OPTIMIZING THE LARGE-SCALE PRODUCTION OF SAW1 AND THE SAW1-RAD1-

RAD10 NUCLEASE COMPLEX FOR STRUCTURAL STUDIES

OPTIMIZING THE LARGE-SCALE PRODUCTION OF SAW1 AND THE SAW1-RAD1-RAD10 NUCLEASE COMPLEX FOR STRUCTURAL STUDIES

By MARGARITA A. RASHEV, B.Sc. (Honours)

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree of Master of Science

McMaster University MASTER OF SCIENCE (2017) Hamilton, Ontario (Biochemistry)

TITLE: Optimizing the large-scale production of Saw1 and the Saw1-Rad1-Rad10 nuclease complex for structural studies

AUTHOR: Margarita A. Rashev, B.Sc. (University of Waterloo)

SUPERVISOR: Dr. Alba Guarné

NUMBER OF PAGES: 128

ABSTRACT

Yeast Rad1-Rad10 is a structure specific nuclease that processes branched double-strand break (DSB) repair intermediates; the persistence of which can impede normal DNA metabolism. The single strand annealing (SSA) mechanism of DSB repair acts when homologous repeats flank both sides of the DSB. End resection from the 5′ ends of the break exposes complementary sequences at the flanking repeats, which are annealed to form 3′ non-homologous flap structures. Saw1 recruits Rad1-Rad10 recruits to these 3′ non-homologous flaps, where Rad1-Rad10 incises the DNA and removes the flap. Saw1 has affinity towards branched DNA structures and forms a stable complex with Rad1-Rad10. The mechanism of both structure specific recruitment and nucleolytic activity of the Saw1-Rad1-Rad10 complex is currently unknown. To study this nuclease complex, we need to produce large quantities of pure, stable, and active recombinant protein. Using dynamic light scattering (DLS) and differential scanning fluorimetry (DSF)-based high throughput thermal stability assays, we have developed a method for large-scale production of recombinant Saw1. This optimized method has increased the stability and yield of protein, thereby allowing for future biochemical investigation of Saw1. Similarly, we have optimized the large-scale production of the higher molecular-weight complex (Saw1-Rad1-Rad10) and improved the homogeneity of the recombinant complex. We have also biochemically characterized the minimal branched DNA substrates for both Saw1 and Saw1-Rad1-Rad10. This work allows for biochemical investigation into the molecular mechanism of eukaryotic 3′ non-homologous flap removal during SSA.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION

CHAPTER 2 MATERIALS AND METHODS

CHAPTER 3 RESULTS: OPTIMIZED PURIFICATION OF RECOMBINANT SAW1

CHAPTER 4 RESULTS: PURIFICATION OF SAW1-RAD1-RAD10 COMPLEX

CHAPTER 5 DISCUSSION AND CONCLUSIONS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND SYMBOLS

DNA	Deoxyribonucleic acid
SSB	Single-strand break
DSB	Double-strand break
ICL	Interstrand cross-link
DDR	DNA damage response
MMR	Mismatch repair
ROS	Reactive oxygen species
RNS	Reactive nitrogen species
BER	Base excision repair
NER	Nucleotide excision repair
UV	Ultra violet
GG-NER	Global genome nucleotide excision repair
TC-NER	Transcription coupled nucleotide excision repair
IR	Ionizing radiation
NHEJ	Non-homologous end-joining
c-NHEJ	Canonical-NHEJ
Alt-EJ	Alternative end joining
MRX	Mre11-Rad50-Xrs2
MRN	MRE11-RAD50-NBS1
rNTP	Ribonucleoside triphosphate
dNTP	Deoxyribonucleoside triphosphate
Pol μ	Polymerase μ
TCR	T-cell receptor
HR	Homologous recombination
ssDNA	Single stranded DNA
dsDNA	Double stranded DNA
CDK1	Cyclin-dependent kinase 1
nt	nucleotide
bp	Base pair
RPA	Replication protein A
HJ	Holliday junction
SDSA	Synthesis-dependent strand annealing
BIR	Break induced replication
SSA	Single strand annealing
HSV-1	Herpes Simplex Virus 1
EM	Electron microscopy
MMEJ	Microhomology mediated end-joining
FA	Fanconi Anemia
SUMO	Small Ubiquitin-like Modifer
XP	Xeroderma Pigmentosum
HhH	Helix-hairpin-Helix
IPTG	Isopropyl β-D-1-thiogalactopyranoside
OD_{600}	Optical density (600 nm)
βME	β-mercaptoethanol

PMSF	Phenylmethylsulfonyl fluoride
LDAO	Lauryldimethylamine-N-Oxide
EDTA	Ethylenediaminetetraacetic acid
MW	Molecular weight
DLS	Dynamic light scattering
DTT	Dithiothreitol
R	Arginine
E	Glutamate
RE	Arginine-Glutamate
DSF	Differential scanning fluorimetry
°C	Degrees Celsius
NaAce	Sodium Acetate
NaPhos	Sodium Phosphate
KPhos	Potassium Phosphate
AmmAce	Ammonium Acetate
Imid.	Imidazole
6-FAM	6-Carboxyfluorescein
BSA	Bovine serum albumin
His-Saw1	Hexa-histidine tagged Saw1
PDI	Polydispersion index
IMAC	Immobilized metal affinity chromatography
IEC	Ionic exchange chromatography
SEC	Size exclusion chromatography
PEG	Polyethylene glycol
kDa	Kilo-Dalton
EMSA	Electrophoretic mobility shift assay
Sc	Saccharomyces cerevisiae
Sm	Saccharomyces mikatate
Sp	Saccharomyces paradoxus
Sca	Saccharomyces castelli
His-Rad1	Hexa-histidine tagged Rad1
tRNA	Transfer RNA
SDS	Sodium dodecyl sulphate
PAGE	Polyacrylamide gel electrophoresis
TG	Tris-Glycine
TE	Tris-EDTA

DECLARATION OF ACADEMIC ACHIEVEMENT

I optimized the expression and purification of *Saccharomyces cerevisiae* Saw1 using dynamic light scattering and differential scanning fluorimetry to screen buffer components that stabilize the protein. This work has recently been accepted for publication: M. Rashev, J.A. Surtees, A Guarné. Large-scale production of recombinant Saw1 in *Escherichia coli*, *Protein Expression and Purification* (2017), doi: 10.1016/j.pep.2017.02.014. I performed EMSAs to determine the minimal 3′-flap and splayed-Y substrates for Saw1. I performed initial crystallization screens of Saw1 alone and in complex with 3 different splayed DNA substrates. I performed limited proteolysis of Saw1 to determine possible domain boundaries. Amanda Khoo (4th-year Undergraduate Thesis student) optimized expression and purification of C1 Saw1 construct and performed EMSAs with 3′-flap DNA. Eduardo Rojas (3rd-year Undergraduate Thesis student) performed cloning, expression/solubility assays, and preliminary purifications of C2, C3, and C4 Saw1 constructs. Eduardo Rojas also performed the Saw1 fungal homolog alignment and chose which homologs to pursue. He cloned and performed expression/solubility assays for each homolog.

I optimized the expression and purification of *Saccharomyces cerevisiae* Saw1-Rad1-Rad10 complex. This included optimizing the separation of two species with anion exchange, and determining optimal buffers using differential scanning fluorimetry. I also performed the EMSAs to determine the minimal splayed-Y substrate that maintained binding to Saw1-Rad1-Rad10. I developed a method of sample preparation for electron microscopy and viewed and captured images of the complex on a JEOL TEMSCAN microscope (Electron Microscopy Facility, Faculty of Health Sciences, McMaster University).

CHAPTER 1

INTRODUCTION

1.1 DNA damage: Causes and cellular consequences

The integrity of the genome is constantly being challenged by an array of DNA damaging agents. DNA damage can be caused by both exogenous (radiation, genotoxic chemicals) and endogenous (oxidative stress, replication stress, spontaneous hydrolysis and deamination, misincorporation of bases) sources (Hoeijmakers, 2001; Ciccia and Elledge, 2010). This array of genotoxic agents leads to a variety of types of DNA damage, including base modifications, single and double strand breaks (SSB and DSB, respectively), interstrand crosslinks (ICLs), and base mismatches (Hoeijmakers, 2001). The consequences of DNA damage can be sorted into two general groups. The first involves damage that impairs DNA replication and transcription, often resulting in cell cycle arrest and cell death. The second involves damage that results in permanent alterations to the genome, including mutations and chromosomal aberrations, allowing the cell to survive and pass these alterations to the next generation. These permanent alterations are closely linked to the development of cancer, through the activation of oncogenes or inactivation of tumour suppressor genes (Hoeijmakers, 2009). The DNA-damage response (DDR) encompasses a series of mechanisms that cells have evolved to cope with various types of DNA lesions (van Gent *et al.*, 2001; Jackson and Bartek, 2009).

1.2 Repair of base modifications

The DNA can be modified spontaneously (hydrolytic removal of bases, deamination), through reactions with metabolic products (reactive oxygen and nitrogen species (ROS, RNS)),

and through exogenous physical and chemical agents (Hoeijmakers, 2009). When the replication fork encounters unrepaired DNA lesions, it can stall and lead to genomic instability. To avoid deleterious replication fork stalling, more promiscuous polymerases are recruited to bypass the lesion through error-prone DNA synthesis. This mechanism, called translesion synthesis, is a mode of DNA damage tolerance (Zeman and Cimprich, 2014; Chalissery *et al.*, 2017).

Some specialized repair enzymes can directly reverse base modifications. For example, O^6-methylguanine-DNA methyltransferase directly reverses O^6-methylguanine modifications (alkylation of guanine) (Iyama and Wilson, 2013). Other subtle DNA modifications can be repaired through the base excision repair (BER) pathway. In this pathway, the damaged base is removed, creating an abasic site. Then, an incision is made at this site so that a phosphodiesterase can remove the abasic nucleotide. The gap is then filled by a DNA polymerase, and the strand is re-ligated to complete repair (Lindahl, 1993). Single stranded breaks (SSB) can be caused by ROS or as intermediates of repair or other enzyme activity. To repair these breaks, the site is modified to create compatible 3′-OH and 5′-phosphate termini for synthesis and ligation (Iyama and Wilson, 2013).

Nucleotide excision repair (NER) removes bulky, helix-distorting lesions, which involves more extensive excision of several nucleotides, in contrast to BER. These lesions include UV-induced pyrimidine dimers and photoproducts, base adducts caused by chemical agents, and cyclopurines caused by ROS (Iyama and Wilson, 2013). The damage is recognized either through the general surveillance of the genome, called global genome-NER (GG-NER) or during transcription when RNA polymerase stalls upon encountering the lesion, called transcription coupled-NER (TC-NER). After damage recognition, the DNA is unwound around the lesion so that the damaged strand can be excised by endonucleolytic incisions on either side of the lesion.

Finally, a polymerase fills in the gap and the strand is ligated to complete repair (Iyama and Wilson, 2013).

Though rare, replication polymerases can also misincorporate bases during replication, leading to base mismatches or small insertion/deletion loops. The mismatch repair (MMR) pathway removes and corrects this type of damage. First, the mismatch recognition proteins recognize the mismatch and recruit an endonuclease that incises the newly synthesized strand. Exonucleases can then resect and remove the error-containing strand. Finally, a polymerase elongates the strand and a ligase joins the ends to complete repair (Li, 2008; Iyama and Wilson, 2013).

1.3 Double strand DNA breaks (DSB) and repair

Double strand breaks are less frequent than other types of damage (an estimated 25 DSBs per day per cell (Tubbs and Nussenzweig, 2017)), but they are the most deleterious form of damage since both strands are broken and they cannot be easily repaired using a template strand. In the case of a DSB, the free DNA ends can lead to recombination events, resulting in large chromosomal translocations, deletions, insertions, and duplications – rearrangements that affect many genes at once. The genomic instability caused by DSBs often results in the development of a cancerous cell (van Gent *et al.*, 2001). Pathologic DSBs can arise from replication across a single-stranded chromatid break, exposure to reactive oxygen species causing SSBs in close proximity, exposure to ionizing radiation (IR) (including gamma rays and X-rays), inadvertent activity of nuclear enzymes (i.e. incomplete activity of type II topoisomerases), and mechanical stress on DNA (Hoeijmakers, 2001; Frankenberg-Schwager *et al.*, 2008; Ciccia and Elledge, 2010; Lieber, 2010). Cells also induce DSBs for programmed cellular recombination events (i.e.

V(D)J recombination in immune cells and meiotic recombination) (Lieber, 2010). There are several different mechanisms used to repair DSBs, the main pathways being non-homologous end joining (NHEJ), homologous recombination (HR), and single strand annealing (SSA). The pathway choice depends on several factors, including the cell (cell type and cell cycle phase), homologous template availability (sister chromatid or homologous chromosome), and the nature of the broken DNA ends (the state of end resection, competitive protein binding, flanking repeat sequences, chromatin state) (Preston *et al.*, 2002).

1.3.1 Non-homologous End Joining (NHEJ) and Alternative End Joining (Alt-EJ)

In G1 phase, where no sister chromatid is available to use as a template for repair, mammalian cells prefer the NHEJ mechanism of DSB repair, where broken ends are re-ligated. This type of repair has the potential to be mutagenic since processing of the DNA ends may be required to form suitable substrates for re-ligation. (Lieber *et al.*, 2003; Radhakrishnan *et al.*, 2014).

In canonical-NHEJ (c-NHEJ), Ku complexes (made of the Ku70/80 heterodimer) bind to the broken DNA ends to facilitate the processing on each side of the break. The Ku complexes recruit DNA-PKcs and, together, bridge the DNA ends (Radhakrishnan *et al.*, 2014). This complex is central in the recruitment of other NHEJ factors. There are several factors thought to contribute to end processing, and the requirement of each depends on the nature of the broken ends. For instance, artemis is an end-processing factor with hairpin opening ability, and is important in V(D)J recombination (Lieber, 2010; Radhakrishnan *et al.*, 2014). Other enzymes, such as polynucleotide kinases/phosphatases, are involved in end processing to form compatible ends for ligation (Radhakrishnan *et al.*, 2014). Polymerases μ and λ are also involved in end

processing (Radhakrishnan *et al.*, 2014) and have very flexible activity. For example, Pol μ can incorporate rNTPs (in an environment where dNTPs are depleted), transverse discontinuous template strands, and synthesize DNA in a template-independent manner. These activities are thought to allow temporary end joining via short, complementary base-pairing until ligation can occur (Lieber, 2010). Finally, the XLF-XRCC4-DNA ligase IV complex is recruited and performs the ligation step. It is a flexible ligase that has the ability to ligate a range of substrates, including incompatible ends and ligating across gaps (Lieber, 2010) (**Fig 1.1B**).

There is a subset of alt-EJ pathways of repair that are independent of the c-NHEJ factors and often (but not always) rely on homology to tether broken DNA ends (**Fig 1.1C**). For homology-dependent EJ events, end resection is a necessary step to expose complementary regions. In G1, Ku and DNA-PKcs bind and protect DNA ends from resection, whereas, in G2/S, end resection machinery (including the human MRE11-RAD50-NBS1 (MRN) complex, CtIP) out-compete Ku for DSB ends and promote end-resection. This promotes resection-dependent repair mechanisms, including homology-directed repair (described in sections 1.3.2 and 1.3.3) as well as alt-EJ pathways that require some extent of complementary base-pairing (Bennardo *et al.*, 2008) (**Fig 1.1**).

Beyond repair of double-strand breaks, NHEJ plays a significant role in regular cellular activities. During immunoglobulin and T-cell receptor (TCR) generation in B and T lymphocytes, programmed DSBs are induced in regions encoding the antigen-binding portion of immunoglobulin and are repaired with the NHEJ machinery to generate diversity in these regions. Defects in NHEJ machinery not only lead to repair deficiencies, but also to immune-deficiency (Jackson and Bartek, 2009).
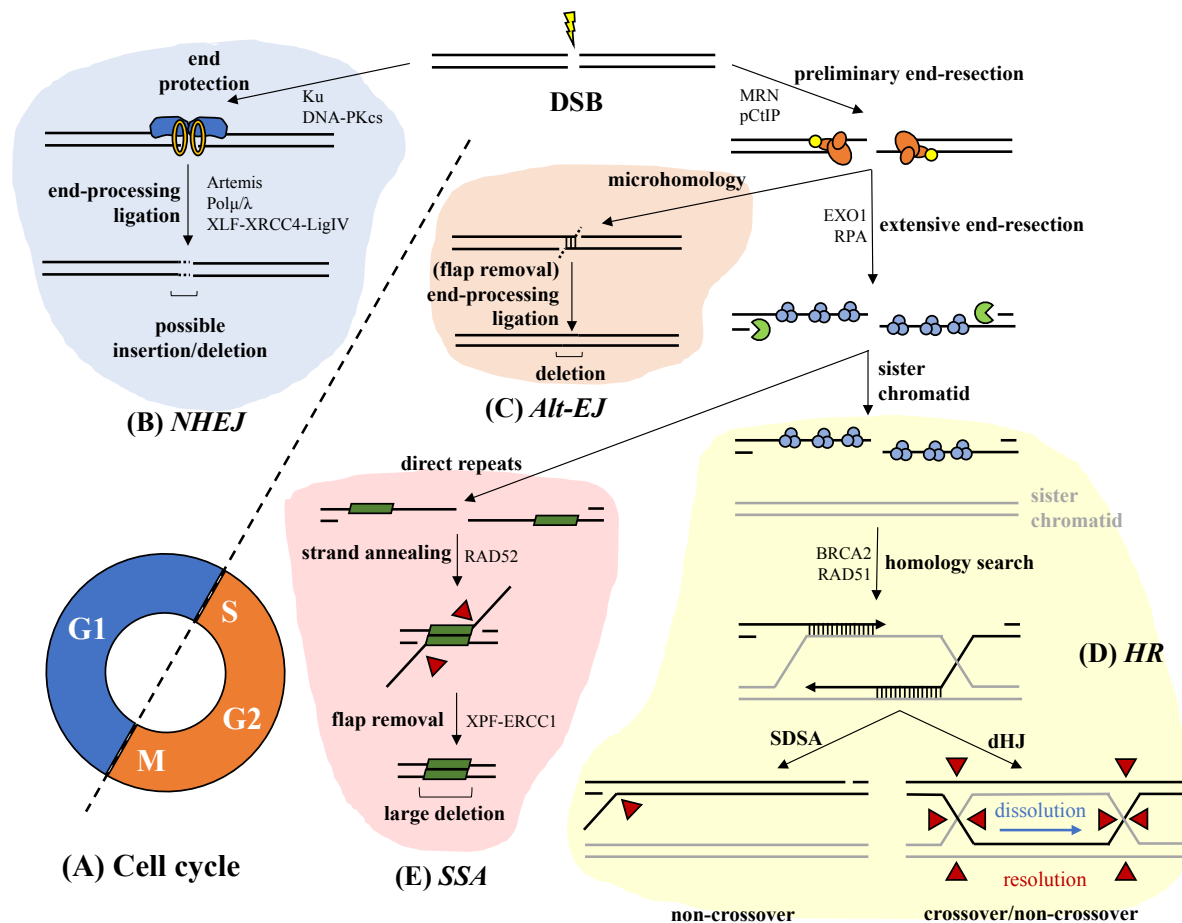
Figure 1.1 *DSB repair pathways and determinants of pathway choice in humans.*
The cell cycle (A) dictates when DSB ends are protected or resected. Protection in G1 by Ku (orange rings) and DNA-PKcs (blue) leads to NHEJ (B), resulting in a possible insertion or deletion at the site of repair. Preliminary end resection in S/G2/M by MRN (orange) and CtIP (yellow) can lead to complementary pairing between terminal microhomology regions, and repair via Alt-EJ (C) (resulting in deletion) or it can lead to more extensive end resection via EXO1 exonuclease (green) (or DNA2 with WRN or BLM helicases) and coating of ssDNA with RPA (blue trimer). If a sister chromatid (shown in grey) is available, BRCA2/RAD51-mediated homology search is carried out and leads to conservative repair via HR (D) which can be resolved through SDSA or HJ-resolution/dissolution. If repeats are exposed during resection (green bars), RAD52-mediated strand annealing can occur, leading to repair via SSA (E) and resulting in a large deletion at the site of repair. Nucleases that resolve joint/junctional molecules are represented by red triangles (ie. dHJ resolution and non-homologous flap removal).

1.3.2 Homologous Recombination (HR) repair

Conversely, a sister chromatid is present in S and G2 phases and is used as an exact template for repair via the Homologous Recombination (HR) repair pathway (**Fig 1.1D**). In this

repair process, broken ends are resected, revealing 3′-single-stranded DNA (ssDNA) which is then used to find a homologous template for repair (Grabarz *et al.*, 2012; Krejci *et al.*, 2012).

End-resection occurs as a two-step process (**Fig 1.1**). The first involves the MRN-CtIP complex, which performs initial resection of 5′ DNA termini at the break, resecting ~200-300 nucleotides (nt), and also helps process complex DNA ends (ex. protein-DNA adducts) (Anand *et al.*, 2016). The second involves EXO1 or DNA2 with WRN or BLM helicases, which perform extensive end resection (Sturzenegger *et al.*, 2014; Symington, 2014). As tracts of ssDNA are exposed, Replication Protein A (RPA), the ssDNA-binding protein in eukaryotes, coats the DNA to prevent secondary structure formation, protect from nucleases, and promote continued end-resection and recruitment of repair factors (Chen *et al.*, 2013; Symington, 2014).

Next, BRCA2-PALB2 recruits RAD51 to ssDNA, promoting nucleofilament formation for strand invasion (Jensen *et al.*, 2010; Ceccaldi *et al.*, 2016). It acts with RAD51 to displace RPA from ssDNA, and is essential for HR (Jensen *et al.*, 2010). Once the presynaptic filament has been formed, RAD51 begins its search for homology, stimulated by the RAD54 translocase (Godin *et al.*, 2016). Homologous pairing ensues and leads to invasion of the acceptor duplex (normally a sister chromatid) by the RAD51-coated filament, forming a D-loop structure. RAD54 removes RAD51 from the DNA, allowing strand synthesis to occur (Symington, 2002; Morrical, 2015).

Homologous recombination can have three different results (**Fig 1.1D**). The first is the canonical HR pathway where, after strand invasion, the other end of the DSB also invades the acceptor molecule in what is termed "second-end capture", possibly by the annealing activity of RAD52 (Jensen *et al.*, 2010). Yeast Rad52 is implicated in second end capture and this suggests an analogous process in humans (Nimonkar *et al.*, 2009). This results in a double Holliday

Junction (HJ), which is resolved by structure selective nucleases (ie. MUS81-EME1, GEN1, SLX1-SLX4) leading to either crossover or non-crossover products (Mazón *et al.*, 2010; Svendsen and Harper, 2010). HJs can also be dissolved by BLM-TOP3-RMI1/2, leading to non-crossover events only (Wu and Hickson, 2003). Second, after strand invasion and synthesis, the invading strand can be displaced and re-annealed with the other end of the resected DSB; an event mediated by the ssDNA annealing activity of RAD52 (Jensen *et al.*, 2010; Morrical, 2015). This is called synthesis-dependent strand annealing (SDSA) and results in non-crossover products. Lastly, break-induced replication (BIR) can occur if the DSB is one ended, where the invading strand is extended until it reaches the end of the chromosome. This pathway can lead to the duplication of an entire chromosome (Symington, 2002; Barlow and Rothstein, 2010; Mazón *et al.*, 2010; Morrical, 2015). BIR may be involved in replication restart when the replication fork encounters a SSB which results in a one-ended DSB (Barlow and Rothstein, 2010).

1.3.3 Single strand annealing (SSA)

There is a third, kinetically distinct DSB repair mechanism called single-strand annealing (SSA) (**Fig 1.1E**). SSA operates primarily in S phase (Frankenberg-Schwager *et al.*, 2009), but it does not require a sister chromatid or homologous chromosome as a template for repair. Instead, direct homologous repeats on either side of the DSB are used to join the DNA ends through annealing of these complementary regions. (Ivanov *et al.*, 1996; Paques and Haber, 1999). It is predominant in highly repetitive regions of DNA, such as ribosomal DNA (Li *et al.*, 2008). SSA repair is an intrachromosomal repair event (compared to HR) (Morrical, 2015) and proceeds through a four-step process: 1) exposing direct repeats through 5′-3′ end-resection by the HR end-processing machinery (MRN-CtIP, EXO1); 2) annealing exposed, complementary repeats

through the single strand-annealing activity of RAD52, forming a branched DNA substrate with 3′ non-homologous flaps; 3) removing 3′ non-homologous flaps through the structure-specific nuclease activity of the XPF-ERCC1 complex; and 4) strand elongation and ligation to complete repair. Notably, SSA always results in a deletion of the sequence between the repeats as well as one of the repeats (Paques and Haber, 1999; Symington, 2002). The steps of SSA are described in detail in section 1.6.

Apart from its role in maintaining genomic integrity, SSA also has implications in the pathogenesis of Herpes Simplex Virus-1 (HSV-1) (Schumacher *et al.*, 2012). HSV-1 is known to manipulate the host DNA repair machinery during infection. In 2012, Schumacher *et al.* showed that HSV-1 infection stimulated host SSA repair where all other repair mechanisms were inhibited. The HSV-1 genome encodes the 5′-3′ endonuclease, UL12, which interacts with several homology-dependent repair factors. Through these interactions, UL12 streams DSB repair away from NHEJ towards HR, and more specifically towards SSA. They suggested that SSA could be implicated in viral DSB repair, concatemerization of the viral genome in the early stages of infection, or in restarting stalled replication forks in an SSA-dependent manner that gives rise to recombinant DNA molecules. It is possible that SSA is also important for the evolution and pathogenesis of HSV-1 by promoting high levels of variation through genetic exchange during infection (Schumacher *et al.*, 2012).

1.4 DSB repair pathway choice

Several factors govern the choice of DSB repair pathway, including the cell type/developmental stage (Fiorenza *et al.*, 2001; Preston *et al.*, 2006), complexity of the DSB and location in the genome, the cell cycle phase, and the kinetics of the respective pathways

(Belov *et al.*, 2015). The pathway preference also differs between organisms. For example, the hierarchy of DSB repair pathway choice in yeast and in mammals differs in that, while yeast prefers HR, mammalian cells prefer NHEJ and also employ HR (Sugawara and Haber, 1992; Pastink *et al.*, 2001). When we compare the genomes of yeast and mammals, we can speculate as to why they differ in pathway preference. The yeast genome is much more compact than that of mammals, and contains fewer introns and repetitive sequences than their mammalian counterparts. Therefore, we would expect yeast to avoid any potentially mutagenic form of repair (SSA and NHEJ) as any alterations to the genome will be more likely to result in a functional mutation. Mammalian genomes, however, are rich in repetitive and non-coding DNA. Therefore, deleterious repair mechanisms are less likely to have a severe effect (favours NHEJ). HR is promoted only when a sister chromatid is present (S and G2 phase) to promote conservative repair (Liang *et al.*, 1998).

When comparing SSA and NHEJ activities in mouse cells at two different developmental stages (dictyate oocytes and early pre-implantation embryos), oocytes had high SSA levels and low NHEJ levels, whereas embryos showed low SSA and high NHEJ levels (Fiorenza *et al.*, 2001). A similar effect is seen during the development of *Drosophila melanogaster* where SSA is a primary repair mechanism for DSB in early stages of development, whereas classical HR and NHEJ were more prevalent in later stages (Preston *et al.*, 2006). Together, these indicate that the stage of development is also a factor for repair mechanism selection.

Throughout the cell cycle, the MRN complex is a major player in the initial steps of DSB recognition and repair. However, its nuclease activity is what drives the choice between c-NHEJ and homology-directed repair. Nuclease-active MRN promotes resection-dependent repair (HR, SSA, alt-EJ), whereas the inactive nuclease promotes c-NHEJ (Grabarz *et al.*, 2012). The

nuclease activity of MRN is controlled by the cell cycle, specifically by CDK1 whose activity increases in S/G2 compared to G1 (Trovesi *et al.*, 2013) (**Fig 1.1A**). CDK1 phosphorylates CtIP, a co-factor of the MRN complex, activating the end-resection activity of the complex and promoting resection-dependent repair (HR, SSA, alt-EJ) (**Fig 1.1C-E**) (Huertas and Jackson, 2009). This restricts HR to phases of the cell cycle where a sister chromatid is available as a perfect template for repair, and avoids gene conversion and loss of heterozygosity during G1. This mechanism of cell cycle-dependent end-resection control is conserved in budding yeast where the CtIP homolog, Sae2, is phosphorylated by CDK (Huertas *et al.*, 2008). During G1, Ku and XRCC4 (required for NHEJ) protect ends from resection by MRN (Bennardo *et al.*, 2009; Grabarz *et al.*, 2012), but are displaced during S/G2 by resection machinery (Grabarz *et al.*, 2012). Modulation of 53BP1 and BRCA1 also promote end protection and end resection, respectively, and their activity is also influenced by the cell cycle phase (Grabarz *et al.*, 2012; Ceccaldi *et al.*, 2016).

The extent of end resection is also a factor in pathway selection. Initial end resection is catalyzed by MRE11 and CtIP nucleases where "end clipping" is performed. This exposes enough micro-homology for annealing through alt-EJ (**Fig 1.1C**). Processive nucleases and helicases perform more extensive resection, promoting HR and SSA by providing larger stretches of homology for annealing (**Fig 1.1DE**). Therefore, the choice between alt-EJ and HR/SSA is dependent on the extent of resection, but this level of repair regulation is not fully understood (Grabarz *et al.*, 2012; Ceccaldi *et al.*, 2016).

The cell favours resection-dependent mechanisms once it enters S phase, but this preference does not directly depend on the presence of a sister chromatid (Saleh-Gohari and Helleday, 2004). The use of classical HR in early S phase before the synthesis of a sister

chromatid can pose the same risks as its use during G1 (loss of heterozygosity via gene conversion). It is therefore possible that the error-prone SSA and alt-EJ pathways play a more significant role in early S phase repair of DSB to avoid such events (Bhargava *et al.*, 2016; Ceccaldi *et al.*, 2016).

The interplay between the factors involved in each of the respective DSB repair pathways determines the choice of pathway. For example, Ku suppresses homology-directed repair pathways, and loss-of-function mutants cause an increase in SSA and HR repair (Stark *et al.*, 2004). The choice between HR and SSA seems to depend on the strand invasion and strand annealing factors, respectively. For strand invasion in HR, RAD51-nucleofilament formation requires BRCA2. In SSA, homologous strand annealing requires RAD52. Both BRCA2 and RAD52 interact with the RPA coated ssDNA produced by extensive end resection. RAD52 depletion in mouse cells attenuates SSA and promotes HR, whereas deficiencies in BRCA2 or RAD51 streamline repair into SSA (Larminat *et al.*, 2002; Stark *et al.*, 2004; Mansour *et al.*, 2008). Similarly, in yeast, Rad52 plays both the roles of human RAD52 and BRCA2, and Rad51-dependent nucleofilament formation prevents the ssDNA annealing activity of Rad52 (Ceccaldi *et al.*, 2016).

Although SSA promotes error-prone repair of DSBs, it plays a role in certain repair contexts. In two different studies, Frankenberg-Schwager *et al.* looked at the contributions of NHEJ, HR, and SSA in DSB repair in G1 or in S phase of CHO cells using H_2O_2 or ionizing radiation (X-rays or alpha particles) (Frankenberg-Schwager *et al.*, 2008; Frankenberg-Schwager *et al.*, 2009). SSA was negligible for repair of H_2O_2 induced damage, but was important for repair of complex DSBs caused by alpha particle irradiation within S phase. Therefore, damage complexity may dictate repair pathway choice.

One must also take into account the genomic context of the break. HR and SSA compete for the same substrate: extensively resected 3′-single stranded tails coated with RPA. When a break occurs between two repeated sequences, the substrate is preferentially repaired through SSA. As the distance between the repeats increases, the proportion of repair events moves towards classical HR repair. This is due to the kinetics of repair, specifically end resection to expose homologous repeats. The longer it takes to commit to SSA, the more likely it is that classical HR repair will take over (Fishman-Lobell *et al.*, 1992; Sugawara *et al.*, 2000). Since repetitive DNA accounts for 10% of the human genome (Jurka, 1998), SSA is likely to be a major pathway of repair in humans (Bhargava *et al.*, 2016).

1.5 Misregulation of DSB repair and its link to cancer and cancer treatment

Repair-deficient cancer cells channel repair events into other, often more error-prone, mechanisms. Since SSA is a mutagenic repair pathway, its promotion can be quite deleterious to the cell. Upregulation of SSA can be caused by mutations in the DSB repair regulation machinery. BRCA2 deficient mammalian cells have increased levels of RAD52-dependent SSA repair of DSB (Lok *et al.*, 2013). Through promotion of error-prone repair, BRCA2 mutations cause the accumulation of genetic mutations that often lead to cancer development and tumour progression (Tutt *et al.*, 2001).

We can take advantage of this channeling of repair into other pathways. Inhibition of these secondary pathways can lead to targeted cancer cell death, termed synthetic lethality. The mechanistic understanding of each repair process, as well as the regulation governing pathway choice, will allow for the discovery of compounds for highly targeted chemotherapy (Ceccaldi *et al.*, 2016).

1.6 Steps of eukaryotic SSA repair of DSBs

SSA has been demonstrated in flies (Do *et al.*, 2014), yeast (Ivanov *et al.*, 1996), plants (Orel *et al.*, 2003), worms (Pontier and Tijsterman, 2009), and mice (Tutt *et al.*, 2001). It has been most thoroughly studied in budding yeast where the major factors have been identified and their roles elucidated (Bhargava *et al.*, 2016). The human system is known to involve the human homologs of Rad52 (RAD52) for annealing, and Rad1-Rad10 (XPF-ERCC1) for 3′-flap removal, but is otherwise still highly uncharacterized (Ceccaldi *et al.*, 2016).

