencing. When dealing with low-abundance proteins or low starting material, more sensitive methods should be chosen, usually relying on optimised detection systems (such as specific antibodies like anti-TOP1cc) or employing amplification techniques. Whenever possible, using two or more techniques is recommended to consolidate the results.

A common challenge across all DPC preparations is dealing with proteins that bind tightly to DNA, albeit not covalently. Lysis methods with enhanced stringency, especially the inclusion of chaotropic salts, will minimise the presence of contaminating non-crosslinked proteins, which in turn reduces the background noise and enhances the sensitivity of the assay. For example, additional lysis steps were implemented in the ARK method to improve the precipitation of the KCl-SDS technique. The visualisation of DPCs in intact cells is also challenging. Currently, cell lysis is necessary for the detection of all DPCs except TOP1, where a monoclonal has been developed specifically to detect the covalently bound protein. Visualising DPCs in intact cells provides valuable information on localisation, such as whether they cluster or migrate, and on interactions with other factors. Developing a 'DPC reporter' and/or identifying cellular markers for microscopy screenings (akin to γH2Ax for DSBs) would significantly advance the field. Moreover, large-scale screening methods, such as those for identifying DPC repair factors, are currently challenging and time-consuming, impeding our ability to gather comprehensive data on DPCs. Therefore, there is clear potential for the development of innovative techniques to overcome these limitations and enhance our understanding of DPCs.

Materials and methods

Immunodetection of DNA topoisomerase 1 cleavage complexes

HeLa cells (CCL-2) obtained from ATCC were grown on coverslips and treated for 1 hour with 50 nM CPT. Recovery was performed in growing media: DMEM containing 10% foetal bovine serum (FBS) and 5% Penicillin/Streptomycin. Cells were fixed in 4% formaldehyde in PBS for 15 min at room temperature; washed with PBS; permeabilised in 0.5% Triton X-100 in PBS for 15 min at 4 °C; and blocked in 5% BSA/PBS for 1 h at 37°C. To render the DNA–protein crosslinks more accessible to the antibody, the coverslips were incubated in 1% SDS/PBS at room temperature (RT) for 5 min, then washed five times with a wash buffer (0.1% BSA, 0.1% Triton X-100 in PBS) and twice with PBS. Incubation with TOP1cc primary antibody (Millipore Cat# MABE 1084, RRID AB_2756354) diluted at 1:250 in 2.5% BSA/PBS was performed for 1 hour at RT. Before subsequent incubation, coverslips were washed once with the wash buffer for 3 min with gentle shaking and twice with PBS. Incubation with another primary antibody can be performed at this step if needed. Coverslips were washed with PBS and incubated with secondary antibody (Thermo Fisher Scientific Cat# A-31570, RRID AB_2536180) and DAPI, both diluted at 1:500 in 2.5% BSA/PBS for 1 hour at room temperature. Coverslips were mounted onto slides using Fluoromount G (ThermoFisher). Images were taken using a Zeiss 710 LSM microscope utilising a Plan-apochromat 63x lens with a 1.4 NA and oil immersion. Images were collected sequentially to avoid any overlap between dyes whilst ensuring the same MBS filter set was maintained for all acquisitions. Images were gathered in 1024 × 1024 pixel format at approximately 55% Nyquist sampling, with full Nyquist sampling not being appropriate for this experiment. Images were acquired at 12-bit with 4× averaging being utilised to help with spurious noise within the images. Analysis was carried out using ImageJ Fiji and CellProfiler™. The pipeline used for TOP1cc quantification per cell in CellProfiler™ is provided in the source data. A Two-way ANOVA test was performed for the analysis.

Data availability

The data underlying this article are available in the article and in its online supplementary material.

Supplementary data

Supplementary Data are available at NAR Online.

Conflict of interest statement

None declared.

