Functional Roles of Nucleases in DNA Metabolism and Genome Stability

Justin Lee Sparks
Washington University in St. Louis

Follow this and additional works at: https://openscholarship.wustl.edu/etd

Recommended Citation
Sparks, Justin Lee, "Functional Roles of Nucleases in DNA Metabolism and Genome Stability" (2014). All Theses and Dissertations (ETDs). 1349.
https://openscholarship.wustl.edu/etd/1349

This Dissertation is brought to you for free and open access by Washington University Open Scholarship. It has been accepted for inclusion in All Theses and Dissertations (ETDs) by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.
WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences

Molecular Cell Biology

Dissertation Examination Committee:

Peter M. Burgers, Chair
Roberto Galletto
Susana Gonzalo
Nima Mosammamparast
Sheila Stewart
Zhongsheng You

Functional Roles of Nucleases in DNA Metabolism and Genome Stability

by

Justin Sparks

A dissertation presented to the Graduate School of Arts and Sciences of Washington University in partial fulfillment of the requirements for the degree of Doctor of Philosophy

August 2014

St. Louis, Missouri
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acknowledgements</td>
<td>iii</td>
</tr>
<tr>
<td>Abstract</td>
<td>iv</td>
</tr>
<tr>
<td><strong>Chapter I.</strong> Introduction</td>
<td>1</td>
</tr>
<tr>
<td><strong>Chapter II.</strong> Characterization of the Exonuclease 5 family of novel single-strand specific bi-directional exonucleases</td>
<td>25</td>
</tr>
<tr>
<td>A. Yeast Exonuclease 5 is essential for mitochondrial genome maintenance</td>
<td>28</td>
</tr>
<tr>
<td>B. Human Exonuclease 5 is a novel sliding exonuclease required for genome stability</td>
<td>39</td>
</tr>
<tr>
<td>C. <em>S. pombe</em> Exonuclease 5 plays a role in maintenance of both nuclear and mitochondrial genomes</td>
<td>51</td>
</tr>
<tr>
<td><strong>Chapter III.</strong> Reconstitution of RNase H2-initiated Ribonucleotide Excision Repair (RER)</td>
<td>80</td>
</tr>
<tr>
<td>A. RNase H2-initiated Ribonucleotide Excision Repair (RER)</td>
<td>82</td>
</tr>
<tr>
<td>B. RNase H2 roles in genome integrity revealed by unlinking its activities</td>
<td>95</td>
</tr>
<tr>
<td><strong>Chapter IV.</strong> Reconstitution of Top1-initiated ribonucleotide removal with associated mutagenesis</td>
<td>122</td>
</tr>
<tr>
<td><strong>Chapter V.</strong> Future aims and directions</td>
<td>145</td>
</tr>
<tr>
<td><strong>References.</strong></td>
<td>156</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

I would like to begin by saying that there are no words to fully describe how thankful I am to so many people in my life. I could fill the volume of this dissertation with all of the people who made this possible and who touched my life, but I will highlight several people who need to be acknowledged specifically.

First, I would like to thank Peter for being the most amazing mentor I could imagine. In my mind, he epitomizes a true scientist. His enjoyment in not only thinking about but in actually doing experiments will always be something that I aspire to in my own career. Not only is he an amazing scientist with the ability to ask the important questions and the knowledge and experience to find a way to address them, he is almost more interested in being a mentor. My development from student to scientist is due in no small part to his guidance and his personal interest in helping me not just accumulate more data, but to have a better understanding of how to ask the right questions and think like a scientist rather than a student. I will always strive to be a mentor like him.

Thank you to the members of my committee who have been helpful and insightful throughout my thesis research. I want to specifically thank Zhongsheng You for help with the human Exo5 work and the many discussions we shared about the project. Also I would like to thank Roberto Galletto for assistance with several experiments, and also for many helpful suggestions. Additionally, I would like to thank John Majors for many discussions and suggestions about my project and whom I consider another member of my committee.

I would like to thank all of the members of the Burgers lab past and present. I have to thank Sandeep and Vasu for all their help and suggestions, and for keeping lab a fun place to work. I would also like to thank Joe for always being available to discuss our latest experiments and brainstorming crazy possible ideas. I think all of our discussions helped me become a better scientist. I have also appreciated all of the conferences that we attended together and the amusing diversions we always seemed to find--like the ten-mile hike we made around Tunnel Mountain. Sjoerd, Paulina, and Alena: Thank you for always being helpful and full of thought-provoking discussions. I also have to thank Bonnie for being a wonderful person who is so caring and has been amazing to my family and me. Finally, I have to thank Carrie who taught me almost all of the techniques that I used in my research. She has been a great teacher and friend in lab.

I would also like to thank my wife Megan for always supporting me and helping me edit anything that I write. She has been there to share all my good and rough times. She is an amazing wife, and friend. I also have to thank our daughter Esther who has made these last several months of graduate school even more exciting, and has given me a new reason to want to come home after a long day in lab. I love both of them very much.

Last but not least, I need to thank my wonderful parents who have always pushed me to be my best, and nothing more. I have learned so much from your example and cannot express how much you both mean to me. I have always tried to make you proud and let you know that you inspire me to use the talents that you helped nurture through all those years. I also have to thank my siblings for each being unique and special to me in so many ways. I love you all and thank you for always supporting me.
ABSTRACT OF THE DISSERTATION

Functional Roles of Nucleases in DNA Metabolism and Genome Stability

by

Justin Lee Sparks

Doctor of Philosophy in Biology and Biomedical Sciences

Molecular Cell Biology

Washington University in St. Louis, 2014

Professor Peter Burgers, Chair

The work outlined in this dissertation focuses on two distinct areas that are important for genome stability. Both areas focus on DNA repair pathways that require the action of nucleases, specifically Exonuclease 5 and Ribonuclease H2. First, I describe the biochemical and molecular characterization of the novel Exonuclease 5 family of enzymes from *S. cerevisiae*, *S. pombe*, and humans. The Exo5 family consists of bidirectional single-strand DNA specific exonucleases that all contain an iron-sulfur cluster as a structural motif and all have various roles in DNA metabolism. In the *Saccharomycetales* order that includes the budding yeast, *S. cerevisiae*, Exo5 is a mitochondrial protein that is essential for mitochondrial genome maintenance. In an unrelated yeast species, *Schizosaccharomyces pombe*, Exo5 is important for both nuclear and mitochondrial DNA metabolism. The human ortholog is important for nuclear genome stability, and for DNA repair. The work outlined in Chapter II of this Dissertation establishes Exo5 as a protein that is important for DNA metabolism.

The second area of study outlined in Chapters III and IV is related to the phenomenon of ribonucleotide incorporation into the genome by replicative polymerases,
and these chapters focus on the enzymes that remove these noncanonical nucleotides. Ribonucleotides are incorporated into DNA by the replicative DNA polymerases at frequencies of about 2 per kb, which makes them by far the most abundant form of potential DNA damage in the cell. Their removal is essential for restoring a stable intact chromosome. In Chapter III, I present a complete biochemical reconstitution of the ribonucleotide excision repair (RER) pathway with enzymes purified from *Saccharomyces cerevisiae*. I highlight the requirement for RNase H2 in the process of RER and investigate the redundancies at different steps of repair. Also outlined in this dissertation is the dissection of the different functions of RNase H2 in RER and in the removal of RNA-loops in DNA, and implications for genome instability in human diseases that are affected for these activities.

Chapter IV of this dissertation discusses work on an alternative pathway for ribonucleotide removal from the genome by Topoisomerase I. In *S. cerevisiae*, deletion of *rnh201*, the catalytic subunit of RNase H2, results in the persistence of ribonucleotides remain in the genome, which leads to ~100-fold increase in the frequency of 2-5 bp deletions at di-nucleotide repeat sequences. These deletions are dependent on topoisomerase I (Top1) activity. Here we present an *in vitro* reconstitution of the mechanism of Top1-dependent deletions at di-nucleotide repeat sequences and a mechanism for Top1-initiated removal of ribonucleotides outside of the context of these repeat sequences in *S. cerevisiae*. Top1 attack at a ribonucleotide leads to the formation of a 2’, 3’ cyclic phosphate terminated ssDNA nick, followed by subsequent formation of a Top1-cleavage complex (Top1-cc) upstream of the 2’, 3’ cyclic phosphate. If the ribonucleotide is in the context of a di-nucleotide repeat, there can be realignment of the
DNA allowing for religation and release of Top1, leading to a 2-nucleotide deletion. If the ribonucleotide resides outside a repeat sequence, the realignment is not possible and a different pathway must repair the Top1-cc. Tdp1-dependent repair of Top1-cc requires prior proteolytic processing of the Top1-cc before it can be removed leaving a 3’-phosphate that can be removed by Tpp1, Apn1, or Apn2 forming a substrate suitable for repair by DNA polymerase δ, FEN1 and DNA ligase.
CHAPTER I

INTRODUCTION
The vast majority of DNA metabolic pathways, including replication and repair, require specific regulated DNA cutting, trimming, and nucleotide removal. The first observation of the enzymatic degradation of nucleic acids was reported in 1903 (Araki 1903). DNA repair pathways are essential for maintaining the integrity of the genetic material of the cell through both repair of damaging events and also through DNA metabolic processes such as replication. DNA repair pathways require nucleases to remove aberrant structures and lesions from the damaged DNA, and also are critical for creating DNA structures that are substrates for downstream repair activities such as template extension or homologous recombination. Nucleases are critical components involved in a variety of functions in maintaining the integrity and fidelity of the genome.

The widely varying metabolic processes requiring DNA or RNA to be removed in some capacity has necessitated that cells utilize a toolbox of nucleases with different activities. First, nucleases can be divided into two groups based on their backbone sugar, either ribonucleases (RNA) or deoxyribonuclease (DNA). A second division of nucleases is the requirement of a free DNA/RNA end. Exonucleases require a free end of DNA/RNA to be able to load, allowing access to the active site whereas endonucleases bypass this requirement and are able to load internally on the molecule. Exonucleases can be subdivided further by the directionality of the enzyme’s activity. Directionality of the enzyme comes from the inherent polarity of DNA/RNA. DNA/RNA is a polar molecule that has two distinct ends, either a 5’-phosphate or a 3’-hydroxyl. Most nucleases degrade the nucleic acid in a 5’-3’ or 3’-5’ direction, although there are several exonuclease that are bi-directional. Exonuclease 5, discussed in Chapter II is one example of a bidirectional exonuclease. Nucleases can also have a structural specificity. Examples of structure specific nucleases include Flap endonuclease 1 (FEN1) and Mus81-mms4 endonuclease (Kao,
Henricksen et al. 2002; Bastin-Shanower, Fricke et al. 2003). Nucleases are an integral part of maintaining the genome and their necessity can be appreciated from the degree of redundancy that is inherent in DNA repair pathways. Many pathways can utilize several different nucleases to accomplish the same function, making a clear, genetic delineation of the role of a specific nuclease difficult. Much work has been carried out to determine how nucleases function in DNA metabolism and a short overview will be presented here.

**Replication**

Nucleases are intimately involved in DNA metabolism during replication of the genome. Nucleases actively function on the lagging strand of DNA. Lagging strand synthesis is carried out in a discontinuous manner: Polymerase α primase (Pol α) lays down a RNA primer (Okazaki fragment) followed by a complex of polymerase δ (Pol δ) with the processivity clamp PCNA (proliferating cell nuclear antigen), and clamp loader RFC performs the extension of the RNA primer leading to Okazaki fragments around 200 nt in length in eukaryotes (Ogawa and Okazaki 1980). A graphic representation of a replication fork is outlined in Figure 1. When Pol δ encounters the 5’-terminus of the adjacent Okazaki fragment it initiates strand displacement synthesis to form RNA/DNA flaps that are removed by two partially redundant pathways (Figure 2C). The predominant pathway is called the short flap pathway utilizing Flap endonuclease 1 (scFEN1, spRad2, hFen1), a structure specific 5’-3’ exonuclease that cleaves short RNA or DNA flaps (Figure 2C) (Bambara, Murante et al. 1997). FEN1 also has a PCNA

![Figure 1. The eukaryotic replication fork (Nick McElhinny, Watts et al. 2010)]
interacting motif that allows for recruitment and coordination of the activities of enzymes of Okazaki fragment processing (Li, Li et al. 1995; Kao and Bambara 2003). In the short flap pathway Pol δ displaces a short flap less than ten nucleotides that is cleaved by FEN1, and this process is iterative until the RNA primer has been removed and the DNA is finally ligated by DNA ligase 1 (Bambara, Murante et al. 1997; Ayyagari, Gomes et al. 2003). The alternative pathway arises when Pol δ displaces a flap longer than ~22 nucleotides that becomes a substrate for the stable binding of the single strand binding protein RPA, subsequently blocking the ability of FEN1 to remove the flap (Rossi and Bambara 2006); the alternative pathway is therefore called the long flap pathway of Okazaki fragment processing (Figure 2C). The long flap pathway utilizes a nuclease/helicase containing protein called Dna2 to degrade the long flap to 5-6 nucleotides, thereby removing the bound RPA and allowing for FEN1 and the short flap

Figure 2. Stages of lagging strand replication. (A) Lagging strand replication Pol α lays down an RNA primer and is then handed off to PCNA and Pol δ along with FEN1. (B) Pol δ extends the RNA primer until it reaches the 5’-end of adjacent Okazaki fragment. (C) Maturation of Okazaki fragments is carried out by at least two pathways either short flap pathway requiring FEN1 or if flaps grow longer than ~20 nt and are bound by RPA the long flap pathway requiring both Dna2 and FEN1 (Garg and Burgers 2005).
Dna2 is a bi-directional exonuclease with weak 5’-3’ helicase activity (Kang, Lee et al. 2010). In *S. cerevisiae* DNA2 is an essential gene, and it has been shown to have many functions including Okazaki fragment processing (Budd, Tong et al. 2005), 5’-3’ resection during homologous recombination (Zhu, Chung et al. 2008), and checkpoint activation (Kumar and Burgers 2013) among others. It has been proposed that the long flap pathway is necessary because of the action of a helicase Pif1, at least in *S. cerevisiae* where Pif1 augments the strand displacement synthesis activity of Polδ leading to increased frequency of long flaps. Deletion of *PIF1* in *S. cerevisiae* suppresses the lethality of a dna2Δ, supporting a model where deletion of Pif1 removes the requirement of Dna2 to perform long flap Okazaki processing (Budd, Reis et al. 2006). Proper maturation of Okazaki fragments is essential for faithful DNA replication and genome maintenance. Replication fork stalling and collapse is another aspect of replication that requires coordinated activities of several nucleases. Fork stability and collapse will be discussed later in the context of homologous recombination. How cells stabilize and restart stalled replication forks is a very intense area of research, and is essential to understanding how cells deal with the task of replicating their genomes.

**Base Excision Repair (BER)**

Base excision repair is important for repair of damaged bases that occur from endogenous damage resulting in AP-site formation. AP-sites are thought to be one of the most abundant forms of DNA damage with estimates of ~10,000 sites per mammalian cell per day (Lindahl and Andersson 1972; Friedberg, Walker et al. 2006).

Base excision repair pathway, particularly in yeast, shares many features with Okazaki fragment maturation. BER is utilized for the repair of a wide variety of damaged DNA bases that maintain the DNA structure with little distortion including 8-oxoguanine (oxidation of guanine),
uracil (deamination of thymine), and many others. These lesions are thought to occur by both endogenous and exogenous sources. These bases are first recognized by DNA glycosylases that cleave the glycosidic bond between the DNA base and the deoxyribose sugar backbone leaving an abasic site (AP-site). There are many DNA glycosylases that have various damaged base substrates. Once, an AP-site has been formed it is subsequently cleaved by an AP-endonuclease, in yeast Apn1 and Apn2 (Figure 3). These enzymes cleave the phosphate backbone at AP-sites leaving a free 3’-hydroxyl that is a substrate for extension by a DNA polymerase. Also, in yeast several DNA glycosylases can cleave AP-sites including Ntg1, Ntg2, and Ogg1 (Figure 3) (Girard and Boiteux 1997). These DNA glycosylases leave a 3’-phosphate that must be removed by Apn1, Apn2, or Tpp1 before extension by a polymerase (Vance and Wilson 2001). There are
two main pathways that can be utilized to fill in the ssDNA gap (Figure 3). First is the short patch mechanism where Pol β insert a single nucleotide and the nick is ligated by DNA ligase III. Alternatively, the long patch mechanism results from Pol δ carrying out strand displacement synthesis leading to the formation of a 5'‐flap that is cleaved by FEN1 similar to short-flap Okazaki fragment maturation. The latter pathway is the only one operating in yeast.

Another specialized form of base excision repair has recently been characterized in S. cerevisiae: Ribonucleotide Excision Repair (RER). Replicative DNA polymerases have a strong preference for the incorporation of deoxynucleotide (dNMPs) over ribonucleotides (rNMP) into DNA, but in cells the concentration of rNTPs are in 10‐100‐fold excess of dNTPs leading to significant number of rNMPs incorporated into the genome (Nick McElhinny, Kumar et al. 2010; Nick McElhinny, Watts et al. 2010). These incorporated rNMPs are the most abundant non‐canonical nucleotides incorporated into the genome (Nick McElhinny, Watts et al. 2010). In vitro experiments using a template with a single ribonucleotide causes Pol δ and Polε to pause leading to slower replication (Watt, Johansson et al. 2011). RER is responsible for removal of these incorporated rNMPs from the genome and is important for maintaining genome stability. RER is initiated by RNase H2, a heterotrimeric RNase H protein complex that cleaves on the 5'‐

Figure 4. Aicardi Goutières Syndrome is an autoimmune disorder. Mutations in several nucleases have been shown to cause this syndrome including Rnase H2 (Lee-Kirsch, Wolf et al. 2013).
side of a ribonucleotide in a RNA/DNA hybrid. RNase H2 can cleave at sites of a single rNMP or at a long stretch in a RNA/DNA hybrid (Eder and Walder 1991; Rychlik, Chon et al. 2010). In yeast, RNase H2 is not an essential gene, but in mice RNase H2 is essential for growth after gastrulation and its deletion causes rNMPs to remain in the genome (Reijns, Rabe et al. 2012). Bi-allelic hypomorphic mutations in RNase H2 and several other nucleic acid metabolism proteins including a 3’-exonuclease Trex1 and a protein of unknown function, SamHD1, have been shown to be the cause of the human disease Aicardi-Goutières syndrome (Figure 4) (Crow, Leitch et al. 2006). RNase H2 cleaves on the 5’-side of the rNMP leaving a 5’-phosphate and a 3’-hydroxyl that is a substrate for extension. The rNMP is removed by the concerted work of either Polδ or ε performing strand displacement synthesis and forming short flaps that are cleaved by FEN1, in an analogous fashion to Okazaki fragment maturation mentioned earlier (Sparks, Chon et al. 2012).

In *S. cerevisiae* deletion of RNase H2 leads to a small increase in the mutation rate but leads to a significant increase in short 2-5 bp deletions, specifically at sites of di-nucleotide repeats that are dependent on Topoisomerase 1 (Top1) (Nick McElhinny, Kumar et al. 2010; Kim, Huang et al. 2011; Williams, Smith et al. 2013). Both human and vaccinia viral Top1 homologs were shown to possess endoribonuclease activity on rNMPs in a RNA/DNA hybrid (Sekiguchi and Shuman 1997; Kim, Huang et al. 2011). A model has been proposed for the Top1 induced mutagenesis based on these studies in which Top1 first cleaves the incorporated rNMP leaving a 2’, 3’-cyclic phosphate terminated nick within or near a di-nucleotide repeat sequence. Subsequently, a second Top1 cleavage occurs within six nucleotides to the 5’-side of the nick, allowing the diffusion of the short intervening oligonucleotide, including the rNMP, and leading to the formation of a Top1-DNA covalent complex (Top1-cc) (Cho, Kim et al. 2013). Resolution
of this Top1-cc would result from realignment in the non-cleaved DNA strand allowing proximity of the 5’-hydroxyl to the 3’-phosphate permitting religation of the DNA and release of the Top1 protein (Cho, Kim et al. 2013). The mismatch repair or the subsequent round of replication would lead to the 2-5 bp deletion in one of the daughter cells. No direct biochemical studies have been undertaken to test this model in vitro. Chapter IV of this dissertation is the first direct biochemical reconstitution rNMP removal and di-nucleotide repeat deletions caused by Top1.

Resolution of Top1-cc has been an intense focus of study for some time because several classes of therapeutic cancer drugs target Top1 and stabilize the cleavage complex leading to DNA damage. Several pathways have been shown to resolve Top1-cc in cells. One pathway that has been extensively studied is the TDP1 (tyrosyl-DNA phosphodiesterase 1) pathway. Biochemical studies support a model in which Top1-cc must be partially unfolded and/or partially proteolyzed before Tdp1 is able to remove Top1. Cell biological studies support a role of the proteasome in the Top1 proteolysis (Lin, Ban et al. 2008). The partially proteolized Top1-peptide can be removed by Tdp1 (Debethune, Kohlhagen et al. 2002). Tdp1 activity leaves a 3’-phosphate that must then be removed by one of several redundant pathways. These pathways include protein phosphatases Tpp1, Apn1, and Apn2 (Vance and Wilson 2001). Tpp1 (Three prime phosphatase 1) is a 3’-phosphatase that is distantly related to the T4 polynucleotide kinase, but it does not have the kinase domain. Apn1 and Apn2 are apurinic/apyrimidinic endonucleases that have endonuclease, 3’-diester, and 3’-5’ exonuclease activity. Genetic evidence supports a redundant role for these proteins in removal of the 3’-phosphate after Tdp1 removal of the Top1-peptide (Vance and Wilson 2001). There have been several genetic studies that have implicated other pathways for resolving the Top1-cc including roles for the Rad1-Rad10 endonuclease as
well as the homologous recombination machinery (Liu, Pouliot et al. 2002). In the absence of RNase H2 in *S. cerevisiae*, a Top1-dependent pathway is able to remove a small portion of the incorporated rNMPs from the genome (Williams, Smith et al. 2013). Further studies will be needed to resolve the mechanistic details of these other pathways.

**Mismatch Repair**

Mis-incorporations of nucleotide bases during replication that are not removed by the 3’ to 5’ proofreading domain of the replicative polymerases, are another form of damage that must be repaired for sequence fidelity to be maintained. The repair pathway mostly responsible for the repair of these mismatches is the mismatch repair pathway (MMR). Much of our understanding of mismatch repair system comes from studies of the *E. coli* MutHLS system. In this prokaryotic system the strand discrimination is mediated by MutH endonuclease that cleaves hemimethylated *dam* sites on the unmethylated daughter strand. In eukaryotes there are multiple MutL and MutS homologs, but no MutH homologs (Culligan and Hays 2000). Strand discrimination signals in eukaryotic systems still remain to be established. *In vitro* studies of eukaryotic systems demonstrated that a ssDNA nick can signal strand discrimination (Thomas, Roberts et al. 1991) (Holmes, Clark et al. 1990). DNA nicks are abundant in the lagging strand during DNA synthesis, and some evidence supports this model for discrimination on the lagging strand (Pavlov, Newlon et al. 2002; Nick McElhinny, Kissling et al. 2010). The leading strand of DNA replication is thought to be fairly continuous and it was unknown how strand discrimination was achieved. Recently, two studies have provided evidence that the cleavage by RNase H2 of rNMPs mis-incorporated into the leading strand during replication may in part serve as a strand discrimination signal for the leading strand (Figure 4) (Lujan, Williams et al.; Ghodgaonkar, Lazzaro et al. 2013). Deletion of RNase H2 in *S. cerevisiae* leads to a minor
reduction in the efficiency of MMR (Ghodgaonkar, Lazzaro et al. 2013; Lujan, Williams et al. 2013). In *S. cerevisiae* the mismatch is recognized by either a complex of Msh2-Msh6 (MutSα) or Msh2-Msh3 (MutSβ), along with the major MutL complex of Mlh1-Pms1 (Prolla, Christie et al. 1994). Pms1 has been reported to have endonuclease activity that may allow entry of the downstream exonuclease Exo1. Both Msh3 and Msh6 have PCNA interacting motifs that are important for mismatch recognition of the MutS complexes (Flores-Rozas, Clark et al. 2000). DNA single-strand nicks either at Okazaki fragment junctions (lagging strand) or RNase H2 cleavage at incorporated rNMPs (leading strand), and by additional mechanism(s) to be determined, allow loading of PCNA and allow loading of Exonuclease 1 (Exo1), which is a 5’ to 3’ exonuclease activity. Exo1 is responsible for degradation of the mismatch strand, leaving a gap that is filled by polymerase δ and finally ligated to leave the repaired site. Mismatch repair is important for maintaining the fidelity of genome from prokaryotes to eukaryotes.

**Nucleotide Excision Repair (NER)**

Nucleotide excision repair (NER) is the main pathway of repair DNA damage events such as UV radiation induced damage and other bulky DNA lesions. NER was first discovered in 1964 and early studies were important for this pathway and human disease, specifically the

---

**Figure 5. Eukaryotic rNMP repair directed MMR.**

MutSα binds mismatch and MutLα loads. Incision by RNase H2 at incorporated rNMPs is the signal for strand discrimination in eukaryotic system (Williams and Kunkel 2014).
genetic disorder xeroderma pigmentosum (XP) (Boyce and Howard-Flanders 1964; Pettijohn and Hanawalt 1964; Rasmussen and Painter 1964; Setlow and Carrier 1964; DiGiovanna and Kraemer 2012). XP is characterized by patients having high degree of sensitivity to sunlight and a significant increased disposition to acquiring skin cancer (DiGiovanna and Kraemer 2012). Study of XP patient cells facilitated the determination of proteins that are essential for NER and delineation of the pathway. Further studies in several model systems finally led to the reconstitution of the NER pathway from purified components (Wood, Aboussekhra et al. 1993; Aboussekhra, Biggerstaff et al. 1995). The NER pathway is a series of sequential steps that is initiated by detection distortions to the B-form DNA helix. The DNA helical distortions are detected by a protein XPC, Rad4 in S. cerevisiae, and bind to non-damaged bases that thermodynamically destabilize, allowing them to flip out of the helix and recognize a wide variety of lesions (Sugasawa, Okamoto et al. 2001). Next the TFIIH complex, which has roles in both transcription and NER, is recruited with two ATPase/helicases XPB and XPD (Evans, Fellows et al. 1997). XPB is thought to open the DNA and allow for loading of

Figure 6. Proposed model for NER. XPC binds damaged nucleotide. Followed by binding TFIIH, XPD, XPB leading to DNA unwinding recruiting XPF-ERCC1 and XPG leading to incisions and followed by replication and ligation (Scharer 2014).
XPD. Then XPD is responsible for opening up the DNA duplex around the lesion and translocates along the DNA, being stalled at the lesion (Winkler, Araujo et al. 2000). Following the opening of a DNA bubble and XPD stalling at the lesion, XPA is able to bind and recruit structure specific endonuclease XPF-ERCC1 (scRad1-Rad10, spRad16-swi10) while TFIH recruits XPG (scRad2, spRad13); these nucleases are responsible for cleavage of the damaged strand. XPF-ERCC1 is responsible for cleavage on the 5’-side of the bubble and XPG on the 3’-side. The removed DNA patch is then filled by DNA polymerase δ and repair is finished by strand displacement synthesis and subsequent ligation, allowing for repair of the damage site.

**Double-Strand Break Repair**

DNA double-strand breaks (DSBs) are one of the most detrimental forms of DNA damage because of the possibility of loss or translocation of large portions of chromosomes. As a result, repairing double-strand DNA breaks is imperative for cell survival. Eukaryotes have evolved two major pathways for the repair of these lesions—Non-Homologous End Joining (NHEJ) and Homologous Recombination (HR). There are also alternative end joining pathways, one of which is called Microhomology-Mediated End Joining. The choice of repair pathway depends on many factors including the phase of the cell cycle and the structure of the double strand break that must be repaired.

**Non-Homologous End Joining**

Classical non-homologous end joining is the major pathway of repair when there is no sister chromatid present as a template, but there are also alternative end joining mechanisms that are also important for DSB repair. Classical NHEJ involves processing of the ends of the broken DNA to allow for religation of the broken ends of DNA, which has the potential to be a mutagenic process leading to the loss of genetic information or to genomic rearrangements.
DSBs are first bound by the MRX complex in yeast (Mre11-Rad50-Xrs2) or MRN in humans (Mre11-Rad50-Nbs1), and by the heterdimeric Ku complex (Yku70-Yku80) (Walker, Corpina et al. 2001; Lisby, Barlow et al. 2004). The Ku complex DNA binding protects the DNA ends from degradation and allows time for religation and facilitates recruitment of other NHEJ proteins (Aylon, Liefshitz et al. 2004; Clerici, Mantiero et al. 2008). Ku interacts with Dnl4, whereas Xrs2 interacts with Lif1, allowing for recruitment of the DNA ligase IV (Teo and Jackson 2000). Lif1 has specific interactions with Xrs2 and ku80 with Mre11 (Chen, Trujillo et al. 2001; Daley, Palmbos et al. 2005). The MRX complex is able to stimulate the ligase activity of Dnl4-Lif1 in vitro indicating that the MRX complex plays a role in NHEJ (Chen, Trujillo et al. 2001). Once the DNA ligase IV is in place the DNA ends must be aligned and paired to allow for ligation to occur, possibly one strand at a time (Pascal, O'Brien et al. 2004). If the ends are not suitable for religation there can be recruitment of DNA processing enzymes, such as FEN1, that are able to remove ssDNA flaps; however, the kinetics parameters of this process are unknown. Also, the kinetics for switching to the Homologous Recombination pathway are also unknown; the pathway choice for how a DSB is made will be discussed later.

Recently, there have been several alternative end joining pathways described in yeast including single-strand annealing (SSA) and microhomology mediated end joining (MMEJ) that can be used to repair DSBs, although these mechanisms require resection of the 5’-end of the DNA ends. The SSA mechanism requires 5’-resection that exposes short regions of homology that are able to anneal, and end processing by an endonuclease Rad1-Rad10 that allows for the removal of non-homologous DNA ends; the DNA is then religated (Ivanov and Haber 1995). SSA results in the loss of genetic information in the search for homology (Daley, Laan et al. 2005).
MMEJ is a pathway similar to SSA that is characterized by microhomology at the break junction and is associated with deletion of genetic material (Boulton and Jackson 1996). MMEJ does not require the canonical NHEJ protein Ku but does require the activity of the MRX complex (Ma, Kim et al. 2003). The mechanism is similar to SSA in which the break must be resected to allow for the homology search and the intervening DNA between the microhomology is processed out, leading to deletions. These alternative-NHEJ pathways are mutagenic mechanisms for the repair of DSB that may cause genomic instability.

Homologous Recombination

Homologous Recombination is considered a non-mutagenic pathway of repair because it requires the use of a homologous sister chromatid as a template. Much of our understanding of homologous recombination has come from studies in *S. cerevisiae* in which several nucleases are important, including: Exo1, Dna2/Sgs1, Mre11, and Sae2 (Mimitou and Symington 2008; Zhu, Chung et al. 2008). After a DSB, the cell must make the choice how to repair the break, either using NHEJ or HR. The switch that initiates HR is the 5’-end resection that eliminates NHEJ repair and allows for the formation of long 3’-ssDNA tails that can be used for Rad51 dependent homology search. In *S. cerevisiae*, a two-step resection model has been proposed. First, resection is initiated by the MRX complex along with the endonuclease activity of Sae2 removing a short oligonucleotide from the 5’-strand leading to limited processing (Mimitou and Symington 2008; Zhu, Chung et al. 2008). MRX and Sae2 are only required when the DNA ends are modified either chemically or protein-DNA crosslinks. At DSBs with clean ends MRX and Sae2 are not essential. Many studies of DSB repair in *S. cerevisiae* take advantage of rare-cutting endonucleases such as the HO-endonuclease or I-SceI that allow site specific formation of a DSB that can be analyzed directly. MRX and Sae2 initiate slow processing up to around 100
nucleotides away from the break (Mimitou and Symington 2008; Zhu, Chung et al. 2008). This initial processing is then handed off to two different long-range resection pathways. It is not known how this exchange from MRX-Sae2 is accomplished, but the pathways depend on Exo1 and Dna2/STR (Sgs1-Top3-Rmi1), respectively, and they play redundant roles in the 5’ to 3’ resection away from the break (Mimitou and Symington 2008; Zhu, Chung et al. 2008). Exo1 is a 5’-3’ dsDNA exonuclease that is stimulated by MRX and Sae2 in vitro (Nicolette, Lee et al. 2010; Nimonkar, Genschel et al. 2011). Dna2 is a bidirectional exonuclease that also has weak helicase activity and interacts with the STR complex (Zhu, Chung et al. 2008). It was shown in vitro that in the presence of RPA, Dna2 becomes restricted to being a 5’-3’ exonuclease (Niu, Chung et al.; Cejka, Cannavo et al. 2010; Niu, Chung et al. 2010). Sgs1 is a 3’-5’ RecQ family DNA helicase that interacts with two other proteins Top3 and Rmi1 and this STR complex plays an important role in maintaining the genome (Bernstein, Gangloff et al. 2010). The human STR complex (BLM-Top3α-RMI1-RMI2) suppresses mitotic crossovers and in vitro is able to dissolve double holiday junctions using the topoisomerase activity of Top3α (Bernstein, Gangloff et al. 2010). Top3-

Figure 8. Two-step mechanism of 5’-resectioning. The first step is short-range resection mediated by MRX and Sae2. Free DNA ends do not require MRX, but chemically modified ends require the activity of MRX and Sae2. The second step is mediated either by Dna2 or Exo1 pathway (Symington and Gautier 2011).
Rmi1 complex is not essential to the resection activity of the STR-Dna2 pathway but they do stimulate the resection activity (Cejka, Cannavo et al. 2010; Niu, Chung et al. 2010). Long range resection is not essential for HR and in cells defective for Sgs1-DNA and Exo1, MRX is able to form short 3’-ssDNA tails ~100-700 nucleotides in length that still allow for limited recombination (Gravel, Chapman et al. 2008; Mimitou and Symington 2008; Zhu, Chung et al. 2008). One function of long-range resection carried out by the Exo1 and Sgs1-Dna2 pathways is to create long stretches of RPA coated ssDNA that can activate the DNA damage checkpoint (Mimitou and Symington 2008; Zhu, Chung et al. 2008; Chung, Zhu et al. 2010). Another reason for long-range resection is to expose longer stretches of ssDNA to increase the possible sequence for homology search. There is a high level of functional redundancy of nucleases in the cells, which is expected because of the importance of maintaining the fidelity of the genome (Mimitou and Symington 2009). The pathway of homologous recombination intimately requires the work of nucleases for proper resection and ultimately for the repair of double-strand breaks. Following resection, the 3’-ssDNA tail must have the RPA replaced by Rad51, a homolog of the RecA protein in E. coli. Rad51 requires the function of mediator proteins to displace RPA from the ssDNA, and these mediators assist in the formation of Rad51 nucleofilaments by several mechanisms including assisting in loading, stabilizing, and protecting these nucleofilaments during HR. These mediator proteins include Rad52 and the Rad51 paralogues, Rad55, and Rad57. The Rad51 filament is able to form synapsis with the DNA substrate and form a D-loop that is essential for the homology search. Once homology is established the invading 3’-tail can be extended by a polymerase and several mechanisms can be used to resolve the DNA. First, the invading 3’-tail can be religated to the 5’-end of the broken DNA (second end capture) leading to the formation of a double holiday junction that can be resolve in two ways; either by resolution,
using structure specific endonucleases or by dissolution, by the use of the Sgs1-Top3-Rmi1 complex to untangle the DNA strands (Szostak, Orr-Weaver et al. 1983). The second pathway for resolution of the D-loop is by synthesis-dependent strand-anneling (SDSA), where the D-loop is dissolved before the “second end capture” by a helicase, either Srs2 or Mph1 (Nassif, Penney et al. 1994; Allers and Lichten 2001), and the single-strand DNA overhangs anneal and are able to be ligated together. Finally, the D-loop can be converted into a replication fork that can be extended to the end of the chromosome, a pathway known as break-induced replication BIR (Malkova, Ivanov et al. 1996). This choice of pathway for resolution of the Rad51 D-loops is determined by several factors, including the structure of the DSB, and by the many Rad51 regulators that must be coordinated to lead to repair.

**DSB Repair Pathway Choice**

NHEJ is thought of as a more mutagenic form of repair where the two ends of the double-strand break are made compatible with religation and are then ligated back together, possibly leading to loss of genetic material. Homologous recombination, on the other hand, is thought of as a high-fidelity repair pathway that is non-mutagenic, and requires the sister chromatid to be used as a template to refill the break. In a process termed 5’-resection, the HR pathway requires several nucleases to produce long 3’ DNA tails used to invade the sister

![Figure 9. Model for DSB repair choice. 5’-resection is the major determinate controlling the pathway choice (Chiruvella, Liang et al. 2014).](image-url)
chromatid template and as primers for replicating across the break. Homologous recombination is essential during S phase where collapsed replication forks form single ended double strand breaks that have no partner to religate and must be repaired via a HR pathway. The critical step for determining the pathway that is utilized to repair a double strand break is the initiation of the 5’-resection that commits the cell to repair by homologous recombination or MMEJ (Ira, Pelliccioli et al. 2004; Zhang, Shim et al. 2009). The NHEJ DNA end binding complex Ku blocks the DNA ends from being processed and Ku must be removed before 5’-resection and HR can take place. In cells lacking a functional Ku complex HR is up-regulated, but cells lacking HR do not have increased NHEJ activity, implying that cells first attempt to repair the break using a Ku dependent NHEJ and subsequently allow resection (Allen, Kurimasa et al. 2002; Frank-Vaillant and Marcand 2002; Zhang, Shim et al. 2009). How the cell determines the efficiency of resection initiation is cell cycle dependent and is strictly controlled.

The cell has evolved mechanisms for regulating 5’-resection on a cell cycle basis, decreasing 5’-resection capacity in G1 when there is no sister chromatid to be used as a template for HR, and increasing resection efficiency in S and G2 phases of the cell cycle to stimulate HR activity (Aylon, Liefshitz et al. 2004; Jazayeri, Falck et al. 2006; Barlow, Lisby et al. 2008; Zhang, Shim et al. 2009). CDK (Cdc28) is the cyclin-dependent protein kinase that regulates the cell cycle, and also plays a key role in DSB repair pathway choice by stimulating resection and driving the repair into the HR pathway. CDK activity is low in the G1 phase of the cell cycle, and the activity of NHEJ Ku complex leads to reduced 5’-resection and therefore less HR repair (Aylon, Liefshitz et al. 2004; Clerici, Mantiero et al. 2008; Zhang, Shim et al. 2009). In S and G2 phases of the cell cycle 5’-resection is stimulated by the activity of CDK, and several proteins in the HR pathway are targets for phosphorylation. The CDK target that seems to control the switch
required to turn on resection is Sae2 in *S. cerevisiae* (CtIP human). It is phosphorylated at Ser 267, and mutation of S267A leads to a phenotype that is similar to a *sae2* deletion, which is defective in HR (Huertas, Cortes-Ledesma et al. 2008). Sae2 has endonuclease activity *in vitro*, and HR defects of a *sae2*-deletion are suppressed by concurrent deletion of the Ku complex, suggesting that Sae2 may remove Ku from DNA ends and facilitate resection initiation (Mimitou and Symington 2010). Sae2 activity may lead to a substrate that allows for recruitment of the long-range resection machinery and flip the switch from a substrate suitable for NHEJ to one that facilitates homology directed mechanisms.

**Interstrand Cross link (ICL) Repair**

DNA Interstrand cross-links (ICLs) are genotoxic lesions that covalently link the single-strands of the DNA duplex stably together. The ability to destabilize and separate DNA strands is an essential step in the processes of DNA replication and transcription. ICLs form potent obstructions of these fundamental processes leading to potent genotoxicity that must be efficiently repaired for cell viability. Importantly, ICLs can arise from exogenous agents such as mitomycin C but also from endogenous sources in the form of by-products of metabolic processes (Minko, Harbut et al. 2008). Repair of an ICL lesion is complicated by the inclusion of both strands of DNA in the lesion; thus, precluding the removal of the lesion and use of the opposite strand as an error-free template.

It has been previously appreciated that proteins from several classical DNA repair pathways must be coordinated to efficiently repair ICLs; including proteins involved in nucleotide excision repair (NER), translesion synthesis (TLS), homologous recombination (HR), and in many higher eukaryotes the proteins of the Fanconi Anemia (FA) pathway.
Figure 10. Models for replication-coupled repair of ICLs from studies in *Xenopus laevis* and mammalian systems (Zhang and Walter 2014).

