Understanding the Role of the SNM1B and EXD2 Nucleases in DNA Damage Repair



Thesis submitted for the degree of Doctor of Philosophy

Hannah Tamara Baddock

University College

Department of Oncology

University of Oxford

Trinity Term 2017

Abstract

Understanding the Role of the SNM1B and EXD2 Nucleases in DNA Damage Repair

Hannah Baddock, University College

Thesis submitted for the degree of Doctor of Philosophy, Trinity Term 2017

Unrepaired, or misrepaired, DNA damage can be carcinogenic or mutagenic; thus functional DNA damage repair pathways are essential for the safeguarding of the genome.

SNM1B is a 5' to 3' exonuclease implicated in the repair of damaged DNA, particularly the repair of interstrand crosslinks. Genetic studies have identified SNPs in the SNM1B gene as related to cancer risk. One of these (rs11552449) encodes a single amino acid change, H61Y. This study shows that WT and H61Y SNM1B have comparable *in vitro* biochemical and biophysical characteristics. The structures of both WT and H61Y C-terminally truncated SNM1B (Δ-SNM1B) were solved to 2.8 and 3.1 Å, respectively, and reveal similar structural properties. The structure of WT Δ-SNM1B was also reported (to 1.8 Å) with two 2'-deoxy-5'-adenosine monophosphate molecules in the active site. The structure of SNM1B shows an accessible extended active site, which may facilitate the binding of a variety of non-canonical DNA substrates. Accordingly, *in vitro,* WT and H61Y SNM1B are able to exonucleolytically process a wide range of structurally diverse DNA substrates. By utilising SNM1B depleted cell lines, this study also shows that SNM1B is required for DNA repair in response to treatment with DNA-crosslinking genotoxic agents (including cisplatin and SJG-136).

This study also identifies the novel double strand break repair factor, EXD2, as having intrinsic 3' to 5' exonuclease activity. EXD2 was shown to have enzymatic activity on a variety of substrates *in vitro*, including replication fork intermediates, 'nicked' or 'gapped' DNA duplexes, and RNA based substrates. Together with the cellular data this suggests a role for EXD2 in nucleolytically processing RNA or DNA-based intermediates in damage repair pathways.

Acknowledgements

First, thank you to my supervisors, Prof. Peter McHugh and Dr Opher Gileadi, your knowledge, support, and guidance has been invaluable. Thank you particularly to Peter, who agreed to take a student from halfway around the world, that he'd only met over Skype at unsociable hours. Thank you both for your scientific rigour, enthusiasm, and encouragement.

I am extremely grateful to everyone in Peter's and Opher's labs. Thanks must go to the following: Sook, for being the world's best Superpal, and for teaching me how to do all of the biochemical assays; Lonnie, for your support, encouragement, and guiding me to become a better thinker; Lianne, for being my fellow DPhil student/circuits pal, and only dropping me on the floor once; Anna, for always being kind, helpful, and bringing me vegetables from your garden; Sanja, for all the happiness, hugs, and rowing adventures; Will, for being my SNM1 pal; and Jamie, for your knowledge and input. Enormous thanks must also go to Joe and Hazel for helping me with protein purification and crystallography. Thank you particularly to Joe for assisting with large parts of my project, explaining crystallography concepts to me, and patiently proof-reading Chapter 4, twice.

Thank you also to the Niedzwiedz lab for the work on EXD2. Thank you to Jadwiga, Ronan, and Wojciech for the scientific discussions and the opportunity to collaborate on this new and interesting project. Thank you also to the Blackford lab for being excellent labbies and always offering advice when needed. Thank you particularly to Sam for helping me with all of my formatting, and generally being my thesis-writing go-to guy.

Thanks must also go to my OUWBC, rowing, and Univ friends for the support outside the lab, and particular thanks to Nadine and Dora, for four years of JR lunch club. Thank you also to my family back home in New Zealand – your endless support and the care packages of tea and chocolate, were essential for the writing of this thesis. Thank you also to Jonathan – we matriculated together, and you've been there ever since. The completion of this thesis would have been much tougher without your support and hugs.

I am also very grateful to CRUK for funding my studies over the last four years and for providing me with this opportunity.

List of Abbreviations

(6-4)-PP	6-4 photoproducts
2'-deoxy-5-AMP	2'-deoxyadenosine-5'-monophosphate
53BP1	p53 binding protein 1
8-oxo-G	8-oxo-guanine
ALDH2	aldehyde dehydrogenase 2
ALT	alternative lengthening of telomeres
alt-NHEJ	alternative-NHEJ
AP	apurinic/ apyrimidinic
APLF	aprataxin-PNKP-like factor
APTX	aprataxin
ATM	ataxia telangiectasia mutated
ATP	adenosine triphosphate
ATR	ataxia telangiectasia and Rad3 related
ATRIP	ATR-interacting protein
BER	base excision repair
BIR	break induced replication
BLM	Bloom syndrome protein
BRCA1/2	breast cancer type 1/2 susceptibility protein
BSA	bovine serine albumin
c-NHEJ	classical-NHEJ
$CaCl_2$	calcium chloride
Cas9	CRISPR-associated protein 9
CASP	CPSF, Artemis, SNM1A, and PSO2
CDK	cyclin dependent kinase
CHK1/2	checkpoint kinase 1/2
CHO	Chinese hamster ovary
CMAHP	cytodine monophosphate-N-acetylneuraminic acid hydroxylase-like protein
CMG	Cdc45-MCM2-7-GINS
CO	crossover
CO_2	carbon dioxide
CPD	cyclobutane pyrimidine dimer
CPSF-100	cleavage and polyadenylation specific factor-100

CPSF-73	cleavage and polyadenylation specific factor-73
CRISPR	clustered regularly spaced short palindromic repeats
CS	Cockayne syndrome
CSB	Cockayne syndrome group B protein
CtIP	C-terminal-binding protein interacting protein
D-loop	displacement loop
dATP	deoxyadenosine triphosphate
DCLRE1BA/B/C	DNA crosslink repair 1A/B/C
DDP	diamminedichloroplatinum
DDR	DNA damage response
dHJ	double Holliday junction
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
DNA-PK	DNA-dependent protein kinase
DNA-PKcs	DNA-dependent protein kinase, catalytic subunit
DNA2	DNA replication ATP-dependent helicase/nuclease
dNTP	deoxynucleotide triphosphate
DPC	DNA protein crosslink
DSB	double strand break
DSBR	double strand break repair
dsDNA	double stranded DNA
DTT	dithiothreitol
E. coli	Escherichia coli
EDTA	ethylenediaminetetraacetic acid
EGTA	ethylene glycol tetraacetic acid
ELAC1/2	ElaC ribonuclease Z 1/2
EME1/2	essential meiotic structure-specific endonuclease 1/2
EMSA	electrophoretic mobility shift assay
ERCC1	endonuclease non-catalytic subunit
ESI-TOF-MS	electrospray ionisation time of flight mass spectrometry
ETHE1	ethylmalonic encelphalopathy 1 protein
EXD2	exonuclease 3’ – 5’ domain containing protein 2
EXO1	exonuclease 1
FA	Fanconi anaemia
FAAP20	FA core complex associated protein of 20 kDa

FAAP24	FA core complex association protein of 24 kDa
FAN1	FANCD2/FANCI associated nuclease 1
FANC	Fanconi anaemia complementation group
FANCD2	Fanconi anaemia complementation group D2
FCS	foetal calf serum
$FeCl_2$	iron chloride
FEN1	flap structure-specific endonuclease 1
GAPDH	glyceraldehyde 3-phosphate dehydrogenase
gDNA	genomic DNA
GFP	green fluorescent protein
GWAS	genome wide association study
HAGH	hydroxyacylglutathione hydrolase
HAGH-L	hydroxyacylglutathione hydrolase-like
HCl	hydrochloric acid
HCS70	heat-shock cognate 70
HDR	homology directed repair
HEK	human embryonic kidney
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HH	Hoyeraal-Hreidarsson
His_{10}	deca-histidine
His_6	hexa-histidine
HR	homologous recombination
HSP60/70/72	heat shock protein 60/70/72
hTERT	human telomerase reverse transciptase
ICL	interstrand crosslink
Int11	Integrator complex subunit 11
Int9	Integrator complex subunit 9
IPTG	isopropyl β-D-1-thiogalactopyranoside
IR	ionising radiation
KCl	potassium chloride
KIN	karyomegalic interstitial nephritis
LACTB2	lactamase β 2
LB	lysogeny broth
LIG1/3/4	DNA ligase 1/3/4
LOH	loss of heterozygosity

M1G	pyrimido(1,2-a)purin-10(3H)-one
MBL	metallo-β-lactamase
MBLAC1/2	metallo-β-lactamase domain-containing protein 1/2
MDC1	mediator of DNA damage checkpoint protein 1
MEF	mouse embryonic fibroblast
$MgCl_2$	magnesium chloride
MHF1/2	Mph1-associated histone-fold protein 1/2
min	minutes
MMC	mitomycin C
MMR	mismatch repair
$MnCl_2$	manganese chloride
MRN	MRE11-RAD50-NBS1
mRNA	messenger RNA
MS	mass spectrometry
NaCl	sodium chloride
NaOH	sodium hydroxide
NAPE-PLD	N-acyl-phosphatidylethanolamine phospholipase D
NCO	non-crossover
NER	nucleotide excision repair
NHEJ	non-homologous end joining
$NiCl_2$	nickel chloride
PALB2	partner and localiser of BRCA2
PAM	protospacer adjacent motif
PARP1	poly(ADP-ribose) polymerase 1
PBS	phosphate-buffered saline
PCNA	proliferating cell nuclear antigen
PCR	polymerase chain reaction
PEG	polyethylene glycol
PMSF	phenylmethane sulfonyl fluoride
PNK	polynucleotide kinase
PNKP	polynucleotide kinase 3’ phosphatase
Pol	polymerase
POT1	protection of telomeres protein 1
PSO	sensitive to psoralen
qRT-PCR	quantitative real time polymerase chain reaction

RAP1	repressor/activator protein
RMI1	RecQ mediated genome instability 1
RMSD	root mean squared deviation
RNA	ribonucleic acid
RNAi	RNA interference
ROS	reaction oxygen species
RPA	replication protein A
RTEL1	regulator of telomere elongation helicase 1
SCE	sister chromatid exchange
SD	standard deviation
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SDSA	synthesis-dependent strand annealing
sec	second
SEM	standard error of the mean
sgRNA	single guide RNA
SNM1	sensitive to nitrogen mustard 1
SNM1A/B/C	sensitive to nitrogen mustard A/B/C
SNP	single nucleotide polymorphism
SOC	super optimal broth with catabolite repression
SSA	single strand annealing
SSB	single strand break
ssDNA	single stranded DNA
t-loop	telomere loop
TB	terrific broth
TCEP	tris(2-carboxyethyl)phosphine
TdT	terminal deoxynucleotidyl transferase
TEV	tobacco etch virus
TFIIH	transcription factor II H
TIF	telomere dysfunction induced foci
TIN2	TRF1 and TRF2 interacting nuclear factor 2
TLS	translesion synthesis
T_m	melting temperature
Top1/2	DNA topoisomerase 1/2
Topo3α	DNA topoisomerase 3α
TRF1/2	telomeric repeat-binding factor 1/2

TRFH	TRF2 homology
UBZ	ubiquitin binding zinc finger
UV	ultraviolet
v/v	volume per volume
w/v	weight per volume
WRN	Werner protein
WT	wild type
XLF	XRCC4-like factor
XP	xeroderma pigmentosum
XPA/B/C/G	xeroderma pigmentosum group A/B/C/G-complementing protein
XPF	xeroderma pigmentosum group F-complementing protein
XRCC1/4	X-ray repair cross-complementing protein 1/4
XRF	X-ray fluorescence
$ZnCl_2$	zinc chloride
γ-H2AX	phosphorylated histone variant H2AX at serine 139
Δ-EXD2	truncated EXD2 protein
Δ-SNM1B	C-terminally truncated SNM1B protein

Table of Contents

1. Introduction 1

1.1. Threats to genome integrity 1

1.1.1. Endogenous sources of DNA damage 2

1.1.2. Exogenous sources of DNA damage 4

1.2. The DNA damage response 6

1.2.1. The role of ATM in the DDR 7

1.2.2. The role of ATR in the DDR 8

1.3. DNA damage repair pathways 10

1.4. Double strand break (DSB) repair 13

1.4.1. Non-homologous end-joining 15

1.4.2. Homologous recombination 20

1.5. Interstrand crosslink repair 26

1.5.1. Endogenous ICL inducing agents 27

1.5.2. Exogenous ICL inducing agents 29

1.5.3. ICL repair in yeast 37

1.5.4. ICL repair in vertebrates 37

1.6. SNM1B 48

1.6.1. The *SNM1* family 48

1.6.2. Structural characteristics of SNM1B 50

1.6.3. The role of SNM1B at telomeres 56

1.6.4. The role of SNM1B in DNA damage repair 63

1.6.5. The role of SNM1B in human disease 69

1.7. The Aims of the current project 71

2. Methods and Materials 73

2.1. Cloning into baculovirus for WT and mutant SNM1B 73

2.1.1. Site directed mutagenesis PCR 73

2.1.2. DpnI digest 73

2.1.3. Transformations into chemically competent XL1-Blue *E. coli* 74

2.1.4. Plasmid DNA purification 74

2.1.5. PCR to add extensions for LIC 74

2.1.6. T4 polymerase treatment of PCR products 75

2.1.7. Restriction digest of the transfer vector 75

2.1.8. T4 polymerase treatment of the transfer vector 76

2.1.9. Annealing into the appropriate transfer vector 76

2.1.10. Transformation into chemically competent MACH-1 *E. coli* 76

2.1.11. Colony PCR screen 77

2.1.12. Agarose gel electrophoresis 77

2.1.13. Bacmid generation by transposition 78

2.1.14. Bacmid PCR screen 78

2.1.15. Transfection of Sf9 insect cells 79

2.2. Purification of WT and mutant Δ-SNM1B 80

2.2.1. SDS-PAGE to determine the presence of WT and mutant SNM1B .. 81

2.3. Biochemical characterisation of WT and mutant SNM1B 82

2.3.1. Generation of 3’ radiolabelled DNA substrates 82

2.3.2. Gel-based nuclease assay 83

2.3.3. Electrophoretic mobility shift assay 83

2.3.4. Real-time fluorescence nuclease assay 84

2.3.5. Fluorescence polarisation DNA binding assay 85

2.3.6. Thermal stability assay 85

2.4. Crystallisation of WT and H61Y SNM1B 86

2.4.1. Crystallisation of apo-form WT and H61Y Δ-SNM1B 86
2.4.2. Co-crystallisation of WT Δ-SNM1B with 2'-deoxy-5'-AMP 86
2.5. Analysis of the cellular role of SNM1B 87
2.5.1. Generation of plasmids for CRISPR-Cas9 87
2.5.2. Culture of HeLa1.2.11 cells 88
2.5.3. Transfection and sorting of CRISPR cells 88
2.5.4. PCR screen of CRISPR clones 89
2.5.5. Determining mRNA expression by quantitative real-time PCR 89
2.5.6. Storage of genotoxic drugs 90
2.5.7. Resazurin viability assays 91
2.5.8. Clonogenic survival assays 91
2.6. Purification of WT and D108A/E110A Δ-EXD2 92
2.7. Biochemical characterisation of WT and D108A/E110A EXD2 93
2.7.1. Generation of 5' radiolabelled substrates 93
2.7.2. Gel-based nuclease assay 94
2.7.3. Fluorescence polarisation DNA binding assay 94
3. A Biochemical Analysis of WT and mutant SNM1B 95
3.1. Introduction 95
3.2. Results 96
3.2.1. Cloning of WT and mutant SNM1B 96
3.2.2. Purification of WT and mutant SNM1B constructs 106
3.2.3. DNA binding and nuclease activity of WT SNM1B 118
3.2.4. Initial biochemical and biophysical characterisation of WT and mutant SNM1B 124
3.2.5. Analysis of WT and H61Y nuclease activity on structurally diverse DNA substrates 129

3.2.6. Analysis of WT and H61Y SNM1B on telomeric DNA substrates ... 130
3.2.7. Analysis of WT and H61Y SNM1B on 8-oxo-guanine containing substrates ... 133
3.2.8. Processivity of WT and H61Y SNM1B ... 135
3.2.9. Substrate length preference of WT and H61Y SNM1B ... 137
3.3. Conclusions ... 139
4. Crystallographic analysis of WT and H61Y SNM1B ... 142
4.1. Introduction ... 142
4.2. Results ... 144
4.2.1. Generation of WT and H61Y SNM1B crystal structures ... 144
4.2.2. Comparison of the crystal structure of WT and H61Y SNM1B ... 146
4.2.3. Generation of a WT SNM1B:DNA co-crystal structure ... 153
4.2.4. WT SNM1B structure with 2'-deoxy-5'-AMP ... 160
4.2.5. Understanding the interaction of SNM1B with DNA ... 163
4.2.6. The importance of the phosphate binding pocket in the extended active site of WT SNM1B ... 165
4.2.7. Understanding the catalytic core of WT SNM1B ... 167
4.2.8. The interaction of SNM1B with a modelled DNA substrate ... 170
4.2.9. Is there a structural basis for RNA v DNA selectivity for SNM1B? ... 175
4.2.10. The active site of SNM1B accommodates a damaged substrate ... 177
4.2.11. The requirements for metal ion cofactor(s) for catalysis of SNM1B 182
4.3. Conclusion ... 189
5. Analysis of the cellular role of WT SNM1B ... 193
5.1. Introduction ... 193
5.2. Results ... 195
5.2.1. Generation of an SNM1B disrupted cell line using CRISPR-Cas9 . 195

5.2.2. Validation of an SNM1B disrupted cell line 200
5.2.3. Survival assays 219
5.2.4. Clonogenic survival assays 221
5.3. Conclusions 225
6. A Biochemical Analysis of the Novel Nuclease, EXD2 229
6.1. Introduction 229
6.2. Results 230
6.2.1. Cloning the EXD2 gene 230
6.2.2. Purification of Δ-EXD2 (K76–K283) and (K76–G286) 237
6.2.3. Purification of Δ-EXD2 (K76–S589) 245
6.2.4. Purification of D108A/E110A Δ-EXD2 (K76–S589) 254
6.2.5. Analysis of *in vitro* nuclease activity and DNA binding of EXD2 261
6.2.6. Active Site Metal Ion Analysis for EXD2 266
6.2.7. Investigation of the nuclease activity of EXD2 on structurally diverse DNA substrates 270
6.2.8. Nuclease activity of EXD2 on nicked DNA substrates 274
6.2.9. Analysis of the nuclease activity of EXD2 on different length DNA substrates 279
6.2.10. Analysis of the processivity of EXD2 282
6.2.11. A novel role for EXD2 in processing RNA substrates 285
6.3. Conclusion 286
7. Discussion and future work 291
8. Bibliography 300
Appendices 327

1. Introduction

1.1. Threats to genome integrity

The maintenance of genomic integrity is of critical importance for normal cellular processing. However, the genome is subjected to an onslaught of DNA damage events (estimated at 10 000 lesions per cell, per day) (Lindahl, 1993), and these can arise from both endogenous and environmental sources. These lesions must be repaired accurately to allow DNA and RNA polymerases to faithfully replicate, and express, genetic information. Unrepaired, or misrepaired, DNA damage can have negative cellular consequences and the accumulation of DNA lesions has been implicated in the aetiology of a number of different pathologies; including cancer (Jackson and Bartek, 2009), neurodegenerative disorders (Kulkarni and Wilson, 2008), and ageing (Schumacher et al., 2008). As such, the cell has evolved a number of different DNA damage repair responses and pathways to maintain the integrity of the genome. Common DNA damage lesions are represented in Figure 1.1.

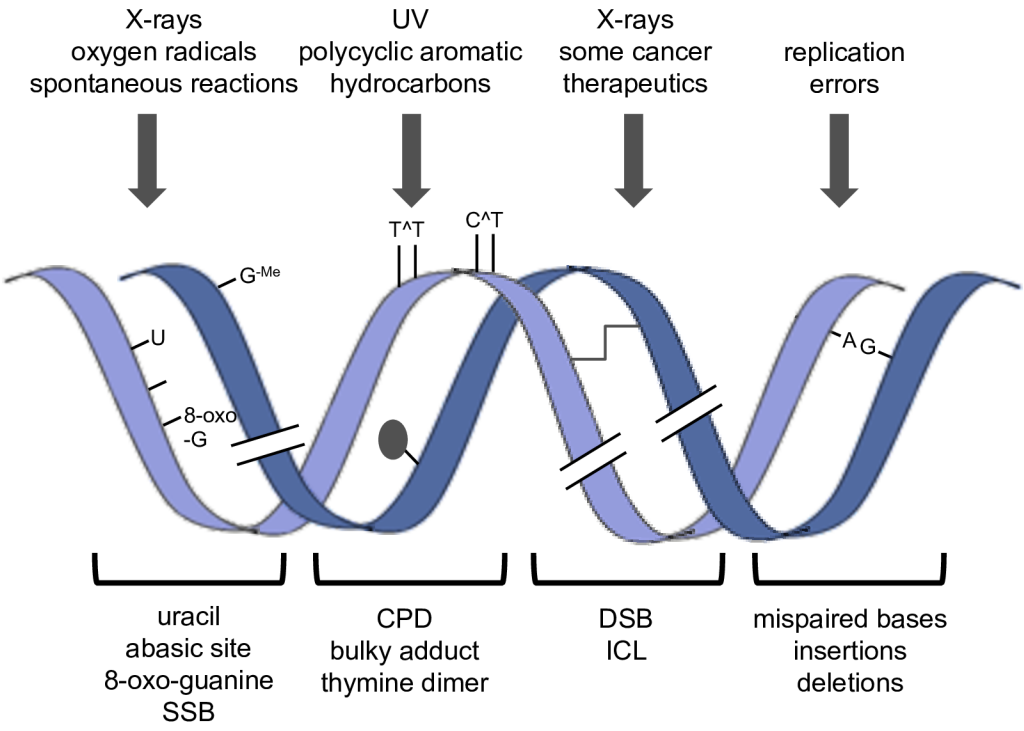


Figure 1.1 Schematic representing common DNA damaging agents and the lesions they induce.

ICL = interstrand crosslink. DSB = double strand break. SSB = single strand break. T^T = thymine dimer. G^{-Me} = O^6-methylguanine. T^C = cyclobutane pyrimidine dimer (CPD). The dark grey circle represents a bulky base adduct. Figure adapted from (Hoeijmakers, 2001).

1.1.1. Endogenous sources of DNA damage

There are a number of sources of endogenous DNA damage; however, the majority of these arise through interactions between chemically active DNA and either water or reactive oxygen species, via hydrolytic or oxidative mechanisms.

In an aqueous environment, the glycosidic bonds in DNA can be hydrolysed, leading to depurination or depyrimidination, although the former is ~500 times

more frequent than the latter (Lindahl and Karlstrom, 1973, Loeb and Preston, 1986). This results in the formation of abasic, or AP (apurinic/apyrimidinic) sites, which are inherently unstable and can chemically convert to single strand breaks (SSBs) during repair, via a β-elimination reaction that attacks the 3' phosphodiester bond 5' to the AP site (Bailly and Verly, 1988). Similarly, spontaneous base deamination is a common hydrolytic event, whereby bases lose their exocyclic amine. Deamination of cytosine and 5-methyl-cytosine is the most common, and can lead to the introduction of point mutations due to aberrant uracil base pairing (Lindahl, 1993, Yonekura et al., 2009). Reaction oxygen species (ROS) (primarily hydrogen peroxide, superoxide radicals, and the hydroxyl radical) are generated through normal metabolic processes, and can interact with DNA to induce myriad lesions; two of the most common being SSBs, and the formation of 8-oxo-guanine (8-oxo-G) (Henle and Linn, 1997, Kasai and Nishimura, 1984, Cadet and Wagner, 2014). Likewise, lipid peroxidation (the oxidative degradation of lipid molecules by hydroxyl radicals), results in the formation of aldehyde products, such as malondialdehyde or 4-hydroxynonenal, which can form mutagenic base adducts (Marnett, 1999, Plastaras et al., 2000, VanderVeen et al., 2001). Similarly, formaldehyde arises from a number of metabolic processes (including amino acid metabolism, lipid peroxidation, and P450-mediated demethylation), as well as being a common environmental contaminant (e.g. in cigarette smoke). Formaldehyde can induce a number of DNA damage lesions, primarily N-hydroxymethyl monoadducts on guanine, adenine, and cytosine, as well as intra- and interstrand crosslinks between purines, and DNA-protein crosslinks (Kawanishi et al., 2014).

Enzymes present within the cell, for example S-adenosylmethionine (SAM), which is involved in methyl transfer reactions, can also modify DNA. SAM spontaneously generates N3-methyladenine, N7-methylguanine, and O^6-methylguanine (Rydberg and Lindahl, 1982, Holliday and Ho, 1998). However, the cytotoxicity of these lesions varies – N3-methyladenine can inhibit DNA replication, whereas N7-methylguanine is essentially harmless, albeit in a small proportion of cases where it can react further with the DNA to generate an AP site (Loechler et al., 1984, Larson et al., 1985, Singer et al., 1989). Mutations can also arise as a result of erroneous physiological processes; for example, replication errors can result in single base substitutions, insertions, or deletions, as well as longer insertions or deletions from strand slippage at repetitive sequences (Kunkel, 2009, Viguera et al., 2001). Aberrant topoisomerase activity during replication and transcriptional processes can also result in the formation of protein-bound DNA strand breaks (Pommier et al., 2006).

1.1.2. Exogenous sources of DNA damage

DNA damage can arise from a number of exogenous sources; examples include damage resulting from exposure to ionising radiation (IR) or ultraviolet (UV) radiation. UV-B radiation from the sun can cause the formation of cyclobutane pyrimidine dimers (CPDs), and pyrimidine-pyrimidone (6-4) photoproducts ((6-4)-PPs) and, as bulky base adducts, these lesions pose a significant barrier to both replication and transcriptional processes (Mitchell and Nairn, 1989, Sinha and Hader, 2002). Ionising radiation (encompassing alpha, beta, and gamma radiation, as well as neutron radiation, and X-rays) is highly reactive and can damage DNA both directly and indirectly (Wardman, 2009, Vignard et al., 2013). Accordingly, IR

induces a wide variety of different lesions, including oxidative damage, SSBs, and double strand breaks (DSBs) (Henner et al., 1982, Obe et al., 1992, Hutchinson, 1985).

In addition to radiation, other exogenous sources of DNA damage include a range of carcinogenic environmental compounds. Aromatic amines (for example, from cigarette smoke, industrial dyes, and coal) are metabolised by the cytochrome P450 enzymes to carcinogenic alkylating agents, which can form C8-guanine adducts (Hammons et al., 1997, Eckel and Krugh, 1994). Similarly, environmental alkylating agents can react with DNA. Polycyclic aromatic hydrocarbons (PAHs) are common pollutants arising from incomplete combustion of fossil fuels and tobacco smoke. Initially inert, these compounds are metabolised to more reactive species, and can intercalate with DNA inducing a variety of DNA adducts (Yu, 2002). Additional alkylating environmental pollutants are the *N*-nitrosamine family of compounds. These arise from tobacco smoke, are present in preserved meat products, and can be both mutagenic and carcinogenic (Bartsch and Montesano, 1984, Hecht, 1999, Tricker and Preussmann, 1991).

There are also many more environmental carcinogenic compounds, including, but not limited to; aflatoxins (Bennett and Klich, 2003), synthetic oestrogen (Cavalieri et al., 2000), butylparaben (Oishi, 2002), and plant protection products (Kasiotis et al., 2012). A further significant source of DNA damage, linked to both exogenous and endogenous factors, is viral infection (Chen et al., 2014).

It is also important to note that a number of DNA damaging agents have been exploited for either diagnostic or therapeutic potential. For example, the

radioactive isotopes iodine-131 and technetium-99 are used in the diagnosis and therapeutic management of some thyroid diseases (Buscombe, 2007); and psoralen + UV-A is used in the treatment of psoriasis and other skin disorders (Situm et al., 2014). Furthermore, a number of DNA damaging agents are utilised as cancer chemotherapeutics, as the lesions they induce can be cytotoxic (Jackson and Bartek, 2009).

1.2. The DNA damage response

In response to cellular recognition of a DNA damage lesion, the DNA damage response (DDR) is activated. The DDR is a highly complex, coordinated, and spatiotemporally controlled process; lesions are specifically recognised, the appropriate factors are recruited, downstream effectors are activated, and the lesion is subsequently repaired. There are myriad different DNA lesions; however, the DDR is principally mediated by members of the phosphatidylinositol 3-kinase-like protein kinase (PIKKs) family (ATM, ATR, and DNA-PK), and by members of the poly(ADP-ribose) polymerase (PARP) family. PARP enzymes are primarily involved in the repair of SSBs (Harper and Elledge, 2007), and DNA-PK in the repair of DSBs via non-homologous end-joining (NHEJ) (Blackford and Jackson, 2017). However, ATM and ATR are the best characterised. Each are capable of phosphorylating hundreds of substrate proteins, mediating the global DDR, and initiating a signal cascade that results in checkpoint activation, lesion repair, alterations in gene expression, and apoptosis or senescence. Below is a very brief summary of the roles of ATM and ATR in the DDR (also represented schematically in Figure 1.2).

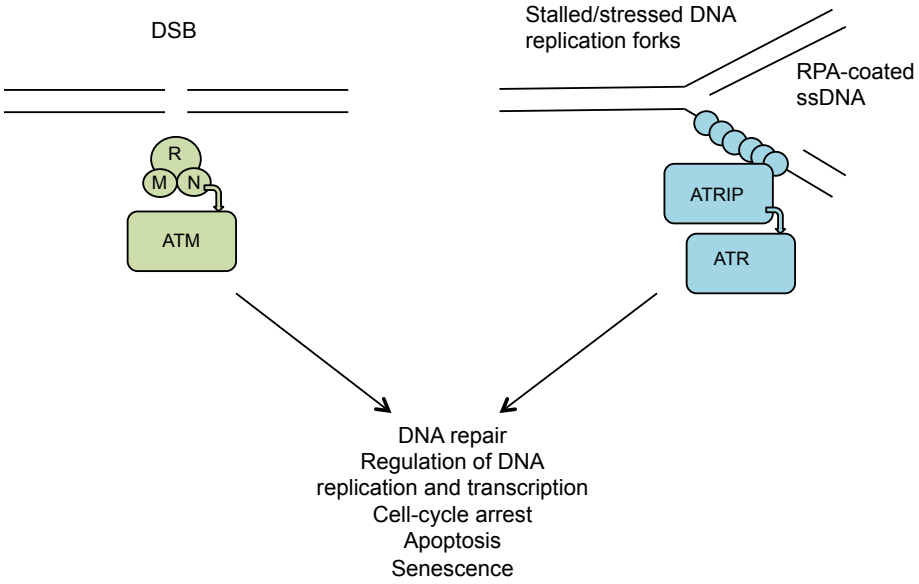


Figure 1.2 Schematic representing the recruitment and activation of DDR pathways involving ATM and ATR.

ATM is recruited to DSBs by the MRE11-RAD50-NBS1 (MRN) complex. ATR is recruited to regions of RPA-coated ssDNA by its binding partner ATRIP. ATM and ATR then phosphorylate downstream targets, and these effectors are important for regulating a number of cellular outcomes necessary for the maintenance of genomic integrity. Figure adapted from (Marechal and Zou, 2013) and (Blackford and Jackson, 2017).

1.2.1. The role of ATM in the DDR

ATM (ataxia telangiectasia mutated) is a serine/threonine kinase, and is recruited to chromatin in response to double strand breaks (DSBs), in a process mediated by the interaction of ATM binding to the C-terminal region of NBS1, a component of the MRE11-RAD50-NBS1 (MRN complex) (Andegeko et al., 2001, Bekker-Jensen et al., 2006). Once recruited to the site of damage, ATM dissociates from the dimeric to the monomeric form and becomes phosphorylated on multiple

residues. ATM is then able to phosphorylate any number of its hundreds of downstream substrates, and also activates second wave of phosphorylation through its primary effector kinase, CHK2 (Matsuoka et al., 1998, Liu et al., 2000). Activation of ATM, and subsequent phosphorylation of its downstream substrates facilitates a number of cellular outcomes including the activation and regulation of DNA damage repair pathways, inducing cell-cycle arrest, and apoptosis.

With regards to DNA damage repair, activation of ATM triggers a number of events including the activation of BRCA1, phosphorylation of the histone variant H2AX, and accumulation of other DNA repair proteins and chromatin remodelers at sites of damage (Savic et al., 2009, Stiff et al., 2004, Lukas et al., 2011). Regarding cell cycle arrest and apoptosis, the tumour suppressor p53 is phosphorylated on multiple sites by ATM, and thus stabilised, allowing it to activate a transcriptional pathway resulting in G_1 cell cycle arrest, apoptosis, or senescence (Shiloh and Ziv, 2013).

1.2.2. The role of ATR in the DDR

ATR (ataxia telangiectasia and Rad3 related) is a serine/threonine kinase, and is recruited to sites of DNA damage and stalled/stressed replication forks, via its binding partner ATRIP, by the presence of replication protein A (RPA) coated ssDNA (Zou and Elledge, 2003). The presence of ATR is essential for the continued survival of proliferating cells (de Klein et al., 2000, Brown and Baltimore, 2000), and this is likely due to its role in regulating replication fork progression, and responding to the intrinsic replication stresses that arise during S-phase. Unlike ATM, which is activated primarily in response to DSBs, ATR is activated by

a variety of genotoxic stressors, due to RPA-ssDNA tracts being generated as repair intermediates in the resolution of an array of lesions (Raderschall et al., 1999), as well as the aforementioned role of ATR in regulating replication fork progression, even under unperturbed conditions.

For optimal ATR activation, additional activator proteins are required, and many of these are found at the junctions between ss- and dsDNA regions. The best characterised of these proteins is TopBP1 which possesses an ATR-activation domain that stimulates the kinase activity of ATR (Kumagai et al., 2006).

Following activation, ATR is able to phosphorylate its downstream targets, including its effector kinase, CHK1 (Liu et al., 2000, Guo et al., 2000, Zhao and Piwnica-Worms, 2001). Subsequent to phosphorylation, CHK1 promotes the degradation of CDC25A, thus reducing the activity of the cyclin dependent kinases (CDKs). This has the effect of inhibiting progression through the cell cycle, due to G_2/M cell cycle arrest (Bartek et al., 2004). ATR activation can also result in p53 phosphorylation, thus contributing to G_1/S cell cycle arrest (Tibbetts et al., 1999). This allows time for repair of the DNA damage lesion, or, if the damage is too great, for the cell to undergo apoptosis or senescence.

Although ATM and ATR are activated in response to structurally diverse DNA damage lesions, and each has their own unique substrates, there is also a significant amount of functional redundancy between the two. For example, DSBs result in the initial activation of ATM; however, subsequent to the resection requisite for HR, there is activation of ATR, primarily due to the generation of RPA coated ssDNA (Adams et al., 2006, Cuadrado et al., 2006, Jazayeri et al., 2006).

This activation of ATR is essential for mediating the intra-S and G_2/M checkpoint responses. Similarly, ATR is also able to phosphorylate H2AX in response to replication stress, thus recruiting ATM to replication associated DNA lesions (Ward and Chen, 2001). There is also some evidence that ATM and ATR may be able to directly phosphorylate one another (Stiff et al., 2006); and they certainly phosphorylate many of the same substrates, thus effecting similar cellular outcomes (Matsuoka et al., 2007).

However, ATM and ATR can each phosphorylate unique substrates. One example is the role of ATR in the Fanconi Anaemia (FA) mediated repair of interstrand crosslinks (ICLs), where ATR is able to promote the monoubiquitination of FANCD2, an essential step in FA mediated ICL repair (Andreassen et al., 2004).

1.3. DNA damage repair pathways

To maintain genome integrity, a number of cellular pathways have evolved to specifically repair DNA lesions. However, the simplest method of repair is by chemical reversal of the lesion. Only three mechanisms for direct reversal have been documented to date and they are utilised in the specific reversal of UV photolesions (by photolyases) and alkylated bases. For example, O^6-alkylguanine-DNA-alkyltransferase (AGT/MGMT) can reverse *O*-alkylated DNA bases, such as, O^6-methyl, ethyl, and benzyl groups, as well as the pyridyloxobutyl guanine adducts. Each of these can be removed in a one-step reaction as an AGT/MGMT molecule transfers the alkyl group from the oxygen of the DNA base to an internal cysteine residue present in the catalytic core (Kaina et al., 2007, Tubbs et al., 2007). Similarly, the AlkB family dioxygenases can reverse *N*-alkylated base

adducts. To date, nine AlkB family demethylases have been identified in human cells, but only a subset function as DNA/RNA repair enzymes. For oxidative demethylation, AlkB family members use an iron(II) site to hydroxylate the aberrant alkyl group, which is then released as formaldehyde (Drablos et al., 2004, Falnes et al., 2007).

However, other DNA lesions are not so simply repaired. Thus, the following pathways exist for the repair of more complex lesions: base excision repair (BER); nucleotide excision repair (NER); mismatch repair (MMR); single strand break (SSB) repair; and double strand break repair, which consists primarily of homologous recombination (HR), and non-homologous end-joining (NHEJ); as well as interstrand crosslink (ICL) repair. These repair pathways are initiated after cellular sensing and signalling as part of the DDR. If the DNA damage is not repaired appropriately this can lead to either cell death or senescence, an observation that has been exploited by both cancer radiotherapy and chemotherapy (Jackson and Bartek, 2009). It is also important to note that the aforementioned pathways are not always clearly delineated, and there is a significant degree of overlap and redundancy.

Of the different DNA damage repair pathways, only double strand break repair and interstrand crosslink repair are pertinent for this study. Thus, other pathways will only be mentioned very briefly below.

Base excision repair (BER) is the process whereby non-helix-distorting, chemically modified DNA bases are removed. As noted in 1.1.1, DNA bases can undergo spontaneous depurination or deamination; or they can be chemically modified by a

variety of mechanisms, including oxidation, or alkylation. Thus, BER functions at all stages of the cell cycle to remove lesions such as 8-oxo-G, 7-methylguanosine, or hypoxanthine (Dianov and Hubscher, 2013).

Nucleotide excision repair (NER) is the pathway for the removal of a variety of lesions, particularly those that are bulky and helix distorting, such as CPDs, and (6-4)-PPs from exposure to UV radiation, as well as M1G from lipid peroxidation. The severity of NER deficient disorders – xeroderma pigmentosum (XP), Cockayne syndrome (CS), and trichothiodystrophy (TTD) highlight the importance of this pathway (Vermeulen and Fousteri, 2013). It is important to note that both BER and NER occur at all stages of the cell cycle (Dianov and Hubscher, 2013, Fagbemi et al., 2011).

Conversely, mismatch repair (MMR) is active only in the S/G_2 stages of the cell cycle, and is a post-replicative repair pathway that is highly important for replication fidelity (Kunkel, 2009). MMR facilitates the removal and repair of mismatched bases, as well as insertions or deletions that have arisen due to replicative processes (Jiricny, 2006).

Translesion synthesis (TLS) is an error-prone repair pathway whereby a DNA lesion is bypassed to prevent replication fork stalling or collapse. This requires the activity of the TLS polymerases, a group of specialised polymerases that are able to progress past lesions (Sale, 2013). However, this activity is achieved with relatively lower fidelity than the replicative DNA polymerases, in part because most TLS polymerases lack the proof reading ability of the replicative polymerases (McCulloch and Kunkel, 2008) and, due to the more open nature of their active

sites, they are less able to discriminate between bases, and thus are more likely incorporate an incorrect nucleotide (Sale, 2013). It is also important to note that TLS plays an important role in the repair of ICLs, and this will be discussed further below.

Single strand breaks (SSBs) are one of the most common DNA lesions and often arise from oxidative DNA damage, from abasic sites, or from the erroneous activity of enzymes, such as topoisomerase I (TOP1) (Caldecott, 2008). Due to the varied nature of SSBs that arise, there are subtypes of SSB repair that exist, although most are repaired by a global SSB repair process which is initiated by the recognition of the lesions by PARP1 (D'Amours et al., 1999). As the lesion is processed, subsequent repair steps can be classified as either short patch, or long patch repair. SSBs may also be repaired via a transcription-coupled or replication-coupled process (Caldecott, 2008).

1.4. Double strand break (DSB) repair

Double strand breaks (DSBs) are one of the most cytotoxic DNA lesions. Repair of DSBs is uniquely challenging, as it is inherently a bimolecular process, and inappropriate or inaccurate re-joining can lead to the introduction of mutations, or to chromosomal rearrangements.

DSBs can arise from a number of exogenous and endogenous sources. Regarding endogenous sources, DSBs can form via reactions with ROS, or as a result of fork collapse during DNA replication when the replicative machinery encounter unrepaired DNA lesions, or as intermediates in other DNA repair

pathways (Pfeiffer et al., 2000). DSBs also arise during specific, programmed cellular events; they are generated during meiotic recombination in gamete formation (Keeney and Neale, 2006), and during V(D)J recombination as part of antigen receptor diversification (Dudley et al., 2005). Regarding exogenous sources, DSBs are generated after exposure to UV radiation, IR, or some cancer chemotherapeutics. The timely and accurate repair of DSBs is critical because, if left unresolved, they are implicated in a number of different human disorders and cancers (Jackson and Bartek, 2009).

In normal cells, DSBs are repaired by one of two major pathways, homologous recombination (HR), which functions primarily during S and G_2-phases of the cell cycle; or non-homologous end-joining (NHEJ), which is able to function at all cell stages. Which repair pathway is utilised is determined by a number of factors, including cell-cycle stage, the nature of the DSB, and the availability of the DNA ends. HR is restricted to S/G_2 phases of the cell cycle, and this temporal constraint is achieved, in part, by CDK mediated activation of the HR machinery (Aylon et al., 2004). For example, CtIP, which is necessary factor in controlling DSB end resection to generate the ssDNA overhangs requisite for HR, is activated by phosphorylation in a CDK dependent manner (You and Bailis, 2010, Huertas and Jackson, 2009). The nature of the DSB is also important for determining pathway choice, with one-ended DSBs being preferentially repaired via HR, due to the lack of an appropriate DNA end for ligation by NHEJ (Arnaudeau et al., 2001). Further contributing to repair pathway choice, is the availability of the DNA ends. This is regulated by the antagonistic relationship between 53BP1 and BRCA1. 53BP1 has a number of roles in regulating DSB repair (Panier and Boulton, 2014); however, one important role is to promote NHEJ by restricting the DSB end-resection

associated with HR (Bunting et al., 2010, Bothmer et al., 2010). However, in S/G_2 phase, BRCA1 is able to counteract the activity of 53BP1 by promoting 53BP1 dephosphorylation, thus promoting HR, and suppressing NHEJ (Zhang et al., 2016, Isono et al., 2017).

1.4.1. Non-homologous end-joining

Non-homologous end-joining (NHEJ) is a pathway for repair, whereby the two ends of the DSB are directly ligated to one another. Unlike HR, it requires little (less than ~10 bp) to no homology between the two ends to guide repair. Within NHEJ there are two subtypes – 'classical NHEJ' (c-NHEJ), characterised by the dependence on Ku, DNA ligase 4, and other associated factors; or 'alternative NHEJ' (alt-NHEJ) which is not reliant on the function of these proteins (Chiruvella et al., 2013). NHEJ is responsible for repairing the majority of two-ended DSBs, and it is a critical part of the aforementioned V(D)J recombination process. Accordingly, deficiencies in NHEJ factors, such as Artemis and ligase 4 (LIG4), cause a severe immunodeficiency phenotype (O'Driscoll et al., 2004).

The NHEJ pathway for repair is initiated by DSB formation, which is recognised by the ring-shaped Ku (Ku70 and Ku80) heterodimer (Pang et al., 1997). Ku70 and Ku80 are constituent subunits of the DNA-dependent protein kinase (DNA-PK), along with the catalytic subunit (DNA-PKcs). Binding of Ku70/80 to the ends of the DSB results in recruitment of DNA-PKcs and activates its kinase activity, resulting in phosphorylation of many of the downstream substrates necessary for NHEJ repair (Dvir et al., 1992, Gottlieb and Jackson, 1993). The Ku70/80 heterodimer is also highly important for stabilising the DNA ends, preventing end resection, and

keeping them in proximity to one another (Mimitou and Symington, 2010, Mari et al., 2006, Chiruvella et al., 2013). Subsequent to activation of the DNA-PK complex, other NHEJ repair factors are recruited including, XRCC4 (X-ray cross complementation group 4), LIG4, XLF (XRCC4-like factor), APLF (aprataxin-PNKP-like factor), Artemis, and TdT (terminal deoxynucleotidyl transferase) (Nick McElhinny et al., 2000, Costantini et al., 2007, Yano et al., 2008, Grundy et al., 2013). The order in which these components are recruited, and also which specific factors are required, depends on the nature of the lesion (Yano and Chen, 2008, Mari et al., 2006). The recruitment of these factors translocates the Ku70/80 heterodimer away from the DSB, and other factors, including XRCC4 and XLF, participate in the proximal tethering and bridging of the two free DSB ends (Yoo and Dynan, 1999, Malivert et al., 2010, Hammel et al., 2011, Andres et al., 2012).

ROS and IR can induce a variety of chemically modified ends, and small ssDNA overhangs are also common. Therefore, prior to ligation the two DNA strands must be processed to generate the requisite 5' phosphate and 3' hydroxyl ends. The enzyme(s) involved in this process depend on the nature of the DSB. The enzyme classically involved is Artemis (also known as SNM1C), which possesses structure-specific endonuclease activity (Povirk et al., 2007, Kurosawa and Adachi, 2010). Artemis is essential for repairing IR induced DSBs, presumably by cleaving chemically modified damaged ends (Woodbine et al., 2011). Other end processing enzymes include; polynucleotide kinase 3' phosphatase (PNKP), which has a dual function in phosphorylating 5' ends, and dephosphorylating 3' ends (Coquelle et al., 2011, Garces et al., 2011), and APLF, which possesses endonuclease and 3' exonuclease activity (Li et al., 2011). Similarly, aprataxin (APTX) is able to remove AMP groups from the 5' terminus that abort DNA ligation (Rass et al., 2008, Gong

et al., 2011, Tumbale et al., 2011). The dual exonuclease and endonuclease, MRE11 has also been implicated in end processing (Xie et al., 2009), as has the exonucleolytic activity of the Werner protein (WRN) (Cooper et al., 2000, Kusumoto et al., 2008). Interestingly, Ku70 has also been shown to possess intrinsic lyase activity, thus allowing it to cleave abasic sites and also contribute to this process (Roberts et al., 2010). Topoisomerases can also become adducted to DSB ends, when replication forks stall and subsequently collapse. These can be cleaved by the tyrosyl-DNA phosphodiesterases, TDP1 and TDP2 (Bahmed et al., 2010, Cortes Ledesma et al., 2009).

Once end-processing has taken place, any gaps that have been introduced are subsequently filled in by the family X polymerases in either a template-dependent (Pol μ), or template-independent manner (Pol λ) (Ramadan et al., 2004). The two DNA ends are then ligated together by the XRCC4-LIG4 complex (Grawunder et al., 1997). This process is illustrated schematically in Figure 1.3.

However, in the absence of classical-NHEJ, alternative-NHEJ is activated (Figure 1.3). C-NHEJ is relatively accurate and error-free, whereas alt-NHEJ is a highly error prone pathway that often introduces larger deletions and translocations (Betermier et al., 2014). The specific steps that constitute the alt-NHEJ pathway are also less well elucidated than for c-NHEJ. It is also not clear whether there is more than one alt-NHEJ pathway, or simply multiple overlapping mechanisms. However, the pathway that is discussed briefly below is one form of alt-NHEJ, often described as microhomology-mediated end joining.

Alt-NHEJ differs from c-NHEJ, as it is independent of DNA-PK and is also suppressed by Ku (Audebert et al., 2004, Wang et al., 2006). Instead, alt-NHEJ is initiated by the presence of poly(ADP-ribose) polymerase 1 (PARP1) at DSB sites. This interaction leads to the activation of PARP1, which is then able to modify itself and downstream factors with the addition of branched polymers of ADP-ribose (Audebert et al., 2004, Ali et al., 2012). Alt-NHEJ is characterised by the generation of microhomologous regions to guide repair. In some cases this arises due to local template extensions of the 3' strand of the DSB (Yu and McVey, 2010). However, more commonly these regions are generated by limited nucleolytic resection. Nucleases implicated in this process include MRE11, EXO1, and FEN1 (Sharma et al., 2015). Following resection, these short ssDNA overhangs are annealed, guided through sequence microhomology, and any mispaired bases are corrected by MMR. This process results in the formation of 3' heterologous flaps and these are cleaved by structure-specific endonucleases, such as XPF/ERCC1 (Ahmad et al., 2008). Any missing nucleotides are subsequently filled by DNA polymerases: polymerase θ has been shown to be necessary for this process (Kent et al., 2015); although the DNA polymerases β, μ or λ have also been shown to participate in gap-filling at the DNA ends, albeit with varying degrees of efficacy (Crespan et al., 2012). Ligation between the two strands is subsequently catalysed by either XRCC1/LIG3 (Audebert et al., 2004) or LIG1 (Liang et al., 2008). Also implicated in alt-NHEJ, although perhaps not specifically to microhomology-mediated end-joining, is WRN (Shamanna et al., 2016, Chan et al., 2010). However, more work is required to fully elucidate and further understand these alternative NHEJ pathways.

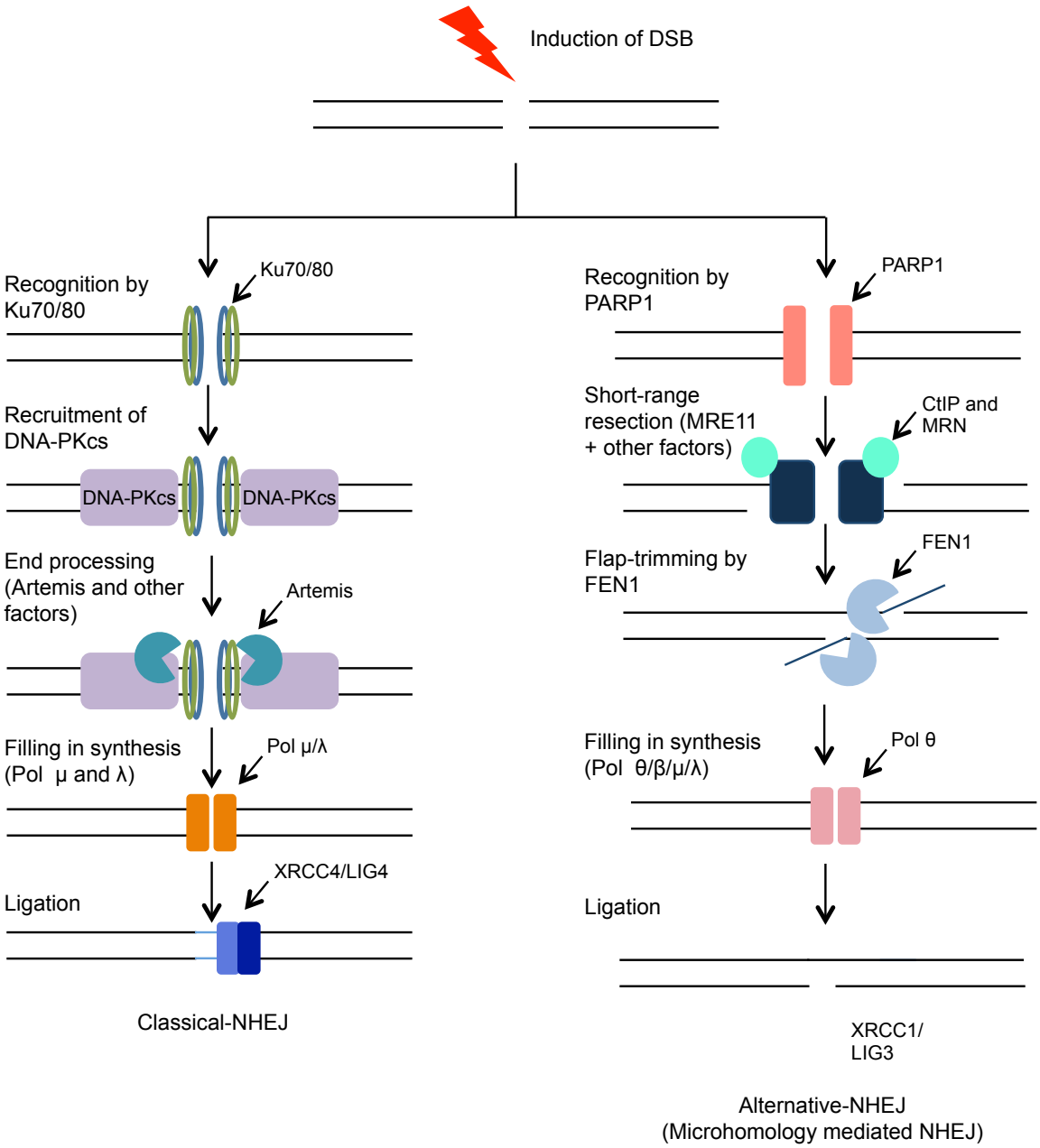


Figure 1.3 Schematic representing the pathways of classical and alternative-NHEJ.

The pathway represented for alternative-NHEJ can also be considered microhomology mediated-NHEJ. Constituent protein factors are as labelled and the blue line represents DNA that may be newly synthesised due to 'filling in' by repair polymerases.

1.4.2. Homologous recombination

Homologous recombination (HR) is an important mechanism for the high-fidelity resolution of DSBs that utilises strand invasion and template-directed repair. Due to this reliance on a homologous chromosome, this pathway is primarily active during G_2 and S-phases of the cell cycle. The importance of a functional HR pathway for genome integrity is highlighted by increased tumourigenesis in HR defective individuals. Examples of this include: Bloom's syndrome, where mutations in BLM result in early onset of cancer (amongst other phenotypic characteristics) (Chu and Hickson, 2009); and increased breast and ovarian cancer risk in individuals who possess mutations in the BRCA1 and BRCA2 genes (Jasin, 2002). HR also plays an important cellular role in the resolution of programmed DSBs, such as those induced during meiotic recombination by SPO11 (Keeney, 2008).

It would perhaps be more appropriate to consider HR as consisting of a number of sub-pathways – the initial resection steps are shared with the alt-NHEJ pathway discussed above, and more extensive resection is common with the molecular process of single strand annealing (Bhargava et al., 2016). Later steps involving strand invasion, can be resolved via either synthesis-dependent strand annealing (SDSA), or the more canonical double strand break repair (DSBR) route (Li and Heyer, 2008). BIR (break induced replication) is another important HR mediated mechanism for the repair of one-ended DSBs, and this pathway is also the basis for the alternative lengthening of telomeres (ALT) observed in some cancer cells (Dilley et al., 2016, Sakofsky and Malkova, 2017). However, all share the initial recognition and processing steps.

HR mediated repair of a DSB is initiated when the MRN complex (MRE11-RAD50-NBS1) recognises the lesion and binds to the DNA. MRN then recruits ATM and TIP60 (Sun et al., 2010, Stracker and Petrini, 2011). Upon recruitment, TIP60 acetylates histone H4, resulting in chromatin modification, and also triggers the activation of ATM. (Sun et al., 2005). ATM then phosphorylates itself and downstream substrates, including H2AX (resulting in the formation of phospho-H2AX or γH2AX) (Bhatti et al., 2011). This promotes the recruitment of MDC1 (mediator of DNA damage checkpoint protein 1), which is, in turn, phosphorylated by ATM (Stucki et al., 2005). Phosphorylated MDC1 subsequently acts a molecular scaffold for the ubiquitin E3 ligases, RNF8 and RNF168 (Altmeyer and Lukas, 2013). RNF8 and RNF168 then ubiquitinate γH2AX (amongst other targets), which is able to bind both 53BP1 and BRCA1; however, in S/G_2 phase, BRCA1 plays an important role in counteracting 53BP1, thus stimulating HR (Zhang et al., 2016, Isono et al., 2017). BRCA1 also ubiquitinates its binding partner CtIP (C-terminal binding protein interacting protein) enhancing its interaction with DNA (Yu et al., 2006). Initial resection then occurs in the 5' to 3' direction, mediated by the nuclease activity of MRE11 and dependent on CtIP (Mimitou and Symington, 2011, Sartori et al., 2007). This initial resection generates short ssDNA 3' overhangs and appears to be similar between the alt-NHEJ pathway, described previously, and HR. As MRN possesses 3' to 5' exonuclease activity, this polarity is discordant with the observation that resection occurs in the 5' to 3' direction. However, experimental data suggest that a bidirectional resection model may be more accurate. In this schema, MRN makes a series of endonucleolytic incisions proximal to the DSB, and the 5' to 3' exonucleolytic activity of MRE11 then acts to digest back towards the DSB end.

These endonucleolytic incisions also provide entry points for the long-range exonucleolytic resection necessary to generate the 3' ssDNA overhangs necessary for subsequent strand invasion (Garcia et al., 2011, Cannavo and Cejka, 2014).

Experimental data suggest that long range resection is dependent on either the 5' to 3' exonucleolytic activity of EXO1 or the DNA2 5' to 3' helicase/nuclease, in conjunction with the BLM (Bloom's syndrome protein) helicase (Nimonkar et al., 2011). Interestingly, the activity of EXO1 is also stimulated by BLM, even though the two pathways appear to function independently (Nimonkar et al., 2008). The nuclease activity of EXO1 is also stimulated by the presence of RPA (Cejka et al., 2010).

As nucleolytic resection occurs, either by EXO1/BLM or DNA2/BLM, this generates tracts of ssDNA that are quickly coated by the heterotrimeric protein, RPA. This triggers the recruitment and activation of ATR and subsequent phosphorylation of downstream substrates, including CHK1, thus triggering the intra-S and G_2/M checkpoint responses (Bartek et al., 2004).

At this point, it is possible for single stranded annealing (SSA) to occur, thus diverging from the canonical homology directed repair (HDR) pathway (Figure 1.4). The exposed ssDNA can anneal in regions of homology, mediated by RAD51, generating a synapsed intermediate (Rothenberg et al., 2008, Symington, 2002). The non-homologous 3' ssDNA tails are removed by endonucleolytic activity of XPF-ERRC1, any gaps are filled by polymerase activity, and the DNA ends are subsequently ligated (Motycka et al., 2004, Bhargava et al., 2016).

However, this can be a particularly error prone pathway, due to the insertions and deletions that can occur due to inaccurate annealing, excessive resection, or inappropriate gap-filling.

With regards to canonical HDR, the RPA bound to the ssDNA is displaced by RAD51 in a process mediated by BRCA2 and PALB2, with the assistance of RAD51B, RAD51C, and RAD51D (Zhang et al., 2009, Nickoloff, 2017). The RAD51-DNA nucleoprotein filament is able to perform a homology search and invade the appropriate duplex DNA (either a sister chromatid or homologous chromosome), forming a displacement-loop (D-loop) (Jasin and Rothstein, 2013). This process is aided by the ATPase dependent activity of RAD54, which stabilises the nucleoprotein filament and stimulates strand invasion (Mazin et al., 2010). The 3' OH group of the invading strand is able to prime synthesis by DNA polymerases. A number of polymerases have been implicated in this process, both replicative (Pol δ) and translesional (Pol η, κ, ζ) and the presence of PCNA (proliferating cell nuclear antigen) enhances their processivity (Li et al., 2013, Sebesta et al., 2013, Sharma et al., 2012, Kawamoto et al., 2005).

This extended strand invasion intermediate can be resolved in a number of ways, generating either a crossover (CO), or non-crossover (NCO) product. Most simply, it is possible for the newly synthesised strand to dissociate from the D-loop and anneal to the original DNA strand, in a process termed synthesis-dependent strand annealing (SDSA) (Jasin and Rothstein, 2013). This process is dependent on the helicase activities of RTEL1 and BLM, results in NCO repair, and is a primary repair pathway in mitotic cells (Uringa et al., 2011, Jasin and Rothstein, 2013, Adams et al., 2003).

Conversely, the second DNA strand can become captured and form a D-loop, thus generating a double Holliday-junction (dHJ). This can be resolved by DSBR to generate either a CO or NCO product (Figure 1.4), although this process appears to be biased towards the formation of CO products (Allers and Lichten, 2001). At this point, it is also possible for Holliday junction dissolution to occur, mediated by the branch migration and topoisomerase activity of the BLM/TOP3α/RMI1 complex, resulting in a NCO product (Figure 1.4) (Mimitou and Symington, 2009). In mitotic cells, the presence of BLM encourages dHJ dissolution, rather than DSBR dHJ resolution, which is more common in meiotic cells. Accordingly, in Bloom's syndrome, due to hypomorphic or loss of function mutations in BLM, dHJs must be resolved via DSBR, resulting in an increased number of CO products, thus, the phenotypic characteristics of increased sister chromatid exchanges (SCEs) and loss of heterozygosity (LOH) observed in Bloom's syndrome (Wu and Hickson, 2003). dHJ resolution can be catalysed by the activity of a number of structure-specific endonucleases, including MUS81/EME1, GEN1, and SLX1 and SLX4 (Svendsen and Harper, 2010, Ho et al., 2010). Which endonucleases are utilised in this process may be impacted by cell cycle stage, and each resolve the dHJ slightly differently (Matos et al., 2011).

In BIR, the initial invading strand is hypothesised to establish a replication fork that copies the remainder of the template chromosome. This is the primary repair pathway for one ended double strand breaks and the molecular basis for ALT, however, it does commonly result in significant loss of genetic information (Li and Heyer, 2008).

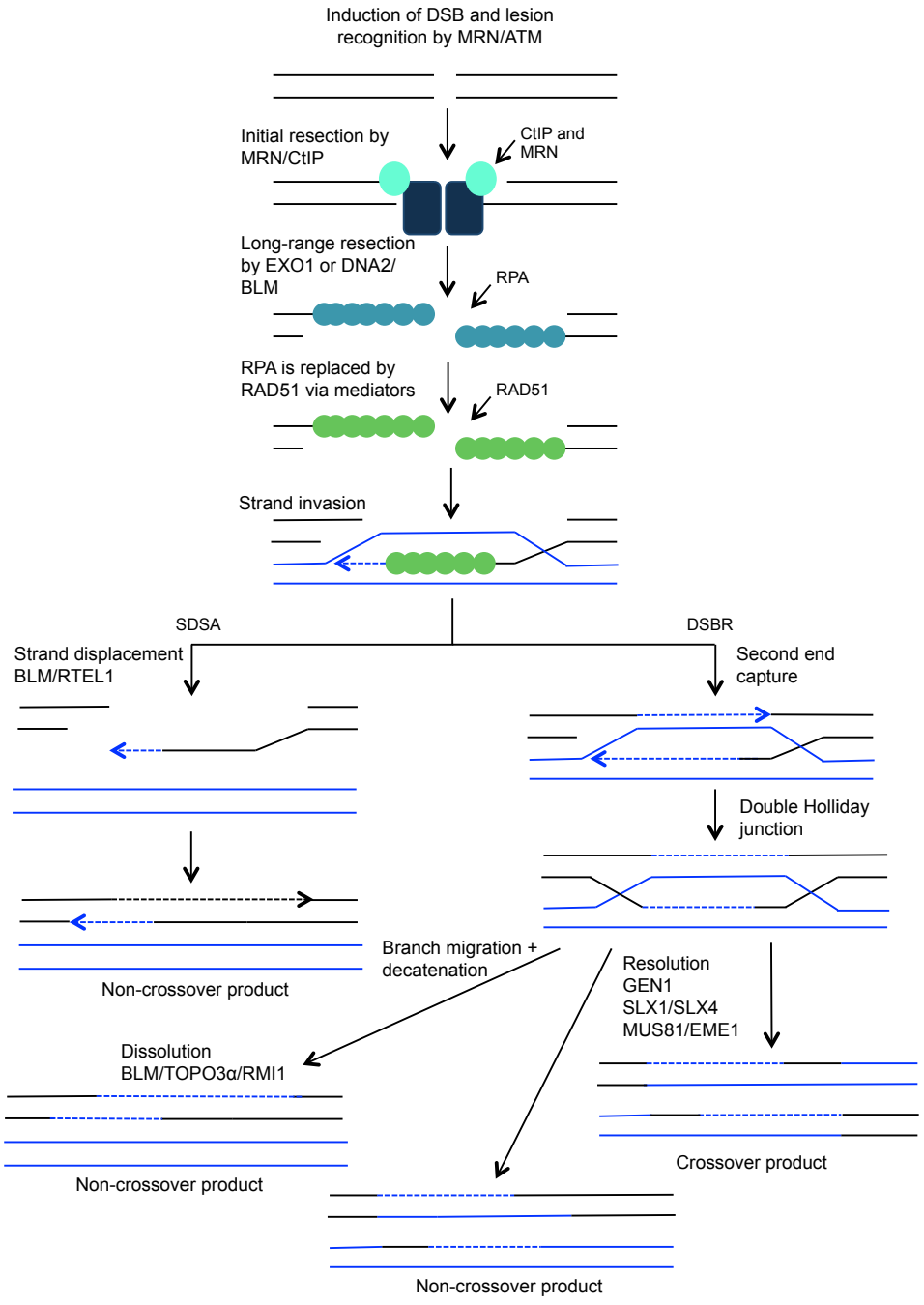


Figure 1.4 Repair of DNA double strand breaks by DBSR and SDSA.

DSBs can be repaired by several homologous recombination (HR)-mediated pathways, including double strand break repair (DSBR) and synthesis-dependent strand annealing (SDSA). Pathway steps are as indicated and the dashed lines represent newly synthesised DNA.

1.5. Interstrand crosslink repair

Interstrand crosslinks (ICLs) are a particularly cytotoxic form of DNA damage that covalently tether the two strands of the DNA, thus inhibiting the critical cellular processes of transcription and replication. This inhibition of transcription/replication is likely the major cause of cytotoxicity and, 1–2 unrepaired ICLs are sufficient to kill a yeast cell, whereas 2–4 unrepaired ICLs are lethal in mammalian cells (Magana-Schwencke et al., 1982, Clauson et al., 2013). This cytotoxicity has been exploited in anticancer therapy, where ICL inducing agents (including mitomycin C, cisplatin, and the nitrogen mustards) are used as chemotherapeutic agents.

The pathways that contribute to ICL repair are still not clearly defined. There have been many studies of ICL repair in bacteria and yeast, where Cole's model generally applies. This model was posited in the 1970s after experiments in *E. coli*, and relies on the functional collaboration of NER and HR factors in ICL resolution (Cole, 1973). Briefly, this pathway constitutes the following: recognition of the lesion by NER factors; endonucleolytic incision either side of the crosslink resulting in 'unhooking;' exonucleolytic gap generation; and gap repair either by homologous recombination, or by lesion bypass synthesis. This process converts the ICL to a monoadduct, which can subsequently be processed by conventional NER pathways (Cole, 1973). Many of these steps and processes are conserved in mammalian cells; however, the presence of other factors and sub-pathways makes mammalian ICL repair significantly more complex.

Unrepaired, ICLs can be mutagenic, clastogenic, and lead to chromosome missegregation, and mitotic catastrophe; thus, their repair is extremely important.

Accordingly, there is a strong link between ICL repair defects and genomic instability, and this is exemplified by the inherited disorder Fanconi anaemia (FA). FA is a rare genetic disorder characterised by developmental abnormalities, bone marrow failure, and an increased incidence of cancer (Ceccaldi et al., 2016). FA arises from germline mutations in any one of the 21 identified FA genes (FANCA to FANCV) (Chatterjee and Walker, 2017), and the defining cellular feature of FA is sensitivity to ICL inducing agents, thus this can be used as a diagnostic criterion (Auerbach, 2015, D'Andrea, 2010). Furthermore, large-scale genetic studies have identified a number of somatic hypomorphic/loss of function mutations in FA genes in breast, ovarian, and pancreatic tumour cells (Ceccaldi et al., 2016). The 21 FA genes encode proteins that, together with other repair factors, orchestrate the detection and repair of ICLs, as well as activating the global DNA damage response.

1.5.1. Endogenous ICL inducing agents

Most research into ICL repair mechanism has utilised exogenous ICL inducing agents. However, it seems unlikely that an ICL repair pathway that is (at least partly) conserved from bacteria to higher eukaryotes, evolved to deal exclusively with (primarily chemotherapeutic) exogenous agents (Clauson et al., 2013). However, understanding cellular ICL sources has proved complex and most research to date has utilised *in vitro* systems, or assessment of chemical mutagenicity (Pang and Andreassen, 2009).

Aldehydes, such as acrolein, malondialdehyde, and acetaldehyde, generated from lipid peroxidation and alcohol metabolism, have been indicated as a potential

source of endogenous ICLs. In accordance with this, studies have shown that ICL repair (or FA) defective mice are sensitive to aldehyde treatment (Langevin et al., 2011). Furthermore, FA deficient mice and people, which also lack the aldehyde metabolising enzyme, ALDH2, exhibit a more severe phenotype than those with FA alone (Garaycoechea et al., 2012, Hira et al., 2013). Cells deficient in BRCA1, BRCA2, and RAD51 (factors in both HR and ICL repair) also exhibit cell cycle arrest and fail to proliferate in response to ALDH2 inhibition (Tacconi et al., 2017). These data suggest that FA proteins mediate the repair of acetaldehyde-induced DNA lesions and that, if left unrepaired, these lesions can contribute to genomic instability. *In vivo,* acetaldehyde is highly reactive and can induce a variety of lesions. The most common DNA adduct from reaction with acetaldehyde is N2-ethyl-2'-guanosine, but this cannot form an ICL. However, it is possible for two acetaldehyde molecules to react with guanine to form 1-N2-propano-2'-deoxyguanosine, and this can exist in the open chain or cyclic aldehyde form. The cyclic configuration is mutagenic, whereas in the open chain configuration the free aldehyde group can react to form secondary lesions, resulting in both ICLs and DNA protein crosslinks (Brooks and Theruvathu, 2005).

As mentioned previously (1.1.1), the metabolic process of lipid peroxidation generates a variety of reactive α,β-unsaturated aldehydes. Examples include malondialdehyde and acrolein, both of which can also react to form ICLs between guanine residues in a manner that does not significantly disrupt the double helix (Stone et al., 2008, Kozekov et al., 2003).

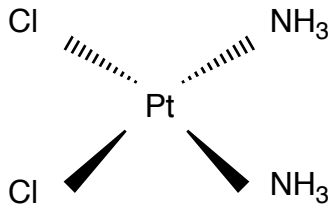
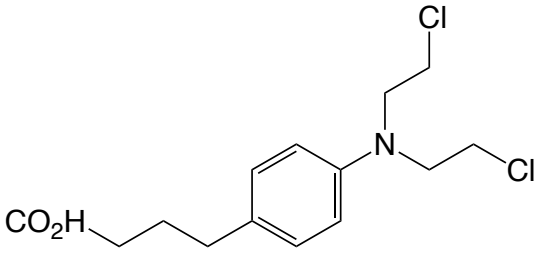
It is also possible for ICLs to arise from abasic (AP) sites. If the deoxyribose sugar of the AP site ring opens to its aldehyde form, it can react with the amino group on

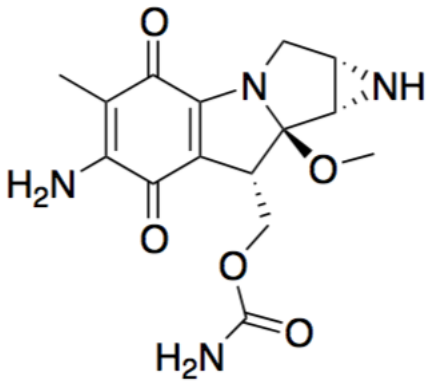
an opposing strand guanine residue, thus generating an ICL (Dutta et al., 2007). Many AP sites undergo spontaneous strand cleavage, and this can leave an α,β-unsaturated aldehyde residue within the nicked duplex. This can further react with an opposing adenine residue, producing an ICL flanked by a single strand break (Yang et al., 2017b).

1.5.2. Exogenous ICL inducing agents

There are a number of exogenous ICL inducing agents, some of which are synthetic and often used as chemotherapeutic agents, and some of which occur naturally, for example, psoralen + UVA. The chemical structures of some common ICL inducing agents are displayed in Table 1.1. All possess bifunctional reactive groups, requisite to interact with both strands of the DNA helix, and most react with opposing guanine bases in the DNA duplex to induce ICLs.

Table 1.1 Therapeutically relevant ICL inducing agents, their clinical applications, and example chemical structures. Adapted from (Deans and West, 2011).

Drug	Clinical Application	Example of Chemical Structure
Platinum-based		
Cisplatin	Primarily testicular, ovarian, and non-small cell (NSC) lung cancer	 Cisplatin
Carboplatin	Primarily ovarian cancer	
Oxaliplatin	Primarily colorectal cancer	
Satraplatin	Prostate and breast cancer	
Nitrogen mustards		
Cyclophosphamide	Lymphoma	 Chlorambucil
Melphalan	Multiple myeloma, melanoma, and ovarian cancer	
Chlorambucil	Chronic lymphocytic leukaemia	
Ifosfamide	NSC lung cancer and sarcomas	

Miscellaneous		
Mitomycin C	Oesophageal and bladder cancer	 Mitomycin C
Psoralen + UVA radiation	Cutaneous T-cell lymphoma and psoriasis	
Pyrrolobenzodiazepines	Phase II clinical trials for acute myeloid leukaemia and ovarian cancer	

The specific reaction mechanisms and lesions induced by common ICL agents will be discussed briefly below.

Nitrogen Mustards

Nitrogen mustards can be both monofunctional and bifunctional alkylating agents; however, only those that are bifunctional can induce ICLs and are thus used therapeutically as anticancer agents (Table 1.1). Nitrogen mustard ICLs form primarily with guanine residues in 5'-GNC-3' DNA sequences. The reactive moiety of nitrogen mustards is the *N,N*-bis(2-chloroethyl)amine group which reacts with N7 on the guanine base, via the aziridinium intermediate, generating an alkylated guanine derivative. This monoadduct can then form a second aziridinium ion, and react with a guanine base on the opposing strand to generate an ICL (Rink, 1992). This process is illustrated schematically in Figure 1.5. These lesions are helix distorting and the mechlorethamine crosslink generated bends the DNA by approximately 14^{o} (Rink and Hopkins, 1995). It is important to note, however, that

the ICLs described above occur only ~5% of the time, and much more common are DNA monoadducts (~90–95%) (Muniandy et al., 2010).

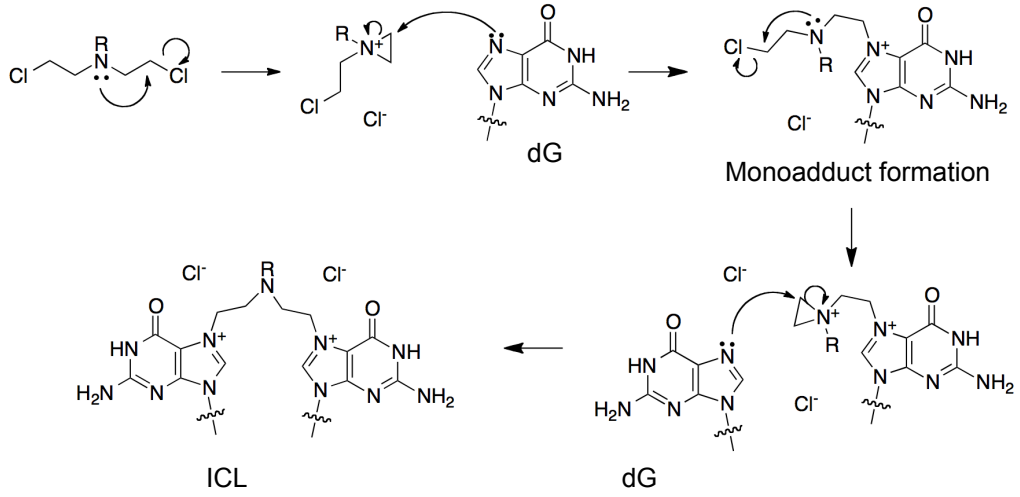


Figure 1.5 ICL formation by a bifunctional nitrogen mustard compound.

The loss of chloride leaving groups of nitrogen mustards results in formation of a reactive positively charged aziridinium intermediate that can alkylate DNA to form an ICL between the N7 positions of guanines. The R specifies the generic alkyl group.

Cisplatin

Although a number of platinum based ICL inducing agents are utilised as chemotherapeutics, cisplatin is the best characterised and most common. Cisplatin, or *cis*-diamminedichloroplatinum (*cis*-DDP) is a bifunctional crosslinking agent, that forms both intrastrand and interstrand crosslinks (Noll et al., 2006). The active form of *cis*-DDP is generated when both chloride ions are displaced by water (Reishus and Martin, 1961). *Cis*-DDP is able to form intrastrand crosslinks between guanine bases in 5'-GG-3' and 5'-GNG-3' DNA sequences, and between adenine and guanine in 5'-AG-3' sequences, and these are the most common

lesions, accounting for more than 90% of adducts formed. Interstrand crosslinks are relatively rare (2–5% of lesions) and are usually formed between guanines in 5'-GC-3' sequences (Kartalou and Essigmann, 2001). Cisplatin induced ICLs cause a significant distortion in the DNA helix, unwinding it by 79^{o} and bending it by 45^{o} (Malinge et al., 1994). The process by which cisplatin induces ICLs is illustrated in Figure 1.6.

Trans-CDDP, the trans stereoisomer of *cis*-CDDP, also reacts with DNA to induce both N7-guanine monoadducts and interstrand crosslinks between the N7-nitrogen of guanine and the N3-nitrogen of an opposing cytosine. However, *trans*-CDDP is not an effective chemotherapeutic agent, possibly because the formation of *trans*-ICLs occurs too slowly, and monoadducts can be scavenged by cellular nucleophiles such as glutathione (Brabec and Leng, 1993, Eastman and Barry, 1987).

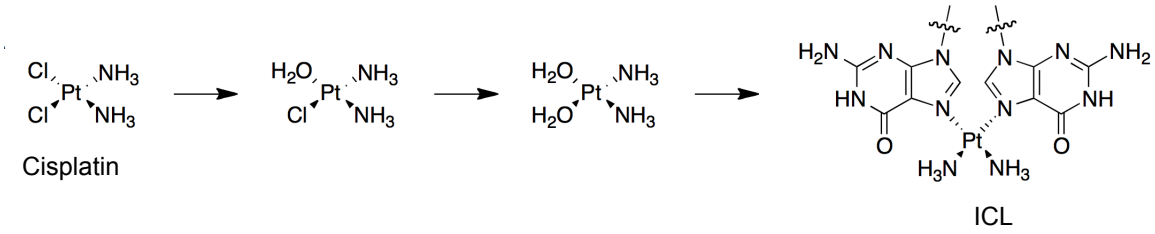


Figure 1.6 ICL formation by cisplatin between N7 positions of guanines.

Mitomycin C

Mitomycin C (MMC) is a bifunctional alkylating agent that induces ICLs. Unlike both cisplatin and the nitrogen mustards, MMC forms crosslinks across the minor groove of the DNA, reacting with the exocyclic amino groups of guanine bases in 5'-CG-3' sequences. MMC is derived from the mould, *Streptomyces caespitosus,*

and possesses little inherent reactivity. Instead the quinone ring of MMC must be enzymatically or chemically reduced to initiate alkylation (Noll et al., 2006).

The reaction mechanism by which MMC induces ICLs is relatively complex and is summarised in Figure 1.7. Briefly, MMC is activated by two-electron reduction of the quinone ring, resulting in the elimination of the methoxy group, and the formation of a hydroquinone intermediate. This tautomerises and reacts with the exocyclic amino group of a guanine base resulting in formation of a monoadduct. The carbamoyl group of the monoadduct is then eliminated, generating a reactive vinylogous hydroquinone methide intermediate, which further reacts with the exocyclic amino group of a guanine on the opposite strand of the DNA duplex (Rajski and Williams, 1998). This MMC-ICL causes very little distortion to the DNA duplex, causing only a slightly opening of the minor groove (Norman et al., 1990).

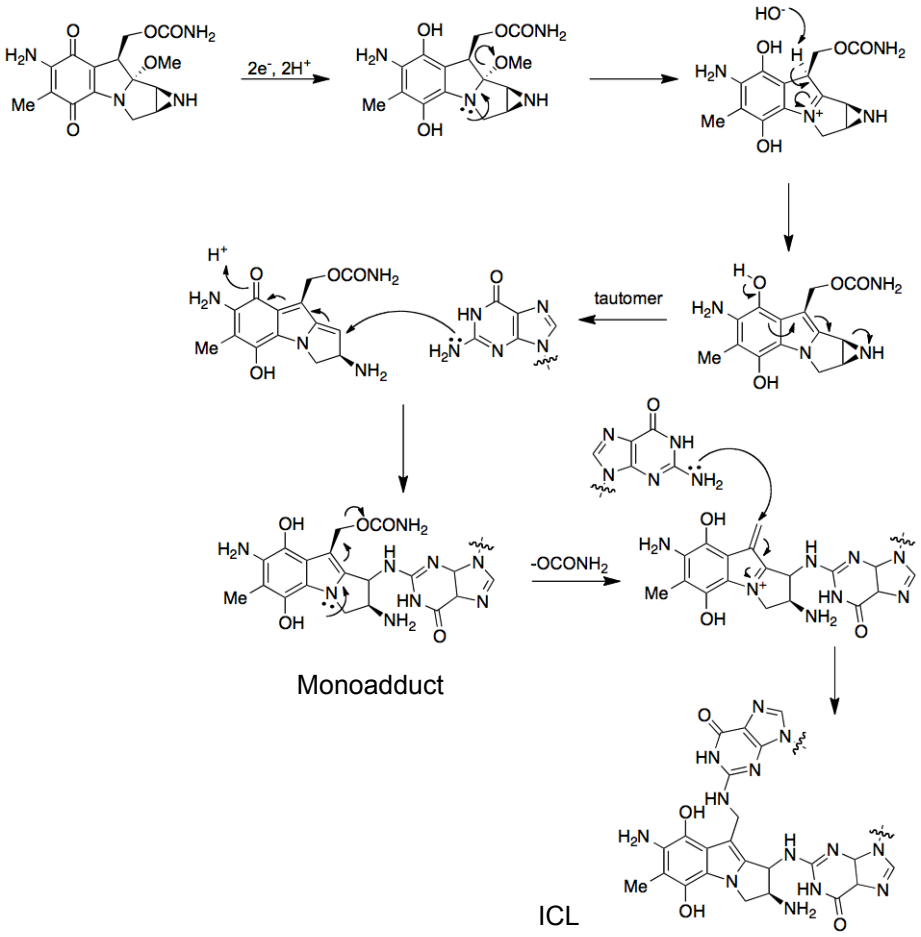


Figure 1.7 ICL formation by mitomycin C (MMC).

MMC undergoes two-electron reduction to generate a DNA crosslinking species. ICLs are formed between N2 positions of guanine residues.

SJG-136

SJG-136 is pyrrolobenzodiazepine (PBD) derivative that has been developed as an anti-tumour therapeutic. The first identified pyrrolobenzodiazepine, anthramycin, was discovered in the 1960s as a natural product of *Streptomyces* (Leimgruber et al., 1965) and, since then, a number of derivatives have been developed due to their genotoxic potential (Hartley, 2011). SJG-136 reacts with the exocyclic amine groups of guanine bases and forms ICLs across the minor

groove, with little perturbation to the DNA helix (Figure 1.8) (Hopton and Thompson, 2011). Unlike other ICL inducing agents, namely the nitrogen mustards and cisplatin, SJG-136 forms primarily interstrand crosslinks; however, intrastrand crosslinks and monoadducts are also generated (Rahman et al., 2009).

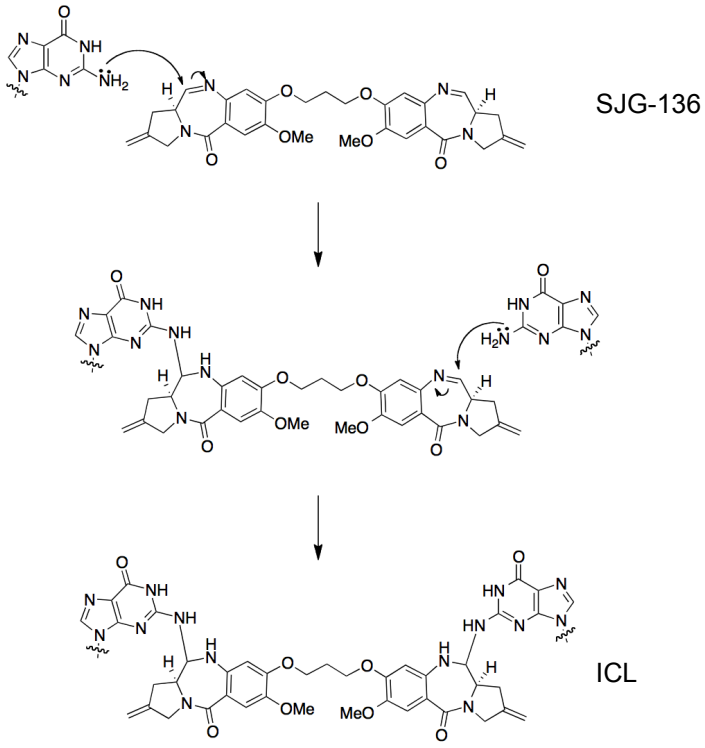


Figure 1.8 ICL formation by SJG-136.

It is worthy of note that there are many other exogenous ICL inducing agents, notably the psoralens, nitrosoureas, and diepoxybutane, for which the specific reaction mechanisms have not been discussed. However, many of the aforementioned ICL inducing agents have been essential in elucidating cellular ICL repair pathways.

1.5.3. ICL repair in yeast

ICL repair pathways were first identified in bacterial cells, however, research then turned to budding yeast as a eukaryotic model organism. Genetic studies in *Saccharomyces cerevisiae* revealed that factors that contribute to ICL repair in yeast belong to one of three epistasis groups. These epistasis groups were found to correspond with NER, HR, and TLS factors (Lehoczky et al., 2007). Further studies revealed that ICL repair in yeast could generally be delineated into two major pathways – the error-prone pathway, relying on NER and TLS factors, and the error-free pathway dependent on NER and HR. Which repair pathway is utilised appears to be influenced by cell cycle phase and the characteristics of the ICL (Noll et al., 2006, Lehoczky et al., 2007). Studies of ICL repair in yeast also identified many of the key proteins that have orthologous counterparts in higher eukaryotic ICL repair.

1.5.4. ICL repair in vertebrates

ICL repair in vertebrate cells shares many conserved characteristics with repair in yeast. The specific processes that constitute vertebrate ICL repair have not been well elucidated, however, it is clear that it is a complex and highly coordinated process involving proteins from a number of different pathways including NER, HR, and TLS. Evidence suggests that ICL repair is highly cell phase specific, and can occur in a replication-dependent, or replication-independent manner (Williams et al., 2013). Most ICLs are detected and repaired during S-phase, mediated by the FA pathway proteins (Kottemann and Smogorzewska, 2013). A simplified overview of the mechanistic steps of ICL repair is shown in Figure 1.9.

ICL repair in S-phase is initiated when the replication machinery encounters an ICL and stalls. Experimental data utilising *Xenopus* extracts show that replication fork convergence is required for subsequent downstream processing steps (Zhang et al., 2015), however, it is not yet clear whether one or two forks are required under physiological conditions. Subsequent detection of fork stalling is mediated by FANCM, which acts as an anchor for the recruitment of the remainder of the FA core complex (comprised of FANCA, B, C, E, F, G, L, M, T, as well as FAAP100, MHF1, MHF2, FAAP20, and FAAP24) (Niedernhofer, 2007). FANCM is recruited to the damage lesion with its constitutive binding partner, FAAP24 (FA associated protein of 24 kDa), which facilitates FANCM DNA binding (Ciccia et al., 2007). The Bloom syndrome complex also promotes the recruitment of FANCM to the stalled replication fork (Ling et al., 2016). Electron microscopy and DNA binding studies have shown that both FANCM and FAAP24 are recruited to, and preferentially bind crosslinked or stalled fork lesions (Coulthard et al., 2013, Xue et al., 2008). Also recruited are MHF1 and 2 (Mph1-associated histone-fold protein 1/2), which load onto the chromatin, forming a heterotetramer (Nishino et al., 2012), and stimulate replication fork remodelling (Yan et al., 2010b). The FANCM-MHF complex promotes branch migration of the replication fork by the ATP-dependent translocase activity of ATM (Gari et al., 2008a, Gari et al., 2008b). This exposes regions of ssDNA, which are quickly bound by RPA, thus activating ATR, resulting in CHK1 phosphorylation and activation of the cell cycle checkpoint. ATR also phosphorylates FANCD2 (Andreassen et al., 2004), FANCI (Smogorzewska et al., 2007), and MRN (Duquette et al., 2012), whilst CHK1 phosphorylates FANCE (Wang et al., 2007). In an, as of yet unidentified process, other FA core complex components (FANCA, B, C, E, F, G, L, T) are recruited to the lesion site and the

phosphorylated FANCD2-FANCI heterodimer is activated by FANCL mediated monoubiquitination (Smogorzewska et al., 2007).

The monoubiquitination of FANCD2/FANCI is an essential event in FA pathway mediated ICL repair and is required for both unhooking of the lesion and translesion synthesis (Knipscheer et al., 2009, Raschle et al., 2008). Monoubiquitinated FANCD2/FANCI is able to specifically interact with other repair factors, namely the TLS polymerase DNA Pol v (Moldovan et al., 2010), FAN 1 (FA-associated nuclease 1) (Huang and D'Andrea, 2010), and SLX4 (Yamamoto et al., 2011). Each of these proteins interacts specifically with the ubiquitin moiety of monoubiquitinated FANCD2 via an ubiquitin binding zinc finger (UBZ) domain. Monoubiquitination of FANCD2 also facilitates downstream FANCD2 focus formation. Deubiquitination of FANCD2 is necessary to complete ICL repair and this is achieved by USP1 (ubiquitin carboxyl-terminal hydrolase 1) in conjunction with UAF (USP1-associated factor) (Nijman et al., 2005, Cohn et al., 2007). However, the specific details regarding the timing and function of this ubiquitination/deubiquitination process are still not well understood.

Nucleolytic incision is required to unhook the ICL at the site of the converged replication fork. Experimental data suggest that each of the replication forks pause 20–40 base pairs away from the ICL, likely due to the steric hindrance imposed by the CMG replicative helicase (comprised of Cdc45, MCM2-7, and GINS). The CMG complex is subsequently displaced from the DNA in a BRCA1 dependent manner, allowing one leading strand to approach the ICL (Long et al., 2014). Subsequent nucleolytic incision is controlled by ubiquitinated FANCD2/FANCI and

a number of nucleases have been implicated in this process. These will be discussed briefly, in turn, below.

XPF (xeroderma pigmentosum group F-complementing protein) and ERCC1 together form a heterodimer with structure-specific endonuclease activity, which specifically cleaves the phosphodiester backbone at junctions between single and double stranded DNA (Sijbers et al., 1996). A number of studies have implicated the endonucleolytic activity of XPF-ERCC1 in the specific repair of ICLs. Mutations in the XPF gene were found in patients with previously unclassified FA; XPF was thus identified as a complementation group of FA, and denoted FANCQ (Bogliolo et al., 2013). XPF-ERCC1 depleted cells are acutely sensitive to ICL inducing agents when compared to WT cells, or those deficient in other NER factors (Niedernhofer et al., 2004), and comet assays show that XPF-ERCC1 depleted cells are unable to complete ICL unhooking (De Silva et al., 2000). Experimental data utilising the *Xenopus* cell free system identified that the presence of XPF was essential for ICL unhooking, and that the presence of molecular scaffold, SLX4, was required for the recruitment and appropriate positioning for catalytic activity of XPF-ERCC1 (Klein Douwel et al., 2014, Klein Douwel et al., 2017). Furthermore, *in vitro* biochemical studies have demonstrated that recombinant XPF-ERCC1 is able to incise a stalled replication fork substrate, 5' to the ICL, in a manner that is dependent on the presence of RPA (Abdullah et al., 2017).

The structure-specific endonuclease heterodimer MUS81-EME1 (essential meiotic structure-specific endonuclease 1) is also implicated in ICL repair. MUS81-EME1 preferentially cleaves 3' flap and replication fork substrates, and depleted mammalian cells exhibit hypersensitivity to ICL inducing agents (Abraham et al.,

2003). The absence of MUS81-EME1 also precludes DSB formation in response to crosslinking agents (Hanada et al., 2006, Hanada et al., 2007). The data, therefore suggest, that both XPF-ERRC1 and MUS81-EME1 are required for ICL unhooking, and reconciling this requires further research. It has been posited that XPF-ERCC1 may be the primary nuclease involved in 5' unhooking of the ICL and that MUS81-EME1 instead functions in the unhooking of stalled or regressed replication forks (Clauson et al., 2013). In support of this, MUS81-EME1 has a defined role in the resolution of Holliday junctions (Chen et al., 2001). Interestingly, MUS81-EME2 (essential meiotic structure-specific endonuclease 2) exhibits broader substrate specificity, and is required for nucleolytic processing and restart of a stalled replication fork (Pepe and West, 2014b, Pepe and West, 2014a). Therefore, it may be that MUS81-EME2 also plays a role in the resolution of ICL stalled replication forks.

Another structure-specific endonuclease implicated in ICL repair is the heterodimer SLX1-SLX4. SLX4 depleted cells exhibit hypersensitivity to ICL inducing agents, and SLX4 was also identified as a complementation group of FA (therefore designated FANCP) (Kang, 2011, Munoz et al., 2009). SLX1-SLX4 has endonucleolytic activity on static and migrating Holliday junctions, thus it acts as a HJ resolvase (Svendsen et al., 2009). However, the role of SLX4 in ICL repair may not necessarily be linked specifically to its functional interaction with SLX1. SLX4 also interacts with a number of other ICL repair nucleases, including XPF, MUS81, and possibly SNM1B (Andersen et al., 2009, Fekairi et al., 2009, Salewsky et al., 2012). SLX4 is specifically recruited to ICLs via its direct interaction with ubiquitinated FANCD2 (Yamamoto et al., 2011). Accordingly, SLX4 may function

as a molecular scaffold for the recruitment of the aforementioned nucleases, thus mediating ICL repair (Munoz et al., 2009).

FAN1 is a structure-specific endonuclease and 5' to 3' exonuclease that interacts specifically with monoubiquitinated FANCD2 and FANCI, and FAN1 depleted cells exhibit hypersensitivity to ICL inducing agents (Liu et al., 2010, Kratz et al., 2010, MacKay et al., 2010, Smogorzewska et al., 2010). Interestingly, germline FAN1 mutations are not associated with FA, but rather, karyomegalic interstitial nephritis (KIN) (Zhou et al., 2012). It also been noted that FANCD2 recruits FAN1 for the restart of stalled replication forks, in a function not required for ICL repair (Lachaud et al., 2016). The data are not clear, but it may be that FAN1 is redundant with other ICL repair nucleases, or that it interacts with specific lesions outside the canonical pathway of FA mediated ICL repair (Ceccaldi et al., 2016)

The 5' to 3' exonucleases SNM1A and SNM1B (sensitive to nitrogen mustard 1A and 1B) have also been implicated in ICL repair. Both SNM1A and SNM1B are human orthologues of yeast SNM1/PSO2, so named for their identification in screens for mutants sensitive to nitrogen mustard and psoralen. Similarly, SNM1A or SNM1B depleted cells exhibit hypersensitivity to ICL inducing agents (Yan et al., 2010a). Cellular experiments suggest that SNM1A is epistatic to XPF-ERCC1 in the resolution of ICLs, and *in vitro* biochemical data show that SNM1A is able to digest past an ICL contained in a duplex that has priorly been incised 5' to the lesion by XPF (Abdullah et al., 2017, Wang et al., 2011). Together, these data suggest that the combined activities of XPF-ERCC1 and SNM1A are enough to process the ICL, so that it becomes a viable substrate for downstream lesion bypass by TLS polymerases. This would effectively abrogate the need for a

second endonucleolytic incision 3' to the ICL. The role of the SNM1B in DNA damage repair will be discussed in detail below.

Subsequent to ICL unhooking, TLS and HR are necessary for lesion bypass and replication fork restart. TLS polymerases are essential in this process for bypassing an unhooked ICL to provide an intact template for HR mediated repair. In mammalian cells, experimental data suggest that Pol ζ (consisting of the REV3 and REV7 subunits) and REV1 are the primary mediators in ICL repair, as cells depleted for either are hypersensitive to ICL inducing agents (Nojima et al., 2005, Gan et al., 2008). Pol ζ and REV1 have also been implicated in both replication-dependent (Sonoda et al., 2003) and replication-independent (Shen et al., 2006, Enoiu et al., 2012) ICL repair. It is likely that ICL repair requires more than one TLS polymerase for lesion bypass. Current models suggest that there is a two-step repair process, whereby a Y family polymerase adds a nucleotide opposite the damaged base, and then a second polymerase completes the extension step (Prakash and Prakash, 2002). This second polymerase is often Pol ζ due to its ability to bypass a range of bulky lesions, and REV1 is often required for regulating the process of polymerase switching (Guo et al., 2003, Sharma and Canman, 2012).

The role of other TLS polymerases in ICL repair is not as clearly defined. Pol η can bypass unhooked ICLs *in vitro* and Pol η depleted cells are sensitive to cisplatin and psoralen (Ho et al., 2011, Albertella et al., 2005, Misra and Vos, 1993). It was also shown that Pol η is involved, but not required, for the replication coupled repair of MMC and psoralen ICLs (Wang et al., 2001, Zheng et al., 2003). Both Pol κ and Pol ν deficient cells exhibit hypersensitivity to MMC (Zietlow et al., 2009).

Pol κ can bypass minor groove ICLs *in vitro* (Minko et al., 2008), but exhibits reduced activity on major groove ICLs (Ho et al., 2011). Conversely, Pol ν can bypass major groove ICLs *in vitro*, but shows very low efficiency on psoralen ICLs (Zietlow et al., 2009). Furthermore, Pol ν interacts with FANCD2 and FANCI suggesting a role in the FA pathway of replication coupled ICL repair (Moldovan et al., 2010).

Replication coupled ICL repair involves the generation of DSBs via nucleolytic incision of the lesion. Thus, subsequent to gap filling by TLS polymerases, homologous recombination is required for replication fork restart. Experimental evidence suggests that there is a significant degree of crosstalk or overlap between the FA mediated repair of ICLs and HR (thus sometimes it is referred to as the FA/BRCA pathway), and mutations in the HR proteins BRCA2, PALB2, BRCA1, and RAD51 have all been identified as FA patient mutations (Ceccaldi et al., 2016). Accordingly, depletion of RAD51, BRCA1, or BRCA2 in human cells results in sensitivity to MMC (Bhattacharyya et al., 2000), and HR-deficient CHO cells exhibit hypersensitivity to nitrogen mustard and cisplatin, whereas NHEJ-deficient cells are only mildly sensitive (De Silva et al., 2000, De Silva et al., 2002). RAD51 has also been shown to be essential for ICL repair in *Xenopus* egg extracts (Long et al., 2011), and BRCA2 and PALB2 were found to be enriched in the chromatin fraction of human cells after crosslink damage in S/G_2 cell cycle phase (Shen et al., 2009). The pathway of HR is initiated by resection of MRN and CtIP, as described in 1.4.2, and MRN forms subnuclear foci after MMC treatment, in a FANCC dependent manner (Pichierri et al., 2002). Take together, these results suggest that HR plays an important role in replication coupled ICL repair.