References

1. Champoux,J.J. (2001) DNA topoisomerases: structure, function, and mechanism. *Annu. Rev. Biochem.*, **70**, 369–413.
2. 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.
3. 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.
4. Pommier,Y. (2006) Topoisomerase I inhibitors: camptothecins and beyond. *Nat. Rev. Cancer*, **6**, 789–802.
5. Weickert,P. and Stingele,J. (2022) DNA–protein crosslinks and their resolution. *Annu. Rev. Biochem.*, **91**, 157–181.
6. Sun,Y., Saha,S., Wang,W., Saha,L.K., Huang,S.N. and Pommier,Y. (2020) Excision repair of topoisomerase DNA–protein crosslinks (TOP-DPC). *DNA Repair (Amst.)*, **89**, 102837.
7. Thompson,P.S. and Cortez,D. (2020) New insights into abasic site repair and tolerance. *DNA Repair (Amst.)*, **90**, 102866.
8. Kojima,Y. and Machida,Y.J. (2020) DNA–protein crosslinks from environmental exposure: mechanisms of formation and repair. *Environ. Mol. Mutagen.*, **61**, 716–729.
9. Vaz,B., Popovic,M. and Ramadan,K. (2017) DNA–protein crosslink proteolysis repair. *Trends Biochem. Sci*, **42**, 483–495.
10. Stingele,J., Bellelli,R. and Boulton,S.J. (2017) Mechanisms of DNA–protein crosslink repair. *Nat. Rev. Mol. Cell Biol.*, **18**, 563–573.
11. 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.
12. Connelly,J.C. and Leach,D.R. (2004) Repair of DNA covalently linked to protein. *Mol. Cell*, **13**, 307–316.
13. Ide,H., Shoulkamy,M.I., Nakano,T., Miyamoto-Matsubara,M. and Salem,A.M. (2011) Repair and biochemical effects of DNA–protein crosslinks. *Mutat. Res.*, **711**, 113–122.
14. Stingele,J. and Jentsch,S. (2015) DNA–protein crosslink repair. *Nat. Rev. Mol. Cell Biol.*, **16**, 455–460.
15. Hiroshi,I., Toshiaki,N., Mahmoud,I.S. and Amir,M.H.S. (2015) In: Clark,C.C. (ed.) *Advances in DNA Repair*. IntechOpen, Rijeka, pp. Ch. 2.
16. Murai,J., Huang,S.Y., Das,B.B., Renaud,A., Zhang,Y., Doroshow,J.H., Ji,J., Takeda,S. and Pommier,Y. (2012) Trapping of PARP1 and PARP2 by clinical PARP inhibitors. *Cancer Res.*, **72**, 5588–5599.
17. Murai,J., Huang,S.Y., Renaud,A., Zhang,Y., Ji,J., Takeda,S., Morris,J., Teicher,B., Doroshow,J.H. and Pommier,Y. (2014) Stereospecific PARP trapping by BMN 673 and comparison with olaparib and rucaparib. *Mol. Cancer Ther.*, **13**, 433–443.
18. Kiianitsa,K., Zhang,Y. and Maizels,N. (2020) Treatment of human cells with 5-aza-dC induces formation of PARP1-DNA covalent adducts at genomic regions targeted by DNMT1. *DNA Repair (Amst.)*, **96**, 102977.
19. Barker,S., Weinfeld,M. and Murray,D. (2005) DNA–protein crosslinks: their induction, repair, and biological consequences. *Mutat. Res.*, **589**, 111–135.
20. Tretyakova,N.Y., Groehler,A.t. and Ji,S. (2015) DNA–protein cross-links: formation, structural identities, and biological outcomes. *Acc. Chem. Res.*, **48**, 1631–1644.
21. 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.
22. Xu,X., Muller,J.G., Ye,Y. and Burrows,C.J. (2008) DNA–protein cross-links between guanine and lysine depend on the mechanism of oxidation for formation of C5 vs C8 guanosine adducts. *J. Am. Chem. Soc.*, **130**, 703–709.
23. Kamps,J.J.A.G., Hopkinson,R.J., Schofield,C.J. and Claridge,T.D.W. (2019) How formaldehyde reacts with amino acids. *Communications Chemistry*, **2**, 126.
24. Hopkinson,R.J., Barlow,P.S., Schofield,C.J. and Claridge,T.D.W. (2010) Studies on the reaction of glutathione and formaldehyde using NMR. *Org. Biomol. Chem.*, **8**, 4915–4920.
25. John,T., Pires,E., Hester,S.S., Salah,E., Hopkinson,R.J. and Schofield,C.J. (2023) Formaldehyde reacts with N-terminal proline residues to give bicyclic aminals. *Commun. Chem.*, **6**, 12.
26. Shishodia,S., Zhang,D., El-Sagheer,A.H., Brown,T., Claridge,T.D.W., Schofield,C.J. and Hopkinson,R.J. (2018) NMR analyses on N-hydroxymethylated nucleobases – implications for formaldehyde toxicity and nucleic acid demethylases. *Org. Biomol. Chem.*, **16**, 4021–4032.
27. Loeber,R., Rajesh,M., Fang,Q., Pegg,A.E. and Tretyakova,N. (2006) Cross-linking of the human DNA repair protein O6-alkylguanine DNA alkyltransferase to DNA in the presence of 1,2,3,4-diepoxybutane. *Chem. Res. Toxicol.*, **19**, 645–654.
28. Ming,X., Groehler,A.t., Michaelson-Richie,E.D., Villalta,P.W., Campbell,C. and Tretyakova,N.Y. (2017) Mass spectrometry based proteomics study of cisplatin-induced DNA–protein cross-linking in Human fibrosarcoma (HT1080) cells. *Chem. Res. Toxicol.*, **30**, 980–995.
29. Gajewski,E. and Dizdaroglu,M. (1990) Hydroxyl radical induced cross-linking of cytosine and tyrosine in nucleohistone. *Biochemistry*, **29**, 977–980.
30. 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**, 63–77.
31. Glumac,M., Polovic,M., Batel,A., Gelemanovic,A., Macek,B., Velic,A. and Marinovic-Terzic,I. (2023) SPRTN-dependent DPC degradation precedes repair of damaged DNA: a proof of concept revealed by the STAR assay. *Nucleic Acids Res.*, **51**, e35.
32. 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.
33. Weickert,P., Li,H.Y., Gotz,M.J., Durauer,S., Yaneva,D., Zhao,S., Cordes,J., Acampora,A.C., Forne,I., Imhof,A., *et al.* (2023) SPRTN patient variants cause global-genome DNA–protein crosslink repair defects. *Nat. Commun.*, **14**, 352.
34. Gherezghiher,T.B., Ming,X., Villalta,P.W., Campbell,C. and Tretyakova,N.Y. (2013) 1,2,3,4-Diepoxybutane-induced DNA–protein cross-linking in human fibrosarcoma (HT1080) cells. *J. Proteome Res.*, **12**, 2151–2164.
35. Groehler,A.t., Villalta,P.W., Campbell,C. and Tretyakova,N. (2016) Covalent DNA–protein cross-linking by phosphoramide mustard and nornitrogen mustard in Human cells. *Chem. Res. Toxicol.*, **29**, 190–202.
36. Michaelson-Richie,E.D., Ming,X., Codreanu,S.G., Loeber,R.L., Liebler,D.C., Campbell,C. and Tretyakova,N.Y. (2011) Mechlorethamine-induced DNA–protein cross-linking in human fibrosarcoma (HT1080) cells. *J. Proteome Res.*, **10**, 2785–2796.