1.6.1 End-resection and Rad52-dependent annealing of homologous repeats

The first step in all homology-directed pathways of repair is 5′-3′ resection of the broken DNA ends. This is a process that is highly similar in both yeast and mammalian cells. Therefore, any factors that may promote or inhibit HR at the stage of DSB-recognition and resection (described in the sections above) will also promote or inhibit SSA. In yeast, the MRX (Mre11-Rad50-Xrs2) complex (MRN in humans) detects DSBs and catalyzes initial 5′-3′ resection of the DNA ends, followed by more processive resection by Exo1 (EXO1 in humans). This step is shared between SSA and conservative HR (Symington and Gautier, 2011) (**Fig 1.2**).
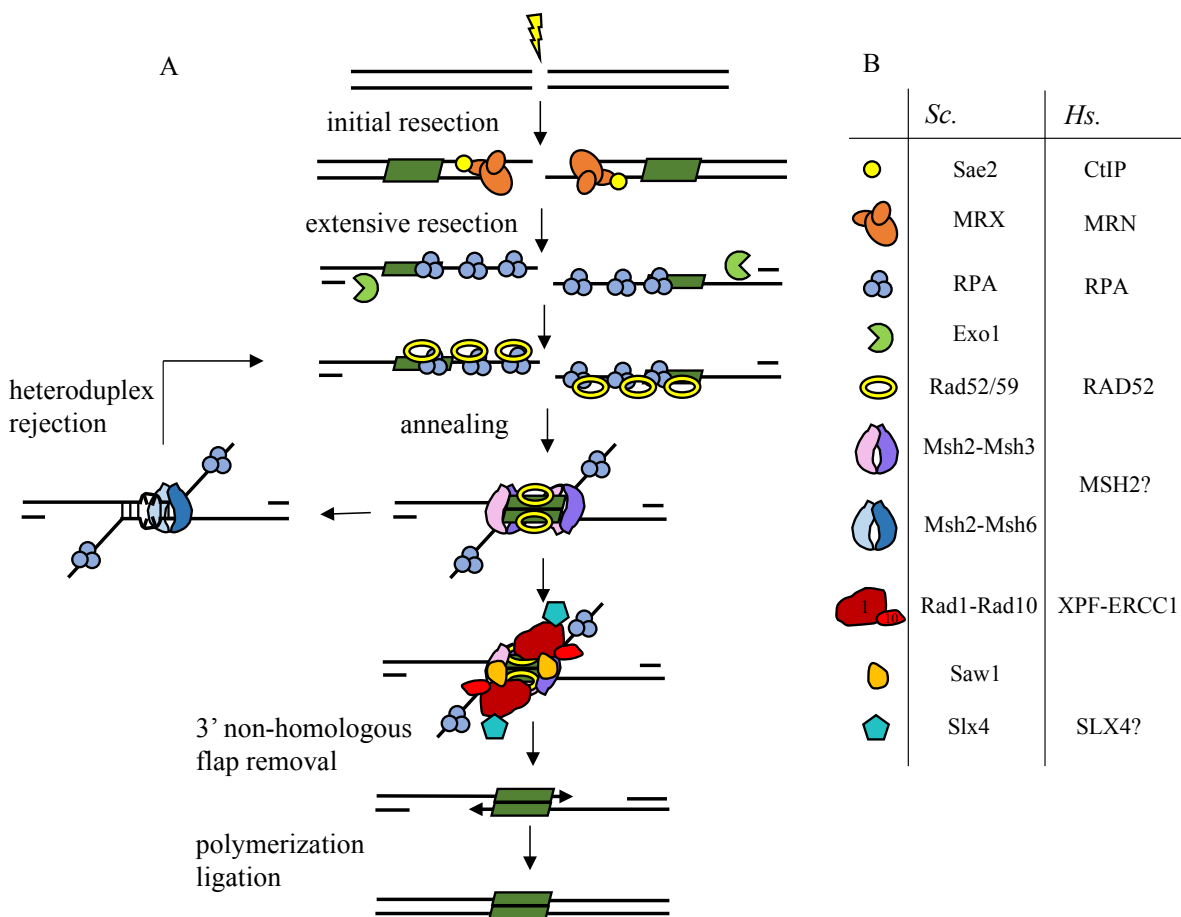
Figure 1.2 *Model of yeast SSA repair of a DSB.*
(A) Steps of *Saccharomyces cerevisiae* (*Sc.*) SSA, including initial and extensive resection, Rad52-mediated annealing of complimentary strands, duplex stabilization/heteroduplex rejection by MMR factors, non-homologous flap removal, and polymerization and ligation to complete repair. Direct homologous repeats are represented by green boxes. All known factors involved in yeast SSA are pictured. (B) Names of yeast (*Sc.*) factors as well as the human (*Hs.*) homologs known to participate in human SSA.

In both yeast and mammals, Rad52 is a ssDNA-binding protein that promotes complementary strand annealing (Mortensen *et al.*, 1996; Davis and Symington, 2001; Sugiyama and Kantake, 2009). Rad52 interacts with RPA (which binds to exposed ssDNA) to stimulate homologous strand annealing during SSA (**Fig 1.2**) (Sugiyama *et al.*, 1998; Ceccaldi *et al.*, 2016). SSA in yeast also requires a second strand-annealing factor, Rad59. Rad59 has ssDNA-

binding affinity as well as strand annealing ability (like Rad52) but is not dependent on RPA (unlike Rad52) (Petukhova *et al.*, 1999; Davis and Symington, 2001).

Although the yeast and mammalian Rad52 are structurally and biochemically similar, they differ in that human RAD52 mutant cells are viable whereas the yeast mutation is lethal. This is thought to be because BRCA2 in humans mediates RAD51-dependent strand invasion during classical HR, and human RAD52 only promotes ssDNA annealing between complementary sequences (Jensen *et al.*, 2010). The yeast Rad52 has both of these functions and is therefore vital for yeast survival (Lok and Powell, 2012). In mice, disruption of both RAD51 and BRCA2 result in growth retardation, genomic instability, sensitivity to genotoxic agents, and early embryonic lethality (Lim and Hasty, 1996; Tsuzuki *et al.*, 1996; Abbott *et al.*, 1998; Patel *et al.*, 1998; Moynahan *et al.*, 2001)

The crystal structure of the single-strand annealing domain of human RAD52 (N-terminus of the protein) revealed an undecameric (11)-subunit ring with a large positively charged groove running along the surface. This suggests that the ssDNA is bound by RAD52 with its bases pointing away from the surface, presenting the bases for annealing to complementary ssDNA (Singleton *et al.*, 2002; Grimme *et al.*, 2010). This activity and structure are shared with single-strand annealing proteins in other organisms, namely RecT in *E. coli*, and β protein in bacteriophage λ (Van Dyck *et al.*, 2001). Extensive complementary base-pairing occurs through a collision mechanism and the favourable energy change causes the release of dsDNA (Sugawara *et al.*, 2000; Singleton *et al.*, 2002; Grimme *et al.*, 2010).

SSA efficiency and pathway choice greatly depend on the length of the homologous repeats as well as the distance between them. In yeast, when sequences of homology on either side of the break are short (1-2 kilobase pairs (kbp)), Rad52 is essential for annealing the strands.

However, with longer regions of homology, Rad52 becomes dispensable for SSA. Repeats as short as 29 base pairs (bp) can be annealed by the cooperative efforts of both Rad52 and Rad59, the latter having a larger role with smaller repeats (Sugawara *et al.*, 2000).

There is evidence of a microhomology-mediated end joining (MMEJ) pathway that requires SSA factors for the early steps of repair. Studies in budding yeast showed a requirement for Rad52- and Rad59-mediated annealing of very small microhomology regions, termed micro-SSA. This process is distinct from SSA, as it still requires end-joining factors for the later steps of repair. The HR end processing machinery (EXO1, MRE11) is also involved in micro-SSA (Decottignies, 2013).

1.6.2 Mismatch Repair factors in SSA

Once the heteroduplex is formed during SSA, the mismatch-recognition machinery from the mismatch repair pathway is responsible for, first, ensuring recombination does not take place between divergent sequences, and second, stabilizing the heteroduplex for downstream processing. (**Fig 1.2**) (Sugawara *et al.*, 2004; Goldfarb and Alani, 2005). During mismatch repair, the Msh2-Msh6 complex recognizes single base mismatches, whereas the Msh2-Msh3 complex recognizes large insertion and deletion loops. After mismatch recognition, both complexes recruit downstream mismatch repair factors to complete repair (Johnson *et al.*, 1996).

In yeast, a 3% sequence divergence over 205 bp repeats decreases SSA repair 6-fold compared to repair with identical repeats (Sugawara *et al.*, 2004). This is due to heteroduplex rejection, which is the process that prevents annealing of imperfect repeats and allows for further homology searching. Heteroduplex rejection is thought to be mediated by the mismatch-recognition protein Msh2-Msh6, as well as the action of the Sgs1 helicase (Sugawara *et al.*,

2004; Goldfarb and Alani, 2005). During SSA, it is thought that this complex recognizes mismatches and recruits the Sgs1 helicase to unwind the annealed duplex (Goldfarb and Alani, 2005).

Conversely, the mismatch-recognition Msh2-Msh3 complex stabilizes the annealed, branched substrate, and is necessary for the removal of 3′ non-homologous flaps (Sugawara *et al.*, 1997; Surtees and Alani, 2006). Msh2 interacts with the Rad1-Rad10 nuclease, which is responsible for 3′-flap removal during SSA (Bertrand *et al.*, 1998; Lyndaker and Alani, 2009). In SSA, Msh2-Msh3 binds specifically to the ds-ssDNA junctions of the annealed heteroduplex with 3′-flaps. Once bound, it changes the conformation of the branched substrate, possibly allowing for downstream processing of the flaps (Surtees and Alani, 2006). The requirement of Msh2-Msh3 for SSA depends on the length of the annealed sequences, where it is required for short annealed sequences (250 bp) and dispensable for long ones (1 kb). This reinforces the hypothesis that Msh2-Msh3 contributes to the 3′-flap removal complex by stabilizing the annealed duplexes (Sugawara *et al.*, 2000; Li *et al.*, 2013). This contribution may also be conserved in the mammalian system as an interaction between the Rad10 human homolog (ERCC1) and human MSH2 has been detected (Lan *et al.*, 2004).

1.6.3 Removal of 3′ non-homologous flaps and completion of SSA

The next step of SSA is the removal of 3′ non-homologous flaps that form through strand annealing. In yeast, this is carried out by the Rad1-Rad10 complex, which has structure selective nuclease activity towards 3′-single stranded flaps at ds-ssDNA junctions (**Fig 1.2**) (Davies *et al.*, 1995; Ivanov and Haber, 1995). This complex is required for both SSA and HR, but only during recombination events that produce long regions of 3′ non-homology (Ivanov and Haber, 1995).

In mammalian cells, the human homolog of Rad1-Rad10 (XPF-ERCC1) is responsible for cleavage of 3′ non-homologous ends in both SSA and in the synthesis-dependent strand annealing mechanism of gene conversion (Al-Minawi *et al.*, 2008).

In SSA, yeast Rad1-Rad10 only transiently associates with its substrate, and requires another protein to target it to repair intermediates. Saw1 (Single strand Annealing Weakened 1) was discovered through a microarray-based screen for SSA genes (Li *et al.*, 2008) and was later found to target Rad1-Rad10 to 3′-flap substrates and promote its nuclease activity (**Fig 1.2**) (Li *et al.*, 2013). Saw1 is a structure specific DNA-binding protein that binds 3′-flap substrates and forms a stable complex with Rad1-Rad10 through an interaction with Rad1. A Rad1-binding motif has been identified on the N-terminus of Saw1 through mutational analysis (Li *et al.*, 2008; Li *et al.*, 2013), but details of the physical interaction remain unknown. Currently, a stretch of 6 positively charged residues in its C-terminus forms a putative DNA-binding motif (conserved across fungal homologs) that, when mutated, abolishes DNA-binding. *In vitro,* it has slight affinity towards ssDNA and tailed substrates, and high affinity for splayed-Y, 3′- and 5′ flap, and replication fork like structures. Its affinity towards junctional substrates also depends on the length of the flap. Saw1 had significantly reduced binding to a 3′-flap substrate with a 20-nt flap (compared to a 30-nt flap) and had no affinity when the flap was less than 10 nt (Li *et al.*, 2013). It also interacts with other SSA intermediates, namely Rad52 and Msh2-Msh3. (Li *et al.*, 2008; Li *et al.*, 2013).

Motycka *et al.* (2004) showed that human XPF-ERCC1 formed a very stable ternary complex with human RAD52, interacting through the N-terminal domains of RAD52 and XPF. Formation of this complex in human cell extracts promoted the nuclease activity and also decreased the DNA annealing activity of RAD52, indicating that this complex functions to

recruit the nuclease to the site of repair for cleavage of non-homologous 3′-flaps as well as to attenuate the strand annealing activity of RAD52 (Motycka *et al.*, 2004). Although there is no human homolog of the yeast Saw1, recruitment of Saw1 (and thus Rad1-Rad10) to SSA intermediates in yeast was found to be dependent on Rad52 (Li *et al.*, 2013). Therefore, it is possible that Saw1 interacts with yeast Rad52 to recruit Rad1-Rad10 to sites of repair, similar to the interaction between XPF and RAD52 in humans.

Saw1 is also implicated in repair in other damage contexts including base lesions, protein-DNA adducts, and UV lesions. The first two require the canonical role of Saw1 in recruiting Rad1-Rad10 to sites of repair. However, in the UV damage context, Saw1 plays a Rad1-Rad10-independent role in repair of these lesions. In this context, the SUMOylation of Saw1 promotes its association with the Slx1-Slx4 structure specific nuclease. Its exact role in the repair of UV induced damage is still unknown, but it seems as though Saw1 contributes to several damage contexts as a versatile and programmable scaffold (Sarangi *et al.*, 2014).

In yeast, Slx4 plays an important role in Rad1-Rad10-mediated 3′-flap cleavage of SSA intermediates, independent of formation of the Slx1-Slx4 nuclease complex. It is not necessary for recruitment, but it stimulates the endonuclease activity of Rad1-Rad10 (**Fig 1.2**) (Li *et al.*, 2013). Slx4 is phosphorylated upon the induction of DSBs and therefore could link SSA to DNA damage checkpoint control (Toh *et al.*, 2010). In humans, SLX4 interacts with the XPF-ERCC1 complex, as is the case in yeast. Its role of 3′-flap cleavage stimulation could therefore be conserved in humans (Lyndaker *et al.*, 2008). The mechanism of action of Slx4 stimulation of Rad1-Rad10 activity in both yeast and mammals has yet to be elucidated and could reveal another layer of repair regulation/pathway choice following a DSB.

After hydrolysis, a SUMO E3 ligase SUMOylates Rad1-Rad10 on a lysine residue outside of its DNA-binding and nuclease domains, which alters its DNA binding affinity. This allows Rad1-Rad10 to dissociate from the repair site and continue repairing other DNA lesions, improving the efficiency of repair. Similarities between the human and yeast nucleases suggest that XPF-ERCC1 is regulated in a similar manner (Sarangi *et al.*, 2014).

After dissociation, a polymerase extends the resulting 3′-OH using the complementary strand as a template and a ligase joins the ends to complete repair. The factors involved in the gap fill-in synthesis and ligation steps are still unknown (Bhargava *et al.*, 2016). As seen in **Fig 1.2**, the steps in this pathway lead to a deletion of one of the repeats as well as the intervening sequence between the repeats (Symington, 2002).

1.7 XPF/MUS81 family proteins

Various DNA-repair pathways require processing of branched repair intermediates and, like XPF-ERCC1 homologs, nucleases with structure-selective activity towards these branched substrates are responsible for this processing. Many 3′-flap nucleases belong to the XPF/MUS81 family of proteins, including XPF-ERCC1 and MUS81-EME1 homologs, and are found in eukaryotes and archaea (**Fig 1.3**). Nucleases in this family are necessary for various DNA repair pathways and deficiencies in these proteins lead to increased risk of cancer and premature ageing (Ciccia *et al.*, 2008).

1.7.1 Role of XPF-ERCC1 and its homologs in multiple repair pathways

Rad1-Rad10 and XPF-ERCC1 were initially discovered as factors involved in nucleotide excision repair (NER) of UV-induced DNA lesions that distort the DNA helix. In this process,

after damage recognition and repair bubble formation, Rad14 targets Rad1-Rad10 to NER repair sites through an interaction between Rad1 and Rad14 (Guzder *et al.*, 2006). In humans, XPF-ERCC1 also requires an adapter protein for recruitment in NER, but the interaction is between the human homolog of Rad10 (ERCC1) and the human homolog of Rad14 (XPA) (Park and Sancar, 1994; Tripsianes *et al.*, 2007). Once recruited, the nuclease incises the lesion-containing strand on the 5′ side of the damage, and Rad2 (XPG in humans) incises the 3′ side (Davies *et al.*, 1995; Park and Choi, 2006). This dual incision releases the damaged fragment, allowing for strand elongation and completion of repair. Patients with XPF mutations present with Xeroderma pigmentosum (XP) where symptoms include photosensitivity, freckling, and enhanced risk of skin cancer, which are all typical of defective NER of UV lesions (Gregg *et al.*, 2011).

Mutations in XPF that affect the localization of the complex (Ahmad *et al.*, 2010), or mutations in ERCC1 that impair complex formation (Jaspers *et al.*, 2007; Faridounnia *et al.*, 2015) can result in more severe phenotypes including neurological deterioration, accelerated aging, and disorders such as cerebro-oculo-facio-skeletal syndrome (Gregg *et al.*, 2011). Mice homozygous for a truncation mutation in XPF have developmental defects, die within 3 weeks of birth, and show sensitivity to various DNA-damaging agents (Tian *et al.*, 2004). These severe phenotypes are evidence of the role of XPF-ERCC1 outside of NER, where the complex protects the cell against chromosomal aberrations through its activity in homology-directed repair.

Rad1 and Rad10 mutants only result in HR deficiencies when long regions of non-homology are involved in the recombination event (Schiestl and Prakash, 1988; Schiestl and Prakash, 1990; Fishman-Lobell and Haber, 1992). A similar phenotype was seen with ERCC1 mutants (Adair *et al.*, 2000; Ahmad *et al.*, 2008; Al-Minawi *et al.*, 2008). This is evidence that, not only is this nuclease necessary for homology-directed repair via SSA, but is necessary

whenever 3′ non-homologous flaps are produced during recombination, and acts on these substrates as it does during SSA. These types of substrates can arise during the SDSA pathway of HR, where long-tract synthesis can produce 3′-flap structures when the newly synthesized strand re-anneals with its original partner. These flaps need to be removed before repair can be completed (Al-Minawi *et al.*, 2008) (see **Fig 1.1DE**). In yeast, Saw1 is required for Rad10 recruitment to flap substrates generated during SDSA as well as for efficient removal of flaps during SDSA repair (Diamante *et al.*, 2014). This indicates that the nuclease activity of XPF-ERCC1 (or Rad1-Rad10) depends on the recognition of this specific branched DNA structure rather than being specific to certain repair contexts.

Cells deficient in XPF-ERCC1 are hypersensitive to cross-linking agents, as this complex plays a central role in interstrand crosslink (ICL) repair (De Silva *et al.*, 2000). The yeast homologs Rad1 and Rad10 have also been implicated in the repair of ICLs (Hodskinson *et al.*, 2014). ICLs are caused by chemicals (chemotherapeutics, products of dietary metabolism and cigarette smoke) that form covalent linkages between bases on opposite strands of the DNA duplex. This type of DNA damage is extremely genotoxic as it impairs all DNA metabolism (transcription, replication). In humans, the Fanconi anaemia (FA) pathway repairs these types of DNA lesions. This is done through the concerted efforts of multiple proteins and is primarily active during S-phase at converging replication forks. Defects in these proteins lead to the FA disease (Ceccaldi *et al.*, 2016). Once the core FA complex detects the lesion, it recruits structure selective nucleases, including XPF-ERCC1, that incise one of the strands on either side of the lesion. These incisions "unhook" the cross-link from the other strand. Replication of the strand containing the lesion continues using translesion synthesis polymerases and the DSB produced through unhooking is repaired via homologous recombination (Ceccaldi *et al.*, 2016).

In all of these cases (nucleotide excision repair, single strand annealing, and synthesis-dependent strand annealing), XPF-ERCC1 and its homologs act on branched substrates with the aid of other repair factors specific to each repair mechanism (Bennardo *et al.*, 2009).

1.7.2 Domain distribution of XPF-ERCC1 homologs

The proteins in the XPF/MUS81 family all share a highly-conserved domain structure. Nuclease-active family members carry the conserved ERCC4 nuclease motif GDX_nERKX_3D, which is the active site of the family. A structure of the homodimeric archaeal Hef nuclease from *Pyrococcus furiosis* revealed that the nuclease site has a highly similar fold to that of type II restriction endonucleases, which hydrolyze DNA through a 2-metal ion mechanism (Nishino *et al.*, 2003; Nowotny and Gaur, 2016).

XPF-ERCC1 and its eukaryotic and archaeal homologs contain the signature domain structure of an N-terminal Superfamily 2 helicase domain, a central (nuclease/ERCC4) domain, and two tandem C-terminal Helix-hairpin-Helix $((HhH)_2)$ DNA-binding domains (**Fig 1.3**). In archaea, the helicase domain is functional, whereas those belonging to the human and yeast orthologs are inactive. Some archaea possess only the nuclease and HhH domains, along with a PCNA-interacting motif at their C-terminus (Ciccia *et al.*, 2008). These helicase/PCNA-interacting motifs are thought to contribute to the ability of these archaeal XPFs to process fork-like structures. The archaeal XPF homologs function as homodimers and have catalytic selectivity towards 3′ flap and replication fork structures, similar to the substrate preference of vertebrate MUS81-EME1. The eukaryotic homologs function as heterodimers of a catalytic and non-catalytic subunit (Ciccia *et al.*, 2008). The inactive helicase domains in eukaryotic XPF homologs are thought to contribute to cleavage efficiency but not to selectivity (Tsodikov *et al.*,

2005).
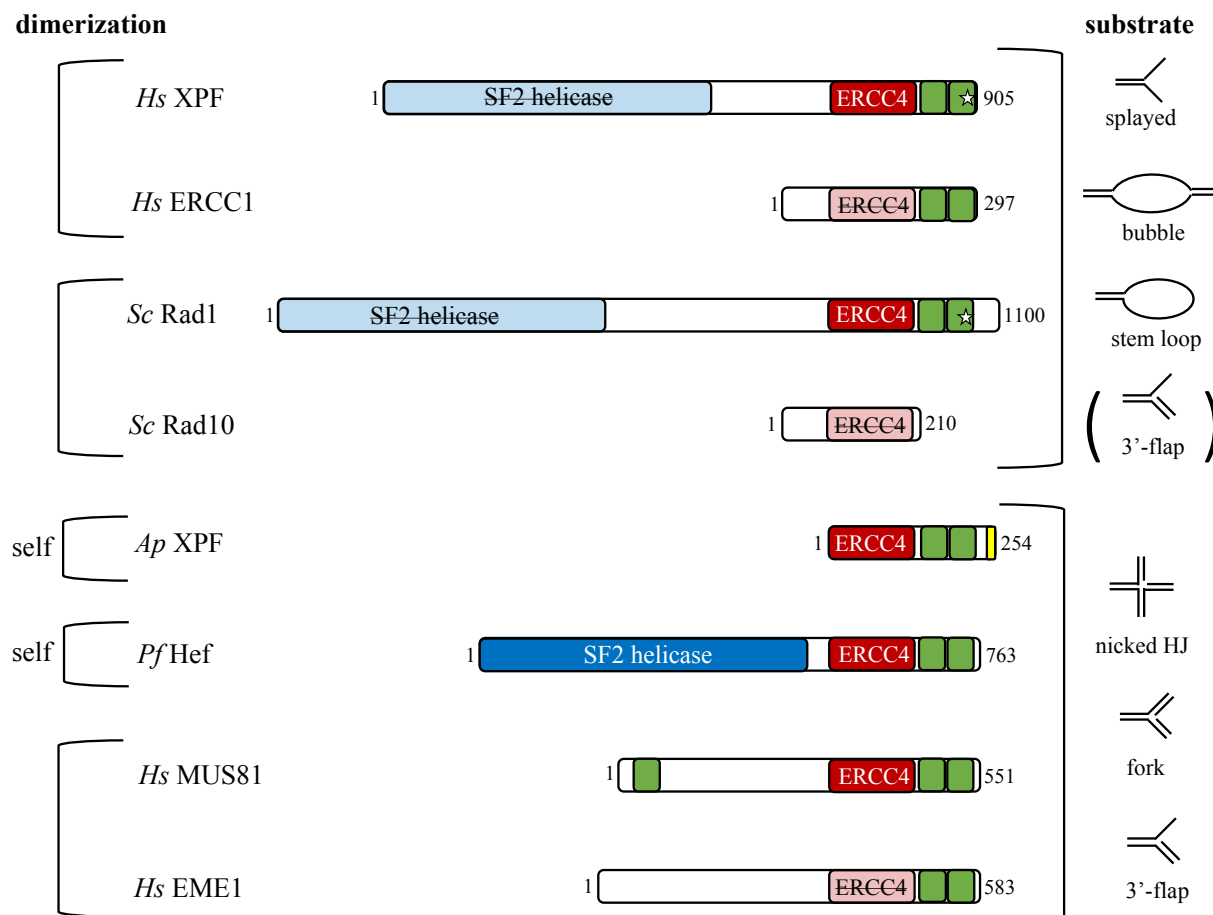
Figure 1.3 *Domain distribution of XPF/MUS81 family of structure-specific nucleases.*
XPF/MUS81 family members from *Homo sapiens (Hs), Saccharomyces cerevisiae (Sc), Aeropyrum pernix (Ap),* and *Pyrococcus furiosis (Pf)*. Central nuclease/ERCC1 domains containing the ERKX$_3$D active site are shown as red boxes, while inactive domains are shown in light red with strikethrough text. Super family II helicase domains are shown in blue, while inactive helicase domains are shown in light blue with strikethrough text. Helix-hairpin-Helix (HhH) domains are shown in green. A star in the HhH domain represents the missing residue that produces a Helix-turn-helix (as seen in *Hs* XPF). The PCNA-interacting motif of *Ap* XPF is shown in yellow. Preferred substrates are shown on the right, and dimerization partners are shown on the left.

In eukaryotes, the Rad10/ERCC1 partners to these nuclease-active subunits share high sequence similarity with the nuclease domain and HhH domains of Rad1/XPF, although they contain a disrupted nuclease domain and are therefore the non-catalytic subunit of the complex. These proteins are also missing the helicase-like domain of the XPF/Rad1 subunits.

1.7.3 Catalytic complex of an XPF/MUS81 family member: MUS81-EME1

The most complete picture of structure selectivity of XPF/MUS81 family members is of the human MUS81-EME1 nuclease in complex with a 3′-flap DNA substrate (Nowotny and Gaur, 2016). MUS81-EME1 is one of the nucleases responsible for HJ resolution and also plays roles in ICL repair and meiosis (Ciccia *et al.*, 2008). The heterodimeric complex contains the same domains as the XPF and Rad1 proteins, although in a different order (**Fig 1.3**). The active MUS81 protomer contains a C-terminal $(HhH)_2$ domain as well as a single HhH domain at its N-terminus, and a C-terminal ERCC4/nuclease domain (Fadden *et al.*, 2013). The non-catalytic EME1 protomer has a single functional HhH domain at its C-terminus, and an inactive ERCC4/nuclease domain.

The structure shows the upstream and downstream duplex-interacting surfaces and the positioning of the incision point of the substrate at the active site of MUS81. MUS81-EME1 bends the DNA duplex on either side of the discontinuity through interactions with the HhH domains. Upon DNA-binding the HhH domains move 40º relative to the nuclease domains. Both of these actions position the 3′ flap within the active site of the MUS81 nuclease domain, and also position the 5′-end after the discontinuity into the hydrophobic wedge. This wedge contains a pocket for the 5′ end that is proximal to the incision point in 3′-flap and nicked-HJ structures. Altogether, this structure shows how MUS81-EME1 selects for 3′-flaps and nicked-HJs and not 5′-flaps or splayed substrates, which are incompatible with this structure (**Fig 1.4A**) (Gwon *et al.*, 2014).

1.7.4 Substrate specificity of XPF-ERCC1 homologs: Archaea

The archaeal XPF homodimers share substrate specificity with the MUS81-EME1 nuclease (**Fig 1.3B**) (Ciccia *et al.*, 2008) and they seem to share many of the same features that dictate this selectivity. Archaeal XPF homodimers were the first subjects of structural investigation of the molecular determinants of complex formation and structure specificity of XPF homologs. A structure of Hef nuclease (*Pyrococcus furiosis*) ERCC4-(HhH)$_2$ homodimer showed that it dimerizes through independent interactions between both the ERCC4 domains and the (HhH)$_2$ domains of each protomer (Nishino *et al.*, 2003). While only one nuclease site is required for cleavage, both (HhH)$_2$ domains are required for DNA binding. By mapping where various parts of a forked DNA substrate made contact with selected residues on the protein, Nishino *et al.* (2005) constructed a model of DNA-binding. This model showed how Hef nuclease bent the DNA between the two (HhH)$_2$ domains, which placed the junction of the DNA substrate at the catalytic site of one of the protomers (**Fig 1.4B**).

The structure of another archaeal homodimeric XPF homolog, this one from *Aeropyrum pernix* in complex with duplex DNA, further solidified these findings. The duplex interacted with the (HhH)$_2$ of one protomer and, through the apparent symmetry of the (HhH)$_2$ domains of both protomers, they modeled-in a second duplex onto the second protomer. By modeling a link between one strand of both duplexes, they speculated that, like Hef nuclease, this XPF homodimer could bend a nicked DNA substrate so that the discontinuity would be placed near the active site of one of the protomers. Upon DNA-binding, the nuclease domain closes (in an asymmetrical form) onto the junction or discontinuity of the DNA substrate (**Fig 1.4B**) (Newman *et al.*, 2005). DNA structures with ss-dsDNA junctions are good substrates for nucleases within this family as they are susceptible to distortion. Although no structures exist of the catalytic

complex, the apparent bending of the DNA on either side of the discontinuity, as well as the conformational change upon DNA-binding, indicates that archaeal XPF homodimers share similar substrate selection mechanisms as the eukaryotic MUS81-EME1 nuclease.
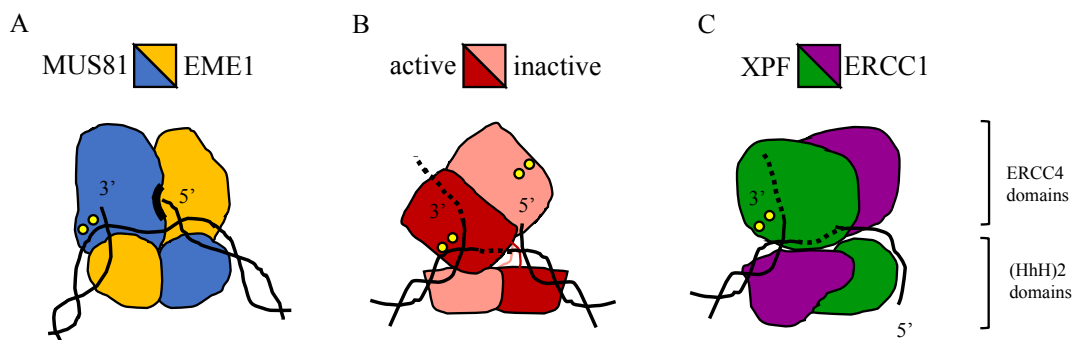
Figure 1.4 *Simplified models of substrate selectivity of XPF/MUS81 family nucleases*
(A) Model of substrate recognition by human MUS81-EME1 showing bending of the DNA substrate at the junction, coordination of the 5′ terminus (post-nick), and placement of the pre-nick 3′-terminus at the active site of MUS81 (dark blue) through a large conformational change. (B) Hypothetical model of 3′-flap substrate recognition of archaeal XPF homologs showing potential DNA bending at the junction and placement of the active site of one of the protomers at the incision site of the pre-nick 3′-terminus. (C) Hypothetical model of possible splayed substrate recognition by human XPF-ERCC1 showing how the substrate may be specifically coordinated both upstream and downstream of the junction. The active site is shown by the two yellow circles. Dotted lines represent proposed, interpolated positions of nucleic acid.