(McCabe, Olson et al. 2009; Moldovan and D'Andrea 2009). Fanconi Anemia is a human inherited cancer predisposition disorder, and patients show hypersensitivity to interstrand cross-linking agents (Wang 2007). The FA pathway is utilized during the S-phase of the cell cycle and is believed to be involved in the resolution of DNA lesions including DSBs brought on through replication stress and ICL (Diffley, 2000). The pathway consists of a core complex of proteins FANCA, B, C, D, E, F, G, L, M, and also several associated proteins that create a scaffold for the ubiquitin ligase FANCL (Ciccia, Ling et al. 2007; Collis, Ciccia et al. 2008). The core complex then leads to the monoubiquitination of a heterodimeric complex of FANCD2 and FANCI (ID-complex) (Garcia-Higuera, Taniguchi et al. 2001; Taniguchi, Garcia-Higuera et al. 2002). Recently, the monoubiquitinated ID-complex has been shown to recruit the nuclease FAN1 to
sites of ICL lesions through its monoubiquitin-binding domain (MacKay, Declais et al. 2011). *Schizosaccharomyces pombe* has a Fan1 homolog and it recently has been shown to work in a pathway with the FANCM homolog in *S. pombe* Fml1 (Fontebasso, Etheridge et al. 2013). The role of FAN1 in the repair of ICLs is still not clear, but it has been shown that the unhooking of ICLs requires monoubiquitinated FANCD2 (Knipscheer, Raschle et al. 2009).

The clearest mechanistic studies of ICL repair have been carried out in the *Xenopus laevis* egg extract system in which a site specific ICL is made in a plasmid that can be replicated in the extract (Raschle, Knipscheer et al. 2008). These detailed biochemical studies are now unraveling the mechanistic details of ICL repair. The model that has come from these studies begins with the convergence and stalling of two replication forks at the ICL with the leading strands of both forks stalling ~30 nucleotides and the lagging strands stalling ~50 nucleotides from the ICL (Fu, Yardimci et al. 2011). The stalling leads to the unloading of one of the MCM helicases from the DNA allowing for the approach of the partner leading strand polymerases to within one nucleotide of the ICL (Knipscheer, Raschle et al. 2009). The repair begins by inducing incisions of the opposing single strand on either side of the ICL termed “unhooking,” carried out by structure specific endonucleases: XPF-ERCC1 along with its scaffolding protein SLX4 (Klein Douwel, Boonen et al. 2014). Following the unhooking of the lesion, the uncleaved strand is first extended by one nucleotide by a yet to be identified translesion polymerase, and is subsequently extended by polymerase ζ (Raschle, Knipscheer et al. 2008; Knipscheer, Raschle et al. 2009). The unhooking process also creates a double-strand breaks (DSBs) that are repaired by the HR-mediate pathway. The resolution of the HR-mediated repair requires the SLX4 complex to resolve the DNA-intermediates of the pathways. The SLX4 complex sets up a scaffold for several nuclease including SLX1, MUS81-EME1, and XPF-ERCC1 (Garner and Smogorzewska
There are also other nucleases that have been shown to be involved in repair of ICLs, but their specific role in the repair process remains elusive. First, through genetic studies in both *S. cerevisiae* and *S. pombe*, the deletion of the nuclease *PSO2/pso2* leads to hypersensitivity to interstrand cross-linking agents. In *S. cerevisiae* it has been shown that Δpso2 strains have normal incision (unhooking) of the ICLs, but fail to fully resolve the double-strand breaks (Magana-Schwencke, Henriques et al. 1982; Li, Hejna et al. 2005). The specific mechanism of repair is still not well characterized because of the complexity of the repair pathways, and biochemical techniques for studying the system have only recently been developed.

ICL repair predominately occurs during DNA replication, initiated by one or dual collision of a replication fork(s) at site of ICLs (Raschle, Knipscheer et al. 2008). There is still much discussion about which of the two models is more relevant *in vivo* (MacKay, Declais et al.; Raschle, Knipscheer et al. 2008; Kratz, Schopf et al. 2010; MacKay, Declais et al. 2011). One of the main questions of the single-collision model is that in metazoan cells there is no known mechanism for reloading the MCM helicase during S phase, as seen for other treatments that collapse replication forks and require converging forks for rescue (Petermann, Orta et al. 2010). There is no conclusive *in vitro* evidence that a single replication fork collision can initiate productive repair of and ICL. The critics of the dual-collision is that in metazoan cells the replication origins are spaced roughly 100 kb apart and a single-fork stalled at an ICL will have to wait a long time before a second replication fork converges (Duderstadt, Reyes-Lamothe et al. 2014). These factors seem to point to the possibility that both of these models may take place in cells, and there may be different mechanisms of repair based on the convergence of the forks. Recently, a third model has been proposed, based on observations using DNA combing and fluorescently labeled ICLs where ~60% of ICLs had been bypassed by a single replication fork
without unhooking the ICL (Huang, Liu et al. 2013). This model has been called the “transverse model”, and was shown to require the Fanconi Anemia helicase FANCM, but the mechanism of repair of the ICL is still not understood (Huang, Liu et al. 2013). The repair of ICLs has only recently been explored and there is the possibility of many new discoveries about how all of these processes are integrated and regulated during repair.
CHAPTER II

Characterization of Exonuclease 5 family of novel single-strand specific bi-directional exonucleases
This chapter describes the biochemical and molecular characterization of the novel Exonuclease 5 family of enzymes from *S. cerevisiae*, *S. pombe*, and humans. Section IIA is focused on the biochemical and *in vivo* function of the *S. cerevisiae EXO5* homolog. In *S. cerevisiae* Exo5 is essential for mitochondrial genome maintenance, but there is no apparent nuclear function. The biochemical characterization demonstrated that scExo5 is a 5’ to 3’ exonuclease leaving di-nucleotides as its major product. Peter Burgers performed the majority of the biochemical work, while I performed the remaining experiments. Bonita Yoder and Carrie Stith carried out the strain construction and *in vivo* experiments as well as protein purification.

Section IIB outlines the work that characterized the human Exo5 protein *in vitro* and *in vivo*. biochemical analysis of the human homolog demonstrates that the human protein has a different *in vitro* and *in vivo* activities compared with the *S. cerevisiae* homolog. Human Exo5 is a bi-directional, single-strand, DNA-specific exonuclease that has a propensity to slide along ssDNA prior to cleavage, and has a direct interaction with human single-strand binding protein complex RPA that enforces a 5’-3’ directionality onto Exo5 *in vitro*. Also, in this study using a new oligonucleotide system, we are able to show that the *S. cerevisiae* homolog is also a bi-directional exonuclease. The *in vivo* experiments demonstrate that human Exo5 is localized to the nucleus and cytoplasm and knockdown of Exo5 in human cells leads to a significant increase in sensitivity to a variety of DNA damaging agents, particularly interstrand crosslinking agents. I carried out all of the biochemical assays for this work and most of the protein purification, with early purifications carried out by Carrie Stith. Our collaborators in Tej Pandita’s lab, Rakesh Kumar and Mayank Singh, carried out the *in vivo* experiments. Also, wild type and several
human RPA mutants were provided by Marc Wold’s lab, while Tim Lohman provided several oligonucleotides used in this section.

Section IIC describes the genetic and biochemical characterization of the *S. pombe* Exo5 homolog. *S. pombe* Exo5 works in an identical fashion to the human protein, but it is ~100-fold higher in activity. Our genetic characterization reveals the knockout of Exo5 in *S. pombe* recapitulates the sensitivity to DNA damaging agents as seen by knockdown in human cells. The *exo5Δ* *S. pombe* strains are sensitive to UV-irradiation, methyl methanesulfonate (MMS), and, in particular, ICL forming drugs. Epistasis experiments demonstrate that *exo5Δ rad2Δ* (FEN1) is synergistically more sensitive to MMS and cis-platin, an ICL forming drug, than either of the single mutants; this is consistent with a model where Exo5 and Rad2 are in different pathways that compete for the same lesion. Also, we see a negative genetic interaction where *exo5Δ rad2Δ* cells are slow growing and cells are elongated, indicating checkpoint activation. Exogenous overexpression of Exo5 is lethal to the cell, and cell elongation indicating strong checkpoint activation. The observed lethality is not caused by the nuclease activity of the protein. Exo5 physically interacts with RPA, and PCNA and our hypothesis is that overexpression of Exo5 titrates away these proteins from an essential function in replication leading to checkpoint activation and lethality. I conducted all of the strain and plasmid construction for this study and did most of the biochemical and cell biological assays. Anna Ballard assisted with several of the biochemical and DNA damage sensitivity assays. We received many *S. pombe* strains from a variety of labs and their generosity allowed this work to be performed.
CHAPTER IIA

Yeast Exonuclease 5 is essential for mitochondrial genome maintenance
Yeast Exonuclease 5 Is Essential for Mitochondrial Genome Maintenance

Peter M. Burgers,* Carrie M. Stith, Bonita L. Yoder, and Justin L. Sparks

Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, Missouri 63110

Received 30 September 2009/Returned for modification 9 November 2009/Accepted 5 January 2010

Yeast exonuclease 5 is encoded by the YBR163w (DEMI) gene, and this gene has been renamed EXO5. It is distantly related to the Escherichia coli RecB exonuclease class. Exo5 is localized to the mitochondria, and EXO5 deletions or nuclease-defective EXO5 mutants invariably yield petites, amplifying either the ori3 or ori5 region of the mitochondrial genome. These petites remain unstable and undergo continuous rearrangement. The mitochondrial phenotype of exo5Δ strains suggests an essential role for the enzyme in DNA replication and recombination. No nuclear phenotype associated with EXO5 deletions has been detected. Exo5 is a monomeric 5′ exonuclease that releases dinucleotides as products. It is specific for single-stranded DNA and does not hydrolyze RNA. However, Exo5 has the capacity to slide across 5′ double-stranded DNA or 5′ RNA sequences and resumes cutting two nucleotides downstream of the double-stranded-to-single-stranded junction or RNA-to-DNA junction, respectively.

Endonucleases and exonucleases are intimately involved in all aspects of DNA metabolism in the cell. In mitochondria, several constitutive nucleases have been identified that contribute to the proper maintenance of the mitochondrial genome through replication and recombination pathways. In addition, nucleases can localize to mitochondria in response to DNA stress in order to mediate appropriate DNA repair. Among the constitutive mitochondrial nucleases in Saccharomyces cerevisiae are the Nuc1 nuclease that contributes to DNA recombination efficiency and functions in apoptosis (4, 38) and the Cce1 endonuclease that resolves recombination intermediates (29). The Din7 endonuclease is a mitochondrially located 5′ flap endonuclease related to FEN1 (20). While deletion of the gene for either of these enzymes produced marginal mitochondrial phenotypes, more severe phenotypes were observed when combined deletions of these nuclease genes were studied or when they were combined with deletions of other genes involved in DNA recombination or repair, such as MIH1 or MSH1 (20, 22, 27). Recently, human Dna2 was shown to localize to both the nuclear and mitochondrial compartments and to participate in mitochondrial DNA replication and base excision repair (11, 39). Its function in yeast mitochondrial DNA maintenance has not been studied in detail. Finally, the 5′ flap endonuclease FEN1, which normally functions in primer RNA degradation during Okazaki fragment maturation in the nucleus, also localizes to the mitochondrion in response to DNA damage, participating in long-patch base excision repair (19, 23).

Since mitochondrial function is not essential to yeast survival, dysfunction caused by mutations of the mitochondrial genome can be readily detected as a loss of respiration function, which is scored as the inability to grow on nonfermentable carbon sources. A defect in the mitochondrial DNA polymerase γ MIP1 results in complete loss of the mitochondrial DNA, and the mutant fails to grow on glycerol-containing media lacking glucose (14). Such cells are designated ρ−. Genome maintenance defects can also result in the generation of petite mutants that still contain mitochondrial DNA. Generally, most of the mitochondrial genome has been deleted, and a small origin-containing region has been amplified (ρA). S. cerevisiae contains eight such origin regions that are highly similar in sequence and are distributed over the 86-kb mitochondrial genome (8, 9, 15). Petites that have amplified the ori5 region have been studied more extensively (16, 22).

While the nucleases listed above participate in the proper maintenance of the mitochondrial genome through their replication and/or recombination functions, none appears to be essential for the integrity of the mitochondrial genome. One reasonable explanation for these observations is functional redundancy. Indeed, functional nuclease redundancy is quite common; it has been observed in the process of DNA degradation during mismatch repair in Escherichia coli, during Okazaki fragment maturation in yeast, and during the resection of double-stranded breaks in yeast (7, 25, 33). However, the possibility remains that an additional nuclease(s) is active in the mitochondrion. The present paper describes an essential mitochondrial exonuclease that is distantly related to the nuclease domain of RecB, a subunit of the bacterial RecBCD recombination. This nuclease was discovered over 2 decades ago during a biochemical chromatographic survey of yeast exonucleases and was called exonuclease 5 (3). Initial studies with a partially purified enzyme preparation showed it to be a 5′ exonuclease specific for single-stranded DNA (ssDNA). Here we report the identification of the EXO5 gene and describe comprehensive biochemical and genetic studies that show a critical role for EXO5 in mitochondrial DNA maintenance, presumably through the processing of replication intermediates. Upon deletion of EXO5 or inactivation of its nuclease activity, only ρ− mutants could be recovered. EXO5 has previously been characterized as DEM1 (defects in morphology)

* Corresponding author. Mailing address: Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, MO 63110. Phone: (314) 362-3872. Fax: (314) 362-7183. E-mail: burgers@biochem.wustl.edu.

† Published ahead of print on 19 January 2010.
because the deletion mutant shows defects in growth and in mitochondrial morphology (10, 12). No nuclear defect associated with an Exo5 deletion has been detected.

MATERIALS AND METHODS

Plasmids and oligonucleotides. Plasmid pBL253 contains the YBR163w (Exo5) open reading frame (ORF) together with 249 nucleotides (nt) of 5′ untranslated sequence and 300 nt of 3′ untranslated sequence cloned into centromere vector yCplac22 (ARSI CEN4 URA3). Plasmid pBL256 contains the same Exo5 region cloned into centromere vector yCpaa2 (-ARSI CEN4 TRPI). Plasmids pBL256-270 and -320, with active-site mutations exo5-D270A and exo5-D320A, respectively, were made by site-directed mutagenesis, and the correct sequences were confirmed by sequencing of the entire gene. Plasmid pBL254 (bluescript 2µm ori HIS3 M13 ori GAL1-10 GST-EXO5) contains the Schistosoma japonicum glutathione S-transferase (GST) gene fused to the N terminus of the Exo5 gene in vector pRS424-GALGST5 (5). The GST tag is separated from the Exo5 gene by a recognition sequence for the human rhinoviral 3C protease (LEVLFQ/GP). After cleavage by the protease, the N-terminal sequence of the GST-Exo5 was determined by automated Edman degradation.

Oligonucleotides were purchased from IDT (Coralville, IA) and purified by high-performance liquid chromatography or urea–polyacrylamide gel electrophoresis (PAGE) (V31, GCCATCAACGGCTTCGAGGCAAC, C1, GTTCTGGAAGGAGGGGATCTTGG; C42, GTTCTGGAAGGAGGGGATCTTGG; C44, GTTCTGGAAGGAGGGGATCTTGG; C45, GTTCTGGAAGGAGGGGATCTTGG). Exo5 region cloned into centromere vector yCplac22 (ARSI CEN4 TRPI). Plasmids pBL256-270 and -320, with active-site mutations exo5-D270A and exo5-D320A, respectively, were made by site-directed mutagenesis, and the correct sequences were confirmed by sequencing of the entire gene. Plasmid pBL254 (bluescript 2µm ori HIS3 M13 ori GAL1-10 GST-EXO5) contains the Schistosoma japonicum glutathione S-transferase (GST) gene fused to the N terminus of the Exo5 gene in vector pRS424-GALGST5 (5). The GST tag is separated from the Exo5 gene by a recognition sequence for the human rhinoviral 3C protease (LEVLFQ/GP). After cleavage by the protease, the N-terminal sequence of the GST-Exo5 was determined by automated Edman degradation.

Enrichment of Exo5 from yeast extracts. Yeast single-ORF deletion derivatives of BY4741 were grown in 1 liter of YPD (2% peptone, 1% yeast extract, 2% glucose) to mid-log phase. The cells were harvested, and extracts were prepared by precipitation with glass beads, polynucleotide precipitation, and ammonium sulfate precipitation as described previously (3). Protein (1 mg) in buffer A was gently stirred with 200 µl of SP-Sepharose plus 20 µl of SP-Sepharose in a total volume of 1.5 ml. Under these buffer conditions, Exo5 does not bind to either matrix. Beads were spun down, and the supernatant was used for nuclease assays with 5′-32P-labeled dT12 substrate.

Enrichment of Exo5 from yeast extracts. Yeast single-ORF deletion derivatives of BY4741 were grown in 1 liter of YPD (2% peptone, 1% yeast extract, 2% glucose) to mid-log phase. The cells were harvested, and extracts were prepared by precipitation with glass beads, polynucleotide precipitation, and ammonium sulfate precipitation as described previously (3). Protein (1 mg) in buffer A was gently stirred with 200 µl of SP-Sepharose plus 20 µl of SP-Sepharose in a total volume of 1.5 ml. Under these buffer conditions, Exo5 does not bind to either matrix. Beads were spun down, and the supernatant was used for nuclease assays with 5′-32P-labeled dT12 substrate.

Exonuclease assays. The standard 20 µl assay mixture contained 200 nM HEPS-NaOH (pH 7.6), 20 µg/ml bovine serum albumin, 1 mM DTT, 5 mM Mg-acetate, 50 mM NaCl, 100 fmol of 32P-labeled oligonucleotide substrate, and enzyme. Incubations were at 30°C for the indicated times. Variations from the standard assay are indicated in the legends to the figures. Reactions were stopped with 10 µl of (final concentration) EDTA plus 0.2% SDS and analyzed by thin-layer chromatography (TLC) on polyethyleneimine-cellulose in 0.7 M LiCl or stopped with 10 µl of (final concentration) EDTA plus 40% formamide and analyzed by 20% PAGE–7 M urea electrophoresis. After the TLC plate or gel was dried, it was subjected to PhosphorImager analysis.

PCR and Southern analysis. Total cellular chromosomal DNA preparations were made from 5-ml cultures grown overnight in selective media (17). These were subjected to 20 cycles of PCR in a 50 µl reaction mixture with 100 ng of chromosomal DNA and 50 pmol each of primers ori-in (GGGGGCTCCATATTTTTC and TAGGGGGGAGGGGTTG), ori-out (primers complementary to ori-in: GAAAAATTAATTTGGAGCCACCC and ACCACCCCTG CTGAAAGAGATGTTG), oriCOX2 (GCCGAGATCCCTTCTCGAGGCAAC and GAACCTTCGTCGAACTTTAAAACCACTTATCT, and oriCOX2 (GCCGAGATCCCTTCTCGAGGCAAC and GAACCTTCGTCGAACTTTAAAACCACTTATCT, oriCOX2 (GCCGAGATCCCTTCTCGAGGCAAC and GAACCTTCGTCGAACTTTAAAACCACTTATCT, and oriCOX2 (GCCGAGATCCCTTCTCGAGGCAAC and GAACCTTCGTCGAACTTTAAAACCACTTATCT). Analysis was on a 2.5% metaphor agarose gel. To label the ori sequences, the isolated ors3 or ors5 PCR fragment was reamplified by PCR with ori-in primers but with 10 µM [α-32P]dATP and the other three deoxynucleotide triphosphates at 200 µM. The labeled ors3 and ors5 fragments were mixed and used as a probe of EcoRV-cut total cellular DNA. EcoRV cuts either origin near the end of the fragment (see Fig. 6).

RESULTS

Identification of the S. cerevisiae EXO5 gene. Previously, we had partially purified Exo5 from yeast and established some of the properties of the enzyme (3). Specifically, we noted that the enzyme preferentially degrades ssDNA and produced dinucleotides as the main products of digestion. In order to further studies of this unique enzyme and gain an understanding of its physiological role in the cell, we first identified the gene for Exo5.

The enzyme was partially purified from yeast as described before, through four chromatographic steps (3). This preparation was further fractionated on a MonoQ column. Unfortunately, sodium dodecyl sulfate (SDS)-PAGE analysis still did...
not allow positive identification of a protein band that comigrated with exonuclease activity. Therefore, the fraction with the highest exonuclease activity and a neighboring fraction with much reduced enzyme activity but with a similar protein banding pattern were fractionated by SDS-PAGE and each of the two lanes was cut into 15 slices and subjected to liquid chromatography-tandem mass spectrometry analysis. Six yeast proteins (Ade3, Tos9, Ngl1, Rfx1, Dem1, and Gon3) that were present in greater abundance in the active lane than in the lane without activity were identified. In order to determine which one corresponded to Exo5, we obtained deletion strains for each one of these genes (none of the genes identified was essential for growth). Extracts were prepared, partially enriched for Exo5, and assayed for activity. Only the extract of the dem1Δ mutant strain lacked a dinucleotide-producing activity (Fig. 1A). This analysis suggests that the DEM1 gene is required for Exo5 activity, and the simplest explanation is that it actually encodes Exo5.

An analysis of yeast protein databases revealed some information about YBR163w = EXO5 = DEM1. The deletion mutant is respiration deficient and shows morphological defects of the mitochondrion, hence the original gene designation DEM1 (defects in morphology) (12). Consistent with a mitochondrial function for EXO5/DEM1, the localization of the protein is mitochondrial (18). In support of this observed localization, the protein sequence contains a strong mitochondrial localization signal, using several prediction programs (MitoProt, Predotar), predicting cleavage at Ser26/Leu27 during mitochondrial import.

Overproduction and purification of Exo5 and Exo5 mutants. Exo5/Dem1 was overproduced in yeast from a multicopy plasmid with the EXO5/DEM1 gene placed under the control of the galactose-inducible GAL1-10 promoter. A cleavable GST tag was added to aid in purification (see Materials and Methods). Following glutathione affinity column chromatography, the GST tag was proteolytically cleaved and Exo5/Dem1 was further purified by heparin-agarose chromatography to more than 98% purity (Fig. 1B). Upon SDS-PAGE analysis, the protein migrated as a 64-kDa protein, in agreement with the size calculated from the ORF sequence (67 kDa). A gel filtration analysis showed that Exo5 migrated as a 58-kDa protein, indicating that Exo5 is a compact monomeric enzyme (Fig. 1C).

An analysis of the gel filtration fractions, using a 5′-labeled oligonucleotide as the substrate, showed potent exonuclease activity that comigrated with the protein peak (data not shown). In agreement with previous studies of the partially purified enzyme, dinucleotides were the main products of digestion (Fig. 2C and 3). Since these studies strongly suggest that the DEM1 gene actually encodes exonuclease 5, we will from now on refer to this gene as EXO5.

We subjected EXO5 to a PSI-BLAST analysis, followed by a protein threading analysis (http://toolkit.tuebingen.mpg.de/hhpred) (32). The PSI-BLAST analysis identified EXO5 as the member of a poorly conserved protein family. Interestingly, the threading analysis that was carried out with the consensus sequence obtained from the PSI-BLAST analysis yielded bacterial RecB-type nuclease as top-scoring structural homologs of EXO5 (Fig. 2A). Among the putative homologs in other model eukaryotic organisms are ORFC185.02 in Schizosaccharomyces pombe and C1orf117 in humans, both uncharacterized ORFs. While the conservation with E. coli RecB was restricted to just three small motifs in the C-terminal nuclear domain of RecB (Fig. 2A), the members of the eukaryotic Exo5 family showed conservation of a large number of small motifs over a core domain of ~250 amino acids (alignment not shown). Remarkably, a set of conserved cysteine residues in the C-terminal domain of the Exo5 members was conserved with the AddAB nuclease domain of the Bacillus subtilis AddAB recombinase (36). In AddB, these cysteines coordinate an iron-sulfur cluster, suggesting that the Exo5 class may also possess this structural domain.

The first and second motifs conserved with the E. coli RecB nuclease are in two β strands that contain the two aspartates that chelate the divalent metal ion in the active site of the nuclease (31). The third conserved motif is in an α helix that overlies the active-site aspartates (Fig. 2B). Although the function of this α helix is not known, the proximity of the invariant glutamine and tyrosine to the active site is remarkable. Previous studies of RecB had shown that mutation of D1080, one of

![Diagram](http://example.com/diagram.png)
the aspartates that ligate the divalent metal ion, abolishes the exonuclease activity of RecBCD (37). Therefore, we mutated either putative active-site residue Asp270 or Asp320 to alanine, overproduced and purified the mutant proteins analogously to the wild type (Fig. 1B), and determined the resulting exonuclease activity. No exonuclease activity was detectable in Exo5-D320A, while that of Exo5-D270A was less than 10⁻⁴ times that of the wild type (Fig. 2C). This analysis not only confirms that Exo5 actually encodes the exonuclease but also lends strength to the threading analysis indicating that Exo5 belongs to the RecB family of nucleases.

Exo5 is a single-strand-specific 5’ exonuclease. Previous enzymatic studies with a partially purified preparation of Exo5 showed that the enzyme is specific for ssDNA and releases primarily dinucleotides as products. (3). We carried out a more thorough investigation of the enzymatic properties of the pure enzyme. First, we tested whether Exo5 has endonuclease activity. A 34-mer linear oligonucleotide was circularized, and activity on the circular substrate was compared with that of the linear substrate. No activity was detected on the circular oligonucleotide under the exact same conditions generated a ladder of intermediates of odd lengths, consistent with a model in which the enzyme sequentially releases dinucleotides from the 5’ terminus of the 21-mer oligonucleotide (Fig. 4B, lanes 4 to 7). The starting 21-mer oligonucleotide was contaminated with ~10% 22-mer due to the addition of two dAMP residues during labeling (Fig. 4B, lane 3). This contamination likely accounted for the presence of small amounts of even-sized oligonucleotides in the digest. These data strongly indicate that Exo5 is a 5’ exonuclease and releases dinucleotides as main products of catalysis.

Exo5 showed no detectable nuclease activity when the 5’-²⁵³P oligonucleotide was completely double stranded (Fig. 4A, lane 12). Activity at the 5’-²⁵³P-labeled oligonucleotide was not affected when just the 3’ end was made double stranded by hybridization of a 14-mer, leaving 6 nt single stranded at the 5’ end (Fig. 4A, lanes 7 to 10). However, no dinucleotide was produced when only the 5’ end was made double stranded (Fig. 4A, lane 11). Interestingly, a 16-mer product was made at an ~50-fold-reduced rate compared to the single-stranded control. This 16-mer could not have resulted from endonucleolytic activity since Exo5 has none (Fig. 3). However, generation of the 16-mer product could have resulted from 3’ loading of Exo5 with low efficiency, followed by cutting 4 nt in from the terminus or, alternatively, Exo5 could have loaded at the double-stranded end at a reduced rate and, after sliding across the double-stranded DNA (dsDNA) region, cut 2 nt downstream from the dsDNA-ssDNA junction. Based upon studies with the
low efficiency and slide across dsDNA, followed by initiation of consistent with the ability of Exo5 to load at dsDNA ends with \( \text{-labeled substrate} \), these data are most produced at an \( \text{H11032} \) with the data from the \( \text{5} \) with the single-stranded oligonucleotide, but they were produced, similar to the results observed \( \text{H11032} \) to 11). When just the \( \text{5} \) by hybridizing a 12-mer oligonucleotide across the EcoRV site, structures contained \( \text{5} \) nM circular or linear bluescript SK2 ssDNA (linear-converted efficiency at the dsDNA-ssDNA junction (Fig. 4B, lanes 8 and 9 with lanes 2 and 3). However, when the \( \text{5} \) end was blocked with a biotin-streptavidin moiety, only residual nuclease activity was observed, indicating that Exo5 has to load at the \( \text{5} \) RNA terminus and sliding across RNA do not constitute a rate-limiting step in the reaction. This type of analysis was repeated with a chimeric oligonucleotide containing 10 ribonucleotides, followed by 11 deoxyribonucleotides, and in addition a \( \text{5} \) biotin label that allows blocking of the \( \text{5} \) end by streptavidin binding (Fig. 5B). With the unblocked \( \text{RNA} \) oligonucleotide, the major cut site was at position 12, \( \text{2} \) nt downstream of the RNA-DNA junction, and the rate of cutting was reduced to about \( \text{50%} \) of that of the comparable DNA oligonucleotide (compare lanes 8 and 9 with lanes 2 and 3). However, when the \( \text{5} \) end was blocked with a biotin-streptavidin moiety, only residual nuclease activity was observed, indicating that Exo5 has to load at the \( \text{5} \) RNA end and slide across the RNA prior to cutting downstream DNA (lanes 14 and 15). As demonstrated above, converting the oligonucleotides to a double-stranded state arrogated all nuclease activity (lanes 4 and 5, lanes 10 and 11, and lanes 16 and 17). This specificity suggests that if Exo5 were involved in the processing of RNA-primed DNA replication intermediates, it would require prior strand displacement synthesis by the DNA polymerase or the participation of a \( \text{3} ' - \text{5} ' \) helicase, e.g., Hm1, in order to generate the appropriate substrate for digestion (30).

Mitochondrial defects of \( \text{EXO5} \) mutants. The large-scale gene deletion project already indicated that \( \text{EXO5} \) (\( \text{DEMI} \)) deletions are respiration deficient (12). To confirm this mitochondrial phenotype, we made a \( \text{his}3\Delta / \text{his}3\Delta \) diploid strain heterozygous for \( \text{EXO5/EXO5}\Delta : \text{HIS3} \). After sporulation, tet-

\[
\text{FIG. 3. Exo5 lacks endonuclease activity. (A) Standard nuclease assay mixtures contained either 10 nM linear or circular 34-mer oligonucleotide} \quad 32\text{P-c81} \quad \text{(see Materials and Methods) and 1 nM, 10 nM, or 100 nM Exo5. Analysis was on a 7 M urea–20\% polyacrylamide gel. Note that the circular (c) oligonucleotide migrates slower than the linear form and is contaminated with} \quad 5\% \quad \text{linear form} \quad \text{asterisk}. \quad \text{This contaminating linear form is converted into the dinucleotide by Exo5. PDE, snake venom phosphodiesterase ladder. (B) Standard assay mixtures contained 5 nM circular or linear bluescript SK2 ssDNA (linearized by hybridizing a 12-mer oligonucleotide across the EcoRV site, followed by cutting with EcoRV and heating to restore complete single strandedness) and 200 nM Exo5. Aliquots taken after the indicated times were analyzed on a 1.2\% agarose gel. Note that under these conditions the linear form migrates slightly slower than the circular form.}
\]

3'-labeled substrate discussed below, we conclude that the latter model is most likely.

Next, we assayed enzymatic activity on 3'-end-labeled substrates. No nuclease activity was detected on the fully double-stranded substrate (Fig. 4B, lane 13). When only the 3' end was made double stranded, the primary product was 17 nt in length and was followed by a 15-mer at a reduced rate. This again indicates that Exo5 digests from the 5' end but cuts with limited efficiency at the dsDNA-ssDNA junction (Fig. 4B, lanes 8 to 11). When just the 5' end was made double stranded, di- and trimucleotides were produced, similar to the results observed with the single-stranded oligonucleotide, but they were produced at an ~50-fold reduced rate (Fig. 4B, lane 12). Together with the data from the 3'-labeled substrate, these data are most consistent with the ability of Exo5 to load at dsDNA ends with low efficiency and slide across dsDNA, followed by initiation of degradation at a position 2 nt down from the dsDNA-ssDNA junction.

To determine Exo5 activity at or close to junctions, we carried out a comprehensive analysis with different partially double-stranded substrates (summarized in Table 1). A partially dsDNA substrate with a 4-nt 5' overhang is cut as efficiently as ssDNA (compare entry 5 with entry 1); however, a 2-nt 5' overhang shows only 2.8% activity and a single-nucleotide overhang is inactive (entries 4 and 3). Exo5 also cuts pseudoforks with a similar substrate preference, showing full activity as long as 4 nt of the 5' strand are single stranded (entry 7). As expected from the substrate specificity displayed by the enzyme, model Holiday junctions were not cut by Exo5 (data not shown).

Exo5 slides across RNA regions to engage downstream DNA. Incubation of Exo5 with labeled RNA yielded little or no digestion products, indicating that the enzyme shows sugar specificity (data not shown). However, how does the enzyme react with chimeric RNA-DNA molecules such as those that might arise as a result of RNA-primed DNA synthesis? Exo5 was incubated with oligonucleotides with increasing-length sections of 5' RNA (Fig. 5A). Remarkably, while incubation with pTr12 released the dinucleotide pTpT, incubation with pUT11, pU3T9, and pU3T9, yielded tri-, tetra-, and penta-nucleotides, respectively, indicating that the enzyme cuts selectively 2 nt downstream of the RNA-DNA junction. Since we know from the analyses described above that Exo5 shows neither endonuclease activity nor 3' exonuclease activity, we conclude that Exo5 binds to 5' RNA termini and, after sliding across the RNA substrate, cuts the DNA 2 nt from the RNA-DNA junction. The rates of hydrolysis do not decrease significantly with the addition of increasing 5' RNA sections, indicating that loading of Exo5 at the 5' terminus and sliding across RNA do not constitute a rate-limiting step in the reaction. This type of analysis was repeated with a chimeric oligonucleotide containing 10 ribonucleotides, followed by 11 deoxyribonucleotides, and in addition a 5' biotin label that allows blocking of the 5' end by streptavidin binding (Fig. 5B). With the unblocked \( \text{RNA} \) oligonucleotide, the major cut site was at position 12, \( \text{2} \) nt downstream of the RNA-DNA junction, and the rate of cutting was reduced to about \( \text{50%} \) of that of the comparable DNA oligonucleotide (compare lanes 8 and 9 with lanes 2 and 3). However, when the 5' end was blocked with a biotin-streptavidin moiety, only residual nuclease activity was observed, indicating that Exo5 has to load at the 5' RNA end and slide across the RNA prior to cutting downstream DNA (lanes 14 and 15). As demonstrated above, converting the oligonucleotides to a double-stranded state arrogated all nuclease activity (lanes 4 and 5, lanes 10 and 11, and lanes 16 and 17). This specificity suggests that if Exo5 were involved in the processing of RNA-primed DNA replication intermediates, it would require prior strand displacement synthesis by the DNA polymerase or the participation of a 3'-5' helicase, e.g., Hm1, in order to generate the appropriate substrate for digestion (30).
rads were dissected onto YPD plates. A 2:2 segregation was observed for histidine prototrophy, and all His<sup>+/H11001</sup> colonies failed to grow on YPG plates, i.e., containing 5% glycerol as a non-fermentable carbon source, while all His<sup>+/H11002</sup> colonies did grow on YPG (data not shown). The tetrad analysis was repeated, but this time spores were directly germinated onto YPG plates. Each of the 14 tetrads dissected yielded a maximum of two colonies that were all His<sup>+/H11002</sup>. Therefore, even under selective conditions, cells lacking EXO5 are respiration deficient.

In order to determine whether the nuclease activity of Exo5 was required to maintain respiration proficiency, we first made strain PY209, which is exo5<sup>/H9004</sup>::HIS3 but is wild type ([rho<sup>-/H11001</sup>]) because of the presence of a complementing EXO5 plasmid with a URA3 selectable marker. We then transformed this strain with a TRP1 plasmid with wild-type EXO5, exo5<sup>-/H11001</sup>-D270A, exo5<sup>-/H11001</sup>-D320A, or the empty vector and asked whether the URA3 plasmid could be evicted by growth on 5-fluoroorotic acid (FOA) media, either on plates containing glucose or on plates containing glycerol. While all four strains produced colonies on FOA-glucose plates, only the strain with wild-type EXO5 produced colonies on FOA-glycerol plates (Fig. 2D). This analysis shows that the nuclease activity of Exo5 is essential for respiration proficiency and therefore indicates an essential role for Exo5 exonuclease activity in the maintenance of the mitochondrial genome.

### TABLE 1. DNA substrate specificity for Exo5<sup>a</sup>

<table>
<thead>
<tr>
<th>Expt</th>
<th>DNA</th>
<th>Sequence</th>
<th>% Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>v31</td>
<td>5'-GCCCAT----</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>c41</td>
<td>3'-CGGGTA----</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>v31</td>
<td>5'-GCCCAT----</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>c42</td>
<td>3'-GGGTA----</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>v31</td>
<td>5'-GCCCAT----</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>c43</td>
<td>3'-GGTA----</td>
<td>2.6</td>
</tr>
<tr>
<td></td>
<td>v31</td>
<td>5'-GCCCAT----</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>c44</td>
<td>3'-TA----</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>v31</td>
<td>5'-GCCCAT----</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>c45</td>
<td>3'-TT&lt;sub&gt;10&lt;/sub&gt;CGGGTA----</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>v31</td>
<td>5'-GCCCAT----</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>c48</td>
<td>3'-TT&lt;sub&gt;10&lt;/sub&gt;TA----</td>
<td>113</td>
</tr>
</tbody>
</table>

<sup>a</sup> Standard assay mixtures contained 100 fmol single-stranded 5'-labeled oligonucleotide v31 (experiment 1) or 100 fmol 5'-labeled oligonucleotide v31 hybridized to 200 fmol of the indicated oligonucleotide and either 10, 100, or 1,000 fmol of Exo5 for 1, 3, or 10 min (see Materials and Methods). One hundred percent activity corresponds to 0.14 dinucleotide released per second per Exo5 molecule.

FIG. 4. Substrate specificity of Exo5. (A) 5'-labeled substrates. 5'-<sup>32</sup>P-labeled v31 was either single stranded (a) or hybridized to a twofold excess of oligonucleotide c50 (b), oligonucleotide c49 (c), or oligonucleotide c41 (d). (B) 3'-labeled substrates. 3'-<sup>32</sup>P-v31-<sup>32</sup>P-dA was either single stranded (a) or hybridized to a twofold excess of oligonucleotide cT50 (b), oligonucleotide c49 (c), or oligonucleotide cT41 (d). An asterisk indicates the position of the label. All of the assay mixtures contained 10 nM labeled substrate, 200 mM NaCl, and 10 nM Exo5 for the indicated times at 30°C. Assay mixtures were analyzed on 7 M urea-20% polyacrylamide gels. PDE, partial digestion with snake venom phosphodiesterase, a 3' exonuclease; RecJ, partial digestion with E. coli RecJ, a 5' exonuclease. A dotted guide line is added in panel B. The values on the left of each panel are oligomer sizes.
Analysis of petite phenotypes in exo5Δ yeast. Yeast strains can become petite because of rearrangement or partial deletions of the mitochondrial genome (\([\text{rho}^-]\)) or because of the deletion of all mitochondrial DNA (\([\text{rho}^0]\)). The latter phenotype is, for example, observed in \(\text{mip1}\) mutant strains defective for the mitochondrial DNA polymerase. In petites that arise from gross deletions in the genome (\([\text{rho}^-]\)), the DNA region retained is rapidly amplified, resulting in a total mitochondrial genome. Control strains were wild type (\([\text{rho}^+]\)). The colonies arising from gross deletions in the genome (\([\text{rho}^-]\)) or to the \(\text{COX2}\) gene (coordinates 73.8 to 75.2 kb). We also used a primer set (ori-in) that amplifies the various mitochondrial ORI sequences because of their identical G/C-rich border sequences (30). In the wild-type strain, PCR products were observed with all three primer sets while PCR of total DNA isolated from the \(\text{mip1}\) mutant strain showed no detectable PCR product with any of the primer sets. Ten out of 11 petite isolates yielded an abundant PCR product with the ori-in primer set, compared to the wild type. However, none gave a PCR product with the \(\text{AI1}\) primer set or the \(\text{COX2}\) primer set (data not shown). We conclude that loss of Exo5 nuclease leads to deletion of major genomic regions and amplification of the ORI region(s). The petite isolates were grown further for a cumulative 90 generations, and the total DNA was analyzed again. Two out of the 11 isolates now failed to produce a PCR product with the ori-in primers, suggesting a complete loss of mitochondrial DNA in these two isolates (Fig. 6B). Sequence analysis of the ~250-nt PCR products of the nine remaining clones showed that each isolate contained one single ORI sequence, either \(\text{ori3}\) (coordinates 54.6 to 54.8 kb; three isolates) or \(\text{ori5}\) (coordinates 82.3 to 82.6 kb; six isolates).

Amplification of a particular ORI sequence in petites generally produces head-to-tail repeats of the ORI sequence with variable neighboring DNA. These neighboring sequences should be identifiable by PCR analysis with a primer set complementary to the ori-in primers (ori-out, Fig. 6A). Indeed, PCR with the ori-out primers showed prominent products for each isolate, confirming that the generation of petites had occurred through head-to-tail amplification of an ORI sequence-containing region. Remarkably, the ori-out PCR consistently identified products with a very distributive size distribution (Fig. 6B). This diffuse distribution could result either from a variability in the repeat lengths generated during establishment of the petite genome or from a general genomic instability in the exo5Δ petite cells during continued growth (Fig. 6B). A Southern blot analysis was carried out with total mitochondrial DNA isolated from these clones and digested with restriction endonuclease EcoRV by using \(\text{ori3}\) and \(\text{ori5}\) DNAs as hybridization probes. This analysis confirmed the pattern of strong mitochondrial repeat length variability (Fig. 6C). Some isolates also showed a
reduction in hybridization strength, particularly isolates 2 and 11, which is presumably due to partial loss of the mitochondrial genome in subpopulations of cells. Furthermore, as expected, isolates 4 and 7, which were \( \rho^0 \), showed no hybridization signal.