37. Kiianitsa,K. and Maizels,N. (2020) The "adductome": a limited repertoire of adducted proteins in human cells. *DNA Repair (Amst.)*, **89**, 102825.

38. 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.

39. 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.

40. 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.

41. 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.

42. Sun,Y., Saha,L.K., Saha,S., Jo,U. and Pommier,Y. (2020) Debulking of topoisomerase DNA–protein crosslinks (TOP-DPC) by the proteasome, non-proteasomal and non-proteolytic pathways. *DNA Repair (Amst.)*, **94**, 102926.

43. de Graaf,B., Clore,A. and McCullough,A.K. (2009) Cellular pathways for DNA repair and damage tolerance of formaldehyde-induced DNA–protein crosslinks. *DNA Repair (Amst.)*, **8**, 1207–1214.

44. 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.

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

46. Hiom,K. (2014) SPRTN is a new player in an old story. *Nat. Genet.*, **46**, 1155–1157.

47. Ramadan,K., Halder,S., Wiseman,K. and Vaz,B. (2017) Strategic role of the ubiquitin-dependent segregase p97 (VCP or Cdc48) in DNA replication. *Chromosoma*, **126**, 17–32.

48. 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.

49. Halder,S., Torrecilla,I., Burkhalter,M.D., Popovic,M., Fielden,J., Vaz,B., Oehler,J., Pilger,D., Lessel,D., Wiseman,K., *et al.* (2019) SPRTN protease and checkpoint kinase 1 cross-activation loop safeguards DNA replication. *Nat. Commun.*, **10**, 3142.

50. 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.

51. 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.

52. 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.

53. 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.

54. Kim,M.S., Machida,Y., Vashisht,A.A., Wohlschlegel,J.A., Pang,Y.P. and Machida,Y.J. (2013) Regulation of error-prone translesion synthesis by Spartan/C1orf124. *Nucleic Acids Res.*, **41**, 1661–1668.

55. 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.

56. Fielden,J., Ruggiano,A., Popovic,M. and Ramadan,K. (2018) DNA protein crosslink proteolysis repair: from yeast to premature ageing and cancer in humans. *DNA Repair (Amst.)*, **71**, 198–204.

57. 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.

58. 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.

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

60. Krastev,D.B., Li,S., Sun,Y., Wicks,A.J., Hoslett,G., Weekes,D., Badder,L.M., Knight,E.G., Marlow,R., Pardo,M.C., *et al.* (2022) The ubiquitin-dependent ATPase p97 removes cytotoxic trapped PARP1 from chromatin. *Nat. Cell Biol.*, **24**, 62–73.

61. Liu,J.C.Y., Kuhbacher,U., Larsen,N.B., Borgermann,N., Garvanska,D.H., Hendriks,I.A., Ackermann,L., Haahr,P., Gallina,I., Guerillon,C., *et al.* (2021) Mechanism and function of DNA replication-independent DNA–protein crosslink repair via the SUMO-RNF4 pathway. *EMBO J.*, **40**, e107413.

