Investigation and characterisation of novel poly(ADP-ribose)-binding proteins in *Dictyostelium discoideum*

A thesis submitted to the Board of the Medical Sciences Division, University of Oxford, in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Alasdair Gunn

St Hugh's College

Trinity term 2015

Word count: 46,112

Investigation and characterisation of poly(ADP-ribose)-binding proteins in *Dictyostelium discoideum*

Alasdair Gunn – St Hugh's College – Trinity 2015

Submitted for the degree of Doctor of Philosophy

The genome is under continual assault from endogenous and exogenous sources of DNA damage. The cell has therefore evolved a number of pathways to identify, signal, and ultimately repair DNA lesions. Poly(ADP-ribosyl)ation, the addition of multiple ADP-ribose moieties to proteins to form poly(ADP-ribose) (PAR) chains, has been implicated in several DNA repair pathways. One function of these PAR chains is to act as a scaffold for the recruitment of downstream repair proteins, suggesting the existence of protein domains that specifically bind PAR. Several of these domains have been identified, the best characterised being the PBZ and macro domains.

We utilise *in silico* genome-wide searches to identify novel proteins containing PAR-binding domains in the simple eukaryotic model organism *Dictyostelium discoideum*. We identity three proteins with unannotated macro domains in *Dictyostelium*: DNA ligase III (Lig3), an aprataxin-like protein (APL), which also contains a PBZ domain, and Q54R54. The macro domain of APL was found to be circularly permuted compared to the other human and *Dictyostelium* macro domains; however, structure prediction by homology modelling indicated that it had retained the structure of a classical macro domain. We performed *in vitro* PAR-binding assays that indicated that the macro domains of both Lig3 and APL bind to PAR chains. Lig3 is enriched on DNA following the infliction of base damage, indicating its role in the SSBR pathway. However, the dependence of this

pathway on PARylation has not been determined. In contrast, APL was found to be enriched on chromatin in response to DNA inter-strand cross-links and S-phase-associated double-strand breaks, which was reduced following mutation of both the PBZ and macro domains of APL. Furthermore, APL was found to be mono-ubiquitinated in response to DNA inter-strand cross-links, an event that is dependent on its macro domain region. These data indicate a role for APL in the repair of DNA inter-strand cross-links, or S-phase associated DNA damage, which would not be predicted from its homology to aprataxin, thereby suggesting a protein with novel characteristics in the DDR in *Dictyostelium*.

Acknowledgements

Firstly, I would like to thank my supervisor Nick Lakin, for his support and guidance throughout my DPhil, particularly in the reading of drafts of this thesis. I would also like to thank Catherine Pears for her support and counsel too, alongside Chris Ponting and Luis Sanchez-Pulido, for their instruction in the field of bioinformatics. Furthermore, the work in this thesis would not have been completed as fully without support from Ivan Ahel (APL macro domain experiments), Ben Thomas (mass spectrometry) and KJ Patel (provision of FA-deficient strains).

Importantly, I would like to pay tribute to Anne-Marie Couto, Duen Wei Hsu, Amanda Unsworth, Eric Liang, and Nick Crump, whose advice, teaching, support and friendship throughout my time in the lab I am most thankful for. I would also like to thank Peggy Paschke and Mehera Emrich for performing some of the experiments outlined in this thesis. Furthermore, I would like to thank my colleagues and friends in the Lakin, Pears, Cohn and Mahadevan labs, with special mentions to Alina Rakhimova, Lena Kolb, George Ronson, Laura Mathews, Fu-sheng Chang, Seiji Ura, Iza Bombik, Ellie Warren and Huajiang Xiong.

I would like to thank the friends who I shared this journey with: my DTC colleagues, members of the St Hugh's MCR, and housemates from 20 Hurst Street, 216b Abingdon Road, and 126&127 Magdalen Road. My DPhil would have been a much worse place without you. I would also like to thank my friends from my undergraduate and school days, who I know I can always count on. I am truly thankful to have been surrounded by such amazing people.

Some of the people mentioned above deserve special recognition, as they were there to help me through the darkest passage of my life, and without whom, this DPhil may never

have been completed. I therefore extend my most heartfelt thanks to Rachel James, Will Cooke, Kat Coyte, Lena Kolb, Jia Tsing Ng, and Anne-Marie Couto, to whom I will always be in debt.

A very special mention has been reserved for the person whose companionship I have been privileged to enjoy for the past year. Kate, you are the most wonderful person I have ever met, and my time with you has been the best of my life.

Lastly and most importantly, I dedicate this thesis to my parents, Donald and Catherine, and to my sister, Sarah. I could not have asked for a better family, and their contribution to this thesis and my life is immeasurable.

Abbreviations

4-NQO	4-nitroquinoline-1-oxide
5’-dRP	5’ deoxyribose phosphate
Adprt	ADP-ribosyl-transferase
Alt-NHEJ	Alternative non-homologous end-joining
APE	AP-endonuclease
APL	Aprataxin-like protein
APLF	Aprataxin and PNKP-like factor
AP-site	Apurinic/Apyrimidinic site
APTX	Aprataxin
ARH	ADP-ribosyl-hydrolase
BER	Base excision repair
BIR	Break-induced replication
BRCA1	Breast cancer-associated protein 1
BRCA2	Breast cancer-associated protein 2
BRCT	BRCA1 C-terminus
Bsr	Blasticidin resistance
CAM	Camptothecin
cAMP	Cyclic adenosine monophosphate
CHO	Chinese hamster ovary
CIS	Cisplatin
CtiP	CtBP-interacting protein
DDR	DNA damage response
dHJ	Double Holliday junction

D-loop	Displacement loop
DNA-PKcs	DNA-dependent protein kinase catalytic subunit
DSB	Double-strand break
DSBR	Double-strand break repair
FA	Fanconi Anaemia
FHA	Forkhead-associated
Fen1	Flap Endonuclease 1
GG-NER	Global genome NER
GST	Glutathione S-transferase
HMM	Hidden Markov model
HR	Homologous recombination
ICL	Inter-strand cross-link
ICLR	Inter-strand cross-link repair
IR	Ionising radiation
MACROD1	Macro domain containing-protein 1
MACROD2	Macro domain containing-protein 2
MAR	Mono(ADP-ribose)
mART	Mono-ADP-ribosyl-transferase
MARylation	Mono(ADP-ribosyl)ation
MEF	Mouse embryonic fibroblast
MMS	Methyl methanesulphonate
MRN	Mre11/Rad50/Nbs1
MRX	Mre11/Rad50/Xrs2
NAD^+	Nicotinamide adenosine dinucleotide
NER	Nucleotide excision repair

NHEJ	Non-homologous end-joining
PAR	Poly(ADP-ribose)
PARG	Poly(ADP-ribose) glycohydrolase
PARP	Poly(ADP-ribose) polymerase
PARylation	Poly(ADP-ribosyl)ation
PBZ	Poly(ADP-ribose)-binding zinc finger
PCNA	Proliferating cellular nuclear antigen
PCR	Polymerase chain reaction
PDB	RCSB Protein Data Bank
PHLEO	Phleomycin
PNKP	Polynucleotide kinase phosphatase
Pol X	DNA polymerase X
REMI	Restriction enzyme-mediated integration
ROS	Reactive oxygen species
RPA	Replication protein A
SDSA	Synthesis-directed strand annealing
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SSB	Single-strand break
SSBR	Single-strand break repair
TC-NER	Transcription-coupled NER
TDP1	Tyrosyl-DNA phosphodiesterase 1
Ung	Uracil DNA glycosylase
UV	Ultraviolet
XRCC1	X-ray cross-complementing protein 1
XRCC4	X-ray cross-complementing protein 4

Table of Contents

1. Introduction

1.1 The DNA Damage Response

The cell is under continuous attack from genotoxic agents, resulting in a constant stream of damage to its DNA. Such agents arise from both exogenous and endogenous sources, and cause various types of DNA lesions, including: single-strand breaks (SSBs), base damage or mismatches, double-strand breaks (DSBs), intra-strand and inter-strand crosslinks (ICLs). If left unchecked, these damages would accumulate, ultimately resulting in a high level of genetic mutation and gross chromosomal rearrangements. Therefore, the cell has evolved a complex network of pathways to signal and repair the damage, which are collectively referred to as the DNA damage response (DDR). Failure of the DDR leads the cell to an unstable genetic state, which is observed in many clinical conditions, including neurological disorders and cancer[1]. Therefore, a greater understanding of the molecular functioning and regulation of the DDR may further our knowledge of a number of disease states, and potentially provide therapeutic targets for drugs[2].

The pathways of the DDR share the same structure as many other biological signal transduction pathways. The DNA lesion is first detected by specialised sensor proteins, which then recruit downstream transducers to signal the damage. The targets of this signal are the effector proteins, which perform the variety of functions required for an effective response to the damage: repair, cell-cycle regulation and, potentially, apoptosis. Damage signalling utilises the diverse range of protein post-translational modifications available to the cell, such as phosphorylation and acetylation. For example, the histone variant H2AX is phosphorylated at S139 by three known phosphatidylinositol 3-kinase-like kinases

(PIKKs): ATM, ATR and DNA-PKcs, in response to DNA DSBs[3–6]. This rapid phosphorylation is required for the recruitment of downstream DSBR factors[7]. Moreover, acetylation of histone H4 in yeast was shown to be required for DSBR, which has since been observed in mammals[8,9]. These are a small subset of the known protein post-translational modifications implicated in DNA repair. In addition to phosphorylation and acetylation, ubiquitination, SUMOylation, and ADP-ribosylation have also been shown to be functional components of many DNA repair pathways. The role of these post-translational modifications, particularly ADP-ribosylation, will be discussed throughout this thesis.

1.2. ADP-ribosylation

ADP-ribosylation is a protein post-translational modification in which ADP-ribose moieties are covalently attached to target proteins. The enzymes responsible for catalysing these reactions are ADP ribosyltransferases (ARTs), which use NAD^+ as a substrate. ADP-ribosylation events can be divided into two categories: mono(ADP-ribosyl)ation (MARylation) and poly(ADP-ribosyl)ation (PARylation), in reactions that are catalysed by mono-ADP-ribosyltransferases (MARTs) and poly(ADP-ribose) polymerases (PARPs), respectively. MARylation is defined as the addition of one ADP-ribose moiety to a target, whereas PARylation is the addition of multiple units to form a linear or branched chain. PAR chains have been observed to be up to 200 ADP-ribose units in in length, with multiple branching points[10]. Typically, glutamic acid, aspartic acid, arginine and lysine residues in target proteins are ADP-ribosylated[11]. The formation of poly(ADP-ribose) chains is a rapid process, and these chains are turned over on the timescale of seconds to minutes[12]. Both PARylation and MARylation have been implicated in numerous cellular processes, including the response to stress, transcriptional regulation, and the DDR[13–16].

The well-established role of ADP-ribosylation in DNA repair will be described in Sections 1.3-1.7, in the context of the specific repair pathways in which it is utilised.

1.2.1. Metabolism of ADP-ribose

In humans, there are seventeen known ARTs, which are identified through conservation of a PARP catalytic domain (or ADP-ribosyltransferase fold)[17]. This domain bears strong structural and functional conservation with diphtheria toxin (DT), which is the pathogenic toxin secreted by *Corynebacterium diphtheria* from which ADP-ribosylation was first identified[18]. Among their conserved features is a triad of amino acids, H-Y-E (His-Tyr-Glu) in DT, which is important for the binding of NAD^+ and catalytic activity of the protein (Figure 1.1)[19–21]. Human ARTs can be divided into two classes: those which contain the H-Y-E triad (ARTD1-ARTD6), and those where the glutamic acid of the triad is replaced by a leucine, isoleucine, valine or threonine (ARTD7-ARTD17). ARTD9 and ARTD13 possess additional point mutations to the histidine residue of the triad, and are believed to be catalytically inactive as a result[11]. The glutamic acid residue of the triad in ARTD1 (E988) is vital for PAR chain elongation, as mutations of this amino acid abrogated this function[22]. This led to the prediction that the H-Y-E group of ARTs are PARPs, and the others are MARTs[23]. Experimental evidence has largely been gathered by studying the auto-modification activity of the ARTs, and largely supports the triad predictions (Table 1.1)[24–30]. However, ARTD3 and ARTD4 have both been shown to act as MARTs, contracting these predictions. Recently, other structural elements of the NAD^+ binding pocket have been shown to affect catalytic activity. The role donor loop, which interacts with NAD^+ directly, and the acceptor loop, which is thought to mediate the accepting and incorporation of ADP-ribose units (Figure 1.1), have been elucidated through the generation of ART-chimeras[31,32]. This analysis showed the formation of PAR

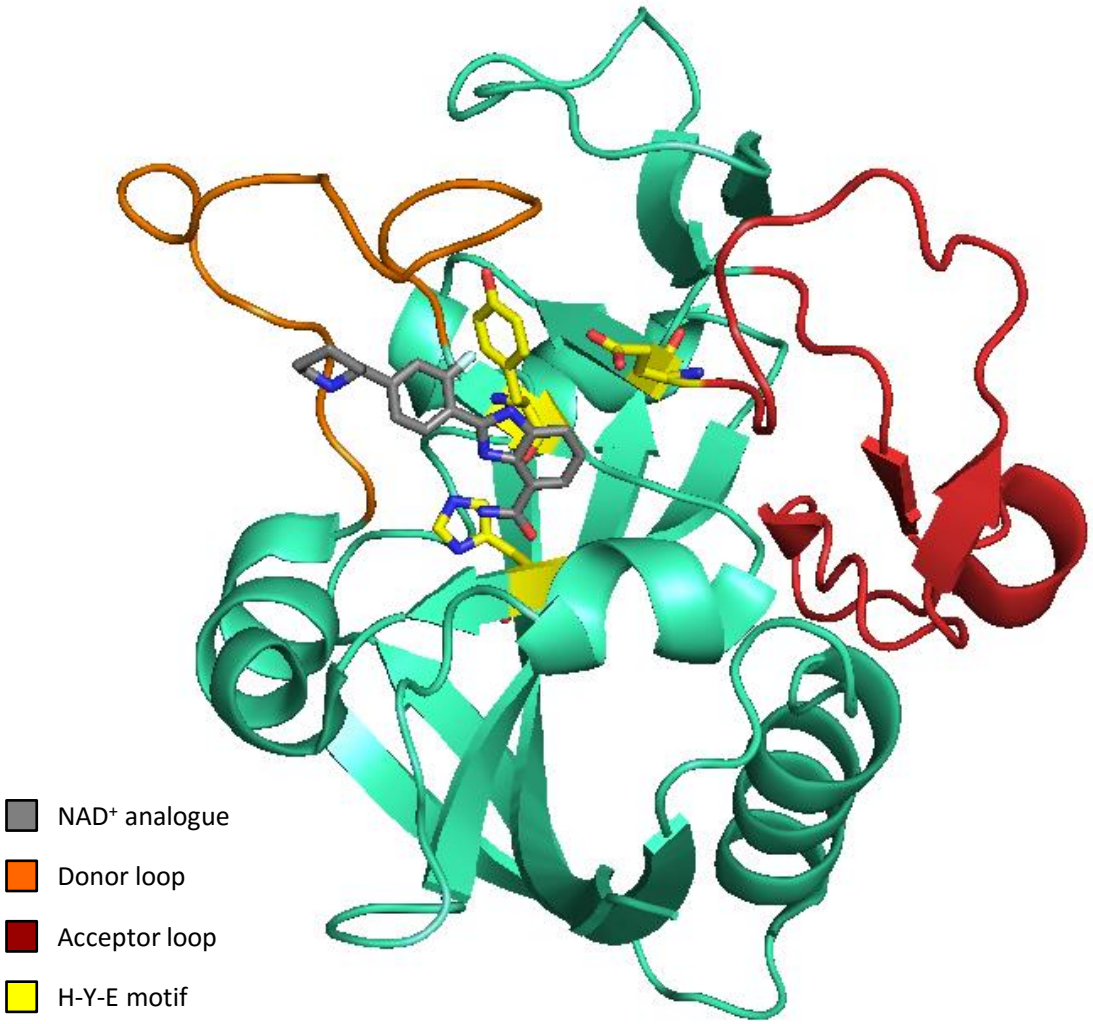


Figure 1.1 – The structure of the catalytic domain of human ARTD1. The crystal structure of the catalytic domain of human ARTD1 (PDB ID: 3L3M), bound to an NAD^+ analogue inhibitor (A927929, grey). The H-Y-E triad motif is shown in yellow, while the donor loop and the acceptor loop are shown in orange and red, respectively. Adapted from Vyas & Cheng (2014).

ADP-ribosyltransferase	Triad	Auto-modification
ARTD1 (PARP1)	H-Y-E	PARylation (conf.)
ARTD2 (PARP2)	H-Y-E	PARylation (conf.)
ARTD3 (PARP3)	H-Y-E	MARylation (conf.)
ARTD4 (PARP4, vPARP)	H-Y-E	MARylation (conf.)
ARTD5 (PARP5a, TNKS)	H-Y-E	PARylation [non-branched] (conf.)
ARTD6 (PARP5b, TNKS2)	H-Y-E	PARylation [non-branched] (post.)
ARTD7 (PARP15, BAL3)	H-Y-L	MARylation (conf.)
ARTD8 (PARP14, BAL2)	H-Y-L	MARylation (conf.)
ARTD9 (PARP9, BAL1)	Q-Y-T	Not detected
ARTD10 (PARP10)	H-Y-I	MARylation (conf.)
ARTD11 (PARP11)	H-Y-I	MARylation (post.)
ARTD12 (PARP12)	H-Y-I	MARylation (post.)
ARTD13 (PARP13, ZAP1)	Y-Y-V	Not detected
ARTD14 (PARP7, TiPARP)	H-Y-I	MARylation (post.)
ARTD15 (PARP16)	H-Y-Y	MARylation (conf.)
ARTD16 (PARP8)	H-Y-I	MARylation (post.)
ART17 (PARP6)	H-Y-I	MARylation (post.)

Table 1.1 – Summary of human ADP-ribosyltransferases. The current nomenclature and historical names of the 17 human ADP-ribosyltransferases (ARTs). The catalytic activity of each ART is thought to be predicted by its triad motif (second column), with H-Y-E indicating a poly(ADP-ribose) polymerase (PARP); however, exceptions to this rule have been identified. Variants in the glutamic acid of the triad are unable to elongate PAR chains, and are therefore predicted to be mono-ADP-ribosyltransferases (MARTs). Additional amino acid changes render the catalytic domain inactive. ARTs are characterised by studying their auto-modification activity, which is listed in the third column as confirmed (conf.) or postulated (post.). ARTD5 and ARTD6 are found to synthesise unbranched PAR chains of up to 20 amino acids only. References for activity are in the main text. Adapted and updated from Hottiger (2010).

polymers or MAR is determined by the donor loop structure, whereas the acceptor loop affects kinetics of chain formation[33].

Although the functions of the entire family of human ARTs have not yet been uncovered, those which have been characterised have been implicated in a wide range of cellular pathways. Aside from their catalytic domains, these proteins carry a wide range of additional domains that enable them to function in these diverse processes[11,17]. ARTD1 (PARP), ARTD2 (PARP) and ARTD3 (MART), have been well characterised as catalysing the ADP-ribosylation of proteins in response to DNA damage[15,25,34]. More recently, ARTD8 (MART), ARTD9 (inactive) and ARTD10 (MART) have also been shown to be DDR factors[35–37]. Additionally, ARTD8 has a role in transcriptional regulation, while ARTD10 functions in signalling pathways and cell proliferation[38–40]. ARTD5 and ARTD6 are tankyrases that have roles in cell signalling and telomere length maintenance[41]. ARTD13 (inactive) and ARTD15 (MART) are involved in antiviral activity and the response to ER stress, respectively[40,42,43]. The specific functions of these proteins will be described in Sections 1.3-1.7, in relation to their domain architecture.

ADP-ribosylation is a reversible modification with a rapid turnover, implying that proteins exist with catalytic activity opposite to that of the ARTs[12]. Several proteins have been identified which catabolise PAR chains. Poly(ADP-ribose) glycohydrolase (PARG) and ADP-ribosylhydrolase 3 (ARH3) are two such proteins that break ribose-ribose bonds in PAR chains, but are unable to cleave the bond between the base ADP-ribose moiety and the target amino acid[44–46]. Therefore, this catabolism leaves PARylated proteins MARylated. Proteins that are able to break the bond between the base ADP-ribose unit and the target modified amino acid will be discussed later.

1.2.2. Signalling of DNA damage by ADP-ribosylation

Perhaps the best defined role of ADP-ribosylation is in signalling DNA damage. This was first suggested following the observation that cellular NAD levels were reduced after exposure to DNA damaging agents[47–50]. Subsequently, it was demonstrated that inhibition of ARTs resulted in an elevated sensitivity to DNA alkylating agents, and NAD depletion impeded the joining of DNA strand breaks[51]. These observations implicated ADP-ribosylation in the DDR, and paved the way for future research. In general, PARylation in DNA repair is hypothesised to have two roles in damage signalling. Firstly, due to the negative charge of PAR chains, their formation is believed to relax chromatin at DNA lesions through electrostatic repulsion, thereby allowing repair proteins access to the lesion[52,53]. Secondly, PAR chains are also thought to directly recruit downstream repair proteins through their interaction with specific PAR-binding domains, or to stabilise protein-protein or protein-DNA interactions at DNA breaks[15,34,54]. A large volume of research has been directed towards identifying the role of individual ARTs in specific DNA repair pathways.

1.3. ADP-ribosylation in DNA single-strand break repair

1.3.1. Sources of single-strand breaks

DNA SSBs are one of the most common forms of DNA damage. Such lesions typically arise from endogenously produced reactive oxygen species (ROS), such as hydrogen peroxide, or as intermediates of other cellular processes[55]. Damaged DNA bases are removed by the base excision repair (BER) pathway, which excises the damaged base to leave a SSB[56]. Similarly, the abortive action of DNA topoisomerase 1, which introduces transient SSBs to relax DNA during replication and transcription, also leads to SSBs[57].

1.3.2. Base excision repair (BER)

As previously introduced, small base modifications, such as base alkylation, are repaired in a manner which results in the generation of a SSB, via the BER pathway. In BER, a specific DNA glycosylase cleaves the covalent bonds between the damaged base and the DNA backbone, removing the damaged base completely, leaving an abasic (apurinic/apyrimidinic; AP) site[58]. DNA glycosylates exist in two subfamilies: monofunctional and bifunctional. Both subfamilies perform the action of removing the damaged base from DNA; however, bifunctional DNA glycosylates also display AP lyase activity, and cleave the DNA backbone to generate a SSB[59]. In the case of monofunctional DNA glycosylases, the AP site is processed into a SSB by AP endonuclease 1 (APE1)[60–62]. A large family of DNA glycosylases exist to excise a wide variety of damaged bases, and they are highly specific to their substrates[63,64].

1.3.3. Mechanism of SSBR

Detection of SSBs is performed by ARTD1, which is activated upon binding to the lesion. Once catalytically active, ARTD1 PARylates itself and target proteins proximal to the SSB, facilitating the recruitment of downstream proteins[65,66]. One protein recruited to SSBs in this manner is the scaffold protein x-ray cross complementing factor 1 (XRCC1), which has no catalytic activity, but recruits further downstream proteins and increases the efficiency of single-strand break repair (SSBR) in mammals[67–70]. DNA ends are often chemically modified as a result of DNA damage, and therefore require processing prior to SSB ligation. This end processing is performed by numerous enzymes, with a range of catalytic activity to match the diverse nature of ends encountered. AP endonuclease (APE1) and tyrosyl-DNA phosphodiesterase 1 (TDP1) both remove 3'-phosphoglycolate groups, whereas 3'-phosphate and 5'-AMP groups are removed by polynucleotide kinase

3'-phosphatase (PNKP) and Aprataxin (APTX), respectively[71–76]. 5'-AMP groups are thought to arise as products of adenylation in aborted ligation attempts. Additionally, TDP1 has been shown to hydrolyse 3'-phosphotyrosyl bonds formed from the abortive activity of DNA topoisomerase 1[77,78].

Following DNA end processing, the procedure known as gap filling replaces the absent nucleotides of the DNA sequence, using the intact complementary strand as a template for error-free repair. For gaps of a single nucleotide (short patch repair), gaps are filled by DNA polymerase beta (Polβ), which also possesses 5-deoxyribose phosphate lyase activity[79–81]. The polymerase dependence of long patch repair, which fills longer gaps, is less well understood. Polβ has been implicated in concert with flap endonuclease 1 (FEN1) and ARTD1, whereas proliferating cell nuclear antigen (PCNA) appears to initiate gap filling utilising DNA polymerases delta and epsilon (Polδ and Polε)[82,83].

The final step of SSBR is ligation to restore the DNA backbone. This is performed by nuclear DNA Ligase 3 (Lig3α, hereafter referred to as Lig3) and DNA Ligase 1 (Lig1). Lig3 is recruited to the damage site through an interaction with XRCC1[69,84,85]. Additionally, Lig3 possesses a zinc finger domain homologous to those found in ARTD1, which specifically binds to nicked DNA, thereby indicating that Lig3 may be directly recruited to some SSBs[68,86]. An alternative isoform of DNA Ligase 3 (Lig3β) is localised to the mitochondria, where it performs BER[87]. Deletion of Lig3β is lethal in human and mouse cell lines, presumably due to the fact that it is the only mitochondrial DNA ligase[88,89].

1.3.4. The role of ARTD1 and ARTD2 in SSBR

As previously introduced, the detection and signalling of SSBs in humans is performed by ARTD1 (PARP1). The role of ARTD1 in the repair of DNA SSBs was first identified in

ARTD1$^{-/-}$ mice, which displayed survival defects when exposed to DNA damaging agents[90,91]. Moreover, cells derived from these *ARTD1*$^{-/-}$ mice showed elevated sensitivity to alkylating agents[92]. These data supported earlier observations that treatment of CHO cells with ART inhibitors led to increased levels of DNA damage following treatment with alkylating agents[93]. HeLa cells depleted of ARTD1 show slower kinetics of SSBR in G1-phase of the cell cycle, and the recruitment of XRCC1 to DNA SSBs was also abrogated in these cells[94]. Furthermore, depletion of ARTD1 in human A549 cells leads to slower kinetics of repair of SSBs, and elevated sensitivity to oxidising agents[95]. Mechanistically, ARTD1 contains two DNA-binding zinc-finger motifs that bind to specific structures of damaged DNA: the first binds to DSBs, whereas the second has a high affinity for gapped DNA[96,97]. Binding to SSBs induces a conformational change in ARTD1, activating its catalytic activity[98]. Active ARTD1 subsequently auto-PARylates itself and this automodification mediates the interaction of ARTD1 with downstream SSBR proteins, such as XRCC1 and Polβ[66,67,69,99]. This PARylation leads also leads to the dissociation of ARTD1 from DNA SSBs[65].

Despite being transcribed at significantly lower levels than ARTD1, it is also thought that ARTD2 has a role in PARylation at SSBs. ARTD2 was identified as being responsible for PARylation detected at SSBs in the absence of ARTD1 in MEF[25]. Cells derived from *ARTD2*$^{-/-}$ mice show increased sensitivity to genotoxic agents, including alkylating agents. Furthermore, *ARTD1*$^{-/-}$*/ARTD2*$^{-/-}$ mice are embryonic lethal, indicating an overlap of the functions of these enzymes[100]. However, depletion of ARTD2 did not significantly elevate the sensitivity of A549 cells to oxidative damage, even when co-depleted with ARTD1[95]. ARTD2 has been shown to bind to known SSBR factors, including XRCC1 and Polβ, and also to form heterodimers with ARTD1[54]. Additionally, ARTD2 displays similar autoPARylation activity to ARTD1, and is also activated by DNA breaks[101]. However, the

precise function of ARTD2 in SSBR remains unclear. It has been shown that ARTD2 displays different recruitment kinetics to SSBs compared to ARTD1. ARTD1 is enriched on chromatin immediately and transiently after DNA damage, whereas ARTD2 recruitment occurs in a slower and more persistent manner. Furthermore, despite its binding to XRCC1, XRCC1 recruitment to SSBs is not impaired in the absence of ARTD2, in a clear functional difference to ARTD1[102]. These data suggest that ARTD2 may be required for a unique role in the latter stages of SSBR, but such a role has not yet been identified.

1.4. ADP-ribosylation in DNA double-strand break repair

1.4.1. Sources of DNA DSBs

DNA DSBs are widely regarded as the most toxic form of DNA damage due to the loss of Watson-Crick base-pairing, and can arise from both exogenous and endogenous sources. The primary source of exogenous DSBs is cellular exposure to ionising radiation, alongside other genotoxic chemicals. Endogenous DSBs are generated from reactions with ROS generated by the cell, or from mechanical stress to the DNA. Additionally, one-sided DSBs arise when a replication fork encounters a SSB[103]. DNA DSBs are also induced in a tightly regulated manner in some cellular processes. Type II DNA topoisomerases act by cleaving both strands of a DNA duplex, thereby allowing another duplex to pass through. The action of these DNA topoisomerases is important for many biological functions, including mitosis and meiosis[104]. V(D)J recombination, which occurs in the maturation of T and B cells of the immune system, also functions via the induction of DSBs[105].

1.4.2. DSBR pathways

Classically, there are two major pathways implicated in the repair of DSBs: non-homologous end-joining (NHEJ) and homologous recombination (HR). A third DSBR pathway, alternative non-homologous end-joining (alt-NHEJ), was originally identified as a back-up pathway for NHEJ; however, its importance in DSBR is becoming progressively illuminated[106,107]. HR requires a homologous template for accurate DNA repair, whereas NHEJ simply involves ligation of the DSB ends in an error-prone manner. Alt-NHEJ utilises regions of local micro-homology to serve as a template, which can result in chromosomal translocations and other errors. NHEJ is also involved in mammalian V(D)J recombination; thereby individuals with NHEJ-protein mutations present with immune system disorders[108,109].

The regulation and choice of these DSBR pathways is of paramount importance to the cell, particularly in the case of programmed DSB induction, such as in V(D)J recombination and in meiosis (utilising NHEJ and HR, respectively)[110,111]. There is a strong cell-cycle dependence on the choice of repair pathway. Homologous recombination is active primarily in S and G2 phases, due to the presence of a sister chromatid to act as a template, whereas NHEJ functions primarily in G1[112]. NHEJ factors are present throughout the cell-cycle; however, its action is suppressed in S and G2 phases as its error-prone nature is toxic to the cell[113].

1.4.3. Homologous recombination

Homologous recombination provides the cell with a high fidelity procedure with which to repair DSBs in S and G2 cell-cycle phases, where it utilises the recently synthesised sister chromatid as a template to protect against information loss (Figure 1.2). The first step of HR involves a complex of three proteins: Mre11, Rad50 and Nbs1 (MRN complex), which

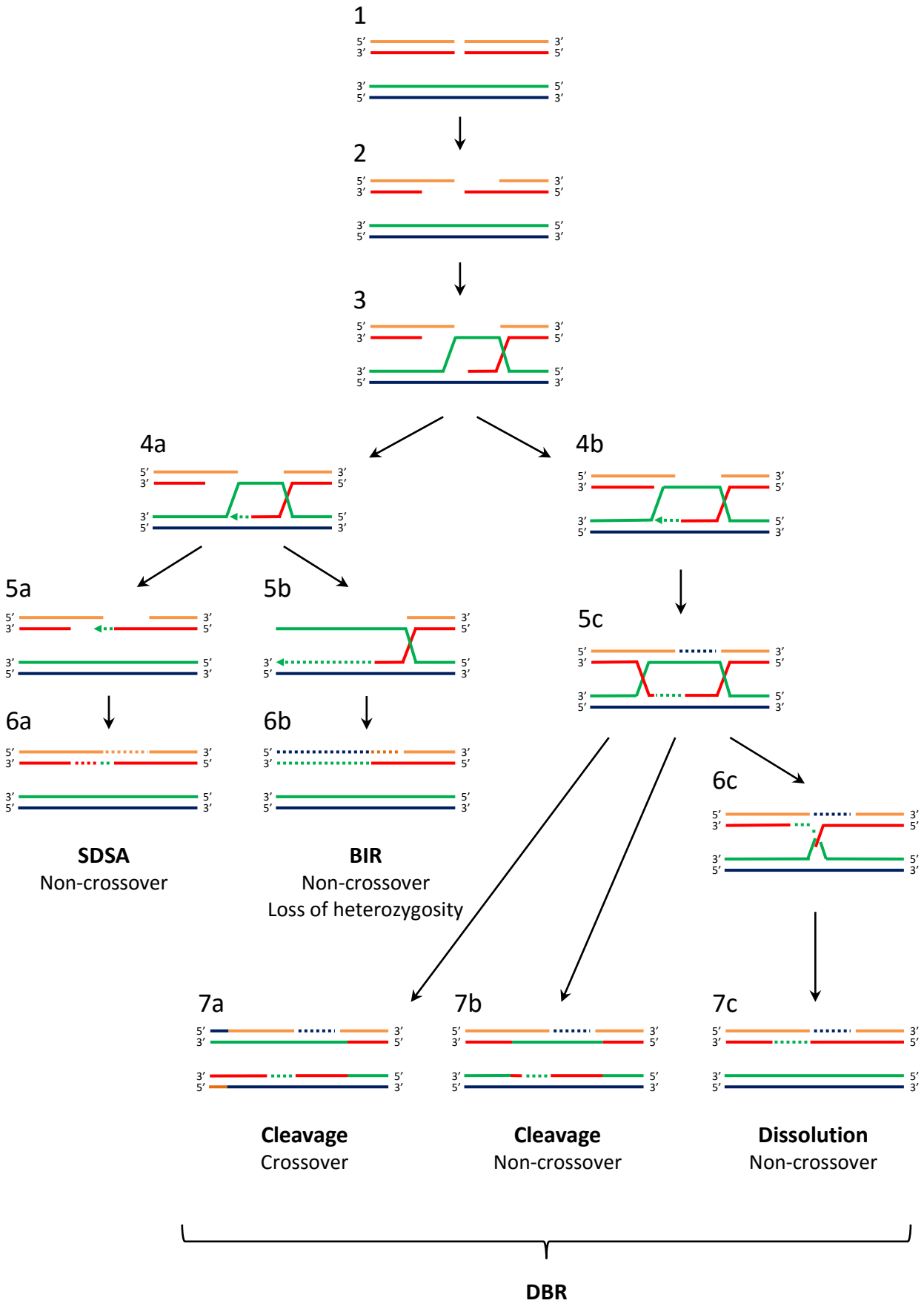

1
2
3
4a
4b
5a
5b
5c
6a
6b
6c
SDSA
Non-crossover
BIR
Non-crossover
Loss of heterozygosity
7a
7b
7c
Cleavage
Crossover
Cleavage
Non-crossover
Dissolution
Non-crossover
DBR

Figure 1.2 – The resolution of DNA DSBs by homologous recombination. The mechanism of repair of DNA DSBs by homologous recombination (HR) yields multiple end products. HR is active primarily in S- and G2-phases of the cell cycle, when a sister chromatid is available as a template. A DNA DSB is suffered on one chromatid (red and orange) and is detected (1). In the presence of a sister chromatid (green and blue), HR is the predominant pathway, and so the DNA ends are resected 5' to 3' by the exonuclease activity of EXO1, leaving a 3' overhang (2). The ssDNA is coated with Rad51, and this nucleoprotein filament performs strand invasion and homology searching of the sister chromatid, resulting in D-loop formation (3). The HR pathway now diverges, proceeding with either synthesis-directed strand annealing (SDSA), break-induced replication (BIR) or double-strand break repair (DBR) to resolve the lesion. In SDSA (4a-5a-6a), DNA synthesis from the sister chromatid is performed until just after the DSB, before the template is switched back to undamaged DNA from the original chromatid. In BIR (4a-5b-6b), the D-loop structure is formed into a replication fork, and DNA synthesis from the sister chromatid proceeds until the end of the sister chromatid. This results in a loss of heterozygosity between the sister chromatids. Both of these pathways result in repaired DNA without crossover. In DBR (4c-5c-onwards), strand invasion is performed on both sides of the DSB, resulting in double Holliday junction (dHJ) formation. Resolution of this dHJ proceeds by resolvase-dependent cleavage, leading to crossover and non-crossover products (7a and 7b), or by dissolution (6c-7c). Dotted lines indicate synthesised DNA, coloured to match the homologous strand. Adapted from Li and Heyer (2008).

is quickly recruited to DSBs[114]. Mre11 possesses endonuclease and exonuclease activity, but not 5' to 3' exonuclease activity at DSBs[115]. Rad50 is thought to interact directly with the DNA ends to stabilise the MRN complex at the lesion, whereas Nbs1 is responsible for nuclear localisation of the complex and interaction with the checkpoint kinase ATM[116–120]. Thereby, the MRN complex initially processes the DSB ends and activates ATM, resulting in phosphorylation of multiple protein targets involved in damage signalling and cell-cycle checkpoint activation[121]. DNA resection is promoted by the binding of CtIP to the MRN complex, which also promotes the HR pathway[122–124].

Homology searching, a core feature of HR, is performed by a single-stranded-DNA-protein filament. Following initial processing by the MRN complex, single-stranded DNA (ssDNA) is generated by the 5' to 3' exonuclease activity of Exo1 and Dna2, to form a 3' overhang[125–127]. The formation of unwanted secondary structure in the ssDNA is avoided by the binding of RPA, which is subsequently replaced by RAD51 in a process known as RAD51 loading[128–131]. RAD51 loading onto RPA-coated ssDNA is mediated by several proteins, including RAD52 and BRCA2[132–136]. Mutation of BRCA2 is linked to an increased risk of breast and ovarian cancers in humans[137]. The ssDNA-RAD51 nucleoprotein filament performs homology searching and strand invasion of the sister chromatid, aided by RAD54, leading to the formation of a D-loop intermediate DNA structure and Holliday junction formation[138]. This marks the commencement of DNA synthesis from the 3' end of the invading strand.

From the D-loop intermediate, HR can proceed down multiple paths: synthesis-dependent strand annealing (SDSA), break-induced replication (BIR) and double-strand break repair (DBR). In both SDSA and BIR, DNA synthesis proceeds from one Holliday junction; however, the template for DNA synthesis differs. SDSA initially uses the sister chromatid until the region of the break is passed, before the template DNA strand is disengaged and

the original complementary strand is switched to for synthesis[139]. In BIR, the entire sister chromatid downstream of the break is used, leading to a loss of heterozygosity between the sister chromatids[140]. Both of these pathways result in products without crossover. The DBR pathway differs in that DNA synthesis proceeds from Holliday junctions on either side of the lesion. Resolution of these double Holliday junctions is achieved through three possible pathways, mediated by different resolvase proteins: dissolution (BLM-TopIIIα), symmetric resolution (GEN1) and asymmetric cleavage (MUS81-EME1)[141–143].This processing leads to a multitude of crossover and non-crossover products.

1.4.4. Non-homologous end-joining

The NHEJ pathway repairs DNA DSBs by directly ligating the broken DNA ends, with limited end processing. The first step of the pathway is recognition of the DNA ends by the Ku70/Ku80 (Ku) heterodimer, which serves to both recruit downstream repair factors and also to protect the DNA ends from resection or degradation by nucleases[144]. The Ku heterodimer forms a toroidal ring structure around each DNA end, with one heterodimer either side of the break[145]. Ku binding to DNA leads to the recruitment of DNA-dependent protein kinase catalytic subunit (DNA-PKcs), which is catalytically active in its bound state[146–148]. Active DNA-PKcs auto-phosphorylates itself and phosphorylates other target proteins, such as histone H2AX at Serine-139, which is critical for repair complex formation[6,7,149]. Ku has also been shown to directly interact with aprataxin and PNPK-like factor (APLF) at DSBs[150,151]. APLF increases the stability of the repair complex at the DSB and accelerates the process of NHEJ[15].

The second step of NHEJ is end processing, in order to make the DNA ends compatible for ligation. This end processing inevitably leads to loss of genetic information at a DSB repaired by NHEJ due to the lack of a homologous DNA strand, and therefore contributes

to the error-prone nature of the pathway. Much of the machinery utilised for end processing is shared with that of SSBR, with the enzymes TDP1, APE1, PNKP and APTX performing the same functions[71–78]. In addition to its functionality as a scaffold protein, Ku has also been shown to have 5′-deoxyribose-5-phosphate (5′-dRP)/AP lyase activity, indicating that it may play a role in end processing[152]. Limited resection may be performed by nucleases such as exonuclease 1 (Exo1), Mre11, Artemis and Werner syndrome ATP-dependent helicase (WRN), while DNA polymerases μ and λ (Polμ and Polλ) may perform gap-filling[153–158].

The final step of NHEJ is the ligation of the two DNA ends. This is performed by DNA Ligase 4 (Lig4), in complex with the structurally-related XRCC4 and XLF, which are thought to interact with and stabilise Lig4[159,160]. Similarly to XRCC1 in SSBR, XRCC4 has no catalytic activity, but recruits end-processing factors APLF, PNKP and APTX, in addition to Lig4[161–164]. XLF also does not possess catalytic activity, but promotes the activity of Lig4 on non-cohesive ends, in addition to functioning in the re-adenylation of Lig4[165–167]. A paralog of XRCC4 and XLF (PAXX) has recently been discovered to function in NHEJ. PAXX has been shown to bind to Ku directly *in vitro* via its C-terminus, and stabilises NHEJ proteins on damaged chromatin[168].

1.4.5. Alternative non-homologous end-joining (alt-NHEJ)

Recently, the role of a third DSBR pathway, alternative non-homologous end-joining (alt-NHEJ), has become more strongly understood. Alt-NHEJ is a back-up DNA repair pathway which primarily functions if either NHEJ or HR is absent, particularly in G2-phase[169–171]. Repairing breaks more slowly than the classical NHEJ pathway described previously, alt-NHEJ directly ligates DNA ends in an error-prone manner, but does utilise some level of resection and local regions of homology (microhomologies), although this is

not required[172,173]. The use of microhomologies has led to the pathway being known as microhomology-mediated end-joining (MMEJ). There is a wide array of proteins implicated in alt-NHEJ, leading to the suggestion that multiple sub-pathways may be functioning within it. However, a clear mechanism for alt-NHEJ has not yet been proposed. In addition to its role in DSBR, recent data also suggests that the alt-NHEJ pathway also plays a major role in telomere maintenance[174,175]. DNA polymerase θ (Polθ) has been identified as the key DNA polymerase in alt-NHEJ. HR-deficient cells have been shown to be further sensitised to DSBs by Polθ knock-down, emphasising that alt-NHEJ also serves as a back-up pathway to HR[176–178]. Both Lig1 and Lig3 are thought to act in a redundant manner for the final ligation step of alt-NHEJ; however, recent data suggest that they may operate in different sub-pathways, with Lig3 mediating ligation with the use of microhomologies[179,180]. Furthermore, Alt-NHEJ is additionally thought to be promoted by CtIP and BRCA1, and a role for the MRN complex has also been suggested, observations that complement the role of resection in the pathway[154,181,182].

1.4.6. The role of ARTs in DSBR

ADP-ribosylation has been implicated in each of the three DSBR pathways introduced. It is known in mammalian cells that HR is activated at replication forks stalled by DNA lesions[183]. ARTD1 has been shown to bind to stalled replication fork structures *in vitro*, and is enriched on DNA following damage inflicted in S-phase *in vivo*[16]. Moreover, *ARTD1*$^{-/-}$ and ARTi-treated MEFs display hypersensitivity to chemicals that induced replication stress[184]. Furthermore, depletion of ARTD1 in HeLa cells results in elevated sensitivity to IR in S-phase, indicating a role for this protein in human HR[94]. Depletion of ARTD1 leads to a cellular defect in replication fork restart after DNA damage in S-phase, and a defect in Mre11 recruitment and resection activity[185]. Furthermore, Rad51 foci

formation at stalled replication forks is greatly reduced in both ARTD1 and ARTD2 knock-down cells, indicating abrogated HR levels. The current model for ARTD1 and ARTD2 in HR is that they are required to recruit Mre11 to stalled forks, thereby facilitating resection and ssDNA generation. This subsequently leads to Rad51 loading and HR, allowing replication fork restart[16].

A recent study has suggested a role for ARTD8, a MART, in the HR pathway in response to replication stress. ARTD8 knockdown HeLa cell lines display an increased sensitivity to DSBs, deoxyribonucleotide depletion and replicative DNA polymerase inhibition. Mechanistically, ARTD8 MARylates Rad51 in response to induction of DSBs, and lack of this MARylation leads to persistent Rad51 foci at these lesions that are thought to block efficient repair. ARTD8 is also found to associate with PCNA, a component protein of DNA replication forks[35].

In NHEJ, the recruitment of APLF to DSBs has been shown to be affected by ADP-ribosylation activity of ARTD3. Despite the sequence of the catalytic domain of ARTD3 suggesting that it may be a PARP, ARTD3 displays autoMARylation activity. This activity is stimulated by the presence of DNA DSBs, while knock-down of ARTD3 slows the kinetics of repair of DNA DSBs. The enrichment of APLF on damaged chromatin is dependent on the catalytic activity of ARTD3, and the data suggest that these proteins both function to accelerate DSB resolution by NHEJ. It is thought that a major chromatin target of ARTD3 ADP-ribosylation activity is histone H1. *In vitro,* histone H1.2 is found to be primarily MARylated, but also PARylated, by ARTD3. PARylated histone H1 is hypothesised to be responsible for APLF recruitment to DSBs[15]. Furthermore, ARTD3 has been shown to promote NHEJ through its interaction with the Ku-heterodimer. Depletion of ARTD3 reduces the binding of Ku80 at DNA DSBs, and leads to elevated levels of resection[186].

In contrast, ARTD1 has been implicated in promoting DSBR by HR at the expense of NHEJ. *ARTD1*$^{-/-}$ DT40 cells display reduced HR to repair DSBs, which restored by subsequent deletion of *ku70*, implying that a role of ARTD1 is to mediate NHEJ[187]. However, ARTD1 was implicated in a DSBR pathway utilising the XRCC1-Lig3 heterodimer, which has since been identified as the alt-NHEJ pathway[188]. As previously introduced, the first zinc finger of ARTD1 (aa: 9-91) has been shown to bind to DSB ends[97]. It is currently thought that ARTD1 and Ku compete for DSB ends *in vivo*, and act to promote alt-NHEJ and NHEJ, respectively. Both proteins have been shown to bind to DSB ends *in vitro*, but Ku has a much higher binding affinity. In NHEJ-deficient cells, ARTD1 inhibition further reduces the ability of cells to repair IR-induced DSBs, However, ARTD1 inhibition in NHEJ-proficient cells has no effect on DSB repair kinetics, suggesting that ARTD1 promotes a back-up pathway[189]. Furthermore, the recruitment of Polθ to DSBs is reduced in U2OS cells treated with PARP inhibitors, or after deletion of ARTD1 by RNAi, providing additional evidence of the role of ARTD1 in alt-NHEJ[178].

1.5. ADP-ribosylation in translesion synthesis

1.5.1. Translesion synthesis

During DNA replication in S-phase, replication forks often encounter unresolved DNA breaks or other damages. These lesions block the progression of the replication fork, causing replication fork stalling. Bypass of these lesions is performed by switching from a replicative DNA polymerase to a translesion synthesis (TLS) DNA polymerase[190]. TLS polymerases are a family of DNA polymerases with five currently well-established members: Rev1, Polζ, Polκ, Polη, and Polι. These are specialised DNA polymerases that are able to synthesise DNA across lesions, albeit sacrificing their fidelity and proofreading

activity[191–201]. This action of the TLS polymerases is vital for the restart of stalled replication forks and the cellular tolerance of DNA damage.

Polymerase switching from a replicative to a TLS polymerase is achieved through the sliding clamp PCNA. PCNA is associated with the progressing replication fork, and upon stalling, it is mono-ubiquitinated by Rad6 and Rad18[202,203]. This mono-ubiquitination is observed in response to numerous forms of induced DNA damage, including that which is specifically inflicted in S-phase. The TLS polymerases contain ubiquitin binding domains, and are therefore recruited to mono-ubiquitinated PCNA[204–208]. This direct interaction with the stalled replication fork, at the DNA lesion, leads to the TLS polymerase performing synthesis across the lesion[209]. Replicative DNA polymerases resume DNA synthesis following PCNA de-ubiquitination by USP1[210].

1.5.2. ARTD10 in translesion synthesis

ARTD10 is a MART which was initially characterised as a protein involved in cell proliferation[29,40]. However, ARTD10 has recently been found to interact with mono-ubiquitinated PNCA, through the combined action of its PCNA-binding PIP domain and ubiquitin-binding motif. ARTD10 knock-down cell lines are hypersensitive to DNA damaging agents that cause replication fork stalling, including those that act exclusively in S-phase, in addition to displaying markedly slower repair kinetics of this form of damage. Furthermore, ARTD10 knockdown is observed to reduce the recruitment of the TLS polymerase Rev1 to UV-induced DNA lesions. These data suggest a role for ARTD10 in promoting TLS. Moreover, complementation of ARTD10 knock-down cell lines with catalytically-dead ARTD10 does not alter the TLS phenotype, indicating that MARylation by ARTD10 is important for efficient TLS[37]. However, the protein targets of MARylation by ARTD10 remain undiscovered.

1.6. ADP-ribosylation in nucleotide excision repair (NER)

1.6.1. Sources of complex base damage

The cell employs a different pathway for repairing bases with bulky adducts, the removal of UV-induced damage such as 6-4 photoproducts (6-4 PPs), and DNA intra-strand cross-links. These damages only affect one DNA strand; however, they are larger in nature than those repaired by BER, which is restricted to repairing small base modifications. The pathway utilised for repairing these lesions is nucleotide excision repair (NER). Unlike BER, where specific glycosylases excise bases with specific types of damage, the core NER machinery is capable of excising a wide range of complex and bulky damages[211].

1.6.2. NER Mechanism

The NER pathway takes two forms: global-genome (GG-NER) and transcription-coupled (TC-NER)[212,213]. GG-NER is performed across the entire genome, whereas TC-NER is an accelerated process occurring at actively transcribed regions only. The two sub-pathways share a core reaction, but differ in key areas. For example, detection of lesions in GG-NER is performed by a complex of XPC-RAD23B, whereas TC-NER is activated by a stalled RNA polymerase (in addition to other TC-NER proteins). XPC is able to detect a broad range of bulky adducts by binding to two nucleotides on the undamaged strand of DNA, opposite the lesion. The damage to the other DNA strand causes thermodynamic destabilisation of the duplex, causing the nucleotides on the undamaged strand to display a more single-stranded nature, which is detected by XPC[214,215]. For certain types of UV damage, detection is performed by UV-DDB[216].

Following XPC binding, the protein complex TFIIH, which contains DNA helicases XPB and XPD, is recruited through direct interaction with the XPC-RAD23B complex[217–219].

The XPB and XPD helicases unwind the DNA at the lesion, but have distinctive roles. It is thought that the ATPase activity of XPB, but not its helicase activity, is required for functional NER[220]. In this model, the ATPase activity leads to the binding of XPD, which unwinds the DNA. It has been shown that the action of XPD is blocked by sites of DNA damage, potentially providing an additional sensing mechanism[221,222]. Following helix unwinding, XPA, replication protein A (RPA) and XPG then bind at the lesion. XPA interacts with both TFIIH and RPA; however, the recruitment of XPA is independent of these interactions[223,224]. XPA is thought to bind to DNA on the 5' side of the lesion, whereas RPA binds to the ssDNA opposite the lesion[225]. RPA coordinates the process of incision on either side of the lesion, firstly on the 5' end by ERRC1-XPF, and then the 3' end by XPG, excising a region of approximately 30 bp[226–229]. ERRC1-XPF is recruited by XPA, consistent with its positioning on the 5' side of the lesion[225,230].

The gap is filled by either standard DNA replication polymerases Polδ and Polε, or by the translesion synthesis (TLS) polymerase Polκ[231–234]. This latter finding was an unexpected result due to the lower fidelity of TLS polymerases, and suggests multiple pathways may be involved in the late stages of NER. This is further implied by the requirement for different interacting factors with each of the DNA polymerases. Polδ acts with PCNA, whereas Polκ interacts with mono-ubiquitinated PCNA, as it does in TLS[231]. DNA ligation is performed by either Lig1 or a complex of XRCC1-Lig3, in a cell-cycle dependent manner: Lig1 is thought to be preferred in proliferating cells[235]. However, the redundant nature of these two ligases in NER is yet to be understood fully.

1.6.3. ARTD1 in NER

Recently, ARTD1 has been identified in a screen for novel GG-NER proteins. ARTD1 was shown to interact with DDB2, a subunit of UV-DDB, in a UV damage-dependent manner.

This interaction leads to DDB2-dependent PARylation at the pre-incision stage of NER, which is abrogated by chemical inhibition of ARTD1. Additionally, PARylation by ARTD1 at the post-incision stage is also detected, which is thought to be as a result of lesion excision leading to a SSB. DDB2, and its partner subunit DDB1, are both targets of PARylation by ARTD1, and this PARylation increases the stability of DDB2 at UV-induced lesions by protecting it from ubiquitin-dependent degradation by the proteasome. To further implicate the role of ARTs in NER, normal human fibroblasts treated with PARP inhibitors display increased sensitivity to UV-induced DNA lesions[236].

1.7 PAR-binding domains

PARylation is a protein post-translational modification that has been implicated in multiple cellular processes, including transcription, stress responses and DNA repair[13]. In DNA repair, PARylation is generally thought to have two roles: chromatin relaxation and protein recruitment[52,66,237]. This latter function, where chains of PAR act as a scaffold for protein binding, suggests the idea that the recruited proteins may have specific PAR-binding domains. Indeed, at the start of this study, four domains had been identified as non-covalent PAR-binding modules: the PAR-binding motif (PBM), the macro domain, the zinc finger CCHH (zf-CCHH or PBZ) and the WWE domain[238–242]. Each of these domains, aside from the PBM, has an identified method of PAR interaction, utilising different features of PAR chains. A large number of human proteins have been identified that contain one or more of these domains.

1.7.1. The PAR-binding motif (PBM)

The first domain identified to be PAR-binding is a 20 amino acid motif known as the PAR-binding motif (PBM). The PBM is formed of two regions: a series of hydrophobic

and basic amino acids, and an adjacent cluster of basic amino acids. This motif was identified by sequence homology searches in several human proteins that are implicated in the DDR, including: Ku70 (aa: 243-264; Figure 1.3A), DNA-PKcs (aa: 2728-2752), XPA (aa: 215-237), Lig3 (aa: 12-34) and XRCC1 (aa: 379-400), in addition to a host of other proteins, including p53 (Figure 1.4). The relevant peptides from these proteins displayed *in vitro* PAR binding[238]. Both Ku70 and DNA-PKcs have well described roles in NHEJ. PARylation has been implicated in human NHEJ through the activity of APLF; however, direct association of Ku70 or DNA-PKcs with PAR chains *in vivo* has not yet been shown[15]. XPA, which functions in NER, has been shown to associate with ARTD1 in a PAR-dependent manner *in vivo*, in response to exposure to UV radiation[243]. However, the requirement of the PBM for this interaction has not been ascertained. XRCC1 and Lig3 exist in a complex and function in concert in BER and NER. Lig3 has been shown to preferentially bind to PARylated ARTD1 *in* vitro; however, deletion of its PBM does not fully abrogate binding[244]. Additionally, XRCC1 has been shown to bind to ARTD1 *in vitro* and *in vivo*. Deletion of the central BRCT domain of XRCC1 (aa: 315-403), a known phosphopeptide recognition motif first identified in BRCA1, abolishes this interaction[34,67,245]. Interestingly, the PBM of XRCC1 is located within this central BRCT domain. Similarly, the PBM of Lig3 is located within a DNA-binding zinc finger motif. This zinc finger motif is homologous with the second zinc finger of ARTD1, where it is thought to act as a nick sensor to detect SSBs[68]. It is thought that the presence of the PBM may augment the functionality or regulation of these domains[238].

Further work on the PBM refined its consensus motif to $[HKR]_1$-X_2-X_3-$[AIQVY]_4$-$[KR]_5$-$[KR]_6$-$[AILV]_7$-$[FILPV]_8$. A comprehensive *in sil*ico search identified this motif in 526 human proteins, including many associated in DNA processing. Of these 526, which largely do not overlap with the original study, 25 were shown to bind PAR through

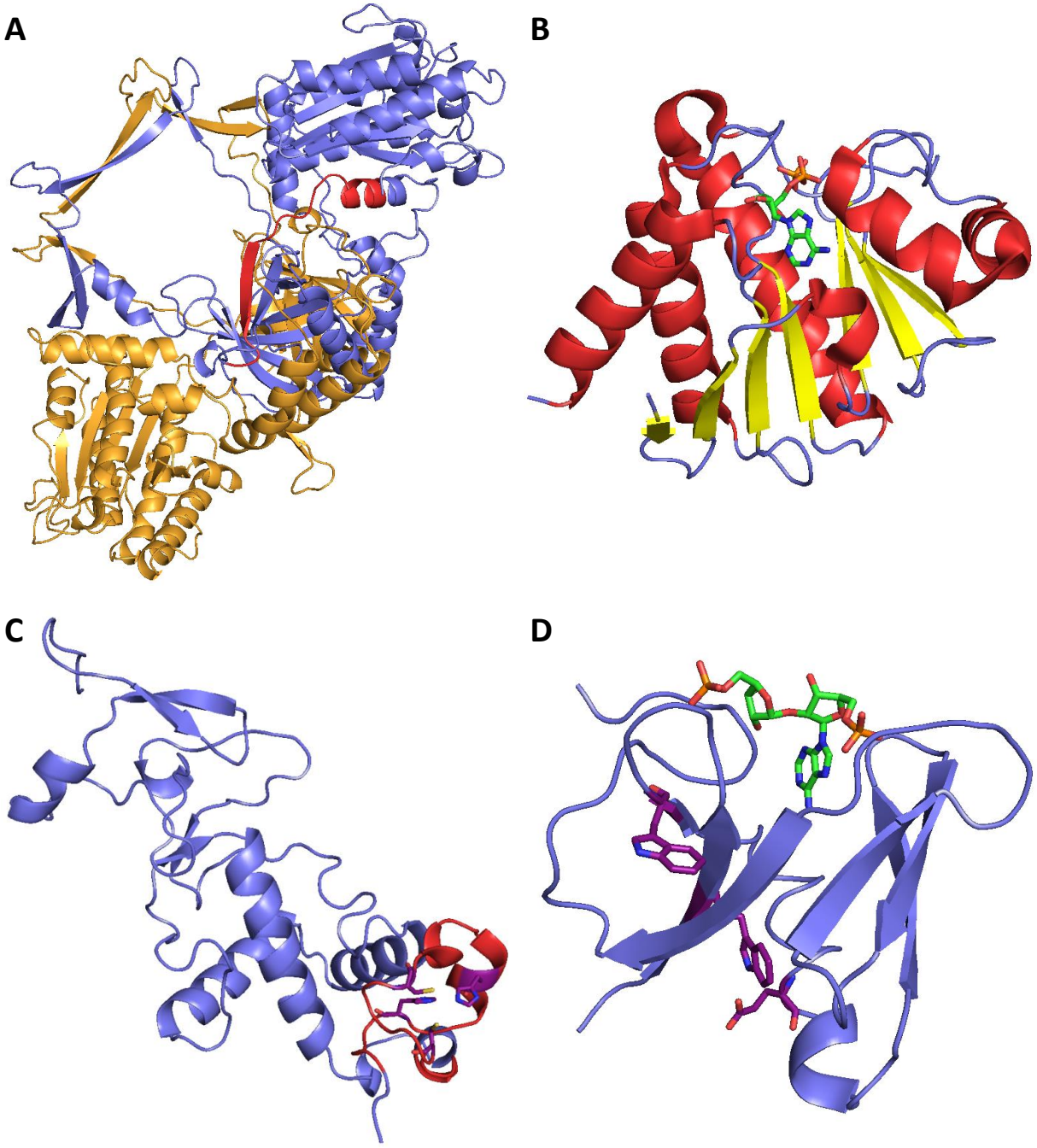


Figure 1.3 – Solved crystal structures of known PAR-binding domains. A: The structure of the human Ku70/Ku80 heterodimer (blue/orange), clearly showing the loop formed around the DNA duplex (PDB ID: 1JEQ). The identified PBM in Ku70 (aa: 243-264) is highlighted in red. **B:** The structure of the second macro domain of human ARTD8 binding to adenosine-5-diphosphate, coloured by secondary structure (PDB ID: 4D86). **C:** The structure of the C-terminus of human CHFR, with its PBZ domain indicated (red). The functionally important residues that define the PBZ domain (CCHH) are shown in purple (PDB ID: 2XP0). **D:** The structure of the WWE domain in human RNF146 binding to an iso-ADP-ribose ligand (PDB ID: 3V3L). The amino acids conserved in domains of this family (Trp-Trp-Glu) are located away from the binding pocket in this structure (purple).

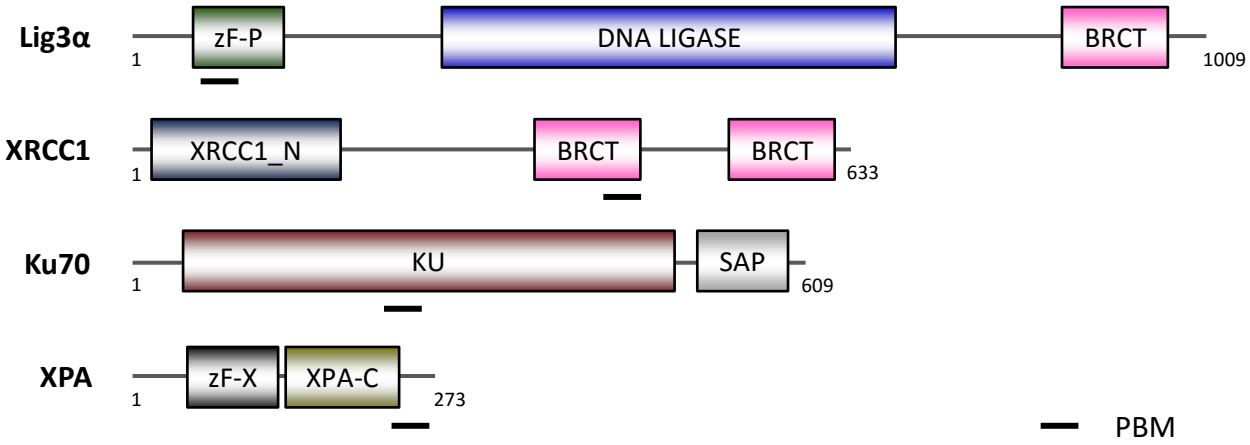


Figure 1.4 – Domain architecture of a selection of human DDR proteins with PBMs. Domain structures of selected human DDR proteins with PBMs identified using bioinformatics techniques: Lig3α (SSBR, NER, alt-NHEJ), XRCC1 (SSBR, NER), Ku70 (NHEJ) and XPA (NER). PBMs (black lines) found in these proteins fully or partially (XPA) overlap with other known domains. Domain abbreviations: PARP-type zinc finger (zf-P), BRCA1 C-terminus (BRCT), XRCC1-type zinc finger (zF-X).

immunoprecipitation and mass spectrometry analysis[239]. However, despite the growing list of PAR-binding proteins containing a PBM, the mode of binding of this motif is yet to be established and structural comparison of the regions of proteins containing a PBM yields little similarity. However, the presence of this motif in DDR proteins associated with PARylation suggests that it may indeed have an important function in the cellular response to DNA damage.

1.7.2. The macro domain superfamily

The macro domain is an ancient, evolutionary conserved domain that is found in all forms of life, with characterised examples of macro domain-containing proteins found in eukaryotes, bacteria, archaea and viruses[240]. Typically, the domain is between 130 and 190 amino acids in length, with no simple consensus sequence that has yet been identified. In fact, the family of macro domains is classed as a superfamily due to the diverse nature of the sequences of its members, and the diverse array functions to which they have been found to possess[246]. In addition to various forms of ADP-ribose that macro domains are found to bind: PAR, MAR and O-acetyl-ADP-ribose (OAADPR), macro domains have also been found to bind poly(A)[247].

Despite the lack of clear sequence homology across the macro domain superfamily, the tertiary structure of the domain is well conserved, including a binding pocket for ligand interactions. Macro domains are formed of a core of seven β-sheets, of mixed orientation, surrounded by four α-helices (Figure 1.3B). This is similar in construction to other nucleotide-binding proteins[248]. The binding pocket itself is hydrophobic cleft, and the interaction with ADP-ribose is formed by multiple hydrogen bonds with amide bonds in the protein backbone, in addition to amino acid side chains. The protein-ligand complex is stabilised by aromatic stacking of the adenine ring with an aromatic amino acid side chain.

As such, the binding pocket is highly specific for ADP-ribose. In PAR-binding macro domains, the domain binds to the terminal ADP-ribose moiety, interacting with multiple chemical groups across the molecule, perhaps suggesting how the macro domain fold is capable of binding different forms of ADP-ribose[240].

1.7.2.1. Macro domains in human proteins

There are currently 10 human proteins with identified macro domains (Figure 1.5). Of these proteins, the clearest example of a PAR-binding DNA repair protein is ALC1, which was first discovered as a protein displaying elevated levels of transcription in some liver cancers[249]. ALC1 was identified as a potential chromatin remodeller due to the presence of an N-terminal helicase domain, and it was postulated that this function may be dependent on its recruitment to chromatin though its macro domain. Investigating this function revealed that ALC1 is quickly recruited to laser-induced DNA damage, in a manner dependent on the PAR-binding nature of its macro domain. Deletion of the macro domain or the point mutation D723A, which was shown to abrogate PAR binding *in vitro,* severely diminishes this recruitment. The same study also demonstrated that the macro domain of ALC1 mediates interactions with the PAR chains of auto-PARylated ARTD1, which stabilises ALC1 at the nucleosome. In addition, ALC1-depleted U2OS cells show increased sensitivity to DSB-inducing agents and sources of replication stress[250]. More recently, ALC1 has been shown to be recruited to UV-induced DNA lesions, in a DDB2- and ARTD1-dependent manner. Additionally, ALC1-deficient normal human fibroblasts are sensitive to UV-induced damage[236].

Traditionally, macro domains were considered to be ligand binding modules. However, three macro domains have recently been characterised as catalytically active domains, functioning in the process of catabolism of ADP-ribose chains. PAR chains are known to

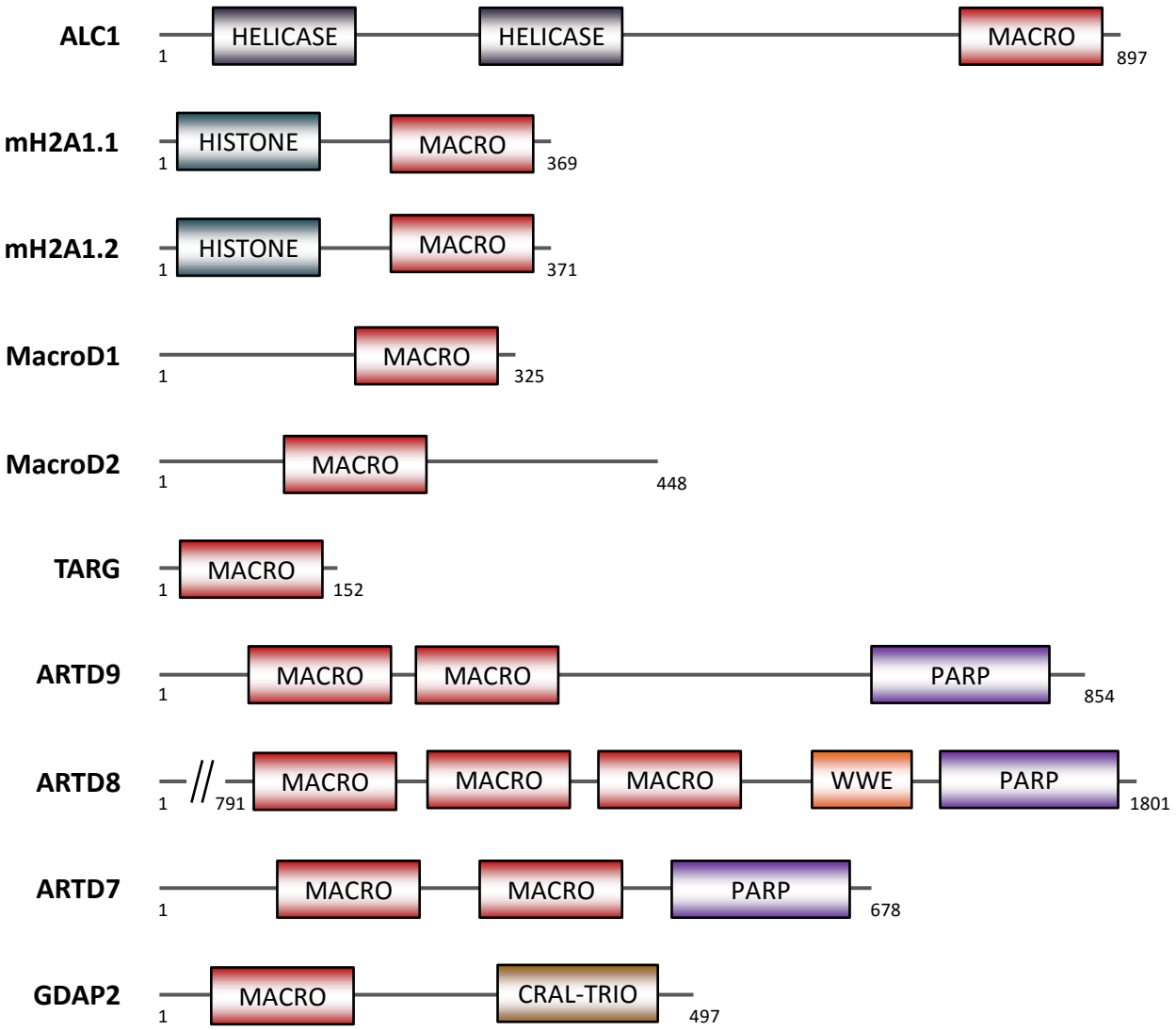


Figure 1.5 – Domain architecture of human proteins with annotated macro domains. Domain structures of human proteins identified to contain members of the macro domain superfamily. Proteins containing this domain include the chromatin remodeller ALC1, two histone H2A variants (H2A1.1 and H2A1.2) and GDAP2. The macro domains of MacroD1, MacroD2 and TARG have recently been shown to be catalytically active mono-ADP-ribosylhydrolases. Also shown is the family of macroPARPs, which are ARTs additionally containing macro domains. Domain abbreviations: PARP catalytic (PARP).

be catabolised by either PARG or ARH-3; however, neither of these two enzymes is capable of removing the base ADP-ribose unit from the chain[45,46]. As a result, the PAR hydrolysis activities of PARG and ARH-3 leave proteins MARylated. ARH-1 is known to de-MARylate arginine residues in humans and other species; however, the reversibility of MARylation on other types of amino acids was not understood[251,252]. The macro domain-containing proteins MacroD1, MacroD2 and TARG (C6orf130), which show high levels of structural conservation in their core domains, have recently been shown to be catalytically active in removing ADP-ribose from MARylated glutamic acid and aspartic acid residues in proteins[253]. These acidic amino acids have been shown to be the primary targets of ARTD1 auto-PARylation, perhaps linking these macro domain-containing proteins in PAR turnover in response to DNA damage[99]. However, the recent application of MacroD1 and MacroD2 as mono-ADP-ribosylhydrolases against a wide range of proteins in *in vitro* assays may imply that this is coincidental.

In the family of identified human proteins with macro domains, the majority only contain a single macro domain. However, three proteins of this family contain multiple macro domains. Interestingly, these proteins also share one other domain, the PARP catalytic domain that identifies and gives primary function to the family of ADP-ribosyltransferases (Table 1.1). These proteins are known as the macroPARPs: ARTD9 (PARP9; BAL1), ARTD8 (PARP14; BAL2) and ARTD7 (PARP15; BAL3)[28,254]. ARTD9 has been implicated in the DDR, where it is recruited to DSBs in an ARTD1-dependent manner. Recruitment to sites of DNA damage has been shown to be dependent on the second macro domain of ARTD9 (aa: 296-528), through its binding of auto-PARylated ARTD1[36]. BBAP, a ubiquitin E3 ligase that is known to bind ARTD9, has been shown to ubiquitinate histone H4 (H4K91), the absence of which negatively affects the recruitment of 53BP1 at damage sites[255,256]. The function of the first macro domain (aa: 1-291), has not yet been

discovered. Intriguingly, the catalytic domain of ARTD9 lacks conservation of the histidine amino acid of the catalytically important H-Y-E triad, and consequently does not PARylate or MARylate itself *in vitro*. In contrast to ARTD9, ARTD7 and ARTD8 both exhibit active PARP catalytic domains, and are observed to auto-MARylate themselves *in vitro*[28]. As described previously, ARTD8 has recently been implicated in the HR pathway in response to replication stress, and the macro domains of this protein have been shown to be important for this function. Following MARylation of Rad51 by ARTD8 in S-phase, ARTD8 then binds to MARylated Rad51 through its second macro domain (aa: 1003-1190). The binding of the second macro domain of ARTD8 is thought to inhibit HR and replication fork restart in a dominant negative manner. This behaviour highlights the interesting domain structure of ARTD8, as it can bind to target proteins following catalytic action by itself or other MARTs[35]. In addition its role in HR, ARTD8 has previously been identified as a transcriptional switch, repressing transcription under normal conditions, and promoting it when enzymatically activated through ADP-ribosylating itself and other proteins[38]. The function of the macro domains in this behaviour is still at a hypothetical stage. A precise role for ARTD7 has not yet been established; however, it is believed to function in regulating transcription[28].

Further to their roles in PAR catabolism and DNA repair, macro domains have also been found in proteins involved in other processes associated with ADP-ribosylation. Two histone variants containing macro domains have been identified: MacroH2A1 and MacroH2A2. MacroH2A1 is a macro domain-containing histone variant of H2A that is found primarily in heterochromatin, where it suppresses transcription by preventing remodelling by transcription factors and inhibiting histone acetylation[257,258]. The presence of the macro domain has been shown to be responsible for both of these functions[259]. Additionally, MacroH2A1 is also found to be enriched on silenced X chromosomes[260].

MacroH2A1 undergoes alternative splicing to produce two isoforms, with only one product encoding an NAD metabolite-binding protein. MacroH2A1.1, though interactions with its macro domain, binds to ADP-ribose and OAADPR specifically, whereas no ADP-ribose binding has been detected with MacroH2A1.2[261]. MacroH2A2 is a H2A histone variant with a 68% sequence identity with MacroH2A1.2. Similarly to MacroH2A1, MacroH2A2 is localised to the inactive X chromosome, although its expression pattern across multiple human cell types is markedly different[262].

GDAP2 (ganglioside-induced differentiation-associated protein 2) is a protein containing a single macro domain that displays no affinity for ADP-ribose polymers. Instead, this macro domain binds to poly(A), further highlighting the variation found within the macro domain superfamily[247]. GDAP2 additionally contains a lipid binding CRAL-TRIO domain[263].