In contrast to the canonical model of replication-dependent ICL repair, a study in 2013 showed that an ICL is not always a complete barrier to replication. The data show that the DNA translocase FANCM, together with MHF, was able to facilitate replication fork traverse of an ICL without repairing the lesion, and this was promoted by the interaction between FANCM and PCNA (Huang et al., 2013, Rohleder et al., 2016). In cases of replication fork traverse, the ICL would likely be repaired post-replicatively; however, the intermediate nucleolytic processing steps may be similar to replication-coupled ICL repair. An incision independent ICL repair pathway was also posited when it was shown that the BER glycosylase, NEIL3, was able to unhook interstrand crosslinks derived from psoralen or an abasic site, effectively negating the need for nucleolytic incision (Yang et al., 2017a).

It is also clear that ICL repair can occur in a replication-independent manner. This process was first observed in ICL repair reporter assays with plasmids that contained no replication origins required for replication-dependent repair to occur (Wang et al., 2001, Enoiu et al., 2012, Zheng et al., 2003). It has been shown in both yeast and mammalian cells that ICLs can be resolved in G_1 phase of the cell cycle, or in non-cycling cells, in an NER dependent manner (Sarkar et al., 2006, Muniandy et al., 2009). NER repair factors, XPA, XPB, XPC, and XPF, are also recruited to sites of psoralen + UVA induced ICLs, and repair of these lesions is dependent on Pol ζ and XPC (Muniandy et al., 2009, Sarkar et al., 2006). Most ICL lesions appear to be recognised by both global genome NER and transcription-coupled NER machinery; however, the repair of cisplatin ICLs appears to be entirely dependent on CSB, a transcription-coupled NER factor (Enoiu et al., 2012). The experimental data suggest that XPF-ERCC1, XPG, and

SNM1A may be the nucleases involved in incising and processing the ICL, leaving only a monoadducted base (Iyama et al., 2015, Enoiu et al., 2012). This, therefore, provides a substrate for TLS polymerases (as discussed above) and subsequent repair of the lesion. The molecular processes of replication-independent ICL repair are not well elucidated, however, a simplified version of the proposed scheme is illustrated in Figure 1.9.

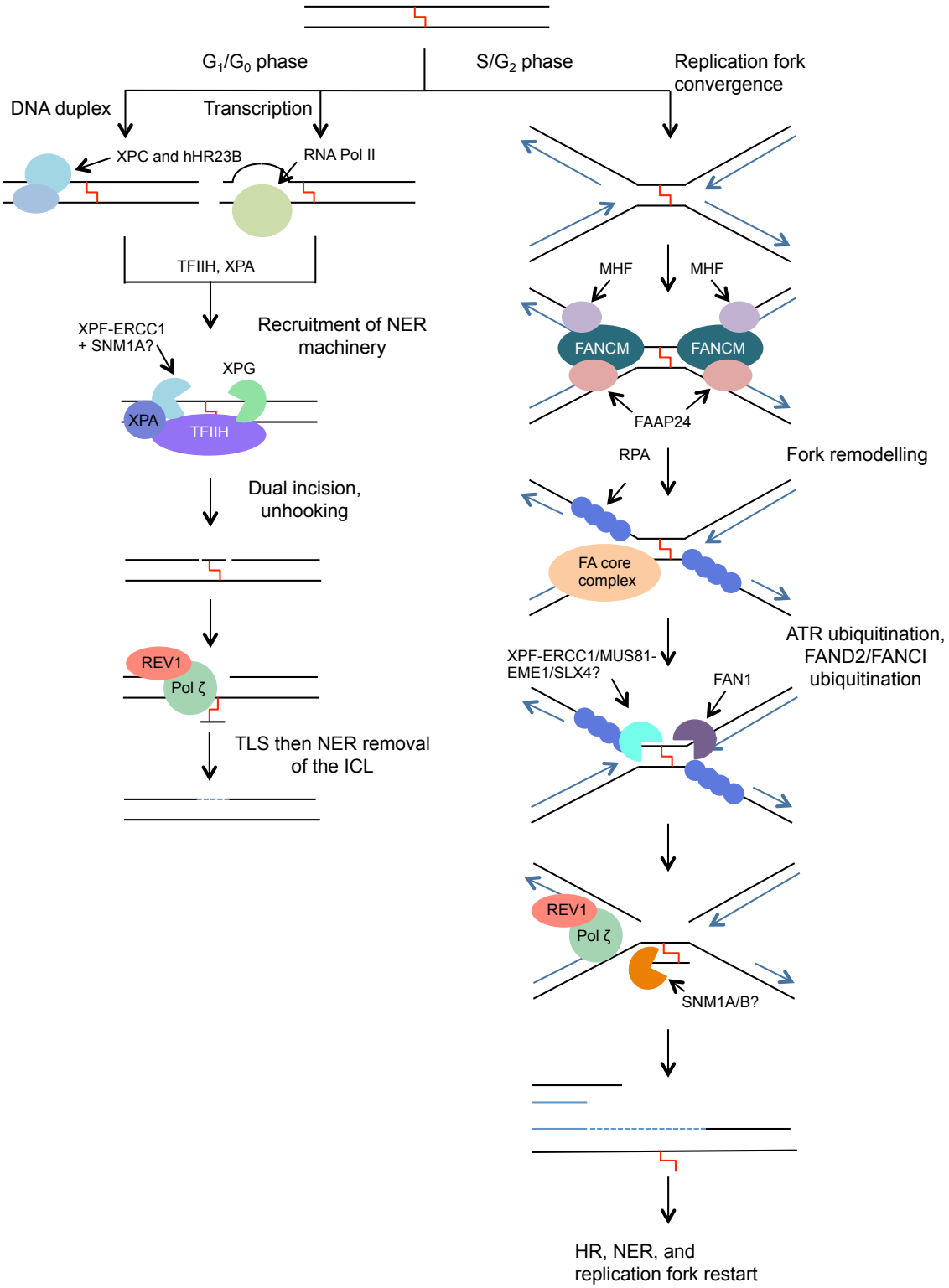


Figure 1.9 Schematic representing ICL repair.

Constituent protein factors are as labelled, and the dashed line represents DNA that may be newly synthesised due to repair by TLS polymerases.

1.6. SNM1B

1.6.1. The *SNM1* family

The *SNM1/PSO2* gene was first identified in the early 1980s when genetic screens in *Saccharomyces cerevisiae* revealed mutants were sensitive to bifunctional alkylating agents, but not to monofunctional alkylating agents, ionising radiation, or UV light (Henriques and Moustacchi, 1980, Henriques and Moustacchi, 1981, Ruhland et al., 1981a, Ruhland et al., 1981b). The mutated genes in these screens, *PSO2* (sensitive to psoralen) and *SNM1* (sensitive to nitrogen mustard) were later found to be allelic (Cassier-Chauvat and Moustacchi, 1988). Ten *PSO* genes were identified, and were hence labelled *PSO1-10*; each of these was shown to play a role in DNA damage repair, but the gene, *PSO2,* encoded a protein (Pso2p) that was found to be specifically essential for ICL repair (Brendel et al., 2003). Yeast Pso2p has been shown to have both 5' to 3' exonuclease activity, and structure-specific endonuclease activity (Tiefenbach and Junop, 2012, Li et al., 2005).

Three vertebrate orthologues of *PSO2/SNM1* have been identified, and these are SNM1A (KIAA0086), SNM1B (Apollo), and SNM1C (Artemis) (Demuth and Digweed, 1998, Aravind, 1999, Nagase et al., 1995). The human gene names are, respectively: DCLRE1A, DCLRE1B, and DCLRE1C. All are members of the MBL (metallo-β-lactamase) fold containing superfamily of enzymes, and are further classified into the self-defining β-CASP (CPSF, Artemis, SNM1, PSO2) family of nucleic acid processing MBLS (Aravind, 1999, Callebaut et al., 2002).

SNM1A has the highest level of amino acid sequence similarity with yeast Pso2p at 48%, and is the only human orthologue where ectopic expression is able to partially restore the resistance of *PSO2* mutants to ICL inducing agents (Demuth and Digweed, 1998, Hazrati et al., 2008). SNM1A possesses 5' to 3' exonuclease activity and is able to digest a variety of DNA substrates, including those that contain various DNA adducts (Sengerova et al., 2012, Allerston et al., 2015). The cellular role of SNM1A is not well elucidated, however it is clear that SNM1A does play a role in ICL repair and this is discussed above (see 1.5.4).

SNM1C is a structure-specific endonuclease, which plays an important role in NHEJ DSB repair, and in V(D)J recombination. Accordingly, germline mutations in SNM1C result in radiosensitive severe combined immunodeficiency (Moshous et al., 2001). When complexed to, and phosphorylated by, DNA-PKcs, SNM1C has the ability to nucleolytically cleave 5' and 3' overhangs, flaps, hairpins, and gapped structures (Ma et al., 2002, Ma et al., 2005). The role of SNM1C in NHEJ is described in 1.4.1.

SNM1B is the subject of this study and, therefore the structural, biochemical, and cellular aspects will be discussed below. The SNM1B gene consists of four exons, is situated on chromosome 1p13.1–13.3, and encodes an open reading frame of 532 amino acids. The SNM1B protein exhibits 33% sequence similarity with Pso2p in its N-terminal catalytic domain (Schmiester and Demuth, 2017). Two isoforms of SNM1B exist – the longer full-length isoform, and a shorter isoform lacking exon 2. The significance of this shorter isoform is not clear and it is likely to lack most of the MBL domain (Demuth et al., 2004), suggesting it wouldn't be catalytically active. Interestingly, expression of the shorter transcript has been associated with

an increased risk of breast cancer (Caswell et al., 2015). SNM1B possesses 5' to 3' exonuclease activity, is essential for normal telomere maintenance, and plays a role in the cellular response to genotoxic agents.

1.6.2. Structural characteristics of SNM1B

As mentioned above, SNM1B is a member of the metallo-β-lactamase superfamily. The 'true' MBLs are a subclass of the bacterial β-lactamases (BL), which are responsible for the hydrolysis of β-lactam antibiotics, such as penicillins and cephalosporins. The MBLs are distinct from the serine-β-lactamases as they require either one or two active site metal ions and an activated water molecule for hydrolysis (Palzkill, 2013).

The MBL family is defined by structural similarity, and all members possess the characteristic α/β/β/α MBL fold. This molecular scaffold coordinates the metal ions necessary for hydrolysis, and most MBLs appear to bind zinc (II) in the active site, although iron (II) or (III) or manganese (II) have also been reported. MBL fold containing enzymes catalyse an array of reactions, including the hydrolysis of phosphodiester bonds and thioesters, as well as redox reactions (Bebrone, 2007, Palzkill, 2013). Most MBL family members share five highly conserved motifs: where motif 1 is an acidic residue; motif 2, the HxHxDH sequence; motif 3, a histidine residue; motif 4, an acidic or cysteine residue; and motif 5, another cysteine. These amino acid residues define the active site of the enzymes and participate in metal ion coordination and hydrolysis (Callebaut et al., 2002). To date, 18 human MBL fold containing proteins have been categorised and they possess a variety of functions, including small molecule metabolism and nucleic

acid processing, and these are summarised in Table 1.2 (Daiyasu et al., 2001, Garau et al., 2004, Pettinati et al., 2016).

Members of the MBL family have been classified into 16 sub-groups, based on biological function (Daiyasu et al., 2001). The nucleic acid processing MBLs were thus classified into group 6 and group 7. RNA processing enzymes, such as CPSF (cleavage and polyadenylation specificity factor) and RNase J, were categorised as group 6; and DNA processing enzymes, such as SNM1A, B, C, were categorised as group 7 (Daiyasu et al., 2001). There are nine group 6 and 7 human MBLs that have been identified; these are SNM1A/B/C, CPSF-73, CPSF-100, ELAC1, ELAC2, Int9, and Int11 (Table 1.2).

Table 1.2 Human MBL-fold containing proteins.

Protein name	Enzymatic function	Presence of β-CASP domain	Crystal structure
Glyoxalase II Subfamly			
HAGH	Glyoxalase	No	Yes (PDB: 1QH3/5)
HAGHL	Redox	No	No
ETHE1	Redox	No	Yes (PDB: 4CHL)
LACTB2	RNase	No	Yes (PDB: 4AD9)
MBLAC1	Unidentified	No	No
MBLAC2	Unidentified	No	No
PNKD	Unidentified	No	No
The DNA/RNA Interacting Subfamily			
SNM1A	DNase	Yes	Yes (PDB: 4B87)
SNM1B	DNase	Yes	Yes (PDB: 5AHO)
SNM1C	DNase	Yes	No
ELAC1	RNase	No	Yes (PDB: 3ZWF)
ELAC2	RNase	No	No
CPSF-73	RNase	Yes	Yes (PDB: 2I7T)
CPSF-100	No enzymatic function	Yes	No
Int9	No identified enzymatic function	Yes	No
Int11	RNase	Yes	No
Other human MBL enzymes			
NAPE-PLD	Phospholipase	No	Yes (PDB: 4QN9)
CMAHP	No identified enzymatic function in humans	No	No

In vitro, CPSF-73 possesses both 5' to 3' exoribonuclease and endoribonuclease activity (Yang et al., 2009). *In vivo*, CPSF-73 dimerises with CPSF-100 (which is catalytically inactive) to form part of the CPSF complex responsible for 3' end processing of mRNA (Mandel et al., 2006, Dominski et al., 2005a). Int11 (Integrator complex subunit 11) possesses endoribonucleolytic activity and dimerises with Int9 (Integrator complex subunit 9) to form part of the Integrator complex. Similar to CPSF-100, Int9 possesses no catalytic activity. The Integrator complex is responsible for the 3' end processing of small nuclear RNA (snRNA) (Albrecht and Wagner, 2012, Dominski et al., 2005b). ELAC1 and ELAC2 are two isoforms of RNase Z, which catalyse the removal of 3' extensions from tRNA precursors, thus mediating an essential step in tRNA biogenesis. ELAC1 is localised in the cytosol and its cellular function is not well elucidated (Rossmanith, 2011). ELAC2 is primarily located in the nucleus and mitochondria, and plays a role in nuclear and mitochondrial tRNA processing (Bogenhagen et al., 2014, Brzezniak et al., 2011).

A subset of the nucleic acid processing MBLs is delineated into the self-defining β-CASP family of enzymes (Table 1.2) (Callebaut et al., 2002). This family includes Int11, CPSF-73, and SNM1A/B/C, and all members contain the MBL domain and an additional β-CASP domain. This domain consists of a fully parallel β-sheet, flanked by α-helices, which inserts within the MBL domain (see Figure 1.10 for the structure of SNM1B). The β-CASP domain contains three conserved motifs: motif A is an acidic amino acid residue at the end of a stretch of hydrophobic residues; motif B is a histidine residue at the end of an amphiphilic β-strand that is linked to an α-helix; and motif C is a conserved histidine (RNA processing MBLs) or valine (DNA processing MBLs) residue. It has been suggested that motif C replaces motif

5 of the MBL domain in this family of enzymes (Callebaut et al., 2002). These three motifs, as well as the four motifs in the MBL domain, coordinate the active site metal ions, the nucleophilic water, and the DNA substrate. The role of these motifs in the reaction mechanism of SNM1B is displayed in Figure 1.10.

Of the β-CASP-MBLs a handful of structures have been solved, including: human CPSF-73; yeast CPSF-100 (Mandel et al., 2006); bacterial RNase J and RNase J1 (Dorleans et al., 2011, Newman et al., 2011); bacterial MTH1203 (Silva et al., 2011); and human SNM1A and SNM1B (Allerston et al., 2015). The overall architecture of these proteins is similar and all show the characteristic bilobar MBL and β-CASP domain structure. The structure of SNM1B is shown in Figure 1.10 and the MBL and β-CASP domains are as labelled. This structure is of a truncated form of SNM1B (amino acids 1–335) containing both the MBL and β-CASP regions, but lacking the C-terminal region. The MBL domain consists of two mixed β-sheets, flanked by α-helices, with the di-metal binding site located at the edge of the β-sandwich. The β-CASP is inserted between strands 10 and 11 of the MBL domain, and consists of a four-stranded parallel β-sheet, flanked by α-helices. The catalytic core (containing the metal ion-binding site and presumed substrate binding site) is interfaced between the two domains (Allerston et al., 2015). The structures of SNM1A and SNM1B are very similar, and both show a positively charged face, hypothesised to mediate binding to the DNA substrate. Interestingly, SNM1A exhibits a more extensive region of positive charge than SNM1B, and it was suggested that this explained the greater processivity of SNM1A, particularly on higher molecular weight substrates. It was also observed that both the SNM1A and SNM1B structures had a negatively charged ‘reverse’ face, although the reasons for this are not clear (Allerston et al., 2015). However, both the SNM1A

and SNM1B structures are of truncated proteins (containing only the known catalytic domains) and it may be that full-length proteins fold in a different manner.

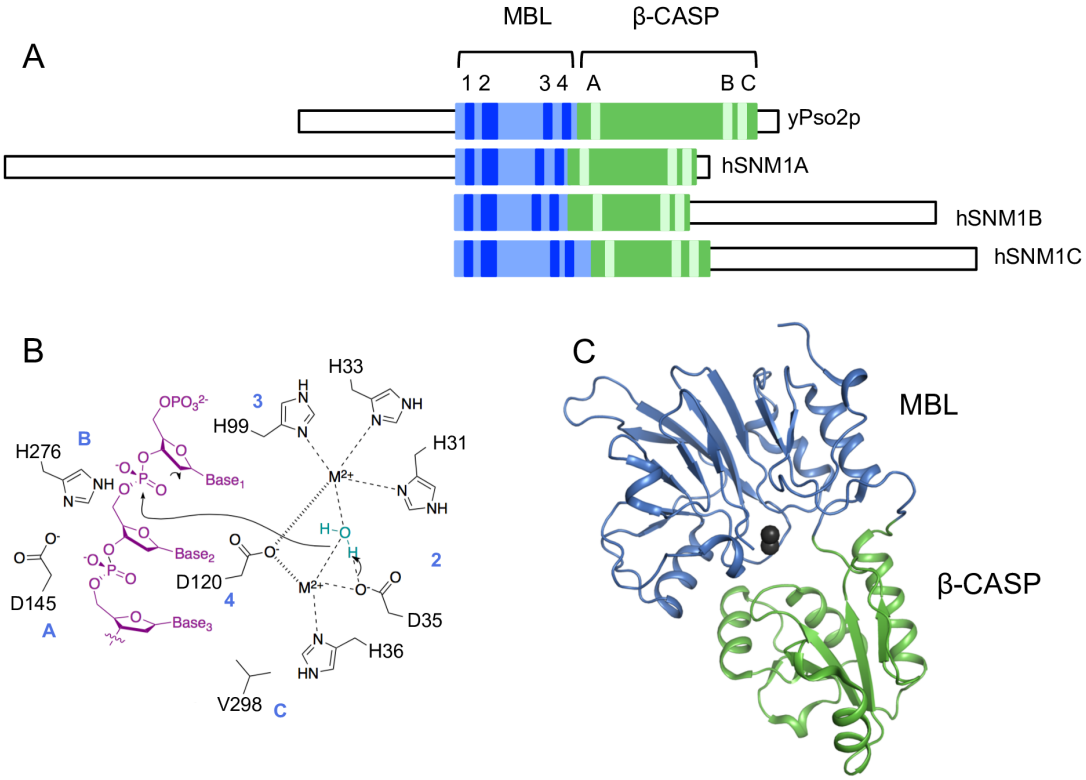


Figure 1.10 The MBL and β-CASP domains of SNM1B.

A: Linear representation of the primary amino acid sequences of yeast Pso2p, SNM1A, SNM1B, and SNM1C. The MBL domains are in light blue, and the conserved motifs in dark blue; the β-CASP domains are in green, with the conserved motifs in light green. B: 2D representation of active site of SNM1B showing the conserved motif residues as labelled. The motifs are numbered in blue, 2–4 are from the MBL domain, and A–C from the β-CASP domain. C: The structure of truncated WT SNM1B (PDB: 5AHO) represented in cartoon form. The MBL and β-CASP domains are as labelled and the metal ions are depicted as grey spheres. Panel A and B are adapted from (Sengerova et al., 2012).

This structure of SNM1B provides important information regarding the overall architecture, and suggests how DNA substrate interactions may be mediated via the positively charged putative DNA binding cleft. The configuration of the active

site also allows deduction of the catalytic mechanism (represented in Figure 1.10) and reveals a relatively open active site, which may explain the ability of SNM1B to digest past DNA adducts *in vitro* (Allerston et al., 2015). However, in the absence of a co-crystal structure of SNM1B with DNA it is difficult to understand precisely how SNM1B interacts with DNA in the active site and how the structure of SNM1B is related to substrate selectivity.

1.6.3. The role of SNM1B at telomeres

Perhaps the best characterised cellular role of SNM1B is in telomere maintenance, where its 5' to 3' exonuclease activity is responsible for maintaining the overhangs necessary for t-loop formation at newly replicated leading strand telomeres.

Mammalian telomeres are complex nucleoprotein structures that exist at the ends of linear chromosomes. They consist of arrays of TTAGGG repeats (5–15 kb) that terminate with a 3' single stranded overhang of the G-rich strand, consisting of 300–4000 nucleotides. This overhang folds back and forms a t-loop structure, whereby the 3' single stranded DNA invades the duplex DNA, thus stabilising the telomeric ends. Telomeric repeat sequences (TTAGGG) are specifically recognised by several telomeric proteins, which bind to these regions, regulate telomere length, and protect chromosome ends. These proteins form a complex called shelterin which consists of six members: TRF1 (telomeric repeat-binding factor 1), TRF2 (telomeric repeat-binding factor 2), POT1 (protection of telomeres 1), TIN2 (TRF1 and TRF2 interacting nuclear factor 2), TPP1, and RAP1 (repressor/activator protein). TRF1 and TRF2 specifically recognise, and bind to, telomeric repeat DNA; whilst POT1 specifically binds to the single stranded 3'

overhang. TRF1 and TRF2 recruit TIN2, RAP1, and TPP1 to telomeric sequences (de Lange, 2005, Xin et al., 2008, Maestroni et al., 2017).

The shelterin complex is essential for repressing the DNA damage response at telomeres and inhibiting unscheduled DNA damage repair. This is achieved in part, by formation of the t-loop, as it shields telomeric ends and prevents cells from recognising blunt ended chromosomes as DSBs and inappropriately activating DNA repair pathways, particularly NHEJ. If the shelterin complex is inhibited, and t-loops do not form, then a telomere dysfunction phenotype persists, evidenced by telomere dysfunction induced foci (TIFs), chromosome end-to-end fusions, sequence alterations due to one-ended HR, and cell cycle arrest (de Lange, 2009). In addition to the shelterin components, there are a number of other proteins that are located at telomere ends. These do not fulfil the criteria of a shelterin protein (i.e. they are less abundant or transiently located at telomeres, or they also fulfil other cellular functions), and are instead denoted as shelterin associated proteins (Diotti and Loayza, 2011).

A decade ago, four groups independently identified that SNM1B binds to TRF2, and is localised to telomeres under normal cellular conditions. Evidence for both points is manifold: SNM1B was shown to co-immunoprecipitate with TRF2 from 293T cells, the two proteins co-localised at telomeres (Freibaum and Counter, 2006), and the interaction with TRF2 stabilised cellular SNM1B (Freibaum and Counter, 2008); SNM1B and TRF2 interacted in a GST pull-down assay, and TRF2 was required to localise SNM1B to telomeric regions (Lenain et al., 2006); a proteomic mass spectrometry approach identified SNM1B as an interacting partner of TRF2, and this was validated by co-immunoprecipitation from 293T cells

(van Overbeek and de Lange, 2006); SNM1B and TRF2 were also shown to co-immunoprecipitate from HEK293T cells, and immunofluorescence indicated that SNM1B and TRF2 co-localised at telomeres (Demuth et al., 2008).

This interaction was mapped to the TRFH (TRF homology) domain of TRF2, and the C-terminal region of SNM1B (Lenain et al., 2006, Freibaum and Counter, 2006, van Overbeek and de Lange, 2006). To further characterise this interaction, Chen and colleagues used isothermal titration calorimetry and X-ray crystallography. The structure of the TRF2 TRFH domain and a peptide of SNM1B (496–532) revealed that amino acids 498 to 509 (in SNM1B) mediated binding, and contained the typical TRFH-binding signature motif (Y/F/H-X-L-X-P). When L506 and P508 (the latter part of the motif) were each mutated to alanine, the interaction of SNM1B with TRF2 was abrogated and SNM1B no longer localised to telomeres (Chen et al., 2008). The location of the TRF2 binding domain in the amino acid sequence of SNM1B is represented schematically in Figure 1.11.

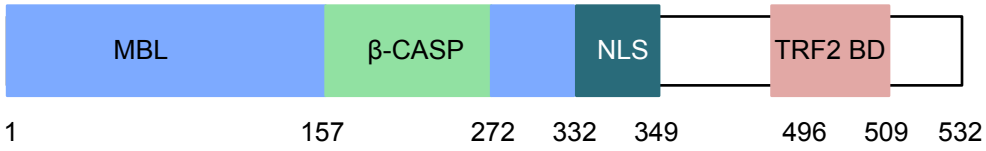


Figure 1.11 Linear representation of the protein sequence of SNM1B.

The MBL, β-CASP, nuclear localisation signal (NLS) and TRF2 binding domain (TRF2 BD) are as indicated.

Lagging strand telomeres are replicated in a discontinuous manner, and their replication results in a small 3' overhang due to the removal of the final RNA primer. Conversely, leading strand telomeres are replicated continuously and terminate in a blunt ended manner. When SNM1B was found to interact with TRF2 and localise to telomeres it was hypothesised that SNM1B, due to its nucleolytic activity, may be involved in resection and generating the 3' overhang requisite for t-loop formation (Lenain et al., 2006, Freibaum and Counter, 2006). Accordingly, experimental data from SNM1B depleted cells suggested that SNM1B plays an important role in regulating replication at telomeres: shRNA against SNM1B in human cells resulted in an increased number of TIFs during S-phase, accumulation of telomere doublets indicating impaired replication, and induction of a senescent phenotype (van Overbeek and de Lange, 2006). It was also shown that TRF2 and SNM1B cooperate with topoisomerase 2α to relieve topological stress during telomere replication, and the nuclease domain of SNM1B is essential for this process (Ye et al., 2010).

The role of SNM1B at telomeres was further elucidated by (Lam et al., 2010) and (Wu et al., 2012), each of whom showed a role for SNM1B in generation of the 3' overhang. Lam *et al.* showed that SNM1B$^{-/-}$ MEFs (mouse embryonic fibroblasts)

exhibited an increased number of TIFs, an increased number of leading strand telomere chromatid fusions, and a 65% reduction in the telomeric 3' overhang signal. Each of these phenotypes could not be rescued by two 'nuclease dead' SNM1B mutants, implicating the nucleolytic activity of SNM1B as essential in preventing telomeric defects. Furthermore, deletion of Ku70 decreased the prevalence of chromatid fusions in SNM1B$^{-/-}$ MEFs, suggesting that end-joining was mediated by NHEJ (Lam et al., 2010). Wu *et al.* showed that the nuclease activity of SNM1B was essential for preventing TIF formation, ATM activation, and CHK2 phosphorylation at telomeres, as well as preventing chromatid fusion at leading strand telomeres. Depletion of SNM1B also resulted in a 30–40% reduction in the telomeric overhang signal, near to the 50% expected if the overhang of leading and lagging strand telomeres were considered to be equal. It was also shown that if the interaction between SNM1B and TRF2 was abrogated, SNM1B failed to localise to telomeres, and reproduced the phenotype of SNM1B null cells (Wu et al., 2010). Further experiments by Wu *et al*. examined the regulation of resection at telomeres and they showed that the shelterin component, POT1b (mice have two POT1 paralogues), inhibited hyper-resection by SNM1B. It was also shown that EXO1 contributed to resection at both leading strand and lagging strand telomeres (Wu et al., 2012).

Recruitment of SNM1B to telomeres by TRF2 appears to be controlled by cell cycle phase and regulated by the phosphorylation status of NBS1. Phosphorylation of NBS1 at S433 by CDK2 occurs in S/G_2 phase of the cell cycle and disrupts the NBS1-TRF2 interaction, thus promoting the TRF2-SNM1B interaction. NBS1 and SNM1B both bind to the TRFH domain of TRF2, and NBS1 is able to outcompete SNM1B for binding. MEFs expressing dephosphorylated

NBS1 exhibit a dramatic increase in chromatid fusions, and SNM1B exhibits significantly less recruitment to telomeres when S433A NBS1 (a variant that cannot be phosphorylated) was present, when compared to WT NBS1 (Rai et al., 2017).

Together, the data show that SNM1B is recruited to telomeres via a specific interaction with TRF2 during S-phase, where it is required for maintaining the 3' overhang at newly replicated leading strand telomere ends (Figure 1.12). Some experimental results are somewhat contradictory (for example, regarding the relationship to ATM activation and downstream phosphorylation; (Wu et al., 2010) showed that, in the absence of SNM1B, NHEJ mediated repair was ATM dependent, whereas (Lam et al., 2010) suggested it was ATM independent) and further work is required to elucidate this. However, it is clear that the role of SNM1B at telomeres is essential for normal cellular functioning. $SNM1B^{-/-}$ mice exhibit perinatal lethality due to multiple organ defects and profound developmental delay (Akhter et al., 2010). In humans, an individual with Hoyeraal-Hreidarsson (HH) syndrome was found to have a splice variant of SNM1B that lacked the TRF2 interaction domain and was unable to localise to telomeres. This variant exhibited a dominant negative effect, and patient fibroblasts exhibited a severe telomere deprotection phenotype and failed to proliferate. The patient experienced developmental defects, severe bone marrow failure, and premature death (Touzot et al., 2010).

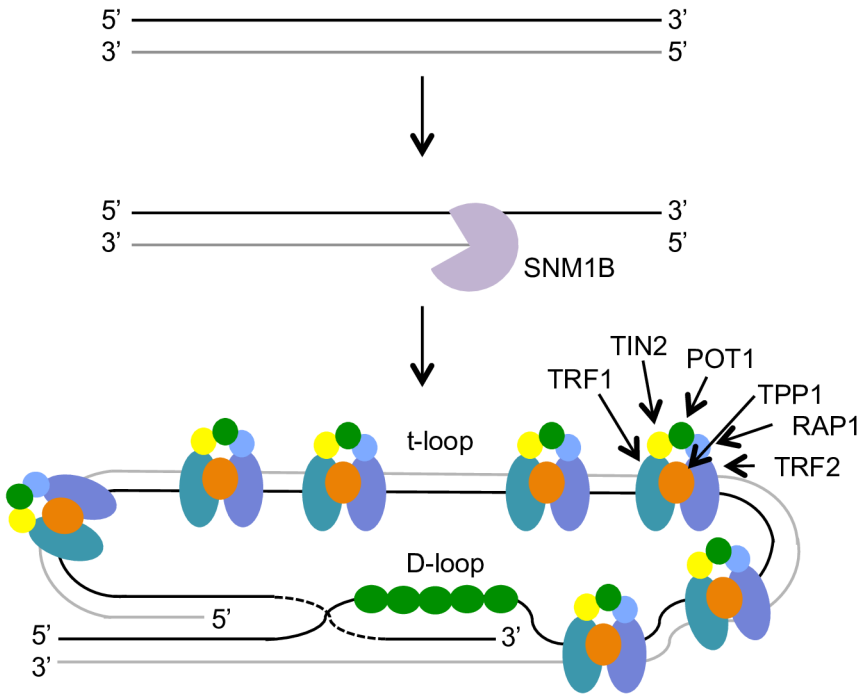


Figure 1.12 Schematic of the role of SNM1B at leading strand telomere ends.

Newly replicated leading strand telomeres are blunt ended. SNM1B is recruited to telomere ends via interaction with TRF2. SNM1B then resects, via its 5' to 3' exonuclease activity. This generates the 3' single stranded DNA which loops back (forming a t-loop) and invades the DNA duplex, forming a displacement (or D-) loop. This structure sequesters telomeric ends, and is stabilised by the binding of shelterin proteins (POT1, TIN2, TPP1, TRF1, TRF2, and RAP1).

The role of SNM1B at telomeres is not a focus of this study, however, there is significant overlap between proteins involved in telomere maintenance, and those that contribute to DNA damage repair pathways. For example, TRF2 localises to double strand breaks elsewhere in the genome (Mao et al., 2007), FAND2 localises to telomeres via interaction with TRF1 (Fan et al., 2009), FANCJ also localises to telomeres (Dejardin and Kingston, 2009), MRN plays a role in telomeric end resection (Lamarche et al., 2010), SLX4 is recruited to telomeres by TRF2, and SLX1 is essential for telomeric trimming (Kim, 2014).

1.6.4. The role of SNM1B in DNA damage repair

The current literature also supports a role for SNM1B in DNA damage repair, particularly the repair of ICLs. Depletion of SNM1B in DT40 cells results in sensitivity towards MMC and cisplatin (Ishiai et al., 2004, Nojima et al., 2005). RNAi based depletion of SNM1B renders human cells hypersensitive to MMC, cisplatin, and psoralen + UVA with regards to both short-term viability, and in longer-term clonogenic survival assays (Bae et al., 2008, Mason and Sekiguchi, 2011, Demuth et al., 2004). Depletion of SNM1B also resulted in an increased number of chromosomal aberrations (including gaps, breaks, and radial structures) after treatment with ICL inducing agents (Bae et al., 2008, Mason and Sekiguchi, 2011). Sensitivity of SNM1B depleted cells to IR was contradictory, with Bae *et al.* reporting no sensitivity in HEK293T cells, whereas Demuth *et al.* found increased sensitivity in GM00639 cells (Bae et al., 2008, Demuth et al., 2004). The latter study also showed that there was a small but significant increase in the number of SNM1B foci per nucleus after IR exposure in immunofluorescence experiments, and that GFP tagged SNM1B localised to laser micro-irradiation induced DSBs in GM00639 cells by 10 seconds post-irradiation (Demuth et al., 2004). The data clearly implicate SNM1B in ICL repair, and possibly DSB repair; however, the function of SNM1B in these pathways is not yet well elucidated.

The phenotype of SNM1B depleted cells in response to ICL inducing agents is evocative of FA cells. Accordingly, studies have provided insight to the relationship between SNM1B and the FA pathway. Experimental data suggest that SNM1B functions epistatically to FANCD2 in ICL repair: survival assays after MMC

treatment show similar survival rates for single vs double depletion of SNM1B and FANCD2; and SNM1B was also shown to function epistatically with FANCD2 and FANCI in the suppression of ICL induced aberrant chromosomal structures (Mason and Sekiguchi, 2011). SNM1B is not required for FANCD2 ubiquitination (Demuth et al., 2004, Bae et al., 2008), but is required for the appropriate localisation of FANCD2, BRCA1, and RAD51 to subnuclear repair foci following MMC treatment (Mason and Sekiguchi, 2011, Mason et al., 2013). FANCD2 and SNM1B also co-immunoprecipitate from HEK293T cells and although this interaction is not direct, it does suggest they might form part of the same complex (Bae et al., 2008). SNM1B has also been shown to directly interact with another FA repair factor, SLX4; and in U2OS cells SLX4 and SNM1B were also shown to function epistatically in response to treatment with MMC or IR (Salewsky et al., 2012, Schmiester and Demuth, 2017).

SNM1B also appears to play a role in the repair of stalled replication forks. Faithful and timely replication of the genetic material is essential for normal cellular functioning. However, the presence of DNA lesions, secondary DNA structures, or transcription/replication conflicts can contribute to replication stress and fork stalling. There are cellular mechanisms to restart stalled forks, however, replication fork collapse due to replication run-off or inadequate repair leads to the formation of double strand breaks (Berti and Vindigni, 2016). Stalled replication forks at ICLs are often collapsed in a controlled manner by endonucleolytic incision to bypass the ICL and allow replication fork restart. The DSBs that arise due to collapsed replication forks can be mutagenic or clastogenic if not accurately repaired. SNM1B depleted HEK293T cells exhibit a decreased ability to form DSBs in response to ICL treatment and SNM1B was shown to directly interact, via

co-immunoprecipitation experiments, with the structure-specific endonuclease MUS81-EME1. This interaction was mediated via the MBL domain of SNM1B to MUS81. It was therefore hypothesised that SNM1B and MUS81-EME1 collaborate in the nucleolytic replication fork collapse necessary at ICLs (Bae et al., 2008). It is possible, for example, that in a regressed fork, SNM1B might play a role in digesting the lagging strand, thus generating a suitable substrate for MUS81-EME1. It has also been shown via co-immunoprecipitation experiments that SNM1B directly interacts with both MRE11 and RAD50, constituent components of the MRN complex (Bae et al., 2008). As MRN and SLX4 have been implicated in the initial processing of ICLs to generate DSBs (see 1.5.4), the interaction of SNM1B with these factors suggests that SNM1B may also be involved in this process.

With regards to processing ICLs, it has also been hypothesised that both SNM1A and SNM1B may play a role in digesting past the crosslinked nucleotide after incision 5' to the ICL. Evidence for this role is stronger for SNM1A, as SNM1A was shown to be epistatic with XPF in ICL repair (Wang et al., 2011). Interestingly, cells depleted for both SNM1A and SNM1B exhibit a more severe phenotype in response to ICL inducing agents, than when depleted for either gene alone, thus suggesting there might be a degree of functional redundancy between the two (Nojima et al., 2005, Ishiai et al., 2004). Furthermore, both SNM1A and SNM1B possess the ability to digest past an SJG-136 crosslinked substrate *in vitro*, albeit SNM1B with reduced efficiency compared to SNM1A (Allerston et al., 2015).

In addition to the repair of ICLs, FA factors have been shown to contribute to a repair independent pathway, whereby they protect stalled replication forks from

degradation (Schlacher et al., 2012). To investigate if SNM1B played a role in this process, Mason *et al.* examined the role of SNM1B in repair or stabilisation of aphidicolin stalled replication forks. It was observed that SNM1B depleted fibroblasts showed decreased survival in response to aphidicolin treatment. SNM1B was not involved in the initial recognition and signalling of fork stalling, as PCNA ubiquitination and ATR phosphorylation were unaffected, but was required for the recruitment of FANCD2 and BRCA1 to the stalled fork. The recruitment of FANCD2 and BRCA1 was shown to be dependent on the nuclease activity of SNM1B as a 'nuclease dead' mutant of SNM1B (D14N) was unable to rescue the phenotype, whilst WT SNM1B was. This suggests that SNM1B exonucleolytically processes the stalled fork in some way, prior to the recruitment of downstream repair factors. Interestingly, it was also observed that SNM1B depleted cells displayed increased fragile site instability, both spontaneously, and following aphidicolin treatment (Mason et al., 2013). The data are supportive of a role for SNM1B in relieving replication stress, although the precise mechanisms are not clear. It has been hypothesised that SNM1B, in cooperation with other nucleases, is able to process nascent DNA on the lagging strand, thus generating ssDNA regions that can then be bound by FANCD2 and BRCA1 (Mason et al., 2013, Schmiester and Demuth, 2017).

SNM1B is also implicated in the activation of cell cycle checkpoints. Experimental data suggest that SNM1B is required for the G_2/M checkpoint in response to IR induced DNA damage, and for the S phase checkpoint in response to MMC treatment (Bae et al., 2008, Demuth et al., 2008). SNM1B depleted HeLa cells also exhibit a defective prophase checkpoint, and this was shown to be mediated by the interaction between SNM1B and the microtubule binding protein, astrin (Liu

et al., 2009). Correspondingly, phosphorylation of ATM and ATR, and CHK1 and CHK2 was disrupted in some SNM1B depleted cell lines, in response to DNA damaging agents. Subsequent to MMC treatment, SNM1B depleted HEK293T cells showed attenuated phosphorylation of ATM and CHK2, whilst CHK1 was unaffected. CHK2 phosphorylation, in these cells, was not affected after treatment with IR (Bae et al., 2008). Similarly, SNM1B depletion in HCT116 cells did not affect CHK1 phosphorylation after aphidicolin treatment (Mason et al., 2013). SNM1B depleted GM00637 cells also exhibited decreased ATM autophosphorylation and phosphorylation of downstream target proteins (such as p53, H2AX, and SMC), in response to exposure to IR (Demuth et al., 2008). In contrast to the above results, Anders *et al.* reported diminished CHK1 phosphorylation in SNM1B depleted GM00637 cells after treatment with UV-C (Anders et al., 2009). Together, these results suggest that SNM1B is involved in ATM signalling in response to DNA damage, and perhaps also has a role in ATR signalling. Given that ATR is the primary kinase involved in ICL repair, it therefore suggests that SNM1B either functions downstream, or independently of this step. Nevertheless, ATM also functions in ICL repair and phosphorylates a number of substrates (including FANCD2 (Taniguchi et al., 2002)), and there is a significant amount of crosstalk and overlap between the two pathways.

SNM1B also interacts with a number of other factors known to be involved in DNA damage repair pathways. SNM1B was shown to co-immunoprecipitate with PSF2 from 293T cells (Stringer and Counter, 2012). PSF2 is a member of the GINS complex, necessary for DNA replication and initiation. Interestingly, depletion of PSF2 sensitises cells to ICL inducing agents (Tumini et al., 2011). The interaction between PSF2 and SNM1B also stimulated the interaction between MUS81 and

SNM1B, suggesting they may collaborate together in the replication-coupled repair of DNA damage lesions (particularly ICLs). SNM1B was also shown to interact with HSC70, HSP70, HSP72, HSP60, and class II β-tubulin via tandem affinity purification from HEK293T cells. The interaction with HSP72 was also required for the correct localisation of SNM1B to telomeric ends (Anders et al., 2009). HSC70, HSP70, HSP72, and HSP60 are all heat shock proteins, which act as molecular chaperones to facilitate correct synthesis and folding. There is also evidence that the HSP70 proteins function within DNA damage repair pathways (Niu et al., 2006, Yamashita et al., 2007).

Together, the data strongly support a role for SNM1B in DNA damage repair, particularly the repair of ICLs. It's possible that SNM1B functions within the FA pathway – particularly given its functional relationship with FANCD2 and SLX4. At this stage, it seems likely that SNM1B functions in the upstream processing steps of stalled replication forks, either due to ICLs, or other lesions/stressors. Perhaps SNM1B, through its intrinsic nuclease activity, contributes to replication fork collapse (in conjunction with other nucleases), which leads to the activation of ATM regulated cell cycle arrest. However, the molecular function of SNM1B at these lesions is not yet clear. There are also some incongruities in the current literature – the role of SNM1B in ATM and ATR phosphorylation, and the role of SNM1B in homology directed repair (Bae et al., 2008, Mason and Sekiguchi, 2011) are two examples, and these may be reconciled by further experimentation.

1.6.5. The role of SNM1B in human disease

It is clear that SNM1B is essential for normal cellular functioning – loss of SNM1B is incompatible with life in mice; and a truncated version of SNM1B, found in a patient with HH syndrome, resulted in severe bone marrow failure, developmental defects, and premature death (Touzot et al., 2010, Akhter et al., 2010). To date, there have been no identified SNM1B null alleles in humans. However, polymorphisms in the genetic sequence of SNM1B have been shown to affect cancer risk in genome wide association studies (GWAS) (Michailidou et al., 2013, Caswell et al., 2015, Liang et al., 2012, Karami et al., 2016).

A meta-analysis of GWAS, in 2013, identified an SNP (rs11552449) within the SNM1B gene that was associated with increased breast cancer risk (with genome-wide significance) (Michailidou et al., 2013). This SNP (rs11552449) was also identified in another GWAS, where it was significantly associated with breast cancer risk. However, this was also confounded by the observation that this SNP was associated with alternative splicing of SNM1B, and the neighbouring protein *PHTF1* (Caswell et al., 2015). Another SNP (rs12046289) was identified in SNM1B, which showed suggestive association (gene specific p = 0.0006) with cutaneous malignant melanoma, after Bonferroni correction (Liang et al., 2012). Karami *et al,* identified two novel loci in SNM1B: rs974404, which is associated with increased SNM1B expression, and was associated with prostate and lung cancer risk; and rs12144215, which was associated with colorectal, breast, and ovarian cancer risk (Karami et al., 2016). Of the four SNPs identified in SNM1B that mediate cancer risk only one, rs11552449, is in a coding region, and results in an amino acid change – H61Y. This substitution is within the MBL domain of

SNM1B, and is adjacent to the putative active site in the SNM1B crystal structure (Allerston et al., 2015). Strikingly, this polymorphism is also relatively common, with an allele frequency of 17% in the study population (Michailidou et al., 2013). However, the effect of this variant on the DNA binding and nuclease activity of SNM1B has not yet been elucidated.

There is also some evidence that oncogenic activation may result in increased or deregulated activity of SNM1B. In hepatocellular carcinoma cells, overexpression of HULC (highly upregulated in liver cancer) and MALAT1 (a highly conserved long non-coding RNA) alters the promoter region of TRF2, resulting in its overexpression and also activation by post-translational modifications. This stimulates the formation of shelterin, and recruitment of SNM1B to telomeres, thus stabilising telomeric ends. In these cells, hTERT activity was upregulated and thus upregulated recruitment and activity of SNM1B (and other shelterin proteins) was essential for protecting the chromosome ends, allowing uncontrolled cell growth, and preventing cellular apoptosis or senescence (Wu et al., 2016). Similar results were observed by An *et al*. where expression of IKKα and IKKβ (constituent components of the IKK protein kinase) in liver cancer cells, enhanced the recruitment and of telomeric factors (including SNM1B), increased TERRA expression, and promoted hTERT activity (An et al., 2016).

The dual cellular function of SNM1B makes it a potential therapeutic target. Firstly, the recruitment of SNM1B to telomeres is increased in some liver cancer cell lines, and this helps to sequester and protect telomeric ends in the presence of upregulated telomerase activity. Inhibition of SNM1B in this case may result in telomeric deprotection, induction of a telomeric dysfunction phenotype, and

increased cell death or senescence. Secondly, SNM1B depleted cells exhibit sensitivity to ICL inducing chemotherapeutics, thus suggesting that SNM1B might be an appropriate target for cancer chemosensitisation. Accordingly, research by Dr Sook Yee Lee (WIMM) found that cephalosporins are competitive inhibitors for both SNM1A and SNM1B and this avenue may be further explored in target validation studies (Lee et al., 2016).

1.7. The Aims of the current project

It is clear that SNM1B is critically important for genome maintenance; however, many facets of SNM1B function are not well understood. With regards to biochemical and structural studies, it is not clear how the structure of SNM1B facilitates DNA binding, nuclease activity, and confers substrate selectivity. Understanding this is important for understanding how SNM1B might bypass DNA adducts *in vitro* and *in vivo*, and it may also provide information to guide inhibitor design. The handful of genetic studies in the last four years, have provided important genetic information regarding SNPs in SNM1B and the relationship to disease. However, the biochemical and biophysical impacts of the H61Y breast cancer associated SNP variant have not been established, and these may have important *in vivo* connotations. The cellular role of SNM1B is still not well elucidated, and this also requires careful further analysis. In addition to the work on SNM1B, during the course of this study, an opportunity arose to investigate a novel factor involved in double strand break repair. The resection that occurs prior to strand invasion in HR is a well-described process, although the molecular steps are not fully understood. Thus, the aims of this study are as follows:

1. Perform a biochemical characterisation of WT and H61Y SNM1B.

2. Obtain a crystal structure of WT and H61Y SNM1B, and a co-crystal structure of WT SNM1B with a DNA substrate or substrate analogue.

3. Generate an SNM1B disrupted human cell line and perform a basic characterisation as a tool for further cellular analysis.

4. Identify the biochemical function of a novel DSB repair nuclease, EXD2.

2. Methods and Materials

2.1. Cloning into baculovirus for WT and mutant SNM1B

2.1.1. Site directed mutagenesis PCR

Mutant SNM1B gene constructs (D35A/H36A and H61Y) were generated using site-directed mutagenesis PCR. Primers (Invitrogen) were designed to flank the region of interest and to contain the desired sequence mutation(s) (see Appendix I for primer sequences). *Pfu Ultra II* high-fidelity polymerase (Agilent, UK) was used, and the reaction mixture consisted of 250 µM each dNTP, 100 ng DNA template, 5 µM each primer, 1 µL *Pfu Ultra II* Fusion Polymerase, in the provided buffer, in a total volume of 50 µL. Cycling parameters were as follows: 95^{o}C for 2:00 min, followed by 30 cycles of: 95^{o}C for 0:20 sec, annealing at T_m – 5^{o}C for 0:20 sec and extension at 72^{o}C for 8:00 min; the reaction was then held at 72^{o}C for 10 min before returning to 4^{o}C. All PCRs were carried out in a PTC-200 Peltier Thermal Cycler (MJ Research, Bio-Rad Laboratories).

2.1.2. DpnI digest

To cleave parental methylated DNA, the PCR product was treated with 10 units of *DpnI* (New England Biolabs, NEB) and incubated at 37^{o}C for 1 hour.

2.1.3. Transformations into chemically competent XL1-Blue *E. coli*

Plasmids were transformed into chemically competent XL1-Blue *E. coli.* 200 µL of XL1-Blue cells were thawed on ice, 20–100 ng DNA was added, gently mixed, and returned to ice for 30 minutes. Cells were heat shocked at 37^{o}C for 2 minutes, before 200 µL LB-media was added, and cells were incubated, shaking, at 37^{o}C for 1 hour. The transformed cells were then plated onto LB-ampicillin (100 µg/mL) agar plates, and incubated overnight at 37^{o}C.

2.1.4. Plasmid DNA purification

Single colonies were selected from the transformed plates and used to inoculate 5 mL of LB-ampicillin broth (100 µg/mL), and cultures were grown overnight, shaking, at 37^{o}C. Isolation of the plasmid DNA from 5 mL cultures was performed using Plasmid Purification Mini Kits (Qiagen).

2.1.5. PCR to add extensions for LIC

Primers (Invitrogen, UK) were designed to add 15 (for cloning into pFB-CT10HF-LIC) or 17 (for cloning into pFB-LIC-BseR1) nucleotide extensions to provide the overlap for specific annealing into the appropriate vector via LIC (Applendix I). PCRs were performed using *Herculase II* Fusion DNA Polymerase (Agilent Technologies, UK) and a touchdown PCR protocol was used. The reaction mixture consisted 200 µM each dNTP, 2.5 units of *Herculase II*, 6.25 ng DNA template, 100 ng of each primer, either with or without 3% DMSO, in the appropriate

reaction buffer, made up to 25 µL. PCR cycling parameters were as follows: 95°C for 10:00 min; (95°C for 0:30 sec, 68°C for 0:30 sec, 68°C for 1:30 min) x 5 cycles; (95°C for 0:30 sec, 60°C for 0:30 sec, 68°C for 1:30 min) x 5 cycles; (95°C for 0:30 sec, 55°C for 0:30 sec, 68°C for 1:30 min) x 5 cycles; (95°C for 0:30 sec, 50°C for 0:30 sec, 68°C for 1:30 min) x 5 cycles; (95°C for 0:30 sec, 50°C for 0:30 sec, 68°C for 1:30 min) x 15 cycles. PCR products were analysed by agarose gel electrophoresis (as below). PCR products were purified using PureLink PCR Purification Kit (Invitrogen) before downstream use.

2.1.6. T4 polymerase treatment of PCR products

PCR products were treated with T4 Polymerase to generate the overhangs for LIC. Reactions contained 2.5 mM dGTP (for cloning into the pFB-CT10HF-LIC vector) or 2.5 mM dCTP (for cloning into the pFB-LIC-Bse vector), 5mM DTT, 1% BSA, 1.5 U T4 DNA Polymerase (Novagen) and 5 µL DNA PCR product, in the manufacturer's supplied buffer. Samples were incubated for 30 min at 22°C and 20 min at 75°C.

2.1.7. Restriction digest of the transfer vector

To generate the complementary overhangs for correct annealing, the transfer vectors were treated with the appropriate restriction endonuclease, followed by T4 polymerase treatment. For restriction digest of pFB-LIC-Bse-R1, 5 µg of plasmid was incubated with 31.25 U of BseR1 (NEB), in a total volume of 100 µL, for 2 hours at 37°C. For pFB-CT10-HF, 5 µg of plasmid was incubated with 5 U of

BfuA1 (NEB), in a total volume of 50 µL for 2 hours at 50°C. Plasmid DNA was subsequently purified using a PureLink PCR purification kit (Invitrogen).

2.1.8. T4 polymerase treatment of the transfer vector

Digested and purified plasmid products were treated with T4 polymerase. Reactions contained 50 µL of plasmid DNA, 1% BSA, 5 mM DTT, 15 U T4 DNA Polymerase (Novagen), and 2.5 mM dGTP (for pFB-LIC-BseR1), or 2.5 mM dCTP (for pFB-CT10-HF). Samples were incubated at 22°C for 30 minutes, then 75°C for 20 minutes.

2.1.9. Annealing into the appropriate transfer vector

The appropriate vector (either pFB-CT10HF-LIC or pFB-LIC-Bse) was mixed with the corresponding treated insert at each a 1 µL:1 µL, 1:2, 1:3 and 1:4 volume ratio (total volume, 3 µL), incubated at 37°C for 10 minutes, and room temperature for 20 min.

2.1.10. Transformation into chemically competent MACH-1 *E. coli*

Annealed vector and insert products were transformed into chemically competent MACH-1 *E. coli.* Samples were mixed with 40 µL of chilled competent cells, incubated on ice for 30 minutes, then heat-shocked at 42°C for 45 seconds. Samples were briefly returned to ice, 100 µL of SOC medium was added, and

incubated, shaking, at 37°C for 1 hour. Cultures were plated onto LB-agar plates, containing 5% sucrose and 200 µg/mL ampicillin, and incubated overnight at 37°C.

2.1.11. Colony PCR screen

Transformed colonies were screened by PCR. Using a 1 µL loop, a single colony was used to inoculate the PCR mix, then immediately to inoculate 1 mL of LB-ampicillin broth. PCRs contained 0.5 µM each primer (see Appendix I) and 0.1 µL MyTaq DNA Polymerase, in MyTaq Reaction Buffer (Bioline), in a total volume of 20 µL, Cycling parameters were 94°C for 10 minutes, 25 cycles of 94°C for 30 seconds, 50°C for 30 seconds, 72°C for 30 seconds, then a final extension of 72°C for 5 minutes. PCR products were analysed by agarose gel electrophoresis and the appropriate plasmid DNA was purified using the PureLink Quick Plasmid Miniprep Kit (Invitrogen). Purified plasmids were sequenced by Source Bioscience (Oxford); primers are indicated in Appendix II.

2.1.12. Agarose gel electrophoresis

Agarose gel electrophoresis was used as a tool to validate a number of steps in the cloning process. Samples were run on a 1% agarose (w/v) gel in 1xTAE (0.04 M Tris-acetate, 0.001 M EDTA), with the addition of ethidium bromide (1 µg/mL). Five parts DNA product were mixed with one part 6 x loading dye (Fermentas) and loaded along with a 1 kb Plus DNA ladder (Invitrogen) as a size reference. Electrophoresis was performed at 100 V for 40 minutes. The AlphaImager HP

Transilluminator (Alpha Innotech) was used to visualize DNA bands under UV light, and gel images documented.

2.1.13. Bacmid generation by transposition

The transfer vectors containing the SNM1B constructs were transposed into the bacmids for viral infection. 3 µL of purified recombinant DNA was added to 30 µL of DH10Bac *E. coli* (Life Technologies) and incubated on ice for 30 minutes before being heat-shocked at 42^{o}C for 45 seconds. 900 µL pre-warmed SOC medium (containing 50 µg/mL kanamycin, 10 µg/mL tetracycline and 1 µg/mL gentamycin) was added to each sample and incubated for 5 hours, at 37^{o}C, shaking at 700 rpm. 50 µL of the transformed DH10Bac cells were spread on LB-agar plates (containing 50 µg/mL kanamycin, 10 µg/mL tetracycline and 1 µg/mL gentamycin) and incubated at 37^{o}C for 48 hours, protected from light. After 48 hours white colonies were selected, streaked to dilution, and incubated at 37^{o}C overnight. The recombinant white colonies were used to inoculate 2 mL of 2xLB broth (containing 50 µg/mL kanamycin, 10 µg/mL tetracycline, and 1 µg/mL gentamycin), and incubated overnight at 37^{o}C, shaking at 700 rpm. Bacmid DNA was purified using a PureLink HiPure Plasmid DNA Purification Kit (Invitrogen).

2.1.14. Bacmid PCR screen

Purified bacmid DNA was diluted 1 in 50, in MilliQ H_2O, and a PCR screen was performed to ensure correct transposition. The reaction mix was as described in 2.1.11 and cycling conditions were as follows: 94^{o}C for 5 minutes, followed by 30

cycles of: 94^{o}C for 45 seconds, 50^{o}C for 45 seconds, 72^{o}C for 2.5 minutes; followed by a final extension of 72^{o}C for five minutes.

2.1.15. Transfection of Sf9 insect cells

Sf9 cells were prepared one day in advance (to ensure they were mid-log phase for transfection) by diluting the cell count to 1x10^{6} cells/mL in SF-900 II medium (ThermoFisher Scientific). For transfection, 2x10^{5} mid-log phase cells were aliquoted to a 24-well sterile plate (Corning) and incubated at 27^{o}C for one hour to allow adhesion. Per well, 2 µL JetPRIME transfection reagent (Polyplus Transfection) was mixed with 50 µL JetPRIME buffer, added to 6 µL of recombinant bacmid DNA (~0.5 µg), and incubated at room temperature for 30 minutes. The JetPRIME:DNA solution was then added to the Sf9 cells at 27^{o}C for 4 hours, after which the media was replaced with 0.8 mL SF-900 II media supplemented with 2% (v/v) FCS (ThermoFisher Scientific) and 0.1% (w/v) penicillin/streptomycin. Cells were incubated at 27^{o}C for 96 hours, after which the media was removed, centrifuged at 1500xg for 20 minutes, and the supernatant was stored at 4^{o}C at the P0 virus stock. To further amplify the virus, 120 µL of the P0 virus stock was added to 6x10^{6} Sf9 cells (in 3 mL) in SF-900 II media, supplemented as previously, and incubated at 27^{o}C, shaking at 450 rpm, for 66–72 hours. Cells were then harvested by centrifugation at 1500xg for 20 minutes, and cell pellets were used for expression testing. The supernatants (P1 virus stock) were further amplified by infecting 50 mL of Sf9 cells (2x10^{6} cells/mL for 72 hours); and the P2 virus supernatant was stored at 4^{o}C in the dark, and used for preparative infections.

2.2. Purification of WT and mutant Δ-SNM1B

WT and mutant truncated (1/2–335) SNM1B proteins were purified from Sf9 insect cells. 3 mL of P2 virus stock were used to infect 1 L cultures of Sf9 insect cells ($2x10^6$ cells/mL in 5 L shaker flasks), and the culture was grown for 72 hours at 27^{o}C. 4 litres of infected cells were harvested by centrifugation in a JLA8.1 rotor (Beckmann-Coulter) at 5000 rpm for 30 minutes. The cell pellet was re-suspended in 150 mL of cold lysis buffer (50 mM HEPES, pH 7.5, 500 mM NaCl, 10 mM imidazole, 5% glycerol (v/v), 1 mM TCEP, supplemented with protease inhibitors (cOmplete, Mini, EDTA-free Protease Inhibitor Cocktail, Roche)). Cells were lysed by sonication on ice for 15 minutes (5 seconds on, 10 seconds off), before being centrifuged at 23 000 rpm for 40 minutes in a JA25.5 rotor (Beckmann Coulter) to remove cellular debris. The supernatant was mixed with 10 mL Ni-Sepharose slurry (prewashed with lysis buffer; GE Healthcare) and incubated, on a rotating mixer, at 4^{o}C for 1–2 hours. Samples were centrifuged at 7000 x g for 5 minutes, the supernatant was decanted, the Ni-Sepharose beads were washed twice with 50 mL lysis buffer, resuspended in 30 mL wash buffer (50 mM HEPES, pH 7.5, 300 mM NaCl, 45 mM imidazole, 5% glycerol (v/v), 1 mM TCEP), and added to a 25 mL gravity flow column. The column was washed with 10 column volumes of wash buffer, and eluted in 10 column volumes elution buffer (50 mM HEPES, pH 7.5, 300 mM NaCl, 500 mM imidazole, 5% glycerol (v/v), 1 mM TCEP). Protein containing fractions (as determined by absorbance at 280 nm and SDS-PAGE) were pooled, and the histidine tag was cleaved by incubation with TEV protease (at a 20:1 SNM1B:TEV mass ratio) overnight, at 4^{o}C, whilst being dialysed using a 3.5 kDa molecular weight cut-off snakeskin membrane (ThermoFisher Scientific)

into a buffer containing 50 mM HEPES, 500 mM NaCl, 5% glycerol, 1 mM TCEP. The dialysed sample was added to a 5 mL (bed volume) Ni-Sepharose column pre-equilibrated in dialysis buffer. The column was washed with buffer (as for dialysis) containing increasing amounts of imidazole (10 mM, 45 mM, 80 mM, 150 mM), before being eluted in ten column volumes of buffer containing 300 mM imidazole. SNM1B proteins were further purified by size exclusion chromatography using an AKTA Express S200 Superdex column (GE, LifeSciences) in a buffer containing 50 mM HEPEs, pH 7.5, 500 mM NaCl, 5% glycerol, 1 mM TCEP. At each aforementioned step, samples were analysed for the presence of SNM1B by SDS-PAGE and verified by ESI-TOF-MS. For ESI-TOF-MS (electrospray-ionisation time-of-flight mass-spectrometry), 1 µg of protein was diluted in 60 µL 0.1% formic acid (v/v) prior to injection. All MS was performed with the assistance of the Mass Spectrometry Facility, SGC, Oxford. Δ-SNM1B protein concentration was determined by denaturation in 6 M guanidinium hydrochloride, 50 mM sodium phosphate, pH 7.5, and absorbance measurement at 280 nm (Nanodrop Spectrophotometer), using the calculated molecular mass and extinction coefficient.

2.2.1. SDS-PAGE to determine the presence of WT and mutant SNM1B

NuPAGE LDS Sample Buffer (Novex; Life Technologies) and 250 mM DTT were added to protein containing samples, and boiled at 95^{o}C for five minutes. SDS-PAGE (sodium dodecyl sulphate polyacrylamide gel electrophoresis) was performed with NuPAGE 4–2% Bis-Tris Protein Gels (Invitrogen) with 1xNuPAGE MES SDS Running Buffer (ThermoFisher), at 150 V for 40 minutes. A Precision

Plus Unstained Protein Marker (BioRad) was used as an indicator of molecular weight. Gels were stained with InstantBlue Protein Stain (Expedeon) to visualise protein bands.

2.3. Biochemical characterisation of WT and mutant SNM1B

The sequences of the oligonucleotide substrates used in this study are listed in Appendix V.

2.3.1. Generation of 3' radiolabelled DNA substrates

100 pmol of single-stranded DNA (Eurofins MWG Operon, Germany) were labelled with 3.3 pmol of α-^{32}P-dATP (Perkin Elmer) by terminal deoxynucleotidyl transferase (Tdt, 20 U; ThermoFisher Scientific), incubated together at 37^{o}C for 1 hour. This solution was passed through a P6 Micro Bio-Spin chromatography column (BioRad), and the radiolabelled DNA was annealed with the appropriate unlabelled oligonucleotides (1:3 molar ratio of labelled to unlabelled oligonucleotide) (see Appendices V and VI) by boiling at 95^{o}C and cooling to below 30^{o}C in annealing buffer (10 mM Tris-HCl; pH 7.5, 100 mM NaCl, 0.1 mM EDTA). Correct annealing was verified by 10% non-denaturing polyacrylamide gel electrophoresis.

2.3.2. Gel-based nuclease assay

Standard exonuclease assays were carried out in 10 µL reactions containing, 20 mM HEPES-KOH, pH 7.9, 50 mM KCl, 0.5 mM DTT, 10 mM $MgCl_2$, 0.05% Triton X-100, 0.1 mg/mL BSA, 5% glycerol, and the indicated amounts of SNM1B. The reactions were started by the addition of DNA substrate (100 nM), incubated at 37^{o}C for the indicated time period, and quenched by the addition of 2 µL stop solution (95% formamide, 10 mM EDTA, 0.25% xylene cyanol, 0.25% bromophenol blue) and boiling at 95^{o}C for 3 minutes.

Reactions were analysed by 20% denaturing polyacrylamide gel electrophoresis (40% solution of 19:1 acrylamide:bis-acrylamide, BioRad) and 7 M urea (Sigma Aldrich)) in 1 x TBE (Tris-borate EDTA) buffer. Electrophoresis was carried out at 525 V for 1.5 hours; gels were subsequently fixed for 40 minutes in fixing solution (50% methanol, 10% acetic acid), and dried at 60^{o}C for four hours under a vacuum. The dried gels were exposed to a Kodak phosphorimager screen, and scanned using a Typhoon 9400 instrument (GE).

2.3.3. Electrophoretic mobility shift assay

Electrophoretic mobility shift assays (EMSAs) were carried out in 10 µL reactions, containing 20 mM HEPES, pH 7.9, 50 mM KCl, 10, 0.5 mM DTT, 0.05% Triton-X, 0.1 mg/mL BSA, 5% glycerol, and the indicated amount of SNM1B. The reaction was started with the addition of DNA substrate (1 nM) and incubated at 37^{o}C for 5 minutes before being returned to ice and 2 µL loading dye was added (50%

glycerol, 1 mg/mL bromophenol blue (Sigma Aldrich), 1 mg/mL xylene cyanol (Sigma Aldrich). Reactions were analysed by 10% polyacrylamide native gel electrophoresis in 0.5 x TBE at 100V for 3 hours, at 4^{o}C. The gel was fixed in fixing solution for 15 minutes, and dried under vacuum for 2 hours at 80^{o}C. Dried gels were exposed to a phosphorimager screen and scanned as above.

2.3.4. Real-time fluorescence nuclease assay

Real-time fluorescence nuclease assays were carried out as previously described ((Sengerova et al., 2012) and Appendix IV). Reactions were carried out in black 384-well microplates (Corning, USA) with a total volume of 15 µL per well. To investigate enzyme kinetics, reactions contained between 10 and 2500 nM DNA substrate, 1.0 nM SNM1B, in a reaction buffer containing 20 mM HEPES, pH 7.9, 50 mM KCl, 10 mM $MgCl_2$, 0.5 mM DTT, 0.05% Triton-X100, 0.1 mg/mL BSA, 5% glycerol. The reaction was started by the addition of Δ-SNM1B and the emission spectra were measured (excitation at 495 nm, emission at 525 nm, cutoff at 515 nm) with 6 readings taken at 7-second intervals for 6 minutes. All measurements were made using a SpectraMax M2e fluorescent plate reader, using SoftMaxPro software (Molecular Devices) to control the settings. The fluorescence intensity of each well was plotted against time, and the rate of increase in fluorescence was determined. This rate was taken to represent enzymatic activity, which was plotted against DNA concentration. Michaelis-Menten enzyme kinetics were calculated using Prism software (GraphPad Software Inc., La Jolla, CA, USA).

2.3.5. Fluorescence polarisation DNA binding assay

Reactions (total volume 30 µL) were carried out in black 384 well plates (Corning, USA). SNM1B was serially diluted two-fold from 1 µM to 0.288 nM, and incubated with 10 nM fluorescein-labelled DNA substrate in buffer containing 20 mM HEPES, pH 7.9, 50 mM KCl, 0.5 mM DTT, 0.01 mg/mL BSA, 5% glycerol, at room temperature for 5 minutes. Fluorescence polarisation was measured with excitation at 285 nm and emission at 520 nm using a PHERAstar FSX plate reader (BMG Labtech). Data were fitted using a sigmoidal curve in GraphPad Prism Software.

2.3.6. Thermal stability assay

1 µM Δ-SNM1B was mixed with 1 µL SPYRO Orange Protein Gel Stain (diluted from 5000x stock solution) (Life Technologies), in a total volume of 20 µL, in a buffer containing 20 mM HEPES, pH 7.9, 50 mM KCl, 5% glycerol, 0.5 mM DTT. Fluorescence readings (excitation 470 nm/emission 570 nm) were taken in triplicate between 25^{o}C and 95^{o}C and the temperature was increased linearly at 1^{o}C/min. Melting curves for each dataset were exported to GraphPad Prism Software and a Baltzmann curve was fitted to determine the melting temperature.

2.4. Crystallisation of WT and H61Y SNM1B

2.4.1. Crystallisation of apo-form WT and H61Y Δ-SNM1B

For crystallisation of the apo form of WT and H61Y Δ-SNM1B, the proteins were concentrated to 15 and 12 mg/mL, respectively, in a 30 000 molecular weight cut-off (MWCO) centrifugal concentrator (Amicon Ultra Centrifugal Filter, Merck). Crystallisation was achieved by sitting drop vapour diffusion, at 4^{o}C, in conditions containing 6% PEG 20 K, 0.1 M bicine, pH 7.3 for WT Δ-SNM1B; and 6% PEG 3350 K, 0.1 M HEPES, pH 7.0, 0. 2 M sodium formate for H61Y Δ-SNM1B. Crystals were cryoprotected in the reservoir solution supplemented with 20% ethylene glycol, and plunged into liquid nitrogen.

2.4.2. Co-crystallisation of WT Δ-SNM1B with 2'-deoxy-5'-AMP

WT Δ-SNM1B (at 15 mg/mL) was incubated with 50 mM 2'-deoxy-5'-AMP (Sigma Aldrich; adjusted to pH 7.5 with 1 M NaOH) for 30 minutes on ice. Crystallisation was achieved by sitting drop vapour diffusion at 20^{o}C, in conditions containing: 20% PEG 3350, 0.2 M ammonium chloride. Crystals were cryoprotected and stored, as above (2.4.1).

2.5. Analysis of the cellular role of SNM1B

2.5.1. Generation of plasmids for CRISPR-Cas9

Generation of plasmids for CRISPR-Cas9 transfection was performed as outlined in (Ran et al., 2013). Guide oligonucleotides were designed using the CRISPR design tool (Zhang Lab, MIT, 2015) and sequences are displayed in Appendix III. Guide oligonucleotides (Invitrogen) were phosphorylated and annealed in a reaction containing: 10 µM each top and bottom guide sequences, 10 U polynucleotide kinase (PNK, NEB) in T4 ligation buffer (NEB). Reactions were incubated at 37^{o}C for 30 minutes, 95^{o}C for 5 minutes, and ramped down to 25^{o}C at 5^{o}C/min. Phosphorylated and annealed oligonucleotides were diluted 1 in 200 in MilliQ H_2O.

100 ng of the pX330 and pX458 CRISPR plasmids (a kind gift from Dr Philip Hublitz, WIMM) were incubated with 10 U BbsI-HF (NEB) in the appropriate buffer for 1 hour at 37^{o}C, before being dephosphorylated by incubation with 1 U shrimp alkaline phosphatase (rSAP; NEB) for 1 hour at 37^{o}C. Digested and dephosphorylated plasmids were purified by agarose gel electrophoresis (2.1.12) and gel bands were purified using the QIAquick Gel Extraction Kit (Qiagen). Diluted phosphorylated and annealed oligonucleotides were ligated with the appropriate vector (either pX330 or pX458) at 1:3 vector to insert molar ratio, by overnight incubation at 16^{o}C with 1 µL T4 DNA ligase (NEB) in the manufacturer's buffer. 1–5 µL of the ligated reaction were transformed into competent XL1-Blue *E. coli* (2.1.3), and plasmid DNA was purified (2.1.4). To ensure that ligation was correct, sequencing was performed by Source Biosciences (Oxford, UK).

2.5.2. Culture of HeLa1.2.11 cells

HeLa1.2.11 cells (a kind gift from Professor Madalena Tarsounas, Oxford) were cultured in tissue culture flasks (Corning, USA) containing GIBCO Dulbecco's Modified Eagle's Medium (DMEM; Invitrogen), supplemented with 20% foetal calf serum (FCS; Invitrogen), and incubated at 37^{o}C in a 5% CO_2 humidified atmosphere in a Heraeus cell culture incubator (ThermoFisher Scientific). Cells were subcultured by trypsinisation with 0.05% Trypsin/EDTA (ThermoFisher Scientific) when cultures reached 80–90% confluency.

2.5.3. Transfection and sorting of CRISPR cells

pX330 and pX458 CRISPR plasmids containing the appropriate guide sequences were concurrently transfected into HeLa1.2.11 cells. Lipofectamine 2000 Transfection Reagent (ThermoFisher Scientific) was used, as per the manufacturer's protocol. One day prior to transfection, HeLa1.2.11 cells were plated to a 6-well tissue culture dish (Corning, USA), and allowed to adhere overnight. Cells were transfected at ~80% confluency: 1 µg each plasmid DNA was mixed with 100 µL GIBCO-OptiMEM media (ThermoFisher Scientific), and 5 µL Lipofectamine 2000 was mixed with 100 µL GIBCO-OptiMEM media; these were each incubated for 5 minutes at room temperature, before being mixed together and incubated for a further 20 minutes at room temperature; 200 µL of the transfection complex was then added, drop-wise to the assigned wells. Cells were incubated at 37^{o}C for 24 hours, before being harvested by trypsinisation (as in 2.5.2). Cells were subsequently single-cell sorted, for those fluorescing both

GFP and Ruby positive, into a 96-well plate (Corning, USA) containing 100 µL DMEM. Cell sorting was performed with the assistance of the Flow Cytometry Facility, WIMM, Oxford. Single cells were cultured in 96-well plates until 80% confluency was reached, when they were passaged to 24-well plates (Corning, USA).

2.5.4. PCR screen of CRISPR clones

Genomic DNA (gDNA) was extracted from confluent cells growing in 24 well plates using the DNeasy Blood and Tissue Kit (Qiagen), and DNA concentration was quantified using a NanoDrop-1000 Spectrophotometer (ThermoFisher Scientific). gDNA was analysed in a PCR screen with the following reaction mixture: 0.2 mM dNTP mix, 0.5 µM each primer (Appendix I), between 10 ng and 1 µg gDNA (100 ng was always used initially), and 0.5 U Q5-High Fidelity Polymerase (NEB), in a reaction volume of 25 µL Cycling parameters were: 98^{o}C for 30 seconds, followed by 35 cycles of: 98^{o}C for 10 seconds, 62^{o}C for 30 seconds, 72^{o}C for 30 seconds; a final extension of 72^{o}C for 2 minutes. PCR products were subsequently analysed by agarose gel electrophoresis (2.1.12) for genetic disruption. Clones that showed potential genetic disruption were sequenced by Source Bioscience (Oxford).

2.5.5. Determining mRNA expression by quantitative real-time PCR

Total RNA from cells was purified using Qiagen RNeasy kit and concentration was quantified by NanoDrop Spectrophotometry. Equal amounts of RNA (1 µg) was reverse transcribed to cDNA using the High-Capacity cDNA Reverse Transcription

(RT) Kit (Applied Biosystems) in a reaction containing 4 mM dNTP mix, 1xRT random primers, 1 µL MultiScribe RT, in the manufacturer's buffer conditions, in a total volume of 20 µL. This solution was incubated at 25^{o}C for 10 minutes, 37^{o}C for 120 minutes, and 85^{o}C for 5 minutes.

The mRNA levels of genes of interest were calculated by quantitative real-time PCR (qRT-PCR), using the appropriate primers listed in Appendix I, SYBR dye, and hot-start Taq-Polymerase (SensiMix SYBR and Fluoroscein Kit, Bioline). PCR mixes were pipetted by the automated pipetting system epMotion 5070 (Eppendorf), and reactions were carried out in an Applied Biosystems 7500 real-time PCR system. Data were analysed using the comparative CT method.

2.5.6. Storage of genotoxic drugs

Cisplatin (Ebewe Pharmaceuticals) was stored at 1 mg/mL, protected from light, at room temperature. For use in resazurin viability assays, mitomycin C (MMC; Sigma Aldrich) was solubilised in DMSO (dimethyl sulfoxide; Sigma Aldrich) to 15 mg/mL; and for clonogenic survival assays was dissolved to 0.5 mg/mL in MilliQ H_2O. Zeocin (Sigma Aldrich) and hydroxyurea (HU; Sigma Aldrich) were each solubilised and stored in MilliQ H_2O at 70 mM and 100 mM, respectively. Etoposide (Sigma Aldrich) and camptothecin (Sigma Aldrich) were solubilised in DMSO to 50 mM and 10mM, respectively. Formaldehyde (Sigma Aldrich) was stored as a 37% (w/v) solution in H_2O. SJG-136 (a kind gift from Professor John Hartley, UCL) was solubilised in DMSO, at a concentration of 10 mM. All drugs

were diluted in DMEM immediately prior to use in assays, and final DMSO concentration was kept below <2%.

2.5.7. Resazurin viability assays

All assays were carried out in black 96-well microplates (Corning). 2000 cells/well were seeded per well and allowed to adhere overnight. Drugs were added at the indicated concentrations, and incubated for various times: cisplatin was incubated for 4 hours at 37^{o}C, and MMC for 1 hour, after which the media was removed and replaced with fresh DMEM. All other drugs were treated continuously for the duration of the assay. After 72 hours, media was removed; cells were washed with PBS-A, 100 µL of complete, GIBCO phenol-free DMEM (ThermoFisher Scientific) and 20 µL of 0.15 mg/mL resazurin (Sigma Aldrich) was added per well, and incubated at 37^{o}C for 1 hour. The amount of resorufin was subsequently quantified by measuring fluorescence spectra at 560 nm excitation / 590 nm emission. Cell survival (as determined by the absorbance of resorufin) was calculated as a percentage of untreated cells, and graphed in GraphPad Prism.

2.5.8. Clonogenic survival assays

1000 cells were plated, in 10 mL media, to 10 cm tissue culture dishes (Corning). Cells were allowed to adhere, and return to the log phase of cell growth overnight. Cells were then treated with 10 mL of media containing different dosages of drug after the plating media was removed. MMC and formaldehyde were incubated for 1 hour, and cisplatin for 4 hours (after which the drug containing media was removed, and DMEM was added); all other drugs were treated continuously, and

the media was not changed. Cells were incubated at 37°C for 10–12 days to allow colony formation. At the end of this period, media was removed; the plates were rinsed with PBS-A, before staining in Coomassie Brilliant Blue solution (0.1% Brilliant Blue R-250; Sigma Aldrich, 7% acetic acid, 50% methanol). Residual dye was removed by rinsing with MilliQ H_2O, and plates were allowed to air-dry. Colonies were counted using a ColCount Colony Counter (Oxford Optronix). Data were exported, expressed as a proportion of untreated cells, and graphed in GraphPad Prism.

2.6. Purification of WT and D108A/E110A Δ-EXD2

WT and D108A/E110A Δ-EXD2 were purified from BL21(DE3)-R3-pRARE1 *E. coli.* 40 mL starter cultures were grown overnight, in Terrific Broth (TB) supplemented with 50 µg/mL each chloramphenicol and kanamycin. These were used to inoculate 4 L of TB (supplemented to 50 µg/mL kanamycin), which were grown at 37°C, shaking, until the OD (A600 nm) was 1.5. The cultures were then chilled to 18°C, 0.5 mM IPTG (isopropyl β-D-1-thiogalactopyranoside) was added to induce protein expression, and cultures were further grown overnight at 18°C. Cells were harvested by centrifugation at 5000xg for 20 minutes in a JLA8.1 rotor (Beckmann Coulter). Four volumes of lysis buffer (v/w) were added to the cell pellets; buffer as in 2.2, with the addition of 0.5% Triton-X100 (Sigma Aldrich), 1 mM PMSF (phenylmethylsulfonyl fluoride; Sigma Aldrich), and 2 mM benzamidine (Sigma Aldrich). The cells were lysed by sonication (15 minutes; 5 seconds on, 10 seconds off), and PEI (polyethyleneimine) added to a final concentration of 0.15% (w/v) from a 5% stock solution (pH 7.5). Lysates were centrifuged at 16500xg for

45 minutes in a JA17 rotor (Beckmann Coulter) to remove DNA and cellular debris. The supernatant was subsequently incubated with 10 mL Ni-Sepharose slurry (5 mL bed volume, pre-equilibrated in lysis buffer) and incubated at 4^{o}C, tumbling, for 1–2 hours. The Ni-Sepharose beads were centrifuged, washed, and added to a 25 mL gravity flow column. This was performed following the same protocol as in 2.2, with the addition of 0.5% Triton-X100, 1 mM PMSF, and 2 mM benzamidine to the wash and elution buffers. The eluate was analysed by SDS-PAGE for the presence of Δ-EXD2 (WT or D108A/E110A), and relevant fractions were pooled, concentrated, and further purified by two sequential AKTA Express S200 Superdex gel filtration chromatography columns (GE, LifeSciences) in a buffer containing 50 mM HEPEs, pH 7.5, 500 mM NaCl, 5% glycerol, 1 mM TCEP, 1 mM PMSF, and 2 mM benzamidine. The presence of Δ-EXD2 was confirmed (as in 2.2) by SDS-PAGE and ESI-TOF-MS.

2.7. Biochemical characterisation of WT and D108A/E110A EXD2

2.7.1. Generation of 5’ radiolabelled substrates

10 pmol of single-stranded DNA (Eurofins MWG Operon, Germany) was incubated with 6.8 pmoles γ-^{32}P-dATP (Perkin Elmer), and 10 U T4 PNK (ThermoFisher Scientific) at 37^{o}C for 1 hour. This solution was passed through a P6 Micro Bio-Spin chromatography column (BioRad), and the radiolabelled DNA was annealed with the appropriate unlabelled oligonucleotide(s) (Appendix V), as described in 2.3.1.

2.7.2. Gel-based nuclease assay

Gel-based exonuclease assays with Δ-EXD2 (WT and D108A/E110A) were carried out as described in 2.3.2; however, the buffer contained 5 mM $MnCl_2$, instead of 10 mM $MgCl_2$, and the concentration of DNA substrate was 10 nM, unless otherwise indicated.

2.7.3. Fluorescence polarisation DNA binding assay

Fluorescence polarisation based DNA binding assays with Δ-EXD2 were performed as in 2.3.5.

3. A Biochemical Analysis of WT and mutant SNM1B

3.1. Introduction

Understanding the biochemical characteristics of WT SNM1B and the role of key amino acid residues, and regions, in the enzymatic activity of the SNM1B protein is important for elucidating how it might function *in vivo* and the role it might play in disease pathogenesis.

Over-expression of full-length SNM1B can be toxic to mammalian cells and result in premature cell death (Freibaum and Counter, 2008). Over-expression of the MBL and β-CASP domains in *E. coli* appears to be better tolerated, but results in very poor yield, even from large amounts of culture (Sengerova et al., 2012). To this end, a protocol for purifying recombinant SNM1B from Sf9 insect cells had previously been established (Allerston et al., 2015, Sengerova et al., 2012) and the truncated SNM1B protein (Δ-C SNM1B (1–335)) produced and used in these studies was shown to have robust 5' to 3' exonuclease activity (Sengerova et al., 2012). However, purification of Δ-C SNM1B (1–335) remained challenging, as the N-terminal hexa-histidine tag was difficult to proteolytically cleave and precluded efficient purification (preliminary experiments; data not shown). As such, SNM1B constructs were redesigned for cloning and purification.

In addition to WT SNM1B, two mutated versions of SNM1B were cloned and purified – D35A/H36A SNM1B, a previously established 'nuclease dead' variant (Sengerova et al., 2012) and H61Y SNM1B. As discussed in Chapter 1, this H61Y

variant of SNM1B was associated with a slightly increased risk of breast cancer in a meta-analysis of GWAS (Michailidou et al., 2013), the crystal structure (PDB: 5AHO) showed that the His-61 residue was located adjacent to the active site, and previous work showed that an H61A mutation slightly reduced catalytic activity when compared to the WT protein (Lee, 2015).

Thus, the WT, D35A/H36A, and H61Y SNM1B genes were cloned, expressed and each respective SNM1B protein was purified from Sf9 insect cells. This subsequent recombinant, purified protein was used for a number of biochemical and biophysical studies.

3.2. Results

3.2.1. Cloning of WT and mutant SNM1B

H61Y and D35A/H36A mutant SNM1B constructs were generated using a site-directed mutagenesis PCR approach, and sequencing analysis validated introduction of the correct mutation (Figure 3.1).

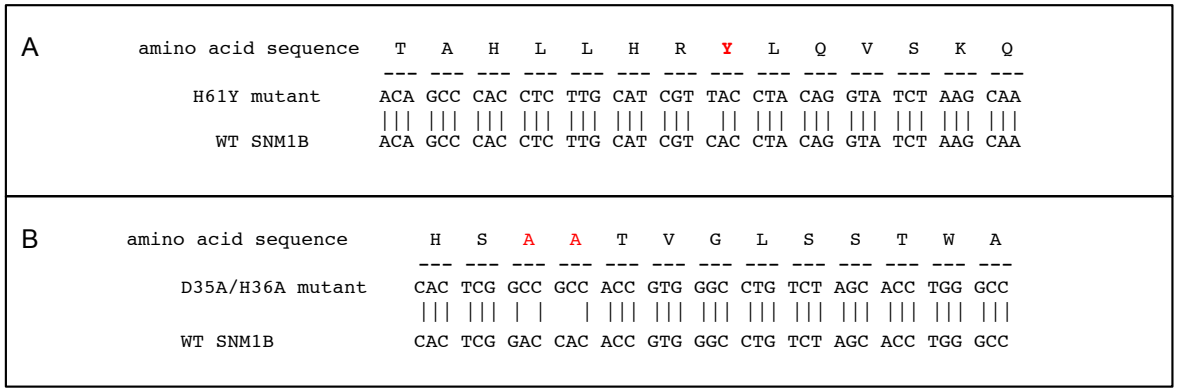


Figure 3.1 Successfully generated mutant SNM1B clones.

Sequencing results showing the desired mutations in the nucleotide sequence of the SNM1B gene and the corresponding amino acid change. Mutated amino acid residues are shown in red. A: H61Y mutation, which corresponds to C181T in the nucleotide coding sequence. B: D35A/H36A double mutation, which corresponds to A104C, C106G and A107C. Note that the entire SNM1B gene was sequenced, however only the mutation containing regions are shown.

Due to the aforementioned difficulties purifying Δ-C SNM1B (1–335) with an N-terminal 6-histidine tag, WT, D35A/H36A, and H61Y SNM1B (Δ-C SNM1B (1/2–335)) were subcloned into two different vectors in an attempt to improve purification. The cloning of these constructs into the baculovirus expression vector was performed using the method outlined by (Shrestha et al., 2008) and briefly summarised in Figure 3.2. The initial step involved the cloning of the relevant SNM1B construct into the appropriate transfer vectors.

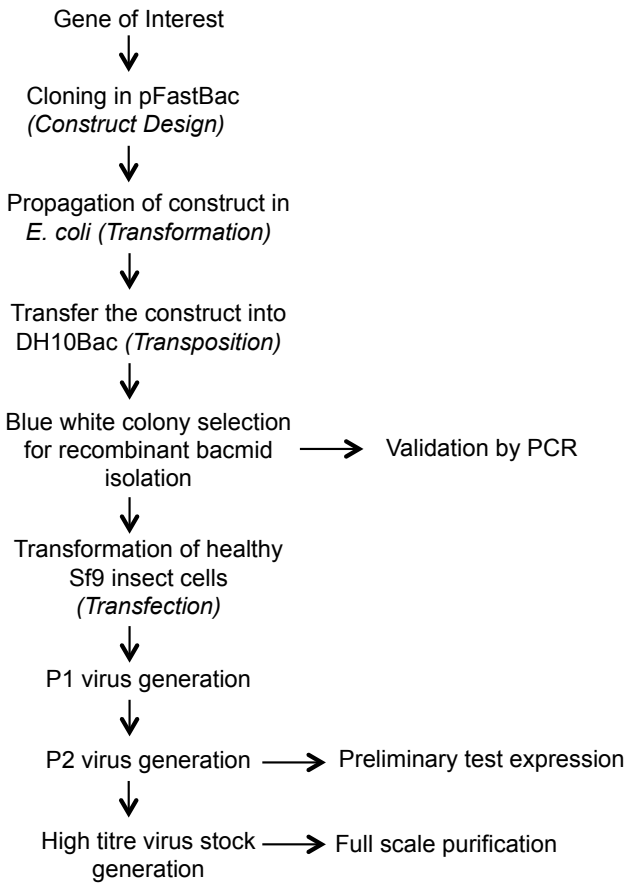


Figure 3.2 An overview of the steps involved in cloning for baculovirus expression in insect cells for the purposes of recombinant protein purification.

Adapted from Shrestha et al, 2008.

WT, D35A/H36A, and H61Y Δ-C SNM1B (1/2–335) were cloned into each of the pFB-CT10HF-LIC and pFB-LIC-BseR1 transfer vectors. Cloning into the pFB-CT10HF-LIC vector resulted in the addition of a C-terminal 10-histidine and FLAG tag to the truncated purified protein (Figure 3.3). Whereas, cloning into the pFB-LIC-BseR1 resulted in the addition of an N-terminal 6-histidine tag followed by a short glycine rich linker (Figure 3.4). Each of the tags were designed to be spaced from the SNM1B protein sequence by a tobacco etch virus (TEV) cleavage site, thus allowing efficient cleavage of the histidine tag during the purification process. The schematic for these constructs is shown in Figure 3.3 and Figure 3.4. The

WT, D35A/H36A and H61Y gene sequences were cloned into these two vectors for protein purification because it was unknown which construct would result in higher overexpression, or better purification.

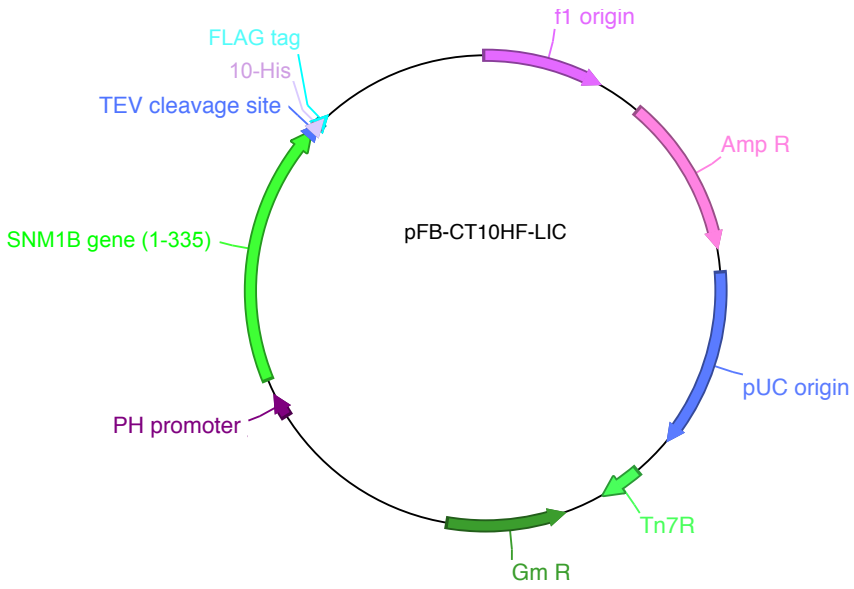


Figure 3.3 Schematic of the truncated SNM1B gene (amino acids 1–335) cloned into the pFB-CT10HF-LIC vector.

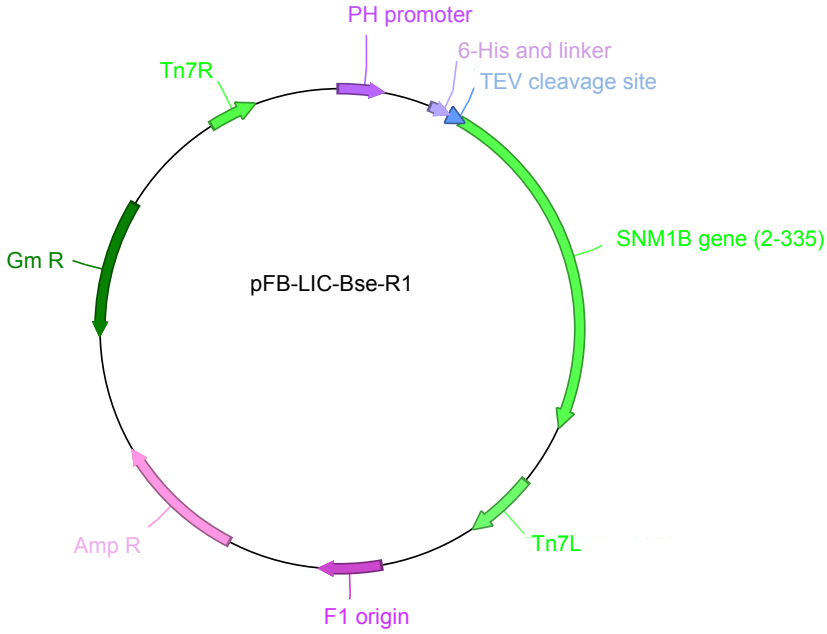


Figure 3.4 Schematic of the truncated SNM1B gene (amino acids 2–335) cloned into the pFB-LIC-Bse-R1 vector.

Cloning into the transfer vector was performed by a ligation independent cloning (LIC) protocol. As such, initial PCRs were performed with primers designed to add 15–17 nucleotide extensions that were complementary to the cloning sites in the vector (Figure 3.5).

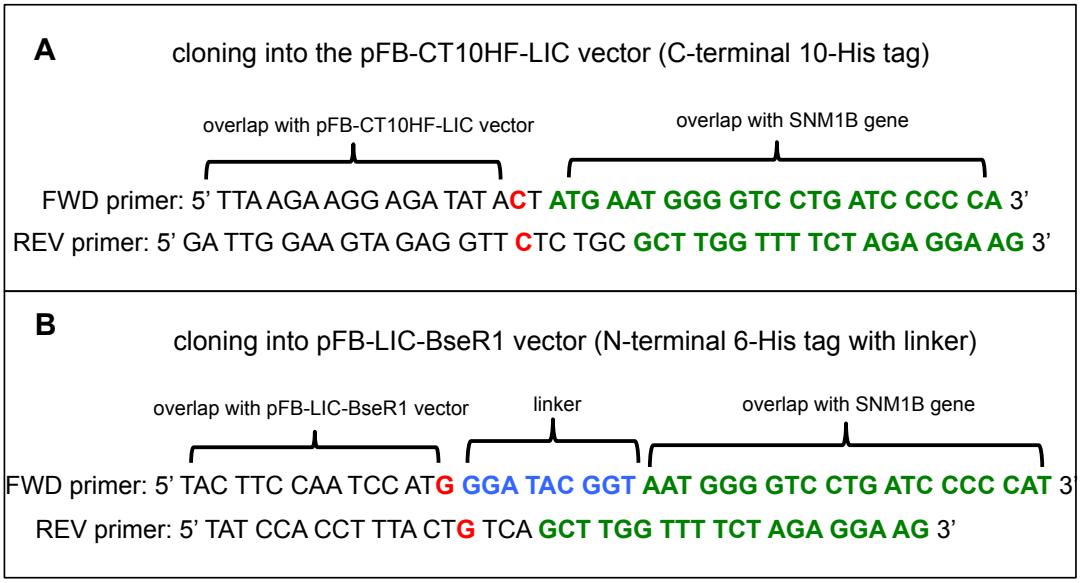


Figure 3.5 Primer design for ligation independent cloning.

Two constructs were designed for purification, both containing the desired truncated SNM1B gene: one with a C-terminal 10-His tag (using the pFB-CT10HF-LIC plasmid); and one with an N-terminal 6-His tag followed by a short glycine rich linker (using the pFB-LIC-Bse plasmid). A: primers for cloning into the pFB-CT10HF-LIC vector. B: primers for cloning into the pFB-LIC-BseR1 vector. The sequence that is complementary for the vector is used for LIC. The 5' to 3' exonuclease activity of T4 DNA polymerase resects the non-coding strand of the PCR amplicon to the point of the red nucleotide, thus generating 15–17 nucleotide overhangs for annealing. The same process was performed with the vector, allowing annealing of the insert and vector and subsequent transformation to propagate the recombinant plasmid.

The correct addition of these extensions was verified by agarose gel electrophoresis (Figure 3.6). Following PCR, both the empty vector and relevant SNM1B insert fragment were treated with T4 DNA polymerase to facilitate resection, generating the necessary overhangs for specific annealing, requisite for LIC.

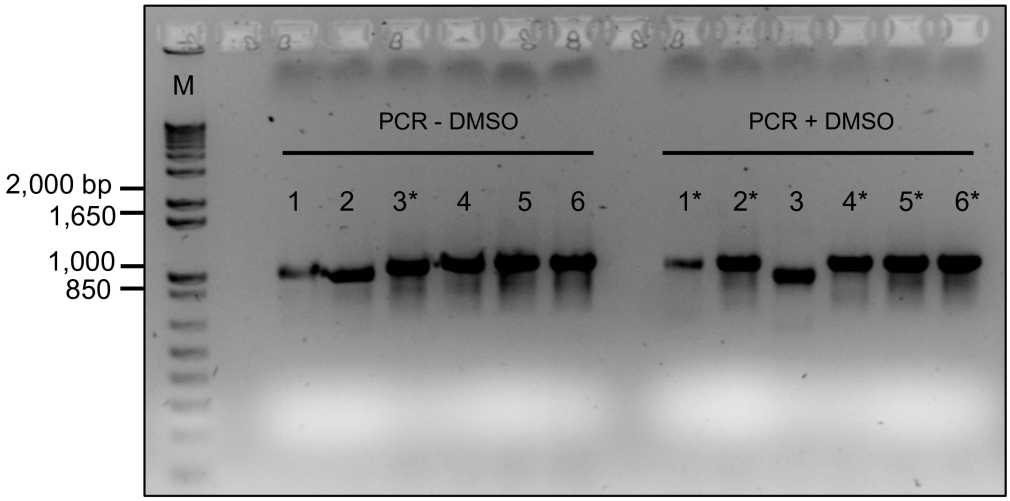


Figure 3.6 Successful amplification of SNM1B inserts for ligation independent cloning.

PCRs were performed using forward and reverse primers designed to introduce the overhangs for LIC and to amplify the truncated SNM1B gene product (1–1005 nt) (Figure 3.5). Reactions were performed using *Herculase II* DNA Polymerase, 6.25 ng DNA template, 100 ng of each primer and 200 µM each dNTP either with or without the addition of 3% DMSO. Samples are labelled as follows: 1 = WT, 2 = H61Y, 3 = D35A/H36A (all with primers designed for cloning into the pFB-CT10HF-LIC vector) 4 = WT, 5 = H61Y, 6 = D35A/H36A (all with primers designed for cloning into the pFB-LIC-Bse vector). A 1 kb Plus ladder was used and approximate sizes in base pairs are indicated. Samples of the appropriate size were marked with an asterisk (either 1020 or 1022 nucleotides, depending on whether 15 or 17 nt overhangs were added) were deemed successful and used in downstream applications. PCR amplicons were run on a 1.5% agarose gel and stained with ethidium bromide.