62. Sun,Y., Chen,J., Huang,S.N., Su,Y.P., Wang,W., Agama,K., Saha,S., Jenkins,L.M., Pascal,J.M. and Pommier,Y. (2021) PARylation prevents the proteasomal degradation of topoisomerase I DNA–protein crosslinks and induces their deubiquitylation. *Nat. Commun.*, **12**, 5010.

63. 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

64. 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.

65. 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.

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

67. 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.

68. Perry,M., Biegert,M., Kollala,S.S., Mallard,H., Su,G., Kodavati,M., Kreiling,N., Holbrook,A. and Ghosal,G. (2021) USP11 mediates repair of DNA–protein cross-links by deubiquitinating SPRTN metalloprotease. *J. Biol. Chem.*, **296**, 100396.

69. Zhao,S., Kieser,A., Li,H.Y., Reinking,H.K., Weickert,P., Euteneuer,S., Yaneva,D., Acampora,A.C., Gotz,M.J., Feederle,R., *et al.* (2021) A ubiquitin switch controls autocatalytic

inactivation of the DNA–protein crosslink repair protease SPRTN. *Nucleic Acids Res.*, **49**, 902–915.

70. Leng,X. and Duxin,J.P. (2022) Targeting DNA–protein crosslinks via post-translational modifications. *Front Mol. Biosci.*, **9**, 944775.

71. Barker,S., Murray,D., Zheng,J., Li,L. and Weinfeld,M. (2005) A method for the isolation of covalent DNA–protein crosslinks suitable for proteomics analysis. *Anal. Biochem.*, **344**, 204–215.

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

73. Qin,H. and Wang,Y. (2009) Exploring DNA-binding proteins with in vivo chemical cross-linking and mass spectrometry. *J. Proteome Res.*, **8**, 1983–1991.

74. Celis,J.E., Fink,M. and Kaltoft,K. (1976) On the use of ultraviolet light to study protein-DNA crosslinking. *Nucleic Acids Res.*, **3**, 1065–1071.

75. Meselson,M., Stahl,F.W. and Vinograd,J. (1957) Equilibrium sedimentation of macromolecules in density gradients. *Proc. Natl. Acad. Sci. U.S.A.*, **43**, 581–588.

76. Gueranger,Q., Kia,A., Frith,D. and Karran,P. (2011) Crosslinking of DNA repair and replication proteins to DNA in cells treated with 6-thioguanine and UVA. *Nucleic Acids Res.*, **39**, 5057–5066.

77. Muller,M.T., Pfund,W.P., Mehta,V.B. and Trask,D.K. (1985) Eukaryotic type I topoisomerase is enriched in the nucleolus and catalytically active on ribosomal DNA. *EMBO J.*, **4**, 1237–1243.

78. Subramanian,D., Furbee,C.S. and Muller,M.T. (2001) ICE bioassay. Isolating in vivo complexes of enzyme to DNA. *Methods Mol. Biol.*, **95**, 137–147.

79. Subramanian,D., Rosenstein,B.S. and Muller,M.T. (1998) Ultraviolet-induced DNA damage stimulates topoisomerase I-DNA complex formation in vivo: possible relationship with DNA repair. *Cancer Res.*, **58**, 976–984.

80. Subramanian,D., Kraut,E., Staubus,A., Young,D.C. and Muller,M.T. (1995) Analysis of topoisomerase I/DNA complexes in patients administered topotecan. *Cancer Res.*, **55**, 2097–2103.

81. Nitiss,J.L., Soans,E., Rogojina,A., Seth,A. and Mishina,M. (2012) Topoisomerase assays. *Curr. Protoc. Pharmacol.*, **Chapter 3**, Unit 3.3.

82. Nakano,T., Moriwaki,T., Tsuda,M., Miyakawa,M., Hanaichi,Y., Sasanuma,H., Hirota,K., Kawanishi,M., Ide,H., Tano,K., *et al.* (2022) SPRTN and TDP1/TDP2 independently suppress 5-Aza-2'-deoxycytidine-induced genomic instability in Human TK6 cell line. *Chem. Res. Toxicol.*, **35**, 2059–2067.

83. Meng,L.H., Meng,Z., Miao,Z.H., Veenstra,T.D. and Pommier,Y. (2008) Cytokeratin-RNA cross-linking mediated by the antitumor aminoflavone, 5-amino-2,3-fluorophenyl-6,8-difluoro-7-methyl-4H-1-benzopyran-4-one. *J. Pharmacol. Exp. Ther.*, **325**, 674–680.

84. Shoulkamy,M.I., Nakano,T., Ohshima,M., Hirayama,R., Uzawa,A., Furusawa,Y. and Ide,H. (2012) Detection of DNA–protein crosslinks (DPCs) by novel direct fluorescence labeling methods: distinct stabilities of aldehyde and radiation-induced DPCs. *Nucleic Acids Res.*, **40**, e143.

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

86. Kiianitsa,K. and Maizels,N. (2014) Ultrasensitive isolation, identification and quantification of DNA–protein adducts by ELISA-based RADAR assay. *Nucleic Acids Res.*, **42**, e108.

