Ribonuclease H Function in *Bacillus subtilis*

by

Justin R. Randall

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Molecular, Cellular and Developmental Biology)
in The University of Michigan
2018

Doctoral Committee:

Associate Professor Lyle A. Simmons, Chair
Professor Ursula Jakob
Assistant Professor Jayakrishnan Nandakumar
Professor Nils Walter

Justin R. Randall

justrand@umich.edu

ORCID iD: 0000-0002-5429-8995

DEDICATION

For my mother and father. Without your love and support this document, and come to think of it I myself, would not exist.

ACKNOWLEDGEMENTS

I've read that Oscar Wilde once said, "Success is a science; if you have the conditions, you get the result." Oscar Wilde wasn't a scientist, but in my humble opinion Lyle Simmons sets up pretty damn favorable conditions. From the moment I first met Lyle I knew I wanted to work in his lab. After introducing myself to him outside his office he said to me while I shook his hand, "I need to go to the bathroom want to come with me?" If I could go back I would have responded with a confident, "Sure." But instead I awkwardly replied, "Umm…I think I'll wait." That was an awesomely odd first encounter which blossomed into a wonderful mentor/mentee and social relationship. If I were to try to engineer a dream Principle Investigator (PI), they would be to Lyle's specs. I really could not have asked for a better mentor to lead me through my Ph.D. Thank you so much for all of your help and support throughout the last 5 years. I hope to one day develop your kindness, patience, humor, and scientific mind.

To everyone else past and present who have called the Simmons lab home, I thank you for your shared knowledge, advice, and entertainment. Lenhart, I have never met someone who works harder than you. Brian, I leaned on your friendship and humor to get through rough science times. Jeremy, your mentorship and patience meant so much to my development as a scientist. Pete,

I’ve never had a more confident and knowledgeable co-worker. Lindsay and Taylor, thank you for helping to keep me sane; whether through shared sushi and crossword solving or providing supportive advice when experiments weren’t going well. Katie and Lieke, though our time together was more brief, you both have very bright minds and I’m sure you will both be successful in the future. To Amy and Heather, thank you for putting up with my sarcasm and terrible humor. It was a pleasure having the opportunity to teach and spend time with you both. You are both terrific people and scientists. Good luck in your future endeavors!

To my thesis committee, I can’t thank you enough for giving me your time and feedback. I know you are all very busy and I very much appreciate you making time to give me valuable feedback on my work and helping me stay on the correct path to complete my degree. I hope this document meets your expectations.

To my friends, the last five and a half years would not have been nearly as fun nor gone by so fast without each of your friendships. Thank you all for the good times, too many drinks, disc golf, brewery tours, and general tomfoolery. I wish you all success in your new careers and hope to stay in touch throughout the years to come.

To my family, I would have never made it this far without each of you. I became the person I am because of our shared experience, and I wouldn’t change a single thing about it. I love you all and hope I have made each of you proud.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

ABSTRACT

RNA is commonly found included in chromosomal DNA forming RNA-DNA hybrids. RNA becomes embedded in DNA through DNA polymerase errors, Okazaki fragments, or annealed to DNA in the form of R-loops. Efficient processing of RNA-DNA hybrids is critical for cell survival and genome stability. RNase H enzymes are responsible for recognizing RNA-DNA hybrids and hydrolyzing the RNA containing strand. In this dissertation, I show that single ribonucleotides incorporated as DNA polymerase errors are corrected by RNase HII in a process known as ribonucleotide excision repair (RER). I show that RNase HII from *B. subtilis* cleaves 5' to single rNMPs embedded in DNA and that Pol I efficiently extends from an RNase HII processed substrate, reconstituting the minimal set of proteins for RER on a linear substrate *in vitro*. To determine the mutagenic cost that occurs in the absence of RNase HII (*rnhB)*, mutation accumulation lines were completed demonstrating a 2-fold increase in GC → AT transitions in a strand- and sequence-context dependent manner. Using purified proteins, I demonstrate that DnaE can access a gap but not an RNase HII-dependent nick and that DnaE is ~2-fold more mutagenic than Pol I when replicating over the 3'-GCC(C/T)T-5' sequence context identified as mutagenic *in vivo*. This work suggests that in the absence of RNase HII a secondary pathway

removes the ribonucleotide, creating a gap allowing for DnaE access and mutagenesis. To understand how RNase HII and HIII activity is regulated, I measured the activity of each RNase H enzyme on several different RNA-DNA hybrid substrates *in vitro*. I show that although RNase HII and HIII are capable of incising all RNA-DNA hybrids tested, the activity of RNase HII and RNase HIII is dependent on the specific divalent metal ion available to the enzyme *in vitro*. I demonstrate that RNase HIII from three Gram-positive bacteria are proficient for cleavage at single rNMPs embedded in DNA and that an RNase HIII nick can facilitate Pol I extension. Importantly, I show that under physiologically relevant Mg^{2+} and Mn^{2+} concentrations RNase HII efficiently cleaves RNA-DNA hybrids containing RNA-DNA junctions while RNase HIII efficiently cleaves junction-less hybrids such as R-loops. Lastly, I identify a striking sensitivity of RNase HIII deficient cells to the chemotherapeutic agent hydroxyurea (HU). Further, expression of RNase HII does not rescue the RNase HIII deficient phenotype, demonstrating different functions for these RNase H proteins *in vivo*. In this work, I conclude that RNase HII and Pol I are responsible for RER while RNase HIII is critical for R-loop resolution *in vivo*. Based on these results, I suggest that substrate specificity of RNase HII and HIII is regulated *in vivo* by intracellular divalent metal ion concentrations, dictating the RNA-DNA hybrids they act upon in *B. subtilis*.

CHAPTER I

I. Introduction

1.1 DNA/RNA structure

Deoxyribonucleic acid (DNA) is the fundamental genetic molecule for all cellular life. The structure of DNA not only makes it extraordinarily well suited for the storage of genetic material, but remarkably easy to replicate and pass on to future generations (Kornberg and Baker, 1992). DNA is composed of polymers of deoxyribonucleoside monophosphates (dNMPs). Each dNMP contains three molecular units: a phosphate, a deoxyribose sugar, named so because of the missing oxygen atom at the 2' position, and one of four nitrogenous bases. The bases in DNA include: adenine (A), thymine (T), guanine (G), or cytosine (C) (Kornberg and Baker, 1992). The structure of ribonucleic acid (RNA) is remarkably similar. It differs atomically from DNA by only one oxygen atom attached to the 2' carbon of the ribose sugar (Figure 1.1A). That one atom causes RNA to be over 100,000 times more likely to spontaneously hydrolyze than DNA (Thompson et al., 1995) making RNA much less well suited for the long-term storage of genetic information. Instead RNA plays several other pivotal roles including acting as the transient intermediate between DNA and protein

synthesis. The remarkable structural similarity between the two molecules, and their ability to hydrogen bond with one another, creates several problems *in vivo* resulting in the formation of RNA-DNA hybrid molecules (Figure 1.1B). RNA-DNA hybrid molecules have greater stability than double-stranded DNA (dsDNA) (Roberts and Crothers, 1992), making RNA-DNA hybrids occur commonly *in vivo*. In addition, RNA is covalently incorporated into genomic DNA during replication causing genome instability if left unrepaired (Figure 1.2C). Here we discuss the causes of both covalently and non-covalently linked RNA-DNA hybrids *in vivo,* what is known about how these hybrids form, and how they are resolved.

1.2 rNTP incorporation by DNA polymerases

During DNA replication, DNA polymerases must incorporate deoxyribonucleoside triphosphates (dNTPs) into newly synthesized DNA rapidly and DNA polymerases need to avoid using ribonucleoside triphosphates (rNTPs) in place of dNTPs. DNA polymerase fidelity with respect to rNTP/dNTP ratios has been well studied *in vitro* (Brown and Suo, 2011; Joyce, 1997). Discrimination between sugars by replicative and some non-replicative DNA polymerases can be attributed to a large steric gate residue: usually a tyrosine, phenylalanine, or glutamic acid (Brown and Suo, 2011; Joyce, 1997). The steric gate residue is located in the DNA polymerase active site and physically clashes with the 2' OH of the ribose sugar in rNTPs limiting their incorporation during DNA polymerase extension. (Astatke et al., 1998; Beck et al., 2002; Bonnin et al., 1999; Brown et

al., 2010; Cases-Gonzalez et al., 2000; DeLucia et al., 2003; Gao et al., 1997; Gardner and Jack, 1999; Kasiviswanathan and Copeland, 2011; Patel and Loeb, 2000; Yang et al., 2002). DNA polymerases that contain a steric gate residue show high discrimination for dNTPs over rNTPs with selectivity as high as a million-fold. These more selective polymerases are only likely to misincorporate single ribonucleoside monophosphates (rNMPs) (Figure 1.1B). Conversely, there are several low fidelity non-replicative DNA polymerases, which lack a steric gate residue causing little to no rNTP/dNTP selectivity (Ordonez et al., 2014). A common hypothesis is that DNA polymerases incapable of discriminating between rNTPs and dNTPs could synthesize patches of RNA during repair or during other stress conditions such as lowered dNTP pools (Ordonez et al., 2014) (Figure 1.1B).

The task of dNTP/rNTP DNA polymerase selectivity is further complicated by the difference in cellular nucleotide concentrations. *In vivo* rNTP concentrations vastly exceed that of dNTPs in both prokaryotes and eukaryotes (Figure 1.2A). Ribonucleotide Reductase (RNR) is responsible for converting rNTPs into the dNTPs later used for DNA replication [for review see (Torrents, 2014). In *E. coli* nucleotide ratios range from 1.8-fold (rUTP/dTTP) to as high as 600-fold (rATP/dATP) (Buckstein et al., 2008). In *S. cerevisiae* ratios range from 36-fold (rCTP/dCTP) to 190-fold (rATP/dATP) (Nick McElhinny et al., 2010a) similar differences in nucleotide pools exist in other organisms including mammalian cells (Traut, 1994). Below, I discuss in detail the relevant DNA

polymerase contributions to single rNMP incorporations and embedded patches of RNA in genomic DNA.

1.2.1 Replicative DNA polymerases

Replication of eukaryotic genomic DNA is primarily dependent on two different B-family DNA polymerases. Leading strand replication is performed by Pol ε, while lagging strand replication is performed by Pol δ (Braithwaite and Ito, 1993; Pursell et al., 2007). In the model organism *S. cerevisiae* rNTP misincorporation of Pols ε, δ, and α has been shown *in vitro* and *in vivo* (Nick McElhinny et al., 2010b; Sparks et al., 2012) [for review (Williams and Kunkel, 2014)]. These studies estimate that replicative DNA polymerases are responsible for the incorporation of 10,000 rNMPs per round of replication based on *in vitro* incorporation rates. This work infers that as many as 3 million rNMPs could be placed into the human genome by replicative DNA polymerases based on the rates for yeast and mouse DNA polymerases *in vitro* (Nick McElhinny et al., 2010a; Reijns et al., 2012). When compared with other DNA polymerase errors including mismatches, which occur tens of thousands of fold less often, it becomes clear that ribonucleotide incorporations are by far the most frequent error made by replicative DNA polymerases.

Bacterial estimates of rNTP incorporation have been determined in *Escherichia coli.* Replication in *E. coli* differs from eukaryotes in that *E. coli* only uses one DNA polymerase to replicate its genome, DNA Pol III, a C-family replicase (Johansson and Dixon, 2013). Recent studies estimate that DNA Pol III

is responsible for one rNMP incorporation every 2,300 base pairs replicated corresponding to about 2000 misincorporations per round of replication for *E. coli* (Figure 2.1C) (Yao et al., 2013). In contrast, the mismatch error rate for non-Watson-Crick pairings is estimated to be one misincorporation every 15 rounds of replication (Schroeder et al., 2016). This again makes ribonucleotide errors by far the most abundant mistake introduced by the replicative DNA polymerases in bacteria. Of the rNTP incorporations that occur in *E. coli,* 1500 of the 2300 errors are estimated to be rAMP. This is attributed to the inflated rATP/dATP concentration relative to the other nucleotide ratios. (Yao et al., 2013). In contrast to *E. coli*, the model organism *Bacillus subtilis* uses a two DNA polymerase system for genome replication, a mechanism more reminiscent of eukaryotic replication (Dervyn et al., 2001; Sanders et al., 2010). The rNTP incorporation rates for the two *B. subtilis* replicative DNA polymerases, DnaE and PolC, have not been determined although the rates are likely similar to those measured for *E. coli* Pol III given each are C-family replicases. Considering the results above, it seems clear that the incorporation of RNA into DNA during replication is frequent when compared to other DNA polymerase errors and rNMP incorporation appears to be conserved from bacteria to humans and across several families of DNA polymerases. What is known about how these incorporations are recognized and replaced with dNTPs is addressed later in this chapter (see Ribonucleotide excision repair). The process of repairing single rNMP errors and the consequences of deficient repair are examined for *B. subtilis* in Chapter II of this dissertation.

1.2.2 DNA polymerase I

Bacterial DNA Pol I serves in many replication and repair capacities. These include Okazaki fragment maturation, DNA repair, and ribonucleotide excision repair (RER) (see below) (Kornberg and Baker, 1992; Vaisman et al., 2014). It seems that *in vitro,* rNTP incorporation for *E. coli* Pol I depends on the catalytic metal bound and the rNTP/dNTP ratios provided. Pol I incorporates more rNTPs when Mn^{2+} is used in place of Mg^{2+}. This has been attributed to a distortion of the active site in Pol I when Mn^{2+} is bound allowing for promiscuous nucleotide binding (Tabor and Richardson, 1989; Van de Sande et al., 1972). When provided with rNTP concentrations 1000 times that of dNTPs, *E. coli* Pol I readily incorporates rCTP and rGTP but still has high fidelity against rUTP and rATP (Astatke et al., 1998; Ide et al., 1993). One notable mutation which affects Pol I incorporation of rNTPs is a change to the steric gate residue mentioned above. This type of mutation drastically increases the number of rNTPs incorporated (Astatke et al., 1998). The contribution of DNA polymerase I to genomic ribonucleotide incorporation is currently unknown; although because Pol I is critical for replication and repair processes, it is likely that Pol I is responsible for at least some rNMP inclusion into genomic DNA (Figure 1.2D). Further, it is unclear if Pol I is involved in the process of RER. The role of *B. subtilis* DNA Pol I in the repair of ribonucleotides is investigated in Chapter II.

1.2.3 Y-family polymerases

The Y-family of DNA polymerases act in lesion bypass and their expression is often induced by DNA damage [for review see (Sale et al., 2012)]. Y-family DNA polymerase-dependent incorporation of rNTPs has been studied for many DNA polymerases and ranges widely from 2.5-fold to several thousand-fold (DeLucia et al., 2003; Ordonez et al., 2014). Though Y-family DNA polymerases are not highly expressed during normal replication they could be involved in a limited amount of replication and therefore could contribute to rNTP incorporation, particularly during stress responses. Y-family pols are well known for their role in lesion bypass. During lesion bypass, Y-family polymerases are capable of incorporating rNTPs, although to what degree would depend on the organism and specific polymerase examined. For highly promiscuous Y-family polymerases such as *M. smegmatis* DinB2, stretches of rNMPs referred to as "ribopatches" could be incorporated, especially when dNTP concentrations are limited in non-growing states (Ordonez et al., 2014). Such "ribopatches" would result in covalently linked RNA-DNA hybrids with both 5' and 3' RNA-DNA junctions (Figure 1.1B) and would need to be removed and replaced with dNMPs in a similar manner to Okazaki fragments (discussed below). The recognition of such stretches of embedded patches of rNMPs is investigated for the *B. subtilis* enzymes in Chapter III.

1.3 Primase and Okazaki fragments

During the process of DNA replication the leading and lagging strand must be primed for DNA polymerase extension with RNA (Kornberg and Baker, 1992). In bacteria this process is catalyzed by DnaG (primase) a DNA dependent RNA polymerase which travels with the replisome synthesizing primers of 10-15 ribonucleotides approximately every 1.5 kbps (Figure 1.2B) (Corn and Berger, 2006; van der Ende et al., 1985; Rowen and Kornberg, 1978a, 1978b). The action of DnaG accounts for the vast majority of all RNA covalently placed into DNA with approximately 23,000 ribonucleotides placed into the *E. coli* chromosome per round of replication (Schroeder et al., 2014). Estimates for the human nuclear genome are around 150 million ribonucleotides per replication event (Williams and Kunkel, 2014). This process results in the formation of Okazaki fragments, especially on the lagging strand where replication is discontinuous. Okazaki fragments have a 3' RNA-DNA junction and the RNA must be removed and replaced with DNA in a process called Okazaki fragment maturation. Below I discuss the current model for how Okazaki fragments are matured into a continuous strand of DNA.

1.3.1 Okazaki fragment maturation

Eukaryotic Okazaki fragment maturation is thought to occur primarily by a three-enzyme complex with the nick sealed by DNA ligase. The primary model is referred to as the FEN1 model, although other models do exist (Kao and Bambara, 2003). In the FEN1 model after Pol α has extended an RNA primer Pol

δ, bound to the processivity clamp PCNA, takes over lagging strand synthesis. Upon reaching the 5' end of an Okazaki fragment, Pol δ displaces the RNA primer producing a single stranded RNA-DNA flap cleaved by the flap endonuclease FEN1 (Bambara et al., 1997). After the RNA primer is removed, DNA ligase can then seal the remaining nick completing Okazaki fragment maturation into a continuous strand of DNA [for review see (Balakrishnan and Bambara, 2013). Eukaryotic Okazaki fragment maturation has been reconstituted *in vitro* and therefore the process is fairly well understood. Okazaki fragment maturation has not been nearly as well studied for the bacterial enzymes and lacks the same level of mechanistic understanding.

In contrast, bacterial Okazaki fragment maturation is thought to take place with DNA polymerase I removing RNA primers in conjunction with RNase HI (see RNase H enzymes below). Pol I replaces the rNMP primer with dNMPs which is later ligated by DNA Ligase (Ogawa and Okazaki, 1984) (Figure 1.1D). Even though RNase HI operating with DNA polymerase I provides the most accepted model for Okazaki fragment maturation in bacteria, this process is not well understood due to the difficulty in measuring Okazaki fragment maturation *in vivo* and *in vitro*. Further, cells deficient in RNase HI do not show striking deficiencies in Okazaki fragment maturation suggesting RNase HI is not required. In Gram-positive organisms like *B. subtilis* which lack RNase HI and instead have RNase HIII, it is assumed that HIII acts in RNA primer removal from the lagging strand although this has not been shown experimentally. It is also possible that RNase HII contributes to Okazaki fragment processing. Okazaki fragment processing

and other potential enzymes that could contribute to RNA removal from Okazaki fragments in *B. subtilis* are examined and discussed further in the appendix of this dissertation.

1.4 R-loops

R-loops have recently gained attention as causing genome instability in organisms ranging from bacteria to humans [for review see (Aguilera and García-Muse, 2012; Santos-Pereira and Aguilera, 2015; Sollier and Cimprich, 2015)]. R-Loops are three stranded structures where a portion of dsDNA has been invaded by a single strand of RNA (Figure 1.1B). The base-pairing of RNA with one DNA strand forces the displacement of the other DNA strand into a single stranded bubble causing the formation of an RNA-DNA hybrid lacking an RNA-DNA junction. This is in contrast to Okazaki fragments discussed above where the RNA is covalently linked to the DNA molecule through a 3'- junction. R-loops are most often formed during transcription when the transcript generated by RNA polymerase II hybridizes with the complementary template strand of DNA (Drolet et al., 1994). One mechanism for R-loop formation in bacteria is through RecA catalyzed strand invasion of the duplex DNA with ssRNA (Kirkpatrick and Radding, 1992; Kirkpatrick et al., 1992). R-loops are also promoted by several DNA structural features including negative DNA supercoiling, G-rich stretches, and DNA nicks (Duquette et al., 2004; Roy and Lieber, 2009; Roy et al., 2010).

Recently R-loops have been implicated as a threat to genome stability and have also been identified as having a possible role in gene transcription and

chromatin structure remodeling (Chédin, 2016). Long RNA-DNA hybrids such as R-loops adopt a hybrid intermediate between the A and B form helical structure (Shaw and Arya, 2008) resulting in a more stable structure than dsDNA (Roberts and Crothers, 1992). These R-loops are structurally different from both dsDNA and dsRNA, are difficult to remove, and can act as a blockage to DNA replication. In bacteria R-loops are capable of serving to prime DNA synthesis away from the origin during constitutively stable DNA replication (cSDR) (Kogoma, 1997). R-loops have also been implicated as important for mutagenesis related to head-on transcription (Lang et al., 2017). This occurs when the replisome collides with the transcription machinery during DNA replication in a head-on orientation and leads to increased mutagenesis (Paul et al., 2013). R-loop resolution and possible contributions to genome instability are explored in Chapter III of this document using *B. subtilis*. Now that we have considered the types of RNA-DNA hybrids found inside of cells, I will discuss in more detail what is known about the enzymes responsible for recognition and cleavage of RNA-DNA hybrids in bacteria.

1.5 RNase H family of enzymes

RNase H enzymes cleave the RNA containing strand in RNA-DNA hybrids. RNase H enzymes are conserved amongst all kingdoms of life and are categorized into two groups: RNase H type 1 and RNase H type 2 based on primary structure [for review see (Cerritelli and Crouch, 2009; Tadokoro and Kanaya, 2009)]. In bacteria, there are three RNase H enzymes. RNase HI is a

type 1 enzyme while RNase HII and HIII are type 2 enzymes. Organisms usually code for two of these three enzymes, either RNase HI and RNase HII, or RNase HII and RNase HIII (Ohtani et al., 1999a) (Figure 1.3). RNase HI and RNase HIII have overlapping *in vitro* specificity and it has been suggested that they are redundant, which may explain why HI and HIII appear to be mutually exclusive in most genomes (Kochiwa et al., 2007).

1.5.1 RNase HI

RNase HI enzymes cleave polymers of 4 or more rNMPs hybridized to DNA leaving a 5'-PO_4^- (Berkower et al., 1973; Hogrefe et al., 1990). RNase HI enzymes are discussed in two groups: The archaeal hybrid binding domain RNase HI (HBD) and *E. coli* type I coded for by the *rnhA* gene (Tadokoro and Kanaya, 2009). *E. coli* RNase HI is the most well studied bacterial RNase H enzyme. RNase HI from *E. coli* cleaves RNA with a minimum of four consecutive rNMPs hybridized to DNA and has been implicated in both the resolution of R-loops and Okazaki fragment maturation *in vivo* (Hogrefe et al., 1990; Hong et al., 1995; Kogoma and von Meyenburg, 1983; Kogoma et al., 1993; Ogawa and Okazaki, 1984). Single rNMP incorporations by the replicative DNA polymerase are unlikely to be cleaved by RNase HI although ribopatches of $\geq$ 4 rNMPs possibly resulting from Y-family polymerase synthesis, could serve as a substrate for HI *in vivo*. Not all bacteria contain RNase HI. Instead, many bacteria contain RNase HIII, which is proposed to have a similar function (discussed below).

1.5.2 RNase HII

The most well studied RNase HII enzyme is coded for by the *rnhB* gene of *E. coli*. Biochemically, RNase HII from *E. coli* and *B. subtilis* cleave 5' to both single rNMPs embedded in DNA and polymers of rNMPs hybridized to DNA (Itaya, 1990; Ohtani et al., 1999b, 2000). This means that RNase HII could contribute to resolving R-loops, Okazaki fragments, ribopatches, or be involved in removal of single rNMPs during RER (see below). Deletion of the *rnhB* gene coding for RNase HII from bacteria shows an increased susceptibility of genomic DNA to alkaline treatment representing an assay used to measure the accumulation of ribonucleotides *in vivo* (Lu et al., 2012a; McDonald et al., 2012; Yao et al., 2013). In humans, mutations in the gene coding the catalytic domain of RNase H2 cause a neurodegenerative disease known as Aicardi-Goutierres Syndrome (AGS) and deletions of the same gene in mice are embryonically lethal [for review see (Crow, 2013; Crow; Reijns and Jackson, 2014)]. However, in bacteria loss of RNase HII does not cause a growth defect or strong phenotype. In *B. subtilis*, a deletion of *rnhB* (RNase HII) causes a 2-fold increase in spontaneous mutagenesis (Yao et al., 2013). Interestingly an *rnhB* deletion in *E. coli* does not have an increase in mutagenesis, suggesting a fundamental difference in the secondary ribonucleotide excision repair pathway between *E. coli* and *B. subtilis* (Yao et al., 2013). The catalytic activity of RNase HII is dependent on the ability of four negatively charged residues (DEDD) to coordinate a divalent metal cation necessary for hydrolysis (Figure 1.3). The ability of RNase HII to recognize single ribonucleotides in DNA is attributed to a

GRG motif which directly interacts with the 2'-OH of the single ribonucleotide (Rychlik et al., 2010). This GRG motif was recently expanded to G(R/K)G because *C. pneumoniae* RNase HIII has a GKG motif and was recently discovered to cleave single rNMPs with manganese (see below) (Lu et al., 2012b, 2012a). It is unclear if cleavage of single rNMPs by RNase HIII is biologically conserved. Bacterial RNase HII seems to cleave most types of RNA-DNA hybrids *in vitro* but the Δ*rnhB* strains lack a phenotype, making it difficult to study the function of RNase HII *in vivo*. Chapter II of this dissertation investigates the consequences of an RNase HII deletion in *B. subtilis* and its effect on genome stability. Chapter III investigates how divalent metal ions regulate RNase HII activity *in vitro*.

1.5.3 RNase HIII

In bacterial genomes RNase HI and RNase HII are commonly coincident. However several organisms including those in the *Chlamydophila*, *Staphylococcus*, *Streptococcus*, *Enterococcus* and *Bacillus* genera*,* code for an RNase HII gene and an RNase HIII gene (Ohtani et al., 1999b; Randall et al., 2017). RNase HIII is considered a type 2 RNase H enzyme and contains a large N-terminal domain implicated in substrate binding and a DEDE motif for metal coordination (Figure 1.3) (Chon et al., 2006; Jongruja et al., 2012; Miyashita et al., 2011). Even though RNase HIII is a type 2 enzyme, biochemical characterization suggests it may be functionally analogous to *E. coli* RNase HI (Lu et al., 2012a; Ohtani et al., 1999b). RNase HIII enzymes are not known to

cleave at single rNMPs with one interesting exception. *Chlamydophila pneumoniae* RNase HIII (Cpn-HIII) complements an *E. coli rnhB* knockout when cells are grown in the presence of high Mn^{2+} concentrations (Lu et al., 2012a). The enzyme was characterized biochemically and demonstrated activity on a single embedded rNMP with Mn^{2+} but not Mg^{2+}. This activity was attributed to a serine residue permitting a nearby GKG motif to hydrogen bond with the 2'OH (Lu et al., 2012b). This activity is similar to the GRG motif described above for RNase HII enzymes. It is not known whether other RNase HIII enzymes also have this function *in vitro* and whether it is physiologically relevant *in vivo*. RNase HIII has also recently been implicated in the processing of R-loops in *B. subtilis* (Lang et al., 2017). The appendix of this dissertation contributes to our understanding of how RNase HIII impacts genome stability, as well as the ability of RNase HIII to cleave R-loop structures *in vitro*.

1.6 Ribonucleotide Excision Repair

Above, I discussed the high frequency of DNA polymerase rNTP incorporation and the enzymes responsible for their recognition. Here I will discuss how these incorporations are repaired. Ribonucleotide excision repair (RER) refers to the removal and replacement of rNMPs covalently incorporated into genomic DNA. Most commonly RER is used to describe the process of repairing single rNMP incorporations by replicative DNA polymerases through an RNase HII-dependent pathway. When left unrepaired rNMPs incorporated into DNA are highly unstable and likely to hydrolyze at the 3' end resulting in a 2', 3'-

cyclic phosphate, causing genome instability (Oivanen et al., 1998). Unrepaired rNMPs in the template strand can stall or slow DNA polymerases (Yao et al., 2013) further affecting genome integrity. Below, I discuss what is known of the RER repair pathways in eukaryotes and bacteria.

1.6.1 Eukaryotic RER

It has been shown in numerous organisms that when RNase HII is non-functional rNMPs accumulate in genomic DNA (Lu et al., 2012a; McDonald et al., 2012; Nick McElhinny et al., 2010a; Reijns et al., 2012; Yao et al., 2013). In eukaryotes, this pathway has been reconstituted using *S. cerevisiae* proteins (Sparks et al., 2012). First, RNase HII nicks 5' to the rNMP nested in DNA leaving a 5'-PO_4^- and 3'-OH. The replicative DNA polymerases Pol δ or ε then catalyze strand displacement of the rNMP containing portion while replacing it with dNMPs in conjunction with the processivity clamp (PCNA). The remaining flap containing the rNMP is then removed by FEN1 or Exo I followed by DNA ligase sealing the remaining nick (Sparks et al., 2012). Therefore eukaryotic RER is reminiscent of Okazaki fragment maturation (Sparks et al., 2012). In an RNase H2 deletion, single rNMPs can be recognized and incised by Topoisomerase I (Topo I) (Kim et al., 2011; Sekiguchi and Shuman, 1997). In this backup pathway Topo I incision results in a 2', 3'-cyclic phosphate. This cyclic phosphate is ultimately removed, resulting in 2-5 bp deletions (Sparks et al., 2012). Topo I-dependent deletions are mitigated by Srs2 helicase in combination with Exo I (Kim et al., 2011; Potenski et al., 2014; Sekiguchi and Shuman, 1997; Williams et

al., 2013). Therefore, the process of RER and the backup pathways that repair single rNMP errors in the absence of RNase H2 are well studied. Much less is known in *E. coli* and the pathway is completely unstudied in *B. subtilis,* which is thought to use a completely different set of enzymes.

1.6.2 Bacterial RER

It is unclear how RER takes place in bacteria. Further, RER is hypothesized to be completed by at least two pathways and probably more. Some insight was gained through the expression of a Pol V (UmuC-Y11A) mutant in *E. coli* which has a mutated steric gate residue allowing for adept rNMP incorporation. This Pol V variant readily incorporates rNMPs when overexpressed *in vivo* (Vaisman et al., 2014). These experiments highlight the importance of RNase HII and the nick translation activity of Pol I in mitigating Pol V-dependent mutagenesis. It also suggests that Pol I nick translation activity could be carried out jointly by other enzymes including Exo I and Pol III. Further, using the same Pol V variant, it was suggested that nucleotide excision repair (NER) serves as a backup pathway for rNMP recognition in RNase HII deficient cells (Cai et al., 2014; Vaisman et al., 2013).

Little is known about RER in Gram-positive bacteria. *B. subtilis* RNase HII has been implicated in the pathway (Yao et al., 2013), but how an RNase HII-incised substrate is repaired remains unknown and the consequences of ribonucleotides left uncorrected in the genome has not been determined. Further,

there is the possibility that RNase HIII may also contribute to RER. These questions are addressed in both Chapter II and Chapter III of this document.

