STRUCTURAL AND FUNCTIONAL CHARACTERIZATION OF HUMAN SNM1A

STRUCTURAL AND FUNCTIONAL CHARACTERIZATION OF HUMAN SNM1A

By SIMON Y.M. HUANG, B.Sc.

A Thesis submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree of Master of Sciences

McMaster University © Copyright by Simon Y.M. Huang, November 2013

MASTER OF SCIENCES (2013) McMaster University

(Biochemistry and Biomedical Sciences) Hamilton, Ontario

TITLE: Structural and Functional Characterization of Human SNM1A:

AUTHOR: Simon York-Ming Huang, B.Sc. (McMaster University)

SUPERVISOR: Dr. Murray S. Junop

NUMBER OF PAGES: ix; 119

ABSTRACT

DNA interstrand cross-links (ICLs) occur when various chemical agents bind to chromosomal DNA and form a covalent bond between adjacent strands, preventing unwinding of the DNA double helix. The formation of an ICL is therefore extremely toxic to cells and necessitates quick removal and subsequent repair. Human SNM1A is a 5'-phosphate-dependent exonuclease that has been shown to be selectively involved in ICL repair; however the mechanism by which it processes ICL substrates remains unclear. To address this, our research is focused on the structural and functional characterization of SNM1A to determine this mechanism of substrate processing. In this thesis, we report the purification of human $SNM1A_{698-1040}$ as a His_6-NusA tagged protein from 4 L of *E. Coli* cell culture. This protein was found to possess 5'-phosphate-dependent exonuclease activity, and demonstrated a preference for ssDNA. Additionally, electrophoretic mobility shift assays performed with a D736A/H737A mutant suggest that the binding of SNM1A to DNA is independent of the presence of a 5' phosphate. Lastly, we have developed an *in vitro* fluorescence-based assay for inhibitor screening of $SNM1A_{698-1040}$ with a Z' factor of 0.88. Collectively, these results provide insight into the mechanism of SNM1A substrate processing in ICL repair, and establish a platform for future studies of this protein.

ACKNOWLEDGEMENTS

I would like to start this thesis by thanking my supervisor, Dr. Murray Junop, who graciously provided me the opportunity to perform research during my undergraduate and graduate studies. During this time, he has been an inspirational model, and has repeatedly demonstrated patience and a willingness to allow me to pursue my curiosity as a scientist. His mentorship and guidance have been indispensable during these years, and have been instrumental in shaping my growth as a scientist.

I would also like to thank my committee members, Dr. Radhey Gupta and Dr. Daniel Yang for their guidance and feedback during the course of this project. Their careful criticism and valuable advice have imparted me with the knowledge and wisdom that I needed to succeed in my graduate studies.

Additionally, I would like to thank the Junop lab for their friendly and exuberant attitude during my time in the lab. The level of support that they have provided during these three years has been phenomenal. I would like to especially thank Wilson Lee, Beverlee Buzon and Mac Mok, who have been there to help me in difficult times. Special thanks also to my undergraduate students: Daniel Newsted and Cameron Rzadki, for the hard work they have performed.

Finally, I would like to thank my family. My parents: Doris Ngai and Jimmy Huang; my sister: Jennifer Huang; and my brother-in-law: Ken Tang. Your love and support have been instrumental in shaping me as a person, and guiding me to where I am today.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

5' OH	5' hydroxyl
5' PO_4	5' phosphate
6-FAM	6-fluoresceninamidite
ATR	Ataxia telangiectasia and Rad3-related
BHQ1-dt	Black hole quencher 1-derivatized deoxythymidine
BRCA1	Breast cancer type 1 susceptibility protein
BSA	Bovine serum albumin
Ca^{2+}	Calcium ion
CHK1	Checkpoint kinase 1
Co^{2+}	Cobalt (II) ion
CPSF	Cleavage and polyadenylation specificity factor
Da-dt	Dabycl-derivatized deoxythymidine
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
DNA-pKC	DNA dependent protein kinase catalytic subunit
DSB(s)	Double-stranded break(s)

DTT	Dithiothreitol
dsDNA	Double-stranded DNA
E. Coli	*Escherichia coli*
EDTA	Ethylene-diaminetetraacetic acid
EGTA	Ethylene glycol tetraacetic acid
EME1	Essential meiotic endonuclease 1
EMSA	Electrophoretic mobility shift assay
ERCC1	Excision repair cross-complementation protein 1
ESI	Electrospray ionization
FA	Fancomi Anemia
FAN1	Fanconi Anemia Nuclease 1
FANC	Fanconi Anemia Complementation Group
FAAP	Fanconi Anemia Associated Protein
Fe^{3+}	Iron (III) ion
Fl-dt	Fluorescein-derivatized deoxythymidine
FPLC	Fast protein liquid chromatography
FTMS	Fourier transform mass spectrometry
FWHM	Full width at half maximum
GST	Glutathione-S-transferase
His_6	Hexa-histidine
HR	Homologous recombination
HTS	High throughput screening
ID	FANCD2-FANCI
IPTG	Isopropyl β-D-thiogalactopyranoside
IR	Ionizing radiation
IRES	Internal ribosome entry site
k_{cat}	Turnover number
K_d	Dissociation constant
K_m	Michaelis constant
LB	Luria broth
LC	Liquid chromatography
ICL(s)	Interstrand cross-link(s)
IEC	Ion exchange chromatography
IMAC	Immobilized metal affinity chromatography
MBL	Metallo-β-lactamase
MBP	Maltose binding protein
Mg^{2+}	Magnesium ion
MHF1/2	Histone fold proteins 1/2
MMC	Mitomycin C
Mn^{2+}	Manganese (II) ion
mRNA	Messenger RNA
MS/MS	Tandem mass spectrometry
MSA	Multi-sequence alignment
NEM	N-ethylmaleimide

NER	Nucleotide excision repair
NFR	Nuclear foci region
NHEJ	Non-homologous end joining
Ni^{2+}	Nickel (II) ion
NLR	Nuclear localization region
nt	Nucleotide
OD_{600}	Optical Density (600 nm)
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PARP1	Poly(ADP-ribose) polymerase 1
PBS	Phosphate buffered saline
PCNA	Proliferating cell nuclear antigen.
PCR	Polymerase chain reaction
PDB	Protein databank
PIP	PCNA interacting protein
PMF	Peptide mass fingerprinting
Pol ζ	DNA polymerase ζ
PSO2	Sensitivity to Psoralen
PTH	Phenylthiohydantoin
PVDF	Polyvinylidene difluoride
QTOF	Quadropole time-of-flight
RING	Really interesting new gene
RNA	Ribonucleic acid
RPA	Replication protein A
SDS	Sodium dodecyl sulfate
SGC	Structural Genomics Consortium
SNM1	Sensitivity to nitrogen mustard
ssDNA	Single-stranded DNA
TEV	Tobacco etch virus
Tm	Melting temperature
TOF	Time-of-flight
TLS	Translesion synthesis
tRNA	Transfer RNA
UAF1	USP1 associated factor
UBE	Ubiquitin-conjugating enzyme
UBZ	Ubiquitin-binding zinc finger
UMP	Uridine 5'-monophosphate
USP1	Ubiquitin specific peptidase 1
UV	Ultraviolet
V(D)J	Variable diverse joining
XPF	Xeroderma pigmentosum
Zn^{2+}	Zinc (II) ion

CHAPTER 1 – INTRODUCTION

1.1 DNA Interstrand Cross-links

DNA interstrand cross-links (ICLs) are toxic lesions formed when various chemical agents bind to chromosomal DNA and form a covalent bond between opposing strands, preventing unwinding of the double helix (Hlavin *et al*, 2010). The presence of an ICL therefore blocks replication or transcription of the affected DNA and leads to cell cycle arrest. If unrepaired, an ICL may result in the formation of a DNA double-stranded break (DSB), mutation or chromosomal rearrangement, leading to cell death (Akkari *et al*, 2000; Bree *et al*, 2004). For these reasons, an ICL represents one of the most toxic forms of DNA damage, with an estimated forty ICLs being lethal to repair-deficient mammalian cells (Lawley & Phillips, 1996). Interestingly, this form of DNA damage is unique; as it is solely caused by exposure to ICL-inducing agents, most of which are only encountered in a clinical setting (Schrarer, 2005).

ICL-inducing agents are capable of reacting with DNA to create a wide variety of lesions. These include DNA monoadducts, which affect a single base; intrastrand cross-links, which occur between adjacent bases on the same strand; and ICLs (Scharer, 2005). Of these, ICLs typically constitute 1-5% of the total adducts formed, with the majority being monoadducts and intrastrand cross-links. In spite of this, ICLs are largely responsible for the cytotoxicity of these compounds, since they are more resistant to removal by DNA repair pathways (O'Connor & Kohn, 1990). It should be noted, that while all cross-linking agents share these common features, they differ in both their mechanism of action and

the structure of the ICLs they form (figure 1.1). As such, these cross-linking agents can be categorized based on their mechanism of action, and fall into four main classes: the platinums, the alkylating agents, the furocoumarins and mitomycin C (Deans & West, 2011).

The platinums and the alkylating agents represent two classes of cross-linking agents that are synthetic in origin. In the case of the platinums, this encompasses compounds such as cisplatin that are defined by a central platinum (II) atom (Siddik, 2003). Alkylating agents, on the other hand, are bifunctional electrophiles and can be sub-divided into the nitrogen mustards and the nitrosoureas (Scharer, 2005). These compounds contain chloroethyl groups that act as electrophiles, and result in the alkylation of DNA. Both the platinums and alkylating agents have a similar mechanism of action, in which they are activated in the cell by the sequential displacement of two chloride ions by water molecules (Deans & West, 2011). Following this, nucleophilic substitution from two bases on adjacent DNA strands generates the cross-link. It should be noted that the formation of these ICLs is DNA sequence specific, with platinum-based compounds and nitrogen mustards favoring adduct formation at the N-7 of two guanines (Huang *et al*, 1995; Rink *et al*, 1993). Whereas the ICLs generated by the nitrosoureas occur between the N-3 of cytosine and the N-1 of guanine (Ludlum, 1997; Fischhaber *et al*, 1999).

In contrast to the platinums and alkylating agents, the furocoumarins and mitomycin C are two natural sources of ICL-inducing agents. Furocoumarins,

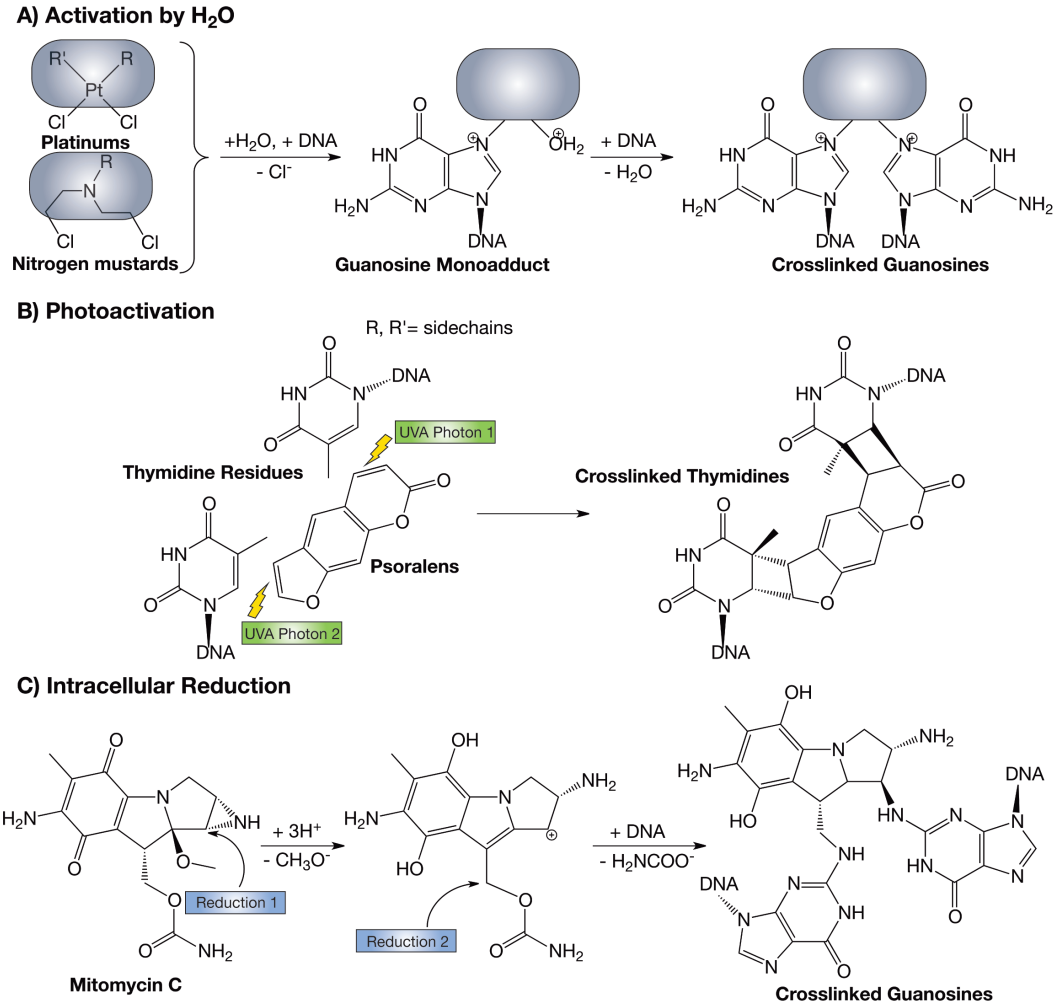

Figure 1.1: Mechanisms of ICL formation for different cross-linking agents. **(A)** The platinums and nitrogen mustards react in a two-step mechanism, where they are activated by water to form a monoadduct with guanosine. A subsequent reaction with another guanosine forms the cross-link. **(B)** In the presence of UVA light, psoralen reacts with two thymidine residues in a cycloaddition reaction. **(C)** Mitomycin C is reduced into an active bis-electrophile and proceeds to form a cross-link with two guanosine residues (Deans & West, 2011).

such as psoralen, are natural products that can be found in a variety of plants, which include bergamot, celery and parsley (Ashwood-Smith *et al*, 1980; Manderfeld *et al*, 1997). While mitomycin C is an antitumor antibiotic isolated from *Streptomyces caespitosus* and *Streptomyces lavendulae* (Danshiitsoodol *et al*, 2006). Furocoumarin ICL adducts are unique, since they are formed through a photon-mediated cycloaddition reaction. In the case of psoralen, exposure to long-wave ultraviolet light (UVA 315-400 nm) causes cross-linking to occur between two thymine residues (Zajdela & Bisagni, 1981; Hearst, 1989). Mitomycin C, on the other hand, requires reduction from cellular processes to convert it into an active bis-electrophile form (Tomasz, 1995). This subsequently reacts with DNA to form an ICL between the N-2 of two guanine residues.

In addition to these exogenous cross-linking agents, certain endogenous compounds are capable of generating ICL adducts. Lipid peroxidation, for example, generates malondialdehyde, crotonaldehyde and acrolein (Niedernhofer *et al*, 2003; Kozekov *et al,* 2003). These compounds contain either two aldehyde groups (malondialdehyde) or a α-β unsaturated aldehyde (crotonaldehyde & acrolein), and function similarly to the alkylating agents. Furthermore, malondialdehyde, crotonaldehyde and acrolein share a common site-specificity, and selectively cross-link two guanine residues at the N-2 position (Dooley *et al*, 2003; Kozekov *et al,* 2003). It is currently unclear how biologically relevant these cross-links are, since studies have shown that they are formed infrequently, and are unstable relative to other ICLs.

1.2 ICL Damage and Chemotherapy

Since ICLs are particularly lethal to dividing cells, DNA cross-linking agents are commonly used in the treatment of cancer. The discovery of these compounds as chemotherapeutics originates from the use of chemical weapons during the Second World War (Deans & West, 2011). Specifically, physicians noted that victims of sulphur mustard attacks had decreased levels of leukocytes, and hypothesized that these agents could be used in the treatment of leukaemia. The first paper on nitrogen mustard treatment of lymphomas was subsequently published in 1946 (Goodman *et al*, 1946). Since then, compounds such as cisplatin, psoralen and mitomycin C have been added to the repertoire of cross-linking drugs, and represent a first line of treatment for solid and haemotological malignancies. In particular, cisplatin is clinically used for testicular and ovarian cancers (Knox *et al*, 1986); mitomycin C is used for oesophageal and bladder tumors (Stoll *et al*, 1985); psoralen is used for cutaneous T cell lymphomas (Querfeld *et al*, 2005); and nitrogen mustard derivatives are used for chronic lymphocyte leukaemia (Rai *et al*, 2000).

While the advantages of ICL-based chemotherapy are obvious, there are notable drawbacks of using these drugs. Cancer cells often develop resistance to cross-linking agents during the course of treatment. Cisplatin is a prominent example of this, since it is effective initially in treating ovarian and small cell lung cancers, but has a high relapse rate of 95% (Giaccone, 2000). There are several mechanisms for this resistance, which include increased drug efflux, increased

detoxification and deactivation of apoptosis pathways (Siddik, 2003). Interestingly, studies have shown that many tumor cells resistant to cross-linking drugs also have upregulated levels of ICL repair proteins (Dabholkar *et al*, 1994; Lee *et al*, 1993). One example of this is the excision repair cross-complementation 1 protein (ERCC1), which is expressed at higher levels in tumor cells resistant to cisplatin (Lee *et al*, 1993). It has therefore been suggested that inhibitors of ERCC1 or other ICL repair proteins could represent a potential therapeutic, since they may resensitize cancer cells to cross-linking agents (Deans & West, 2011).

1.3 ICL Repair in Humans

ICL repair in higher eukaryotes can be categorized based on the context in which the lesion is detected (Hlavin *et al*, 2010). During S-phase, the ICL is recognized when one or two replication forks stall upon encountering the ICL lesion. By contrast, ICLs can be detected in other phases of the cell cycle through the stalling of RNA polymerase during transcription, or by the recognition of DNA distortion caused by the lesion (Muniandy *et al*, 2009). With this in mind, ICLs are predominantly removed during S-phase in a replication-dependent manner (Kim & D'Andrea, 2012). This process involves the interaction of the Fanconi Anemia (FA), nucleotide excision repair (NER), translesion synthesis (TLS) and homologous recombination (HR) pathways in a concerted mechanism to repair an ICL (Kim & D'Andrea, 2012).

1.3.1 Activation of Coordinated ICL Repair

In the first step of ICL repair, the FA pathway is activated during S-phase, when one or two replication forks stall upon encountering an ICL (figure 1.2) (Niedernhofer, 2007). The resulting structure is recognized by FANCM, which associates with the Fanconi anemia associated protein 24 (FAAP24) and histone fold proteins 1 and 2 (MHF1/2) to bind to the stalled replication fork (Ciccia *et al*, 2007; Singh *et al*, 2010). Once bound, FANCM stabilizes the replication fork by using its ATPase-dependent helicase activity to remodel the replication fork and promote fork regression (Gari *et al*, 2008). This is required to prevent promiscuous HR events caused by continued DNA synthesis. FANCM then activates the ATR (ataxia telangiectasia and Rad3-related)-mediated checkpoint by recruiting ATR to the replication fork. Following this, the ATR and CHK1 kinases initiate a phosphorylation cascade and induce cell cycle arrest until the ICL lesion can be removed (Collis *et al*, 2008).

Once the replication fork is stabilized and the ATR cell cycle checkpoint initiated, FANCM then recruits the FA core complex (consisting of the seven FA proteins: FANCA, B, C, E, F, G and L) to the site of the ICL (Niedernhofer, 2007). While most of the proteins within the FA core complex do not have any known catalytic function individually, the entire ensemble as a whole serves as a multi-subunit RING E3 ubiquitin ligase (Alpi *et al*, 2009). The FA core complex has also been shown to interact with Fanconi Anemia associated proteins FAAP100 and FAAP20 for proper function (Ling *et al*, 2007; Kim *et al*, 2012).

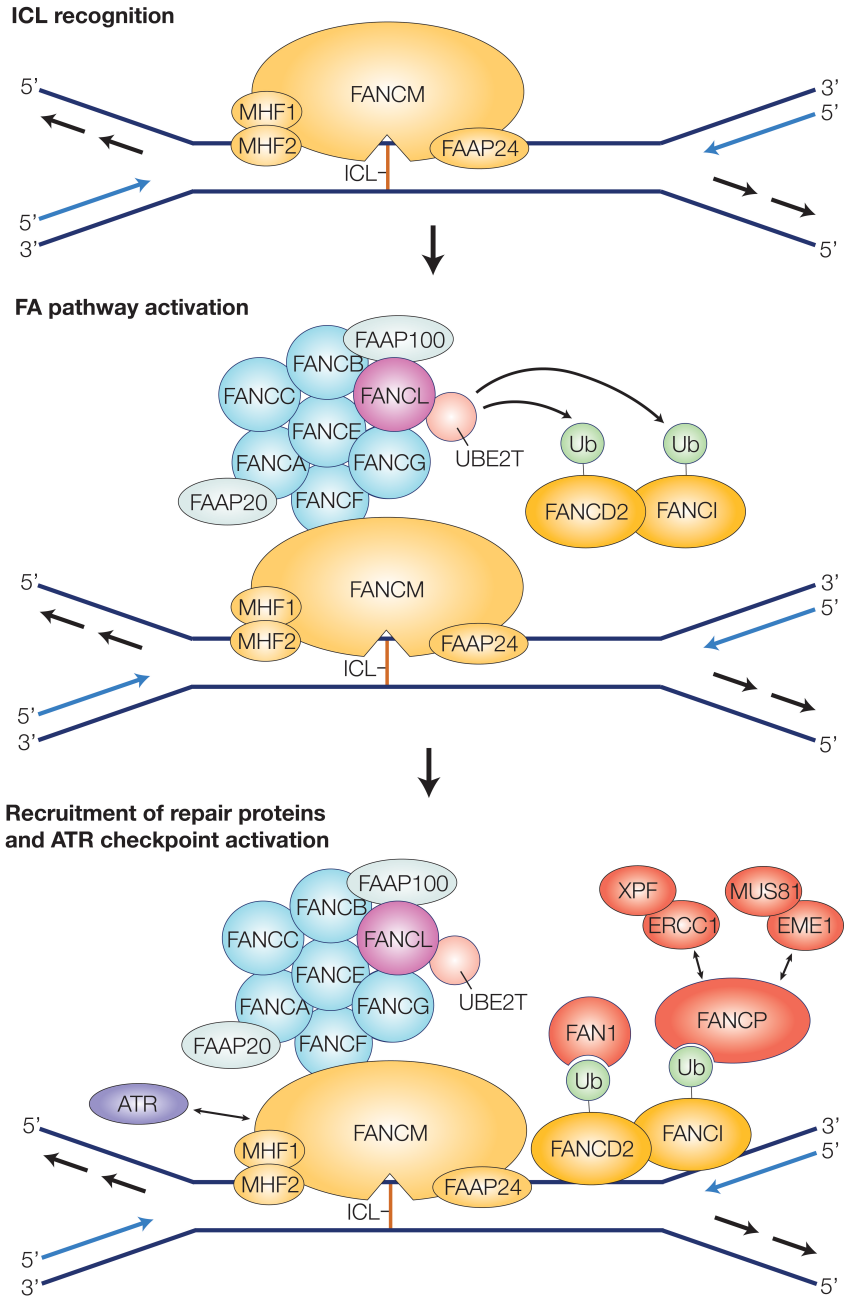

Figure 1.2: The Fanconi anemia pathway initiates DNA interstrand cross-link repair. The ICL lesion is detected upon replication fork stalling by the FANCM-FAAP24-MHF complex. Following this, the FA core complex (FANCA, B, C, E, F, G, and L) along with FAAP100 and FAAP24, associates with FANCM to monoubiquitinate the FANCD2-FANCI (ID) complex. FANCM also activates the ATR checkpoint to cause cell cycle arrest. The FAN1, XPF-ERCC1, and MUS81-EME1 endonucleases are then localized to the ICL lesion, where they participate in the first step of ICL repair.

Within this complex, FANCL acts as the catalytic E3 ubiquitin ligase and associates with the E2 ubiquitin-conjugating enzyme Ube2 to monoubiquitinate FANCD2 and FANCI (Alpi *et al*, 2006). Interestingly, it has been shown that FANCL and Ube2 are sufficient for FANCD2-FANCI monoubiquitination *in vitro*, but are insufficient *in vivo*, with the entire FA core being required for activity. It has therefore been suggested that the other FA core subunits play a role in mediating interactions with FANCM and stabilizing the overall complex (Alpi *et al*, 2009).

The monoubiquitination of FANCD2-FANCI (ID) is the key event in the FA pathway, as the activated complex is required for downstream repair events (Alpi *et al*, 2009). Upon being monoubiquitinated, the ID complex is relocated to the ICL stalled replication fork, where it recruits FANCP/SLX4 and the Fanconi anemia nuclease 1 (FAN1) (Smogorzewska *et al*, 2010; Stoepker *et al*, 2011). Both FAN1 and FANCP/SLX4 interact with the activated ID complex using their ubiquitin-binding zinc finger 4 (UBZ4) domains. FAN1 is an endonuclease that is involved with excising the ICL adduct, while FANCP/SLX4 acts as a scaffold for recruiting the endonucleases XPF-ERCC1 and MUS81-EME1. Overall, the ID complex is required for recruiting and coordinating the endonucleases required in the first step of repair (Kim & D'Andrea, 2012).

1.3.2 Coordination of ICL Unhooking

Once the FA pathway has initiated repair, endonucleases are required to incise 5' and 3' of the ICL on a single strand in order to excise the lesion (figure

1.3) (Raschle *et al*, 2008). This process is known as ICL unhooking and generates two DNA substrates: a DNA double-stranded break (DSB) and an unhooked oligonucleotide tethered to the opposite strand by the cross-link. Currently, three endonucleases are known to function in the unhooking process – XPF-ERCC1, FAN1, and MUS81-EME1 (Smogorzewska *et al*, 2010; McNeil *et al*, 2012; Hanada *et al*, 2006). All three are structure-specific endonucleases that cleave DNA duplexes with a 5' or 3' single-stranded DNA (ssDNA) overhang. These structures are referred to as 5' or 3' flaps and are commonly found within stalled replication forks. Specifically, XPF-ERCC1 and MUS81-EME1 are both 3' flap endo-nucleases, while FAN1 is a 5' flap endonuclease (Smogorzewska *et al*, 2010; McNeil *et al*, 2012; Hanada *et al*, 2006). Given that incisions must be made 5' and 3' to the ICL, a combination of these endonucleases is required for the unhooking reaction.

Although the endonucleases involved with ICL unhooking have been identified, the coordination and regulation of these nucleases is still unknown. The identity of the nuclease required for the initial incision in particular, is highly debated. Functional studies have shown that cells deficient in XPF-ERCC1, MUS81-EME1 and FAN1 are all hypersensitive to cross-linking agents (Smogorzewska *et al*, 2010; McNeil *et al*, 2012; Hanada *et al*, 2006). However, only MUS81-EME1-deficient cells show a significant reduction of ICL-induced DSBs, suggesting its involvement with the first incision of the stalled replication

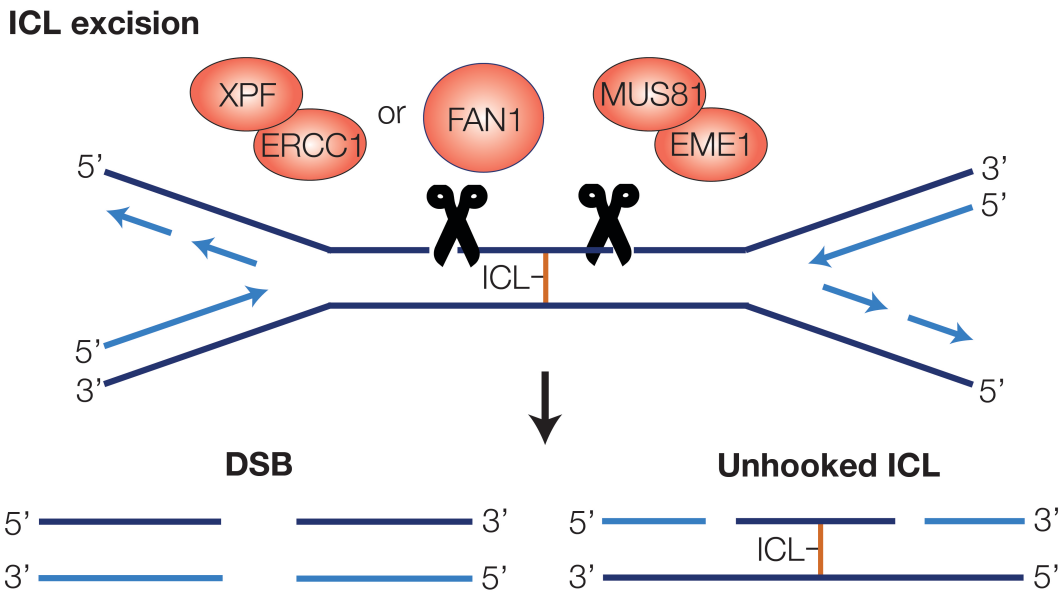

Figure 1.3: ICL excision by DNA endonucleases. Upon being recruited to the ICL by the ID complex, the XPF-ERCC1, FAN1, and MUS81-EME1 endonucleases are coordinated to remove the ICL lesion. This requires incisions to be made both 5' and 3' to the ICL adduct on one strand. MUS81-EME1 incises 3' to the ICL, while XPF-ERCC1 and/or FAN1 incise 5' to the ICL to generate a double-stranded break (DSB) and an "unhooked" ICL intermediate.

fork (Hanada *et al*, 2006). Subsequent incision by XPF-ERCC1 and/or FAN1 then generates the unhooked oligonucleotide intermediate. Additional studies have shown MUS81-EME1 incises 3' to the ICL, while XPF-ERCC1 and FAN1 incise 5' to the ICL. Although the incisions made by FAN1 and MUS81-EME1 are in accordance with their flap endonuclease activities, XPF-ERCC1 incises ICL substrates with an opposite polarity than what would be expected. It is possible that the ICL substrate is somehow structurally distinct from a flap substrate, resulting in XPF-ERCC1 cleavage 5' to the lesion.

