8-23-2022 2:00 PM

Dual Functions of Interstrand Crosslink Repair Nuclease SNM1A

Ryan Grainger, *The University of Western Ontario*

Supervisor: Junop, Murray S., *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Biochemistry

Abstract

Interstrand crosslinks (ICL) are a highly cytotoxic form of DNA damage, covalently linking opposing strands of DNA. ICLs disrupt essential cellular processes requiring strand separation, including transcription and replication. Consequently, lesion recognition and removal are critical to prevent chromosomal aberrations, mitotic catastrophe and apoptosis. ICL repair requires the coordination of a complex network of nucleases necessary for remodelling, unhooking and resolving repair intermediates. While many nucleases participate, little is known about where and when each nuclease acts. SNM1A is a dual-function exonuclease and endonuclease necessary for ICL repair. Where SNM1A is absent, cells accumulate irreparable double-strand breaks and exhibit reduced survival following treatment with ICL-inducing compounds. Although essential for fidelitous repair of ICLs, it is unclear where SNM1A functions and which intermediate(s) it processes.

The primary objectives of this thesis were to examine the capacity and preferences of SNM1A nuclease activities *in vitro*, investigate which nuclease activities contribute to ICL repair and develop small molecule inhibitors of SNM1A. To examine functional preferences, we characterized SNM1A nuclease activities on various potential repair intermediates. While SNM1A exonuclease activity was generally more robust than the endonuclease function, translesional processing constituted the rate-limiting step during digestion of an ICL-containing stalled replication fork mimic. Further, structural models of SNM1A and its yeast homolog were generated to enable mutagenic isolation of nuclease functions. SNM1A exonuclease and endonuclease processing were selectively disrupted by substituting residues in the phosphate-binding pocket and novel DNA binding groove, respectively. In a yeast model, neither separation-of-function mutant was sufficient to facilitate ICL repair, indicating that both nuclease activities are necessary. Finally, an *in silico* high-throughput screen identified four specific inhibitors of SNM1A with low micromolar potency. Cumulatively, experiments presented in this thesis expand the potential roles of SNM1A in ICL repair and provide promising lead compounds to target SNM1A *in vivo*.

Summary for a Lay Audience

Interstrand crosslinks (ICLs) are formed by chemicals that irreversibly bind both strands of a DNA helix. ICL damage prevents the separation of DNA strands, disrupting essential cellular activities, including reading genes to produce protein and copying DNA for cell division. As such, failure of a cell to quickly find and remove ICL damage can result in significant loss of genetic material, cancer or cell death. Removal of ICL damage requires the recruitment of multiple nucleases, which are molecular scissors that cleave distinct DNA structures. While different nucleases have been shown to participate in ICL repair, questions remain regarding where, when and how these nucleases function. SNM1A is a dual-function nuclease, able to cut DNA from a free end (exonuclease activity) or within a strand of unpaired DNA (endonuclease activity). Although previous reports demonstrated that SNM1A is needed for ICL repair, it remains unclear what DNA structure(s) SNM1A cuts, and with which activity. The primary objectives of this thesis were to determine how DNA structure impacts SNM1A processing and investigate the SNM1A nuclease activities required for repair. Experiments presented here demonstrate that both nuclease functions are necessary for repair. Further, identified substrate preferences of SNM1A suggest potential intermediates SNM1A may act on during repair. Finally, four molecules were found to specifically inhibit SNM1A nuclease activities in a test tube. Developing small molecules will help in future experiments to determine where and when SNM1A is acting in a cell.

Keywords

SNM1A, Pso2, DNA Repair, Interstrand Crosslinks, Fanconi Anemia, Nuclease, High-Throughput Screen, Inhibitors, Mutagenesis, Separation-of-Function

Co-Authorship Statement

All experiments and data analysis presented in this thesis were completed by the author, with exceptions noted below. All text and figures were generated by the author, with input and editing from Dr. Murray Junop.

Chapter 2: Dr. Simon Huang generated the initial bacterial expression plasmids. Sam Chu assisted with the generation of baculovirus for SNM1A expression in insect cells. Dr. Beverlee Buzon assisted in substrate design, crosslinking and purification of DNA in Figure 2.9.

Chapter 3: The high-throughput screen to identify SNM1A bioactive inhibitors was completed at McMaster Center for Microbial Chemical Biology by Dr. Beverlee Buzon and Cameron Rzadki. Optimization of SNM1B purification was completed by Braeden Medeiros. SNM1A-specific *in silico* screening was completed by Patrick Melo.

Chapter 4: Dr. Tyler Blue and Kristina Timcevska assisted in the purification of select SNM1A mutants in Figure 4.3. Claire Zhang assisted with yeast survival experiments in Figure 4.14.

Acknowledgments

Completing this thesis required the support, knowledge and creativity of many people. I would like first to thank my supervisor, Dr. Murray Junop. For years, he supplied endless ideas, advice and optimism. He has worked hard to create a space for science to thrive, and I am lucky to have him as a mentor. My projects would not have succeeded without the help of team ICL, who spent many late nights purifying proteins that refused to cooperate. Thank you to Dr. Beverlee Buzon for the many brainstorming (and subsequent troubleshooting) sessions. And to the students and colleagues who helped make those experiments come to life, including Dr. Tyler Blue, Nikki Case, Claire Zhang, Justin Ching-Johnson, and soon-to-be Drs. Braeden Medeiros and Patrick Melo. Thank you to Kun Zhang, Brianna Kaplanis, Michael Kovacevic and Tedson Nie for helping me realize the dream of SNM1A electron density and to Drs. Mac Mok, Chris Brown and Rob Szabla for teaching me how to transform diffraction patterns into protein structures. Finally, I would like to thank my family. Jackie, thank you for your patience, moral support and fantastic editing skills. While I buried myself in the lab and library, you kept the rest of the world up and running. And to my parents, who continue to inspire me to reach higher, work harder and be better.

Table of Contents

List of Figures

List of Tables

List of Appendices

Summary of Abbreviations

53BP1	P53 Binding Protein 1
ADME	Absorption, Distribution, Metabolism and Excretion
ALDH2	Aldehyde Dehydrogenase 2
APC3	Anaphase Promoting Complex Subunit 3
APE1	DNA-Apurinic(Apyrimidinic) Endonuclease
ATM	Ataxia Telangiectasia Mutated
ATP	Adenosine Triphosphate
ATR	Ataxia Telangiectasia and Rad3-Related
β-CASP	β-CPSF-Artemis-SNM1A-Pso2
BER	Base Excision Repair
BIR	Break-Induced Repair
BLM	Bloom Syndrome Helicase
BRCA 1,2	Breast Cancer Genes
BSA	Bovine Serum Albumin
CDC 27, 45, 48	Cell Division Cycle
Chk1	Checkpoint Kinase 1
CMCB	Centre for Microbial Chemical Biology
CMG	CDC45, MCM2-7, GINS
CPD	Cyclobutane Pyrimidine Dimers
CPSF	Cleavage and Polyadenylation Specificity Factor
CRL	Cullin/RING Ubiquitin Ligase
CS A, B	Cockayne Syndrome Group
CtIP	C-Terminal Binding Protein
CV	Column Volume
DBB2	Damage Specific DNA Binding 2
DIDS	4,4'-Diisothiocyanostilbene-2,2'-disulfonic Acid
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid

dNTP	Deoxyribonucleoside Triphosphate
ds	Double Strand
DSB	Double Strand Break
DSF	Differential Scanning Fluorimetry
DTT	Dithiothreitol
ECL	Enhanced Chemiluminescence
EDTA	Ethylenediaminetetraacetic Acid
EGCG	Epigallocatechin Gallate
EGFP	Enhanced Green Fluorescent Protein
ELAC2	ElaC Homolog Protein 2
EME1	Essential Meiotic Structure Specific Endonuclease 1
ERCC 1, 4	Excision Repair-Defective Cross-Complementation Groups
EtBr	Ethidium Bromide
Exo1	Exonuclease 1
FA	Fanconi Anemia
FAAP 24, 100	FA-Associated Proteins
FAN1	FA-Associated Nuclease 1
FANC A-C, D1, D2, E-G, I, L, M, P, U, T, Q, S	FA Complementation Groups
FBS	Fetal Bovine Serum
FEN1	Flap Structure-Specific Endonuclease 1
GG	Global-Genome
GINS	Go-Ichi-Ni-San
GroEL	Growth Defect in λ Protein E (Long Transcript)
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic Acid
HJ	Holliday Junction
HMGB1	High Mobility Group Box 1
HR	Homologous Recombination
HRP	Horseradish Peroxidase

Hrq1	Homologous to RecQ 1
HTS	High Throughput Screen
IC_{50}	Half-Maximal Inhibitory Concentration
ICL	Interstrand Crosslink
I-Dmo1	*Desulfurococcus mobilis* homing endonuclease 1
Ig G	Immunoglobulin G
IMAC	Immobilized Metal Affinity Chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IR	Ionizing Radiation
IRES	Internal Ribosome Entry Site
LB	Lennox Broth
M 1,2	Metal Coordination Sites
MALS	Multi-Angle Light Scattering
MβL	Metallo-β-Lactamase
MCE	Mixed Cellulose Esters
MCM 2-7	Minichromosome Maintenance Complex Component 2-7
m-DHPG	Methyl-7-deshydroxypyrogallin-4-carboxylic Acid
mer	Oligomer
MHF 1, 2	FANCM Interacting Histone-Fold
MLH1	MutL Homolog 1
MMC	Mitomycin C
MMR	Mismatch Repair
MMS	Methyl Methanesulfonate
MRN	Mre11, Rad50, Nbs1 Complex
mRNA	Messenger RNA
MSH 2, 3, 6	MutS Homologs
MUS81	MMS-UV Sensitive Gene Clone 81
MWCO	Molecular Weight Cut Off
NEIL3	Nei-Like DNA Glycosylase 3

NER	Nucleotide Excision Repair
NGC	Next Generation Chromatography
NHEJ	Non-Homologous End Joining
NLS	Nuclear Localization Signal
nt	Nucleotide
OD	Optical Density
PAGE	Polyacrylamide Gel Electrophoresis
PAINS	Pan-Assay Interference Compounds
PAR	Poly-(ADP-Ribose)
PBS	Phosphate Buffered Saline
PBZ	PAR-Binding Zinc Finger
PCNA	Proliferating Cell Nuclear Antigen
PCR	Polymerase Chain Reaction
PEG	Polyethylene Glycol
PI3K	Phosphoinositide 3-Kinase
PIAS1	Protein Inhibitor of Activated STAT 1
PIP	PCNA-Interacting Peptide
PKcs	Protein Kinase Catalytic Subunit
PMS2	Post-Meiotic Segregation 2
PMSF	Phenylmethylsulfonyl Fluoride
Pol I, 32, δ, ε, ζ, η, κ	DNA Polymerase
PrimPol	Primase and DNA-Directed Polymerase
Pso2	Psoralen-Sensitive 2
PTM	Post Translational Modification
PVDF	Polyvinylidene Difluoride
RAD 1, 2, 3, 10, 18, 23B, 25, 26, 51, 52	Radiation-Sensitive Mutants
RDR	Replication-Dependent Repair
RecA	Recombination-Deficient Mutant A

RecQL4	RecQ-Like Helicase 4
REOS	Rapid Elimination of Swill
Rev 1,3	Reversionless Mutant
RIR	Replication-Independent Repair
RNA	Ribonucleic Acid
RNAP	RNA Polymerase
RPA	Replication Protein A
RPM	Rotations Per Minute
RRI	Robarts Research Institute
Sak1	SNF1-Activating Kinase 1
SAN1	Sanataxin-Associated Nuclease 1
SAR	Structure-Activity Relationship
SDF	Structure Data File
SDS	Sodium Dodecyl Sulfate
SEC	Size Exclusion Chromatography
SFM	Serum-Free Media
siRNA	Small-Interfering RNA
SLX 1, 4	Synthetic Lethal of Unknown Function
SNM1 A B C	Sensitive to Nitrogen Mustard 1
SOD1	Superoxide Dismutase 1
ss	Single Strand
SSA	Single Strand Annealing
SUMO	Small Ubiquitin-Like Modifier
TB	Terrific Broth
TBS	Tris-Buffered Saline
TC	Transcription-Coupled
TCE	2,2,2-Trichloroethanol
TCEP	Tris(2-carboxyethyl)phosphine
TEV	Tobacco Etch Virus

TFDG	Theaflavin Digallate
TFII H, S	Transcription Factor II
TLS	Translesion Synthesis
T_m	Half-Maximal Melting Temperature
TRAIP	TRAF Interacting Protein
UBZ	Ubiquitin Binding Zinc Finger
UDS	Unscheduled DNA Synthesis
UHRF1	Ubiquitin-Like with PHD and Ring Finger Domain 1
UTR	Untranslated Region
UV	Ultraviolet
UvrABC	UV-Sensitive Mutants
UVSSA	UV Stimulated Scaffold Protein A
XP A-G	Xeroderma Pigmentosum Complementation Groups
XRCC 2, 3	X-Ray Repair Cross-Complementing
YPD	Yeast-Extract Peptone Dextrose
ZBG	Zinc-Binding Group

Chapter 1
Introduction

1.1 Thesis Outline

Interstrand crosslink repair requires the complex coordination of disparate repair pathways to resolve covalent linkages between opposite strands in a DNA helix. As such, successful and timely repair depends on the contribution of a diverse set of factors. This thesis aims to improve our understanding of how one such factor, SNM1A, participates in ICL repair. **Chapter 1** consists of an extensive literature review describing the current research landscape regarding ICL damage, repair and SNM1A. As the role of SNM1A in repair is unclear, outlining past research in detail facilitates later alignment of newly identified *in vitro* SNM1A preferences with potential biological functions. In **Chapter 2**, *in vitro* investigation of SNM1A substrate preferences and molecular mechanism

enables functional characterization, informing potential roles of SNM1A in repair. In **Chapter 3**, specific small molecule inhibitors of SNM1A are identified. The characterization of multiple lead compounds serves as a platform for the further development of effective *in vivo* inhibitors. In **Chapter 4**, separation-of-function mutants establish that both exonuclease and endonuclease activity of SNM1A are necessary for repair. Finally, in **Chapter 5**, questions stemming from work presented throughout this thesis are discussed with approaches to address them. The work presented here provides critical insight into how SNM1A contributes to repair and how those roles can be exploited to overcome chemoresistance with small molecule inhibitors.

1.2 Interstrand Crosslink Damage

Interstrand crosslinks (ICL) are a particularly lethal form of DNA damage. ICLs link two opposing strands of DNA, either directly between opposing bases or offset by one or more nucleotides[1,2]. The DNA distortion induced by covalent linkage of the strands ranges from absent to extreme, depending on the chemistries of individual crosslinkers and where they react on the nucleobases[3,4]. ICLs induce a severe block on all cellular processes relying on the separation of DNA strands, primarily replication and transcription. ICLs are characteristically clastogenic, inducing gross chromosomal abnormalities such as radial chromosomes, deletions and translocations[5]. While any bifunctional chemical can induce an ICL in principle, biologically relevant crosslinks arise primarily from limited sources.

1.2.1 Exogenous Sources

ICLs induced in rapidly transcribing and dividing cells are particularly cytotoxic[2,6]. Therefore, ICLs are most frequently associated with purposeful exogenous exposure to treat hyperproliferative conditions, primarily as chemotherapy. There are multiple

distinct classes of ICL-inducing chemotherapeutics, generated either synthetically or by exploiting natural crosslinking compounds.[2,7]

Among the most frequently used chemotherapeutics include synthetic crosslinkers such as nitrogen mustards and platins. Nitrogen mustards, including mechlorethamine, chlorambucil and cyclophosphamide, comprise a common class of ICL-inducing drugs currently used to treat leukemia, multiple myeloma and lymphoma[2]. Nitrogen mustards contain a bifunctional N,N-bis-(2-chloroethyl)amine, which reacts at N^7-dG to form a 1,3 ICL across 5'GNC sequences (**Figure 1.1A**)[5,7]. As the crosslinker is too short to span three nucleotides, it necessarily induces a modest bend in the damaged DNA[7,8]. Platinum compounds form another class of ICL-inducing chemotherapeutics. Clinically, many derivatives are used including oxaliplatin, carboplatin and cisplatin, in particular, to treat testicular, ovarian and colorectal cancers, among others[2]. Cisplatin (cis-diamminedichloroplatinum) is a square planar platinum coordination complex containing two reactive cis-chlorides[5,7]. Like nitrogen mustards, cisplatin reacts predominantly at N^7-dG, instead forming a 1,2 ICL within 5'GC sequences (**Figure 1.1B**). Cisplatin ICLs produce substantial helical distortion, inducing extrahelical displacement of opposing cytosines[5,7,8].

Crosslinking compounds are also formed naturally as defense mechanisms in bacteria and plants. Mitomycin C (MMC) is a natural antibiotic produced by *Streptomyces caespitosus* and is frequently used to treat gastrointestinal cancers including gastric, pancreatic, biliary and colorectal[5]. MMC requires intracellular activation via quinone reduction, where it can then react with the exocyclic N^2-dG amine to form a 2,1 ICL across 5'-CG (**Figure 1.1C**). MMC induces slight distortion to accommodate the compound within the DNA minor groove[5,7,8]. Furocoumarins comprise another class of naturally occurring ICL-inducing compounds produced by some plant species. Clinically, psoralen (and other furocoumarin analogues) are used to treat proliferative skin disorders, including psoriasis and vitiligo[5,7]. Psoralen is a tricyclic planar compound, which forms interstrand crosslinks via cycloaddition with the 5,6' unsaturated bonds of opposing thymine bases following DNA intercalation and UV-A

photoactivation. Psoralen forms a stable 1,2 ICL within 5'TA sequences, resulting in moderate helical destabilization and minimal DNA distortion (**Figure 1.1D**)[5,7,8].

All crosslinking compounds used clinically result in a variety of DNA damage, where ICLs often reflect only a small minority (as low as <1% of total damage). For instance, cisplatin preferentially forms intrastrand crosslinks and monoadducts[7]. Though this damage does require repair, it does not present as significant a roadblock to the replisome[9]. The severe cytotoxicity of ICL-inducing compounds specifically arises from the interstrand crosslink. This specific cytotoxicity has been repeatedly observed as increased cellular tolerance of ICL-incompetent analogues, which can generate adducting DNA damage but cannot form ICLs[5].

1.2.2 Endogenous Sources

Genetic disruption of ICL repair results in a severe increase in cancer predisposition and accumulation of chromosomal abnormalities[2,10]. This intrinsic chromosomal instability demonstrates that there must be a consistent source of endogenous ICLs. Significant research effort has identified numerous cellular by-products which generate crosslinks *in vitro*. Importantly, ICLs represent a small minority of the DNA damage induced by these metabolic intermediates or by-products. As such, their biological relevance remains somewhat ambiguous, where many of these ICLs are unstable *in vitro*.

Aldehydes, in a variety of forms, induce ICLs *in vitro*[11–14] and *in vivo*[15–17]. Aldehydes result directly as metabolic intermediates of ethanol[18], lipid peroxidation (frequently as enals)[12,13] and DNA maintenance[19,20]. Though chemical diversity of aldehydes is significant *in vivo*, crosslinking reactions often proceed similarly – targeting nucleobase exocyclic amines (frequently N^2-dG). Damage induced by simple aldehydes result in aldimine intermediates, which in turn facilitate attack by a second aldehyde equivalent, tautomerized to its enolate form[11,14,21]. More complex α,β-unsaturated aldehydes instead directly react via Michael addition[12,13,21]. Each process culminates in a

monoadduct terminated by a reactive aldehyde. In contrast, abasic sites generated during DNA maintenance by glycosylases produce a deoxyribose in equilibrium between the aldose and furanose form. As a result, the terminal aldehyde of the aldose sugar is positioned to directly crosslink opposing nucleobases (**Figure 1.1E**)[19,20]. Aldehyde-induced crosslinks form between opposing exocyclic amines and exist in equilibrium between hemiaminal and aldimine. *In vitro*, chemical reduction of the crosslink dramatically improves stability, though the biological relevance is unclear[13]. The chemical necessity for correct positioning of exocyclic amines within the DNA gives rise to specific sequence requirements, depending on the reactive aldehyde.

Studies using combination genetic disruptions in mice clearly demonstrate the connection between ICL repair and the cellular turnover of endogenous aldehydes. Disruption of ICL repair (Fancc$^{-/-}$ or Fancd2$^{-/-}$) and increased oxidative stress (Sod1$^{-/-}$)[22] or decreased aldehyde oxidation (Aldh2$^{-/-}$)[17,18,23] dramatically stimulates chromosomal abnormalities and cellular hypersensitivity to simple aldehydes. These severe phenotypes were moderated or absent in individual disruptions, suggesting that these pathways function cooperatively. Collectively, aldehyde detoxification lowers the cellular load of ICL-inducing by-products, where ICL repair recognizes and removes any residual damage. In mice, disruption of a single mechanism was tolerated while dual-disruption was frequently embryonically lethal[10].

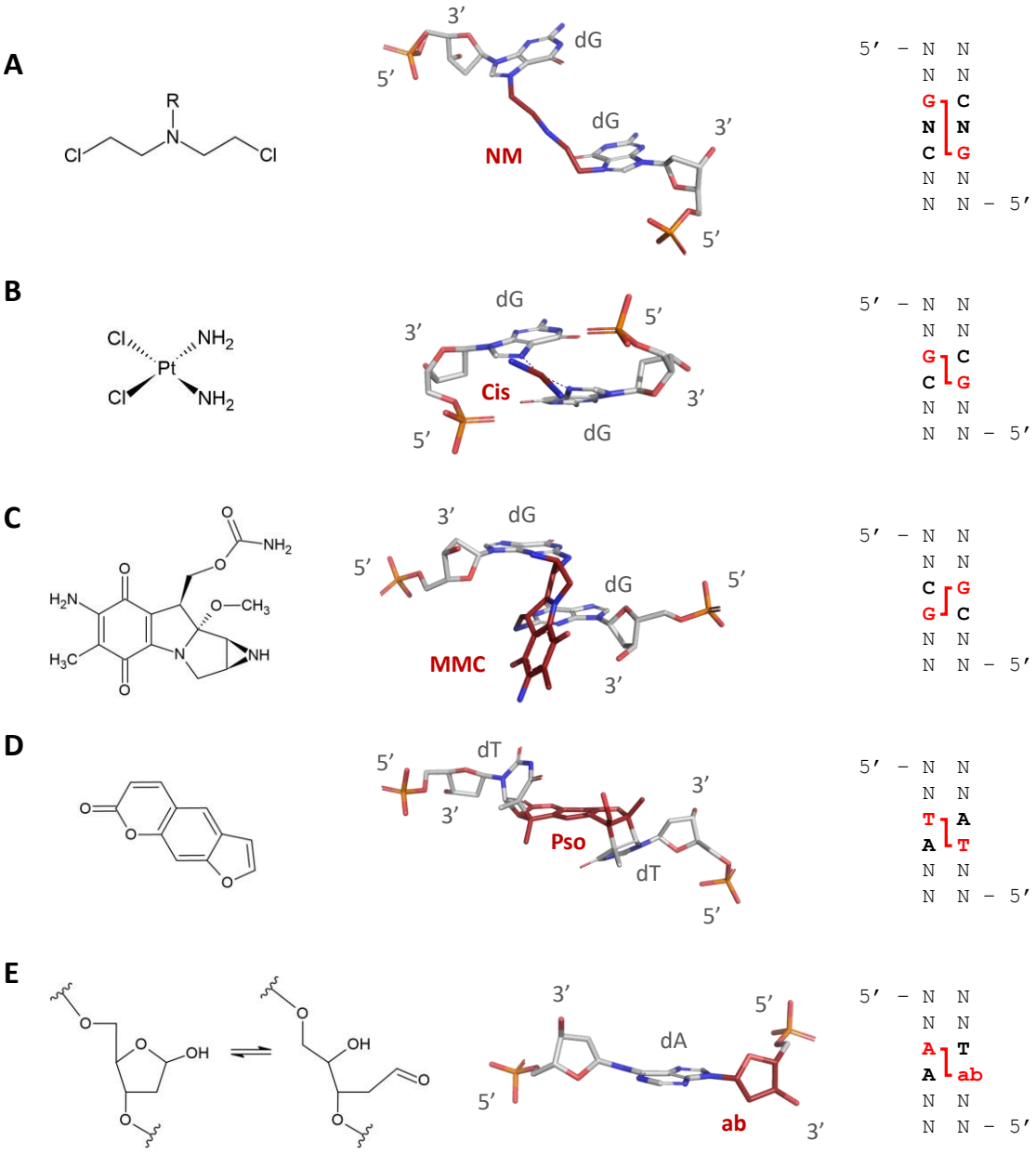


Figure 1.1. Chemical and DNA structure of common crosslinkers.

[A] Nitrogen mustard (NM, model of crosslink), [B] cisplatin (Cis, PDB: 1DDP)[24], [C] mitomycin C (MMC, model of crosslink adapted from PDB: 199D)[25], [D] psoralen (Pso, PDB: 204D)[26] and [E] abasic site (ab; PDB: 6XAH)[27]. Crosslinking compounds are coloured in red.

1.3 Interstrand Crosslink Repair

Most DNA lesions are repaired by a dedicated, linear pathway. Interestingly, examination of different interstrand crosslinkers using distinct experimental systems identified roles for nearly all DNA repair pathways during ICL repair. This includes enzymes from base excision repair (BER)[28,29], mismatch repair (MMR)[30,31], nucleotide excision repair (NER)[32,33], single-strand annealing (SSA) or break-induced repair (BIR)[34,35], translesion synthesis (TLS)[36,37] and homologous recombination (HR)[38–40]. Further, higher eukaryotes have evolved the Fanconi Anemia (FA) pathway to help coordinate ICL lesion resolution[1,2,6]. The complexity of repair seems to reflect both the toxicity and diversity of ICL lesions. The chemical diversity of crosslinkers has a direct impact on how the lesion is recognized and repaired. In general, the mechanism of ICL repair depends on the helical distortion induced by the crosslink[3], the genomic location of the damage[41] and replicative status of the cell[42].

This sub-chapter has been organized into three sections. As ICL repair requires substantial coordination with the NER pathway, section **1.3.1** discusses global-genome (GG-NER) and transcription-coupled (TC-NER) nucleotide excision repair. Next, section **1.3.2** summarizes mechanisms of replication-independent ICL repair (RIR), including GG- and TC-ICL repair pathways. Finally, section **1.3.3** summarizes the current understanding of replication-dependent ICL repair (RDR), including the FA-pathway and alternative mechanisms of ICL resolution.

1.3.1 Nucleotide Excision Repair

DNA in an average mammalian cell experiences thousands of individual damage events per day[43]. Damage resulting in bulky adducts disrupts elongation by both RNA and DNA polymerases. Rapidly identifying damaged or modified nucleotides is essential for proper transcription, DNA replication and maintenance of genomic fidelity. Nucleotide excision repair (NER) encompasses a highly conserved pathway primarily

responsible for recognizing and removing bulky adducts on a single strand of DNA. Most base damage outside of nucleobase mismatch, smaller alkylation/oxidation products or base deamination are removed by NER[44]. Ideal lesions requiring NER processing include UV radiation-induced photoproducts, intrastrand crosslinks and other large adducting chemicals[43,44].

1.3.1.1 Global-Genome Nucleotide Excision Repair (GG-NER)

Global-genome nucleotide excision repair (GG-NER) is activated by a recognition complex, constantly probing the genome for sources of helical distortion[45,46]. The presence of damage, especially bulky damage, impacts the thermodynamic stability of the DNA helix. XPC, in complex with RAD23B, recognizes this local destabilization and forcibly melts the helix, producing a small region of single-stranded DNA[46,47]. XPC loads onto the DNA opposite the damaged base and serves as an anchor for NER factor recruitment and complex assembly (**Figure 1.2A**)[48]. Indirect lesion recognition by XPC explains the versatility of NER, using a single factor to recognize a wide array of damage. As a direct consequence of the sensing mechanism, damage which does not result in destabilization of the DNA helix is not removed by NER.

Once anchored across from a site of damage, XPC directly recruits the multi-protein complex TFIIH. TFIIH is a critical remodelling factor involved in transcription initiation and NER, functionally composed of a 5 to 3' helicase (XPD) and DNA-dependent ATPase/ translocase (XPB)[49,50]. The small XPC-induced bubble is remodelled by TFIIH; opened further 3' by the XPD helicase while wedged open 5' by XPB translocase activity. The XPD helicase is particularly sensitive to damaged nucleotides and facilitates primary damage verification. TFIIH loads XPD onto the damaged strand, 5' of the potential lesion. XPD stalling at the damage site is required for pre-incision complex assembly to continue[43,44]. If XPC aberrantly interacts with undamaged DNA, the failure of XPD to stall will terminate complex assembly, preventing unnecessary DNA incision[50].

Following remodelling around the lesion, sufficient ssDNA is available in the bubble for binding of XPA-RPA[43,50]. RPA displaces XPC on the undamaged strand and coordinates XPA binding 3' of the lesion, at the ss-dsDNA junction[51]. Together, XPA-RPA form the central scaffold of the pre-incision complex. While RPA shields the undamaged strand from incision, XPA carries out secondary damage verification through direct interaction with the 'damaged' nucleobase prior to recruitment of the structure-specific endonucleases XPF and XPG[52–54]. Gap excision is initiated by XPF-ERCC1, which incises the 3' splayed fork structure on the damaged strand (5' of the lesion)[55]. This initial incision releases a 3' hydroxyl, which is extended across the undamaged strand by polymerase δ, ε or κ (depending on cell cycle status), while bound to PCNA[43]. Strand synthesis across the cleaved bubble generates a favourable substrate for XPG, which cleaves toward the ss-dsDNA junction of the now 5' flap intermediate[55]. Critically, XPF endonuclease processing requires the presence of XPG in the pre-incision complex. A structural dependence between both nucleases serves as a fail-safe, ensuring incision only occurs when complex assembly is complete[54,55]. The dual incisions expel a damage containing ~24-32 oligomer and allow for completion of strand synthesis and nick ligation[44,56].

Interestingly, an alternative pathway competes for the canonical XPG 5' fork intermediate. Disruption of processing downstream of XPF incision can slow the XPG-dependent resolution of the 5' fork substrate. This can result from damage overload depleting necessary repair factors, dNTP dysregulation disrupting synthesis or cleavage-disrupting missense mutations in XPG[44]. Regardless, persistence of the XPF-generated 5' phosphate promotes Exo1-dependent processive 5-3' exonuclease digestion. In addition to resolving the stalled NER intermediate, Exo1 continues past the would-be XPG incision site to degrade kilobases of undamaged DNA[57]. Extended single-strand regions are rapidly recognized and coated by RPA, which is a potent trigger for ATR signalling[57,58]. Damage-induced ATR signalling (through its effector kinase Chk1) promotes cell cycle arrest, inhibition of replication origin firing and/or stabilization of

replication fork intermediates[58]. Collectively, Exo1 outcompeting XPG results in a DNA damage response, indicating that repair is not proceeding as normal.

Proper functioning GG-NER is required for maintaining genomic integrity. Dysfunction results in the severe genetic disorder, Xeroderma pigmentosum (XP). Typically XP results from genetic disruption, through frameshift or missense mutations in critical factors responsible for recognition (XPC), pre-incision complex assembly (XPA, XPD or XPB) or excision (XPF or XPG)[43,59], among others. Clinical presentation includes extreme sensitivity to light resulting in thousand-fold increases in the incidence of skin and eye cancers[44]. The cellular hallmark of XP is a reduction in unscheduled DNA synthesis, resulting from the absence of patch repair[44,60].

1.3.1.2 Transcription-Coupled Nucleotide Excision Repair (TC-NER)

Damage occurring in transcribed genes is repaired faster than elsewhere in the genome. Such differences in repair kinetics require active gene expression and is specific to damage on the transcribed strand[60–62]. Increased repair is primarily observed for lesions able to block RNA polymerase-dependent synthesis. More subtle damage such as oxidation, smaller alkylation or nucleotide mismatches can be directly bypassed and do not trigger repair[43]. Together, this demonstrates a specific mechanism to maintain gene expression by coupling transcriptional stalling with a dedicated repair pathway.

Transcription-coupled nucleotide excision repair (TC-NER) depends on many of the same factors as GG-NER, with some important differences. First, repair is initiated by stalling of an RNA polymerase in a transcription complex, not by sensing factors like XPC (**Figure 1.2B**)[41,60,63]. Though most research has focused on initiation of TC-NER by RNAPII during mRNA synthesis, similar repair-coupling exists with other RNA polymerases[64]. Second, disassembly of the repair complex must be coupled with transcription-restart (**Figure 1.2C**)[65,66].

Transcription is coupled to repair directly by two key factors, TFIIH and CSB. TFIIH functions in both transcription initiation by stimulating promoter melting, and NER by repair bubble remodelling and damage verification[49,50]. CSB functions as both a chromatin remodelling ATPase and a platform for multi-protein complex assembly[67–70]. Throughout transcription, CSB transiently associates with RNAPII. Regardless of damage, this interaction stabilizes into a longer-lived complex following RNAPII stalling[71,72]. CSB ATPase activity differentiates between natural pause sites and bulky damage by inducing forward-translocation of the RNAPII transcription complex. TC-NER is initiated when CSB-mediated translocation fails, indicating that RNAPII is truly blocked by damage[73,74]. Persistent RNAPII stalling triggers a complex series of DNA damage signalling, relying primarily on ubiquitination and SUMOylation dependent conformational change and factor recruitment[73,75]. This process involves CSB, CRL^{CSA}, UVSSA and RNAPII itself[66,72]. Though the exact mechanism is not yet clear, these signals alter the structure of the stalled intermediate, allowing for association of repair factors and access to DNA for endonuclease-mediated cleavage.

Immediately on stalling, the RNAP directly precludes access to the damage site[67]. To promote assembly of repair factors, the RNAP position must be altered, either by forced backtracking along the transcribed strand[65,68] or through inducing an open conformation[67]. Regardless, once the lesion site has been made available, TFIIH stimulates assembly of the canonical NER pre-incision complex. Concurrent with gap-filling and ligation, sequential incisions by XPF and XPG complete repair[60,67,73]. Importantly, XPG serves an additional structural role in TC-NER, forming essential interactions with CSB and RNAPII[67]

Transcription restart is critical for the resolution of TC-NER[65]. RNA elongation requires the 3’ hydroxyl of the transcript to be in the polymerase active site. Any required RNA repositioning occurs via TFIIS-dependent ribonucleolytic transcript cleavage[65]. Transcription complex degradation is a time-dependent process, so quick resolution of the damage is necessary for transcription rescue. Extended stalling of RNA

polymerases results in more robust damage signalling and acts as a potent apoptotic trigger[58,65,75].

Proper functioning of TC-NER is crucial for the maintenance of efficient gene expression. Disruption of this coupled-repair process manifests in the severe genetic disorder Cockayne Syndrome (CS)[59]. CS is characteristically associated with genetic disruption of factors specifically involved in the initial activation of TC-NER, including CSA, CSB, XPB, XPD and specific XPG nonsense mutations[59,60]. Patients with CS present with an increased incidence of skin cancers, and a less severe UV-sensitive phenotype than those with XP. Uniquely, these patients demonstrate progeria (rapid pre-mature aging) and significant physical and mental developmental deficiencies[44,59,69]. The cellular hallmark of CS is accumulation of stalled RNA polymerase complexes, indicating a failure of transcription restart[60]. Where XP primarily features a dramatic cancer risk through failure of fidelitous lesion repair, CS features progeria resulting from a generalized elevation of apoptosis.

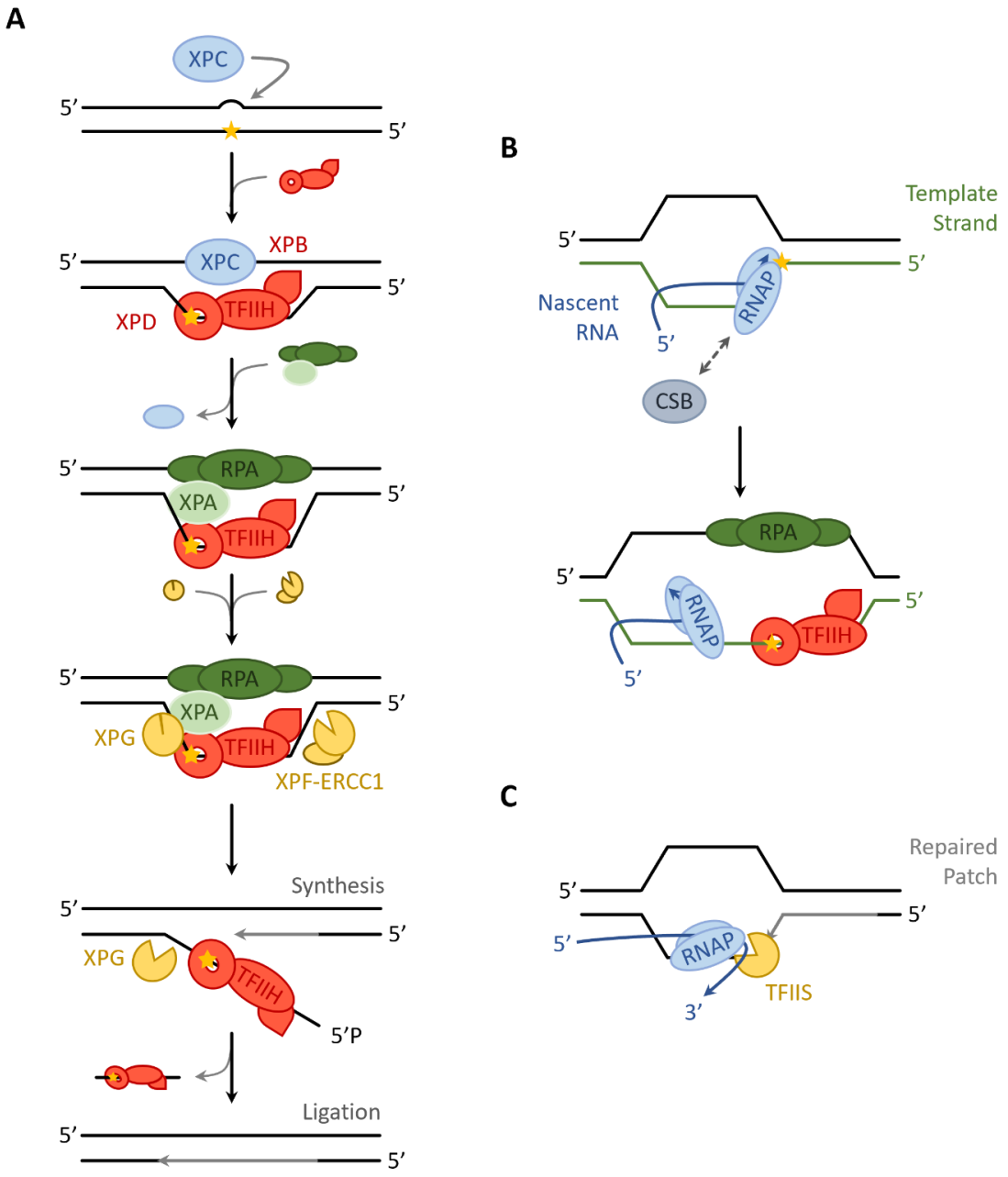


Figure 1.2. Schematic of Nucleotide Excision Repair.

[A] General schematic of GG-NER. XPC senses local instability in DNA, binds opposite the damage and initiates assembly of the pre-incision complex. TFIIH subunits XPB and XPD verify damage and recruit XPA and RPA. XPF initiates excision 5′ of damage, licencing XPG incision concurrently with strand synthesis. Gap filling and ligation completes repair, regenerating an undamaged helix. **[B]** Schematic of TC-NER initiation. RNA polymerase stalling promotes CSB recruitment. Failure of CSB-dependent forward translocation initiates assembly of incision complex. **[C]** Schematic of transcription restart following TC-NER. TFIIS degrades the nascent RNA such that the 3′ hydroxyl is in position for elongation by RNAP.

1.3.2 Replication-Independent ICL Repair

1.3.2.1 Global-Genome Interstrand Crosslink (GG-ICL) Repair

Repair of interstrand crosslinks has been observed in transcription- and replication-independent systems[33,76]. In mammalian cells, these pathways likely dominate in quiescent or differentiated, non-replicative cells[41]. The accumulation of unrepaired ICLs promotes genetic instability beyond what is introduced during polymerase-dependent processes, through disruption of chromosome topology and supercoiling[41]. This necessitates mechanisms of recognition, independent from a stalled replisome or RNA polymerase complex. Studies specifically tailored to monitor this form of ICL repair have identified multiple competing pathways.

ICL repair in bacterial and lower eukaryote systems has been more comprehensively characterized. In *E. coli*, ICL damage is removed in a three-step process depending on the coordination of NER and HR[1]. Repair is initiated by asymmetric incisions flanking the crosslink, made by the NER UvrABC endonuclease complex[77,78]. Dual incisions unhook the crosslinked strands and produce a substrate for the 3' to 5' exonuclease activity of PolI, generating a small gapped-DNA intermediate 3' of the ICL. RecA-mediated strand invasion of the gapped repair intermediate displaces the remaining oligonucleotide, still crosslinked to the unbroken strand[1,78]. UvrABC then recognizes the unique, short triplex DNA structure and excises the second strand. Concurrent strand synthesis (using the homologous template) regenerates the undamaged duplex while maintaining fidelity, without generation of a DSB intermediate[78]. Unfortunately, bacterial models have limited explanatory power in mammalian systems where ICL repair can persist in the absence of homologous recombination[41].

Though the NER-HR mechanism is generally conserved between bacteria and yeast, homology-dependent repair (coupled with DNA replication) proceeds through a DSB intermediate in *S. cerevisiae*[79]. Though the primary ICL repair mechanism in yeast depends on DNA replication, repair persists in stationary haploid cells. In contrast with

homology-dependent repair, homology-independent repair does not generate DSB intermediates[79]. This pathway is instead facilitated by NER dual incision flanking the ICL, followed by attempted gap-filling via PCNA-dependent Pol32 (Polδ). Polymerase stalling at the crosslink site triggers ubiquitination of PCNA, promoting polymerase-switching and mutagenic Rev3 (Polζ) translesion synthesis[79,80]. Concomitant with strand synthesis, Pso2(SNM1) exonuclease processing of the oligonucleotide adduct is thought to resolve the short, triplex DNA structure[80,81]. This pathway established the importance of replication-/ homology-independent ICL repair mechanisms in eukaryotes. As multicellular organisms are composed of many non-replicative differentiated cells, it is likely that higher-order eukaryotes have a more complex network of replication-independent recognition and repair processes.

Initial studies of mammalian ICL repair used cell extracts and short (~150 nucleotide) duplex DNA substrates containing an internal site-specific ICL lesion[82,83]. As these DNA substrates lacked replication origins and promoters, the observed repair must have been facilitated exclusively by GG-ICL repair mechanisms. These researchers identified NER-dependent dual incisions, *5' of the lesion*, excising an undamaged ~24 oligomer[82]. The resulting gapped DNA intermediate served as a substrate for Polη strand synthesis, followed by ligation[83]. Detection of this futile repair cycle confirmed that direct recognition of a psoralen ICL results in NER-dependent endonuclease incisions, though how it promoted repair was unclear[82,83]. These observations highlight a potential limitation of NER processing of ICLs. If NER depends on the formation of a bubble surrounding the lesion, would an ICL be suitable for direct NER processing?

Using similar mammalian cell extract experiments, work by Miller and colleagues identified incisions *flanking* the ICL in a duplex DNA substrate[41]. These incisions occurred within a few nucleotides 5 and 3' of the ICL and were independent of the NER endonucleases XPF and XPG. The frequency of NER-independent incisions was proportional to the crosslink-induced helical distortion[3], establishing a plausible mechanism of direct recognition of an ICL lesion. Regeneration of non-damaged linear DNA demonstrated that complete repair was supported by the cell extracts, though it

was relatively inefficient[3]. Stepwise characterization of this process determined the resulting unhooked adduct contained of a single residual nucleotide, crosslinked to the non-targeted strand. Generation of this intermediate would require processing of the unhooked oligonucleotide (originally ~10 nucleotides) by an unknown factor, likely an intrinsic polymerase-dependent exonuclease or translesion exonuclease (such as SNM1A)[84]. Gap-filling efficiency of Polη-dependent synthesis depended on the chemical nature of the crosslink. Where crosslinks disrupt the hydrogen-bonding face of the nucleobase (such as at N^2-dG), TLS was required to complete repair. As translesion synthesis was inactive in the cell extracts, non-disrupting crosslinks (such as at N^7-dG) were bypassed efficiently by Polη, maintaining sequence fidelity[84]. Sensing factors responsible for initiating this process were not directly identified[41].

Counter to the NER-independent process identified by Miller and colleagues, other groups have demonstrated an alternative surveillance mechanism, dependent on NER[77,85]. The primary lesions generated by ICL-inducing compounds *in vivo* are monoadducts. These monoadducts are, generally, ideal substrates for NER. Because NER is required for repairing the non-ICL damage induced by standard ICL treatments, deconvoluting the role of NER in GG-ICL repair has proved difficult. Specific recognition of ICLs by key NER complexes provide strong evidence that absent transcription and replication, NER does play a role in ICL repair. Specifically, XPC-RAD23B[33,86], and XPA-RPA[87] in combination with HMGB1[88] have both been shown to directly recognize ICL containing DNA, *in vitro* and *in vivo*.

To further characterize the role of NER, repair in response to laser-localized psoralen ICLs was contrasted with an obligate monoadducting analogue, angelicin, in G_1 stalled mammalian cells. Temporal treatment with an RNA synthesis inhibitor, actinomycin D, allowed differentiation of transcription-dependent effects. Together, this study monitored transcription- and replication-independent repair[33]. Here, repair of psoralen ICLs required all core NER factors. Using immunofluorescence, the kinetics of recruitment and repair were monitored. As with GG-NER, XPC-RAD23B is recruited to the damage site immediately, followed by formation of the pre-incision complex.

Importantly, ICL unhooking (a key metric for repair progression) was absent in XPC-deficient cells. Resolution of psoralen repair foci was significantly slower compared to angelicin, demonstrating that ICL repair is substaintially slower than NER[33].

Intuitively, an ICL would be a poor substrate for NER because the covalent crosslink should prevent formation of the required repair bubble. The practical importance of this limitation to NER efficiency is unclear, as UvrABC can generate flanking incisions surrounding an ICL in *E. coli*[78]. Further translating the canonical NER process into a GG-ICL repair context, XPC recognition of an ICL must result in binding adjacent to the damage site, instead of directly opposite. Because both strands are damaged by an ICL, XPC could form four degenerate complexes with the DNA (top or bottom strand, 5' or 3' of the lesion). However, only two of these complexes should have productive repair potential owing to the polarity of TFIIH-XPD (3 to 5' helicase activity) complex remodelling. XPD interaction with the damage is required for assembly of the complete NER complex[50]. This is consistent with the previously discussed observations of NER-dependent incisions on linear DNA, where futile repair was not strand specific, however was only observed 5' of the lesion[82,83]. Subsequent assembly of the pre-incision complex should position XPG to cut at/ near the lesion. As NER has been demonstrated to be important in GG-ICL repair[33], either (i) XPG makes an incision in the duplex, past (3') the ICL or (ii) another factor is required to complete unhooking (for example, a translesional exonuclease such as SNM1A). If the results originally described by Bessho *et al.* are broadly generalizable (XPG-dependent incision occurs ~6 nucleotides 5' of the ICL[82]), the latter seems more likely.

In addition to NER participation in GG-ICL repair, multiple groups have demonstrated a role for mismatch repair factors (in combination or in isolation with NER factors). Using recombinantly purified MutSβ and HeLa cell extracts, MutSβ (MSH2-MSH3) was shown to directly recognize psoralen ICLs[30,89]. Further, MutSβ interacted cooperatively with NER factors XPA and XPC in lesion recognition[33,90]. Unexpectedly, the mismatch endonuclease subunit of MutLα was shown to function not in incision, but in initiating ICL-induced checkpoint activation[91]. The direct relevance of these studies is

unclear, as they were done using homology/ replication-proficient systems, obscuring their direct application in GG-ICL repair.

Using transcription- and replication-incompetent *Xenopus* extracts, MutSα (MSH2-MSH6) was required for the repair of a site-specific SJG-136 crosslink[31]. Here, sensing factor MutSα interacted directly with the ICL lesion and recruited the endonuclease MutLα (MLH1-PMS2), exonuclease Exo1, PCNA and RPA. Repair of the ICL depended specifically on the nuclease activity of PMS2 and Exo1[31]. This repair process required two distinct nuclease events, stimulated by a 5' nick close to the ICL (reminiscent of the product from an un-ligated 'futile NER repair cycle', described above[82,83])[31]. Though repair of ICLs required key MMR factors, the complete process was not characterized. In particular, as Exo1 processing is unable to bypass ICL lesions *in vitro*, it was unclear how it contributed to intermediate resolution[31,92]. Again, ICL repair seems to require participation of additional nucleases.

Interestingly, the GG-ICL repair process triggers large unscheduled DNA synthesis (UDS)[93]. Using plasmids with a single site-specific ICL, repair in replication-incompetent *Xenopus* extracts elicited synthesis upwards of 300 nucleotides[93]. While this exceeds expected UDS associated with canonical NER by nearly 10-fold[84], it would be in-line with long-patch MMR[94] or other mechanisms dependent on processive exonuclease digestion. Extended excision of nucleotides surrounding an ICL may function to activate a DNA damage response, like that observed when Exo1 degrades stalled NER-intermediates[58,93].

Although the findings of many studies seem contradictory, both NER- dependent and independent repair likely function as separate pathways, competing for overlapping substrates. The exact dynamic has not been clearly established, as research is limited by the lower efficiency of GG-ICL repair. Comparable estimates of replication-dependent and -independent repair demonstrate that GG-ICL repair is a much slower process. In replication competent *Xenopus* extracts, site-specific ICLs were almost completely repaired after 2 hours[39]. In contrast, using replication-incompetent *Xenopus* extracts,

less than 15% of ICLs were removed over the same time[93]. The differential kinetics are likely the result of recognition; while replisomes rapidly initiate repair[95], GG-ICL repair depends on the efficiency of direct sensors. MutSα[31], XPC-RAD23B and XPA-RPA-HMGB1[33,86–88] recognize ICLs with affinity proportional to the ICL-induced helical distortion. This is consistent with the findings that more-distorting ICLs are repaired faster[3,4,31]. Interestingly, UHRF1, a newly described ICL sensor, more efficiently recognized less distorting ICL damage[96]. Thus, a web of GG-ICL sensors may distinguish repair depending on cell type, chromatin state, genomic location in addition to helical distortion.

1.3.2.2 Transcription-Coupled Interstrand Crosslink (TC-ICL) Repair

In the absence of replication, ICLs formed within actively transcribed regions are recognized and repaired faster than those in non-transcribed regions[3,4]. Transcription promotes DNA repair through two distinct mechanisms. Indirectly, transcription activation decreases chromatin condensation, thereby increasing potential for the direct sensing of lesions. Directly, stalling or blocking of the transcription complex facilitates damage recognition[41].

Transcription-coupled ICL repair occurs through a distinct pathway than that associated with global-genome repair, sharing many characteristics of TC-NER. To isolate TC-ICL repair, studies used site-specific ICLs incorporated within a transcribed region under the control of a strong promoter[4,32,36,97,98]. Removal of the ICL was inferred by reactivation of the lesion-disrupted reporter gene. Using this experimental design, contributions from replication-dependent or global-genome repair were minimized or excluded altogether. Contributing repair factors were determined using patient-derived cell lines containing specific gene disruptions (and their complements) for many known DNA repair proteins[4,32,36,97,98]. Collectively, a wide range of ICLs were used to monitor

TC-ICL repair including site-specific incorporation of: psoralen[32], MMC[97] analogues[36], nitrogen mustard[4] or cisplatin[98].

As anticipated, disruption or depletion of factors specifically required for homologous recombination (FANCU/XRCC2, XRCC3 and Rad52)[4,32,36,97] and the Fanconi Anemia pathway (FANCA, FANCL, FANCD2 and Polζ)[36,93] were not required for repair. Instead, repair relied on TC-specific CSB[4,97,98], CSA[97], as well as the NER core pre-incision complex (XPA, XPB, XPD and RPA)[32,97,98] and endonucleases (XPF-ERCC1 and XPG)[4,32,97,98]. Involvement of XPC-RAD23B was inconsistent; not required in some systems[4,98] and associated with a mild phenotype in others[32,97]. Interestingly, XPC was not required for initial ICL unhooking. It instead may be important for downstream recognition of the monoadduct, generated by transcription-coupled repair[4]. Multiple Y-family translesion polymerases (Rev1 and pol η or κ) have been implicated in synthesis following unhooking[36,97]. Which polymerase is recruited is likely a reflection of remaining adduct structure.

Unlike other forms of ICL repair (GG or RDR) or GG-NER, there is a distinct polarity to TC repair processes (**Figure 1.3**)[97]. Polymerases extend the 3' end of DNA, thus transcription moves 5 to 3', with RNAPII using the complement (3 to 5') DNA strand as a template. An ICL will act as a complete block of the transcribing RNA polymerase, resulting in transcription stalling up to the -1 nucleotide from the lesion[67]. The stalled transcription bubble forms a pseudo-Y, splayed arm structure, with the ICL lesion just upstream of the ss-dsDNA junction, within the duplex. Incorporating the elongated RNA molecule, the intermediate formed resembles a 5' flap, where the 3' arm is an RNA:DNA hybrid helix. In replication-dependent or global-genome repair, the damage is symmetric and repair can be initiated on either strand[3]. Plasmid reactivation assays, monitoring site-specific repair of ICLs, identified asymmetry in transcription-initiated repair through a strong strand preference for incision[4,97]. Analysis of mutational spectra following repair identified that the major mutagenic event occurred directly opposite to the damaged base of the non-transcribed strand[97]. Assuming this reflects translesion

synthesis across the damaged base, the primary incision events must occur on the transcribed strand.

Similar to NER-dependent GG-ICL repair, it is not clear whether the NER incision complex is able to directly unhook an ICL in the absence of additional factors. It is conceivable that initial processing produces an intermediate for ICL-specific factors to digest. Although TC-ICL repair is responsible for removing only a fraction of ICLs in most cells, its function is essential for maintenance of transcription[39,73]. When disrupted, significant RNA polymerase stalling promotes dysfunctional gene expression and apoptosis. As such, clear elucidation of the differences in transcription-coupled processing of an ICL, relative to other bulky adducts is necessary.

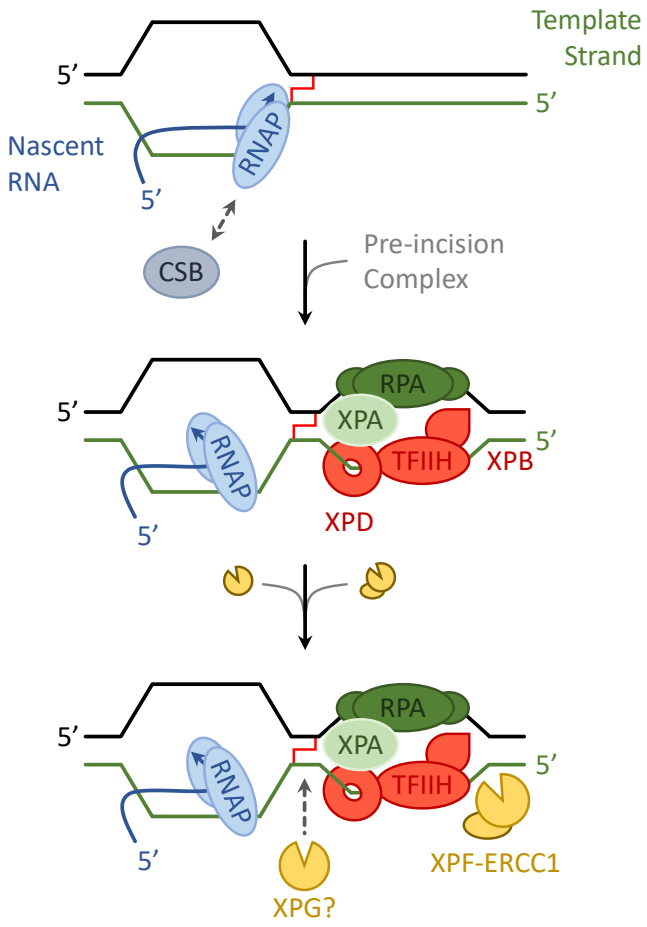


Figure 1.3. Schematic of transcription dependent ICL repair.

TC-ICL repair is initiated by stalling of an RNA polymerase at an ICL. Failed forward translocation of the RNA polymerase by CSB promotes assembly of the NER pre-incision complex. Incisions occur on the transcribed strand, requiring TFIIH to load downstream of the lesion. As in TC-NER, excision is initiated by XPF incision of the 3’ splayed substrate. While XPG is required for TC-ICL repair, how the incision promotes unhooking may depend on the identity of the crosslinker.

1.3.3 Replication-Dependent ICL Repair

The presence of a homologous template dramatically stimulates repair of an ICL[38]. *Ex vivo* repair of plasmids containing a site-specific psoralen ICL is stimulated approximately 20-fold in HeLa whole-cell extracts when incubated with an undamaged homologous plasmid. This process is distinct from the GG- or TC-ICL repair discussed above, occurring independent of most NER factors (XPA, XPC and XPG)[30,38] or the transcription-coupling factor CSB[99]. Critically, XPF-ERCC1 endonuclease plays a central role, independent from NER, in a recombination-dependent repair pathway[30,54,99]. Repair requires significant DNA synthesis on both the damaged and undamaged templates, dependent on recombinatorial proteins Rad51-like XRCC2 (FANCU) and XRCC3[38,100]. Unlike RIR, recombination-dependent ICL repair generates a double-strand break intermediate[39,76,99]. The repair of the DSBs occur independent of non-homologous end joining (NHEJ), relying on homologous recombination or break-induced repair pathways[38,99,101].

The majority of ICL lesions in dividing cells are recognized and repaired during replication[1,102]. Specifically, mammalian cells will stall in late S-phase, independent of when in the cell cycle the ICL damage is induced[102]. Further, ICL repair-induced DSB intermediates are only observed in S-phase cells[76,99]. Because ICLs act as a significant block to replication, cell division requires resolution of these lesions. Fanconi Anemia (FA)-associated genes uniquely facilitate replication-dependent ICL repair (RDR), functioning throughout initiation, signalling, ICL processing, recombination and substrate resolution[2,6]. Many of these genes are specific to the repair of ICLs during replication[76,102] and facilitate the stabilization of stalled replication forks[10,103].

Absence of an intact FA-pathway has severe cellular and phenotypic consequences[10]. Fanconi Anemia is a rare, recessive and heterogenous genetic disorder caused primarily by the disruption of RDR. Cells lacking FANC-designated genes share common hallmarks, though each deficiency can manifest in unique disease conditions[10,104]. Primary cells derived from patients with dysfunctional RDR share a

hypersensitivity to crosslinking agents and severe chromosomal instability marked by the accumulation of radial chromosomes[10]. Patients are typically diagnosed either (i) at birth, due to congenital abnormalities resulting from chromosomal aberration in progenitor cells during development, or (ii) in childhood, resulting from severe pancytopenia, bone marrow failure and/or cancer diagnoses[2,10,104].

The FA-pathway is conserved among higher eukaryotes, while a prototypic version has also been described in yeast[105]. FA-associated genes fall into three broad categories: (i) an E3-ubiquitin ligase complex, (ii) incision scaffold and nucleases and (iii) downstream repair and resolution factors[2,6,10,106]. Many of the factors in groups II and III have overlapping roles with other DNA repair pathways, primarily NER, HR and TLS. Group I factors play a less defined, but specific role within ICL repair.

Lesion sensing is critical for repair initiation. FANCM is a multi-subunit protein, which interacts specifically with branched oligonucleotides, favouring junction-specific interactions[106]. Sandwiched between two DNA binding domains, FANCM contains FANCF-interacting and MHF-interacting regions[106,107]. While MHF1 and MHF2 promote chromatin association of the entire complex, FANCF serves as a scaffold for assembly of the remaining FA core complex. Though disruption of the FANCM complex limits FA-associated repair, it does not abolish it[106]. This suggests that FANCM serves as a primary, though not the sole, molecular sensor in FA.

The FA core complex is a higher order assembly of multiple ternary complexes localized to the site of damage by FANCM[106,108]. Collectively, FANCA-C and FANCE-G along with associated factors FAAP20 and FAAP100 form the substrate selectivity module of the E3-ubiquitin conjugating ligase FANCL[2,108]. FANCL, charged by the E2 enzyme UBE2T (FANCT), monoubiquitinates the major signalling effector of the FA-pathway, FANCD2[106,109]. While in a heterodimer with FANCI, ubiquitination of FANCD2 fixes the scaffold on DNA proximal to the ICL lesion[110,111]. It is not yet clear why such extreme complexity would be required for the regulation of FANCD2 ubiquitination. Intuitively, replication-initiated repair should provide the cell with a simple recognition

mechanism. This suggests that there are alternative/ additional functions facilitated by the core complex which are not fully understood. Research aimed at uncovering those diverse functions is currently a major focus in the field[35,103,112,113].

Outside the activation step of FA by ubiquitination of FANCD2, the clearest models of RDR come from a series of studies from Walter and colleagues, using *Xenopus* egg extracts. *Xenopus* extracts provide a replication competent system to evaluate ICL repair initiated by replication[114]. Preparation of the extracts provide temporal control of the licencing and initiation of a single round of replication. Further, this system enables efficient depletion and/or complementation of endogenous proteins, and introduction or recovery of plasmid DNA[39].

Replication initiation generates symmetric forks which diverge along the plasmid DNA. Incoming DNA is driven apart at the leading edge of the replisome by the CMG (Cdc45, MCM2-7, GINS) helicase, which surrounds the leading strand template[9]. Polymerases ε and δ are closely coupled behind CMG, extending the 3' hydroxyl of the nascent leading and lagging strands, respectively[9,115]. Generally, replication is stalled by the presence of an ICL. The collision of each replication fork can be directly observed in the extracts (peaking ~10-20 minutes following replication initiation). Replication using α^{32}P dNTPs show that leading strand synthesis stalled ~20 nucleotides from crosslink[39], sterically blocked by the CMG helicases abutting the ICL[9]. Reflecting the diverse primase activity between plasmids, lagging synthesis was more heterogenous, leaving ~70-290 nucleotides as ssDNA[39]. Productive repair of the ICL depends on replication fork convergence, resulting in an X-structure surrounding the lesion (**Figure 1.4A**)[95]. Because inter-origin distance in mammals is much larger than the plasmid constructs in these assays, convergence could take significantly longer. Critically, single stalled replication forks were stable for more than 3 hours, remaining competent for repair following arrival of the second fork[95].

Repair is initiated by post-translational modification of CMG by TRAIP-dependent ubiquitination[116]. TRAIP is an E3-ubiquitin ligase constitutively associated with the

replisome[117,118]. At a converged fork, TRAIP catalyzes the poly-ubiquitylation of the opposing CMG molecule. Significant extension of these ubiquitin chains promotes time-delayed recruitment of the p97/Cdc48 segregase, which in turn ejects the helicase[116,119]. Without the steric block, Polε extends the nascent leading strand up to the crosslinked nucleotide (**Figure 1.4B**)[9,39,116]. Approach of the leading strand occurs on a single fork, initiating asymmetric fork reversal[39,120]. The specific DNA remodeling proteins which drive fork reversal are unclear, however endonuclease incisions required for lesion unhooking depend on the reversal of the X-structure[120,121]. Reversal produces a 3' tail comprised of the nascent strands and regenerates an apparent single stalled fork structure adjacent to the ICL. These structures are separated by reannealed DNA and a Holliday junction[120]. Without branch migration, the intervening duplex DNA should reflect the ssDNA gap on the lagging strand of the reversed fork (**Figure 1.4C**).

Incisions flanking an ICL lesion on a single stand unhook the crosslinked DNA. Coupled with replication, unhooking generates a two-ended double strand break[39,99,122]. Using *Xenopus* extracts, incision-dependent DSBs can be directly monitored[39,95,123]. The mechanism of ICL unhooking remains poorly characterized, however XPF-ERCC1 is unambiguously required. XPF(FANCQ)-ERCC1, functioning independent of NER[54], is recruited in complex with the scaffold protein SLX4 (FANCP) by the ubiquitinated FANCD2-FANCI heterodimer[122–125]. It is unclear whether XPF makes its incision 5 or 3' of the ICL and whether additional nucleases are required to complete unhooking. Depletion of XPF in the extracts completely abolished unhooking incisions[122], suggesting either (i) XPF completes unhooking alone or (ii) actions between nucleases are somehow coordinated. Coordination would require either an additional endonuclease (similar to NER) or that XPF generates the substrate for a translesional exonuclease (similar to MMR).

Following unhooking, a residual adduct would remain crosslinked on the unbroken strand. If dual endonuclease incisions result in unhooking, the adduct would consist of the intervening DNA between incisions. Depending on where these incisions form, the expected adduct would contain ~4-10 nucleotides[84,122,126–128]. Alternatively, if

unhooking is accomplished through coupling of endonuclease initiation with translesional exonuclease digestion, the adduct should consist of 1-3 nucleotides[39,84,92,121,129,130]. Direct observations in distinct systems have exclusively identified a single nucleotide in the unhooked adduct[39,84]. It is unclear if this results directly from unhooking or reflects secondary processing by an exonuclease. Regardless, exonuclease trimming of the oligonucleotide adduct is essential for efficient insertion and translesion bypass[128,131].

When incisions are generated in the presence of a reversed fork, the 5’ incision does not immediately result in a double strand break. Secondary processing, which unhooks the crosslink, generates a first-end DSB and residual Y-structure that is held together by the Holliday junction. The second-end DSB is released on resolution of the junction by branch migration toward the ICL or some other mechanism (**Figure 1.4D**)[120].

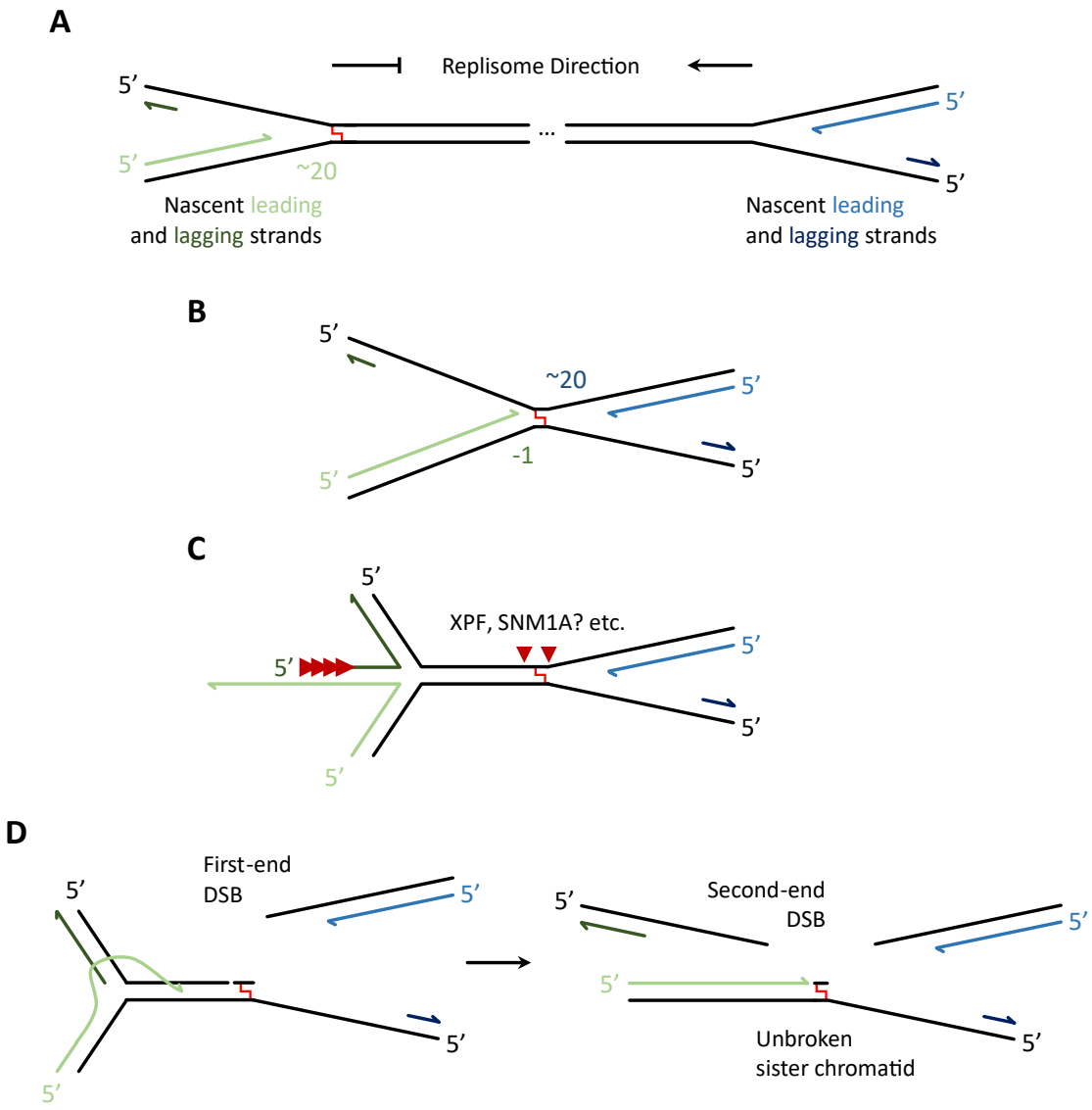


Figure 1.4. Schematic of replication dependent ICL recognition and unhooking.