1.7.3. The zinc-finger CCHH (zf-CCHH; PBZ) domain

The PBZ domain is a zinc finger domain with a 24 amino acid consensus sequence of $[K/R]X_2CX[F/Y]GX_2CXBBX_4HX_3[F/Y]XH$, with B and X representing basic and unconstrained amino acids, respectively. It is present only in eukaryotic organisms, excluding yeast, which directly mirrors the presence of ARTs. This domain has recently been identified as a PAR binding module, and is characterised by four completely conserved residues: two cysteine and two histidine amino acids[241]. This configuration of amino acids gives the domain specificity towards PAR chains, where it interacts with the α(1-2) O-glycosidic bond between two ADP-ribose moieties[264,265]. The domain is also known as the zf-CCHH, in order to distinguish it from the classical zf-C2H2 zinc finger domain, with consensus sequence $CX_{2-4}CX_{12}HX_{2-6}H$, most of which are found to bind nucleotides in DNA and RNA[266]. Three human proteins with PBZ domains have been

identified: APLF, with tandem PBZ domains, SNM1A (the analogue of yeast PSO2), and CHFR (Figure 1.6)[15].

1.7.3.1. Human proteins with PBZ domains

APLF was first implicated in the DDR through its interaction with phosphorylated XRCC1 and XRCC4 in yeast two-hybrid (Y2H) assays. *In vivo*, APLF interacts with CK2-phosphorylated XRCC1 and XRCC4 through its N-terminal forkhead-associated (FHA) domain, a phosphopeptide recognition motif, and this interaction enriches APLF on damaged chromatin[162]. In the absence of XRCC1, APLF is recruited to damaged chromatin through its tandem PBZ domains, which bind to PAR[241,267]. The presence of tandem PBZ domains is unique in human proteins, and results in a marked increase in the affinity of APLF for PAR chains[268]. APLF co-localises with sites of PAR synthesis, particularly at sites of ARTD1 PARylation. In response to DSBs, APLF has been shown to directly bind to the NHEJ factor Ku80, via a novel motif, which stabilises APLF at the break site[150]. Additionally, APLF has been shown to act in the same pathway as ARTD3 in NHEJ, suggesting ARTD3 ADP-ribosylated proteins are targets of APLF[15]. At the DSB, APLF promotes the stability of XRCC4 and accelerates the action of Lig4 in ligating the DNA ends. Cells depleted of endogenous APLF display slower repair kinetics of SSBs and DSBs than wild-type cells, and hypersensitivity to ionising radiation (IR)[151]. PSO2 was first identified as a DNA repair protein in *Saccharomyces cerevisiae*. Cells lacking this protein were initially found to display sensitivity to photo-induced ICLs; however, *Pso2*$^{-}$ cells have since been shown to be sensitive to all forms of induced cross-links[269]. PSO2 has 5'-exonuclease activity that is dependent on its beta-lactamase domain[270]. Epistatic analysis placed PSO2 in the RAD3-dependent pathway of yeast DNA repair, which is the NER pathway[271]. However, more recent studies focusing on the nuclease activity of yeast

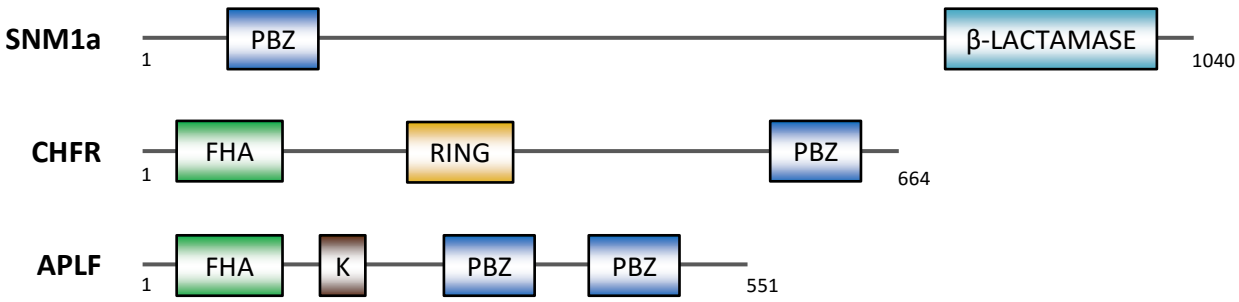


Figure 1.6 – Domain architecture of human proteins with annotated PBZ domains. Domain structures of the three human proteins with identified CCHH-type zinc finger (PBZ) domains, all of which are involved in the cellular response to DNA damage. SNM1a is a nuclease involved in ICL repair, whereas CHFR is a G2/M checkpoint protein whose function is abrogated by deletion of its PBZ domain. APLF is implicated in the stabilisation of NHEJ complexes at DSBs, utilising its tandem PBZ domains. Domain abbreviations: forkhead associated (FHA), RING finger (RING), Ku80 interacting motif (K).

PSO2 have suggested that it may process ICL-induced DSBs, independent of NER[272]. In fact, yeast cells lacking PSO2 display normal incision activity at sites of DNA inter-strand cross-links, but fail to process and resolve the resulting DSB intermediates[272,273].

There are three mammalian homologs of PSO2: SNM1A, SNM1B and SNM1C/Artemis, all of which have been shown to display nuclease activity[274,275]. SNM1A has 5'-exonuclease activity, and is considered to be the closest mammalian homolog to Pso2. Both mouse *SNM1a*$^{-/-}$ embryonic stem cells and human SNM1A knock-down fibroblasts show similar sensitivity phenotypes in response to ICL-inducing agents as yeast PSO2[276]. Furthermore, survival of Fanconi Anaemia-deficient (FA-deficient) cells to ICLs is further reduced by SNM1a knock-down, whilst mono-ubiquitination of FANCD2 is unperturbed by the knock-down, providing evidence that SMN1a functions in an alternative ICL repair pathway to the FA pathway[277]. Human SNM1A possesses a single PBZ domain; however, the role of this domain in the function of the protein, in addition to its binding activity, has not yet been determined[241].

CHFR has been identified as a tumour suppressing protein through its action as a cell cycle regulator. CHFR prevents cells that have undergone mitotic stress from entering metaphase[278]. As such, it is found to be mutated in several human cancer cell lines, and *chfr*$^{-/-}$ mice display increased levels of tumorigenesis[279]. In humans, it is known that CHFR acts as a ubiquitin E3 ligase through a central RING finger domain. In addition to ubiquitinating itself, CHFR ubiquitinates Polo-like kinase 1 (Plk1), a regulator of other cell-cycle proteins, resulting in its degradation. The degradation of Plk1 results in the arrest of the cell cycle at the G2/M checkpoint[280,281].

A C-terminus PBZ domain has been identified in CHFR, and this domain was shown to be PAR-binding *in vitro* and *in vivo* (Figure 1.3C). Furthermore, CHFR co-localises with

PAR foci in undamaged cells in a PBZ domain-dependent manner, analogous to the behaviour of APLF. Deletion of its PBZ domain renders CHFR unable to perform its function in the G2/M checkpoint; however, it does not affect its auto-ubiquitination activity[241].

1.7.4. The WWE domain

A third domain identified with PAR-binding activity is the WWE domain, named for its three completely conserved and functionally important amino acids (Trp, Trp, Glu). Similarly to the PBZ domain, it is found only in eukaryotes[242]. The solved structure of a WWE domain indicates that it consists of a half beta-barrel structure, made of six anti-parallel β-sheets surrounding an α-helix (Figure 1.3D). The WWE domain has been shown to bind to the region of PAR chains containing the ribose-ribose glycosidic bond, termed iso-ADP-ribose (iso-ADPr). This mode of binding, which has been verified by a solved NMR structure (PDB ID: 4QPL), suggests specificity for PAR chains over MAR units[282]. Human proteins with annotated WWE domains fall into two functional classes: those involved in PARylation, and those involved in ubiquitination[242]. The domain structures of these proteins are shown in Figure 1.7.

1.7.4.1. WWE domains in human proteins

The currently identified proteins involved in PARylation that contain WWE domains are all members of the ART family: ARTD8 (PARP14), ARTD11 (PARP11), ARTD12 (PARP12), ARTD13 (PARP13; ZAP; ZC3HAV1), and ARTD14 (PARP7; tiPARP). The role of ARTD8 in transcription regulation and DNA repair has already been described, through its MAR-binding macro domain; however, the role of its WWE domain is yet to be established[35]. ARTD12, ARTD13 and ARTD14 can be classed as a subfamily of ARTs as they share one other domain: a CCCH-type zinc finger (zf-C3H1). This domain is

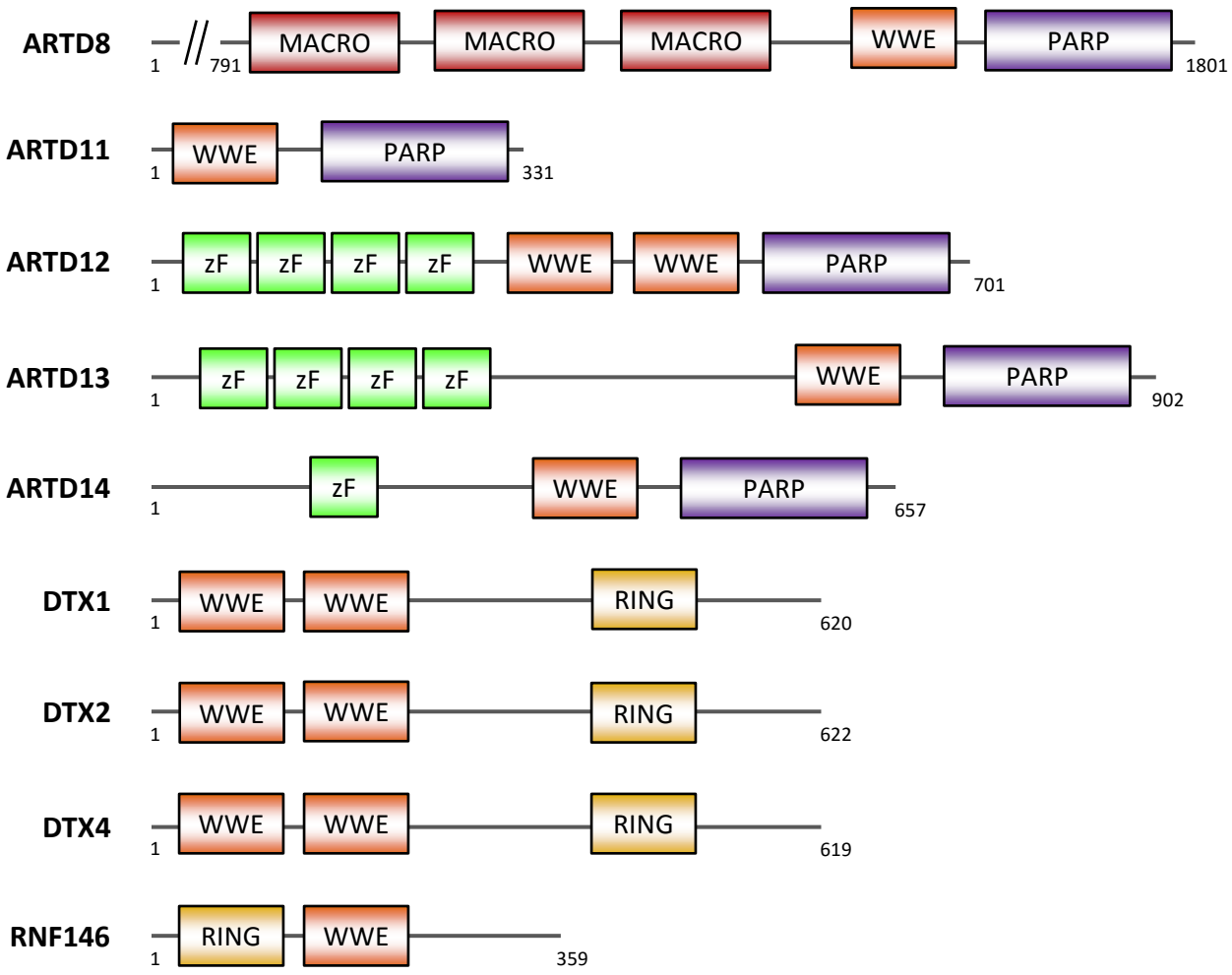


Figure 1.7 – Domain architecture of human proteins with annotated WWE domains. Human proteins with identified WWE domains fall into two functional categories: PARylation and ubiquitination. Five human ARTs (ARTD8, ARTD11, ARTD12, ARTD13 and ARTD14) contain WWE domains alongside their PARP catalytic domains. The Deltex family of proteins (DTX1, DTX2 and DTX4) are ubiquitin E3 ligases which function in the Notch signalling pathway. Additionally, a WWE domain is found in the ubiquitin E3 ligase RNF146, which is implicated in PARylation-dependent ubiquitination (PARdU). Domain abbreviations: PARP catalytic (PARP), C3H1-type zinc finger (zF), RING finger (RING).

thought to be RNA binding as it is identified in several RNA binding proteins, although its precise function is poorly understood[283]. Analysis of the catalytic domains of these ARTs leads to the prediction that they act as MARTs; however, there is no experimental evidence to support this prediction[11]. While there is currently no data available on the functions of ARTD12 and ARTD14, ARTD13 has been identified as an antiviral protein. ARTD13 targets viral mRNAs, including from retroviruses such as HIV-1, and recruits the cell's RNA degradation machinery[42]. ARTD13 exists as two isoforms, ARTD13-L (long) and ARTD13-S (short). Both isoforms contain the WWE domain, but only the long isoform retains its PARP catalytic domain. It is found that ARTD13-L is a more potent antiviral protein than ARTD13-S, implicating ADP-ribosylation in this process[284]. A requirement for the WWE domain in the function of ARTD13 has not yet been established. ARTD11 contains a PARP catalytic domain, which is predicted to MARylate proteins, and a WWE domain only[11]. There is currently no indication as to the cellular role of this protein.

WWE domains have also been identified in proteins implicated in ubiquitination pathways, and thereby serve as a link between ADP-ribosylation and ubiquitination. Furthermore, the presence of a WWE domain in ubiquitin E3 ligases suggest a mechanism for PARylation-dependent ubiquitination (PARdU)[282,285]. RNF146 is a RING domain-containing ubiquitin E3 ligase that has been shown to ubiquitinate its targets in a PAR-dependent manner. The iso-ADPr-binding WWE domain of RNF146 binds to PARylated proteins, activating the E3 ligase activity of RNF146. RNF146 interacts with ARTD5 (Tankyrase 1) and ARTD6 (Tankyrase 2), which are ARTs involved in the Wnt signalling pathway and telomere length maintenance. In this PARdU mechanism, a Tankyrase PARylates a protein, which then becomes a target for RNF146 ubiquitination. In the Wnt signalling pathway, PARdU of Axin by RNF146 results in its degradation, thereby promoting signalling[286,287].

The Deltex family of proteins: DTX1, DTX2, DTX3 and DTX4, are a set of ubiquitin E3 ligases thought to regulate the Notch signalling pathway. This function has been inferred from similarity to the *Drosophila melanogaster* protein Deltex, in addition to the identification of the individual protein targets of ubiquitination[288–290]. Deltex has tandem WWE domains, and has been shown to bind to the Ankyrin domain of the Notch receptor through interaction with these domains[291]. To date, PARylation has not been implicated in this process, despite a single WWE being an iso-ADPR-binding module. DTX1, DTX2 and DTX4 all contain tandem WWE domains; however, the function of these domains in humans has not been confirmed. In keeping with a role in the Notch signalling pathway, the individual members of the Deltex family have been implicated in the ubiquitination and regulation of other Notch-associated proteins[292,293].

1.7.5. Recently discovered PAR binding domains

Recently, two domains with previously well-established functions in other processes have been implicated as independently PAR-binding in the response to DNA damage. These are the previously introduced FHA and BRCT domains, which are classically thought of as phosphopeptide recognition motifs that mediate protein-protein interactions[245,294]. The FHA domain, a 65-100 amino acid motif, was first identified in a DNA repair protein in *Saccharomyces cerevisiae,* where the cell-cycle checkpoint protein Rad53 has two such domains. Rad53 was shown to interact with phosphorylated Rad9. Disruption or deletion of the FHA domain by mutation abrogates Rad53 phosphorylation and abolishes cell-cycle arrest at the G2/M checkpoint following DNA damage[295,296]. Since that discovery, FHA domains have been identified in many more DDR proteins in yeast and higher eukaryotes. For example, FHA domains mediate the phospho-specific protein-protein interactions between APLF and XRCC1 or XRCC4 in response to single- or double- strand breaks,

respectively[162]. Additionally an FHA domain is found in NBS1, a key component of the MRN complex, which facilitates phospho-specific binding with HR proteins CtIP and MDC1[297,298].

The BRCT domain was first identified in BRCA1, and more such domains were subsequently found in DNA repair and cell-cycle checkpoint proteins, such as XRCC1, NBS1, 53BP1, MDC1, Lig3 and Lig4[299]. BRCT domains are typically found as either single modules or in tandem. The BRCT domains of BRCA1 were identified as functionally important as it is found that clinically relevant mutations of the protein destabilise the structure of the tandem BRCT domains[300]. It was subsequently found that the BRCT domains bind to phosphorylated BACH1, a DEAH helicase family protein that is implicated in the DDR[301,302]. Functions for the BRCT domains in other DNA repair factors have since been discovered. As previously described, the BRCT domain of Lig3 mediates its damage-induced interaction with XRCC1 in BER, whereas the BRCT domains of 53BP1 bind to p53, although in the latter case the deletion of its BRCT domains does not appear to affect its role in DNA repair[69,303–305]. Additionally, the BRCT domains of MDC1, a cell-cycle checkpoint protein, facilitate its binding to γH2AX in response to DSBs. This leads to the formation of DSBR complexes at the lesion, and is a critical step for the resolution of this form of DNA damage[306].

It has recently been shown that several proteins interact with PAR chains through either BRCT or FHA domains. It was found that the fast recruitment of BRCA1 to sites of laser microirradiation was dependent on the PAR-binding behaviour of the tandem BRCT domains of its partner, BARD1. BARD1 and BRCA1 interact through their RING domains. In addition to introducing domain-inactivating mutations, this fast recruitment was shown to be abrogated by the PARP inhibitor PJ34[307]. Furthermore, the early recruitment of APTX (FHA), PNKP (FHA), NBS1 (FHA and BRCT), Lig4 (BRCT) and

XRCC1 (BRCT) have also been shown to be dependent on the PAR-binding nature of the bracketed domains. In the case of NBS1, G2/M checkpoint activation is defective if the interaction between PAR and NBS1 is disrupted[308]. The lack of co-evolution of the both the FHA domain and BRCT domain with ARTs, as seen by the presence of FHA domains in yeast and BRCT domains in prokaryotes, suggests that the binding of ADP-ribose chains is not the primary function of these domains. However, mounting evidence indicates that there may indeed be a subset of these domains for which this function is both present and required for operational DNA repair pathways.

Finally, the oligonucleotide/oligosaccharide-binding (OB) fold, which is classically seen as an RNA or ssDNA binding motif, has been shown to be PAR binding *in vitro* and *in vivo*[309,310]. The protein hSSB1 is a ssDNA binding protein thought to be one of the first proteins to bind at the sites of DNA DSBs. It interacts with the MRN complex directly, though NBS1, which recruits it to DSBs and promotes its endonuclease activity. Cells lacking hSSB1 are radiosensitive and display the hallmarks of genetic instability[311–313]. It has recently been shown that the early recruitment of hSSB1 to sites of DNA damage is dependent on the PAR-binding activity of its OB fold. The early recruitment of hSSB1 is abrogated by treatment of U2OS cells with the PARP inhibitor olaparib, in U2OS cells transfected with ARTD1 siRNA (but not ARTD2), and by introducing point mutations to the OB fold. To date, this is the only example of a PAR-binding OB fold. However, due to the presence of the fold in both prokaryotes and eukaryotes, it is clear that the OB fold has not co-evolved with ARTs and this function may be coincidental[310].

1.8. The use of model organisms for studying ADP-ribosylation

As is true in many fields in the biological sciences, a substantial volume of information on DNA repair pathways has been gathered through the use of eukaryotic model organisms.

These organisms, such as *S. cerevisiae* (budding yeast), have been vital for identifying DNA repair pathways and proteins due to their genetic tractability and suitable of level conservation between themselves and higher eukaryotes, particularly in the case of proteins involved in recombination[314]. However, the study of ADP-ribosylation in *S. cerevisiae* is not possible as there is an absence of proteins containing PARP catalytic domains in the genome of this species, in stark contrast with the 17 human ARTs described previously[315]. This lack of ARTs leads to the conclusion that PARylation and MARylation, two post-translational modifications strongly implicated in the DDR, are absent in yeast, thereby making this model unsuitable for studying these processes. Therefore, the use of other model organisms was required for this purpose.

The action of ARTD1 in DNA repair in chicken DT40 cells has been well studied[187,316]. However, despite containing orthologs of the majority of human ARTs, these cells do not have an ortholog of ARTD2[317]. This has advantages for studying the function of ARTD1, as redundancy with ARTD2 is not present; however, direct comparison with human cells cannot be made for this reason. Homologs of ARTD1 and ARTD2 have been identified in the nematode worm *C. elegans*, and these have been shown to be active in the response to DNA damage[318,319]. However, the genome of this organism lacks homologs of many ARTs found in humans. Similarly, the *Drosophila melanogaster* genome contains genes for only two ARTs[320]. Therefore, these organisms are not suitable models for ADP-ribosylation in human cells. *In silico* analysis of the genome of *Dictyostelium discoideum*, a simple eukaryotic slime mould, revealed 15 proteins with putative ART catalytic domains. Moreover, several of these proteins displayed strong homology with the DNA damage-responsive PARPs[315,321]. Therefore, *Dictyostelium discoideum* became recognised as a model organism for studying ADP-ribosylation in DNA repair.

1.9. *Dictyostelium* as a model organism for studying the DDR

1.9.1. The genetic model organism *Dictyostelium discoideum*

Dictyostelium discoideum (hereafter referred to as *Dictyostelium*) is a simple eukaryotic, soil-dwelling social amoeba that has been modified so as to grow in laboratory conditions. These axenic strains, of which Ax2 and Ax4 are among the most common, grow in shaking suspension at 22°C with a doubling time of approximately eight hours[322–324]. The *Dictyostelium* nuclear genome is compact at 34 Mb in length, split over six chromosomes. The genome is very AT-rich, with a GC content of 22.4%, a striking difference to mammalian genomes. Similarly different is the structure of *Dictyostelium* genes, which have an average intron number of 1.85 per gene compared with 7.88 in humans[325–327]. Importantly, chromatin structure and the conservation of histone variants in *Dictyostelium* appears to be well conserved with vertebrate organisms[328].

Dictyostelium is a haploid organism; however, in an exponentially growing, vegetative cell population, 55-70% of the cells are in G2-phase of the cell-cycle and 5-10% of the cells are in S-phase. In *Dictyostelium*, M-phase typically lasts for 10-20 minutes, and is immediately followed by S-phase. There is no identifiable G1-phase, and therefore the G1/S checkpoint is also thought to be absent[329,330]. This cell-cycle is in clear contrast with higher eukaryotic organisms. In practical terms, *Dictyostelium* cells can be synchronised in G2-phase by cold incubation, but obviously not in G1-phase[331,332]. Regardless of these differences, the genetic tractability of *Dictyostelium*: it being a haploid organism with a single copy of a large number of genes, makes it an attractive model to work with.

Dictyostelium has long been studied as a model for development and chemotaxis. Under conditions of starvation, single-cellular *Dictyostelium* undergo a remarkable, cooperative,

developmental cycle[333]. The first stage of this cycle is aggregation into a multicellular body, which is mediated by waves of secreted cAMP. Individual cells secrete cAMP, and its detection by neighbouring cells stimulates them to secrete further cAMP, resulting in a spiral wave of cAMP secretion. *Dictyostelium* use chemotaxis to move to regions of higher cAMP concentrations, thereby aggregating in a mound. The end point of this development cycle is the dispersal of spores from the tip of a stalk, called a fruiting body. The cells of the mound are able to move to more favourable dispersal conditions as a slug, during which decisions are made as to which cells will become spores and which will be sacrificed in the stalk. Following slug taxis, the fruiting body is formed and the spores are released. It was originally unclear as to the cell-cycle phase of spores, primarily due to a lack of cell-cycle markers. It is currently thought that spores are in G2-phase, the same phase as the majority of vegetative, single-cellular *Dictyostelium*[334].

1.9.2. The DNA damage response in *Dictyostelium*

The most attractive feature of using *Dictyostelium* as a model organism for studying the DNA damage response is the striking conservation of DNA repair pathways and proteins with humans. Despite *Dictyostelium* cells being highly resistant to many forms of DNA damage, perhaps due to being able to engage HR for high fidelity repair in most of the cells of a population, it appears that the processes it uses for DNA repair are highly similar to those found in mammalian cells[335,336]. Early work in this organism, prior to the sequencing of its genome, identified complementation groups of genes, mutations of which resulted in elevated sensitivity to specific DNA damaging agents[337–339]. Furthermore, individual DDR proteins, including ARTs, were purified and analysed biochemically, thereby constructing the knowledge of the DNA repair pathways of this organism[340,341].

Homologs of many of the proteins involved in mammalian BER and NER have been found in *Dictyostelium*. Homologs of core BER proteins such as XRCC1, Polβ, Polδ, Lig1 and Lig3 have been identified *in silico*[321]. Additionally, an AP endonuclease was identified by homology and subsequent deletion of the gene encoding this protein resulted in sensitivity to DNA damage[342]. Furthermore, a uracil DNA-glycosylase has been purified and characterised *in vitro*[343]. Both XRCC1 and Lig3 are absent in yeasts. NER is not as strongly conserved as BER; however, homologs of XPB, XPC, XPD, XPE and XPF have been found[344–347]. Additionally, XPB$^-$ and XPF$^-$ (unpublished data) cells are sensitive to UV exposure and the UV-mimetic drug 4-NQO, respectively[348]. The TLS polymerases Rev1 and Rev3 (a subunit of Pol ζ) have also been identified, along with a homolog of the nuclease Mus81. Therefore, *Dictyostelium* can be used as a good model of both of these DNA repair pathways.

In addition to the conservation of many of the proteins involved in SSBR, *Dictyostelium* also has homologs of many factors involved in the resolution of DSBs. HR is strongly conserved across all forms of life, and *Dictyostelium* is no different in this regard[349]. Homologs of Rad51, Exo1, RPA, BRCA2, Mre11 and Rad50 have all been identified *in silico*; however, a homolog of the final component of the MRN complex, Nbs1, is yet to be found[350–353]. It has been difficult to assess the contribution of HR in *Dictyostelium* as HR knock-out mutations have mostly proved to be lethal. However, *exo1*$^-$ cells are viable, despite a significant growth defect, and are sensitive to DSB-inducing agents[351]. The viability of these cells is thought to be possible due to redundancy of Exo1 with other exonucleases, potentially Dna2 or Mre11.

In the case of NHEJ, the core components of the pathway are found in simple eukaryotes and in some prokaryotes, but notably not *Escherichia coli* (*E. coli*), which appears to rely completely on HR for DSBR[354]. In contrast to HR, there is not a complete conservation of

NHEJ repair proteins between simple eukaryotes and mammalian cells, and it is often found that alternative proteins are employed in the NHEJ pathway[355]. In *S. cerevisiae*, Artemis and DNA-PKcs are absent, and the Mre11/Rad50/Xrs2 (MRX) complex, the yeast equivalent of the MRN complex, is implicated in end-processing[356,357]. Additionally, Pol4 (the yeast homolog of Polβ) is utilised, but no such role for mammalian Polβ has been found. *Dictyostelium* has homologs of Ku70, Ku80, Artemis, DNA-PKcs, XRCC4 and Lig4, making it an attractive model for mammalian NHEJ[353]. Vegetative *Dictyostelium* cells display active HR and NHEJ pathways; however, NHEJ-defective strains are not sensitive to DSB-inducing agents[358]. However, plentiful evidence suggests that NHEJ is functional in these cells. One method to assess NHEJ cells is to monitor the level of restriction-endonuclease mediated integration (REMI) of an antibiotic resistance cassette. In this assay, a linearised plasmid is transfected into *Dictyostelium* cells that bears a low level of DNA sequence homology to any region the *Dictyostelium* genome, alongside the restriction enzyme that is used to digest the plasmid. Due to the lack of a homologous DNA sequence, and the restriction enzyme generating compatible DNA ends in the genomic DNA, integration of this construct proceeds by NHEJ[359]. The construct encodes an antibiotic resistance cassette, which is used for selection. The assessment of REMI in NHEJ-defective *Dictyostelium* strains (*ku70*$^-$, *ku80*$^-$, *dnapkcs*$^-$ and *dclre1*$^-$) display reduced REMI efficiency, indicating that NHEJ is functioning in vegetative cells[351,358]. Furthermore, Ku70 and Ku80 are enriched on chromatin following the induction of DSBs, indicating an active NHEJ pathway[360]. One theory to explain the lack of sensitivity of NHEJ-deficient *Dictyostelium* strains to DSBs is the cell-cycle distribution of vegetative cells described earlier: the majority of cells are in G2 and so will favour DSBR by HR[330]. Interestingly, NHEJ-defective germinating spores are sensitive to DSB-inducing agents, suggesting a higher reliance on NHEJ in this developmental stage[358,360]. It was thought at

one stage that this indicated that spores were in G1-phase, but recent experimental evidence has pointed to spores being in G2-phase, and it is unclear why cells at this developmental stage show stronger reliance on NHEJ[334]. Despite the clear difference in the reliance of *Dictyostelium* and mammalian cells on HR and NHEJ, it is clear from its levels of HR and NHEJ protein conservation that *Dictyostelium* is a highly suitable model for studying both pathways in a simple eukaryotic organism.

1.9.3. ADP-ribosylation in *Dictyostelium* DNA repair

As previously discussed, the existence of proteins involved in ADP-ribosylation in *Dictyostelium* was known long before the sequencing of its genome, due to the purification of ARTs[340,341]. Analysis of the *Dictyostelium* genome reveals 15 putative ARTs, identified by conservation of the PARP catalytic domain[315,321]. Importantly, nuclear PARylation is observed following treatment of *Dictyostelium* cells with DNA damaging agents, and PARylated proteins can be found to be enriched on damaged chromatin, indicating that PARylation is a functional component of the DNA damage response[360,361]. In addition, commercially available PARP inhibitors, such as NU1025 and benzamide, abrogate this response[360]. Utilising a combined approach of bioinformatics and genetic manipulations, putative homologs of human ARTD1, ARTD2 and ARTD3 were identified (Adprt2, Adprt1b and Adprt1a, respectively). Experimental analysis showed that *adprt2*$^{-}$ and *adprt1b*$^{-}$ cells are sensitive to SSB-inducing agents, whereas *adprt1a*$^{-}$ cells have an NHEJ defect[360]. Interestingly, deletion of both *adprt1a* and *adprt2* further increases cellular sensitivity to SSBs, implying that these lesions are channelled into NHEJ in the absence of SSBR[361].

Study of the NHEJ proteins in *Dictyostelium* revealed that its homolog of Ku70 contains a C-terminal PBZ domain, which is absent in human Ku70. Endogenous *Dictyostelium*

Ku70 binds to PAR polymers *in vitro*, and deletion of its PBZ domain abolishes this PAR-binding activity. Furthermore, this deletion of the PBZ domain reduces the stability of Ku70 at DSBs. This has profound consequences for the NHEJ, as cells expressing only the PBZ-deletion mutant of Ku70 display highly reduced levels of NHEJ, measured by REMI efficiency[360].

1.10 Aims

Interestingly, the PBZ domain located at the C-terminus of *Dictyostelium* Ku70, which is required for the stability of NHEJ factors to DNA DSBs, is absent in the human ortholog of Ku70[360]. In human NHEJ, the stability of NHEJ proteins at DNA lesions has been shown to be mediated by the tandem PBZ domains of APLF, as previously described[15]. A homolog of APLF has not been identified in *Dictyostelium*. These data suggest that while dependence of the functioning NHEJ pathway on PARylation is conserved between the humans and *Dictyostelium*, the individual proteins and domains required for this may not be conserved (Figure 1.8). It appears that, over the course of evolution, the PBZ domains responsible for stabilisation of NHEJ factors at DSBs have become localised on different core NHEJ proteins in humans and *Dictyostelium*. We hypothesise that this event may not be a unique occurrence, and that other PAR-binding domains may have evolved so as to be found on different proteins in these two organisms. Therefore, we will use known PAR-binding domains as surrogate markers for novel DNA repair proteins in *Dictyostelium*, the identification of which could lead to further understanding or discovery of novel human DNA repair proteins. For example, the discovery of a PAR-binding domain on a *Dictyostelium* protein with a human homolog that is involved in DNA repair may elucidate an additional role for ADP-ribosylation in the DNA repair pathway in which the human protein operates. Moreover, if a novel *Dictyostelium* PAR-binding protein is found and

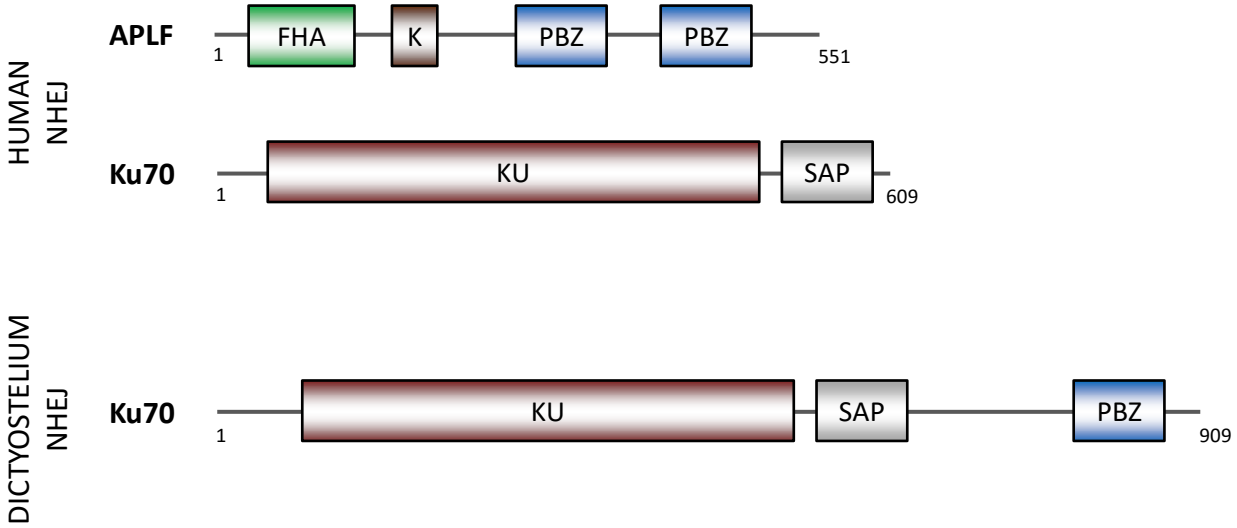


Figure 1.8 – Stability of human and *Dictyostelium* NHEJ complexes is achieved through PBZ domains in different proteins. In humans, following the binding of the Ku70/Ku80 heterodimer to DSB ends, APLF is recruited to the lesion (interacting with Ku80 via its specific binding motif). Though its PBZ domains, APLF increases the stability of the complex of NHEJ proteins at the lesion, thereby facilitating efficient repair. In *Dictyostelium*, deletion of the PBZ domain from Ku70 decreases the stability of NHEJ proteins at DSBs, and vastly reduces the ability of cells to perform NHEJ. There is no *Dictyostelium* APLF. Domain abbreviations: forkhead associated (FHA), C2H2-type zinc finger (PBZ), Ku80 interacting motif (K).

implicated in the response to DNA damage, this might indicate the presence of further human proteins with PAR binding domains, or additional functionality to those that have already been characterised.

The overall aim of this thesis was to identify *Dictyostelium* proteins with previously unannotated PAR-binding domains, and then investigate the function of these proteins in the DDR in this organism. We employed advanced bioinformatics techniques to perform an *in silico,* genome-wide search for novel proteins with domains that may bind to ADP-ribose chains, and we identified three *Dictyostelium* and one human protein with unannotated macro domains. We then performed experimental analysis of two of these putative *Dictyostelium* DDR proteins: DNA Ligase III (Lig3) and an aprataxin-like protein (APL), to assess if these proteins have a role in DNA repair, and the function of their putative PAR-binding domains in this. Our results show that the macro domains of both proteins bind to PAR *in* vitro. Furthermore, the results presented in this thesis implicate *Dictyostelium* Lig3 and APL in the response to MMS-induced DNA damage and DNA inter-strand cross-links, respectively.

2. Materials and Methods

2.1. Materials

Church Buffer

0.5 M $NaHPO_4$ (pH 7.2)

1% (w/v) bovine serum albumin fraction V

1 mM EDTA (pH 8)

7% (w/v) SDS

[Filter sterilised]

DNA loading dye

0.25% bromophenol blue

40% sucrose

H50 Buffer

50 mM KCl

20 mM HEPES

10 mM NaCl

5 mM $NaHCO_3$

1 mM $NaH_2PO_4.H_2O$

1 mM $MgSO_4.7H_2O$

[pH 7.0; filter sterilised]

HL5

53 mM maltose

1.4% bacteriological peptone

0.7% yeast extract

4 mM Na_2HPO_4

3.5 mM KH_2PO_4

340 μM dihydrostreptomycin sulphate

74 μM vitamin B12 (cyanocobalamin)

82 μM biotin

532 μM riboflavin

[pH 6.5; autoclaved; 25 ml 20% glucose added to 475 ml HL5 immediately before use]

IP Beads Buffer

50 mM Tris (pH 8.0)

150 mM NaCl

1% Triton X-100

IP DUB Buffer

50 mM Tris (pH 8.0)

150 mM NaCl

5 mM DTT

Protease inhibitor cocktail (Roche)

IP Lysis Buffer

50 mM Tris (pH 8.0)

150 mM NaCl

1% Triton X-100

1 mM DTT

2 mM $MgCl_2$

100 units/ml Benzonase (Sigma)

Phosphatase inhibitor cocktail 2 & 3 (Sigma)

Protease inhibitor cocktail (Roche)

[Supplemented with 10 mM benzamide (Sigma) and 200 μM DEA (Trevigen) for PARPi/PARGi]

IP Wash Buffer

50 mM Tris (pH 8.0)

150 mM NaCl

1% Triton X-100

1 mM DTT

2 mM $MgCl_2$

Phosphatase inhibitor cocktail 2 & 3 (Sigma)

Protease inhibitor cocktail (Roche)

[Supplemented with 10 mM benzamide (Sigma) and 200 μM DEA (Trevigen) for PARPi/PARGi]

KK2

19 mM KH_2PO_4

3.6 mM K_2HPO_4

[Autoclaved; supplemented with 10 mM benzamide (Sigma) and 200 μM DEA (Trevigen) for PARPi/PARGi]

LB Agar

1% bactotryptone

0.5% yeast extract

85 mM NaCl

1.5% agar

[Autoclaved]

LB Broth

1% bactotryptone

0.5% yeast extract

85 mM NaCl

[Autoclaved]

Nuclear Extraction Buffer

50 mM HEPES (pH 7.5)

150 mM NaCl

1 mM EDTA

Phosphatase inhibitor cocktail 2 & 3 (Sigma)

Protease inhibitor cocktail (Roche)

10 mM Benzamide (Sigma)

200 μM DEA (Trevigen)

Phosphate buffered saline (PBS)

One PBS tablet dissolved in 200 ml ddH_2O yields:

10 mM phosphate buffer

2.7 mM KCl

137 mM NaCl

[pH 7.4; autoclaved]

Pt Buffer

3 mM NaCl

1 mM $NaPO_4$

[pH 6.5; filter sterilised]

SDS-PAGE Running Buffer

25 mM Tris

192 mM glycine

0.1% SDS

SDS-PAGE Transfer Buffer

25 mM Tris

192 mM glycine

20% methanol

SDS Loading Buffer (2×)

50 mM Tris (pH 6.8)

20% glycerol

4% SDS

0.2% bromophenol blue

100 mM DTT

SM Agar

1% peptone

56 mM glucose

0.1% yeast extract

16 mM KH_2PO_4

5.5 mM $K_2HPO_4.3H_2O$

4 mM $MgSO_4$

1.7% agar

[Autoclaved]

SSC (20×)

3 M NaCl

300 mM sodium citrate tribasic ($C_6H_5Na_3O_7.2H_2O$)

[Autoclaved]

TAE (50×)

2 M Tris Base

50 mM EDTA

5.71% glacial acetic acid

[pH 8.2]

Tris-buffered saline supplemented with Tween-20 (TBS-T)

2.48 mM Tris

137 mM NaCl

[pH 7.4; autoclaved]

2.2. Methods

2.2.1. Bioinformatics

2.2.1.1. Multiple sequence alignments

Protein amino acid sequences were obtained from UniProt or dictyBase[362,363]. Alignments of multiple protein sequences were performed using MUSCLE or T-Coffee, and visualised in Belvu[364–366]. DNA sequences were aligned using the MultAlin interface[367].

2.2.1.2. Homology searching

In silico searches for PAR-binding domains were performed within the non-redundant UniRef50 database[362]. Proteins containing known PAR-binding domains were identified in the Pfam and SUPERFAMILY protein databases, in addition to dictyBase[363,368,369]. Initial local homology searches were formed using BLAST[370]. Profile hidden Markov models (profile-HMMs) were generated using HMMer2 and HMMer3, which were also used for profile-sequence homology searches[371]. Searches with HMMer 2 and 3 were iterated up to 40 times. HHPred was employed for profile-profile homology searches, including secondary structure prediction[372].

2.2.1.3. Protein structure prediction by homology modelling

Homology modelling was performed using the Chimera interface to MODELLER[373,374]. Protein structures were presented and manipulated using PyMOL[375]. The template structure of MACROD1 was obtained from the PDB (PDB ID: 2X47), and the multiple sequence alignment was produced with T-Coffee[365,376,377].

2.2.2. *Dictyostelium* cell culture and genetic manipulation

2.2.2.1. Cell culture

Dictyostelium cells were grown axenically in HL5 medium, at 22 °C in the dark. 6-, 12-, 24-, and 96-well plates were used where appropriate, or 90 mm dishes. For growth in shaking suspension, cells were placed in flasks in HL5 medium and rotated at 220 rpm, at 22 °C. Cells were maintained between 5×10^5 and 6×10^6 cells/ml, and used for all experiments in the exponential growth range ($2\text{-}6\times10^6$ cells/ml). Cell density was measured using a haemocytometer and counting cells under a Motic AE21 trinocular microscope. Antibiotic selection was maintained by adding the appropriate antibiotic: blasticidin (Enzo) or G418 (Sigma) at 10μg/ml. *Dictyostelium* cells were also grown on SM agar plates, in association with *Klebsiella aerogenes*.

2.2.2.2. Transformation

Dictyostelium cells were counted and 5×10^7 cells were collected by centrifugation (3000 rpm for 3 minutes). The cell pellet was washed twice in ice cold H50 buffer, before resuspension to a density of 5×10^7 cells/ml in H50 buffer. $5x10^6$ cells were then mixed with 3-8 μg of plasmid DNA, and left on ice for 5 minutes. For transfections with PDXA-3C vectors, and an equal amount of the pREP plasmid simultaneously added. The cells and plasmid mix was then transferred to a 1 mm gap electroporation cuvette (Bio-Rad) and subjected to two pulses (0.65 kV and 25 μF, separated by 5 seconds) from the Gene Pulser Xcell (Bio-Rad). The cells were left to recover on ice for 5 minutes, before transfer to HL5. For transfections with disruption or knock-out constructs, cells were diluted into 50 ml HL5, and subsequent 1:10 and 1:100 dilutions were performed. Cells were then transferred to 96-well plates, and 10 μg/ml blasticidin (Enzo) was added 24 hours later.

For transfection of extrachromosomal vectors (PDXA-3C), cells were transferred to a 90 mm dish in 10 ml HL5, and 10 μg/ml G418 (Sigma) was added 24 hours later. The plates were incubated in the dark at 22 °C.

2.2.2.3. Generating disruption strains

Identified homologous arms for each disruption were ligated into the pLPBLP vector, flanking a blasticidin-resistance cassette (*bsr* cassette), by standard techniques. A suitable concentration of the pLPBLP disruption vector was then obtained by Maxi-preparation (Macherey-Nagel), following transformation of the vector into XL1-Blue *E. coli* cells. The disruption construct was excised by restriction digestion, and purified using phenol-chloroform extraction (see Section 2.2.3.1). The disruption construct was then transformed into *Dictyostelium* cells as described in Section 2.2.2.2. Growth after 14 days in individual wells of the 96-well plates indicated potential disruptions, which were screened by PCR (Section 2.2.2.4). Those cells which were positive in the screen were spread on SM agar, in association with *K. aerogenes*, for clonal selection. Clones were subsequently screened by PCR, as before.

2.2.2.4. PCR screening of disruption strains

Genomic DNA was extracted from potential disruption strains by the Wizard® genomic DNA purification kit (Promega), by the provided protocol. Amplification of selected DNA fragments was guided by synthesised primers (Sigma), the sequences of which are given in Appendix A. PCR was formed using 100 μg DNA and PCRBIO HiFi Polymerase (PCR Biosystems), using the manufacturer's protocol.

2.2.2.5. Cre-loxing

Transformation of *Dictyostelium* cells was performed as described in Section 2.2.2.2, with the PDXA-NLS-Cre plasmid encoding Cre recombinase, with the exception that pREP was not added in this case, to ensure transient expression of Cre recombinase. After transformation, the cells were diluted in 50 ml HL5 medium, and further diluted 1:10, 1:100 and 1:1000, in HL5, prior to transfer to 96-well plates. 20 μg/ml G418 (Sigma) was added after 24 hours, and the cells were incubated at 22 °C, in the dark, for 5-7 days. The medium was then removed and replaced with HL5 containing no selection. Cells that are observed to be growing were replica-plated into HL5 containing 20 μg/ml G418 or 10 μg/ml blasticidin, and those identified to be killed by both antibiotics were identified as successful strains. These were spread on SM agar plates in association with *K. aerogenes* to select clones, which were subsequently verified by PCR.

2.2.2.6. Development and spore germination

Exponentially growing *Dictyostelium* cells were washed three times in KK2, prior to resuspension to 1×10^8 cells/ml in KK2. The cells were then spread onto filter paper to a density of 3.5×10^6 cells/cm^2, and incubated in the dark at 22 °C. After 24 hours, development of fruiting bodies was observed; however, a further 48 hours of development was required before spore harvesting. For spore harvesting, fruiting bodies were transferred to KK2 containing 0.1% NP-40 alternative (Calbiochem), and disrupted by passing through a 19.5 gauge needle. Spores were then enriched by centrifuging at 1500 ×g for 3 minutes, before washing twice in KK2 with 0.1% NP-40 and once with KK2. Spores were then resuspended at 2×10^7 cells/ml in KK2. Spore germination was induced by heating to 45 °C for 30 minutes, with shaking. Following heating, cells were diluted to 1×10^6 cells/ml in room temperature HL5 medium.

2.2.2.7. Storage of strains

Permanent stocks of strains were generated by removing exponential *Dictyostelium* cells from HL5 and resuspending them to 5×10^7 cells/ml in Freezing Media. Cells were then immediately transferred to cryotubes (Thermo Fisher Scientific) or as 100 µl aliquots in 0.5 ml tubes, placed on ice for 5 minutes, then at -20 °C for 1-2 hours, and finally at -80 °C for long-term storage. Cells were recovered from long-term storage by thawing aliquots and transferring to HL5 medium in a 90 mm dish. After 30 minutes, to allow *Dictyostelium* to adhere to the base of the dish, the HL5 was replaced with fresh HL5, with appropriate antibiotic selection (described in Section 2.2.2.1).

2.2.3. Molecular biology

2.2.3.1. Phenol-chloroform DNA extraction

Between $2\text{-}4\times10^8$ exponentially growing *Dictyostelium* cells were removed from HL5 by centrifugation (3000 rpm for 3 minutes) and washed twice with KK2. The cell pellets were resuspended to a density of 1×10^8 cells/ml in nuclear lysis buffer (10 mM MgOAc, 10 mM NaCl, 30 mM HEPES pH 7.5, 10% sucrose, 2% NP-40 alternative), and incubated on ice for 10 minutes. Nuclei were then collected by centrifugation at 9000 ×g for 10 minutes at 4 °C. The pellet was then resuspended in MHK buffer (40 mM MgOAc, 20 mM KCl, 50 mM HEPES pH 7.5), in twice the volume of NLB used, before addition of NaCl to 200 mM and SDS to 1% final concentrations. After gentle inversion, phenol-chloroform-isoamyl alcohol (Sigma) was added in a 1:1 ratio, prior to centrifugation at 20,000 ×g for 5 minutes. The upper aqueous layer was then transferred to a fresh tube. If this layer was clear, chloroform was added in a 1:1 ratio, before centrifugation at 20,000 ×g for 5 minutes. If the aqueous layer was not clear, the phenol-chloroform step was repeated. The

upper aqueous layer obtained after addition of chloroform was transferred to a fresh tube containing two volumes of 100% ethanol and 100 mM sodium acetate. DNA was precipitated either at -20 °C overnight, or at -80 °C for 30 minutes. Following precipitation, the DNA was pelleted by centrifugation at 20,000 ×g for 5 minutes. The pellet was then washed with 70% ethanol, before the centrifugation step was repeated. The 70% ethanol was then aspirated and the pellet was air-dried, prior to resuspension of the DNA in a minimal volume ddH_2O. The DNA concentration was quantified using the NanoDrop ND-1000 spectrophotometer.

2.2.3.2. Southern blotting

Genomic DNA was purified and quantified as described in Section 2.2.3.1 and 50 μg was digested overnight with 500 units of the desired restriction enzyme. Following digestion, the DNA-enzyme mix was purified using the phenol-chloroform method (Section 2.2.3.1), and finally resuspended in a volume of ddH_2O suitable for gel electrophoresis. The DNA was then loaded into a 0.7% TAE-agarose/ethidium bromide gel and separated by electrophoresis. The gel was then washed in 0.2 N HCl for 10 minutes, rinsed in dH_2O, then washed in 1.5 M NaCl and 0.5 M NaOH for 45 minutes, before rinsing in dH_2O again. The gel is then further washed twice in 1 M Tris (pH 7.4) for 30 and then 15 minutes, and rinsed in dH_2O in between. The DNA was then transferred overnight by capillary action onto a nylon Hybond-N^+ membrane (Amersham), using 10× SSC, and subsequently washed with 2× SSC and air-dried. The DNA then covalently cross-linked to the membrane by UV irradiation using a UV Stratalinker (Stratagene).

The probe for Southern blot was amplified from genomic Ax2 DNA by PCR, and radioactively-labelled with [α^{32}P]-dATP (Hartmann Analytic), using the Decaprime™ II DNA Labelling Kit (Thermo Fisher Scientific), according to the provided protocol.

Unincorporated [α^{32}P]-dATP was removed using SigmaSpin Post-reaction Clean-up columns (Sigma), according to the provided instructions.

Prior to hybridisation, the membrane was incubated in Church buffer for 1 hour with rotation, at 65 °C. The probe was denatured by boiling at 100 °C for 5 minutes, before addition to the Church buffer. Hybridisation proceeded overnight at 65 °C, with rotation. The membrane was then washed twice with 2× SSC with 0.1% SDS, and then twice with 0.2× SSC with 0.1% SDS at 65 °C, with rotation (both wash solutions were pre-heated to 65 °C). Detection of radioactively-labelled bands was performed using autoradiographic film (Hyperfilm, GE Healthcare). Blots were stripped by incubation with boiling 0.1% SDS, which was repeated twice, before re-probing using the same protocol as above.

2.2.3.3. RNA Extraction and cDNA synthesis

RNA was extracted from 1×10^7 *Dictyostelium* cells by resuspension in 1 ml TRI reagent (Sigma), where the solution was repeatedly mixed by pipetting. After incubation at room temperature for 5 minutes, 0.2 ml chloroform was added and the mixture was vigorously shaken, before a further 2-15 minute incubation at room temperature. The mixture was then centrifuged at 12,000 ×g for 15 minutes at 4 °C, and the colourless upper phase containing the RNA was removed to a fresh tube containing 0.5 ml isopropanol. This was incubated at 4 °C for 5-10 minutes before centrifugation at 12,000 ×g for 20 minutes, and the resulting pellet was washed with 70% ethanol, before further centrifugation at 7,500 ×g for 5 minutes at 4 °C. The RNA pellet was briefly air-dried (5-10 minutes) before resuspension in ddH2O, which was facilitated by repeated pipetting at 55-60 °C for 10-15 minutes. The concentration of RNA was quantified using the NanoDrop ND-1000 spectrophotometer. RNA stocks were stored at -80 °C or used immediately for cDNA synthesis.

The synthesis of cDNA from 5 µg RNA was performed using the SUPERSCRIPT II Reverse Transcriptase (Invitrogen), Oligo(dT) and random hexamers, following the provided protocol. The final product was used in PCR reactions, as described in Section 2.2.2.4.

2.2.3.4. Preparation of nuclear extracts

Exponentially growing *Dictyostelium* cells (5×10^8) were harvested by centrifugation (3000 rpm for 3 minutes) and washed with 10 ml KK2. The cells were lysed in 2 ml NIB1 buffer (50 mM Tris (pH 7.8), 5 mM MgOAc, 10% sucrose, 2% NP-40 alternative (Calbiochem), protease inhibitor cocktail (Roche)) for 1 minute, on ice. Nuclei were then collected by centrifugation (2300 ×g for 5 minutes at 4 °C), before resuspension in 400 µl HS buffer (20 mM HEPES (pH 7.6), 25% glycerol, 420 mM NaCl, 0.2 mM EDTA, 0.5 mM DTT). Nuclei were then disrupted by passing them through a 26G needle, before mixing of the solution by gentle rotation for 30 minutes (at 4 °C). Debris was removed by centrifugation at 14,000 ×g for 5 minutes (4 °C), and the supernatant (nuclear extract) was collected. The supernatant was then mixed with 2× SDS loading buffer and boiled at 100 °C for 5 minutes.

2.2.3.5. DNA manipulations

Standard DNA manipulation techniques of restriction digestion, ligation, PCR, ethanol and isopropanol precipitation and TAE gel electrophoresis were performed using well-established protocols. Plasmid DNA was transformed into XL1-Blue competent *E. coli* for the generation of large quantities by Mini and Maxi-preparation (Macherey Nagel), using the manufacturer's instructions. Gel extraction and PCR purification (Qiagen) were performed using the provided protocols. DNA fragments were sequenced by Sanger

sequencing, which was performed locally by Source Bioscience, using synthesised primers (Sigma).

2.2.4. Phenotypic analysis

2.2.4.1. Sensitivity assay

Exponentially growing *Dictyostelium* cells were diluted to 1×10^6 cells/ml in HL5, and separated into 1 ml aliquots. The aliquots were then exposed to the indicated concentrations of genotoxic agents (MMS (Sigma), phleomycin (Sigma) or 4-NQO (Sigma)) or mock-treated. Phleomycin was resuspended in ddH2O to a stock concentration of 20 mg/ml, 4-NQO was resuspended in DMSO to a concentration of 10 mg/ml (and protected from light). Cells were incubated in shaking suspension at 100 rpm for 1 hour, in the dark in the case of 4-NQO. After incubation, 1×10^4 cells were diluted 1:100 in KK2 and 250 cells were mixed with 350 µl *K. aerogenes* and spread on to 140 mm SM agar plates. The plating was performed twice for each condition. The plates were incubated in the dark and survival was assessed by observing plaque formation after 3, 4, 5 and 6 days.

For cisplatin-treated cells, cells were resuspended in Pt buffer to 1×10^6 cells/ml, before the addition of cisplatin (Sigma). Cisplatin stock was prepared by resuspending the powder to 3 mM in Pt buffer immediately before the experiment, and protecting from light. Incubation of cells with cisplatin was performed in the dark, for 5 hours, before plating out as described previously.

2.2.4.2. Spore sensitivity assay

Spores were developed and their germination was induced by the protocol outlined in Section 2.2.2.6. The germinating spores were immediately resuspended to 1×10^6 cells/ml in a HL5/KK2 mixture (1:5 ratio), and the indicated doses of phleomycin (Sigma) were

administered. Cells where then incubated in shaking suspension for 16 hours at 22 °C, before plating out as described in Section 2.2.4.1. Spore germination rates in untreated samples were assessed by direction observation using a haemocytometer and Motic AE21 trinocular microscope.

2.2.4.3. Chromatin extraction

Exponentially growing *Dictyostelium* cells were resuspended to a density of 5×10^6 cells/ml in HL5, and incubated with genotoxic agents with rotation. For MMS (Sigma), phleomycin (Sigma) and 4-NQO (Sigma), incubation was for 1 hour (4-NQO-treated cells were incubated in the dark). For camptothecin (Enzo), which is resuspended 10 mg/ml in DMSO, incubation was in the dark for 3 hours. For cisplatin (Sigma), cells are resuspended to 5×10^6 cells/ml in Pt buffer instead of HL5, and the incubation was in the dark for 5 hours. Following incubation, the cells were washed with KK2 and resuspended in Nuclear Extraction Buffer (NEB) with 0.1% Triton X-100 to a density of 1×10^7 cells/ml. This mixture was incubated on ice for 15 minutes, before centrifugation at 14,000 ×g for 3 minutes at 4 °C. The pellet was then resuspended in the same volume of NEB with 0.1% Triton X-100 again, and incubated on ice for a further 15 minutes, before centrifugation at 14,000 ×g for 3 minutes at 4 °C. The pellet was then resuspended in NEB with 200 μg/ml RNase A, and incubated for 30 minutes at room temperature, with rotation, before the centrifugation step was repeated. The final pellet was resuspended to 6.25×10^4 cells/ml in 2× SDS loading buffer, followed by boiling at 100 °C for 5 minutes. Analysis of chromatin extracts was performed by SDS-PAGE and Western blotting (Sections 2.2.5.3 and 2.2.5.4). Whole-cell extracts were also prepared following treatment and washing in KK2, and were resuspended to 2.5×10^4 cells/ml in 2× SDS loading buffer, followed by boiling at 100 °C for 5 minutes.

2.2.4.4. Co-immunoprecipitation of Myc-APL

Exponentially growing *Dictyostelium* cells were resuspended to a density of 5×10^6 cells/ml, and 1×10^7 cells per condition were treated with genotoxic agents (as described in Section 2.2.4.3). Following treatment, the cells were washed twice in KK2, before resuspension to 1×10^7 cells/ml in IP Lysis Buffer, before incubation for 30 minutes at 4 °C, with gentle rotation. Cell lysates were centrifuged at 14,000 rpm for 10 minutes at 4 °C, and the supernatant was then transferred to 50 µl pre-washed Protein G Sepharose® beads (in a 1:1 ratio with IP Beads buffer, see below for pre-washing method; Sigma). This was incubated for 1 hour at 4 °C, with gentle rotation. A 20 µl aliquot of the supernatant was also boiled in 4× SDS loading buffer with 0.1 M DTT (100 °C for 5 minutes), as the input to the co-immunoprecipitation. The Protein G Sepharose® beads were separated from the solution by centrifugation (500 rpm for 2 minutes at 4 °C), and the supernatant was transferred to a fresh tube containing 50 µl pre-washed Anti-c-Myc agarose beads (in a 1:1 ratio with IP beads buffer; Sigma), and subsequently incubated for 90 minutes at 4 °C, with gentle rotation. The Anti-c-Myc agarose beads are then removed from the solution by centrifugation (500 rpm for 2 minutes at 4 °C), and a 20 µl aliquot of the supernatant was boiled with 4× SDS loading buffer with 0.1 M DTT (flow-through). The remaining supernatant is discarded. The Anti-c-Myc agarose beads are then washed 3 times with 1 ml IP Wash buffer, with each wash separated by centrifugation (500 rpm for 2 minutes at 4 °). Elution is performed by addition of 30 µl 4× SDS loading buffer with 0.1M DTT, followed by boiling at 100 °C for 5 minutes. Analysis is performed by silver staining (SilverQuest™ Silver Staining Kit; Thermo Fisher Scientific), following the manufacturer's instructions, or by SDS-PAGE and Western blotting (Sections 2.2.5.3 and 2.2.5.4).

For treatment of the samples with a deubiquitinating enzyme (USP2core, LifeSensors), after the third wash, an additional wash is performed with 1 ml IP DUB buffer, followed by resuspension of the beads with 200 µl IP DUB buffer. USP2core was added to a final concentration of 61 nM (500 ng), and the beads were then incubated at 37 °C for 2 hours, with shaking. The beads were then centrifuged at 500 rpm for 2 minutes, before the IP DUB buffer is removed and elution is performed as described previously.

Protein G Sepharose® and Anti-c-Myc agarose beads were both prepared by washing 5 times with 1 ml IP Beads buffer, prior to resuspension in a 1:1 ratio with IP Beads buffer.

2.2.5. Protein techniques

2.2.5.1. GST-purification of proteins

E. coli BL21(DE3) cells were transformed with pGEX-4T-2 (GE Healthcare) vectors containing *Dictyostelium* cDNA sequences to be expressed with N-terminal GST-tags. Induction of expression of GST-proteins was mediated by addition of 1 mM IPTG (Sigma) to the LB broth (containing 50 µg/ml ampicillin (Sigma)). *E. coli* cells were grown overnight at 25 °C (to an OD_{600} of 0.6-0.8), and the cells were harvested by centrifugation at 4000 ×g for 10 minutes. The cell pellet was resuspended in 10 ml lysis buffer (1× PBS, 1% Triton X-100, protease inhibitor cocktail (Roche)), and lysis was performed by sonication (4 times for 30 seconds each). The resulting solution is centrifuged at 13,300 rpm for 20 mins (4 °C), and the supernatant is added to 0.5 ml pre-washed Glutathione Sepharose 4B beads (GE Healthcare), before incubation at 4 °C for 1 hour (with gentle rotation). Glutathione Sepharose 4B beads are prepared by washing 3 times with 1 ml lysis buffer, and are then resuspended in a 1:1 ratio with lysis buffer. After incubation, the beads are removed from the solution by centrifugation (3000 rpm for 5 minutes at 4 °C) and the supernatant is removed. The beads are then washed twice with lysis buffer, and then twice

with 1× PBS and protease inhibitor cocktail only. Following the final wash and centrifugation, the GST-protein is eluted from the beads by the addition of 300 µl Tris (pH 8.0) with 5 mM reduced glutathione, and incubation for 30 minutes at 4 °C (with rotation). The beads are then centrifuged to separate them from the supernatant, and the supernatant (containing the GST-protein) is removed. The elution step is repeated three times. The protein content of the elution fractions is verified and quantified by comparison by SDS-PAGE and comparison with BSA standards (Section 2.2.5.3).

2.2.5.2. *In vitro* determination of PAR-binding

GST-tagged proteins are serially diluted, and increasing concentrations of each protein were either slot-blotted or dot-blotted onto a nitrocellulose membrane (5 µl per dot). Concentrations ranged from 0.625 pmol to 2.5 pmol. The membrane was subsequently blocked with 5% milk in TBS-T, before incubation with PAR polymers (1:5000 dilution; Trevigen) for 1 hour at room temperature. The membrane is then washed with TBS-T, before 4 washes with TBS-T with 1 M NaCl, and a further TBS-T wash. Detection was then performed by Western blotting with anti-PAR and anti-GST antibodies (see Section 2.2.5.3).

2.2.5.3. Poly-acrylamide gel electrophoresis and Western blotting

Protein samples were analysed in sodium dodecyl sulphate polyacrylamide gels, run at 200 V alongside pre-strained protein markers (PageRuler Plus, Fermentas). Gels were either pre-cast 4-15% gradient gels (Thermo Fisher Scientific), or single-percentage gels. Samples were then transferred onto PVDF membrane (Millipore, 0.45 µM pore size) at 100 V for 1 hour. The membrane was then blocked for 1 hour with 5% milk in TBS-T (blocking buffer), before incubation with the primary antibody which was diluted in blocking buffer. Primary antibody dilutions and conditions used are given in Table 2.1.

Antibody	Description	Manufacturer	Dilution
Anti-Myc	Mouse, monoclonal	Santa Cruz	1:1000
Anti-H3	Goat, polyclonal	Abcam	1:2000
Anti-γH2AX	Rabbit, polyclonal	Abcam	1:1000
Anti-Actin	Goat, polyclonal	Santa Cruz	1:2000
Anti-APL	Sheep, polyclonal	See Section 2.2.5.4	1:2500
Anti-PAR	Rabbit, polyclonal	Trevigen	1:1000
Anti-GST	Rabbit, polyclonal	Sigma	1:1000

Table 2.1: Primary antibodies used for Western blotting on PVDF membranes.

Following incubation with the primary antibody, the membrane was washed 3 times with TBS-T, for 10 minutes, before incubation with the relevant horseradish-peroxidase-conjugated (HRP) secondary antibody, which was diluted to 1:4000 in blocking buffer (1 hour at room temperature). A further three TBS-T washes are performed (10 minutes each), before detection with ECL (Millipore), following the provided protocol.

Analysis of protein content of SDS-PAGE gels was also provided by Coomassie staining with SimplyBlue^{TM} SafeStain (Life Technologies), using the provided protocol.

2.2.5.4. Anti-APL antibody generation

The antigen for raising the Anti-APL antibody, purified His-tagged macro domain region of APL (aa: 343-563; ΔAPL), was kindly provided by Ivan Ahel, in 50 mM HEPES (pH 7.0), 150 mM NaCl and 2 mM DTT. Aliquots of this antigen were prepared as requested, and were then injected into a sheep, following the immunisation schedule shown in Table 2.2. Crude serum, containing the anti-APL antibody, was sent to us for analysis and purification. This service was provided by the Scottish National Blood Transfusion Service.

Date	Action
25-Sept-2013	Immunisation
23-Oct-2013	Injection
30-Oct-2013	Donation
20-Nov-2013	Injection
27-Nov-2013	Donation
18-Dec-2013	Injection
25-Dec-2013	Donation

Table 2.2: Immunisation schedule for the generation of the anti-APL antibody in

2.2.5.5. Affinity purification of the anti-APL antibody with GST-APL

The purification column (Poly-Prep® Chromatography Column, Bio-Rad) was formed by coupling GST-APL to CNBr-Activated Sepharose 4B beads (GE Healthcare), following the manufacturer's instructions. Briefly, 0.25 mg of GST-APL was dialysed into coupling buffer (0.1 M $NaHCO_3$ (pH 8.3), 0.5 M NaCl) using a Slide-A-Lyzer™ G2 Dialysis Cassette (Thermo Fisher Scientific), overnight at 4 °C. CNBr-Activated Sepharose 4B beads were prepared as instructed to give a final volume of 1.25 ml, before addition of GST-APL and incubation for 1 hour at room temperature (with gentle rotation). The beads were then washed with 5 gel volumes of coupling buffer, before blocking with 0.1 M Tris-HCl (pH 8.0) for 2 hours. The beads were then washed three times with a cycle of 0.1 M acetic acid (pH 4.0) and 0.5 M NaCl, followed by 0.1 M Tris-HCl (pH 8.0) and 0.5 M NaCl.

Immediately prior to purifying the anti-APL antibody serum, the beads were washed with 10 bed-volumes of 10 mM Tris (pH 7.5), followed by 10 bed-volumes of 100 mM glycine (pH 2.5), and then 10 bed-volumes of 10 mM Tris (pH 8.8). The pH was then checked to

ensure that it was pH 8.8. 10 bed-volumes of 100 mM trimethylamine (pH 11.5) were then added, before addition of 10 bed-volumes of 10 mM Tris (pH 7.5), and the pH was checked to ensure it was 7.5. Anti-APL serum (5 ml) was diluted 1:10 with 10 mM Tris (pH 7.5) and passed through the column three times (at 4 °C). The column was then washed with 20 bed-volumes of 10 mM Tris pH (7.5), and then 20 bed-volumes of 10 mM Tris (pH 7.5) with 500 mM NaCl. Elution was performed by addition of 10 bed-volumes of 100 mM glycine (pH 2.5), and collected in 1.5 ml tubes containing 1 bed-volume of 1 M Tris (pH 8.0). The column was then washed with 10 mM Tris (pH 8.8), before a second elution with 10 bed-volumes of 100 mM trimethylamine (pH 11.5), into 1.5 ml tubes containing 1 bed-volume of 1 M Tris (pH 8.0). The protein content of the elution fractions was assessed by SDS-PAGE and Coomassie staining (Section 2.2.5.3), and those fractions containing the highest concentrations of purified anti-APL antibody were combined and dialysed with 1× PBS and 5% glycerol, overnight at 4 °C, before snap-freezing on dry ice and storage at -80 °C.

3. *In silico* identification of novel *Dictyostelium* proteins with PAR-binding domains

3.1. Introduction

The overall aim of this thesis is to identify and characterise putative *Dictyostelium* proteins that are involved in the cellular response to DNA damage, using PAR-binding domains as surrogate markers of such DDR proteins. Several human proteins with known PAR-binding domains: PBMs, macro domains, PBZ domains and WWE domains, have been discussed previously. These proteins have roles in many cellular pathways, including the DDR. Therefore, the presence of a PAR-binding domain alone is not indicative of a role for the protein in the DDR; however, ADP-ribosylation is an increasingly apparent and important post-translational modification in DNA repair. Furthermore, DNA repair proteins contain an elevated number of these domains compared to proteins involved in other processes: all of the human PBZ domain-containing proteins are implicated in the DDR[241]. ADP-ribosylation is conserved in *Dictyostelium*, with three ARTs: Adprt1a, Adprt1b and Adprt2, currently being identified to be activated in response to DNA damage. The targets of these ARTs are still being identified, and the aim of this chapter is to identify novel proteins containing domains that may bind to these targets. The identification of such proteins could lead to the discovery of such protein targets, and help to build a more complete picture of the ADP-ribosylation network in *Dictyostelium*.

3.1.1. Known *Dictyostelium* proteins with PAR-binding domains

To further our understanding of PARylation in Dictyostelium, it is important to identify proteins that have the potential to bind to ADP-ribose or PAR. Such proteins would likely contain the same PAR-binding domains found in mammalian proteins. There are currently three *Dictyostelium* proteins with annotated macro domains (Figure 3.1). Two of these have been identified by sequence homology to be orthologs of human macro domain-containing proteins, whereas the third is uncharacterised. GDAP2 is the first identified homolog. The function of human GDAP2 is not known; however, the mouse homolog of the protein has been implicated in cell differentiation, and is transcribed strongly in the adult mouse brain[378]. The macro domain found in human GDAP2 binds to poly(A), not ADP-ribose or PAR[247]. The second homolog is a *Dictyostelium* ART, named pARTg. It has been inferred from similarity of its PARP catalytic domain that this ART is the homolog of ARTD8, one of the human macroPARPs that has recently been implicated in DSBR in response to replication stress[35]. However, there is currently no experimental evidence linking pARTg to a cellular function. In addition to the single macro domain, the *Dictyostelium* pARTg protein contains a WWE domain and a RING domain, whereas human ARTD8 has three macro domains and a WWE domain[321]. The single macro domain of pARTg makes it unlike the three human macroPARPs, which all have multiple macro domains. The final macro domain containing protein is uncharacterised (UniProt accession: Q54YH9), and contains one macro domain and three WWE domains. Due to the diverse nature of the macro domain superfamily, it is likely that there are more *Dictyostelium* proteins with currently unannotated macro domains, and potentially more human proteins too. Additionally, there is a clear difference between the number of macro domain-containing proteins in humans and *Dictyostelium*, suggesting that more may be found in the latter organism.

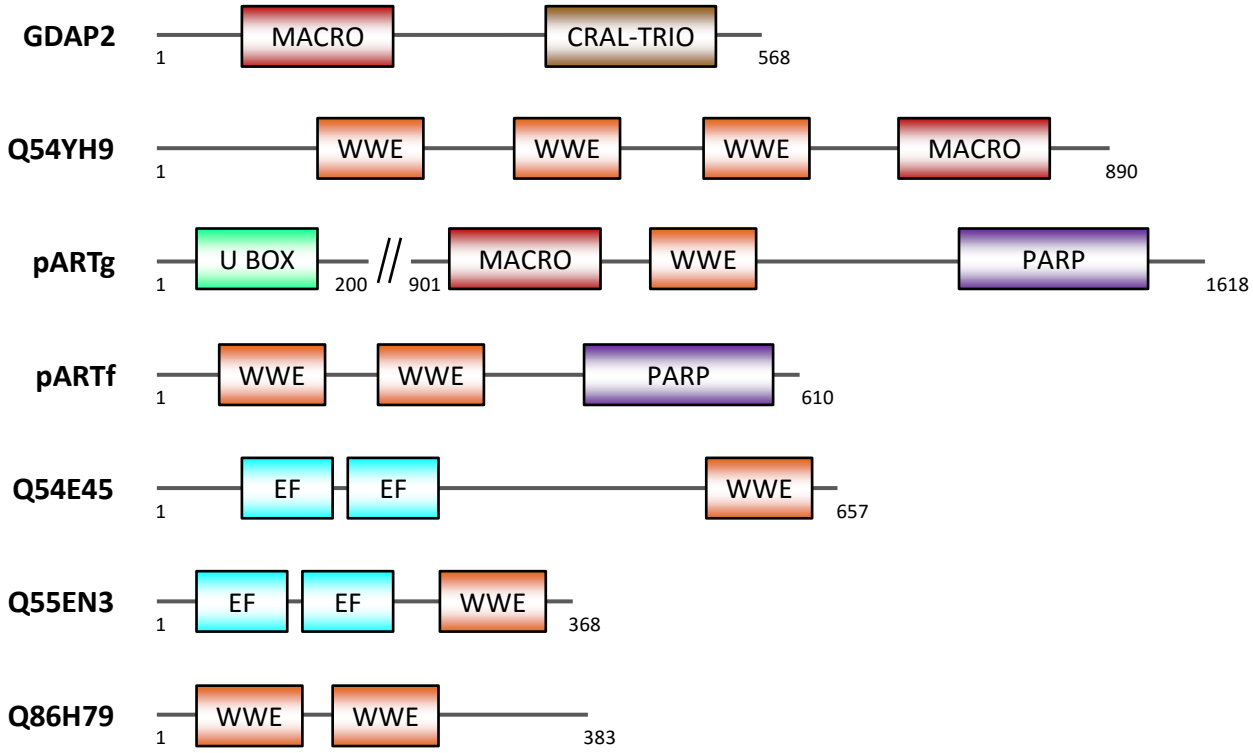


Figure 3.1 – Domain architecture of *Dictyostelium* macro- and WWE-domain-containing proteins. The domain structure of the *Dictyostelium* proteins with currently annotated macro or WWE domains. Three proteins have been identified as containing macro domains: a homolog of human GDAP2, an uncharacterised protein (Q54YH9), and an ART (pARTg). The latter two proteins additionally contain three and one WWE domains, respectively. Another ART (pARTf) contains two WWE domains, as does another uncharacterised protein (Q86H79). Additionally, two predicted calcium ion-binding proteins (Q54E45 and Q55EN3) contain a single WWE domain each. Domain abbreviations: PARP catalytic (PARP), EF-hand (EF).

There are six *Dictyostelium* proteins with automatically annotated WWE domains (Figure 3.1). Two of these are found in ARTs: the previously described pARTg, and pARTf, the catalytic domain of which displays homology to that of ARTD6 (BLAST results, not shown). Aside from the PARP catalytic domain, there are no additional domains in pARTf, and no experimental evidence addressing its cellular function[363]. Additionally, there is a protein identified with the UniProt accession Q54YH9, with three WWE domains, which has been introduced previously due to its macro domain. The remaining three proteins (UniProt accession numbers: Q54E45, Q55EN3 and Q86H79) are experimentally uncharacterised; however, both Q54E45 and Q55EN3 have a pair of calcium binding EF hand domains, in addition to their single WWE domain[379]. The protein Q86H79 contains two WWE domains and no other recognised domains. The WWE domain annotations found here have all been performed automatically, so there may be additional proteins in *Dictyostelium* with WWE domains.