1.3.3 Regulation of TLS and lesion bypass

Following the unhooking of the ICL adduct, the resulting gap must be filled in by DNA polymerases and ligated to restore an intact strand. This is a simple process in normal NER, since the complementary strand remains intact as a template for DNA synthesis. However, in the context of ICL repair, the complementary strand contains a lesion and is also damaged (Harder *et al*, 1976). This necessitates the use of special low fidelity polymerases capable of translesion synthesis (TLS) (Lange *et al*, 2011). These polymerases are capable of bypassing a lesion due to their unconstrained active site, which can accommodate distorted bases and base pair mismatches. Previous studies have shown that the TLS polymerases Rev1 and Pol ζ are crucial for the bypass of an ICL, as cells deficient in Rev1 or Pol ζ exhibit ICL hypersensitivity. Furthermore, the genes encoding Rev1 and Pol ζ are epistatic for MMC sensitivity, suggesting that Rev1 and Pol ζ cooperate in a single pathway for TLS (Mirchandani *et al*, 2008).

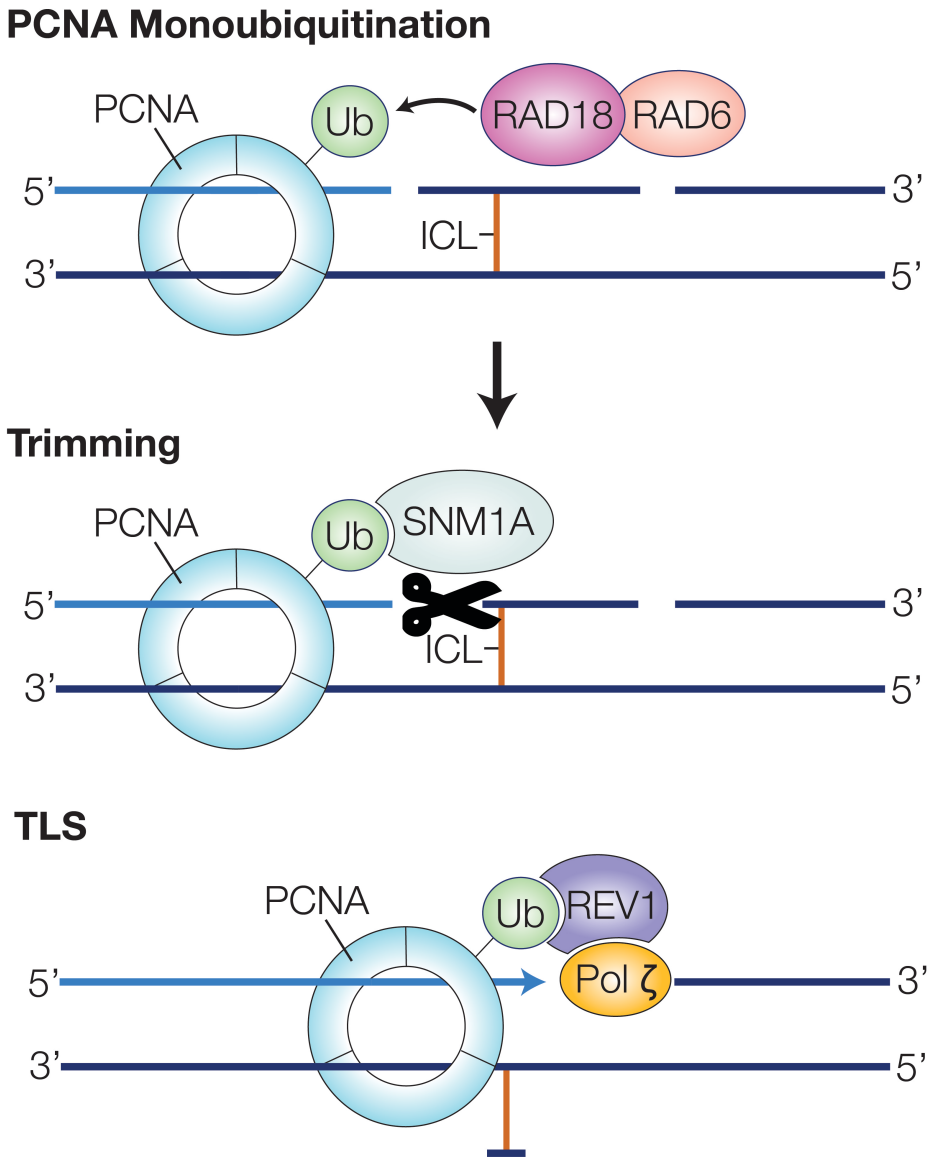

Figure 1.4: Translesion synthesis (TLS). The TLS pathway is activated by RAD18-RAD6 monoubiquitination of proliferating nuclear cell antigen (PCNA). Monoubiquitinated PCNA recruits SNM1A to the lesion, where it 'trims' the ICL-tethered oligonucleotide using its 5' exonuclease activity to generate a favorable substrate for TLS. PCNA then recruits Pol ζ through Rev1 and lesion bypass is performed by Pol ζ to restore the DNA strand. Following this, nucleotide excision repair (NER) removes the ICL adduct from the opposite strand to restore the DNA duplex.

The process of TLS during ICL repair is initiated when the E3 ubiquitin ligase RAD18-RAD6 monoubiquitinates proliferating cell nuclear antigen (PCNA), a clamp-like processivity factor associated with DNA polymerases during replication (figure 1.4) (Geng *et al*, 2010). The monoubiquitination of PCNA triggers the release of the processive DNA polymerase and the recruitment of Rev1 in a process known as polymerase switching. The FA pathway has also been suggested to regulate this step, as recent studies have shown that FAAP20 of the FA core complex assists in localizing Rev1 to the ICL lesion (Kim *et al*, 2012; Masuda *et al*, 2003). It is important to note that Rev1 is not technically a DNA polymerase, but rather a deoxycytidine transferase that inserts a cytosine base into a nascent strand regardless of template sequence (Masuda *et al*, 2003). Interestingly, this transferase activity is not required for TLS, as catalytically inactive Rev1 is still capable of promoting lesion bypass (Ross *et al*, 2005). Instead, Rev1 acts as a scaffold for Pol ζ and other possible TLS polymerases for the lesion bypass step (Ross *et al*, 2005; Guo *et al*, 2003). Pol ζ subsequently extends the nascent DNA strand up to the ICL lesion, where it inserts a random base before continuing synthesis. This random base insertion is often mutagenic, but is a necessary compromise to prevent chromosomal rearrangements that may occur in the absence of lesion bypass (Lange *et al*, 2011). Following this, nucleotide excision repair (NER) excises the remaining ICL adduct from the opposite strand to restore the DNA duplex (Kim & D'Andrea, 2012).

While Pol ζ and Rev1 are the key elements in TLS, the SNM1A exonuclease has also been shown to be involved in this step of ICL repair (Wang *et al*, 2011). Studies have shown that SNM1A interacts with monoubiquitinated PCNA using its UBZ4 domain to localize to the unhooked ICL adduct (Yang *et al*, 2010). Following this, SNM1A degrades the unhooked oligonucleotide to a single covalently linked nucleotide using its 5' exonuclease activity in a process known as trimming (Wang *et al*, 2011). This allows for efficient TLS, as the single nucleotide is not base-paired to the complementary strand and is more readily displaced by Pol ζ.

1.3.4 HR and DSB Repair

In addition to the unhooked ICL intermediate, ICL excision generates a DSB, which must be repaired for continued cell survival. Within eukaryotes there are two pathways by which DSBs are repaired – HR and non-homologous end joining (NHEJ) (Lieber *et al*, 2003). The HR pathway uses a sister chromatid as a template for repairing DSBs, while NHEJ directly ligates two double-stranded DNA ends together without the need for homology. Of these two pathways, HR is selectively involved in the repair of ICL-induced DSBs, since cells deficient in NHEJ repair proteins do not exhibit any hypersensitivity to ICL-inducing agents (Frankenberg-Schwager *et al*, 2005).

HR is initiated by resection of the DSB on each side to generate 3' ssDNA overhangs, which are subsequently coated with replication protein A (RPA) to prevent annealing (figure 1.5) (Garcia-Higuera *et al*, 2001). RPA then activates

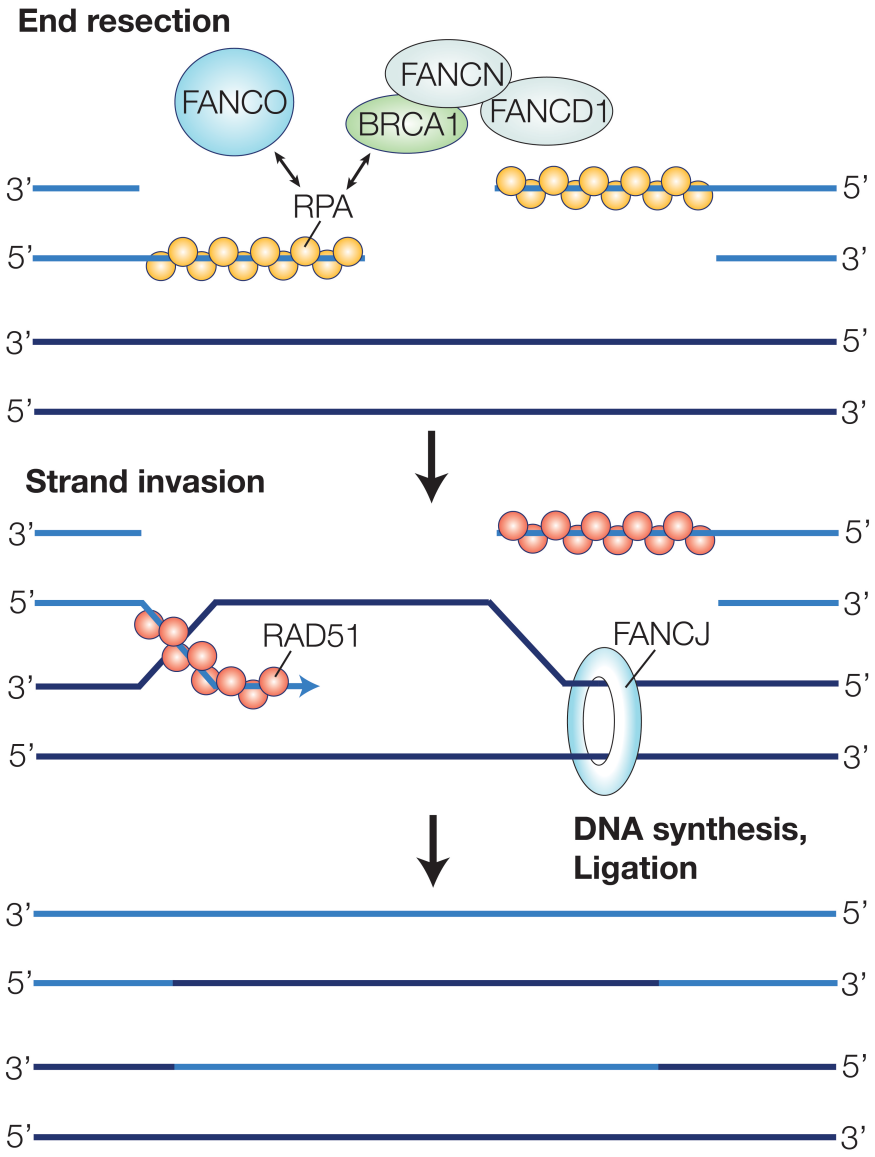

Figure 4: The FA and BRCA proteins participate in homologous recombination (HR) repair of the DSB. The DSB is resected on each side to generate 3' single-stranded DNA overhangs, and coated with replication protein A (RPA) to prevent annealing. RPA then recruits BRCA1, FANCN, FANCD1, and FANCO to facilitate the formation of the RAD51-DNA filament, which initiates strand invasion. The resulting D-loop structure is unwound by FANCJ, allowing for DNA synthesis. Second end capture and branch migration then occurs, followed by Holliday resolution and DNA ligation to restore the DSB.

and recruits the BRCA1-FANCN-FANCD1 complex along with FANCO, which facilitates the displacement of RPA and the loading of RAD51 onto the ssDNA strand. The resulting nucleoprotein filament initiates strand invasion and the formation of a D-loop structure. FANCJ then unwinds the DNA duplex ahead of the invading strand using its 5' to 3' helicase activity to allow access to the template. Second strand capture then occurs, followed by DNA synthesis and Holliday junction resolution to restore the DSB. Following this, the deubiquitinating enzyme USP1-UAF1 removes ubiquitin from the FANCD2-FANCI and PCNA to deactivate the FA, NER, TLS and HR pathways and complete the ICL repair process (Kim *et al, 2009*).

1.4 SNM1A and ICL Repair

While the removal of an ICL involves the recruitment of proteins from other DNA repair pathways, the SNM1A exonuclease is unique as it appears to function solely in ICL repair. This protein was first identified as a member of the β-CASP family of nucleases, which are named after a commonly shared metallo-β-lactamase (MBL) domain and a homology domain from the representative members **C**PSF, **A**rtemis, **S**NM1 and **P**SO2 (figure 1.6) (Callebaut *et al*, 2002). The first member of this family, PSO2 (also known as SNM1 due to *pso2* and *snm1* being allelic) from *Saccharomyces cerevisiae*, was identified in genetic screens (Cassier *et al*, 1980). Specifically, it was found that *pso2* knockouts were highly sensitive to cross-linking agents, but exhibited little sensitivity to other forms of DNA damage. It was therefore suggested that PSO2 played a critical

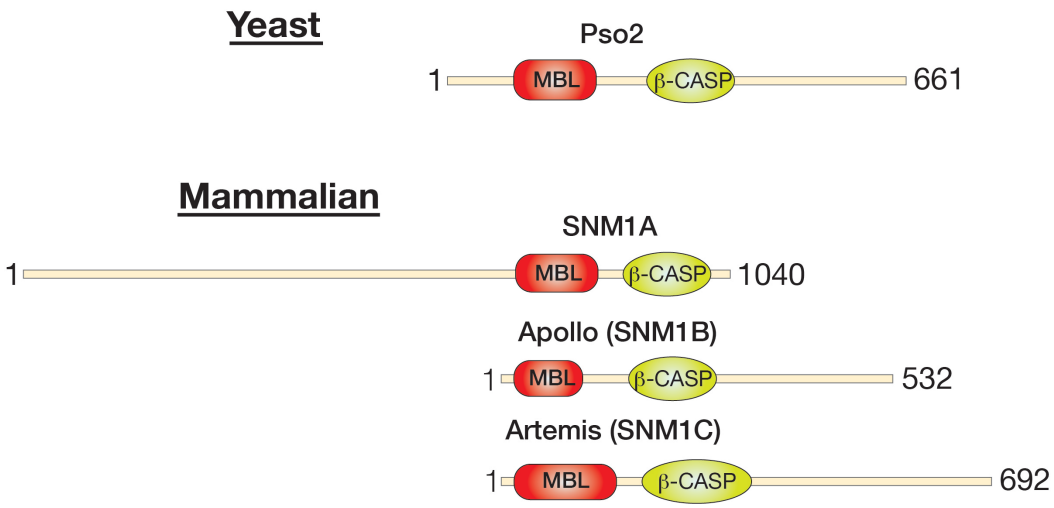

Figure 1.6: The β-CASP family of nucleic acid processing enzymes. All members of this family are 5' exonucleases and are defined by the MBL and β-CASP homology domains. Collectively, the MBL and β-CASP domain are responsible for catalytic activity in these proteins.

role in the repair of DNA cross-links in *S. cerevisiae* that was unaccounted for in previous models of ICL repair (Cassier *et al*, 1980).

Homologues of PSO2 have been found in all eukaryotes examined, with three in mammals: SNM1A, SNM1B/Apollo and SNM1C/Artemis (Cattell *et al*, 2010). All mammalian PSO2 homologues possess 5' exonuclease activity, with PSO2 and Artemis having additional hairpin opening endonuclease function (Tiefenbach & Junop, 2012; Goodarzi *et al*, 2006). Despite having a common exonuclease activity, the biological function of these enzymes is significantly different. Apollo, for example, is involved with telomere maintenance and is suggested to play a minor role in ICL repair (Freibaum & Counter, 2008; Ishiai *et al*, 2004). While Artemis is involved in V(D)J recombination and NHEJ repair of DSBs (Moshous *et al*, 2001; Ma *et al*, 2002). Of these three homologues, however, only SNM1A is able to complement yeast *pso2* knockouts to partially restore ICL repair function (Hazrati *et al*, 2008). This suggests that SNM1A is the functional homologue of PSO2 and may have a similar role in ICL repair in mammals.

Although SNM1A is capable of complementing yeast *pso2* knockouts, its importance in the context of mammalian ICL repair is still unclear (Hazrati *et al*, 2008). Mouse *SNM1A* knockouts demonstrate a two-fold increase in sensitivity to MMC, but not other ICL inducing agents such as melphalan or cisplatin (Dronkert *et al*, 2000). While disruption of *SNM1A* in chicken DT-40 cells results in sensitivity to both cisplatin and MMC (Ishiai *et al*, 2004). This effect is more

pronounced in a double disruption of *SNM1A* and *SNM1B*, suggesting a redundancy for these genes in ICL repair. Human fibroblast *SNM1A* knockdowns show a similar sensitivity to MMC compared to mutant mice cells, with increased ICL-induced radial chromosome formation (Ahkter *et al*, 2005). Collectively, the evidence indicates a mild phenotype for SNM1A depleted cell lines, with sensitivity to select few cross-linking agents. It has been suggested that in the absence of SNM1A, other functionally redundant exonucleases such as SNM1B or FAN1 may be able to partially complement (Cattell *et al*, 2010). Similarly, the selective ICL sensitivity is thought to be from SNM1A's specificity for a subset of ICL adducts, particularly those generated in the minor groove of DNA (Wang *et al*, 2011).

To date, most studies investigating the role of SNM1A in ICL repair have been focused on the human homologue. Full-length human SNM1A is a 116.5 kDa monomeric protein that is expressed at low levels throughout the cell cycle, which is controlled by an internal ribosome entry site (IRES) (Ahkter *et al*, 2004; Zhang *et al*, 2002). Expressed protein is segregated in one or two bodies within the nucleus (Richie *et al*, 2002). Upon induction of DNA damage with IR or cross-linking agents, SNM1A is localized to foci corresponding with formed DSBs. As previously described, this localization is mediated by the RAD18 monoubiquitination of PCNA, which interacts with SNM1A to recruit it to the site of the ICL damage (Yang *et al*, 2010). Once recruited, SNM1A is involved in processing the ICL, following the unhooking reaction from XPF-ERCC1 and

other NER endonucleases. Specifically, a recent study has shown that SNM1A is capable of loading at a nick 5' to an ICL and subsequently digest past the lesion *in vitro* (Wang *et al*, 2011). It is therefore thought that SNM1A acts as a translesional exonuclease trimmer of the unhooked oligonucleotide, thereby providing a better substrate for TLS by Pol ζ (figure 1.4). Interestingly, this novel translesional activity has not been previously reported in the literature for SNM1A or any other nuclease. Moreover, it remains unseen whether or not SNM1A is capable of performing this function *in vivo* (Wang *et al*, 2011).

1.4.1 The Biochemistry of SNM1A

Human SNM1A is a 5' phosphate-dependent exonuclease with a preference for DNA over RNA substrates (Hejna *et al*, 2007; Sengerova *et al*, 2012). This 5' phosphate dependence is characteristic of other members of the β-CASP family, and is shared by PSO2, Apollo and Artemis (Catell *et al*, 2010). Full-length SNM1A was observed to readily hydrolyze ssDNA oligomers, but exhibited greatly reduced activity on dsDNA (~1% of single-stranded) (Hejna *et al*, 2007). A subsequent study with a C-terminal truncation of SNM1A, spanning residues 608 to 1040, demonstrated increased activity compared to the full-length protein (Wang *et al*, 2011; Sengerova *et al*, 2012). Specifically, $SNM1A_{608-1040}$ was found to have a k_{cat} of 25.6 min^{-1} and a K_m of 5 nM for an ssDNA 21-mer, with a k_{cat} of 13.8 min^{-1} and a K_m of 97 nM for a dsDNA 21-mer (Sengerova *et al*, 2012). Since the turnover number (k_{cat}) for both substrates are similar, the preference of $SNM1A_{608-1040}$ for ssDNA is thought to be due to its tighter binding

(K_m). Based on competition assays, it was also determined that $SNM1A_{608-1040}$ is not processive and repeatedly binds and releases the DNA oligomer after each hydrolysis reaction. Interestingly, $SNM1A_{608-1040}$ was found to have increased efficiency for high molecular weight DNA substrates, being capable of digesting 0.5 kilobases of linearized or gapped plasmid DNA over an hour. This reaction, in contrast to those with the shorter substrates, was noted to be processive, as a large excess of unlabeled DNA was needed to inhibit $SNM1A_{608-1040}$ (Wang *et al*, 2011).

In addition to the previous substrates, SNM1A has also been tested with DNA containing flap, hairpin and ICL structures. Both full-length and truncated SNM1A have not demonstrated endonuclease activity on either flap or hairpin substrates (Hejna *et al*, 2007; Sengerova *et al*, 2012). This is an interesting result, since its functional homologue in yeast, PSO2 has been reported to possess hairpin opening endonuclease activity (Tiefenbach & Junop, 2011). In spite of this, a novel exonuclease activity for SNM1A has been reported for a dsDNA substrate containing a single ICL (Wang *et al*, 2011). This substrate, which was generated using the site-specific cross-linking agent SJG-136, was translesionally digested by SNM1A in the presence of a 5' phosphate or a nick produced by XPF-ERCC1. It is important to note that while SNM1A is capable of digesting through the lesion, it is neither an efficient or processive process, since it pauses two nucleotides after the cross-link (Wang *et al*, 2011). It is currently unclear as to why or how SNM1A ceases hydrolysis after proceeding through the ICL adduct.

Full-length SNM1A has been found to function optimally at a slightly alkaline pH and low salt, and requires a divalent metal cation for hydrolysis (Hejna *et al*, 2007). Both Mg^{2+} and Mn^{2+} were found to stimulate SNM1A exonuclease activity, while other tested metal ions such as Fe^{3+}, Zn^{2+}, Co^{2+}, Ca^{2+} and Ni^{2+} were inhibitory at high concentrations (Hejna *et al*, 2007; Sengerova *et al*, 2012). It should be noted that Mn^{2+} was found to partially restore the catalytically inactive SNM1A point mutants D736A and D736A/H737A, which disrupt metal cation binding within the active site (Sengerova *et al*, 2012). It is therefore thought that Mn^{2+} functions by fitting into the empty active site, due to its less stringent coordination requirements. The Mg^{2+} requirement of SNM1A and its inhibition with Zn^{2+} is intriguing, since the active sites of MBL enzymes usually involve the coordination of one or two Zn^{2+} ions (Cattell *et al*, 2010). To assess the identity of the metal ion within the active site, a recent study used the metal chelators ethylene-diaminetetraacetic acid (EDTA), ethylene glycol tetra-cetic acid (EGTA) and *o*-phenanthroline (Sengerova *et al*, 2012). EDTA was used as a general metal chelator, while EGTA and *o*-phenanthroline were used to preferentially chelate Ca^{2+} and Zn^{2+} respectively. Of these three chelating agents, *o*-phenanthroline was found to inhibit strongly inhibit SNM1A; suggesting Zn^{2+} is the metal ion present in the active site (Sengerova *et al*, 2012).

1.4.2 The Structure of SNM1A

Human SNM1A is 1040 amino acids in length, and consists of a long extended N-terminus, with C-terminal MBL and β-CASP domains (figure 1.7)

(Hejna *et al*, 2007). The N-terminal region of this protein is predicted to be disordered, but has been suggested to localize SNM1A within the nucleus. Based on deletion analysis of SNM1A, a nuclear localization region (NLR) and a nuclear foci region (NFR) were identified to span residues 1-190 and 396-614 respectively (Richie *et al*, 2002). The NLR was found to transport SNM1A into the nucleus, while the NFR was required for localizing SNM1A to sites of DNA damage. A sequence analysis also revealed the presence of a UBZ4 motif and a PCNA interacting protein (PIP) box at residues 118-145 and 556-563 respectively (Yang *et al*, 2010). The PIP box was found to be essential for SNM1A to bind to PCNA and form nuclear foci. Furthermore, this interaction was found to be spontaneous, with approximately 20% of cells exhibiting nuclear foci in the absence of DNA damage. By contrast, mutations in the UBZ4 motif did not abolish PCNA-SNM1A interaction, but instead abrogated the ability of SNM1A to localize to nuclear foci in response to DNA damage (Yang *et al*, 2010). Collectively, the data indicate that the UBZ4 motif and PIP box cooperate to bind monoubiquitinated PCNA to localize SNM1A to sites of DNA repair.

The C-terminal MBL and β-CASP domains are responsible for SNM1A exonuclease activity. As the name suggests, the MBL fold was first identified in bacterial enzymes responsible for hydrolyzing β-lactam rings in a metal ion dependent reaction (Wang *et al*, 1999). These enzymes typically require two Zn^{2+} ions as co-factors, and catalyze the breakdown of antibiotics. By contrast, the β-CASP family has evolved from the MBL enzymes to hydrolyze the

phosphodiester backbone of DNA or RNA (Catell *et al*, 2010). Members of this family have an additional β-CASP homology domain inserted into the canonical MBL fold. This β-CASP domain is thought to confer these proteins their nuclease activity, though it is unclear as to how this is accomplished. A sequence analysis of the β-CASP family has also found seven sequence motifs (Callebaut *et al*, 2002). The first four motifs are found within the MBL domain and are as follows: aspartic or glutamic acid (motif 1), HxHxDH (motif 2), histidine (motif 3) and histidine (motif 4). The remaining three motifs are found within the β-CASP insertion: aspartic or glutamic acid (motif A), histidine (motif B) and histidine or valine (motif C). Overall, these sequence motifs are thought to be involved in the binding of Zn^{2+} ions for catalytic activity.

The crystal structure of a C-terminal truncation of human SNM1A (residues 676-1040) was recently solved by the Structural Genomics Consortium (SGC) of Toronto and deposited into the Protein Data Bank (PDB) (figure 1.8). This structure, which spans amino acids 696-1040, encapsulates both the MBL and β-CASP domains (PDB ID 4B87). As figure 1.8 shows, $SNM1A_{696\text{-}1040}$ consists of the MBL and β-CASP domains oriented on a hinge around a Zn^{2+} ion to form the active site. The MBL domain begins at the N-terminus, which consists of an internal β-sandwich (β1–β2–β3–β4–β5–β6–β7–β8–β9, β15) that is protected from the solvent by a series of amphipathic helices (α1–α2–α3, α10–α11) and loops. The β-CASP domain is then inserted following the β9 strand, and consists of an internal parallel β-sheet (β10–β11–β12–β13–β14) protected from

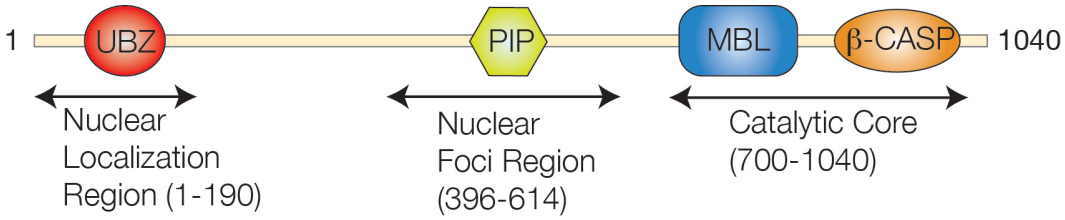

Figure 1.7: Domain boundaries of full-length SNM1A. The long extended N-terminus is involved in the regulation and localization of SNM1A, while the C-terminal MBL and β-CASP domain are responsible for exonuclease activity. A UBZ4 motif (residues 556-563) and a PIP box (residues 118-145) are involved in binding monoubiquitinated PCNA to recruit SNM1A to sites of DNA damage.