[A] ICL repair is initiated by convergence of two replisomes at an ICL. Nascent leading strand synthesis temporarily stalls ~20 nucleotides from the lesion. **[B]** Approach. Following dual replisome convergence, one fork is extended up to the ICL by the replicative DNA polymerase. **[C]** Reversal. The approached fork is remodelled, generating a reversed tail with nascent strands. Exonuclease digestion of the reversed lagging strand produces a 3’ tail. Unhooking of the ICL generates the first-end DSB. **[D]** Invasion. Strand invasion by the 3’ tail collapses unhooking intermediate, releasing the second-end DSB and an unbroken sister chromatid.

Only nascent leading strand approach on the reversed fork would lead to productive TLS across the damaged nucleotide. Resolution of the reversed fork places the 3' hydroxyl of the nascent leading strand (from the 3' tail of reversed fork) directly adjacent to the damaged nucleotide (**Figure 1.5A**)[120]. Recruitment of PCNA and a Y-family translesion polymerase facilitates insertion across the damaged base, followed by extension through the opposite lagging strand gap. REV1 with Polζ were required for nucleotide insertion and extension, respectively, across cisplatin and nitrogen mustard ICLs[37,39]. Once bypassed, the remaining unhooked adduct becomes a substrate for GG-NER. Ligation of the nick on the unbroken sister chromatid regenerates a substrate sufficient to facilitate HR-mediated repair of the two-ended double strand break (**Figure 1.5B**).

Canonically, DSBs destined for HR are recognized and resected by the MRN (Mre11, Rad50, Nbs1) complex, CtIP and Exo1. BRCA1 (FANCS) promotes end resection, indirectly generating extensive 3' tails which are rapidly bound by RPA[132,133]. BRCA2 (FANCD1) then stimulates RPA displacement through RAD51 nucleofilament polymerization. RAD51 is recombinogenic, initiating homology search and strand invasion of the unbroken sister chromatid[40,132].

Uniquely in ICL repair, BRCA1 (FANCS) function precedes double strand break generation, such that significant RAD51 nucleofilament formation occurs prior to endonucleolytic incisions[40,134]. Pre-emptive RAD51 loading would facilitate near concomitant HR following restoration of the reversed fork. It is possible that RAD51 filament formation occurs on the RPA-coated lagging strand of the stalled fork, however the lagging strand which generates the second-end DSB should be annealed within the reversed fork[120]. Alternatively, RAD51 may be necessary for fork reversal or resolution[40,135,136]. Fork reversal, in combination with BRCA1, has been shown to stimulate nascent lagging strand resection (~1-2 kilobases) to produce a long single-strand 3' tail, ideal for RAD51 nucleofilament polymerization[120]. Genetic studies in human cells suggest this process may be facilitated by ICL-specific early functioning of MRN-CtIP resection, prior to lesion unhooking[137].

RDR is completed by HR-mediated DSB repair. Regardless of the exact mechanism, processing of both break ends to generate long 3' tails facilitates RAD51 polymerization, promoting recombinogenic strand invasion of the unbroken sister chromatid[40]. The non-invading tail can anneal to the extended D-loop, likewise promoting DNA synthesis. Dual strand synthesis, extending both DNA breaks, produces a Holliday junction between the chromatids. Resolution of the HJ completes the DSB repair and resolves the remaining ICL-induced intermediates (**Figure 1.5C**)[40,135].

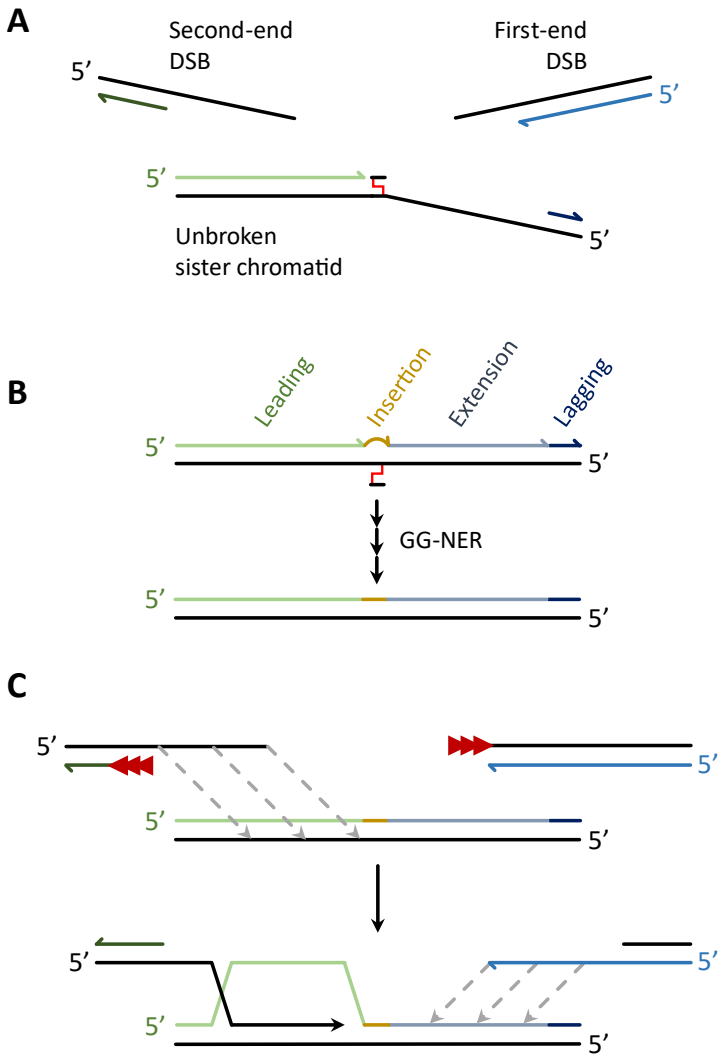


Figure 1.5. Schematic of replication dependent ICL resolution.

[A] Collapse of the unhooking intermediate produces a two-ended DSB and a gapped sister chromatid, not suitable for HR. **[B]** TLS inserts a nucleotide across from the damaged base (insertion) and extends the leading strand to fill the gap left by the other stalled replisome (extension). Residual damage is removed by GG-NER, regenerating a template for HR. **[C]** End resection of each break promotes HR-mediated DSB repair.

1.3.3.1 Traverse Model of Repair

Though experiments using *Xenopus* extracts have significantly increased our understanding of ICL repair mechanisms, it is not clear how comprehensively the system mirrors human repair. Recent work in human cells observed global replication fork slowing/ stalling in response to ICL damage, dependent on ATR signalling[136]. As such, initial fork convergence at one lesion lowers the likelihood of convergence at others.

To directly test the relevance of dual-fork convergence, multi-layered nascent strand labelling with nucleotide analogues was used in conjunction with antigen-conjugated psoralen. Together, these techniques permitted direct observation of replication adjacent to ICL lesions *in vivo*[112,136,138]. Following damage, only 15-20% of ICLs were bracketed by converged replication forks. Instead, the majority were directly bypassed by a single fork, without unhooking. While the rate of bypass was cell type dependent, most ICLs were traversed too rapidly to require dual-fork convergence (<15 minutes)[112,136]. Emergence of ICL lesion bypass was in direct conflict with the prevailing notion that ICLs act as an absolute block to replication.

FANCM plays a primary instigating role in activation of the FA-pathway[2,106]. The ATPase-dependent translocase activity of FANCM, while dispensable for FA[139], was required for bypass of an ICL[112,136,138,140]. The remaining FA core-complex was not required for lesion bypass, though it likely functions later in lesion repair[136]. Activation of ATR damage signalling following fork stalling triggers, among many other events, the phosphorylation of FANCM. Phosphorylated FANCM facilitates CMG remodelling, culminating in the ejection of the GINS subunit[138]. The exact mechanistic consequences of FANCM-facilitated remodelling is not fully understood, though the 'unlocking' of the replisome helicase could allow accommodation of the ICL structure[112,138]. Together, with BLM helicase action past the ICL and PrimPol-dependent re-priming of leading and lagging strands, DNA replication can restart downstream of the ICL lesion without firing of a dormant origin[140,141].

Though lacking CMG-dependent recruitment of any processing factors, the residual DNA structure should closely mimic converged forks, where dual fork convergence generates a (nearly) symmetric X-structure[39]. However, replication traversing the ICL would maintain nascent leading synthesis on the same parent strand[112]. Whether this manifests in differences in unhooking and downstream processing remains unclear.

1.3.3.2 Alternative RDR Pathways

1.3.3.2.1 Glycosylase-Dependent Processing

Dual replication fork convergence at an ICL initiates pathway choice within eukaryotes. Stalled CMG helicases at an ICL are progressively ubiquitylated by TRAIP[116]. Extensive polyubiquitination triggers p97-induced CMG unloading[119], however short chain ubiquitination serves as a scaffold for the recruitment of the NEIL3 glycosylase[116]. NEIL3 canonically functions in BER, removing oxidized nucleobases via digestion of the N-glycosyl bond of the damaged nucleotide[142]. When recruited to an ICL, NEIL3 directly catalyzes the unhooking of a crosslink, producing an abasic site and a nucleobase-/ crosslinker-containing monoadduct[29]. The damage specificity of this repair branch is unclear, though it is likely determined by substrate selectivity of NEIL3 itself. In experiments using *Xenopus* extracts, glycosylase-dependent unhooking was the primary repair pathway for abasic and psoralen ICLs (but not nitrogen mustard and cisplatin)[29,39]. Using nucleotide analogues containing non-hydrolyzable N-glycosyl bonds, repair of psoralen and abasic ICLs were instead funneled into the FA pathway[29]. As such, glycosylase-coupled ICL repair likely represents a primary RDR mechanism, where its failure culminates in initiation of the more versatile FA pathway. Importantly, NEIL3 unhooking negates DSB formation, minimizing the potential for genomic instability[6,29,143].

1.3.3.2.2 Direct Lesion Processing

Mechanistic analysis of an acetaldehyde derived ICL recently identified another novel pathway for repair in *Xenopus* extracts[144]. Convergence of two replication forks at the ICL immediately initiated unhooking that was independent of CMG unloading. Unlike psoralen or abasic ICLs however[29], unhooking did not depend on NEIL3 or any glycosidic processing[144]. Instead, they observed direct enzymatic digestion of the acetaldehyde crosslinker. Though the enzyme has yet to be identified, unhooking generated an undamaged guanine and a N^2-propanoguanine monoadduct. Hydrolysis at the aldimine would reverse crosslink formation, generating the observed products. In fact, chemical reduction of the aldimine prior to repair abolished the direct processing, triggering the slower FA-dependent unhooking[144]. The specific moiety targeted by the unknown enzyme should be conserved in crosslinks from many bifunctional aldehydes[11,21].

Abasic sites and metabolic aldehydes are likely to be major endogenous sources of ICLs, given their abundance in cells[17,19,20,23]. It is especially interesting that distinct, non-overlapping, pathways exist to rapidly unhook each in a DSB-independent manner[29,144]. The efficiency of each pathway further bolsters the physiological relevance of both crosslink sources.

1.3.3.3 Unhooking Associated Nucleases

Regardless of how ICL repair is initiated, nucleases play critical roles in the unhooking of crosslinked strands. With the exception of glycosylase and direct-processing repair[29,144], all ICLs require nucleolytic unhooking reactions. Determining which nucleases are responsible for this processing has proven extremely difficult, reflecting heterogeneity in repair[42], redundancy between nucleases[124,145] and situational plasticity of individual nuclease preferences[121,127,146]. As discussed above, the cell cycle status and genomic location result in substantially different DNA intermediates requiring nuclease digestion[39,41,42]. Biochemical characterization of individual nucleases

has highlighted comprehensive functional overlap *in vitro* and *in vivo*. Adding to the complexity, nuclease activity and substrate preferences are often modified by post-translational modification and interacting partners.

Accumulation of prominent nucleases required for ICL repair was directly observed within *Xenopus* extracts using time-resolved mass spectroscopy analysis of psoralen-crosslinked chromatin. As expected, XPF-ERCC1 and the associated scaffold protein SLX4 accumulate at psoralen lesions concomitant with incision. Although many other nucleases have been separately implicated, FAN1 and SNM1A were similarly recruited to the psoralen damage[147].

As discussed previously, XPF-ERCC1 is indispensable for FA-dependent incisions in *Xenopus* extracts[122,125]. Similarly, XPF is required for unhooking and resolution of ICLs in mammalian cells[76,99,127,148,149]. Localization of XPF to ICL lesions depends on its association with SLX4, which in turn requires FANCD2 monoubiquitination[123]. SLX4 is a large scaffolding protein that coordinates recruitment of multiple nucleases to DNA[150,151]. In addition to XPF-ERCC1, SLX4 directly interacts with and coordinates an atypical HJ resolvase, composed of SLX1 and MUS81-EME1[146,152,153]. Though SLX4 deficiency results in severe ICL hypersensitivity[127,154], truncations lacking MUS81 and SLX1 interaction domains are able to rescue the repair deficiency[127]. Instead, MUS81/SLX1 appear to primarily function in late G2, resolving residual junctions or replication forks in order to ensure sister chromatid segregation during mitosis[146,155].

XPF-ERCC1 endonuclease activity has been extensively characterized *in vitro*. XPF primarily functions as a 3' flap endonuclease in NER and SSA. XPF-ERCC1 alone incises a 3' flap near the ss-dsDNA junction[124,145]. Critically, XPF incision specificity is modulated *in vitro* by interactions with RPA[121] and SLX4[127]. When fork or pseudo-Y DNA substrates are co-incubated with RPA, XPF incision is instead directed 6 nucleotides within the upstream duplex DNA[121]. XPF-RPA coordination is a requirement for NER function[54] and should also be present at a stalled replication fork. In complex with SLX4 however, XPF is dramatically stimulated (>100-fold) to make dual incisions approximately 1 and 5

nucleotides within upstream duplex DNA in a 3’ pseudo-Y substrate. This stimulated dual cleavage occurred even in DNA containing an ICL at the branch point[127]. Collectively then, XPF action is determined by its binding partners, enabling XPF to generate 5 and/or 3’ incision(s) within the duplex DNA (**Figure 1.6**). Though compelling, the exact location of XPF incision within ICL repair has yet to be firmly established *in vivo*. Importantly, additional nucleases have been implicated in ICL unhooking, potentially functioning in conjunction with XPF.

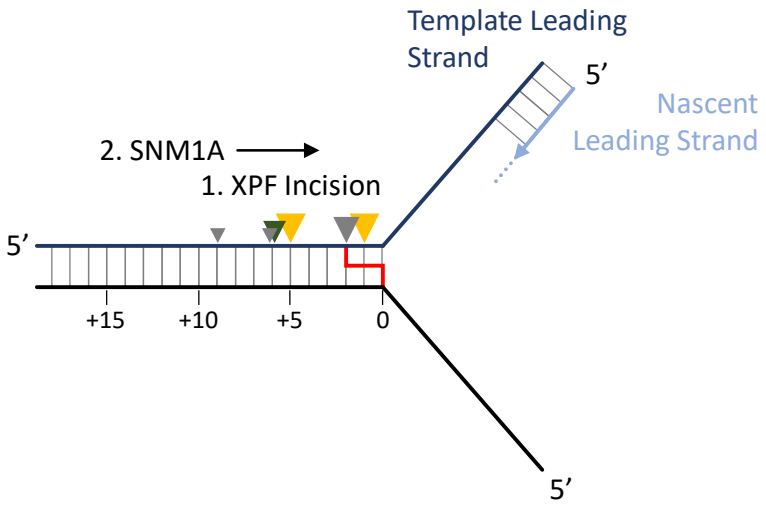


Figure 1.6. XPF incisions depend on DNA structure and binding partners.

XPF incision(s) upstream of the junction depend on its interacting partner. Arrowheads indicate expected location(s) of XPF incision(s): XPF alone (grey), XPF/SLX4 (yellow), XPF/RPA (green). Relative sizes indicate expected frequency. Incision location and frequency depend on size and nature of the crosslinker. Incomplete unhooking may provide SNM1A access to a 5’ phosphate (arrow indicates direction of digestion).

FAN1 is a dual-function 5' flap endonuclease[130] and 5-3' exonuclease[156], which participates in ICL repair[157–159] and fork stalling/ restart in response to broader replication stress[152,160]. FAN1 exonuclease processing is ideally suited for resolution of ICLs. Spatial separation between the DNA binding pocket and active site results in a flexible preference for generating triplet nucleotide products[130]. This unique catalytic mechanism, coupled with an open active site accommodates efficient bypass of crosslinks *in vitro*[92,130].

Depleted or knockout FAN1 mammalian cells exhibit a relatively mild survival hypersensitivity following exposure to ICL-inducing agents[154,157,159,161,162]. FAN1 is recruited to stressed replication forks through its ubiquitin binding (UBZ) and PCNA-interacting (PIP) domains via $FANCD2^{Ub}$ and $PCNA^{Ub}$, respectively[158,160,163]. Despite a direct interaction between FAN1 and the central FA effector ($FANCD2^{Ub}$), the relevance of FAN1 in the FA-pathway remains controversial. Epistasis experiments conflict, suggesting that FAN1 and FANCD2 function cooperatively at stressed forks, but not within ICL repair[154,159,161]. Further, FAN1 was first described in a screen identifying MutLα interactors, hinting at potential functions in RIR[164]. While FAN1 nuclease function is required for ICL repair, how it contributes is currently unclear. Importantly, FAN1 nuclease activities significantly overlap *in vitro* with SNM1A[154,165] and SAN1[162]. Redundant functions between these nucleases may be confounding experiments designed to uncover the biological function of each.

1.4 SNM1A

1.4.1 Structure

SNM1A is a dual-function nuclease. Multi-species alignment shows high amino acid conservation exists primarily in the C-terminal ~340 residues (**Figure 1.7**), encoding a nuclease domain composed of a metallo-β-lactamase (MβL) fold and β-CASP insertion casette[166]. The MβL domain belongs to a large superfamily of evolutionarily conserved enzymes which broadly catalyze ester hydrolysis. MβL-containing enzymes facilitate a variety of biological roles, including β-lactam hydrolysis-mediated antibiotic resistance, redox, thioesterase and phosphoryl transfer reactions, among others[167]. The MβL encodes a conserved binuclear Zn-binding active site (with the exception of Fe-binding redox MβLs), coordinated primarily by conserved histidine/ aspartic acid motifs[168,169].

Though the MβL domain catalyzes disparate biological reactions, insertion of the β-CASP fold confers nucleic acid specificity[167]. Dual MβL: β-CASP structures form a nuclease domain conserved in five human enzymes: SNM1A, SNM1B (Apollo), SNM1C (Artemis), CPSF-73 and ELAC2. These enzymes have distinct roles in nucleic acid processing and are broadly classified as RNA or DNA degrading enzymes. CPSF-73[170] and ELAC2[171] both preferentially degrade RNA and function in mRNA and tRNA maturation, respectively. SNM1 paralogs preferentially degrade DNA and function predominantly in ICL repair (SNM1A), telomere maintenance (Apollo)[172] or DSB end-processing (Artemis)[173,174]. Though sequence conservation between nuclease domains is quite poor (<30% identity), the tertiary structures are remarkably well conserved.

Initial characterization of the SNM1A nuclease domain described a robust 5-3' exonuclease activity, strictly requiring a terminal 5' phosphate[175–177]. Importantly, SNM1A exonuclease digestion can accommodate bulky nucleobase damage, including interstrand crosslinks[129,177–179]. Both full length[175,176] and the isolated nuclease domain[177] demonstrate a strong preference for single-strand over double-strand DNA. Further, SNM1A acts as a processive exonuclease on long DNA substrates (60-3000 nucleotides)[177,178], but not on shorter oligonucleotides (<60 nucleotides). Nuclease

activity is broadly stimulated by divalent cations, specifically Mg^{2+} and Mn^{2+}, as well as low concentrations of Zn^{2+} and Ca^{2+} [175,177]. Previous work in our lab identified novel SNM1A single-strand specific endonuclease activity. SNM1A endonuclease substrate preference is driven more by sequence (pyrimidine over purine) than DNA structure[179]. Unlike the hairpin opening activity of Artemis[173] or the yeast homolog Pso2[180], SNM1A endonuclease activity only requires stretches of single strand DNA. Importantly, endonuclease processing of single-strand DNA is significantly less efficient than SNM1A exonuclease processing. It is unclear whether the differential catalytic efficiency is modulated by post-translational modification (PTM) or binding partners, similar to DNA-PKcs stimulation of Artemis endonuclease activity[173].

1.4.2 Regulation

SNM1A activity and localization are strictly regulated *in vivo*. An extended, structured, 5' untranslated region (UTR) present in the primary SNM1A transcript (~900 nucleotides) dramatically reduces SNM1A translation[181]. Stable secondary structures within the UTR prevent translation initiation through cap-mediated ribosome scanning. Instead, SNM1A translation is facilitated by an internal ribosome entry site (IRES), predominantly associated with mitotic translation[181]. Interestingly, a secondary SNM1A transcript produced by alternative splicing of a non-coding exon significantly shortens the 5' UTR[182], though the direct impact on expression is unknown.

Additional spatial regulation of SNM1A in mammalian cells has been observed directly by immunofluorescence[183–185]. SNM1A nuclear localization is facilitated by an N-terminal (1-190) nuclear localization sequence (NLS), where SNM1A is sequestered into one or two large nuclear bodies[186]. These bodies are reversibly (dis)assembled, dependent on the cell-cycle or presence of DNA damage[183]. Following exposure to IR[183,184], UV[185] or ICL[183,185] damage, SNM1A-containing nuclear bodies disassemble, facilitating SNM1A mobilization to repair sites. SNM1A recruitment to IR and UV repair

foci suggests SNM1A either has an uncharacterized role in these processes or reflects the very small portion of IR/ UV damage which culminates in an ICL.

SNM1A contains two well-characterized interaction motifs (**Figure 1.7B**): the ubiquitin-binding zinc finger (UBZ; 119-149) and PCNA-interacting motif (PIP box; 556-561). Each motif directly interacts with ubiquitin and PCNA *in vitro*, respectively[185]. Substitution of the Zn-coordinating residues within the UBZ domain (C122, C125, H139 or C143) disrupts damage-inducible SNM1A localization to repair foci, while targeting the PIP box abolishes SNM1A foci formation entirely. Further, affinity of the SNM1A: PCNA interaction has been shown to improve synergistically following ubiquitination of PCNA by the Rad18 E3-ubiquitin ligase[185]. RAD18 completes an early signalling step, mono-ubiquitinating PCNA in response to replication stress or DNA damage. Though hundreds of proteins contain a PIP box[187], cells lacking RAD18 specifically failed to form damage inducible SNM1A repair foci[185]. Collectively then, this suggests that SNM1A is localized to DNA damage or replication stress following PCNA ubiquitination.

Spatiotemporal organization of SNM1A implies coordination with binding partners or PTM. Multiple binding partners have been inferred through co-immunoprecipitation (53BP1[183], APC3/CDC27[188] and ATM[184]) or yeast two-hybrid (PIAS1[189] and CSB[190]) experiments. Though the structural nature of these complexes is unknown, only interactions with CSB and ATM have been directly validated using purified recombinant proteins. Additionally, multiple whole-proteome analyses using human cell lines have catalogued numerous SNM1A PTMs. In response to DNA damaging agents, multiple sites undergo phosphorylation, ubiquitination and SUMOylation (**Figure 1.7B**). Similar mass spectroscopy experiments have not (yet) identified any SNM1A poly(ADP-Ribosyl)ation in response to damage[191,192]. Alternatively, an uncharacterized PAR-binding zinc finger (PBZ; 150-183) was recently identified in SNM1A, suggesting that it may instead read PARylation signalling[193–195].

Significant phosphorylation-mediated signalling cascades are coordinated by the PI3K-related kinases in response to DNA damage (ATM, DNA-PKcs) and replication stress

(ATR)[58,196]. ATM directly phosphorylates the N-terminus of SNM1A (1-393) *in vitro*. Though the specific modified sites were not identified, the SNM1A N-terminus contains (S/Q)T clusters canonically targeted by the ATM kinase[197]. Additional phosphorylation sites have been identified across SNM1A (highest abundance: S191[198,199], S238, S490, S590, S653)[200], though their functional relevance has yet to be established.

Modification with ubiquitin or small ubiquitin-like modifier (SUMO) commonly functions by altering protein stability, localization or binding partner interactions[196]. Early studies with SNM1A suggested that nuclear sequestration may be mediated via SUMOylation, as SNM1A interacted with the E3-SUMO ligase, PIAS1[189]. Though proteomic analysis of the SUMOylome has identified at least four sites within SNM1A (K202[199], K429[199], K488[199,201] and K508[199,202]), PIAS1 does not appear to be the responsible E3 ligase[203]. Interestingly, K202 SUMOylation is coordinated with tandem phosphorylation at S191 in response to MMC damage[199]. Moreover, competition between ubiquitination and SUMOylation at K488 and K508 has also been observed[199,204,205]. Mass spectroscopy-based experiments have provided a wealth of new, high confidence, modification sites. It is notable that very little modification has been detected within the nuclease domain, in any context. Instead, the vast majority of reported PTMs occur within disordered regions of SNM1A (residues ~200-600, **Figure 1.7C**). Future validation and characterization of individual (or paired) modifications will be required to understand how they function to regulate repair.

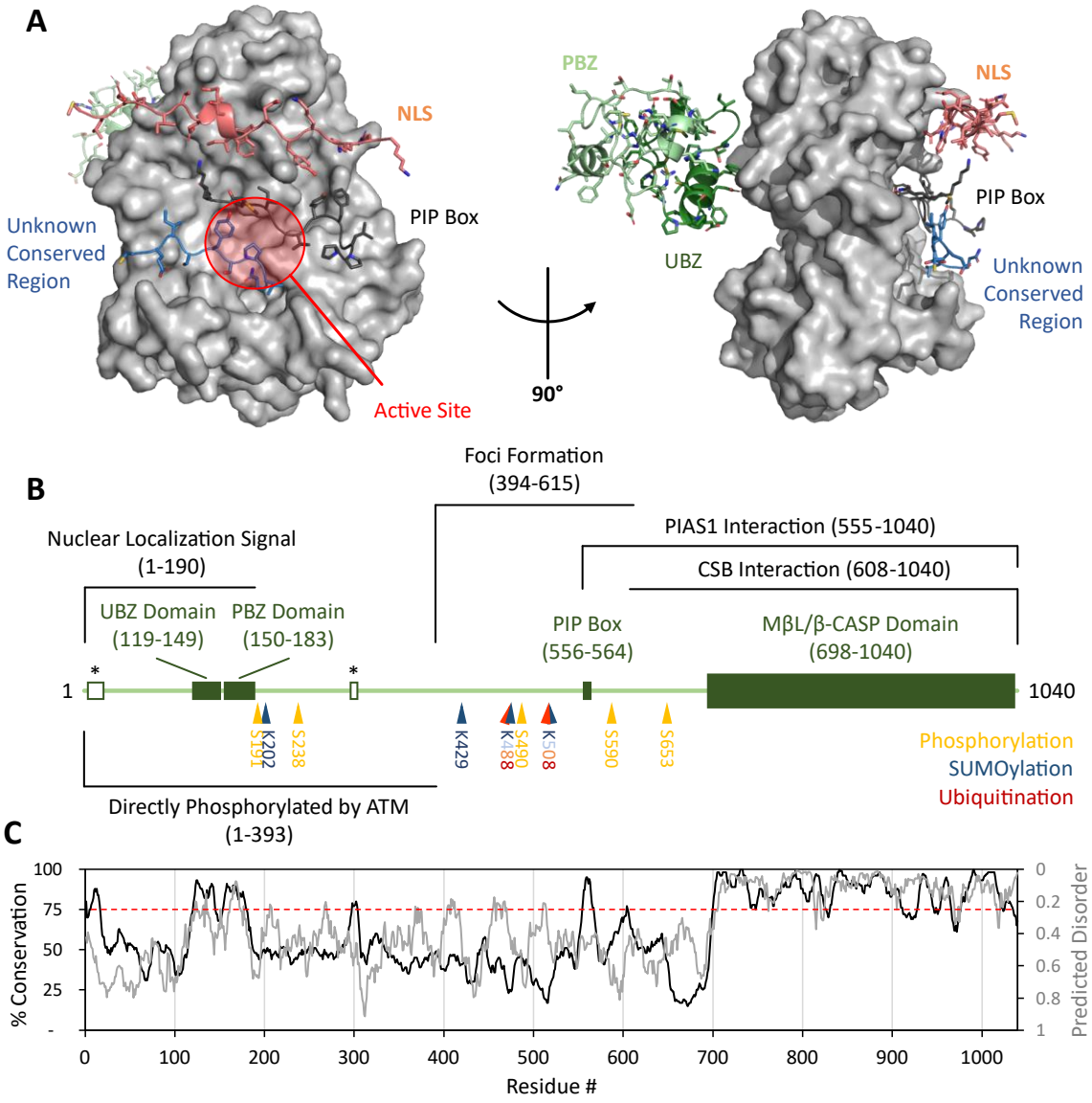


Figure 1.7. Schematic of SNM1A domain structure, post-translational modification, interspecies conservation and disorder.

Structural and sequence conservation of SNM1A. **[A]** Conserved regions of the AlphaFold tertiary structure prediction including a portion of the NLS (peach), UBZ domain (dark green), PBZ domain (light green), conserved region (294-308) of unknown function (blue), PIP box (black) and nuclease domain (grey, shown in surface). **[B]** Schematic of SNM1A interaction and domain organization. High confidence post-translational modification sites are indicated in yellow (phosphorylation), blue (SUMOlyation) and red (ubiquitination). * indicates regions of high conservation lacking explicit function. **[C]** Amino acid identity conservation between species (black line)[206] and PrDOS protein disorder prediction (grey line)[207].

1.4.3 Biological Function

SNM1A has been shown to function in the cell cycle, apoptosis, DNA damage response and the repair of ICLs, however defects in SNM1A are primarily associated with DNA damage repair dysfunction. Though there are no diseases known to be caused by SNM1A point mutations, SNM1A function has been modelled using mouse knockouts. Homozygous SNM1A knockout mice are viable, born at expected mendelian ratios and fertile. Simultaneously, knockout mice display advanced tumorigenesis, sensitivity to MMC and selective susceptibility to infection[186,197,208]. Though severity differs between cohorts, mice predominantly die of hematologic cancers between 12-16 months old[197]. Double knockout mice with p53 ($SNM1A^{-/-}$, $p53^{-/-}$) significantly increase tumorigenesis, suggesting SNM1A functions as a tumor suppressor[197].

The DNA damage response is a complex biological network of signalling molecules and post-translational modifications. In addition to specific response factors, many enzymes directly involved in lesion repair have additional signalling functions. These signalling networks coordinate repair with DNA replication, the cycle cell and apoptosis[209]. Cells lacking SNM1A demonstrate micronucleation and rapid aneuploidy following exposure to spindle poisons in the absence of any exogenous DNA damaging agents[188]. Though mechanistically unclear, SNM1A contributes to a prometaphase checkpoint, distinct from the well characterized metaphase spindle checkpoint. Furthermore, SNM1A was shown to interact by co-immunoprecipitation with 53BP1[183,184] and APC3/CDC27[184]. Whether SNM1A directly interacts with either protein is unclear, though it suggests SNM1A coordinates with cell cycle regulation and checkpoint activation through modulation of or by the anaphase promoting complex.

Knockout[186,189,197,208] or siRNA depletion of SNM1A in chicken[189], mice[154,186,197,208] or human[129,162,208] cells elicits a mild survival hypersensitivity following exposure to ICL-inducing agents, but not UV, MMS or IR damage[186]. Further, SNM1A deficient cells treated with ICL damage exhibit substantial chromosomal instability, primarily through accumulation of DNA double strand breaks[129] and radial chromosomes[208]. As SNM1A-

deficient cells are not sensitive to other DSB-inducing drugs, SNM1A appears required for processing an ICL-specific break intermediate. Where cells lack SNM1A processing, 'DSB repair' of this intermediate fails.

ICL repair is coordinated through a network of competing/ overlapping pathways (as outlined previously). Therefore, determining where SNM1A functions requires careful epistatic mapping. Initial analysis using chicken DT40 combination knockouts identified potential redundancy with SNM1B and non-epistasis with FA ($FANCC^{-/-}$), TLS ($RAD18^{-/-}$) and HR ($XRCC3^{-/-}$) following ICL exposure. In combination with SNM1A, each of the non-epistatic groups demonstrated additive hypersensitivity to cisplatin[189]. Cytogenic analysis of SNM1A depletion, in FANCA-deficient human fibroblasts, resulted in a synergistic increase in radial chromosome accumulation, beyond what was observed following ERCC1 disruption. In contrast, co-depletion of ERCC1 or SLX4 with SNM1A in HeLa and U2OS cell lines highlighted a potential epistatic relationship between the two nucleases (XPF-ERCC1 and SNM1A) within the FA pathway[129]. Consequently, coordinated unhooking facilitated by XPF and SNM1A could occur in ICL repair.

FA-dependent monoubiquitination of FANCD2 is a prerequisite for SLX4-XPF recruitment to ICL damage in RDR[123]. Concurrently, PCNA is monoubiquitinated in response to replicative stress. SNM1A is directly recruited to $PCNA^{Ub}$ via the PCNA interacting motif (PIP Box) and ubiquitin binding domain (UBZ) in response to ICL damage[185]. Suggestive *in vitro* models have shown that XPF can mediate unhooking alone[127] or in concert with SNM1A processing from an initial XPF-dependent 5' incision (**Figure 1.6**)[120,121]. It is possible that both models function *in vivo*, where XPF-dual incision unhooking depends on the identity of the crosslinker. If 3' XPF incision fails, coupling with a 5-3' translesional exonuclease like SNM1A would resolve the partially unhooked intermediate.

XPF is also recruited to ICL lesions recognized by RNAP stalling through assembly of the NER incision complex[54,73]. Furthermore, a subset of cellular SNM1A is

constitutively associated with CSB, the essential RNAP-associated translocase and chromatin remodeler[190]. A similar dynamic as in RDR may function here where unhooking is mediated by XPF and XPG coordinated incisions. If XPF-XPG fail to complete unhooking, SNM1A may be required to resolve the intermediate.

SNM1A demonstrates a relatively unique ability to exonucleolytically digest DNA substrates past an ICL lesion[129]. Translesional activity has also been observed in FAN1[92,130] and (more recently) SAN1[162]. Though the biological role of each nuclease may be distinct, epistasis experiments identified synergistic increases in survival hypersensitivity between each gene pair[154,162]. This may reflect either distinct activity in separate, parallel pathways or functional redundancy. Both SAN1 and FAN1 seem to primarily function outside of the FA-pathway[122,162]. As SNM1A has been implicated in both FA-dependent[129] and FA-independent[189,190] repair, it remains to be determined whether (or how) the nuclease functions overlap. It is possible that ICL intermediates are processed by a 'pool' of competing translesional exonucleases, each with distinct preferences to collectively repair diverse subsets of unhooking products.

1.5 Thesis Objectives

SNM1A participation in ICL repair is well established, though its actual role remains ambiguous. Experiments presented in this thesis address key questions surrounding SNM1A: characterization of dual-function nuclease activities, identification of specific small molecule inhibitors and validation of the nuclease functions in repair.

Careful characterization of SNM1A nuclease function is necessary to understanding where/how SNM1A may be contributing to repair. Recent work in our lab identified a previously undescribed single-strand dependent endonuclease activity. To contextualize this function, significant extension of prior *in vitro* SNM1A characterization was necessary. The objective of **Chapter 2** was to thoroughly characterize differences in the nuclease functions. Here, a robust purification of the SNM1A nuclease domain was

established using heterologous expression in *E. coli*. Functional analysis involving a variety of DNA substrates was conducted to explore relative binding affinities, nuclease efficiencies and substrate preferences. Gaining a more complete understanding of ideal substrates enable better models to be constructed of SNM1A action in repair.

Inhibition of SNM1A is a high priority target, due to its suggested coordination with XPF in ICL repair. To that end, our lab completed a high-throughput screen of bioactive compounds to identify potential candidates[210]. The objective of **Chapter 3** was to further examine and characterize affinity, mode of action and specificity of SNM1A bioactive inhibitors. This chapter also discusses the results of a complementary *in silico* approach to improve SNM1A inhibitor specificity.

Recent identification of SNM1A dual nuclease functions expanded the potential role(s) of SNM1A in ICL repair. SNM1A nuclease function is required for its role in repair, though the single active site catalyzes two distinct DNA substrates. The objective of **Chapter 4** is to examine the biological relevance of SNM1A exonuclease and endonuclease activity individually within repair. To accomplish this, structural models of SNM1A were generated leveraging well-established characteristics of a phosphoryl transferase 2-metal ion reaction to approximate expected binding of short, single-stranded exonuclease and endonuclease DNA substrates. These models were exploited to identify single or multiple amino acid substitutions that resulted in separation-of-function. Using a yeast model system, separation-of-function mutants were recapitulated in the SNM1A yeast homolog, Pso2. Here we show that both exonuclease and endonuclease functions of Pso2 are essential for its role in ICL repair.

Chapter 2
Functional characterization of SNM1A nuclease activity

2.1 Preface

Dr. Simon Huang generated the initial bacterial expression plasmid. Sam Chu assisted with generation of baculovirus for SNM1A expression in insect cells. Dr. Beverlee Buzon assisted in substrate design, crosslinking and purification of DNA in Figure 2.9. I completed all other experiments presented below.

Data in Figure 2.9 has been previously published and is reproduced here with copyright permission:

Buzon B., Grainger R., Huang S. *et al.* (2018) Nucleic Acids Res, 46(17): 9057-9066
DOI: 10.1093/nar/gky759

2.2 Purpose

SNM1A plays a critical role in human interstrand crosslink repair, though how it contributes remains ambiguous. The purpose of this chapter is to establish an efficient method to recombinantly generate pure and stable SNM1A protein enabling analysis of models describing SNM1A in repair through expanded functional characterization of metal-dependence and substrate preferences.

2.3 Introduction

Interstrand crosslinks (ICLs) are a form of DNA damage resulting from the covalent linkage of opposing DNA strands[5,7]. ICLs are particularly cytotoxic because they prevent strand separation, disrupting both transcription and replication across the lesion. Failure to efficiently repair ICL damage results in the accumulation of clastogenic chromosomal breaks, often prompting either apoptosis or tumorigenesis[2,211]. Nucleases play critical roles in the initiation, excision, and resolution of ICL repair intermediates[145,155]. As such, significant effort has been invested in identifying the individual roles of many nucleases implicated in ICL repair. This effort has been complicated by (i) functional redundancy between nucleases and (ii) multi-functional nucleases with broad substrate specificities.

SNM1A is a dual-function 5’ phosphate-dependent, 5-3’ exonuclease[175,176] and single-strand dependent endonuclease[179] required for the repair of interstrand crosslinks[186,208]. Disruption of SNM1A results primarily in sensitivity to ICL-inducing drugs and accumulation of chromosomal aberrations (DSBs and radial chromosomes)[129,208]. SNM1A nuclease activity participates directly in repair, as active site mutants fail to rescue ICL-associated phenotypes of SNM1A knockout/ knockdown cells[176,189]. Because both SNM1A exo- and endo-nuclease activities are catalyzed by a common active site, current evidence does not distinguish which function participates in repair. Early reports identified a robust exonuclease activity *in vitro*, particularly on large

plasmid substrates[177]. Interestingly, SNM1A can also digest translesionally past crosslinked DNA strands[129,177]. This relatively unique translesional capacity implies a potential biological role where SNM1A could process DNA 5' of an ICL to unhook or trim past the lesion. Recent work published from our lab identified an additional single-strand dependent endonuclease activity. In contrast to other structure-specific endonucleases involved in ICL repair, SNM1A endonuclease activity had little directional or structural preference on a variety of DNA intermediates including bubble or replication fork substrates[179]. Importantly, metal availability impacts relative SNM1A nuclease activities and preferences. Proper definition of the metal requirements of SNM1A is necessary for extrapolating any endogenous function. Further biochemical characterization is necessary to elucidate the biological functions facilitated by SNM1A.

Significant ambiguity remains regarding the biological role of SNM1A in ICL repair. SNM1A has been broadly implicated in repair, functioning in both RIR and RDR pathways. Extended discussion in section **1.4.3** outlines potential mechanisms of action for SNM1A within the context of global-genome, transcription-coupled and replication-dependent repair pathways. Briefly, GG-ICL repair is comprised of multiple competing pathways which recognize and remove ICL lesions[36,93]. As discussed in **1.3.2.1**, multiple putative pathways involving NER- or MMR-dependent excision produce a small gap adjacent to the ICL lesion[31,83]. Where these gaps are generated 5' of the ICL, the 5-3' translesional exonuclease activity of SNM1A (or similar nuclease) may play a critical role in lesion unhooking. SNM1A function within TC-ICL repair has been implied through direct interaction with, and stimulation by, the repair coupling factor CSB[190]. While common bulky damage is removed by flanking XPF and XPG endonuclease incisions[54,55], the mechanism of transcription-coupled removal of an ICL is less clear. As discussed in **1.3.2.2**, both XPF and XPG are required for TC-ICL repair, however the nature of the product depends on where XPG acts relative to the ICL[4,32,97,98]. Where XPG endonuclease incision is abrogated by the ICL structure, SNM1A exonuclease activity could be essential. Regardless, as RIR does not generate a DSB intermediate[76,99], direct involvement of SNM1A endonuclease activity would be expectedly limited.

Within RDR, SNM1A functions epistatically with XPF[129]. While XPF-dependent endonuclease incision(s) have been directly demonstrated using *Xenopus* extracts[122], the role of SNM1A remains speculative. FA-activation promotes monoubiquitination of FANCD2[110], which in turn localizes XPF to the site of damage through its scaffold partner, SLX4[123]. XPF can incise ICLs 5', 3' or both, to partially or completely unhook an ICL adduct[120,121,127,212]. SNM1A deficiency results in the accumulation of unrepairable double-strand breaks[129,208], suggesting that SNM1A may act on these products to generate intermediates amenable to downstream homologous recombination. Whether SNM1A acts via translesion exonuclease, hairpin opening, or other endonuclease processing is unclear.

Direct evidence to validate proposed roles of SNM1A function in any ICL repair pathway is lacking. In this chapter, SNM1A activities on various DNA structures and ICL repair intermediates are examined. Characterization of SNM1A exonuclease, translesional and endonuclease activity on these substrates define the relative efficiency of each and how they may collaborate in repair. Further, affinity and metal titration studies identify distinct catalytic mechanisms for exonuclease and endonuclease processing. Extending SNM1A characterization will help to better establish where SNM1A functions and which intermediate(s) it processes in repair.

2.4 Materials and Methods

2.4.1 Primers and DNA Substrates

Primer sequences used in all cloning, sequencing and mutagenesis experiments are summarized in **Primer Information**. All DNA substrates and oligonucleotides are summarized in **DNA Substrate Diagrams** and **DNA Oligonucleotide Information**, respectively. All DNA concentrations were determined using NanoDrop™ 2000c Spectrophotometer (Thermo Scientific).

2.4.2 Construct Design and Cloning

2.4.2.1 Bacterial Expression Constructs

Full-length SNM1A (UniProt KB ID# Q6PJP8) was codon optimized for expression in *E. coli* (**Appendix 2.A**), synthesized and cloned into a pET24a plasmid (Geneaid). SNM1A bacterial expression constructs were generated using Gateway™ cloning (Invitrogen). SNM1A was flanked between attB1/2 recombination sites and truncated to the nuclease domain (698-1040) by PCR amplification of 10 ng $SNM1A^{1-1040}$ (in pET24a) with 100 ng of N697.SNM1A-F and N697.SNM1A-R primers, and 2x PCR Bestaq Mastermix (Diamed). N697.SNM1A-F created a flexible linker and TEV protease recognition site on the N-terminus of SNM1A (**Appendix 2.B**). SNM1A PCR product (100 ng) was recombined into pDONR-201 (100 ng, Thermofisher) with 0.7 U Gateway™ BP Clonase™ II enzyme mix (Invitrogen; BP Clonase) at room temperature for 4 hours. The reaction was terminated with Proteinase K (0.6 U, Thermofisher) at 37 °C for 10 minutes and transformed by heat shock into One Shot™ TOP10 chemically competent *E. coli* (TOP10, Invitrogen). Cells were plated on Lennox Broth (LB, Bioshop) with 50 µg/mL kanamycin selection at 37 °C overnight. Individual transformants were grown in LB liquid culture under kanamycin selection overnight at 37 °C and 220 rpm to saturation. Plasmids were extracted using Presto™ Mini Plasmid Kit (Geneaid) following

manufacturer instructions. Constructs were validated by sequencing with pDONR201-F or pDONR201-R primers (DNA Sequencing Facility, Robarts Research Institute; RRI).

Sequence verified wildtype SNM1A$^{\Delta N697}$ (further referred to as SNM1A) in pDONR201 (150 ng) was recombined with pDEST-527 (150 ng) and 0.7 U Gateway™ LR Clonase™ II (LR Clonase, Invitrogen) at room temperature for 2 hours. Bacterial expression plasmid pDEST-527 was a gift from Dominic Esposito (Addgene plasmid #11518). Reactions were terminated with 0.6 U Proteinase K at 37 °C for 10 minutes. Recombination products were transformed into TOP10 and plated on LB with 100 µg/mL ampicillin. Successful transformants were individually grown to saturation in liquid LB cultures with ampicillin selection and plasmids were extracted using the Presto™ Mini Plasmid Kit (Geneaid).

2.4.2.2 Insect Cell Expression Constructs

SNM1A (698-1040) cloned in pFB-LIC-Bse was a gift from Opher Gileadi (SGC, Oxford University; pFB-LIC-Bse backbone, Addgene plasmid #26108). Plasmid was verified by sequencing with pFB-F or pFB-R primers (DNA Sequencing Facility, RRI). pFB-LIC-Bse containing N-terminal His-tagged SNM1A$^{\Delta N697}$ (10 ng) was transformed into MAX Efficiency™ DH10Bac Competent Cells (Thermofisher) by heat shock and grown for 4 hours in Super Optimal Broth (Thermofisher) at 37 °C and 225 rpm. The cell suspension was plated on LB under 50 µg/mL kanamycin, 7 µg/mL gentamycin and 10 µg/mL tetracycline selection, with 100 µg/mL X-gal and 40 µg/mL Isopropyl β-D-1-thiogalactopyranoside (IPTG), and grown at 37 °C for 48 hours. White colonies were individually grown to saturation in liquid LB culture under identical selection. Baculovirus shuttle vector (Bacmid DNA, bMON14272) was purified from each using PureLink™ HiPure Plasmid DNA Miniprep Kit (Thermofisher). Successful Tn7-dependent recombination between pFastBac and Bacmid was verified by PCR amplification of each purified Bacmid (100 ng) with 100 ng of pUC/M13-F and pUC/M13-R primers. Validated Bacmid constructs were stored at 4 °C in TE pH 8.0 until transfection.

Spodoptera frugiperda (Sf21) insect cells (Gibco) were grown in Sf-900^{TM} II SFM media (Gibco) at 27 °C and 135 rpm until mid-log phase (2x10^6 cells/mL). Cell density was determined by hemocytometer count using 0.04% v/v Trypan Blue to differentiate live cells. Sf21 cells were pelleted by centrifugation at 130 x g for 3 minutes and resuspended in unsupplemented Grace's media (Gibco) to seed 6-well plates to 8x10^5 cells. Following seeding, cells were allowed to adhere at 27 °C for 15 min. SNM1A-containing Bacmid (1 µg) and 8 µL Cellfectin II (Gibco) were individually mixed with 100 µL of unsupplemented Grace's media, prior to combining. The transfection mixture was incubated at room temperature for 30 minutes and added dropwise to the adherent Sf21 cells. Following incubation for 4 hours at 27 °C, media was refreshed with Grace's media supplemented with 10% v/v Fetal Bovine Serum (FBS, Wisent). Cells were grown at 27 °C for an additional 72 hours before media was collected and clarified by centrifugation at 500 x g for 5 minutes. Clarified media contained p1 SNM1A-Baculovirus stock and was stored at -80 °C until needed for expression.

2.4.3 Protein Expression

2.4.3.1 Bacterial Expression

SNM1A$^{\Delta N697}$ pDEST-527 was transformed by heat shock into Rosetta^{TM} (DE3) pLysS *E. coli* (Novagen) and plated on LB with 100 µg/mL ampicillin. Liquid LB cultures under ampicillin selection were inoculated by multi-colony streak of transformants and incubated at 37 °C overnight at 225 rpm. LB or Terrific Broth (TB) cultures were inoculated with the saturated overnight (1:100 v/v) and incubated at 37 °C and 225 rpm. Cultures were grown to 1.2 or 1.8 OD_{600} (for LB or TB, respectively), cooled on ice to room temperature and induced with 1 mM IPTG before incubation at 25 °C for 12 hours at 225 rpm. Ethanol (5% v/v) was added to room temperature samples requiring stress-induced expression, 10 minutes prior to addition of IPTG. Cells were collected by centrifugation at 3,315 x g for 15 minutes, resuspended with bacterial NiA buffer (bNiA:

500 mM NaCl, 50 mM Tris 7.5 or Sodium Phosphate buffer pH 6.8, 10% glycerol and 0.5 mM TCEP), flash frozen in liquid nitrogen and stored at -80 °C until purification.

2.4.3.2 Insect Cell Expression

High Five™ *Trichoplusia ni* cells (High Five, Thermofisher) in Sf-900™ II SFM media were grown to an approximate density of $1x10^6$ cells/mL prior to infection with 1:1000 v/v SNM1A-containing baculovirus (p1 viral titre). Cells were incubated for 96 hours at 27 °C and 135 rpm post-infection. Samples were collected by centrifugation at 500 x g for 5 minutes, resuspended into insect NiA buffer (iNiA: 500 mM NaCl, 50 mM HEPES pH 7.5, 5% v/v glycerol, 1 mM TCEP, 10 mM imidazole), flash frozen in liquid nitrogen and stored at -80 °C until purification.

2.4.4 SNM1A Purification

2.4.4.1 SNM1A (Bacterial Expression)

Cell suspensions were thawed from -80 °C and diluted to at least 1:8 m/v (g pellet/ buffer volume) with bNiA containing 0.01% v/v Triton X-100 and protease inhibitors (3 µM aprotinin, 1 µM pepstatin A, 1 mM benzamidine, 1 µM leupeptin and 1 mM PMSF). Purification buffers contained Tris pH 7.5 or Sodium Phosphate pH 6.8 prior to or after optimization (**Figure 2.3**), respectively. Cells were lysed with three passes at 10,000 psi by cell disruptor (Avestin Emulsiflex C3), clarified by centrifugation at 48,384 x g for 60 minutes and vacuum filtered (0.45 µm Triton-Free MCE, Millipore). Clarified lysate was loaded on a 5 mL EDTA-resistant Ni-Penta™ IMAC (Marvelgent) column using the NGC Chromatography System (BioRad), washed with bNiA containing 15 mM imidazole and step-eluted with 300 mM imidazole. Eluted protein was buffer exchanged (HiPrep 26/10, GE Healthcare) into bNiA, diluted below 0.5 mg/mL protein and incubated with 1:10 m/m TEV protease at 4 °C overnight. Treated sample was diluted to bNiA containing 5 mM imidazole and loaded on a 1 mL nickel-chelating IMAC HisTrap HP (GE Healthcare) column. The sample was washed sequentially with 5, 20, 35, 50 and 300

mM imidazole. Samples containing SNM1A were pooled and diluted with SA buffer (50 mM Tris pH 7.5 or Sodium Phosphate pH 6.8, 10% v/v glycerol, 0.5 mM TCEP) to 250mM NaCl and loaded onto a 1mL HiTrap SP HP (GE Healthcare) column. The sample was eluted with a gradient using SB buffer (1 M NaCl, 50 mM Tris pH 7.5 or 50 mM Sodium Phosphate pH 6.8, 10% v/v glycerol) from 250 to 750 mM at 8.3 mM/ Column Volume (CV). Pure SNM1A was exchanged (HiPrep 26/10, GE Healthcare) into Storage Buffer (200 mM NaCl, 10 mM Tris pH 7.5 or Sodium Phosphate pH 6.8, 5% glycerol, 0.5 mM TCEP), concentrated with a Vivaspin Turbo15 10 kDa MWCO Centricon (Sartorius) at 1000 x g to 128 µM (5 mg/mL) and aliquoted into thin-walled PCR tubes, flash frozen in liquid nitrogen and stored at -80 °C until needed.

2.4.4.2 SNM1A (Insect Expression)

Cell suspensions were thawed from -80 °C and resuspended into 25mL iNiA Buffer/ 250mL culture, with protease inhibitors (same as above). Cells were lysed by sonication (QSonica Q155) with three passes at 40% amplitude, 1 s pulse rate for 1 minute. Lysate was clarified by centrifugation at 21,100 x g at 4 °C for 75 minutes and loaded onto a 1mL nickel-chelating IMAC HisTrap HP (GE Healthcare). Protein was washed and eluted with iNiA Buffer containing 15.8, 24.5, and 39 mM and 213 mM imidazole, respectively. 1:6 m/m TEV protease was added to eluted protein and dialyzed (1:100 v/v) serially twice with into 500 mM NaCl, 50 mM HEPES pH 7.5, 5% v/v glycerol and 1 mM TCEP at 4 °C, overnight. Treated sample was diluted with iNiB Buffer (500 mM NaCl, 50 mM HEPES pH 7.5, 5% v/v glycerol, 1 mM TCEP, 300 mM imidazole) to iNiA Buffer and loaded onto a 1 mL nickel-chelating IMAC HisTrap HP. Sample was washed with 24.5, 31.75, 29, 82.5, 126 and 300 mM imidazole and fractions containing pure SNM1A were combined. Pure SNM1A was concentrated with Vivaspin Turbo15 10 kDa MWCO Centricon at 1000 x g to 25 µM (1 mg/mL) and aliquoted into thin-walled PCR tubes, flash frozen in liquid nitrogen and stored at -80 °C until needed.

2.4.5 Western Blot

Purification samples (**Figure 2.1**) were separated by 15% SDS-PAGE and visualized with 0.5% 3,3,3'-trichloroethanol (TCE) stain free imaging by GelDoc EZ Imager (BioRad). Samples were transferred to activated PVDF (Immobilon-P, Millipore) and incubated in a blocking solution with 5% Bovine Serum Albumin (BSA) in Tris-buffered saline and 0.1% v/v Tween-20 for 1 hour. The blot was then incubated with mouse IgG monoclonal anti-polyhistidine antibodies (1:2000, H1029 Sigma) in 3% BSA TBS-T for 2 hours at room temperature. Following 3 washes, the blot was incubated with rabbit anti-mouse IgG-HRP conjugated antibodies (1:4000, A9044 Sigma) in 3% BSA TBS-T. The blot was visualized following treatment with enhanced chemiluminescent solution (100 mM Tris pH 8.8, 2 mM 4-iodophenylboronic acid, 1.25 mM luminol and 5.4 mM H_2O_2) using a ChemiDoc MP Imager (BioRad), detecting chemiluminescence.

2.4.6 Thermal Shift Assay

SNM1A (25 µM) was mixed 1:1 v/v in a 96-well plate with conditions from Durham pH (MD1-101) or Salt (MD1-102) screens and 1:2000 v/v SYPRO™ Orange Dye (Thermofisher), up to 20 µL final volume. Fluorescence emission at 595 nm was monitored across a temperature gradient from 20-90 °C at 0.05 °C/s using the QuantStudio™ 3.0 Real-Time PCR System (Applied Biosciences). SNM1A Melting Temperature (T_m) in each condition was determined as the temperature at which the rate of change in F_{595} (dF_{595}) was maximal. Derivative plots of fluorescence were determined and analyzed using Protein Thermal Shift™ software v1.2 (Applied Biosciences).

2.4.7 Size Exclusion Chromatography (SEC) and Multi-Angle Light Scattering (MALS)

Linear regression between known High Molecular Weight Markers (GE healthcare) molecular weight and observed retention volumes in Storage Buffer was

used to calibrate a Superdex 200 10/300 GL (GE Healthcare). Indicated concentrations of SNM1A (25 or 125 µM) were injected onto a Superdex 200 10/300 GL equilibrated with Storage Buffer, run at 1 mL/min. Absorbance was monitored at 280 nm (**Figure 2.6**, **Figure 2.8**) or 255 nm (**Figure 2.12**) between void and column volume.

Figure 2.8. In-line Multi-Angle Light Scattering (MALS) instrument and differential refractometer (DAWN, Wyatt Technologies Inc), in combination with ASTRA analysis software (Wyatt Technologies), detected and analyzed the apparent molecular weight of the SNM1A (25 µM) peak.

Figure 2.12. SNM1A$^{D736A/H737A}$ (25 µM) was incubated with or without **F.14T** DNA (2.5 µM) at room temperature for 15 minutes. SNM1A only, DNA only and SNM1A-DNA samples were injected separately and absorbance traces (255 nm) were overlaid.

2.4.8 Oligo Purification and Annealing

Lyophilized DNA oligos were designed and ordered from BioBasic Inc (Markham, Ontario). Unlabelled oligos were resuspended to 100 µM in water. Fluorophore (6-FAM) labelled oligos were resuspended in Formamide Buffer (95% formamide, 5 mM EDTA), and purified using 22.5% denaturing PAGE. The band of interest was excised, submerged in Elution Buffer (200 mM NaCl, 10 mM Tris pH 7.5, 5 mM EDTA), and incubated at 37 °C and 225 rpm overnight, protected from light. The eluent was ethanol precipitated with 70% v/v ethanol and 75 mM sodium acetate at -20 °C, washed with 70% v/v ethanol and allowed to air dry. Dried pellets were resuspended with water and stored at -20 °C until needed.

Multi-oligonucleotide substrates were annealed with 1 µM labelled DNA and 1.2 µM cold DNA(s) oligonucleotide(s) in Annealing Buffer (100 mM NaCl, 10 mM Tris pH 7.5, 10 mM $MgCl_2$) at 90 °C for 10 minutes. Annealing reactions were then slowly cooled to room temperature over ~1 hour. Annealing efficiency was confirmed by Native PAGE

(10-20% depending on substrate size) at 100 V. Concentration of cold DNA was successively increased, as necessary, to achieve sufficient substrate annealing efficiency.

Hairpin DNA substrates (1 µM) were incubated in Annealing Buffer at 90 °C for 10 minutes and snap-cooled on ice. Putative hairpin substrates were verified by diagnostic digestion with T7 Endonuclease I (New England Biolabs). Substrates (100 nM) were incubated with 1 U of T7 Endonuclease and NEB2 buffer at 37 °C for 30 minutes. Digested products were analyzed using 22.5% denaturing PAGE and visualized with a ChemiDoc MP Imager. Substrates with correct secondary structure were specifically degraded at the hairpin apex.

2.4.9 SJG Substrate Crosslinking

DNA substrates were engineered to have a single SJG crosslinking site (5'-GATC-3') and annealed to 10 µM, as described above. Annealed substrates were crosslinked with 40x molar excess of SJG-136 at 37 °C overnight in 25 mM triethanolamine, 1 mM EDTA, pH 7.2. Crosslinked reactions were ethanol precipitated, resuspended into Formamide Buffer and purified using 22.5% denaturing PAGE. Gels were visualized by ChemiDoc MP Imager and the crosslinked band was excised and eluted, as described above. Eluent was ethanol precipitated and resuspended to 1 µM in water. Crosslinking efficiency was determined using 22.5% denaturing PAGE analysis of each resuspended crosslinked substrate (1 pmol). 'Percent Substrate Crosslinked' was determined using **Equation 1**, using pixel integration of crosslinked and uncrosslinked bands for each substrate in ImageLab (Biorad).

Equation 1

$$\%\ Substrate\ Crosslinked = \frac{Crosslinked\ Band\ Intensity}{Total\ Lane\ Intensity} \times 100\%$$

2.4.10 Nuclease Assays

2.4.10.1 Protein Dose-Dependent Assays

SNM1A was serially diluted from 25 µM using Storage Buffer to indicated concentrations. SNM1A dilutions in Nuclease Reaction Buffer (50 mM Tris-acetate pH 7.2, 75 mM potassium acetate, 10 mM $MgCl_2$, 1 mM DTT and 100 µg/mL BSA) were incubated with 100 nM DNA substrate at 37 °C for 30 minutes. Reactions were terminated with Formamide Buffer and products were separated by 22.5% denaturing PAGE. Gels were visualized by ChemiDoc MP Imager detecting fluorescence at 526 nm.

2.4.10.2 Time Dependent Assays

SNM1A in Nuclease Reaction Buffer was incubated with 100 nM DNA substrates at 37 °C for indicated times. Reactions were terminated with Formamide Buffer, products were analyzed by 22.5% denaturing PAGE and visualized by ChemiDoc MP Imager detecting fluorescence at 526 nm.

Figure 2.9. SNM1A (40 nM) was incubated with DNA substrates, with or without a flap or SJG ICL, for 2 or 60 minutes.

Figure 2.10. SNM1A (5 nM) was incubated with **P.Nicked** or **3' Overhang** DNA substrates for 0-40 minutes. Quantification reflects the mean of three independent replicates. 'Percent DNA Processed' was determined by **Equation 2**, using band integration in ImageLab.

Equation 2

$$\%\ DNA\ Processed = 1 - \frac{Remaining\ Substrate\ Intensity}{DNA\ Alone\ Intensity} \times 100\%$$

Figure 2.11. SNM1A (100 nM) was incubated with **F.xT** DNA substrates (*x* refers to the length of substrate and is indicated for each substrate) for 30 minutes.

2.4.10.3 Metal Titration Assays

SNM1A (purified as described in section **2.4.4.1**) was incubated in 5% v/v DMSO with or without 50 mM EDTA and 5 mM o-phenanthroline on ice for 2 hours. Each treatment sample was then buffer exchanged (HiPrep 26/10, GE Healthcare) into Storage Buffer, aliquoted into single-use tubes and flash frozen in liquid nitrogen.

SNM1A (treated or untreated) in EMSA Buffer (50 mM Tris acetate pH 7.2, 75 mM potassium acetate, 1 mM DTT and 100 µg/mL BSA) was incubated with or without indicated metals (1-10,000 µM) for 5 minutes at room temperature. SNM1A concentrations used in metal titration studies were optimized by protein-dose dependent nuclease assays such that untreated SNM1A digested ~50% of the DNA substrate (5 nM, exonuclease; 100 nM, endonuclease). Nuclease reactions were initiated by addition of exonuclease (**P.20TF**) or endonuclease (**F.20T**) DNA substrate (100 nM) and incubated at 37 °C for 15 or 30 minutes, respectively. Nuclease reactions were terminated with Formamide Buffer and products were analyzed by 22.5% denaturing PAGE and visualized by ChemiDoc MP Imager at 526 nm.

2.4.11 Mutagenesis

Primers were constructed according to the one-step, site-direction mutagenesis protocol described by Liu and Naismith (2008)[213]. $SNM1A^{WT}$ in pDONR201 (as described in **2.4.2.1**) template (10 ng) was PCR amplified with 100 ng D736A/H737A-F and D736A/H737A-R primers and 2x Bestaq Mastermix (Diamed). PCR products were exchanged by PCR Cleanup Kit (Geneaid) into water, according to manufacturer instructions. PCR products were then incubated with 1x Tango Buffer (Thermo Scientific) and 10 U Dpn1 (Thermo Scientific) at 37 °C for 60 minutes. Products were transformed by heat shock into TOP10, plated on LB with 50 µg/mL kanamycin selection and grown at 37 °C overnight. Individual transformants were grown to saturation in liquid LB media under kanamycin selection, and mutated plasmids were extracted using Presto™ Mini Plasmid Kit (Geneaid). Mutations were confirmed by DNA sequencing of the entire gene

with pDONR201-F and pDONR201-R primers (DNA Sequencing Facility, RRI). Subsequent LR recombination into pDEST-527 expression vector was completed as described in section **2.4.2.1** and mutant purification as described in section **2.4.4.1.**

2.4.12 Electrophoretic Mobility Shift Assay (EMSA)

SNM1A$^{D736A/H737A}$ was serially diluted from 25 µM to indicated concentrations using Storage Buffer and incubated with 50 nM DNA in EMSA Buffer (50 mM Tris-acetate pH 7.2, 75 mM potassium acetate, 1 mM DTT and 100 µg/mL BSA) for 15 minutes at room temperature. EMSA Loading Buffer (10 mM EDTA, 50% v/v glycerol) was added to the complexes to facilitate analysis with 10% (**Figure 2.10**) or 20% (**Figure 2.11**, **Figure 2.14**) native PAGE. Quantification of '% DNA Bound' was determined using integration of remaining substrate band pixel intensity in ImageLab, normalized to the DNA only control, according to:

Equation 3

$$\%\ DNA\ Bound = 1 - \frac{Remaining\ Substrate\ Intensity}{DNA\ Alone\ Intensity} \times 100\%$$

SNM1A substrate affinity (K_D) was determined with variable slope, non-linear regression in GraphPad PRISM v8.4.3. Standard error of the mean was calculated from three independent replicates.

2.5 Results

2.5.1 Development of a robust SNM1A purification scheme

To extend characterization of SNM1A function, a robust and reliable purification scheme was necessary. To that end, the nuclease domain of SNM1A (698-1040; further referred to as SNM1A$^{\Delta N697}$ or SNM1A) was codon optimized for expression in *E. coli* and cloned into an IPTG-inducible expression vector containing an N-terminal His_6 tag. This construct exhibited sufficient expression but poor solubility (~40 µg protein recovered/ L culture, 24% purity; **Figure 2.1**). Significant optimization was therefore required to improve the purification procedure. Co-elution of His-SNM1A with GroEL chaperone complex (confirmed by mass spectroscopy; data not shown) was observed consistently during purification. This interaction suggested that *E. coli* chaperones may mediate recombinant SNM1A folding. Overexpression of chaperones, induced through activation of cell stress response, has been shown to increase recovery of many marginally soluble proteins[214,215]. Activating stress response through the addition of ethanol (5% v/v) prior to inducing expression with IPTG improved SNM1A solubility and increased recovery nearly 10-fold (**Figure 2.2**).