87. Dheekollu,J., Wiedmer,A., Ayyanathan,K., Deakyne,J.S., Messick,T.E. and Lieberman,P.M. (2021) Cell-cycle-dependent EBNA1-DNA crosslinking promotes replication termination at oriP and viral episome maintenance. *Cell*, **184**, 643–654.

88. Meroni,A., Grosser,J., Agashe,S., Ramakrishnan,N., Jackson,J., Verma,P., Baranello,L. and Vindigni,A. (2022) NEDDylated Cullin 3 mediates the adaptive response to topoisomerase 1 inhibitors. *Sci. Adv.*, **8**, eabq0648.

89. Zhang,H., Xiong,Y., Su,D., Wang,C., Srivastava,M., Tang,M., Feng,X., Huang,M., Chen,Z. and Chen,J. (2022) TDP1-independent pathways in the process and repair of TOP1-induced DNA damage. *Nat. Commun.*, **13**, 4240.

90. Marzi,L., Sun,Y., Huang,S.N., James,A., Difilippantonio,S. and Pommier,Y. (2020) The Indenoisoquinoline LMP517: a novel antitumor agent targeting both TOP1 and TOP2. *Mol. Cancer Ther.*, **19**, 1589–1597.

91. Jo,U., Murai,Y., Agama,K.K., Sun,Y., Saha,L.K., Yang,X., Arakawa,Y., Gayle,S., Jones,K., Paralkar,V., *et al.* (2022) TOP1-DNA trapping by Exatecan and combination therapy with ATR inhibitor. *Mol. Cancer Ther.*, **21**, 1090–1102.

92. Pan,M., Wright,W.C., Chapple,R.H., Zubair,A., Sandhu,M., Batchelder,J.E., Huddle,B.C., Low,J., Blankenship,K.B., Wang,Y., *et al.* (2021) The chemotherapeutic CX-5461 primarily targets TOP2B and exhibits selective activity in high-risk neuroblastoma. *Nat. Commun.*, **12**, 6468.

93. Flor,A., Wolfgeher,D., Li,J., Hanakahi,L.A. and Kron,S.J. (2021) Lipid-derived electrophiles mediate the effects of chemotherapeutic topoisomerase I poisons. *Cell Chem Biol*, **28**, 776–787.

94. Saha,S., Sun,Y., Huang,S.N., Baechler,S.A., Pongor,L.S., Agama,K., Jo,U., Zhang,H., Tse-Dinh,Y.C. and Pommier,Y. (2020) DNA and RNA cleavage complexes and repair pathway for TOP3B RNA- and DNA–protein crosslinks. *Cell Rep.*, **33**, 108569.

95. Chen,X., Legrand,A.J., Cunniffe,S., Hume,S., Poletto,M., Vaz,B., Ramadan,K., Yao,D. and Dianov,G.L. (2018) Interplay between base excision repair protein XRCC1 and ALDH2 predicts overall survival in lung and liver cancer patients. *Cell Oncol (Dordr)*, **41**, 527–539.

96. Gao,Y., Guitton-Sert,L., Dessapt,J., Coulombe,Y., Rodrigue,A., Milano,L., Blondeau,A., Larsen,N.B., Duxin,J.P., Hussein,S., *et al.* (2023) A CRISPR-Cas9 screen identifies EXO1 as a formaldehyde resistance gene. *Nat. Commun.*, **14**, 381.

97. 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.

98. Li,J., Nagy,N., Liu,J., Gupta,S., Frisan,T., Hennig,T., Cameron,D.P., Baranello,L. and Masucci,M.G. (2021) The Epstein-Barr virus deubiquitinating enzyme BPLF1 regulates the activity of topoisomerase II during productive infection. *PLoS Pathog.*, **17**, e1009954.

99. Kuzin,V., Wiegard,A., Cameron,D.P. and Baranello,L. (2022) TOP1 CAD-seq: a protocol to map catalytically engaged topoisomerase 1 in human cells. *STAR Protoc*, **3**, 101581.

100. Sinha,D., Kiianitsa,K., Sherman,D.R. and Maizels,N. (2020) Rapid, direct detection of bacterial topoisomerase 1-DNA adducts by RADAR/ELISA. *Anal. Biochem.*, **608**, 113827.

101. O'Connor,P.M. and Fox,B.W. (1987) Comparative studies of DNA cross-linking reactions following methylene dimethanesulphonate and its hydrolytic product, formaldehyde. *Cancer Chemother. Pharmacol.*, **19**, 11–15.

102. Ruggiano,A., Vaz,B., Kilgas,S., Popovic,M., Rodriguez-Berriguete,G., Singh,A.N., Higgins,G.S., Kiltie,A.E. and Ramadan,K. (2021) The protease SPRTN and SUMOylation coordinate DNA–protein crosslink repair to prevent genome instability. *Cell Rep.*, **37**, 110080.