Correct annealing of the SNM1B gene insert into the relevant transfer vector (pFB-CT10HF-LIC or pFB-LIC-BseR1) was verified by a PCR screen (Figure 3.7) and sequencing. Subsequently, the annealed vector and inserts were propagated in DH5α *E. coli.*

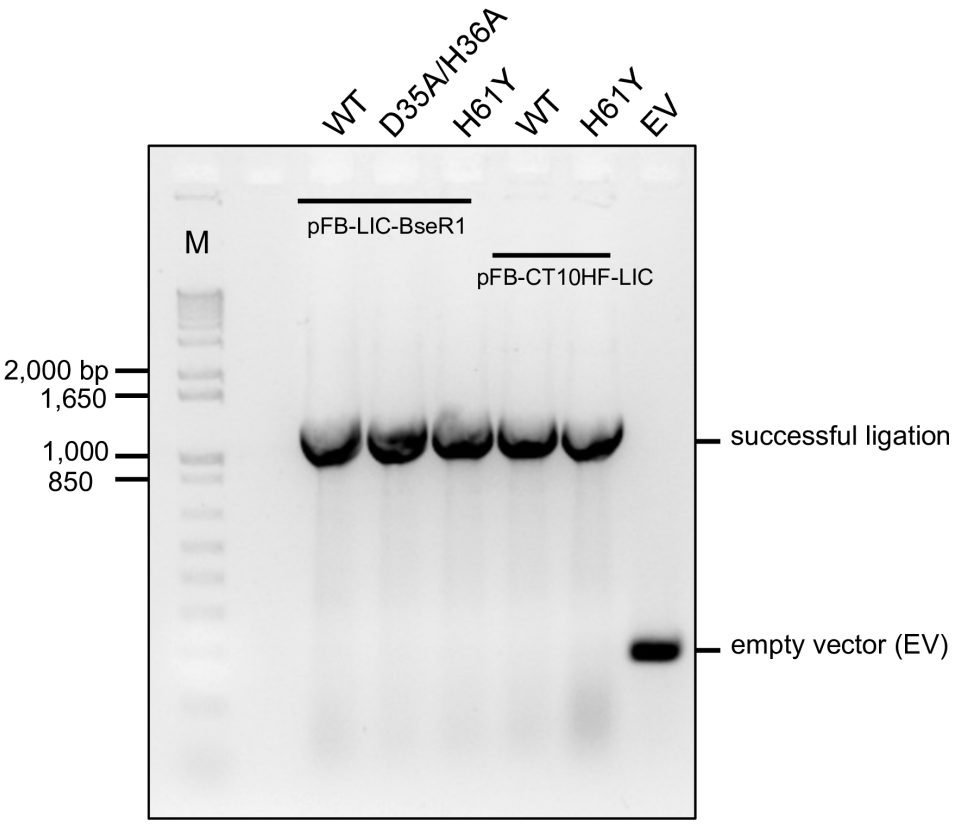


Figure 3.7 Successful annealing of the SNM1B gene insert into the appropriate vector.

The SNM1B inserts were annealed with the appropriate vectors (either pFB-CT10HF-LIC or pFB-LIC-Bse-R1) and transformed into competent *E coli*. PCRs were performed to verify correct annealing. Primers were designed to flank the cloning site of the vector and add 200 bp to the PCR fragment. Successful ligations were therefore regarded as being 1225 bp. Samples are as labelled and EV denotes empty vector. A 1 kb Plus ladder was used and approximate sizes in base pairs are indicated. PCR amplicons were run on a 1.5% agarose gel and stained with ethidium bromide.

The second step was the site-specific transposition of the cloned fragment into the baculovirus genome. This involved the expression of the transfer vector, the 'helper' plasmid, and the bacmid. Successful transposition resulted in reconstitution of the recombinant bacmid and disruption of the *lacZ* cassette, allowing blue/white colony selection. Successful transposition was confirmed by PCR of the bacmid DNA (Figure 3.8).

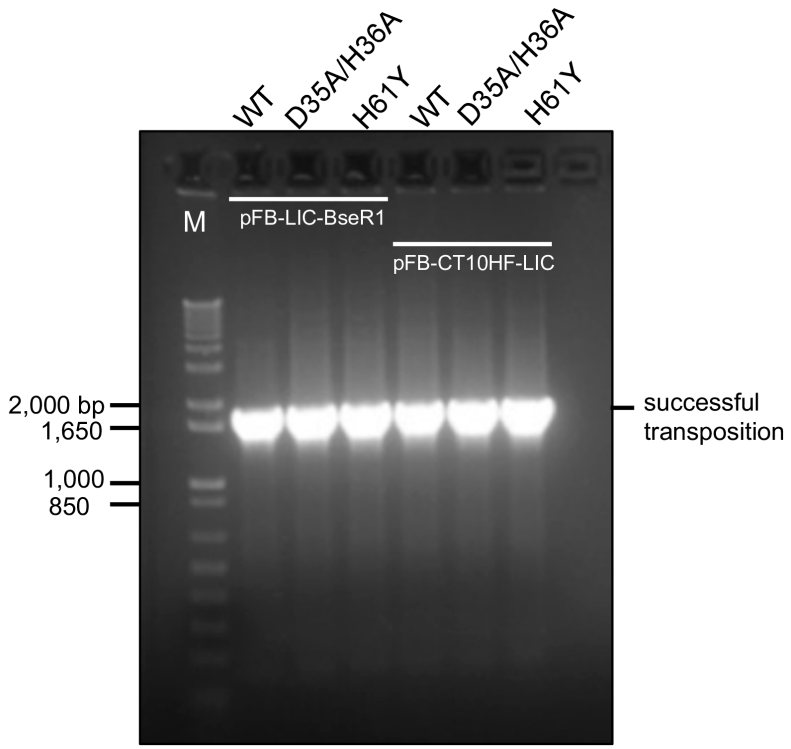


Figure 3.8 Successful transposition of the SNM1B construct from the indicated transfer vector to the bacmid.

Recombinant baculovirus were generated by site-specific transposition in DH10Bac *E. coli.* Transposition occurred by elements of the transfer vector to the parent bacmid in the presence of transposition proteins provided by a helper plasmid. Successful transposition disrupts the *LacZ* cassette in the parent bacmid thus allowing selection of white colonies. These were then screened by PCR with primers designed to flank the transposition site and add 700 bp to the truncated SNM1B gene resulting in an amplicon of 1725 bp if transposition was successful. SNM1B constructs are as labelled and it is indicated whether the gene was in the pFB-CT10HF-LIC vector or the pFB-LIC-BseR1 vector prior to transposition. A 1 kb Plus ladder was used and approximate sizes in base pairs are indicated. PCR amplicons were run on a 1.5% agarose gel and stained with ethidium bromide.

Successfully generated bacmids were transfected into Sf9 insect cells and the viral supernatant was used to infect further cells. A small-scale expression study (a single Ni-Sepharose column from 3 mL culture) was performed and initial studies indicated that three SNM1B constructs were successfully overexpressed and were able to be partially purified. This was evidenced by the presence of a ~38 kDa band in the final elution step of the test purification protocol (Figure 3.9) and was validated by electrospray ionisation time of flight (ESI-TOF) mass spectrometry (MS; data not shown).

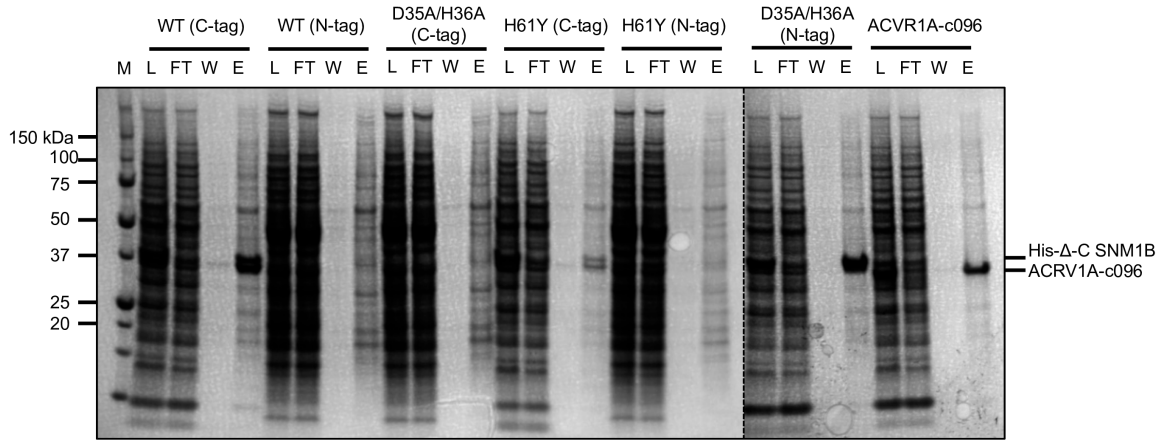


Figure 3.9 Test purification of WT and mutant Δ-C SNM1B (1/2-335).

SDS-PAGE of each step of the test purification protocol. Each construct (as labelled) has four samples from the following steps of the test purification: L = lysate, FT = flow through, W = wash, E = eluate. C-tag denotes a C-terminal 10-histidine tag, N-tag denotes an N-terminal 6-histidine tag. The ACVR1A-c096 sample was included as a highly-expressing positive control protein, Mw = 37015 Da. A Precision Plus Unstained Protein Marker was used and approximate molecular weights are as labelled.

The three proteins that were successfully overexpressed were WT Δ-C SNM1B (1–335) and H61Y Δ-C SNM1B (1–335) with a C-terminal 10-histidine tag, and

D35A/H36A Δ-C SNM1B (2–335) with an N-terminal 6-histidine tag with a linker. These viral cultures were scaled up for larger purification. Each of these purified proteins ran as a doublet on SDS-PAGE (Figure 3.9) and MS data indicates that both the tagged and untagged versions of the protein are present; this may be due to residual protease contamination, although it has not been investigated further.

3.2.2. Purification of WT and mutant SNM1B constructs

C-terminally truncated (1/2–335) WT, H61Y and D35A/H36A SNM1B proteins were purified from Sf9 insect cells. All had the addition of a histidine tag (either 6- or 10-His), spaced by a TEV cleavage site, at either the N-terminus or the C-terminus. Thus, the purification protocol involved an initial Ni-Sepharose chromatography column to selectively bind the protein through interaction with the histidine tag (Figure 3.10), and subsequent cleavage of the tag by incubation with the TEV protease. Complete removal of the tag was validated by ESI-TOF-MS (Figure 3.11, Figure 3.12, Figure 3.13) and a secondary Ni-Sepharose column was performed to capture the TEV protease and any residual uncleaved His-Δ-C SNM1B (Figure 3.14, Figure 3.15, Figure 3.16). Δ-C SNM1B containing eluate fractions (the 45 mM imidazole and 80 mM imidazole wash fractions from the second Ni-Sepharose column) were pooled, concentrated, and further purified by Superdex S200 size exclusion chromatography. The protein containing eluate fractions were identified by absorbance at A280 and these fractions were analysed by SDS-PAGE for the presence of the relevant Δ-C SNM1B (1/2–335). The purest fractions were subsequently pooled and concentrated for use in further studies (Figure 3.17,Figure 3.18, Figure 3.19). The final protein product was validated by

ESI-TOF-MS (Figure 3.20). Final protein yield was consistently between 15–20 mg from 2 L of Sf9 insect cell culture for each WT, D35A/H36A, and H61Y Δ-C SNM1B (1/2–335).

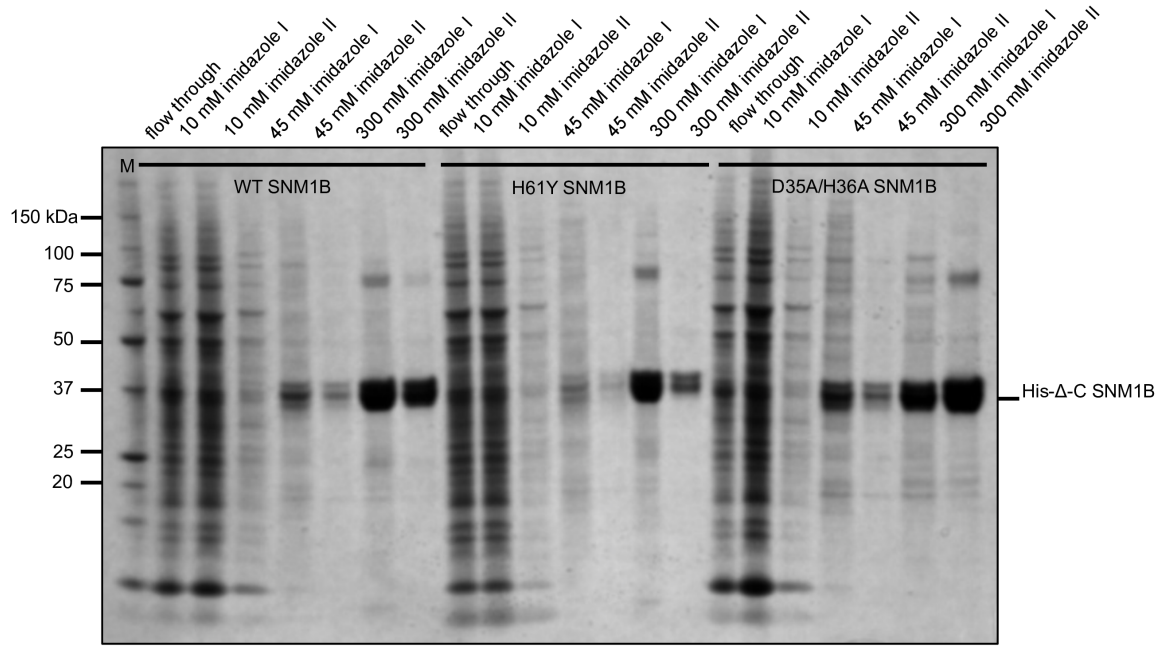


Figure 3.10 SDS-PAGE analysis of the initial Ni-Sepharose column for the purification of WT, D35A/H36A, and H61Y Δ-C SNM1B (1/2–335).

Samples were collected at all stages of the initial Ni-Sepharose column and were analysed by SDS-PAGE for the presence of the relevant Δ-C SNM1B (1/2–335). Cell lysates were added to the initial Ni-Sepharose column and the flow-through was collected. The column was washed with 10 column volumes (CV) of 10 mM imidazole containing buffer, 10 CV of 45 mM imidazole containing buffer, and the protein was eluted in 300 mM imidazole containing buffer. WT, H61Y, and D35A/H36A Δ-C SNM1B (1/2–335) are as labelled. A Precision Plus Unstained Protein Marker was used and approximate molecular weights are indicated.

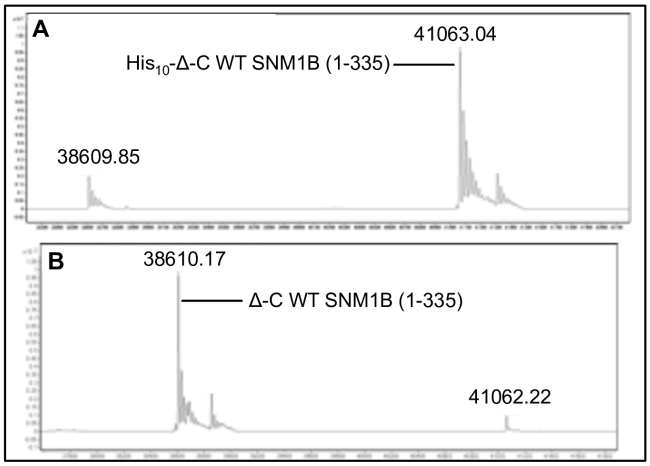


Figure 3.11 ESI-TOF-MS validating TEV protease cleavage of Δ-C WT SNM1B.

Mass spectrum analysis (counts vs deconvoluted mass (amu)) of Δ-C WT SNM1B (1–335) before (panel A) and after (panel B) treatment with TEV protease to remove the 10-His tag. A: His_{10}-Δ-C WT SNM1B, expected mass = 41020 Da, measured mass = 41063.04 Da. B = Δ-C WT SNM1B, expected mass = 38566.6 Da, measured mass = 38610.17 Da.

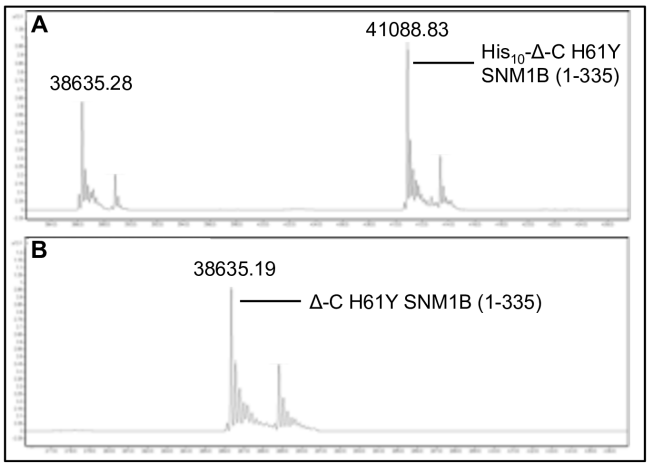


Figure 3.12 ESI-TOF-MS validating TEV protease cleavage of Δ-C H61Y SNM1B.

Mass spectrum analysis (counts vs deconvoluted mass (amu)) of Δ-C H61Y SNM1B (1–335) before (panel A) and after (panel B) treatment with TEV protease to remove the 10-His tag. A: His_{10}-Δ-C H61Y SNM1B, expected mass = 41046, Da, measured mass = 41088.83 Da. B = Δ-C H61YSNM1B, expected mass = 38592.6 Da, measured mass = 38635.19 Da.

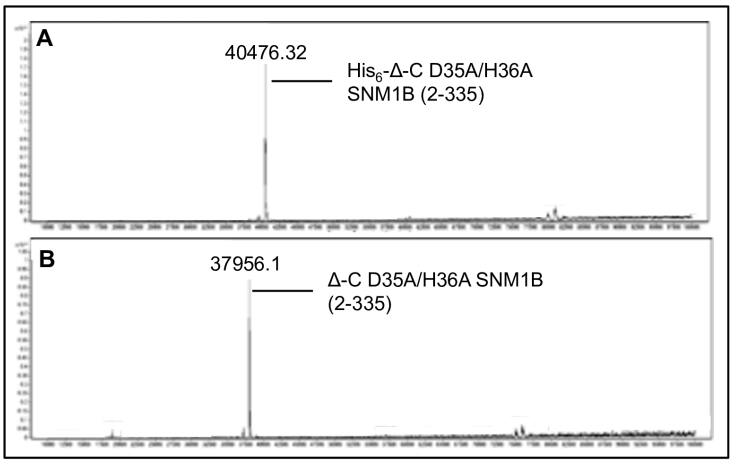


Figure 3.13 ESI-TOF-MS validating TEV protease cleavage of Δ-C D35A/H36A SNM1B.

Mass spectrum analysis of Δ-C D35A/H36A SNM1B (2–335) before (panel A) and after (panel B) treatment with TEV protease to remove the 6-His tag. A: His_6-Δ-C D35A/H36A SNM1B, expected mass = 40477.6 Da, measured mass = 40476.32 Da. B = Δ-C D35A/H36A SNM1B, expected mass = 37954.9 Da, measured mass = 37 956.1 Da.

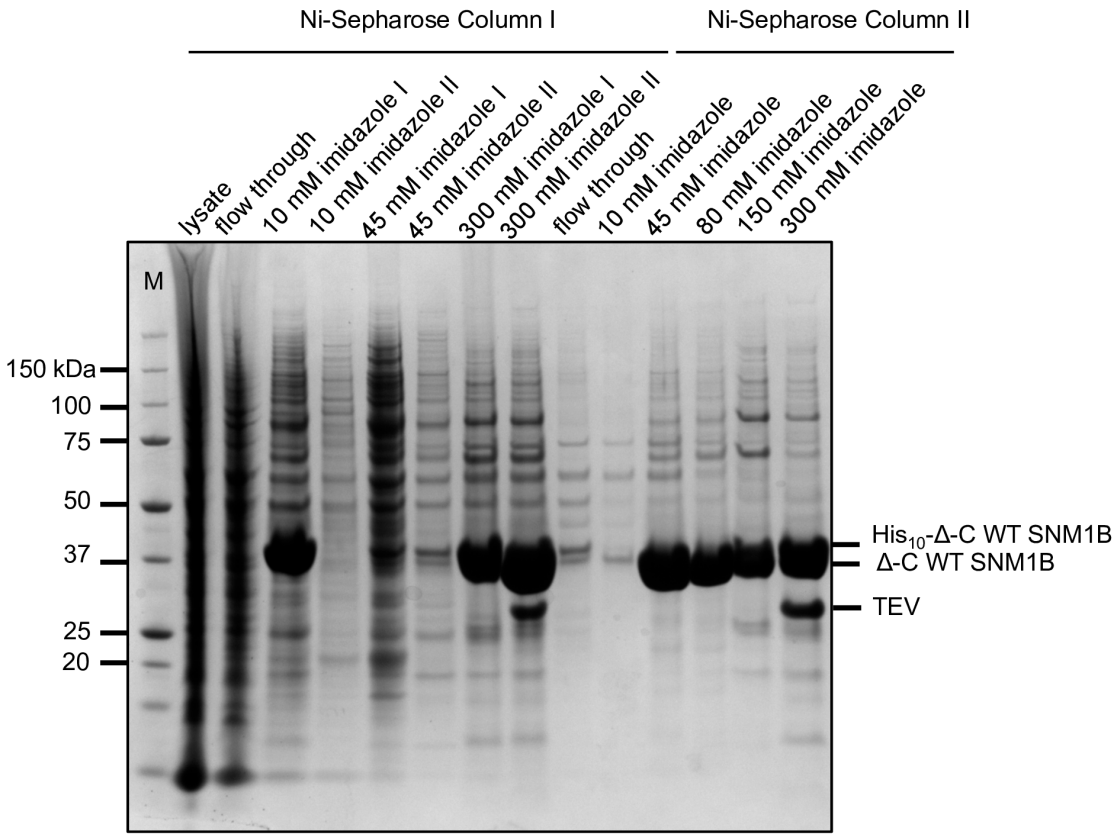


Figure 3.14 SDS-PAGE of both Ni-Sepharose columns for the purification of Δ-C WT SNM1B (1–335).

Samples were collected at all steps of both Ni-Sepharose columns and analysed by SDS-PAGE for the presence of Δ-C WT SNM1B (1–335), indicated by an over-expressed band at ~37 kDa. The samples from Ni-Sepharose column I are as specified in Figure 3.10. After treatment with the TEV protease to cleave the 10-His tag, the second Ni-Sepharose column was performed and relevant steps are as labelled. The Δ-C WT SNM1B sample (after TEV cleavage) was added to the column and the flow-through was collected. The column was then washed with 10 CV of each 10, 45, 80, 150, 300 mM imidazole containing buffer and all fractions were collected. The 45, 80, and 150 mM imidazole fractions were collected, pooled, and concentrated for subsequent S200 gel filtration chromatography.

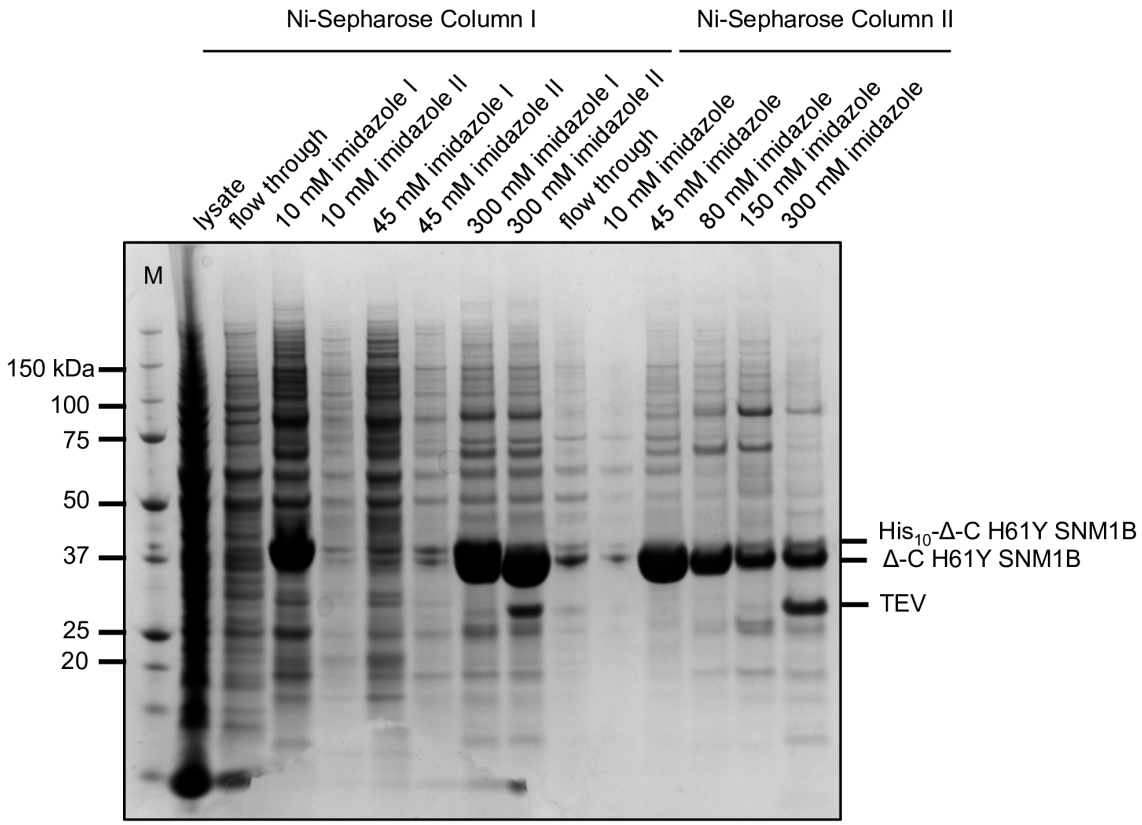


Figure 3.15 SDS-PAGE analysis of both Ni-Sepharose columns for the purification of Δ-C H61Y SNM1B (1–335).

Samples were collected at all steps of both Ni-Sepharose columns and analysed by SDS-PAGE for the presence of Δ-C H61Y SNM1B (1–335), indicated by the presence of an over-expressed band at ~37 kDa. The samples from Ni-Sepharose column I are as specified in Figure 3.10. After treatment with the TEV protease to cleave the 10-His tag, the second Ni-Sepharose column was performed and relevant steps are as labelled. The Δ-C H61Y SNM1B sample (after TEV cleavage) was added to the column and the flow-through was collected. The column was then washed with 10 CV of each 10, 45, 80, 150, 300 mM imidazole containing buffer and all fractions were collected. The 45, 80, and 150 mM imidazole fractions were collected, pooled, and concentrated for subsequent S200 gel filtration chromatography.

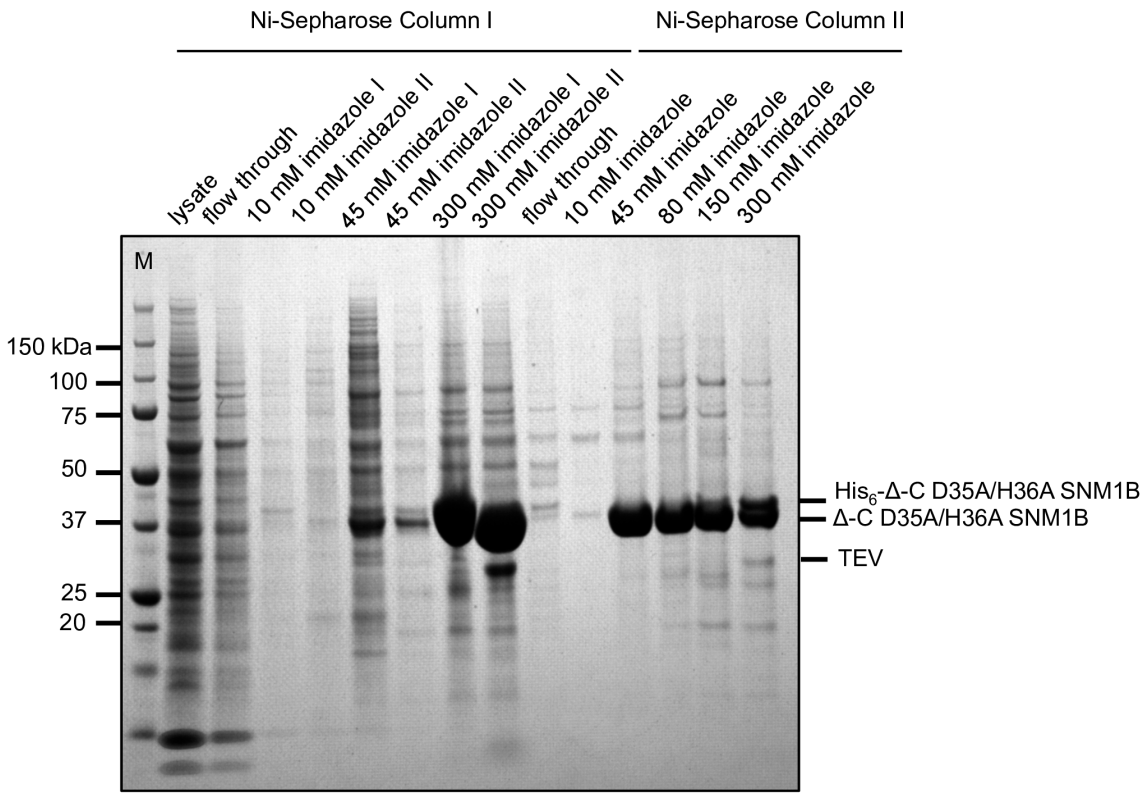


Figure 3.16 SDS-PAGE analysis of both Ni-Sepharose columns for the purification of Δ-C D35A/H36A SNM1B (1–335).

Samples were collected at all steps of both Ni-Sepharose columns and analysed by SDS-PAGE for the presence of Δ-C D35A/H36A SNM1B (2–335), indicated by the presence of an over-expressed band at ~37 kDa. The samples from Ni-Sepharose column I are as specified in Figure 3.10. After treatment with the TEV protease to cleave the 10-His tag, the second Ni-Sepharose column was performed and relevant steps are as labelled. The Δ-C D35A/H36A SNM1B sample (after TEV cleavage) was added to the column and the flow-through was collected. The column was then washed with 10 CV of each 10, 45, 80, 150, 300 mM imidazole containing buffer and all fractions were collected. The 45, 80, and 150 mM imidazole fractions were collected, pooled, and concentrated for subsequent S200 gel filtration chromatography.

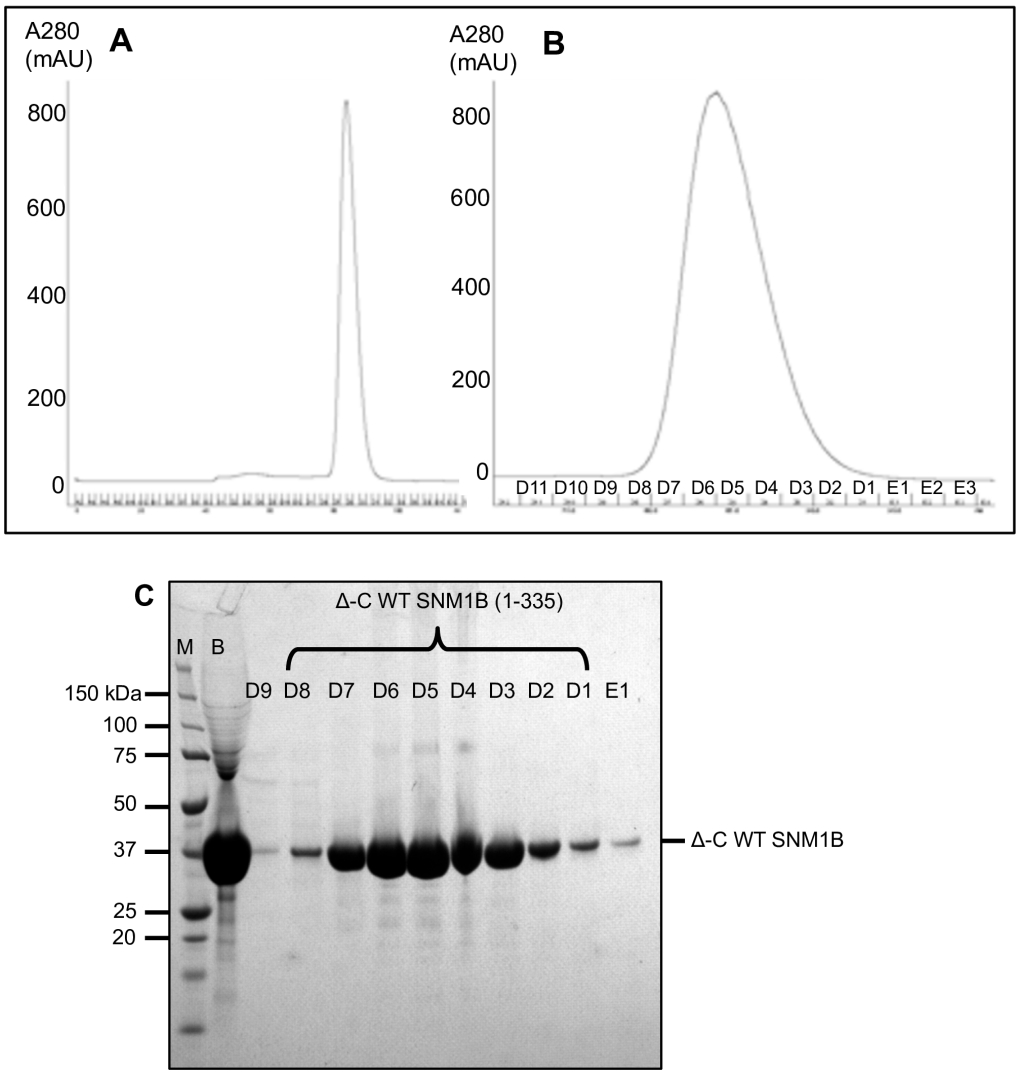


Figure 3.17 Superdex S200 gel filtration for Δ-C WT SNM1B (1–335).

Panels A and B display absorbance at 280 nm across collected fractions for Δ-C WT SNM1B (1–335) after S200 Superdex gel filtration chromatography. Panel A indicates all fractions, and panel B is enlarged to show the protein containing fractions. Panel C: SDS-PAGE of peak protein containing fractions, as indicated in B. The presence of a ~38 kDa band across the peak fractions indicates the presence of Δ-C WT SNM1B (1–335). The bracket indicates purified protein fractions that were pooled and concentrated for use in further assays. M denotes molecular weight marker, and B, the protein sample before it was loaded onto the S200 Superdex gel filtration column.

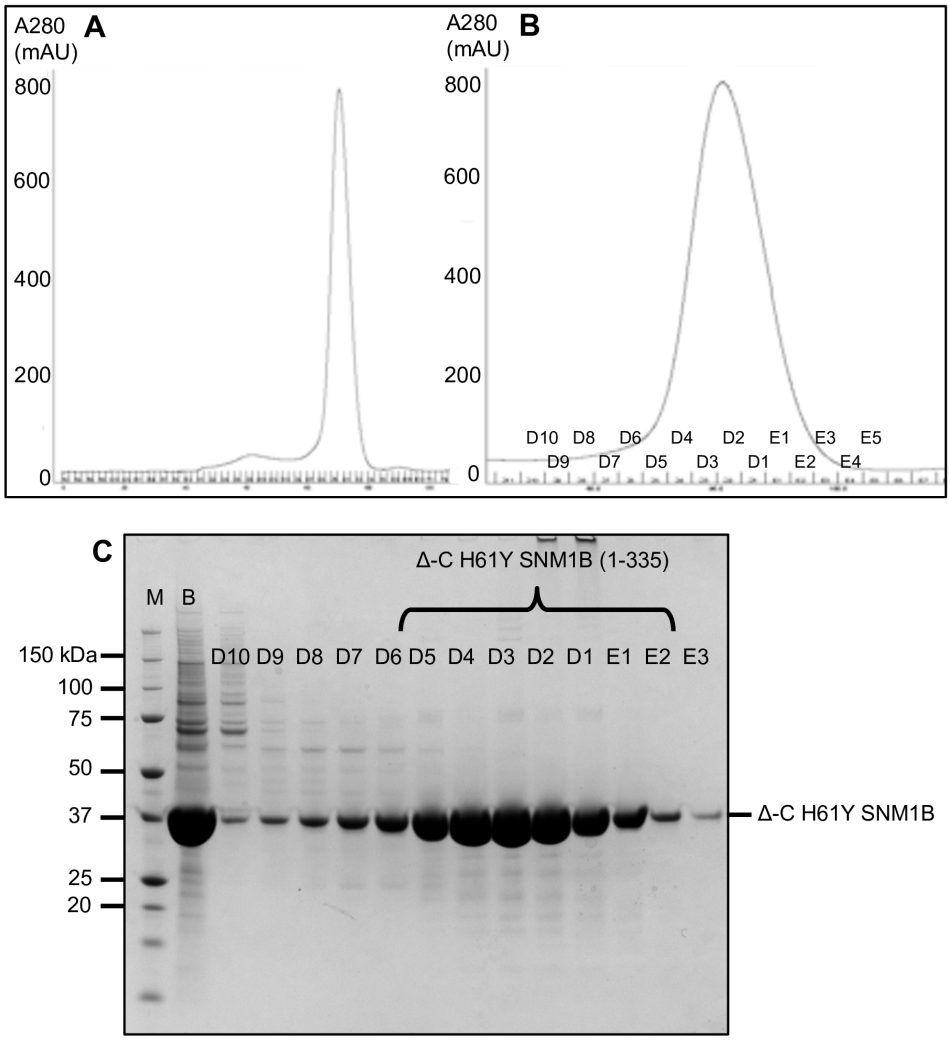


Figure 3.18 Superdex S200 gel filtration for Δ-C H61Y SNM1B (1–335).

Panels A and B display absorbance at 280 nm across collected fractions for Δ-C H61Y SNM1B (1–335) after S200 Superdex gel filtration chromatography. Panel A indicates all fractions, and panel B is enlarged to show the protein containing fractions. Panel C: SDS-PAGE of peak protein containing fractions, as indicated in B. The presence of a ~38 kDa band across the peak fractions indicates the presence of Δ-C H61Y SNM1B (1–335). The bracket indicates purified protein fractions that were pooled and concentrated for use in further assays. M denotes molecular weight marker, and B, the protein sample before it was loaded onto the S200 Superdex gel filtration column.

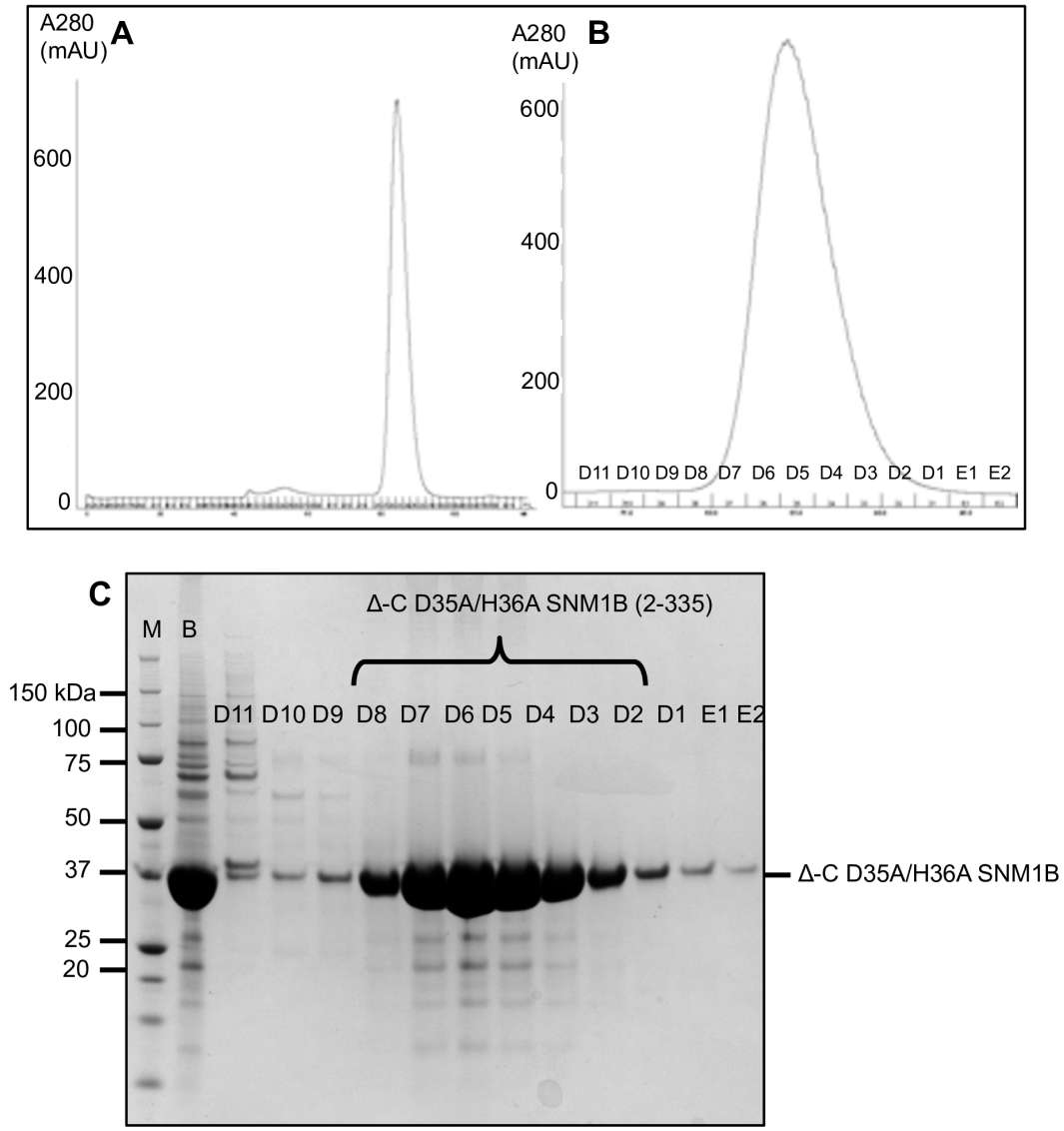


Figure 3.19 Superdex S200 gel filtration for Δ-C D35A/H36A SNM1B (1–335).

Panels A and B display absorbance at 280 nm across collected fractions for Δ-C D35A/H36A SNM1B (2–335) after S200 Superdex gel filtration chromatography. Panel A indicates all fractions, and panel B is enlarged to show the protein containing fractions. Panel C: SDS-PAGE of peak protein containing fractions, as indicated in B. The presence of a ~38 kDa band across the peak fractions indicates the presence of Δ-C D35A/H36A SNM1B (2–335). The bracket indicates purified protein fractions that were pooled and concentrated for use in further assays. M denotes molecular weight marker, and B, the protein sample before it was loaded onto the S200 Superdex gel filtration column.

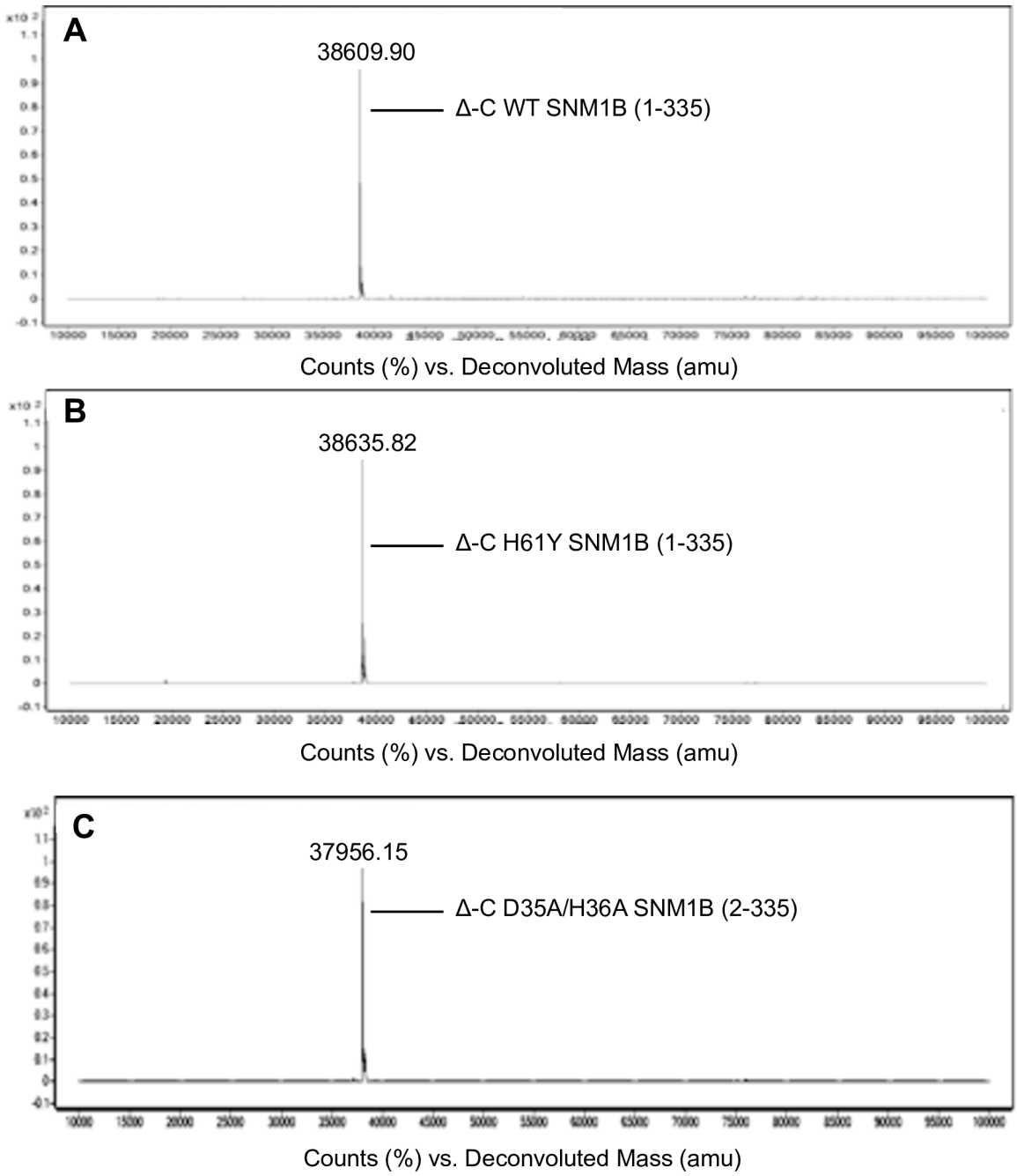


Figure 3.20 ESI-TOF-MS validation of final purified Δ-C SNM1B (1/2–335).

Mass spectrometry analysis of WT, H61Y, and D35A/H36A Δ-C SNM1B (1/2–335). A: WT Δ-C SNM1B (1–335), expected mass = 38566.6 Da, measured mass = 38609.9 Da. B: H61Y Δ-C SNM1B (1–335), expected mass = 38592.6 Da, measured mass = 38635.82 Da. C: D35A/H36A Δ-C SNM1B (2–335), expected mass 337954.9 Da, measured mass = 37956.15 Da.

Each of the proteins, WT, D35A/H36A, and H61Y Δ-C SNM1B (1/2–335) elute in the monomeric form from the S200 Superdex gel filtration column. For each protein, the primary protein peak (as evidenced by A280 and SDS-PAGE) eluted at ~80 mL (Figure 3.17, Figure 3.18, Figure 3.19). This is consistent with the monomeric form when analysed using the equation y = -0.0415 x + 5.1829 (where y is the log(Mw)) and x is the retention volume) from the standard curve generated by running the BioRad molecular weight standards (thyroglobulin (670 kDa), γ-globulin (158 kDa), ovalbumin (44 kDa), myoglobin (17 kDa), vitamin B12 (1.3 kDa)). It has previously been reported that SNM1B can interact with itself in cellular immunoprecipitation experiments (van Overbeek and de Lange, 2006). However, given that SNM1B elutes as a monomer from size exclusion chromatography, it would seem that the truncated SNM1B (1/2–335) lacks the specific domain or residues required for self-interaction, or this is an event that only occurs under specific cellular conditions.

For two of the samples – WT and H61Y Δ-C SNM1B (1–335), there was a +43 Da difference from the predicted mass to the measured mass, whereas the measured mass of D35A/H36A Δ-C SNM1B (2–335) precisely matched the expected mass (Figure 3.20). This mass discrepancy was noted at all steps of the purification – i.e. both before, and after, treatment with the TEV protease, and after the final S200 size exclusion column. Both the WT and H61Y SNM1B constructs were cloned for purification in the same vector (pFB-CTHF-LIC) and were tagged at the C-terminus. This is unlike the D35A/H36A construct, which was cloned into the pFB-LIC-Bse vector and tagged at the N-terminus. Post-translational modification (PTM) mapping of both the WT and H61Y Δ-C SNM1B (1–335) confirmed that this

mass differential was due to acetylation of the start methionine (+42 Da), a particularly common type of modification (Varland et al., 2015) and deamination of the second amino acid, asparagine N2 (+1 Da). PTM mapping was performed by proteolytic digestion of both WT and H61Y Δ-C SNM1B (1–335) (with trypsin, chymotrypsin, and pepsin) and subsequent tandem MS (MS/MS).

Interestingly, the MS results for all WT, D35A/H36A, and H61Y Δ-C SNM1B (1/2–335) indicate the presence of a small second peak ~248 Da larger than the main species of the correct molecular weight (Figure 3.11, Figure 3.12, Figure 3.13, Figure 3.20). Proteolytic digest MS/MS and MS from relevant gel bands excised from SDS-PAGE showed only the presence of the SNM1B protein, so it seems most likely that this is an unidentified post-translational modification, rather than a contaminating protein. Additionally, it can be seen (Figure 3.11, Figure 3.12) that, prior to TEV cleavage, for both WT and H61Y Δ-C SNM1B (1–335) there is a small amount of the protein that is already untagged. Again, the reason for this is not known.

3.2.3. DNA binding and nuclease activity of WT SNM1B

As an exonuclease with 5’ to 3’ directionality, SNM1B has previously been shown to bind to, and digest, a variety of DNA substrates *in vitro* (Sengerova et al., 2012, Lenain et al., 2006). To determine if SNM1B has any structure-based substrate selectivity we designed a number of structurally diverse DNA substrates, particularly those that mimic stalled replication forks or replication intermediates. The nuclease activity of SNM1B was examined on these structures, as they are

common intermediates in the repair of interstrand crosslinks (ICLs) and are thus perhaps physiologically relevant, given the involvement of SNM1B in mediating ICL repair. All substrates were based on a single stranded (ss-) 51 oligonucleotide radio-labelled with α-^{32}P-dATP on the 3' end. These were fully annealed to complementary oligonucleotide sequences to generate the substrates represented in Figure 3.21.

Electrophoretic mobility shift assays (EMSAs) were performed to assess SNM1B DNA substrate binding. WT Δ-C SNM1B (1–335) (henceforth denoted WT SNM1B) shows binding to all DNA substrates at a concentration of 5.0 nM and above (Figure 3.22). The binding of SNM1B to DNA also appears to be stoichiometrically 1:1, as there is a single shift that occurs when SNM1B binds to the DNA, and no secondary 'supershift' (Figure 3.22). EMSAs are a semi-quantitative method for analysing DNA protein interactions, and the data suggest that WT SNM1B is able to bind to the tested range of structurally diverse DNA substrate with comparable efficiency.

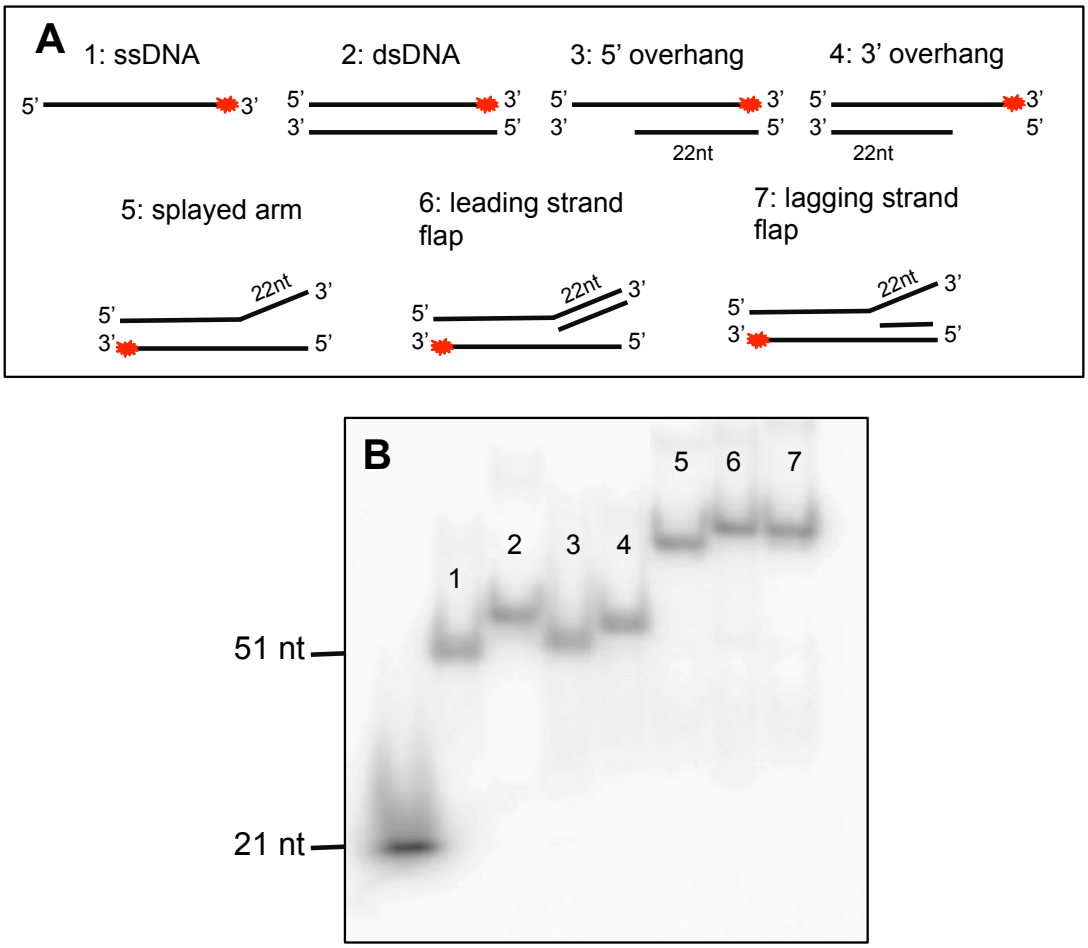


Figure 3.21 Generation of a variety of DNA substrates.

A: Single-stranded 51 nt oligonucleotides radiolabelled at the 3' end were annealed with the appropriate corresponding oligonucleotide to generate the substrates outlined in schematic form (see Appendix V for DNA sequences). The red stars denote the α-^{32}P-dATP label. B: Native polyacrylamide gel electrophoresis demonstrating correct annealing of oligonucleotide substrates. Labelled as follows: 1 = ssDNA, 2 = dsDNA (annealed strand = 50 nt), 3 = 5' overhang, 4 = 3' overhang, 5 = splayed arm, 6 = leading strand flap, 7 = lagging strand flap.

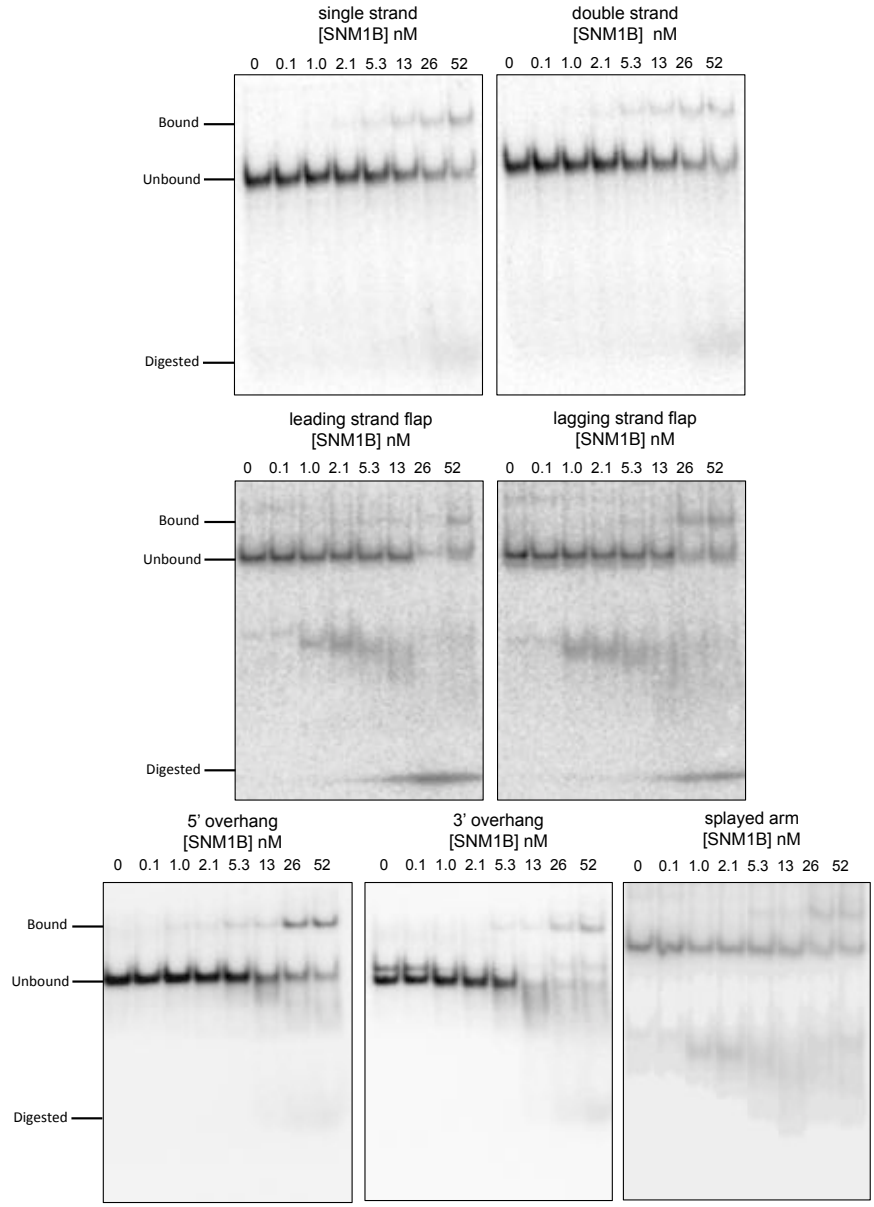


Figure 3.22 WT SNM1B binds to a variety of structurally diverse DNA substrates.

EMSAs showing binding of WT SNM1B to DNA substrates. 1 nM of the appropriate DNA substrate was incubated for 5 minutes at 37^{o}C with WT SNM1B (concentrations as labelled), before being analysed by 10% native polyacrylamide gel electrophoresis. Unbound DNA substrate, DNA substrate bound to SNM1B, and digestion products are as labelled. DNA substrate structures are as indicated. Each gel is representative of at least three individual experiments.

Radiolabelled nuclease assays have previously been used to examine the enzymatic activity of SNM1B *in vitro* (Lenain et al., 2006, Sengerova et al., 2012). As such, a similar approach was used in the present study. Initial optimisation experiments were performed and time points of up to 90 minutes, and an enzyme concentration between 0.8–1.0 nM resulted in reaction rates that could be conveniently monitored in the gel-based assays.

SNM1B exhibited effective digestion of all substrates over a 90 minute time period (Figure 3.23). The single stranded DNA (ssDNA) displayed the most complete digestion and this preference for ssDNA is in accordance with previous research (Sengerova et al., 2012, Lenain et al., 2006). The 3’ overhang substrate also appears to be more efficiently processed than each double stranded (ds-), 5’ overhang, and splayed arm DNA substrates. The leading strand flap and the lagging strand flap exhibit the least effective digestion over time (Figure 3.23).

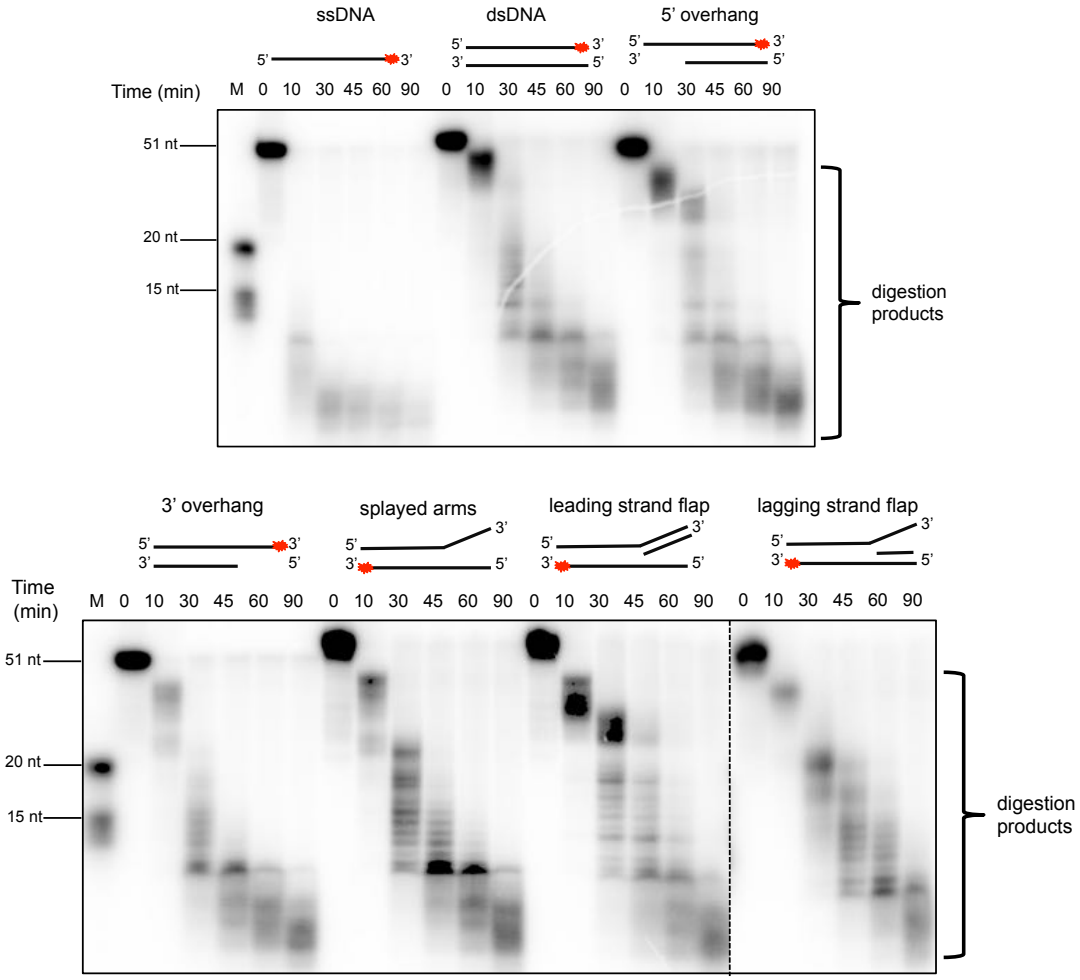


Figure 3.23 WT SNM1B exhibits efficient exonucleolytic digestion over time of a variety of structurally diverse DNA substrates.

100 nM of DNA substrate (as labelled) was added to 0.8 nM SNM1B and incubated at 37°C for increasing time. The reaction was stopped with the addition of 2 µL stop solution and boiling. Samples were then analysed by 20% polyacrylamide denaturing gel electrophoresis. DNA substrate structures are as indicated. Molecular weight markers are as labelled. Each gel is representative of at least three individual experiments.

3.2.4. Initial biochemical and biophysical characterisation of WT and mutant SNM1B

To analyse the effect of the H61Y mutation on SNM1B function a number of DNA binding and nuclease assays were performed. Initial biochemical assays revealed that H61Y Δ-C SNM1B (1–335) (henceforth denoted H61Y SNM1B) had similar activity on a 51 nucleotide ssDNA substrate, when compared to the WT protein (Figure 3.24). As anticipated, the D35A/H36A Δ-C SNM1B (2–335) entirely lacks nuclease activity on a ssDNA substrate (Figure 3.24).

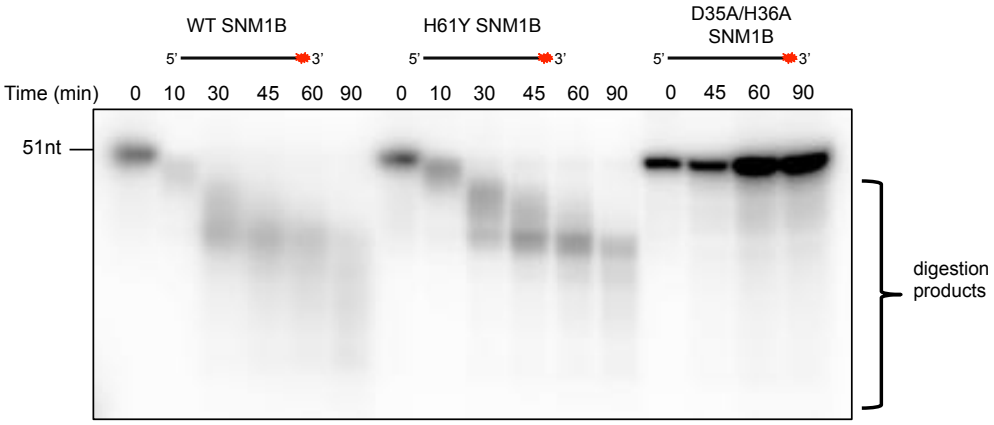


Figure 3.24 WT and H61Y, but not D35A/H36A, SNM1B exhibit nuclease activity on a ssDNA substrate.

100 nM of DNA substrate (as labelled) was added to 0.8 nM SNM1B in a total reaction volume of 10 µL and incubated at 37°C for increasing time. The reaction was stopped with the addition of 2 µL stop solution and boiling. Samples were then analysed by 20% polyacrylamide denaturing gel electrophoresis. Each gel is representative of at least three individual experiments.

This lack of activity was to be expected, as this has previously been reported (Sengerova et al., 2012), and these two residues form the latter part of the

conserved HxHxDH motif, and have been shown to directly coordinate one of the zinc metal ions in the active site (Allerston et al., 2015). It also serves as an important negative control in protein purification, as it provides evidence that the measured enzymatic activity of SNM1B is intrinsic to SNM1B and not due to a contaminating nuclease.

To rule out the possibility that the H61Y mutation is destabilising to the structure of SNM1B, a thermal shift assay was performed. Both WT and H61Y SNM1B showed similar rates of protein denaturation as temperature increased (WT T_m = 57.4°C and H61Y T_m = 57.2°C), suggesting that the H61Y mutation does not significantly perturb the structural conformation of the enzyme (Figure 3.25).

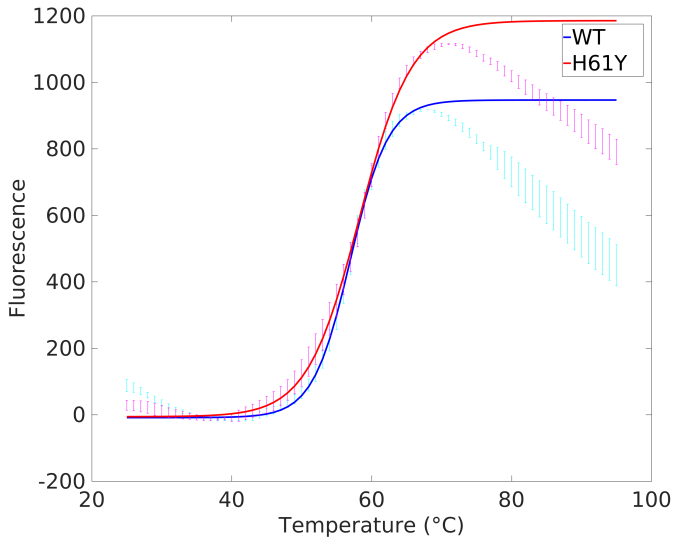


Figure 3.25 The H61Y does not disrupt the thermal stability, compared to the WT SNM1B protein.

Temperature was increased over time and, as the protein denatured, fluorescence was indicative of SYPRO Orange dye binding to hydrophobic protein surfaces. WT T_m = 57.4°C and H61Y T_m = 57.2°C. The dashed lines represent the measured fluorescence, and the continuous lines represent the modelled curve.

To further validate the following biochemical analyses, both WT and H61Y SNM1B were purified a total of four times (Figure 3.26). These independent purifications were shown to have comparable nuclease activity on a 21 nt ssDNA substrate, and at least two purifications were tested in all subsequent assays. When all four purifications were tested side by side, WT and H61Y SNM1B were shown to have comparable nuclease activity on 21 nucleotide ssDNA (Figure 3.26).

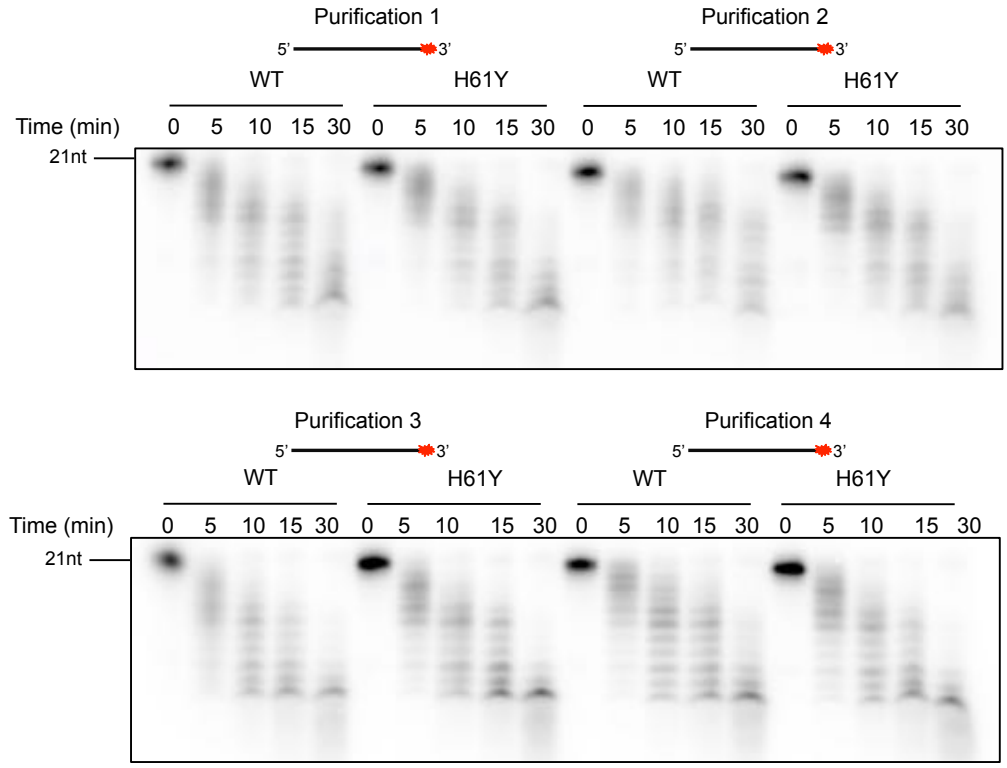


Figure 3.26 The nuclease activity of WT and H61Y SNM1B from four separate purifications.

100 nM of 21 nucleotide ss-DNA substrate was incubated with 1.0 nM SNM1B at 37°C for increasing time. The reaction was stopped with the addition of 2 µL stop solution and boiling. Samples were analysed by 20% polyacrylamide gel electrophoresis. Each gel is representative of at least three individual experiments.

The crystal structure of SNM1B (PDB: 5AHO) (Allerston et al., 2015) revealed that the His-61 residue was located adjacent to the active site, and it was hypothesised

that the positively charged imidazole side chain may interact with the negatively charged phosphodiester backbone of the DNA substrate. For this reason the DNA binding ability of H61Y SNM1B was assessed. In contrast with previous binding analysis, this was done via a fluorescence anisotropy based assay instead of a gel shift assay, as the former is more quantifiable than the latter (Heyduk et al., 1996). However, both WT and H61Y show very similar DNA binding affinities – WT SNM1B K_d = 153 nM ± 20 nM and H61Y SNM1B K_d = 209 nM ± 15 nM (Figure 3.27).

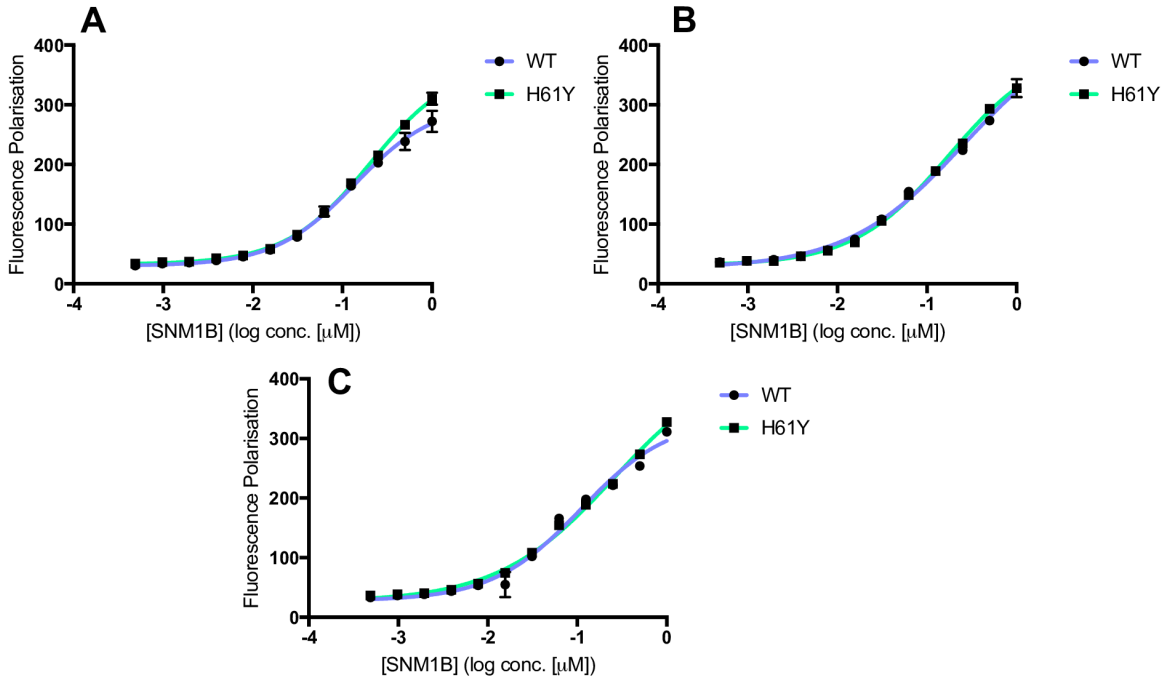
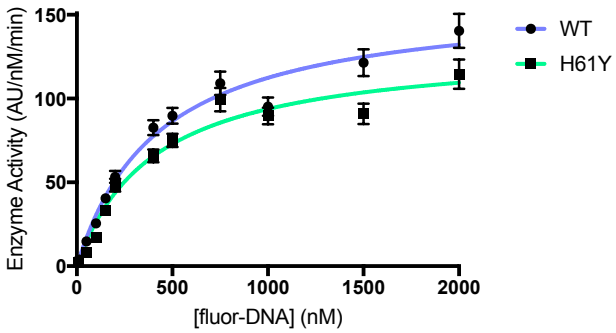


Figure 3.27 Fluorescence anisotropy DNA binding assay for WT and H61Y SNM1B.

A, B, and C denote three separate purifications for both WT and H61Y SNM1B. 10 nM single stranded 21 nt DNA substrate was incubated with serially diluted concentrations of SNM1B (as indicated) for five minutes at room temperature. Each reaction was in quadruplicate and at least two experimental repeats were performed. Error bars indicate the SD. WT SNM1B K_d = 153 nM ± 20 and H61Y SNM1B K_d = 209 nM ± 15.

In addition to the gel-based nuclease assay, the enzymatic activity of both WT and H61Y SNM1B was determined using a real-time fluorescence-based assay. This real-time fluorescence assay (described in Chapter 2) tests the ability of SNM1B to hydrolyse the first phosphodiester bond at the 5' end of a fluor-DNA substrate (Appendix IV). Michaelis-Menten curves were generated from this assay (Figure 3.28) and the K_M and k_{cat} values for WT and H61Y show that there is no significant difference in catalytic activity between the two (Figure 3.28).

A



B

SNM1B	K_M (nM)	k_{cat} (AU/nM/ min)
WT	428 ± 37	160 ± 5
H61Y	402 ± 43	132 ± 5

Figure 3.28 Michaelis-Menten enzyme kinetics for WT and H61Y SNM1B.

A: Graph showing Michaelis-Menten curves for WT and H61Y SNM1B. Fluorescence based nuclease assay with 20 nt fluorescein labelled ssDNA substrate. Data obtained from experiments from four purifications of both WT and H61Y SNM1B with at least three experimental repeats for each. Error bars indicate the SEM. B: table showing K_M and k_{cat} of WT and H61Y SNM1B. Values determined by fitting Michaelis-Menten curves using Graphpad-Prism software.

3.2.5. Analysis of WT and H61Y nuclease activity on structurally diverse DNA substrates

H61Y SNM1B appears to have comparable nuclease activity and DNA binding activity to the WT protein on ssDNA substrates (Figure 3.24, Figure 3.26, Figure 3.27), and this would suggest that this mutation has no effect on the ability of SNM1B to process simple substrates. However, the ability of H61Y SNM1B to effectively digest a range of structurally diverse substrates was not known. As such, gel-based nuclease assays were performed with both WT and H61Y using the DNA structures outlined in Figure 3.21.

WT and H61Y exhibited comparable nuclease activity on all substrates tested (Figure 3.29). Both WT and H61Y SNM1B showed a strong preference for a ssDNA substrate, and some slight preference for a 3' overhang substrate, however, all substrates showed effective digestion over the 60 minute incubation period. This is in accordance with previous results for WT SNM1B (Figure 3.23). Thus, it would seem that H61Y has similar *in vitro* nuclease activity to WT SNM1B on different DNA structures.

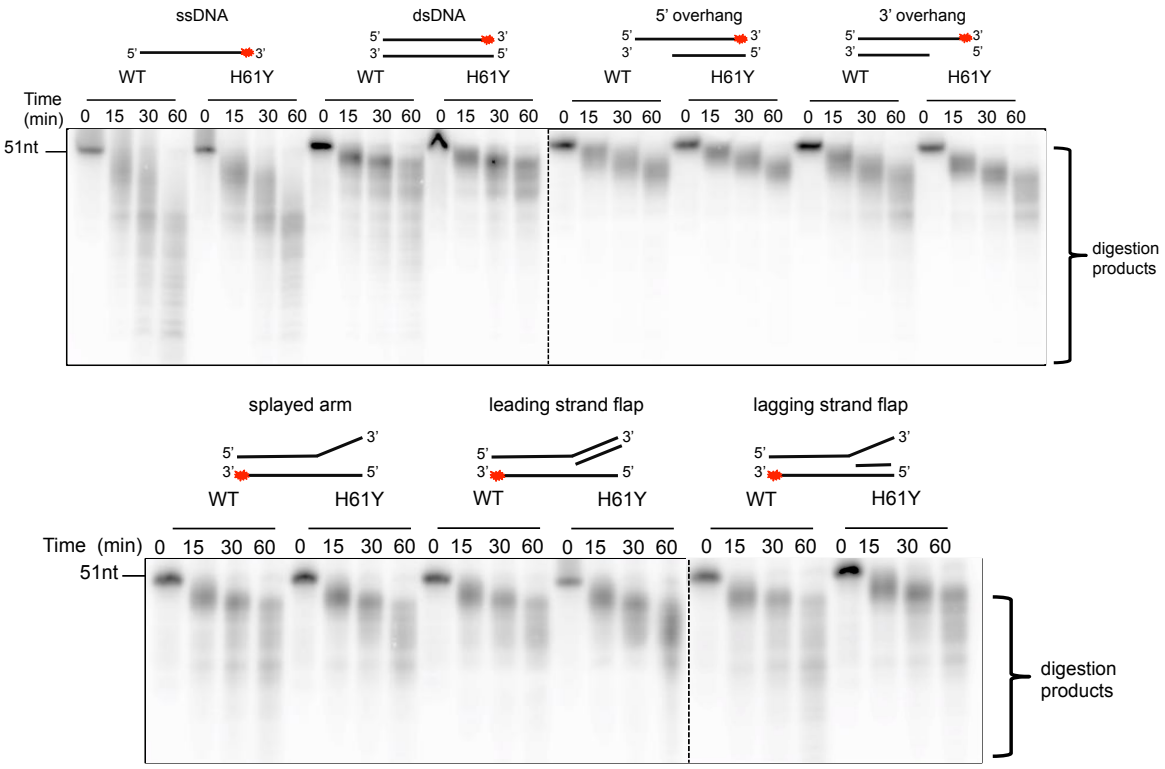


Figure 3.29 Both WT and H61Y SNM1B digest a variety of DNA substrates.

Nuclease assay showing digestion of various DNA substrates over time. 100 nM of DNA substrate (as labelled) was added to 1.0 nM SNM1B enzyme and incubated at 37°C, as previously. The reaction was stopped with the addition of 2 µL stop solution and boiling. Samples were then analysed by 20% polyacrylamide gel electrophoresis. Each gel is representative of at least three individual experiments from two separate purifications.

3.2.6. Analysis of WT and H61Y SNM1B on telomeric DNA substrates

The primary cellular function of SNM1B is to resect the C-rich strand at newly replicated leading strand telomeres to generate the 3' overhang necessary for t-loop formation (Lam et al., 2010). As such, the nuclease activity of WT and H61Y was examined on telomeric sequence substrates. The telomeric sequences used

were adapted from those outlined in (Lenain et al., 2006) and contained GGTTAG repeats on the top strand and a CTAACC sequence on the lower strand (see Appendix V). WT and H61Y SNM1B are able to digest both the standard DNA substrates and telomeric DNA substrates effectively. Again, there appeared to be no significant difference between WT and H61Y SNM1B on either of the telomeric substrates (Figure 3.30).

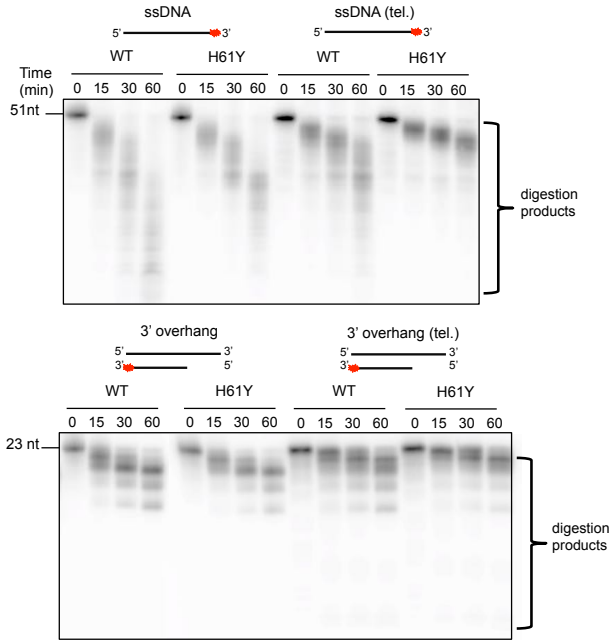


Figure 3.30 Both WT and H61Y SNM1B ss- and 3' overhang DNA substrates containing telomeric repeat sequences.

Gel-based nuclease assay showing digestion of ss- and 3' overhang DNA substrates over time. Telomeric substrates (those labelled tel.) are the same length as the control DNA substrates and are similarly annealed, however, contain TTAGGG nucleotide repeats. 100 nM of DNA substrate (as labelled) was added to 1.0 nM SNM1B enzyme and incubated at 37^{o}C . The reaction was stopped with the addition of 2 µL stop solution and boiling. Samples were then run on a 20% polyacrylamide denaturing gel and each gel is representative of at least three individual experiments from two separate purifications.

Interestingly, both WT and H61Y SNM1B appear to exhibit slightly decreased nuclease activity on the telomeric ssDNA substrate compared to the control ssDNA substrate (Figure 3.30). The reason for this is not immediately clear, although, the G rich nature of these oligonucleotide sequences may allow for secondary structure formation. It is possible for even relatively short telomeric G-rich sequences (for example, d[GGG(TTAGGG)$_3$]) to form intramolecular G-quadruplexes through Hoogsteen base-pairing (Zaug et al., 2005). Formation of G-rich DNA into a G-quadruplex requires the presence of monovalent cations, such as Na^+ or K^+, to stabilise the structure (Sen and Gilbert, 1990) and the structure of the G-quadruplex differs depending on which ion is present (Antonacci et al., 2007). The telomeric oligonucleotide sequence used in this assay, [ATCGTCCTAGCAAGG(ggttag)$_6$], in a reaction buffer containing 50 mM KCl, at a temperature of 37^{o}C may have the ability to form a G-quadruplex *in vitro*. However, in this reaction the [KCl] is significantly lower than the minimum (~100 mM KCl) that is usually required to stabilise G-quadruplex structures of a comparable length (Antonacci et al., 2007, Lane et al., 2008). Although, other studies suggest that KCl concentrations as low as 50 mM are able to facilitate G-quadruplex formation even with relatively short ssDNA sequences *in vitro* (Gray and Chaires, 2008).

Taken together, the data suggest that it may be possible for the telomeric ssDNA substrate to form (at least partially) into G-quadruplex structures in the given reaction conditions. This might provide as explanation as to why the observed SNM1B activity is decreased on the telomeric ssDNA substrate when compared to the control ssDNA substrate (Figure 3.30). However, given that the presence of

secondary G-quadruplex structure was not assessed, the preceding paragraphs only offer a possible explanation as to the experimental results.

3.2.7. Analysis of WT and H61Y SNM1B on 8-oxo-guanine containing substrates

The current literature provides a significant amount of evidence that SNM1B plays a role in the repair of DNA damage lesions *in vivo.* The H61Y mutation did not affect the DNA binding and nuclease activity of SNM1B on undamaged substrates; however, damaged lesions may be more difficult to process. As such, the ability of WT and H61Y SNM1B to digest past an 8-oxo-guanine (8-oxo-G) *in vitro* was examined (Figure 3.31).

It is unlikely that SNM1B plays a specific role in the base excision repair of 8-oxo-G lesions (as these are easily processed by the DNA glycosylase, OGG1). However, 8-oxo-G is one of the most common types of DNA damage, readily arising from reactive oxygen species. Ionising radiation (IR) induced double strand breaks (DSBs) often have chemically modified, or 'dirty ends,' some of which may contain an 8-oxo-G (amongst other lesions). It has also been reported that SNM1B depleted human cells exhibit hypersensitivity to IR (Demuth et al., 2004).

The oxidation of guanine to 8-oxo-guanine is a simple chemical modification (Figure 3.31). However, it is enough to inhibit the nuclease activity of the 5' to 3' exonuclease EXO1, whilst SNM1B, and the related exonuclease SNM1A, were previously shown to efficiently digest 8-oxo-G containing substrates (Figure 3.31).

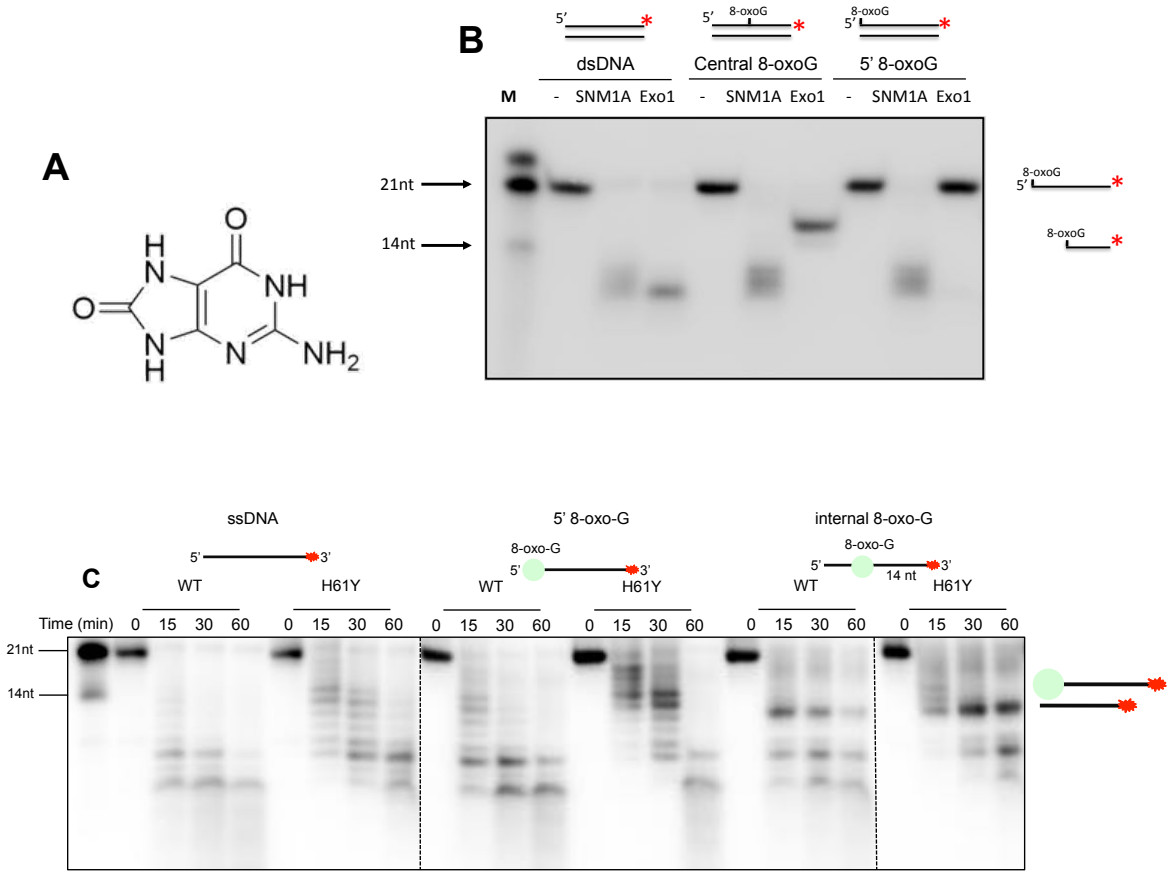


Figure 3.31 Both WT and H61Y SNM1B possess the ability to digest past a 5' 8-oxo-G or an internal 8-oxo-G.

A: schematic of the structure of 8-oxo-guanine. B: EXO1 is inhibited by the inclusion of an 8-oxo-G in a dsDNA substrate. EXO1 is unable to initiate digestion on a substrate containing a 5' 8-oxo-G, and is unable to bypass an internal 8-oxo-G, pausing at the site of the lesion, 15 nt from the 3' end. Experimental data from Dr Sook Yee Lee. C: nuclease assay showing digestion of ssDNA over time. 100 nM of DNA substrate (either undamaged or containing an 8-oxo-G as indicated) was added to 1.0 nM WT or H61Y SNM1B enzyme and incubated at 37°C for increasing time. The approximate size of a substrate digested up to the point of the 8-oxo-G is indicated, as is digestion past the 8-oxo-G. Samples were analysed by 20% polyacrylamide gel electrophoresis and each gel is representative of three individual experiments from two separate purifications.

Both WT and H61Y are able to effectively digest ssDNA substrates containing either a 5' 8-oxo-G or an internal 8-oxo-G, at a rate only slightly slower than that for undamaged ssDNA (Figure 3.31). The nuclease activity appears to be comparable between WT and H61Y SNM1B on all substrates, with the exception of perhaps slightly decreased activity for H61Y SNM1B on the 5' 8-oxo-G substrate. Interestingly, when acting on the internal 8-oxo-G substrate, both WT and H61Y SNM1B appear to pause immediately after the site of the lesion, at ~14 nucleotides form the 3' label. Given the nature of the DNA binding pocket of SNM1B (discussed in detail in Chapter 4), it is possible that a slightly larger damaged nucleotide would not be released quite as easily from the active site of the enzyme, thus precluding efficient rebinding and nuclease activity. However, this does not appear to be a significant hindrance to the activity of the enzyme.

3.2.8. Processivity of WT and H61Y SNM1B

The location of the His-61 is adjacent to the active site, and in the absence of a co-crystal with a DNA oligonucleotide, it is unknown how it might interact with a DNA substrate. It is clear from the results of this study that the H61Y mutation does not affect DNA binding or nuclease activity of SNM1B on the analysed substrates *in vitro.* However, how the H61Y mutation might affect the processivity of the SNM1B exonuclease on the DNA substrate was not known. As such, a competition-based processivity assay was performed.

The nuclease reaction (for both WT and H61Y SNM1B) was allowed to proceed for 10 minutes, before the addition of unlabelled DNA substrate, and then allowed to continue for a further 50 minutes (Figure 3.32).

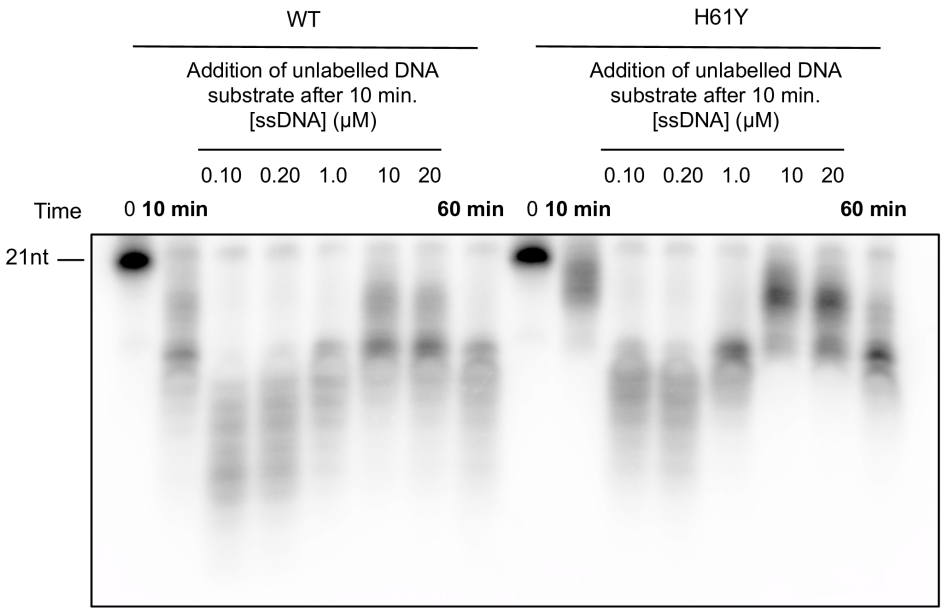


Figure 3.32 The nuclease activity of both WT and H61Y SNM1B is inhibited by the addition of 'cold' excess ssDNA substrate.

Gel based 'competition' assay whereby 100 nM of DNA substrate was added to 1.0 nM SNM1B and incubated at 37^{o}C for 10 min, at which point the reaction was either stopped, or increasing amounts of unlabelled ssDNA substrate was added to the reaction mix, and these were then incubated for a total period of 60 min. A single 60 min reaction was included where no additional unlabelled substrate was added. Samples were analysed by 20% polyacrylamide gel electrophoresis and each gel is representative of at least three individual experiments from two separate purifications.

In accordance with previous research (Sengerova et al., 2012), WT SNM1B shows little processivity, with the reaction being partially inhibited by the addition of 1.0 μM unlabelled ssDNA (a ten-fold molar excess), and completely inhibited by the

addition of 10 μM unlabelled substrate (a hundred-fold molar excess) (Figure 3.32). This phenomenon was also observed with H61Y SNM1B, and there was no appreciable difference between the two (Figure 3.32).

3.2.9. Substrate length preference of WT and H61Y SNM1B

WT SNM1B has previously been shown to have no activity on higher molecular weight (i.e. plasmid DNA) (Sengerova et al., 2012) and this is also evident in the present study, where neither WT nor H61Y exhibited digestion of a 2.7 kb plasmid DNA substrate that was either supercoiled, gapped (by prior treatment with Nb.BbvCI), or linearised (by prior treatment with HindIII) (Figure 3.33).

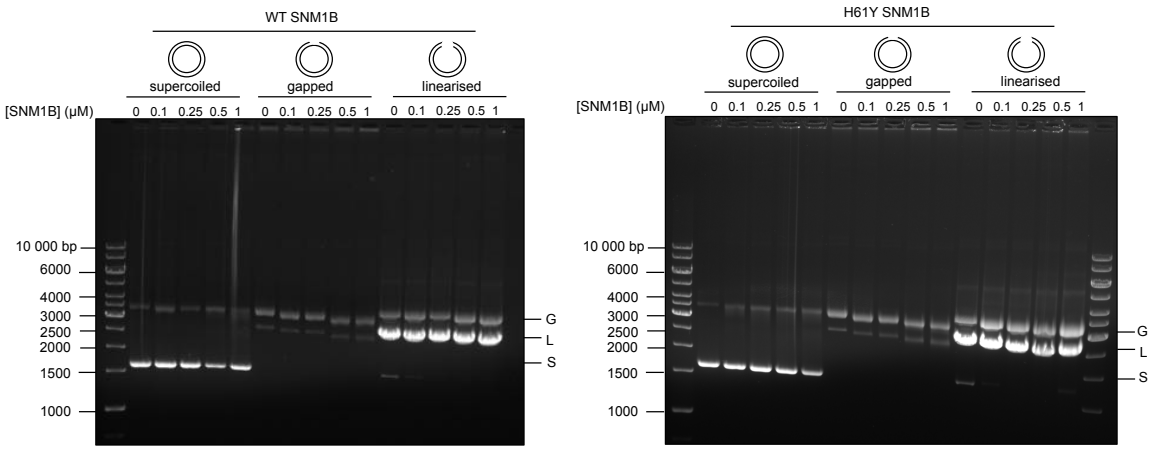


Figure 3.33 Neither WT nor H61Y SNM1B are able to effectively digest higher molecular weight DNA substrates.

Nuclease assay showing activity of WT and H61Y SNM1B on supercoiled (S), gapped (G), or linearised (L) 2.7 kb plasmid DNA. Gapped and linearised plasmids were priorly treated with either Nb.BbvCI or HindIII respectively. 325 ng of plasmid DNA was incubated with 10 nM SNM1B at 37^{o}C for 1 h. Samples were analysed on a 1% agarose gel, stained with ethidium bromide.

Interestingly, both WT and H61Y SNM1B show more activity on a shorter (21 nt) than a longer (61 nt) ssDNA substrate (Figure 3.34). There appears, however, to be equivalent nuclease activity between WT and H61Y SNM1B on both substrates (Figure 3.34).