The study that first identified the PBZ domain as a PAR-binding module also performed *in silico* homology searches to find proteins containing this domain[241]. They uncovered a PBZ domain in three human proteins, previously described, and seven *Dictyostelium* proteins (shown in Figure 3.2). Of these seven proteins, CHFR was the only putative homolog of a human PBZ domain-containing protein; however, there is no experimental data published to confirm that this protein is indeed CHFR. Putative homologs of DDR proteins Rad17, Chk2 and an FHA domain-containing uracil DNA-glycosylase (Ung) were identified. Human Rad17 is a checkpoint protein that is phosphorylated by ATM and ATR in response to various forms of DNA damage[380]. Rad17 has been implicated in the loading on the Rad9-Rad1-Hus1 (9-1-1) complex onto chromatin, leading to cell-cycle checkpoint activation[381–383]. Chk2 is a protein kinase involved in the cell-cycle arrest in response to detection of DSBs. Phosphorylation by ATM activates Chk2, which then phosphorylates

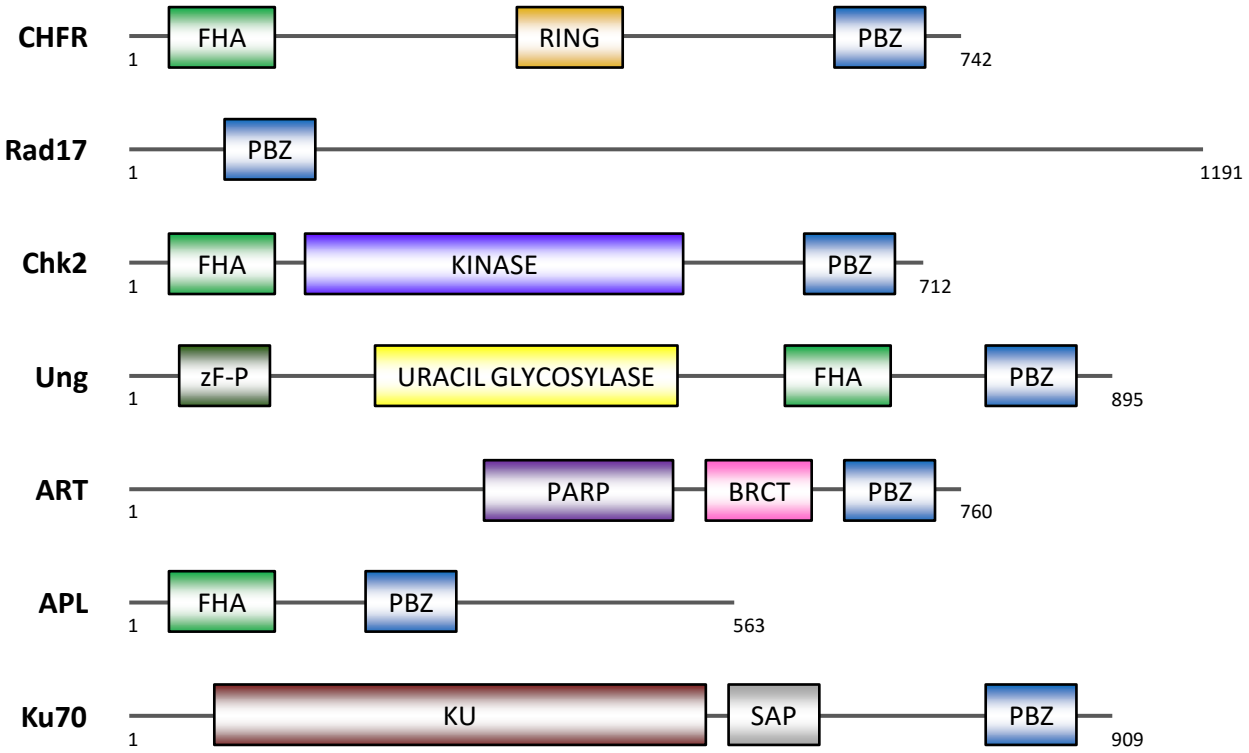


Figure 3.2 – Domain architecture of *Dictyostelium* PBZ-domain-containing proteins. The domain structure of the seven *Dictyostelium* proteins with currently annotated PBZ domains. CHFR, Rad17, Chk2 and Ku70 are homologs of human DDR proteins; however, CHRF is the only human protein of these with a PBZ domain itself. Ung is an identified uracil glycosylase, whereas ART is an uncharacterised PARP-domain-containing protein. APL is a protein without known function, containing only an FHA domain alongside a PBZ domain. Domain abbreviations: forkhead associated (FHA), RING finger (RING), C2H2-type zinc finger (PBZ), PARP-type zinc finger (zF-P), PARP catalytic (PARP), BRCA1 C-terminus (BRCT).

downstream targets, such as p53 and BRCA1, resulting in cell-cycle arrest[384–386]. PARylation has not been implicated in the function of Rad17 or Chk2 in humans. Additionally, a PBZ domain is found in a *Dictyostelium* ART (unclassified), alongside WWE and BRCT domains. There is no human ART with an identified PBZ domain, but it may play a role similar to the macro domains of the macro PARPs. However, the role of this ART in *Dictyostelium* has not been ascertained. Also identified in the homology search was a protein with no human homolog, but containing an N-terminal FHA domain with sequence homology to the human DDR proteins APLF, APTX and PNKP (hereby referred to as aprataxin-like protein (APL); UniProt accession: Q54B72). The presence of this FHA domain may indicate a role for this protein in the DDR; however, there is no experimental evidence to support this.

The only *Dictyostelium* PBZ-domain-containing protein to be experimentally studied is Ku70, which contains a C-terminus, PAR-binding PBZ domain that is absent in the human homolog. As described previously, this PBZ domain has been shown to stabilise NHEJ factors at DSBs, and its deletion greatly reduces the capability of cells to perform NHEJ[360]. In humans, the tandem PBZ domains of APLF perform this function, and it appears that some kind of domain switching event has occurred during the course of evolution as humans and *Dictyostelium* diverged[15]. While the overall mechanism of PAR-dependent stability of NHEJ complexes has been conserved, but the domain responsible for this is found in different proteins. We hope that this example is not the only occurrence of domain switching, and that we may identify novel DNA repair proteins by genome-wide, *in silico* protein homology searches. The sequence variability of the macro domain superfamily in particular makes such searches difficult, so therefore we will be applying advanced bioinformatics techniques.

3.1.2. Methods of *in silico* protein homology searching

In parallel with the increasing size and availability of protein and DNA sequence databases, there has also been an increasing demand for tools with which to search them. Upon encountering a new protein, one of the earliest indications of function can lie in the domains or regions present in the protein, which can be identified if they display homology to already characterised or annotated domains in other proteins. To perform this analysis, algorithms for protein homology searching have been developed, and are continuously increasing in their sophistication and speed.

Protein homology searching tools fall into two general categories: global and local. Global homology searching algorithms look for complete matches of the input (or seed) sequence with database proteins, whereas local algorithms match regions of the seed sequence with regions of the database proteins[387,388]. Therefore, global algorithms are less able to match regions of high homology to the seed in database proteins if they are surrounded by regions of low homology, which is often the case for evolutionarily distant enzymes sharing an active site. Local regions of homology shared between proteins typically represent domains, making local algorithms much more suitable for identifying specific domains. Therefore, local homology searching tools will be employed for searching for proteins with novel PAR-binding domains.

One of the most commonly encountered local homology searching algorithms is the Basic Local Alignment Search Tool (BLAST)[370]. BLAST directly compares the input sequence with those in the database in a pairwise manner, seeking regions of maximal local homology. The output of BLAST is a list of proteins with identified regions of homology with some or all of the input sequence, alongside assigned expectation values (E-values). The E-value of a homology match represents the chance that the match could have

occurred randomly, taking into account the input sequence length and database size. Therefore, the closer the E-value is to zero, the more confidence can be taken that the regions of the input and matched sequences are homologous.

BLAST utilises a comparatively simple and fast algorithm, making it an attractive choice for many applications. However, homology searching within domain families containing more varied sequences requires more complicated mathematical and statistical tools. One main class of such tools use profile hidden Markov models (profile-HMMs) for homology searching[389]. The input of such tools is a profile-HMM instead of a single sequence. A profile-HMM is constructed from a multiple sequence alignment of homologous domain sequences, which are converted into a mathematical construct representing the amino acid variations at every point of the sequences as a probabilistic model (including insertions and deletions). Therefore, a profile contains substantially more information about the input domain than a single sequence, and exploiting this variation allows for more variant members of a family of domains to be identified. Initially, profile-HMMs were used to search sequence databases in profile-sequence searches[371]. However, greater sensitivity can be achieved by using profile HMMs to search databases consisting of other profile-HMMs, in profile-profile comparison[390]. Thereby, the sequence variation encoded in the profile is applied to both sides of the search.

Typically, protein sequence homology searching tools are applied together to maximise their effectiveness and to minimise false positives. A sample workflow is shown in Figure 3.3. To identify novel proteins containing a known domain, a well characterised sequence from this domain family is used as an input to a BLAST search. The sequences of the statistically significant hits of this search are then aligned and used to form the profile-HMM for more sensitive profile-sequence and profile-profile searches. The sequence

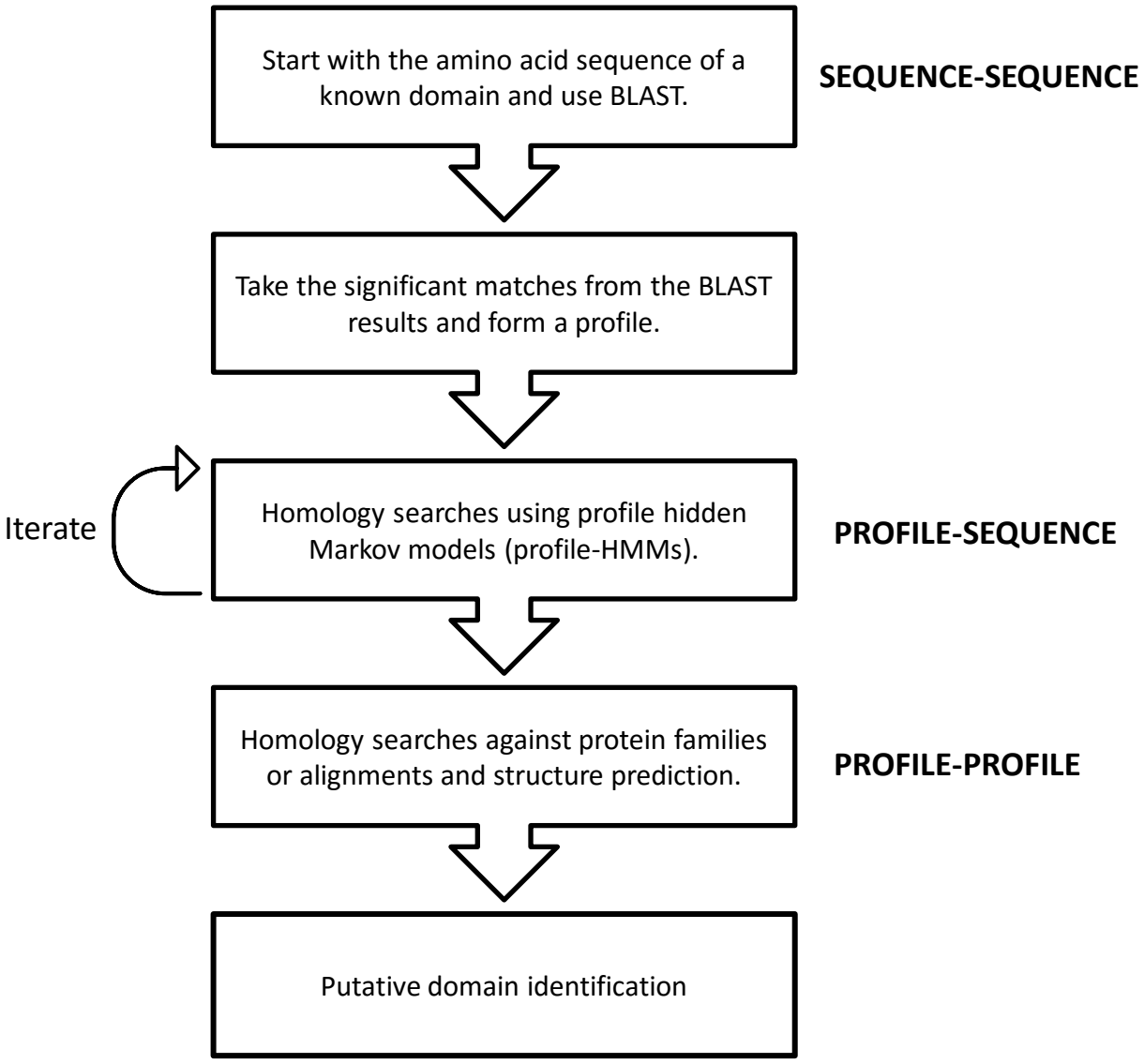


Figure 3.3 – Overview of the *in silico* approach to identify previously undiscovered PAR-binding domains. The starting point of this workflow is the amino acid sequence of a known domain, which is used in BLAST sequence-sequence searches. The results of this BLAST are used to form a profile, which is the input for the more sophisticated profile-HMM based algorithms. Profiles are first used to search sequence databases in an iterative manner, with the output of one search is used as the input for the next. Following this, profile-profile searches are employed to incorporate the advantages of the profile on both sides of the search.

output of these can then be aligned and resubmitted in an iterative manner to provide a more powerful search.

The above homology detection algorithms are based around comparison of the primary structure of proteins. However, the conservation of function of a domain is a much more valuable discovery, and function is determined by the tertiary structure of a protein. Multiple algorithms exist for comparison of known domain structures with those in structural databases, and these use various metrics for calculating the similarity of two structures and aligning them[391,392]. Of course, the structural alignment of a domain requires a solved structure of that domain, and a suitably large database of structures from which to search. These are currently the major limiting factors of these techniques. Alternatively, tools have been developed for the prediction of both secondary and tertiary structures of proteins. Two categories of tertiary structure modelling tools are *de novo* and comparative (or homology) protein modelling. Briefly, *de novo* modelling utilises physical laws to build stable protein conformations (energy minimums) from knowledge of only the primary sequence of the protein or domain in question, whereas homology modelling bases its prediction on a known template structure, and requires a multiple sequence alignment of the query and template sequence[374,393–396]. *De novo* methods are typically computationally expensive; whereas homology modelling relies on an expectation of what the domain being modelled looks similar too, and will not predict truly novel structures. Several secondary structure prediction tools are also available, relying on inputs of single sequences and multiple sequence alignments, and utilising neural network and machine learning algorithms in their prediction[397,398].

3.1.3. Aims

In this chapter, the previously introduced methods of *in silico* homology searches will be applied to identify novel proteins containing PAR-binding domains in *Dictyostelium*. As has been illustrated with the PBZ domains in *Dictyostelium* Ku70 and human APLF, PAR-binding domains may have evolved in different proteins in these two organisms, and may serve as surrogate markers for DNA repair proteins. The searches will be limited to the PBZ and macro domains, as these are found more frequently in human DDR proteins than WWE domains. The PBM will be excluded from this procedure as its 8 amino acid consensus motif is too short for currently available homology searching algorithms to provide statistically reliable results, as false and true positives cannot be distinguished[399]. Therefore, the *in silico* approach is questionable for this motif. Advanced techniques employing profile-HMMs will be used to search for diverse members of the macro domain superfamily, which may not be identified using BLAST. Putative PAR-binding *Dictyostelium* proteins will be bioinformatically analysed to gain as much information about their possible function prior to the commencement of experimental assays.

3.2. Results

3.2.1. Analysis of PBZ domain-containing proteins in *Dictyostelium*

We firstly performed genome-wide homology searches for novel proteins with PBZ domains. PBZ domains have specificity for binding to PAR chains, and so identifying a PBZ domain in a protein would be a strong indicator of that protein binding to PARylated proteins *in vivo*. The consensus sequence of the PBZ domain is a well-defined, twenty-four amino acid sequence of $[K/R]X_2CX[F/Y]GX_2CXBBX_4HX_3[F/Y]XH$, with B and X

representing basic and unconstrained amino acids, respectively[241]. This sequence is quite short for homology searching algorithms, but is suitable for this analysis.

A comprehensive study and identification of PBZ domains in proteins has already been performed, and this study highlighted the apparent increased reliance of *Dictyostelium* DNA repair on proteins containing PBZ domains, compared to humans: seven *Dictyostelium* proteins have thus far been found to contain PBZ domains, in comparison to three human proteins[241]. An alignment of the known human and *Dictyostelium* proteins is shown in Figure 3.4. As *in silico* investigations have already been performed for this domain, we decided to start with the more sophisticated profile-sequence and profile-profile HMM based methods for homology searching (HMMer2/HMMer3, HHpred)[371,372,390]. The alignment in Figure 3.4 serves as the input for these processes, and the profile-HMMs were formed from this. Despite multiple iterations and extensive searching, we were unable to find any additional proteins in *Dictyostelium* or human that contain novel PBZ domains.

3.2.2. Identification of novel *Dictyostelium* proteins with macro domains

Unlike the PBZ domain, which has a strict consensus sequence, there is much more sequence variability across the macro domain superfamily, and macro domains have been shown to bind to a variety of ligands[246]. The multiple sequence alignment of currently annotated human and *Dictyostelium* proteins is shown in Figure 3.5, and although this alignment does display blocks of homolog, these blocks do not stretch across all members of the superfamily. This variability increases the likelihood that *Dictyostelium* proteins may exist with macro domains that have escaped automatic annotation. However, the lack of a defined consensus sequence makes identification of these domains more challenging.

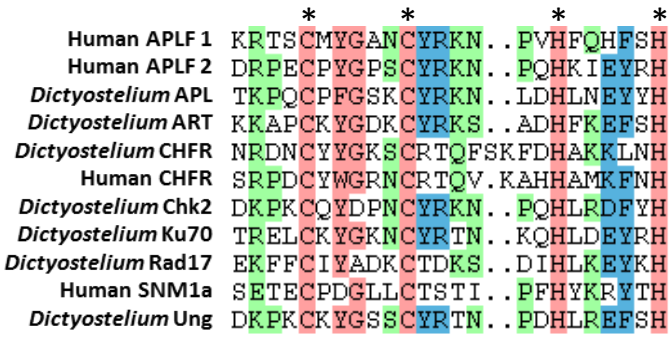


Figure 3.4 – Multiple sequence alignment of known human and *Dictyostelium* PBZ domains. The sequences of the PBZ domain-containing regions of human and *Dictyostelium* proteins were aligned using T-Coffee. The completely conserved cysteine and histidine residues that define the domain are marked by asterisks. Sequences are coloured by BLOSUM62 score: 0.5-1.5 – green, 1.5-3 – blue, 3< – red, where a higher number represents stronger conservation of an amino acid.

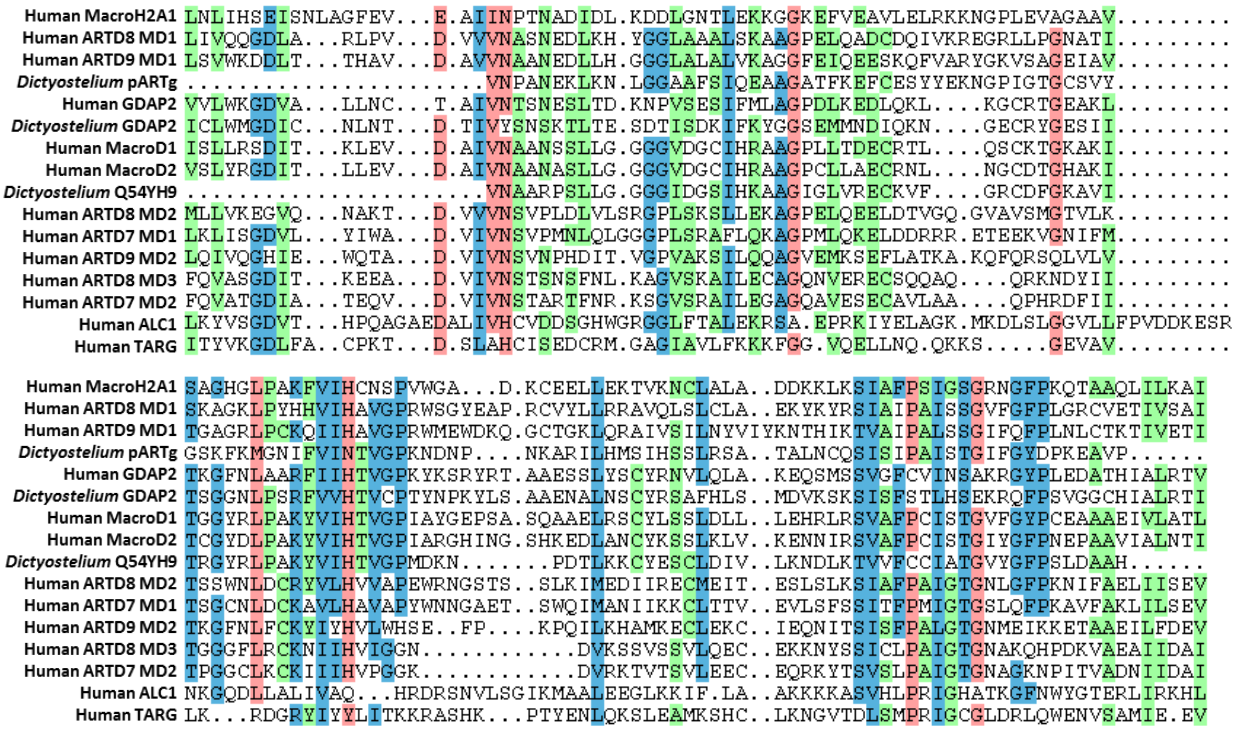


Figure 3.5 – The alignment of all known *Dictyostelium* and human macro domains. This multiple sequence alignment contains the sequences of all of the currently known macro domains in humans and *Dictyostelium*. This alignment is used to form a profile-HMM for further homology searches for novel macro domain-containing proteins. Sequences are coloured by BLOSUM62 score: 0.5-1.5 – green, 1.5-3 – blue, 3< – red, where a higher number represents stronger conservation of an amino acid.

Therefore, the more sophisticated approach utilising profile-HMMs is necessary for the *in silico* search for novel macro domains.

As we are searching for macro domains that act as PAR-binding domains, a relevant starting point for the search is the sequence of a known PAR-binding macro domain. It would be hoped that an undiscovered, novel PAR-binding domain would show more sequence similarity to such a domain, as opposed to one that binds to poly(A), for example. The macro domain of ALC1 is a well-characterised PAR-binding macro domain that is crucial for the function of ALC1 as a helicase in the DDR in humans[250]. Therefore, the sequence of this macro domain is a highly suitable starting point for our homology searches. Additionally, there is no currently identified homolog of ALC1 in *Dictyostelium*, and it is possible that a putative homolog may be currently unannotated. We also performed BLAST searches using other macro domains found in human proteins as their starting point, but these did not yield any additional *Dictyostelium* proteins.

Firstly, the primary sequence of the macro domain of human ALC1 was used as the input to a BLAST search. The results of this homology search, which are shown aligned in Figure 3.6, uncovered a previously unannotated macro domain in *Dictyostelium* DNA Ligase III (Lig3; UniProt accession: Q54QM4), with an E-value of 0.02. Human Lig3 is a well-characterised human DDR protein, with confirmed functions in SSBR, NER and alt-NHEJ[68,180,235]. This E-value is higher than might be expected to imply that the result is significant; however, this is likely due to the diversity within the macro domain superfamily. Despite Lig3 not being a novel DDR protein in *Dictyostelium*, the discovery of a central macro domain is novel as no human DNA ligase has a macro domain; however, DNA Ligase III orthologs from other species have been found to contain macro domains (*Naegleria gruberi* - amoeba-flagellate) or PBZ domains (*Strongylocentrotus purpuratus* – sea urchin)[400]. Interestingly, none of the other currently annotated human or

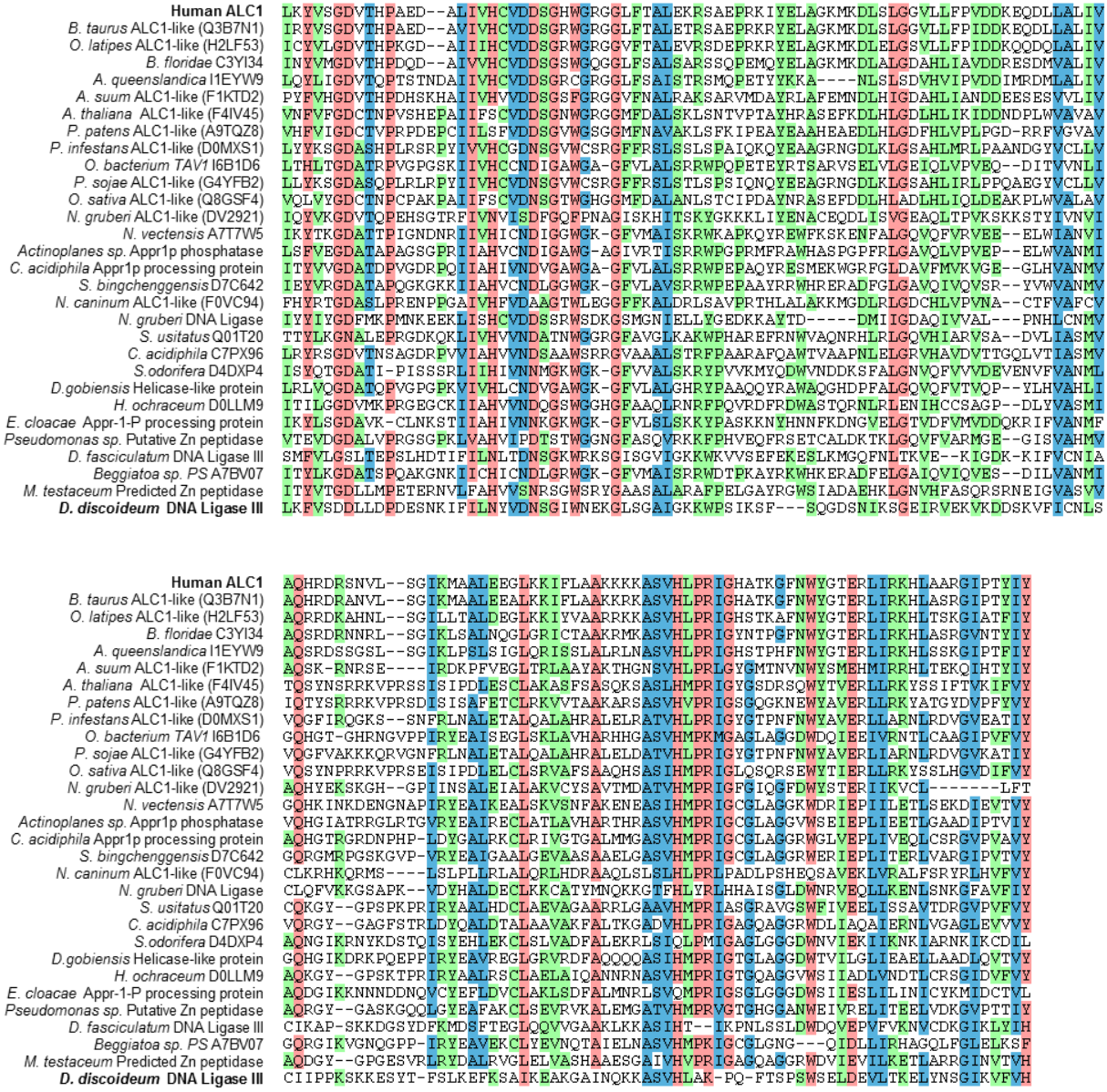
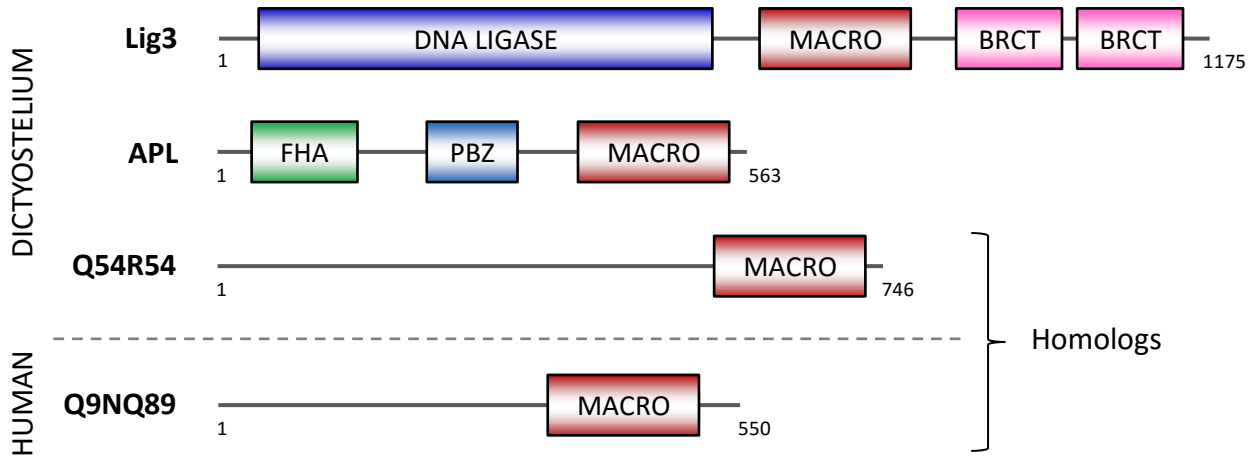


Figure 3.6 – A BLAST search conducted with the macro domain of ALC1 as an input reveals a macro domain in *Dictyostelium* DNA Ligase 3. A multiple sequence alignment of the statistically significant (E value < 0.05) results of a BLAST search using the sequence of the macro domain of ALC1 as its input, ordered by sequence similarity to human ALC1. This alignment was subsequently used to form a profile for future *in silico* searches. Proteins are labelled by UniProt accession numbers. Sequences are coloured by BLOSUM62 score: 0.5-1.5 – green, 1.5-3 – blue, 3< – red, where a higher number represents stronger conservation of an amino acid.

Dictyostelium macro domain containing-proteins appeared in the BLAST results from the macro domain of ALC1. This illustrates the combined challenge of the limitations of BLAST with the diversity of the macro domain superfamily.

Following the BLAST search with the macro domain of ALC1, we wished to employ techniques using profile-HMMs. Profile-HMMs encapsulate the natural variation of amino acid sequences with a domain family, and will be more suited to exploring the space of possible macro domains. Therefore, the sequences that resulted from the BLAST search with an E-value less than 0.05 were aligned (Figure 3.6) and converted into a profile. Additionally, a profile consisting of an alignment of the sequences of all of the known human and *Dictyostelium* macro domains was formed (Figure 3.5). Using these two profiles in profile-sequence local homology searches with HMMER3 and HMMER2 revealed a further two *Dictyostelium* proteins with unannotated macro domains. These proteins are the previously introduced, PBZ-domain-containing protein APL (UniProt accession: Q54B72), and a completely uncharacterised protein with UniProt accession number Q54R54. We also identified a human homolog of protein Q54R54, also containing an unannotated macro domain (UniProt accession: Q9NQ89). Therefore, as a result of this *in silico* genome-wide search, three *Dictyostelium* and one human protein with a novel macro domain have been identified. The domain structures and an updated alignment of known macro domains in human and *Dictyostelium* are shown in Figure 3.7. This alignment further illustrates the high variation in amino acid sequence between the different macro domains, as at many relatively well-conserved amino acid positions in the previous alignment (Figure 3.5), the newly discovered sequences do not display conservation.

A



B

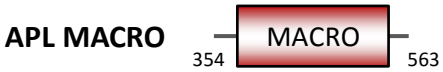
```
                    *
Dictyostelium APL   DD.A..TEISA...N.AISEYLHFH.......EK.EDDIKLKMMVEKSIYSDNL...IQSFKKH.FNDKWDKRFE...II
Dictyostelium GDAP2 GDIC...NLNT...D.TIVY............SN.SKTLTE.....SDTISDKI...F.KYGGSEMMNDIQKN....GEC
Dictyostelium DNA Ligase 3 SDDE...SNKI.....FILN.....YVDNSGIWN.....EKGLSGAI...G.KKWPS.IPKSFS.QGD.SNII
Dictyostelium Q54R54 NMSTELHF...DSIDS.QIH.....................TFKQSL...L.KEEEREK.EENNN.NNNNIKL
Dictyostelium Q54YH9 ..................VN.....AA.RPSLLG.....GGGIDGSI...H.KAAGIGLVRECKVF....GRC
Dictyostelium pARTg  ..................VN.....PA.NEKLKN.....LGGAAFSI...Q.EAAGATFKEFCESYYEKNGPI
Human MacroD2       GDIT...LLEV...D.AIVN.....AA.NASLLG.....GGGVDGCI...H.RAAGPCLLAECRNL....NGC
Human MacroH2A1     SEISNLAGFEV...E.AIIN.....PT.NADIDL.....KDDLGNTL...E.KKGGKEFVEAVLELRKKNGPL
Human ALC1          GDVT...HPQAGAEDALIVH.....CV.DDSGHW....GRGGLFTAL...E.KRSA.EPRKI.YELAGKMKDL
Human MacroD1       SDIT...KLEV...D.AIVN.....AA.NSSLLG.....GGGVDGCI...H.RAAGPLLTDECRTL....QSC
Human Q9NQ89        QECTDFHF...PRIEE.QLE....VVQQVVLYAR.TQ........RRSKLKESLDSGN.QNGGNDD.KTKNA.ERNYLNV
Human GDAP2         GDVA...LLNC...T.AIVN.....TS.NESLTD.....KNPVSESI...F.MLAGPDLKEDLQKL....KGC
Human TARG          GDLF..ACPKT...D.SLAH.....CI.SEDCRM.....GAGIAVLF...K.KKFGG.VQELLN.Q.....QK
Human ARTD7 MD1     GDVL...YIWA...D.VIVN.....SV.PMNLQL....GGGPLSRAF...L.QKAGPMLQKELDDRRR.ETEE
Human ARTD8 MD1     GDLA...RLPV...D.VVVN.....AS.NEDLKH.....YGGLAAAL...S.KAAGPELQADCDQIVKREGRL
Human ARTD9 MD1     DDLT...THAV...D.AVVN.....AA.NEDLLH.....GGGLALAL...V.KAGGFEIQEESKQFVARYGKV
Human ARTD7 MD2     GDIA...TEQV...D.VIVN.....ST.ARTFNR.....KSGVSRAI...L.EGAGQAVESECAVLA....AQ
Human ARTD8 MD2     EGVQ...NAKT...D.VVVN.....SV.PLDLVL....SRGPLSKSL...L.EKAGPELQEELDTVGQ.GVAV
Human ARTD9 MD2     GHIE...WQTA...D.VIVN.....SV.NPHDIT.....VGPVAKSI...L.QQAGVEMKSEFLATKA.KQFQ
Human ARTD8 MD3     GDIT...KEEA...D.VIVN.....ST.SNSFNL.....KAGVSKAI...L.ECAGQNVERECSQQA....QQ

Dictyostelium APL   KIENSNS.........LEQFNL........GC.KL...F..ATESTWRL...K...K.T.PQN....KQLYEML.DTG..
Dictyostelium GDAP2 RYGESII.........TSGGNL........PS.RF...VVHTVCPTYNP...K..YL.S.AAENALNSCYRSAF.HLS..
Dictyostelium DNA Ligase 3 KSGEIRV.........E...KVKDDSISTNKK.VFICNLSCIIPPKSKK...E..SY.T.FSLKEFKSAIKEAKGAIN..
Dictyostelium Q54R54 SNGDFFI.........TKHSNL........SDVQV...VFHLVIDSKNRTPLGGDNQ.L.STTSDIAKGLRNIL.LTA..
Dictyostelium Q54YH9 DFGKAVI.........TRGYRL........PA.KY...VIHTVGPMDKN.............PDTLKKCYESCL.DIV..
Dictyostelium pARTg GTGCSVY.........GSKFKM........GN.IF...VINTVGPKNDN...P.......NKARILHMSIHSSL.RSA..
Human MacroD2       DTGHAKI.........TCGYDL........PA.KY...VIHTVGPIARG...H..IN.G.SHKEDLANCYKSSL.KLV..
Human MacroH2A1     EVAGAAV.........SAGHGL........PA.KF...VIHCNSPVWGA.........D.KCEELLEKTVKNCL.ALA..
Human ALC1          SLGGVLLFPVDDKESRNKGQDL........LA.LI.......VAQHRDR...S..NVLSGIKMAALEEGLKKIF....LA
Human MacroD1       KTGKAKI.........TGGYRL........PA.KY...VIHTVGPIAYG...E..PS.A.SQAAELRSCYLSSL.DLL..
Human Q9NQ89        LPGEFYI.........TRHSNL........SEIHV...AFHLCVDDHVKS..GNI.....TARDPAIMGLRNIL.KVC..
Human GDAP2         RTGEAKL.........TKGFNL........AA.RF...IIHTVGPKYKS...R..YR.T.AAESSLYSCYRNVL.QLA..
Human TARG          KSGEVAV.........L...KR........DG.RY...IYYLITKKRAS...H..KP.T...YENLQKSLEAMK.SHC..
Human ARTD7 MD1     KVGNIFM.........TSGCNL........DC.KA...VLHAVAPYWNN...G..A..E.TSWQIMANIIKKCL.TTV..
Human ARTD8 MD1     LPGNATI.........SKAGKL........PY.HH...VIHAVGPRWSG...Y..EA.P.RCVYLLRRAVQLSL.CLA..
Human ARTD9 MD1     SAGEIAV.........TGAGRL........PC.KQ...IIHAVGPRWME...W..DK.Q.GCTGKLQRAIVSIL.NYVIY
Human ARTD7 MD2     PHRDFII.........TPGGCL........KC.KI...IIHVPGGK.................DVRKTVTSVL.EEC..
Human ARTD8 MD2     SMGTVLK.........TSSWNL........DC.RY...VLHVVAPEWRN...G..S..T.SSLKIMEDIIRECM.EIT..
Human ARTD9 MD2     RSQLVLV.........TKGFNL........FC.KY...IYHVLWHSE.....F.......PKPQILKHAMKECL.EKC..
Human ARTD8 MD3     RKNDYII.........TGGGFL........RC.KN...IIHVIGGN.................DVKSSVSSVL.QEC..

Dictyostelium APL   TFEKVTKNLYPNCGKIGKVY....P
Dictyostelium GDAP2 MDVKSKSISFSTLHSEKRQF....P
Dictyostelium DNA Ligase 3 QKKASVHLAKPQFTS....P....S
Dictyostelium Q54R54 SKYGIGNITIPICLTESE.LDL...
Dictyostelium Q54YH9 LKNDLKTVVFCCIATGVYGF....P
Dictyostelium pARTg TALNCQSISIPAISTGIFGY....D
Human MacroD2       KENNIRSVAFPCISTGIYGF....P
Human MacroH2A1     DDKKLKSIAFPSIGSGRNGF....P
Human ALC1          AKKKKASVHLPRIGHATKGF....N
Human MacroD1       LEHRLRSVAFPCISTGVFGY....P
Human Q9NQ89        CTHDITTISIPLLLVHDMSEEMTIP
Human GDAP2         KEQSMSSVGFCVINSAKRGY....P
Human TARG          LKNGVTDLSMPRIGCGLDRL....Q
Human ARTD7 MD1     EVLSFSSITFPMIGTGSLQF....P
Human ARTD8 MD1     EKYKYRSIAIPAISSGVFGF....P
Human ARTD9 MD1     KNTHIKTVAIPALSSGIFQF....P
Human ARTD7 MD2     EQRKYTSVSLPAIGTGNAGK....N
Human ARTD8 MD2     ESLSLKSIAFPAIGTGNLGF....P
Human ARTD9 MD2     IEQNITSISFPALGTGNMEI....K
Human ARTD8 MD3     EKKNYSSICLPAIGTGNAKQ....H
```

Figure 3.7 – Homology searching for novel *Dictyostelium* proteins with macro domains yields three results. A: The domain structures of the three *Dictyostelium* proteins, and one human protein, that have been identified to contain previously undetected macro domains. Proteins are identified by name or UniProt accession number. Domain abbreviations: Brca1 C-terminus (BRCT). **B:** Multiple sequence alignment of all known human and *Dictyostelium* macro domains, including the findings of this study. The position of the D723 amino acid in human ALC1, which is vital for PAR-binding and protein function, is marked by an asterisk. Sequences are coloured by BLOSUM62 score: 0.5-1.5 – green, 1.5-3 – blue, 3< – red, where a higher number represents stronger conservation of an amino acid.

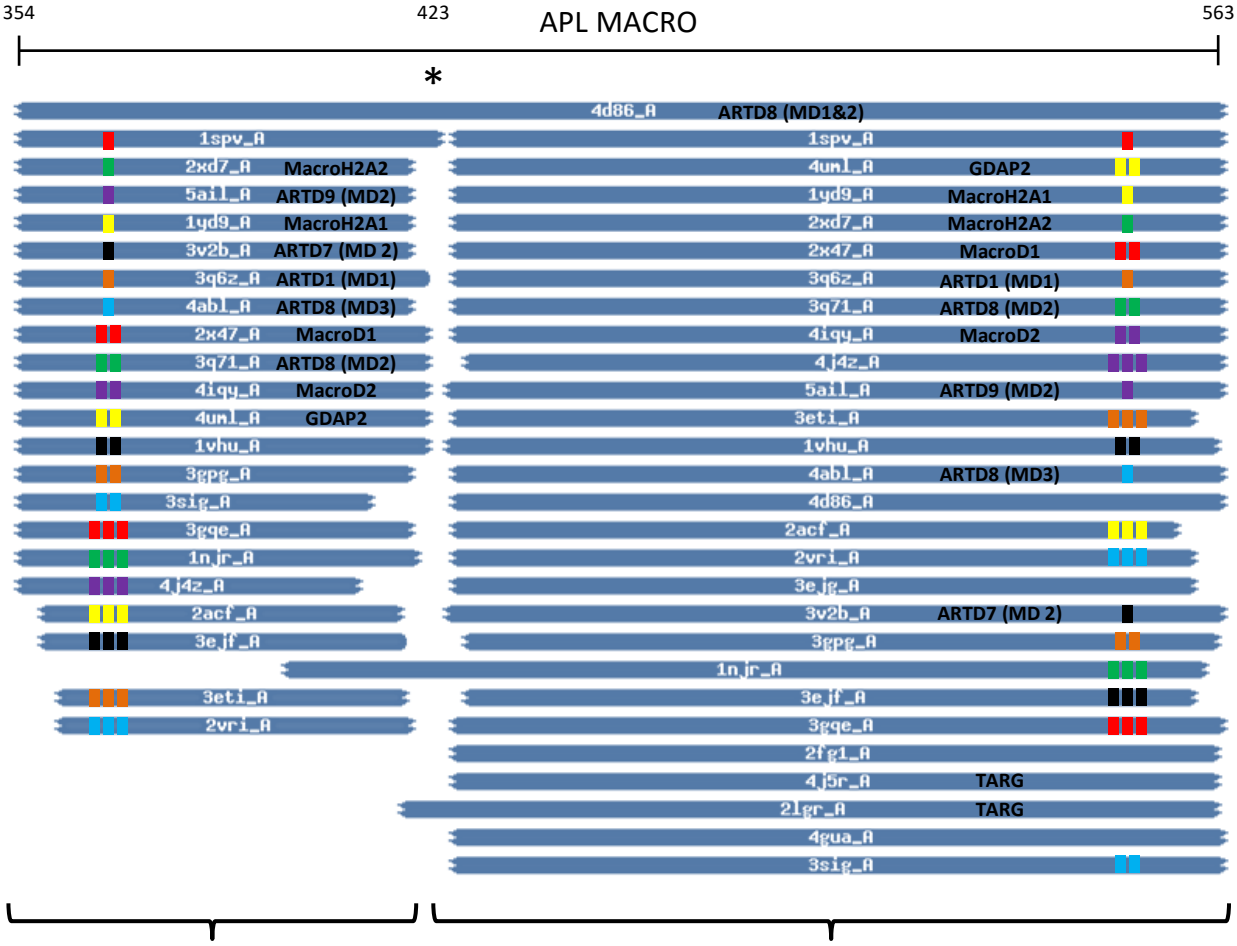
3.2.3. The macro domain of APL is circularly permuted

Following the identification of these novel putative PAR-binding domains, we subsequently employed the profile-profile homology comparison tool HHpred, which employs structural prediction in addition to HMM-HMM comparison. The sequence of the macro domain region of APL (aa: 354-563; Figure 3.8A) was used as the input to HHpred, and the statistically significant results that were returned were all from known macro domains. This further strengthens evidence that a macro domain does exist in APL. While this approach yielded no further novel macro-domain-containing proteins, it unveiled an interesting characteristic about the macro domain identified in APL: it is circularly permuted compared to the other macro domains found in human and *Dictyostelium* proteins. This was detected in the HHpred results as the sequences of the known macro domains did not align in a continuous manner with that of APL: the N-terminal and C-terminal ends of the domain were exchanged (Figure 3.8B). The only result for which this is not true is that with the smallest E-value, and corresponds to the structure of the first and second macro domains of ARTD8 (Figure 3.8C). In this alignment, the N-terminus of the macro domain in APL aligns with the C-terminus of the first macro domain of ARTD8, while the C-terminus of the macro domain of APL aligns with the adjacent N-terminus of the second macro domain of ARTD8. The macro domain of APL therefore aligns with the macro domains in ARTD8 in a continuous manner, but this alignment still represents a permuted macro domain. The HHpred result, and subsequent alignment of ARTD8 and the macro domain of APL (Figure 3.8D) indicate that the N-terminus of the macro domain of APL aligns with the C-terminus of the first macro domain of ARTD8, and the C-terminus aligns with the N-terminus of the second macro domain (with an insertion between them). Sequence alignments of the macro domain of APL with the individual macro domains of

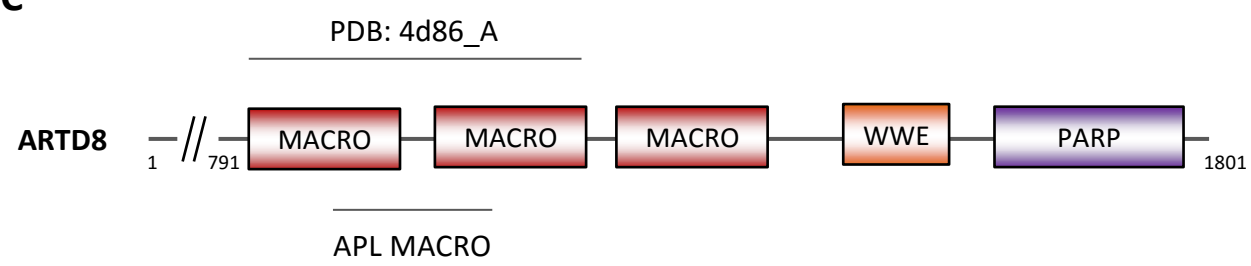
A



B



C



D

```
Dictyostelium APL  SLAFPFISTSTFGFNIDDATEISANAISEYLHFHEKE...DDIKLKMMVEKSI..YSDNLIQSFKKHFND..........
Human ARTD8        SIAIPAISSGVFGFPLGRCVETIVSAIKENFQFKKDGHCLKEIYLVDVSEKTVEAFAEAVKTVFKATLPDTAAPPGLPPA

Dictyostelium APL  .......KWDK.......RFEIIKIENSNSLEQFNLGCKLFATESTWRLKKTPQNKQLYEMLD...TGTFEKVTKNLYPN
Human ARTD8        AAGPGKTSWEKGSLVSPGGLQMLLVK..EGVQNAKTDVVVNSVPLDLVLSRGPLSKSLLEKAGPELQEELDTVGQGVAVS

Dictyostelium APL  CGKIGKVYPISLQ.................NNKQLVNSILHKEYGI.......DIVILVLGV.NMNPNK.........PDA
Human ARTD8        MGTVLKTSSWNLDCRYVLHVVAPEWRNGSTSSLKIMEDIIRECMEITESLSLKSIAFPAIGTGNLGFPKNIFAELIISEV

Dictyostelium APL  FKENSELAKPLLLETYHSLFNA........LDNF
Human ARTD8        FKFSSKNQLKTLQEVHFLLHPSDHENIQAFSDEF
```

Figure 3.8 – The macro domain of APL has undergone a circular permutation. A: Diagrammatic representation of the input to the HHpred server, used for profile-profile-HMM homology searching and secondary structure prediction. **B:** A representation of the results from HHpred. The black line at the top represents the sequence of the macro domain of APL, and the significant matches to it are shown as blue boxes below. The matches are fragments of known PDB structures (white text), which align with the macro domain of APL. Those structures representing human macro domains are also named (black text). All of the significant matches to the macro domain of APL are structures of proteins containing macro domains. Interestingly, aside from the top result (which is the structure of the first and second macro domains of human ARTD8), the rest of the aligned macro domains have been split in two, with the N- and C-terminal ends of the structures swapped (pairs are indicated by coloured bars). This indicates a circular permutation in the macro domain of APL, at the point marked by an asterisk. Selected PDB IDs from human proteins: 4d86 – ARTD8 (macros 1 and 2), 4uml – GDAP2, 1yd9 – MacroH2A1, 2xd7 – MacroH2A2, 2x47 – MacroD1, 3q6z – ARTD8 (macro 1), 3q71 – ARTD8 (macro 2), 4abl – ARTD8 (macro 3), 4iqy- MacroD2, 3v2b – ARTD7 (macro 2), 5ail – ARTD9 (macro 2), 4j5r/2lgr – TARG. **C:** Domain structure of ARTD8, indicating the region of the protein represented by the structure 4d86, and the alignment of the macro domain of APL calculated by HHpred. This further shows that the macro domain is permuted. **D:** Alignment of the macro domain of APL and ARTD8, as shown in **C**, by T-Coffee. Amino acids are coloured by BLOSUM62 score, which in the case of two sequences colours identical amino acids red, and those of a similar type green.

ARTD8 indicate that it shares the most sequence identity with the second macro domain, but it is less similar to that than to the hybrid of the first and second macro domains.

These data indicate that the macro domain of APL has undergone a circular permutation. In such a permutation, the subdomains of a protein or domain are rearranged in a circular manner relative to the other members of its family (Figure 3.9A). Such events are thought to arise through duplication events or by the differential fission of subdomains over the course of evolution (Figure 3.9B)[401,402]. In the case of the macro domain of APL, there are two subdomains, and the effect of the permutation is to swap the position of the two subdomains (Figure 3.9C). Following this permutation, the N- and C-termini of the non-permuted macro domain become the centre of the domain, and the centre of the non-permuted domain becomes the new N- and C-termini (labelled N' and C', respectively). We wished to identify if permuted macro domains could be found in other proteins, as these would be more difficult to detect by the methods applied so far. A BLAST search with the sequence of the permuted macro domain of APL reveals some protein sequences with high levels of homology to it (Figure 3.9D). Aside from APL homologs in the closely related *Dictyostelium purpureum*, *Dictyostelium fasciculatum* and *Polysphondylium pallidum*, the permuted macro domain is also identified in a protein in *Arabidopsis thanalia* (Unipot accession: Q9M041). This has been identified as a transcription factor by homology, but does not share any other domains in common with APL.

3.2.4. Structural prediction of the macro domain of APL

The gross sequence rearrangement associated with a circular permutation could have significant impact on the structure of the domain of APL, and therefore would also affect and potentially disrupt its function. In order to determine whether or not the macro domain of APL has retained its structure, we attempted to predict the structure of the domain. We

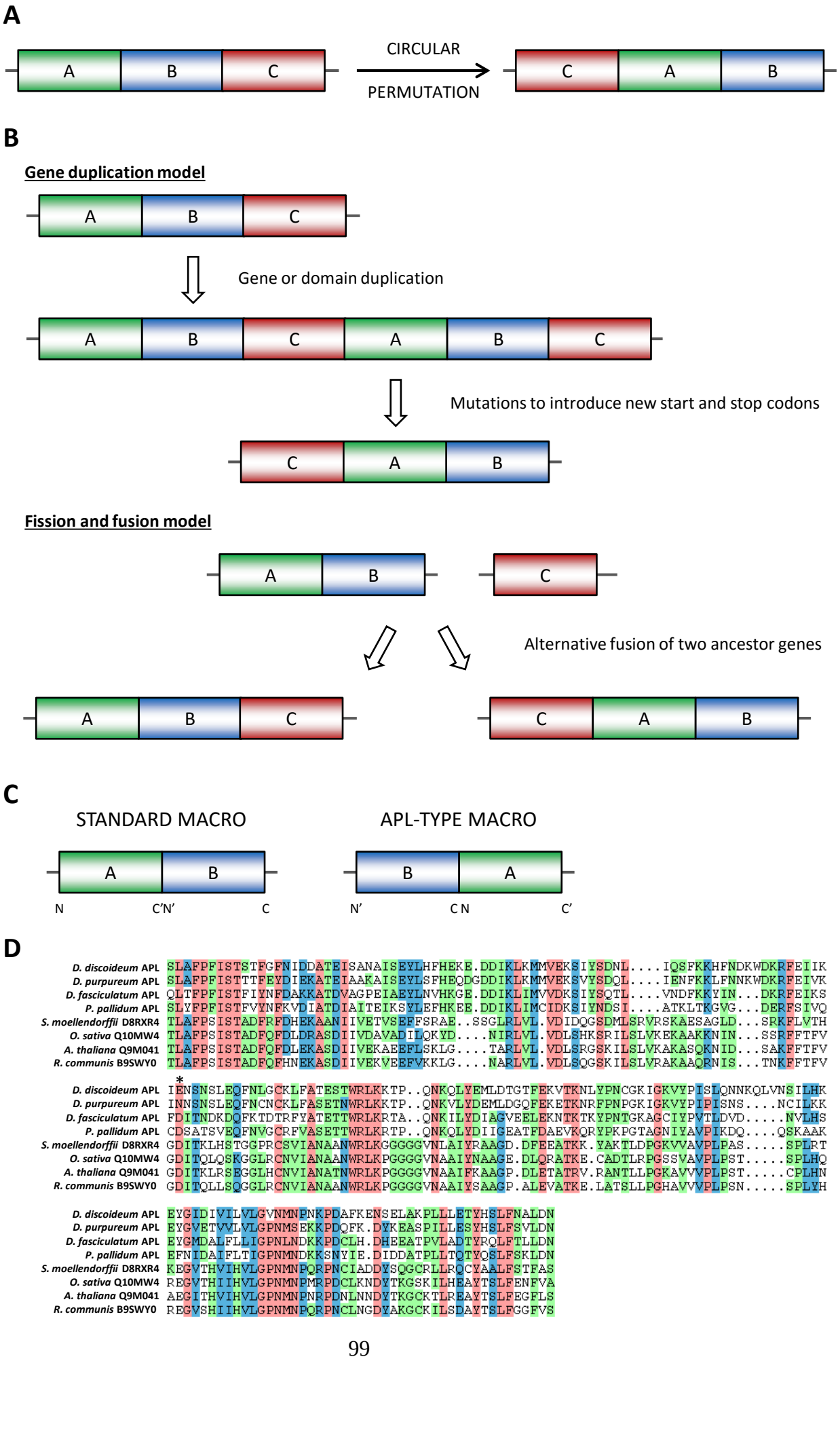
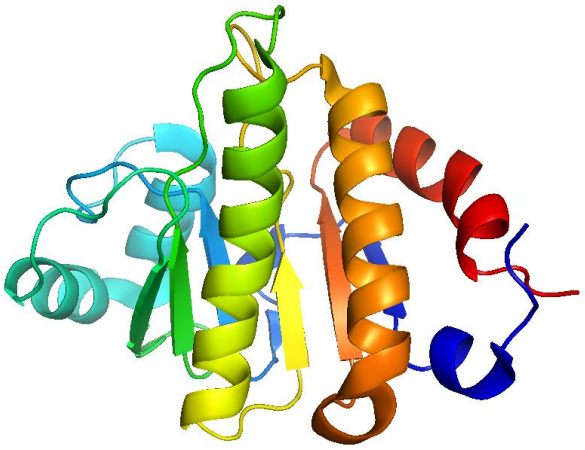

A
A
B
C
CIRCULAR
PERMUTATION
C
A
B
B
Gene duplication model
A
B
C
Gene or domain duplication
A
B
C
A
B
C
Mutations to introduce new start and stop codons
C
A
B
Fission and fusion model
A
B
C
Alternative fusion of two ancestor genes
A
B
C
C
A
B
C
STANDARD MACRO
APL-TYPE MACRO
A
B
B
A
N
C'N'
C
N'
C N
C'
D
D. discoideum APL SLAFPFISTSTFGFNIDDATEISANAISEYLHFHEKE.DDIKLKMMVEKSIYSDNL....IQSFKKHFNDKWDKRFEIIK
D. purpureum APL SLAFPFISTTTFEYDIEKATEIAAKAISEYLSFHEQDGDDIKLKMMVEKSVYSDQL....IENFKKLFNNKWDKRFEIVK
D. fasciculatum APL QLTFPFISTFIYNFDAKKATDVAGPEIAEYLNVHKGE.DDIKLIMVVDKSIYSQTL....VNDFKKYIN...DKRFEIKS
P. pallidum APL SLYFPFISTFVYNFKVDIATDIAITEIKSYLEFHKEE.DDIKLIMCIDKSIYNDSI....ATKLTKGVG...DERFSIVQ
S. moellendorffii D8RXR4 TLAFPSISTADFRFDHEKAANIIVETVSEFFSRAE..SSGLRLVL.VDIDQGSDMLSRVRSKAESAGLD...SRKFLVTH
O. sativa Q10MW4 TLAFPSISTADFQFDLDRASDIIVDAVADILQKYD....NIRLVL.VDLSHKSRILSLVKEKAAKKNIN...SSRFFTFV
A. thaliana Q9M041 TLAFPSISTADFQFDLEKASDIIVEKAEEFLSKLG....TARLVL.VDLSRGSKILSLVKAKASQKNID...SAKFFTFV
R. communis B9SWY0 TLAFPSISTADFQFHNEKASDIIVEKVEEFVKKLG....NARLVL.VDLSQGSKILSLVRAKAAQRNIS...TNKFFTFV
*
D. discoideum APL IENSNSLEQFNLGCKLFATESTWRLKKTP..QNKQLYEMLDTGTFEKVTKNLYPNCGKIGKVYPISLQNNKQLVNSILHK
D. purpureum APL INNSNSLEQFNCNCKLFASETNWRLKKTP..QNKVLYDEMLDGQFEKETKNRFPNPGKIGKVYPIPISNS....NCILKK
D. fasciculatum APL FDITNDKDQFKTDTRFYATETTWRLKRTA..QNKILYDIAGVEELEKNTKTKYPNTGKAGCIYPVTLDVD.....NVLHS
P. pallidum APL CDSATSVEQFNVGCRFVASETTWRLKRTP..QNKQLYDIIGEATFDAEVKQRYPKPGTAGNIYAVPIKDQ....QSKAAK
S. moellendorffii D8RXR4 GDITKLHSTGGPRCSVIANAANWRLKGGGGGVNLAIYRAAGD.DFEEATKK.YAKTLDPGKVVAVPLPAS.....SPLRT
O. sativa Q10MW4 GDITQLQSKGGLRCNVIANAANWRLKPGGGGVNAAIYNAAGE.DLQRATKE.CADTLRPGSSVAVPLPST.....SPLHQ
A. thaliana Q9M041 GDITKLRSEGGLHCNVIANATNWRLKPGGGGVNAAIFKAAGP.DLETATRV.RANTLLPGKAVVVPLPST.....CPLHN
R. communis B9SWY0 GDITQLLSQGGLRCNVIANAANWRLKPGGGGVNAAIYSAAGP.ALEVATKE.LATSLLPGHAVVVPLPSN.....SPLYH
D. discoideum APL EYGIDIVILVLGVNMNPNKPDAFKENSELAKPLLLETYHSLFNALDN
D. purpureum APL EYGVETVVLVLGPNMSEKKPDQFK.DYKEASPILLESYHSLFSVLDN
D. fasciculatum APL EYGMDALFLLIGPNLNDKKPDCLH.DHEEATPVLADTYRQLFTLLDN
P. pallidum APL EFNIDAIFLTIGPNMNDKKSNYIE.DIDDATPLLTQTYQSLFSKLDN
S. moellendorffii D8RXR4 KEGVTHVIHVLGPNMNPQRPNCIADDYSQGCRLLRQCVAALFSTFAS
O. sativa Q10MW4 REGVTHIIHVLGPNMNPMRPDCLKNDYTKGSKILHEAYTSLFENFVA
A. thaliana Q9M041 AEGITHVIHVLGPNMNPNRPDNLNNDYTKGCKTLREAYTSLFEGFLS
R. communis B9SWY0 REGVSHIIHVLGPNMNPQRPNCLNGDYAKGCKILSDAYTSLFGGFVS

Figure 3.9 – Circularly-permuted macro domains exist in other species. A: An illustration of a circular permutation for a protein (or domain) with three sub-domains. **B:** Diagrammatical representation of the gene duplication and fission and fusion models of circular permutation. **C:** An illustration of the circular permutation that has occurred in the macro domain of APL compared to classical macro domains. **D:** The results of a BLAST search with the circularly permuted macro domain of APL, indicating that circularly permuted macro domains are observed in proteins in other species. Alongside the orthologs of APL in other closely related slime moulds: *D. purpureum* (F0Z946), *D. fasciculatum* (F4QDB1) and *P. pallidum* (D3BHP0), a permuted macro domain is identified in a putative transcription factor in *Arabidopsis thanalia* (Q9M041). Proteins are identified by their UniProt accession numbers, and coloured according to amino acid conservation (BLOSUM62 score): : 0.5-1.5 – green, 1.5-3 – blue, 3< – red, where a higher number represents stronger conservation of an amino acid. Conservation of amino acids in these pileups can be used to identify functionally important residues, such as the one marked with an asterisk here (E439 in *Dictyostelium discoideum* APL).

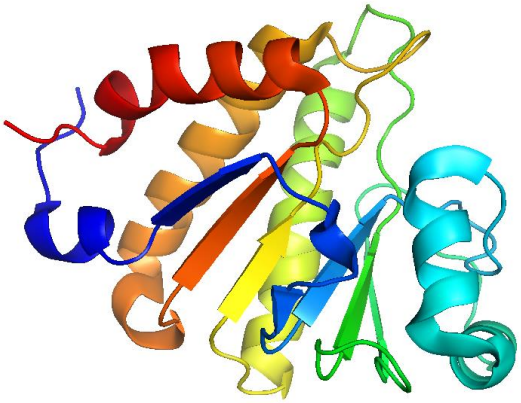
employed the technique of homology modelling for this purpose, which models domains from a template structure of a solved domain, and a multiple sequence alignment of the target and query domains[403]. The structures of several human macro domains have been solved; however, that of the PAR-binding macro domain of ALC1 has not. The structure of this macro domain would be most suitable to complement the *in silico* search we have performed. In the absence of this structure, we chose the structure of the human protein MacroD1 (PDB ID: 2x47; Figure 3.10A), as it displays the highest sequence similarity to that of APL, once it has been manually unpermuted. The structure of MacroD1 was modified in two ways to make it suitable for modelling the permuted macro domain of APL. Firstly, a break in the protein backbone at the approximate location of the permutation site was made, and secondly, the chain numbers were reordered so as to mimic those of a permuted macro domain. The judgement of location of the permutation site was informed by the alignment shown in Figure 3.8B, of which MacroD1 is a member, and was also chosen so as to avoid the secondary structure elements in the solved structure of MacroD1. These structure modifications were made to assist the modelling algorithms, and should not affect the convergence of the modelling algorithm. A multiple sequence alignment of the sequence of the macro domain of MacroD1 and the permuted macro domain of APL was made, but displayed only 18% sequence identity due to the diversity of the macro domain superfamily. Such a low level of sequence identify is generally unsuitable for homology modelling; however, we continued as the model produced would serve only as a guide[404]. The result of the homology modelling is shown in Figure 3.10B, and this displays a high degree of secondary and tertiary structure similarity with the macro domain of MacroD1, shown by an overlay of the predicted and template structures in Figure 3.10C. Additionally, the binding pocket of the macro domain of MacroD1 is present in the predicted structure of the macro domain of APL (Figure 3.11A). Therefore,

A

MacroD1 - Front

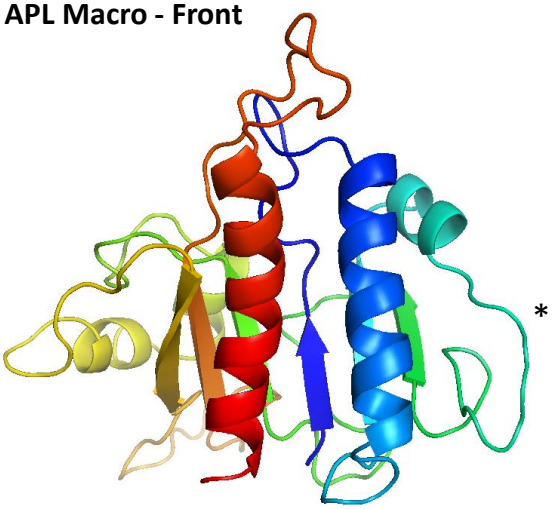


MacroD1 - Back

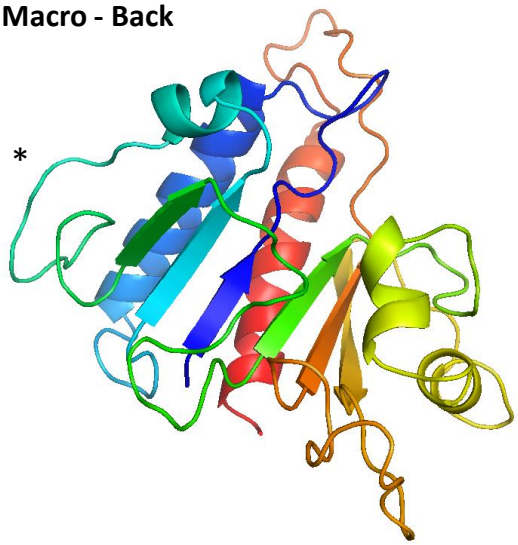


B

APL Macro - Front

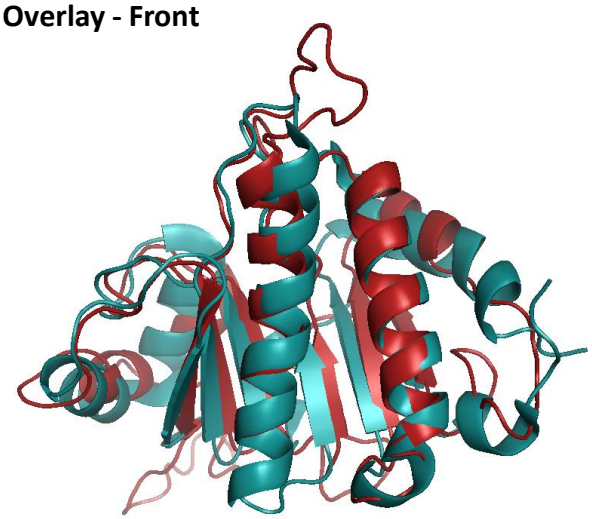


APL Macro - Back

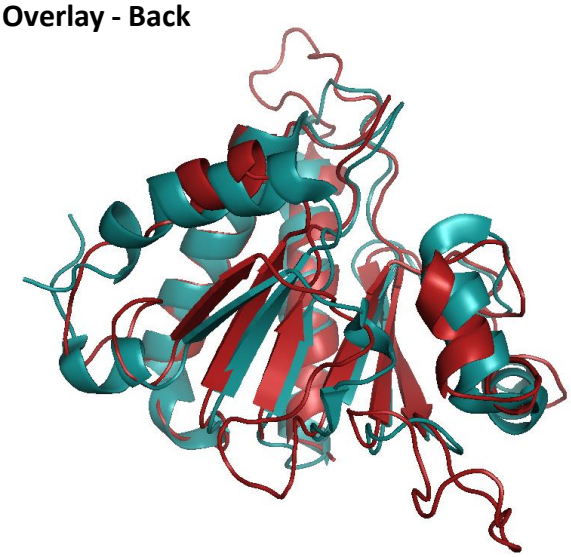


C

Overlay - Front



Overlay - Back



- MacroD1

- APL Macro

Figure 3.10 – Structural prediction of the permuted macro domain of APL by homology modelling suggests that the domain has retained the structure of a classical macro domain. A: Arbitrarily chosen front and back views of the solved structure of MacroD1 (PDB: 2x47), which was chosen to be the template for the homology modelling of the macro domain of APL. The secondary structure of the protein is displayed, with rainbow colouring from N- to C-terminus (blue to red). **B:** Front and back views of the predicted structure of the macro domain of APL (APL Macro), generated by homology modelling. The secondary structure of the predicted macro domain is shown, rainbow coloured from N- to C-terminus (blue to red). The change of location of N- and C-termini is clearly apparent, as is the loop predicted over the N- and C-termini of MacroD1 (marked by an asterisk).The structures of MacroD1 in **A** and of APL Macro were aligned in Pymol and so each view is taken from the same place for both structures. **C:** An overlay of the template structure of MacroD1 (turquoise) and the predicted structure of APL Macro (red), showing secondary structure.

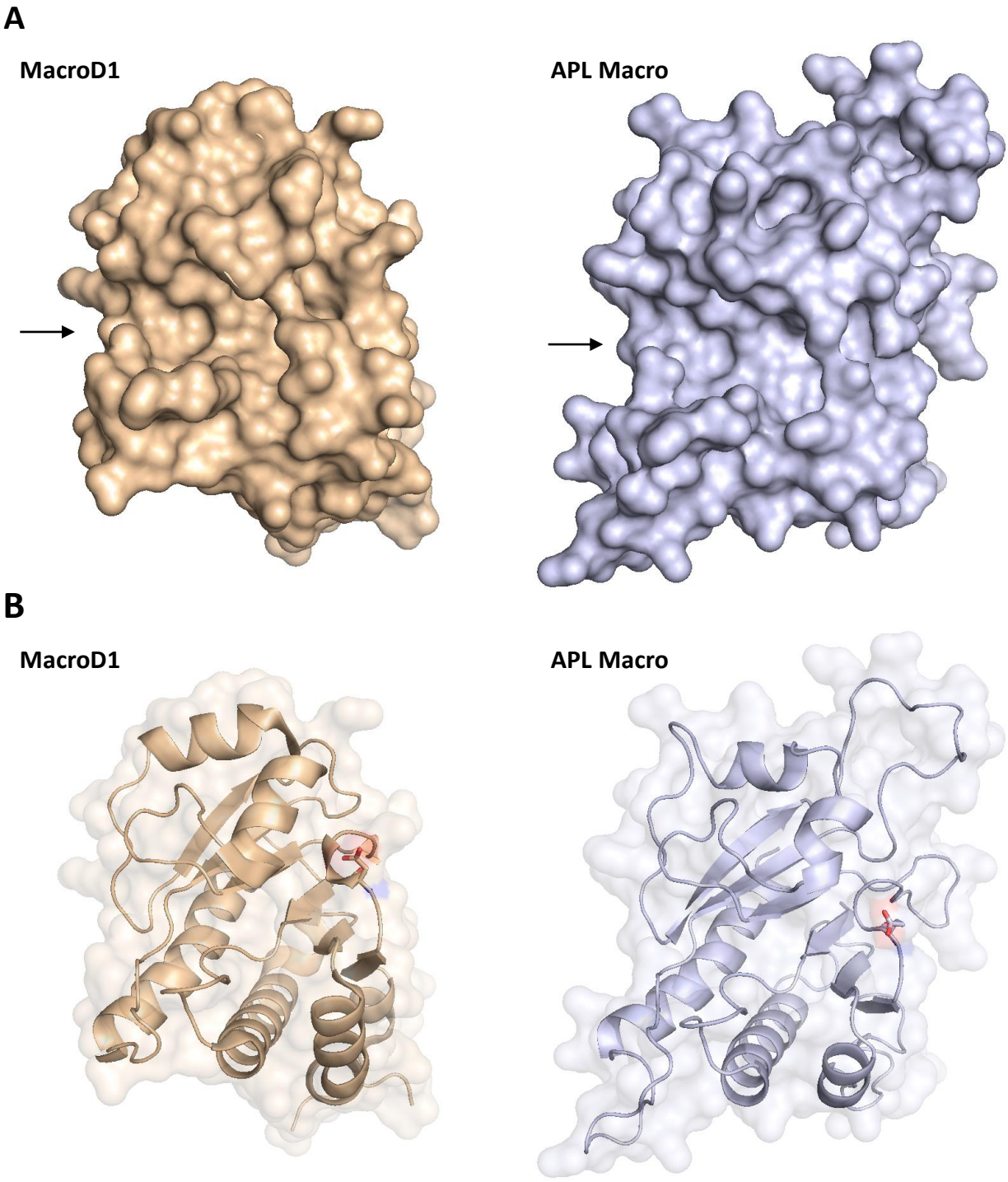

A
MacroD1
APL Macro
B
MacroD1
APL Macro
C
- MacroD1
- APL Macro

Figure 3.11 – Predicted conservation of the binding pocket of the permuted macro domain of APL. A: Surface visualisation of the solved structure of the macro domain of MacroD1 (left; orange) and the predicted structure of the macro domain of APL (right; purple). The binding pocket is shown by an arrow in both structures. **B:** The same surface visualisations as **A**, but with secondary structure superimposed. The side chain of the catalytically important D160 amino acid in MacroD1 is shown on the right-hand side of the binding pocket. The side chain of the identified E439 amino acid in the macro domain of APL is also shown, and is located with a similar position and orientation to D160. **C:** An overlay of the known structure of MacroD1 and the predicted structure of the macro domain of APL, focusing on the region of the D160 and E439 amino acids of MacroD1 and APL, respectively.

this predicted structure suggests that the macro domain of APL may be functional, and still resembles the rest of the macro domain superfamily, despite its permutation. The obvious caveat of homology modelling is that, if the modelling converges to a structure, that structure will be highly similar to the template. However, if the sequence inputted is not suitable then there will be no convergence at all, and the result will be a string of amino acids with no secondary structure elements. The convergence of the structure of the permuted macro domain of APL is therefore encouraging.

In addition to predicting the structure of the macro domain of APL, we wanted to use this structure to inform future experiments. The circular permutation, coupled with the diversity of the macro domain superfamily, makes it difficult to form and draw conclusions from multiple sequence alignments, and therefore the prediction of catalytic or key amino acids must be done by other methods. Firstly, the alignment of permuted macro domains (Figure 3.9D) gives some indication of functionally important regions of the macro domain of APL, as these regions will be more resistant to the accumulation of mutations over the course of evolution. Secondly, analysis of the amino acids around the binding pocket of the predicted structure is also highly informative. In ALC1, the mutation of the aspartic acid at position 723 (D723) to alanine abrogates the PAR-binding ability of its macro domain[250]. The position of this amino acid is marked in Figure 3.7B; however, the aspartic acid aligned with this in the macro domain of APL is not proximal to the predicted binding pocket. The equivalent amino acid in the structure of MacroD1, D160, is located in the binding pocket of the macro domain of this protein (Figure 3.11B). The mutation of this amino acid to alanine reduces the catalytic activity of the domain in deacetylating *O*-acetyl-ADP-ribose[377]. Comparison of the region of the structure of MacroD1 containing D160 with the macro domain of APL reveals a glutamic acid (E439) in a similar position and orientation, which may perform the same role in PAR-binding as

D723 in ALC1 (Figure 3.11B-C). Moreover, the glutamic acid at position 439 in APL is a conserved acidic residue across the permuted macro domains (Figure 3.9D), further implicating it as a functional amino acid.