We reasoned that in the absence of Exo5, each mitochondrial replication cycle might be associated with a high probability of genome rearrangements and duplications, leading to a dynamic variability in the sizes of the ampicons by PCR and by Southern analysis. We sequenced the ori-out PCR products from the nine clones shown in Fig. 6B (right panel). Overlapping sequences were identified in all of the clones, prohibiting accurate sequence assignment. However, for several isolates, a specific sequence predominated and could be mapped to regions directly to the left and right of either ori3 or ori5. Importantly, this analysis showed that the sequences flanking ori3 or ori5 contained additional internal repeat sequences between 50 and 200 nt in length, suggesting that recombination frequently occurred in the \( \rho^0 \)/H9004 clones (Fig. 6D). We confirmed the high-instability phenotype of \( \rho^0 \)/H9004 mutants by carrying out an experiment in which two independently isolated \( \rho^0 \)/H9004 clones, one with ori3 amplified and one with ori5 amplified, were serially propagated and expanded through successive growth cycles, and individual clones were analyzed by PCR. The two clones produced after loss of \( EXO5 \) persistently altered their ori-out PCR patterns upon serial propagation (Fig. 6E).

We next tested whether the reintroduction of \( EXO5 \) into the petite clones would cause stabilization of the petite genomes.
FIG. 7. exo5Δ mutant strains are not damage sensitive. All strains were miplΔ ([rho0]) and, in addition, had the indicated genotype. Serial 10-fold dilutions of late-log-phase cells, from 10⁵ to 10 cells per spot, were spotted onto YPD plates or YPD plates containing the indicated concentrations of hydroxyurea (HU, mM) or camptothecin (CPT, μg/ml). Some YPD plates were irradiated with the indicated fluency of UVc. Plates were grown for 3 days at 30°C and photographed. WT, wild type.

An EXO5-complementing plasmid was transformed back into randomly selected petite isolates, and the transformants were analyzed by PCR analysis with the ori-in and ori-out primers (Fig. 6F). Remarkably, the ori-out PCR product distribution had narrowed drastically, indicating that a stable repeat pattern had been established in the Exo5-containing transformants. Sequence analysis of the repeat unit identified unique wild-type sequences that border ori3 or ori5 on the left and right. The junction between the upstream and downstream ORI sequences varied from clone to clone. However, invariably, this junction was formed by a sequence 7 to 13 nt in length that was present in both the upstream and downstream mitochondrial genomic sequences. This observation strongly suggests that junction formation had occurred through recombination between these small identical sequence motifs.

Petite mutants can exhibit zygotic suppressiveness, a phenomenon in which, upon mating with a wild-type [rho+] strain, the wild-type mitochondrial genome is excluded from diploids; i.e., the diploids are also [rho-] (13). Hypersuppressiveness is a consequence of the large number of ORI sequence repeats in the [rho+] strain. Both RNA-primed replication and recombination mechanisms contribute to hypersuppressivity (22, 24). The exo5Δ clones lacking mitochondrial DNA (clones 4 and 7 in Fig. 6B) showed no hypersuppressiveness upon mating with a [rho+] strain; i.e., all diploids were wild type [rho+]. In contrast, the other clones showed high degrees of hypersuppressiveness, from 80% to 100%; i.e., up to 100% of the diploids were [rho-], in agreement with similar findings by others (22, 24).

EXO5 deletion shows no nuclear phenotypes. We explored the possibility that Exo5 may also reside in the yeast nucleus at low levels, even though it was only detected in the mitochondria (18). However, our genetic analysis identifying a potential nuclear function for Exo5 was uniformly negative. To eliminate phenotypes due to mitochondrial dysfunction, we carried out all of our studies in a [rho+] miplΔ mutant background. We compared the growth and DNA damage survival of exo5Δ miplΔ double mutants with miplΔ mutants as a control. In addition, we combined these mutants with deletions of the additional nuclease genes RAD27, EXO1, MUS81, and MRE11. The damage sensitivities of the triple mutants were compared with those of the relevant double mutants (Fig. 7). FEN1, encoded by RAD27, is primarily involved in Okazaki fragment maturation and base excision repair, Exo1 is involved in mismatch repair and double-strand break repair, Mus81 is involved in the processing of stalled replication forks, and Mre11 is involved in double-strand break repair (reviewed in references 2, 6, and 34). Serial dilutions were plated on YPD plates containing hydroxyurea or camptothecin, or plates were UV irradiated. No significant differences in damage sensitivity were observed between a given miplΔ nuclease double mutant and the exo5Δ miplΔ nuclease triple mutant. Therefore, we conclude that either EXO5 is not involved in nuclear DNA maintenance or a very efficient redundancy with other nucleases exists in the DNA damage repair pathways tested.

DISCUSSION

Exo5 has the remarkable ability to slide over RNA and less efficiently over dsDNA to initiate cutting at a position 2 nt downstream of the RNA-DNA or dsDNA-ssDNA junction. The ability to differentiate between RNA and DNA but cut ssDNA flaps efficiently through sliding may be important for the maturation of intermediates during mitochondrial DNA replication. Since Exo5 does not degrade dsDNA or dsRNA, the generation of such single-stranded flaps would have to be accomplished either through strand displacement synthesis by the mitochondrial DNA polymerase γ or through the participation of a 3'-5' helicase. The first is an important Okazaki maturation mechanism in the nucleus, where substrates for cutting by FEN1 and by Dna2 are generated through strand displacement synthesis by polymerase δ (reviewed in reference 2). Alternatively, the single-stranded flap could be generated through helicase action. The Hml1 helicase has the correct directionality (3'-5') to generate substrates for Exo5, and it is also essential for maintenance of the wild-type mitochondrial genome but not for that of [rho-] petites (21, 26). Unfortunately, very little is known about the mechanism of replication of the yeast mitochondrial genome. Replication may occur through several mechanisms, including initiation by RNA polymerase and by double-strand break-induced recombinational mechanisms that can mature into rolling-circle replication (22, 28, 35). Sufficient replicative ability must remain in an exo5Δ mutant strain to allow replication of the ORI repeat sequences, since these cells are [rho+] and not [rho0]. However, the observations that 2 out of the 11 petite isolates studied lost their mitochondrial DNA and 2 more were severely depleted for mitochondrial DNA indicate that even maintenance of the [rho+] genome is severely compromised (Fig. 6B and C).

An alternative function for Exo5 could be in recombination. We think this is unlikely because of the extreme repeat variability displayed in the exo5Δ mutant strains that can only be a consequence of very active recombination. All of the [rho+] clones that we generated by eviction of EXO5 had amplified either ori3 or ori5, together with variable-length sequences surrounding either ORI sequence. These regions showed high variability during propagation, indicative of very active

<table>
<thead>
<tr>
<th>WT</th>
<th>70-UV</th>
<th>15-CPT</th>
<th>30-HU</th>
</tr>
</thead>
<tbody>
<tr>
<td>YPD</td>
<td>exo5Δ</td>
<td>rad27Δ</td>
<td>exo5Δ</td>
</tr>
<tr>
<td>rad27Δ</td>
<td>exo1Δ</td>
<td>exo5Δ</td>
<td>exo1Δ</td>
</tr>
<tr>
<td>exo1Δ</td>
<td>exo5Δ</td>
<td>exo5Δ</td>
<td>miplΔ</td>
</tr>
</tbody>
</table>

VOL. 30, 2010 YEAST MITOCHONDRIAL Exo5 1465
recombination (Fig. 6B, C, and E). In addition, sequences inside the repeat unit were subject to additional partial duplication events, suggesting a hyperrecombination phenotype in the mutants (Fig. 6D).

The generation of conditional mutations in EXO5 should be of invaluable help in delineating its contribution to mitochondrial genome maintenance. As Exo5 is conserved in higher organisms, including humans (Fig. 2A), resolving the mechanism of yeast EXO5 may have important implications for understanding genome stability in higher organisms.

ACKNOWLEDGMENTS

We thank David Phillips for help with the tetrat analysis, Françoise Foury for advice and strains, and John Majors for critical discussion during the progress of this study. This work was supported in part by grant GM032431 from the National Institutes of Health.

REFERENCES

CHAPTER IIB

Human Exonuclease 5 is a novel sliding exonuclease required for genome stability
Human Exonuclease 5 Is a Novel Sliding Exonuclease Required for Genome Stability*

Justin L. Sparks‡, Rakesh Kumar§, Mayank Singh§, Marc S. Wold‡, Tej K. Pandita§1, and Peter M. Burgers‡2

From the‡Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, Missouri 63110, §Department of Radiation Oncology, University of Texas Southwestern Medical Center, Dallas, Texas 75390, and §Department of Biochemistry, University of Iowa Carver College of Medicine, Iowa City, Iowa 52242

Required for Genome Stability*

Background: Deoxiribonucleases are key DNA metabolic enzymes, but their functions remain ill defined.

Results: Human EXO5 is a novel bidirectional single strand-specific sliding exonuclease; however, RPA enforces a 5′-directionality of nuclease activity.

Conclusion: hEXO5 functions in nuclear genome maintenance and interstrand cross-link repair.

Significance: Nucleases are important in directing pathway choices at DNA damage and replication blocks.

Previously, we characterized Saccharomyces cerevisiae exonuclease 5 (EXO5), which is required for mitochondrial genome maintenance. Here, we identify the human homolog (C1orf176; EXO5) that functions in the repair of nuclear DNA damage. Human EXO5 (hEXO5) contains an iron-sulfur cluster. It is a single-stranded DNA (ssDNA)-specific bidirectional exonuclease with a strong preference for 5′-ends. After loading at an ssDNA end, hEXO5 slides extensively along the ssDNA prior to cutting, hence the designation sliding exonuclease. However, the single-stranded binding protein human replication protein A (hRPA) restricts sliding and enforces a unique, species-specific 5′-directionality onto hEXO5. This specificity is lost with a mutant form of hRPA (hRPA-t11) that fails to interact with hEXO5. hEXO5 localizes to nuclear repair foci in response to DNA damage, and its depletion in human cells leads to an increased sensitivity to DNA-damaging agents, in particular interstrand cross-link repair. Depletion of hEXO5 also results in an increase in spontaneous and damage-induced chromosome abnormalities including the frequency of triradial chromosomes, suggesting an additional defect in the resolution of stalled DNA replication forks in hEXO5-depleted cells.

Exonucleases are versatile processors of metabolic intermediates during DNA metabolism involving replication and recombination and the various pathways that function during the response to DNA damage in the cell (1). By definition, exonucleases need a DNA end for activity, and the vast majority generates mononucleotides as the products of nuclelease action. However, some exonucleases with specialized functions can cut internally of the DNA end that they initially engage. An example is the flap exonuclease Fen1 that cuts precisely at the base of a 5′-flap after having loaded at the 5′ terminus during the process of Okazaki fragment maturation (2, 3). The novel exonuclease described in this study, human EXO5 (hEXO5),3 loads at ssDNA ends and then slides along the ssDNA prior to cutting (sliding exonuclease). Our studies suggest that hEXO5 functions in the repair of DNA damage, in particular ultraviolet (UV) irradiation and interstrand cross-link (ICL) damage.

Interstrand cross-links are genotoxic lesions that covalently link the two strands of the DNA duplex together. DNA strand separation is an essential step in the processes of DNA metabolism including transcription, replication, and recombinational repair. Interstrand cross-links create complete obstructions of these fundamental DNA metabolic processes, leading to potent genotoxicity. Importantly, ICLs can arise from exogenous agents such as the anticancer drugs cisplatin and mitomycin C (MMC) and endogenous sources in the form of by-products of metabolic processes (4). In general, these agents also form intrastrand cross-links (5). However, the interstrand cross-links are generally considered to be much more genotoxic. ICL repair predominantly occurs during S phase while DNA replication takes place and is initiated by the convergence of replication forks at sites of ICLs (6).

Here, we describe the biochemical properties and biological function of a novel human bidirectional exonuclease, which is a sequence homolog of Saccharomyces cerevisiae Exo5 that we characterized previously as an exonuclease essential for mitochondrial genome maintenance (7). However, through a fascinating turn of evolutionary events, it appears that the distinct mitochondrial function of Exo5 prevails only in the Saccharomyces order of fungi because they possess a strong mitochondrial localization signal. Other fungi and organisms including mammals lack such a localization signal and may have dedicated Exo5 to the task of maintaining nuclear genome stability (see Fig. 1A). All members of the Exo5 family share some common characteristics beyond that of primary amino acid sequence. They possess an iron-sulfur cluster that is structurally important in linking the N terminus to the C terminus of

*This work was supported, in whole or in part, by National Institutes of Health Grants GM032431 and GM083970 (to P. M. B.); CA123232, CA129537, CA154320, and U19A1091175 (to T. K. P.); and GM44722 (to M. S. W).
1To whom correspondence may be addressed. Tel.: 214-648-1918; Fax: 214-648-4995; E-mail: tej.pandita@utsouthwestern.edu.
2To whom correspondence may be addressed. Tel.: 314-362-3872; Fax: 314-372-7183; E-mail: burgers@biochem.wustl.edu.
3The abbreviations used are: hEXO5, human EXO5; Exo5, exonuclease 5; ssDNA, single-stranded DNA; RPA, replication protein A; HRPA, human replication protein A; ICL, interstrand cross-link; MMC, mitomycin C; mEXO5, mouse EXO5.
Role of hEXO5 in DNA Repair

the enzyme, thereby likely creating a cavity that may encircle the ssDNA. A similar structural motif was previously identified in the Bacillus subtilis recombinase AddAB (8). However, our study shows that the detailed biochemical properties of hEXO5 are quite different from those of the budding yeast enzyme, and this may reflect their distinct functions in the nucleus versus mitochondrion, respectively. Depletion of hEXO5 in human 293 cells leads to hypersensitivity to genotoxic agents including UV irradiation and particularly to ICL-inducing agents. However, sensitivity to ionizing radiation exposure was not observed in hEXO5-depleted cells. Depletion of hEXO5 also leads to the accumulation of a higher percentage of chromosome aberrations either spontaneously or after treatment with cross-linking agents. In particular, an accumulation of triradial chromosomes was observed at metaphase that is indicative of unresolved and collapsed replication forks (9). These biochemical and genetic results suggest that hEXO5 plays a role in genome stability in general.

EXPERIMENTAL PROCEDURES

Plasmids and Oligonucleotides—Plasmid pBL277 contains the Schistosoma japonicum glutathione S-transferase (GST) gene fused to the N terminus of the human EXO5 (C1orf176) gene in vector pRS424–GALGST (10). The GST tag is separated from the N terminus of hEXO5 by a recognition sequence for the human rhinoviral 3C protease (LEVLFQDEL) (11). Following cleavage by the protease, the N-terminal sequence of hEXO5 is extended with the GPEF sequence. All variants and mutants were made in pBL277. Plasmid pBL276 contains the GST tag fused to the N-terminal 220-amino acid domain of the Escherichia coli GyrB gene (11) followed by a six-amino acid linker fused to the N terminus of hEXO5. Plasmid pBL272 is a plasmid for mammalian expression with a C-terminal GFP-hEXO5 construct. Plasmids and sequences are available upon request.

The following oligonucleotides were purchased from Integrated DNA Technologies, Inc. (Coralville, IA) and purified by urea-polyacrylamide gel electrophoresis (PAGE): c81, TTGC-GCATGAACCTTTTTTTTTAGTACAGACCTT; v81, AAGGTCTCCATCAAAAAAAAAAGTTCCATCGGCAA. The po-lymerase chain reaction (PCR) primers were synthesized by Integrated DNA Technologies, Inc. (Coralville, IA) and purified by PAGE. The primers were annealed against a buffer of 30 mM HEPES (pH 7.4), 100 mM NaCl, 500 μM MgCl₂, 0.1 mM EDTA, 1 mM dithiothreitol, and 0.005% polyoxyethylene (10) lauryl ether. 5 μM sodium bisulfite, 5 μM pepstatin A, and 5 μM leupeptin; subscript number indicates the mM sodium acetate concentration) until the lystate conductivity was equal to that of buffer A₁₀₀. The lystate was then used for batch binding to 1 ml of glutathione-Sepharose 4B beads (GE Healthcare), equilibrated with buffer A₄₀₀ and gently rotated at 4 °C for 2 h. The beads were collected at 1,000 rpm in a swinging bucket rotor followed by batch washes (3 × 20 ml of buffer A₁₀₀). The beads were transferred to a 10-ml column and washed at 2.5 ml/min with 100 ml of buffer A₄₀₀. The second wash was with 50 ml of buffer A₄₀₀ containing 5 mM magnesium acetate and 1 mM ATP, and the third wash used 50 ml of buffer A₄₀₀ and 30 ml of buffer A₂₀₀. Elution was carried out at a flow rate of 0.2 ml/min with buffer A₂₀₀ containing 20 mM glutathione (pH adjusted to 8.0). Fractions containing EXO5 were combined and diluted with buffer A₀ to equal buffer A₁₀₀ and then loaded onto a 1-ml Mono Q column. Protein was eluted with a linear gradient of buffer A₁₀₀ to A₁₂₀₀. The fractions containing pure protein were incubated overnight at 4 °C with 30 units of rhinoviral 3C protease, diluted with A₀ to equal A₁₀₀ and loaded onto a 1-ml Mono Q column. Pure EXO5 protein was eluted at 300–400 mM sodium acetate. mEXO5 and hEXO5 variants and mutants were overexpressed and purified similarly.

Analytical Ultracentrifugation—Purified protein was dialyzed against a buffer of 30 mM HEPES (pH 7.4), 100 mM sodium acetate, 5% glycerol, and 5 mM β-mercaptoethanol overnight at 0 °C. We used a Beckman Optima XL-A analytical ultracentrifuge and ran a velocity experiment at 40,000 rpm for 6 h with 5 μM hEXO5. The traces were analyzed using the Sedfit program.

Exonuclease Assays—The standard 10-μl assay mixture contained 100 mM Tris-HCl (pH 7.8), 500 μg/ml bovine serum albumin, 5 mM DTT, 5 mM magnesium acetate, 50 mM NaCl, 50–100 fmol of ³²P-end-labeled oligonucleotide substrate, and enzyme. Incubations were carried out at 30 °C for the indicated time periods. Deviations from the standard assay conditions are indicated in the legends of the figures. Reactions were stopped with 10 mM final concentration of EDTA in addition to 40% formamide and analyzed by 7 μm urea-18% PAGE. Dried gels were subjected to phosphorimaging analysis.

Forced Dimerization Assay—The standard 10-μl assay mixture was the same as for the standard exonuclease assay; however, all assays were carried out under initial linear rate conditions in which <30% of the substrate was hydrolyzed. Coumermycin (Sigma-Aldrich) was dissolved in DMSO and further diluted in H₂O, and novobiocin (Sigma-Aldrich) was dissolved in H₂O. Reactions were stopped as described above. The reactions were analyzed by 7 μm urea-18% PAGE. Dried gels were analyzed by phosphorimaging.

Electrophoretic Mobility Shift Assay (EMSA)—The standard 20-μl EMSA assay mixture contained 100 mM Tris-HCl (pH 7.8), 500 μg/ml bovine serum albumin, 5 mM EDTA, 1 mM DTT, 50 mM NaCl, 20% glycerol, 50 fmol of ³²P-end-labeled oligonucleotide substrate or unlabeled oligonucleotide, and enzyme. Incubations were carried out at 25 °C for the indicated periods. Reactions were loaded on 5% non-denaturing polyacrylamide gels prerun for 2 h at 4 °C. Either gels were dried and
subjected to phosphorimaging analysis, or the proteins were transferred to nitrocellulose for Western analysis with rabbit antibodies to hEXO5. Deviations from the standard assay conditions are indicated in the legends to the figures.

Human Cell Growth and Transfection—Human 293 and GM5849 cells were maintained by previously published procedures (13). Full-length FLAG-tagged hEXO5 cDNA cloned in pcDNA3.1 was transfected into cells. hEXO5 small interfering RNA (siRNA) and control luciferase siRNA were obtained from Integrated DNA Technologies, Inc. or Dharmacon Research (Lafayette, CO). The hEXO5 siRNAs were as follows: ORF176-1, ACUCAGAACUGUGUGAUCUC + GUUCA-CACCAUGUUCAGUUC; ORF176-2, CUGUGAACUGU-UUGGGUGAUU + UCACCCAAAGACUUCAGUUC. RNA interference (RNAi) treatment of 293 cells was performed as described previously (14). Cells were used 72 h after transfection for all experimental purposes.

Human Damage Sensitivity Assays—Clonogenic survival was determined using human 293 cells. Cells after 72 h of transfection with control or hEXO5 siRNA were seeded at known densities onto 60-mm dishes in 5.0 ml of medium, incubated for 16 h, and washed with 1× phosphate-buffered saline (PBS) prior to UV or ionizing radiation or exposure to the indicated doses of mitomycin C for 24 h or cisplatin for 1 h. Cells were washed and incubated in fresh medium for ~12 days and then fixed in methanol-acetic acid (3:1) prior to staining with crystal violet. Only colonies containing >50 cells were counted. Each experiment was repeated three to four times. The S.E. is given in the figures. Chromosome aberrations were analyzed at metaphases, which were prepared by standard procedures (15, 16). Cells were treated with cisplatin or mitomycin C, and metaphases were collected after different time points of drug treatment.

Immunostaining—Cells grown in chamber slides were exposed to irradiation (10 J/m²) and incubated at 37 °C prior to fixation. Cells were fixed in 2% paraformaldehyde for 15 min, washed with 1× PBS, permeabilized for 5 min on ice in 0.2% Triton X-100, and blocked in PBS with 1% bovine serum albumin. The procedure used for immunostaining is the same as that described previously (17–19).

RESULTS

hEXO5 Contains a Conserved Iron-Sulfur Cluster—The catalytic domain of eukaryotic Exo5 is distantly related to the E. coli RecB nuclease domain of the bacterial RecBCD recombinases (Fig. 1, B and C) (8). We recently reported that S. cerevisiae Exo5 is essential for mitochondrial genome maintenance (7). However, a phylogenetic and cellular localization prediction analysis of 95 eukaryotic Exo5 proteins indicates that strong mitochondrial localization prevails only in the Saccharomycetales order of fungi (Fig. 1A). This phylogenetic analysis along with our biochemical and functional cellular studies suggests that in other fungi and in metazoans Exo5 has a role in maintaining nuclear genome stability. Therefore, we carried out a biochemical and genetic analysis of hEXO5. The cDNA for human C1orf176 was cloned, and the gene was designated EXO5. hEXO5 with an N-terminal GST tag was overexpressed in a yeast overexpression system, and the recombinant protein was purified to apparent homogeneity by a combination of affinity and ion exchange chromatography (see “Experimental Procedures”). An analysis of hEXO5 by analytical ultracentrifugation showed that the protein exists in a monomeric form (Fig. 1D). The homolog from Mus musculus (RP23-182D20.4, Exo-5) was also cloned, and the enzyme was purified analogously (data not shown).

Upon purification of hEXO5, we noticed that it was yellow-brown in color, suggesting the presence of an iron-sulfur cluster. Indeed, the UV-visible spectrum of hEXO5 showed a peak at 410 nm, indicative of the presence of an iron-sulfur cluster, most likely of the [4Fe-4S] form (Fig. 1E). From aligning several Exo5 homologs, we noticed a set of four conserved cysteine residues, one N-terminal with respect to the catalytic core and three C-terminal of the catalytic core. This is a unique arrangement for coordination of the Fe-S cluster that was previously identified in the B. subtilis AddB helicase-nuclease (Fig. 1C) (8). This arrangement is conserved in mEXO5, which also purifies as a yellow protein with a UV-visible peak at 410 nm. To test whether the conserved cysteines are responsible for coordination of the iron-sulfur cluster, we purified a hEXO5 form with cysteines at 343 and 346 mutated to alanines (hEXO5-CCAA), and the mutant hEXO5 protein was colorless and lacked absorption at 410 nm (Fig. 1E). Lack of the iron-sulfur cluster also led to an 80–90% loss of catalytic activity of hEXO5, indicating that this structural motif is important but not essential for enzymatic function (Fig. 1G).

Both human and mouse Exo5 degraded ssDNA; however, their activities were both quite low with turnover numbers in the order of min⁻¹ (Fig. 1F). When mammalian Exo5 was overexpressed in E. coli, the persistent presence of contaminating E. coli nucleases made a reliable characterization problematic (data not shown). Therefore, we exploited the yeast overexpression system (described under “Experimental Procedures”), which dramatically reduced the level of contaminating nuclease activities. Considering the very low catalytic activity of hEXO5, it was important to establish that the observed activity was not caused by a contaminating yeast nuclease. Based upon the alignment with the RecB nuclease domain (20), we predicted that Asp-182 and Glu-196 bind the divalent metal ion that is essential for catalysis (Fig. 1, B and C). The E196A form of hEXO5 was purified and shown to display negligible nuclease activity (<1% of wild type), indicating that the observed wild-type activity albeit low is valid (Fig. 1G). We also purified two naturally occurring variants of hEXO5, either with a D115N or V172G change, and these variants had activity similar to that of the reference hEXO5 (Asp-115 and Val-172), which we designate as wild type (Fig. 1G). Below we explore the basis for the low enzymatic activity of mammalian Exo5.

The Exo5 Orthologs Are Sliding Bidirectional Exonucleases—A variety of oligonucleotide structures were analyzed as substrates for hEXO5. The enzyme is specific for DNA as single-stranded RNA was not degraded (data not shown). DNA structures such as Holliday junctions, replication forklike structures, and DNA flap substrates showed no activity unless these structures contained a free ssDNA tail, and none of these structures showed a higher activity than a simple ssDNA substrate did (Fig. 2, A and C, and data not shown). Thus, hEXO5 is a
Role of hEXO5 in DNA Repair

FIGURE 1. Catalytic activity of hEXO5. A, phylogenetic analysis. Using PSI-BLAST, 108 eukaryotic Exo5 sequences were identified in GenBank. This set was culled by imposing the following requirements: (i) an N-terminal methionine and (ii) the diagnostic WCE and CK,CK,C motifs N-terminal and C-terminal of the catalytic domain that constitute the iron-sulfur cluster coordinating cysteines. The remaining 95 sequences were subjected to PREDATOR to identify classical N-terminal mitochondrial localization sequences and to Phylogeny.fr for phylogenetic analysis using MUSCLE multiple alignment and PhyML phylogeny analysis. All 15 Saccharomycetales Exo5 proteins showed a medium to strong mitochondrial localization probability (0.2–0.92). The mitochondrial localization probability of all 80 other Exo5 proteins was 0–0.02 with the exception of S. pombe, variants and mutants. Standard assay mixtures used 10 nM 5'-o- or 3'-exonuclease domain of hEXO5 with conserved active site residues and iron-sulfur coordinating cysteine residues indicated. C, sequence alignment of critical residues in the Exo5 family. Active site Asp/Glu are shown in red, Gin and Tyr of a highly conserved α helix are shown in green, and four conserved Cys residues that coordinate the Fe-S cluster are shown in brown. Hsap, Homo sapiens; Mmus, M. musculus; Spom, S. pombe; Scer, S. cerevisiae; RecB, nuclease domain of E. coli RecBCD recombinase; AddB, nuclease domain of B. subtilis AddA8 recombinase. D, purified hEXO5 (5 μM) in buffer containing 100 mM sodium acetate was run in a Beckman Optima XL-A analytical ultracentrifuge as described under “Experimental Procedures.” Sedim. coeff., sedimentation coefficient. E, UV spectrum of hEXO5 and the C143A,C146A (CCAA) mutant. Traces were set to 1.0 at 279 nm.

single strand-specific deoxyribonuclease, and fully double-stranded DNAs (dsDNAs) are resistant to degradation (Fig. 2A, lanes 5–7 versus 11–13). The enzyme also requires a free DNA end for activity as the circularized form of the single-stranded oligonucleotide was not cleaved (lanes 8–10).

To determine whether hEXO5 is a 5'- or 3'-exonuclease, we utilized an ssDNA substrate hybridized with a 12-nucleotide complementary oligonucleotide to form a dsDNA block at either the 5'-end or the 3'-end. Although the 5'-blocked substrate showed a large decrease in activity (Fig. 2A, lanes 14–16), the 3'-blocked substrate showed only marginal inhibition (lanes 17–19), suggesting that hEXO5 is primarily a 5'-exonuclease, but it may also have 3'-exonuclease activity (Fig. 2B). We designed two partially double-stranded oligonucleotides with either a 5'- or 3'-ss-(d15) flaps (Fig. 2C). Both the 5'- and 3'-flip substrates were degraded by hEXO5 with the 5'-flip substrate being the preferred substrate. These data suggest that hEXO5 is a bidirectional single strand-specific exonuclease with a preference for 5'-ends. However, an alternative explanation that the enzyme slides onto the dsDNA ends of the flap substrates prior to cutting into the ssDNA region could not be excluded by these experiments.

To test the exonuclease polarity problem in an unambiguous manner, we used a set of oligo(dT) substrates in which the chain polarity is switched in the middle of the oligonucleotide by either a 3'-3' or 5'-5'-dinucleotide linkage so that the resulting oligo(dT) has either two 5'- or two 3'-ends, respectively (Fig. 2D). The use of homopolymeric oligo(dT) substrates also eliminates potential problems in activity relating to sequence context (Fig. 2A) or secondary structure formation. Control studies showed that the proofreading 3'-exonuclease activity of DNA polymerase δ degraded the substrate with only 3'-ends but not the substrate with only 5'-ends (Fig. 2D). Conversely, a 5'-exonuclease, a 5'-nuclease, degraded the substrate with only 5'-ends...
but not the substrate with only 3'-ends (data not shown). hEXO5 degraded both polarity switch oligonucleotides, indicating that its activity is bidirectional in agreement with the studies of the flap substrates (Fig. 2C).

Using these diagnostic substrates, we reinvestigated the polarity preference of *S. cerevisiae* Exo5, which we previously had designated to be a 5'-exonuclease (7). Surprisingly, this enzyme also showed bidirectionality but still with a preference for 5'-ends. However, in contrast to hEXO5, which generates larger oligonucleotide products, yeast Exo5 predominantly generates dinucleotide products from the 5'-end and tri- and tetranucleotides from the 3'-end (Fig. 2D). We conclude that our previous inability to detect a 3'-directionality with yeast Exo5 may have been caused by the sequence context of the oligonucleotides used in that study.

Remarkably, hEXO5 releases oligonucleotides rather than mononucleotides as products of exonuclease action. The results from a single hit kinetic analysis of hEXO5 on (dT)₆₀ are consistent with a model in which hEXO5 loads at a DNA end and then stochastically translocates along the ssDNA prior to random cutting (Fig. 2E). In a single hit kinetic analysis, each DNA molecule undergoes one cleavage event upon binding by hEXO5. Under our experimental conditions, hydrolysis at early time points was <10%. Importantly, the ratio of the different length products formed between 2 and 4 min of reaction remained constant. The ratio of 2–10-nucleotide products

---

**FIGURE 2.** hEXO5 is a single-stranded DNA-specific bidirectional exonuclease. A, standard assays containing a 10 nM concentration of the indicated 5’-labeled DNA were incubated with 3, 10, or 21 nM hEXO5 for 8 min. PDE, ladder from partial digestion with snake venom phosphodiesterase, a 3’-exonuclease. B, quantification of activity of hEXO5 on ssDNA, dsDNA, and 5’- and 3’-dsDNA substrates from A. C, standard assays containing a 10 nM concentration of the indicated 5’- or 3’-labeled DNA substrate were incubated with 5, 15, 50, or 150 nM hEXO5 for 8 min at 30 °C. hydrotol, hydrolysis. D, standard assays containing a 10 nM concentration of 5’-labeled (dT)₆₀ or double 5’-ended (dT)₇₀, or 3’-labeled double 3’-ended (dT)₇₀ were incubated with 5, 15, 50, 150, or 300 nM hEXO5 for 4 min or with 0.05, 0.15, or 5 nM *S. cerevisiae* (S. cer) Exo5 for 30 s at 30 °C. Controls contained DNA polymerase (Pol) δ where indicated with an arrow under the gels. E, standard assays contained 10 nM 5’-labeled (dT)₆₀ and 5 nM hEXO5 for the indicated times. All samples were analyzed on 7 M urea-18% polyacrylamide gels and quantified by phosphorimaging. The asterisks indicate the position of the ³²P label.
made after 4 min compared with 2 min of reaction time was 1.89 ± 0.2. In comparison, the ratio of 11–20-nucleotide products was 1.95 ± 0.2. Multiple cleavage events would have led to an accumulation of shorter products with increasing time. This was not observed until the 15-min time point, indicating that at the initial times single hit kinetics prevailed.

The sliding model predicts an equal probability of cutting along the length of the DNA across which sliding occurs as indeed was observed for the (dT)$_{50}$ substrate with some preference for cutting near the 5’-end, which is the preferable loading site. mExo5 also exhibited this behavior of a sliding exonuclease (data not shown). However, S. cerevisiae Exo5, which functions in mitochondria, predominantly released dinucleotides as products from the 5’-end (Fig. 2D and Ref. 7). Therefore, the biochemistry of mammalian Exo5 is quite different from that of budding yeast Exo5.

**hExo5 Activity Is Greatly Increased by Forced Dimerization**—The extremely low mammalian Exo5 catalytic activity suggests that its activity may be regulated. Interestingly, we noted that under standard assay conditions with 10 nm substrate DNA the activity of hExo5 carrying the GST purification tag was approximately 30-fold higher than that of the protein after proteolytic removal of the GST tag (data not shown). However, the product distribution remained very similar (data not shown). We surmised that the increased activity of GST-hExo5 was caused by the tendency of GST domains to form homodimers (10), thereby effectively dimerizing the exonuclease domain. A likely explanation for the increased activity of the dimeric enzyme would be an increased binding affinity for DNA. A full kinetic analysis of the two forms of hExo5 showed that the native enzyme bound ssDNA over 100-fold more poorly than did GST-hExo5 (Fig. 3A). At saturating ssDNA concentrations, the catalytic activities of both species differed less than 2-fold. In agreement with this conclusion, we observed no detectable binding of hExo5 to ssDNA in an EMSA, whereas GST-hExo5 did form a detectable complex (Fig. 3B). The GST domain itself showed no DNA binding (data not shown).

To determine whether the increased exonuclease activity was due to protein dimerization or a consequence of the presence of an N-terminal fusion, we reinvestigated this problem using a system of forced chemical dimerization. The N-terminal domain of the E. coli gyrase B subunit binds the drugs novobiocin and coumermycin with high affinity (21). Although novobiocin binds the GyrB domain in a 1:1 complex, coumermycin, which has two GyrB binding sites, forms a 1:2 complex with GyrB. We fused the GyrB domain to hExo5 and measured its activity at 10 nm ssDNA concentrations with or without drugs (Fig. 3C). In the absence of antibiotic, the GyrB-hExo5 fusion protein had very low catalytic activity similar to that of monomeric hExo5 alone. Next, we titrated either novobiocin or coumermycin into a nuclease assay that contained a constant concentration of GyrB-hExo5 (215 nm). The addition of increasing concentrations of novobiocin had no effect on the nuclease activity of GyrB-hExo5. However, increasing concentrations of coumermycin led to a maximal ~10-fold increase in nuclease activity at 100 nm drug, an ~1:2 ratio of coumermycin to GyrB-hExo5, followed by a gradual decrease in activity at higher concentrations of coumermycin (Fig. 3C).

The latter observation is consistent with a 1:2 equilibrium of coumermycin-GyrB-hExo5 complexes that is driven to 1:1 complexes at higher coumermycin concentrations. Therefore, we conclude that high affinity DNA binding activity of hExo5 requires that it be in the form of a dimer.

**Human Replication Protein A (hRPA) Enforces a 5’-Directionality onto hExo5**—RPA is a highly conserved heterotrimeric complex that is essential for DNA replication and repair. It
acts as a hub protein that recruits repair proteins to sites of DNA replication stress and DNA damage (22). To determine whether hRPA shows a functional interaction with hEXO5, we performed an activity titration experiment with increasing concentrations of either hRPA, *S. cerevisiae* RPA, or *E. coli* single-stranded DNA-binding protein (SSB) on oligo(dT)$_{65}$ in the presence of a constant hEXO5 concentration. Increased coating of the ssDNA with either *S. cerevisiae* RPA or *E. coli* single-stranded DNA-binding protein progressively inhibited hEXO5 activity, whereas hRPA stimulated hEXO5 activity (Fig. 4, A and B). The concentration of the hEXO5 is 50 nM. The concentrations of WT hRPA, hRPA-t11, *S. cerevisiae* RPA (scRPA), and *E. coli* single-stranded DNA-binding protein (SSB) from left to right are 0, 5, 10, 20, and 40 nM. The reactions were carried out at 30 °C for 4 min. The results were analyzed on a 7 M urea-18% polyacrylamide gel. B, quantification of the gel from A. C, standard assay mixtures using 10 nM 5’-labeled (dT)$_{65}$ or (dT)$_{70}$ containing either a 3’-3’ or 5’-5’-dinucleotide linkage ssDNA substrate labeled at the 5’- or 3’-ends respectively. The concentration of the hEXO5 is 50 nM. The concentrations of human RPA from left to right are 0, 2.5, 5, 10, 20, and 40 nM. The reactions were carried out at 30 °C for 4 min. The results were analyzed on a 7 M urea-18% polyacrylamide gel. Graphs to the right show the distribution of hEXO5 cleavage products without hRPA or with 40 nM hRPA. The asterisks indicate the position of the $^{32}$P label.

FIGURE 4. hEXO5 is an RPA-directed 5’-exonuclease. A, standard assay mixtures using 10 nM 5’-labeled (dT)$_{65}$. The concentration of the hEXO5 is 50 nM. The concentrations of WT hRPA, hRPA-t11, *S. cerevisiae* RPA (scRPA), and *E. coli* single-stranded DNA-binding protein (SSB) from left to right are 0, 5, 10, 20, and 40 nM. The reactions were carried out at 30 °C for 4 min. The results were analyzed on a 7 M urea-18% polyacrylamide gel. B, quantification of the gel from A. C, standard assay mixtures using 10 nM 5’-labeled (dT)$_{65}$ or (dT)$_{70}$ containing either a 3’-3’ or 5’-5’-dinucleotide linkage ssDNA substrate labeled at the 5’- or 3’-ends respectively. The concentration of the hEXO5 is 50 nM. The concentrations of human RPA from left to right are 0, 2.5, 5, 10, 20, and 40 nM. The reactions were carried out at 30 °C for 4 min. The results were analyzed on a 7 M urea-18% polyacrylamide gel. Graphs to the right show the distribution of hEXO5 cleavage products without hRPA or with 40 nM hRPA. The asterisks indicate the position of the $^{32}$P label.

Role of hEXO5 in DNA Repair

DECEMBER 14, 2012 • VOLUME 287 • NUMBER 51

JOURNAL OF BIOLOGICAL CHEMISTRY

42779

46

The DNA, coating of the DNA with RPA shifted the cutting probability strongly toward the 5’-end (Fig. 4C, right panel and traces). This observation is most consistent with two simultaneous activities of hRPA. (i) It inhibits sliding of hEXO5 and stimulates cutting near the site of 5’-loading, and (ii) it inhibits 3’-loading of hEXO5.

To test this model, we used the set of oligo(dT) polarity switch oligonucleotides described in Fig. 2D. Coating of the 5’-double-ended substrate with hRPA enhanced activity and shifted cleavage to smaller products similarly to what was observed with the natural (dT)$_{65}$. In contrast, cutting of the 3’-double-ended substrate was strongly inhibited by hRPA (Fig. 4C). Thus, RPA inhibits 3’-activity of hEXO5 and thereby enforces a 5’-directionality onto the enzyme.
Role of hEXO5 in DNA Repair

To further examine the specificity of the hEXO5-hRPA interaction, we took advantage of a previously characterized mutation in the N-terminal domain of the RPA70 subunit, RPA70-K45E or RPA-t11. This allele was first identified in S. cerevisiae as a mutant (rfa1-t11) that caused sensitivity to DNA damage (23), and the human mutant exhibits recombination defects (24). Mutant hRPA-t11 failed to stimulate hEXO5. In fact, it inhibited hEXO5 activity to a degree similar to that of S. cerevisiae RPA, suggesting that the mutation abrogated stimulatory interactions with hEXO5 (Fig. 4A).