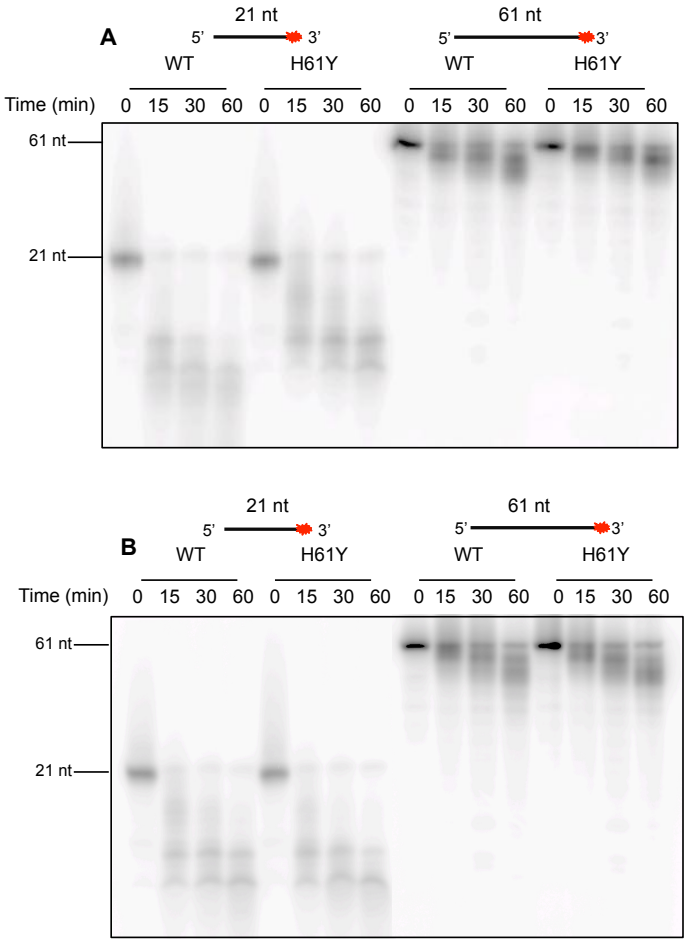


Figure 3.34 Both WT and H61Y SNM1B digest more efficiently a shorter ssDNA substrate (21 nt) than a longer substrate (61 nt).

Nuclease assay with two different purifications (designated A and B) of WT and H61Y showing digestion of ssDNA over time. 100 nM of DNA substrate (either 21 nt or 61 nt) was added to 1.0 nM WT or H61Y SNM1B enzyme and incubated at 37°C for increasing time. Samples were then run on a 20% polyacrylamide gel. A and B present two separate purifications of each WT and H61 SNM1B, and each gel is representative of at least three individual experiments.

Why SNM1B is able to more effectively digest shorter substrates (20–50 nt) than longer ones (upwards of two kilobases) is not immediately obvious. It was posited by (Sengerova et al., 2012) that this was due to the lack of a positively charged 'face' in the SNM1B enzyme, that was present in the related exonuclease, SNM1A, and was hypothesised to explain the latter's enhanced processivity on higher molecular weight DNA and effective digestion of plasmid based substrates. However, this characteristic of SNM1B remains to be further understood.

3.3. Conclusions

This study describes an effective method for purifying WT and mutant ΔC-SNM1B from Sf9 insect cells with high purity and yield. These can be subsequently utilised in biochemical, biophysical, and structural studies.

The biochemical and biophysical characterisation showed no demonstrable difference between WT and H61Y SNM1B. The thermal stability of the two SNM1B proteins was very similar (Figure 3.25) indicating no major structural disruption. There was also no significant difference with regards to DNA binding (Figure 3.27), enzyme kinetics (Figure 3.28), or nuclease activity (Figure 3.26). WT and H61Y SNM1B also showed a similar ability to effectively digest a variety of structurally diverse DNA substrates (Figure 3.29), telomeric substrates (Figure 3.30), and those that contain an 8-oxo-G (Figure 3.31). Both enzymes showed little processivity on a 21 nt substrate (Figure 3.32) and were equally unable to digest higher molecular weight substrates (2.7 kb plasmid DNA; Figure 3.33).

All of these analyses were performed *in vitro* with purified, truncated SNM1B protein. However, given that truncated SNM1B (similar to the amino acid truncation used in this study) has previously been shown to have comparable nuclease activity to the full-length protein (Sengerova et al., 2012) it does seem likely that there is little difference between WT and H61Y SNM1B with regards to DNA binding, and nucleolytic processing of substrates. This also seems probable to be the case *in vivo*.

The experimental interest in the H61Y SNP (rs11552449) was piqued due to a meta-analysis of GWAS which found this SNP was associated with an increased risk of breast cancer (Michailidou et al., 2013). However, during the course of this study, two further GWAS (Karami et al., 2016, Caswell et al., 2015) were published, both of which examined the relationship between this SNP and cancer risk. One of these (Karami et al., 2016) found no statistical significance between rs11552449 and cancer risk. Interestingly, the other study (Caswell et al., 2015), noted that the rs11552449 SNP was significantly associated with differential transcript isoforms of the *SNM1B* gene, and the subsequent inclusion of exon 2, and this inclusion of exon 2 was positively associated with a significantly increased risk of breast cancer. However, their study was somewhat confounded by the fact that rs11552449 also affected splicing of the nearby gene, *PHTF1*. rs11552449 was associated with the inclusion of exons 1 and 2 in the transcript of *PHTF1*, and this also was associated with an increased risk of breast cancer. Thus, while this SNP appears to have an effect in mediating cancer risk they could not delineate which of the two genes was causally associated (Caswell et al., 2015).

It may therefore be that rs11552449 is not associated with increased cancer risk because of the amino acid substitution in the metallo-β-lactamase (MBL) domain of SNM1B, but rather due to differential splicing effects. However, there is still a possibility that the H61Y SNM1B protein has some cellular effects, such as altering protein-protein interactions, or *in vivo* protein stability. The only protein-protein interaction presently mapped to this region, is the interaction with astrin, which binds to the MBL domain of SNM1B (Liu et al., 2009).

The ability of SNM1A and SNM1B to digest a variety of diverse DNA substrates, particularly those that contain DNA damage, is a distinguishing feature of these nucleases. Prior crystallographic analysis (Allerston et al., 2015) suggested that this might be due to the 'open' nature of the active site, and this will be discussed further in Chapter 4.

The results from this study raise some interesting questions about the biochemical role of SNM1B at telomeres. The ability of SNM1B to digest telomeric substrates is relevant (Figure 3.30), and further work to elucidate whether SNM1B can digest telomeric G-quadruplex containing structures is pertinent. The preference of SNM1B for digesting ssDNA substrates over other structures (Figure 3.23 and Figure 3.29) raises the possibility that SNM1B may function *in vivo* with other proteins that affect the structure of DNA, for example the helicase, RTEL1, which can separate the two strands of DNA, disassemble t-loops, and destabilise G-quadruplex structures (Vannier et al., 2012).

4. Crystallographic analysis of WT and H61Y SNM1B

4.1. Introduction

As discussed in Chapter 1, SNM1B is a member of the Group 7 category of the metallo-β-lactamase (MBL) superfamily, and is further categorised into the β-CASP family of enzymes (so named for its constituent members, CPSF, Artemis, SNM1, and PSO2). Accordingly, it contains an MBL fold, consisting of four conserved motifs, as well as the distal β-CASP domain, comprising three conserved motifs (Allerston et al., 2015, Callebaut et al., 2002). These two domains confer both the DNA binding and catalytic activity of SNM1B and encompass amino acids 1–299 of the SNM1B protein. The remaining characterised domain, the TRF2 binding motif (TBM) is mapped to amino acids 496–511 (Lenain et al., 2006, van Overbeek and de Lange, 2006, Demuth et al., 2008, Freibaum and Counter, 2008, Chen et al., 2013).

The structures of these different domains of the WT SNM1B protein have been solved by X-ray crystallography. The structure of an N-terminal peptide of SNM1B (amino acids 492–532) has been solved in complex with its binding partner, TRF2 (Chen et al., 2013). Additionally, the catalytic core of SNM1B (amino acids 1–335, encompassing both the MBL and β-CASP domains) has also been crystallised by previous members of the Gileadi and McHugh labs (Allerston et al., 2015). This structure (PDB: 5AHO) shows the two catalytic domains – the MBL domain, and the β-CASP domain. The conformation of the MBL domain is typical to other members of the superfamily and consists of a four-layered β-sandwich (α/β-β/α)

with two mixed β-sheets, flanked by two α-helices on each side. The β-CASP domain consists of a four-stranded parallel β-sheet, flanked by three α-helices on one side, and one on the other, is inserted between strands 10 and 11 of the MBL domain (Allerston et al., 2015).

The aforementioned structure of SNM1B (which will henceforth be referred to by its PDB ID: 5AHO) indicated the location of the His-61 residue as adjacent to the active site (Figure 4.1). The role of the H61Y breast cancer associated variant has been discussed extensively in the previous chapter, and biochemical analyses led to the conclusion that this mutation did not significantly affect DNA binding or nuclease activity *in vitro*. However, the effect of the H61Y variation on the structure of SNM1B was unknown, and thus this was investigated, particularly given the proximity to the active site.

Whilst the structure of SNM1B has previously been reported, there has been no described co-crystal structure of SNM1B bound to a DNA substrate. Of the eukaryotic β-CASP subgroup of nucleic-acid processing MBLs, there have been a handful of crystal structures solved – human SNM1A, human SNM1B, human CPSF-73, and yeast CPSF-100 (Allerston et al., 2015, Mandel et al., 2006). However, to date, no co-crystal structures of any SNM1 family members with the relevant catalytic substrate have been described. Thus, achieving a co-crystal of SNM1B with DNA was one of the aims of this current project.

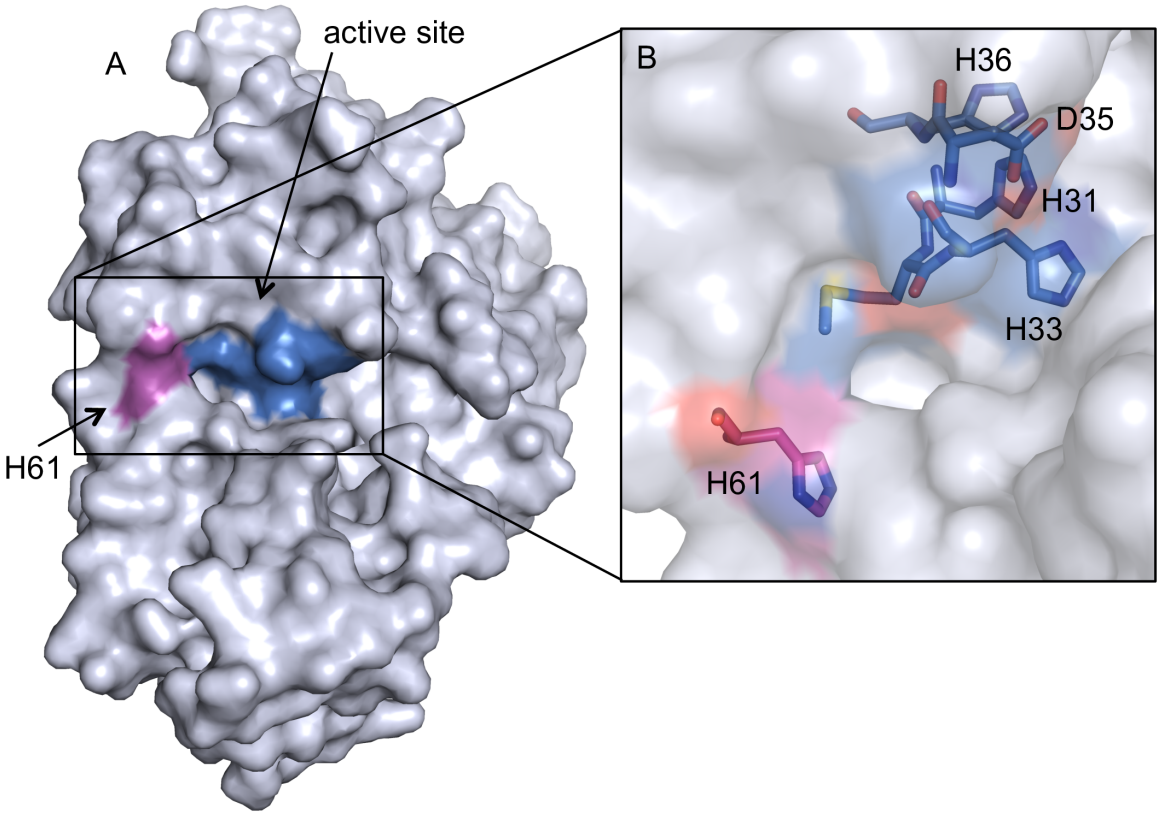


Figure 4.1 Surface representation of WT SNM1B (PDB: 5AHO) showing the location of the H61 residue.

A: the H61 residue is indicated in pink, and the active site in blue. The represented active site is the conserved HxHxDH motif involved in co-ordinating the metal ions. B: closer view of H61 and the active site with relevant amino acid residues labelled.

4.2. Results

4.2.1. Generation of WT and H61Y SNM1B crystal structures

For the purposes of crystallisation, both WT and H61Y SNM1B were purified as outlined in Chapters 2 and 3. Crystallisation screening was performed using the sitting drop vapour diffusion method and five core screens from the SGC were

used to screen for potential hits. These are summarised in (Ng et al., 2016) and were as follows: the Hampton Research Index Screen; Hampton Research Crystal Screen (both from Hampton Research); JCSG+ Screen (Molecular Dimensions, modified); Ligand Friendly Screen (LFS, designed in-house); Basic ChemSpace Screen (designed in-house). For crystallisation, 50–100 nL of WT or H61Y SNM1B was mixed with 100–50 nL reservoir solution, SNM1B protein concentration was varied between 10 and 20 mg/mL, and plates were incubated at both 4^{o}C and 20^{o}C. The initial screening process identified potential hits and follow-up screens were designed around these conditions.

Following optimisation, crystals that diffracted to the highest quality were grown in the following conditions. WT SNM1B: 6% PEG 20 K, 0.1 M bicine, pH 7.3. H61Y SNM1B: 24% PEG 3350 K, 0.1 M HEPES, pH 7.0, 0.2 M sodium formate. The protein concentrations were 15 mg/mL and 12 mg/mL for WT and H61Y, respectively, and both were incubated at 4^{o}C (Figure 4.2). Subsequently, crystals were mounted and taken to the Diamond Light Source Synchrotron at Harwell, and data sets were collected from those that diffracted to the highest resolution. Dr Joseph Newman, SGC, performed crystal mounting and data collection.

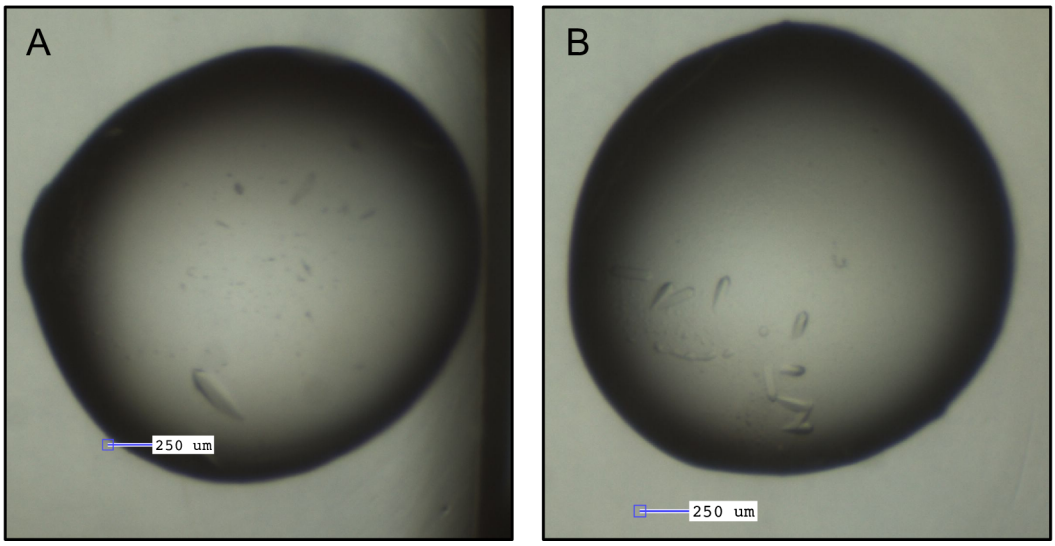


Figure 4.2 WT SNM1B and H61Y SNM1B crystals growing in a sitting drop arrangement. A = WT SNM1B (form I); B = H61Y.

4.2.2. Comparison of the crystal structure of WT and H61Y SNM1B

The crystals obtained in 4.2.1 diffracted to 2.8 Å and 3.2 Å for WT and H61Y, respectively. The structure of each WT and H61Y SNM1B was solved by Dr Joseph Newman, using molecular replacement, with the previously deposited structure of WT SNM1B (PDB: 5AHO) used a search model. For clarity, the newly solved SNM1B structures will be referred to as follows: WT (form I) and H61Y.

Initial analysis of the structures of WT (form I) and H61Y revealed that the overall architecture between the two was extremely similar with a root mean squared deviation (RMSD) of 0.48 Å (312 residues to 312 residues). The two structures are overlaid in Figure 4.3, and they exhibit a very similar global conformation. Both WT (form I) and H61Y possess two primary structural domains: the characteristic MBL fold; and the β-CASP domain, the latter typical to nucleic acid processing MBLs.

The structure around the active site is also very similar, and there appears to be no disruption in the structure of H61Y compared to WT. There is no observed effect on the neighbouring residues (R60, R206, E209), no perturbation to the α-helix that Y61 resides in, and also no disruption to the adjacent helix, which contains many of the nearby residues (Figure 4.3).

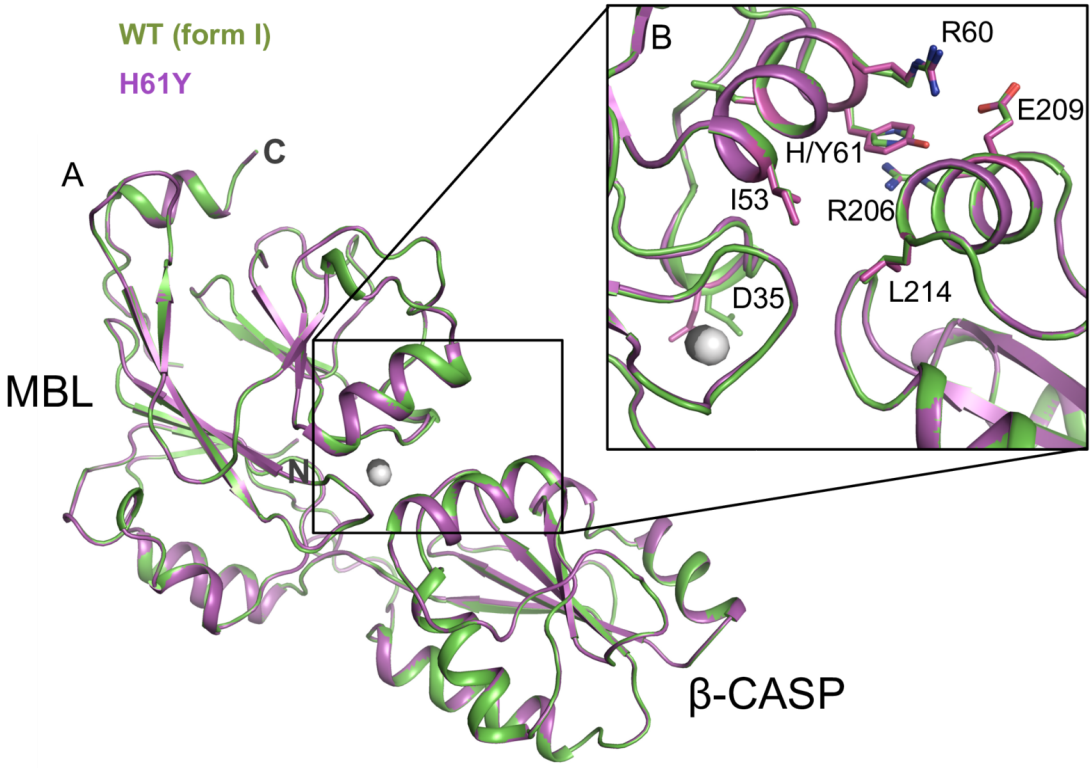


Figure 4.3 Overlay of WT (form I) and H61Y SNM1B represented in cartoon form.

WT SNM1B (form I) is represented in green, and H61Y in magenta. The active site metal ion is represented in light grey for WT SNM1B, and dark grey for H61Y SNM1B. The N- and C-termini are as indicated. A shows the overall structure of WT and H61Y SNM1B and the MBL and β-CASP domains are as indicated. B shows the location and conformation of H/Y61 and neighbouring residues and regions.

Taken together with the results of the previous chapter, the data suggest that the H61Y mutation has little effect on the structural conformation of SNM1B, and thus, has no discernible effect on the DNA binding or nuclease activity. However, there is of course the caveat that these results are all from *in vitro* experiments and may not necessarily reflect normal cellular processes. It is also important to note that mutations can have effects on the energetics of binding or catalysis and this might not be observed in a static crystal structure. Nevertheless, it does seem likely that the H61Y SNM1B has equivalent structural characteristics, DNA binding, and nuclease activity to WT SNM1B. As WT and H61Y SNM1B are very similar structurally, further analysis focussed on the WT SNM1B structure obtained in this study.

The MBL and β-CASP catalytic domains for WT (form I) are arranged as in previous WT SNM1B structural analyses, and this is displayed in Figure 4.4.

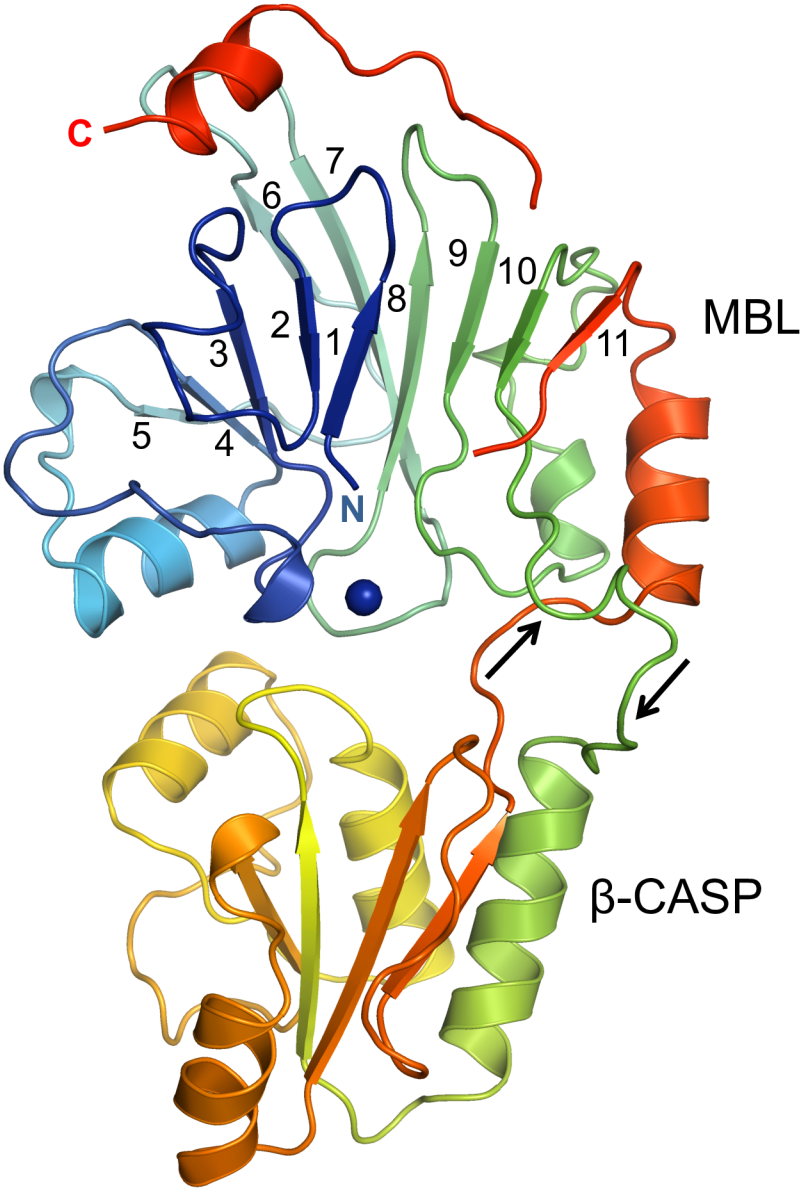


Figure 4.4 Overall structural features of WT (form I) shown in cartoon representation.

The N- and C- termini are as indicated, and the structure is accordingly coloured from blue to red. The nickel ion is indicated as a blue sphere. The 11 β-strands of the MBL domain are as labelled, the β-CASP domain is inserted between strands 10 and 11, and the directionality is indicated by the arrows.

The WT (5AHO) and WT (form I) structures are overlaid in Figure 4.5, and they are highly similar. This similarity is reflected in an RMSD of 1.27 Å (314 residues to 312 residues). This characteristic structural arrangement of the MBL and β-CASP domains is similar to the structures observed for other MBL-β-CASP family members including SNM1A (Allerston et al., 2015), CPSF-73 (Mandel et al., 2006), and RNaseJ (Newman et al., 2011).

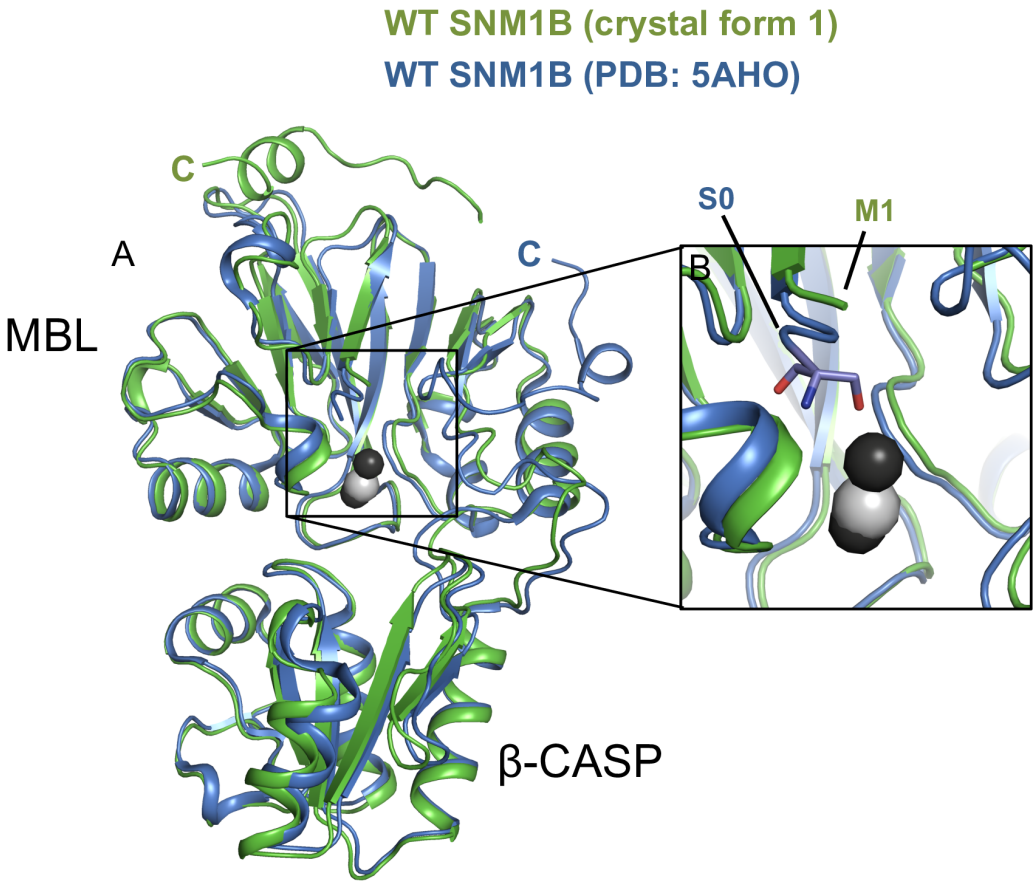


Figure 4.5 Overlay of WT (form I) and WT (5AHO) represented in cartoon form.

WT SNM1B (form I) is represented in green, and WT (5AHO) in blue. The zinc ions present in WT (5AHO) are indicated as black spheres, and the nickel ion present in WT (form I) as a grey sphere. A shows the overall conformation for both structures and the C-termini are as labelled. B shows the location and conformation of the N-terminus and relevant amino acid residues. S0 indicates the serine prior to the start methionine in WT (5AHO), and M1 denotes the start methionine in WT (form I).

However, there are two appreciable differences between the WT (form I) structure obtained in this study, and WT (5AHO). Firstly, there is a difference between the two structures in the positioning of the C-terminus. The two structures are highly similar up to the point of isoleucine 297, and thereafter they diverge. In the WT (5AHO) structure, the sequence forms a loop and an α-helix, however, after R313 the structure is poorly resolved, presumably because it is disordered, and is thus not visible in the electron density. Conversely, for WT (form I), the structure between I297 and I314 is not resolved, but after I314 there is an extended loop, followed an α-helix that forms part of the MBL domain. This helix sits between the two sheets of the β-sandwich, contacting β-strands two and three; the strand then terminates at S329 (Figure 4.5). This observed difference between WT (5AHO) and WT (form I) in the positioning of their C-termini is likely due to protein construct differences. WT (5AHO) was purified with an N-terminal hexa-histidine tag, whereas WT (form I) was purified with a C-terminal deca-histidine tag; and both constructs spanned amino acids 1–335 in the SNM1B sequence. In each instance the tag was cleaved in the purification process; however, amino acids from the TEV cleavage recognition sequence subsequently remain. Therefore, the WT (form I) protein contains an additional seven amino acids at the C-terminus, and this likely affects the structural conformation in this region. This is unlikely to have a significant effect on DNA binding and nuclease activity, as this region is located away from the active site. Accordingly, both WT SNM1B proteins exhibit robust nuclease activity (as evidenced by the previous chapter, and (Allerston et al., 2015)).

Secondly, there was a difference in the metal ion co-ordination, and the conformation of the N-terminus. WT (5AHO) was modelled with two Zn^{2+} ions coordinated in the active site (Allerston et al., 2015); however, this was not observed in the current study. The electron density map for WT (form I) revealed that only one metal binding site was occupied, and an X-ray fluorescence (XRF) scan of the crystal revealed a mixture of iron, nickel, and zinc. As nickel gave a slightly more prominent peak in the XRF spectra it was subsequently modelled in the active site (Figure 4.4 and Figure 4.6).

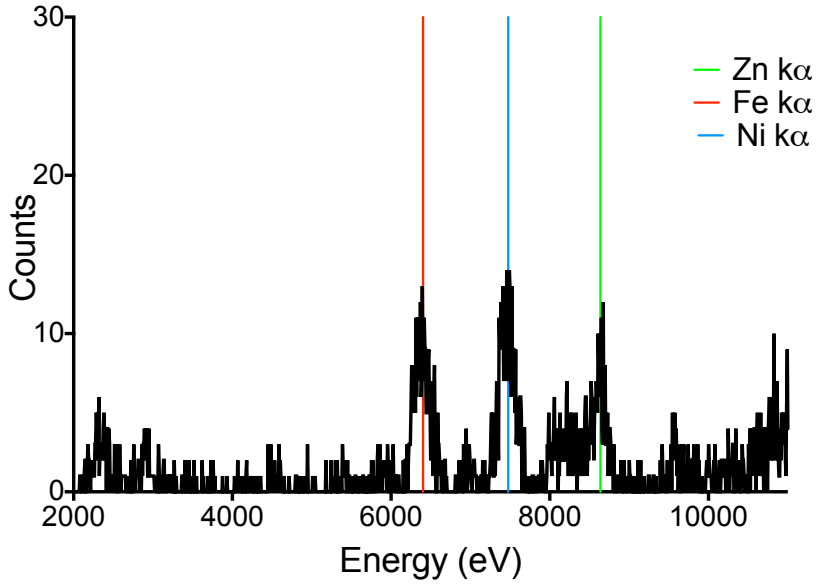


Figure 4.6 X-ray fluorescence analysis for the WT (form I) crystal.

The k-α emission peaks are labelled for zinc, iron, and nickel.

It may be that this difference in metal ion coordination between the two structures was influenced by the slightly different conformation of the N-terminus. As previously, WT (5AHO) and WT (form I) had different histidine tags for purification. WT (5AHO) contained an N-terminal tag and, as such, an additional serine

residue, constituent of the TEV recognition sequence is present immediately prior to the start methionine. This serine is located closely to D35, which coordinates the second metal ion (Figure 4.5). In contrast, the sequence of WT (form I) simply begins with the start methionine, and this is positioned slightly differently in the structure. Thus, it might be that the respective location of these N-terminal amino acid residues affects the coordination of the active site metal ion.

Nonetheless, these differences in metal ion coordination may also simply reflect different purifications and the inherent variability in this process, or that these were two different crystal forms, grown in different conditions. Both WT SNM1B purifications possess robust nuclease activity and this does not appear to be affected by the observation that there are differing catalytic metal ion(s) present. However, it's possible the sub-set of the SNM1B proteins that crystallise may not be representative of the WT SNM1B population in solution. Accordingly, the relevant metal ion(s) requisite for SNM1B catalysis will be investigated further in 4.2.11.

4.2.3. Generation of a WT SNM1B:DNA co-crystal structure

At the time of this study there were no reported structures of any of the SNM1 family members co-crystallised with the relevant nucleic acid substrate. As such, a co-crystal structure of SNM1B with DNA may provide specific information regarding DNA binding, catalytic mechanism, and the importance of specific residues, and regions, around the active site. Thus, the intention was to obtain a

co-crystal structure of SNM1B with DNA, and several approaches were taken to achieve this.

Firstly, it was attempted to soak single stranded (ss-) DNA substrates into already formed SNM1B protein crystals. As outlined in 4.2.1, a method for crystallising the apo SNM1B protein had already been developed. Accordingly, DNA substrates between 5 and 12 nucleotides in length were ‘soaked’ into the already formed crystals. However, this was unsuccessful as the subsequent crystals no longer diffracted after soaking. Presumably this was due to DNA binding inducing a conformational change in the structure of SNM1B and thus disrupting the crystal lattice, or the soaking procedure itself damaged the diffraction of the crystal.

Crystallisation of an SNM1B-DNA complex was also attempted. As the optimal DNA substrate for co-crystallisation was unknown, it was attempted with a range of different oligonucleotides. Single stranded DNA sequences varied in length from 5–12 nucleotides, and double stranded and overhang DNA structures varied between 8–15 nucleotides. However, this approach raised several problems. For optimal crystallisation conditions, the solution is required to be near homogeneous (Benvenuti and Mangani, 2007). Over time, SNM1B will catalytically digest a DNA substrate, thus rendering the solution heterogeneous (or depleting the DNA entirely); therefore, several attempts were made to mitigate this. DNA substrates containing a 3’ hydroxyl group, as opposed a 3’ phosphate group were incubated with SNM1B, as SNM1B will weakly bind to, but not digest DNA with a 3’ hydroxyl moiety (Figure 4.7 and Figure 4.8).

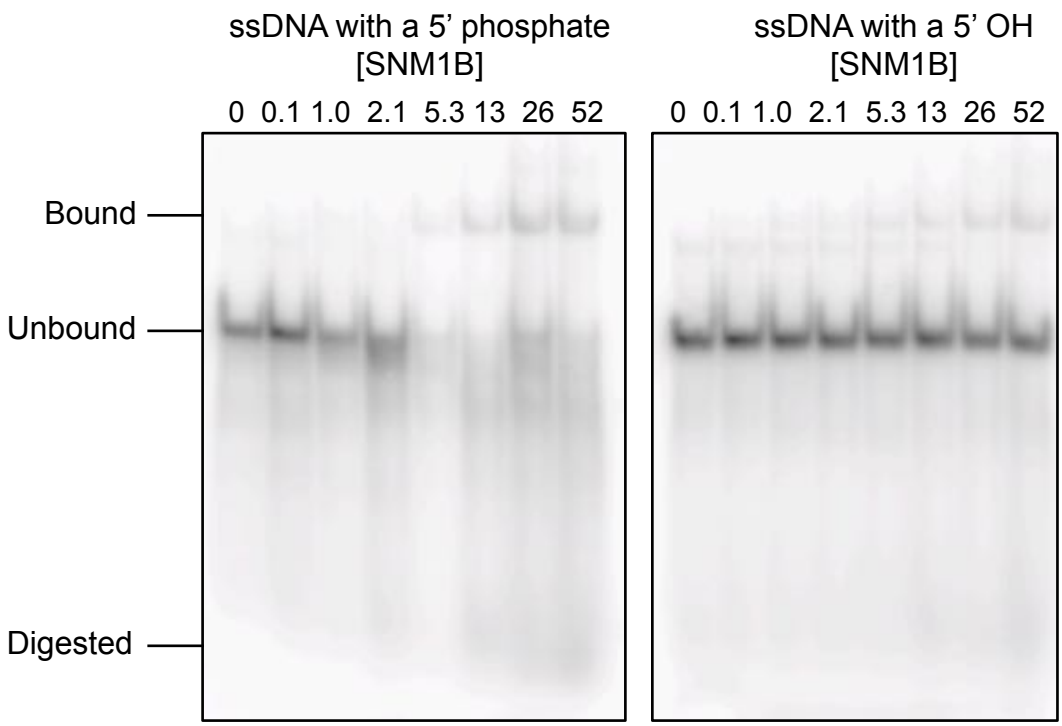


Figure 4.7 WT SNM1B preferentially binds to ssDNA with a 5’ phosphate, compared with a ssDNA with a 5’ hydroxyl moiety.

EMSA showing binding of WT SNM1B to ssDNA substrate either with a 5’ phosphate group, or a 5’ hydroxyl group. 1 nM of the appropriate DNA substrate was incubated for 5 minutes at 37^{o}C with WT SNM1B (concentrations as labelled), before being analysed by 10% native polyacrylamide gel electrophoresis. Unbound DNA substrate, DNA substrate bound to SNM1B, and digestion products are as labelled. Each gel is representative of at least three individual experiments.

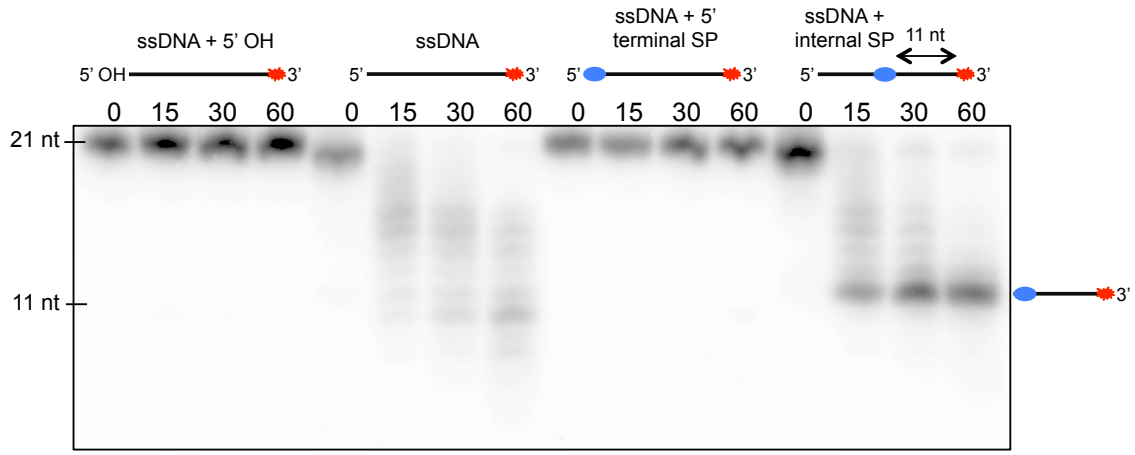


Figure 4.8 WT SNM1B will exonucleolytically process ssDNA with a terminal 5' monophosphate, but not a ssDNA substrate with a terminal 5' hydroxyl group, nor past a phosphorothioate bond.

100 nM of DNA substrate (as labelled) was added to 0.8 nM SNM1B and incubated at 37°C for increasing time. Samples were then analysed by 20% polyacrylamide gel electrophoresis. DNA substrate constituted the following (as labelled left to right above): ssDNA with a 5' hydroxyl group; ssDNA with a 5' phosphate group (standard ssDNA); ssDNA with a 5' terminal phosphorothioate bond (denoted SP); ssDNA with an internal phosphorothioate bond, 11 nucleotides from the 3' end. Each gel is representative of at least three individual experiments.

Additionally, phosphorothioate containing oligonucleotide sequences were also utilised. Phosphorothioate bonds can be substituted for phosphodiester bonds in the backbone of the DNA substrate by replacing the non-bridging oxygen atom of the sugar moiety with a sulphur atom. This substitution renders the DNA substrate non-hydrolysable (at that point) to many nucleases (Stein et al., 1988), including SNM1B. SNM1B does not show any digestion of a ssDNA substrate that contains a phosphorothioate bond between the first and second nucleotide; and nor will

SNM1B digest past an internal phosphorothioate bond (Figure 4.8). Thus, phosphorothioate bonds were substituted at different points within the oligonucleotide sequence. It was hypothesised that SNM1B may catalytically process up to the point of the non-hydrolysable bond and 'pause' thereby allowing the digestion to be captured at a homogeneous point. However, there is also the consideration that SNM1B binding to DNA is relatively transient (as shown by the processivity analysis in the previous chapter), thus meaning that SNM1B may release the DNA at any point, including when it is no longer a catalytic substrate.

It has previously been reported that crystallisation of a stable protein-DNA complex can be optimised by altering the metal ions present in the crystallisation solution, or by the addition of chelating agents (Hollis, 2007). This is based on the premise that these conditions facilitate binding, but diminish catalytic degradation of the DNA substrate. However, SNM1B is able to catalytically digest a ssDNA substrate in the presence of a variety of metal ions (Mg^{2+}, Mn^{2+}, Zn^{2+}, Ca^{2+}), and also when no exogenous metal ion has been added to the reaction buffer (pilot experiments, data not shown; and (Sengerova et al., 2012)). Previous analysis (Sengerova et al., 2012), had shown that concentrations of $ZnCl_2$ above 0.1 mM inhibited the *in vitro* nuclease activity of SNM1B (when [SNM1B] was 43.8 nM); however, in the current study, fluorescence anisotropy based DNA binding assays showed that the addition of $ZnCl_2$ also inhibited DNA binding (Figure 4.9). Whether this is due to Zn^{2+} specific interactions with the SNM1B metallo-binding core, non-specific interactions with the protein structure, or interactions between Zn^{2+} ions and the DNA substrate, is not clear. There is some evidence that Zn^{2+} ions are able to cause protein precipitation *in vitro* (Lovrien and Matulis, 2001, Matulis,

2016), and given that Zn^{2+} ions were present in molar excess to SNM1B this may be a factor (Figure 4.9 and (Sengerova et al., 2012)). There is also some *in vitro* evidence that the positively charged Zn^{2+} can interact with the negatively charged phosphodiester backbone of the DNA, and thus alter its conformation (Ihara et al., 2015, Morris, 2014).

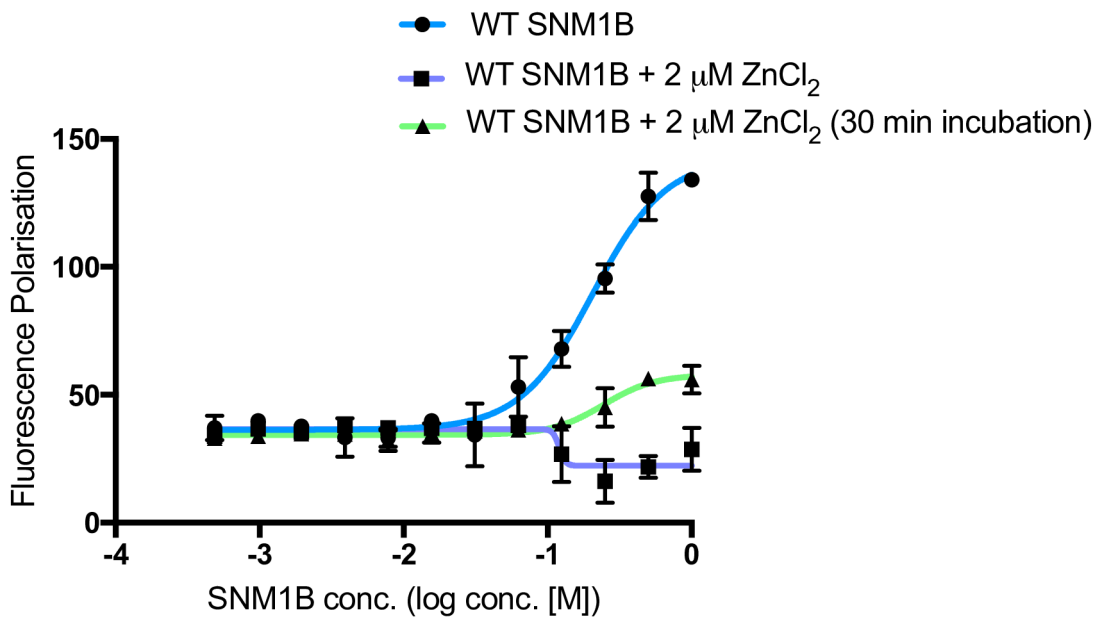


Figure 4.9 WT SNM1B binding to DNA is inhibited by the addition of 2 μM $ZnCl_2$.

1 μM of 20 nucleotide ssDNA was incubated with serially diluted concentrations of SNM1B (as indicated) for five minutes at room temperature. The solution either contained no divalent metal ion (WT SNM1B) or 2 μM $ZnCl_2$. In the case of WT SNM1B + 2 μM $ZnCl_2$ (30 min incubation), 2 μM WT SNM1B was incubated with 2 μM $ZnCl_2$, before the excess $ZnCl_2$ was removed, and SNM1B was serially diluted as above. Each reaction was in quadruplicate and at least two experimental repeats were performed. Error bars indicate the SD.

Ultimately, none of the aforementioned methods for achieving a co-crystal of SNM1B with an oligonucleotide substrate were successful. Instead, a new

approach, whereby SNM1B was co-crystallised with a single nucleotide (2'-deoxyadenosine-5'-monophosphate) was exploited. As ligand binding to protein can result in a conformational change, and thus a different crystal form (Hassell et al., 2007), *ab initio* crystallisation conditions were optimised. This was performed in the same manner as in 4.2.1.

It is worthy of note that two different crystal forms grew from the same conditions, and often within the same drop – 'plate' crystals and 'cubic' crystals (Figure 4.10). These could be distinguished on the basis of their shape, and also by the fact that the cubic crystals (as a consequence of their isomeric nature) did not appear birefringent under cross polarisation. The plate crystals diffracted to a much higher resolution than the cubic crystals, and were thus used in further analysis.

Crystals were mounted and taken to the Diamond Light Source Synchrotron at Harwell, and data sets were collected from those that diffracted to the highest resolution. Data mounting and data collection was again performed in collaboration with Dr Joseph Newman, SGC. The crystal that diffracted to the highest resolution was grown from the following conditions: 20% PEG 3350 K, 0.2 M ammonium chloride. Prior to being mixed with the reservoir solution, 15 mg/mL WT SNM1B was incubated with 50 mM 2'-deoxy-5'-AMP (pH 7.5). Plates were incubated at 20^{o}C.

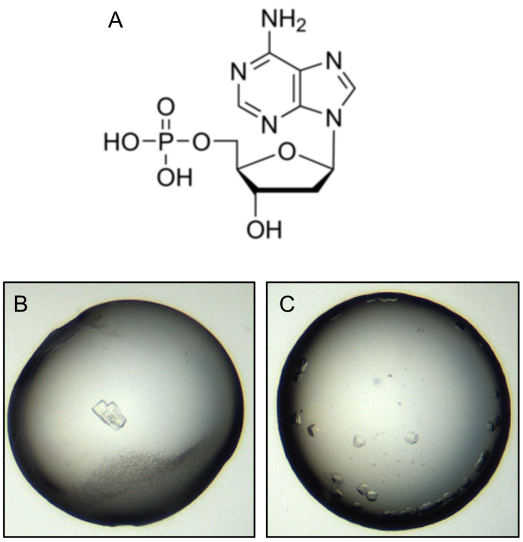


Figure 4.10 WT SNM1B was crystallised in the presence of 2'-deoxy-5'-AMP.

A: Schematic of the structure of 2'-deoxy-5'-AMP. B: WT SNM1B + AMP crystals growing in the plate form. C: WT SNM1B + AMP crystals growing in the cube form.

4.2.4. WT SNM1B structure with 2'-deoxy-5'-AMP

Of the crystals obtained above, the best diffracted to 1.8 Å. Dr Joseph Newman, SGC, solved the structure using molecular replacement, and the WT (form I) structure generated in 4.2.1 was used as a search model. It is worthy of note that these SNM1B crystals had a different appearance; and the crystal packing and space group differed from the previously obtained WT (form I) crystals. The structure obtained in 4.2.3, denoted WT (nucleotide form), also contained two 2'-deoxy-5'-AMP molecules in the active site. The data collection and refinement statistics for each WT (form I), H61Y, and WT (nucleotide form) can be found in Appendix VII.

Interestingly, it can be seen that there is no striking global conformational change that occurs as a result of nucleotide binding to SNM1B. The structures of WT (form I) and WT (nucleotide form) are overlaid in Figure 4.11, and the overall architecture is very similar. There is a slight shift in one of the helices in the β-CASP domain between the two WT structures, and the loop that connects the α-helix at the C-terminus is better resolved in the WT (nucleotide form) structure (Figure 4.11). However, these differences are relatively minor, and it appears that active site of SNM1B is relatively open in both the bound and unbound form.

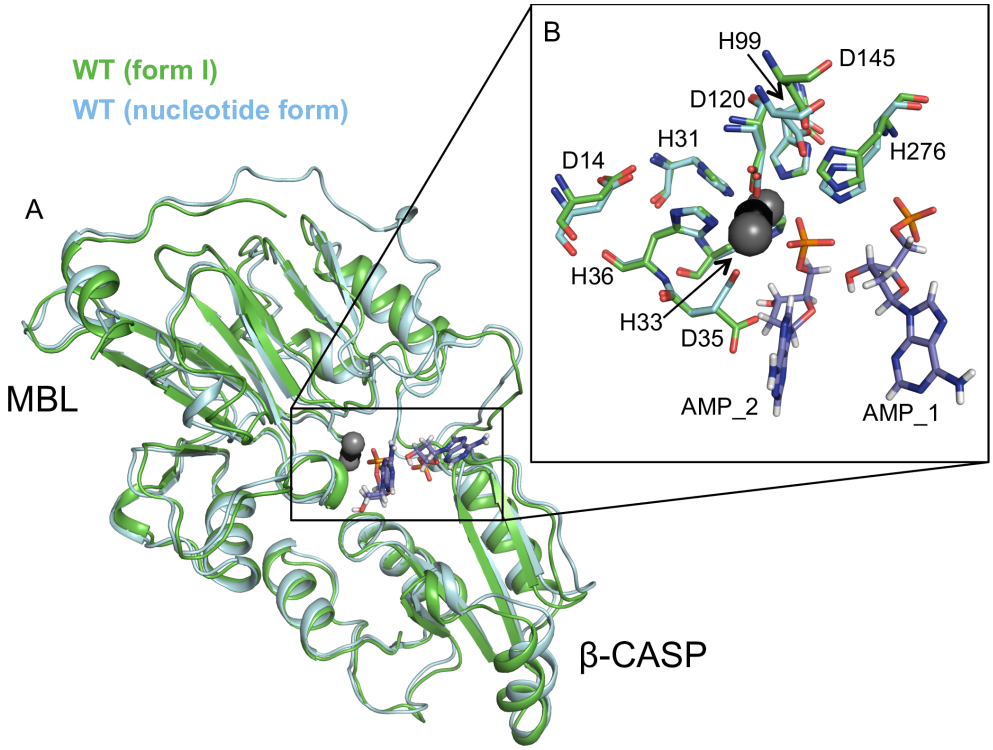


Figure 4.11 Overlay of the structures of WT (form I) and WT (nucleotide form).

A shows the overall structure of both WT SNM1B crystal forms. The MBL and β-CASP domains are as indicated. The iron(II) and nickel metal ions in crystal form 3 are represented as grey spheres, and the nickel ion present in crystal form 1, as a black sphere. The two 2'-deoxy-5'-AMP molecules are represented in stick form. B shows an enlargement of the active site of both WT SNM1B structures. The metal ion co-ordinating residues are as labelled.

There is, however, a difference in metal ion co-ordination between WT (form I) and WT (nucleotide form). As previously noted, in WT (form I) only one metal ion-binding site was occupied, and this was modelled as a Ni^{2+} ion. However, the electron density data for WT (nucleotide form) revealed that both metal ion-binding sites were occupied. XRF analysis revealed a mixture of predominantly iron and nickel present in the crystal, with only a very small amount of zinc (Figure 4.12). Based on the coordination environment and electron density levels, the final structure was modelled with both Ni^{2+} (present in the first metal ion-binding site) and Fe^{2+} (present in the second metal ion-binding site).

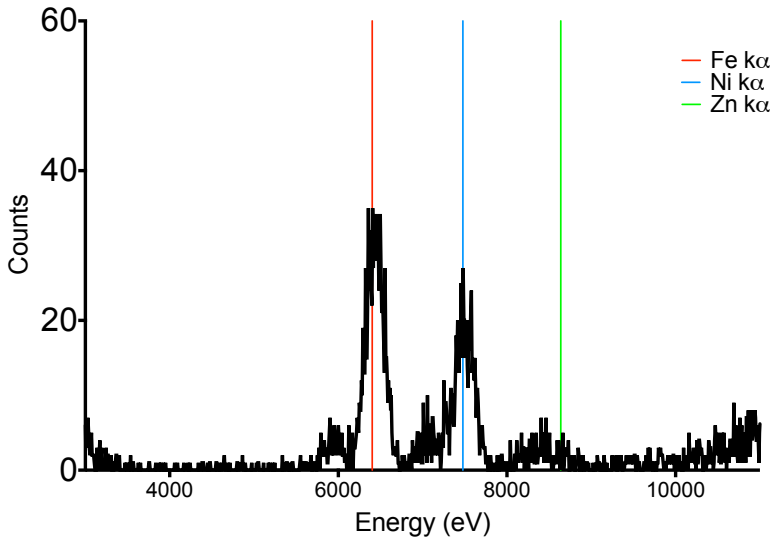


Figure 4.12 X-ray fluorescence analysis for the WT (nucleotide form) crystal.

The k-α emission peaks are labelled for zinc, iron, and nickel.

Interestingly, it can be seen in Figure 4.11, the positioning of the metal ion and the coordination profile differs when either one (as in WT, form I) or two metal ions (as in WT, nucleotide form) are present. In WT (nucleotide form), the Ni^{2+} ion is coordinated in the first metal ion-binding site by the side chains of H31, H33, H99, and D120; and the Fe^{2} ion is coordinated in the second metal ion-binding site by

the side chains of D120, D35, and H36 (Figure 4.11). Similarly, in WT (form I) the single Ni^{2+} ion is coordinated in the first metal ion-binding site. However, in the absence of the second metal ion-binding site being occupied, the Ni^{2+} receives an extra ligand from H36 and is shifted a little towards the second metal ion-binding site (Figure 4.11) In the structure of WT (form I) D35 is not involved in coordinating any metal ion and its side chain is subsequently oriented away from the metal binding site. Interestingly, when the two structures are overlaid (as in Figure 4.9), the D35 side chain of WT (form I) is located very close to the ribose sugar of AMP_2 in WT (nucleotide form). This, therefore, may occlude DNA binding when only the first metal ion site is occupied. However, whether the one metal bound form of SNM1B possesses catalytic activity will be discussed further in 4.2.11.

4.2.5. Understanding the interaction of SNM1B with DNA

The two 2'-deoxy-5'-AMP (AMP_1 and AMP_2) molecules are arranged in a similar way as might be expected for a two nucleotide ssDNA substrate. The distance between the two phosphate groups is 6.1 Å, and this is commensurate with the distance between two phosphate groups of adjacent nucleotides in DNA (approximately ~6.6 Å). As such, they can be regarded as an appropriate tool for understanding the DNA binding and catalytic mechanism of SNM1B. The electron density map of the two 2'-deoxy-5'-AMP molecules bound to the active site of WT SNM1B is portrayed in Figure 4.13.

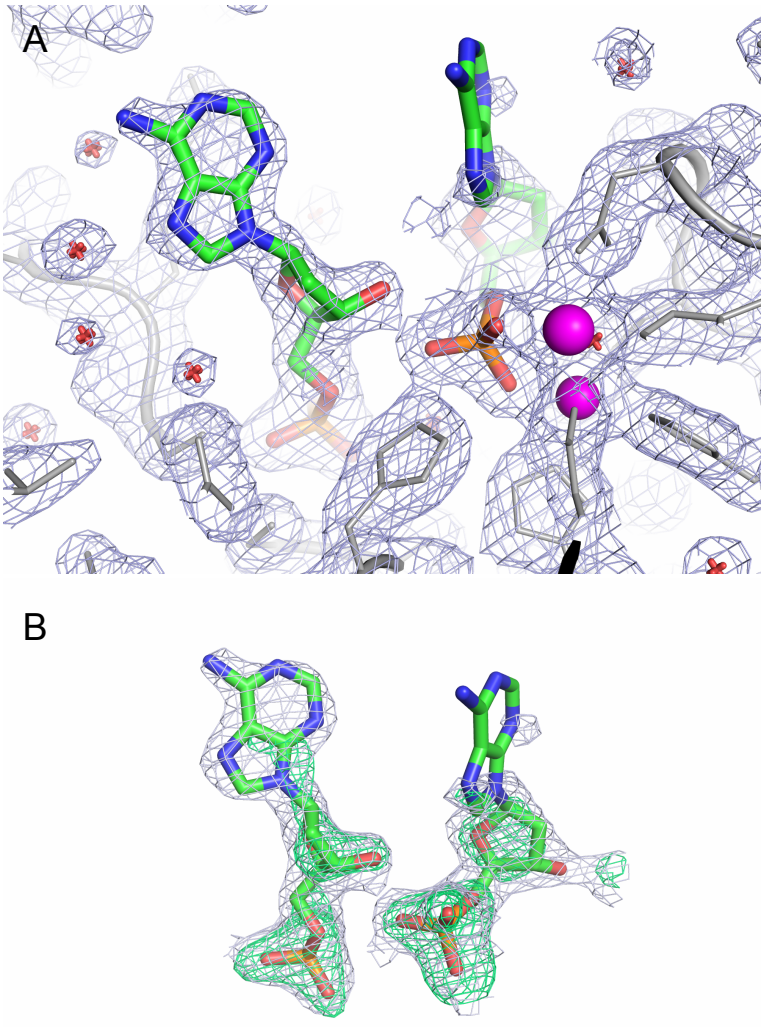


Figure 4.13 Electron density map of the active site of WT SNM1B (nucleotide form).

A: 2Fo-1Fc electron density map of WT SNM1B nucleotide complex contoured at 1 sigma in the vicinity of the active site region, with the nucleotides in green and the active site metals in pink. B: zoomed in view of the density for the nucleotides with the 2Fo-1Fc map shown in blue (contoured at 1.0 sigma) and the omit Fo-Fc map shown in green (contoured at 3.0 sigma). The density for the nucleotides is of good quality, with the exception of the adenine base of the second nucleotide, which is presumably disordered. Figure generated by Dr Joseph Newman, SGC.

4.2.6. The importance of the phosphate binding pocket in the extended active site of WT SNM1B

Both AMP molecules are bound in the 'extended active site' of SNM1B; and this was defined by Allerston *et al.* as consisting of the di-metal catalytic centre, the 5' phosphate pocket, and the DNA binding groove. The 5' phosphate of AMP_1 makes a number of interactions with nearby residues (Figure 4.14, A); hydrogen bonds are made with the main chain amide of D275, the side chain hydroxyl of S274, the side chain of T257, the hydroxyl group of Y182, the side chain of K186, and with both the main chain and side chain of S258. The 5' phosphate of AMP_1 appears to be bound to, and define, the phosphate binding pocket of the extended active site. For efficient DNA binding, SNM1B prefers a substrate possessing a terminal 5' phosphate; and for catalysis, this requirement is absolute. This can be seen in Figure 4.8 where WT SNM1B will not digest a ssDNA substrate possessing a terminal 5' hydroxyl group. This selectivity is likely conferred by these interactions in the phosphate binding pocket, and this interaction is likely necessary to appropriately position the relevant DNA substrate for exonucleolytic activity. The ribose sugar of AMP_1 is well resolved in the electron density, but doesn't appear to form any specific interactions. Interestingly, the adenine base of AMP_1 is stabilised by base stacking with the planar guanidium side chains of two arginine residues – R301 and R259 (Figure 4.14, B). This is likely to enhance substrate binding in a non-sequence specific manner.

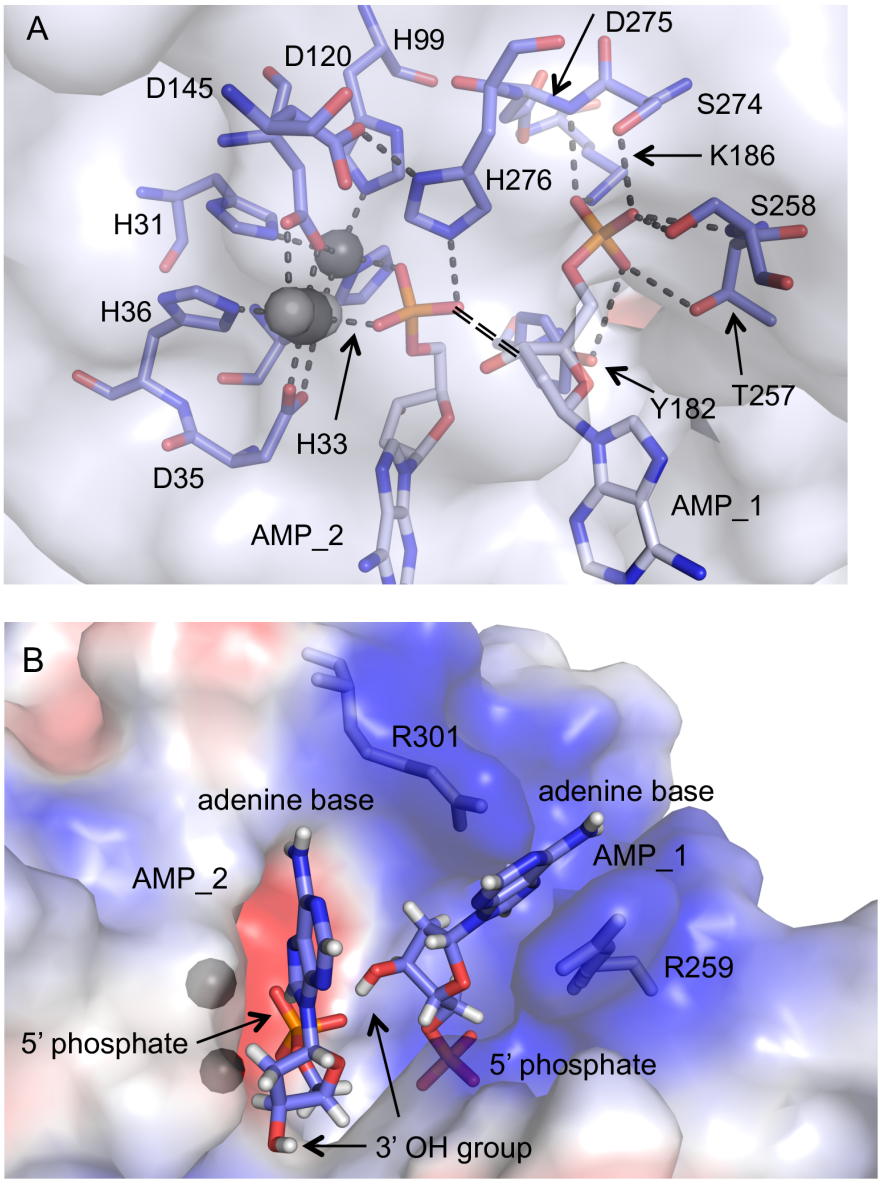


Figure 4.14 Conformation of the active site of WT SNM1B (nucleotide form).

A: The catalytic core of WT SNM1B. The two AMP molecules are as labelled, the two metal ions (iron(II) and nickel) are depicted as dark grey spheres, and the two water molecules, as light grey spheres. The hydrogen bonding network is depicted in dashed form. The double dashed line represents where the phosphodiester bond between the two AMP molecules would probably be, prior to hydrolysis. B: The adenine base of AMP_1 is stabilised by interactions with R301 and R259. The electrostatic surface of the active site of SNM1B is indicated, and blue represents positively charged regions, and red, negatively charged.

Of the residues that constitute the phosphate binding pocket, T257 and S258 are located in the β-CASP domain, and the remainder are in the MBL domain. A similar phosphate binding pocket was observed in the crystal structure of RNase J, and was hypothesised to confer the exonucleolytic preference for an RNA substrate with a 5' monophosphate (Dorleans et al., 2011, Li de la Sierra-Gallay et al., 2008). Interestingly, S274 is highly conserved amongst MBL-β-CASP family members and, in the published structures of RNase J and RNase J1 structures, the equivalent residues are involved in binding the 5' monophosphate of RNA (Newman et al., 2011, Dorleans et al., 2011). Furthermore, when this serine residue was mutated to leucine in RNase J1 (*B. subtilis*) it markedly reduced exonucleolytic activity (Li de la Sierra-Gallay et al., 2008).

Of the three human SNM1 family members, both SNM1A and SNM1B require a terminal 5' phosphate for exonucleolytic activity. Conversely, SNM1C, which exhibits primarily structure-specific endonucleolytic activity, does not possess the same requirement. Interestingly, S274, T257, Y182, and K186 are all conserved in the SNM1A amino acid sequence, whereas only S274 is conserved in the SNM1C sequence. This therefore, may suggest that these analogous amino acid residues also form a phosphate binding pocket in the extended active site of SNM1A, but are not required for SNM1C substrate binding.

4.2.7. Understanding the catalytic core of WT SNM1B

AMP_2 is coordinated in the catalytic centre of the extended active site, and the oxygen atoms of the 5' phosphate group are coordinated to each of the metal ions

and form a hydrogen bond with the imidazole side chain of H276 (Figure 4.14). Neither the ribose sugar, nor the adenine base appear to make specific interactions, and the adenine base is somewhat disordered in the electron density.

The metal ion(s) in the active site of SNM1B are essential for catalysis. The Ni^{2+} in the first metal ion-binding site is coordinated in an octahedral arrangement, by three histidine side chains H31 and H33 (from motif 2 in the MBL domain), H99 (motif 3, from the MBL domain), and the carboxylate oxygen of D120 (motif 4, from the MBL domain), and a water molecule. The coordination is completed by the 5' phosphate oxygen (O3) of AMP_2 (Figure 4.14).

The Fe^{2+} ion modelled in the second metal ion-binding site is also coordinated in an octahedral arrangement (Figure 4.14). D35 and H36A (motif 2, in the MBL domain) each contribute a co-ordination bond, D120 (motif 4, in the MBL domain) makes an additional bond, and another bond is made with the oxygen (O1) of the 5' phosphate group of AMP_2. Two further co-ordination bonds are made with two water molecules.

The binding of AMP_1 and AMP_2 in the active site of SNM1B provides some elucidation of the enzymatic mechanism. As they are two separate molecules, their positioning and coordination network accordingly resembles the product of an exonucleolytic cleavage reaction. The two AMP molecules are in close proximity to one another and the double dashed line in Figure 4.14 indicates the location of the phosphodiester bond that would have been cleaved if they were considered as a two-nucleotide substrate.

The 5’ phosphate group of AMP_2 is in the hydrolysis position (where the phosphodiester bond would be, in the case of a DNA strand). Two oxygen atoms (O1 and O3) from the 5’ phosphate are liganded to the two metal ions. The central water (which, in models of the catalytic mechanism, is assumed to be activated as a hydroxide ion) is also coordinated to both metal ions, and thus, in the reaction process, is likely to nucleophilically attack the phosphodiester bond, resulting in exonucleolytic cleavage of terminal nucleotide.

The third oxygen atom (O2) of the 5’ phosphate group of AMP_2 is hydrogen bonded to H276 (motif B, in the β-CASP domain) (Figure 4.14); and H276 is, in turn, hydrogen bonded to D145 (motif A, β-CASP domain). A role in acid base catalysis is likely, whereby H276 accepts a proton from the catalytic hydroxide ion and, subsequent to hydrolysis, protonates the 3’ hydroxyl group of the ribose sugar of the nucleotide reaction product (AMP_1).

The structure obtained in this study, WT (nucleotide form), and the geometry of the active site, thus supports the accepted reaction mechanism of SNM1B, posited by (Allerston et al., 2015) and (Sengerova et al., 2012). This reaction mechanism is also equivalent to what has been suggested from the structures of other MBL- β-CASP family members, including RNase J (Newman et al., 2011, Dorleans et al., 2011) and CPSF-73 (Mandel et al., 2006).

It is also worthy of note that the overall architecture of WT (nucleotide form) and WT (5AHO) is very similar. The two metal ions are coordinated in the same location, by the same residues, and the two tartrate molecules (present in WT:

5AHO) occupy the same binding sites as the 5' phosphate groups of each of the AMP molecules (present in WT: nucleotide form).

4.2.8. The interaction of SNM1B with a modelled DNA substrate

The structure of WT (nucleotide form), because of its similarity to DNA, could be used as a guide for modelling the interaction of SNM1B with a relevant substrate. Across the front surface of SNM1B, adjacent to the active site are a cluster of positively charged amino acid residues. These form a prominent charged patch, with appropriate dimensions for a DNA binding groove. Based on the proximity to the active site, and the location of the two AMP molecules, it was therefore possible to model a DNA substrate positioned with respect to this positively charged region. Accordingly, B-form duplex DNA with a two nucleotide 3' overhang was modelled, and the phosphate groups of each of the two nucleotides in the active site were utilised as a reference for DNA positioning and orientation (Figure 4.13). This model was generated by Dr Joseph Newman.

As can been seen in Figure 4.15, dsDNA was permitted up to the point of the 5' terminal nucleotide without disrupting base pairing between the two strands of the duplex. At the point of the terminal nucleotide, base pairing is partially disrupted (only one hydrogen bond remains), and the base of the terminal nucleotide is stabilised by the aforementioned interactions with R259 and R301.

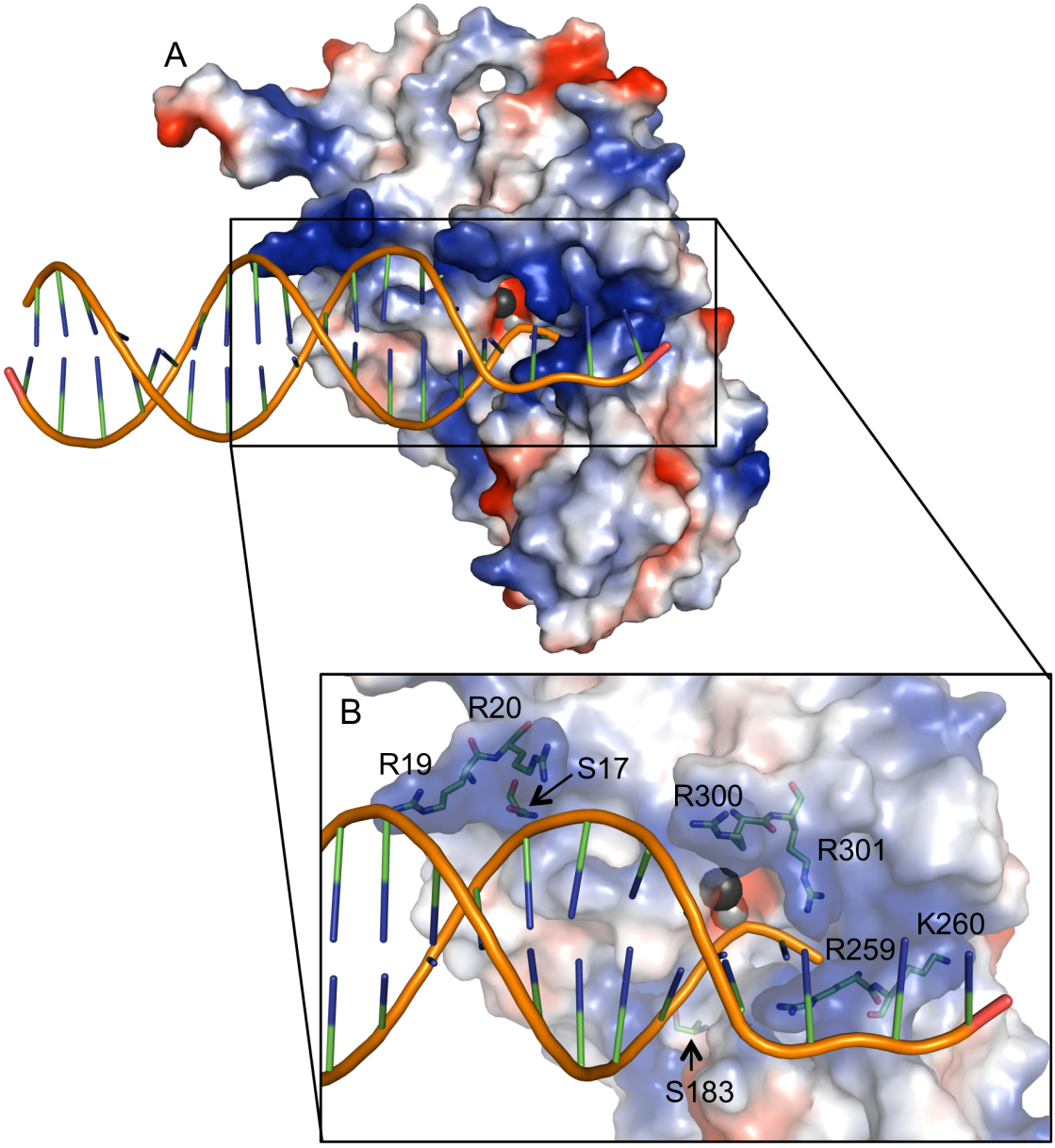


Figure 4.15 Model showing a duplex DNA substrate interacting with the positively charged DNA binding cleft of SNM1B.

A: WT SNM1B with electrostatic surface potential – blue represents positively charged regions, and red, negatively charged. Duplex DNA is modelled in the active site and in the putative DNA binding cleft. B: the specific residues that contribute to the DNA binding cleft are as indicated.

There are a number of residues along this putative DNA binding cleft that contribute to the positively charged surface (Figure 4.15). These include the following: S17, R19, and R20, which are all located between β-strands two and three of the MBL domain; R259 and K260, located in a loop between two parallel β-strands of the β-CASP domain; S183, located in a loop between a β-strand and an α-helix in the β-CASP domain; and R300 and R301, located after β-strand 11 of the MBL domain in a loop region, prior to the α-helix at the C-terminus.

In this model (Figure 4.15), R300 interacts with the non-hydrolysed DNA strand of the duplex, as well as contributing to the depth of the phosphate binding pocket. R301, as previously mentioned, is oriented towards the hydrolysed DNA strand and the guanidium side chain stabilises the 5' terminal nucleotide through base stacking interactions with the adenine base (Figure 4.14). Interestingly, in the structures of WT (form I) and H61Y, the loop that contains R300 and R301 is poorly resolved, and presumably disordered. However, this flexible region is present in the previous structure of WT (5AHO), and the active sites of WT (nucleotide form) and WT (5AHO) are overlaid in Figure 4.16. The side chain of R300 is oriented similarly between the two structures, perhaps being slightly closer to the DNA substrate in WT (nucleotide form). However, the side chain of R301 is oriented differently between the two structures: in WT (nucleotide form), it is proximal (~4 Å) to the adenine base of AMP_1; whereas, in WT (5AHO), it is oriented further away (~8 Å) from the location of the adenine base in the overlaid structure. This is perhaps not surprising given that the structure of WT SNM1B (PDB: 5AHO) contains only the tartrate molecules in the active site, and thus no base interactions are present. Together, these data suggest that, upon DNA

substrate binding, van der Waals interactions between R301 and the relevant base alter the position of the R301 side chain, thus stabilising the DNA substrate for catalysis.

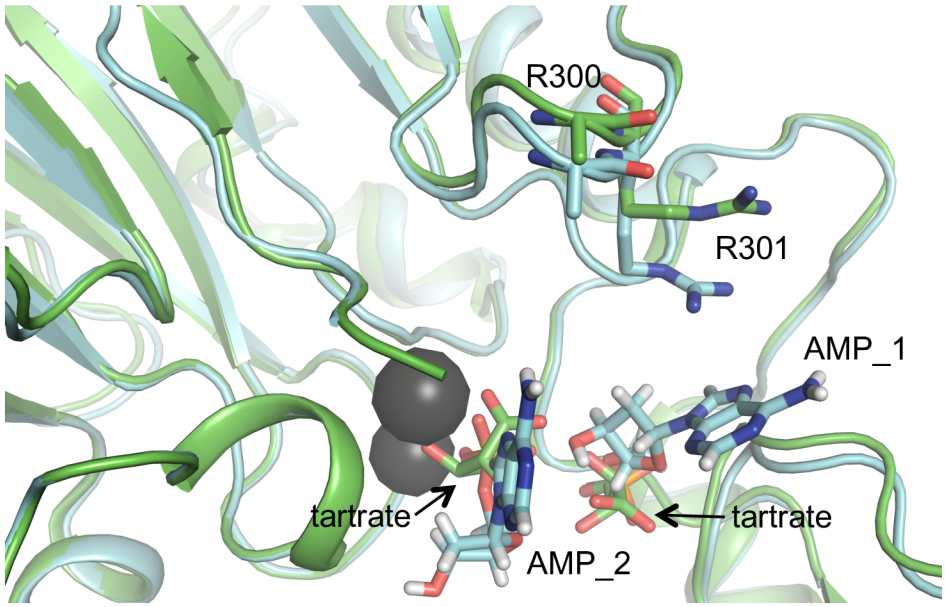


Figure 4.16 Overlay of WT (5AHO) and WT (nucleotide form) showing the location of R300 and R301.

Features of the active site of WT (5AHO), green, and WT (nucleotide form), cyan, shown in cartoon representation. The two AMP molecules of WT (nucleotide form) are indicated, as are the tartrate molecules of WT (5AHO). The two metal ions are indicated in grey. R300 and R301 are as labelled.

The active site architecture of SNM1B likely only permits one strand of duplex DNA to enter the catalytic core (Figure 4.15). As this strand is exonucleolytically processed, a 3' overhang is generated, and this is included in the model. Interestingly, the 'stripe' of positive charge across the face of SNM1B extends slightly beyond the catalytic core; and R259 and K260 provide a positively charged surface that may be important for binding to the phosphodiester backbone of the

newly generated ssDNA overhang. Interestingly, 3' overhang DNA represents the physiologically relevant telomeric substrate for SNM1B (Akhter et al., 2010), and thus, these residues may play an important role *in vivo.*

The presence of a positively charged DNA binding cleft characteristic to both SNM1B and SNM1A was first reported by (Allerston et al., 2015). Mutational analysis of this region included *in vitro* analysis of S183A SNM1B. This mutation resulted in reduced DNA binding and nuclease activity when compared to WT SNM1B (Allerston et al., 2015, Lee, 2015). These data suggest the importance of the positively charged cleft for facilitating DNA binding and subsequent nuclease activity. However, further mutational analysis is required to analyse the importance of the remaining amino acids that contribute to the DNA binding cleft. Suggestive to their possible importance; R19, S183, and R259 are conserved across chicken, fish, mouse, human, chimpanzee, and pig. Furthermore, R20 and K260 exhibit conservation of charge (variation to K or R, respectively) across the aforementioned species.

A second important observation from the model of duplex DNA with SNM1B (Figure 4.15) is that there is no apparent mechanism by which the hydrolysed nucleotide can vacate the active site whilst SNM1B remains bound to the DNA. This suggests that the DNA substrate that is being processed must be released after each exonucleolytic event to allow the 5' terminal nucleotide to vacate the active site. Together with the processivity experiments from the previous chapter, this hypothesis seems likely. However, further analysis is warranted. Interestingly, this is in direct contrast to the observed structures of RNase J (another MBL-β-

CASP family member). These structures of RNase J reveal a highly specific 'exit tunnel' for the hydrolysed nucleotide to be removed from the active site (Dorleans et al., 2011, Li de la Sierra-Gallay et al., 2008, Newman et al., 2011).

4.2.9. Is there a structural basis for RNA v DNA selectivity for SNM1B?

SNM1B possesses rather promiscuous exonucleolytic activity, and is able to degrade most DNA-based substrates (providing they have a terminal 5' phosphate); however, shows very little nuclease activity on RNA-based substrates (Sengerova et al., 2012).

The nucleic acid processing MBL-β-CASP family members exhibit substrate selectivity for either RNA (as is the case for RNase J, and CPSF) or DNA (as is the case for SNM1A/B/C). The preference of nucleic acid processing MBL-β-CASP family members for either RNA or DNA was initially hypothesised to be the result of the presence of either valine (for DNA nucleases) or histidine (for RNA nucleases) at motif C, in the β-CASP domain (Callebaut et al., 2002). However, mutation of the relevant valine residue to histidine (V1061H) in SNM1A had no impact on its substrate preference, or catalytic efficiency (Sengerova et al., 2012).

As can be seen in Figure 4.17, V298 (motif C, β-CASP domain) is not directly involved in binding either of the metal ions, or either of the AMP molecules. This is in contrast to the structure of human CPSF-73 where the motif C histidine (H418) is involved in co-ordinating the second zinc ion in the active site (Mandel et al., 2006). Consequently, the DNA processing β-CASP-MBLs (SNM1A/B/C) have one

fewer metal ion coordinating residue. It was hypothesised by Allerston *et al.* that this resulted in the second metal ion being bound less tightly for these enzymes and a more open active site conformation. It is not clear, however, whether V298 confers some necessary enzymatic property, or whether it simply defines an evolutionary marker that delineates the RNA and DNA processing β-CASP-MBLs.

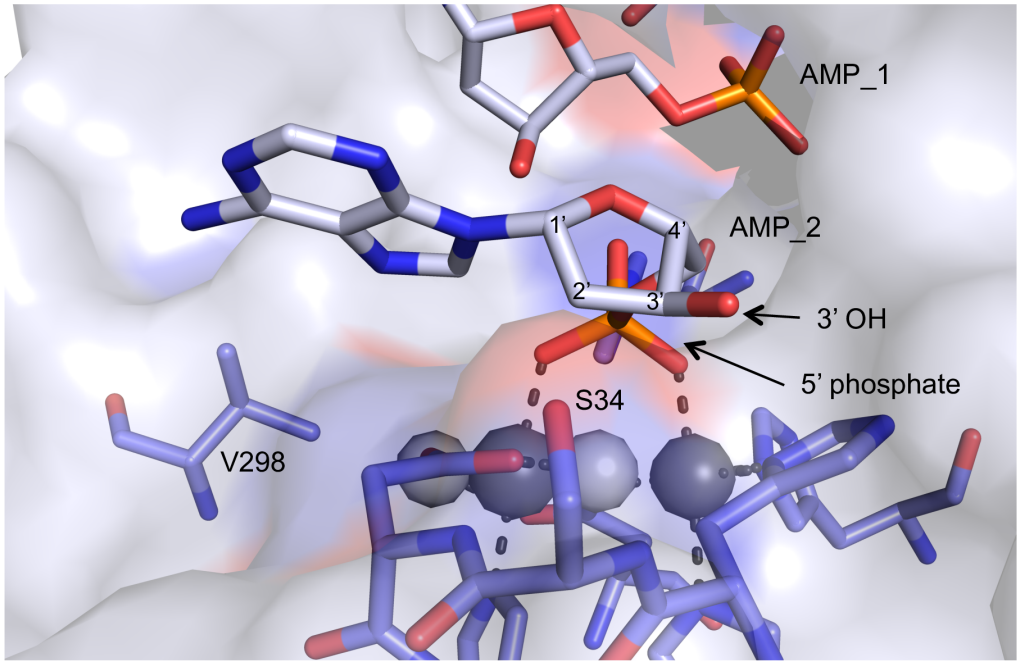


Figure 4.17 The role of V298 and S34 in binding an RNA substrate in the active site.

The active site of SNM1B (crystal form 3) showing the two AMP molecules, the two metal ions (dark grey spheres), and the two coordinated waters (light grey spheres). The residues that coordinate the metal ions, and the 5' phosphate group of AMP_2 are present and the coordination network is represented by dashed lines. Valine 298 and serine 34 are as labelled. The distance between the 2' C of the ribose ring of 2'-AMP_2 and the hydroxyl side chain of S34 is 3.2 Å.

The preference exhibited by SNM1B for DNA over RNA may be conferred by properties of the active site. The hydroxyl side chain of S34 is located adjacent (~3 Å) to the second carbon of the ribose ring of AMP_1 (Figure 4.17). This, therefore,

may inhibit effective catalysis of an RNA substrate as the addition of an oxygen at the 2' position on the ribose ring would cause a steric clash; although it is possible that this may be accommodated by a small movement of the S34 side chain. Interestingly, S34 is conserved in SNM1A and yeast Pso2p, but not with the RNA-processing β-CASP-MBLs.

Furthermore, RNA and DNA differ structurally, and this may affect substrate interactions with the positively charged binding cleft. Putative interactions between the DNA substrate and the extended active site in the model, for example, those mediated by the positively charged residues S17, R19 and R20 may be dependent on the pitch of the helix, and these might not be well replicated for an RNA substrate. RNA is more often single stranded than DNA, however (due to the addition of extra OH group of the ribose sugar), ssRNA is more structurally constrained than ssDNA. Furthermore, dsRNA is more often found in the A-form, whereas DNA is most often found in B-form. The helix of A-form RNA has a larger diameter (23 Å) than B-form dsDNA (18 Å), and, as a result, there are more base pairs per helix turn. Accordingly, these structural differences between RNA and DNA might alter the interaction with the substrate binding cleft of SNM1B, and thus alter observed exonucleolytic efficiency. In support of this, WT SNM1B shows very little nuclease activity on a dsRNA substrate (Sengerova et al., 2012).

4.2.10. The active site of SNM1B accommodates a damaged substrate

As previously discussed (4.2.8), the extended active site of SNM1B is relatively open and permissive to substrate binding. This is unusual, both amongst β-CASP-

MBL family members, and amongst nuclease structures considered more globally. As SNM1B has been reported to digest past DNA damage lesions *in vitro* including 8-oxo-guanine (8-oxo-G) containing substrates (Figure 3.31) and a site-specific, SJG-136 interstrand crosslink (ICL) (Sengerova et al., 2012), it was attempted to model each of these interacting with SNM1B.

In Figure 4.18, two 8-oxo-G bases were modelled in the position of the two terminal 5' nucleotides in the DNA strand. In each of the nucleotide positions the slightly bulkier 8-oxo-G bases were well tolerated. This likely contributes to the observation that 8-oxo-G containing substrates are easily processed by SNM1B.

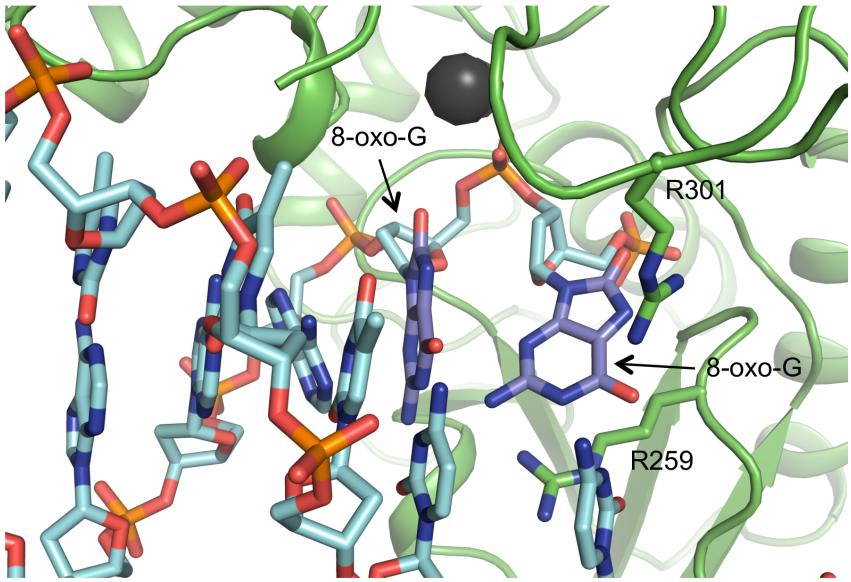


Figure 4.18 Model of a duplex DNA substrate containing two 8-oxo-guanines in the active site of SNM1B.

The structure of WT (nucleotide form) is in green, and represented in cartoon form. The two metal ions (iron(II) and nickel) are depicted as grey spheres. The duplex DNA modelled bound to SNM1B is depicted in stick form in cyan. The two 8-oxo-guanine bases are depicted in purple. The base stacking residues, R301 and R259 are as labelled.

The SJG-ICL was modelled similarly to previously published (Hopton and Thompson, 2011). However, it was not possible to fit duplex DNA containing the ICL to the DNA binding cleft, with the 5' end being accessible to the active site. SJG-136 interacts with the exocyclic NH_2 groups of guanine bases to form either intrastrand or interstrand crosslinks across the minor groove, with little perturbation to the DNA helix (Hopton and Thompson, 2011, Rahman et al., 2009). However, in the model of SNM1B with duplex DNA (Figure 4.15), the minor groove opens a little towards the 5' end to accommodate the slight strand separation necessary for the substrate strand to enter the active site. The addition of SJG-136 crosslink at

this point restricts any potential opening of the DNA duplex, and thus prevents the 5' DNA strand from binding to the active site in the manner predicted by the model. However, it has previously been documented that SNM1B can digest past an SJG-136 ICL, albeit with reduced efficiency when compared to an undamaged DNA substrate (Sengerova et al., 2012). Therefore, it may be that when the movement of the DNA helix is constrained, the substrate interacts with SNM1B in a different (or non-optimal) manner, and, subsequently, catalytic efficiency is reduced, although not entirely abrogated.

The relative accessibility of the active site of SNM1B is in direct contrast to RNA processing β-CASP-MBLs, for example, RNase J and CPSF. The structures of both RNase J1 (*B. subtilis*) and RNase J (*T. thermophilus*) show a highly charged and relatively restricted RNA binding 'tunnel' that would not accommodate a double stranded substrate (Newman et al., 2011, Dorleans et al., 2011). The structures of CPSF-73 and CPSF-100 have only been solved in the apo form (Mandel et al., 2006), however, they also appear to possess a more closed DNA binding site.

There are a number of other DNA repair nucleases for which the structures of the enzyme in complex with DNA have been solved, and each of these reveal a structural basis for substrate preference. For example, the 5' to 3' exonuclease, and flap endonuclease, EXO1 has a relatively restricted active site. Briefly, with dsDNA substrates, base pairing is disrupted, and duplex fraying (of 2–3 nucleotides) at the 5' terminus is required to generate the ssDNA substrate for exonucleolytic activity. Once this ssDNA region is in the active site the final base

‘flipped out’ and thus positioned for catalysis (Orans et al., 2011). Furthermore, the specificity for nicked substrates is specified by bending the DNA in such a way that is only tolerated at the junction of single stranded and double stranded DNA (Orans et al., 2011). As a result of the ‘crowded’ active site, there is arguably little space for a damage lesion at the position of the flipped terminal base. Additionally, the requisite bending of the DNA substrate may not be permitted when lesions are present that constrain backbone flexibility (e.g. intrastrand, or interstrand crosslinks). Accordingly, EXO1 cannot digest past a nitrogen mustard ICL *in vitro* (Pizzolato et al., 2015). On the other hand, the structure of the 5’ to 3’ exonuclease/5’ flap endonuclease, FAN1, in complex with DNA reveals a more accessible active site, and FAN1 can digest past ICLs *in vitro* (Wang et al., 2014, Pizzolato et al., 2015). However, binding of the DNA substrate to the FAN1 nuclease results in the DNA backbone becoming bent at an angle of 76^{o}. This, therefore, restricts binding to ssDNA, nicked or flap duplex DNA substrates (Wang et al., 2014).

It would appear that both FAN1 and EXO1 modify the DNA substrate to a greater extent than SNM1B to allow enzyme catalysis. This probably contributes to the greater substrate diversity of SNM1B. However, it is important to note that the structures of SNM1B have all been obtained from truncated proteins containing only the MBL and β-CASP domains. Thus, additional domains in the full-length protein may result in different structural conformations, and may affect the accessibility of the active site, and subsequent substrate binding/nuclease activity.

4.2.11. The requirements for metal ion cofactor(s) for catalysis of SNM1B

It is clear that the enzymatic activity of WT SNM1B is dependent on its ability to coordinate the relevant metal ion cofactor(s). This is evidenced by the observation that the addition of metal ion chelators entirely abrogates DNA binding and inhibits nuclease activity (Figure 4.19 and (Sengerova et al., 2012)). However, it is not immediately clear which metal ion(s) are used by SNM1B for catalysis *in vivo*. Furthermore, the three crystal structures of WT SNM1B obtained to date: two from this study (WT, form I and nucleotide form); and one from previous work (Allerston et al., 2015), each show a different metal ion coordination profile. Accordingly the di-metal binding core and the coordination network for each is displayed in Figure 4.20.

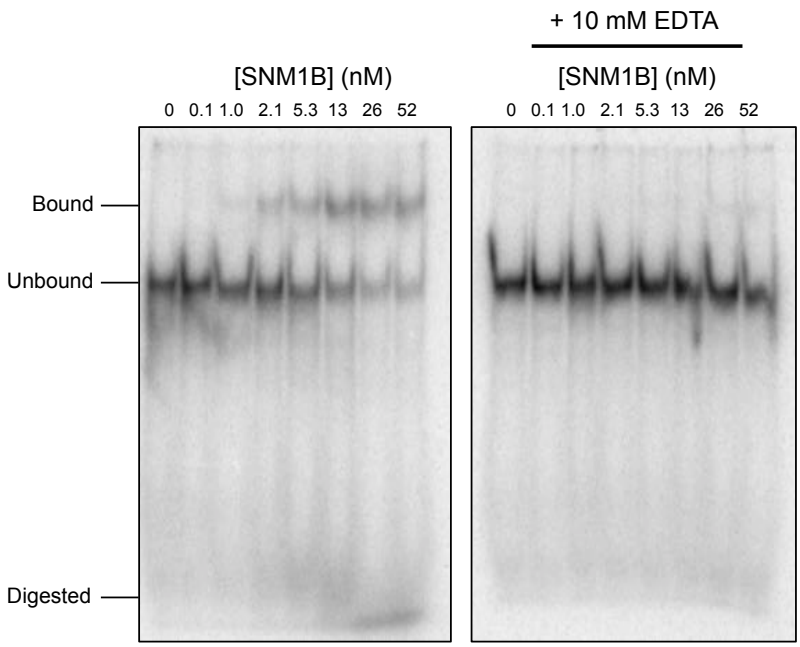


Figure 4.19 The addition of the divalent metal ion chelator, EDTA, abrogates SNM1B DNA binding.

EMSA showing binding of WT SNM1B to ssDNA substrate in either the presence or absence of 10 mM EDTA (as labelled). 1 nM of single stranded, 50 nucleotide DNA substrate was incubated for 5 minutes at 37^{o}C with WT SNM1B (concentrations as labelled), before being analysed by 10% native polyacrylamide gel electrophoresis. Unbound DNA substrate, DNA substrate bound to SNM1B, and digestion products are as indicated. Each gel is representative of at least three individual experiments.

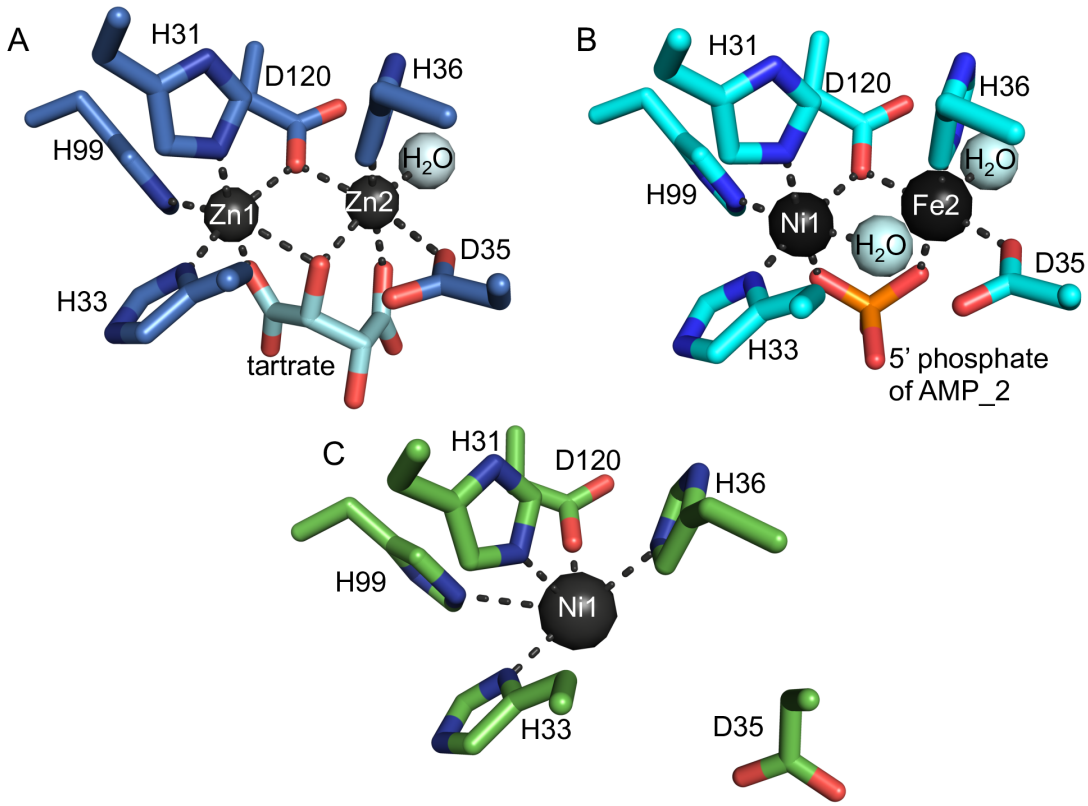


Figure 4.20 The metal binding regions for three WT SNM1B structures.

The metal ion coordination network for: A – WT (5AHO); B – WT (nucleotide form); C – WT (form I). The coordinating amino acids are as labelled, as are the active site tartrate (for WT 5AHO) and the 5' phosphate of AMP_2 (for WT nucleotide form). The coordination network is indicated by dashed lines. The active site metal ions are as labelled.