3.3. Discussion

In this chapter, we have identified three additional *Dictyostelium* proteins with putative PAR-binding domains: Lig3, APL and Q54R54 (and its human homolog Q9NQ89). These proteins have all been found to contain macro domains. Additionally, the macro domain of APL was found to have undergone a circular permutation with respect to the other identified human and *Dictyostelium* macro domains. However, prediction of the structure of this permuted macro domain suggested that it has retained its structure and may still be functional. The search for proteins containing previously unannotated PBZ domains did not build on the already known list of proteins[241]. Therefore, this study has reinforced the suggestion that *Dictyostelium* DNA repair has a stronger reliance on PBZ domains compared to humans, but there are still substantially more human proteins containing macro domains.

Following this *in silico* identification, the next step of the analysis is to assess whether or not these proteins are involved in the DDR, and to determine the role of PARylation in their function. It is impractical to study all of the proteins simultaneously, and therefore we need to identify which of the candidate proteins are more likely to be involved in the DDR from the information available in protein databases. The domain structures of the three *Dictyostelium* proteins with previously unannotated macro domains are shown in Figure 3.7A. *Dictyostelium* Lig3 shares similar domain structure with human Lig3. Aside from the presence of a macro domain in *Dictyostelium* Lig3, there are two other notable differences: *Dictyostelium* Lig3 lacks the N-terminal PARP-type zinc finger present in

human Lig3, and has two BRCT domains instead of one[68]. Aside from its C-terminal macro domain, APL has an FHA domain and a PBZ-domain, the latter of which is predicted to bind to PAR and was identified prior to this study[241]. The FHA domain is a phospho-specific protein-protein interaction motif that is found in many DNA repair proteins[405]. As previously introduced, the FHA domain of APL displays sequence similarity to the FHA domains found in human APTX, PNKP and APLF, which are proteins involved in the human DDR (Figure 3.12)[15,74,406]. However, no human protein has the same domain structure as APL. Furthermore, no annotated human protein contains both a PBZ and a macro domain. Finally, despite extensive homology searching, the *Dictyostelium* protein Q54R54 and its human homolog Q9NQ89 do not contain any other characterised domains aside from their macro domains. While the presence of an uncharacterised human ortholog makes Q54R54 an interesting candidate, macro domains do not all bind to PAR and their presence certainly does not implicate a protein in the DDR. Q54R54 could therefore be involved in a wide range of cellular processes, and in the absence of known domains that may implicate the protein in DNA repair, we decides that Q54R54 would be an impractical choice for study in this thesis. However, the role of the macro domain and the overall function of the protein are certainly worth further investigation in either organism. Therefore, from initial observations of the domain structures of the candidate proteins, we have two proteins that we will investigate experimentally: Lig3 and APL.

Lig3 is already implicated in the DDR in humans, but the variation of the domain structure of *Dictyostelium* Lig3 could indicate an alternative mechanism of recruitment to SSBs. The traditional model of human Lig3 recruitment to SSBs is achieved through its interaction with XRCC1, which is itself recruited to the PAR scaffold synthesised by ARTD1[34,68]. Recently, however, it has been suggested that Lig3 is able to detect and bind

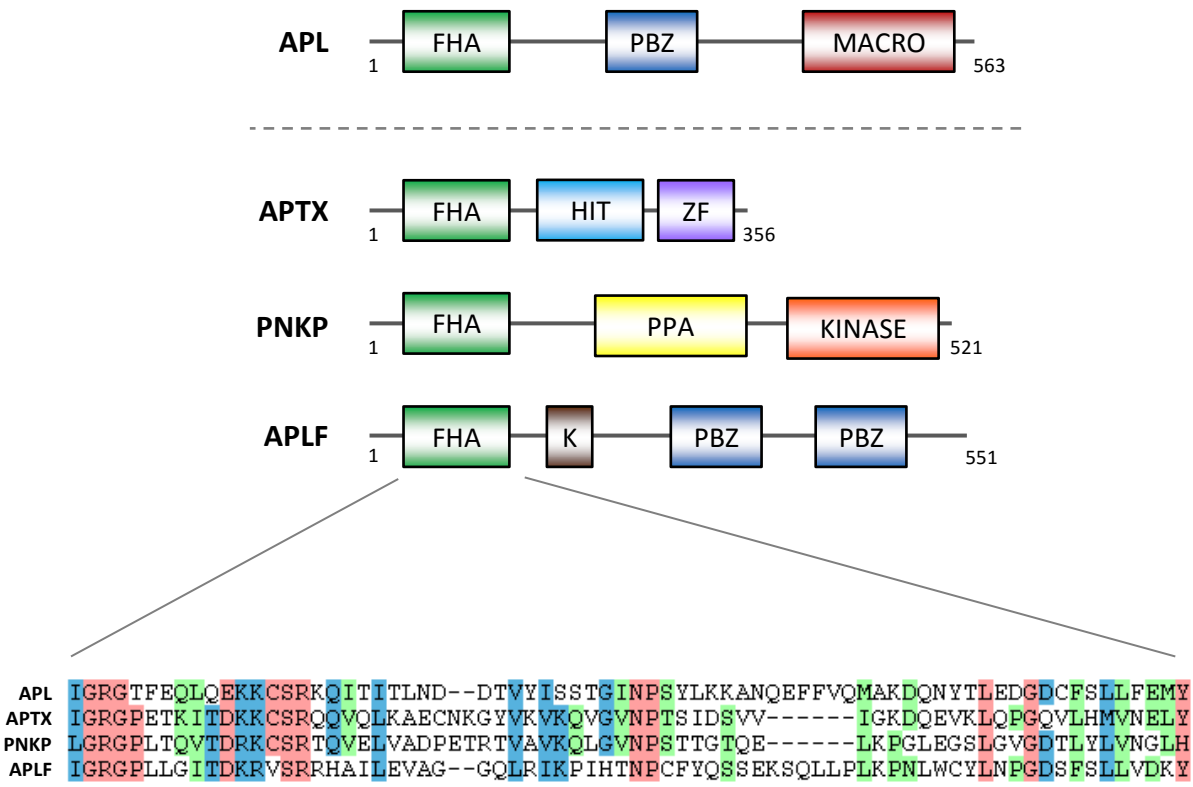


Figure 3.12 – The FHA domain of APL is found in three human DDR proteins. Performing a BLAST search with the FHA domain of APL as the input yields three statistically significant matches to human proteins: DNA end-processing factors APTX, PNKP and APLF. The domain structure of these proteins is shown, along with an alignment of their FHA domains (below). The amino acids are coloured by BLOSUM62 score: 0.5-1.5 – green, 1.5-3 – blue, 3< – red, where a higher number represents stronger conservation of an amino acid. Domain abbreviations: histidine triad (HIT), C2H2-type zinc finger (ZF), phosphatase (PPA), Ku binding motif (K).

to DNA nicks directly via its PARP-type zinc finger. This interaction is thought to be XRCC1- and PAR-independent[86]. As *Dictyostelium* Lig3 has no PARP-type zinc finger, it would appear that this mechanism of direct recruitment to DNA breaks may be absent. In this regard, the presence of a macro domain may indicate a stronger dependence of *Dictyostelium* SSBR, and potentially NER, on PARylation (if the macro domain is found to bind to PAR). The macro domain may also augment the function of *Dictyostelium* Lig3. This macro domain bears the strongest homology to that of human ALC1, a factor not yet identified in the *Dictyostelium* genome. In a manner analogous to the domain switching of the PBZ domain of human Ku70 and APLF, the macro domain of *Dictyostelium* Lig3 may serve the same PAR-binding function as that of ALC1, or may recruit and stabilise additional proteins not yet implicated in DNA repair[250]. As no human DNA ligase has been found to contain a macro domain, the role of the macro domain of *Dictyostelium* Lig3 could provide useful mechanistic insights into the DDR.

APL is a *Dictyostelium* protein with novel domain structure and no clear human homolog, making it an attractive candidate for a DNA repair protein in *Dictyostelium*. The presence of an FHA domain at the N-terminus of APL, which displays homology to the FHA domains found in the human DNA repair proteins APTX, PNKP and APLF, provides further evidence that it may function in the DDR. As FHA domains mediate phospho-specific protein-protein interactions, the FHA domain of APL may bind to the same DDR factors as the human proteins containing the same domain. Such interactions would strongly implicate APL in the DDR. Furthermore, APL additionally contains a PBZ domain, which is predicted to bind PAR[241]. Therefore, APL has two potential PAR-binding domains in its PBZ and macro domains. In APLF, tandem PBZ domains act to increase the affinity of the protein to PAR chains, and this could also happen with these two domains in APL[268]. Such cooperativity is conceivable due to the complementary,

differential binding of PBZ and macro domains to PAR chains: PBZ domains can bind to the junction between ADP-ribose moieties, whereas macro domains bind to the terminal ADP-ribose unit[240,264]. This would be the first example of two different PAR-binding domains acting cooperatively. Alternatively, the macro domain may bind to mono-(ADP-ribosyl)ated proteins, or possess catalytic activity to break down PAR. The presence of a PAR-binding PBZ domain and either a MAR-binding or catalytically active macro domain would suggest an interesting mechanism of action of the protein, and potentially dual-functionality. The behaviour of the individual domains of APL will be addressed in later chapters.

Interestingly, the permuted macro domain of APL shows strongest homology to the C-terminus of the first and N-terminus of the second macro domains of ARTD8, a human MART recently implicated in HR at stalled replication forks[35]. The second macro domain of ARTD8 has been shown to strongly bind ADP-ribose *in vitro*, and auto-modified ARTD10 *in vivo*, while the independent function of the first macro domain is less clear[407]. The alignment of the macro domain of APL with those of ARTD8 suggests that the macro domain of APL may be a MAR-binding domain. Furthermore, this could also implicate APL in the cellular response to replication-associated DNA damage. *Dictyostelium* has a putative homolog of ARTD8, previously referred to as pARTg. Similar to human ARTD8, pARTg contains a PARP catalytic domain and a WWE domain, but only one macro domain (human ARTD8 has three). As previously discussed for *Dictyostelium* Lig3 and Ku70, domain switching could have occurred between APL and pARTg, thereby transferring some of the functionality associated with human ARTD8 to a novel *Dictyostelium* protein, in this case APL. One possible explanation for the permutation of the macro domain of APL is that it has been caused by a duplication event, or the fusing of two adjacent non-permuted macro domains. The homology between the first and second

macro domain region of ARTD8 and APL and the lack of multiple macro domains in pARTg would be explained by this hypothesis.

We have identified two *Dictyostelium* proteins with previously unannotated macro domains, which we will proceed to experimentally characterise in the future chapters of this thesis. Lig3, a known DDR protein in humans, may display stronger reliance on PARylation in this organism, while APL is an intriguing protein with novel domain structure and several links to human DDR proteins. We will first characterise Lig3, before moving on to the novel protein APL.

4. The role of the macro domain of *Dictyostelium* DNA ligase III in DNA repair

4.1. Introduction

4.1.1. Human DNA ligases in DNA repair

In humans, three genes have been identified that encode DNA ligases: DNA ligase I (Lig1), DNA ligase III (Lig3) and DNA ligase IV (Lig4)[408]. However, alternative splicing of the gene encoding Lig3 leads to a nuclear and mitochondrial form of the protein, Lig3α and Lig3β, respectively[87]. The domain structure of the nuclear DNA ligases is shown in Figure 4.1. All of these DNA ligases have been implicated in the DNA damage response, where they act in the final step in many repair pathways.

Mammalian DNA Ligase I was first identified as the major contributor to the DNA ligation activity of calf thymus extracts, and is found to be induced in proliferating cells[409,410]. This latter observation implied a role for Lig1 in DNA replication, and Lig1 was subsequently identified to be recruited to replication forks via an interaction with PCNA[411,412]. Cells encoding a mutant form of Lig1 displayed defective Okazaki fragment joining, which was then shown to be dependent on the interaction of Lig1 and PNCA[413,414]. Despite strong evidence implicating Lig1 in DNA replication, experiments designed to assess the involvement of Lig1 in DNA repair have yielded mixed results. Early reports implicated mutant forms of Lig1 as the cause of Bloom's syndrome (BS)[415]. Cells from patients with BS are found to suffer increased levels of chromosomal rearrangements and elevated sensitivity to genotoxic agents. However, Lig1 was found to be normal in BS

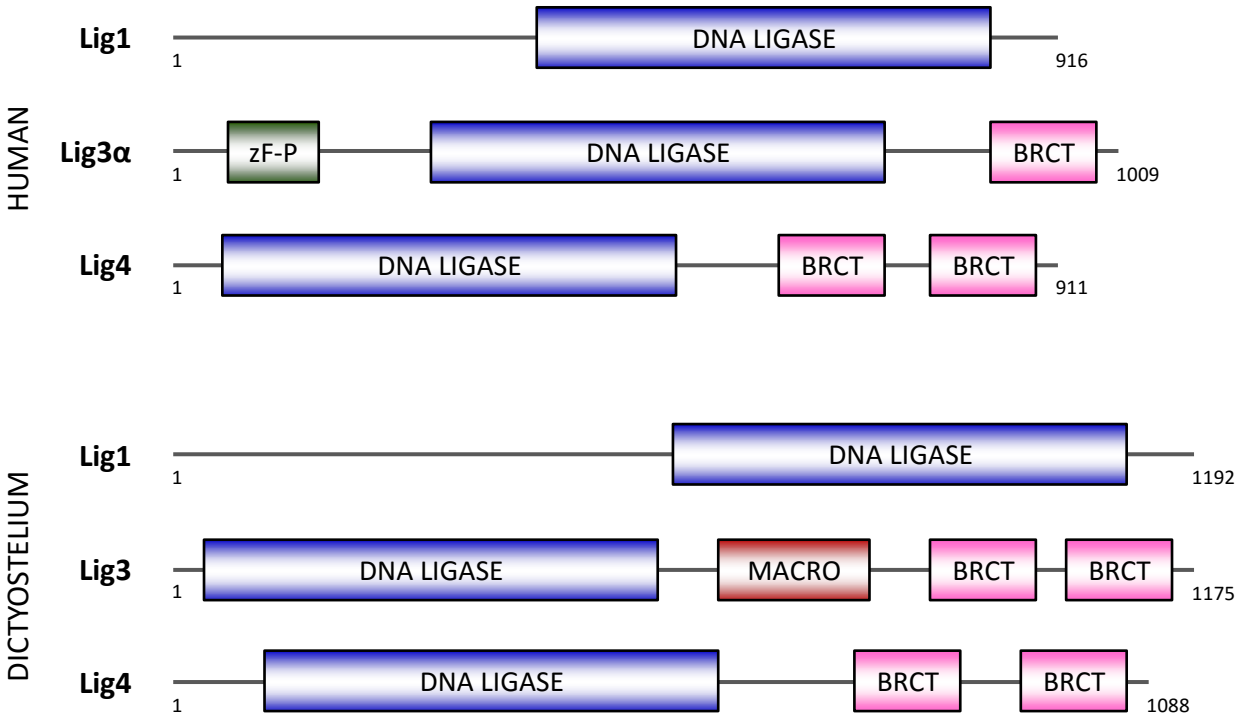


Figure 4.1 – Domain structures of the ATP-dependent nuclear ligases found in humans and *Dictyostelium*. The known domain structures of the nuclear DNA ligases in humans and *Dictyostelium*. In humans, the *lig3* gene encodes two isoforms of Lig3: Lig3α (nuclear) and Lig3β (mitochondrial). In addition to the three *Dictyostelium* DNA ligases shown here, *Dictyostelium* also encodes an NAD^+-dependent homolog of *E. coli* LigA. Domain abbreviations: PARP-type zinc finger (zF-P).

cells, although patients with a similar phenotype and Lig1 mutations have been observed[416–418]. Lig1 has been purified alongside Polβ in a BER-complex, and is thought to act in long-patch BER[419,420]. Lig1 has also been used to reconstitute the NER pathway *in vitro*[233]. Cells harbouring a mutation in Lig1, derived from a patient with immune-deficiency, displayed hypersensitivity to alkylating agents, UV irradiation, and IR[421]. However, more recent studies on the repair kinetics in these cells indicate a repair defect in response to UV light only[422]. Furthermore, Lig1 knock-out mice show deficiencies in DNA replication but not DNA repair[423]. These results indicate that there may be some redundancy or interplay between the mammalian DNA ligases.

DNA ligase 3 (Lig3) was discovered after Lig1, and orthologs have been identified in a smaller subset of species[424,425]. In humans, the *lig3* gene encodes two isoforms of Lig3: Lig3α and Lig3β, which localise to the nucleus and mitochondria, respectively[87,89,426]. Deletion of the *lig3* gene is lethal, but it has been demonstrated that this lethality stems from a requirement for Lig3β in mitochondria, and that the nuclear form of the protein is dispensable[89]. Lig3β is the only DNA ligase found in mitochondria, which is believed to be the reason for its necessity. Interestingly, depletion of both forms of Lig3 can be rescued by expression of the NAD^+-dependent *E.coli* LigA[88]. Lig3α is found in a heterodimer with XRCC1, via an interaction mediated by the BRCT domain of Lig3α[69,427,428]. In the widely accepted model of SSBR, the XRCC1-Lig3α complex is recruited to SSBs via an XRCC1-mediated interaction with auto-PARylated ARTD1, and Lig3α is implicated in performing the final ligation step in SSBR[34,429]. However, a direct interaction with PAR polymers and autoPARylated ARTD1 *in vitro* has been observed. Lig3α contains a PAR-binding motif (PBM) in the vicinity of its zinc finger motif (aa: 12-34); however, the binding to PAR and ARTD1 is independent of this region[238,244]. Furthermore, the zinc finger motif was identified as being similar to that found in ARTD1

(PARP-type), and has been shown to be a DNA nick sensor[68]. Recently, a PARP-independent model of Lig3α recruitment to DNA SSBs has been proposed, and is supported by data showing residual enrichment of Lig3α at SSBs in the absence of XRCC1[86]. In addition to functioning in SSBR, Lig3α has been implicated in alt-NHEJ, a DSBR pathway promoted by ARTD1 and normally supressed by Ku[189,430,431]. In this pathway, Lig3α is able to act independently of XRCC1[171]. Functioning in this pathway, Lig3α has been strongly implicated in mediating end-joining events at telomeres. However, despite its role in DNA repair, cells lacking Lig3α do not display enhanced sensitivity to SSB- or DSB-inducing agents, clearly indicating redundancy between the mammalian DNA ligases[88,89].

In addition to containing *lig1* and *lig3* genes, the human genome encodes for another DNA ligase, Lig4. The homology of the active site of Lig4 to Lig1 was utilised to identify and subsequently purify Lig4, but it was quickly discovered to possess different *in vitro* substrate specificity to Lig1 and Lig3α[425,432]. It was found that Lig4 associates with XRCC4, a scaffold protein functioning in NHEJ, thereby implicating it in the repair of DSBs via this pathway. The interaction of these two proteins is mediated by the tandem BRCT domains of Lig4[433,434]. This association is analogous to the association of Lig3α and XRCC1 in BER and NER[235,429]. Cells depleted or encoding mutant forms of Lig4 display sensitivity to ionising radiation and impaired V(D)J recombination, a process for which NHEJ is required[435–437]. The action of Lig4 in these pathways cannot be compensated for by either Lig1 or Lig3, and Lig4 knock-out mice are found to be lethal, highlighting the lack of redundancy of this DNA ligase[438,439].

4.1.2. *Dictyostelium* DNA ligases

The *Dictyostelium* genome contains genes for four DNA ligases. These have been identified by homology to be orthologs of eukaryotic DNA ligase I (Lig1), nuclear DNA ligase III (Lig3), DNA ligase IV (Lig4), and also a homolog of a bacterial NAD^+-dependent DNA Ligase (LigA)[321,440]. As previously introduced, *E. coli* LigA has been shown to complement the depletion of the essential human mitochondrial DNA ligase Lig3β[88]. Therefore, it would be reasonable to hypothesise that *Dictyostelium* LigA may be localised to and function in mitochondria; however, as yet there is no experimental evidence to support this. Making this assumption, this leaves three ATP-dependent, nuclear DNA ligases in *Dictyostelium*, in agreement with situation found in human cells (Figure 4.1). However, despite the recent study of *Dictyostelium* DNA repair, particularly NHEJ, the specific role of any of these DNA ligases in the DDR of this organism has not yet been established[441].

As is clear from the domain structure diagrams of the human and *Dictyostelium* DNA ligases (Figure 4.1), there is a notable difference in domain structure between human Lig3α and *Dictyostelium* Lig3, while the domains present in Lig1 and Lig4 are the same in both organisms. In addition to the shared ATP-dependent ligase domain, human Lig3α contains a BRCT domain and a PARP-type zinc finger. In contrast, the *Dictyostelium* ortholog contains two BRCT domains and the macro domain revealed by our *in silico* investigation. As Lig3 has not yet been characterised, the function of the BRCT domains has not yet been ascertained; however, it would be surprising if at least one of them did not mediate an interaction with XRCC1, as is the case in the human DDR.

4.1.3. Aims

The aim of this chapter is to identify the role of the newly discovered macro domain of *Dictyostelium* Lig3 in DNA repair. No human DNA ligases contain a macro domain, so the presence of one in *Dictyostelium* Lig3 might suggest a novel interaction of this protein to support its function in the DDR. The macro domain of Lig3 bears strongest homology to that of human ALC1. In ALC1, the macro domain is required for the recruitment of the protein to DNA lesions, and this is abrogated by the D723A mutation[250]. Therefore, one might hypothesise that the macro domain in Lig3 is responsible for its localisation to DNA lesions. There is no *Dictyostelium* protein currently annotated with the same domain structure of human ALC1, and the presence of the macro domain in Lig3 may indicate a further example of domain shuffling, previously observed with the PBZ domain of human APLF and *Dictyostelium* Ku70[360]. In this chapter, we will test this hypothesis, and also assess the role of Lig3 in *Dictyostelium* DNA repair in general.

4.2. Results

4.2.1. The macro domain of *Dictyostelium* Lig3 binds to PAR *in vitro*

We identified Lig3 as a putative PAR-binding protein through a novel macro domain in its central region. Homology searches with BLAST revealed that this macro domain is most similar to the PAR-binding macro domain found in human ALC1, a chromatin-remodelling helicase involved in DNA repair[250]. However, the PAR-binding nature of the macro domain in Lig3 is not known, as members of the macro domain superfamily bind to a variety of ligands[246]. Therefore, we first wanted to determine whether or not this macro domain does bind to PAR, and if it does, if we could identify any amino acids within the domain that can be mutated to abrogate this binding. As the macro domain is central to

Lig3, introducing point mutations is the most practical way of stopping the domain from functioning in future experiments.

We decided to investigate the PAR-binding ability of the macro domain of Lig3 in an *in vitro* assay, as this would allow us to focus solely on Lig3 and avoid the associated technical difficulties with investigating *in vivo* PARylation. To do this, we purified a GST-tagged form of the macro domain of Lig3 (Figure 4.2A). Primers were designed to amplify the macro domain of Lig3 from Ax2 cDNA by PCR, with suitable restriction sites added (BamHI and XhoI) so as to allow for ligation of the macro domain sequence into the pGEX-4T-2 vector, a bacterial expression vector for proteins with an N-terminal GST-tag. The resulting pGEX-4T-2-Lig3-MACRO vector (Figure 4.2B) was verified by Sanger sequencing and confirmed to be mutation-free. The pGEX-4T-2-Lig3-MACRO vector was transformed into *E. coli* BL21(DE3) cells by standard procedures, and expression of the protein was induced by addition of 0.1 mM IPTG to the nutrient broth. GST-tagged proteins were purified by standard GST purification procedures, and the purity of the extracts at each stage was monitored SDS-PAGE and Coomassie staining (Figure 4.2C). The predicted molecular weight of the macro domain and GST-tag is 45 kDa, matching the highlighted band in the eluted fractions, the concentration of which was subsequently quantified by comparison with BSA standards (data not shown).

We assessed the *in vitro* PAR-binding ability of the macro domain of Lig3 by dot-blotting increasing concentrations of GST-Lig3-MACRO onto a nitrocellulose membrane. Alongside the GST-Lig3-MACRO peptide, we also included GST alone as a negative control, and GST-Ku70-C, which is the C-terminus region of *Dictyostelium* Ku70. This region of Ku70 contains a PBZ domain known to bind to PAR chains, and therefore served as our positive control[360]. Following addition of the peptides, the membrane was incubated with synthesised PAR polymers, prior to washing and incubation with an anti-PAR

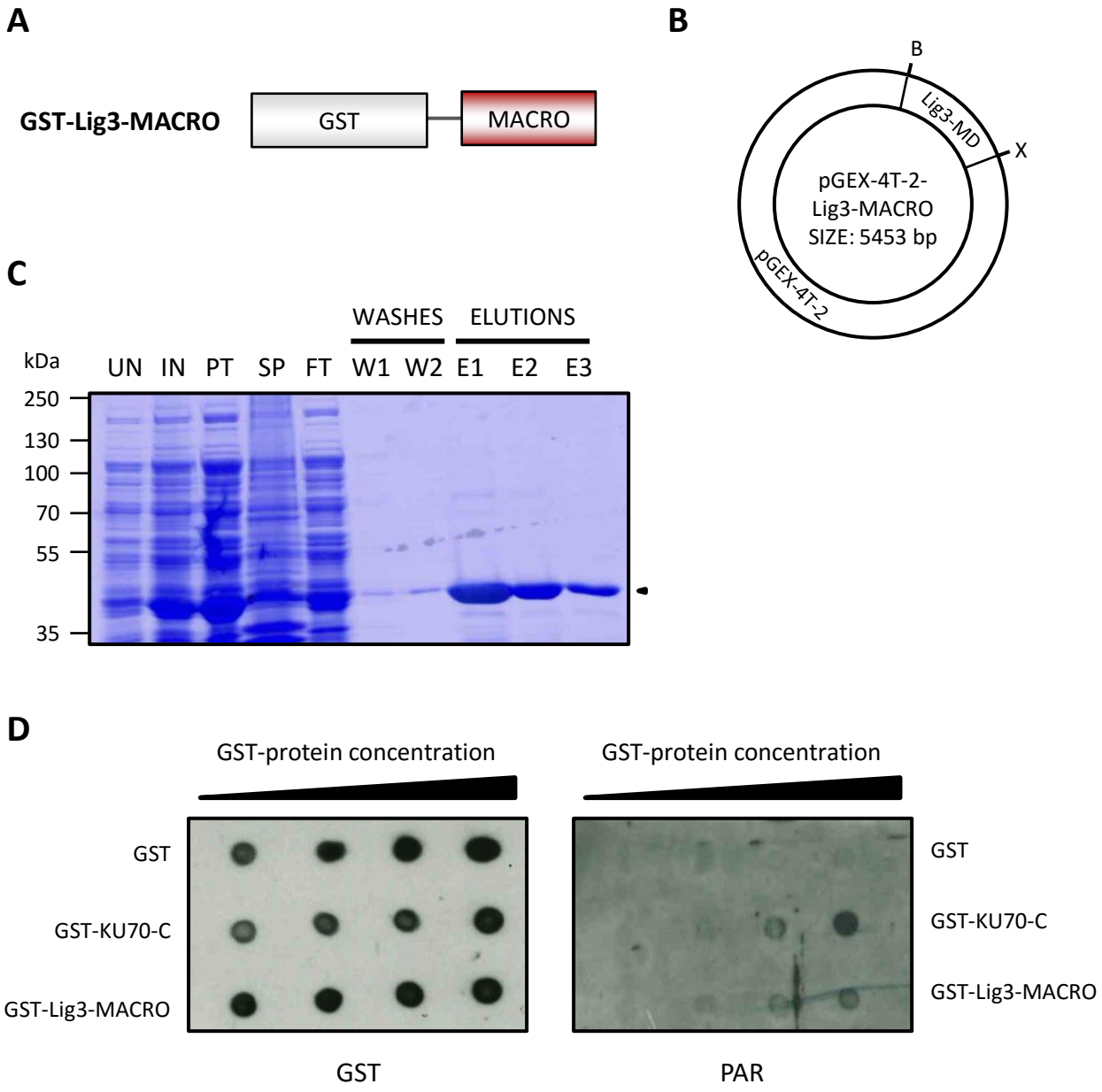


Figure 4.2 – The macro domain of *Dictyostelium* Lig3 binds PAR *in vitro*. A: Domain structure of the GST-tagged Lig3 macro domain construct. **B:** Plasmid map of the pGEX-4T-2-Lig3-MACRO vector, containing the macro domain sequence from Lig3 (Lig3-MD). Transformation of this vector into BL21(DE3) *E. coli* and subsequent IPTG-induction leads to expression of N-terminally GST-tagged macro domain of Lig3. Restriction sites used for cloning, BamHI (B) and XhoI (X), are shown. **C:** Purification of GST-Lig3-MACRO protein (44 kDa) by GST purification. Abbreviations: uninduced (UN), induced (IN), pellet (PT), supernatant (SP), flow-through (FT). **D:** *In vitro* assessment of the PAR-binding ability of the GST-Lig3-MACRO peptide. Increasing concentrations of GST and GST-Lig3-MACRO were dot-blotted onto nitrocellulose membrane, and incubated with PAR-polymers. The amount of bound PAR was assessed by Western blotting with an anti-PAR antibody. The C-terminus, PBZ domain-containing region of *Dictyostelium* Ku70 (GST-Ku70-C), was included as a positive control. Experiments in this figure were performed by Peggy Paschke.

antibody. This assay indicated that the macro domain of Lig3 does indeed bind to PAR, albeit not as strongly as the PBZ domain of Ku70 (Figure 4.2D). This further verifies our *in silico* screen, and also the parallels between the macro domain Lig3 and that found in human ALC1.

4.2.2. Lig3 is enriched on chromatin in response to base methylation

While it is well known that DNA Ligase III is involved in DNA repair in humans, it has not yet been studied in *Dictyostelium*. Therefore, to characterise the function of the macro domain of *Dictyostelium* Lig3, we must first identify the role of Lig3 in the DDR of this organism. We hypothesised that the macro domain found in Lig3 performs a similar function to that of human ALC1, which is required for the recruitment of ALC1 to DNA lesions in a PAR-dependent manner[250]. We have so far shown that the macro domain of ALC1 binds to PAR *in vitro*, but we have not yet assessed its function *in vivo*. Human Lig3α has been found to function in the SSBR pathway[429]. Therefore, we decided to investigate the enrichment of Lig3 to chromatin following MMS treatment. If Lig3 is involved in SSBR, enrichment of this protein on chromatin following induced damage should be observed, despite potential redundancy with Lig1. In *Dictyostelium*, the macro domain may be required for the recruitment or stability of Lig3 at DNA lesions, and mutating it may negatively affect the kinetics of the enrichment.

As no antibody is currently raised against *Dictyostelium* Lig3, and there is insufficient sequence homology between the *Dictyostelium* and human proteins to use an antibody against the human protein, we chose to express an N-terminally Myc-tagged form of Lig3 in cells (shown in Figure 4.3A). To generate this, we amplified the Lig3 coding DNA sequence from Ax2 cDNA by PCR, with an attached Myc-tag and compatible restriction sites, and ligated this construct into an empty pDXA-3C vector. The map of this pDXA-

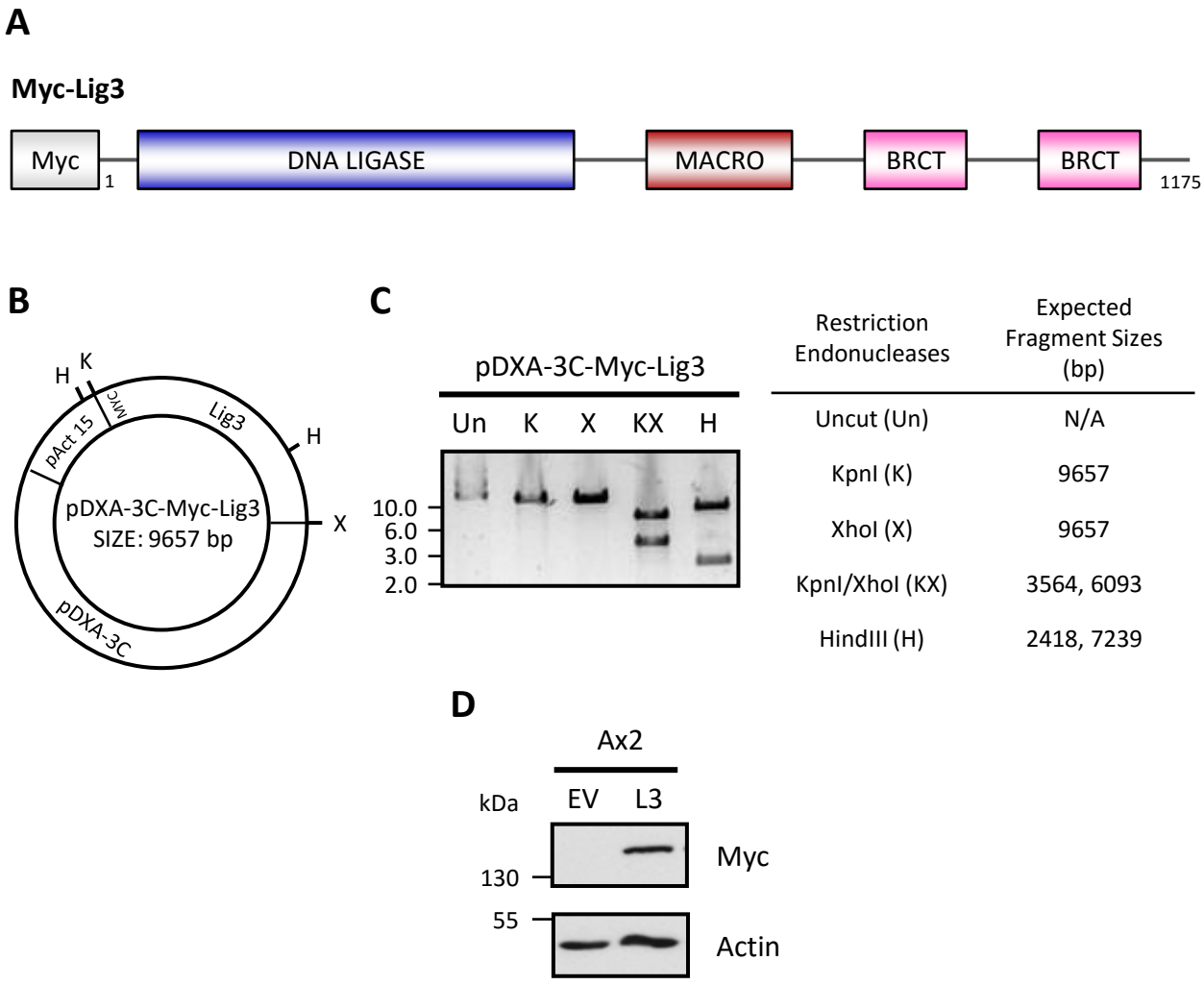


Figure 4.3 – Generation of a Myc-Lig3 *Dictyostelium* overexpression vector. A: Domain structure of the N-terminally Myc-tagged Lig3, which will be expressed in *Dictyostelium* cells. **B:** Map of the pDXA-3C-Myc-Lig3 *Dictyostelium* overexpression vector. The coding sequence for Lig3 was amplified from Ax2 cDNA by PCR, with primers to incorporate an N-terminal Myc-tag, and ligated into the empty pDXA-3C vector, using the KpnI (K) and XhoI (X) restriction sites. In this vector, Myc-Lig3 is placed under the transcriptional control of an Actin15 promotor (pAct15), and is highly expressed *in vivo*. Also indicated are the HindIII (H) restriction sites used for verification. **C:** Verification of the pDXA-3C-Myc-Lig3 vector by restriction digestion. Restriction enzymes used and expected fragments are shown in the table. **D:** Ax2 cells were transfected with empty pDXA-3C vector (EV) or pDXA-3C-Myc-Lig3 (L3) using standard procedures. Whole-cell extracts were then prepared and Myc-Lig3 expression was assessed by Western blotting with the indicated antibodies. The expected size of Myc-Lig3 is 133.2 kDa.

3C-Myc-Lig3 vector and restriction sites used are shown in Figure 4.3B. The pDXA-3C vector places the transcription of Myc-Lig3 under the control of an Actin15 promoter, which is highly expressed, and also encodes a gene for resistance to the antibiotic G418, for selection purposes. The cDNA sequence of Lig3 was analysed by Sanger sequencing to ensure that it carried no point mutations, and the pDXA-3C-Myc-Lig3 vector was verified by restriction digestion (Figure 4.3C).

Once the pDXA-3C-Myc-Lig3 vector was completely verified, it was transfected into Ax2 *Dictyostelium* strains by standard procedures, and the expression level of Myc-Lig3 was verified by Western blotting (Figure 4.3D). We then exposed Myc-Lig3 expressing cells to 5 mM MMS for up to 90 minutes. After treatment, the chromatin from these cells was extracted and analysed by SDS-PAGE and Western blotting. The chromatin extracts indicated an enrichment of Myc-Lig3 on chromatin over the course of the experiment, which implicates *Dictyostelium* Lig3 in the BER/SSBR pathway (Figure 4.4).

4.2.3. Vegetative *lig3*$^{-}$ cells do not display a defect in DNA SSBR

We have now shown that *Dictyostelium* Lig3 is enriched on chromatin following the induction of base alkylation, implicating Lig3 in the repair of DNA SSBs in this organism. However, we have not yet assessed the necessity of Lig3 for the survival of induced DNA damage. Human Lig3α has been found to act in a redundant manner with Lig1 in both SSBR and alt-NHEJ, and its depletion does not lead to a survival defect following DNA damage[88,179,442,443]. However, the redundancy of the DNA ligases in *Dictyostelium* has not been investigated, and the macro domain may have brought additional functionality and necessity to Lig3.

To address this question, we generated a *Dictyostelium lig3*$^{-}$ strain. We devised a strategy of targeted homologous recombination to disrupt the lig3 gene after its first 49 nucleotides,

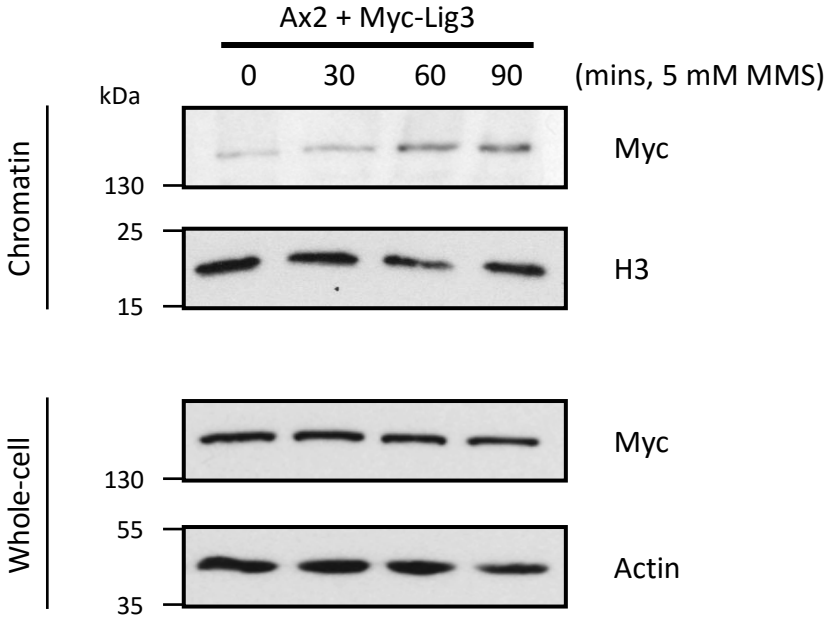


Figure 4.4 – Myc-Lig3 is enriched on chromatin following MMS treatment. *Dictyostelium* Ax2 cells expressing Myc-Lig3 were exposed to 5 mM MMS for the indicated durations. Following incubation, the chromatin fraction was extracted and protein levels were assessed by SDS-PAGE and Western blotting with the indicated antibodies. Chromatin and whole-cell loading are assessed by anti-H3 and anti-Actin antibodies, respectively.

thereby preventing transcription of the vast majority of the gene (Figure 4.5A). Each of the indicated homologous arms was amplified by PCR with the primers shown, and ligated into an empty pLPBLP vector with the indicated restriction sites. The pLPBLP vector contains a cassette encoding a gene for resistance to blasticidin, an antibiotic toxic to Ax2 cells. The two homologous arms were ligated either side of this blasticidin resistance (*bsr*) cassette, to form the pLPBLP-Lig3 disruption vector (Figure 4.5B). The arms were ligated in a reversed manner due to the availability of restriction sites at the *lig3* genomic locus, which does not affect the function of the cassette in any way. The homologous arms were analysed by Sanger sequencing to check for mutations, and the final pLPBLP-Lig3 vector was verified by restriction digestion to ensure it was correctly formed (Figure 4.5C). Prior to transfection into *Dictyostelium* Ax2 cells, the disruption construct was excised from the plasmid by BamHI/PvuI double digestion, and gel purified (Figure 4.5D). This strategy deletes the N-terminal ligase domain from the genome, but leaves the macro and BRCT domains of Lig3. However, these domains should not be transcribed due to the presence of the *bsr* cassette, which additionally contains stop codons in all possible reading frames.

Ax2 cells were transfected with the Lig3 disruption construct using standard procedures, and cell growth in blasticidin was observed after two weeks. The intended recombination event following transfection is shown in Figure 4.6A; however, a random integration of the construct would also yield resistance to blasticidin. Therefore, we employed a screening strategy to identify *lig3*$^{-}$ cells. Firstly, we extracted genomic DNA from potential *lig3*$^{-}$ clones and performed PCR verification. In addition to the positive control for DNA content (primers EX1L and EX1R, in the *exo1* gene), we chose primer combinations internal to the deleted region of the *lig3* gene, and spanning from inside the BSR cassette to a primer external to the region of homologous recombination (shown in Figure 4.6A). These latter primer combinations are unique to a successful disruption of *lig3*. The results

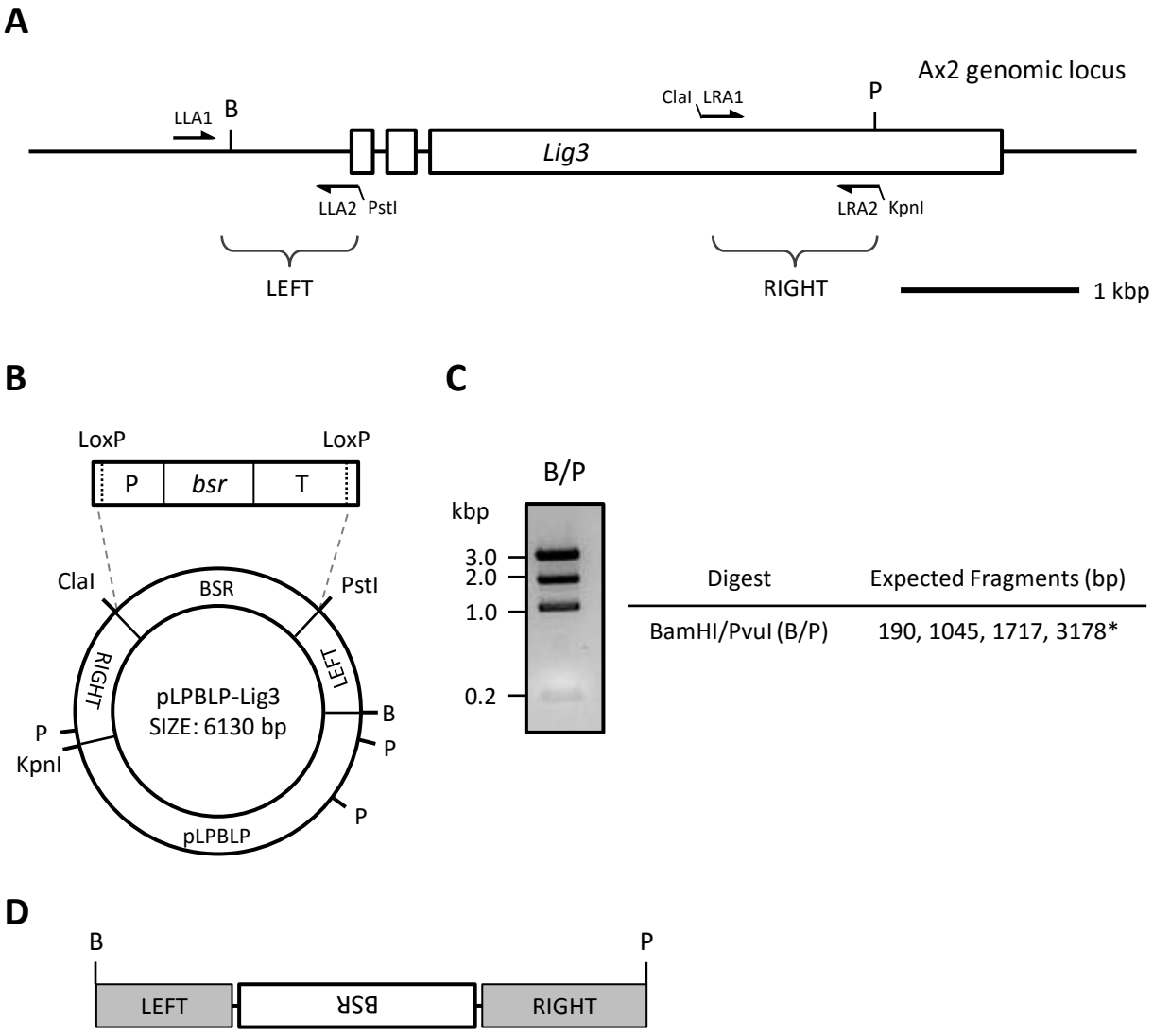
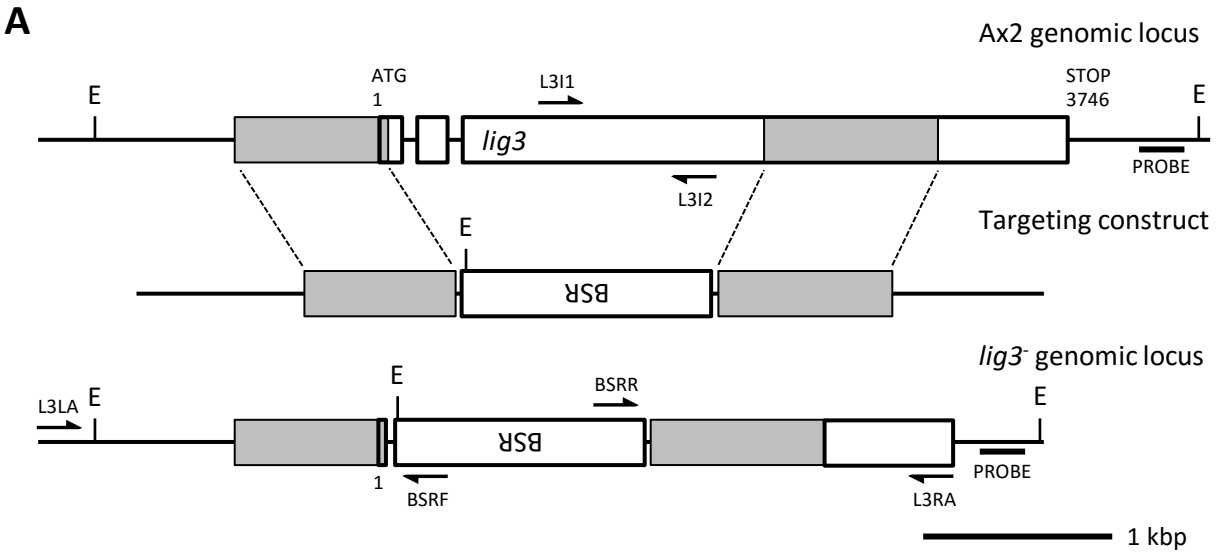
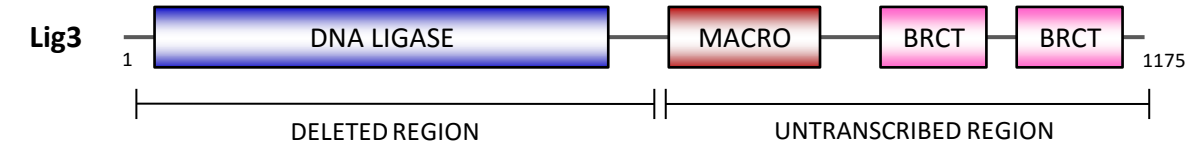
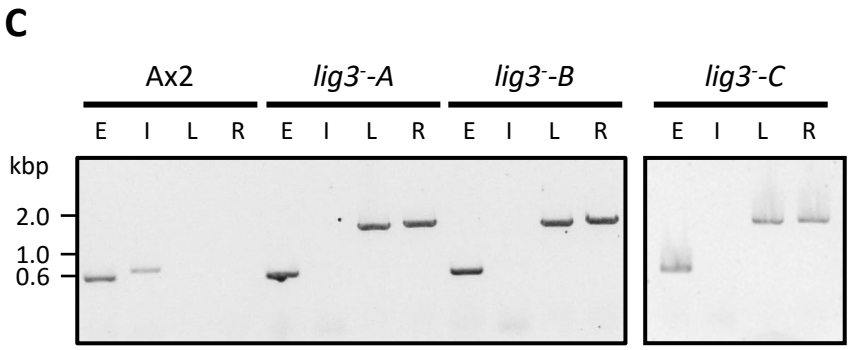
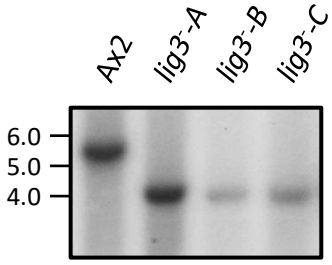


Figure 4.5 – Strategy for disruption of the *lig3* gene by targeted homologous recombination. A: Diagrammatical representation of the *lig3* genomic locus in *Dictyostelium* cells, indicating the location of the homologous arms to be used for disruption of the *lig3* gene by targeted homologous recombination. Location of the primers used, and attached restriction sites, are indicated. Endogenous BamHI (B) and PvuI (P) restriction sites, marking the ends of the homologous region, are also shown. **B:** Map of the pLPBLP-Lig3 disruption construct. The homologous arms in **A** were ligated into the pLPBLP vector, either side of a cassette encoding blasticidin resistance (BSR; blasticidin is an antibiotic toxic to the Ax2 strain). The restriction sites used for this, and verification of the plasmid, are shown (BamHI (B), PvuI (P)). The structure of the blasticidin resistance cassette is shown, containing the *bsr* gene, Actin6 promoter (P) and terminator (T). LoxP sites for excision of the cassette by Cre recombinase are also indicated. **C:** Verification of the pLPBLP-Lig3 strain by restriction digestion with BamHI and PvuI, simultaneously (B/P). Expected fragments are shown in the table, with the asterisk marking the fragment containing the homologous arms. **D:** Final construct for the disruption of the *lig3* gene in Ax2 cells, obtained through restriction digestion of pLPBLP-Lig3 with BamHI (B) and PvuI (P).









Primers	Expected PCR Products (bp)	
	Ax2	*lig3*-
EX1L/EX1R (E)	500	500
L3I1/L3I2 (I)	646	N/A
L3LA/BSRF (L)	N/A	1544
BSRR/L3RA (R)	N/A	1565

Strain	Expected Fragments (kbp)
Ax2	6.06
lig3-	4.25

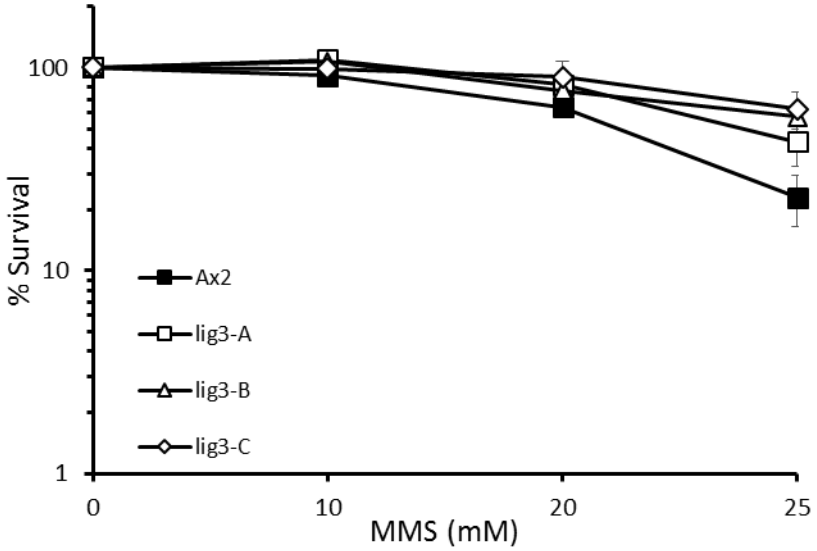
Figure 4.6 – Generation and verification of *lig3*$^{-}$ *Dictyostelium* strains. A: Representation of the intended recombination event to generate *lig3* disruption strains (*lig3*$^{-}$) by integration of a cassette encoding resistance to blasticidin (BSR). Also indicated are the primers subsequently used for verification of the *lig3*$^{-}$ strains by PCR, and the EcoRV (E) restriction sites and probe location for verification by Southern blotting. **B:** The domain map of Lig3, indicating the domains deleted in successful *lig3*$^{-}$ strains. Stop codons present in the BSR cassette prevent transcription of the C-terminal region of the gene. **C:** Verification of three putative *lig3*$^{-}$ strains by PCR, using the primers indicated in **A**. Primer combinations and expected products are shown in the table. **D:** Further verification of *lig3*$^{-}$ strains by Southern blotting. Genomic DNA was extracted from these strains and digested with EcoRV (E), at the sites shown in **A**. A probe was chosen so as to generate a detectable size-shift following digestion (**A**)**,** with fragment sizes shown in the table.

of the PCR screen indicated three *lig3*$^{-}$ clones (Figure 4.6B). This was supported by Southern blot analysis of the same genomic DNA preparations (Figure 4.6C). These data indicate that we have successfully disrupted the *lig3* gene in three independent clones.

Firstly, we investigated the sensitivity of these *lig3*$^{-}$ clones to base methylation, which requires the BER/SSBR pathway for its resolution. In humans, Lig3α has been shown to function in this pathway[429]. Therefore, we exposed Ax2 cells and the three independent *lig3*$^{-}$ clones to a range of MMS concentrations and investigated their survival following this treatment by assessing post-treatment plaque growth on agar. This assay showed that there is no survival defect of the *lig3*$^{-}$ cells to MMS compared to Ax2 cells, and potentially a slight resistance at higher doses (Figure 4.7). This was not completely unexpected as the same phenotype is observed in mammalian cells depleted of Lig3, due to the ability of Lig1 to substitute for Lig3 in SSBR[419].

4.2.4. Absence of Lig3 does not sensitise DSBR-defective cells to DSBs

In addition to its role in SSBR, we also wanted to investigate the role of *Dictyostelium* Lig3 in DSBR. Lig3 is implicated in alt-NHEJ, a DSBR pathway thought to function in the absence of NHEJ or HR (depending on cell-cycle phase)[169,177]. Therefore, the sensitivity of cells following loss of Lig3 to induced DSBs was assessed. Firstly, we considered the survival of *lig3*$^{-}$ cells to DSBs. The depletion of Lig3α in human cells does not result in a survival deficiency to induced DSBs, as NHEJ and HR are able to resolve the induced breaks, and we wished to see if this is true in *Dictyostelium*. Therefore, we performed the same plaque growth survival assay as before, except we treated Ax2 and *lig3*$^{-}$ cells with phleomycin instead of MMS, to induce DSBs. In addition to these strains, we also assessed the sensitivity of *dnapkcs*$^{-}$ and *lig3*$^{-}$*dnapkcs*$^{-}$ cells to DSBs (the *lig3*$^{-}$*dnapkcs*$^{-}$ strain was generated by Lena Kolb). In vegetative *Dictyostelium* cells, NHEJ appears dispensable



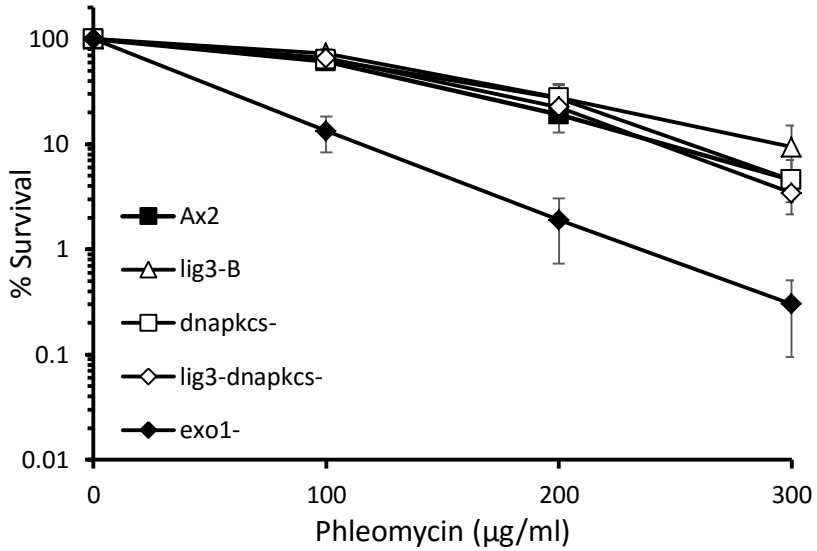
Strain	D_{50} (mM)	SE	P-value (Ax2)
Ax2	21.4	0.77	N/A
lig3⁻-A	24.3	1.12	0.10
lig3⁻-B	27.9	2.79	0.15
lig3⁻-C	26.1	1.84	0.10

Figure 4.7 – Vegetative *lig3⁻ Dictyostelium* cells are not sensitive to MMS-induced DNA damage. Ax2 and three *lig3⁻* strains were exposed to the indicated concentrations of MMS for 1 hour, before washing and plating onto SM agar. Survival was assessed by plaque growth after 3-6 days, relative to an untreated control. The error bars represent the standard error of three independent experiments. D_{50} toxicity values are provided in the table, with standard errors (SE). P-values are calculated with respect to the bracketed strain.

when considering survival, and it is known that vegetative *dnapkcs*$^{-}$ cells are not hypersensitive to DSBs compared to Ax2[358]. The reason for this is hypothesised to be due to the reliance of *Dictyostelium* cells in vegetative culture on HR, an idea which is supported by the high percentage of the population in G2 phase of the cell cycle[330]. However, the role of alt-NHEJ in the survival of vegetative *Dictyostelium* cells has not been elucidated, and this pathway may also compensate for the deficiency in NHEJ in *dnapkcs*$^{-}$ cells. As predicted, the HR-defective *exo1*$^{-}$ cells included as a positive control displayed increased sensitivity to DSBs in this assay; however, *lig3*$^{-}$ cells do not when compared to Ax2 cells, indicating that Lig3 is not required for survival of induced-DSBs in this genetic background (Figure 4.8A). Furthermore, the *lig3*$^{-}$*dnapkcs*$^{-}$ strains do not display increased sensitivity to phleomycin-induced DSBs compared to either Ax2 or *dnapkcs*$^{-}$ cells. These data either assign HR an even stronger function in resolving DSBs in *Dictyostelium* vegetative cells, or highlights the redundancy of Lig1 and Lig3 in alt-NHEJ.

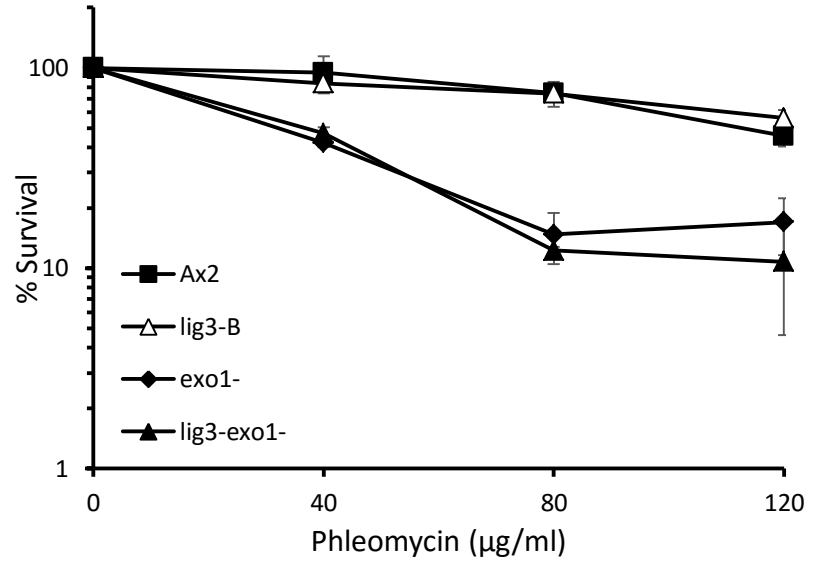
These experiments only consider the role of Lig3 in the repair of DSBs in genetic backgrounds of Ax2 cells or *dnapkcs*$^{-}$ cells, which both display the same level survival following treatment with DSB-inducing agents in vegetative culture. Alt-NHEJ was discovered as a back-up DSBR pathway, and therefore should be studied in genetic backgrounds in which the primary DSBR pathways of the cell are compromised[106,107]. In vegetative *Dictyostelium* cells, there is a strong reliance on HR for DSBR. Although alt-NHEJ was originally believed to be a back-up pathway for classical NHEJ, it has recently been discovered that this pathway functions in the absence of HR in mammalian cells[177]. Therefore, we hypothesised that disruption of *lig3* in a HR-deficient strain may lead to further sensitisation of vegetative *Dictyostelium* cells to DSBs. To address this, we created a *lig3*$^{-}$*exo1*$^{-}$ strain to assess the contribution of alt-NHEJ to DSBR in a HR-deficient background (strain generated by Lena Kolb). We then exposed vegetative *lig3*$^{-}$*exo1*$^{-}$ cells

A



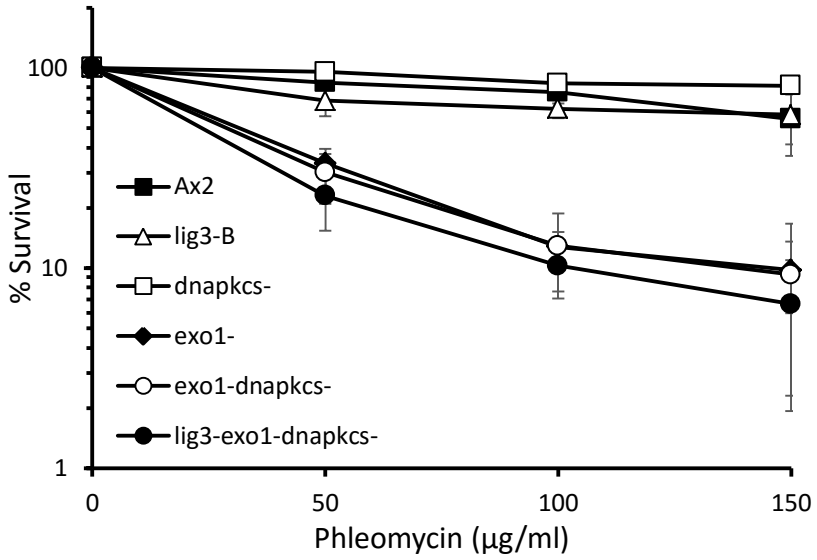
Strain	D_{50} (µg/ml)	SE	P-value (Ax2)	P-value (*dnapkcs*$^-$)
Ax2	117	22.1	N/A	N/A
lig3$^-$	134	20.3	0.60	N/A
dnapkcs$^-$	141	26.3	0.54	N/A
lig3$^-$*dnapkcs*$^-$	132	13.3	N/A	0.78
exo1$^-$	45.6	5.13	0.11	N/A

B



Strain	D_{50} (µg/ml)	SE	P-value (Ax2)	P-value (*exo1*$^-$)
Ax2	102	15.2	N/A	N/A
lig3$^-$	156	31.1	0.21	N/A
exo1$^-$	33.5	2.60	0.047 (*)	N/A
lig3$^-$*exo1*$^-$	35.3	2.31	N/A	0.63

C



Strain	D_{50} (µg/ml)	SE	P-value (Ax2)	P-value (*exo1*$^-$)	P-value (*exo1*$^-$*dnapkcs*$^-$)
Ax2	169	19.1	N/A	N/A	N/A
lig3$^-$	141	35.1	0.53	N/A	N/A
dnapkcs$^-$	289	65.2	0.22	N/A	N/A
exo1$^-$	33.4	2.20	0.02 (*)	N/A	N/A
exo1$^-$*dnapkcs*$^-$	35.2	8.69	N/A	0.86	N/A
lig3$^-$*exo1*$^-$*dnapkcs*$^-$	28.3	4.16	N/A	N/A	0.36

Figure 4.8 – Vegetative *lig3*$^-$ *Dictyostelium* cells are not sensitive to DSBs in NHEJ- or HR-defective backgrounds. The indicated strains were exposed to various concentrations of phleomycin for 1 hour, before washing and plating on SM agar plates. Survival was assessed by observing plaque formation after 3-7 days, relative to mock-treated controls. **A:** The survival of *lig3*$^-$ cells and *lig3*$^-$*dnapkcs*$^-$ were compared to Ax2 and NHEJ-defective *dnapkcs*$^-$ cells, respectively, while HR-defective *exo1*$^-$ cells were included as a positive control. **B:** The survival of *lig3*$^-$*exo1*$^-$ cells was compared to that of the HR-defective *exo1*$^-$ cells. **C:** The effect on survival of the disruption of Lig3 in the HR- and NHEJ-defective *exo1*$^-$*dnapkcs*$^-$ strain. The error bars represent the standard error of three independent experiments. D_{50} toxicity values with standard errors (SE) are provided in the table. P-values are calculated with respect to the bracketed strain. The *lig3*$^-$*dnapkcs*$^-$, *lig3*$^-$*exo1*$^-$, *exo1*$^-$*dnapkcs*$^-$ and *lig3*$^-$*exo1*$^-$*dnapkcs*$^-$ strains were generated by Lena Kolb.

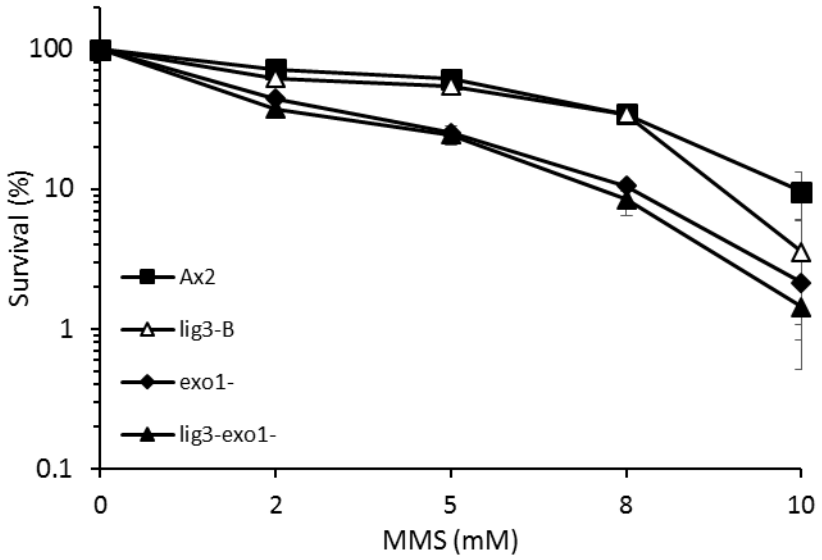
to phleomycin in a plaque-forming survival assay. However, the absence of Lig3 did not further sensitise *exo1*$^{-}$ cells to DSBs (Figure 4.8B). This indicates that Lig3 is not required in the absence of *exo1*$^{-}$ for survival of DSBs, and therefore, it is still not clear which DSBR pathway is responsible for resolving DSBs in *exo1*$^{-}$ cells. It is known that NHEJ factors are enriched on chromatin following DSBs in vegetative cells, so that pathway does appear to be active[360]. To address the role of NHEJ in DSBR in *exo1*$^{-}$ cells, we generated an *exo1*$^{-}$ *dnapkcs*$^{-}$ strain and a *lig3*$^{-}$*exo1*$^{-}$*dnapkcs*$^{-}$ strain (both generated by Lena Kolb). The latter strain will assess the role of Lig3 in a NHEJ- and HR-defective background, where alt-NHEJ is believed to be the only active DNA DSB repair pathway. These strains were assessed by plaque formation after phleomycin exposure, along with the single-mutant strains. However, both the *exo1*$^{-}$*dnapkcs*$^{-}$ and *lig3*$^{-}$*exo1*$^{-}$*dnapkcs*$^{-}$ cells do not display elevated sensitivity to DSBs compared to *exo1*$^{-}$ strains (Figure 4.8C). These data indicate that NHEJ does not contribute to the survival of vegetative *Dictyostelium* cells following exposure to DSBs, and Lig3 is also dispensable in the absence of HR and NHEJ.

The majority of vegetative *Dictyostelium* cells are in G2-phase of the cell cycle; however, in mammalian cells, alt-NHEJ has been shown to function in the absence of HR in S-phase cells. Therefore, the defect associated with Lig3 deletion may only be uncovered upon encountering S-phase associated damage. In order to probe the *Dictyostelium* DDR in S-phase, we chose to incubate cells with a low dose of MMS for 4 hours. In this system, in addition to more cells passing through S-phase and suffering DNA damage, unrepaired lesions inflicted in other cell-cycle phases may also be carried into S-phase and act as blocks to replication forks[444]. It has been shown that these MMS-induced, replication-associated DSBs are dependent on HR for repair, making this system suitable for our purpose[445]. Therefore, we exposed *lig3*$^{-}$*exo1*$^{-}$ cells to this MMS treatment and assessed survival by plaque formation. Again, *lig3*$^{-}$*exo1*$^{-}$ cells were no more sensitive to this

treatment than *exo1*$^{-}$ cells; however, the increased sensitivity of *exo1*$^{-}$ cells indicates repair by HR is occurring in Ax2 cells in these conditions (Figure 4.9). Collectively, these data indicate that Lig3 is not required for DSBR in vegetative cells deficient in NHEJ, HR or both pathways.

4.3. Discussion

In this chapter, we have shown that the macro domain of *Dictyostelium* Lig3 binds to PAR chains *in vitro*. This implicates PARylation directly in the role of Lig3 in this organism. In humans, Lig3α has been shown to interact with PAR chains, and auto-PARylated ARTD1, *in vitro*; however the dependency of this binding on a particular domain has not been fully established[244]. Additionally, Lig3 homologs in other organisms have been found to contain macro domains (*Naegleria gruberi* - amoeba-flagellate) or PBZ domains (*Strongylocentrotus purpuratus* – sea urchin), in a parallel study that also identified the macro domain in *Dictyostelium* Lig3[400]. However, the binding characteristics of these domains have not been experimentally analysed. Regardless, the presence of these domains is suggestive that a direct mode of binding of Lig3 to PAR chains is present in a variety of organisms. PARylation plays an important role in all of the pathways in which Lig3 has been implicated. In SSBR, ARTD1 (or ARTD2) acts as a sensor protein for DNA lesions[446]. Upon detection and activation, ARTD1 auto-PARylates itself and PARylates proteins proximal to the lesion, leading to the recruitment of the XRCC1-Lig3 heterodimer[68]. In this model of SSBR, recruitment of Lig3 requires XRCC1, and direct interactions of Lig3 with PARylated ARTD1 are not implicated. An alternative model of SSBR, in which Lig3 acts as the sensor of DNA nicks and is recruited first, has been proposed[86]. This model is PARylation-independent, requiring instead the PARP-type zinc finger found at the N-terminus of human Lig3. *Dictyostelium* Lig3, like the *Drosophila*



Strain	D_{50} (mM)	SE	P-value (Ax2)	P-value (*exo1*$^-$)
Ax2	5.60	0.69	N/A	N/A
lig3$^-$	4.46	0.32	0.23	N/A
exo1$^-$	1.73	0.08	0.03 (*)	N/A
lig3$^-$*exo1*$^-$	1.49	0.18	N/A	0.32

Figure 4.9 – Deletion of *lig3* does not further sensitise *exo1*$^-$ cells to DNA damage in S-phase. A: The indicated strains were exposed to various concentrations of MMS for 4 hours, before washing and plating out on SM agar. Survival was assessed by observing plaque formation after 3-7 days, relative to an untreated control. The error bars represent the standard error from three independent experiments. D_{50} toxicity values with standard errors are shown in the table. P-values are calculated with respect to the bracketed strain. The *lig3*$^-$*exo1*$^-$ strain was generated by Lena Kolb.

melanogaster homolog, lacks this zinc finger domain, and so this mode of recruitment to DNA nicks appears to not be conserved in these organisms[400]. However, the PAR-binding macro domain of *Lig3* in *Dictyostelium* may bind to auto-PARylated Adprt2, the *Dictyostelium* ortholog of ARTD1, and this interaction could serve to recruit or stabilise Lig3 on damaged chromatin[360]. This model would indicate a much stronger reliance of the function of Lig3 on PARylation. We have directly observed the enrichment of Lig3 on chromatin following induction of DNA base methylation, implicating the *Dictyostelium* protein in SSBR. However, we have yet to assess the dependence of PARylation on this, or the macro domain. Using a sequence alignment of the macro domain of Lig3 with that of human ALC1, we have predicted several amino acids that may be necessary for the PAR-binding behaviour of the domain, but they have yet to be assessed *in vitro*. Abrogation of PAR-binding by point mutation is a much more desirable approach to deletion of the macro domain region of Lig3, or deletion of *Dictyostelium* ARTs, and so assessing the effect of these mutations will be important in establishing the function of the macro domain *in vivo*.

We also assessed whether *Dictyostelium lig3*$^{-}$ cells display a survival defect following treatment with a base-alkylating agent, and found that they did not. In fact, all three *lig3*$^{-}$ strains displayed a slight, but significant, resistance to this genotoxic agent at high doses. The lack of sensitivity of Lig3-depleted strains has been observed in other organisms, including human cells[88]. In these organisms, it is thought that Lig1 is able to act in a redundant manner with Lig3. Both Lig1 and human LigIIIα have been shown to function in BER, in the long- and short-patch variants, respectively[442,443]. However, Lig1 can function in short-patch repair *in vitro*[419]. We hypothesise that this redundancy also exists in Dictyostelium, thereby allowing cells to repair the damage induced by base alkylation in a Lig3-independent manner. Depletion of both Lig1 and Lig3 is lethal; however, it would be

interesting to observe the role of Lig3 in a *lig1*$^{-}$ background. Lig1 is thought to be recruited to SSBs though a protein-protein interaction with PCNA, and is not currently associated with PARylation[84]. In a *lig1*$^{-}$ background, it would be hypothesised that SSBR would display an even stronger, non-redundant, reliance on Lig3, and potentially its PAR-binding macro domain. Therefore, an alternative approach to studying the role of Lig3 in *Dictyostelium* would be to first delete *lig1*, and then perform knockdown or conditional knockout of Lig3. This approach would be combined with expressing macro domain mutant forms of Lig3 in the *lig1*$^{-}$ background to specifically assess the role of that domain in the function of Lig3.