Thermal shift, or differential scanning fluorimetry (DSF), analysis of semi-pure SNM1A was used to optimize buffering component and pH/ ionic strength of lysis and purification solutions. SNM1A demonstrated low thermal stability in initial Tris-based lysis and purification buffers (T_m = 39 °C). Stability improved below pH 7, though the effect was buffer specific. While citrate buffers (pH 5.5 and 6.0) prompted the largest increase in T_m, sodium phosphate dramatically improved stability under more neutral conditions (pH 6.8; **Figure 2.3A**). Trends between different ionic strength conditions were less clear. SNM1A was most stable in monovalent and ammonium-based salts, however increasing concentrations did not proportionally improve T_m. Because large increases of NaCl resulted in little change in stability, the initial 500 mM NaCl condition was maintained (**Figure 2.3B**). Collectively, ethanol stress-induction, optimized buffering conditions and finally, supplementing growth media with terrific broth, synergistically

improved the yield of recovered SNM1A (5.4mg/ L culture, 93% purity; **Figure 2.4**). Using this procedure to generate His-SNM1A, purity was increased by subsequent removal of the His-tag with TEV protease followed by a second immobilized nickel affinity column and cation exchange (1.8mg/ L culture final, ~99% purity; **Figure 2.5**). Size-exclusion chromatography with multi-angle light scattering confirmed that SNM1A was monomeric and monodisperse, with a calculated molecular weight corresponding closely to the theoretical value ($MW_{Calculated}$=40.6±2.4 kDa, $MW_{Theoretical}$=38.9 kDa; **Figure 2.6**).

Previous characterizations of SNM1A completed in other labs relied on SNM1A generated with the eukaryotic expression system, *Spodoptera frugiperda*[177,178]. Although structural studies have not indicated any post-translational processing that would prohibit expression using *E. coli*, we wanted to confirm that SNM1A samples generated by each system were directly comparable. To accomplish this, baculovirus was modified to express an equivalent construct of SNM1A (698-1040) and was used to infect High Five insect cells. Following expression, SNM1A-containing insect lysate was purified using an established protocol (**Figure 2.7**)[178]. SNM1A samples purified from either insect cells or *E. coli* were identical when analyzed by size-exclusion chromatography (**Figure 2.8A**). Critically, both preparations exhibited similar activity in protein dose-dependent nuclease assays using single and double-stranded DNA substrates designed to monitor exonuclease (5'P-3'F ssDNA: **P.DSF-T** or **P.20TF,** and 5'P-3'F dsDNA: **P.DS**) and endonuclease (5'OH-3'F ssDNA: **OH.20TF** and 5'F ssDNA: **F.20T**) activity (**Figure 2.8B-C**, detailed sequence and structure information of all DNA substrates presented in this thesis are included in the **Supplemental Information**). These findings establish an efficient procedure for expression and purification of SNM1A from *E. coli* with improved yield and purity while maintaining directly comparable nuclease activities.

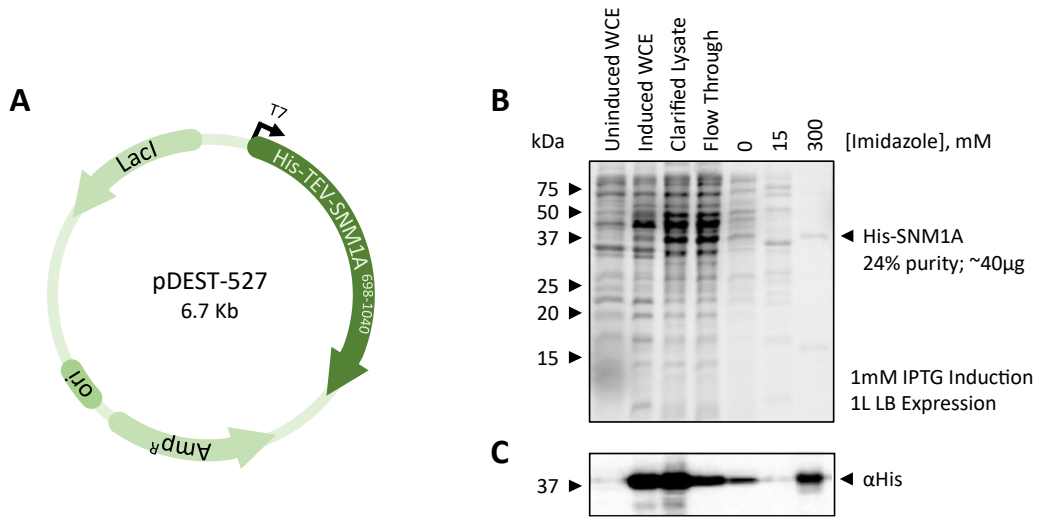


Figure 2.1. Standard His-SNM1A$^{\Delta N697}$ expression generates limited protein in *E. coli*.

[A] Schematic of pDEST-527 expression vector used to generate TEV-cleavable His_6-tagged SNM1A$^{\Delta N697}$. **[B]** SDS-PAGE analysis of SNM1A purification. Expression was induced with 1 mM IPTG in transformed Rosetta pLysS pRARE cells for 16 hours at 25 °C. Clarified supernatant was purified by nickel-affinity chromatography. Purification samples were analyzed by 15% SDS-PAGE and visualized using 0.5% TCE stain-free imaging by GelDoc EZ Imager (515 nm). Purity was determined using band integration in ImageLab, as a ratio with total lane intensity. **[C]** Western blot to confirm SNM1A expression and elution. Western blot was incubated with mouse monoclonal anti-polyhistidine (1:2000, H1029 Sigma) primary antibodies, followed by secondary incubation with rabbit anti-mouse IgG-HRP (1:4000, A9044 Sigma). His-SNM1A was visualized after exposure with ECL-substrate using a ChemiDoc MP Imager.

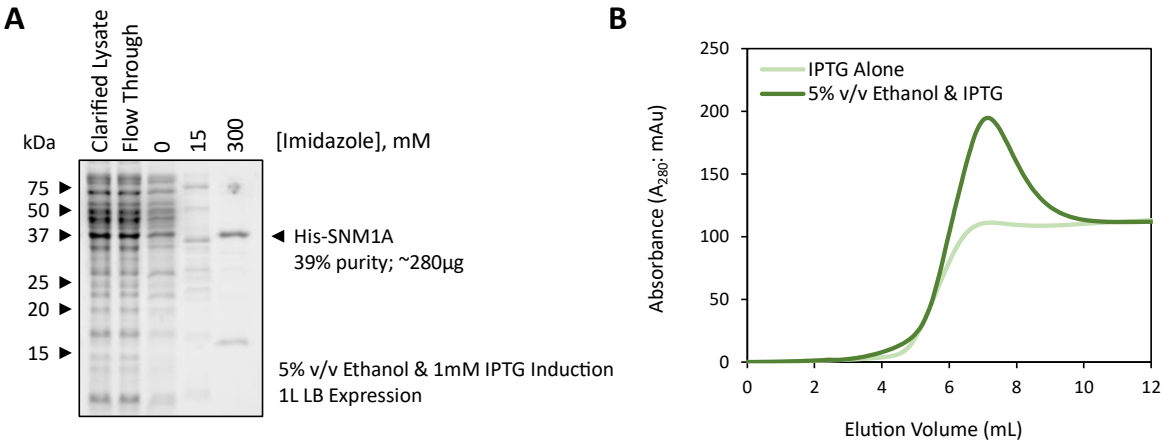


Figure 2.2. Ethanol stress induction of SNM1A$^{\Delta N697}$ improves purification yield.

Rosetta DE3 pLysS pRARE were transformed with the expression construct from **Figure 2.1A**. Expression was induced with 1 mM IPTG at 25 °C for 16 hours following a 10 min incubation with (dark green) or without (light green) 5% v/v ethanol. Lysed and clarified cells were purified using nickel-affinity chromatography and eluted with 300 mM imidazole. **[A]** Purification samples from stress induction were analyzed by 15% SDS-PAGE and visualized using 0.5% v/v TCE stain free imaging by GelDoc EZ Imager. **[B]** Absorbance trace (A_{280}) of SNM1A elution following nickel affinity chromatography. Absorbance of eluted protein was monitored using NGC chromatography system.

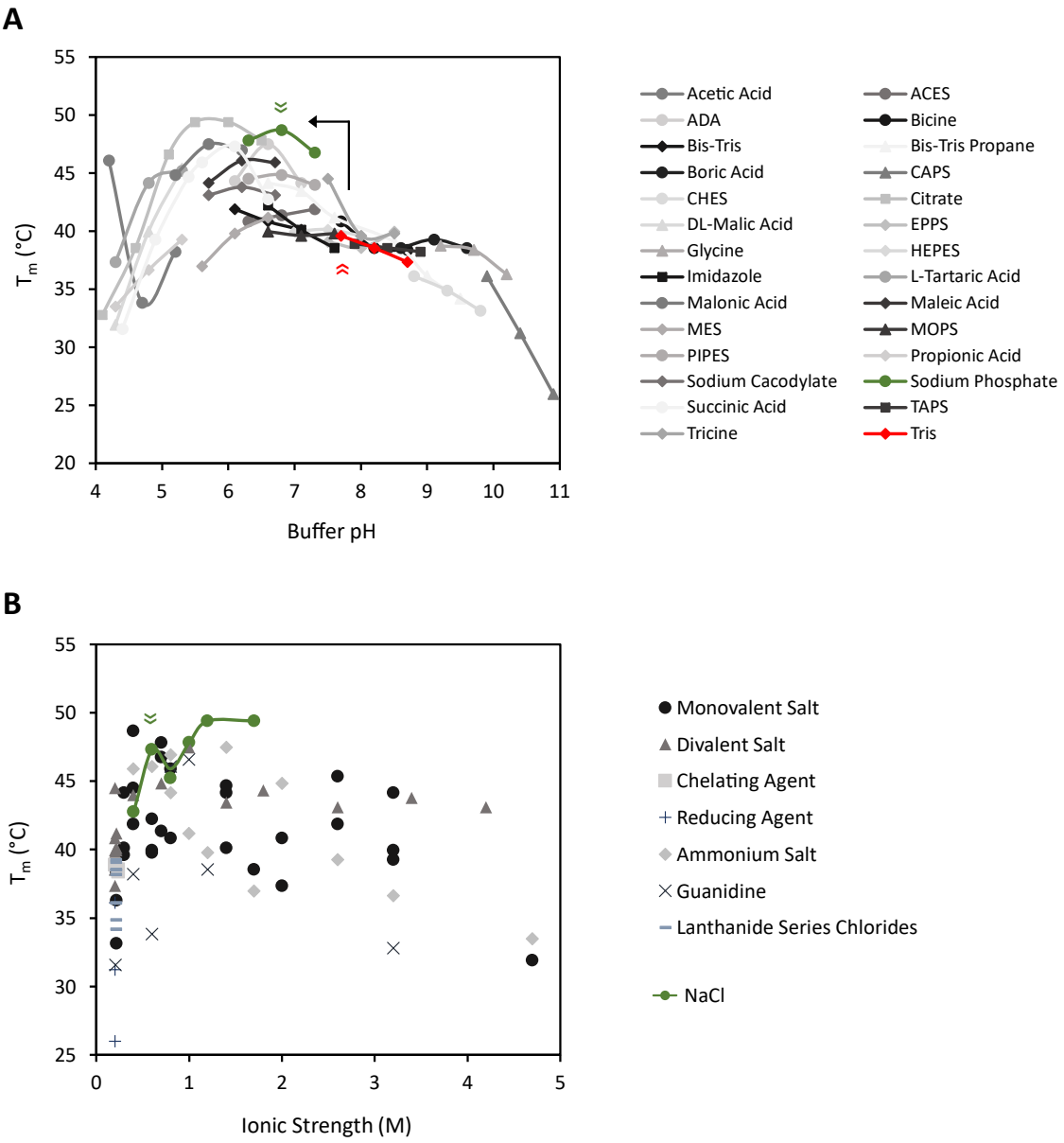


Figure 2.3. Thermal shift analysis of purified SNM1A identified optimal conditions for purification.

SNM1A$^{\Delta N697}$ (25 µM) was mixed 1:1 with the Durham **[A]** pH or **[B]** salt screen. Fluorescence in each condition from Sypro Orange dye (595 nm) was monitored across 20-90 °C using QuantStudio3. Melt curves and T_m were generated through data analysis with Protein Thermal Shift Software v1.3. Initial buffer (Tris) is indicated in red, optimized buffer (Sodium Phosphate) is indicated in green. » Indicates conditions used in purification.

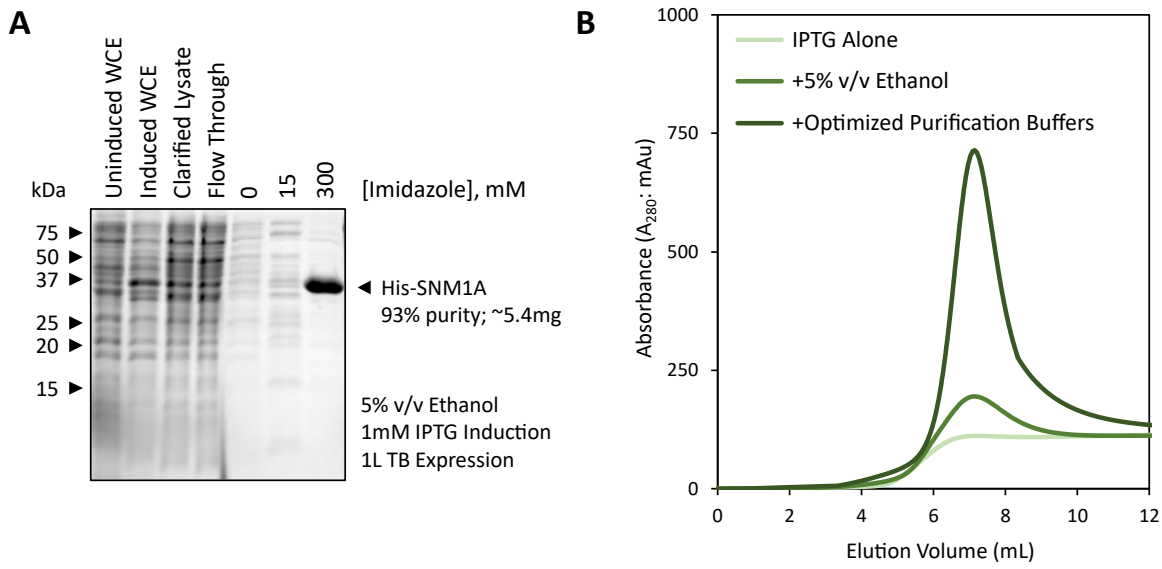


Figure 2.4. Optimizing buffer pH and ionic strength in purification buffers improves solubility of SNM1A$^{\Delta N697}$.

[A] Immobilized nickel affinity purification using optimized buffers from **Figure 2.3**. Cells were grown in terrific broth and expressed using the ethanol stress induction protocol established in **Figure 2.2**. Purification samples were analyzed by 15% SDS-PAGE with 0.5% v/v TCE using stain-free GelDoc EZ Imager. **[B]** Absorbance trace (A_{280}) of SNM1A elution following nickel affinity chromatography. Stress induction expression chromatogram traces from **Figure 2.2** are included.

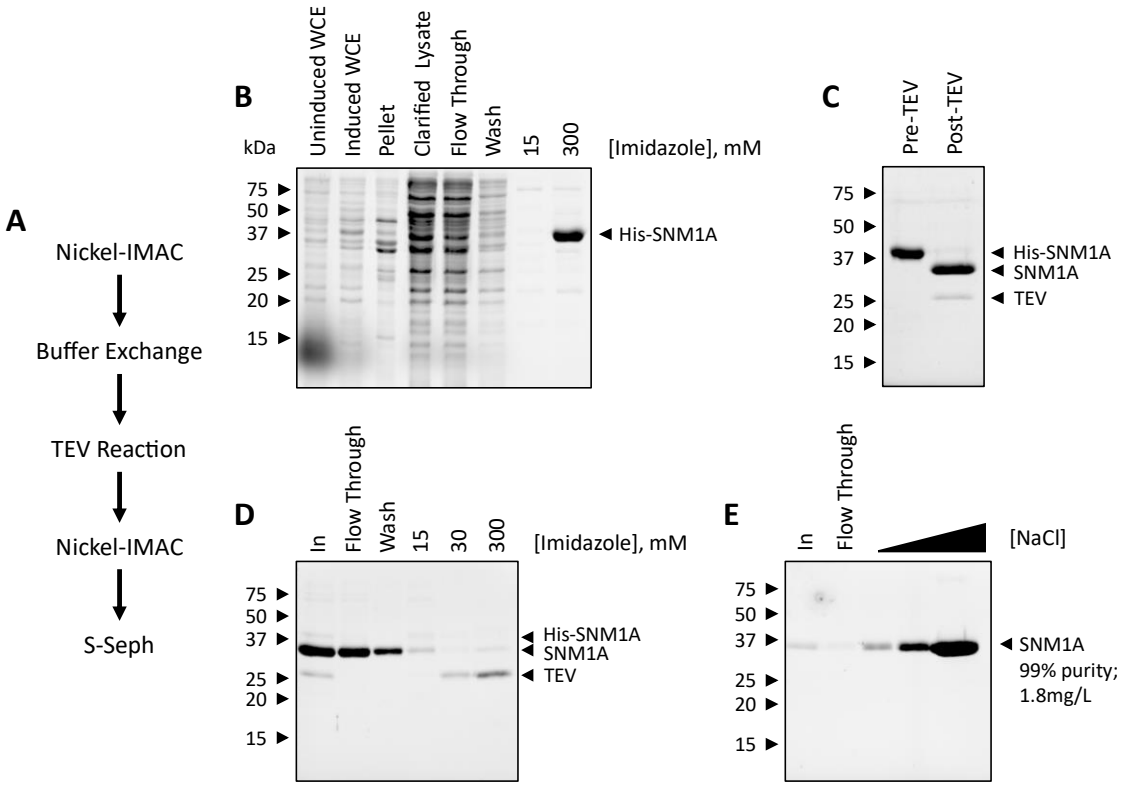


Figure 2.5. Complete optimized purification scheme of SNM1A$^{\Delta N697}$.

[A] Schematic of complete SNM1A purification. Purification fractions from **[B]** Nickel-IMAC, **[C]** buffer exchange and TEV reaction, **[D]** Nickel-IMAC, and **[E]** cation exchange were analyzed on 15% SDS-PAGE and visualized with 0.5% v/v TCE using stain-free GelDoc EZ Imager. Pure SNM1A was exchanged into storage buffer, concentrated to 25 µM and flash frozen.

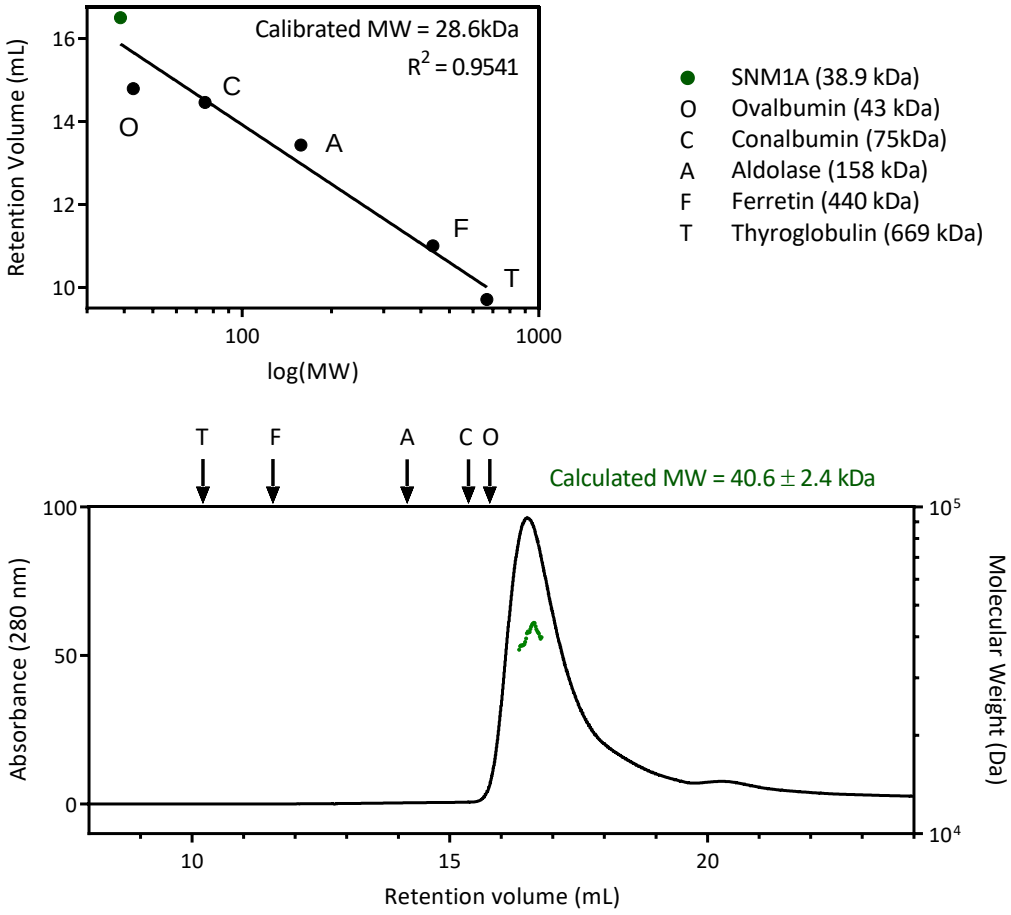


Figure 2.6. Purified SNM1A$^{\Delta N697}$ is monomeric and monodisperse in solution.

Analysis of SNM1A oligomeric state by calibrated SEC and MALS. A 1 mL sample was injected on a Superdex 200 10/300 GL calibrated with high-molecular weight standards. Calibrated molecular weight was calculated using linear regression of standard retention volumes. In-line MALS instrument and differential refractometer was used to confirm the molecular weight (green line, expected: 38.9 kDa, SEC estimate: 28.6 kDa, MALS estimate: 40.6 kDa).

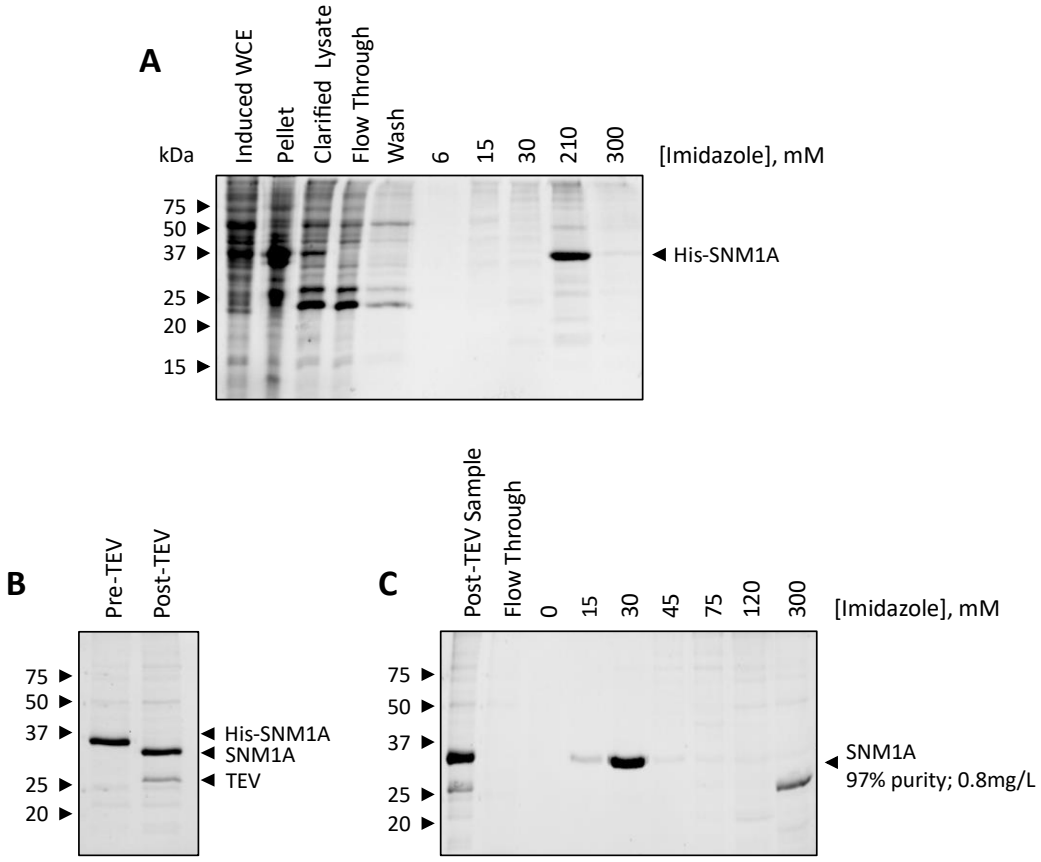


Figure 2.7. Purification of SNM1A$^{\Delta N697}$ from insect cells.

Baculovirus was generated through Tn7-dependent recombination between His-SNM1A$^{\Delta N697}$ pFastBac and DH10 Bacmid (Thermofisher). Modified SNM1A-Bacmid was transfected into Sf21 (Thermofisher) and p1 viral titres were collected. Suspended High Five (Thermofisher) insect cells were infected with baculovirus (1:1000 v/v). Cells were collected following 96-hour incubation at 27 °C and 125 rpm. **[A]** Collected cells were lysed by sonication and purified using Nickel-IMAC followed by **[B]** buffer exchange and TEV cleavage at 4 °C overnight. **[C]** Cleaved fractions were purified again using Nickel-IMAC. Pure fractions (15-30 mM imidazole elutions) were pooled, exchanged into storage buffer, concentrated to 25 µM and flash frozen. All samples were analyzed on 15% SDS-PAGE and visualized with 0.5% v/v TCE using stain-free GelDoc EZ Imager.

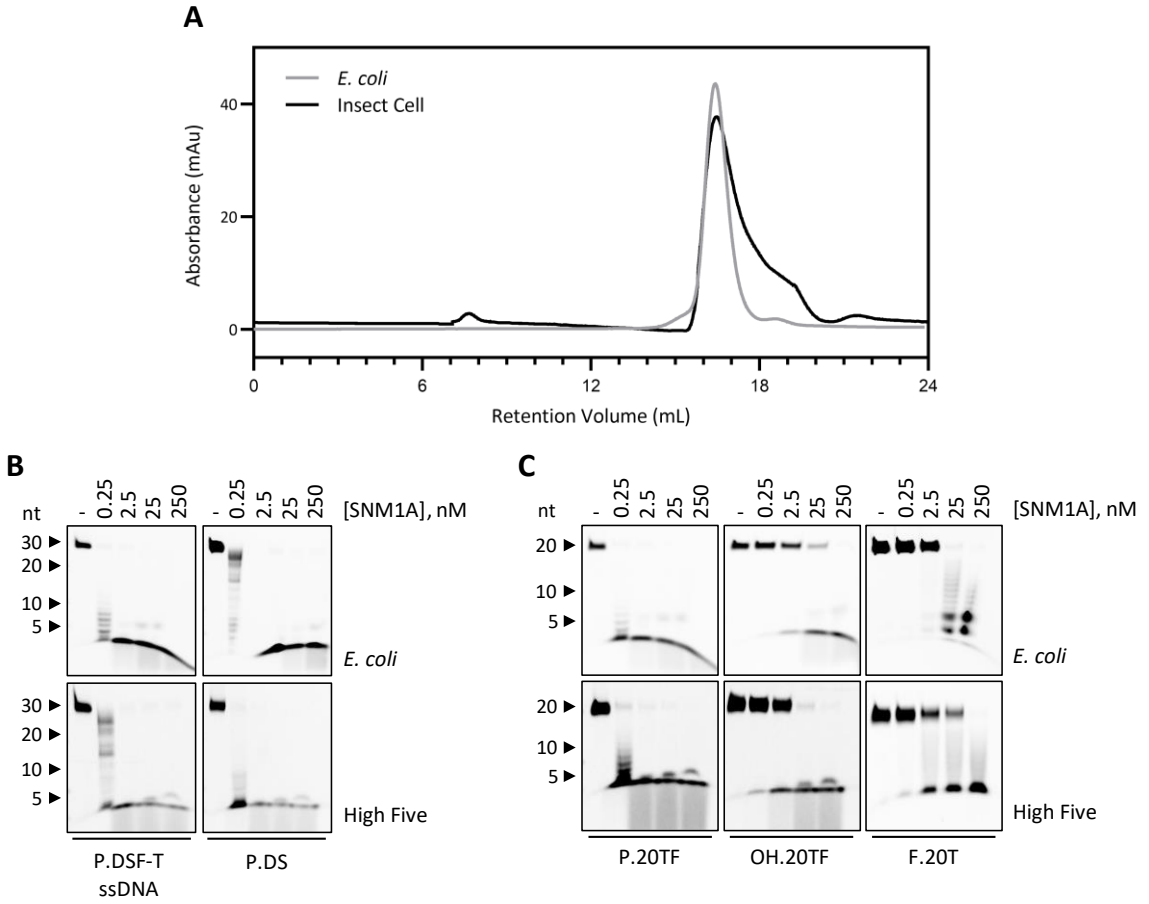


Figure 2.8. Comparison of purified SNM1A$^{\Delta N697}$ expressed in *E. coli* or High Five insect cells.

[A] Absorbance trace of size exclusion (Superdex 200 10/300) analysis of *E. coli* (grey) and insect cell (black) expressed SNM1A. Stock concentrations from both preparations were standardized to 25 µM by Bradford assay. **[B]** Protein-dose dependent assay comparing activity using **P.DS** (single- and double-stranded). **[C]** Protein dose-dependent assay comparing activity using **P.20TF**, **OH.20TF** and **F.20T**. Reactions with SNM1A (0.25-250 nM) were initiated with 100 nM of indicated DNA substrates and incubated at 37 °C for 30 minutes. Products were analyzed using 22.5% denaturing PAGE and visualized with the ChemiDoc MP Imager.

2.5.2 *In vitro* model of SNM1A nuclease capabilities in interstrand crosslink repair

Though different models have been proposed in the literature, the exact function of SNM1A in repair has not been determined. Until our publication establishing SNM1A as a single-strand dependent endonuclease[179], models were limited to SNM1A exonuclease processing[120,121,129]. To investigate SNM1A dual nuclease activities on potential ICL intermediates, a series of DNA substrates mimicking a stalled replication fork were designed. Simple fork structures (**ICL.Fork**), with or without a nascent leading strand (**ICL-F**), were damaged with an SJG-136 site-specific ICL (**Figure 2.9A**). SJG-136 crosslinks DNA at 5'-GATC sequences to form a heat reversible 1,4 ICL[129,216]. Following damage, substrates were analyzed by denaturing PAGE to ensure efficient crosslinking (>95% crosslinked, **Figure 2.9B**). SNM1A exonuclease activity requires a terminal 5' phosphate[177]. Forks containing a 5'P were expected to be engaged by SNM1A as an exonuclease (5'P; **P.ICLF-T**). Forks lacking a phosphate (5'OH; **OH.ICLF-T**) or blocked with a 5' fluorophore (5'F; **F.ICL-T**) require initiation by SNM1A endonuclease processing[179].

SNM1A nuclease activity was observed on all substrates, irrespective of the nascent leading strand (**Figure 2.9C**). With an available 5' phosphate, SNM1A rapidly degraded the 5' flap DNA up to the crosslink. Accumulation at the crosslink was evident after as little as 2 minutes (lanes 5 and 11, narrow dashes). Following a 60-minute incubation, undamaged substrates were almost completely degraded (lanes 3 and 9), while crosslinked substrates remained largely blocked at the ICL (lanes 6 and 12). SNM1A-dependent translesional processing past the ICL was observed, though significantly less robust than anticipated (lanes 6 and 12, wide dashes). In the absence of a 5' phosphate (5'OH or 5'F), SNM1A processed the single-strand flap as an endonuclease, with a slight preference toward the ss-dsDNA junction. As SNM1A exonuclease activity proceeds 5-3', products generated on the 5' labelled substrate (5'F) exclusively capture the initial nuclease event. In contrast, substrates with 5' hydroxyl and 3' labels (5'OH) enabled visualization of secondary SNM1A exonuclease degradation. Though substrates requiring endonuclease initiation demonstrated a clear reduction in overall activity, identical intermediates accumulated at the ICL. Taken

together, in the presence of free single-strand DNA (template lagging strand), SNM1A can initiate and completely unhook an ICL, *in vitro*.

Significant stalling of SNM1A exonuclease activity at the ss-dsDNA junction was only observed in the presence of the nascent leading strand (**Figure 2.9C**; 5'P; compare lanes 5 and 11). This unanticipated stalling suggests SNM1A has a reduced turnover rate, specifically on nicked DNA. To quantify this preference, a series of duplex DNA substrates with an internal phosphate-containing nick (**P.Nicked**) or gap (**1nt P.Gap** or **2nt P.Gap**) was generated. Time-course nuclease assays demonstrated that SNM1A exonuclease processing of a nicked DNA substrate is approximately 20-fold slower compared to the equivalent **3' overhang** or **ssDNA** substrate. To a lesser extent, the same slowing was observed on a duplex substrate with a 1nt gap but was entirely absent with a 2nt gap (**Figure 2.10A-B**). Similarly, SNM1A bound the nicked substrate with 5-fold reduced affinity compared to the equivalent 3' overhang (**Figure 2.10C-D**). To prevent digestion of DNA substrates, a catalytically deficient SNM1A mutant (D736A/H737A) was used for affinity quantification by EMSA.

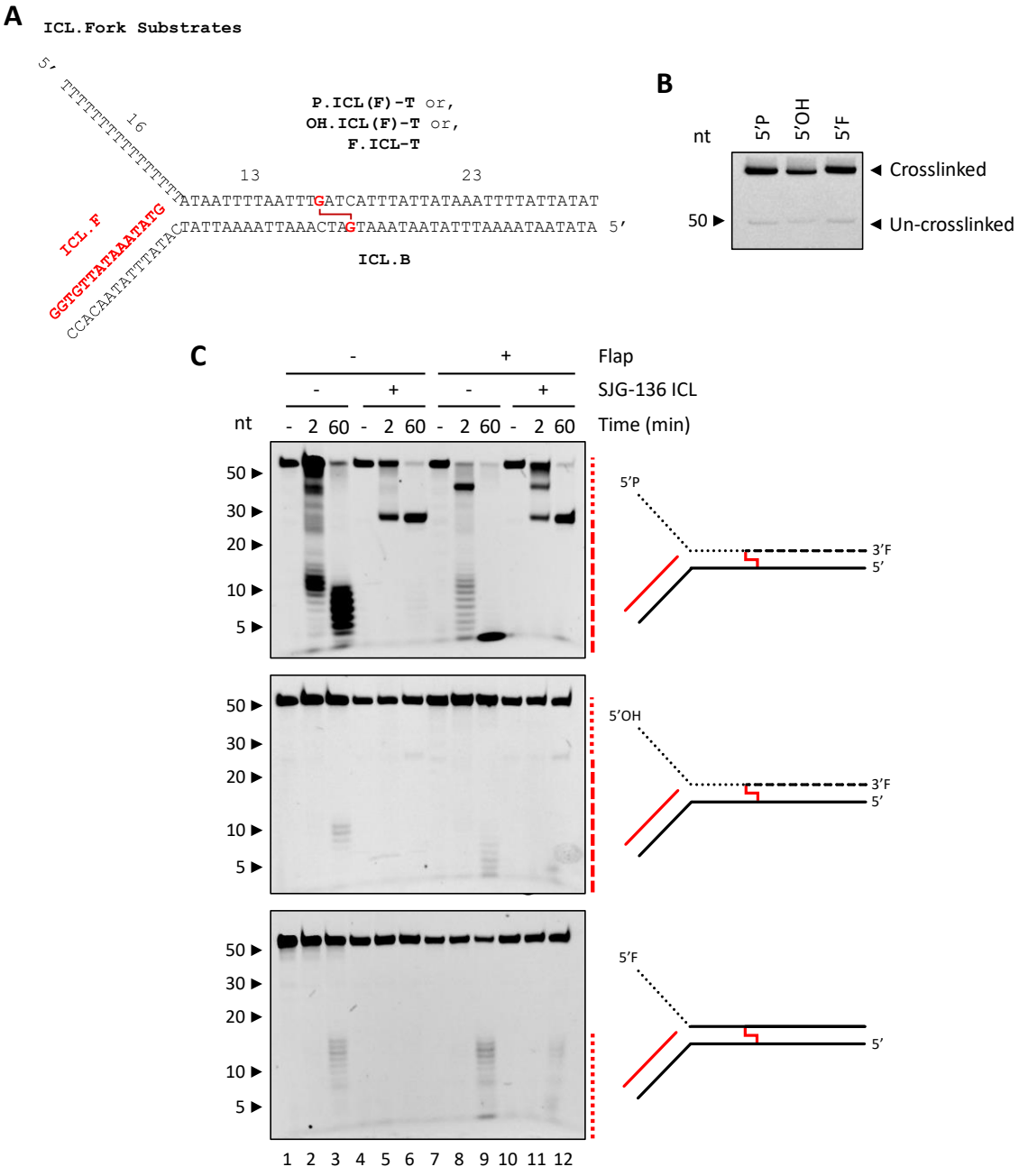


Figure 2.9. SNM1A can directly initiate processing to unhook ICL damage.

[A] Schematic of ICL repair intermediate substrates. **[B]** Purity of crosslinked DNA substrates analyzed by 22.5% denaturing PAGE. **[C]** Nuclease assays of SNM1A (40 nM) incubated with ICL intermediate DNA substrates (100 nM) labelled with top strands: 5'P (**P.ICL-T)**, 5'OH (**OH.ICLF-T)** or 5'F (**F.ICLF-T)**, with or without nascent leading strand (**ICL-F)** at 37 °C for 2 or 60 minutes. Products were boiled for 5 minutes and analyzed with 22.5% denaturing PAGE. All gels were visualized by ChemiDoc MP Imager. Dashes indicate digested regions of each substrate.

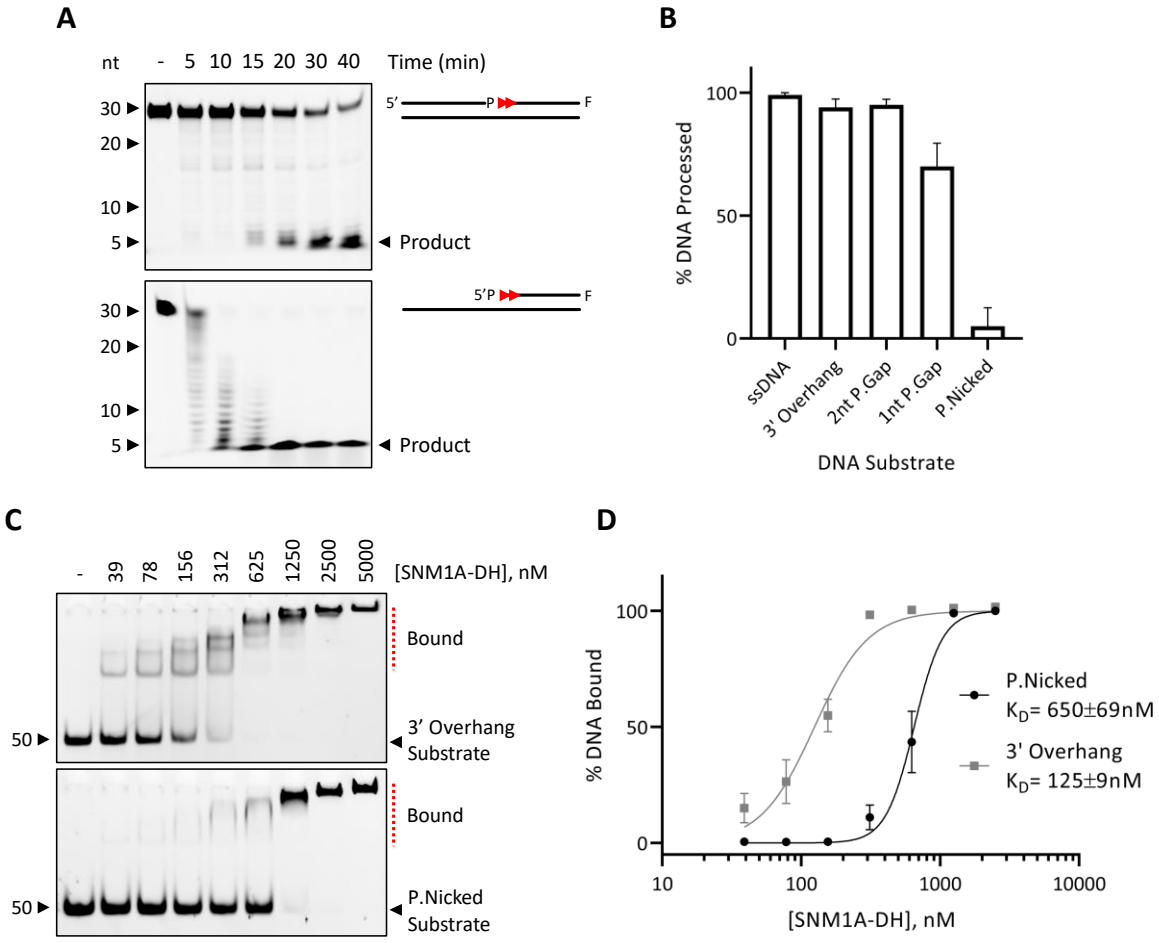


Figure 2.10. SNM1A activity and affinity are significantly reduced on bulky exonuclease substrates.

[A] SNM1A (5 nM) was incubated with 100 nM DNA (**P.Nicked** or **3' Overhang** duplex substrates) for 0-40 minutes at 37 °C. Products were analyzed using 22.5% denaturing PAGE and visualized by ChemiDoc MP Imager. Red arrows indicate expected initial nuclease events. P indicates phosphorylation, F indicates fluorophore placement. **[B]** Quantification of reaction progress after 20 minutes. Reaction progress was determined by integration of remaining substrate normalized to no protein controls using ImageLab. **[C]** Electrophoretic mobility shift assay using SNM1A$^{D736A/H737A}$ (39 nM – 5 µM) with 50 nM of either **3' Overhang** or **P.Nicked** DNA. Complexes were formed over 15 minutes at room temperature, analyzed using 10% native PAGE and visualized by ChemiDoc MP Imager. **[D]** Quantification of EMSA results. Integration of remaining substrate was used to determine % DNA bound. Four-variable sigmoidal curve fitting in GraphPad PRISM v8.4.3 was used to calculate K_D for each substrate. Error bars represent standard error of the mean from three independent replicates.

2.5.3 Characterization of SNM1A substrate preferences and implications in repair

Work in our lab identified SNM1A endonuclease activity on single-strand DNA substrates[179], in direct contrast to earlier reports suggesting SNM1A was an exclusive 5-3' exonuclease[175–177]. This broadening of SNM1A substrate specificity increases the potential intermediates SNM1A may process in DNA repair or other cellular functions. Therefore, better defining substrate preferences and limitations will inform which repair intermediate(s) SNM1A endonuclease activity may act on. Though there was no strict sequence specificity, SNM1A endonuclease activity was more robust on poly-pyrimidine sequences. Further, we showed that endonuclease processing was active on all substrates containing regions of significant single-strand DNA, regardless of local DNA secondary structure (3 or 5' overhangs, forks, flaps, heterologous loops and gaps)[179].

As SNM1A exonuclease activity strictly requires a free 5' phosphate, blocking the 5' end with a 6-FAM fluorophore disrupts initial exonuclease processing. Labelling of the 5' end also allows visualization of only endonuclease events, as subsequent 5-3' exonuclease digestion occurs exclusively on the unlabelled reaction products. To explore the length of single-strand DNA required for endonuclease processing, SNM1A was incubated with different lengths (6-20 nucleotides) of 5' fluorophore labelled polyT DNA. Under conditions tested, SNM1A was unable to digest DNA containing less than approximately 4 nucleotides (**Figure 2.11A**). Endonuclease processing was dependent on strand length. Where <5% of 6 polyT substrates was digested by SNM1A in 30 minutes, ~100% of 20 polyT DNA was processed under identical conditions. Further, SNM1A binding affinity mirrored the endonuclease preference for longer DNA substrates. SNM1A was unable to bind substrates 6 nucleotides or shorter while affinity increased proportionally with substrate length up to 20 nucleotides (**Figure 2.11B-C**). Size exclusion chromatography monitoring absorption at 255 nm further demonstrated that SNM1A endonuclease binding to longer substrates was stable in solution at high concentration of both protein (25μM) and DNA (2.5 μM; **Figure 2.12**). Collectively, these results identify a minimum binding footprint of 4-6 nucleotides, limiting potential

functions of SNM1A endonuclease activity to intermediates with more than 4 exposed nucleotides.

Long stretches of single-strand DNA do not often exist in a cellular context. *In vivo*, sources of extended ssDNA can arise as hairpin and gapped DNA intermediates. We have previously established that SNM1A endonuclease activity can digest hairpins and gaps (with 10 polyT apex or gap)[179]. Defining limitations of SNM1A endonuclease activity on these substrates could further refine potential intermediates available for SNM1A processing. Activity on all hairpin and gapped endonuclease substrates was quite inefficient, in some cases requiring low micromolar concentrations of SNM1A to detect digestion. SNM1A endonuclease activity increased with the size of hairpin apex, demonstrating very little processing on hairpins containing a 0 or 3T apex (**Figure 2.13A-B**). Similarly, SNM1A required at least a 3T gap to digest the gapped DNA substrates, becoming more efficient on a larger stretch of single-strand DNA (**Figure 2.13C-D**). Notably, a strong 3' cleavage preference was observed on all substrates; however, the bias was weakened with increased apex/ gap size, where SNM1A endonuclease activity was more robust.

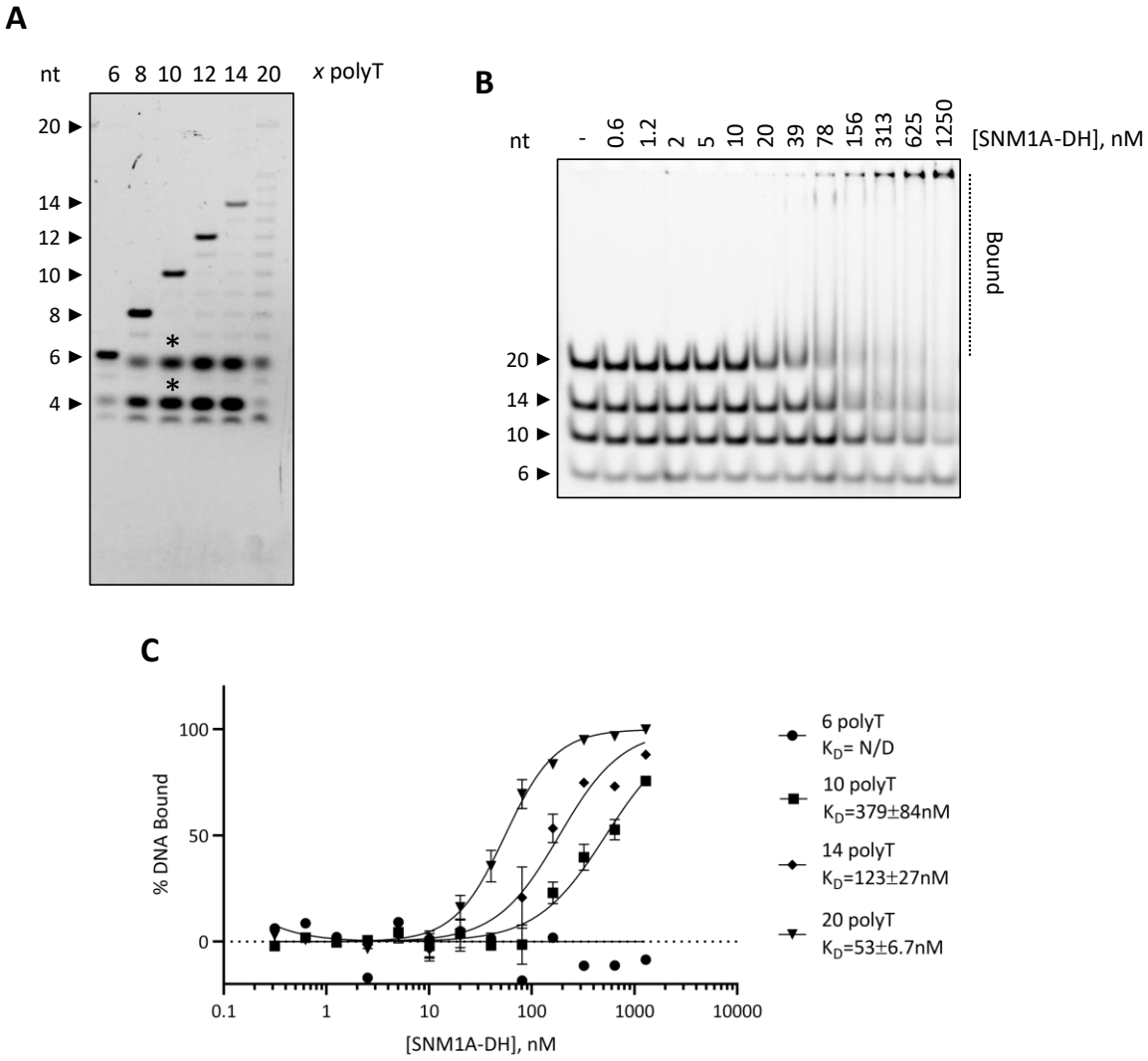


Figure 2.11. Affinity and activity of SNM1A endonuclease function are substrate length dependent.

[A] Nuclease assay of SNM1A (100 nM) incubated with 5' labelled DNA substrates (**F.6T**, **F.8T**, **F.10T**, **F.12T**, **F.14T**, **F.20T**; 100 nM) for 30 minutes at 37 °C. Products were analyzed using 22.5% denaturing PAGE **[B]** Combined electrophoretic mobility shift assay of SNM1A$^{D736A/H737A}$ (1250-0.6 nM) with 5' labelled DNA substrates (**F.20T**, **F.14T**, **F.10T** or **F.6T**; 50 nM). Protein and DNA were incubated at room temperature for 15 minutes before complexes were analyzed using 20% native PAGE. **[C]** EMSA quantification on indicated DNA lengths. Complex formation was measured using integration of remaining substrate in ImageLab and K_D was calculated using 4-variable sigmoidal curve fitting in GraphPad PRISM v8.4.3. Error bars represent standard error of the mean from three independent replicates. All gels were visualized with ChemiDoc MP imager. *denotes artifacts of aberrant 6FAM processing (**Appendix 2.C**).

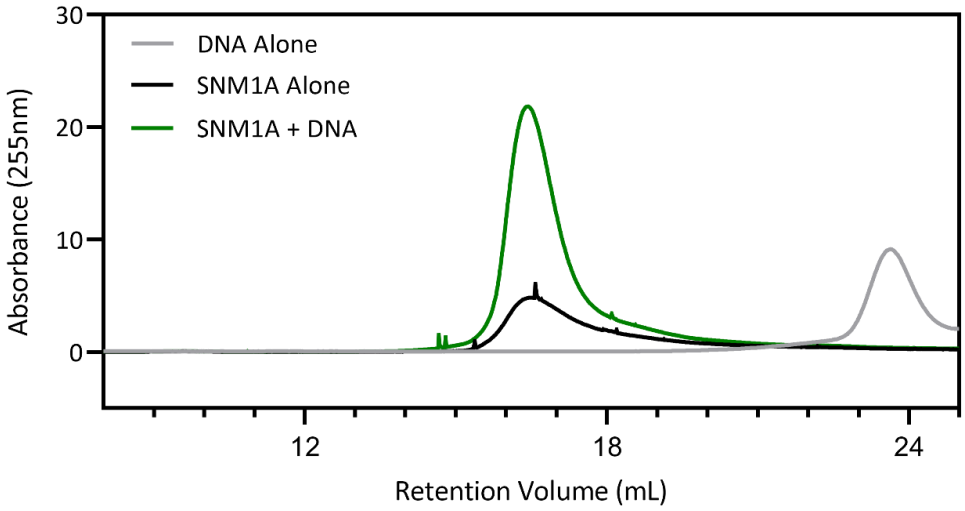


Figure 2.12. SNM1A interaction with an endonuclease DNA substrate is stable at high concentrations in solution.

SNM1A$^{D736A/H737A}$ (25 µM) was incubated with **F.14T** DNA substrate (2.5 µM) for 15 minutes at room temperature. SNM1A$^{D736A/H737A}$, DNA alone or SNM1A$^{D736A/H737A}$ and DNA were analyzed using a Superdex 200 10/300 (GE). Absorbance was measured at 255 nm and plotted using GraphPad PRISM.

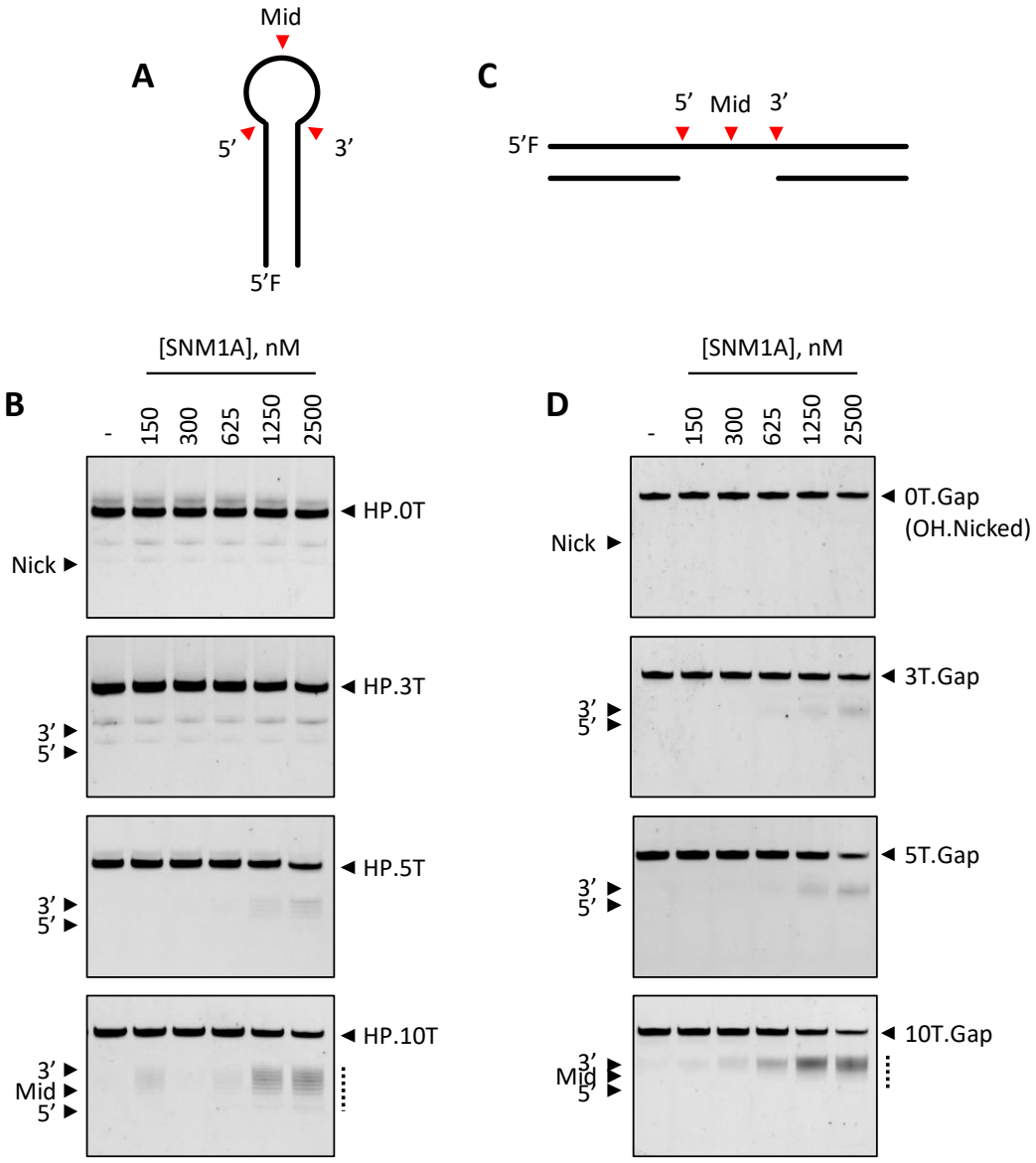


Figure 2.13. SNM1A has limited endonuclease activity on substrates with small regions of available single strand DNA.

[A] Schematic of hairpin DNA substrates. 5′ labelled, apex composed of varied length of polyT. Boundaries of expected nuclease products are indicated with red arrows. **[B]** SNM1A nuclease assays on hairpin DNA. SNM1A (150 nM – 2.5 µM) was incubated with hairpin DNA substrates (100 nM) at 37 °C for 60 minutes. **[C]** Schematic of gapped DNA substrate. 5′ labelled duplex DNA separated by varying lengths of single-strand polyT. **[D]** SNM1A nuclease assays on gapped DNA. SNM1A (150 nM – 2.5 µM) was incubated with gapped DNA substrates (100 nM) at 37 °C for 30 minutes. All products were analyzed on 22.5% denaturing PAGE and visualized by ChemiDoc MP Imager.

2.5.4 SNM1A dual nuclease activities rely on distinct metal-dependence

Though SNM1A is a dual-function nuclease, the exonuclease catalytic efficiency significantly exceeds that of the endonuclease activity[179]. It is unclear whether this differential is a result of disparities between binding affinity or turn-over rate. Electrophoretic mobility shift assays were used to measure binding affinity of SNM1A with equivalent polyT DNA substrates, modified to promote engagement as either an endonuclease (5'F; **F.20T**), exonuclease (5'P; **P.20TF**) or both (5'OH; **OH.20TF**). No significant difference was observed in SNM1A substrate affinity regardless of whether it engaged as an exonuclease or endonuclease (all K_D within ~100-200 nM; **Figure 2.14**). Therefore, the difference between SNM1A endonuclease and exonuclease efficiency must not be driven primarily by differences in substrate binding affinity. Instead, these results suggest that differences in turn-over rate are responsible for the observed differential activity.

SNM1A catalysis is metal-dependent, though which metals (and how many) are required has not been clearly identified[177,178]. Recombinant SNM1A can digest DNA substrates in the absence of additional exogenous metals in the reaction buffer. Residually bound metals are likely derived artifactually from recombinant expression and purification and may have confounded previous mechanistic studies examining the role of metals in catalysis. Therefore, following purification, SNM1A was treated with chelators (50 mM EDTA and 5 mM o-phenanthroline) and exchanged into storage buffer. Treated SNM1A was analyzed by SEC to confirm no significant disruption in structural stability following metal chelation (data not shown). To determine whether differences in metal-dependence were responsible for the catalytic preference for exonuclease processing, changes in SNM1A nuclease activity were monitored following titration with different metals (**Figure 2.15**). Prior to initiating nuclease reactions with exonuclease (**P.20TF**, lanes 1-13) or endonuclease (**F.20T**, lanes 14-26) DNA substrates, SNM1A was incubated with individual metals (Mg^{2+}, Zn^{2+}, Mn^{2+}, Ca^{2+}, Cu^{2+}, or Ni^{2+}) at a range of concentrations (1 µM - 10 mM). As expected, untreated SNM1A maintained

significant residual activity on each DNA substrate in the absence of metals in the buffer, in contrast with de-metallated SNM1A (compare lanes 2 to 8 and 15 to 21).

SNM1A exonuclease activity was observed with all divalent metals tested, at relatively low concentrations (100 µM, lane 11). As previously reported, addition of Zn^{2+} (as well as Cu^{2+}) was inhibitory at millimolar concentrations (lanes 11-13)[175,177]. Mn^{2+} and Ni^{2+} were significantly stimulatory at 100 µM, while comparable Mg^{2+}-dependent activation required substantially higher concentrations (1-10 mM). In contrast, endonuclease activity was only stimulated at low concentrations by Zn^{2+} (100 µM, lane 24). While addition of Mn^{2+} or Ni^{2+} recovered activity at much higher concentrations (1-10 mM, lanes 25-26), Mg^{2+} and Cu^{2+} were completely incompatible with endonuclease processing under the conditions tested. Taken together, even though SNM1A exonuclease and endonuclease activity is facilitated by a single active site, these findings suggest differing catalytic mechanisms.

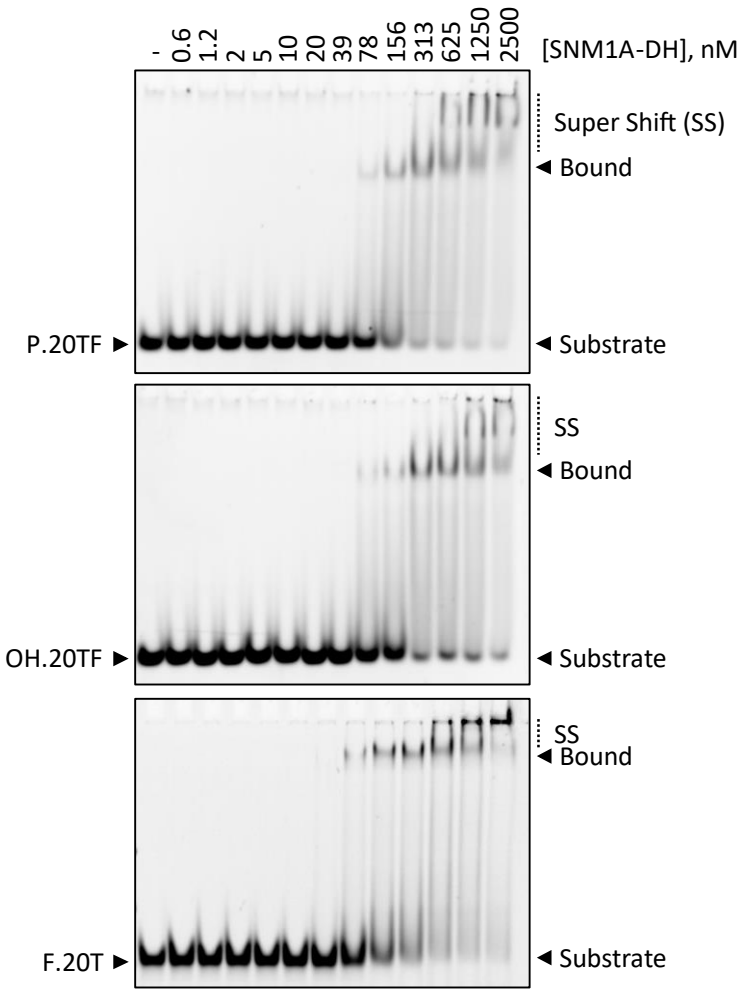


Figure 2.14. Modification of 5' DNA ends does not alter affinity of SNM1A interaction.

SNM1A$^{D736A/H737A}$ (0.6-2500 nM) was incubated with 20 polyT DNA substrates (100 nM) modified with a 5' phosphate, 5' hydroxyl or 5' fluorophore label for 15 minutes at room temperature. Complex formation was analyzed using 20% native PAGE and visualized using a ChemiDoc MP Imager. All gels are representative of three independent replicates.

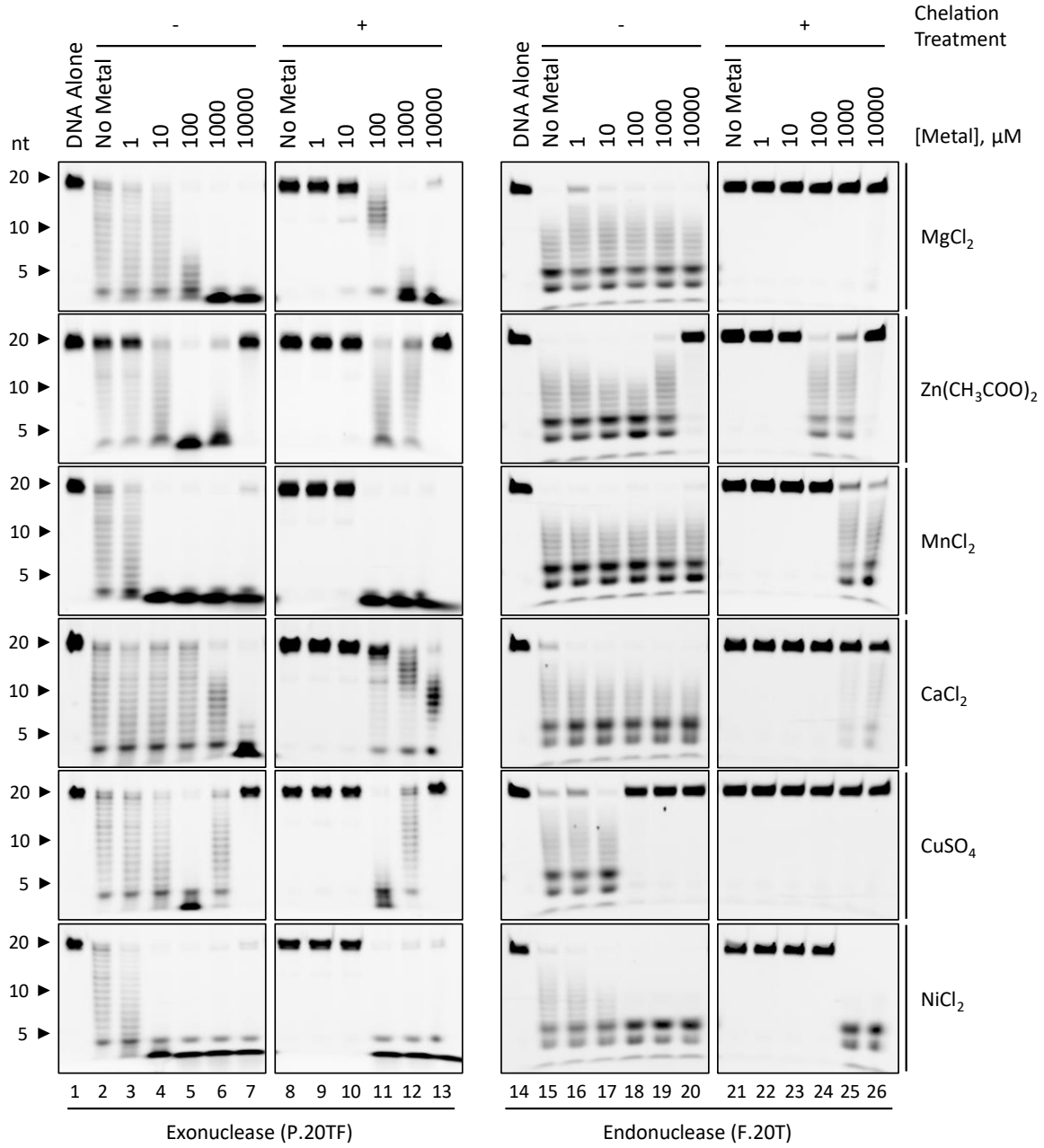


Figure 2.15. Metal-dependence is distinct for SNM1A endo- and exonuclease activity.

Purified SNM1A was incubated with or without chelators (50 mM EDTA, 5 mM o-phenanthroline) for 2 hours. Following treatment, both samples were buffer exchanged into storage buffer. Exonuclease (5'P, 3'F: **P20TF**) or endonuclease (5'F: **F20T**) 20 polyT DNA substrates (100 nM) were incubated with SNM1A (exonuclease; 5 nM or endonuclease; 100 nM) for 15 or 30 minutes, respectively. Each reaction was incubated with the indicated concentrations of metal. Products were analyzed on 22.5% denaturing PAGE and visualized by ChemiDoc MP imager. Gels are representative of three independent replicates.