The metal ion(s) present in each of the WT (form I) and WT (nucleotide form) structures was modelled based on a number of factors. For WT (form I), only the first metal binding site was occupied, and XRF revealed iron, nickel, and zinc present in the crystal, with nickel being the most abundant (Figure 4.6). The profile and geometry of the co-ordinating atoms consists of four nitrogens from histidine residues, as well as one oxygen from the carboxyl side chain of D120, arranged in

an approximately octahedral conformation (the final bond being contributed by an oxygen atom from a water molecule). This bonding network and preference for oxygen and nitrogen is characteristic for both Zn^{2+} and Ni^{2+} (Zheng et al., 2008). The approximate co-ordination distances (between 2.4 and 3 Å) are concomitant with each Ni^{2+}, Zn^{2+}, Fe^{2+} (Zheng et al., 2008). Although the structure is modelled with Ni^{2+}, it is likely that all three (Ni^{2+}, Zn^{2+}, Fe^{2+}) are present in the first metal binding site throughout the crystal.

For WT (nucleotide form) the electron density map revealed that both metal ion-binding sites were occupied, and XRF analysis revealed that a mixture of nickel and iron were present in the crystal in roughly equal proportions (Figure 4.12). Based on the coordination profile and the electron density, Ni^{2+} was modelled in the first metal ion-binding site. Considering the electron density, the XRF analysis, and the coordination profile, Fe^{2+} was modelled in the second metal ion-binding site. Fe^{2+} is coordinated (in an octahedral arrangement) by two oxygen atoms from the carboxyl side chains of D35 and D120; a nitrogen atom from the imidazole side chain of H36; one phosphate oxygen from AMP_2; and an oxygen atom from a water molecule (Figure 4.16 and Figure 4.10). Both Ni^{2+} and Fe^{2+} exhibit a strong preference for interacting with histidine residues; however, Fe^{2+} displays a stronger preference for oxygen ligands than Ni^{2+} (Zheng et al., 2008). The approximate bond length (~2.2 Å for interactions with both histidine, and aspartate residues) was again consistent with Fe^{2+} (Zheng et al., 2008).

However, how the above WT SNM1B structures represent the catalytically, or physiologically, relevant form of SNM1B is not clear. Firstly, the small proportion of

WT SNM1B that crystallised in the given conditions (4.2.1 and 4.2.3) is not necessarily representative of the WT SNM1B population in solution. Secondly, WT and H61Y SNM1B were both purified using an initial Ni-Sepharose chromatography column, which may have introduced Ni^{2+} to the active site. Furthermore, all crystallised proteins were purified in the truncated form from an over-expressed heterologous system, which does not accurately recapitulate physiological conditions.

These results, however, raise three interesting questions: is WT SNM1B catalytically active with just one metal ion-binding site occupied; what metal ions best support SNM1B catalysis; and which metal ions are physiologically relevant?

SNM1B (along with the other β-CASP-MBL nucleases) has previously been considered as a two metal-ion-dependent nuclease (Yang, 2011). Indeed, the structures of other β-CASP-MBL family members, including RNase J1 (Newman et al., 2011), RNase J (Dorleans et al., 2011, Li de la Sierra-Gallay et al., 2008), CPSF-73, and CPSF-100 (Mandel et al., 2006), have been solved with two Zn^{2+} ions coordinated in the active site. However, the structures of other family members, including SNM1A (Allerston et al., 2015) have been solved with just one Zn^{2+} ion present in the active site. Within the broader MBL superfamily, members have also been shown to coordinate a single metal ion, including the class of bacterial B2 MBLs (Palzkill, 2013). However, it is not clear if the single metal ion form of SNM1B is compatible with catalysis, and elucidating this will require further research.

Previous research has been performed to reveal which divalent metal ions support SNM1B catalysis. SNM1B has been shown to have robust nuclease activity in the presence of 0.01–10 mM of $MgCl_2$, $MnCl_2$, and $CaCl_2$ (Sengerova et al., 2012, Lenain et al., 2006). $ZnCl_2$ was stimulatory at low concentrations (0.01 mM), and strongly inhibitory at higher concentrations (0.1–10 mM) (Sengerova et al., 2012). The nuclease activity of SNM1B was also inhibited by the addition of metal ion chelators, with a particular sensitivity towards o-phenanthroline, which is relatively selective for Zn^{2+} ions (Slater et al., 1971, Williams and Loeb, 1973). This led the authors to suggest that having one metal ion-binding site occupied by Zn^{2+} was stimulatory to enzymatic activity; however, having both metal ion-binding sites occupied by Zn^{2+} may be inhibitory (Sengerova et al., 2012). However, these results were confounded because the co-purified, endogenous metal ion was not removed prior to nuclease assay analysis.

To further clarify this, nuclease assays were performed whereby 100 nM WT SNM1B was pre-incubated with 10 mM EDTA for 30 minutes, after which the excess EDTA was removed. This was an attempt to remove the any co-purified endogenous metal ion(s). Subsequent to incubation with EDTA, a nuclease assay was performed with increasing concentrations of $FeCl_2$, $NiCl_2$, $MnCl_2$, $MgCl_2$, $ZnCl_2$, and $CaCl_2$, both individually and in combination. Unfortunately, there was some aggregation at higher concentrations of $FeCl_2$ and $NiCl_2$ and the DNA failed to enter the gel and became trapped in the loading wells (Figure 4.21).

However, it is worthy of note that nuclease activity was only supported in the presence of 1.0–10 mM $MnCl_2$ (Figure 4.21). Perhaps surprisingly, when both

$MnCl_2$ and $ZnCl_2$, or $MgCl_2$ and $ZnCl_2$ were present, nuclease activity was still inhibited. However, concentrations of $ZnCl_2$ at 0.05 mM (as in the combined $MnCl_2$/$MgCl_2$ + $ZnCl_2$) may still be inhibitory to SNM1B activity.

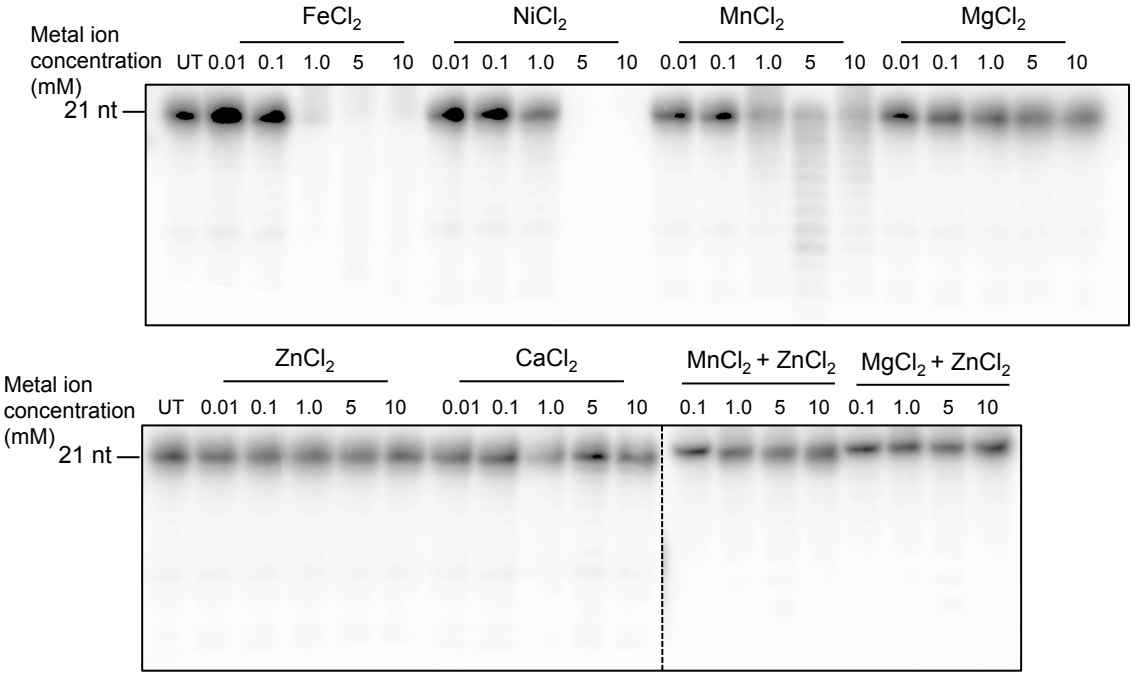


Figure 4.21 The nuclease activity of WT SNM1B is supported by $MnCl_2$ after pre-treatment of SNM1B with EDTA.

100 nM SNM1B was incubated with 10 mM EDTA for 30 minutes, before the excess EDTA was removed by spinning through a P6 gel chromatography column. 50 nM of WT SNM1B was then added to 100 mM 21 nucleotide ssDNA substrate in the presence of concentrations of $MgCl_2$, $MnCl_2$, $NiCl_2$, $FeCl_2$, $ZnCl_2$, or $CaCl_2$, either alone or in combination. For $MnCl_2$/$MgCl_2$ + $ZnCl_2$, the indicated concentration (mM) refers to the final total concentration of divalent metal ions in the solution. The reaction was incubated at 37°C for 1 hour, and samples were then analysed by 20% polyacrylamide gel electrophoresis. UT indicates substrate DNA with no enzyme added. Gel is representative of three individual experiments.

These results are in accordance with previous research (Sengerova et al., 2012) insofar as they validate the presence of Mn^{2+} as the most permissive metal ion for catalysis. Perhaps surprisingly, the lack of nuclease activity in the presence of other metal ions suggests that they are unlikely to support catalysis. However, it is also possible that chelation of the endogenous metal ion(s) with EDTA led to the SNM1B becoming irreversibly denatured, or altering the structure in some way that rendered it incompatible with enzymatic activity. Accordingly, further analyses are required to establish which metal ion cofactors are necessary for SNM1B nuclease activity, and which of those are physiologically relevant.

4.3. Conclusion

This study describes the crystal structure of the apo form of WT and H61Y SNM1B, as well as the structure of WT SNM1B bound to two AMP molecules. A comparison of the structures of WT (form I) and H61Y reveals negligible difference between the two. The overall architecture is extremely similar, as is the conformation of the active site (Figure 4.3). The only discrepancy is the orientation of the respective Y or H61 residue. Taken together with the results of Chapter 3, it would seem that the H61Y breast cancer associated variant of SNM1B has the equivalent structure to WT SNM1B and thus there is no discernable difference in DNA binding or nuclease activity.

The structure of WT (nucleotide form) provides useful insight into some of the key features of the extended active site of SNM1B. It allowed it to be determined which residues define the 5’ phosphate binding pocket and which are likely to contribute

to catalysis. On the basis of the nucleotide structure, a model of duplex DNA bound to SNM1B was generated (Figure 4.15). This model suggested the importance of the positively charged putative DNA binding cleft and also showed how the conformation of the active site is permissive to certain DNA damage lesions (including 8-oxo-G). Future experiments involving site directed mutagenesis of amino acids that are suggested to constitute each of these features may provide further elucidation as to their role.

Of the two crystal structures of WT SNM1B obtained in this study (WT (form I) and WT (nucleotide form)), each showed a different metal ion co-ordination pattern. Although both of these were solved with Ni^{2+} in the first metal ion-binding site, it is unlikely that this is physiologically relevant, but rather an artefact of the purification process. It may be more likely that, in keeping with the majority of the MBL family, and considering the atomic nature and orientation of the coordinating, and highly conserved, histidine residues, this site is occupied by a Zn^{2+} ion (Zheng et al., 2008, Baldwin et al., 1979, Francisco J Fernandez, 2011).

The metal ion co-factor that occupies the second binding site is not clear; and the crystal structures of WT (form I) and H61Y raise the possibility that SNM1B is able to function with only metal ion-binding site occupied. Interestingly, the B2 class of bacterial MBLs possess optimal hydrolytic activity when one zinc ion is bound, but are inhibited when both sites are occupied. However, in this family of enzymes, the single Zn^{2+} is occupied in the second metal ion-binding site (Palzkill, 2013). Furthermore, the substrate specificity of the B2 class of MBLs is also highly restricted and, in general, single metal ion nucleases tend to possess stricter

substrate specificity (Palzkill, 2013). Thus, it seems more likely that SNM1B (and SNM1A) have one high affinity metal binding site, and one lower affinity metal binding site. As the metal ion in the second binding site is more loosely coordinated it thus might not always be present. Accordingly, this has been observed in the crystal structures for both SNM1B (Figure 4.4) and SNM1A (Allerston et al., 2015)

The structure of WT (5AHO) was modelled with two Zn^{2+} ions in the active site, however, this was not necessarily supported by experimental evidence – for example, XRF scans were not reported to be performed on the crystal. As such, it is likely that both metal ions were modelled as Zn^{2+} based on similarity with the structures of other MBL family members. However, it is important to note that the di-zinc RNA processing β-CASP-MBLs have an additional histidine coordination residue at the second metal ion site, which gives an additional nitrogen ligand at this position. For SNM1B, the atomic nature of the coordinating residues at this site is more predominated by oxygen ligands (one histidine nitrogen, two carboxylate oxygens, one phosphate oxygen, and two water oxygens). This may be compatible with Zn^{2+}; however, both Fe^{2+} and Mn^{2+} are more receptive to coordination by oxygen ligands (Zheng et al., 2008).

Fe^{2+} occupied the second metal ion-binding site in the structure of WT (nucleotide form) and this may be physiologically or catalytically relevant. Within the bacterial MBLs, glyoxalase II contains a mixed Fe^{2+} and Zn^{2+} binuclear catalytic centre, and the type A flavoproteins contain two Fe^{2+} ions in the active site (Gomes et al.,

2002). Thus, coordination of Fe^{2+} would not be unprecedented within the context of the MBL superfamily.

As the addition of $MnCl_2$ stimulates the nuclease activity of SNM1B (Sengerova et al., 2012) and, is most able to rescue the activity of SNM1B after metal ion chelation (Figure 4.20), it may be that Mn^{2+} is coordinated in the second metal ion-binding site. Again, this would not be entirely unusual within the MBL family, as the bacterial protein UlaG was shown to preferentially bind a single Mn^{2+} ion in the active site (Garces et al., 2010). Additionally, Mn^{2+} is requisite for the catalysis of other nucleases, including the dual exonuclease/endonuclease, MRE11. Unlike the SNM1 family members, MRE11 belongs to the protein phosphatase 2B (PP2B) family. However, there are many analogous features between the β-CASP family members and MRE11, including a similar structural topology and similar active site catalytic residues (Francisco J Fernandez, 2011). Furthermore, Mn^{2+} does possess the Lewis acid character that (although not as strong as Zn^{2+}) makes it a possible candidate for catalysis (Bock et al., 1999).

However, defining the precise metal ion(s) requirements for the enzymatic activity of SNM1B is technically challenging Furthermore, the requirements *in vitro* may differ from the exact metal ion(s) that occupy the protein *in vivo*, as this is dependent on many additional factors. Ultimately, further experimentation is required to elucidate this.

5. Analysis of the cellular role of WT SNM1B

5.1. Introduction

As discussed in Chapter 1, SNM1B is critically important for normal cellular functioning, and a number of studies have elucidated some of the cellular roles of SNM1B by utilising either knockout or depleted cell lines.

To date, reported cell lines deficient in SNM1B include those generated from chicken DT40 (Ishiai et al., 2004, Nojima et al., 2005), and mouse embryonic fibroblasts (MEFs; cultured from SNM1B homozygous null mice) (Wu et al., 2012, Akhter et al., 2010, Lam et al., 2010). There have also been a number of studies that have exploited RNA interference techniques (either siRNA or shRNA) to examine the effects of SNM1B depletion in human cell lines. These studies have utilised the following cell lines: HeLa (Demuth et al., 2004), HEK293 (Liu et al., 2009, Bae et al., 2008, Mason and Sekiguchi, 2011, Mason et al., 2013), U2OS, and GM00637 (Demuth et al., 2008).

Taken together, experimental results from the aforementioned cellular studies show broadly similar phenotypic characteristics, thus suggesting no major cell line or species dependent differences. SNM1B null, or depleted, cells exhibit hypersensitivity to DNA damaging agents (particularly those that induce ICLs) (Demuth et al., 2004, Ishiai et al., 2004, Nojima et al., 2005, Bae et al., 2008), significant telomeric defects (as evidenced by end-to-fusions and telomere-dysfunction induced foci, (Akhter et al., 2010, Lenain et al., 2006, van Overbeek

and de Lange, 2006)), and enhanced susceptibility to replication stressors (Mason et al., 2013).

Given that RNA interference based techniques can be transient (depending on the particular technique utilised), and that depletion of protein expression is often incomplete (Medema, 2004), one of the aims of the present study was to generate an SNM1B disrupted human cell line. This would therefore provide an extremely useful tool for further elucidating the cellular roles of SNM1B in human DNA damage and telomeric maintenance pathways. As such, the generation of an SNM1B disrupted cell line was attempted utilising the CRISPR-Cas9 genome editing technique first described by (Jinek et al., 2012) and (Cong et al., 2013).

It is also important to note that there is the possibility that SNM1B is an essential gene, at least in some cell lines. To date, there have been no reports of an SNM1B knockout in a human cell line, there have been no observed cases of individuals lacking the SNM1B protein, and loss of SNM1B results in perinatal lethality in mice (Akhter et al., 2010). Nonetheless, whilst loss of SNM1B renders primary MEFs unable to divide and grow normally (Akhter et al., 2010), this effect was not observed with SV40 immortalised MEFs (Lam et al., 2010, Wu et al., 2012). Furthermore, SNM1B was not identified as an essential gene in either HAP1 or KBM7 cells (Blomen et al., 2015, Wang et al., 2015). Along with the RNAi results from the literature, the data therefore suggest that complete loss of SNM1B should be viable in, at least some, immortalised cell lines.

5.2. Results

5.2.1. Generation of an SNM1B disrupted cell line using CRISPR-Cas9

SNM1B disrupted cell lines were generated using the CRISPR-Cas9 genome editing technique (Jinek et al., 2012, Cong et al., 2013). This is a relatively recently developed method and is derived from the type II CRISPR (clustered regularly interspaced short palindromic repeats) bacterial adaptive immune system. The Cas9 protein (CRISPR-associated protein 9) is an RNA guided, DNA endonuclease. The specificity of the nuclease activity is directed by a guide sequence, within an RNA-duplex, that interacts, via Watson-Crick base pairing, with a target DNA sequence. This binding of the guide RNA sequence to the DNA enables the Cas9 nuclease to cut, adjacent to the target DNA sequence, thus inducing a double strand break (DSB) (Jinek et al., 2012, Cong et al., 2013).

In unmodified type II CRISPR bacterial systems the combination of the crRNA (CRISPR RNA) and the second tracrRNA (transactivating CRISPR RNA) are required to generate the RNA duplex necessary for endonucleolytic cleavage. However, for the purposes of genome engineering the crRNA and the tracrRNA are synthetically engineered as a single guide RNA (sgRNA) (Jinek et al., 2012, Cong et al., 2013). This sgRNA consists of two necessary features: a specific sequence at the 5' end that interacts, via Watson-Crick base pairing, with the target DNA sequence; and at the 3' end, a duplex RNA structure that binds to the Cas9 nuclease, therefore localising it to the DNA. The Cas9 nuclease then endonucleolytically cleaves adjacent to the protospacer adjacent motif (PAM) in the target DNA sequence (Cong et al., 2013). The canonical PAM consists of the

sequence 5’ -NGG- 3’, and therefore, a guide sequence can be directed anywhere in the genome, as long as it is located adjacent to a PAM (Doudna and Charpentier, 2014). This process is illustrated schematically in Figure 5.1.

Endonucleolytic cleavage by the Cas9 nuclease results in the formation of a DSB. Subsequently, there are two main pathways for repair – homologous recombination (HR), or non-homologous end joining (NHEJ). This is discussed in detail in Chapter 1, but pathway choice is primarily dependent on the cell cycle stage (Sonoda et al., 2006). If a sister chromatid is not available (i.e. the cell is not in either S- or G_2-phase), then the lesion is likely to be repaired by NHEJ. NHEJ is therefore the major DSB repair pathway in mammalian cells (Sonoda et al., 2006). This process is potentially error-prone and can result in insertion or deletion mutations (indels) (Figure 5.1) (Heidenreich et al., 2003). These indels can lead to a frameshift mutation and the introduction of premature stop codon, resulting in a truncated mRNA transcript, and nonsense mediated decay of that transcript, thus resulting in the loss of expression of the final protein product (Popp and Maquat, 2016). This, therefore, is the premise of generating a knockout cell line using the CRISPR-Cas9 system (Doudna and Charpentier, 2014).

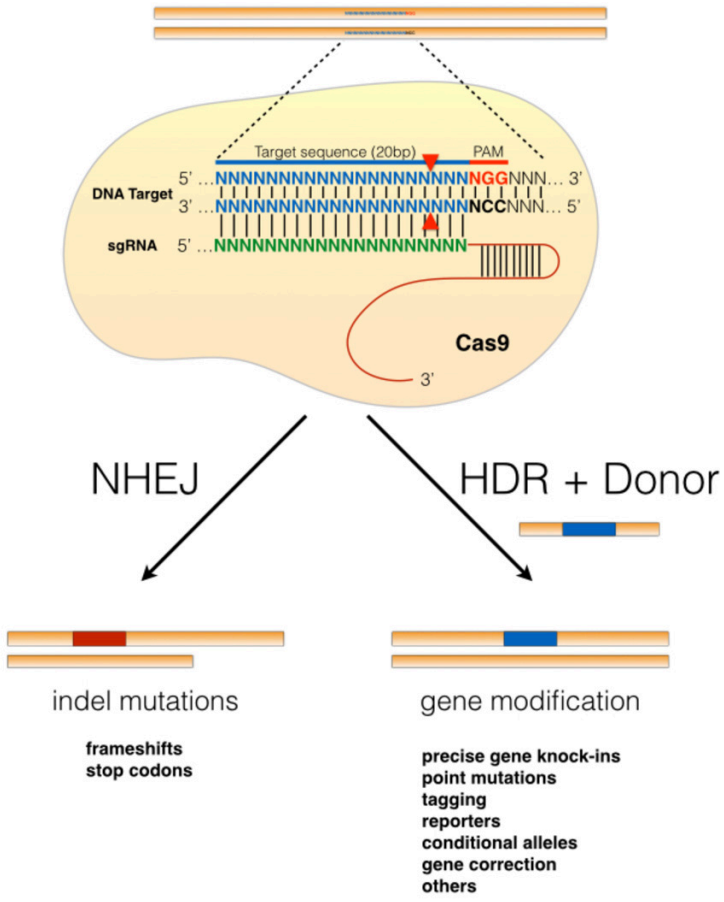


Figure 5.1 Schematic of CRISPR-Cas9 genome editing.

The sgRNA sequence from the transfected CRISPR plasmid undergoes Watson-Crick base pairing with the target DNA sequence, and this directs the Cas9 nuclease to the desired site. The Cas9 protein then endonucleolytically cleaves the DNA target sequence, adjacent to the PAM. This generates a DSB that is either repaired through NHEJ or homology directed repair (HDR). However, this study relies only on NHEJ. The NHEJ process is potentially error-prone which can result in indel mutations thus leading to frameshift mutations, the introduction of a premature stop codon, and an invalid protein product. This image is taken from (Sanchez-Rivera and Jacks, 2015) and is reused with permission from Professor Tyler Jacks, MIT.

The intention in this study was to remove the majority of exon 1 of the SNM1B gene, thereby resulting in a truncation, and the deletion of 57 amino acids from the catalytic metallo-β-lactamase (MBL) domain. As such, specific single-guide RNA (sgRNA) sequences were designed to bind at both the beginning of exon 1 and the beginning of intron 1, with the intention of deleting the intervening genetic information (Figure 5.2). The guide sequences were generated with the CRISPR design tool from the Zhang lab (MIT), and sequences can be found in Appendix III. Each guide sequence was cloned into either an eGFP-expressing pX330 plasmid, or a Ruby-expressing pX458 plasmid (Figure 5.2). A successful cut by the Cas9 nuclease at both guide sites would result in the loss of ~170 or ~250 bp, depending on the combination of guide sequences used (Figure 5.2).

As the focus of this study was on the role of SNM1B in DNA damage repair pathways, rather than its telomeric role, disruption of SNM1B was attempted on three different cell lines, each with a different telomeric phenotype. This was to try and mitigate the possible telomere dysfunction effects of SNM1B depletion, thus allowing the study to focus on the role of SNM1B in DNA damage repair. As such, HeLa1.2.11 cells (a subclone of HeLa) were utilised because they possess extra long telomeres (~15–40 kb) (Saltman et al., 1993, Smogorzewska et al., 2000); U2OS cells, because they rely on the alternative lengthening of telomeres (ALT) pathway (Bryan et al., 1997); and HAP1 cells, because they are near haploid (Kotecki et al., 1999) and thus rely on fewer concomitant gene disruption events.

The plasmids containing the sgRNA sequences were transfected into cells, which were single-cell sorted for those fluorescing both green and red. Subsequently, analysis of SNM1B gene disruption was performed.



Figure 5.2 SNM1B partial gene sequence showing the CRISPR guide sequence cut sites.

Exons 1 and 2 are indicated by uppercase, bold letters, and the intervening intron is indicated by lowercase letters. The first guide sequence, cloned into the eGFP-expressing pX330 plasmid is indicated in green; the second guide sequence(s), cloned into the Ruby-expressing pX335 plasmid are indicated in red. Note that there are two second guide sequences and these were cloned separately and co-transfected with the eGFP-expressing plasmid. Guide sequences were obtained using the CRISPR design tool from the Zhang lab (MIT), and selected with the help of Dr Philip Hublitz, from the Genome Engineering Centre, WIMM. All guides had predicted scores of greater than 90%.

5.2.2. Validation of an SNM1B disrupted cell line

Earlier experiments in this study were unable to detect endogenous SNM1B by Western blots, even after prior enrichment by immunoprecipitation with an SNM1B antibody (pilot experiments; data not shown). This observation is in accordance with previous studies that have reported endogenous expression of SNM1B too low to be detected by Western blot using available antibodies (Demuth et al., 2008, Demuth et al., 2004, van Overbeek and de Lange, 2006). As such, a PCR screen was used as an initial assessment of SNM1B gene disruption.

PCR primers were designed to flank exon 1 and to amplify the region around which the guide RNA sequences were designed to bind, and the Cas9 nuclease to subsequently cleave (Figure 5.2). An undisrupted genetic sequence would result in an amplicon of 510 bp, and a double cut, at the sites where each of the RNA guide sequences bind, would result in an amplicon between ~250–330 bp (depending on the specific guide sequences used). As can be seen in Figure 5.3, initial screening identified only non-disrupted and heterozygous clones (upward of 150 clones were analysed, and only a subset are exhibited). Consequently, clones exhibiting heterozygous disruption (for example clone 4, from Figure 5.3) were used for an iterative round of transfection, sorting, and screening.

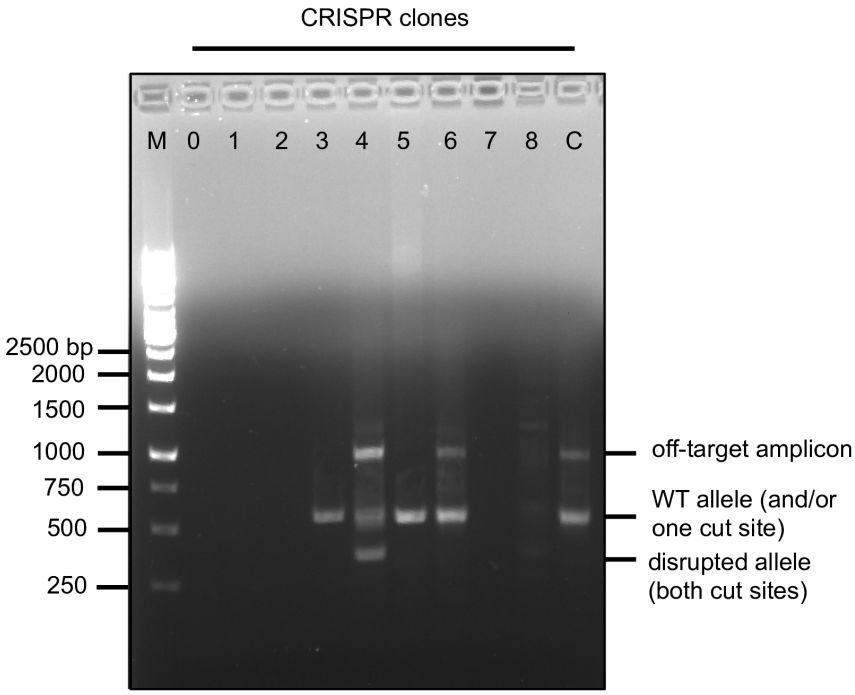


Figure 5.3 Initial PCR screen for transfected and sorted CRISPR clones.

Genomic DNA was extracted from the indicated clones and a PCR screen was performed, as outlined in Chapter 2. CRISPR clones are as indicated and C denotes control WT HeLa1.2.11 cells. Primers were designed to flank the CRISPR cut sites in the SNM1B gene, and the resulting WT amplicon consisted of 510 bp. A disrupted allele at both cut sites resulted in a shorter amplicon (between ~170–~250 bp). An off-target amplicon at 1000 bp was also observed, however, this did not contain the SNM1B genetic sequence. PCR amplicons were analysed by 1% agarose gel electrophoresis and visualised by ethidium bromide. A 1 kb Plus Marker was run and approximate molecular weights, in base pairs, are indicated

Subsequent genomic PCR screening identified a number of clones that appeared to be heterozygous (i.e. there were two PCR amplicons – one 510 bp, and one ~250 or ~330 bp), and for all of these clones, the PCR bands were excised and sequenced. Of these, two clones appeared to show the greatest degree of

disruption – clone 49 and clone 50. Agarose gel electrophoresis of the gDNA PCR products for both clone 49 and 50, revealed three significant amplicons (Figure 5.4).

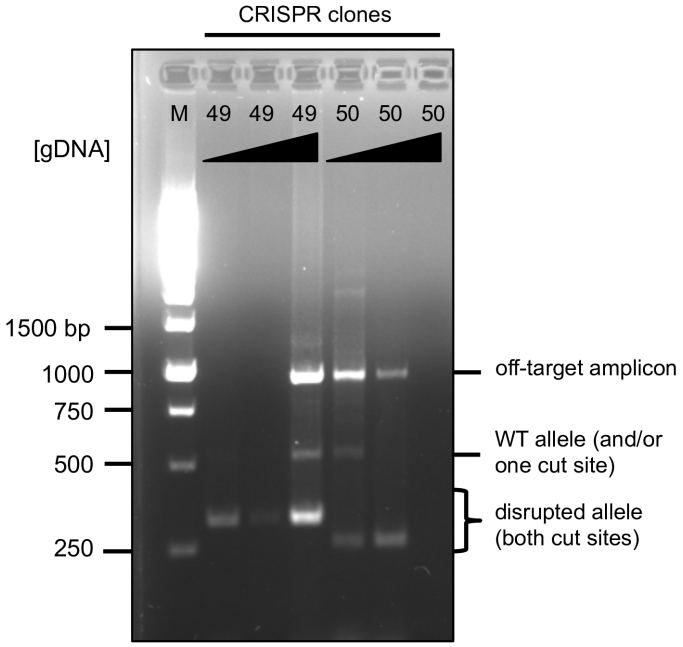


Figure 5.4 gDNA PCR screen for clones 49 and 50.

Genomic DNA was extracted from clone 49 and clone 50 and a PCR screen was performed, as outlined in Chapter 2, with increasing concentrations of genomic DNA. Clone 49 and clone 50 are as indicated. PCR primers and subsequent amplicon sizes are as described in Figure 5.3. PCR amplicons were analysed by 1% agarose gel electrophoresis and visualised by ethidium bromide. A 1 kb Plus Marker was run and approximate molecular weights, in base pairs, are indicated.

The largest amplicon was the ~1000 bp product, which sequencing revealed to be an off-target product. The second was the ~500 bp product, and the third was the ~330 bp (clone 49) or ~250 bp (clone 50) product (Figure 5.4). It is important to note, that for both clone 49 and clone 50, varying concentrations of genomic DNA were used for the PCR (Figure 5.4). This was an attempt to mitigate the

preferential amplification problem, whereby alleles of a gene in a heterozygous sample are amplified at differing levels relative to one another (Walsh et al., 1992).

To further examine the genetic disruption for both clone 49 and 50, the PCR products were excised, gel purified, and analysed by sequencing. The services of Eurofins Genomics were utilised for all sequencing analysis; their method utilises dideoxy-chain termination sequencing.

Sequencing analysis for clone 49 revealed that the ~330 bp product resulted from a loss of genetic information, corresponding with a dual cut at both guide sequence recognition sites (Figure 5.5). This resulted in a frameshift mutation and the introduction of a premature stop codon. The putative translated protein product from this genetic sequence thus showed significant truncation, compared to the WT SNM1B protein (Figure 5.6, Figure 5.5).

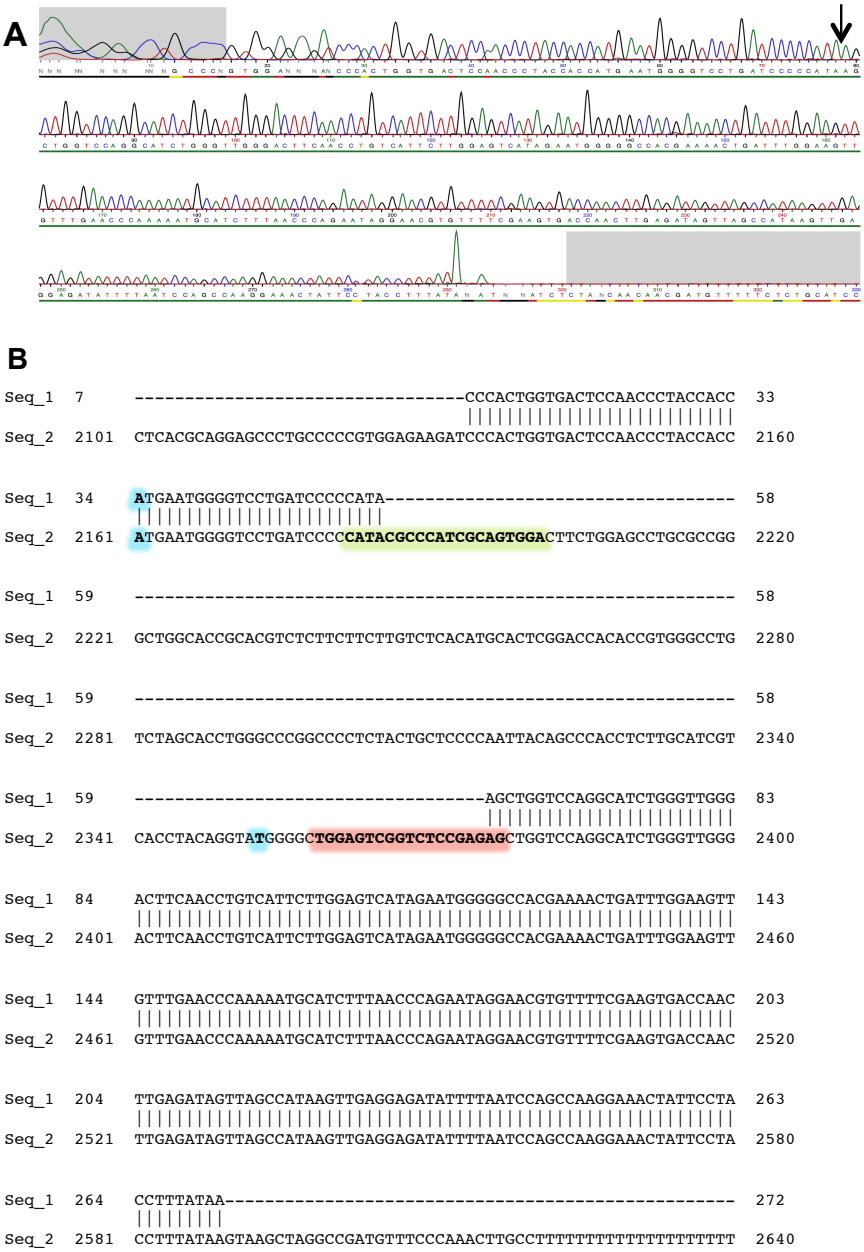


Figure 5.5 Sequencing analysis for clone 49 (~330 bp fragment).

Section A displays the quality report for the sequencing result. The arrow indicates the site of the truncation. Section B displays the clone 49 sequencing results (Seq_1) aligned with the SNM1B genomic DNA sequence (Seq_2). The green and red sequences indicate the target DNA, for which the guide sequences were cloned into either the eGFP or Ruby expressing vector. The beginning and end of exon 1 are indicated in blue. Sequence alignment was performed using Serial Cloner software.

A

M N G V L I P H S W S R H L G W D F N L S F L E S Stop

B

```
Clone_49          1 MNGVLIPHS------WSRHLGWDFNLSFLES------------------     25
                    ||||||||:      ||........|.||..
WT_SNM1B          1 MNGVLIPHTPIAVDFWSLRRAGTARLFFLSHMHSDHTVGLSSTWARPLYC     50

Clone_49         26 --------------------------------------------------     25

WT_SNM1B         51 SPITAHLLHRHLQVSKQWIQALEVGESHVLPLDEIGQETMTVTLLDANHC    100

Clone_49         26 --------------------------------------------------     25

WT_SNM1B        101 PGSVMFLFEGYFGTILYTGDFRYTPSMLKEPALTLGKQIHTLYLDNTNCN    150

Clone_49         26 --------------------------------------------------     25

WT_SNM1B        151 PALVLPSRQEAAHQIVQLIRKHPQHNIKIGLYSLGKESLLEQLALEFQTW    200

Clone_49         26 --------------------------------------------------     25

WT_SNM1B        201 VVLSPRRLELVQLLGLADVFTVEEKAGRIHAVDHMEICHSNMLRWNQTHP    250

Clone_49         26 --------------------------------------------------     25

WT_SNM1B        251 TIAILPTSRKIHSSHPDIHVIPYSDHSSYSELRAFVAALKPCQVVPIVSR    300

Clone_49         26 --------------------------------------------------     25

WT_SNM1B        301 RPCGGFQDSLSPRISVPLIPDSVQQYMSSSSRKPSLLWLLERRLKRPRTQ    350

Clone_49         26 --------------------------------------------------     25

WT_SNM1B        351 GVVFESPEESADQSQADRDSKKAKKEKLSPWPADLEKQPSHHPLRIKKQL    400

Clone_49         26 --------------------------------------------------     25

WT_SNM1B        401 FPDLYSKEWNKAVPFCESQKRVTMLTAPLGFSVHLRSTDEEFISQKTREE    450

Clone_49         26 --------------------------------------------------     25

WT_SNM1B        451 IGLGSPLVPMGDDDGGPEATGNQSAWMGHGSPLSHSSKGTPLLATEFRGL    500

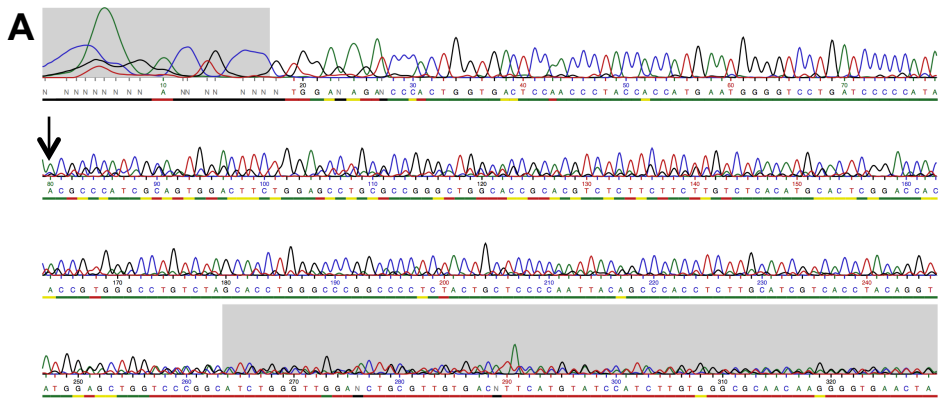
Clone_49         26 --------------------------------     25

WT_SNM1B        501 ALKYLLTPVNFFQAGYSSRRFDQQVEKYHKPC    532
```

Figure 5.6 Putative protein product from the ~330 bp PCR amplicon of clone 49.

Section A: the nucleotide sequence from the ~330 bp fragment translated into amino acid sequence. Section B: alignment of the truncated clone 49 amino acid sequence (Clone_49) with the full length SNM1B sequence (WT_SNM1B). Alignment was performed using the Pairwise Sequence Alignment tool from EMBL-EBI.

Interestingly, sequencing results from the ~500 bp product suggested that there had been a successful cut at each of the guide sequence recognition sites (Figure 5.7).



B

```
Seq_1   28     ----------------------------------CCCACTGGTGACTCCAACCCTACCAC   53
                                                 ||||||||||||||||||||||||||
Seq_2   2100   CCTCACGCAGGAGCCCTGCCCCCGTGGAGAAGATCCCACTGGTGACTCCAACCCTACCAC   2159

Seq_1   54     CATGAATGGGGTCCTGATCCCCCATAACGCCCATCGCAGTGGACTTCTGGAGCCTGCGCC   113
               ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Seq_2   2160   CATGAATGGGGTCCTGATCCCCCATA-CGCCCATCGCAGTGGACTTCTGGAGCCTGCGCC   2218

Seq_1   114    GGGCTGGCACCGCACGTCTCTTCTTCTTGTCTCACATGCACTCGGACCACACCGTGGGCC   173
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2   2219   GGGCTGGCACCGCACGTCTCTTCTTCTTGTCTCACATGCACTCGGACCACACCGTGGGCC   2278

Seq_1   174    TGTCTAGCACCTGGGCCCGGCCCCTCTACTGCTCCCCAATTACAGCCCACCTCTTGCATC   233
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2   2279   TGTCTAGCACCTGGGCCCGGCCCCTCTACTGCTCCCCAATTACAGCCCACCTCTTGCATC   2338

Seq_1   234    GTCACCTACAGGTATGG--------------------AGCTGGTCCCGGCATCTGGGTTG   273
               ||||||||||||||||                    |||||||| ||||||||||||||
Seq_2   2339   GTCACCTACAGGTATGGGGCTGGAGTCGGTCTCCGAGAGCTGGTCCAGGCATCTGGGTTG   2398
```

Figure 5.7 Sequencing analysis for clone 49 (~500 bp fragment).

Section A displays the quality report for the sequencing result. Section B displays the clone 49 sequencing results (Seq_1) aligned with the SNM1B genomic DNA sequence (Seq_2). The green and red sequences indicate the target DNA, for which the guide sequences were cloned into either the eGFP or Ruby expressing vector. The beginning and end of exon 1 are indicated in blue and the arrow indicates the site of the additional nucleotide. Sequence alignment was performed using Serial Cloner software.

This cut at the first target DNA sequence resulted in the addition of a single adenine nucleotide (Figure 5.7). At the second target DNA sequence this resulted in resection and a loss of 20 bp. However, this resection at the second target DNA site is beyond exon 1, in intron 1 (Figure 5.7). Thus, this loss of genetic information is unlikely to affect the coding DNA sequence. Nonetheless, the single insertion of an adenine, in the first target DNA sequence, resulted in a frameshift mutation and the introduction of premature stop codon. The putative translated protein product from this genetic sequence again showed an extreme truncation, when compared to the WT protein (Figure 5.8).

A

M N G V L I P H N A H R S G L L E P A P G W H R T S L L
L V S H A L G P H R G P V **Stop**

B

```
Clone_49          1 MNGVLIPH--------NAHRSGLLEPAPGWHRTSLLLVSHALGPHRGPV-     41
                    ||||||||        :..|:|         ...|..:||....|...:
WT_SNM1B          1 MNGVLIPHTPIAVDFWSLRRAG---------TARLFFLSHMHSDHTVGLS     41

Clone_49         42 --------------------------------------------------     41

WT_SNM1B         42 STWARPLYCSPITAHLLHRHLQVSKQWIQALEVGESHVLPLDEIGQETMT     91

Clone_49         42 --------------------------------------------------     41

WT_SNM1B         92 VTLLDANHCPGSVMFLFEGYFGTILYTGDFRYTPSMLKEPALTLGKQIHT    141

Clone_49         42 --------------------------------------------------     41

WT_SNM1B        142 LYLDNTNCNPALVLPSRQEAAHQIVQLIRKHPQHNIKIGLYSLGKESLLE    191

Clone_49         42 --------------------------------------------------     41

WT_SNM1B        192 QLALEFQTWVVLSPRRLELVQLLGLADVFTVEEKAGRIHAVDHMEICHSN    241

Clone_49         42 --------------------------------------------------     41

WT_SNM1B        242 MLRWNQTHPTIAILPTSRKIHSSHPDIHVIPYSDHSSYSELRAFVAALKP    291

Clone_49         42 --------------------------------------------------     41

WT_SNM1B        292 CQVVPIVSRRPCGGFQDSLSPRISVPLIPDSVQQYMSSSSRKPSLLWLLE    341

Clone_49         42 --------------------------------------------------     41

WT_SNM1B        342 RRLKRPRTQGVVFESPEESADQSQADRDSKKAKKEKLSPWPADLEKQPSH    391

Clone_49         42 --------------------------------------------------     41

WT_SNM1B        392 HPLRIKKQLFPDLYSKEWNKAVPFCESQKRVTMLTAPLGFSVHLRSTDEE    441

Clone_49         42 --------------------------------------------------     41

WT_SNM1B        442 FISQKTREEIGLGSPLVPMGDDDGGPEATGNQSAWMGHGSPLSHSSKGTP    491

Clone_49         42 ------------------------------------------     41

WT_SNM1B        492 LLATEFRGLALKYLLTPVNFFQAGYSSRRFDQQVEKYHKPC    532
```

Figure 5.8 Putative protein product from the ~500 bp PCR amplicon of clone 49.

Section A: the nucleotide sequence from the ~500 bp fragment translated into amino acid sequence. Section B: alignment of the truncated clone 49 amino acid sequence (Clone_49) with the full length SNM1B sequence (WT_SNM1B). Alignment was performed using the Pairwise Sequence Alignment tool from EMBL-EBI.

Sequencing analysis of clone 50 revealed that the ~250 bp product had loss of genetic information resulting from dual cleavage at both guide sequence recognition sites (Figure 5.9).

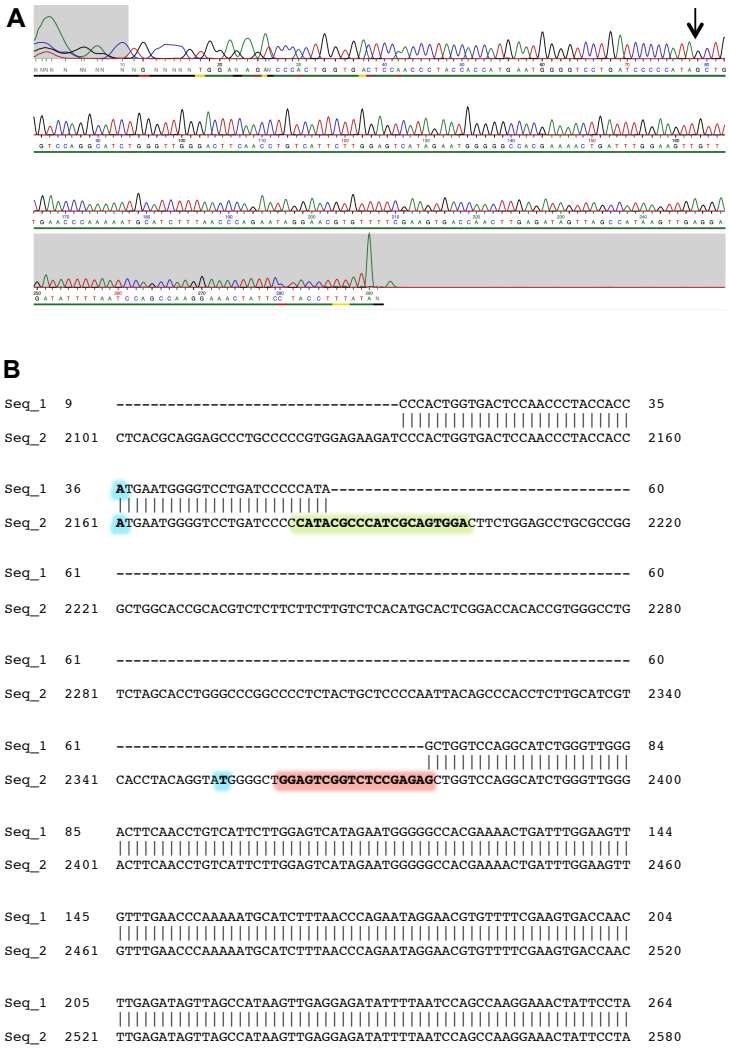
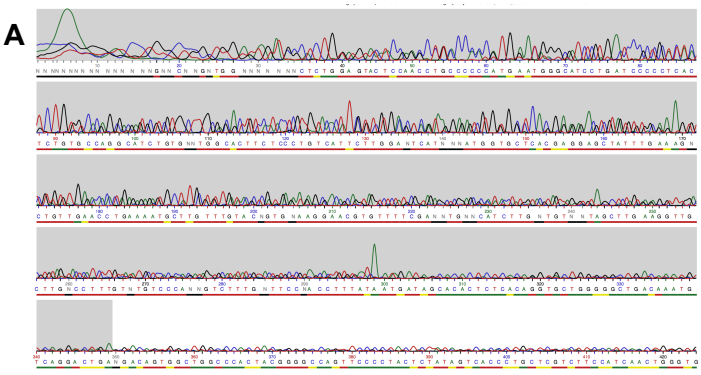


Figure 5.9 Sequencing analysis for clone 50 (~250 bp fragment).

Section A displays the quality report for the sequencing result. The arrow indicates the site of the truncation. Section B displays the clone 49 sequencing results (Seq_1) aligned with the SNM1B genomic DNA sequence (Seq_2). The green and red sequences indicate the target DNA, for which the guide sequences were cloned into either the eGFP or Ruby expressing vector. Sequence alignment was performed using Serial Cloner software.

However, sequencing results from the ~500 bp product, revealed that instead of containing the SNM1B genetic sequence, this amplicon instead exhibited a 96% match with the genetic sequence located at chromosome 7, GRCh38.p7 (Figure 5.10). It is worth noting that this is a non-coding region of the genome. However, it was unclear whether this was due to a chromosomal rearrangement or an off-target effect as a result of the CRISPR-Cas9 process, or as an off-target sequencing amplicon. It is also evident (Figure 5.10) that the sequencing result was of very poor quality, and suggested the presence of multiple genetic sequences. This was repeated a number of times, however, did not significantly improve.



B

```
Query  9        GATAGCACACTCTCACAGGTGCTGGGGGCTGACAAATGTCAGGTCTGAGGACAGTGGCTG  68
Sbjct  6934656  GATAGCACACTCTCACAGGTGCTGGGGGCTGACAAATGTCAGGTCTGAGGACAGTGGCTG  6934715
Query  69       GCCCACTACGGGGCCAGTTCCCCTACTCTATAGTCACCCTGCTCATCTTCCATCAACTGG  128
Sbjct  6934716  GCCCACTACGGGGCCAGTTCCCCTACTCTATAGTCACCCTGCTCGTCTTCCATCAACTGG  6934775
Query  129      GTGCTCAGGACAGTGGCATGGTGGATCCGCCTGTACAGCCTGTGCTCCAGCGTCCTGCAG  188
Sbjct  6934776 GTGCTCAGGACAGTGGCATGGTGGATCCGCCTGTACAGCCTGTGCTCCAGCGTCCTGCAG  6934835
Query  189      GCCACAACGGCGTCC  203
Sbjct  6934836 GCCACAGCGGTGTCC  6934850
```

Figure 5.10 Sequencing analysis for clone 50 (~500 bp fragment)

Section A displays the quality report for the sequencing result. Section B shows the results of a nucleotide sequence search (using the NCBI BLAST tool). The clone 50 nucleotide sequence (denoted 'Query') aligns with 96% similarity with the 'Subject' sequence, which corresponds to chromosome 7, GRCh38.p7.

Sequencing results for clone 49 were simpler to delineate than for clone 50, as the latter potentially contained a genetic insertion, or chromosomal rearrangement. Therefore, clone 49 was used for additional genetic analysis, subsequent assays, and phenotypic characterisation.

To further analyse the degree of SNM1B genetic disruption in clone 49, a qRT-PCR analysis was performed. Accordingly, two sets of qRT-PCR primers were designed and these are illustrated schematically in Figure 5.11. Both sets of primers (denoted SNM1B primer pair #1 and SNM1B primer pair #2, respectively) were designed to span the region of CRISPR-Cas9 mediated genetic disruption.

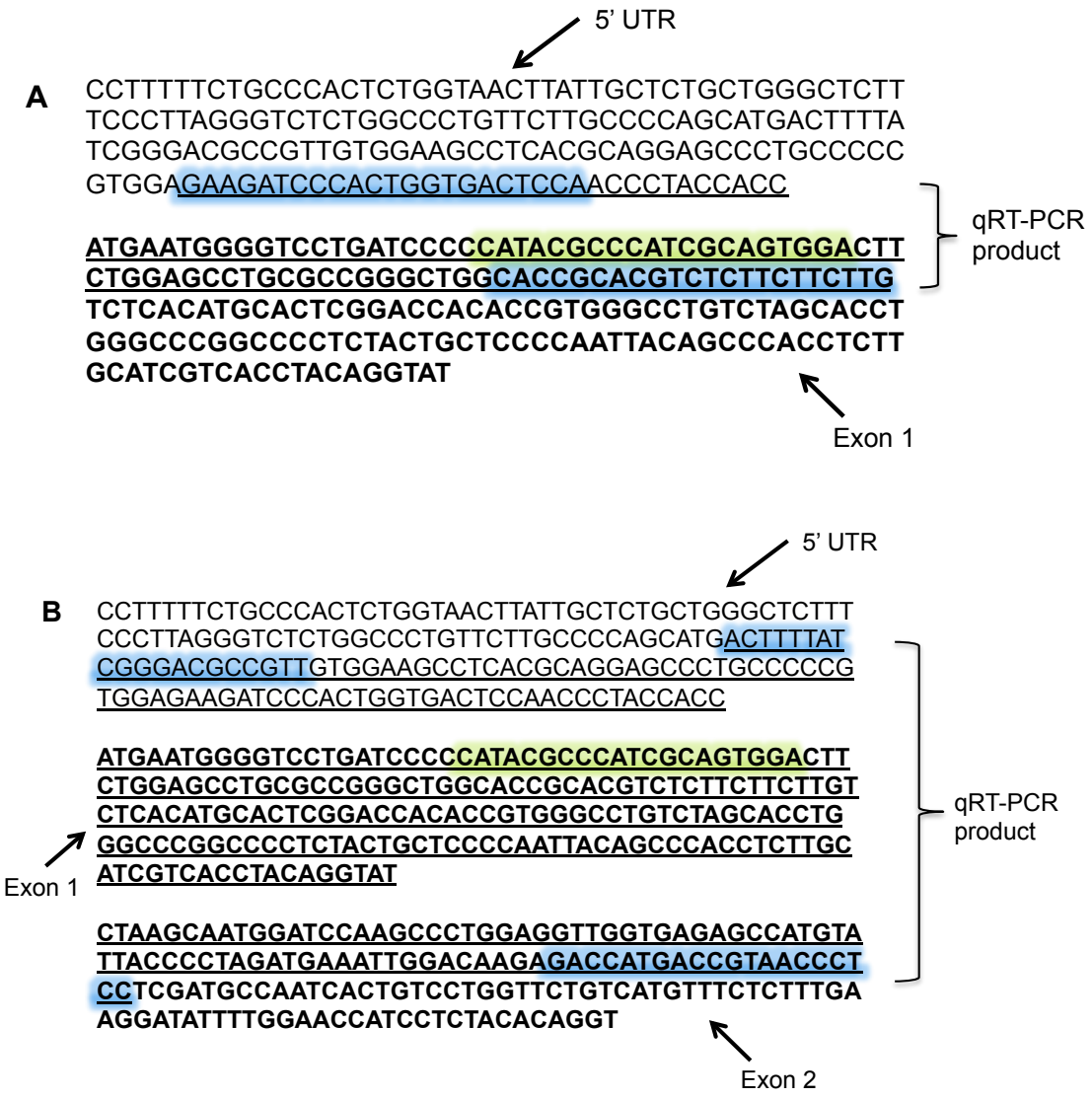


Figure 5.11 qRT-PCR primer design.

Section A shows the design of SNM1B Primer Pair #1. The location of the primers (both forward and reverse) are indicated in blue. The CRISPR-Cas9 DNA target site in exon 1 is indicated in green, and the qRT-PCR amplicon product is underlined. The expected amplicon size is 121 nt. Section B shows the design of SNM1B Primer Pair #2. The location of the primers (both forward and reverse) are indicated in blue. The CRISPR-Cas9 DNA target site in exon 1 is indicated in green, and the qRT-PCR amplicon product is underlined. The expected amplicon size is 372 nt. For each, the 5' UTR, exon 1, and exon 2 are as indicated.

Figure 5.12 details the results of the qRT-PCR analysis of clone 49, in comparison with a WT HeLa1.2.11 cell population. The percentage expression for each primer pair, 53% and 76%, respectively, reflects only a partial reduction in the levels of mature mRNA transcript (Figure 5.12).

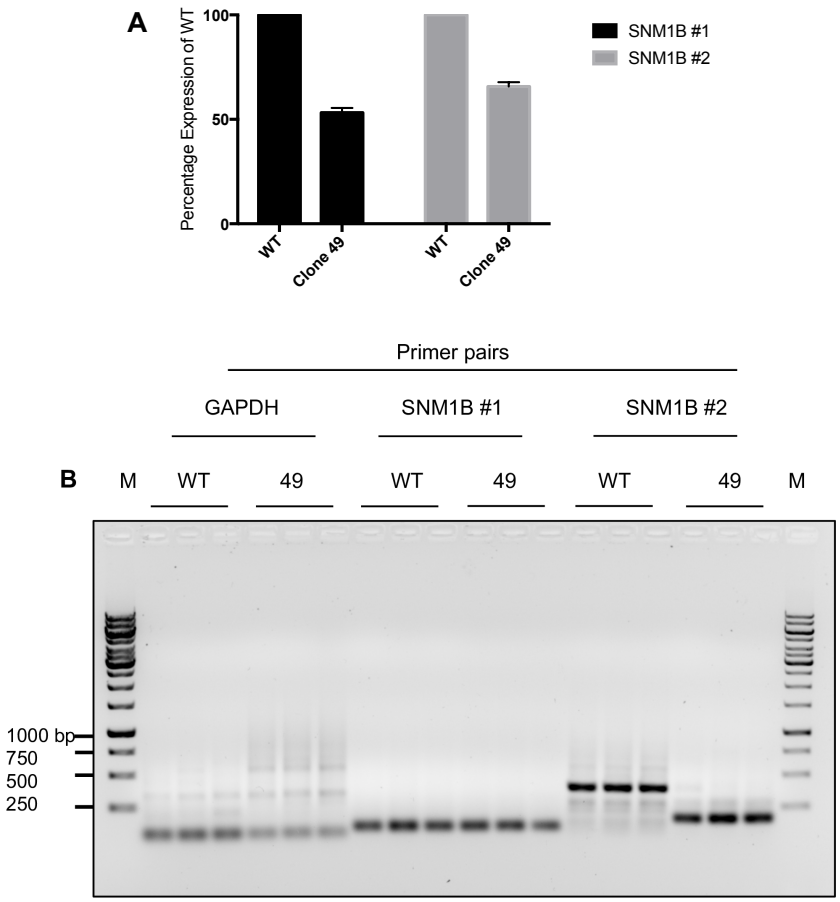


Figure 5.12 qRT-PCR analysis of WT and clone 49 HeLa1.2.11 cells.

A: qRT-PCR detection of mRNA levels of SNM1B in both WT and clone 49 HeLa1.2.11 cells. Two sets of primer pairs were used and are as indicated. Expression was standardised to GAPDH, and normalised to WT expression of SNM1B mRNA transcript levels. The qRT-PCR was performed three times and the error bar indicates the SEM. B: agarose gel electrophoresis of the indicated qRT-PCR products. qRT-PCR amplicons were analysed on a 1% agarose gel and visualised with ethidium bromide. A 1 kb Plus Marker was run and approximate molecular weights, in base pairs, are indicated.

However, agarose gel electrophoresis of the PCR amplicons and subsequent sequencing analysis provides a more nuanced evaluation. qRT-PCR amplicons generated from clone 49 using SNM1B primer pair #1 show no obvious size difference to the WT amplicons when analysed by gel electrophoresis (both appear to be the expected size of 116 nt) (Figure 5.12). Nonetheless, sequencing results show that, for clone 49, there was a specific insertion of a single adenine nucleotide, consistent with the site of insertion in the target DNA sequence when the genomic DNA was analysed (Figure 5.13). This insertion is again consistent with a frameshift mutation and the introduction of a premature stop codon. Experimental data suggest mRNA transcripts that contain premature stop codons, and thus encode incomplete protein sequences, are likely to undergo nonsense-mediated mRNA decay, leading to translational repression, and the loss of functional protein expression (Baker and Parker, 2004). This, therefore, is likely to be case for the clone 49 mRNA transcript. However, in the absence of being unable to directly assess protein levels and thus validate this assumption, it is also possible that exon skipping and aberrant protein production may occur (Sharpe and Cooper, 2017). However, if this were the case then loss of exon 1 would result in an SNM1B protein lacking the majority of the catalytic MBL core (amino acids 1–157), and would be unlikely to possess any nuclease activity. The WT amplicon was also sequenced, and was found to contain no genetic disruptions (data not shown).

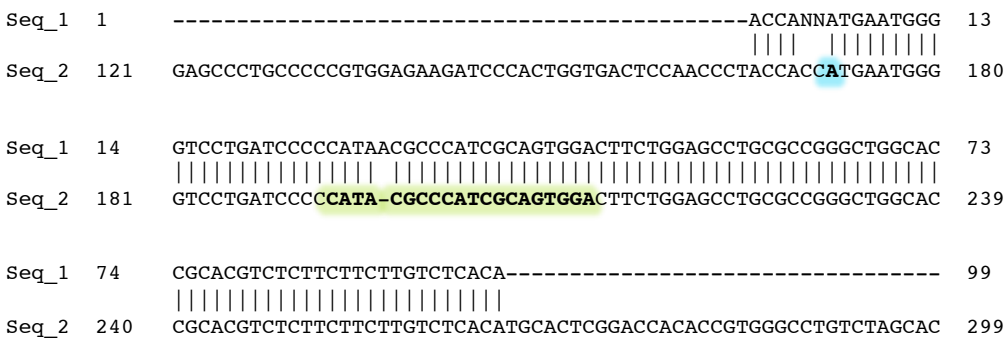


Figure 5.13 qRT-PCR sequencing results with primer pair #1 show disruption to the SNM1B mRNA transcript in clone 49.

qRT-PCR amplicon products were sent for sequencing analysis. The clone 49 qRT-PCR sequencing results from primer pair #1 (denoted Seq_1) are aligned with the SNM1B WT cDNA sequence (Seq_2). The green sequence indicates the guide sequence target DNA site, and the blue indicates the start codon for exon 1. The SNM1B WT cDNA sequence was obtained from a PubMed Nucleotide search and is categorised as follows: cDNA clone MGC:29835 IMAGE: 5001181. Sequence alignment was performed using Serial Cloner software.

For primer pair #2, agarose gel electrophoresis revealed that the clone 49 amplicon was considerably smaller than the WT amplicon (Figure 5.12). The WT amplicon was the correct size of 383 nt, and the clone 49 amplicon appeared to be ~150 nt. Sequencing analysis revealed that this truncation reflected a reduction in the mRNA transcript (Figure 5.14).

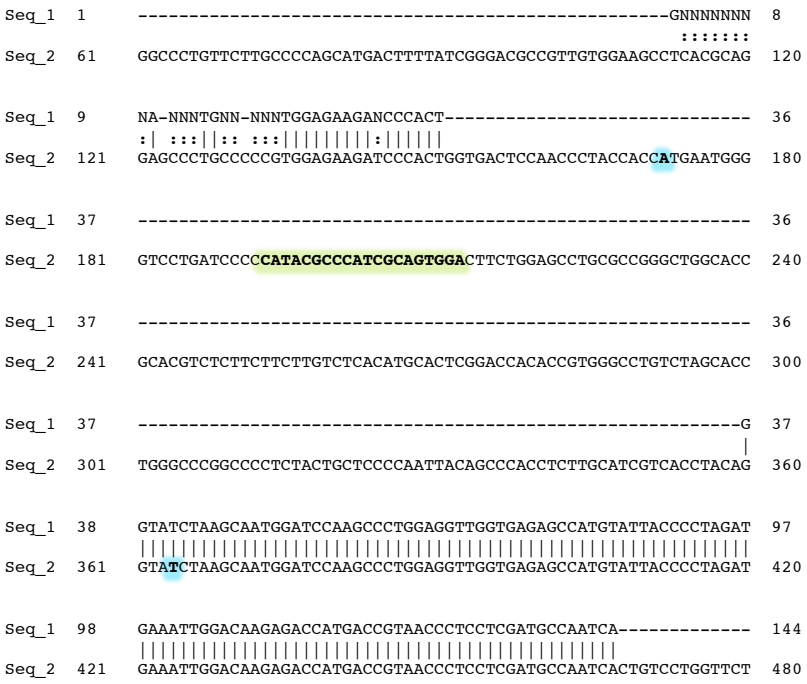


Figure 5.14 qRT-PCR sequencing results with primer pair #2 show disruption to the SNM1B mRNA transcript in clone 49.

qRT-PCR amplicon products were sent for sequencing analysis. The clone 49 qRT-PCR sequencing results from primer pair #2 (denoted Seq_1) are aligned with the SNM1B WT cDNA sequence (Seq_2). The green sequence indicates the guide sequence target DNA site, and the blue indicates both the beginning and end nucleotides for exon 1. The SNM1B WT cDNA sequence was obtained from a PubMed Nucleotide search and is categorised as follows: cDNA clone MGC: 29835 IMAGE: 5001181. Sequence alignment was performed using Serial Cloner software.

Interestingly, there is a greater loss of genetic information at the transcript level when compared to the genomic level (as can be observed in Figure 5.5). Figure 5.14 clearly shows that there is loss of genetic information before the target DNA site prior to the start codon in the qRT-PCR sequencing results that is not lost in the gDNA sequencing results (Figure 5.5); however, the reasons for this are

unclear. It has been observed that if the canonical ATG start codon is deleted, translation can start from a subsequent downstream ATG site (Sharpe and Cooper, 2017). However, in this case, the mRNA transcript exhibits such a loss in genetic information that, even if a downstream in-frame start codon were utilised, the translated product would still lack the first 123 amino acid residues. This would consist nearly the entirety of the catalytic MBL domain, and it is highly unlikely that any expressed protein from this mRNA transcript would possess any substantial nuclease activity. Again, the WT qRT-PCR amplicon was also sequenced and showed no genetic disruption (data not shown).

Taken together, the data show that there is clear disruption to at least two alleles of the SNM1B gene in clone 49 (Figure 5.4). Sequencing of the genomic DNA revealed a loss of genetic information in both PCR amplicons (Figure 5.5 and Figure 5.7) each resulting in a frameshift mutation and a premature stop codon. qRT-PCR analysis revealed a slight decrease in transcript levels (Figure 5.12), and, more importantly, a shorter transcript when the PCR primers spanned the CRISPR cut site. Furthermore, sequencing of the qRT-PCR products closely matched the genomic DNA sequencing data, suggesting genetic disruption that resulted in consistent alterations at the mRNA transcript level (Figure 5.13 and Figure 5.14).

It is worthy of note, however, that it is unclear exactly how many alleles of the SNM1B gene are present in the clone 49 population. The SNM1B gene is located on chromosome 1 (1p13.2) and, in most HeLa cells there appear to be three copies of this chromosome (Landry et al., 2013). However, the HeLa population is

extremely genetically heterogeneous, and individual clones exhibit a large degree of variation (Frattini et al., 2015, Mittelman and Wilson, 2013). For HeLa1.2.11 cells, the data are a little unclear, and no genetic analysis or karyotyping has been performed.

It is therefore entirely possible, for clone 49, that not all alleles of the SNM1B gene have been disrupted. There is no evidence of an undisrupted WT allele when analysed by genomic DNA PCR (and subsequent sequencing), or by qRT-PCR. However, in the absence of quantitative analysis by Western blot, it is difficult to accurately assess protein expression levels. Nonetheless, the data suggest that clone 49 is an SNM1B depleted cell line (if not fully disrupted) and, therefore, is henceforth denoted SNM1B^{-}.

It is perhaps noteworthy that more than 600 clones were sorted, screened, and analysed for genetic disruption, and for both U2OS and HeLa1.2.11 cell lines a substantial number of heterozygous disruptions were observed, however, of the two clones that showed potential disruption to both alleles (clone 49 and clone 50), both were from a HeLa1.2.11 background. Whether this is because depletion of SNM1B is tolerated differently in different cell lines, or whether it is simply that not enough clones have yet been screened is not yet clear. The efficiency of CRISPR-Cas9 editing varies greatly depending on genomic region (Richardson et al., 2016), so it’s not possible to speculate on whether the observed rate of disruption is greater or lesser than expected.

5.2.3. Survival assays

SNM1B has been implicated in a number of cellular processes to maintain genomic integrity. Prior research has indicated that SNM1B depleted cells are sensitive to DNA damaging agents including cisplatin (Nojima et al., 2005, Ishiai et al., 2004, Demuth et al., 2004, Bae et al., 2008), mitomycin C (MMC) (Demuth et al., 2004, Bae et al., 2008), and (in some cases) ionising radiation (Demuth et al., 2004); as well as to the DNA polymerase inhibitor, aphidicolin (Mason et al., 2013).

Accordingly, the cellular survival of SNM1B$^-$ cells treated with a range of DNA damaging agents was examined. These were as follows: cisplatin and MMC, each of which induce ICLs; zeocin, a radiomimetic that predominantly induces chemically modified DSBs; hydroxyurea, which depletes the intracellular dNTP pool and induces replication stress; and etoposide and camptothecin, which inhibit topoisomerase II and I, respectively, also inducing replicative stress. The mechanism of action for each of these drugs is summarised in (Cheung-Ong et al., 2013) and (Chankova et al., 2007).

Short-term survival of both WT and SNM1B$^-$ HeLa1.2.11 cell lines, after treatment with the indicated drugs, was examined by a resazurin viability assay. This assay relies on the observation that viable cells that are actively metabolising will reduce the cell permeable redox indicator, resazurin, into resorufin, which is coloured and fluorescent (Riss et al., 2004). The reduction of resazurin to resorufin is considered proportional to the number of cells, and the amount of resorufin can be

quantified by measuring fluorescence at 560 nm excitation / 590 nm emission (Riss et al., 2004). Initial experiments were performed and both optimal plating density (2000 cells/well), and incubation time with resazurin (60 minutes), was established (pilot experiments; data not shown).

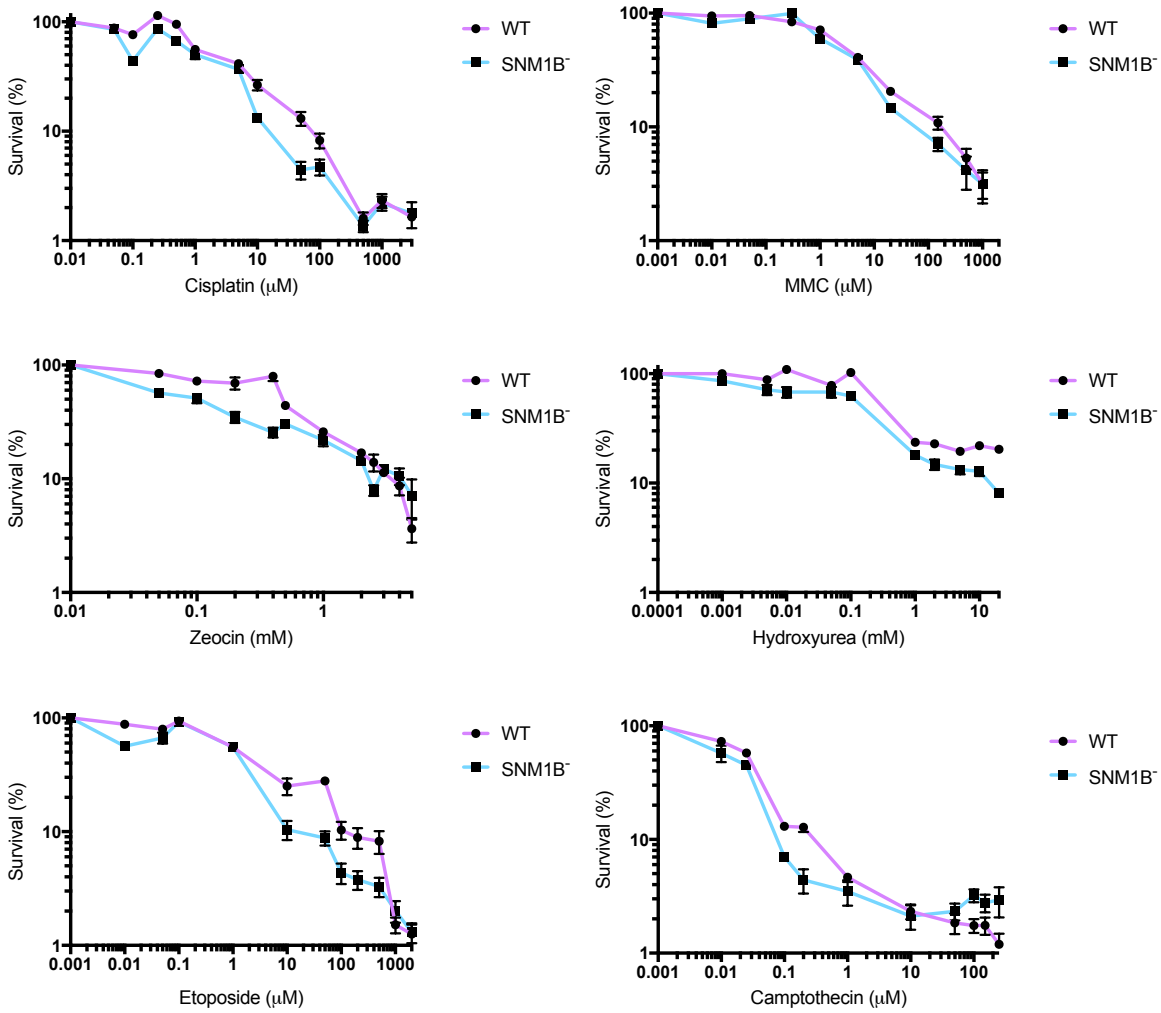


Figure 5.15 Sensitivity profiles of WT and SNM1B$^-$ HeLa1.2.11 cells in response to a variety of DNA damaging agents.

WT or SNM1B$^-$ HeLa1.2.11 cells were plated to 96 well plates and treated with the indicated drug dosage. 72 hours post-treatment cellular viability was assessed with a resazurin survival assay. Percentage survival compared with untreated samples was determined. The mean of at least three independent experiments is indicated and the error bars represent the SEM.

However, as the data in Figure 5.15 show, SNM1B$^-$ HeLa1.2.11 cells exhibit no striking hypersensitivity to any of the examined DNA damaging agents, when compared to WT cells. The reason for this is unclear, but there are a couple of possibilities. Firstly, there may still be some residual SNM1B protein that is being expressed and thus no loss-of-function phenotype was observed. Secondly, cellular viability over a short-term period may not necessarily associate with long-term survival and colony formation. It has been reported that drug treatment and examination of short-term survival is not necessarily correlated with drug-induced cytotoxicity, as measured by the ability to survive, divide iteratively, and form a colony over a longer period of time (Brown, 1997). As such, colony survival assays, where long-term survival and proliferation is measured over a period of 10+ days, is often considered a better model for reproducing *in vivo* conditions (Brown, 1997). In accordance with this, a number of studies that examined the sensitivity of SNM1B depleted cells in response to DNA damaging agents utilised colony formation assays (Nojima et al., 2005, Demuth et al., 2004, Ishiai et al., 2004, Bae et al., 2008). In fact, the only exception was the analysis of the sensitivity profile of SNM1B depleted cells to aphidicolin (Bae et al., 2008).

5.2.4. Clonogenic survival assays

Due to the aforementioned reasons, the sensitivity profile of SNM1B$^-$ cells to a variety of DNA damaging agents was examined by a clonogenic survival assay. This was performed according to the method outlined in Chapter 2.

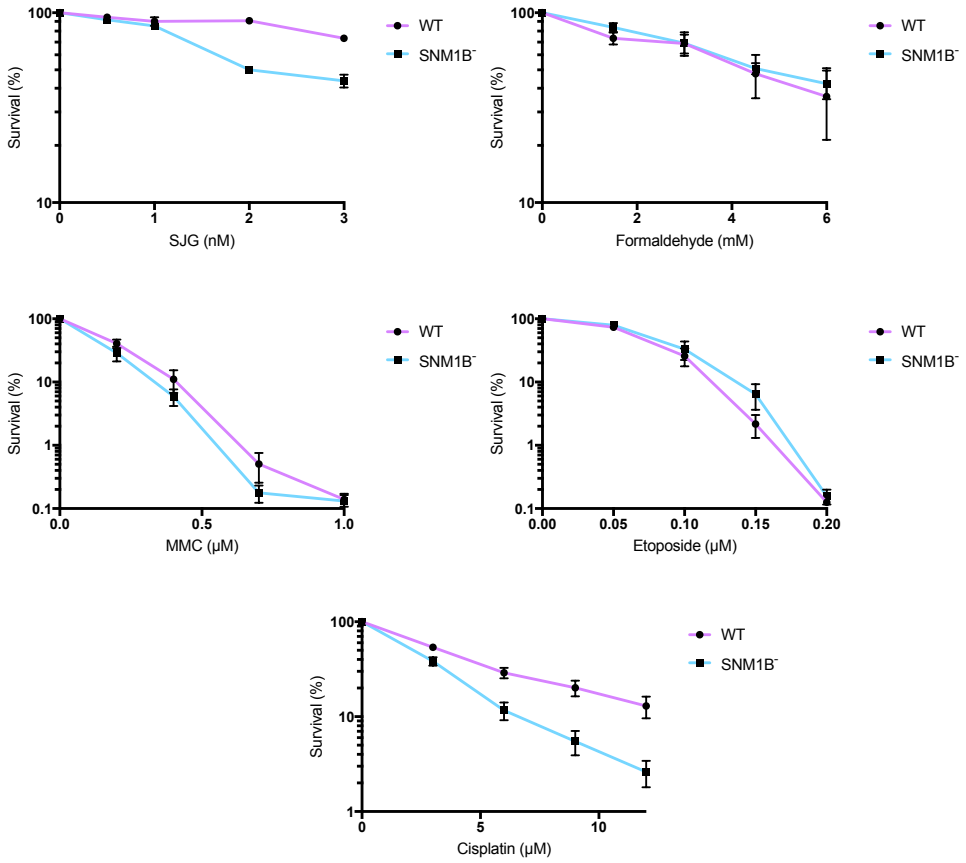


Figure 5.16 Clonogenic survival assays of WT and SNM1B$^-$ HeLa1.2.11 cells in response to DNA damaging agents.

WT or SNM1B$^-$ HeLa1.2.11 cells were plated to 10 cm dishes and subsequently treated with the indicated drug dosage. 10–12 days post-treatment, colony formation was analysed by staining with Coomassie Blue. Colonies were counted and percentage survival compared with untreated samples was determined. Experimental repeats were performed in duplicate, and at least three independent experiments were performed. The mean of these was calculated and graphed, and the error bars represent the SEM.

Interestingly, SNM1B$^-$ HeLa1.2.11 cells did not show any striking sensitivity to etoposide, MMC, or formaldehyde (Figure 5.16), when compared to the WT

HeLa1.2.11 population. However, SNM1B$^-$ cells did exhibit hypersensitivity to both cisplatin and the sequence selective crosslinking agent, SJG-136 (Figure 5.16).

It is of note that SNM1B$^-$ cells exhibit hypersensitivity to the ICL inducing agents, cisplatin and SJG-136, but not to MMC, which also induces ICLs. This differential response to ICL inducing agents is intriguing and it may reflect the type of damage that each of these drugs respectively induces. These differing types of damage are briefly summarised below.

MMC induces ICLs across the minor groove of the DNA, with the near entirety of crosslinks occurring between the exocyclic amino groups of guanine residues of 5'-d(CG) sequences. This results in minimal distortion of the DNA double helix (Noll et al., 2006).

This contrasts with cisplatin which forms major groove adducts. However (as discussed in Chapter 1), cisplatin forms both intrastrand and interstrand crosslinks – greater than 90% of crosslinks that are formed are intrastrand, and only 2–5% are interstrand (Kartalou and Essigmann, 2001). A cisplatin induced ICL also causes significant distortion to the DNA helix – unwinding it by 79^o and bending it by 45^o (Malinge et al., 1994).

SJG-136, another ICL inducing agent, forms crosslinks between guanine residues at the minor groove, with minimal perturbation to the DNA helix (Hopton and Thompson, 2011). Unlike cisplatin, the majority of lesions induced are interstrand crosslinks, as opposed to intrastrand crosslinks (Rahman et al., 2009).

The observed hypersensitivity of SNM1B$^-$ cells to cisplatin is in accordance with previous research (Nojima et al., 2005, Ishiai et al., 2004, Demuth et al., 2004, Bae et al., 2008). The hypersensitivity of SNM1B$^-$ cells to SJG has not been reported previously, however, prior experimental analysis revealed that purified SNM1B is able to digest past an SJG-136 crosslink *in vitro* (Sengerova et al., 2012). Taken together, these results suggest that SNM1B is involved in the cellular repair of SJG-136 induced lesions. Additionally, because SNM1B$^-$ cells are sensitive to SJG-136, a relatively specific interstrand crosslinking agent, this provides some evidence that it is the interstrand crosslinking activity of cisplatin (rather than the intrastrand activity) that renders SNM1B$^-$ cells sensitive.

In contrast to previous research (Demuth et al., 2004, Bae et al., 2008), this study observed no sensitivity to MMC in SNM1B$^-$ HeLa1.2.11 cells. This may be due to differences in phenotype in different cell lines, although (Demuth et al., 2004) also utilised HeLa cells. It is also important to note that MMC is not inherently active; instead it is 'activated' after reduction of its quinone ring by cellular reducing agents. Thus, these differences in phenotype may represent metabolic variability between cellular populations. It may also be that SNM1B is not involved in the repair of all types of interstrand crosslinks, or that its activity is functionally redundant with other DNA repair nucleases (e.g. SNM1A or FAN1).

There is also the possibility that cell clones that survive SNM1B disruption adapt to the deficiency by altering other pathways, perhaps overexpressing some repair

genes. This is a risk with genomically disrupted clones, when compared with short-term disruption or inhibition.

5.3. Conclusions

This study describes the process of generating an SNM1B disrupted HeLa1.2.11 cell line. Initial phenotypic analyses revealed sensitivity to cisplatin and SJG-136, but not formaldehyde, etoposide, and MMC, in a clonogenic survival assay. This sensitivity to cisplatin and SJG-136 is in accordance previous research, which implicates SNM1B in interstrand crosslink repair pathways.

The lack of sensitivity to etoposide, formaldehyde, and MMC is also intriguing. The phenotypic profile of SNM1B^{-} cells in response to MMC treatment has been discussed above. Previous research indicated that SNM1B depleted cells were hypersensitive to the DNA polymerase inhibitor, aphidicolin, which induces replication stress (Mason et al., 2013). Similarly, camptothecin induces replication stress, albeit through a different mechanism. It covalently binds to topoisomerase-I and DNA, thus preventing strand religation and providing a barrier to replication fork progression (Hertzberg et al., 1989). To date, the literature has indicated that some ICL repair factors (such as BRCA1, BRCA2, and FANCJ) participate in the resolution of stalled replication forks, including those induced by aphidicolin and camptothecin (Michl et al., 2016, Matsuzaki et al., 2015). Given the experimental data provided by (Mason et al., 2013), it is entirely possible that SNM1B plays a role in the resolution of some stalled forks, but not others, perhaps those that

require nucleolytic processing, for example, the protein adducted DNA lesions generated by camptothecin.

The sensitivity of SNM1B$^-$ cells to formaldehyde was examined, as the highly reactive nature of formaldehyde induces a variety of DNA-protein crosslinks (DPCs) (Hoffman et al., 2015), and the repair of DPCs in humans is not fully characterised. Smaller DPCs (where the crosslinked protein is <16 kDa) can be processed by NER machinery (Minko et al., 2002). However, only a few proteins have been identified as playing a role in the resolution of larger DPCs. Perhaps most important is the DNA dependent metalloprotease, SPRTN, which enzymatically digests the proteinaceous part of the DPC (Lopez-Mosqueda et al., 2016). The tyrosyl-DNA phosphodiesterases 1 and 2 (TDP1 and 2, respectively) are also required for the repair of topoisomerase-DNA crosslinks (Debethune et al., 2002, Interthal and Champoux, 2011). Additionally, there is some evidence that the nuclease activity of MRE11 (as part of the MRN complex) is involved in the final stages of DPC processing (Aparicio et al., 2016). However, given that SNM1B$^-$ cells did not show any sensitivity to formaldehyde treatment when compared to WT (Figure 5.16), it is unlikely that SNM1B is necessary for the specific repair of a formaldehyde induced DPCs.

Generation of an SNM1B$^-$ HeLa1.2.11 cell line also provides an extremely useful tool for further experimental analysis. The genetic relationship between SNM1B and other DNA damage repair factors is something to be considered in further research, perhaps by concomitant siRNA mediated knockdown and subsequent phenotypic analysis. Particularly interesting is the possible functional redundancy

between SNM1B and other DNA damage repair exonucleases, such as SNM1A, EXO1, or FAN1 in the repair of different DNA lesions. Also of interest, is investigating the role that SNM1B plays in controlling progression through the cell cycle, as reported by (Bae et al., 2008) and (Liu et al., 2009).

There is also the intriguing possibility that SNM1B plays a role in the specific DNA damage response at telomeres. It has recently been established that DNA damage repair mechanisms at telomeres differ markedly from global DNA damage repair responses (Webb et al., 2013). As SNM1B is localised to telomeres under normal cellular conditions (through its interaction with TRF2) this is another interesting avenue for further research.

The cellular analysis in this chapter is presently ongoing and further work is required to definitively validate the SNM1B disruption. It is imperative to confirm that the phenotype of the SNM1B^{-} cells is due to the loss of SNM1B, and not due to a clonal effect. As such it is necessary to restore WT SNM1B expression (perhaps transiently) and assess whether this recapitulates the WT phenotype. It is also possible that there is still some residual SNM1B being expressed (if not all alleles were successfully disrupted) and therefore the process of immunoprecipitation and Western blotting with novel SNM1B antibodies is being optimised.

With regards to future work, it would also be of value to generate an additional SNM1B disrupted clone, perhaps in an alternative cell line. U2OS cells provide an interesting candidate due to their aforementioned reliance on the ALT pathway to

lengthen their telomeres. In ALT cells, such as U2OS, a substantial proportion of telomeric DNA is extrachromosomal and can be either linear (Tokutake et al., 1998, Ogino et al., 1998), or circular (Cesare and Griffith, 2004, Wang et al., 2004), in structure. However, ALT cells do exhibit t-loop formation and normal telomeric localisation of shelterin proteins (Cesare and Griffith, 2004). It may therefore interesting to understand how the DNA damage repair function of SNM1B might differ as a result of different telomeric structures – i.e. is SNM1B able to localise to sites of global DNA damage, when its localisation may be constrained extra-chromosomally?

It is possible that this chapter elicits more questions than answers. However, it is clear that SNM1B is critically important for normal cellular function, and there is much that remains to be further understood and there are many interesting avenues for future research. Accordingly, this study provides an SNM1B disrupted human cell line, a necessary tool for subsequent experimental analyses.

6. A Biochemical Analysis of the Novel Nuclease, EXD2

6.1. Introduction

The role of EXD2 in DNA damage repair has only recently begun to be investigated. EXD2 (or EXDL2) was first identified as a potential candidate in the processes of DNA damage repair when an shRNA screen identified that EXD2 depleted cells exhibited hypersensitivity to the DNA damaging agents mitomycin C (MMC), camptothecin, and the alkylating agent, MMS (Smogorzewska et al., 2010).

Further experimentation in the lab of Dr Wojciech Niedzwiedz (WIMM, Oxford) identified EXD2 as a novel interacting partner of the DNA damage repair factor, CtIP. At the cellular level EXD2 was identified as a key component in facilitating homologous recombination (HR) directed repair of double strand breaks (DSBs). EXD2 was shown to functionally interact with the MRE11-RAD50-NBS1 (MRN) complex and to play a critical role in end resection prior to RAD51 loading and subsequent strand exchange in HR. EXD2 was also recruited to chromatin in a DNA damage-dependent manner and conferred resistance to DNA damaging agents including ionising radiation, camptothecin, and phleomycin (Broderick et al., 2016).

As EXD2 was only recently identified to play a role in DNA damage repair, very little was known about its biological function. At the time of this study, Niedzwiedz and colleagues had amassed a significant amount of preliminary functional data

from cellular biology experiments. However, the biochemical activity of EXD2 had not been defined, and there existed a complete paucity of data regarding its enzymology. As such, a biochemical analysis of the EXD2 protein was carried out.

6.2. Results

6.2.1. Cloning the EXD2 gene

The EXD2 protein contains a predicted exonuclease domain and this spans residues 155–247 in the amino acid sequence (Figure 6.1). This predicted exonuclease fold shows significant sequence similarity with the 5' to 3' exonuclease domain of the Werner protein (WRN) (Figure 6.1).

The region of sequence similarity between EXD2 and WRN maps to amino acids 99–257 in the protein sequence of EXD2. Interestingly, this encompasses an additional 56 residues in the EXD2 protein sequence, prior to the projected exonuclease fold (Figure 6.1). WRN also contains an additional 3' to 5' helicase/ATPase domain; however, there is little similarity with EXD2 in this region. Of note, the EXD2 protein contains no other predicted functional domains, and no protein-protein interaction sites have been mapped.

It is worthy of note that there are two isoforms of the EXD2 protein – the longer 'canonical' 621 amino acid isoform 1, and the shorter 496 amino acid isoform 2 (lacking residues 1–125). As isoform 1 was the primary subject of analysis in the Niedzwiedz lab, all constructs in the present study were also based on this isoform.

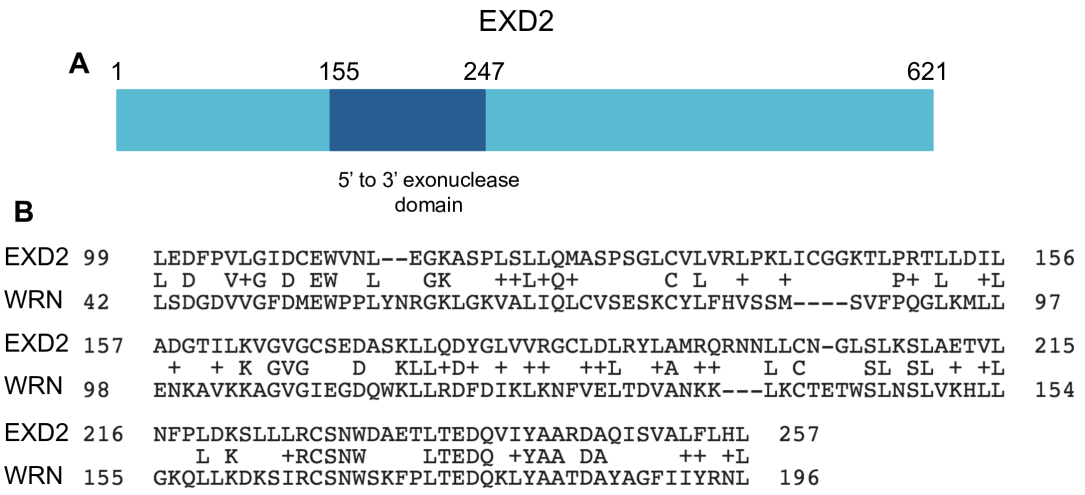


Figure 6.1 The EXD2 protein has a predicted exonuclease domain and shows sequence homology to the WRN protein nuclease/helicase.

A: Linear representation of the EXD2 protein. The predicted 5' to 3' exonuclease fold is coloured dark blue. B: Sequence similarity of the EXD2 protein with the WRN protein. The area of homology spans amino acid residues 99–257 in the EXD2 protein, and maps to the 3' to 5' exonuclease domain of the WRN protein.

Dr Jadwiga Nieminuszczy (from the Niedzwiedz lab) developed a method for purifying very small amounts of full-length GST tagged EXD2 from *E. coli* and this was subsequently utilised in nuclease assays (Figure 6.2). As the predicted nuclease activity was directional, with 3' to 5' polarity, substrates were radiolabelled on the 5' end with γ^{32}P-dATP. As in Chapter 3, a ssDNA substrate was labelled, and annealed to specific oligonucleotide sequences to generate different DNA substrate structures (see Appendix V for DNA sequences). Initial nuclease activity was examined on ss- and splayed arm DNA substrates (Figure 6.3) and these were purposefully selected. The former, because ssDNA is the simplest substrate for nucleolytic digestion. The latter, because EXD2 has sequence similarity with the exonuclease domain of WRN, however, WRN also

possesses intrinsic ATP-dependent helicase activity and can therefore unwind DNA substrates to facilitate digestion. As such, it was hypothesised that EXD2 might catalytically act on a splayed arm substrate as this structurally mimics an unwound dsDNA substrate.

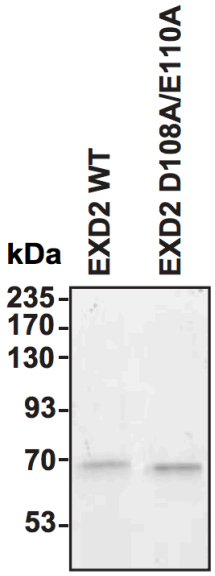


Figure 6.2 Purified full-length WT and D108A/E110A EXD2.

Coomassie stained SDS-PAGE of WT and D108A/E110A EXD2 ectopically expressed in *E. coli* and purified to homogeneity for use in *in vitro* biochemical analysis. Figure obtained from (Broderick et al., 2016).

It was also not known which divalent metal ion would support the putative exonuclease activity of EXD2. As such, increasing concentrations (0.1–10 mM) of candidate divalent metal ions were present in the reaction buffer (Figure 6.3).

In accordance with its predicted activity, purified full-length EXD2 protein showed significant nucleolytic digestion on both ss- and splayed arm DNA substrates in the

presence of 1–10 mM Mn^{2+}, with a preference for a splayed arm substrate compared to ssDNA (Figure 6.3).

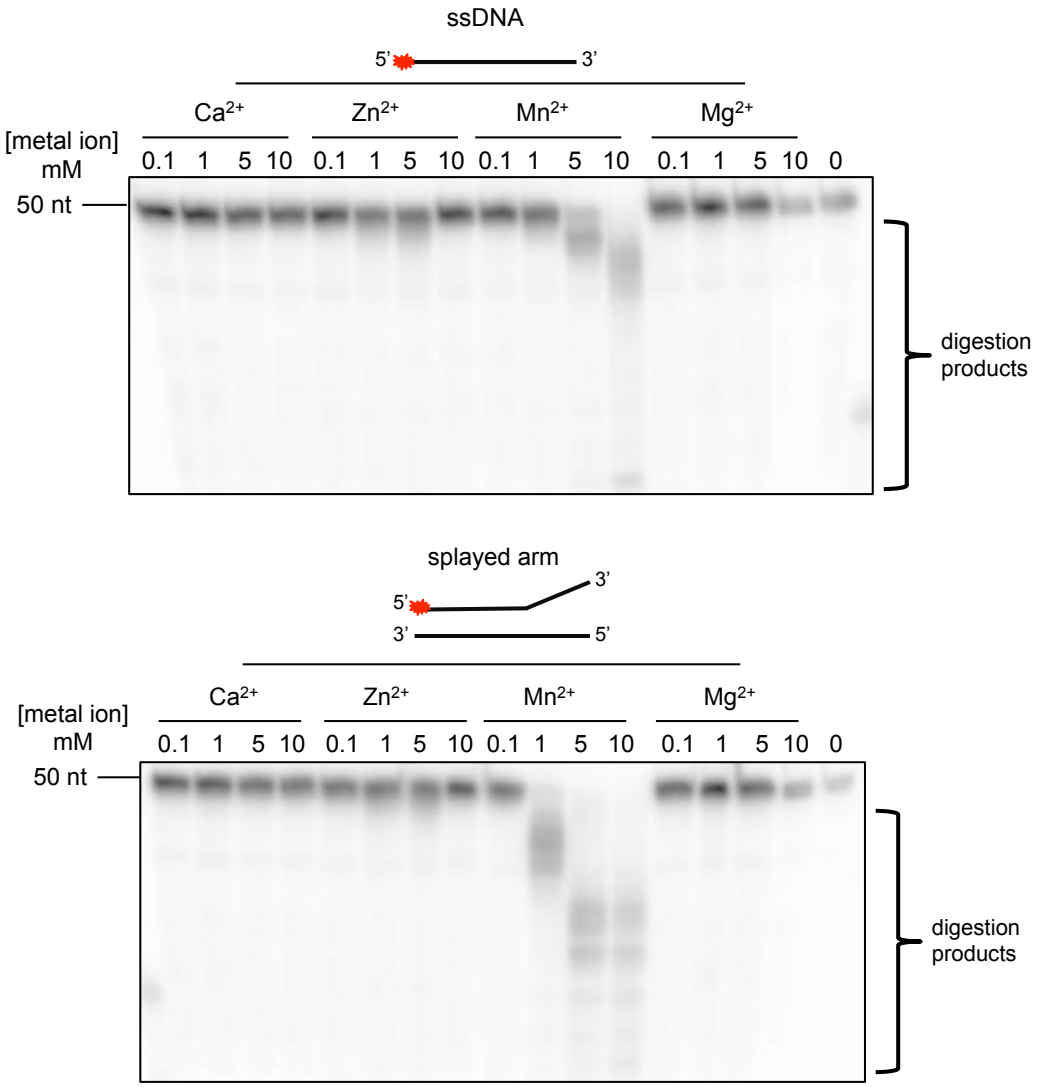


Figure 6.3 Full-length EXD2 exhibits nuclease activity on ss- and splayed arm DNA substrates in the presence of Mn^{2+}.

10 nM of DNA substrate (as labelled) was added to 100 nM EXD2 and incubated at 37°C for 1 hour in the presence of different concentrations of divalent metal ions. The reaction was stopped with the addition of 2 µL stop solution and boiling. Samples were then analysed by 20% polyacrylamide gel electrophoresis. DNA substrate structures are as indicated.

However, results from the purification of the full-length protein from *E. coli* were variable, and neither purity nor yield was optimal, and this precluded efficient biochemical analysis of EXD2. To this end, the EXD2 gene was subcloned for large-scale purification to generate truncated protein for use in biochemical assays.