103. Wilhelm,L., Burmann,F., Minnen,A., Shin,H.C., Toseland,C.P., Oh,B.H. and Gruber,S. (2015) SMC condensin entraps chromosomal DNA by an ATP hydrolysis dependent loading mechanism in Bacillus subtilis. *eLife*, **4**, e06659.

104. Lin,X., Zhuang,Z. and Costa,M. (1992) Analysis of residual amino acid–DNA crosslinks induced in intact cells by nickel and chromium compounds. *Carcinogenesis*, **13**, 1763–1768.

105. Kumar,K. and Woolum,K. (2021) A novel reagent for radioiodine labeling of new chemical entities (NCEs) and biomolecules. *Molecules*, **26**, 4344.

106. Tsomides,T.J. and Eisen,H.N. (1993) Stoichiometric labeling of peptides by iodination on tyrosyl or histidyl residues. *Anal. Biochem.*, **210**, 129–135.

107. Zhuang,Z. and Costa,M. (1994) Development of an 125I-postlabeling assay as a simple, rapid, and sensitive index of DNA–protein cross-links. *Environ. Health Perspect.*, **102**, 301–304.

108. Lai,Y., Yu,R., Hartwell,H.J., Moeller,B.C., Bodnar,W.M. and Swenberg,J.A. (2016) Measurement of endogenous versus exogenous formaldehyde-induced DNA–protein crosslinks in animal tissues by stable isotope labeling and ultrasensitive mass spectrometry. *Cancer Res.*, **76**, 2652–2661.

109. Cloos,P.A., Christensen,J., Agger,K. and Helin,K. (2008) Erasing the methyl mark: histone demethylases at the center of cellular differentiation and disease. *Genes Dev.*, **22**, 1115–1140.

110. Hou,H. and Yu,H. (2010) Structural insights into histone lysine demethylation. *Curr. Opin. Struct. Biol.*, **20**, 739–748.

111. Morellato,A.E., Umansky,C. and Pontel,L.B. (2021) The toxic side of one-carbon metabolism and epigenetics. *Redox. Biol.*, **40**, 101850.

112. Dingler,F.A., Wang,M., Mu,A., Millington,C.L., Oberbeck,N., Watcham,S., Pontel,L.B., Kamimae-Lanning,A.N., Langevin,F., Nadler,C., *et al.* (2020) Two aldehyde clearance systems are essential to prevent lethal formaldehyde accumulation in mice and humans. *Mol. Cell*, **80**, 996–1012.

113. Heck,H.D., Casanova,M. and Starr,T.B. (1990) Formaldehyde toxicity–new understanding. *Crit. Rev. Toxicol.*, **20**, 397–426.

114. Serbyn,N., Bagdiul,I., Noireterre,A., Michel,A.H., Suhandynata,R.T., Zhou,H., Kornmann,B. and Stutz,F. (2021) SUMO orchestrates multiple alternative DNA–protein crosslink repair pathways. *Cell Rep.*, **37**, 110034.

115. Hecht,S.S. (1999) Tobacco smoke carcinogens and lung cancer. *J. Natl. Cancer Inst.*, **91**, 1194–1210.

116. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans (2008) IARC monographs on the evaluation of carcinogenic risks to humans. Volume 97. 1,3-butadiene, ethylene oxide and vinyl halides (vinyl fluoride, vinyl chloride and vinyl bromide). *IARC Monogr. Eval. Carcinog. Risks Hum.*, **97**, 3–471.

117. Morrow,N.L. (1990) The industrial production and use of 1,3-butadiene. *Environ. Health Perspect.*, **86**, 7–8.

118. Trask,D.K. and Muller,M.T. (1983) Biochemical characterization of topoisomerase I purified from avian erythrocytes. *Nucleic Acids Res.*, **11**, 2779–2800.

119. Trask,D.K., DiDonato,J.A. and Muller,M.T. (1984) Rapid detection and isolation of covalent DNA/protein complexes: application to topoisomerase I and II. *EMBO J.*, **3**, 671–676.

120. Liu,L.F., Rowe,T.C., Yang,L., Tewey,K.M. and Chen,G.L. (1983) Cleavage of DNA by mammalian DNA topoisomerase II. *J. Biol. Chem.*, **258**, 15365–15370.

121. 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.

122. Hu,Q., Klages-Mundt,N., Wang,R., Lynn,E., Kuma Saha,L., Zhang,H., Srivastava,M., Shen,X., Tian,Y., Kim,H., *et al.* (2020) The ARK assay is a sensitive and versatile method for the global detection of DNA–protein crosslinks. *Cell Rep.*, **30**, 1235–1245.