2.6 Discussion

To elucidate the biological role(s) of SNM1A, careful *in vitro* characterization of substrate preferences is necessary. SNM1A possess an efficient 5' phosphate-dependent 5-3' exonuclease activity[175,176], which can bypass DNA crosslink damage[121,177,178]. SNM1A additionally catalyzes regions of single-strand DNA as an endonuclease[179]. Recent work from our lab expanded the repertoire of SNM1A catalytic activities, necessitating further analysis of how exonuclease, translesional and endonuclease activities compare and cooperate. Work presented in this chapter established a reliable and robust purification scheme necessary for *in vitro* characterization. Nuclease assays monitoring endo- and exo-nuclease activities, individually or collaboratively, further extend our understanding of SNM1A capacity.

2.6.1 Recombinant expression and purification of SNM1A

Characterization of full-length SNM1A has been limited due to its insolubility and instability in a variety of recombinant expression systems (*E. coli*, *S. cerevisiae*[176], *S. fugiperda*[175]). Overexpression in human and other mammalian cells is extremely toxic, further limiting the use of native expression systems[183,186,217]. Full-length SNM1A is predicted to be a largely disordered protein outside of the conserved nuclease domain (**Figure 1.7**)[195]. Consequently, structural and functional characterization of SNM1A has been focused almost exclusively on the core nuclease domain[177–179]. Prior functional work with SNM1A in our lab was facilitated by purification of truncated SNM1A, expressed with a protease-removable NusA fusion tag[179,218]. This construct was identified following expression and solubility testing of multiple truncations with different solubility fusions at the N- or C-terminus. Though this purification scheme was sufficient to generate limited quantities of pure SNM1A, it was relatively inefficient and inconsistent[219]. Significant optimization discussed in **Figures 2.1 – 2.6** addressed issues with expression, solubility, stability and reproducibility such that SNM1A$^{\Delta N697}$ could be generated absent a fusion tag using an *E. coli* expression system. Coincident with this

work, another lab reported structural and functional analysis of SNM1A exonuclease activity using an insect cell expression system[178]. To confirm that SNM1A generated from *E. coli* and insect cells were directly comparable, we purified each in parallel. Analysis by SEC-MALS confirmed identical monomeric behaviour in solution (**Figure 2.8A**). Critically, protein-dose dependent nuclease assays demonstrated that each SNM1A preparation exhibited similar exo- and endo-nuclease activities on single- and double-stranded DNA substrates (**Figure 2.8B-C**).

2.6.2 Limitations of SNM1A exonuclease activity

Previously published models require SNM1A to engage the freed 5'P product of an XPF incision within duplex DNA[120,121,129]. Here, we showed that while SNM1A can function as an exonuclease from an internal nick, its activity and affinity are significantly reduced relative to the equivalent ssDNA or 3' overhang substrate (**Figure 2.10B, D**). This preference disappears with even a small gap abutting the phosphate. This finding is contradictory to a prior report[145], which concluded SNM1A has no preference between blunt dsDNA, 3' overhang or nicked DNA containing an internal 5' phosphate. SNM1A engages exonuclease DNA substrates via interaction between the required 5' phosphate and a phosphate-binding pocket, directly adjacent to the active site (see **Chapter 4**). It is not unexpected then that bulkier nicked DNA substrates would be more difficult for SNM1A to bind. These findings suggest that SNM1A is ideally suited to initiate exonuclease processing from a small DNA gap (such as intermediates in MMR-dependent GG-ICL repair) or 5' flap (such as intermediates in TC-ICL repair), instead of an internal nick. Alternatively, SNM1A binding affinity for DNA increases as the nick is extended to an open overhang (**Figure 2.10D**), implying that SNM1A should behave processively as it extends from a nick. This processivity is apparent from the absence of intermediate products observed on nicked, but not overhang, DNA substrates in **Figure 2.10A**.

Models frequently focus on the relatively unique ability of SNM1A to bypass an ICL as a translesional exonuclease[120,121]. Despite this focus, systematic analysis of SNM1A translesional activity remains very limited. It is not known how/ whether variation in crosslinker chemistry, induced helical distortion, nucleotide off-set or polarity impacts the efficiency or ability of SNM1A to bypass an ICL. It seems likely that SNM1A may have differing preferences on 5' leading ICLs (such as nitrogen mustards, cisplatin or psoralen)[7] over 3' leading (such as MMC)[7] or blunt ICLs (such as a synthetic triazole)[121]. The number of nucleotides spanned by the crosslinker (nucleotide off-set) may further impact SNM1A translesional activity. To date, SNM1A translesional exonuclease activity has been directly observed across only an SJG-136 (1,4: 5' GATC)[129,177–179] and triazole (1,1: 5'A)[121] ICL. Previous reports by other groups showed robust translesional processing past SJG-136 crosslinked DNA. In contrast, SNM1A translesional activity on SJG-136 damaged substrates reported here is extremely limited (**Figure 2.9C**, see 5'P, lanes 6 and 12), barely exceeding the uncrosslinked substrate background (**Figure 2.9B**)[179]. The cause of this discrepancy is unclear, as the assays reported here have excess SNM1A (40 nM compared to ~1 nM).

Differences in substrate stability may impact the observed translesional activity. To improve instability of the SJG-136 crosslink we encountered *in vitro*, we embedded the crosslink 13 nucleotides into 40-mer duplex (**Figure 2.9A**). SNM1A exonuclease processing up to the ICL produces a stalled 3' overhang structure. The remaining downstream duplex DNA in our substrates remains stable under reaction conditions (T_m > 45 °C). Earlier reports, however, embedded SJG-136 in shorter substrates, where the remaining duplex DNA within stalled overhang intermediates was likely unstable (T_m < 37 °C, even considering the ICL acting as a 'hairpin'). SNM1A translesional nuclease activity appears inversely proportional to the stability of the DNA downstream of the SJG-136 crosslink[179]. The biological consequence of this on SNM1A translesional processing in repair is ambiguous. An ICL would be embedded in an extended genomic sequence during RIR, implying SNM1A translesional activity could be limited. XPF 5' incision of the reversed fork in the FA-pathway, in contrast, would leave an unstable

short stretch of annealed DNA, held together only by the ICL. Alternatively, recent work with Pso2 (yeast homolog) showed that translesion exonuclease activity was stimulated through functional interaction with RecQ4 helicases[81]. It is possible a yet-to-be-described interaction likewise stimulates SNM1A activity in humans. Collectively, this highlights the need for a more detailed characterization of SNM1A translesional activity on different ICLs, in differing DNA contexts.

2.6.3 Characterization of SNM1A endonuclease activity

Initial characterization of SNM1A endonuclease function revealed broad activity on single-strand DNA, with little specificity for DNA overhangs, forks, flaps, gaps and bubbles[179]. Each substrate contained 20 polyT single-strand stretches of DNA, allowing SNM1A to engage and degrade the substrates. Unlike many structure-specific endonucleases involved in DNA repair[124,145], SNM1A endonuclease processing demonstrated only a mild skew toward the ss-dsDNA junctions of each substrate. This finding suggests that the SNM1A nuclease domain does not specifically engage structured elements in these substrates; instead, it requires only a stretch of non-basepaired nucleotides. Here, we further refine the requirements of SNM1A endonuclease activity, showing that processing of a 5' labelled single-strand DNA substrate requires an absolute minimum of 4 nucleotides. The effect of increasing substrate length on endonuclease activity was striking, where a 6 polyT and 20 polyT were degraded <5% and ~100% under equivalent conditions, respectively. Changes in activity likely resulted from similar trends in substrate affinity, where SNM1A was unable to bind 6 polyT DNA under even micro-molar concentrations (**Figure 2.11**). Interestingly, SNM1A can degrade dinucleotides as an exonuclease (seen in products of **Figures 2.8, 2.9 and 2.10**). The necessary affinity for short exonuclease substrates is likely facilitated primarily by the 5' phosphate-binding pocket. As endonuclease substrates lack this interaction, increased substrate length may provide comparable affinity through extended non-specific interactions with the DNA backbone. Further

affinity increases between 10 polyT and 20 polyT suggest endonuclease substrate interaction could be more extensive than anticipated and/or reflect multiple binding events per substrate.

SNM1A (truncated to its nuclease domain) demonstrates more robust exonuclease processing. Our published estimates of single turn-over rates on each substrate suggest a single exonuclease catalytic event occurs approximately 1000-fold faster than a comparable endonuclease event[179]. The work presented in this thesis further shows ~100-fold difference in SNM1A concentration is required to equivalently digest endo- over exo-nuclease DNA substrates. While exact rate comparisons vary depending on reaction conditions and DNA structure/ sequences, SNM1A exonuclease function is universally more efficient *in vitro*. A biological role for SNM1A endonuclease activity would likely require stimulation, presumably by post-translational modification or modulation by a binding partner. These findings are reminiscent of SNM1A human homolog, Artemis. *In vitro*, Artemis alone possesses predominantly 5-3' exonuclease activity[220,221]. When activated through phosphorylation by DNA-PKcs, Artemis exhibits a robust endonuclease activity on DNA overhangs and hairpins[220,222,223]. This endonuclease activity dominates the biological role of Artemis in end-processing and V(D)J recombination[173]. SNM1A latent endonuclease activity demonstrated in **Figure 2.13** exhibited very similar 3' targeted processing of small apex hairpins and DNA gaps, with the caveat that activity required significantly higher concentrations of SNM1A (high nanomolar to low micromolar). Both SNM1A and Pso2 can function as hairpin-opening endonucleases in yeast, suggesting this activity may be more pronounced in a cellular context[179,180,224]. Though SNM1A is not redundant with Artemis and has no known role in NHEJ[189], their endonuclease activities may behave similarly in non-overlapping DNA repair pathways.

2.6.4 Differential metal-dependent mechanisms of SNM1A

Work presented in this chapter has demonstrated that the difference in catalytic rates of SNM1A exonuclease and endonuclease activity is driven primarily by differences in mechanism (**Figure 2.15**), not substrate binding affinity (**Figure 2.14**). SNM1A contains a single binuclear active site, conserved within the broader MβL superfamily[167,169]. The high-affinity metal site 1 (M1) is coordinated by four conserved residues (comprising MβL motifs II, III and IV) while the low-affinity metal site 2 (M2) is coordinated by three residues (motifs II and IV). Although X-ray structures of SNM1A identified only a single metal ion bound at the active site[178,225,226], homologs including Apollo[178,227], Artemis[174], CPSF-73[170] and others[132,228–230] display dual-zinc dependence. Further, substitution of conserved residues in SNM1A (D736 and/or H737) spatially oriented to interact with unoccupied M2 abolishes both exonuclease[177] and endonuclease activity[179]. Therefore, SNM1A was expected to facilitate nuclease reactions via a 2-metal catalytic mechanism (**Figure 2.16**).

Previous characterization of SNM1A metal-dependence is complicated by two factors: (i) the presence of artifactual metals resulting from recombinant purification and (ii) an exclusive focus on exonuclease processing. High-resolution crystal structures have consistently shown a single Ni ion bound in the high-affinity metal binding site of SNM1A (M1)[225,226]. Further, EDTA titrations revealed a contradictory impact on SNM1A function, where low concentrations stimulate and high concentrations inhibit exonuclease activity[177]. Both findings suggest residual metals, presumably resulting from purification, may have confounded previous metal titration experiments. To address this, SNM1A was treated with a combination of chelators prior to metal characterization. Optimization of chelation treatments demonstrated that high concentrations of multiple chelators (50 mM EDTA and 5 mM o-phenanthroline) were required to fully disrupt SNM1A nuclease activity, reflecting the high affinity of the M1 site.

The ability of all tested divalent metals to recover SNM1A exonuclease activity, at least partially, revealed surprising catalytic flexibility. In contrast, only Zn^{2+}, Mn^{2+} and Ni^{2+} were able to rescue endonuclease activity (**Figure 2.15**). Interestingly, both Mn^{2+} and Ni^{2+} dramatically stimulated both exo- and endonuclease functions. Similar non-specific stimulation by exogenous Mn^{2+} has been reported among many enzymes which rely on a 2-metal catalytic mechanism, commonly reflecting chemical flexibility inherent to manganese, as opposed to an implied biological function[231–234]. Broad stimulation by Ni^{2+} underscores the importance of chelation treatment prior to *in vitro* characterization. Residual Ni^{2+}, pulled from IMAC purification, may alter SNM1A substrate preferences and kinetic parameters *in vitro*.

Two-metal catalysis refers to a broad reaction mechanism conserved across species in enzymes ranging from polymerases, transposases and nucleases[234–236]. Each metal serves a distinct function within the binuclear active site, where M2 coordinates the activated nucleophile and M1 coordinates the scissile phosphate. Together, the metals position the incoming DNA backbone O^3-P bond for an in-line attack from the nucleophile to produce a 5' phosphate and 3' hydroxyl product (**Figure 2.16A**)[234,235,237]. While this mechanism is often supported by two Mg^{2+} ions[234,235], canonical MβL nucleases instead typically use Zn^{2+} [167]. The anticipated mechanism would be similar, as both Zn^{2+} and Mg^{2+} share a similar ionic radius and propensity for octahedral geometry, differing primarily in preferred coordinating ligands reflecting differences in relative acidity[169,238]. Metal titration results reported here showed SNM1A exonuclease activity was recovered by Zn^{2+} but dramatically stimulated by high concentrations of Mg^{2+}. This indicated that Mg^{2+} could fully substitute for Zn^{2+} in exonuclease processing, but not the reverse. That excess Mg^{2+} alone was both necessary and sufficient to fully stimulate exonuclease processing suggests a specific role for the ion in the exonuclease catalytic mechanism. Recent work using time-resolved crystallography to examine reaction mechanisms of polymerases η[239,240] and β[241], as well as nucleases Exo1[242] and I-DmoI[243] has identified a transient third metal binding site necessary for driving forward the phosphoryltransfer reaction from transition state to products. Molecular dynamic

simulations have additionally suggested third-metal mechanisms may be especially important in 5’ phosphate-dependent nucleases[237,244]. Natively, this ‘product metal ion’ is predominantly coordinated either by α/ β-phosphates of an incoming dNTP in polymerases, or scissile and 5’ phosphates in nucleases[245]. The distinction between Zn-dependent recovery and Mg-dependent stimulation of SNM1A exonuclease activity may reflect an uncharacterized role of a transient product stabilizing third metal ion (**Figure 2.16B**).

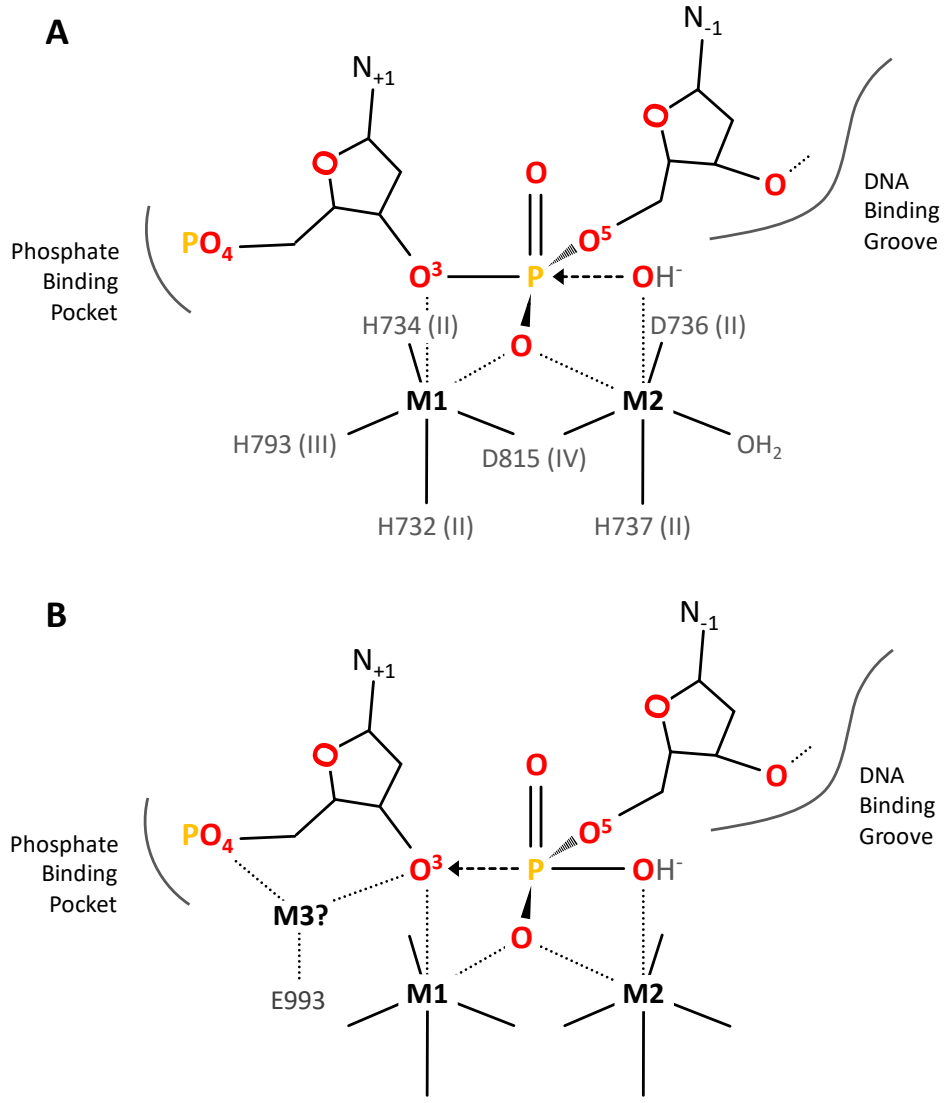


Figure 2.16. Hypothetical catalytic mechanism of SNM1A.

[A] Schematic of the 2-metal ion-induced transition state during exonuclease processing. M1 and M2 coordinate the O^3 – P bond and nucleophilic hydroxide, respectively. Nucleophilic attack generates a penta-covalent transition state via an in-line S_N2 reaction. **[B]** Schematic of the transition state resolution mechanism, stimulated by transient coordination of M3. Metal coordinating residues are labelled. Arrows indicate the direction of electron movement during reaction. Metal coordinating residues (MβL motif) indicated in [A].

As described above, SNM1A endonuclease catalysis lacked Mg stimulation, possibly due to an exo-specific uncharacterized M3 site. Further, rescue of SNM1A endonuclease activity was uniquely specific to low concentrations of Zn^{2+} alone. These results demonstrate distinct metal-dependent catalytic mechanisms for SNM1A exo- and endo-nuclease activities, both supported by the same active site. Where divalent metals promiscuously functioned in exonuclease processing – Mg^{2+}, Ca^{2+} and Cu^{2+} were entirely unable to facilitate endonuclease digestion (and Mn^{2+} and Ni^{2+} only at millimolar concentrations). Here, the dual-zinc MβL alone was sufficient for SNM1A endonuclease processing – presumably via the conventional 2-metal ion catalytic mechanism. It seems likely that SNM1A functions endogenously as an exo-/ and endo-nuclease with Zn^{2+} bound at M1 and M2. Though the mechanism of Mg-dependent stimulation has not been directly observed in SNM1A, the inherent efficiency of the exonuclease activity may be stimulated by an uncharacterized M3 site driving the transition state to products. It will be important (moving forward) to reflect these combinatorial metal requirements in reaction buffers when characterizing SNM1A function (instead of Mg^{2+} alone[121,177–179]).

2.6.5 Implications for SNM1A in Repair

Prevailing models of ICL repair suggest that SNM1A exclusively engages the product of an endonuclease incision (made by another factor) 5’ of the ICL and digests up to, and past, the ICL as an exonuclease[120,121]. In this way, SNM1A is thought to either complete unhooking or trim the oligonucleotide-adduct following unhooking by flanking endonuclease incisions. In support of this model, SNM1A was shown to function epistatically with the 3’ fork structure-specific endonuclease XPF in response to MMC or SJG-136 ICL damage[129]. XPF is a critical unhooking endonuclease involved in both FA- and TC-dependent ICL repair.

XPF-ERCC1 endonuclease incisions are heavily influenced by DNA structure[124,145], presence of a nascent leading strand[120,121], chemical nature of the crosslink[127] and

binding partner(s)[125,127,146] (see section **1.3.3.3**). Importantly, in *Xenopus* extracts, no FA-dependent unhooking occurred in cisplatin damaged DNA when XPF was not present[122]. This strongly suggests that XPF either (i) unhooks the crosslink directly with dual flanking incisions, or (ii) XPF action is coupled with another nuclease. Primary endonuclease incision(s) made by XPF-ERCC1 on 3' fork substrates occur within the duplex on the template leading strand. Depending on particular experimental circumstances, XPF typically makes one or two incisions between ~2-9 nucleotides into the duplex DNA[120,121,145,146,212]. *In vitro* models have convincingly shown that XPF in complex with SLX4 can unhook an ICL alone, by making dual incisions 1 and 5 nucleotides from the ss-dsDNA junction[127]. In contrast, when XPF is in complex with RPA, it makes a single incision 6 nucleotides from the junction (**see Figure 1.6**)[121]. In either situation, SNM1A could engage as an exonuclease and degrade from the 5'P released by the XPF incision made 5' of the ICL. Importantly, disruption of SNM1A reliably produces a more mild survival hypersensitivity in response to ICL-inducing damage than does disruption of XPF[129,208] or FANC core complex[154,162,189]. This indicates that SNM1A is likely processing only a subset of XPF-mediated incision products. Where the chemistry, position or helical distortion of an ICL prevents dual XPF endonuclease incisions, coordination with SNM1A may complete unhooking.

How SNM1A endonuclease activity may contribute to these repair functions is unclear. Previous work establishing a role for SNM1A nuclease activity in ICL repair relied on mutating metal-coordinating residues in the active site[176,189]. We now know that these mutants disrupt not only SNM1A exonuclease, but also endonuclease activity as well. Whether SNM1A function in repair depends on its exo-/ or endo-nuclease activity is an open question, though the relative preference for exonuclease substrates *in vitro* suggests functional relevance. During RDR, substantial stretches of single-strand DNA persist surrounding the stalled replication forks (template lagging strands) which could be a substrate for SNM1A endonuclease processing. Alternatively, future identification of an SNM1A endonuclease modulator (potentially a PTM or protein-interaction) may help determine where exactly SNM1A may be participating. Work

presented in **Chapter 4** builds on the characterization discussed here, defining whether both nuclease activities of SNM1A contribute to its role in repair.

2.7 Appendix

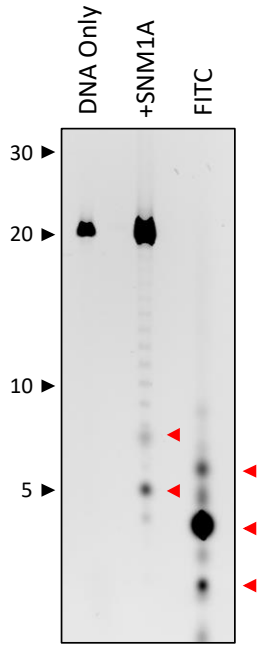
```
aa#                                                                              nt#
                                                                     AAA AAA       6
698                                                                   K   K
      ACC TGC CCG TTC TAC AAA AAA ATC CCT GGC ACC GGT TTT ACC GTT GAT GCA TTT CAG TAT      66
700    T   C   P   F   Y   K   K   I   P   G   T   G   F   T   V   D   A   F   Q   Y
      GGT GTT GTT GAA GGT TGT ACC GCA TAT TTT CTG ACC CAT TTT CAT TCC GAT CAT TAT GCC     126
720    G   V   V   E   G   C   T   A   Y   F   L   T   H   F   H   S   D   H   Y   A
      GGT CTG TCC AAA CAT TTT ACC TTT CCG GTT TAT TGC AGC GAA ATT ACC GGC AAT CTG CTG     186
740    G   L   S   K   H   F   T   F   P   V   Y   C   S   E   I   T   G   N   L   L
      AAA AAC AAA CTG CAT GTT CAA GAG CAG TAT ATT CAT CCG CTG CCG CTG GAT ACC GAA TGT     246
760    K   N   K   L   H   V   Q   E   Q   Y   I   H   P   L   P   L   D   T   E   C
      ATT GTT AAT GGT GTT AAA GTG GTT CTG CTG GAT GCA AAT CAT TGT CCG GGT GCA GTT ATG     306
780    I   V   N   G   V   K   V   V   L   L   D   A   N   H   C   P   G   A   V   M
      ATT CTG TTT TAT CTG CCG AAT GGC ACC GTT ATT CTG CAT ACC GGT GAT TTT CGT GCA GAT     366
800    I   L   F   Y   L   P   N   G   T   V   I   L   H   T   G   D   F   R   A   D
      CCG AGC ATG GAA CGT AGT CTG CTG GCA GAT CAG AAA GTT CAC ATG CTG TAT CTG GAT ACC     426
820    P   S   M   E   R   S   L   L   A   D   Q   K   V   H   M   L   Y   L   D   T
      ACC TAT TGT AGT CCG GAA TAT ACA TTT CCG AGC CAG CAA GAA GTT ATT CGC TTT GCA ATT     486
840    T   Y   C   S   P   E   Y   T   F   P   S   Q   Q   E   V   I   R   F   A   I
      AAC ACC GCA TTT GAA GCA GTT ACC CTG AAT CCG CAT GCA CTG GTT GTT TGT GGC ACC TAT     546
860    N   T   A   F   E   A   V   T   L   N   P   H   A   L   V   V   C   G   T   Y
      AGC ATT GGT AAA GAA AAA GTG TTT CTG GCC ATT GCA GAT GTT CTG GGT AGC AAA GTT GGT     606
880    S   I   G   K   E   K   V   F   L   A   I   A   D   V   L   G   S   K   V   G
      ATG AGC CAA GAA AAA TAC AAA ACC CTG CAG TGT CTG AAC ATT CCG GAA ATT AAC AGC CTG     666
900    M   S   Q   E   K   Y   K   T   L   Q   C   L   N   I   P   E   I   N   S   L
      ATT ACC ACC GAT ATG TGT AGC AGC CTG GTG CAT CTG CTG CCG ATG ATG CAG ATT AAC TTT     726
920    I   T   T   D   M   C   S   S   L   V   H   L   L   P   M   M   Q   I   N   F
      AAA GGT CTG CAG AGC CAC CTG AAA AAA TGT GGT GGC AAA TAT AAC CAG ATT CTG GCA TTT     786
940    K   G   L   Q   S   H   L   K   K   C   G   G   K   Y   N   Q   I   L   A   F
      CGT CCG ACC GGT TGG ACC CAT AGC AAC AAA TTC ACC CGT ATT GCC GAT GTT ATT CCG CAG     846
960    R   P   T   G   W   T   H   S   N   K   F   T   R   I   A   D   V   I   P   Q
      ACC AAA GGT AAC ATT AGC ATT TAT GGT ATC CCG TAT AGC GAA CAT AGC AGC TAT CTG GAA     906
980    T   K   G   N   I   S   I   Y   G   I   P   Y   S   E   H   S   S   Y   L   E
      ATG AAA CGT TTT GTT CAG TGG CTG AAA CCG CAG AAA ATT ATC CCG ACC GTT AAT GTT GGC     966
1000   M   K   R   F   V   Q   W   L   K   P   Q   K   I   I   P   T   V   N   V   G
      ACC TGG AAA AGC CGT AGC ACC ATG GAA AAA TAC TTT CGT GAA TGG AAA CTG GAA GCC GGT    1029
1020   T   W   K   S   R   S   T   M   E   K   Y   F   R   E   W   K   L   E   A   G
      TAT                                                                            1032
1040   Y
```

Appendix 2.A. *E.coli* codon optimized nucleotide sequence of SNM1A (698-1040) with corresponding amino acids.

```
                          HHHHHH RSDITSLYKKAGL DYDIPTT ENLYFQ^S
698  KKTCPFYKKIPGTGFTVDAFQYGVVEGCTAYFLTHFHSDHYAGLSKHFTFPVYCSEITGN
     LLKNKLHVQEQYIHPLPLDTECIVNGVKVVLLDANHCPGAVMILFYLPNGTVILHTGDFR
     ADPSMERSLLADQKVHMLYLDTTYCSPEYTFPSQQEVIRFAINTAFEAVTLNPHALVVCG
     TYSIGKEKVFLAIADVLGSKVGMSQEKYKTLQCLNIPEINSLITTDMCSSLVHLLPMMQI
     NFKGLQSHLKKCGGKYNQILAFRPTGWTHSNKFTRIADVIPQTKGNISIYGIPYSEHSSY
     LEMKRFVQWLKPQKIIPTVNVGTWKSRSTMEKYFREWKLEAGY*                1040
```

Appendix 2.B. Amino acid sequence of optimized SNM1A (698-1040) expression construct.

Highlighted N-terminal fusions include His_6 tag (yellow), attB1 scar (green), flexible linker (blue) and TEV protease recognition site (underlined). ^ denotes peptide bond cleaved by TEV protease, * indicates the C-terminal stop.



Appendix 2.C. 6FAM labelled DNA substrates can be aberrantly processed by SNM1A nuclease activity.

SNM1A (50 nM) was incubated with **F.20T** DNA (100 nM) substrate for 15 minutes at 37 °C. Products were analyzed on 22.5 % denaturing PAGE alongside fluorescein isothiocyanate (FITC; 0.5 pmol) and visualized with ChemiDoc MP imager. Red arrow indicate diffuse artifacts of free fluorophore derivatives.

Chapter 3
Inhibition of SNM1A: sensitizing cancer and probing function

3.1 Preface

The high-throughput screen of SNM1A bioactive inhibitors was completed at McMaster Center for Microbial Chemical Biology by Dr. Beverlee Buzon and Cameron Rzadki. Optimization of SNM1B purification was completed with Braeden Medeiros. SNM1A-specific *in silico* screening was completed by Patrick Melo. I completed all other experiments presented below.

Grainger R., Buzon B., Rzadki C. *et al.* (2021) *ACS Omega* 6(14), 9352-9361
DOI: 10.1021/acsomega.0c03528

3.2 Purpose

The purpose of this chapter is to identify effective small molecule inhibitors of SNM1A. Putative SNM1A inhibitors were characterized to determine potency, mode of action and specificity. These inhibitors can be used to (i) inhibit SNM1A action within ICL repair, sensitizing cells to ICL-inducing chemotherapeutics and (ii) as a tool to probe SNM1A function in a cellular context.

3.3 Introduction

Standard therapy following primary tumour identification consists of surgical resection followed by radiation or chemotherapy[2,246]. Cytotoxicity associated with chemotherapy is generally achieved through induction of DNA damage or replication stress[247–249]. Specifically, many common chemotherapeutics (such as cisplatin, nitrogen mustards and MMC) function through the generation of DNA interstrand crosslinks. ICLs covalently link opposing strands of DNA, resulting in a severe block on transcription and replication[2]. All cells rely on a complex network of DNA damage signalling to coordinate primary lesion detection with repair, and if necessary, cell cycle stalling and/or apoptosis[58,247]. One enabling hallmark associated with most tumorigenic events is dysregulation of DNA repair and replication. While this decoupling promotes rapid proliferation, it can result in severe genomic instability[250]. Recognition of an ICL in a non-cancerous cell promotes global replication fork stalling to prevent clastogenesis, concurrent with cell cycle stalling to prevent mitotic catastrophe[58,136]. Governed by a dysfunctional DNA damage response, cancerous cells often instead progress into mitosis irrespective of repair[250,251]. Damage response disruption most frequently occurs through inactivation of surveillance mechanisms (p53, PI3K-related kinases etc.) or DNA repair factors (BRCA1/2, MSH1/2, ERCC1 etc.)[247,248,250,252,253]. As such, ICL-inducing compounds exploit a therapeutic window resulting from the differential response between healthy and cancerous tissues.

ICL-inducing chemotherapies result in multiple forms of DNA damage, including DNA-protein crosslinks, intrastrand crosslinks, monoadducts – in addition to interstrand crosslinks. *In vivo*, ICLs represent only a small minority (1-10%) of damage following exposure to common 'ICL-inducing' drugs[7,254]. Though they represent a small minority of damage, the substantial cytotoxicity associated with these compounds depends specifically on ICL formation. This has been repeatedly demonstrated in studies directly comparing ICL-inducing compounds with their monofunctional analogues[2,5]. Though unrepaired monoadducts and intrastrand crosslinks can slow replication, translesion bypass of the damage is a relatively rapid process[37]. In contrast, interstrand crosslinks act as significant blocks to replication[39]. Where cancer cells lack functional DNA damage responses, replication and mitosis continue unimpeded, driving clastogenesis and mitotic catastrophe[247–249].

Resistance to ICL-inducing therapy significantly limits patient response and survival. Though ICL-inducing drugs are often extremely effective, some patients demonstrate an innate resistance. The frequency of innate resistance is often correlated with the origin of tumorigenic tissue[255]. Moreover, following primary treatment with ICL-inducing chemotherapy, many tumours develop an acquired resistance[253,256–259]. Innate and acquired resistance to ICL-including chemotherapy is multifactorial, arising from cellular alterations with respect to pre-target, on-target or post-target processes[257,259]. ICL-inducing chemotherapy selectively promotes cytotoxicity by crosslinking DNA, provoking aneuploidy and mitotic catastrophe in tumour cells[2,247]. Pre-target resistance can be accomplished by preventing drug access to DNA, commonly through disruption of active import mechanisms, upregulation of multi-drug resistance efflux pumps or cytosolic inactivation of the compound itself[257–260]. Post-target resistance mechanisms rely on the disruption of intracellular apoptotic activation and signalling[250,261]. Where programmed cell death is dysfunctional, cells will continue accumulating otherwise intolerable genomic instability. Finally, increased efficiency of DNA lesion recognition and repair through the hyperactivation or overexpression of DNA repair enzymes promotes increased tolerance to ICL-inducing drugs[248,253,257,259,262].

Each source of resistance lowers the therapeutic window available for chemotherapeutic success. As such, mutation or altered expression of genes involved in intake, efflux, apoptosis and repair have prognostic value for patient outcomes.

Developing tools to counteract the existence or development of resistance is a key priority for improving chemotherapeutic outcomes. Where innate or acquired resistance is driven by overexpression of DNA repair factors, specific small molecule inhibitors of key ICL repair enzymes will re-sensitize chemoresistant tumours. XPF-ERCC1 is a key structure-specific endonuclease facilitating transcription- and replication-mediated ICL repair (see **Chapter 1**). Expression levels (mRNA and protein) of ERCC1 have been shown to predict sensitivity, *ex ante*, in cervical[263], ovarian[264], testicular[265], bladder[265], melanoma[266] and lung cancers[255,267,268]. Tissues with innately elevated ERCC1 expression demonstrate an intrinsic resistance to cisplatin and other ICL-inducing agents[255,263,265,267,269]. Further, increased ERCC1 expression was directly shown to mediate chemoresistance, whether endogenously upregulated following primary treatment[266] or through exogenous overexpression[265]. As such, multiple small molecule inhibitors of XPF-ERCC1 have been developed with sufficient potency to abolish endonuclease activity *in vitro*, and XPF-mediated NER and ICL repair *in vivo*[270–278]. Treating ICL-resistant cancers concurrently with XPF-ERCC1 inhibitors and cisplatin has repeatedly demonstrated successful cisplatin sensitization, both in culture and mouse xenograft experiments[270,271,274,276].

Inhibition of XPF-ERCC1 will disrupt not only ICL repair, but also XPF-mediated NER and SSA[279]. Though effective in principle, this suggests XPF-ERCC1 inhibitors will simultaneously disrupt multiple, tangential DNA repair pathways. Where chemotherapeutic cytotoxicity is induced predominantly by ICL formation, these off-target impacts would be problematic. Instead, developing small molecule inhibitors against a target more specifically involved only in ICL repair could provoke similar mitotic catastrophe in cancer cells, without off-target effects. SNM1A has been shown to be epistatic with XPF-ERCC1, collaborating to unhook ICLs and generate substrates for downstream HR-mediated DSB repair[129]. As SNM1A functions in both replication[129] and

transcription[190] coupled repair, targeting SNM1A for inhibition should (like XPF-ERCC1) broadly disrupt the two major ICL repair pathways. Importantly, SNM1A has been identified as a prognostic indicator for chemotherapeutic response in ovarian[280] and advanced colorectal cancers[281]. Where increased expression of SNM1A was associated with poor prognosis, inhibition could sensitize these tumours to ICL-inducing compounds.

Reflecting the potential of SNM1A as a target for chemo-sensitization, several efforts have been initiated to develop small molecule inhibitors. To date, inhibitors of SNM1A have been identified in screens containing β-lactam antibiotics[282] and nucleoside analogues[283–287]. Multiple cephalosporin β-lactams inhibit SNM1A with low-to-mid micromolar potency, however, low membrane permeability limited potential therapeutic use[282]. In parallel, extensive SNM1A inhibitor development exploring modification of nucleoside analogues have been reported. A combination of thymidine or uridine scaffolds with 3' monodentate[283,284] or bidentate[285–287] zinc-binding groups (ZBG) effectively inhibit SNM1A at high micromolar concentrations. Further, 5' phosphorylation of ZBG-modified nucleosides dramatically improved potency, as high as 20 μM[286]. While promising, nucleoside analogues are expected to lack necessary specificity to be clinically useful. To directly address both membrane permeability and specificity, our lab completed a *de novo* high-throughput screen (HTS) of 3941 bioactive small molecules to identify novel inhibitors of SNM1A[210].

This chapter presents significant progress toward the development of specific and potent bioactive SNM1A inhibitors. Putative hits were validated, and inhibitor potency was quantified against SNM1A exonuclease and endonuclease activity using dose-dependent nuclease assays. Characterization of inhibitor mode of action was assessed using assays to monitor metal-chelation, colloidal aggregation and direct interaction with DNA. Further, to address limited specificity observed in initial bioactive inhibitors, an *in silico* HTS was completed to identify lead compounds with improved SNM1A selectivity. Leveraging structural differences surrounding SNM1A and SNM1B active sites, inhibitors were filtered to identify only those with an expected preference

for SNM1A. *In vitro* validation of these inhibitors confirmed that new lead compounds afforded inhibition of SNM1A with increased specificity.

3.4 Materials and Methods

3.4.1 DNA Substrates

All DNA oligonucleotides and substrates are summarized in **DNA Oligonucleotide Information** and **DNA Substrate Diagrams** in Supplemental Information, respectively. Sequences of all cloning and sequencing primers used are detailed in **Primer Information**. All DNA oligonucleotide concentrations were determined by NanoDrop™ 2000c Spectrophotometer (Thermo Scientific).

3.4.2 Compounds

All compounds from **Figure 3.1** were solubilized to 10 mM in DMSO. Bioactive inhibitors were purchased from Center of Microbial Chemical Biology at McMaster University (Hamilton, Ontario). Selective SNM1A inhibitors from **Appendix 3.A** and **Appendix 3.B** were purchased as a powder from MolPort (Beacon, New York) and dissolved to 10 mM in DMSO. All dilutions were made using DMSO and added to nuclease reactions to 5% v/v final. Compounds were stored at -80 °C until needed.

3.4.3 Nuclease Assays

3.4.3.1 IC_{50} Determination

SNM1A (0.2 nM in **Figure 3.2** or 100 nM **Figure 3.3**) was incubated with increasing concentrations of indicated inhibitors (1 nM - 100 µM) at room temperature for 20 minutes in Nuclease Reaction Buffer (50 mM Tris-acetate pH 7.2, 75 mM potassium acetate, 10 mM $MgCl_2$, 1 mM DTT and 100 µg/mL BSA). Serial dilutions of inhibitors solubilized in DMSO were diluted, in reaction, to 5% v/v final. Reactions were initiated by addition of either 110 nM exonuclease (**P.1F**) or 30 nM endonuclease (**Gapped.DS**) DNA substrates and incubated at 37 °C for 60 or 150 minutes, respectively. Reactions were terminated by addition of Formamide Buffer (95% formamide, 5 mM

EDTA) and products were analyzed by 22.5% denaturing PAGE. Gels were visualized by ChemiDoc MP Imager (BioRad).

Reaction progression was quantified using band integration in ImageLab (BioRad). Product intensity from each reaction (reported as 'Percent Activity') was normalized against a No Protein (low control) and No Inhibitor (high control) according to:

Equation 4

$$\%\ Activity = \frac{Product - Low\ Control}{High\ Control - Low\ Control} \times 100\%$$

Inhibitor potency (IC_{50}) was determined using curve fitting by variable slope, non-linear regression in GraphPad PRISM v8.4.3. Error bars reflect standard error of the mean from at least three independent replicates.

3.4.3.2 Metal-Dependent Reversible Inhibition

SNM1A (2 nM) was incubated in EMSA Buffer (50 mM Tris-acetate pH 7.2, 75 mM potassium acetate, 1 mM DTT and 100 µg/mL BSA) containing 1 mM $MgCl_2$ or 50 µM zinc acetate, and increasing concentrations of either zinc acetate (5 – 250 µM) or $MgCl_2$ (25 – 2000 µM), respectively, with or without indicated inhibitor. All inhibitors were present at indicated concentrations and incubated with protein for 10 minutes at room temperature prior to initiation with 100 nM DNA (**P.1F**). Reactions were incubated at 37 °C for 15 minutes and terminated by addition of Formamide Buffer. Products were analyzed using 22.5% denaturing PAGE and visualized by ChemiDoc MP Imager (BioRad).

Reaction progress was quantified using band integration in ImageLab, where product accumulation was normalized to 'Inhibitor' (low control) and 'No Inhibitor' (high control), using **Equation 4**. Percent Activity was plotted with and without inhibitor in reaction.

3.4.3.3 Detergent-Dependent Reversible Inhibition

SNM1A (2.5 nM) was incubated with inhibitor (5% v/v) at indicated concentrations in Nuclease Reaction Buffer with or without 0.01% v/v Triton X-100 at room temperature for 10 minutes. Reactions were initiated with addition of 100 nM DNA (**P.20TF**) and incubated at 37 °C for 15 minutes. Reactions were terminated with Formamide Buffer. Products were analyzed by 22.5% denaturing PAGE and visualized with ChemiDoc MP Imager (BioRad).

3.4.3.4 Inhibitor Nuclease Specificity

All nuclease reactions contained SNM1A inhibitors at 2x IC_{50} concentrations reported in **Figure 3.2**. SNM1B and FAN1 were purified as described below in sections **3.4.8.2** and **3.4.8.3**, respectively. Each nuclease was incubated with indicated inhibitors (5% v/v final DMSO) for 10 minutes at room temperature. Reactions were initiated with 100 nM DNA substrate for indicated time and temperature. Reactions were terminated with Formamide Buffer and products were separated using 22.5% denaturing PAGE. Gels were visualized by ChemiDoc MP Imager (BioRad).

SNM1A (2.5 nM) and SNM1B (13 nM) reactions were prepared in Nuclease Reaction Buffer with **P.1F** DNA substrate and incubated at 37 °C for 15 minutes. FAN1 (170 nM) reactions were incubated in 20 mM Tris pH 7.5, 15 mM KCl, 1 mM DTT, 5 mM $MgCl_2$ and 50 µg/mL BSA with **P.(1)Flap** DNA substrate at 37 °C for 60 minutes. hAPE1 (4 mU, New England Biolabs) reactions were prepared in 50 mM potassium acetate, 20 mM Tris pH 7.9, 10 mM magnesium acetate and 1 mM DTT with **F.AB** DNA substrate at 37 °C for 45 minutes. *Thermococcus* FEN1 (64 mU, New England Biolabs) was incubated in 20 mM Tris-HCl pH 8.8, 10mM $(NH_4)_2SO_4$, 10 mM KCl, 2 mM $MgSO_4$ and 0.1% Triton X-100 with **F.(6)Flap** DNA substrate at 45 °C for 60 minutes.

Reaction quantification was completed using band integration with ImageLab. Reaction progress (reported as '% Inhibition') was normalized to 'No Protein' (high control) and 'No Inhibitor' (low control) according to a modified **Equation 4**.

3.4.4 Compound Analysis

3.4.4.1 ADME Prediction

Pharmacological features (Administration, Distribution, Metabolism and Excretion) were predicted in **Tables 3.1 and 3.2** using the online tool, SwissADME (Swiss Institute of Bioinformatics)[288].

3.4.4.2 PAINS Prediction

SNM1A bioactive inhibitors were screened for problematic chemical motifs against the Pan-Assay Interference Compounds (PAINS) database, using the online webserver ZINC15[289].

3.4.4.3 Colloidal Aggregation Prediction

SNM1A bioactive inhibitors were cross-referenced against the ZINC15 database of reported chemical aggregators. Inhibitors were flagged as potential colloidal aggregators if they (or a similar compound) had been previously reported to aggregate in a high throughput screen. Similarity was defined as having a Tanimoto similarity coefficient greater than 0.95.

3.4.4.4 Ligand Efficiency Calculation

Ligand efficiencies (LE) of bioactive SNM1A inhibitors were calculated according to:

Equation 5

$$Ligand\ Efficiency\ (\Delta G/atom) = \frac{-R\,T\ln(IC_{50})}{N}$$

where: *R* is the universal gas constant (kcal K^{-1} mol^{-1}), *T* is temperature (K), IC_{50} is the half-maximal inhibitor concentration of each SNM1A inhibitor (mol L^{-1}) (as measured against both exonuclease and endonuclease activity), and *N* refers to the number of non-hydrogen atoms in the inhibitor.

3.4.5 Ethidium Bromide Displacement Assay

Duplex DNA (**P.DS**, 5 pmol) was incubated with ethidium bromide (15 pmol) in phosphate-buffered saline (PBS) at 37 °C for 10 minutes. Inhibitors were added at IC_{50} concentrations reported in **Figure 3.2**, to 10% v/v DMSO in reaction. Fluorescence at 595 nm was monitored in 384-well plates (Corning 3575) using a Synergy H1 plate reader (BioTek) at 37 °C for 15 minutes. Fluorescence from each displacement reaction was normalized with No DNA (high control) and No Inhibitor (low control) according to (where F refers to fluorescence):

Equation 6

$$\%\ Displacement = \frac{F_{Experimental} - F_{No\ Inhibitor}}{F_{No\ DNA} - F_{No\ Inhibitor}}$$

3.4.6 Inhibitor-Induced Precipitation Assay

SNM1A (1.5 µM) was incubated with indicated inhibitors (50 µM) in Nuclease Reaction Buffer with or without 0.01% v/v Triton X-100 at room temperature for 20 minutes. Samples were centrifuged at 14,800 x g and supernatant was analyzed by 15% SDS PAGE. Gels were visualized with 0.5% v/v TCE stain-free by GelDoc EZ Imager (BioRad).

3.4.7 Small Molecule Docking

The full, purchasable HTS compound library from the Center for Microbial Chemical Biology (McMaster University; Hamilton, Ontario) was converted into digital 3D structure data files (SDF) and filtered with the MCULE webserver. Filters removed compounds that (i) violate two or more of Lipinski's Rule of 5, (ii) contain motifs predicted to be highly reactive (Rapid Elimination of Swill [REOS] filter) and (iii) were highly similar compounds (sharing Tanimoto coefficient greater than 0.85). A curated list

of 100,000 compounds was docked to targets using AutoDock Vina on the MCULE webserver.

All ligands and water (except active site metals) were removed from target structures: SNM1A (PDB: 5Q2A) and SNM1B (PDB: 5AHO). The SNM1A structure was modified to incorporate two zinc ions in the active site (refer to section **4.4.2**). An additional SNM1A target file was modified to truncate flexible residues (to β- or γ-carbons), adjacent to the active site (**Appendix 3.C**). Compounds were sequentially docked to ε-nitrogen of H994 or H276 of SNM1A and SNM1B, respectively. Compounds predicted to bind SNM1A with the highest affinity (top 5000 hits) were cross-docked to SNM1B. Hits with a predicted preferential binding affinity to SNM1A with or without truncated residues, exceeding 1 or 2 kcal mol^{-1}, respectively, were purchased for *in vitro* validation.

3.4.8 Purification

3.4.8.1 SNM1A

SNM1A$^{\Delta N697}$ was purified as described previously in section **2.4.4.1**.

3.4.8.2 SNM1B

SNM1B (Uniprot KB ID# 9QH816) was truncated to 1-335 (SNM1B$^{\Delta C336}$, further referred to as SNM1B) and cloned into pDEST-527 expression vector using Gateway™ cloning (as described with SNM1A in section **2.4.2.1**) containing an N-terminal, TEV cleavable His_6 tag. SNM1B in pDEST-527 was transformed by heat shock into Rosetta™ DE3 pLysS (Novagen) *E. coli* cells and plated on LB agar containing 100 µg/mL ampicillin selection at 37 °C overnight. Liquid LB with ampicillin selection was inoculated by multi-colony streak of transformants and grown to saturation at 37 °C and 225 rpm. Saturated cultures were used to inoculate fresh LB with ampicillin selection which were grown at 37 °C and 225 rpm until OD_{600} 1.5. Cultures were incubated with 5% v/v ethanol for 10 minutes, cooled to room temperature on ice and induced with 1 mM IPTG at 25 °C,

overnight. Cells were harvested by centrifugation at 3,315 x g for 45 minutes, resuspended in NiA^{S1B} Buffer (500 mM NaCl, 50 mM Tris pH 7.5, 10% v/v glycerol, 0.5 mM TCEP, 10 mM imidazole), flash frozen in liquid nitrogen and stored at -80 °C.

Cell suspensions were thawed and diluted to 1:10 m/v (g pellet/ mL buffer) with 0.01% v/v Triton X-100 and protease inhibitors (3 µM aprotinin, 1 µM pepstatin A, 1 mM benzamidine, 1 µM leupeptin and 1 mM PMSF). Cells were lysed by cell disrupter (Avestin Emulsiflex C3) with three passes at 10,000 psi. Lysate was clarified by centrifugation at 48,384 x g for 45 minutes, vacuum filtered (0.45 µm Triton-free MCE, Millipore) and loaded onto a Nickel-chelating HisTrap HP column (GE Healthcare). Following extensive washing with NiA^{S1B} containing 10 and 55 mM imidazole, SNM1B was eluted with NiB^{S1B} Buffer (NiA^{S1B} containing 300 mM imidazole). SNM1B was buffer exchanged (HiPrep 26/10, GE Healthcare) into 500 mM NaCl, 50 mM Tris pH 7.5, 10% glycerol and 0.5 mM TCEP, and incubated with 1:1 m/m TEV protease at 4 °C for 48 hours. Cleaved protein was loaded onto a 1 mL Nickel-chelating HisTrap HP column (GE Healthcare) and washed with NiA^{S1B} containing 15, 30, 45 and 300 mM imidazole. SNM1B containing fractions were combined, exchanged (HiPrep 26/10) into Storage Buffer (200 mM NaCl, 50 mM Tris pH 7.5, 10% v/v glycerol and 0.5 mM TCEP) and concentrated using Vivaspin Turbo15 10 kDa MWCO Centricon (Sartorius) at 1000 x g to 27 µM (1 mg/mL). Concentrated SNM1B was aliquoted into single-use thin-walled PCR tubes, flash frozen in liquid nitrogen and stored at -80 °C until needed.

3.4.8.3 FAN1

FAN1 (Uniprot KB ID# Q9Y2M0) was truncated to 371-1017 ($FAN1^{\Delta N370}$, further referred to as FAN1) and cloned into pDEST-566 expression vector containing an N-terminal TEV cleavable His_6-MBP tag using Gateway™ cloning as described in section **2.4.2.1**. pDEST-566 was a gift from Dominic Esposito (Addgene plasmid #11517). FAN1 in pDEST-566 was transformed into BL21 DE3 (Invitrogen) *E. coli* by heat shock and plated overnight at 37 °C on LB agar with 100 µg/mL ampicillin selection. Liquid LB with ampicillin selection was inoculated by a multi-transformant streak and grown to

saturation at 37 °C and 225 rpm. Saturated cultures were used to inoculate (1:100 v/v) fresh LB with ampicillin selection. Cultures were grown at 37 °C and 225 rpm until OD_{600} 0.5, cooled to room temperature on ice, and induced with 50 µM IPTG at 16 °C at 225 rpm overnight. Cells were harvested by centrifugation at 3,315 x g, resuspended in NiA^{FAN1} Buffer (500 mM NaCl, 50 mM Tris pH 8.0, 5 mM imidazole), flash frozen in liquid nitrogen and stored at -80 °C until purification.

Cell suspensions were thawed and diluted to 1:10 m/v (g pellet/ mL buffer) with NiA^{FAN1} containing protease inhibitors (same as section **3.4.8.2**). Cells were lysed by cell disrupter (Avestin Emulsiflex C3) with three passes at 10,000 psi. Lysate was clarified by centrifugation at 48,384 x g for 45 minutes, vacuum filtered (0.45 µm Triton-free MCE, Millipore) and loaded onto a 5 mL EDTA-resistant Ni-Penta^{TM} IMAC (Marvelgent). Following extensive washing with NiA^{FAN1} containing 27 and 49.25 mM imidazole, FAN1 was eluted with NiB^{FAN1} Buffer (500 mM NaCl, 50 mM Tris pH 8.0, 300 mM imidazole). Eluted protein was buffer exchanged (HiPrep 26/10) into 150 mM NaCl, 50 mM Tris pH 8.0 and 1 mM DTT, and incubated with 1:3 m/m TEV protease at 4 °C, overnight. Cleaved protein was exchanged into NiA^{FAN1} Buffer and loaded on an EDTA-resistant Ni-Penta^{TM} IMAC column. The column was washed with NiA^{FAN1} Buffer containing 5, 27.5, 49.25 and 300 mM imidazole. Samples containing FAN1 were combined and concentrated using Vivaspin Turbo15 10 kDa MWCO Centricon (Sartorius) at 1000 x g to 1 mL. Concentrated FAN1 was injected over a size exclusion column (Superdex 200 10/300 GL, GE Healthcare) equilibrated with FAN1 Storage Buffer (150 mM NaCl, 20 mM Tris pH 8.0 and 10% v/v glycerol) and purified at 0.5 mL/min. Pure FAN1 fractions were pooled, diluted with FAN1 Storage Buffer to 14 µM (1 mg/mL), flash frozen in liquid nitrogen and stored at -80 °C.

3.5 Results

3.5.1 Characterization of bioactive inhibitors of SNM1A

An *in vitro* high-throughput screen (HTS) of SNM1A was previously conducted in our lab against the 3941 compound bioactive library from the Centre for Microbial Chemical Biology (McMaster University)[210]. The HTS relied on plate reader detection of SNM1A exonuclease processing of a short single-stranded DNA substrate (containing a fluorophore-quencher pair). Compounds which reduced the generation of free fluorophore (lower SNM1A nuclease activity) were deemed putative SNM1A inhibitors. Initial hits (52 compounds) were tested for dose-dependence (22/52 compounds) and validated with a gel-based secondary screen (9/22 compounds). The identity and chemical structures of validated inhibitors of SNM1A (Series A) are shown in **Figure 3.1**. Following identification and initial validation, further characterization was required to determine potency, drug-likeness, mechanisms of action and specificity.

Though SNM1A is a dual-function nuclease, the HTS only tested inhibition on exonuclease activity. An inhibitor dose-dependent, gel-based nuclease assay was therefore, used to quantify potency with SNM1A exonuclease (**Figure 3.2**) and endonuclease (**Figure 3.3**) activity. Potency between inhibitors ranged nearly 1000-fold, from 80 nM (Compound A40) to 68 µM (A7) and 1.6 µM (A20) to 195 µM (A7) with SNM1A exonuclease and endonuclease activity, respectively (summarized in **Figure 3.4**). Nearly all inhibitors (except A13 and A27) exhibited increased potency on exonuclease activity, with four inhibitors demonstrating significant exonuclease preference (>5-fold; A40, A20, A30, A53). Notably, compound A61 showed no detectable inhibition against SNM1A endonuclease activity.

In silico tools were used to analyze each inhibitor for lead-likeness, ligand efficiency (**Table 3.1**), pharmacokinetics and common HTS issues (**Table 3.2**). Collectively, these attributes highlight anticipated potential issues likely to arise when translating *in vitro* inhibition to an *in vivo* system. Many compounds found in large screening libraries contain motifs known to be promiscuous or highly reactive (PAINS).

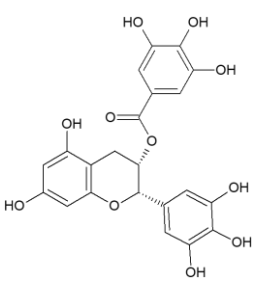
Presence of these motifs suggest the compound may not be useful for further lead development (or requires modification) because the inhibition is either non-specific, indirect or artifactual. In particular, four SNM1A inhibitors (A7, A27, A40, A61) contain catechol moieties and another four (A24, A27, A30, A40) are (or closely resemble) reported colloidal aggregators.

Catechols are 1,2-dihydroxybenzene moieties which commonly interfere with enzyme activity directly or indirectly by chelating active site or other metals required for catalysis. To directly test whether any catechol-containing SNM1A inhibitor functions by non-specific metal chelation, Mg^{2+} or Zn^{2+} were titrated into nuclease reactions and rescue of inhibited SNM1A activity was monitored. If any inhibitor functioned through non-specific metal chelation, incubating an inhibited SNM1A reaction with increasing amounts of divalent metals would restore SNM1A nuclease activity. In contrast, no rescue was observed for the catechol-containing inhibitors (A7, A27, A40, A61), clearly demonstrating that they do not function exclusively through a chelation-dependent mechanism (**Figure 3.5**).

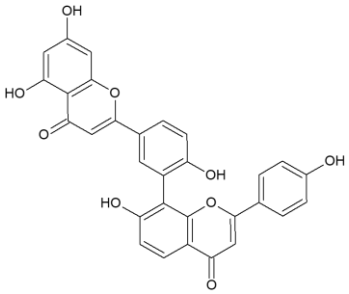
Colloidal aggregation is a common feature of many organic small molecules present in screening collections. These aggregates form at the interface of DMSO (used to solubilize screening compounds) and the aqueous reaction. Colloid surfaces non-specifically interact with proteins, sequestering and/or partially unfolding them, resulting in inhibition[290]. These small molecule aggregates can be disrupted by the addition of a detergent. SNM1A inhibitors were screened against the ZINC15 database[289] to identify previous reports of colloidal behaviour. Four SNM1A inhibitors (A24, A27, A30, A40) are (or are similar to) reported/predicted aggregators (**Table 3.2**). To determine whether any flagged inhibitors function via this non-specific mechanism, aggregate-induced protein precipitation or nuclease inhibition was tested using Triton X-100. Although no inhibitor-induced precipitation was observed (**Figure 3.6A**), compounds A24 and A30 showed clear detergent-dependent reversal of inhibition (**Figure 3.6B**). This demonstrates that compounds A24 and A30 primarily function through non-specific colloidal aggregation.

Many small molecules can interact directly with DNA substrates. When this interaction promotes helical distortion or disrupts enzyme-DNA contacts, the small molecule will non-specifically inhibit nuclease catalysis. Any inhibitor which functions by modifying DNA conformation or blocking DNA interaction with SNM1A would have limited utility *in vivo* due to expected lack of specificity. These compounds would not inhibit SNM1A, but instead inhibit anything from interacting with DNA. To test whether any SNM1A inhibitor functions via non-specific DNA interaction, ethidium bromide (EtBr) displacement was monitored with and without each inhibitor. Displacement assays exploit the differential fluorescence of EtBr, either in complex with DNA or free in solution. Monitoring changes in EtBr fluorescence informs whether an inhibitor directly impacts the association of EtBr with the DNA substrate. Significant displacement of EtBr intercalation was observed when DNA-EtBr complexes were incubated with compounds A20, A27 and A30 at exonuclease IC_{50} concentrations (**Figure 3.7**; IC_{50} values summarized in **Figure 3.4**). Displacement demonstrates that these compounds likely (at least in part) function through non-specific DNA interaction.

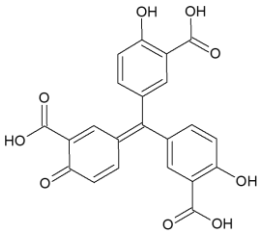
Finally, to determine the specificity of SNM1A inhibitors, each compound was tested for inhibition of four other nucleases. Though perfect specificity is not required for successful drug development, specificity is inversely proportional to expected off-target (ie. side-effects) activity *in vivo*. Nucleases tested include human homolog SNM1B, unrelated human DNA repair nucleases FAN1 and APE1, and *Thermococcous* DNA repair nuclease FEN1. Each nuclease was optimized for its preferred substrate and incubated with inhibitors at double the SNM1A exonuclease IC_{50} concentrations (summarized in **Figure 3.4**). All compounds demonstrated equivalent or increased potency with at least one other nuclease, though to different extents (**Figure 3.8**). Unfortunately, this lack of specificity is likely to limit the development of Series A inhibitors into useful drugs targeting SNM1A and, more broadly, ICL repair.



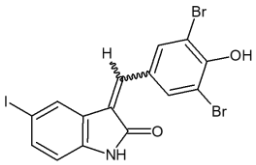
Epigallocatechin gallate (A7)



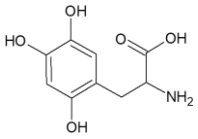
Amentoflavone (A13)



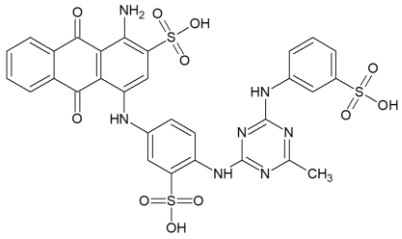
Aurintricarboxylic acid (A20)



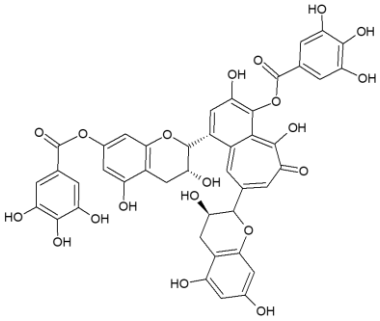
GW5074 (A24)



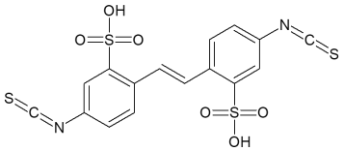
6-Hydroxy-DL-DOPA (A27)



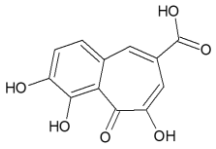
Reactive Blue 2 (A30)



Theflavin digallate (A40)



4,4’-diisothiocyanostilbene-2,2’-disulfonic acid (A53)



Methyl-7-deshydroxypyrogallin-4-carboxylic acid (A61)

Figure 3.1. Chemical structure of SNM1A bioactive inhibitors.

Validated inhibitors of SNM1A identified in HTS campaign (Series A). Inhibitor ID indicated in brackets.

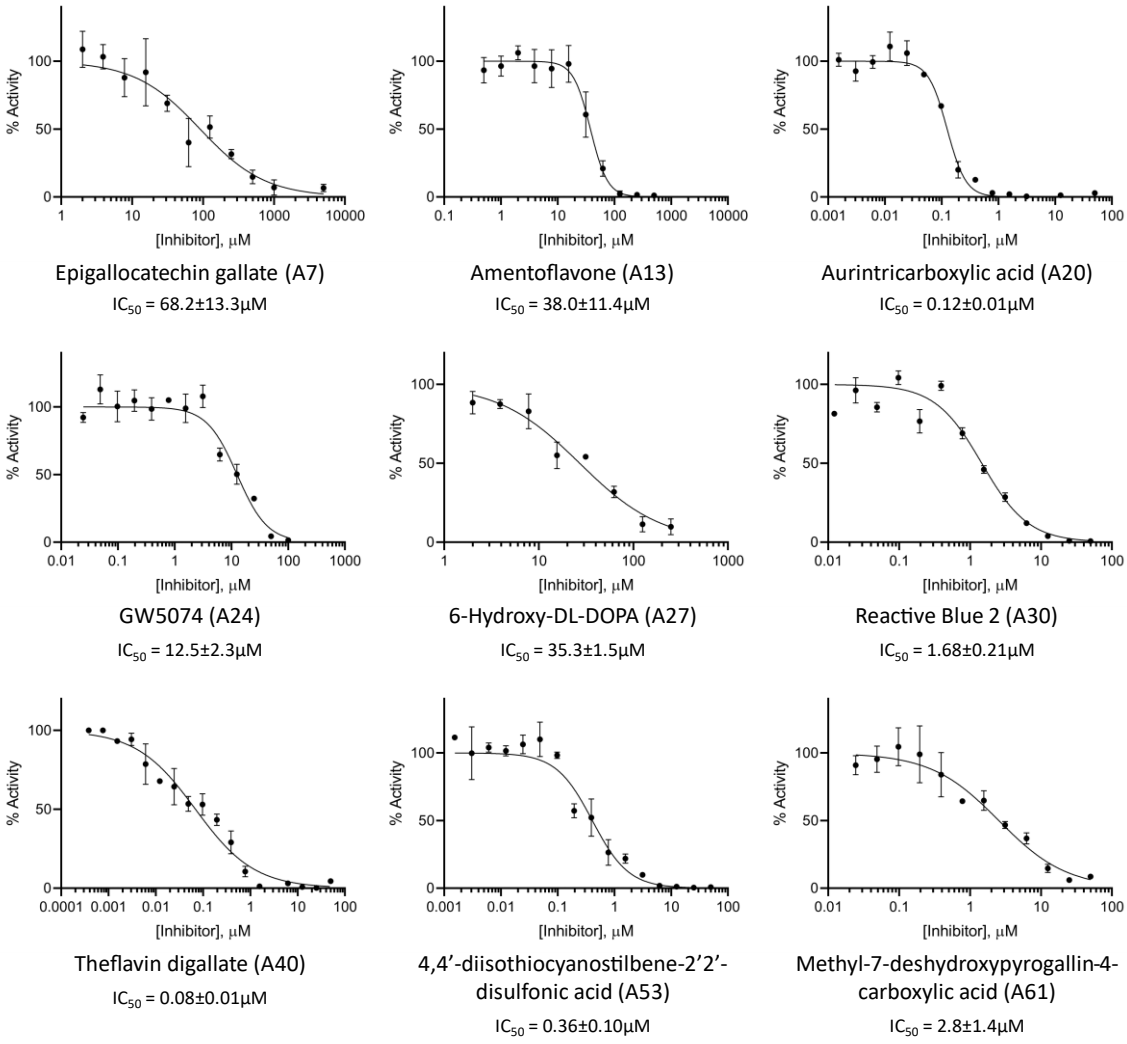


Figure 3.2. Exonuclease IC_{50} quantification of SNM1A bioactive inhibitors.

SNM1A (0.2 nM) was incubated with increasing inhibitors (1 nM – 100 μM) for 20 minutes at room temperature. Nuclease reactions were initiated with **P.1F** (110 nM) and incubated at 37 °C for 60 minutes. Products were analyzed using 22.5% denaturing PAGE and visualized with ChemiDoc MP Imager. Integration of product intensity was determined using ImageLab and '% Activity' was determined following normalization to a low 'No Protein' and high 'No Inhibitor' control. Selected inhibitor concentration ranges are displayed above. Four-variable sigmoidal curve fitting in GraphPad PRISM v8.4.3 was used to calculate IC_{50}. Error bars represent standard error of the mean from at least three independent replicates.