1.7.5 Substrate specificity of XPF-ERCC1 homologs: Humans

The substrate preference of eukaryotic XPF-ERCC1 differs from the above nucleases, where it prefers splayed-Y substrates over 3′-flaps or nicked HJs (Ciccia *et al.*, 2008). Several groups have examined structures of fragments of human XPF and ERCC1 to understand the substrate selectivity of this nuclease. Several structures of human XPF-ERCC1 (HhH)$_2$ dimerization domains (Tripsianes *et al.*, 2005) and ERCC4-(HhH)$_2$ domains (Tsodikov *et al.*, 2005) have been published, showing that the heterodimer forms through extensive hydrophobic interactions within the (HhH)$_2$ domains of each protomer, similar to the archaeal homodimers (de

Laat *et al.*, 1998). The presence of XPF is required as a scaffold for proper ERCC1 folding, which explains the instability of either protomer in the absence of the other (Sijbers *et al.*, 1996; Tripsianes *et al.*, 2005). Tsodikov *et al.* (2005) saw no stable interaction between the central domains, unlike the archaeal homodimers. The groups that published these structures also characterized the *in vitro* DNA-binding of the complex, but the results of these studied disagreed. Tripsianes *et al.* (2005) found that ERCC1 was the sole contributor to DNA binding and that XPF did not show any affinity for a stem-loop substrate. However, Tsodikov *et al.* (2005) proposed a model where the $(HhH)_2$ domains of both protomers recognized ssDNA of a splayed substrate (with the ERCC1 $(HhH)_2$ domain preferentially recognizing the 5′ overhang) and this placed the catalytic site of the XPF central domain at the site of incision. The ERCC1 specificity for ssDNA reflects the original finding that Rad10 binds ssDNA preferentially over dsDNA (Sung *et al.*, 1992).

Later, Tripsianes *et al.* (2007) found that the ERCC1 central domain has ssDNA binding activity. They suggested that the ERCC1 $(HhH)_2$ domain coordinates the upstream duplex and the central domain plays a role in coordinating the downstream ssDNA which places the junction in the active site of the XPF central domain (Tripsianes *et al.*, 2007). This is similar to the archaeal models where the $(HhH)_2$ domains of each protomer coordinate both the upstream and downstream duplexes of a junctional substrate. However, the role of XPF in DNA coordination remained elusive.

Most recently, the role of human XPF in DNA coordination was uncovered through structural analysis of the human XPF homodimer (Das *et al.*, 2008) bound to ssDNA (Das *et al.*, 2012). In the human heterodimeric structure of the $(HhH)_2$ domains of XPF-ERCC1, the highly-conserved hairpin structure within the 2nd HhH motif of ERCC1 (necessary for contacting the

minor groove of duplex DNA) was replaced by a turn in XPF through a single missing residue within this hairpin region (Tripsianes *et al.*, 2005; Das *et al.*, 2008). Rather than abolishing DNA binding, Das *et al* (2012) showed that this turn simply changed the specificity of XPF from dsDNA to ssDNA. They determined the solution structure of the ERCC4-(HhH)$_2$ domains of human XPF homodimers bound to two 10 nt oligonucleotides. The (HhH)$_2$ domains of both protomers bound the ssDNA fragments in the same manner, making contacts with the phosphate backbone of the ssDNA. This gave the ssDNA an extended conformation – one that could not occur with dsDNA (Das *et al.*, 2012). Based on the archaeal and human XPF homodimeric structures, they proposed a model of substrate recognition. The (HhH)$_2$ of the ERCC1 domain, which is structurally analogous to that of the archaeal XPF (HhH)$_2$ domains, recognizes the upstream duplex region, and the XPF (HhH)$_2$ domain and central domain of ERCC1 recognize the downstream ssDNA of a splayed substrate (Tripsianes *et al.*, 2007; Das *et al.*, 2012). XPF (HhH)$_2$ was also found to prefer binding to the 5′ overhang or, in the context of NER, the undamaged strand (Das *et al.*, 2012; Das *et al.*, 2016) (**Fig 1.4C**).

This was a very exciting advance in our understanding of the substrate selectivity of XPF-ERCC1 and its homologs. However, we still do not understand the mechanism of polarity selection of a splayed substrate and we have a limited understanding of the protein-DNA interactions near the cleavage site/junction of these molecules.

1.7.6 Budding yeast homologs of XPF-ERCC1: Rad1-Rad10

Like XPF-ERCC1, the yeast Rad1-Rad10 form a highly stable and specific complex (Bailly *et al.*, 1992). Bardwell *et al.* (1992) initially mapped this interaction to the C-termini of both Rad1 and Rad10 (outside of the HhH motifs), and later defined the hydrophobic regions on

each protein that were responsible for the interaction (residues 90-210 of Rad10, residues 809-997 of Rad1) (Bardwell *et al.*, 1993). Inhibition of this interaction through mutations in Rad1 results in a UV sensitivity phenotype similar to *rad1Δ* and *rad10Δ* mutants, showing that the interaction is necessary for the repair function of these proteins (Bailly *et al.*, 1992; Siede *et al.*, 1993). This difference in complex formation between the XPF-ERCC1 and Rad1-Rad10 complexes is most likely due to the fact that, unlike other eukaryotic homologs, Rad10 is missing the C-terminal tandem HhH domains typical of this family (**Fig 1.3**) (van Duin *et al.*, 1986; Aravind *et al.*, 1999). However, Rad1 maintains the conserved domain distribution seen in XPF, including the one missing nucleotide that converts the hairpin of the 2nd HhH domain into a turn (**Fig 1.3**) (Das *et al.*, 2008).

A second difference between the human and yeast homologs is that Rad1-Rad10 requires the additional DNA-binding activity of Saw1 for efficient recruitment to sites of SSA repair (Li *et al.*, 2008; Li *et al.*, 2013). Saw1 depletion in yeast prevents recruitment of Rad1-Rad10 to sites of repair and prevented 3′-flap removal. Though Saw1-dependent recruitment of Rad1-Rad10 to 3′ flaps is necessary for 3′ flap removal, the mechanism of this recruitment is still unknown.

The Rad1-Rad10 complex specifically cleaves branched substrates with 3′ flaps or tails, including NER bubble-like structures, splayed substrates, and has a lower affinity for duplex DNA with 3′ flaps (but not structures with 5′ flaps, tails, or HJs). It incises within the duplex region of the junctional substrate, +2 to the ds-ssDNA junction, always on the strand with the 3′ tail (Bardwell *et al.*, 1994; Davies *et al.*, 1995; Rodriguez *et al.*, 1996). Cleavage by Rad1-Rad10 produces a 3′-OH which primes the strand for elongation by a polymerase. This is necessary for both NER and recombination repair after removal of the 3′ flap (Bardwell *et al.*, 1994). Despite the differences in dimerization of the yeast and human complexes, both homologs share the same

conserved nuclease site residues, have the same substrate preference, and have the same *in vivo* activity (**Fig 1.3**) (Schiestl and Prakash, 1988; Schiestl and Prakash, 1990; Sijbers *et al.*, 1996; Sargent *et al.*, 2000; Chan and Schiestl, 2009).

1.8 Thesis Objective and Rationale

So far, we only have a fragmented view of substrate recognition/binding and cleavage by XPF-ERCC1 homologs. As of yet, there is no structure of the catalytic complex with a proper branched substrate, and we continue to construct mechanistic models from partial views of this protein complex.

In light of the differences between the human and yeast nucleases, they maintain the same substrate preference and endonucleolytic activity (Ciccia *et al.*, 2008). It would therefore be interesting to see how yeast and humans evolved these differences yet still maintain the same substrate preference and nuclease activity. It is possible that, since Saw1 has structure-specific 3′-flap DNA binding affinity, it may share a similar DNA binding mode with the human XPF-ERCC1 nuclease, or may lead us to a structurally or functionally analogous protein in the human SSA system. Since there are no structures of the yeast Saw1-Rad1-Rad10 complex, its characterization would be of use in understanding this divergence. The yeast system is an attractive model since, not only is it highly similar to the human system, but it also allows for more genetic, biochemical, and cellular biology studies. SSA has also been most extensively studied in yeast.

In order to study these repair factors, we need to produce large quantities of stable, pure, and homogeneous recombinant protein. This is not always straightforward, and optimization is often required at both the expression and purification levels of recombinant protein production.

Therefore, the first objective of this thesis was to explore the conditions in which recombinant Saw1, which is normally unstable when not in complex with Rad1-Rad10 (Li *et al.*, 2013), can be produced with high purity, stability, and at high concentrations, while maintaining its structure-specific substrate affinity (Chapter 3). The second objective of this thesis was to perform preliminary characterization of recombinant Saw1-Rad1-Rad10 complex by exploring conditions that enhance the stability and homogeneity of the complex (Chapter 4). Establishing effective large-scale production methods for these proteins will allow for future robust biochemical characterization of this essential repair complex.

CHAPTER 2

MATERIALS AND METHODS

2.1 His-Saw1 expression

The plasmid for Saw1 expression was kindly provided by Dr. Jennifer Surtees, Department of Biochemistry, School of Medicine and Biomedical Sciences, University of Buffalo, Buffalo, NY, USA. The *SAW1* gene was encoded on the expression plasmid pAG8815 with a pET28a backbone, flanked by *BamHI* and *XhoI* restriction sites. The vector allows for expression of the Saw1 protein (residues 1-261) with a removable N-terminal His-tag (6 Histidine residues) with a thrombin cleavage site and contains the Kanamycin resistance gene. To test solubility, $CaCl_2$ competent BL21 Star pRARE, BL21-CodonPlus, and ArcticExpress pRARE *Escherichia coli* cells were transformed with the pAG8815 plasmid and transformed cells were selected for in LB media. Cells were grown to OD_{600} at 37°C with aeration and expression was induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and incubated at 30°C (4 hours), 20°C (6 hours), and, in the case of ArcticExpress pRARE, 12°C (16 hours). Auto-induction of Saw1 was performed with BL21 Codon+ cells to take advantage of leaky expression of Saw1 in the pET expression system using an extended incubation time (63 hours at 30°C). Protein solubility was tested by lysing cell pellets (resuspended in 20 mM Tris-HCl pH 8.0, 1.4 mM β-mercaptoethanol (βME)) on ice with 1 mg/mL lysozyme for 25 minutes, followed by lysis with 90 mM KCl, 10 mm $MgCl_2$, 0.05% lauryldimethylamine-oxide (LDAO) for 30 minutes. The lysate was then treated with 20 units of DNaseI (1U/μL, Thermo Scientific) for 20 minutes at room temperature to shear chromosomal DNA. Lysates were then clarified by

centrifugation. Soluble proteins in the supernatant were analyzed by 15 % SDS-PAGE, alongside samples of total protein before and after induction.

For large-scale expression of soluble Saw1, BL21 Star pRARE *E. coli* cells transformed with pAG 8815 were grown to an OD_{600} of 0.4 at 37°C with agitation, induced with 0.5 mM IPTG and incubated for 4 hours at 20°C. Pelleted cells from 1L preparations of His-Saw1 were washed with PBS and stored at -80°C until ready to be purified.

2.2 Initial purification of His-Saw1

Pellets were resuspended on ice in 20 mL Ni Buffer (50 mM Tris-HCl pH 8, 500 mM NaCl, 5 mM β-mercaptoethanol (βME), 10 mM imidazole, 10% glycerol), a protease inhibitor cocktail (157 μg/mL Benzamidine, 174 μg/mL PMSF, 0.7 μg/mL Pepstatin A, and 5 μg/mL Leupeptin), and 0.05% LDAO. Cells were disrupted by sonication with twenty-second pulses. The cell lysate was clarified by centrifugation at 39,191 g and the supernatant applied to a 5-mL HiTrap^{TM} Chelating HP column (GE Healthcare) charged with $NiCl_2$ and equilibrated with Ni Buffer. After washing with Ni Buffer, and then with Ni Buffer supplemented with 53.5 mM imidazole, His-Saw1 was eluted with a gradient to 300 mM imidazole. Fractions containing His-Saw1 were collected and pooled, diluted to adjust the salt concentration to 200 mM NaCl, and finally loaded onto a Mono S 5/50 GL cation exchange column (GE Healthcare). The column was washed with Mono S buffer (50 mM Tris-HCl pH8, 0.5 mM EDTA pH8, 5 mM βME, 10% glycerol) supplemented with 200 mM NaCl and His-Saw1 was eluted with a gradient to 500 mM NaCl. Fractions containing His-Saw1 were pooled and concentrated to <50 μL with a 10 kDa molecular weight (MW) cut-off concentrator (PES membrane, Sartorius).

NaOH washes of the concentrator membrane were performed as follows: 50 µL 1M NaOH was pipetted onto the concentrator membrane (washed 3X with 200 µL storage buffer) and left for 10 minutes at room temperature. The NaOH was recovered and neutralized with concentrated HCl. 20 µL of this membrane wash was applied to a 15% SDS-PAGE gel.

2.3 pH screen with Dynamic Light Scattering

All Dynamic Light Scattering experiments performed with a Zetasizer NanoS with a 633-nm laser. The pH screen was based on the screen described by Jancarik *et al.* (2004). For screening pH, 1.7 µL of protein at 0.3 mM was diluted 1/15 with 23.3 µL of 1.07X buffer to a final buffer composition of 50 mM Buffering agent, 50 mM Arginine-Glutamate (RE) mix, 500 mM NaCl, 0.1 mM EDTA, 10% glycerol, 1.4 mM βME. Samples were incubated for 1 hour at $4^\textbf{o}$C, centrifuged at 1570 g for 10 minutes at $4^\textbf{o}$C, and 18 µL used for DLS measurements. DLS measurements were done at $4^\textbf{o}$C, with 120 second equilibration time, attenuator set to maximum (max = 11), manual number of runs set to 30. The buffers tested include K_2HPO_4 (pH 5.0), K_2HPO_4 (pH 6.0), MES (pH 6.0), Tris-HCl (pH 7.0), HEPES (pH 7.0), Tris-HCl (pH 7.5), HEPES (pH 7.5), Tris-HCl (pH 8.0), Tris-HCl (pH 8.5), CHES (pH 9.0).

For screening additives, His-Saw1 in MES buffer (50 mM MES pH6, 50 mM Arginine/Glutamate mixture (Arg/Glu), 500 mM NaCl, 1.4 mM βME, 0.1 mM EDTA, 10% glycerol) was diluted with an appropriate MES buffer-based solution to produce the desired concentrations of the additive components, in a final sample volume of 25 µL. All other buffer components remain the same as MES buffer unless otherwise indicated. Samples were incubated for 3 hours at $4^\textbf{o}$C, centrifuged at 1570 x g, 10 minutes, and subjected to DLS measurement. The resulting PDI for Saw1 in each additive is represented as a difference from the PDI of the

corresponding dilution control (His-Saw1 diluted in MES buffer) in **Fig 3.5**. Saw1 in 5 mM DTT, 5 mM βME, 10 mM βME, 0.4 M NaCl, 0.3 M NaCl was prepared by diluting 15 μL of 151 μM His-Saw1 in the required buffer (final protein concentration of 90.6 μM). Saw1 in 5% glycerol was prepared by diluting 12.5 μL of 151 μM His-Saw1 in the required buffer (final protein concentration of 75.5 μM). Saw1 in 0.2 M NaCl was prepared by diluting 10 μL of 151 μM His-Saw1 in the required buffer (final protein concentration of 60.4 μM).

The polydispersity index (PDI) is determined through the cumulants analysis of the DLS measurement. This is the equation of a polynomial fit to the log of the G1 correlation function, where the PDI is equal to $2c/b^2$: $Ln[G1] = a + bt + ct^2$

2.4 Differential Scanning Fluorimetry (*Thermofluor* assay)

Thermofluor assay conditions were prepared in 96-well PCR plates as described by Boivin *et al.* (2013). Plates were pre-chilled before 2 μL of 50 μM His-Saw1 and 2 μL of 62X stock of SYPRO-Orange dye were simultaneously added at 4°C. The plate was immediately placed in a BioRad Thermocycler CFX96 (Centre for Microbial Chemical Biology, McMaster University) and subjected to a temperature gradient from 4-95°C, increasing at a rate of 1°C/minute. Fluorescence intensity was measured with a SYBR-Green filter, and melting curves were generated and grouped into pH screen, salt screen, etc. The concentrations of buffer components are described in **Table 2.1**. Boxes shaded in blue contain 50 mM of the indicated buffer. Boxes shaded in orange contain 50 mM of the indicated buffer and 0.25 M NaCl. Buffers shaded in grey also contain 50 mM MES pH 6.0 and 0.25 M NaCl. Boxes shaded in yellow also contain 50 mM MES pH 6.0. Boxes shaded in green also contain 50 mM Tris-HCl pH 8.0. Abbreviations are as follows: β-mercaptoethanol (β-ME), Arginine (R), Glutamate (E).

Melting temperatures of Saw1 (**Table 3.1**) were determined using the first derivative of the reported melt curves. The temperature at which the derivative of the melting curve is at a minimum value is reported as the T_m., but if two points were found flanking the apex of the peak, the median temperature of these two data points was taken as the T_m.

Table 2.1: *Layout of 96-well DSF Thermofluor buffer screen for Saw1*

	1	2	3	4	5	6	7	8	9	10	11	12
A	Water	Citrate pH 4.0	Sodium Acetate pH 4.5	Citrate pH 5.0	MES pH 6.0	Potassium Phosphate pH 6.0	Citrate pH 6.0	BisTris pH 6.5	Sodium Cacodylate pH 6.5	Sodium Phosphate pH 7.0	Potassium Phosphate pH 7.0	HEPES pH 7.0
B	MOPS pH 7.0	Ammonium Acetate pH 7.3	Tris-HCl pH 7.5	Sodium Phosphate pH 7.5	HEPES pH 8.0	Tris-HCl pH 8.0	Tris-HCl pH 8.5	CHES pH 9.0	0.01M MES pH 6.0	0.05M MES pH 6.0	0.1M MES pH 6.0	0.25M MES pH 6.0
C	0.25M NaCl	Citrate pH 4.0	Sodium Acetate pH 4.5	Citrate pH 5.0	MES pH 6.0	Potassium Phosphate pH 6.0	Citrate pH 6.0	BisTris pH 6.5	Sodium Cacodylate pH6.5	Sodium Phosphate pH 7.0	Potassium Phosphate pH 7.0	HEPES pH 7.0
D	MOPS pH 7.0	Ammonium Acetate pH 7.3	Tris-HCl pH 7.5	Sodium Phosphate pH 7.5	HEPES pH 8.0	Tris-HCl pH 8.0	Tris-HCl pH 8.5	CHES pH 9.0	1.6 μM ssDNA	1.6 μM splayed-Y	1.6 μM flap	1.6 μM fork
E	0.1 M NaCl	0.2 M NaCl	0.35 M NaCl	0.5 M NaCl	0.7 M NaCl	1 M NaCl	0.1 M NaCl	0.2 M NaCl	0.35 M NaCl	0.5 M NaCl	0.7 M NaCl	1 M NaCl
F	0.05 M Imidazole	0.1 M Imidazole	0.25 M Imidazole	0.5 M Imidazole	50 mM R	50 mM RE	0.4 M RE	50 mM E	1% glycerol	5% glycerol	10% glycerol	20% glycerol
G	1 mM TCEP	5 mM TCEP	10 mM TCEP	1 mM 2-ME	5 mM 2-ME	10 mM 2-ME	1 mM DTT	5 mM DTT	10 mM DTT	0.1 mM EDTA	1 mM EDTA	5 mM EDTA
H	10 mM MgCl$_2$	10 mM CaCl$_2$	1 mM MnCl$_2$	1 mM NiCl$_2$	1 mM FeCl$_2$	1 mM ZnCl$_2$	1 mM CoCl$_2$	0.1M LiCl	0.1M NaCl	0.1M KCl	0.1M NH$_4$Cl	MES pH 6.0, 0.25 M NaCl

2.5 Revised purification of His-Saw1

Pellets were resuspended in Ni Buffer (50 mM BisTris pH6.5, 500 mM NaCl, 5 mM βME, 10 mM imidazole, 5% glycerol), including a protease inhibitor cocktail (157 μg/mL Benzamidine, 174 μg/mL PMSF, 0.7 μg/mL Pepstatin A, and 5 μg/mL Leupeptin), and 0.05% LDAO. Samples were lysed by sonication and the lysates clarified by centrifugation at 39,191 g. The supernatants were applied to a 5-mL HiTrap^{TM} Chelating HP column (GE Healthcare)

charged with NiCl$_2$ equilibrated with Ni Buffer. After washing with 24.5 mM imidazole, Saw1 was eluted with a gradient up to 300 mM imidazole. Fractions containing His-Saw1 were pooled, diluted to 250 mM NaCl, and finally loaded onto a Mono S 5/50 GL cation exchange column (GE Healthcare). The column was washed with Mono S buffer (50 mM BisTris pH 6.5, 0.5 mM EDTA, 5 mM βME, 5% glycerol) supplemented with 250 mM NaCl and His-Saw1 was eluted with a NaCl gradient to 1 M NaCl. Fractions containing eluted His-Saw1 were collected, pooled, and concentrated with a 10 kDa MW cut-off concentrator (PES membrane, Sartorius). Purified His-Saw1 was applied to a Superdex 75 10/300 GL size exclusion column (GE Healthcare) equilibrated with Saw1 storage buffer (50 mM BisTris pH 6.5, 0.1 mM EDTA, 300 mM NaCl, 5 mM βME, 5% glycerol). Size exclusion chromatography fractions containing His-Saw1 were pooled and concentrated with a 10 kDa MW cut-off concentrator (PES membrane, Sartorius).

2.6 DNA substrate preparation

Lyophilized oligonucleotides (IDT DNA) were dissolved in 50 μL of sterile water. DNA substrates were annealed by mixing purified oligonucleotides at a 1:1.2 molar ratio of labeled (5′ 6-Carbofluorescein (6-FAM) label, Biobasic Inc.; or 5′ P^{32} labeled DNA, Perkin Elmer) to unlabeled (IDT DNA) oligonucleotide, adding 10X annealing buffer (100 mM Tris-HCl pH 7.5, 500 mM NaCl) to a final concentration of 1X, heating the mixture in 1 L of boiling water for 5 minutes, and slowly cooling to room temperature (or 4°C for very short regions of complementarity) overnight. Substrates were prepared as described in **Table 2.2**. Sequences of each oligonucleotide are shown in **Table 2.3**. All oligonucleotides were designed based on original oligonucleotides used by Li *et al.* (2013) for their DNA-binding assays (MR1, MR2, MR3, and MR4 are Oligonucleotides 1, 2s, 3s, and 4 from this study). Sequences are written 5′ -

3′. Nucleotides in brackets [] in **Table 2.3** are excluded when used as a substrate for crystallization or for EM (ie. only used in DNA-binding assays). The asterisk represents where the radioisotope label or fluorescent label was added for DNA-binding assays. Structures of each substrate are shown in **Fig 2.1**.

Table 2.2: *Preparation of DNA substrates from annealed oligonucleotides*

Substrate	Oligos annealed
60-mer ssDNA	MR1
60-mer dsDNA	MR1+MR5
60-mer splayed-Y	MR1+MR4
60-mer 3′-flap	MR1+MR3+MR4
60-mer fork	MR1+MR2+MR3+MR4
40-mer splayed-Y	MR6+MR7
30-mer splayed-Y	MR8+MR9
10-mer splayed-Y	MR10+MR11
40-mer 3′-flap	MR12+MR13 (MR13 forms hairpin)
30-mer(10nt) 3′-flap	MR14+MR15 (MR15 forms hairpin)
20-mer 3′-flap	MR16 (hairpin structure)

Table 2.3: *Sequence of oligonucleotides used to prepare DNA substrates*

Name	Sequence (5' → 3')
MR1	*GACGCTGCCGAATTCTACCAGTGCCTTGCTAGGACATCTTTGCCCACCTGCAGGTTCACCC
MR2	TGGGTGAACCTGCAGGTGGGCAAAGATGTCC
MR3	CATGGAGCTGTCTAGAGGATCCGACTATCGA
MR4	ATCGATAGTCGGATCCTCTAGACAGCTCCATGTAGCAAGGCACTGGTAGAATTCGGCAGCGT
MR5	GGGTGAACCTGCAGGTGGGCAAAGATGTCCTAGCAAGGCACTGGTAGAATTCGGCAGCGT
MR6	*[G]ATTCTACCAGTGCCTTGCTAGGACATCTTTGCCCACCTGC
MR7	ATCCTCTAGACAGCTCCATGTAGCAAGGCACTGGTAGAAT
MR8	*[G]ACCAGTGCCTTGCTAGGACATCTTTGCCCA
MR9	CTAGACAGCTCCATGTAGCAAGGCACTGGT
MR10	*[G]TGCCTTGCTAGGACATCTTT
MR11	CAGCTCCATGTAGCAAGGCA
MR12	*AATTCTACCAGTGCCTTGCTAGGACATCTTTGCCCACCTGC
MR13	CATGGAGCTGTCTAGAGGATTTTATCCTCTAGACAGCTCCATGTAGCAAGGCACTGGTAGAAT
MR14	*AATTCTACCAGTGCCTTGCTAGGACATCTTT
MR15	CATGGAGCTGTTTCAGCTCCATGTAGCAAGGCACTGGTAGAAT
MR16	*CATGGAGCTGTTTCAGCTCCATGGAGCAAGGCATTTTGCCTTGCTCGGACATCTTT

* 5′-radioisotope or fluorescent label
[] nucleotide included in EMSAs only
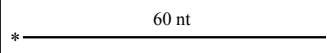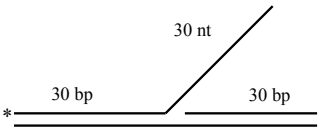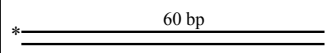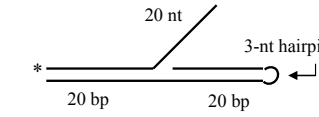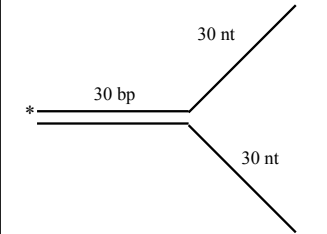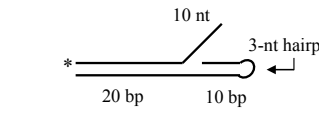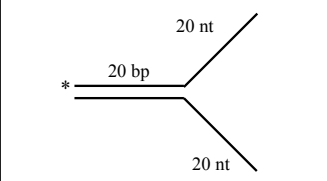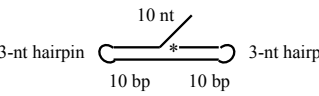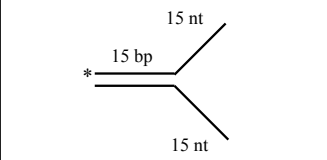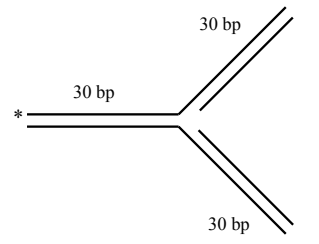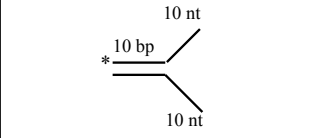
Substrate Name	Substrate Structure	Substrate Name	Substrate Structure
60-mer ssDNA		60-mer 3′-flap	
60-mer dsDNA			
60-mer splayed-Y		40-mer 3′-flap	
40-mer splayed-Y		30-mer(10nt) 3′-flap	
30-mer splayed-Y		20-mer 3′-flap	
20-mer splayed-Y		60-mer fork	

Figure 2.1: *DNA substrates*
Name and structure of all DNA substrates. Length of each arm of the branched substrates, either in nucleotides (nt) for ssDNA or base-pairs (bp) for dsDNA, is indicated. The top strand is read 5′-3′, left to right respectively. The position of the radioisotope or fluorescent label is indicated with an asterisk (*).

2.7 Electrophoretic Mobility Shift Assays

In 15 μL reactions, 10 nM labeled DNA substrate was incubated with a range of Saw1 concentrations, from 10-280 nM, in EMSA reaction buffer (20 mM Tris-HCl pH 8.0, 50 mM KCl, 1 mM DTT, 5 mM EDTA, 1.5 μM BSA, 15% glycerol) at 4°C for 30 minutes. These reactions were then loaded onto 6%Tris-Glycine (TG) polyacrylamide gels and electrophoresed on ice for 60 minutes at 80V with cold 1X TG running buffer. All gels were imaged through fluorescence imaging with a Typhoon Trio+ (GE Healthcare, Centre for Microbial Chemical Biology, McMaster University).

2.8 Salt additive screen for solubilization of Saw1-DNA mixtures for crystallography

This assay is based on one described by Pryor *et al.* (2012). In 24-well crystallization trays, 2 μL of precipitated Saw1-DNA complex and 2 μL of S75 buffer (50 mM BisTris pH 6.5, 5% glycerol, 300 mM NaCl, 0.1 mM EDTA, 5 mM β-mercaptoethanol) supplemented with 10, 50, or 100 mM of each additive were mixed and incubated for 2.5 hours at 4°C. The extent of precipitation was examined under a microscope and scored relative to a drop containing no additive. Additives included LiCl, NaCl, KCl, $MgCl_2$, $CaCl_2$, Li_2SO_4, $MgSO_4$, $LiCH_3COO$, $NaCH_3COO$, KCH_3COO, $Mg(CH_3COO)_2$, NaCHOO, $Mg(CHOO)_2$. Final concentrations of these additives were 5, 25, and 50 mM within the drop.

2.9 Proteolysis of His-Saw1 with trypsin and IMAC of digestion products

Stocks of trypsin were prepared by 1:2 serial dilutions of a 200 μg/mL stock with Tris-EDTA buffer (TE). For small-scale digestion, 4 μL purified His-Saw1 at 12.2 μM was mixed with 1 μL 50 mM $MgCl_2$ and 1 μL trypsin (at 1.6, 3.1, 6.3, 12.5, 25, 50, 100, and 200 μg/mL),

and topped-up to 10 μL with buffer (50 mM BisTris pH 6.5, 0.3M NaCl, 5 mM βME, 5% glycerol). Samples were incubated for 1 hour at room temperature. Digestion was attenuated with 10 μL SDS-PAGE loading buffer and boiling for 8 minutes. Degradation products were analyzed with 15% SDS-PAGE. A final concentration of 0.63 μg/mL trypsin was selected for large scale digestion of His-Saw1, where digestion was performed as described, with each component volume scaled-up ~300X. Here, digestion was attenuated with 157 μg/mL Benzamidine and injected onto a 1-mL Ni-Column (HiTrap^{TM} Chelating HP column) equilibrated with imidazole-free Ni Buffer. Digestion products were eluted and collected by washing with Ni Buffer supplemented with 30 mM, 60 mM, and 300 mM imidazole.

2.10 Expression of Saw1-(His-Rad1)-Rad10

The plasmid encoding the Rad1-Rad10 complex, and the plasmid encoding tagless Saw1 were kindly provided by Dr. Jennifer Surtees (Department of Biochemistry, School of Medicine and Biomedical Sciences, University of Buffalo, Buffalo, NY, USA). Full-length Rad1 and Rad10 (pAG 9068) were cloned in the multicloning site of pET15b (Ampicillin resistance), each with an independent T7 promoter and ribosome-binding site, with flanking *NdeI* (5'-end of Rad1) and *BamHI* (3'-end of Rad10) restriction sites. Full length, tagless Saw1 (pAG 9069) was cloned in a pET28a plasmid (Kanamycin resistance) using the *NcoI* and *HindIII* restriction sites. His-Rad1-Rad10 and Saw1 were co-expressed in BL21 Star pRARELysS, grown to an OD$_{600}$ of 0.6, induced with 0.5 mM IPTG and incubated for 6 hours at 20°C. Pelleted cells from 1L preparations of Saw1-Rad1-Rad10 were washed with PBS and stored at -80°C until ready to be purified.

2.11 Purification of Saw1-(His-Rad1)-Rad10

Cell pellets containing Saw1-Rad1-Rad10 were resuspended in 20 mL Ni Buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mM βME, 10 mM imidazole, 5% glycerol), including a protease inhibitor cocktail (157 µg/mL Benzamidine, 174 µg/mL PMSF, 0.7 µg/mL Pepstatin A, and 5 µg/mL Leupeptin), and 0.05% LDAO. Cells were disrupted by sonication with 30-second pulses. The lysate was centrifuged at 39,191 g and supernatant applied to a 5-mL Ni-Column (HiTrap^{TM} Chelating HP column) equilibrated with Ni Buffer. After washing with 59.3 mM imidazole, Saw1-Rad1-Rad10 was co-eluted with a gradient to 300 mM imidazole. Fractions containing eluted Saw1-Rad1-Rad10 were collected, pooled, and loaded onto a Mono Q 5/50 GL anion exchange column. The column was washed with MonoQ buffer (50 mM Tris-HCl pH 8.0, 5 mM βME, 0.1 mM ETDA, 5% glycerol) supplemented with 150 mM NaCl and Saw1-Rad1-Rad10 was eluted with a NaCl gradient to 500 mM NaCl. Fractions containing the <u>first</u> eluted Saw1-Rad1-Rad10 peak were collected and pooled, and concentrated with a 100 kDa MW cut-off concentrator.

2.12 Differential Scanning Fluorimetry (*ProteoPlex Assay*)

All steps were performed at 4ºC unless otherwise specified. 16 µL of 1.54 mg/mL 40mer splayed DNA-Saw1-Rad1-Rad10 (set at a ratio of 1:1.2, protein:DNA and incubated at 4ºC overnight) was aliquoted into the wells of a 96-well PCR plate (BioRad). 2 µL of 1M buffer stocks and 2µL of 100X SYPRO Orange dye were pipetted onto the side of the well and added simultaneously to the protein-DNA sample by centrifugation at 1000 g. The plate was immediately placed in a BioRad Thermocycler CFX96 (Centre for Microbial Chemical Biology, McMaster University) and subjected to a temperature gradient from 4-95°C, increasing at a rate

of 1°C/minute. Fluorescence intensity is read out with a FRET filter. The assay was performed as described previously (Chari *et al.*, 2015). **Table 2.4** describes the layout of the 96-well *ProteoPlex* screen where the final concentration of each condition is 100 mM. The long-range buffers (SPG, MMT) were prepared as follows: 1 M SPG: 0.125 M Succinic acid, 0.5 M NaH$_2$PO$_4$, 0.375 M Glycine. 1M MMT: 0.2 M DL-Malic acid, 0.4 MES monohydrate, 0.4 M Tris. Both were pH-adjusted with either concentrated HCl or NaOH. Grey boxes indicate empty wells. A11, B11, C11, and D11 contain controls where 18 µL of the indicated sample and 2 µL of 100X SYPRO-Orange were mixed.