123. Matsuzaki,K., Kumatoriya,K., Tando,M., Kometani,T. and Shinohara,M. (2022) Polyphenols from persimmon fruit attenuate acetaldehyde-induced DNA double-strand breaks by scavenging acetaldehyde. *Sci. Rep.*, **12**, 10300.

124. Hager,D.A. and Burgess,R.R. (1980) Elution of proteins from sodium dodecyl sulfate-polyacrylamide gels, removal of sodium dodecyl sulfate, and renaturation of enzymatic activity: results with sigma subunit of Escherichia coli RNA polymerase, wheat germ DNA topoisomerase, and other enzymes. *Anal. Biochem.*, **109**, 76–86.

125. Ostling,O. and Johanson,K.J. (1984) Microelectrophoretic study of radiation-induced DNA damages in individual mammalian cells. *Biochem. Biophys. Res. Commun.*, **123**, 291–298.

126. Singh,N.P., McCoy,M.T., Tice,R.R. and Schneider,E.L. (1988) A simple technique for quantitation of low levels of DNA damage in individual cells. *Exp. Cell. Res.*, **175**, 184–191.

127. Hartmann,A. and Speit,G. (1995) Genotoxic effects of chemicals in the single cell gel (SCG) test with human blood cells in relation to the induction of sister-chromatid exchanges (SCE). *Mutat. Res.*, **346**, 49–56.

128. Speit,G. and Hartmann,A. (1995) The contribution of excision repair to the DNA effects seen in the alkaline single cell gel test (comet assay). *Mutagenesis*, **10**, 555–559.

129. Tice,R.R., Strauss,G.H. and Peters,W.P. (1992) High-dose combination alkylating agents with autologous bone-marrow support in patients with breast cancer: preliminary assessment of DNA damage in individual peripheral blood lymphocytes using the single cell gel electrophoresis assay. *Mutat. Res.*, **271**, 101–113.

130. Tice,R.R., Yager,J.W., Andrews,P. and Crecelius,E. (1997) Effect of hepatic methyl donor status on urinary excretion and DNA damage in B6C3F1 mice treated with sodium arsenite. *Mutat. Res.*, **386**, 315–334.

131. Merk,O. and Speit,G. (1998) Significance of formaldehyde-induced DNA–protein crosslinks for mutagenesis. *Environ. Mol. Mutagen.*, **32**, 260–268.

132. Merk,O. and Speit,G. (1999) Detection of crosslinks with the comet assay in relationship to genotoxicity and cytotoxicity. *Environ. Mol. Mutagen.*, **33**, 167–172.

133. Klingelfus,T., Disner,G.R., Voigt,C.L., Alle,L.F., Cestari,M.M. and Leme,D.M. (2019) Nanomaterials induce DNA–protein crosslink and DNA oxidation: a mechanistic study with RTG-2 fish cell line and Comet assay modifications. *Chemosphere*, **215**, 703–709.

134. Lagroye,I., Anane,R., Wettring,B.A., Moros,E.G., Straube,W.L., Laregina,M., Niehoff,M., Pickard,W.F., Baty,J. and Roti Roti,J.L. (2004) Measurement of DNA damage after acute exposure to pulsed-wave 2450 MHz microwaves in rat brain cells by two alkaline comet assay methods. *Int. J. Radiat. Biol.*, **80**, 11–20.

135. Morocz,M., Gali,H., Rasko,I., Downes,C.S. and Haracska,L. (2013) Single cell analysis of human RAD18-dependent DNA post-replication repair by alkaline bromodeoxyuridine comet assay. *PLoS One*, **8**, e70391.

136. Padget,K., Carr,R., Pearson,A.D., Tilby,M.J. and Austin,C.A. (2000) Camptothecin-stabilised topoisomerase I-DNA complexes in leukaemia cells visualised and quantified in situ by the TARDIS assay (trapped in agarose DNA immunostaining). *Biochem. Pharmacol.*, **59**, 629–638.

137. Hsiang,Y.H., Liu,L.F., Wall,M.E., Wani,M.C., Nicholas,A.W., Manikumar,G., Kirschenbaum,S., Silber,R. and Potmesil,M. (1989) DNA topoisomerase I-mediated DNA cleavage and cytotoxicity of camptothecin analogues. *Cancer Res.*, **49**, 4385–4389.

138. Patel,A.G., Flatten,K.S., Peterson,K.L., Beito,T.G., Schneider,P.A., Perkins,A.L., Harki,D.A. and Kaufmann,S.H. (2016) Immunodetection of human topoisomerase I-DNA covalent complexes. *Nucleic Acids Res.*, **44**, 2816–2826.

139. Liao,C., Beveridge,R., Hudson,J.J.R., Parker,J.D., Chiang,S.C., Ray,S., Ashour,M.E., Sudbery,I., Dickman,M.J. and El-Khamisy,S.F. (2018) UCHL3 Regulates topoisomerase-induced chromosomal break repair by controlling TDP1 proteostasis. *Cell Rep.*, **23**, 3352–3365.