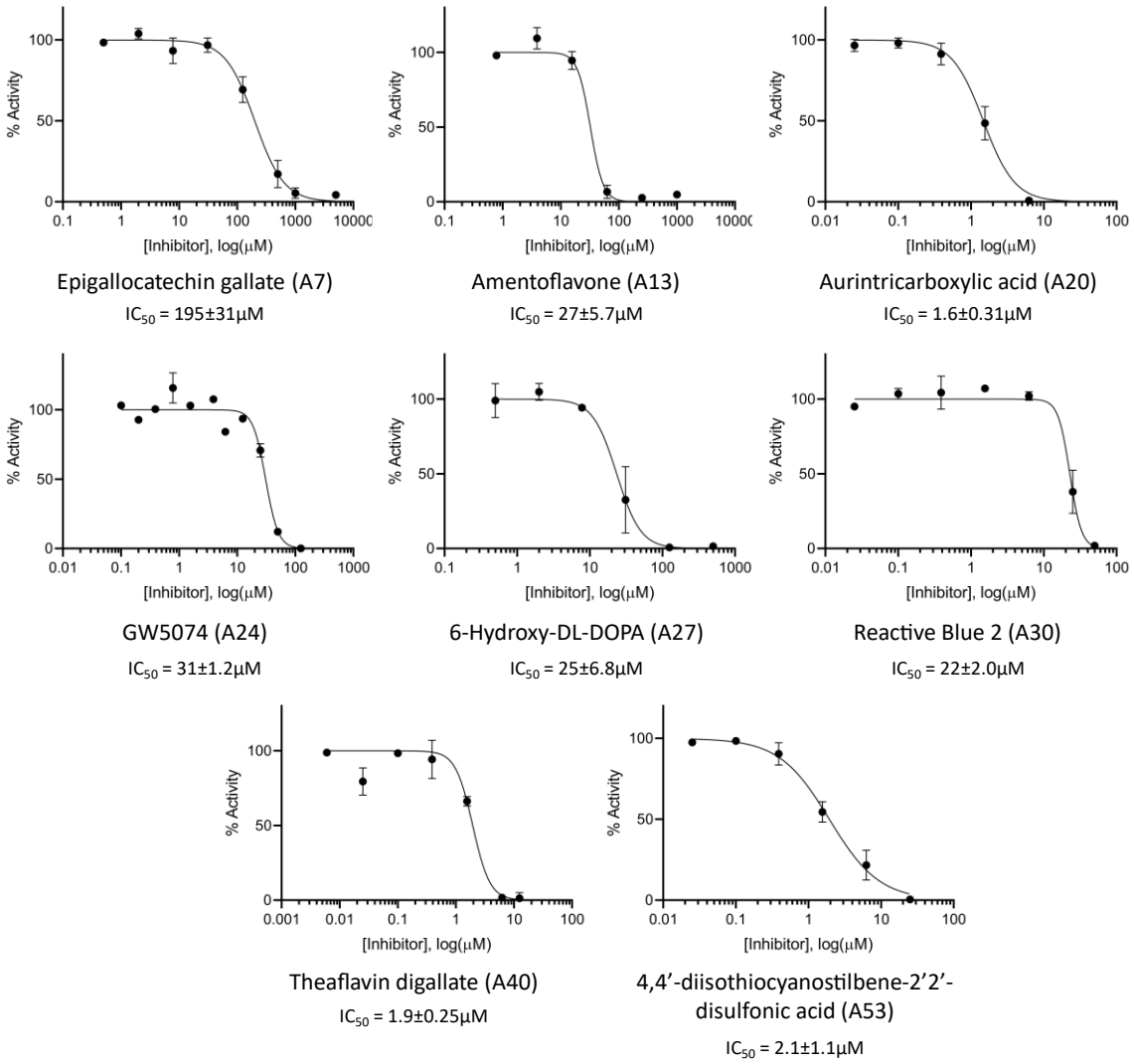
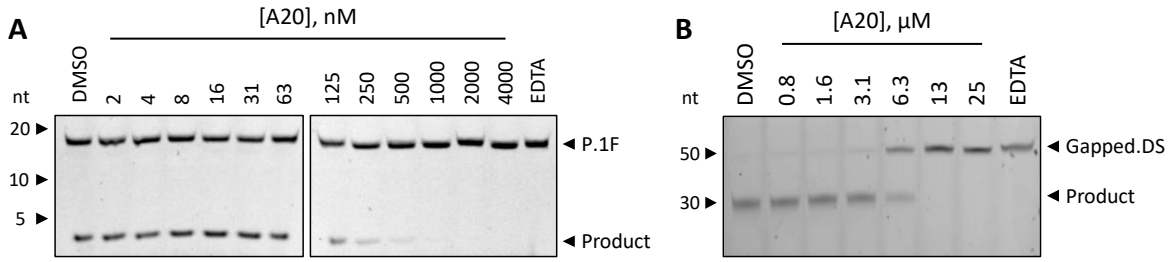


Figure 3.3. Endonuclease IC_{50} quantification of SNM1A bioactive inhibitors.

SNM1A (200 nM) was incubated with increasing inhibitors (30 nM – 5 mM) for 20 minutes at room temperature. Nuclease reactions were initiated with **Gapped.DS** (30 nM) and incubated at 37 °C for 150 minutes. Products were analyzed using 22.5% denaturing PAGE and visualized with ChemiDoc MP Imager. Integration of product intensity was determined using ImageLab and '% Activity' was determined following normalization to a low 'No Protein' and high 'No Inhibitor' control. Selected inhibitor concentration ranges are displayed above. Four-variable sigmoidal curve fitting in GraphPad PRISM v8.4.3 was used to calculate IC_{50}. Error bars represent standard error of the mean from at least three independent replicates.



C

Inhibitor	Exonuclease IC_{50} (µM)	Endonuclease IC_{50} (µM)	Fold Difference Endo/ Exo
A40	0.08 ± 0.01	1.9 ± 0.3	23.8
A20	0.12 ± 0.01	1.6 ± 0.3	13.3
A53	0.36 ± 0.10	2.1 ± 1.1	5.8
A30	1.7 ± 0.2	22 ± 2.0	12.9
A61	2.8 ± 1.4	N/D	-
A24	12 ± 2.3	31 ± 1.2	2.6
A27	35 ± 1.5	25 ± 6.8	0.7
A13	38 ± 11	27 ± 5.7	0.7
A7	68 ± 13	195 ± 31	2.9

Figure 3.4. Representative SNM1A bioactive inhibitor potency assays.

Representative gels of inhibitor IC_{50} determination with SNM1A **[A]** exonuclease and **[B]** endonuclease activity. Reactions were carried out as described in **Figure 3.2** and **Figure 3.3**. Products were analyzed using 22.5% denaturing PAGE and visualized with ChemiDoc MP Imager. Gels are representative of three independent replicates. **[C]** Summary of bioactive inhibitor IC_{50} quantification.

A

Inhibitor ID	MW (g/mol)	H-Bond Acceptors	H-Bond Donors	$logP_{o/w}$*
A7	458.37	11	8	1
A13	522.46	9	5	3.86
A20	422.34	9	5	1.96
A24	520.94	2	2	4.27
A27	213.19	6	5	-1.12
A30	840.1	14	4	2.62
A40	868.7	20	13	1.48
A53	498.48	6	8	4.12
A61	262.21	6	3	1.2
RO5†	<500	<10	<5	<5

B

		Ligand Efficiency	
Inhibitor ID	# Non-H Atoms	Exonuclease	Endonuclease
A7	33	0.17	0.15
A13	39	0.15	0.16
A20	31	0.30	0.26
A24	21	0.32	0.29
A27	15	0.41	0.42
A30	50	0.16	0.13
A40	63	0.15	0.12
A53	28	0.31	0.28
A61	19	0.40	NA

Table 3.1. *In silico* prediction of SNM1A bioactive inhibitor lead-likeness and ligand efficiencies.

[A] Analysis of inhibitor violations of Lipinski's Rule of 5†. *Lipophilicity ($logP_{o/w}$) is quantified using multiple, distinct models and reported as an average prediction determined by SwissADME. **[B]** Ligand efficiency (**Equation 5**) calculation for bioactive inhibitors

Inhibitor ID	ADME Prediction				Problematic Chemical Motifs	
	GI Absorption	Blood-Brain Barrier Permeability	Substrate of P-glycoprotein	Cytochrome P450 Inhibitor? (# of Isoforms)	PAINS Motif Present?	Colloidal Aggregator?
A7	Low	No	No	No	Catechol	No
A13	Low	No	No	Yes (1)	No	No
A20	Low	No	No	Yes (1)	No	No
A24	High	Yes	No	Yes (3)	No	Yes
A27	High	No	No	No	Catechol	Yes
A30	Low	No	Yes	No	Anthanil, Quinone	Yes
A40	Low	No	Yes	Yes (1)	Catechol	Yes
A53	Low	No	Yes	No	No	No
A61	High	No	No	No	Catechol	No

Table 3.2. *In silico* prediction of SNM1A bioactive inhibitor pharmacokinetics and chemical PAINS.

In silico pharmacokinetic (ADME: absorption, digestion, metabolism and excretion) parameters using SwissADME online tool, problematic chemical motifs including PAINS and similarity to known aggregators (Tanimoto coefficient: T_C >0.95) using ZINC virtual screening database.

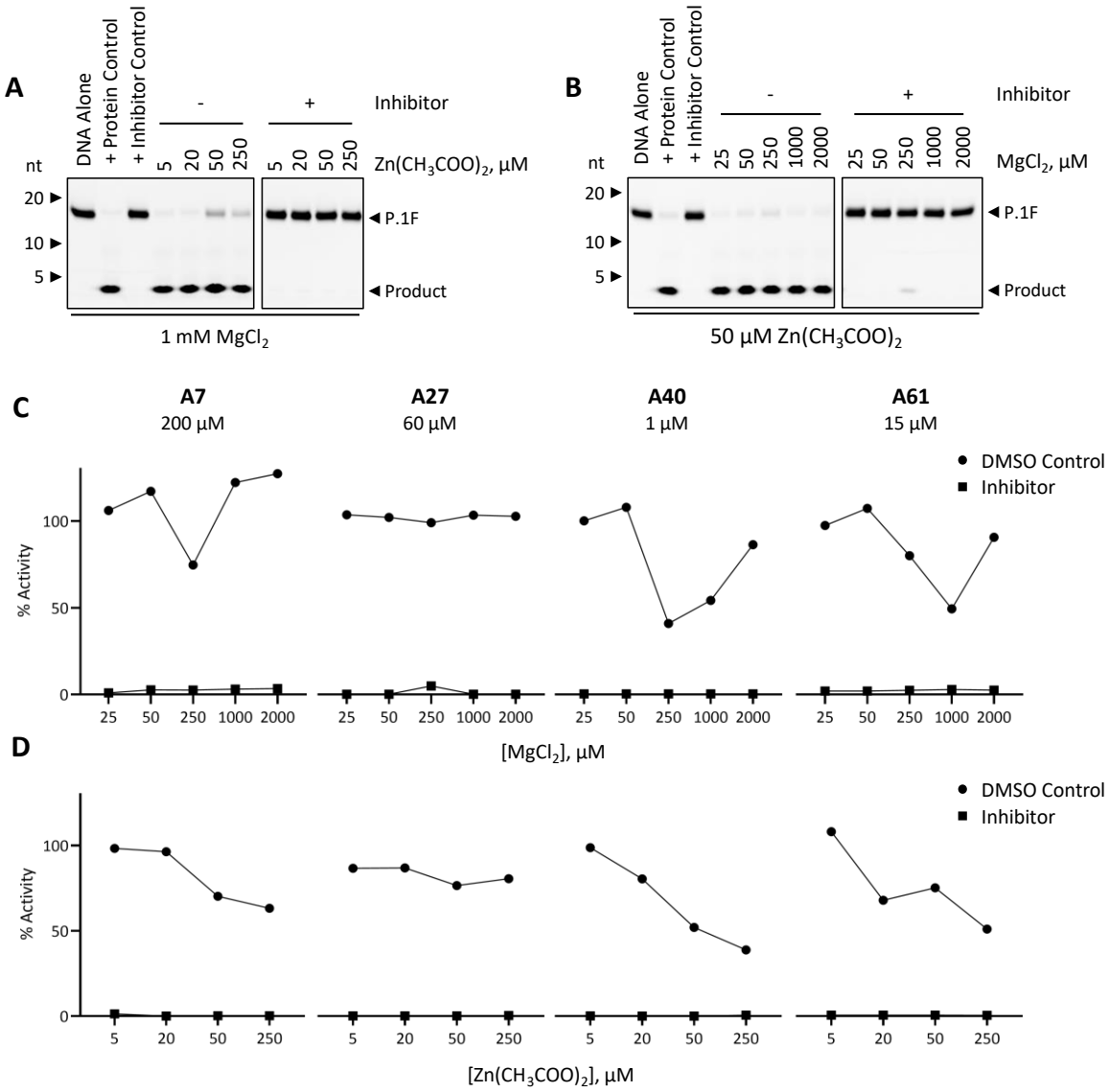


Figure 3.5. SNM1A bioactive inhibitors do not function through non-specific metal chelation.

Analysis of inhibitors containing a catechol motif identified in **Table 3.2**. Representative gels of metal rescue assays titrating increasing amounts of **[A]** $Zn(CH_3COO)_2$ (5 – 250 µM) or **[B]** $MgCl_2$ (25 – 2000 µM). SNM1A (2 nM) was incubated with inhibitor (indicated concentrations) and metal in a minimal nuclease activity buffer (containing 1 mM $MgCl_2$ or 50 µM $Zn(CH_3COO)_2$) for 10 minutes at room temperature. Reactions were initiated with addition of 100 nM **P.1F** and incubated at 37 °C for 15 minutes. Products were analyzed using 22.5% denaturing PAGE and visualized on ChemiDoc MP Imager. Quantification of **[C]** $MgCl_2$ and **[D]** $Zn(CH_3COO)_2$ metal rescue assays. Integration of product intensity was determined using ImageLab and '% Activity' was normalized with high activity '+ Protein control' and low '+ Inhibitor Control'. Black circles and squares represent % Activity with No Inhibitor and Inhibitor, respectively.

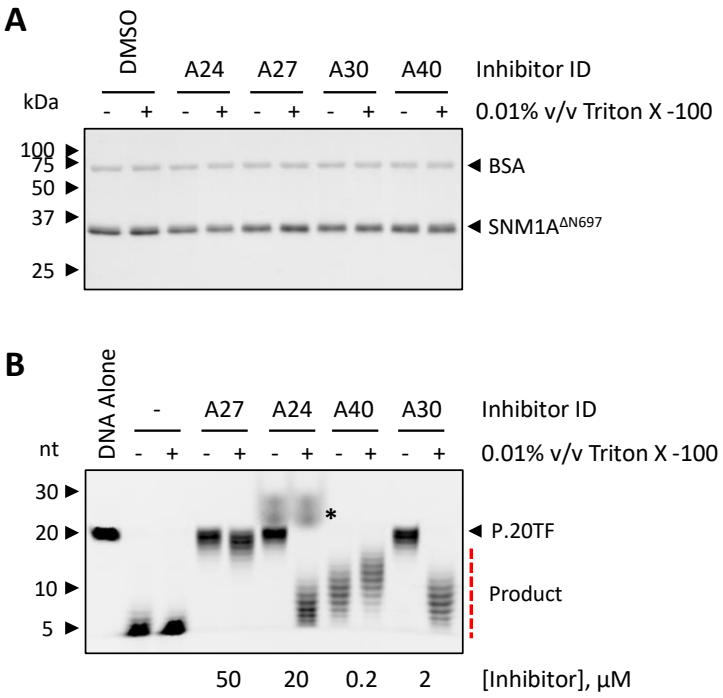


Figure 3.6. SNM1A bioactive inhibitors A24 and A30 likely function through non-specific colloidal aggregation.

Analysis of bioactive inhibitors identified in **Table 3.2** as potential non-specific aggregators. **[A]** Assessment of inhibitor-induced SNM1A precipitation. SNM1A (1.5 µM) was incubated with indicated inhibitors (50 µM) in nuclease activity buffer with or without 0.01% v/v Triton X-100 for 20 minutes at room temperature. Samples were clarified by centrifugation, analyzed using 15% SDS PAGE and visualized by GelDoc EZ Imager. **[B]** Rescue assay of SNM1A inhibition with Triton X-100. SNM1A (2.5 nM) was incubated with inhibitors (indicated concentrations), with or without 0.01% v/v Triton X-100 at room temperature for 10 minutes. Reactions were initiated by the addition of **P.20TF** (100 nM) and incubated at 37 °C for 15 minutes. Products were analyzed by 22.5% denaturing PAGE and visualized using ChemiDoc MP Imager. * indicates signal from internal inhibitor fluorescence.

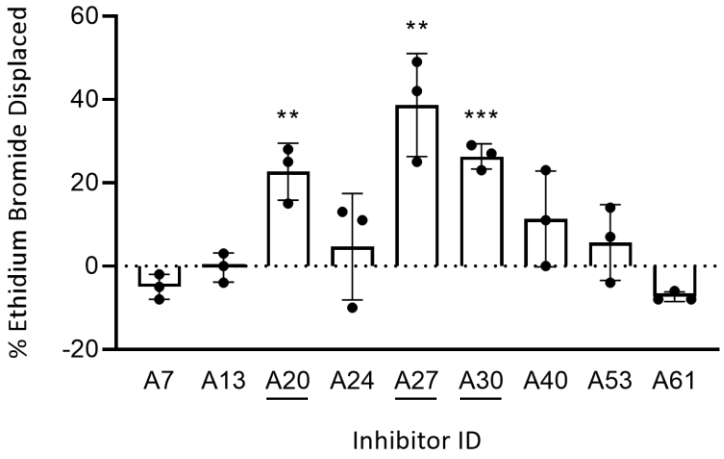


Figure 3.7. Three SNM1A bioactive inhibitors interact non-specifically with DNA.

Inhibitors (at exonuclease IC_{50} concentrations; summarized in **Figure 3.4**) were added to short double-stranded DNA (**P.DS**, 5 pmol), preincubated with ethidium bromide (15 pmol), and incubated at 37 °C for 10 minutes. Fluorescence was measured before and after inhibitor addition with Synergy H1 plate reader. Changes in fluorescence were expressed as '% EtBr Displacement'; normalized with No Inhibitor (0% displacement) and No DNA (100% 'displacement') controls. Error bars represent standard error of the mean from three independent experiments. ** denotes p-value < 0.01, *** < 0.001 using a one-tailed t-test compared with no inhibitor control. Underline indicates identified non-specific DNA interactors.

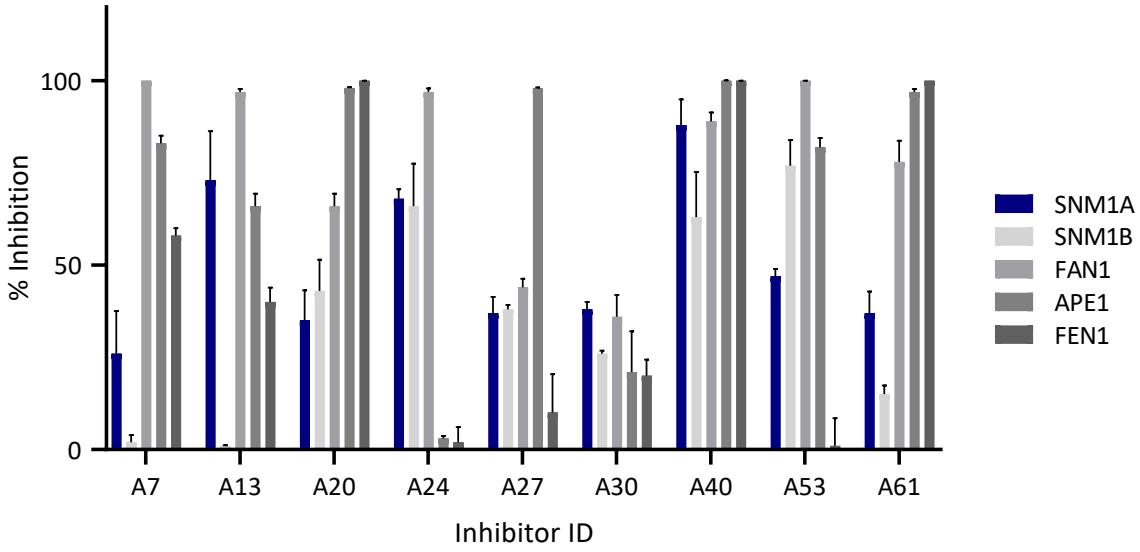


Figure 3.8. Validated bioactive inhibitors are not specific to SNM1A.

SNM1A (2.5 nM) was incubated with DMSO or inhibitors at double exonuclease IC_{50} concentrations (summarized in **Figure 3.4**) for 10 minutes at room temperature. Reactions were initiated with addition of **P.1F** (100 nM) and incubated at 37 °C for 15 minutes. $SNM1B^{1-335}$ (13 nM), $FAN1^{371-1017}$ (170 nM), and APE1 (4 mU, NEB) were incubated with **P.1F**, **P.(1)Flap** or **F.AB** (100 nM) for 20, 60 or 45 minutes, respectively at 37 °C. *Thermococcus* FEN1 (64 mU, NEB) was incubated with **F.(6)Flap** (100 nM) for 60 minutes at 45 °C. All products were analyzed by 22.5% denaturing PAGE and visualized with ChemiDoc MP Imager. Reaction progress was calculated by band integration using ImageLab. '% Inhibition' was determined following normalization with 'No Inhibitor' (low) and 'No Protein' (high) controls. Error bars reflect standard deviation of the mean from three independent replicates.

3.5.2 In silico identification of SNM1A-specific inhibitors

To improve inhibitor specificity, we aimed to exploit the similarities and differences between SNM1A (blue) and its close homolog SNM1B (grey). Alignment of the two nuclease domains (1.52Å RMSD) reveals the extensive global (**Figure 3.9A**) and active site (**Figure 3.9B**) conservation. Despite this structural conservation, limited sequence identity (~35%) and similarity (~50%) result in substantial changes to surface electrostatics (**Figure 3.9C-D**). These differences provide an opportunity to identify SNM1A-specific inhibitors *in silico* by cross-docking top predicted SNM1A binders to SNM1B. Enriching putative inhibitors with a significant affinity preference for SNM1A over SNM1B should increase the likelihood these compounds would be specific against more distinct nucleases.

To accomplish this, an *in silico* high-throughput screen was completed, docking 264,158 compounds (purchasable library from CMCB, McMaster University) to the active site of SNM1A (PDB: 5Q2A, H994 εN). Compounds (top 5000) predicted to have the highest binding affinity for SNM1A were then cross-docked to the active site of SNM1B (PDB: 5AHO, H276 εN) (**Figure 3.10A**). Screening was completed using AutoDock Vina, on the MCULE webserver. Compounds with a predicted binding score preference for SNM1A exceeding 2 kcal/mol (dotted line) were selected for *in vitro* validation (hits indicated in green, **Figure 3.10B**, **Appendix 3.A**).

Many residues adjacent to the active site contain solvent-exposed, flexible side chains. To prevent the crystal conformation of these side chains from biasing our docking results, potentially excluding viable compounds, these side chains were truncated (to reflect a high degree of flexibility; **Appendix 3.C**) and the screen was repeated. Unsurprisingly, these resulting hits demonstrated less predicted specificity for SNM1A over SNM1B. Any compound with a predicted preference for SNM1A exceeding 1 kcal/mol was selected for *in vitro* validation (**Appendix 3.B**).

Compounds (Series B; **Appendix 3.A** and **Appendix 3.B**) were tested for SNM1A inhibition in triplicate, using an exonuclease DNA substrate. All compounds showing

>50% inhibition at 50 μM and dose-dependent inhibition at the concentrations tested (10 or 50 μM) were considered successful inhibitors (hits indicated in green, target region in green shading; **Figure 3.11**). Notably, 9 of 74 (12%) compounds predicted *in silico* demonstrated significant SNM1A inhibition at low micromolar concentrations. This compares exceptionally well with our original *in vitro* HTS hit rate of 1.3%[210]. All nine validated SNM1A inhibitors were then tested for potential inhibition of SNM1B. Excitingly, four compounds (B23, B29, B68, and B74) demonstrated no inhibition of SNM1B at concentrations tested. Except for B58, all other compounds maintained a preference for SNM1A inhibition, with apparent IC_{50} values up to ~5-fold more potent against SNM1A than SNM1B (**Figure 3.12**).

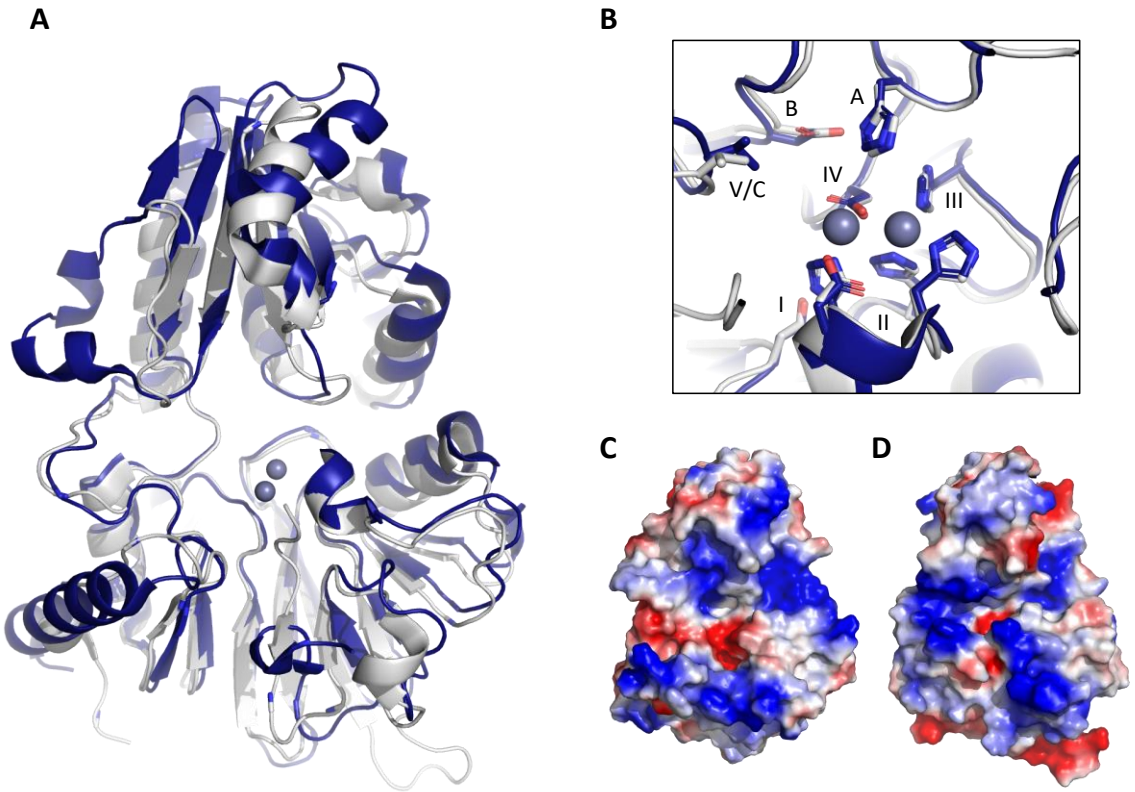


Figure 3.9. Specific inhibitors of SNM1A can be designed to exploit structural differences between SNM1A and SNM1B.

[A] Global structural alignment (1.52 Å RMSD) of SNM1A (blue, PDB: 5Q2A) and SNM1B (grey, PDB: 5AHO). **[B]** Zoom of aligned active site. Side chains of conserved motifs of SNM1 family shown (MβL motifs: I-V, β-CASP motifs: A-C). Electrostatic surface potential of **[C]** SNM1A (± 58.384 kT/e) and **[D]** SNM1B (± 61.733 kT/e).

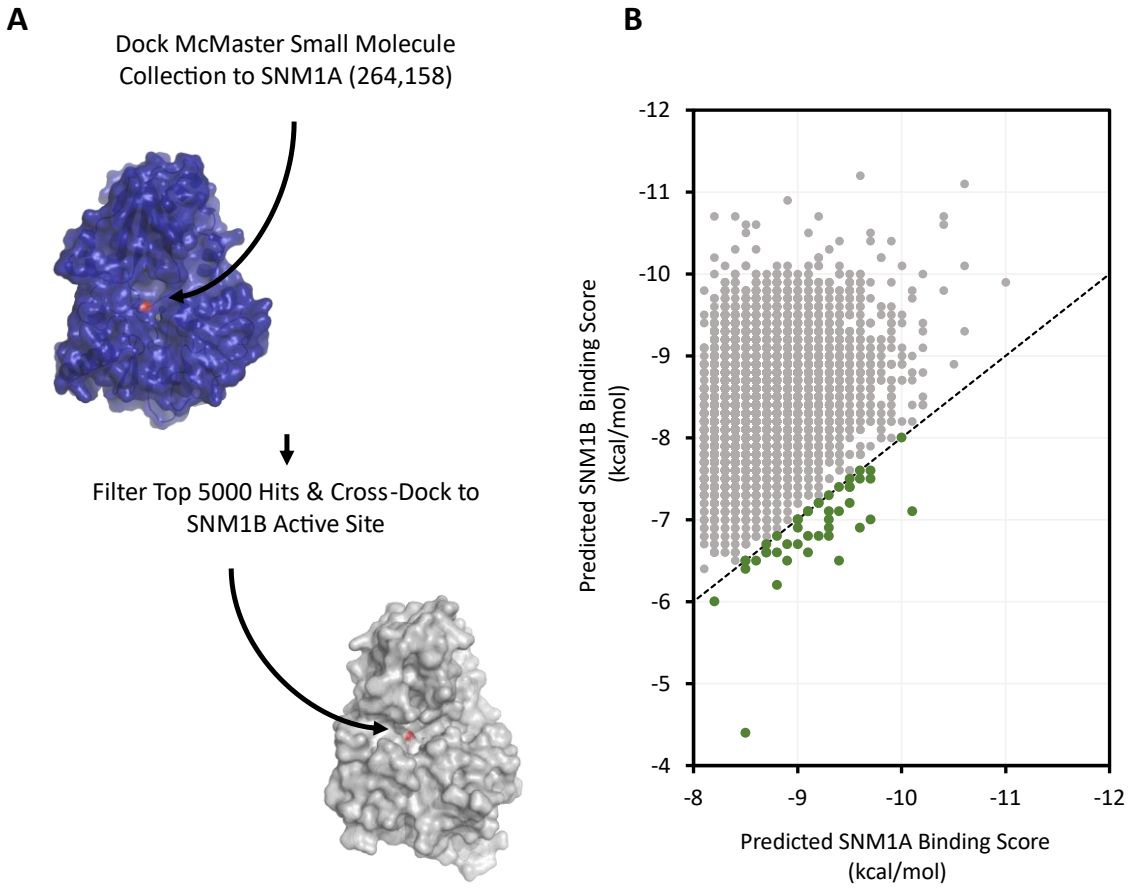


Figure 3.10. *In silico* screening for SNM1A-selective inhibitors.

[A] Schematic for *in silico* screen. A virtual library of purchasable compounds was filtered to remove problematic motifs. Remaining compounds were docked to the active site (indicated in red, H994 εN) of SNM1A (blue, PDB: 5Q2A) using AutoDock Vina. Compounds with the highest predicted affinity were then docked to the active site (indicated in red, H276 εN) of SNM1B (grey, PDB: 5AHO). **[B]** Results of SNM1A-SNM1B cross-dock plotted according to predicted binding scores (kcal/mol). Hit cut-off (SNM1A - SNM1B Binding Score ≥ 2.0) indicated with black dashed line. Hits are indicated by green dots.

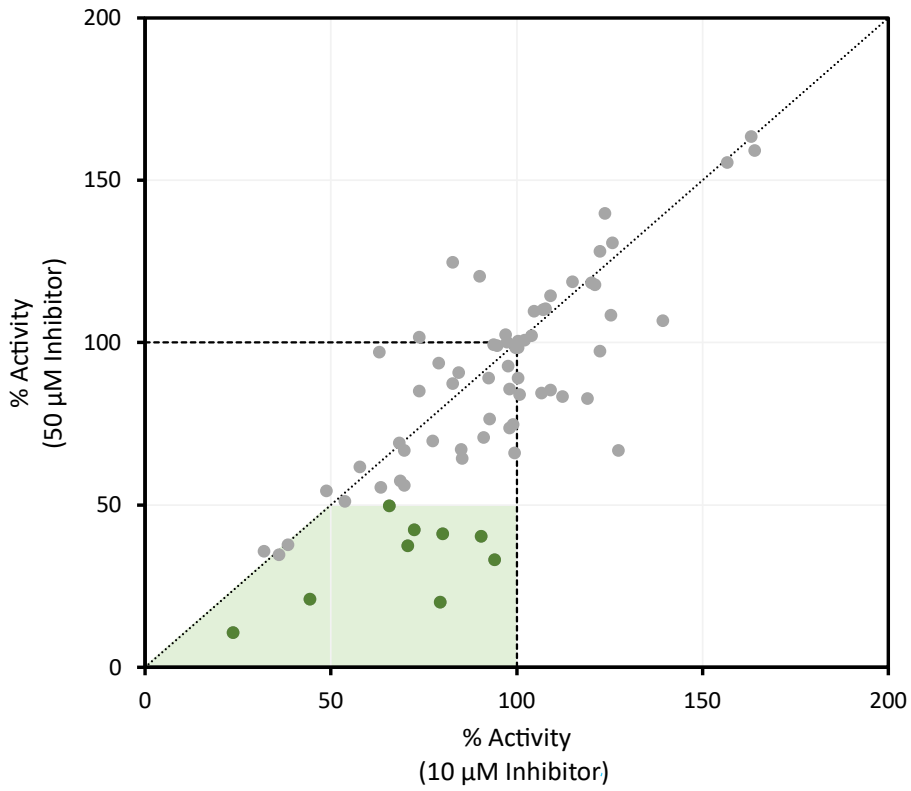


Figure 3.11. *In vitro* validation of SNM1A-specific inhibitors.

Compounds (Series B) identified as preferential SNM1A inhibitors with the *in silico screen* (**Appendix 3.A** and **Appendix 3.B**) were tested for inhibition of SNM1A exonuclease activity. SNM1A (1 nM) was incubated with each compound (10 or 50 µM) for 10 minutes at room temperature. Reactions were initiated with **P.1F** (100 nM) and incubated for 20 minutes at 37 °C. Reaction products were analyzed using 22.5% denaturing PAGE and visualized by ChemiDoc MP Imager. Reaction progression (% Activity) was determined following normalization of activity to No Compound (high) and DNA Only (low) controls. Validated inhibitors (green dots) were identified as compounds showing dose-dependent inhibition (below the dotted line), with activity reduced by 50% in the high condition (area shaded green). Each dot represents the mean of at least three independent replicates.

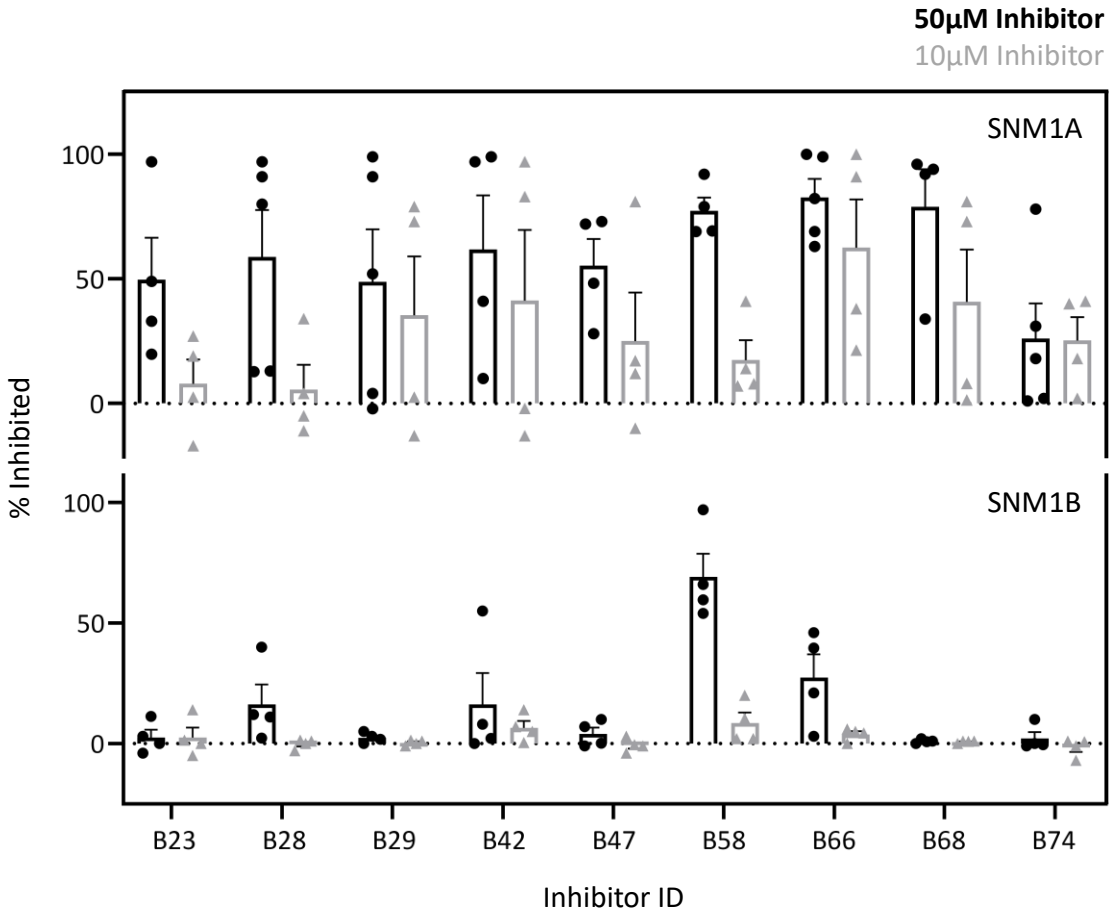


Figure 3.12. Cross-validation of SNM1A inhibitors with SNM1B.

Compounds which exhibited SNM1A dose-dependent inhibition in **Figure 3.11** were cross-tested with SNM1B. $SNM1B^{1-335}$ (9 nM) was incubated with compound (10 or 50 µM) for 10 minutes at room temperature. Reactions were initiated with **P.1F** (100 nM) exonuclease DNA and incubated at 37 °C for 45 minutes. Products were analyzed by 22.5% denaturing PAGE and visualized with the ChemiDoc MP Imager. Reaction progress was determined following normalization using DNA Only (high) and No Compound (low) controls. Data for SNM1A was generated as described in **Figure 3.11**. Error bars represent standard error of the mean from at least three independent replicates.

3.6 Discussion

Tumorigenesis is often associated with dysregulation of genetic surveillance, canonically tasked with slowing replication following DNA damage to prevent chromosomal instability[250]. Where many cancers continue through replication into mitosis despite damage, a healthy cell will stall to facilitate repair[247,248]. This distinction generates a therapeutic window, such that equivalent DNA damage preferentially induces genetic instability and mitotic catastrophe in cancerous cells. Similarly, small molecule inhibition of ICL repair factors can successfully sensitize cancers displaying innate or acquired chemoresistance. Past reports developing small molecule inhibitors targeting the FA-pathway[291–293] or XPF-ERCC1[270,271,274] have synergistically improved tumour response to ICL-inducing chemotherapy. As SNM1A function is critical for multiple, major ICL repair pathways, identifying potent and specific SNM1A inhibitors will likewise reduce cellular ICL repair capacity. To this end, the work presented here characterized SNM1A inhibitors identified by *in vitro* HTS (Series A) or *in silico* docking (Series B). Experiments elucidating the mechanism of inhibition informed which compounds represent promising lead scaffolds for future development. Further, we demonstrated that filtering initial hits from an *in silico* HTS, using a homolog for cross-docking, enriches the identification of SNM1A selective inhibitors.

3.6.1 Characterization of SNM1A inhibitors

A high-throughput screen (previously conducted by our lab) identified 9 compounds which inhibited SNM1A exo-/ and endo-nuclease activity (**Figure 3.1**)[210]. Inhibitor potency ranged from mid-nanomolar to mid-micromolar (**Figure 3.4**). None displayed specific inhibition against SNM1A, differentially inhibiting multiple human and non-human nucleases (**Figure 3.8**). This was particularly unfortunate, as the panel was selected to test closely (SNM1B) and distantly (FAN1, APE1, *Thermococcus* FEN1) related DNA repair nucleases. This lack of specificity was likely driven in A20, A27 and A30 by direct interaction with DNA (not SNM1A) and in A24 and A30 by colloidal aggregation.

As both mechanisms are independent of the nuclease itself, the lack of specificity is, in hindsight, not surprising and explains why some have been reported as hits from HTSs of other human nucleases[294,295]. The mechanism of inhibition of A7, A13, A40, A53 and A61 is not clear, and represent leads for further study. In particular, A53 and A61 were most drug-like and had IC_{50} values near 1 μM.

Lipinski's Rule of 5 outlines the criteria for compounds resembling drug-like molecules. Only one compound (A61) adhered to all criteria, though violations do not preclude further drug development. The more a compound violates these guidelines, the less likely it is to be clinically useful[296]. Most inhibitors had fewer than two violations, and only one (A40) had three. Ligand efficiency (measuring ΔG/atom) is often used for tracking the impact of rationally designed modifications to a compound through structure-activity-relationship (SAR) studies[297,298]. Efficiencies presented in **Table 3.1** represent the starting point for future improvement. Finally, many inhibitors were predicted to: (i) have limited oral availability (A7, A13, A20, A30, A40, A53), (ii) interact with multi-drug efflux pumps, like P-glycoprotein (A30, A40, A53) and (iii) inhibit cytochrome P450 liver enzymes (A13, A20, A24, A40). These issues would need to be addressed and/or limit the development of these inhibitors into clinically useful drugs.

Compounds which inhibited SNM1A via non-specific mechanisms likely represent poor leads for the development of therapeutically relevant compounds or molecular probes *in vivo*, precluding A20, A24, A27 and A30 from further structural and activity-relationship studies. The remaining 5 compounds identified in the *in vitro* HTS represent putative lead compounds to be developed further. Each of these compounds have been implicated in other cellular processes and have been examined in different contexts. Future *in vivo* work examining these (or improved) inhibitors would need to be conducted with these implications in mind.

Each of A7 (EGCG), A40 (TFDG) and A61 (m-DHPG) are common polyphenolic compounds produced by a broad range of plant species. Each consist of constituent moieties which are associated with diverse biological outcomes, including anti-

microbial/viral and antioxidant activity, along with altered gene expression, kinase signalling, protein homeostasis, cell growth and migration[299–301]. Specifically, m-DHPG is comprised of a core pyrogallin, modified with a methyl ester carboxylic acid. EGCG, instead, is composed of a pyrogallol (1,2,3-trihydroxybenzene) substituted epicatechin core, esterified with gallic acid (3,4,5-trihydroxybenzoic acid). Interestingly, TFDG is a chimera of each consisting of two equivalents of gallic acid decorating a di-catechin substituted pyrogallin core. These common elements suggest similar modes of inhibition. The chemical specificity observed in the HTS demonstrates that substitution and stereoselectivity critically influence inhibition. For example, while TFDG was an extremely potent inhibitor, close analogues in the HTS differing only by gallic acid substitution (Theaflavin alone or as a monogallate) failed entirely to inhibit SNM1A. Similarly, mid-micromolar concentrations of EGCG inhibited SNM1A while 10 stereo- or constitutional isomers (with or without an esterified gallic acid) did not.

EGCG is a major component of green tea extract, associated with broad therapeutic benefits[301]. Multiple studies have reported that concurrent treatment of cancer cells (in culture or xenograft) with EGCG and cisplatin sensitizes cytotoxicity. This sensitization is apparently provoked through modulation of multiple biological pathways, including drug import upregulation[302], increased autophagy[303], and DNA repair inhibition[276]. Interestingly, an *in vitro* HTS of XPF-ERCC1 identified EGCG (and a close analogue) acting as a potent endonuclease inhibitor (IC_{50} ~20 nM). Patrick and colleagues demonstrated that EGCG treatment, in conjunction with cisplatin, directly disrupted NER and ICL repair in lung cancer cells. Further, they showed a poly-acetylated EGCG maintained its sensitization in a mouse xenograft model[276]. Work by our lab has likewise identified increased cytotoxicity in HeLa cells following co-administration of EGCG with a sub-lethal dose of cisplatin[179]. The relatively low affinity of EGCG for SNM1A indicates these cellular effects are not likely mediated by SNM1A inhibition.

Amentoflavone (A13) belongs to a different plant metabolite family, also generally associated with a range of traditional therapeutic applications[304]. Amentoflavone is a biflavonoid, comprised of two flavonoid monomers, apigenin and

7,4'-dihydroxyflavone. Numerous flavonoids have been implicated in inhibition of other human nucleases such as APE1[294] and FEN1[295]. Importantly, SNM1A inhibition by amentoflavone was structure specific, as 36 other (bi)flavonoid compounds within the bioactive library had no impact on activity, including constitutional isomers and the amentoflavone component flavonoids themselves.

DIDS (A53) belongs to a class of stilbene compounds which inhibit anion exchangers (primarily chloride-bicarbonate antiporters)[305] and alkylate primary amine-/ thiol-containing proteins[306]. Though the mechanism of inhibition was not determined, the role of the sulfate substitutions appear critical. The bioactive library contained 18 compounds with an identical trans-stilbene scaffold, including multiple with a mono-sulfonic acid substituent. Despite the structural diversity, only DIDS resulted in inhibition of SNM1A. Inhibition was not likely mediated by direct alkylation (as was reported with DIDS inhibition of Caspase-3[306]), as Michaelis-Menten kinetic analysis from our lab suggested a competitive mechanism[218]. Depending on conformation, the inter-sulfate distance resembles that of adjacent DNA phosphates (~6Å), likewise suggesting direct substrate competition.

The bioactive library used in the initial *in vitro* HTS is composed of common off-patent FDA drugs and drug-like small molecules curated to optimize the 'hit to therapeutic' pipeline. Reflecting this, most compounds within the library are expected to have multiple cellular effects. The larger collection of compounds used in the *in silico* HTS increase both chemical diversity and novelty. As such, significantly less is known about the cellular bioavailability, toxicity and anticipated off-target effects of Series B SNM1A inhibitors (**Appendix 3.D**)[289,307]. The specificity assay reported above (**Figure 3.12**) identified 4 compounds (B23, B29, B68 and B74) which dose-dependently inhibited SNM1A while having no impact on SNM1B nuclease processing at 50 µM. While lacking common motifs, each inhibitor is generally structured as an elongated chain connecting small heterocyclic rings. To understand how these motifs (really all SNM1A inhibitors) facilitate SNM1A inhibition, future structure-activity relationship studies are required. Quantitatively monitoring inhibition by libraries of inhibitor

fragments and analogues of each SNM1A inhibitor will inform which chemical structures are critical for inhibition. Further, structure determination of SNM1A by X-ray crystallography has been previously reported[178,225,226]. Co-crystallization of each inhibitor (Series A and B) with the nuclease domain of SNM1A will guide attempts to improve potency and specificity.

3.6.2 Consideration for future inhibitor development

High throughput screens of SNM1A, whether *in vitro* or *in silico*, require assumptions surrounding the native catalytic mechanism. For instance, *in vitro* SNM1A nuclease reactions depend, in part, on the metal constitution of the reaction buffer. The original high-throughput SNM1A screen was completed in the presence of 10 mM $MgCl_2$[210]. As discussed in **Chapter 2**, the metal status of recombinantly purified SNM1A may artifactually alter nuclease processing. The *in silico* screen discussed above relied on a structural model of SNM1A, where the active site was coordinated by two Zn-ions. Each of these assumptions should be carefully considered in future experiments designed to identify biologically relevant inhibitors of SNM1A.

Metal titration studies reported in **Figure 2.15** suggest that the core MβL of SNM1A coordinates two Zn^{2+} ions in its active state. Further, the exonuclease activity relied on an additional metal to maximize catalytic efficiency, best enabled by Mg^{2+}. This functional insight is in contrast with reported SNM1A structures coordinating a single divalent metal[178,225,226], likely reflecting a differential in binding affinity between the M1 and M2 binding sites. We have found that even when soaked in high concentrations of Zn^{2+} or Mg^{2+}, the SNM1A M2 site remains unoccupied (data not shown). This suggests that SNM1A nuclease activity may be limited by the DNA substrate binding. Residues positioned to coordinate an M2 site (D736 and H737) require a small conformational change (from the unbound state) to coordinate the second Zn ion. Through an induced fit, DNA substrate binding may stimulate such a conformational change, thereby increasing M2 affinity and driving the generation of the transition state. This theoretical

mechanism of nuclease regulation has been reported in a diverse set of nucleases, and serves as a licence for DNA cleavage[308].

As predictive docking software improves, incorporating machine learning and artificial intelligence, *in silico* compound screens remain limited by the accuracy of the target enzyme structure. Further characterization of this proposed reaction mechanism would inform whether the target structure of a future inhibitor screen should resemble the one- or two-metal bound state of SNM1A. If M2 affinity is driven by substrate binding, a competitive inhibitor should be designed against a one-metal bound SNM1A structure. In principle, this would target the SNM1A 'ground state', preventing DNA-induced activation during repair. By contrast, if the native state of SNM1A is a two-metal coordinate complex, then in silico screens must reflect this. Deconvoluting the mechanism of SNM1A metal catalysis and activation will be essential for designing an effective inhibitor *in vivo*

Broad inhibition of ICL repair represents a promising target for improving the efficacy of chemotherapeutics which rely on the formation of ICL damage to induce cytotoxicity. Further, where tumour resistance is driven by increased capacity of ICL repair, inhibition can promote sensitization. Work presented in this chapter identified and developed small molecule inhibitors of SNM1A, exploiting its critical role in both major ICL repair pathways. Future optimization of each chemically distinct lead compound will enable improvements in potency and specificity. Finally, *in vivo* validation will be required to demonstrate that SNM1A inhibitors act on target in a cellular context and to quantify the impact of SNM1A inhibition on overall ICL repair capacity.

3.7 Appendix

Inhibitor ID	ZINC ID	InChI Key	Predicted SNM1A Docking Score	Predicted SNM1B Docking Score	Difference
B52	ZINC000016134268	VTVRVLXGDMBGTQ-MEKGRNQZSA-N	-8.5	-4.4	-4.1
B66	ZINC000006188202	ZYZTUDAPINZSSE-UHFFFAOYSA-N	-9.6	-6.8	-2.8
B72	ZINC000012997364	VAECNKKGBICBTF-UHFFFAOYSA-N	-9.6	-6.9	-2.7
B59	ZINC000002757117	MITZUAAAUUXRPM-UHFFFAOYSA-N	-9.3	-6.9	-2.4
B64	ZINC000001036755	NIAWRJJDBZPOSG-UHFFFAOYSA-N	-9.4	-7.1	-2.3
B53	ZINC000001214905	GBOQIVVTDJBANC-UHFFFAOYSA-N	-8.8	-6.6	-2.2
B54	ZINC000004893137	QTVYYJSLPKBNMK-SSGKUCQKSA-N	-9.3	-7.1	-2.2
B70	ZINC000008344198	LHQLIGUAACKXPS-UHFFFAOYSA-N	-8.8	-6.6	-2.2
B71	ZINC000023968440	ZEKYRUUODQOVPV-LJQANCHMSA-N	-8.9	-6.7	-2.2
B56	ZINC000001136614	ZQSBWBGTQTZCAK-UHFFFAOYSA-N	-9.2	-7.1	-2.1
B57	ZINC000000450127	SAWWNUZFWJQRKX-UHFFFAOYSA-N	-9	-6.9	-2.1
B61	ZINC000003157720	ZIDXTXSTXUAQOB-UHFFFAOYSA-N	-8.6	-6.5	-2.1
B68	ZINC000009508648	XECNZTGLDNUQRW-UHFFFAOYSA-N	-9.7	-7.6	-2.1
B74	ZINC000025323570	RLOFWASZIOZWLP-UHFFFAOYSA-N	-9.5	-7.4	-2.1
B55	ZINC000002086651	YGNYZXZNAITSLG-UHFFFAOYSA-N	-8.7	-6.7	-2
B58	ZINC000019851406	LGKSYAXWCXQISG-UHFFFAOYSA-N	-9.6	-7.6	-2
B60	ZINC000001297626	CTIWTDACUSIQMK-UHFFFAOYSA-N	-9.9	-7.9	-2
B62	ZINC000008618192	AGWPIHNMFIMIPO-UHFFFAOYSA-N	-10	-8	-2
B63	ZINC000008636629	CPKHXDWRUCPNSY-UHFFFAOYSA-N	-10.1	-8.1	-2
B65	ZINC000000756538	UGRPKUFZHQKDCT-VWLOTQADSA-N	-9.3	-7.3	-2
B67	ZINC000002633162	FUPBHBDTCKTADZ-UHFFFAOYSA-N	-8.8	-6.8	-2
B69	ZINC000007737074	QIOALFZVDWZFNO-UHFFFAOYSA-N	-9.4	-7.4	-2
B73	ZINC000025323561	FKYAOJMYIIPPPL-UHFFFAOYSA-N	-9.6	-7.6	-2

Appendix 3.A. Predicted SNM1A-specific inhibitors.

List of compounds (Series B) predicted to preferentially bind SNM1A (hits from **Figure 3.10**). Out of stock compounds were excluded from table.

Inhibitor ID	ZINC ID	InChI Key	Predicted SNM1A Docking Score	Predicted SNM1B Docking Score	Difference
B51	ZINC000100089496	ZCGOMHNNNFPNMX-KYTRFIICSA-N	-9.1	-6.9	-2.2
B25	ZINC000004543115	HXZSVIVGZJULFH-UHFFFAOYSA-N	-9.1	-7.2	-1.9
B27	ZINC000001033840	DYKJRNUQGHFVSE-UHFFFAOYSA-N	-9	-7.2	-1.8
B12	ZINC000001000992	BROMLIUYQDBEEG-UHFFFAOYSA-N	-9.2	-7.6	-1.6
B24	ZINC000015932036	ITNBJMJWSHEMRC-UHFFFAOYSA-N	-9.5	-7.9	-1.6
B33	ZINC000015724303	HVPMCJCTPOMRKO-UHFFFAOYSA-N	-8.2	-6.6	-1.6
B38	ZINC000009217653	WIJOLPSZZAVITB-UHFFFAOYSA-N	-8	-6.4	-1.6
B49	ZINC000022069702/4	DASPVRWVRKDAEM-UHFFFAOYSA-N	-8.3	-6.8	-1.5
B6	ZINC000001229506	QWWLASKIRMRUDU-UHFFFAOYSA-N	-8.4	-6.9	-1.5
B8	ZINC000001166126	IPYCCEWROIAUHH-UHFFFAOYSA-N	-8.9	-7.4	-1.5
B41	ZINC000012880861	SLDILHGXKCAYPU-UHFFFAOYSA-N	-7.5	-6	-1.5
B19	ZINC000001164859	VRUKSEHJZPEHPZ-UHFFFAOYSA-N	-8.7	-7.2	-1.5
B40	ZINC000008747334	FMGWMGUMHNFSST-UHFFFAOYSA-N	-7.5	-6.1	-1.4
B46	ZINC000032824298	VDVSPQQUMDGZHT-UHFFFAOYSA-N	-8	-6.6	-1.4
B21	ZINC000019952457	JVSLUQILMQTIQA-UHFFFAOYSA-N	-8.1	-6.7	-1.4
B39	ZINC000003256948	NDGXZKNLVPBOFL-UHFFFAOYSA-N	-8.7	-7.3	-1.4
B10	ZINC000015229010	DPZPKXRGESRIMF-UHFFFAOYSA-N	-8.8	-7.5	-1.3
B17	ZINC000002946360	AMKUQBRKJHVYMN-UHFFFAOYSA-N	-8.4	-7.1	-1.3
B34	ZINC000020574566	HETLCTGYACENPA-UHFFFAOYSA-N	-8.8	-7.5	-1.3
B28	ZINC000003129022	ZNOOVZHUWBQVMJ-UHFFFAOYSA-N	-8.5	-7.2	-1.3
B29	ZINC000003045888	POFIZXNOWXSAPP-UHFFFAOYSA-N	-8.6	-7.3	-1.3
B31	ZINC000009597578	PTGPKCGTYDLBIN-UHFFFAOYSA-N	-8	-6.7	-1.3
B35	ZINC000021157461/5	HUMSFBAWWSJSMP-UHFFFAOYSA-N	-8.1	-6.8	-1.3
B42	ZINC000012903485/90	HQZIFQSPWJNSNV-UHFFFAOYSA-N	-7.8	-6.5	-1.3
B47	ZINC000008026593	MWEZVYLTMLNNMO-UHFFFAOYSA-N	-8.5	-7.2	-1.3
B5	ZINC000001231839	WGKLUJYDUOFYJC-UHFFFAOYSA-N	-8.7	-7.4	-1.3
B13	ZINC000010077399	POEUJFKRSJQBGQ-UHFFFAOYSA-N	-8.7	-7.4	-1.3
B15	ZINC000000315823	JBKHRQVVULIWRF-UHFFFAOYSA-N	-8.8	-7.6	-1.2
B23	ZINC000012868391	BIYOOQOSTQULGA-UHFFFAOYSA-N	-9.3	-8.1	-1.2
B9	ZINC000001145478	OQUGMTHNSGMLMR-UHFFFAOYSA-N	-8.9	-7.7	-1.2
B16	ZINC000001166461	PWMHWVUFPQIIOD-UHFFFAOYSA-N	-8.9	-7.7	-1.2
B26	ZINC000003126526	PWRLXBPVCQLMQH-UHFFFAOYSA-N	-8.4	-7.2	-1.2
B45	ZINC000012941609/14	WPDAIXYOCUWVPT-UHFFFAOYSA-N	-9	-7.8	-1.2
B3	ZINC000000677891	NJKMQAYNFHYEKB-UHFFFAOYSA-N	-9.5	-8.3	-1.2
B4	ZINC000002755341-4	BVCBDDCXBSAZHB-UHFFFAOYSA-N	-8.1	-6.9	-1.2
B11	ZINC000001165406	PTHOGHNYFPWGPY-UHFFFAOYSA-N	-8.7	-7.5	-1.2
B37	ZINC000008881553	RQLKHVIQFKMPON-UHFFFAOYSA-N	-8.1	-6.9	-1.2
B44	ZINC000012527867	SQXLSOLJYXNERR-UHFFFAOYSA-N	-8.6	-7.4	-1.2
B7	ZINC000004845278	FREOXDKIGUGACF-UHFFFAOYSA-N	-8.4	-7.3	-1.1
B14	ZINC000008756907	QMGYOGIIPJVHDF-UHFFFAOYSA-N	-8.8	-7.7	-1.1
B48	ZINC000008026599/603	KGPYPOLMSYGEOX-UHFFFAOYSA-N	-8.3	-7.2	-1.1
B1	ZINC000001209624	CUMULSDIMHJUPR-UHFFFAOYSA-N	-8.7	-7.6	-1.1
B2	ZINC000004833241	NRUMSQGCGTYIJX-UHFFFAOYSA-N	-8.6	-7.5	-1.1
B18	ZINC000001104854	PJDHZPDDEODATR-UHFFFAOYSA-N	-8.5	-7.4	-1.1
B20	ZINC000001149902	VMJLEGWNOZYHPX-UHFFFAOYSA-N	-8.5	-7.4	-1.1
B22	ZINC000002211560	ZQLRGLXUFGYWGU-UHFFFAOYSA-N	-8.2	-7.1	-1.1
B30	ZINC000020114369	JDQNUEOWXXAZHW-UHFFFAOYSA-N	-8	-6.9	-1.1
B32	ZINC000009117007	RJQZRFBZXQRRKS-UHFFFAOYSA-N	-8.5	-7.4	-1.1
B36	ZINC000006703381	PDJJNWDYWVBQGZ-UHFFFAOYSA-N	-8.1	-7	-1.1
B43	ZINC000012520828	WAKNQHVEFHBYOK-UHFFFAOYSA-N	-8.7	-7.6	-1.1
B50	ZINC000012541406-7	WIWWIIIPKLHPFF-UHFFFAOYSA-N	-8.6	-7.5	-1.1

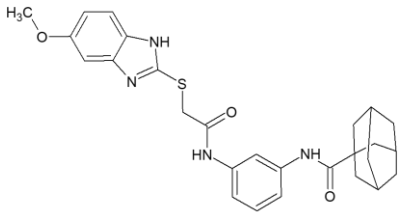
Appendix 3.B. Additional predicted SNM1A-specific inhibitors.

List of compounds (Series B) predicted to preferentially bind SNM1A identified in secondary in silico screen. Out of stock compounds were excluded from the table.

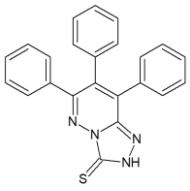
SNM1A (5Q2A)		SNM1B (5AHO)	
Residue	Truncated To	Residue	Truncated To
Y841	γ-carbon	L153	β-carbon
Y846	β-carbon	Y182	β-carbon
Y879	β-carbon	K186	γ-carbon
K883	γ-carbon	R259	β-carbon
R960	γ-carbon	Y273	γ-carbon
Y991	γ-carbon	D275	β-carbon
E993	β-carbon	R301	β-carbon

Appendix 3.C. Flexible residues truncated in secondary SNM1A *in silico* screen.

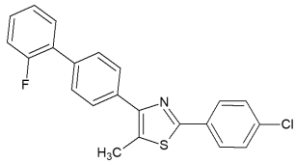
List of residues in SNM1A and SNM1B truncated at indicated carbon for secondary *in silico* SNM1A inhibitor screen.



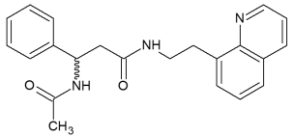
ZINC000012868391 (B23)



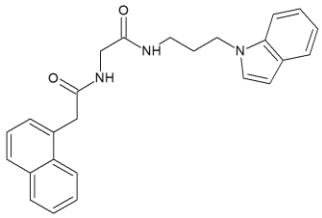
ZINC000003129022 (B28)



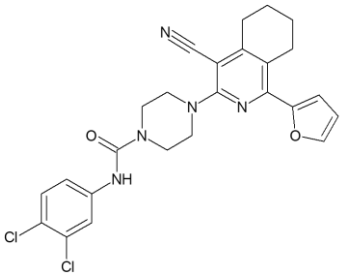
ZINC000003045888 (B29)



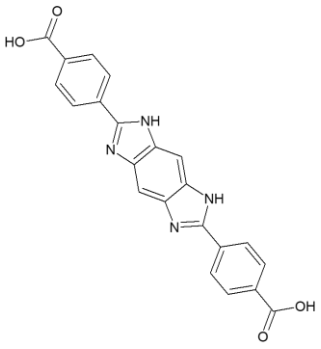
ZINC000012903485/90 (B42)



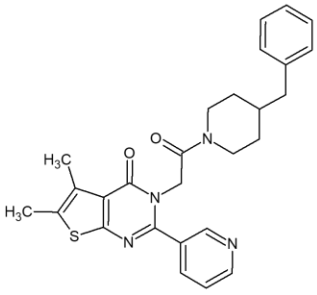
ZINC000008026593 (B47)



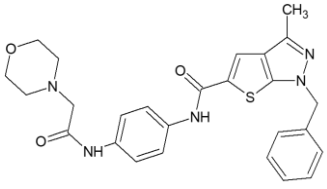
ZINC000019851406 (B58)



ZINC000006188202 (B66)



ZINC000009508648 (B68)



ZINC000025323570 (B74)

Appendix 3.D. Chemical structures of validated SNM1A inhibitors.

Putative inhibitors (Series B) from **Figure 3.12**.

Chapter 4
Interrogating the biological relevance of SNM1 dual nuclease function

4.1 Preface

Dr. Tyler Blue and Kristina Timcevska assisted in purification of select SNM1A mutants in Figure 4.3. Claire Zhang assisted with yeast survival experiments in Figure 4.14. I performed all other experiments presented below.

4.2 Purpose

The purpose of this chapter is to develop models of DNA-bound SNM1A and leverage these models to identify mutants that selectively diminish SNM1A exo- or endo-nuclease activity. Such separation-of-function mutants could then be used to determine if any biological relevance exists for both nuclease functions within ICL repair.

4.3 Introduction

SNM1A is a critical nuclease involved in the repair of interstrand crosslinks[186,208]. SNM1A disrupted cells demonstrate a specific survival hypersensitivity to ICL-inducing agents and accumulate gross chromosomal aberrations – radial chromosomes, in particular[129,208]. SNM1A has been shown to function directly in replication- and transcription-dependent repair, demonstrated via epistasis with the FA pathway[129,162,309] or CSB[190], respectively. While the exact intermediate(s) SNM1A acts on is(are) unclear, it primarily processes an unknown ICL unhooking intermediate to facilitate downstream homologous recombination. Biochemically, SNM1A is a dual function 5' phosphate dependent 5-3' exonuclease[175–177] and a single-strand dependent endonuclease[179]. SNM1A relies on a single binuclear Zn active site to accomplish both nuclease functions[166,167,178]. While active site mutants have been used to demonstrate that SNM1A nuclease activity is necessary for its role in ICL repair[176,189], whether this role is mediated by the exo-, endo- or both nuclease activities remains unclear.

Due to the robust exonuclease activity *in vitro*, SNM1A has traditionally been thought to participate in repair exclusively as an exonuclease[120,121,129,194]. SNM1A exonuclease activity can act translesionally, bypassing ICLs and other forms of base damage[121,129,177]. Correspondingly, SNM1A exonuclease activity could facilitate repair directly by completing unhooking initiated by a nick 5' of an ICL[120,121], or indirectly by trimming an unhooked oligonucleotide adduct to promote translesion synthesis[128,131]. In contrast, SNM1A endonuclease is 100- to 1000-fold weaker than its exonuclease

activity, depending on the substrate size, sequence and structure (see **Chapter 2**). Whether *in vitro* catalytic efficiency reflects biological function is unclear. For instance, Artemis (human homolog of SNM1A) demonstrates poor endonuclease activity *in vitro*, though its primary function in NHEJ and V(D)J recombination relies on endonuclease end-processing and hairpin opening. Artemis endonuclease activity is dramatically stimulated following phosphorylation by DNA-PKcs[220,222]. Analogously, SNM1A is phosphorylated by ATM[184], although the impact on nuclease activity is unknown. ATM modification (or another uncharacterized event) may function as a licence for SNM1A endonuclease activity *in vivo*. To define the role(s) of SNM1A, we must first establish the biological relevance of each nuclease function. Demonstrating a biological role of SNM1A exo- and/or endonuclease activity will help further define which intermediate(s) SNM1A may act on, thereby improving our understanding of SNM1A-dependent ICL repair.

SNM1A appears to function in poorly characterized, overlapping or redundant pathways with other nucleases including SNM1B[189], FAN1[154] and SAN1[162]. Correspondingly, disruption of SNM1A provokes a mild survival sensitivity following exposure to ICL-inducing agents in mammalian cells. In *S. cerevisiae*, Pso2 (yeast homolog of SNM1A) alone facilitates a more central role in ICL repair[310–312]. Since *S. cerevisiae* lacks additional SNM1 homologs[166,313], FAN1[165] and SAN1[162] nucleases, it makes a good model system to investigate SNM1A exo- and endo-nuclease contributions to ICL repair.

Similar to SNM1A, Pso2 is a dual function 5’ phosphate dependent 5-3’ exonuclease[314] and single-strand dependent endonuclease[179,180]. Although Pso2 exonuclease activity has a limited capacity to translesionally bypass ICLs alone[81,180], this activity is dramatically stimulated by Hrq1 (RecQ4-family helicase), both *in vitro* and *in vivo*[81,315]. In addition to processing extended regions of ssDNA, Pso2 endonuclease activity demonstrates a preference for opening paired hairpins[179,180]. Disruption of Pso2 generates severe sensitivity to interstrand crosslinking agents[310–312,316], along with elevated chromosomal aberrations and breaks[316–319]. Reminiscent of SNM1A[129,208], cells

lacking Pso2 generate unrepairable DSBs that are not recognized by canonical HR and NHEJ factors[319–322]. Importantly, exogenous expression of SNM1A (but not Apollo or Artemis) rescues both ICL hypersensitivity and DSB accumulation in Pso2 knockout yeast[176]. Thus, Pso2-dependent repair represents a useful system to evaluate the individual nuclease activities of SNM1A.

In this chapter, structural data of SNM1A (and its homologs) were used to generate models of SNM1A bound to simple exonuclease and endonuclease DNA substrates. These models were then used to identify residues for mutagenesis. Quantification of mutagenic induced changes to exonuclease and endonuclease activity facilitated the identification of separation-of-function mutants, while validating and improving DNA-bound models of SNM1A. To exploit the simpler yeast system, analogous separation-of-function mutants were extended into Pso2 and assessed for their capacity to rescue ICL survival hypersensitivity in knockout yeast. Failure of monofunctional exo- or endo-nuclease mutants to rescue the knockout suggests that Pso2 (and SNM1A, by extension) process two distinct intermediates within ICL repair, both as an exonuclease *and* an endonuclease.