1.7 Summary

RNA bound to DNA through hydrogen bonding, covalent linkage, or both exist in the form of DNA polymerase incorporations, Okazaki fragments, and R-loop formation. RNA-DNA hybrids are common and if not resolved can have devastating effects on cell viability and genome integrity. RNA bound to DNA is recognized and cleaved by the RNase H family of enzymes. In *B. subtilis*, there are four putative RNase H enzymes. Some of these enzymes have been characterized biochemically; however recent work has shown more promiscuous substrate recognition than previously shown. Furthermore, the bacterial pathways in which each RNase H enzyme works and the contribution these may have in genome maintenance remain unclear.

This dissertation aims to examine the complete *in vitro* functions of the *B. subtilis* RNase H enzymes and reveal their *in vivo* functions through the study of cells deficient in RNase H activity.

1.8 Notes and Acknowledgements

Modified parts of this chapter were published in Critical Reviews in Biochemistry and Molecular Biology including versions of Figures 1.2 and 1.3 (Schroeder et al., 2014).

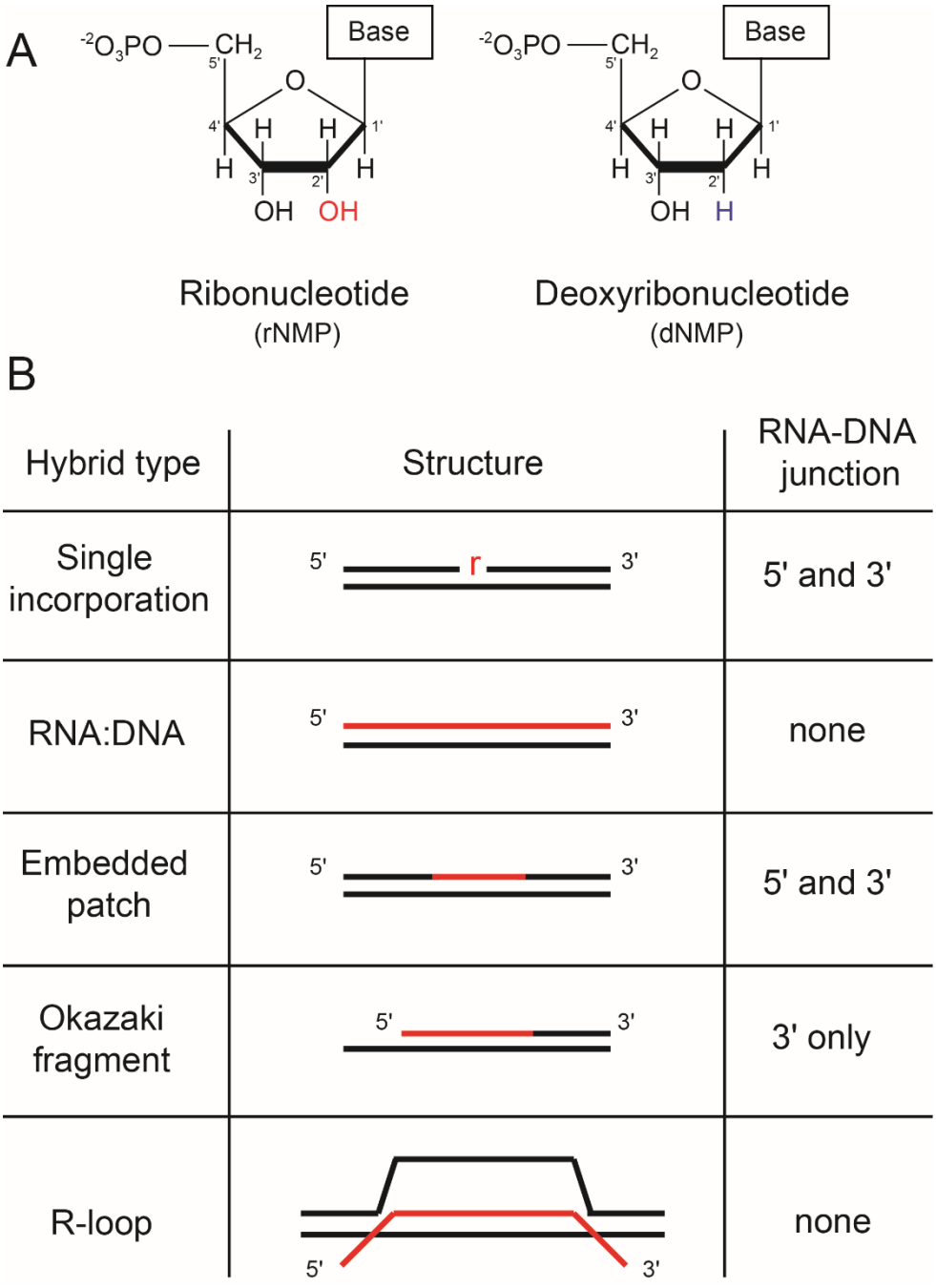

Figure 1.1 RNA-DNA hybrid structures. (A) Comparison of a ribonucleotide and a deoxyribonucleotide. All carbons are labelled with the 2' position in red (rNMP) or blue (dNMP). (B) A list of relevant RNA-DNA hybrids with representative structures. The polarity of the RNA containing strand is indicated and the presence of any RNA-DNA junctions are described. Red “r” and lines represent ribonucleotides.

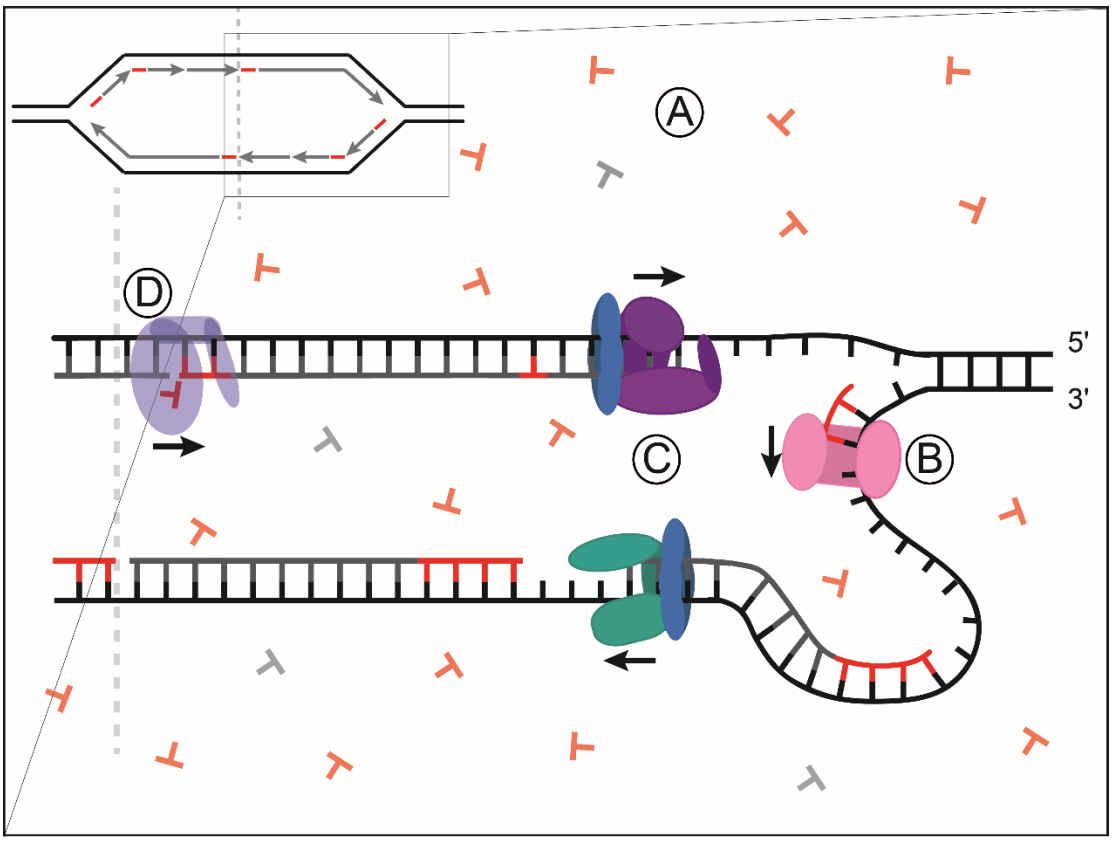

Figure 1.2 Incorporation of RNA during replication. Examples of different RNA-DNA hybrids. (A) During replication of DNA, concentrations of rNTPs exceed those of dNTPs, increasing the likelihood of rNTP incorporation by a replicative DNA polymerase. Gray = dNTP, red = rNTP (B) Primase (pink), a DNA dependent RNA polymerase, synthesizes RNA primers. (C) Replicative DNA polymerases (purple, teal, β-clamp blue) incorporate rNTPs during replication of the newly synthesized strand (in gray) (D) Pol I (lavender) removes RNA primers left by primase after DNA polymerase loading and extension.

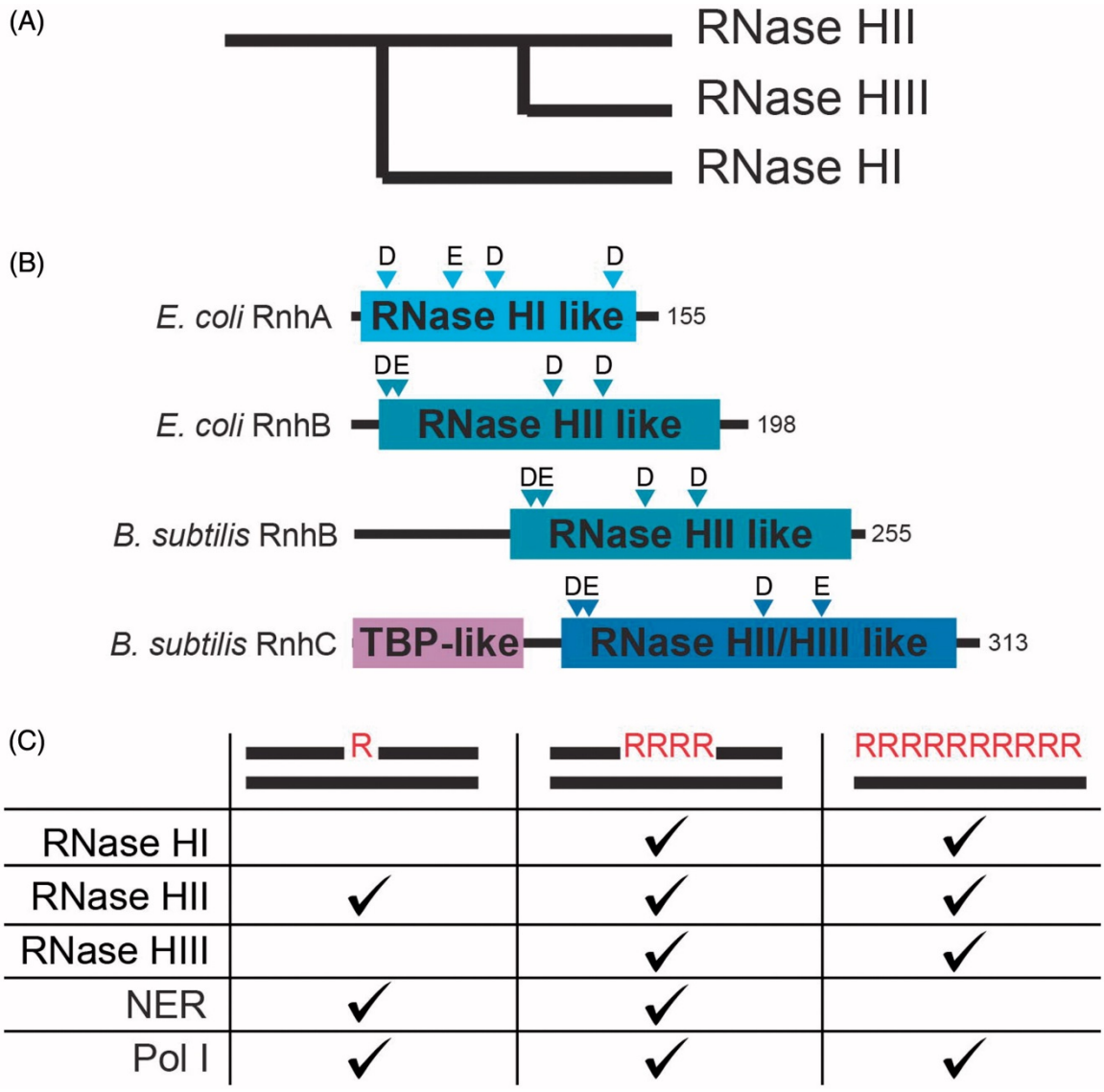

Figure 1.3 Bacterial recognition and removal of rNMPs. (A) Schematic evolutionary history of the family of RNase H enzymes with respect to sequence similarity. (B) Domain structures of functional *E. coli* and *B. subtilis* RNases H. Triangles denote the location and identity of the residues stabilizing the divalent metal ions necessary for catalytic activity. (C) The known contribution of RNases H, NER and Pol I to recognize or engage in repair of different RNA/DNA hybrids.

CHAPTER II

II. Mutagenic cost of ribonucleotides in bacterial DNA

2.1 Abstract

Replicative DNA polymerases misincorporate ribonucleoside triphosphates (rNTPs) into DNA approximately once every 2,000 base pairs synthesized. Ribonucleotide excision repair (RER) removes rNMPs from genomic DNA, replacing the error with the appropriate deoxyribonucleoside triphosphate. Ribonucleotides represent a major threat to genome integrity with the potential to cause strand breaks. Furthermore, it has been shown in the bacterium *Bacillus subtilis* that loss of RER increases spontaneous mutagenesis. Despite the high rNTP error rate and the effect on genome integrity, the mechanism underlying mutagenesis in RER-deficient bacterial cells remains unknown. We performed mutation accumulation lines and genome-wide mutational profiling of *B. subtilis* lacking RNase HII, the enzyme that incises at single rNMP residues initiating RER. We show that loss of RER in *B. subtilis* causes strand and sequence-context dependent GC→AT transitions. Using purified proteins, we show that the replicative polymerase DnaE is mutagenic

within the sequence context identified in RER-deficient cells. We also found that DnaE does not perform strand displacement synthesis on its own. Given the use of nucleotide excision repair (NER) as a backup pathway for RER in RNase HII-deficient cells and the known mutagenic profile of DnaE, we propose that misincorporated ribonucleotides are removed by NER followed by error-prone resynthesis with DnaE.

2.2 Significance

DNA polymerases frequently incorporate ribonucleotides in place of deoxyribonucleotides during genome replication. RNase HII is responsible for initiating the removal of ribonucleotide errors across all three domains of life. Ribonucleotides that persist in genomic DNA due to defects in RNase HII result in strand breaks, mutagenesis, and neurodevelopmental disease in humans. Here, we define the proteins important for ribonucleotide excision repair in *Bacillus subtilis* and use genome-wide mutational profiling to determine the mutagenic cost of ribonucleotides in RNase HII-deficient cells. We show that the absence of RNase HII yields error-prone ribonucleotide correction via a pathway that relies on an essential DNA polymerase. We further demonstrate that error-prone ribonucleotide removal causes sequence context dependent GC → AT transitions on the lagging strand.

2.3 Introduction

Replicative DNA polymerases duplicate genomes with high fidelity (Kunkel and Bebenek, 2000). In bacteria, it is estimated that base-pairing errors between deoxyribonucleoside triphosphates (dNTP) occur approximately once every 60 million properly paired bases (Schroeder et al., 2016). Such a high degree of accuracy is due to the intrinsic fidelity of the DNA polymerases from factors including induced fit in the active site and 3' to 5' exonuclease “proofreading” activity (Kunkel and Bebenek, 2000). Interestingly, replicative DNA polymerases are far more likely to incorporate sugar errors as opposed to dNTP base-pairing errors (Nick McElhinny et al., 2010a; Yao et al., 2013). Sugar errors represent the insertion of a ribonucleoside triphosphate (rNTP) in place of its corresponding dNTP (Brown and Suo, 2011). With the exception of rUTP, the difference between each dNTP and its corresponding rNTP is the presence of a single oxygen atom at the 2' position of the ribose sugar. Many DNA polymerases have a steric gate residue, which limits the use of rNTPs as a substrate (DeLucia et al., 2003). The steric gate is often a bulky amino acid side chain that clashes with the 2'-OH on the ribose sugar of rNTPs, limiting their incorporation into DNA (Brown and Suo, 2011; DeLucia et al., 2003). The intracellular abundance of rNTPs presents a challenge for sugar specificity during DNA replication as polymerases needed to select the proper dNTP are outnumbered 10- to 100-fold by rNTPs (Nick McElhinny et al., 2010a; Traut, 1994). The imbalance in nucleotide concentration causes rNTPs to be incorporated into genomic DNA in eukaryotes and bacteria (Kim et al., 2011; Yao et al., 2013).

In *Saccharomyces cerevisiae*, the DNA polymerases required for genome replication combine for an error rate of ~10,000 sugar errors per round of replication (Nick McElhinny et al., 2010a). Specifically, the DNA polymerases δ and Ɛ are responsible for incorporating 1 rNMP every 5,000 and 1,250 base pairs replicated, respectively (Nick McElhinny et al., 2010a). In *S. cerevisiae*, ribonuclease (RNase) H2 is responsible for initiating removal of single rNMPs from DNA as part of the ribonucleotide excision repair (RER) pathway (Chon et al., 2013; Sparks et al., 2012). During RER in *S. cerevisiae*, RNase H2 incises the sugar-phosphate backbone 5' to a single rNMP followed by strand displacement synthesis with Pol δ or Pol Ɛ. The resulting 5' flap is removed by FEN1 or Exo I with the nick sealed by DNA ligase (Sparks et al., 2012). Loss of RER is mutagenic in *S. cerevisiae*, resulting in topoisomerase I-dependent 2- to 5-bp deletions (Kim et al., 2011).

Much less is known about rNTP incorporation, RER, or the consequences of rNMPs nested in bacterial DNA. In *Escherichia coli*, DNA polymerase III holoenzyme incorporates sugar errors at a rate of one misinserted rNMP every 2.3 kb replicated (Yao et al., 2013). Based on the selectivity of Pol III for dNTPs relative to rNTPs, it is expected that ~2,000 rNMPs would be incorporated per round of replication (Yao et al., 2013). Furthermore, Y-family translesion DNA polymerases have been shown to readily incorporate rNMPs in place of dNMPs (McDonald et al., 2012; Ordonez et al., 2014). Therefore, rNMPs likely represent the most frequent nucleotide targeted for removal from bacterial DNA.

Although it is clear that rNMPs are frequently incorporated into DNA, the consequences of embedded rNMPs for genome integrity in bacteria remain unclear. Interestingly, RNase HII is broadly conserved in eubacteria (Kochiwa et al., 2007), yet loss of RER in *E. coli* does not increase mutagenesis; however *B. subtilis* cells lacking RNase HII (Δ*rnhB*) show an increase in mutation rate (Yao et al., 2013). In *E. coli,* nucleotide excision repair (NER) removes ribonucleotides from DNA in cells lacking RNase HII (Δ*rnhB*) (Vaisman et al., 2013). Therefore, in *E. coli* it seems that RNase HII-dependent RER provides the primary pathway for rNMP removal and that NER serves as a backup (Vaisman et al., 2013). Furthermore, genetic evidence taking advantage of a Pol V variant adept at rNMP incorporation showed that DNA polymerase I is important for RER in *E. coli* with other gene products providing redundant functions (Vaisman et al., 2014). *E. coli* and *B. subtilis* differ in that loss of RNase HII is mutagenic in *B. subtilis*, but not in *E. coli*, suggesting a fundamental difference in RER or the backup pathways used (Yao et al., 2013). The source of mutagenesis due to rNMPs in genomic DNA is unknown, and the overall RER pathway has not been defined genetically or reconstituted *in vitro* for *B. subtilis*.

To understand how unrepaired rNMPs impact genome stability, we performed mutation accumulation lines in RER-deficient (Δ*rnhB*) *B. subtilis* cells. We found that persistent rNMPs in genomic DNA result in a mutagenic signature that is caused by error-prone gap-filling. Furthermore, to understand how rNMPs in genomic DNA are replaced, we reconstituted the minimal set of proteins to replace an rNMP with a dNMP in a primer extension reaction *in vitro*.

2.4 Results

2.4.1 RNase HII incises at single rNMPs in duplex DNA

B. subtilis is known to have the RNase H enzymes RNase HII (*rnhB*) and RNase HIII (*rnhC*) (Itaya et al., 1999; Ohtani et al., 1999b). Based on prior studies, we expected that *B. subtilis* RNase HII would be able to incise DNA 5' to a single ribonucleotide in dsDNA (Haruki et al., 2002; Ohtani et al., 1999b). To be certain, we purified *B. subtilis* RNases HII and HIII alongside catalytically inactive variants (DE-AA) to serve as controls (Figure 2.1). Each protein was purified with an N-terminal His_6-SUMO tag that was removed during the purification process, yielding recombinant proteins without any additional amino acids (Materials and Methods). We incubated each protein with a dsDNA substrate containing four contiguous rNMPs or a single rNMP with the rNMP-containing strand labeled on the 5' end. Following incubation, each reaction was resolved in a 20% urea–polyacrylamide gel. We show that RNases HII and HIII cleave the four-rNMP substrate while each catalytically inactive variant failed to show activity (Figure 2.2A). Under the conditions tested, RNase HII cleaved the substrate containing a single rNMP, while RNase HIII did not (Figure 2.2B). This work supports prior results showing that RNase HII initiates removal of single ribonucleotides from DNA, whereas RNase HIII may be more important for initiating repair of longer stretches of rNMPs embedded in DNA (Haruki et al., 2002; Ohtani et al., 1999b).

2.4.2 RER-deficient *B. subtilis* cells accumulate GC → AT transitions

We performed mutation accumulation (MA) lines followed by whole-genome sequencing in cells without RNase HII (Δ*rnhB*) to determine the consequences to genome-wide mutation rate and mutation spectrum. Genome-wide MA was compared with a compendium of wild-type *B. subtilis* MA line data that we have previously compiled (Schroeder et al., 2016). Eighty-one individual Δ*rnhB* (RNase HII) lines were completed for this work, each of which underwent 3,610 generations (Materials and Methods). An overall summary of the MA line results is presented in Table 2.1, with all variants presented in Dataset S1 [see (Schroeder et al., 2017)]. A total of 462 mutations were detected, 420 of which were base-pair substitutions (BPSs). We also detected 42 insertions/deletions (indels) in Δ*rnhB* lines (Table 2.1). In agreement with prior results using a mutational reporter (Yao et al., 2013), loss of *rnhB* yielded an increase in the overall genome-wide mutation rate of ∼1.5-fold compared with wild type (Table 2.1). BPSs were increased in Δ*rnhB* (Figure 2.3A), with GC → AT transitions occurring approximately twofold more frequently in Δ*rnhB* than in wild type (Figure 2.3B). From these data, we conclude that ribonucleotides persistent in genomic DNA result in a specific genome-wide mutagenic signature of GC → AT transitions. The mutation spectrum reported here is distinct from the 2- to 5-bp deletion spectrum observed for an RNase H2 deficiency in *S. cerevisiae* (Kim et al., 2011).

After observing the accumulation of transitions along the genome as cumulative distributions, it became clear that the increase in GC → AT transitions

in Δ*rnhB* cells showed a strand dependence (Figure 2.3C). The reference sequence of the right replichore represents the lagging-strand template and the reference sequence of the left replichore represents the leading-strand template. Therefore, the effect of persistent ribonucleotides on the GC → AT transition rate must be strand-dependent. Specifically, genomic loci were at an increased risk of undergoing a transition when RER was inactivated and guanosine was in either the leading strand or the lagging-strand template. We tested whether transitions due to misinserted rNMPs were more likely to occur in certain contexts by calculating conditional mutation rates for transitions in the 16 possible dinucleotide sequence contexts (Figure 2.4A). All sequence contexts were considered from the perspective of the lagging-strand template, and the analysis was performed separately for 5' (Figure 2.5) and 3' contexts (Figure 2.4A). The effects of 5' neighboring nucleotides on the rNMP-induced transition rate were subtle; however, the 3' neighboring context had a strong effect (Figure 2.4A). Specifically, if guanosine is present in the lagging-strand template (or possibly the leading strand) followed 3' by cytidine, the guanosine will undergo a transition ~4 times more frequently in Δ*rnhB* than in wild type (Figure 2.4A and Figure 2.5).

Upon finding that GC → AT transitions are strand- and sequence-context–dependent, we further investigated the effect of the 3' local sequence context on G→A transitions in the lagging-strand template using logistic regression. Logistic regression was performed genome-wide to determine if sequence context in the lagging-strand template at distances up to five nucleotides 3' to a guanosine could influence mutation occurrence at the position

of the guanosine (Materials and Methods). Impressively, nucleotide identity up to five nucleotides 3' to a guanosine in the lagging-strand template was associated with G→A transition occurrence in Δ*rnhB* (Fig. 2.4C). Bioinformatics analysis (Bailey and Elkan, 1994) determined a sequence motif to be present at G → A transitions in the lagging-strand template of Δ*rnhB* as 5'-GCC(T/C)T-3'. The underlined guanosine indicates the position that underwent a transition to adenosine (Figure 2.4). Therefore, genome-wide MA lines show a strong sequence-context–dependent increase in GC → AT transitions in RNase HII (Δ*rnhB*)-deficient cells (Figures 2.3, 2.4 and 2.5).

NER is able to remove ribonucleotides from DNA when RER is deficient by excising a tract of DNA extending from eight phosphodiester bonds 5' to the rNMP through four or five phosphodiester bonds 3' to the rNMP (Vaisman et al., 2013). Adenosine is by far the most frequently misinserted ribonucleotide by *E. coli* Pol III (Yao et al., 2013). The motif that we identified to be associated with G → A transitions in the lagging-strand template includes thymidine three to four nucleotides (a distance of four to five phosphodiester bonds) 3' to the G → A transition. In the absence of RNase HII, if rAMP were misinserted across from a thymidine in 5'-GCC(T/C)T-3' during lagging-strand synthesis, NER could excise the rAMP-containing strand (Figure 2.4D). NER-dependent removal would generate a gap that must be filled. Because high-GC content near a given position contributes to a higher rate of mispairing by DNA polymerases (Petruska and Goodman, 1985), we hypothesized that gap filling after rNMP removal is more error-prone in part due to the high-GC content at the motif near the G→A

transitions in the lagging-strand template. Below, we test the hypothesis that error-prone resynthesis results in mutagenesis.

2.4.3 DNA polymerase I participates in RNase HII-dependent RER in *B. subtilis*

In *E. coli*, DNA polymerase I (Pol I) has been shown to participate in RER *in vivo* (Vaisman et al., 2014), while in *B. subtilis* the DNA polymerase participating in RER is unknown. *B. subtilis* has three enzymes that we considered candidates for resynthesis of the DNA from an RNase HII incision. These polymerases are the replicative enzymes PolC, DnaE, and Pol I. *B. subtilis* also has two translesion DNA polymerases, PolY1 and PolY2 (Sung et al., 2003). Due to the specific nature of the GC → AT transition in Δ*rnhB* cells, the observed genome-wide spectrum is inconsistent with activity of PolY1 and PolY2, which would include transversion mutations (Sung et al., 2003).

We began by testing which DNA polymerase(s) could function in the canonical RNase HII-dependent RER pathway. We purified Pol I and received DnaE and PolC from Charles McHenry for extension reactions (Materials and Methods). The purity of Pol I, DnaE, and PolC was verified (Figure 2.6). All three *B. subtilis* Pols were active in a positive control assay for primer extension (Figure 2.7A). We show a schematic to determine the Pol capable of extending a 3'-OH end generated after RNase HII incision at a single rNMP (Figure 2.7B). We incubated a 5' end-labeled RNase HII-incised dsDNA substrate with each DNA polymerase. The RNase HII reactions were quenched after 45 min and Pol I,

DnaE, or PolC were added to the RNase HII reaction followed by quenching at 1, 5, 20, and 60 minutes (Figure 2.7 C–F). Pol I extended the substrate after RNase HII incision at the single rNMP (Figure 2.7C). Furthermore, the Pol I product (Figure 2.7C, lane 8) was refractory to alkaline hydrolysis, demonstrating that the rNMP was removed and replaced. As a control, we show that the substrate with the embedded rNMP from lane 3 (Figure 2.7C) is alkali-sensitive.

Neither DnaE nor PolC were able to catalyze DNA synthesis from a nick generated by RNase HII (Figure 2.7 D and E). Pol I was able to extend 78% of an RNase HII-nicked substrate after 60 min whereas DnaE and PolC were able to extend only 3% and 2% of the substrate, respectively (Figure 2.7F). We conclude that RNase HII and Pol I work in conjunction to remove and replace a single misincorporated rNMP in DNA and that PolC and DnaE are not effective at strand displacement synthesis from an RNase HII-incised nick.

2.4.4 DnaE catalyzes error-prone resynthesis of a gapped substrate resulting in a G → A transition

After establishing that Pol I participated in RNase HII-dependent RER and that PolC and DnaE did not, we asked if Pol I, DnaE, or PolC catalyzed mutagenic resynthesis *in vitro*. Because we identified the sequence context [5'-GCC (T/C) T-3'] *in vivo* (Figure 2.4), we assembled a 5' end-labeled substrate to more closely model the resynthesis step following NER-mediated excision (Figure 2.8). We provided all four dNTPs to demonstrate that each polymerase could extend the primer during a 5-minute reaction (Figure 2.8B). With activity

established, we tested the ability of Pol I, DnaE, and PolC to generate a mismatch and then fully extend the mismatched product in the absence of deoxycytidine triphosphate (dCTP) (Fig. 2.8B). Pol I and DnaE were able to misincorporate and extend from a mismatch in the absence of dCTP. *B. subtilis* PolC has intrinsic 3' to 5' exonuclease proofreading activity (Sanjanwala and Ganesan, 1991) while *B. subtilis* DnaE lacks a known proofreading-associated protein (Le Chatelier et al., 2004). Furthermore, although *E. coli* Pol I possesses proofreading (Setlow et al., 1972), *B. subtilis* Pol I lacks proofreading activity (Duigou et al., 2005). Quantification of extension from a mismatch showed that DnaE caused an approximately two-fold increase in error relative to Pol I (Figure 2.8 B and C). We found that DnaE extended 13% of the substrate, whereas Pol I extended ~7% of the substrate when reactions lacked dCTP (Figure 2.8C). PolC was ineffective at mismatch formation and extension, with only 1.3% of the primer extended after 5 min (Figure 2.8C). Therefore, DnaE was approximately twice as likely to form a mispair across from guanosine in the 5'-GCC(T/C)T-3' sequence context than Pol I and 10 times more likely than PolC (Figure 2.8C). Interestingly, in the absence of RNase HII (Δ*rnhB*), we observed a two-fold increase in GC → AT transitions, a result supportive of the error rate that we observed *in vitro* following DnaE extension. We propose that, following excision of rNMPs in cells that lack RNase HII activity (Δ*rnhB*), the gap is filled by DnaE, resulting in an increased genome-wide BPS rate.