Table 2.4: *Layout of 96-well DSL Proteoplex buffer screen for DNA-Saw1-Rad1-Rad10.*

	1	2	3	4	5	6	7	8	9	10	11
A	SPG pH 5.6	MMT pH 5.6	MES pH 5.5	BisTris pH 5.8	Citrate pH 5.5	Sodium Phosphate pH 5.8	Potassium Phosphate pH 5.8	HEPES pH 6.8	Imidazole pH 6.2	Tris-HCl pH 7.5	Complex in Storage Buffer
B	SPG pH 6.0	MMT pH 6.0	MES pH 5.7	BisTris pH 6.0	Citrate pH 5.75	Sodium Phosphate pH 6.1	Potassium Phosphate pH 6.1	HEPES pH 7.0	Imidazole pH 6.45	Tris-HCl pH 7.7	Saw1-Rad1-Rad10 in Storage Buffer
C	SPG pH 6.4	MMT pH 6.4	MES pH 5.9	BisTris pH 6.2	Citrate pH 6.0	Sodium Phosphate pH 6.4	Potassium Phosphate pH 6.4	HEPES pH 7.2	Imidazole pH 6.7	Tris-HCl pH 7.9	20-mer DNA in Storage Buffer
D	SPG pH 6.8	MMT pH 6.8	MES pH 6.1	BisTris pH 6.4	Citrate pH 6.25	Sodium Phosphate pH 6.7	Potassium Phosphate pH 6.7	HEPES pH 7.4	Imidazole pH 6.95	Tris-HCl pH 8.1	Storage Buffer
E	SPG pH 7.2	MMT pH 7.2	MES pH 6.3	BisTris pH 6.6	Citrate pH 6.5	Sodium Phosphate pH 7.0	Potassium Phosphate pH 7.0	HEPES pH 7.6	Imidazole pH 7.2	Tris-HCl pH 8.3	
F	SPG pH 7.6	MMT pH 7.6	MES pH 6.5	BisTris pH 6.8	Citrate pH 6.75	Sodium Phosphate pH 7.3	Potassium Phosphate pH 7.3	HEPES pH 7.8	Imidazole pH 7.45	Tris-HCl pH 8.5	
G	SPG pH 8.0	MMT pH 8.0	MES pH 6.7	BisTris pH 7.0		Sodium Phosphate pH 7.6	Potassium Phosphate pH 7.6	HEPES pH 8.0	Imidazole pH 7.7	Tris-HCl pH 8.8	
H	SPG pH 8.4	MMT pH 8.4	MES pH 6.9	BisTris pH 7.2		Sodium Phosphate pH 7.9	Potassium Phosphate pH 7.9	HEPES pH 8.2	Imidazole pH 7.95	Tris-HCl pH 9.0	

2.13 Sample Preparation and Negative Stain Electron Microscopy

After buffer optimization with *ProteoPlex,* Saw1-Rad1-Rad10 was purified by the same method as in section 2.10, except 50 mM Tris-HCl pH 8.0 was substituted with 50 mM HEPES pH 6.8. Concentrated Saw1-Rad1-Rad10 from peak 1 of anion exchange chromatography was incubated with 1.2X molar excess of DNA substrate and injected onto a Superdex 200 100/300

GL size exclusion chromatography column (GE Healthcare) equilibrated with Saw1-Rad1-Rad10 SEC buffer (50 mM Citrate pH 6.75, 150 mM NaCl, 5 mM β-mercaptoethanol, and 0.1 mM EDTA). Fractions at the apex of the peak containing the nucleoprotein complex (~11.4 mL retention volume) were used for negative stain electron microscopy. 5 µL of sample at ~60 nM was applied to glow discharged copper grids (Electron Microscopy Sciences) with an added layer of carbon, incubated 2 minutes, and blotted dry. Grids were then stained with 1% uranyl acetate for 1 minute, blotted dry, and air-dried further before storage. Images are taken between 250,000 and 300,000 X magnification with a JEOL TEMSCAN Electron Microscope (Electron Microscopy Facility, Faculty of Health Sciences, McMaster University).

2.14 SEC of complexes with varying ratios of Saw1-Rad1-Rad10 to 40-mer splayed-Y DNA

For each run, 130 µL of protein-DNA sample was prepared by incubating 5 µM protein with varying concentrations of 40-mer splayed-Y DNA: 1.67 µM (3:1, protein:DNA), 2.5 µM (2:1), 5 µM (1:1), 10 µM (1:2), 20 µM (1:4), 30 µM (1:6) at 4°C for 1 hour. 100 µL Saw1-Rad1-Rad10 alone (1:0) at 5 µM or protein-DNA complex samples were injected onto a Superose 6 Increase 10/300 GL (GE Healthcare) equilibrated with Saw1-Rad1-Rad10 SEC buffer and run for one column volume.

CHAPTER 3

RESULTS: OPTIMIZED PURIFICATION OF RECOMBINANT SAW1

3.1 Optimizing Solubility of His-Saw1 over-expressed in *Escherichia coli*

The biochemical characterization of Saw1 requires high quantities of pure, stable, and active protein. However, Saw1 is highly unstable in the absence of Rad1 or Rad10 (Li *et al.*, 2013) and this makes it difficult to obtain the required yields and quality of recombinant protein. The first challenge was expressing soluble Saw1 in *E. coli* over-expression strains.

Expression and solubility of His-Saw1 was tested at a range of temperatures and in a variety of *E. coli* expression cell lines. In all cases, cells expressed only slightly soluble His-Saw1 (**Fig 3.1A-C**). Auto-induction was also performed with BL21 Codon+ cells to take advantage of leaky expression of pET expression plasmids. This technique did not yield soluble protein (**Fig 3.1D**). To assess the relationship between cell density at induction and the production of soluble protein, the solubility of recombinant His-Saw1 was assessed after expression in BL21 Star pRARE cells that were induced with IPTG at different cell densities and incubated at a constant temperature (20°C) (**Fig 3.2**). His-Saw1 was almost entirely soluble when expressed by cells grown to a lower density prior to induction (OD_{600} 0.4), although levels of expression were significantly lower than cells grown to higher densities. These optimized parameters were applied to large-scale 1 L preparations of His-Saw1 for purification.
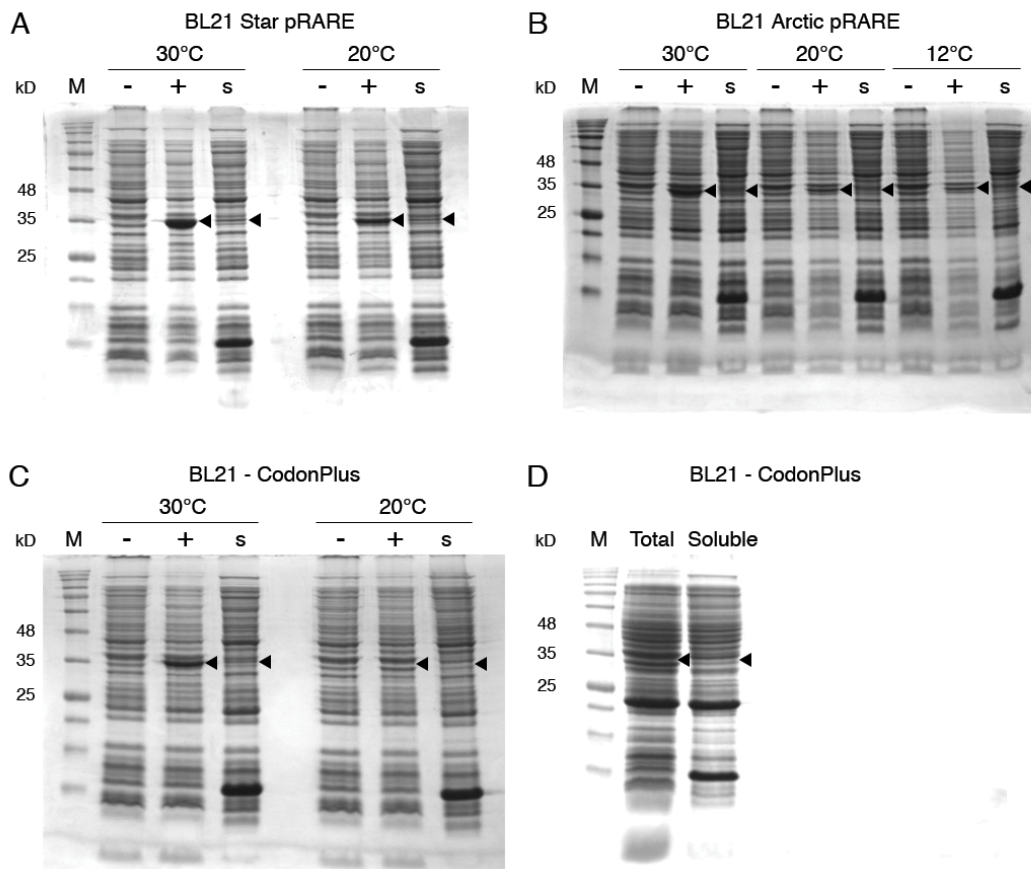
Figure 3.1 *Recombinant Saw1 is mostly insoluble when expressed in E. coli expression cell lines.*
Expression and solubility of His-Saw1 in (A) BL21 Star pRARE, (B) BL21 Arctic pRARE, and (C) BL21 CodonPlus expression cell lines. (D) Auto-induction of His-Saw1 in BL21 CodonPlus. "-" represents proteins present before induction with IPTG, "+" represents the total expressed protein after induction with IPTG, and "s" represents the soluble fraction of proteins expressed after induction with IPTG. Black triangles indicate bands representing His-Saw1, or where His-Saw1 bands are expected to be. "M" represents the BluEye Protein Marker (GeneDirex) with relevant bands identified (in kDa) to the left.
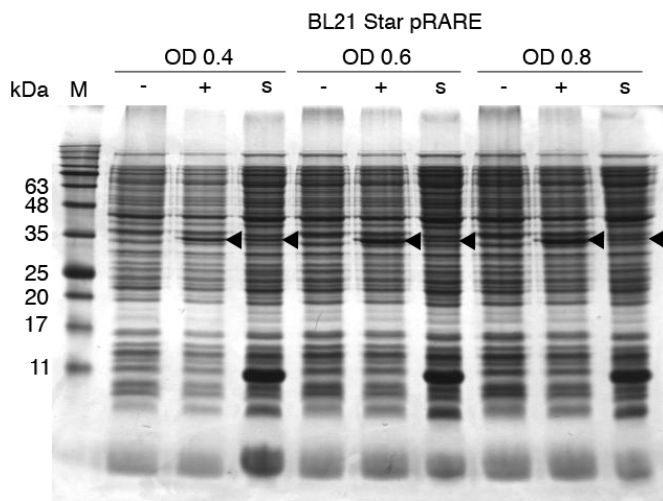
Figure 3.2 *Effect of cell density on solubility of His-Saw1*
Expression and solubility of His-Saw1 in BL21 Star pRARE cells induced at varying OD_{600}. "-" represents proteins present before induction with IPTG, "+" represents the total expressed protein after induction with IPTG, and "s" represents the soluble fraction of proteins expressed after induction with IPTG. Black triangles indicate bands representing His-Saw1. "M" represents the BluEye Protein Marker (GeneDirex) with relevant bands identified (in kDa) to the left.

3.2 Initial Purification of His-Saw1

Initially, we purified His-Saw1 as described by Li *et al.* (2013) where buffers were prepared with Tris-HCl pH 8.0. A two-step purification process using immobilized metal affinity chromatography (IMAC) and subsequent ion exchange chromatography (IEC) yielded pure recombinant Saw1 (**Fig 3.3**). This purification produced very unstable protein that tended to aggregate and precipitate during concentration of pooled IEC fractions. This led to up to 90% protein losses. Washing the concentrator membrane with 1 M NaOH showed a large quantity of protein adhering to the membrane after sample concentration. The final purification yield was 60 µg of His-Saw1 per 1 L of cell culture. In terms of total protein, this yield is similar to that observed by Li and co-workers (<0.1 mg per liter of cell culture), despite using chromatography columns providing a more stringent separation of contaminants than those used in the original

study (Li *et al.*, 2013). This low yield limited our ability to study this protein biochemically and structurally.
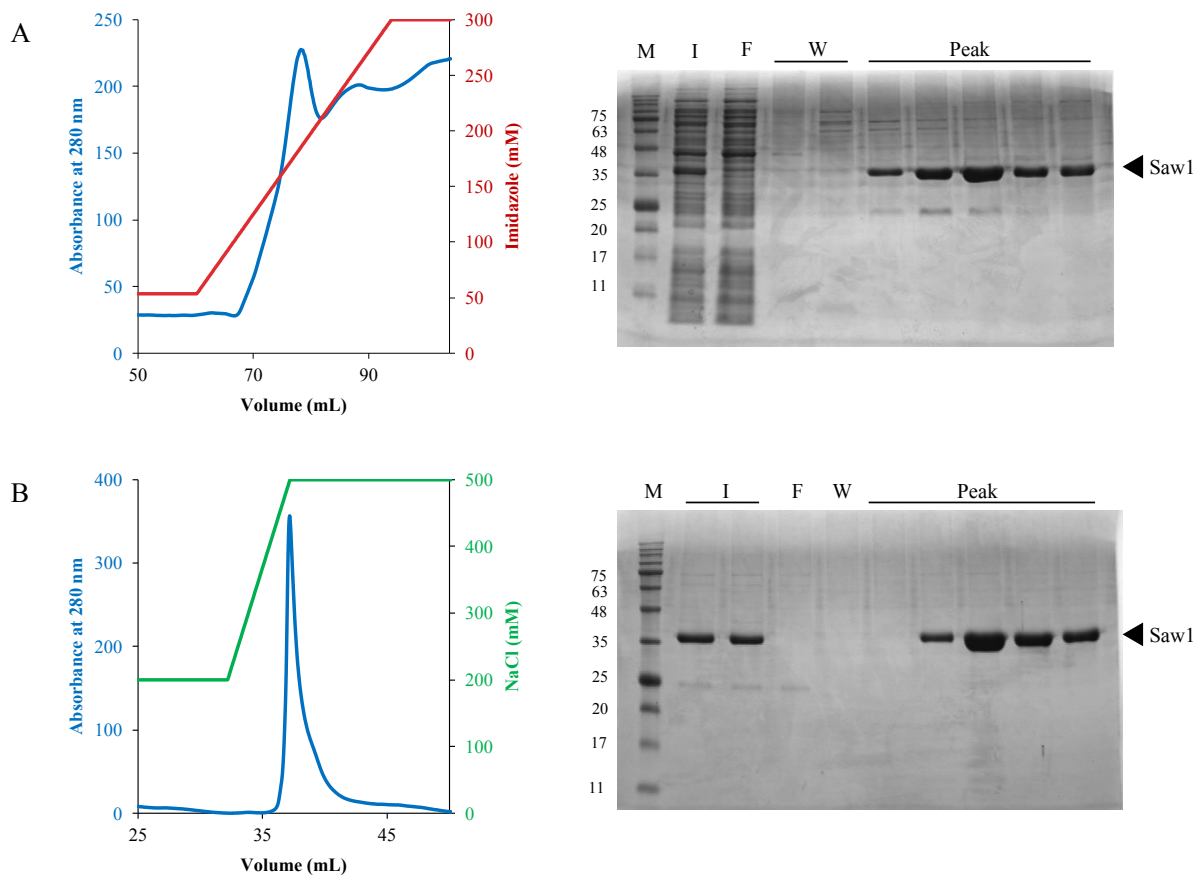
Figure 3.3 *Two-step Purification of His-Saw1.*
(A) Elution profile of Saw1 from Nickel affinity chromatography with an imidazole gradient (left) and the corresponding 15% SDS-PAGE gel (right). (B) Elution profile of Saw1 from the cation exchange chromatography with an NaCl gradient (left) and the corresponding 15% SDS-PAGE gel (right). "M" represents the BluEye Protein Marker (GeneDirex) with relevant bands identified (in kDa) to the left. The gels contain samples representative of the protein loaded into ("I") the column, the loading flow-through ("F"), buffer washes ("W"), and fractions from the eluted peak seen in the chromatogram on the left.

3.3 Preliminary optimization of His-Saw1 purification

One technique which has been found to help stabilize proteins is the use of 50 mM Arginine and 50 mM Glutamate in the protein storage buffer (Golovanov *et al.*, 2004). Other

factors that potentially decrease yield loss include reducing the centrifugation speed of concentration to 1,400 g, short centrifugation times (5 minute spins), and resuspending the protein solution after each spin during concentration. These techniques prevent the formation of a high local concentration of protein at the surface of the concentrator membrane, which could potentially cause aggregation, precipitation, and interactions with the membrane. Although these changes were implemented, His-Saw1 yield increased only marginally. These yields were still too low for subsequent protein characterization, and there was still significant precipitation during concentrating.

Next, we sought to improve stability by using a SUMO-tagged construct of Saw1. The SUMO fusion tag is a 100-amino acid polypeptide that has been shown to improve the solubility of otherwise insoluble or poorly soluble proteins (Steinmetz, 2011). Although this construct was soluble, dynamic light scattering showed that the majority of the sample aggregated immediately after elution from IMAC.

3.4 Optimizing His-Saw1 buffer conditions with Dynamic Light Scattering

Since neither the Arg/Glu additive nor the solubility tag had any significant effect on stability of Saw1, we began exploring global buffer parameters, such as the pH and formulation of the purification buffer as well as buffer additives. To this end, we used Dynamic Light Scattering (DLS) as a tool to assess the extent of aggregation of His-Saw1 in a range of buffer conditions (Jancarik *et al.*, 2004). DLS relates the fluctuations in intensities of scattered light through a sample to the size of particles within that sample. A size distribution curve can then be generated where the relative amounts of different species in the sample can be determined. Folded and stable proteins are represented by a peak between 5-10 nm (radius), and large

oligomers or aggregates are represented by peaks at a larger radius. The presence of larger oligomeric assemblies or aggregates indicates destabilization of the protein of interest, possibly due to partial unfolding which can lead to unwanted interactions. The DLS measurement also includes a calculated Polydispersion Index (PDI) representing the level of polydispersity of particles in the sample. High polydispersity (larger PDI) refers to a wide distribution of different sizes of particles in a sample. For biochemical and structural studies, it is necessary to have a monodispersed sample with a single species and as little variation as possible.

Based on an assay developed by Jancarik *et al.* (2004), we first screened a range of pH buffering systems to establish an optimal pH for Saw1 stability. With the exception of K_2HPO_4 pH 5.0, a clear trend was seen: with increasing pH, the intensity peak representing small (>10 nm radius) species decreased while the intensity peak representing aggregates increased. Of the ten pH buffers tested, four are shown in **Figure 3.4** to show the trend in pH-dependent protein stability. The least aggregation was seen in the sample with MES pH 6.0 whereas the most aggregation was seen with CHES pH 9.0. Saw1 precipitated at pH below 6.0, precluding the DLS analysis in buffers at lower pH. Indeed, the DLS profile measured in potassium phosphate at pH 5.0 showed no peaks, as His-Saw1 heavily precipitated in this buffer and it was likely removed during the centrifugation step prior to the DLS measurement.
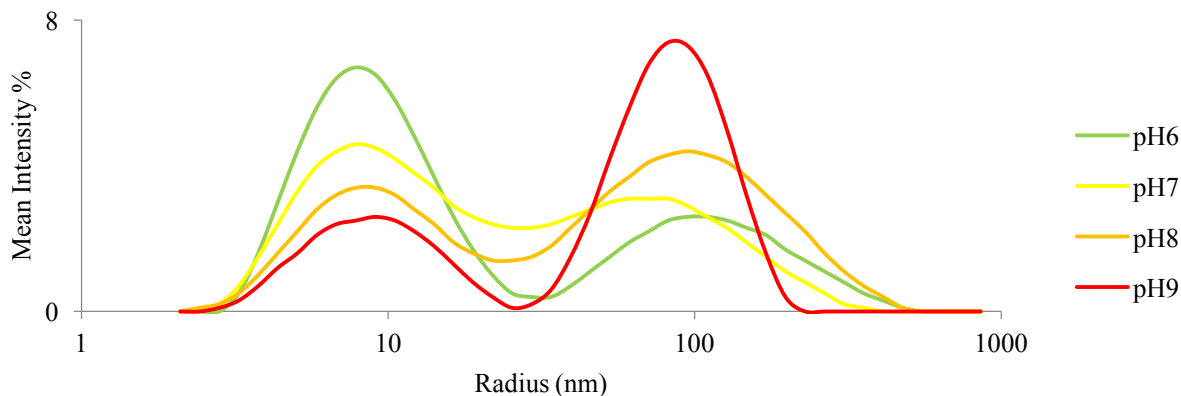
Figure 3.4: *The effect of pH on the stability of Saw1 through Dynamic Light Scattering.*
DLS profiles of His-Saw1 diluted in buffers containing MES pH 6.0 (green), HEPES pH 7.0 (yellow), Tris-HCl pH 8.0 (orange), CHES pH 9.0 (red). The size distribution intensity curves show the distribution of particles between two main species: small particles with ~10 nm radius, and large assemblies with ~100 nm radius.

Since MES pH 6.0 had a highly stabilizing effect, the remainder of the purified His-Saw1 was buffer exchanged into a MES pH 6.0-based buffer. We then used the same assay to assess the stabilizing effect of various additives in the buffer, including reducing agents (DTT, βME), glycerol, and NaCl concentration. In this case, analysis of the PDI generated for each sample was used to ascertain whether the condition had a stabilizing or destabilizing effect on His-Saw1 compared to a control (**Figure 3.5**). If the PDI decreases, the additive has a stabilizing effect on Saw1 and should be incorporated into the purification buffers. For example, both 5 mM DTT/βME and 10 mM βME have similar stabilizing effects compared to 1.4 mM βME. Decreasing the glycerol concentration from 10% to 5% has a negligible increase in the PDI and we concluded that His-Saw1 remains stable at lower glycerol concentration. Finally, decreasing the salt concentration seems to have a stabilizing effect on His-Saw1 since the PDI at 0.4, 0.3, and 0.2 M NaCl are all lower than the control.
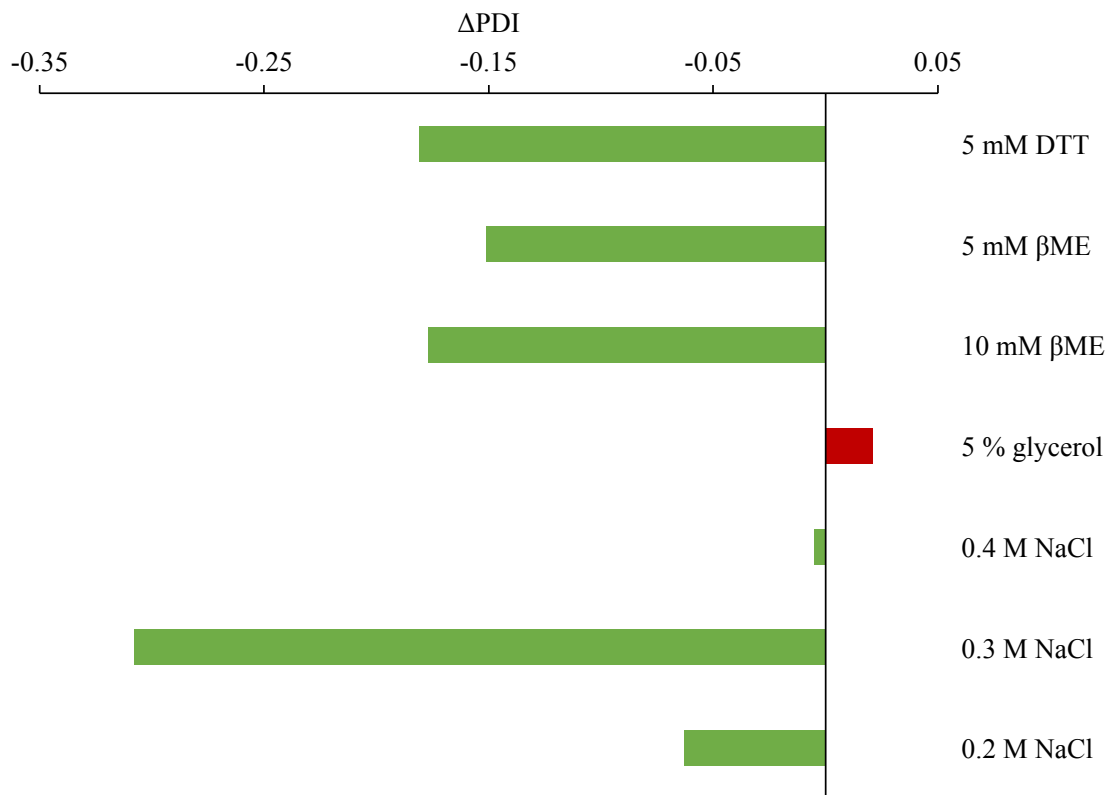
Figure 3.5: *The effect of buffer additives on the polydispersity of His-Saw1*
Differences in PDI of His-Saw1 in various buffer additives relative to a dilution control of His-Saw1 in MES buffer (50 mM MES pH6, 50 mM RE, 500 mM NaCl, 1.4 mM βME, 0.1 mM EDTA, 10% glycerol). Green bars represent a decrease in PDI (stabilizing effect) and red bars represent an increase in PDI (destabilizing effect).

3.5 High throughput buffer condition screen of His-Saw1 with *Thermofluor*

Screening different pH buffers using DLS revealed that the stability of His-Saw1 was drastically influenced by the pH of the buffer. However, only a select number of buffers could be tested with DLS due to the amount of protein required for these measurements. We wanted to screen a wider range of buffers (both pH and formulation) to perhaps find the optimal buffering agent for His-Saw1 purification. To do this, we used differential scanning fluorimetry (DSF) to monitor the thermal stability of His-Saw1 in different conditions. The *Thermofluor* assay is done in a 96-well and only requires 2 μL of protein in the micromolar concentration range per measurement. This setup maximizes the throughput of the experiment while using minimal

amounts of sample. We used SYPRO-Orange because it fluoresces when bound to hydrophobic pockets of proteins and, therefore, fluorescence intensity can be used to monitor temperature-dependent protein unfolding, generating a melting curve. In these curves, a stabilizing condition results in a shift of the melting curve to higher temperatures and *vice versa* (Boivin *et al.*, 2013). Using this approach, we could readily assess the effect of buffer formulation and pH, salt, and other additives on the thermal stability of His-Saw1 (**Table 2.1 – Materials and Methods**).

We screened a broad range of pH buffers, including those assayed in the DLS screen. Between these four pH buffers (MES pH 6.0, HEPES pH 7.0, Tris-HCl pH 8.0, CHES pH 9.0), protein stability decreased with increasing pH, consistent with DLS (**Fig 3.6A**). We assayed 19 different buffering agents in a pH range from 4 to 9. As previously shown, the stability of His-Saw1 generally decreased with increasing pH, showing optimal stability at pH 6. This trend applies to even lower pH as pH 4 and 5 buffers (except for Citrate pH 5.0) showed severe destabilization of His-Saw1 (**Fig 3.6C**), corroborating the DLS result of His-Saw1 in Potassium phosphate pH 5.0. Interestingly, Citrate pH 5.0 showed the same thermal stabilization of His-Saw1 as MES pH 6.0. The optimal buffer according to this assay was Citrate pH 6.0, followed by BisTris pH 6.5, and Potassium phosphate pH 6.0 (**Fig 3.6B**).
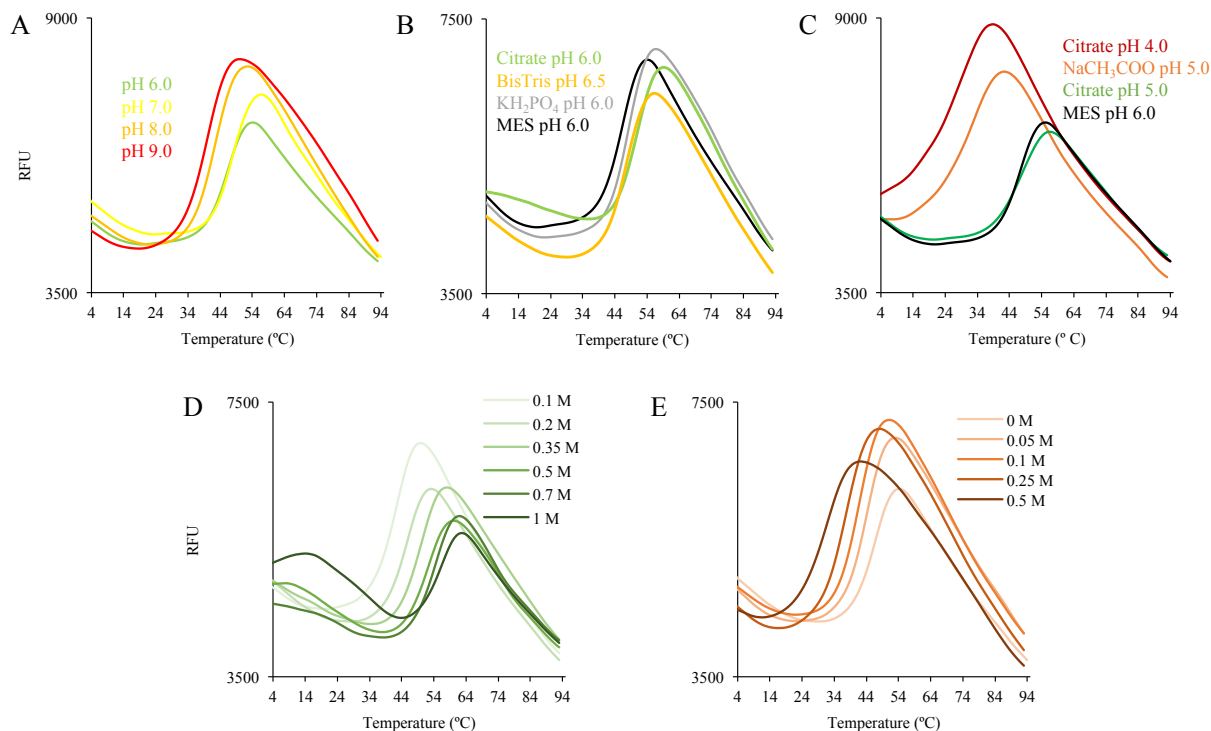
Figure 3.6: *Stabilization of Saw1 by buffer composition in the Thermofluor assay*
(A) Melting curves of His-Saw1 in buffers screened with DLS (**Fig 3.4**): MES pH 6.0 (green), HEPES pH 7.0 (yellow), Tris-HCl pH 8.0 (orange), CHES pH 9.0 (red) (all in the presence of 250 mM NaCl). (B) Melting curves of Saw1 in Citrate pH 6.0, BisTris pH 6.5, and Potassium Phosphate (KH_2PO_4) pH 6.0 compared to MES pH 6.0 (all in the presence of 250 mM NaCl). (C) Destabilizing effect of Citrate pH 4.0 and Sodium Acetate ($NaCH_3COO$) pH 5.0, compared to MES pH 6.0 and Citrate pH 5.0 (all in the presence of 250 mM NaCl). (D) Effect of NaCl on His-Saw1 stability (all in the presence of 50 mM MES pH 6.0). (E) Effect of increasing concentration of imidazole on His-Saw1 stability (all in the presence of 50 mM MES pH 6.0 and 200 mM NaCl). The Y-axis represents the Relative Fluorescence Units (RFU) and the X-axis represents the temperature (^oC).

The ionic strength of purification buffers can have a significant effect on the solubility and stability of purified proteins. We thought this to be the case with His-Saw1, as some precipitation occurred when diluting the sample recovered from IMAC to a lower salt concentration, prior to its application to IEC. We, therefore, included a range of NaCl concentrations in our *Thermofluor* screen. **Figure 3.6D** shows a shift in the melting curve to lower temperatures with decreasing salt concentration. His-Saw1 seems to be most stable at high salt concentrations (contrary to the DLS screen results), and the T_m decreases significantly

between 0.35 M and 0.2 M NaCl ($\Delta 4.5^{\circ}$C) (**Table 3.1k, l**). This suggests that the salt concentration should be kept relatively high during His-Saw1 purification, when possible, to avoid destabilization. There was also no difference in the effect of NaCl versus KCl on His-Saw1 stability.