We have also assessed the role of Lig3 in alt-NHEJ, another DNA repair pathway dependent on ADP-ribosylation in humans. In this pathway, ARTD1 detects the DSB and directly competes with the Ku heterodimer[189]. In this manner, Ku promotes NHEJ and suppresses alt-NHEJ in normal circumstances, while ARTD1 promotes alt-NHEJ. The molecular mechanism of alt-NHEJ is yet to be fully understood, and it is not currently clear if it operates as a single pathway or as multiple sub-pathways[180]. However, one might hypothesise that auto-PARylated ARTD1 and Lig3 may interact in the early stages of this pathway. In vegetative *Dictyostelium* cells, we found that disruption of the *lig3* gene did not elevate the sensitivity of Ax2 cells to DSBs. As alt-NHEJ is observed to be important for survival in the absence of NHEJ or HR, the lack of a sensitivity of *lig3*$^{-}$ cells alone is not surprising. It is known that vegetative *Dictyostelium* cells do not require functional NHEJ to survive DSBs, and cells that are HR-defective display hypersensitivity to DSB-inducing agents[351]. However, it was not known to what degree alt-NHEJ plays a role in the tolerance of DSBs in a NHEJ-deficient background. Similarly, it was not known which DSBR pathway was responsible for the survival of cells in the HR-defective, *exo1*$^{-}$ strain, following exposure to DSB-inducing agents. Our experiments with *lig3*$^{-}$*dnapkcs*$^{-}$ and *lig3*$^{-}$

exo1^{-} cells show no elevated sensitivity compared to the dnapkcs^{-} and *exo1*$^{-}$ cells, respectively. Therefore, Lig3 is not implicated in the survival of these strains to DSBs; however, we are unable to rule out alt-NHEJ due to the known redundancy between human Lig1 and Lig3α in this pathway[171,179,447]. A model has been proposed in which Lig1 and Lig3α act in different pathways of alt-NHEJ, with Lig3α ligating DSBs with the use of microhomologies; however, they may also act in a redundant manner in a single pathway[180]. The assessment of the survival of cells performed here does not distinguish between these two models. Furthermore, the same lack of additional sensitivity of *lig3*$^{-}$ *exo1*$^{-}$ was observed when inducing S-phase-associated DNA damage. Interestingly, *exo1*$^{-}$ *dnapkcs*$^{-}$ cells were no more sensitive to DSBs than *exo1*$^{-}$ cells, indicating that NHEJ does not act as a backup DSBR pathway in this genetic context.

Overall, these data can be explained by three possible theories. In the first, Lig3 is not active in alt-NHEJ, and so its mutation is irrelevant for DSBR. We have not yet assessed the enrichment of Lig3 on chromatin following the induction of DSBs in any genetic background, so this remains a possibility, although the volume of literature on Lig3 functioning in alt-NHEJ and the generally high level of conservation of *Dictyostelium* DNA repair proteins makes this unlikely[171,179,321,441]. Secondly, the function of Lig3 is completely redundant with that of Lig1 in this context, thereby leaving alt-NHEJ fully functional in *lig3*$^{-}$ cells. This redundancy is observed in other organisms, as previously described, and is the most plausible of the three theories presented here. However, it has been shown in DT40 cells that deletion of Lig3 does result in slower kinetics of DSBR[179]. We have not yet assessed repair rates in our *lig3*$^{-}$ strains. The final possible model would be residual HR accounting for the survival of *exo1*$^{-}$ cells to DSBs. Exo1 may function in a redundant manner with other exonucleases, for example Dna2 or Mre11, and this may be sufficient for survival[448]. In *Dictyostelium,* deletion of most HR genes is results in loss of

viability, with *exo1*$^{-}$ cells being an exception[347,358]. This would also be explained by the presence of redundant exonuclease activity, and suggests that it may be present. This model would assign alt-NHEJ, and Lig3, a very low contribution to survival of DSBs, considering *exo1*$^{-}$ cells themselves already display a substantial survival defect[351].

The lack of an observable survival phenotype of *lig3*$^{-}$ cells in response to DSBs in any genetic background has posed an obstacle to assessing the role of Lig3 in alt-NHEJ in *Dictyostelium*, and we have not yet studied whether or not Lig3 is enriched on chromatin following DSBs. The only function for which Lig3 has been observed to act non-redundantly is in mediating fusions of unprotected telomeres via the alt-NHEJ pathway; however, telomeres are poorly characterised in *Dictyostelium*[174,327]. Therefore, we have not been able to characterise the function of the macro domain of Lig3 *in vivo*. In addition to showing that this macro domain binds to PAR chains, we could expand our *in vitro* assays to assess any interaction between the macro domain of Lig3 and auto-modified Adprt2, which might further implicate Lig3 in either SSBR or alt-NHEJ. However, to uncover the *in vivo* function of Lig3, a system which allows simultaneous or conditional depletion of Lig1 may be required.

5. Initial characterisation of APL

5.1. Introduction

5.1.1. Aprataxin-like protein (APL)

APL is a 563 amino acid protein with three currently annotated domains: an FHA domain, a PBZ domain (previously identified), and a circularly permuted macro domain[241]. All three of these domains, including the circularly permuted macro domain, which is predicted to retain the structure of an unpermuted domain, are potential PAR-binding domains. There is no human protein yet discovered to contain both a PBZ and a macro domain, which could suggest a novel mode of action for the protein if both domains are indeed functional. Multiple sequence alignments with the macro domain of APL show that it displays similarity to the first and second macro domains of those found in human ARTD8, a MART recently implicated in homologous recombination in response to replication-associated DSBs. The second macro domain of ARTD8 binds to MARylated proteins, and this may be true of the macro domain identified in APL[35]. Interestingly, the FHA domain of APL is found in three human DDR proteins: aprataxin (APTX), aprataxin and PNK-like factor (APLF), and polynucleotide phosphatase/kinase (PNKP). FHA domains are well-established mediators of phospho-specific protein-protein interactions, and the presence of this FHA domain in APL could implicate the protein in binding to the same DNA repair factors as APTX, APLF and PNKP[405]. The specific roles of these DDR proteins will be discussed below, in the context of them being the most closely related human proteins to APL, and may therefore provide an indication of its role in the DDR in *Dictyostelium*.

5.1.2. Potential human homologs of APL

5.1.2.1. Aprataxin

Aprataxin (APTX) in humans was first identified as a product of a gene that is mutated in the neurodegenerative condition ataxia-ocular apraxia 1 (AOA1)[449,450]. The role of APTX in the DDR was suggested by the discovery that it binds to XRCC1 and XRCC4, *in vitro* and *in vivo*, thereby linking it to both SSBR and NHEJ. This interaction occurs through its FHA domain. Assessment of the sensitivity of AOA1 positive or APTX-deficient cells to both DSB- and SSB-inducing agents have produced inconsistent results, and it is still not clear if loss of functional APTX results in a cellular survival defect following DNA damage[161,451]. It has since been shown that APTX encodes a nuclear and mitochondrial isoform, and that depletion of APTX results in mitochondrial dysfunction[452].

Since its initial placement in the DDR, the specific role of APTX has since been uncovered. Alongside its FHA domain, APTX also has a histidine triad (HIT) domain and a DNA-binding zinc finger (zf-C2H2 type) domain. In other proteins, the HIT domain has nucleotide hydrolysis and transferase functions[453]. The function of APTX has been shown to remove 5'-AMP groups from DNA ends, which are typically present following aborted ligation attempts, thereby classifying APTX as a DNA end-processing enzyme[74]. In addition to APTX, at least two other enzymes have been shown to be able to process 5'-AMP ends *in vitro*, albeit with less subtlety than APTX. Polβ and FEN1 have both been shown to be able to compensate for APTX deficiency in SSBR, perhaps suggesting the reason for the lack of clear DNA damage sensitivity in APTX-deficient cells[81].

In contrast to APTX, APL has no HIT domain. However, its permuted macro domain may possess catalytic function, allthough no human macro domain has been shown to target 5'-AMP groups. *Dictyostelium* has no clear homolog of APTX; however, there is one protein

with a HIT domain similar to that of APTX, albeit lacking an FHA domain (UniProt accession: Q558W0). Therefore, APL does not appear to be a direct homolog of APTX, although it may function in a complex with other proteins to achieve the same overall functionality.

5.1.2.2. Polynucleotide kinase/phosphatase (PNKP)

Early experiments indicated that PNKP had 5'-DNA kinase and 3'-phosphatase activity *in vitro* and that it acts to process 3' DNA ends *in vivo* following SSBs[75,76]. PNKP has since been shown to act to remove 3'-phosphate groups and to phosphorylate 5'-hydroyl groups, thereby processing DNA ends to make them suitable for ligation. Interactions with the scaffold proteins XRCC1 and XRCC4 have been shown to be important for this end-processing activity, in SSBR and NHEJ, respectively. As is true with APTX, phospho-specific interactions of PNKP with XRCC1 and XRCC4 are mediated by the N-terminal FHA domain of PNKP[454,455]. More recently, mutations in PKNP have been linked with a neurodevelopmental disorder. Cells from patients diagnosed with this condition displayed slower kinetics of repair of SSBs and S-phase-induced DSBs, with some SSBs remaining unresolved at the conclusion of the experiment[406].

The only link between the domain structure of APL and that of PNKP is the N-terminal FHA domain, which binds to XRCC1 and XRCC4. APL has no annotated domains that would be predicted to process DNA ends, unless the macro domain has novel catalytic activity. Additionally, a putative *Dictyostelium* PNKP has been identified *in silico*; however, this protein does not contain the FHA domain required for targeting human PNKP to DNA lesions (UniProt accession: Q54U78). Furthermore, a role for PARylation in the end processing activity of PNKP has not been discovered. These data suggest that APL is unlikely to be a direct homolog of PNKP.

5.1.2.3. Aprataxin- and PNK-like factor (APLF)

Aprataxin- and PNK-like factor (APLF) was identified as a putative DDR protein due to the similarity of its FHA domain to that of APTX and PNKP, which were known to bind to XRCC1 and XRCC4 in a phospho-specific manner. The same phospho-specific interactions were subsequently found to exist between APLF and both XRCC1 and XRCC4, *in vitro* and *in vivo*, thereby placing APLF in the DDR. It has also been demonstrated that APLF displays endo- and exonuclease activity on damaged DNA *in vitro*, although the domain-dependence of this activity is not clear[456]. *In vivo*, APLF was observed to be recruited to sites of SSBs and DSBs, with the former being mediated partially by XRCC1. Phenotypically, depletion of APLF results in slower repair kinetics of SSBs and DSBs, and APLF deletion has also been reported to increase cellular sensitivity to these forms of DNA damage[162,457].

It was subsequently identified that the enrichment of APLF at DNA strand breaks was dependent on PARylation, through interactions mediated by the tandem PBZ domains of APLF[241]. In response to UVA laser damage, APLF enrichment on chromatin is strongly correlated with the activity of ARTD1[267]. APLF increases the stability of NHEJ proteins at DSBs, thereby accelerating the pathway. It does this in concert with ARTD3, which PARylates histone H1 variants and makes them targets for APLF binding[15]. Additionally, APLF was shown to directly bind to Ku80 *in vivo*, in a manner dependent on a specialised Ku-binding motif in APLF[151,458].

The domain structure of APLF more closely resembles that of APL. As is also true for APTX, the FHA of APLF also shows a strong level homology with that of APL, providing more weight to the argument that APL may function in SSBR or DSRB. Aside from its Ku-binding motif, which is absent in APL, APLF has tandem PAR-binding PBZ domains

that act together to increase the affinity of APLF to PAR chains[268]. However, it is not clear which domain of APL would possess nuclease activity. As no *Dictyostelium* APLF has been identified, it is possible that APL could perform the role of this protein, with a PAR-binding macro domain substituting for the second PBZ domain of APLF. However, the existence of APLF in *Dictyostelium* would conflict with the current model of NHEJ in this organism, where stability of NHEJ factors at DNA DSBs in dependent on the C-terminal PBZ domain of Ku70[360].

5.1.3. Aims

The first aim of this chapter is to identify if APL is a PAR-binding protein, and if so, through which domains that it binds to PAR chains. This will be assessed *in vitro* using purified proteins and commercially available PAR polymers. While this analysis will not shed any light on the cellular function of the protein, as PARylation is implicated in numerous cellular processes, it will serve to verify the *in silico* screen previously performed. Characterisation of the domain-dependence of any identified PAR-binding may provide insight into the mode of action of the protein, and may also indicate whether or not the macro domain is functional.

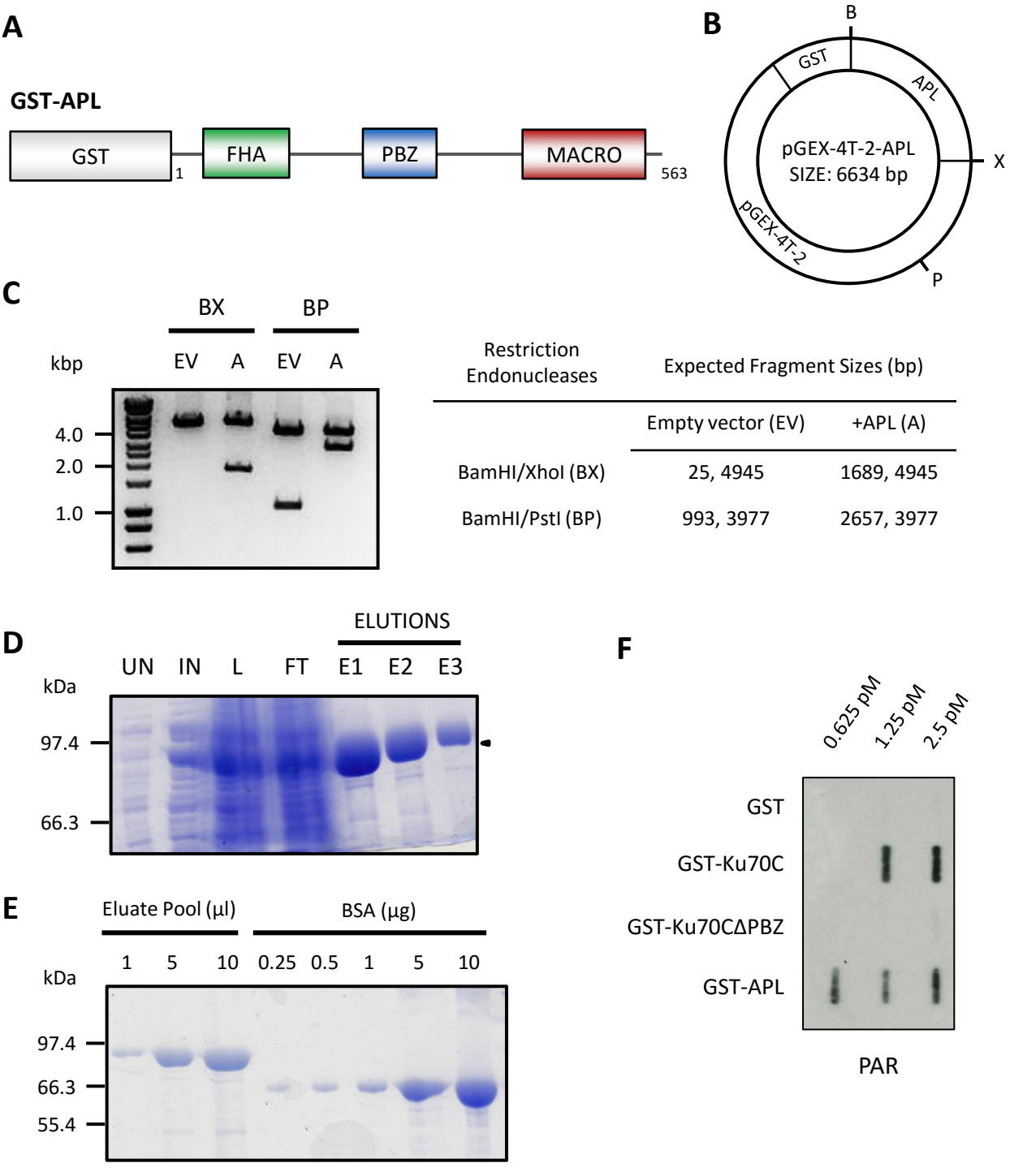
Secondly, we will address the role of APL in the DDR. The reverse genetics approach adopted by this thesis means that APL could play any role in any DNA repair pathway, and so a range of genotoxic chemicals will be employed to activate a wide range of responses. We have chosen two experimental readouts for the initial characterisation of APL: sensitivity of *apl*$^{-}$ cell lines and enrichment of APL on chromatin following exposure to various DNA damaging agents. A positive result in either of these assays will implicate APL in the response to DNA damage. This will allow for more thorough analysis of the function of APL in subsequent assays.

5.2. Results

5.2.1. APL displays *in vitro* PAR-binding activity

The putative DDR protein APL was identified in an *in silico* screen for proteins with novel PAR-binding domains. Therefore, the first stage of the functional analysis of this protein is to determine if APL indeed binds to PAR chains. Despite the loss of physiological relevance associated with *in vitro* assays, the turbulent nature of *in vivo* PARylation and lack of known stimulus leaves this type of assay as the most sensible option for a screen. A construct for the expression of APL was designed and ligated into the pGEX-4T-2 vector, a bacterial protein expression vector with an N-terminal GST-tag, using standard techniques (Figure 5.1A-B). The resulting plasmid was analysed by restriction digestion and Sanger sequencing to ensure that the vector was mutation-free (Figure 5.1C). Once transformed into *E.* coli, transcription from the vector leads to the expression of APL with an N-terminal GST-tag (GST-APL). The position of this tag was chosen so as to maximise its distance from the potentially PAR-binding macro domain, which will be examined in future assays. Purification of GST-APL provided a suitable yield, despite incomplete elution from the GST-beads (Figure 5.1D, beads not shown). Quantification by comparison with known BSA standards (Figure 5.1E) and by Bradford assay (data not shown) indicated a concentration of 0.53 mg/ml.

To investigate the PAR-binding nature of APL *in vitro*, increasing concentrations of GST-APL were slot-blotted onto nitrocellulose membrane. In addition to GST-APL, the same concentrations of two forms of the C-terminus of *Dictyostelium* Ku70 were slot-blotted in an identical manner for assay verification and comparison. These two forms were the endogenous C-terminus of Ku70, containing a PBZ domain known to bind PAR, and the C-terminus with this PBZ domain deleted (GST-Ku70C and GST-Ku70CΔPBZ,



Restriction Endonucleases	Expected Fragment Sizes (bp)	
	Empty vector (EV)	+APL (A)
BamHI/XhoI (BX)	25, 4945	1689, 4945
BamHI/PstI (BP)	993, 3977	2657, 3977

Figure 5.1 – APL is a PAR-binding protein *in vitro*. A: The domain structure of the N-terminally GST-tagged APL construct. **B:** A map of the product of ligating pGEX-4T-2, a GST-protein bacterial expression vector, and APL cDNA. Digestion with BamHI (B) and XhoI (X) excises the APL cDNA, whereas the PstI site (P) is used for plasmid verification. **C:** Verification of pGEX-4T-2-APL by restriction digestion. Enzymes used for digestion and expected fragment sizes are shown in the table. **D:** Purification of GST-APL (90.3 kDa) from BL21(DE3) *E. coli*. Cell equivalent samples were separated by SDS-PAGE and detected by Coomassie staining. Abbreviations: uninduced (UN), induced (IN), lysis (L), flow-through (FT). **E:** Quantification of the concentration of GST-APL eluted, by comparison with BSA standards. The indicated volumes and amounts of GST-APL and BSA, respectively, were separated by SDS-PAGE and detected by Coomassie staining. **F:** *In vitro* determination of the PAR-binding nature of GST-APL. Increasing concentrations of peptides were slot-blotted onto a nitrocellulose membrane and then incubated with PAR chains. Detection by Western blotting and comparison of GST and GST-APL indicates that APL does bind to PAR. GST-Ku70C and GST-Ku70CΔPBZ were included as positive and negative controls, respectively. Slot-blot experiment (F) performed by Peggy Paschke.

respectively)[360]. Incubation of this membrane with PAR polymers and subsequent Western blotting reveals that full length GST-APL does bind to PAR (Figure 5.1F). As expected, GST alone and GST-Ku70CΔPBZ display negligible and highly reduced binding, respectively. The slot-blot experiment was performed by Peggy Paschke.

Following the successful identification of APL as a PAR-binding protein, we wished to determine which domains were responsible for this behaviour. APL contains three domains: an FHA domain, a PBZ domain, and a permuted macro domain. Examples of all of these domain types have been shown to be PAR binding, with varying degrees of conservation of this function across their families[241,246,308]. To assess the domain-dependence of the PAR-binding of APL, a series of constructs of the isolated domains of APL were designed, with N-terminal GST tags. This was performed by amplifying those domains from the pGEX-4T-2-GST-APL vector by PCR, and ligating those products into empty pGEX-4T-2 vectors, which were then verified by Sanger sequencing. The proteins encoded by these constructs were expressed following transfection of the pGEX-4T-2 vectors in BL21(DE3) *E.coli*, and purified by GST purification. The purification process was monitored by SDS-PAGE and Coomassie staining (Figure 5.2A). The isolated domains of APL were then assessed for PAR-binding activity by dot-blotting onto nitrocellulose, followed by incubation with PAR polymers. Subsequent detection by Western blotting identified the PBZ (GST-PBZ) and macro (GST-MACRO) domains as both displaying PAR-binding to a similar magnitude as full length GST-APL, which is again seen to be lower than GST-Ku70C (Figure 5.2B). The FHA domain (GST-FHA) has negligible PAR-binding activity.

To further elucidate the PAR-binding nature of APL, we also assessed the ability of domain-deletion mutants to the protein to bind PAR (Figure 5.3A). This analysis would be predicted to corroborate the data from the isolated domains of APL. As previously, the

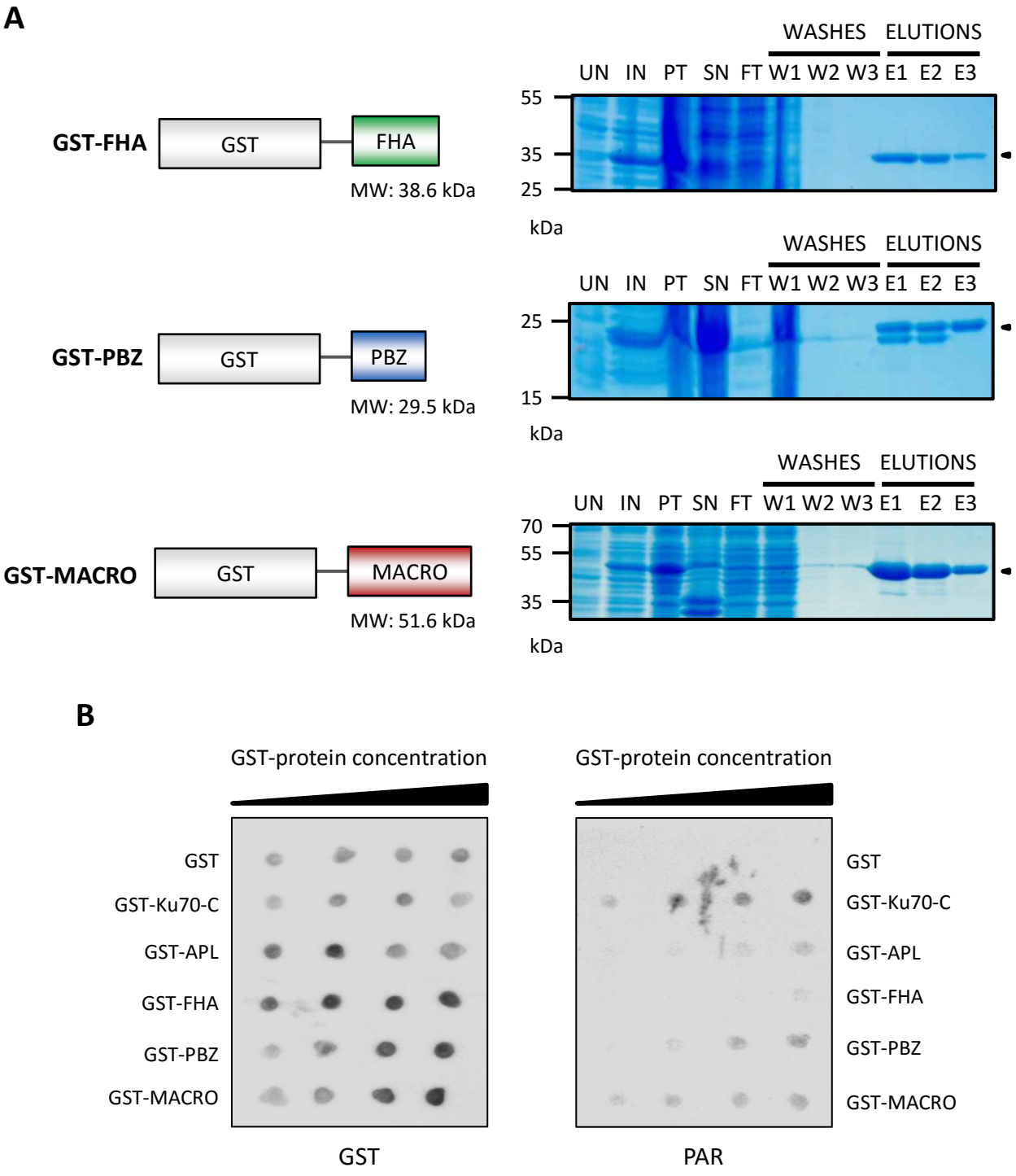


Figure 5.2 – The isolated PBZ and macro domains of APL bind to PAR *in vitro*. A: Domain structures and purification of the N-terminally GST-tagged individual domains of APL. The indicated domains of APL were amplified by PCR from Ax2 cDNA, and ligated into pGEX-4T-2 vectors. These vectors were transfected into BL21(DE3) *E. coli* cells, and were expressed with N-terminal GST-tags, before purification. The GST-purification procedure was monitored by SDS-PAGE and Coomassie staining. Abbreviations: uninduced (UN), induced (IN), pellet (PT), supernatant (SN), flow-through (FT). **B:** Determination of the PAR-binding abilities of the isolated domains of APL. Increasing concentrations of each peptide were dotted onto a nitrocellulose membrane and incubated with PAR chains, followed by detection by Western blotting. GST-Ku70C was included as a positive control. Experiments performed by Peggy Paschke and Mehera Emrich.

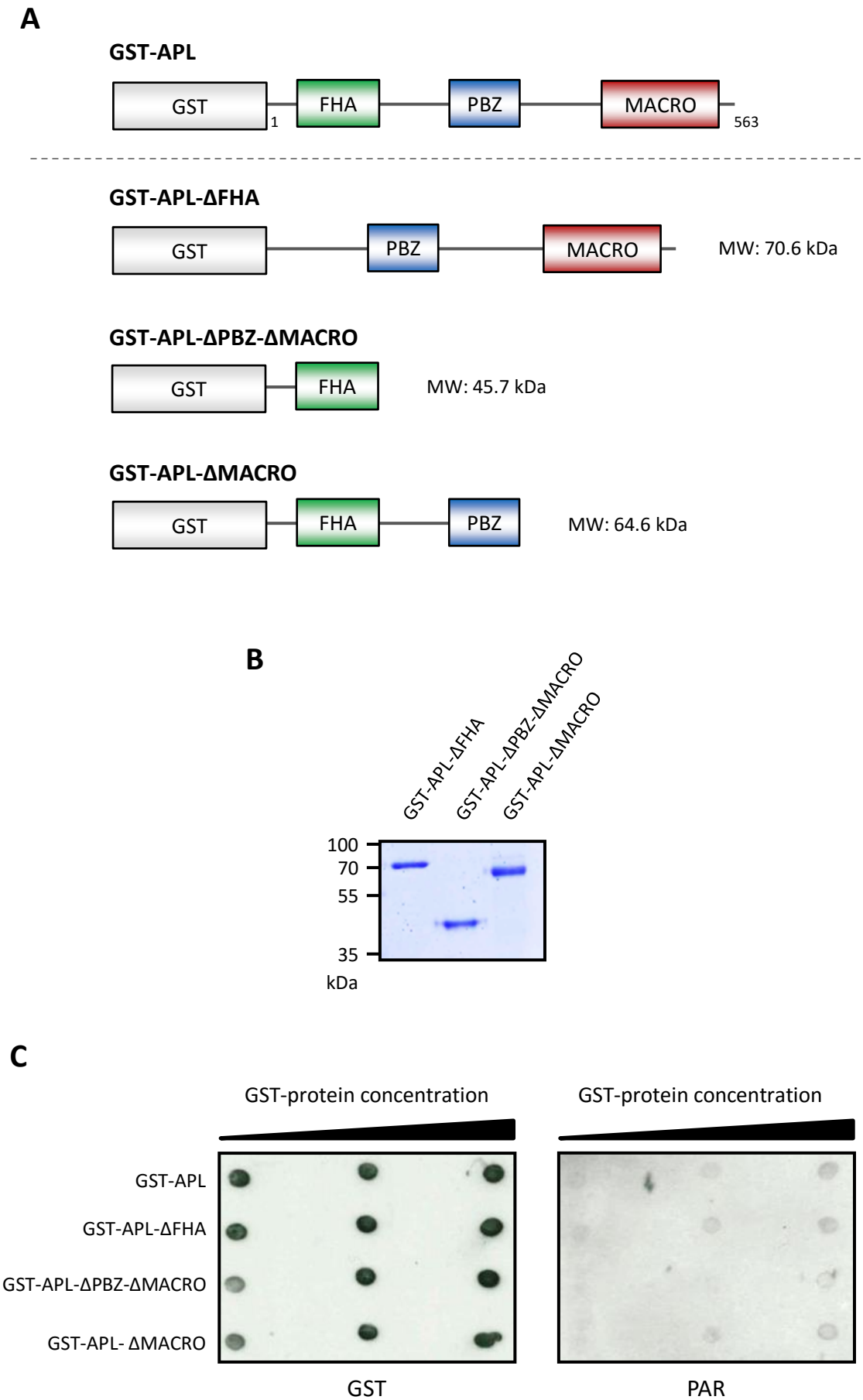

A
GST-APL
GST
1
FHA
PBZ
MACRO
563
GST-APL-ΔFHA
GST
PBZ
MACRO
MW: 70.6 kDa
GST-APL-ΔPBZ-ΔMACRO
GST
FHA
MW: 45.7 kDa
GST-APL-ΔMACRO
GST
FHA
PBZ
MW: 64.6 kDa
B
GST-APL-ΔFHA
GST-APL-ΔPBZ-ΔMACRO
GST-APL-ΔMACRO
100
70
55
35
kDa
C
GST-protein concentration
GST-protein concentration
GST-APL
GST-APL-ΔFHA
GST-APL-ΔPBZ-ΔMACRO
GST-APL- ΔMACRO
GST
PAR

Figure 5.3 – The PAR-binding ability of APL shows strongest reliance on its PBZ domain. A: Domain structures of the N-terminally GST-tagged truncated forms of APL, with that of full-length APL provided as a reference. The indicated regions of APL were amplified by PCR from Ax2 cDNA, and ligated into pGEX-4T-2 vectors. These vectors were transfected into BL21(DE3) *E. coli* cells, and were expressed with N-terminal GST-tags, before purification. **B:** The elution fractions from the GST purifications performed in **A** were analysed using SDS-PAGE and Coomassie staining. **C:** Determination of the PAR-binding abilities of the truncated forms of APL. Increasing concentrations of each peptide were dotted onto a nitrocellulose membrane and incubated with PAR chains, followed by detection by Western blotting. Experiments performed by Peggy Paschke and Mehera Emrich.

DNA sequence of the desired domain-deletion was amplified by PCR, and ligated into a pGEX-4T-2 vector. Following transfection of this vector into BL21(DE3) *E. coli*, the GST-tagged recombinant proteins were purified by GST purification. The purity of the samples was analysed by SDS-PAGE and Coomassie staining (Figure 5.3B). The PAR-binding nature domain-deletion mutants of APL were then assessed by the same dot-blotting assay as used previously (Figure 5.3C). Qualitatively similar levels of PAR-binding were observed between full-length APL, APL lacking its FHA domain (GST-APL-ΔFHA) and APL lacking just its macro domain (GST-APL-ΔMACRO). Deletion of both macro and PBZ domains (GST-APL-ΔPBZΔMACRO) reduces the PAR-binding of this peptide compared to the other mutants. These data indicate that, while both the PBZ and macro domains have PAR-binding activity, the PBZ domain may be the dominant PAR-binding domain of APL. Design of the pGEX-4T-2 vectors, GST-protein purification, and dot-blots were performed by Peggy Paschke and Mehera Emrich.

5.2.2. Generation and verification of an *apl*$^-$ strain

Following the confirmation that APL is a PAR-binding protein *in vitro*, and thereby validating the *in silico* screen, we wished to ascertain whether or not APL functions in the DDR of *Dictyostelium*. To do this, we assessed whether *Dictyostelium* cells lacking APL displayed any phenotype consistent with the protein acting in the DDR. The first approach that we took to identify the *in vivo* function of APL was to assess the sensitivity of *apl*$^-$ strains to a variety of DNA damaging agents, chosen to activate a range of DNA repair pathways. Therefore, we generated a *Dictyostelium* strain disrupted in the *apl* gene. The strategy employed for this disruption by targeted homologous recombination is shown in Figure 5.4A, where the primer sets indicate the regions of the *Dictyostelium* genome upstream and downstream of the *apl* gene that are used to target the knock-out construct.

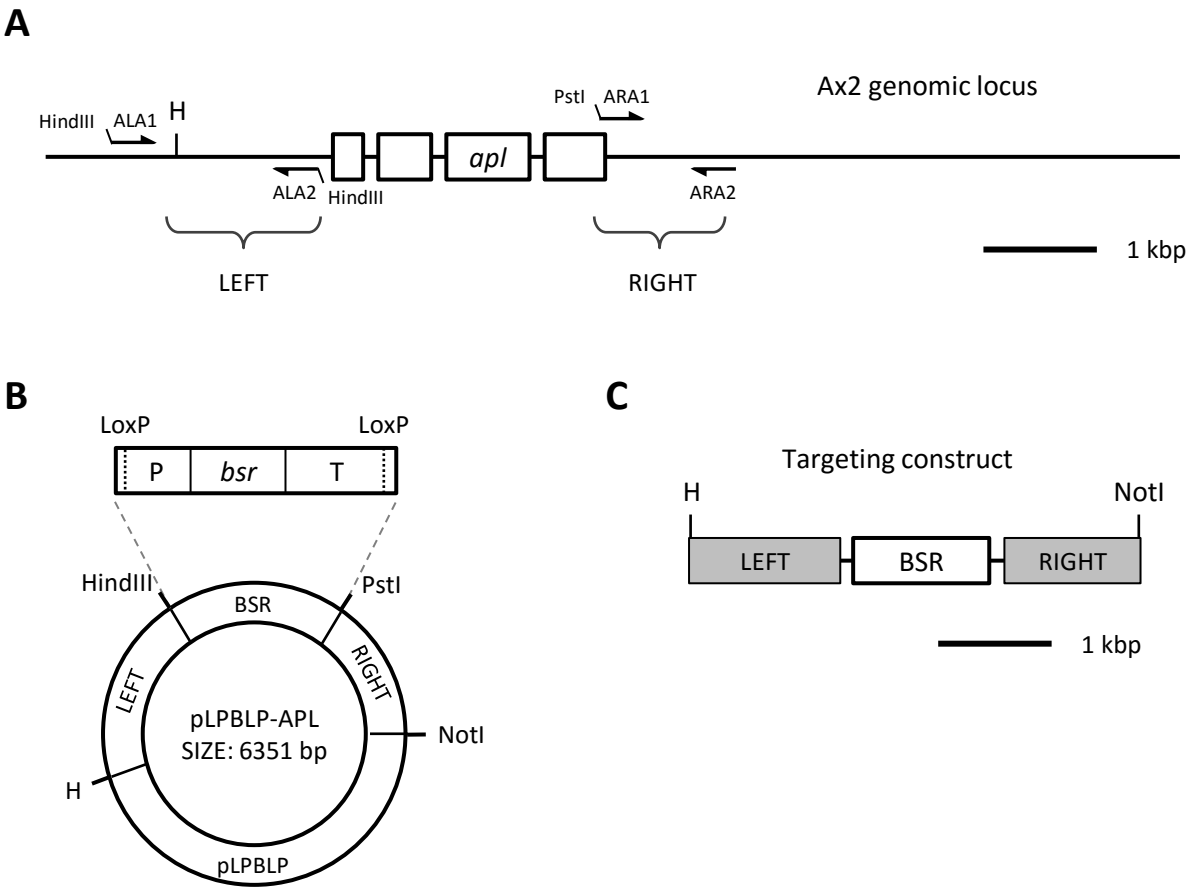
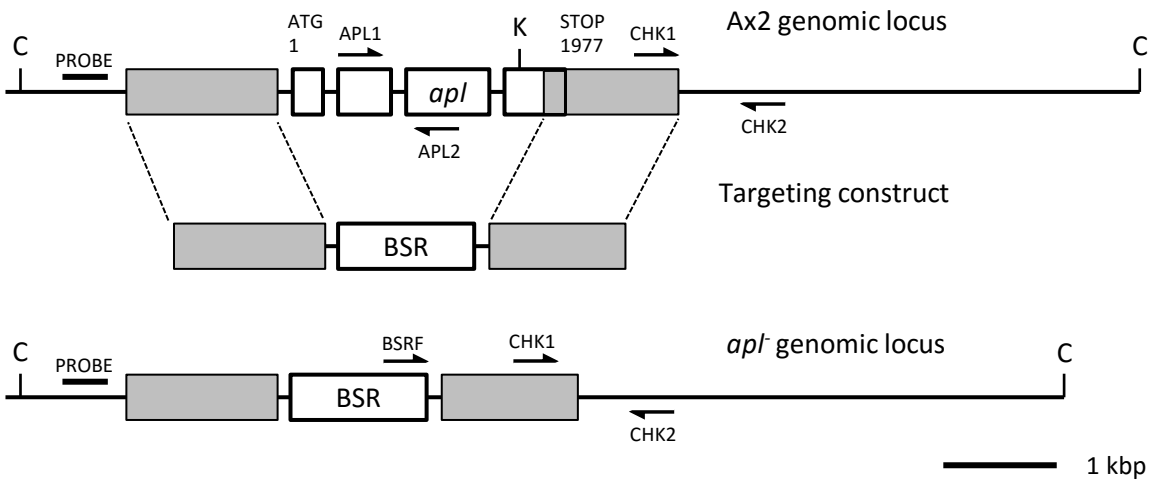


Figure 5.4 – Knock-out construct generation for *apl*$^-$ cells. A: The structure of the *apl* gene, displaying its four exons (solid boxes). Primer locations for the two homologous arms (LEFT and RIGHT) used for *apl* deletion are indicated. The slight overlap of the RIGHT homologous arm with the 3' end of the *apl* gene was necessary due to the AT-rich nature of the genome downstream of the *apl* gene. Restriction sites used for the cloning procedure are shown. The endogenous HpaI (H) restriction site is annotated. **B:** Map of the pLPBLP-APL vector used for making an *apl*$^-$ *Dictyostelium* strain. The homologous arms are ligated into this vector flanking a cassette encoding a gene for resistance to blasticidin *(bsr*), an antibiotic normally toxic to *Dictyostelium*. The indicated restriction sites for excision of the knock-out construct, HpaI (H) and NotI, are shown. The NotI site is present in the multiple cloning site of the pLPBLP vector. The map of the *bsr* cassette is also shown, indicating its promotor (P), terminator (T), and LoxP sites for excision of the cassette by Cre recombinase. **C:** The useful product of digesting pLPBLP-APL with HpaI and NotI, which can be used to knock-out the *apl* gene by targeted homologous recombination. Strategy and knock-out construction performed by Duen-Wei Hsu.

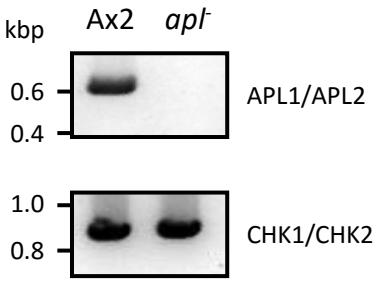
The two homologous arms were amplified by PCR and ligated into the pLPBLP vector, either side of a blasticidin resistance cassette (Figure 5.4B). The blasticidin resistance cassette is formed of a gene for blasticidin resistance (*bsr*), flanked by an Actin promotor and terminator. LoxP sites are placed upstream and downstream of the cassette to enable it to be excised from the *Dictyostelium* genome by the action of Cre recombinase, if desired. Restriction digestion with HpaI (H) and NotI remove the *apl k*nock-out construct from the pLPBLP backbone, for transfection into *Dictyostelium* Ax2 cells (Figure 5.4C). The *apl* knock-out construct was transfected into Ax2 cells by a standard protocol, and cells were left to grow in blasticidin-containing media until cell growth was directly observable, indicating resistance to blasticidin attained by integration of the construct.

The generation of blasticidin-resistant *Dictyostelium* cells could be achieved by the intended recombination event (Figure 5.5A), or by random integration of the construct. Therefore, we verified the *apl*$^{-}$ strains at the genetic level by PCR, RT-PCR and Southern blot. The location of the primers used for PCR and RT-PCR, in both Ax2 and *apl*$^{-}$ genetic backgrounds, are shown in Figure 5.5A. The verification of *apl*$^{-}$ strains by PCR was performed by two combinations of primers: APL1/APL2, in a region of the *apl* gene deleted by the intended recombination event, and BSRF/CHK2, which amplifies a genetic sequence unique to correctly generated *apl*$^{-}$ cells. The results of both of these PCRs are shown in Figure 5.5B-C, and they indicate the successful generation of an *apl*$^{-}$ clone. The conclusion of the PCR verification was supported by a Southern blot. The ClaI (C) and KpnI (K) restriction sites used for the digestion of Ax2 and *apl*$^{-}$ genomic DNA are shown in Figure 5.5A, alongside the location of the probe. The strategy was designed so that deletion of the KpnI site by the recombination event leads to an increase in the length of the DNA fragment bound by the probe in *apl*$^{-}$ cells. The result of the Southern blot indicated that the putative *apl*$^{-}$ strain had undergone the correct recombination event

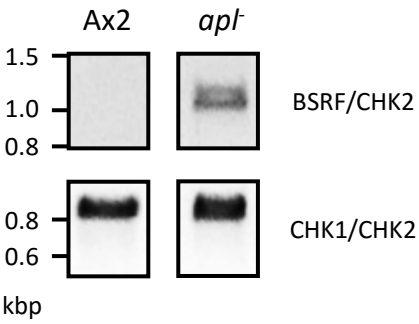
A



B

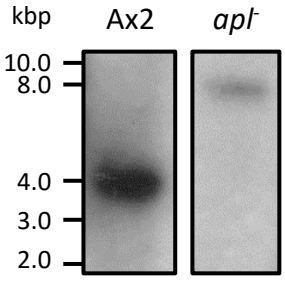


C



Primers	Expected PCR Products (Genomic DNA; bp)	
	Ax2	*apl*-
APL1/APL2	618	N/A
BSRF/CHK2	N/A	1049
CHK1/CHK2	849	849

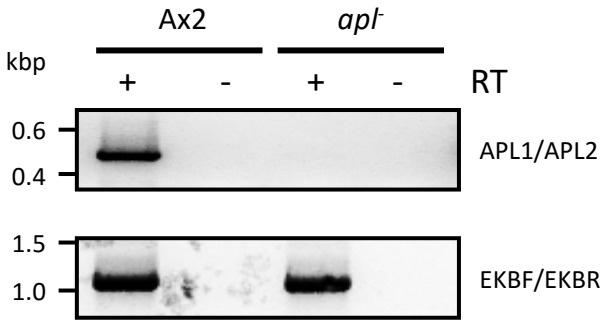
D



KpnI (K)/ClaI (C)

Strain	Expected Fragments (bp)
Ax2	4020
apl-	9772

E



Primers	Expected PCR Products (cDNA; bp)	
	Ax2	*apl*-
APL1/APL2	516	N/A
EKBF/EKBR	1110	1110

Figure 5.5 – Verification of *apl*$^-$ knock-out strains. A: The strategy for *apl* deletion by targeted homologous recombination, displaying the intended recombination event at the *apl* genetic locus (dotted lines). The expected integration of the blasticidin resistance cassette (BSR) and alteration of the Ax2 genome is also shown. Primers used for verification of the *apl*$^-$ strains are indicated, as are the location of the probe and restriction sites for Southern blot analysis (ClaI (C) and KpnI (K)). **B:** Verification of the *apl*$^-$ strain by PCR. The APL1/APL2 primer combination shown in **A** was used to amplify genomic DNA extracted from AX2 and an *apl*$^-$ strain. **C:** Further PCR verification of the *apl*$^-$ strain, using BSRF/CHK2 primer combination shown in **A**. The primers CHK1 and CHK2 were used as a positive control in both **B** and **C**. **D:** Southern blot verification of the *apl*$^-$ strain. Parental Ax2 and putative *apl*$^-$ genomic DNA was digested by the indicated restriction enzymes, with a noticeable size-shift introduced by the deletion of a KpnI site present in the *apl* gene. **E:** Verification of the *apl*$^-$ strain by RT-PCR. The APL1 and APL2 primers shown in **A** were used to amplify cDNA synthesised from RNA extracted from Ax2 and a putative *apl*$^-$ strain. Two primers in the *erkB* housekeeping gene, EKBF and EKBR, were used as a positive control. For **B-E**, the expected fragment sizes are shown in the respective tables.

(Figure 5.5D). Furthermore, the transcription of the *apl* gene was assessed in Ax2 and *apl*$^-$ cells by RT-PCR, with the APL1/APL2 primer combination. The cDNA generated from *apl*$^-$ cells indicated that APL is no longer being transcribed (Figure 5.5E). Cumulatively, these data indicate the successful generation of an *apl*$^-$ strain.

To further verify this strain, at the protein level, we raised an antibody against the macro domain region of APL, with an N-terminal 6xHis tag (Figure 5.6A; ΔAPL; purified protein kindly provided by Ivan Ahel). The antibody was raised in sheep, following a procedure involving multiple injections of the ΔAPL antigen and subsequent bleeds (outlined in the Materials and Methods section). The bleeds were provided as crude serum, which we assayed for anti-APL reactivity. Additionally, we were sent a pre-immunisation (PI) bleed to use as a negative control. Initial tests of the crude antibody serum by SDS-PAGE and Western blotting indicated a strong increase in the affinity for the antigen between the PI and first bleed, which was further increased following the second antibody injection and bleed (Figure 5.6B). Subsequent injections did not further increase the affinity of the crude serum for the antigen, so we proceeded to affinity purify the serum from the second bleed. Affinity purification selects only the antibodies from the serum whose epitope is the macro domain region of APL, which is a small subset of the total proteins in the serum. As the antigen for raising the antibody was N-terminally 6xHis-tagged, we performed affinity purification with GST-APL, so as to remove any proteins with affinity for the 6xHis-tag region of the antigen (Figure 5.6C). The affinity purification procedure was monitored by SDS-PAGE and Coomassie staining of samples at each stage of the purification, with the final elution containing only the denatured heavy and light chain antibody proteins (Figure 5.6D). Comparison of the antigen-binding abilities of the crude serum and purified anti-APL (αAPL) antibody by Western blotting indicated a clear increase in affinity following the purification process (Figure 5.6E). To further assess the function of the anti-APL

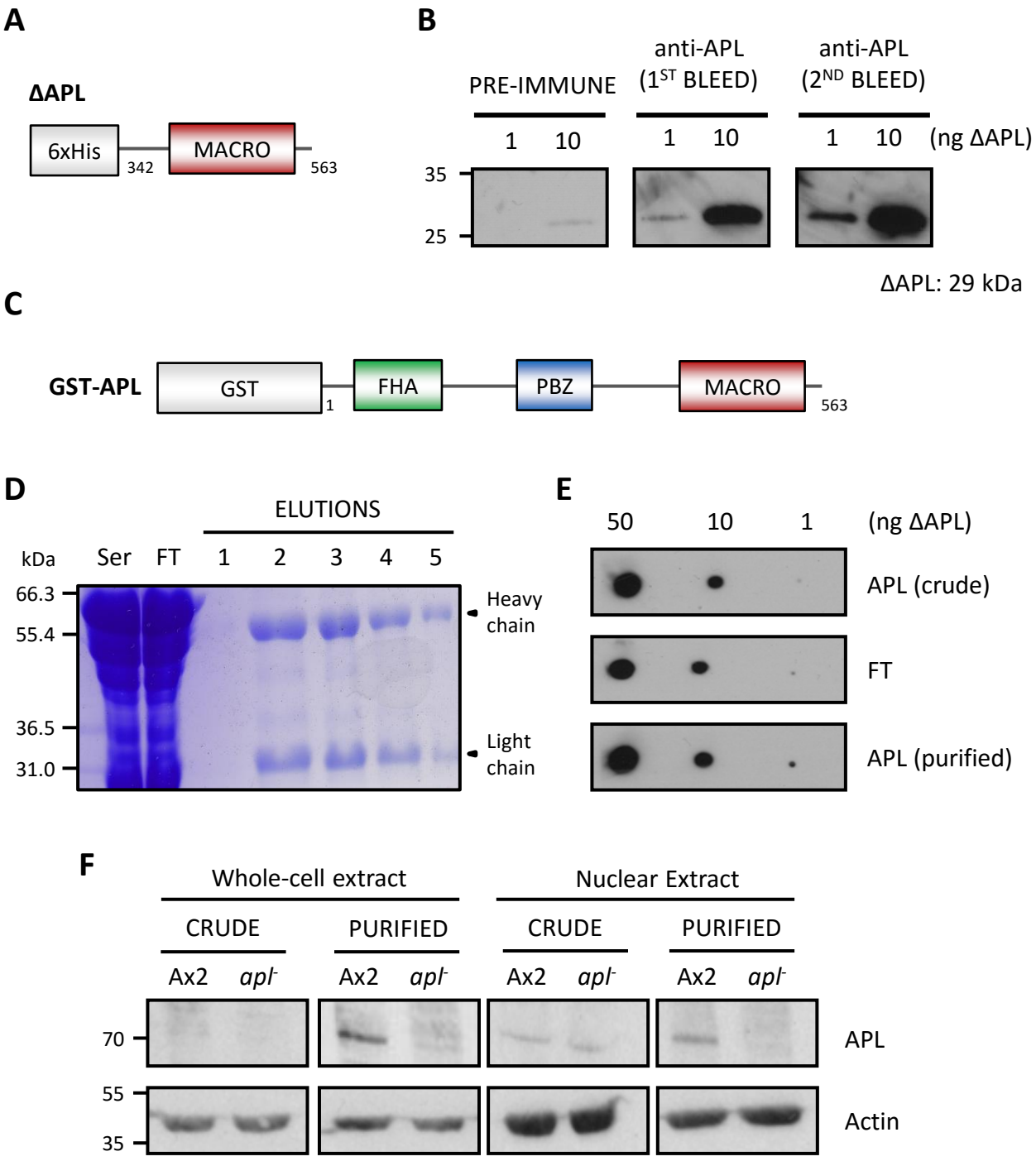

A
ΔAPL
6xHis
342
MACRO
563
B
PRE-IMMUNE
anti-APL (1ST BLEED)
anti-APL (2ND BLEED)
1 10 1 10 1 10 (ng ΔAPL)
35
25
ΔAPL: 29 kDa
C
GST-APL
GST
1
FHA
PBZ
MACRO
563
D
ELUTIONS
kDa Ser FT 1 2 3 4 5
66.3
55.4
36.5
31.0
Heavy chain
Light chain
E
50 10 1 (ng ΔAPL)
APL (crude)
FT
APL (purified)
F
Whole-cell extract
Nuclear Extract
CRUDE
PURIFIED
CRUDE
PURIFIED
Ax2 apl- Ax2 apl- Ax2 apl- Ax2 apl-
70
55
35
APL
Actin

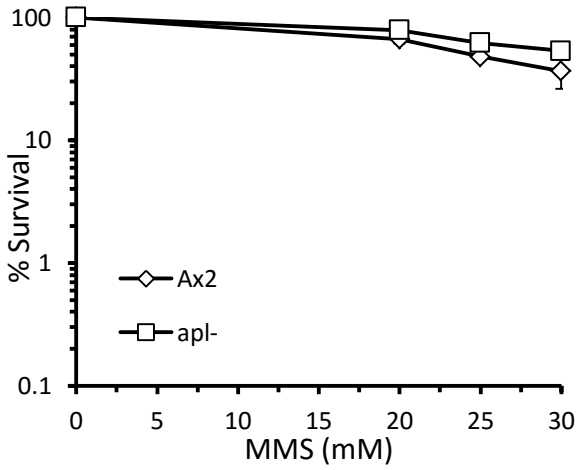
Figure 5.6 – Raising an antibody against endogenous APL. A: Domain structure of the 6xHis-tagged macro domain region of APL, named ΔAPL, which was used as an antigen for raising an anti-APL antibody. ΔAPL was purified and kindly sent to us by Ivan Ahel. **B:** Assessing the affinity of the crude antibody bleeds. The indicated amounts of the antigen, ΔAPL, were run on a gel by SDS-PAGE, and Western blotting was performed with crude sera from the different anti-APL bleeds. The pre-immune serum, taken before antigen injection, provides a negative control. **C:** Domain structure of the GST-APL construct used for affinity purification of the second bleed anti-APL serum. **D:** Monitoring the process of affinity purification of the anti-APL antibody. Samples of serum were taken at each stage of the purification procedure and separated by SDS-PAGE. Detection was performed with Coomassie staining. The two bands in the elution fractions (arrows), represent the heavy and light antibody chains. Abbreviations: diluted crude serum (Ser), flow-through (FT). **E:** The effect of the affinity purification procedure on the anti-APL antibody was first determined by observing the affinity of the purified antibody to the antigen. Increasing amounts of ΔAPL were dot-blotted onto nitrocellulose membrane, and Western blotting was performed with the indicated fraction. **F:** More rigorous comparison of the crude serum and purified anti-APL antibody. Whole-cell and nuclear Ax2 and *apl*$^-$ extracts were prepared and separated by SDS-PAGE. Detection by Western blotting with either the crude serum (CRUDE) or purified anti-APL antibody (PURIFIED) was then performed. The absence of a band in the *apl*$^-$ extracts also serves to further verify the strain.

antibody, *Dictyostelium* Ax2 and *apl*$^-$ whole-cell and nuclear extracts were prepared and analysed by SDS-PAGE and Western blotting (Figure 5.6F). Following detection, Ax2 whole-cell and nuclear extracts both produced a clear band at approximately 70 kDa (APL has a molecular weight of 64.3 kDa). Moreover, this band was not present in the *apl*$^-$ strain, providing further verification of this strain. Probing the extracts with both crude serum and purified anti-APL antibody again highlights the clear increase in affinity gained through the purification procedure.

5.2.3. Vegetative *Dictyostelium apl*$^-$ cells are not sensitive to DNA damage

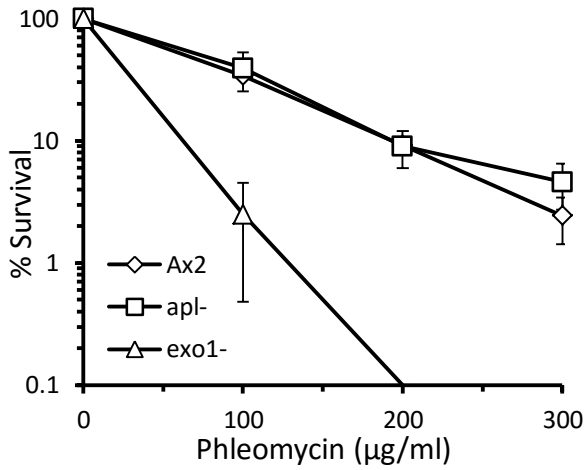
With an *apl*$^-$ strain generated and fully verified, we wished to determine whether or not these cells were sensitive to DNA damage. DDR protein deletion in *Dictyostelium* can lead to noticeable growth phenotypes under standard growth conditions. However, initial observations of growth rates in shaking suspension and plaque formation on agar showed no difference in comparison to parental cells. To analyse the correct functioning of specific DNA repair pathways, we subjected *apl*$^-$ cells to treatment with various genotoxic chemicals, including: methyl methanesulphonate (MMS), phleomycin, the UV-mimetic 4-nitroquinoline-1-oxide (4-NQO), and cisplatin. These agents induce base methylation (base damage), DSBs, bulky adducts repaired by the NER pathway, and DNA cross-links (both intra- and inter-strand), respectively. In humans, the repair of each of these forms of damage, with the exception of cisplatin-induced damage, has been shown to have some level of dependence on ADP-ribosylation[17,171,459]. Following treatment, the cells were washed before being spread onto SM agar plates. Survival was measured by assessing plaque formation, compared to untreated control cells. The survival of the *apl*$^-$ strain to MMS was found to be no different to that shown by parental Ax2 cells (Figure 5.7A). Similarly, while *exo1*$^-$ cells displayed a strong survival defect following phleomycin

A



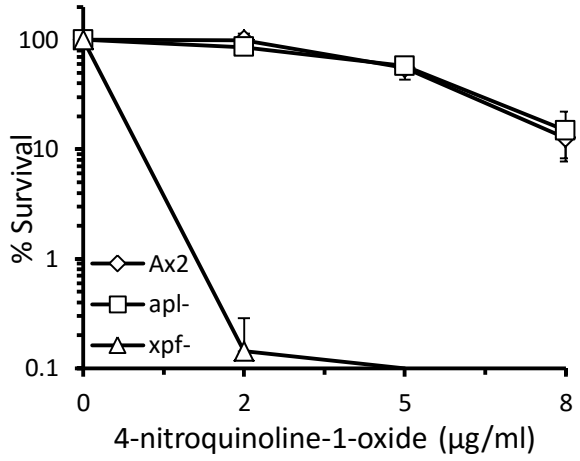
Strain	D_{50} (mM)	SE	P-value (Ax2)
Ax2	25.5	0.36	N/A
apl^{-}	29.7	1.22	0.08

B



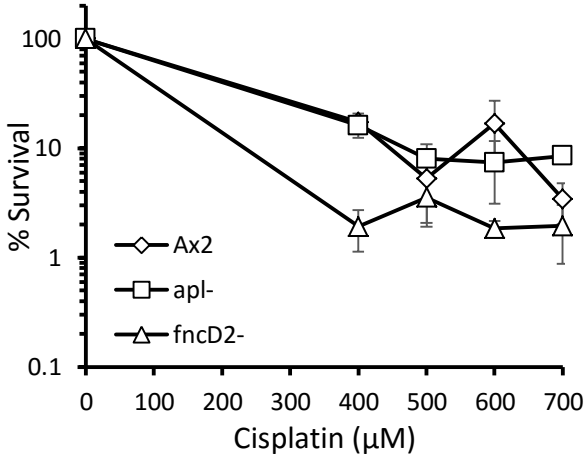
Strain	D_{50} (µg/ml)	SE	P-value (Ax2)
Ax2	80.3	3.75	N/A
apl^{-}	80.8	17.4	0.55
exo1^{-}	39.7	1.21	0.009 (**)

C



Strain	D_{50} (µg/ml)	SE	P-value (Ax2)
Ax2	5.59	0.97	N/A
apl^{-}	5.38	0.19	0.85
xpf^{-}	0.81	0.002	0.03 (*)

D



Strain	D_{50} (µM)	SE	P-value (Ax2)
Ax2	198	17.1	N/A
apl^{-}	189	3.04	0.64
fncD2^{-}	159	3.41	0.15

Figure 5.7 – The sensitivity of *Dictyostelium* Ax2 cells to DNA damaging agents is not increased in the absence of APL. Ax2, *apl*$^{-}$ and a relevant positive control were exposed to **A:** MMS, **B:** phleomycin, **C:** 4-NQO, for one hour, or **D:** cisplatin, for five hours. Following treatment and washing, 500 cells were plated out on SM agar in association with *Klebsiella aerogenes*, and plaque formation was monitored over seven days. Cellular survival is measured relative to an untreated control. For phleomycin treatment, *exo1*$^{-}$ cells were included as a positive control, whereas *xpf*$^{-}$ cells served this purpose for 4-NQO, and *fncD2*$^{-}$ cells did for cisplatin. The error bars represent the standard error of three independent experiments. D_{50} toxicity values for all strains are shown in the tables, with standard errors (SE). P-values on these are calculated with respect to the bracketed strain.

exposure, Ax2 and *apl*$^{-}$ cells displayed the same survival phenotypes (Figure 5.7B). As expected, the NER-deficient *xpf*$^{-}$ strain was hypersensitive to 4-NQO treatment; however, deletion of APL did not increase the sensitivity of *Dictyostelium* cells to this genotoxic agent (Figure 5.7C). Finally, the FA-deficient *fncD2*$^{-}$ strain showed slightly increased sensitivity to cisplatin compared to Ax2 cells, but the *apl*$^{-}$ strain did not (Figure 5.7D). These data indicate that APL is not a critical component of the DNA repair pathways probed, as its deletion does not appear to affect the overall survival of the cells. However, this does not rule out a role of APL in these DNA repair pathways, as it may function in a redundant manner. Furthermore, it is known that vegetative *Dictyostelium* cells are strongly reliant on HR for DSBR, and that deletion of core NHEJ components does not affect overall survival to phleomycin treatment[358]. Therefore, we proceeded to assess the role of APL in NHEJ in *Dictyostelium*.

5.2.4. *Dictyostelium apl*$^{-}$ germinating spores are not NHEJ-deficient

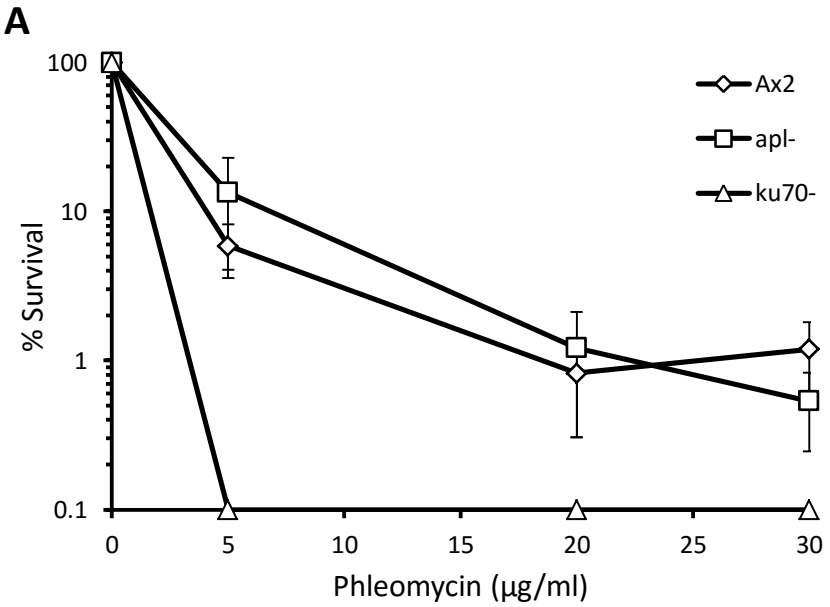
As previously mentioned, deletion of core *Dictyostelium* NHEJ proteins does not lead to a cellular sensitivity to DSBs in vegetative cell cultures. This is primarily thought to be due to the fact that the majority of *Dictyostelium* cells are in either S- or G2-phases of the cell cycle, where HR is the dominant DSBR pathway[330,460]. Therefore, if APL is a NHEJ factor, its action in DSBR will not be uncovered by assessing the survival of a vegetative culture of *apl*$^{-}$ *Dictyostelium* cells to DSBs. It is known that germinating spores of *Dictyostelium*, which are the dispersed from fruiting bodies in response to unfavourable conditions, are more reliant on NHEJ, despite also being in G2-phase[334,358]. To address the role of APL in NHEJ, we generated *Ax2*, *apl*$^{-}$ and *Ku70*$^{-}$ spores, induced their germination by gentle heating, and subsequently treated them with phleomycin for 16 hours. Following treatment, the cells were plated onto SM agar, and their survival was measured by plaque

formation (Figure 5.8A). In parallel, the germination rate of the spores was monitored by visualising the cells at a set number of time points. All of the strains assessed germinated at the same rate, and were fully germinated by the end of the treatment time (Figure 5.8B). The resulting plaque formation indicated the expected survival defect in the NHEJ-deficient *ku70*$^{-}$ strain; however, there was no enhanced sensitivity of *apl*$^{-}$ cells compared to Ax2. Therefore, APL does not appear to have a critical role in NHEJ.

5.2.5. APL is enriched on chromatin in response to DNA damage

Despite the lack of a survival defect of *apl*$^{-}$ cells to the DNA damaging agents tested so far, a role for APL in the DDR has not been ruled out. For example, the contribution of APL may be to increase the efficiency of specific DNA repair pathways, or the sensitivity increase provided by its deletion may be too small to be detected by the survival assays employed. Additionally, APL may act in a redundant manner with another protein, thereby masking its role in the DDR. In order to determine if APL has a more subtle role in the DDR, we chose to assess the enrichment of APL on chromatin following various forms of DNA damage. Such recruitment, if detected, would implicate a role for APL in the response to specific forms of DNA damage.

To monitor the enrichment of APL on chromatin, we chose to overexpress an N-terminally Myc-tagged version of APL (Myc-APL), shown in Figure 5.9A. This approach was chosen due to the relatively low expression of endogenous APL, and the lack of specificity of the anti-APL antibody to proteins on chromatin, the combined result of which led to APL being virtually indistinguishable from background bands (data not shown). The coding sequence of full-length APL was amplified from Ax2 cDNA, with primers designed so as to attach a Myc-tag to the N-terminus of the protein. This construct was ligated into the pDXA-3C vector, a *Dictyostelium* expression vector, under the control of an Actin15



Strain	D_{50} (µg/ml)	SE	P-value (Ax2)
Ax2	49.6	8.91	N/A
apl$^-$	41.8	1.49	0.48
ku70$^-$	38.2	0	0.33

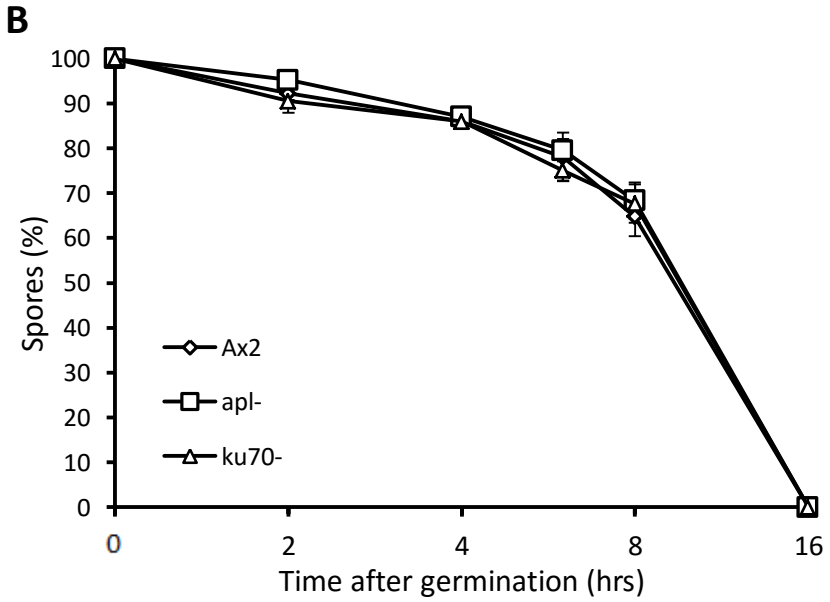


Figure 5.8 – Cells lacking APL do not display defective NHEJ. A: The survival of germinating *Dictyostelium* spores following exposure to phleomycin for 16 hours. Ax2, *apl*$^-$ and *ku70*$^-$ spores were grown and harvested using standard techniques. Germination was induced by gentle heating, and the spores were exposed to the indicated concentrations of phleomycin immediately following this. After 16 hours, the cells were washed and 500 cells were plated out for each condition. Plaque formation was monitored over the following six days. Survival was measured relative to an untreated control. The NHEJ-defective strain *ku70*$^-$ was included as a positive control. D_{50} toxicity values are provided in the table, with standard errors (SE). P-values are calculated with respect to the bracketed strain. **B:** The rate of spore germination of the strains used in **A**, as determined by direct observation. The error bars in both graphs represent the standard error of three independent experiments.

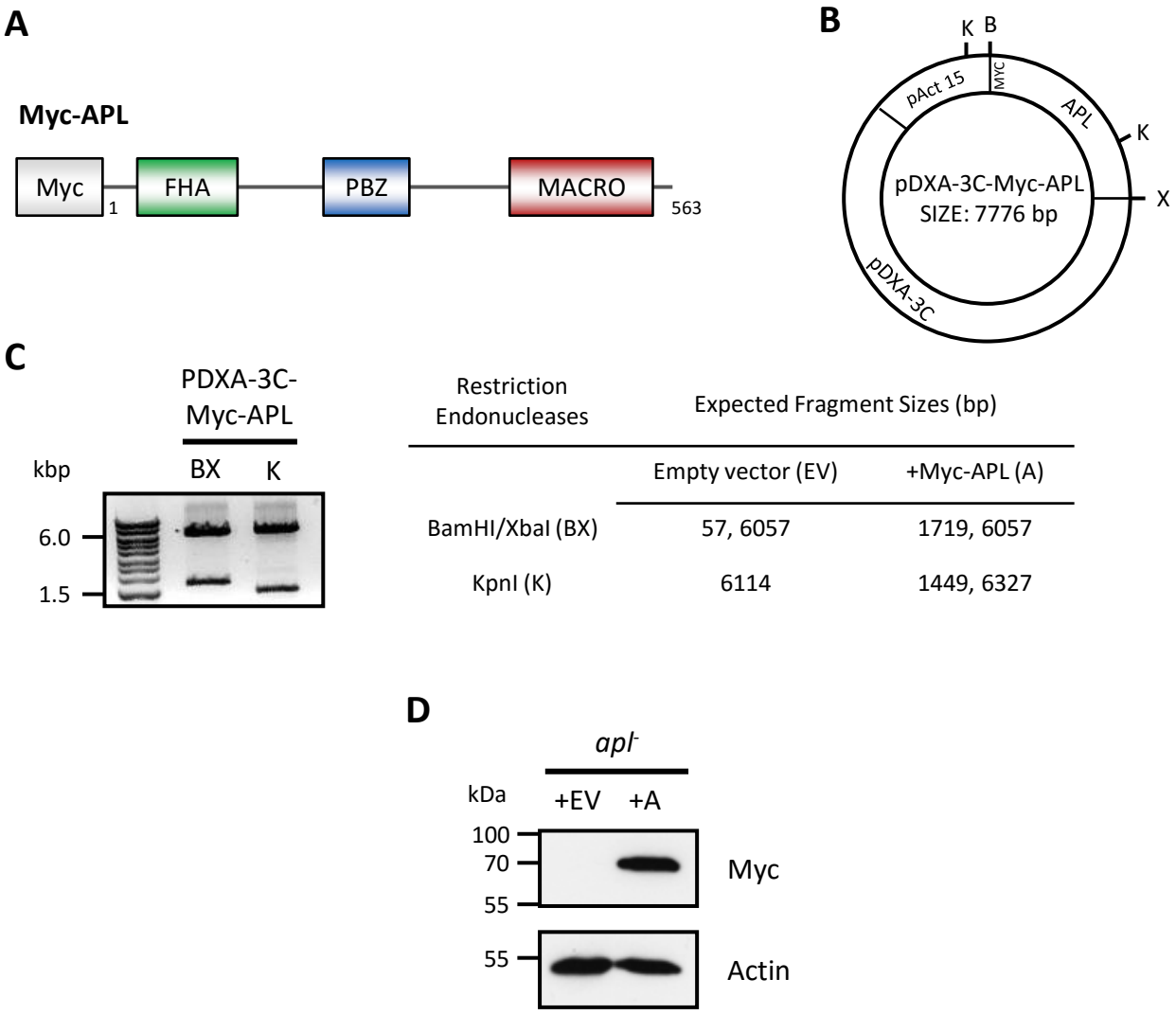


Figure 5.9 – Construction of a vector for overexpression of Myc-tagged APL. A: Domain structure of the N-terminally Myc-tagged form of APL (Myc-APL) to be expressed in *Dictyostelium* cells. **B:** Map of the plasmid resulting from the ligation of the pDXA-3C *Dictyostelium* overexpression vector and Myc-APL cDNA. This vector places Myc-APL under the control of an Actin15 promoter (pAct 15), and also encodes a gene for resistance to G418, an antibiotic normally toxic to *Dictyostelium*. Digestion with BamHI (B) and XbaI (X) excise the Myc-APL cDNA fragment, whereas the KpnI sites (K) are used for plasmid verification. **C:** Verification of the pDXA-3C-Myc-APL vector by restriction digestion, with the indicated enzymes. The expected fragment sizes for each digestion are provided in the table. **D:** Following transfection of pDXA-3C-Myc-APL into *apl*$^-$ cells, expression of Myc-APL was verified by Western blotting. Whole-cell extracts of *apl*$^-$ cells transfected with either empty pDXA-3C (+EV) or pDXA-3C-Myc-APL (+A) were separated by SDS-PAGE and Western blotting with an anti-Myc antibody was performed. The expected size of Myc-APL is 65.5 kDa.

promoter (Figure 5.9B). Additionally, pDXA-3C encodes a gene for G418 resistance, thereby providing a means for selection of *Dictyostelium* cells post-transfection. The pDXA-3C-Myc-APL vector was verified by restriction digestion and Sanger sequencing (Figure 5.9C). Following transfection of *apl*$^-$ cells with either empty pDXA-3C or pDXA-3C-Myc-APL, expression was verified by Western blotting with an anti-Myc antibody (Figure 5.9D).

We exposed Myc-APL-expressing *apl*$^-$ cells to a range of DNA damaging agents: MMS, phleomycin, 4-NQO, cisplatin and camptothecin (to induce DSBs in S-phase only), and then extracted chromatin from these cells. The level of Myc-APL enriched with chromatin in these cells was then determined by SDS-PAGE and Western blotting (Figure 5.10). We monitored and observed the induction of phosphorylated histone H2AX (γH2AX) in response to all types of DNA damage, indicating that all of these genotoxic agents cause DNA damage in *Dictyostelium*. The level of Myc-APL associated with chromatin was not increased following treatment with MMS or 4-NQO, indicating that APL is not involved in the response to the forms of DNA damage induced by these agents. However, we observed strong enrichment of Myc-APL on chromatin following induction of cross-links (cisplatin) and replication-induced DSBs (camptothecin). We also observed weaker enrichment of Myc-APL in response to cell-cycle independent DSBs (phleomycin). This implies that APL has some involvement in the DDR in *Dictyostelium*, particularly in response to forms of damage suffered or detected in S-phase.