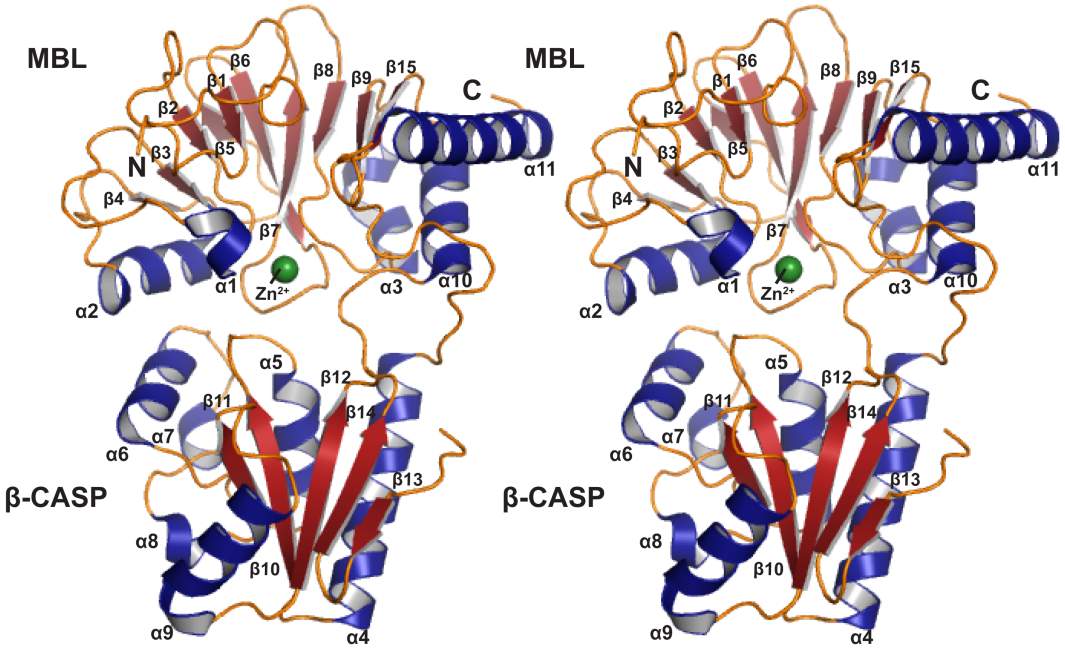

Figure 1.8: Stereo image of SNM1A$_{696-1040}$. This truncation exists as a monomer, which is comprised of the MBL and β-CASP domains. The Zn^{2+} containing active site is found at the interface between the MBL and β-CASP domains (PDB ID 4B87).

the solvent by several amphipathic helices (α4–α5–α6–α7–α8–α9) and loops. Following the β12 strand, an α–β–α (α10–β15–α11) motif extends back into the MBL domain to complete the structure. It should be noted that the loop region connecting β12 and β13 and residues 676-695, are unaccounted for in this structure.

1.4.3 The Mechanism of SNM1A

Although SNM1A has been characterized structurally and biochemically, a mechanism for its role in ICL repair has not been established. Based on studies of other MBL proteins, it is currently thought that SNM1A hydrolyzes the phosphodiester bond of DNA in a two-metal-ion dependent reaction (Cattell *et al*, 2010). In this model, two Zn^{2+} ions are coordinated in the active site by aspartic acid and histidine residues. These in turn, bind to a bridging water or hydroxide ion that is thought to function as the nucleophile in the hydrolysis reaction. Interestingly, only one Zn^{2+} is present within the crystal structure of SNM1A, which is coordinated by residues H732, H734, H793 and D815. This is contrast to other known structures of the β-CASP family, which have two Zn^{2+} ions in the active site. A structural alignment of SNM1A with Apollo (PDB ID 3ZDK) indicates that this missing Zn^{2+} is likely an artifact of the crystal structure, and is furthermore coordinated by the residues D736, H737 and D815 (figure 1.9). This is supported by previous studies that have shown that the D736A and H737A point mutants are catalytically inactive. The H994 residue is also in close proximity to the active site, but does not appear to be involved with Zn^{2+} binding.

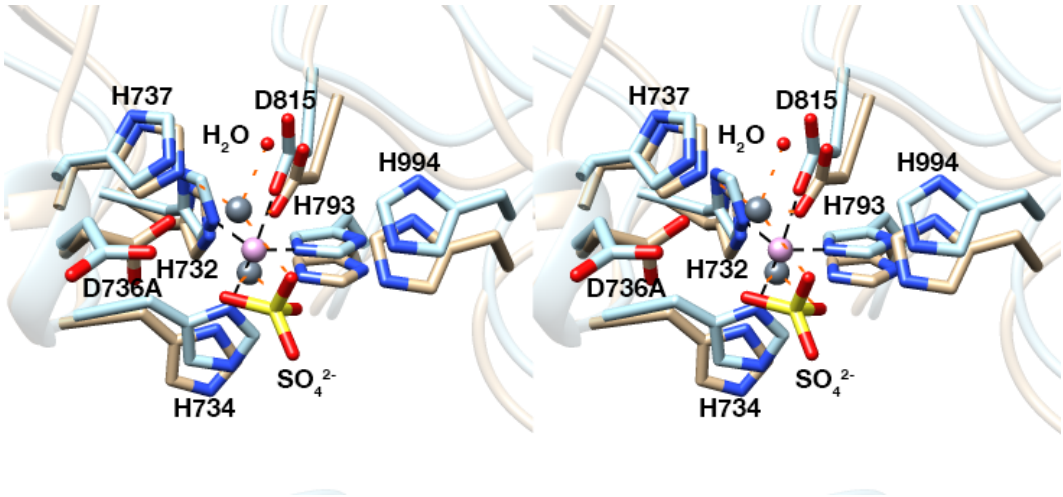

Figure 1.9: Stereo image of the alignment between the SNM1A and Apollo active sites. SNM1A and Apollo are coloured blue and beige respectively, with SNM1A residues labeled. The Zn^{2+} atom of SNM1A is coloured pink, while Apollo's Zn^{2+} atoms are coloured in gray. Metal-ion bonds observed in the SNM1A structure are shown in black, while those found solely within Apollo are orange. A water molecule and a sulfate ion from Apollo are also shown.

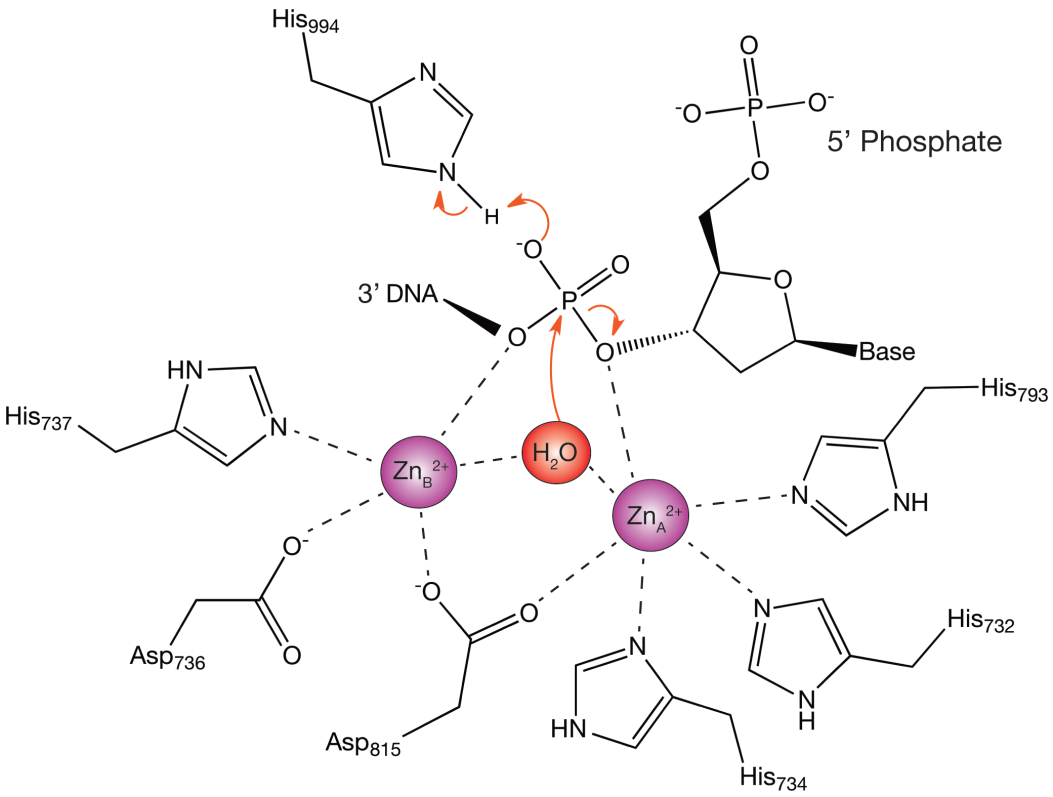

Figure 1.10: Hypothetical mechanism of SNM1A hydrolysis. Residues H732, H734, H793 and D815 coordinate Zn_A^{2+}(as seen in the crystal structure), while residues H732, H734, H793 and D815 coordinate Zn_B^{2+} (predicted). H994 is thought to participate in this reaction as a general acid catalyst. These Zn^{2+} atoms bind to DNA and a water molecule to catalyze hydrolysis. Cleavage of the phosphodiester bond generates a 5'-phosphate substrate, which is required for a subsequent nucleolytic reaction.

It has instead been suggested that this histidine may be involved in general acid catalysis (Catell *et al*, 2010). Interestingly, a cysteine residue is also thought to be involved in this reaction, since the addition of N-ethylmaleimide (NEM) inactivates SNM1A (Hejna *et al*, 2007). Based on this information, a mechanism for SNM1A can be postulated, and is shown in figure 1.10.

While this mechanism is consistent with other MBL enzymes, it does not explain the 5'-phosphate-dependence of SNM1A. Based on analysis of other β-CASP structures, there appears to be two possible functions for this 5'-phosphate. The crystal structure of *Thermus thermophilus* RNase J in complex with uridine 5'-monophosphate (UMP) reveals the presence of a phosphate-binding pocket, suggesting that the 5'-phosphate may be required for correct orientation of the substrate within the active site (Li de la Sierra-Gallay *et al*, 2008). Alternatively, there is also evidence to suggest that the 5'-phosphate is required to coordinate Zn^{2+} or Mg^{2+} in the active site. An example of this can be seen in the crystal structure of Apollo, where an analogous sulfate ion is bound to both Zn^{2+} ions (figure 1.9). To date, however, a clear role for this 5'-phosphate has not been established for the β-CASP nucleases.

The missing metal ion in the SNM1A active site is another question that remains to be addressed. Since other β-CASP nucleases are known to possess two Zn^{2+} ions, it is assumed that this is the missing ion in SNM1A (Cattell *et al*, 2010). However, SNM1A is known to require Mg^{2+} and is inhibited by elevated Zn^{2+} concentrations (Hejna *et al*, 2007; Sengerova *et al*, 2012). This suggests a

role for Mg^{2+} in the mechanism of SNM1A exonuclease activity – possibly as the second metal ion in the active site. Alternatively, it is plausible that Mg^{2+} functions as a third cofactor for SNM1A, though this is less likely, since there have been only two reported nucleases in the literature with a three-metal-ion mode of catalysis: nuclease P1 and *E. Coli* Endo IV (Garcin *et al*, 2008; Ivanov *et al*, 2007; Romier *et al*, 1998). Assuming this is the case, it is possible that high concentrations of Zn^{2+} may inhibit SNM1A by displacing Mg^{2+} from the active site.

While SNM1A has been found to possess exonuclease activity for ssDNA, dsDNA and an ICL substrate, its mode of DNA binding remains to be determined. Specifically, it is unclear how SNM1A is able to accommodate these structurally diverse substrates within its constrained active site. Recently, the crystal structures of *T. thermophilus* RNase J bound to UMP and a 4 nucleotide (nt) RNA molecule have provided some insight into this issue (Dorleans *et al*, 2011) (figure 1.11). Comparing the UMP and RNA bound structures of RNase J, the loops connecting the β-CASP and MBL domains were found to act as a hinge, widening the catalytic cleft by 4°. Surprisingly, this conformational change was enough to accommodate a 4 nt RNA molecule into the active site. With this in mind, it is likely that SNM1A has a similar mode of binding for ssDNA, though it is uncertain how it would process dsDNA or an ICL substrate.

1.5 Research Outline

When the work on this project began, human SNM1A was reported to

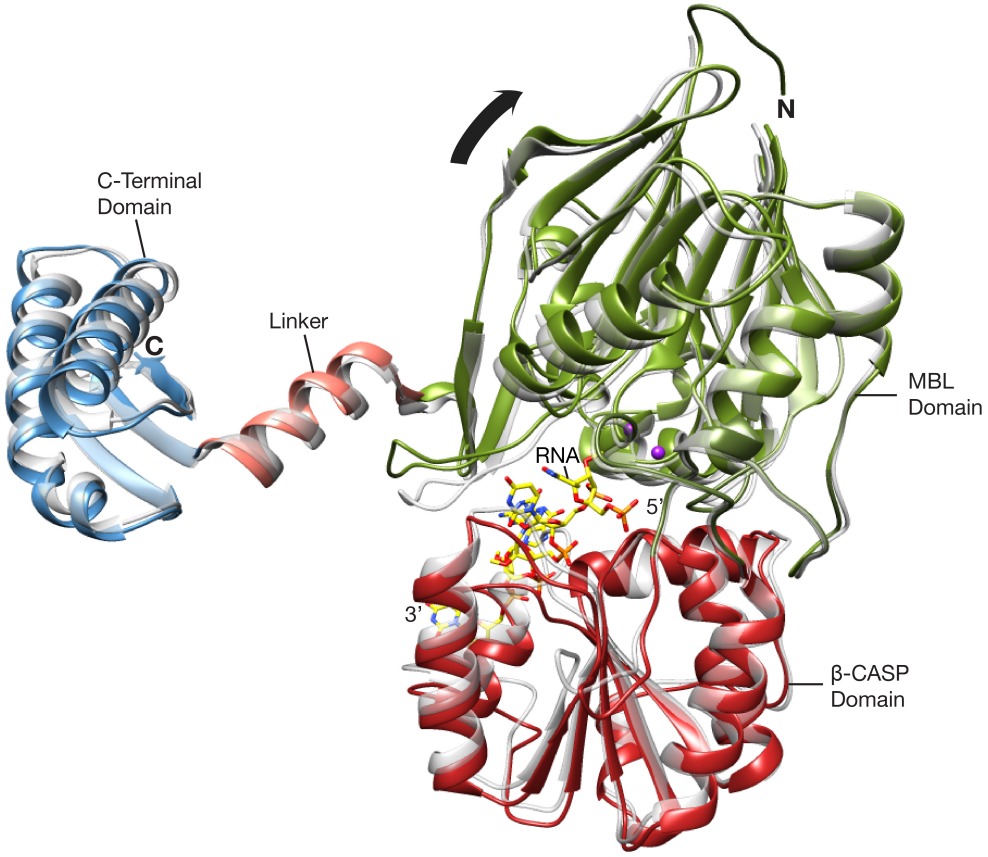

Figure 1.10: Comparison of RNase J structures. The structure of RNase J bound to RNA (coloured according to domain) is superimposed over the UMP bound structure (gray). A RNA 4-mer binds in a cleft between the MBL and β-CASP domains. A widening of this cleft by 4° is sufficient to accommodate this substrate. Structures of RNA-bound and UMP-bound RNase J are from PDB IDs 3T3O and 3BK2 respectively.

have been purified from an *E. Coli* bacterial system, and was shown to possess translesional exonuclease activity on an ICL substrate *in vitro* (Wang *et al.*, 2011). Prior to this, SNM1A was known to be insoluble in bacterial systems and toxic in human cell lines, making it difficult to characterize (Hejna *et al.*, 2007). Therefore, the initial goal of this project was to purify soluble SNM1A from *E. Coli*. Once this was done, it could be biochemically characterized using nuclease assays, and structurally characterized using X-ray crystallography. The objective of this work was to elucidate the mechanism of SNM1A substrate processing in DNA ICL repair. Specifically, determining how SNM1A binds to DNA and the function that the 5' phosphate serves in the hydrolysis reaction. With this in mind, an outline of this thesis is as follows. In Chapter 2, we report the purification of $SNM1A_{698\text{-}1040}$ from *E. Coli* and discuss the advantages and disadvantages of this method. Chapter 3 describes the subsequent biochemical and structural characterization of SNM1A, which includes our attempts at crystallizing this protein. In Chapter 4, we report the development of a fluorescence-based nuclease assay and how it might be used to screen for SNM1A inhibitors. In the final chapter, we summarize the research within this thesis, and discuss future work to be performed.

CHAPTER 2 – PURIFICATION OF SNM1A

2.1 Rationale and Experimental Design

The main goal of this project was to purify human SNM1A from *E. Coli* in sufficient quantities for protein crystallography. Due to previous difficulties encountered in our lab with purifying SNM1A, we decided to approach the problem by adopting a semi-high throughput approach to the purification of SNM1A. First, multi-sequence alignments and secondary structure prediction analyses were performed on the human SNM1A open reading frame (ORF). This was done in order to identify possible truncations of SNM1A that may prove more amenable to purification. Each truncation would then be cloned into five different expression vectors containing different fusion proteins to enhance expression, stability and solubility. Once a library of SNM1A expression constructs was generated, they were systematically tested for expression, solubility and yield, using a small-scale purification assay. Any viable expression constructs identified using this assay would then be scaled up and purified in larger amounts to obtain SNM1A for subsequent structural and function analysis.

2.1.1 Construct Design and Cloning Strategy

SNM1A truncations were designed based on information from sequence alignments and secondary structure predictions. First, a BLAST search was performed using human SNM1A to identify sufficiently distant homologues. A ClustalW2 multiple sequence alignment was then performed with fourteen of these homologues in order to identify regions of high conservation (figure 2.1). In

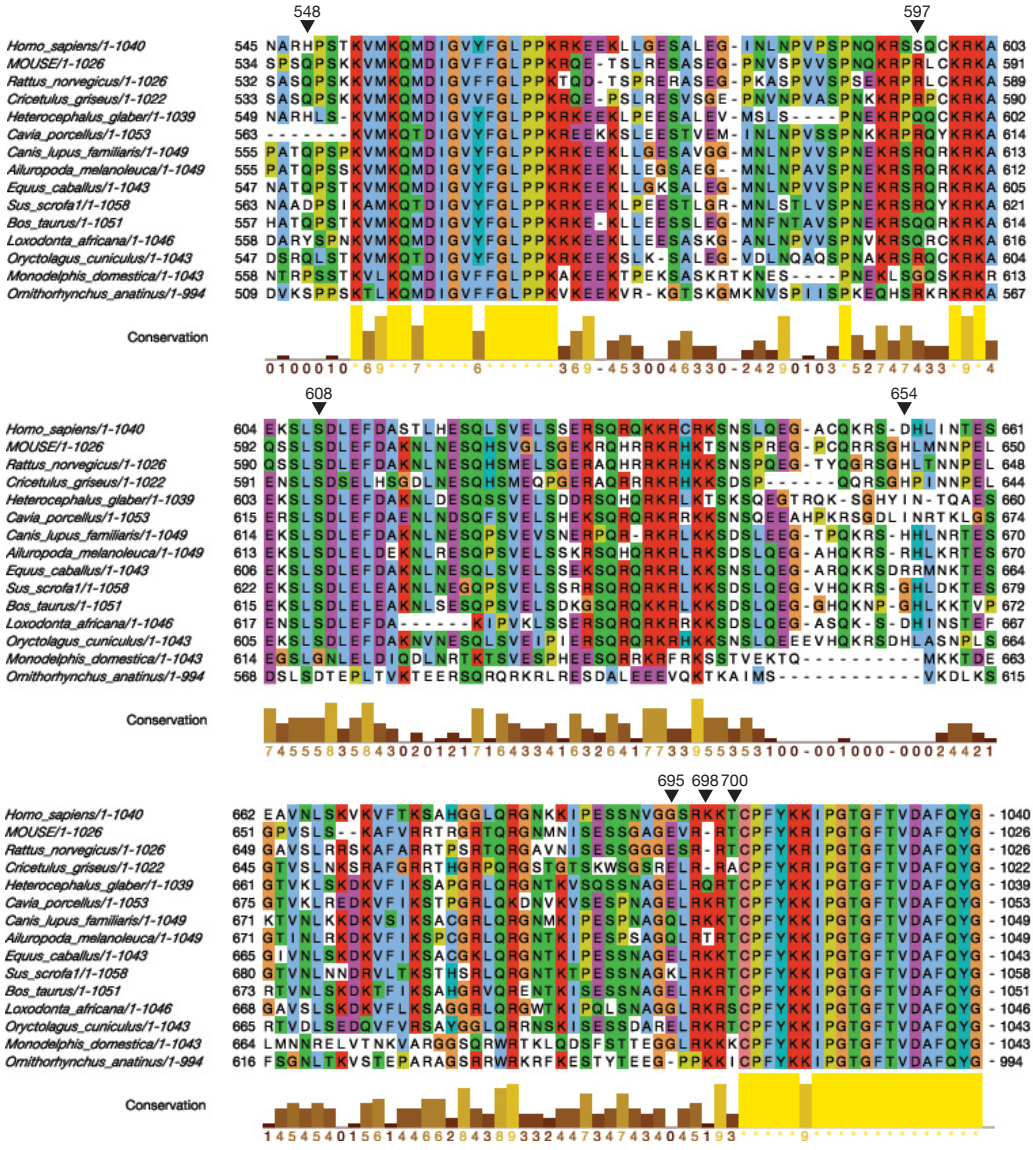

Figure 2.1: SNM1A multiple sequence alignment of residues 545 to 720. Fourteen homologues of SNM1A were aligned using the ClustalW2 server (http://www.ebi.ac.uk/Tools/msa/clustalw2/). Highly conserved residues are annotated, while amino acids are coloured based on similar chemical properties. Triangles indicate start positions for the seven SNM1A truncations: 548H, 597S, 608S, 654D, 695G, 698K and 700T.

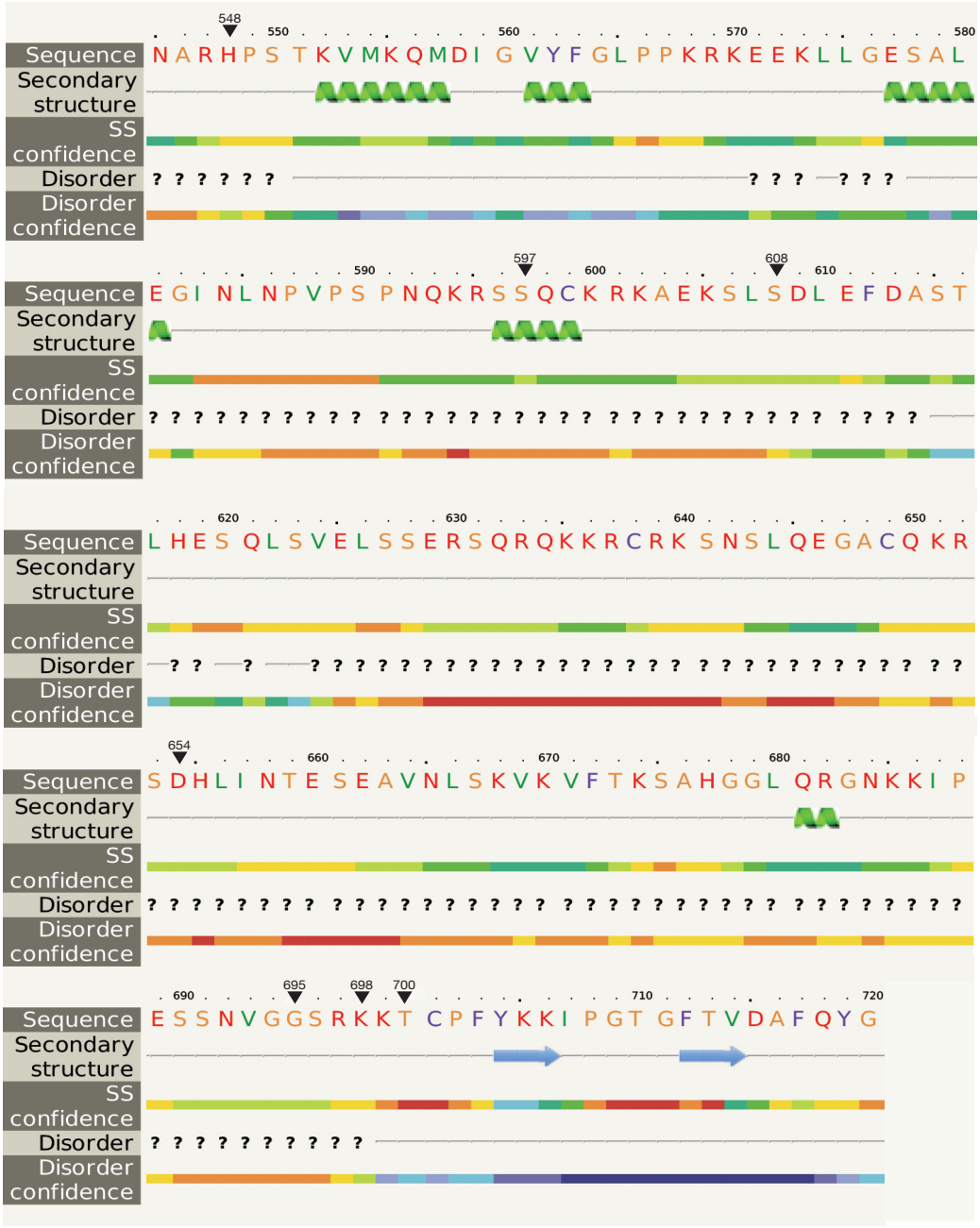

Figure 2.2: Secondary structure prediction of SNM1A residues 545 to 720. A secondary structure analysis of human SNM1A was performed using the Phyre2 server (www.sbg.bio.ic.ac.uk/phyre2/). Predicted secondary structure elements are annotated above the primary amino acid sequence. High and low confidence regions are coloured on a scale from red to blue respectively. Triangles indicate start positions for the seven SNM1A truncations: 548H, 597S, 608S, 654D, 695G, 698K and 700T.

addition to this, a Phyre2 analysis was run on human SNM1A to find well-defined regions of secondary structure (figure 2.2). From this data, the boundary for the MBL and β-CASP domains was predicted to span from residue 700 to the C-terminus at residue 1040. We therefore designed seven truncations of varying length that encapsulated this catalytic domain, spanning residues 548-1040, 597-1040, 608-1040, 654-1040, 695-1040, 698-1040 and 700-1040. Special care was taken to ensure that these truncations did not cut into any regions of high conservation or predicted secondary structure, as doing so may render the resulting protein unstable and/or insoluble. It should be noted that three truncations, namely 695-1040, 698-1040 and 700-1040, are closely spaced together near the predicted domain boundary. This was done to reduce the risk of over-truncating the MBL and β-CASP domains. Lastly, hydrophobic residues were generally avoided in starting positions, with a preference for polar or charged residues instead.

To improve the chances of obtaining soluble protein, each of the seven SNM1A truncations was cloned in parallel into five different expression vectors (figure 2.3). First, the seven truncations were PCR amplified with and without C-terminal His-tags to generate a total of fourteen PCR products. C-terminal His-tags were included to see if they would improve binding to a nickel affinity column. Each of these fourteen PCR products was then cloned into five different expression vectors for a total of seventy constructs. These five vectors contained

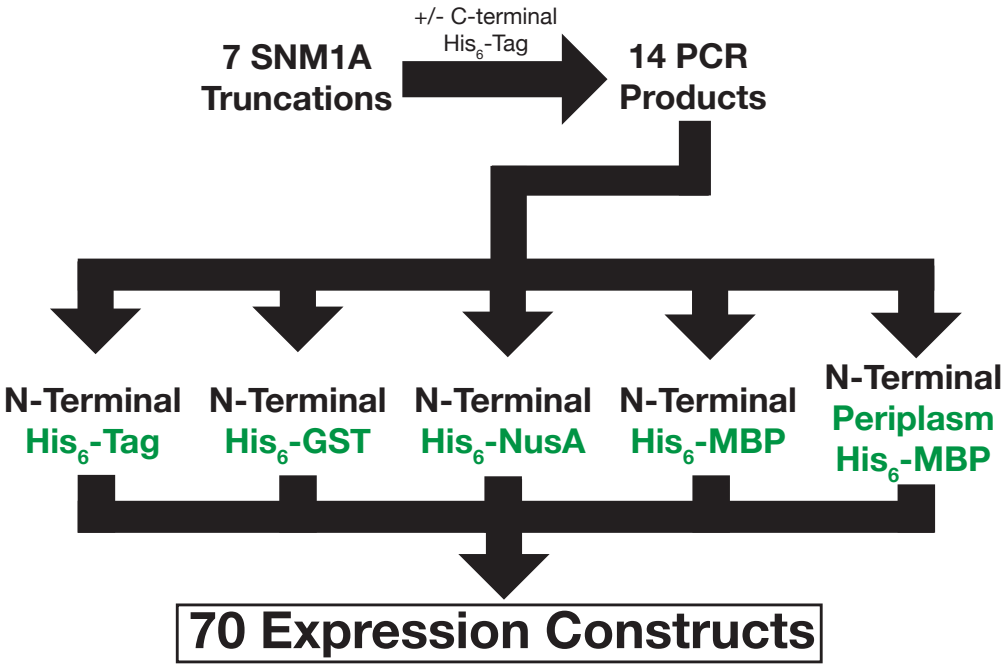

Figure 2.3: Strategy for cloning SNM1A constructs. Seven truncations of human SNM1A were PCR amplified with and without a C-terminal His-tag. These were subsequently cloned into five different expression vectors for a total of seventy SNM1A expression constructs. A TEV site was engineered into every construct to allow for cleavage of N-terminal fusions.