Table 3.1: *His-Saw1 relative melting temperatures from Thermofluor DSF assay*

Condition	T_m (°C)
(a) Citrate pH 4	30
(b) Acetate pH 5	35.5
(c) Citrate pH 5	50
(d) MES pH 6	48.5
(e) Citrate pH 6	53
(f) BisTris pH 6.5	49.5
(g) HEPES pH 7	49
(h) Tris-HCl pH 8	44.5
(i) CHES pH 9	42
(j) 0.1 M NaCl	44.5
(k) 0.2 M NaCl	48
(l) 0.35 M NaCl	52.5
(m) 0.5 M NaCl	54.5
(n) 0.7 M NaCl	57
(o) 1.0 M NaCl	58.5
(p) 50 mM Arg/Glu	50
(q) 5% glycerol	51
(r) 0.05 M Imidazole	46
(s) 0.1 M Imidazole	43.5
(t) 0.25 M Imidazole	40
(u) 0.5 M Imidazole	34
(v) 10 mM $MnCl_2$	47
(w) 10 mM $CaCl_2$	48
(x) 10 mM $MgCl_2$	48
(y) 10 mM $ZnCl_2$	28

Certain buffer additives have the ability to stabilize purified proteins. Of the additives screened, only glycerol had a significantly stabilizing effect on Saw1. The addition of 5% glycerol increased the T_m by 2.5 °C (relative to **Table 3.1e**) and there was no significant

difference between the addition of 5%, 10%, and 20% glycerol. The presence of 50 mM Arg/Glu mix only slightly increased the T_m (by 1 ºC). This additive was found to interfere with downstream DNA binding experiments and is thus undesirable as a buffer additive. EDTA and reducing agents had no significant effect on the melting temperature. Contradictory to this, we later found that exchanging purified His-Saw1 into a buffer containing 1.4 mM βME from one with 5 mM βME caused severe destabilization resulting in precipitation. Therefore, the βME concentration was maintained at 5 mM in all steps of purification. Most metal additives destabilized Saw1, except for $MnCl_2$, $CaCl_2$, and $MgCl_2$, which had no effect on its thermal stability (**Table 3.1w**, **x**, **y** compared to **e**). Of the metals tested, $ZnCl_2$ had the most destabilizing effect, decreasing the T_m to 28 ºC (**Table 3.1z**).

We tested the effect of imidazole concentration on the stability of His-Saw1, since imidazole is used to elute His-Saw1 during IMAC. *Thermofluor* showed that, with increasing concentrations of imidazole, the melting curve of His-Saw1 shifts to lower temperatures **(Fig 3.6E).** This indicates that high concentrations of imidazole have a destabilizing effect on His-Saw1, especially when concentrations are greater than 0.25 M. In subsequent purifications, the imidazole concentration was kept below 0.25 M and contact was limited as much as possible.

3.6 Optimization of tagless Saw1 expression and purification in Citrate pH 6.0

Although Citrate pH 6.0 showed the greatest stabilizing effect out of all buffers screened with the *Thermofluor* assay, its use introduced some challenges to the purification protocol. Citrate is a known metal chelator and is not compatible with IMAC. Since the histidine tag was no longer of use, we decided to use a tagless construct of Saw1.

We first assessed the solubility of Saw1 in Citrate buffer after overexpression in BL21 Star pRARE cells induced at different cell density (OD_{600} 0.4, 0.6, 0.8). For each cell density, all of

the expressed protein was soluble, unlike the lysis of His-Saw1 in Tris-HCl pH 8.0 buffer. This allowed us to grow the cells to higher OD before inducing expression.

Since IMAC was not compatible with Citrate buffer, we decided to use ammonium sulfate precipitation followed by dialysis and IEC to purify tagless Saw1. However, we were unable to reproduce the precipitation conditions from purification to purification. Furthermore, much of the precipitated Saw1 was unable to be re-solubilized when dialyzed into a buffer suitable for IEC. As a result of these inconsistencies, downstream purification steps were also inconsistent and yielded either no protein or highly impure Saw1. The impurities could not be resolved by affinity chromatography (heparin column) or IEC (MonoQ column).

Due to the difficulties in optimizing tagless Saw1 purification in Citrate pH 6.0, we decided that metal affinity chromatography was an important initial purification step. As a result, we designed a purification protocol using the next best buffer for His-Saw1 based on the *Thermofluor* screen – BisTris pH 6.5 (**Fig 3.6B**).

3.7 Optimization of His-Saw1 purification in BisTris pH 6.5

Unlike Citrate, BisTris buffer is compatible with IMAC. We therefore designed a 2-step purification protocol, similar to the initial purification, where clarified cell lysate is applied to IMAC and subsequently to IEC, all in a BisTris pH 6.5 buffer. Purified, concentrated protein was stored in BisTris pH 6.5 with a relatively high concentration of NaCl (0.3 M), as suggested by the *Thermofluor* salt screen (**Fig 3.6D**).

The final yield of this purification protocol was approximately 820 µg of His-Saw1 per 1 L cell preparation. This is a 13X improvement from the original Tris-HCl pH 8.0 purification (~60 µg/1 L culture). Where yields were previously too low, this purification allowed for a final polishing step with size exclusion chromatography (SEC) to improve the monodispersity of the

sample (**Fig 3.7**). **Figure 3.7B** shows concentrated His-Saw1 pre- and post- SEC with an S75. We are able to resolve and remove several higher molecular weight contaminants, eluting between 8-10 mL (**Fig 3.7A**), as seen as the disappearance of the higher molecular weight bands in **Figure 3.7B**. The result is a very pure (**Figure 3.7B** "Post"), stable, and monodispersed protein sample, shown by DLS (**Fig 3.7C**).
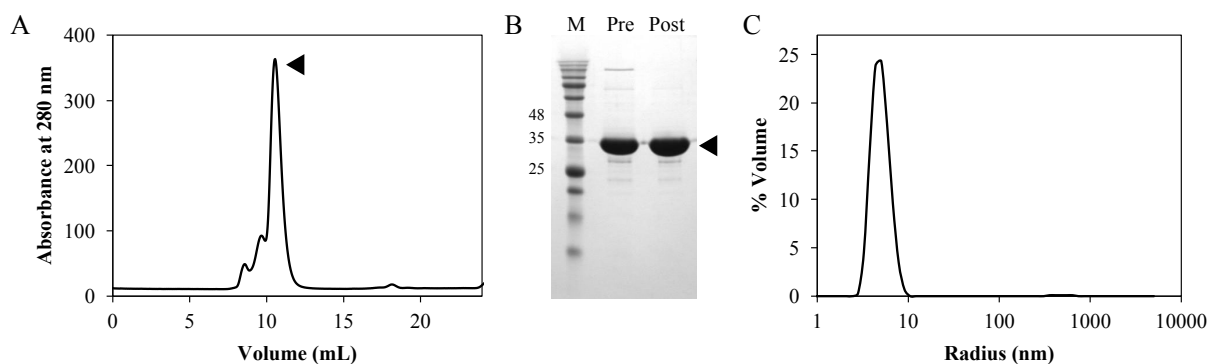
Figure 3.7: *Improved quality of His-Saw1 purified in BisTris pH 6.5.*
(A) Size exclusion chromatography (S75) elution profile of His-Saw1 (black triangle) . (B) SDS-PAGE of concentrated His-Saw1 (black triangle) before (Pre) and after (Post) polishing with gel filtration. (C) DLS size distribution curve (by volume) of purified, concentrated His-Saw1 after gel filtration.

3.8 Recombinant Saw1 binds 3′-flap DNA structures at pH 6.5

The initial purification protocol (Tris-HCl pH 8.0) and the optimized purification protocol (BisTris pH 6.5) yielded very different products in terms of stability and quantity. We compared the binding activity of both sets of purified proteins, using the 60-mer 3′-flap DNA substrate (see Figure 2.1 for the structure of this substrate as well as all others used for binding assays). We found that the protein in BisTris pH 6.5 showed improved binding than that purified in Tris-HCl pH 8.0 (Fig 3.8A). At 1:1 ratio of DNA:protein, we see a distinct band shift with the optimized protein (BisTris pH 6.5), but not with the initial protein preparation (Tris-HCl pH 8.0). We can also see an improvement in the protein stability. In the EMSA with Tris-HCl pH 8.0-

purified His-Saw1, we see DNA remaining in the wells of the gel, even at low concentrations of protein (Fig 3.8A). This could be because His-Saw1 in Tris-HCl pH 8.0 is unstable, causing aggregation that cannot migrate from the wells, and this retains some of the DNA in the wells. We only see this effect at much higher concentrations with the optimized His-Saw1 preparation (BisTris pH 6.5), indicating a more stable preparation.

Using the same substrate and our optimized His-Saw1 (purified in BisTris pH 6.5), we also compared DNA-binding activity in EMSA buffer prepared with either Tris-HCl pH 8.0 or BisTris pH 6.5 to see if the pH of the buffer had an effect on affinity. There were no differences in DNA binding between pH 8.0 and pH 6.5 (both resembling the affinity seen in **Fig 3.8A**, bottom gel), implying that the changes we saw in **Figure 3.8A** are due to an improvement in protein quality after purification. Since there was no difference between the use of these two EMSA buffer conditions, we performed experiments with fluorescently labeled DNA substrates in Tris-HCl pH 8.0. This is because lower pH reduces the fluorescence intensity of the 6-FAM fluorescent label, which could interfere with our interpretation of these binding experiments.
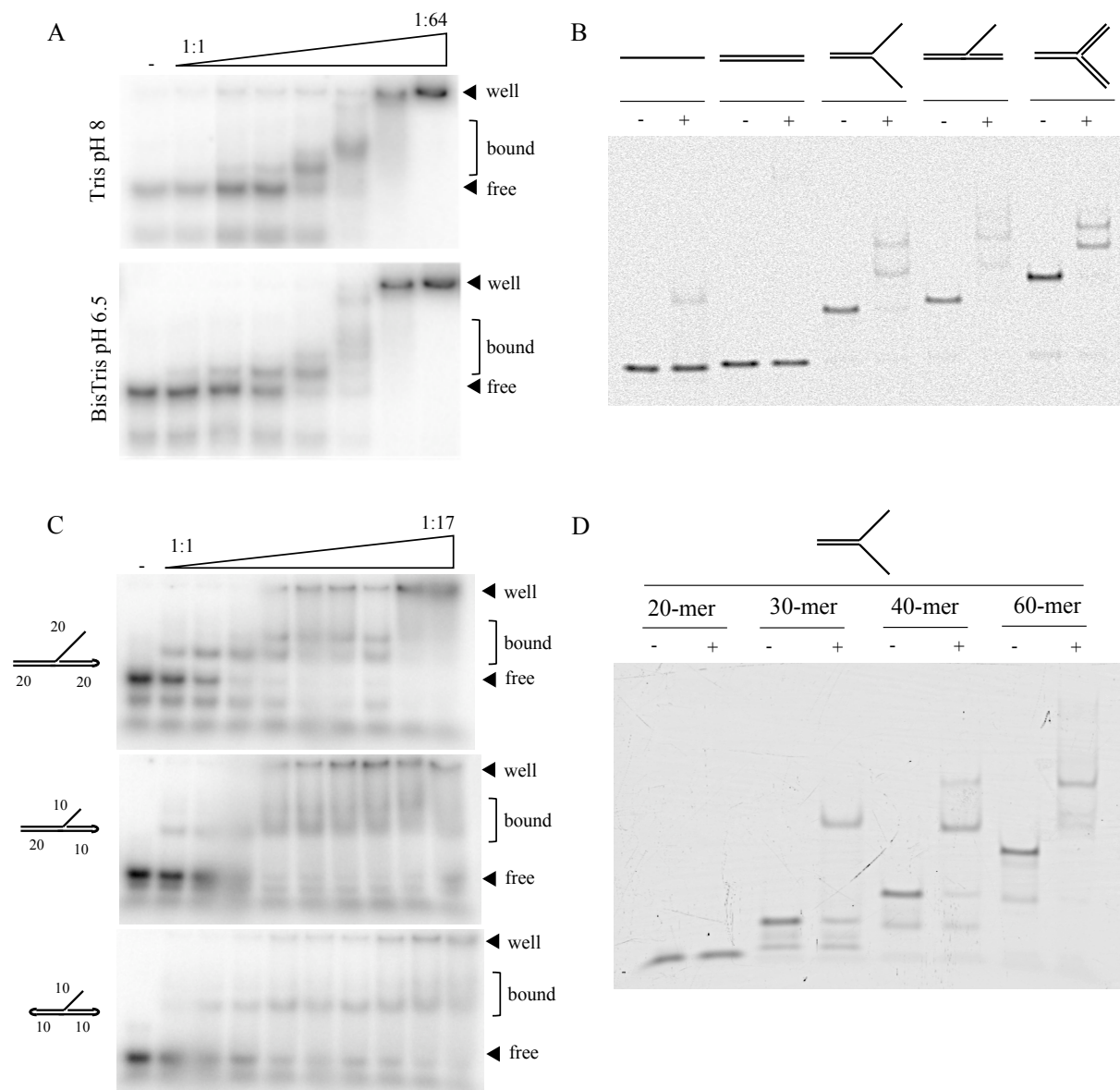
Figure 3.8: *Recombinant Saw1 DNA binding specificity and minimal substrate determination*
(A) EMSA of a 60-mer 3′-flap DNA substrate without (-) or with increasing molar ratios of His-Saw1 purified in Tris-HCl pH 8 buffer (top) or BisTris pH 6.5 buffer (bottom). (B) EMSA of 60 nt ssDNA, 60 bp dsDNA, 60-mer splayed, 60-mer 3′-flap, and 60-mer fork with (+) or without (-) 80 nM His-Saw1. (C) EMSA of 40-mer, 30-mer (10 nt flap), and 10-mer 3′-flap DNA without (-) or with increasing molar ratios of His-Saw1. (D) EMSA of 60-mer, 40-mer, 30-mer, and 20-mer splayed-Y DNA substrates without (-) or with (+) His-Saw1. Bound and free DNA are indicated as well as the position of the well (where some protein-DNA complex remains trapped) in (A) and (C). (B)-(D) were performed with His-Saw1 purified in BisTris pH 6.5 buffer.

3.9 Structure-selective DNA binding activity of recombinant Saw1

Saw1 binds splayed DNA, 3'-flap DNA, and fork structures with high affinity, and has little to no affinity for ssDNA and dsDNA, respectively (Li *et al.*, 2013). We wanted to ensure our purified, recombinant His-Saw1 maintains this specific binding for branched substrates. We performed EMSAs of His-Saw1 with 60-mer ssDNA, 60-mer duplex DNA, 60-mer splayed-Y, 60-mer 3′-flap, and 60-mer fork DNA. His-Saw1 bound to the splayed, flap, and fork substrates, and had no affinity for duplex DNA and weak affinity for ssDNA (**Fig 3.8B**).

When designing substrates for structural studies, like X-ray crystallography, one must take into account the size of the substrate. Long, flexible ends of DNA can impair crystal formation (Rhodes, 2010). Since we can now produce sufficient amounts of stable His-Saw1, we wanted to define its DNA binding specificity for future biochemical and structural characterization. We first began exploring minimal 3'-flap substrates that maintained binding to His-Saw1. **Figure 3.8C** shows the progressive shortening of the left/right arms and 3′-flap of the DNA substrate from a 40-mer 3′-flap substrate to a 20-mer 3′-flap hairpin substrate. Shorter substrates seem to favour a single mode of binding, seen with the 20-mer 3′-flap hairpin as a single shift versus the two distinct shifts observed for the 40-mer 3′-flap substrate. The 20-mer hairpin substrate never fully shifts from the "free" state in this concentration range, unlike its larger counterparts. These "hairpin" structures also pose a problem in terms of mis-annealing and higher-molecular weight product formation, since there are complementary strands that could anneal both intra- and inter-molecularly. Even though intramolecular annealing is more favourable, intermolecular annealing can still occur, introducing heterogeneity to the DNA substrate.

Of the splayed-Y, 3'-flap, and fork substrates, the splayed-Y substrates demonstrated the most homogeneity and had the least number of mis-annealed products. This is because it only requires the annealing of two oligonucleotides, whereas the flap and fork substrates require the annealing of 3 and 4 oligonucleotides, respectively. Native gels of these substrates collected from SEC revealed that there were several annealed products in the flap and fork DNA samples, whereas the splayed had minimal mis-annealing (data not shown).

Due to the above observations, and since it is more physiologically relevant in the context of SSA, we decided to focus on the splayed-Y substrate. To define the minimal splayed Y substrate for which Saw1 maintained affinity, we assayed the binding of His-Saw1 to 20-mer, 30-mer, 40-mer, and 60-mer splayed-Y substrates (**Fig 3.8D**). Although His-Saw1 showed no affinity for 20-mer splayed DNA at a ratio of 8:1 protein:DNA, it bound to a 30-mer splayed substrate. His-Saw1 showed slightly reduced affinity for this substrate compared to the 40- and 60-mer splayed DNA.

3.10 Preliminary crystallization screens of His-Saw1

After defining a promising minimal splayed substrate, we were encouraged to begin preliminary crystallization trials to structurally investigate the interaction between Saw1 and its DNA substrate. Crystallization of protein-DNA complexes can often prove to be difficult, as was the case with His-Saw1. Pre-incubation of purified His-Saw1 and splayed DNA substrates prior to setting crystal screens resulted in immediate and heavy precipitation of the sample. To remedy this, we used an assay described by Pryor *et al.* (2012) which assessed various salt additives for their ability to stabilize or solubilize protein-DNA mixtures for crystallography. In this assay, a precipitated mixture of His-Saw1 and 20-mer splayed-Y DNA was mixed with an equal volume

of storage buffer containing varying concentrations of different salts. These drops were incubated on crystallization trays and checked for solubilization the following day. All soluble drops were those incubated with 50 mM Magnesium salts (MgCl$_2$, MgSO$_4$, Mg(CH$_3$COO)$_2$, Mg(CHOO)$_2$). As a result, we began supplementing SEC buffer with 25 mM Magnesium Acetate (Mg(CH$_3$COO)$_2$). With this, no precipitation occurred when mixing Saw1 with 20-mer splayed-Y DNA. However, this precipitation seemed to be related to the size of the substrate, as longer splayed Y substrates (40-mer versus 30-mer) produced more precipitation, and was more difficult to re-solubilize with the described method. Only by increasing the concentration of Magnesium Acetate to 100 mM would it re-solubilize. At this concentration, the salt may have an inhibitory effect on DNA binding and should therefore be used with caution.

Initial crystallization screens of His-Saw1 with and without DNA showed a tendency to precipitate in conditions containing polyethylene glycols (PEGs) of various lengths. In several screens that employed PEG of varying molecular weight as the precipitant, His-Saw1 with and without DNA precipitated in every drop, regardless of the concentration of protein-DNA complex. In other screens, non-PEG precipitants were more likely to yield soluble drops. In general, the majority of the precipitation observed in crystallization drops is granular and not indicative of destabilization or unfolding. As of yet, no crystals have been successfully produced, and this line of investigation is ongoing.

3.11 Design and characterization of Saw1 C-terminal constructs

To understand the DNA-binding activity of Saw1, our investigation needs to be done in a Rad1-independent manner. As previously shown, full length Saw1 is very unstable in the absence of its protein binding partners, Rad1 and Rad10. It is possible that the presence of Rad1

is necessary for proper folding of Saw1 and, in its absence, Saw1 becomes unstable. The Rad1-interacting region is at the N-terminus of Saw1 whereas the proposed DNA-binding motif is at the C-terminus (**Fig 3.11**) (Li *et al.*, 2013). By removing the N-terminus of Saw1, we propose that the DNA-binding function of Saw1 can be maintained, and the interaction between Saw1 and 3′-flap DNA can be studied in the absence of Rad1-Rad10. In parallel to studies with full length Saw1, we began to design and characterize C-terminal DNA-binding domains of Saw1.

To determine the boundaries of the C-terminal Saw1 constructs, we performed limited proteolysis on full-length His-tagged Saw1 with trypsin, in the presence (**Fig 3.9A**) and absence (**Fig 3.9B**) of 40-mer 3′-flap DNA. When 2.5 molar excess of DNA was added to the reaction prior to digestion, both the band representing full length His-Saw1 as well as some of the digestion products gained some resistance to proteolytic digestion (**Fig 3.9B**). This shows the protective nature of DNA once bound by His-Saw1, potentially through a conformational change upon binding, or through direct protection of Saw1 surfaces by DNA.
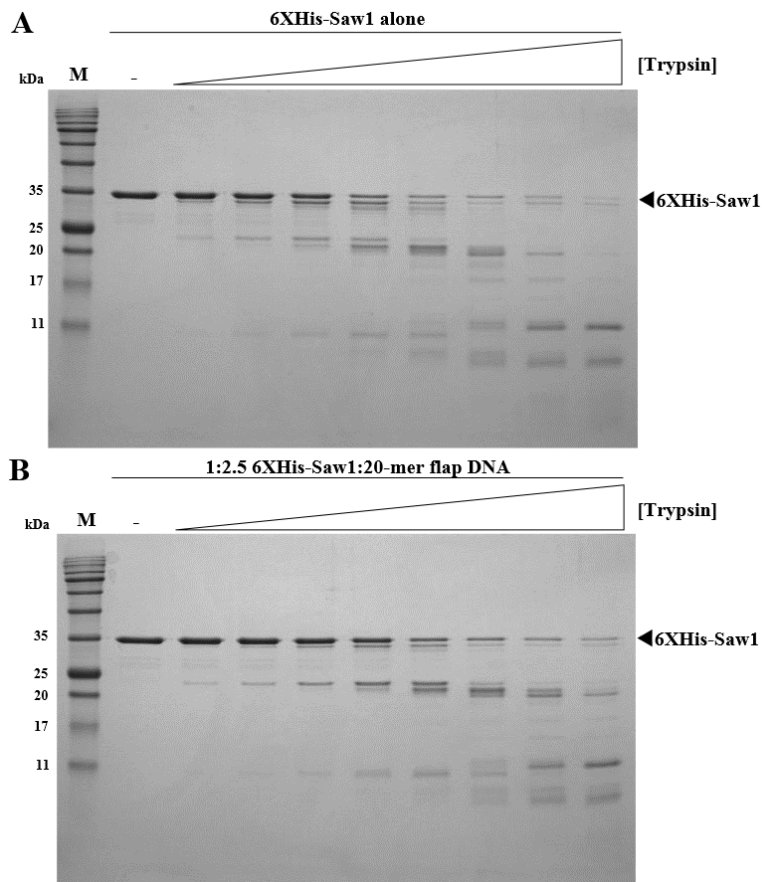
Figure 3.9: *Limited proteolysis of Saw1 with and without 3'-flap DNA.*
SDS-PAGE of degradation products from purified His-Saw1 at constant concentration with (A) or without (B) 2.5 molar excess 20-mer 3′ flap DNA, titrated with increasing concentrations of trypsin, from 0.16 µg/mL – 20 µg/mL. Each consecutive lane has double the trypsin concentration as the previous lane. Full-length His-Saw1 is indicated by black triangles.

Under both of these conditions, two major bands (~25 kDa and ~11 kDa) persisted to mid-range concentrations of trypsin (**Fig 3.9**). This indicated that His-Saw1 possibly consists of two domains connected by an unstructured linker region that is exposed to digestion by trypsin. To determine which of these bands corresponds to the N-terminus (Rad1 interaction) or the C-terminus (DNA-binding) of Saw1, a large-scale digestion of His-Saw1 with 0.63 µg/mL trypsin was carried out and applied to IMAC **(Fig 3.10)**. The fragment containing the N-terminus of the protein should also have a N-terminal hexa-Histidine tag and would therefore bind to the column, whereas the C-terminal fragment should not. **Figure 3.10B** shows that the smaller

fragment elutes at 30 mM imidazole, whereas the other fragments elute both at low and high concentrations of imidazole. The appearance of the large fragment at low concentrations of imidazole could be because the His-tag is also being cleaved from a portion of the population and loses affinity for the column (blue triangle in **Fig 3.10B**). The same is possible for full length Saw1, eluting both at low and high concentrations of imidazole (black triangle in **Fig 3.10B**). We therefore concluded that the last ~100 amino acids (~11 kDa) of the C-terminus of Saw1, consisting of the proposed DNA-binding motif, can fold into a stable domain.
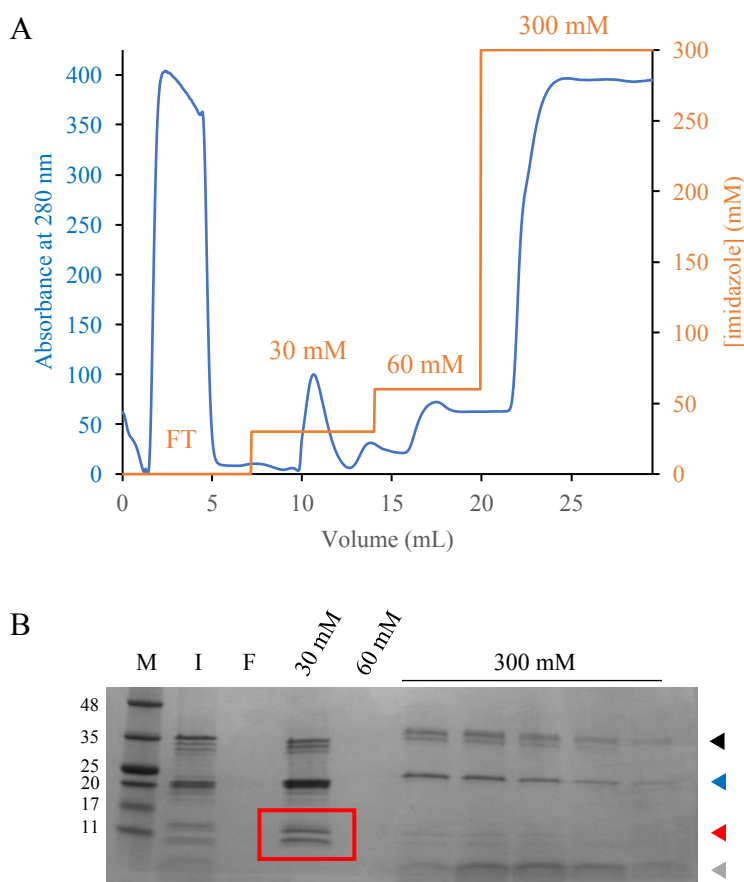
Figure 3.10: *Identifying stable domains of Saw1 through limited proteolysis.*
(A) IMAC elution profile of His-Saw1 digested with 0.63 µg/mL trypsin. (B) SDS-PAGE gel of representative samples from IMAC of digested Saw1 loaded into ("I") the column, the loading flow-through ("F"), the 30 mM, 60 mM, and 100 mM imidazole washes. The black triangle indicates full length Saw1 (possibly -/+ His-tag), the blue triangle indicates possible N-terminal fragment of Saw1 (possibly -/+ His-tag), the red triangle indicates the tagless C-terminal domain fragments (bands also highlighted by the red box in 30 mM wash), and the grey triangle indicates possibly cleaved His-tags. "M" represents the BluEye Protein ladder (GeneDirex), relevant sizes of bands indicated to the left of the gel (kDa).

Initial investigation into the minimal DNA binding domain of Saw1 was carried out by Amanda Khoo (4th-year undergraduate student). We based our initial construct boundaries on the results of the trypsin digest/IMAC purification of the fragments, also taking into account the predicted secondary structure of Saw1 (**Fig 3.11**). Residues 180-261 were chosen as the boundaries for the first Saw1 C-terminal (C1) construct. Although we could express C1, very little of it was soluble, and our attempted purifications had very low yields. We purified C1 by solubilizing the insoluble fraction of cell lysates with urea, and then applied it to IMAC. Pure C1 was refolded by removing urea through rounds of dialysis, and appeared to be stable and monodispersed when analyzed with DLS. However, the purified C1 had no affinity towards 3'-flap DNA substrates when an EMSA was performed with 5′-fluorescently labeled 60-mer 3′-flap DNA substrate. A mobility shift was only seen at very high concentrations of protein, indicating non-specific binding. We concluded that, although the stretch of positive residues at the C-terminus of Saw1 may be necessary, it is not sufficient for DNA binding. Another reason for the loss in DNA-binding is that C1 was possibly incorrectly re-folded.
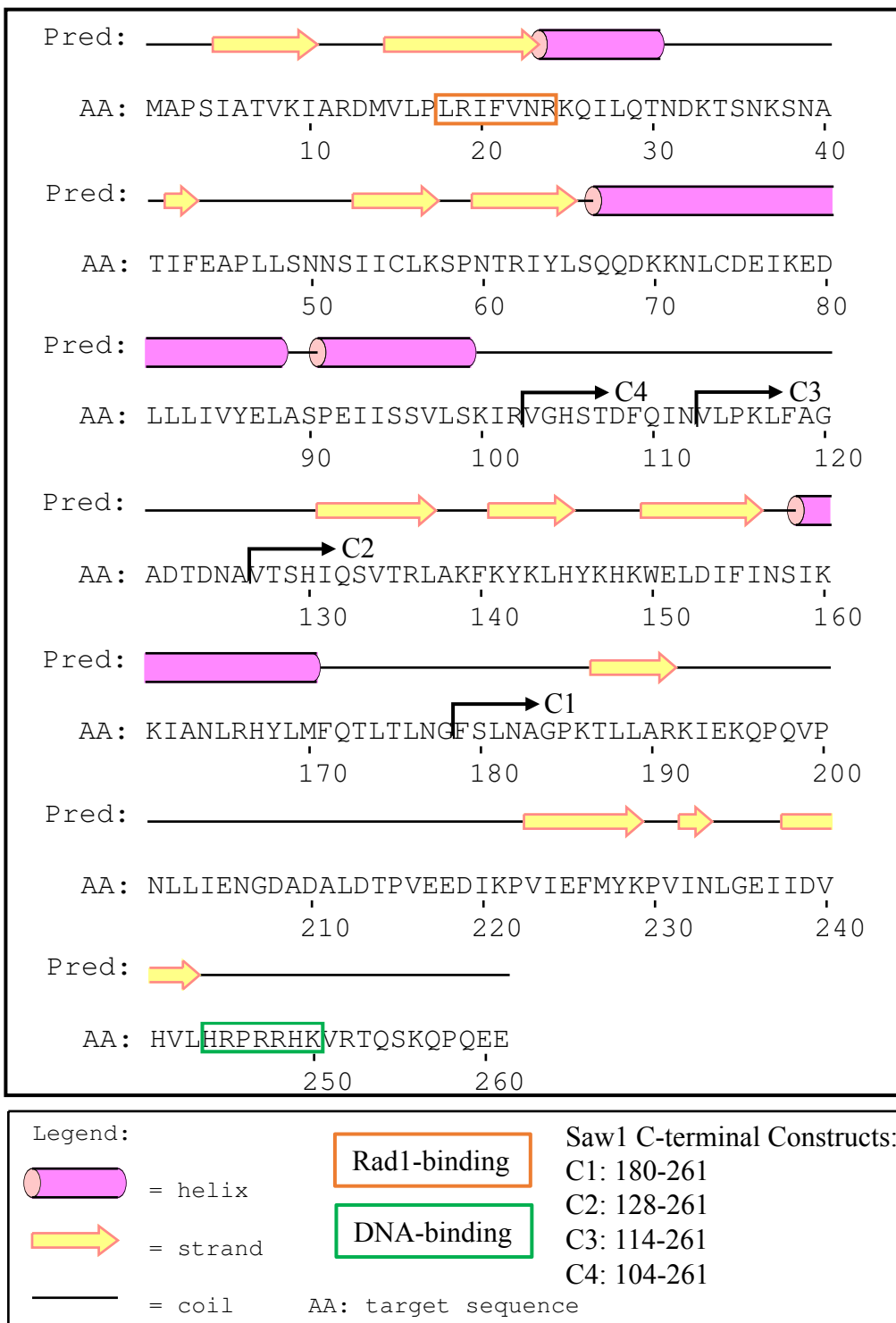
Figure 3.11: *Design of Saw1 C-terminal constructs*
Secondary structure prediction generated with the PSIPRED Protein Sequence Analysis Workbench (PSIPRED v3.3). The N-terminal boundaries of the various Saw1 C-terminal constructs are indicated by back arrows, and the exact variants are indicated.

We continued to investigate the minimal DNA-binding domain of Saw1 by designing larger C-terminal fragments. Using the predicted secondary structure of Saw1 (**Fig 3.10**), Eduardo Rojas (3rd year undergraduate student), aided in the design and production of larger His-tagged Saw1 constructs: C2 (residues 128-261), C3 (residues 114-261), and C4 (residues 104-261). Soluble C2 was expressed in BL21 Star pRARELysS cells, but a two-step purification of this construct (IMAC followed by IEC) yielded only trace amounts of protein. Optimization of this purification is required to produce enough C2 for DNA binding experiments. C3 and C4 have been assayed for expression and solubility. C3 was insoluble in BL21, BL21 Star, BL21 Star pRARE, and BL21 Star pRARELysS cells under all conditions tested. C4 was expressed and soluble in BL21 Star pRARE cells. An initial purification was attempted with this construct but, surprisingly, the majority of C4 was found in the insoluble fraction after cell lysis. This line of investigation is ongoing and a summary of the progress of the production of C-terminal constructs of Saw1 is outlined below in **Table 3.2**. In this table, green indicates successful completion, orange indicates negative results (i.e. no solubility or no DNA binding activity), yellow requires optimization, and grey indicate phases that have not yet been tested.

Table 3.2: *Summary of progress made with various C-terminal Saw1 constructs.*

	Expression/Solubility	Purification	DNA-binding
C1	BL21 pRARE - insoluble	Refolding NiCol, Anion X	No
C2	BL21 Star pRARELysS (20 ºC)	Needs troubleshooting	
C3	No		
C4	BL21 Star pRARE (30ºC)	Needs troubleshooting	

CHAPTER 4

RESULTS: PURIFICATION OF SAW1-RAD1-RAD10 COMPLEX

4.1 Co-expression and purification optimization of Saw1-His-Rad1-Rad10

To study the recruitment of Rad1-Rad10 to 3′-flaps via Saw1, Li *et al.* (2013) produced recombinant Saw1-Rad1-Rad10 by over-expression in *E. coli* and purification via IMAC. Although this preparation exhibited DNA-binding and nuclease activity, it was highly impure and had an apparent molecular weight much higher than expected for a trimeric complex. This preparation allowed for preliminary characterization of the complex, but characterizing the molecular determinants of nuclease recruitment to sites of repair will require protein of much higher quality. For any future study of this complex, we need to determine a method of recombinant Saw1-Rad1-Rad10 production that yields pure, stable, homogeneous, and active protein.