Dr Opher Gileadi (SGC) performed initial cloning and test expression analysis. Twenty-five constructs were generated, whereby different truncations of the EXD2 gene (each spanning the predicted 3' to 5' exonuclease domain) were subcloned into the pNIC28-Bsa4 vector. Cloning into this vector is schematically represented in Figure 6.4, and resulted in the addition of a 6-His tag at the N-terminus of the truncated EXD2 protein. The 6-His tag was spaced from the protein by a tobacco etch virus (TEV) protease recognition site, thus allowing efficient cleavage of the tag in the purification process. These constructs were subsequently analysed for overexpression in *E. coli.*

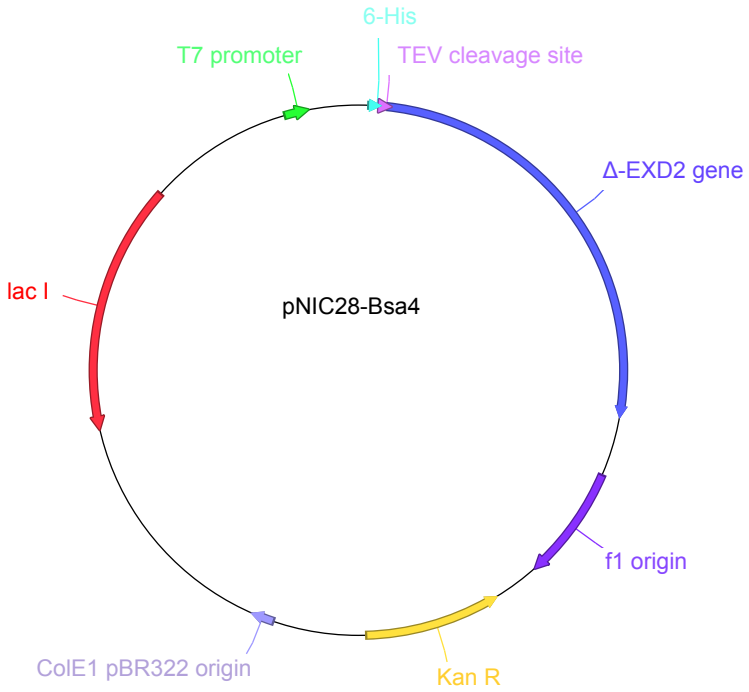


Figure 6.4 Schematic of the truncated EXD2 gene (various truncations) cloned into the pNIC28-Bsa4 vector.

Only three truncated EXD2 proteins were successfully overexpressed (Figure 6.4), and as such, these were scaled up for a large-scale purification. Each of the truncated EXD2 proteins (Δ-EXD2) spanned slightly different regions of the EXD2 protein, and are henceforth denoted as follows: Δ-EXD2 (K76–K283), Δ-EXD2 (K76–G286), Δ-EXD2 (K76–S589) (Figure 6.5).

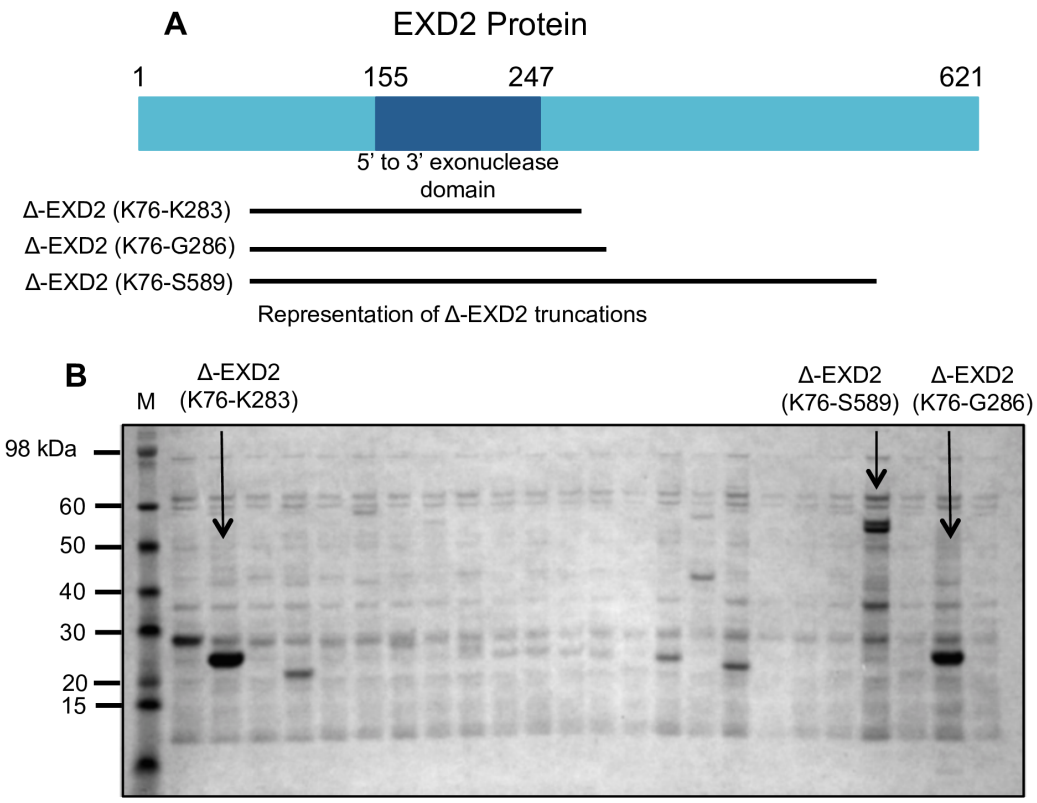


Figure 6.5 Design of successful Δ-EXD2 constructs and SDS-PAGE of subsequent test expression.

A: Linear representation of the EXD2 protein with the predicted exonuclease domain coloured dark blue. The truncated amino acid sequences that constitute each of the successfully overexpressed Δ-EXD2 proteins are as indicated. The sequence of each of the three successfully Δ-EXD2 proteins spans the predicted exonuclease domain. B: SDS-PAGE showing successful overexpression of relevant Δ-EXD2 proteins. For the test expression, an initial Ni-Sepharose column was performed from 1 mL *E. coli* culture, and the elution fraction was analysed by SDS-PAGE for the presence of over-expressed Δ-EXD2. Panel B was produced by Dr Opher Gileadi.

6.2.2. Purification of Δ-EXD2 (K76–K283) and (K76–G286)

Truncated EXD2 proteins were expressed and purified from 4 L of *E. coli* culture. As Δ-EXD2 (K76–K283) and Δ-EXD2 (K76–G286) exhibited the most pronounced over-expression in the test expression protocol (Figure 6.5), they were initially scaled up for larger purification. As each had the addition of an N-terminal 6-His tag, the purification steps were as summarised below (and schematically in Figure 6.6). An initial Ni-Sepharose chromatography column was performed to selectively bind His_6-Δ-EXD2 through interaction with the histidine tag. The presence of Δ-EXD2 (K76–K283) or Δ-EXD2 (K76–G286) in the relevant eluate fractions was validated by the presence of a ~23 kDa band on SDS-PAGE (Figure 6.7). The histidine tag was subsequently cleaved by incubation with the TEV protease, and removal of the tag was confirmed by ESI-TOF-MS (Figure 6.7). A secondary Ni-Sepharose column was performed to remove the TEV protease and any residual uncleaved His_6-Δ-EXD2, and the presence of Δ-EXD2 in eluate fractions was validated by the presence of a ~23 kDa band on SDS-PAGE (Figure 6.7). Δ-EXD2 containing eluate fractions were then pooled, concentrated, and further purified by S75 gel filtration chromatography. Protein containing eluate fractions were identified by absorbance at 280 nm and the presence of the relevant Δ-EXD2 protein was indicated by SDS-PAGE (Figure 6.8). Δ-EXD2 containing fractions were pooled, concentrated, and further purified by MonoQ anion exchange chromatography. Again, protein containing eluate fractions that corresponded with absorbance at 280 nm were analysed by SDS-PAGE (Figure 6.9 and Figure 6.10) and Δ-EXD2 containing fractions were pooled and concentrated for use in biochemical studies. It is worthy of note that both Δ-EXD2 (K76–K283) and Δ-

EXD2 (K76–G286) had a significant shoulder to the left of the main peak eluting from the MonoQ column (Figure 6.9 and Figure 6.10). However, MS analysis and SDS-PAGE revealed no difference between the shoulder and the main peak, so all fractions were pooled. Presence of the correct molecular weight product was validated by ESI-TOF-MS (Figure 6.11).

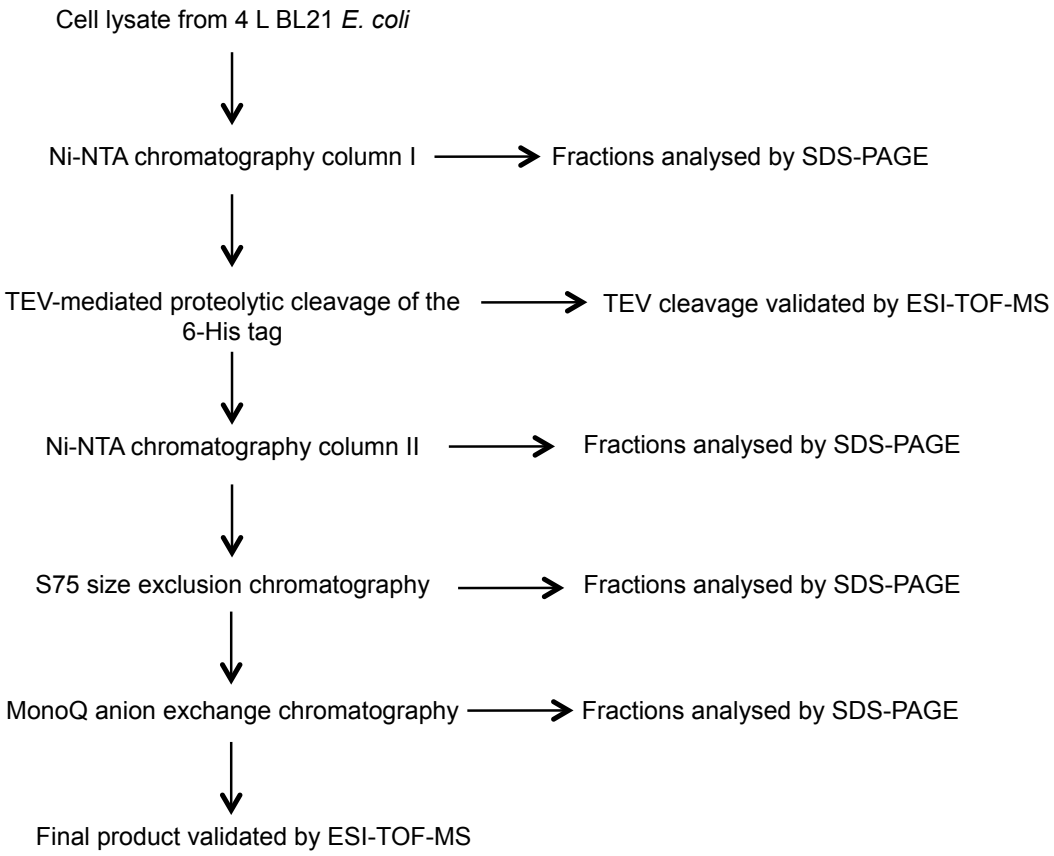


Figure 6.6 Schematic of purification protocol for Δ-EXD2 (K76–K283) and (K76–G286).

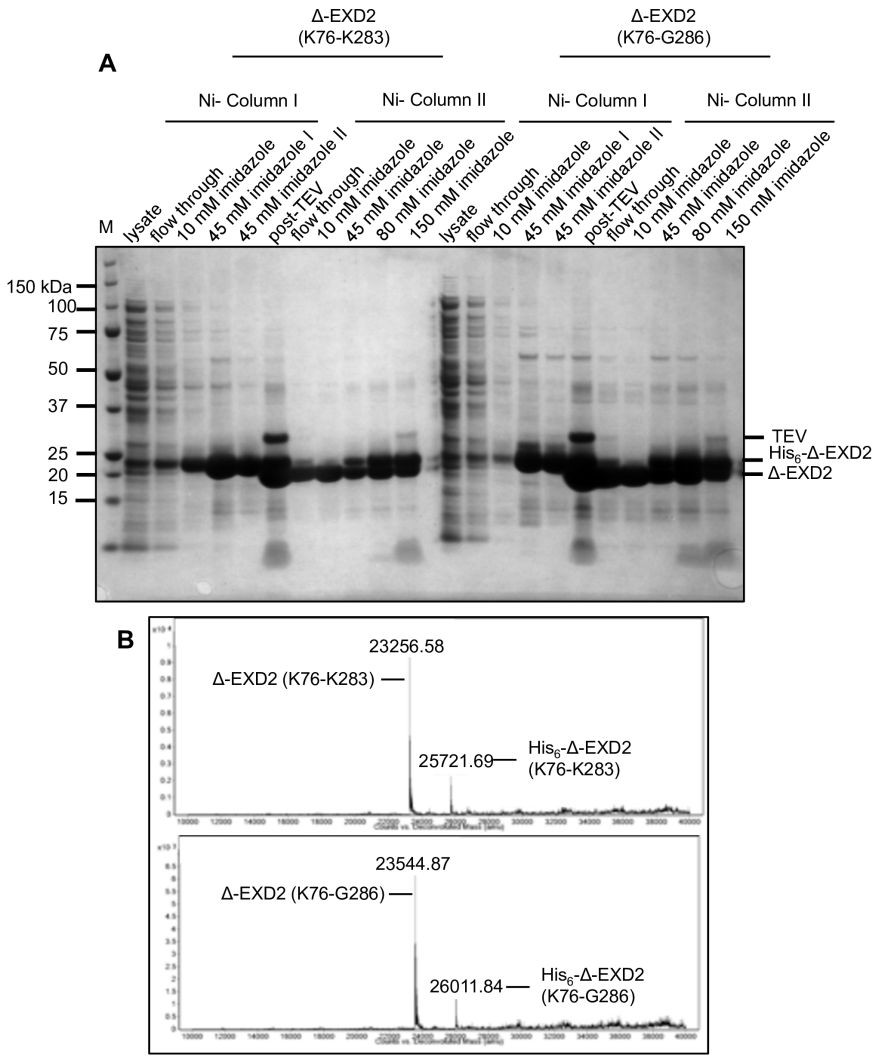


Figure 6.7 Purification of Δ-EXD2 (K76–K283) and (K76–G 286) from *E. coli.*

A: SDS-PAGE of key steps from the cell lysis, initial Ni-Sepharose column, TEV cleavage, and the second Ni-Sepharose column. Samples were collected and analysed by SDS-PAGE for the presence of the relevant Δ-EXD2 indicated by an overexpressed band at ~23 kDa. The approximate molecular weight of His_6-Δ-EXD2, Δ-EXD2, and TEV are all as indicated.

B: ESI-TOF-MS showing cleavage of the 6-His tag after treatment with TEV protease. Δ-EXD2 (K76–K283) expected mass = 23257 Da, measured mass = 23256.58 Da, Δ-EXD2 (K76–G286) expected mass = 23545.4 Da, measured mass = 23544.87 Da.

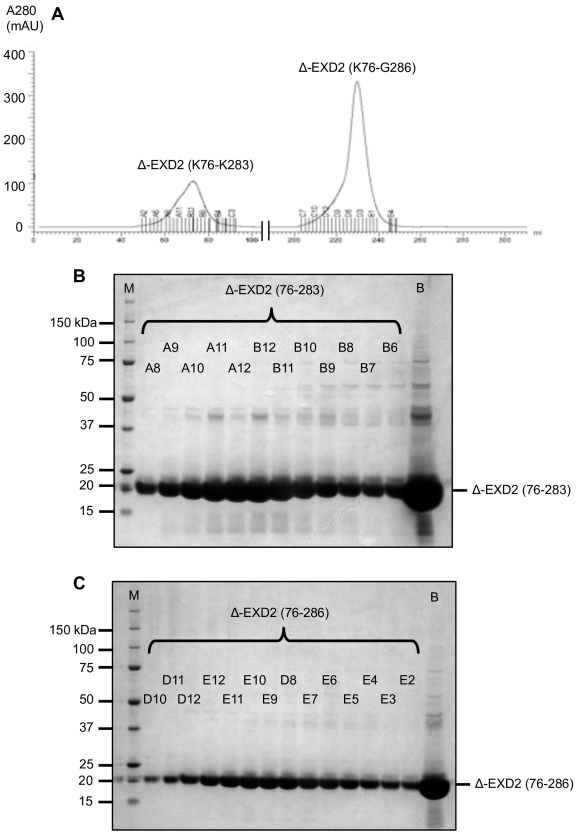


Figure 6.8 Superdex S75 gel filtration for Δ-EXD2 (K76–K283) and for Δ-EXD2 (K76–G286).

Panel A displays absorbance at 280 nm across collected fractions for Δ-EXD2 (76–283) and Δ-EXD2 (76–286) after S75 Superdex gel filtration chromatography. Note that both proteins were purified on the same column consecutively. B: SDS-PAGE of peak protein containing fractions, as indicated in A for Δ-EXD2 (76–283). C: SDS-PAGE of peak protein containing fractions, as indicated in A for Δ-EXD2 (76–286). The presence of a ~23 kDa band across the peak fractions indicates the presence Δ-EXD2 (76–283) or Δ-EXD2 (76–286), as indicated. The bracket indicates purified protein fractions that were pooled and concentrated for MonoQ ion exchange chromatography. M denotes molecular weight marker, and B, the protein sample before it was loaded onto the S75 Superdex gel filtration column.

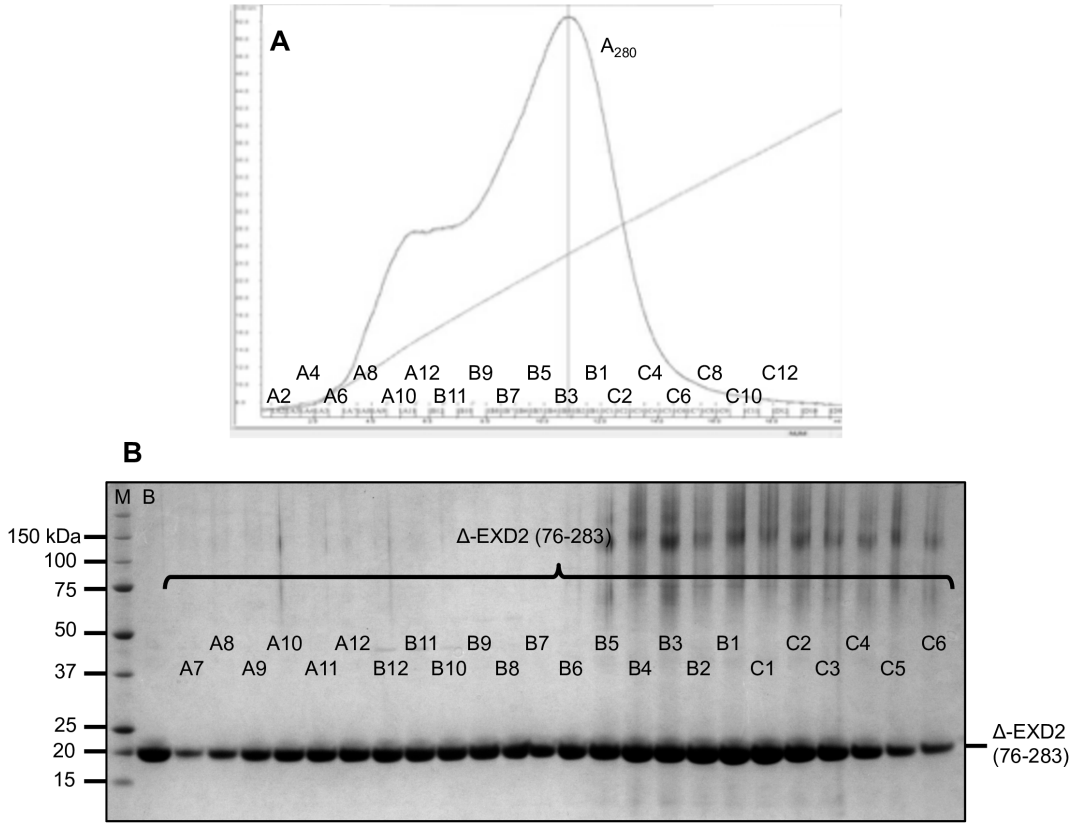


Figure 6.9 MonoQ anion exchange chromatography for Δ-EXD2 (K76–K283).

A: Absorbance at 280 nm across collected fractions for Δ-EXD2 (76–283) after MonoQ anion exchange chromatography. Note that only every second fraction is labelled. B: SDS-PAGE of peak protein containing fractions, as indicated in A for Δ-EXD2 (76–283). The presence of a ~23 kDa band across the peak fractions indicates the presence Δ-EXD2 (76–283). The bracket indicates purified protein fractions that were pooled and concentrated for use in biochemical assays. M denotes molecular weight marker, and B, the protein sample before it was loaded onto the MonoQ column.

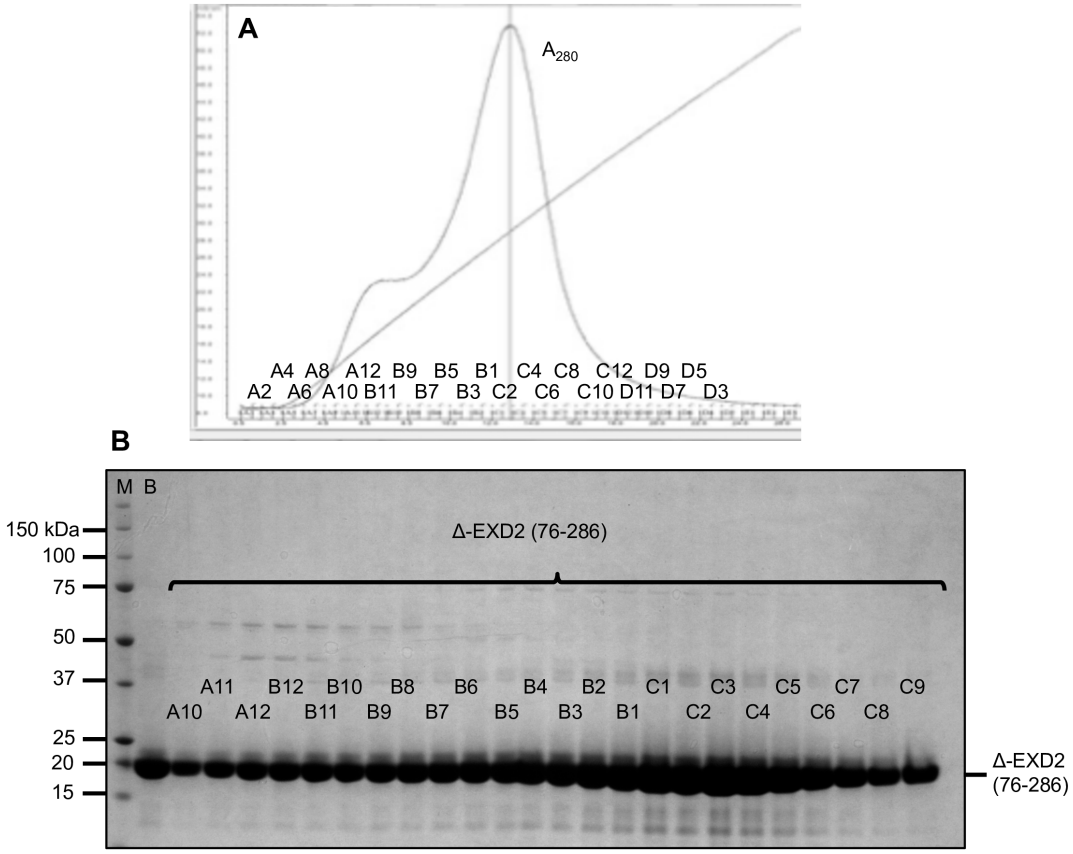


Figure 6.10 MonoQ anion exchange chromatography for Δ-EXD2 (K76–G286).

A: Absorbance at 280 nm across collected fractions for Δ-EXD2 (76–286) after MonoQ anion exchange chromatography. Note that only every second fraction is labelled. B: SDS-PAGE of peak protein containing fractions, as indicated in A for Δ-EXD2 (76–286). The presence of a ~23 kDa band across the peak fractions indicates the presence Δ-EXD2 (76–286). The bracket indicates purified protein fractions that were pooled and concentrated for use in biochemical assays. M denotes molecular weight marker, and B, the protein sample before it was loaded onto the MonoQ column.

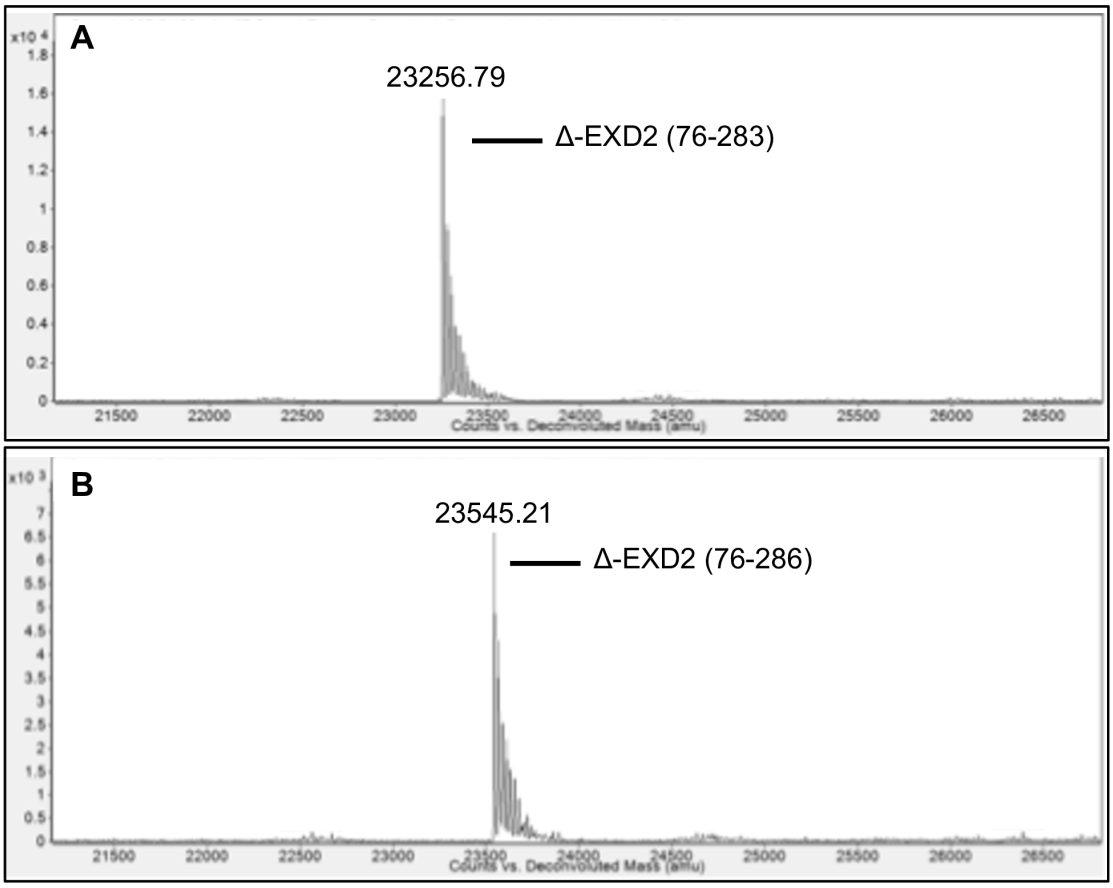


Figure 6.11 ESI-TOF-MS validation of final purified Δ-EXD2 (76–283) and Δ-EXD2 (76–286).

A: Mass spectrum analysis of Δ-EXD2 (76–283) – expected mass = 23257.0 Da, measured mass = 23256.79 Da. B: Mass spectrum analysis of Δ-EXD2 (76–286) – expected mass = 23545.4 Da, measured mass = 23545.21 Da.

Unfortunately, when assessed in biochemical nuclease assays, both Δ-EXD2 (K76–K283) and Δ-EXD2 (K76–G286) failed to show any significant digestion of a 50 nt ssDNA substrate in the presence of either Mn^{2+} or Mg^{2+} (concentrations between 0.1 and 10 mM; Figure 6.12). There was some observed 'smudging' of the DNA substrate at higher Mn^{2+} concentrations (Figure 6.12), however, that is likely to be due to interactions between Mn^{2+} ions and the phosphodiester backbone of the DNA substrate (Polyanichko et al., 2004). Nucleolytic activity of

(K76–K283) and (K76–G286) Δ-EXD2 was further examined on a variety of substrates; ds-, 3' overhang, 5' overhang, splayed arm, leading flap, lagging flap, and replication fork DNA structures. None of these showed any nuclease activity (data not shown).

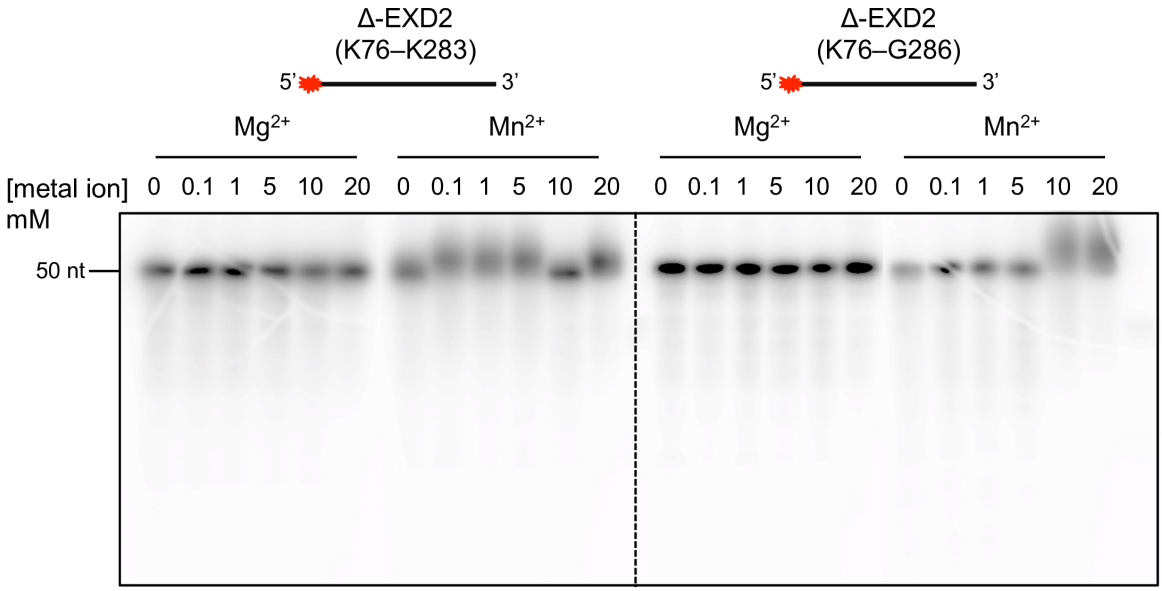


Figure 6.12 Neither Δ-EXD2 (K76–K283) nor Δ-EXD2 (K76–G286) show any nuclease activity on a ssDNA substrate in the presence of Mg^{2+} or Mn^{2+}.

10 nM of DNA substrate (as labelled) was added to 100 nM Δ-EXD2 and incubated at 37°C for 1 hour in the presence of increasing concentrations of Mg^{2+} or Mn^{2+}. The reaction was stopped with the addition of 2 µL stop solution and boiling. Samples were then analysed by 20% polyacrylamide gel electrophoresis. DNA substrate structures are as indicated.

As both Δ-EXD2 (K76–K283) and Δ-EXD2 (K76–G286) span both the predicted exonuclease domain, and the region of sequence similarity with WRN nuclease (mapped to 99 to 257 in the EXD2 protein sequence), this result was somewhat intriguing. The data therefore suggest that additional region or residues, outside that of the predicted exonuclease core, are required for effective enzymatic activity

of EXD2. Which region(s), or residues, are required and whether these are requisite for appropriate folding, DNA binding, or catalytic activity remains to be elucidated.

6.2.3. Purification of Δ-EXD2 (K76–S589)

As the only remaining overexpressed construct from the test expression protocol (Figure 6.5), Δ-EXD2 (K76–S589) was subsequently purified.

However, purification of Δ-EXD2 (K76–S589) proved to be challenging. Δ-EXD2 (K76–S589) was particularly prone to proteolytic cleavage during the purification process and this precluded efficient purification using the protocol outlined in Figure 6.6. As such, the methodology was modified, protease inhibitors were included in all steps of the purification, and the revised protocol is outlined below (and summarised in Figure 6.13).

6 L of BL21 *E. coli* culture expressing Δ-EXD2 (K76–S589) were harvested and the cell lysate was partially purified by an initial Ni-Sepharose column to selectively bind His_6-Δ-EXD2 (K76–S589) through specific interaction with N-terminal histidine tag. Eluate fractions were subsequently analysed by SDS-PAGE for the presence of His_6-Δ-EXD2 (K76–S589) as evidenced by the presence of an overexpressed band at ~57 kDa (Figure 6.14). The presence of His_6-Δ-EXD2 (K76–S589) was additionally confirmed by ESI-TOF-MS (Figure 6.14). The histidine tag was not cleaved by TEV, due to the aforementioned problems with proteolytic degradation. As such, the His_6-Δ-EXD2 (K76–S589) containing eluate

fractions from the Ni-Sepharose column were concentrated, pooled, and further purified by S200 gel filtration chromatography. Protein containing eluate fractions (as indicated by absorbance at 280 nm) were analysed by SDS-PAGE for the presence of His_6-Δ-EXD2 (K76–S589) (Figure 6.15).

The purest fractions were pooled, concentrated, and further purified by a subsequent S200 gel filtration chromatography column. Protein containing eluate fractions were again indicated by absorbance at 280 nm and the presence of His_6-Δ-EXD2 (K76–S589) was validated by SDS-PAGE (Figure 6.16). The purest fractions were pooled and concentrated for use in further biochemical analysis.

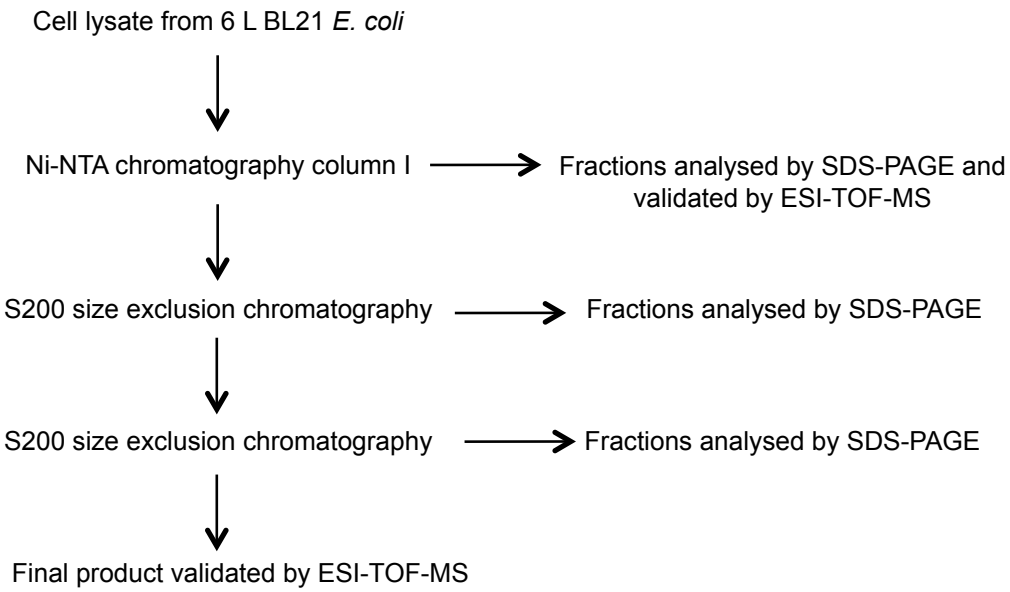


Figure 6.13 Schematic of purification protocol for Δ-EXD2 (K76–S589).

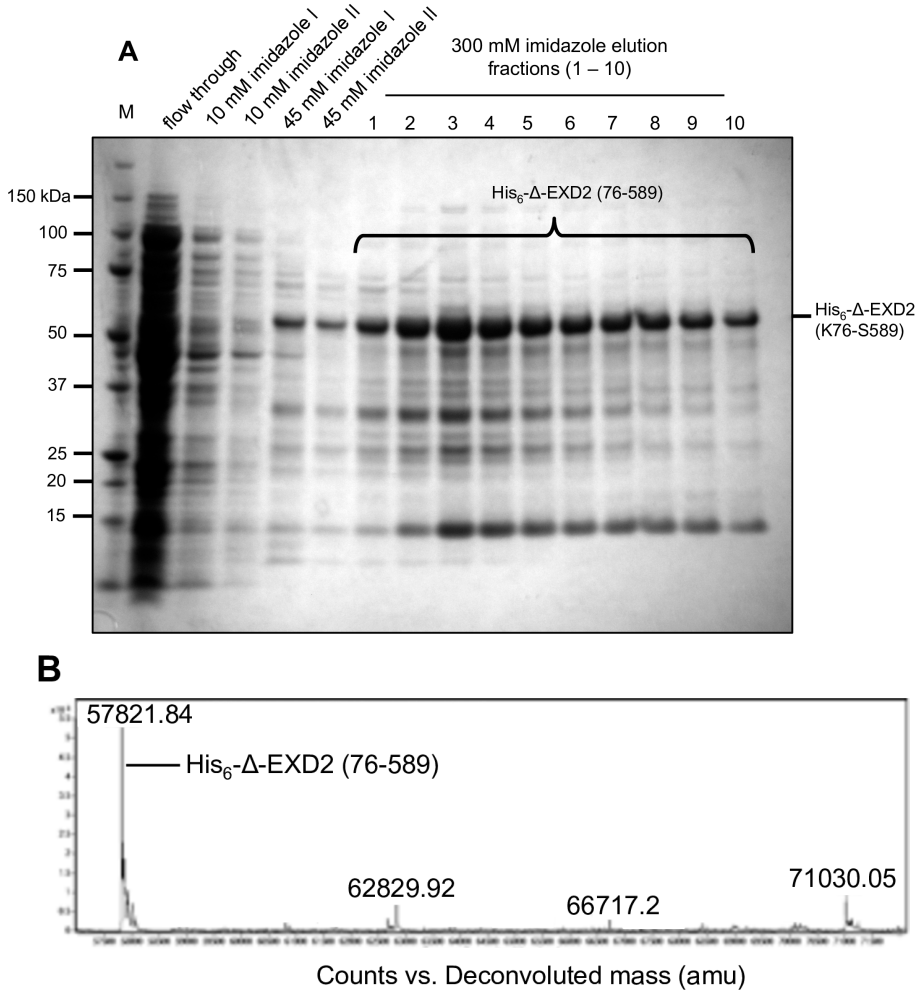


Figure 6.14 Ni-Sepharose column for purification of Δ-EXD2 (K76–S589).

A: Samples were collected at all stages of the initial Ni-Sepharose column and were analysed by SDS-PAGE for the presence of His_6-Δ-EXD2 (K76–S589). Cell lysates were added to the initial Ni-Sepharose column and the flow-through was collected. The column was then washed with 10 column volumes (CV) with 10 mM imidazole containing buffer, 10 CV of 45 mM imidazole containing buffer, and then eluted in 10 x 2 mL fractions with 300 mM imidazole containing buffer. The presence of His_6-Δ-EXD2 (K76–S589) was indicated by a band at ~57 kDa, and the bracket indicates fractions that were pooled and concentrated for further purification. B: ESI-TOF-MS validating the presence of Δ-EXD2 (K76–S589) after Ni-Sepharose chromatography. Expected mass = 60804 Da, measured mass = 57821.84 Da.

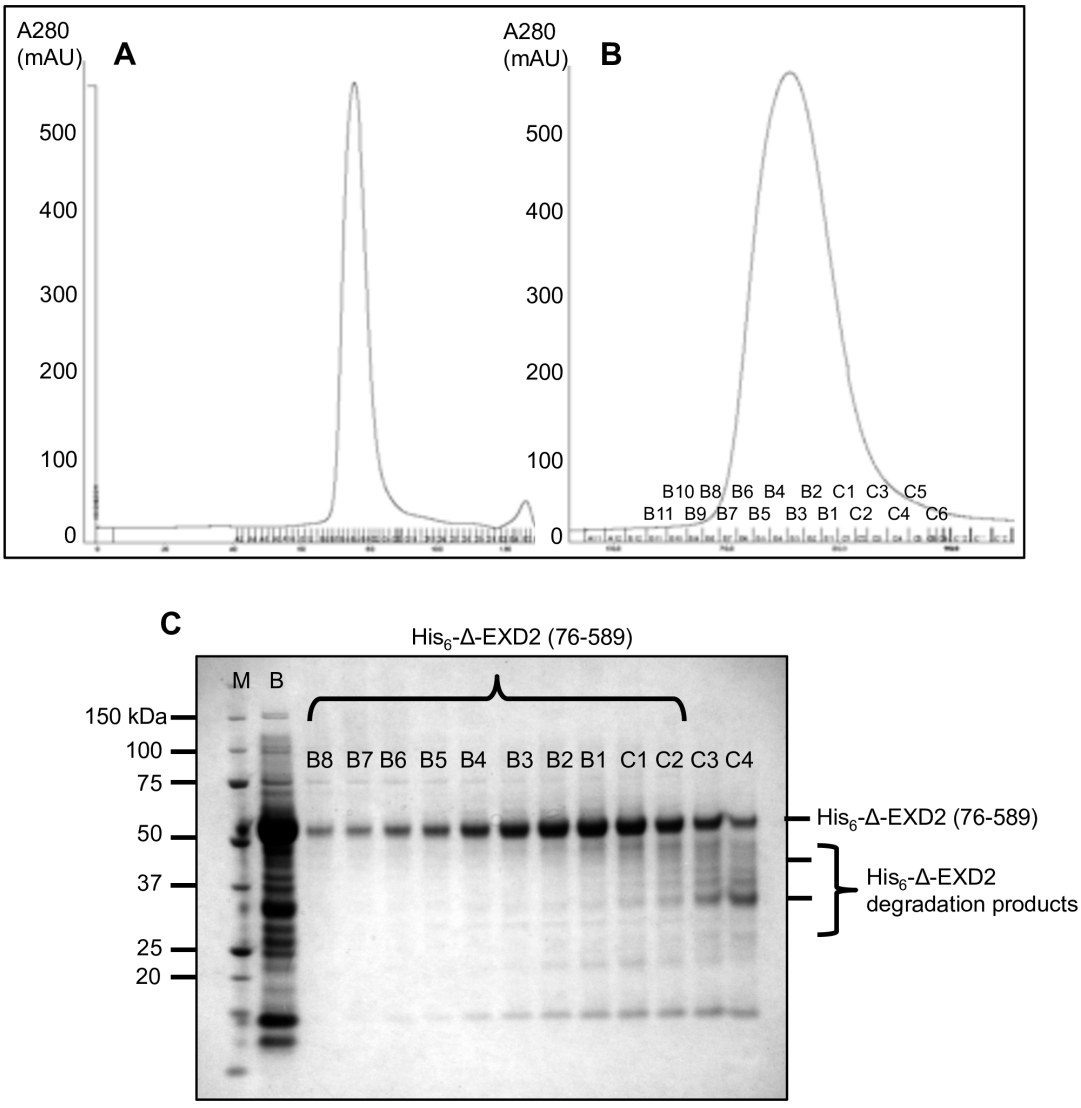


Figure 6.15 Superdex S200 gel filtration for Δ-EXD2 (K76–S589).

Panels A and B display absorbance at 280 nm across collected fractions for His_6-Δ-EXD2 (76–589) after S200 Superdex gel filtration chromatography. Panel A indicates all fractions, and panel B is enlarged to show the protein containing fractions. Panel C: SDS-PAGE of peak protein containing fractions, as indicated in B. The presence of a ~57 kDa band across the peak fractions indicates the presence His_6-Δ-EXD2 (76–589). The bracket indicates purified protein fractions that were pooled and concentrated for a second S200 gel filtration column. M denotes molecular weight marker, and B, the protein sample before it was loaded onto the S200 Superdex gel filtration column.

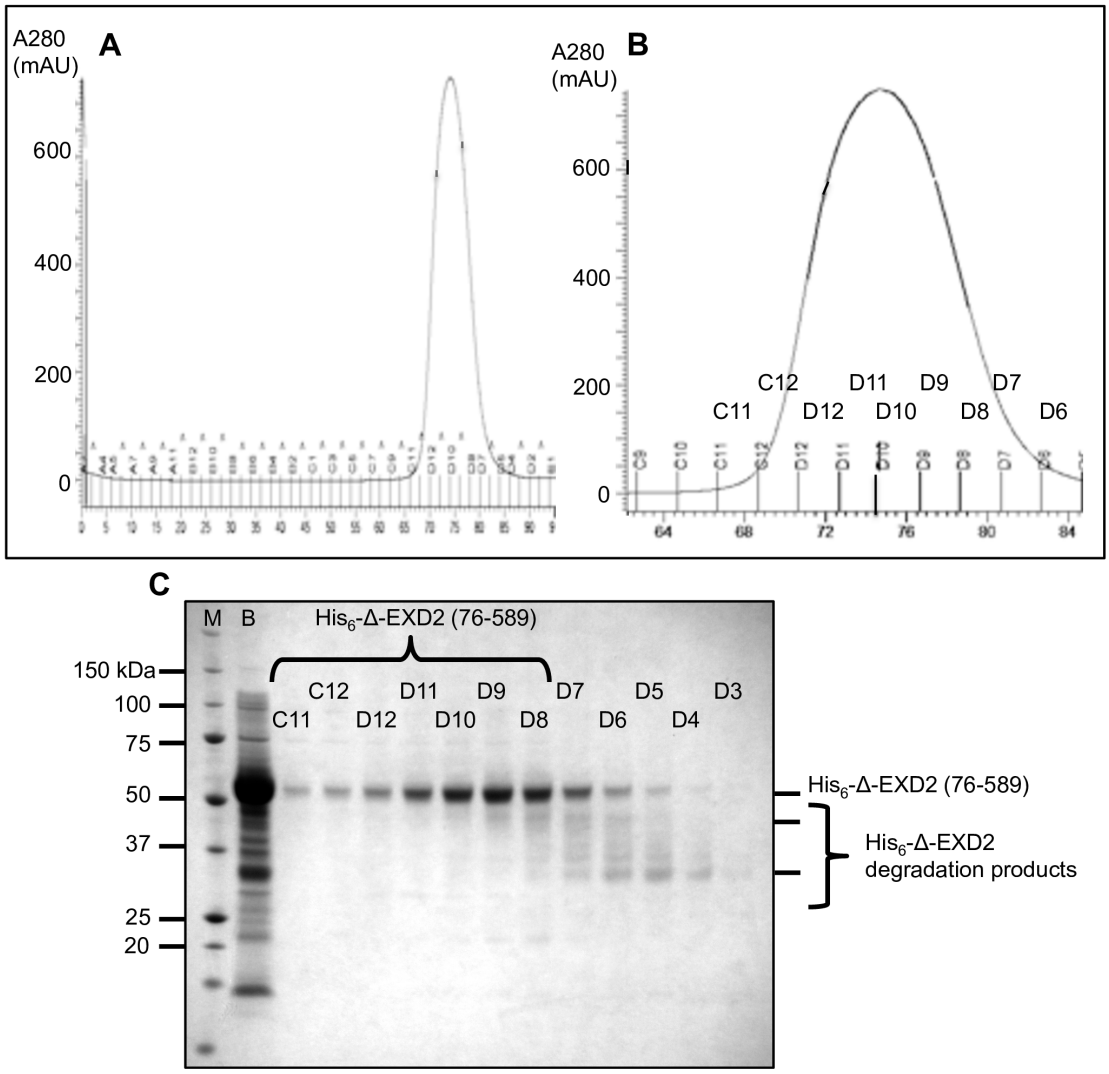


Figure 6.16 Second Superdex S200 gel filtration for Δ-EXD2 (K76–S589).

Panels A and B display absorbance at 280 nm across collected fractions for His_6-Δ-EXD2 (76–589) after S200 Superdex gel filtration chromatography. Panel A indicates all fractions, and panel B is enlarged to show the protein containing fractions. Panel C: SDS-PAGE of peak protein containing fractions, as indicated in B. The presence of a ~57 kDa band across the peak fractions indicates the presence His_6-Δ-EXD2 (76–589). The bracket indicates purified protein fractions that were pooled and concentrated for use in biochemical assays. M denotes molecular weight marker, and B, the protein sample before it was loaded onto the S200 Superdex gel filtration column.

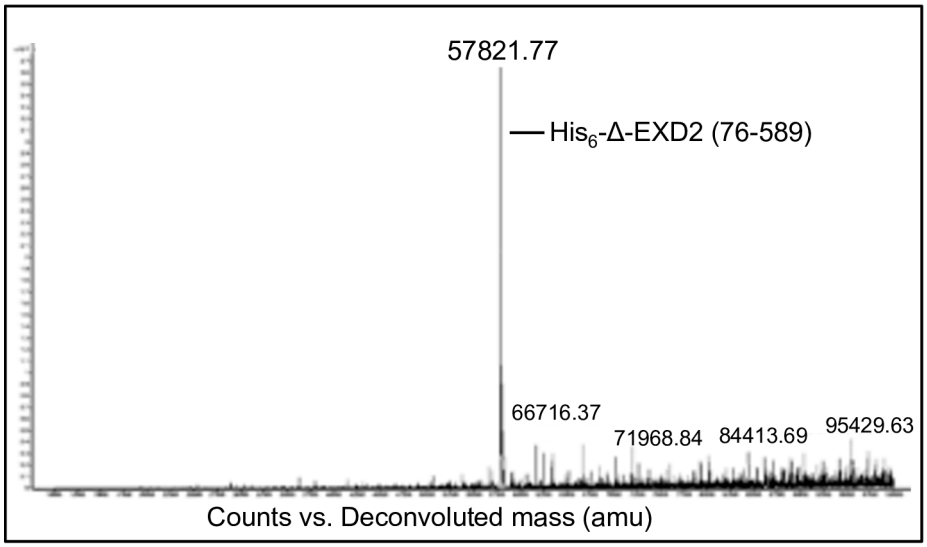


Figure 6.17 ESI-TOF-MS validation of final purified His_6-Δ-EXD2 (76–589).

Mass spectrum analysis of His_6-Δ-EXD2 (76–589). Expected mass = 60804 Da, measured mass = 57821.77 Da. Note that the signal to noise ratio is slightly low and this is likely due to the inclusion of 0.1–0.5% Triton-X100 in all purification buffers.

It is worthy of note that SDS-PAGE analysis after each of the S200 gel filtration chromatography columns display a series of protein products between ~35–50 kDa (Figure 6.15 and Figure 6.16). The main His_6-Δ-EXD2 (K76–S589) band, and the smaller bands (as indicated) were excised and analysed by tryptic digest and tandem MS (MS/MS), and this revealed that they were protein fragments of His_6-Δ-EXD2 (K76–S589). Therefore, the data suggest they are products of proteolytic degradation rather than a contaminating protein, and this is not entirely unexpected given the propensity of His_6-Δ-EXD2 (K76–S589) to undergo fragmentation during the purification process.

The purified product was confirmed to be His_6-Δ-EXD2 (K76–S589) by ESI-TOF-MS (Figure 6.17). However, there was a significant difference between the

expected mass (60804.0 Da) and the measured mass (57821.77 Da) and this was observed at all steps of the purification (Figure 6.14 (B) and Figure 6.17). Sequencing analysis of the cloned EXD2 gene revealed no mutations or undesired truncations, and the excised band from SDS-PAGE (Figure 6.15 and Figure 6.16) was confirmed as EXD2 protein by tryptic digest MS/MS. Subsequent *in silico* analysis using the PAWS software (Genomic Solutions) revealed a number of amino acid truncations that could correspond with the measured mass (57821.77 Da) (Figure 6.18). The predicted mass (57822.1 Da) that most accurately matched the measured mass was the result of a C-terminal truncation, and resulted in a Δ-EXD2 protein that spanned from K76–V564. In support of this, as the His_6-Δ-EXD2 was able to selectively bind the Ni-Sepharose resin as part of the purification protocol (Figure 6.14) it seemed most likely that the tag remained intact.

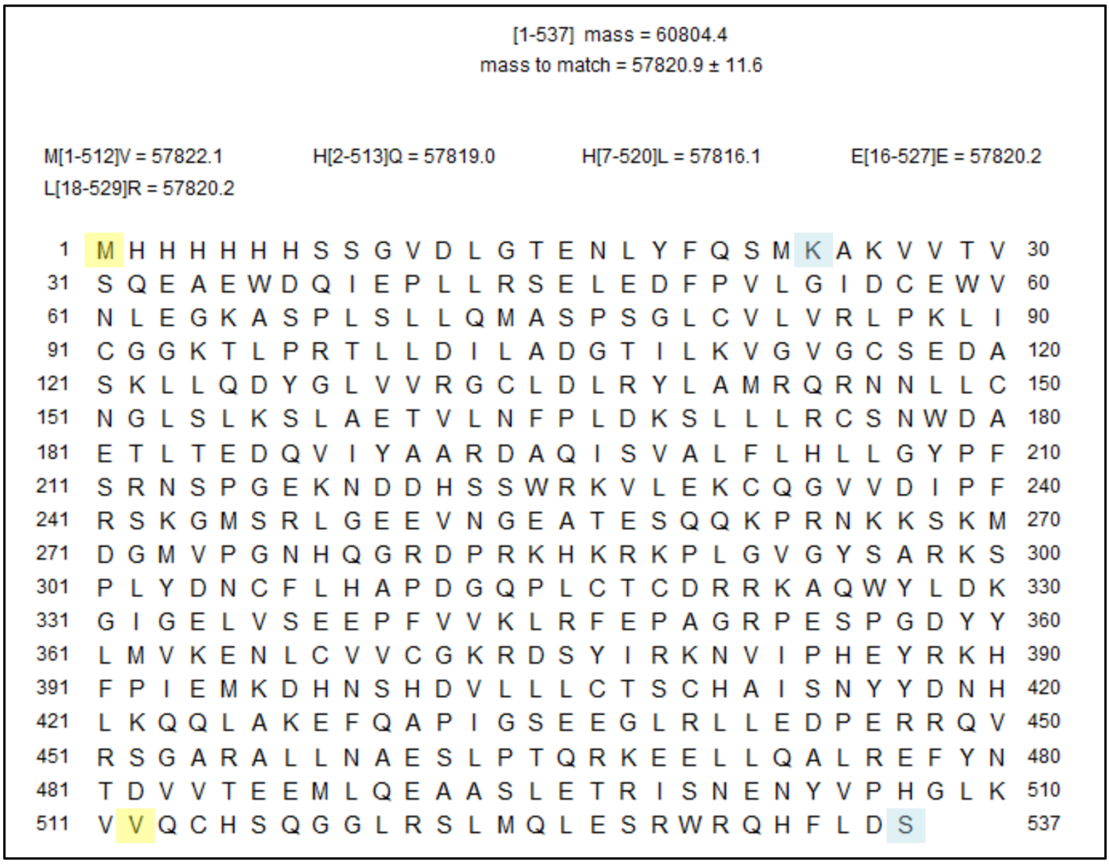


Figure 6.18 Analysis of the Δ-EXD2 amino acid sequence using PAWS software.

The expected sequence of His_6-Δ-EXD2 (76–589) was input into the PAWS software. The start M is highlighted in yellow and is followed by the 6-His tag, and the TEV recognition sequence. K76 is highlighted in blue and corresponds to the first amino acid in the EXD2 sequence. The blue highlighted 'S' indicates S589, the residue at which the Δ-EXD2 construct was designed to terminate. The expected mass from the start M to S589 was 60804.4 Da, and the input search mass was 57820.9 Da. M1–V512 was considered to be the most likely truncation, and V512 is highlighted in yellow. This truncation results in a protein product: His_6-Δ-EXD2 (K76–V564).

An initial biochemical validation of His_6-Δ-EXD2 (K76–V564) (henceforth denoted WT EXD2), revealed significant nuclease activity in the presence of 1.0–10 mM Mn^{2+}, and no activity with Mg^{2+} (Figure 6.19).

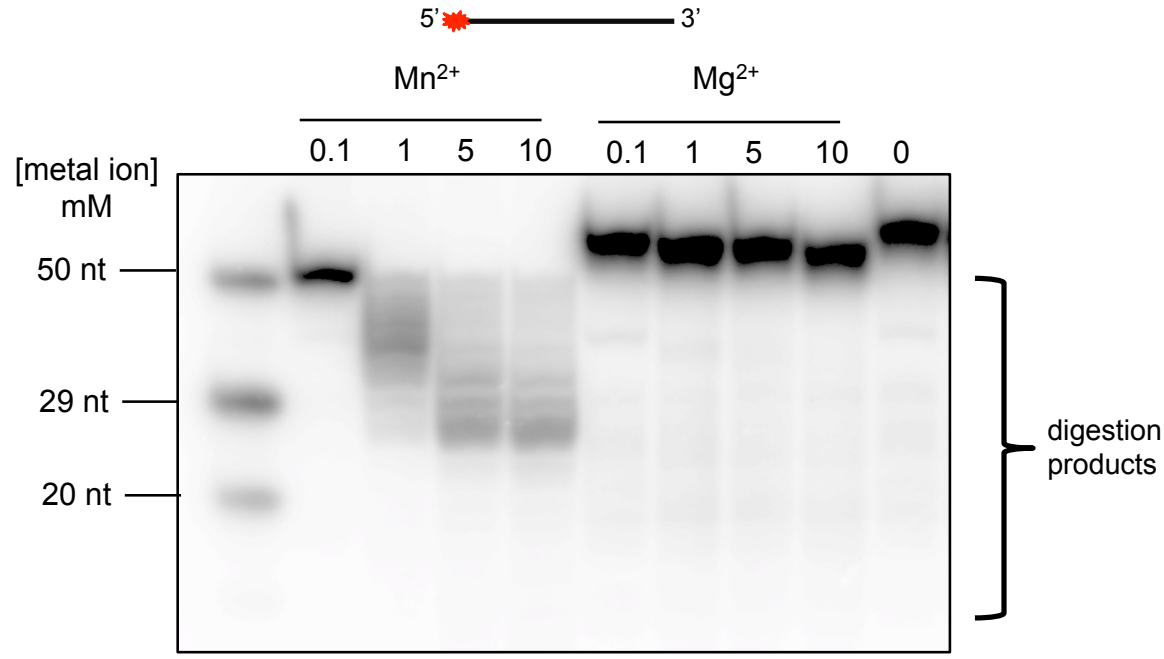


Figure 6.19 WT Δ-EXD2 (K76–V564) exhibits robust nuclease activity on a ssDNA substrate in the presence of Mn^{2+}.

10 nM of ssDNA substrate was added to 100 nM EXD2 and incubated at 37°C for 1 hour in the presence of increasing concentrations of Mn^{2+} or Mg^{2+}. The reaction was stopped with the addition of 2 µL stop solution and boiling. Samples were analysed by 20% polyacrylamide gel electrophoresis. Gel is representative of at three individual experiments.

6.2.4. Purification of D108A/E110A Δ-EXD2 (K76–S589)

To ensure that WT EXD2 is a *bona fide* nuclease it was important to confirm that the observed nucleolytic activity (Figure 6.3 and Figure 6.19) was intrinsic to the EXD2 enzyme and not due to a contaminating nuclease. As such, a 'nuclease dead' EXD2 mutant was designed, subcloned, and purified in the same manner as WT EXD2, as it could serve as an important negative control in subsequent nuclease assays.

Analysis of the sequence alignment between EXD2 and WRN revealed a number of equivalent amino acids (Figure 6.1). Two of these amino acids in the WRN protein (D82 and E84) are highly conserved within the family of DnaQ type exonucleases (Figure 6.20) and coordinate one of the metal ions in the active site (Choi et al., 2007). Mutation of these residues in WRN, and subsequent disruption of metal ion-binding, abrogates nuclease activity *in vitro* (Li et al., 2009, Xue et al., 2002). As such, the analogous residues in the EXD2 sequence (D108 and E110) were both mutated to alanine using a site-directed mutagenesis approach (Broderick et al., 2016).



Figure 6.20 Alignment of the partial amino acid sequences from human and vertebrate analogues of EXD2, and human WRN.

The aligned sequences are from the conserved 3' to 5' exonucleolytic core from each xenopus, chicken, mouse, and human EXD2 and human WRN. Entirely conserved amino acid residues are denoted with an asterisk (*). A colon (:) indicates conservation between residues with strongly similar properties, and a single full stop (.) indicates conservation between residues with weakly similar properties. The conserved metal ion coordinating residues (numbered D108 and E110 in the human EXD2 sequence) are highlighted in yellow. This figure is adapted from Broderick *et al*. and alignment was performed using the ClustalW tool from EMBL-EBI.

The D108A/E110A EXD2 gene construct was subsequently subcloned into the pNIC28-Bsa4 vector and the resulting protein (His_6-Δ-EXD2 D108A/E110A (K76–S589) was purified in exactly the same manner as WT EXD2 (Figure 6.13). Importantly, the WT EXD2 was re-purified at the same time to ensure reproducibility of function in subsequent biochemical assays.

Briefly, His_6-Δ-EXD2 D108A/E110A (K76–S589) was purified as follows. An initial Ni-Sepharose column was performed to selectively bind His_6-Δ-EXD2 D108A/E110A (K76–S589) through interaction with the histidine tag. His_6-Δ-EXD2 D108A/E110A (K76–S589) was eluted from the column with 300 mM imidazole

containing buffer and the presence of His_6-Δ-EXD2 D108A/E110A (K76–S589) in each of the eluate fractions was examined by SDS-PAGE (Figure 6.21). His_6-Δ-EXD2 D108A/E110A (K76–S589) containing eluate fractions were pooled, concentrated, and further purified by S200 gel filtration chromatography. Protein containing fractions (as indicated by absorbance at 280 nm) were analysed by SDS-PAGE for the presence of His_6-Δ-EXD2 D108A/E110A (K76–S589), as evidenced by a ~57 kDa molecular weight band (Figure 6.22). The purest fractions (as indicated) were pooled, concentrated, and further purified by an additional S200 size exclusion chromatography column (Figure 6.22). Protein containing eluate fractions from the second S200 chromatography column were indicated by absorbance at 280 nm and the presence of His_6-Δ-EXD2 D108A/E110A (K76–S589) was validated by SDS-PAGE (Figure 6.23). The purest fractions were pooled and concentrated for use in downstream biochemical assays.

Similarly to WT EXD2, SDS-PAGE during the purification of His_6-Δ-EXD2 D108A/E110A (K76–S589), revealed a number of protein bands between ~30 and 50 kDa (Figure 6.22 and Figure 6.23). These gel bands were excised and analysed by tryptic digest MS/MS and confirmed to be degradation products of His_6-Δ-EXD2 D108A/E110A (K76–S589) rather than a contaminating protein.

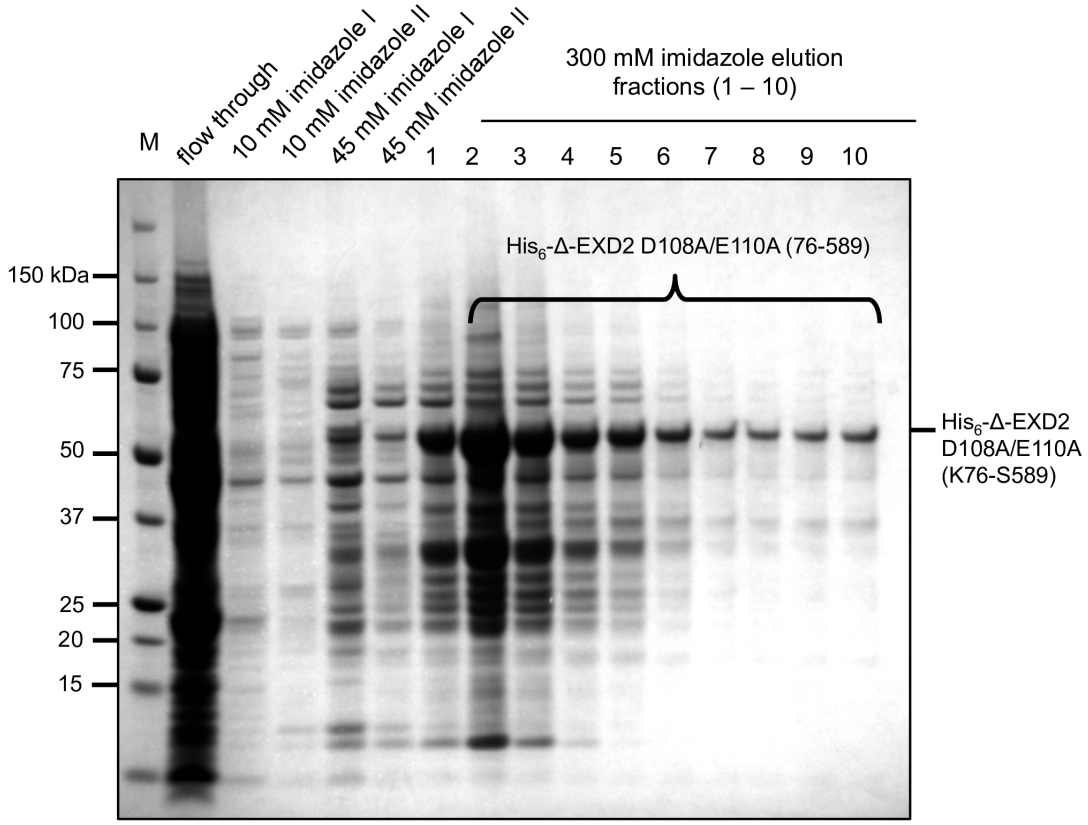


Figure 6.21 SDS-PAGE analysis of the Ni-Sepharose column for the purification of Δ-EXD2 D108A/E110A (K76–S589).

A: Samples were collected at all stages of the initial Ni-Sepharose column and were analysed by SDS-PAGE for the presence of the His_6-Δ-EXD2 D108A/E110A (K76–S589). Cell lysates were added to the initial Ni-Sepharose column and the flow-through was collected. The column was then washed with 10 column volumes (CV) with 10 mM imidazole containing buffer, 10 CV of 45 mM imidazole containing buffer, and then was eluted in 10 x 2 mL fractions with 300 mM imidazole containing buffer. The presence of His_6-Δ-EXD2 D108A/E110A (K76–S589) was indicated by a band at ~57 kDa. A Precision Plus Unstained Protein Marker was used and approximate molecular weights are indicated.

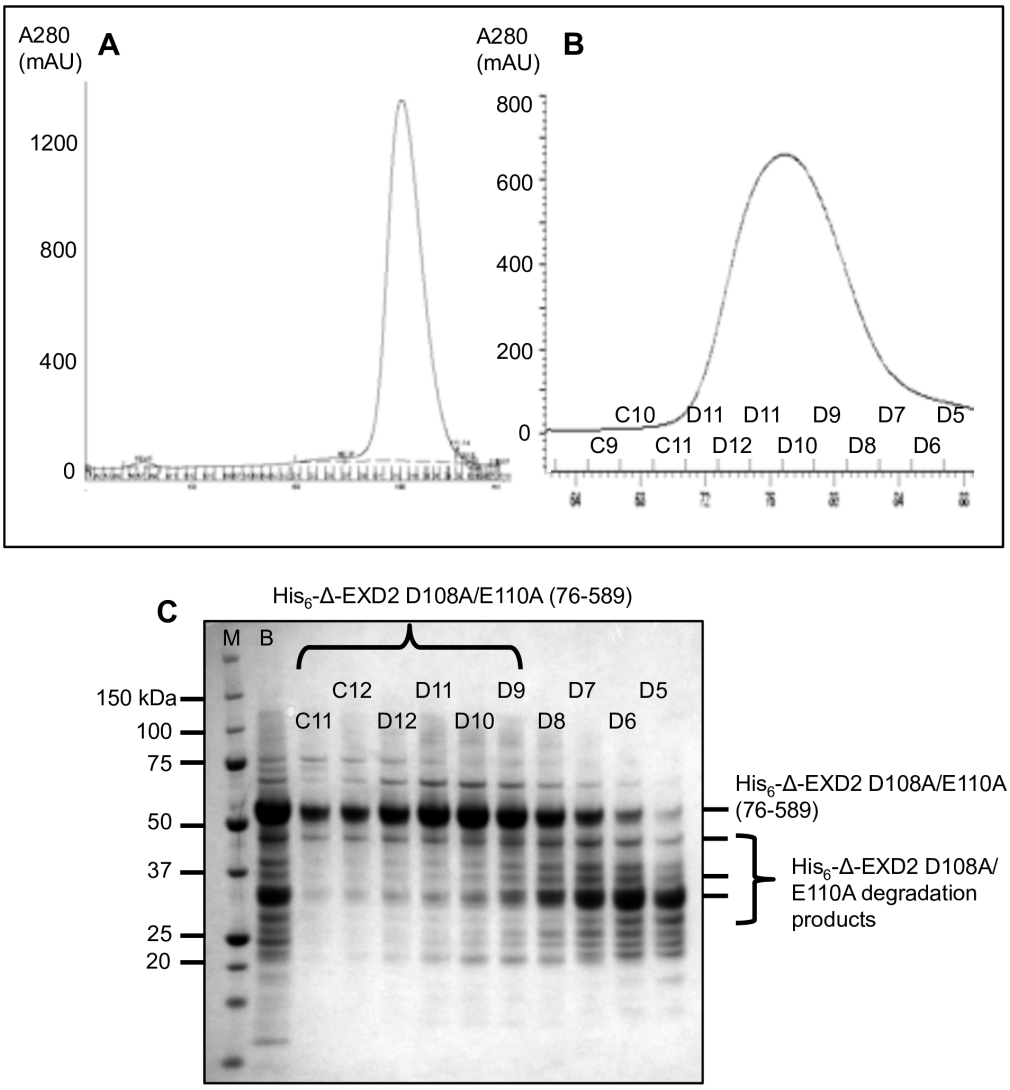


Figure 6.22 Superdex S200 gel filtration for Δ-EXD2 D108A/E110A (K76–S589).

Panels A and B display absorbance at 280 nm across collected fractions for His_6-Δ-EXD2 (76–589) after S200 Superdex gel filtration chromatography. Panel A indicates all fractions, and panel B is enlarged to show the protein containing fractions. Panel C: SDS-PAGE of peak protein containing fractions, as indicated in B. The presence of a ~57 kDa band across the peak fractions indicates the presence His_6-Δ-EXD2 D108A/E110A (76–589). The bracket indicates purified protein fractions that were pooled and concentrated for a second S200 gel filtration column. M denotes molecular weight marker, and B, the protein sample before it was loaded onto the S200 Superdex gel filtration column.

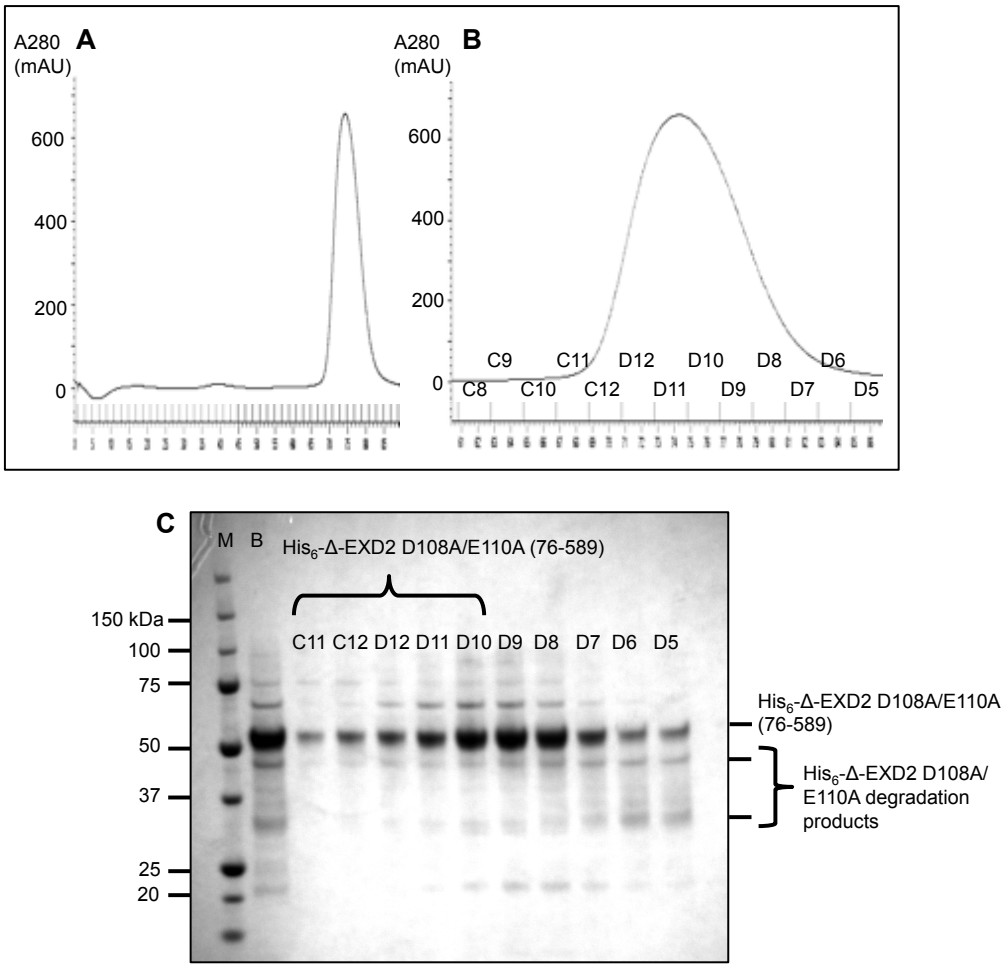


Figure 6.23 Second Superdex S200 gel filtration for Δ-EXD2 D108A/E110A (K76–S589).

Panels A and B display absorbance at 280 nm across collected fractions for His_6-Δ-EXD2 D108A/E110A (76–589) after S200 Superdex gel filtration chromatography. Panel A indicates all fractions, and panel B is enlarged to show the protein containing fractions. Panel C: SDS-PAGE of peak protein containing fractions, as indicated in B. The presence of a ~57 kDa band across the peak fractions indicates the presence His_6-Δ-EXD2 D108A/E110A (76–589). The bracket indicates purified protein fractions that were pooled and concentrated for use in biochemical assays. M denotes molecular weight marker, and B, the protein sample before it was loaded onto the S200 Superdex gel filtration column.

The final purification product, His_6-Δ-EXD2 D108A/E110A (K76–S589) (henceforth denoted D108A/E110A EXD2), was validated by ESI-TOF-MS (Figure 6.24). As was observed with WT EXD2 (Figure 6.14 and Figure 6.17), D108A/E110A EXD2 displayed a significant differential from the expected mass (60598.4 Da) to the measured mass (57716.1 Da) (Figure 6.24). Again, this corresponded to the loss of amino acids 565–589 at the C-terminus, resulting in a protein that consisted of amino acids K76–V564.

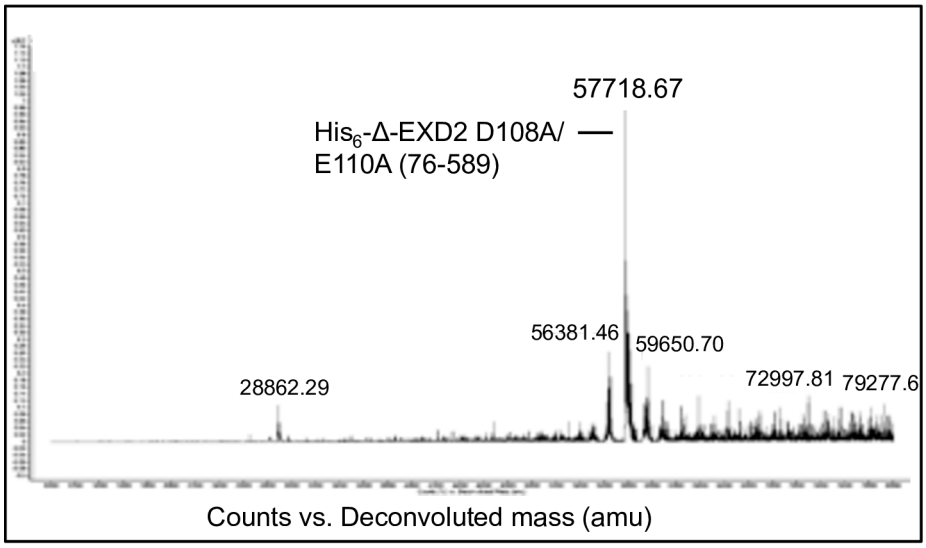


Figure 6.24 ESI-TOF-MS validation of final purified His_6-Δ-EXD2 D108A/E110A (76–589).

Mass spectrum analysis of His_6-Δ-EXD2 D108A/E110A (76–589). Expected mass = 60598.4 Da, measured mass = 57716.1 Da. Note that the signal to noise ratio is slightly low and this is likely due to the inclusion of 0.1–0.5% Triton-X100 in all purification buffers.

6.2.5. Analysis of *in vitro* nuclease activity and DNA binding of EXD2

At the initiation of this study, EXD2 was a novel exonuclease and only a very preliminary validation of nuclease activity had been performed (Figure 6.3). As discussed above, a method was developed for purifying recombinant, truncated WT and D108A/E110A EXD2 from *E. coli* with good yield and purity. The enzymatic activity of these proteins was subsequently examined in a series of nuclease assays and DNA binding experiments to determine the biochemical characteristics of EXD2.

As EXD2 is an exonuclease with 3' to 5' polarity, DNA substrates were radiolabelled on the 5' end with γ-^{32}P-dATP. A ssDNA substrate was initially labelled and subsequently annealed to specific oligonucleotide sequences to generate the structures that are schematically represented at the top of each figure. All nucleotide sequences used in this study can be found in Appendix V.

It is important to note that the exonucleolytic directionality of EXD2 was validated experimentally. Thin-layer chromatography experiments, performed by Dr Jadwiga Nieminuszczy, showed release of a single terminal 3' labelled nucleotide from an α-^{32}P-dATP labelled substrate. This clearly showed that a single nucleotide was released from each catalytic event, and that EXD2 functioned only in the 3' to 5' direction (Broderick et al., 2016).

To determine the specific reaction conditions for catalysis, the nuclease activity of EXD2 was examined with a range of different metal ions at increasing

concentrations (0.01–10 mM) (Figure 6.24). In accordance with full-length EXD2 (Figure 6.3), truncated WT EXD2 exhibited nuclease activity on ssDNA in the presence of 0.1–10 mM Mn^{2+}, with maximal activity at 5 mM (Figure 6.25). Moreover, there was no observed nuclease activity in the presence of 0.1–10 mM Mg^{2}, Zn^{2+}, or Ca^{2+} (Figure 6.25). As such, 5 mM Mn^{2+} was used in all subsequent nuclease assay reactions and further analysis of the active site metal ion for EXD2 will be discussed in 6.2.6.

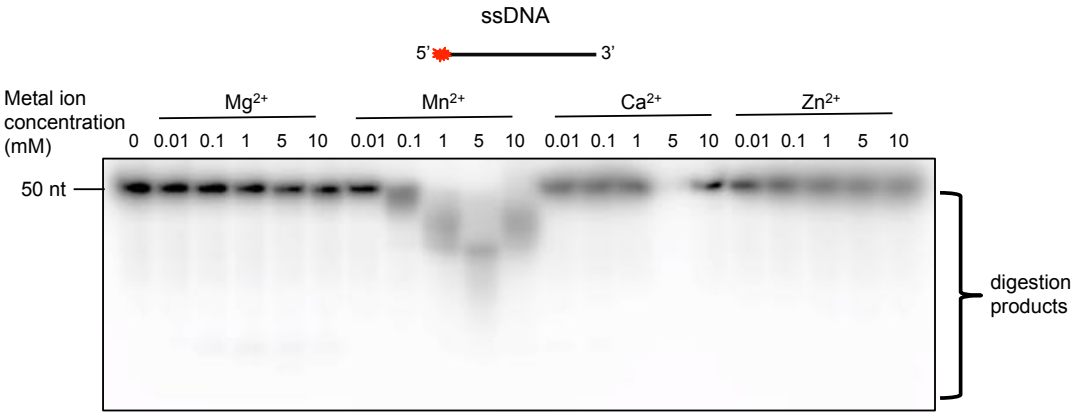


Figure 6.25 WT EXD2 exhibits exonucleolytic activity in the presence of 0.1–10 mM Mn^{2+}.

10 nM of 50 nucleotide ssDNA substrate was added to 100 nM WT EXD2 and incubated at 37°C for 1 hour in the presence of increasing concentrations of Mg^{2+}, Mn^{2+}, Ca^{2+}, or Zn^{2+}. The reaction was stopped with the addition of 2 µL stop solution and boiling. Samples were then analysed by 20% polyacrylamide gel electrophoresis. DNA substrate structures are as indicated. Gel is representative of at least thee individual experiments.

To further investigate the optimal catalytic conditions for further analysis, nuclease activity over a pH range was examined. Like many human enzymes, EXD2 exhibited nuclease activity over a physiological pH range, with significant digestion

occurring from pH 6.0–8.0, and optimal nucleolytic activity was supported at pH 7.0 (Figure 6.26).

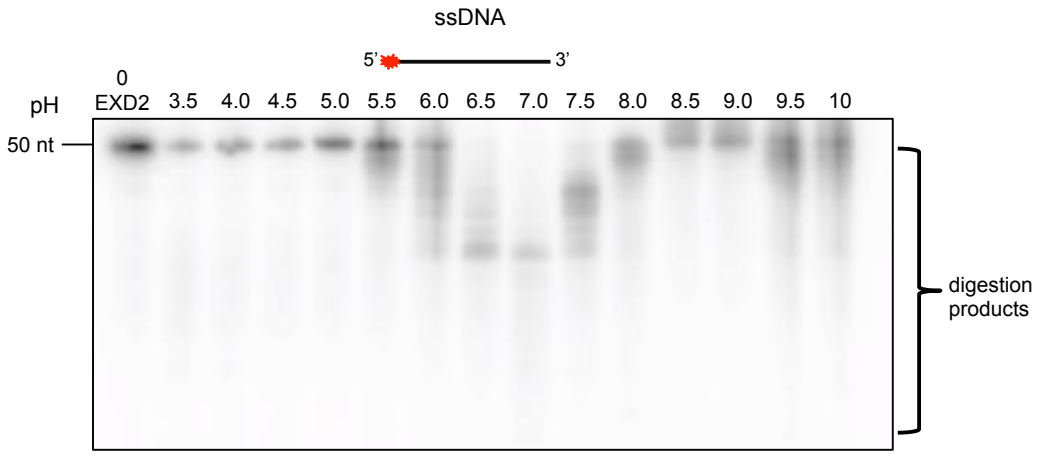


Figure 6.26 WT EXD2 exhibits exonucleolytic activity at a pH range of 6.0–8.0.

10 nM of 50 nucleotide ssDNA substrate was added to 100 nM WT EXD2 and incubated at 37°C for 1 hour in HEPES buffer over a pH range from 3.5–10 (as indicated). The reaction was stopped with the addition of 2 µL stop solution and boiling. Samples were then analysed by 20% polyacrylamide gel electrophoresis. 0 EXD2 simply indicates that no enzyme was added to this reaction. Gel is representative of at least thee individual experiments.

The nuclease activity of WT and D108A/E110A EXD2 was then examined on both a ss- and dsDNA substrate (Figure 6.27). WT EXD2 showed significant digestion of a ssDNA substrate, but was unable to digest a blunt-ended dsDNA substrate. D108A/E110A EXD2 exhibited negligible digestion on either substrate (Figure 6.27). The observed phenomenon that D108A/E110A does not possess nuclease activity is important for two reasons. Firstly, it provides some confirmation that the exonucleolytic activity of WT EXD2 is intrinsic to the enzyme, and not due to a contaminating nuclease, as both WT and D108A/E110A EXD2 were purified in the same manner, at the same time. Secondly, it supports the notion that these

residues – D108 and E110 – are important for co-ordinating one (or more) of the metal ions in the active site. These two residues constitute the former part of the conserved DEDD motif, characteristic to the family of DnaQ-like exonucleases (Yang, 2011). This suggests that the active site of EXD2 is likely to be structurally similar to other family members, including WRN, and thus probably relies on very similar catalytic mechanisms.

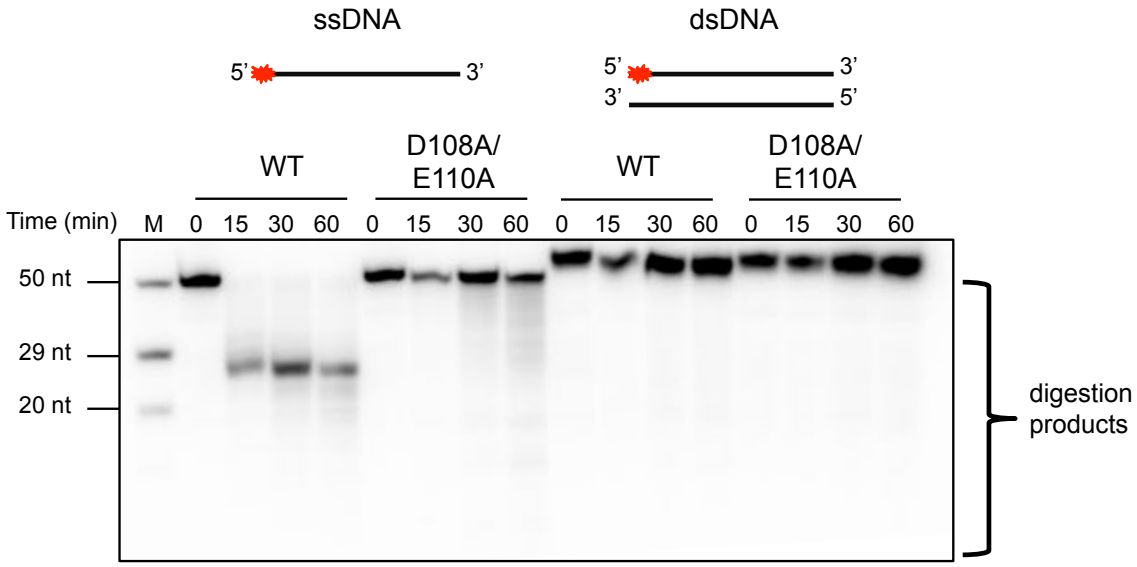


Figure 6.27 WT EXD2 is able to efficiently digest ss- but not dsDNA.

10 nM of DNA substrate (as labelled) was added to 100 nM WT or D108A/E110A EXD2 and incubated at 37°C for increasing time. The reaction was stopped with the addition of 2 µL stop solution and boiling. Samples were analysed by 20% polyacrylamide gel electrophoresis. DNA substrate structures are as indicated. Molecular weight markers are as labelled. Each gel is representative of at least three individual experiments.

The DNA binding capability of WT EXD2 to a ssDNA substrate, using a fluorescence anisotropy based method (Heyduk et al., 1996) was also analysed (Figure 6.28). WT EXD2 binds ssDNA with relatively high affinity, and the K_d of binding was 11 nM ± 4 nM. Interestingly, the K_d of binding for EXD2 (11 nM) is similar to that observed for the binding of WRN to a variety of different DNA structures (Kamath-Loeb et al., 2012), perhaps suggesting a similar mode of DNA interaction and binding.

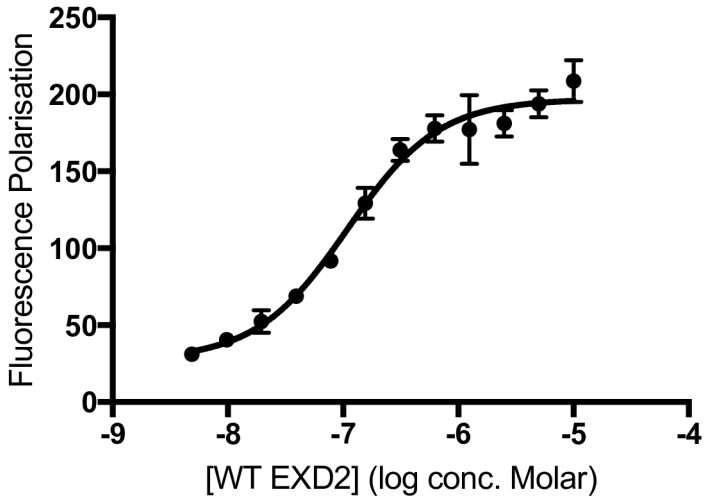


Figure 6.28 Fluorescence anisotropy DNA binding assay for WT EXD2.

10 nM single stranded 21 nt fluorescent DNA substrate was incubated with serially diluted concentrations of WT EXD2 (as indicated) for five minutes at room temperature. Each reaction was in quadruplicate and at least two experimental repeats were performed. Error bars indicate the SD. WT EXD2 K_d = 11 nM ± 4 nM.

6.2.6. Active Site Metal Ion Analysis for EXD2

Whilst examining the nuclease activity of WT EXD2 in the presence of different metal ions (Figure 6.25) was informative both for designing optimal reaction conditions, and understanding the necessary co-factors for catalysis, it didn't fully elucidate what may be the physiological active site metal ion(s).

To further understand this, non-denaturing MS was performed. Unfortunately, non-denaturing MS of His_6-Δ-EXD2 (K76–V564) was unsuccessful. However, native MS of the shorter Δ-EXD2 (K76–K283) revealed a mixed population with either no, one, or two zinc ion(s) coordinated in the active site (Figure 6.29). The presence of Δ-EXD2 (K76–K283) with no bound metal ion is indicated by the species at 23257.01 Da; the second, dominant species at 23320.40 Da is indicative of Δ-EXD2 (K76–K283) with one bound Zn^{2+} ion; and the third species, at 23382.52 is indicative of Δ-EXD2 (K76–K283) with two bound Zn^{2+} ions (Figure 6.29).

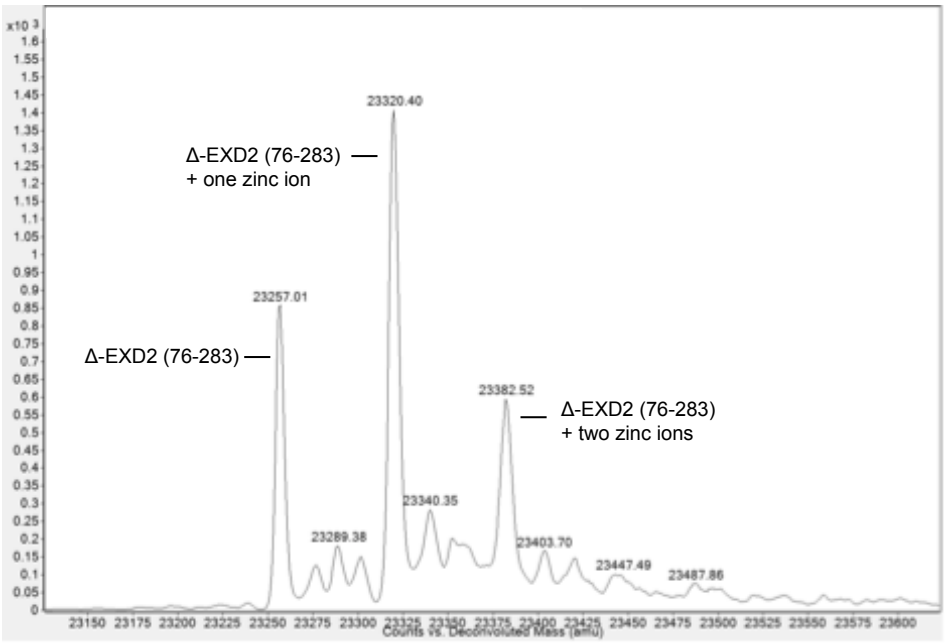


Figure 6.29 Native MS of Δ-EXD2 (K76–K283) showing either no, one, or two zinc ions bound to the protein.

Spectra from native MS of Δ-EXD2 (K76–K283), counts vs deconvoluted mass (amu). The three major species: Δ-EXD2 (K76–K283) with no metal bound; Δ-EXD2 (K76–K283) with one Zn^{2+} ion bound; and Δ-EXD2 (K76–K283) with two Zn^{2+} ions bound, are as indicated. Δ-EXD2 (K76–K283) expected mass = 23257 Da, measured mass = 23257.01 Da. The monoisotopic mass of Zn^{2+} = 63.92 Da; and the difference in mass between Δ-EXD2 (K76–K283) and Δ-EXD2 (K76–K283) with one Zn^{2+} measured 63.39 Da; the difference in mass between Δ-EXD2 (K76–K283) with one Zn^{2} and Δ-EXD2 (K76–K283) with two Zn^{2+} measured 62.12 Da. As such, these peaks reflect the addition of Zn^{2+} ions (either one or two respectively). It is worthy of note that the smaller peaks simply represent Na^{+} adducts. Native MS was performed in collaboration with Dr Rod Chalk, SGC.

This result is important for two reasons. Firstly, it suggests that Δ-EXD2 (K76–K283) may be correctly folded, as it is able to coordinate the metal ion(s) in the active site. This suggests that the absence of nuclease activity observed for these

two purified proteins Δ-EXD2 (K76–K283) and Δ-EXD2 (K76–G286) is due to the lack of specific residues or regions required for DNA interaction or catalysis, rather than incorrect protein folding or conformation. Secondly, this provides evidence that there may be at least one Zn^{2+} ion coordinated in the active site of WT EXD2. However, there is also the caveat that these EXD2 proteins are purified using an overexpression system from *E. coli*, thus this may not entirely represent *in vivo* physiological conditions.

To further explore the identity of the active site metal ion(s), nuclease reactions were carried out in the presence of a variety of metal ion chelating agents. The effects of three compounds were analysed: EDTA, a general metal ion chelator; EGTA, which possesses a higher affinity for calcium, and a lower affinity for magnesium ions (Qin et al., 1999); and o-phenanthroline, which is relatively selective for zinc ions (Williams and Loeb, 1973, Slater et al., 1971).

As can be seen in Figure 6.30, the nuclease activity of WT EXD2 is preferentially inhibited by the addition of o-phenanthroline. At 0.1 mM o-phenanthroline (equimolar to the concentration of WT EXD2), there is strong inhibition of nuclease activity when compared to the same concentration of EDTA and EGTA. Similarly, at 1.0 mM concentration of o-phenanthroline there is almost complete inhibition of nuclease activity, whereas there is little observed effect for both EDTA and EGTA (Figure 6.30). At concentrations of 5 mM and above, the addition of each of the three metal ion chelators completely inhibits nuclease activity, however, at this point, they may also be chelating the exogenous metal ion (5 mM Mn^{2+}) in the reaction buffer. There is also some aggregation that occurs at the higher

concentrations of o-phenanthroline (10 and 25 mM) and this is evidenced by the aggregated products becoming unable to enter the gel from the loading wells (Figure 6.30)

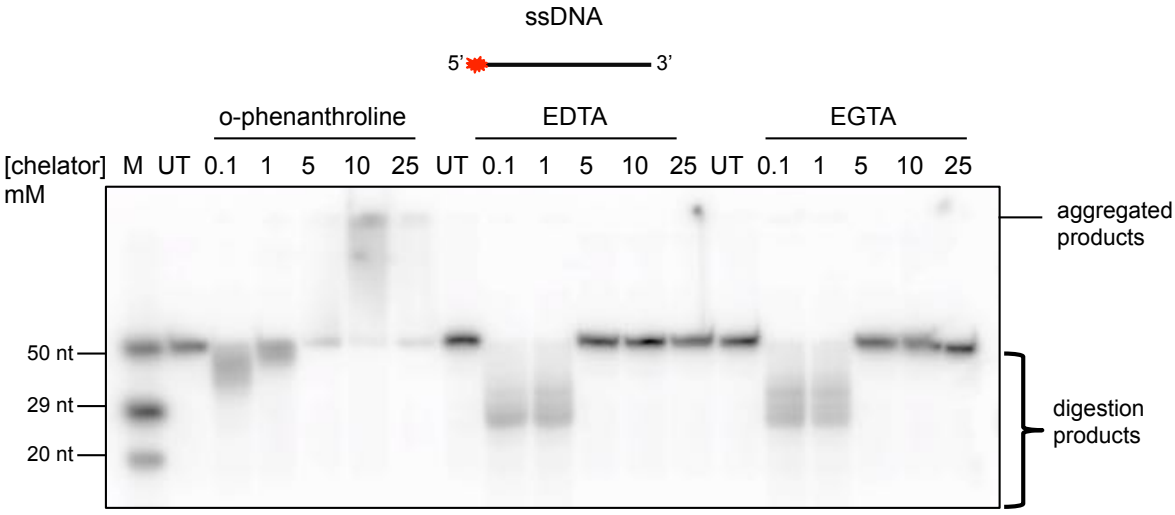


Figure 6.30 The exonuclease activity of WT EXD2 is inhibited with the addition of metal ion chelating agents, with a particular sensitivity to o-phenanthroline.

10 nM of ssDNA was added to 100 mM WT EXD2 and incubated at 37°C in the presence of increasing concentrations (0.1–25 mM) of metal ion chelating compounds; o-phenanthroline, EDTA, and EGTA, as indicated. The reaction was stopped after 60 minutes with the addition of 2 µL stop solution and boiling. Samples were then analysed by 20% polyacrylamide gel electrophoresis. DNA substrate structures are as indicated. Molecular weight markers (M) are as labelled. UT denotes untreated control. Each gel is representative of at least three individual experiments.

Amino acid sequence analysis categorises EXD2 into the DnaQ-like exonuclease super-family (Figure 6.1) (Broderick et al., 2016). Characteristic to this super-family is the DEDD motif, aptly named for the four acidic residues that coordinate the two metal ions required for catalysis. All members of this superfamily can thus be categorised as two-metal-ion-dependent nucleases (Yang, 2011). However, delineating which of the divalent metal cations occupy these two sites *in vivo* can be somewhat challenging, particularly in the absence of a crystal structure.

Nonetheless, the native MS data (Figure 6.29) and the sensitivity to o-phenanthroline (Figure 6.30) suggest that there is at least one zinc ion in the active site of WT EXD2. Furthermore, as WT EXD2 only exhibits nuclease activity in the presence of manganese ions (Figure 6.3 and Figure 6.25) it is entirely possible that Mn^{2+} might occupy the second metal ion-binding site, and that having these two sites occupied is requisite for catalysis.

6.2.7. Investigation of the nuclease activity of EXD2 on structurally diverse DNA substrates

Thus far, EXD2 exhibited nuclease activity on ss- and splayed arm DNA substrates, but showed no activity on blunt-ended dsDNA (Figure 6.3 and Figure 6.26). To further determine if WT EXD2 had any substrate specificity, nuclease activity on a number of structurally diverse DNA structures was examined. WT EXD2 displayed nuclease activity on all substrates, except blunt-ended dsDNA, however there is a clear preference for substrates with a free 3' end – namely, ss-, 3' overhang, splayed arm, and lagging flap DNA substrates (Figure 6.31).