N-terminal His_6, His_6-GST, His_6-MBP, periplasm His_6-MBP, and His_6-NusA tags respectively. GST, MBP, and NusA fusions were used to improve the solubility of recombinant protein, while the inclusion of a periplasmic MBP fusion was used to assess whether an oxidative redox environment would facilitate proper protein folding. Lastly, a TEV site was inserted between the N-terminal fusion and protein for each construct to allow for separation of the tag after IMAC purification.

2.2 Materials and Methods

2.2.1 Purification of $SNM1A_{608-1040}$

The human $SNM1A_{608-1040}$ gene was synthesized by Geneaid Biotech and subsequently incorporated into the pET24a vector (MJ4806). This gene was codon optimized for expression in *E. Coli* and a TEV site was inserted between the His_6-tag and protein. $SNM1A_{608-1040}$ was expressed as a His_6-tagged protein in *E. Coli* Rosetta™ (DE3) pLysS competent cells (Novagen). Bacteria were grown in standard LB medium supplemented with 0.1 mg ml^{-1} ampicillin at 37°C with shaking (225 rev min^{-1}) until the absorbance at 600 nm (OD_{600}) reached 0.5. Protein expression was induced using 1.0 mM IPTG at an incubation temperature of 25°C. After an 8-hour induction time, cells were harvested by centrifugation at 3 315 x G for 15 minutes at 4°C. Collected pellets were resuspended in phosphate buffered saline (PBS) for transfer to storage tubes and subsequently flash frozen in liquid nitrogen and stored at -80°C. Prior to lysis using a French press, cell pellets (4L) were resuspended to a final volume of 60 mL using nickel A buffer

(500 mM NaCl, 50 mM HEPES pH 7.5, 10 mM imidazole). Aprotinin, pepstatin A, benzamidine-HCl, and leupeptin protease inhibitors were added to a final concentration of 3.0 μM, 1 μM, 1 mM and 1 μM respectively. After lysis, samples were centrifuged at 48 384 x G for 45 minutes at 4°C and the supernatants sieved through a membrane filter (0.45 μm pore size). Filtered samples were then applied to a HiTrap nickel affinity column (GE Healthcare) that was set on a FPLC system (AKTA FPLC, Amersham Pharmacia Biotech). Bound proteins were eluted over a gradient of 48-200 mM imidazole, following sequential washes with 10, 29, and 34.7 mM imidazole. Fractions were pooled and protein purity was assessed through SDS-PAGE, while an anti-His_6 Western blot was used to detect the presence of SNM1A. Final protein concentrations were determined using the Bradford assay.

2.2.2 Cloning of SNM1A Expression Vectors

N-terminal truncations of human SNM1A were generated by PCR amplification. Primers were designed as instructed in the Gateway® Technology cloning manual, with the addition of a TEV digestion site in the forward primers. Furthermore, half of the reverse primers had additional codons for a His_6 tag at the C-terminus. PCR was performed using iNtRON Biotechnology *i*-Pfu DNA polymerase with the appropriate primers and plasmids. The codon optimized $SNM1A_{608-1040}$ gene (MJ4806) was used as a template for truncations starting at residues 608, 654, 695, 698, and 700. While the non-codon optimized full-length SNM1A gene (MJ4301) was used as a template for the remaining truncations

starting at residues 548 and 597. Reactions were prepared as instructed in the manufacturer's manual and PCR was performed as follows: initial denaturation (95°C for 5 minutes), followed by 25 cycles of denaturation (95°C for 30 seconds), annealing (60°C for 1 minute), and elongation (68° for 2 minutes 30 seconds), and a final elongation step (72°C for 10 minutes).

PCR products were incorporated into the entry vector pDONR™-201 using the BP Clonase® II Enzyme mix as instructed in the Gateway® Technology cloning manual. The resulting DNA samples were subsequently treated with Proteinase K (Invitrogen) for 15 minutes at 37°C before transformation into chemically competent *E. Coli* One Shot® TOP10 cells (Invitrogen). Cells were plated onto LB agar plates containing 25 μg ml^{-1} kanamycin and transformants were screened for the presence of the insert using colony PCR and agarose gel electrophoresis. A single colony for each construct was picked for the inoculation of 10 ml of LB media supplemented with 25 μg ml^{-1} kanamycin. Plasmids were then extracted from the liquid cultures as outlined in the High-Speed Plasmid Mini Kit Manual (Geneaid Biotech). Entry clones were sent for DNA sequencing at the Mobix Facilities at McMaster University, to verify the integrity of the gene insert.

SNM1A expression constructs were generated by the LR recombination reaction as instructed in the Gateway® Technology cloning manual. Five different expression vector backbones were used: pDEST™-17, pDEST-566, pDEST-periHisMBP, pDEST-565, and pDEST-544. These encoded the His_6, His_6-MBP,

periplasmic His_6-MBP, His_6-GST, and His_6-NusA fusions respectively. The entry clones for each truncation were incubated with each of the five different expression vectors and LR Clonase® II Enzyme mix was added to each reaction. Samples were subsequently treated with Proteinase K (Invitrogen) for 15 minutes at 37°C before transformation into chemically competent *E. Coli* One Shot® TOP10 cells (Invitrogen). Cells were plated onto LB agar plates containing 0.1 mg ml^{-1} ampicillin and transformants were screened for the presence of the insert using colony PCR and agarose gel electrophoresis. A single colony for each construct was picked for the inoculation of 10 mL of LB media supplemented with 0.1 mg ml^{-1} ampicillin. Plasmids were then extracted from the liquid cultures as outlined in the High-Speed Plasmid Mini Kit Manual (Geneaid Biotech). Final expression constructs were verified using agarose gel electrophoresis.

2.2.3 SNM1A Expression Experiments

SNM1A constructs were tested for expression using different temperatures, cell-lines, and incubation times. Briefly, selected expression constructs were introduced by the heat shock method into the following chemically competent *E. Coli* cell-lines: BL21 (DE3) (Novagen), Rosetta™ 2 (DE3) pLysS (Novagen), Rosetta-gami™ 2 (DE3) pLysS (Novagen), OverExpress™ C41 (DE3) (Lucigen), and BL21 Star™ (DE3) pLysS pRARE (This was created by transforming the pRARE plasmid into the original BL21 Star™ (DE3) pLysS from Invitrogen). Cells were plated onto LB agar plates

containing 0.1 mg ml^{-1} ampicillin and several colonies were used for the inoculation of 10 mL of LB media supplemented with 0.1 mg ml^{-1} ampicillin. Bacteria were grown at 37°C with shaking (225 rev min^{-1}) until the absorbance at 600 nm (OD_{600}) reached 0.5. A 1 ml aliquot was withdrawn from each culture at this point for use as an uninduced sample. Protein expression was then induced using 1.0 mM IPTG at incubation temperatures of 37°C, 25°C, and 16°C, with 3 hour, 6 hour and 16 hour induction times respectively. Following this, a 400 µl aliquot was withdrawn from each culture for use as an induced sample. Collected cell samples were processed by resuspending in 2 x SDS-Dye, boiling for 5 minutes, and centrifuging at 16 000 x G. Protein content was assessed through SDS-PAGE analysis, while the presence of SNM1A was determined using an anti-His_6 Western Blot.

2.2.4 Small-Scale Purification Experiments

In order to screen the library of SNM1A constructs, a semi-high throughput, small-scale purification assay was used. Constructs were transformed by the heat shock method into BL21 Star™ (DE3) pLysS pRARE cells and plated onto LB agar plates containing 0.1 mg ml^{-1} ampicillin. These were used for the inoculation of 50 mL LB cultures supplemented with 0.1 mg ml^{-1} ampicillin. Bacteria were grown at 37°C with shaking (225 rev min^{-1}) until the absorbance at 600 nm (OD_{600}) reached 0.5. A 1 ml aliquot was withdrawn from each culture at this point for use as an uninduced sample. Protein expression was induced using 1.0 mM IPTG at an incubation temperature of 25°C. After a 16-hour induction

time, a 400 µl aliquot was withdrawn from each culture for use as an induced sample. The remaining cell cultures were harvested by centrifugation at 3 315 x G for 15 minutes at 4°C and subsequently flash frozen in liquid nitrogen and stored at -80°C.

During purification, cell pellets were treated with 5 ml of chemical lysis buffer (500 mM NaCl, 50 mM KH_2PO_4 pH 8.0, 10 mM imidazole, 10% (v/v) glycerol, 0.03% LDAO, 1 x Bugbuster® from Novagen) and incubated on ice for 30 minutes. Aprotinin, pepstatin A, benzamidine-HCL, and leupeptin protein inhibitors were added to a final concentration of 3.0 µM, 1 µM, 1 mM, and 1 µM respectively. Lysates were clarified by centrifugation at 12 000 x G for 15 minutes. Samples were applied to Ni-NTA spin columns from Qiagen by centrifugation at 890 x G for 2 minutes. Columns were then rinsed twice by centrifuging at 270 x G for 5 minutes with 600 µl of wash buffer (500 mM NaCl, 50 mM KH_2PO_4 pH 8.0, 30 mM imidazole). Bound protein was eluted by centrifugation at 890 x G for 2 minutes with 200 µl of elution buffer (500 mM NaCl, 50 mM KH_2PO_4 pH 8.0, 500 mM imidazole, 10% (v/v) glycerol). Collected fractions were analyzed by SDS-PAGE and Western blot.

2.2.5 Purification of SNM1A$_{698-1040}$

SNM1A$_{698-1040}$ was expressed and purified as a His_6-NusA fusion from *E. Coli* BL21 Star™ (DE3) pLysS pRARE cells. Bacteria were grown in standard LB medium supplemented with 0.1 mg ml^{-1} ampicillin at 37°C with shaking (225 rev min^{-1}) until the OD_{600} reached 0.8. Protein expression was induced using 1.0

mM IPTG at an incubation temperature of 25°C. After a 16-hour induction time, cells were harvested by centrifugation at 3 315 x G for 15 minutes at 4°C. Collected pellets were resuspended in PBS for transfer to storage tubes and subsequently flash frozen in liquid nitrogen and stored at -80°C. Prior to lysis using a French press, cell pellets (4L) were resuspended to a final volume of 180 mL using nickel A buffer (1 M KCl, 50 mM HEPES pH 7.5, 10 mM imidazole, 0.5 mM TCEP). Aprotinin, pepstatin A, benzamidine-HCl, leupeptin, and PMSF protease inhibitors were added to a final concentration of 3.0 μM, 1 μM, 1 mM, 1 μM, and 1 mM respectively. After lysis, samples were subjected to centrifugation at 48 384 x G for 60 minutes at 4°C and the supernatants sieved through a membrane filter (0.45 μm pore size). Filtered samples were then applied to a HiTrap nickel affinity column (GE Healthcare) that was set on a FPLC system (AKTA FPLC, Amersham Pharmacia Biotech). Bound proteins were eluted using 300 mM imidazole after washes with 10, 19.7, and 39 mM imidazole, which were accomplished by mixing appropriate ratios of nickel A and nickel B buffer (1 M KCl, 50 mM HEPES pH 7.5, 300 mM imidazole, 0.5 mM TCEP). Eluted protein was then diluted with QA buffer (20 mM HEPES pH 7.5 and 0.5 mM TCEP) to lower the KCl concentration to 150 mM. This sample was then applied to a Mono Q 10/100 GL anion exchange column (GE Healthcare) and subjected to an elution gradient of 150-600 mM KCl, at a flow rate of 0.2 ml min^{-1} over the course of 450 minutes. Collected fractions were analyzed by SDS-PAGE and fractions corresponding to a peak at approximately 350 mM KCl were pooled together.

$SNM1A_{698-1040}$ was then separated from the His_6-NusA fusion by overnight incubation at 4°C with TEV protease in a 1:15 (w/w) ratio. Digested samples were diluted with SA buffer (20 mM HEPES pH 7.5 and 0.5 mM TCEP) to lower the KCl concentration to 150 mM. This was then applied to a S-Sepharose HP cation exchange column (GE Healthcare) and subjected to an elution gradient of 150-1000 mM KCl at a flow rate of 1 ml min^{-1} over the course of 40 minutes. Final protein samples were concentrated using a Vivaspin ultrafiltration column with a 30 kDa cutoff (GE healthcare) and stored in 50% (v/v) glycerol at -20°C for future experiments. Protein purity was assessed through SDS-PAGE analysis, while concentrations were determined through the Bradford assay.

2.2.6 NusA-$SNM1A_{698-1040}$ Solubility Assay

An assay was developed to investigate the effects of buffer pH and charged amino acids on NusA-$SNM1A_{698-1040}$ solubility. Specifically, IMAC purified NusA-$SNM1A_{698-1040}$ (1 M NaCl, 50 mM HEPES pH 7.5, 300 mM imidazole, 0.5 mM TCEP) was mixed with various buffer solutions from the JBS solubility kit (Jena Bioscience) to see if they resulted in aggregation. For these experiments, the final protein concentration was 0.5 mg ml^{-1}, while the concentrations of NaCl and buffering agent were kept constant at 250 mM and 50 mM respectively. A second set of experiments was also performed in the presence of 50 mM arginine and 50 mM glutamic acid. After mixing, protein samples were incubated overnight at room temperature. These were then transferred to a clear-bottom 96-well plate, and the absorbances at 340 nm were

measured using a BioTek Synergy 4 Hybrid Microplate Reader. A series of buffer only controls, with and without arginine and glutamic acid, were also measured at 340 nm. These values were then subtracted from the previously obtained results to correct for any buffer absorbance.

2.3 Results

2.3.1 Purification of $SNM1A_{608-140}$

$SNM1A_{608-1040}$ was purified by nickel affinity chromatography from 4L of bacterial cell culture. This protein was noted to bind poorly to the nickel column, and eluted gradually over a gradient of 48-67 mM imidazole. As shown in figure 2.4, the obtained $SNM1A_{608-1040}$ has a molecular weight of 51 kDa, and is approximately 5% pure by SDS-PAGE. Protein identity was confirmed using an anti-His_6 Western blot. This was further confirmed, when a TEV digestion was performed, as the 51 kDa band disappears on SDS-PAGE and Western blot, to be replaced with another band at 48 kDa. Protein yield was low, with a total of 0.20 mg of $SNM1A_{608-1040}$ obtained (4 mg of total protein). A subsequent purification of this sample using ion exchange chromatography and a second nickel affinity column was unsuccessful, and did not significantly increase purity. Furthermore, additional exonuclease assays indicated $SNM1A_{608-1040}$ was inactive.

2.3.2 SNM1A Expression requires a pLysS Cell Line

The optimal expression conditions for SNM1A were determined by growing sets of cultures in different cell lines, and inducing protein production

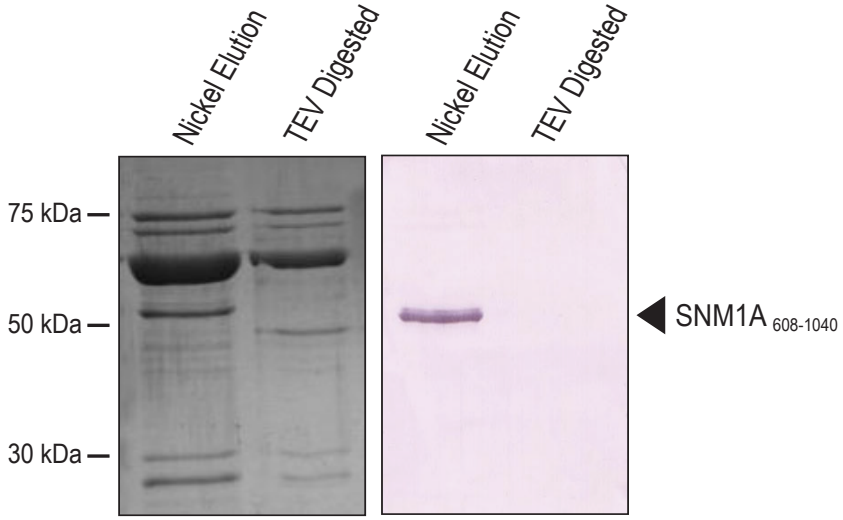

Figure 2.4: Purification of SNM1A$_{608-1040}$. SNM1A$_{608-1040}$ eluted gradually over a gradient of 48-67 mM imidazole. These fractions were pooled and subsequently TEV digested. Samples were run on a 12% SDS-PAGE gel and visualized using Blue-Silver stain, while an anti-His$_6$ Western blot was used to confirm protein identity.

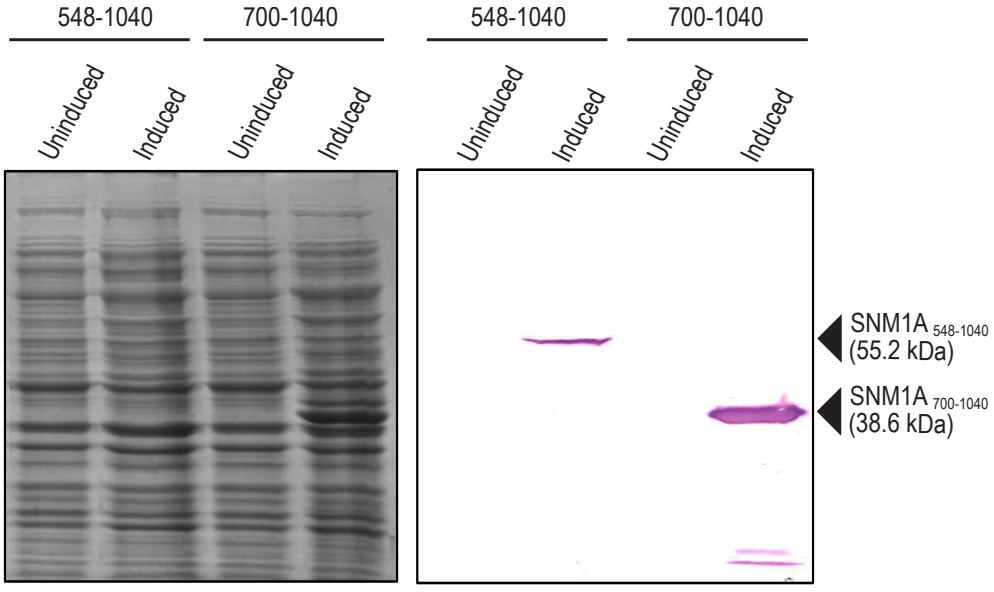

Figure 2.5: Shorter truncations of SNM1A express at higher levels. Uninduced and induced cell samples were run on 12% SDS-PAGE gels and visualized using Blue-Silver stain, while an anti-His$_6$ Western blot was used to identify SNM1A truncations.

under different temperatures and incubation times. SDS-PAGE and Western Blot analyses were subsequently performed on the collected uninduced and induced cell samples. SNM1A constructs were found to require *E. Coli* cell lines containing the pLysS plasmid for expression. Additionally, it was discovered that maximum protein production occurred when cultures were induced at 25°C for 16 hours. Another observation was that shorter truncations of SNM1A expressed at higher levels compared to their longer counterparts. This can be seen in figure 2.5, where $SNM1A_{700\text{-}1040}$ is expressed approximately five-fold compared to $SNM1A_{548\text{-}1040}$.

2.3.3 NusA-SNM1A Fusions have Higher Levels of Expression

A total of seventy SNM1A constructs were screened using a small-scale purification assay. This was done to identify constructs that could produce soluble protein and could furthermore bind effectively to a nickel affinity column. From these experiments, a large majority of SNM1A constructs produced insoluble protein or suffered from expression issues. Specifically, it was found that the His_6-MBP and periplasmic His_6-MBP constructs expressed poorly, with no proteins corresponding to the appropriate molecular weight being found. By contrast, the His_6, His_6-GST and His_6-NusA constructs were readily expressed, but were found to be insoluble under the given conditions. Despite this, it was observed that the His_6-NusA constructs were expressed at relatively high levels when compared to other constructs. This was especially true when the shorter truncations of SNM1A were conjugated to the NusA fusion.

2.3.4 Purification of NusA-SNM1A$_{698-1040}$ Produces Soluble Protein

NusA-SNM1A$_{698-1040}$ was purified from 4L of bacterial cell culture by nickel affinity chromatography. As shown in figure 2.6, the protein was eluted off the column at an imidazole concentration of 300 mM, with a purity of approximately 30%. This sample was further purified using a Mono Q anion exchange column, using a salt gradient from 150-600 mM KCl (figure 2.7). The observed elution profile of NusA-SNM1A$_{698-1040}$ was complex, with multiple fractions containing the protein. However, the peak corresponding to 350 mM KCl was of highest purity (85%) and therefore collected for subsequent purification. A TEV digestion was performed and the resulting sample was applied to a S-Sepharose cation exchange column (figure 2.8). A salt gradient from 150-1000 mM KCl was then performed and SNM1A$_{698-1040}$ was found to elute in two peaks at 700 mM and 750 mM KCl respectively. Interestingly, despite having a molecular weight of 38.8 kDa, the band corresponding to SNM1A$_{698-1040}$ appears to be smaller on a SDS-PAGE gel, with a molecular weight closer to 35 kDa. Final protein was 95% pure by SDS-PAGE, with a total yield of 0.25 mg from 4L of cell pellet. For additional details, a protein purification table can be seen in table 2.1.

2.3.5 NusA-SNM1A$_{698-1040}$ is Soluble at Higher pH

The solubility of NusA-SNM1A$_{698-1040}$ under varying buffer pH conditions was assessed using turbidimetry, where protein aggregation was correlated with

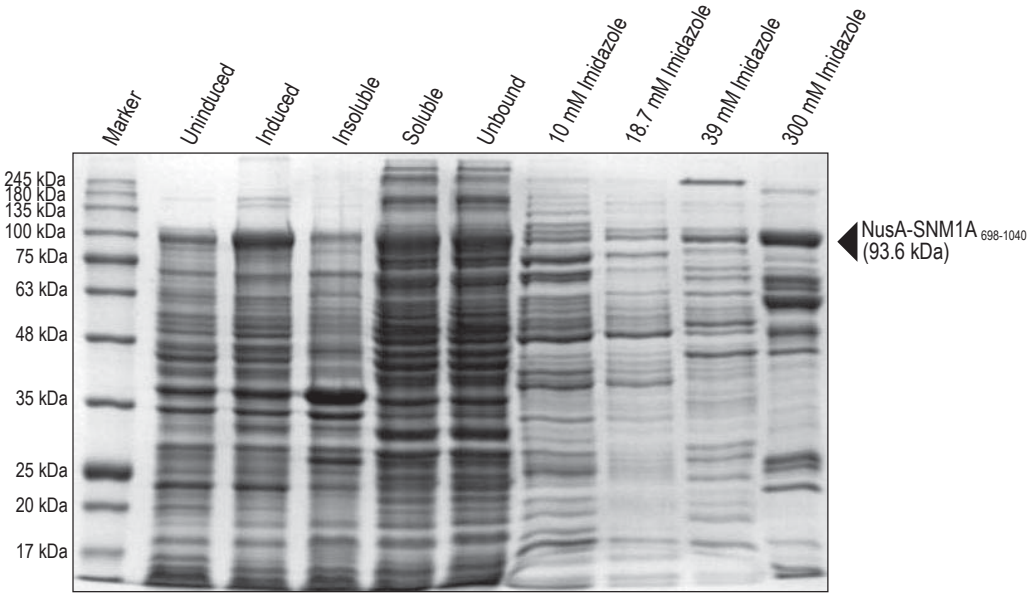

Figure 2.6: IMAC purification of NusA-SNM1A$_{698-1040}$. Fractions were run on a 12% SDS-PAGE gel and visualized using Blue-Silver stain.

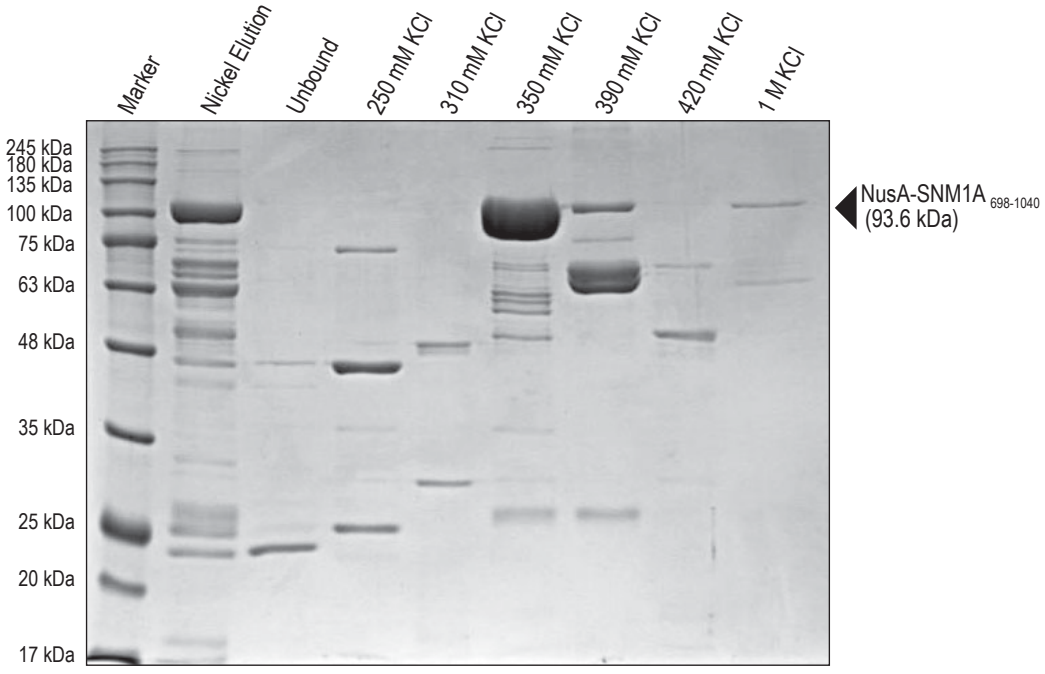

Figure 2.7: Mono Q purification of NusA-SNM1A$_{698-1040}$. IMAC purified protein was loaded on a Mono Q column and subjected to a 150-600 mM KCl gradient, at a flow rate of 0.2 ml min^{-1} over the course of 450 minutes. Fractions were run on a 12% SDS-PAGE gel and visualized using Blue-Silver stain.

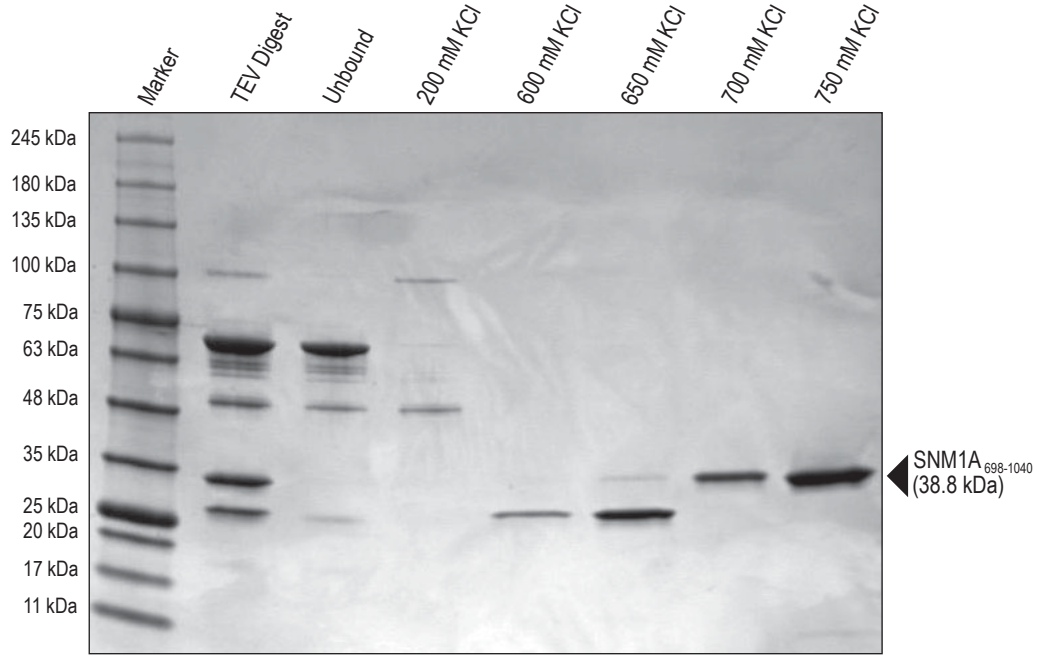

Figure 2.8: S-Sepharose IEC purification of SNM1A$_{698-1040}$. MonoQ purified NusA-SNM1A$_{698-1040}$ was TEV digested overnight at 4°C and loaded on a S-Sepharose column. This was then subjected to a 150-1000 mM KCl gradient at a flow rate of 1 ml min^{-1} over the course of 40 minutes. Fractions were run on a 12% SDS-PAGE gel and visualized using Blue-Silver stain.