As described by Li *et al.* (2013), Saw1, His-Rad1, and Rad10 were co-expressed using a plasmid containing both His-Rad1 and Rad10 cDNA, and a plasmid containing tagless Saw1 cDNA. All three proteins were soluble when co-expressed (**Fig 4.1**). However, expression of the complex was highly inconsistent. This prevented us from further studying this complex.
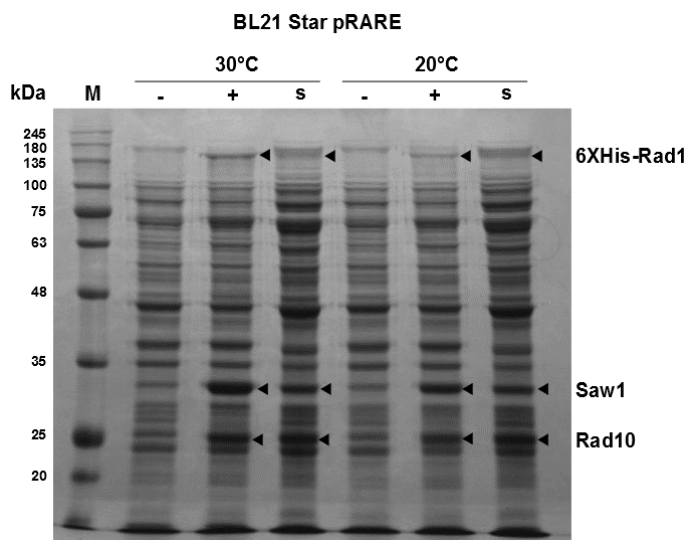
Figure 4.1: *Over-expression of soluble Saw1-His-Rad1-Rad10 in E. coli.*
Expression and solubility of Saw1-His-Rad1-Rad10 in BL21 Star pRARE expression cell line at two different induction temperatures. "-" represents proteins present before induction with IPTG, "+" represents the total expressed protein after induction with IPTG, and "s" represents the soluble fraction of proteins expressed after induction with IPTG. Black triangles indicate bands of interest. "M" represents the BluEye Protein Marker (GeneDirex) with relevant bands identified (in kDa) to the left.

To determine what factors influenced Rad1-Rad10 and Saw1 expression, we assessed the expression and solubility of (Saw1-)Rad1-Rad10 in BL21 Star pRARE cells under different conditions. First, we assessed the effect of media composition on protein expression. Of the five types of media (LB Miller, LB Lennox, SOB, TB, and SB), none showed expression of Rad1-Rad10. Furthermore, the concentration of IPTG used to induce expression had no influence on Rad1-Rad10 expression.

Cells transformed with the Rad1-Rad10 plasmid were grown in media with two antibiotics for selection, and a third for cells transformed with the plasmid carrying Saw1. Although we have used standard concentrations of antibiotics for culturing, it is possible that the presence of multiple antibiotics hinders initial cell growth. Lowering the concentrations of antibiotics could allow the cells to overcome this initial growth barrier. We compared the effect of using standard antibiotic concentrations (100 μg/mL Ampicillin, 30 μg/mL Kanamycin, 25

µg/mL Chloramphenicol) and low concentrations (50 µg/mL Ampicillin, 19.8 µg/mL Kanamycin, 15 µg/mL Chloramphenicol) on the expression of (Saw1-)Rad1-Rad10 and found that neither had an effect. It is possible that cells require multiple rounds of selection to further enrich for the antibiotic resistant and healthy cells. To this end, we compared the Rad1-Rad10 expression of cells subjected to one overnight growth to cells subjected to two overnight growths (the first overnight is used to seed the second), refreshing the media and antibiotics in between. This also had no effect on the expression, although two overnight growths produced more consistent cell preparations with respect to the growth rate of the cells. We therefore used this technique for cell culture preparation for expression.

We also investigated the integrity of the pAG 9068 (Rad1-Rad10 co-expression plasmid) and pAG 9069 (tagless Saw1) plasmids. Both inserts were sequenced and both were found to be correct. Next, we obtained fresh DNA from our collaborators (Dr. Jennifer Surtees, University at Buffalo) and transformed this into BL21 Star pRARE cells. We were then able to express Saw1-Rad1-Rad10, but only when we made fresh plasmid DNA stocks (purified from *E. coli* DH5α cells) before transforming. This was only a temporary solution and did not maintain expression for long.

We hypothesized that, overtime, leaky protein expression could be toxic to the cells and they cease to express the protein of interest, even when expression is induced with IPTG. We transformed BL21 Star pRARELysS cells with Saw1-, His-Rad1-, and Rad10-encoding plasmids. This cell line, along with expressing rare tRNAs, expresses lysozyme which is an inhibitor of T7 polymerase and limits leaky expression until it is induced by the addition of IPTG. These cells were able to express Saw1-Rad1-Rad10 and did not require fresh DNA preparations before each transformation.

4.2 Anionic exchange resolves two different populations of Saw1-Rad1-Rad10

Cell pellets containing Saw1-(His-Rad1)-Rad10 were purified in similar buffer conditions described by Li *et al.* (2013) through IMAC followed by IEC. In both separation steps, all three proteins eluted together, presumably as a complex. The anion exchange step yielded two peaks which were concentrated separately. Both peaks contained Saw1-Rad1-Rad10, where the first was pure and the second had many contaminating proteins (**Fig 4.2A**). The two peaks also differed in their particle size distribution as measured by DLS (**Fig 4.2B**). The first peak contained a more monodispersed distribution of smaller particles (radius of ~6 nm) and the second peak contained a much more polydispersed distribution of larger particles (radius of ~15 nm). When the first peak was concentrated, and applied to SEC, the complex eluted at a retention volume of 11.4 mL. This corresponds to an approximate molecular weight of ~282 kDa which is ~100 kDa larger than the theoretical size of the Saw1-Rad1-Rad10 heterotrimer (~180 kDa). This could be due to flexibility in the complex affecting its shape and, therefore, its retention in the column; or it could be indicative of a subunit composition that is not 1:1:1. The second peak eluted at ~8 mL (void volume of the column) indicating large oligomeric assemblies (>700 kDa) or aggregates (**Fig 4.2C**).

Any changes to the above protocol favoured production of a sample quality similar to that of the second peak, containing many more impurities, showing a polydisperse particle distribution, and resulting in the formation of large oligomers or aggregates (eluting in the void volume during SEC). These changes included implementing a step-elution to 213 mM imidazole (instead of a gradient) or increasing the NaCl concentration of buffers to 300 mM NaCl during IMAC (compared to 150 mM NaCl), or washing the column with 200 mM NaCl before eluting with a NaCl gradient during IEC (compared to 150 mM NaCl). The difference is also very clear

when looking at the particles under negative-stain electron microscopy (EM), specifically the overall size of particles, as well as the distribution of particles of different sizes. The former protocol yields smaller particles (approximately the expected size of the complex) of more uniform size, whereas the above-mentioned changes yield larger particles/aggregates that highly vary in size (data not shown).
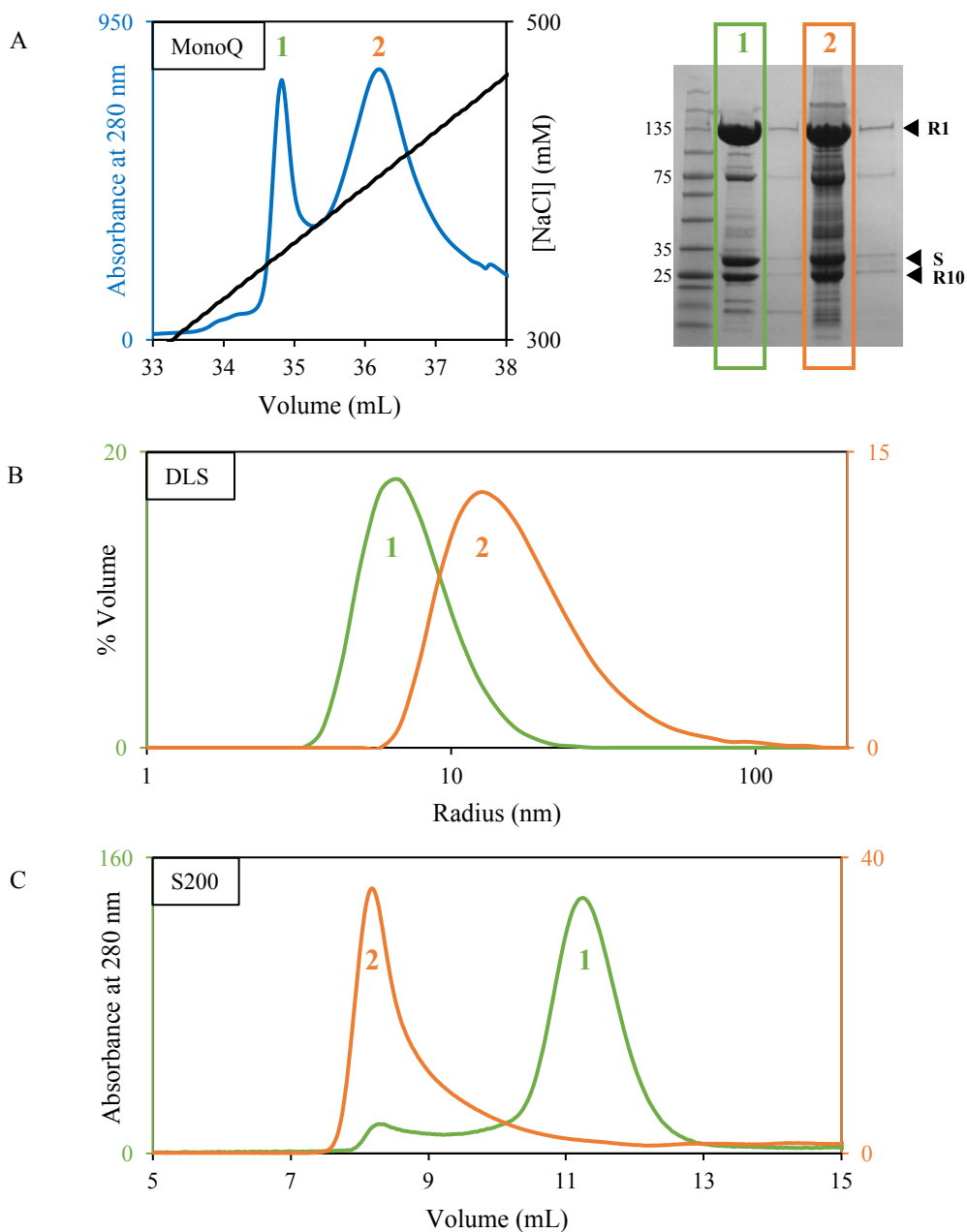
Figure 4.2: *Resolving two species of Saw1-Rad1-Rad10 through anion exchange chromatography.* (A) Anion exchange chromatography (MonoQ) elution profile (left) with two peaks. SDS-PAGE of the two peaks (right) where the green box indicates concentrated peak 1, and the orange box indicates concentrated peak 2 (lanes to the right of either of these boxes represent flow-through during concentrating, and representative bands from the BluEye protein ladder (GeneDirex) are labeled in kDa) Saw1, Rad1 and Rad10 are indicated by "S", "R1" and "R10", respectively. (B) DLS profiles of concentrated peak 1 (green) and peak 2 (orange). (C) Gel filtration elution (S200) profiles of peak 1 (green) and peak 2 (orange).

We assessed the DNA-binding specificity of the recombinant Saw1-Rad1-Rad10 from peak 1 by performing EMSAs with fluorescently labeled 60-mer ssDNA, 60-mer dsDNA, 60-mer splayed DNA, 60-mer 3′-flap DNA and 60-mer fork DNA substrates (see **Figure 2.1** for a description of the substrates). This Saw1-Rad1-Rad10 maintains its specificity for branched DNA substrates, in particular to splayed-Y and 3′-flap substrates, with low to no affinity for ssDNA and duplex DNA (**Fig 4.3A**).

Although Saw1-Rad1-Rad10 cannot be purified in large scale for crystallographic analysis, the complex is a relatively large (~180 kDa). Therefore, one approach to gain the molecular mechanism information of Saw1-Rad1-Rad10 recruitment to DNA is using electron microscopy (EM). Glycerol in sample buffers can interfere with EM visualization of macromolecules by reducing the sample contrast (Thompson *et al.*, 2016). We assessed whether Saw1-Rad1-Rad10 remains stable in a glycerol-free buffer for future structural investigation with EM. Through SEC, we determined that the protein complex is stable in glycerol-free buffer, and that it also retains stability after one freeze-thaw cycle (data not shown).

Altogether, preparations of Saw1-Rad1-Rad10 complex through this protocol yields pure and stable protein that has structure-specific DNA-binding activity, with a yield of ~ 50 mg per liter of cell culture.

4.3 Methods of improving sample homogeneity for biochemical studies

The improved quality of purified recombinant Saw1-Rad1-Rad10 allows for more rigorous biochemical characterization of the complex and its interaction with SSA repair intermediates. As part of our exploratory characterization, we visualized the complex using negative stain EM to gain insight into its general shape and architectural features when bound to a preferred DNA substrate. When we first viewed Saw1-Rad1-Rad10 incubated with a 60-mer

splayed-Y substrate using negative-stain EM, we noticed that the particle shape distribution was quite heterogeneous. This heterogeneity could be due to several things, including having a mixture of different particles within the sample, or possibly due to destabilization of the complex under the experimental conditions used. We therefore explored two techniques of improving the sample homogeneity for downstream structural studies: defining an optimal DNA substrate, and assessing the effect of buffer composition on stability.

4.3.1 Optimizing DNA substrate to improve sample homogeneity

Heterogeneity can arise from having a combination of protein-DNA complexes, protein alone, and DNA alone in the sample. We wanted to separate protein-DNA complexes from other species using SEC in order to increase the homogeneity of our sample.

Since splayed DNA showed the least mis-annealed products on a native gel, we decided that this would be the optimal substrate for analysis with EM. We chose the 60-mer splayed-Y DNA, as Saw1-Rad1-Rad10 had very high affinity for this substrate. Unfortunately, this substrate eluted at the same retention volume as Saw1-Rad1-Rad10 alone. Native gels of these fractions showed a mobility shift in the DNA band, indicating that the DNA was being bound by Saw1-Rad1-Rad10, but these fractions contained a mixture of protein-DNA complex, protein alone, and DNA alone. As a result, we looked for smaller splayed substrates that maintained binding to Saw1-Rad1-Rad10 and were small enough to be resolved from the protein peak. Saw1-Rad1-Rad10 bound 30-mer and 40-mer splayed DNA with only slightly less affinity than to the 60-mer splayed DNA (**Fig 4.3B**). These substrates both had increased retention volumes during SEC compared to that of Saw1-Rad1-Rad10 (and the 60-mer splayed DNA), and could therefore be resolved from this peak (**Fig 4.3C**). We decided that the 40-mer splayed substrate

was optimal since Saw1-Rad1-Rad10 had higher affinity for this substrate than for the 30-mer splayed DNA.
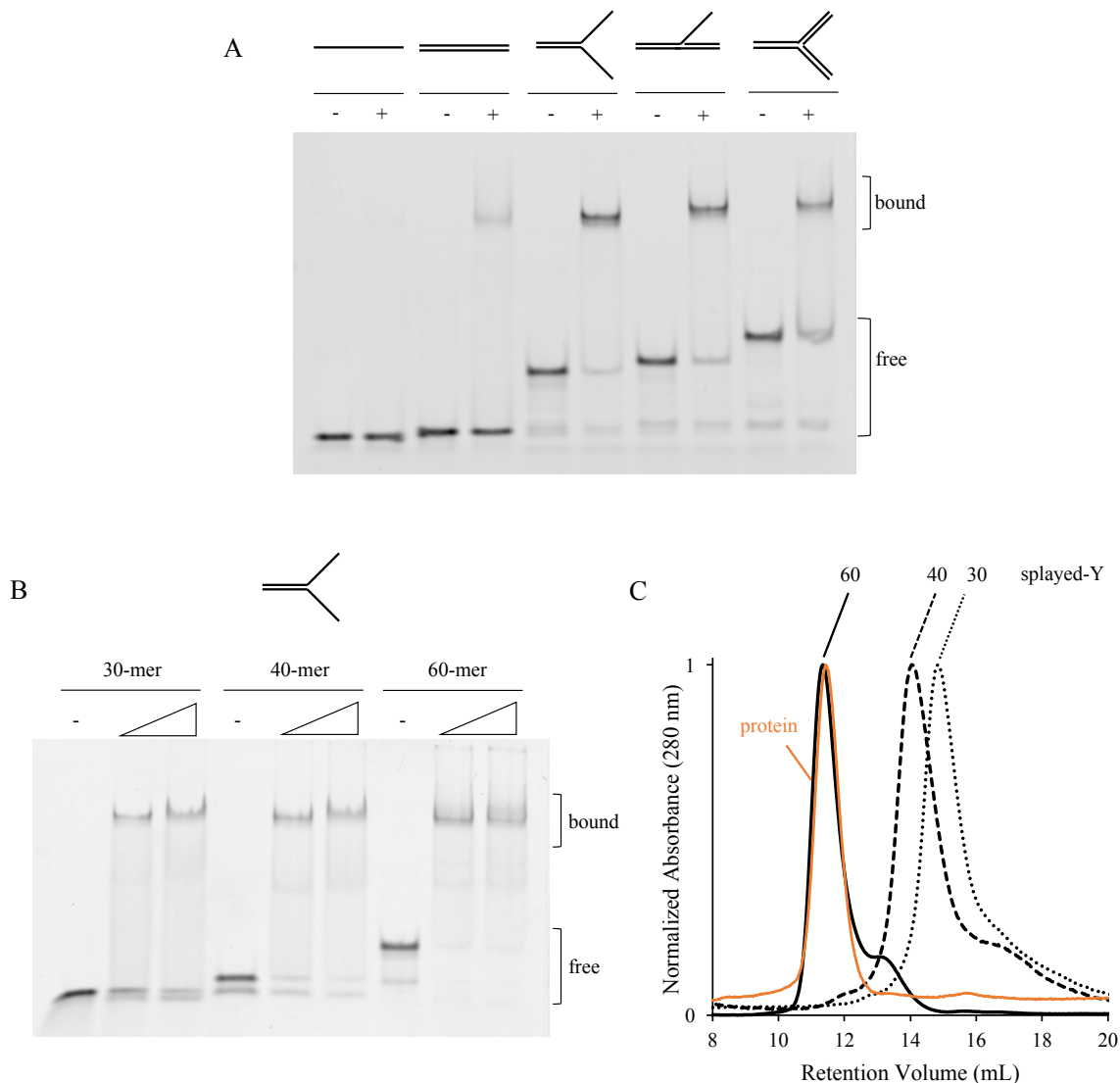
Figure 4.3: *Characterizing the structure specific binding of recombinant Saw1-Rad1-Rad10*
(A) EMSA of 60-mer ssDNA, 60-mer dsDNA, 60-mer splayed-Y DNA, 60-mer 3′-flap DNA, and 60-mer fork DNA, alone (-) or incubated with 40 nM Saw1-Rad1-Rad10 (+). Protein-bound and free DNA are indicated. (B) EMSA of 30-mer, 40-mer, and 60-mer splayed-Y DNA without (-), or with 40nM, or 80 nM Saw1-Rad1-Rad10. Protein-bound and free DNA are indicated. (C) Gel filtration (S200) elution profiles of 60-mer, 40-mer, and 30-mer splayed-Y DNA (black solid, dashed, and dotted lines, respectively). The elution profile of Saw1-Rad1-Rad10 is shown in orange. The absorbance of each curve has been normalized for ease of comparison of elution volumes.

4.3.2 Improving sample homogeneity with *ProteoPlex* buffer screening

The stability of macromolecules and macromolecular complexes greatly depends on the solvent pH (Chari *et al.*, 2015), as we have previously shown with Saw1 and the DSF-based *Thermofluor* assay (Chapter 3). Stability and monodispersity are imperative for biochemical studies, and are particularly important for structural studies. DSF is a powerful tool for improving the stability of monomeric proteins, but multi-subunit complexes often generate complex melting curves that are difficult to interpret (Chari *et al.*, 2015). To overcome this limitation, Chari *et al.* (2015) developed the *ProteoPlex* assay, which is a DSF-based comprehensive sparse-matrix screen of buffer formulation and pH for improving the stability and homogeneity of macromolecular complexes EM studies. Stabilizing buffers show both an increase in melting temperature (T_m) and a sharp transition between folded and unfolded states, demonstrating a two-state unfolding process.

To improve the homogeneity and stability of Saw1-Rad1-Rad10 in complex with the 40-mer splayed-Y substrate, we performed the *ProteoPlex* buffer screen as outlined in Section 2.11, **Table 2.4**. Overall, there was a clear relationship between the thermal stability of DNA-Saw1-Rad1-Rad10 and the pH, where the T_m of the complex decreased with increasing pH. The melting curves from the range of MMT buffer pH (from pH 5.6 to pH 8.4) show this very clearly (**Fig 4.4A**). MMT is a long-range buffering system and represents the general pH dependency of the thermal stability of the complex. Also, Tris-HCl buffers from pH 7.5 – 9.0 produced the lowest T_m of all buffers in the screen (**Table 4.1**). This includes Tris-HCl pH 7.9 and 8.1 which are essentially the buffer used previously in Saw1-Rad1-Rad10 purifications (Tris-HCl pH 8.0). The most stabilizing buffers ranged from pH 5.5 (Citrate) to pH 6.8 (HEPES), with Citrate pH 6.75 producing both the highest T_m and the sharpest unfolding transition (**Fig 4.4C, Table 4.1**).

Table 4.1 shows the 7 most stabilizing buffers (based on T_m) and their ranking in terms of both

T_m and unfolding transition, determined using the derivative curve of the thermal denaturation

curves (showing the inflection point/unfolding transition as a minima). The absolute value of the

height of the minima in the derivative curves reflects the "steepness" of the unfolding transition.

The greater the absolute value of the height of the minima, the "faster" the unfolding transition,

which is an indicator of complex stabilization.
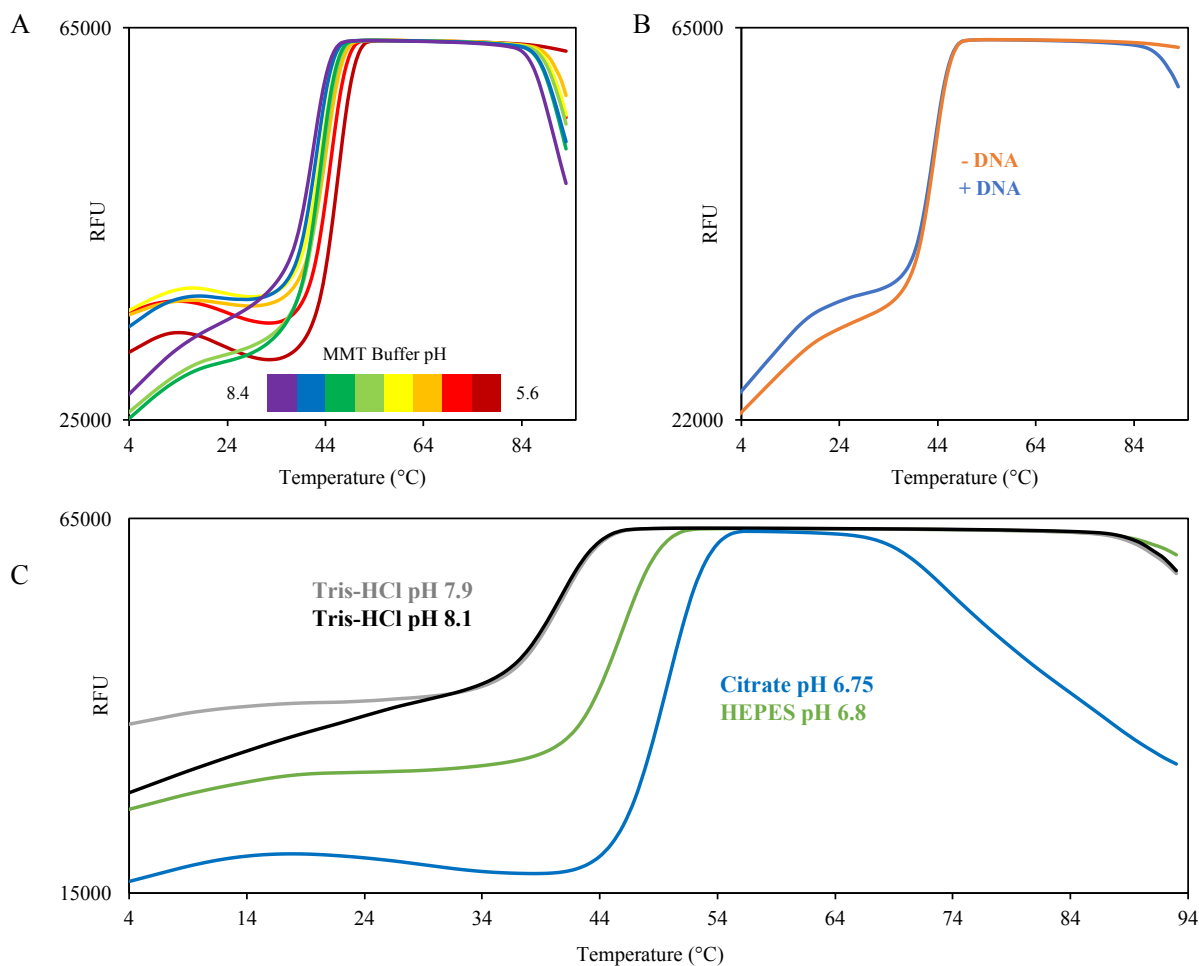
Figure 4.4: *Thermal stability of DNA-Saw1-Rad1-Rad10 in different conditions using ProteoPlex.*
(A) Melting curves of DNA-Saw1-Rad1-Rad10 in MMT buffer ranging from pH 5.6 (dark red) to 8.4 (dark blue). (B) Melting curves of Saw1-Rad1-Rad10 with (blue) and without (orange) pre-incubation with a 40-mer splayed-Y DNA substrate, in Tris-HCl pH 8 storage buffer. (C) Melting curves of DNA-Saw1-Rad1-Rad10 in destabilizing buffers Tris-HCl pH 7.9 (grey), Tris-HCl pH 8.1 (black), and stabilizing buffers HEPES pH 6.8 (green), and Citrate pH 6.75 (blue). "RFU" represents Relative fluorescence units.

Citrate, however, is not compatible with Nickel affinity chromatography (as previously discussed) and therefore another buffer is necessary for the initial purification steps. Although HEPES pH 6.8 is ranked 6th in terms of T_m, it is only within 1°C of the T_m of the second best SPG pH 5.6, and this is not a significant difference. Its unfolding transition is 4th in the ranking, behind citrate pH 6.75, SPG pH 6.0 (this T_m is ranked last), and SPG 5.6. Although it is lower in the ranking than several other buffers, HEPES pH 6.8 is a good candidate for Saw1-Rad1-Rad10 purification buffer since it is simple, inert, and compatible with various chromatography columns.

Table 4.1: *Comparison of 7 most stabilizing buffers for DNA-Saw1-Rad1-Rad10 from ProteoPlex*

Rank	by T_m	T_m	by Transition
1	Citrate pH 6.75	49.8	Citrate pH 6.75
2	SPG pH 5.6	47.8	SPG pH 6.0
3	Citrate pH 5.5	47.5	SPG pH 5.6
4	K Phosphate pH 5.8	47.5	HEPES pH 6.8
5	Na Phosphate pH 5.8	47.1	Citrate pH 5.5
6	HEPES pH 6.8	46.9	Na Phosphate pH 5.8
7	SPG pH 6.0	45.3	K Phosphate pH 5.8
	Tris-HCl pH 7.9	40.9	
	Tris-HCl pH 8.1	40.7	

Lastly, there was no difference in the melting temperature or unfolding transition between Saw1-Rad1-Rad10 alone and in complex with DNA in a Tris-HCl pH 8.0-based storage buffer (**Fig 4.4B**). This is possibly because there is no significant conformational change upon DNA binding, or that the relative difference is too small to observe, which could also account for gel filtration experiments where there is no difference in retention volume between protein alone and protein-DNA complex. It could also be due to nucleolytic cleavage and release of the substrate during the assay, or that this assay is not sensitive enough to detect conformational changes upon DNA-binding.

4.4 Sample preparation using optimized DNA substrate and optimized buffer

We combined the two sample improvement strategies discussed previously to produce a more stable and more homogeneous sample of protein-DNA complex for downstream biochemical and structural studies. Saw1-Rad1-Rad10 was purified in HEPES pH 6.8 buffer in a protocol similar to the one previously described. Then, after incubating Saw1-Rad1-Rad10 with 1.2X 40-mer splayed DNA, the complex was applied to SEC equilibrated with either a HEPES pH 6.8-based buffer or a Citrate pH 6.75-based buffer. Eluted fractions were analyzed via native gel to determine if the protein-DNA complex was forming. Although complex formation was seen in HEPES pH 6.8, the excess DNA seemed to interact with the column and only eluted after exchanging the buffer in the column from HEPES pH 6.8 to Citrate pH 6.75. HEPES pH 6.8 should therefore be avoided during SEC in case it has adverse effects on DNA binding. During SEC with Citrate pH 6.75 buffer, the free-DNA peak eluted as expected, and binding was also evident in the native gels as the Coomassie blue- and SYBR Gold-stained bands overlap, and show a mobility shift, indicating the presence of both protein and DNA in a higher molecular weight complex (**Fig 4.5A**). Fractions containing the protein-DNA complex were visualized with negative stain EM (**Fig 4.5B**) at a concentration of ~ 60 nM (~ 11 μg/mL). These images show no obvious improvement or difference in sample homogeneity when compared to images taken before optimizing the DNA substrate and sample buffer.
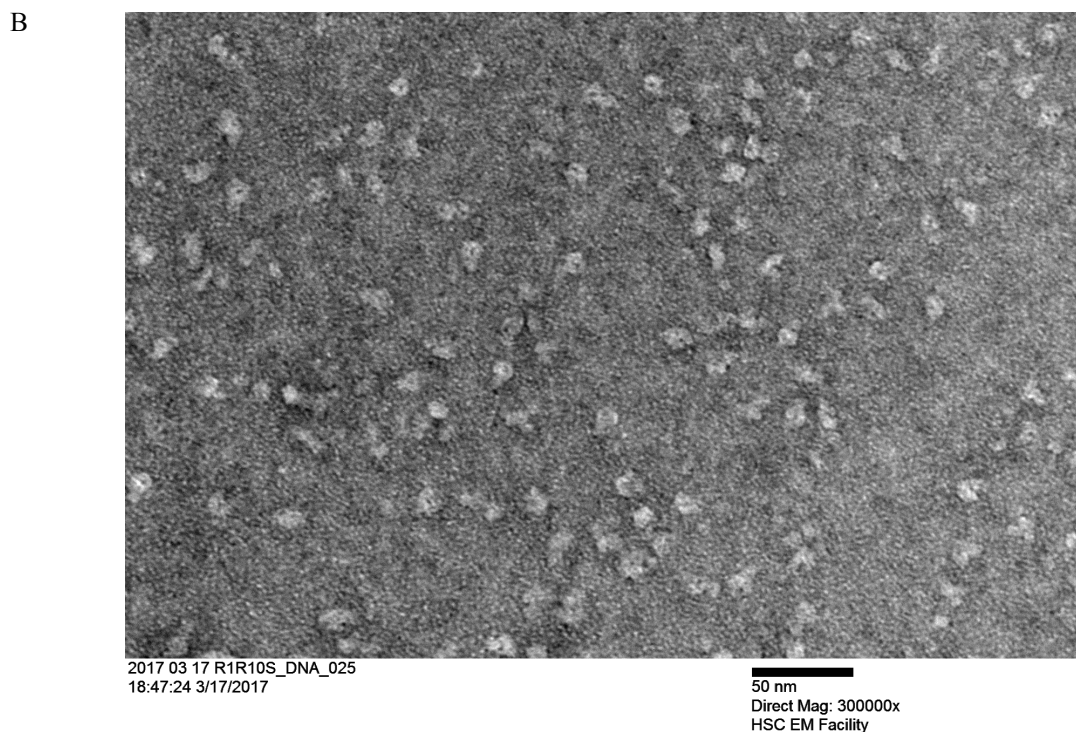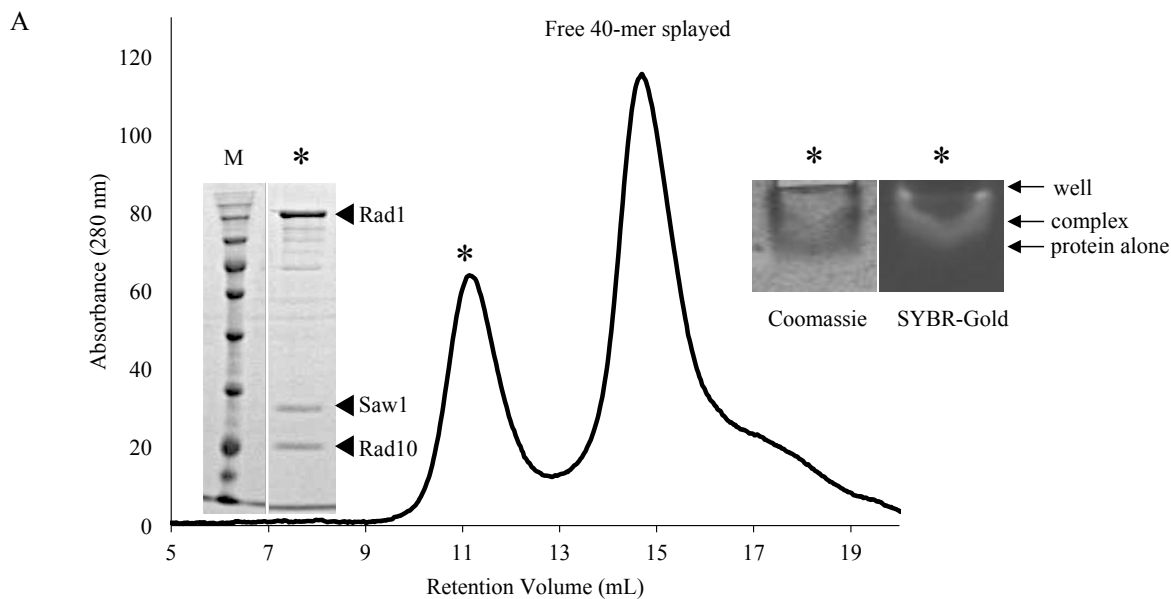
Figure 4.5: *Protein-DNA complex preparation for negative stain EM*
(A) Gel filtration elution profile of pre-incubated Saw1-Rad1-Rad10 and 40-mer splayed-Y DNA. Excess 40-mer splayed substrate is indicated. The fraction at the apex of the complex peak (*) was analyzed via denaturing (left) and native (right) PAGE. Native gels were stained with both SYBR-Gold for nucleic acid, and Coomassie Blue for protein. The well is indicated, as well as overlapping DNA/protein bands (complex), and protein alone bands. (B) Negative stain EM image of 1/5 dilution of this fraction (*) observed under 300,000X magnification on a JEOL TEMSCAN (Electron Microscopy Facility, Faculty of Health Sciences, McMaster University).