2.4.5 Loss of NER alters the mutation spectrum in Δ*rnhB*

To test if *rnhB* and *uvrA* undergo a genetic interaction, we used rifampin resistance as an indicator for mutation rate (Table 2.2). We found that cells with Δ*rnhB* or Δ*uvrA* showed a mutation rate 1.7- and 1.6- fold higher than wild type, respectively (Table 2.2). The mutation rate of the double mutant (Δ*rnhB*, Δ*uvrA*) was increased 3.9-fold relative to wild type, suggesting an additive or synergistic affect. Because the mutation rate suggests genetic interaction between *rnhB* and NER, we tested the mutation spectrum using a selection for a reporter gene to determine if Δ*uvrA* altered the mutation spectrum of Δ*rnhB*. Loss-of function mutations in genes encoding thymidylate synthase yield resistance to the drug trimethoprim (Dutra and Lovett, 2006). Because *B. subtilis* encodes two thymidylate synthases (Neuhard et al., 1978), we disrupted *thyB* and determined the mutation spectrum of *thyA* mutants in trimethoprim resistant colonies. We tested strains that were Δ*rnhB*, Δ*uvrA*, Δ*rnhB*; Δ*uvrA*, or otherwise wild type. Dataset 2 [see (Schroeder et al., 2017)] includes information on all *thyA* variants detected. Comparison of the *thyA* mutation spectrum from each strain showed that the only significant difference in mutation spectra was between Δ*rnhB* and Δ*rnhB*; Δ*uvrA* strains (P = 0.04) (Figure 2.9). These data support the conclusion that the mutation spectrum in Δ*rnhB* cells is altered when NER is inactivated (Δ*uvrA*), suggesting that NER may provide a backup role for rNMP removal as shown for *E. coli* (Vaisman et al., 2013).

4.6 Discussion

4.6.1 Model for RER in *B. subtilis*.

The data that we present shows that RNase HII is responsible for incision at a single embedded rNMP and that Pol I is efficient at resynthesis. Based on these results, we propose that the minimal set of proteins for RER in *B. subtilis* includes RNase HII, Pol I, and DNA ligase to seal the remaining nick. Our results indicate that PolC and DnaE are largely incapable of replication from an RNase HII-incised nick. NER has been shown to function as a backup pathway for removal of rNMPs in *E. coli* lacking RNase HII (Vaisman et al., 2013), and we provide evidence suggesting that *B. subtilis* NER functions in a similar capacity. Our findings indicate that, in the absence of RNase HII, the backup pathway for removal of rNMPs in the *B. subtilis* genome comes at a mutagenic cost (Yao et al., 2013). Here, we show that DnaE is capable of causing the increase in genome-wide BPS rate in *B. subtilis* cells lacking RNase HII. Therefore, we propose that in wild-type cells RNase HII incises 5' to a single rNMP followed by removal and resynthesis by Pol I. In the absence of RNase HII, NER likely removes the rNMP, leaving a gap allowing access and mutagenic resynthesis by DnaE. DnaE is inactive from a nick (Figure 2.7), but should be active on a gapped substrate (Figure 2.8). Considering data from prior studies (Vaisman et al., 2013), we propose that NER provides the five to eight nucleotide gap needed for resynthesis by DnaE resulting in mutagenesis *in vivo*. Because the mutation rate increased in cells deficient for NER and RER, it remains possible that, in

Δ*rnhB uvrA*+ cells, mutations are generated by DnaE, but in Δ*rnhB* Δ*uvrA* strains, the mutations that are observed are generated by alternate mechanisms.

Our prior work and results herein measured spontaneous mutagenesis for Δ*rnhB B. subtilis* and *E. coli* cells (Table 2.2) (Yao et al., 2013). These results show that Δ*rnhB* cells have a 1.7- 2 fold higher mutation rate than that of wild type. In contrast, the mutation rate of Δ*rnhB E. coli* cells was indistinguishable from that of the wild-type control (Yao et al., 2013). We propose that the difference in mutagenic cost of Δ*rnhB B. subtilis* cells relative to *E. coli* is due to differences in proofreading activity associated with DnaE. *E. coli* uses DnaE for replication of the leading and lagging strands with Ɛ providing proofreading (Johnson and O'Donnell, 2005). In contrast, *B. subtilis* uses proofreading-proficient PolC for replication of the leading and lagging strands after DnaE extends the RNA primer (Sanders et al., 2010). We propose that, because *B. subtilis* DnaE lacks proofreading, it causes mutagenic resynthesis of gapped repair intermediates. During RNase HII-dependent RER, DnaE does not cause mutagenesis because DnaE is unable to extend the 3' end of a nicked substrate. In contrast, *E. coli* Δ*rnhB* cells do not show an increase in mutagenesis because, if Pol III gains access to the repair intermediate, it has an associated proofreading domain that limits errors during the resynthesis step. In further support of this model, it has been shown that Pol III-dependent resynthesis during NER is mutagenic when proofreading is inactivated in *E. coli* (Vaisman et al., 2014).

2.6.2 Embedded ribonucleotides are unlikely to provide significant participation in nascent strand recognition during mismatch repair

E. coli uses DNA methylation as a marker for strand recognition during mismatch repair (MMR) (Lahue et al., 1989). Most bacteria and all eukaryotes lack a methylation-directed pathway (Lenhart et al., 2016). Biochemical results show that a nick will direct excision to the incised strand for mismatch removal (Genschel and Modrich, 2003). It is reasonable to hypothesize that 5' and 3' termini of Okazaki fragments serve as nascent-strand signals for correction of errors in the lagging strand. The signal that directs mismatch excision to the leading strand is unclear. It has recently been proposed that removal of rNMPs from DNA in eukaryotes could provide a strand discrimination signal for mismatch correction (Ghodgaonkar et al., 2013; Lujan et al., 2013). We considered that bacteria like *B. subtilis*, which lack a methylation-directed MMR pathway, could also use RNase HII incision at rNMP errors as a mechanism to direct mismatch repair to the leading strand (Yao et al., 2013). If rNMPs provided such a signal, spontaneous mutagenesis of the Δ*rnhB* and Δ*rnhB;* Δ*uvrA* strains should be higher than 1.5- to 3.9-fold. Furthermore, an MMR deficiency should cause an increase in transition rates for A·T → G·C and G·C → A·T in a variety of sequence contexts. To the contrary, we observed genome-wide that G·C → A·T transitions are increased within a specific sequence context in RNase HII (Δ*rnhB*).

With these results we argue against the model that excision at embedded rNMPs provides a substantial strand discrimination signal for MMR in bacteria

that lack a methylation-directed pathway. For *B. subtilis*, it has been shown that MutL-dependent incision is stimulated by interaction with the replication sliding clamp (β-clamp) (Pillon et al., 2015). Furthermore, MutL variants unable to interact with β-clamp show defects in MMR *in vivo* and fail to nick a linear substrate *in vitro* (Pillon et al., 2015). Studies of the eukaryotic proteins showed that MutLα incision was also stimulated by the replication sliding clamp, and PCNA variants of MutLα impaired for interaction with PCNA prevented MMR *in vitro* and *in vivo* (Genschel et al., 2017). It seems that the strand discrimination mechanism in organisms lacking a methylation signal may instead rely on orientation imparted to the MMR machinery by replication sliding clamps, allowing for stimulation of strand-dependent incision (Pluciennik et al., 2010).

2.7 Materials and Methods

2.7.1 RNase H cleavage assays

The 5' IR dye-labeled substrates with either one or four consecutive rNMPs were prepared by incubating complementary strands (either oJR209 or oJR210 with oJR145) in a 98 °C water bath for 1 minute in a buffer containing 20 mM Tris-HCl pH 8, 100 mM NaCl, 10 mM $MgCl_2$, and 1 mM DTT. Strands were then annealed by slowly cooling the solution to room temperature. Reactions were performed in the same buffer with 1 µM substrate and 200 nM protein for 10 minutes at 30 °C in 10 µL total volume. For NaOH samples, 1 µM of substrate was placed into 300 mM NaOH and incubated at 55 °C for 30 min, followed by neutralization with 2 M Tris-HCl pH 7.5. Reactions were stopped by addition of

10 µL formamide loading dye (95% formamide, 20 mM EDTA, 1% SDS, and 0.01% bromophenol blue), followed by denaturation at 100 °C for 2 min and immediate snap cooling in an ice-water bath. Following cooling, 2 µL of each reaction was electrophoresed in a denaturing 20% urea–polyacrylamide gel followed by visualization with the LI-COR Odyssey imager.

2.7.2 DNA polymerase activity assay

A 5′ IR dye-labeled primer was annealed to the template strand by incubating oJS895 (1 µM) and oJR283 (1.25 µM) together in a 98 °C water bath for 1 minute in a buffer containing 40 mM Tris–acetate pH 7.8, 12 mM magnesium acetate, 200 µM dNTPs, 300 mM potassium glutamate, 3 µM $ZnSO_4$, 2% wt/vol PEG, and 1 mM DTT and then allowing the mixture to cool slowly to room temperature. Reactions were carried out in a 10-µL volume in the same buffer with 100 nM polymerase (Pol I, DnaE, PolC, or Klenow fragment) and 100 nM substrate at 25 °C for 5 min. Reactions were quenched with 10 µL of formamide dye (see above), denatured at 100 °C for 2 min, and immediately snap cooled in an ice-water bath. Products were resolved by electrophoresis through a 17% urea–polyacrylamide gel, followed by visualization with the LI-COR Odyssey imager.

2.7.3 Protein purification

Proteins were isolated as described (Liao et al., 2015). Briefly, cultures with each plasmid were grown in 3 L of terrific broth (1.2% tryptone, 2.4% yeast

extract 72 mM K_2HPO_4, 17 mM KH_2PO_4, 0.4% glycerol) with constant shaking at 37 °C to an OD of ≈ 0.75 and expression was induced by addition IPTG to a final concentration of 250 µM. Cultures were left for an additional 3 hours at 37 °C with constant shaking followed by harvesting of cells with centrifugation. Cell pellets were frozen in liquid nitrogen before further processing. Pellets were thawed and resuspended in 50 mL of lysis buffer (50 mM Tris-HCl pH 8, 300 mM NaCl, 10% sucrose, 10 mM imidazole) and lysed via sonication. Cell debris was pelleted via centrifugation and the supernatant was applied to a 2 mL Ni^{2+}-NTA agarose gravity-flow column. The column was washed (wash buffer: 50 mM Tris-HCl pH 8, 2M NaCl, 10 mM imidazole) and then eluted (elution buffer: 50 mM Tris-HCl pH 8, 50 mM NaCl, 500 mM imidazole). The SUMO protease was added to the eluate, which was dialyzed into SUMO Protease buffer (50 mM Tris-HCl pH 8, 150 mM NaCl, 1 mM DTT) 15-18 hours at 4 °C to remove the affinity tag. The dialyzed protein solution was then applied to a 2 mL Ni^{2+}-NTA agarose gravity-flow column and fractions were collected and analyzed for purity via SDS-PAGE. Fractions containing the desired protein were pooled and each protein was applied to an anion exchange column (GE product number 17-5156-01) followed by elution with a 50-500 mM NaCl gradient. Fractions were analyzed via SDS-PAGE with the pure fractions pooled and concentrated into protein storage buffer (50 mM Tris-HCl pH 8, 150 mM NaCl, and 25% glycerol) and frozen in liquid nitrogen. PolC and DnaE proteins were generous gifts from Dr. Charles McHenry at the University of Colorado (Sanders et al., 2010). The Klenow

fragment used in this work was purchased from New England Biolabs (product number M0210S).

2.7.4 Calculation of conditional mutation rate

Transitions rates for each base and replichore were calculated essentially as described (Schroeder et al., 2016):

$$\frac{M_{b,r}}{R_{b,r}xG}$$

where $M_{b,r}$ is the number of transitions from base b in the reference sequence of replichore r, $R_{b,r}$ is the total number of occurrences of base b in the reference sequence of replichore r, and G is the total number of generations surpassed during the mutation accumulation procedure. Transition rates in each dinucleotide sequence context were calculated essentially as described (Schroeder et al., 2016; Sung et al., 2015):

$$\frac{M_{d}}{R_{d}xG}$$

where M_d is the number of transitions at the 5' position of a given dinucleotide in the lagging strand template, R_d is the total number of occurrences of the given dinucleotide in the lagging strand template, and G is the total number of generations the MA lines underwent during the mutation accumulation procedure.

2.7.5 Logistic regression of sequence context effect on transition rate

To determine the impact of local sequence context on guanosine to adenosine transitions in the lagging strand template, the logit of the probability of obtaining a guanosine to adenosine transition against nucleotide base identity up to five nucleotide positions in the 3′ direction of a guanosine in the lagging strand template:

$$\ln(p/1-p) = \alpha + \beta Base_{+1} + \beta Base_{+2} + \beta Base_{+3} + \beta Base_{+4} + \beta Base_{+5}$$

where p is the probability of guanosine in the lagging strand template having a transition occur and $Base_i$ represents either adenosine, cytidine, thymidine or guanosine at position i 3' to the guanosine in question. Log-odds ratios presented in Figure 2.4B were calculated using adenosine as the baseline.

2.7.6 Motif identification at lagging strand template G to A transitions

A multifasta file containing the local sequence from all 162 sites where a G in the lagging strand template was transitioned to an A in Δ*rnhB* was prepared along with a multifasta file containing local sequence contexts at 1000 randomly selected G in the lagging strand template to control for random local sequence context. Specifically, the sequence contexts for motif detection extended one position 5' through six positions 3' to the transition site, from the perspective of the lagging strand template. MEME was applied in discriminative mode using the 162 sequence contexts surrounding G→A transitions in the lagging strand template as the primary sequences and the file containing 1000 negative control sequence contexts as the control sequences with the minimum possible motif

width set to two (Bailey and Elkan, 1994). All other parameters were default values. The resulting MEME logo was trimmed to include only positions one through four nucleotides 3' to the transition site.

2.7.7 RNase HII nicked substrate extension

A 5' IR dye-labeled, 40-mer oligonucleotide containing one embedded rNMP (oJR297) was annealed to a complementary 40-mer oligonucleotide (oJR298) as described (see polymerase activity assay). RNase HII reactions were performed in a 200 µL reaction in the same buffer with 1 µM substrate and 200 nM RNase HII for 45 minutes at 37 °C. Reactions were stopped by heating the sample to 95 °C for 5 minutes. 45 µL of this reaction was then placed into three different microcentrifuge tubes containing 5 µL of either 1 µM Pol I, DnaE, or PolC and mixed by pipetting. Reactions were incubated at 25 °C and 5 µL of the reaction was removed and quenched with formamide loading dye at 1, 5, 20, and 60 minutes and immediately placed on ice. After 60 minutes, NaOH was added to a final concentration of 300 mM and incubated at 55 °C for 45 minutes to verify the rNMP was completely hydrolyzed. As a control the same extension reaction was performed without RNase HII cleavage. All reactions were then denatured at 100 °C for two minutes and immediately snap cooled in an ice water bath. Products were resolved by electrophoresis through a 17% urea-polyacrylamide gel and visualized with a LI-COR Odyssey imager. Three independent reactions were performed. The percent full product extended was calculated by dividing the extended product in lane 8 by the cleaved substrate

band in lane 4 and multiplying by 100. Plots show the mean and 95% confidence interval of all three reactions at each time point with each polymerase.

2.7.8 Sequence context-dependent mutagenic resynthesis

A 5' IR dye-labeled primer was annealed as described above (see polymerase activity assay) to a complementary template strand intended to modeling an NER-processed rNMP incorporation at a 5′-GCCTT-3′ sequence (oJR895 and oJR283). Extension reactions were performed in 10 µL reactions in the same buffer. Each reaction contained 1 µM substrate without polymerase or with Pol I, DnaE, or PolC. These reactions were performed with all dNTPs at 200 µM, or in the absence of dCTP. Reactions were stopped with 10 µL of formamide loading dye at 5 and 45 minutes. Reactions were denatured at 100 °C for two minutes and immediately snap cooled in an ice water bath. Products were resolved via electrophoresis through 17% urea polyacrylamide gels and visualized with a LI-COR Odyssey imager. Two independent reactions were performed and the relative primer extension was calculated by dividing the intensity of the band from 5 minute extension with no dCTP by the intensity of the band from the 5 minute extension with all dNTPs and multiplying by 100. Bar height and error bars represent the mean and range, respectively, of two independent experiments.

2.7.9 Gene reporter sequencing

A total of 48 independent 100 µL cultures of JWS328, JWS329, JWS330 or JWS331 were grown in 96-well plates at 37 °C with constant shaking in S750 minimal medium supplemented with 0.1 µM final concentration each of tryptophan and phenylalanine and 200 µM final concentration thymidine. The entirety of each culture was plated on S750 minimal medium agar plates supplemented with 0.1 µM final concentration each of tryptophan and phenylalanine, 200 µM final concentration thymidine, 34 µM final concentration trimethoprim and 0.2% (w/v) final concentration casaminoacids. Plates were incubated at 37 °C for two days, at which time fifteen colonies from each plate were inoculated into 96-well plates containing LB supplemented with 200 µM thymidine and 34 µM trimethoprim, for a total of 720 trimethoprim-resistant colonies of each genotype. 96-well plates were incubated with constant shaking for 8 hours at 37°C. Plates were stored at 4 °C overnight. Cultures were diluted 16-fold in water. A Phusion Hot Start II DNA polymerase PCR master mix was prepared for amplification of the *thyA* locus using primers oJS858 and oJS859. For amplification 19 µL of the PCR master mix was deposited into each well of 96-well plates. For template, 1 µL of each diluted culture was added to each PCR reaction and mixed by pipetting. PCR amplification was performed and products were purified using 1x volume of MagNA bead slurry consisting of 2% (v/v) Sera-Mag SpeedBeads, 18% (w/v) PEG-8000, 1 M NaCl, 10 mM Tris HCl, 1 mM EDTA, 0.05% (v/v) Tween 20 (Baym et al., 2015; DeAngelis et al., 1995; Rohland and Reich, 2012). The concentration of each purified PCR product was

determined using the QuantIT – High sensitivity kit (Fisher Scientific, Q33120). The fifteen PCR products that originated from colonies from a common plate were pooled at an equimolar concentration to achieve a combined final DNA concentration of 0.5 ng/µL. Sequencing libraries were prepared from the pooled PCR products according to the method of Baym et al. (Baym et al., 2015). DNA was sequenced at the University of Michigan DNA Sequencing Core using an Illumina Hi-seq 4000.

Paired-ended 150 base reads were trimmed to remove Illumina sequencing adapters. Alignment to the *thyA* open reading frame and the 500 bases flanking each side was performed using bwa (Li and Durbin, 2009) and alignment files prepared using samtools (Li et al., 2009). Variants present in greater than 2% of the aligned reads at a given location were detected using Freebaye (Garrison and Marth, 2012). Potential variants were then stringently filtered, retaining only variants with a quality score greater than 500 for further analysis. The final list of variants included 181 from JWS328, 266 from JWS329, 185 from JWS330, and 193 from JWS331. Testing for differences in mutation spectra was performed using a Chi-square test.

2.7.10 Mutation accumulation line protocol

MA lines on Δ*rnhB*, sequence alignments, and variant detection and conditional mutation rate were performed as previously described (Schroeder et al., 2016). Wild-type (*B. subtilis* PY79) MA line data were previously published (Schroeder et al., 2016).

2.7.11 Data-sharing plan

High-throughput sequencing data used in this study have been deposited in the Sequence Read Archive under accession no. SRP117359. Equations are described in the Materials and Methods; all code used is available upon request.

2.7.12 Statistical analysis

Statistical analysis and plotting were performed using the statistical computing software R. Throughout this work, *** denotes $P \leq 0.001$, ** denotes $P \leq 0.01$, and * denotes $P \leq 0.05$.

2.8 Notes and Acknowledgements

We thank Dr. Charles McHenry for his generous gifts of DnaE and PolC, Heather Schroeder for help with statistical analyses, and Peter Burby for constructing the Δ*uvrA* strain. This work was supported in part by NIH National Research Service Award T32 GM007544, NIH Cellular Biotechnology Training Grant (T32 GM008353) and a pre-doctoral fellowship from Rackham Graduate School, University of Michigan. Additionally, this work was supported by NIH Grant R01 GM107312.

This work is published in the Proceedings of the National Academy of Sciences (Schroeder et al., 2017). Jeremy Schroeder, Lyle Simmons, Will Hirst and I designed and performed the research. Jeremy Schroeder and William Hirst carried out MA lines and Jeremy Shroeder performed all sequencing data

analysis. He also performed the trimethoprim mutation spectrum assay and analysis. Lyle Simmons performed Rifampin resistance assays. I purified all RNase H enzymes as well as Pol I and carried out all *in vitro* RNase H and DNA polymerase experiments and analysis. Jeremy Schroeder, Lyle Simmons, Mike O'Donnell and I analyzed the data and wrote the paper. None of the Authors have any conflicts of interest.

2.9 Figures and Tables

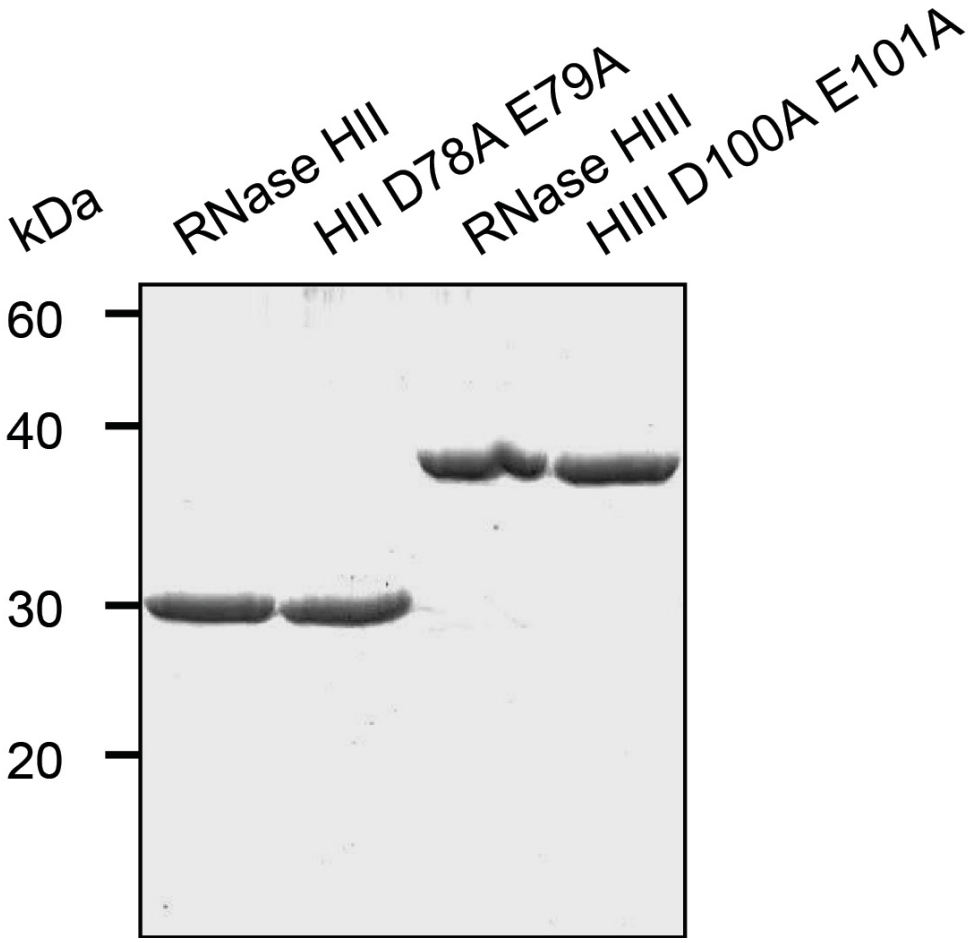

Figure 2.1 SDS-PAGE of RNase HII, HIII and catalytically inactive variants. Shown is an SDS-PAGE with 2 µg of purified RNase HII, RNase HIII, RNase HII D78A E79A, and RNase HIII D100A E101A followed by coomassie blue staining.

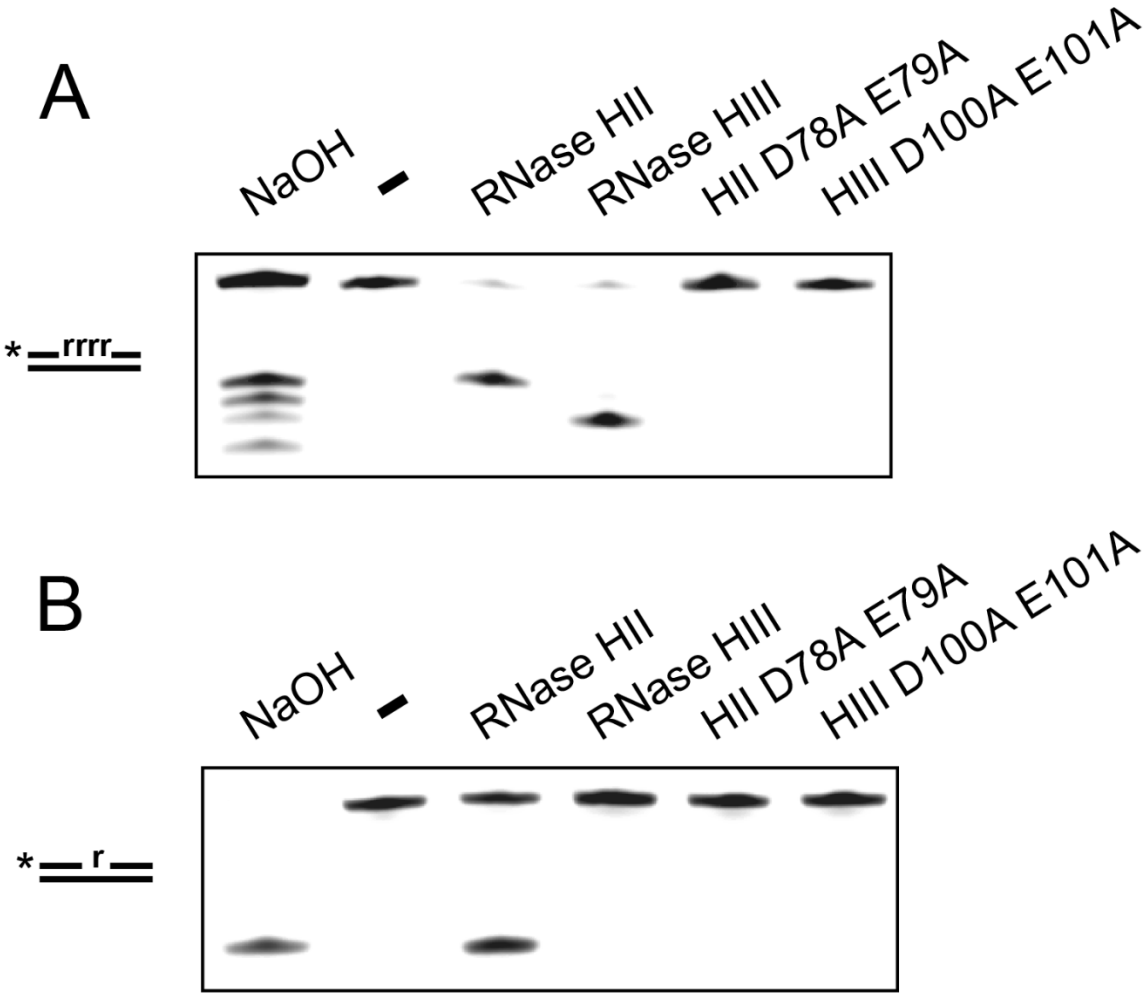

Figure 2.2 RNase HII cleaves a 2.single rNMP in hybrid substrates. (A and B) A 5′-labeled substrate with four rNMPs (A) or a single rNMP (B) in DNA was incubated under the indicated conditions for 5 min followed by electrophoresis in a 20% denaturing urea–PAGE.

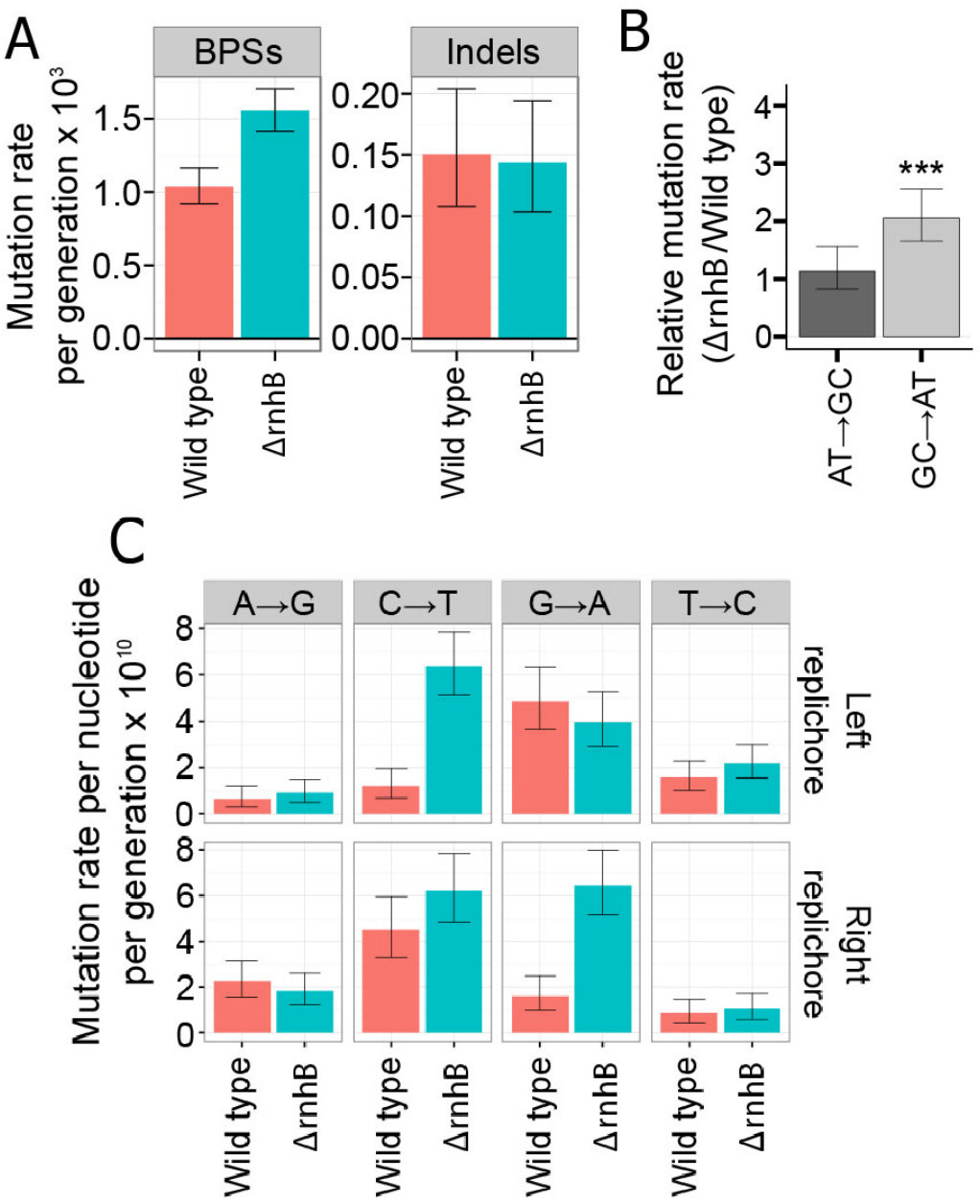

Figure 2.3 Persistent ribonucleotides cause strand-dependent transitions. (A) Base-pair substitution rate increases in the absence of RNase HII. (B) GC → AT transition rates are increased approximately two-fold in Δ*rnhB* relative to wild-type. (C) Bar plots show the mutation rates of the indicated transitions from the perspective of the reference sequence. Persistent ribonucleotides cause GC → AT transitions in a strand-dependent manner.