Table 2.1: Purification table for SNM1A$_{698-1040}$

Step	Total Protein (mg)[b]	Target Protein (mg)[c]	Yield (%)	Purity (%)[d]
Crude Lysate[a]	1600	80 (34)	100	5
Crude Extract	1200	12 (5.1)	15	1
Nickel Column	11	3.3 (1.4)	4.1	30
Mono Q Column	3.1	2.6 (1.1)	3.3	85
TEV Digestion	4.0	(0.6)	1.8	15
S-Sepharose Column	0.26	(0.25)	0.74	95

[a] From 21.4 g of wet weight E. Coli cell pellet (from 4L of bacterial culture)
[b] Protein concentrations determined by Bradford assay using BSA as a standard
[c] The mg of SNM1A$_{698-1040}$ is shown in parentheses
[d] As estimated by SDS-PAGE

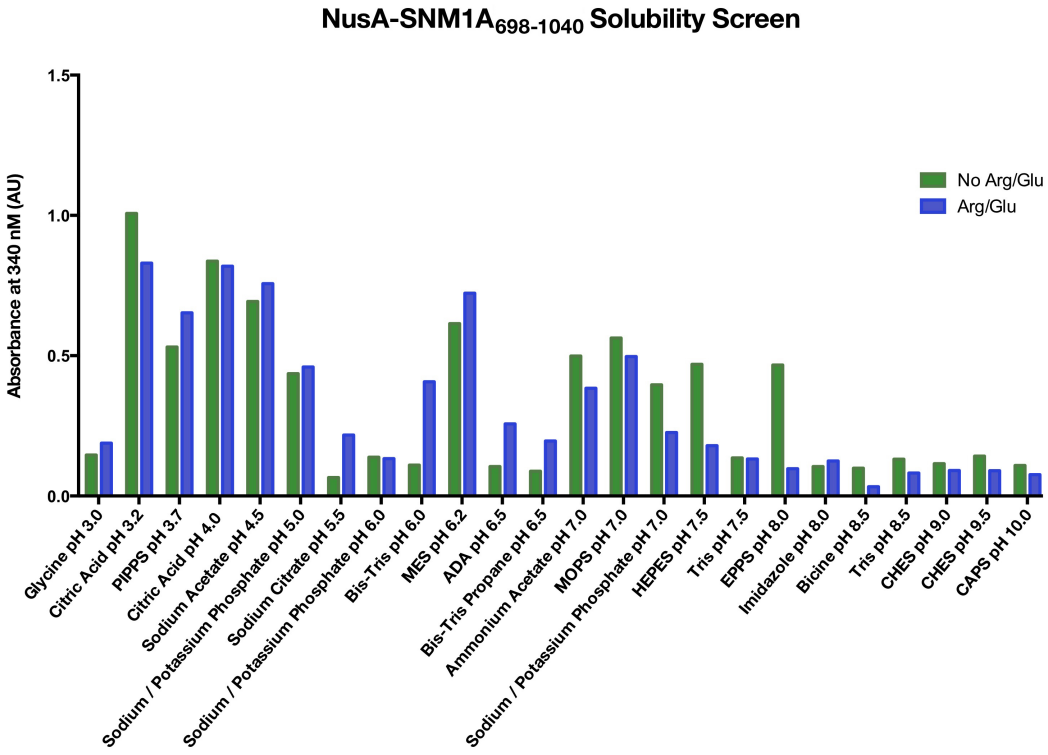

Figure 2.9: Solubility assay for NusA-SNM1A$_{698-1040}$. IMAC purified NusA-SNM1A$_{698-1040}$ was screened against different buffer solutions from the JBS solubility kit. In these assays, final protein, buffer agent, and NaCl concentrations were kept constant at 0.5 mg/ml, 50 mM, and 250 mM respectively. Samples were incubated overnight at room temperature and the absorbances were measured at 340 nm to assess for protein aggregation.

higher absorbance at 340 nm. Samples with an absorbance above and below the arbitrary threshold of 0.250 were considered to be soluble or insoluble respectively. Under constant salt and protein concentrations of 250 mM NaCl and 0.5 mg ml^{-1} respectively, NusA-$SNM1A^{698-1040}$ appeared to be more soluble at a buffer pH greater than 8.0 (figure 2.9). Several other conditions also had decreased absorbance at 340 nm, specifically: Glycine pH 3.0, Sodium Citrate pH 5.5, Sodium/Potassium Phosphate pH 5.5, Bis-Tris pH 6.0, ADA pH 6.5, Bis-Tris Propane pH 6.5 and Tris pH 7.5. Furthermore, the addition of 50 mM arginine and 50 mM glutamic acid did not appear to significantly improve solubility in most conditions tested. The sole exceptions being HEPES pH 7.5 and EPPS pH 8.0, where the addition of arginine and glutamic acid reduces absorbance values by 61.8% and 79.2% respectively.

2.4 Discussion

In order for biochemical and crystallography experiments to be performed on human SNM1A, it is necessary to obtain soluble, active protein of high purity. This is difficult to accomplish, as previous studies have shown that SNM1A is toxic upon overexpression in human cell lines, and highly insoluble in bacterial expression systems (Hejna *et al.*, 2007). To date full-length SNM1A has been partially purified from both yeast and insect cell lines, albeit with low yield and purity (Hejna *et al.*, 2007; Hazrati *et al.*, 2008). Recently, a N-terminal truncation of human SNM1A, spanning residues 608-1040, was purified from *E. Coli* (Wang *et al*, 2011). This truncation contained both the MBL and β-CASP domains and

was noted to possess 5'exonuclease activity. These results established that purifying a truncation of SNM1A from *E. Coli* is feasible. This is particularly advantageous, as bacterial expression systems are robust, easy to use, and low in cost. With this in mind, we attempted to replicate the purification of $SNM1A_{608-1040}$ from *E. Coli*. However, this was difficult in practice, as the authors did not include details about their purification within their paper (Wang *et al*, 2011).

$SNM1A_{608-1040}$ was purified from 4L of cell culture with a yield of 0.20 mg and a purity of 5% (figure 2.4). This truncation was noted to bind poorly to a nickel affinity column, and eluted over a gradient of 48-67 mM imidazole. Since this sample was of low purity, an effort was made to further enrich $SNM1A_{608-1040}$ using additional purification steps. Specifically, a TEV digestion was performed on this protein to remove the His_6-tag, and a second nickel affinity column was used to remove the tag and any other co-eluting contaminants. A S-Sepharose cation exchange column was also used, since at the buffer pH of 7.5, $SNM1A_{608-1040}$ is positively charged (pI of 9.01). However, both of these methods were unsuccessful, as they were unable to separate $SNM1A_{608-1040}$ from the large contaminating species at 63 kDa. Furthermore, $SNM1A_{608-1040}$ appeared to be inactive based on nuclease assays performed using a 5'PO_4 DNA substrate. Based on these results, we concluded that $SNM1A_{608-1040}$ might be at least partially unfolded and/or bound to a chaperone protein (63 kDa contaminant). Consequently, it was decided that this construct would be unsuitable for future experiments.

Due to the difficulties encountered with $SNM1A_{608-1040}$, we decided to re-approach the problem by identifying an optimal construct for purification. Studies have shown that cloning and screening multiple expression constructs early in the project provides a higher chance of obtaining soluble protein (Graslund *et al*, 2008). This is in contrast to more traditional methods, which focus on optimizing the purification of a single construct. With this in mind, we decided to create a library of seventy SNM1A expression constructs in order to maximize the chances of obtaining soluble protein. This was accomplished by cloning seven truncations of SNM1A, with and without a C-terminal His_6-tag, into five different expression vectors (section 2.1.1). Once this was done, experiments were performed to identify a set of standardized expression conditions for testing constructs. Similar to $SNM1A_{608-1040}$, all the constructs tested were found to require an *E. Coli* cell line with the pLysS plasmid, which suppresses basal expression of the recombinant protein (Studier, 1991). This is necessary for the expression of toxic proteins, and is consistent with SNM1A's profile as a nuclease, which is likely deleterious upon expression in a bacterial cell. Additionally, it was also determined that an induction at 25°C for 12-16 hours was optimal for maximal protein production.

Following this, each construct was expressed from *E. Coli* BL21 Star™ (DE3) pRARE pLysS cells. This cell line was chosen as it contains a mutation in RNaseE (*rne*131), which improves the stability of mRNA transcripts (Lopez *et al*, 1999). While the inclusion of the pRARE plasmid encodes tRNAs for codons

rarely found in *E. Coli* (Novy *et al*, 2001). Constructs were then purified using a small-scale assay and evaluated based on their expression, solubility and nickel column binding. In general, most SNM1A constructs were found to be insoluble (section 2.3.3). The exceptions to this were the His_6-MBP and periplasmic His_6-MBP constructs, which did not express. Interestingly, an anti-His_6 Western blot of induced cell samples from these constructs revealed multiple low molecular weight species. This suggests that the lack of expression is due to degradation or incomplete translation of mRNA transcripts, though it is unclear as to why this would be the case. The His_6, His_6-GST and His_6-NusA constructs, on the other hand, were readily expressed but were insoluble under the given conditions.

While none of the SNM1A constructs produced soluble protein, two important observations were made during these experiments. The first was that shorter truncations of SNM1A were produced at higher levels when compared to their longer counterparts (figure 2.5). The second observation was that NusA-SNM1A constructs were expressed in higher amounts compared to other fusion proteins. Furthermore, these two effects appeared to be synergistic with one another. Taking this information into consideration, large-scale purifications were performed on several of the shorter NusA-SNM1A constructs. It was reasoned that for a highly expressed construct, some fraction of the protein would be soluble. Scaling up the purification would therefore compensate for this partial solubility and result in an acceptable yield of SNM1A.

From the constructs tested in large-scale purifications, it was found that $SNM1A_{698-1040}$ was partially soluble when expressed as a His_6-NusA tagged protein. NusA-$SNM1A_{698-1040}$ was successfully purified from 4L of bacterial cell culture by nickel affinity chromatography (figure 2.6). As seen in table 2.2, approximately 15% of NusA-$SNM1A_{698-1040}$ was recovered in the soluble fraction, with the rest of the protein presumably insoluble. Additionally, NusA-$SNM1A_{698-1040}$ was found to bind to the nickel column with low affinity, as the yield of protein decreased by more than three-fold after this step. Following this, a Mono Q anion exchange column was used to achieve further purity (figure 2.7). This was done, since at a buffer pH of 7.5, NusA and $SNM1A_{698-1040}$ are negatively and positively charged respectively (pIs of 4.44 and 8.76). This step was found to improve the purity of the sample from 30% to 85%, with a 21% reduction in yield. Lastly, the His_6-NusA tag was removed by TEV digestion and $SNM1A_{698-1040}$ was isolated using a S-Sepharose cation exchange column (figure 2.8). The final yield of $SNM1A_{698-1040}$ from this purification was 0.25 mg with a purity of 95%.

Although we have been successful in obtaining soluble and pure SNM1A, the low yield of protein is not optimal for structural studies. We are currently optimizing the purification of $SNM1A_{698-1040}$ to address this issue. To date three parameters have been found to affect the behavior of this protein: the OD_{600} for induction, buffer salt concentration and purification temperature. Like most proteins, it has been found that IPTG induction during exponential phase

increases the yield of NusA-SNM1A$_{698-1040}$ by a factor of three to five-fold. However, this has the caveat of producing a highly viscous sample that readily precipitates. It is possible that the insoluble SNM1A$_{698-1040}$ truncation causes aggregation at higher concentrations, despite being conjugated to a NusA fusion. Dilution of this sample was also been unable to prevent precipitation, as aggregation was found to occur at protein concentrations of less than 0.5 mg ml^{-1}. Furthermore, this effect is exacerbated when NusA-SNM1A$_{698-1040}$ is purified at low temperatures or low buffer salt concentrations. As a result we have been unable to purify NusA-SNM1A$_{698-1040}$ in higher quantities, since all attempts to do so resulted in precipitation of the sample.

Since NusA-SNM1A$_{698-1040}$ precipitates at high concentrations with low salt, it is necessary to find an alternate method of solubilizing this protein. To facilitate this, the crystal structure of SNM1A$_{696-1040}$ was examined to see if any insight could be made regarding its insolubility. From the structure, it was apparent that SNM1A$_{696-1040}$ is polarized, with both positively and negatively charged surfaces (figure 2.10). While these charged surfaces might be used to interact with DNA, it is possible that the oppositely charged faces cause SNM1A molecules to interact electrostatically and therefore aggregate together. By changing the pH of the buffer, we hypothesized that the charge on these surfaces might be altered through the deprotonation or protonation of exposed side chains. In addition to this, we also considered treating SNM1A$_{698-1040}$ with arginine and glutamic acid, since previous studies have shown that the addition of these amino

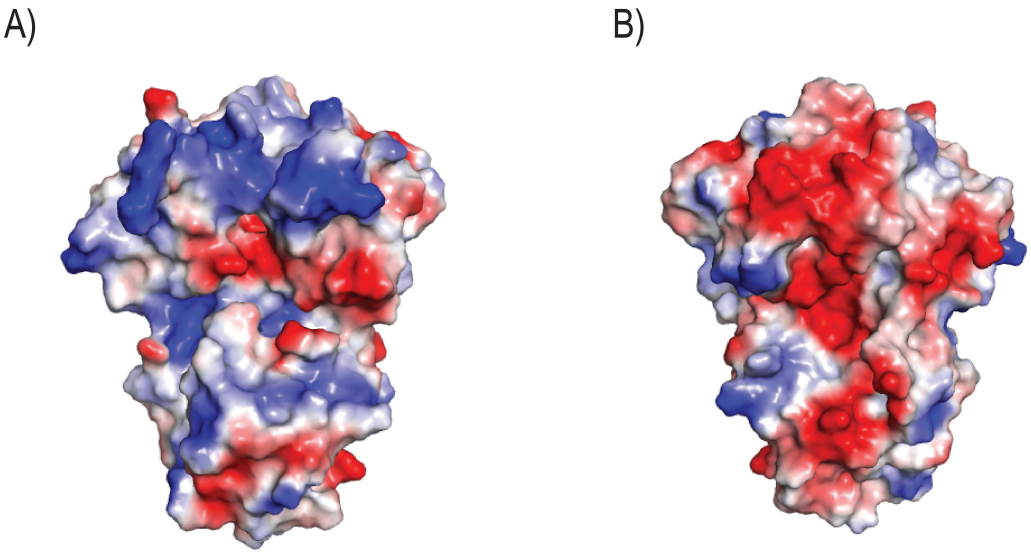

Figure 2.10: Surface representation of SNM1A$_{698-1040}$. 'Top' (A) and 'bottom' (B) faces of SNM1A$_{698-1040}$. Positive and negative electrostatic charge potentials are shown in blue and red respectively (PDB ID 4B87).

acids can greatly increase solubility (Golovanov *et al*, 2004).

The solubility of NusA-SNM1A$_{698-1040}$ with varying pH was tested using a turbidimetric assay. Protein precipitation was correlated with increased absorbance at 340 nm, since aggregates strongly scatter light at this wavelength (Stranks *et al*, 2009). From this data, NusA-SNM1A$_{698-1040}$ was found to be more soluble at a buffer pH greater than 8 (figure 2.9). This is consistent with our model, as a pH higher than the isoelectric points of NusA and SNM1A$_{698-1040}$ would cause the entire protein to be more negatively charged, therefore preventing aggregation. Interestingly, the addition of arginine and glutamic acid did not appear to improve protein solubility in the majority of conditions tested. This is surprising, as it was expected that the addition of these charged amino acids would enhance solubility by shielding the surface charges of NusA-SNM1A$_{698-1040}$. Instead, the effect of these amino acids appeared to vary for each condition. Most notably, the addition of arginine and glutamic acid was found to solubilize protein in the HEPES pH 7.5 condition, which is also the same buffer used in our purifications.

Although the purification of SNM1A$_{698-1040}$ from *E. Coli* is not ideal for high yield applications such as crystallography, it is a significant improvement compared to the recently reported purification of SNM1A$_{608-1040}$. Briefly, in their protocol, SNM1A$_{608-1040}$ is expressed as a His$_6$-thioredoxin tagged protein from 24 L of cell culture (Sengerova *et al*, 2012). This is purified stepwise using a phosphocellulose P11 resin, a HisTrap column and a S75 gel filtration column.

Following this, $SNM1A_{608-1040}$ is TEV digested and further purified using a Heparin column and a S-Sepharose cation exchanger. Seeing as how our purification requires 4 L of cell culture, and half the number of steps, it is evident that we have successfully developed a more efficient method of purifying SNM1A from *E. Coli.*

CHAPTER 3 – CHARACTERIZATION OF SNM1A

3.1 Rationale and Experimental Design

The main goal of this project was to structurally and functionally characterize human $SNM1A_{698-1040}$. Since this was recently purified from *E. Coli*, it was deemed necessary to verify its activity, prior to any additional experiments. To this end, biochemical assays were performed on $SNM1A_{698-1040}$ to establish whether it possessed 5'-phosphate-dependent exonuclease activity. In addition to this, the D736A and D736A/H737A point mutants were generated by site-directed mutagenesis, and subsequently purified to act as controls. This would ensure that any observed nuclease activity would be attributable to $SNM1A_{698-1040}$ and not any contaminants. Once the appropriate controls were done, we proceeded to characterize the nuclease activity of $SNM1A_{698-1040}$ on ssDNA and dsDNA substrates. Additionally, electrophoretic mobility shift assays (EMSAs) were performed using the D736A/H737A mutant, to see if a 5'phosphate was required for DNA binding.

$SNM1A_{698-1040}$ was structurally characterized using mass spectrometry. Specifically, a trypsinization and liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis was performed to confirm the identity of the protein, while electrospray ionization time of flight mass spectrometry (ESI-TOF-MS) was used to determine molecular mass. Furthermore, Edman sequencing was used to identify the first six residues at the N-terminus. These experiments were performed to confirm the identity of $SNM1A_{698-1040}$, and to investigate the anomalous molecular mass of this truncation as seen by SDS-PAGE. Lastly, we

have screened for crystallization conditions for $SNM1A_{698-1040}$, in an attempt to obtain a crystal structure of this protein.

3.2 Materials and Methods

3.2.1 Site-directed Mutagenesis for $SNM1A_{698-1040}$

The D736A and D736A/H737A mutants were generated by designing the appropriate primers to introduce point mutations within the $SNM1A_{698-1040}$ gene. Primers were designed using the PrimerX web tool and mutagenesis was performed on the $SNM1A_{698-1040}$ entry clone (MJ4829) using the iNtRON Biotechnology *i*-Pfu DNA polymerase. Reactions were prepared as instructed in the manufacturer's manual and PCR was performed as follows: initial denaturation (95°C for 5 minutes), followed by 25 cycles of denaturation (95°C for 30 seconds), annealing (60°C for 1 minute), and elongation (68° for 10 minutes), and a final elongation step (72°C for 20 minutes). The resulting DNA samples were treated with DpnI from ThermoScientific for 1 hour at 37°C before transformation into chemically competent *E. Coli* One Shot® TOP10 cells (Invitrogen). Transformants were selected for using kanamycin, and plasmids were propagated and extracted using a High-Speed Plasmid Mini Kit Manual (Geneaid Biotech). Entry clones were then sent for DNA sequencing at the Mobix Facilities at McMaster University, to verify that the appropriate point mutations had been introduced.

The D736A and H737A expression vectors were subsequently generated using the LR recombination reaction as instructed in the Gateway® Technology

cloning manual. Both entry clones were incubated with the pDEST-544 plasmid encoding the His_6-NusA fusion, along with the LR Clonase® II Enzyme mix. Samples were subsequently treated with Proteinase K (Invitrogen) for 15 minutes at 37°C before transformation into chemically competent *E. Coli* One Shot® TOP10 cells (Invitrogen). Transformants were selected for using ampicillin, and plasmids were propagated and extracted using a High-Speed Plasmid Mini Kit Manual (Geneaid Biotech). Final expression constructs were verified using agarose gel electrophoresis. The resulting D736A and D736A/H737A mutants were expressed and purified in an identical procedure to $SNM1A_{698-1040}$ as described in section 2.2.5.

3.2.2 Preparation of DNA Substrates

DNA oligonucleotides used in nuclease assays and EMSAs were labeled at the 3' end with a 6-fluoresceinamidite (6-FAM) label (Biobasic). Two fluorescently labeled polythymidine 20-mers were synthesized with a 5'PO_4 and a 5'OH respectively for use in EMSAs. While another two fluorescently labeled ssDNA 20-mers with the sequence 5'-TACCCGCCCGGCCCGCCCAT-3' were synthesized with a 5'PO_4 and a 5'OH for nuclease assays. Prior to performing experiments, oligonucleotides were purified by gel purification. This was done by loading the oligonucleotides onto a 20% denaturing polyacrylamide gel (containing 7 M urea) and performing electrophoresis at 900 V for 5 hours using a vertical gel system from CBS Scientific. The gels were visualized by fluorescence using a Typhoon Imager (GE Healthcare) with excitation and

emission wavelengths of 485 nm and 526 nm respectively. Following this, the appropriate gel bands were excised and the oligonucleotides eluted by an overnight incubation of the gel pieces in 200 mM NaCl, 10 mM Tris pH 7.5, 1 mM EDTA at 37°C with shaking (225 rev min^{-1}). Oligonucleotides were subsequently ethanol precipitated and resuspended in double-deionized water (ddH_2O), and DNA was quantified by measuring the absorbance at 260 nm.

A dsDNA substrate was prepared by annealing the phosphorylated 5'-TACCCGCCCGGCCCGCCCAT-3' oligonucleotide with its unlabeled and unphosphorylated complement 5'ATGGGCGGGCCGGGCGGGTA-3' (IDT). This was done by mixing the oligonucleotides, with a slight excess of the unlabeled complement (1:1.2 molar ratio), and performing the annealing reaction using a thermocycler. This mixture was heated to 95°C for two minutes, and then reduced to a temperature of 25°C at a rate of 1°C min^{-1}. Following this, the ssDNA and dsDNA substrates were run on a 20% native polyacrylamide gel and visualized by fluorescence, to verify for annealing. A control consisting of dsDNA incubated at 37°C for 1 hour was also performed, to ensure this substrate remained annealed at higher temperatures.

3.2.3 Nuclease Assays

$SNM1A_{698-1040}$ was tested for 5'-phosphate-dependent exonuclease activity by performing a time course assay with 5'PO4 and 5'OH ssDNA oligonucleotides (figure 3.1). These reactions were carried out in 90 µl of 1 x Buffer F (50 mM Tris-acetate pH 7.2, 10 mM Mg acetate, 75 mM potassium

acetate, 1 mM DTT) supplemented with 100 μg ml^{-1} BSA and incubated at 37°C. The final concentrations of $SNM1A_{698-1040}$ and DNA were 33.33 nM and 66.66 nM respectively. 15 μl aliquots were withdrawn and quenched with 2 x denaturing stop dye (50% glycerol, 0.25% xylene cyanol, 0.25% bromophenol blue, 7 M urea, 50 mM EDTA) at the following time points: 0 s, 30 s, 2 min, 6 min and 14 min. Reaction products were subsequently resolved on a 20% denaturing polyacrylamide gel running at 900 V for 4 hours. Gels were visualized by fluorescence using a Typhoon Imager (GE Healthcare) with excitation and emission wavelengths of 485 nm and 526 nm respectively.

An additional gel consisting of nuclease assay controls was also run. Specifically, 5'PO4 and 5'OH ssDNA oligonucleotides were incubated with $SNM1A_{698-1040}$ and the D736A and D736A/H737A mutants. Reactions were performed in 15 μl of 1 x Buffer F supplemented with 100 μg ml^{-1} BSA and incubated at 37°C for 30 minutes. Final concentrations of enzyme and DNA were 33.33 nM and 66.66 nM respectively. All reactions were quenched by adding 15 ul of 2 x denaturing stop dye. Samples were loaded on 20% denaturing polyacrylamide gels and resolved by electrophoresis at 900 V for 4 hours. Gels were subsequently visualized by fluorescence.

$SNM1A_{698-1040}$ time course assays were performed using ssDNA and dsDNA substrates. Reactions were run in 160 μl volumes of 1 x Buffer F supplemented with 100 μg ml^{-1} BSA, and incubated at 37°C. Concentrations of SNM1A and DNA were held at 8.33 nM and 133.32 nM respectively. 15 μl

aliquots were withdrawn and quenched with 2 x denaturing stop dye at the following time points: 0 s, 30 s, 1 min, 2 min, 4 min, 8 min, 16 min, 32 min and 64 min. These samples were loaded on 25% denaturing polyacrylamide gels and resolved by electrophoresis at 900 V for 6 hours. Gels were subsequently visualized by fluorescence.

3.2.4 DNA EMSAs

DNA EMSAs were performed with 5'PO_4 and 5'OH polythymidine 20-mers. Purified D736A/H737A in storage buffer (20 mM HEPES pH 7.5, 750 mM KCl, 0.5 mM TCEP pH 7.5, 50% (v/v) glycerol) was buffer exchanged into 2 x EMSA buffer (40 mM HEPES pH 7.5, 300 mM KCl, 2 mM dithiothreitol (DTT), 30% (v/v) glycerol). This was done by repeatedly concentrating in a 30 kDa cutoff Nanosep® column (Pall Corporation), and diluting with 2 x buffer. EMSA reactions were then performed by adding 15 µl of $SNM1A_{698-1040}$ of varying concentrations in 2 x EMSA buffer, to 15 µl of DNA substrate in ddH_2O. DNA substrates were kept constant at concentration of 15 nM, while the concentrations of D736A/H737A were 0 nM, 75 nM, 375 nM and 1875 nM respectively. After a 30-minute incubation time, samples were run on a 10% native PAGE gel at 100 V for 30 minutes. This gel was subsequently visualized using by fluorescence.

3.2.5 Peptide Mass Fingerprinting

$SNM1A_{698-1040}$ was purified as described in section 2.2.5 for identification by peptide mass fingerprinting (PMF). After purification, $SNM1A_{698-1040}$ was

concentrated to 10 uM using a Vivaspin ultrafiltration column with a 30 kDa cutoff (GE healthcare). $SNM1A_{698-1040}$ was then extensively dialyzed into a low salt buffer, consisting of 20 mM HEPES pH 7.5, 300 mM KCl and 1.0 mM TCEP. This was done using a 10 kDa cutoff, Slide-A-Lyzer MINI Dialysis Unit (Thermo Scientific). Prepared $SNM1A_{698-1040}$ samples were sent to Li Zhang at the Advanced Protein Technology Centre (APTC) at the University of Toronto for analysis.

PMF analysis was performed using LC-MS/MS after an in solution trypsin digestion. $SNM1A_{698-1040}$ was evaporated and resuspended in 100 µl of 6 M urea for denaturation. Disulfide bonds were reduced using 10 mM DTT and incubating for 60 minutes at room temperature. Cysteine residues were then alkylated using 100 mM iodoacetamide and incubating for 15 minutes in the dark at room temperature. DTT was subsequently added to a concentration of 40 mM to reduce any unreacted iodoacetamide. Following this, the sample was diluted in water to reduce the concentration of urea to 0.6 M. Digestion was performed by adding 20 µg of trypsin (Promega) and incubating overnight at 37°C. This reaction was quenched by the addition of 5% formic acid to lower the pH. The resulting sample was evaporated and reconstituted in 5 µl of 0.1% formic acid.

Following this, the digested peptides were loaded onto a 150 µm ID pre-column (Magic C18, Michrom Biosciences) at 4 µl/min and separated over a 75 µm ID analytical column packed into an emitter tip containing the same packing material. The peptides were eluted over 60 min at 300 nl/min. using a 0 to 40%

acetonitrile gradient in 0.1% formic acid using an EASY n-LC nano-chromatography pump (Proxeon Biosystems, Odense Denmark). The peptides were eluted into an LTQ-Orbitrap hybrid mass spectrometer (Thermo-Fisher, Bremen, Germany) operated in a data dependent mode. MS was acquired at 60,000 full width at half maximum (FWHM) resolution in the Fourier transform mass spectrometry (FTMS) and MS/MS was carried out in the linear ion trap. 6 MS/MS scans were obtained per MS cycle. The raw data was searched using Mascot 2.3.02 (Matrix Sciences, London UK). The search result was analyzed using Scaffold 3.4.3 (Proteome Software Inc, Portland, US).

3.2.6 Mass Determination by ESI-TOF-MS

$SNM1A_{698-1040}$ was prepared as described in section 3.2.5 and sent to Li Zhang at the APTC for ESI-TOF-MS mass determination. This sample was desalted prior to ESI-TOF-MS as fol1ows. 1 µl of 1% trifluoroacetic acid was added to a 10 µl aliquot of $SNM1A_{698-1040}$ and mixed. Sample was then bound to a $ZipTip_{C18}$™ (Millipore) chromatography pipette tip pre-equilibrated with 0.1% trifluoroacetic acid in water. After washing with 10 µl of 0.1% trifluoroacetic acid, $SNM1A_{698-1040}$ was eluted in 5 µl of 50% acetonitrile in water. Eluate was directly injected onto a QSTAR XL Pulsar quadrupole time-of-flight (QTOF) mass spectrometer (Applied Biosystems/MDS Sciex, Foster City, CA). MS was acquired in the positive ion TOF/MS mode from *m/z* 1200-2500. These were subsequently deconvoluted using the Bayesian protein reconstruct tool

(Bioanalyst QS 1.1 software package) to determine the zero charge mass of the intact protein.