We used EMSA to assess interactions among hEXO5, hRPA, and ssDNA (Fig. 5). In these studies, DNA was detected by phosphorimaging of the 5′-32P label and hEXO5 by Western analysis. The (dT)8 oligonucleotide used in this study can bind either one or two hRPAs depending on the concentration (complex 1 in lane 9 and complex 2 in lanes 10 and 11). A complex between hEXO5 and hRPA was readily detectable by Western analysis (lane 2). Under these conditions, no complex between hEXO5 and ssDNA was detectable (lane 4; see also Fig. 3B). However, addition of increasing hRPA shifted hEXO5 into hRPA-DNA complex 1 (lane 5) or complex 2 (lanes 6 and 7). These data suggest that hEXO5 is recruited to ssDNA through binding to DNA-bound hRPA. Given the lack of stimulation of hEXO5 activity exhibited by mutant hRPA-t11 (Fig. 4A), we also determined interactions of the mutant protein with hEXO5. Indeed, hRPA-t11 did not form a complex with hEXO5 in the absence of DNA (compare lanes 12 and 4), and although hRPA-t11 alone showed robust binding to ssDNA (lane 13), it did not recruit hEXO5 into the RPA-ssDNA complex (compare lanes 14 and 7). Therefore, hRPA-t11 is defective for binding hEXO5 on or off the DNA. These data together indicate that hEXO5 physically interacts with hRPA as a mechanism to increase the affinity for ssDNA. Also, hRPA enforces 5′-directionality onto hEXO5.

hEXO5 Functions in UV Irradiation and Interestrand Cross-link Repair—Budding yeast Exo5 has a consensus mitochondrial localization signal, and the exo5Δ mutant shows catastrophic mitochondrial defects but no detectable nuclear defects (7). However, most other EXOS genes including the mammalian gene lack a mitochondrial localization signal, suggesting that their primary function may be nuclear (Fig. 1A). We tested the localization of a C-terminal GFP fusion to hEXO5 in transfected HEK 293 human cells. GFP fluorescence was distributed over both the cytosol and the nucleus (Fig. 6A). However, approximately 4 h after exposure of cells to UV irradiation, we observed distinct nuclear GFP foci, suggesting that hEXO5 is recruited to sites of DNA damage.

To determine the effect of hEXO5 on cell survival, cells with or without depletion by hEXO5-specific siRNAs were exposed to various genotoxic agents. RT-PCR was used to determine the level of hEXO5 mRNA knockdown to >80% (data not shown). Unfortunately, several independently raised antisera against purified hEXO5 failed to recognize the protein in human cell extracts. Therefore, to test the level of protein knockdown elicited by siRNAs, a FLAG-tagged hEXO5 construct containing the same RNA sequence that is targeted by the siRNAs was expressed. Extracts of hEXO5 siRNA-transfected cells were analyzed for FLAG-hEXO5 by Western analysis with anti-FLAG antibodies. From these studies, an ∼80% knockdown of FLAG-tagged hEXO5 was consistently achieved (Fig. 6B). From these studies, we conclude that levels of native hEXO5 are similarly decreased by the siRNAs.

Treatment of hEXO5-depleted cells with DNA-damaging agents caused an increased sensitivity to various DNA-damaging agents but not to γ-irradiation as determined by clonogenic survival assays. We consider off-target effects of the siRNAs unlikely because neither siRNA1 nor siRNA2 caused an increase in sensitivity to γ-irradiation (Fig. 6C), but both siRNA1 and siRNA2 showed strong and similar sensitivity to treatment of cells with cisplatin (Fig. 7A). hEXO5 depletion in 293 cells also caused an increased sensitivity to cell killing after UV irradiation and to the alkylating agent methylmethane sulfonate (Fig. 6, D and E). An analysis of the rate of appearance and disappearance of γ-H2AX foci associated with UV damage showed a delay in their disappearance, suggesting that the rate of UV damage repair is slowed down in hEXO5-depleted cells (Fig. 6F).

Interestingly, hEXO5-depleted cells are very sensitive to either interstrand cross-linking agent cisplatin or MMC as determined by clonogenic survival assays (Fig. 7, A and B). Defects in ICL repair such as in Fanconi anemia patients are associated with an increase in metaphases with chromosomal aberrations. In particular, there is an increased occurrence of triradial chromosomes that is attributed to a failure in the resolution of stalled DNA replication forks (25). Indeed, we observed that chromosome abnormalities were significantly increased in hEXO5-depleted cells even in the absence of external damage-inducing agents (Fig. 7, C, panel b, D, and E). Furthermore, treatment with either interstrand cross-linking agent cisplatin (Fig. 7D) or mitomycin C (Fig. 7E) led to a large increase in the frequency of metaphases with chromosomal aberrations in hEXO5-depleted cells compared with controls. Interestingly, these included a higher frequency of triradial chromosomes (Fig. 7C, panel d). Therefore, hEXO5 depletion results in an increase in genomic instability and an increased sensitivity to various DNA-damaging agents as determined by clonogenic survival, metaphase aberrations, and appearance and disappearance of repair foci.
DISCUSSION

The EXOS gene is widely conserved in archaeal and eukaryotic organisms including mammals, fungi, plants, and protozoa; however, surprisingly, it appears to be absent from insects and worms. A phylogenetic analysis of all 95 eukaryotic complete EXOS genes deposited in GenBank™ shows that the proteins from organisms that fall in the Saccharomycetales order are more distantly removed from those in other orders and phyla, and they show a strong mitochondrial localization signal (Fig. 1A). This phylogenetic and bioinformatics analysis is supported by our experimental studies, which show severe mitochondrial but no nuclear defects for S. cerevisiae exo5Δ (7). In contrast, other fungi and the mammalian forms generally lack a mitochondrial localization signal, and hEXO5-depleted human cells show strong defects in nuclear genome stability.

The basic structure of Exo5 in which the N terminus is linked to its C terminus via an iron-sulfur cluster is derived from the AddB nuclease domain of the B. subtilis AddAB recombinase, which in turn is related to the better known E. coli RecBCD recombinase (8, 26). The linkage of the N terminus to the C terminus of the protein through an Fe-S cluster could provide a cavity through which the ssDNA threads. This model would explain our observation that circular ssDNA is inactive for hEXO5 (Fig. 2A). The Okazaki fragment-processing exonuclease Fen1 also contains a hole through which the 5'-end of the ssDNA threads, although that hole is elaborated through a flexible loop (3). Both Exo5 and Fen1 show a propensity to slide along ssDNA; however, the hydrolytic activity of Fen1 is specific for the base of a flap substrate. No such preference has been found for Exo5 (Fig. 2C). In fact, Exo5 is structurally related to the Dna2 nuclease-helicase, which has a similar iron-sulfur cluster arrangement (27). Another biochemical similarity between Dna2 and Exo5 is that RPA enforces a 5’–3’ directionality onto the enzyme (Fig. 4 and Ref. 28). We show here that this enforcement of directionality requires species-specific hRPA-hEXO5 interactions as neither yeast RPA nor the interaction-defective hRPA-t11 mutant fulfill this function (Figs. 4 and 5). Directionality enforcement by RPA has been demonstrated previously for enzymes and factors involved in nucleotide excision repair and in the DNA damage checkpoint, re-emphasizing the importance in carrying out these type of in vitro studies with the cognate form of RPA (29, 30).

What might be the biochemical role of hEXO5 during DNA repair? Given the sensitivity of hEXO5-depleted cells to UV irradiation, methylmethane sulfonate, and ICL agents, the enzyme may function in multiple pathways with partial redundancy with other nucleases. hEXO5-depleted cells are very sensitive to cisplatin and MMC, which may indicate that they have a defect in the repair of interstrand cross-links. Alternatively, because ICLs are generally detected during DNA replication (6, 31), hEXO5-depleted cells may also be defective in the processing of stalled replication forks in general. The observation that chromosome abnormalities such as triaridals, which arise during S phase, are increased in hEXO5-depleted cells in the absence of damage supports this hypothesis.

Given its absolute requirement for single-stranded DNA, hEXO5 can be visualized as functioning together with a DNA helicase that generates 5’-ssDNA through unwinding or with a DNA polymerase that generates 5’-ssDNA through strand displacement synthesis. The identification of hEXO5-interacting...
FIGURE 7. Chromosome aberrations in hEXO5-depleted cells. A, clonogenic survival of control 293 cells or those depleted for hEXO5 using either siRNA1 or siRNA2 followed by exposure to the indicated concentrations of cisplatin. B, survival after exposure to MMC. C, representative images of metaphases with chromosome aberrations. Exponentially growing cells with or without depletion of hEXO5 were treated without or with 15 μg/ml cisplatin for 1 h, cells were washed and incubated in fresh medium for 24 h, and then Colcemid was added to collect metaphases. Thin arrows indicate tri- or quadriradial chromosomes, and thick arrows indicate breaks and gaps. Panel a, control siRNA without cisplatin treatment; panel b, hEXO5 siRNA without cisplatin treatment; panel c, hEXO5 siRNA irradiated with 2 grays; panel d, hEXO5 siRNA with cisplatin treatment. D and E, histograms showing the frequency of metaphases with aberrations in control, control siRNA-, or hEXO5 siRNA-transfected cells after cisplatin (D) or MMC exposure (E) as described under “Experimental Procedures.” The data presented are the mean of three experiments, and for each experiment, 100 metaphases were counted (significance according to Student’s t test: *, p < 0.05; **, p < 0.01). Error bars represent S.E.
proteins beyond human RPA identified here will aid in a further understanding of the specific roles that this novel exonuclease displays during the maintenance of genome integrity. In addition, our phylogenetic analysis suggests that fission yeast Exo5 may be more closely related to the human than to the budding yeast enzyme, and in fact, a proteomics study has indicated both nuclear and mitochondrial localization for Schizosaccharomyces pombe Exo5 (32). We have initiated a study of Exo5 in this genetically more tractable organism.

Acknowledgments—We thank Carrie Stith for expert technical assistance in plasmid construction and protein purification, Arun Gupta for technical help, Tim Lohman for critical discussions and a gift of reverse polarity oligonucleotides, and Roberto Galletto for critical distance in plasmid construction and protein purification, Arun Gupta

REFERENCES

CHAPTER IIC

*S. pombe* Exonuclease 5 plays a role in maintenance of both nuclear and mitochondrial genomes
Schizosaccharomyces pombe Exonuclease 5 Functions in both Nuclear and Mitochondrial Genome Maintenance

Justin L. Sparks¹ and Peter M. Burgers¹*

¹Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, MO 63110, USA
ABSTRACT

Previously we described the biochemical and molecular characterization of the novel Exonuclease 5 family of enzymes from *S. cerevisiae* and humans. The Exo5 family consists of bi-directional, single-strand, DNA-specific exonucleases that all contain an iron-sulfur cluster as a structural motif, and all have various roles in DNA metabolism. In the *Saccharomycetales* order, that includes the budding yeast *S. cerevisiae*, Exo5 is a mitochondrial protein that is essential for mitochondrial genome maintenance. In contrast, the human ortholog is important for nuclear genome stability and DNA repair. Here, we identify the Exo5 ortholog in an unrelated yeast species, *Schizosaccharomyces pombe*. *S. pombe* Exo5 (spExo5) is a bi-directional, single-strand, DNA-specific exonuclease with approximately 100-fold higher activity than the human enzyme, but functions similarly *in vitro*. Both RPA and PCNA co-immunoprecipitate with spExo5 and these interactions mediate the lethality of spExo5 overexpression in *S. pombe*. Similar to results from knockdown experiments in human cells, DNA damage sensitivity assays show exo5Δ *S. pombe* strains are sensitive to various DNA damaging agents, particularly interstrand crosslinking agents. SpExo5 is in a redundant pathway with spRad2 (FEN1) in both nuclear and mitochondria genome maintenance. Deletion of both enzymes lead to depletion of the mitochondrial DNA, indicating spExo5 and spRad2 are essential for mitochondrial DNA metabolism.
INTRODUCTION

Exonuclease 5 was discovered over two decades ago in a biochemical screen for exonucleases in *S. cerevisiae*; however, Exo5 was only recently analyzed (Burgers, Bauer et al. 1988; Burgers, Stith et al. 2010). *S. cerevisiae*, mouse and human Exo5 orthologs are bi-directional, single-strand, DNA-specific exonucleases. The human and mouse Exo5 nucleases slide along ssDNA prior to cleavage, but the *S. cerevisiae* enzyme does not slide; rather, it preferentially cleaves di-nucleotides as its major product. There is also apparent divergence in function of Exo5 in *Saccharomycetales* order, which includes *S. cerevisiae*, in which all the Exo5 orthologs have strong consensus mitochondrial localization sequences (Burgers, Stith et al. 2010; Sparks, Kumar et al. 2012). However, most orthologs outside the *Saccharomycetales* order lack a mitochondrial localization sequence, indicating their main function may be nuclear (Sparks, Kumar et al. 2012). In *S. cerevisiae* Exo5 is essential for mitochondrial genome maintenance but has no apparent nuclear function (Burgers, Stith et al. 2010), whereas studies of the human Exo5 ortholog revealed it to be a nuclear protein that is important for nuclear genome stability and DNA repair (Sparks, Kumar et al. 2012). We were interested a more genetically tractable model system to investigate the *in vivo* nuclear functions of Exo5 outside of the *Saccharomycetales* order, to better understand what DNA metabolism pathways human Exo5 may be involved.

*Schizosaccharomyces pombe*, the fission yeast, was an attractive candidate with previous proteome studies clearly demonstrating that the *S. pombe* Exo5 (spExo5) ortholog is localized to both the nucleus and mitochondria (Matsuyama, Arai et al. 2006). Additionally, knockdown of Exo5 in human cells hyposensitized to interstrand crosslinking agents compare to control cells (Sparks, Kumar et al. 2012). DNA Interstrand cross-links (ICLs) are genotoxic lesions that covalently link the single-strands of the DNA duplex stably together. The ability to destabilize and separate DNA strands is an essential step in the processes of DNA replication and transcription. ICLs form potent obstructions of these fundamental processes leading to potent genotoxicity that must be efficiently repaired for cell viability. In *S. pombe* there are at least three pathways for repair of interstrand crosslinks (ICLs) that have been genetically investigated. The first pathway is considered to be a primitive Fanconi anemia pathway that requires the activity of the nuclease Fan1, human Fan1 ortholog, the helicase Fml1 that is the human FANCM homolog, and Rhp51 (Rad51) in *S. pombe* (Fontebasso, Etheridge et al. 2013). The second pathway requires the nuclease Pso2 and the nucleotide excision repair endonuclease Rad13, as well as Rhp51 (Fontebasso, Etheridge et al. 2013). Finally, the third pathway for repair of ICLs in *S. pombe* is dependent on the sumo ligase Pli1
(Fontebasso, Etheridge et al. 2013). These previous studies provide a framework for the investigation of the function of Exo5 in ICL repair. The mechanistic details of these pathways have not been studied and more work will be required to determine what other proteins are involved in these pathways.

In this paper we describe the biochemical and genetic characterization of the *Schizosaccharomyces pombe* Exo5 homolog. We observe that spExo5 is a bi-directional single-strand DNA specific exonuclease that has strikingly similar activities to the previously characterized human and mouse homolog. We demonstrate that spExo5 physically interacts with both the single-stranded, DNA-binding protein RPA, and the processivity clamp Pcn1 (PCNA), and that these interactions are important for the spExo5 overexpression phenotype of checkpoint activation and cell death. The *exo5Δ* strains are sensitive to UV-irradiation, methyl methanesulfonate (MMS) and to interstrand crosslinking agents in particular, closely recapitulating the previous Exo5 knockdown experiment in human cells. Finally, we demonstrate that spExo5 contains an internal ribosome entry site at methionine 58 that produces a short isoform of spExo5 that is important for mitochondrial genome maintenance, and that this short isoform plays a redundant role with spRad2 (FEN1) in the mitochondria. An *exo5Δ rad2Δ* double deletion strain of *S. pombe* is deficient for growth on plates containing glycerol as the sole carbon source, and genomic DNA digestion from this strain indicates an approximate 10-fold depletion of mitochondrial DNA, indicating these proteins perform a redundant function that is essential for mitochondrial genome maintenance.

**RESULTS**

*S. pombe* Exo5 (spExo5) is a bi-directional, ssDNA-specific exonuclease, homolgous to the human Exo5.

Comparison of the biochemical activity of human, mouse, and *S. pombe* Exo5 homologs on a 5’-labeled ssDNA oligonucleotide reveal that these proteins have identical sequence preferences in their cleavage patterns (Fig 1A). Similar to other Exo5 family members spExo5 is a bi-directional exonuclease able to degrade single-stranded oligonucleotides engineered to have only free 5’-ends or 3’-ends respectively, leaving a distribution of products similar to the human enzyme (Fig 1B). The spExo5 seems to slide along DNA prior to cleavage and to have divergent activity from the *S. cerevisiae* Exo5 ortholog (Fig 1B). These results indicate that spExo5 is a bi-directional, ssDNA-specific exonuclease that possesses similar sliding activities to the human homolog (Fig 1B). Mutations of the conserved active site aspartates at amino acid 176 or 207 to alanines (D176A, D207A) abrogate the nuclease activity of spExo5 (Fig 1C). A full kinetic
analysis of both human Exo5 and spExo5 shows that the human enzyme binds ssDNA over 100-fold more poorly than spExo5 (Fig. 1D). At saturating ssDNA concentrations, the catalytic activity of spExo5 is 100-fold greater than the human enzyme. From these experiments we hypothesized that *S. pombe* would be a more genetically tractable model system to study Exo5 function in vivo.

*Exo5 interacts with RPA and PCNA in vivo.*

Previously we demonstrated a species-specific physical interaction between human Exo5 and the human single-strand DNA binding complex RPA (Sparks, Kumar et al. 2012). In an attempt to link Exo5 to a specific DNA repair pathway, we performed co-immunoprecipitation experiments to find Exo5 interacting proteins. We overexpressed Exo5 with a C-terminal 3XFLAG tag to perform our pull-downs. Cell extracts were subjected to co-immunoprecipitation with 3XFLAG-beads (Sigma) and extensively washed prior to elution with the 3XFLAG peptide (Sigma). The elutions were analyzed by SDS-PAGE followed by commassie staining (Fig 2A). The upper and lower sections of the Exo5-3XFLAG lane were cut out and sent for mass spectrometry analysis (Fig 2A). Mass finger printing identified several DNA replication and repair proteins with the full list outlined in Supplemental Table 1. Both the processivity clamp PCNA (spPcn1) and the single-strand binding protein RPA are identified as co-purifying with spExo5. To further test this interaction we obtained antibodies against the *S. pombe* Ssb1 (RPA70) and Pcn1 (PCNA) proteins. Indeed both RPA and PCNA are detected by western blot analysis in the Exo5 co-IPs but not in the controls, and the interaction is not mediated through DNA as treatment with DNase I does not eliminate these interactions (Figure 2B, C). The results from these Co-IP experiments identify spExo5 as possibly being involved at the replication fork through its interaction with the processivity clamp PCNA and RPA, but further work will be need to characterize this interaction. Exo5 does not have an apparent PCNA interaction motif and the PCNA interacting motif is within Exo5 is still unknown. The RPA interaction seems to be conserved from *S. pombe* to humans, but we still need to determine if spRPA loading on ssDNA has the same effect on Exo5 by enforcing a 5'-3' directionality on the exonuclease activity as seen with the human homologs. We hypothesize these interactions are relevant in vivo and is consistent with Exo5 being involved in DNA replication and repair in the nucleus.

*Overexpression of Exo5 in S. pombe leads to checkpoint activation and cell death by titrating away RPA and PCNA from their function in replication.*
Concomitant with our spExo5-co-IP experiments we made the observation that overexpression of spExo5 was lethal to the cells and this lethality was not due to the nuclease activity of the protein, as demonstrated by overexpression of a nuclease-deficient mutant (D207A) showed similar lethality (Fig 3A). Examination of the cell morphology revealed that the cells are elongated, which is indicative of checkpoint activation in *S. pombe* (Fig 3B). The checkpoint is mediated through the Rad3 (ATR) kinase, where by overexpression of spExo5 in a *rad3Δ* background eliminates the cell elongation phenotype, but does not suppress the lethality of overexpression (Fig 3A, B). We hypothesized that overexpression of spExo5 mediated checkpoint activation and lethality could be the result of titrating away an essential replication factor from performing its essential function during replication. To test this hypothesis we made a series of point mutation in spExo5 to look for mutations that suppressed the lethality of overexpression. Two motifs in the N-terminus of the protein are found to be essential for the lethality of overexpression of spExo5 (Fig 3C). First, we noticed the very N-terminus of the protein has an acidic patch from amino acid 2-11 where there are three sets of di-acidic amino acids, and mutating any set of these di-acidic pairs to alanines suppresses the checkpoint activation and lethality of overexpression of Exo5 (Fig 3C, D). Recently a di-acidic motif was identified in several proteins that mediate the interaction with the basic cleft of the F-domain of RPA70 (Frank, Vangamudi et al. 2014). More work will be needed to determine if this motif in Exo5 is the RPA binding surface. The other motif that is essential for both checkpoint activation and lethality of Exo5 occurs just prior to the first cystiene of the iron-sulfur cluster (Fig 3C). Mutation of either Y74A or D79A also suppresses both the checkpoint activation and lethality of overexpression (Fig 3C, D). A co-immunoprecipitation experiment demonstrates that these mutations disrupt the ability of Exo5 to pull-down both RPA and PCNA as shown by western blot analysis (Fig 3E). The results from our experiments indicate that there may be two RPA interacting domains in spExo5. Our data would be consistent with a model where overexpression of Exo5 leads to squelching of the free RPA and PCNA in the cell, blocking some essential function that causes checkpoint activation and ultimately cell death. Further studies will be needed in order to fully characterize these interactions.

*S. pombe* Exo5 is important for DNA repair.

In order to determine possible spExo5 DNA repair pathways, we carried out *in vivo* survival assays to various DNA damaging agents using *exo5Δ* strains. We detect no sensitivity to γ-irradiation (Sup. Fig 1A), however the deletion strain is sensitive to UV-irradiation (Fig 4B), alkylating agents (Fig 4B,C), and crosslinking agents such as cis-platin (Fig 4B) and 8-
methoxypsoralen (Fig 4A). Remarkably, these results largely recapitulate the results from the knockdown experiments in human cells (Sparks, Kumar et al. 2012). These results differ from the results of experiments in \textit{S. cerevisiae} exo5\textDelta strains where no sensitive was observed to any DNA damaging agents tested (Burgers, Stith et al. 2010). These results allowed us to use \textit{S. pombe} as a genetic tool to look for genetic interactions with known components of various DNA repair pathways that spExo5 might be involved.

We followed the damage sensitivity experiments by conducting a large set of epistasis experiments with deletions of known DNA repair genes. The results of these experiments are outlined in Supplementary Table 2. In \textit{S. pombe}, \textit{rad2}+ (FEN1) nuclease is required for Okazaki fragment maturation and long patch base-excision repair (Liu, Kao et al. 2004). Construction of a \textit{rad2}\textDelta exo5\textDelta deletion reveals a synthetic sickness phenotype, even in absence of DNA damage treatment (Fig 5A). Treatment of the \textit{rad2}\textDelta exo5\textDelta strain with either MMS or cis-platin reveals strong hypersensitivity to these DNA damaging agents that is consistent with a synergistic interaction (Fig 4C). Experiments with \textit{exo1}+, the multifunctional nuclease Exo1, showing partial redundancy for mismatch repair, double strand break repair, Okazaki fragment maturation, and interstrand crosslink repair (Tran, Erdeniz et al. 2004; Barber, Ward et al. 2005; Mimitou and Symington 2009); additionally, \textit{pso2}+, a nuclease required for interstrand crosslink repair (Lambert, Mason et al. 2003), reveals a strong hypersensitivity of the \textit{exo1}\textDelta exo5\textDelta double mutant to UV-irradiation and MMS (Fig 4B), and of \textit{exo1}\textDelta exo5\textDelta and the \textit{pso2}\textDelta exo5\textDelta double mutants to cis-platin (Fig 4B,D). Experiments with \textit{hrq1}+ the \textit{S. pombe} ortholog of the human RecQ4L helicase, that is involved in NER and ICL repair, reveals hypersensitivity of the \textit{hrq1}\textDelta exo5\textDelta mutant to cis-platin (Fig 4E). In contrast, we find little or no enhancement in damage sensitivity when the \textit{exo5} deletion combines with \textit{mre11}\textDelta, \textit{rad13}\textDelta, \textit{slx1}\textDelta, \textit{fan1}\textDelta, \textit{fmi1}\textDelta, \textit{mus81}\textDelta, \textit{swi10}\textDelta, \textit{rhp51}\textDelta, \textit{rqh1}\textDelta, or \textit{uve1}\textDelta (Supp. Fig. 1, 2, 3; Supp. Table 1). The results from these experiments are consistent with \textit{exo5}+ playing a redundant role with \textit{rad2}+ in DNA metabolism, even in the absence of exogenous DNA damage. The genetic interactions observed with \textit{exo5}+ and \textit{exo1}+, \textit{pso2}+ or \textit{hrq1}+ are consistent with additive interactions indicating these proteins are in different pathways that target different DNA lesions. These experiments are not sufficient when it comes to determining in what pathway or pathways Exo5 may be involved given that many pathways can utilize several different nucleases to accomplish the same function, making a clear, genetic delineation of the role of a specific nuclease difficult. Nevertheless, the integral role nucleases have in maintaining the genome can be appreciated from the degree of redundancy that is inherent in DNA repair pathways.
In *S. pombe* there are at least three distinct redundant pathways for repair of ICLs. One pathway was shown to involve *fan1*, *fml1*, and *rhp51*, a second requires *psd2*, *rad13*, and *rhp51*, and a third is dependent on the E3-sumo ligase *pli1* (Fontebasso, Etheridge et al. 2013). Our data is consistent with these previously published results, and we have expanded these pathways (Fig 2A,B). We made triple and quadruple mutants to attempt to overcome the inherent redundancies of the nucleases in these pathways (Supp. Fig 2, 3, 4). Our results indicate that Hrq1 and Swi10-Rad16 are epistatic with the Pso2-dependent ICL repair pathway and non-epistatic with the Fan1-dependent pathway (Supp. Fig 3A, B). Experiments with Rqh1, the BLM homolog in *S. pombe*, demonstrate synergistic sensitivity to cis-platin when combined with the Fml1 helicase of the Fan1 dependent pathway (Supp. Fig 3C). On the other hand, we observe that Slx1 and Exo1 are epistatic with the Fan1 pathway and non-epistatic with the Pso2 pathway (Supp. Fig 2A, 4). The pathway that utilizes Exo5 is still unclear. The *exo5Δ* is epistatic with the Fan1 pathway including Fan1 and Fml1 (Supp. Fig 2), but is also epistatic with several proteins in the Pso2 pathway including Swi10 and Rad13 (Supp. Fig 2, 3). It may be possible that Exo5 is involved in the Pli1 pathway, but more experiments are necessary to test that hypothesis. These experiments further delineate the proteins involved in either the Pso2 or Fan1 pathways of ICL repair, and builds context to place Exo5 in the repair of ICLs. Our proposed model for how these proteins fit into the previously characterized pathways of ICL repair are outlined in Figure 6A, B.

*S. pombe* Exo5 and Rad2 play a redundant role in mitochondrial DNA metabolism.

A proteomic localization study indicated both nuclear and mitochondrial localization for *S. pombe* Exo5 (Matsuyama, Arai et al. 2006). *S. pombe* is a petite negative yeast which requires a functional mitochondrial genome for growth (Haffter and Fox 1992). *S. pombe* *exo5Δ* strains are viable, indicating that spExo5 is not absolutely required for mitochondrial genome stability. Indeed, an *exo5Δ* strain transformed with an *exo5*+-containing plasmid grows indistinguishable from one transformed with a control vector on media containing glycerol as the non-fermentable carbon source (Fig 5A). Creation of an *exo5Δ rad2Δ* reveals a negative genetic interaction between *exo5*+ and *rad2*+, leading to the double-deletion strains showing synthetic sickness at all temperatures tested (Fig 5B). Further analysis reveals that the double mutant fails to grow on plates containing glycerol as the sole carbon source (Fig 5C). Exogenous expression of *exo5*+ (Fig 5C) or *rad2*+ (data not shown) restores the growth of these strains on glycerol plates, and this restoration requires the nuclease activity of Exo5 (Fig 5C).

*S. pombe* Exo5 contains several internal methionines that could serve as alternative translational start sites. Previously, human FEN1 was shown to have an alternative translation
start site leading to a short FEN1 isoform, and this shorter isoform is physiologically relevant and specifically localizes to the mitochondria, whereas the full-length protein localizes to the nucleus (Kazak, Reyes et al. 2013). When we express spExo5 in cells we observe two prominent Exo5 bands by western blotting analysis; the first is consistent with the predicted molecular weight of spExo5-3XFLAG protein (51 kDa), but the lower band is consistent with a protein starting at methionine 58 (44 kDa) (Fig 2D). Expression of exo5-M58A mutant with methionine 58 changed to an alanine eliminates the detection of the shorter isoform by western blot, and expression of an N-terminal truncation mutant exo5-(58-409) is the correct molecular weight and is consistent with the lower isoform starting at methionine 58 (Fig 2D). Surprisingly, the expression of the short isoform (exo5-(58-409)) is able to restore growth of the exo5Δ rad2Δ on glycerol plates, but expression of exo5-M58A mutant, which eliminates the translation of the short spExo5 isoform, does not restore growth. This result indicates that the short isoform is important for mitochondrial genome maintenance but the full-length protein is dispensable (Fig 5C). These data support a model in which the short isoform, derived from the internal start at methionine 58, is localized to the mitochondria (mitochondrial isoform) and full-length spExo5 isoform is not based on the fact that the M58A mutant is not able to restore growth on glycerol plates (nuclear isoform). It is also interesting that neither of the mutants, neither the mitochondrial isoform nor the M58A mutant, fully complement the DNA-damage sensitivity of the exo5Δ rad2Δ (Fig 5C). Further investigation is needed to determine the localization of these isoforms by microscopy.

Restriction digestion of total DNA from exo5Δ rad2Δ strains reveals the ~10-fold loss of mitochondrial DNA in these strains compared to the wild type strain, demonstrating that Exo5 and Rad2 are important for mitochondrial DNA metabolism (Fig 5D). The results from these experiments establish a redundant function in the mitochondria of Exo5 and Rad2, and demonstrate that two isoforms of Exo5 are produced in S. pombe, with the short isoform starting at methionine 58 is important for mitochondrial DNA metabolism.

DISCUSSION

The data presented here indicates that S. pombe is a good model system to study the in vivo function of Exo5, and that these studies have the potential to give insight into the function of the human enzyme. The similarity between the sliding activity of the human and S. pombe enzyme indicates that this activity may be conserved and may have functional implications that are still unclear. The identification of both the single-strand DNA binding protein complex RPA and processivity clamp PCNA could support a model where Exo5 is involved at the replication fork, but more work will be needed to test this model. The RPA interaction appears to be
conserved from humans to *S. pombe* (Sparks, Kumar et al. 2012). There are two motifs that are important for this interaction. First is an acidic patch at the extreme N-terminus of the protein and the second is a hydrophobic patch in the vicinity of the first cysteine of the iron-sulfur complex. It still needs to be determined if this interaction is required for its function in DNA metabolism. The PCNA interacting motif is unclear in Exo5, with the protein not containing a canonical PCNA interacting motif (PIP). It will be important to investigate the physiological role these interactions play in the *in vivo* function of Exo5.

*S. pombe* has provided a good genetic tool for investigating the DNA repair pathways that Exo5 may play a role. The DNA damage sensitivity of an *exo5Δ* strain closely recapitulates the hExo5 knockdown experiments in human cells. The *exo5Δ* strains were particularly sensitive to DNA damaging agents that require replication-coupled repair including interstrand cross-linking agents. The epistasis analysis is giving us the first glimpse of the pathways in which Exo5 is involved. The *exo5Δrad2Δ* mutants show synergistic interactions, which is consistent with these proteins working in different pathways that target the same lesions. Rad2 (FEN1) is required for Okazaki fragment maturation and long patch base-excision repair (Liu, Kao et al. 2004). It is possible that, similarly to the exonuclease/helicase Dna2, Exo5 plays a redundant function in Okazaki fragment maturation when ssDNA flaps become bound by RPA and Rad2 is no longer able to act (Murante, Rust et al. 1995; Bae, Bae et al. 2001; Ayyagari, Gomes et al. 2003; Kao, Veeraraghavan et al. 2004). Much is known about the nuclear function of Rad2, but very little is known about mitochondrial function of this enzyme. Exo5 and Rad2 play an essential redundant function in the mitochondria in *S. pombe*. The *exo5Δrad2Δ* strain has a ten-fold depletion of the mitochondrial DNA that leads to its inability to grow on plates with glycerol as the sole carbon source. Exogenous expression of either of these proteins restores growth on glycerol leading to a model where these proteins may play a redundant role in mitochondrial DNA replication or repair. One issue is the lack of understanding of the mechanism of mitochondrial replication in *S. pombe*. Both enzymes are single-strand specific exonucleases that may be required to remove single-strand DNA flaps during mitochondrial DNA replication. Another interesting aspect of the mitochondrial function of spExo5 is the apparent use of an alternative translation start at methionine 58, and the human Exo5 protein has a methionine that is similar to the *S. pombe* methionine 58. It is a possibility that the human enzyme has this conserved alternative translation start site that makes a shorter variant of the enzyme that is specific for localization to the mitochondria. It is also interesting that the short isoform still maintains all the cysteines required for forming the iron-sulfur cluster. Further work will be needed to determine the function of these enzymes in the mitochondria.
Exo5 deficient strains are sensitive to interstrand crosslinking agents such as cis-platin and 8-methyloxypsoralen. The repair of ICLs has only recently been investigated in *S. pombe*. These studies have begun to delineate the pathways for repair of ICLs, but there are still many questions that need to be investigated. There are at least three pathways that have been identified in *S. pombe* for ICL repair. One of these pathways involves the nuclease Fan1 and the helicase Fml1, which are Fanconi anemia homologs; the second pathway is dependent on the nuclease Pso2 and the nucleotide excision repair endonuclease Rad13 (XPG); the third pathway requires the sumo ligase Pli1 (Fontebasso, Etheridge et al. 2013). Exo5 is epistatic with the proteins of the Fan1 dependent pathway of repair and is non-epistatic with the Pso2 nuclease. Although Exo5 is epistatic with the Rad13 and Swi10-Rad16 endonucleases that are a part of the Pso2 dependent pathway, it is still a possibility that Exo5 is involved in the Pli1 dependent pathway. The Pli1 pathway of repair has not been studied to date. There are also many other DNA repair proteins that have been implicated in ICL repair that have not been tested for their role in these pathways. We demonstrate that Exo1 is the multifunctional exonuclease (Tran, Erdeniz et al. 2004; Barber, Ward et al. 2005; Mimitou and Symington 2009) and Slx1, an endonuclease that forms a complex with Slx4, are epistatic with the Fan1 pathway and non-epistatic with the Pso2 pathway and indicating these proteins function with Fan1 in the repair of ICLs. On the other hand, Swi10-Rad16 (XPF), an endonuclease of the NER pathway that has been shown in *Xenopus laeveis* egg extracts to be essential for the unhooking step of ICL repair (Klein Douwel, Boonen et al. 2014), and the RecQ helicase Hrq1, involved in nucleotide excision and ICL repair (Groocock, Prudden et al. 2012), are epistatic with the Pso2 pathway and non-epistatic with the Fan1 pathway. These genetic interactions expand the proteins that are involved in these two pathways of ICL repair (Fig 6). Further work will be needed to establish the pathway of ICL repair that involves Exo5.

ACKNOWLEDGEMENTS

The authors thank Nick Rhind, Susan Forsburg, for providing plasmids. The authors would also like to thank Nick Rhind, Susan Forsburg, Anthony Carr, Hiroshi Iwasaki, Lynda Groocock, Albert Pastink, Nancy Walworth, Daniel Teasley, Emily Higuchi, Michael Boddy, Randy Hyppa, Paul Russell, Toru Nakamura for providing *S. pombe* strains. The authors also thank Hiroshi Iwasaki for providing an antibody used in this manuscript.

EXPERIMENTAL PROCEDURES
**DNA damaging agents.** Drugs used were all from SIGMA: cis-platin (diamminecis-platinum(II) dichloride), cat. no. P4394; methyl methanesulfonate (MMS), cat. no. 129925; 8-MOP (8-methoxypsoralen), cat. no. M3501.

**S. pombe strains.** The exo5Δ strain was obtained from Bioneer Corporation. Double disruption mutants were obtained by crossing the exo5Δ strain with strains containing other repair gene disruptions. Genotypes of the double deletion strains were confirmed by diagnostic PCR.

**Plasmids and oligonucleotides.** Plasmid pBL281 (*S. pombe*), pBL277 (human), and pBL282 (mouse) contains the Schistosoma japonicum glutathione S-transferase (GST) gene fused to the N terminus of the Exo5 gene in vector pRS424-GALGST (Walker, Crowley et al. 1993). The GST tag is separated from the N terminus of the *S. pombe*, human, and mouse Exo5 gene by a recognition sequence for the human rhinoviral 3C protease (LEVLFQ/GP). Following cleavage by the protease, the N-terminal sequence of *S. pombe* Exo5 is extended with the GPEF sequence. Plasmid pBL281-207 has an active site mutation D207. Plasmid pBL281-176 has an active site mutation D176. *S. pombe* expression plasmids pREP3x-spExo5 with spExo5 under the control of the strong nmt promoter. Plasmids and sequences are available upon request.

Oligonucleotides were purchased from IDT (Coralville, IA) and purified by urea-polyacrylamide gel electrophoresis (PAGE): c81, TTGCCGATGAAGTTTTTTTTTTTGATCGAGACCTT; v81 AAGGTCTCCATCAAAAAAAAAAGTTTCATCGGCAA. The 5’-32P-label was introduced on oligonucleotides c81 or 5’-5’ polarity switch oligonucleotide using [γ-32P] ATP and T4 Polynucleotide kinase. While the 3’-32P-label was added to 3’-3’ polarity switch oligonucleotide by incubation with [α-32P] dATP with terminal deoxynucleotidyl transferase under manufacturers’ conditions.

**Exo5 overproduction and purification.** Overproduction was carried out in *S. cerevisiae* strain FM113 (MATa ura-3-52 trp1-289 leu2-3112 prb1-1122 prc1-407 pep4-3) transformed with plasmid pBL277 (human Exo5), pBL281 (*S. pombe* Exo5-), pBL281-207 (*S. pombe* Exo5-D207A), pBL281-176 (*S. pombe* Exo5-D176), or pBL282 (mouse Exo5). Growth, induction, and extraction were similar to the procedures described previously (Bylund, Majka et al. 2006). Cells were harvested and resuspended in 1/2 the volume of 3x buffer A (buffer A: 60 mM HEPES-NaOH [pH 7.8], 0.4 M sodium acetate, 0.1 mM EDTA, 0.005% polyoxyethylene (10) lauryl ether, 50 mM sodium bisulfite, 50 µM pepstatin A, 50 µM leupeptin), then frozen in liquid nitrogen. The frozen cell pellets were then blended in dry ice powder. All further preparation was carried out at 0-4 °C. After the thawing of the lysate, 10% glycerol, 1 mM dithiothreitol (DTT), 0.05 mM phenylmethylsulfonyl fluoride (100 mM stock), and 150 mM ammonium sulfate (4 M stock), 0.45 % polymin P (10% stock, pH 7.3) were added to the lysate. The mixture was stirred
for 15 minutes, the lysate was cleared at 40,000 × g for 30 minutes, and the supernatant was precipitated with 0.31 g/ml solid ammonium sulfate. The precipitate was collected at 40,000 × g for 30 minutes and then redissolved in buffer A₀ (subscript indicates the sodium acetate concentration) 10% glycerol, and 1 mM dithiothreitol until the lysate conductivity was equal to that of buffer A₄₀₀. The lysate was then used for batch binding to 1 ml of glutathione-Sepharose 4B beads (GE Healthcare), equilibrated with buffer A₄₀₀, and gently rotated at 4 °C for two hours. GST-beads were collected at 1,000 rpm in a swinging-bucket rotor, followed by batch washes (3 × 20 ml of buffer A₄₀₀). The beads were transferred to a 10 ml column, and washed at 2.5 ml/min with 100 ml of buffer A₄₀₀. The second washing was with 50 ml buffer A₄₀₀ containing 5 mM Mg-acetate and 1 mM ATP. And the third washing used 50 ml of buffer A₄₀₀ and 30 ml of buffer A₂₀₀. Elution was carried out with a flow rate of 0.2 ml/min with buffer A₂₀₀ containing 20 mM glutathione (pH adjusted to 8.0). The protein was eluted into four fractions. Fractions containing Exo5 were combined and diluted with buffer A₀ to equal buffer A₁₀₀, then loaded onto a 1-ml MonoQ column. Protein was eluted with a linear gradient of buffer A₁₀₀ to A₁₂₀₀. The fractions containing pure protein were incubated over night at 4 °C with 30U of rhinoviral 3C protease, and diluted with A₀ to equal A₁₀₀, loaded on a 1-ml Mono Q column. Pure Exo5 protein was eluted at 300mM to 400mM sodium acetate. Exo5 mutants were purified similarly throughout the procedures described.