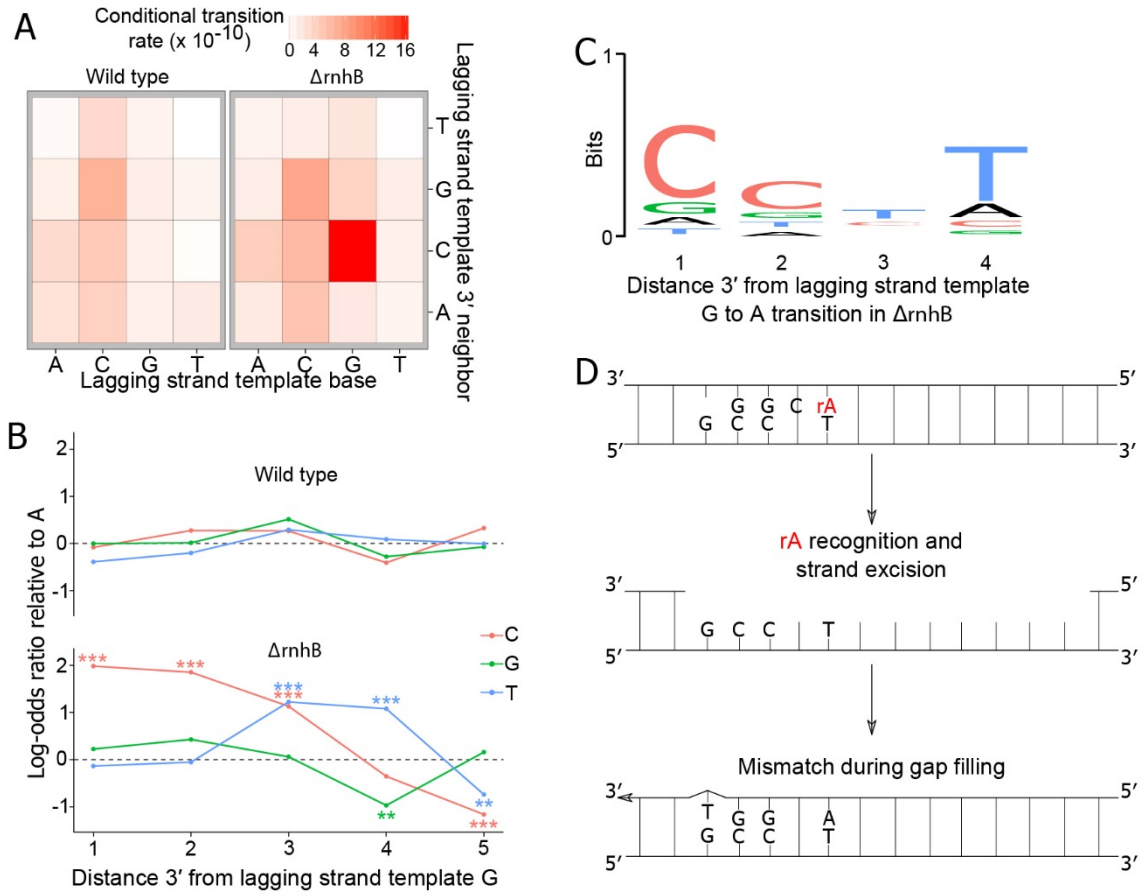

Figure 2.4 Persistent ribonucleotides cause context-dependent transitions. (A) Heat map showing the transition rates for the 16 possible dinucleotides normalized to the abundance of each dinucleotide in the lagging-strand template. If guanosine is followed 3′ by cytidine in the lagging-strand template, the guanosine is more likely to undergo transition. (B) Logistic regression to determine the effect of sequence context 3′ to guanosine on transition occurrence at a guanosine in the lagging-strand template. All lagging-strand template guanosine positions were included in the regression analysis, and the log-odds of guanosine undergoing transition to adenosine were regressed against nucleotide identities up to five positions 3′ to the guanosine. The log-odds ratios are presented with adenosine as the baseline. (C) A motif logo was generated using the four positions 3′ to all guanosine-to-adenosine transitions in the lagging-strand template in Δ*rnhB*. The motif that is enriched 3′ to guanosine-to-adenosine transitions in Δ*rnhB* is 5′-CC (T/C) T-3′. (D) A model for ribonucleotide-mediated mutagenesis in the lagging-strand template 5′-GCC (T/C) T-3′

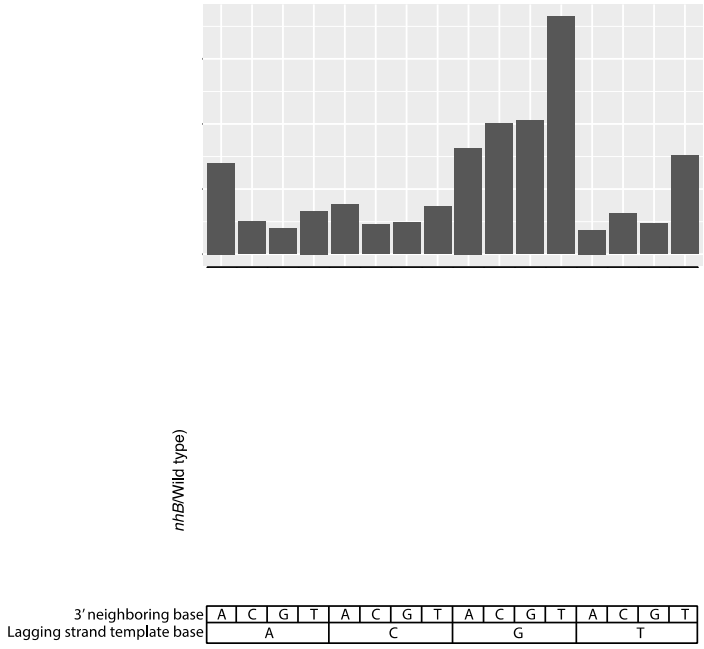

Figure 2.5 Context-dependence of transition rate in the lagging strand template is specific to 3' neighboring bases. (A) A heat map displaying the transition rate for a nucleotide with the base given on the x-axis with the 5′ neighboring base indicated in the y-axis. (B) A bar plot displaying the relative transition rate in Δ*rnhB*/wild type from the indicated base with the given 5′ neighbor showing the increased mutation rate at guanosine nucleotides is largely independent of 5′ neighboring nucleotide identity. (C) A bar plot showing the relative transition rate in Δ*rnhB*/wild type from the indicated base with the given 3′ neighbor shows the increase in transitions at guanosine nucleotides is highly dependent on 3′ sequence context.

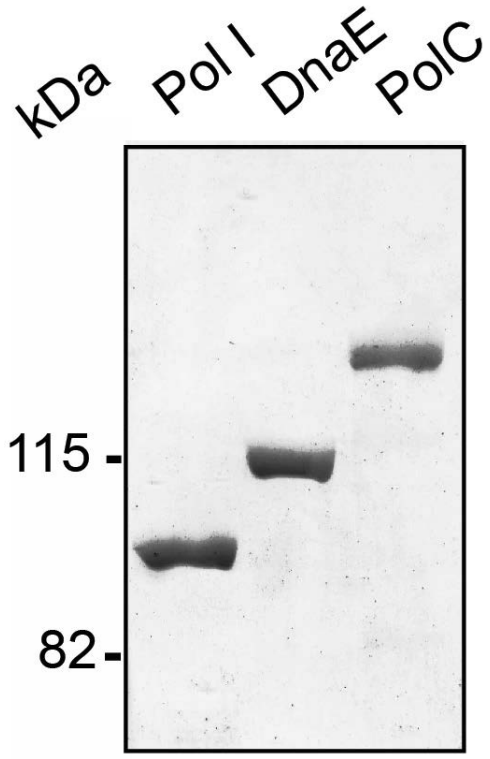

Figure 2.6 Purified DNA polymerases. Coomassie blue-stained 8% SDS-polyacrylamide gel with 2 µg each of Pol I, DnaE and PolC.

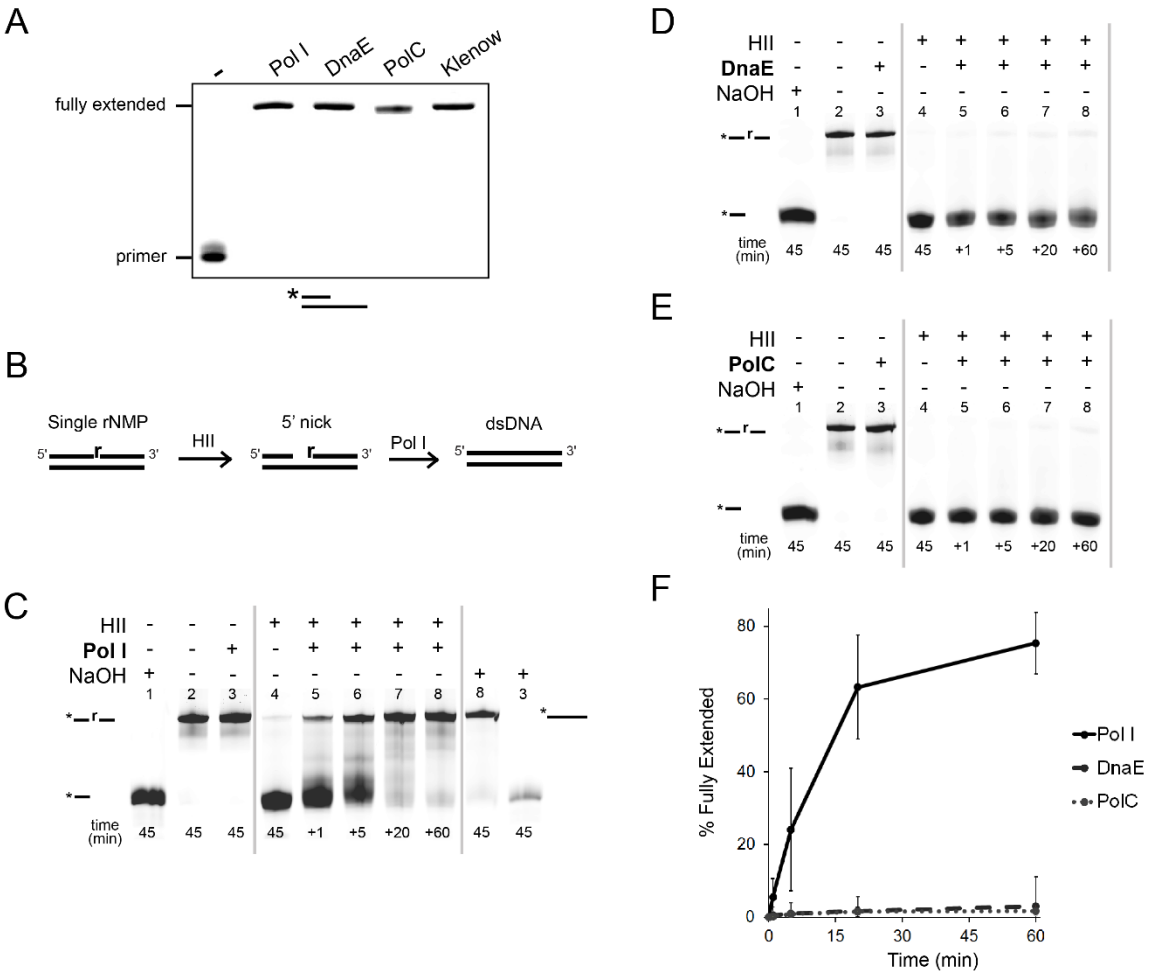

Figure 2.7 Pol I removes and replaces a single rNMP after processing by RNase HII. (A) Primed DNA substrate was incubated with Pol I, DnaE, PolC, or Klenow fragment. Replication products were resolved using 17% urea–PAGE. (B) Schematic representation of substrate, RNase HII treatment, and subsequent DNA synthesis used in C–E. (C–E) A 5′-labeled RNA/DNA hybrid substrate was incubated with 300 mM NaOH, no protein, or the indicated DNA polymerase: (C) Pol I, (D) DnaE, or (E) PolC. (C) Pol I, (D) DnaE, or (E) PolC was then added to the RNase HII reaction and stopped at 1, 5, 20, and 60 min. (C) Samples from lanes 8 and 3 were then incubated with 300 mM NaOH for 45 min followed by resolution using a denaturing urea–PAGE. (F) Mean percentage of substrate extended is presented for triplicate samples for all polymerases at the indicated times. Error bars represent 95% CI.

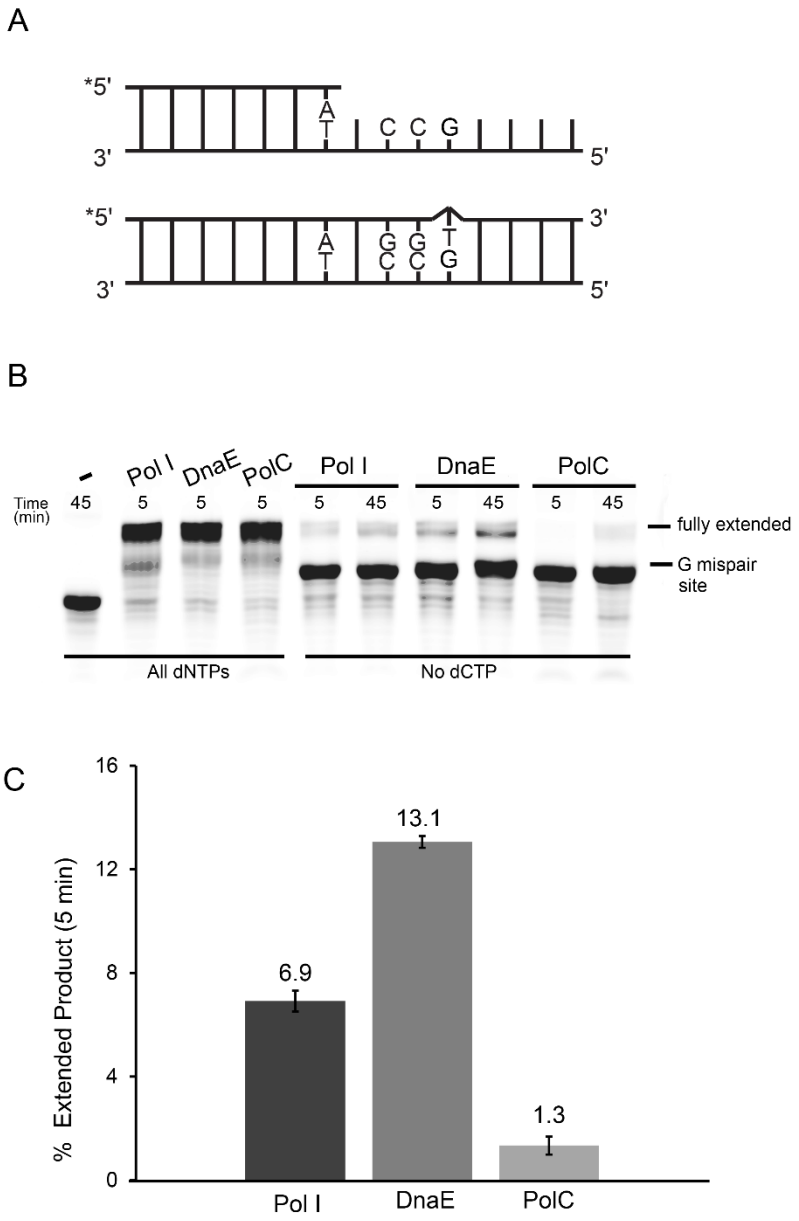

Figure 2.8 DnaE and Pol I mispair T across from G in a 5′-GCCTT-3′ sequence context. (A, Top) A schematic of the 5′ end-labeled substrate designed to mimic NER-mediated excision of the rA-containing strand across from 5′-GCC (C/T) T-3′. (A, Bottom) The fully extended primer in the absence of dCTP. (B) The substrate shown in A was treated with no protein, Pol I, DnaE, or PolC with all four dNTPs or with dCTP omitted from the reactions for the indicated times followed by electrophoresis through a 17% urea–PAGE. (C) A bar plot showing the mean of two independent experiments with the proportion of primer fully extended after 5 minutes in the absence of dCTP relative to reactions in which all nucleotides were present. Bar height and error bars represent the mean and range, respectively.

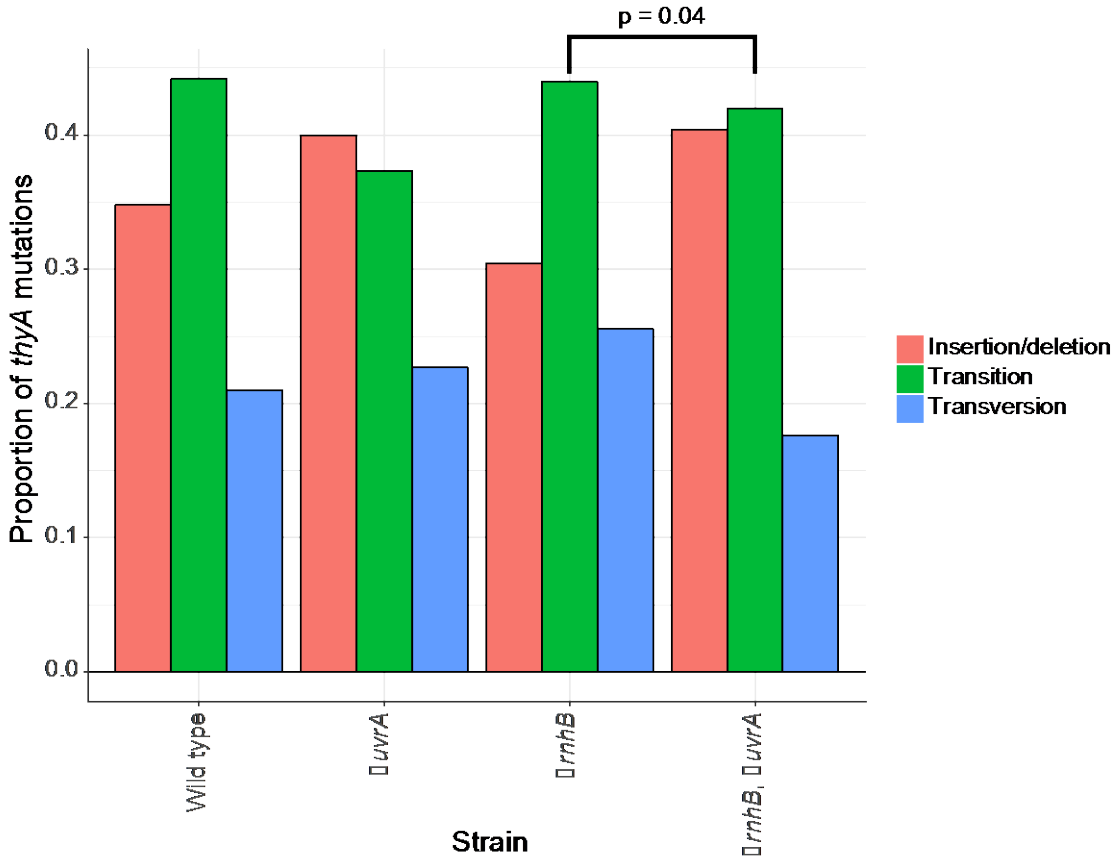

Figure 2.9 Mutation spectra of cells with *rnhB* and *uvrA* deletions. Shown is a bar plot with the mutation spectra for the *thyA* gene in trimethoprim-resistant mutants of the indicated strains in a *thyB::erm* genetic background. The spectrum of Δ*rnhB* is significantly different from that of the Δ*rnhB*, Δ*uvrA* strain. The *thyA* reporter represents a selection for mutations at a specific locus with a limited sequence context. In contrast, mutation accumulation lines measure mutations that occur in the near absence of selection sampling sequence context on a genome wide scale. Therefore, the spectra observed at the *thyA* locus differ from what we observed genome-wide in the mutation accumulation lines because spectra are sequence context dependent. The p-value is the result of a Chi-square test. All variants are listed in Dataset 2.

Table 2.1 Overview of the variants detected in mutation accumulation line

	Wild type [a]	Δ*rnhB*
Number of lines	75	81
Total mutations	295	462
BPSs	254	420
Indels	41	42
Transitions	198	363
AT to GC	74	90
GC to AT	124	273
Transversions	56	57
AT to CG	25	21
AT to TA	10	13
GC to CG	4	3
GC to TA	17	20
Generations per line	3636	3610
Mutation rate per genome replication ($x10^{3}$) $[95\% CI]_b$	1.1 [1.0-1.2]	1.6 [1.4-1.7]
Mutation rate per nucleotide ($x10^{10}$) $[95\% CI]_b$	2.7 [2.4-3.0]	4.0 [3.6-4.3]

[a] Wild type data were published in (Schroeder et al., 2016).

Table 2.2 Mutation rate of cells with *rnhB* and *uvrA* deletions

Strain	No. of cultures	Mutations per culture [±95%CI]	Mutation rate (Mutations per generation 10^{-8} ± [95%CI]	Relative mutation rate
Wild Type	30	0.5 [0.3-0.5]	1.15 [0.7-1.2]	1
Δ*rnhB*	30	0.9 [0.5-1.2]	1.9 [1.0-2.5]	1.7
ΔuvrA	27	1.0 [0.5-1.5]	1.8 [0.8-2.6]	1.6
Δ*rnhB, ΔuvrA*	30	1.6 [0.7-2.4]	4.45 [1.9-6.7]	3.9

Mutagenesis assays were done as described using Rif^R as an indicator. Mutation rate and mutation per culture were calculated using the Ma-Sandri-Sarkar Maximum Likelihood Estimator with the web-based tool FALCOR (Hall et al., 2009).

Table 2.3 Oligonucleotides used in this study

Name	Sequence
oJR46	TCGAGCACCACCACCACCACCACTGAG
oJR47	ACCTCCAATCTGTTCGCGGTGAGCCTCAATAATATCG
oJR88	CCGCGAACAGATTGGAGGTGTGAATACATTAACCGTAAA GGACATTAAAGACC
oJR89	TGGTGGTGGTGGTGCTCGATTATCTGAAAGATTGAACAG GAGCG
oJR90	CCGCGAACAGATTGGAGGTGTGTCCCATTCAGTGATAAA AGTATCG
oJR91	TGGTGGTGGTGGTGCTCGACTATGAACGTTTTTTATCAG CAAGGCG
oJR94	GTTGcCGcGGTCGGCCG
oJR95	GACCgCGgCAACACCTGCAATC
oJR96	TCTGcCGcAGTCGGAACCGG
oJR97	GACTgCGgCAGAACCGATAACAGACATTCC
oJR102	CCGCGAACAGATTGGAGGTATGACGGAACGAAAAAAATT AGTGC
oJR103	TGGTGGTGGTGGTGCTCGATTATTTCGCATCGTACCAAG ATGG
oJR145	GCAGAGCTAGCTTACGATCG
oJR209	/5IRD800CWN/CGATCGTAArGCTAGCTCTGC
oJR210	/5IRD800CWN/CGATCGTArArGrCrUAGCTCTGC
oJR283	AACTTAGCCTTGCGCATCAGCTGCAG
oJS895	/5IRD800CWN/CTGCAGCTGATGCGCA
oJR297	/5IRD800CWN/GTACTCGGTGATCCGTACATrGGCGCATC AGCTGCAGTAG
oJR298	CTACTGCAGCTGATGCGCCATGTACGGATCACCGAGTAC
oJS858	ACCGCAATATCAAACCATTTCGT
oJS859	TGGTCATAGTTCGTTTCTTCAGC

Lowercase letters in oligonucleotide sequences indicate sites of mutagenesis for catalytic mutants. /5IRD800CWN/ preceding the oligonucleotide sequence indicates the oligonucleotide was 5′ end labeled by Integrated DNA Technologies with IR Dye CW800 (NHS ester) for imaging on a Li-COR imaging system.

Table 2.4 Plasmids used in this study

Name	Vector	Insert
pJR17	pE-SUMO	*rnhB*
pJR18	pE-SUMO	*rnhB (D78A, E79A)*
pJR19	pE-SUMO	*rnhC*
pJR20	pE-SUMO	*rnhC (D100A, E101A)*
pJR22	pE-SUMO	*polA*

Table 2.5 Strains used in this study

Name	Relevant Genotype	Source
PY79	Wild type SPβ^{o}	(Youngman et al., 1984)
JWS105	Δ*rnhB*	(Yao et al., 2013)
PEB307	Δ*uvrA*	
JWS305	Δ*rnhB,* Δ*uvrA*	This work
JWS328	*thyB::erm*	This work
JWS329	Δ*rnhB*, *thyB::erm*	This work
JWS330	Δ*uvrA*, *thyB::erm*	This work
JWS331	Δ*rnhB*, Δ*uvrA*, *thyB::erm*	This work

CHAPTER III

III. Substrate specificity for bacterial RNase HII and HIII is influenced by metal availability

3.1 Abstract

We tested the activity of four predicted RNase H enzymes including two RNase HI-type enzymes in addition to RNase HII (RnhB) and RNase HIII (RnhC) on several RNA-DNA hybrid substrates with different divalent metal cations. We found that the two RNase HI-type enzymes YpdQ and YpeP failed to show activity on the three substrates tested. RNase HII and RNase HIII cleaved all substrates tested although activity was dependent on the divalent metal cation made available. We show that *B. subtilis* RNase HII and RNase HIII are both able to incise 5' to a single rNMP. We show that RNase HIII incision at a single rNMP occurs most efficiently with Mn^{2+}, an activity we found to be conserved among other Gram-positive RNase HIII enzymes. Characterization of RNase HII and HIII with metal concentrations in the physiological range showed that RNase HII can cleave at single rNMPs embedded in DNA while RNase HIII is far less

effective. Further, using metal concentrations within physiological range, RNase HIII efficiently cleaved longer RNA-DNA hybrids lacking an RNA-DNA junction while RNase HII is much less effective. Phenotypic analysis shows that cells with an *rnhC* deletion are sensitive to hydroxyurea (HU). In contrast, cells with an *rnhB* deletion show wild type growth in the presence of HU supporting the hypothesis that RNase HII and HIII have distinct substrate specificities *in vivo*. This work demonstrates how divalent metal availability influences substrate recognition and activity of RNase HII and HIII providing insight into their functions *in vivo*.

3.2 Significance

Ribonuclease H (RNase H) represents a class of proteins that cleave RNA-DNA hybrids helping resolve R-loops and Okazaki fragments as well as initiating the process of ribonucleotide excision repair (RER). We investigated the activity of four *Bacillus subtilis* RNase H enzymes finding that only RNase HII and HIII have activity and that their substrate preference is dependent on divalent metal availability. To understand factors that contribute to RNase HII and RNase HIII substrate preference, we show that in the presence of metal concentrations within physiological range, RNase HII and HIII have distinct activities on different RNA-DNA hybrids. This work provides insight into how RNase HII and HIII repair the broad range of RNA-DNA hybrids that form in Gram-positive bacteria.

3.3 Introduction

RNA is often covalently incorporated into duplex DNA by a variety of processes. For example, DNA replication initiates from an RNA primer synthesized by primase incorporating approximately 10-15 consecutive ribonucleotides to begin each Okazaki fragment (Rowen and Kornberg, 1978a, 1978b). The lagging strand RNA primer is removed during replication as the lagging strand matures to become a continuous strand of DNA (Ogawa and Okazaki, 1984). Further, single ribonucleotides are misincorporated into genomes by DNA polymerases during replication requiring replacement of the ribonucleotide with the appropriate deoxyribonucleoside triphosphate (dNTP) (Nick McElhinny et al., 2010a). In addition to RNA covalently nested in genomic DNA, RNA can also persist base-paired with DNA. RNA transcripts can base-pair with DNA displacing a single strand of DNA forming an R-loop, a structure lacking an RNA-DNA junction [for review (Santos-Pereira and Aguilera, 2015)]. R-loops can form by several mechanisms including aborted transcription. If not removed, R-loops can compromise genome integrity (Lang et al., 2017; Santos-Pereira and Aguilera, 2015).

In the case of single ribonucleoside triphosphate (rNTP) misincorporations, it is becoming clear that sugar errors are abundant in genomic DNA. In eukaryotes it has been shown that replicative DNA polymerases form sugar errors at an alarmingly high rate. For example, the budding yeast *Saccharomyces cerevisiae* incorporates approximately 10,000 rNMPs in place of their cognate dNMP per round of genome replication while in mouse cells the

error rate is 1 million rNMPs per replication cycle (Nick McElhinny et al., 2010a; Reijns et al., 2012).

Bacterial DNA polymerases have also been shown to incorporate rNMP errors in place of dNMPs. *Escherichia coli* DNA polymerase III was shown to incorporate one sugar error every 2,300 base-pairs replicated *in vitro* using primed ΦX174 DNA as a template (Yao et al., 2013). This work estimates approximately 2,000 rNMP errors would be incorporated per round of replication for the *E. coli* chromosome (Yao et al., 2013). In addition to the rNTP error rate by eukaryotic and bacterial replicative machines many other DNA polymerases including bacterial UmuD′$_2$C, DinB, and eukaryotic Pol μ incorporate rNTPs in place of dNTPs adding to the need for rNMP removal from genomic DNA (McDonald et al., 2012; Nick McElhinny et al., 2010a; Ordonez et al., 2014; Vaisman et al., 2013, 2014). The translesion polymerase-based rNMP errors likely result in polymers of four or more rNMPs embedded in DNA providing a substrate differing from single misincorporation events (McDonald et al., 2012; Ordonez et al., 2014; Vaisman et al., 2013, 2014).