Although there is obvious binding in these fractions, there is still a significant amount of excess 40-mer splayed DNA (**Fig 4.5A**) and not all of the Saw1-Rad1-Rad10 is forming a complex, as seen in the native gel where the majority of the band stained by Coomassie only (protein alone) does not align with the slightly shifted band stained by both Coomassie and SYBR-Gold (protein-DNA complex) (**Fig 4.5A**). Since we cannot resolve protein-alone and protein-DNA peaks with SEC, this results in sample heterogeneity. Since we have defined the optimal DNA substrate for the complex using EMSA (**Fig 4.5**), we thought that this erroneous complex formation was possibly due to non-stoichiometric ratios of protein complex to DNA substrate. The stoichiometry of the nucleoprotein complex has not been elucidated and, for this reason, we wanted to identify the optimal protein to DNA ratio that favours the DNA-bound state of Saw1-Rad1-Rad10.

To do this, Saw1-Rad1-Rad10 (at constant concentration) was incubated with varying concentrations of 40-mer splayed-Y DNA and applied to SEC (Superose 6 Increase 10/300 GL, GE Healthcare). Complex formation was monitored through changes in the absorbance ratio at 260 nm (A260, representing DNA) and 280 nm (A280, representing protein) (**Fig 4.6**). Ratios of protein to DNA (protein:DNA) that were tested include 3:1, 2:1, 1:1, 1:2, 1:4, 1:6, compared to Saw1-Rad1-Rad10 alone (1:0). In all cases, a large excess of DNA was still observed, and A260/A280 ratios did not increase significantly with excess Saw1-Rad1-Rad10 (**Fig 4.6A**) or with excess splayed DNA (**Fig 4.6B**), compared to the 1:1 ratio. In fact, the A260/A280 ratio remained similar to that of Saw1-Rad1-Rad10 alone. Only when a significant excess of DNA was added (1:4, 1:6) did we see more substantial (but still very small) increases in the A260/A280 ratio (**Fig 4.6B**). However, in these samples, we also began to see evidence of potential concentration-dependent, non-specific interactions between molecules in the sample

(peaks representing larger DNA species, indicated with black arrows in **Fig 4.6B** inlet). Therefore, these changes in A260/A280 could have been due to concentration-dependent artifacts in the experiment. With this experimental set-up, there was no indication that complex formation could be favoured by changing the ratio of protein to DNA.
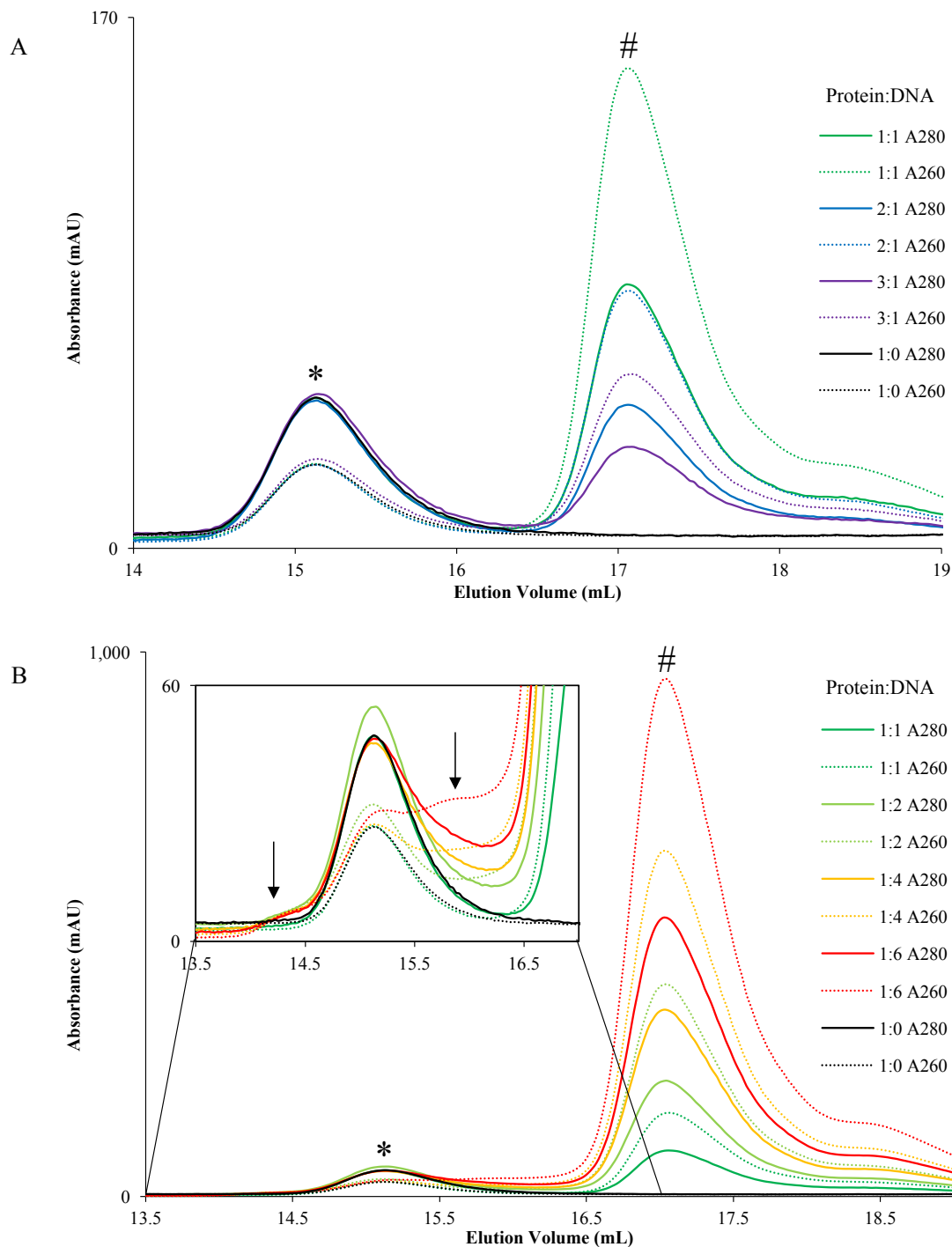
Figure 4.6: *Changing the stoichiometry of the nucleoprotein complex does not affect DNA-binding*
(A) SEC elution profile of molar excesses of Saw1-Rad1-Rad10 incubated with 40-mer splayed-Y DNA.
(B) SEC elution profile of molar excesses of 40-mer splayed-Y DNA incubated with Saw1-Rad1-Rad10.
Inlet shows peaks representing larger molecular weight DNA species (arrows) that appear with significant
excess DNA (1:4 and 1:6, protein:DNA). * represents Saw1-Rad1-Rad10 peak, and # represents 40-mer
splayed-Y DNA alone peak.

CHAPTER 5

CONCLUDING REMARKS AND FUTURE DIRECTIONS

5.1 Intrinsic instability of Saw1 in the absence of Rad1-Rad10

The apparent instability of over-expressed and purified recombinant Saw1 is representative of its *in vivo* role as a molecular scaffold. In the context of 3′-tail removal, Saw1 is normally found in complex with the Rad1-Rad10 nuclease and is unstable in its absence (Li *et al.*, 2013). This complex is strong as it is resistant to three consecutive chromatography steps and remains as a complex during ultracentrifugation with a 100 kDa cut-off membrane. Hydrophobic interfaces are a common feature of strong protein-protein interactions (Nooren and Thornton, 2003) and Saw1 likely interacts with Rad1 through one such interface. Recombinant Saw1 is likely intrinsically unstable in solution due to exposure of hydrophobic surfaces, making it difficult to purify on a large scale. These exposed hydrophobic surfaces also contribute to protein aggregation at higher concentrations.

In other damage contexts, SUMOylation of Saw1 weakens its interaction with Rad1 but strengthens its interaction with Slx4. This allows it to exchange one binding partner for another to respond to different DNA damage situations (Sarangi *et al.*, 2014). This indicates that, although it is not constitutively interacting with Rad1-Rad10, Saw1 may always be in complex with another protein or protein complex, and is unstable on its own.

It is also possible that Saw1 requires Rad1 as a scaffold for correct folding. This would explain the absence of Saw1 when not in complex with Rad1-Rad10. In the human system, the ERCC1 subunit requires the presence of the XPF subunit as a scaffold for folding and is otherwise unstable (Tripsianes *et al.*, 2005). Therefore, it is possible that the yeast Rad1 subunit

acts as a surface to facilitate proper folding of its interacting partners and, in its absence, these proteins are unstable.

Since there is no information on the structure of Saw1, we initially performed secondary and tertiary structure predictions with PSIPRED and Phyre[2] software, respectively. Secondary structure revealed several long stretches of unstructured regions, indicating intrinsic flexibility in the protein. Tertiary structure prediction did not yield any confident predictions, indicating that this protein has no known or common structural folds and is most likely highly flexible. The instability index, determined through the ProtParam tool (ExPASy), classified Saw1 to likely be unstable. Altogether, *in silico* analysis has coincided with our experimental observations, where Saw1 alone is unstable and difficult to purify. To combat this instability, we have performed high through-put screening of buffer conditions that stabilize Saw1 in the absence of its binding partners, and have successfully identified optimal conditions for large scale production of stable protein.

5.2 Discrepancies between DLS- and DSF-based buffer screens

Through this work, both DLS and DSF have proven to be useful tools in optimizing the large-scale production of recombinant proteins. Notably, both techniques allowed us to characterize the general pH dependency of His-Saw1 stability. However, there are two cases where the results of either screen differed.

First, the *Thermofluor* assay showed no change in the stability of His-Saw1 with the addition of reducing agents (βME, DTT, TCEP). However, His-Saw1 aggregated and later precipitated when the concentration of βME was reduced from 5 mM to 1.4 mM during SEC. Initial additives screening with DLS reflected this phenomenon, as the PDI decreased when

increasing the concentration of βME from 1.4 mM to 5 or 10 mM. This discrepancy between the two assays reflects the concentration of protein used in each case. The SEC experiment and DLS screening use relatively high concentrations of protein, compared to the *Thermofluor* assay. The greater concentration increases the potential for intermolecular interactions between particles in the sample. Saw1 has two cysteine residues which, if they are surface exposed, can be oxidized to form intermolecular disulfide bonds. At higher concentrations of protein and lower concentrations of reducing agent, this can lead to aggregation of Saw1. Therefore, it is necessary to maintain a reducing environment during large scale production of Saw1 to prevent this aggregation.

DLS also showed that lower salt concentrations decrease the polydispersity of His-Saw1. This result is unexpected, since proteins are often more stable at higher salt concentration. Also, the decrease in PDI is negligible at 0.4 M NaCl, decreases by a significant amount at 0.3 M NaCl, and decreases a small amount at 0.2 M NaCl. These results do not follow a distinct trend, and thus make it difficult to confidently assess the effect of ionic strength on His-Saw1 with DLS. Conversely, *Thermofluor* showed a clear and consistent trend where increasing the concentration of NaCl enhanced the thermal stability of His-Saw1. This result corroborates the observation that diluting the salt concentration between IMAC and IEC purification steps causes some precipitation. Here is an instance where, although the concentration of protein used in the *Thermofluor* assay is relatively low, the clear trend in salt-dependent stability gives us confidence that this result is true, similar to the cases of pH and imidazole.

These two examples illustrate the variability of results based on the differences in concentration between DLS and DSF experiments. DSF is extremely useful for initial characterization of general buffer components like buffer pH and formulation as well as salt,

when only small amounts of recombinant protein can be produced. This allows for preliminary optimization of purification buffers for increasing yields of protein. In the case of reducing agents and additives that stabilize proteins through attenuating intermolecular interactions, it is perhaps better to assess their effect on stability through DLS at a concentration more similar to actual working conditions. Taken together, these methods are complimentary to each other and provide an effective and efficient means for optimizing large scale production of recombinant protein.

5.3 Multiple approaches for structural characterization of Saw1-DNA interactions

Since structural studies of flexible proteins can be difficult, we have taken a 3-pronged approach for our investigation of the structural determinants of Saw1 recruitment to SSA intermediates. In parallel, we are exploring the use of full length *S. cerevisiae* (Sc) Saw1, truncations of Sc Saw1, and fungal homologs of Saw1 to maximize the probability of success in structural investigation of this interaction.

5.3.1 Substrate design for Saw1-DNA complex crystallization

First and foremost, we focused on optimizing the production of high quantity and quality full length His-Saw1. Now that production of recombinant Saw1 is amenable to structural study using X-ray crystallography (Rashev *et al.*, 2017), we were able to characterize the minimal branched DNA substrates to which Saw1 maintains binding. The minimal 3′-flap substrate that we found agrees with Li *et al.* (2013) who also determined that Saw1 did not have observable affinity for structures with 3′-flaps shorter that 10-nt. These minimal-length experiments reflect how Rad1-Rad10 is dispensable for flap removal *in vivo* when flaps are short (less than 30 nt), in

which case they are thought to be removed through the proof-reading activity of polymerase δ (Paques and Haber, 1997).

Using these lower limits for substrate length, we can now perform high-throughput crystallization experiments with branched DNA substrates ranging from 20-mer to 40-mer splayed-Y or 3′-flap DNA. While the 30-mer and 40-mer splayed-Y substrates yielded no crystals in preliminary screens, our approach will be to continue screening many different size variations of the splayed substrate within the 30-mer to 40-mer range. By independently altering the length of the duplex, 3′-, and 5′-tails of this substrate, we can determine an optimal substrate for crystallization.

5.3.2 Assessing the DNA-binding ability of C-terminal constructs of Saw1

Our second approach was to design Saw1 truncations that may be more amenable to crystallography in the absence of its stabilizing partners, Rad1-Rad10, using limited proteolysis coupled with affinity chromatography. We have designed several C-terminal Saw1 constructs and are systematically probed for expression/solubility in *E. coli*, ability to be purified, and affinity for branched DNA. This approach will also give insight into the boundaries of the minimal region of Saw1 required for structure-specific DNA-binding.

5.3.3 Use of Saw1 homologs for enhanced stability for structural studies

Our third approach is the use of homologs, as slight differences in the surface residues may affect the stability of the homolog, making it more amenable to crystallography (Savchenko *et al.*, 2003). Saw1 homologs have been found in other fungal species. Although these have not been biochemically or genetically characterized, they share high sequence homology with budding yeast Saw1. In particular, their proposed DNA-binding motifs are highly conserved (Li

et al., 2013). Therefore, alongside our investigations with full length budding yeast Saw1 and C-terminal constructs, we have also selected three fungal homologs of Saw1 from a fungal alignment of Sc Saw1 with predicted orthologs in other fungal species, performed through the *Saccharomyces* Genome Database. These include Saw1 from *S. mikatae* (Sm), *S. paradoxus* (Sp), and *S. casteilli* (Sca). These were selected due to the high homology and the similarities between the predicted secondary structure (PSIPRED) of the orthologs compared to Sc Saw1.

So far, expression and solubility have been assessed for each homolog in *E. coli* BL21, BL21 Star, BL21 Star pRARE, and BL21 Star pRARELysS cells, during induction at 16°C, 20°C, and 30°C. Sm Saw1 was insoluble in all of the above cell lines over the range of temperatures assayed. Sp Saw1 was slightly soluble in BL21 Star pRARE cells induced at 16°C. Sca Saw1 was slightly soluble in BL21 Star pRARE cells in all temperatures and was soluble in BL21 Star pRARELysS cells induced at 16°C and 20°C. Purifications have not been attempted for any of the Saw1 homologs, and this line of investigation is still ongoing.

Using these techniques as well as performing crystallization screens of different DNA substrates with full-length Sc Saw1, we hope to increase our chances of structure determination of Saw1 for insight into its interaction with branched DNA substrates.

5.4 Separation of two populations of recombinant Saw1-Rad1-Rad10

Through initial purifications of Saw1-Rad1-Rad10, we were able to isolate two different populations of the complex during IEC. The first peak that eluted from the column contained a population of Saw1-Rad1-Rad10 that had a radius of ~ 6 nm (DLS), eluted as a complex of ~282 kDa, and showed a more uniform distribution of smaller particles in negative-stain EM. Conversely, the second peak contained larger particles (~ 15 nm radius) which eluted in the void

volume of a Superdex 200 SEC column, indicating a particle size greater than 669 kDa (size of thyroglobulin, function test of Superdex 200). This population reflects the finding by Li *et al.* (2013) that Saw1-Rad1-Rad10 co-eluted during SEC with a corresponding molecular weight of 500-700 kDa. This was likely a mixture of the two populations we are now able to separate. This separation is extremely important and necessary for biochemical and structural analysis of this complex.

This mixture in the Saw1-Rad1-Rad10 population is most likely due to a stability issue during expression or purification of the complex, causing it to aggregate. As we have previously seen in the case of Saw1, the stability of recombinant proteins can be highly dependent on the pH and formulation of the purification buffers. Through the DSF-based Proteoplex assay, we also saw a similar pH-dependency of the Saw1-Rad1-Rad10 complex, where stability of the complex increased with decreasing pH and was optimal at around pH 6. If the observed aggregation was due to buffer-dependent instability of the complex, we would expect that, by using a more optimal buffer, we would see a decrease in the aggregated population (peak 2) during IEC. However, no difference in the distribution of different populations of Saw1-Rad1-Rad10 was seen when purification buffers were changed to HEPES pH 6.8. This indicates that the pH of the buffer was not responsible for the apparent aggregation of the complex.

Aggregation can also occur due to changes in salt concentration before applying recombinant protein to IEC. However, IMAC fractions containing Saw1-Rad1-Rad10 were simply pooled and applied directly to IEC with no dilution or other alteration. This indicates that the aggregated complex (seen in peak 2) is also present during or just after IMAC. As we have seen with Saw1, the imidazole concentration used to elute recombinant His-tagged proteins can adversely affect the stability of the protein. This may also be the destabilized by imidazole

during IMAC, and produces the aggregated population seen in the second peak of IEC. This can be assessed by DSF-based *Thermofluor* screen where the effect of imidazole and other buffer additives on the stability of the complex can be assessed. Normally, complexes formed of several subunits tend to be difficult to analyze via DSF since they can have complex melting curves (Boivin *et al.*, 2013). However, as seen in the *Proteoplex* assay, Saw1-Rad1-Rad10 demonstrates a two-state melting transition, indicating that this complex would be amenable to high throughput DSF-based screens of buffer additives.

It is also possible that this aggregated population arises from initial steps of purification, such as cell lysis via sonication. Sonication for cell lysis also shears chromosomal DNA to which proteins with affinity towards DNA may bind. This shearing prevents the removal of DNA-binding proteins from soluble cell lysates during the process of pelleting cell debris as well as chromosomal DNA. It is possible that the sonication process is too harsh and is causing a subset of the Saw1-Rad1-Rad10 complex to over-heat and aggregate. It is also possible that the DNA shearing during sonication is insufficient and is leaving larger DNA fragments to which Saw1-Rad1-Rad10, and potentially other endogenous *E. coli* proteins, stay bound. This would account for the increased impurity seen in peak 2, as well as the larger size and variation in these particles. It may therefore be necessary to use enzymatic or chemical digestion of chromosomal DNA during cell lysis to efficiently cleave chromosomal DNA into more appropriate fragment sizes. During and after sonication, the lysate remains highly viscous and this could be indicative of insufficient chromosomal DNA shearing.

Although sonication is desirable for its ability to shear chromosomal DNA, the high-frequency sound waves emitted by the instrument heats the solution. We limit this heating by lysing cells with short pulses while keeping the cells on ice, but heat denaturation of proteins in

the lysate can still occur and lead to protein aggregation or precipitation. Other mechanical methods of cell disruption, including Homogenizers and French press, pass cells through small spaces to shear cell membranes and do not produce heat during this process (Middelberg, 1995). Although they do not facilitate DNA shearing, these methods of cell lysis can be coupled with DNaseI treatment to digest chromosomal DNA. Together, they present possible alternatives to sonication that prevents destabilization of proteins during lysis and effectively degrades DNA.

Finally, these two populations could arise during expression. Decreasing the rate of expression (for example, through reduced incubation temperatures) is known to aid in the solubility and stability of recombinant proteins. It is possible that the current protein expression parameters are partially influencing the stability of the recombinant Saw1-Rad1-Rad10 complex. Since this process has already been very difficult to optimize, all other avenues should be investigated (buffer additives, cell lysis protocols) before attempting to optimize the expression and solubility of the complex in *E. coli* expression cell lines.

5.5 Optimizing sample homogeneity of Saw1-Rad1-Rad10 for future structural study

Any heterogeneity in samples for structural investigation can prevent determination of high-resolution structures. When first viewing complexes of Saw1-Rad1-Rad10 (from IEC peak 1) and DNA with negative-stain EM, the distribution of particles seemed heterogeneous, in terms of the shape of the particles. In an attempt to further improve the homogeneity of DNA-Saw1-Rad1-Rad10 complexes, we have optimized various aspects of sample preparation and monitored their effect on homogeneity through visualization with negative-stain EM.

5.5.1 Complex formation and isolation: Effects of substrate size and concentration

First, we sought to establish a protocol for setting protein-DNA complex to isolate the complex from all other species (unbound protein, unbound DNA) using gel filtration. Although we are at the moment unable to separate nucleoprotein complex from protein alone, we have found that 40-mer splayed-Y substrates (or smaller) allow for separation of unbound DNA from higher molecular weight species (**Fig 4.3C**).

We also consistently see a large excess of free DNA during SEC of the protein-DNA complex. Although EMSAs with Saw1-Rad1-Rad10 showed strong affinity towards the 40-mer splayed-Y substrate (using nanomolar concentrations), this complex is known to be dispensable for shorter flaps *in vivo*. It is possible that larger splayed-Y substrates may be required (> 60-mer splayed-Y) in order to separate protein-DNA complexes from protein alone and DNA alone with SEC.

Or, it is possible that the molar ratios of protein:DNA need to be optimized in order to favour complex formation. Initial attempts have been made to this end, where different ratios of protein to DNA have been incubated and run through SEC, monitoring complex formation through changes in the A260/A280 ratio (**Fig 4.6**). However, with this experimental set-up, we were unable to see any distinct differences in complex formation over the range of protein:DNA ratios tested. Another SEC-based method for determining appropriate molar ratios for DNA-binding could be done by monitoring the disappearance of the free-DNA peak upon incubating with varying concentrations of protein. This does not depend on the A260/A280 ratio, and this method reflects the experimental set-up of the EMSAs which showed high affinity for this splayed DNA substrate.

5.5.2 Buffer-dependent stability of Saw1-Rad1-Rad10 and its effect on heterogeneity

The stability, and thus homogeneity of proteins can be highly dependent on the pH and formulation of purification buffers, as we have seen with recombinant Saw1. We therefore explored buffer conditions that stabilize the DNA-Saw1-Rad1-Rad10 complex using a sparse-matrix DSF screen described by Chari *et al.* (2015). The *ProteoPlex* assay showed a significant pH-dependency in the thermal stability of the complex. This relationship is identical to that of Saw1 in the *Thermofluor* assay, where stability increases with decreasing pH and is optimal at around pH 6. However, protein prepared with the optimized buffers from this screen (IMAC and IEC in HEPES pH 6.8, SEC in Citrate pH 6.75) showed no significant difference in the level of heterogeneity of the protein-DNA complex through visualization with negative-stain EM. Most of the examples reported by Chari *et al.* describe protein complexes that are initially aggregated (or partially aggregated) or broken/fragmented and are optimized to a more homogeneous and stable distribution of particles. These examples often show a difference in the melting curves between the pre- and post-optimization conditions that go from a complex, multi-step melting transition to a two-step melting transition, respectively. Even at Tris-HCl pH 8, DNA-Saw1-Rad1-Rad10 demonstrates a two-step transition state. It is possible that the heterogeneity in the Saw1-Rad1-Rad10 sample cannot be improved upon to the same extent as many of the examples given in the *ProteoPlex* report.

Chari *et al.* (2015) also gave an example of sample optimization where the buffer drastically improves the sample quality, but the melting curves and T_m determined through *ProteoPlex* do not change. It is therefore possible that, although we see a high T_m, these buffers may not have a significant effect on the heterogeneity of the complex. Ultimately, it is necessary

to assess the heterogeneity of the complex with negative-stain EM using several of the optimal buffers determined through the *ProteoPlex* screen.

Improvement in the T_m and the "sharpness" of the melting transition was seen in several buffers around pH 6, specifically in Citrate pH 6.75. Interestingly, Citrate pH 6.75 did not follow the same trend as the rest of the Citrate buffers tested, where Citrate pH 5.5 showed the greatest T_m and Citrate pH 6.5 showed the lowest T_m. It is possible that Citrate was not a real stabilizer of the complex and is an outlier. The next best stabilizing buffer was SPG pH 5.6, and it would be worth assessing if this buffer has an effect on the level of heterogeneity of Saw1-Rad1-Rad10.

5.5.3 Conformational heterogeneity due to efficient nuclease activity

It is also possible that the heterogeneity seen with negative-stain EM was not due to buffer-dependent destabilization, but rather from a mixture of Saw1-Rad1-Rad10 in varying states of DNA binding (conformational heterogeneity). When fractions containing 40-mer splayed-Y DNA-Saw1-Rad1-Rad10 complex eluted from SEC were electrophoresed through 5% non-denaturing TG-gels, there was evidence of free-DNA that had similar mobility to that seen in DNA-alone fractions. Since we were able to efficiently separate Saw1-Rad1-Rad10 from the 40-mer splayed-Y DNA, this was unexpected. There are two explanations for the presence of free DNA in these fractions.

Firstly, it's possible that the interaction between protein and DNA is dynamic or weak. This is unlikely since EMSAs showed a distinct band shift, indicative of a stable interaction. The other possibility is that Saw1-Rad1-Rad10 is binding and cleaving the splayed-Y substrate and releasing a 5′-tailed product which migrates approximately the same distance as the splayed-Y substrate on non-denaturing PAGE. Interestingly, the retention volume of the splayed-Y DNA alone (**Fig 4.3C**) and that of the "free-DNA" peak during sample preparation (**Fig 4.5A**) from

SEC differ slightly (14.04 mL versus 14.46 mL, respectively). If, during incubation prior to separation through SEC, Saw1-Rad1-Rad10 is able to efficiently bind, cleave and release its substrate, the free-DNA peak we see could represent a 5′-tailed product of endonucleolytic cleavage. However, these two SEC experiments were done in Tris-HCl pH 8 and Citrate pH 6.75 buffers, respectively. This difference in pH can possibly affect the physical properties of the substrate and the difference in retention volume observed can be due to this effect. This theory is also reinforced by the lack of shift in T_m between protein alone, and protein incubated with DNA in the *ProteoPlex* experiment (**Fig 4.4B**). Upon DNA binding, we would expect a conformational change that may have an effect on the T_m and this is not seen in this experiment.

We do not see indication of cleavage in the EMSAs of Saw1-Rad1-Rad10 with 40-mer splayed DNA, but this is most likely due to the relatively high EDTA concentration used in these experiments (5 mM EDTA). Since Rad1-Rad10 is a metal-dependent nuclease, EDTA can chelate the catalytic metal, removing it from the active site, and thus inhibit nuclease activity. Only 0.1 mM EDTA is used in the SEC and *ProteoPlex* experiments, and a much higher protein concentration is present in these samples. It is therefore possible that the nuclease is active and cleaving the substrate. This can account for the heterogeneity seen in the negative-stain EM samples. To test this theory, SEC can be performed in a range of EDTA concentrations where, at high enough concentrations, nuclease activity should be inhibited and this, in turn should improve the large-scale DNA-binding and complex formation for structural studies. Other ways to prevent cleavage include using a nuclease-dead variant of Rad1, or using a non-hydrolysable substrate.

5.6 Final remarks

This body of work has allowed us to produce the quality and quantity of recombinant Saw1 and Saw1-Rad1-Rad10 complex that are required for biochemical and structural investigation, which would otherwise be near impossible due to low yields, instability, and polydispersity. Although DNA-Saw1-Rad1-Rad10 complex formation is still in need of optimization, we have outlined several possible routes to ameliorate sample homogeneity for eventual structural analysis with cryo-EM. We now have the tools to explore the protein-protein and protein-DNA interactions that are determinants for the structural specificity of this nuclease complex. This work brings us closer to understanding the mechanism of Rad1-Rad10 recruitment to non-homologous DNA flaps by Saw1, as well as the mechanism of flap-removal and how Saw1 stimulates this process in yeast. Although there have been several structures of the XPF/Rad1 family of nucleases, there has yet to be a structure of a catalytic complex to show how structure selectivity and endonucleolytic cleavage of splayed and 3′-flap structures are carried out. These improvements in sample preparation pave the way for the biochemical and structural characterization of the Saw1-Rad1-Rad10 complex and they will, in turn, provide mechanistic insight into the repair of double-strand breaks by single-strand annealing.

REFERENCES

Abbott, D. W., Freeman, M. L. and Holt, J. T. (1998). "Double-strand break repair deficiency and radiation sensitivity in BRCA2 mutant cancer cells." J Natl Cancer Inst **90**(13): 978-985.

Adair, G. M., Rolig, R. L., Moore-Faver, D., Zabelshansky, M., Wilson, J. H. and Nairn, R. S. (2000). "Role of ERCC1 in removal of long non-homologous tails during targeted homologous recombination." Embo j **19**(20): 5552-5561.

Ahmad, A., Enzlin, J. H., Bhagwat, N. R., Wijgers, N., Raams, A., Appledoorn, E., Theil, A. F., JH, J. H., Vermeulen, W., NG, J. J., Scharer, O. D. and Niedernhofer, L. J. (2010). "Mislocalization of XPF-ERCC1 nuclease contributes to reduced DNA repair in XP-F patients." PLoS Genet **6**(3): e1000871.

Ahmad, A., Robinson, A. R., Duensing, A., van Drunen, E., Beverloo, H. B., Weisberg, D. B., Hasty, P., Hoeijmakers, J. H. and Niedernhofer, L. J. (2008). "ERCC1-XPF endonuclease facilitates DNA double-strand break repair." Mol Cell Biol **28**(16): 5082-5092.

Al-Minawi, A. Z., Saleh-Gohari, N. and Helleday, T. (2008). "The ERCC1/XPF endonuclease is required for efficient single-strand annealing and gene conversion in mammalian cells." Nucleic Acids Res **36**(1): 1-9.

Anand, R., Ranjha, L., Cannavo, E. and Cejka, P. (2016). "Phosphorylated CtIP Functions as a Co-factor of the MRE11-RAD50-NBS1 Endonuclease in DNA End Resection." Mol Cell **64**(5): 940-950.

Aravind, L., Walker, D. R. and Koonin, E. V. (1999). "Conserved domains in DNA repair proteins and evolution of repair systems." Nucleic Acids Res **27**(5): 1223-1242.

Bailly, V., Sommers, C. H., Sung, P., Prakash, L. and Prakash, S. (1992). "Specific complex formation between proteins encoded by the yeast DNA repair and recombination genes RAD1 and RAD10." Proc Natl Acad Sci U S A **89**(17): 8273-8277.

Bardwell, A. J., Bardwell, L., Johnson, D. K. and Friedberg, E. C. (1993). "Yeast DNA recombination and repair proteins Rad1 and Rad10 constitute a complex in vivo mediated by localized hydrophobic domains." Mol Microbiol **8**(6): 1177-1188.

Bardwell, A. J., Bardwell, L., Tomkinson, A. E. and Friedberg, E. C. (1994). "Specific cleavage of model recombination and repair intermediates by the yeast Rad1-Rad10 DNA endonuclease." Science **265**(5181): 2082-2085.

Bardwell, L., Cooper, A. J. and Friedberg, E. C. (1992). "Stable and specific association between the yeast recombination and DNA repair proteins RAD1 and RAD10 in vitro." Mol Cell Biol **12**(7): 3041-3049.