**SpExo5 Co-immunoprecipitation**

Co-IPs were carried out in *S. pombe* strain PYP102 (exo5Δ) by transformation of pREP3x-spExo5 or mutant. Cells were grown in Edinburgh minimal media supplemented with 5mM thiamine overnight. Cells were then harvested and resuspended in 1/2 the volume of 3x buffer A (buffer A: 60 mM HEPES-NaOH [pH 7.8], 150 mM sodium chloride, 0.1 mM EDTA, 0.05% polyoxyethylene (10) lauryl ether, 10 mM sodium bisulfite, 10 µM pepstatin A, 10 µM leupeptin), then frozen in liquid nitrogen. The frozen cell pellets were then blended in dry ice powder. All further preparation was carried out at 0-4 °C. After the thawing of the lysate, 10% glycerol, 1 mM dithiothreitol (DTT), and 0.5 mM phenylmethylsulfonyl fluoride (100 mM stock) were added to the lysates. The lysates were cleared at 40,000 × g for 30 minutes, and the cleared lysate was then used for batch binding to FLAG-M2-beads (SIGMA), equilibrated with buffer A₁₅₀, and gently rotated at 4 °C for two hours. FLAG-M2-beads were collected at 1,000 rpm in a swinging-bucket rotor, followed by batch washes (3 × 500 µl of buffer A₁₅₀). The second washing was with 500 µl buffer A₁₅₀ containing 5 mM Mg-acetate and 1 mM ATP. And the third washing used (3 × 500 µl of buffer A₁₅₀). Elution was carried out in buffer A₁₅₀ containing 150 µg/ml
3XFLAG peptide (SIGMA). The protein was eluted into two fractions and analyzed by SDS-PAGE followed by comassie Brilliant Blue staining.

**Exonuclease assays.** The standard 10 µl assay mixture contains 100 mM Tris-HCl (pH 7.8), 500 µg/ml bovine serum albumin, 5 mM DTT, 5 mM Mg-acetate, 50 mM NaCl, 50-100 fmol of ³²P-end-labeled oligonucleotide substrate, and enzyme. Incubations were carried out at 30 °C for the indicated time periods. Deviations for the standard assay conditions are indicated in the legends of the figures. Reactions were stopped with 10 mM final concentration of EDTA in addition to 40% formamide and analyzed on a 17% PAGE-7 M urea electrophoresis. After the gels were dried they were subjected to phosphorImager analysis.

**S. pombe damage sensitivity assays.** Strains were inoculated in 5 ml of liquid yeast extract supplemented (YES) medium and grown overnight at 30 °C with shaking. The cells were then washed in phosphate-buffered saline and the cell density was determined by spectroscopy at 595nm. Serial dilutions were carried out and plated on YES plates supplemented with various genotoxic agents, or irradiated with UV or X-rays. The plates were then incubated for 2-3 days at 30 °C.


Figure 1. SpExo5 is a bi-directional ssDNA specific exonuclease that slides along ssDNA prior to cleavage. (A) Comparison of activity of human, mouse, and S. pombe Exo5 on 34-nucleotide oligonucleotide. Standard assay mixtures used 10 nM 5'-32P-labeled cexo81 a 34-nucleotide ssDNA substrate. The concentration of human Exo5 from left to right are 1.5, 5, 15, 50, 150 nM; mouse Exo5 from left to right are 1.5, 5, 15, 50, 150 nM; S. pombe Exo5 from left to right are 0.15, 0.5, 1.5, 5, 15 nM. The reactions were carried out at 30 °C for 4 min. The results were analyzed on a 7 M urea-17% polyacrylamide gel. (B) Comparison of activity of S. pombe and S. cerevisiae Exo5 on polarity switch poly-dT oligonucleotide. Standard assays conditions containing 10nM concentration of 5’-labeled (dT)_{65} or double 5’-ended (dT)_{70}, or 3’-labeled double 3’-ended (dT)_{70} were incubated with 0.5, 0.15, 5, or 15 nM spExo5 for 30 s or with 0.05, 0.15, 5, or 15 nM S. cerevisiae (S. cer) Exo5 for 30 s at 30 °C. Controls contained DNA polymerase Polδ. All samples were analyzed on 7 M urea-18% polyacrylamide gels and quantified by phosphorimaging. The asterisks indicate the position of the 32P label. (C) Activity of S. pombe Exo5 catalytic dead mutants on 34-nucleotide oligonucleotide. Standard assay mixtures used 10 nM 5'-32P-labeled cexo81 a 34-nucleotide ssDNA substrate. The concentration S. pombe Exo5 from left to right are 0.15, 0.5, 1.5, 5, 15 nM. The reactions were carried out at 30 °C for 4 min. The results were analyzed on a 7 M urea-17% polyacrylamide gel. (D) Dependence of hExo5 and spExo5 molar activity on DNA concentration. Data were fit to a Michaelis-Menten model. For spEXO5, $K_m = 26 \pm 2.2$ nM and $V_{max} = 60 \pm 1.5$ min$^{-1}$; for hEXO5, $K_m = 3,200 \pm 800$ nM and $V_{max} = 0.7 \pm 0.1$ min$^{-1}$.

Figure 2. SpExo5 physically interacts with spRPA and spPCNA in vivo. (A) Co-IP of spExo5-3XFLAG from S. pombe. Co-IP elution were analyzed on a 10%-SDS PAGE gel and stained with comassie brilliant blue. The red boxes indicate sections cut out of gel and sent for mass spectrometry footprinting. (B) Western analysis of co-IP samples either treated or mock treated with DNasel. Antibodies used against 3XFLAG (spExo5) (SIGMA), spSSB1 (Hiroshi Iwasaki lab), and spPcn1 (Santa Cruz). Elutions were analyzed on a 10%-SDS PAGE gel and transferred to a PVDF membrane for western analysis. (C) DNase I control. Plasmid DNA was either incubated with mock treated or DNase I treated co-IP and sample was analyzed on a 1% agarose gel stained with ethidium bromide. (D) spExo5 mutant expression comparison by western analysis. spExo5-3XFLAG mutants were IP using FLAG-M2-beads (SIGMA) and elutions were analyzed by western analysis.
Figure 3. Overexpression of SpExo5 causes checkpoint activation and cell death mediated through interaction with spRPA and spPCNA. (A) Overexpression of spExo5 leads to cell death in S. pombe cells regardless of checkpoint proficiency. Cells either WT or rad3Δ containing pREP3x-spExo5 or pREP3x-spExo5-D207A were grown overnight in EMM-leu supplemented with 5µM thiamine. Cells were washed serially diluted on plates either containing 15 µM or 0 µM thiamine and incubated 2-3 day at 30 °C. (B) Light microscopy of cells from (A) either grown on plates containing 15 µM or 0 µM thiamine for 2-3 days. (C) SpExo5 amino acid sequence illustrating point mutations used in (D). Point mutations are indicated by arrows and an A indicating the amino acid was change to an alanine. (D) Mutations that suppress the checkpoint activation and cell death of overexpression of spExo5. exo5Δ cells containing pREP3x-exo5Δ or exo5 mutants were grown overnight in EMM-leu supplemented with 5µM thiamine. Cells were washed serially diluted on plates either containing 15 µM or 0 µM thiamine and incubated 2-3 day at 30 °C. (E) Suppressor mutations eliminate spExo5 interactions with RPA and PCNA. Co-IP elutions from expression of spExo5-3XFLAG or mutants was run on 10% SDS-PAGE gels and transferred to PVDF membrane for western analysis against 3XFLAG, ssb1, or Pcn1.

Figure 4. Exo5 genetic interactions. (A) Sensitivity of exo5Δ or pso2 strains to acute psoralen treatment. Cells were grown to an OD of 0.4 and then either mock treated (-) with water or with 5 µg/ml 8-MOP (+) and cells incubated for 15 minutes at 30 °C. Following incubation with 8-MOP, cells were irradiated with 365 nm light to activate the 8-MOP for various time point. The cells were then washed twice with PBS buffer and diluted for spotting on rich media (YES) plates. (B) Epistasis experiment using exo5Δ, exolΔ, and exo5Δexo1Δ S. pombe strains as indicated, testing sensitivity to UV-irradiation, MMS and cis-platin. Cells were spotted in ten-fold dilutions (C) Epistasis experiment using exo5Δ, rad2Δ, and exo5Δrad2Δ S. pombe strains as indicated, testing sensitivity MMS and cis-platin. Cells were spotted in ten-fold dilutions (D) Epistasis experiment investigating interaction of exo5Δ and pso2Δ, testing the sensitivity of the strains to cis-platin. (E) Epistasis experiment investigating interaction of exo5Δ, exolΔ, and hrq1Δ, testing the sensitivity of the strains to cis-platin. (F) Graphic representation of the genetic interactions revealed in this study between spExo5 and other known DNA repair proteins. The arrows indicate epistatic interactions, arrows with + indicate additive interactions, and arrows with side arrow indicate synergistic interactions.

Figure 5. A short isoform of Exo5 and the Rad2 play an essential redundant function in mitochondrial genome maintenance. (A) exo5Δ strain transformed either with vector or
pREP3x-spExo5 streaked on YEG plates and grown for 7 days at 30 °C (B) Epistasis experiment using exo5Δ, rad2Δ, and exo5Δrad2Δ S. pombe strains as indicated, testing growth on YES plates at various temperatures. Cells were spotted in ten-fold dilutions (C) Epistasis experiment using exo5Δ, rad2Δ, and exo5Δrad2Δ S. pombe strains transformed with either pREP3x, pREP3x-spExo5, pREP3x-spexo5-D207A, pREP3x-spexo5-M58A, pREP3x-spexo5-58-409 as indicated, testing sensitivity to cis-platin and YEG. Cells were spotted in ten-fold dilutions (D) Genomic DNA digested with xhoI, nheI, mlul, and ageI; followed by analysis on a 0.6% agarose gel and ethidium bromide staining.

**Figure 6. Model for ICL repair and possible role of spExo5 and overview of spExo5 genetic interactions.** Described in the text.

**SUPPLEMENTAL INFORMATION**

**Supplemental Table 1. Mass spectrometry fingerprint of Exo5 Co-IP.** Results acquired from co-IP of interacting proteins of spExo5-3XFLAG in Figure 2.

**Supplemental Table 2. Genetic interactions matrix revealed by epistasis analysis.** Data collected from epistasis experiments are outlined in this table. (++) Indicates synergistic interaction, (L) Lethal interaction, (+) indicates additive interactions, (-) indicates epistatic interactions, and (n.t.) indicates that the interaction was not tested.

**Supp. Figure 1. Epistasis analysis.** (A) Epistasis experiment using exo5Δ, mus81Δ, and exo5Δmus81Δ S. pombe strains as indicated, testing sensitivity to cis-platin. Cells were spotted in ten-fold dilutions. (B) Epistasis experiment using exo5Δ, exo1Δ, and exo5Δexo1Δ S. pombe strains as indicated, testing sensitivity to ionizing radiation. Cells were spotted in ten-fold dilutions. (C) Epistasis experiment using exo5Δ, rhp51Δ, and exo5Δrhp51Δ S. pombe strains as indicated, testing sensitivity to MMS. Cells were spotted in ten-fold dilutions. (D) Epistasis experiment using exo5Δ, mre11Δ, and exo5Δmre11Δ S. pombe strains as indicated, testing sensitivity to ionizing radiation. Cells were spotted in ten-fold dilutions. (E) Epistasis experiment using exo5Δ, rad13Δ, and exo5Δrad13Δ S. pombe strains as indicated, testing sensitivity to UV-irradiation and cis-platin. Cells were spotted in ten-fold dilutions. (F) Epistasis experiment using exo5Δ, uve1Δ, and exo5Δuve1Δ S. pombe strains as indicated, testing sensitivity to UV-irradiation. Cells were spotted in ten-fold dilutions.
**Supp. Figure 2. SpExo5 is Epistatic with Fan1 pathway and non-epistatic with Pso2.** (A) Epistasis experiment using \textit{exo5Δ, fan1Δ, pso2Δ, exo1Δ}, and the pair wise double and triple mutants and the quadruple mutant \textit{S. pombe} strains as indicated, testing sensitivity to cis-platin. Cells were spotted in ten-fold dilutions. (B) Epistasis experiment using \textit{exo5Δ, exo1Δ, fml1Δ, pso2Δ}, and all the pair wise double and triple mutant \textit{S. pombe} strains as indicated, testing sensitivity to cis-platin. Cells were spotted in ten-fold dilutions.

**Supp. Figure 3. Genetic interactions with Exo5, Hrq1, Slx1, and Rqh1.** (A) Epistasis experiment using \textit{exo5Δ, hrq1Δ, pso2Δ, exo1Δ, fan1Δ, fml1} and the pair wise double mutant \textit{S. pombe} strains as indicated, testing sensitivity to cis-platin. Cells were spotted in ten-fold dilutions. (B) Epistasis experiment using \textit{exo5Δ, swi10Δ, exo1Δ, fml1Δ, pso2Δ, fan1Δ} and the pair wise double and triple mutant \textit{S. pombe} strains as indicated, testing sensitivity to cis-platin. Cells were spotted in ten-fold dilutions. (C) Epistasis experiment using \textit{exo5Δ, rqh1Δ, fml1Δ}, and the pair wise double mutant \textit{S. pombe} strains as indicated, testing sensitivity to cis-platin. Cells were spotted in ten-fold dilutions.

**Supp. Figure 4. Genetic interactions with Exo5, Fml1, and Slx1.** (A) Epistasis experiment using \textit{exo5Δ, slx1Δ, fml1} and the pair wise double mutant \textit{S. pombe} strains as indicated, testing sensitivity to cis-platin. Cells were spotted in ten-fold dilutions.
Figure 1

Figure 2
Figure 3

A

<table>
<thead>
<tr>
<th></th>
<th>Repression</th>
<th>Overexpression</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>WT</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vector</td>
<td></td>
</tr>
<tr>
<td></td>
<td>WT</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D207A</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>rad3Δ</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vector</td>
<td></td>
</tr>
<tr>
<td></td>
<td>WT</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D207A</td>
<td></td>
</tr>
</tbody>
</table>

B

**WT**

**rad3Δ**

Vector

exo5+

exo5-D207A

C

Spom MEEEDFE---------MEDLSEL 59 YLpVTDLy1plWCEV 321
Hsap MAETREETSASAHFSDLSD- 58 YLpVTDLy1QNWCEL 276

D

<table>
<thead>
<tr>
<th>spExo5</th>
<th>Repression</th>
<th>Overexpression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vector</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D207A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E2A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E3A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E5A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T11A</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

E

**IP: 3XFLAG**

<table>
<thead>
<tr>
<th>vector</th>
<th>WT</th>
<th>D207A</th>
<th>Y74A</th>
<th>D79A</th>
</tr>
</thead>
</table>

- α ssb1
- α 3XFLAG
- α pcn1

73
Figure 5

A

vector

exo5^{

YE-Glycerol

B

\*C

24

30

34

WT

WT

exo5Δ

exo5Δ

rad2Δ

rad2Δ

exo5Δ rad2Δ

exo5Δ rad2Δ

C

YES

0.5 mM cisplatin

YEG

WT

exo5Δ

rad2Δ

vector

exo5^{

d207A

M58A

58-409

exo5Δ rad2Δ

D

% WT mtDNA

100

69

100

15

Figure 6

Pso2

Rad13

Swi10

Rhp51

Hrq1

Fan1

Exo1

Sxl1

Exo5

Pli1

mtD

rDN

75
### Supplementary Table 1

<table>
<thead>
<tr>
<th>Score</th>
<th>Protein Name</th>
<th>MW</th>
<th>% Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Upper Band</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>616</td>
<td>Exo5</td>
<td>48365</td>
<td>53.3</td>
</tr>
<tr>
<td>412</td>
<td>DNA replication factor A subunit Ssb1</td>
<td>68154</td>
<td>41.1</td>
</tr>
<tr>
<td>123</td>
<td>AAA family ATPase Rvb1</td>
<td>50021</td>
<td>12.1</td>
</tr>
<tr>
<td>102</td>
<td>AAA family ATPase Rvb2</td>
<td>51529</td>
<td>22.4</td>
</tr>
<tr>
<td>86</td>
<td>AAA family ATPase Cdc48</td>
<td>90069</td>
<td>8.8</td>
</tr>
<tr>
<td>42</td>
<td>single-stranded telomeric binding protein Tgc1</td>
<td>37980</td>
<td>16.9</td>
</tr>
<tr>
<td>25</td>
<td>TFIH regulator Mms19</td>
<td>114597</td>
<td>0.9</td>
</tr>
<tr>
<td>21</td>
<td>crm1-N1 protein</td>
<td>123650</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td><strong>Lower Band</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>Exo5</td>
<td>48365</td>
<td>54.3</td>
</tr>
<tr>
<td>266</td>
<td>14-3-3 protein Rad24</td>
<td>30357</td>
<td>29.5</td>
</tr>
<tr>
<td>190</td>
<td>14-3-3 protein Rad25</td>
<td>30350</td>
<td>24.4</td>
</tr>
<tr>
<td>174</td>
<td>DNA polymerase delta processivity factor</td>
<td>28951</td>
<td>42.7</td>
</tr>
<tr>
<td>160</td>
<td>single-stranded DNA binding protein Ssb2</td>
<td>30372</td>
<td>45.5</td>
</tr>
<tr>
<td>71</td>
<td>Rim1</td>
<td>17129</td>
<td>25.3</td>
</tr>
</tbody>
</table>

### Supplementary Table 2

<table>
<thead>
<tr>
<th></th>
<th>Exo5</th>
<th>Fan1</th>
<th>Rad2</th>
<th>Exo1</th>
<th>Fml1</th>
<th>Swi10</th>
<th>Rad13</th>
<th>Rhp51</th>
<th>Pso2</th>
<th>Slx1</th>
<th>Hrq1</th>
<th>Rqh1</th>
<th>Mus81</th>
<th>Mre11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exo5</td>
<td>-</td>
<td>++</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Fan1</td>
<td>-</td>
<td>n.t.</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>n.t.</td>
<td>++</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
</tr>
<tr>
<td>Rad2</td>
<td>++</td>
<td>n.t.</td>
<td>L</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>L</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
</tr>
<tr>
<td>Exo1</td>
<td>+</td>
<td>-</td>
<td>L</td>
<td>-</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>++</td>
<td>n.t.</td>
<td>-</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
</tr>
<tr>
<td>Fml1</td>
<td>-</td>
<td>-</td>
<td>n.t.</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>n.t.</td>
<td>++</td>
<td>-</td>
<td>++</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
</tr>
<tr>
<td>Swi10</td>
<td>-</td>
<td>++</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>+</td>
<td>-</td>
<td>n.t.</td>
<td>-</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
</tr>
<tr>
<td>Rad13</td>
<td>-</td>
<td>++</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>-</td>
<td>n.t.</td>
<td>-</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
</tr>
<tr>
<td>Rhp51</td>
<td>-</td>
<td>-</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>-</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
</tr>
<tr>
<td>Pso2</td>
<td>+</td>
<td>++</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
</tr>
<tr>
<td>Slx1</td>
<td>-</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>-</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
</tr>
<tr>
<td>Hrq1</td>
<td>+</td>
<td>n.t.</td>
<td>n.t.</td>
<td>-</td>
<td>++</td>
<td>-</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
</tr>
<tr>
<td>Rqh1</td>
<td>-</td>
<td>n.t.</td>
<td>L</td>
<td>n.t.</td>
<td>++</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
</tr>
<tr>
<td>Mus81</td>
<td>-</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
</tr>
<tr>
<td>Mre11</td>
<td>-</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
</tr>
</tbody>
</table>

76
Supplementary Figure 3
CHAPTER III

Reconstitution of RNase H2-initiated Ribonucleotide Excision Repair (RER)
This chapter of the dissertation discusses the phenomenon of ribonucleotide incorporation into the genome by replicative polymerases, and how a pathway initiated by the heterotrimeric enzyme RNase H2 repairs these non-canonical bases. Section IIIA relates the study of ribonucleotide incorporation and repair by biochemical reconstitution of the system in vitro using a 7.3 kb primed single-strand plasmid. The rate of incorporation by either Polδ or Polε is approximately two-rNMPs per kilobase replicated. The RNase H2 dependent pathway of repair of genomic rNMPs requires RNase H2 to cleave at the ribonucleotide, and RNase H1 is unable to complement for RNase H2. It also requires the replication machinery RFC, PCNA, and either Polδ or Polε. Additionally, it requires either FEN1 or Exo1 to perform strand displacement synthesis along with the polymerases, and finally DNA ligase to close the nick. Peter Burgers and I performed all the biochemical assays for this work. Other labs provided purified protein, Robert Crouch’s lab provided RNase H2 and RNase H1 and Eric Johansson provided purified Polε for this study.

Section IIIB of this chapter describes how the different functions of RNase H2 contribute to the phenotype of an RNase H2 deletion in yeast. Making mutants of RNase H2 that are either only able to process single-rNMPs or sites of multiple-rNMPs (R-loops) allow for correlation of the two functions with different phenotypes of RNase H2 deletion in S. cerevisiae. RNase H2’s ability to process single-rNMPs from the genome is responsible for preventing small deletions at di-nucleotide repeats in the genome; RNaseH2’s ability to process sites of multiple-rNMPs is important in the context of an sgs1Δ in yeast. I contributed one of the biochemical assays and analysis with Peter Burgers’ assistance for the attached paper.
CHAPTER IIIA

RNase H2-initiated Ribonucleotide Excision Repair (RER)
RNase H2-Initiated Ribonucleotide Excision Repair

Justin L. Sparks,1 Hyongi Chon,2 Susana M. Cerritelli,2 Thomas A. Kunkel,3 Erik Johansson,4 Robert J. Crouch,2 and Peter M. Burgers1,*

1Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, MO 63110, USA
2Program in Genomics of Differentiation, Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, MD 20892, USA
3Laboratory of Molecular Genetics and Laboratory of Structural Biology, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709, USA
4Department of Medical Biochemistry and Biophysics, Umeå University, SE-901 87, Umeå, Sweden

*Correspondence: burgers@biochem.wustl.edu
http://dx.doi.org/10.1016/j.molcel.2012.06.035

SUMMARY

Ribonucleotides are incorporated into DNA by the replicative DNA polymerases at frequencies of about 2 per kb, which makes them by far the most abundant form of potential DNA damage in the cell. Their removal is essential for restoring a stable intact chromosome. Here, we present a complete biochemical reconstitution of the ribonucleotide excision repair (RER) pathway with enzymes purified from Saccharomyces cerevisiae. RER is most efficient when the ribonucleotide is incised by RNase H2, and further excised by the flap endonuclease FEN1 with strand displacement synthesis carried out by DNA polymerase δ, the PCNA clamp, its loader RFC, and completed by DNA ligase I. We observed partial redundancy for several of the enzymes in this pathway. Exo1 substitutes for FEN1 and Pol ε for Pol δ with reasonable efficiency. However, RNase H1 fails to substitute for RNase H2 in the incision step of RER.

INTRODUCTION

DNA polymerases show an extraordinary ability to distinguish deoxyribonucleotide from ribonucleotide precursors for DNA replication (reviewed in Joyce [1997]). Replicative DNA polymerases typically show a specificity ratio of \( \sim 10^{-4} - 10^{-5} \) (rNTPs incorporated per dNTPs incorporated). However, because the cellular concentrations of rNTPs in yeast are estimated to be 10- to 100-fold higher than those of dNTPs (Nick McElhinny et al., 2010b), the actual misincorporation of ribonucleotides into genomic DNA is likely to be much higher as well and could reach ratios on the order of \( 10^{-3} \). Compared to a deoxyribonucleotide, a ribonucleotide has a reactive 2'-hydroxyl on the sugar moiety that renders the DNA backbone more susceptible to strand cleavage. This could potentially reduce genome stability and affect its replication during a subsequent cell cycle (Nick McElhinny et al., 2010a; Watt et al., 2011). Therefore, repair mechanisms exist to deal with ribo-damage. The one initiated by ribonuclease H2 (RNase H2), called ribonucleotide excision repair (RER), is considered to be the most common repair pathway.

Ribonuclease H enzymes cut RNA/DNA hybrids. RNase H1 is most active on RNA/DNA hybrids containing several consecutive ribonucleotides, and no incision activity has been observed on a DNA substrate with a single embedded ribonucleotide (Cerritelli and Crouch, 2009). Consecutive RNA residues are not expected to result from the occasional, stochastic insertion of ribonucleotides by DNA polymerases, or at least these occurrences should be extremely rare. On the other hand, RNase H2 shows prominent activity on a single ribonucleotide embedded within double-stranded DNA (Eder and Walder, 1991; Rychlik et al., 2010). RNase H2 incises the DNA 5'-end of the ribonucleotide, generating DNA containing 3'-hydroxyl and 5'-phospho-ribonucleotide ends (see Figure 4). This product is expected to be an ideal substrate for the Okazaki fragment maturation system that normally deals with 5'-RNA-terminated DNA fragments (Burgers, 2009), however, whether repair actually proceeds by this or a similar pathway remains to be determined. Ribonucleotide-containing DNA can also be incised by Topoisomerase I (Kim et al., 2011; Sekiguchi and Shuman, 1997), whose reaction mechanism proceeds through a 3'-phospho-tyrosyl intermediate. If this covalent intermediate is localized at the ribonucleotide, it can be resolved aberrantly through nucleophilic attack by the vicinal 2'-OH group resulting in the generation of a 2'-3'-cyclic phosphate with concomitant release of Topo I generating a single-strand break. Currently, it is not clear whether or how these cyclic phosphate intermediates are processed. However, based on genetic analysis in yeast, Topo I-mediated cleavage is likely a minor pathway for ribo-repair (Kim et al., 2011).

Mutations in RNase H2 are associated with Aicardi-Goutières syndrome, an autosomal recessive disorder. Patients have progressive microencephaly and show complex neurological defects in motor function and altered interferon response (Crow et al., 2006). However, defects in other enzymes, such as the 3'-exonuclease TREX1 (Stetson et al., 2008), and SAMHD1, a dNTP triphosphatase (Powell et al., 2011), also cause Aicardi-Goutières syndrome, making a direct connection between the syndrome and ribonucleotide excision repair less straightforward. A homozygous null mouse for RNase H2 shows embryonic lethality at the gastrulation stage of development (Reijns et al., 2012). On the other hand, yeast RNase H2 null mutants are viable. Nonetheless, growth defects and even lethality result
when a m<sup>r</sup>201 mutation is combined with other defects in DNA metabolism (Budd et al., 2005; Lazzaro et al., 2012).

Ribonucleotide excision repair (RER) initiated by RNase H2 has so far not been studied in a reconstituted, purified system. Earlier studies in crude yeast extracts already implicated RNase H2 and the flap endonuclease FEN1 in RER. Reduced incision at a ribonucleotide in DNA was observed in extracts from strains with a deletion of RNH201, encoding the catalytic subunit of RNase H2, or a deletion of RAD27, encoding FEN1 nuclease that is involved in Okazaki fragment maturation (Rydberg and Game, 2002). However, these studies did not identify additional components of the repair pathway. Here, we present a complete biochemical analysis of the RER pathway with recombinant proteins. We used, as substrate for RER, DNA replicated by the replicative DNA polymerases in reactions containing physiological concentrations of dNTPs and rNTPs, rather than using preformed DNA substrates with embedded ribonucleotides. In this fashion, context effects due to the use of a specific oligonucleotide substrate are eliminated, and events that are most likely to occur during DNA replication are substrates for RER. Our data show (1) that ribonucleotide misincorporation during leading and lagging strand DNA replication in vitro is comparable; (2) that RNase H2-initiated RER can be efficiently reconstituted with replicative DNA polymerases at these physiologically relevant rNTP/dNTP ratios using defined oligonucleotide template primer as substrates, we found that sequence contexts greatly affected rNMP incorporation probabilities at individual nucleotide positions (Nick McElhinny et al., 2010a). Therefore, in this study, to examine RER of true ribonucleotide replication errors, we first determined the frequency of ribonucleotide incorporation during the replication of a 7.3 kb long single-stranded (ss) viral DNA template (ssM13mp18) under physiological rNTP and dNTP concentrations. Under these conditions, many sequence contexts should be sampled, including those most likely to be made during replication, and a mean distribution of ribonucleotide incorporation can be obtained. We carried out this analysis with both the leading strand polymerase, Pol ε, and the lagging strand polymerase, Pol δ, under conditions of processive DNA replication by interaction of the polymerase with the replication clamp PCNA (proliferating cell nuclear antigen). The replicated DNA was treated with 0.3M NaOH at 55°C, which cleaves DNA at the ribonucleotide positions, and sized by electrophoresis on a denaturing agarose gel.

Replication of the viral ssM13mp18 DNA with or without rNTPs was compared. The ssDNA was coated with yeast RPA in all experiments. Replication assays without rNTPs actually contained 100 μM ATP, in order to effect efficient loading of PCNA, by its clamp loader replication factor C (RFC), which is an ATP-dependent process. Replication assays with rNTPs contained a total of 5.9 mM rNTPs, and additional magnesium was added so that the free magnesium concentration in the assay remained the same. Inclusion of rNTPs in the replication assay reduced the rate of processive replication by Pol δ or by Pol ε to about 40% (Figure S1A). However, complete replication of the entire 7.3 kb M13mp18 circle was accomplished after 6 min by Pol δ, and after 15 min by Pol ε (Figure S1B). Alkaline treatment of M13mp18 circles replicated by either Pol δ or by Pol ε without rNTPs showed no sensitivity to alkali as indicated by the presence of predominantly full-length products (Figure 1A). In sharp contrast, M13mp18 circles replicated in the presence of physiological rNTPs were sensitive to treatment with alkali, resulting in a broad distribution of products. In order to convert the radioactivity distribution into product length distribution, the radioactivity distribution was divided by the length distribution in kb. From this normalized distribution, the median length of alkali-stable DNA produced by Pol δ was determined at 0.61 kb and was not significantly different from that produced by Pol ε (0.54 kb, Figure 1B). The average frequency of ribonucleotide incorporation can be obtained by also taking into consideration that the maximum linear product length is 7.3 kb and, second, the

**RESULTS**

The cellular concentrations of each of the dNTP and rNTPs under normal growth conditions have been determined (Nick McElhinny et al., 2010b). The rNTP/dNTP ratios range from 35 for [CTP]/[dCTP] up to almost 200 for [ATP]/[dATP]. When we examined the probability of stable ribonucleotide incorporation by replicative DNA polymerases at these physiologically relevant rNTP/dNTP ratios using defined oligonucleotide template primer as substrates, we found that sequence contexts greatly affected rNMP incorporation probabilities at individual nucleotide positions (Nick McElhinny et al., 2010a). Therefore, in this study, to examine RER of true ribonucleotide replication errors, we first determined the frequency of ribonucleotide incorporation during the replication of a 7.3 kb long single-stranded (ss) viral DNA template (ssM13mp18) under physiological rNTP and dNTP concentrations. Under these conditions, many sequence contexts should be sampled, including those most likely to be made during replication, and a mean distribution of ribonucleotide incorporation can be obtained. We carried out this analysis with both the leading strand polymerase, Pol ε, and the lagging strand polymerase, Pol δ, under conditions of processive DNA replication by interaction of the polymerase with the replication clamp PCNA (proliferating cell nuclear antigen). The replicated DNA was treated with 0.3M NaOH at 55°C, which cleaves DNA at the ribonucleotide positions, and sized by electrophoresis on a denaturing agarose gel.

Replication of the viral ssM13mp18 DNA with or without rNTPs was compared. The ssDNA was coated with yeast RPA in all experiments. Replication assays without rNTPs actually contained 100 μM ATP, in order to effect efficient loading of PCNA, by its clamp loader replication factor C (RFC), which is an ATP-dependent process. Replication assays with rNTPs contained a total of 5.9 mM rNTPs, and additional magnesium was added so that the free magnesium concentration in the assay remained the same. Inclusion of rNTPs in the replication assay reduced the rate of processive replication by Pol δ or by Pol ε to about 40% (Figure S1A). However, complete replication of the entire 7.3 kb M13mp18 circle was accomplished after 6 min by Pol δ, and after 15 min by Pol ε (Figure S1B). Alkaline treatment of M13mp18 circles replicated by either Pol δ or by Pol ε without rNTPs showed no sensitivity to alkali as indicated by the presence of predominantly full-length products (Figure 1A). In sharp contrast, M13mp18 circles replicated in the presence of physiological rNTPs were sensitive to treatment with alkali, resulting in a broad distribution of products. In order to convert the radioactivity distribution into product length distribution, the radioactivity distribution was divided by the length distribution in kb. From this normalized distribution, the median length of alkali-stable DNA produced by Pol δ was determined at 0.61 kb and was not significantly different from that produced by Pol ε (0.54 kb, Figure 1B). The average frequency of ribonucleotide incorporation can be obtained by also taking into consideration that the maximum linear product length is 7.3 kb and, second, the
RER Requires RNase H2 and FEN1

A previous study showed defects in the incision of ribonucleotide-containing DNA in extracts from strains deleted for RNH201 and RAD27 (Ryberg and Game, 2002). Specifically, the incision step at the ribonucleotide position was severely inhibited in rnh201J extracts, while the subsequent liberation of this ribonucleotide was attenuated in rad27J extracts. However, actual repair of the DNA was not demonstrated in this study. To establish whether PCNA affected ribonucleotide misincorporation frequencies, we carried out replication by Pol δ without PCNA (Figure 1C). In order to get full-length synthesis of the M13mp18 template by Pol δ alone, we had to reduce the salt concentration from 100 to 20 mM NaCl and increase enzyme concentration and incubation time (Figure S1C). No significant differences in rNTP misincorporation were observed between Pol δ alone and the processive PCNA-Pol δ complex at 20 mM NaCl, when compared to the standard assay at 100 mM NaCl (Figure 1C).

Role of PCNA in RER

All four enzymes involved in RER show interactions with PCNA. PCNA mediates processive DNA synthesis by Pol δ and increases the nuclease efficiency of FEN1 (Li et al., 1995). On the other hand, DNA ligase I binds PCNA, but ligation efficiency is not significantly enhanced by PCNA (Vijayakumar et al., 2007). Rather, the PCNA-binding domain of DNA ligase aids in localizing the enzyme to replication foci (Montecucco et al., 1998). RNase H2 is a three-subunit enzyme. The largest subunit is the catalytic subunit, and the second subunit has a PCNA-binding domain.
domain with unknown function (Chon et al., 2009). The PCNA-binding domain of RNase H2 is essential for localizing it to replication foci in mammalian cells (Bubeck et al., 2011). The activity of an archaeal form of RNase H2 that also contains a PCNA-interaction motif is stimulated by PCNA (Bubeck et al., 2011). However, in model studies with oligonucleotides, we failed to see such stimulation for either the yeast or the human enzyme by its cognate PCNA (Chon et al., 2009).

In order to test the potential importance of PCNA in RER, we synthesized ribonucleotide-containing DNA as described above, isolated the DNA, and then used this substrate in a subsequent assay with Pol δ, RNase H2, FEN1, DNA ligase, either with or without PCNA (and its loader RFC) present. In parallel, the same assay was carried out with RNase H2 lacking the PCNA-interacting motif (ΔPI). While PCNA greatly stimulated RER, we observed no dependence on the PCNA-interacting motif of RNase H2 (Figure 3A). Sometimes, a role for PCNA in stabilizing the factor under investigation can be revealed by carrying out the assay under suboptimal conditions, e.g., at low concentrations of the factor and/or at high salt. However, under limiting conditions for RNase H2, at which all other factors were saturating and the rate of DNA repair was dependent on the RNase H2 concentration, both mutant and wild-type RNase H2 showed a comparable inhibition of the rate of DNA repair at increasingly higher concentrations of NaCl (Figure S3). We conclude that either RNase H2 can access the ribonucleotide substrate efficiently without prior interaction with PCNA, or that our studies have failed to reveal a functional interaction in vitro.

**RNase H1 Does Not Substitute for RNase H2 in RER**
Both RNase H1 and RNase H2 incise RNA-containing DNA; however, while RNase H2 incises efficiently at single ribonucleotide positions, RNase H1 is only active on double-stranded DNA containing three or more consecutive ribonucleotides (Tadokoro and Kanaya, 2009). Potentially, interactions of RNase H1 with other repair factors might aid in promoting incision at single ribonucleotides. Therefore, we compared the efficiency of RNase H1 and RNase H2 in RER. RNase H1 was completely defective in promoting RER, even at the highest concentrations tested (Figure 3B).

**FEN1 and Exo1, and Pol δ and Pol ε Show Redundancy for RER**
In our proposed scheme for RER, nick translation is an important step in removing the incised ribonucleotide residue (Figure 4). Exo1 can partially substitute for FEN1 during Okazaki fragment maturation, as both nucleases can promote nick translation by Pol δ (Stith et al., 2008; Tran et al., 2002). Indeed, substituting FEN1 by Exo1 in the RER reaction allowed complete repair of misincorporated ribonucleotides (Figure 3C). However, the rate of repair promoted by FEN1 was about 2- to 3-fold that of Exo1. In contrast, Dna2 nuclease, which only acts on long 5'-flaps, is completely defective for RER. This agrees with our understanding of Dna2 function during Okazaki fragment maturation, wherein long flaps trimmed by Dna2 still require further processing by FEN1 (Burgers, 2009). Similarly, Pol ε can also function in Okazaki fragment maturation, albeit less efficiently than Pol δ (Garg et al., 2004). In a reconstituted reaction, Pol ε carried out RER at a rate about half of that of Pol δ (Figure 3D).

**DISCUSSION**
Here we have reconstituted, for the first time, the excision repair of ribonucleotides in a highly purified system. Our results extend a previous study that used crude extracts to show incisions at single ribonucleotides in DNA by RNase H2 and FEN1. We demonstrate complete and efficient repair of ribonucleotides incorporated into DNA by the major leading and lagging strand replicases for the nuclear genome. The basic mechanism outlined in Figure 4 shows an employment of PCNA, Pol δ, FEN1, and DNA ligase that is analogous to that used during Okazaki
RNase H2-Initiated Ribonucleotide Excision Repair

Figure 4. Model for Ribonucleotide Excision Repair

Redundant functions of FEN1 with Exo1 and Pol δ with Pol ε are indicated. See text for details.

![Model of Ribonucleotide Excision Repair](image)

RNase H2-Initiated Ribonucleotide Excision Repair

sites in the yeast chromosomal DNA isolated from an "mh201" strain that centered in the 2–20 kb range (Nick McElhinny et al., 2010a). Interestingly, a higher frequency of alkali-sensitive sites was detected in a double mutant of "mh201" with a Pol ε mutant that in vitro was shown to display an increased frequency of ribonucleotide incorporation. Importantly, chromosomal DNA isolated from a strain that is wild-type for RNase H2 is much less alkali-labile, indicating that RNase H2 is critical for the repair of ribonucleotides in DNA.

There are two potential reasons why the in vivo occurrence of alkali-sensitive sites in chromosomal DNA from a "mh201" strain, 1 per several kb, is less than the in vitro occurrence, 1 per ~0.7 kb. First, the local concentration of dNTPs at replication foci could be higher than that of the overall cellular concentration that we used in this study, e.g., through the localization of dNTP precursor enzymes at replication sites. Currently, this idea lacks conclusive experimental support, either in yeast or in human cells (Lee and Elledge, 2006; Pontarin et al., 2008). Second, alternative repair systems may exist for the repair of ribonucleotides. Topoisomerase I can incise ribonucleotide-containing DNA, via the covalent 3'-phosphotyrosyl linkage formed at the ribose moiety, into a 2',3'-cyclic phosphate terminated chain (Kim et al., 2011; Sekiguchi and Shuman, 1997). In the cell, this intermediate can be aberrantly processed to yield 2–5 nt deletions within short tandem repeats, suggesting that topoisomerase I may initiate a secondary pathway that can ultimately result in ribonucleotide removal. The enzymes that act downstream of Topoisomerase I to ultimately repair ribonucleotides with consequent mutagenesis are as yet unknown. The complexity of processing ribonucleotides in DNA is further revealed by the fact that survival of yeast in the absence of both RNase H1 and H2 partly depends on two lesion tolerance pathways, Rad5-dependent template switching, and translesion synthesis by Pol ζ (Lazzaro et al., 2012). This is interesting, given our evidence here that RNase H1 does not initiate repair of ribonucleotides incorporated by Pols δ and ε in vitro. Collectively, the results suggest that the cytotoxic lesions in strains defective in both yeast RNase H1 and H2 could be (1) an R-loop formed during transcription (Huertas and Aguilera, 2003); (2) a stretch of several consecutive ribonucleotides placed into DNA during priming of Okazaki fragments or perhaps more rarely by a DNA polymerase; or (3) multiple interrupted ribonucleotides incorporated by a DNA polymerase within a short tract of DNA.