If rNMP errors are not corrected then genome stability can be compromised (Pizzi et al., 2015; Reijns et al., 2012). The 2′-OH is reactive and can cause breaks in the DNA leaving 2′, 3′ cyclic phosphates (Oivanen et al., 1998). The 2′, 3′ cyclic phosphate is recalcitrant to ligation requiring removal or healing of the cyclic phosphate end prior to ligation (Das and Shuman, 2013; Das et al., 2014). Furthermore, rNMPs in DNA have been shown to slow DNA synthesis for both the *E. coli* and *S. cerevisiae* replicative DNA polymerases

when the rNMP is present in the template strand (Yao et al., 2013). Recent studies of *B. subtilis* cells found that when rNMPs accumulated in genomic DNA cells suffered a mutagenic cost (Yao et al., 2013). Therefore, a failure to repair rNMPs has the potential to cause DNA breaks, mutagenesis, and slow replication fork progression, all of which have the potential to impact genome integrity and evolution.

The proteins responsible for removing RNA from DNA are RNase H enzymes [for review (Tadokoro and Kanaya, 2009)]. RNase H enzymes fall into two classes. Class 1 is represented by RNase HI. These enzymes have been extensively characterized *in vitro* to cleave within RNA-DNA hybrids with four or more consecutive rNMPs (Cerritelli and Crouch, 2009). Class 2 enzymes encompass RNase HII and RNase HIII. Classically, RNases HII have been shown to cleave 5′ to single embedded rNMPs or 3′ to the first rNMP at an RNA-DNA junction when multiple rNMPs are present (Haruki et al., 2002). Traditionally, RNase HIII enzymes have been described to cleave RNA-DNA hybrids with three or more consecutive rNMPs much like RNase HI even though RNase HIII is more closely related to RNase HII (Ohtani et al., 1999b). Phylogenic analysis shows that most organisms have RNase HI and HII, or HII and HIII although some archaeal organisms have only RNase HII (Ohtani et al., 1999a). Because very few organisms contain RNase HI and HIII it has been suggested that these genes are functionally redundant and mutually exclusive due to their overlapping biochemical activities (Kochiwa et al., 2007). However, recently it was shown that *E. coli* RNase HI is capable of cleaving at a single

rNMP (Tannous et al., 2015) and that RNase HIII from *Chlamydophila pneumonia* is also able to cleave at single rNMPs in DNA in the presence of Mn^{2+} (Lu et al., 2012a). Therefore, although RNase H enzymes have been well-characterized biochemically, recent work suggests that these enzymes may have more plasticity in substrate recognition than initially described. Further, for organisms like *B. subtilis* and other Firmicutes, which contain RNase HI, HII and HIII it is unclear if HI and HIII are both functional (Kochiwa et al., 2007) or how metal binding regulates the activity of RNase HII and HIII (Lu et al., 2012a; Tannous et al., 2015).

In this work, we investigate the ability of four RNase H enzymes from *B. subtilis* to incise different RNA-DNA hybrids. Our work demonstrates that substrate specificity of RNase HII and HIII enzymes is influenced by divalent metal availability. Furthermore, we identified a previously undescribed phenotype for *B. subtilis* cells with an *rnhC* deletion. Our results provide insight into how substrate specificity may be achieved by RNase HII and RNase HIII *in vivo*.

3.4 Results

3.4.1 *B. subtilis* has four putative RNase H enzymes with two that show activity *in vitro*

Prior work identified four *B. subtilis* genes encoding putative RNase H enzymes based on their conserved primary structures (Fukushima et al., 2007). These genes include *rnhB*, *rnhC*, *ypdQ* and *ypeP.* RNase HII (RnhB), RNase HIII (RnhC) and YpdQ have been characterized biochemically (Haruki et al., 2002;

Ohtani et al., 1999b). While RNase HII and RNase HIII have activity on RNA-DNA hybrids, RNase H activity was not observed for YpdQ and it was suggested that YpdQ lacked the required amino acids for catalysis (Ohtani et al., 1999b) (Fig. 1A). YpeP has not been characterized biochemically although overexpression of this protein has been shown to reduce a cell elongation phenotype described for cells with the Δ*rnhB*, Δ*rnhC,* Δ*ypeP* alleles (Fukushima et al., 2007). Therefore, because the contribution of *B. subtilis* RNase HII, HIII, YpdQ and YpeP is uncertain, we choose to examine the activity of all four proteins under the same experimental conditions *in vitro*.

We purified each protein using a His_6- SUMO tag ("Materials and Methods"). Following cleavage with SUMO protease RNase HII, RNase HIII, YpeP and YpdQ were isolated without any additional tag. Each protein was purified to homogeneity (Figure 3.1). We began by assaying for *in vitro* RNase H activity on an RNA-DNA hybrid of 20 rNMPs (Figure 3.1C). All four proteins were assayed on an RNA-DNA hybrid lacking an RNA-DNA junction with three different divalent metal cations. Each protein was added to the reaction followed by resolution by denaturing urea-PAGE to detect fragment cleavage. The substrate was also hydrolyzed using alkaline conditions with NaOH to control for cleavage and show the location of each rNMP within the substrate ("Materials and Methods"). RNase HII showed poor incision of the substrate with Mn^{2+} and no incision with the other metals (Figure 3.1C left panel). This activity becomes more apparent if enzyme concentration is increased (Figure 3.2A left panel). RNase HIII incised the RNA-DNA substrate with all three metals showing

approximately equal activity with Mg^{2+} and Co^{2+} (Figure 3.1C left panel). We did not detect activity with YpeP or YpdQ with any of the three divalent metals tested even with a 125-fold increase in protein concentration relative to RNase HII and HIII (Figure 3.1C right panel).

As a second test for substrate specificity we assayed for activity on an RNA-DNA hybrid substrate with four rNMPs embedded into DNA. RNase HII incised the substrate with all three metals, and had approximately equal activity with Mg^{2+} and Co^{2+} while RNase HIII showed an equal amount of incision with all three metals (Figure 3.1D left panel). Again, we failed to observe activity with YpeP or YpdQ on the substrate with four embedded rNMPs (Figure 3.1D right panel).

As a third test for substrate specificity, we assayed for activity on an RNA-DNA hybrid with a single embedded rNMP representing a replicative DNA polymerase error (Yao et al., 2013). We increased protein concentration from 4 nM to 50 nM because during the incubation time used, RNase HII and HIII did not incise at a single rNMP with a protein concentration of 4 nM. RNase HII incised with all three divalent metals, preferring Mn^{2+} and Co^{2+} and HIII preferred Mn^{2+} and Co^{2+} showing very little activity with Mg^{2+} (Figure 3.1E left panel). Again, YpeP and YpdQ failed to show activity on a substrate with a single embedded rNMP (Figure 3.1E right panel). To determine if purified YpdQ and YpeP are folded and contain secondary structure, we generated a circular dichroism (CD) spectrum for each protein showing that both are folded and adopt a β-sheet rich secondary structure (Figure 3.3). Our result showing that YpdQ

lacks activity is consistent with the phylogenic study suggesting that RNase HI and HIII are mutually exclusive (Kochiwa et al., 2007; Ohtani et al., 1999b). We cannot exclude the possibility that YpeP or YpdQ are functional *in vivo* under a condition we could not replicate *in vitro*, however under the conditions tested here we did not detect incision on any RNA-DNA hybrid substrate with either YpeP or YpdQ.

As controls, we were unable to detect activity with RNase HII and RNase HIII on any substrate without a metal demonstrating that incision is metal-dependent (Figure 3.2B). Moreover, RNase HII and HIII have a conserved DEDD or DEDE catalytic motif responsible for divalent metal coordination [for review (Tadokoro and Kanaya, 2009)]. As an added control for incision, we purified variants of RNase HII and HIII with the first two residues (DE) mutated to alanines (AA). We assayed each variant and showed that the catalytically inactive proteins were unable to incise any of the substrates tested even with a divalent metal cation supplied in the reaction (Figure 3.2C).

With these results, we conclude that RNase HII and RNase HIII are active for incision of RNA-DNA hybrids and likely represent the enzymes critical for RNase H activity and ribonucleotide excision repair (RER) in *B. subtilis*. We also conclude that RNase HII and HIII have metal dependent activity on each of the three substrates examined.

3.4.2 RNases HIII from Gram-positive bacteria incise at a single rNMP with manganese

To understand how ribonucleotide errors are repaired, RNase HII and RNase HIII were tested more comprehensively on a substrate with a single rNMP embedded in DNA for the purpose of modeling a single DNA polymerase misincorporation event. Each *B. subtilis* protein was titrated into reactions containing either Mg^{2+} (1 mM) or Mn^{2+} (1 mM) followed by resolution of the 5′ end labeled product on a denaturing PAGE. We used a lower concentration of metals in these reactions as oppose to the 2 mM concentration used above to more easily distinguish differences in activity. With Mg^{2+} only RNase HII was capable of substantial incision of a substrate with a single rNMP (Figure 3.4A top panel). To control for RNase H activity in our protein preparations we again used purified variants of RNase HII and RNase HIII predicted to be catalytically inactive after changing the conserved DE motif to AA ("DE/AA"). Neither catalytically altered variant showed activity on the single rNMP-bearing substrate with Mg^{2+} or Mn^{2+} present in the reaction (Figure 3.2C).

It has been shown that RNase HIII from *Chlamydophila pneumoniae* is able to incise at a single rNMP in DNA with Mn^{2+} (Lu et al., 2012b). Further, this work suggested that RNase HIII proteins with a G (R/K) G motif, commonly found among RNase HII enzymes, was critical for conferring activity on DNA substrates with a single rNMP (Lu et al., 2012b). *B. subtilis* RNase HIII lacks the G (R/K) G motif (Figure 3.4B). Nevertheless, we found that *B. subtilis* RNase HIII incised a substrate with a single rNMP with Mn^{2+} or Co^{2+} (Figure 3.1E) therefore, we

investigated this result more comprehensively. In Figure 3.4A (bottom panel) we show that *B. subtilis* RNase HIII was able to cleave at a single rNMP in DNA with Mn^{2+}. Interestingly, we found that RNase HIII appears to have a lower fidelity on the single rNMP than RNase HII as we observe faint smaller fragments indicating incision adjacent to the rNMP within the DNA. We suggest that Mn^{2+} may change substrate binding in a way that allows for the enzyme to be more promiscuous during substrate recognition or incision.

Because we found that the *B. subtilis* RNase HIII enzyme had activity on a single embedded rNMP in duplex DNA, we asked if this activity was more broadly conserved among other RNase HIII enzymes lacking the G (R/K) G motif. To this end we cloned and purified *Geobacillus stearothermophilus* and *Staphylococcus aureus* RNase HIII after overexpression in *E. coli*. We show how closely related *G. stearothermophilus* and *S. aureus* RNase HIII enzymes are to those from *B. subtilis* and *C. pneumoniae* in a simplified phylogenetic tree where we aligned the N-terminal region (Figure 3.4C). We used the same purification strategy as with the *B. subtilis* proteins, isolating RNase HIII proteins without a tag after purification (“Materials and Methods”). *G. stearothermophilus* and *S. aureus* RNase HIII were purified and resolved by SDS-PAGE with each protein migrating to their predicted sizes (Figure 3.5). The *G. stearothermophilus* and *S. aureus* RNase HIII enzymes were then tested for RNase H activity with Mg^{2+} and Mn^{2+} as cofactors on DNA substrates with a single or four embedded rNMPs alongside *B. subtilis* RNase HIII as a control. Both *G. stearothermophilus* and *S. aureus* RNase HIII proteins showed activity with both metals on a substrate

containing four rNMPs (Figure 3.4D top panel). As we observed with *B. subtilis* RNase HIII, *G. stearothermophilus* and *S. aureus* RNase HIII were able to incise a substrate containing a single rNMP hybridized to DNA with Mn^{2+} as a cofactor, but not with Mg^{2+} (Figure 3.4C bottom). RNase HIII enzymes have an N-terminal domain and a C-terminal catalytic domain containing four catalytic residues (DEDE). As mentioned above, the ability to recognize and cleave at a single rNMP hybridized to DNA has been attributed to a G (R/K) G motif found in RNase HII enzymes a motif absent from most RNase HIII enzymes (Lu et al., 2012b). None of the three RNase HIII enzymes tested here contain the G (R/K) G motif (Figure 3.4B). With these results, we suggest that RNase HIII enzymes with Mn^{2+} have activity on a single rNMP substrate and that this activity is conserved among organisms that code for RNase HIII. Further, we conclude that Mn^{2+}-stimulated cleavage is certainly not restricted to G (R/K) G containing RNase HIII enzymes.

3.4.3 RNase HII is effective on embedded rNMPs at *in vivo* concentrations of Mg^{2+} and Mn^{2+}

In vitro experiments thus far have been unable to differentiate between the possible *in vivo* functions of both RNase HII and HIII. Further, prior work assaying *B. subtilis* RNase HII and HIII used divalent metal concentrations well above the physiological range (Haruki et al., 2002). To determine which substrates RNase HII might act upon *in vivo* we conducted experiments titrating Mg^{2+}, Co^{2+}, Zn^{2+} or Mn^{2+}, with all three substrates to determine differences in

activity. We then compared these activities with those observed at the reported physiological range for both free metals from *B. subtilis*, (1 mM Mg^{2+} and 10 µM Mn^{2+}) [for review (Helmann, 2014)]. For Zn^{2+} and Co^{2+}, we are not sure of the physiological range therefore we titrated these metals over the same concentrations as done for Mn^{2+} and Mg^{2+}.

RNase HII did not show appreciable activity on a 20 mer RNA-DNA hybrid at physiologically relevant concentrations of Mg^{2+} or Mn^{2+} (Figure 3.6A). Therefore, at predicted *in vivo* concentrations of both metals RNase HII was unable to cleave the RNA-DNA hybrid lacking an RNA-DNA junction. At an RNase HII concentration 500-fold higher than previously used, RNase HII had activity cleaving 69.8% ± 0.02% of the substrate at predicted *in vivo* concentration of Mg^{2+} while showing no activity at an *in vivo* relevant Mn^{2+} concentration (Figure 3.7, Table 1). A mixture of Mg^{2+} and Mn^{2+} yielded an activity closely matching the activity observed with 1 mM Mg^{2+} of 71.4% ± 2.0% suggesting that RNase HII shows very weak activity on a 20 mer RNA-DNA hybrid with Mg^{2+} (Figure 3.7, Table 1).

When testing a substrate with four consecutive rNMPs embedded in DNA RNase HII showed the highest activity at 10 µM Mn^{2+} and 1 mM Mg^{2+}, 79.2% ± 2.0% and 87.9% ± 0.3% respectively. At metal concentrations in the physiological range using a mixture, we observed 86.8% ± 1.4% of the substrate was cleaved (Figure 3.6B). With these results we suggest that RNase HII can effectively cleave four embedded rNMPs with Mn^{2+} or Mg^{2+} while showing a slight preference for Mg^{2+} (Table 1). Interestingly, RNase HII activity with Mn^{2+} sharply

decreased at concentrations greater than 10 µM suggesting an inhibitory effect (Figure 3.6B).

RNase HII cleaved a single embedded rNMP with either Mg^{2+} or Mn^{2+}. RNase HII cleaved the substrate with greatest efficiency at Mn^{2+} concentrations greater than or equal to 10 µM (82.5% ± 1.01%) and at 10 mM Mg^{2+} concentrations well above the physiological range (78.1% ± 0.9%) (Figure 3.6C). Interestingly with metals predicted in the physiological range, RNase HII cleaved 83.3% ± 2.8% of the single embedded rNMP substrate suggesting Mn^{2+} or Co^{2+} (see below) might be preferred for this substrate (Table 1 and Figure 3.8).

In addition, we quantified the RNase HII activity with Co^{2+} and Zn^{2+} using the same metal titration assay as with Mg^{2+} and Mn^{2+}. Using a 20 mer RNA-DNA hybrid neither metal supported RNase HII cleavage. Co^{2+} supported RNase HII activity on a substrate with four consecutive rNMPs with a maximal cleavage of 78.0% ± 0.6% at 1 mM $CoCl_2$. Cleavage of a single embedded rNMP with Co^{2+} closely resembled Mn^{2+}activity with 89.4% ± 0.9% of the substrate cleaved at 1 mM $CoCl_2$. Cleavage on this substrate was also observed with Zn^{2+} but with much lower activity than with Mn^{2+} or Co^{2+} (Figure 3.8).

3.4.4 RNase HIII is most effective on RNA/DNA hybrids with four or more consecutive rNMPs using metal concentrations in the physiological range

To determine which substrates RNase HIII might act upon *in vivo* we repeated the same titration assays. Using a 20 mer RNA-DNA hybrid RNase HIII

was most active with Mg^{2+} concentrations greater than 10 µM cleaving 79.3% ± 0.3% at 1 mM. Activity with Mn^{2+} was also observed at concentrations greater than 10 µM but at about half that of Mg^{2+} (Figure 3.9A). Using concentrations of metals in the physiological range RNase HIII cleaved 77.2% ± 1.9% suggesting Mg^{2+} is preferred for incision of a substrate without an RNA-DNA junction. Further, RNase HIII was much more active on the 20 mer RNA-DNA hybrid than RNase HII suggesting that this may also represent an *in vivo* relevant substrate for RNase HIII (Table 1). In support of our *in vitro* findings, it was recently shown that cells with an *rnhC* deletion accumulate R-loops suggesting RNase HIII degrades RNA-DNA hybrids that lack an RNA-DNA junction *in vivo* (Lang et al., 2017).

Using a substrate with four embedded rNMPs RNase HIII showed activity at all concentrations of Mn^{2+} with a maximal activity of 89.3% ± 0.8% at 10 mM, well above the physiological range. RNase HIII was active with Mg^{2+} concentrations greater than 10 µM with a maximal activity of 87.6% ± 1.5% at 10 mM (Figure 3.9B). Using a mixture of Mn^{2+} and Mg^{2+} in the physiological range, RNase HIII cleaved 81.4% ± 2.7% of the substrate. In the presence of 1 mM Mg^{2+} 84.0% ± 0.9% of the substrate was cleaved. With these results we suggest that RNase HIII prefers Mg^{2+} or Co^{2+} (see below) for a substrate with four consecutive rNMPs (Table 1).

RNase HIII showed poor activity on a single embedded rNMP with any concentration of Mg^{2+} tested. Some activity was observed with Mn^{2+} starting at 10 µM and peaking at 1 mM with 50.9% ± 1.9% of the substrate cleaved (Figure

3.9C). At *in vivo* Mn^{2+} concentrations only 12.7% ± 2.2% of the substrate was cleaved suggesting that RNase HIII bound to Mn^{2+} is capable of weak activity on a single embedded rNMP.

Again, we assayed the activity of RNase HIII with Co^{2+} and Zn^{2+}. On a 20 mer RNA-DNA hybrid Zn^{2+} only permitted low levels of cleavage (Figure 3.10). Co^{2+} allowed for maximal activity at 10 mM concentration with 46.7% ± 0.6% of the substrate cleaved. On a substrate with four embedded rNMPs, RNase HIII showed similar activity with Co^{2+} as with Mn^{2+} and was generally ineffective with Zn^{2+}. Using a substrate with a single embedded rNMP both metals showed poor stimulation with activity peaking at 18.2% ± 9.2% with 1 mM Co^{2+} (Figure 3.10). With these results we conclude that RNase HIII does not make a substantial contribution to incision of single rNMPs embedded in DNA (Table 1). In support of this, cells with an RNase HIII deletion (*ΔrnhC*) do not appear to accumulate single rNMPs *in vivo* (Yao et al., 2013). Further, we conclude that RNase HIII is most active on substrates that lack an RNA-DNA junction.

3.4.5 DNA polymerase I can extend from an RNase HIII incised substrate in duplex DNA

Not much is known about the biochemistry of RNase HIII-mediated repair. To determine if Pol I was capable of extension from an RNase HIII incised substrate we purified DNA polymerase I (Pol I) (Figure 3.11) and assayed for extension following incision with RNase HII, RNase HIII or after alkaline hydrolysis simulating spontaneous cleavage *in vivo*.

For this assay we used a substrate with a single embedded rNMP because it provides a method to test for Pol I activity and rNMP removal by assaying for alkali insensitivity of the reaction product. Incorporated rNMPs are normally nicked on the 5' side by RNase HII leaving a clean 5′-phosphate and 3′-OH (Eder et al., 1993). If an incorporated rNMP is not removed the ribonucleotide can be subjected to spontaneous hydrolysis resulting in a 2′, 3′ cyclic phosphate, which requires further processing before removal can be achieved (Li and Breaker, 1999). We incubated the substrate with RNase HII, RNase HIII or treated it with alkaline conditions to mimic spontaneous hydrolysis *in vivo.* After cleavage with RNase HII, RNase HIII or NaOH the cleaved products were then purified and incubated with Pol I over a time course of 5 to 75 minutes. If Pol I extends the cleaved substrate then we would expect to observe the appearance of a full-length product that is refractory to alkaline treatment indicating that the rNMP was removed and replaced with a dNMP.

We found that substrates with a single rNMP cleaved by alkaline hydrolysis (NaOH) were refractory to Pol I extension (Figure 3.12) as expected due to the 2′, 3′-cyclic phosphate group. As a control, we show that Pol I extends a primed substrate treated with NaOH (Figure 3.13). The same substrate incised by RNase HII or RNase HIII was efficiently extended indicating that both cleaved intermediates served as suitable substrates for Pol I (Figure 3.6B). To ensure that the ribonucleotide was indeed removed by Pol I, we isolated the full-length product and tested for rNMP removal by subjecting the product to alkaline hydrolysis. We found that the RNase HII and RNase HIII incised substrates

yielded a product recalcitrant to alkaline hydrolysis following extension with Pol I (Figure 3.13). These results demonstrate removal of the rNMP and correction with a dNMP. With these results we conclude that in a defined system Pol I can extend and remove a single embedded rNMP in DNA following incision with RNase HII or RNase HIII.

3.4.6 Cells with an *rnhC* deletion are strongly sensitized to hydroxyurea

Hydroxyurea (HU) inhibits ribonucleotide reductase (RNR) and can cause an imbalance in the rNTP/dNTP pools (Navarra and Preziosi, 1999). HU may also cause DNA damage through other mechanisms (Kuong and Kuzminov, 2009). We constructed deletions of RNase HII (*rnhB*), RNase HIII (*rnhC*), *ypeP*, *ypdQ* and tested for growth inhibition on HU. We found that Δ*rnhC* cells were exquisitely sensitive to HU as determined by a cell spot titer assay (Figure 3.14A). We observed three to four logs of sensitivity for Δ*rnhC* cells relative to wild type. To control for growth inhibition resulting from Δ*rnhC* we expressed RNase HIII from an ectopic chromosomal locus complementing the Δ*rnhC* sensitivity (Figure 3.14A). None of the other predicted RNase H gene deletions conferred sensitivity to HU. Double mutants of Δ*rnhB,* Δ*ypeP* and Δ*ypdQ* in combination with Δ*rnhC* showed no further increase in growth inhibition relative to Δ*rnhC* alone.

To test whether RNase HII or YpeP, were capable of substituting for RNase HIII *in vivo, rnhB* (HII) or *ypeP* were overexpressed in the absence of RNase HIII (Δ*rnhC*). Neither *rnhB* (HII) nor *ypeP* were able to restore growth to

Δ*rnhC* cells (Figure 3.14B). We did find that overexpression of *rnhB* provided some feeble rescue of Δ*rnhC,* but nowhere near the rescue observed with ectopic expression of *rnhC* (please compare last row Figure 3.14A with third row Figure 3.14B). As a control, we show that RNase HII (RnhB) was indeed overexpressed *in vivo* (Figure 3.14C). With these results we conclude that *rnhB* cannot substitute for *rnhC* during survival to HU in *B. subtilis*. Further, with this data we suggest that RNase HII and HIII have distinct substrate specificities *in vivo*.

3.5 Discussion

Here we show that *B. subtilis* RNase HII and HIII are functional on several RNA-DNA hybrids while putative RNase H enzymes, YpeP and YpdQ are not. Further, the activity of RNase HII and HIII overlaps on the three substrates examined although enzyme activity is dependent on the concentration and the divalent metal cation used.

Although the activity and substrate specificity of *B. subtilis* RNase HII and RNase HIII has been partially examined (Haruki et al., 2002; Ohtani et al., 1999b), recent work has shown that *C*. *pneumonia* RNase HIII can incise a substrate with a single rNMP when bound to Mn^{2+} (Lu et al., 2012a). This study motivated us to test the *B. subtilis* protein, which was previously shown to lack activity on a substrate with a single rNMP (Haruki et al., 2002). We found that RNase HIII from three Gram-positive bacteria was indeed able to incise a substrate at a single rNMP when bound to Mn^{2+} even though all three of the

RNase HIII enzymes we tested lack the G(R/K)G motif, a sequence implicated in single rNMP recognition by RNase HII enzymes and *C. pneumonia* RNase HIII (Lu et al., 2012b). We speculate that most if not all bacterial RNase HIII enzymes are biochemically capable of incision at a single embedded rNMP with Mn^{2+} or Co^{2+}.

We also investigated the ability of Pol I to concomitantly remove and extend a substrate with a single rNMP following alkaline hydrolysis or incision with RNase HIII. We show that a 2′, 3′-cyclic phosphate resulting from alkaline hydrolysis cannot be extended by Pol I, as expected, suggesting that it must be removed or healed to provide a suitable substrate for Pol I. In *E. coli,* RtcB is a cyclic phosphodiesterase that heals the 2′, 3′-cyclic phosphate end allowing for ligation followed by RER (Tanaka and Shuman, 2011). We have been unable to identify an RtcB homolog or ortholog in *B. subtilis* that we could test for end healing activity. We did show that RNase HII and RNase HIII cleavage products are efficiently used by Pol I for removal of the rNMP and resynthesis with dNTPs. The work we present here suggests that RNase HIII is biochemically capable of performing a similar role to RNase HII however when Mn^{2+} is in the physiological range, RNase HIII yields weak incision suggesting that RNase HIII provides very little if any contribution to the correction of single rNMP errors *in vivo*. In support of our findings, *B. subtilis* cells with Δ*rnhB* show an increase in mutation rate (Yao et al., 2013) and genomic DNA isolated from Δ*rnhB* cells have increased alkaline sensitivity while Δ*rnhC* do not. In addition, Δ*rnhB,* Δ*rnhC* cells show no further sensitivity relative to Δ*rnhB* (Yao et al., 2013). Therefore, based on

several lines of evidence we find it very unlikely that RNase HIII contributes to single rNMP removal *in vivo*. This work does show that Pol I can operate from an RNase HIII incised nick, which could implicate the cooperation of RNase HIII and Pol I in removing longer stretches of RNA embedded in DNA including Okazaki fragments.

Additionally, at physiological metal concentrations it becomes clear that RNase HII cannot efficiently cleave RNA-DNA hybrids lacking an RNA-DNA junction. This could represent substrates such as R-loops *in vivo*. Whereas, we show using metals in the physiological range that RNase HIII efficiently cleaves substrates lacking an RNA-DNA junction further suggesting that R-loops or similar structures represent the primary *in vivo* substrate of RNase HIII. In our experiments RNase HIII shows the most activity with physiological concentrations of Mg^{2+} on a substrate lacking an RNA-DNA junction. In further support of our results we show that RNase HIII deficient cells are sensitive to HU while RNase HII cells are not. This sensitivity cannot be fully complemented with the overexpression of RNase HII, although a very modest rescue is observed (Figure 3.14). This work suggest RNase HIII has a specific role *in vivo,* where metal availability is limited. Recently RNase HIII was shown to remove R-loops from head-on genes (Lang et al., 2017). This serves as a possible explanation for our observation that RNase HIII deficient cells are sensitive to HU. Inhibition of RNR with HU could deplete dNTP pools slowing the rate of replication. Slowed or impaired replication due to HU challenge would make it increasingly difficult for the replisome to bypass R-loops that would accumulate in cells bearing an *rnhC*

deletion. At metal concentrations in the physiological range, RNase HII is unable to use Mn^{2+} (Fig. 3.6A) and therefore would be unlikely to resolve R-loops. At increased concentrations of RNase HII weak activity is observed with Mg^{2+} (Figure 3.7), which could account for the modest rescue of the phenotype when RNase HII is overexpressed (Figure 3.14B). *B. subtilis* manganese levels can be altered *in vivo* through deletion of *mntR*, a manganese transport regulator; however at concentrations necessary for RNase HII activity on an R-loop like substrate (1 mM), Mn^{2+} affects growth making this hypothesis difficult to test further *in vivo* (Que and Helmann, 2000).

We did find that RNase HII and HIII efficiently cleave substrates with four rNMPs nested in DNA. Both enzymes showed almost identical activity with Mg^{2+} in the physiological range or when incubated with Co^{2+}. Since we could not distinguish between RNase HII and HIII on the substrate with four rNMPs we suggest that both enzymes are likely to have activity on this substrate *in vivo*. A stretch of rNMPs in DNA could be generated during “ribonucleotide patch” repair when cells are in non-growing states or when dNTP pools are low (Ordonez et al., 2014). Further, although this substrate is not an Okazaki fragment, it does provide a stretch of rNMPs with a DNA junction like an Okazaki fragment. Therefore, we suggest that RNase HII and HIII may act during Okazaki fragment maturation *in vivo*. With these data we conclude that RNase HII and HIII are specialized for activity on single embedded ribonucleotides or substrates lacking an RNA-DNA junction based on the available concentration of metals regulating

their activity. We also show that RNase HII and HIII have overlapping activity on stretches of four or more ribonucleotides.

3.6 Materials and Methods

3.6.1 Residue alignments

FASTA sequences of each protein were obtained from the NCBI database and entered into the Clustal Omega sequence alignment website and aligned for conserved residues. (http://www.ebi.ac.uk/Tools/msa/clustalo/)

3.6.2 CD spectroscopy

YpdQ and YpeP proteins were dialyzed into a 10 mM sodium phosphate pH 8, 10 mM NaCl buffer and diluted to a final concentration of 0.133 mg/ml and 0.2 mg/ml respectively. Far-UV CD spectra were recorded in the same buffer at 30°C using a Jasco-J1500 spectropolarimeter. Spectra were measured after buffer correction between 260-190 nm and molar ellipticity was calculated.

3.6.3 RNase HIII phylogenetic tree

The phylogenetic tree of RNase HIII proteins was generated from the NCBI database cluster cd14796 representing only RNase HIII proteins with the N-terminal domain extension. This list is not comprehensive.