3.2.7 Edman Sequencing

$SNM1A_{698-1040}$ was prepared as described in section 3.2.5 and sent to Reynaldo Interior at the APTC for Edman sequencing. This sample was spotted onto a methanol activated polyvinylidene difluoride (PVDF) membrane (Millipore) and washed with ddH_2O to remove salt and other buffer components. $SNM1A_{698-1040}$ was then directly sequenced from the membrane, using a Procise cLC 492 protein sequencer (Applied Biosystems) employing automated Edman degradation chemistry.

3.2.8 Crystallization Experiments

$SNM1A_{698-1040}$ and NusA-$SNM1A_{698-1040}$ were screened for crystallization conditions through hanging drop vapor diffusion. Samples containing NusA-$SNM1A_{698-1040}$ were subjected to *in situ* proteolysis by adding TEV protease in a 1:15 (w/w) ratio prior to setting drops. The conditions tested were from the MCSG I (Microlytic), MCSG II (Microlytic), MCSG III (Microlytic), MCSG IV (Microlytic), Nextal Classics (Qiagen), Nextal Cryos (Qiagen) and Wizard (Emerald BioSystems) broad screens. Crystallization buffers for $SNM1A_{698-1040}$ and NusA-$SNM1A_{698-1040}$ were as follows: 750 mM KCl, 20 mM HEPES pH 7.5 and 0.5 mM TCEP pH 7.5; and 1 M KCl, 50 mM HEPES pH 7.5 and 0.5 mM TCEP pH 7.5. Protein samples (1.0 µl) were mixed with different broad screen

conditions (1.0 µl) and suspended over 500 µl of 1.5 M ammonium sulfate. These samples were organized into trays, which varied in protein concentrations in the range of 2 to 3 mg ml^{-1} and were alternatively incubated at 4°C and 20°C respectively. Drops were periodically monitored under a light microscope for crystal formation. Observed crystals were screened for protein content using a HighFlux HomeLab™ X-ray diffraction system (Rigaku).

3.3 Results

3.3.1 $SNM1A_{698-1040}$ has 5'-Phosphate-Specific Exonuclease Activity

Based on exonuclease assays performed with 5'PO_4 and 5'OH DNA substrates, $SNM1A^{698-1040}$ was found to have 5`phosphate-dependent exonuclease activity. As shown in figure 3.1, $SNM1A_{698-1040}$ selectively hydrolyzes DNA substrates with a 5'PO_4, while showing no activity for substrates with a 5'OH. This reaction occurred quickly, with the majority of DNA hydrolyzed to 2-3 nucleotides after 14 minutes. Interestingly, the rate of DNA hydrolysis appeared to slow down after 2 minutes, suggesting that $SNM1A_{698-1040}$ is more efficient at degrading longer DNA substrates.

To ensure that this observed exonuclease activity was from purified $SNM1A_{698-1040}$ and not a contaminating nuclease, the D736A and D736A/H37A point mutants were generated by site-directed mutagenesis. These mutants were purified identically to $SNM1A_{698-1040}$ and incubated with 5'PO_4 DNA substrates under identical reaction conditions. While both the D736A and D736A/H37A mutants have residual exonuclease activity, this is greatly reduced when compared

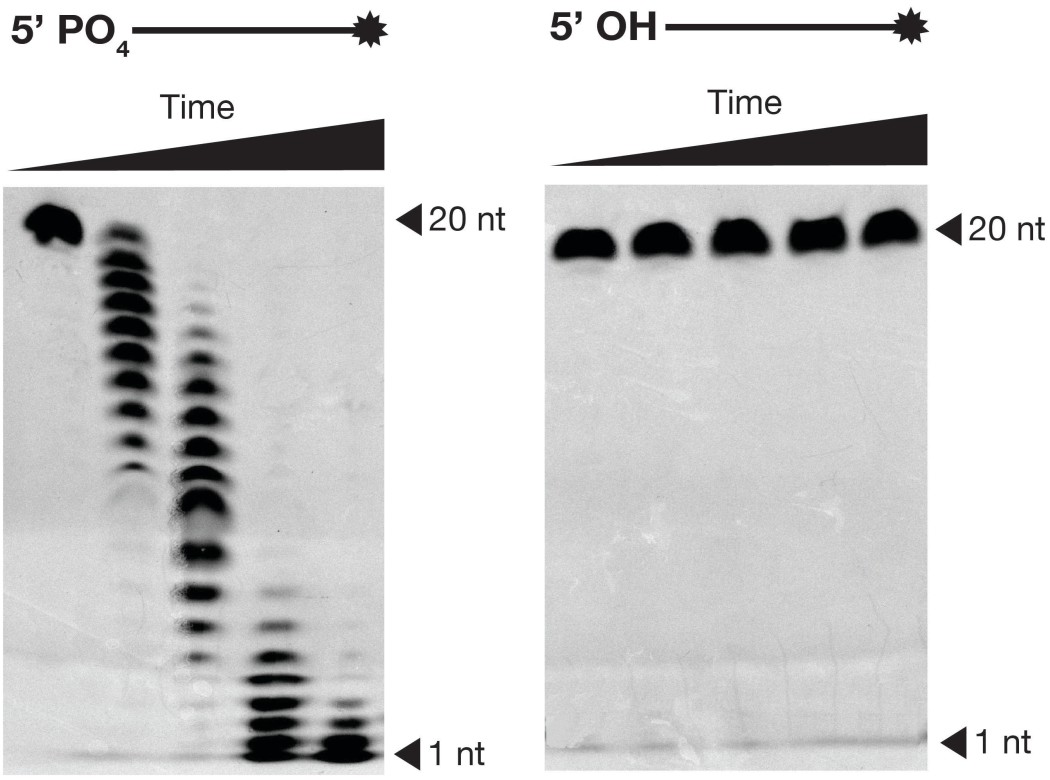

Figure 3.1: SNM1A$_{698-1040}$ has 5'phosphate-dependent exonuclease activity. Time courses were performed on 5'PO_4 and 5'OH DNA substrates. Protein and DNA concentrations were 33.33 nM and 66.66 nM respectively. Samples were taken at 30 seconds, 2 minutes, 6 minutes and 14 minutes. All products were resolved on a 20% denaturing PAGE gel and imaged by fluorescence.

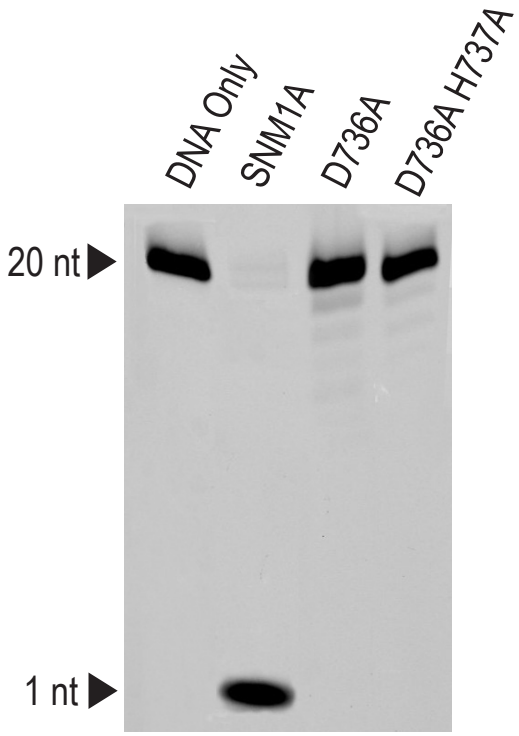

Figure 3.2: D736A and D736A/H737A mutants have reduced exonuclease activity. Exonuclease assays were performed on a 5'PO_4 ssDNA substrate. Reactions were allowed to incubate at 37°C for 30 minutes. Protein and DNA concentrations were 33.33 nM and 66.66 nM respectively. All samples were resolved on a 20% denaturing PAGE gel and imaged by fluorescence.

to wild type SNM1A$_{698-1040}$ (figure 3.2). This confirms that purified SNM1A$_{698-1040}$ is active and is responsible for the observed nuclease activity. It should be noted that the effect of the D736A and H737A mutations appears to be additive, since the D736A/H737A mutant appears to be less active than D736A.

3.3.2 SNM1A$_{698-1040}$ Preferentially Hydrolyzes ssDNA

In order to test SNM1A$_{698-1040}$ exonuclease activity on dsDNA, two complementary 20-nt oligonucleotides with a T_m of 75°C were annealed. This was confirmed by native-PAGE gel electrophoresis, and the dsDNA complex was found to be stable at the reaction temperature of 37°C (figure 3.3). With this validated, SNM1A$_{698-1040}$ exonuclease time courses were performed on both ssDNA and dsDNA substrates (figure 3.4). These assays were performed using SNM1A$_{698-1040}$ and DNA concentrations of 8.33 nM and 133.3 nM respectively. Under these conditions SNM1A$_{698-1040}$ is capable of hydrolyzing ssDNA and dsDNA to 1-2 nt and 1-5 nt after 64 minutes. A closer examination of the two time courses suggests that SNM1A$_{698-1040}$ degrades ssDNA approximately twice as fast compared to dsDNA. This can be seen at the 1 minute time point for ssDNA and the 2 minute time point for dsDNA, where the amount of hydrolysis at these two times is comparable. Similar to previous time courses, the rate of DNA hydrolysis slows down as the oligonucleotide reaches a certain length. In the case of ssDNA, SNM1A$_{698-1040}$ removes the first 12 nt in 16 minutes, with the remaining 8 nt hydrolyzed over the remaining 48 minutes.

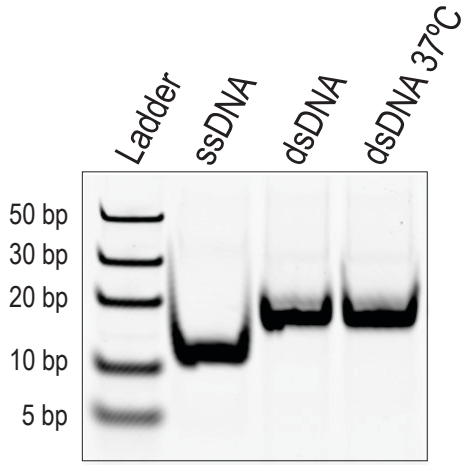

Figure 3.3: Annealed DNA substrate is stable at 37°C. DNA substrates were resolved on a 20% native PAGE gel and visualized by fluorescence.

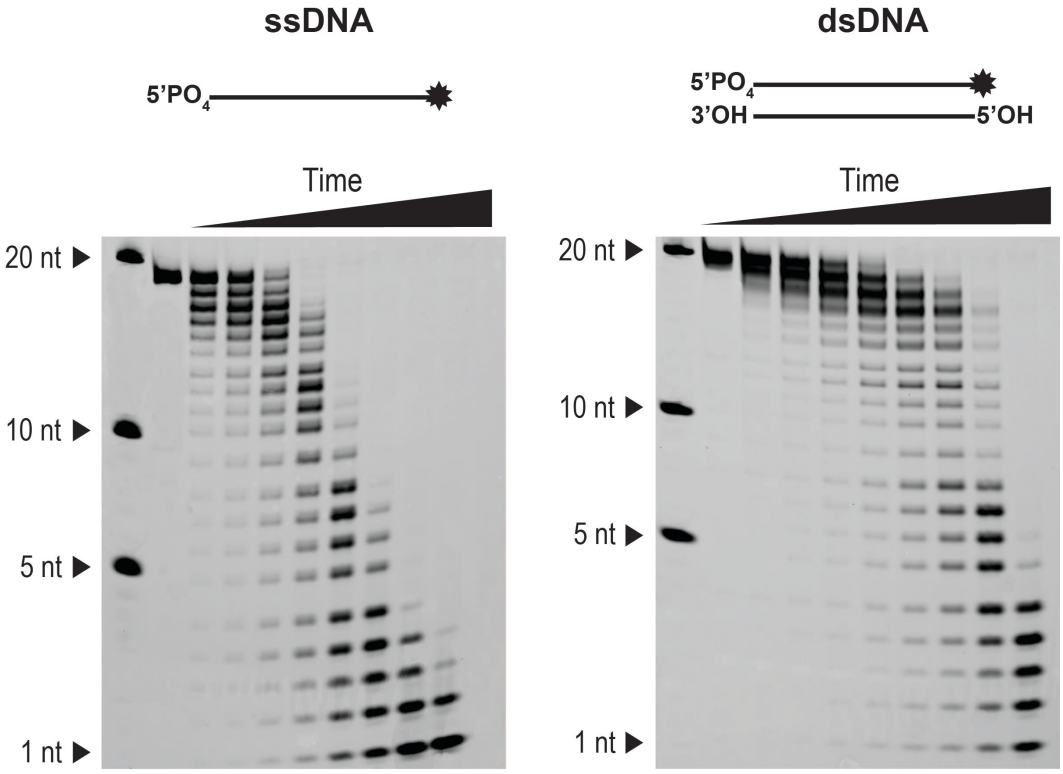

Figure 3.4: SNM1A$_{698\text{-}1040}$ has a preference for ssDNA over dsDNA. Protein and DNA concentrations were 8.33 nM and 133.33 nM respectively. Time points were taken at 30 seconds, 1 minute, 2 minutes, 4 minutes, 8 minutes, 16 minutes, 32 minutes and 64 minutes. All samples were resolved on 25% denaturing PAGE gel and imaged by fluorescence.

3.3.3 D736A/H737A DNA Binding does not require a 5' PO_4

To determine whether $SNM1A_{698-1040}$ is capable of binding to DNA in the absence of a 5' PO_4, EMSAs were performed with both 5' PO_4 and 5' OH ssDNA substrates. In these experiments, the D736A/H737A mutant was used instead of wild-type SNM1A to prevent hydrolysis of the 5' PO_4 substrate. As figure 3.5 shows, the D736A/H737A mutant appears to bind to both the 5' PO_4 and 5' OH substrates. Specifically, at a D736A/H737A concentration of 375 nM, the bands corresponding to these substrates decrease in intensity and are replaced with a supershifted smear. Further increasing the concentration of this mutant, results in a complete disappearance of this smear, possibly due to saturation of the DNA substrates. It is important to note, however, that a discrete supershifted band was not seen in these experiments. Tentatively, these results suggest that SNM1A-DNA binding is independent of the 5' PO_4.

3.3.4 Identification of $SNM1A_{698-1040}$ by PMF

$SNM1A_{698-1040}$ was digested in solution by trypsinization and the resulting peptides were analyzed by LC-MS/MS. The obtained peptide mass fingerprint was then compared against theoretical peptide masses arising from trypsin cleavage of proteins in the human genome. Based on this analysis, this sample was identified with high confidence to be human SNM1A, with a total of 22 unique peptides and 42 unique spectra (figure 3.6). These peptides collectively span from residues 703 to 1026, and cover 26% of the full-length SNM1A amino

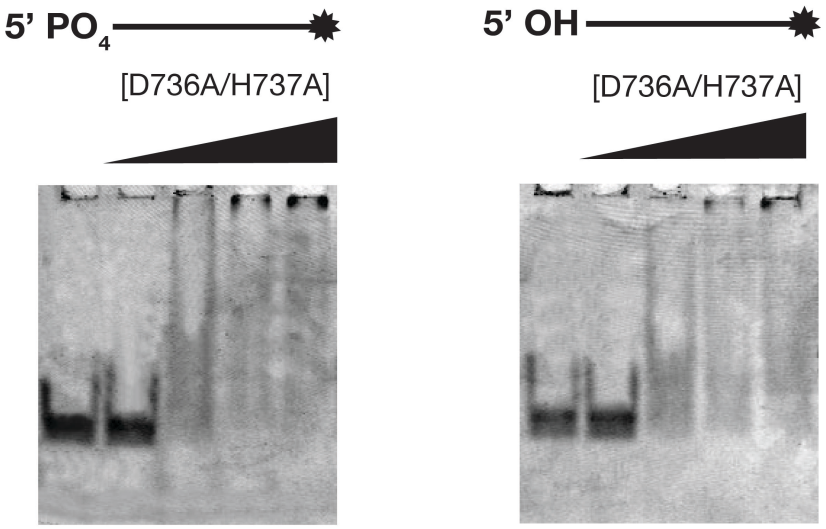

Figure 3.5: D736A/H737A DNA binding does not require a 5' phosphate. EMSAs were performed with the D736A/H737A mutant on ssDNA substrates containing a 5'phosphate or 5'hydroxyl. The concentration of DNA was kept constant at 15 nM, while the concentrations of D736A/H737A were 75 nM, 375 nM, 7875 nM and 12604 nM respectively. All products were run on 10% native PAGE gels and visualized by fluorescence.

acid sequence, or 80% of SNM1A$_{698-1040}$. Peptides and protein assignments were made with threshold confidence limits of 80% and 99% respectively. Table 3.1 shows the unique peptides identified and their representative spectra.

3.3.5 SNM1A$_{698-1040}$ has an Exact Molecular Mass of 38,932 Da

The exact molecular weight of SNM1A$_{698-1040}$ was determined by ESI-TOF-MS to be 38,934 Da (figure 3.7). By contrast the calculated molecular weight of SNM1A$_{698-1040}$ is 38,874.11 Da. However, this discrepancy can be reconciled by the fact that the cleavage of the TEV digestion site leaves a glycine residue at the N-terminus of SNM1A$_{698-1040}$, increasing its molecular mass to 38,931.16 Da. The observed peak therefore correlates with our predicted mass, with an error of 0.84 Da. This suggests that the SNM1A$_{698-1040}$ truncation is stable and does not undergo any additional cleavage after purification.

3.3.6 The N-terminus of SNM1A$_{698-1040}$ is Intact

Based on Edman sequencing of the N-terminus of SNM1A$_{698-1040,}$ the first four amino acids are glycine, lysine, lysine and threonine. The fifth amino acid is ambiguous, as no phenylthiohydantoin (PTH) derivative was detected, suggesting this residue was a cysteine or a modified amino acid. The last identified amino acid in this sequence was proline. This sequence is identical to that of our SNM1A$_{698-1040}$ truncation, which has an N-terminal GKKTCP sequence. These results indicate that the N-terminus of SNM1A$_{698-1040}$ is stable and is not cleaved off after purification.

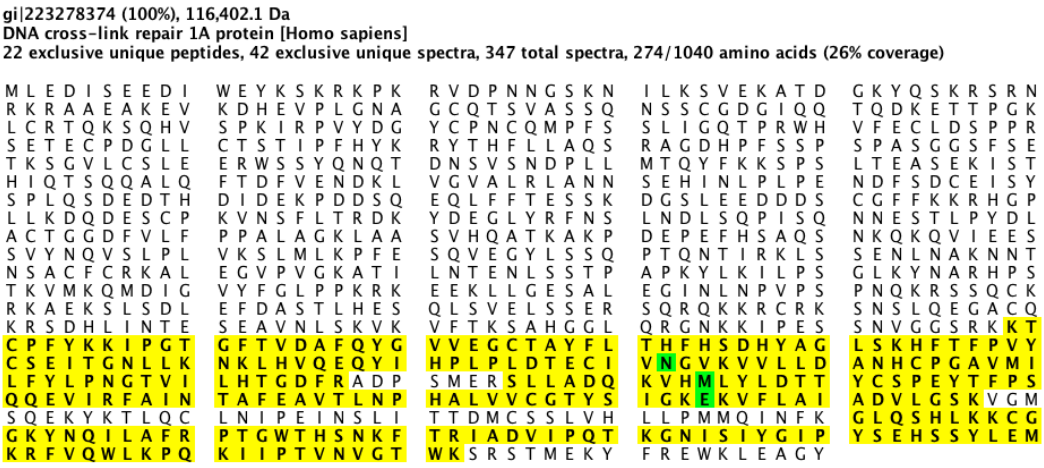

Figure 3.6: Peptide mass fingerprint of SNM1A$_{698-1040}$. Sequence coverage was 26% for full-length SNM1A and 80% for SNM1A$_{698-1040}$. Observed peptide fragments (yellow) span from residue 703 to residue 1026. Green residues indicate a deamidated asparagine, methionine sulfoxide and pyroglutamic acid respectively.

Table 3.1: Identified SNM1A$_{698-1040}$ peptides

Peptide	Sequence	m/z[a]	z	Actual Mass (Da)[b]	Error (ppm)[c]	Probability[d]
1	CGGKYNQILAFRPTGWTHSNK	1218.11	2+	2434.20	1.8	99%
2	EKVFLAIADVLGSK	745.44	2+	1488.86	0.59	100%
3	FAINTAFEAVTLNPHALVVCGTYSIGK	965.17	3+	2892.48	0.013	100%
4	FAINTAFEAVTLNPHALVVCGTYSIGKEK	788.42	4+	3149.63	3.0	95%
5	FVQWLKPQK	587.34	2+	1172.67	3.5	98%
6	GLQSHLKK	455.78	2+	909.54	1.2	100%
7	GNISIYGIPYSEHSSYLEMK	1144.55	2+	2287.09	1.3	100%
8	GNISIYGIPYSEHSSYLEMKR	815.40	3+	2443.19	1.4	95%
9	HFTFPVYCSEITGNLLK	676.01	3+	2025.00	-0.6	100%
10	HFTFPVYCSEITGNLLKNK	756.72	3+	2267.15	4.7	99%
11	IADVIPQTK	492.79	2+	983.57	2.1	100%
12	IIPTVNVGTWK	614.36	2+	1226.70	0.96	100%
13	KIPGTGFTVDAFQYGVVEGCTAYFLTHFHSDHYAGLSK	704.34	6+	4220.02	3.3	100%
14	KTCPFYK	472.24	2+	942.46	0.86	98%
15	LHVQEQYIHPLPLDTECIVNGVK	901.47	3+	2701.40	2.3	100%
16	RFVQWLKPQK	443.93	3+	1328.78	2.6	100%
17	SLLADQKVHMLYLDTTYCSPEYTFPSQQEVIR	1278.28	3+	3831.83	-4.7	100%
18	VFLAIADVLGSK	616.87	2+	1231.72	3.6	100%
19	VHMLYLDTTYCSPEYTFPSQQEVIR	1026.48	3+	3076.43	-0.74	100%
20	VVLLDANHCPGAVMILFYLPNGTVILHTGDFR	1184.96	3+	3551.86	4.2	100%
21	YNQILAFRPTGWTHSNK	1017.02	2+	2032.03	2.4	100%
22	YNQILAFRPTGWTHSNKFTR	813.09	3+	2436.25	2.3	100%

[a] Observed representative MS peak
[b] Calculated theoretical mass of peptide
[c] Mass accuracy associated with parent ion in parts per million (ppm)
[d] Confidence of peptide assignment to human SNM1A

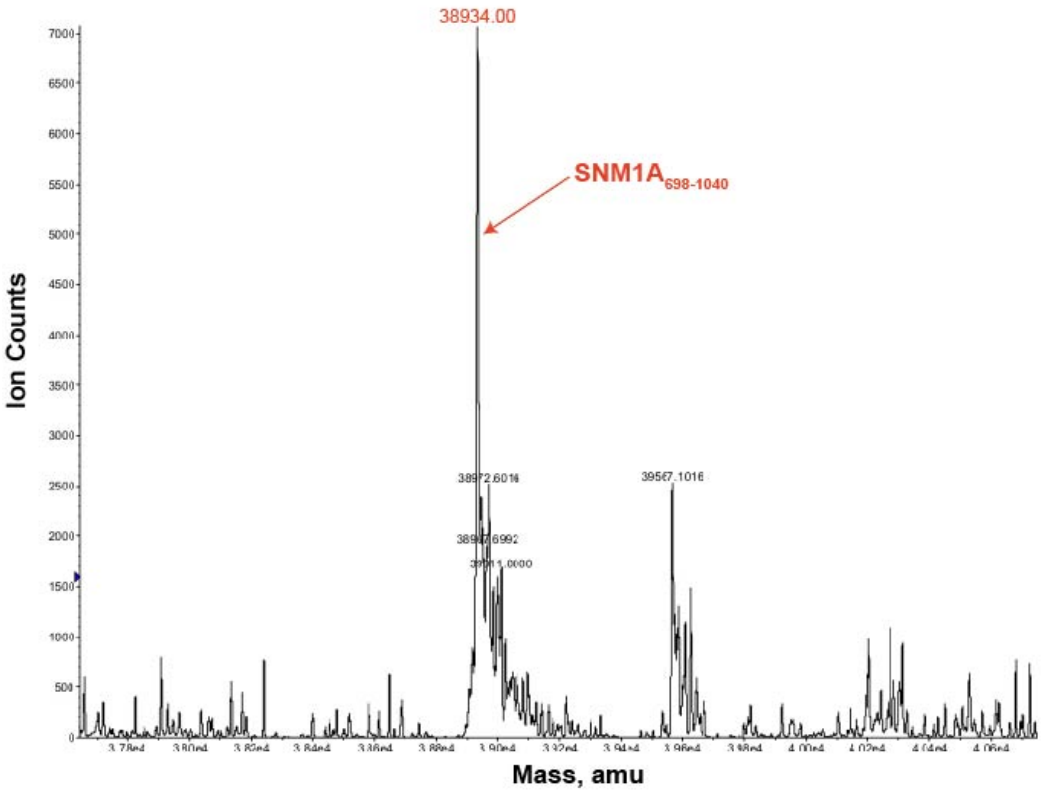

Figure 3.7: SNM1A$_{698-1040}$ ESI-TOF mass spectrum. The peak corresponding to SNM1A$_{698-1040}$ is assigned and coloured in red.

3.3.7 Crystallization attempts of SNM1A$_{698-1040}$

To date 672 crystallization conditions have been screened for SNM1A$_{698-1040}$. Since the yield of SNM1A$_{698-1040}$ is prohibitively low for crystallization, the majority of these conditions were screened with NusA-SNM1A$_{698-1040}$ that had been partially purified using a nickel affinity column. Specifically, these samples were treated with TEV protease to promote *in situ* proteolysis. The rationale for this being that the highly insoluble SNM1A$^{698-1040}$ truncation would more readily crystallize relative to the NusA or any other contaminants. Overall, from these screens, crystals have been observed in approximately 62 conditions. However, it should be noted that all crystals so far have tested negative for protein content, based on X-ray diffraction analysis.

3.4 Discussion

Although SNM1A exonuclease activity has been described in previous studies, there is merit in conducting similar experiments on SNM1A$_{698-1040}$, since it represents a novel truncation that has yet to be characterized (Hejna *et al*, 2007; Hazrati *et al*, 2007; Sengerova *et al*, 2012). With this in mind, we performed nuclease assays on this protein, to see how its biochemical activity compared to full-length SNM1A or SNM1A$_{608-1040}$. As expected, SNM1A$_{698-1040}$ demonstrates 5'-phosphate-specific exonuclease activity, showing no hydrolysis for DNA substrates with a 5' OH. To ensure that this activity was from SNM1A$_{698-1040}$ and not from a contaminating nuclease, the D736A and D736A/H737A mutants were generated by site-directed mutagenesis. These mutants have previously been

reported in the literature to be catalytically inactive (Wang *et al*, 2011; Sengerova *et al*, 2012; Hazrati *et al*, 2007, Hejna *et al*, 2008). Testing these mutants on 5'PO_4 ssDNA, they were found to have greatly reduced exonuclease activity compared to wild type $SNM1A_{698-1040}$. Interestingly, the D736A/H737A double mutant was found to be less active than D736A, indicating that the effect of these mutations is additive.

Once this 5'-phosphate-specific exonuclease activity was established, $SNM1A_{698-1040}$ was tested on dsDNA. Previous studies have shown SNM1A has a preference for ssDNA, however the extent of its exonuclease activity on dsDNA appears to be dependent on truncation length. Full-length SNM1A has been shown with greatly reduced activity on dsDNA (~1% of ssDNA), while $SNM1A_{608-1040}$ has been found to be much more active (Hazrati *et al*, 2007; Sengerova *et al*, 2012). Given this discrepancy, we performed ssDNA and dsDNA time course assays with $SNM1A_{698-1040}$. A dsDNA substrate with a high *Tm* of 75°C was used to ensure that it would remain annealed at a reaction temperature of 37°C. Based on these experiments, $SNM1A_{698-1040}$ was found to hydrolyze ssDNA twice as fast compared to dsDNA. This is similar to $SNM1A_{608-1040}$, which has been reported with k_{cat} values of 25.6 min^{-1} and 13.8 min^{-1} for ssDNA and dsDNA respectively. Since full-length SNM1A appears to be far less efficient with dsDNA, it is possible that the extended N-terminus of this protein sterically inhibits dsDNA from accessing the active site. Assuming this is true, it

is somewhat logical that $SNM1A_{698-1040}$ and $SNM1A_{608-1040}$ would have comparable kinetics, given their similar truncation lengths.