EXPERIMENTAL PROCEDURES

Proteins

Sources are listed in Supplemental Information.
RER Assay

The repair assay is generally carried out in two stages. In the first stage, single-stranded M13mp18 DNA is replicated at physiological cellular concentrations of rNTPs and dNTPs. After replication is complete, the DNA is either isolated or incubated at 54°C for 20 min at 17,000 g, and the pellet was washed with 70% ethanol, dried, and redissolved in TE with 0.1% SDS/50 mM NaCl. The samples were analyzed by electrophoresis on a 1% alkaline agarose gel at 1.3 V/cm for 16 hr. The gel was neutralized, dried, and analyzed by Phosphorimaging. Size distribution determinations were carried out as described in Supplemental Information.

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures, Supplemental Experimental Procedures, and Supplemental References and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2012.06.035.

ACKNOWLEDGMENTS

The authors thank Carrie Stith and Else-Britt Lundström for protein purification and Bonnie Yoder for strain construction. This work was supported in part by Grant GM032431 from the National Institutes of Health to P.M.B.; the Swedish research council and Cancerfonden to E.J.; by Project Z01 ES065089 from the National Institute on Environmental Health Sciences, to T.A.K.; and the Eunice Kennedy Shriver National Institute of Child Health and Human Development, NIH to R.J.C. and T.A.K.

Received: May 7, 2012
Revised: June 14, 2012
Accepted: June 28, 2012
Published online: August 2, 2012

REFERENCES


Checklist of Supplemental Information

1. Additional experimental procedures - **Protein sources.**

2. Additional experimental procedures - **Determination of average ribonucleotide misincorporation frequency.**

3. Supplementary figure - **Fig. S1. Associated with Fig. 1. Replication of M13mp18 DNA in the presence of rNTPs.** This figure shows the slow-down of DNA replication rates by the presence of rNTPs, and the extent of complete replication obtained for each experiment.

4. Supplementary figure - **Fig. S2. Associated with Fig. 2. RER requires the catalytic function of RNase H2.** This figure shows that a mutant form of RNase H2 defective for its catalytic activity, is also defective for RER.

5. Supplementary figure - **Fig. S3. Associated with Fig. 3. The PCNA-interaction motif of RNase H2 is not required for efficient RER.** This figure shows that the PCNA-binding domain of RNase H2 does not aid in RER even at high salt concentrations.
Supplemental Material to:

RNase H2-Initiated Ribonucleotide Excision Repair

Justin L. Sparks¹, Hyongi Chon², Susana M. Cerritelli², Tom A. Kunkel³, Erik Johansson⁴, Robert J. Crouch² and Peter M. Burgers¹*

¹Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, MO 63110; ²Program in Genomics of Differentiation, Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda MD 20892; ³NIEHS; ⁴Department of Medical Biochemistry and Biophysics, Umeå University, SE-901 87, Umeå, Sweden.

EXPERIMENTAL PROCEDURES

Proteins - RPA (Henrickson et al., 1994), PCNA (Eissenberg et al., 1997), RFC (Gomes et al., 2000), FEN1 (Gomes and Burgers, 2000), RNase H1 (Cerritelli et al., 1993), and RNase H2 (Jeong et al., 2004), and RNase H2 mutants (Chon et al., 2009) were purified from *E. coli* overexpression systems, while Pol δ (Fortune et al., 2006), Pol ε (Chilkova et al., 2003), Dna2 (Ayyagari et al., 2003), Exo1 (Gelperin et al., 2005), and DNA ligase (Ayyagari et al., 2003) were purified from yeast overexpression systems.

Determination of average ribonucleotide misincorporation frequency - This frequency follows from the median size of alkali-resistant replication products. First, the radioactivity distribution, between 10 kb and 0.2 kb, was divided up in ~500 segments and the radioactivity value in each segment divided by the size of the DNA in each segment (determined from size markers). This essentially converts the radioactivity distribution into a size distribution of alkali-resistant DNA molecules, from which the median length of the DNA population can be determined. Second, we need to account for the presence of a nick in the completely replicated DNA, and for DNA breakage due to alkaline treatment and handling that is unrelated to the presence of incorporated ribonucleotides. Since the size distributions, on a neutral gel, of M13mp8 DNA replicated with or without rNTPs are the same (Fig. S1B), the median size of alkali-resistant replication products synthesized without rNTPs forms a suitable control. From this we can determine the frequency of ribonucleotide incorporation as:

\[ \text{rNTP incorporation frequency} = \frac{a}{b} - 1 \]

Under the conditions used:

for Pol δ: \( a = 4.1 \text{ kb}, b = 0.61 \text{ kb} \rightarrow \text{rNTP incorporation frequency} = 1/720 \)

for Pol ε: \( a = 3.5 \text{ kb}, b = 0.54 \text{ kb} \rightarrow \text{rNTP incorporation frequency} = 1/640 \)
SUPPLEMENTARY FIGURES

Fig. S1. Associated with Fig. 1. Replication of M13mp18 DNA in the presence of rNTPs. (A) rNTPs slow down replication rates. DNA replication assays were carried out as described in Materials and Methods under standard conditions with PCNA and either Pol δ or Pol ε, with or without rNTPs. Aliquots were taken and acid-precipitable radioactivity determined. (B,C) Complete replication of M13mp18 DNA. (B) Selected aliquots from (A) were electrophoresed on a 1% agarose gel. Note the decrease in the size of replication products at the early time points in the presence of rNTPs. Marker positions are SS-M13mp18 DNA and nicked DS-M13mp18 DNA, and gapped DS-M13mp18 DNA, as shown. (C) The standard assay (lane 3) contains dNTPs plus rNTPs, 100 mM NaCl, 1.5 mM primed M13mp18 DNA, RPA, 15 nM PCNA, 3 nM RFC and 3 nM of Pol δ for 8 min. In lane 1, PCNA and RFC were omitted, NaCl was reduced to 20 mM, Pol δ increased to 500 nM, and the incubation time to 180 min. In lane 2, the assay was the same as standard except for 20 mM NaCl. Replication products were analyzed on a 1% neutral agarose gel (shown here), or digested with NaOH and analyzed on an alkaline agarose gel (shown in Fig. 1C). Marker positions are as indicated.
**Fig. S2.** Associated with Fig. 2. RER requires the catalytic function of RNase H2. Primed M13mp18 DNA was replicated with dNTPs plus rNTPs, RPA, PCNA, RFC and Pol δ for 8 min. Then FEN1, DNA ligase and either wild-type RNase H2 or catalytic-dead RNase H2-D39A was added as indicated, and incubation continued for an additional 4 min. Replication products were digested with NaOH and analyzed on an 1% alkaline agarose gel. Size markers are shown.

**Fig. S3.** Associated with Fig. 3. The PCNA-interaction motif of RNase H2 is not required for efficient RER. Primed M13mp18 DNA was replicated with dNTPs plus rNTPs, RPA, PCNA, RFC and Pol δ for 8 min. Then, the indicated levels of NaCl were added followed by 3 nM FEN1, 15 nM DNA ligase I, and 0.6 nM RNase H2 or RNase H2-ΔPIP, and incubation continued for 3 min. Replication products were digested with NaOH and analyzed on an 1% alkaline agarose gel. Size markers are shown.
SUPPLEMENTARY REFERENCES


CHAPTER IIIB

RNase H2 roles in genome integrity revealed by unlinking its activities
RNase H2 roles in genome integrity revealed by unlinking its activities

Hyongi Chon¹, Justin L. Sparks², Monika Rychlik³, Marcin Nowotny³, Peter M. Burgers², Robert J. Crouch¹,* and Susana M. Cerritelli¹

¹Program in Genomics of Differentiation, Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, MD 20892, USA, ²Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, MO 63110, USA and ³Laboratory of Protein Structure, International Institute of Molecular and Cell Biology, 4 Trojdena Street, 02-109 Warsaw, Poland

Received November 8, 2012; Revised December 21, 2012; Accepted January 1, 2013

ABSTRACT

Ribonuclease H2 (RNase H2) protects genome integrity by its dual roles of resolving transcription-related R-loops and ribonucleotides incorporated in DNA during replication. To unlink these two functions, we generated a Saccharomyces cerevisiae mutant that can resolve R-loops but cannot cleave single ribonucleotides in DNA. This mutant definitively correlates the 2–5 bp deletions observed in strains with single R-loops in DNA. It also establishes a connection between R-loops and sgs1-mediated replication reinitiation at stalled forks and identifies R-loops uniquely processed by RNase H2. In mouse, deletion of any of the genes coding for RNase H2 results in embryonic lethality, and in humans, RNase H2 hypomorphic mutations cause Aicardi–Goutiéres syndrome (AGS), a neuroinflammatory disorder. To determine the contribution of R-loops and R-loops to the defects observed in AGS, we characterized in yeast an AGS-related mutation, which is impaired in processing both substrates, but has sufficient R-loop degradation activity to complement the defects of sgs1 strains. However, this AGS-related mutation accumulates 2–5 bp deletions at a very similar rate as the deletion strain.

INTRODUCTION

RNA and DNA worlds are not entirely separated and frequently come together during replication and transcription, forming transient RNA–DNA structures that when become stable can jeopardize genome integrity. These structures can be divided into two groups: (i) RNA, consisting of one or more ribonucleotides, covalently attached to DNA and opposite to DNA (RNA/DNA, later referred as RpR/DNA), in vivo as part of an R-loop or reverse transcripts (1–3). The former are present during replication, as part of the primers for lagging strand synthesis, or when ribonucleotides are erroneously incorporated into DNA by the replicative polymerases, which, if not removed, leads to mutagenesis (4–6). R-loops are formed during transcription when the RNA exiting the RNA polymerase fails to associate with the posttranscriptional machinery and instead remains annealed to the DNA (7–9). R-loops have the template DNA strand annealed to the RNA, whereas the other DNA strand is in a single-stranded form. Stable R-loops constitute a barrier to transcription and replication fork progression; stalling replication and inducing fork collapse (10,11).

Ribonucleases H (RNase H) comprise a group of enzymes devoted to the removal of both types of RNA–DNA structures. There are two main classes of RNases H, grouped by primary sequence and substrate specificity (12). Type 1 RNases H are able to cleave R-loops (RpR/DNA), but because the enzyme recognizes the RNA via contacts with the 2'-OH residues of four ribonucleotides, they are unable to hydrolyze single R-loops (13). Type 2 RNases H recognize the transition from ribonucleotide to deoxyribonucleotide (RpD), hydrolyzing at the 5'-end of the ribonucleotide, leaving it attached to the 5'-end of the DNA (14). Eukaryotic RNases H2 cleave both RpD/DNA and RpR/DNA structures with similar efficiencies, in contrast to bacterial RNases H2, which have a clear preference for RpD substrates (15). Both types of RNases H are dispensable for viability in bacteria and single-cell eukaryotes (16,17) although genomic instability increases in their absence (18). In higher eukaryotes, both enzymes are essential. RNase H1 is present in nuclei and mitochondria, and it is necessary for mitochondrial DNA replication (19). RNase H2 has been associated with ribonucleotide removal from genomic DNA in yeast and mouse, where it is required for embryonic development (20,21).

*To whom correspondence should be addressed. Tel: +1 301 496 4082; Fax: +1 301 496 0243; Email: Robert_crouch@nih.gov

Published by Oxford University Press 2013.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Mutations in the three genes encoding human RNase H2 (Hu-RNase H2) leads to Aicardi–Goutières syndrome (AGS), a severe neuroinflammatory disorder, with characteristics similar to in utero viral infection (22).

The crystal structure of Thermotoga maritima RNase H2 (Tm-RNase H2) in complex with a DNA containing a single ribonucleotide revealed the crucial role of a conserved tyrosine in the recognition and hydrolysis of the RpD structure (23). This residue displaces the strand to be cleaved and positions the phosphate between the ribose and deoxyribose to coordinate one of the two divalent metal ions required for catalysis. This displacement could not occur efficiently with an RpR structure, and accordingly Tm-RNase H2 cleaves RNA/DNA hybrids very poorly. The conserved Tyr in eukaryotic and accordingly Tm-RNase H2 cleaves RNA/DNA could not occur efficiently with an RpR structure, and accordingly Tm-RNase H2 cleaves RNA/DNA hybrids very poorly. The conserved Tyr in eukaryotic RNases H2 may have a slightly different position allowing both RpD and RpR cleavages.

Some organisms, such as Bacillus stearothermophilus, have two proteins with primary amino acid sequences (Figure 1A) and three-dimensional structures similar to type 2 RNases H (24). However, one (RNase H2) has enzymatic properties similar to bacterial RNases H2, recognizing an RpD transition and cleaving 5' of the ribonucleotide attached to DNA, whereas the other (RNase H3) has properties similar to RNases H1, requiring a short string of ribonucleotides for hydrolysis. We reasoned that we could use the cocrystal structure of Tm-RNase H2 as a guide to superimpose the structure of B. stearothermophilus RNase H3 (Bst-RNase H3) and thereby identify the residues involved in single and multiple rNMPs recognition. Using our model, we made changes in the catalytic subunit of Saccharomyces cerevisiae RNase H2 (Sc-Rnh201) to create an enzyme that retains the ability to hydrolyze RpR/DNA hybrids but is unable to cleave RpD substrates. Constructing such a mutant allowed us to unlink the two activities of RNase H2 and assign in vivo phenotypes to each activity.

MATERIALS AND METHODS

Modeling the structure of Bst-RNase H3 in complex with substrates

To model a structure of Bst-RNase H3 in complex with a substrate, we used the crystal structures of Bst-RNase H3 (PDB: 2D0B), human RNase H1 in complex with RNA/DNA hybrid (PDB: 3PUF) and Tm-RNase H2 in complex with duplex DNA containing single ribonucleotide (PDB: 3O3G). The structures of Bst-RNase H3 and human RNase H1-substrate complexes were superimposed onto that of Tm-RNase H2-substrate complex using DaliLite pairwise comparison (http://www.ebi.ac.uk/Tools/dali-lite/). In addition, the substrate in human RNase H1 was moved to fit on the substrate of Tm-RNase H2 based on the two nucleotides around the cleavage sites using Pymol (http://www.pymol.org/). This program was also used to display structures in designing mutants.

Purification of yeast RNase H2

Previously, we coexpressed and purified the three subunits of yeast RNase H2 with all the subunits C-terminally fused to an His-tag (25). In this study, we constructed an expression system to facilitate purification of the trimeric complex, in which only Rnh201p subunit was C-terminally tagged with 6XHis. The polycistronic coexpression system (pET-yH2ABC) was constructed as described previously for Hu-RNase H2 (15). For overexpression, Escherichia coli BL21(DE3) was transformed with pET-yH2ABC and cultured in Luria-Bertani (LB) media at 30°C. When the optical density at 600 nm (OD$_{600}$) reached around 0.5, 1 mM IPTG was added to the culture medium and cultivation was continued for additional 3 h. Cells were harvested by centrifugation at 5000 rpm for 10 min, resuspended in buffer A (20 mM Tris–HCl, pH 7.5, 1 mM EDTA and 1 mM DTT), disrupted by sonication and centrifuged at 15000 rpm for 30 min. The supernatant was loaded onto a Hitrap Heparin HP column (1 ml) (GE Healthcare) equilibrated with buffer A. The protein was eluted from the column with a linear gradient of 0–1 M NaCl. The fractions containing yeast RNase H2 trimeric complex were collected and applied to the Histrap crude column (1 ml) (GE Healthcare) equilibrated with buffer B (20 mM Tris–HCl pH 7.5, 500 mM NaCl and 15 mM imidazole). The protein was eluted from the column with 60 mM imidazole, dialyzed against 20 mM Tris–HCl pH 7.5 containing 1 mM EDTA, 1 mM DTT and 150 mM NaCl and concentrated. The protein purity was confirmed by SDS–PAGE, followed by staining with Coomassie brilliant blue R250. Expression and purification of yeast RNase H2 mutants were carried out as described for wild-type protein.

RNase H assays

Enzymatic reactions were performed in buffer containing 15 mM Tris–HCl (pH 7.9), 50 mM NaCl, 1 mM DTT, 100 μg/ml BSA, 5% glycerol and 10 mM MgCl$_2$. Activity assays using a uniformly $\gamma^{32}$P-ATP-labeled poly-rA/poly-dT substrate and the 5'32P-labeled short RNA/DNA (the sequences of the oligonucleotides are given in Figure 1 and Supplementary Figure S1) were performed as described previously (15).

Crystallization and structure solution

Tm-RNase H2 G21S was purified as previously described (23). Nucleic acid for crystallization was prepared by annealing high-performance liquid chromatography–purified oligonucleotides: cleaved strand 5'-GACACGTGATTG (single ribonucleotide in small caps) and complementary DNA strand 5'-GAATCGGTTGTC (both purchased from Metabion, Martinsried, Germany). Before crystallization, Tm-RNase H2 G21S (2 mg/ml final concentration) was mixed with the oligos at 1:2.1 substrate:protein molar ratio. The complexes were mixed with the reservoir solution [50 mM CaCl$_2$, 45% PEG 200 and 0.1 M MES (pH 6.0)] at equal volume and crystallized by the sitting drop vapor diffusion method at 18°C. For data collection, the crystals were flash frozen in liquid nitrogen. Diffraction data were collected at Berliner Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung synchrotron at beamline MX-14 (26) on a Mar225 CCD.
Figure 1. Comparison of RNase H2 and H3. (A) Alignment was generated for human RNase H2A (Hu2A), *S. cerevisiae* RNase H2A (Sc2A), *T. maritima* RNase H2 (Tm2), *B. stearothermophilus* RNase H3 (Bst3), *Streptococcus pneumoniae* RNase H3 (Spn3) and *Thermovibrio ammonificans* HB-1 RNase H3 (Tam3). α-Helices are indicated with pink letters, and β-sheets are indicated with orange letters. Active site residues are highlighted in yellow. Amino acid substitutions in *S. cerevisiae* RNase H2A are noted above the first row of alignment D39A, G42S, P45D, S or E and Y219A—along with the conserved DSK triplet. The ‘GRG’ motif, DSK and the conserved Tyr of RNase H2 are also highlighted in grey. (B) The 5'-32P-labeled 12 mer substrates indicated above the gels were digested by yeast and Tm-RNase H2 in the presence of 10 mM MgCl2. The lanes marked with 0 contained no enzyme, and lanes marked with triangle contained increasing amount of the proteins (0.16, 1.6, 16 and 160 nM in the case of Tm-RNase H2, 0.011, 0.11, 1.1 and 11 nM in the case of yeast RNase H2 and 0.011, 0.11, 1.1 and 11 nM in the case of Bst-RNase H3). Products of the hydrolysis were analyzed by 20% TBE-urea gels. The sizes of products were measured based on molecular size markers indicated as M (products of digestion of 32P-labeled strands without complementary strand by phosphodiesterase I). Major cleavage sites of the substrates are summarized on the bottom of each gel. (C) Relative activity of Bst-RNase H3 D103N and D103A mutants compared with the wild-type protein was analyzed by liquid RNase H assay using poly-rA/poly-dT substrate as previously done (15).
Yeast expression plasmid for Rnh201 expression from its native promoter

C-terminally FLAG tagged RNH201 gene coding for WT, D39A, G42S or P45D-Y219A, flanked with 5'-terminal 608 bp and 3'-terminal 404 bp genomic sequence, was cloned into HindIII/NgoMIV site of ycplac111 containing ARS/CEN and LEU2 marker to make ycNPH2-FL2 plasmid expressing WT and mutant RNH201. RNH202 wt and ΔPIP (with a stop codon that makes a protein 11 amino acids shorter eliminating the PIP sequence) was also expressed from plasmid yepLac111 using its own promoter.

Construction of yeast strains

Strain Cy8338 (W303-1A but sgs1::HIS3) (31) was used to first replace the RNH201 and RNH202 genes with the CORE-I-SceI cassette from pGSHU (32). Polymerase chain reaction (PCR) products with the cassette flanked by 60 nucleotides of sequence upstream and downstream of the genes open reading frames were introduced in the strain, generating deletion/replacement of the corresponding gene by homologous recombination (HR). Proper replacement was confirmed by PCR and sequencing. Then, to introduce the desired mutations, WT and mutant genes from plasmid ycNPH2-FL2 were PCR amplified and introduced in the deletion strains (SMC306 and SMC307, Supplementary Table S1), which were growing in the presence of galactose to facilitate HR by I-SceI-mediated double-stranded break. Transformants were selected in 5-FOA plates and confirmed by PCR analysis and sequencing.

Yeabs expression plasmid for Rnh1 overexpression under TPI promoter

Yeast RNase H1 gene was cloned into EcoRI and Sall site of pYX242 (16) to make pYX242-YH1. KanMX4 cassette was cloned into SfoI site of pYX242 and pYX242-YH1.

Growth rate determination and microscopic analysis

Overnight cultures grown in liquid Yeast Extract Peptone Dextrose (YPD) were inoculated into fresh YPD medium for OD_600 to be about 0.1. The OD_600 was measured every hour for 10 h to determine the doubling times. Multiple isolates were analyzed for each strain. Yeast cells fixed with 70% ethanol were stained with 4',6-diamidino-2-phenylindole (DAPI). The cells were immobilized on polylysine-coated slides and photographed in a fluorescent microscope.

Analysis of spontaneous mutation in pol2-M644G rnh201A strain transformed with plasmid expressing rnh201 WT and mutants

The pol2-M644G rnh201A yeast strain (5) was transformed with ycNPH2-FL2 WT and derivatives. Cells were grown in SD-Leu medium and plated on SD-Leu lacking arginine and supplemented with 60 μg/ml canavanine. Genomic DNA was isolated from independent canavine-resistant mutants, PCR amplified and sequenced. Spontaneous mutation rates were measured by fluctuation analysis (33), using FALCOR (Fluctuation Analysis CalculatOR) program (http://www.mitochondria.org/protocols/FALCOR.html). Rates of individual mutation classes were calculated by multiplying the proportion of each mutation type by the total mutation rate.

RESULTS

How RNase H3 recognizes RpR, but not RpD in duplex substrates

We first compared the activities of yeast RNase H2, Tm-RNase H2 and Bst-RNase H3 using RpD and RpR substrates (Figure 1B). Sc-RNase H2 showed the same substrate preference as Hu-RNase H2 (34), cleaving RNA_12, RNA_6-DNA_6 and DNA_7-RNA_1-DNA_6 strands with similar efficiencies when annealed to complementary DNA, although it preferred RpD over RpR sequences when present in the same substrate, cleaving 5' of the ribonucleotide attached to DNA (Figure 1B, left). In contrast, Tm-RNase H2 only recognized RpD sequences, cutting 5' of the ribo (Figure 1B, middle), and Bst-RNase H3 hydrolyzes regions containing a stretch of several ribonucleotides and could not process a single ribonucleotide embedded into DNA (Figure 1B, right), even when a large excess of enzyme (110 nM) was used, or when Mn^{2+} replaced Mg^{2+} (data not shown). Hu-RNase H2 recognizes RpR or RpD structures only when hybridized to DNA (34), whereas some bacterial enzymes can cleave the strand containing RpD junction with RNA or DNA in the opposite strand (35). Sc-RNase H2 acted like the human enzyme cleaving only in regions of RNA/DNA (Supplementary Figure S1, left); however, Tm-RNase H2 could cleave RpD substrates even when the noncleaved strand was RNA (Supplementary Figure S1, right), similar to other bacterial RNases H2 from E. coli and Thermus thermophilus (2).

To determine the reason for the different specificity between RNase H2 and RNase H3, we examined which residues of RNase H3 are involved in catalysis by superimposing the structure of Bst-RNase H3 (PDB: 2D0B) and the structure of Tm-RNase H2 in complex with a substrate containing a RpD junction (PDB: 3O3G) (Figure 2A and C). We then replaced the RpD junction substrate with the structure of the RNA/DNA hybrid co-crystallized with Hu-RNase H1 (PDB: 3PUF), and determined the interactions of the Bst-RNase H3/Tm-RNase H2 superimposition (Figure 2 B and D). Instead of the Tm-RNase H2 Tyr residue essential for catalysis, Bst-RNase H3 has an Asp (D103) that,
according to the modeled structure, was within 3.3 and 3.2 Å from the 2'-OH of the ribose residues at positions +1 and +2, respectively. Moreover, the second catalytic residue (Glu-98 of Bst-RNase H3 and Glu-19 of Tm-RNase H2) could interact with the 2'-OH at position -1, suggesting that RNase H3 interacts with at least three consecutive ribonucleotides of the cleaved strand. We tested the activity of Bst-RNase H3 with D103N or D103A substitutions and found that these enzymes have reduced RNase H activity (Figure 1C); especially Bst-RNase H3-D103A that could not form hydrogen bonds with either of the 2'-OH at position +1 or +2, confirming the important role of D103 in substrate recognition and interaction with two consecutive 2'-OH.

Creating an Sc-RNase H2 that cleaves RpR not RpD sequences

In RNase H3, the residue corresponding to the conserved junction-binding tyrosine from RNase H2 is an alanine (Figure 1A). Tyr to Ala in Sc-Rnh201, the catalytic subunit, resulted in an enzyme with less than 0.1% residual activity with all substrates (Figure 3 and Table 1), as previously shown for Hu-RNase H2, indicating that the Tyr is important for cleavages of both RpD and RpR substrates. To see whether substitution of Y219 by amino acids with OH moieties could provide interaction with the RNA 2'-OH, we replaced Y219 with S, D or E. Rnh201-Y219D and Rnh201-219E were inactive with all substrates, whereas Rnh201-Y219S was inactive for the cleavage of RpD substrates but has higher activity than Rnh201-Y219A with hybrids containing RpR junctions (Figure 3A and B). The structure modeled in Figure 2 suggested that D103 in Bst-RNase H3 plays a role similar to Tm-Y163 in contacting the ribose 2'-OH at position +1. D103 of Bst-RNase H3 is in the same spatial location as P45 of Sc-Rnh201 (Figure 1A, 1C and 2D). Because D103 is a Ser or Glu in other RNases H3 (Figure 1A), we changed P45 to D, E or S in the Sc-Rnh201-Y219A. Sc-RNases H2 containing Rnh201-P45D-Y219A or Rnh201-P45E-Y219A had about 40% activity of the wt enzyme for the cleavage of poly-rA/poly-dT substrate, whereas Sc-RNase H2 with Rnh201-P45S-Y219A substitution was less efficient (Figure 3A). In contrast, RpD molecules were essentially not hydrolyzed by yeast RNase H2 enzyme containing Rnh201-Y219A-P45D (Figure 3B). With substrates containing RpR and RpD junctions in the same hybrid (top panel of Figure 3B), the mutant Rnh201-P45D-Y219A showed about 3.1% of the wt activity on RpR structures, while it was less than 0.1% as active as the wt for the RpD cleavage (Table 1). Actually, activity on the 6R substrate may be low due to length of RNA, sequence preferences, or poor stability related to the sequence or the proximity to the end (see Figure 1B). Thus, the accuracy of our model was confirmed. Y219A mutation in Sc-Rnh201 eliminated recognition of RpD and RpR, but the additional replacement of P45 with D provided a key interaction with a 2'-OH of two consecutive ribonucleotides conferring specificity for RpR/DNA cleavage. Therefore, we have converted Sc-RNase H2 into an enzyme that solely cleaves RpR/DNA substrates.

Tm-RNase H2 substrate specificity

To examine whether we could change the specificity of Tm-RNase H2 to cleave RpR sequences similarly to eukaryotic RNases H2, we substituted C24 by D and Y163
Figure 3. Cleavage of poly-rA/poly-dT substrate and oligo substrates with Sc-RNase H2 WT and mutants. (A) The uniformly $^{32}$P-labeled poly-rA/poly-dT substrate (1 μM) was digested by the Sc-RNase H2 WT and mutants indicated above the gel in the presence of 10 mM MgCl₂. The lanes marked with 0 contained no enzyme, and lanes marked with triangle contained increasing amount of the proteins (1 pM, 11 pM, 110 pM, 1.1 nM and 11 nM). Products of the hydrolysis were analyzed by 12% TBE-urea gels and visualized by Phosphorimager. (B) The 5'-$^{32}$P-labeled 12 mer substrates indicated on the left of the gels were digested by Sc-RNase H2 in the presence of 10 mM MgCl₂. The lanes marked with 0 contained no enzyme, and lanes marked with triangle contained increasing amount of the proteins (0.011, 0.11, 1.1 and 11 nM). Products of the hydrolysis were analyzed by 20% TBE-urea gels. The sizes of products were measured based on molecular size markers indicated as M. Cleavage at the 5' of RpD junction is indicated by an arrow.
Table 1. Relative activities of yeast RNase H2 derivatives

<table>
<thead>
<tr>
<th></th>
<th>1R *</th>
<th>6R *</th>
<th>12R *</th>
<th>poly-rA/poly-dT</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Y219A</td>
<td>&lt;0.1</td>
<td>&lt;0.1</td>
<td>&lt;0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>P45D Y219A</td>
<td>&lt;0.1</td>
<td>3.1b (&lt;0.1)c</td>
<td>25</td>
<td>40</td>
</tr>
<tr>
<td>G42S</td>
<td>2.2</td>
<td>5.1</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>D39A</td>
<td>&lt;0.1</td>
<td>&lt;0.1</td>
<td>&lt;0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Rnh202-ΔPIP</td>
<td>109</td>
<td>85</td>
<td>112</td>
<td>118</td>
</tr>
</tbody>
</table>

\*12 mer substrates shown in Figure 1B.
\*Efficiency for the cleavage at RpR.
\*Efficiency for the cleavage at RpD.

by A in Tm-RNase H2, which are the equivalents to the Sc-Rnh201 P45D-Y219A substitutions. The Tm-RNase H2 WT enzyme in the presence of Mg\(^{2+}\) cleaves exclusively at the 5'-end of the ribonucleotide of the RpD junction. In the presence of Mn\(^{2+}\), the WT enzyme is able to cleave RpP structures although with lower affinity than RpD sequences. Tm-RNase H2-Y163A had undetectable activity with Mg\(^{2+}\) and very low activity in the presence of Mn\(^{2+}\) with all the substrates (Supplementary Figure S2). Tm-RNase H2-C24D-Y163A was inactive with Mg\(^{2+}\) and only partially active in Mn\(^{2+}\). For the cleavage of RNA\(_{6c}\) DNA\(_6\)/DNA\(_{12}\) substrate, this enzyme cleaves at the 5'-end of both the RpD and RpP sequences, whereas the WT cuts only 5' of the RpD, suggesting that the mutant lost selectivity for the RpD junction. This indicated that we have identified in the yeast and Tm-RNases H2 the amino acids that confer selectivity for RpD and RpP hydrolysis.

The Gly residue, which in the Tm-RNase H2 co-crystal is part of the GRG motif interacting with the 2'-OH of the RpD substrate, is important for substrate specificity (23). This glycine in the catalytic Hu-RNase H2A is particularly interesting because it is substituted for Ser in patients presenting a severe form of AGS (22). Tm-RNase H2-G21S has low Mn-dependent activity on all substrates and reduced RpD processing (23) in the presence of Mg\(^{2+}\). To understand the effect of the G37S-equivalent mutation in the prokaryotic protein, we solved the crystal structure of Tm-RNase H2-G21S in complex with a duplex DNA containing a single ribonucleotide (Supplementary Figure S3). Surprisingly, in the crystal structure, interactions between the substrate and protein were only modestly perturbed. The main chain of the GRG motif was slightly deformed, still allowing full recognition of the 2'-OH (Supplementary Figure S3C). Moreover, the role of the conserved Tyr in the recognition of the 2'-OH of the scissile phosphate and in the stacking interaction with D of the RpD junction was intact, suggesting Tm-RNase H2-G21S retains the ability to recognize RpD junctions. The mutation affected the dihedral angle of the loop containing G21S and the surrounding residues due to the presence of the side chain of Ser21, which needs to be accommodated at the center of protein–nucleic acid interface. These small perturbations around the active site disturbed the binding of metal ion B. The structure suggested that G21S substitution in the GRG loop results in a general defect in catalysis, not a selective loss of RpD or RpR hydrolysis. We made the corresponding change in the catalytic subunit of Sc-RNase H2 and found that the G42S substitution showed reduced activity with the different substrates, having about half the activity with the single-ribonucleotide embedded duplex DNA sequence than with longer stretches of ribonucleotides (Table 1). Similar results have been reported for the mouse (36–38), human (15,34), yeast (25) and Tm-RNase H2 (23).

Recognition of single ribonucleotide by RNase H2 is essential for in vitro ribonucleotide excision repair assay

We examined the ability of WT and mutant forms of Sc-RNase H2 to perform in an in vitro ribonucleotide excision repair (RER) assay. Three independent experiments reproducibly gave the results shown in Figure 4. This replication assay relies on several proteins to produce mp18 dsDNAs devoid of misincorporated ribonucleotides as shown in Figure 4A. Alkali treatment results in DNA strand scissions at the ribonucleotide locations as seen in Figure 4A. When Sc-RNase H2 is present, the majority of dsDNA becomes insensitive to alkali treatment (Figure 4A, WT lane after 3 min incubation) indicating the misincorporated ribonucleotides are repaired in an RNase H2-dependent manner. Proliferating Cell Nuclear Antigen (PCNA) is loaded onto the circular DNA by RFC and increases the activities of DNA polymerase δ, FEN1 and ligase1 by interacting with their PCNA-interacting peptide (PIP). However, as we have shown before (39), interaction with PCNA is not necessary for the repair activity of RNase H2, because an RNase H2 with an Rnh202 subunit missing the last 11 amino acid including the PCNA-interacting peptide (PIP) was as active as the wild-type enzyme (Figure 4A, ΔPIP lane after 3 min incubation). Accordingly, the in vitro activity of the ΔPIP protein with different substrates is equivalent to the activity of the WT enzyme (Table 1).

The RNase H2 containing the catalytic subunit Rnh201-P45D-Y219A mutation, which could not cleave RpD substrates in vitro (Figure 3 and Table 1), also did not function to repair misincorporated ribonucleotide in this assay (Figure 4A). Sc-RNase H2-G42S exhibited low activity in single ribonucleotide cleavage (Table 1) and in the RER assay showed a modestly larger distribution of products compared with the catalytically dead Sc-RNase H2-D39A and the Sc-RNase H2-P45D-Y219A mutants (Figure 4B). The in vitro activity of these mutants correlates well with their capacity to process rNMPs in the RER assay.

Mutants defective in RpD cleavage have the same mutator effect as deletion of RNH201

Yeast cells lacking RNH201 accumulate 2–5 bp deletions within short tandem repeats in a topoisomerase 1-dependent manner (6). In combination with error-prone replicative DNA polymerase ε mutant (pol2-M644G), deletion of RNH201 increased the spontaneous mutation rate and the frequency of the 2-5 bp deletions (5). We transformed the double mutant strain, pol2-M644G rnh201Δ, with single copy plasmids expressing
RNH201-WT, rnh201-G42S or rnh201-P45D-Y219A driven by their own promoter and a control empty vector. The transformants were analyzed for the spontaneous mutation rate and mutational specificity at the CAN1 locus (Figure 5A). Plasmid-expressed WT RNH201 decreased the mutation rate and showed a mutation spectrum mostly of base substitutions, similar to wild-type cells. However, pol2-M644G rnh201D cells expressing rnh201-P45D-Y219A had similar mutation rates and spectra as the deletion strain pol2-M644G rnh201D harboring an empty vector (Figure 5A). Cells expressing rnh201-G42S had a slightly reduced rate of 2–5 bp deletions, compared with pol2-M644G rnh201A cells expressing rnh201-P45D-Y219A, but still had a much higher accumulation of deletions than wild-type cells. Western analyses demonstrated that WT, G42S and P45D-Y219A Rnh201p were expressed at similar levels in these strains (Supplementary Figure S4).

Mutants defective in RpD cleavage can participate in R-loop processing

Next, we determined whether these enzymes can function in vivo in the processing of RNA/DNA hybrids. We chose a system that requires R-loop processing for viability. Loss of Top1 in combination with RNH1 and RNH201 deletions leads to R-loop accumulation and consequent cell death in yeast (40,41). We used yeast strain YAEH275 (BY4741 but PGAL1-3HA-TOP1 (KanMx6) rnh201D (NatMx6) rnh1D (HphMx6)) was transformed with plasmids expressing RNH201-WT, rnh201-G42S, -P45D-Y219A and -D39A and also an empty vector. The transformants were grown at 30°C in liquid SD-Leu medium containing 2% galactose, and series of dilutions were spotted on SD-Leu plates with 2% galactose or 2% glucose and incubated for 3 days at 30°C.

Figure 5. In vivo phenotypes of strains rnh201Δ carrying plasmids harboring RNH201 WT and mutants. (A) Saccharomyces cerevisiae strain pol2-M644G rnh201Δ was transformed with plasmid expressing RNH201WT and mutants and empty vector. Strains were plated in canavanine plates and the rates of individual mutations types at CAN1 determined. N, number of mutants sequenced; indel, insertion/deletion; BS, base substitution. 95% confidence intervals are in parenthesis below total rates. (B) Saccharomyces cerevisiae strain YAEH275 (BY4741 but PGAL1-3HA-TOP1 (KanMx6) rnh201Δ (NatMx6) rnh1Δ (HphMx6)) was transformed with plasmids expressing RNH201WT, rnh201-G42S, -P45D-Y219A and -D39A and also an empty vector. The transformants were grown at 30°C in liquid SD-Leu medium containing 2% galactose, and series of dilutions were spotted on SD-Leu plates with 2% galactose or 2% glucose and incubated for 3 days at 30°C.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1-bp indel</td>
</tr>
<tr>
<td>WT (N=42)</td>
<td>48.6</td>
</tr>
<tr>
<td>rnh201Δ (N=44)</td>
<td>78.0</td>
</tr>
<tr>
<td>G42S (N=94)</td>
<td>62.9</td>
</tr>
<tr>
<td>P45D-Y219A (N=98)</td>
<td>69.8</td>
</tr>
</tbody>
</table>

Figure 4. In vitro RER assay. (A) In vitro RER assay is schematically shown. The 7.3-kb mp18 single-stranded circle DNA was replicated by Pol6 and cofactors in the presence of the physiological levels of dNTPs, rNTPs and 32P-labeled dATP at 30°C. After 12 min incubation, FEN1, Ligase1 and wild-type or mutant RNase H2 were added, and the reactions were continued. After 1 or 3 min, aliquots were taken and treated with NaOH for hydrolysis at RpD linkage due to misincorporation of ribonucleotide during replication. DNA was extracted from the aliquots and loaded onto 1% alkali agarose gel. The alkali agarose gel was visualized by Phosphorimager. (B) The radioactivity distribution was scanned and divided by the size distribution to obtain a normalized product distribution as described before (39).
substrates of the RNase H2 enzyme containing this mutant protein (Table 1). The catalytically dead rnh201-D39A protein did not rescue lethality in glucose. These results indicate that R-loop removal \textit{in vivo} parallels that seen \textit{in vitro} (Table 1).

\textbf{RNase H2 R-loop processing is sufficient to correct rnh201Δsgs1Δsynthetic sickness}

In \textit{S. cerevisiae}, Sgs1 is crucial for HR repair of double-stranded DNA breaks and for re-establishing of replication after fork collapse (42,43). Deletion of any of the three genes that code for the subunits of the heterotrimeric RNase H2 together with deletion of \textit{SGS1} confers a synthetic-sickness phenotype (44), suggesting that the loss of RNase H2 creates a need for recombinational repair.

To examine whether RER or R-loop processing is necessary in the absence of \textit{SGS1}, we analyzed a set of isogenic strains that were both \textit{sgs1Δ} and mutated in \textit{RNH201} to either eliminate catalytic activity completely (D39A) or eliminate RER but retain R-loop processing (P45D-Y219A or G42S). The growth rate was examined by colony size on YPD plates and by the analysis of doubling time in liquid medium. The doubling time was determined at least four times for each strain. The error bars represent standard deviation (Figure 6A). Chromosomal RNH201-WT, -rnh201-P45D-Y219A or -G42S suppressed the slow growth phenotype, while the strain with active site mutant \textit{rnh201-D39A} showed the same poor growth as the parental \textit{sgs1Δ rnh201Δ} strain (Figure 6A).

These data were confirmed using plasmid complementation assays (Supplementary Figure S5).