3.6.4 RNase H assays

5′ IR dye labeled 1 rNMP, 4 rNMP, and all RNA substrates were annealed by placing oligos (oJR234, oJR235, and oJR227) with oJR145 in a 98°C water bath for 1 minute in a buffer containing 20 mM Tris-HCl pH 8, 50 mM NaCl. Samples were then allowed to cool slowly back to 25^{o}C on the bench top. Reactions were performed in the same buffer with 100 nM substrate and various protein concentrations and metals for 10 min at 25ºC in a total volume of 10 µl. For NaOH samples, 100 nM substrate was placed into 300 mM NaOH and incubated at 45°C for 5-30 min depending on substrate followed by neutralizing the pH with 2M Tris-HCl pH 7. Reactions were quenched by addition of 10 µl formamide loading dye (95% formamide 20 mM EDTA, 1% SDS, 0.01% bromophenol blue). Samples were boiled for 2 minutes at 100°C and snap cooled on ice. Reaction products were separated in denaturing 20% Urea-PAGE followed by visualization with a LI-COR Odyssey imager.

3.6.5 Metal-dependent incision reactions

Assay was performed essentially as described above except: 100 nM of each substrate was used with 4 nM or 50 nM RNase HII or RNase HIII. Metals were titrated from 1 µM to 10 mM followed by incubation of the reactions for 20 minutes at 30°C. Reactions were quenched and imaged as described above. The LI-COR Odyssey imager software was used to quantify total substrate in each lane and the substrate cleaved. Percent substrate cleaved, was reported by dividing substrate cleaved by total substrate and multiplying by 100. The

percentage cleaved with no metal was subtracted from each lane. Reactions were performed in duplicate with the mean reported. Error and error bars represent the range between duplicate experiments.

3.6.6 Pol I extension

Extension assays were performed by first annealing oJR234 and oJR145 as described above (see RNase H assays) to create a 10 µM 5′ IR labeled substrate with 1 embedded rNMP. Hydrolysis reactions were then incubated for 2 hours at 37°C. The same annealing buffer was used with 2 µM of the annealed substrate. The first hydrolysis reaction contained only 1 mM $MnCl_2$ and was labeled ‘None’. The second contained 300 mM NaOH and was labeled ‘NaOH’. The third and fourth both contained 1 mM $MnCl_2$ with the former having 1 µM RNase HII and the latter 1 µM RNase HIII. Each reaction was subsequently stopped using 100 µl of quenching buffer (3.6 M NH_4AC, 20 mg glycogen). A 25-fold excess of ethanol was then added, mixed and each was incubated at -80°C overnight. Precipitated DNA was pelleted at 13k RPM for 15 min, washed in 1 ml of 70% ethanol and pelleted again. Liquid was aspirated and the pellet was air dried before resuspending to a final volume of 20 µl using ddH_2O solubilizing the hydrolyzed substrates. Substrate concentrations were normalized after visualizing each by 20% Urea-PAGE and quantifying the concentration via the LI-COR IR Odyssey imager software. For the primer extension reaction (Fig. 3.12) Primer oJR247 was annealed to oJR251, NaOH treated and purified as described above.

Extension reactions with each substrate were performed with 100 nM each hydrolyzed substrate in extension buffer (40 mM Tris-acetate pH7.8, 12 mM magnesium acetate, 300 mM potassium glutamate, 3 µM $ZnSO_4$, 2% PEG, 1 mM DTT) with or without 1 µM Pol I at 25°C. Samples were removed and quenched with an equal volume of formamide loading dye (95% formamide, 1% SDS, 5 mM EDTA, 0.01% bromophenol blue) at 5, 25, and 75 minutes. Next, 300 mM NaOH was added to the remaining sample and incubated at 45°C for 60 minutes. 1/5 the volume of 2M Tris-HCl pH 7 was added to neutralize the pH and reactions were quenched with an equal volume of formamide loading dye. Samples were boiled for 2 minutes at 100°C and snap cooled in an ice bath. Reactions products were separated by 20% Urea-PAGE followed by imaging with a LI-COR Odyssey imager.

3.6.7 Affinity purification of polyclonal antiserum

SDS-PAGE was performed using 100 µg of purified RNase HII. The resulting gel was transferred to a PVDF membrane and Ponceau stained. The RNase HII band was cut from the membrane and blocked with 5% non-fat milk in PBS for 20 min. The band was washed twice with PBS and incubated with 300 µl of antisera for 1 hour. Antisera was removed and the band was washed twice more with PBS. 300 µl of 5 mM glycine, 150 mM NaCl pH2.4 is added and incubated for 30 seconds to strip the antibody. Supernatant was removed and

added to 100 µl of $NaPO_4$ to neutralize. This solution was used at 1:1000 as α-RnhB from lot MI1369. Custom anti-serum was obtained from Covance.

3.6.8 Western blotting

Strains were grown in 2 ml LB media for two hours at 37°C and 1 ml was pelleted by centrifugation. Cells were resuspended in lysis buffer (10 mM Tris HCl pH 7, 1x protease inhibitors, 1 mM AEBSF, 2 mg/ml lysozyme) to an OD of 1. Resuspended cells were incubated for 15 min at 37°C then SDS was added to 1% and the cells were placed on ice for 30 minutes. SDS-PAGE was performed using each lysate and then transferred onto a nitrocellulose membrane. The membrane was blocked for 30 min in TBST with 2% milk and incubated overnight at 4°C in α-RnhB diluted 1:1000 in TBST with 2% milk. Membrane was washed 3x in TBST then incubated with secondary antibody conjugated to IR800 dye diluted 1:15000 in TBST with 2% milk. The membrane was again washed 3x with TBST and imaged with the LICOR Odyssey IR imager.

3.6.9 Spot-titer assays

A single colony of the indicated strain was used to inoculate 3 ml of LB media and grown to an OD600 between 1 and 1.5. Cultures were then normalized to an OD600 = 1 in a 0.85% saline solution and serial diluted to 10^{-5}. A total volume of 5 µl was spotted for each dilution on LB or LB containing the indicated concentrations of hydroxyurea and or IPTG. Plates were then incubated overnight at 30ºC and imaged the following morning.

3.6.10 Protein Purification

All proteins were purified essentially as described (Schroeder et al., 2017). Briefly, each protein was purified by first transforming *E. coli* BL21 cells with pE-SUMO overexpression plasmids. Cultures of each strain containing the plasmid were grown in 3 L of TB to an OD of ~0.75 shaking at 37°C and then expression was induced by addition of 250 µM IPTG. Cultures were left for an additional 3-4 hours, then pelleted and frozen in liquid nitrogen. Pellets were thawed and resuspended in 50 ml of Lysis Buffer (50 mM Tris-HCl pH8, 300 mM NaCl, 10% sucrose, 10 mM imidazole, 1x protease inhibitors) and lysed via sonication. Cell debris was pelleted via centrifugation and the supernatant was applied to a 2 ml column of Ni^{2+}-NTA agarose bead column equilibrated in lysis buffer. The column was washed with wash buffer (50 mM Tris-HCl pH 8, 2M NaCl, 10 mM imidazole) and then eluted with elution buffer (50 mM Tris-HCl pH 8, 50 mM NaCl, and 500 mM imidazole). Following elution, the elute was dialyzed into SUMO Protease buffer (50 mM Tris-HCl pH 8, 150 mM NaCl, 1 mM DTT) with SUMO Protease and left overnight at 4°C to cut the tag. The dialyzed protein mixture was then added back to a 2 ml column of Ni^{2+}-NTA agarose bead column and purified protein was collected and analyzed via SDS-PAGE. Fractions were pooled, diluted and then each protein was further purified using a Q anion exchange column with a 50-750 mM NaCl gradient. Fractions were analyzed via SDS-PAGE with the purest fractions concentrated into protein storage buffer (50 mM

Tris-HCl pH 8, 150 mM NaCl, and 25% glycerol) aliquoted and frozen in liquid nitrogen.

3.6.11 Strain building

B. subtilis strains JRR25 and JRR26 were built by placing the corresponding gene (*rnhB, rnhC,*) with their upstream and downstream sequences into the pMiniMad plasmid and integrated at the native locus via double crossover. *B. subtilis* strains JRR59 and JRR60 were built by first ordering *B. subtilis* 168 knockout *erm* (BKE) strains from the Bacillus Genetic Stock Center (http://www.bgsc.org). Each strain has the corresponding gene replaced with a *loxP* flanked *erm* cassette. Genomic DNA was purified from these strains, and used to transform PY79. The *erm* cassette was removed with Cre recombinase. *B. subtilis* strains JRR30, JRR28, and WGH29 were built by placing the corresponding genes into the pDR110 plasmid and integrating them at the *amyE* locus of the JRR26 strain (JRR30, JRR62, and WGH28) via Campbell integration. *B. subtilis* strains JRR61 and JRR68 were created by transforming JRR26 with genomic DNA from the Bacillus Genetic Stock Center again removing the *erm* cassette with Cre recombinase.

3.7 Notes and Acknowledgements

A version of this chapter was published in *The Journal of Bacteriology* (Randall et al., 2017). This work was supported by NIH grant R01 GM107312, NIH Cellular Biotechnology Training Grant (T32 GM008353) and a Pre-doctoral

Fellowship from Rackham Graduate School at the University of Michigan. The work was also supported in part by NIH T32 GM007544. We wish to thank Bastian Groitl and Dr. Ursula Jakob for their help and use of their spectrophotometer regarding circular dichroism spectroscopy.

Lyle Simmons and I designed all the research. I performed all of the experiments except the Western blot analysis which was performed by William Hirst. Lyle Simmons and I wrote and edited the paper.

3.8 Figures and Tables

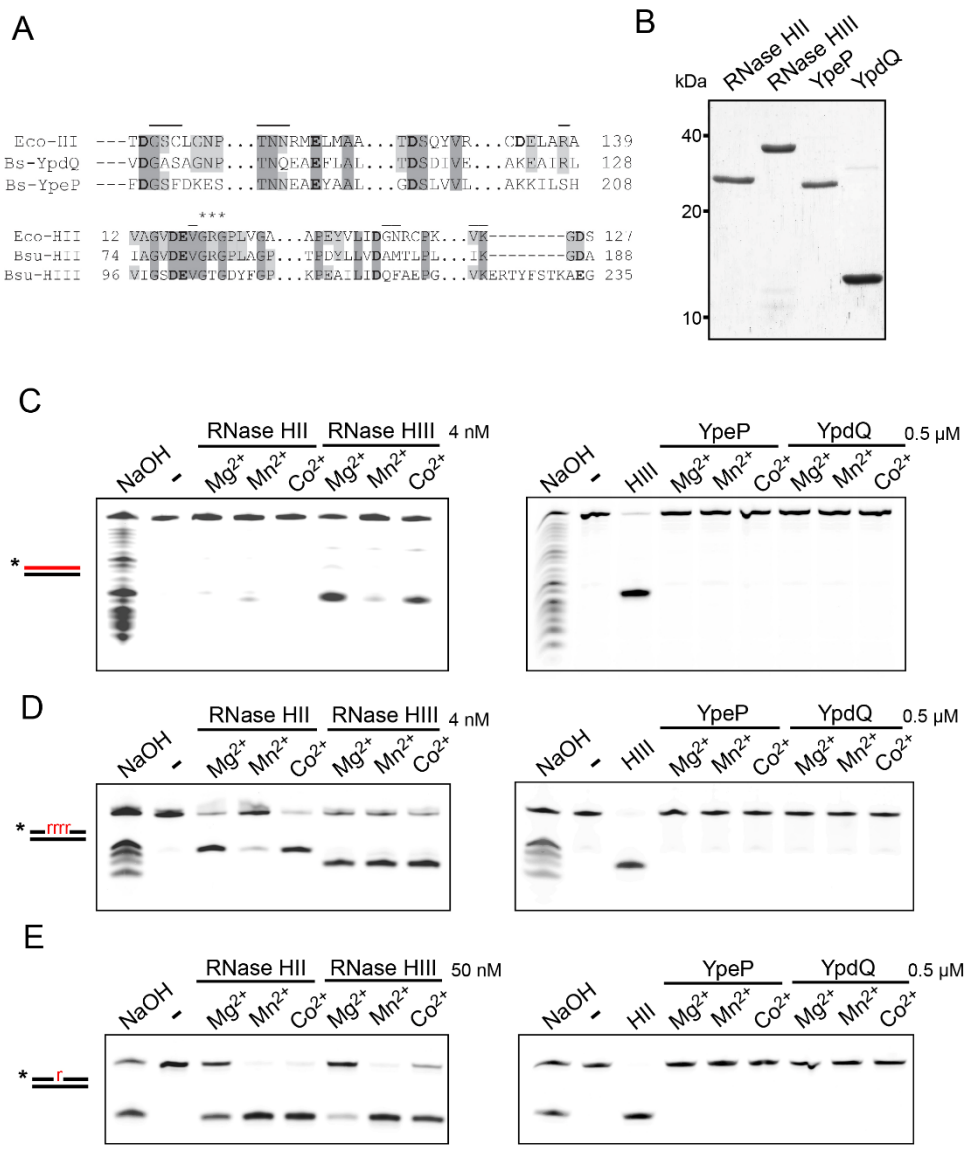

Figure 3.1 *B. subtilis* has two active RNase H enzymes. (A) Alignment of two putative RNase H enzymes YpdQ and YpeP with *E. coli* RNase HI. *Bacillus subtilis* RNase HII and HIII aligned with *E. coli* HII. Conserved residues are in gray catalytic residues are bold, *** denotes GRG single rNMP binding motif, and residues involved in substrate binding have a line above. The alignment was performed using Clustal Omega. (B) SDS-PAGE with 2 μg of the indicated *B. subtilis* RNase H enzymes stained with coomassie blue. (C) Left 4 nM RNase HII and HIII; (C, right) 0.5 μM YpeP and YpdQ cleavage of a 5' end labeled 20 bp RNA/DNA hybrid with Mg^{2+}, Mn^{2+}, or Co^{2+} resolved on 20% Urea-PAGE. (D) Left 4 nM RNase HII and HIII; (D, right) 0.5 μM YpeP and YpdQ cleavage of a 5' end labeled four embedded rNMP hybrid substrate with Mg^{2+}, Mn^{2+}, or Co^{2+} resolved on 20% Urea-PAGE. (E) Left 50 nM RNase HII and HIII; (E, right) 0.5 μM YpeP and YpdQ incubated with a 5' end labeled single embedded rNMP RNA-DNA hybrid with Mg^{2+}, Mn^{2+}, or Co^{2+} resolved on 20% Urea-PAGE. For (C-E) metal concentrations were 2 mM. Red lines and r's represent ribonucleotides. ***** represents IR dye label

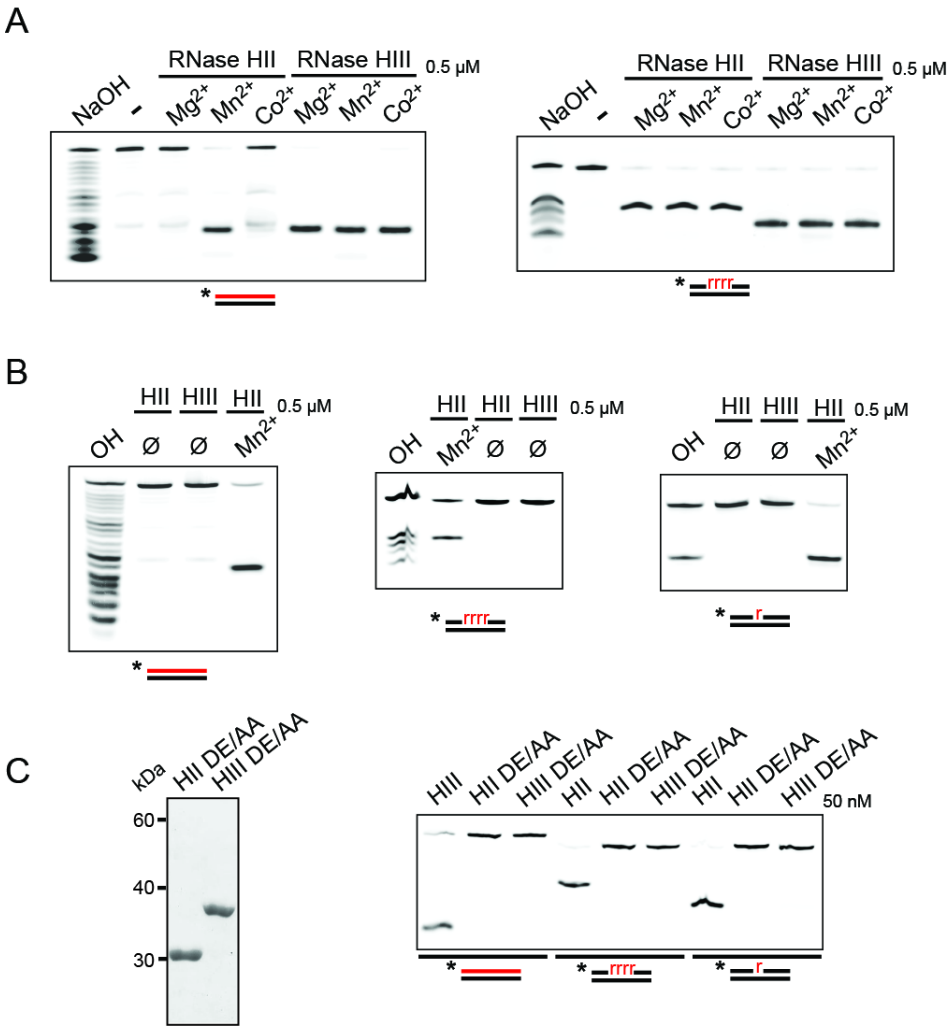

Figure 3.2 RNase HII and RNase HIII require metal for functionality. (A) 0.5 µM RNase HII and HIII cleavage of a 5' end labelled 20 bp RNA/DNA hybrid (left) and four embedded rNMP hybrid (right) with Mg^{2+}, Mn^{2+}, or Co^{2+} resolved on 20% Urea-PAGE. When used metal concentration was 2 mM. (B) RNase HII and HIII cleavage of a 20 bp RNA/DNA hybrid (left), four embedded rNMP hybrid (center) or one embedded rNMP hybrid (right) with no metal present. (C) Left catalytically inactive DE/AA variants including RNase HII D78A, E79A and RNase HIII D100A, E101A were resolved by resolution in a 12% SDS-PAGE and stained with coomassie blue. (C, right) Catalytically inactive RNase HII and RNase HIII as well as wild type RNase HII cleavage of a 5' end labelled 20 bp RNA/DNA hybrid, four embedded rNMP hybrid and one embedded rNMP hybrid with Mn^{2+} resolved on 20% Urea-PAGE. Red lines and r's represent ribonucleotides. ***** represents IR dye label.

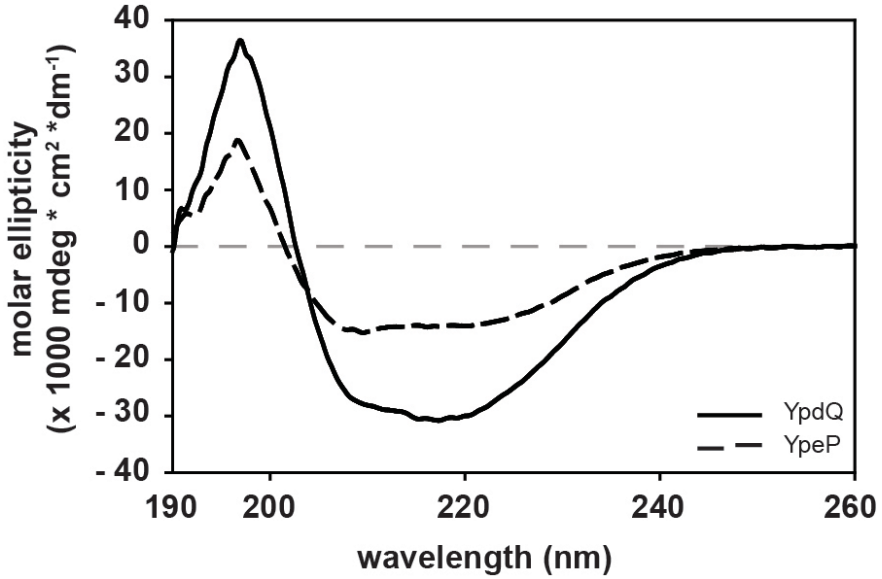

Figure 3.3. Purified YpdQ and YpeP are structured. Shown is the circular dichroism (CD) spectrum for YpdQ (solid) and YpeP (dashed).

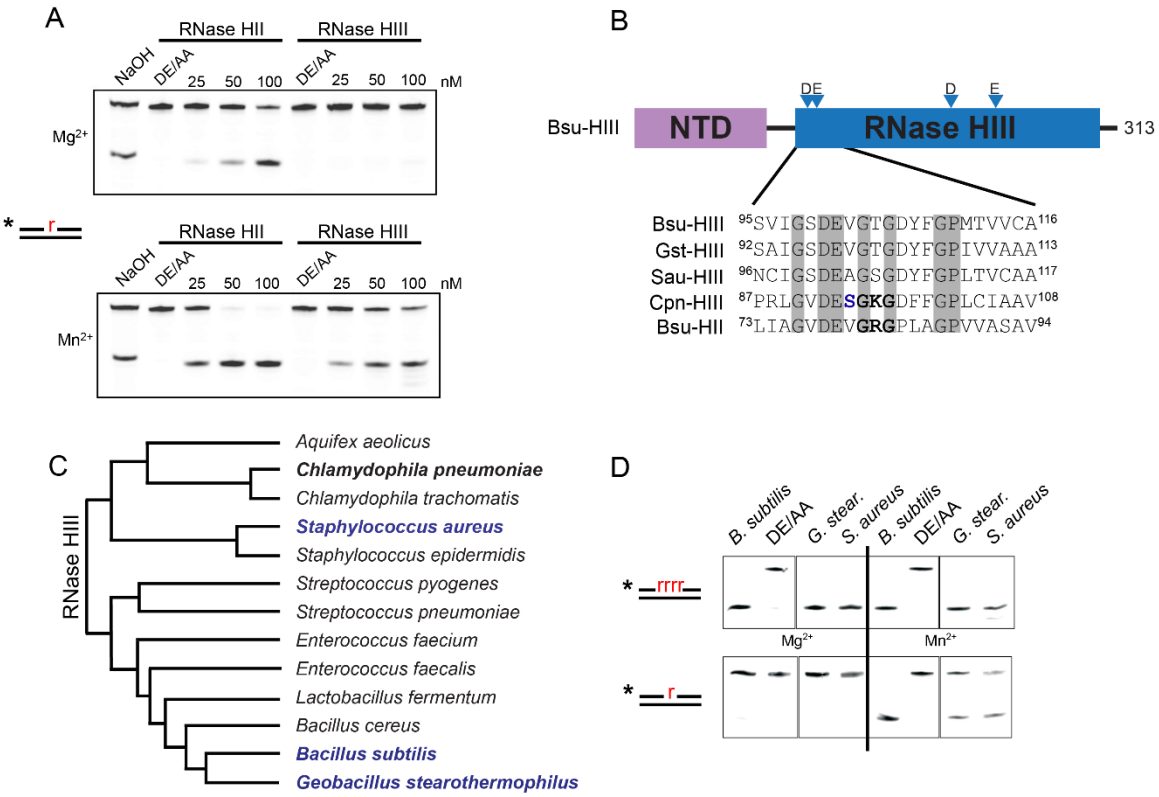

Figure 3.4 Cleavage of a single embedded rNMP with Mn^{2+} is conserved among different RNase HIII enzymes. (A) A 5' end labeled single embedded rNMP substrate cleaved with 25, 50, and 100 nM RNase HII and RNase HIII with Mg^{2+} (top) or Mn^{2+} 633 (bottom). DE/AA corresponds to 1 μM RNase HII D78A, E79A and HIII D100A, E101A included as controls. Metal concentrations were 1 mM. (B) Predicted functional domains of *B. subtilis* RNase HIII. Magnified is an alignment of the region containing the G (R/K)G motif from *Chlamydophila pneumoniae* (bold). *C. pneumoniae* S94 is in blue. (C) Phylogenetic tree of some RNase HIII containing bacterium based on N-terminal domain (NTD) homology. Blue are RNases HIII assayed in D, bolded correspond to RNase HIII published previously (25, 27). (D) Cleavage of a 5' end labeled four embedded rNMP (top) or single embedded rNMP (bottom) hybrid substrate with RNase HIII from *B. subtilis* (wild type), D100A, E101A, *Geobacillus stearothermophilus* and *Staphylococcus aureus* with Mg^{2+} (left half) or Mn^{2+} (right half) using 1 mM metal concentrations. Red lines and "r" represent ribonucleotides. * denotes IR dye label.

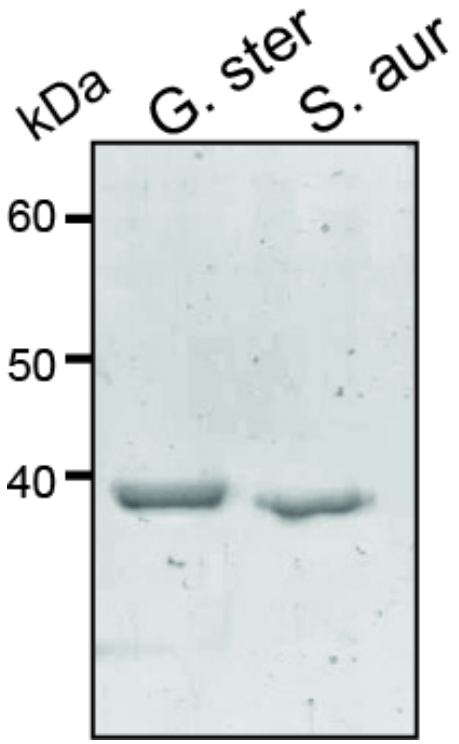

Figure 3.5 *G. stearothermophilus* and *S. aureus* RNase HIII. Shown is 2 µg of *Geobacillus stearothermophilus* and *Staphylococcus aureus* RNase HIII resolved in a 12% SDS-PAGE and stained with coomassie blue.

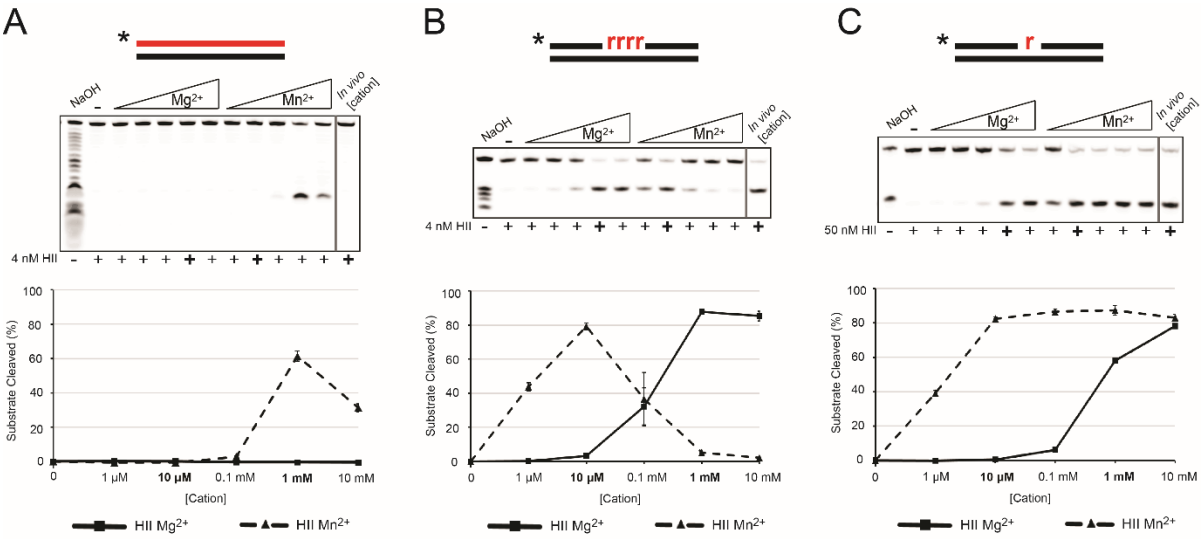

Figure 3.6 RNase HII is effective on substrates with an RNA-DNA junction using metal concentrations in the physiological range. (A) 4 nM RNase HII cleaving a 20 bp RNA-DNA hybrid substrate with varying Mg^{2+} and Mn^{2+} concentrations. (B) 4 nM RNase HII cleaving a substrate with 4 embedded rNMPs in DNA with varying concentrations of Mg^{2+} and Mn^{2+} RNase H activity is influenced by metal availability (C) 50 nM RNase HII cleaving a single embedded rNMP in DNA, representing an RER substrate, with varying Mg^{2+} and Mn^{2+} concentrations. For (A-C) Predicted *in vivo* concentrations (Helmann, 2014) of Mg^{2+} and Mn^{2+} are shown in bold. The percentage of substrate cleaved versus metal concentration is graphed below. Concentrations for Mg^{2+} and Mn^{2+} used were in 10 fold increments from 0.001 mM to 10 mM. The mean is reported with error bars representing the range of duplicate samples. Red lines and "r" represent ribonucleotides. * denotes IR dye label.

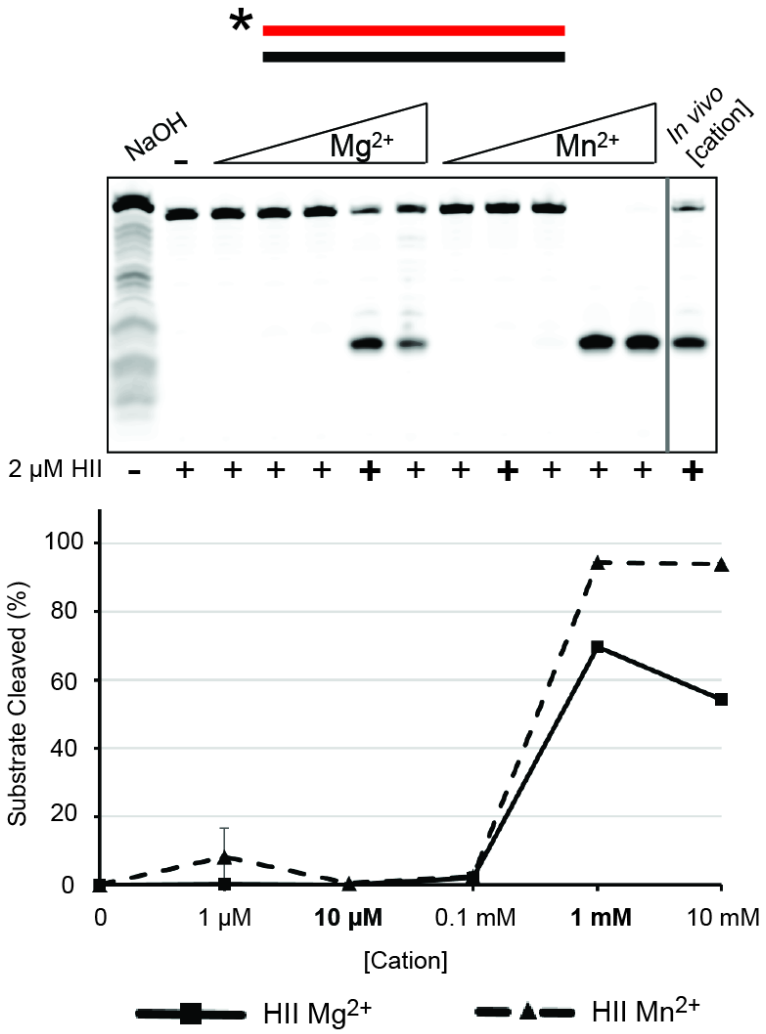

Figure 3.7 RNase HII incision on a 20 mer RNA-DNA hybrid with increased protein concentration. Shown is 2 μM RNase HII incubated with 20 bp RNA-DNA hybrid substrate with varying Mg^{2+} and Mn^{2+} concentrations to detect incision. Predicted *in vivo* concentrations of each metal are bolded. The percentage of substrate cleaved versus concentration of metal is graphed below. Error bars represent the range of duplicate experiments. Red bar represents RNA. * denotes IR dye label.