5.2.6. Chromatin-association of APL is dependent on its macro domain

We have identified that APL is enriched on chromatin following the induction of DNA cross-links or replication-associated damage. However, we do not yet know which domains of APL are responsible for this enrichment. Identifying which domains are

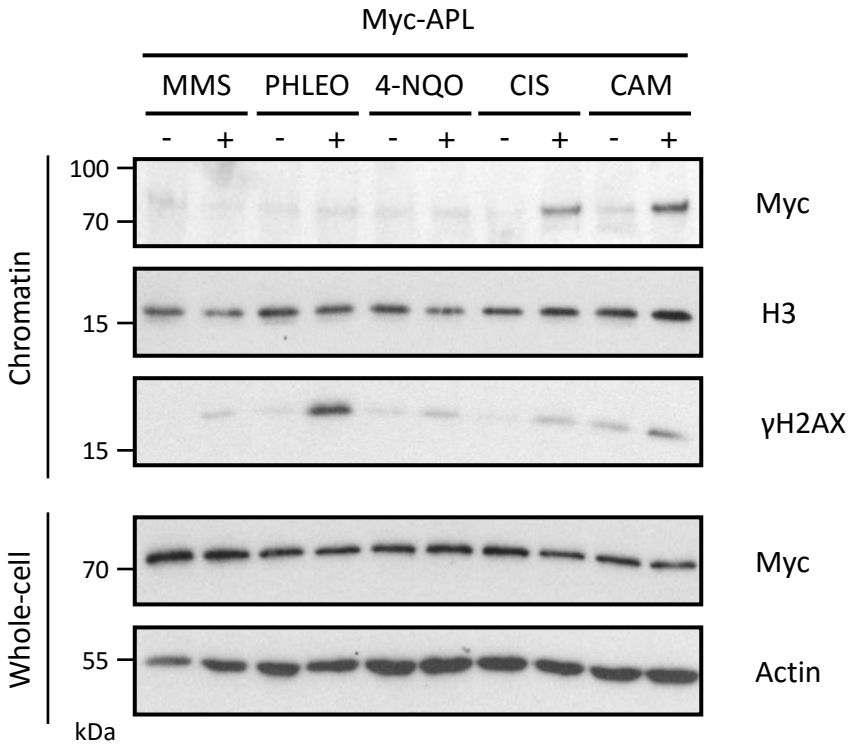


Figure 5.10 – Myc-APL is enriched on chromatin in response to some forms of DNA damage. *Dictyostelium apl*$^{-}$ cells overexpressing Myc-APL were treated either 5 mM MMS (1 hour), 300 μg/ml phleomycin (1 hour; PHLEO), 5 μg/ml 4-NQO (1 hour), 300 μM cisplatin (5 hours; CIS) or 50 μM camptothecin (3 hours; CAM), alongside mock-treated controls. Following incubation, chromatin and whole-cell extracts were prepared and separated by SDS-PAGE. Detection was performed by Western blotting with the indicated antibodies. Histone H3 is a loading control for chromatin, while γH2AX serves as a marker of DNA damage.

responsible for the recruitment will elucidate the mechanism of function of APL. The PBZ or macro domains could mediate PAR-dependent interactions with proteins assembled or present at DNA lesions, as they have both been shown to be PAR-binding *in vitro*, or the FHA domain may interact with other DDR proteins phosphorylated as a result of the DNA damage. To assess this, we introduced a series of mutations into the pDXA-3C-Myc-APL vector in order to mutate or delete the domains of the full-length protein: APL-C174AC180A (PBZ mutant; P*), APL-Δ342-563 (macro deletion, ΔM), APL-E439A (predicted macro mutant; M*), and APL-C174AC180AE439A (PBZ and macro mutant; PM*). The glutamic acid at position 439 was identified in the predicted structure of APL as bearing positional similarity to an acidic residue in MacroD1 responsible for ligand binding (Figure 3.11). The point mutations were introduced by primers encoding those mutations, and were amplified by PCR from the pDXA-3C-Myc-APL vector. Following verification of the mutations by Sanger sequencing, the mutated DNA fragments were ligated into the pDXA-3C-Myc-APL vector, following excision of the original APL DNA sequence by restriction digestion. The new pDXA-3C vectors were subsequently transfected into *apl*$^{-}$ *Dictyostelium* cells, using standard techniques. Generation of pDXA-3C-Myc-APL vectors, and transfections, were performed by Peggy Paschke.

We then exposed these *apl*$^{-}$ cells expressing mutant forms of Myc-APL to cisplatin for five hours, or mock-treated, and subsequently generated chromatin extracts. Analysis of these extracts was performed by SDS-PAGE and Western blotting (Figure 5.11). As observed previously, induction of γH2AX indicates the efficacy of cisplatin in this organism, and we also observed enrichment of full-length APL on chromatin in response to cisplatin treatment. The P*, M* and PM* mutations appear to decrease the enrichment of APL on chromatin following damage, to levels equivalent to untreated cells. Interestingly, deletion of the macro domain of APL disrupts the constitutive association of Myc-APL on

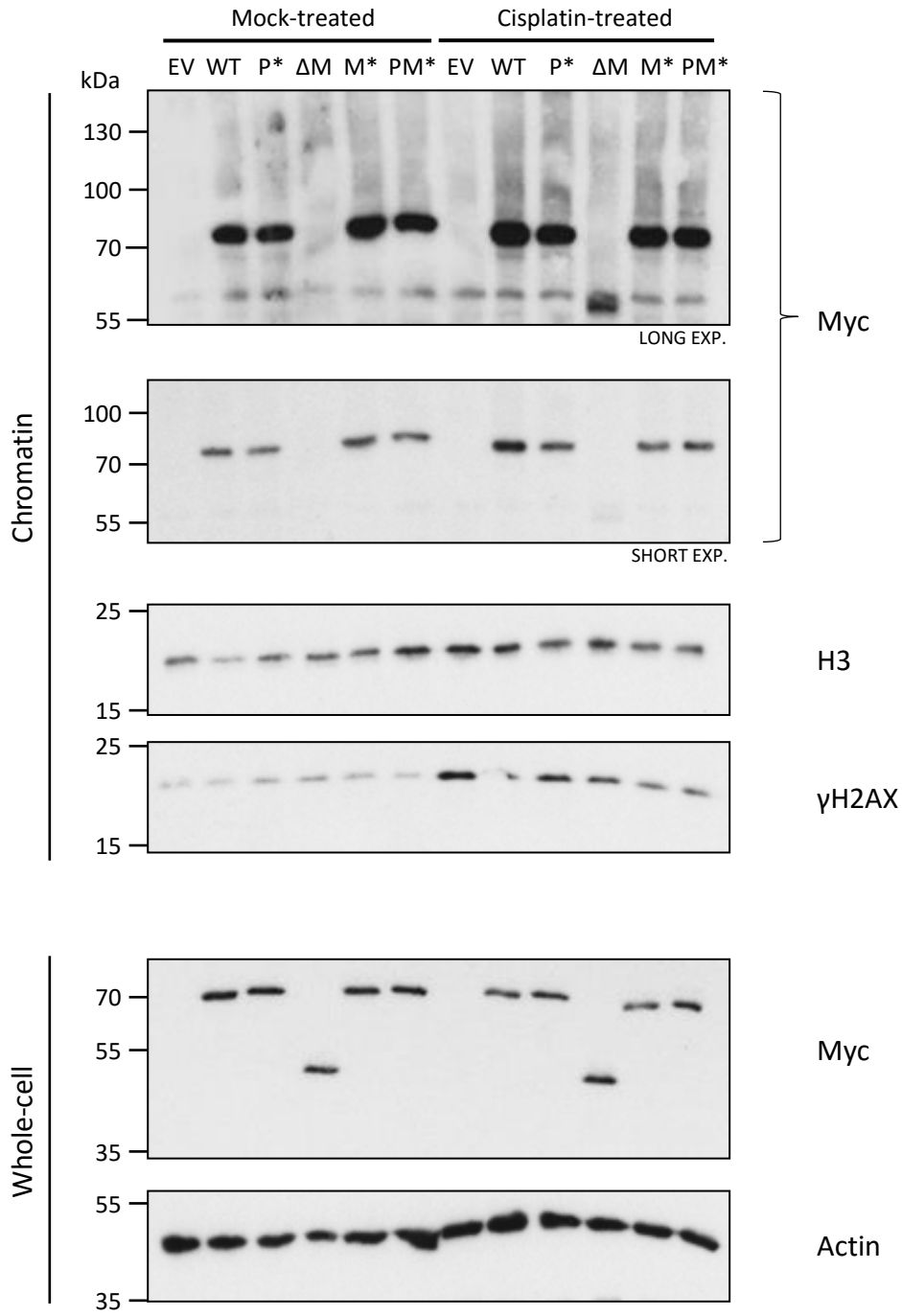


Figure 5.11 – The macro domain of APL is required for the association of APL with chromatin. *Dictyostelium apl*$^{-}$ cells expressing recombinant, mutant forms of Myc-tagged APL were treated with 300 μM cisplatin for 5 hours, or mock treated (empty vector (EV), wild-type APL (WT), APL-C174AC180 (PBZ mutant; P*), APL-ΔMACRO (ΔM), APL-E439A (macro mutant; M*), APL-C174AC180AE439A (PBZ and macro mutant; PM*)). Following incubation, chromatin and whole-cell extracts were prepared and separated by SDS-PAGE. Detection was performed by Western blotting with the indicated antibodies. Histone H3 is a loading control for chromatin, while γH2AX serves as a marker of DNA damage.

chromatin, but did not abrogate the damage-induced enrichment of the protein (Figure 5.11, top panel). Therefore, we have identified a constitutive association of Myc-APL with chromatin, mediated by the macro domain region, and a damage-induced enrichment, which is abrogated following mutation of the PBZ domain or macro domain.

5.3. Discussion

In this chapter, we performed initial characterisation of the *Dictyostelium* protein APL, which was the most likely candidate for a novel DDR protein identified through our *in silico* screen. This screen was designed to find proteins with previously unannotated PAR-binding domains, so we firstly investigated the PAR-binding nature of APL, and which domains are required for this interaction. We found that APL does bind to PAR chains *in vitro*, albeit not as strongly as the PBZ-domain-containing C-terminus region of *Dictyostelium* Ku70. Testing the individual domains of APL showed that both the PBZ domain and permuted macro domain were able to bind to PAR, indicating that the macro domain may have retained its function following the circular permutation. This supports the structural prediction of the macro domain of APL, which predicts that the binding pocket is still present. Moreover, the identification of the macro domain as a PAR-binding domain validates our screen, as this domain was previously unannotated. However, deletion of the macro domain from APL does not appear to reduce the PAR-binding ability of the protein, thereby suggesting that the PBZ domain is primarily responsible for the interaction of APL with PAR chains *in vitro*. The FHA domain of APL, which displays strong homology to the FHA domains of human APTX, APLF and PNKP, does not appear to bind PAR chains. Interestingly, despite the similarities in binding partners of the FHA domains of human APTX and APLF, they have been shown to display different PAR-binding behaviour: APTX does bind whereas APLF does not[308]. This may indicate that the

FHA domain of APL is more similar to that of APLF than APL. However, the role of the FHA domain in mediating protein-protein interaction between APL and its partners will be far more informative as to its cellular role.

Following the discovery that APL is a PAR-binding protein, we performed assays to determine whether or not APL is a DDR protein. Firstly, we observed the cellular survival of *apl*$^-$ cells to various forms of DNA damage: SSBs, DSBs, bulky adducts, and cross-links, but found no defect in comparison to parental Ax2 cells. Specifically with regards to *Dictyostelium*, deletion of NHEJ factors does not lead to a survival defect of vegetative cells to DSB-inducing agents[358]. Such a defect can be observed in germinating spores; however, *apl*$^-$ germinating spores did not show this survival phenotype when compared to Ax2 cells either. The lack of a survival defect indicates that APL is not a key protein in the DNA repair pathways associated with the repair of these types of damage, but does not rule out a role for APL in the DDR. APL may act to improve the efficiency of repair, rather than be necessary for its completion. Moreover, APL may also act in a redundant manner with another protein. In this case, a survival defect might only be uncovered in a double mutant with the redundant protein.

A potential role of APL in the DDR in *Dictyostelium* was uncovered when we assessed the level of Myc-APL on chromatin following multiple forms of DNA damage. We observed the enrichment of Myc-APL on chromatin following treatment with cisplatin and camptothecin. Cisplatin primarily causes DNA intra-strand cross-links, but also causes DNA inter-strand cross-links, whereas camptothecin traps type I DNA topoisomerases on DNA following cleavage, resulting in the formation of DNA DSBs when encountered by replication forks[461,462]. Therefore, camptothecin is a source of DSBs in S-phase only[463]. It is likely that the inter-strand crosslinks induced by cisplatin are responsible for APL enrichment on chromatin, as exposure to the UV-mimetic drug 4-NQO does not induce

this, and both intra-strand cross-links and the DNA lesions inflicted by 4-NQO are repaired by the NER pathway[464]. The detection of inter-strand crosslinks is thought to be coupled to stalled replication forks in S-phase; therefore, these data indicate that APL is involved in the response to S-phase associated damage[465].

Following the observation of the enrichment of APL on chromatin in response to DNA cross-links and S-phase associated DSBs, we identified that the macro domain of APL is responsible for a constitutive association of APL with chromatin. Furthermore, mutation of the PBZ (C174AC180A) or macro domain (E439A) appeared to attenuate the cisplatin-induced enrichment of APL on chromatin. This suggests a role for PARylation in the recruitment of APL to damage sites, as the PBZ domain, in particular, is a PAR-binding module[241]. PARylation by ARTD1 and ARTD2 has previously been implicated in the restart of stalled replication forks and the resolution of replication-associated DNA damage; however, this has not yet been demonstrated in *Dictyostelium*[16]. Interestingly, ADP-ribosylation has not been identified as a component of DNA cross-link repair pathways in humans. The human nuclease SNM1A has a PBZ domain, and has been shown to be involved in the resolution of DNA inter-strand cross-links; however, the function of its PBZ domain has not been established in this process[466]. The constitutive association of APL with chromatin could be an artefact of expressing a recombinant protein at a higher level than the endogenous APL, or could be the recruitment of APL to the background level of DNA lesions present in the cell. This second suggestion is unlikely, as cisplatin-induced enrichment of Myc-APL-Δ342-563 (ΔM) is still observed. However, these data indicate that the macro domain is responsible for a function of APL, perhaps interacting with progressing replication forks. This *in vivo* role further supports the conclusion of our predictive modelling, which indicated that the circular permutation of the macro domain did not cause substantial structural alteration to the domain that may

have disrupted its function. Additionally, we also identified the enrichment of higher molecular weight forms of Myc-APL, indicating that the protein may undergo post-translational modifications, such as poly-ubiquitination or poly-SUMOylation, which have both been implicated in various DNA repair pathways and will be discussed in the next chapter of this thesis[467].

Further work is required to determine the precise function of APL in the DDR, but potential roles could be in the repair of S-phase associated DSBs by HR, which also arise during inter-strand crosslink repair or in TLS[468]. This may explain why no sensitivity of *apl*$^-$ cells is detected, as only 5-10% of vegetative cells are in S-phase[330]. In humans, the macroPARP ARTD8 has recently been implicated in facilitating DSBR by HR in response to replication stress. ARTD8 is a MART that ADP-ribosylates Rad51, which is required for its efficient turnover. Furthermore, the second macro domain of ARTD8 binds to MARylated Rad51[35]. We have not yet fully characterised the binding behaviour of the macro domain of APL, which bears strong homology to the first and second macro domains of ARTD8, and it may also bind to single ADP-ribose units. A putative *Dictyostelium* ARTD8 homolog, pARTg, has been identified, but not yet experimentally assessed for its activity[321]. Further characterisation of the properties of the macro domain of APL will be important for constructing the mechanism of action of this protein, particularly given its role in the association of APL with chromatin. ARTD10, another human MART, has also been implicated in lesion bypass in S-phase[37]. Therefore, the role of ADP-ribosylation is S-phase-associated DNA repair in humans has been shown, and APL may be a novel component of this in *Dictyostelium*. Interesting, transcription of APL is upregulated 5-fold in a retinoblastoma-like gene (*rblA*) disruption strain. In this strain, generally speaking, genes encoding proteins involved in transcription-coupled repair are

upregulated 3- to 6-fold[332]. Albeit not a direct assessment of the function of APL, these data support the results gathered in this chapter.

The role of APL in response to S-phase-associated damage is an unpredicted discovery. Thus far, we have compared APL to the human proteins that share its FHA domain: APLF, APTX and PNKP. As previously stated, PNKP and APTX are end-processing factors in SSBR and NHEJ, whereas APLF stabilises NHEJ repair complexes[15,74,76]. Therefore, these proteins are not associated with DNA repair in S-phase, or the repair of inter-strand crosslinks. Moreover, these proteins are recruited to DNA lesions by interactions of their FHA domains with the scaffold proteins XRCC1 and XRCC4, which are known to function in SSBR and NHEJ, respectively, but not in inter-strand crosslink repair or DSBR in S-phase[469,470]. If APL is involved in the repair of inter-strand crosslinks, this could suggest either a novel binding partner of the FHA domain of APL, which may also be true of the domain in human proteins, or implicate XRCC1 and XRCC4 in additional roles in the DDR. Therefore, identifying constitutive or damage-induced binding partners of APL will be highly informative for the further characterisation of this protein.

6. Investigating APL in DNA inter-strand cross-link repair

6.1. Introduction

6.1.1. The repair of DNA inter-strand cross-links

DNA inter-strand cross-links (ICLs) are highly toxic lesions that prevent the complementary DNA strands from separating during DNA replication and transcription. Consequently, the toxic effects of ICLs are most prominent in proliferating cells. ICLs may also distort the helical structure DNA, preventing proteins from binding[471]. Detection of ICLs is thought to primarily occur in S-phase, following replication fork stalling, and multiple models have been proposed for the repair of ICLs in this manner[466,472]. The repair of ICLs was originally investigated in prokaryotes and lower eukaryotes, such as budding yeast. In *E. coli*, the action of the NER machinery, followed by HR, was shown to be required for ICL repair[473]. In yeast, the genes required for the repair of ICLs were identified in three epistasis groups, each representing a different sub-pathway required for the repair of these complex lesions: NER, TLS (or post-replication repair), and HR. Representative genes from each of these groups: *pso2*, *rev3* and *rad51*, respectively, were deleted in yeast cells, and each strain displayed increased sensitivity to ICL-inducing agents[271]. Mechanistically, the NER machinery makes incisions either side of the ICL, unhooking it, and the resulting intermediate is resolved by TLS and HR[474,475]. For ICLs repaired in G1, in the absence of a homologous template, gap-filling is performed by TLS, and *rev3*-disruptant strains have enhanced sensitivity in this cell-cycle phase[476]. The 5'-3'

exonuclease PSO2 was found to be indispensable for ICL repair in yeast, and is also found to be functionally different to the other members of the NER epistasis group[477]. Unlike cells with NER-deficiency, *pso2*$^{-}$ cells are proficient at performing the initial incisions in ICL-repair, but are unable to perform downstream processing and DSBR to resolve the lesion[272,273]. An ortholog of PSO2 has been identified in humans and mice[276,478]. The human PSO2 ortholog, SNM1A, has previously been discussed as it is one of three human proteins containing a PBZ domain, thereby indicating that this PSO2-dependent ICL-repair pathway exists in higher eukaryotes, and that it may involve ADP-ribosylation[241]. Human cells depleted of SNM1A show increased sensitivity to ICLs, and elevated levels of replication-associated DSBs[466]. Furthermore, there is evidence of an error-prone, HR-independent ICL-repair pathway operating in mammalian cells, utilising NER and TLS machinery[479–481].

A proposed model for replication-associated ICL-repair is shown in Figure 6.1, following the convergence of two replication forks at the ICL[472]. Briefly, the NER machinery is employed to make incisions either side of the ICL, utilising the Mus81-Eme1 and ERCC1-XPF endonuclease complexes[482–484]. SNM1a has a crucial role in digesting across the ICL in a 5' to 3' manner[466]. The crosslink is now unhooked, and is then bypassed by the TLS DNA polymerase Rev1, and the synthesised strand is then extended by Polζ (Rev3/Rev7), another TLS DNA polymerase[485–487]. The ICL is then excised, again by NER machinery, leaving a DSB intermediate on one DNA strand. This DSB is resolved by HR to complete the repair with minimal errors[485,488–490].

6.1.2. The Fanconi Anaemia pathway

The Fanconi Anaemia (FA) pathway, named after the condition that mutation of its core proteins presents, has evolved specifically to repair ICL damage, and is a major ICL-repair

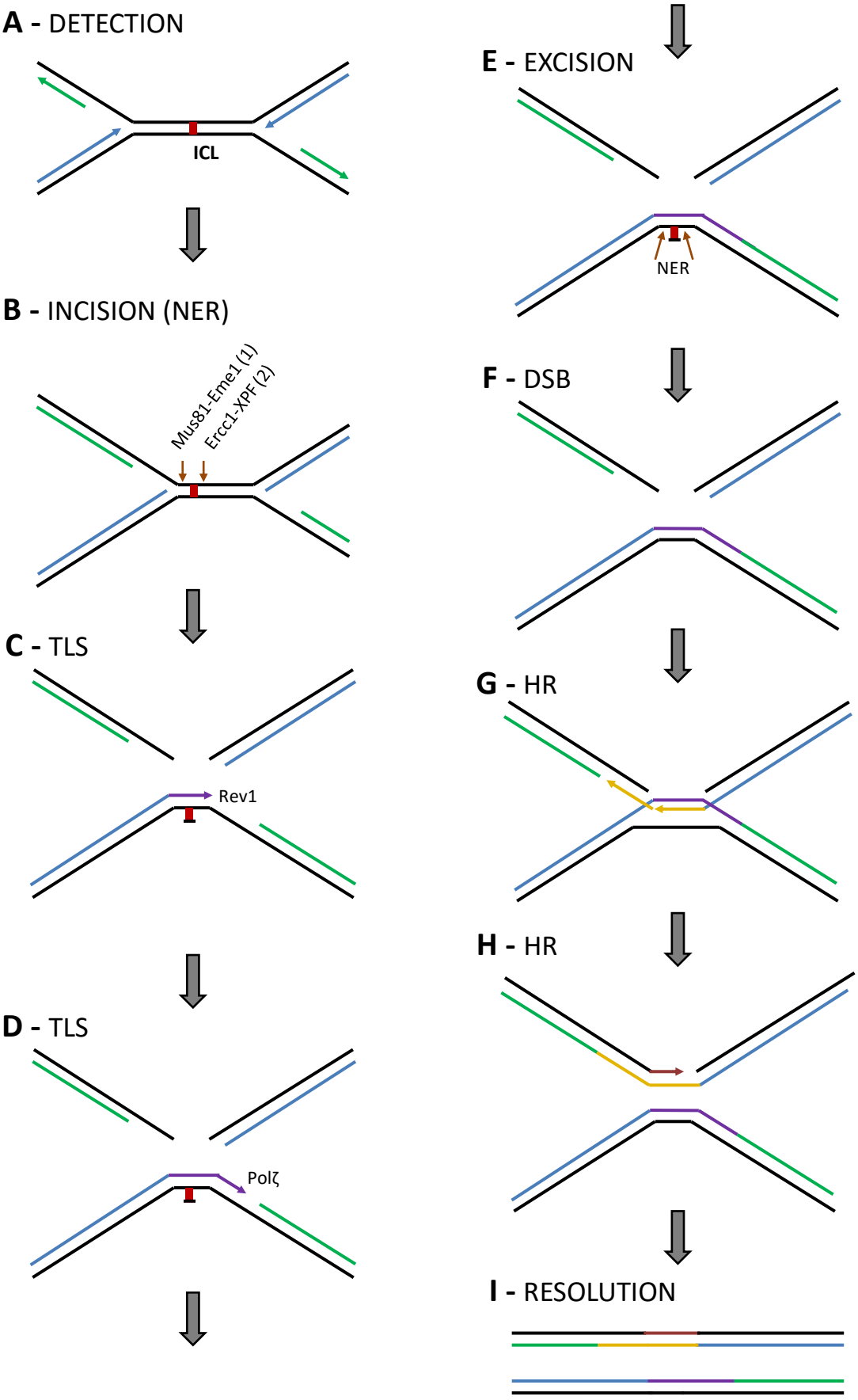

A - DETECTION
ICL
B - INCISION (NER)
Mus81-Eme1 (1)
Ercc1-XPF (2)
C - TLS
Rev1
D - TLS
Polζ
E - EXCISION
NER
F - DSB
G - HR
H - HR
I - RESOLUTION

Figure 6.1 – The replication-coupled repair of ICLs in mammalian cells. A speculative model for the replication-dependent repair of ICLs, highlighting the coordination of multiple DNA repair pathways. The lesion is detected by stalled replication forks (A). The ICL is unhooked by the sequential incisions by the endonuclease complexes Mus81-Eme1 and ERCC1-XPF (B). DNA is synthesised across the lesion by the translesion synthesis (TLS) polymerase Rev1 (C) and is then continued by another TLS polymerase, Polζ (D). The region of DNA containing the ICL is then excised using NER machinery (E) to leave a DSB (F). The HR pathway is then engaged to resolve the DSB, using the newly synthesised DNA as a template (G-I). Adapted from Moldovan and D'Andrea (2009).

pathway in mammalian cells. As such, FA-deficient cells display hypersensitivity to ICL-inducing agents[491]. The core components of FA pathway are involved in detection and signalling of ICLs, which coordinate other DNA repair pathways to facilitate repair, namely: NER, translesion synthesis (TLS), and HR[485]. Study of the FA pathway, for the most part, has been limited to higher eukaryotes, as the pathway was not thought to be conserved in lower eukaryotes such as budding yeast. However, factors involved in a FA-like pathway have recently been discovered in this organism[492].

Of the seventeen proteins so far identified causing Fanconi Anaemia in humans, several are found in one complex, known as the FA core complex. These proteins are FANCA, FANCB, FANCC, FANCE, FANCF, FANCG, FANCL and FANCM [493–500]. Other proteins, yet to be identified as causing Fanconi Anaemia in a patient, but found in the FA core complex, include FAAP100 and FAAP24[501,502]. It is thought that FANCM and FAAP24 are involved in recruiting the FA core complex to ICLs in S-phase, which are marked by stalled replication forks[503]. The FA core complex acts as a ubiquitin E3 ligase, with FANCL as its catalytic subunit and Ube2T as its ubiquitin E2 ligase partner[499,504]. The target for ubiquitination in response to ICL detection is the heterodimer of FANCD2 and FANCI, which are both monoubiquitinated[505–510]. While FANCI ubiquitination is dispensable for FA pathway activation, FANCD2 ubiquitination and its subsequent localisation to chromatin is essential component of this pathway. However, phosphorylation of FANCI by ATR is believed to activate the FA pathway[511]. The activity of the FA pathway is controlled by the ubiquitination state of FANCD2, with USP1 deubiquitinating the protein in normal cellular conditions[512]. Deletion of USP1 leads to sensitivity to cross-link-inducing agents, suggesting that deubiquitination of FANCD2 is also important for functional ICL repair[513,514].

Downstream of FANCD2 ubiquitination is the coordinated action of several DNA repair pathways to resolve the ICL. Interactions between ubiquitinated FANCD2 and proteins involved in these pathways are believed to be necessary for this orchestration. The structure-specific endonuclease subunit SLX4 and nuclease FAN1 are both recruited to ICLs through direct interaction with mono-ubiquitinated FANCD2, and are required for ICL processing[515–522]. Furthermore, the helicase and BRCA1-interacting protein FANCJ (BACH1), the key HR protein identified as FANCD1 (BRCA2), and the BRCA2 interacting factor FANCN (PALB2), also function downstream of FancD2 mono-ubiquitination[302,523–528]. The illegitimate action of NHEJ to repair the DSBs formed in ICL repair is toxic to the cell, as suppression of NHEJ in FA-deficient cells rescues them following exposure to ICL-inducing agents[529]. Thereby, it is thought that one of the major roles of the FA pathway is to ensure that high fidelity repair of these lesions by HR is performed[530].

6.1.3. Conservation of ICL repair proteins in *Dictyostelium*

A homolog of yeast Pso2 (Dclre1) has been identified in *Dictyostelium*, and cells lacking this protein display elevated sensitivity to ICLs, indicating that ICL-repair involving this protein is conserved in this organism (unpublished data). However, deletion of Dclre1 also results in delayed DSBR, indicating similarity to the human protein Artemis[351]. In humans, three homologs of Pso2 are present: SNM1A, SNM1B and SNM1C (Artemis). SNM1A is the closest homolog to yeast PSO2 and functions in ICL-repair, whereas Artemis is a nuclease in NHEJ[155,531]. *Dictyostelium* only has one homolog of PSO2, Dclre1, which would appear to function in multiple DNA repair pathways. The conservation of some NER and TLS factors, such as XPF and Rev3, has been discussed previously, and the presence of these pathways is required for the operation of Dclre1-mediated repair.

Additionally, the components of a functional FA pathway have been discovered in *Dictyostelium*. A minimal FA pathway, consisting of FncD2, FncI, FncL, Ube2T, FncM, FncJ and FncD1/Brca2, has been identified, which appears to function with a reduced core complex compared to mammalian cells. The roles of FncL and Ube2T in the essential mono-ubiquitination of FncD2 are conserved. However, clear functional differences have been found between the mammalian and *Dictyostelium* pathways. Deletion of the components of the FA pathway or TLS polymerases in mammalian cells results in high levels of sensitivity to ICL-inducing agents. In *Dictyostelium*, however, deletion of these factors yields moderate sensitivity, with epistasis shown between the FA and TLS components. Contrastingly, deletion of the NER nuclease XPF in *Dictyostelium* yields extreme sensitivity to ICLs, but not other core NER components such as XPC, indicating that XPF may play an NER-independent role[347]. It is postulated that this difference may have evolved as a consequence of the cell-cycle of *Dictyostelium*. In the FA pathway in mammalian cells, ICLs are primarily detected due to replication fork stalling in S-phase[472]. In *Dictyostelium*, the majority of ICLs will occur in G2-phase, on one sister chromatid. ICLs in G2 can be resolved in a much simpler manner, requiring excision of the lesion on both strands and deployment of HR to seal the gap. This model predicts that HR-deficient cells would also be highly sensitive to ICLs; however, no data is yet available to test this hypothesis.

6.1.4. Aims

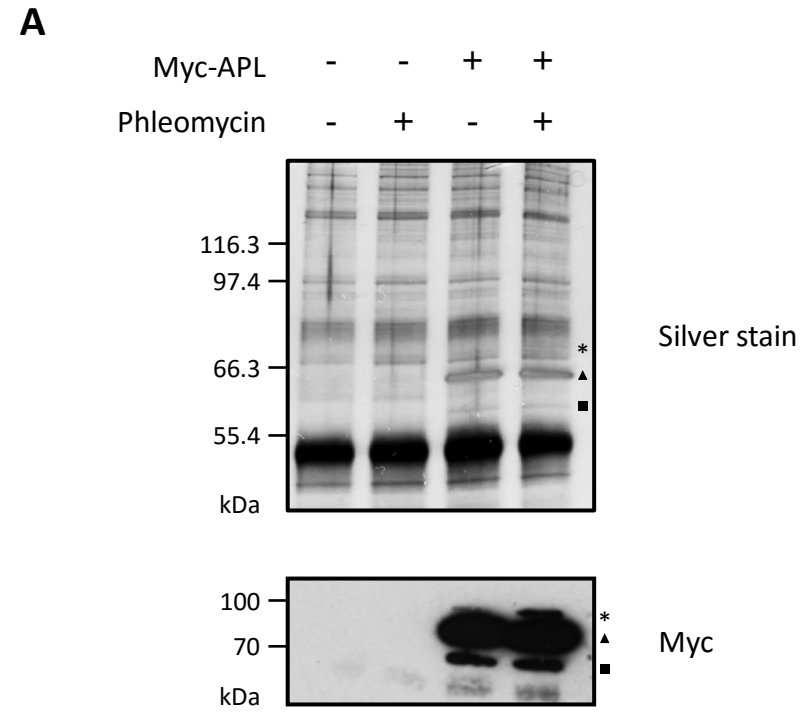
To this point, we have identified a potential PAR-binding protein, APL, through *in silico* methods, and verified this PAR-binding activity through *in vitro* assays. We then assessed the survival of *apl*$^{-}$ cells to various types of DNA damage, and found them to be normal compared to parental Ax2 cells. However, we have observed recruitment of Myc-APL to

chromatin, primarily in response to DNA cross-links, and also S-phase-induced DSBs. The association of APL with chromatin was also shown to be dependent on its macro domain region; however, deletion of this did not fully abrogate the cisplatin-induced enrichment of APL to chromatin. The aim of this chapter is to further investigate the role of APL in the repair of DNA cross-links.

6.2. Results

6.2.1. APL is modified in response to DNA DSBs

Prior to the data on the enrichment of Myc-APL on chromatin (Figure 5.10), we were unclear as to which DNA repair pathway APL may be involved in. To address this, we decided to investigate the DNA damage-induced interacting partners of APL by co-immunoprecipitation, which we hoped would be known DDR factors. We firstly assessed proteins interacting with APL in a DSB-induced manner, as the FHA domain of APL is predicted to bind to the NHEJ factor XRCC4 in response to DSBs[161]. Therefore, we transfected Ax2 cells with empty PDXA-3C vector or PDXA-3C-Myc-APL and exposed them to phleomycin. After treatment, co-immunoprecipitation was performed by incubating the protein extracts with anti-Myc agarose beads. This experimental design pulls down Myc-APL directly with the anti-Myc antibody, which is conjugated to the agarose beads, in conditions favourable to preserving any protein-protein interactions between Myc-APL and any binding partners that it may have. The choice of expressing recombinant Myc-APL instead of directly using the anti-APL antibody previously generated was due to the lack of specificity of the anti-APL antibody. Analysis of the pull-downs was first performed by silver staining (Figure 6.2A), and it revealed a putative damage-induced interacting protein (asterisk), and a putative constitutive interacting protein (square), in the cells expressing Myc-APL. Subsequent analysis of the same


A
Myc-APL - - + +
Phleomycin - + - +
116.3
97.4
66.3
55.4
kDa
Silver stain
100
70
kDa
Myc

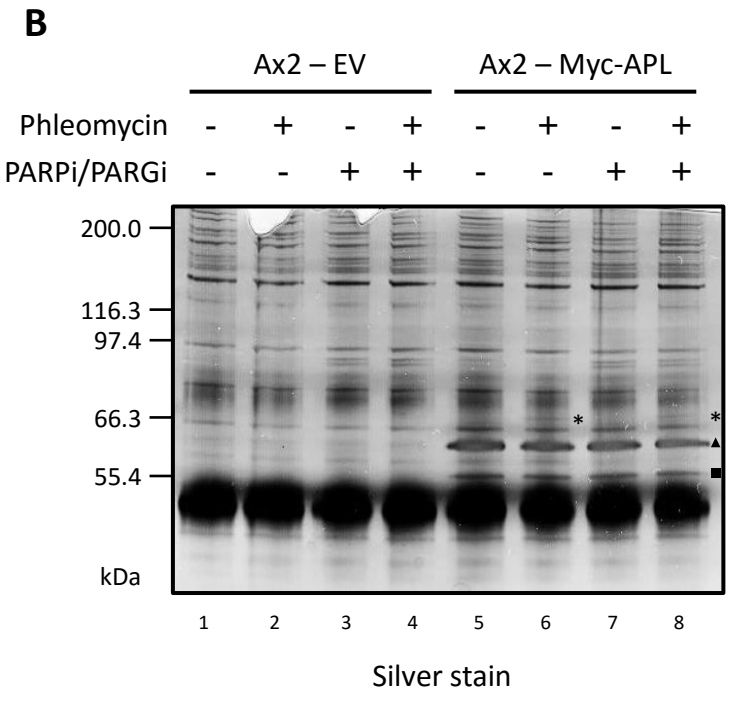

B
Ax2 – EV
Ax2 – Myc-APL
Phleomycin - + - + - + - +
PARPi/PARGi - - + + - - + +
200.0
116.3
97.4
66.3
55.4
kDa
1 2 3 4 5 6 7 8
Silver stain

Figure 6.2 – Co-immunoprecipitation of proteins interacting with Myc-APL reveals a damage-induced species. A: Ax2 cells transfected with empty pDXA-3C vector (Ax2 – EV) or overexpressing Myc-APL (Ax2 – Myc-APL) were treated with 300 μg/ml phleomycin for one hour or mock-treated. Following phleomycin exposure, the cells were washed and incubated with anti-c-Myc agarose beads to pull-down Myc-APL and any proteins interacting with it. Proteins bound to the anti-c-Myc agarose beads were analysed by SDS-PAGE, followed by either silver staining or Western blotting with an anti-Myc antibody. Myc-APL (triangle), a probable degradation product (square), and a damage-induced species (asterisk) are visible. **B:** Ax2 cells transfected with empty pDXA-3C vector (Ax2 – EV) or overexpressing Myc-APL (Ax2 – Myc-APL) were treated as in **A**. Prior to washing, the cells were split and one half were subsequently exposed to buffers containing PARP and PARG inhibitors, while the other half were exposed to standard buffers. These extracts were then separated by SDS-PAGE and the protein bands were visualised by silver staining. The bands are marked as in **A**.

samples by Western blotting with a different anti-Myc antibody to that found on the beads indicated that both of these bands were size-shifted forms of Myc-APL (Figure 6.2A). The most plausible explanation of higher molecular weight form of APL (asterisk) is that it undergoes post-translational modification in response to DSBs. The lower molecular weight band (square), originally identified as a putative constitutive interactor, is therefore identified as a degradation product. To further strengthen this conclusion, this band was only present in a minority of successful transfections with pDXA-3C-Myc-APL.

Although we have identified that APL undergoes some form of modification following induction of DSBs, we have not identified any proteins that interact with APL. Therefore, we addressed the possibility that the original co-immunoprecipitation experiments were not optimised for detecting protein-protein interactions dependent on PARylation. Considering the domain structure of APL, we believed that such interactions are likely to be present. One of the main challenges in studying PARylation is maintaining the stability of PAR chains in extracts, which are rapidly catabolised, alongside minimising the induction of PARylation by extraction procedures. To counter this in our experiments, following treatment of Myc-APL-expressing Ax2 cells with phleomycin, we split the cells in half and added PARP and PARG inhibitors (PARPi/PARGi) to the wash and lysis buffers used in one half, and compared this to normal buffers in the other (see Materials and Methods). Comparison of the samples was performed by silver staining (Figure 6.2B). The addition of PARPi and PARGi is observed to have an effect on the interactions between proteins, as new bands appear in all of the PARPi/PARGi treated samples. However, none of the new bands are DNA damage-induced, and they are also present in cells transfected with an empty PDXA vector as a control (lanes 3 and 4). Therefore, these new bands do not represent damage-induced or constitutive interactions of proteins with

Myc-APL. Importantly, however, the modified band of Myc-APL (asterisk) remains following the inclusion of PARPi and PARGi in the extraction buffers.

6.2.2. APL is modified primarily in response to DNA cross-links

Following the discovery that APL is modified in response to DSBs, we wished to see whether this modification occurs in response to other forms of DNA damage, particularly cross-links, which APL has been implicated in the repair of. Furthermore, after failing to identify any proteins interacting with APL in either a constitutive or DSB-induced manner, we wished to widen our approach to include the induction of different forms of DNA damage. Therefore, we expressed Myc-APL in Ax2 cells and performed co-immunoprecipitation (co-IP) with an anti-Myc antibody, following cellular exposure to MMS, phleomycin, 4-NQO, cisplatin, or camptothecin. Analysis of the immunoprecipitates was performed by SDS-PAGE and silver staining (Figure 6.3), and indicated that the modified form of APL was induced much more strongly in response to cisplatin, compared to phleomycin (marked by asterisks). The modified form was also induced in response to MMS and camptothecin treatment, though much more weakly, but not present in 4-NQO-treated cell extracts. The presence of a strong cisplatin-induced band matches the enrichment of APL on chromatin following cellular exposure to this genotoxic agent (Figure 5.10), providing further evidence that APL is involved in the repair of DNA cross-links.

To further verify that the strong increase in the induction of the upper band in the silver-stained gel in response to DNA cross-links is due solely to the modification, we treated *apl*$^{-}$ cells with the same array of DNA damaging agents as previously, and subsequently generated whole-cell extracts. We then analysed the samples by SDS-PAGE and Western blotting with the anti-Myc antibody (Figure 6.4). The antibody detected the induced band

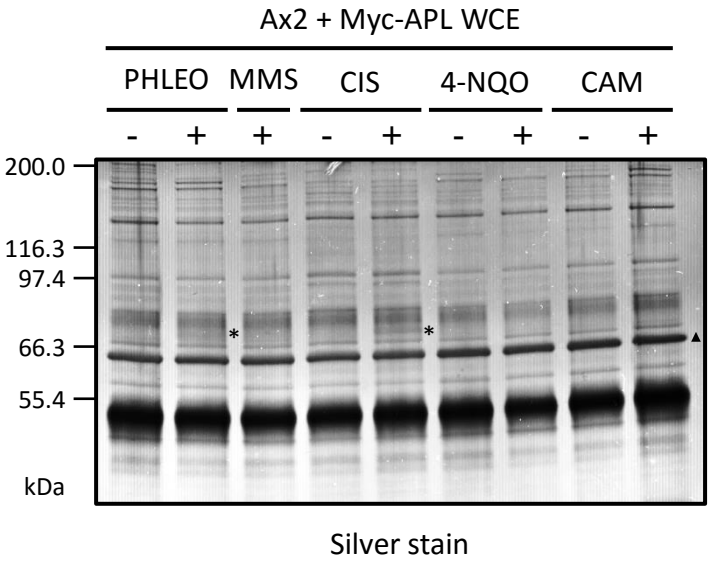


Figure 6.3 – The damage-induced species is present most strongly in response to DNA cross-links. Ax2 cells expressing Myc-APL were exposed to the indicated genotoxic agents, or mock-treated. The cells were exposed to 300 μg/ml phleomycin (PHLEO), 5 mM MMS and 5 μg/ml 4-NQO for one hour, 300 μM cisplatin (CIS) for five hours or 30 μM camptothecin (CAM) for three hours. Following treatment with these agents, the cells were washed and incubated with anti-c-Myc agarose beads to pull-down Myc-APL and any proteins interacting with it. The resulting proteins were denatured by boiling in SDS loading buffer, and were analysed by SDS-PAGE. Protein bands were observed by silver staining. A damage-induced band in response to both cisplatin and phleomycin (asterisk) is observed at slightly higher molecular weight than Myc-APL (triangle; 65.5 kDa). The heavy chain of the anti-c-Myc antibody is clearly visible below the 55.4 kDa marker.

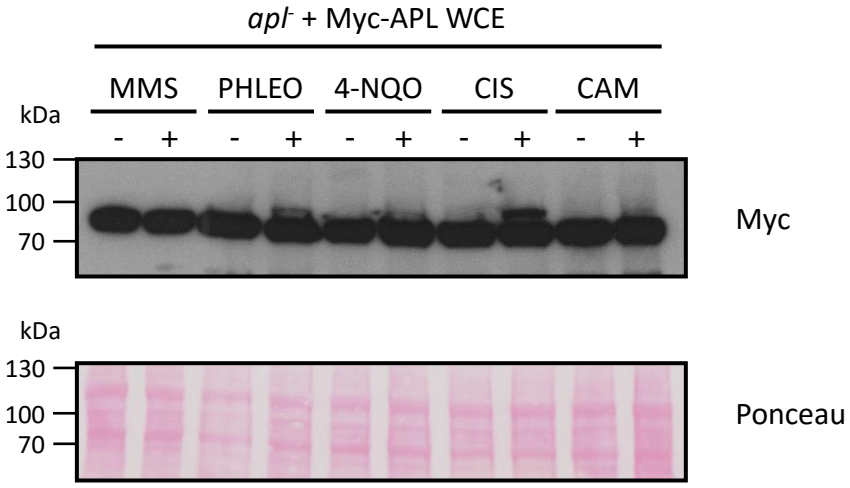


Figure 6.4 – Analysis of whole-cell extracts by Western blotting identifies a damage-induced, higher molecular weight form of Myc-APL. Whole-cell extracts were prepared from *apl*$^-$ cells overexpressing Myc-APL following treatment for one hour (unless specified) with either 5 mM MMS, 300 µg/ml phleomycin, 5 µg/ml 4-NQO, 300 µM cisplatin (CIS, five hours), or 50 µM camptothecin (CAM, three hours). The extracts were then analysed by SDS-PAGE and Western blotting with an anti-Myc antibody. Protein loading is observed by Ponceau staining prior to incubation with the anti-Myc antibody.

strongly in cisplatin-treated samples, indicating that this band is the modified form of APL. We also observed the modification in phleomycin- and camptothecin-treated samples, in agreement with previous data. Additionally, a weak band is observed in response to 4-NQO. Therefore, we have shown that DNA cross-links are the most potent stimulus of the modification of APL.

To further confirm that the induced band is a post-translationally modified form of APL, we analysed it by mass spectrometry. We excised the modified band from the silver-stained co-immunoprecipitation gels and sent the sample for analysis. The most significant match obtained from peptides present in the sample, that were judged to have originated from *Dictyostelium* proteins, were from Myc-APL, providing further evidence that this is a modified form of the protein (Figure 6.5A). Interestingly, the *Dictyostelium* protein identified in the sample with second highest significance was a ubiquitin-ribosomal fusion protein (ubqC), a protein from which ubiquitin is cleaved prior to its use in protein post-translational modifications. The peptides identified originated from the ubiquitin section of this protein (Figure 6.5B). UbqC has a molecular weight of 17.4 kDa, so it is likely that this signal, at approximately 70 kDa, is due to the presence of a ubiquitinated species in the sample. Although this is far from conclusive data that APL is the protein that is ubiquitinated, the size-shift of the modified form relative to unmodified Myc-APL could be explained by a mono-ubiquitination event. Additionally, we analysed whole-lane samples containing all of the proteins from each co-IP: Myc-APL-expressing and control Ax2 cells, each mock-treated or treated with phleomycin, in order to identify any constitutive or damage-induced interactions too weak to be identified by the silver staining. No proteins implicated in DNA repair were identified in any of the samples with statistical significance.

A

Protein Name	Size (kDa)	Probability	Protein Function
APL	64.3	1E-10	Putative DDR protein
ubqC	17.4	2E-05	Ubiquitin/ribosomal protein S27a fusion protein
mhsp70	71.4	0.0001	Mitochondrial heat shock protein
DDB_G0270388	106.9	0.0001	Ubiquitin system component Cue domain-containing protein
DDB_G0284741	156.1	0.0002	Suppressor of ty-induced promotor mutation
DDB_G0276445	72.5	0.0006	Heat shock family protein
DDB_G0273339	134.6	0.0050	Uncharacterised
DDB_G0267572	49.4	0.0126	Glioma tumour suppressor candidate
vwkA	70.2	0.0126	Von Willebrand factor kinase A
DDB_G0284527	56.4	0.0398	Uncharacterised

B

Protein encoded by ubqC:

MQIFIKTLTGK<u>**TITLEVEGSDNIENVK**</u>**AKIQDKEGIPPDQQRLIFAGKQLEDGR**<u>**TLSDYNIQKESTLHLVLR**</u>**LRGG**GGK
KKKKKTYATPKVLKRKLRKVKLAVLKYYKFDENGKIKRVLRECPAETCGAGVFMAQHANRQYCGKCHSTLVKKSK

UBIQUITIN RIBOSOMAL PROTEIN <u>**PEPTIDE IDENTIFIED BY MASS SPECTROMETRY**</u>

Figure 6.5 – Mass spectrometry analysis identifies the damage-induced band as a higher molecular weight form of Myc-APL. A: The damage-induced band present in co-immunoprecipitation experiments with Myc-APL was excised from the gel and analysed by mass spectrometry. The peptides identified by mass spectrometry were assigned to known *Dictyostelium* proteins, which are shown. The results are sorted by the probability that the observed match is a random event. **B:** The amino acid sequence of the ubiquitin/ribosomal fusion protein encoded by the *ubqC* gene. Peptides from this protein were identified in the mass spectrometry analysis of the excised band (underlined).

6.2.3. APL is ubiquitinated in response to DNA damage

The discovery that APL is modified following cisplatin treatment strongly implicates APL in the response to DNA cross-links, while weaker forms of the modification in response to DSBs (both replication-associated and independent) suggest that it may have a broader role in resolving S-phase-associated damage. However, the nature of the modification is still unknown. The presence of a ubiquitinated species in the mass spectrometry data might suggest that the modification is some form of ubiquitination event, but this is circumstantial evidence at this point. There are many different forms of post-translational modifications that are implicated in the DDR; however, some can be ruled as unlikely given the data available thus far. Single phosphorylation, methylation, MARylation and acetylation events do not increase the mass of the target protein significantly enough to explain the band-shift observed with Myc-APL, while multiple additions of these groups would lead to the formation of a laddered array of bands. Furthermore, protein modification by poly-ubiquitination, poly-SUMOylation and PARylation produces chains of varying lengths, and therefore yields multiple bands or smears on SDS-PAGE gels. This suggests that the single band observed is unlikely to be caused by one of these modifications. Both mono-ubiquitination and mono-SUMOylation involve the conjugation of a small, approximately 10 kDa protein to the target, thereby increasing its molecular weight. This size-shift would account for that observed with Myc-APL, and therefore we hypothesised that one of these modifications was responsible for the damage-induced modification of Myc-APL. We chose to investigate the potential ubiquitination of APL first, due to identification of a ubiquitinated species in the mass spectrometry results. Therefore, we performed further pull-downs of Myc-APL with c-Myc agarose beads, following the exposure of Myc-APL-expressing Ax2 cells to cisplatin or mock treatment. Despite the fact that we are no longer searching for Myc-APL-interacting proteins, and the

modification is observable in whole-cell extracts, this approach was required to remove other ubiquitinated proteins from the samples which would mask the signal from Myc-APL. Samples were then analysed by SDS-PAGE and Western blotting with anti-Myc and FK2 antibodies, the latter detecting mono- and poly-ubiquitinated conjugates[532]. The Western blotting identified that the damage-induced band is a mono-ubiquitination event, by alignment of the FK2 and anti-Myc blots (Figure 6.6A, asterisk). Moreover, additional damage-induced bands and smearing in Myc-APL expressing Ax2 cells are also identified by the FK2 antibody, suggesting that the modification of APL may be more than a single mono-ubiquitination event. Unmodified Myc-APL (triangle) is also detected by the FK2 antibody, presumably due to non-specific binding to the high concentration of protein. Additionally, the degradation product is present in the anti-Myc blot (square).

To further verify this mono-ubiquitination event, we treated c-Myc pull-down samples with a deubiquitinating enzyme (DUB). We chose the commercially available core complex from USP2 (usp2cc) as the DUB for this experiment. The protein pull-downs were prepared as previously; however, prior to elution, the samples were split and one half was incubated with the DUB. Subsequent analysis by SDS-PAGE and Western blotting with the anti-Myc antibody confirmed that the modification is mono-ubiquitination, as it is removed by treatment with the DUB (Figure 6.6B). Collectively, these data provide strong evidence that APL is mono-ubiquitinated following induced DNA cross-links, and also weakly poly-ubiquitinated.

6.2.4. APL mono-ubiquitination is dependent on its macro domain region

Having confirmed that APL is mono-ubiquitinated in response to DNA cross-links, we then wished to determine the domain dependence of this damage-induced post-translational modification. We have previously constructed a series of vectors to express

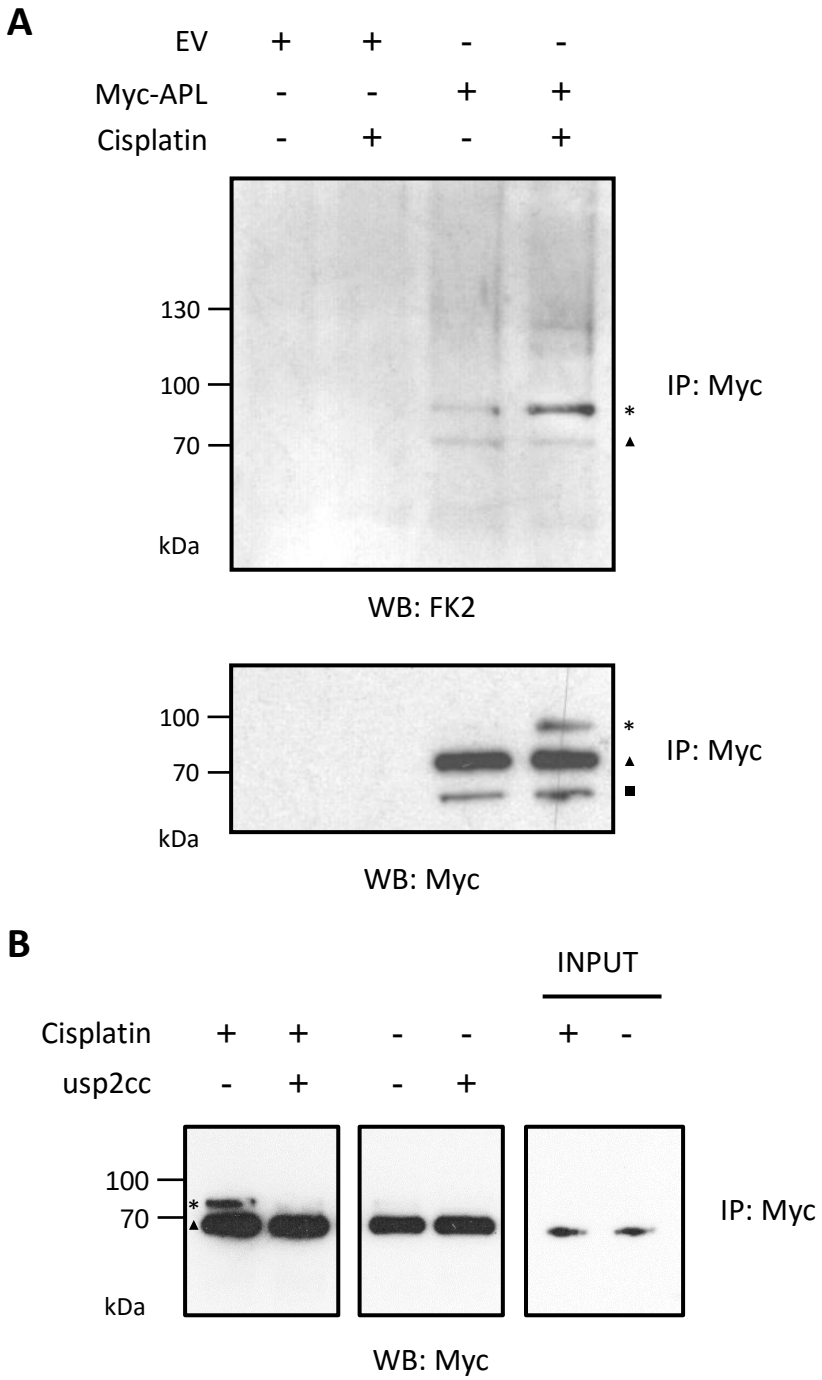


Figure 6.6 – APL is mono-ubiquitinated in response to DNA cross-links. A: Ax2 cells transfected with empty pDXA-3C or pDXA-3C-Myc-APL were treated with 300 µM cisplatin for five hours, or mock-treated. Following treatment, washing and lysis, the cell extracts were incubated with c-Myc agarose beads to immunoprecipitate all forms of Myc-APL. The resulting extracts were analysed by SDS-PAGE and Western blotting with anti-Myc and FK2 (anti-mono- and poly-ubiquitinated protein) antibodies. Myc-APL (triangle), the modified form of Myc-APL (asterisk) and a degradation product of Myc-APL (square) are identified. **B:** Myc-APL expressing Ax2 cells were treated and immunoprecipitated as in **A**. Prior to eluting the proteins from the c-Myc agarose beads, the extracts were treated for two hours with USP2 core (a deubiquitinating enzyme), or mock-treated. Samples were subsequently analysed by Western blotting with an anti-Myc antibody.

mutant forms of recombinant Myc-tagged APL in *apl*$^{-}$ cells (Figure 5.11). We used these strains to address this question. We treated cells expressing Myc-APL, Myc-APL-C174AC180A (PBZ domain mutant; P*), Myc-APL-ΔMACRO (macro domain deletion; ΔM), Myc-APL-E439A (macro domain mutation; M*), Myc-APL-C174AC180AE439A (PBZ and macro domain double mutant; PM*), and control cells with cisplatin (or mock treated), and subsequently generated whole-cell extracts from these cells. Analysis of these extracts by Western blotting with the anti-Myc antibody revealed that the damage-induced ubiquitination event is abrogated in the cells expressing only Myc-APL-ΔMACRO, the macro domain deletion strain (Figure 6.7). This implies that either the macro domain region is ubiquitinated following cisplatin treatment, or an interaction mediated by the macro domain is a necessary precursor to mono-ubiquitination (or both). Interestingly, the E439A mutation, which is predicted to be functionally important for ligand binding in the permuted macro domain of APL, and is observed to reduce the cisplatin-induced enrichment of Myc-APL on chromatin (Figure 5.11), appears to increase the level of modification of APL. This is observed in both the single M* mutant (lane 11), and also in the double PM* cell-line (lane 12). Furthermore, a longer exposure revealed a second, damage-induced form of Myc-APL at high molecular weight (circle). This band was not observed on the silver-stained gels following co-immunoprecipitation of Myc-APL, but was observed retrospectively in previous whole-cell extracts to be induced in response to cisplatin only. The presence of such a high molecular weight band is puzzling, as no known post-translational modification would increase the weight of the protein to that degree. However, this could indicate some form of multiple modifications of APL. Additionally, any dimers or trimers of Myc-APL formed in a damage-induced manner would be broken apart during sample boiling and reduction. Nevertheless, this high molecular weight modification is observed robustly in a cisplatin-dependent manner, and

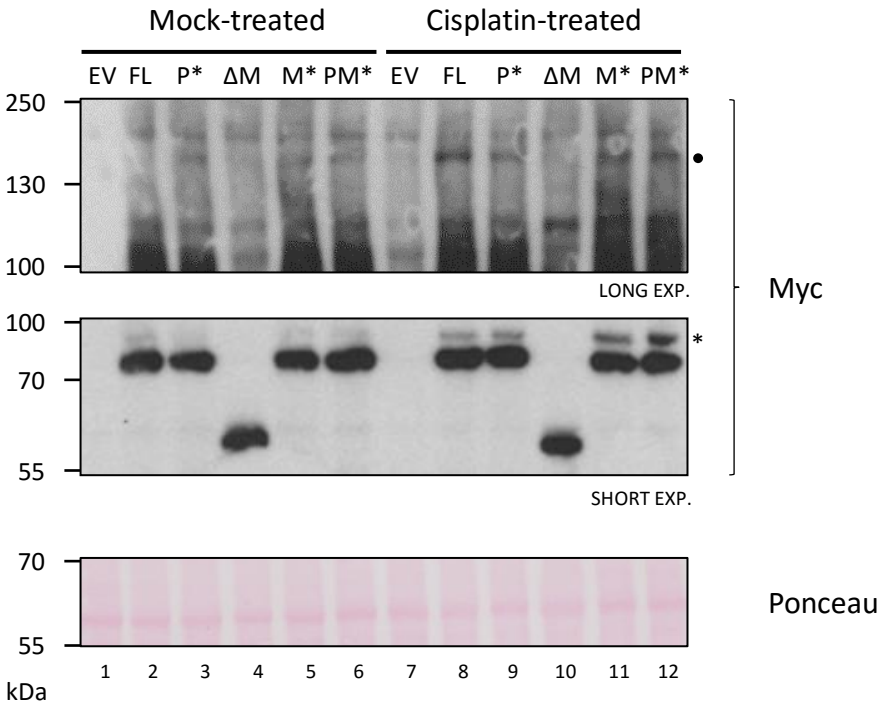


Figure 6.7 – Deletion of the macro domain of APL abrogates its ubiquitination. *apl*$^{-}$ cells over-expressing the following proteins: empty PDXA vector control (EV), full-length Myc-APL (FL), Myc-APL-C174AC180A (PBZ-domain mutant; P*), Myc-APL-ΔMacro (macro domain deletion; ΔM), Myc-APL-E439A (macro domain mutant; M*), and Myc-APL-C174AC180AE439A (PBZ and macro domain double mutant; PM*), were treated with 300 μM cisplatin for 5 hours, or mock treated. Following treatment, whole-cell extracts were prepared and protein content was analysed by Western blotting with an anti-Myc antibody. In addition to the ubiquitinated form of Myc-APL (asterisk), a long exposure revealed a further damage-induced form of Myc-APL at high molecular weight (circle). Protein concentrations necessary to observe the modified forms of Myc-APL were too high for the use of an anti-Actin antibody as a positive control, therefore Ponceau staining was utilised for this purpose.

is not present in cells not expressing Myc-APL. It is unclear from this data, due to the presence of non-specific bands, whether or not the high molecular weight form of Myc-APL is present in the macro domain-deletion mutant (lane 10).

6.2.5. The high molecular weight form of APL is absent in *ube2T*⁻ cells

Thus far, we have implicated APL in the response to DNA-cross-links by observing its ubiquitination and enrichment on chromatin following cisplatin treatment, and also in response to DSBs. This role of APL in response to cisplatin implies that it may act in one of the ICL-repair pathways present in *Dictyostelium*. Despite extensive study of human ICL-repair pathways, no protein with the domain structure of APL has been implicated in the repair of ICLs. Moreover, PARylation has not been associated in ICL-repair, and so the identification of APL as a putative ICL-repair factor is a discovery with novel implications. One such ICL-repair pathway is the Fanconi Anaemia pathway, which has recently been investigated in *Dictyostelium*[347]. In both humans and *Dictyostelium,* the mono-ubiquitination of FancD2 (FncD2 in *Dictyostelium)* is required for FA-pathway activation, and we hypothesised that the mono-ubiquitination of APL may be dependent on the same ubiquitin ligases responsible for FncD2 ubiquitination (Figure 6.8A). Orthologs of the E2- and E3-ubiquitin ligases involved in this modification in human cells, Ube2T and FancL, respectively, have been identified and characterised in *Dictyostelium*[347]. We were kindly provided with strains with the genes encoding these two proteins disrupted, by KJ Patel, alongside their parental controls (Ax2 for *fncL*⁻, and a Tap-tagged FncD2 knock-in for *ube2T*⁻). The pDXA-3C vector encoding Myc-APL was transfected into these strains, and whole-cell extracts were prepared, following treatment with cisplatin. Modification of Myc-APL was subsequently assessed by SDS-PAGE and Western blotting (Figure 6.8B). Interestingly, while the absence of FncL or Ube2T had a negligible effect

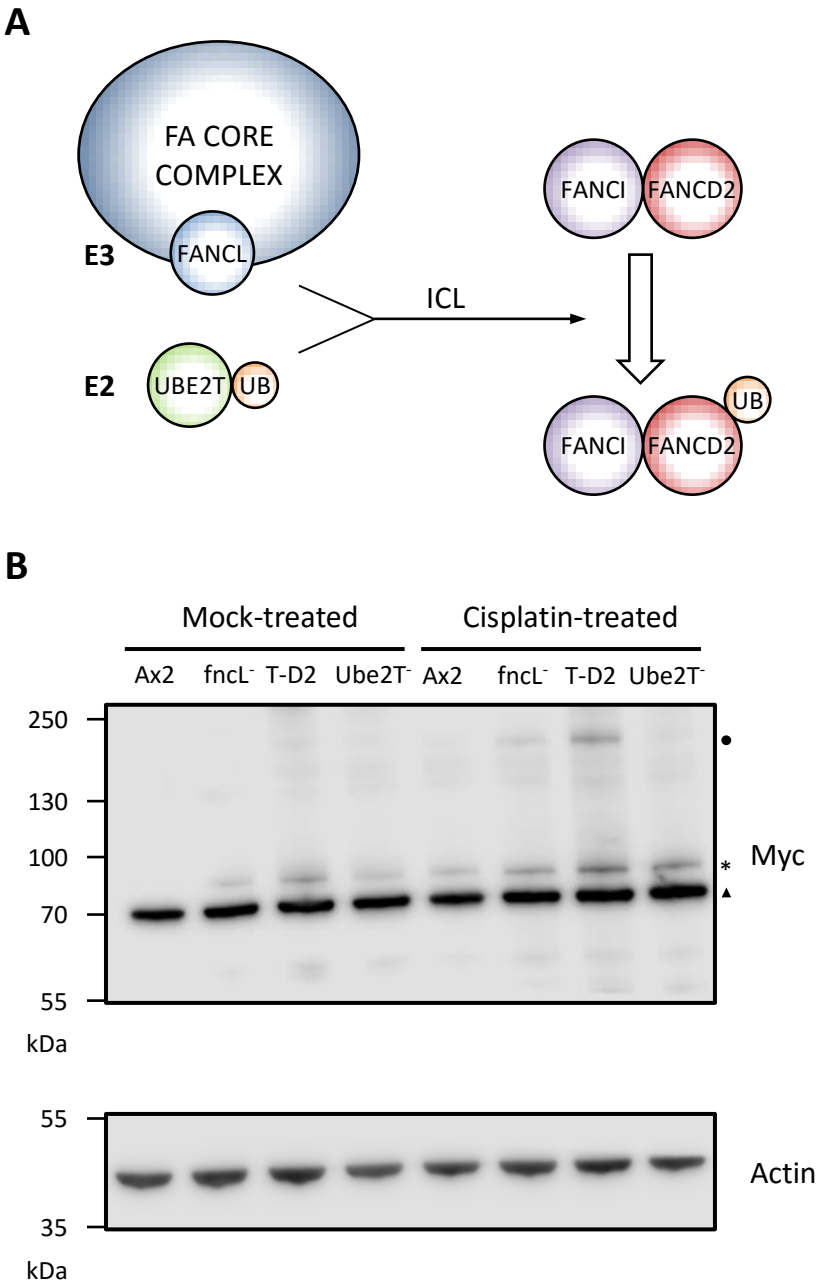


Figure 6.8 – Deletion of Ube2T abrogates the high molecular weight modification of APL. A: A highly simplified schematic of the activation of the human Fanconi Anaemia pathway, identifying Ube2T and FancL as the E2- and E3-ubiquitin ligase pair implicated in the ubiquitination of FancD2. **B:** Myc-tagged APL was over-expressed in *fncL*$^{-}$ and *Ube2T*$^{-}$ strains, alongside Ax2 and *Tap-fncD2* knock-in strains as the respective parental controls. The cells were treated with 300 μM cisplatin for 5 hours, or mock-treated. Subsequently, whole-cell extracts were prepared and protein content was assessed by Western blotting. Myc-APL (triangle) is indicated in all strains, alongside the ubiquitinated (asterisk) and high molecular weight (circle) forms of the protein. All of the strains used in this experiment were kindly provided by KJ Patel (Cambridge MRC LMB).

on the mono-ubiquitination of Myc-APL, the currently uncharacterised high molecular weight form of Myc-APL (circle) is not present in *ube2T*$^-$ cells, but is present in the *fncL*$^-$ strain. Ubiquitin E2 and E3 ligases act in pairs in a highly specific manner to ubiquitinate a target protein, and in humans, FancL is the only known E3 partner of Ube2T. This experiment would indicate that Ube2T in *Dictyostelium* acts independently of FncL in facilitating the high molecular weight modification of APL, either by direct action or by ubiquitinating an upstream protein. However, the FK2 blot of cisplatin-treated cells expressing Myc-APL (Figure 6.6) did not indicate a species of the same molecular weight as this high molecular weight band of APL, reducing the possibility that it is a poly-ubiquitinated form generated by the direct action of Ube2T. The independent action of Ube2T and FncL observed here may indicate that APL functions in a separate pathway to the classical FA pathway.

6.2.6. APL deletion does not further sensitise *fncD2*$^-$ cells to cisplatin

The lack of dependence of the mono-ubiquitination of APL on FncL might indicate that APL does not function in the FA pathway in *Dictyostelium*, and may function in a different ICL-repair pathway, such as that involving Dclre1 (the *Dictyostelium* homolog of yeast Pso2). If APL functions in a separate pathway to FA for ICL repair, then inactivation of both pathways in cells should result in further sensitisation to ICL-inducing agents than observed in FA-deficient cells alone. This would explain the lack of sensitivity of *apl*$^-$ cells to DNA cross-links, as the role of APL may only be unveiled in an FA-deficient background. To produce a FA-deficient strain, we devised a strategy for the disruption of the *fncD2* gene, utilising endogenous ClaI (C) and EcoNI (E) restriction sites to mark the external ends of the homologous arms used for targeting the disruption vector (Figure 6.9A). This strategy mimics that employed by KJ Patel for this disruption[347]. These arms

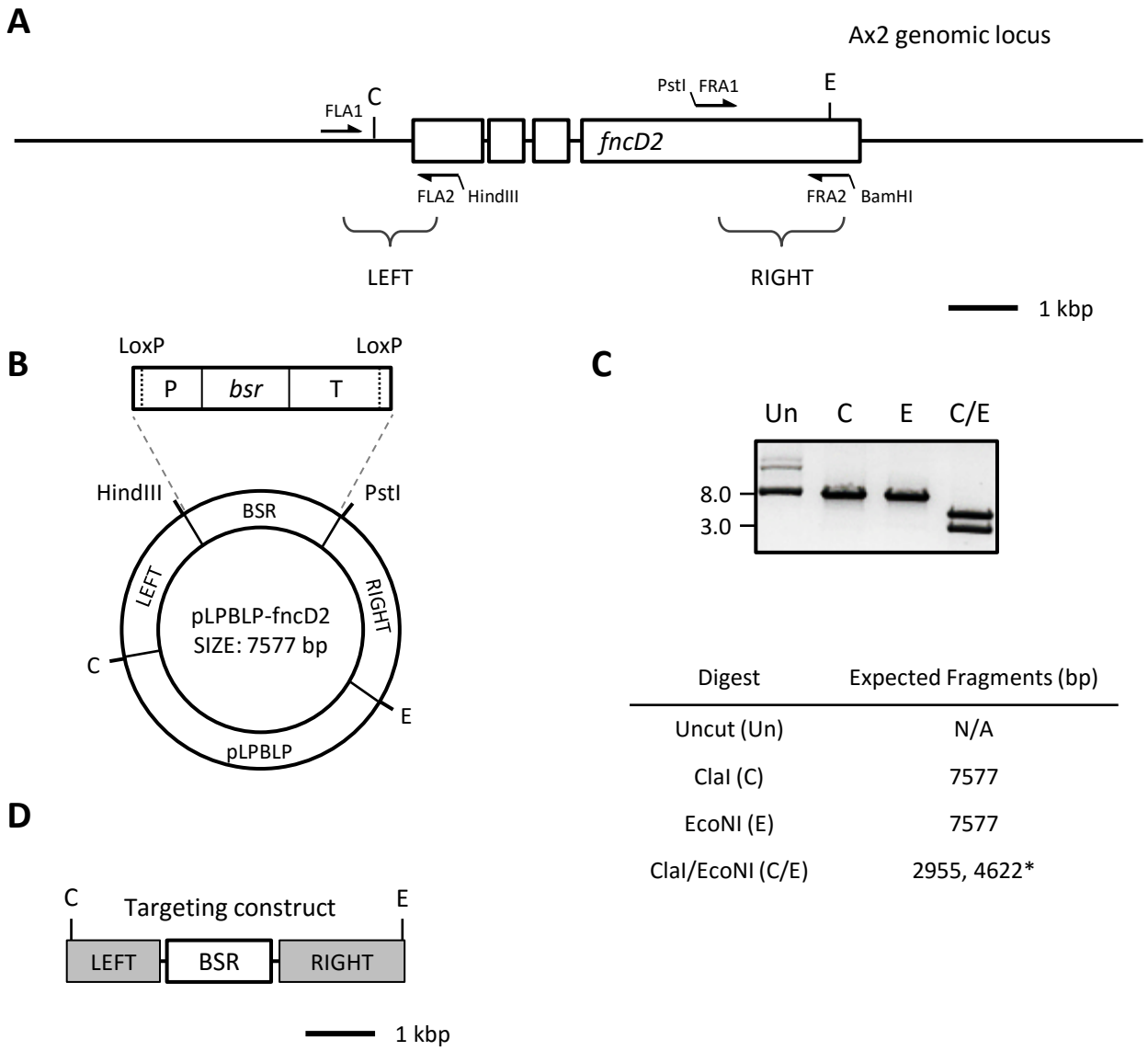
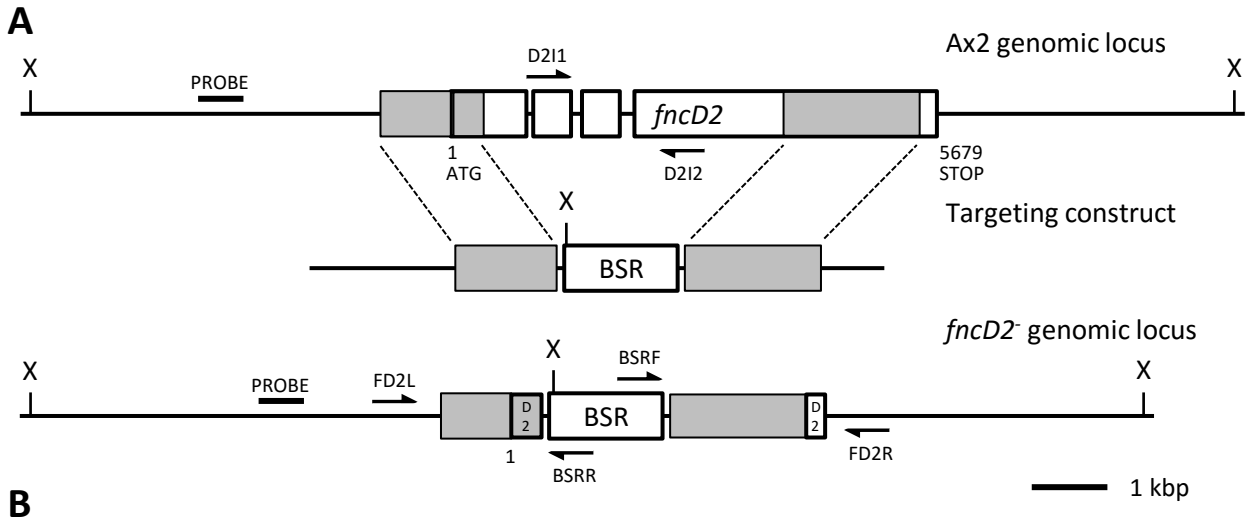
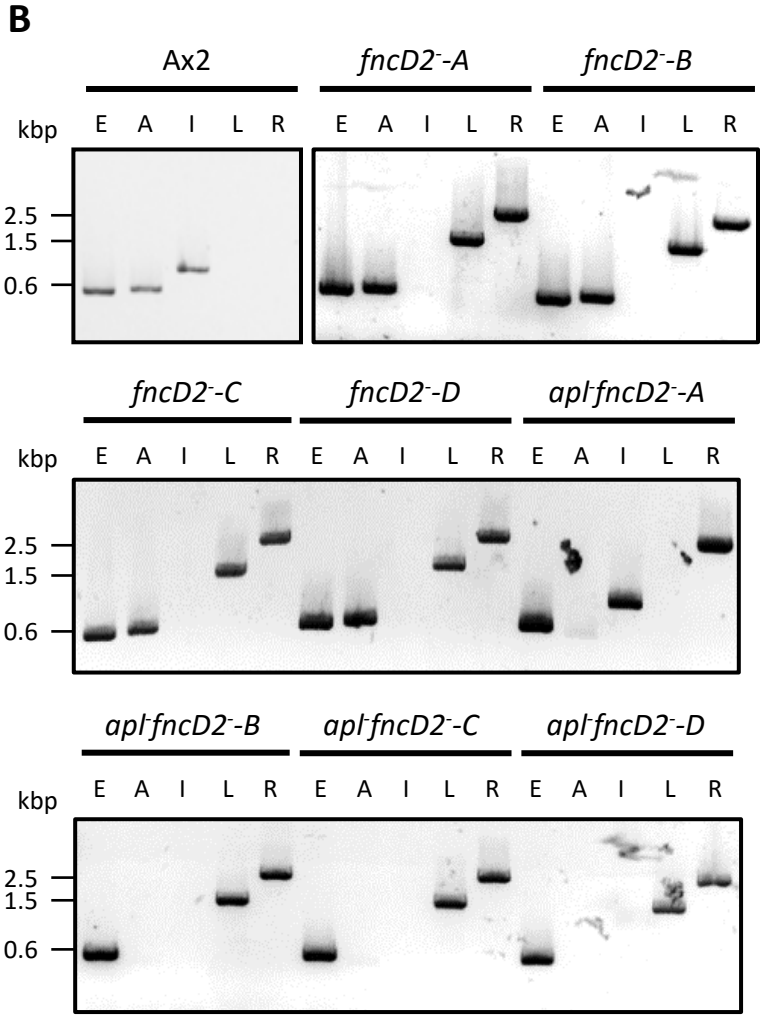


Figure 6.9 – Strategy for the disruption of the *fncD2* gene. A: The genomic locus of the *fncD2* gene. The primers designed for the generation of the homologous arms for the disruption of the gene by targeted homologous recombination are shown, with attached restriction sites where appropriate. Endogenous ClaI (C) and EcoNI (E) restriction sites are indicated. **B:** Map of the pLPBLP-fncD2 vector, constructed by ligating the genomic regions indicated in **A** into pLPBLP. The homologous arms are ligated either side of a cassette encoding a gene for resistance to blasticidin (*bsr* cassette), an antibiotic toxic to Ax2 cells. Indicated restriction sites map to those shown in **A**. The structure of the blasticidin resistance cassette is shown, containing the *bsr* gene, Actin6 promoter (P) and terminator (T). LoxP sites for excision of the cassette by Cre recombinase are also indicated. **C:** Validation of the pLPBLP-fncD2 disruption vector by restriction digest with ClaI (C) and EcoNI (E). Expected fragment sizes are shown in the table. The fragment purified and used for transfection into *Dictyostelium*, containing the homologous arms and *bsr* cassette, is marked with an asterisk. **D:** Diagrammatic representation of the 4622 bp fragment transfected into *Dictyostelium* cells for disruption of the *fncD2* gene.