To eliminate the possibility that an RNase H2-independent suppressor mutation had restored the normal growth of the \textit{sgs1Δ} strain expressing \textit{rnh201-G42S} and -P45D-Y219A, we replaced the \textit{rnh201-G42S} and -P45D-Y219A alleles by KanMX-cassette, recreating the \textit{rnh201Δ} and thereby reproducing the slow growth rate (data not shown). Taken together, these results

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure6.png}
\caption{Effect of RNase H2-mutations in \textit{sgs1Δ} yeast strain. (A) Doubling time of the yeast strains with RNase H2-mutations in genomic DNA was determined at 30°C in YPD medium. Error bars are standard deviations with $P < 0.1$ measured by the Ttest in Excel. (B) Representatives of normal cells and abnormal cells from \textit{sgs1Δ rnh201Δ} background stained with DAPI are shown. (C) Fraction of abnormal cells from log-phase culture was analyzed by gating enlarged cells by FACS, which was correlated with the ratio of abnormal cells observed by microscope (data not shown).}
\end{figure}
suggest that the growth defect of the \textit{sgs1A rnh201A} strain is caused by R-loops that become stable in the absence of RNase H2 and Sgs1 and that the residual activity of the enzymes containing P45D-Y219A and, surprisingly, G42S mutants is sufficient to complement these defects.

In the \textit{sgs1A rnh201A} strain, there was an increase of nonproliferative abnormal cells arrested in G2/M phase (Figure 6B), that seem to be unable to complete the cell cycle, presumably due to replication and repair impairments. By microscopic analysis, we initially noticed increased levels of enlarged dumbbell-shaped cells or cells with abnormal shape as shown in Figure 6B. Next, analysis of the cells for DNA content (propidium iodide (PI) staining) and cell size (side scattering) by Fluorescence-activated cell sorting (FACS) (data not shown) suggested an increased portion of enlarged G2/M-arrested cells in \textit{sgs1A rnh201A} mutant, which is in agreement with the microscopic analysis. In Figure 6C, enlarged cells observed by side scattering of FACS data were used to determine the fraction of abnormal cells.

The strain with \textit{sgs1A rnh201-1D39A} had elevated level of abnormal cells as \textit{sgs1A rnh201A} mutant, whereas strains \textit{sgs1A rnh201-1G42S} and \textit{sgs1A rnh201-1P45D-Y219A} had the same amount of abnormal cells as single mutant \textit{sgs1A} Figure 6C, indicating that RNase H2 R-loop resolution prevents increased formation of abnormal cells.

To determine whether interaction with PCNA was important in this genetic background, we mutated \textit{RNH202} by gene replacement in the strain carrying \textit{sgs1A} to make an RNase H2 enzyme devoid of the PIP sequence. RNase H2-ΔPIP complemented the deletion of \textit{RNH202} (Figure 6A), indicating that interaction with PCNA is not necessary for R-loop processing.

\textbf{RNase H1 overexpression did not complement \textit{sgs1A rnh201A} synthetic sickness}

R-loop processing can be accomplished by either RNases H1 or H2. This seems to be the case for the R-loops in nucleoli (40). To determine whether RNase H1 can also participate in the processing of the substrates that accumulate in the \textit{sgs1A rnh201A} mutant, we transformed the \textit{sgs1A rnh201A} strain with a plasmid that carries the \textit{sgs1A} gene by gene replacement in the strain carrying \textit{sgs1A} to make an RNase H2 enzyme devoid of the PIP sequence. RNase H2-ΔPIP complemented the deletion of \textit{RNH201} (Figure 6A), indicating that interaction with PCNA is not necessary for R-loop processing.

\textbf{DISCUSSION}

Eukaryotic RNases H1 and H2 are important participants in maintaining genome stability by resolving R-loops that form during transcription and by initiating the removal of rNMPs in DNA. Only RNase H2 can process single ribonucleotides in DNA but both enzymes are capable of eliminating RNA/DNA hybrids. Roles of RNase H1 or RNase H2 are usually examined by deleting one or both enzymes, which in \textit{S. cerevisiae} leads to modest increases in sensitivity to DNA damaging agents (16), but in mouse has devastating consequences, causing embryonic lethality (1,20,21). R-loops naturally form due to transcription errors and are removed by RNase H1 and RNase H2. However, under high transcription (40) and when RNA biogenesis factors are defective (8,9,18), R-loop formation is increased, inducing genome instability. Stable R-loops constitute a block to replication fork progression causing fork collapse, recombination and chromosomal breaks (45,46). The deleterious consequences of the R-loops can be at least partially reversed by overexpression of RNase H1, indicating that normal levels of RNase H1 and RNase H2 are insufficient when there are more abundant R-loops.

In contrast to transcription-induced R-loop formation, single ribonucleotides in DNA occur during genome duplication, and RNase H2 is essential to initiate their removal. In the absence of RNase H2, rNMPs in the genome are aberrantly processed by Top1 (6), originating 2–5 base pairs deletions within short tandem repeats. R-loops and single ribonucleotides in DNA are quite distinct structures that created different challenges to genome integrity, and loss of RNase H2 may result in an accumulation of defects in transcription, replication and repair. Therefore, deletion of the \textit{RNH201} gene does not permit the specific assignment of contributions of R-loops and single rNMPs to genome instability.

To assess the separate roles of RNase H1, and particularly RNase H2 in R-loop degradation and RER, we designed a Sc-RNase H2 variant with only two amino acid substitutions in the catalytic subunit (Rnh201-P45D-Y219A), which behaves like a bacterial RNase H3, i.e. the enzyme can participate in R-loop processing but not in the removal of single rNMPs misincorporated in DNA. This mutant constituted an invaluable tool to understand the functions of RNase H2 under normal and pathological condition. This mutant allowed us to conclusively correlate the 2–5 bp deletions observed in \textit{rnh201A} strains (6) with single ribonucleotide misincorporation in the genome. It also suggested a connection between R-loops and Sgs1-mediated replication reinitiation at stalled forks, as well as identifying R-loops that can be processed by RNase H2, but not by RNase H1 (Figure 7).

Genome instability in yeast is greater when both \textit{RNH1} and \textit{RNH201} genes are deleted than when single deletions are present, although in combination with defects of RNA biogenesis factors, the deletion of \textit{RNH1} is more deleterious than that of \textit{RNH201} (18). This seems to indicate that RNases H1 and H2 have partially overlapping specificities in the resolution of R-loops, but RNase H1 is the major player in removing transcription-associated R-loops (Figure 7B). Conversely, our results suggest that there are R-loops that are uniquely processed by RNase H2. These loops might be involved in DNA transactions during replication/repair processes (Figure 7C), and RNase H2 could have access to them by interaction with
PCNA and the replication machinery via the PIP motif of the Rnh202 subunit. Although no phenotype has been associated thus far with the deletion of the PIP motif, it is likely than under stress circumstances or other conditions not yet tested, the PIP might be required for the function of RNase H2. Moreover, the PIP sequence may not be the only means of RNase H2 interaction with DNA replication/repair machines.

Both RNases H complemented the lethal defects caused by the accumulation of R-loops created under high transcription in a topoisomerase-defective \( \text{rnh1}\Delta \text{ rnh201}\Delta \) mutant, as did the Rnh201-P45D-Y219A mutant that still processes R-loops (Figure 5B). However, only RNase H2 was effective in correcting the slow growth of \( \text{sgs1}\Delta \text{ rnh201}\Delta \) strains (Figure 6 and Supplementary Figure S6). The synthetic defect of \( \text{sgs1}\Delta \text{ rnh201}\Delta \) could be due to misincorporated rNMPs in DNA, especially because \( \text{sgs1}\Delta \text{ rnh1}\Delta \) strains do not exhibit a slow growth phenotype. However, the fact that the mutant completely devoid of single-ribonucleotide processing activity could reverse the slow growth defect of \( \text{sgs1}\Delta \text{ rnh201}\Delta \) (Figure 6A) indicated that persistent R-loops have deleterious effects in this strain.

R-loops present from bacteria to humans constitute a block to replication leading to replication fork stalling (10). Replication after fork arrest can be reinitiated by HR (46), for which Sgs1 is essential (47). Sgs1 recruits the intra-S phase checkpoint kinase Rad53 to stalled forks and contributes to replication fork stability by dissolving fold-back structures and Holliday junctions formed at collapsed forks (48). Our study establishes a link between stable R-loops and Sgs1-mediated replication restart.

The forms of RNase H2 containing the AGS-related mutation Rnh201-G42S, which \textit{in vitro} has low activity on RNA/DNA hybrids, could reverse the slow growth and accumulation of abnormal cells in the \( \text{sgs1}\Delta \) background, indicating that even a very impaired enzyme has sufficient activity to process the small number of hybrids formed in this strain. However, the mutant RNase H-G42S was only partially effective in YAEH275 (\( \text{PGAL-\text{TOP1 rnh1}}\Delta \text{ rnh201}}\Delta \)), possibly related to the higher abundance of R-loops in this strain, which might exceed the capacity of the defective RNase H2-G42S. Alternatively, for the resolution of R-loops in the \( \text{sgs1}\Delta \) strain, RNase H2 could be part of a larger complex that may confer stability to the mutant enzyme and consequently increasing enzymatic activity.

\textit{In vitro}, the mutant RNase H2-G42S had low but detectable activity on single rNMP (Table 1) and was
partially effective in the RER assay (Figure 4). \textit{In vivo}, although the RNase H2-Rnh201-P45D-Y219A mutant strain exhibited a 3.9-fold increase over the wt in mutation rate, the same as the Rnh201 deletion, the Sc-RNase H2 containing Rnh201-G42S has an increase slightly lower (3-fold over wt) mutation rate, and fewer 2–5 bp deletions (Figure 5A), indicating that this mutant and presumably the human enzyme with the corresponding mutation are able to remove single rNMPs in the genome at some low frequency (Figure 7A).

In agreement with the \textit{in vitro} and \textit{in vivo} results, the Tm-RNase H2 G21S structure in complex with a single ribonucleotide–substituted duplex DNA showed preserved interactions of the RpD junction with the conserved Tyr and with the GRG (SRG in the mutant) motif, indicating the enzyme has reduced enzymatic activity while maintaining substrate specificity. These data are important to understand the defects associated with the corresponding mutation in the human enzyme, which is responsible for a severe form of AGS, providing a note of caution in interpreting data when both activities are decreased or absent.

Recently, defects in mouse due to deletion or low expression of RNase H2 (20,21) have been attributed to genome instability associated with the presence of rNMPs in DNA, without considering the possible effects of unresolved R-loops, and RNA/DNA hybrids formed during retroelements reverse transcription. However, the deletion of another AGS-inducing gene, \textit{Trex1}, has been shown to cause the accumulation of single-stranded DNA derived from endogenous retroelements (49). In addition, antiretroviral drugs prevented autoimmune response and restored the viability of \textit{Trex1}-knockout mice (50). Also, SAMHD1, a more recently described AGS-associated protein (51), has been found to restrict the infection of dendritic and other myeloid cells by human immunodeficiency virus type 1 (HIV-1) by hydrolyzing intracellular dNTPs (52). Low concentrations of dNTPs, due to the action of SAMHD1, inhibit the synthesis of the viral RNA by reverse transcriptase (RT). In the absence of SAMHD1, higher levels of dNTPs could facilitate reverse transcription of endogenous retroelements. AGS substrates arising via reverse transcription of endogenous RNAs would lead to standard RNA/DNA duplexes that could be resolved by RNase H2 and RNase H1. The fact that RNase H1 does not cause AGS suggests that RNase H2 has unique access to these substrates. The mutant described in this study that could degrade RNA/DNA hybrids formed by RT but could not remove single rNMPs in DNA will be enormously important to address the defects associated with AGS.

Here, we have shown that the multiple activities of RNase H2 can provide quite different functions \textit{in vivo}. Our mutant RNase H2 lacking RpD substrate cleavage allows us and others to precisely analyze the contributions of R-loops and single rNMPs in DNA to the \textit{in vivo} phenotypes until now associated with deletion of RNase H2. It will be particularly interesting to see the pattern of changes in gene expression when the \textit{rnh201}-P45D-Y219A gene is present compared with the 349 changes when RH201 is deleted (53), and how the loss of single-ribonucleotide activity of RNase H2 affects the general genomic instability described by Wahba \textit{et al.} (18).

**ACCESSION NUMBERS**

RCSB ID code rcsb075483 and PDB ID code 4HHT.

**SUPPLEMENTARY DATA**

Supplementary Data are available at NAR Online: Supplementary Figures 1–6, Supplementary Tables 1–3 and Supplementary References [5,31,40,53–55].

**ACKNOWLEDGEMENTS**

The authors thank M. Foiani for strain Cy8338, T. Kunkel for strain Δ(-2)-7B-YUNI300 and A. El Hage and D. Tollervey for strain YAEH275.

**FUNDING**

Intramural Program of the Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health (in part); National Institutes of Health [GM032431, in part] to P.M.B.; Wellcome Trust International Senior Research Fellowship [081760 to M.N.]; M.N. is an HHMI International Early Career Scholar. Funding for open access charge: Intramural Program of the Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health.

Conflict of interest statement. None declared.

**REFERENCES**


SUPPLEMENTARY DATA

RNase H2 roles in genome integrity revealed by unlinking its activities

Hyongi Chon¹, Justin L. Sparks², Monika P. Rychlik³, Marcin Nowotny³, Peter M. Burgers², Robert J. Crouch¹ and Susana M. Cerritelli¹

¹Program in Genomics of Differentiation, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892, USA

²Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, MO 63110, USA

³Laboratory of Protein Structure, International Institute of Molecular and Cell Biology, 4 Trojdena Street, 02-109, Warsaw, Poland
Supplementary Figure S1. Cleavage of short substrates with yeast and Tm-RNases H2.

The 5'-32P labeled 12 mer substrates indicated above the gels were digested by Yeast and Tm-RNase H2 in the presence of 10 mM MgCl2. The lanes marked with 0 contained no enzyme, and lanes marked with triangle contained increasing amount of the proteins (0.16, 1.6, 16 and 160 nM in the case of Tm-RNase H2, 0.011, 0.11, 1.1 and 11 nM in the case of yeast RNase H2). Products of the hydrolysis were analyzed by 20% TBE-urea gels. The sizes of products were measured based on molecular size markers indicated as M (products of digestion of 32P-labeled strands without complementary strand by phosphodiesterase I). Major cleavage sites are summarized on the bottom.
Supplementary Figure S2. Effect of RNase H3-like double mutation in Tm-RNase H2.

The 5'-32P labeled 12 mer substrates schematically shown on the left of the gels were digested by Tm-RNase H2 WT, Y163A and C24D-Y163A mutants in the presence of 10 mM MgCl2 and 1 mM MnCl2. The lanes marked with 0 contained no enzyme, and lanes marked with triangle contained increasing amount of the proteins (1.6, 16, 160 and 1600 nM). Products of the hydrolysis were analyzed by 20% TBE-urea gels. The sizes of products were measured based on molecular size markers indicated as M (products of digestion of 32P-labeled strands without complementary strand by phosphodiesterase I).
Supplementary Figure S3. Crystal structure of Tm-RNase H2 G21S mutant in complex with DNA containing single ribonucleotide.

Overall structures of wild-type and G21S mutant are shown. (A) N-terminal domain is colored with purple for beta-sheet and orange for loop and alpha-helix. C-terminal domain is in yellow. Residues (G21, K47 and Y163) and RpD junctions are shown in sticks. Cartoon model of the substrate is shown in blue for DNA and red for single ribonucleotide. (B) Protein structure except S21 (in purple), Y163 (in yellow) is shown in gray. (C) Close-up view for the recognition of RpD junction by Tm-RNase H2 WT and G21S mutant. G21S mutant is shown in gray except the G21S mutation (in magenta). The scissile phosphate is indicated with an arrow. (D) Close-up view for the correlation between the recognition of RpD junction by Y163 and catalysis. Structure of G21S mutant is shown in gray. Calcium ions of WT complex is shown in green. The scissile phosphate is indicated with an arrow. (E) Sample of electron density of the GRG(S) motif, Y163 and the RNA-DNA junction. The RNA is shown in red and DNA in blue. A 2Fo-Fc map contoured at 1.0 sigma is overlaid on the structure.
Supplementary Figure S4. Expression of C-terminally FLAG-tagged RNase H2A from the plasmid, ycNPH2-FL2 and its variants

(A) Yeast cells expressing C-terminally FLAG-tagged Rnh201 WT (lane 2), D39A (lane 3), G42S (lane 4), P45D-Y219A (lane 5) or no protein (lane 1) from the plasmids with Rnh201 gene flanked by its own native promoter were analyzed with western analysis using anti-FLAG M2 antibody. (B) The PVDF membrane for Figure S4A was stained with Ponceau-S.
Supplementary Figure S5. Complementation of slow growth phenotype of sgs1Δ rnh201Δ strain with plasmids.

Transformants of sgs1Δ rnh201Δ strain with plasmids expressing Rnh201-WT, G42S, P45D-Y219A or D39A were streaked on YPD plate and cultured at 30°C. The table shows doubling time of the yeast strains analyzed in YPD medium.
Supplementary Figure S6. Effect of overexpression of RNase H1 on growth rate of sgs1Δ rnh201Δ strain.

sgs1Δ rnh201Δ strain was transformed with a plasmid overexpressing Sc-RNase H1 and the doubling time was measured in YPD medium containing G418 at 30°C. The relative growth rate is shown.
## Supplementary Table S1. List of yeast strains and plasmids

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>W303-1A</td>
<td>MATa ade2-1 can1-100 his3-11,-15 leu2-3,112 trp1-1 ura3-1</td>
<td>(55)</td>
</tr>
<tr>
<td>Cy8338</td>
<td>W303-1A but sgs1::HIS3</td>
<td>(56)</td>
</tr>
<tr>
<td>SMC306</td>
<td>W303-1A but sgs1::HIS3 rnh201::HYG-URA3-GAL1-IsceI</td>
<td>This study</td>
</tr>
<tr>
<td>SMC311</td>
<td>W303 genotype but sgs1::HIS3 Rnh201-D39A</td>
<td>This study</td>
</tr>
<tr>
<td>SMC308</td>
<td>W303 genotype but sgs1::HIS3 Rnh201-G425</td>
<td>This study</td>
</tr>
<tr>
<td>SMC310</td>
<td>W303 genotype but sgs1::HIS3 Rnh201-P45D-Y219A</td>
<td>This study</td>
</tr>
<tr>
<td>SMC307</td>
<td>W303 genotype but sgs1::HIS3 rnh202::HYG-URA3-GAL1-IsceI</td>
<td>This study</td>
</tr>
<tr>
<td>SMC314</td>
<td>W303 genotype but sgs1::HIS3 Rnh202-ΔPIP::KanMX</td>
<td>This study</td>
</tr>
<tr>
<td>YAEH275</td>
<td>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 PGAL1-3HA-TOP1::KanMx6 rnh201::NatMx6 rnh1::HphMx6</td>
<td>(40)</td>
</tr>
<tr>
<td>Δ[(-2)]-7B-YUNI300</td>
<td>MATa CAN1 his7-2 leu2-Δ::kanMX ura3-Δ trip1-289 ade2-1 lys2-ΔGG28899-2900 pol2-M644G rnh201Δ::HYG-R</td>
<td>(5)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>ycplac111</td>
<td>LEU2, ARS/CEN</td>
<td>Novagen</td>
</tr>
<tr>
<td>ycNPH2-FL2</td>
<td>C-terminally FLAG-tagged Rnh201 flanked by its own native promoter in ycplac111</td>
<td>This study</td>
</tr>
<tr>
<td>pYX242-Kan</td>
<td>LEU2, 2micro, KanMX4 at SfoI site</td>
<td>Novagen</td>
</tr>
<tr>
<td>pYX242-YH1-Kan</td>
<td>Sc-Rnh1 was cloned under TPI promoter of pYX242, KanMX4 at SfoI site</td>
<td>This study</td>
</tr>
</tbody>
</table>
Supplementary Table S2. Crystallographic data collection and refinement statistics for the Tm-RNase H2 G21S-substrate complex.

<table>
<thead>
<tr>
<th>Data collection</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>C 2</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td></td>
</tr>
<tr>
<td>(a, b, c) (Å)</td>
<td>106.95, 47.63, 77.45</td>
</tr>
<tr>
<td>(a, b, g) (°)</td>
<td>90, 131.26, 90</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.91841</td>
</tr>
<tr>
<td>Resolution (Å)*</td>
<td>50.00 – 3.10 (3.15-3.10)</td>
</tr>
<tr>
<td>(R_{merge}) (%)</td>
<td>14.6 (37.3)</td>
</tr>
<tr>
<td>(I / s(I))</td>
<td>9.8 (2.2)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>96.9 (76.3)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>3.3 (2.2)</td>
</tr>
<tr>
<td>Refinement</td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>3.1</td>
</tr>
<tr>
<td>No. reflections</td>
<td>5022</td>
</tr>
<tr>
<td>(R_{work} / R_{free}) (%)</td>
<td>17.8/25.4</td>
</tr>
<tr>
<td>No. atoms</td>
<td>2189</td>
</tr>
<tr>
<td>Protein</td>
<td>1676</td>
</tr>
<tr>
<td>Nucleic acids</td>
<td>487</td>
</tr>
<tr>
<td>Ligands/ions</td>
<td>2</td>
</tr>
<tr>
<td>Water</td>
<td>24</td>
</tr>
<tr>
<td>Avg. B-factor</td>
<td>40.1</td>
</tr>
<tr>
<td>Protein</td>
<td>38.7</td>
</tr>
<tr>
<td>Nucleic acids</td>
<td>45.3</td>
</tr>
<tr>
<td>Ligands/ions</td>
<td>35.9</td>
</tr>
<tr>
<td>Water</td>
<td>30.3</td>
</tr>
<tr>
<td>Root-mean-square deviation</td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.009</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.255</td>
</tr>
</tbody>
</table>

* Values in parentheses are for highest-resolution shell.
Supplementary Table S3. MolProbity.

<table>
<thead>
<tr>
<th>All-Atom Contacts</th>
<th>Clashscore, all atoms: 28.21</th>
<th>88th percentile * (N=59, 2.85Å – 9999Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Clashscore is the number of serious steric overlaps (&gt; 0.4 Å) per 1000 atoms.</td>
<td></td>
</tr>
</tbody>
</table>

### Protein Geometry

<table>
<thead>
<tr>
<th>Poor rotamers</th>
<th>8.67%</th>
<th>Goal: &lt;1%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ramachandran outliers</td>
<td>0.46%</td>
<td>Goal: &lt;0.2%</td>
</tr>
<tr>
<td>Ramachandran favored</td>
<td>93.52%</td>
<td>Goal: &gt;98%</td>
</tr>
<tr>
<td>Cβ deviations &gt;0.25Å</td>
<td>0</td>
<td>Goal: 0</td>
</tr>
<tr>
<td>MolProbity score*</td>
<td>3.07</td>
<td>77th percentile * (N=2114, 3.100Å ± 0.25Å)</td>
</tr>
<tr>
<td>Residues with bad bonds:</td>
<td>0.00%</td>
<td>Goal: 0%</td>
</tr>
<tr>
<td>Residues with bad angles:</td>
<td>0.00%</td>
<td>Goal: &lt;0.1%</td>
</tr>
</tbody>
</table>

### Nucleic Acid Geometry

<table>
<thead>
<tr>
<th>Bad backbone conformations*:</th>
<th>0</th>
<th>Goal: 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residues with bad bonds:</td>
<td>0.00%</td>
<td>Goal: 0%</td>
</tr>
<tr>
<td>Residues with bad angles:</td>
<td>0.00%</td>
<td>Goal: &lt;0.1%</td>
</tr>
</tbody>
</table>

All-atom structure was validated for macromolecular crystallography (5).
REFERENCES


CHAPTER IV

Reconstitution of Top1-initiated Ribonucleotide removal and consequent mutagenesis
This chapter of the dissertation focuses on the role of Topoisomerase 1 in the repair of genomic rNMPs and the mechanism of Topoisomerase induced 2-5 bp deletions at di-nucleotide repeats in the absence of RNase H2 in *S. cerevisiae*. Top1 cleaves at an rNMP in dsDNA leaving a 2’, 3’-cyclic phosphate terminated ssDNA nick. Subsequently, a second Top1 cleavage event happens two-nucleotides to the 5’-side of the cyclic-phosphate and the intervening two-nucleotides can dissociate, trapping Top1 on the DNA. From there, if the trapped Top1 is in a di-nucleotide repeat sequence, the Top1 can lead to the realignment of the DNA to allow for religation causing the deletion of two nucleotides. On the other hand, if the site is not in a di-nucleotide repeat it must be repaired by another pathway. The trapped Top1 can be removed in a TDP1 dependent pathway leading to repair. I carried out all of the biochemical assays for this work and purified most of the protein. Carrie Stith also purified several proteins used in these experiments.
Reconstitution of Topoisomerase 1-Initiated rNMP-Dependent Mutagenesis in *S. cerevisiae*

Justin L. Sparks¹ and Peter M. Burgers¹*

¹Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, MO 63110, USA
Replicative polymerases frequently incorporate ribonucleoside monophosphates (rNMPs) into the genome that are normally repaired by the RNase H2-initiated ribonucleotide excision repair. In *S. cerevisiae* deletion of *rnh201*, the catalytic subunit of RNase H2, results in rNMPs remaining in the genome and leads to ~100-fold increase in the frequency of 2-5 bp deletions at di-nucleotide repeat sequences that are dependent on topoisomerase I (Top1) activity. Top1 has endoribonuclease activity against rNMPs in RNA/DNA hybrids. Here we present an *in vitro* reconstitution of the mechanism of Top1-dependent deletions at di-nucleotide repeat sequences and a mechanism for Top1-initiated removal of rNMPs outside of the context of these repeat sequences in *S. cerevisiae*. Top1 attack of rNMP leads to the formation of a 2’, 3’ cyclic phosphate terminated ssDNA nick, allowing for subsequent formation of a Top1-cleavage complex (Top1-cc). If the rNMP is in the context of a di-nucleotide repeat there can be realignment of the DNA allowing for religation and release of Top1, leading to a small deletion. If the rNMP resides outside a repeat sequence the realignment is not possible and a different pathway must repair the Top1-cc. Tdp1-dependent repair of Top1-cc requires prior proteolytic processing of the Top1-cc before it can be removed, leaves a 3’-phosphate that can be removed by Tpp1, Apn1, or Apn2 and forms a substrate suitable for repair by DNA polymerase δ, FEN1 and DNA ligase.
INTRODUCTION

DNA replicative polymerases misincorporate approximately two ribonucleotides per kilobase of DNA replicate (Sparks, Chon et al. 2012). These incorporated ribonucleotides (rNMPs) constitute the most abundant replication error in cells. Removal of these rNMPs from the genome is important to maintain genomic stability (Nick McElhinny, Kumar et al. 2010). RNase H2-initiated ribonucleotide excision repair is the main pathway that removes ribonucleotides from the genome. In *Saccharomyces cerevisiae*, strains lacking functional RNase H2 are relatively normal but maintain large numbers of ribonucleotides in their genome (Nick McElhinny, Kumar et al. 2010), indicating yeast may have alternative mechanisms for replicating through rNMP containing DNA. In humans, mutations in any of the three subunits of RNase H2 are responsible for a genetic syndrome known as Aicardi-Goutières syndrome (AGS) (Crow, Leitch et al. 2006). The mechanistic details of the pathogenesis are only recently becoming clear, but many questions still remain about what function(s) the RNase H2 mutations disrupt that lead to the disease. Further work in yeast may lead to an understanding of how incorporated rNMPs are processed in the absence of RNase H2. Our data may have relevance for understanding the underlying cause of the pathology of RNase H2 mutations.

RNase H2 deficient yeast strains show a distinct mutational pattern that is characterized by an increase in 2-5 bp deletions, specifically at sites of di-nucleotide repeats (Nick McElhinny, Kumar et al. 2010; Kim, Huang et al. 2011). The increase in 2-5 bp deletions in the RNase H2-deficient cells is dependent on topoisomerase 1 (Top1) (Kim, Huang et al. 2011). Top1 is a type IB topoisomerase characterized by going through a covalent intermediate with the 3’-phosphate of the attacked strand. If Top1 is unable to religate the scissile DNA strand by displacement or loss of the 5’-hydroxyl then the Top1 can become covalently linked to the DNA through the 3’-phosphate. These covalent complexes are termed Top1-cleavage complexes (Top1-cc). Human and vaccinia viral topoisomerase 1 proteins have been shown to possess endoribonuclease activity toward ribonucleotides in RNA/DNA hybrids (Sekiguchi and Shuman 1997; Kim, Huang et al. 2011). The rNMP possesses a 2’-hydroxyl that can attack the covalently linked 3’-phosphate, which releases the Top1 and leaves a nick with the 3’-terminus as an rNMP with a 2’, 3’-cyclic phosphate (Sekiguchi and Shuman 1997). A model has been proposed for the Top1 induced mutagenesis based on these studies in which Top1 first cleaves the incorporated rNMP and leaves a 2’, 3’-cyclic phosphate terminated nick within or near a di-nucleotide repeat sequence. Subsequently, a second Top1 cleavage occurs to the 5’-side of the nick, within six nucleotides, which allows the diffusion of the short intervening oligonucleotide, including the rNMP, and leads to the formation of a Top1-cc (Cho, Kim et al. 2013). Resolution of this Top1-cc would
result from realignment on the non-cleaved DNA strand, allow proximity of the 5’-hydroxyl to the 3’-phosphate, permit religation of the DNA, and release of the Top1 protein (Cho, Kim et al. 2013). The following round of replication or processing by mismatch repair pathway would lead to the 2-5 bp deletion in one of the daughter cells (Strand, Prolla et al. 1993). No direct biochemical studies have been undertaken to test this model in vitro. Top1-dependent rNMP-initiated mutagenesis phenomenon is similar to the Transcription-Associated-Mutagenesis (TAM) in yeast, which also occurs specifically at di-nucleotide repeats in regions of highly transcribed sequences and is dependent on Top1. TAM is thought to occur by collision between the RNA polymerase and Top1 forming Top1-cc that must be resolved (Lippert, Kim et al. 2011). The mechanistic details of how Top1-cc leads to this mutagenesis is still unclear, and in vitro reconstitution will be important in understanding how Top1 causes these short deletions.

Resolution of Top1-cc has been an intense focus of study for some time, as several classes of therapeutic cancer drugs target Top1 and stabilize the cleavage complex leading to DNA damage. Several pathways have been shown to resolve Top1-cc in cells. One pathway has been extensively studied, the TDP1 (tyrosyl-DNA phosphodiesterase 1)-dependent pathway. Biochemical studies support a model in which Top1-cc must be partially unfolded and/or partially proteolyzed before Tdp1 is able to remove Top1 (Debethune, Kohlhagen et al. 2002). Several cell biological studies support a role of the proteasome in the Top1 proteolysis (Lin, Ban et al. 2008). The partially proteolyzed Top1-peptide can be removed by attack of the phospho-tyrosyl bond of the Top1 active site with the 3’-phosphate of the DNA (Debethune, Kohlhagen et al. 2002). Tdp1 activity leaves a 3’-phosphate that must then be removed by several redundant pathways. These pathways include protein phosphatases Tpp1, Apn1, and Apn2 (Vance and Wilson 2001). Tpp1 (three prime phosphatase 1) is a 3’-phosphatase that is distantly related to the T4 polynucleotide kinase but does not have the kinase domain. Apn1 and Apn2 are apurinic/apyrimidinic endonucleases that have endonuclease, 3’-diester, and 3’-5’ exonuclease activity. Genetic evidence supports a redundant role for these proteins in removal of the 3’-phosphate after Tdp1 removal of the Top1-peptide (Vance and Wilson 2001). There have been several genetic studies that have implicated other pathways for resolving the Top1-cc, including roles for the Rad1-Rad10 endonuclease as well as the homologous recombination machinery (Liu, Poulion et al. 2002). Further studies will be needed to resolve the mechanistic detail of these other pathways. Along with mutagenic effect of Top1 at sites of rNMPs a recent study has clearly demonstrated that, in the absence of RNase H2, Top1 leads to the repair of a small portion of the rNMPs which are incorporated into the yeast genome (Williams, Smith et al. 2013). The mechanistic details of how Top1 leads to the repair of incorporated rNMPs are unclear.
In this study, we set out to investigate the mechanism of Top1-initiated repair of incorporated rNMPs using in vitro reconstitution, and to test the model previously proposed for the Top1-initiated mutagenesis. Our data supports a previously proposed model by Kim et al., in which Top1 cleaves on the 3'-face of rNMPs and is released by attack of the phosphate by the 2'-hydroxyl. The resulting product is a single-strand nick with a 3'-terminal 2', 3'-cyclic phosphate, and a 5'-hydroxyl. Top1 subsequently attacks the nicked strand, 2-5 nucleotides to the 5'-side of the ribonucleotide, which leads to the release of a short oligonucleotide including the rNMP. The release of the 2-nt oligonucleotide leads to Top1 covalently linked to the 3'-phosphate of the DNA. If the rNMP resides within a di-nucleotide repeat it leads to realignment of the cleaved strand leading to Top1 religation and subsequent loss of two nucleotides. The position of the rNMP in the di-nucleotide plays a major role in the ability of the realignment and subsequent religation by Top1. For rNMPs cleaved outside di-nucleotide repeats, our results favor a model in which the Top1-cc must be processed either by a Tdp1-dependent pathway or an endonuclease-dependent pathway for repair to occur. We have reconstituted the Tdp1-dependent removal of the Top1-cc leading to the repair of the rNMP. We show that the Tdp1 pathway requires Top1 to be partially unfolded and or degraded before it can be a substrate for Tdp1 as previously reported (Debethune, Kohlhagen et al. 2002). Tdp1 leaves a 3'-phosphate that can be removed by Tpp1 or Apn1, which results in a substrate that can be extended by the replication machinery. We also have evidence that the tRNA ligase is able to resolve the 2’, 3’-cyclic phosphate to a 2’-phosphate, but have been unable to determine how this product is resolved in cells. More work will need to be carried out to determine if there are other pathways involved in resolving the 2’, 3’-cyclic phosphate product.

RESULTS

*S. cerevisiae* Top1 possesses endoribonuclease activity leaving a 3’-terminal 2’, 3’-cyclic phosphate.

We began our study by overexpressing and purifying *S. cerevisiae* Top1 and testing the ability of the purified protein to cleave at incorporated rNMPs. We used an oligonucleotide system that reflects the sequence of a di-nucleotide repeat, identified as a Top1-dependent deletion hotspot in the *S. cerevisiae* *CAN1* gene (Kim, Huang et al. 2011). The double-stranded oligonucleotides contain a single ribonucleotide at position either proximal or internal to the GA di-nucleotide repeat on the 32P-labeled strand. The addition of *S. cerevisiae* Top1 to the 5'-32P-labeled rNMP-proximal substrate yielded two products; the first is consistent with the protein cleaving on the 3’-side of the rNMP, and the second is a high molecular weight product that remains in the well of
the gel (Fig 1A). These products are dependent on the presence of the rNMP in the substrate and are not observed in the absence of rNMP (Supp. Fig 1; lanes 7 and 13). Addition of Top1 to a 3’-32P-labeled substrate again shows a cleavage event on the 3’-side of the incorporated rNMP (Supp. Fig 1; lanes 7 and 8). The 3’-labeled cleavage product is resistant to alkaline phosphatase as is an RNase A cleavage product indicating Top1 cleavage leaves a 5’-hydroxyl terminus (Supp. Fig 1B; lanes 3, 4 and 7, 8). RNase H2, on the other hand, leaves a 5’-phosphate that is susceptible to alkaline phosphatase treatment (Supp. Fig 1B; lanes 5 and 6). These data support our hypothesis that \textit{S. cerevisiae} Top1 has endoribonuclease activity when ribonucleotides are incorporated into DNA. This cleavage leaves a 5’-hydroxyl and a 3’-phosphate. It was previously shown that vaccinia virus topoisomerase I cleaves rNMPs in DNA, leaving a 2’, 3’-cyclic phosphate at the 3’-terminus of the rNMP (Sekiguchi and Shuman 1997). We hypothesized \textit{S. cerevisiae} Top1 cleavage at the rNMP would also result in the formation of the 2’, 3’-cyclic phosphate at the 3’-terminus. We set up a system to investigate the DNA end left by the Top1 cleavage. The 5’-Cy3-label Top1 rNMP cleavage product is resistant to alkaline phosphatase treatment, which is able to remove 2’ or a 3’-phosphate, but unable to remove a 2’, 3’-cyclic phosphate (Fig 1B; lane 3). Also, the yeast 3’-DNA specific phosphatase Tpp1 is not able to remove the 2’, 3’ cyclic phosphate (Fig 1B; lane 4). Additionally, prior treatment with Trl1, the yeast tRNA ligase, is able to unhook the cyclic phosphate leaving a 2’-phosphate that can be removed by alkaline phosphatase but not Tpp1 (Fig 1C; lanes 5, 6 and 7). T4 polynucleotide kinase was used as a control since it was shown to be able to unhook 2’, 3’-cyclic phosphates and subsequently remove the 3’-phosphates from DNA (Fig 1C; lane 8) (Das and Shuman 2013). It is still unclear if Trl1 unhooks a 3’-terminal 2’, 3’-cyclic phosphate in dsDNA \textit{in vivo}, and if the resulting 2’-phosphate is resolved in \textit{S. cerevisiae}. More work will be needed to determine if Trl1 has a role for resolving these lesions \textit{in vivo}. We have been unable to find any other proteins that are able to resolve the cyclic phosphate in \textit{S. cerevisiae}. It will be important to determine if the cyclic phosphate may be a dead-end product of the reaction \textit{in vivo}.

\textit{Top1-dependent rNMP nicks cause the formation of Top1-cc.}

A high molecular weight product is observed that increases with increasing Top1 concentrations, and is dependent on the presence of the rNMP in the substrate (Fig 1A). The 5’-labeled oligonucleotide remains in the wells of the gel, and addition of SDS does not change the running of the high molecular weight product. We hypothesized this product was the formation of a Top1-cc in which Top1 was covalently linked to the DNA through a 3’-phosphotyrosyl bond. The formation of the covalent DNA-Top1 complex increases over time (Fig 1C). The formation of the
Top1-cc results from Top1 cutting within six nucleotides of a persistent nick in the DNA which allows the short oligonucleotide to diffuse away, trapping Top1 (Cho, Kim et al. 2013); we have data about the precise distance from the rNMP nick that we will present later in this manuscript. Our results demonstrate that S. cerevisiae Top1 forms two products that are dependent on the presence of an rNMP in the dsDNA. First, we observe the previously reported 2’, 3’-cyclic phosphate at the rNMP; and, second, a Top1-cc where Top1 is covalently linked to the cleaved strand. We set out to understand the order of the reaction in the formation of these two products.

We hypothesize the formation of the rNMP-terminated nick by formation of a 2’, 3’-cyclic phosphate precedes the formation of the Top1-cc. The formation of both the cyclic phosphate product and the Top1-cc were monitored over time (Fig 1D). The Top1-DNA complex was resolved on the urea-PAGE gel by first degrading the Top1 protein with proteinase K, which gave several products that corresponded to different lengths of Top1-peptide linked to the ssDNA. We observe an early accumulation (5%) and subsequent decrease of the 2’, 3’-cyclic phosphate cleavage product (Fig 1E). The formation of Top1-cc is slower than the 2’, 3’ cyclic phosphate and accumulates over time (81%) (Fig 1E). Our data is consistent with a model where Top1 first cleaves at the rNMP leaving the 2’, 3’-cyclic phosphate. Subsequently, a second Top1 cleavage event occurs upstream of the nick and allows for the diffusion of the short oligonucleotide, including the rNMP, which causes Top1 to be covalently linked to the DNA (Fig 2B) (Kim, Huang et al. 2011; Cho, Kim et al. 2013).

rNMP-induced Top1-cc can be resolved by Tdp1.

One pathway for removal of Top1-cc is through the Tdp1-dependent pathway. We tested the ability of purified S. cerevisiae Tdp1 to remove the covalently linked Top1 from our DNA substrate. As previously reported, the full length Top1-cc is resistant to Tdp1 treatment (data not shown) (Debethune, Kohlhagen et al. 2002). We then set up a reaction where the rNMP containing substrate was first treated with Top1 and subsequently stopped by addition of EDTA, SDS, and proteinase K to degrade Top1 down to small peptides. The purified double stranded oligonucleotide was then used for subsequent reactions. The proteinase K treatment leads to several different Top1-peptide-DNA products corresponding to different peptide lengths (Fig 3A; lane 1). Using this substrate, Tdp1 is efficiently able to remove the Top1-peptides from the 3’-phosphate of the ssDNA (Fig 3A, lane 2). Tpp1 (Fig 3A; lane 3), Apn1 (Supp Fig2; lane 8), or Apn2 (Supp Fig2; lane 9) are able to remove the remaining 3’-phosphate with varying efficiencies. The resulting product is consistent with the second Top1 cleavage; it occurs primarily one-nucleotide 5’ of the rNMP (89%) and leads to the release of a di-nucleotide,
including the rNMP, and leaving a two-nucleotide gap (Fig 3A; lane 3). We also observe a small amount of product that is consistent with the second cleavage occurring two (4%), three (5%), or four (2%) nucleotides from the rNMP-terminated nick and leaving a three, four, or five nucleotide gap, respectively. Our data demonstrates the second Top1-dependent cleavage event happens preferentially, two-nucleotides 5’ to the nick with minor products that are three to five nucleotides 5’ to the nick.