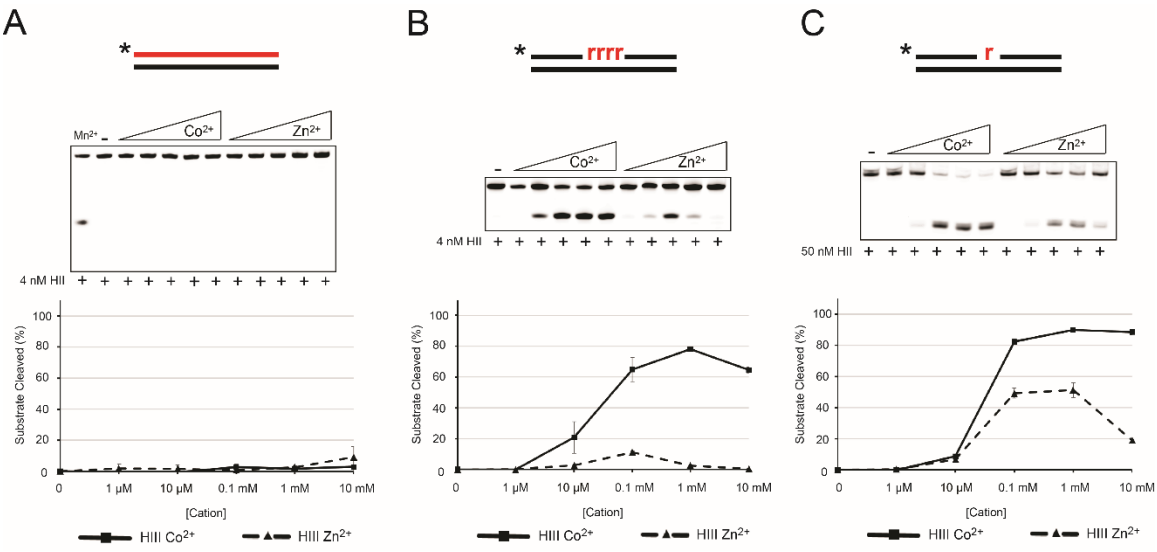

Figure 3.8 RNase HII activity with Co^{2+} and Zn^{2+}. RNase HII incubated with a 20 mer RNA-DNA hybrid (A) four embedded rNMPs (B) or (C) one embedded rNMP in DNA at various $CoCl_2$ and $ZnCl_2$ concentrations. Top panels are representative urea-PAGE. The percentage of substrate cleaved versus metal concentration is graphed below. Error bars represent the range of duplicate experiments. Red lines and "r" represent ribonucleotides. * denotes IR dye label.

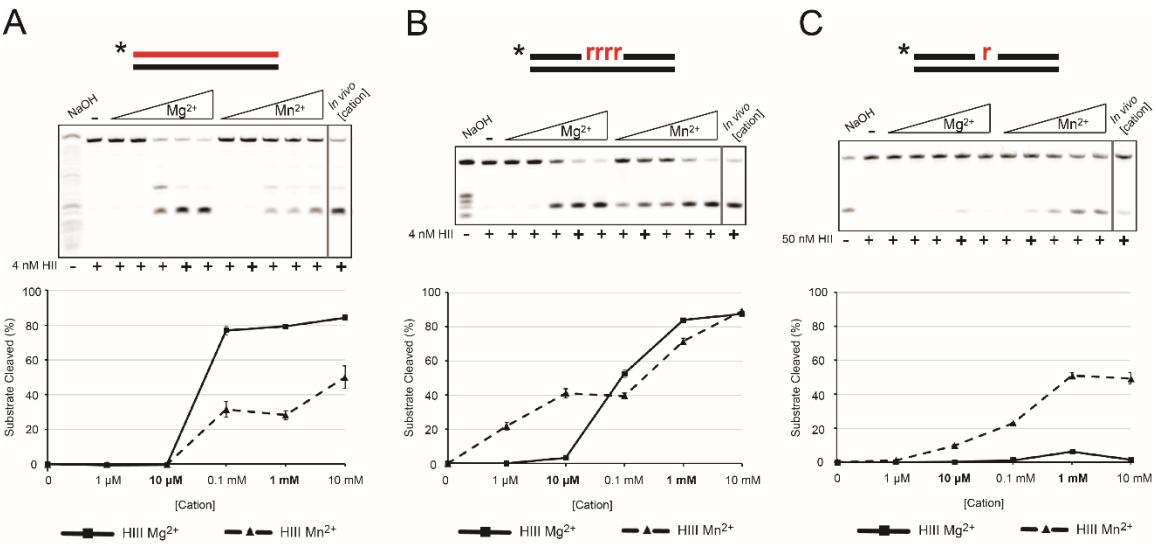

Figure 3.9 RNase HIII is effective on a substrate lacking an RNA-DNA junction using metal concentrations in the physiological range. (A) 4 nM RNase HIII cleaving a 20 bp RNA-DNA hybrid substrate with varying Mg^{2+} and Mn^{2+} concentrations. (B) 4 nM RNase HIII cleaving a substrate with 4 embedded rNMPs in DNA with varying concentrations of Mg^{2+} and Mn^{2+} (C) 50 nM RNase HIII cleaving a single embedded rNMP in DNA, representing an RER substrate, with varying Mg^{2+} and Mn^{2+} concentrations. For (A-C) Predicted *in vivo* concentrations (Helmann, 2014) of Mg^{2+} and Mn^{2+} are in bold. The percentage of substrate cleaved versus metal concentration is graphed below. Concentrations for Mg^{2+} and Mn^{2+} used were in 10-fold increments from 0.001 mM to 10 mM. The mean is reported with error bars representing the range of duplicate samples. Red lines and "r" represent ribonucleotides. * denotes IR dye label.

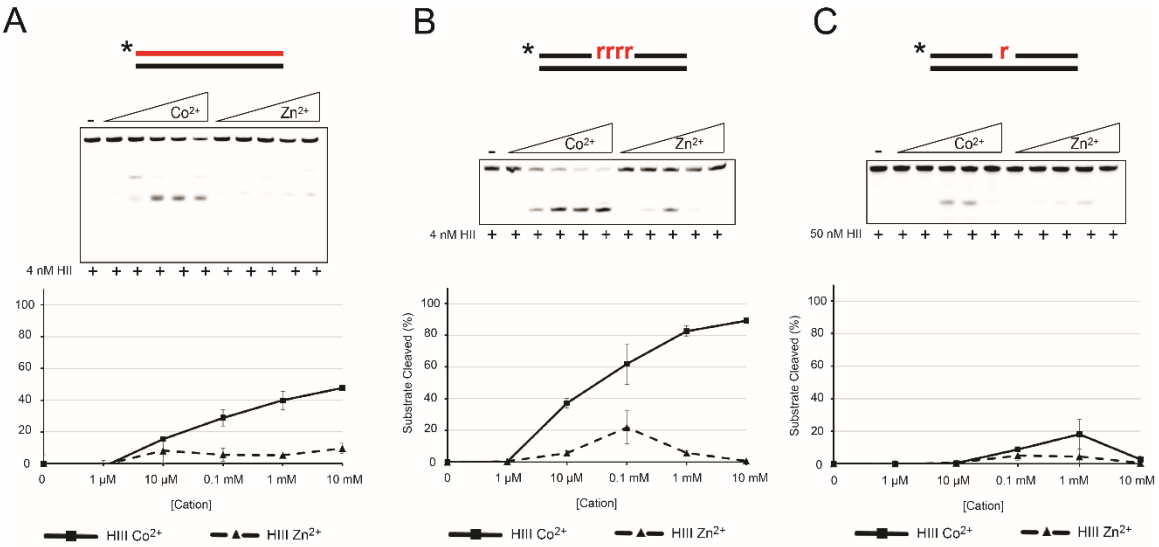

Figure 3.10 RNase HIII activity with Co^{2+} and Zn^{2+}. RNase HIII incubated with a 20mer RNA-DNA hybrid (A) four embedded rNMPs (B) or (C) one embedded rNMP in DNA at various $CoCl_2$ and $ZnCl_2$ concentrations. Top panels are representative urea-PAGE gels. The percentage of substrate cleaved versus metal concentration is graphed below. Error bars represent the range of duplicate experiments. Red lines and "r" represent ribonucleotides. * denotes IR dye label.

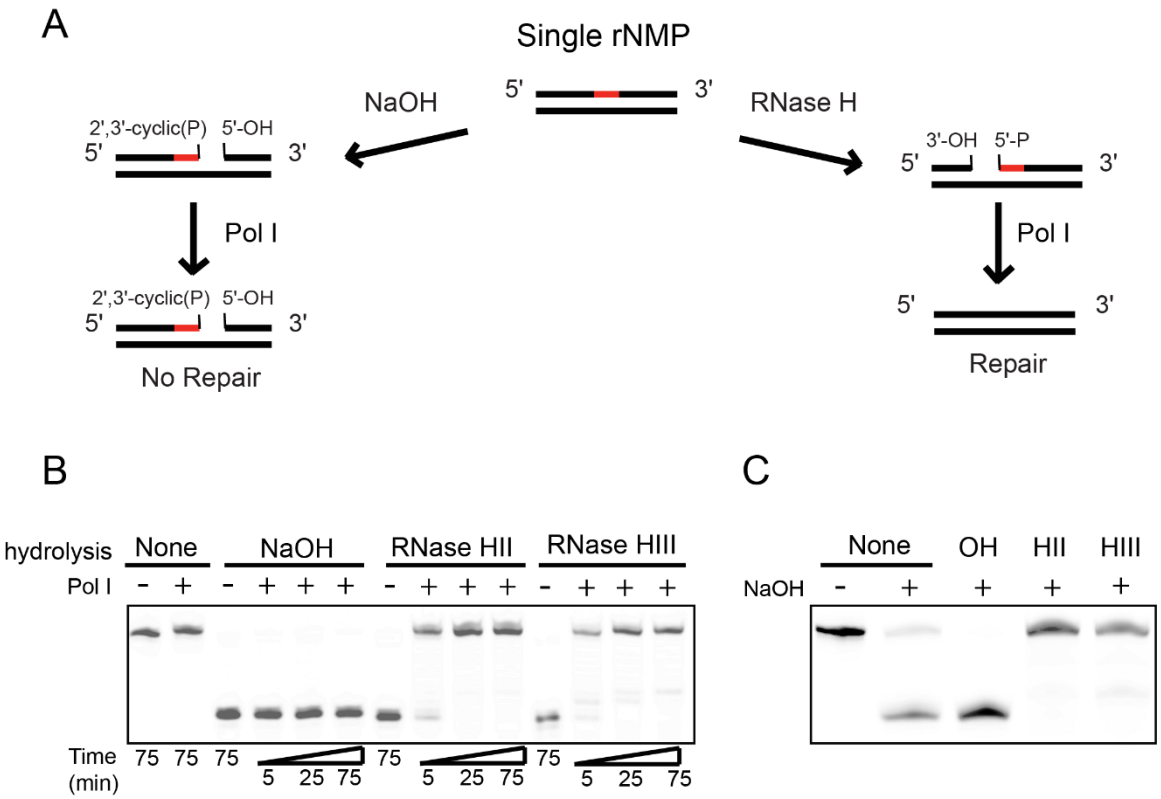

Figure 3.11 DNA polymerase I can extend from an RNase HIII incised substrate. (A) Schematic of the experimental design with a single embedded rNMP in duplex DNA cleaved by RNase H or hydrolyzed by NaOH followed by extension with Pol I. (B) A single embedded rNMP substrate with no cleavage (None), NaOH hydrolysis (NaOH), RNase HII cleavage (RNase HII) or RNase HIII cleavage (RNase HIII) followed by incubation with Pol I for 5, 25 and 75 minutes. For efficient incision RNase HII and HIII were incubated with 2 mM $MnCl_2$. (C) Pol I incubated products from (B) were treated with NaOH for 60 minutes to determine if the product following Pol I extension was sensitive or refractory to alkaline treatment to test for removal and replacement of the rNMP with a dNMP. Red bars represent a single rNMP.

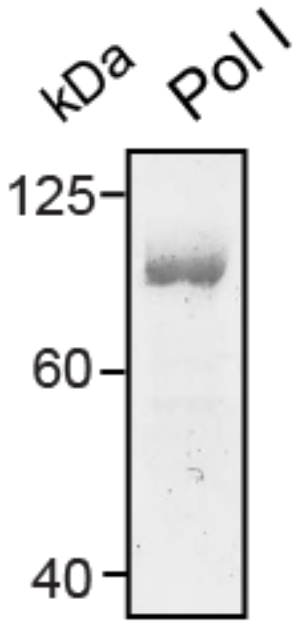

Figure 3.12 Purified DNA polymerase I. Shown is 2 µg of Pol I resolved on an 8% SDS-PAGE and stained with coomassie blue.

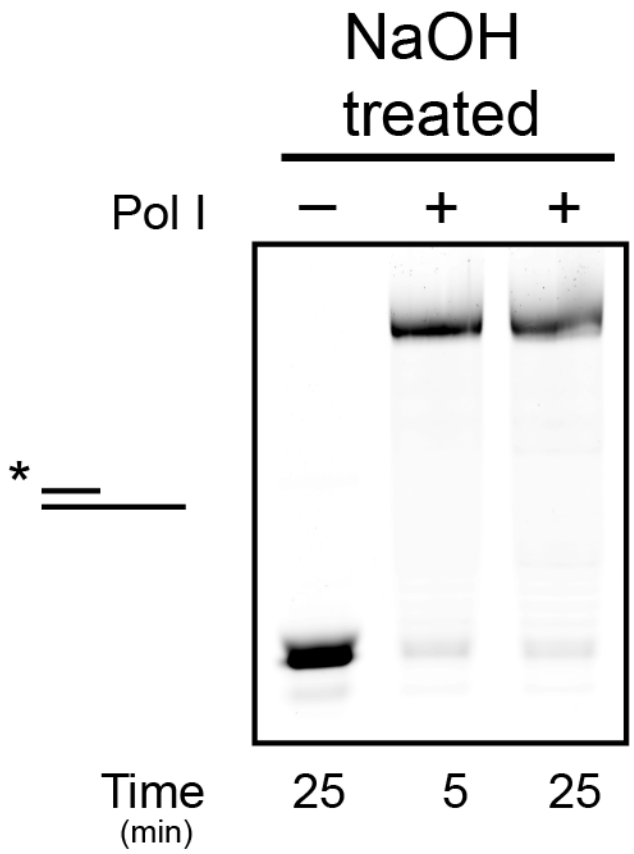

Figure 3.13 Pol I extends a substrate treated with NaOH. A primer extension assay was performed on a primed substrate after treatment with NaOH for 5 and 25 minutes respectively.

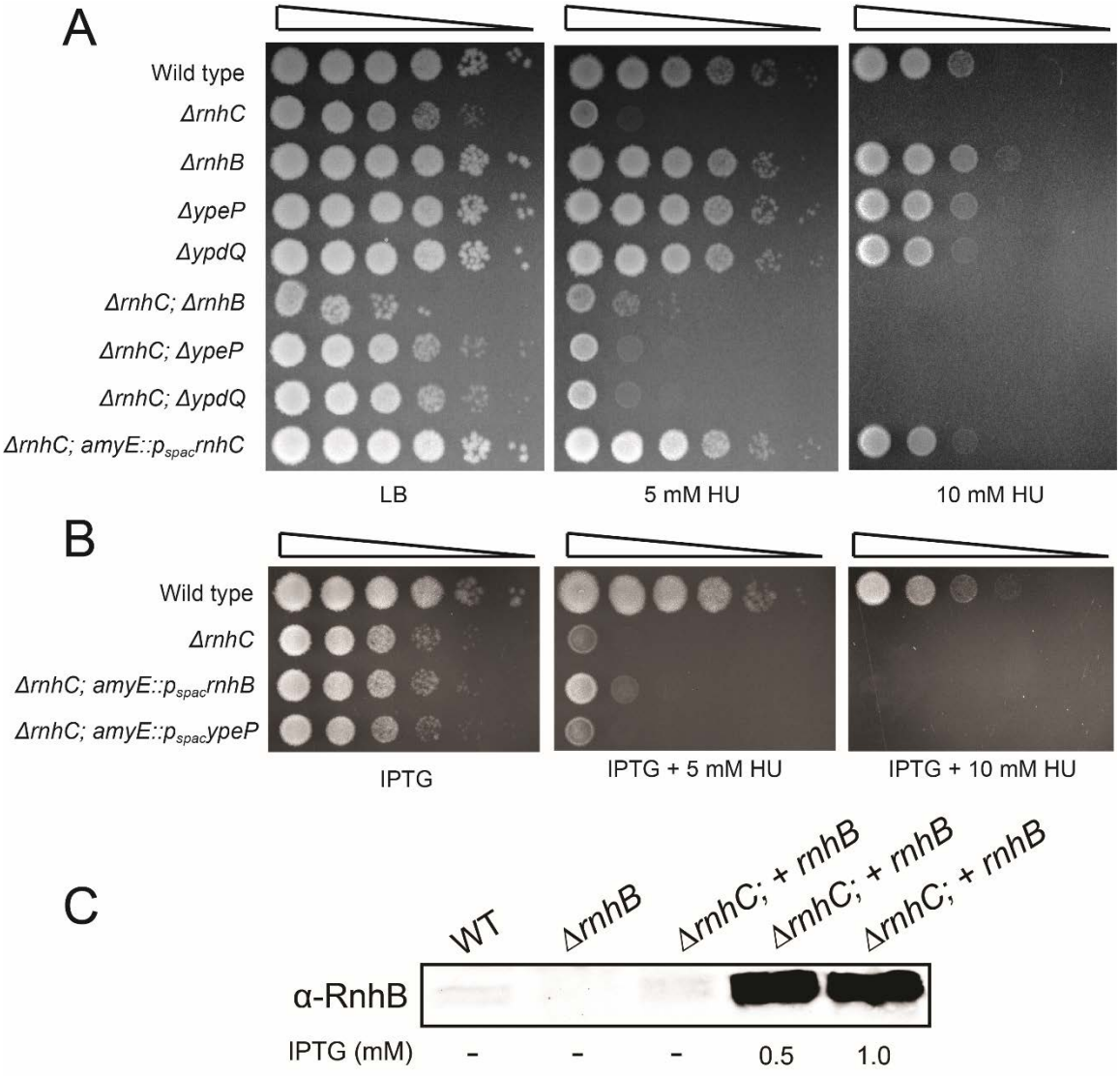

Figure 3.14 Cells lacking RNase HIII are sensitive to HU (A) Spot-titer assay of isogenic *B. subtilis* cells with the indicated genotype were spotted on hydroxyurea (HU). (B) Spot-titer assay of isogenic *B. subtilis* cells with the indicated genotype were spotted on HU. IPTG was included at 1 mM. (C) Immunoblot detection of RNase HII from whole cell extracts. Detection was accomplished using affinity-purified anti-serum raised against RNase HII (RnhB) as the primary antibody.

Table 3.1 Summary of RNase H activity

Substrate	Enzyme	*~In vivo* [Mn^{2+}] 10 µM	*~In vivo* [Mg^{2+}] 1 mM	Mixture
20 mer RNA:DNA	4 nM RNase HII	-0.6% ± 0.3%	-0.4% ± 0.3%	-0.6 ± 1.1%
	2 µM RNase HII	0.4% ± 0.3%	69.8% ± 0.02%	71.4 ± 2.0%
	4 nM RNase HIII	31.6% ± 4.6%	79.3% ± 0.3%	77.2 ± 1.9%
4 embedded rNMPs	4 nM RNase HII	79.2% ± 2.0%	87.9% ± 0.3%	86.8 ± 1.4%
	4 nM RNase HIII	39.5% ± 1.6%	84.0% ± 0.9%	81.4 ± 2.7%
1 embedded rNMP	50 nM RNase HII	82.5% ± 1.0%	58.1% ± 0.1%	83.3 ± 2.8%
	50 nM RNase HIII	9.7% ± 1.0%	6.3% ± 0.7%	12.7 ± 2.2%

*****Error = range of duplicate reactions

Table 3.2 Oligonucleotides used in this study

Primer	Purpose	Sequence
oJR46	pE-SUMO vector	TCGAGCACCACCACCACCACCACTGAG
oJR47	pE-SUMO vector	ACCTCCAATCTGTTCGCGGTGAGCCTCAATAA TATCG
oJR88	rnhB SUMO o/h	CCGCGAACAGATTGGAGGTGTGAATACATTAA CCGTAAAGGACATTAAAGACC
oJR89	rnhB SUMO o/h	TGGTGGTGGTGGTGCTCGATTATCTGAAAGAT TGAACAGGAGCG
oJR90	rnhC SUMO o/h	CCGCGAACAGATTGGAGGTGTGTCCCATTCA GTGATAAAAGTATCG
oJR91	rnhC SUMO o/h	TGGTGGTGGTGGTGCTCGACTATGAACGTTTT TTATCAGCAAGGCG
oJR94	*rnhB DE/AA* mutation	GTTGcCGcGGTCGGCCG
oJR95	*rnhB DE/AA* mutation	GACCgCGgCAACACCTGCAATC
oJR96	*rnhC DE/AA* mutation	TCTGcCGcAGTCGGAACCGG
oJR97	*rnhC DE/AA* mutation	GACTgCGgCAGAACCGATAACAGACATTCC
oJR101	polA SUMO o/h	TGGTGGTGGTGGTGCTCGATTACCACTGTTTT AAAACGACGTTTTTTTGACCC
oJR102	polA SUMO o/h	CCGCGAACAGATTGGAGGTATGACGGAACGA AAAAAATTAGTGC
oJR145	Complementing substrates	GCAGAGCTAGCTTACGATCG
oJR180	S. aur rnhC SUMO o/h	GAGGCTCACCGCGAACAGATTGGAGGTATGG CGAATATCGTTTTTAAATTGTCGG
oJR181	S. aur rnhC SUMO o/h	GTGGTGGTGGTGGTGCTCGATCAAAGAGGCT TTAAAATTTTTTGGG
oJR205	G. stear rnhC SUMO o/h	CCGCGAACAGATTGGAGGTTTGTCAAACTATA TGATTCAAGCCG
oJR185	G. stear rnhC SUMO o/h	GTGGTGGTGGTGGTGCTCGATCACTTCCGGC GCTTGG
oJR200	Upstream rnhC	GGCTTTACACTTTATGCTTCCTAACCCCCAGA CGTTCACATATG
oJR201	Upstream rnhC	GTAATAATCTCCTTTTTTTACACTTTTCGCTG
oJR203	Downstream rnhC	ATGCGTAAGGAGAAAATACCTACGCTTCCTGA GACAGC
oJR211	ypeP pDR110 o/h	GTCGACTAAGGAGGTATACATATGAAGCTCAG ACCGCATTTG
oJR212	ypeP pDR110 o/h	CGAATTAGCTTGCATGCGGCTAGCTTATCCAA TATCGTCATCTCCGTTACG
oJR234	1 rNMP substrate	800IR/CGA TCG TAA rGCT AGC TCT GC
oJR235	4 rNMP substrate	800IR/CGATCGTArArGrCrUAGCTCTGC
oJR227	all RNA substrate	800IR/rCrGrArTrCrGrTrArArGrCrUrArGrCrTrCrTr CrC
oJR262	pDR110	GCTAGCCGCATGCAAGCTAA
oJR263	pDR110	ATGTATACCTCCTTAGTCGACTAAGCTTA
oJR276	ypdQ SUMO o/h	CCGCGAACAGATTGGAGGTATGCCTACAGAA

		ATATATGTAGACGGC
oJR277	ypdQ SUMO o/h	CTCAGTGGTGGTGGTGGTGGTGCTCGATTAA TTCTTTTCA
oJS476	pMiniMad2.0	GGTATTTTCTCCTTACGCATCTGTGC
oJS477	pMiniMad2.0	GGAAGCATAAAGTGTAAAGCCTGGG

***** Lower case bases represent mutation locations. Lower case r before a base indicates a ribonucleotide

Table 3.3 List of plasmids used

Plasmid	Vector	Insert
pJR17	pE-SUMO	*rnhB*
pJR18	pE-SUMO	*rnhB (D78A, E79A)*
pJR19	pE-SUMO	*rnhC*
pJR20	pE-SUMO	*rnhC (D100A, E101A)*
pJR22	pE-SUMO	*polA*
pJR29	pE-SUMO	*ypeP*
pJR40	pE-SUMO	*S. aur rnhC*
pJR42	pE-SUMO	*G. stear rnhC*
pJR46	pUC57	*G. stear rnhC*
pJR49	pDR110	*ypeP*
pWG2	pDR110	*rnhB*
pWG3	pDR110	*rnhC*
pJR55	pE-SUMO	*ypdQ*

Table 3.4 List of strains used

Strain Name	Species	Genotype
JRR1	*E. coli* BL21	*Native*
JRR2	*E. coli* MC1061	*Native*
JRR25	*B. subtilis* PY79	*ΔrnhB*
JRR26	*B. subtilis* PY79	*ΔrnhC*
JRR27	*B. subtilis* PY79	*Native*
JRR30	*B. subtilis* PY79	*ΔrnhC; amyE::p_{spac}ypeP*
JRR33	*B. subtilis* PY79	*ΔrnhC; rnhB::spec*
JRR59	*B. subtilis* PY79	*ypeP::lox scar*
JRR60	*B. subtilis* PY79	*ypdQ::lox scar*
JRR61	*B. subtilis* PY79	*ΔrnhC; ypeP::lox scar*
WGH29	*B. subtilis* PY79	*ΔrnhC; amyE::p_{spac}rnhB*
JRR68	*B. subtilis* PY79	*ΔrnhC; ypdQ::lox scar*
WGH28	*B. subtilis* PY79	*ΔrnhC; amyE::p_{spac}rnhC*

CHAPTER IV

IV. Concluding remarks and future directions

4.1 Introduction

In Chapter I of this dissertation, I reviewed the types of RNA-DNA hybrids that form inside of cells, the mechanisms for how they are created, and what is known about the enzymes and pathways responsible for their resolution. In Chapter II, I discussed single rNMP incorporations, their repair, and the consequences if rNMPs remain in the genome. In Chapter III, I discussed how metal availability inside the cell dictates substrate specificity of bacterial RNase HII and HIII *in vivo*. In this chapter, I will briefly revisit what was learned in Chapters II and III as well as discuss what questions remain unanswered. I will end by discussing the approaches that should be taken to address these outstanding questions.

4.2 Ribonucleotide excision repair

The importance of RNase H2 in the process of ribonucleotide excision repair (RER) was well established in eukaryotic organisms prior to this work. The

RER pathway had been reconstituted using yeast proteins (Sparks et al., 2012) and the consequences of its absence had been studied in yeast, mice, and humans (Hiller et al., 2012; Nick McElhinny et al., 2010b; Reijns et al., 2012). Studies in bacteria have been far less comprehensive and striking phenotypes for strains with an *rnhB* deletion have not been identified, though it was known that genomic DNA from these cells contain more RNA and can cause a mild increase in mutation rate in some bacteria (McDonald et al., 2012; Yao et al., 2013). In *E. coli*, RNase HII was identified as important for bacterial RER (Vaisman et al., 2014), but the function of RNase HII in other bacteria had not been explored.

In Chapter II we demonstrated the effect of RNase HII deficiency on genome stability in the Gram-positive bacterium *Bacillus subtilis*. Loss of RNase HII results in a specific mutation pattern of GC → AT transitions at a distinct sequence (3'-GCC(C/T)T-5') on the lagging strand of the chromosome (Schroeder et al., 2017). We went on to provide a possible mechanism for this type of mutation involving nucleotide excision repair (NER) serving a backup role to RNase HII in RER with DnaE replicating the remaining gap resulting in the observed mutagenesis (Schroeder et al., 2017).

A few questions remain unresolved regarding this topic. Although we provide evidence of NER acting as the backup pathway to RNase HII in RER (Figure 2.9 and Table 2.2), we cannot exclude the possibility that the mutagenesis is not NER dependent. To answer this question, sequencing of mutation accumulation lines from a Δ*rnhB;* Δ*uvrA* (NER) double mutant would

need to be analyzed and show the absence of the sequence-context dependent GC → AT transitions. We provided evidence that the mutation spectrum observed between Δ*rnhB* alone and Δ*rnhB*; Δ*uvrA* are significantly different, but the data do not conclusively show that the sequence and strand context-dependent mutagenesis is no longer present.

Another unresolved question is how the sequence-context dependence mutagenesis is context and lagging strand specific? DnaE is responsible for extending Okazaki fragments so the vast majority of the replication it performs is on the lagging strand (Sanders et al., 2010), possibly explaining the strand-specific dependence. In addition, the bulk of rNMP incorporations expected are rATP (Yao et al., 2013). If rATP were placed across from the thymine in the mutagenic sequence context 3'-GCC(C/T)T-5', DnaE would encounter the two cytidines just before replicating over the guanosine. DNA polymerases show an increase in mutation rate when they replicate through short homopolymer runs or high GC content and the mutagenic sequence motif contains both (Petruska and Goodman, 1985). Further, the most common mismatch made by replicative DNA polymerases is a G:T pairing [for review see (Kunkel, 2004)], possibly explaining this sequence-context mutagenesis. This along with DnaE lacking a proofreading domain and our finding that DnaE is more mutagenic than Pol I over the mutagenic sequence (Figure 2.8) could explain the transitions observed in the Δ*rnhB* MA lines. It also remains possible that rNTPs are placed into the genome more commonly in the 3'-GCC(C/T)T-5' sequence context by one or both of the replicative polymerases.