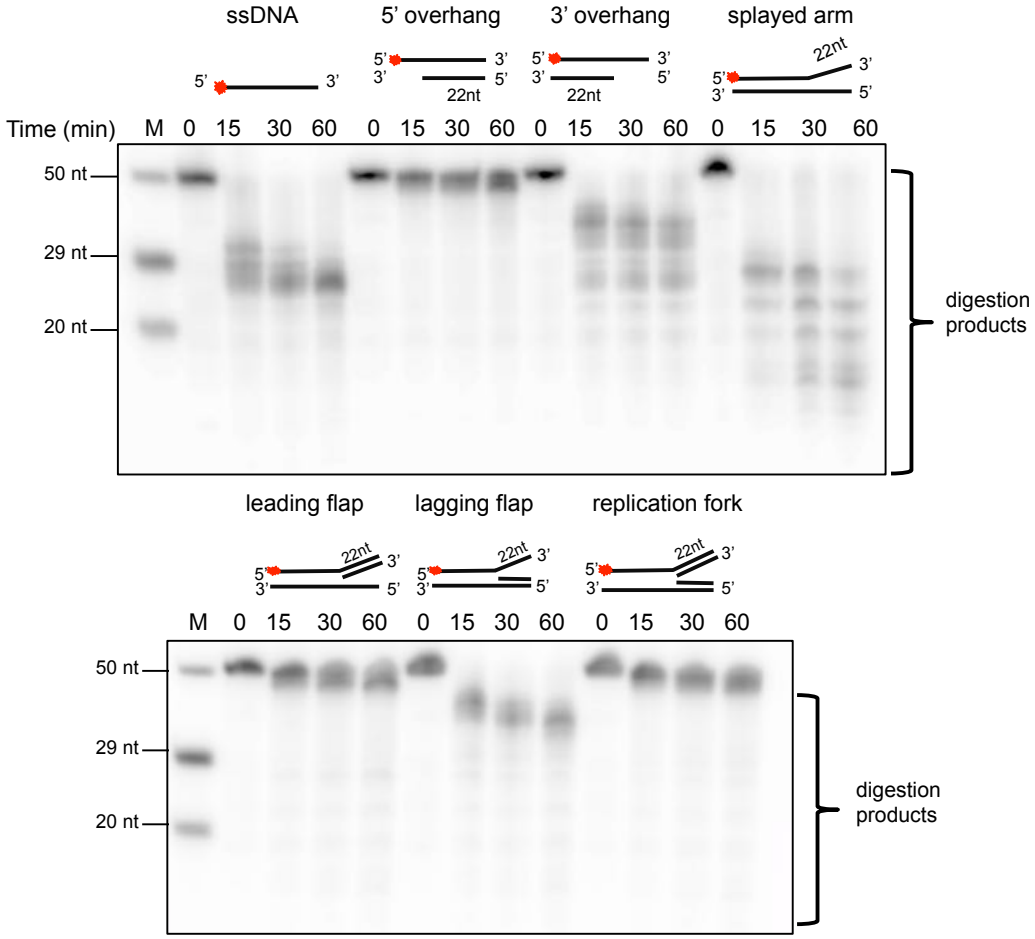


Figure 6.31 Exonuclease activity of WT EXD2 on a variety of structurally diverse DNA substrates.

10 nM of DNA substrate (as labelled) was added to 100 nM WT EXD2 and incubated at 37°C for increasing time. The reaction was stopped with the addition of 2 µL stop solution and boiling. Samples were then analysed by 20% polyacrylamide gel electrophoresis. DNA substrate structures are as indicated. Molecular weight markers are as labelled. Each gel is representative of at least three individual experiments.

To confirm that WT EXD2 was acting as a *bona fide* exonuclease, and not an endonuclease, activity was subsequently examined on the substrate with a biotin moiety on the 3' end. The addition of a biotin molecule occludes the 3' end of the DNA substrate and, therefore, does not allow terminal loading or exonucleolytic activity. This was performed at both 100 nM and 150 nM EXD2 to determine if increasing concentration affected exo-/endonuclease activity. However, results for these were comparable and only data for 150 nM are shown (Figure 6.32).

Increased EXD2 concentration promoted digestion of all DNA substrates, however, activity was still strongest on those with a 3' single stranded region. However, WT EXD2 exhibited no endonucleolytic activity as the presence of a 3' biotin (as opposed to a 3' hydroxyl group) completely abrogated digestion (Figure 6.31).

EXD2 is able to digest a variety of DNA structures with a clear preference for a splayed arm substrate (Figure 6.31). However, there was also significant nuclease activity on a lagging flap substrate (Figure 6.31), and some activity on leading flap and replication fork substrates at a higher EXD2 concentration (Figure 6.32). These three substrates, in particular, mimic stalled replication forks (or intermediates in their repair), and this may be pertinent as EXD2 depleted cells exhibit hypersensitivity to the topoisomerase I inhibitor, camptothecin, which contributes to fork stalling in DNA replication (Smogorzewska et al., 2010). Thus, EXD2 may play a cellular role in resolving stalled replication forks, or partially processed intermediates.

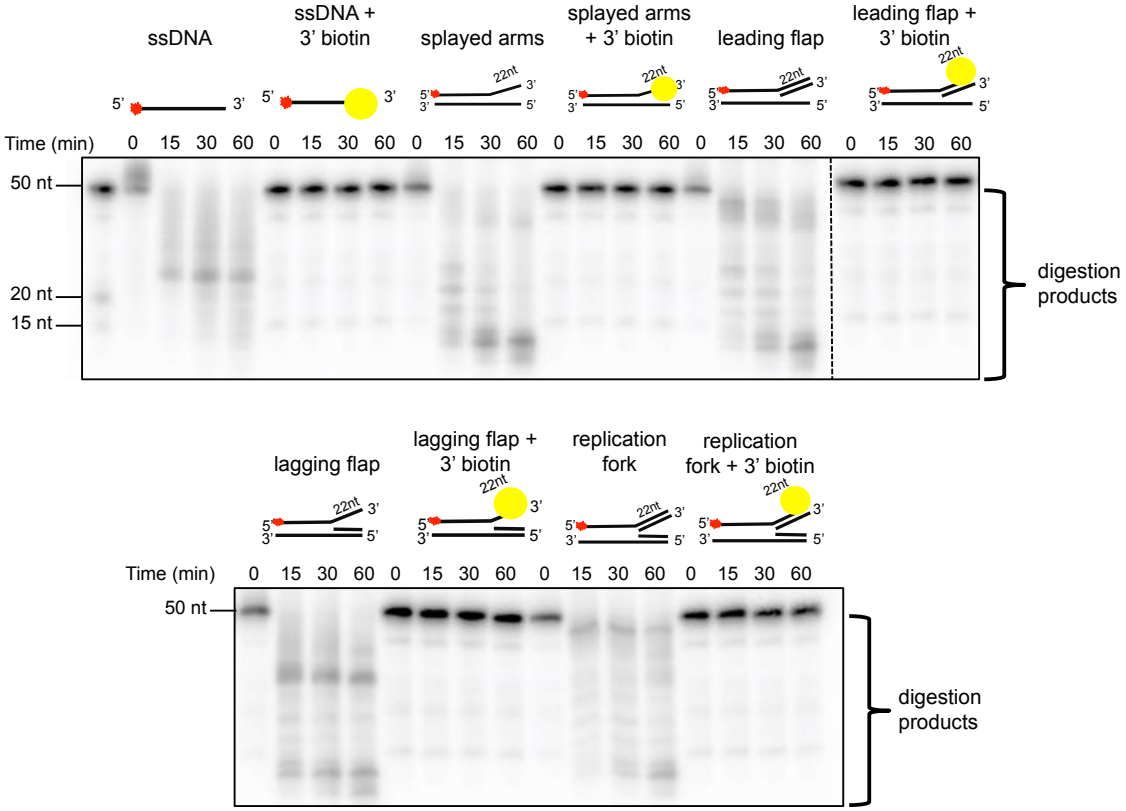


Figure 6.32 WT EXD2 possesses exonucleolytic, but not endonucleolytic activity on a variety of DNA substrates.

10 nM of DNA substrate with either a 3' OH group, or a 3' biotin group (indicated by the yellow circle) was added to 150 nM WT EXD2 and incubated at 37°C for increasing time. The reaction was stopped with the addition of 2 µL stop solution and boiling. Samples were then analysed by 20% polyacrylamide gel electrophoresis. DNA substrate structures are as indicated. Molecular weight markers are as labelled. Each gel is representative of at least three individual experiments.

6.2.8. Nuclease activity of EXD2 on nicked DNA substrates

The repair of DSBs via HR requires processing to generate the ssDNA 3' overhangs that are necessary for RAD51 mediated strand exchange and subsequent recombination events. This is discussed in detail in Chapter 1, however, can be briefly summarised as follows. Initial resection is carried out by the MRN complex and CtIP to generate short regions of ssDNA (Ciccia and Elledge, 2010, Mimitou and Symington, 2011). Further resection, to generate longer ssDNA regions, is then carried out by the EXO1 or DNA2 nucleases, in collaboration with the Bloom's Syndrome helicase (BLM) (Blackwood et al., 2013, Mimitou and Symington, 2008). These ssDNA regions are initially bound by replication protein A (RPA) (Sartori et al., 2007, Zhu et al., 2008), which is subsequently displaced by RAD51, which facilitates the strand exchange process (Esashi et al., 2007).

The work by Dr Ronan Broderick and Dr Jadwiga Nieminuszczy suggested a role for EXD2 in this end resection process, requisite for HR mediated DSB repair. Briefly, evidence for this consisted of the following: EXD2 depleted cells exhibited spontaneous chromosome instability and sensitivity to DSB inducing agents and to the PARP inhibitor, Olaparib; EXD2 depleted cells showed significantly impaired ssDNA, RPA, and RAD51 foci formation following ionising radiation (IR); EXD2 was also shown to be functionally epistatic with MRE11; and the nuclease-dead version of EXD2 (D108A/E110A) could not recapitulate the EXD2 depleted phenotype in cells, suggesting that it was the nuclease activity of EXD2 that was responsible for mediating its role in DSB repair (Broderick et al., 2016).

It has previously been shown that MRE11, which possesses both exonucleolytic and endonucleolytic activity *in vitro*, has the ability to nick the DNA strand that is to be resected at multiple points up to 300 bp distal to the break (Garcia et al., 2011, Jazayeri et al., 2008). Therefore, given the cellular data that implicated EXD2 in this process, it was hypothesised that EXD2 could load onto an MRE11 nicked dsDNA substrate and digest the nicked DNA strand in the 3' to 5' direction, towards the site of the DSB, thus contributing to the formation of the 3' single stranded DNA overhang (Broderick et al., 2016). This is illustrated schematically in Figure 6.33.

To examine this, the *in vitro* nuclease activity of EXD2 on both 'nicked' and 'gapped' DNA substrates, mimicking those resulting from endonucleolytic incision of MRE11, was examined (Figure 6.34). The 'nicked' DNA structure consisted of a break in the phosphodiester backbone on one strand of the duplex substrate, whereas the 'gapped' substrate also possessed a 1 nt gap (Figure 6.34).

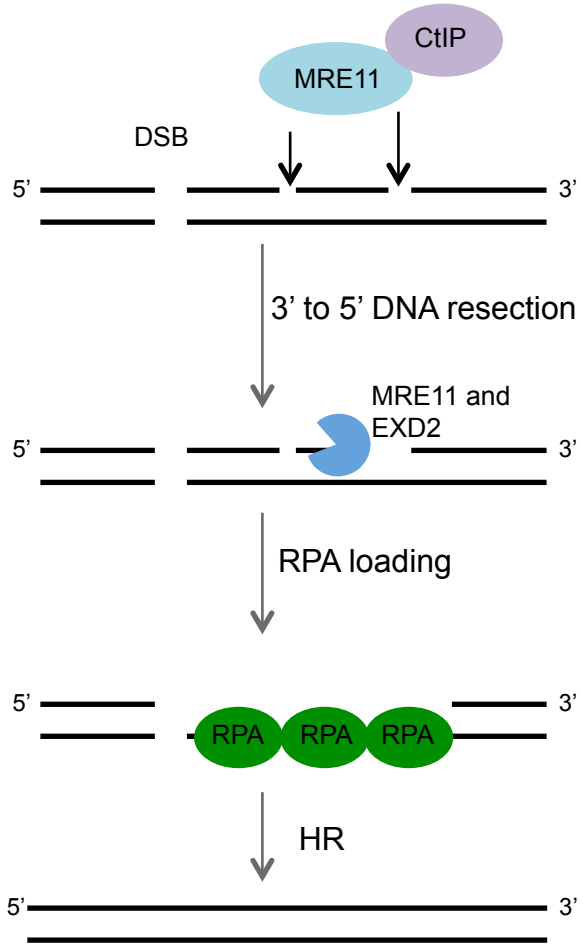


Figure 6.33 Understanding the role of EXD2 in end resection, prior to strand invasion in HR.

A: Schematic of the putative role for EXD2 in end resection, prior to strand invasion in HR. Induction of a DSB results in the recruitment of MRE11 and CtIP. Subsequently, MRE11 (stimulated by CtIP) makes a series of endonucleolytic incisions proximal to the DSB. The exonucleolytic activity of EXD2 (perhaps in conjunction with MRE11) then digests the nicked strand in the 3' to 5' direction. The newly generated ssDNA overhangs are coated with RPA and then the downstream steps of RAD51 displacement, strand invasion, and homologous recombination occur.

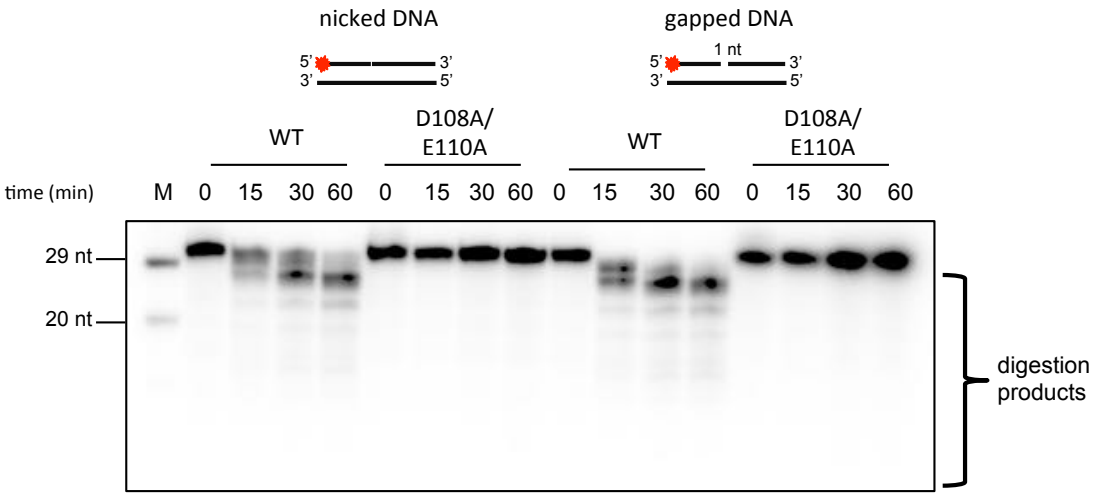


Figure 6.34 WT EXD2, but not D108A/E110A EXD2, is able to effectively digest both a 'nicked' and 'gapped' dsDNA substrate.

10 nM of DNA substrate (either 'nicked' or 'gapped') was added to 100 nM EXD2 and incubated at 37°C for increasing time. The reaction was stopped with the addition of 2 µL stop solution and boiling. Samples were then analysed by 20% polyacrylamide gel electrophoresis. DNA substrate structures are as indicated. Molecular weight markers are as labelled. Each gel is representative of at least three individual experiments.

WT EXD2 exhibited effective exonucleolytic digestion of both the nicked and gapped duplex DNA substrate (Figure 6.34). D108A/110A EXD2 showed no digestion of either substrate. Moreover, if EXD2 were to load onto an MRE11 nicked dsDNA substrate and digest (in the 3' to 5' direction) toward the DSB (as outlined in Figure 6.33), the size of the 'gap' in DNA duplex would correspondingly increase. Accordingly, the nuclease activity of EXD2 was examined on dsDNA substrates with increasing 'gap' size (Figure 6.35).

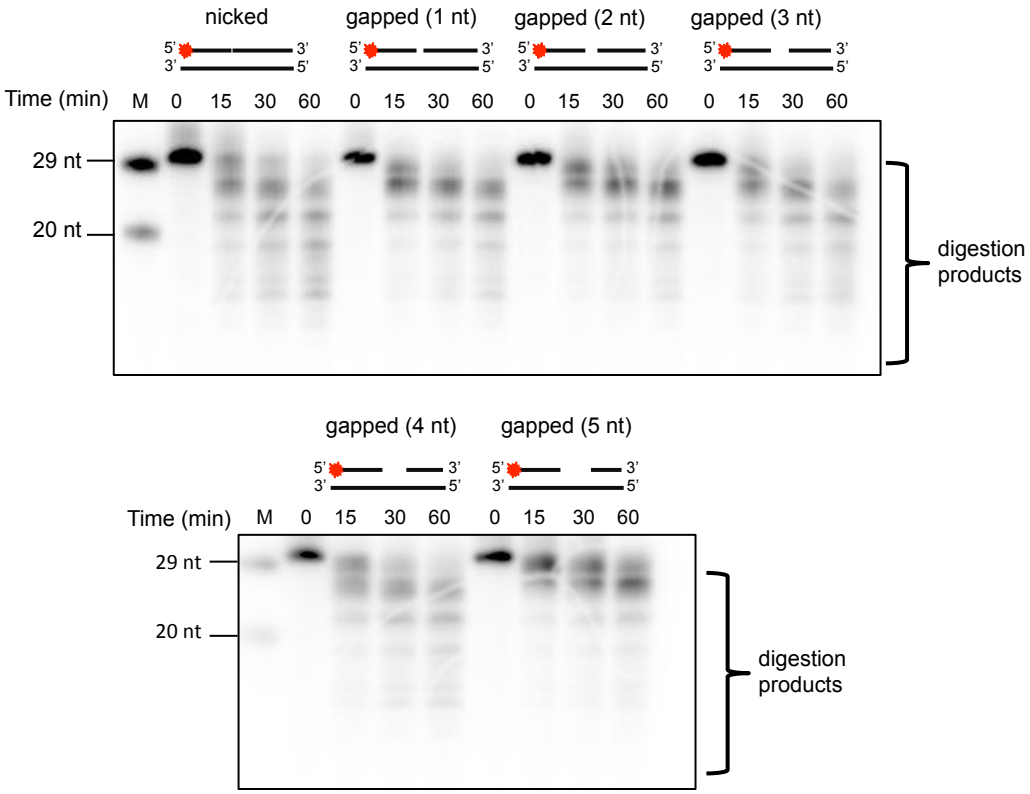


Figure 6.35 WT EXD2 shows effective exonucleolytic digestion of a 'nicked' DNA duplex and 'gapped' DNA duplexes.

10 nM of DNA substrate (as indicated) was added to 100 nM WT EXD2 and incubated at 37°C for increasing time. The reaction was stopped with the addition of 2 µL stop solution and boiling. Samples were then analysed by 20% polyacrylamide gel electrophoresis. DNA substrate structures consist of a 'nicked' substrate (with no nucleotide gap) and 'gapped' substrates, with the size of the 'gap' increasing from 1–5 nt. Molecular weight markers are as labelled. Each gel is representative of at least three individual experiments.

WT EXD2 exhibited effective digestion of all 'gapped' substrates (Figure 6.35). Interestingly, nuclease activity appeared to be strongest on the nicked substrate, and weakest on the 5-nucleotide 'gapped' substrate (Figure 6.35). Although this effect is slight, it may suggest that EXD2 might not necessarily be effective for

long-range resection, or that it needs to load from multiple nicked entry sites. However, it is worthy of note that EXD2 was able to effectively bind and digest a 'nicked' and 'gapped' DNA substrate (even at increasing gap sizes). This is interesting because these substrates are structurally quite different to those containing a free 3' end that prior analysis revealed to be preferred by EXD2 (Figure 6.31).

The data show that EXD2 is able to effectively load and exonucleolytically digest a nicked, or gapped, duplex substrate. As such, it is entirely feasible that EXD2 is able to process MRE11 nicked substrates *in vivo*. Taken together, the biochemical and cellular data (Broderick et al., 2016) strongly suggest a role for EXD2 in end resection requisite for HR. However, further biochemical and cellular analysis is required to fully elucidate this process.

6.2.9. Analysis of the nuclease activity of EXD2 on different length DNA substrates

The data show that WT EXD2 is able to effectively digest a ssDNA substrate (Figure 6.27 and Figure 6.31). However, it was not known whether there was a length preference for EXD2 on ssDNA substrates. To establish this, the nuclease activity of EXD2 on different length ssDNA substrates was examined (Figure 6.36).

WT EXD2 shows clear nuclease activity on ssDNA substrates varying between 20–100 nucleotides, and activity on all substrates appears to be fairly comparable (Figure 6.36).

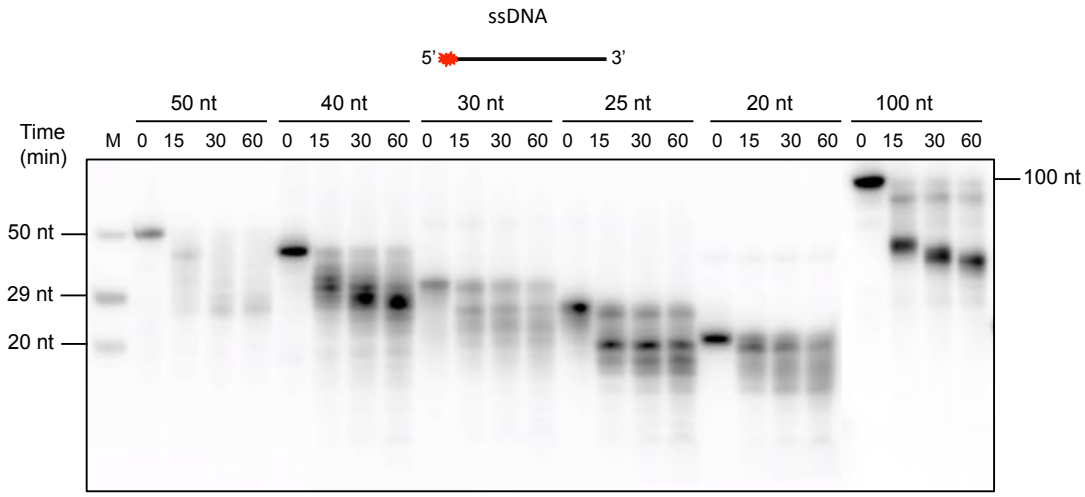


Figure 6.36 WT EXD2 exhibits effective nuclease activity on ssDNA substrates of varying length.

10 nM of DNA substrate (ssDNA, length as indicated) was added to 100 nM WT EXD2 and incubated at 37°C for increasing time. The reaction was stopped with the addition of 2 µL stop solution and boiling. Samples were then analysed by 20% polyacrylamide gel electrophoresis. Molecular weight markers are as labelled. Each gel is representative of at least three individual experiments.

The activity of WT and D108A/E110A EXD2 was also examined on a gapped, linearised, or supercoiled 2.7 kb plasmid (Figure 6.37). Neither WT EXD2 nor D108A/E110A EXD2 showed any appreciable digestion of any the plasmid substrates (Figure 6.37). It can be seen (Figure 6.37) that there is a small amount of 'nicking' occurring in the D108A/E110A treated gapped plasmid, as it shows increasing linearization with increasing concentrations of D108A/E110A EXD2. However, due to the high concentrations of EXD2 present, coupled with the observation that D108A/E110A EXD2 has negligible nuclease activity on shorter substrates, it seems likely that is due to a contaminant.

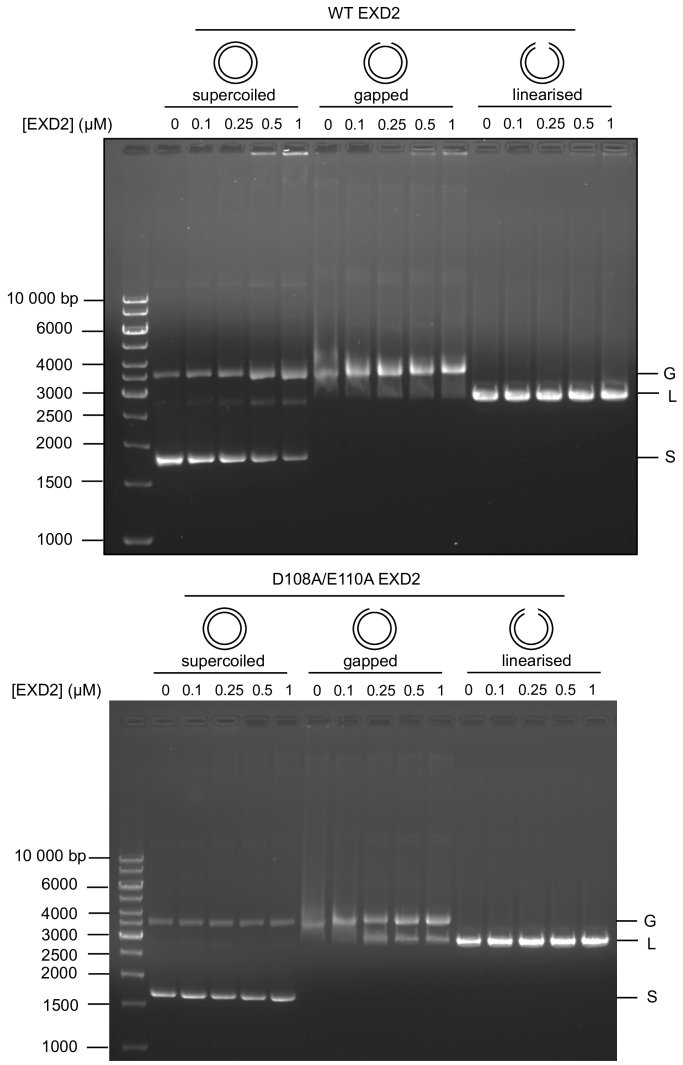


Figure 6.37 Neither WT nor D108A/E110A EXD2 are able to effectively digest nicked, linearised, or supercoiled plasmid DNA substrates.

Nuclease assay showing activity of WT and H61Y SNM1B on supercoiled (S), gapped (G), or linearised (L) 2.7 kb plasmid DNA. Gapped and linearised plasmids were priorly treated with either Nb.BbvCI or HindIII respectively. 325 ng of plasmid DNA was incubated with increasing concentrations of EXD2 (as indicated) at 37^{o}C for 1 h. Samples were analysed on a 1% agarose gel, stained with ethidium bromide.

The inability of WT EXD2 to digest a linearised or supercoiled plasmid is perhaps not surprising given the prior enzymatic analysis – a supercoiled plasmid offers no

point of access from which EXD2 could exonucleolytically digest; and the slight overhang (4 nt) generated from treatment with *HindIII* may not be sufficient to allow EXD2 to bind and digest. Interestingly, WT EXD2 was unable to digest a gapped plasmid, perhaps because the gap was too large to allow EXD2 binding and digestion (Figure 6.37).

Experimental analysis by Dr Jadwiga Nieminuszczy (also utilising His_6-Δ-EXD2 (K76–V564)) showed that EXD2 was able to digest a circular ssDNA substrate (PhiX 174 – length 5386 nucleotides) after prior endonucleolytic incision by MRE11 (Broderick et al., 2016).

Taken together, these data suggest that EXD2 is able to digest higher molecular weight DNA, however, only ssDNA and in the presence of MRE11 (Broderick et al., 2016). Whether this is due to a stimulatory effect of MRE11, or merely that endonucleolytic incision provides a free ssDNA 3' end that is necessary for digestion, remains to be elucidated, and further research is necessary to delineate this.

6.2.10. Analysis of the processivity of EXD2

This study has shown that WT EXD2 binds ssDNA with relatively high affinity (K_d = 11 nM ± 4 nM), however, little was known about the processivity of WT EXD2. As such, a competition-based processivity was performed.

The nuclease reaction was allowed to proceed for 10 minutes, before the addition of unlabelled DNA substrate, and then allowed to continue for a further 50 minutes. This reaction was performed both with ss- and splayed arm DNA, as each are preferred substrates of EXD2 (Figure 6.31). The experimental data show that EXD2 possesses reasonable processivity, as 1.0 μM of unlabelled substrate (a one hundred-fold molar excess) is required to effectively inhibit the nuclease reaction for both (Figure 6.38). Interestingly, processivity appears to be slightly higher on ssDNA, when compared to splayed arm DNA, and this may reflect the effect of structural differences on enzymatic processivity. It may also be, that *in vivo*, accessory factors are required to enhance the processivity of EXD2 – perhaps similar to what is observed for EXO1 and PCNA (Chen et al., 2013).

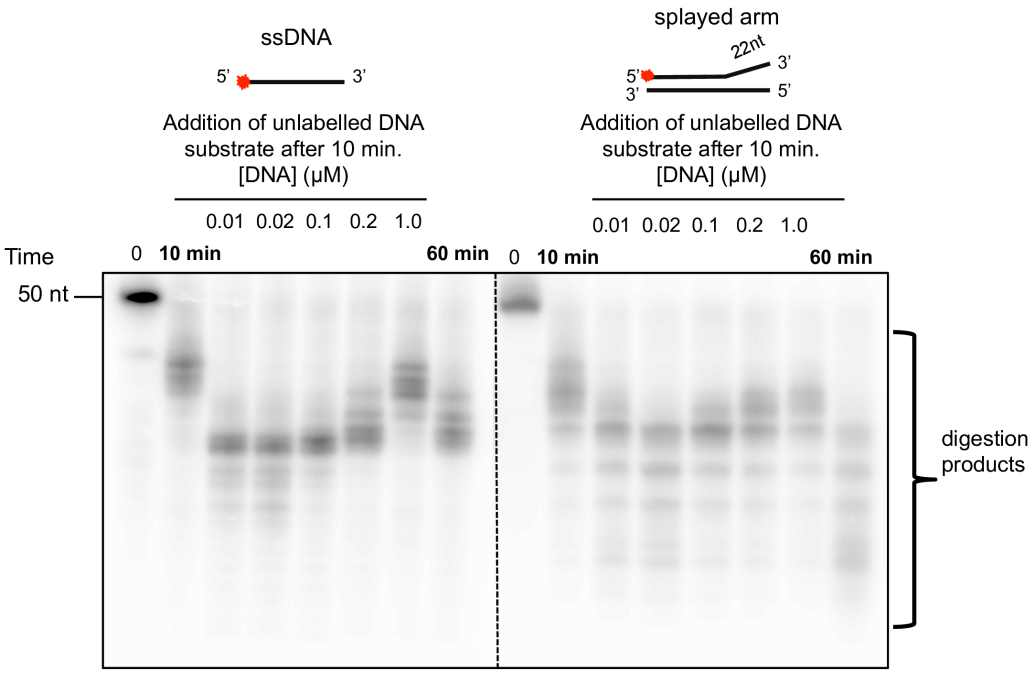


Figure 6.38 EXD2 exhibits some processivity on ss- and splayed arm DNA substrates.

Gel based 'competition' assay whereby 10 nM of DNA substrate (either ss- or splayed arm DNA, as indicated) was added to 100 nM WT EXD2 and incubated at 37°C for 10 min, at which point the reaction was either stopped, or increasing amounts of 'cold', unlabelled DNA substrate (either ss- or splayed arm DNA) was added to the reaction mix, and these were then incubated for a total period of 60 min. A single 60 min reaction was included where no additional unlabelled substrate was added. Samples were then run on a 20% polyacrylamide gel and each gel is representative of at least three individual experiments.

6.2.11. A novel role for EXD2 in processing RNA substrates

As EXD2 is a newly characterised nuclease, the activity on ssRNA and RNA:DNA hybrids was also examined (Figure 6.39).

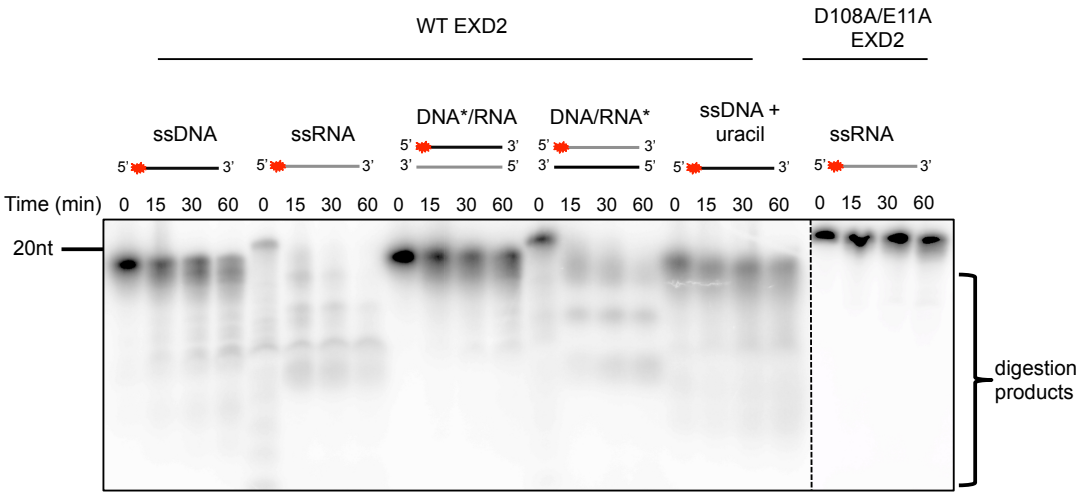


Figure 6.39 WT EXD2 shows significant nuclease activity on RNA-based substrates.

10 nM of DNA, or RNA substrate was added to 100 nM WT EXD2 and incubated at 37°C for increasing time. The reaction was stopped with the addition of 2 µL stop solution and boiling. Samples were then analysed by 20% polyacrylamide gel electrophoresis. DNA is represented by a black bar, and RNA with a grey bar. The asterisk in the annealed DNA:RNA structures indicates the radiolabelled strand. ssDNA + uracil denotes ssDNA where thymine has been substituted with uracil. Gel is representative of three individual experiments.

Intriguingly, WT EXD2 displayed robust nuclease activity on a ssRNA substrate, and showed much more efficient digestion of a ssRNA substrate compared to a ssDNA substrate of the same length (Figure 6.39). Additionally, EXD2 exhibited effective digestion of an RNA:DNA hybrid, when the RNA strand was radiolabelled. This is particularly interesting as WT EXD2 has previously shown to

be unable to digest a blunt-ended dsDNA substrate (Figure 6.27) and exhibits negligible digestion of the DNA strand of an RNA:DNA hybrid, when the DNA strand is radiolabelled (Figure 6.39). Importantly, D108A/E110A EXD2 exhibits no significant nuclease activity on a ssRNA substrate. The activity of D108A/E110A EXD2 was examined on all of these substrates; however, none showed any significant digestion, and only ssRNA is shown (Figure 6.39).

These results are interesting, and may suggest a novel role for EXD2 in processing RNA or RNA:DNA structures in the cell.

6.3. Conclusion

The present study provides a basic biochemical characterisation of the novel exonuclease, EXD2. It has been demonstrated that EXD2 is *bona fide* exonuclease with 3' to 5' polarity and possesses no intrinsic endonuclease activity on the substrates examined. Native MS and chelation studies suggest that there is at least one Zn^{2+} ion present in the active site, and specific reaction conditions suggest that the second metal ion-binding site may be occupied by Mn^{2+} *in vitro* to allow nuclease activity. However, how this reflects physiological conditions is not yet known.

It may be of value to compare the metal ion status of EXD2, with the related nuclease, WRN. However, understanding which metal ion best supports nuclease activity of WRN is somewhat confounded by the fact that the presence of different metal ions can facilitate the switch from helicase to exonuclease function

(Choudhary et al., 2004). Nonetheless, studies have shown that each Zn^{2+}, Mg^{2+}, and Mn^{2+} support the enzymatic activity of the purified exonuclease domain of WRN, with a strong preference for either Mn^{2+} (Perry et al., 2006) or Zn^{2+} (Choudhary et al., 2004) over Mg^{2+}. However, higher concentrations of Zn^{2+} (above 500 μM) were inhibitory to nuclease activity (Choudhary et al., 2004, Choi et al., 2007). The crystal structure of the exonuclease fold of WRN has also been solved, with either two Mn^{2+} ions, two Mg^{2+} ions (Perry et al., 2006) or two Zn^{2+} ions (Choi et al., 2007) coordinated in the active site.

However, with regards to EXD2, it is worthy of note that intracellular concentrations of Mn^{2+} are relatively low (Lyons, 2006), and that the addition of Mn^{2+} can sometimes relax substrate specificity (Yang et al., 2006). Thus, it may be that in this *in vitro* analysis some of the necessary interacting factors (perhaps other proteins) to enhance the nuclease activity of EXD2 are not present, and thus Mn^{2+} is preferred as it allows catalysis, whereas *in vivo,* Mg^{2+} (or another divalent metal ion) may be more physiologically relevant.

EXD2 is also able to digest a variety of DNA substrates and structures and the published literature is indicative for a role in DNA damage repair (Broderick et al., 2016, Smogorzewska et al., 2010), however, there is much about its physiological role(s) that remains to be elucidated. The *in vitro* data from this study suggest that EXD2 may have a role in processing replication fork intermediates, contributing to DSB end resection from a nicked dsDNA substrate, or in the resolution of RNA:DNA hybrids. As these are results from an *in vitro* analysis, their application

to *in vivo* processes is, at present, purely hypothetical, however, they do point to some very interesting avenues for further research.

With regards to substrate preference, the observed nucleolytic activity of EXD2 on RNA based structures is particular interesting, as it may point to a novel cellular role, perhaps in the resolution of RNA:DNA hybrids. The best documented instances of RNA:DNA structures arising in the cell are R-loops. R-loops are non-canonical, three-stranded, nucleic acid structures composed of a nascent RNA strand annealed to the template strand of a DNA bubble. They can result from a variety of circumstances, but most *in vivo* evidence to date suggests that they commonly arise as a result of transcriptional processes (Drolet et al., 1995). The exact mode by which R-loop structures arise is somewhat contentious; however, their formation is dependent on features such as G-rich DNA sequences (Reaban et al., 1994, Li and Manley, 2005, Ginno et al., 2012), DNA nicks on the non-template strand (Roy et al., 2010), and negative supercoiling of dsDNA (Roy et al., 2010).

However, once formed, RNA-loops are particularly persistent, due to the enhanced thermodynamic stability of RNA/DNA associations, compared with DNA/DNA interactions (Roberts and Crothers, 1992). Research has shown that R-loops can be important for a number of cellular processes, such as regulating gene expression (Ginno et al., 2012, Skourti-Stathaki et al., 2014, Sun et al., 2013). Nonetheless, persistent R-loops can pose a threat to genome integrity, and two main mechanisms for this have been suggested. Firstly, their presence can act as a barrier to advancing replication forks, thus promoting replication stress,

subsequent fork collapse, and double strand break formation (Aguilera and Garcia-Muse, 2012). Secondly, the unpaired ssDNA strand in the R-loop is presumably more susceptible to DNA damage, through processes such as deamination of dC to dU, ultimately resulting in the formation of single stranded, or double stranded DNA lesions (Li and Manley, 2006, Aguilera and Garcia-Muse, 2012).

Thus, there are a number of different processes that prevent the unwanted over-accumulation of R-loops *in vivo*. Initially, their formation is tightly controlled by mechanisms including the following: regulating transcriptional elongation (Aguilera, 2002); controlling mRNA export (Strasser et al., 2002); relaxing DNA supercoiling; and controlling histone methylation (Yang et al., 2014).

However, once present and persistent in the genome, R-loops are generally removed by one of two mechanisms. Firstly, RNase H specifically degrades the RNA strand of RNA:DNA hybrids, thus allowing the reformation of standard duplex DNA (Wahba et al., 2011). Secondly, R-loops can be resolved by SETX, a protein that possesses an RNA/DNA helicase domain, and can thus putatively displace the nascent RNA strand, allowing the two DNA strands to reanneal (Skourti-Stathaki et al., 2011).

It is a very intriguing possibility that EXD2 may be involved in the resolution of R-loops, and it may be that it functions in conjunction with one of the aforementioned pathways, or in an entirely alternative manner. However, additional cellular and biochemical research is needed to further elucidate the role of EXD2 in digesting

ssRNA and RNA:DNA structures *in vivo*. With regards to cellular approaches, the accumulation of RNA:DNA hybrids in EXD2 depleted cells may be examined by immunodetection (Boguslawski et al., 1986)

It is also worthy of note, that the preferred DNA substrate for EXD2 is a splayed arm structure as it displays the most complete digestion over the analysed time period (Figure 6.31). This is of interest because of EXD2's sequence similarity to the WRN protein, which also possesses an intrinsic helicase activity and is therefore able to unwind substrates prior to exonucleolytic digestion. This raises the possibility that perhaps EXD2 functions with a helicase *in vivo*, perhaps indicating the BLM helicase due to its role in DSB repair. However, this purely speculative and further work on this topic is warranted.

Although not comprehensive, the evidence to date strongly supports a role for EXD2 in DNA damage repair processes (Broderick et al., 2016, Smogorzewska et al., 2010). However, there is much more to still be elucidated. Further biochemical assays with other DNA repair factors to reconstitute DNA damage repair processes *in vitro* may provide valuable information, and of course, additional cellular experiments will allow a deeper understanding of how EXD2 acts to mitigate genetic damage.

7. Discussion and future work

The accurate and timely repair of DNA damage lesions is essential for the maintenance of genomic integrity, and normal cellular function. This study focussed on two DNA repair factors – SNM1B and EXD2 – to further understand how they contribute to DNA repair pathways.

The first aim of this study was to characterise the H61Y SNM1B variant, with regards to its biochemical and biophysical properties. The H61Y coding variant arises as a result of a common SNP (rs11552449) in the genetic sequence of SNM1B, that has been associated with an increased of breast cancer (Michailidou et al., 2013). To determine if this H61Y mutation had any effect on the enzymatic activity, DNA interaction, or the structure of SNM1B, recombinant, truncated WT and H61Y Δ-SNM1B were purified from insect cells, and a number of *in vitro* assays were performed. The data from this study clearly show that the two proteins have comparable nuclease activity, DNA binding, catalytic efficiency, enzymatic processivity, and thermal stability (see Chapter 3). However, as previously noted, there is the caveat that all of these results are derived from *in vitro* assays with truncated, purified SNM1B from a heterologous system. Nevertheless, it does seem likely that, *in vivo*, WT and H61Y SNM1B would possess similar nuclease activity and DNA binding abilities. It does remain to be seen whether the H61Y mutation disrupts protein-protein interactions or has some other cellular effect. The interaction with astrin has been mapped to this region (amino acids 1–140 in the SNM1B protein sequence) (Liu et al., 2009), and the

effect of H61Y on this interaction could be assessed by immunoprecipitation experiments. Other SNM1B protein-protein interactions are presently unmapped (such as the interactions with SLX4, HSP70, HSP72, HSC70, HSP60, β-tubulin, and the interaction with SNM1B itself (Schmiester and Demuth, 2017)), and it's possible that the H61Y mutation could disrupt these. To accurately determine the cellular effect of this mutation it would be necessary to either complement SNM1B disrupted cells with the H61Y variant, or to use homology directed CRISPR-Cas9 techniques to mutate the endogenous loci. It is arguably more likely that this SNP (rs11552449) is associated with cancer risk due to its correlation with alternative splicing of the SNM1B and *PHTF1* transcripts, each also associated with increased breast cancer risk (Caswell et al., 2015). In which case, this would be an example of the uncertainty of interpreting genetic associations with individual SNPs, even in the case of non-synonymous substitutions in protein-coding sequences.

A plethora of genetic studies have identified polymorphisms in a number of DNA damage response/repair genes as contributing to cancer risk; however, the enzymatic or cellular effects of these SNPs is not always clear. Accordingly, the results from this study are important for informing the functional relationship between the H61Y SNM1B genetic variant, enzymatic activity, and cancer risk.

WT (and H61Y) SNM1B are also able to catalytically digest a wide array of structurally diverse DNA substrates with high efficiency *in vitro*. This suggests that *in vivo* SNM1B may play an important cellular role in processing a wide range of

DNA damage lesions, including (but not limited to) stalled replication forks, oxidative adducts, and interstrand crosslinks.

The second aim of this study was to obtain a crystal structure of WT and H61Y SNM1B, and a co-crystal structure of WT Δ-SNM1B with a DNA substrate. The structures of both WT and H61Y in the apo form were solved to 2.8 and 3.1 Å, respectively. The two structures were extremely similar (Figure 4.3), and this is likely the basis for the aforementioned biochemical similarity. The structure of SNM1B was not successfully solved with a DNA substrate. However, WT Δ-SNM1B did crystallise in the presence of 2'-deoxy-5'-AMP, and the structure was solved to 1.8 Å. This structure (WT nucleotide form) shows the active site of Δ-SNM1B with two AMP molecules bound, and thus it can be inferred which amino acids are required for metal ion coordination, nucleotide binding, and likely to be involved in catalysis (Figure 4.14). Some of these amino acid residues had been previously indicated as being catalytically relevant (such as those that coordinate the metal ions); however, the structure of WT (nucleotide form I) imparts further useful information about the active site of SNM1B. The coordination of the 5' phosphate group of AMP_1 defines the phosphate binding pocket, and it is likely that this confers the exonucleolytic selectivity for substrates possessing a 5' phosphate group (Figure 4.7 and Figure 4.8). The positioning of the two AMP molecules, the binding of the phosphate group of AMP_2, and the coordination of the catalytic hydroxide ion, supports the catalytic mechanism that has previously been suggested for SNM1B (Sengerova et al., 2012, Allerston et al., 2015) and that is well-conserved throughout this family of enzymes. Regarding the binding of the two AMP molecules, the electron density map revealed that, in the crystal

structure, the base of AMP_1 was better resolved than the base of AMP_2, which was presumably more disordered (Figure 4.13). The data suggest that the positioning of the first base is stabilised by interaction with R301 and R259, and, mutational analysis of these residues is necessary to experimentally confirm this hypothesis.

The WT (nucleotide form) structure also provided the basis for modelling the interaction of WT Δ-SNM1B with a duplex DNA substrate. B-form duplex DNA was modelled to interact with the extended active site of Δ-SNM1B; with the two AMP molecules constituting the first two nucleotides of the DNA strand that enters the catalytic core (Figure 4.15), and the backbone of the DNA substrate modelled to interact with the positively charged, putative DNA binding cleft across the front face of Δ-SNM1B (Figure 4.15). To experimentally validate that this region is important for substrate binding, a mutational analysis of these residues (namely, R19, R20, S17, R300, R301, R259 and K260) would be helpful to examine the impact on DNA binding, nuclease activity, and enzymatic processivity. Importantly, prior *in vitro* biochemical analysis of S183A revealed decreased nuclease activity and DNA binding, compared to WT Δ-SNM1B, suggesting to the importance of this residue (Allerston et al., 2015).

The open nature of the extended active site of Δ-SNM1B meant that dsDNA could be accommodated in the model, until the final base pair, where hydrogen bonding would be partially disrupted (Figure 4.15). Two–three nucleotides of the newly generated 3' overhang were then modelled to interact with R259 and K260 (Figure 4.15), and this may be relevant when considering the physiological role of SNM1B

at telomeres. The open nature of the active site meant it was also possible to model an 8-oxo-G containing substrate with no conflict to the structure (Figure 4.18). This supports the biochemical data, where Δ-SNM1B is able to digest past an 8-oxo-G with little reduced efficiency (Figure 3.31). However, it was not possible to model an SJG-136 ICL in the active site of Δ-SNM1B, as the ICL constrained separation of the double helix. Given that Δ-SNM1B can digest past ICLs *in vitro* (albeit with significantly reduced efficiency (Allerston et al., 2015)), it would suggest that SNM1B binds to the ICL containing substrate in a non-optimal manner, and a co-crystal structure of SNM1B with an ICL containing substrate in the active site would further elucidate this.

It was also attempted to determine the catalytically relevant metal ions for SNM1B activity. The data are not conclusive, although it seems likely that SNM1B (like SNM1A) has one tightly bound metal ion-binding site, and one more loosely bound metal ion-binding site. It does seem probable that having both sites occupied is requisite for catalysis, given that mutation of D35 and/or H36 abrogates nuclease activity (Sengerova et al., 2012, Ye et al., 2010), and that the hydrolytic mechanism relies on a di-metal catalytic centre with a bridging hydroxide ion. Conversely, in the B2 class of bacterial MBLs, where only one metal ion-binding site is required for catalysis (and two can be inhibitory), the water molecule is instead liganded by a histidine (Palzkill, 2013).

The identity of the metal ions(s) is not clear, and it remains to be determined whether the metal ion(s) that best support catalysis are physiologically relevant. Homology with other MBL family members, the atomic nature of the first metal ion-

binding site, and the biochemical sensitivity to o-phenanthroline together suggest that Zn^{2+} may occupy this site. However, there is also the observation that catalysis is best supported by the presence of Mn^{2+}, even after treatment with EDTA to chelate the endogenous metal ion(s) ((Sengerova et al., 2012) and Figure 4.21). Reconciling these observations requires further experimental work, and will be technically challenging. However, techniques such as inductively coupled plasma mass spectrometry (ICP-MS) of SNM1B purified from human cells may be useful in this instance. Understanding the physiologically relevant metal ion(s) for SNM1B catalysis may not be particularly pertinent for examining its cellular role, but it is a very important consideration in inhibitor design.

This structure, WT (nucleotide form) is the first reported structure of a human MBL-β-CASP family member co-crystallised with an RNA/DNA substrate, or relevant analogue. This offers an insight into how the accessible active site architecture of SNM1B provides a structural basis for the broad range of catalytically accepted substrates. This is relevant when considering the *in vivo* roles of SNM1B in telomere maintenance and in DNA damage repair, and may be useful in guiding future inhibitor design.

The third aim of this study was to generate an SNM1B disrupted cell line, and to use this as a tool for further research. Accordingly, an $SNM1B^{-}$ cell line was generated, and exhibited hypersensitivity to the crosslinking agents, cisplatin and SJG-136 (Figure 5.16), in accordance with previous research. However, further experimental work is required, and it is necessary to validate, via WT SNM1B

complementation, that the observed phenotype is due to SNM1B disruption and not due to a clonal effect.

The fourth aim of this study was to perform a biochemical characterisation of a novel DNA damage repair factor, EXD2. Cellular data implicate EXD2 as playing a role in DSB repair: EXD2 depleted cells are hypersensitive to DSB-inducing agents (including camptothecin, phleomycin, and IR); EXD2 is recruited to chromatin in a DNA damage dependent manner; interacts with CtIP and MRN; and acts in an epistatic manner to MRE11 in end resection, prior to strand invasion in HR (Broderick et al., 2016). This study developed a method for purifying truncated, recombinant EXD2 (WT and D108A/E110A) from *E. coli,* and showed that WT Δ-EXD2 possessed intrinsic 3' to 5' exonuclease activity, similar to the full-length protein (Figure 6.3, Figure 6.19). The necessary biochemical conditions for *in vitro* catalytic activity were determined, including pH, metal ion dependency, and substrate preference. From MS analysis it seems likely that purified Δ-EXD2 contains at least one (or possibly two), Zn^{2+} ion(s) in the active site. How this represents physiological conditions is unknown. The dependence of purified EXD2 on Mn^{2+} for activity is not unprecedented amongst nucleases (EXO1 being a notable example); however, it's not clear whether Mn^{2+} binds to the active site or is required for catalysis in another way. Δ-EXD2 displays nuclease activity on a variety of substrate structures, although, there is a clear preference for DNA with 3' single-stranded regions (Figure 6.31). This may be due to the accessibility of the active site, and a crystal structure of EXD2 would be useful in inferring substrate selectivity from structure.

Interestingly, Δ-EXD2 also displayed the ability to exonucleolytically digest both nicked and gapped duplex DNA substrates (Figure 6.34 and Figure 6.35). Together with the cellular data, which show that EXD2 facilitates the generation of ssDNA at DSBs, and acts downstream of, and in the same pathway as, MRE11, this led to the hypothesis that EXD2 functionally acts in the MRN-CtIP mediated step of short-range resection in HR (Broderick et al., 2016). According to this hypothesis, EXD2 acts from the endonucleolytic incision by MRE11 distal to the DSB, and then exonucleolytically processes back towards the DSB (Figure 6.33), thus contributing to short-range resection. Why two nucleases with the same polarity (MRE11 and EXD2) would function at the same point during short-range resection in HR is not immediately clear. However, short-range resection is a highly important cellular process, and functional redundancy at this step may promote faster generation of ssDNA tracts, or MRE11/EXD2 might be involved in different molecular circumstances (for example, in the case of chemically modified DSB ends). Similar to MRE11, EXD2 displays relatively weak exonucleolytic activity *in vitro*, and this may be due to a lack of accessory factors interacting with EXD2, or generate a specific residual intermediate by prior processing of the DNA substrate. Biochemical reconstitution in *in vitro* assays with the addition of candidate proteins may help to elucidate this. Interestingly, EXD2 also showed significant exoribonuclease activity (Figure 6.39). This result requires further biochemical and cellular analysis; however, it does suggest that EXD2 may process RNA or RNA:DNA structures in cells; for example, the processing of R-loops, or in the removal of RNA primers from Okazaki fragments. Regarding this, it may be informative to investigate RNA:DNA hybrid accumulation in EXD2 deficient

cells (Boguslawski et al., 1986), perhaps with concomitant depletion with other factors known to contribute to these pathways.

8. Bibliography

ABDULLAH, U. B., MCGOURAN, J. F., BROLIH, S., PTCHELKINE, D., EL-SAGHEER, A. H., BROWN, T. & MCHUGH, P. J. 2017. RPA activates the XPF-ERCC1 endonuclease to initiate processing of DNA interstrand crosslinks. *EMBO J,* 36**,** 2047-2060.

ABRAHAM, J., LEMMERS, B., HANDE, M. P., MOYNAHAN, M. E., CHAHWAN, C., CICCIA, A., ESSERS, J., HANADA, K., CHAHWAN, R., KHAW, A. K., MCPHERSON, P., SHEHABELDIN, A., LAISTER, R., ARROWSMITH, C., KANAAR, R., WEST, S. C., JASIN, M. & HAKEM, R. 2003. Eme1 is involved in DNA damage processing and maintenance of genomic stability in mammalian cells. *EMBO J,* 22**,** 6137-47.

ADAMS, K. E., MEDHURST, A. L., DART, D. A. & LAKIN, N. D. 2006. Recruitment of ATR to sites of ionising radiation-induced DNA damage requires ATM and components of the MRN protein complex. *Oncogene,* 25**,** 3894-904.

ADAMS, M. D., MCVEY, M. & SEKELSKY, J. J. 2003. Drosophila BLM in double-strand break repair by synthesis-dependent strand annealing. *Science,* 299**,** 265-7.

AGUILERA, A. 2002. The connection between transcription and genomic instability. *EMBO J,* 21**,** 195-201.

AGUILERA, A. & GARCIA-MUSE, T. 2012. R loops: from transcription byproducts to threats to genome stability. *Mol Cell,* 46**,** 115-24.

AHMAD, A., ROBINSON, A. R., DUENSING, A., VAN DRUNEN, E., BEVERLOO, H. B., WEISBERG, D. B., HASTY, P., HOEIJMAKERS, J. H. & NIEDERNHOFER, L. J. 2008. ERCC1-XPF endonuclease facilitates DNA double-strand break repair. *Mol Cell Biol,* 28**,** 5082-92.

AKHTER, S., LAM, Y. C., CHANG, S. & LEGERSKI, R. J. 2010. The telomeric protein SNM1B/Apollo is required for normal cell proliferation and embryonic development. *Aging Cell,* 9**,** 1047-56.

ALBERTELLA, M. R., GREEN, C. M., LEHMANN, A. R. & O'CONNOR, M. J. 2005. A role for polymerase eta in the cellular tolerance to cisplatin-induced damage. *Cancer Res,* 65**,** 9799-806.

ALBRECHT, T. R. & WAGNER, E. J. 2012. snRNA 3' end formation requires heterodimeric association of integrator subunits. *Mol Cell Biol,* 32**,** 1112-23.

ALI, A. A. E., TIMINSZKY, G., ARRIBAS-BOSACOMA, R., KOZLOWSKI, M., HASSA, P. O., HASSLER, M., LADURNER, A. G., PEARL, L. H. & OLIVER, A. W. 2012. The zinc-finger domains of PARP1 cooperate to recognize DNA strand breaks. *Nat Struct Mol Biol,* 19**,** 685-692.

ALLERS, T. & LICHTEN, M. 2001. Differential timing and control of noncrossover and crossover recombination during meiosis. *Cell,* 106**,** 47-57.

ALLERSTON, C. K., LEE, S. Y., NEWMAN, J. A., SCHOFIELD, C. J., MCHUGH, P. J. & GILEADI, O. 2015. The structures of the SNM1A and SNM1B/Apollo nuclease domains reveal a potential basis for their distinct DNA processing activities. *Nucleic Acids Res,* 43**,** 11047-60.

ALTMEYER, M. & LUKAS, J. 2013. To spread or not to spread--chromatin modifications in response to DNA damage. *Curr Opin Genet Dev,* 23**,** 156-65.

AN, J., WU, M., XIN, X., LIN, Z., LI, X., ZHENG, Q., GUI, X., LI, T., PU, H., LI, H. & LU, D. 2016. Inflammatory related gene IKKalpha, IKKbeta, IKKgamma cooperates to determine liver cancer stem cells progression by altering telomere via heterochromatin protein 1-HOTAIR axis. *Oncotarget,* 7**,** 50131-50149.

ANDEGEKO, Y., MOYAL, L., MITTELMAN, L., TSARFATY, I., SHILOH, Y. & ROTMAN, G. 2001. Nuclear retention of ATM at sites of DNA double strand breaks. *J Biol Chem,* 276**,** 38224-30.

ANDERS, M., MATTOW, J., DIGWEED, M. & DEMUTH, I. 2009. Evidence for hSNM1B/Apollo functioning in the HSP70 mediated DNA damage response. *Cell Cycle,* 8**,** 1725-32.
ANDERSEN, S. L., BERGSTRALH, D. T., KOHL, K. P., LAROCQUE, J. R., MOORE, C. B. & SEKELSKY, J. 2009. Drosophila MUS312 and the vertebrate ortholog BTBD12 interact with DNA structure-specific endonucleases in DNA repair and recombination. *Mol Cell,* 35**,** 128-35.
ANDREASSEN, P. R., D'ANDREA, A. D. & TANIGUCHI, T. 2004. ATR couples FANCD2 monoubiquitination to the DNA-damage response. *Genes Dev,* 18**,** 1958-63.
ANDRES, S. N., VERGNES, A., RISTIC, D., WYMAN, C., MODESTI, M. & JUNOP, M. 2012. A human XRCC4-XLF complex bridges DNA. *Nucleic Acids Res,* 40**,** 1868-78.
ANTONACCI, C., CHAIRES, J. B. & SHEARDY, R. D. 2007. Biophysical characterization of the human telomeric (TTAGGG)4 repeat in a potassium solution. *Biochemistry,* 46**,** 4654-60.
APARICIO, T., BAER, R., GOTTESMAN, M. & GAUTIER, J. 2016. MRN, CtIP, and BRCA1 mediate repair of topoisomerase II-DNA adducts. *J Cell Biol,* 212**,** 399-408.
ARAVIND, L. 1999. An evolutionary classification of the metallo-beta-lactamase fold proteins. *In Silico Biol,* 1**,** 69-91.
ARNAUDEAU, C., LUNDIN, C. & HELLEDAY, T. 2001. 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**,** 1235-45.
AUDEBERT, M., SALLES, B. & CALSOU, P. 2004. 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.
AUERBACH, A. D. 2015. Diagnosis of Fanconi anemia by diepoxybutane analysis. *Curr Protoc Hum Genet,* 85**,** 8 7 1-17.
AYLON, Y., LIEFSHITZ, B. & KUPIEC, M. 2004. The CDK regulates repair of double-strand breaks by homologous recombination during the cell cycle. *EMBO J,* 23**,** 4868-75.
BAE, J. B., MUKHOPADHYAY, S. S., LIU, L., ZHANG, N., TAN, J., AKHTER, S., LIU, X., SHEN, X., LI, L. & LEGERSKI, R. J. 2008. Snm1B/Apollo mediates replication fork collapse and S Phase checkpoint activation in response to DNA interstrand cross-links. *Oncogene,* 27**,** 5045-56.
BAHMED, K., NITISS, K. C. & NITISS, J. L. 2010. Yeast Tdp1 regulates the fidelity of nonhomologous end joining. *Proc Natl Acad Sci U S A,* 107**,** 4057-62.
BAILLY, V. & VERLY, W. G. 1988. Possible roles of beta-elimination and delta-elimination reactions in the repair of DNA containing AP (apurinic/apyrimidinic) sites in mammalian cells. *Biochem J,* 253**,** 553-9.
BAKER, K. E. & PARKER, R. 2004. Nonsense-mediated mRNA decay: terminating erroneous gene expression. *Curr Opin Cell Biol,* 16**,** 293-9.
BALDWIN, G. S., WALEY, S. G. & ABRAHAM, E. P. 1979. Identification of histidine residues that act as zinc ligands in beta-lactamase II by differential tritium exchange. *Biochem J,* 179**,** 459-63.
BARTEK, J., LUKAS, C. & LUKAS, J. 2004. Checking on DNA damage in S phase. *Nat Rev Mol Cell Biol,* 5**,** 792-804.
BARTSCH, H. & MONTESANO, R. 1984. Relevance of nitrosamines to human cancer. *Carcinogenesis,* 5**,** 1381-93.
BEBRONE, C. 2007. Metallo-beta-lactamases (classification, activity, genetic organization, structure, zinc coordination) and their superfamily. *Biochem Pharmacol,* 74**,** 1686-701.
BEKKER-JENSEN, S., LUKAS, C., KITAGAWA, R., MELANDER, F., KASTAN, M. B., BARTEK, J. & LUKAS, J. 2006. Spatial organization of the mammalian genome

surveillance machinery in response to DNA strand breaks. *J Cell Biol,* 173**,** 195-206.
BENNETT, J. W. & KLICH, M. 2003. Mycotoxins. *Clin Microbiol Rev,* 16**,** 497-516.
BENVENUTI, M. & MANGANI, S. 2007. Crystallization of soluble proteins in vapor diffusion for x-ray crystallography. *Nat Protoc,* 2**,** 1633-51.
BERTI, M. & VINDIGNI, A. 2016. Replication stress: getting back on track. *Nat Struct Mol Biol,* 23**,** 103-9.
BETERMIER, M., BERTRAND, P. & LOPEZ, B. S. 2014. Is non-homologous end-joining really an inherently error-prone process? *PLoS Genet,* 10**,** e1004086.
BHARGAVA, R., ONYANGO, D. O. & STARK, J. M. 2016. Regulation of Single-Strand Annealing and its Role in Genome Maintenance. *Trends Genet,* 32**,** 566-75.
BHATTACHARYYA, A., EAR, U. S., KOLLER, B. H., WEICHSELBAUM, R. R. & BISHOP, D. K. 2000. 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.
BHATTI, S., KOZLOV, S., FAROOQI, A. A., NAQI, A., LAVIN, M. & KHANNA, K. K. 2011. ATM protein kinase: the linchpin of cellular defenses to stress. *Cell Mol Life Sci,* 68**,** 2977-3006.
BLACKFORD, A. N. & JACKSON, S. P. 2017. ATM, ATR, and DNA-PK: The Trinity at the Heart of the DNA Damage Response. *Mol Cell,* 66**,** 801-817.
BLACKWOOD, J. K., RZECHORZEK, N. J., BRAY, S. M., MAMAN, J. D., PELLEGRINI, L. & ROBINSON, N. P. 2013. End-resection at DNA double-strand breaks in the three domains of life. *Biochem Soc Trans,* 41**,** 314-20.
BLOMEN, V. A., MAJEK, P., JAE, L. T., BIGENZAHN, J. W., NIEUWENHUIS, J., STARING, J., SACCO, R., VAN DIEMEN, F. R., OLK, N., STUKALOV, A., MARCEAU, C., JANSSEN, H., CARETTE, J. E., BENNETT, K. L., COLINGE, J., SUPERTI-FURGA, G. & BRUMMELKAMP, T. R. 2015. Gene essentiality and synthetic lethality in haploid human cells. *Science,* 350**,** 1092-6.
BOCK, C. W., KAUFMAN KATZ, A., MARKHAM, G. D. & GLUSKER, J. P. 1999. Manganese as a Replacement for Magnesium and Zinc: Functional Comparison of the Divalent Ions. *Journal of the American Chemical Society,* 121**,** 7360-7372.
BOGENHAGEN, D. F., MARTIN, D. W. & KOLLER, A. 2014. Initial steps in RNA processing and ribosome assembly occur at mitochondrial DNA nucleoids. *Cell Metab,* 19**,** 618-29.
BOGLIOLO, M., SCHUSTER, B., STOEPKER, C., DERKUNT, B., SU, Y., RAAMS, A., TRUJILLO, J. P., MINGUILLON, J., RAMIREZ, M. J., PUJOL, R., CASADO, J. A., BANOS, R., RIO, P., KNIES, K., ZUNIGA, S., BENITEZ, J., BUEREN, J. A., JASPERS, N. G., SCHARER, O. D., DE WINTER, J. P., SCHINDLER, D. & SURRALLES, J. 2013. Mutations in ERCC4, encoding the DNA-repair endonuclease XPF, cause Fanconi anemia. *Am J Hum Genet,* 92**,** 800-6.
BOGUSLAWSKI, S. J., SMITH, D. E., MICHALAK, M. A., MICKELSON, K. E., YEHLE, C. O., PATTERSON, W. L. & CARRICO, R. J. 1986. Characterization of monoclonal antibody to DNA.RNA and its application to immunodetection of hybrids. *J Immunol Methods,* 89**,** 123-30.
BOTHMER, A., ROBBIANI, D. F., FELDHAHN, N., GAZUMYAN, A., NUSSENZWEIG, A. & NUSSENZWEIG, M. C. 2010. 53BP1 regulates DNA resection and the choice between classical and alternative end joining during class switch recombination. *J Exp Med,* 207**,** 855-65.
BRABEC, V. & LENG, M. 1993. DNA interstrand cross-links of trans-diamminedichloroplatinum(II) are preferentially formed between guanine and complementary cytosine residues. *Proc Natl Acad Sci U S A,* 90**,** 5345-9.
BRENDEL, M., BONATTO, D., STRAUSS, M., REVERS, L. F., PUNGARTNIK, C., SAFFI, J. & HENRIQUES, J. A. 2003. Role of PSO genes in repair of DNA damage of Saccharomyces cerevisiae. *Mutat Res,* 544**,** 179-93.

BRODERICK, R., NIEMINUSZCZY, J., BADDOCK, H. T., DESHPANDE, R. A., GILEADI, O., PAULL, T. T., MCHUGH, P. J. & NIEDZWIEDZ, W. 2016. EXD2 promotes homologous recombination by facilitating DNA end resection. *Nat Cell Biol,* 18**,** 271-80.

BROOKS, P. J. & THERUVATHU, J. A. 2005. DNA adducts from acetaldehyde: implications for alcohol-related carcinogenesis. *Alcohol,* 35**,** 187-93.

BROWN, E. J. & BALTIMORE, D. 2000. ATR disruption leads to chromosomal fragmentation and early embryonic lethality. *Genes Dev,* 14**,** 397-402.

BROWN, J. M. 1997. NCI's anticancer drug screening program may not be selecting for clinically active compounds. *Oncol Res,* 9**,** 213-5.

BRYAN, T. M., ENGLEZOU, A., DALLA-POZZA, L., DUNHAM, M. A. & REDDEL, R. R. 1997. Evidence for an alternative mechanism for maintaining telomere length in human tumors and tumor-derived cell lines. *Nat Med,* 3**,** 1271-4.

BRZEZNIAK, L. K., BIJATA, M., SZCZESNY, R. J. & STEPIEN, P. P. 2011. Involvement of human ELAC2 gene product in 3' end processing of mitochondrial tRNAs. *RNA Biol,* 8**,** 616-26.

BUNTING, S. F., CALLEN, E., WONG, N., CHEN, H. T., POLATO, F., GUNN, A., BOTHMER, A., FELDHAHN, N., FERNANDEZ-CAPETILLO, O., CAO, L., XU, X., DENG, C. X., FINKEL, T., NUSSENZWEIG, M., STARK, J. M. & NUSSENZWEIG, A. 2010. 53BP1 inhibits homologous recombination in Brca1-deficient cells by blocking resection of DNA breaks. *Cell,* 141**,** 243-54.

BUSCOMBE, J. R. 2007. Radionuclides in the management of thyroid cancer. *Cancer Imaging,* 7**,** 202-9.

CADET, J. & WAGNER, J. R. 2014. Oxidatively generated base damage to cellular DNA by hydroxyl radical and one-electron oxidants: similarities and differences. *Arch Biochem Biophys,* 557**,** 47-54.

CALDECOTT, K. W. 2008. Single-strand break repair and genetic disease. *Nat Rev Genet,* 9**,** 619-31.

CALLEBAUT, I., MOSHOUS, D., MORNON, J. P. & DE VILLARTAY, J. P. 2002. Metallo-beta-lactamase fold within nucleic acids processing enzymes: the beta-CASP family. *Nucleic Acids Res,* 30**,** 3592-601.

CANNAVO, E. & CEJKA, P. 2014. Sae2 promotes dsDNA endonuclease activity within Mre11-Rad50-Xrs2 to resect DNA breaks. *Nature,* 514**,** 122-5.

CASSIER-CHAUVAT, C. & MOUSTACCHI, E. 1988. Allelism between pso1-1 and rev3-1 mutants and between pso2-1 and snm1 mutants in Saccharomyces cerevisiae. *Curr Genet,* 13**,** 37-40.

CASWELL, J. L., CAMARDA, R., ZHOU, A. Y., HUNTSMAN, S., HU, D., BRENNER, S. E., ZAITLEN, N., GOGA, A. & ZIV, E. 2015. Multiple breast cancer risk variants are associated with differential transcript isoform expression in tumors. *Hum Mol Genet,* 24**,** 7421-31.

CAVALIERI, E., FRENKEL, K., LIEHR, J. G., ROGAN, E. & ROY, D. 2000. Estrogens as endogenous genotoxic agents--DNA adducts and mutations. *J Natl Cancer Inst Monogr***,** 75-93.

CECCALDI, R., SARANGI, P. & D'ANDREA, A. D. 2016. The Fanconi anaemia pathway: new players and new functions. *Nat Rev Mol Cell Biol,* 17**,** 337-49.

CEJKA, P., CANNAVO, E., POLACZEK, P., MASUDA-SASA, T., POKHAREL, S., CAMPBELL, J. L. & KOWALCZYKOWSKI, S. C. 2010. DNA end resection by Dna2-Sgs1-RPA and its stimulation by Top3-Rmi1 and Mre11-Rad50-Xrs2. *Nature,* 467**,** 112-6.

CESARE, A. J. & GRIFFITH, J. D. 2004. Telomeric DNA in ALT cells is characterized by free telomeric circles and heterogeneous t-loops. *Mol Cell Biol,* 24**,** 9948-57.

CHAN, S. H., YU, A. M. & MCVEY, M. 2010. Dual roles for DNA polymerase theta in alternative end-joining repair of double-strand breaks in Drosophila. *PLoS Genet,* 6**,** e1001005.

CHANKOVA, S. G., DIMOVA, E., DIMITROVA, M. & BRYANT, P. E. 2007. Induction of DNA double-strand breaks by zeocin in Chlamydomonas reinhardtii and the role of increased DNA double-strand breaks rejoining in the formation of an adaptive response. *Radiat Environ Biophys,* 46**,** 409-16.

CHATTERJEE, N. & WALKER, G. C. 2017. Mechanisms of DNA damage, repair, and mutagenesis. *Environ Mol Mutagen,* 58**,** 235-263.

CHEN, X., PAUDYAL, S. C., CHIN, R. I. & YOU, Z. 2013. PCNA promotes processive DNA end resection by Exo1. *Nucleic Acids Res,* 41**,** 9325-38.

CHEN, X. B., MELCHIONNA, R., DENIS, C. M., GAILLARD, P. H., BLASINA, A., VAN DE WEYER, I., BODDY, M. N., RUSSELL, P., VIALARD, J. & MCGOWAN, C. H. 2001. Human Mus81-associated endonuclease cleaves Holliday junctions in vitro. *Mol Cell,* 8**,** 1117-27.

CHEN, Y., WILLIAMS, V., FILIPPOVA, M., FILIPPOV, V. & DUERKSEN-HUGHES, P. 2014. Viral carcinogenesis: factors inducing DNA damage and virus integration. *Cancers (Basel),* 6**,** 2155-86.

CHEN, Y., YANG, Y., VAN OVERBEEK, M., DONIGIAN, J. R., BACIU, P., DE LANGE, T. & LEI, M. 2008. A shared docking motif in TRF1 and TRF2 used for differential recruitment of telomeric proteins. *Science,* 319**,** 1092-6.

CHEUNG-ONG, K., GIAEVER, G. & NISLOW, C. 2013. DNA-damaging agents in cancer chemotherapy: serendipity and chemical biology. *Chem Biol,* 20**,** 648-59.

CHIRUVELLA, K. K., LIANG, Z. & WILSON, T. E. 2013. Repair of double-strand breaks by end joining. *Cold Spring Harb Perspect Biol,* 5**,** a012757.

CHOI, J. M., KANG, S. Y., BAE, W. J., JIN, K. S., REE, M. & CHO, Y. 2007. Probing the roles of active site residues in the 3'-5' exonuclease of the Werner syndrome protein. *J Biol Chem,* 282**,** 9941-51.

CHOUDHARY, S., SOMMERS, J. A. & BROSH, R. M., JR. 2004. Biochemical and kinetic characterization of the DNA helicase and exonuclease activities of werner syndrome protein. *J Biol Chem,* 279**,** 34603-13.

CHU, W. K. & HICKSON, I. D. 2009. RecQ helicases: multifunctional genome caretakers. *Nat Rev Cancer,* 9**,** 644-54.

CICCIA, A. & ELLEDGE, S. J. 2010. The DNA damage response: making it safe to play with knives. *Mol Cell,* 40**,** 179-204.

CICCIA, A., LING, C., COULTHARD, R., YAN, Z., XUE, Y., MEETEI, A. R., LAGHMANI EL, H., JOENJE, H., MCDONALD, N., DE WINTER, J. P., WANG, W. & WEST, S. C. 2007. Identification of FAAP24, a Fanconi anemia core complex protein that interacts with FANCM. *Mol Cell,* 25**,** 331-43.

CLAUSON, C., SCHARER, O. D. & NIEDERNHOFER, L. 2013. Advances in understanding the complex mechanisms of DNA interstrand cross-link repair. *Cold Spring Harb Perspect Biol,* 5**,** a012732.

COHN, M. A., KOWAL, P., YANG, K., HAAS, W., HUANG, T. T., GYGI, S. P. & D'ANDREA, A. D. 2007. A UAF1-containing multisubunit protein complex regulates the Fanconi anemia pathway. *Mol Cell,* 28**,** 786-97.

COLE, R. S. 1973. Repair of DNA containing interstrand crosslinks in Escherichia coli: sequential excision and recombination. *Proc Natl Acad Sci U S A,* 70**,** 1064-8.

CONG, L., RAN, F. A., COX, D., LIN, S., BARRETTO, R., HABIB, N., HSU, P. D., WU, X., JIANG, W., MARRAFFINI, L. A. & ZHANG, F. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science,* 339**,** 819-23.

COOPER, M. P., MACHWE, A., ORREN, D. K., BROSH, R. M., RAMSDEN, D. & BOHR, V. A. 2000. Ku complex interacts with and stimulates the Werner protein. *Genes Dev,* 14**,** 907-12.

COQUELLE, N., HAVALI-SHAHRIARI, Z., BERNSTEIN, N., GREEN, R. & GLOVER, J. N. 2011. Structural basis for the phosphatase activity of polynucleotide kinase/phosphatase on single- and double-stranded DNA substrates. *Proc Natl Acad Sci U S A,* 108**,** 21022-7.

CORTES LEDESMA, F., EL KHAMISY, S. F., ZUMA, M. C., OSBORN, K. & CALDECOTT, K. W. 2009. A human 5'-tyrosyl DNA phosphodiesterase that repairs topoisomerase-mediated DNA damage. *Nature,* 461**,** 674-8.

COSTANTINI, S., WOODBINE, L., ANDREOLI, L., JEGGO, P. A. & VINDIGNI, A. 2007. Interaction of the Ku heterodimer with the DNA ligase IV/Xrcc4 complex and its regulation by DNA-PK. *DNA Repair (Amst),* 6**,** 712-22.

COULTHARD, R., DEANS, A. J., SWUEC, P., BOWLES, M., COSTA, A., WEST, S. C. & MCDONALD, N. Q. 2013. Architecture and DNA recognition elements of the Fanconi anemia FANCM-FAAP24 complex. *Structure,* 21**,** 1648-58.

CRESPAN, E., CZABANY, T., MAGA, G. & HUBSCHER, U. 2012. Microhomology-mediated DNA strand annealing and elongation by human DNA polymerases lambda and beta on normal and repetitive DNA sequences. *Nucleic Acids Res,* 40**,** 5577-90.

CUADRADO, M., MARTINEZ-PASTOR, B., MURGA, M., TOLEDO, L. I., GUTIERREZ-MARTINEZ, P., LOPEZ, E. & FERNANDEZ-CAPETILLO, O. 2006. ATM regulates ATR chromatin loading in response to DNA double-strand breaks. *J Exp Med,* 203**,** 297-303.

D'AMOURS, D., DESNOYERS, S., D'SILVA, I. & POIRIER, G. G. 1999. Poly(ADP-ribosyl)ation reactions in the regulation of nuclear functions. *Biochem J,* 342 (Pt 2)**,** 249-68.

D'ANDREA, A. D. 2010. Susceptibility pathways in Fanconi's anemia and breast cancer. *N Engl J Med,* 362**,** 1909-19.

DAIYASU, H., OSAKA, K., ISHINO, Y. & TOH, H. 2001. Expansion of the zinc metallo-hydrolase family of the beta-lactamase fold. *FEBS Lett,* 503**,** 1-6.

DE KLEIN, A., MUIJTJENS, M., VAN OS, R., VERHOEVEN, Y., SMIT, B., CARR, A. M., LEHMANN, A. R. & HOEIJMAKERS, J. H. 2000. Targeted disruption of the cell-cycle checkpoint gene ATR leads to early embryonic lethality in mice. *Curr Biol,* 10**,** 479-82.

DE LANGE, T. 2005. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev,* 19**,** 2100-10.

DE LANGE, T. 2009. How telomeres solve the end-protection problem. *Science,* 326**,** 948-52.

DE SILVA, I. U., MCHUGH, P. J., CLINGEN, P. H. & HARTLEY, J. A. 2000. Defining the roles of nucleotide excision repair and recombination in the repair of DNA interstrand cross-links in mammalian cells. *Mol Cell Biol,* 20**,** 7980-90.

DE SILVA, I. U., MCHUGH, P. J., CLINGEN, P. H. & HARTLEY, J. A. 2002. Defects in interstrand cross-link uncoupling do not account for the extreme sensitivity of ERCC1 and XPF cells to cisplatin. *Nucleic Acids Res,* 30**,** 3848-56.

DEANS, A. J. & WEST, S. C. 2011. DNA interstrand crosslink repair and cancer. *Nat Rev Cancer,* 11**,** 467-80.

DEBETHUNE, L., KOHLHAGEN, G., GRANDAS, A. & POMMIER, Y. 2002. Processing of nucleopeptides mimicking the topoisomerase I-DNA covalent complex by tyrosyl-DNA phosphodiesterase. *Nucleic Acids Res,* 30**,** 1198-204.

DEJARDIN, J. & KINGSTON, R. E. 2009. Purification of proteins associated with specific genomic Loci. *Cell,* 136**,** 175-86.

DEMUTH, I., BRADSHAW, P. S., LINDNER, A., ANDERS, M., HEINRICH, S., KALLENBACH, J., SCHMELZ, K., DIGWEED, M., MEYN, M. S. & CONCANNON, P. 2008. Endogenous hSNM1B/Apollo interacts with TRF2 and stimulates ATM in response to ionizing radiation. *DNA Repair (Amst),* 7**,** 1192-201.

DEMUTH, I. & DIGWEED, M. 1998. Genomic organization of a potential human DNA-crosslink repair gene, KIAA0086. *Mutat Res,* 409**,** 11-6.

DEMUTH, I., DIGWEED, M. & CONCANNON, P. 2004. Human SNM1B is required for normal cellular response to both DNA interstrand crosslink-inducing agents and ionizing radiation. *Oncogene,* 23**,** 8611-8.

DIANOV, G. L. & HUBSCHER, U. 2013. Mammalian base excision repair: the forgotten archangel. *Nucleic Acids Res,* 41**,** 3483-90.
DILLEY, R. L., VERMA, P., CHO, N. W., WINTERS, H. D., WONDISFORD, A. R. & GREENBERG, R. A. 2016. Break-induced telomere synthesis underlies alternative telomere maintenance. *Nature,* 539**,** 54-58.
DIOTTI, R. & LOAYZA, D. 2011. Shelterin complex and associated factors at human telomeres. *Nucleus,* 2**,** 119-35.
DOMINSKI, Z., YANG, X. C. & MARZLUFF, W. F. 2005a. The polyadenylation factor CPSF-73 is involved in histone-pre-mRNA processing. *Cell,* 123**,** 37-48.
DOMINSKI, Z., YANG, X. C., PURDY, M., WAGNER, E. J. & MARZLUFF, W. F. 2005b. A CPSF-73 homologue is required for cell cycle progression but not cell growth and interacts with a protein having features of CPSF-100. *Mol Cell Biol,* 25**,** 1489-500.
DORLEANS, A., LI DE LA SIERRA-GALLAY, I., PITON, J., ZIG, L., GILET, L., PUTZER, H. & CONDON, C. 2011. Molecular basis for the recognition and cleavage of RNA by the bifunctional 5'-3' exo/endoribonuclease RNase J. *Structure,* 19**,** 1252-61.
DOUDNA, J. A. & CHARPENTIER, E. 2014. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science,* 346**,** 1258096.
DRABLOS, F., FEYZI, E., AAS, P. A., VAAGBO, C. B., KAVLI, B., BRATLIE, M. S., PENA-DIAZ, J., OTTERLEI, M., SLUPPHAUG, G. & KROKAN, H. E. 2004. Alkylation damage in DNA and RNA--repair mechanisms and medical significance. *DNA Repair (Amst),* 3**,** 1389-407.
DROLET, M., PHOENIX, P., MENZEL, R., MASSE, E., LIU, L. F. & CROUCH, R. J. 1995. Overexpression of RNase H partially complements the growth defect of an Escherichia coli delta topA mutant: R-loop formation is a major problem in the absence of DNA topoisomerase I. *Proc Natl Acad Sci U S A,* 92**,** 3526-30.
DUDLEY, D. D., CHAUDHURI, J., BASSING, C. H. & ALT, F. W. 2005. Mechanism and control of V(D)J recombination versus class switch recombination: similarities and differences. *Adv Immunol,* 86**,** 43-112.
DUQUETTE, M. L., ZHU, Q., TAYLOR, E. R., TSAY, A. J., SHI, L. Z., BERNS, M. W. & MCGOWAN, C. H. 2012. CtIP is required to initiate replication-dependent interstrand crosslink repair. *PLoS Genet,* 8**,** e1003050.
DUTTA, S., CHOWDHURY, G. & GATES, K. S. 2007. Interstrand cross-links generated by abasic sites in duplex DNA. *J Am Chem Soc,* 129**,** 1852-3.
DVIR, A., PETERSON, S. R., KNUTH, M. W., LU, H. & DYNAN, W. S. 1992. Ku autoantigen is the regulatory component of a template-associated protein kinase that phosphorylates RNA polymerase II. *Proc Natl Acad Sci U S A,* 89**,** 11920-4.
EASTMAN, A. & BARRY, M. A. 1987. Interaction of trans-diamminedichloroplatinum(II) with DNA: formation of monofunctional adducts and their reaction with glutathione. *Biochemistry,* 26**,** 3303-7.
ECKEL, L. M. & KRUGH, T. R. 1994. 2-Aminofluorene modified DNA duplex exists in two interchangeable conformations. *Nat Struct Biol,* 1**,** 89-94.
ENOIU, M., JIRICNY, J. & SCHARER, O. D. 2012. Repair of cisplatin-induced DNA interstrand crosslinks by a replication-independent pathway involving transcription-coupled repair and translesion synthesis. *Nucleic Acids Res,* 40**,** 8953-64.
ESASHI, F., GALKIN, V. E., YU, X., EGELMAN, E. H. & WEST, S. C. 2007. Stabilization of RAD51 nucleoprotein filaments by the C-terminal region of BRCA2. *Nat Struct Mol Biol,* 14**,** 468-74.
FAGBEMI, A. F., ORELLI, B. & SCHARER, O. D. 2011. Regulation of endonuclease activity in human nucleotide excision repair. *DNA Repair (Amst),* 10**,** 722-9.
FALNES, P. O., KLUNGLAND, A. & ALSETH, I. 2007. Repair of methyl lesions in DNA and RNA by oxidative demethylation. *Neuroscience,* 145**,** 1222-32.
FAN, Q., ZHANG, F., BARRETT, B., REN, K. & ANDREASSEN, P. R. 2009. A role for monoubiquitinated FANCD2 at telomeres in ALT cells. *Nucleic Acids Res,* 37**,** 1740-54.

FEKAIRI, S., SCAGLIONE, S., CHAHWAN, C., TAYLOR, E. R., TISSIER, A., COULON, S., DONG, M. Q., RUSE, C., YATES, J. R., 3RD, RUSSELL, P., FUCHS, R. P., MCGOWAN, C. H. & GAILLARD, P. H. 2009. Human SLX4 is a Holliday junction resolvase subunit that binds multiple DNA repair/recombination endonucleases. *Cell,* 138**,** 78-89.

FRANCISCO J FERNANDEZ, M. L.-E., AND M. CRISTINA VEGA 2011. *Nucleases of Metallo-Beta-Lactamase and Protein Phosphatase Families in DNA Repair*.

FRATTINI, A., FABBRI, M., VALLI, R., DE PAOLI, E., MONTALBANO, G., GRIBALDO, L., PASQUALI, F. & MASERATI, E. 2015. High variability of genomic instability and gene expression profiling in different HeLa clones. *Sci Rep,* 5**,** 15377.

FREIBAUM, B. D. & COUNTER, C. M. 2006. hSnm1B is a novel telomere-associated protein. *J Biol Chem,* 281**,** 15033-6.

FREIBAUM, B. D. & COUNTER, C. M. 2008. The protein hSnm1B is stabilized when bound to the telomere-binding protein TRF2. *J Biol Chem,* 283**,** 23671-6.

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

GARAU, G., GARCIA-SAEZ, I., BEBRONE, C., ANNE, C., MERCURI, P., GALLENI, M., FRERE, J. M. & DIDEBERG, O. 2004. Update of the standard numbering scheme for class B beta-lactamases. *Antimicrob Agents Chemother,* 48**,** 2347-9.

GARAYCOECHEA, J. I., CROSSAN, G. P., LANGEVIN, F., DALY, M., ARENDS, M. J. & PATEL, K. J. 2012. Genotoxic consequences of endogenous aldehydes on mouse haematopoietic stem cell function. *Nature,* 489**,** 571-5.

GARCES, F., FERNANDEZ, F. J., MONTELLA, C., PENYA-SOLER, E., PROHENS, R., AGUILAR, J., BALDOMA, L., COLL, M., BADIA, J. & VEGA, M. C. 2010. Molecular architecture of the Mn2+-dependent lactonase UlaG reveals an RNase-like metallo-beta-lactamase fold and a novel quaternary structure. *J Mol Biol,* 398**,** 715-29.

GARCES, F., PEARL, L. H. & OLIVER, A. W. 2011. The structural basis for substrate recognition by mammalian polynucleotide kinase 3' phosphatase. *Mol Cell,* 44**,** 385-96.

GARCIA, V., PHELPS, S. E., GRAY, S. & NEALE, M. J. 2011. Bidirectional resection of DNA double-strand breaks by Mre11 and Exo1. *Nature,* 479**,** 241-4.

GARI, K., DECAILLET, C., DELANNOY, M., WU, L. & CONSTANTINOU, A. 2008a. Remodeling of DNA replication structures by the branch point translocase FANCM. *Proc Natl Acad Sci U S A,* 105**,** 16107-12.

GARI, K., DECAILLET, C., STASIAK, A. Z., STASIAK, A. & CONSTANTINOU, A. 2008b. The Fanconi anemia protein FANCM can promote branch migration of Holliday junctions and replication forks. *Mol Cell,* 29**,** 141-8.

GINNO, P. A., LOTT, P. L., CHRISTENSEN, H. C., KORF, I. & CHEDIN, F. 2012. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol Cell,* 45**,** 814-25.

GOMES, C. M., FRAZAO, C., XAVIER, A. V., LEGALL, J. & TEIXEIRA, M. 2002. Functional control of the binuclear metal site in the metallo-beta-lactamase-like fold by subtle amino acid replacements. *Protein Sci,* 11**,** 707-12.

GONG, Y., ZHU, D., DING, J., DOU, C. N., REN, X., GU, L., JIANG, T. & WANG, D. C. 2011. Crystal structures of aprataxin ortholog Hnt3 reveal the mechanism for reversal of 5'-adenylated DNA. *Nat Struct Mol Biol,* 18**,** 1297-9.

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

GRAWUNDER, U., WILM, M., WU, X., KULESZA, P., WILSON, T. E., MANN, M. & LIEBER, M. R. 1997. Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. *Nature,* 388**,** 492-5.

GRAY, R. D. & CHAIRES, J. B. 2008. Kinetics and mechanism of K+- and Na+-induced folding of models of human telomeric DNA into G-quadruplex structures. *Nucleic Acids Res,* 36**,** 4191-203.

GRUNDY, G. J., RULTEN, S. L., ZENG, Z., ARRIBAS-BOSACOMA, R., ILES, N., MANLEY, K., OLIVER, A. & CALDECOTT, K. W. 2013. APLF promotes the assembly and activity of non-homologous end joining protein complexes. *EMBO J,* 32**,** 112-25.

GUO, C., FISCHHABER, P. L., LUK-PASZYC, M. J., MASUDA, Y., ZHOU, J., KAMIYA, K., KISKER, C. & FRIEDBERG, E. C. 2003. Mouse Rev1 protein interacts with multiple DNA polymerases involved in translesion DNA synthesis. *EMBO J,* 22**,** 6621-30.

GUO, Z., KUMAGAI, A., WANG, S. X. & DUNPHY, W. G. 2000. Requirement for Atr in phosphorylation of Chk1 and cell cycle regulation in response to DNA replication blocks and UV-damaged DNA in Xenopus egg extracts. *Genes Dev,* 14**,** 2745-56.

HAMMEL, M., REY, M., YU, Y., MANI, R. S., CLASSEN, S., LIU, M., PIQUE, M. E., FANG, S., MAHANEY, B. L., WEINFELD, M., SCHRIEMER, D. C., LEES-MILLER, S. P. & TAINER, J. A. 2011. 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.

HAMMONS, G. J., MILTON, D., STEPPS, K., GUENGERICH, F. P., TUKEY, R. H. & KADLUBAR, F. F. 1997. Metabolism of carcinogenic heterocyclic and aromatic amines by recombinant human cytochrome P450 enzymes. *Carcinogenesis,* 18**,** 851-4.

HANADA, K., BUDZOWSKA, M., DAVIES, S. L., VAN DRUNEN, E., ONIZAWA, H., BEVERLOO, H. B., MAAS, A., ESSERS, J., HICKSON, I. D. & KANAAR, R. 2007. The structure-specific endonuclease Mus81 contributes to replication restart by generating double-strand DNA breaks. *Nat Struct Mol Biol,* 14**,** 1096-104.

HANADA, K., BUDZOWSKA, M., MODESTI, M., MAAS, A., WYMAN, C., ESSERS, J. & KANAAR, R. 2006. The structure-specific endonuclease Mus81-Eme1 promotes conversion of interstrand DNA crosslinks into double-strands breaks. *EMBO J,* 25**,** 4921-32.

HARPER, J. W. & ELLEDGE, S. J. 2007. The DNA damage response: ten years after. *Mol Cell,* 28**,** 739-45.

HARTLEY, J. A. 2011. The development of pyrrolobenzodiazepines as antitumour agents. *Expert Opin Investig Drugs,* 20**,** 733-44.

HASSELL, A. M., AN, G., BLEDSOE, R. K., BYNUM, J. M., CARTER, H. L., 3RD, DENG, S. J., GAMPE, R. T., GRISARD, T. E., MADAUSS, K. P., NOLTE, R. T., ROCQUE, W. J., WANG, L., WEAVER, K. L., WILLIAMS, S. P., WISELY, G. B., XU, R. & SHEWCHUK, L. M. 2007. Crystallization of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr,* 63**,** 72-9.

HAZRATI, A., RAMIS-CASTELLTORT, M., SARKAR, S., BARBER, L. J., SCHOFIELD, C. J., HARTLEY, J. A. & MCHUGH, P. J. 2008. Human SNM1A suppresses the DNA repair defects of yeast pso2 mutants. *DNA Repair (Amst),* 7**,** 230-8.

HECHT, S. S. 1999. DNA adduct formation from tobacco-specific N-nitrosamines. *Mutat Res,* 424**,** 127-42.

HEIDENREICH, E., NOVOTNY, R., KNEIDINGER, B., HOLZMANN, V. & WINTERSBERGER, U. 2003. Non-homologous end joining as an important mutagenic process in cell cycle-arrested cells. *EMBO J,* 22**,** 2274-83.

HENLE, E. S. & LINN, S. 1997. Formation, prevention, and repair of DNA damage by iron/hydrogen peroxide. *J Biol Chem,* 272**,** 19095-8.

HENNER, W. D., GRUNBERG, S. M. & HASELTINE, W. A. 1982. Sites and structure of gamma radiation-induced DNA strand breaks. *J Biol Chem,* 257**,** 11750-4.

HENRIQUES, J. A. & MOUSTACCHI, E. 1980. Isolation and characterization of pso mutants sensitive to photo-addition of psoralen derivatives in Saccharomyces cerevisiae. *Genetics,* 95**,** 273-88.

HENRIQUES, J. A. & MOUSTACCHI, E. 1981. Interactions between mutations for sensitivity to psoralen photoaddition (pso) and to radiation (rad) in Saccharomyces cerevisiae. *J Bacteriol,* 148**,** 248-56.

HERTZBERG, R. P., CARANFA, M. J. & HECHT, S. M. 1989. On the mechanism of topoisomerase I inhibition by camptothecin: evidence for binding to an enzyme-DNA complex. *Biochemistry,* 28**,** 4629-38.
HEYDUK, T., MA, Y., TANG, H. & EBRIGHT, R. H. 1996. Fluorescence anisotropy: rapid, quantitative assay for protein-DNA and protein-protein interaction. *Methods Enzymol,* 274**,** 492-503.
HIRA, A., YABE, H., YOSHIDA, K., OKUNO, Y., SHIRAISHI, Y., CHIBA, K., TANAKA, H., MIYANO, S., NAKAMURA, J., KOJIMA, S., OGAWA, S., MATSUO, K., TAKATA, M. & YABE, M. 2013. Variant ALDH2 is associated with accelerated progression of bone marrow failure in Japanese Fanconi anemia patients. *Blood,* 122**,** 3206-9.
HO, C. K., MAZON, G., LAM, A. F. & SYMINGTON, L. S. 2010. Mus81 and Yen1 promote reciprocal exchange during mitotic recombination to maintain genome integrity in budding yeast. *Mol Cell,* 40**,** 988-1000.
HO, T. V., GUAINAZZI, A., DERKUNT, S. B., ENOIU, M. & SCHARER, O. D. 2011. Structure-dependent bypass of DNA interstrand crosslinks by translesion synthesis polymerases. *Nucleic Acids Res,* 39**,** 7455-64.
HOEIJMAKERS, J. H. 2001. Genome maintenance mechanisms for preventing cancer. *Nature,* 411**,** 366-74.
HOFFMAN, E. A., FREY, B. L., SMITH, L. M. & AUBLE, D. T. 2015. Formaldehyde crosslinking: a tool for the study of chromatin complexes. *J Biol Chem,* 290**,** 26404-11.
HOLLIDAY, R. & HO, T. 1998. Gene silencing and endogenous DNA methylation in mammalian cells. *Mutat Res,* 400**,** 361-8.
HOLLIS, T. 2007. Crystallization of protein-DNA complexes. *Methods Mol Biol,* 363**,** 225-37.
HOPTON, S. R. & THOMPSON, A. S. 2011. Nuclear magnetic resonance solution structures of inter- and intrastrand adducts of DNA cross-linker SJG-136. *Biochemistry,* 50**,** 4720-32.
HUANG, J., LIU, S., BELLANI, M. A., THAZHATHVEETIL, A. K., LING, C., DE WINTER, J. P., WANG, Y., WANG, W. & SEIDMAN, M. M. 2013. The DNA translocase FANCM/MHF promotes replication traverse of DNA interstrand crosslinks. *Mol Cell,* 52**,** 434-46.
HUANG, M. & D'ANDREA, A. D. 2010. A new nuclease member of the FAN club. *Nat Struct Mol Biol,* 17**,** 926-8.
HUERTAS, P. & JACKSON, S. P. 2009. Human CtIP mediates cell cycle control of DNA end resection and double strand break repair. *J Biol Chem,* 284**,** 9558-65.
HUTCHINSON, F. 1985. Chemical changes induced in DNA by ionizing radiation. *Prog Nucleic Acid Res Mol Biol,* 32**,** 115-54.
IHARA, T., OHURA, H., SHIRAHAMA, C., FURUZONO, T., SHIMADA, H., MATSUURA, H. & KITAMURA, Y. 2015. Metal ion-directed dynamic splicing of DNA through global conformational change by intramolecular complexation. *Nat Commun,* 6**,** 6640.
INTERTHAL, H. & CHAMPOUX, J. J. 2011. Effects of DNA and protein size on substrate cleavage by human tyrosyl-DNA phosphodiesterase 1. *Biochem J,* 436**,** 559-66.
ISHIAI, M., KIMURA, M., NAMIKOSHI, K., YAMAZOE, M., YAMAMOTO, K., ARAKAWA, H., AGEMATSU, K., MATSUSHITA, N., TAKEDA, S., BUERSTEDDE, J. M. & TAKATA, M. 2004. DNA cross-link repair protein SNM1A interacts with PIAS1 in nuclear focus formation. *Mol Cell Biol,* 24**,** 10733-41.
ISONO, M., NIIMI, A., OIKE, T., HAGIWARA, Y., SATO, H., SEKINE, R., YOSHIDA, Y., ISOBE, S. Y., OBUSE, C., NISHI, R., PETRICCI, E., NAKADA, S., NAKANO, T. & SHIBATA, A. 2017. BRCA1 Directs the Repair Pathway to Homologous Recombination by Promoting 53BP1 Dephosphorylation. *Cell Rep,* 18**,** 520-532.
IYAMA, T., LEE, S. Y., BERQUIST, B. R., GILEADI, O., BOHR, V. A., SEIDMAN, M. M., MCHUGH, P. J. & WILSON, D. M., 3RD 2015. CSB interacts with SNM1A and promotes DNA interstrand crosslink processing. *Nucleic Acids Res,* 43**,** 247-58.