In addition to its preference for ssDNA, the rate of $SNM1A_{698-1040}$ hydrolysis was observed to slow down as the oligonucleotide decreased in length. This is similar to what has previously been reported for full-length SNM1A, which had decreased affinity for an oligonucleotide once it was shorter than an octamer (Hejna *et al*, 2007). Considering that $SNM1A_{608-1040}$ has also been shown to be more efficient at processing high molecular weight DNA substrates, it is appears that there is a correlation between DNA length and SNM1A processivity (Sengerova *et al*, 2012). However, this remains to be conclusively proven, since no DNA binding studies have ever been performed with SNM1A.

Since DNA binding experiments for SNM1A have never been reported in the literature, we performed EMSAs with both 5' PO_4 and 5' OH ssDNA substrates, to see if SNM1A could bind to DNA in the absence of a 5' PO_4. To prevent hydrolysis of the DNA substrates, we performed these assays using the D736A/H737A mutant. Furthermore, Mg^{2+} was also omitted from our buffers, as studies have shown it is necessary for SNM1A activity (Hejna *et al*, 2007). From these experiments, titrating increasing amounts of D736A/H737A resulted in the disappearance of both 5'PO_4 and 5' OH DNA bands, which were replaced with a supershifted smear. Although it is tempting to state that this result was due to D736A/H737A binding to both DNA substrates, it is not conclusive, since a

supershifted band was not observed. Tentatively however, these results suggest that SNM1A-DNA binding may be independent of the 5' PO_4.

In addition to these biochemical studies, $SNM1A_{698-1040}$ was characterized by mass spectrometry and Edman sequencing. These experiments were performed to address the anomalous molecular weight of $SNM1A_{698-1040}$ on a SDS-PAGE gel. Specifically, $SNM1A_{698-1040}$ was observed with an apparent molecular weight of 35 kDa, despite having a calculated molecular weight of 38.9 kDa. Given this result, we hypothesized that $SNM1A_{698-1040}$ may have been proteolytically cleaved after purification, resulting in a smaller band. To investigate this possibility, we performed three different studies. First, a trypsinization and LC-MS/MS would be used to obtain a PMF of $SNM1A_{698-1040}$, allowing us to map peptides to this truncation. Second, an ESI-TOF MS would be used to directly obtain an accurate mass of intact $SNM1A_{698-1040.}$ Lastly, Edman sequencing was used to unambiguously identify the first few amino acids at the N-terminus.

From these experiments, $SNM1A_{698-1040}$ was surprisingly found to be intact, with a molecular weight corresponding to 38,932 Da. This was also verified by Edman sequencing, which showed that the N-terminus was intact, with the first six residues corresponding with our known amino acid sequence. While these results indicate that $SNM1A_{698-1040}$ is not being proteolytically cleaved, it is curious as to why it would run smaller than expected on a SDS-PAGE gel. It has been established, that based on amino acid composition, anomalous migration can

occur. The tumor suppressor protein p53, for example, has an apparent molecular weight of 53 kDa, despite having a molecular mass of 48 kDa (Ziemer *et al*, 1982). This is attributed to its high proline content, which slows its migration through an SDS-PAGE gel. However, the cause for the anomalous apparent molecular weight of $SNM1A_{698-1040}$ remains to be determined.

Although a main goal of this project was to obtain the crystal structure of apo $SNM1A_{698-1040}$, the low yields of protein obtained have made this endeavor difficult. To compensate for this, we attempted to crystallize NusA-$SNM1A_{698-1040}$ using *in situ* proteolysis, due to the higher amounts of protein available. However, one problem with this technique is that many of the observed crystals are formed from contaminating proteins. As such, we are no longer using *in situ* proteolysis for crystallization screening. Additionally, since the SGC has recently solved the structure of apo $SNM1A_{698-1040}$, current efforts are now focused on obtaining a co-crystal structure of this protein bound to DNA. This will be elaborated upon in chapter 5, where the future work for this project will be discussed.

CHAPTER 4 – SNM1A INHIBITOR ASSAYS

4.1 Rationale and Experimental Design

In addition to characterizing human SNM1A, another aspect of my project involved the development of an *in vitro* assay to identify inhibitors of this enzyme. To accomplish this, we have designed a fluorescence-based nuclease assay to measure the inhibition of SNM1A activity. For these assays, a ssDNA substrate containing an internal fluorophore-quencher pair was used. As figure 4.1 shows, under normal circumstances $SNM1A_{698-1040}$ degrades this substrate and frees the fluorophore, creating a fluorescent signal. However, in the presence of an inhibitor, this exonuclease activity is suppressed and therefore results in a loss of fluorescence. To validate this assay, we have performed a series of positive and negative controls to obtain an assay window and a preliminary Z' factor. Furthermore, we have since optimized this assay for high throughput screening (HTS), and have obtained an improved Z' factor as a result. Currently, we are in the process of screening for inhibitors of $SNM1A_{698-1040}$, with the ultimate objective of identifying compounds that can be used in conjunction with ICL-based chemotherapy.

4.2 Materials and Methods

4.2.1 SNM1A Inhibitor Assay

The DNA substrate used for initial inhibitor assays consisted of a 5'-[PO_4] ACTCTTCCTAGC[Fl-dT]A[Da-dT]GGTTCGATCAAGA-3' oligo, where a fluorescein-derivatized deoxythymidine (Fl-dT) and dabycl-derivatized deoxy-thymidine (Da-dT) were used as the fluorophore and quencher respectively (this

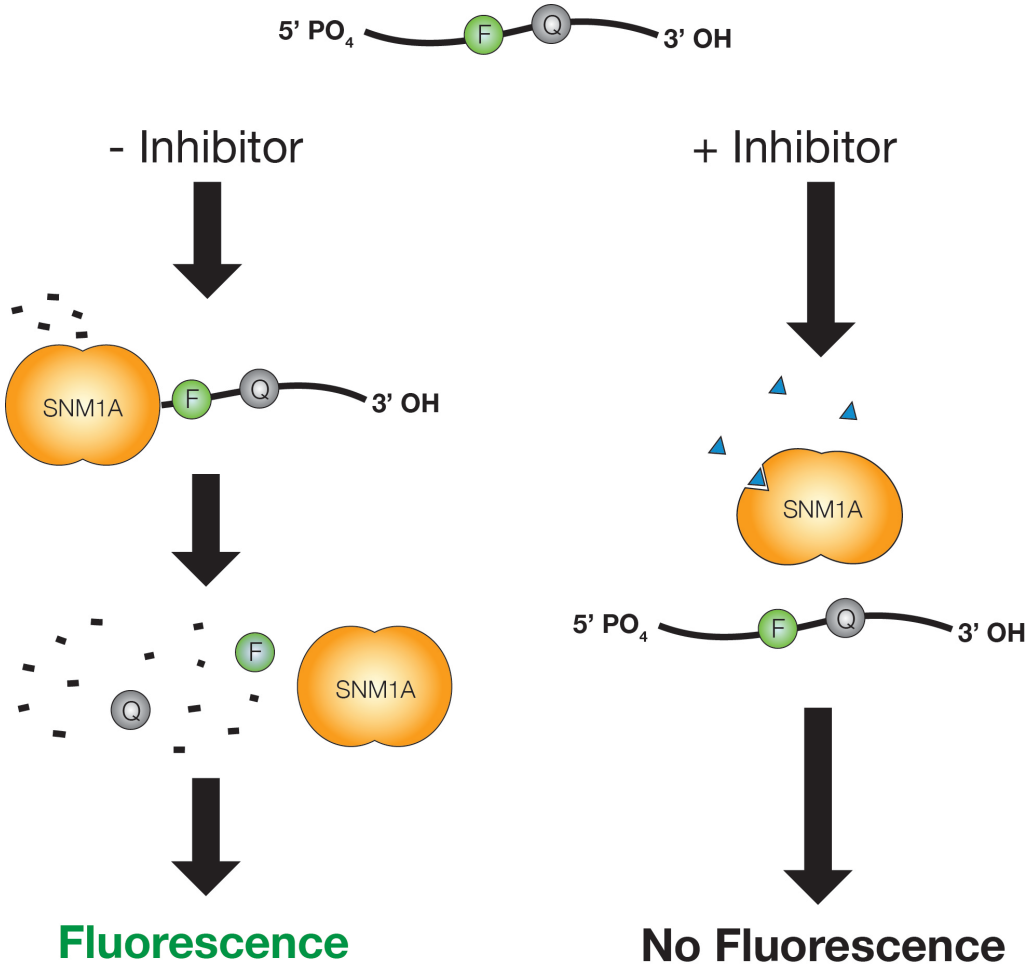

Figure 4.1: *In vitro* inhibitor assay for SNM1A. Under normal conditions, SNM1A will degrade the DNA substrate and separate the fluorophore-quencher pair, leading to an increase in fluorescence. In the presence of an inhibitor, however, the exonuclease activity of SNM1A will be ablated, leading to a loss of fluorescent signal.

substrate was a generous gift from the Li lab). For this assay, negative controls were performed by incubating this DNA substrate (100 nM) with $SNM1A_{698-1040}$ (25 nM) in a total volume of 75 ul of 1 x Buffer F (50 mM Tris-acetate pH 7.2, 10 mM Mg acetate, 75 mM potassium acetate, 1 mM DTT) supplemented with 100 $\mu g\ ml^{-1}$ BSA. Positive controls, on the other hand, consisted of DNA substrate and buffer. These were incubated at room temperature for an hour in black, flat-bottom 96-well plates (Corning) prior to measuring fluorescence using a BioTek Synergy 4 Hybrid Microplate Reader. Excitation and emission wavelengths for these experiments were 485 nm and 526 nm respectively. Positive and negative controls were done in replicates, and a Z' factor was calculated using their corresponding means (μ) and standard deviations (σ). The formula used for the calculation of the Z' factor is as follows: $Z'=1-[(3\sigma_{+c} + 3\sigma_{-c})/|\mu_{-c} - \mu_{+c}|]$ (Zhang *et al*, 1999).

4.2.2 Assay Optimization

In order to optimize this assay for HTS, multiple kinetic experiments were performed with different concentrations of $SNM1A_{698-1040}$ and DNA. A new substrate was designed, consisting of a 5'-[PO_4]AGC[Fl-dT]A[BHQ1-dT]GG-TTCGATCAAG-3' oligonucleotide (MWG Operon). This substrate was shortened at the 5' end and contained a Black hole quencher 1 derivatized deoxythymidine (BHQ1-dT) instead of Da-dT. Kinetic assays were performed in triplicate by incubating variable concentrations of this DNA substrate (100 nM, 150 nM, 200 nM, 250 nM) with increasing amounts of $SNM1A_{698-1040}$ (25 nM, 50

nM, 75 nM, 100 nM) in a total volume of 75 µl of 1 x Buffer F. Samples were mixed in black, flat-bottom 384-well plates (Corning), and fluorescence measurements were recorded every 30 seconds for an hour at 26°C using a BioTek Synergy 4 Hybrid Microplate Reader, with excitation and emission wavelengths of 485 nm and 526 nm respectively. In addition to this, dimethylsulfoxide (DMSO), EDTA and zinc acetate were also added to certain wells to investigate their effect on $SNM1A_{698-1040}$ activity. Specifically, DMSO was added to a final concentration of 1.33% (v/v) to see if $SNM1A_{698-1040}$ could tolerate its use as a solvent for small molecules. While EDTA and zinc acetate were added to final concentration of 6.67 mM and 1 mM respectively, to see if they could be used as positive controls for HTS.

4.2.3 Final Z' Factor Determination

A final Z' factor was determined to accurately represent an actual high throughput screen. These experiments were performed by Cameron Rzadki, with the help of Cecilia Murphy, at the HTS lab in the Centre for Microbial Chemical Biology (CMCB). In this experiment, 1 µl of DMSO was pipetted into 192 wells on a black, flat-bottom 384-well plate (Corning). Following this, 1 µl of zinc acetate (75 mM) was added to half of these wells, while 1 µl of ddH_2O was added to the other half. These steps were performed at a Biomek FX workstation (Beckman-Coulter) using a BioRAPTR (Beckman/Coulter) liquid dispensing system. This plate was then transferred a second Biomek FX workstation, where 63 µl of $SNM1A_{698-1040}$ (59.5 nM) in 1.19 x Buffer F was added to each of the 192

wells using a second BioRAPTR liquid dispensing system. The plate was allowed to incubate at 30 minutes at ambient temperature (26°C), prior to the addition of 10 µl of DNA mix (1.125 uM). The final concentrations of all reagents are as follows: 50 nM $SNM1A_{698-1040}$, 150 nM DNA substrate, 1 mM Zinc Acetate (for positive controls only) and 1 x Buffer F. Reactions were subsequently mixed by pipetting up and down once, and allowed to run for 40 minutes. Fluorescence measurements were performed using an Envision Plate Reader (Perkin Elmer), with excitation and emission wavelengths of 485 nm and 526 nm respectively. Positive and negative controls were run 96 times each, and a Z' factor was calculated as described in section 4.2.1.

4.3 Results

4.3.1 Initial Inhibitor Assay has a Z' Factor of 0.68

The initial $SNM1A_{698-1040}$ inhibitor assay was performed using protein and DNA concentrations of 50 nM and 100 nM respectively. Specifically, the substrate used consisted of a ssDNA 28-mer with a fluorescein-dT as the fluorophore and a dabcyl-dT as the quencher. Since we had not identified a compound capable of inhibiting $SNM1A_{698-1040}$ at the time, our positive controls in these experiments consisted of DNA and buffer only. This assay was conducted by measuring the endpoint fluorescence after an hour incubation at room temperature. Performing this assay with 22 replicates, we found that the means of the negative and positive controls had mean fluorescence values of 613 and 24, with relative standard deviations of 9.82% and 7.75% respectively (figure

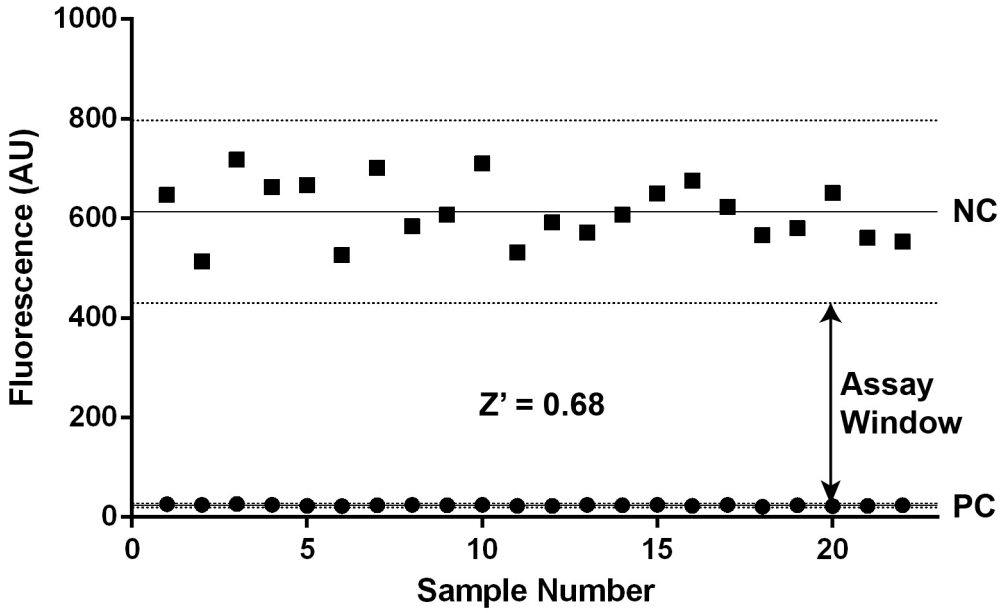

Figure 4.2: Initial inhibitor assay has a Z' factor of 0.68. An assay window was generated by performing the positive and negative controls in replicate and calculating the Z' factor. The positive control (PC) consisted of only the DNA substrate, while the negative control (NC) had both $SNM1A_{698\text{-}1040}$ and DNA substrate. Dotted lines indicate 3 standard deviations from the PC and NC averages.

4.2). Although there is a 25.5-fold difference in signal between these two controls, the high standard deviation associated with the negative control reduces the assay window to 68% (Z'factor = 0.68).

4.3.2 Zinc Acetate and EDTA Inhibit SNM1A

In order to prepare this assay for HTS, several experiments needed to be performed. First, kinetic assays were carried out with varying concentrations of $SNM1A_{698-1040}$ and a new 18-mer DNA substrate. It was found that a concentration of 50 nM $SNM1A_{698-1040}$ and 150 nM DNA was optimal for maximizing signal, while minimizing the amount of reagents used. This was done out of practical considerations, considering the large amounts of protein and substrate that would be consumed over the course of a screening campaign. Additionally, to ensure that DMSO solvent from a compound library would not interfere with this assay, we tested $SNM1A_{698-1040}$ in the presence of 1.33% (v/v) DMSO. At this concentration, $SNM1A_{698-1040}$ functioned normally and was capable of complete hydrolysis in 40 minutes at a temperature of 26°C (figure 4.3). Consequently, a time point of 40 minutes was chosen as the endpoint for subsequent inhibitor assays. Lastly, we tested $SNM1A_{698-1040}$ with both EDTA and zinc acetate to see if they could act as appropriate positive controls. These were found to inhibit $SNM1A_{698-1040}$ exonuclease activity at concentrations of 6.67 mM and 1 mM for EDTA and zinc acetate respectively. Interestingly, the addition of zinc acetate appears to lower fluorescence over time, though it is unclear how this occurs.

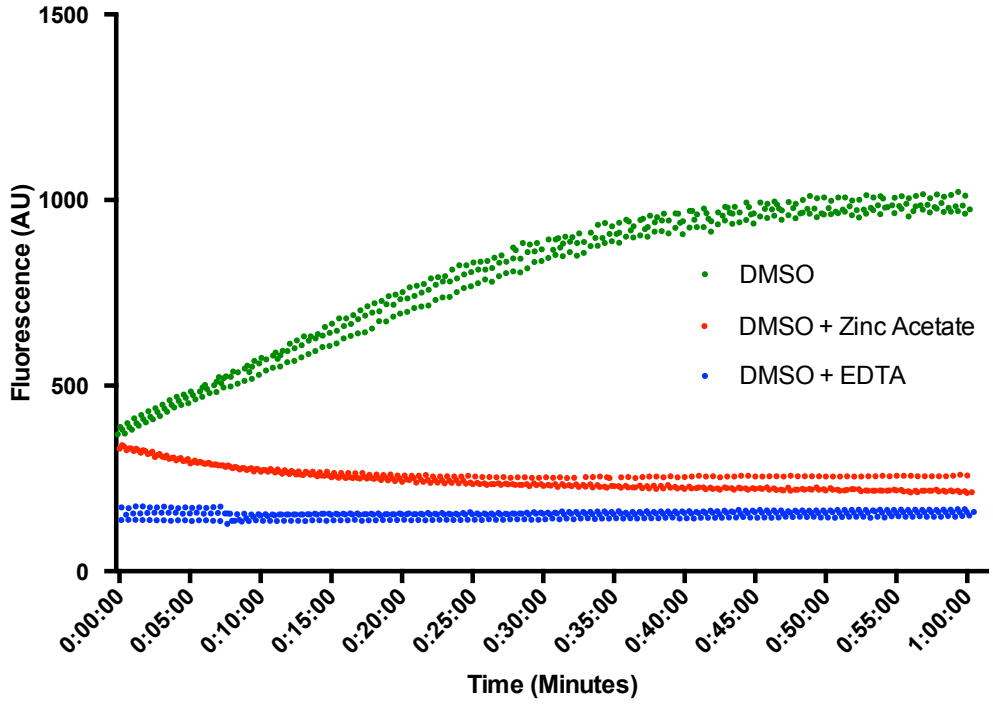

Figure 4.3: $SNM1A_{698-1040}$ is inhibited by zinc acetate and EDTA. Kinetic assays were performed with $SNM1A_{698-1040}$ on a fluorophore-quencher ssDNA substrate. These assays were carried out in the presence of 1.33% (v/v) DMSO and the addition of either 1 mM zinc acetate or 6.67 mM EDTA. In the absence of either of these compounds, $SNM1A_{698-1040}$ is capable of complete hydrolysis in approximately 40 minutes at 26°C.

4.3.3 Optimized Inhibitor Assay has a Z' Factor of 0.88

Taking this information into consideration, an optimized inhibitor assay was performed using the equipment in the HTS lab at the CMCB. DMSO and ddH_2O were used as a negative control, while zinc acetate was used as the positive control. These were added in 2 µl volumes to a dry plate prior to the addition of a protein buffer mix (63 µl). Following this, the plate was allowed to incubate for 30 minutes at room temperature to allow for equilibration of zinc acetate with $SNM1A_{698\text{-}1040}$. 10 µl of DNA substrate was then added to each well and mixed. Reactions were allowed to proceed for 40 minutes at 26°C before a fluorescence endpoint read was taken. With 96 replicates, negative and positive controls had mean fluorescence values of 6.81×10^6 and 4.36×10^5, with relative standard deviations of 3.56% and 4.99% respectively. This corresponds to a 15.6-fold difference between the negative and positive controls, which is lower than our initial assay (25.5-fold). However, the lower standard deviation with these values result in an overall larger assay window of 88% (Z' factor = 0.88).

4.4 Discussion

Due to their high lethality to dividing cells, DNA cross-linking agents are commonly used in chemotherapy (West *et al*, 2011). While this treatment is initially effective, tumors are known to develop resistance to cross-linking agents over time. This can be partially attributed to the upregulation of proteins involved in ICL repair such as ERCC1, which is highly expressed in tumor cells resistant to cisplatin (Lee *et al*, 2003). A new paradigm in chemotherapy therefore consists of

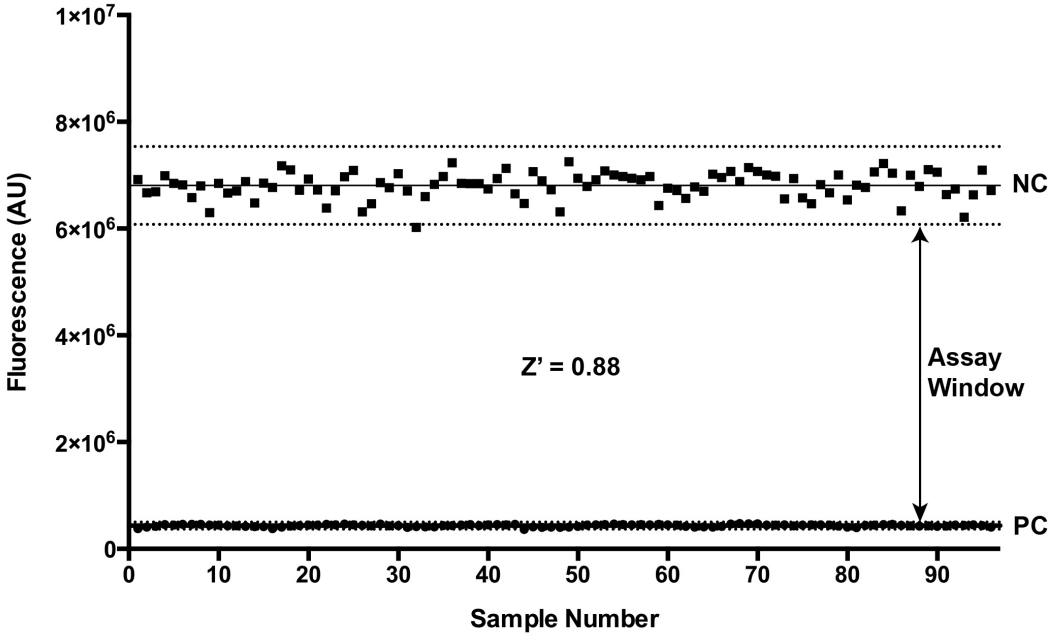

Figure 4.4: Optimized inhibitor assay has a Z' factor of 0.88. The positive control (PC) used was 1 mM zinc acetate, while the negative control (NC) consisted of $SNM1A_{698-1040}$ and DNA substrate. Dotted lines indicate 3 standard deviations from the PC and NC averages.

disrupting DNA repair pathways to resensitize tumor cells to cross-linking agents. This can be seen in the example of the poly(ADP-ribose) polymerase 1 (PARP1) inhibitor ABT-188 (Abott Laboratories), which has been shown to regress tumors when used in combination with cisplatin, carboplatin and cyclophosphamide therapy (Donawho *et al*, 2007). With this in mind, we postulated that SNM1A might represent a target for anti-cancer therapy, since it is selectively involved in the repair of ICLs. Furthermore, since SNM1A has previously been difficult to obtain in appreciable quantities, it represents an under-explored and novel target for inhibition.

Since HTS typically requires large quantities of protein, inhibitor screening for SNM1A has previously been prohibited by low purification yields. However, we have recently developed an improved method for purifying $SNM1A_{698-1040}$ from *E. Coli*, providing us a unique opportunity to perform HTS experiments. With this in mind, we have developed a fluorescence-based nuclease assay to measure the inhibition of SNM1A activity. As previously described, this assay involves the use of a ssDNA substrate containing a fluorophore-quencher (Fl-dT and Da-dT) pair to allow for rapid readout. To validate this assay and establish proof of concept, we performed a Z' factor experiment (Zhang *et al*, 1999). This consisted of measuring the fluorescence of DNA substrate with and without $SNM1A_{698-1040}$, representing the negative and positive controls respectively. Means (μ_{+c}, μ_{-c}) and standard deviations (σ_{+c}, σ_{-c}) were calculated for these controls, and the assay window was defined as the

region between $|\mu_{-c}-3\sigma_{-c}|$ and $|\mu_{+c}+3\sigma_{-c}|$. This represents the 95% confidence limit for any hits during HTS. For 22 replicates, we obtained a Z' factor of 0.68, indicating that 68% of the signal range was available for HTS. Since a Z' factor greater than 0.5 is considered to be excellent, this assay was considered to be a viable platform for HTS with SNM1A.

To improve upon our initial Z' factor, several parameters were optimized. First, a new DNA substrate was designed, consisting of a shortened 5' end, to allow $SNM1A_{698-1040}$ quicker access to the fluorophore-quencher pair. This was done to lower the incubation times required for signal generation. Additionally, a BHQ1-dT quencher was substituted for the original Da-dT. The rationale being that BHQ1-dT's absorbance maximum (534 nm) is better suited for Fl-dT emission (517 nm) compared to Da-dT's absorbance (453 nm). Using this new DNA substrate, we found that the initial enzyme and DNA concentrations of 25 nM and 100 nM were insufficient for a reasonable assay window. This may have been due to the omission of BSA from our buffers, which could have stabilized $SNM1A_{698-1040}$ and prevented surface adhesion. However, we were able to compensate for this by increasing enzyme and DNA concentrations to 50 nM and 150 nM respectively for future experiments.

In addition to these considerations, it was also necessary to verify $SNM1A_{698-1040}$ activity in the presence of DMSO. Since our HTS protocol involved the addition of 1 µl of compound solubilized in DMSO to a total reaction volume of 75 µl, $SNM1A_{698-1040}$ was tested in the presence of 1.33% (v/v) DMSO.

Under these conditions, $SNM1A_{698-1040}$ was active and completely hydrolyzed the fluorophore-quencher pair in approximately 40 minutes. A temperature of 26°C was chosen, since this was the ambient temperature within the HTS facility. Optimizing parameters at this temperature would therefore simplify our HTS assay, as an incubator would not be required for this step. Lastly, we reasoned that EDTA and zinc acetate could be as accurate positive controls. Being a metal-ion-dependent nuclease, EDTA would inhibit $SNM1A_{698-1040}$ by removing Zn^{2+} ions from its active site. Paradoxically, elevated concentrations of Zn^{2+} have also been shown to inhibit SNM1A (Hejna *et al*, 2007; Sengerova *et al*, 2012). Testing both of these compounds, we found that EDTA and zinc acetate were inhibitory at concentrations of 6.67 mM and 1 mM respectively.

With these optimized parameters, we designed a simple HTS protocol using an automated liquid dispenser to reduce pipetting error. Using this protocol, which is described in section 4.2.3, we were able obtain a Z' factor of 0.88 for 96 replicates. This is an extremely high Z' factor and furthermore indicates that our assay is sensitive, robust and reproducible. However, despite this, a limiting factor in these experiments is the amount of $SNM1A_{698-1040}$ required. Currently, for every compound plate tested (384 compounds), 1.44 nmol of $SNM1A_{698-1040}$ is required. Given that an average yield of $SNM1A_{698-1040}$ from 4L of culture is 6-7 nmol, this corresponds to 4-5 screens per purification. Although this is an acceptable compromise, the efficiency of this assay can be improved. The re-addition of BSA into reaction buffers, for example, may allow us to lower

$SNM1A_{698-1040}$ concentrations back to 25 nM. Alternatively, it may be possible to use a more abundant, but less pure version of this enzyme, such as NusA-$SNM1A_{698-1040}$. Despite this, we are currently proceeding to screen for inhibitors of $SNM1A_{698-1040}$.