4.4 Materials and Methods

4.4.1 DNA Substrates

All DNA oligonucleotides and substrates are summarized in **DNA Oligonucleotide Information** and **DNA Substrate Diagrams** in Supplemental Information, respectively. Sequences of all cloning and mutagenic primers used are detailed in **Primer Information**. All DNA oligonucleotide concentrations were determined by NanoDrop™ 2000c Spectrophotometer (Thermo Scientific). DNA substrates were prepared as described in **2.4.8**.

4.4.2 Structural Modeling

Models of DNA-bound SNM1A were generated using structural information of homologs bound to nucleic acid. The conserved active site of SNM1A$^{\Delta N697}$ (PDB: 5Q2A, stripped of water and non-metal ligands) was aligned to SNM1B (PDB: 7AF1) and a monomer of RNAse J (PDB: 4XWW) in Pymol™ v2.3.4. Coordinates of M2, along with scissile and terminal phosphate, were transposed to SNM1A. Multiple DNA conformations were generated for each substrate, limited by established chemical constriants[234,235].

4.4.3 Cloning

4.4.3.1 Bacterial Expression Constructs

Bacterial expression constructs of truncated Pso2$^{211\text{-}661}$ (also referred to as Pso2$^{\Delta 210}$) were generated using Gateway™ cloning (Invitrogen). Full-length Pso2 (Uniprot KB ID# P30620) template (10 ng) was truncated by PCR amplification using 100 ng of N210.PSO2-F and N210.PSO2-R primers with 2x Bestaq Mastermix (Diamed). N210.PSO2-F primer was designed to accommodate an N-terminal fusion, incorporating a flexible linker and TEV protease cleavage site. As described in **2.4.2.1**, the Pso2$^{211\text{-}661}$

amplicon was recombined into pDONR-201 (Invitrogen) followed by the pDEST-527 bacterial expression vector, containing an N-terminal His_6 fusion. Bacterial expression plasmid pDEST-527 was a gift from Dominic Esposito (Addgene plasmid #11518).

4.4.3.2 Yeast Expression Constructs

Yeast expression constructs were generated using Gateway™ cloning (Invitrogen). Full-length Pso2 template (10 ng) was amplified by PCR with 100 ng of FL.PSO2-F and FL.PSO2-R primers and 2x Bestaq Mastermix (Diamed). FL.PSO2-F primer was designed to incorporate a Shine-Dalgarno sequence surrounding the start codon. As described in **2.4.2.1**, the full length Pso2 amplicon was recombined within pDONR-201 (Invitrogen), followed by pYESDEST-52 (Invitrogen). Constructs were cloned lacking a stop codon, in frame with a C-terminal V5 epitope and His_6 tag, encoded on the plasmid backbone.

4.4.4 Mutagenesis

SNM1A and Pso2 mutants identified in **Figure 4.2** or **Figure 4.10**, respectively, were generated using site-direct mutagenesis (refer to section **2.4.11**). Mutagenic primer sequences used are outlined in **Primer Information**. Incorporation of all indicated mutations were confirmed by sequencing (DNA Sequencing Facility, RRI).

4.4.5 Purification

4.4.5.1 SNM1A

$SNM1A^{\Delta N697}$ and mutants were purified as described in section **2.4.4.1**.

4.4.5.2 Pso2

$Pso2^{211-661}$ pDEST-527 was transformed by heat shock into BL21 DE3 (Invitrogen) *E. coli* and plated overnight at 37 °C on LB agar containing 100 µg/mL ampicillin

selection. Liquid LB cultures of transformants under ampicillin selection were grown at 37 °C and 220 rpm overnight. Saturated cultures were used to inoculate fresh LB (1:100 v/v) under ampicillin selection, and grown at 37 °C and 220 rpm until OD_{600} 0.75. Cells were spiked with 5% v/v ethanol and allowed to cool to room temperature for 10 minutes prior to induction with 1 mM IPTG. Induced cultures were grown at 25 °C and 220 rpm for 12 hours. Cells were pelleted by centrifugation at 3,315 x g for 15 minutes, resuspended in bNiA Buffer (500 mM NaCl, 50 mM Sodium Phosphate pH 6.8, 10% glycerol and 0.5 mM TCEP), flash frozen in liquid nitrogen and stored at -80 °C until purification.

All Pso2 purification procedures were completed at 4 °C. Cell suspensions were thawed from -80 °C and diluted to at least 1:10 m/v (pellet/ buffer) with bNiA containing 0.01% v/v Triton X-100 and protease inhibitors (3 µM aprotinin, 1 µM pepstatin A, 1 mM benzamidine, 1 µM leupeptin and 1 mM PMSF). Cells were lysed by three passes at 10,000 psi by cell disruptor (Avestin Emulsiflex C3), clarified by centrifugation at 48,384 x g for 60 minutes and vacuum filtered (0.45 µm Triton-Free MCE, Millipore). Clarified lysate was loaded onto an EDTA-resistant Ni-Penta™ (Marvelgent) column using an Econo Pump (EP-1, BioRad), washed with bNiA containing 15 mM imidazole and eluted with bNiB (bNiA with 300 mM imidazole). Pso2-containing eluent was mixed with concentrated TEV protease (1:3 m/m) and sequentially dialyzed (1:100 v/v) twice into 300 mM NaCl, 25 mM Sodium Phosphate pH 6.8 overnight.

Cleaved Pso2 was spiked with 5 mM imidazole and loaded onto a 1 mL nickel-chelating HisTrap HP (GE Healthcare) column. The column was sequentially washed with bNiA containing 5, 20, 35, 50 and 300 mM imidazole. Samples containing Pso2 were pooled, concentrated with a Vivaspin Turbo15 10 kDa MWCO Centricon (Sartorius) at 1000 x g to 38 µM (2 mg/mL) and sequentially dialyzed (1:100 v/v) twice into Pso2 Storage Buffer (300 mM NaCl, 25 mM Sodium Phosphate pH 6.8, 5% glycerol and 0.5 mM TCEP) overnight. Dialyzed Pso2 fractions were aliquoted into single use, thin-walled PCR tubes, flash frozen in liquid nitrogen and stored at -80 °C until needed.

4.4.6 Thermal Shift Analysis

Purified SNM1A (6.25 µM) or Pso2 (18 µM) was mixed individually in a 96-well plate with Sypro™ Orange Dye (Thermofisher) to a final volume of 20 µL. Thermal stability was monitored and quantified as described in **2.4.6**. Error bars represent standard error of the mean, as determined from three independent replicates.

4.4.7 Nuclease Assay

SNM1A (wildtype and mutants) were serially diluted 2-fold using Storage Buffer (200 mM NaCl, 10 mM Sodium Phosphate pH 6.8, 5% v/v glycerol, 0.5 mM TCEP). Indicated concentrations of SNM1A (5 pM – 135 nM for exonuclease or 5-500 nM for endonuclease substrates) in Nuclease Reaction Buffer (50 mM Tris-acetate pH 7.2, 75 mM potassium acetate, 10 mM $MgCl_2$, 1 mM DTT and 100 µg/mL BSA) were initiated with 100 nM of exonuclease (**P.1F**) or endonuclease (**F.20T**) DNA substrates. Reactions were incubated at 37 °C for 30 minutes and terminated with equivolume Formamide Buffer (95% formamide, 5 mM EDTA).

Pso2 (wildtype and mutants) were serially diluted 2-fold in Pso2 Storage Buffer (300 mM NaCl, 25 mM Sodium Phosphate pH 6.8, 5% glycerol, 0.5 mM TCEP). Indicated concentrations of Pso2 (0.5-66 nM for exonuclease or 1-132 nM for endonuclease substrates) in Pso2 Nuclease Reaction Buffer (10 mM Tris pH 7.9, 10 mM $MgCl_2$, 50 mM NaCl and 1 mM DTT) were incubated with 100 nM of exonuclease (**P.1F**) or endonuclease (**F.20T**) DNA substrates. Reactions were incubated at 30 °C for 30 minutes and terminated with equivolume of Formamide Buffer.

Nuclease products were analyzed using 22.5% denaturing PAGE and visualized by ChemiDoc MP Imager (BioRad). Reaction progression (% Activity) was determined using band integration in ImageLab (BioRad). Product accumulation was normalized against a ‘No Protein’ (low) control. Standard error of the mean was calculated from three independent experiments. Four-variable curve fitting using GraphPad PRISM v8.4.3 was

used to determine protein concentration required to process 50% of the DNA substrate, denoted $[\text{SNM1A or Pso2}]_{50\%}$. Fold-change in $[\text{Protein}]_{50\%}$ was calculated according to:

Equation 7

$$Fold\ Decrease\ in\ Activity = \frac{[Mutant]_{50\%}}{[Wildtype]_{50\%}}$$

4.4.8 Yeast Survival Assay

4.4.8.1 Transformation

Survival assays were completed using *Saccharomyces cerevisiae* strains obtained from the yeast deletion collection[323]; wildtype BY4741 (MATa *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*) and Pso2 knockout (BY4741 with *pso2::kanMX4*). Yeast peptone media containing 2% v/v glucose (YPD) was inoculated with a single colony of wildtype or knockout yeast and grown to saturation at 30 °C and 220 rpm. Fresh YPD media was inoculated with 1:20 of saturated overnight cultures and grown at 30 °C and 220 rpm for approximately 6 hours. Yeast were pelleted by centrifugation at 750 x g for 5 minutes and washed twice by resuspension in Yeast Wash Buffer (100 mM lithium acetate, 10 mM Tris pH 8 and 1 mM EDTA). Washed pellets were resuspended in 1% of initial volume with Yeast Wash Buffer, and aliquoted with 100 µg UltraPure™ Salmon Sperm DNA Solution (Thermofisher) and 1 µg plasmid DNA. Cell mixtures were incubated in a 30 °C water bath for 15 minutes, vortexed and diluted 10:1 with Yeast Wash Buffer containing 40% w/v PEG 400. Following secondary incubation at 30 °C for 30 minutes in a water bath, transformation mixtures were spiked with 8% v/v DMSO and incubated at 42 °C for 15 minutes. Cells were pelleted by centrifugation at 1900 x g for 4 minutes, the transformation mixture was partially decanted and remaining cell suspension was plated on uracil dropout agar with 2% v/v glucose. Transformants were grown at 30 °C for 2 days prior to multi-colony re-streak.

4.4.8.2 Cisplatin Survival Assay

Uracil dropout media containing 1% v/v raffinose was inoculated with multi-colony streaks of yeast transformed with individual pYESDEST-52 Pso2 or empty control constructs. Liquid cultures were grown at 30 °C and 220 rpm for 18 hours prior to induction. Expression was induced with 2% v/v galactose and incubated at 30 °C and 220 rpm for 2 (controls, WT, D252A/H253A, K446A and S609A) or 3 (K446A/S609A, T378F and V633F) hours. Cultures were split, washed with Phosphate Buffer Saline (PBS; 137 mM NaCl, 2.7 mM KCl, Na_2HPO_4, 1.8 mM KH_2PO_4), resuspended in fresh inducing media with or without indicated concentrations of cisplatin (1, 1.5, 2 mM), and incubated at 28 °C and 220 rpm for 90 minutes. Cells were then pelleted by centrifugation at 750 x g, washed twice by resuspension and serially diluted 5-fold in PBS. Dilutions were spread plated on uracil dropout agar containing 2% galactose and incubated at 30 °C for 3 days. Surviving cells per mL of culture was determined using colony counts from plated dilutions with ~30-300 colonies and normalized relative to survival measured on matched, undamaged controls.

4.4.9 Western Blot

Yeast transformed with individual Pso2 constructs or controls were collected following induction, described above in **4.4.8.2**. Yeast (1 mL) were pelleted by centrifugation at 2000 x g for 5 minutes and resuspended sequentially into 100 µL of water and equivolume 0.2 M NaOH. Cells were incubated at room temperature for 5 minutes, pelleted by centrifugation, resuspended in SDS Loading Buffer (100 mM Tris pH 6.8, 4% w/v SDS, 0.2% w/v bromophenol blue, 20% v/v glycerol and 200 mM β-mercaptoethanol) and boiled for 3 minutes. Once cooled, cell debris was pelleted at 15,000 x g for 10 minutes and supernatant used for lysate analysis. Yeast lysate (10 µL) was separated using 12% SDS-PAGE and visualized with 0.5% v/v TCE on a GelDoc EZ Imager (BioRad). The gel was transferred to an activated PVDF (Immobilin-P, Millpore) membrane and treated as described in **2.4.5**. Blots were probed using mouse IgG

monoclonal anti-V5 primary (1:2500 v/v, R960-25 Invitrogen) and rabbit anti-mouse IgG-HRP secondary (1:4000 v/v, A9044 Sigma) antibodies.

4.5 Results

4.5.1 Structure-based mutagenesis of SNM1A

The dual exo-/ endo-nuclease function of SNM1A relies on a single metal-dependent active site. Substitution of metal binding or catalytic residues in the active site abolishes all activity[177,179], significantly complicating identification of SNM1A separation-of-function mutants. In the absence of DNA-bound SNM1A structural data, we instead developed models using existing structural data and the substrate preferences characterized in **Chapter 2**. These models combined minimum substrate requirements with well-established chemical constraints of phosphoryl transfer reactions[234,235]. Multiple SNM1A residues were found to be positioned to potentially facilitate either exonuclease or endonuclease substrate engagement and were therefore selected as candidates for mutagenesis.

Structures of SNM1B (human homolog)[227] and RNase J (*D. radiodurans* homolog)[229] have previously been determined in complex with dAMP or 5' monophosphorylated polyU RNA, respectively. Both homologs share near complete conservation of active site residues, despite limited overall sequence similarity (<35% within the nuclease domain) with SNM1A. Global alignment (**Figure 4.1A**) of SNM1B-dAMP (gold, PDB: 7A1F) and RNase J-RNA (green, PDB: 4XWW) to SNM1A (grey, PDB: 5Q2A) revealed likely positioning of the terminal (exonuclease substrate) and scissile phosphate (both substrates). The alignment further allowed for positioning of both metals in the anticipated native dual-zinc active site within the SNM1A model (**Figure 4.1B**).

Relying on the conserved DNA binding channel identified in the RNase J crystal structure, a trinucleotide could confidently be modelled 'feeding' into the active site. The trinucleotide begins (5') with the 'scissile' phosphate in coordination with the active site zinc ions and binds along the incoming DNA binding groove (mimicking the product of a nuclease reaction). This interaction is expected to be conserved between exonuclease and endonuclease substrate binding, with substrate engagement differing

‘upstream’ of the active site. As one nucleotide spans ~6 Å (phosphate to phosphate), key residues involved in driving substrate differentiation were selected within 6Å of the scissile phosphate for exonuclease and between ~6-18Å for endonuclease substrates. Using these criteria, 28 candidate residues were selected for mutation individually or in combination (labelled and indicated in red; **Figure 4.2**).

The impact of each mutant on the exonuclease and endonuclease activity was monitored as a ratio describing the change in the concentration of SNM1A required to digest 50% of each DNA substrate relative to wildtype (**Figure 4.3**). Initial results shown in **Figure 4.3C** summarize the relative efficiency of SNM1A exonuclease and endonuclease activity of each mutant. Though most mutations did not have significant impacts on either nuclease function, two specific regions necessary for exo- or endonuclease activity were identified (an anticipated 5’ phosphate binding pocket and an extended DNA binding channel).

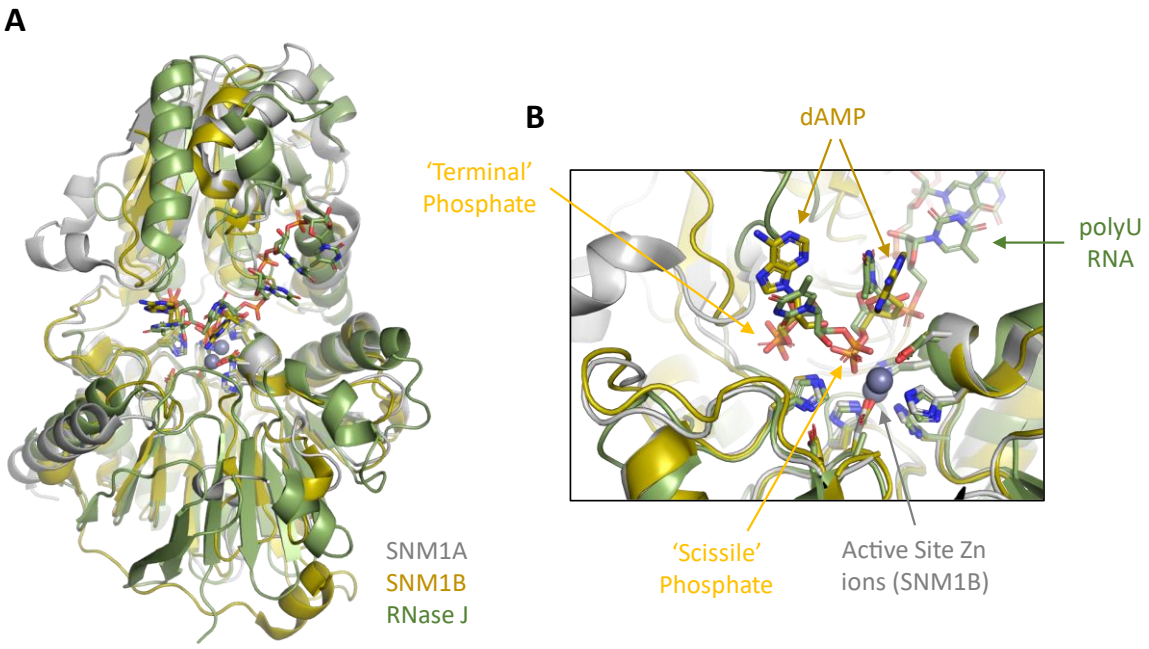


Figure 4.1. Alignment of SNM1A homologs bound to nucleic acid substrates.

[A] Global alignment of RNA-bound *D. radiodurans* RNAse J (3.14Å RMSD, PDB: 4XWW, green) and dAMP-bound human SNM1B (1.52Å RMSD, PDB: 7A1F, gold) to SNM1A (PDB: 5Q2A, grey). **[B]** Zoom of aligned active site (side chains of conserved MβL and β-CASP motifs shown). Position of terminal and scissile phosphate for SNM1A model indicated in orange.

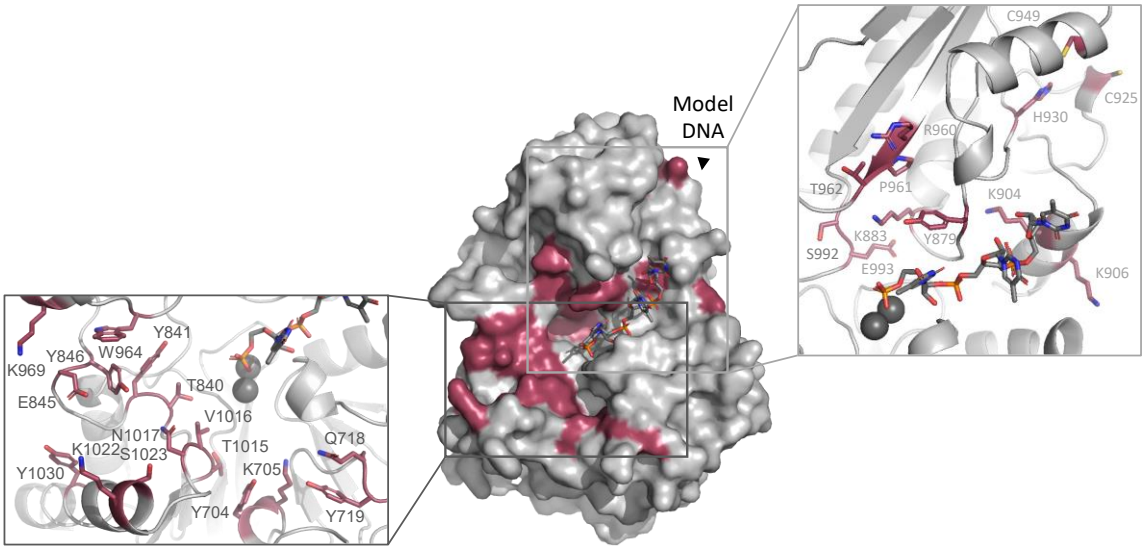


Figure 4.2. Overview of residues mutated for structure-function analysis of SNM1A nuclease activity.

Model of SNM1A nuclease domain (PDB: 5Q2A) bound to a short nucleic acid substrate modified from **Figure 4.1**. Residues of interest, mutated individually or in combinations, are labelled and coloured red.

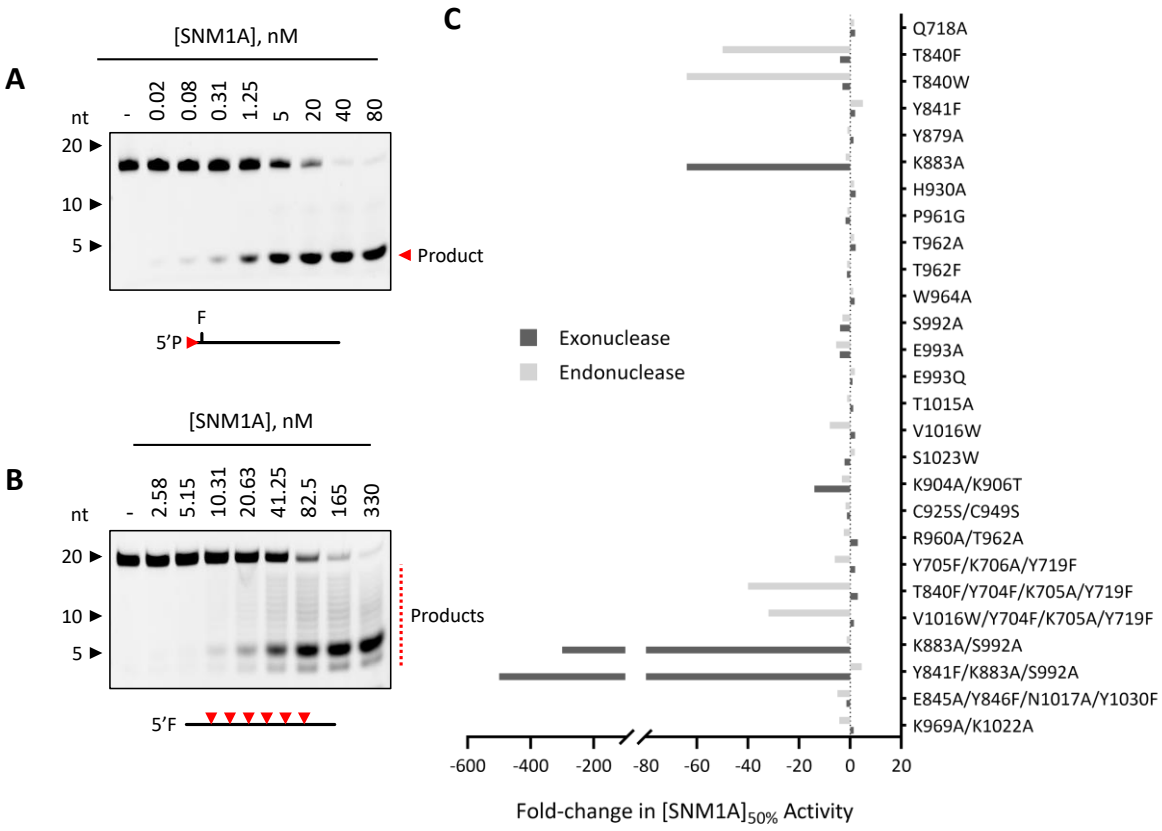


Figure 4.3. Initial analysis of SNM1A mutant nuclease activity.

Representative (SNM1A^{WT}) protein dose-dependent **[A]** exonuclease or **[B]** endonuclease activity assay. Indicated concentrations of SNM1A were incubated with **P.1F** or **F.20T** (100 nM) at 37 °C for 30 minutes, respectively. Products were analyzed with 22.5% denaturing PAGE and visualized by ChemiDoc MP Imager. **[C]** Graph of SNM1A mutant nuclease activity change (relative to wildtype). $[SNM1A]_{50\%}$ indicates the concentration of SNM1A required to digest 50% of the indicated substrate, under conditions tested. $[SNM1A]_{50\%}$ was determined using '% Substrate Remaining' in mutant protein dose-dependent assays. Fold-change represents a ratio of $[Mutant]_{50\%}$ over $[Wildtype]_{50\%}$. Data represents results from a single replicate.

4.5.2 Characterization of SNM1A separation-of-function mutants

4.5.2.1 Exonuclease separation-of-function

SNM1A exonuclease activity absolutely requires a terminal 5' phosphate[177]. Initial mutational analysis demonstrated that disruption of K883, S992 and Y841 abrogated exonuclease processing (**Figure 4.3C**). These three residues form a pocket, perfectly positioned to interact with the terminal phosphate, situated ~6Å from the scissile phosphate (**Figure 4.4** and **Figure 4.5A**). This model of exonuclease terminal phosphate binding agrees with structural evidence from SNM1B and RNase J. As shown in global alignments in **Figure 4.1,** the phosphate of a dAMP molecule in SNM1B and the 5' phosphate of the RNA bound to RNase J both interact with this well-conserved pocket.

Before quantifying the impact on SNM1A exonuclease activity, mutant purity (**Figure 4.5B**) and stability (**Figure 4.5C**) was assessed. Thermal stability (comparing the melting temperature of the mutants) was used to identify whether any mutant resulted in structural/ conformation changes large enough to disrupt global protein folding. All mutants exceeded 95% purity and maintained wildtype stability.

Analysis of nuclease reactions with SNM1A 'phosphate-binding pocket' mutants clearly demonstrated that the proposed phosphate interaction is critical for maintaining exonuclease function. Substitution of K883 severely limited exonuclease processing, while S992 and Y841 moderately reduced activity, under the conditions tested. Combination mutants (K883A/S992A and K883A/S992A/Y841F) further reduced exonuclease processing (**Figure 4.5D**), while remaining endonuclease competent. Unexpectedly, Y841 mutants provoked an increase in endonuclease processing, whether individually or in combination (**Figure 4.5E**). In all cases, significant separation-of-function was observed. To quantify separation-of-function, a protein-dose dependent nuclease assay with an exo- or endo-nuclease substrate was completed. Fitting sigmoidal curves to the resulting '% Activity' graphs (**Figure 4.6A-B**) established the concentration of each mutant required to digest 50% of substrate. Separation was

measured as the relative 'fold-decrease' in exonuclease activity, determined as the ratio of $[Mutant]_{50\%}$ over $[Wildtype]_{50\%}$ (**Figure 4.6C**). In line with the qualitative results in **Figure 4.5D-E**, combination mutants K883A/S992A and K883A/S992A/Y841F resulted in greater than 1000-fold reduction in exonuclease processing, effectively abolishing all exonuclease activity at relevant protein concentrations.

4.5.2.2 Endonuclease separation-of-function

Initial mutational data also demonstrated that substituting T840 or V1016 for bulkier amino acids selectively disrupted SNM1A endonuclease activity (**Figure 4.3C**). Characterization of SNM1A endonuclease activity in **Chapter 2** demonstrated that SNM1A engagement of endonuclease substrates requires a footprint of approximately 4-6 nucleotides (**Figure 2.11** and **Figure 2.13**). This implies that from the scissile phosphate (**Figure 4.7B**), at least three nucleotides must extend past the active site. Our working model limited the putative DNA binding groove to three possible channels (**Figure 4.7**), where the substrate bends up (red), down (green) or continues straight (blue). Despite testing many mutants targeting residues in these regions that were predicted to stabilize the sugar-phosphate backbone of the extended nucleotides, only amino acid substitutions that sterically blocked the junction between 'straight-through' (T840) and 'bent-down' (V1016) channels interfered with endonuclease processing. Attempts to disrupt electrostatic interactions were unable to significantly impact catalysis, thus sterically blocking binding appeared to be superior in reducing substrate affinity (**Figure 4.8**).

Prior to characterization, each mutant was analyzed for purity (**Figure 4.8B**) and probed for changes to global stability, as measured by thermal shift (**Figure 4.8C**). Analysis of the exo- and endo-nuclease activities of steric mutants on DNA substrates clearly showed selective disruption of endonuclease processing (**Figure 4.8D-E**). Unfortunately, combination steric mutants had limited solubility, preventing additive/ synergistic improvement of any observed separation. Quantitative analysis of changes in exonuclease (**Figure 4.9A**) and endonuclease (**Figure 4.9B**) activity was completed using

protein-dose dependent nuclease assays. Though less disruptive than the exonuclease deficient mutants, T840F/W and V1016W significantly abrogated endonuclease processing (**Figure 4.9C**). Notably, mutants at position 840 each generated slight (relative to the change in endonuclease activity) reduction in exonuclease processing.

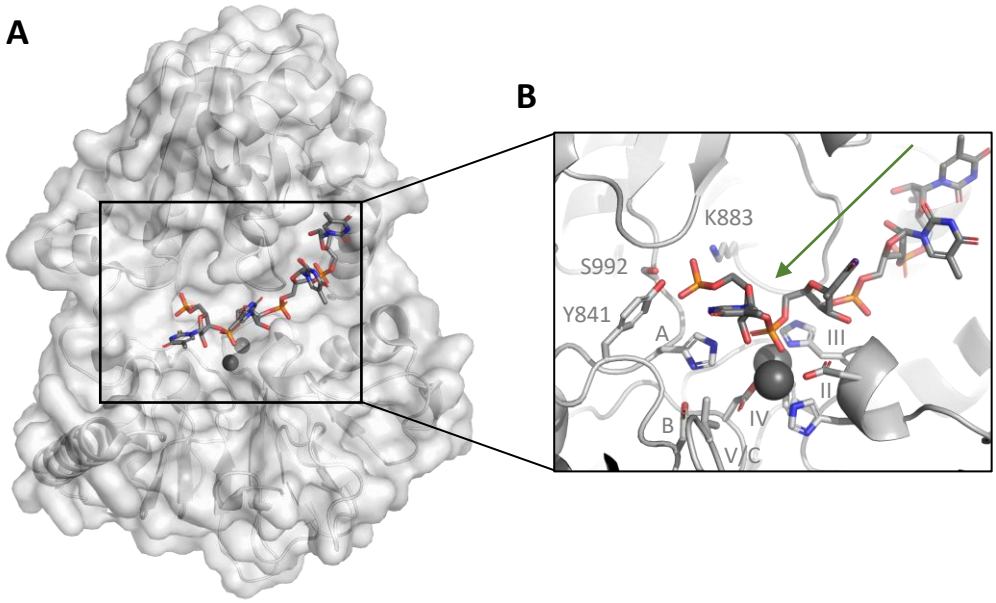


Figure 4.4. Model of SNM1A bound to an exonuclease DNA substrate.

[A] Model of modified surface and cartoon structure of SNM1A (PDB: 5Q2A) bound to a four poly dT exonuclease DNA substrate (using conservation from **Figure 4.1**). **[B]** Zoomed active site with conserved structural motifs (II-V, A-C) and potential interacting side chains (Y841, K883, S992) labelled. Green arrow indicates direction of (5-3') DNA movement between nuclease digestion. Motif II and the bound Zn ion were modified using SNM1B active site (PDB: 7A1F).

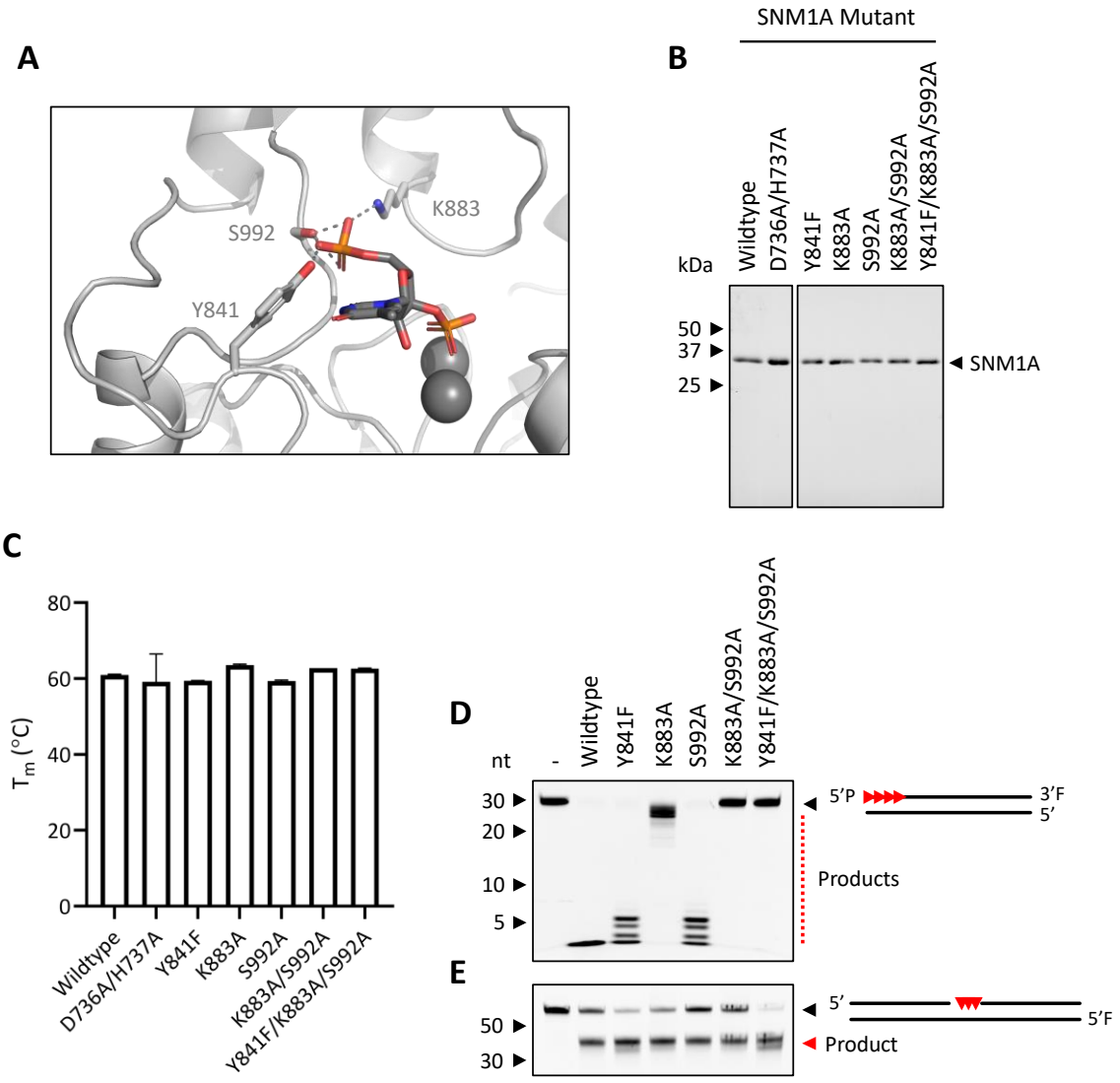
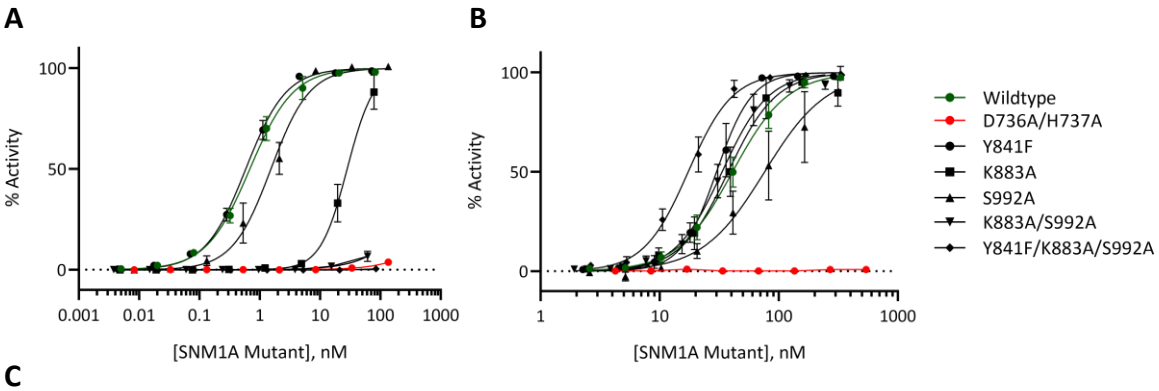


Figure 4.5. Targeting key interactions with the 5' phosphate disrupts SNM1A exonuclease activity.

[A] Simplified model of exonuclease substrate binding (**Figure 4.4**) with predicted terminal phosphate interactions labelled. **[B]** Normalized samples of purified (according to scheme in **Figure 2.5A**) SNM1A mutants analyzed by 15% SDS-PAGE. **[C]** Melting temperature of SNM1A mutants (6.25 µM). Fluorescence from Sypro Orange dye was monitored across 20-90 °C. T_m was determined by the maximal dF_{595}. Error bars represent standard error of the mean from two independent experiments. Indicated SNM1A mutants (25 or 100 nM) were incubated at 37 °C for 60 minutes with **[D] P.DS** or **[E] Gapped.DS** (100 nM), respectively. Products were analyzed using 22.5% denaturing PAGE and visualized by ChemiDoc MP Imager. Red arrows indicate expected initial cleavage products.



C

Mutant	Exonuclease (P.1F)		Endonuclease (F.20T)		**Adjusted Fold-Separation in Activity**
	$[SNM1A]_{50\%}$ (nM)	Fold-Decrease	$[SNM1A]_{50\%}$ (nM)	Fold-Decrease	
Wildtype	0.69 ± 0.11	1.00	40.94 ± 4.85	1.00	-
D736A/H737A	ND*	>1000	ND	>1000	-
Y841F	0.59 ± 0.06	0.85	30.04 ± 3.48	0.73	**1.17**
K883A	28.16 ± 4.78	41.08	37.64 ± 7.04	0.92	**44.68**
S992A	1.53 ± 0.38	2.24	76.15 ± 22.32	1.86	**1.20**
K883A/S992A	698.05 ± 356.21	1018.31	33.46 ± 4.02	0.82	**1245.95**
Y841F/K883A/S992A	ND	>1000	17.37 ± 1.80	0.42	**>1000**

Figure 4.6. Quantification of exonuclease deficiency in SNM1A phosphate-binding pocket mutants.

Quantification of mutant SNM1A nuclease activity. **[A]** SNM1A mutants (5 pM – 135 nM) were incubated with **P.1F** (100 nM) at 37 °C for 30 minutes. **[B]** SNM1A mutants (5-500 nM) were incubated with **F.20T** (100 nM) at 37 °C for 60 minutes. Reaction products were resolved by 22.5% denaturing PAGE and visualized by ChemiDoc MP Imager. Integration of product intensity was determined using ImageLab and '% Activity' was normalized with a 'No Protein' (low) control. Four-variable sigmoidal curve fitting in GraphPad PRISM v8.4.3 was used to determine mutant concentration required to process 50% of substrate. Error bars indicate standard error of the mean, determined from three independent experiments. **[C]** Calculation of nuclease deficiency in SNM1A mutants. 'Fold-decrease' in activity was determined as the ratio between [Mutant] over [WT] for each activity. *ND indicates there was insufficient activity for curve fitting.

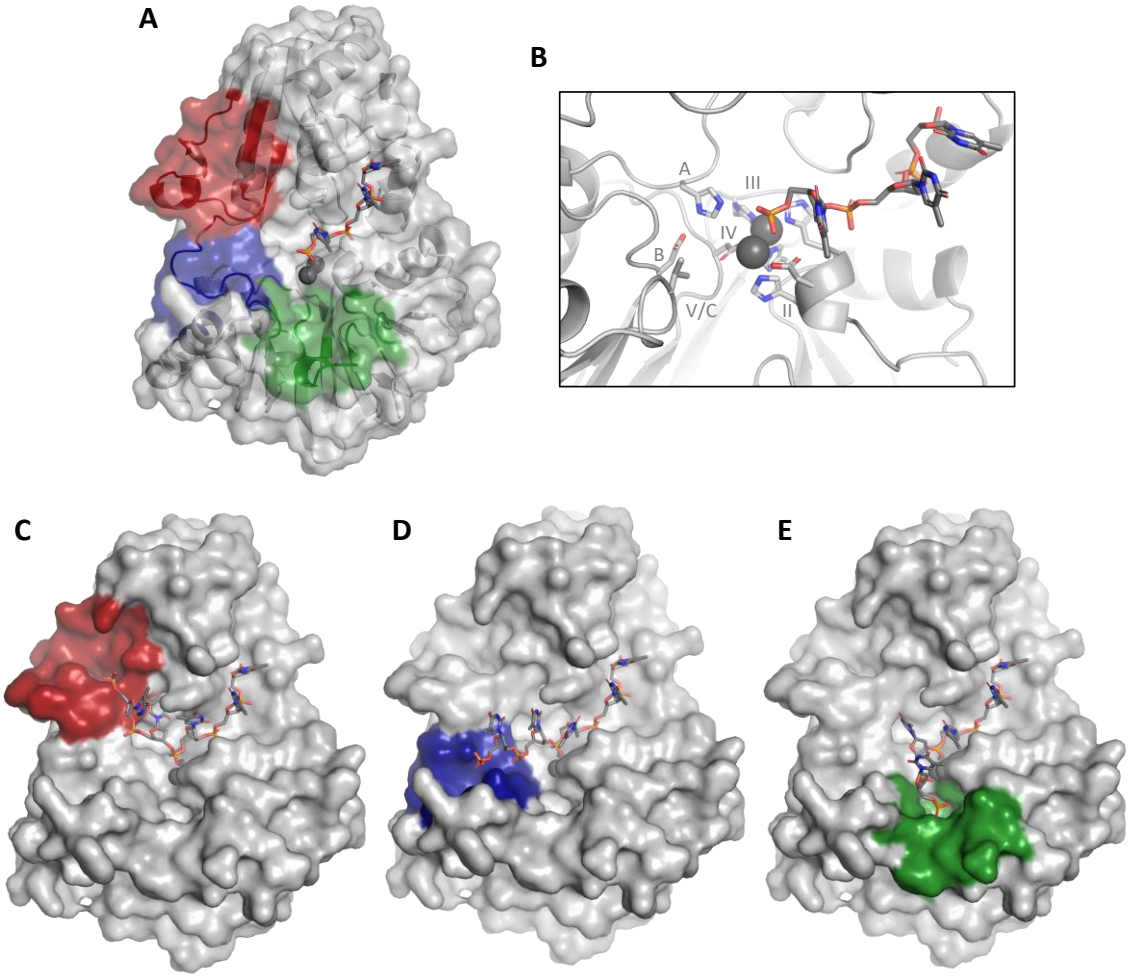


Figure 4.7. Models of SNM1A bound to an endonuclease DNA substrate.

[A] Modified surface and cartoon structure of SNM1A (PDB: 5Q2A) bound to a three poly dT DNA substrate (as shown in **Figure 4.2**). **[B]** Zoomed active site with conserved structural motifs (II-V, A-C) labelled. Scissile phosphate positioned directly above active site Zn ions. Potential DNA binding grooves modelled with 5' dinucleotide extension from the scissile phosphate in [B]. DNA modelled to **[C]** bend up (red), **[D]** continue straight (blue) or **[E]** bend down (green).

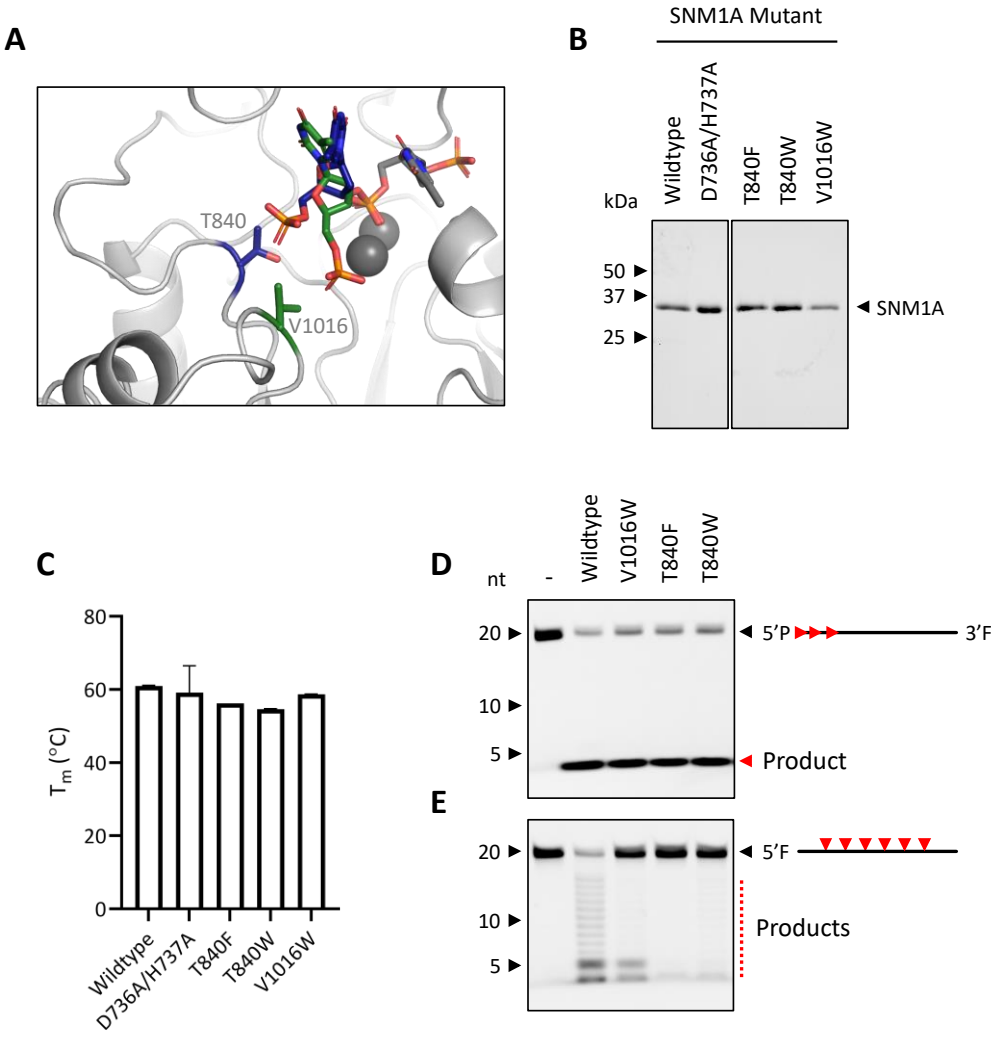
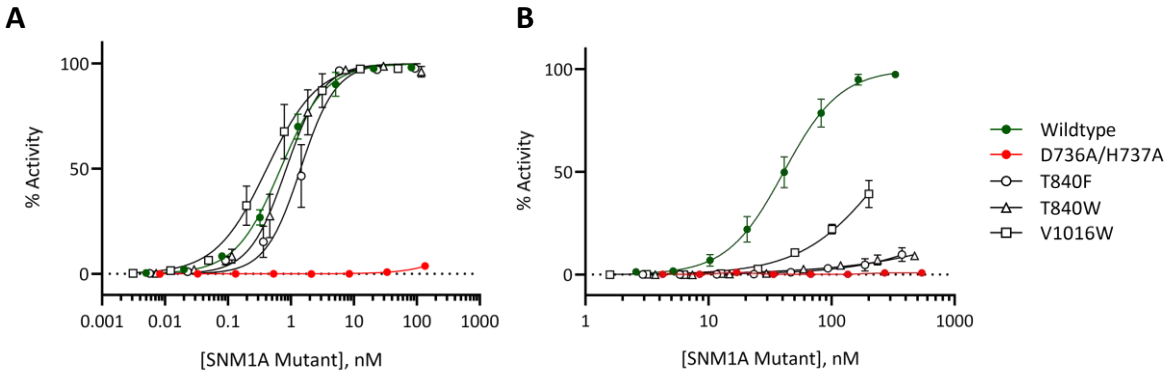


Figure 4.8. Steric disruption of predicted DNA binding groove diminishes SNM1A endonuclease activity.

[A] Superimposed model of endonuclease substrate binding (**Figure 4.7D-E**). **[B]** Normalized samples of purified (according to scheme in **Figure 2.5A**) SNM1A mutants analyzed by 15% SDS-PAGE. **[C]** Melting temperature of SNM1A mutants (6.25 µM). Error bars indicate standard error of the mean from two independent experiments. Indicated SNM1A mutants (10 or 100 nM) were incubated at 37 °C for 60 minutes with **[D] P.20TF** or **[E] F.20T** (100 nM), respectively. Products were analyzed using 22.5% denaturing PAGE and visualized by ChemiDoc MP Imager. Red arrows indicate expected cleavage pattern.



C

Mutant	Exonuclease (P.1F)		Endonuclease (F.20T)		**Adjusted Fold-Separation in Activity**
	$[SNM1A]_{50\%}$ (nM)	Fold-Decrease	$[SNM1A]_{50\%}$ (nM)	Fold-Decrease	
Wildtype	0.69 ± 0.11	1.00	40.94 ± 4.85	1.00	-
D736A/H737A	ND*	>100	ND	>100	-
T840F	2.79 ± 0.39	4.07	3295 ± 2261.50	80.48	**19.77**
T840W	1.79 ± 0.20	2.61	ND	>100	**>100**
V1016W	0.87 ± 0.14	1.27	277.30 ± 53.90	6.77	**5.34**

Figure 4.9. Quantification of endonuclease deficiency in SNM1A DNA binding mutants.

Quantification of mutant SNM1A nuclease activity. **[A]** SNM1A mutants (5 pM – 135 nM) were incubated with **P.1F** (100 nM) at 37°C for 30 minutes. **[B]** SNM1A mutants (5-500 nM) were incubated with **F.20T** (100 nM) at 37 °C for 60 minutes. Reaction products were analyzed by 22.5% denaturing PAGE and visualized by ChemiDoc MP Imager. Quantification was completed as in **Figure 4.6**. Error bars represent the standard error of the mean, derived from three independent experiments. **[C]** Calculation of nuclease deficiency in SNM1A mutants. 'Fold-Decrease' in activity was defined as the ratio between [Mutant] over [WT] for each activity. *ND indicates there was insufficient activity for curve fitting.

4.5.3 Probing SNM1A dual-nuclease biological relevance using a yeast model

To evaluate the biological significance of exo- and endo-nuclease activities on ICL repair *in vivo,* we turned to a yeast model. Like SNM1A, Pso2 is a dual-function nuclease that is critical for the repair of ICL damage[180]. SNM1A is the functional homolog of Pso2 and is able to rescue the ICL hypersensitivity of a Pso2 yeast knockout[176]. Since yeast lack the significant nuclease redundancy observed in mammalian ICL repair, the sensitivity associated with Pso2 knockout is enhanced in yeast making it ideal for testing possible contributions of exo- and endo-nuclease activities.

To avoid complications of heterologous complementation, SNM1A separation-of-function mutants were recapitulated in Pso2. Critically, all residues that separated nuclease functions in SNM1A were conserved in Pso2 (**Figure 4.10D**). In the absence of available structural information, a tertiary structure model of Pso2 was generated using AlphaFold v2.0 (**Figure 4.10A**). The model of Pso2 suggested that residues targeted for mutagenesis were also structurally conserved with SNM1A (**Figure 4.10B-C**)[195,324]. Using a modified schematic of the SNM1A purification (**Figure 4.11A;** stress induction & TB growth media), Pso2 (wildtype and mutants) were purified (**Figure 4.11B-C**) and tested for thermal stability (**Figure 4.11D**). Targeting these conserved Pso2 residues resulted in disruption of specific nuclease activities (**Figure 4.12**). While individual 5' phosphate-binding pocket disrupting mutants prompted limited (K446A) or no (S609A) separation-of-function, a combination K446A/S609A synergistically diminished exonuclease processing by ~17-fold. Substitution of T378 and V633 to tryptophan severely limited yield during purification, likely reflecting reduced global stability (data not shown). Instead, equivalent substitution to phenylalanine maintained stability while simultaneously disrupting Pso2 endonuclease processing (**Figure 4.12**).

To interrogate the biological relevance of each nuclease function, Pso2 mutants were assessed for their capacity to rescue the ICL hypersensitivity associated with *Δpso2* in *S. cerevisiae*. It was reasoned that if either activity was dispensable within Pso2-dependent ICL repair, mutants selectively diminishing that activity would fully rescue

knockout phenotypes. To that end, Pso2 was cloned under the control of the galactose inducible GAL1 promoter, in-frame with a C-terminal V5 epitope to facilitate detection (**Figure 4.13A**). To control for any potential growth changes due to induction, expression of all Pso2 mutants (and controls) was normalized using western blot following induction with galactose (**Figure 4.13B-C**) and each ICL-treated culture was matched with a control.

To establish the yeast survival assay, colony formation following cisplatin damage was monitored in wildtype (dark green) and Pso2 knockout (dark red) strains (**Figure 4.14**). Following treatment with 2 mM cisplatin, an approximate 5-fold reduction in survival (From 83 to 17%, relative to undamaged controls) was observed. Full-length $Pso2^{WT}$ was able to rescue the ICL hypersensitivity (54% survival, light green), while catalytic dead $Pso2^{D252A/H253A}$ reduced survival beyond the knockout (1.3%, light red). Full phenotypic rescue by $Pso2^{WT}$ may have been limited by the C-terminal fusion, similarly observed in a previous report[325]. Monofunctional Pso2 mutants were unable to rescue the knockout, all of which (except S609A) had equivalent survival compared to the knockout control. $Pso2^{S609A}$ resulted in a partial rescue, likely due to the minimal impact on either exo-/ or endo-nuclease activity (1.65-fold, **Figure 4.12**).

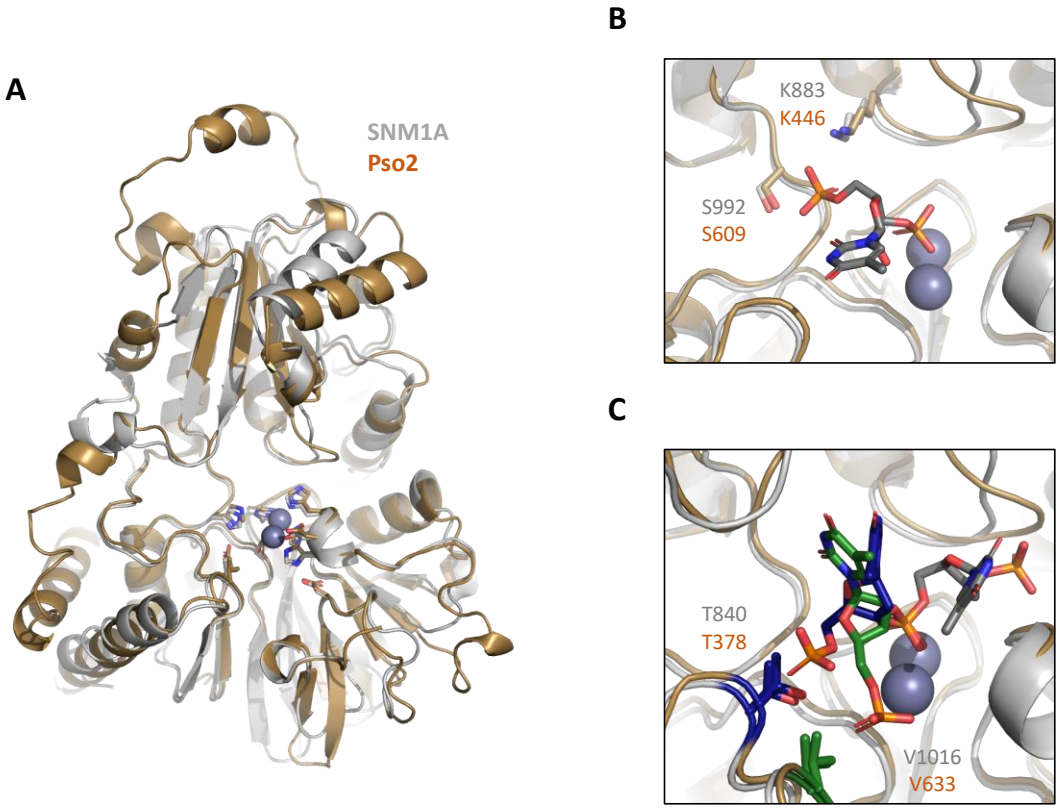


Figure 4.10. Residues that separate SNM1A nuclease activities are conserved in Pso2.

[A] Global alignment of AlphaFold tertiary structure prediction of *Saccharomyces cerevisiae* Pso2$^{211\text{-}661}$ (brown) with SNM1A model (grey, generated from **Figure 4.1**). Conserved MβL and β-CASP residues shown with Zn ions. **[B]** Zoomed alignment of residues critical for exonuclease activity in SNM1A (exonuclease substrate-bound model from **Figure 4.4**). **[C]** Zoomed alignment of residues critical for endonuclease activity in SNM1A (endonuclease substrate-bound model from **Figure 4.7**). **[D]** Partial sequence alignment of β-CASP family members. Critical residues for separation-of-function are highlighted yellow (*/. denote identical and similar residues, respectively).

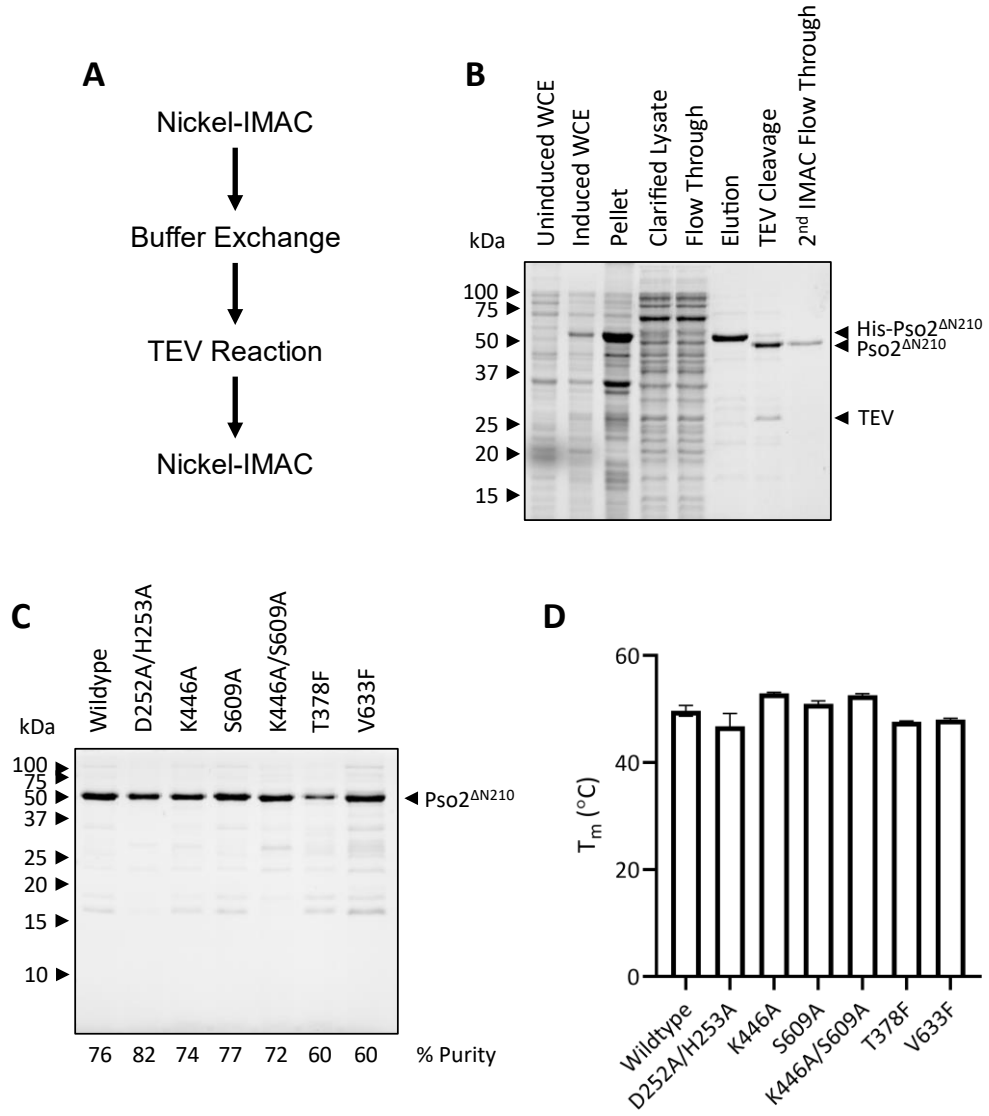
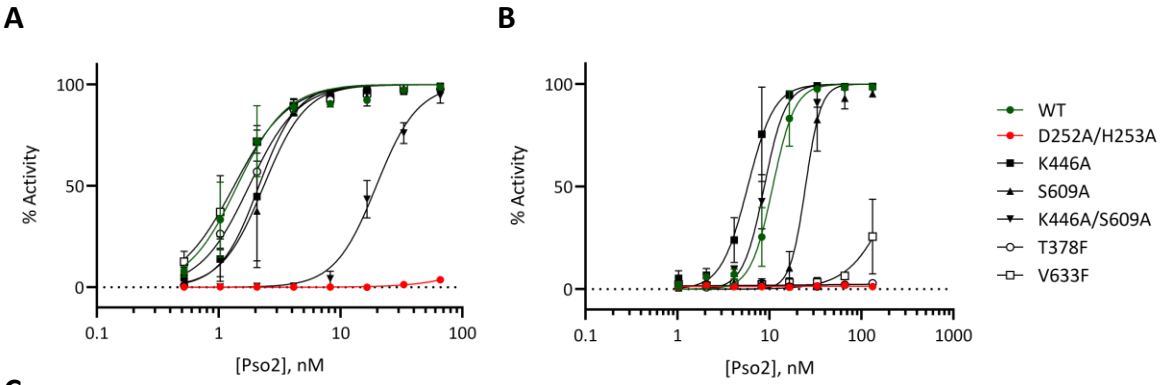


Figure 4.11. Conserved mutants of Pso2$^{\Delta N210}$ are stable and can be purified.

[A] Schematic of Pso2 purification. **[B]** Representative purification gel of Pso2^{WT} (211-661, expected size: 52.8 kDa). Fractions from protein induction, Nickel-IMAC, TEV cleavage and second Nickel-IMAC purification were analyzed by 12% SDS-PAGE and visualized using coomassie stain by GelDoc EZ Imager. **[C]** SDS-PAGE analysis of all purified Pso2 mutants. **[D]** Melting temperature of Pso2 mutants (18 µM). Error bars indicate standard error of the mean from three independent experiments.



C

Mutant	Exonuclease (P.1F)			Endonuclease (F.20T)		
	$[Pso2]_{50\%}$ (nM)	Fold-Decrease	**Adjusted Fold-Separation**	$[Pso2]_{50\%}$ (nM)	Fold-Decrease	**Adjusted Fold-Separation**
Wildtype	1.42 ± 0.33	1.00	**1.00**	10.86 ± 1.56	1.00	**1.00**
D252A/H253A	ND*	>100	**>100**	ND	>100	**>100**
K446A	2.14 ± 0.59	1.51	**2.82**	5.80 ± 1.22	0.53	NA
S609A	2.34 ± 0.52	1.65	**0.73**	24.62 ± 6.75	2.27	NA
K446A/S609A	19.67 ± 2.15	13.87	**17.15**	8.78 ± 0.95	0.81	NA
T378F	1.78 ± 0.27	1.25	NA	ND	>100	**>100**
V633F	1.34 ± 0.22	0.94	NA	240.50 ± 125.60	22.15	**23.50**

Figure 4.12. Pso2 nuclease functions can be isolated by targeting residues critical for SNM1A activity.

Quantification of mutant Pso2 nuclease activity. **[A]** Pso2 mutants (0.5-66 nM) were incubated with **P.1F** (100 nM) at 30 °C for 30 minutes. **[B]** Pso2 mutants (1-132 nM) were incubated with **F.20T** (100 nM) at 30 °C for 30 minutes. Reaction products were analyzed by 22.5% denaturing PAGE and visualized by ChemiDoc MP Imager. Quantification was completed as in **Figure 4.6**. Error bars represent standard error of the mean from three independent replicates. **[C]** Calculation of normalized nuclease deficiency of Pso2 mutants. 'Fold-Decrease' was defined as the ratio between [Mutant] over [WT] for each activity. *ND indicates where insufficient activity was observed for accurate curve fitting.

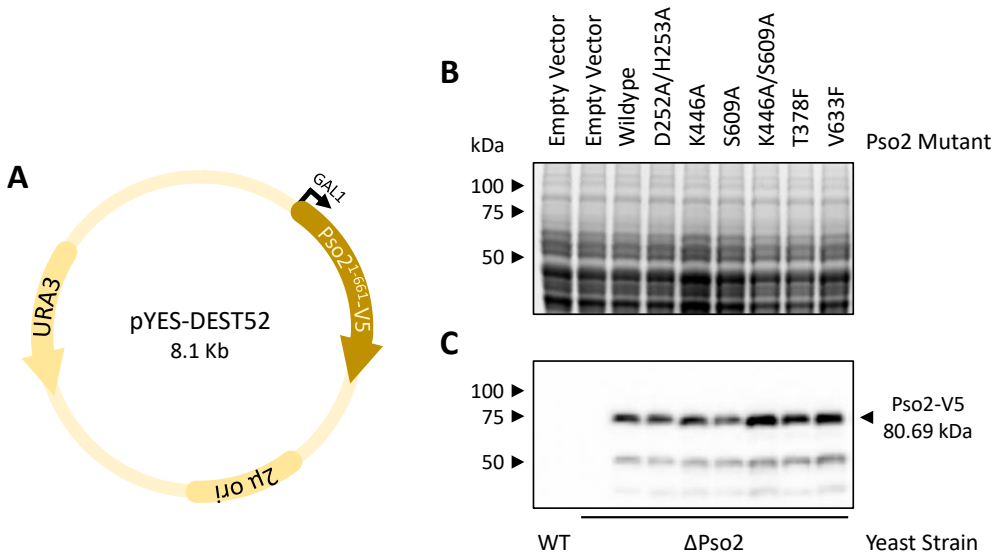


Figure 4.13. Full length Pso2 can be inducibly expressed in *Δpso2 Saccharomyces cerevisiae*.

[A] Simplified schematic of pYESDEST-52 yeast expression vector used to express Pso2 constructs in **Figure 4.14**. Full length Pso2 (wildtype or mutants) were cloned with a C-terminal V5 epitope under the control of the galactose-inducible GAL1 promoter. Yeast transformed with each Pso2 mutant were grown to saturation at 30 °C and 220 rpm prior to addition of 2% v/v galactose for 2 or 3 hours. Cells were collected, lysed and whole cell extracts were normalized by **[B]** SDS-PAGE and expression was visualized by **[C]** western blot. Western blot was probed with mouse monoclonal anti-V5 (1:2500, R960-25 Thermofisher) primary antibodies, followed by secondary incubation with rabbit anti-mouse IgG-HRP (1:4000, A9044 Sigma). Pso2-V5 was visualized for chemiluminescence by ChemiDoc MP Imager following treatment with ECL-substrate.