140. Tsuda,M., Kitamasu,K., Kumagai,C., Sugiyama,K., Nakano,T. and Ide,H. (2020) Tyrosyl-DNA phosphodiesterase 2 (TDP2) repairs topoisomerase 1 DNA–protein crosslinks and 3'-blocking lesions in the absence of tyrosyl-DNA phosphodiesterase 1 (TDP1). *DNA Repair (Amst.)*, **91-92**, 102849.

141. Tan,Y., Yao,L., Gamliel,A., Nair,S.J., Taylor,H., Ohgi,K., Aggarwal,A.K. and Rosenfeld,M.G. (2023) Signal-induced enhancer activation requires Ku70 to read topoisomerase1-DNA covalent complexes. *Nat. Struct. Mol. Biol.*, **30**, 148–158.

142. Kohorst,M.A., Flatten,K.S., Peterson,K.L., Schneider,P.A., Correia,C., Pratz,K.W., Smith,B.D. and Kaufmann,S.H. (2018) Topoisomerase I-DNA covalent complexes in myeloid malignancies: a potential biomarker for Topoisomerase I inhibitor sensitivity. *Blood*, **132**, 5146–5146.

143. Ji,Y., Dang,X., Nguyen,L.N.T., Nguyen,L.N., Zhao,J., Cao,D., Khanal,S., Schank,M., Wu,X.Y., Morrison,Z.D., *et al.* (2019) Topological DNA damage, telomere attrition and T cell senescence during chronic viral infections. *Immun Ageing*, **16**, 12.

144. Palminha,N.M., Dos Santos Souza,C., Griffin,J., Liao,C., Ferraiuolo,L. and El-Khamisy,S.F. (2022) Defective repair of topoisomerase I induced chromosomal damage in Huntington's disease. *Cell. Mol. Life Sci.*, **79**, 160.

145. 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.

146. Aditi Downing,S.M., Schreiner,P.A., Kwak,Y.D., Li,Y., Shaw,T.I., Russell,H.R. and McKinnon,P.J. (2021) Genome instability independent of type I interferon signaling drives neuropathology caused by impaired ribonucleotide excision repair. *Neuron*, **109**, 3962–3979.

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

148. Tallis,M., Morra,R., Barkauskaite,E. and Ahel,I. (2014) Poly(ADP-ribosyl)ation in regulation of chromatin structure and the DNA damage response. *Chromosoma*, **123**, 79–90.

149. Farmer,H., McCabe,N., Lord,C.J., Tutt,A.N., Johnson,D.A., Richardson,T.B., Santarosa,M., Dillon,K.J., Hickson,I., Knights,C., *et al.* (2005) Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature*, **434**, 917–921.

150. Lord,C.J. and Ashworth,A. (2017) PARP inhibitors: synthetic lethality in the clinic. *Science*, **355**, 1152–1158.

151. 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.

152. 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.

153. 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.

154. 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.

155. Jackson,D.A. and Pombo,A. (1998) Replicon clusters are stable units of chromosome structure: evidence that nuclear organization contributes to the efficient activation and propagation of S phase in human cells. *J. Cell Biol.*, **140**, 1285–1295.

156. Merrick,C.J., Jackson,D. and Diffley,J.F. (2004) Visualization of altered replication dynamics after DNA damage in human cells. *J. Biol. Chem.*, **279**, 20067–20075.

157. Wang,Y., Chastain,P., Yap,P.T., Cheng,J.Z., Kaufman,D., Guo,L. and Shen,D. (2011) In: *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. IEEE, Chicago, IL, USA, pp. 1349–1352.

158. Quinet,A., Carvajal-Maldonado,D., Lemacon,D. and Vindigni,A. (2017) DNA Fiber analysis: mind the gap!. *Methods Enzymol.*, **591**, 55–82.

159. Chastain,P.D. 2nd, Nakamura,J., Rao,S., Chu,H., Ibrahim,J.G., Swenberg,J.A. and Kaufman,D.G. (2010) Abasic sites preferentially form at regions undergoing DNA replication. *FASEB J.*, **24**, 3674–3680.

160. Bianco,J.N., Poli,J., Saksouk,J., Bacal,J., Silva,M.J., Yoshida,K., Lin,Y.L., Tourriere,H., Lengronne,A. and Pasero,P. (2012) Analysis of DNA replication profiles in budding yeast and mammalian cells using DNA combing. *Methods*, **57**, 149–157.

161. Bensimon,A., Simon,A., Chiffaudel,A., Croquette,V., Heslot,F. and Bensimon,D. (1994) Alignment and sensitive detection of DNA by a moving interface. *Science*, **265**, 2096–2098.

162. Shen,Y., Aoyagi-Scharber,M. and Wang,B. (2015) Trapping poly(ADP-ribose) polymerase. *J. Pharmacol. Exp. Ther.*, **353**, 446–457.