JACKSON, S. P. & BARTEK, J. 2009. The DNA-damage response in human biology and disease. *Nature,* 461**,** 1071-8.
JASIN, M. 2002. Homologous repair of DNA damage and tumorigenesis: the BRCA connection. *Oncogene,* 21**,** 8981-93.
JASIN, M. & ROTHSTEIN, R. 2013. Repair of strand breaks by homologous recombination. *Cold Spring Harb Perspect Biol,* 5**,** a012740.
JAZAYERI, A., BALESTRINI, A., GARNER, E., HABER, J. E. & COSTANZO, V. 2008. Mre11-Rad50-Nbs1-dependent processing of DNA breaks generates oligonucleotides that stimulate ATM activity. *EMBO J,* 27**,** 1953-62.
JAZAYERI, A., FALCK, J., LUKAS, C., BARTEK, J., SMITH, G. C., LUKAS, J. & JACKSON, S. P. 2006. ATM- and cell cycle-dependent regulation of ATR in response to DNA double-strand breaks. *Nat Cell Biol,* 8**,** 37-45.
JINEK, M., CHYLINSKI, K., FONFARA, I., HAUER, M., DOUDNA, J. A. & CHARPENTIER, E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science,* 337**,** 816-21.
JIRICNY, J. 2006. The multifaceted mismatch-repair system. *Nat Rev Mol Cell Biol,* 7**,** 335-46.
KAINA, B., CHRISTMANN, M., NAUMANN, S. & ROOS, W. P. 2007. MGMT: key node in the battle against genotoxicity, carcinogenicity and apoptosis induced by alkylating agents. *DNA Repair (Amst),* 6**,** 1079-99.
KAMATH-LOEB, A., LOEB, L. A. & FRY, M. 2012. The Werner syndrome protein is distinguished from the Bloom syndrome protein by its capacity to tightly bind diverse DNA structures. *PLoS One,* 7**,** e30189.
KANG, M. H. 2011. Cancel all Hollidays for SLX4 mutations: identification of a new Fanconi anemia subtype, FANCP. *Clin Genet,* 80**,** 28-30.
KARAMI, S., HAN, Y., PANDE, M., CHENG, I., RUDD, J., PIERCE, B. L., NUTTER, E. L., SCHUMACHER, F. R., KOTE-JARAI, Z., LINDSTROM, S., WITTE, J. S., FANG, S., HAN, J., KRAFT, P., HUNTER, D. J., SONG, F., HUNG, R. J., MCKAY, J., GRUBER, S. B., CHANOCK, S. J., RISCH, A., SHEN, H., HAIMAN, C. A., BOARDMAN, L., ULRICH, C. M., CASEY, G., PETERS, U., AMIN AL OLAMA, A., BERCHUCK, A., BERNDT, S. I., BEZIEAU, S., BRENNAN, P., BRENNER, H., BRINTON, L., CAPORASO, N., CHAN, A. T., CHANG-CLAUDE, J., CHRISTIANI, D. C., CUNNINGHAM, J. M., EASTON, D., EELES, R. A., EISEN, T., GALA, M., GALLINGER, S. J., GAYTHER, S. A., GOODE, E. L., GRONBERG, H., HENDERSON, B. E., HOULSTON, R., JOSHI, A. D., KURY, S., LANDI, M. T., LE MARCHAND, L., MUIR, K., NEWCOMB, P. A., PERMUTH-WEY, J., PHAROAH, P., PHELAN, C., POTTER, J. D., RAMUS, S. J., RISCH, H., SCHILDKRAUT, J., SLATTERY, M. L., SONG, H., WENTZENSEN, N., WHITE, E., WIKLUND, F., ZANKE, B. W., SELLERS, T. A., ZHENG, W., CHATTERJEE, N., AMOS, C. I., DOHERTY, J. A., GECCO, THE GAME-ON NETWORK: CORECT, D. E. F. & TRICL 2016. Telomere structure and maintenance gene variants and risk of five cancer types. *Int J Cancer,* 139**,** 2655-2670.
KARTALOU, M. & ESSIGMANN, J. M. 2001. Recognition of cisplatin adducts by cellular proteins. *Mutat Res,* 478**,** 1-21.
KASAI, H. & NISHIMURA, S. 1984. Hydroxylation of deoxyguanosine at the C-8 position by ascorbic acid and other reducing agents. *Nucleic Acids Res,* 12**,** 2137-45.
KASIOTIS, K. M., KYRIAKOPOULOU, K., EMMANOUIL, C., TSANTILA, N., LIESIVUORI, J., SOUKI, H., MANAKIS, S. & MACHERA, K. 2012. Monitoring of systemic exposure to plant protection products and DNA damage in orchard workers. *Toxicol Lett,* 210**,** 182-8.
KAWAMOTO, T., ARAKI, K., SONODA, E., YAMASHITA, Y. M., HARADA, K., KIKUCHI, K., MASUTANI, C., HANAOKA, F., NOZAKI, K., HASHIMOTO, N. & TAKEDA, S. 2005. Dual roles for DNA polymerase eta in homologous DNA recombination and translesion DNA synthesis. *Mol Cell,* 20**,** 793-9.

KAWANISHI, M., MATSUMOTO, T. & TAKASHI, Y. 2014. Genotoxicity of formaldehyde: molecular basis of DNA damage and mutation. *Frontiers in Environmental Science* 2**,** 1-8.

KEENEY, S. 2008. Spo11 and the Formation of DNA Double-Strand Breaks in Meiosis. *Genome Dyn Stab,* 2**,** 81-123.

KEENEY, S. & NEALE, M. J. 2006. Initiation of meiotic recombination by formation of DNA double-strand breaks: mechanism and regulation. *Biochem Soc Trans,* 34**,** 523-5.

KENT, T., CHANDRAMOULY, G., MCDEVITT, S. M., OZDEMIR, A. Y. & POMERANTZ, R. T. 2015. Mechanism of microhomology-mediated end-joining promoted by human DNA polymerase theta. *Nat Struct Mol Biol,* 22**,** 230-7.

KIM, Y. 2014. Nuclease delivery: versatile functions of SLX4/FANCP in genome maintenance. *Mol Cells,* 37**,** 569-74.

KLEIN DOUWEL, D., BOONEN, R. A., LONG, D. T., SZYPOWSKA, A. A., RASCHLE, M., WALTER, J. C. & KNIPSCHEER, P. 2014. XPF-ERCC1 acts in Unhooking DNA interstrand crosslinks in cooperation with FANCD2 and FANCP/SLX4. *Mol Cell,* 54**,** 460-71.

KLEIN DOUWEL, D., HOOGENBOOM, W. S., BOONEN, R. A. & KNIPSCHEER, P. 2017. Recruitment and positioning determine the specific role of the XPF-ERCC1 endonuclease in interstrand crosslink repair. *EMBO J,* 36**,** 2034-2046.

KNIPSCHEER, P., RASCHLE, M., SMOGORZEWSKA, A., ENOIU, M., HO, T. V., SCHARER, O. D., ELLEDGE, S. J. & WALTER, J. C. 2009. The Fanconi anemia pathway promotes replication-dependent DNA interstrand cross-link repair. *Science,* 326**,** 1698-701.

KOTECKI, M., REDDY, P. S. & COCHRAN, B. H. 1999. Isolation and characterization of a near-haploid human cell line. *Exp Cell Res,* 252**,** 273-80.

KOTTEMANN, M. C. & SMOGORZEWSKA, A. 2013. Fanconi anaemia and the repair of Watson and Crick DNA crosslinks. *Nature,* 493**,** 356-63.

KOZEKOV, I. D., NECHEV, L. V., MOSELEY, M. S., HARRIS, C. M., RIZZO, C. J., STONE, M. P. & HARRIS, T. M. 2003. DNA interchain cross-links formed by acrolein and crotonaldehyde. *J Am Chem Soc,* 125**,** 50-61.

KRATZ, K., SCHOPF, B., KADEN, S., SENDOEL, A., EBERHARD, R., LADEMANN, C., CANNAVO, E., SARTORI, A. A., HENGARTNER, M. O. & JIRICNY, J. 2010. Deficiency of FANCD2-associated nuclease KIAA1018/FAN1 sensitizes cells to interstrand crosslinking agents. *Cell,* 142**,** 77-88.

KULKARNI, A. & WILSON, D. M., 3RD 2008. The involvement of DNA-damage and -repair defects in neurological dysfunction. *Am J Hum Genet,* 82**,** 539-66.

KUMAGAI, A., LEE, J., YOO, H. Y. & DUNPHY, W. G. 2006. TopBP1 activates the ATR-ATRIP complex. *Cell,* 124**,** 943-55.

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

KUROSAWA, A. & ADACHI, N. 2010. Functions and regulation of Artemis: a goddess in the maintenance of genome integrity. *J Radiat Res,* 51**,** 503-9.

KUSUMOTO, R., DAWUT, L., MARCHETTI, C., WAN LEE, J., VINDIGNI, A., RAMSDEN, D. & BOHR, V. A. 2008. Werner protein cooperates with the XRCC4-DNA ligase IV complex in end-processing. *Biochemistry,* 47**,** 7548-56.

LACHAUD, C., MORENO, A., MARCHESI, F., TOTH, R., BLOW, J. J. & ROUSE, J. 2016. Ubiquitinated Fancd2 recruits Fan1 to stalled replication forks to prevent genome instability. *Science,* 351**,** 846-9.

LAM, Y. C., AKHTER, S., GU, P., YE, J., POULET, A., GIRAUD-PANIS, M. J., BAILEY, S. M., GILSON, E., LEGERSKI, R. J. & CHANG, S. 2010. SNMIB/Apollo protects leading-strand telomeres against NHEJ-mediated repair. *EMBO J,* 29**,** 2230-41.

LAMARCHE, B. J., ORAZIO, N. I. & WEITZMAN, M. D. 2010. The MRN complex in double-strand break repair and telomere maintenance. *FEBS Lett,* 584**,** 3682-95.

LANDRY, J. J., PYL, P. T., RAUSCH, T., ZICHNER, T., TEKKEDIL, M. M., STUTZ, A. M., JAUCH, A., AIYAR, R. S., PAU, G., DELHOMME, N., GAGNEUR, J., KORBEL, J. O., HUBER, W. & STEINMETZ, L. M. 2013. The genomic and transcriptomic landscape of a HeLa cell line. *G3 (Bethesda),* 3**,** 1213-24.

LANE, A. N., CHAIRES, J. B., GRAY, R. D. & TRENT, J. O. 2008. Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res,* 36**,** 5482-515.

LANGEVIN, F., CROSSAN, G. P., ROSADO, I. V., ARENDS, M. J. & PATEL, K. J. 2011. Fancd2 counteracts the toxic effects of naturally produced aldehydes in mice. *Nature,* 475**,** 53-8.

LARSON, K., SAHM, J., SHENKAR, R. & STRAUSS, B. 1985. Methylation-induced blocks to in vitro DNA replication. *Mutat Res,* 150**,** 77-84.

LEE, S. Y. 2015. *Structure and Inhibition of the Human SNM1A and SNM1B/Apollo DNA Repair Exonuclease.* Doctor of Philosophy University of Oxford.

LEE, S. Y., BREM, J., PETTINATI, I., CLARIDGE, T. D., GILEADI, O., SCHOFIELD, C. J. & MCHUGH, P. J. 2016. Cephalosporins inhibit human metallo beta-lactamase fold DNA repair nucleases SNM1A and SNM1B/apollo. *Chem Commun (Camb),* 52**,** 6727-30.

LEHOCZKY, P., MCHUGH, P. J. & CHOVANEC, M. 2007. DNA interstrand cross-link repair in Saccharomyces cerevisiae. *FEMS Microbiol Rev,* 31**,** 109-33.

LEIMGRUBER, W., STEFANOVIC, V., SCHENKER, F., KARR, A. & BERGER, J. 1965. Isolation and characterization of anthramycin, a new antitumor antibiotic. *J Am Chem Soc,* 87**,** 5791-3.

LENAIN, C., BAUWENS, S., AMIARD, S., BRUNORI, M., GIRAUD-PANIS, M. J. & GILSON, E. 2006. The Apollo 5' exonuclease functions together with TRF2 to protect telomeres from DNA repair. *Curr Biol,* 16**,** 1303-10.

LI, B., REDDY, S. & COMAI, L. 2009. Sequence-specific processing of telomeric 3' overhangs by the Werner syndrome protein exonuclease activity. *Aging (Albany NY),* 1**,** 289-302.

LI DE LA SIERRA-GALLAY, I., ZIG, L., JAMALLI, A. & PUTZER, H. 2008. Structural insights into the dual activity of RNase J. *Nat Struct Mol Biol,* 15**,** 206-12.

LI, J., HOLZSCHU, D. L. & SUGIYAMA, T. 2013. PCNA is efficiently loaded on the DNA recombination intermediate to modulate polymerase delta, eta, and zeta activities. *Proc Natl Acad Sci U S A,* 110**,** 7672-7.

LI, S., KANNO, S., WATANABE, R., OGIWARA, H., KOHNO, T., WATANABE, G., YASUI, A. & LIEBER, M. R. 2011. Polynucleotide kinase and aprataxin-like forkhead-associated protein (PALF) acts as both a single-stranded DNA endonuclease and a single-stranded DNA 3' exonuclease and can participate in DNA end joining in a biochemical system. *J Biol Chem,* 286**,** 36368-77.

LI, X., HEJNA, J. & MOSES, R. E. 2005. The yeast Snm1 protein is a DNA 5'-exonuclease. *DNA Repair (Amst),* 4**,** 163-70.

LI, X. & HEYER, W. D. 2008. Homologous recombination in DNA repair and DNA damage tolerance. *Cell Res,* 18**,** 99-113.

LI, X. & MANLEY, J. L. 2005. Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell,* 122**,** 365-78.

LI, X. & MANLEY, J. L. 2006. Cotranscriptional processes and their influence on genome stability. *Genes Dev,* 20**,** 1838-47.

LIANG, L., DENG, L., NGUYEN, S. C., ZHAO, X., MAULION, C. D., SHAO, C. & TISCHFIELD, J. A. 2008. 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.

LIANG, X. S., PFEIFFER, R. M., WHEELER, W., MAEDER, D., BURDETTE, L., YEAGER, M., CHANOCK, S., TUCKER, M. A., GOLDSTEIN, A. M. & YANG, X. R. 2012. Genetic variants in DNA repair genes and the risk of cutaneous malignant melanoma in melanoma-prone families with/without CDKN2A mutations. *Int J Cancer,* 130**,** 2062-6.

LINDAHL, T. 1993. Instability and decay of the primary structure of DNA. *Nature,* 362**,** 709-15.
LINDAHL, T. & KARLSTROM, O. 1973. Heat-induced depyrimidination of deoxyribonucleic acid in neutral solution. *Biochemistry,* 12**,** 5151-4.
LING, C., HUANG, J., YAN, Z., LI, Y., OHZEKI, M., ISHIAI, M., XU, D., TAKATA, M., SEIDMAN, M. & WANG, W. 2016. Bloom syndrome complex promotes FANCM recruitment to stalled replication forks and facilitates both repair and traverse of DNA interstrand crosslinks. *Cell Discov,* 2**,** 16047.
LIU, L., AKHTER, S., BAE, J. B., MUKHOPADHYAY, S. S., RICHIE, C. T., LIU, X. & LEGERSKI, R. 2009. SNM1B/Apollo interacts with astrin and is required for the prophase cell cycle checkpoint. *Cell Cycle,* 8**,** 628-38.
LIU, Q., GUNTUKU, S., CUI, X. S., MATSUOKA, S., CORTEZ, D., TAMAI, K., LUO, G., CARATTINI-RIVERA, S., DEMAYO, F., BRADLEY, A., DONEHOWER, L. A. & ELLEDGE, S. J. 2000. Chk1 is an essential kinase that is regulated by Atr and required for the G(2)/M DNA damage checkpoint. *Genes Dev,* 14**,** 1448-59.
LIU, T., GHOSAL, G., YUAN, J., CHEN, J. & HUANG, J. 2010. FAN1 acts with FANCI-FANCD2 to promote DNA interstrand cross-link repair. *Science,* 329**,** 693-6.
LOEB, L. A. & PRESTON, B. D. 1986. Mutagenesis by apurinic/apyrimidinic sites. *Annu Rev Genet,* 20**,** 201-30.
LOECHLER, E. L., GREEN, C. L. & ESSIGMANN, J. M. 1984. In vivo mutagenesis by O6-methylguanine built into a unique site in a viral genome. *Proc Natl Acad Sci U S A,* 81**,** 6271-5.
LONG, D. T., JOUKOV, V., BUDZOWSKA, M. & WALTER, J. C. 2014. BRCA1 promotes unloading of the CMG helicase from a stalled DNA replication fork. *Mol Cell,* 56**,** 174-85.
LONG, D. T., RASCHLE, M., JOUKOV, V. & WALTER, J. C. 2011. Mechanism of RAD51-dependent DNA interstrand cross-link repair. *Science,* 333**,** 84-7.
LOPEZ-MOSQUEDA, J., MADDI, K., PRGOMET, S., KALAYIL, S., MARINOVIC-TERZIC, I., TERZIC, J. & DIKIC, I. 2016. SPRTN is a mammalian DNA-binding metalloprotease that resolves DNA-protein crosslinks. *Elife,* 5.
LOVRIEN, R. E. & MATULIS, D. 2001. Selective precipitation of proteins. *Curr Protoc Protein Sci,* Chapter 4**,** Unit 4 5.
LUKAS, J., LUKAS, C. & BARTEK, J. 2011. More than just a focus: The chromatin response to DNA damage and its role in genome integrity maintenance. *Nat Cell Biol,* 13**,** 1161-9.
LYONS, T. J. A. E., D. J. 2006. Transport and storage of metal ions in biology. . *In:* BERTINI, I., GRAY, H. B., STIEFEL, E. I., VALENTINE, J. S. (ed.) *Biological Inorganic Chemistry: Structure and Reactivity* Herndon, VA: University Science Books.
MA, Y., PANNICKE, U., SCHWARZ, K. & LIEBER, M. R. 2002. Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell,* 108**,** 781-94.
MA, Y., SCHWARZ, K. & LIEBER, M. R. 2005. The Artemis:DNA-PKcs endonuclease cleaves DNA loops, flaps, and gaps. *DNA Repair (Amst),* 4**,** 845-51.
MACKAY, C., DECLAIS, A. C., LUNDIN, C., AGOSTINHO, A., DEANS, A. J., MACARTNEY, T. J., HOFMANN, K., GARTNER, A., WEST, S. C., HELLEDAY, T., LILLEY, D. M. & ROUSE, J. 2010. Identification of KIAA1018/FAN1, a DNA repair nuclease recruited to DNA damage by monoubiquitinated FANCD2. *Cell,* 142**,** 65-76.
MAESTRONI, L., MATMATI, S. & COULON, S. 2017. Solving the Telomere Replication Problem. *Genes (Basel),* 8.
MAGANA-SCHWENCKE, N., HENRIQUES, J. A., CHANET, R. & MOUSTACCHI, E. 1982. The fate of 8-methoxypsoralen photoinduced crosslinks in nuclear and mitochondrial yeast DNA: comparison of wild-type and repair-deficient strains. *Proc Natl Acad Sci U S A,* 79**,** 1722-6.

MALINGE, J. M., PEREZ, C. & LENG, M. 1994. Base sequence-independent distorsions induced by interstrand cross-links in cis-diamminedichloroplatinum (II)-modified DNA. *Nucleic Acids Res,* 22**,** 3834-9.
MALIVERT, L., ROPARS, V., NUNEZ, M., DREVET, P., MIRON, S., FAURE, G., GUEROIS, R., MORNON, J. P., REVY, P., CHARBONNIER, J. B., CALLEBAUT, I. & DE VILLARTAY, J. P. 2010. Delineation of the Xrcc4-interacting region in the globular head domain of cernunnos/XLF. *J Biol Chem,* 285**,** 26475-83.
MANDEL, C. R., KANEKO, S., ZHANG, H., GEBAUER, D., VETHANTHAM, V., MANLEY, J. L. & TONG, L. 2006. Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature,* 444**,** 953-6.
MAO, Z., SELUANOV, A., JIANG, Y. & GORBUNOVA, V. 2007. TRF2 is required for repair of nontelomeric DNA double-strand breaks by homologous recombination. *Proc Natl Acad Sci U S A,* 104**,** 13068-73.
MARECHAL, A. & ZOU, L. 2013. DNA damage sensing by the ATM and ATR kinases. *Cold Spring Harb Perspect Biol,* 5.
MARI, P. O., FLOREA, B. I., PERSENGIEV, S. P., VERKAIK, N. S., BRUGGENWIRTH, H. T., MODESTI, M., GIGLIA-MARI, G., BEZSTAROSTI, K., DEMMERS, J. A., LUIDER, T. M., HOUTSMULLER, A. B. & VAN GENT, D. C. 2006. Dynamic assembly of end-joining complexes requires interaction between Ku70/80 and XRCC4. *Proc Natl Acad Sci U S A,* 103**,** 18597-602.
MARNETT, L. J. 1999. Lipid peroxidation-DNA damage by malondialdehyde. *Mutat Res,* 424**,** 83-95.
MASON, J. M., DAS, I., ARLT, M., PATEL, N., KRAFTSON, S., GLOVER, T. W. & SEKIGUCHI, J. M. 2013. The SNM1B/APOLLO DNA nuclease functions in resolution of replication stress and maintenance of common fragile site stability. *Hum Mol Genet,* 22**,** 4901-13.
MASON, J. M. & SEKIGUCHI, J. M. 2011. Snm1B/Apollo functions in the Fanconi anemia pathway in response to DNA interstrand crosslinks. *Hum Mol Genet,* 20**,** 2549-59.
MATOS, J., BLANCO, M. G., MASLEN, S., SKEHEL, J. M. & WEST, S. C. 2011. Regulatory control of the resolution of DNA recombination intermediates during meiosis and mitosis. *Cell,* 147**,** 158-72.
MATSUOKA, S., BALLIF, B. A., SMOGORZEWSKA, A., MCDONALD, E. R., 3RD, HUROV, K. E., LUO, J., BAKALARSKI, C. E., ZHAO, Z., SOLIMINI, N., LERENTHAL, Y., SHILOH, Y., GYGI, S. P. & ELLEDGE, S. J. 2007. ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science,* 316**,** 1160-6.
MATSUOKA, S., HUANG, M. & ELLEDGE, S. J. 1998. Linkage of ATM to cell cycle regulation by the Chk2 protein kinase. *Science,* 282**,** 1893-7.
MATSUZAKI, K., BOREL, V., ADELMAN, C. A., SCHINDLER, D. & BOULTON, S. J. 2015. FANCJ suppresses microsatellite instability and lymphomagenesis independent of the Fanconi anemia pathway. *Genes Dev,* 29**,** 2532-46.
MATULIS, D. 2016. Selective Precipitation of Proteins. *Curr Protoc Protein Sci,* 83**,** 4 5 1-4 5 37.
MAZIN, A. V., MAZINA, O. M., BUGREEV, D. V. & ROSSI, M. J. 2010. Rad54, the motor of homologous recombination. *DNA Repair (Amst),* 9**,** 286-302.
MCCULLOCH, S. D. & KUNKEL, T. A. 2008. The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Res,* 18**,** 148-61.
MEDEMA, R. H. 2004. Optimizing RNA interference for application in mammalian cells. *Biochem J,* 380**,** 593-603.
MICHAILIDOU, K., HALL, P., GONZALEZ-NEIRA, A., GHOUSSAINI, M., DENNIS, J., MILNE, R. L., SCHMIDT, M. K., CHANG-CLAUDE, J., BOJESEN, S. E., BOLLA, M. K., WANG, Q., DICKS, E., LEE, A., TURNBULL, C., RAHMAN, N., BREAST, OVARIAN CANCER SUSCEPTIBILITY, C., FLETCHER, O., PETO, J., GIBSON, L., DOS SANTOS SILVA, I., NEVANLINNA, H., MURANEN, T. A., AITTOMAKI, K., BLOMQVIST, C., CZENE, K., IRWANTO, A., LIU, J., WAISFISZ, Q.,

MEIJERS-HEIJBOER, H., ADANK, M., HEREDITARY, B., OVARIAN CANCER RESEARCH GROUP, N., VAN DER LUIJT, R. B., HEIN, R., DAHMEN, N., BECKMAN, L., MEINDL, A., SCHMUTZLER, R. K., MULLER-MYHSOK, B., LICHTNER, P., HOPPER, J. L., SOUTHEY, M. C., MAKALIC, E., SCHMIDT, D. F., UITTERLINDEN, A. G., HOFMAN, A., HUNTER, D. J., CHANOCK, S. J., VINCENT, D., BACOT, F., TESSIER, D. C., CANISIUS, S., WESSELS, L. F., HAIMAN, C. A., SHAH, M., LUBEN, R., BROWN, J., LUCCARINI, C., SCHOOF, N., HUMPHREYS, K., LI, J., NORDESTGAARD, B. G., NIELSEN, S. F., FLYGER, H., COUCH, F. J., WANG, X., VACHON, C., STEVENS, K. N., LAMBRECHTS, D., MOISSE, M., PARIDAENS, R., CHRISTIAENS, M. R., RUDOLPH, A., NICKELS, S., FLESCH-JANYS, D., JOHNSON, N., AITKEN, Z., AALTONEN, K., HEIKKINEN, T., BROEKS, A., VEER, L. J., VAN DER SCHOOT, C. E., GUENEL, P., TRUONG, T., LAURENT-PUIG, P., MENEGAUX, F., MARME, F., SCHNEEWEISS, A., SOHN, C., BURWINKEL, B., ZAMORA, M. P., PEREZ, J. I., PITA, G., ALONSO, M. R., COX, A., BROCK, I. W., CROSS, S. S., REED, M. W., SAWYER, E. J., et al. 2013. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet,* 45**,** 353-61, 361e1-2.

MICHL, J., ZIMMER, J. & TARSOUNAS, M. 2016. Interplay between Fanconi anemia and homologous recombination pathways in genome integrity. *EMBO J,* 35**,** 909-23.

MIMITOU, E. P. & SYMINGTON, L. S. 2008. Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing. *Nature,* 455**,** 770-4.

MIMITOU, E. P. & SYMINGTON, L. S. 2009. Nucleases and helicases take center stage in homologous recombination. *Trends Biochem Sci,* 34**,** 264-72.

MIMITOU, E. P. & SYMINGTON, L. S. 2010. Ku prevents Exo1 and Sgs1-dependent resection of DNA ends in the absence of a functional MRX complex or Sae2. *EMBO J,* 29**,** 3358-69.

MIMITOU, E. P. & SYMINGTON, L. S. 2011. DNA end resection--unraveling the tail. *DNA Repair (Amst),* 10**,** 344-8.

MINKO, I. G., HARBUT, M. B., KOZEKOV, I. D., KOZEKOVA, A., JAKOBS, P. M., OLSON, S. B., MOSES, R. E., HARRIS, T. M., RIZZO, C. J. & LLOYD, R. S. 2008. Role for DNA polymerase kappa in the processing of N2-N2-guanine interstrand cross-links. *J Biol Chem,* 283**,** 17075-82.

MINKO, I. G., ZOU, Y. & LLOYD, R. S. 2002. Incision of DNA-protein crosslinks by UvrABC nuclease suggests a potential repair pathway involving nucleotide excision repair. *Proc Natl Acad Sci U S A,* 99**,** 1905-9.

MISRA, R. R. & VOS, J. M. 1993. Defective replication of psoralen adducts detected at the gene-specific level in xeroderma pigmentosum variant cells. *Mol Cell Biol,* 13**,** 1002-12.

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

MITTELMAN, D. & WILSON, J. H. 2013. The fractured genome of HeLa cells. *Genome Biol,* 14**,** 111.

MOLDOVAN, G. L., MADHAVAN, M. V., MIRCHANDANI, K. D., MCCAFFREY, R. M., VINCIGUERRA, P. & D'ANDREA, A. D. 2010. DNA polymerase POLN participates in cross-link repair and homologous recombination. *Mol Cell Biol,* 30**,** 1088-96.

MORRIS, D. L., JR. 2014. DNA-bound metal ions: recent developments. *Biomol Concepts,* 5**,** 397-407.

MOSHOUS, D., CALLEBAUT, I., DE CHASSEVAL, R., CORNEO, B., CAVAZZANA-CALVO, M., LE DEIST, F., TEZCAN, I., SANAL, O., BERTRAND, Y., PHILIPPE, N., FISCHER, A. & DE VILLARTAY, J. P. 2001. Artemis, a novel DNA double-strand break repair/V(D)J recombination protein, is mutated in human severe combined immune deficiency. *Cell,* 105**,** 177-86.

MOTYCKA, T. A., BESSHO, T., POST, S. M., SUNG, P. & TOMKINSON, A. E. 2004. Physical and functional interaction between the XPF/ERCC1 endonuclease and hRad52. *J Biol Chem,* 279**,** 13634-9.

MUNIANDY, P. A., LIU, J., MAJUMDAR, A., LIU, S. T. & SEIDMAN, M. M. 2010. DNA interstrand crosslink repair in mammalian cells: step by step. *Crit Rev Biochem Mol Biol,* 45**,** 23-49.

MUNIANDY, P. A., THAPA, D., THAZHATHVEETIL, A. K., LIU, S. T. & SEIDMAN, M. M. 2009. Repair of laser-localized DNA interstrand cross-links in G1 phase mammalian cells. *J Biol Chem,* 284**,** 27908-17.

MUNOZ, I. M., HAIN, K., DECLAIS, A. C., GARDINER, M., TOH, G. W., SANCHEZ-PULIDO, L., HEUCKMANN, J. M., TOTH, R., MACARTNEY, T., EPPINK, B., KANAAR, R., PONTING, C. P., LILLEY, D. M. & ROUSE, J. 2009. Coordination of structure-specific nucleases by human SLX4/BTBD12 is required for DNA repair. *Mol Cell,* 35**,** 116-27.

NAGASE, T., MIYAJIMA, N., TANAKA, A., SAZUKA, T., SEKI, N., SATO, S., TABATA, S., ISHIKAWA, K., KAWARABAYASI, Y., KOTANI, H. & ET AL. 1995. Prediction of the coding sequences of unidentified human genes. III. The coding sequences of 40 new genes (KIAA0081-KIAA0120) deduced by analysis of cDNA clones from human cell line KG-1. *DNA Res,* 2**,** 37-43.

NEWMAN, J. A., HEWITT, L., RODRIGUES, C., SOLOVYOVA, A., HARWOOD, C. R. & LEWIS, R. J. 2011. Unusual, dual endo- and exonuclease activity in the degradosome explained by crystal structure analysis of RNase J1. *Structure,* 19**,** 1241-51.

NG, J. T., DEKKER, C., REARDON, P. & VON DELFT, F. 2016. Lessons from ten years of crystallization experiments at the SGC. *Acta Crystallogr D Struct Biol,* 72**,** 224-35.

NICK MCELHINNY, S. A., SNOWDEN, C. M., MCCARVILLE, J. & RAMSDEN, D. A. 2000. Ku recruits the XRCC4-ligase IV complex to DNA ends. *Mol Cell Biol,* 20**,** 2996-3003.

NICKOLOFF, J. A. 2017. Paths from DNA damage and signaling to genome rearrangements via homologous recombination. *Mutat Res*.

NIEDERNHOFER, L. J. 2007. The Fanconi anemia signalosome anchor. *Mol Cell,* 25**,** 487-90.

NIEDERNHOFER, L. J., ODIJK, H., BUDZOWSKA, M., VAN DRUNEN, E., MAAS, A., THEIL, A. F., DE WIT, J., JASPERS, N. G., BEVERLOO, H. B., HOEIJMAKERS, J. H. & KANAAR, R. 2004. The structure-specific endonuclease Ercc1-Xpf is required to resolve DNA interstrand cross-link-induced double-strand breaks. *Mol Cell Biol,* 24**,** 5776-87.

NIJMAN, S. M., HUANG, T. T., DIRAC, A. M., BRUMMELKAMP, T. R., KERKHOVEN, R. M., D'ANDREA, A. D. & BERNARDS, R. 2005. The deubiquitinating enzyme USP1 regulates the Fanconi anemia pathway. *Mol Cell,* 17**,** 331-9.

NIMONKAR, A. V., GENSCHEL, J., KINOSHITA, E., POLACZEK, P., CAMPBELL, J. L., WYMAN, C., MODRICH, P. & KOWALCZYKOWSKI, S. C. 2011. BLM-DNA2-RPA-MRN and EXO1-BLM-RPA-MRN constitute two DNA end resection machineries for human DNA break repair. *Genes Dev,* 25**,** 350-62.

NIMONKAR, A. V., OZSOY, A. Z., GENSCHEL, J., MODRICH, P. & KOWALCZYKOWSKI, S. C. 2008. Human exonuclease 1 and BLM helicase interact to resect DNA and initiate DNA repair. *Proc Natl Acad Sci U S A,* 105**,** 16906-11.

NISHINO, T., TAKEUCHI, K., GASCOIGNE, K. E., SUZUKI, A., HORI, T., OYAMA, T., MORIKAWA, K., CHEESEMAN, I. M. & FUKAGAWA, T. 2012. CENP-T-W-S-X forms a unique centromeric chromatin structure with a histone-like fold. *Cell,* 148**,** 487-501.

NIU, P., LIU, L., GONG, Z., TAN, H., WANG, F., YUAN, J., FENG, Y., WEI, Q., TANGUAY, R. M. & WU, T. 2006. Overexpressed heat shock protein 70 protects cells against DNA damage caused by ultraviolet C in a dose-dependent manner. *Cell Stress Chaperones,* 11**,** 162-9.

NOJIMA, K., HOCHEGGER, H., SABERI, A., FUKUSHIMA, T., KIKUCHI, K., YOSHIMURA, M., ORELLI, B. J., BISHOP, D. K., HIRANO, S., OHZEKI, M., ISHIAI, M., YAMAMOTO, K., TAKATA, M., ARAKAWA, H., BUERSTEDDE, J. M., YAMAZOE, M., KAWAMOTO, T., ARAKI, K., TAKAHASHI, J. A., HASHIMOTO, N., TAKEDA, S. & SONODA, E. 2005. Multiple repair pathways mediate tolerance to chemotherapeutic cross-linking agents in vertebrate cells. *Cancer Res,* 65**,** 11704-11.

NOLL, D. M., MASON, T. M. & MILLER, P. S. 2006. Formation and repair of interstrand cross-links in DNA. *Chem Rev,* 106**,** 277-301.

NORMAN, D., LIVE, D., SASTRY, M., LIPMAN, R., HINGERTY, B. E., TOMASZ, M., BROYDE, S. & PATEL, D. J. 1990. NMR and computational characterization of mitomycin cross-linked to adjacent deoxyguanosines in the minor groove of the d(T-A-C-G-T-A).d(T-A-C-G-T-A) duplex. *Biochemistry,* 29**,** 2861-75.

O'DRISCOLL, M., GENNERY, A. R., SEIDEL, J., CONCANNON, P. & JEGGO, P. A. 2004. An overview of three new disorders associated with genetic instability: LIG4 syndrome, RS-SCID and ATR-Seckel syndrome. *DNA Repair (Amst),* 3**,** 1227-35.

OBE, G., JOHANNES, C. & SCHULTE-FROHLINDE, D. 1992. DNA double-strand breaks induced by sparsely ionizing radiation and endonucleases as critical lesions for cell death, chromosomal aberrations, mutations and oncogenic transformation. *Mutagenesis,* 7**,** 3-12.

OGINO, H., NAKABAYASHI, K., SUZUKI, M., TAKAHASHI, E., FUJII, M., SUZUKI, T. & AYUSAWA, D. 1998. Release of telomeric DNA from chromosomes in immortal human cells lacking telomerase activity. *Biochem Biophys Res Commun,* 248**,** 223-7.

OISHI, S. 2002. Effects of butyl paraben on the male reproductive system in mice. *Arch Toxicol,* 76**,** 423-9.

ORANS, J., MCSWEENEY, E. A., IYER, R. R., HAST, M. A., HELLINGA, H. W., MODRICH, P. & BEESE, L. S. 2011. Structures of human exonuclease 1 DNA complexes suggest a unified mechanism for nuclease family. *Cell,* 145**,** 212-23.

PALZKILL, T. 2013. Metallo-beta-lactamase structure and function. *Ann N Y Acad Sci,* 1277**,** 91-104.

PANG, D., YOO, S., DYNAN, W. S., JUNG, M. & DRITSCHILO, A. 1997. Ku proteins join DNA fragments as shown by atomic force microscopy. *Cancer Res,* 57**,** 1412-5.

PANG, Q. & ANDREASSEN, P. R. 2009. Fanconi anemia proteins and endogenous stresses. *Mutat Res,* 668**,** 42-53.

PANIER, S. & BOULTON, S. J. 2014. Double-strand break repair: 53BP1 comes into focus. *Nat Rev Mol Cell Biol,* 15**,** 7-18.

PEPE, A. & WEST, S. C. 2014a. MUS81-EME2 promotes replication fork restart. *Cell Rep,* 7**,** 1048-55.

PEPE, A. & WEST, S. C. 2014b. Substrate specificity of the MUS81-EME2 structure selective endonuclease. *Nucleic Acids Res,* 42**,** 3833-45.

PERRY, J. J., YANNONE, S. M., HOLDEN, L. G., HITOMI, C., ASAITHAMBY, A., HAN, S., COOPER, P. K., CHEN, D. J. & TAINER, J. A. 2006. WRN exonuclease structure and molecular mechanism imply an editing role in DNA end processing. *Nat Struct Mol Biol,* 13**,** 414-22.

PETTINATI, I., BREM, J., LEE, S. Y., MCHUGH, P. J. & SCHOFIELD, C. J. 2016. The Chemical Biology of Human Metallo-beta-Lactamase Fold Proteins. *Trends Biochem Sci,* 41**,** 338-55.

PFEIFFER, P., GOEDECKE, W. & OBE, G. 2000. Mechanisms of DNA double-strand break repair and their potential to induce chromosomal aberrations. *Mutagenesis,* 15**,** 289-302.

PICHIERRI, P., AVERBECK, D. & ROSSELLI, F. 2002. DNA cross-link-dependent RAD50/MRE11/NBS1 subnuclear assembly requires the Fanconi anemia C protein. *Hum Mol Genet,* 11**,** 2531-46.

PIZZOLATO, J., MUKHERJEE, S., SCHARER, O. D. & JIRICNY, J. 2015. FANCD2-associated nuclease 1, but not exonuclease 1 or flap endonuclease 1, is able to unhook DNA interstrand cross-links in vitro. *J Biol Chem,* 290**,** 22602-11.

PLASTARAS, J. P., RIGGINS, J. N., OTTENEDER, M. & MARNETT, L. J. 2000. Reactivity and mutagenicity of endogenous DNA oxopropenylating agents: base propenals, malondialdehyde, and N(epsilon)-oxopropenyllysine. *Chem Res Toxicol,* 13**,** 1235-42.

POLYANICHKO, A. M., ANDRUSHCHENKO, V. V., CHIKHIRZHINA, E. V., VOROB'EV, V. I. & WIESER, H. 2004. The effect of manganese(II) on DNA structure: electronic and vibrational circular dichroism studies. *Nucleic Acids Res,* 32**,** 989-96.

POMMIER, Y., BARCELO, J. M., RAO, V. A., SORDET, O., JOBSON, A. G., THIBAUT, L., MIAO, Z. H., SEILER, J. A., ZHANG, H., MARCHAND, C., AGAMA, K., NITISS, J. L. & REDON, C. 2006. Repair of topoisomerase I-mediated DNA damage. *Prog Nucleic Acid Res Mol Biol,* 81**,** 179-229.

POPP, M. W. & MAQUAT, L. E. 2016. Leveraging Rules of Nonsense-Mediated mRNA Decay for Genome Engineering and Personalized Medicine. *Cell,* 165**,** 1319-1322.

POVIRK, L. F., ZHOU, T., ZHOU, R., COWAN, M. J. & YANNONE, S. M. 2007. Processing of 3'-phosphoglycolate-terminated DNA double strand breaks by Artemis nuclease. *J Biol Chem,* 282**,** 3547-58.

PRAKASH, S. & PRAKASH, L. 2002. Translesion DNA synthesis in eukaryotes: a one- or two-polymerase affair. *Genes Dev,* 16**,** 1872-83.

QIN, N., OLCESE, R., BRANSBY, M., LIN, T. & BIRNBAUMER, L. 1999. Ca2+-induced inhibition of the cardiac Ca2+ channel depends on calmodulin. *Proc Natl Acad Sci U S A,* 96**,** 2435-8.

RADERSCHALL, E., GOLUB, E. I. & HAAF, T. 1999. Nuclear foci of mammalian recombination proteins are located at single-stranded DNA regions formed after DNA damage. *Proc Natl Acad Sci U S A,* 96**,** 1921-6.

RAHMAN, K. M., THOMPSON, A. S., JAMES, C. H., NARAYANASWAMY, M. & THURSTON, D. E. 2009. The pyrrolobenzodiazepine dimer SJG-136 forms sequence-dependent intrastrand DNA cross-links and monoalkylated adducts in addition to interstrand cross-links. *J Am Chem Soc,* 131**,** 13756-66.

RAI, R., HU, C., BROTON, C., CHEN, Y., LEI, M. & CHANG, S. 2017. NBS1 Phosphorylation Status Dictates Repair Choice of Dysfunctional Telomeres. *Mol Cell,* 65**,** 801-817 e4.

RAJSKI, S. R. & WILLIAMS, R. M. 1998. DNA Cross-Linking Agents as Antitumor Drugs. *Chem Rev,* 98**,** 2723-2796.

RAMADAN, K., SHEVELEV, I. V., MAGA, G. & HUBSCHER, U. 2004. De novo DNA synthesis by human DNA polymerase lambda, DNA polymerase mu and terminal deoxyribonucleotidyl transferase. *J Mol Biol,* 339**,** 395-404.

RAN, F. A., HSU, P. D., WRIGHT, J., AGARWALA, V., SCOTT, D. A. & ZHANG, F. 2013. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc,* 8**,** 2281-2308.

RASCHLE, M., KNIPSCHEER, P., ENOIU, M., ANGELOV, T., SUN, J., GRIFFITH, J. D., ELLENBERGER, T. E., SCHARER, O. D. & WALTER, J. C. 2008. Mechanism of replication-coupled DNA interstrand crosslink repair. *Cell,* 134**,** 969-80.

RASS, U., AHEL, I. & WEST, S. C. 2008. Molecular mechanism of DNA deadenylation by the neurological disease protein aprataxin. *J Biol Chem,* 283**,** 33994-4001.

REABAN, M. E., LEBOWITZ, J. & GRIFFIN, J. A. 1994. Transcription induces the formation of a stable RNA.DNA hybrid in the immunoglobulin alpha switch region. *J Biol Chem,* 269**,** 21850-7.

REISHUS, J. W. & MARTIN, D. S. 1961. cis-Dichlorodiammineplatinum(I1). Acid Hydrolysis and Isotopic Exchange of the Chloride Ligands. *Journal of the American Chemical Society,* 83**,** 2457-2462.

RICHARDSON, C. D., RAY, G. J., BRAY, N. L. & CORN, J. E. 2016. Non-homologous DNA increases gene disruption efficiency by altering DNA repair outcomes. *Nat Commun,* 7**,** 12463.

RINK, S. M. & HOPKINS, P. B. 1995. A mechlorethamine-induced DNA interstrand cross-link bends duplex DNA. *Biochemistry,* 34**,** 1439-45.

RINK, S. M., SOLOMON, M. S., TAYLOR, M. J., RAJUR, B. S., MCLAUGHLIN, L. W., HOPKINS, P. B. 1992. Covalent Structure of a Nitrogen Mustard-Induced DNA Interstrand Cross-Link: An N7-to-N7 Linkage of Deoxyguanosine Residues at the Duplex Sequence 5'-d(GNC). *Journal of the American Chemical Society,* 115**,** 2551-2557.

RISS, T. L., MORAVEC, R. A., NILES, A. L., DUELLMAN, S., BENINK, H. A., WORZELLA, T. J. & MINOR, L. 2004. Cell Viability Assays. *In:* SITTAMPALAM, G. S., COUSSENS, N. P., BRIMACOMBE, K., GROSSMAN, A., ARKIN, M., AULD, D., AUSTIN, C., BAELL, J., BEJCEK, B., CHUNG, T. D. Y., DAHLIN, J. L., DEVANARYAN, V., FOLEY, T. L., GLICKSMAN, M., HALL, M. D., HASS, J. V., INGLESE, J., IVERSEN, P. W., KAHL, S. D., KALES, S. C., LAL-NAG, M., LI, Z., MCGEE, J., MCMANUS, O., RISS, T., TRASK, O. J., JR., WEIDNER, J. R., XIA, M. & XU, X. (eds.) *Assay Guidance Manual.* Bethesda (MD).

ROBERTS, R. W. & CROTHERS, D. M. 1992. Stability and properties of double and triple helices: dramatic effects of RNA or DNA backbone composition. *Science,* 258**,** 1463-6.

ROBERTS, S. A., STRANDE, N., BURKHALTER, M. D., STROM, C., HAVENER, J. M., HASTY, P. & RAMSDEN, D. A. 2010. Ku is a 5'-dRP/AP lyase that excises nucleotide damage near broken ends. *Nature,* 464**,** 1214-7.

ROHLEDER, F., HUANG, J., XUE, Y., KUPER, J., ROUND, A., SEIDMAN, M., WANG, W. & KISKER, C. 2016. FANCM interacts with PCNA to promote replication traverse of DNA interstrand crosslinks. *Nucleic Acids Res,* 44**,** 3219-32.

ROSSMANITH, W. 2011. Localization of human RNase Z isoforms: dual nuclear/mitochondrial targeting of the ELAC2 gene product by alternative translation initiation. *PLoS One,* 6**,** e19152.

ROTHENBERG, E., GRIMME, J. M., SPIES, M. & HA, T. 2008. Human Rad52-mediated homology search and annealing occurs by continuous interactions between overlapping nucleoprotein complexes. *Proc Natl Acad Sci U S A,* 105**,** 20274-9.

ROY, D., ZHANG, Z., LU, Z., HSIEH, C. L. & LIEBER, M. R. 2010. Competition between the RNA transcript and the nontemplate DNA strand during R-loop formation in vitro: a nick can serve as a strong R-loop initiation site. *Mol Cell Biol,* 30**,** 146-59.

RUHLAND, A., HAASE, E., SIEDE, W. & BRENDEL, M. 1981a. Isolation of yeast mutants sensitive to the bifunctional alkylating agent nitrogen mustard. *Mol Gen Genet,* 181**,** 346-51.

RUHLAND, A., KIRCHER, M., WILBORN, F. & BRENDEL, M. 1981b. A yeast mutant specifically sensitive to bifunctional alkylation. *Mutat Res,* 91**,** 457-62.

RYDBERG, B. & LINDAHL, T. 1982. Nonenzymatic methylation of DNA by the intracellular methyl group donor S-adenosyl-L-methionine is a potentially mutagenic reaction. *EMBO J,* 1**,** 211-6.

SAKOFSKY, C. J. & MALKOVA, A. 2017. Break induced replication in eukaryotes: mechanisms, functions, and consequences. *Crit Rev Biochem Mol Biol,* 52**,** 395-413.

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

SALEWSKY, B., SCHMIESTER, M., SCHINDLER, D., DIGWEED, M. & DEMUTH, I. 2012. The nuclease hSNM1B/Apollo is linked to the Fanconi anemia pathway via its interaction with FANCP/SLX4. *Hum Mol Genet,* 21**,** 4948-56.

SALTMAN, D., MORGAN, R., CLEARY, M. L. & DE LANGE, T. 1993. Telomeric structure in cells with chromosome end associations. *Chromosoma,* 102**,** 121-8.

SANCHEZ-RIVERA, F. J. & JACKS, T. 2015. Applications of the CRISPR-Cas9 system in cancer biology. *Nat Rev Cancer,* 15**,** 387-95.

SARKAR, S., DAVIES, A. A., ULRICH, H. D. & MCHUGH, P. J. 2006. DNA interstrand crosslink repair during G1 involves nucleotide excision repair and DNA polymerase zeta. *EMBO J,* 25**,** 1285-94.

SARTORI, A. A., LUKAS, C., COATES, J., MISTRIK, M., FU, S., BARTEK, J., BAER, R., LUKAS, J. & JACKSON, S. P. 2007. Human CtIP promotes DNA end resection. *Nature,* 450**,** 509-14.

SAVIC, V., YIN, B., MAAS, N. L., BREDEMEYER, A. L., CARPENTER, A. C., HELMINK, B. A., YANG-IOTT, K. S., SLECKMAN, B. P. & BASSING, C. H. 2009. Formation of dynamic gamma-H2AX domains along broken DNA strands is distinctly regulated by ATM and MDC1 and dependent upon H2AX densities in chromatin. *Mol Cell,* 34**,** 298-310.

SCHLACHER, K., WU, H. & JASIN, M. 2012. A distinct replication fork protection pathway connects Fanconi anemia tumor suppressors to RAD51-BRCA1/2. *Cancer Cell,* 22**,** 106-16.

SCHMIESTER, M. & DEMUTH, I. 2017. SNM1B/Apollo in the DNA damage response and telomere maintenance. *Oncotarget,* 8**,** 48398-48409.

SCHUMACHER, B., GARINIS, G. A. & HOEIJMAKERS, J. H. 2008. Age to survive: DNA damage and aging. *Trends Genet,* 24**,** 77-85.

SEBESTA, M., BURKOVICS, P., JUHASZ, S., ZHANG, S., SZABO, J. E., LEE, M. Y., HARACSKA, L. & KREJCI, L. 2013. Role of PCNA and TLS polymerases in D-loop extension during homologous recombination in humans. *DNA Repair (Amst),* 12**,** 691-8.

SEN, D. & GILBERT, W. 1990. A sodium-potassium switch in the formation of four-stranded G4-DNA. *Nature,* 344**,** 410-4.

SENGEROVA, B., ALLERSTON, C. K., ABU, M., LEE, S. Y., HARTLEY, J., KIAKOS, K., SCHOFIELD, C. J., HARTLEY, J. A., GILEADI, O. & MCHUGH, P. J. 2012. Characterization of the human SNM1A and SNM1B/Apollo DNA repair exonucleases. *J Biol Chem,* 287**,** 26254-67.

SHAMANNA, R. A., LU, H., DE FREITAS, J. K., TIAN, J., CROTEAU, D. L. & BOHR, V. A. 2016. WRN regulates pathway choice between classical and alternative non-homologous end joining. *Nat Commun,* 7**,** 13785.

SHARMA, S. & CANMAN, C. E. 2012. REV1 and DNA polymerase zeta in DNA interstrand crosslink repair. *Environ Mol Mutagen,* 53**,** 725-40.

SHARMA, S., HICKS, J. K., CHUTE, C. L., BRENNAN, J. R., AHN, J. Y., GLOVER, T. W. & CANMAN, C. E. 2012. REV1 and polymerase zeta facilitate homologous recombination repair. *Nucleic Acids Res,* 40**,** 682-91.

SHARMA, S., JAVADEKAR, S. M., PANDEY, M., SRIVASTAVA, M., KUMARI, R. & RAGHAVAN, S. C. 2015. Homology and enzymatic requirements of microhomology-dependent alternative end joining. *Cell Death Dis,* 6**,** e1697.

SHARPE, J. J. & COOPER, T. A. 2017. Unexpected consequences: exon skipping caused by CRISPR-generated mutations. *Genome Biol,* 18**,** 109.

SHEN, X., DO, H., LI, Y., CHUNG, W. H., TOMASZ, M., DE WINTER, J. P., XIA, B., ELLEDGE, S. J., WANG, W. & LI, L. 2009. Recruitment of fanconi anemia and breast cancer proteins to DNA damage sites is differentially governed by replication. *Mol Cell,* 35**,** 716-23.

SHEN, X., JUN, S., O'NEAL, L. E., SONODA, E., BEMARK, M., SALE, J. E. & LI, L. 2006. 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.

SHILOH, Y. & ZIV, Y. 2013. The ATM protein kinase: regulating the cellular response to genotoxic stress, and more. *Nat Rev Mol Cell Biol,* 14**,** 197-210.

SHRESTHA, B., SMEE, C. & GILEADI, O. 2008. Baculovirus expression vector system: an emerging host for high-throughput eukaryotic protein expression. *Methods Mol Biol,* 439**,** 269-89.

SIJBERS, A. M., DE LAAT, W. L., ARIZA, R. R., BIGGERSTAFF, M., WEI, Y. F., MOGGS, J. G., CARTER, K. C., SHELL, B. K., EVANS, E., DE JONG, M. C., RADEMAKERS, S., DE ROOIJ, J., JASPERS, N. G., HOEIJMAKERS, J. H. & WOOD, R. D. 1996. Xeroderma pigmentosum group F caused by a defect in a structure-specific DNA repair endonuclease. *Cell,* 86**,** 811-22.

SILVA, A. P., CHECHIK, M., BYRNE, R. T., WATERMAN, D. G., NG, C. L., DODSON, E. J., KOONIN, E. V., ANTSON, A. A. & SMITS, C. 2011. Structure and activity of a novel archaeal beta-CASP protein with N-terminal KH domains. *Structure,* 19**,** 622-32.

SINGER, B., CHAVEZ, F., GOODMAN, M. F., ESSIGMANN, J. M. & DOSANJH, M. K. 1989. Effect of 3' flanking neighbors on kinetics of pairing of dCTP or dTTP opposite O6-methylguanine in a defined primed oligonucleotide when Escherichia coli DNA polymerase I is used. *Proc Natl Acad Sci U S A,* 86**,** 8271-4.

SINHA, R. P. & HADER, D. P. 2002. UV-induced DNA damage and repair: a review. *Photochem Photobiol Sci,* 1**,** 225-36.

SITUM, M., BULAT, V., MAJCEN, K., DZAPO, A. & JEZOVITA, J. 2014. Benefits of controlled ultraviolet radiation in the treatment of dermatological diseases. *Coll Antropol,* 38**,** 1249-53.

SKOURTI-STATHAKI, K., KAMIENIARZ-GDULA, K. & PROUDFOOT, N. J. 2014. R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature,* 516**,** 436-9.

SKOURTI-STATHAKI, K., PROUDFOOT, N. J. & GROMAK, N. 2011. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol Cell,* 42**,** 794-805.

SLATER, J. P., MILDVAN, A. S. & LOEB, L. A. 1971. Zinc in DNA polymerases. *Biochem Biophys Res Commun,* 44**,** 37-43.

SMOGORZEWSKA, A., DESETTY, R., SAITO, T. T., SCHLABACH, M., LACH, F. P., SOWA, M. E., CLARK, A. B., KUNKEL, T. A., HARPER, J. W., COLAIACOVO, M. P. & ELLEDGE, S. J. 2010. A genetic screen identifies FAN1, a Fanconi anemia-associated nuclease necessary for DNA interstrand crosslink repair. *Mol Cell,* 39**,** 36-47.

SMOGORZEWSKA, A., MATSUOKA, S., VINCIGUERRA, P., MCDONALD, E. R., 3RD, HUROV, K. E., LUO, J., BALLIF, B. A., GYGI, S. P., HOFMANN, K., D'ANDREA, A. D. & ELLEDGE, S. J. 2007. Identification of the FANCI protein, a monoubiquitinated FANCD2 paralog required for DNA repair. *Cell,* 129**,** 289-301.

SMOGORZEWSKA, A., VAN STEENSEL, B., BIANCHI, A., OELMANN, S., SCHAEFER, M. R., SCHNAPP, G. & DE LANGE, T. 2000. Control of human telomere length by TRF1 and TRF2. *Mol Cell Biol,* 20**,** 1659-68.

SONODA, E., HOCHEGGER, H., SABERI, A., TANIGUCHI, Y. & TAKEDA, S. 2006. Differential usage of non-homologous end-joining and homologous recombination in double strand break repair. *DNA Repair (Amst),* 5**,** 1021-9.

SONODA, E., OKADA, T., ZHAO, G. Y., TATEISHI, S., ARAKI, K., YAMAIZUMI, M., YAGI, T., VERKAIK, N. S., VAN GENT, D. C., TAKATA, M. & TAKEDA, S. 2003. Multiple roles of Rev3, the catalytic subunit of polzeta in maintaining genome stability in vertebrates. *EMBO J,* 22**,** 3188-97.

STEIN, C. A., SUBASINGHE, C., SHINOZUKA, K. & COHEN, J. S. 1988. Physicochemical properties of phosphorothioate oligodeoxynucleotides. *Nucleic Acids Res,* 16**,** 3209-21.

STIFF, T., O'DRISCOLL, M., RIEF, N., IWABUCHI, K., LOBRICH, M. & JEGGO, P. A. 2004. ATM and DNA-PK function redundantly to phosphorylate H2AX after exposure to ionizing radiation. *Cancer Res,* 64**,** 2390-6.

STIFF, T., WALKER, S. A., CEROSALETTI, K., GOODARZI, A. A., PETERMANN, E., CONCANNON, P., O'DRISCOLL, M. & JEGGO, P. A. 2006. ATR-dependent phosphorylation and activation of ATM in response to UV treatment or replication fork stalling. *EMBO J,* 25**,** 5775-82.

STONE, M. P., CHO, Y. J., HUANG, H., KIM, H. Y., KOZEKOV, I. D., KOZEKOVA, A., WANG, H., MINKO, I. G., LLOYD, R. S., HARRIS, T. M. & RIZZO, C. J. 2008. Interstrand DNA cross-links induced by alpha,beta-unsaturated aldehydes derived from lipid peroxidation and environmental sources. *Acc Chem Res,* 41**,** 793-804.

STRACKER, T. H. & PETRINI, J. H. 2011. The MRE11 complex: starting from the ends. *Nat Rev Mol Cell Biol,* 12**,** 90-103.

STRASSER, K., MASUDA, S., MASON, P., PFANNSTIEL, J., OPPIZZI, M., RODRIGUEZ-NAVARRO, S., RONDON, A. G., AGUILERA, A., STRUHL, K., REED, R. & HURT, E. 2002. TREX is a conserved complex coupling transcription with messenger RNA export. *Nature,* 417**,** 304-8.

STRINGER, J. R. & COUNTER, C. M. 2012. Snm1B interacts with PSF2. *PLoS One,* 7**,** e49626.

STUCKI, M., CLAPPERTON, J. A., MOHAMMAD, D., YAFFE, M. B., SMERDON, S. J. & JACKSON, S. P. 2005. MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. *Cell,* 123**,** 1213-26.

SUN, Q., CSORBA, T., SKOURTI-STATHAKI, K., PROUDFOOT, N. J. & DEAN, C. 2013. R-loop stabilization represses antisense transcription at the Arabidopsis FLC locus. *Science,* 340**,** 619-21.

SUN, Y., JIANG, X., CHEN, S., FERNANDES, N. & PRICE, B. D. 2005. A role for the Tip60 histone acetyltransferase in the acetylation and activation of ATM. *Proc Natl Acad Sci U S A,* 102**,** 13182-7.

SUN, Y., JIANG, X. & PRICE, B. D. 2010. Tip60: connecting chromatin to DNA damage signaling. *Cell Cycle,* 9**,** 930-6.

SVENDSEN, J. M. & HARPER, J. W. 2010. GEN1/Yen1 and the SLX4 complex: Solutions to the problem of Holliday junction resolution. *Genes Dev,* 24**,** 521-36.

SVENDSEN, J. M., SMOGORZEWSKA, A., SOWA, M. E., O'CONNELL, B. C., GYGI, S. P., ELLEDGE, S. J. & HARPER, J. W. 2009. Mammalian BTBD12/SLX4 assembles a Holliday junction resolvase and is required for DNA repair. *Cell,* 138**,** 63-77.

SYMINGTON, L. S. 2002. Role of RAD52 epistasis group genes in homologous recombination and double-strand break repair. *Microbiol Mol Biol Rev,* 66**,** 630-70, table of contents.

TACCONI, E. M., LAI, X., FOLIO, C., PORRU, M., ZONDERLAND, G., BADIE, S., MICHL, J., SECHI, I., ROGIER, M., MATIA GARCIA, V., BATRA, A. S., RUEDA, O. M., BOUWMAN, P., JONKERS, J., RYAN, A., REINA-SAN-MARTIN, B., HUI, J., TANG, N., BRUNA, A., BIROCCIO, A. & TARSOUNAS, M. 2017. BRCA1 and BRCA2 tumor suppressors protect against endogenous acetaldehyde toxicity. *EMBO Mol Med*.

TANIGUCHI, T., GARCIA-HIGUERA, I., XU, B., ANDREASSEN, P. R., GREGORY, R. C., KIM, S. T., LANE, W. S., KASTAN, M. B. & D'ANDREA, A. D. 2002. Convergence of the fanconi anemia and ataxia telangiectasia signaling pathways. *Cell,* 109**,** 459-72.

TIBBETTS, R. S., BRUMBAUGH, K. M., WILLIAMS, J. M., SARKARIA, J. N., CLIBY, W. A., SHIEH, S. Y., TAYA, Y., PRIVES, C. & ABRAHAM, R. T. 1999. A role for ATR in the DNA damage-induced phosphorylation of p53. *Genes Dev,* 13**,** 152-7.

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

TOKUTAKE, Y., MATSUMOTO, T., WATANABE, T., MAEDA, S., TAHARA, H., SAKAMOTO, S., NIIDA, H., SUGIMOTO, M., IDE, T. & FURUICHI, Y. 1998. Extra-chromosomal telomere repeat DNA in telomerase-negative immortalized cell lines. *Biochem Biophys Res Commun,* 247**,** 765-72.

TOUZOT, F., CALLEBAUT, I., SOULIER, J., GAILLARD, L., AZERRAD, C., DURANDY, A., FISCHER, A., DE VILLARTAY, J. P. & REVY, P. 2010. Function of Apollo (SNM1B) at telomere highlighted by a splice variant identified in a patient with Hoyeraal-Hreidarsson syndrome. *Proc Natl Acad Sci U S A,* 107**,** 10097-102.

TRICKER, A. R. & PREUSSMANN, R. 1991. Carcinogenic N-nitrosamines in the diet: occurrence, formation, mechanisms and carcinogenic potential. *Mutat Res,* 259**,** 277-89.

TUBBS, J. L., PEGG, A. E. & TAINER, J. A. 2007. DNA binding, nucleotide flipping, and the helix-turn-helix motif in base repair by O6-alkylguanine-DNA alkyltransferase and its implications for cancer chemotherapy. *DNA Repair (Amst),* 6**,** 1100-15.

TUMBALE, P., APPEL, C. D., KRAEHENBUEHL, R., ROBERTSON, P. D., WILLIAMS, J. S., KRAHN, J., AHEL, I. & WILLIAMS, R. S. 2011. Structure of an aprataxin-DNA complex with insights into AOA1 neurodegenerative disease. *Nat Struct Mol Biol,* 18**,** 1189-95.

TUMINI, E., PLEVANI, P., MUZI-FALCONI, M. & MARINI, F. 2011. Physical and functional crosstalk between Fanconi anemia core components and the GINS replication complex. *DNA Repair (Amst),* 10**,** 149-58.

URINGA, E. J., YOUDS, J. L., LISAINGO, K., LANSDORP, P. M. & BOULTON, S. J. 2011. RTEL1: an essential helicase for telomere maintenance and the regulation of homologous recombination. *Nucleic Acids Res,* 39**,** 1647-55.

VAN OVERBEEK, M. & DE LANGE, T. 2006. Apollo, an Artemis-related nuclease, interacts with TRF2 and protects human telomeres in S phase. *Curr Biol,* 16**,** 1295-302.

VANDERVEEN, L. A., HASHIM, M. F., NECHEV, L. V., HARRIS, T. M., HARRIS, C. M. & MARNETT, L. J. 2001. Evaluation of the mutagenic potential of the principal DNA adduct of acrolein. *J Biol Chem,* 276**,** 9066-70.

VANNIER, J. B., PAVICIC-KALTENBRUNNER, V., PETALCORIN, M. I., DING, H. & BOULTON, S. J. 2012. RTEL1 dismantles T loops and counteracts telomeric G4-DNA to maintain telomere integrity. *Cell,* 149**,** 795-806.

VARLAND, S., OSBERG, C. & ARNESEN, T. 2015. N-terminal modifications of cellular proteins: The enzymes involved, their substrate specificities and biological effects. *Proteomics,* 15**,** 2385-401.

VERMEULEN, W. & FOUSTERI, M. 2013. Mammalian transcription-coupled excision repair. *Cold Spring Harb Perspect Biol,* 5**,** a012625.

VIGNARD, J., MIREY, G. & SALLES, B. 2013. Ionizing-radiation induced DNA double-strand breaks: a direct and indirect lighting up. *Radiother Oncol,* 108**,** 362-9.

VIGUERA, E., CANCEILL, D. & EHRLICH, S. D. 2001. Replication slippage involves DNA polymerase pausing and dissociation. *EMBO J,* 20**,** 2587-95.

WAHBA, L., AMON, J. D., KOSHLAND, D. & VUICA-ROSS, M. 2011. RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. *Mol Cell,* 44**,** 978-88.

WALSH, P. S., ERLICH, H. A. & HIGUCHI, R. 1992. Preferential PCR amplification of alleles: mechanisms and solutions. *PCR Methods Appl,* 1**,** 241-50.

WANG, A. T., SENGEROVA, B., CATTELL, E., INAGAWA, T., HARTLEY, J. M., KIAKOS, K., BURGESS-BROWN, N. A., SWIFT, L. P., ENZLIN, J. H., SCHOFIELD, C. J., GILEADI, O., HARTLEY, J. A. & MCHUGH, P. J. 2011. Human SNM1A and XPF-ERCC1 collaborate to initiate DNA interstrand cross-link repair. *Genes Dev,* 25**,** 1859-70.

WANG, M., WU, W., WU, W., ROSIDI, B., ZHANG, L., WANG, H. & ILIAKIS, G. 2006. PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways. *Nucleic Acids Res,* 34**,** 6170-82.

WANG, R., PERSKY, N. S., YOO, B., OUERFELLI, O., SMOGORZEWSKA, A., ELLEDGE, S. J. & PAVLETICH, N. P. 2014. DNA repair. Mechanism of DNA interstrand cross-link processing by repair nuclease FAN1. *Science,* 346**,** 1127-30.

WANG, R. C., SMOGORZEWSKA, A. & DE LANGE, T. 2004. Homologous recombination generates T-loop-sized deletions at human telomeres. *Cell,* 119**,** 355-68.

WANG, T., BIRSOY, K., HUGHES, N. W., KRUPCZAK, K. M., POST, Y., WEI, J. J., LANDER, E. S. & SABATINI, D. M. 2015. Identification and characterization of essential genes in the human genome. *Science,* 350**,** 1096-101.

WANG, X., KENNEDY, R. D., RAY, K., STUCKERT, P., ELLENBERGER, T. & D'ANDREA, A. D. 2007. Chk1-mediated phosphorylation of FANCE is required for the Fanconi anemia/BRCA pathway. *Mol Cell Biol,* 27**,** 3098-108.

WANG, X., PETERSON, C. A., ZHENG, H., NAIRN, R. S., LEGERSKI, R. J. & LI, L. 2001. 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.

WARD, I. M. & CHEN, J. 2001. Histone H2AX is phosphorylated in an ATR-dependent manner in response to replicational stress. *J Biol Chem,* 276**,** 47759-62.

WARDMAN, P. 2009. The importance of radiation chemistry to radiation and free radical biology (The 2008 Silvanus Thompson Memorial Lecture). *Br J Radiol,* 82**,** 89-104.

WEBB, C. J., WU, Y. & ZAKIAN, V. A. 2013. DNA repair at telomeres: keeping the ends intact. *Cold Spring Harb Perspect Biol,* 5.

WILLIAMS, H. L., GOTTESMAN, M. E. & GAUTIER, J. 2013. The differences between ICL repair during and outside of S phase. *Trends Biochem Sci,* 38**,** 386-93.

WILLIAMS, R. O. & LOEB, L. A. 1973. Zinc requirement for DNA replication in stimulated human lymphocytes. *J Cell Biol,* 58**,** 594-601.

WOODBINE, L., BRUNTON, H., GOODARZI, A. A., SHIBATA, A. & JEGGO, P. A. 2011. Endogenously induced DNA double strand breaks arise in heterochromatic DNA regions and require ataxia telangiectasia mutated and Artemis for their repair. *Nucleic Acids Res,* 39**,** 6986-97.

WU, L. & HICKSON, I. D. 2003. The Bloom's syndrome helicase suppresses crossing over during homologous recombination. *Nature,* 426**,** 870-4.

WU, M., LIN, Z., LI, X., XIN, X., AN, J., ZHENG, Q., YANG, Y. & LU, D. 2016. HULC cooperates with MALAT1 to aggravate liver cancer stem cells growth through telomere repeat-binding factor 2. *Sci Rep,* 6**,** 36045.

WU, P., TAKAI, H. & DE LANGE, T. 2012. Telomeric 3' overhangs derive from resection by Exo1 and Apollo and fill-in by POT1b-associated CST. *Cell,* 150**,** 39-52.

WU, P., VAN OVERBEEK, M., ROONEY, S. & DE LANGE, T. 2010. Apollo contributes to G overhang maintenance and protects leading-end telomeres. *Mol Cell,* 39**,** 606-17.

XIE, A., KWOK, A. & SCULLY, R. 2009. Role of mammalian Mre11 in classical and alternative nonhomologous end joining. *Nat Struct Mol Biol,* 16**,** 814-8.

XIN, H., LIU, D. & SONGYANG, Z. 2008. The telosome/shelterin complex and its functions. *Genome Biol,* 9**,** 232.

XUE, Y., LI, Y., GUO, R., LING, C. & WANG, W. 2008. FANCM of the Fanconi anemia core complex is required for both monoubiquitination and DNA repair. *Hum Mol Genet,* 17**,** 1641-52.

XUE, Y., RATCLIFF, G. C., WANG, H., DAVIS-SEARLES, P. R., GRAY, M. D., ERIE, D. A. & REDINBO, M. R. 2002. A minimal exonuclease domain of WRN forms a hexamer on DNA and possesses both 3'- 5' exonuclease and 5'-protruding strand endonuclease activities. *Biochemistry,* 41**,** 2901-12.

YAMAMOTO, K. N., KOBAYASHI, S., TSUDA, M., KURUMIZAKA, H., TAKATA, M., KONO, K., JIRICNY, J., TAKEDA, S. & HIROTA, K. 2011. Involvement of SLX4 in interstrand cross-link repair is regulated by the Fanconi anemia pathway. *Proc Natl Acad Sci U S A,* 108**,** 6492-6.

YAMASHITA, T., ODA, T. & SEKIMOTO, T. 2007. Hsp90 and the Fanconi anemia pathway: a molecular link between protein quality control and the DNA damage response. *Cell Cycle,* 6**,** 2232-5.

YAN, Y., AKHTER, S., ZHANG, X. & LEGERSKI, R. 2010a. The multifunctional SNM1 gene family: not just nucleases. *Future Oncol,* 6**,** 1015-29.

YAN, Z., DELANNOY, M., LING, C., DAEE, D., OSMAN, F., MUNIANDY, P. A., SHEN, X., OOSTRA, A. B., DU, H., STELTENPOOL, J., LIN, T., SCHUSTER, B., DECAILLET, C., STASIAK, A., STASIAK, A. Z., STONE, S., HOATLIN, M. E., SCHINDLER, D., WOODCOCK, C. L., JOENJE, H., SEN, R., DE WINTER, J. P., LI, L., SEIDMAN, M. M., WHITBY, M. C., MYUNG, K., CONSTANTINOU, A. &

WANG, W. 2010b. A histone-fold complex and FANCM form a conserved DNA-remodeling complex to maintain genome stability. *Mol Cell,* 37**,** 865-78.
YANG, W. 2011. Nucleases: diversity of structure, function and mechanism. *Q Rev Biophys,* 44**,** 1-93.
YANG, W., LEE, J. Y. & NOWOTNY, M. 2006. Making and breaking nucleic acids: two-Mg2+-ion catalysis and substrate specificity. *Mol Cell,* 22**,** 5-13.
YANG, X. C., SULLIVAN, K. D., MARZLUFF, W. F. & DOMINSKI, Z. 2009. Studies of the 5' exonuclease and endonuclease activities of CPSF-73 in histone pre-mRNA processing. *Mol Cell Biol,* 29**,** 31-42.
YANG, Y., MCBRIDE, K. M., HENSLEY, S., LU, Y., CHEDIN, F. & BEDFORD, M. T. 2014. Arginine methylation facilitates the recruitment of TOP3B to chromatin to prevent R loop accumulation. *Mol Cell,* 53**,** 484-97.
YANG, Z., NEJAD, M. I., VARELA, J. G., PRICE, N. E., WANG, Y. & GATES, K. S. 2017a. A role for the base excision repair enzyme NEIL3 in replication-dependent repair of interstrand DNA cross-links derived from psoralen and abasic sites. *DNA Repair (Amst),* 52**,** 1-11.
YANG, Z., PRICE, N. E., JOHNSON, K. M., WANG, Y. & GATES, K. S. 2017b. Interstrand cross-links arising from strand breaks at true abasic sites in duplex DNA. *Nucleic Acids Res*.
YANO, K. & CHEN, D. J. 2008. Live cell imaging of XLF and XRCC4 reveals a novel view of protein assembly in the non-homologous end-joining pathway. *Cell Cycle,* 7**,** 1321-5.
YANO, K., MOROTOMI-YANO, K., WANG, S. Y., UEMATSU, N., LEE, K. J., ASAITHAMBY, A., WETERINGS, E. & CHEN, D. J. 2008. Ku recruits XLF to DNA double-strand breaks. *EMBO Rep,* 9**,** 91-6.
YE, J., LENAIN, C., BAUWENS, S., RIZZO, A., SAINT-LEGER, A., POULET, A., BENARROCH, D., MAGDINIER, F., MORERE, J., AMIARD, S., VERHOEYEN, E., BRITTON, S., CALSOU, P., SALLES, B., BIZARD, A., NADAL, M., SALVATI, E., SABATIER, L., WU, Y., BIROCCIO, A., LONDONO-VALLEJO, A., GIRAUD-PANIS, M. J. & GILSON, E. 2010. TRF2 and apollo cooperate with topoisomerase 2alpha to protect human telomeres from replicative damage. *Cell,* 142**,** 230-42.
YONEKURA, S., NAKAMURA, N., YONEI, S. & ZHANG-AKIYAMA, Q. M. 2009. Generation, biological consequences and repair mechanisms of cytosine deamination in DNA. *J Radiat Res,* 50**,** 19-26.
YOO, S. & DYNAN, W. S. 1999. Geometry of a complex formed by double strand break repair proteins at a single DNA end: recruitment of DNA-PKcs induces inward translocation of Ku protein. *Nucleic Acids Res,* 27**,** 4679-86.
YOU, Z. & BAILIS, J. M. 2010. DNA damage and decisions: CtIP coordinates DNA repair and cell cycle checkpoints. *Trends Cell Biol,* 20**,** 402-9.
YU, A. M. & MCVEY, M. 2010. Synthesis-dependent microhomology-mediated end joining accounts for multiple types of repair junctions. *Nucleic Acids Res,* 38**,** 5706-17.
YU, H. 2002. Environmental carcinogenic polycyclic aromatic hydrocarbons: photochemistry and phototoxicity. *J Environ Sci Health C Environ Carcinog Ecotoxicol Rev,* 20**,** 149-83.
YU, X., FU, S., LAI, M., BAER, R. & CHEN, J. 2006. BRCA1 ubiquitinates its phosphorylation-dependent binding partner CtIP. *Genes Dev,* 20**,** 1721-6.
ZAUG, A. J., PODELL, E. R. & CECH, T. R. 2005. Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro. *Proc Natl Acad Sci U S A,* 102**,** 10864-9.
ZHANG, F., FAN, Q., REN, K. & ANDREASSEN, P. R. 2009. PALB2 functionally connects the breast cancer susceptibility proteins BRCA1 and BRCA2. *Mol Cancer Res,* 7**,** 1110-8.
ZHANG, H., LIU, H., CHEN, Y., YANG, X., WANG, P., LIU, T., DENG, M., QIN, B., CORREIA, C., LEE, S., KIM, J., SPARKS, M., NAIR, A. A., EVANS, D. L., KALARI, K. R., ZHANG, P., WANG, L., YOU, Z., KAUFMANN, S. H., LOU, Z. &

PEI, H. 2016. A cell cycle-dependent BRCA1-UHRF1 cascade regulates DNA double-strand break repair pathway choice. *Nat Commun,* 7**,** 10201.

ZHANG, J., DEWAR, J. M., BUDZOWSKA, M., MOTNENKO, A., COHN, M. A. & WALTER, J. C. 2015. DNA interstrand cross-link repair requires replication-fork convergence. *Nat Struct Mol Biol,* 22**,** 242-7.

ZHAO, H. & PIWNICA-WORMS, H. 2001. ATR-mediated checkpoint pathways regulate phosphorylation and activation of human Chk1. *Mol Cell Biol,* 21**,** 4129-39.

ZHENG, H., CHRUSZCZ, M., LASOTA, P., LEBIODA, L. & MINOR, W. 2008. Data mining of metal ion environments present in protein structures. *J Inorg Biochem,* 102**,** 1765-76.

ZHENG, H., WANG, X., WARREN, A. J., LEGERSKI, R. J., NAIRN, R. S., HAMILTON, J. W. & LI, L. 2003. Nucleotide excision repair- and polymerase eta-mediated error-prone removal of mitomycin C interstrand cross-links. *Mol Cell Biol,* 23**,** 754-61.

ZHOU, W., OTTO, E. A., CLUCKEY, A., AIRIK, R., HURD, T. W., CHAKI, M., DIAZ, K., LACH, F. P., BENNETT, G. R., GEE, H. Y., GHOSH, A. K., NATARAJAN, S., THONGTHIP, S., VETURI, U., ALLEN, S. J., JANSSEN, S., RAMASWAMI, G., DIXON, J., BURKHALTER, F., SPOENDLIN, M., MOCH, H., MIHATSCH, M. J., VERINE, J., READE, R., SOLIMAN, H., GODIN, M., KISS, D., MONGA, G., MAZZUCCO, G., AMANN, K., ARTUNC, F., NEWLAND, R. C., WIECH, T., ZSCHIEDRICH, S., HUBER, T. B., FRIEDL, A., SLAATS, G. G., JOLES, J. A., GOLDSCHMEDING, R., WASHBURN, J., GILES, R. H., LEVY, S., SMOGORZEWSKA, A. & HILDEBRANDT, F. 2012. FAN1 mutations cause karyomegalic interstitial nephritis, linking chronic kidney failure to defective DNA damage repair. *Nat Genet,* 44**,** 910-5.

ZHU, Z., CHUNG, W. H., SHIM, E. Y., LEE, S. E. & IRA, G. 2008. Sgs1 helicase and two nucleases Dna2 and Exo1 resect DNA double-strand break ends. *Cell,* 134**,** 981-94.

ZIETLOW, L., SMITH, L. A., BESSHO, M. & BESSHO, T. 2009. Evidence for the involvement of human DNA polymerase N in the repair of DNA interstrand cross-links. *Biochemistry,* 48**,** 11817-24.

ZOU, L. & ELLEDGE, S. J. 2003. Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. *Science,* 300**,** 1542-8.

Appendices

Appendix I. Table of PCR primers used in this study.

Site directed mutagenesis PCR primers for the introduction of point mutations	
H61Y forward	5' CCACCTCTTGCATCGTTACCTACAGGTATCTAAG 3'
H61Y reverse	5' TTAGATACCTGTAGGTAACGATGCAAGAGGTGGG 3'
D35A/H36A forward	5' ACATGCACTCGGCCGCCACCGTGGGCCTGTCTAG 3'
D35A/H36A reverse	5' GGAAAAAGTTCACTGCAGTCTCAAGATATTTGAGTGCTAGAC 3'
Primers for ligation independent cloning of the SNM1B gene into pFB-CT10HF-LIC/pFB-LIC-BseR1	
pFB-CT10HF-LIC forward	5' TTAAGAAGGAGATATACTATGAATGGGGTCCTGATCCCCCA 3'
pFB-CT10HF-LIC reverse	5' GATTGGAAGTAGAGGTTCTCTGCGTTGGTTTTCTAGAGGAAG 3'
pFB-LIC-BseR1 forward	5' TACTTCCAATCCATGGGATACGGTAATGGGGTCCTGATCCCCCAT 3'
pFB-LIC-BseR1 reverse	5' TATCCACCTTTACTGTCAGCTTGGTTTTCTAGAGGAAG 3'
Primers for colony PCR screen following annealing of the SNM1B gene into pFB-CT10HF-LIC/pFB-LIC-BseR1	
Forward	5' TATTCATACCGTCCCACCA 3'
Reverse	5' GGGAGGTTTTTTAAAGCAAGTAAA 3'
Primers for bacmid PCR screen to validate correct transposition	
Forward	5' GGATTATTCATACCGTCCCACCA 3'
Reverse	5' CAGGAAACAGCTATGAC 3'
Primers for gDNA PCR screen to detect SNM1B disrupted clones	
Forward:	5' TTTATCGGGACGCCGTTGTGGA 3'
Reverse:	5' ATAAAGGTAGGAATAGTTTCCT 3'

qRT-PCR primers used in quantification of SNM1B mRNA levels	
GAPDH forward	5’ AACAGCGACACCCATCCTC 3’
GAPDH reverse	5’ CATACCAGGAAATGAGCTTGACAA 3’
SNM1B #1 forward	5’ CACCTCTTGCATCGTCACCT 3’
SNM1B #1 reverse	5’ GGAGGGTTACGGTCATGGTC 3’
SNM1B #1 forward	5’ ACTTTTATCGGGACGCCGTT 3’
SNM1B #1 reverse	5’ GATTGGCATCGAGGAGGGTT’

Appendix II. Table of sequencing primers used in this study.

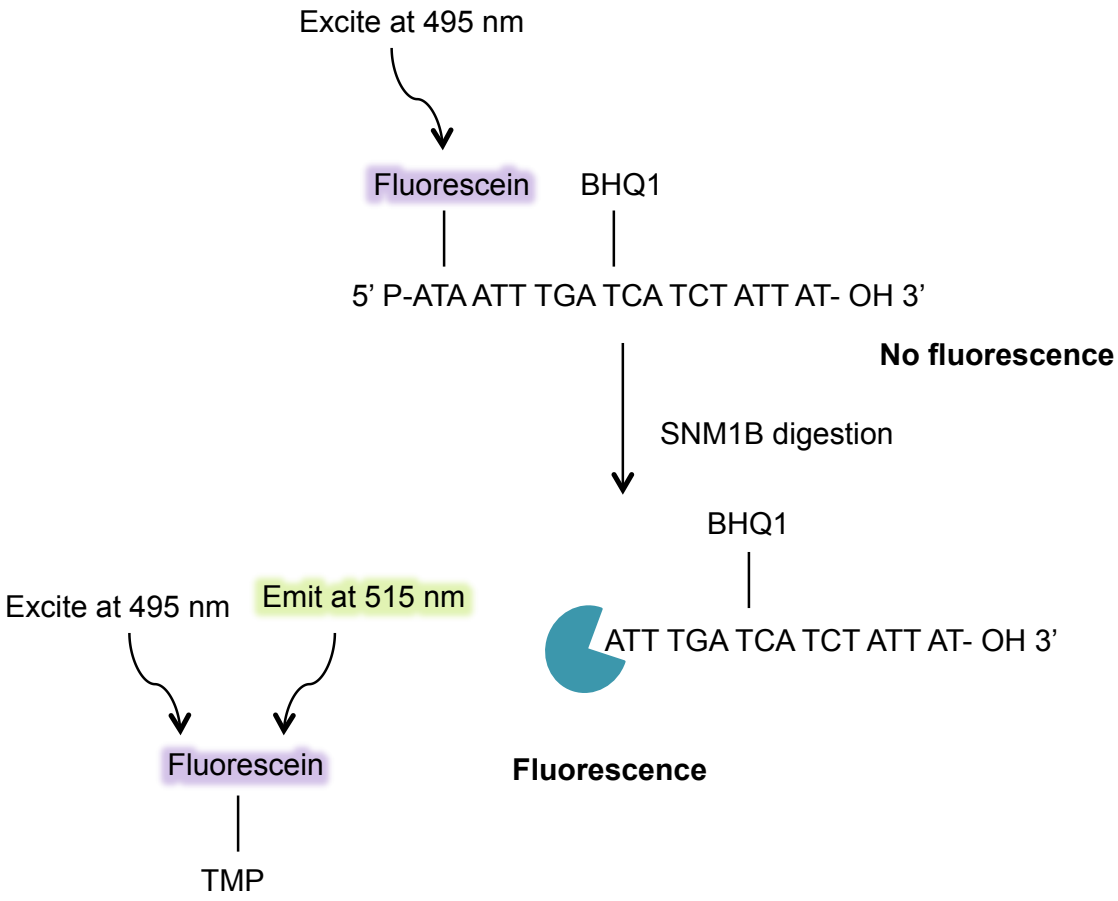
Sequencing of the entire SNM1B gene to confirm point mutations (H61Y and D35A/H36A) and ensure the fidelity of the genetic sequence	
SNM1B forward 2	5’ CTTCCCACAAGCCGAAAA 3’
SNM1B forward 3	5’ GGGTCCTCGTGCCTATGG 3’
SNM1B forward 4	5’ GGGGATTTAAAAGAGTTTTCTGC 3’
SNM1B reverse 5	5’ CCACCCAGGTCTGAAACTCC 3’
SNM1B reverse 6	5’ CAAAACTGCAGTGTGGGCTC 3’
SNM1B reverse 7	5’ TGGCTCCTTGGTCTCTTCCT 3’
Sequencing to validate correct annealing of the truncated SNM1B gene into pFB-CT10-HF-LIC/pFB-LIC-BseR1 vectors	
Forward	5’ TATTCATACCGTCCCACCA 3’
Reverse	5’ GGGAGGTTTTTTAAAGCAAGTAAA 3’
Sequencing to validate correct sgRNA sequence ligation into pX330 and pX458 CRISPR plasmids	
pX330 forward	5’ ACTATCATATGCTTACCGTAA 3’
pX330 reverse	5’ CCCATCGCTGCACAAAATAATT 3
Sequencing of CRISPR-Cas9 disrupted gDNA PCR products	
pX458 forward	5’ GCCTGGTATCTTTATAGTCC 3’
pX458 reverse	5’ CCGCCCCGCCTGGCGCGCGC 3’

Appendix III. sgRNA sequences for CRISPR-Cas9 mediated disruption of SNM1B.

Guide oligonucleotide sequences cloned into pX330	
Guide 1 top	5’ GTCCACTGCGATGGGCGTAT 3’
Guide 1 bottom	5’ ATACGCCCATCGCAGTGGACC 3’
Guide 5 top	5’ GCCACTGCGATGGGCGTATG 3’
Guide 5 bottom	5’ CATACGCCCATCGCAGTGGCC 3’
Guide 9 top	5’ GGCTCCAGAAGTCCACTGCGA 3’
Guide 9 bottom	5’ TCGCAGTGGACTTCTGGAGCC 3’
Guide oligonucleotide sequences cloned into pX458	
Guide 3 top	5’ GTGGAGTCGGTCTCCGAGAGC 3’
Guide 3 bottom	5’ GCTCTCGGAGACCGACTCCAC 3’
Guide 6 top	5’ GGGGGGCCACGAAAACTGATT 3’
Guide 6 bottom	5’ AATCAGTTTTCGTGGCCCCCC 3’

Appendix IV. Schematic representation of the real-time fluorescence-based nuclease assay.

Adapted from Lee, 2015.



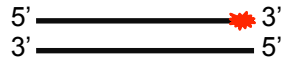
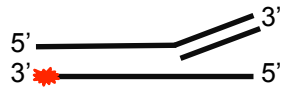
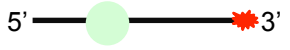
Appendix V. The sequences of DNA oligonucleotides used for DNA binding and nuclease assays.

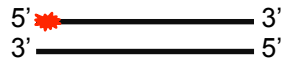
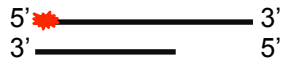
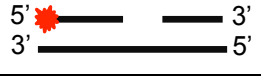
Code	DNA oligonucleotide sequence
1 (5' phosphate)	5' P-ATA AAT ATT TTT TAT TAA TAA TAG ATC ACC TTT CTT TCT CTT CTC CCC TT-OH 3'
1A (5' OH)	5' OH-AT AAA TAT TTT TTA TTA ATA ATA GAT CAC CTT TCT TTC TCT TCT CCC CTT-OH 3'
1B (3' biotin)	5' OH-ATA AAT ATT TTT TAT TAA TAA TAG ATC ACC TTT CTT TCT CTT CTC CCC TT-biotin 3'
2 (UA1)-	5' OH-AAG GGG AGA AGA GAA AGA AAG GTG ATC TAT TAT TAA TAA AAA ATA TTT AT-OH 3'
3 (UA2)-	5' OH-AAG GGG AGA AGA GAA AGA AAG G-OH 3'
4 (UA3) (5' OH)	5' OH-ATT ATT AAT AAA AAA TAT TTA T -OH 3'
4A (UA3) (5' phosphate)	5' P-ATT ATT AAT AAA AAA TAT TTA T -OH 3'
5 (5' OH) (BS41)	5' OH-TTC CCC TCC TCT CCT TCC TTC CTG ATC TAT TAT TAA TAA AAA ATA TTT AT-OH 3'
5A (5' phosphate)	5' P-TTC CCC TCC TCT CCT TCC TTC CTG ATC TAT TAT TAA TAA AAA ATA TTT AT-OH 3'
6 (HB1)-	5' OH-AAG GGG AGA AGA GAA AGA AAG G-OH 3'
DNA 20 (5' phosphate)	5' P-AT AAT AGA TGA TCA AAT TAT-OH 3'
DNA 20 (5' OH)	5' OH-AT AAT AGA TGA TCA AAT TAT-OH 3'
RNA 20-	5' OH-AUA AUU UGA UCA UCU AUU AU-OH 3'
DNA (U) 20-	5' OH-AUA AUU UGA UCA UCU AUU AU-OH 3'
Gap (left)	5' OH-AT AAA TAT TTT TTA TTA ATA ATA GATC ACC-OH 3'
Gap (right)	5' OH-TTT CTT TCT CTT CTC CCC TT-OH 3'
Gap (-1)	5' OH-TT CTT TCT CTT CTC CCC TT-OH 3'
Gap (-2)	5' OH-T CTT TCT CTT CTC CCC TT-OH 3'

Gap (-3)	5’ OH-CTT TCT CTT CTC CCC TT-OH 3’
Gap (-4)	5’ OH-TT TCT CTT CTC CCC TT-OH 3’
Gap (-5)	5’ OH-T TCT CTT CTC CCC TT –OH 3’
Tel top (5’ phosphate)	5’ P-ATC GTC CTA GCA AGG GGT TAG GGT TAG GGT TAG GGT TAG GGT TAG GGT TA-OH 3’
Tel top (5’ OH)	5’ OH-ATC GTC CTA GCA AGG GGT TAG GGT TAG GGT TAG GGT TAG GGT TAG GGT TA-OH 3’
Tel bottom	5’ P-CCT AAC CCC TTG CTA GGA CGA T-OH 3’
DNA 100	5’ OH-AT AAA TAT TTT TTA TTA ATA ATA GATC ACC TTT CTT TCT CTT CTC CCC TTAT AAA TAT TTT TTA TTA ATA ATA GATC ACC TTT CTT TCT CTT CTC CCC TT-OH 3’
DNA 60 (5’ OH)	5’ OH-TT ATA AAT AAA ATA TTA AAACCA ATT TGA TCA TCT ATT ATA TAT TTA TTA ATT TAT TTT T– OH 3’
DNA 60 (5’ phosphate)	5’ P-TT ATA AAT AAA ATA TTA AAACCA ATT TGA TCA TCT ATT ATA TAT TTA TTA ATT TAT TTT T– OH 3’
DNA 40	5’ OH-T TTA TTA ATA ATA GATC ACC TTT CTT TCT CTT CTC CCC TT-OH 3’
DNA 30	5’ OH-ATA GATC ACC TTT CTT TCT CTT CTC CCC TT-OH 3’
DNA 25	5’ OH-TC ACC TTT CTT TCT CTT CTC CCC TT-OH 3’
DNA 20	5’ OH-TTT CTT TCT CTT CTC CCC TT-OH 3’
8oxoG (CTRL)	5’ P-ATA ATT TGA TCA TCT ATT ATA- OH 3’
8-oxo-G (internal)	5’ P-ATA ATT T[8oxoG]A TCA TCT ATT ATA-OH 3’
8-oxo-G (5’)	5’ P-[8oxoG]TA ATT TGA TCA TCT ATT ATA-OH 3’
Fluorescein-DNA	5' P- ATA ATT TGA TCA [Fluorescein-T]CT ATT ATA -OH 3'
Fluor-BHQ-DNA	5'P- A[Fluorescein-T]A ATT TGA [BHQ1-T]CA TCT ATT AT -OH 3'
Phosphorothioate (terminal)	5’P-A(phosphorothioate)TA AAT TGG ATC ATC TAT TA-OH 3’
Phosphorothioate (internal)	5’P-ATA AAT TTG A(phosphorothioate)TC ATC TAT TA-OH 3’

Appendix VI. Schematic representation of the DNA substrates utilised in this study.

The code for the oligonucleotide sequences used to assemble the structures are displayed in Appendix V. The asterisk indicates the radiolabelled strand.

3' labelled substrates (α-^{32}P-dATP)			
Annealed DNA sequences	**Substrate Structure**	**Description**	**Figure(s)**
1*		Single-stranded 51 nt DNA	3.21, 3.22, 3.23, 3.24, 3.29, 3.30, 4.7, 4.18
1A*		Single-stranded 51 nt DNA with a 5' OH	4.7
1* + 2		dsDNA	3.21, 3.22, 3.23, 3.29
1* + 3		5' overhang	3.21, 3.22, 3.23, 3.29
1* + 4		3' overhang	3.21, 3.22, 3.23, 3.29
1A + 5A*		Splayed arm	3.21, 3.22, 3.23, 3.29
1A + 5A* + 3		Leading strand flap	3.21, 3.22, 3.23, 3.29
1A + 5A* + 6		Lagging strand flap	3.21, 3.22, 3.23, 3.29
DNA 20 (5' phosphate)*		Single-stranded 21 nt DNA	3.26, 3.31. 3.32, 3.34, 4.8, 4.20
DNA 20 (5' OH)*		Single-stranded 21 nt DNA with 5' OH	4.8
Tel. top (5' phosphate)*		Single-stranded 51 nt DNA (telomeric)	3.30
1 + 4A*		3' overhang with 4A labelled	3.30
Tel. top (5' OH) + tel. bottom*		3' overhang (telomeric) with tel. bottom labelled	3.30
8-oxo-G (internal)*		Single stranded 21 nt DNA with internal 8-oxo-G	3.31
8-oxo-G (5')*		Single stranded 21 nt DNA with 5' 8-oxo-G	3.31
DNA 60 (5' phosphate)*		Single stranded 61 nt DNA	3.34

Phosphorothioate (terminal)		Single stranded 21 nt DNA with a 5’ terminal phosphorothioate bond	4.8
Phosphorothioate (internal)		Single stranded 21 nt DNA with an internal phosphorothioate bond	4.8
5’ labelled substrates (γ-^{32}P-dATP)			
1A*		Single-stranded 50 nt DNA	6.2, 6.11, 6.18, 6.24, 6.25 6.26, 6.29, 6.30, 6.37
1A* + 2		dsDNA	6.26, 6.31
1A* + 3		5’ overhang	6.30
1A* + 4		3’ overhang	6.30
1A* + 5		Splayed arm	6.2, 6.30, 6.31, 6.37
1A* + 5 + 3		Leading strand flap	6.30, 6.31
1A* + 5 + 6		Lagging strand flap	6.30, 6.31
1A* + 5 + 3 + 6		Replication fork	6.30, 6.31
1B*		1B* anneals to generate all above substrate structures, with the addition of a 3’ biotin on the labelled strand	6.31
Gap (left)* + gap (right) + 2		Nicked duplex DNA	6.33, 6.34
Gap (left)* + gap (-1) + 2		Gapped duplex DNA	6.33, 6.34
Gap (left)* + gap (-2) + 2		2 nt gapped duplex DNA	6.34
Gap (left)* + gap (-3) + 2		3 nt gapped duplex DNA	6.34
Gap (left)* + gap (-4) + 2		4 nt gapped duplex DNA	6.34

Gap (left)* + gap (-5) + 2		5 nt gapped duplex DNA	6.34
1A (5OH)*, DNA 60 (5’ OH)*, DNA 100*, DNA 40*, DNA 30*, DNA 25*, DNA 20*		Single stranded DNA of varying lengths	6.35
DNA 20 (5’ OH)*		Single stranded 20 nt DNA	6.38
RNA 20*		Single stranded 20 nt RNA	6.38
DNA 20 (5’ OH)* + RNA 20		DNA* RNA hybrid	6.38
DNA 20 (5’ OH) + RNA 20*		DNA RNA* hybrid	6.38
DNA (U) 20*		Single stranded 20 nt DNA with uracil	6.38
Fluorescent substrates			
Fluorescein-DNA		Single stranded 21 nt DNA with internal fluorescein-T	3.27, 6.27
Fluor-BHQ-DNA	See Appendix IV.	Single stranded 20 nt DNA substrate with terminal fluoresceinT and internal BHQ1	3.28

Appendix VII. Data collection and refinement statistics for WT (form I), H61Y, and WT (nucleotide form).

	WT form I	H61Y	WT nucleotide form
Space group	P 4_1 2_1 2	P 4_1 2_1 2	P 1
Cell dimensions, *a,b,c* (Å)	90.4, 90.4, 104.2	90.7, 90.7, 104.1	50.1, 54.2, 72.1
Angles α,β,γ (°)	90, 90, 90	90, 90, 90	75.4, 74.3, 63.1
Wavelength (Å)	0.979	0.979	0.979
Resolution (Å)	45.2 – 2.80 (2.95 – 2.80)	45.4 – 3.10 (3.31 – 3.10)	68.6 – 1.80 (1.84 - 1.80)
R_{merge}	0.18 (2.13)	0.15 (2.17)	0.06 (0.62)
$R_{p.i.m.}$	0.04 (0.44)	0.05 (0.61)	0.05 (0.52)
I/σI	11.8 (1.6)	13.8 (1.3)	7.9 (1.3)
CC1/2	0.99 (0.67)	0.99 (0.58)	0.99 (0.65)
Completeness (%)	100 (100)	100 (100)	93.3 (95.5)
Multiplicity	25.3 (25.3)	12.7 (13.3)	2.2 (2.3)
No. Unique reflections	11188 (1588)	8386 (1474)	55714 (3368)
Refinement statistics			
Resolution	45.2 – 2.80	45 – 3.1	44 – 1.80
R_{work}/R_{free} (%)	27.8/33.6	27.7/33.5	17.7/22.3
No. atoms			
Protein	2459	2460	5200
Solvent	-	-	566
Ligand/ion	1	1	139
Average B factors (Å^2)			
All atoms	76	88	32
Protein	76	88	31
Solvent	-	-	38
Ligand/ion	120	157	49
Wilson B	86	101	24
R.M.S. deviations			
Bond lengths (Å)	0.002	0.005	0.007
Bond angles (°)	0.6	0.7	0.9
Ramachandran plot			
Favoured (%)	94.8	96.1	97.7
Allowed (%)	99.0	99.3	100