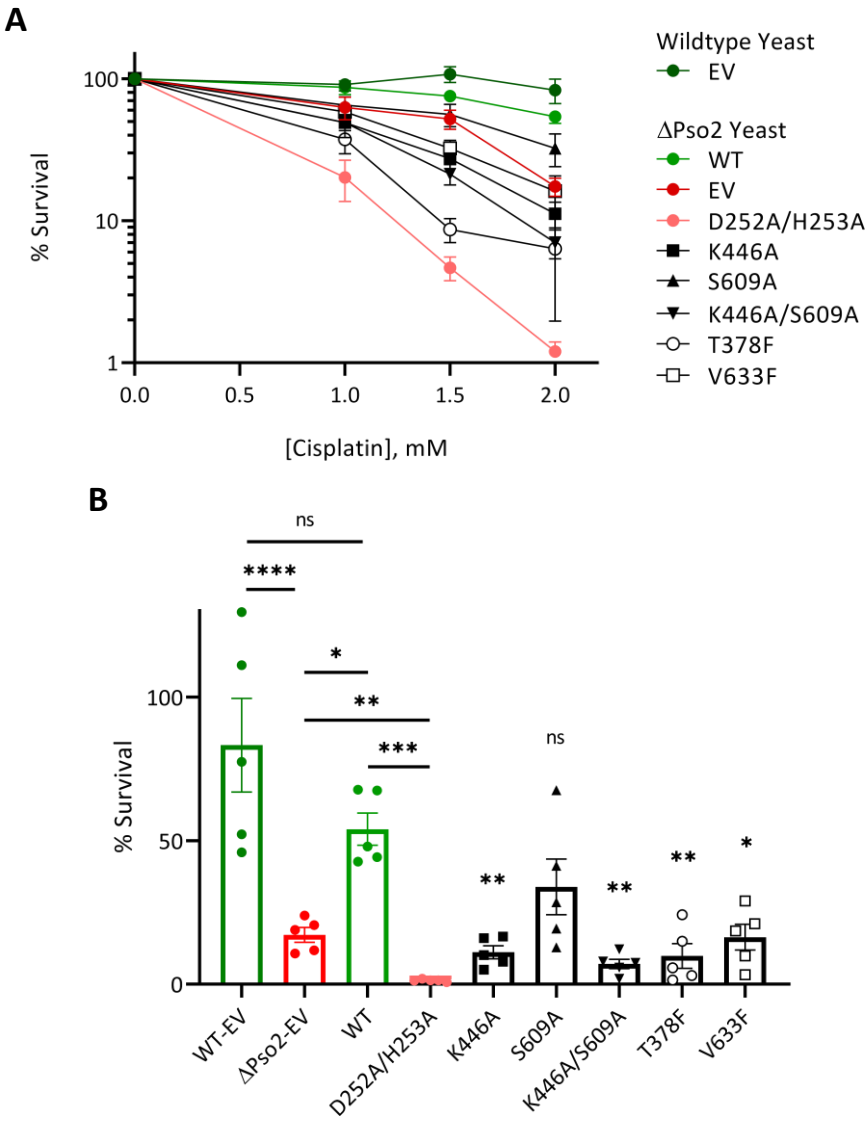


Figure 4.14. Pso2-dependent ICL repair requires both endonuclease and exonuclease activity.

Quantification of yeast survival assay. *Δpso2* yeast complemented with indicated Pso2 mutants or controls were grown to saturation at 30 °C and 220 rpm prior to induction with 2% v/v galactose for 2 or 3 hours. Yeast were damaged with **[A]** 0-2 mM cisplatin for 90 minutes at 28 °C (data at 2 mM cisplatin expanded in **[B]**). Serial dilutions of each Pso2 mutant were plated on auxotrophic media and colonies counted after incubation for 3 days at 30 °C. '% Survival' was determined following normalization to equivalent undamaged cultures. EV denotes pYESDEST-52 empty vector controls. Significant differences were identified between groups using the one-way ANOVA test ($F = 14.57$, $p < 0.0001$). Pairwise significance was determined by Tukey's test. Unless indicated, significance indicates difference compared to $Pso2^{WT}$ complement. Error bars represent the standard error of the mean from 5 independent replicates.

4.6 Discussion

Despite significant progress in understanding the genetic and biochemical interactions of SNM1A, models of how it contributes to ICL repair remain largely speculative. This chapter primarily addresses the ambiguity surrounding which nuclease activity SNM1A relies on to process ICL intermediates. To strategically isolate individual SNM1A nuclease functions using mutagenesis, we needed to identify residues which differentially contribute to each activity. Despite significant effort, our lab has not been able to generate structural data for substrate-bound SNM1A that would directly inform our understanding of how SNM1A engages with exonuclease and endonuclease DNA substrates. In the absence of direct data, glimpses of the molecular mechanism of each activity rely on biochemical evidence reported by our lab[179] and others[177,178]. This biochemical data can be further contextualized using structural data available for SNM1A and its homologs. At the outset of this work, Gileadi and colleagues had reported apo-structures of SNM1A (PDB: 5AHR) and Apollo (PDB: 5AHO) nuclease domains[178]. Interestingly, two tartaric acid molecules from the Apollo crystallization buffer formed stable 'phosphate-mimicking' interactions directly coordinating the two active site Zn ions and a pocket directly adjacent. In addition, the structure of RNase J (homolog from *Deinococcus*) had been reported in complex with a 5' phosphorylated 6 polyU ssRNA substrate[230]. In recent months, further structural models have been published of Apollo[227], and Artemis[174,326] bound to dAMP (PDB: 7A1F) and a short 3' overhang DNA substrate (PDB: 7ABS), respectively. While initial structures were leveraged to constrain models of DNA-bound SNM1A, these new structures have further validated our modelled DNA conformations and provided new insights into potential protein interactions with the non-targeted DNA strand.

4.6.1 SNM1A mutagenesis and implications for substrate engagement

SNM1A exonuclease activity absolutely requires a 5' phosphate[177]. As SNM1A binds DNA substrates with equivalent affinity regardless of phosphorylation (**Figure**

2.14), the terminal phosphate is unnecessary for substrate binding. Instead, the phosphate seems to promote substrate orientation to favour catalysis. Since endonuclease substrates lack a terminal phosphate, this strict requirement can be exploited to selectively abolish exonuclease activity. Coincident alignment of the terminal phosphate in RNase J and 'pseudo-phosphate' tartaric acid in Apollo structures suggested that residues within this well conserved pocket likely promote exonuclease substrate catalysis. Targeting specific residues deep within the putative binding pocket had differential effects on the exonuclease function (**Figure 4.5A**). While targeting E993 had no impact on activity, disruption of K883 and S992 selectively diminished exonuclease activity, both individually and in combination.

SNM1A[175,179], Pso2[180,314], Apollo (unpublished results) and Artemis[220,221] are homologous dual function exo-/ endo-nucleases. While SNM1A, Pso2 and Apollo demonstrate a strong exonuclease preference *in vitro*, Artemis functions predominantly as an endonuclease (following phosphorylation). SNM1A mutagenesis results reported here may illustrate some of the evolutionary drivers of these preferences. The phosphate-binding pocket is essential for exonuclease processing. Structure (**Figure 4.15**) and sequence (**Figure 4.10**) alignment demonstrates that critical residues positioned to coordinate the terminal phosphate are conserved between SNM1A (K883, S992 and E993), Pso2 (K446, S609, E610) and Apollo (K186, S274, D275). We showed that substitution of the Lys and Ser in SNM1A significantly disrupts exonuclease digestion. Recent work by McHugh and colleagues confirmed a similar dependence on the phosphate-binding pocket in Apollo[227]. Evolutionary disruption of this pocket in Artemis (Y212, S317, F318) likely accounts for its inefficient exonuclease activity, as both Tyr and Phe sterically preclude access to the phosphate-binding pocket. Further, we report above that elimination of the p-hydroxy group of Y841 in SNM1A actually stimulated endonuclease processing. It is possible Y841 (p-hydroxy, in particular) in SNM1A steers the incoming phosphate into the binding pocket, thereby promoting exonuclease function (proposed Y841 'steering', **Figure 4.15A**). It is notable that F168 is structurally equivalent in Artemis. If this residue acts as a driver for exonuclease activity

in SNM1A, the evolutionary switch to Phe in Artemis may further explain its preference for endonuclease processing (**Figure 4.15D**).

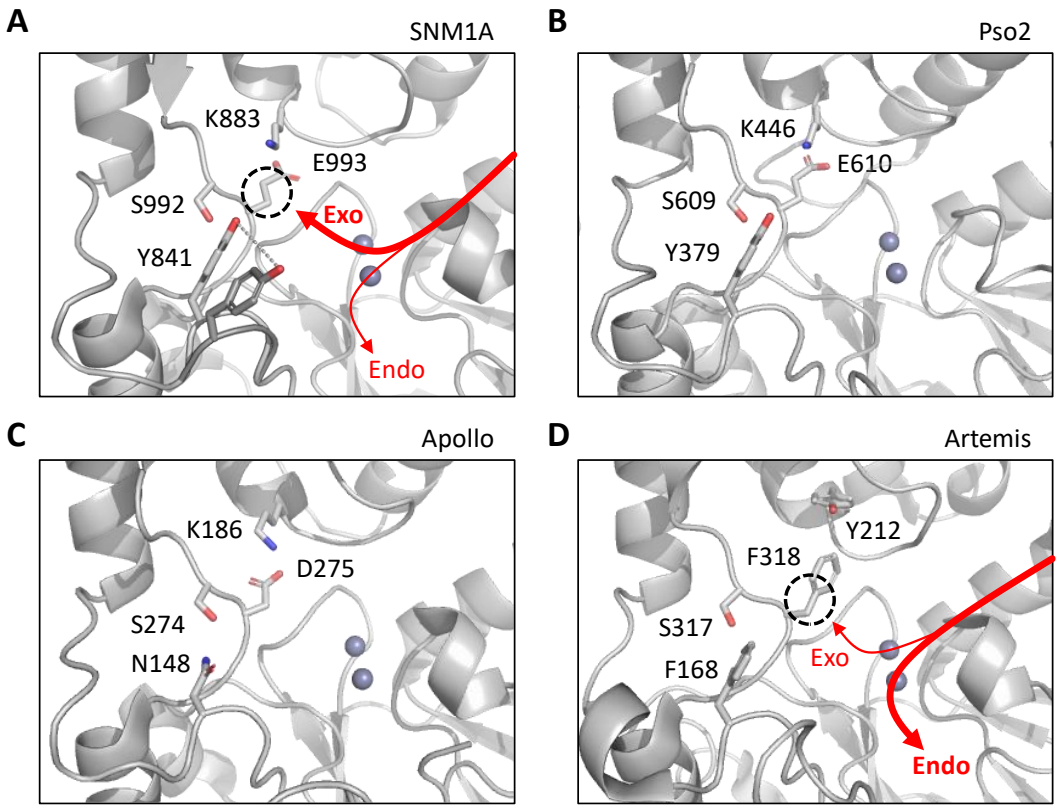


Figure 4.15. Divergence in phosphate-binding pocket residues may drive catalytic preferences.

Global structural alignment of **[A]** SNM1A (PDB: 5Q2A) with **[B]** Pso2 (AlphaFold structure prediction), **[C]** Apollo (PDB: 7A1F) and **[D]** Artemis (PDB: 7A1F). Coordinates of Zn ions were extracted from SNM1B. Dashed black lines indicates the phosphate-binding pocket. Red arrowheads indicated the 5' end of incoming DNA, while width represents catalytic efficiency. Hypothetical 'phosphate steering' mechanism by Y841 in [A] depicted by movement of the flexible loop (dark grey).

Predicting how SNM1A interacts with an endonuclease substrate was complicated by broad substrate specificity. As demonstrated in **Chapter 2**, SNM1A endonuclease function has limited structural and sequence preferences. This suggested that SNM1A-DNA interaction is mediated primarily by electrostatic interactions with the sugar-phosphate backbone, and only limited contacts with the nucleobases. Further, SNM1A required at least 4 nucleotides to bind and therefore catalyze an endonuclease reaction. A requirement for longer DNA regions likely produces the affinity necessary to correctly orient the endonuclease substrate. This necessitates at least 3 nucleobases interacting 'past' the active site, as SNM1A endonuclease activity on short DNA substrates exclusively occurred on the phosphodiester bonds toward the 3' end. We anticipated that catalysis should be closely conserved between exo-/ and endo-nuclease activity. As such, the polarity and position of the 'incoming' DNA would likely be degenerate between the two substrates. Absent any large-scale conformational rearrangement following DNA binding, three potential binding grooves on the surface of SNM1A were identified that might facilitate this interaction. Connecting the DNA binding channels 5 and 3' of cleavage, 'upstream' nucleotides must (red) bend up toward the flexible loop, (blue) continue straight or (green) bend down along the MβL (as modelled in **Figure 4.7**). To ascertain which channel facilitated endonuclease substrate binding, a series of mutants were designed either to (i) disrupt potential electrostatic interactions with DNA or (ii) sterically exclude DNA binding in the channel.

Targeting residues, either individually or in combination, within the predicted red (T962, W964 and R960) or blue (E845, Y846, K969, T1015, N1017, K1022 and Y1030) DNA binding grooves failed to disrupt endonuclease catalysis. In contrast, mutagenesis targeting the green channel demonstrated that a triple mutant (Y704F/K705A/Y719F) slightly diminished endonuclease activity. Targeting potential electrostatic interactions was insufficient to substaintially disrupt the endonuclease activity. This tolerance likely reflects the extensive network of hydrogen bonds between SNM1A and DNA backbone, such that abrogation of a few contacts has negligible effects on overall substrate affinity.

To sterically exclude a DNA substrate from binding in one of the three potential binding grooves, residues positioned at the entrance of each channel were mutated to tryptophan or phenylalanine. Channels were sterically disrupted by substitution of T962 (red), T840 or S1023 (blue) and V1016 (green). Where targeting of electrostatics was largely ineffective, steric mutants significantly diminished endonuclease processing. Although mutation of T840 (blue) had a more marked effect on activity than V1016 (green), targeting Y704, K705 and Y719 (green) improved disruption of endonuclease activity (**Figure 4.3C**). Additionally, targeting S1023 (blue) had no impact on nuclease activity. Although further structural work is required, data presented here suggests endonuclease DNA substrates interact primarily through the 'bent-down' (green) channel (**Figure 4.7E**).

4.6.2 Other SNM1A mutagenic insights

Although not all mutants were useful for separation-of-function, additional results from mutagenesis experiments reported here provide insights into how the structure of SNM1A impacts its function:

1. The putative phosphate-binding pocket is formed at an interface between MβL and β-CASP, capped by a large flexible loop (P961-P978). While apo-structures of SNM1A lack electron density for this loop (indicating a high degree of flexibility), some inhibitors have been demonstrated to promote loop stabilization[225]. As such, loop flexibility was anticipated to promote substrate binding, selection or stabilization. Residues within this loop were targeted individually for substitution, to determine whether they form contacts with the terminal phosphate necessary for catalysis (see residue positions in **Figure 4.2** and resulting activity quantification in **Figure 4.3**). None of the targeted residues within the flexible loop (P961, T962, W964) had any significant effect on substrate processing, indicating that loop stabilization is likely not a driver of substrate orientation or catalysis. Likewise,

while Y879 is proximal to this binding pocket, it also had no impact on either nuclease activity, consistent with a previous report[178].

2. V1016 is the conserved β-CASP motif C. In RNA processing β-CASP nucleases, motif C is an M2-coordinating histidine[166]. In contrast, DNA processing β-CASP nucleases encode a valine. While a previous mutagenesis report demonstrated that reversion of V1016 to His did not alter the sugar preference of SNM1A[177], an instinctive consequence of losing the His metal-coordination is lower M2 affinity (bound by 3 instead of 4 residue ligands). As reported here, substitution of V1016 to Phe or Trp significantly diminished SNM1A endonuclease processing. This interdependence of a more flexible active site with DNA-specific endonuclease processing is unclear.
3. Recent structural characterization of Artemis identified a novel Zn-binding motif within the β-CASP fold[174,326]. Co-crystallization with a dsDNA substrate suggests this motif makes critical interactions with the non-target DNA strand[174]. While sequence alignments between β-CASP homologs suggest that this motif is not conserved in SNM1A or Apollo, closer inspection of the analogous region within SNM1A identified residues consistent with a His_2Cys_2 zinc binding finger. Though no evidence of Zn binding at this site was identified in previous SNM1A structures, we mutated either H930 or C925/C949 to determine whether this may represent a conserved functional motif. Interestingly, neither mutant prompted a change in nuclease processing (**Figure 4.3**). This suggests this β-CASP Zn binding finger may be unique to Artemis.

4.6.3 Extending separation-of-function to Pso2

Targeting the 5' phosphate-binding pocket in SNM1A selectively diminished the exonuclease activity (**Figure 4.6**). While Y841F and S992A only modestly reduced activity by approximately 20% each, K883A dramatically disrupted exonuclease processing (44-fold). Maintaining wildtype endonuclease activity, $SNM1A^{K883A}$ residual exonuclease function closely mimics that observed with a previously reported 'catalytic' mutant

targeting metal-coordination by D736A[179]. Combinations of each mutant synergistically disrupted exonuclease activity, further substantiating that all identified residues are likely to facilitate phosphate-binding. Sequence conservation between SNM1A and Pso2 allowed for recapitulation of exonuclease disrupting mutants. While generally producing a similar effect, all were less severe than their SNM1A equivalent. Specifically, S609A demonstrated no significant separation-of-function, however K446A and K446A/S609A produced 2.8- and 17.2- fold disruption in exonuclease activity, respectively (**Figure 4.12**).

Selective disruption of endonuclease activity required targeting residues positioned at the opening of a newly described *extended* DNA-binding groove (**Figure 4.7**). To this effect, SNM1A mutants V1016W, T840F and T840W diminished endonuclease processing by 5.3-, 19.8- and >100- fold, respectively. Though substitution of V1016W resulted in more modest endonuclease disruption, exonuclease processing was entirely unaffected. In contrast, while targeting T840 selectively diminished endonuclease activity, exonuclease activity was reduced 2.5- (T840F) and 4-fold (T840W). Extended in Pso2, V633F and T378F provoked more dramatic endonuclease disruption (23.5- and 100- fold, respectively). The increased phenotype severity in Pso2 may reflect the increased baseline preference for endonuclease substrates[179,180]. Conserved impact of separation-of-function mutants indicates similar substrate binding and catalytic mechanisms between SNM1A and Pso2, as expected for functional homologs[176]. Collectively, *in vitro*, this series of Pso2 mutants demonstrate a 'dose-dependent' selective disruption of exonuclease (S609A < K446A < K446A/S609A) and endonuclease (V633F < T378F) activity, sufficient for probing the biological relevance of each activity within yeast ICL repair.

4.6.4 Separation-of-function and biological relevance of SNM1A nuclease activities

Examining the behaviour of separation-of-function mutants in a cellular repair context can provide direct evidence for Pso2 participating as an exonuclease and/or

endonuclease. Further, exploiting the differing severity of Pso2 mutants provides a mechanism for probing the sensitivity of each function in repair.

Prior to investigating the ability of each mutant to rescue the ICL hypersensitivity exhibited by Pso2 knockout yeast, we needed to establish controls. In particular, defining the extent of exogenous wildtype ($Pso2^{WT}$) and inactive ($Pso2^{D252A/H253A}$) complementation relative to baseline wildtype and Pso2 knockout yeast. While expression of $Pso2^{WT}$ resulted in a near complete rescue, catalytic dead $Pso2^{H252A/D253A}$ exacerbated the survival hypersensitivity following cisplatin treatment (**Figure 4.14**). This dominant negative phenotype suggests that the presence of non-functional Pso2 is more detrimental to ICL repair than the absence of Pso2. Interestingly, this phenotype was not observed in other mutants targeting metal coordination (D252A[325] or H611A[81,180]). This may reflect the dramatic reduction in activity exhibited by the double mutant used here (~0.1% of $Pso2^{WT}$), as compared with single mutants, which demonstrate residual activity (~1-5% of $Pso2^{WT}$) *in vitro*[180,314]. Since $Pso2^{D252A/H253A}$ was expressed at similar levels (**Figure 4.13**), this phenotype must be driven by interactions in repair. While the mechanism driving the dominant negative phenotype is unclear, inactivation of Hrq1 helicase (Pso2-interacting repair factor) results in a similar effect[81]. It seems likely that non-functional components of the Pso2-dependent ICL repair pathways actively prevent access to competing sub-pathways, such as MutSα[319,320] and/or proto-FA[105,321].

Expression of nearly all separation-of-function Pso2 mutants failed to complement the survival hypersensitivity with *Δpso2* yeast. $Pso2^{S609A}$ did partially rescue the ICL sensitivity to a similar extent as $Pso2^{WT}$, likely reflecting the lack of nuclease disruption. Yeast survival in cultures expressing $Pso2^{K446A}$ and $Pso2^{K446A/S609A}$ both phenocopied Pso2 knockout yeast, highlighting that even minor disruption (K883A induced ~3-fold reduction in exonuclease activity) prevents robust ICL repair. Similarly, both endonuclease disrupting mutants (T378F, V633F) failed to rescue the knockout hypersensitivity. Neither exo-/ or endo-nuclease deficient mutants reproduced the dominant negative phenotype demonstrated by non-functional Pso2, suggesting that

Pso2 may function in two or more discrete (sub)pathways of ICL repair, each requiring a distinct nuclease function.

4.6.5 Distinctions between yeast and mammalian ICL repair

S. cerevisiae is a common model system used to infer how human (or other mammalian) cells recognize and remove crosslink damage[6,327]. While sharing many commonalities, there are significant distinctions between the two systems which may limit direct extrapolation of the *in vivo* yeast data reported in this chapter. Understanding these distinctions will be critical for extending our model to SNM1A in human repair. Similar to mammalian cells, ICL repair in yeast is primarily accomplished by either transcription-coupled or replication-dependent repair.

TC-ICL repair closely resembles the mammalian equivalent (described in **1.3.2.2**), depending on the coordination between transcribing polymerases and the full NER exinuclease. In *S. cerevisiae*, homologs of TFIIH factors Rad25 (hXPB) and Rad3 (hXPD) initiate NER incision complex formation following failed forward translocation of the RNA polymerase by Rad26 (hCSB)[74,328]. Unhooking depends on the presence of Rad1-Rad10 (hXPF-hERCC1) and Rad2 (hXPG) endonucleases. Unlike RDR, transcription-initiated repair does not proceed via DSB formation/ repair[79,329]. TC-ICL repair is Pso2-dependent, as stationary cells[310] (or G1 synchronized cells[320]) lacking Pso2 maintain a hypersensitivity to ICL-inducing agents. While RDR remains the dominant repair process, yeast rely more extensively on TC-ICL repair than most mammalian cells[317,320,330]. The yeast genome contains approximately 6000 genes encoded within 12 Mbp across 16 chromosomes[331]. While expression depends largely on environmental conditions of the yeast, nearly 85% of the genome can be transcriptionally active[332], suggesting that most ICL lesions will be encountered by transcription machinery. In contrast, the human genome is significantly larger and relatively less transcriptionally active.

Significant differences exist between yeast and mammalian RDR. While replication-dependent repair in higher eukaryotes depends primarily on the FA-pathway

(refer to **Figure 1.4**), *S. cerevisiae* lacks necessary core factors[105,321]. Instead, yeast RDR relies on sequential collaboration between NER and HR. While partially overlapping pathways exist in this network, unhooking typically requires the complete assembly of the NER incision complex[105,327]. Flanking incisions by the NER exinuclease generate DSBs, which are then repaired by canonical HR components. Critically, Pso2 is required to bridge the gap between NER and HR processing[317,318,320,325]. While cells lacking Pso2 accumulate unrepairable DSBs, it is unclear whether these arise because Pso2 contributes to unhooking or is required to generate HR-competent DNA ends.

Despite differences in ICL repair pathways, yeast remain an excellent system for understanding the role of human SNM1A. *S. cerevisiae* lacks many of the confounding/ competing nucleases which have obscured direct study of SNM1A[165]. Pso2 functions at a critical bottleneck in the dominant RDR pathway, where SNM1A appears to compete for a subset of FA-dependent ICL repair intermediates. We therefore were able to exploit the phenotypic clarity in Pso2 knockout yeast to quantify the behaviour of our mutant constructs within repair. Though Pso2 and SNM1A likely play degenerate roles in TC-ICL repair, it is unclear whether they process identical RDR intermediates. Regardless, biochemical and cellular characterization suggests that the enzymes can substitute (at least partially) each other in the process[176].

Work presented in this chapter demonstrates, for the first time, that Pso2 endonuclease function directly contributes to ICL repair. Whether these Pso2 exonuclease- and endonuclease-dependent parallel pathways exist within RDR, or reflect differences between RDR and RIR will require further investigation. Many questions remain. Could Pso2 endonuclease activity be required to process a subset of 'dirty ends' generated by incomplete unhooking, acting in series with its translesion exonuclease activity? Or, does Pso2 endonuclease activity function in a distinct repair pathway? Might translesional exonuclease processing be necessary during TC-ICL repair, but dispensable during RDR? Careful re-analysis of genetic interactions, leveraging separation-of-function mutants characterized above may help to provide answers. Likewise, further work in mammalian systems is required to translate these findings to

SNM1A function. Though SNM1A is the functional homolog of Pso2, direct interrogation of SNM1A separation-of-function mutants in human cells will be required to demonstrate whether both nuclease activities contribute to ICL repair in mammals.

4.7 Appendix

HHHHHH RSDITSLYKKAGL DYDIPTT ENLYFQ^S

211 KLVLPSFKIIKFNNGHEIVVDGFNYKASETISQYFLSHFHSDHYIGLKKSWNNPDENPIK
KTLYCSKITAILVNLKFKIPMDEIQILPMNKRFWITDTISVVTLDANHCPGAIIMLFQEF
LANSYDKPIRQILHTGDFRSNAKMIETIQKWLAETANETIDQVYLDTTYMTMGYNFPSQH
SVCETVADFTLRLIKHGKNKTFGDSQRNLFHFQRKKTLTTHRYRVLFLVGTYTIGKEKLA
IKICEFLKTKLFVMPNSVKFSMMLTVLQNNENQNDMWDESLLTSNLHESSVHLVPIRVLK
SQETIEAYLKSLKELETDYVKDIEDVVGFIPTGWSHNFGLKYQKKNDDDENEMSGNTEYC
LELMKNDRDNDDENGFEISSILRQYKKYNKFQVFNVPYSEHSSFNDLVKFGCKLKCSEVI
PTVNLNNLWKVRYMTNWFQCWENVRKTRAAK* 661

Appendix 4.A. Amino acid sequence of Pso2 (211-661) expression construct.

Highlighted N-terminal fusions include His_6 tag (yellow), attB1 scar (green), flexible linker (blue) and TEV protease recognition site (underlined). ^ denotes peptide bond cleaved by TEV protease, * indicates the C-terminal stop.

Chapter 5
Closing remarks & future directions

Interstrand crosslinks represent a severe block to essential cellular processes, particularly transcription and replication. A complex network of repair pathways ensures that ICL damage is rapidly recognized and removed to maintain genomic fidelity. Due to the clastogenic nature of ICL damage, genetic disruption of repair factors result in Fanconi Anemia. Notably, the severe clastogenesis of ICLs has been leveraged in chemotherapy. Where cancers lack the necessary DNA damage response, ICL-inducing chemotherapy provokes specific cytotoxicity to eliminate or reduce tumour growth[2,248]. Understanding how ICL repair factors contribute to lesion recognition and removal is essential for improving the outcomes of patients with FA and cancer alike. The work presented in this thesis significantly expands our understanding of one such factor, SNM1A.

5.1 Biochemical characterization of SNM1A

How SNM1A nuclease activities collaborate *in vivo* to facilitate ICL repair is unclear. To understand what role the exonuclease and endonuclease activity of SNM1A may play within ICL repair, we needed to thoroughly characterize its function *in vitro*. First, we developed a robust purification scheme by heterologous expression using *E. coli*. The optimized procedure enabled us to explore SNM1A catalytic mechanism and substrate preferences. Using a series of bulky DNA substrates, we observed significant stalling of SNM1A exonuclease activity. While some models predict SNM1A engages its biological substrate from an endonuclease nick, we showed that SNM1A processing was reduced nearly 20-fold when the terminal phosphate is obscured at a nick or small gap. Next, we demonstrated that SNM1A endonuclease activity was most active on substrates containing extended regions of ssDNA. Both substrate affinity and the catalytic rate increased dramatically as the size of ssDNA increased from ~6 to 20 nucleotides. Reminiscent of Artemis and Pso2, SNM1A endonuclease activity exhibited a limited 3' cleavage preference within hairpin or gapped DNA substrates. Finally, we determined that the differential efficiencies of SNM1A exonuclease and endonuclease processing result from distinct metal-dependent reaction mechanisms.

The relative efficiency of SNM1A nuclease activities were apparent on DNA substrates designed to mimic potential ICL repair intermediates. SNM1A was able to completely process an ICL-containing stalled replication fork, initiated by endonuclease processing of a 5' flap followed by exonuclease digestion toward (and past) the ICL. Surprisingly, translesional exonuclease digestion past the ICL was the rate-limiting step, not endonuclease initiation. Experiments presented in this chapter raise many new questions about SNM1A function and mechanism. Increasing exonuclease processivity from nicked substrates raises the possibility that SNM1A exonuclease activity functions to resect long stretches of DNA. Limited translesion exonuclease activity reported here suggests that crosslink structure and its DNA context may determine the efficiency of

SNM1A lesion bypass. Finally, the distinct metal-dependence of SNM1A catalytic activity suggests a mechanism for differential regulation of each nuclease activity.

Previous characterization of SNM1A exonuclease activity identified increased processivity, specifically on large DNA substrates (>50 nucleotides)[178]. Here, we observed that SNM1A exonuclease processing initiated from a nicked dsDNA substrate is similarly processive. These data suggest that SNM1A processing from a nick may promote extended DNA resection, significantly exceeding what would be necessary at or around an ICL. Future work should examine extended resection around an ICL and any potential links to DNA damage response. SNM1A-dependent resection of large stretches of DNA during repair could function to exacerbate a DNA damage response via ATR activation. Reminiscent of Exo1-XPG competition during stalled NER (as introduced in **1.3.1.1**)[57], extended SNM1A exonuclease processing may produce a necessary signal indicating to the cell that repair is not proceeding as expected. In this scenario, SNM1A would only be required to process a subset of intermediates when repair is dysfunctional. As previous reports connected SNM1A with DNA damage signalling and mitotic checkpoints[184,197], SNM1A may be functioning indirectly in DDR to trigger replication and cell cycle stalling in response to improper repair.

SNM1A translesional exonuclease processing is relatively unique and likely plays a critical role during ICL repair. In contrast to other ICL translesional exonucleases FAN1 and SAN1, SNM1A translesion activity seems to depend on crosslink structure and/or the stability of the adjacent DNA. Where FAN1 and SAN1 digest past ICLs with no (or limited) stalling[92,162], we observed that ICL context impacts the efficiency of SNM1A translesional activity[179]. Examining what factors license SNM1A translesional exonuclease activity will be critical for determining when it functions in repair. This will require quantifying the ability of SNM1A to bypass structurally distinct ICLs, in different DNA contexts. Further, recent work identified Hrq1 helicase-dependent activation of Pso2 translesional processing. Determining whether SNM1A binding partners (or PTMs) alter the efficiency of translesion bypass will contextualize how and where SNM1A acts in repair.

Finally, we demonstrated that distinct metal-dependent mechanisms support SNM1A exonuclease and endonuclease activities. Where Zn^{2+} was sufficient to support endonuclease processing, supplemental Mg^{2+} was required to fully stimulate the exonuclease activity. As a single active site enables both activities, this suggests differences in catalytic efficiency of each nuclease activity result from the discrete metal-dependence. In **2.6.4**, I discuss a model where Mg-dependent stimulation may occur through an uncharacterized substrate-coordinated M3 metal binding site. While suggestive, other models of catalysis may explain the discrepancy between metal-dependence. For example, models of SNM1A bound to short exonuclease and endonuclease DNA substrates presented in **Chapter 4** suggest that the function of M1 and M2 sites differ during exonuclease and endonuclease catalysis. Where M1 coordinates the O^3-P bond and M2 coordinates the nucleophile during exonuclease digestion (refer to schematic in **Figure 2.16**), the active site may invert for endonuclease processing (M2 coordinates O^3-P and M1 coordinates the nucleophile). Active site inversion is necessary for the endonuclease substrate to engage with the lower DNA binding channel, consistent with our mutagenesis experiments (refer to model in **Figure 4.7**). If the endonuclease substrate forms additional contacts with the M2 metal (absent during exonuclease digestion), this may also explain why endonuclease processing specifically depends on Zn^{2+}. Further work is necessary to differentiate whether the Mg-dependent stimulation occurs through a hypothetical M3 site or an alternative mechanism.

5.2 Inhibition of SNM1A

Increased DNA repair capacity is a major driver of innate or acquired resistance to ICL-inducing chemotherapies. High expression of critical ICL repair factors in target tissue and tumours reduces the therapeutic value of these treatments. As such, significant research has focused on developing specific inhibitors against ICL repair factors, particularly XPF[270,271,277] and ATR[248]. In contrast to XPF and ATR, SNM1A

represents a more specific target for the disruption of ICL repair. In addition to therapeutic aims, specific inhibitors would enable *in vivo* functional assays to explore SNM1A function in repair by leveraging dose-dependent, temporal and reversible inhibition. As molecular probes, small molecule inhibitors have been extremely useful for elucidating protein function that would otherwise be difficult to uncover.

Here, we report initial characterization of Series A bioactive SNM1A inhibitors, identified by *in vitro* HTS. Further, using an *in silico* approach, we identified new Series B inhibitors which specifically inhibit SNM1A over SNM1B *in vitro*. Primary characterization focused on Series A inhibitors, using a collection of assays to determine potency and mechanism of inhibition. Historical HTS characterization studies have identified recurring non-specific mechanisms of inhibition, often associated with particular chemical moieties. As such, we conducted assays to identify inhibitors which function through metal competition, colloidal aggregation or direct interactions with DNA. While no Series A inhibitor displayed specific inhibition of SNM1A, four functioned via a non-specific mechanism of action. To address the limited specificity of Series A inhibitors, we conducted an *in silico* HTS. Here, we improved inhibitor specificity *in vitro* by enriching for compounds predicted to have a high affinity for SNM1A but not a close homolog. Collectively, we identified nine promising SNM1A inhibitors (5 from Series A, 4 from Series B), which represent lead compounds for further development.

To develop useful inhibitors for the lab or clinic, further characterization is required *in vitro, ex vivo* and in culture. First, additional *in vitro* characterization is needed to better elucidate mechanisms of SNM1A inhibition. Next, the potency of each inhibitor needs to be improved. Both aims can be accomplished using structure determination; co-crystallization of SNM1A with each inhibitor. Structure determination would directly answer: how does each compound inhibit SNM1A nuclease activity? And, how can we improve inhibitor potency? In most cell types, SNM1A expression is very low. This implies effective inhibition *in vivo* would require high affinity inhibitors. While inhibitors reported here demonstrate a low micromolar affinity for SNM1A, future SAR

experiments could significantly improve potency. Current work by our lab to crystallize inhibitor-bound SNM1A to that end is ongoing.

An effective inhibitor must interact with its target in a cell. Further development of our lead compounds must determine whether inhibitors (i) interact with SNM1A in a cellular context and (ii) provoke broader inhibition of ICL repair. First, a direct interaction between SNM1A and inhibitors can be established using an *ex vivo* assay such as cellular thermal shift[333]. Inhibitor binding promotes a change in the thermal stability of its target. Using human cell lysates, we can monitor changes (or lack thereof) to the thermal stability of SNM1A produced by each inhibitor. Next, assays designed to monitor the *in vivo* efficiency of ICL repair will be necessary to evaluate the consequences of SNM1A inhibition. Multiple assays should be used to establish whether SNM1A inhibitors provoke a survival hypersensitivity (by clonogenic survival assay) and chromosomal instability (by comet assay) following exposure of ICL-inducing compounds. More direct plasmid reactivation assays have been used to monitor *in vivo* efficacy of XPF inhibitors. As discussed in **1.3.2.2**, plasmid reactivation studies monitor expression levels of an easily detected protein (such as EGFP), where expression correlates with the resolution of a site-specific ICL positioned to interrupt transcription. Unlike chromosomal instability or survival hypersensitivity, plasmid reactivation would directly demonstrate that an inhibitor exerts its effect via inhibition of ICL repair.

5.3 SNM1A as a dual-function nuclease in repair

SNM1A nuclease activity is essential for ICL repair. As SNM1A catalyzes two distinct reactions with a single active site, it was unclear whether both exonuclease and endonuclease activity participate in repair. To elucidate which intermediate(s) SNM1A acts on, we first needed to determine whether SNM1A functions in repair as an endonuclease, exonuclease or both. To that end, we generated models of SNM1A bound to either substrate. These models highlighted residues positioned to preferentially promote engagement of one substrate but not the other. Targeting these

residues by mutagenesis successfully separated the nuclease functions of SNM1A, allowing individual interrogation of each activity in repair. We then extended these separation-of-function mutants to the yeast homolog, Pso2. Using Pso2 mutants, we showed that both exonuclease and endonuclease processing is necessary for repair. Experiments presented in this thesis raise further questions about the role of Pso2 and SNM1A in ICL repair. Do both nuclease activities act concertedly in repair? Or in distinct pathways? Is the catalytic efficiency of Pso2/ SNM1A endonuclease activity stimulated *in vivo*? How conserved is the role of Pso2 in yeast with SNM1A in human ICL repair?

Previous reports suggest that Pso2 function is essential in more than one ICL repair pathway. It is unclear whether these pathways co-exist during replication to resolve distinct intermediates or reflect concurrent RIR and RDR pathways[320]. Further, genetic experiments indicate that Pso2 functions with the Hrq1 helicase or Sak1 kinase in subsets of ICL repair[81,334]. Data presented here demonstrate that both Pso2 exonuclease and endonuclease activities are required for ICL repair. Future work to improve our understanding of Pso2 function in repair should focus on resolving these three layers of findings. Does Pso2 function with Hrq1 exclusively in RIR? Or Sak1 in RDR? Hrq1 stimulates Pso2 exonuclease activity[81]. Can we use Pso2 separation-of-function mutants to determine whether exonuclease activity is only required in RIR? Does Sak1-dependent phosphorylation stimulate Pso2 endonuclease activity? While we do not know exactly how the exonuclease and endonuclease functions of Pso2 contribute to ICL repair, future experiments designed to delineate these genetic interactions could determine whether the nuclease activities participate in distinct repair pathways.

Pso2 and SNM1A both exhibit relatively weak endonuclease activity *in vitro*. With evidence that Pso2 endonuclease activity directly contributes to ICL repair, it will be important to understand how this activity is regulated *in vivo*. Are there uncharacterized interactions with binding partners or PTMs that stimulate endonuclease processing by Pso2 and SNM1A? As discussed in **1.4.2**, mass-spectroscopy experiments have identified phosphorylation, SUMOylation and ubiquitination of SNM1A in response to DNA

damage. Additionally, direct phosphorylation of Pso2 and SNM1A has been observed by Sak1 and ATM, respectively[184,334]. Future work is needed to determine any functional consequences of these PTMs *in vitro* and *in vivo*. Uncovering regulatory changes which license SNM1A/ Pso2 endonuclease activity will allow us to determine when, where and under what circumstances it may be functioning in repair.

Perhaps most importantly, future work should evaluate how our findings in yeast translate to human ICL repair. To confirm whether SNM1A also functions as both an exonuclease and endonuclease in repair, we need to complete complementation assays to probe the ability of SNM1A separation-of-function mutants to rescue knockout-associated cellular phenotypes. Like the yeast experiments presented here, monitoring the rescue of cellular defects associated with SNM1A knockout will clarify which activities are necessary for repair. Notably, quantifying the accumulation of chromosomal aberrations following treatment with ICL-inducing drugs may provide a clearer metric for observing rescue, as SNM1A knockout prompts a limited survival hypersensitivity. Because SNM1A is the functional homolog of Pso2, we anticipate our findings will be consistent in human repair. Regardless, directly demonstrating that SNM1A exonuclease and endonuclease activity are essential for repair will significantly expand the proposed role of SNM1A.

Interstrand crosslinks induce severe stress on cellular processes requiring strand separation. In response, cells have evolved a complex web of redundant, competing repair pathways responsible for the rapid recognition and removal of ICLs. SNM1A is a critical dual-function nuclease which contributes to repair through an unclear mechanism. While many questions remain regarding the role of SNM1A, the work presented in this thesis has significantly advanced our understanding of SNM1A action. *In vitro* biochemical characterization determined substrate preferences of SNM1A endonuclease and exonuclease activity. Yeast models indicated that both nuclease functions are necessary for repair. Finally, *in vitro* and *in silico* high-throughput screens identified novel small molecule inhibitors of SNM1A. These compounds exhibit improved potency and specificity relative to previously reported inhibitors. Collectively,

this work describes expanded roles for SNM1A in ICL repair and developed tools to begin exploiting SNM1A dependency for improving chemosensitivity.

Supplemental Information

Buffer Recipes

Buffer Name	Components
Formamide Buffer	95% Formamide, 5 mM EDTA
Elution Buffer	200 mM NaCl, 10 mM Tris pH 7.5, 5 mM EDTA
Annealing Buffer	100 mM NaCl, 10 mM Tris pH 7.5, 10 mM $MgCl_2$
EMSA Buffer	50 mM Tris-acetate pH 7.2, 75 mM potassium acetate, 1 mM DTT and 100 µg/mL BSA
EMSA Loading Buffer	10 mM EDTA, 50% v/v glycerol
Nuclease Reaction Buffer	50 mM Tris-acetate pH 7.2, 75 mM potassium acetate, 10 mM $MgCl_2$, 1 mM DTT and 100 µg/mL BSA
ECL Solution	100 mM Tris pH 8.8, 2 mM 4-iodophenylboronic acid, 1.25 mM luminol and 5.4 mM H_2O_2
Yeast Wash Buffer	100 mM lithium acetate, 10 mM Tris pH 8.0, 1 mM EDTA
bNiA	500 mM NaCl, 50 mM Tris pH 7.5 or 50 mM Sodium Phosphate pH 6.8, 10% glycerol, 0.5 mM TCEP
iNiA	500 mM NaCl, 50 mM HEPES pH 7.5, 5% glycerol, 1 mM TCEP, 10 mM Imidazole
NiA^{S1B}	500 mM NaCl, 50 mM Tris pH 7.5, 10% v/v glycerol, 0.5 mM TCEP, 10 mM imidazole
NiA^{FAN1}	500 mM NaCl, 50 mM Tris pH 8.0, 5 mM imidazole
bNiB	500 mM NaCl, 50 mM Tris pH 7.5 or 50 mM Sodium Phosphate pH 6.8, 10% glycerol, 0.5 mM TCEP, 300 mM imidazole
iNiA	500 mM NaCl, 50 mM HEPES pH 7.5, 5% glycerol, 1 mM TCEP, 300 mM Imidazole
NiB^{S1B}	500 mM NaCl, 50 mM Tris pH 7.5, 10% v/v glycerol, 0.5 mM TCEP, 300 mM imidazole
NiB^{FAN1}	500 mM NaCl, 50 mM Tris pH 8.0, 300 mM imidazole
SA	50mM Tris pH 7.5 or 50 mM Sodium Phosphate pH 6.8, 10% glycerol, 0.5mM TCEP
SB	1 M NaCl, 50mM Tris pH 7.5 or 50 mM Sodium Phosphate pH 6.8, 10% glycerol, 0.5mM TCEP
SNM1A Storage Buffer	200mM NaCl, 50mM Tris pH 7.5 or Sodium Phosphate pH 6.8, 5% glycerol, 0.5mM TCEP
FAN1 Storage Buffer	150 mM NaCl, 20 mM Tris pH 8.0, 10% glycerol
Pso2 Storage Buffer	300 mM NaCl, 25 mM Sodium Phosphate pH 6.8, 5% glycerol, 0.5 mM TCEP

Primer Information

Table S.1. Sequence information for cloning primers used in all experiments. Underlined nucleotides anneal to template sequence.

Primer Name	Nucleotide Sequence (5-3')
N697.SNM1A-F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAGATTACGATATCCCAACGACCGAAAACCT GTATTTTCAGAGCAAAAAAAACCTGCCCGTTCTACAAA
N697.SNM1A-R	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTATTAATAACCGGCTTCCAGTTTCCA
C336.SNM1B-F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAGATTACGATATCCCAACGACCGAAAACCT GTATTTTCAGTCCATGAATGGGGTCCTGATCCCCCAT
C336.SNM1B-R	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAGCTTGGTTTTCTAGAGGAAGAACTCAT
N210.PSO2-F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAGAAAACTGTATTTTCAGGGCAAACTTGTT TTACCAAGTTTCAAGATAATTAAGTTCAATAATGG
N210.PSO2-R	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTATTTAGCCGCCCGCGTTTTCCTAACG
FL.PSO2-F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAAAAATGTCTTCAAGGAAATCTATAGTGCA AATAAGAAGA
FL.PSO2-R	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTTAGCCGCCCGCGTTTTCCT

Table S.2. Sequence information for sequencing primers used in all experiments.

Primer Name	Nucleotide Sequence (5-3')
pDONR201-F	TTAACGCTAGCATGGATCT
pDONR201-R	AACATCAGAGATTTTGAGACAC
pUC/M13-F	CCCAGTCACGACGTTGTAAAACG
pUC/M13-R	AGCGGATAACAATTTCACACAGG
pFB-F	TATTCATACCGTCCCACCA
pFB-R	GGGAGGTTTTTTAAAGCAAGTAAA
PSO2.98-F	GAACACACATACATCCTCT
PSO2.663-F	GTTCAATAATGGCCACG
PSO2.1349-F	CATCAAGATATGTGAATTTTTG

Table S.3. Sequence information for primers used in SNM1A mutagenesis experiments. Superscripts indicate primers used sequentially to generate combination mutations.

Primer Name	Nucleotide Sequence (5-3')
D736A/H737A-F	ATTCCGCTGCTTATGCCGGTCTGTCCAAACATTTTACCT
D736A/H737A-R	CGGCATAAGCAGCGGAATGAAAATGGGTCAGAAAATATGCG
Q718A-F	TGCATTTGCGTATGGTGTTGTTGAAGGTTGTACCG
Q718A-R	CACCATACGCAAATGCATCAACGGTAAAACCGG
T840F-F[1]	CTGGATACCTTCTATTGTAGTCCGGAATATACATTTCCGA
T840F-R[1]	TACAATAGAAGGTATCCAGATACAGCATGTGAACT
T840W-F	CTGGATACCTGGTATTGTAGTCCGGGAATATACATTTCCGA
T840W-R	ACTACAATACCAGGTATCCAGATACAGCATGTGAACT
Y841F-F[2,3]	ACCACCTTTTGTAGTCCGGAATATACATTTC
Y841F-R[2,3]	GACTACAAAAGGTGGTATCCAGATACAGC
Y879A-F	GCACCGCTAGCATTGGTAAAGAAAAAGTGTTTCTG
Y879A-R	GCTAGCGGTGCCACAAACAACCAGTGCATG
K883A-F[2,3]	GCATTGGTGCAGCCCCCGTGTTTCTGGCCATTG
K883A-R[2,3]	CACTTTTTCTGCACCAATGCTATAGGTGCCAC
H930A-F	GCAGAGCGCCCTGAAAAAATGTGGTGGCAA
H930A-R	TTTTTTCAGGGCGCTCTGCAGACCTTTAAAGTTAATCTGCA
P961G-F	GCATTTCGTGGGACCGGTTGGACCCATAG
P961G-R	AACCGGTCCCACGAAATGCCAGAATC
T962A-F	CGTCCGGCCGGTTGGACCCATAGC
T962A-R	CCAACCGGCCGGACGAAATGCCAG
T962F-F	CGTCCGATTTGTTGGACCCATAGCAACAAATTCACC
T962F-R	TCCAACCAAACGGACGAAATGCCAGAATCTGGTT
W964A-F	CCGGTGCAACCCATAGCAACAAATTCACCCG
W964A-R	TATGGGTTGCACCGGTCGGACGAAATGCCAGAA
S992A-F[2,3]	CCGTATGCCGAACATAGCAGCTATCTGGAAA
S992A-R[2,3]	TATGTTCGGCATACGGGATACCATAAATGCTAATG
E993A-F	CGTATAGCGCACATAGCAGCTATCTGGAAATG
E993A-R	GCTGCTATGTGCGCTATACGGGATACC
E993Q-F	CGTATAGCCAACATAGCAGCTATCTGGAAATG
E993Q-R	GCTGCTATGTTGGCTATACGGGATACC
T1015W-F	ATCCCGTGGGTTAATGTTGGCACCTGG
T1015W-R	AACATTAACCCACGGGATAATTTTCTGCGGTTTCAG
V1016W-F	CCCGACCTGGAATGTTGGCACCTGGAAAAG
V1016W-R	CAACATTCCAGGTCGGGATAATTTTCTGCGGTT
S1023W-F	TGGAAATGGCGTAGCACCATGGAAAAATACTT
S1023W-R	GCTACGCCATTTCCAGGTGCCAACAT
K904A/K906T-F	CCAAGAAGCATACACAACCCTGCAGTGTCTGAA
K904A/K906T-R	GCAGGGTTGTGTATGCTTCTTGGCTCATACCAAC
C925S-F[4]	GATATGTCTAGCAGCCTGGTGCATCTGCTGCC

C925S-R[4]	GGCTGCTAGACATATCGGTGGTAATCAGGCTGTTA
C949S-F[4]	CCTGAAAAAATCTGGTGGCAAATATAACCAGATTC
C949S-R[4]	CACCAGATTTTTTCAGGTGGCTCTGCAGACCTTTAA
R960A/T962A-F	GCATTTGCTCCGGCCGGTTGGACCCAT
R960A/T962A-R	CGGCCGGAGCAAATGCCAGAATCTGGTTATATTTGC
Y704F/K705A-F[1,5]	CCGTTCTTCGCTAAAATCCCTGGCACCGGTTTTACC
Y704F/K705A-R[1,5]	AGGGATTTTAGCGAAGAACGGGCAGGTTTTTTTGCCC
Y719F-F[1,5]	GCATTTCAGTTTGGTGTTGTTGAAGGTTGTACCGCAT
Y719F-R[1,5]	CAACACCAAACTGAAATGCATCAACGGTAAAACCG
E845A/Y846F-F[6]	TCCGGCATTTACATTTCCGAGCCAGCAAGAAG
E845A/Y846F-R[6]	GAAATGTAAATGCCGGACTACAATAGGTGGTATCCAGAT
N1017A-F[6]	CCGTTGCTGTTGGCACCTGGAAAAGCC
N1017A-R[6]	CCAACAGCAACGGTCGGGATAATTTTCTGCGG
Y1030F-F[6]	GGAAAAATTCTTTCGTGAATGGAAACTGGAAGCCGG
Y1030F-R[6]	ACGAAAGAATTTTTCCATGGTGCTACGGCTT
K969A-F[7]	GCAACGCATTCACCCGTATTGCCGATGTTA
K969A-R[7]	GGTGAATGCGTTGCTATGGGTCCAACCGGTC
K1022A-F[7]	ATGGAAGCATACTTTCGTGAATGGAAACTGGAAGCC
K1022A-R[7]	CGAAGTATGCTTCCATGGTGCTACGGCTTT

Table S.4. Sequence information for primers used in Pso2 mutagenesis experiments. Superscripts indicate primers used sequentially to generate combination mutations.

Primer Name	Nucleotide Sequence (5-3')
D252A/H253A-F	TTTTCATTCTGCTGCCTATATTGGACTCAAAAAATCTTGG
D252A/H253A-R	CCAATATAGGCAGCAGAATGAAAATGAGAAAGAAAGTAC
K446A-F	ATTGGAGCAGAGAAGCTAGCCATCAAGATATGTGAATTTTT
K446A-R	CTTCTCTGCTCCAATAGTATATGTTCCCACAAGAAATAGGACCC
S609A-F	CCATATGCTGAACACAGTAGTTTTAACGATTTAGTGAAATTTGGTTGT
S609A-R	TGTGTTCAGCATATGGTACGTTAAAAACTTGAAATTTGTTGTACT
T378F-F	TAGATACATTGTATATGACTATGGGATACAACTTTCCTTCACAACATTCTG
T378F-R	AGTCATATAGAATGTATCTAAATAAACTTGGTCAATAGTTTCATTCGCTGTC
V633F-F	CCCACTTTCAATCTAAACAATTTATGGAAAGTGAGGTATATGACG
V633F-R	TTTAGATTGAAAGTGGGAATGACTTCAGAACACTTCAATTTACAACCAA

DNA Oligonucleotide Information

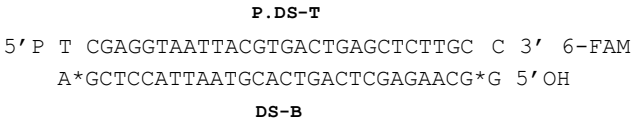
Table S.5. Sequence and modification information for all oligos used in above experiments. * denotes phosphorothioate. †, ‡, ¶, § Indicates oligos that anneal to form complex substrates (letter superscript to indicate alternate oligos in a set).

Oligo Name	Nucleotide Sequence	Modification(s)
P.DS-T†	TCGAGGTAATTACGTGACTGAGCTCTTGCC	5’ P, 3’ 6-FAM
DS-B†	G*GCAAGAGCTCAGTCACGTAATTACCTCG*A	
P.20TF	TTTTTTTTTTTTTTTTTTT	5’ P, 3’ 6-FAM
OH.20TF	TTTTTTTTTTTTTTTTTTT	3’ 6-FAM
F.20T	TTTTTTTTTTTTTTTTTTT	5’ 6-FAM
F.ICL-T$^{\ddagger a}$	T*T*TTTTTTTTTTTTTATAATTTTAATTTGATCATTTATTATAAATTTTATTAT*A*T	5’ 6-FAM
P.ICLF-T$^{\ddagger b}$	TTTTTTTTTTTTTTTATAATTTTAATTTGATCATTTATTATAAATTTTATTAT*A*T	5’ P, 3’ 6-FAM
OH.ICLF-T$^{\ddagger c}$	T*T*TTTTTTTTTTTTTATAATTTTAATTTGATCATTTATTATAAATTTTATTAT*A*T	3’ 6-FAM
ICL-B‡	A*T*ATAATAAAATTTATAATAAATGATCAAATTAAAATTATCATATTTATAACA*C*C	
ICL-F‡	G*G*TGTTATAAATA*T*G	
P.BX-T¶	TCGAGGTAATTACGTGACTGAGCTCTTGCC	5’ P, 3’ 6-FAM
BX-B¶	G*GCAAGAGCTCAGTCACGTAATTACCTCGATTAAACGTACGTACTAAGC*C	
2G.BX-T$^{\P b}$	G*GCTTAGTACGTACGTT*T	
1G.BX-T$^{\P c}$	G*GCTTAGTACGTACGTTT*A	
NG.BX-T$^{\P d}$	G*GCTTAGTACGTACGTTTA*A	
F.6T	TTTTTT	5’ 6-FAM
F.8T	TTTTTTTT	5’ 6-FAM
F.10T	TTTTTTTTTT	5’ 6-FAM
F.12T	TTTTTTTTTTTT	5’ 6-FAM
F.14T	TTTTTTTTTTTTTT	5’ 6-FAM
HP.0T	TGGACTCTAGAACACCGGCGTTCTAGAGTCCA	5’ 6-FAM
HP.3T	TGGACTCTAGAACGCCTTTGGCGTTCTAGAGTCCA	5’ 6-FAM
HP.5T	TGGACTCTAGAACGCCTTTTTGGCGTTCTAGAGTCCA	5’ 6-FAM
HP.10T	TGGACTCTAGAACGCCTTTTTTTTTTGGCGTTCTAGAGTCCA	5’ 6-FAM
F.NG-T$^{\S a}$	G*AGGGCGAGCCCGATTTTTTCCGCTAAGATTTTTTGCAGATACTTAAC*A*C	5’ 6-FAM
F.3TG-T$^{\S b}$	G*AGGGCGAGCCCGATTTTTTCCGCTTTTAAGATTTTTTGCAGATACTTAAC*A*C	5’ 6-FAM
F.5TG-T$^{\S c}$	G*AGGGCGAGCCCGATTTTTTCCGCTTTTTTAAGATTTTTTGCAGATACTTAACA*C	5’ 6-FAM
F.10NG-T$^{\S d}$	GAGGGCGAGCCCGATTTTTTCCGCTTGACCCAAGTAAGATTTTTTGCAGATACTTAACAC	5’ 6-FAM
G-B1§	A*GCGGAAAAAATCGGGCTCGCCCT*C	
G-B2§	G*TGTTAAGTATCTGCAAAAAATCT*T	
P.1F$^{\&}$	/T-F/TGACACACATTCGTAA	5’ P
F.AB-T^{+}	GTTAAGTTGGGATTG-[spacer]-GACACACATTCGTAA	3’ 6-FAM
AB-B$^{\&, +, \|\|}$	TTACGAATGTGTGTCACAATCCCAACTTAAC	
F.(6)Flap-T$^{\|\|}$	CTAACTTGACACACATTCGTAA	3’ 6-FAM
(1)Flap-T$^{\&, \|\|}$	GTTAAGTTGGGATTGG	

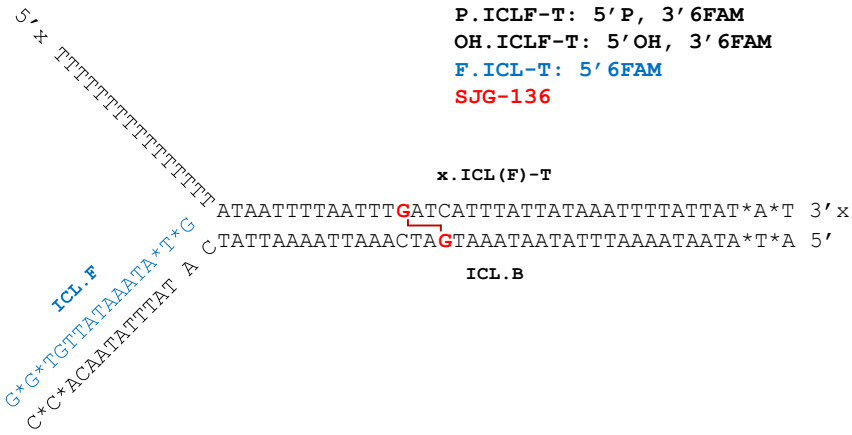
DNA Substrate Diagrams

P.DS

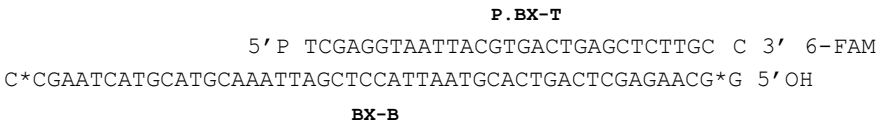
*P.DS used in Figure 3.7 did not contain a 6-FAM label.



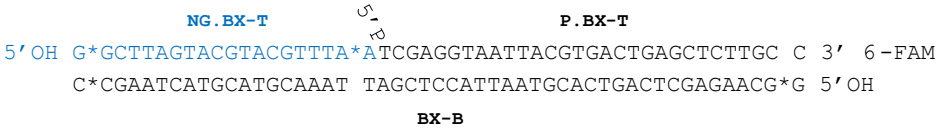
ICL.Fork Series



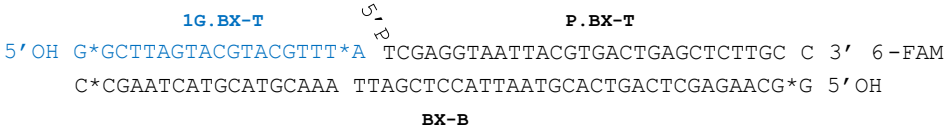
3′ Overhang



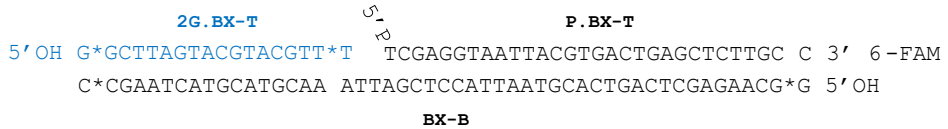
P.Nicked



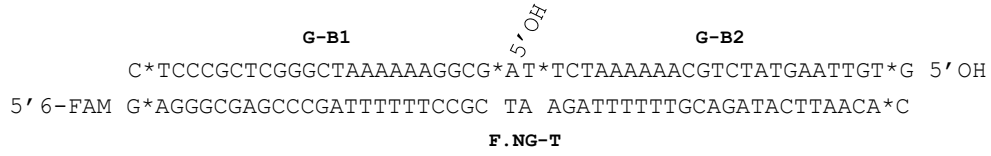
1nt P.Gap



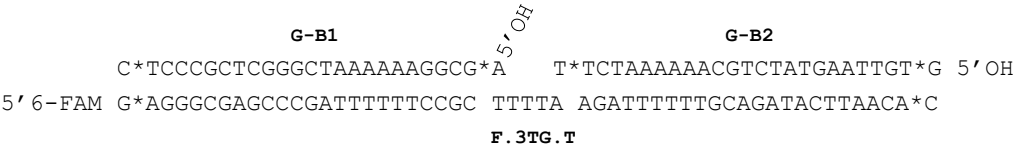
2nt P.Gap



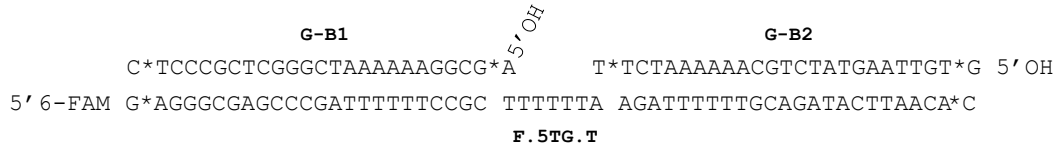
OH.Nicked



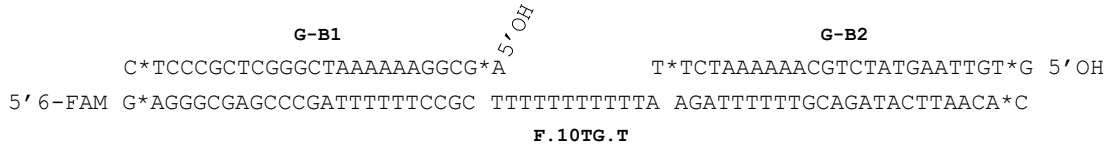
3T.Gap



5T.Gap

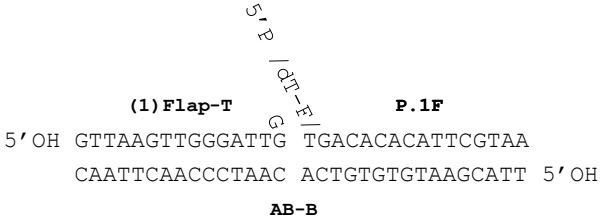


10T.Gap

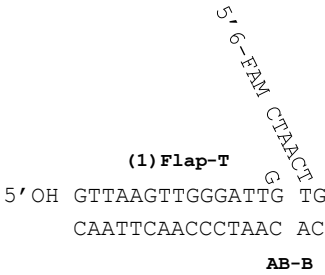


Gapped.DS

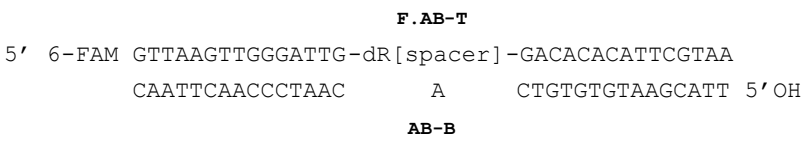
P.(1)Flap



F.(6)Flap


5'-6-FAM CTAACT
G
(1)Flap-T
F.(6)Flap.T
5'OH GTTAAGTTGGGATTG TGACACACATTCGTAA
CAATTCAACCCTAAC ACTGTGTGTAAGCATT 5'OH
AB-B

F.AB


F.AB-T
5' 6-FAM GTTAAGTTGGGATTG-dR[spacer]-GACACACATTCGTAA
CAATTCAACCCTAAC A CTGTGTGTAAGCATT 5'OH
AB-B

References

1. Dronkert, M. L. G. & Kanaar, R. Repair of DNA interstrand cross-links. *Mutat. Res. Repair* **486**, 217–247 (2001).

2. Deans, A. J. & West, S. C. DNA interstrand crosslink repair and cancer. *Nat. Rev. Cancer* **11**, 467–480 (2011).

3. Smeaton, M. B. *et al.* Distortion-dependent unhooking of interstrand cross-links in mammalian cell extracts. *Biochemistry* **47**, 9920–9930 (2008).

4. Hlavin, E. M., Smeaton, M. B., Noronha, A. M., Wilds, C. J. & Miller, P. S. Cross-link structure affects replication-independent DNA interstrand cross-link repair in mammalian cells. *Biochemistry* **49**, 3977–3988 (2010).

5. Noll, D. M., Mason, T. M. & Miller, P. S. Formation and Repair of Interstrand Cross-Links in DNA. *Chem. Rev.* **106**, 277–301 (2006).

6. Semlow, D. R. & Walter, J. C. Mechanisms of Vertebrate DNA Interstrand Cross-Link Repair. *Annu. Rev. Biochem.* **90**, 107–135 (2021).

7. Guainazzi, A. & Schärer, O. D. Using synthetic DNA interstrand crosslinks to elucidate repair pathways and identify new therapeutic targets for cancer chemotherapy. *Cell. Mol. Life Sci.* **67**, 3683–3697 (2010).

8. Lopez-Martinez, D., Liang, C. C. & Cohn, M. A. Cellular response to DNA interstrand crosslinks: the Fanconi anemia pathway. *Cellular and Molecular Life Sciences* vol. 73 (2016).

9. Fu, Y. V. *et al.* Selective bypass of a lagging strand roadblock by the eukaryotic replicative DNA helicase. *Cell* **146**, 931–941 (2011).

10. Kottemann, M. C. & Smogorzewska, A. Fanconi anaemia and the repair of Watson and Crick DNA crosslinks. *Nat. 2013 4937432* **493**, 356–363 (2013).

11. Wang, M. *et al.* Identification of DNA Adducts of Acetaldehyde. *Chem. Res. Toxicol.* **13**, 1149–1157 (2000).

12. Kozekov, I. D. *et al.* DNA interchain cross-links formed by acrolein and crotonaldehyde. *J. Am. Chem. Soc.* **125**, 50–61 (2003).

13. Stone, M. P. *et al.* Interstrand DNA cross-links induced by α,β-unsaturated aldehydes derived from lipid peroxidation and environmental sources. *Acc. Chem. Res.* **41**, 793–804 (2008).

14. Theruvathu, J. A., Jaruga, P., Nath, R. G., Dizdaroglu, M. & Brooks, P. J. Polyamines stimulate the formation of mutagenic 1,N2-propanodeoxyguanosine adducts from acetaldehyde. *Nucleic Acids Res.* **33**, 3513–3520 (2005).

15. Mechilli, M., Schinoppi, A., Kobos, K., Natarajan, A. T. & Palitti, F. DNA repair

deficiency and acetaldehyde-induced chromosomal alterations in CHO cells. *Mutagenesis* **23**, 51–56 (2008).

16. Marietta, C., Thompson, L. H., Lamerdin, J. E. & Brooks, P. J. Acetaldehyde stimulates FANCD2 monoubiquitination, H2AX phosphorylation, and BRCA1 phosphorylation in human cells in vitro: implications for alcohol-related carcinogenesis. *Mutat. Res.* **664**, 77–83 (2009).

17. Garaycoechea, J. I. *et al.* Alcohol and endogenous aldehydes damage chromosomes and mutate stem cells. *Nat. 2018 5537687* **553**, 171–177 (2018).

18. Langevin, F., Crossan, G. P., Rosado, I. V., Arends, M. J. & Patel, K. J. Fancd2 counteracts the toxic effects of naturally produced aldehydes in mice. *Nature* **475**, 53–59 (2011).

19. Price, N. E. *et al.* Interstrand DNA-DNA cross-link formation between adenine residues and abasic sites in duplex dna. *J. Am. Chem. Soc.* **136**, 3483–3490 (2014).

20. Housh, K. *et al.* Formation and Repair of an Interstrand DNA Cross-Link Arising from a Common Endogenous Lesion. *J. Am. Chem. Soc.* **143**, 15344–15357 (2021).