When an rNMP is in the context of di-nucleotide repeat sequence Top1 alone can lead to 2-nucleotide deletions in vitro.

Previously it was demonstrated that Top1 was capable of cleaving at incorporated rNMPs, and that in the context of a rnh201 deletion in S. cerevisiae Top1 is able to remove a small portion of incorporated rNMPs, but causes short deletions that specifically occur at di-nucleotide repeat sequences in the process (Kim, Huang et al. 2011; Williams, Smith et al. 2013). We set out to test whether the location of the rNMP, either internal or proximal to a di-nucleotide repeat, affects the processing activity of Top1 treatment. The treatment of two, 32-nucleotide rNMP containing dsDNA substrates, with the rNMP either internal (GAGArGA) or proximal (GAGAGArU) to the di-nucleotide repeat sequence, reveals different Top1-dependent products (Fig 3A). First, Top1 treatment of the rNMP internal substrate (GAGArGATGC) leads to the 2’, 3’-cyclic phosphate cleavage and the Top1-cc, but surprisingly we also observe the appearance of a small fraction of a 30-nucleotide religation product (5%) at 90 minutes (Fig 3A). Top1 treatment of the rNMP proximal substrate (GAGAGArUGC) also leads to the formation of the 2’, 3’-cyclic phosphate terminated nick and the Top1-cc, but no religation products are observed (Fig 3A). This is the first biochemical reconstitution of the Top1-dependent deletions at di-nucleotide repeats. To further test the importance of the sequence context of the di-nucleotide repeat we increased the length of the di-nucleotide repeat by one repeat unit (GAGAGArGATGC), and repeated the Top1 treatment. As expected, the increase in repeat-length increases the kinetics of religation, and the fraction of religation product increases from 5% to 46% after 90 minutes (Fig 3A). We then hypothesized that changing ribo-guanine to a ribo-cytosine would disrupt the realignment of the Top1-cc and decrease religation; in fact, this is what we observe with the change of the rGA to a rCA. The efficiency is drastically reduced from 46% religation product to 5% after 90 minutes. A comparison of the accumulation of Top1-dependent products between the rGA and rCA sequences demonstrates that the kinetics of the first Top1 cleavage event and formation of the Top1-cc are identical; but, on the rGA substrate the Top1-cc is then converted to the 32-mer religation product, whereas the rCA substrate Top1-cc remains stable with little religation (Fig
3B). Similar religation activity is seen with a different di-nucleotide repeat (TATA) and the rNMP positioned internally, so that realignment of the DNA 3’-oligonucleotide is required. Treatment of repeat sequence (GArUATGC) allows for possible realignment on the 3’-side of the 2-nucleotide gap, formed after the formation of the Top1-cc, and is also able to form the religation product with faster kinetics and even more efficiency (81%) after 90 minutes (Fig 3A). The starting substrate is susceptible to cleavage by NaOH, consistent with the presence of an rNMP in the substrate; whereas the Top1-dependent religation product is resistant to NaOH, which supports our hypothesis that the rNMP is removed from the religation product by Top1 (Fig 3C). The removal of the rNMP by Top1 would be one possible mechanism of rNMP removal previously observed (Williams, Smith et al. 2013). Our data supports the previously proposed model by Chon et al. (Fig 5). These results are also relevant for another Top1-dependent mutagenic phenomenon, transcription associated mutagenesis (TAM), characterized by accumulation of short deletions specifically at di-nucleotide repeat sequences in highly transcribed regions of the genome (Lippert, Kim et al. 2011). We were also interested in how Top1 would lead to repair of incorporated rNMPs outside of the context of di-nucleotide sequences, as these incorporation sites would represent the majority of rNMP incorporations in the genome. We began by investigating the ability of the Tdp1 dependent pathway to remove Top1-cc that are stable when the ssDNA ends are unable to realign to allow for religation by Top1.

_Tdp1-pathway leads to substrate for strand displacement synthesis by Pol δ._

We used the Top1-peptide linked oligonucleotide and hybridized it to a template that contained both a 5’ and 3’-biotin to allow for binding to streptavidin. The streptavidin does not allow RFC loaded PCNA to slide off of the DNA and allows more efficient replication by polymerase δ (Pol δ). Therefore, preloading PCNA onto this substrate with dNTPs does not allow the addition of Pol δ to extend the peptide-linked oligonucleotide primer (Fig 4; lane 2). Neither is it able to extend the oligonucleotide treated with Tdp1 or Tpp1 alone (Fig4; lanes 4 and 6). Pol δ is only able to extend the oligonucleotide when both Tdp1 and Tpp1 are both added to the reaction leaving a 3’-hydroxyl, and the extended product is resistant to treatment with sodium hydroxide which is consistent with the rNMP being removed from the substrate (Fig 4; lane 8 and 9). We also observe that both Apn1 and Apn2 are able to remove the Top1- peptides from the DNA with different efficiencies and leaving a 3’-hydroxyl with subsequent 3’-exonuclease activity, as previously observed (Supp. Fig 2) (Liu, Pouliot et al. 2002). The genetic data from previous studies indicate that Apn1 and Apn2 may not remove Top1 from DNA in vivo because epistasis
experiments demonstrate that Tdp1 is epistatic with Apn1 and Apn2, indicating these proteins are in the same pathway (Liu, Pouliot et al. 2002). The authors believe Apn1 and Apn2 play a redundant role with Tpp1 in removal of the 3’-phosphate left by Tdp1 activity. Our data demonstrates that a Tdp1-dependent pathway can efficiently remove the Top1-cc caused by the rNMP. First, the covalently linked Top1 must be partially unfolded and degraded, leaving DNA-linked peptides that can be removed by Tdp1 and a 3’-phosphate that can be subsequently removed by Tpp1, Apn1, or Apn2, resulting in a 3’-hydroxyl. The replication clamp PCNA and Polδ are then able to carry out strand displacement synthesis with the help of FEN1 and ligase that allow for repair of the ssDNA gap (Fig 5). There are several pathways for removal of Top1-cc that have been described genetically and all of these pathways will likely play a role in removal of the Top1-cc left at after attack of an rNMP in vivo. More work will be needed to reconstitute these pathways in vitro.

DISCUSSION

Topoisomerase 1 has been clearly demonstrated to cause short-deletions at di-nucleotide repeat sequences in yeast, including transcription-associated mutagenesis (TAM) and rNMP-induced mutagenesis, in the absence of a functional RNase H2 complex. How Top1 leads to these deletions has been a mystery, and it was unclear if other factors were involved in this process. Chon et al. proposed a model for Top1-dependent deletions based on genetic evidence. Here we present the first reconstitution of the Top1-mediated deletions using purified S. cerevisiae Top1. We believe that the data presented here is relevant to both Top1-dependent mutagenic phenomena, both being mediated through a Top1-cc formed in a di-nucleotide repeat sequence.

Our data supports a model where Top1 first cuts at the rNMP in the context of a repeat sequence, leaving a 2’, 3’-cyclic phosphate terminated nick and allowing for the formation of a Top1-cc two nucleotides to the 5’-side of the nick. Next, the repeat sequences allow for realignment of the cleaved strand with the terminal nucleotides perfectly base paired to allow for religation of the cleaved strand by Top1 (Fig 5). It is possible that other factors influence, either in a positive or negative fashion, the creation of these deletions but more work will be needed to test that hypothesis. Our model explains the requirement of repetitive DNA sequence to allow for realignment of the DNA ends to allow for perfect base pairing, and bulging of nucleotides downstream from the religation. We have also reconstituted one mechanism of repair of rNMPs outside these repeat sequences where we see no evidence of Top1-dependent religation. Our results support a model where Top1-cc must be processed by a repair pathway such as the Tdp1-dependent pathway. The majority of rNMP incorporation events likely occur outside of the
context of a repeat sequence leading to the majority of Top1-dependent rNMP removal leading to a stable Top1-cc that must be repaired. These lesions will most likely be deleterious to the cell which fits nicely with in vivo studies in the context of an rnh201Δ and pol2-M644G, a Polɛ mutant that incorporates rNMPs at a higher frequency than wild type Polɛ; these cells have constitutive checkpoint activation and increased sensitivity to replication stress that is dependent of Top1 (Williams, Smith et al. 2013).

Further studies will need to be carried out to determine if there are other factors that contribute to the Top1-dependent deletions. It will also be interesting to determine if Trl1 is able to lead to the removal and religation of the 2', 3'-cyclic phosphate terminated intermediate formed by Top1 attack at an rNMP.

ACKNOWLEDGEMENTS
The authors thank Carrie Stith for protein purification.

EXPERIMENTAL PROCEDURES

Proteins and oligonucleotides- RPA (Henricksen, Umbricht et al. 1994), PCNA (Eisenberg, Ayyagari et al. 1997), RFC (Gomes, Gary et al. 2000), FEN1 (Gomes and Burgers 2000), and RNaseH2 (Jeong, Backlund et al. 2004) were purified from E. coli overexpression systems, while Pol δ (Fortune, Stith et al. 2006) were purified from yeast overexpression systems. Top1 was overexpressed by pRS425-GGT-TOP1 containing Schistosoma japonicum glutathione S-transferase (GST) gene fused to the N-terminus of the TOP1 gene in vector pRS425-GALGST-term (Walker, Crowley et al. 1993). The GST tag is separated from the N-terminus of the Top1 by a recognition sequence for the human rhinoviral 3C protease (LEVLFQ/GP). Following cleavage by the protease, the N-terminal sequence of Top1 is extended with the GPEFDIKL sequence. Tpp1, Trl1, Apn1, and Apn2 proteins were overexpress from the Open Biosystems S. cerevisiae open reading frame library in the vector backbone BG1805.

Oligonucleotides were purchased from IDT (Coralville, IA) and purified by HPLC chromatography: NTS-GCT, TTTGAAATACCGTGGCATCTCTCGTGACGAGT; TS-GCT, ACTCGTCACGAGAGATGCCACCGGTATTTCAAA; NTS-GCT-Biotin, /biotin/CTACAGCACAGATTCAGCAATTAAGCTCTAATTTGAAATACCGTGGCATCTCTCGTGACGAGT/biotin/; TS-TC-rU5'Cy3, /Cy3/ACTCGTCACGAGAGATGCCACCGGTATTTCAAA; TS-TC-rG5'Cy3, /Cy3/ACTCGTCACGAGAGATGCCACCGGTATTTCAAA; NTS-GAGAGAg, TTTGAAATACCGTGGCATCTCTCGTGACGAGT; TS-GAGAGArGA5’Cy3,
/Cy3/ACTCGTCACGAGAGArGATGCCACCGGTATTTCAAA; NTS-GAGAGArU,
TTTGAATACCCTGGCATATCTTCTCGTGACGAGT; TS-GAGAGArU,
/Cy3/ACTCGTCACGAGAGArUAAGCCACGGTATTTCAAA; NTS-CTCTCTATTCTTCTTCTCGTGACGAGT; TS-GAGAGArUA5’Cy3,
/Cy3/ACTCGTCACGAGAGArUAAGCCACGGTATTTCAAA; NTS-CTCTCTATTCTTCTTCTCGTGACGAGT; TS-GAGAGArUA5’Cy3,
/Cy3/ACTCGTCACGAGAGArUAAGCCACGGTATTTCAAA; NTS-CTCTCTATTCTTCTTCTCGTGACGAGT; TS-GAGAGArUA5’Cy3,
/Cy3/ACTCGTCACGAGAGArUAAGCCACGGTATTTCAAA; NTS-CTCTCTATTCTTCTTCTCGTGACGAGT; TS-GAGAGArUA5’Cy3,
/Cy3/ACTCGTCACGAGAGArUAAGCCACGGTATTTCAAA; NTS-CTCTCTATTCTTCTTCTCGTGACGAGT; TS-GAGAGArUA5’Cy3,
/Cy3/ACTCGTCACGAGAGArUAAGCCACGGTATTTCAAA; NTS-CTCTCTATTCTTCTTCTCGTGACGAGT; TS-GAGAGArUA5’Cy3,
The 5'-32P-label was introduced on oligonucleotides using [γ-32P] ATP and T4 Polynucleotide kinase, while the 3'-32P-label was by incubation with [α-32P] dATP and terminal deoxynucleotide transferase under manufacturers’ conditions. Labeled oligonucleotides were hybridized with a two-fold excess of the relevant complimentary oligonucleotide.

**Top1 overexpression and purification**- Top1 overproduction was carried out in *S. cerevisiae* strain FM113 (MATa ura-3-52 trp1-289 leu2-3112 prb1-1122 prc1-407 pep4-3) transformed with plasmid pRS425-GGT-TOP1. Growth, induction, extraction preparation, and ammonium sulfate precipitation (0.3 g/ml) were similar to the procedures described previously (Bylund, Majka et al. 2006). The ammonium sulfate precipitate was resuspended in buffer A(o) (buffer A: 60 mM HEPES-NaOH [pH 7.4], 10% glycerol, 1 mM dithiothreitol (DTT), 1 mM EDTA, 0.01% polyoxyethylene (10) lauryl ether, 1 mM sodium bisulfite, 1 µM pepstatin A, 1 µM leupeptin; subscript indicates the mM sodium chloride concentration), until the lysate conductivity was equal to that of buffer A(400). The lysate was then used for batch binding to glutathione-Sepharose 4B beads (GE Healthcare), equilibrated with buffer A(400), and gently rotated at 4 °C for two hours. The beads were collected at 1,000 rpm in a swinging-bucket rotor, followed by batch washes (3 × 20 ml of buffer A(400)). The beads were transferred to a 10 ml column, and washed at 2.5 ml/min with 100 ml of buffer A(400). The second washing was with 50 ml buffer A(400) containing 5 mM Mg-acetate and 1 mM ATP. And the third washing used 50 ml of buffer A(400) and 30 ml of buffer A(200). Elution was carried out at a flow rate of 0.2 ml/min with buffer A(150) containing 30 mM glutathione (pH adjusted to 8.1). Fractions containing Top1 were combined and incubated overnight at 4 °C with 30 U of rhinoviral 3C protease. The following day the Top1 protein was loaded on a heparin column in buffer A(150) without protease inhibitors. The column was washed with 10 column volumes of buffer A(300), the protein was eluted with buffer A(750). Fractions containing pure Top1 were collected and dialyzed overnight to A(200) without protease inhibitors.

**Protein overexpression and purification**- Trl1, Tpp1, Apn1, and Apn2 were purified in an identical manner. Overproduction was carried out in *S. cerevisiae* strain FM113 (MATa ura-3-
52 trp1-289 leu2-3112 prb1-1122 prc1-407 pep4-3) transformed with plasmid BG1805 containing TRL1, TPP1, APN1, or APN2 respectively. Growth, induction, and extraction preparation were similar to the procedures described previously (Bylund, Majka et al. 2006).

Extraction and wash buffer B300 except extraction buffer used HEPES-NaOH [pH7.8](Buffer B: 50 mM HEPES-NaOH [pH 7.4], 10% glycerol, 5 mM dithiothreitol (DTT), 4 mM MgCl, 1 mM EDTA, 0.02% NP40, 1 mM sodium bisulfite, 1 µM pepstatin A, 1 µM leupeptin; subscript indicates the mM sodium chloride concentration). Cell extracts were cleared by centrifugation at 18000K for 30 min. The cleared cell extracts were batch bound with B300 equilibrated Igg beads () rotating at 4 °C for two hours. The beads were then collected and washed with twenty column volumes of B300. Followed by a second wash with buffer B300 containing 5 mM Mg-acetate and 1 mM ATP. The beads were then equilibrated with cleavage buffer (50 mM HEPES-NaOH [pH 7.4], 10% glycerol, 1 mM dithiothreitol (DTT), 1 mM EDTA, 0.01% NP40, 500 mM NaCl). The beads were then rotated overnight in one column volume of cleavage buffer with 30U of rhinoviral 3C protease at 4 °C. The following day the supernatant was eluted from the beads and a second cleavage step was carried in one column volume of cleavage buffer.

**rNMP repair assay**- The standard 10 µl assay mixture contains 100 mM Tris-HCl (pH 7.8), 500 µg/ml bovine serum albumin, 5 mM DTT, 5 mM Mg-acetate, 50 mM NaCl, 100 fmol of 32P-end-labeled oligonucleotide substrate or 250 fmol Cy3-end-labeled oligonucleotide, and enzyme. Incubations were carried out at 30 °C for the indicated time periods. Deviations for the standard assay conditions are indicated in the legends of the figures. Reactions were stopped with stop buffer containing 10 mM final concentration of EDTA, 0.05% SDS, and 40% formamide. Samples were analyzed on a 17% 7 M urea-PAGE electrophoresis. After the gels were dried they were subjected to phosphorImager analysis, or directly imaged for Cy3-signal.

Repair assays were generally carried out in two stages. In the first stage, the rNMP containing substrate was treated with two-fold excess concentration of Top1 to produce a mixture of full length substrate, rNMP 2’, 3’-cyclic phosphate nicked substrate, and Top1-covalently linked substrate. The reactions were stopped with stop buffer as previously described. To purify the 2’, 3’-cyclic phosphate containing substrate the DNA was ethanol precipitated and used for subsequent reactions. The Top1-covalent linked substrate was purified by first treatment with proteinase K at 50 °C for 30 minutes followed by phenol-chloroform extraction and ethanol precipitation. The replication substrates were purified as previously stated, but then the ssDNA was purified by 7M-urea-PAGE purification. The oligo was then hybridized to NTS-GCT-Biotin template strand.
REFERENCES


LEGENDS TO FIGURES

Figure 1. Top1 is an endoribonuclease forming a 3'-terminal 2', 3'-cyclic phosphate followed by formation of Top1-cc.
(A) Activity of Top1 on dsDNA with single rNMP. Standard assay mixtures used 10 nM 5’-32P-labeled 32-oligonucleotide in dsDNA substrate. The concentrations of Top1 from left to right are 1.5, 5, 15, 50, 150 nM. The reactions were carried out at 30 °C for 10 min. The results were analyzed on a 7 M urea-17% polyacrylamide gel. (B) 5’-cy3-labeled ssDNA with a single-rNMP in dsDNA substrate was treated with Top1 under standard assay conditions for 1 hour at 30 °C. The DNA containing 2’, 3’-cyclic phosphate terminated nick was then isolated and purified, and the DNA was incubated in a second reaction with 25 nM DNA substrate supplemented with T4 polynucleotide kinase, alkaline phosphatase, Tpp1 (25 nM), Trl1 (25 nM), or in combination as indicated for 10 minutes at 30 °C, except T4 PNK treatment was 60 minutes at 37 °C. The results were analyzed on a 7 M urea-17% polyacrylamide gel. (C) Formation of Top1-cc on rNMP containing dsDNA substrate. Standard assay mixtures used 25 nM 5’-cy3-labeled 32-nucleotide ssDNA oligonucleotide containing a single rNMP in dsDNA substrate. The concentration of Top1 used was 100 nM. The reactions were carried out at 30 °C for times indicated. The reactions were results were analyzed by 11% PAGE. (D) and (E) Top1 first forms 2’, 3’-cyclic phosphate prior to formation of Top1-cc. Standard assay mixtures used 25 nM 5’-cy3-labeled 32-nucleotide ssDNA oligo containing a single rNMP in dsDNA substrate. The concentration of Top1 was 100 nM. The reactions were carried out at 30 °C for times indicated. The reactions were then stopped with 10 mM EDTA and 0.5% SDS and treated with proteinase K to release ssDNA from Top1 protein. The reactions were results were analyzed by 7 M urea-17% polyacrylamide gel. (E) Quantitation of gel in (D) left y-axis represents percent cyclic phosphate and the right y-axis represents percent Top1-cc formation.

Figure 2. Top1-cc formed by Top1 cleavage two-nucleotides to the 5'-side of 2’, 3’-cyclic phosphate. (A) 5’-cy3-labeled ssDNA with a single-rNMP in dsDNA substrate was treated with Top1 under standard assay conditions for 1 hour at 30 °C. The DNA containing 2’, 3’-cyclic phosphate terminated nick and Top1-cc was then treated with proteinase K at 42 °C for 30 minutes followed by isolation and purification. Then DNA was incubated in a second reaction with 25 nM DNA substrate supplemented with Tdp1 (25 nM), Tpp1 (25 nM), Apn1 (25 nM), Apn1-D192 (nM), or Apn2 (25 nM) or in combination as indicated for 10 minutes at 30 °C. The results were analyzed on a 7 M urea-17% polyacrylamide gel. (B) Our model for the formation of 2’, 3’-cyclic phosphate terminated nick dsDNA, and subsequent formation of Top1-cc. See text for details.

Figure 3. Top1 leads to deletions when rNMP is in the context of repetitive DNA sequences. (A) The Top1 dependent deletions are observed through the religation activity of Top1 on dsDNA with single rNMP either internal or proximal to di-nucleotide repeat sequence. Standard assay mixtures used 25 nM 5’-cy3-labeled oligonucleotide with a single rNMP either internal or proximal to a GA or TA di-nucleotide repeat hybridized to its template strand substrate. The concentration of Top1 used was 250 nM. The reactions were carried out at 30 °C for the time indicated. The results were analyzed on a 7 M urea-17% polyacrylamide gel. (B) Graphic representation of the quantification of the four products of the Top1 religation reactions. The quantitation of the rGA reaction from (A) is represented in solid curves, where the quantitation of the rCA reaction from (A) is represented in dashed curves. The 34-mer is the starting substrate (green), where the 32-mer is the religation product (red), the Top1-DNA is the Top1-cc (blue), and finally the 16-mer represents the cyclic-phosphate product formed by Top1 cleavage at the ribonucleotide (black). The rGA and rCA 16-mer curves were identical and combined for clarity.
purposes. (C) The rNMP is removed in the Top1 religation products. Standard assay mixtures used 25 nM 5’-cy3-labeled oligonucleotide with a single rNMP internal GA di-nucleotide repeat hybridized to its template strand substrate. The concentration of Top1 used is 250 nM. The reactions were carried out at 30 °C for the time indicated. Samples as indicated were treated with 0.3 M NaOH at 50 °C for 30 minutes. Then the samples were neutralized with 0.3 M acetic acid. The results were analyzed on a 7 M urea-17% polyacrylamide gel.

**Figure 4. Reconstitution of the Tdp1-dependent pathway of Top1-cc resolution.** 5’-cy3-labeled 34-nucleotide oligonucleotide with a single-rNMP was hybridized with ssDNA template. The substrate was treated with Top1 under standard assay conditions for 1 hour at 30 °C. The DNA containing Top1-cc was then treated with proteinase K at 42 °C for 30 minutes followed by isolation and purification by 7 M urea-17% polyacrylamide gel. Then DNA was then hybridized to a 64-nucleotide template contain both 5’ and 3’ biotin bumpers was incubated in a second reaction with 25 nM DNA substrate supplemented with streptavidin (100 nM), PCNA (50 nM), RFC (100 nM), RPA (50 nM), 100 μM ATP as well as Tdp1 (25 nM), Tpp1 (25 nM), or in combination as indicated for 10 minutes at 30 °C either in the presence of absence of 50 nM Polδ. Samples treated with 0.3 M NaOH at 55 °C for 30 minutes and quenched with acetic acid 0.3 M are indicated. The results were analyzed on a 7 M urea-17% polyacrylamide gel.

**Figure 5. Model for Top1-dependent rNMP removal either in the context of di-nucleotide repeat sequence or in non-repetitive DNA.** See text for details.

**Supplemental Figure 1.** Top1 cleavage is dependent on the presence of rNMP in dsDNA and the cleavage leaves 5’-hydroxyl terminus. Activity of Top1 on a 3’-labeled single-rNMP containing dsDNA substrate. Standard assay mixtures used 10 nM 3’-32P-labeled 32-nucleotide dsDNA substrate either with or without single rNMP in labeled strand. The concentration used in the assay for Top1 50 nM and RNase H2 1 nM. The reactions were carried out at 30 °C for 10 min. The results were analyzed on a 7 M urea-17% polyacrylamide gel.

**Supplemental Figure 2.** Apn1 and Apn2 are capable of cleaving phospho-tyrosyl bond of Top1 peptides. 5’-cy3-labeled ssDNA with a single-rNMP in dsDNA substrate was treated with Top1 under standard assay conditions for 1 hour at 30 °C. The DNA containing 2’, 3’-cyclic phosphate terminated nick and Top1-cc was then treated with proteinase K at 42 °C for 30 minutes followed by isolation and purification. Then DNA was incubated in a second reaction with 25 nM DNA substrate supplemented with Tdp1 (25 nM), Tpp1 (25 nM), Apn1 (50 nM), or Apn2 (50 nM) or in combination as indicated for 10 minutes at 30 °C. The results were analyzed on a 7 M urea-17% polyacrylamide gel.
Figure 5

Supplemental Figure 1
Supplemental Figure 2
CHAPTER V

Future aims and directions
Exonuclease 5

Exonuclease 5 was first discovered in S. cerevisiae by Peter Burgers, over twenty-five years ago, as an unknown gene product that cleaved ssDNA leaving a di-nucleotide as a major product (Burgers, Bauer et al. 1988). There were no other investigations into what gene encoded EXO5, and no biochemical studies had been carried out. Chapter II presented here identified EXO5 to be encoded by the gene DEM1 (Defects in Morphology) in S. cerevisiae (Burgers, Stith et al. 2010). Deletion of DEM1 led to defects in the morphology of the mitochondria, and caused respiratory loss in the cells (Entian, Schuster et al. 1999; Dimmer, Fritz et al. 2002). DEM1 homologs were identified in other species. The work presented in this dissertation describes our work to elucidate the biochemical activity of this protein family and its role in DNA metabolism. All of the Exo5 homologs tested are bi-directional single-strand DNA specific exonucleases. The S. cerevisiae homolog is essential for mitochondrial genome maintenance, but does not play an apparent role in the nucleus. This strong mitochondrial localization seems to be restricted to the Saccharomycetales order. The human homolog is a nuclear and cytoplasmic protein that is important for nuclear genome maintenance. The S. pombe homolog is important for both nuclear and mitochondrial genome maintenance in a redundant role with Rad2 (FEN1). There are many questions that still need to be answered, both about the biochemical activity of these proteins and their role in DNA metabolism.

Specific questions that need to be addressed

S. pombe

The studies in S. pombe are still in progress and we are preparing a manuscript for publication. There are several questions that need attention that I will address here. First, the protein-protein interactions found from the co-immunoprecipitation of spExo5. There are several interactions that
we have not characterized including with two 14-3-3 proteins, Rad24 and Rad25, as well as an interaction with the RuvB like proteins Rvb1 and Rvb2. The importance of these interactions still needs to be investigated. We also set out to further investigate the spExo5 RPA and PCNA interactions by two-hybrid analysis. The results have been unclear. It is possible that the introduction of the N-terminal Gal4 domains is occluding the interaction surfaces. We plan to change the tag to the C-terminus to test this possibility. It is also possible that the interaction is mediated by some post-translation modification that does not occur when these proteins are expressed in *S. cerevisiae*. Our two-hybrid analysis has yielded one interesting possible interaction between spExo5 and the covalent modification Sumo. This result is only preliminary and more work will be needed to further investigate this possible interaction. spExo5 has several possible Sumo interaction motifs, but we still need to carry out more experiments to test this hypothesis.

**Future Directions**

The first area of discussion is of future investigation into the biochemical activities of Exo5 family members. One aspect that needs to be further examined is the sliding activity of the human and *S. pombe* homologs. The kinetics of sliding seems to be greater than the kinetics of cleavage. We have begun a collaboration with the lab of Tim Lohman to use single molecule approach to investigate the sliding of the human homolog of Exo5 similar to his lab studies of *E. coli* SSB sliding along ssDNA (Zhou, Kozlov et al. 2011). It is still not known what function the sliding activity has *in vivo*, or if it is regulated in cells by either post-translation modifications or protein-protein interactions. Another aspect of human homolog is to determine why the activity is approximately 100-fold lower than the *S. pombe* or *S. cerevisiae* homologs. It is possible that there are post-translational modification(s) that are required to activate the human protein *in vivo*.
that are not present in our purified protein, or again it is possible there are protein interactions that activate the nuclease activity.

One interesting aspect of the Exo5 family of proteins is the presence of a 4Fe-4S iron-sulfur cluster complex. The presence of the iron-sulfur complex is important for the nuclease activity of these proteins. Use of iron-sulfur complexes as structural motifs in proteins is an ancient mechanism for stabilizing protein structures, but recently a hypothesis has been put forward that some iron-sulfur clusters are redox active and this redox activity is important for the protein function in vivo (Nunez, Holmquist et al. 2001). There is a growing list of DNA metabolism proteins that contain iron-sulfur clusters including the replicative polymerases, many helicases, and helicase-nuclease (Wu and Brosh 2012). It will be interesting to investigate if the 4Fe-4S complex of the Exo5 family is redox active and if the biochemical properties of the enzyme changes in different redox states. These are some of the questions that should be addressed as to the biochemical function of the Exo5 family nucleases.

There are also many questions about the in vivo function of the Exo5 family of proteins. First, in S. cerevisiae how Exo5 functions in the mitochondria is still not known. One caveat to trying to understand the role of Exo5 in mitochondrial genome maintenance is that very little is known in general about mitochondrial replication and repair in S. cerevisiae.

The studies in S. pombe are still in progress and we are preparing a manuscript for publication. There are several questions that need attention that I will address here. First, the protein-protein interactions found from the co-immunoprecipitation of spExo5. There are several interactions that we have not characterized including with two 14-3-3 proteins, Rad24 and Rad25, as well as an interaction with the RuvB like proteins Rvb1 and Rvb2. The importance of these interactions still needs to be investigated. We also set out to further investigate the spExo5
RPA and PCNA interactions by two-hybrid analysis. The results have been unclear. It is possible that the introduction of the N-terminal Gal4 domain is occludes the interaction surfaces. We plan to change the tag to the C-terminus to test this possibility. It is also possible that the interaction is mediated by some post-translation modification that does not occur when these proteins are expressed in *S. cerevisiae*. Our two-hybrid analysis has yielded one interesting possible interaction between spExo5 and the covalent modification Sumo. This result is only preliminary and more work will be needed to further investigate this possible interaction. spExo5 has several possible Sumo interaction motifs, but we still need to carry out more experiments to test this hypothesis.

The function of Exo5 in *S. pombe* seems to have a dual function in both mitochondrial and nuclear genome maintenance; although, in *S. pombe* Exo5 is not essential and we have shown that there is functional redundancy with Rad2. There also seems to be a difference in the phenotype of the loss of Exo5 on the mitochondrial DNA. In *S. cerevisiae exo5Δ* leads to large deletions of parts of the mitochondrial genome, whereas in *S. pombe exo5Δ rad2Δ* leads to depletion of the mitochondrial genome without gross deletion(s) and expansion of the DNA. It is possible that the replication mechanisms are different between these two species, and that may explain the difference in phenotype of *exo5Δ*, but very little work has been carried out to describe the mitochondrial replication mechanism in yeast. It will be important to determine if the mitochondrial or nuclear functions of Exo5 and Rad2 are responsible for the synergistic sensitivity to DNA damaging agents. Our data now seem to indicate that both the nuclear and mitochondrial functions are important in that neither the nuclear nor the mitochondrial isoform complemented the *exo5Δrad2Δ* hypersensitivity to either MMS or cis-platin. To further test the
requirement of both functions I will co-express both isoforms in an exo5Δrad2Δ strain to see if co-expression allows for complementation.

The nuclear function of Exo5 in *S. pombe* is also still an area that will need to be a focus of future work. It will be important to continue to expand the epistasis analysis of the exo5Δ with other known DNA repair proteins. We have evidence that spExo5 epistatic with the Fan1-Fml1 pathway, but we still have not tested the possibility that Exo5 is important for the Pli1 pathway of ICL repair. I am currently making the Pli1 deletion strains. Finally, the overexpression of *S. pombe* Exo5 leads to checkpoint activation and cell death. We hypothesize the interaction of Exo5 with RPA is important for this phenotype. It will be of interest to determine how RPA influences Exo5 activity *in vitro*, and to understand why the Exo5-RPA interaction causes checkpoint activation.

Many aspects of how the human homolog works *in vivo* still to be determined. Initially, it will be interesting to investigate if the human protein also has a mitochondrial function as is seen in *S. pombe*. We have not carefully investigated the localization of Exo5 in human cells. Also, in collaboration with Zhongsheng You, we have examined the recruitment of Exo5-GFP to site directed laser stripping (data not shown). The GFP tagged protein localization was variable from cell to cell with some cells having nuclear Exo5 and other only cytoplasmic. After laser stripping, Zhongsheng You saw accumulation of Exo5-GFP at the site of cutting, but the kinetics was very slow with accumulation peaking at around 60 minutes after cutting. Several questions need to be addressed; first, is Exo5 localization cell cycle dependent or is it DNA damage dependent? It will be of interest to determine the mechanism of recruitment and nuclear import; the pathway that Exo5 is involved in human cells is still unclear.
There are many areas and many questions that have yet to be explored, and these questions could lead to important insight into genome stability and disease. This thesis outlines my work on this novel exonuclease family from yeast to humans.

**RNase H2-initiated RER**

The consequence of the disparity in the rNTP vs dNTP concentrations in the cell (rNTP~100 fold higher than dNTP) has only recently been appreciated. The high rNTP concentration leads to large amounts of rNMPs being incorporated into the genome by replicative polymerases *in vitro* and *in vivo* (Nick McElhinny, Kumar et al. 2010; Nick McElhinny, Watts et al. 2010; Sparks, Chon et al. 2012). RNase H2 is able to cleave at sites of RNA in a RNA/DNA hybrid. It has the ability to cleave both at sites of a single-ribonucleotide or in long stretches of ribonucleotide such as R-loops (Eder, Walder et al. 1993; Murante, Henricksen et al. 1998). The work presented in Chapter IIIa of this dissertation outlines the reconstitution of the RNase H2-initiated ribonucleotide excision repair pathway, and created a system to investigate the rate of rNMP incorporation by replicative polymerases using a large plasmid based method to decrease sequence bias. In this reconstitution we find several redundant steps in the repair pathway. Both Polδ and Polε are able to carry out strand displacement synthesis with reasonable efficiencies, and both FEN1 and Exo1 are capable of flap removal during strand displacement synthesis in this system. There are still questions about how this pathway works *in vivo*, and other functions it may play in genome maintenance and disease.

**Future Directions**

One important reason for studying the RNase H2-initiated RER pathway comes from the finding that mutations in RNase H2 in humans causes disease Aicardi Goutières syndrome (Crow, Leitch et al. 2006). Aicardi Goutières is a systemic autoimmune childhood disease that is
triggered by IFN (Type I interferon) activation in response to endogenous nucleic acids, usually from viral infection (Aicardi and Goutieres 1984; Kawai and Akira 2006). RNase H2 is essential for development in mice, and the patients with mutations in RNase H2 have decreased but not complete loss of activity (Crow, Leitch et al. 2006; Reijns, Rabe et al. 2012). Though it has not been investigated whether patients with these RNase H2 mutations have increased genomic rNMPs, it will be interesting to determine if increases in genomic rNMPs plays a role in the disease. Testing patient genomic DNA compared to controls sample to determine the approximate level of rNMPs remaining in their genomes by treatment with NaOH or RNase H2 will be important to understand if genomic rNMPs contribute to the pathology of AGS. There evidence from the RNase H2 knockout mice that removal of genomic rNMP may not be the function of RNase H2 that is causative of the autoimmunity mediated through INF activation. The RNase H2-knockout mouse embryos did not have the characteristic activation of INF that is seen in other models of the disease in mice (Hiller, Achleitner et al. 2012; Reijns, Rabe et al. 2012). One possible explanation is that the IFN activation phenotype occurs later in development, and is not seen because of the early lethality caused by deletion of RNase H2. To test this hypothesis it will require making mutants of RNase H2 that mimic the mutations in human patients rather than using a knockout. It will also be of interest to determine which activity of RNase H2 is responsible for preventing the activation of IFN, cleaving a single rNMP or at sites of multiple rNMPs. Using the knowledge gained from our study on the unlinking of RNase H2 activities may allow for make mouse knock-in RNase H2 mutants that have only single rNMP cleavage activity or only multiple rNMP cleavage activity. It may be the function of cleavage of multiple ribonucleotides that may mimic substrates produced during viral infection.
that prevents the activation of IFN in vivo. There are many aspects of RNase H2 function in mammalian cells that are still to be elucidated.

**Top1-initiated Ribonucleotide repair**

Both human and vaccina viral Topoisomerase I homologs have endoribonuclease activity in vitro against RNA in RNA/DNA hybrids (Sekiguchi and Shuman 1997; Kim, Huang et al. 2011). In *S. cerevisiae* RNase H2 deficient strains show a particular mutagenesis spectrum in which there is a significant increase in 2-5 bp deletions at sites of di-nucleotide repeats (Nick McElhinny, Kumar et al. 2010; Kim, Huang et al. 2011; Williams, Smith et al. 2013). Topoisomerase I is responsible for causing these 2-5 bp deletions in the absence of RNase H2 (Kim, Huang et al. 2011). Also, Top1 is able to remove a small portion of ribonucleotides from the yeast genome in the absence of RNase H2 (Williams, Smith et al. 2013). The data we present in Chapter IV outlines our in vitro reconstitution of Top1 dependent removal of rNMPs from dsDNA. Also Chapter IV outlines the first in vitro biochemical reconstitution of Top1-mediated 2-5 bp deletions observed in the absence of RNase H2.

**Future Directions**

There have been many in vivo studies investigating the formation of the 2-5 bp deletions that are dependent on Top1 in the context of genomic rNMPs and in transcription-associated mutagenesis (TAM)(Cho, Kim et al. 2013). In our in vitro system when the rNMP is in the context of a di-nucleotide repeat sequence Top1 is able to lead to a two-nucleotide deletion. The deletion product is resistant to NaOH treatment supporting our hypothesis that the rNMP is removed by the Top1-dependent deletions, but we still need to purify and sequence the resulting
oligonucleotide to gain more evidence about where the deletion occurs in our rNMP-containing strand.

The role of Top1 in repair of rNMPs seems to be a back up system that is only observed in the absence of RNase H2, but it is possible that at low levels Top1 may attack remaining genomic rNMPs in wild-type cells. Top1 attack of an rNMP forming a cyclic phosphate terminated nick at the rNMP blocks the ability of RNase H2 to act (data not shown). Further, Top1 mutagenesis may play a role in the detrimental effect of RNase H2 mutations in AGS. It would be interesting to investigate if AGS patients have increased deletions at sites of di-nucleotide repeats in their genome.

It will also be important to investigate what pathways contribute to the repair Top1-DNA complexes that are not resolved by religation of the DNA. There are several proteins that have been implicated in repair including Tdp1, the NER endonuclease Rad1-Rad10 (XPF), and Mus81-mus7 endonuclease (Vance and Wilson 2001). It may be possible that these Top1-cc repair pathways help to prevent Top1- religation events that lead to increased mutagenesis. Work from this thesis outlined a mechanism for Top1-DNA repair that requires the Top1 protein to be partially degraded prior its removal by Tdp1, but the mechanism for this degradation *in vivo* is not well understood. The proteins that are responsible for Top1 degradation are not known in yeast and need to be investigated further. Several publications implicate the proteosome in Top1-cc degradation in mammals, but it has not been investigated in yeast (Lin, Ban et al. 2009; Lee-Kirsch, Wolf et al. 2013). These are some of the questions that still need to be investigated.

**Summary**

The work presented in this dissertation focuses on the role of nucleases in DNA metabolism. First, it describes the biochemical characterization of a novel nuclease family,
Exonuclease 5. Exo5 members all contain an iron-sulfur cluster and are bi-directional ssDNA specific exonucleases. The *S. cerevisiae* homolog is essential for mitochondrial genome maintenance, but lacks an apparent nuclear function. The *S. pombe* homolog is both nuclear and mitochondrial, and is subsequently important for both nuclear and mitochondrial genome maintenance, maintaining proper mitochondrial DNA levels, and for repair of DNA damage. *S. pombe* Exo5 interacts with both PCNA and RPA *in vivo*. The human homolog also physically interacts with the human RPA complex, and is important for nuclear genome maintenance and genome stability.

The second area of this dissertation discusses our work on the incorporation of rNMPs into the genome by replicative polymerase and their subsequent removal by two pathways. First, the ribonucleotide excision repair pathway that is initiated by RNase H2 cleaves at the rNMP, allowing loading of PCNA by the RFC complex that allows strand displacement synthesis by either Polδ or Polε. The resulting ssDNA flap containing the ribonucleotide is removed by either FEN1 or Exo1 and the DNA is ligated by DNA ligase I. The other pathway for rNMP removal is dependent on Top1 and can lead to small deletions if the rNMP is in the context of a di-nucleotide repeat sequence, If the rNMP is outside the context of a repeat sequence it must be repair by a pathway of Top1-DNA covalent complexes removal.
REFERENCES