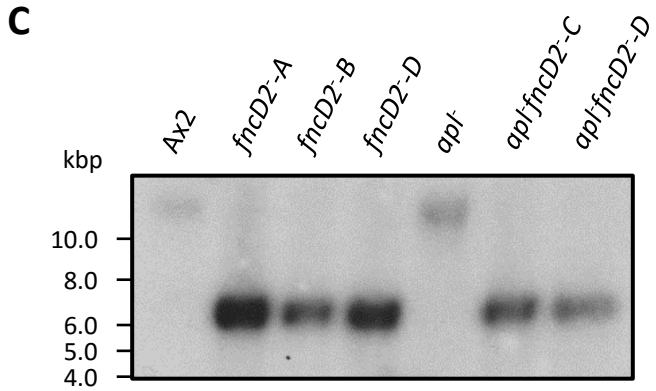
were amplified by PCR using the indicated primers, and ligated either side of a cassette encoding resistance to the antibiotic blasticidin (*bsr* cassette). A diagrammatic map of this vector is shown in Figure 6.9B. The DNA sequence of these arms was verified by Sanger sequencing (not shown), and by restriction digestion with ClaI and EcoNI (Figure 6.9C). The DNA fragment used for transfection of Ax2 and cre-loxed *apl*$^-$ cells is shown in Figure 6.9D.

Performing the transfections described above yielded both *fncD2*$^-$ and *apl*$^-$*fncD2*$^-$ mutants, which needed to be screened. The initial selection of potential mutants was performed by growing cells in media containing blasticidin, which were observed two weeks after transfection. The growth of cells in blasticidin indicates that either the intended recombination event (Figure 6.10A) has occurred, or that the cassette has been integrated randomly. Screening of both putative *fncD2*$^-$ and *apl*$^-$*fncD2*$^-$ strains was initially performed by PCR, with primer combinations amplifying over both homologous arms and internal to the deleted region of the *fncD2* gene (primer locations are shown in Figure 6.10A). Importantly, the primers labelled FD2L and FD2R are external to the region of recombination. The results of the PCRs with four putative *fncD2*$^-$ strains and four putative *apl*$^-$*fncD2*$^-$strains are shown in Figure 6.10B, alongside Ax2 cells as a control strain. These PCRs indicate that all four *fncD2*$^-$ strains were successful disruptions; while three out of four *apl*$^-$*fncD2*$^-$ double mutants were generated correctly (apl$^-$fncd2$^-$A appears to have only integrated the disruption construct at one end of the *fncD2* gene). Further verification was provided by Southern blotting. The relevant XbaI restriction sites and probe location are shown in Figure 6.10A. The BSR cassette contains an additional XbaI site, thereby breaking up the 17.5 kB fragment expected in Ax2 and *apl*$^-$ cells. The result of the Southern blot (Figure 6.10C) supports the PCR results for the indicated strains, therefore confirming the successful generation of at least two *fncD2*$^-$ and *apl*$^-$*fncD2*$^-$ strains. This





Primers	Expected PCR Products (bp)	
	Ax2	*fncD2*-
Ex1L/Ex1R (E)	500	500
APL1/APL2 (A)	618	618
D2I1/D2IR (I)	870	N/A
FD2L/BSRR (L)	N/A	1463
BSRF/FD2R (R)	N/A	2530



Strain	Expected Fragments (kbp)
Ax2, *apl*-	17.5
fncD2-	7.3

Figure 6.10 – Generation and verification of *fncD2* disruption strains. A: The strategy for disruption of the *fncD2* gene by targeted homologous recombination, thereby inserting a gene for resistance to blasticidin (BSR). Primers used for verification by PCR and the relevant XbaI (X) restriction sites and probe location for Southern blotting are indicated. **B:** Verification of *fncD2* disruption in Ax2 and *apl*$^-$ backgrounds (*fncD2*$^-$ and *apl*$^-$*fncD2*$^-$) strains by PCR. Expected fragment sizes are shown in the table. **C:** Verification of *fncD2*$^-$ and *apl*$^-$*fncD2*$^-$ strains by Southern blotting. Digestion of genomic DNA with XbaI (X) and selecting the region indicated in **A** for use as a probe gives a substantial fragment size-shift (shown in the table).

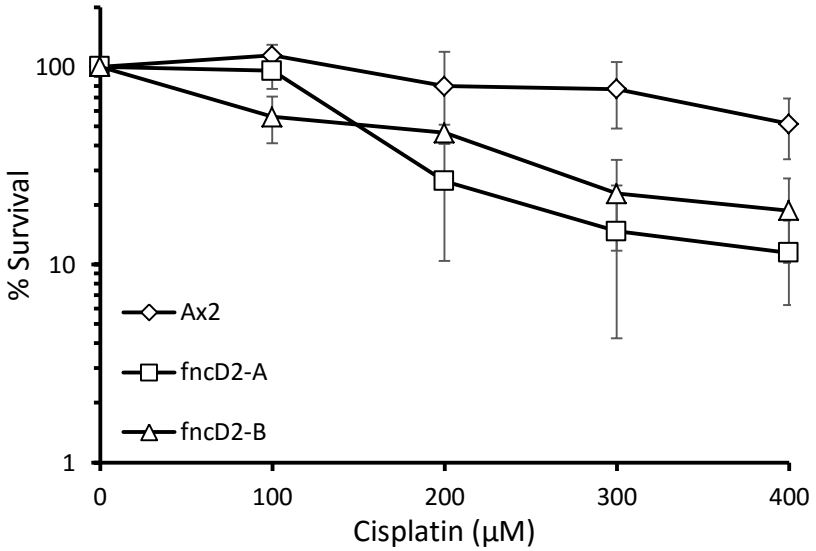
also provides an indication that *apl*$^{-}$ cells have functional homologous recombination, which is expected as they do not display elevated sensitivity to DSBs.

Before assessing the sensitivity of the *apl*$^{-}$*fncD2*$^{-}$ double mutants to cisplatin treatment, we first needed to phenotypically verify the *fncD2*$^{-}$ single mutants, which would also serve as a calibration for future experiments. Therefore, Ax2 cells and two *fncD2*$^{-}$ strains were exposed to a range of cisplatin concentrations for 5 hours, and their survival was quantified by plaque growth in the following 7 days post-incubation and washing. Both *fncD2*$^{-}$ strains displayed increased sensitivity to cisplatin over the dose range indicated (Figure 6.11A). Therefore, these *fncD2*$^{-}$ cells have the expected FA-deficiency. To assess the contribution of APL to cisplatin-resistance in the absence of FncD2, we performed the same survival assay with Ax2, *apl*$^{-}$, *fncD2*$^{-}$, and two *apl*$^{-}$*fncD2*$^{-}$ strains (Figure 6.11B). These experiments showed no additional survival defect of either *apl*$^{-}$*fncD2*$^{-}$ strain compared to *fncD2*$^{-}$ cells. This indicates that APL is not required for survival of cisplatin-induced damage in the absence of the FA pathway, but does not rule out a role for the protein in the repair of ICL-induced DNA damage.

6.3. Discussion

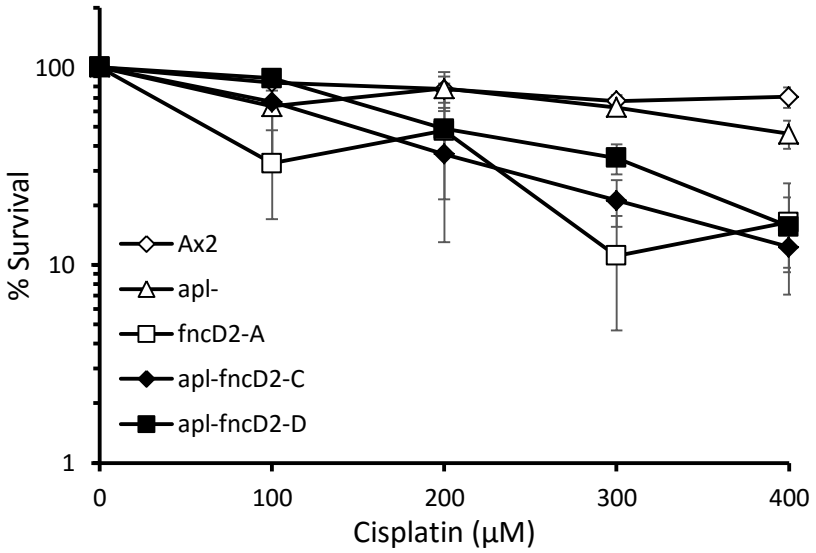
In this chapter, we have further elucidated the role of APL in the DDR of *Dictyostelium*, specifically with regard to induced DNA cross-links. We firstly tried to identify proteins that interact with APL in a damage-induced or constitutive manner, as we believed that these proteins would lead to a further understanding of the function of APL. We uncovered a potential cisplatin- and DSB-induced interacting factor by co-immunoprecipitation, but this was found by mass spectrometry analysis and Western blotting to be a mono-ubiquitinated form of APL. Analysis of mutant forms of APL identified that the mono-ubiquitination event was dependent on the C-terminus, macro domain region of the

A



Strain	D_{50} (μM)	SE	P-value (Ax2)
Ax2	344	48.23	N/A
fncD2⁻-A	178	38.1	0.03 (*)
fncD2⁻-B	160	26.9	0.02 (*)

B



Strain	D_{50} (μM)	SE	P-value (Ax2)	P-value (*fncD2⁻-A*)
Ax2	619	153.7	N/A	N/A
apl⁻	375	39.2	0.26	N/A
fncD2⁻-A	131	31.3	0.09	N/A
apl⁻fncD2⁻-C	171	49.3	N/A	0.54
apl⁻fncD2⁻-C	228	30.3	N/A	0.09

Figure 6.11 – *Dictyostelium apl*$^-$ cells are not further sensitised to DNA cross-links by disruption of the *fncD2* gene. A: Verification of *fncD2* disruption (*fncD2*$^-$) strains by assessing their sensitivity to DNA cross-links. Ax2 cells and two *fncD2*$^-$ disruption strains were exposed to the indicated cisplatin concentrations for 5 hours. Survival was assessed by observing plaque formation 3-7 days post-incubation. **B:** Similarly, Ax2, *apl*$^-$, *fncD2*$^-$ and two *apl*$^-$*fncD2*$^-$ strains were exposed to the indicated concentrations of cisplatin for 5 hours, and sensitivity was assessed by counting plaques formed between 3 and 7 days post-incubation. For each strain, survival is measured relative to an untreated control. In both experiments, error bars represent the standard error from three independent experiments. P-values were calculated from the mean D_{50} toxicity values, which are indicated in the tables with standard errors (SE). P-values are calculated with respect to the bracketed strain.

protein. This mono-ubiquitination event is clear response to DNA damage; however, the source and function of this modification remain unclear. The ubiquitination of proteins in response to DNA damage has been established for almost thirty years. The first example of this was found in yeast post-replication repair, which is the equivalent of human TLS. In this pathway, the ubiquitin E2 ligase Rad6 and E3 ligase Rad18 mono-ubiquitinate PCNA[533]. This mono-ubiquitination is detected by ubiquitin-binding modules on TLS DNA polymerases, which are switched-to from normal DNA polymerases in order to bypass a blockage to DNA replication[202,534]. In ICL repair, the mono-ubiquitination of FancD2 by Ube2T and FancL is required for FA pathway activation, while its heterodimeric partner FancI is also mono-ubiquitinated[506,510]. Histone H2B is mono-ubiquitinated by a heterodimer of RNF20 and RNF40 in response to DSBs, in an ATM-dependent manner. This ubiquitination is thought to promote chromatin condensation and efficient recruitment of NHEJ and HR factors[535,536]. Therefore, the presence of mono-ubiquitination of proteins in TLS, ICL-repair, and DSBR has been well established, as has the functional importance of these modifications. The mono-ubiquitination of APL may be required to mediate its interaction with other DDR proteins. The macro domain of APL has been shown to be required for the constitutive association of APL with chromatin, and the macro domain region of APL is also required for the ubiquitination of the protein, either containing the ubiquitination site, or mediating an interaction with a protein necessary for the mono-ubiquitination event. One theory that would explain the current data is that the mono-ubiquitination of APL occurs when the protein is on chromatin. The damage-induced enrichment of APL on chromatin appears to be independent of the ubiquitination, as this still occurs when the macro domain region of APL is deleted. Furthermore, if APL was ubiquitinated as a pre-requisite for its recruitment to chromatin, we would not observe the enrichment of the unmodified form of the protein on chromatin, and we do observe

this. Additionally, the mono-ubiquitination could directly impact the function of the macro domain, via conformational change or steric clash, either activating or abrogating its binding function.

The mono-ubiquitination of APL was found to be independent of the E2- and E3-ubiquitin ligases known to be involved in the activation of the FA pathway in response to inter-strand cross-links: Ube2T and FncL, and we have not yet identified which enzymes are responsible. APL has been associated with S-phase associated DNA damage, with DNA cross-links providing the strongest stimulus for its mono-ubiquitination. ICL-repair pathways all employ multiple DNA repair pathways to resolve ICLs, including TLS and DSBR, and APL may be ubiquitinated by proteins involved in these pathways. Homologs of the ubiquitin E2- and E3-ligases that mono-ubiquitinate PCNA in TLS, Rad6 and Rad18, have been automatically annotated in *Dictyostelium*. We hypothesise that these enzymes may be responsible for the mono-ubiquitination of APL, but we have yet to assess this. However, many of the other ubiquitin ligases in other organisms discussed here do not have clear *Dictyostelium* orthologs, and E2- and E3-ubiquitin ligase pairs act in a highly specific manner. Therefore, high-throughput methodologies may be required to identify the enzymes responsible.

Investigation of the mono-ubiquitination of APL using an FK2 antibody (Figure 6.6A) also highlighted the possibility that APL is poly-ubiquitinated. Poly-ubiquitination of proteins has been shown to be a modification of paramount importance in several DNA repair pathways. In DSBR, the ubiquitin E3 ligase RNF8 catalyses the formation of ubiquitin chains to promote the assembly of repair proteins, including the additional ubiquitin E3-ligase RNF168[537]. The ubiquitination of proteins by both RNF8 and RNF168, including H2AX in a DNA-damage specific manner on K13/15, leads to the recruitment or retention of many DSBR proteins on chromatin, including 53BP1, RAP80, BRCA1, and HERC2,

and thereby promotes both NHEJ and HR[538–544]. Furthermore, FAAP20, a component of the FA core complex, contains a ubiquitin-binding domain, and ubiquitination by RNF8 and its partner UBC13 is required for the recruitment of FAAP20 (and the FA core complex) to inter-strand cross-links[545]. Poly-ubiquitination of APL could indicate that it serves as a docking site for other DDR proteins with ubiquitin-binding domains, or that it is degraded, depending on the ubiquitin linkage. Additionally, it is not clear if the mono-ubiquitination and poly-ubiquitination are sequential or separate events, as we were unable to abrogate the mono-ubiquitination event by deletion of *ube2T* and *fncL* genes. From the examples given previously, RNF8 and RNF168 are strong candidates for catalysing the DNA cross-link-dependent ubiquitination of APL. However, orthologs of these two proteins have not yet been identified in *Dictyostelium*.

Analysis of cisplatin-treated whole-cell extracts revealed a high molecular weight species of APL, which cannot be explained by dimerisation or the addition of a single ubiquitin molecule or other known single-molecule post translational modification (Figure 6.7). Additionally, this band is not detected by the FK2 antibody following co-immunoprecipitation of Myc-APL, suggesting that this is not a poly-ubiquitinated form of APL. Interestingly, the induction of this high molecular weight form of APL is dependent on the ubiquitin E2-ligase Ube2T, but not its known E3 partner FncL (Figure 6.8). This indicates an indirect dependence of the modification on ubiquitination, potentially the poly-ubiquitination of APL, which we haven't assessed in the *ube2T*⁻ background. Other post-translational modifications that may be responsible for the higher molecular weight band are PARylation and SUMOylation, although the presence of one distinct band and lack of characteristic smearing detected by the anti-Myc antibody would argue against these possibilities. PARylation and its role in DNA repair have been discussed previously, and it has not previously been implicated in the repair of cisplatin-induced damage.

However, while the PAR-binding nature of APL *in vitro* might implicate PARylation in the repair of DNA cross-links in *Dictyostelium*, the PARylation of APL itself would be mechanistically interesting, and unexpected. If APL is PARylated, it may serve as some kind of adaptor or amplifier of the PARylation signal, depending on the behaviour of its domains; however, there are no direct comparisons of this behaviour in human PAR-binding proteins. Human SNM1a contains a PBZ domain that is predicted to bind to PAR, which implicates PARylation in ICL-repair[241]. However, this domain has not been experimentally characterised. SUMOylation is also implicated in several DNA repair pathways, including BER, HR and NHEJ[546–550]. Analysis of the amino acid sequence of APL reveals a putative SUMOylation site in the macro domain region of the protein (K437)[551]. It is unclear, due to the presence of background bands, if the high molecular weight band is present in extracts from cells expressing APL-ΔMACRO only, which lacks the putative SUMOylation site. The E439A mutation (M*) is predicted to disrupt the consensus motif of the SUMOylation site (ΨKXD/E, where Ψ is a hydrophobic residue, K is the SUMOylated lysine, X is any amino acid, and D/E is an acidic residue); however, this mutation is not observed to abrogate the induction of the high molecular weight, damage-induced band[551]. These data suggest that this band is not a SUMOylated form of APL, although we have not yet assessed this directly by Western blotting. A further explanation of this band would be the conjugation of a large protein to APL, in a manner analogous to ubiquitination and SUMOylation. While a post-translational modification involving the conjugation of such a large protein has not previously been observed, the existence of such a modification should not be excluded at this stage.

So far, we have addressed the modification and potential mechanism of chromatin association of APL, but we have not been able to identify specifically where and when APL acts in DNA repair. The strongest stimulus of both the modification of APL, and its

enrichment on chromatin, has been cisplatin. Cisplatin induces primarily intra-strand cross-links, with as low as 1% induced inter-strand cross-links reported as a result of exposing DNA to this chemical[461,462]. The repair of intra-strand cross-links has been shown to be performed by the NER pathway; however, the data gathered on APL following cellular exposure to 4-NQO suggests that APL is not functional in this pathway[552]. Therefore, it is likely that APL functions in the repair of inter-strand cross-links. APL is also modified in response to DSBs, and enriched on chromatin following the induction of replication-associated DSBs. This is not a contradictory observation, as DSBs are known intermediates of ICL repair (Figure 6.1)[468,553]. We assessed the mono-ubiquitination state of APL in cells deficient in FncD2 ubiquitination (*fncL*$^{-}$ and *Ube2T*$^{-}$), and found that it was unperturbed in these strains[347]. This suggests that APL is not a FA protein, as the FA pathway is not active in these strains, unless APL mono-ubiquitination occurs upstream of FncD2 mono-ubiquitination. We have not investigated the ubiquitination of FncD2 in *apl*$^{-}$ cells; however, we would expect this to be normal as *apl*$^{-}$ cells do not bear the survival defect associated with FA-deficient cells. If APL does function upstream of FncD2 mono-ubiquitination, it is dispensable for FA pathway activation. A more likely hypothesis would be that APL functions in an ICL-repair pathway independent of the FA pathway. The comparison between human SNM1A and APL has already been made, as both proteins contain a PBZ domain. However, APL does not contain any domains that would infer nuclease activity, although it may serve to recruit additional nucleases for this purpose. If APL does act in an alternative pathway for ICL repair, we hypothesised that the repair activity of the FA pathway may be masking the deficiency resulting from *apl* deletion. However, deletion of the *apl* gene in FA-deficient cells did not further sensitise those cells to cisplatin. Therefore, if APL does function in an alternative pathway for ICL repair, its contribution is either too minor to be detected by our assay, or it may be

redundant. An ICL repair deficiency of *apl*$^{-}$ cells may also be uncovered by treating cells with an alternative ICL-inducing agent, such as the alkylating chemical nitrogen mustard (HN2), as this would avoid the large variation associated with using cisplatin.

One hypothesis that would account for the lack of a survival defect associated with loss of APL is that APL functions in a sub-pathway of ICL-repair that mediates the resolution of intermediate DSBs by NHEJ. In the current model of ICL-repair, HR is responsible for the repair of the majority of DSB intermediates, which would imply that the requirement for APL in this process would only become apparent in HR-deficient cells. The FHA domain of APL is predicted to interact with XRCC4, and this interaction may be required to recruit NHEJ factors specifically to the intermediates of ICL repair. So far, we have made the assumption that the mono-ubiquitination of APL activates or promotes the activity of the protein, while it could function to suppress it in response to DNA lesions that would be better repaired by HR. In yeast, NHEJ has been shown to be involved in the resolution of DSB intermediates; however, this function is only detectable in HR-deficient cells[553]. NHEJ has also been implicated in ICL-repair in yeast *pso2*$^{-}$ cells[492]. Mammalian cells do not appear to be dependent on NHEJ for the resolution of such lesions, despite a minority of conflicting reports[489,554,555]. Furthermore, it is also reported that NHEJ is toxic to FA-deficient cells[529,530]. However, we did not observe a significant increase in survival of *apl*$^{-}$ *fncD2*$^{-}$ cells compared to *fncD2*$^{-}$ cells.

The data obtained on APL so far may also indicate a role for APL in TLS, which would explain the enrichment of APL on chromatin in response to cisplatin and camptothecin, and would also account for the weak mono-ubiquitination of APL in response to cell-cycle independent DSBs, as only 5-10% of cells in a vegetative *Dictyostelium* population are in S-phase[330]. The process of TLS (or post-replicative repair) is reliant on the ubiquitination of PCNA in yeast and humans, and has also been recently linked to ADP-ribosylation in

humans by the implication of ARTD8 and its macro domains[35,202]. The macro domain of APL may bind to MARylated proteins at sites of stalled replication forks, which would also be present during the repair of ICLs in S-phase. Despite the knowledge gained in this chapter, further work on the function of APL is required to determine its role in the DDR.

7. General discussion

In this thesis, we adopted an *in silico* approach to identify *Dictyostelium* proteins with previously unannotated PAR-binding domains, as we hypothesised that the presence of such domains may infer that the protein is involved in the DDR in this organism. This hypothesis was formed after considering the example of the NHEJ protein Ku70. The *Dictyostelium* ortholog of Ku70, has a C-terminus PAR-binding PBZ domain, and the role of this domain is to aid the stability and retention of NHEJ factors at DSBs[360]. Human Ku70 has no PBZ domain, and the stability of NHEJ proteins at DSBs in this organism is mediated by tandem PBZ domains in APLF, an ortholog of which has not been identified in *Dictyostelium*[15]. The presence of these PBZ domains on different proteins in the two organisms, but performing the same function, led us to speculate the search for unannotated PAR-binding domains in proteins might uncover novel DDR proteins.

The *in silico* search focused on two of the known PAR-binding modules: the PBZ and the macro domain. In humans and *Dictyostelium,* PBZ domains have so far been identified exclusively in DNA repair and cell-cycle regulating proteins, and those which have been experimentally characterised have all displayed PAR-binding activity[162,386,466]. Before this study, three human and seven *Dictyostelium* proteins had been identified to contain these domains, and while we replicated this result, we were unable to identify any novel PBZ domains[241]. In contrast, macro domains comprise a diverse superfamily, which members have been shown to bind to multiple ligands, including multiple forms of ADP-ribose[246,247]. Macro domains also do not obey a strict consensus motif, thereby justifying the employment of advanced bioinformatics tools incorporating profile-HMMs. Utilising this approach, we identified three *Dictyostelium* and one human protein with previously

unannotated macro domains. Despite these findings, the human genome still encodes an elevated number of macro domain containing proteins when compared to *Dictyostelium*: eleven to six, and even more individual macro domains as the three macroPARPs each contain at least two of these domains[28]. This suggests that either human cells are more reliant on the use of macro domains, or that macro domains with higher sequence divergence in the *Dictyostelium* genome are yet to be identified. The opposite case is true for PBZ domains; however, the strict consensus sequence of this domain makes it less likely that proteins containing this domain remain unannotated.

The proteins identified to contain previously unannotated macro domains in *Dictyostelium* were DNA Ligase III (Lig3), APL, and Q54R54. Although at this stage uncharacterised in this species, the role of Lig3 in DNA repair in humans is well known, and APL displayed homology of its FHA domain with the human DDR proteins APTX, PNKP and APLF[89,188,235,430]. Therefore, both proteins were taken forward for experimental analysis. Interestingly, the macro domain of APL appeared to have undergone a circular permutation, but predictive modelling of this domain using the solved structure of human MACROD1 as a template suggested that the domain was still functional[377,402]. The first stage of the analysis was to investigate if the macro domains of these two proteins displayed PAR-binding activity *in vitro,* which they both did in isolation. This observation led to characterisation of both proteins in the DDR in *Dictyostelium.*

The human protein Q9NQ89 was also identified to contain a macro domain, and appeared to be an ortholog of Q54R54 due to the level of homology of their macro domains. However, the lack of any additional identifiable domains of known function in these two proteins led to them not being experimentally characterised in this thesis. Regardless of this, the function of these proteins is certainly worth further study. However, as macro

domains are implicated in multiple cellular processes, investigations may be required outside of the context of DNA repair.

In this *in silico* approach, the motif first identified to bind PAR, the PBM, was excluded because its consensus motif was too short for current homology searching techniques. These algorithms would be unable to distinguish between true and false positives in the results[399]. Furthermore, the WWE domain was not considered as it is commonly encountered in protein degradation pathways, not DNA repair[242]. However, as is highlighted by the apparent stronger reliance of *Dictyostelium* on PBZ domains, and human cells on macro domains, *Dictyostelium* may utilise PAR-binding WWE domains for novel functions, perhaps mediating some aspects of the DDR of this organism. Overall, these data validate the *in silico* methods applied, as both of the domains that we assessed displayed PAR-binding activity, and this was the criteria for our search.

In humans, Lig3α, the nuclear isoform of Lig3, is implicated in SSBR and alt-NHEJ, and we hypothesised that Lig3 in *Dictyostelium* would act in the same DNA repair pathways. We observed the enrichment of recombinant Myc-Lig3 to chromatin following base alkylation, indicating a role for the protein in the BER/SSBR pathway. However, we have yet to address the role of the macro domain in this function. In humans, the recruitment of Lig3α to DNA SSBs is mediated by XRCC1[69]. Lig3α and XRCC1 exist in a heterodimer, and XRCC1 interacts with PARylated ARTD1 at the DNA lesion, thereby recruiting Lig3α to SSBs[34,68]. The interaction between Lig3α and XRCC1 is mediated by the BRCT domain of Lig3α[303,427]. *Dictyostelium* Lig3 has two BRCT domains, and it is likely that these also mediate an interaction with XRCC1 to localise Lig3 to DNA SSBs. Therefore, we cannot currently state that the macro domain of Lig3 is responsible for the enrichment of the protein on chromatin, as there are multiple potential modes of recruitment. The identification of point mutations that abrogate the PAR-binding ability of the macro

domain will allow us to address this question. Furthermore, the interaction of the macro domain of Lig3 and PAR chains *in vivo* may serve to stabilise the protein, and SSBR-complexes, on chromatin, in a manner analogous to the PBZ domain of Ku70 in NHEJ[360]. In this case, cells expressing Lig3 macro domain-mutants only may display slower repair kinetics. However, the *in silico* analysis suggested that the macro domain of *Dictyostelium* Lig3 displayed strongest homology to that of human ALC1, the macro domain of which is required for its recruitment to sites of DNA damage[250]. This may imply that the role of the macro domain in Lig3 is for direct recruitment to DNA breaks, if the function has been conserved.

It has recently been suggested that human Lig3α is directly recruited to DNA breaks in an ARTD1-independent manner[86]. Lig3α also contains a PARP-type zinc finger domain, which displays specificity for binding DNA nicks (whereas the second zinc finger of ARTD1 has higher binding affinity for DNA gaps)[68,446,556]. It was proposed that Lig3α would be recruited to a subset of SSBs through an interaction of this zinc finger domain. This was supported by the observation of recruitment of Lig3α to SSBs in the absence of XRCC1[86]. *Dictyostelium* Lig3 does not have a zinc finger domain, or any annotated DNA-binding motif, and therefore lacks this potential avenue for recruitment to SSBs. The presence of a PAR-binding macro domain may indicate an alternative mode of recruitment of this protein, albeit displaying a stronger reliance on PARylation, which may indicate functional differences in SSBR in this organism. Putative PAR-binding domains have been identified in Lig3 homologs in a range of species, indicating that the binding of Lig3 to PARylated proteins through specific domains may be conserved[400]. A PBM has been identified in human Lig3α, and this protein has been shown to interact directly with PARylated ARTD1 *in vitro*. However, this binding was not solely dependent on the PBM, and the physiological relevance of this interaction has not been ascertained[244].

Superficially, the presence of a direct interaction of Lig3α with PARylated ARTD1 would question the necessity of XRCC1 for the recruitment of Lig3 to SSBs; however, the role of XRCC1 in this has been strongly characterised[69,429,557]. This may indicate that the PAR-binding of Lig3α is only required in alt-NHEJ, where XRCC1 has been shown to be dispensable[171]. Therefore, the study of the macro domain of Lig3 could yield novel mechanistic details about the function of Lig3α in different repair pathways.

In a separate approach, we also investigated the role of Lig3 in the DDR by assessing the survival of *lig3*$^{-}$ strains to exposure to genotoxic agents[88,179]. Deletion of *lig3* is lethal in humans due to the requirement for the mitochondrial isoform Lig3β; however, the presence of a bacterial LigA homolog in *Dictyostelium* indicated that Lig3 may not be localised to the mitochondria in this organism[87–89]. In humans, cells depleted of nuclear Lig3α do not display elevated sensitivity to the induction of SSBs or DSBs, which is thought to be due to redundancy with Lig1. We hypothesised that the different domain structure of Lig3 in *Dictyostelium* may reduce the level of redundancy between the two proteins. However, *lig3*$^{-}$ cells did not display survival defect in response to either SSBs or DSBs, in agreement with the data obtained from human cells. The role of alt-NHEJ is uncovered in NHEJ- or HR-deficient cells, and the NHEJ factor Ku70 competes with ARTD1 to promote classical NHEJ over alt-NHEJ[107,177,189]. Therefore, we assessed the effect of *lig3* deletion in cells with compromised DSBR pathways. However, the deletion of *lig3* had no effect on the survival of these cells. We had hypothesised that the macro domain of Lig3 may be responsible for mediating interactions with ARTD1 in this repair pathway, and this interaction may be specific to Lig3, as Lig1 does not have any annotated PAR-binding domains. However, these data are indicative that Lig1 may be able to act in a completely redundant manner with Lig3 in *Dictyostelium*, in both SSBR and alt-NHEJ, which correlates to the observations in other species[88,179]. Unlike SSBR, in which we have

observed that Lig3 is enriched on chromatin following SSB induction, we do not have any experimental evidence in *Dictyostelium* that Lig3 is involved in alt-NHEJ. However, the level of conservation of DNA repair pathways in *Dictyostelium* in general is suggestive that this pathway is operational, and probably uses a similar set of protein orthologs to the human pathway[441].

The lack of a survival phenotype of *lig3*$^{-}$ cells means that the role of the macro domain of Lig3 cannot be investigated in this context. One approach to assess the role of Lig3 is to perform the experiments in a *lig1*$^{-}$ genetic background; however, the *lig1*$^{-}$*lig3*$^{-}$ double knock-out is believed to be lethal. Therefore, the generation of a conditional knock-out of Lig3 may be required to proceed with the investigation of the function of the macro domain of Lig3. It has been proposed that the Lig1 and Lig3 act in two sub-pathways of alt-NHEJ, as opposed to both performing the same function in a single pathway. These sub-pathways are distinguishable by the use of local microhomology to facilitate repair, as Lig3 is thought to repair DSBs in this manner more than Lig1[180]. Therefore, the use of plasmid-based assays and sequencing could be employed to determine the relative contribution of Lig1 and Lig3 in DSBR, and ultimately the role of the macro domain in this process.

We also experimentally characterised a second *Dictyostelium* protein in this thesis, APL. APL was identified to contain a C-terminus macro domain, in addition to a PBZ domain that had already been annotated[241]. Furthermore, it contains an N-terminus FHA domain that displays high levels of homology with the FHA domains of APTX, PNKP and APLF, which interact with XRCC1 and XRCC4, in SSBR and DSBR, respectively[161,454–456]. Both the macro domain and the PBZ domain of APL were identified to bind to PAR chains in isolation, *in vitro*. However, deletion of the macro domain from full-length APL had a negligible effect on the overall PAR-binding of the protein, suggesting that the PBZ

domain is dominant in this regard. The macro and PBZ domains may cooperate in PAR-binding, as in the case of the tandem PBZ domains of APLF, to increase the overall binding affinity of APL to PAR[268]. However, we have not considered the full range of forms of ADP-ribose that macro domains have been shown to bind[247]. For example, the macro domain of APL may have a stronger binding affinity for single ADP-ribose units than PAR chains, and may be its *in vivo* binding target. This would be supported by our *in silico* study, which indicated that the macro domain of APL displayed strongest homology to the first and second macro domains of ARTD8. The second macro domain of ARTD8 binds to MAR[407]. Furthermore, the homology of APL to this region of ARTD8 suggested a mechanism by which the circular permutation of the macro domain may have arisen, in a manner analogous to the gene duplication model described earlier, in which two adjacent genes become one through the introduction of stop codons by mutation[401,402].

The characterisation of APL in the DDR in *Dictyostelium* began with the evaluation of the sensitivity of *apl*$^-$ cells to various forms of DNA damage. However, no survival defect was detected compared to parental Ax2 cells in response to any of the genotoxic agents applied. While these data indicate that APL is not a vital member of the DNA repair pathways activated by the genotoxic agents, it does not exclude APL from the DDR as it may act redundantly with another protein or pathway. Therefore, we assessed the enrichment of recombinant Myc-APL on chromatin following different forms of DNA damage, and we observed that it was enriched following treatment with cisplatin and camptothecin. Cisplatin induces intra- and inter-strand cross-links; however, Myc-APL was not enriched on chromatin following 4-NQO treatment, indicating that the inter-strand cross-links induced by cisplatin were most likely responsible for the observed enrichment of Myc-APL[461,462]. Both inter-strand cross-links and the S-phase associated DSBs induced by camptothecin result in stalled replication forks in S-phase, implicating APL in the

response to replication-associated damage[472,558]. ARTD8 has been implicated in mediating HR at stalled replication forks, and ARTD10 has been shown to interact with PCNA[35]. Additionally, depletion of ARTD10 results in cellular sensitivity to inter-strand cross-links and replication stress in HeLa cells, which is an overlap with the types of damage that result in APL enrichment on chromatin[37]. ARTD1 and ARTD2 have also been shown to facilitate HR at stalled replication forks, and replication fork restart[16]. Therefore, ADP-ribosylation has been implicated in S-phase associated DNA repair in humans. Furthermore, no protein with the domain structure of APL has been identified in humans, and the predicted function of the protein inferred from its FHA domain is not S-phase associated damage, making APL a novel protein in many regards.

The expression of recombinant domain deletion or mutation versions of APL in cells further elucidated the behaviour of APL. Observing the chromatin enrichment of these mutants indicated two forms of association of Myc-APL with chromatin: constitutive and inter-strand cross-link induced. The constitutive interaction was abrogated completely by deletion of the macro domain region of APL (Myc-APL-ΔMACRO); however, this deletion did not completely abrogate the damage-induced enrichment of Myc-APL on chromatin. The damage-induced enrichment was reduced by mutation of either the macro domain (E439A), which was predicted from structural modelling to be the equivalent acidic amino acid as D160 in MACROD1 and D723 in ALC1, or the PBZ domain. These data are slightly conflicting, as it would be expected that macro domain deletion and mutation would result in the same phenotype, but this is not observed. One theory would be that the macro domain mediates two interactions with chromatin, and the E439A mutation only abrogates one of these. We have not yet determined whether the damage-induced enrichment of Myc-APL-ΔMACRO is reduced compared to the full-length protein, as this may indicate that the macro domain deletion has the same effect as the

mutation in this regard. However, the reduction of the enrichment of the PBZ domain mutant indicates that PARylation may be involved in the recruitment of APL to chromatin in response to inter-strand cross-links, as no alternative function of a PBZ domain has yet been discovered. PARylation has not been directly implicated in inter-strand cross-link repair. In humans, the inter-strand cross-link repair factor SNM1a contains a PBZ domain; however, the role of this domain in the function of SNM1a has not yet been established[241,466].

The post-translational modification of proteins is well known to be a vital component of the DDR, where they are observed to mediate DNA damage signalling and protein-protein interactions. We have observed that APL undergoes mono-ubiquitination, and weak poly-ubiquitination, in response to cisplatin treatment. We have also observed weaker mono-ubiquitination of APL in response to DSBs. These damage-induced post-translational modifications indicate that the function of APL is altered upon DNA damage, further implicating this protein in the DDR in *Dictyostelium*. These data indicate that the mono-ubiquitination and enrichment of APL on chromatin are both stimulated by DNA inter-strand cross-links induced by cisplatin, despite very weak induction of the mono-ubiquitinate species observed in response to 4-NQO treatment. In the FA pathway, FancD2 is mono-ubiquitinated by Ube2T and FancL; however, we have shown that these proteins are not required for APL mono-ubiquitination, indicating that APL may not act in the FA pathway[499,504]. Ubiquitination has also been strongly linked to TLS and HR, which are utilised as sub-pathways in the repair of DNA inter-strand cross-links. In TLS, the ubiquitin E2- and E3-ligases Rad6 and Rad18 mono-ubiquitinate PCNA to facilitate polymerase-switching[190,202]. Orthologs of these proteins have been identified *in silico* in *Dictyostelium*, and they are strong candidates for mediating the mono-ubiquitination of APL. Moreover, RNF8 and Ubc13 have been implicated in ubiquitinating proteins in the

repair of inter-strand cross-links[545]. Furthermore, RNF8 and RNF168 ubiquitinate targets in response to DNA DSBs, leading to the recruitment of HR factors, such as BRCA1[538,541,544]. In HR, BRCA1 acts as ubiquitin E3-ligase in a heterodimer with BARD1, and this function is required for the recruitment of BRCA2 and Rad51[559]. However, orthologs of RNF168 and RNF8 have not been identified in *Dictyostelium*. The mono-ubiquitination of APL may mediate its interaction with other DNA repair factors, which might explain why we did not observe any interacting proteins co-immunoprecipitating with recombinant Myc-APL in response to various forms of DNA damage, as the vast majority of APL is unmodified in these assays. We also observed a cisplatin-induced, high molecular weight form of APL, with was abrogated following deletion of *ube2T*, but not *fncL*. While the function of Ube2T and FncL in mono-ubiquitinating FncD2 is conserved in *Dictyostelium*, the disappearance of the high molecular weight form of APL in *ube2T*¯ cells only suggests that Ube2T has an alternative ubiquitin E3-ligase partner in *Dictyostelium*[347]. However, we have not yet been able to identify what this form of APL is, or its functional importance in the response to DNA damage.

The mono-ubiquitination of APL was observed to be abrogated by deletion of its macro domain. This indicates that either the ubiquitination site of APL is situated within the macro domain, or that an interaction mediated by the macro domain is required for ubiquitination of APL. There have been no previous reports of ubiquitination occurring within, or affecting the function of a macro domain, which suggests further novelty of APL if this is shown to be the case. PARylation and ubiquitination pathways have previously been linked in PARylation-dependent ubiquitination (PARdU). WWE domains have been identified in ubiquitin E3-ligases, such as RNF146, which is recruited to PARylated proteins through its WWE domain, and subsequently ubiquitinates these

proteins, resulting in their degradation[242,285–287]. APL may also be ubiquitinated in a PARylation-, or ADP-ribosylation-dependent manner, if such modifications are shown to be the *in vivo* binding partners of the macro domain of APL. Inactivation of the PBZ domain of APL by mutation did not abrogate the mono-ubiquitination of the protein. However, the role of the FHA domain of APL has not yet been elucidated. The FHA domain was predicted to bind to SSBR and NHEJ proteins in a phospho-specific manner; however, the implication of APL in the repair of DNA inter-strand cross-links or S-phase-associated DSBs suggests that it may bind to novel protein targets. NHEJ has been implicated in the repair of inter-strand cross-links in yeast, but is only required for survival in the absence of HR, which might account for the lack of sensitivity of *apl*$^{-}$ cells to inter-strand cross-links[553]. Nevertheless, we are yet to assess whether interactions mediated by the FHA domain of APL are responsible for the ubiquitination of the protein.

Overall, the search for novel PAR-binding, DNA repair proteins in *Dictyostelium* yielded two candidates that were both shown to bind to PAR *in vitro*, and that gave indications that they are involved in the DDR in experimental assays. Both Lig3 and APL do not appear to be required for the survival of exposure to genotoxic agents in the conditions tested, suggesting some level of redundancy is present. Particularly in the case of APL, the lack of a survival defect of *apl*$^{-}$ cells to DNA inter-strand cross-links is a significant barrier to further investigation of this protein in the DDR. However, both Lig3 and APL are enriched on chromatin following DNA damage, and APL is also ubiquitinated in a DNA damage-dependent manner. While these data are promising, we have not yet ascertained the role of PARylation in the function of these proteins in the DDR, which could provide novel insights into human DNA repair.

8. References

1. Jackson, S. P. & Bartek, J. The DNA-damage response in human biology and disease. *Nature* **461,** 1071–8 (2009).

2. Helleday, T., Petermann, E., Lundin, C., Hodgson, B. & Sharma, R. A. DNA repair pathways as targets for cancer therapy. *Nat. Rev. Cancer* **8,** 193–204 (2008).

3. Rogakou, E. P., Pilch, D. R., Orr, A. H., Ivanova, V. S. & Bonner, W. M. DNA Double-stranded Breaks Induce Histone H2AX Phosphorylation on Serine 139. *J. Biol. Chem.* **273,** 5858–5868 (1998).

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

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

6. Stiff, T. *et al.* ATM and DNA-PK function redundantly to phosphorylate H2AX after exposure to ionizing radiation. *Cancer Res.* **64,** 2390–6 (2004).

7. Paull, T. T. *et al.* A critical role for histone H2AX in recruitment of repair factors to nuclear foci after DNA damage. *Curr. Biol.* **10,** 886–95

8. Bird, A. W. *et al.* Acetylation of histone H4 by Esa1 is required for DNA double-strand break repair. *Nature* **419,** 411–5 (2002).

9. Gupta, A. *et al.* Involvement of human MOF in ATM function. *Mol. Cell. Biol.* **25,** 5292–305 (2005).

10. Alvarez-Gonzalez, R. & Jacobson, M. K. Characterization of polymers of adenosine diphosphate ribose generated in vitro and in vivo. *Biochemistry* **26,** 3218–24 (1987).

11. Hottiger, M. O., Hassa, P. O., Lüscher, B., Schüler, H. & Koch-Nolte, F. Toward a unified nomenclature for mammalian ADP-ribosyltransferases. *Trends Biochem. Sci.* **35,** 208–19 (2010).

12. Alvarez-Gonzalez, R. & Althaus, F. R. Poly(ADP-ribose) catabolism in mammalian cells exposed to DNA-damaging agents. *Mutat. Res. Repair* **218,** 67–74 (1989).

13. Gibson, B. A. & Kraus, W. L. New insights into the molecular and cellular functions of poly(ADP-ribose) and PARPs. *Nat. Rev. Mol. Cell Biol.* **13,** 411–24 (2012).

14. Caldecott, K. W. Single-strand break repair and genetic disease. *Nat. Rev. Genet.* **9,** 619–31 (2008).

15. Rulten, S. L. *et al.* PARP-3 and APLF Function Together to Accelerate Non-Homologous End Joining. *Molecular Cell* (2010).

16. Bryant, H. E. *et al.* PARP is activated at stalled forks to mediate Mre11-dependent replication restart and recombination. *EMBO J.* **28,** 2601–15 (2009).

17. Schreiber, V., Dantzer, F., Ame, J.-C. & de Murcia, G. Poly(ADP-ribose): novel functions for an old molecule. *Nat. Rev. Mol. Cell Biol.* **7,** 517–28 (2006).

18. Holbourn, K. P., Shone, C. C. & Acharya, K. R. A family of killer toxins. Exploring the mechanism of ADP-ribosylating toxins. *FEBS J.* **273,** 4579–93 (2006).

19. Bell, C. E. & Eisenberg, D. Crystal structure of diphtheria toxin bound to nicotinamide adenine dinucleotide. *Biochemistry* **35,** 1137–49 (1996).

20. Penning, T. D. *et al.* Optimization of phenyl-substituted benzimidazole carboxamide poly(ADP-ribose) polymerase inhibitors: identification of (S)-2-(2-fluoro-4-(pyrrolidin-2-yl)phenyl)-1H-benzimidazole-4-carboxamide (A-966492), a highly potent and efficacious inhibitor. *J. Med. Chem.* **53,** 3142–53 (2010).

21. Vyas, S. & Chang, P. New PARP targets for cancer therapy. *Nat. Rev. Cancer* **14,** 502–509 (2014).

22. Marsischky, G. T., Wilson, B. A. & Collier, R. J. Role of glutamic acid 988 of human poly-ADP-ribose polymerase in polymer formation. Evidence for active site similarities to the ADP-ribosylating toxins. *J. Biol. Chem.* **270,** 3247–54 (1995).

23. Rolli, V., O'Farrell, M., Ménissier-de Murcia, J. & de Murcia, G. Random mutagenesis of the poly(ADP-ribose) polymerase catalytic domain reveals amino acids involved in polymer branching. *Biochemistry* **36,** 12147–54 (1997).

24. Desmarais, Y., Ménard, L., Lagueux, J. & Poirier, G. G. Enzymological properties of poly(ADP-ribose)polymerase: characterization of automodification sites and NADase activity. *Biochim. Biophys. Acta* **1078,** 179–86 (1991).

25. Ame, J.-C. *et al.* PARP-2, A Novel Mammalian DNA Damage-dependent Poly(ADP-ribose) Polymerase. *J. Biol. Chem.* **274,** 17860–17868 (1999).

26. Loseva, O. *et al.* PARP-3 is a mono-ADP-ribosylase that activates PARP-1 in the absence of DNA. *J. Biol. Chem.* **285,** 8054–60 (2010).

27. Rippmann, J. F., Damm, K. & Schnapp, A. Functional characterization of the poly(ADP-ribose) polymerase activity of tankyrase 1, a potential regulator of telomere length. *J. Mol. Biol.* **323,** 217–24 (2002).

28. Aguiar, R. C. T., Takeyama, K., He, C., Kreinbrink, K. & Shipp, M. A. B-aggressive lymphoma family proteins have unique domains that modulate

transcription and exhibit poly(ADP-ribose) polymerase activity. *J. Biol. Chem.* **280,** 33756–65 (2005).

29. Kleine, H. *et al.* Substrate-assisted catalysis by PARP10 limits its activity to mono-ADP-ribosylation. *Mol. Cell* **32,** 57–69 (2008).

30. Di Paola, S., Micaroni, M., Di Tullio, G., Buccione, R. & Di Girolamo, M. PARP16/ARTD15 is a novel endoplasmic-reticulum-associated mono-ADP-ribosyltransferase that interacts with, and modifies karyopherin-ß1. *PLoS One* **7,** e37352 (2012).

31. Wahlberg, E. *et al.* Family-wide chemical profiling and structural analysis of PARP and tankyrase inhibitors. *Nat. Biotechnol.* **30,** 283–8 (2012).

32. Ruf, A., Rolli, V., de Murcia, G. & Schulz, G. E. The mechanism of the elongation and branching reaction of poly(ADP-ribose) polymerase as derived from crystal structures and mutagenesis. *J. Mol. Biol.* **278,** 57–65 (1998).

33. Vyas, S. *et al.* Family-wide analysis of poly(ADP-ribose) polymerase activity. *Nat. Commun.* **5,** 4426 (2014).

34. El-Khamisy, S. F., Masutani, M., Suzuki, H. & Caldecott, K. W. A requirement for PARP-1 for the assembly or stability of XRCC1 nuclear foci at sites of oxidative DNA damage. *Nucleic Acids Res.* **31,** 5526–33 (2003).

35. Nicolae, C. M. *et al.* A novel role for the mono-ADP-ribosyltransferase PARP14/ARTD8 in promoting homologous recombination and protecting against replication stress. *Nucleic Acids Res.* **43,** 3143–53 (2015).

36. Yan, Q. *et al.* BAL1 and its partner E3 ligase, BBAP, link Poly(ADP-ribose) activation, ubiquitylation, and double-strand DNA repair independent of ATM, MDC1, and RNF8. *Mol. Cell. Biol.* **33,** 845–57 (2013).

37. Nicolae, C. M. *et al.* The ADP-ribosyltransferase PARP10/ARTD10 interacts with proliferating cell nuclear antigen (PCNA) and is required for DNA damage tolerance. *J. Biol. Chem.* **289,** 13627–37 (2014).

38. Mehrotra, P. *et al.* PARP-14 functions as a transcriptional switch for Stat6-dependent gene activation. *J. Biol. Chem.* **286,** 1767–76 (2011).

39. Verheugd, P. *et al.* Regulation of NF-κB signalling by the mono-ADP-ribosyltransferase ARTD10. *Nat. Commun.* **4,** 1683 (2013).

40. Yu, M. *et al.* PARP-10, a novel Myc-interacting protein with poly(ADP-ribose) polymerase activity, inhibits transformation. *Oncogene* **24,** 1982–93 (2005).

41. Cook, B. D., Dynek, J. N., Chang, W., Shostak, G. & Smith, S. Role for the related poly(ADP-Ribose) polymerases tankyrase 1 and 2 at human telomeres. *Mol. Cell. Biol.* **22,** 332–42 (2002).

42. Zhu, Y. *et al.* Zinc-finger antiviral protein inhibits HIV-1 infection by selectively targeting multiply spliced viral mRNAs for degradation. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 15834–9 (2011).

43. Jwa, M. & Chang, P. PARP16 is a tail-anchored endoplasmic reticulum protein required for the PERK- and IRE1α-mediated unfolded protein response. *Nat. Cell Biol.* **14,** 1223–30 (2012).

44. Oka, S., Kato, J. & Moss, J. Identification and characterization of a mammalian 39-kDa poly(ADP-ribose) glycohydrolase. *J. Biol. Chem.* **281,** 705–13 (2006).

45. Slade, D. *et al.* The structure and catalytic mechanism of a poly(ADP-ribose) glycohydrolase. *Nature* **477,** 616–20 (2011).

46. Mueller-Dieckmann, C. *et al.* The structure of human ADP-ribosylhydrolase 3 (ARH3) provides insights into the reversibility of protein ADP-ribosylation. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 15026–31 (2006).

47. Campagnari, F., Whitfield, J. F. & Bertazzoni, U. The effect of x-irradiation on the nicotinamide adenine dinucleotides (NAD-NADH) content of rat thymocytes. *Exp. Cell Res.* **42,** 646–56 (1966).

48. Goodwin, P. M., Lewis, P. J., Davies, M. I., Skidmore, C. J. & Shall, S. The effect of gamma radiation and neocarzinostatin on NAD and ATP levels in mouse leukaemia cells. *Biochim. Biophys. Acta* **543,** 576–82 (1978).

49. Davies, M. I., Halldorsson, H., Nduka, N., Shall, S. & Skidmore, C. J. The involvement of poly(adenosine diphosphate-ribose) in deoxyribonucleic acid repair. *Biochem. Soc. Trans.* **6,** 1056–7 (1978).

50. Whish, W. J., Davies, M. I. & Shall, S. Stimulation of poly(ADP-ribose) polymerase activity by the anti-tumour antibiotic, streptozotocin. *Biochem. Biophys. Res. Commun.* **65,** 722–30 (1975).

51. Durkacz, B. W., Omidiji, O., Gray, D. A. & Shall, S. (ADP-ribose)n participates in DNA excision repair. *Nature* **283,** 593–6 (1980).

52. Poirier, G. G. Poly(ADP-Ribosyl)ation of Polynucleosomes Causes Relaxation of Chromatin Structure. *Proc. Natl. Acad. Sci.* **79,** 3423–3427 (1982).

53. Mathis, G. & Althaus, F. R. Release of core DNA from nucleosomal core particles following (ADP-ribose)n-modification in vitro. *Biochem. Biophys. Res. Commun.* **143,** 1049–54 (1987).

54. Schreiber, V. *et al.* Poly(ADP-ribose) polymerase-2 (PARP-2) is required for efficient base excision DNA repair in association with PARP-1 and XRCC1. *J. Biol. Chem.* **277,** 23028–36 (2002).

55. Pogozelski, W. K. & Tullius, T. D. Oxidative Strand Scission of Nucleic Acids: Routes Initiated by Hydrogen Abstraction from the Sugar Moiety. *Chem. Rev.* **98,** 1089–1108 (1998).

56. Hegde, M. L., Hazra, T. K. & Mitra, S. Early steps in the DNA base excision/single-strand interruption repair pathway in mammalian cells. *Cell Res.* **18,** 27–47 (2008).

57. Wang, J. C. Cellular roles of DNA topoisomerases: a molecular perspective. *Nat. Rev. Mol. Cell Biol.* **3,** 430–40 (2002).

58. Lindahl, T. An N-Glycosidase from Escherichia coli That Releases Free Uracil from DNA Containing Deaminated Cytosine Residues. *Proc. Natl. Acad. Sci.* **71,** 3649–3653 (1974).

59. Jacobs, A. L. & Schär, P. DNA glycosylases: in DNA repair and beyond. *Chromosoma* **121,** 1–20 (2012).

60. Wilson, D. M., Takeshita, M., Grollman, A. P. & Demple, B. Incision activity of human apurinic endonuclease (Ape) at abasic site analogs in DNA. *J. Biol. Chem.* **270,** 16002–7 (1995).

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

62. Marenstein, D. R., Wilson, D. M. & Teebor, G. W. Human AP endonuclease (APE1) demonstrates endonucleolytic activity against AP sites in single-stranded DNA. *DNA Repair (Amst).* **3,** 527–33 (2004).

63. Krokan, H. E., Standal, R. & Slupphaug, G. DNA glycosylases in the base excision repair of DNA. *Biochem. J.* **325 (Pt 1,** 1–16 (1997).

64. Krokan, H. E., Nilsen, H., Skorpen, F., Otterlei, M. & Slupphaug, G. Base excision repair of DNA in mammalian cells. *FEBS Lett.* **476,** 73–77 (2000).

65. Satoh, M. S. & Lindahl, T. Role of poly(ADP-ribose) formation in DNA repair. *Nature* **356,** 356–8 (1992).

66. Chou, D. M. *et al.* A chromatin localization screen reveals poly (ADP ribose)-regulated recruitment of the repressive polycomb and NuRD complexes to sites of DNA damage. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 18475–80 (2010).

67. Masson, M. *et al.* XRCC1 Is Specifically Associated with Poly(ADP-Ribose) Polymerase and Negatively Regulates Its Activity following DNA Damage. *Mol. Cell. Biol.* **18,** 3563–71 (1998).

68. Caldecott, K. W., Aoufouchi, S., Johnson, P. & Shall, S. XRCC1 polypeptide interacts with DNA polymerase beta and possibly poly (ADP-ribose) polymerase, and DNA ligase III is a novel molecular 'nick-sensor' in vitro. *Nucleic Acids Res.* **24,** 4387–4394 (1996).

69. Caldecott, K. W., McKeown, C. K., Tucker, J. D., Ljungquist, S. & Thompson, L. H. An interaction between the mammalian DNA repair protein XRCC1 and DNA ligase III. *Mol. Cell. Biol.* **14,** 68–76 (1994).

70. Caldecott, K. W. XRCC1 and DNA strand break repair. *DNA Repair (Amst).* **2,** 955–969 (2003).

71. Parsons, J. L., Dianova, I. I. & Dianov, G. L. APE1 is the major 3'-phosphoglycolate activity in human cell extracts. *Nucleic Acids Res.* **32,** 3531–6 (2004).

72. Izumi, T. *et al.* Requirement for human AP endonuclease 1 for repair of 3'-blocking damage at DNA single-strand breaks induced by reactive oxygen species. *Carcinogenesis* **21,** 1329–34 (2000).

73. Inamdar, K. V *et al.* Conversion of phosphoglycolate to phosphate termini on 3' overhangs of DNA double strand breaks by the human tyrosyl-DNA phosphodiesterase hTdp1. *J. Biol. Chem.* **277,** 27162–8 (2002).

74. Ahel, I. *et al.* The neurodegenerative disease protein aprataxin resolves abortive DNA ligation intermediates. *Nature* **443,** 713–6 (2006).

75. Karimi-Busheri, F. *et al.* Molecular Characterization of a Human DNA Kinase. *J. Biol. Chem.* **274,** 24187–24194 (1999).

76. Jilani, A. *et al.* Molecular cloning of the human gene, PNKP, encoding a polynucleotide kinase 3'-phosphatase and evidence for its role in repair of DNA strand breaks caused by oxidative damage. *J. Biol. Chem.* **274,** 24176–24186 (1999).

77. Raymond, A. C., Staker, B. L. & Burgin, A. B. Substrate specificity of tyrosyl-DNA phosphodiesterase I (Tdp1). *J. Biol. Chem.* **280,** 22029–35 (2005).

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

79. Pascucci, B., Russo, M. T., Crescenzi, M., Bignami, M. & Dogliotti, E. The accumulation of MMS-induced single strand breaks in G1 phase is recombinogenic in DNA polymerase beta defective mammalian cells. *Nucleic Acids Res.* **33,** 280–8 (2005).

80. Fortini, P., Pascucci, B., Belisario, F. & Dogliotti, E. DNA polymerase beta is required for efficient DNA strand break repair induced by methyl methanesulfonate but not by hydrogen peroxide. *Nucleic Acids Res.* **28,** 3040–6 (2000).

81. Cağlayan, M., Batra, V. K., Sassa, A., Prasad, R. & Wilson, S. H. Role of polymerase β in complementing aprataxin deficiency during abasic-site base excision repair. *Nat. Struct. Mol. Biol.* **21,** 497–9 (2014).

82. Fortini, P. Two Pathways for Base Excision Repair in Mammalian Cells. *J. Biol. Chem.* **271,** 9573–9578 (1996).

83. Prasad, R., Dianov, G. L., Bohr, V. A. & Wilson, S. H. FEN1 stimulation of DNA polymerase beta mediates an excision step in mammalian long patch base excision repair. *J. Biol. Chem.* **275,** 4460–6 (2000).

84. Tom, S., Henricksen, L. A., Park, M. S. & Bambara, R. A. DNA ligase I and proliferating cell nuclear antigen form a functional complex. *J. Biol. Chem.* **276,** 24817–25 (2001).

85. Pascucci, B. *et al.* Reconstitution of the base excision repair pathway for 7,8-dihydro-8-oxoguanine with purified human proteins. *Nucleic Acids Res.* **30,** 2124–30 (2002).

86. Abdou, I., Poirier, G. G., Hendzel, M. J. & Weinfeld, M. DNA ligase III acts as a DNA strand break sensor in the cellular orchestration of DNA strand break repair. *Nucleic Acids Res.* **43,** 875–92 (2015).

87. Lakshmipathy, U. & Campbell, C. The human DNA ligase III gene encodes nuclear and mitochondrial proteins. *Mol. Cell. Biol.* **19,** 3869–76 (1999).

88. Simsek, D. *et al.* Crucial role for DNA ligase III in mitochondria but not in Xrcc1-dependent repair. *Nature* **471,** 245–8 (2011).

89. Gao, Y. *et al.* DNA ligase III is critical for mtDNA integrity but not Xrcc1-mediated nuclear DNA repair. *Nature* **471,** 240–4 (2011).

90. De Murcia, J. M. *et al.* Requirement of poly(ADP-ribose) polymerase in recovery from DNA damage in mice and in cells. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 7303–7 (1997).

91. Wang, Z.-Q. *et al.* PARP is important for genomic stability but dispensable in apoptosis. *Genes Dev.* **11,** 2347–2358 (1997).

92. Trucco, C., Oliver, F. J., de Murcia, G. & Ménissier-de Murcia, J. DNA repair defect in poly(ADP-ribose) polymerase-deficient cell lines. *Nucleic Acids Res.* **26,** 2644–9 (1998).

93. Park, S. D., Kim, C. G. & Kim, M. G. Inhibitors of poly(ADP-ribose) polymerase enhance DNA strand breaks, excision repair, and sister chromatid exchanges induced by alkylating agents. *Environ. Mutagen.* **5,** 515–25 (1983).

94. Godon, C. *et al.* PARP inhibition versus PARP-1 silencing: different outcomes in terms of single-strand break repair and radiation susceptibility. *Nucleic Acids Res.* **36,** 4454–4464 (2008).

95. Fisher, A. E. O., Hochegger, H., Takeda, S. & Caldecott, K. W. Poly(ADP-ribose) polymerase 1 accelerates single-strand break repair in concert with poly(ADP-ribose) glycohydrolase. *Mol. Cell. Biol.* **27,** 5597–605 (2007).

96. Eustermann, S. *et al.* The DNA-binding domain of human PARP-1 interacts with DNA single-strand breaks as a monomer through its second zinc finger. *J. Mol. Biol.* **407,** 149–70 (2011).

97. Langelier, M.-F., Planck, J. L., Roy, S. & Pascal, J. M. Structural basis for DNA damage-dependent poly(ADP-ribosyl)ation by human PARP-1. *Science* **336,** 728–32 (2012).

98. Langelier, M.-F. & Pascal, J. M. PARP-1 mechanism for coupling DNA damage detection to poly(ADP-ribose) synthesis. *Curr. Opin. Struct. Biol.* (2013). doi:10.1016/j.sbi.2013.01.003

99. Chapman, J. D., Gagné, J.-P., Poirier, G. G. & Goodlett, D. R. Mapping PARP-1 auto-ADP-ribosylation sites by liquid chromatography-tandem mass spectrometry. *J. Proteome Res.* **12,** 1868–80 (2013).

100. Ménissier de Murcia, J. *et al.* Functional interaction between PARP-1 and PARP-2 in chromosome stability and embryonic development in mouse. *EMBO J.* **22,** 2255–63 (2003).

101. Kutuzov, M. M. *et al.* Interaction of PARP-2 with DNA structures mimicking DNA repair intermediates and consequences on activity of base excision repair proteins. *Biochimie* **null,** (2013).

102. Mortusewicz, O., Amé, J.-C., Schreiber, V. & Leonhardt, H. Feedback-regulated poly(ADP-ribosyl)ation by PARP-1 is required for rapid response to DNA damage in living cells. *Nucleic Acids Res.* **35,** 7665–75 (2007).

103. Khanna, K. K. & Jackson, S. P. DNA double-strand breaks: signaling, repair and the cancer connection. *Nat. Genet.* **27,** 247–54 (2001).

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

105. Sadofsky, M. J. The RAG proteins in V(D)J recombination: more than just a nuclease. *Nucleic Acids Res.* **29,** 1399–1409 (2001).

106. Yan, C. T. *et al.* IgH class switching and translocations use a robust non-classical end-joining pathway. *Nature* **449,** 478–82 (2007).

107. Corneo, B. *et al.* Rag mutations reveal robust alternative end joining. *Nature* **449,** 483–6 (2007).

108. Taccioli, G. *et al.* Impairment of V(D)J recombination in double-strand break repair mutants. *Science (80-.).* **260,** 207–210 (1993).

109. Blunt, T. *et al.* Defective DNA-dependent protein kinase activity is linked to V(D)J recombination and DNA repair defects associated with the murine scid mutation. *Cell* **80,** 813–823 (1995).

110. Rooney, S., Chaudhuri, J. & Alt, F. W. The role of the non-homologous end-joining pathway in lymphocyte development. *Immunol. Rev.* **200,** 115–31 (2004).

111. Keeney, S. & Neale, M. J. Initiation of meiotic recombination by formation of DNA double-strand breaks: mechanism and regulation. *Biochem. Soc. Trans.* **34,** 523–5 (2006).

112. Kadyk, L. C. & Hartwell, L. H. Sister chromatids are preferred over homologs as substrates for recombinational repair in Saccharomyces cerevisiae. *Genetics* **132,** 387–402 (1992).

113. Takata, M. *et al.* Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. *EMBO J.* **17,** 5497–508 (1998).

114. Mirzoeva, O. K. & Petrini, J. H. DNA damage-dependent nuclear dynamics of the Mre11 complex. *Mol. Cell. Biol.* **21,** 281–8 (2001).

115. Paull, T. T. & Gellert, M. The 3′ to 5′ Exonuclease Activity of Mre11 Facilitates Repair of DNA Double-Strand Breaks. *Mol. Cell* **1,** 969–979 (1998).

116. Anderson, D. E., Trujillo, K. M., Sung, P. & Erickson, H. P. Structure of the Rad50 x Mre11 DNA repair complex from Saccharomyces cerevisiae by electron microscopy. *J. Biol. Chem.* **276,** 37027–33 (2001).

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

118. De Jager, M. *et al.* Human Rad50/Mre11 is a flexible complex that can tether DNA ends. *Mol. Cell* **8,** 1129–35 (2001).

119. Tauchi, H., Matsuura, S., Kobayashi, J., Sakamoto, S. & Komatsu, K. Nijmegen breakage syndrome gene, NBS1, and molecular links to factors for genome stability. *Oncogene* **21,** 8967–80 (2002).

120. Kobayashi, J. *et al.* NBS1 localizes to gamma-H2AX foci through interaction with the FHA/BRCT domain. *Curr. Biol.* **12,** 1846–51 (2002).

121. Uziel, T. *et al.* Requirement of the MRN complex for ATM activation by DNA damage. *EMBO J.* **22,** 5612–21 (2003).

122. You, Z. & Bailis, J. M. DNA damage and decisions: CtIP coordinates DNA repair and cell cycle checkpoints. *Trends Cell Biol.* **20,** 402–9 (2010).

123. You, Z. *et al.* CtIP Links DNA Double-Strand Break Sensing to Resection. *Mol. Cell* **36,** 954–969 (2009).

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

125. Fiorentini, P., Huang, K. N., Tishkoff, D. X., Kolodner, R. D. & Symington, L. S. Exonuclease I of Saccharomyces cerevisiae functions in mitotic recombination in vivo and in vitro. *Mol. Cell. Biol.* **17,** 2764–73 (1997).

126. Lee, B. I. & Wilson, D. M. The RAD2 domain of human exonuclease 1 exhibits 5' to 3' exonuclease and flap structure-specific endonuclease activities. *J. Biol. Chem.* **274,** 37763–9 (1999).

127. Zhu, Z., Chung, W.-H., Shim, E. Y., Lee, S. E. & Ira, G. Sgs1 helicase and two nucleases Dna2 and Exo1 resect DNA double-strand break ends. *Cell* **134,** 981–94 (2008).

128. Sleeth, K. M. *et al.* RPA mediates recombination repair during replication stress and is displaced from DNA by checkpoint signalling in human cells. *J. Mol. Biol.* **373,** 38–47 (2007).

129. Sugiyama, T., Zaitseva, E. M. & Kowalczykowski, S. C. A single-stranded DNA-binding protein is needed for efficient presynaptic complex formation by the Saccharomyces cerevisiae Rad51 protein. *J. Biol. Chem.* **272,** 7940–5 (1997).

130. Sung, P. Catalysis of ATP-dependent homologous DNA pairing and strand exchange by yeast RAD51 protein. *Science* **265,** 1241–3 (1994).

131. Conway, A. B. *et al.* Crystal structure of a Rad51 filament. *Nat. Struct. Mol. Biol.* **11,** 791–6 (2004).

132. Park, M. S., Ludwig, D. L., Stigger, E. & Lee, S. H. Physical interaction between human RAD52 and RPA is required for homologous recombination in mammalian cells. *J. Biol. Chem.* **271,** 18996–9000 (1996).

133. Sugawara, N., Wang, X. & Haber, J. E. In vivo roles of Rad52, Rad54, and Rad55 proteins in Rad51-mediated recombination. *Mol. Cell* **12,** 209–19 (2003).

134. Tarsounas, M., Davies, D. & West, S. C. BRCA2-dependent and independent formation of RAD51 nuclear foci. *Oncogene* **22,** 1115–23 (2003).

135. Yang, H., Li, Q., Fan, J., Holloman, W. K. & Pavletich, N. P. The BRCA2 homologue Brh2 nucleates RAD51 filament formation at a dsDNA-ssDNA junction. *Nature* **433,** 653–7 (2005).

136. Yang, H. *et al.* BRCA2 function in DNA binding and recombination from a BRCA2-DSS1-ssDNA structure. *Science* **297,** 1837–48 (2002).

137. Wooster, R. *et al.* Identification of the breast cancer susceptibility gene BRCA2. *Nature* **378,** 789–92 (1995).

138. Petukhova, G., Stratton, S. & Sung, P. Catalysis of homologous DNA pairing by yeast Rad51 and Rad54 proteins. *Nature* **393,** 91–4 (1998).

139. Nassif, N., Penney, J., Pal, S., Engels, W. R. & Gloor, G. B. Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair. *Mol. Cell. Biol.* **14,** 1613–25 (1994).

140. Malkova, A., Ivanov, E. L. & Haber, J. E. Double-strand break repair in the absence of RAD51 in yeast: a possible role for break-induced DNA replication. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 7131–6 (1996).

141. Wu, L. & Hickson, I. D. The Bloom's syndrome helicase suppresses crossing over during homologous recombination. *Nature* **426,** 870–4 (2003).

142. Ip, S. C. Y. *et al.* Identification of Holliday junction resolvases from humans and yeast. *Nature* **456,** 357–61 (2008).

143. Gaskell, L. J., Osman, F., Gilbert, R. J. C. & Whitby, M. C. Mus81 cleavage of Holliday junctions: a failsafe for processing meiotic recombination intermediates? *EMBO J.* **26,** 1891–901 (2007).

144. Mimori, T. & Hardin, J. A. Mechanism of interaction between Ku protein and DNA. *J. Biol. Chem.* **261,** 10375–9 (1986).

145. Walker, J. R., Corpina, R. A. & Goldberg, J. Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature* **412,** 607–14 (2001).

146. Smith, G. C., Divecha, N., Lakin, N. D. & Jackson, S. P. DNA-dependent protein kinase and related proteins. *Biochem. Soc. Symp.* **64,** 91–104 (1999).

147. Hammel, M. *et al.* Ku and DNA-dependent protein kinase dynamic conformations and assembly regulate DNA binding and the initial non-homologous end joining complex. *J. Biol. Chem.* **285,** 1414–23 (2010).

148. Thompson, L. H. Recognition, signaling, and repair of DNA double-strand breaks produced by ionizing radiation in mammalian cells: The molecular choreography. *Mutat. Res.* **751,** 246–158 (2012).

149. An, J. *et al.* DNA-PKcs plays a dominant role in the regulation of H2AX phosphorylation in response to DNA damage and cell cycle progression. *BMC Mol. Biol.* **11,** 18 (2010).

150. Shirodkar, P., Fenton, A. L., Meng, L. & Koch, C. A. Identification and functional characterization of a Ku-binding motif in aprataxin polynucleotide kinase/phosphatase-like factor (APLF). *J. Biol. Chem.* **288,** 19604–13 (2013).

151. Grundy, G. J. *et al.* APLF promotes the assembly and activity of non-homologous end joining protein complexes. *EMBO J.* **32,** 112–125 (2012).

152. Roberts, S. A. *et al.* Ku is a 5′-dRP/AP lyase that excises nucleotide damage near broken ends. *Nature* **464,** 1214–1217 (2010).

153. Bahmed, K., Seth, A., Nitiss, K. C. & Nitiss, J. L. End-processing during non-homologous end-joining: a role for exonuclease 1. *Nucleic Acids Res.* **39,** 970–8 (2011).

154. Xie, A., Kwok, A. & Scully, R. Role of mammalian Mre11 in classical and alternative nonhomologous end joining. *Nat. Struct. Mol. Biol.* **16,** 814–8 (2009).

155. Moshous, D. *et al.* Artemis, a Novel DNA Double-Strand Break Repair/V(D)J Recombination Protein, Is Mutated in Human Severe Combined Immune Deficiency. *Cell* **105,** 177–186 (2001).

156. Weterings, E. *et al.* The Ku80 carboxy terminus stimulates joining and artemis-mediated processing of DNA ends. *Mol. Cell. Biol.* **29,** 1134–42 (2009).

157. Kusumoto, R. *et al.* Werner protein cooperates with the XRCC4-DNA ligase IV complex in end-processing. *Biochemistry* **47,** 7548–56 (2008).

158. Ramsden, D. A. & Asagoshi, K. DNA polymerases in nonhomologous end joining: are there any benefits to standing out from the crowd? *Environ. Mol. Mutagen.* **53,** 741–51 (2012).

159. Hammel, M. *et al.* XRCC4 protein interactions with XRCC4-like factor (XLF) create an extended grooved scaffold for DNA ligation and double strand break repair. *J. Biol. Chem.* **286,** 32638–50 (2011).

160. Gu, J. *et al.* XRCC4:DNA ligase IV can ligate incompatible DNA ends and can ligate across gaps. *EMBO J.* **26,** 1010–23 (2007).

161. Clements, P. M. *et al.* The ataxia-oculomotor apraxia 1 gene product has a role distinct from ATM and interacts with the DNA strand break repair proteins XRCC1 and XRCC4. *DNA Repair (Amst).* **3,** 1493–502 (2004).

162. Iles, N., Rulten, S., El-Khamisy, S. F. & Caldecott, K. W. APLF (C2orf13) is a novel human protein involved in the cellular response to chromosomal DNA strand breaks. *Mol. Cell. Biol.* **27,** 3793–803 (2007).

163. Mani, R. S. *et al.* Dual modes of interaction between XRCC4 and polynucleotide kinase/phosphatase: implications for nonhomologous end joining. *J. Biol. Chem.* **285,** 37619–29 (2010).