Barlow, J. H. and Rothstein, R. (2010). "Timing is everything: cell cycle control of Rad52." Cell Div **5**: 7.

Belov, O. V., Krasavin, E. A., Lyashko, M. S., Batmunkh, M. and Sweilam, N. H. (2015). "A quantitative model of the major pathways for radiation-induced DNA double-strand break repair." J Theor Biol **366**: 115-130.

Bennardo, N., Cheng, A., Huang, N. and Stark, J. M. (2008). "Alternative-NHEJ is a mechanistically distinct pathway of mammalian chromosome break repair." PLoS Genet **4**(6): e1000110.

Bennardo, N., Gunn, A., Cheng, A., Hasty, P. and Stark, J. M. (2009). "Limiting the persistence of a chromosome break diminishes its mutagenic potential." PLoS Genet **5**(10): e1000683.

Bertrand, P., Tishkoff, D. X., Filosi, N., Dasgupta, R. and Kolodner, R. D. (1998). "Physical interaction between components of DNA mismatch repair and nucleotide excision repair." Proc Natl Acad Sci U S A **95**(24): 14278-14283.

Bhargava, R., Onyango, D. O. and Stark, J. M. (2016). "Regulation of Single-Strand Annealing and its Role in Genome Maintenance." Trends Genet **32**(9): 566-575.

Boivin, S., Kozak, S. and Meijers, R. (2013). "Optimization of protein purification and characterization using Thermofluor screens." Protein Expr Purif **91**(2): 192-206.

Ceccaldi, R., Rondinelli, B. and D'Andrea, A. D. (2016). "Repair Pathway Choices and Consequences at the Double-Strand Break." Trends Cell Biol **26**(1): 52-64.

Ceccaldi, R., Sarangi, P. and D'Andrea, A. D. (2016). "The Fanconi anaemia pathway: new players and new functions." Nat Rev Mol Cell Biol **17**(6): 337-349.

Chalissery, J., Jalal, D., Natour, Z. A. and Hassan, A. H. (2017). "Repair of Oxidative DNA Damage in Saccharomyces cerevisiae." DNA Repair (Amst).

Chan, C. Y. and Schiestl, R. H. (2009). "Rad1, rad10 and rad52 mutations reduce the increase of microhomology length during radiation-induced microhomology-mediated illegitimate recombination in saccharomyces cerevisiae." Radiat Res **172**(2): 141-151.

Chari, A., Haselbach, D., Kirves, J. M., Ohmer, J., Paknia, E., Fischer, N., Ganichkin, O., Moller, V., Frye, J. J., Petzold, G., Jarvis, M., Tietzel, M., Grimm, C., Peters, J. M., Schulman, B. A., Tittmann, K., Markl, J., Fischer, U. and Stark, H. (2015). "ProteoPlex: stability optimization of macromolecular complexes by sparse-matrix screening of chemical space." Nat Methods **12**(9): 859-865.

Chen, H., Lisby, M. and Symington, L. S. (2013). "RPA coordinates DNA end resection and prevents formation of DNA hairpins." Mol Cell **50**(4): 589-600.

Ciccia, A. and Elledge, S. J. (2010). "The DNA Damage Response: Making It Safe to Play with Knives." Molecular Cell **40**(2): 179-204.

Ciccia, A., McDonald, N. and West, S. C. (2008). "Structural and functional relationships of the XPF/MUS81 family of proteins." Annu Rev Biochem **77**: 259-287.

Das, D., Faridounnia, M., Kovacic, L., Kaptein, R., Boelens, R. and Folkers, G. E. (2016). "Single-strand DNA Binding by the Helix-Hairpin-Helix Domain of XPF Contributes to Substrate Specificity of ERCC1-XPF." J Biol Chem.

Das, D., Folkers, G. E., van Dijk, M., Jaspers, N. G., Hoeijmakers, J. H., Kaptein, R. and Boelens, R. (2012). "The structure of the XPF-ssDNA complex underscores the distinct roles of the XPF and ERCC1 helix- hairpin-helix domains in ss/ds DNA recognition." Structure **20**(4): 667-675.

Das, D., Tripsianes, K., Jaspers, N. G., Hoeijmakers, J. H., Kaptein, R., Boelens, R. and Folkers, G. E. (2008). "The HhH domain of the human DNA repair protein XPF forms stable homodimers." Proteins **70**(4): 1551-1563.

Davies, A. A., Friedberg, E. C., Tomkinson, A. E., Wood, R. D. and West, S. C. (1995). "Role of the Rad1 and Rad10 proteins in nucleotide excision repair and recombination." J Biol Chem **270**(42): 24638-24641.

Davis, A. P. and Symington, L. S. (2001). "The yeast recombinational repair protein Rad59 interacts with Rad52 and stimulates single-strand annealing." Genetics **159**(2): 515-525.

de Laat, W. L., Sijbers, A. M., Odijk, H., Jaspers, N. G. and Hoeijmakers, J. H. (1998). "Mapping of interaction domains between human repair proteins ERCC1 and XPF." Nucleic Acids Res **26**(18): 4146-4152.

De Silva, I. U., McHugh, P. J., Clingen, P. H. and Hartley, J. A. (2000). "Defining the roles of nucleotide excision repair and recombination in the repair of DNA interstrand cross-links in mammalian cells." Mol Cell Biol **20**(21): 7980-7990.

Decottignies, A. (2013). "Alternative end-joining mechanisms: a historical perspective." Front Genet **4**: 48.

Diamante, G., Phan, C., Celis, A. S., Krueger, J., Kelson, E. P. and Fischhaber, P. L. (2014). "SAW1 is required for SDSA double-strand break repair in S. cerevisiae." Biochem Biophys Res Commun **445**(3): 602-607.

Do, A. T., Brooks, J. T., Le Neveu, M. K. and LaRocque, J. R. (2014). "Double-strand break repair assays determine pathway choice and structure of gene conversion events in Drosophila melanogaster." G3 (Bethesda) **4**(3): 425-432.

Fadden, A. J., Schalbetter, S., Bowles, M., Harris, R., Lally, J., Carr, A. M. and McDonald, N. Q. (2013). "A winged helix domain in human MUS81 binds DNA and modulates the endonuclease activity of MUS81 complexes." Nucleic Acids Res **41**(21): 9741-9752.

Faridounnia, M., Wienk, H., Kovacic, L., Folkers, G. E., Jaspers, N. G., Kaptein, R., Hoeijmakers, J. H. and Boelens, R. (2015). "The Cerebro-oculo-facio-skeletal Syndrome Point Mutation F231L in the ERCC1 DNA Repair Protein Causes Dissociation of the ERCC1-XPF Complex." J Biol Chem **290**(33): 20541-20555.

Fiorenza, M. T., Bevilacqua, A., Bevilacqua, S. and Mangia, F. (2001). "Growing dictyate oocytes, but not early preimplantation embryos, of the mouse display high levels of DNA homologous recombination by single-strand annealing and lack DNA nonhomologous end joining." Dev Biol **233**(1): 214-224.

Fishman-Lobell, J. and Haber, J. E. (1992). "Removal of nonhomologous DNA ends in double-strand break recombination: the role of the yeast ultraviolet repair gene RAD1." Science **258**(5081): 480-484.

Fishman-Lobell, J., Rudin, N. and Haber, J. E. (1992). "Two alternative pathways of double-strand break repair that are kinetically separable and independently modulated." Mol Cell Biol **12**(3): 1292-1303.

Frankenberg-Schwager, M., Becker, M., Garg, I., Pralle, E., Wolf, H. and Frankenberg, D. (2008). "The role of nonhomologous DNA end joining, conservative homologous recombination, and single-strand annealing in the cell cycle-dependent repair of DNA double-strand breaks induced by H(2)O(2) in mammalian cells." Radiat Res **170**(6): 784-793.

Frankenberg-Schwager, M., Gebauer, A., Koppe, C., Wolf, H., Pralle, E. and Frankenberg, D. (2009). "Single-strand annealing, conservative homologous recombination, nonhomologous DNA end joining, and the cell cycle-dependent repair of DNA double-strand breaks induced by sparsely or densely ionizing radiation." Radiat Res **171**(3): 265-273.

Godin, S. K., Sullivan, M. R. and Bernstein, K. A. (2016). "Novel insights into RAD51 activity and regulation during homologous recombination and DNA replication." Biochem Cell Biol **94**(5): 407-418.

Goldfarb, T. and Alani, E. (2005). "Distinct roles for the Saccharomyces cerevisiae mismatch repair proteins in heteroduplex rejection, mismatch repair and nonhomologous tail removal." Genetics **169**(2): 563-574.

Golovanov, A. P., Hautbergue, G. M., Wilson, S. A. and Lian, L. Y. (2004). "A simple method for improving protein solubility and long-term stability." J Am Chem Soc **126**(29): 8933-8939.

Grabarz, A., Barascu, A., Guirouilh-Barbat, J. and Lopez, B. S. (2012). "Initiation of DNA double strand break repair: signaling and single-stranded resection dictate the choice between

homologous recombination, non-homologous end-joining and alternative end-joining." Am J Cancer Res **2**(3): 249-268.

Gregg, S. Q., Robinson, A. R. and Niedernhofer, L. J. (2011). "Physiological consequences of defects in ERCC1-XPF DNA repair endonuclease." DNA Repair (Amst) **10**(7): 781-791.

Grimme, J. M., Honda, M., Wright, R., Okuno, Y., Rothenberg, E., Mazin, A. V., Ha, T. and Spies, M. (2010). "Human Rad52 binds and wraps single-stranded DNA and mediates annealing via two hRad52-ssDNA complexes." Nucleic Acids Res **38**(9): 2917-2930.

Guzder, S. N., Sommers, C. H., Prakash, L. and Prakash, S. (2006). "Complex formation with damage recognition protein Rad14 is essential for Saccharomyces cerevisiae Rad1-Rad10 nuclease to perform its function in nucleotide excision repair in vivo." Mol Cell Biol **26**(3): 1135-1141.

Gwon, G. H., Jo, A., Baek, K., Jin, K. S., Fu, Y., Lee, J. B., Kim, Y. and Cho, Y. (2014). "Crystal structures of the structure-selective nuclease Mus81-Eme1 bound to flap DNA substrates." Embo j **33**(9): 1061-1072.

Hodskinson, M. R., Silhan, J., Crossan, G. P., Garaycoechea, J. I., Mukherjee, S., Johnson, C. M., Scharer, O. D. and Patel, K. J. (2014). "Mouse SLX4 is a tumor suppressor that stimulates the activity of the nuclease XPF-ERCC1 in DNA crosslink repair." Mol Cell **54**(3): 472-484.

Hoeijmakers, J. H. (2001). "Genome maintenance mechanisms for preventing cancer." Nature **411**(6835): 366-374.

Hoeijmakers, J. H. (2009). "DNA Damage, Aging, and Cancer." N Engl J Med **361**(15): 1475-1485.

Huertas, P., Cortes-Ledesma, F., Sartori, A. A., Aguilera, A. and Jackson, S. P. (2008). "CDK targets Sae2 to control DNA-end resection and homologous recombination." Nature **455**(7213): 689-692.

Huertas, P. and Jackson, S. P. (2009). "Human CtIP mediates cell cycle control of DNA end resection and double strand break repair." J Biol Chem **284**(14): 9558-9565.

Ivanov, E. L. and Haber, J. E. (1995). "RAD1 and RAD10, but not other excision repair genes, are required for double-strand break-induced recombination in Saccharomyces cerevisiae." Mol Cell Biol **15**(4): 2245-2251.

Ivanov, E. L., Sugawara, N., Fishman-Lobell, J. and Haber, J. E. (1996). "Genetic requirements for the single-strand annealing pathway of double-strand break repair in Saccharomyces cerevisiae." Genetics **142**(3): 693-704.

Iyama, T. and Wilson, D. M., 3rd (2013). "DNA repair mechanisms in dividing and non-dividing cells." DNA Repair (Amst) **12**(8): 620-636.

Jackson, S. P. and Bartek, J. (2009). "The DNA-damage response in human biology and disease." Nature **461**(7267): 1071-1078.

Jancarik, J., Pufan, R., Hong, C., Kim, S. H. and Kim, R. (2004). "Optimum solubility (OS) screening: an efficient method to optimize buffer conditions for homogeneity and crystallization of proteins." Acta Crystallogr D Biol Crystallogr **60**(Pt 9): 1670-1673.

Jaspers, N. G., Raams, A., Silengo, M. C., Wijgers, N., Niedernhofer, L. J., Robinson, A. R., Giglia-Mari, G., Hoogstraten, D., Kleijer, W. J., Hoeijmakers, J. H. and Vermeulen, W. (2007). "First reported patient with human ERCC1 deficiency has cerebro-oculo-facio-skeletal syndrome with a mild defect in nucleotide excision repair and severe developmental failure." Am J Hum Genet **80**(3): 457-466.

Jensen, R. B., Carreira, A. and Kowalczykowski, S. C. (2010). "Purified human BRCA2 stimulates RAD51-mediated recombination." Nature **467**(7316): 678-683.

Johnson, R. E., Kovvali, G. K., Prakash, L. and Prakash, S. (1996). "Requirement of the yeast MSH3 and MSH6 genes for MSH2-dependent genomic stability." J Biol Chem **271**(13): 7285-7288.

Jurka, J. (1998). "Repeats in genomic DNA: mining and meaning." Curr Opin Struct Biol **8**(3): 333-337.

Krejci, L., Altmannova, V., Spirek, M. and Zhao, X. (2012). "Homologous recombination and its regulation." Nucleic Acids Res **40**(13): 5795-5818.

Lan, L., Hayashi, T., Rabeya, R. M., Nakajima, S., Kanno, S., Takao, M., Matsunaga, T., Yoshino, M., Ichikawa, M., Riele, H., Tsuchiya, S., Tanaka, K. and Yasui, A. (2004). "Functional and physical interactions between ERCC1 and MSH2 complexes for resistance to cis-diamminedichloroplatinum(II) in mammalian cells." DNA Repair (Amst) **3**(2): 135-143.

Larminat, F., Germanier, M., Papouli, E. and Defais, M. (2002). "Deficiency in BRCA2 leads to increase in non-conservative homologous recombination." Oncogene **21**(33): 5188-5192.

Li, F., Dong, J., Eichmiller, R., Holland, C., Minca, E., Prakash, R., Sung, P., Yong Shim, E., Surtees, J. A. and Eun Lee, S. (2013). "Role of Saw1 in Rad1/Rad10 complex assembly at recombination intermediates in budding yeast." Embo j **32**(3): 461-472.

Li, F., Dong, J., Pan, X., Oum, J. H., Boeke, J. D. and Lee, S. E. (2008). "Microarray-based genetic screen defines SAW1, a gene required for Rad1/Rad10-dependent processing of recombination intermediates." Mol Cell **30**(3): 325-335.

Li, G. M. (2008). "Mechanisms and functions of DNA mismatch repair." Cell Res **18**(1): 85-98.

Liang, F., Han, M., Romanienko, P. J. and Jasin, M. (1998). "Homology-directed repair is a major double-strand break repair pathway in mammalian cells." Proc Natl Acad Sci U S A **95**(9): 5172-5177.

Lieber, M. R. (2010). "The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway." Annu Rev Biochem **79**: 181-211.

Lieber, M. R., Ma, Y., Pannicke, U. and Schwarz, K. (2003). "Mechanism and regulation of human non-homologous DNA end-joining." Nat Rev Mol Cell Biol **4**(9): 712-720.

Lim, D. S. and Hasty, P. (1996). "A mutation in mouse rad51 results in an early embryonic lethal that is suppressed by a mutation in p53." Mol Cell Biol **16**(12): 7133-7143.

Lindahl, T. (1993). "Instability and decay of the primary structure of DNA." Nature **362**(6422): 709-715.

Lok, B. H., Carley, A. C., Tchang, B. and Powell, S. N. (2013). "RAD52 inactivation is synthetically lethal with deficiencies in BRCA1 and PALB2 in addition to BRCA2 through RAD51-mediated homologous recombination." Oncogene **32**(30): 3552-3558.

Lok, B. H. and Powell, S. N. (2012). "Molecular pathways: understanding the role of Rad52 in homologous recombination for therapeutic advancement." Clin Cancer Res **18**(23): 6400-6406.

Lyndaker, A. M. and Alani, E. (2009). "A tale of tails: insights into the coordination of 3' end processing during homologous recombination." Bioessays **31**(3): 315-321.

Lyndaker, A. M., Goldfarb, T. and Alani, E. (2008). "Mutants defective in Rad1-Rad10-Slx4 exhibit a unique pattern of viability during mating-type switching in Saccharomyces cerevisiae." Genetics **179**(4): 1807-1821.

Mansour, W. Y., Schumacher, S., Rosskopf, R., Rhein, T., Schmidt-Petersen, F., Gatzemeier, F., Haag, F., Borgmann, K., Willers, H. and Dahm-Daphi, J. (2008). "Hierarchy of nonhomologous end-joining, single-strand annealing and gene conversion at site-directed DNA double-strand breaks." Nucleic Acids Res **36**(12): 4088-4098.

Mazón, G., Mimitou, E. P. and Symington, L. S. (2010). "SnapShot: Homologous Recombination in DNA Double-Strand Break Repair." Cell **142**(4): 648.e641-648.e642.

Middelberg, A. P. J. (1995). "Process-scale disruption of microorganisms." Biotechnol Adv **13**(3): 491-551.

Morrical, S. W. (2015). "DNA-pairing and annealing processes in homologous recombination and homology-directed repair." Cold Spring Harb Perspect Biol **7**(2): a016444.

Mortensen, U. H., Bendixen, C., Sunjevaric, I. and Rothstein, R. (1996). "DNA strand annealing is promoted by the yeast Rad52 protein." Proc Natl Acad Sci U S A **93**(20): 10729-10734.

Motycka, T. A., Bessho, T., Post, S. M., Sung, P. and Tomkinson, A. E. (2004). "Physical and functional interaction between the XPF/ERCC1 endonuclease and hRad52." J Biol Chem **279**(14): 13634-13639.

Moynahan, M. E., Pierce, A. J. and Jasin, M. (2001). "BRCA2 is required for homology-directed repair of chromosomal breaks." Mol Cell **7**(2): 263-272.

Newman, M., Murray-Rust, J., Lally, J., Rudolf, J., Fadden, A., Knowles, P. P., White, M. F. and McDonald, N. Q. (2005). "Structure of an XPF endonuclease with and without DNA suggests a model for substrate recognition." Embo j **24**(5): 895-905.

Nimonkar, A. V., Sica, R. A. and Kowalczykowski, S. C. (2009). "Rad52 promotes second-end DNA capture in double-stranded break repair to form complement-stabilized joint molecules." Proc Natl Acad Sci U S A **106**(9): 3077-3082.

Nishino, T., Komori, K., Ishino, Y. and Morikawa, K. (2003). "X-ray and biochemical anatomy of an archaeal XPF/Rad1/Mus81 family nuclease: similarity between its endonuclease domain and restriction enzymes." Structure **11**(4): 445-457.

Nishino, T., Komori, K., Ishino, Y. and Morikawa, K. (2005). "Structural and functional analyses of an archaeal XPF/Rad1/Mus81 nuclease: asymmetric DNA binding and cleavage mechanisms." Structure **13**(8): 1183-1192.

Nooren, I. M. and Thornton, J. M. (2003). "Diversity of protein-protein interactions." Embo j **22**(14): 3486-3492.

Nowotny, M. and Gaur, V. (2016). "Structure and mechanism of nucleases regulated by SLX4." Curr Opin Struct Biol **36**: 97-105.

Orel, N., Kyryk, A. and Puchta, H. (2003). "Different pathways of homologous recombination are used for the repair of double-strand breaks within tandemly arranged sequences in the plant genome." Plant J **35**(5): 604-612.

Paques, F. and Haber, J. E. (1997). "Two pathways for removal of nonhomologous DNA ends during double-strand break repair in Saccharomyces cerevisiae." Mol Cell Biol **17**(11): 6765-6771.

Paques, F. and Haber, J. E. (1999). "Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae." Microbiol Mol Biol Rev **63**(2): 349-404.

Park, C. H. and Sancar, A. (1994). "Formation of a ternary complex by human XPA, ERCC1, and ERCC4(XPF) excision repair proteins." Proc Natl Acad Sci U S A **91**(11): 5017-5021.

Park, C. J. and Choi, B. S. (2006). "The protein shuffle. Sequential interactions among components of the human nucleotide excision repair pathway." Febs j **273**(8): 1600-1608.

Pastink, A., Eeken, J. C. and Lohman, P. H. (2001). "Genomic integrity and the repair of double-strand DNA breaks." Mutat Res **480-481**: 37-50.

Patel, K. J., Yu, V. P., Lee, H., Corcoran, A., Thistlethwaite, F. C., Evans, M. J., Colledge, W. H., Friedman, L. S., Ponder, B. A. and Venkitaraman, A. R. (1998). "Involvement of Brca2 in DNA repair." Mol Cell **1**(3): 347-357.

Petukhova, G., Stratton, S. A. and Sung, P. (1999). "Single strand DNA binding and annealing activities in the yeast recombination factor Rad59." J Biol Chem **274**(48): 33839-33842.

Pontier, D. B. and Tijsterman, M. (2009). "A Robust Network of Double-Strand Break Repair Pathways Governs Genome Integrity during C. elegans Development." Curr Biol **19**(16): 1384-1388.

Preston, C. R., Engels, W. and Flores, C. (2002). "Efficient repair of DNA breaks in Drosophila: evidence for single-strand annealing and competition with other repair pathways." Genetics **161**(2): 711-720.

Preston, C. R., Flores, C. C. and Engels, W. R. (2006). "Differential usage of alternative pathways of double-strand break repair in Drosophila." Genetics **172**(2): 1055-1068.

Pryor, E. E., Jr., Wozniak, D. J. and Hollis, T. (2012). "Crystallization of Pseudomonas aeruginosa AmrZ protein: development of a comprehensive method for obtaining and optimization of protein-DNA crystals." Acta Crystallogr Sect F Struct Biol Cryst Commun **68**(Pt 8): 985-993.

Radhakrishnan, S. K., Jette, N. and Lees-Miller, S. P. (2014). "Non-homologous end joining: emerging themes and unanswered questions." DNA Repair (Amst) **17**: 2-8.

Rashev, M., Surtees, J. A. and Guarne, A. (2017). "Large-scale production of recombinant Saw1 in Escherichia coli." Protein Expr Purif. doi: 10.1016/j.pep.2017.02.014.

Rhodes, G. (2010). Crystallography made crystal clear: a guide for users of macromolecular models, Academic press.

Rodriguez, K., Wang, Z., Friedberg, E. C. and Tomkinson, A. E. (1996). "Identification of functional domains within the RAD1.RAD10 repair and recombination endonuclease of Saccharomyces cerevisiae." J Biol Chem **271**(34): 20551-20558.

Saleh-Gohari, N. and Helleday, T. (2004). "Conservative homologous recombination preferentially repairs DNA double-strand breaks in the S phase of the cell cycle in human cells." Nucleic Acids Res **32**(12): 3683-3688.

Sarangi, P., Altmannova, V., Holland, C., Bartosova, Z., Hao, F., Anrather, D., Ammerer, G., Lee, S. E., Krejci, L. and Zhao, X. (2014). "A versatile scaffold contributes to damage survival via sumoylation and nuclease interactions." Cell Rep **9**(1): 143-152.

Sarangi, P., Bartosova, Z., Altmannova, V., Holland, C., Chavdarova, M., Lee, S. E., Krejci, L. and Zhao, X. (2014). "Sumoylation of the Rad1 nuclease promotes DNA repair and regulates its DNA association." Nucleic Acids Res **42**(10): 6393-6404.

Sargent, R. G., Meservy, J. L., Perkins, B. D., Kilburn, A. E., Intody, Z., Adair, G. M., Nairn, R. S. and Wilson, J. H. (2000). "Role of the nucleotide excision repair gene ERCC1 in formation of recombination-dependent rearrangements in mammalian cells." Nucleic Acids Res **28**(19): 3771-3778.

Savchenko, A., Yee, A., Khachatryan, A., Skarina, T., Evdokimova, E., Pavlova, M., Semesi, A., Northey, J., Beasley, S., Lan, N., Das, R., Gerstein, M., Arrowmith, C. H. and Edwards, A. M. (2003). "Strategies for structural proteomics of prokaryotes: Quantifying the advantages of studying orthologous proteins and of using both NMR and X-ray crystallography approaches." Proteins **50**(3): 392-399.

Schiestl, R. H. and Prakash, S. (1988). "RAD1, an excision repair gene of Saccharomyces cerevisiae, is also involved in recombination." Mol Cell Biol **8**(9): 3619-3626.

Schiestl, R. H. and Prakash, S. (1990). "RAD10, an excision repair gene of Saccharomyces cerevisiae, is involved in the RAD1 pathway of mitotic recombination." Mol Cell Biol **10**(6): 2485-2491.

Schumacher, A. J., Mohni, K. N., Kan, Y., Hendrickson, E. A., Stark, J. M. and Weller, S. K. (2012). "The HSV-1 exonuclease, UL12, stimulates recombination by a single strand annealing mechanism." PLoS Pathog **8**(8): e1002862.

Siede, W., Friedberg, A. S. and Friedberg, E. C. (1993). "Evidence that the Rad1 and Rad10 proteins of Saccharomyces cerevisiae participate as a complex in nucleotide excision repair of UV radiation damage." J Bacteriol **175**(19): 6345-6347.

Sijbers, A. M., de Laat, W. L., Ariza, R. R., Biggerstaff, M., Wei, Y. F., Moggs, J. G., Carter, K. C., Shell, B. K., Evans, E., de Jong, M. C., Rademakers, S., de Rooij, J., Jaspers, N. G., Hoeijmakers, J. H. and Wood, R. D. (1996). "Xeroderma pigmentosum group F caused by a defect in a structure-specific DNA repair endonuclease." Cell **86**(5): 811-822.

Sijbers, A. M., van der Spek, P. J., Odijk, H., van den Berg, J., van Duin, M., Westerveld, A., Jaspers, N. G., Bootsma, D. and Hoeijmakers, J. H. (1996). "Mutational analysis of the human nucleotide excision repair gene ERCC1." Nucleic Acids Res **24**(17): 3370-3380.

Singleton, M. R., Wentzell, L. M., Liu, Y., West, S. C. and Wigley, D. B. (2002). "Structure of the single-strand annealing domain of human RAD52 protein." Proc Natl Acad Sci U S A **99**(21): 13492-13497.

Stark, J. M., Pierce, A. J., Oh, J., Pastink, A. and Jasin, M. (2004). "Genetic steps of mammalian homologous repair with distinct mutagenic consequences." Mol Cell Biol **24**(21): 9305-9316.

Steinmetz, E. (2011). "Expresso[reg] Cloning and Expression Systems: Expressioneering[trade] Technology streamlines recombinant protein expression." Nat Meth **8**(6).

Sturzenegger, A., Burdova, K., Kanagaraj, R., Levikova, M., Pinto, C., Cejka, P. and Janscak, P. (2014). "DNA2 cooperates with the WRN and BLM RecQ helicases to mediate long-range DNA end resection in human cells." J Biol Chem **289**(39): 27314-27326.

Sugawara, N., Goldfarb, T., Studamire, B., Alani, E. and Haber, J. E. (2004). "Heteroduplex rejection during single-strand annealing requires Sgs1 helicase and mismatch repair proteins Msh2 and Msh6 but not Pms1." Proc Natl Acad Sci U S A **101**(25): 9315-9320.

Sugawara, N. and Haber, J. E. (1992). "Characterization of double-strand break-induced recombination: homology requirements and single-stranded DNA formation." Mol Cell Biol **12**(2): 563-575.

Sugawara, N., Ira, G. and Haber, J. E. (2000). "DNA length dependence of the single-strand annealing pathway and the role of Saccharomyces cerevisiae RAD59 in double-strand break repair." Mol Cell Biol **20**(14): 5300-5309.

Sugawara, N., Paques, F., Colaiacovo, M. and Haber, J. E. (1997). "Role of Saccharomyces cerevisiae Msh2 and Msh3 repair proteins in double-strand break-induced recombination." Proc Natl Acad Sci U S A **94**(17): 9214-9219.

Sugiyama, T. and Kantake, N. (2009). "Dynamic regulatory interactions of rad51, rad52, and replication protein-a in recombination intermediates." J Mol Biol **390**(1): 45-55.

Sugiyama, T., New, J. H. and Kowalczykowski, S. C. (1998). "DNA annealing by RAD52 protein is stimulated by specific interaction with the complex of replication protein A and single-stranded DNA." Proc Natl Acad Sci U S A **95**(11): 6049-6054.

Sung, P., Prakash, L. and Prakash, S. (1992). "Renaturation of DNA catalysed by yeast DNA repair and recombination protein RAD10." Nature **355**(6362): 743-745.

Surtees, J. A. and Alani, E. (2006). "Mismatch repair factor MSH2-MSH3 binds and alters the conformation of branched DNA structures predicted to form during genetic recombination." J Mol Biol **360**(3): 523-536.

Svendsen, J. M. and Harper, J. W. (2010). "GEN1/Yen1 and the SLX4 complex: Solutions to the problem of Holliday junction resolution." Genes Dev **24**(6): 521-536.

Symington, L. S. (2002). "Role of RAD52 epistasis group genes in homologous recombination and double-strand break repair." Microbiol Mol Biol Rev **66**(4): 630-670, table of contents.

Symington, L. S. (2014). "End resection at double-strand breaks: mechanism and regulation." Cold Spring Harb Perspect Biol **6**(8).

Symington, L. S. and Gautier, J. (2011). "Double-strand break end resection and repair pathway choice." Annu Rev Genet **45**: 247-271.

Thompson, R. F., Walker, M., Siebert, C. A., Muench, S. P. and Ranson, N. A. (2016). "An introduction to sample preparation and imaging by cryo-electron microscopy for structural biology." Methods **100**: 3-15.

Tian, M., Shinkura, R., Shinkura, N. and Alt, F. W. (2004). "Growth retardation, early death, and DNA repair defects in mice deficient for the nucleotide excision repair enzyme XPF." Mol Cell Biol **24**(3): 1200-1205.

Toh, G. W., Sugawara, N., Dong, J., Toth, R., Lee, S. E., Haber, J. E. and Rouse, J. (2010). "Mec1/Tel1-dependent phosphorylation of Slx4 stimulates Rad1-Rad10-dependent cleavage of non-homologous DNA tails." DNA Repair (Amst) **9**(6): 718-726.

Tripsianes, K., Folkers, G., Ab, E., Das, D., Odijk, H., Jaspers, N. G., Hoeijmakers, J. H., Kaptein, R. and Boelens, R. (2005). "The structure of the human ERCC1/XPF interaction domains reveals a complementary role for the two proteins in nucleotide excision repair." Structure **13**(12): 1849-1858.

Tripsianes, K., Folkers, G. E., Zheng, C., Das, D., Grinstead, J. S., Kaptein, R. and Boelens, R. (2007). "Analysis of the XPA and ssDNA-binding surfaces on the central domain of human ERCC1 reveals evidence for subfunctionalization." Nucleic Acids Res **35**(17): 5789-5798.

Trovesi, C., Manfrini, N., Falcettoni, M. and Longhese, M. P. (2013). "Regulation of the DNA damage response by cyclin-dependent kinases." J Mol Biol **425**(23): 4756-4766.

Tsodikov, O. V., Enzlin, J. H., Scharer, O. D. and Ellenberger, T. (2005). "Crystal structure and DNA binding functions of ERCC1, a subunit of the DNA structure-specific endonuclease XPF-ERCC1." Proc Natl Acad Sci U S A **102**(32): 11236-11241.

Tsuzuki, T., Fujii, Y., Sakumi, K., Tominaga, Y., Nakao, K., Sekiguchi, M., Matsushiro, A., Yoshimura, Y. and MoritaT (1996). "Targeted disruption of the Rad51 gene leads to lethality in embryonic mice." Proc Natl Acad Sci U S A **93**(13): 6236-6240.

Tubbs, A. and Nussenzweig, A. (2017). "Endogenous DNA Damage as a Source of Genomic Instability in Cancer." Cell **168**(4): 644-656.

Tutt, A., Bertwistle, D., Valentine, J., Gabriel, A., Swift, S., Ross, G., Griffin, C., Thacker, J. and Ashworth, A. (2001). "Mutation in Brca2 stimulates error-prone homology-directed repair of DNA double-strand breaks occurring between repeated sequences." Embo j **20**(17): 4704-4716.

van Duin, M., de Wit, J., Odijk, H., Westerveld, A., Yasui, A., Koken, M. H. M., Hoeijmakers, J. H. J. and Bootsma, D. (1986). "Molecular characterization of the human excision repair gene ERCC-1: cDNA cloning and amino acid homology with the yeast DNA repair gene RAD10." Cell **44**(6): 913-923.

Van Dyck, E., Stasiak, A. Z., Stasiak, A. and West, S. C. (2001). "Visualization of recombination intermediates produced by RAD52-mediated single-strand annealing." EMBO Rep **2**(10): 905-909.

van Gent, D. C., Hoeijmakers, J. H. and Kanaar, R. (2001). "Chromosomal stability and the DNA double-stranded break connection." Nat Rev Genet **2**(3): 196-206.

Wu, L. and Hickson, I. D. (2003). "The Bloom's syndrome helicase suppresses crossing over during homologous recombination." Nature **426**(6968): 870-874.

Zeman, M. K. and Cimprich, K. A. (2014). "Causes and Consequences of Replication Stress." Nat Cell Biol **16**(1): 2-9.