21. Klein, D. *Organic Chemistry*. (John Wiley & Sons, Ltd, 2012).

22. Hadjur, S. *et al.* Defective hematopoiesis and hepatic steatosis in mice with combined deficiencies of the genes encoding Fancc and Cu/Zn superoxide dismutase. *Blood* **98**, 1003–1011 (2001).

23. Garaycoechea, J. I. *et al.* Genotoxic consequences of endogenous aldehydes on mouse haematopoietic stem cell function. *Nature* **489**, 571–575 (2012).

24. Huang, H., Zhu, L., Reid, B. R., Drobny, G. P. & Hopkins, P. B. Solution Structure of a Cisplatin-Induced DNA Interstrand Cross-Link. *Science (80-.).* **270**, 1842 (1995).

25. Sastry, M., Fiala, R., Lipman, R., Tomasz, M. & Patel, D. J. Solution Structure of the Monoalkylated Mitomycin C–DNA Complex. *J. Mol. Biol.* **247**, 338–359 (1995).

26. Spielmann, H. P., Dwyer, T. J., Hearst, J. E. & Wemmer, D. E. Solution Structures of Psoralen Monoadducted and Cross-Linked DNA Oligomers by NMR Spectroscopy and Restrained Molecular Dynamics. *Biochemistry* **34**, 12937–12953 (1995).

27. Kellum, A. H. *et al.* Structure of a Stable Interstrand DNA Cross-Link Involving a β-N-Glycosyl Linkage between an N6-dA Amino Group and an Abasic Site. *Biochemistry* **60**, 41–52 (2021).

28. Macé-Aimé, G., Couvé, S., Khassenov, B., Rosselli, F. & Saparbaev, M. K. The Fanconi anemia pathway promotes DNA glycosylase-dependent excision of interstrand DNA crosslinks. *Environ. Mol. Mutagen.* **51**, 508–519 (2010).

29. Semlow, D. R., Zhang, J., Budzowska, M., Drohat, A. C. & Walter, J. C. Replication-Dependent Unhooking of DNA Interstrand Cross-Links by the NEIL3 Glycosylase.

Cell **167**, 498-511.e14 (2016).

30. Zhang, N., Lu, X., Zhang, X., Peterson, C. A. & Legerski, R. J. hMutSβ Is Required for the Recognition and Uncoupling of Psoralen Interstrand Cross-Links In Vitro. *Mol. Cell. Biol.* **22**, 2388–2397 (2002).

31. Kato, N. *et al.* Sensing and Processing of DNA Interstrand Crosslinks by the Mismatch Repair Pathway. *Cell Rep.* **21**, 1375–1385 (2017).

32. Wang, X. *et al.* Involvement of Nucleotide Excision Repair in a Recombination-Independent and Error-Prone Pathway of DNA Interstrand Cross-Link Repair. *Mol. Cell. Biol.* **21**, 713–720 (2001).

33. Muniandy, P. A., Thapa, D., Thazhathveetil, A. K., Liu, S. ting & Seidman, M. M. Repair of laser-localized DNA interstrand cross-links in G1 phase mammalian cells. *J. Biol. Chem.* **284**, 27908–27917 (2009).

34. Zheng, H., Wang, X., Legerski, R. J., Glazer, P. M. & Li, L. Repair of DNA interstrand cross-links: interactions between homology-dependent and homology-independent pathways. *DNA Repair (Amst).* **5**, 566–574 (2006).

35. Benitez, A. *et al.* FANCA Promotes DNA Double-Strand Break Repair by Catalyzing Single-Strand Annealing and Strand Exchange The single-strand annealing activity of FANCA plays a direct role in DSB repair Molecular Cell FANCA Promotes DNA Double-Strand Break Repair by Catalyz. *Mol. Cell* **71**, 621–628 (2018).

36. Williams, H. L., Gottesman, M. E. & Gautier, J. Replication-Independent Repair of DNA Interstrand Crosslinks. *Mol. Cell* **47**, 140–147 (2012).

37. Budzowska, M., Graham, T. G., Sobeck, A., Waga, S. & Walter, J. C. Regulation of the Rev1–pol ζ complex during bypass of a DNA interstrand cross-link . *EMBO J.* **34**, 1971–1985 (2015).

38. Li, L., Peterson, C. A., Lu, X., Wei, P. & Legerski, R. J. Interstrand Cross-Links Induce DNA Synthesis in Damaged and Undamaged Plasmids in Mammalian Cell Extracts. *Mol. Cell. Biol.* **19**, 5619–5630 (1999).

39. Räschle, M. *et al.* Mechanism of replication-coupled DNA interstrand crosslink repair. *Cell* **134**, 969–980 (2008).

40. Long, D. T. *et al.* Mechanism of RAD51-dependent DNA interstrand cross-link repair. *Science (80-.).* **333**, 84–7 (2011).

41. Hlavin, E. M., Smeaton, M. B. & Miller, P. S. Initiation of DNA Interstrand Cross-link Repair in Mammalian Cells. *Environ. Mol. Mutagen.* **51**, 604 (2010).

42. Williams, H. L., Gottesman, M. E. & Gautier, J. The differences between ICL repair during and outside of S phase. *Trends Biochem. Sci.* **38**, 386–393 (2013).

43. Marteijn, J. A., Lans, H., Vermeulen, W. & Hoeijmakers, J. H. J. Understanding

nucleotide excision repair and its roles in cancer and ageing. *Nat. Rev. Mol. Cell Biol.* (2014) doi:10.1038/nrm3822.

44. Schärer, O. D. Nucleotide Excision Repair in Eukaryotes. *Encycl. Biol. Chem. Second Ed.* 341–344 (2013) doi:10.1016/B978-0-12-378630-2.00491-6.

45. Sugasawa, K. *et al.* A multistep damage recognition mechanism for global genomic nucleotide excision repair. *Genes Dev.* **15**, 507–521 (2001).

46. Maillard, O., Camenisch, U., Clement, F. C., Blagoev, K. B. & Naegeli, H. DNA repair triggered by sensors of helical dynamics. *Trends Biochem. Sci.* **32**, 494–499 (2007).

47. Cheon, N. Y., Kim, H. S., Yeo, J. E., Schärer, O. D. & Lee, J. Y. Single-molecule visualization reveals the damage search mechanism for the human NER protein XPC-RAD23B. *Nucleic Acids Res.* **47**, 8337–8347 (2019).

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

49. Compe, E. & Egly, J. M. TFIIH: When transcription met DNA repair. *Nat. Rev. Mol. Cell Biol.* **13**, 343–354 (2012).

50. Tsutakawa, S. E. *et al.* Envisioning how the prototypic molecular machine TFIIH functions in transcription initiation and DNA repair. *DNA Repair (Amst).* **96**, (2020).

51. Topolska-Woś, A. M. *et al.* A key interaction with RPA orients XPA in NER complexes. *Nucleic Acids Res.* **48**, 2173–2188 (2020).

52. Li, L., Peterson, C. A., Lu, X. & Legerski, R. J. Mutations in XPA that prevent association with ERCC1 are defective in nucleotide excision repair. *Mol. Cell. Biol.* **15**, 1993–1998 (1995).

53. Wakasugi, M. & Sancar, A. Assembly, subunit composition, and footprint of human DNA repair excision nuclease. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 6669–6674 (1998).

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

55. Staresincic, L. *et al.* Coordination of dual incision and repair synthesis in human nucleotide excision repair. *EMBO J.* **28**, 1111–1120 (2009).

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

57. Giannattasio, M. *et al.* Exo1 competes with repair synthesis, converts NER intermediates to long ssDNA gaps, and promotes checkpoint activation. *Mol. Cell* **40**, 50–62 (2010).

58. Saldivar, J. C., Cortez, D. & Cimprich, K. A. The essential kinase ATR: Ensuring faithful duplication of a challenging genome. *Nat. Rev. Mol. Cell Biol.* **18**, 622–636 (2017).

59. Vermeulen, W. & Fousteri, M. Mammalian Transcription-Coupled Excision Repair. *Cold Spring Harb Perspect Biol.* **5**, (2013).

60. Fousteri, M. & Mullenders, L. H. F. Transcription-coupled nucleotide excision repair in mammalian cells: Molecular mechanisms and biological effects. *Cell Res.* **18**, 73–84 (2008).

61. Leadon, S. A. & Lawrence, D. A. Preferential repair of DNA damage on the transcribed strand of the human metallothionein genes requires RNA polymerase II. *Mutat. Res. Repair* **255**, 67–78 (1991).

62. Duan, M., Speer, R. M., Ulibarri, J., Liu, K. J. & Mao, P. Transcription-coupled nucleotide excision repair: New insights revealed by genomic approaches. *DNA Repair (Amst).* **103**, (2021).

63. Lainé, J.-P. & Egly, J.-M. Initiation of DNA repair mediated by a stalled RNA polymerase IIO. *EMBO J.* **25**, 387–397 (2006).

64. Daniel, L. *et al.* Mechanistic insights in transcription-coupled nucleotide excision repair of ribosomal DNA. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E6770–E6779 (2018).

65. Sigurdsson, S., Dirac-Svejstrup, A. B. & Svejstrup, J. Q. Evidence that Transcript Cleavage Is Essential for RNA Polymerase II Transcription and Cell Viability. *Mol. Cell* **38**, 202–210 (2010).

66. Son, K. & Schärer, O. D. Repair, Removal, and Shutdown: It All Hinges on RNA Polymerase II Ubiquitylation. *Cell* **180**, 1039–1041 (2020).

67. Sarker, A. H. *et al.* Recognition of RNA polymerase II and transcription bubbles by XPG, CSB, and TFIIH: Insights for transcription-coupled repair and Cockayne syndrome. *Mol. Cell* **20**, 187–198 (2005).

68. Cho, I., Tsai, P. F., Lake, R. J., Basheer, A. & Fan, H. Y. ATP-Dependent Chromatin Remodeling by Cockayne Syndrome Protein B and NAP1-Like Histone Chaperones Is Required for Efficient Transcription-Coupled DNA Repair. *PLoS Genet.* **9**, (2013).

69. Spyropoulou, Z. *et al.* Cockayne syndrome group B (CSB): The regulatory framework governing the multifunctional protein and its plausible role in cancer. *Cells* **10**, (2021).

70. Iyama, T. & Wilson, D. M. Elements That Regulate the DNA Damage Response of Proteins Defective in Cockayne Syndrome. *J. Mol. Biol.* **428**, (2016).

71. Van Den Boom, V. *et al.* DNA damage stabilizes interaction of CSB with the transcription elongation machinery. *J. Cell Biol.* **166**, 27–36 (2004).

72. van der Weegen, Y. *et al.* The cooperative action of CSB, CSA, and UVSSA target TFIIH to DNA damage-stalled RNA polymerase II. *Nat. Commun.* **11**, 1–16 (2020).

73. Lans, H., Hoeijmakers, J. H. J., Vermeulen, W. & Marteijn, J. A. The DNA damage response to transcription stress. *Nat. Rev. Mol. Cell Biol.* **20**, 766–784 (2019).

74. Xu, J. *et al.* Structural basis for the initiation of eukaryotic transcription-coupled DNA repair. *Nat. Publ. Gr.* (2017) doi:10.1038/nature24658.

75. Liebelt, F. *et al.* Transcription-coupled nucleotide excision repair is coordinated by ubiquitin and SUMO in response to ultraviolet irradiation. *Nucleic Acids Res.* **48**, 231–248 (2020).

76. Rothfuss, A. & Grompe, M. Repair Kinetics of Genomic Interstrand DNA Cross-Links: Evidence for DNA Double-Strand Break-Dependent Activation of the Fanconi Anemia/BRCA Pathway. *Mol. Cell. Biol.* **24**, 123–134 (2004).

77. Wood, R. D. Mammalian nucleotide excision repair proteins and interstrand crosslink repair. *Environ. Mol. Mutagen.* NA-NA (2010) doi:10.1002/em.20569.

78. Shen, X. & Li, L. Mutagenic repair of DNA interstrand crosslinks. *Environ. Mol. Mutagen.* **51**, 493–499 (2010).

79. Mchugh, P. J., Sones, W. R. & Hartley, J. A. Repair of Intermediate Structures Produced at DNA Interstrand Cross-Links in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **20**, 3425–3433 (2000).

80. Sarkar, S., Davies, A. a, Ulrich, H. D. & McHugh, P. J. DNA interstrand crosslink repair during G1 involves nucleotide excision repair and DNA polymerase zeta. *EMBO J.* **25**, 1285–1294 (2006).

81. Rogers, C. M. *et al.* The yeast Hrq1 helicase stimulates Pso2 translesion nuclease activity and thereby promotes DNA interstrand crosslink repair. *J. Biol. Chem.* **295**, 8945–8957 (2020).

82. Bessho, T., Mu, D. & Sancar, A. Initiation of DNA interstrand cross-link repair in humans: the nucleotide excision repair system makes dual incisions 5' to the cross-linked base and removes a 22- to 28-nucleotide-long damage-free strand. *Mol. Cell. Biol.* **17**, 6822–6830 (1997).

83. Mu, D. *et al.* DNA Interstrand Cross-Links Induce Futile Repair Synthesis in Mammalian Cell Extracts. *Mol. Cell. Biol.* **20**, 2446–2454 (2000).

84. Smeaton, M. B. *et al.* Effect of cross-link structure on DNA interstrand cross-link repair synthesis. *Chem. Res. Toxicol.* **22**, 1285–1297 (2009).

85. Vasquez, K. M. Targeting and processing of site-specific DNA interstrand crosslinks. *Environ. Mol. Mutagen.* **51**, 527–539 (2010).

86. Thoma, B. S., Wakasugi, M., Christensen, J., Reddy, M. C. & Vasquez, K. M. Human

XPC-hHR23B interacts with XPA-RPA in the recognition of triplex-directed psoralen DNA interstrand crosslinks. *Nucleic Acids Res.* **33**, 2993–3001 (2005).

87. Mustra, D. J., Warren, A. J., Wilcox, D. E. & Hamilton, J. W. Preferential binding of human XPA to the mitomycin C-DNA interstrand crosslink and modulation by arsenic and cadmium. *Chem. Biol. Interact.* **168**, 159–168 (2007).

88. Mukherjee, A. & Vasquez, K. M. HMGB1 interacts with xpa to facilitate the processing of DNA interstrand crosslinks in human cells. *Nucleic Acids Res.* **44**, 1151–1160 (2016).

89. Wu, Q., Christensen, L. A., Legerski, R. J. & Vasquez, K. M. Mismatch repair participates in error-free processing of DNA interstrand crosslinks in human cells. *EMBO Rep.* **6**, 551–557 (2005).

90. Zhao, J., Jain, A., Iyer, R. R., Modrich, P. L. & Vasquez, K. M. Mismatch repair and nucleotide excision repair proteins cooperate in the recognition of DNA interstrand crosslinks. *Nucleic Acids Res.* **37**, 4420–4429 (2009).

91. Wu, Q. & Vasquez, K. M. Human MLH1 Protein Participates in Genomic Damage Checkpoint Signaling in Response to DNA Interstrand Crosslinks, while MSH2 Functions in DNA Repair. *PLOS Genet.* **4**, e1000189 (2008).

92. Pizzolato, J., Mukherjee, S., Schärer, O. D. & Jiricny, J. 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**, (2015).

93. Ben-Yehoyada, M. *et al.* Checkpoint Signaling from a Single DNA Interstrand Crosslink. *Mol. Cell* **35**, 704–715 (2009).

94. Bowen, N. *et al.* Reconstitution of long and short patch mismatch repair reactions using Saccharomyces cerevisiae proteins. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 18472–18477 (2013).

95. Zhang, J. *et al.* DNA interstrand cross-link repair requires replication-fork convergence. *Nat. Struct. Mol. Biol.* **22**, 242–247 (2015).

96. Liang, C. C. *et al.* UHRF1 is a sensor for DNA interstrand crosslinks and recruits FANCD2 to initiate the Fanconi anemia pathway. *Cell Rep.* **10**, 1947–1956 (2015).

97. Zheng, H. *et al.* Nucleotide Excision Repair- and Polymerase η-Mediated Error-Prone Removal of Mitomycin C Interstrand Cross-Links. *Mol. Cell. Biol.* **23**, 754–761 (2003).

98. Enoiu, M., Jiricny, J. & Schärer, O. D. Repair of cisplatin-induced DNA interstrand crosslinks by a replication-independent pathway involving transcription-coupled repair and translesion synthesis. *Nucleic Acids Res.* **40**, 8953–8964 (2012).

99. De Silva, I. U., McHugh, P. J., Clingen, P. H. & Hartley, J. A. Defining the Roles of Nucleotide Excision Repair and Recombination in the Repair of DNA Interstrand

Cross-Links in Mammalian Cells. *Mol. Cell. Biol.* **20**, 7980–7990 (2000).

100. Li, L., Peterson, C. A., Zhang, X. & Legerski, R. J. Requirement for PCNA and RPA in interstrand crosslink-induced DNA synthesis. *Nucleic Acids Res.* **28**, 1424–1427 (2000).

101. Thongthip, S., Conti, B. A., Lach, F. P. & Smogorzewska, A. Suppression of non-homologous end joining does not rescue DNA repair defects in Fanconi anemia patient cells. *Cell Cycle* **19**, 2553–2561 (2020).

102. Akkari, Y. M. N., Bateman, R. L., Reifsteck, C. A., Olson, S. B. & Grompe, M. DNA Replication Is Required To Elicit Cellular Responses to Psoralen-Induced DNA Interstrand Cross-Links. *Mol. Cell. Biol.* **20**, 8283–8289 (2000).

103. Schwab, R. A. *et al.* The Fanconi Anemia Pathway Maintains Genome Stability by Coordinating Replication and Transcription. *Mol. Cell* **60**, 351–361 (2015).

104. Niedernhofer, L. J., Lalai, A. S. & Hoeijmakers, J. H. J. Fanconi Anemia (Cross)linked to DNA Repair. *Cell* **123**, 1191–1198 (2005).

105. McHugh, P. J., Ward, T. A. & Chovanec, M. A prototypical Fanconi anemia pathway in lower eukaryotes? *Cell Cycle* **11**, 3739–3744 (2012).

106. Walden, H. & Deans, A. J. The Fanconi anemia DNA repair pathway: structural and functional insights into a complex disorder. *Annu. Rev. Biophys.* **43**, 257–278 (2014).

107. Xue, X., Sung, P. & Zhao, X. Functions and regulation of the multitasking FANCM family of DNA motor proteins. *Genes Dev.* **29**, 1777 (2015).

108. Swuec, P. *et al.* The FA Core Complex Contains a Homo-dimeric Catalytic Module for the Symmetric Mono-ubiquitination of FANCI-FANCD2. *Cell Rep.* **18**, 611–623 (2017).

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

110. van Twest, S. *et al.* Mechanism of Ubiquitination and Deubiquitination in the Fanconi Anemia Pathway. *Mol. Cell* **65**, 247–259 (2017).

111. Li, L., Tan, W. & Deans, A. J. Structural insight into FANCI–FANCD2 monoubiquitination. *Essays Biochem.* **64**, 807–817 (2020).

112. Huang, J. *et al.* The DNA Translocase FANCM/MHF Promotes Replication Traverse of DNA Interstrand Crosslinks. *Mol. Cell* **52**, 434–446 (2013).

113. Ceccaldi, R., Sarangi, P. & D, A. D. The Fanconi anaemia pathway: new players and new functions. (2016) doi:10.1038/nrm.2016.48.

114. Walter, J., Sun, L. & Newport, J. Regulated Chromosomal DNA Replication in the Absence of a Nucleus. *Mol. Cell* **1**, 519–529 (1998).

115. Yang, W., Seidman, M. M., Rupp, W. D. & Gao, Y. Replisome structure suggests mechanism for continuous fork progression and post-replication repair. *DNA Repair (Amst).* **81**, 102658 (2019).

116. Wu, R. A. *et al.* TRAIP is a master regulator of DNA interstrand crosslink repair. *Nature* **567**, 267–272 (2019).

117. Villa, F. *et al.* CUL2 LRR1 , TRAIP and p97 control CMG helicase disassembly in the mammalian cell cycle. *EMBO Rep.* **22**, (2021).

118. Wu, R. A., Pellman, D. S. & Walter, J. C. The Ubiquitin Ligase TRAIP: Double-Edged Sword at the Replisome. *Trends Cell Biol.* **31**, 75–85 (2021).

119. Fullbright, G., Rycenga, H. B., Gruber, J. D. & Long, D. T. p97 Promotes a Conserved Mechanism of Helicase Unloading during DNA Cross-Link Repair. *Mol. Cell. Biol.* **36**, 2983–2994 (2016).

120. Amunugama, R. *et al.* Replication Fork Reversal during DNA Interstrand Crosslink Repair Requires CMG Unloading. *Cell Rep.* **23**, 3419–3428 (2018).

121. Abdullah, U. B. *et al.* RPA activates the XPF-ERCC1 endonuclease to initiate processing of DNA interstrand crosslinks. *EMBO J.* **36**, 2047–2060 (2017).

122. Klein Douwel, D. *et al.* XPF-ERCC1 Acts in Unhooking DNA Interstrand Crosslinks in Cooperation with FANCD2 and FANCP/SLX4. *Mol. Cell* **54**, 460–471 (2014).

123. Knipscheer, P. *et al.* The Fanconi anemia pathway promotes replication-dependent DNA interstrand cross-link repair. *Science* **326**, 1698–1701 (2009).

124. Zhang, J. & Walter, J. C. Mechanism and regulation of incisions during DNA interstrand cross-link repair. *DNA Repair (Amst).* **19**, 135–142 (2014).

125. Klein Douwel, D., Hoogenboom, W. S., ACM Boonen, R. & Knipscheer, P. Recruitment and positioning determine the specific role of the XPF-ERCC1 endonuclease in interstrand crosslink repair. *EMBO J.* **36**, 2034–2046 (2017).

126. Kumaresan, K. R., Hwang, M., Thelen, M. P. & Lambert, M. W. Contribution of XPF Functional Domains to the 5‘ and 3‘ Incisions Produced at the Site of a Psoralen Interstrand Cross-Link. *Biochemistry* **41**, 890–896 (2001).

127. Hodskinson, M. R. G. *et al.* Mouse SLX4 Is a Tumor Suppressor that Stimulates the Activity of the Nuclease XPF-ERCC1 in DNA Crosslink Repair. *Mol. Cell* **54**, 472–484 (2014).

128. Roy, U. & Schärer, O. D. Involvement of translesion synthesis DNA polymerases in DNA interstrand crosslink repair. *DNA Repair (Amst).* **44**, 33–41 (2016).

129. Wang, A. T. *et al.* Human SNM1A and XPF-ERCC1 collaborate to initiate DNA interstrand cross-link repair. *Genes Dev.* **25**, 1859–1870 (2011).

130. Wang, R. *et al.* Mechanism of DNA Interstrand Crosslink Processing by Repair

Nuclease FAN1. *Science* **346**, 1127 (2014).

131. Ho, T. V., Guainazzi, A., Derkunt, S. B., Enoiu, M. & Schä Rer, O. D. Structure-dependent bypass of DNA interstrand crosslinks by translesion synthesis polymerases. *Nucleic Acids Res.* **39**, 7455–7464 (2011).

132. Krejci, L., Altmannova, V., Spirek, M. & Zhao, X. Homologous recombination and its regulation. *Nucleic Acids Res.* **40**, 5795 (2012).

133. Kaniecki, K., De Tullio, L. & Greene, E. C. A change of view: homologous recombination at single-molecule resolution. *Nat. Rev. Genet. 2017 194* **19**, 191–207 (2017).

134. Bunting, S. F. *et al.* BRCA1 Functions Independently of Homologous Recombination in DNA Interstrand Crosslink Repair. *Mol. Cell* **46**, 125–135 (2012).

135. Long, D. T., Joukov, V., Budzowska, M. & Walter, J. C. BRCA1 promotes unloading of the CMG Helicase from a stalled DNA replication fork. *Mol. Cell* **56**, 174–185 (2014).

136. Mutreja, K. *et al.* ATR-Mediated Global Fork Slowing and Reversal Assist Fork Traverse and Prevent Chromosomal Breakage at DNA Interstrand Cross-Links. *Cell Rep.* **24**, 2629-2642.e5 (2018).

137. Duquette, M. L. *et al.* CtIP Is Required to Initiate Replication-Dependent Interstrand Crosslink Repair. *PLOS Genet.* **8**, e1003050 (2012).

138. Huang, J. *et al.* Remodeling of Interstrand Crosslink Proximal Replisomes Is Dependent on ATR, FANCM, and FANCD2. *Cell Rep.* **27**, 1794-1808.e5 (2019).

139. Blackford, A. N. *et al.* The DNA translocase activity of FANCM protects stalled replication forks. *Hum. Mol. Genet.* **21**, 2005–2016 (2012).

140. Ling, C. *et al.* Bloom syndrome complex promotes FANCM recruitment to stalled replication forks and facilitates both repair and traverse of DNA interstrand crosslinks. *Cell Discov. 2016 21* **2**, 1–18 (2016).

141. González-Acosta, D. *et al.* PrimPol-mediated repriming facilitates replication traverse of DNA interstrand crosslinks. *EMBO J.* **40**, 1–17 (2021).

142. Beard, W. A., Horton, J. K., Prasad, R. & Wilson, S. H. Eukaryotic Base Excision Repair : New Approaches Shine Light on Mechanism. *Annu. Rev.* (2019).

143. Li, N. *et al.* Cooperation of the NEIL3 and Fanconi anemia/BRCA pathways in interstrand crosslink repair. *Nucleic Acids Res.* **48**, 3014–3028 (2020).

144. Hodskinson, M. R. *et al.* Alcohol-derived DNA crosslinks are repaired by two distinct mechanisms. *Nature* **579**, 603–608 (2020).

145. Sengerová, B., Wang, A. T. & McHugh, P. J. Orchestrating the nucleases involved in DNA interstrand cross-link (ICL) repair. *Cell Cycle* **10**, 3999–4008 (2011).

146. Wyatt, H. D. M., Laister, R. C., Martin, S. R., Arrowsmith, C. H. & West, S. C. The SMX DNA Repair Tri-nuclease. *Mol. Cell* **65**, 848-860.e11 (2017).

147. Räschle, M. *et al.* Proteomics reveals dynamic assembly of Repair complexes during bypass of DNA cross-links. *Science (80-.).* **348**, 539–546 (2015).

148. Bhagwat, N. *et al.* XPF-ERCC1 participates in the Fanconi anemia pathway of cross-link repair. *Mol. Cell. Biol.* **29**, 6427–37 (2009).

149. McHugh, P. J. XPF-ERCC1: Linchpin of DNA crosslink repair. *PLoS Genet.* **16**, 1–5 (2020).

150. Cybulski, K. E. & Howlett, N. G. FANCP/SLX4: A Swiss army knife of DNA interstrand crosslink repair. *Cell Cycle* **10**, 1757–1763 (2011).

151. Hoogenboom, W. S., Boonen, R. A. C. M. & Knipscheer, P. The role of SLX4 and its associated nucleases in DNA interstrand crosslink repair. *Nucleic Acids Res.* **47**, 2377–2388 (2019).

152. Castor, D. *et al.* Cooperative control of holliday junction resolution and DNA repair by the SLX1 and MUS81-EME1 nucleases. *Mol. Cell* **52**, 221–33 (2013).

153. Wyatt, H. D. M., Sarbajna, S., Matos, J. & West, S. C. Coordinated actions of SLX1-SLX4 and MUS81-EME1 for holliday junction resolution in human cells. *Mol. Cell* **52**, 234–247 (2013).

154. Thongthip, S. *et al.* Fan1 deficiency results in DNA interstrand cross-link repair defects, enhanced tissue karyomegaly, and organ dysfunction. *Genes Dev.* **30**, 645–659 (2016).

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

156. Zhao, Q., Xue, X., Longerich, S., Sung, P. & Xiong, Y. Structural insights into 5′ flap DNA unwinding and incision by the human FAN1 dimer. *Nat. Commun.* **5**, 5726 (2014).

157. Smogorzewska, A. *et al.* A genetic screen identifies FAN1, a Fanconi anemia-associated nuclease necessary for DNA interstrand crosslink repair. *Mol. Cell* **39**, 36–47 (2010).

158. Kratz, K. *et al.* Deficiency of FANCD2-associated nuclease KIAA1018/FAN1 sensitizes cells to interstrand crosslinking agents. *Cell* **142**, 77–88 (2010).

159. MacKay, C. *et al.* Identification of KIAA1018/FAN1, a DNA Repair Nuclease Recruited to DNA Damage by Monoubiquitinated FANCD2. *Cell* **142**, 65–76 (2010).

160. Porro, A. *et al.* FAN1 interaction with ubiquitylated PCNA alleviates replication stress and preserves genomic integrity independently of BRCA2. *Nat. Commun.* **8**,

1073 (2017).

161. Yoshikiyo, K. *et al.* KIAA1018/FAN1 nuclease protects cells against genomic instability induced by interstrand cross-linking agents. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 21553–21557 (2010).

162. Andrews, A. M., McCartney, H. J., Errington, T. M., D'Andrea, A. D. & Macara, I. G. A senataxin-associated exonuclease SAN1 is required for resistance to DNA interstrand cross-links. *Nat. Commun.* **9**, (2018).

163. Lachaud, C. *et al.* Ubiquitinated Fancd2 recruits Fan1 to stalled replication forks to prevent genome instability. *Science* **351**, 846 (2016).

164. Cannavo, E., Gerrits, B., Marra, G., Schlapbach, R. & Jiricny, J. Characterization of the Interactome of the Human MutL Homologues MLH1, PMS1, and PMS2. *J. Biol. Chem.* **282**, 2976–2986 (2007).

165. Fontebasso, Y., Etheridge, T. J., Oliver, A. W., Murray, J. M. & Carr, A. M. The conserved Fanconi anemia nuclease Fan1 and the SUMO E3 ligase Pli1 act in two novel Pso2-independent pathways of DNA interstrand crosslink repair in yeast. *DNA Repair (Amst).* **12**, 1011–1023 (2013).

166. Callebaut, I., Moshous, D., Mornon, J. P. & De Villartay, J. P. Metallo-β-lactamase fold within nucleic acids processing enzymes: the β-CASP family. *Nucleic Acids Res.* **30**, 3592 (2002).

167. Pettinati, I., Brem, J., Lee, S. Y., McHugh, P. J. & Schofield, C. J. The Chemical Biology of Human Metallo-β-Lactamase Fold Proteins. *Trends Biochem. Sci.* **41**, 338–355 (2016).

168. Callebaut, I., Moshous, D., Mornon, J.-P. & de Villartay, J.-P. Metallo-beta-lactamase fold within nucleic acids processing enzymes: the beta-CASP family. *Nucleic Acids Res.* **30**, 3592–601 (2002).

169. Yang, W. *Nucleases: Diversity of structure, function and mechanism*. *Quarterly Reviews of Biophysics* vol. 44 (2011).

170. Mandel, C. R. *et al.* Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature* **444**, 953–956 (2006).

171. Brzezniak, L. K., Bijata, M., Szczesny, R. J. & Stepien, P. P. Involvement of human ELAC2 gene product in 3' end processing of mitochondrial tRNAs. *RNA Biol.* **8**, 616–626 (2011).

172. Schmiester, M. & Demuth, I. SNM1B/Apollo in the DNA damage response and telomere maintenance. *Oncotarget* **8**, 48398–48409 (2017).

173. Chang, H. H. Y. & Lieber, M. R. Structure-Specific nuclease activities of Artemis and the Artemis: DNA-PKcs complex. *Nucleic Acids Res.* **44**, 4991–4997 (2016).

174. Yosaatmadja, Y. *et al.* Structural and mechanistic insights into the Artemis endonuclease and strategies for its inhibition. *Nucleic Acids Res.* **49**, 9310–9326 (2021).

175. Hejna, J., Philip, S., Ott, J., Faulkner, C. & Moses, R. The hSNM1 protein is a DNA 5'-exonuclease. *Nucleic Acids Res.* **35**, 6115–6123 (2007).

176. Hazrati, A. *et al.* Human SNM1A suppresses the DNA repair defects of yeast pso2 mutants. *DNA Repair (Amst).* **7**, 230–238 (2008).

177. Sengerová, B. *et al.* Characterization of the human SNM1A and SNM1B/Apollo DNA repair exonucleases. *J. Biol. Chem.* **287**, 26254–26267 (2012).

178. Allerston, C. K. *et al.* 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 (2015).

179. Buzon, B., Grainger, R., Huang, S., Rzadki, C. & Junop, M. S. Structure-specific endonuclease activity of SNM1A enables processing of a DNA interstrand crosslink. *Nucleic Acids Res.* **46**, 9057–9066 (2018).

180. Tiefenbach, T. & Junop, M. Pso2 (SNM1) is a DNA structure-specific endonuclease. *Nucleic Acids Res.* **40**, 2131–2139 (2012).

181. Zhang, X., Richie, C. & Legerski, R. J. Translation of hSNM1 is mediated by an internal ribosome entry site that upregulates expression during mitosis. *DNA Repair (Amst).* **1**, 379–390 (2002).

182. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).

183. Richie, C. T. *et al.* hSnm1 Colocalizes and Physically Associates with 53BP1 before and after DNA Damage. *Mol. Cell. Biol.* **22**, 8635–8647 (2002).

184. Akhter, S. & Legerski, R. J. SNM1A acts downstream of ATM to promote the G1 cell cycle checkpoint. *Biochem. Biophys. Res. Commun.* **377**, 236–241 (2008).

185. Yang, K., Moldovan, G.-L. & D'Andrea, A. D. RAD18-dependent Recruitment of SNM1A to DNA Repair Complexes by a Ubiquitin-binding Zinc Finger. *J. Biol. Chem.* **285**, 19085–19091 (2010).

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

187. Horsfall, A. J. *et al.* Unlocking the PIP-box: A peptide library reveals interactions that drive high-affinity binding to human PCNA. *J. Biol. Chem.* **296**, 100773 (2021).

188. Akhter, S. *et al.* Deficiency in SNM1 Abolishes an Early Mitotic Checkpoint Induced by Spindle Stress. *Mol. Cell. Biol.* **24**, 10448–10455 (2004).

189. Ishiai, M. *et al.* DNA Cross-Link Repair Protein SNM1A Interacts with PIAS1 in Nuclear Focus Formation. *Mol. Cell. Biol.* **24**, 10733–10741 (2004).

190. Iyama, T. *et al.* CSB interacts with SNM1A and promotes DNA interstrand crosslink processing. *Nucleic Acids Res.* **43**, 247–258 (2015).

191. Jungmichel, S. *et al.* Proteome-wide identification of poly(ADP-Ribosyl)ation targets in different genotoxic stress responses. *Mol. Cell* **52**, 272–285 (2013).

192. Martello, R. *et al.* Proteome-wide identification of the endogenous ADP-ribosylome of mammalian cells and tissue. *Nat. Commun. 2016 71* **7**, 1–13 (2016).

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

194. Baddock, H. T. *et al.* The SNM1A DNA repair nuclease. *DNA Repair (Amst).* **95**, 102941 (2020).

195. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

196. Dantuma, N. P. & Attikum, H. Spatiotemporal regulation of posttranslational modifications in the DNA damage response. *EMBO J.* **35**, 6–23 (2016).

197. Ahkter, S. *et al.* Snm1-deficient mice exhibit accelerated tumorigenesis and susceptibility to infection. *Mol. Cell. Biol.* **25**, 10071–10078 (2005).

198. Munk, S. *et al.* Proteomics Reveals Global Regulation of Protein SUMOylation by ATM and ATR Kinases during Replication Stress. *Cell Rep.* **21**, 546–558 (2017).

199. Hendriks, I. A. *et al.* Site-specific mapping of the human SUMO proteome reveals co-modification with phosphorylation. *Nat. Struct. Mol. Biol. 2017 243* **24**, 325–336 (2017).

200. Hornbeck, P. V. *et al.* PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.* **43**, D512–D520 (2015).

201. Xiao, Z. *et al.* System-wide analysis of SUMOylation dynamics in response to replication stress reveals novel small ubiquitin-like modified target proteins and acceptor lysines relevant for genome stability. *Mol. Cell. Proteomics* **14**, 1419–1434 (2015).

202. Hendriks, I. A. *et al.* Uncovering global SUMOylation signaling networks in a site-specific manner. *Nat. Struct. Mol. Biol.* **21**, 927–936 (2014).

203. Li, C. *et al.* Quantitative SUMO proteomics identifies PIAS1 substrates involved in cell migration and motility. *Nat. Commun. 2020 111* **11**, 1–14 (2020).

204. Udeshi, N. D. *et al.* Refined preparation and use of anti-diglycine remnant (K-ε-GG) antibody enables routine quantification of 10,000s of ubiquitination sites in

single proteomics experiments. *Mol. Cell. Proteomics* **12**, 825–831 (2013).

205. Akimov, V. *et al.* UbiSite approach for comprehensive mapping of lysine and N-terminal ubiquitination sites. *Nat. Struct. Mol. Biol.* **25**, 631–640 (2018).

206. VA, S. & J, H. PRALINE - a multiple sequence alignment toolbox that integrates homology-extended and secondary structure information. *Nucleic Acids Res.* **33**, 289–294 (2005).

207. Mészáros, B., Erdös, G. & Dosztányi, Z. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* **46**, W329–W337 (2018).

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

209. Petsalaki, E. & Zachos, G. DNA damage response proteins regulating mitotic cell division: double agents preserving genome stability. *FEBS J.* **287**, 1700–1721 (2020).

210. Buzon, B., Grainger, R. A., Rzadki, C., Huang, S. Y. M. & Junop, M. Identification of Bioactive SNM1A Inhibitors. *ACS Omega* **6**, 9352–9361 (2021).

211. Hashimoto, S., Anai, H. & Hanada, K. Mechanisms of interstrand DNA crosslink repair and human disorders. *Genes Environ.* **38**, 1–8 (2016).

212. de Laat, W. L., Appeldoorn, E., Jaspers, N. G. & Hoeijmakers, J. H. DNA structural elements required for ERCC1-XPF endonuclease activity. *J. Biol. Chem.* **273**, 7835–42 (1998).

213. Liu, H. & Naismith, J. H. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol.* **8**, 91 (2008).

214. de Marco, A. Protocol for preparing proteins with improved solubility by co-expressing with molecular chaperones in Escherichia coli. *Nat. Protoc.* **2**, 2632–2639 (2007).

215. Steczko, J., Donoho, G. A., Dixon, J. E., Sugimoto, T. & Axelrod, B. Effect of ethanol and low-temperature culture on expression of soybean lipoxygenase L-1 in Escherichia coli. *Protein Expr. Purif.* **2**, 221–227 (1991).

216. Hartley, J. A. *et al.* SJG-136 (NSC 694501), a Novel Rationally Designed DNA Minor Groove Interstrand Cross-Linking Agent with Potent and Broad Spectrum Antitumor Activity. *Cancer Res.* **64**, 6693–6699 (2004).

217. Li, X., Hejna, J. & Moses, R. E. The yeast Snm1 protein is a DNA 5 -exonuclease. *DNA Repair (Amst).* **4**, 163–170 (2005).

218. Buzon, B. Characterization and inhibition of interstrand crosslink repair nuclease

SNM1A. (McMaster University, 2018).

219. Huang, S. Y. M. Structural and functional characterization of human SNM1A. (McMaster University, 2013).

220. Ma, Y., Pannicke, U., Schwarz, K. & Lieber, M. R. Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell* **108**, 781–794 (2002).

221. Li, S. *et al.* Evidence that the DNA endonuclease ARTEMIS also has intrinsic 5'-exonuclease activity. *J. Biol. Chem.* **289**, 7825–34 (2014).

222. Goodarzi, A. A. *et al.* DNA-PK autophosphorylation facilitates Artemis endonuclease activity. *EMBO J.* **25**, 3880–9 (2006).

223. Yannone, S. M. *et al.* Coordinate 5' and 3' endonucleolytic trimming of terminally blocked blunt DNA double-strand break ends by Artemis nuclease and DNA-dependent protein kinase. *Nucleic Acids Res.* **36**, 3354–3365 (2008).

224. Yu, J., Marshall, K., Yamaguchi, M., Haber, J. E. & Weil, C. F. Microhomology-dependent end joining and repair of transposon-induced DNA hairpins by host factors in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **24**, 1351–64 (2004).

225. Newman, J. A. *et al.* PCSB PDB - 5NZY: Crystal structure of DNA cross-link repair protein 1A in complex with Cefotaxime. *PDB* https://www.rcsb.org/structure/5NZY (2017).

226. Newman, J. A. *et al.* RCSB PDB - 5Q2A: PanDDA analysis group deposition -- Crystal Structure of DCLRE1A after initial refinement with no ligand modelled (structure 1). *PDB* https://www.rcsb.org/structure/5Q2A (2018).

227. Baddock, H. T. *et al.* A phosphate binding pocket is a key determinant of exo- versus endo-nucleolytic activity in the SNM1 nuclease family. *Nucleic Acids Res.* **49**, 9294–9309 (2021).

228. Karsisiotis, A. I., Damblon, C. F. & Roberts, G. C. K. Solution structures of the Bacillus cereus metallo-β-lactamase BcII and its complex with the broad spectrum inhibitor R-thiomandelic acid. *Biochem. J.* **456**, 397 (2013).

229. de la Sierra-Gallay, I. L., Zig, L., Jamalli, A. & Putzer, H. Structural insights into the dual activity of RNase J. *Nat. Struct. Mol. Biol.* **15**, 206–212 (2008).

230. Zhao, Y. *et al.* Structural insights into catalysis and dimerization enhanced exonuclease activity of RNase J. *Nucleic Acids Res.* **43**, 5550–5559 (2015).

231. Kunkel, T. A. & Loeb, L. A. On the fidelity of DNA replication. Effect of divalent metal ion activators and deoxyribonucleoside triphosphate pools on in vitro mutagenesis. *J. Biol. Chem.* **254**, 5718–5725 (1979).

232. Tabor, S. & Richardson, C. C. Effect of manganese ions on the incorporation of

dideoxynucleotides by bacteriophage T7 DNA polymerase and Escherichia coli DNA polymerase I. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 4076–4080 (1989).

233. Allingham, J. S. & Haniford, D. B. Mechanisms of metal ion action in Tn10 transposition. *J. Mol. Biol.* **319**, 53–65 (2002).

234. Yang, W., Lee, J. Y. & Nowotny, M. Making and Breaking Nucleic Acids: Two-Mg2+-Ion Catalysis and Substrate Specificity. *Mol. Cell* **22**, 5–13 (2006).

235. Steitz, T. A. & Steitz, J. A. A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 6498 (1993).

236. Yang, W. An equivalent metal ion in one- and two-metal-ion catalysis. *Nat. Struct. Mol. Biol.* **15**, 1228–1231 (2008).

237. Palermo, G. *et al.* Catalytic Metal Ions and Enzymatic Processing of DNA and RNA. *Acc. Chem. Res.* **48**, 220–228 (2015).

238. Ochiai, E. *Bioinorganic Chemistry: a Survey*. (Academic Press, 2008).

239. Biertümpfel, C. *et al.* Structure and mechanism of human DNA polymerase. *Nature* **465**, 1044–1048 (2010).

240. Gao, Y. & Yang, W. Capture of a third Mg2+ is essential for catalyzing DNA synthesis. *Science (80-.).* **352**, 1334–1337 (2016).

241. Freudenthal, B. D., Beard, W. A., Shock, D. D. & Wilson, S. H. Observing a DNA polymerase choose right from wrong. *Cell* **154**, 157 (2013).

242. Shi, Y., Hellinga, H. W. & Beese, L. S. Interplay of catalysis, fidelity, threading, and processivity in the exo-and endonucleolytic reactions of human exonuclease i. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 6010–6015 (2017).

243. Molina, R. *et al.* Visualizing phosphodiester-bond hydrolysis by an endonuclease. *Nat. Struct. Mol. Biol.* **22**, 65–72 (2015).

244. Donati, E., Genna, V. & De Vivo, M. Recruiting Mechanism and Functional Role of a Third Metal Ion in the Enzymatic Activity of 5′ Structure-Specific Nucleases. *J. Am. Chem. Soc.* **142**, 2823–2834 (2020).

245. Yang, W., Weng, P. J. & Gao, Y. A new paradigm of DNA synthesis: Three-metal-ion catalysis. *Cell Biosci.* **6**, 1–7 (2016).

246. Bouwman, P. & Jonkers, J. The effects of deregulated DNA damage signalling on cancer chemotherapy response and resistance. *Nat. Rev. Cancer* **12**, 587–598 (2012).

247. Curtin, N. J. DNA repair dysregulation from cancer driver to therapeutic target. *Nat. Rev. Cancer* **12**, 801–817 (2012).

248. O'Connor, M. J. Targeting the DNA Damage Response in Cancer. *Mol. Cell* **60**,

547–560 (2015).

249. Motegi, A., Masutani, M., Yoshioka, K. ichi & Bessho, T. Aberrations in DNA repair pathways in cancer and therapeutic significances. *Semin. Cancer Biol.* **58**, 29–46 (2019).

250. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: The next generation. *Cell* **144**, 646–674 (2011).

251. Swift, L. H. & Golsteyn, R. M. Genotoxic Anti-Cancer Agents and Their Relationship to DNA Damage, Mitosis, and Checkpoint Adaptation in Proliferating Cancer Cells. *Int. J. Mol. Sci. 2014, Vol. 15, Pages 3403-3431* **15**, 3403–3431 (2014).

252. Park, C. M. *et al.* Induction of p53-mediated apoptosis and recovery of chemosensitivity through p53 transduction in human glioblastoma cells by cisplatin. *Int. J. Oncol.* **28**, 119–125 (2006).

253. Kauffmann, A. *et al.* High expression of DNA repair pathways is associated with metastasis in melanoma patients. *Oncogene* **27**, 565–573 (2008).

254. Schärer, O. D. DNA interstrand crosslinks: natural and drug-induced DNA adducts that induce unique cellular responses. *Chembiochem* **6**, 27–32 (2005).

255. Olaussen, K. A. *et al.* DNA Repair by ERCC1 in Non–Small-Cell Lung Cancer and Cisplatin-Based Adjuvant Chemotherapy. *http://dx.doi.org/10.1056/NEJMoa060570* **355**, 983–991 (2009).

256. Dong, Q. *et al.* Multiple DNA repair mechanisms and alkylator resistance in the human medulloblastoma cell line D-283 Med (4-HCR). *Cancer Chemother. Pharmacol.* **43**, 73–79 (1999).

257. Panasci, L., Xu, Z. Y., Bello, V. & Aloyz, R. The role of DNA repair in nitrogen mustard drug resistance. *Anticancer. Drugs* **13**, 211–220 (2002).

258. Florea, A. M. & Büsselberg, D. Cisplatin as an anti-tumor drug: cellular mechanisms of activity, drug resistance and induced side effects. *Cancers (Basel).* **3**, 1351–1371 (2011).

259. Galluzzi, L. *et al.* Molecular mechanisms of cisplatin resistance. *Oncogene 2012 3115* **31**, 1869–1883 (2011).

260. Choi, C.-H. ABC transporters as multidrug resistance mechanisms and the development of chemosensitizers for their reversal. (2005) doi:10.1186/1475-2867-5-30.

261. Mohammad, R. M. *et al.* Broad targeting of resistance to apoptosis in cancer. *Semin. Cancer Biol.* **35**, S78 (2015).

262. Martin, L. P., Hamilton, T. C. & Schilder, R. J. Platinum Resistance: The Role of

DNA Repair Pathways. *Clin. Cancer Res.* **14**, 1291–1295 (2008).

263. Britten, R. A., Liu, D., Tessier, A., Hutchison, M. J. & Murray, D. ERCC1 expression as a molecular marker of cisplatin resistance in human cervival tumor cells. *Int. J. Cancer (Pred. Oncol.)* **89**, 453–457 (2000).

264. Li, Q. *et al.* Cisplatin induction of ERCC-1 mRNA expression in A2780/CP70 human ovarian cancer cells. *J. Biol. Chem.* **273**, 23419–23425 (1998).

265. Usanova, S. *et al.* Cisplatin sensitivity of testis tumour cells is due to deficiency in interstrand-crosslink repair and low ERCC1-XPF expression. *Mol. Cancer* **9**, (2010).

266. Song, L., Ritchie, A. M., McNeil, E. M., Li, W. & Melton, D. W. Identification of DNA repair gene Ercc1 as a novel target in melanoma. *Pigment Cell Melanoma Res.* **24**, 966–971 (2011).

267. Lord, R. V. N. *et al.* Low ERCC1 Expression Correlates with Prolonged Survival after Cisplatin plus Gemcitabine Chemotherapy in Non-Small Cell Lung Cancer 1. *Clin. Cancer Res.* **8**, 2286–2291 (2002).

268. Simon, G. R., Sharma, S., Cantor, A., Smith, P. & Bepler, G. ERCC1 expression is a predictor of survival in resected patients with non-small cell lung cancer. *Chest* **127**, 978–983 (2005).

269. Chen, S., Zhang, J., Wang, R., Luo, X. & Chen, H. The platinum-based treatments for advanced non-small cell lung cancer, is low/negative ERCC1 expression better than high/positive ERCC1 expression? A meta-analysis. *Lung Cancer* **70**, 63–70 (2010).

270. Jordheim, L. P. *et al.* Small molecule inhibitors of ERCC1-XPF protein-protein interaction synergize alkylating agents in cancer cells. *Mol. Pharmacol.* **84**, 12–24 (2013).

271. McNeil, E. M. *et al.* Inhibition of the ERCC1–XPF structure-specific endonuclease to overcome cancer chemoresistance. *DNA Repair (Amst).* **31**, 19–28 (2015).

272. Chapman, T. M. *et al.* N-Hydroxyimides and hydroxypyrimidinones as inhibitors of the DNA repair complex ERCC1-XPF. *Bioorg. Med. Chem. Lett.* **25**, 4104–4108 (2015).

273. Chapman, T. M. *et al.* Catechols and 3-hydroxypyridones as inhibitors of the DNA repair complex ERCC1-XPF. *Bioorg. Med. Chem. Lett.* **25**, 4097–4103 (2015).

274. Arora, S. *et al.* Identification of small molecule inhibitors of ERCC1-XPF that inhibit DNA repair and potentiate cisplatin efficacy in cancer cells. *Oncotarget* **7**, 75104–75117 (2016).

275. Yang, L., Ritchie, A. M. & Melton, D. W. Disruption of DNA repair in cancer cells by ubiquitination of a destabilising dimerization domain of nucleotide excision repair protein ERCC1. *Oncotarget* **8**, 55246–55264 (2017).

276. Heyza, J. R. *et al.* Targeting the DNA Repair Endonuclease ERCC1-XPF with Green Tea Polyphenol Epigallocatechin-3-Gallate (EGCG) and Its Prodrug to Enhance Cisplatin Efficacy in Human Cancer Cells. *Nutrients* **10**, (2018).

277. Elmenoufy, A. H. *et al.* Targeting DNA Repair in Tumor Cells via Inhibition of ERCC1-XPF. *J. Med. Chem.* **62**, 7684–7696 (2019).

278. Elmenoufy, A. H. *et al.* Design, synthesis and in vitro cell-free/cell-based biological evaluations of novel ERCC1-XPF inhibitors targeting DNA repair pathway. *Eur. J. Med. Chem.* **204**, 112658 (2020).

279. Ciccia, A., McDonald, N. & West, S. C. Structural and functional relationships of the XPF/MUS81 family of proteins. *Annu. Rev. Biochem.* **77**, 259–287 (2008).

280. Wang, X., Wang, S.-S., Zhou, L., Yu, L. & Zhang, L.-M. A network-pathway based module identification for predicting the prognosis of ovarian cancer patients. *J. Ovarian Res.* **9**, (2016).

281. Laporte, G. A. *et al.* The role of double-strand break repair, translesion synthesis, and interstrand crosslinks in colorectal cancer progression-clinicopathological data and survival. *J. Surg. Oncol.* **121**, 906–916 (2020).

282. Lee, S. Y. *et al.* Cephalosporins inhibit human metallo β-lactamase fold DNA repair nucleases SNM1A and SNM1B/apollo. *Chem. Commun. (Camb).* **52**, 6727–6730 (2016).

283. Dürr, E. M. *et al.* Squaramide-Based 5'-Phosphate Replacements Bind to the DNA Repair Exonuclease SNM1A. *ChemistrySelect* **3**, 12824–12829 (2018).

284. Doherty, W. *et al.* A hydroxamic-acid-containing nucleoside inhibits DNA repair nuclease SNM1A. *Org. Biomol. Chem.* **17**, 8094–8105 (2019).

285. Dürr, E. M. & McGouran, J. F. Probing the binding requirements of modified nucleosides with the dna nuclease snm1a. *Molecules* **26**, 1–25 (2021).

286. Berney, M., Manoj, M. T., Fay, E. M. & McGouran, J. F. 5'-Phosphorylation Increases the Efficacy of Nucleoside Inhibitors of the DNA Repair Enzyme SNM1A. *ChemMedChem* **17**, (2022).

287. Berney, M. *et al.* Synthesis and evaluation of squaramide and thiosquaramide inhibitors of the DNA repair enzyme SNM1A. *Bioorg. Med. Chem.* **46**, 116369 (2021).

288. Daina, A., Michielin, O. & Zoete, V. SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* **7**, 1–13 (2017).

289. Sterling, T. & Irwin, J. J. ZINC 15 - Ligand Discovery for Everyone. *J. Chem. Inf. Model.* **55**, 2324–2337 (2015).

290. Ganesh, A. N., Donders, E. N., Shoichet, B. K. & Shoichet, M. S. Colloidal aggregation: From screening nuisance to formulation nuance. *Nano Today* **19**, 188–200 (2018).

291. Kennedy, R. D. *et al.* Fanconi anemia pathway–deficient tumor cells are hypersensitive to inhibition of ataxia telangiectasia mutated. *J. Clin. Invest.* **117**, 1440 (2007).

292. Chen, C. C., Kennedy, R. D., Sidi, S., Look, A. T. & D'Andrea, A. CHK1 inhibition as a strategy for targeting Fanconi Anemia (FA) DNA repair pathway deficient tumors. *Mol. Cancer* **8**, (2009).

293. Sharp, M. F., Bythell-Douglas, R., Deans, A. J. & Crismani, W. The Fanconi anemia ubiquitin E3 ligase complex as an anti-cancer target. *Mol. Cell* **81**, 2278–2289 (2021).

294. Simeonov, A. *et al.* Identification and Characterization of Inhibitors of Human Apurinic/apyrimidinic Endonuclease APE1. (2009) doi:10.1371/journal.pone.0005740.

295. Mcwhirter, C. *et al.* Development of a high-throughput fluorescence polarization DNA cleavage assay for the identification of FEN1 inhibitors. *J. Biomol. Screen.* **18**, 567–575 (2013).

296. Patrick, G. L. *An Introduction to Medicinal Chemistry*. (Oxford University Press, 2013).

297. Murray, C. W. *et al.* Validity of Ligand Efficiency Metrics. *ACS Med. Chem. Lett.* **5**, 616–618 (2014).

298. Hopkins, A. L., Keserü, G. M., Leeson, P. D., Rees, D. C. & Reynolds, C. H. The role of ligand efficiency metrics in drug discovery. *Nat. Rev. Drug Discov.* **13**, 105–121 (2014).

299. Schuck, A. G., Ausubel, M. B., Zuckerbraun, H. L. & Babich, H. Theaflavin-3,3'-digallate, a component of black tea: An inducer of oxidative stress and apoptosis. *Toxicol. Vitr.* **22**, 598–609 (2008).

300. Lee, B. Il, Ahn, H. J., Han, K. C., Ahn, D. R. & Shin, D. Pyrogallin, an ATP-Competitive Inhibitor of JAK3. *Bull. Korean Chem. Soc.* **32**, 1077–1079 (2011).

301. Chen, D. *et al.* EGCG, green tea polyphenols and their synthetic analogs and prodrugs for human cancer prevention and treatment. *Adv. Clin. Chem.* **53**, 155–177 (2011).

302. Wang, X. *et al.* EGCG Enhances Cisplatin Sensitivity by Regulating Expression of the Copper and Cisplatin Influx Transporter CTR1 in Ovary Cancer. *PLoS One* **10**, (2015).

303. Hu, F. *et al.* EGCG synergizes the therapeutic effect of cisplatin and oxaliplatin

through autophagic pathway in human colorectal cancer cells. *J. Pharmacol. Sci.* **128**, 27–34 (2015).

304. Xiong, X. *et al.* Insights Into Amentoflavone: A Natural Multifunctional Biflavonoid. *Front. Pharmacol.* **12**, 3758 (2021).

305. Himi, T., Ishizaki, Y. & Murota, S. I. 4,4′-Diisothiocyano-2,2′-stilbenedisulfonate protects cultured cerebellar granule neurons from death. *Life Sci.* **70**, 1235–1249 (2002).

306. Benítez-Rangel, E., López-Méndez, M., García, L. & Guerrero-Hernández, A. DIDS (4,4′-Diisothiocyanatostilbene-2,2′-disulfonate) directly inhibits caspase activity in HeLa cell lysates. *Cell Death Discov. 2015 11* **1**, 1–9 (2015).

307. Canny, S. A., Cruz, Y., Southern, M. R. & Griffin, P. R. PubChem promiscuity: A web resource for gathering compound promiscuity data from PubChem. *Bioinformatics* **28**, 140–141 (2012).

308. Tsutakawa, S. E., Lafrance-Vanasse, J. & Tainer, J. A. The cutting edges in DNA repair, licensing, and fidelity: DNA and RNA repair nucleases sculpt DNA to measure twice, cut once. *DNA Repair (Amst).* **19**, 95–107 (2014).

309. Tacconi, E. M. *et al.* Chlorambucil targets BRCA 1/2-deficient tumours and counteracts PARP inhibitor resistance. *EMBO Mol. Med.* **11**, 1–16 (2019).

310. Henriques, J. A. P. & Moustacchi, E. Isolation and characterization of pso mutants sensitive to photo-addition of psoralen derivatives in Saccharomyces cerevisiae. *Genetics* **95**, (1980).

311. Henriques, J. A. P. & Moustacchi, E. Interactions between mutations for sensitivity to psoralen photoaddition (pso) and to radiation (rad) in Saccharomyces cerevisiae. *J. Bacteriol.* **148**, 248–256 (1981).

312. Ruhland, A., Haase, E., Siede, W. & Brendel, M. Isolation of yeast mutants sensitive to the bifunctional alkylating agent nitrogen mustard. *Mol. Gen. Genet.* **181**, 346–351 (1981).

313. Bonatto, D., Brendel, M. & Henriques, J. A. P. The eukaryotic Pso2p/Snm1p family revisited: in silico analyses of Pso2p A, B and Plasmodium groups. *Comput. Biol. Chem.* **29**, 420–433 (2005).

314. Li, X., Hejna, J. & Moses, R. E. The yeast Snm1 protein is a DNA 5′-exonuclease. *DNA Repair (Amst).* **4**, 163–170 (2005).

315. Rogers, C. M. *et al.* Yeast Hrq1 shares structural and functional homology with the disease-linked human RecQ4 helicase. *Nucleic Acids Res.* **45**, 5217–5230 (2017).

316. Wilborn, F. & Brendel, M. Formation and stability of interstrand cross-links induced by cis- and trans-diamminedichloroplatinum (II) in the DNA of Saccharomyces cerevisiae strains differing in repair capacity. *Curr. Genet. 1989*

165 **16**, 331–338 (1989).

317. Magana-Schwencke, N., Pegas Henriques, J. A., Chanet, R. & Moustacchi, E. 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**, 1722–1726 (1982).

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

319. Lam, A. F., Krogh, B. O. & Symington, L. S. Unique and overlapping functions of the Exo1, Mre11 and Pso2 nucleases in DNA repair. *DNA Repair (Amst).* **7**, 655–662 (2008).

320. Barber, L. J., Ward, T. A., Hartley, J. A. & McHugh, P. J. DNA Interstrand Cross-Link Repair in the Saccharomyces cerevisiae Cell Cycle: Overlapping Roles for PSO2 (SNM1) with MutS Factors and EXO1 during S Phase. *Mol. Cell. Biol.* **25**, 2297 (2005).

321. Ward, T. A. *et al.* Components of a Fanconi-like pathway control Pso2-independent DNA interstrand crosslink repair in yeast. *PLoS Genet.* **8**, (2012).

322. Li, Y. *et al.* Nonhomologous End-Joining with Minimal Sequence Loss Is Promoted by the Mre11-Rad50-Nbs1-Ctp1 Complex in Schizosaccharomyces pombe. *Genetics* **206**, 481–496 (2017).

323. Winzeler, E. A. *et al.* Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. *Science (80-.).* **285**, 901–906 (1999).

324. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2016).

325. Li, X. & Moses, R. E. The beta-lactamase motif in Snm1 is required for repair of DNA double-strand breaks caused by interstrand crosslinks in S. cerevisiae. *DNA Repair (Amst).* **2**, 121–9 (2003).

326. Karim, M. F. *et al.* Structural analysis of the catalytic domain of Artemis endonuclease/SNM1C reveals distinct structural features. *J. Biol. Chem.* **295**, 12368–12377 (2020).

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

328. Prakash, S. & Prakash, L. Nucleotide excision repair in yeast. *Mutat. Res. Mol. Mech. Mutagen.* **451**, 13–24 (2000).

329. Lehoczký, P., McHugh, P. J. & Chovanec, M. DNA interstrand cross-link repair in

Saccharomyces cerevisiae. *FEMS Microbiol. Rev.* **31**, 109–133 (2007).

330. Meniel, V., Magana-schwencke, N. & Averbeck, D. Preferential repair in Saccharomyces cerevisiae rad mutants after induction of interstrand cross-links by 8-methoxypsoralen plus UVA. *Mutagenesis* **10**, 543–548 (1995).

331. Cherry, J. M. *et al.* Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res.* **40**, D700 (2012).

332. David, L. *et al.* A high-resolution map of transcription in the yeast genome. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 5320–5325 (2006).

333. Jafari, R. *et al.* The cellular thermal shift assay for evaluating drug target interactions in cells. *Nat. Protoc. 2014 99* **9**, 2100–2122 (2014).

334. Munari, F. M. *et al.* Sak1 kinase interacts with Pso2 nuclease in response to DNA damage induced by interstrand crosslink-inducing agents in Saccharomyces cerevisiae. *J. Photochem. Photobiol. B Biol.* **130**, 241–253 (2014).

Curriculum Vitae

Ryan Grainger

EDUCATION

2016 – 2022 PhD, Biochemistry
Schulich School of Medicine and Dentistry, Western University
Thesis: Dual Functions of Interstrand Crosslink Repair Nuclease SNM1A
Supervisor: Dr. Murray Junop

2012 – 2016 BMSc, Honours Specialization in Biochemistry & Microbiology
Schulich School of Medicine and Dentistry, Western University

SCHOLARSHIPS AND AWARDS

2020 Postgraduate Scholarship – Doctoral (NSERC)
2020 Ontario Graduate Scholarship (Western University, declined)
2019 Ontario Graduate Scholarship (Western University)
2017 Travel Award – Institute Community Support (CIHR)
2016 Canadian Graduate Scholarship – Master's (CIHR)
2016 Ontario Graduate Scholarship (Western University, declined)
2016 Western Graduate Research Scholarship (Western University)
2016 Gold Medal in Biochemistry of Infection and Immunity (Western University)

TEACHING EXPERIENCE

Research Team Leader

- Responsible for training and mentoring 6 undergraduate thesis and 2 Master's students. Taught students to design and conduct experiments, analyze data and present results.

Teaching Assistant, Biological Macromolecules 3381

- Prepared and delivered tutorial lectures for ~200 third year students on protein structure prediction and analysis.
- Designed and evaluated assignments. Conducted office hours to assist student with course content.

PUBLICATIONS

Peer-reviewed:

1. **Grainger R***, Buzon B*, Rzadki C, Huang S, Junop M. (2021) Identification of bioactive SNM1A inhibitors. *ACS Omega.* 6(14): 9352-9361 (*co-author)
2. Buzon B, **Grainger R**, Huang S, Junop M. (2018) SNM1A endonuclease activity enables direct processing of an interstrand crosslink. *Nucleic Acids Research.* 46(17): 9057-9066
3. Heimhalt M, Mukherjee P, **Grainger R**, Szabla R, Brown C, Turner R, Junop M, Berti P. (2021) An Inhibitor-in-pieces approach to DAHP Synthase Inhibition: Potent enzyme and bacterial growth inhibition. *ACS Infect Dis.* 7(12): 3292-3302

In Progress:

1. Balachandran N, **Grainger R**, Rob T, Liuni P, Wilson D, Junop M, Berti P. (2022) DAHP synthase: half-of-sites reactivity, pseudo-twofold symmetry and protein dynamic control of ligand interactions. (manuscript submitted)
2. Mederios B, Goodale D, Roseborough A, Roes M, Bhat V, **Grainger R**, Hearn S, Kitz J, Whitehead S, Dick F, Junop M, Allan A. (2022) Tipifarnib inhibits extracellular vesicle (EV)-mediated lung metastasis of triple negative breast cancer. (manuscript submitted)

PRESENTATIONS

1. **Grainger R,** Buzon B, Blue T, Junop M. (2019) Interrogating the function of SNM1A nuclease activity in DNA repair. *Buffalo Replication and Repair Symposium*. Buffalo, New York
2. **Grainger R**, Buzon B, Junop M. (2019) Identification of bioactive inhibitors of ICL nuclease SNM1A for cisplatin potentiation. *Oncology Research Day*. London, Ontario
3. **Grainger R**, Buzon B, Blue T, Junop M. (2017) Elucidating the role of SNM1A nuclease activity in DNA repair. *London Health Research Day*. London, Ontario
4. Buzon B*, **Grainger R***, Huang S, Dowling M, Junop M. (2015) SNM1A is a single-strand dependent endonuclease: Implications for ICL repair. *Fanconi Anemia Symposium*. Toronto, Ontario

STRUCTURES DEPOSITED IN PROTEIN DATA BANK

8E0Z	**Grainger R**, Junop M, Berti P. (2022) DAHP (3-deoxy-D-arabinoheptulosonate-7-phosphate) Synthase unbound:(bound)$_2$:unbound Conformations
8E0Y	**Grainger R**, Junop M, Berti P. (2022) DAHP (3-deoxy-D-arabinoheptulosonate-7-phosphate) Synthase complexed with DAHP oxime, Pr(III), and P_i in unbound:(bound)$_2$:other Conformations
8E0X	**Grainger R**, Junop M, Berti P. (2022) DAHP (3-deoxy-D-arabinoheptulosonate-7-phosphate) Synthase complexed with Mn(II), PEP, and P_i in unbound:(bound)$_2$:other Conformations
8E0V	**Grainger R**, Junop M, Berti P. (2022) DAHP (3-deoxy-D-arabinoheptulosonate-7-phosphate) Synthase complexed with Mn(II), PEP, and P_i in unbound:(bound)$_2$:other Conformations
8E0U	**Grainger R**, Junop M, Berti P. (2022) DAHP (3-deoxy-D-arabinoheptulosonate-7-phosphate) Synthase complexed with Sulfate in unbound:(bound)$_2$:other conformations
8E0T	**Grainger R**, Junop M, Berti P. (2022) DAHP (3-deoxy-D-arabinoheptulosonate-7-phosphate) Synthase complexed with DAHP Oxime in unbound:(bound)$_2$:other conformations
8E0S	**Grainger R**, Junop M, Berti P. (2022) DAHP (3-deoxy-D-arabinoheptulosonate-7-phosphate) Synthase complexed with DAHP Oxime in unbound:(bound)$_2$:unbound conformations
7RUD	Heimhalt M, Murherjee P, **Grainger R**, Szabla R, Brown C, Turner R, Junop M, Berti P. (2021) DAHP synthase complex with trifluoropyruvate oxime.
7RUE	Heimhalt M, Murherjee P, **Grainger R**, Szabla R, Brown C, Turner R, Junop M, Berti P. (2021) DAHP synthase complexed with trifluoropyruvate semicarbazone.