164. Wu, P.-Y. *et al.* Structural and functional interaction between the human DNA repair proteins DNA ligase IV and XRCC4. *Mol. Cell. Biol.* **29,** 3163–72 (2009).

165. Riballo, E. *et al.* XLF-Cernunnos promotes DNA ligase IV-XRCC4 re-adenylation following ligation. *Nucleic Acids Res.* **37,** 482–92 (2009).

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

167. Lu, H., Pannicke, U., Schwarz, K. & Lieber, M. R. Length-dependent binding of human XLF to DNA and stimulation of XRCC4.DNA ligase IV activity. *J. Biol. Chem.* **282,** 11155–62 (2007).

168. Ochi, T. *et al.* DNA repair. PAXX, a paralog of XRCC4 and XLF, interacts with Ku to promote DNA double-strand break repair. *Science* **347,** 185–8 (2015).

169. Wu, W. *et al.* Repair of radiation induced DNA double strand breaks by backup NHEJ is enhanced in G2. *DNA Repair (Amst).* **7,** 329–38 (2008).

170. Wu, W., Wang, M., Mussfeldt, T. & Iliakis, G. Enhanced Use of Backup Pathways of NHEJ in G 2 in Chinese Hamster Mutant Cells with Defects in the Classical Pathway of NHEJ. *Radiat. Res.* **170,** 512–520 (2008).

171. Soni, A. *et al.* Requirement for Parp-1 and DNA ligases 1 or 3 but not of Xrcc1 in chromosomal translocation formation by backup end joining. *Nucleic Acids Res.* **42,** 6380–92 (2014).

172. Iliakis, G. Backup pathways of NHEJ in cells of higher eukaryotes: cell cycle dependence. *Radiother. Oncol.* **92,** 310–5 (2009).

173. Boulton, S. J. & Jackson, S. P. Saccharomyces cerevisiae Ku70 potentiates illegitimate DNA double-strand break repair and serves as a barrier to error-prone DNA repair pathways. *EMBO J.* **15,** 5093–103 (1996).

174. Jones, R. E. *et al.* Escape from telomere-driven crisis is DNA ligase III dependent. *Cell Rep.* **8,** 1063–76 (2014).

175. Rai, R. *et al.* The function of classical and alternative non-homologous end-joining pathways in the fusion of dysfunctional telomeres. *EMBO J.* **29,** 2598–610 (2010).

176. Chan, S. H., Yu, A. M. & McVey, M. Dual roles for DNA polymerase theta in alternative end-joining repair of double-strand breaks in Drosophila. *PLoS Genet.* **6,** e1001005 (2010).

177. Ceccaldi, R. *et al.* Homologous-recombination-deficient tumours are dependent on Polθ-mediated repair. *Nature* **518,** 258–62 (2015).

178. Mateos-Gomez, P. A. *et al.* Mammalian polymerase θ promotes alternative NHEJ and suppresses recombination. *Nature* **518,** 254–7 (2015).

179. Paul, K. *et al.* DNA ligases I and III cooperate in alternative non-homologous end-joining in vertebrates. *PLoS One* **8,** e59505 (2013).

180. Simsek, D. *et al.* DNA ligase III promotes alternative nonhomologous end-joining during chromosomal translocation formation. *PLoS Genet.* **7,** e1002080 (2011).

181. Della-Maria, J. *et al.* Human Mre11/human Rad50/Nbs1 and DNA ligase IIIalpha/XRCC1 protein complexes act together in an alternative nonhomologous end joining pathway. *J. Biol. Chem.* **286,** 33845–53 (2011).

182. Zhang, Y. & Jasin, M. An essential role for CtIP in chromosomal translocation formation through an alternative end-joining pathway. *Nat. Struct. Mol. Biol.* **18,** 80–4 (2011).

183. Arnaudeau, C., Lundin, C. & Helleday, T. DNA double-strand breaks associated with replication forks are predominantly repaired by homologous recombination involving an exchange mechanism in mammalian cells. *J. Mol. Biol.* **307,** 1235–1245 (2001).

184. Min, W. *et al.* Poly(ADP-ribose) binding to Chk1 at stalled replication forks is required for S-phase checkpoint activation. *Nat. Commun.* **4,** 2993 (2013).

185. Yang, Y.-G., Cortes, U., Patnaik, S., Jasin, M. & Wang, Z.-Q. Ablation of PARP-1 does not interfere with the repair of DNA double-strand breaks, but compromises the reactivation of stalled replication forks. *Oncogene* **23,** 3872–82 (2004).

186. Beck, C. *et al.* PARP3 affects the relative contribution of homologous recombination and nonhomologous end-joining pathways. *Nucleic Acids Res.* **42,** 5616–32 (2014).

187. Hochegger, H. *et al.* Parp-1 protects homologous recombination from interference by Ku and Ligase IV in vertebrate cells. *EMBO J.* **25,** 1305–14 (2006).

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

189. Wang, M. *et al.* PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways. *Nucleic Acids Res.* **34,** 6170–82 (2006).

190. Lehmann, A. R. *et al.* Translesion synthesis: Y-family polymerases and the polymerase switch. *DNA Repair (Amst).* **6,** 891–9 (2007).

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

192. Gibbs, P. E. M. *et al.* The function of the human homolog of Saccharomyces cerevisiae REV1 is required for mutagenesis induced by UV light. *Proc. Natl. Acad. Sci.* **97,** 4186–4191 (2000).

193. Zhang, Y. Response of human REV1 to different DNA damage: preferential dCMP insertion opposite the lesion. *Nucleic Acids Res.* **30,** 1630–1638 (2002).

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

195. Gerlach, V. L., Feaver, W. J., Fischhaber, P. L. & Friedberg, E. C. Purification and characterization of pol kappa, a DNA polymerase encoded by the human DINB1 gene. *J. Biol. Chem.* **276,** 92–8 (2001).

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

197. Masutani, C. *et al.* The XPV (xeroderma pigmentosum variant) gene encodes human DNA polymerase eta. *Nature* **399,** 700–4 (1999).

198. Johnson, R. E., Haracska, L., Prakash, S. & Prakash, L. Role of DNA polymerase eta in the bypass of a (6-4) TT photoproduct. *Mol. Cell. Biol.* **21,** 3558–63 (2001).

199. Zhang, Y., Yuan, F., Wu, X. & Wang, Z. Preferential Incorporation of G Opposite Template T by the Low-Fidelity Human DNA Polymerase iota. *Mol. Cell. Biol.* **20,** 7099–7108 (2000).

200. Kirouac, K. N. & Ling, H. Unique active site promotes error-free replication opposite an 8-oxo-guanine lesion by human DNA polymerase iota. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 3210–5 (2011).

201. Johnson, R. E., Washington, M. T., Haracska, L., Prakash, S. & Prakash, L. Eukaryotic polymerases iota and zeta act sequentially to bypass DNA lesions. *Nature* **406,** 1015–9 (2000).

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

203. Stelter, P. & Ulrich, H. D. Control of spontaneous and damage-induced mutagenesis by SUMO and ubiquitin conjugation. *Nature* **425,** 188–91 (2003).

204. Guo, C. *et al.* REV1 protein interacts with PCNA: significance of the REV1 BRCT domain in vitro and in vivo. *Mol. Cell* **23,** 265–71 (2006).

205. Bienko, M. *et al.* Ubiquitin-binding domains in Y-family polymerases regulate translesion synthesis. *Science* **310,** 1821–4 (2005).

206. Kannouche, P. L., Wing, J. & Lehmann, A. R. Interaction of Human DNA Polymerase η with Monoubiquitinated PCNA. *Mol. Cell* **14,** 491–500 (2004).

207. Bomar, M. G. *et al.* Unconventional ubiquitin recognition by the ubiquitin-binding motif within the Y family DNA polymerases iota and Rev1. *Mol. Cell* **37,** 408–17 (2010).

208. Plosky, B. S. *et al.* Controlling the subcellular localization of DNA polymerases iota and eta via interactions with ubiquitin. *EMBO J.* **25,** 2847–55 (2006).

209. Garg, P. & Burgers, P. M. Ubiquitinated proliferating cell nuclear antigen activates translesion DNA polymerases eta and REV1. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 18361–6 (2005).

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

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

212. Gillet, L. C. J. & Schärer, O. D. Molecular mechanisms of mammalian global genome nucleotide excision repair. *Chem. Rev.* **106,** 253–76 (2006).

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

214. Min, J.-H. & Pavletich, N. P. Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature* **449,** 570–575 (2007).

215. Sugasawa, K. *et al.* A multistep damage recognition mechanism for global genomic nucleotide excision repair. *Genes Dev.* **15,** 507–21 (2001).

216. Moser, J. *et al.* The UV-damaged DNA binding protein mediates efficient targeting of the nucleotide excision repair complex to UV-induced photo lesions. *DNA Repair (Amst).* **4,** 571–82 (2005).

217. Araújo, S. J., Nigg, E. A. & Wood, R. D. Strong functional interactions of TFIIH with XPC and XPG in human DNA nucleotide excision repair, without a preassembled repairosome. *Mol. Cell. Biol.* **21,** 2281–91 (2001).

218. Volker, M. *et al.* Sequential Assembly of the Nucleotide Excision Repair Factors In Vivo. *Mol. Cell* **8,** 213–224 (2001).

219. Yokoi, M. The Xeroderma Pigmentosum Group C Protein Complex XPC-HR23B Plays an Important Role in the Recruitment of Transcription Factor IIH to Damaged DNA. *J. Biol. Chem.* **275,** 9870–9875 (2000).

220. Coin, F., Oksenych, V. & Egly, J.-M. Distinct roles for the XPB/p52 and XPD/p44 subcomplexes of TFIIH in damaged DNA opening during nucleotide excision repair. *Mol. Cell* **26,** 245–56 (2007).

221. Winkler, G. S. *et al.* TFIIH with Inactive XPD Helicase Functions in Transcription Initiation but Is Defective in DNA Repair. *J. Biol. Chem.* **275,** 4258–4266 (2000).

222. Naegeli, H., Bardwell, L. & Friedberg, E. The DNA helicase and adenosine triphosphatase activities of yeast Rad3 protein are inhibited by DNA damage. A potential mechanism for damage- specific recognition. *J. Biol. Chem.* **267,** 392–398 (1992).

223. Nocentini, S., Coin, F., Saijo, M., Tanaka, K. & Egly, J.-M. DNA Damage Recognition by XPA Protein Promotes Efficient Recruitment of Transcription Factor II H. *J. Biol. Chem.* **272,** 22991–22994 (1997).

224. Rademakers, S. *et al.* Xeroderma Pigmentosum Group A Protein Loads as a Separate Factor onto DNA Lesions. *Mol. Cell. Biol.* **23,** 5755–5767 (2003).

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

226. De Laat, W. L. *et al.* DNA-binding polarity of human replication protein A positions nucleases in nucleotide excision repair. *Genes Dev.* **12,** 2598–2609 (1998).

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

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

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

230. Li, L., Elledge, S. J., Peterson, C. A., Bales, E. S. & Legerski, R. J. Specific association between the human DNA repair proteins XPA and ERCC1. *Proc. Natl. Acad. Sci.* **91,** 5012–5016 (1994).

231. Ogi, T. *et al.* Three DNA polymerases, recruited by different mechanisms, carry out NER repair synthesis in human cells. *Mol. Cell* **37,** 714–27 (2010).

232. Shivji, M. K. K., Podust, V. N., Huebscher, U. & Wood, R. D. Nucleotide Excision Repair DNA Synthesis by DNA Polymerase .epsilon. in the Presence of PCNA, RFC, and RPA. *Biochemistry* **34,** 5011–5017 (1995).

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

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

235. Moser, J. *et al.* Sealing of chromosomal DNA nicks during nucleotide excision repair requires XRCC1 and DNA ligase III alpha in a cell-cycle-specific manner. *Mol. Cell* **27,** 311–23 (2007).

236. Pines, A. *et al.* PARP1 promotes nucleotide excision repair through DDB2 stabilization and recruitment of ALC1. *J. Cell Biol.* **199,** 235–49 (2012).

237. Okano, S., Lan, L., Caldecott, K. W., Mori, T. & Yasui, A. Spatial and temporal cellular responses to single-strand breaks in human cells. *Mol. Cell. Biol.* **23,** 3974–81 (2003).

238. Pleschke, J. M., Kleczkowska, H. E., Strohm, M. & Althaus, F. R. Poly(ADP-ribose) binds to specific domains in DNA damage checkpoint proteins. *J. Biol. Chem.* **275,** 40974–80 (2000).

239. Gagné, J.-P. *et al.* Proteome-wide identification of poly(ADP-ribose) binding proteins and poly(ADP-ribose)-associated protein complexes. *Nucleic Acids Res.* **36,** 6959–76 (2008).

240. Karras, G. I. *et al.* The macro domain is an ADP-ribose binding module. *EMBO J.* **24,** 1911–20 (2005).

241. Ahel, I. *et al.* Poly(ADP-ribose)-binding zinc finger motifs in DNA repair/checkpoint proteins. *Nature* **451,** 81–5 (2008).

242. Aravind, L. The WWE domain: a common interaction module in protein ubiquitination and ADP ribosylation. *Trends Biochem. Sci.* **26,** 273–275 (2001).

243. King, B. S., Cooper, K. L., Liu, K. J. & Hudson, L. G. Poly(ADP-ribose) contributes to an association between poly(ADP-ribose) polymerase-1 and xeroderma pigmentosum complementation group A in nucleotide excision repair. *J. Biol. Chem.* **287,** 39824–33 (2012).

244. Leppard, J. B., Dong, Z., Mackey, Z. B. & Tomkinson, A. E. Physical and functional interaction between DNA ligase IIIalpha and poly(ADP-Ribose) polymerase 1 in DNA single-strand break repair. *Mol. Cell. Biol.* **23,** 5919–27 (2003).

245. Yu, X., Chini, C. C. S., He, M., Mer, G. & Chen, J. The BRCT domain is a phospho-protein binding domain. *Science* **302,** 639–42 (2003).

246. Han, W., Li, X. & Fu, X. The macro domain protein family: structure, functions, and their potential therapeutic implications. *Mutat. Res.* **727,** 86–103 (2011).

247. Neuvonen, M. & Ahola, T. Differential activities of cellular and viral macro domain proteins in binding of ADP-ribose metabolites. *J. Mol. Biol.* **385,** 212–25 (2009).

248. Story, R. M. & Steitz, T. A. Structure of the recA protein-ADP complex. *Nature* **355,** 374–6 (1992).

249. Ma, N.-F. *et al.* Isolation and characterization of a novel oncogene, amplified in liver cancer 1, within a commonly amplified region at 1q21 in hepatocellular carcinoma. *Hepatology* **47,** 503–10 (2008).

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

251. Takada, T., Iida, K. & Moss, J. Cloning and site-directed mutagenesis of human ADP-ribosylarginine hydrolase. *J. Biol. Chem.* **268,** 17837–17843 (1993).

252. Moss, J., Oppenheimer, N. J., West, R. E. & Stanley, S. J. Amino acid specific ADP-ribosylation: substrate specificity of an ADP-ribosylarginine hydrolase from turkey erythrocytes. *Biochemistry* **25,** 5408–14 (1986).

253. Rosenthal, F. *et al.* Macrodomain-containing proteins are new mono-ADP-ribosylhydrolases. *Nat. Struct. Mol. Biol.* **20,** 502–7 (2013).

254. Aguiar, R. C. *et al.* BAL is a novel risk-related gene in diffuse large B-cell lymphomas that enhances cellular migration. *Blood* **96,** 4328–34 (2000).

255. Takeyama, K. *et al.* The BAL-binding protein BBAP and related Deltex family members exhibit ubiquitin-protein isopeptide ligase activity. *J. Biol. Chem.* **278,** 21930–7 (2003).

256. Yan, Q. *et al.* BBAP monoubiquitylates histone H4 at lysine 91 and selectively modulates the DNA damage response. *Mol. Cell* **36,** 110–20 (2009).

257. Angelov, D. *et al.* The Histone Variant MacroH2A Interferes with Transcription Factor Binding and SWI/SNF Nucleosome Remodeling. *Mol. Cell* **11,** 1033–1041 (2003).

258. Gamble, M. J., Frizzell, K. M., Yang, C., Krishnakumar, R. & Kraus, W. L. The histone variant macroH2A1 marks repressed autosomal chromatin, but protects a subset of its target genes from silencing. *Genes Dev.* **24,** 21–32 (2010).

259. Doyen, C.-M. *et al.* Mechanism of polymerase II transcription repression by the histone variant macroH2A. *Mol. Cell. Biol.* **26,** 1156–64 (2006).

260. Costanzi, C. & Pehrson, J. R. Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. *Nature* **393,** 599–601 (1998).

261. Kustatscher, G., Hothorn, M., Pugieux, C., Scheffzek, K. & Ladurner, A. G. Splicing regulates NAD metabolite binding to histone macroH2A. *Nat. Struct. Mol. Biol.* **12,** 624–5 (2005).

262. Costanzi, C. & Pehrson, J. R. MACROH2A2, a New Member of the MACROH2A Core Histone Family. *J. Biol. Chem.* **276,** 21776–21784 (2001).

263. Panagabko, C. *et al.* Ligand specificity in the CRAL-TRIO protein family. *Biochemistry* **42,** 6467–74 (2003).

264. Eustermann, S. *et al.* Solution structures of the two PBZ domains from human APLF and their interaction with poly(ADP-ribose). *Nat. Struct. Mol. Biol.* **17,** 241–3 (2010).

265. Oberoi, J. *et al.* Structural basis of poly(ADP-ribose) recognition by the multizinc binding domain of checkpoint with forkhead-associated and RING Domains (CHFR). *J. Biol. Chem.* **285,** 39348–58 (2010).

266. Iuchi, S. Three classes of C2H2 zinc finger proteins. *Cell. Mol. Life Sci.* **58,** 625–635 (2001).

267. Rulten, S. L., Cortes-Ledesma, F., Guo, L., Iles, N. J. & Caldecott, K. W. APLF (C2orf13) is a novel component of poly(ADP-ribose) signaling in mammalian cells. *Mol. Cell. Biol.* **28,** 4620–8 (2008).

268. Li, G.-Y. *et al.* Structure and identification of ADP-ribose recognition motifs of APLF and role in the DNA damage response. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 9129–34 (2010).

269. Henriques, J. A. P. & Moustacchi, E. ISOLATION AND CHARACTERIZATION OF pso MUTANTS SENSITIVE TO PHOTO-ADDITION OF PSORALEN DERIVATIVES IN SACCHAROMYCES CEREVISIAE. *Genetics* **95,** 273–288 (1980).

270. Tiefenbach, T. & Junop, M. Pso2 (SNM1) is a DNA structure-specific endonuclease. *Nucleic Acids Res.* **40,** 2131–9 (2012).

271. Grossmann, K. F., Ward, A. M., Matkovic, M. E., Folias, A. E. & Moses, R. E. S. cerevisiae has three pathways for DNA interstrand crosslink repair. *Mutat. Res. Repair* **487,** 73–83 (2001).

272. Li, X. & Moses, R. E. The β-lactamase motif in Snm1 is required for repair of DNA double-strand breaks caused by interstrand crosslinks in S. cerevisiae. *DNA Repair (Amst).* **2,** 121–129 (2003).

273. Grossmann, K. F., Ward, A. M. & Moses, R. E. Saccharomyces cerevisiae lacking Snm1, Rev3 or Rad51 have a normal S-phase but arrest permanently in G2 after cisplatin treatment. *Mutat. Res. Repair* **461,** 1–13 (2000).

274. Sengerová, B. *et al.* Characterization of the human SNM1A and SNM1B/Apollo DNA repair exonucleases. *J. Biol. Chem.* **287,** 26254–67 (2012).

275. Pannicke, U. *et al.* Functional and biochemical dissection of the structure-specific nuclease ARTEMIS. *EMBO J.* **23,** 1987–97 (2004).

276. Dronkert, M. L. G. *et al.* Disruption of Mouse SNM1 Causes Increased Sensitivity to the DNA Interstrand Cross-Linking Agent Mitomycin C. *Mol. Cell. Biol.* **20,** 4553–4561 (2000).

277. Hemphill, A. W. *et al.* Mammalian SNM1 is required for genome stability. *Mol. Genet. Metab.* **94,** 38–45 (2008).

278. Scolnick, D. M. & Halazonetis, T. D. Chfr defines a mitotic stress checkpoint that delays entry into metaphase. *Nature* **406,** 430–5 (2000).

279. Yu, X. *et al.* Chfr is required for tumor suppression and Aurora A regulation. *Nat. Genet.* **37,** 401–6 (2005).

280. Bothos, J., Summers, M. K., Venere, M., Scolnick, D. M. & Halazonetis, T. D. The Chfr mitotic checkpoint protein functions with Ubc13-Mms2 to form Lys63-linked polyubiquitin chains. *Oncogene* **22,** 7101–7 (2003).

281. Kang, D., Chen, J., Wong, J. & Fang, G. The checkpoint protein Chfr is a ligase that ubiquitinates Plk1 and inhibits Cdc2 at the G2 to M transition. *J. Cell Biol.* **156,** 249–59 (2002).

282. Wang, Z. *et al.* Recognition of the iso-ADP-ribose moiety in poly(ADP-ribose) by WWE domains suggests a general mechanism for poly(ADP-ribosyl)ation-dependent ubiquitination. *Genes Dev.* **26,** 235–40 (2012).

283. Wang, D. *et al.* Genome-wide analysis of CCCH zinc finger family in Arabidopsis and rice. *BMC Genomics* **9,** 44 (2008).

284. Kerns, J. A., Emerman, M. & Malik, H. S. Positive selection and increased antiviral activity associated with the PARP-containing isoform of human zinc-finger antiviral protein. *PLoS Genet.* **4,** e21 (2008).

285. DaRosa, P. A. *et al.* Allosteric activation of the RNF146 ubiquitin ligase by a poly(ADP-ribosyl)ation signal. *Nature* **517,** 223–6 (2015).

286. Callow, M. G. *et al.* Ubiquitin ligase RNF146 regulates tankyrase and Axin to promote Wnt signaling. *PLoS One* **6,** e22595 (2011).

287. Zhang, Y. *et al.* RNF146 is a poly(ADP-ribose)-directed E3 ligase that regulates axin degradation and Wnt signalling. *Nat. Cell Biol.* **13,** 623–9 (2011).

288. Gorman, M. J. & Girton, J. R. A genetic analysis of deltex and its interaction with the Notch locus in Drosophila melanogaster. *Genetics* **131,** 99–112 (1992).

289. Matsuno, K., Diederich, R., Go, M., Blaumueller, C. & Artavanis-Tsakonas, S. Deltex acts as a positive regulator of Notch signaling through interactions with the Notch ankyrin repeats. *Development* **121,** 2633–2644 (1995).

290. Matsuno, K. *et al.* Human deltex is a conserved regulator of Notch signalling. *Nat. Genet.* **19,** 74–8 (1998).

291. Zweifel, M. E., Leahy, D. J. & Barrick, D. Structure and Notch receptor binding of the tandem WWE domain of Deltex. *Structure* **13,** 1599–611 (2005).

292. Izon, D. J. *et al.* Deltex1 Redirects Lymphoid Progenitors to the B Cell Lineage by Antagonizing Notch1. *Immunity* **16,** 231–243 (2002).

293. Yamamoto, N. *et al.* Role of Deltex-1 as a transcriptional regulator downstream of the Notch receptor. *J. Biol. Chem.* **276,** 45031–40 (2001).

294. Durocher, D., Henckel, J., Fersht, A. R. & Jackson, S. P. The FHA Domain Is a Modular Phosphopeptide Recognition Motif. *Mol. Cell* **4,** 387–394 (1999).

295. Hofmann, K. The FHA domain: a putative nuclear signalling domain found in protein kinases and transcription factors. *Trends Biochem. Sci.* **20,** 347–349 (1995).

296. Sun, Z. Rad53 FHA Domain Associated with Phosphorylated Rad9 in the DNA Damage Checkpoint. *Science (80-.).* **281,** 272–274 (1998).

297. Lloyd, J. *et al.* A supramodular FHA/BRCT-repeat architecture mediates Nbs1 adaptor function in response to DNA damage. *Cell* **139,** 100–11 (2009).

298. Williams, R. S. *et al.* Nbs1 flexibly tethers Ctp1 and Mre11-Rad50 to coordinate DNA double-strand break processing and repair. *Cell* **139,** 87–99 (2009).

299. Callebaut, I. & Mornon, J.-P. From BRCA1 to RAP1: a widespread BRCT module closely associated with DNA repair. *FEBS Lett.* **400,** 25–30 (1997).

300. Williams, R. S., Green, R. & Glover, J. N. Crystal structure of the BRCT repeat region from the breast cancer-associated protein BRCA1. *Nat. Struct. Biol.* **8,** 838–42 (2001).

301. Peng, M., Litman, R., Jin, Z., Fong, G. & Cantor, S. B. BACH1 is a DNA repair protein supporting BRCA1 damage response. *Oncogene* **25,** 2245–53 (2006).

302. Cantor, S. B. *et al.* BACH1, a Novel Helicase-like Protein, Interacts Directly with BRCA1 and Contributes to Its DNA Repair Function. *Cell* **105,** 149–160 (2001).

303. Cuneo, M. J., Gabel, S. A., Krahn, J. M., Ricker, M. A. & London, R. E. The structural basis for partitioning of the XRCC1/DNA ligase III-α BRCT-mediated dimer complexes. *Nucleic Acids Res.* **39,** 7816–27 (2011).

304. Derbyshire, D. J. *et al.* Crystal structure of human 53BP1 BRCT domains bound to p53 tumour suppressor. *EMBO J.* **21,** 3863–72 (2002).

305. Ward, I. *et al.* The tandem BRCT domain of 53BP1 is not required for its repair function. *J. Biol. Chem.* **281,** 38472–7 (2006).

306. Stucki, M. *et al.* MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. *Cell* **123,** 1213–26 (2005).

307. Li, M. & Yu, X. Function of BRCA1 in the DNA damage response is mediated by ADP-ribosylation. *Cancer Cell* **23,** 693–704 (2013).

308. Li, M., Lu, L.-Y., Yang, C.-Y., Wang, S. & Yu, X. The FHA and BRCT domains recognize ADP-ribosylation during DNA damage response. *Genes Dev.* **27,** 1752–68 (2013).

309. Murzin, A. G. OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. *EMBO J.* **12,** 861–7 (1993).

310. Zhang, F., Chen, Y., Li, M. & Yu, X. The oligonucleotide/oligosaccharide-binding fold motif is a poly(ADP-ribose)-binding domain that mediates DNA damage response. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 7278–83 (2014).

311. Richard, D. J. *et al.* Single-stranded DNA-binding protein hSSB1 is critical for genomic stability. *Nature* **453,** 677–81 (2008).

312. Richard, D. J. *et al.* hSSB1 interacts directly with the MRN complex stimulating its recruitment to DNA double-strand breaks and its endo-nuclease activity. *Nucleic Acids Res.* **39,** 3643–51 (2011).

313. Richard, D. J. *et al.* hSSB1 rapidly binds at the sites of DNA double-strand breaks and is required for the efficient recruitment of the MRN complex. *Nucleic Acids Res.* **39,** 1692–702 (2011).

314. Krogh, B. O. & Symington, L. S. Recombination proteins in yeast. *Annu. Rev. Genet.* **38,** 233–71 (2004).

315. Citarelli, M., Teotia, S. & Lamb, R. S. Evolutionary history of the poly(ADP-ribose) polymerase gene family in eukaryotes. *BMC Evol. Biol.* **10,** 308 (2010).

316. Pachkowski, B. F. *et al.* Cells deficient in PARP-1 show an accelerated accumulation of DNA single strand breaks, but not AP sites, over the PARP-1-proficient cells exposed to MMS. *Mutat. Res. Mol. Mech. Mutagen.* **671,** 93–99 (2009).

317. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432,** 695–716 (2004).

318. GAGNON, S. N., HENGARTNER, M. O. & DESNOYERS, S. The genes pme-1 and pme-2 encode two poly(ADP-ribose) polymerases in Caenorhabditis elegans. *Biochem. J.* **368,** 263–271 (2002).

319. Dequen, F., Gagnon, S. N. & Desnoyers, S. Ionizing radiations in Caenorhabditis elegans induce poly(ADP-ribosyl)ation, a conserved DNA-damage response essential for survival. *DNA Repair (Amst).* **4,** 814–25 (2005).

320. Otto, H. *et al.* In silico characterization of the family of PARP-like poly(ADP-ribosyl)transferases (pARTs). *BMC Genomics* **6,** 139 (2005).

321. Pears, C. J. *et al.* The role of ADP-ribosylation in regulating DNA double-strand break repair. *Cell Cycle* **11,** 48–56 (2012).

322. SCHWALB, M. & ROTH, R. Axenic Growth and Development of the Cellular Slime Mould, Dictyostelium discoideum. *J. Gen. Microbiol.* **60,** 283–286 (1970).

323. Watts, D. J. & Ashworth, J. M. Growth of myxameobae of the cellular slime mould Dictyostelium discoideum in axenic culture. *Biochem. J.* **119,** 171–4 (1970).

324. Franke, J. & Kessin, R. A defined minimal medium for axenic strains of Dictyostelium discoideum. *Proc. Natl. Acad. Sci. U. S. A.* **74,** 2157–61 (1977).

325. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291,** 1304–51 (2001).

326. Sakharkar, M. K., Chow, V. T. K. & Kangueane, P. Distributions of exons and introns in the human genome. *In Silico Biol.* **4,** 387–93 (2004).

327. Eichinger, L. *et al.* The genome of the social amoeba Dictyostelium discoideum. *Nature* **435,** 43–57 (2005).

328. Stevense, M., Chubb, J. R. & Muramoto, T. Nuclear organization and transcriptional dynamics in Dictyostelium. *Dev. Growth Differ.* **53,** 576–86 (2011).

329. Zada-Hames, I. M. & Ashworth, J. M. The cell cycle during the vegetative stage of Dictyostelium discoideum and its response to temperature change. *J. Cell Sci.* **32,** 1–20 (1978).

330. Weijer, C. J., Duschl, G. & David, C. N. A revision of the Dictyostelium discoideum cell cycle. *J. Cell Sci.* **70,** 111–31 (1984).

331. Araki, T., Nakao, H., Takeuchi, I. & Maeda, Y. Cell-cycle-dependent sorting in the development of Dictyostelium cells. *Dev. Biol.* **162,** 221–8 (1994).

332. Strasser, K. *et al.* A retinoblastoma orthologue is a major regulator of S-phase, mitotic, and developmental gene expression in Dictyostelium. *PLoS One* **7,** e39914 (2012).

333. Strmecki, L., Greene, D. M. & Pears, C. J. Developmental decisions in Dictyostelium discoideum. *Dev. Biol.* **284,** 25–36 (2005).

334. Muramoto, T. & Chubb, J. R. Live imaging of the Dictyostelium cell cycle reveals widespread S phase during development, a G2 bias in spore differentiation and a premitotic checkpoint. *Development* **135,** 1647–57 (2008).

335. Deering, R. A. Dictyostelium discoideum: A Gamma-Ray Resistant Organism. *Science (80-.).* **162,** 1289–1290 (1968).

336. Freim, J. O. . J. & Deering, R. A. Ultraviolet irradiation of the vegetative cells of Dictyostelium discoideum. *J. Bacteriol.* **102,** 36–42 (1970).

337. Bronner, C. E., Welker, D. L. & Deering, R. A. Mutations affecting sensitivity of the cellular slime mold Dictyostelium discoideum to DNA-damaging agents. *Mutat. Res. Repair* **274,** 187–200 (1992).

338. Deering, R. A., Smith, M. S., Thompson, B. K. & Adolf, A. C. Gamma-ray-resistant and -sensitive strains of slime mold (Dictyostelium discoideum). *Radiat. Res.* **43,** 711–28 (1970).

339. Deering, R. A., Guyer, R. B., Stevens, L. & Watson-Thais, T. E. Some repair-deficient mutants of Dictyostelium discoideum display enhanced susceptibilities to bleomycin. *Antimicrob. Agents Chemother.* **40,** 464–467 (1996).

340. Rickwood, D. & Osman, M. S. Characterisation of poly(ADP-Rib) polymerase activity in nuclei from the slime mould dictyostelium discoideum. *Mol. Cell. Biochem.* **27,** 79–84 (1979).

341. Kofler, B. *et al.* Purification and characterization of NAD + :ADP-ribosyltransferase (polymerizing) from Dictyostelium discoideum. *Biochem. J.* **293,** 275–281 (1993).

342. Freeland, T. M., Guyer, R. B., Ling, A. Z. & Deering, R. A. Apurinic/apyrimidinic (AP) endonuclease from Dictyostelium discoideum: cloning, nucleotide sequence and induction by sublethal levels of DNA damaging agents. *Nucleic Acids Res.* **24,** 1950–3 (1996).

343. Guyer, R. B., Nonnemaker, J. M. & Deering, R. A. Uracil-DNA glycosylase activity from Dictyostelium discoideum. *Biochim. Biophys. Acta - Gene Struct. Expr.* **868,** 262–264 (1986).

344. Alexander, H., Lee, S.-K., Yu, S.-L. & Alexander, S. RepE--the Dictyostelium Homolog of the Human Xeroderma Pigmentosum Group E Gene Is Developmentally Regulated and Contains a Leucine Zipper Motif. *Nucleic Acids Res.* **24,** 2295–2301 (1996).

345. Lee, S.-K., Yu, S.-L., Garcia, M. X., Alexander, H. & Alexander, S. Differential developmental expression of the repB and repD Xeroderma pigmentosum related DNA helicase genes from Dictyostelium discoideum. *Nucleic Acids Res.* **25,** 2365–2374 (1997).

346. Yu, S.-L., Lee, S.-K., Alexander, H. & Alexander, S. Rapid changes of nucleotide excision repair gene expression following UV-irradiation and cisplatin treatment of Dictyostelium discoideum. *Nucleic Acids Res.* **26,** 3397–3403 (1998).

347. Zhang, X.-Y. *et al.* Xpf and not the Fanconi anaemia proteins or Rev3 accounts for the extreme resistance to cisplatin in Dictyostelium discoideum. *PLoS Genet.* **5,** e1000645 (2009).

348. Lee, S. K., Yu, S. L., Alexander, H. & Alexander, S. A mutation in repB, the dictyostelium homolog of the human xeroderma pigmentosum B gene, has increased sensitivity to UV-light but normal morphogenesis. *Biochim. Biophys. Acta* **1399,** 161–72 (1998).

349. Cromie, G. A., Connelly, J. C. & Leach, D. R. F. Recombination at Double-Strand Breaks and DNA Ends. *Mol. Cell* **8,** 1163–1174 (2001).

350. McVey, M. Strategies for DNA interstrand crosslink repair: insights from worms, flies, frogs, and slime molds. *Environ. Mol. Mutagen.* **51,** 646–58 (2010).

351. Hsu, D.-W. *et al.* DNA double-strand break repair pathway choice in Dictyostelium. *J. Cell Sci.* **124,** 1655–63 (2011).

352. Wen, X., Khampang, P. & Rutherford, C. L. The glycogen phosphorylase-2 promoter binding protein in Dictyostelium is replication protein A. *J. Mol. Biol.* **284,** 903–13 (1998).

353. Hsu, D.-W. & Gaudet, P. DNA Damage Signalling and Repair in Dictyostelium discoideum. *Cell Cycle* **5,** 702–708 (2006).

354. Weller, G. R. *et al.* Identification of a DNA nonhomologous end-joining complex in bacteria. *Science* **297,** 1686–9 (2002).

355. Della, M. *et al.* Mycobacterial Ku and ligase proteins constitute a two-component NHEJ repair machine. *Science* **306,** 683–5 (2004).

356. Paull, T. T. & Gellert, M. A mechanistic basis for Mre11-directed DNA joining at microhomologies. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 6409–14 (2000).

357. Chen, L., Trujillo, K., Ramos, W., Sung, P. & Tomkinson, A. E. Promotion of Dnl4-Catalyzed DNA End-Joining by the Rad50/Mre11/Xrs2 and Hdf1/Hdf2 Complexes. *Mol. Cell* **8,** 1105–1115 (2001).

358. Hudson, J. J. R. *et al.* DNA-PKcs-dependent signaling of DNA damage in Dictyostelium discoideum. *Curr. Biol.* **15,** 1880–5 (2005).

359. Kuspa, A. & Loomis, W. F. Tagging developmental genes in Dictyostelium by restriction enzyme-mediated integration of plasmid DNA. *Proc. Natl. Acad. Sci.* **89,** 8803–8807 (1992).

360. Couto, C. A.-M. *et al.* PARP regulates nonhomologous end joining through retention of Ku at double-strand breaks. *J. Cell Biol.* **194,** 367–75 (2011).

361. Couto, C. A.-M. *et al.* Nonhomologous end-joining promotes resistance to DNA damage in the absence of an ADP-ribosyltransferase that signals DNA single strand breaks. *J. Cell Sci.* **126,** 3452–61 (2013).

362. Wu, C. H. *et al.* The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Res.* **34,** D187–91 (2006).

363. Kreppel, L. *et al.* dictyBase: a new Dictyostelium discoideum genome database. *Nucleic Acids Res.* **32,** D332–3 (2004).

364. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32,** 1792–7 (2004).

365. Notredame, C., Higgins, D. G. & Heringa, J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302,** 205–17 (2000).

366. Sonnhammer, E. L. L. & Hollich, V. Scoredist: a simple and robust protein sequence distance estimator. *BMC Bioinformatics* **6,** 108 (2005).

367. Corpet, F. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* **16,** 10881–90 (1988).

368. Punta, M. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **40,** D290–301 (2012).

369. Gough, J., Karplus, K., Hughey, R. & Chothia, C. Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *J. Mol. Biol.* **313,** 903–19 (2001).

370. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–10 (1990).

371. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* **39,** W29–37 (2011).

372. Söding, J., Biegert, A. & Lupas, A. N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **33,** W244–8 (2005).

373. Eswar, N. *et al.* Comparative protein structure modeling using MODELLER. *Curr. Protoc. Protein Sci.* **Chapter 2,** Unit 2.9 (2007).

374. Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234,** 779–815 (1993).

375. Delano, W. The PyMOL Molecular Graphics System. (2002). at <http://www.citeulike.org/group/340/article/240061>

376. Berman, H. M. The Protein Data Bank. *Nucleic Acids Res.* **28,** 235–242 (2000).

377. Chen, D. *et al.* Identification of macrodomain proteins as novel O-acetyl-ADP-ribose deacetylases. *J. Biol. Chem.* **286,** 13261–71 (2011).

378. Liu, H., Nakagawa, T., Kanematsu, T., Uchida, T. & Tsuji, S. Isolation of 10 differentially expressed cDNAs in differentiated Neuro2a cells induced through controlled expression of the GD3 synthase gene. *J. Neurochem.* **72,** 1781–90 (1999).

379. Minami, Y., Emori, Y., Kawasaki, H. & Suzuki, K. E-F hand structure-domain of calcium-activated neutral protease (CANP) can bind Ca2+ ions. *J. Biochem.* **101,** 889–95 (1987).

380. Bao, S. *et al.* ATR/ATM-mediated phosphorylation of human Rad17 is required for genotoxic stress responses. *Nature* **411,** 969–74 (2001).

381. Zou, L., Cortez, D. & Elledge, S. J. Regulation of ATR substrate selection by Rad17-dependent loading of Rad9 complexes onto chromatin. *Genes Dev.* **16,** 198–208 (2002).

382. Post, S. *et al.* Phosphorylation of serines 635 and 645 of human Rad17 is cell cycle regulated and is required for G(1)/S checkpoint activation in response to DNA damage. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 13102–7 (2001).

383. Bermudez, V. P. *et al.* Loading of the human 9-1-1 checkpoint complex onto DNA by the checkpoint clamp loader hRad17-replication factor C complex in vitro. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 1633–8 (2003).

384. Lee, J. S., Collins, K. M., Brown, A. L., Lee, C. H. & Chung, J. H. hCds1-mediated phosphorylation of BRCA1 regulates the DNA damage response. *Nature* **404,** 201–4 (2000).

385. Chehab, N. H., Malikzay, A., Appel, M. & Halazonetis, T. D. Chk2/hCds1 functions as a DNA damage checkpoint in G(1) by stabilizing p53. *Genes Dev.* **14,** 278–88 (2000).

386. Shieh, S. Y., Ahn, J., Tamai, K., Taya, Y. & Prives, C. The human homologs of checkpoint kinases Chk1 and Cds1 (Chk2) phosphorylate p53 at multiple DNA damage-inducible sites. *Genes Dev.* **14,** 289–300 (2000).

387. Needleman, S. B. & Wunsch, C. D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48,** 443–53 (1970).

388. Smith, T. F. & Waterman, M. S. Identification of common molecular subsequences. *J. Mol. Biol.* **147,** 195–7 (1981).

389. Eddy, S. R. Profile hidden Markov models. *Bioinformatics* **14,** 755–63 (1998).

390. Söding, J. Protein homology detection by HMM-HMM comparison. *Bioinformatics* **21,** 951–60 (2005).

391. Holm, L. & Sander, C. Dali: a network tool for protein structure comparison. *Trends Biochem. Sci.* **20,** 478–480 (1995).

392. Shindyalov, I. N. & Bourne, P. E. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng.* **11,** 739–47 (1998).

393. Baker, D. & Sali, A. Protein structure prediction and structural genomics. *Science* **294,** 93–6 (2001).

394. Kim, D. E., Chivian, D. & Baker, D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.* **32,** W526–31 (2004).

395. Bradley, P., Misura, K. M. S. & Baker, D. Toward high-resolution de novo structure prediction for small proteins. *Science* **309,** 1868–71 (2005).

396. Wallner, B. & Elofsson, A. All are not equal: a benchmark of different homology modeling programs. *Protein Sci.* **14,** 1315–27 (2005).

397. Heringa, J. Computational methods for protein secondary structure prediction using multiple sequence alignments. *Curr. Protein Pept. Sci.* **1,** 273–301 (2000).

398. McGuffin, L. J., Bryson, K. & Jones, D. T. The PSIPRED protein structure prediction server. *Bioinformatics* **16,** 404–5 (2000).

399. Rost, B. Twilight zone of protein sequence alignments. *Protein Eng. Des. Sel.* **12,** 85–94 (1999).

400. Simsek, D. & Jasin, M. DNA ligase III: a spotty presence in eukaryotes, but an essential function where tested. *Cell Cycle* **10,** 3636–44 (2011).

401. Cunningham, B. A., Hemperly, J. J., Hopp, T. P. & Edelman, G. M. Favin versus concanavalin A: Circularly permuted amino acid sequences. *Proc. Natl. Acad. Sci. U. S. A.* **76,** 3218–22 (1979).

402. Ponting, C. P. & Russell, R. B. Swaposins: circular permutations within genes encoding saposin homologues. *Trends Biochem. Sci.* **20,** 179–180 (1995).

403. Martí-Renom, M. A. *et al.* Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* **29,** 291–325 (2000).

404. Chothia, C. & Lesk, A. M. The relation between the divergence of sequence and structure in proteins. *EMBO J.* **5,** 823–6 (1986).

405. Durocher, D., Henckel, J., Fersht, A. R. & Jackson, S. P. The FHA Domain Is a Modular Phosphopeptide Recognition Motif. *Mol. Cell* **4,** 387–394 (1999).

406. Shen, J. *et al.* Mutations in PNKP cause microcephaly, seizures and defects in DNA repair. *Nat. Genet.* **42,** 245–9 (2010).

407. Forst, A. H. *et al.* Recognition of mono-ADP-ribosylated ARTD10 substrates by ARTD8 macrodomains. *Structure* **21,** 462–75 (2013).

408. Tomkinson, A. E. & Levin, D. S. Mammalian DNA ligases. *Bioessays* **19,** 893–901 (1997).

409. Söderhäll, S. & Lindahl, T. Two DNA ligase activities from calf thymus. *Biochem. Biophys. Res. Commun.* **53,** 910–6 (1973).

410. Söderhäll, S. DNA ligases during rat liver regeneration. *Nature* **260,** 640–2 (1976).

411. Montecucco, A. *et al.* The N-terminal domain of human DNA ligase I contains the nuclear localization signal and directs the enzyme to sites of DNA replication. *EMBO J.* **14,** 5379–86 (1995).

412. Levin, D. S., Bai, W., Yao, N., O'Donnell, M. & Tomkinson, A. E. An interaction between DNA ligase I and proliferating cell nuclear antigen: Implications for Okazaki fragment synthesis and joining. *Proc. Natl. Acad. Sci.* **94,** 12863–12868 (1997).

413. Barnes, D. E. Specific Function of DNA Ligase I in Simian Virus 40DNA Replication by Human Cell-free Extracts Is Mediated by the Amino-terminal Non-catalytic Domain. *J. Biol. Chem.* **272,** 11550–11556 (1997).

414. Levin, D. S., McKenna, A. E., Motycka, T. A., Matsumoto, Y. & Tomkinson, A. E. Interaction between PCNA and DNA ligase I is critical for joining of Okazaki fragments and long-patch base-excision repair. *Curr. Biol.* **10,** 919–S2 (2000).

415. Willis, A. E. & Lindahl, T. DNA ligase I deficiency in Bloom's syndrome. *Nature* **325,** 355–7

416. Webster, A. D., Barnes, D. E., Arlett, C. F., Lehmann, A. R. & Lindahl, T. Growth retardation and immunodeficiency in a patient with mutations in the DNA ligase I gene. *Lancet (London, England)* **339,** 1508–9 (1992).

417. Barnes, D. E., Tomkinson, A. E., Lehmann, A. R., Webster, A. D. B. & Lindahl, T. Mutations in the DNA ligase I gene of an individual with immunodeficiencies and cellular hypersensitivity to DNA-damaging agents. *Cell* **69,** 495–503 (1992).

418. Petrini, J. H., Huwiler, K. G. & Weaver, D. T. A wild-type DNA ligase I gene is expressed in Bloom's syndrome cells. *Proc. Natl. Acad. Sci. U. S. A.* **88,** 7615–9 (1991).

419. Prasad, R. *et al.* Specific Interaction of DNA Polymerase and DNA Ligase I in a Multiprotein Base Excision Repair Complex from Bovine Testis. *J. Biol. Chem.* **271,** 16000–16007 (1996).

420. Balakrishnan, L., Brandt, P. D., Lindsey-Boltz, L. A., Sancar, A. & Bambara, R. A. Long patch base excision repair proceeds via coordinated stimulation of the multienzyme DNA repair complex. *J. Biol. Chem.* **284,** 15158–72 (2009).

421. Teo, I. A., Arlett, C. F., Harcourt, S. A., Priestley, A. & Broughton, B. C. Multiple hypersensitivity to mutagens in a cell strain (46BR) derived from a patient with immuno-deficiencies. *Mutat. Res. Mol. Mech. Mutagen.* **107,** 371–386 (1983).

422. Nocentini, S. Rejoining kinetics of DNA single- and double-strand breaks in normal and DNA ligase-deficient cells after exposure to ultraviolet C and gamma radiation: an evaluation of ligating activities involved in different DNA repair processes. *Radiat. Res.* **151,** 423–32 (1999).

423. Bentley, D. J. *et al.* DNA ligase I null mouse cells show normal DNA repair activity but altered DNA replication and reduced genome stability. *J. Cell Sci.* **115,** 1551–1561 (2002).

424. Tomkinson, A. E., Roberts, E., Daly, G., Totty, N. F. & Lindahl, T. Three distinct DNA ligases in mammalian cells. *J. Biol. Chem.* **266,** 21728–35 (1991).

425. Wei, Y. F. *et al.* Molecular cloning and expression of human cDNAs encoding a novel DNA ligase IV and DNA ligase III, an enzyme active in DNA repair and recombination. *Mol. Cell. Biol.* **15,** 3206–3216 (1995).

426. Akbari, M. *et al.* Overexpression of DNA ligase III in mitochondria protects cells against oxidative stress and improves mitochondrial DNA base excision repair. *DNA Repair (Amst).* **16,** 44–53 (2014).

427. Taylor, R. M., Wickstead, B., Cronin, S. & Caldecott, K. W. Role of a BRCT domain in the interaction of DNA ligase III-alpha with the DNA repair protein XRCC1. *Curr. Biol.* **8,** 877–80 (1998).

428. Nash, R. A., Caldecott, K. W., Barnes, D. E. & Lindahl, T. XRCC1 protein interacts with one of two distinct forms of DNA ligase III. *Biochemistry* **36,** 5207–11 (1997).

429. Cappelli, E. *et al.* Involvement of XRCC1 and DNA Ligase III Gene Products in DNA Base Excision Repair. *J. Biol. Chem.* **272,** 23970–23975 (1997).

430. Oh, S. *et al.* DNA ligase III and DNA ligase IV carry out genetically distinct forms of end joining in human somatic cells. *DNA Repair (Amst).* **21,** 97–110 (2014).

431. Wang, H. *et al.* DNA ligase III as a candidate component of backup pathways of nonhomologous end joining. *Cancer Res.* **65,** 4020–30 (2005).

432. Lindahl, T. DNA Ligase IV from HeLa Cell Nuclei. *J. Biol. Chem.* **271,** 24257–24261 (1996).

433. Critchlow, S. E., Bowater, R. P. & Jackson, S. P. Mammalian DNA double-strand break repair protein XRCC4 interacts with DNA ligase IV. *Curr. Biol.* **7,** 588–98 (1997).

434. Sibanda, B. L. *et al.* Crystal structure of an Xrcc4-DNA ligase IV complex. *Nat. Struct. Biol.* **8,** 1015–9 (2001).

435. Riballo, E. *et al.* Cellular and biochemical impact of a mutation in DNA ligase IV conferring clinical radiosensitivity. *J. Biol. Chem.* **276,** 31124–32 (2001).

436. Riballo, E. *et al.* Identification of a defect in DNA ligase IV in a radiosensitive leukaemia patient. *Curr. Biol.* **9,** 699–S2 (1999).

437. Grawunder, U., Zimmer, D., Fugmann, S., Schwarz, K. & Lieber, M. R. DNA ligase IV is essential for V(D)J recombination and DNA double-strand break repair in human precursor lymphocytes. *Mol. Cell* **2,** 477–84 (1998).

438. Grawunder, U., Zimmer, D., Fugmann, S., Schwarz, K. & Lieber, M. R. DNA Ligase IV Is Essential for V(D)J Recombination and DNA Double-Strand Break Repair in Human Precursor Lymphocytes. *Mol. Cell* **2,** 477–484 (1998).

439. Frank, K. M. *et al.* DNA ligase IV deficiency in mice leads to defective neurogenesis and embryonic lethality via the p53 pathway. *Mol. Cell* **5,** 993–1002 (2000).

440. Block, W. D. & Lees-Miller, S. P. Putative homologues of the DNA-dependent protein kinase catalytic subunit (DNA-PKcs) and other components of the non-homologous end joining machinery in Dictyostelium discoideum. *DNA Repair (Amst).* **4,** 1061–5 (2005).

441. Pears, C. J. & Lakin, N. D. Emerging models for DNA repair: Dictyostelium discoideum as a model for nonhomologous end-joining. *DNA Repair (Amst).* **17,** 121–31 (2014).

442. Sattler, U., Frit, P., Salles, B. & Calsou, P. Long-patch DNA repair synthesis during base excision repair in mammalian cells. *EMBO Rep.* **4,** 363–7 (2003).

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

444. Groth, P. *et al.* Methylated DNA causes a physical block to replication forks independently of damage signalling, O(6)-methylguanine or DNA single-strand breaks and results in DNA damage. *J. Mol. Biol.* **402,** 70–82 (2010).

445. Nikolova, T., Ensminger, M., Löbrich, M. & Kaina, B. Homologous recombination protects mammalian cells from replication-associated DNA double-strand breaks arising in response to methyl methanesulfonate. *DNA Repair (Amst).* **9,** 1050–63 (2010).

446. Gradwohl, G. *et al.* The second zinc-finger domain of poly(ADP-ribose) polymerase determines specificity for single-stranded breaks in DNA. *Proc. Natl. Acad. Sci. U. S. A.* **87,** 2990–4 (1990).

447. Liang, L. *et al.* Human DNA ligases I and III, but not ligase IV, are required for microhomology-mediated end joining of DNA double-strand breaks. *Nucleic Acids Res.* **36,** 3297–310 (2008).

448. Raynard, S., Niu, H. & Sung, P. DNA double-strand break processing: the beginning of the end. *Genes Dev.* **22,** 2903–7 (2008).

449. Moreira, M. C. *et al.* The gene mutated in ataxia-ocular apraxia 1 encodes the new HIT/Zn-finger protein aprataxin. *Nat. Genet.* **29,** 189–93 (2001).

450. Date, H. *et al.* Early-onset ataxia with ocular motor apraxia and hypoalbuminemia is caused by mutations in a new HIT superfamily gene. *Nat. Genet.* **29,** 184–8 (2001).

451. Gueven, N. *et al.* Aprataxin, a novel protein that protects against genotoxic stress. *Hum. Mol. Genet.* **13,** 1081–93 (2004).

452. Sykora, P., Croteau, D. L., Bohr, V. A. & Wilson, D. M. Aprataxin localizes to mitochondria and preserves mitochondrial function. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 7437–42 (2011).

453. Brenner, C. Hint, Fhit, and GalT: Function, Structure, Evolution, and Mechanism of Three Branches of the Histidine Triad Superfamily of Nucleotide Hydrolases and Transferases †. *Biochemistry* **41,** 9003–9014 (2002).

454. Ali, A. A. E., Jukes, R. M., Pearl, L. H. & Oliver, A. W. Specific recognition of a multiply phosphorylated motif in the DNA repair scaffold XRCC1 by the FHA domain of human PNK. *Nucleic Acids Res.* **37,** 1701–12 (2009).

455. Chappell, C., Hanakahi, L. A., Karimi-Busheri, F., Weinfeld, M. & West, S. C. Involvement of human polynucleotide kinase in double-strand break repair by non-homologous end joining. *EMBO J.* **21,** 2827–32 (2002).

456. Kanno, S. *et al.* A novel human AP endonuclease with conserved zinc-finger-like motifs involved in DNA strand break responses. *EMBO J.* **26,** 2094–103 (2007).

457. Bekker-Jensen, S. *et al.* Human Xip1 (C2orf13) is a novel regulator of cellular responses to DNA strand breaks. *J. Biol. Chem.* **282,** 19638–43 (2007).

458. Macrae, C. J., McCulloch, R. D., Ylanko, J., Durocher, D. & Koch, C. A. APLF (C2orf13) facilitates nonhomologous end-joining and undergoes ATM-dependent hyperphosphorylation following ionizing radiation. *DNA Repair (Amst).* **7,** 292–302 (2008).

459. Robu, M. *et al.* Role of poly(ADP-ribose) polymerase-1 in the removal of UV-induced DNA lesions by nucleotide excision repair. *Proc. Natl. Acad. Sci.* **110,** 1658–1663 (2013).

460. Rothkamm, K., Kruger, I., Thompson, L. H. & Lobrich, M. Pathways of DNA Double-Strand Break Repair during the Mammalian Cell Cycle. *Mol. Cell. Biol.* **23,** 5706–5715 (2003).

461. Eastman, A. Characterization of the adducts produced in DNA by cis-diamminedichloroplatinum(II) and cis-dichloro(ethylenediamine)platinum(II). *Biochemistry* **22,** 3927–3933 (1983).

462. Fichtinger-Schepman, A. M. J., Van der Veer, J. L., Den Hartog, J. H. J., Lohman, P. H. M. & Reedijk, J. Adducts of the antitumor drug cis-diamminedichloroplatinum(II) with DNA: formation, identification, and quantitation. *Biochemistry* **24,** 707–713 (1985).

463. Tsao, Y. P., Russo, A., Nyamuswa, G., Silber, R. & Liu, L. F. Interaction between replication forks and topoisomerase I-DNA cleavable complexes: studies in a cell-free SV40 DNA replication system. *Cancer Res.* **53,** 5908–14 (1993).

464. Takebe, H., Furuyama, J.-I., Miki, Y. & Kondo, S. High sensitivity of xeroderma pigmentosum cells to the carcinogen 4-nitroquinoline-1-oxide. *Mutat. Res. Mol. Mech. Mutagen.* **15,** 98–100 (1972).

465. Akkari, Y. M., Bateman, R. L., Reifsteck, C. A., Olson, S. B. & Grompe, M. DNA replication is required To elicit cellular responses to psoralen-induced DNA interstrand cross-links. *Mol. Cell. Biol.* **20,** 8283–9 (2000).

466. Wang, A. T. *et al.* Human SNM1A and XPF-ERCC1 collaborate to initiate DNA interstrand cross-link repair. *Genes Dev.* **25,** 1859–70 (2011).

467. Jackson, S. P. & Durocher, D. Regulation of DNA damage responses by ubiquitin and SUMO. *Mol. Cell* **49,** 795–807 (2013).

468. Moldovan, G.-L. & D'Andrea, A. D. How the fanconi anemia pathway guards the genome. *Annu. Rev. Genet.* **43,** 223–49 (2009).

469. Thompson, L. H., Brookman, K. W., Jones, N. J., Allen, S. A. & Carrano, A. V. Molecular cloning of the human XRCC1 gene, which corrects defective DNA strand break repair and sister chromatid exchange. *Mol. Cell. Biol.* **10,** 6160–71 (1990).

470. Li, Z. *et al.* The XRCC4 gene encodes a novel protein involved in DNA double-strand break repair and V(D)J recombination. *Cell* **83,** 1079–1089 (1995).

471. Schärer, O. D. DNA interstrand crosslinks: natural and drug-induced DNA adducts that induce unique cellular responses. *Chembiochem* **6,** 27–32 (2005).

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

473. Cole, R. S. Repair of DNA containing interstrand crosslinks in Escherichia coli: sequential excision and recombination. *Proc. Natl. Acad. Sci. U. S. A.* **70,** 1064–8 (1973).

474. Miller, R. D., Prakash, L. & Prakash, S. Genetic control of excision of Saccharomyces cerevisiae interstrand DNA cross-links induced by psoralen plus near-UV light. *Mol. Cell. Biol.* **2,** 939–948 (1982).

475. Jachymczyk, W. J., von Borstel, R. C., Mowat, M. R. A. & Hastings, P. J. Repair of interstrand cross-links in DNA of Saccharomyces cerevisiae requires two systems for DNA repair: The RAD3 system and the RAD51 system. *Mol. Gen. Genet. MGG* **182,** 196–205 (1981).

476. Sarkar, S., Davies, A. A., Ulrich, H. D. & McHugh, P. J. DNA interstrand crosslink repair during G1 involves nucleotide excision repair and DNA polymerase zeta. *EMBO J.* **25,** 1285–94 (2006).

477. Li, X., Hejna, J. & Moses, R. E. The yeast Snm1 protein is a DNA 5'-exonuclease. *DNA Repair (Amst).* **4,** 163–70 (2005).

478. Aravind, L., Walker, D. R. & Koonin, E. V. Conserved domains in DNA repair proteins and evolution of repair systems. *Nucleic Acids Res.* **27,** 1223–42 (1999).

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

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

481. Shen, X. *et al.* REV3 and REV1 play major roles in recombination-independent repair of DNA interstrand cross-links mediated by monoubiquitinated proliferating cell nuclear antigen (PCNA). *J. Biol. Chem.* **281,** 13869–72 (2006).

482. Hanada, K. *et al.* The structure-specific endonuclease Mus81-Eme1 promotes conversion of interstrand DNA crosslinks into double-strands breaks. *EMBO J.* **25,** 4921–32 (2006).

483. Kuraoka, I. *et al.* Repair of an interstrand DNA cross-link initiated by ERCC1-XPF repair/recombination nuclease. *J. Biol. Chem.* **275,** 26632–6 (2000).

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

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

486. Simpson, L. J. & Sale, J. E. Rev1 is essential for DNA damage tolerance and non-templated immunoglobulin gene mutation in a vertebrate cell line. *EMBO J.* **22,** 1654–64 (2003).

487. Sonoda, E. *et al.* Multiple roles of Rev3, the catalytic subunit of polzeta in maintaining genome stability in vertebrates. *EMBO J.* **22,** 3188–97 (2003).

488. Yamamoto, K. *et al.* Fanconi anemia protein FANCD2 promotes immunoglobulin gene conversion and DNA repair through a mechanism related to homologous recombination. *Mol. Cell. Biol.* **25,** 34–43 (2005).

489. Nakanishi, K. *et al.* Human Fanconi anemia monoubiquitination pathway promotes homologous DNA repair. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 1110–5 (2005).

490. De Silva, I. U., McHugh, P. J., Clingen, P. H. & Hartley, J. A. Defining the Roles of Nucleotide Excision Repair and Recombination in the Repair of DNA Interstrand Cross-Links in Mammalian Cells. *Mol. Cell. Biol.* **20,** 7980–7990 (2000).

491. SASAKI, M. S. Is Fanconi's anaemia defective in a process essential to the repair of DNA cross links? *Nature* **257,** 501–503 (1975).

492. Ward, T. A. *et al.* Components of a Fanconi-like pathway control Pso2-independent DNA interstrand crosslink repair in yeast. *PLoS Genet.* **8,** e1002884 (2012).

493. Foe, J. R. *et al.* Expression cloning of a cDNA for the major Fanconi anaemia gene, FAA. *Nat. Genet.* **14,** 488 (1996).

494. Meetei, A. R. *et al.* X-linked inheritance of Fanconi anemia complementation group B. *Nat. Genet.* **36,** 1219–24 (2004).

495. Strathdee, C. A., Gavish, H., Shannon, W. R. & Buchwald, M. Cloning of cDNAs for Fanconi's anaemia by functional complementation. *Nature* **358,** 434 (1992).

496. De Winter, J. P. *et al.* Isolation of a cDNA representing the Fanconi anemia complementation group E gene. *Am. J. Hum. Genet.* **67,** 1306–8 (2000).

497. De Winter, J. P. *et al.* The Fanconi anaemia gene FANCF encodes a novel protein with homology to ROM. *Nat. Genet.* **24,** 15–6 (2000).

498. De Winter, J. P. *et al.* The Fanconi anaemia group G gene FANCG is identical with XRCC9. *Nat. Genet.* **20,** 281–3 (1998).

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

500. Meetei, A. R. *et al.* A human ortholog of archaeal DNA repair protein Hef is defective in Fanconi anemia complementation group M. *Nat. Genet.* **37,** 958–63 (2005).

501. Ling, C. *et al.* FAAP100 is essential for activation of the Fanconi anemia-associated DNA damage response pathway. *EMBO J.* **26,** 2104–14 (2007).

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

503. Kim, J. M., Kee, Y., Gurtan, A. & D'Andrea, A. D. Cell cycle-dependent chromatin loading of the Fanconi anemia core complex by FANCM/FAAP24. *Blood* **111,** 5215–22 (2008).

504. Machida, Y. J. *et al.* UBE2T is the E2 in the Fanconi anemia pathway and undergoes negative autoregulation. *Mol. Cell* **23,** 589–96 (2006).

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

506. Garcia-Higuera, I. *et al.* Interaction of the Fanconi anemia proteins and BRCA1 in a common pathway. *Mol. Cell* **7,** 249–62 (2001).

507. Dorsman, J. C. *et al.* Identification of the Fanconi anemia complementation group I gene, FANCI. *Cell. Oncol.* **29,** 211–8 (2007).

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

509. Timmers, C. *et al.* Positional cloning of a novel Fanconi anemia gene, FANCD2. *Mol. Cell* **7,** 241–8 (2001).

510. Sims, A. E. *et al.* FANCI is a second monoubiquitinated member of the Fanconi anemia pathway. *Nat. Struct. Mol. Biol.* **14,** 564–7 (2007).

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

512. Nijman, S. M. B. *et al.* The deubiquitinating enzyme USP1 regulates the Fanconi anemia pathway. *Mol. Cell* **17,** 331–9 (2005).

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

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

515. MacKay, C. *et al.* Identification of KIAA1018/FAN1, a DNA repair nuclease recruited to DNA damage by monoubiquitinated FANCD2. *Cell* **142,** 65–76 (2010).

516. Smogorzewska, A. *et al.* A genetic screen identifies FAN1, a Fanconi anemia-associated nuclease necessary for DNA interstrand crosslink repair. *Mol. Cell* **39,** 36–47 (2010).

517. Liu, T., Ghosal, G., Yuan, J., Chen, J. & Huang, J. FAN1 acts with FANCI-FANCD2 to promote DNA interstrand cross-link repair. *Science* **329,** 693–6 (2010).

518. Kratz, K. *et al.* Deficiency of FANCD2-associated nuclease KIAA1018/FAN1 sensitizes cells to interstrand crosslinking agents. *Cell* **142,** 77–88 (2010).

519. Wang, R. *et al.* DNA repair. Mechanism of DNA interstrand cross-link processing by repair nuclease FAN1. *Science* **346,** 1127–30 (2014).

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

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

522. Yamamoto, K. N. *et al.* Involvement of SLX4 in interstrand cross-link repair is regulated by the Fanconi anemia pathway. *Proc. Natl. Acad. Sci.* **108,** 6492–6496 (2011).

523. Litman, R. *et al.* BACH1 is critical for homologous recombination and appears to be the Fanconi anemia gene product FANCJ. *Cancer Cell* **8,** 255–65 (2005).

524. Levitus, M. *et al.* The DNA helicase BRIP1 is defective in Fanconi anemia complementation group J. *Nat. Genet.* **37,** 934–5 (2005).

525. Howlett, N. G. *et al.* Biallelic inactivation of BRCA2 in Fanconi anemia. *Science* **297,** 606–9 (2002).

526. Hussain, S. *et al.* Direct interaction of FANCD2 with BRCA2 in DNA damage response pathways. *Hum. Mol. Genet.* **13,** 1241–8 (2004).

527. Wang, X., Andreassen, P. R. & D'Andrea, A. D. Functional interaction of monoubiquitinated FANCD2 and BRCA2/FANCD1 in chromatin. *Mol. Cell. Biol.* **24,** 5850–62 (2004).

528. Xia, B. *et al.* Control of BRCA2 cellular and clinical functions by a nuclear partner, PALB2. *Mol. Cell* **22,** 719–29 (2006).

529. Adamo, A. *et al.* Preventing nonhomologous end joining suppresses DNA repair defects of Fanconi anemia. *Mol. Cell* **39,** 25–35 (2010).

530. Pace, P. *et al.* Ku70 Corrupts DNA Repair in the Absence of the Fanconi Anemia Pathway. *Science (80-.).* **329,** 219–223 (2010).

531. Hazrati, A. *et al.* Human SNM1A suppresses the DNA repair defects of yeast pso2 mutants. *DNA Repair (Amst).* **7,** 230–8 (2008).

532. Fujimuro, M., Sawada, H. & Yokosawa, H. Production and characterization of monoclonal antibodies specific to multi-ubiquitin chains of polyubiquitinated proteins. *FEBS Lett.* **349,** 173–180 (1994).

533. Jentsch, S., McGrath, J. P. & Varshavsky, A. The yeast DNA repair gene RAD6 encodes a ubiquitin-conjugating enzyme. *Nature* **329,** 131–4 (1987).

534. Ulrich, H. D. Timing and spacing of ubiquitin-dependent DNA damage bypass. *FEBS Lett.* **585,** 2861–7 (2011).

535. Nakamura, K. *et al.* Regulation of homologous recombination by RNF20-dependent H2B ubiquitination. *Mol. Cell* **41,** 515–28 (2011).

536. Moyal, L. *et al.* Requirement of ATM-dependent monoubiquitylation of histone H2B for timely repair of DNA double-strand breaks. *Mol. Cell* **41,** 529–42 (2011).

537. Panier, S. *et al.* Tandem protein interaction modules organize the ubiquitin-dependent response to DNA double-strand breaks. *Mol. Cell* **47,** 383–95 (2012).

538. Mattiroli, F. *et al.* RNF168 ubiquitinates K13-15 on H2A/H2AX to drive DNA damage signaling. *Cell* **150,** 1182–95 (2012).

539. Gatti, M. *et al.* A novel ubiquitin mark at the N-terminal tail of histone H2As targeted by RNF168 ubiquitin ligase. *Cell Cycle* **11,** 2538–44 (2012).

540. Huen, M. S. Y. *et al.* RNF8 transduces the DNA-damage signal via histone ubiquitylation and checkpoint protein assembly. *Cell* **131,** 901–14 (2007).

541. Bekker-Jensen, S. *et al.* HERC2 coordinates ubiquitin-dependent assembly of DNA repair factors on damaged chromosomes. *Nat. Cell Biol.* **12,** 80–6; sup pp 1–12 (2010).

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

543. Wu, J. *et al.* Histone ubiquitination associates with BRCA1-dependent DNA damage response. *Mol. Cell. Biol.* **29,** 849–60 (2009).

544. Kolas, N. K. *et al.* Orchestration of the DNA-damage response by the RNF8 ubiquitin ligase. *Science* **318,** 1637–40 (2007).

545. Yan, Z. *et al.* A ubiquitin-binding protein, FAAP20, links RNF8-mediated ubiquitination to the Fanconi anemia DNA repair network. *Mol. Cell* **47,** 61–75 (2012).

546. Baba, D. *et al.* Crystal structure of thymine DNA glycosylase conjugated to SUMO-1. *Nature* **435,** 979–82 (2005).

547. Steinacher, R. & Schär, P. Functionality of human thymine DNA glycosylase requires SUMO-regulated changes in protein conformation. *Curr. Biol.* **15,** 616–23 (2005).

548. Morris, J. R. *et al.* The SUMO modification pathway is involved in the BRCA1 response to genotoxic stress. *Nature* **462,** 886–90 (2009).

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

550. Galanty, Y., Belotserkovskaya, R., Coates, J. & Jackson, S. P. RNF4, a SUMO-targeted ubiquitin E3 ligase, promotes DNA double-strand break repair. *Genes Dev.* **26,** 1179–95 (2012).

551. Zhao, Q. *et al.* GPS-SUMO: a tool for the prediction of sumoylation sites and SUMO-interaction motifs. *Nucleic Acids Res.* **42,** W325–30 (2014).

552. O'Donovan, A., Davies, A. A., Moggs, J. G., West, S. C. & Wood, R. D. XPG endonuclease makes the 3' incision in human DNA nucleotide excision repair. *Nature* **371,** 432–5 (1994).

553. McHugh, P. J., Sones, W. R. & Hartley, J. A. Repair of Intermediate Structures Produced at DNA Interstrand Cross-Links in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **20,** 3425–3433 (2000).

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

555. Bhattacharyya, A., Ear, U. S., Koller, B. H., Weichselbaum, R. R. & Bishop, D. K. The breast cancer susceptibility gene BRCA1 is required for subnuclear assembly of Rad51 and survival following treatment with the DNA cross-linking agent cisplatin. *J. Biol. Chem.* **275,** 23899–903 (2000).

556. Petrucco, S. Sensing DNA damage by PARP-like fingers. *Nucleic Acids Res.* **31,** 6689–6699 (2003).

557. Mortusewicz, O., Rothbauer, U., Cardoso, M. C. & Leonhardt, H. Differential recruitment of DNA Ligase I and III to DNA repair sites. *Nucleic Acids Res.* **34,** 3523–32 (2006).

558. Bendixen, C., Thomsen, B., Alsner, J. & Westergaard, O. Camptothecin-stabilized topoisomerase I-DNA adducts cause premature termination of transcription. *Biochemistry* **29,** 5613–5619 (1990).

559. Wu, W., Koike, A., Takeshita, T. & Ohta, T. The ubiquitin E3 ligase activity of BRCA1 and its biological functions. *Cell Div.* **3,** 1 (2008).

9. Appendix

9.1. Appendix A: Primer Sequences

Regions of the primers sequences containing restriction sites are italicised, whereas regions encoding epitope tags are underlined.

General screening primers

EX1L

5’-AACTGCAGGTGTCATCAGAATCTAGTGTACTAGAATC-3’

EL1R

5’-AAGGATCCCACGTGATGAAAATGATTTCCTTGATTG-3’

BSRF

5’-GAAGTTATCATATGCCGCATGG-3’

BSRR

5’-ATGCTATACGAAGTTATCCGTGG-3’

Generating Lig3 constructs

GST-Lig3-MACRO-L (*BamHI*)

5’-CC*GGATCC*ATGTTAAAATTTGTTAGTGATG-3’

GST-Lig3-MACRO-R (*XhoI*)

5’-GG*CTCGAG*TTATGAATGTACAAATACTTTAATAC-3’

Myc-Lig3-L (*KpnI*)

5’-AAA*GGTACC*GAACAAAAATTAATTTCAGAAGAAGATTTAATGTCAGAGGATAAATCAGGATCAT-3’

Myc-Lig3-R (*XhoI*)

5’-AAA*CTCGAG*TTAGAAAAGTTTATAATTTTTTACATCTAAAAGATCACTC-3’

Disruption of the *lig3* gene and screening

LLA1

5’-TGGTATGGATCCAGTTACAAGAGG-3’

LLA2 (*PstI*)

5’-*CTGCAG*CACATAATTTATACATTGAGTAAAATGATCCTG-3’

LRA1 (*ClaI*)

5’-*ATCGAT*AACAGCAACAACAAGAACCATC-3’

LRA2 (*KpnI*)

5’-*GGTACC*GAACCAAAACATTGATCAACCC-3’

L3I1

5’-TGATAATGATAATGGTGATGGCG-3’

L3I2

5’-CACCATTAACAATTGTCACCTCAG-3’

L3LA

5’-GGTATTGGTAGAATTGGTATTGTAGGTTG-3’

L3RA (*XhoI*)

5’-AAA*CTCGAG*GTAGAAAAGTTTATAATTTTTTACATCTAAAAGATCACTC-3’

Generating APL constructs

GST-APL-L (*BamHI*)

5’-AAA*GGATCC*ATGAAAAAAATAAATAATTGTTTAAATATTAAATGTAT-3’

GST-APL-R (*XhoI*)

5’- GCAA*CTCGAG*TTATTCAATTTGATCGAATAAGTCATC-3’

Myc-APL-L (*BamHI*)

5’-AAA*GGATCC*GAACAACCCTTAATTTCAGAAGAAGATTTAATGAAAAAAATAAATAATTGTTTAAATATT

AAATGTAT-3’

Myc-APL-R (*XbaI*)

5’-GCAA*TCTAGA*TTATTCAATTTGATCGAATAAGTCATC-3’

Screening of *apl*⁻ cells

APL1

5’-GGTATTAATCCATCATACTTAAAGAAAGC-3’

APL2

5’-GAACTGTATTTTGTACTTTTGAACCTG-3’

CHK1

5’-TCAGAAGAATTACTACTGCTACCACTAC-3’

CHK2

5’-TGTACCCTTAATTAAATTCCCAA-3’

Disruption of the *fncD2* gene and screening

FLA1

5’-GAGAAATCGATTGAAGTTGATGG-3’

FLA2 (*HindIII*)

5’-TTT*AAGCTT*AAGAACAAGGTGTATCATATAAACGTTG-3’

FRA1 (*PstI*)

5’-AAA*CTGCAG*GAATTGGAAGAAGAATATGATGCTAGC-3’

FRA2 (*BamHI*)

5’-TTTGGATCCCTTTGAGAACCTAAGCTATCGCC-3’

D2I1

5’-ATCATTTAAAATCTCTGTTGTTGATCC-3’

D2I2

5’-GAGTCTATATTAGAATTAGCATTATCACCAGTCAG-3’

FD2L

5’-GAAACTTCTAATTTCGATAATAATGAATCCC-3’

FD2R

5’- AATCAATCATTGTGTTTAGCAGTGG -3’