CHAPTER 5 – CONCLUDING REMARKS

5.1 Summary

Although human SNM1A has been identified for over a decade, it has only been biochemically and structurally characterized within the last few years. Studies have indicated that this enzyme is selectively involved in the repair of DNA cross-links, and operates during the TLS step of ICL repair (Yang *et al*, 2010). Specifically, SNM1A has been shown to function as a translesional exonuclease on an ICL, and is thought to improve the efficiency of TLS (Wang *et al*, 2011). Interestingly, this translesional exonuclease activity has only been shown *in vitro* and has not been reported for any other nuclease to date. Given this unusual biochemical function, we have chosen to study the mechanism of human SNM1A as it relates to ICL repair. More specifically, the main objective of my thesis was to determine the structure of SNM1A and understand how it is uniquely involved in the repair of DNA cross-links.

At the onset of this project, we were unable to obtain SNM1A for biochemical experiments. As such, a major component of my thesis involved the purification of this protein. Here we report the purification of human $SNM1A_{698-1040}$ as a His_6-NusA tagged protein from 4 L of *E. Coli* cell culture. This specific truncation and fusion tag was identified after screening 70 different constructs for solubility and expression, using a semi-high throughput purification assay. While the yield associated with this purification is low (0.25 mg from 4 L of cell culture), this is sufficient for biochemical and inhibitor studies on $SNM1A_{698-1040}$. Furthermore, this represents a significant improvement over previously reported

methods of purifying this protein, as milligram quantities can be obtained from 4 L of cell culture and 3 purification steps. This is in contrast to $SNM1A_{608-1040}$, which requires 24 L and 6 columns for purification (Sengerova *et al*, 2012).

Similar to other β-CASP nucleases, we have shown that the purified $SNM1A_{698-1040}$ possesses 5'-phosphate-dependent exonuclease activity. This was confirmed by the generation of the D736A and D736A/H737A mutants, which exhibit greatly reduced exonuclease activity. $SNM1A_{698-1040}$ was found to hydrolyze ssDNA twice as fast when compared to dsDNA, which is similar to previously reported kinetic parameters for $SNM1A_{608-1040}$ (Sengerova *et al*, 2012). Furthermore, we have also found that $SNM1A_{698-1040}$ has decreased affinity for oligonucleotides shorter than an octamer, which is again consistent with previous reports (Hejna *et al*, 2007). Interestingly, we have obtained preliminary data suggesting that SNM1A-DNA binding is independent of the 5' PO_4. While this remains to be conclusively proven, it implies that the 5' PO_4 plays a role in the catalysis of this enzyme, as opposed to being involved in substrate binding.

Although the original objective of this thesis was to determine the structure of apo SNM1A, we have been unsuccessful in our attempts to crystallize this protein. In particular, our inability to obtain sizable quantities of $SNM1A_{698-1040}$ has made crystallization screening exceptionally difficult. Furthermore, the apo structure of $SNM1A_{676-1040}$ was recently solved by the SGC, making this goal obsolete. Despite this, we have been able to obtain some structural information using mass spectrometry. Specifically, we have found that although $SNM1A_{698-}$

$_{1040}$ has an apparent molecular weight of 35 kDa by SDS-PAGE analysis, it has an actual molecular mass of 38.9 kDa. This is corroborated by Edman sequencing and the PMF of this protein, which have shown that this truncation is stable from proteolytic degradation during purification. At this point in time, however, it is unclear as to why SNM1A$_{698\text{-}1040}$ migrates anomalously on a SDS-PAGE gel.

In addition to this work, we have also developed a fluorescence-based assay for identifying inhibitors of SNM1A$_{698\text{-}1040}$. As described in Chapter 4, this assay uses a ssDNA substrate containing a fluorophore-quencher pair, to measure inhibition of SNM1A activity. Optimizing this assay for HTS, we have been able to obtain a Z' factor of 0.88 for 96 replicates. This is an extremely high Z' factor, indicating that our assay is sensitive, robust and reproducible. It has recently come to our attention that a similar assay has previously been reported in the literature (Sengerova *et al*, 2012). However, since we have been able to purify SNM1A$_{698\text{-}1040}$ in higher quantities, we are in a more advantageous position to perform HTS screening.

Taking into account all the known information on SNM1A, it is difficult to establish a mechanism for its role in ICL repair. Disruptions of the *snm1a* gene have shown a mild phenotype in mouse, chicken and human cell lines, which is selectively sensitive to the MMC and SJG-136 cross-linking agents (Dronkert *et al*, 2000; Ishiai *et al*, 2004, Wang *et al*, 2011). Although it has been hypothesized that SNM1A is selectively involved in the repair of minor groove ICLs, the question of how this substrate would be structurally recognized remains to be

addressed (Wang *et al*, 2011). Furthermore, while biochemical assays have shown that SNM1A is capable of translesionally digesting ICL substrates, it is remarkably inefficient in this function, as it pauses two nucleotides after the cross-link. Recent work in our lab has shown that $SNM1A_{698-1040}$ also possesses hairpin endonuclease activity. This provides a second potential mechanism as to how this enzyme might process an ICL substrate.

5.2 Future Directions

While we have finally succeeded in obtaining soluble and active $SNM1A_{698-1040}$, there is a need for higher purification yields in order to perform crystallization and inhibitor screening. Since the SGC has solved the structure of $SNM1A_{676-1040}$, it can be inferred that they have a method for obtaining sizable quantities of this enzyme. As a result, we have recently cloned this truncation into a His_6-NusA expression vector. This was done, since it is possible that the additional 22 amino acids may promote greater solubility by physically preventing protein molecules from aggregating. Currently, we are attempting to purify this construct under a similar procedure to $SNM1A_{698-1040}$. If this does not work, another alternative would be to move from *E. Coli* to a Baculovirus/insect cell expression system, since the SGC has been known to utilize this system for proteins that are insoluble in *E. Coli.*

Another aspect of this project that must be addressed is the DNA binding studies. Although we have tentative evidence to suggest that D736A/H737A is capable of binding both 5' PO_4 and 5' OH DNA substrates, this must be

characterized in greater detail, before it can be conclusively stated. Currently, we are attempting to use fluorescence polarization as a technique to study this interaction, as it has the distinct advantage of providing quantitative information such as dissociation constants (K_d).

Since there is no crystal structure of SNM1A bound to DNA, this remains as an open avenue of research to pursue. This is particularly important, since the structure of DNA-bound SNM1A could potentially answer questions about the 5' PO_4 dependence of this enzyme. Furthermore, if an ICL substrate could be crystallized with SNM1A, it would provide information about this enzyme's ability to recognize a cross-link. In order to perform these experiments, DNA substrates of variable length must be designed and tested for binding using EMSAs. Once appropriate substrates have been identified, they can be used for co-crystallization experiments with SNM1A.

Since we have adapted our assay for HTS and obtained an acceptable Z' factor, the next step is to begin screening for inhibitors of $SNM1A_{698-1040}$. Once inhibitors are identified, they can be characterized *in vitro* using enzyme kinetics. Additionally, the anti-cancer activity of these inhibitors can be tested *in vivo* using ICL-resistant tumor cell lines. Finally, using X-ray crystallography, structural determination of SNM1A co-crystallized with an inhibitor can be used for structure-based drug design.

REFERENCES:

Ahkter, S., Richie, C.T., Deng, J.M., Brey, E., Zhang, X., Patrick, C. Jr, Behringer, R.R. & Legerski, R.J. (2004) Deficiency in SNM1 abolishes an early mitotic checkpoint induced by spindle stress. *Mol Cell Biol*. **24**: 10448-10455

Ahkter, S., Richie, C.T., Zhang, N., Behringer, R.R., Zhu, C. & Legerski, R.J. (2005) SNM1-deficient mice exhibit accelerated tumorigenesis and susceptibility to infection. *Mol Cell Biol*. **25**: 10071-10078

Alpi, A.F., Pace, P.E., Babu, M.M. & Patel, K.J. (2006) Mechanistic insight into site-restricted monoubiquitination of the FANCD2 by Ube2t, FANCL, and FANCI. *Mol Cell* **23**: 589-596

Alpi, A. F. & Patel, K. J. (2009) Monoubiquitylation in the Fanconi anemia DNA damage response pathway. *DNA Repair*. **8**: 430–435

Akkari, Y.M.N., Bateman, R.L, Reifsteck, C.A., Olson, S.B. & Grompe, M. (2000) DNA replication is required to elicit cellular responses to psoralen-induced DNA interstrand cross-links. *Mol. Cell. Biol.* **20**: 8283-8289

Ashwood-Smith, M. J., Poulton, G. A., Barker, M. & Mildenberger, M. (1980) 5-Methoxypsoralen, an ingredient in several suntan preparations, has lethal, mutagenic and clastogenic properties. *Nature* **285**: 407–409

Bree, R.T., Neary, C., Samali, A. & Lowndes, N.F. (2004) The switch from survival responses to apoptosis after chromosomal breaks. *DNA Repair*. **3**: 989-995

Callebaut, I., Moshous, D., Mornon, J.P. & Villartay, J.P. (2002) Metallo-β-lactamase fold within nucleic acids processing enzymes: the β-CASP family. Nucleic Acids Res. **30**: 3592-3601

Cassier, C., Chanet, R., Henriques, J.A. & Moustachhi, E. (1980) The effects of three PSO genes on induced mutagenesis: a novel class of mutationally defective yeast. *Genetics*. **96**: 841-857

Cattell, E., Sengerova, B. & McHugh, P.J. (2010) The SNM1/Pso2 family of repair nucleases: from yeast to man. *Environ Mol Mutagen*. **51**: 635-645

Ciccia, A., Ling, C., Coulthard, R., Yan, Z., Xue, Y., Meetei, A.R., Laghmani el, H., Joenje, H., McDonald, N., de Winter, J.P., Wang, W. & West, S.C. (2007) Identification of FAAP24, a Fanconi anemia core complex protein that interacts with FANCM. *Mol Cell* **25**: 331–343

Collis, S.J., Ciccia, A., Deans, A.J., Horejsi, Z., Martin, J.S., Maslen, S.L., Skehel, J.M., Elledge, S.J., West, S.C. & Boulton, S.J. (2008) FANCM and FAAP24 function in ATR-mediated checkpoint signaling independently of the Fanconi anemia core complex. *Mol Cell* **32**: 313-324

Dabholkar, M., Vionnet, J., Bostick-Bruton, F., Yu, J.J. & Reed, E. (1994) Messenger RNA levels of XPAC and ERCC1 in ovarian cancer tissue correlate with response to platinum-based chemotherapy. *J Clin Invest.* **94**: 703-708

Danshiitsoodol, N., de Pinho, C.A., Matoba, Y., Kumagai, T., Sugiyama, M. (2006). The mitomycin C (MMC)-binding protein from MMC-producing microorganisms protects from the lethal effect of bleomycin: crystallographic analysis to elucidate the binding mode of the antibiotic to the protein. *J Molec Biol* **360**: 398–408

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

Dooley, P.A., Zhang, M., Korbel, G.A., Nechev, L.V., Harris, C.M., Stone, M.P. & Harris, T.M. (2003) NMR determination of the conformation of a trimethylene interstrand cross-link in an oligodeoxynucleotide duplex containing a 5'-d(GpC) motif. *J Am Chem Soc*. **125**: 62-72

Donawho, C.K., Luo, Y., Penning, T.D., Bauch, J.L., Bouska, J.J., Bontecheva-Diaz, V.D., Cox, B.F., DeWeese, T.L., Dillehay, L.E., Ferguson, D.C., Ghoresihi-Haack, N.S., Girmm, D.R., Guan, R., Han, E.K, Holley-Shanks, R.R., Hristov, B., Idler, K.B., Jarvis, K., Johnson, E.F., Kleinberg, L.R., Klinghofer, V., Lasko, L.M., Liu, X., Marsh K.C., McGonigal, T.P., Meulbroek, J.A., Olson, A.M., Palma, J.P., Rodriguez, L.E., Shi, Y., Stavropoulos, J.A., Tsurutani, A.C., Zhu, G.D., Rosenberg, S.H., Giranda, V.L. & Frost, D.J. (2007) ABT-888, an orally active poly(ADP-ribose) polymerase inhibitor that potentiates DNA-damaging agents in preclinical tumor models. *Clin Cancer Res.* **13**: 2728-2737

Dorleans, A., Li de la Sierra-Gallay, I., Piton, J., Zig, L., Gilet, L., Putzer, H. & Condon, C. (2011) Molecular basis for the recognition and cleavage of RNA by the bifunctional 5'-3' exo/endoribonuclease RNase J. *Structure*. **19**: 1252-1261

Dronkert, M.L., de Wit, J., Boeve, M., Vasconcelos, M.L., van Steeg, H., Tan, T.L., Hoeijmakers, J.H. & Kanar, R. (2000) Disruption of mouse SNM1 causes increased sensitivity to the DNA interstrand cross-linking agent mitomycin C. *Mol Cell Biol*. **20**: 4553-4561

Fischhaber, P.L., Gall, A.S., Duncan, J.A. & Hopkins, P.B. (1999) Direct Demonstration in Synthetic Oligonucleotides that BCNU Cross-Links N1 of Deoxyguanosine to N3 of Deoxycytidine on Opposite Strands of Duplex DNA. *Cancer Res.* **59**, 4363-4368

Frankenberg-Schwager, M., Kirchermeier, D., Greif, G., Baer, K., Becker, M. & Frankenberg, D. (2005) Cisplatin-mediated DNA double-strand breaks in replicating but not in quiescent cells of the yeast Saccharomyces cerevisiae. *Toxicology* **212**: 175–184

Freibaum, B.D. & Counter, C.M. (2008) The protein hSNM1B is stabilized when bound to the telomere-binding protein TRF2. *J Biol Chem.* **283**: 23671-23676

Garcia-Higuera, I., Taniguchi, T., Ganesan, S., Meyn, M.S., Timmers, C., Hejna, J., Grompe, M. & D'Andrea, A.D. (2001) Interaction of the Fanconi anemia proteins and BRCA1 in a common pathway. *Mol Cell* **7**: 249–262

Garcin, E.D., Hosfield, D.J., Desai, S.A., Haas, B.J., Bjoras, M., Cunninghanm, R.P. & Tainer, J.A. (2008) DNA apurinic-apyrimidinic site binding and excision by endonuclease IV. *Nat Struct Mol Biol*. **15**: 515-522

Gari, K., Décaillet, C., Delannoy, M., Wu, L. & Constantinou, A. (2008) Remodeling of DNA replication structures by the branch point translocase FANCM. *Proc Natl Acad Sci USA.* **105**: 16107–16112

Geng, L., Huntoon, C. J. & Karnitz, L. M. (2010) RAD18-mediated ubiquitination of PCNA activates the Fanconi anemia DNA repair network. *J. Cell Biol.* **191**: 249–257

Giaccone, G. (2000) Clinical perspectives on platinum resistance. *Drugs* **59** (Suppl 4): 9-17

Golovanov, A.P., Hautbergue, G.M., Wilson, S.A. & Lian, L.Y. (2004) A simple method for improving protein solubility and long-term stability. *J Am Chem Soc.* **126**: 8933-8939

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

Goodman, L. S., Wintrobe, M.M., Dameshek, W., Goodman, M.J., Gilman, A. & McLennan, M.T. (1946) Nitrogen mustard therapy. Use of methyl-bis(β-

chloroethyl)amine hydrochloride and tris(β-chloroethyl)amine hydrochloride for Hodgkin's disease, lymphosarcoma, leukemia and certain allied and miscellaneous disorders. *J Am Med Assoc*. **132**, 126–132

Graslund, S., Sagemark, J., Bergland, H., Dahlgren, L.G., Flores, A., Hammarstrom, M., Johansson, I., Kotenyova, T., Nilsson, M., Nordlund, P. & Weigelt, J. (2008) The use of systematic N- and C-terminal deletions to promote production and structural studies of recombinant proteins. *Protein Expr. Purif.* **58**: 210-221

Guo, C., Fischhaber, P.L., Luk-Paszyc, M.J., Masuda, Y., Zhou, J., Kamiya, K., Kisker, C. & Friedberg, E.C. (2003) Mouse Rev1 protein interacts with multiple DNA polymerases involved in translesion DNA synthesis. *EMBO J.* **22**: 6621-6630

Hanada, K., Budzowska, M., Modesti, M., Maas, A., Wyman, C., Essers, J. & Kanaar, R. (2006) The structure-specific endonuclease MUS81-EME1 promotes conversion of interstrand DNA crosslinks into double-strands breaks. *EMBO J.* **25**: 4921–4932

Harder, H. C., Smith, R. G. & Leroy, A. F. (1976) Template primer inactivation by cis- and trans-dichlorodiammine platinum for human DNA polymerase α, β, and Rauscher murine leukemia virus reverse transcriptase, as a mechanism of cytotoxicity. *Cancer Res*. **36**: 3821–3829.

Hazrati, A., Ramis-Castelltort, M., Sarkar, S., Barber, L.J., Schofield, C.J., Hartley, J.A. & McHugh, P.J. (2008) Human SNM1A suppresses the DNA repair defects of yeast *pso2* mutants. *DNA Repair.* **7**: 230-238

Hearst, J.E. (1989) Photochemistry of the psoralens. *Chem Res Toxicol*. **2**: 69-75

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

Hlavin, E.M., Smeaton, M.B. & Miller, P.S. (2010) Initation of DNA interstrand cross-link repair in mammalian cells. *Environ. Mol. Mutagen.* **51**: 604-624

Huang, H., Zhu, L., Reid, B. R., Drobny, G. P. & Hopkins, P. B. (1995) Solution structure of a cisplatin-induced DNA interstrand cross-link. *Science.* **270**: 1842–1845

Ishiai, M., Kimura, M., Namikoshi, K. Yamazoe, M., Yamamoto, K., Arakawa, H., Agematsu, K., Matsushita, N., Takeda, S., Buertedde, J.M. & Takata, M.

(2004) DNA cross-link repair protein SNM1A interacts with PIAS1 in nuclear focus formation. *Mol Cell Biol*. **24**: 10733-10741

Ivanov, I., Tainer, J.A. & McCammon, J.A. (2007) Unraveling the three-metal ion catalytic mechanism of the DNA repair enzyme endonuclease IV. *Proc Natl Acad Sci USA*. **104**: 1465-1470

Kim, H. & D'Andrea, A.D. (2012) Regulation of DNA cross-link repair by the Fanconi anemia/BRCA pathway. *Genes Dev*. **26**, 1393-1408

Kim, J. M., Parmar, K., Huang, M., Weinstock, D.M., Ruit, C.A., Kutok, J.L. & D'Andrea, A.D. (2009) Inactivation of murine Usp1 results in genomic instability and a Fanconi anemia phenotype. *Dev Cell* **16**: 314–320

Kim, H., Yang, K., Dejsuphong, D. & D'Andrea, A.D. (2012) Regulation of Rev1 by the Fanconi anemia core complex. *Nat Struct Mol Biol.* **19**: 164-170

Knox, R. J., Friedlos, F., Lydall, D. A. & Roberts, J. J. (1986) Mechanism of cytotoxicity of anticancer platinum drugs: evidence that cis-diamminedichloro-platinum(II) and cis-diammine-(1,1-cyclobutanedicarboxylato) platinum(II) differ only in the kinetics of their interaction with DNA. *Cancer Res.* **46**: 1972–1979

Kozekov, I.D., Nechev, L.V., Moseley, M.S., Harris, C.M., Rizzo, C.J, Stone. M.P. & Harris, T.M. (2003) DNA interchain cross-links formed by acrolein and crotonaldehyde. *J Am Chem Soc.* **125**: 50-61

Lange, S. S., Takata, K. & Wood, R. D. (2011) DNA polymerases and cancer. *Nature Rev Cancer.* **11**: 96–110

Lawley, P.D. & Phillips, D.H. (1996) DNA adducts from chemotherapeutic agents. *Mutat Res*. **355**: 13-40

Lee, K.B., Parker, R.J., Bohr, V., Cornelison, T. & Reed, E. (1993) Cisplatin sensitivity/resistance in UV repair-deficient Chinese hamster ovary cells of complementation groups 1 and 3. *Carcinogenesis*. **14**: 2177-2180

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

Lieber, M. R., Ma, Y., Pannicke, U. & Schwarz, K. (2003) Mechanism and regulation of human non-homologous DNA end-joining. *Nature Rev Mol Cell Biol* **4**: 712–720

Ling, C., Ishiai, M., Ali, A.M., Medhurstm A.L., Neveling, K., Kalb, R., Yan, Z., Xue, Y., Oostra, A.B., Auerbach, A.D., Hoatlin, M.E., Schindler, D., Joenje, H., de Winter, J.P., Takata, M., Meetei, A.R. & Wang, W. (2007) FAAP100 is essential for activation of the Fanconi anemia-assocaited DNA damage response pathway. *EMBO J.* **26**, 2104-2114

Lopez, P.J., Marchand, I., Joyce, S.A. & Dreyfus, M. (1999) The C-terminal half of RNase E, which organizes the *Escherichia coli* degradosome, participates in mRNA degradation but not rRNA processing *in vivo*. *Mol Microbiol.* **33**: 188-199

Ludlum, D.B. (1997) The chloroethylnitrosoureas: sensitivity and resistance to cancer chemotherapy at the molecular level. *Cancer Invest.* **15**: 588-598

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

Manderfeld, M. M., Schafer, H. W., Davidson, P. M. & Zottola, E. A. (1997) Isolation and identification of antimicrobial furocoumarins from parsley. *J Food Prot.* **60**: 72–77

Masuda, Y., Ohmae, M., Masuda, K. & Kamiya, K. (2003) Structure and enzymatic properties of a stable complex of the human REV1 and REV7 proteins. *J Biol Chem*. **278**: 12356–12360

McNeil, E.M. & Melton, D.W. (2012) DNA repair endonuclease ERCC1-XPF as a novel therapeutic target to overcome chemoresistance in cancer therapy. *Nucleic Acids Res.* **40**: 9990-10004

Mirchandani, K. D., McCaffrey, R. M. & D'Andrea, A. D. (2008) The Fanconi anemia core complex is required for efficient point mutagenesis and Rev1 foci assembly. *DNA Repair.* **7**: 902–911

Moshous, D., Callebaut, I., de Chasseval, R., Corneo, B., Cavazzana-Calvo, M., Le Deist, F., Tezcan, I., Sanal, O., Bertrand, Y., Philippe, N., Fischer, A. & de Villartay, J.P. (2001) Artemis, a novel DNA double-stranded break repair/V(D)J recombination protein is mutated in human severe combined immune deficiency. *Cell*. **105**: 177-186

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

Novy, R., Drott, D., Yaeger, K. & Mierendorf, R. (2001) Overcoming the codon bias of *E. Coli* for enhanced protein expression. *in**Nova**tions*. **12**: 1-3

Niedernhofer, L.J., Daniels, J.S., Rouzer, C.A., Greene, R.E. & Marnett, L.J. (2003) Malondialdehyde, a product of lipid peroxidation, is mutagenic in human cells. *J Biol Chem.* **278**: 31426-31433

Niedernhofer, L.J. (2007) The Fanconi anemia signalosome anchor. *Mol Cell* **25**: 487–490

O'Connor, P.M. & Kohn, K.W. (1990) Comparative pharmacokinetics of DNA lesion formation and removal following treatment of L1210 cells with nitrogen mustards. *Cancer Commun.* **2**: 387-394

Querfeld, C., Rosen, S.T., Kuzel, T.M., Kirby, K.A., Roenigk, H.H. Jr., Prinz, B.M. & Guitart, J. (2005) Long-term follow-up of patients with early-stage cutaneous T-cell lymphoma who achieved complete remission with psoralen plus UV-A monotherapy. *Arch Dermatol.* **141**: 305–311

Rai, K. R., Peterson, B.L., Appelbaum, F.R., Kolitz, J., Elias, L., Shepherd, L., Hines, J., Threatte, G.A. (2000) Fludarabine compared with chlorambucil as primary therapy for chronic lymphocytic leukemia. *N Engl J Med*. **343**: 1750–1757

Raschle, M., Knipscheer, P., Enoiu, M., Angelov, T., Sun, J., Griffith, J.D., Ellenberger, T.E., Scharer, O.D. & Walter, J.C. (2008) Mechanism of replication-coupled DNA interstrand crosslink repair. *Cell.* **134**: 969–980

Richie, C.T., Peterson, C., Lu, T., Hittelman, W.N., Carpenter, P.B. & Legerski, R.J. (2002) hSNM1 colocalizes and physically associates with 53BP1 before and after DNA damage. *Mol Cell Biol.* **22**: 8635-8647

Rink, S.M., Solomon, M.S., Taylor, M.J., Rajur, S.B., McLaughlin, L.W. & Hopkins, P.B. (1993) Covalent structure of a nitrogen mustard-induced DNA interstrand cross-link: An N7-to-N7 linkage of deoxyguanosine residues at the duplex sequence 5'-d(GNC) *J Am Chem Soc.* **115**: 2551–2557

Romier, C., Dominguez, R., Lahm, A., Dahl, O. & Suck, D. (1998) Recognition of single-stranded DNA by nuclease P1: high resolution crystal structures of complexes with substrate analogs. *Proteins*. **32**: 414-424

Ross, A.L., Simpson, L.J., & Sale, J.E. (2005) Vertebrate DNA damage tolerance requires the C-terminus but not BRCT or transferase domains of REV1. *Nucleic Acids Res.* **33**, 1280-1289

Scharer, O.D. (2005) DNA interstrand crosslinks: natural and drug-induced DNA adducts that induce cellular responses. *ChemBioChem*. **6**: 27-32

Sengerova, B., Allerston, C.K., Abu, M., Lee, S.Y., Hartley, J., Kiakos, K., Schofield, C.J., Hartley, J.A., Gileadi, O. & McHugh, P.J. (2012) Characterization of the human SNM1A and SNM1B/Apollo DNA repair exonucleases. *J Biol Chem.* **287**: 26254-26267

Smogorzewska, A., Desetty, R., Saito, T.T., Schlabach, M., Lach, F.P., Sowa, M.E., Clark, A.B., Kunkel, T.A., Harper, J.W., Colaiacovo, M.P. & Elledge, S.J. (2010) A genetic screen identifies FAN1, a Fanconi anemia-associated nuclease necessary for DNA interstrand crosslink repair. *Mol Cell.* **39**: 36–47

Siddik, Z.H. (2003) Cisplatin: mode of cytotoxic action and molecular basis of resistance. *Oncogene*. **22**: 7265-7279

Singh, T.R., Saro, D., Ali, A.M., Zheng, X.F., Du, C.H., Killen, M.W., Sachpatzidis, A., Waheng, K., Pierce, A.J., Xiong, Y., Sung, P. & Meetei, A.R. (2010) MHF1-MHF2, a histone-fold-containing protein complex, participates in the Fanconi anemia pathway via FANCM. *Mol Cell.* **37**: 879–886

Stoepker, C., Hain, K., Schuster, B., Hilhorst-Hofstee, Y., Rooimans, M.A., Steltenpool, J., Oostra, A.B., Eirich, K., Korthof, E.T., Nieuwint, A.W., Jaspers, N.G., Bettecken, T., Joenje, H., Schindler, D., Rouse, J. & de Winter, J.P. (2011) SLX4, a coordinator of structure-specific endonucleases, is mutated in a new Fanconi anemia subtype. *Nat Genet.* **43**: 138-141

Stoll, D. B., Lavin, P. T. & Engstrom, P. F. (1985) Hematologic toxicity of cisplatin and mitomycin in combination for squamous cell carcinoma of esophagus. *Am J Clin Oncol*. **8**: 231–234.

Stranks, S.D., Ecroydd, H., Sluyter, S.V., Waters, E.J., Carver, J.A. & Smekal, L.V. (2009) Model for amorphous aggregation processes. *Phys Rev E.* **80**: 1-13

Studier, F.W. (1991) Use of Bacteriophage T7 Lysozyme to Improve an Inducible T7 Expression System. *J. Mol. Biol.* **219**: 37-44

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

Tomasz, M. (1995) Mitomycin C: small, fast and deadly (but very selective). *Chem Biol.* **2**: 575-579

Wang, A.T., Sengerova, B., Cattell, C., Inagawa, T., Hartley, J.M., Kiakos, K., Burgess-Brown, N.A., Swift, L.P., Enzlin, J.H., Schofield, C.J., Gileadi, O., Hartley, J.A. & McHugh, P.J. (2011) Human SNM1A and XPF-ERCC1 collaborate to initiate DNA interstrand cross-link repair. *Genes Dev.* **25**: 1859-1870

Wang, Z., Fast, W., Valentine, A.M. & Benkovic, S.J. (1999) Metallo-beta-lactamase: structure and mechanism. *Curr Opin Chem Biol*. **3**: 614-622

Yang, K., Moldovan, G.L. & D'Andrea, A.D. (2010) RAD18-dependent recruitment of SNM1A to DNA repair complexes by a ubiquitin-binding zinc finger. *J Biol Chem.* **285**: 19085-19091

Zajdela, F, & Bisagni, E. (1981) 5-Methoxypsoralen, the melanogenic additive in suntan preparations, is tumorigenic in mice exposed to 365 nm UV radiation. *Carcinogenesis* **1981**: 121–127

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

Ziemer, M.A., Mason, A. & Carlson, D.M. (1982) Cell-free translations of prolein-rich mRNAs. *J Biol Chem*. **257**: 11176-11180