A few other broad questions still remain. What effect does the loss of RNase HII have on the genomes of other bacteria? It seems plausible that the types of genome instability observed could vary since Δ*rnhB E. coli* show no increase in mutagenesis while *B. subtilis* cells do (Yao et al., 2013) and yeast deficient for RNase H2 show a completely different mutagenic signature (Kim et al., 2011). This could indicate that the backup pathways for RER differ between species. Additionally, how are bacteria seemingly tolerant of an RNase HII deletion while higher order organisms such as mice and humans have drastic phenotypes with an RNase H2 deficiency? This could be because RNase H2 contributes to Okazaki fragment maturation or another unknown process in eukaryotes that is less prominent in bacteria. Another explanation could be that multiple redundancies are built into the RER pathway specific to bacteria. This would explain why loss of RNase HII has little or no effect on mutation rate, and does not appear to be vital for strand discrimination during MMR in bacteria. It is also possible that bacteria are simply more tolerant to RNA left in the genome for other reasons not discussed above.

4.3 Role of RNase HIII *in vivo*

RNase HIII does not exist in eukaryotes and it only appears in the genomes of a subset of Gram-negative and Gram-positive bacteria, including several potentially infectious genera (Randall et al., 2017). RNase HIII had been biochemically characterized and identified as functionally analogous to RNase HI prior to this work (Ohtani et al., 1999a). However, *Chlamydophila pneumoniae*

RNase HIII demonstrated a novel activity for this enzyme with Mn^{2+} as a catalytic cation (Lu et al., 2012a). This brought into question the potential contributions RNase HIII may have *in vivo* and suggested that RNase HIII may not be functionally analogous to RNase HI. In addition, this opened up the possibility that substrate specificity of all RNase H enzymes might be influenced by metal availability more than previously appreciated. Further, in the prior literature RNase H enzymes had not been tested with physiologically relevant metal concentrations making us question some of the reported activities.

This led to the work performed and discussed in Chapter III of this document where we discovered that *B. subtilis* RNase HIII, as well as two other RNase HIII enzymes (*S. aureus* and *G. stearothermophilus*), cleave at single rNMPs embedded in DNA with manganese (Randall et al., 2017). In addition, we demonstrated that RNase HIII processed substrates could also serve as suitable substrates for extension by Pol I (Randall et al., 2017), further questioning the roles of RNase HII and HIII *in vivo*. However, using physiologically relevant metal concentrations *in vitro* we found that only RNase HII could efficiently cleave at single rNMPs and only RNase HIII could efficiently cleave junction-less RNA-DNA hybrids such as R-loops (Randall et al., 2017). We then demonstrated that cells deficient for RNase HIII are sensitive to hydroxyurea (HU) while RNase HII cells were indistinguishable from wild type. These results provide evidence that RNase HII and HIII have distinct contributions to RNA-DNA hybrid resolution *in vivo*. Shortly before publication a manuscript was published suggesting that

RNase HIII resolves R-loops at genes transcribed head-on to replication (Lang et al., 2017), further supporting the conclusions drawn in from our work.

So what is still left to be learned about RNase HIII? First, are R-loops the primary substrate of RNase HIII *in vivo*? RNase HIII has still not been shown to cleave an R-loop substrate *in vitro* (see section 4.3 and Appendix A) and the *in vivo* data supporting a role for RNase HIII resolving R-loops *in vivo* is correlative (Lang et al., 2017). Further, our work demonstrates that RNase HIII can also efficiently cleave longer stretches of RNA-DNA hybrids with RNA-DNA junctions. These include ribopatches and Okazaki fragments, suggesting that RNase HIII resolves a broader range of substrates than just R-loops. The mechanism of Okazaki fragment maturation in bacteria is still not well understood especially in organisms lacking RNase HI, such as *B. subtilis*. Also, in eukaryotes RNase H2 has been shown to contribute to Okazaki fragment maturation (Kao and Bambara, 2003). With these results in mind, RNase HII, RNase HIII or both proteins could be contributing to Okazaki fragment maturation in *B. subtilis.* In addition, the role RNase HIII plays in genome stability still remains unclear. RNase HIII deficient cells have yet to be studied under normal growth conditions and it remains unknown what effect a deficiency may have on chromosome stability or cell stress and survival. Mutation accumulation lines could be performed on RNase HIII deficient cells to help answer some of these questions (see below). Lang et al. also demonstrated that RNase HIII was important for *Listeria* pathogenesis and *B. subtilis* cells deficient in RNase HIII have several sensitivities to a wide range of stressors (Lang et al., 2017). This opens up the

possibility of RNase HIII inhibition as a possible antibiotic target or adjuvant. RNase HIII is common among pathogenic bacteria including *Staphylococcus aureus*, *Bacillus anthracis, Enterococcus faecalis*, *Streptococcus pyogenes* and *Streptococcus pneumonia*. It is certainly worth exploring RNase HIII inhibition with regard to infectious bacteria and antibiotic sensitivity.

4.3 Discussion of ongoing work

The lab continues to study both RNase HII and RNase HIII with respect to many of the ideas mentioned above (see Appendix A). First, we are continuing to study RNase HIII deficient cells, especially with consideration to genome stability under normal growth conditions. RNase HIII appears to be sensitive to many DNA damaging agents, it is enriched for RecA-GFP focus formation and constitutively positive for induction of SOS regulated genes suggesting that RNase HIII deficient cells are under considerable genotoxic stress during normal growth (see Appendix A, Fig A.1 and A.2). In addition we have biochemically verified that RNase HIII cleaves an R-loop structure under physiologically relevant metal concentrations *in vitro* while RNase HII cannot (Fig A.3), further supporting an increase in R-loop formation in RNase HIII deficient cells. The lab also has performed mutation accumulation lines for Δ*rnhC* revealing chromosome instability and high transversion mutation density at head-on genes, especially near the terminus of replication. In collaboration with Dr. Peter Freddolino's Lab at the University of Michigan we are performing DNA-RNA immunoprecipitation sequencing (DRIP-seq) on several *B. subtilis* strains lacking

RNase H, including Δ*rnhC,* to identify the locations of RNA-DNA hybrid enrichment *in vivo*. This will further resolve the intracellular functions of both RNase HII and RNase HIII.

In addition to the R-loop focused work discussed above, the lab has become interested in Okazaki fragment maturation in *B. subtilis*. Determining how Okazaki fragments are processed is of fundamental importance to understanding the process of DNA replication. Because many bacteria lack RNase HI, it remains unclear if RNase HII, RNase HIII or both contribute to Okazaki fragment processing. Both enzymes demonstrate the ability to cleave ribopatches under physiological conditions (see Chapter III) (Randall et al., 2017), but Okazaki fragments, which only contain a 3' RNA-DNA junction, have yet to be published. I present evidence that both RNase HII and RNase HIII are efficient at cleaving an Okazaki fragment (see Fig A.4), however the substrate is processed differently. It would be important to develop a system for reconstituting Okazaki fragment maturation with purified *B. subtilis* proteins. Using this system we could reveal the mechanistic contribution of RNase H enzymes (RNase HII, RNase HIII) and 5'-3' exonucleases (Pol I, ExoA) to the removal of RNA primers during replication in bacteria lacking RNase HI, answering a long-standing question.

APPENDIX

APPENDIX A

A. RNase HIII resolves R-loops

A.1 Ongoing work with *Bacillus subtilis* RNase H enzymes

This section covers ongoing work both *in vivo* and *in vitro* to determine the intracellular functions of RNase HIII. The current results show that RNase HIII is primarily responsible for R-loop removal and that both RNase HII and HIII process Okazaki fragments *in vitro*.

A.1.1 *B. subtilis* Δ*rnhC* cells are sensitive to DNA damage

Previous work has demonstrated that deletion of RNase HIII results in varied sensitivities to stress (Lang et al., 2017; Randall et al., 2017). We expanded our investigation to determine the contribution of RNase HII and HIII to growth in the presence of DNA damage. The goal was to understand if RNase HIII function was specific to the chemotherapeutic agent hydroxyurea or instead was more generally important for growth in the presence of exogenous DNA damage. To this end, isogenic *B. subtilis* strains with Δ*rnhB* (RNase HII) and Δ*rnhC* (RNase HIII) were generated and challenged with mitomycin C (MMC), phleomycin, and methyl methanesulfonate (MMS) in a spot titer assay. These

agents cause several different lesions including inter and intra-strand crosslinks for MMC (Tomasz et al., 1986), single and double stranded DNA breaks for phleomycin (Sleigh, 1976), and DNA methylation for MMS (Bignami et al., 2000). We did not detect any differences in growth for Δ*rnhB* cells relative to the wild type control in the presence of MMC, phleomycin or MMS. Strikingly, Δ*rnhC* cells were highly sensitive to growth in the presence of each DNA damaging agent examined (Fig A.1 rows 1-3). To further investigate the observed sensitivity to DNA damage, RNase HII, DNA polymerase I (Pol I), and RNase HIII were overexpressed from an ectopic locus in Δ*rnhC* cells in an attempt to rescue the growth inhibition. Our results show that only ectopic expression of *rnhC* (RNase HIII) alleviated growth inhibition of all three DNA damaging agents assayed (Fig A.1 bottom row). Overexpression of *rnhB* (RNase HII) showed feeble rescue to phleomycin (Fig A.1 fourth row) while overexpression of *polA* (Pol I) provided no rescue and instead further sensitized cells to MMC and MMS induced damage (Fig A.1 fifth row). With these results, we conclude that Δ*rnhC* cells are highly sensitive to a broad-range of DNA damaging agents in a process independent of RNase HII and Pol I function [Figure A.1 and (Randall et al., 2017)].

A.1.2 The DNA damage response is constitutively induced in *ΔrnhC* cells

Since Δ*rnhC* cells show a significant sensitivity to a variety of DNA damaging agents we asked if the loss of RNase HIII triggered the DNA damage response. In bacteria, the DNA damage response is known as the SOS-response. SOS is activated when RecA binds to excess ssDNA resulting in LexA

auto-cleavage and induction of LexA repressed genes [for review see (Simmons et al., 2008, Lenhart et al., 2012)]. To investigate SOS induction in Δ*rnhC* cells we used a well characterized single cell reporter construct measuring RecA-GFP formation (Lenhart et al., 2014, Simmons et al., 2009, Simmons et al., 2007, Walsh et al., 2014). We asked if Δ*rnhC* cells grown in minimal medium at 30 °C in the absence of exogenous DNA damage showed a higher percentage of RecA-GFP foci as compared with wild type cells. We found that 10.6 ± 4.3 percent of wild type cells contained RecA-GFP foci, the same result published previously (Lenhart et al., 2014, Simmons et al., 2009, Simmons et al., 2007, Walsh et al., 2014). In Δ*rnhC* cells we observed a significant increase in the percentage of cells with RecA-GFP foci to 26.4 ± 6.3 (Fig A.2 A). The increase in the percentage of cells with RecA-GFP foci indicates of a problem with DNA replication or repair arising from endogenous sources in Δ*rnhC* cells. To directly test for SOS induction we performed RNA-seq in Δ*rnhC* cells and observed induction of the SOS-regulon (data not shown).

To investigate how SOS-induction affects Δ*rnhC* growth in the presence of DNA damage we tested *lexA* alleles that were constitutively SOS induced or incapable of SOS induction in the Δ*rnhC* background following challenge with MMC. As shown in Figure A.1, Δ*rnhC* cells are sensitive to MMC relative to wild type (Fig A.2B top two rows). If the growth inhibition of Δ*rnhC* cells is caused by constitutive SOS induction we would expect that combining the LexA non-cleavable allele (*lexA[G92D*]) with an *rnhC* deficiency would rescue the sensitivity. Instead, we observed that *rnhC::erm; lexA[G92D]* cells were severely

growth impaired on LB in the absence of DNA damage, indicating that the ability to turn on the SOS response is important for growth of Δ*rnhC* cells. Further, we show that cells with constitutive SOS induction are not growth impaired as determined by a spot-titer assay with the Δ*lexA* bearing strain, while the *lexA* non-cleavable (*lexA[G92D]*) was slightly sensitized (Fig A.2C last three rows). With these results, we conclude that Δ*rnhC* cells are SOS-induced and that SOS induction is important for cell growth in the absence of RNase HIII.

A.1.3 RNase HIII cleaves R-loops *in vitro*

In Figure A.1 we demonstrated that an RNase HIII deletion (Δ*rnhC*) was sensitive to DNA damage and that the overexpression of either RNase HII or Pol I was unable to mitigate Δ*rnhC* sensitivity. Since RNase HII has been shown to function in RER (Yao et al, 2013; Schroeder et al, 2017) and Pol I serves in several DNA repair processes in addition to Okazaki fragment maturation, (Kornberg and Baker, 1992; Vaisman et al., 2014) we considered the possibility that the Δ*rnhC* sensitivity is due to R-loop formation; especially considering indirect evidence of R-loop accumulation at head-on genes in Δ*rnhC B. subtilis* cells (Lang et al., 2017).

There is no direct evidence demonstrating that RNase HIII cleaves R-loops *in vitro*. To test if RNase HIII was sufficient to cleave R-loops *in vitro* we generated an R-loop substrate using three oligonucleotides with a 22 base pair RNA-DNA hybrid (Fig A.3A) (Tian & Alt, 2000). The substrate was labeled on the strand modeling the DNA coding strand and the oligonucleotide representing the

mRNA on the 5' end (see Fig A.3A). To demonstrate the R-loop was assembled correctly we show that when all three oligos are annealed a super shift is observed and both labels overlapped in a native-PAGE (Fig A.3B). To further test proper assembly of the R-loop the gel purified substrate was treated with mung bean nuclease (MBN), a single stranded RNA and DNA endo/exo-nuclease (Kowalski et al., 1976), followed by denaturing urea-PAGE to resolve the products. The mRNA strand alone was also treated with 0.3 M NaOH to generate a ladder for single nucleotide resolution. The mRNA was susceptible to MBN, but as expected the middle portion was protected in the R-loop structure (Fig A.3C). In addition, the DNA strand shows protection at the ends only when in the R-loop substrate, demonstrating protection of the double stranded DNA regions (Fig A.3C). These results show that the R-loop has been properly assembled.

The R-loop was then incubated with RNase HII (HII), RNase HIII (HIII), and a catalytically inactive version of RNase HIII (HIII DE/AA) with a mixture of Mg^{2+} and Mn^{2+} previously shown to support activity of both proteins *in vitro* (Randall et al., 2017)(Fig A.3D). RNase HII and HIII DE/AA failed to incise the R-loop at 50 nM concentration following a 10 minute reaction at 25°C. However RNase HIII began to cleave the substrate almost immediately (0.1 min) with a 4 nM protein concentration (Fig A.4D). Using NaOH cleavage of the RNA oligonucleotide as a single nucleotide resolution ladder, we mapped the locations of cleavage on the RNA portion of the R-loop (Fig A.4E). This result shows that RNase HIII efficiently cleaves an R-loop, providing direct evidence that RNase HIII is sufficient for R-loop resolution *in vitro*.

A.1.4 RNase HII and RNase HIII cleave an Okazaki fragment like substrate

Although there is evidence RNase HIII resolves R-loops in this document and others (Lang et al., 2017), the possibility exists that the sensitivities observed in Figure A.1 could be due to a deficiency in Okazaki fragment maturation independent of RNase HII and Pol I activity. To investigate this possibility further, we assembled an Okazaki fragment substrate with a 3' label and then tested the ability of RNase HII and HIII to cleave this substrate *in vitro* (Fig A.4). The Okazaki fragment was cleaved by both RNase HII and RNase HIII at 4 nM concentration after a 10-minute incubation. RNase HII showed more efficient incision relative to RNase HIII (Fig A.4A). The substrate was also incubated with 0.3 M NaOH to hydrolyze 3' to each rNMP generating a single nucleotide resolution ladder. We used this ladder to map the sites of cleavage for each enzyme (Fig A.4B). These results show that both RNase HII and HIII cleave Okazaki fragments, further supporting the hypothesis that the observed Δ*rnhC* sensitivities are likely due to a deficiency in R-loop processing and not Okazaki fragment formation since RNase HII should be able to compensate for loss of RNase HIII with respect to RNA removal from Okazaki fragments. Further experiments are necessary to determine if RNase HII and HIII process Okazaki fragments with the same efficiency.

A.2 Discussion

With the data above in mind, it appears RNase HIII has a primary role in R-loop processing while RNase HII and RNase HIII are able to participate in Okazaki fragment maturation. It was purposed in Lang et al. that the induction of stress response genes oriented head-on to replication may cause an increase in R-loops resulting in Δ*rnhC* sensitivity. For DNA damage and the SOS-response this doesn't appear to be the case because when the SOS response is suppressed in the *lexA* non-cleavable background Δ*rnhC* cells have increased sensitivity to MMC (Fig A.2). This suggests that the SOS-response is critical for Δ*rnhC* cell survival rather than detrimental, as suggested for other stress response genes including genes important for osmotic stress as described in Lang et al.

From the data presented in Figures A.3 and A.4 it would seem that under physiological metal concentrations only RNase HIII can process R-loops, but both RNase HII and HIII can cleave Okazaki fragments. This suggests that either HII, HIII, or both may be responsible for Okazaki fragment maturation *in vivo*. This may help to explain the severe growth defect we observe with cells bearing deletions for RNase HII and HIII (Fig 3.14 sixth row). If both enzymes can process Okazaki fragments *in vivo* the double deletion would result in dysfunctional Okazaki fragment maturation and decreased cell viability. It is unclear if RNase HIII activity on Okazaki fragments is contributing to the sensitivities observed here, or in Lang et al., 2017. Future experiments will be

designed to understand if Δ*rnhC* phenotypes are R-loop-mediated or caused by a combination of persistent R-loop formation and immature Okazaki fragments.

A.3 Material and Methods

A.3.1 Spot titer assay

Assays were performed essentially as described in (Randall et al., J Bac. 2017) (3.6.9). Briefly, a single colony of the indicated strain was used to inoculate 3 ml of LB media and grown to an OD600 between 1 and 1.5. Cultures were then normalized to an OD600 = 1 in a 0.85% saline solution and serial diluted to 10^{-5}. A total volume of 5 µl was spotted for each dilution on LB or LB containing the indicated concentrations of MMC, MMS, phleomycin and or IPTG. Plates were then incubated overnight at 30 ºC and imaged the following morning.

A.3.2 RecA-GFP fluorescent imaging

Cells were grown in $S7_{50}$ minimal medium at 30 °C. All imaging and scoring were done as described previously (Lenhart et al., 2014, Simmons et al., 2009, Simmons et al., 2007, Walsh et al., 2014).

A.3.2 Protein Purification

The RNase HII, RNase HIII and RNase HIII DE/AA purifications were completed as published previously (Schroder et al., 2017). The procedure is also described earlier in this document (see 2.6.3).

A.3.4 R-Loop substrate formation

The R-loop substrate, sequence and structure was based on that used in (Tian & Alt, 2000). The structure was achieved by first annealing 10 µM oJR336 and oJR335 in a 50 µl solution of annealing buffer (10 mM Tris-HCl pH 8, 50 mM NaCl) by first heating to 80 °C and then cooling to 25 °C. Next oJR332 was added to 10 µM and the solution was heated to 40 °C and again allowed to cool to 25 °C. Native-PAGE was then performed and the R-loop substrate was extracted from the 8% polyacrylamide gel (see below) and diluted to a 0.5 µM final concentration.

A.3.5 Polyacrylamide gel extraction of the R-loop substrate

The R-loop containing band was identified using a LI-COR Odyssey imager and was extracted from the 8% native polyacrylamide gel and crushed using a razor blade. The crushed extracted band was then incubated in 100 µl of elution buffer rotating overnight at 4°C. The tube was centrifuged for 1 minute at 13,000 rpm the following morning and the supernatant containing the R-loop was removed.

A.3.6 Mung bean nuclease digestion

Mung bean nuclease (MBN) was purchased from NEB (#M0250S). Primers oJR336, oJR332 at 0.5 µM or the R-loop substrate was digested with MBN for 1 minute in a 10 µM reaction and stopped with stop buffer (95% formamide, 5 mM EDTA, 0.01% bromophenol blue). 0.3 M NaOH was added to

oJR336 and used as a ladder. Reactions were heated to 100 °C for 2 minutes and immediately snap cooled in an ice bath. 20% Urea-PAGE was performed on all samples and digested products were imaged using a LI-COR Odyssey imager.

A.3.7 RNase H assays

RNase H assays were performed in 20 µl of RNase H buffer (10 mM Tris-HCl pH 8, 50 mM NaCl, 1 mM $MgCl_2$, 10 µM $MnCl_2$). The model for an Okazaki fragment was annealed in RNase H buffer containing 1 µM oJR339 and 2 µM oJR340 by heating the substrate to 100 °C for 1 minute and allowing it to slowly cool back to 25 °C. Each reaction contained 0.5 µM of the R-loop substrate or Okazaki substrate and the indicated concentrations of RNase HII, HIII, or RNase HIII DE/AA. Reactions were incubated for the indicated times and stopped with stop buffer (described above). Reactions were heated to 100 °C for 2 minutes and immediately snap cooled in an ice water bath. 20% urea-PAGE was performed on all samples and digested products were imaged using a LI-COR Odyssey imager. 0.3 M NaOH digestion of the substrates was used as a ladder to indicate the sites of cleavage.

A.4 Notes and Acknowledgements

William Hirst was the first to identify the Δ*rnhC* sensitivities to DNA damaging agents and created WGH28 and WGH29 for these studies. Jeremy Schroeder created the *lexA* variant strains JWS266, JWS267, and JWS268.

Taylor Timko performed the RecA-GFP fluorescent imaging and analysis. All other data were generated by JRR.

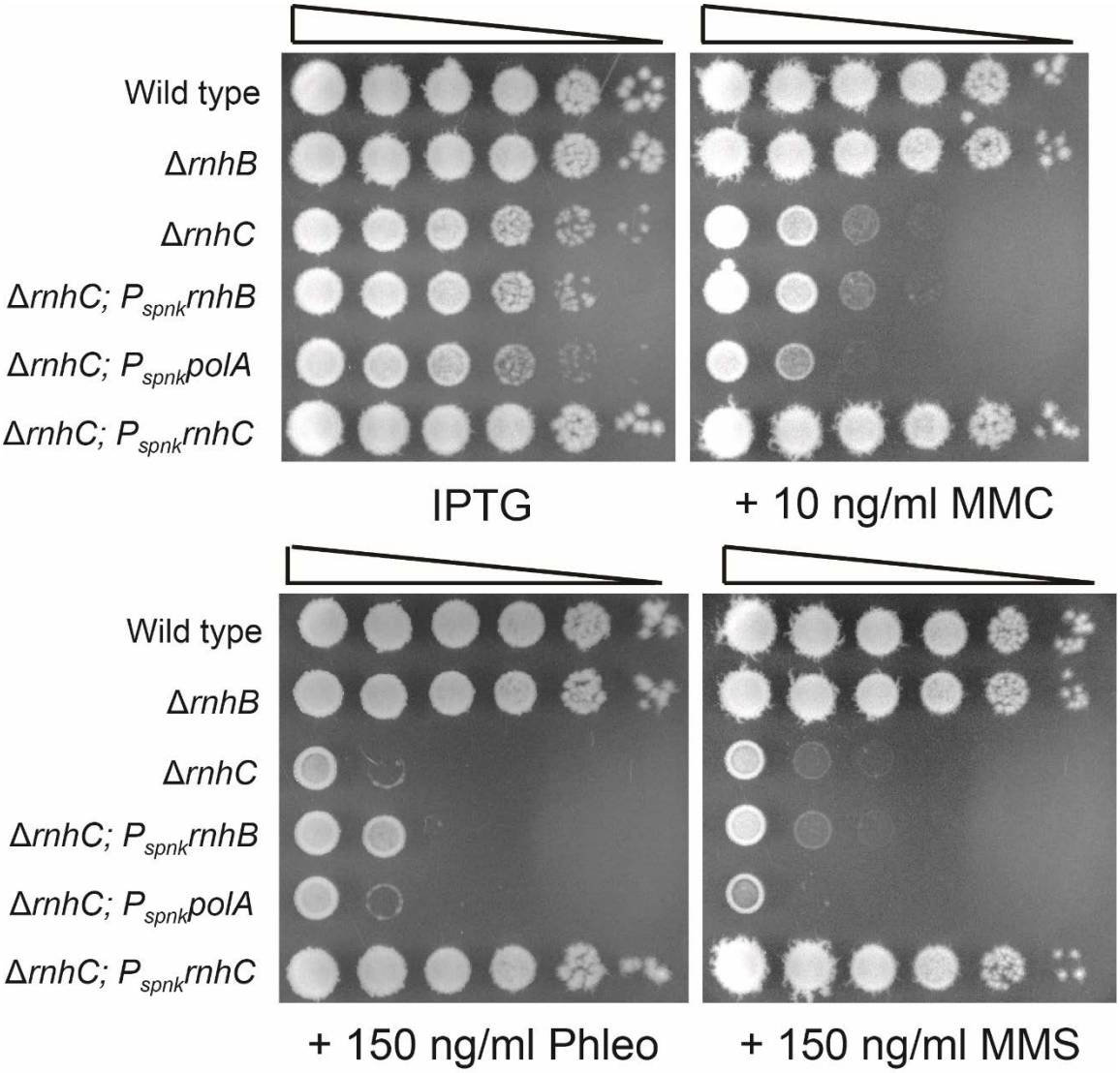

Figure A.1: Δ*rnhC* cells are sensitive to DNA damage. Spot titer assays of various RNase H deficient strains serial diluted 10^{-5} and plated on 1 mM IPTG (top, left) and IPTG plus MMC (top, right) phleomycin (bottom, left) or MMS (bottom, right).

A

Strain	# of cells	Percent RecA-GFP positive
Wild type	1589	10.6 ± 4.3
Δ*rnhC*	1050	26.4 ± 6.3

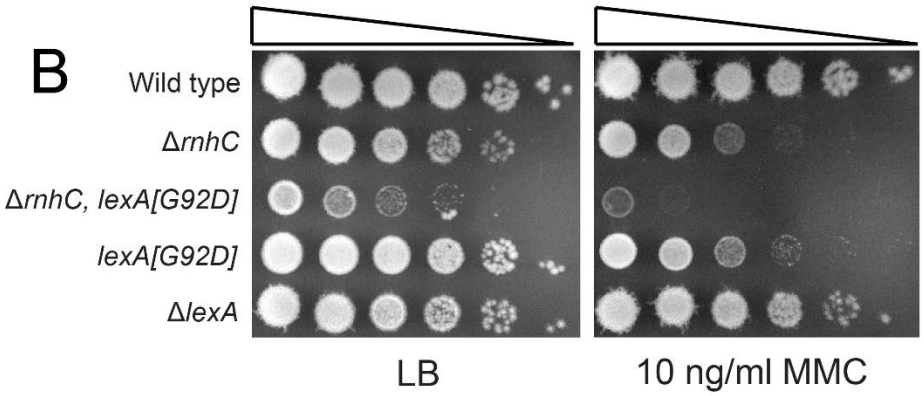

Figure A.2: The DNA damage response is constitutively induced in Δ*rnhC* cells. **(**A) Scoring of wild type and Δ*rnhC* cells for RecA-GFP. The number of cells scored and the percentage of cells with foci are indicated. (B) Spot titer assays of various *lexA* alleles and RNase HIII deficient strains normalized and serial diluted 10^{-5} then plated on LB (left) and 10 ng/ml MMC (right).

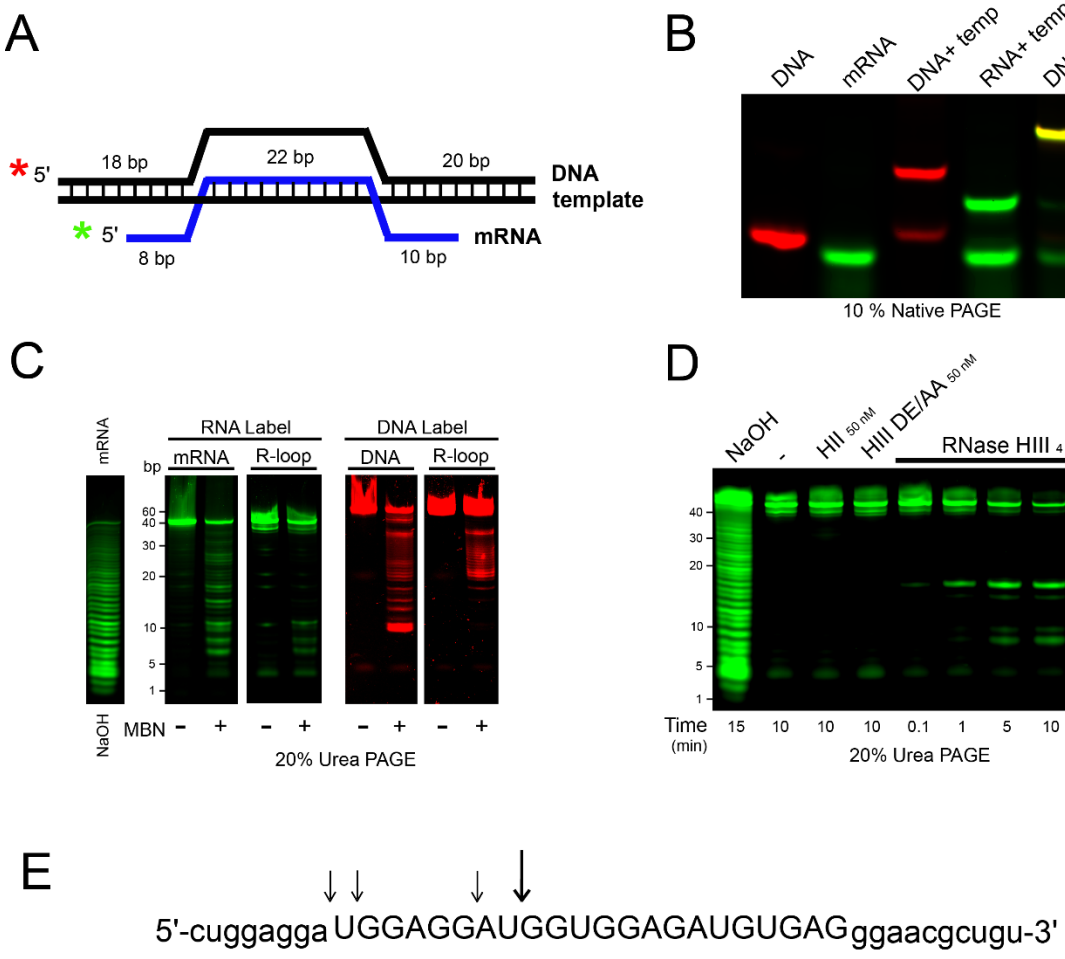

Figure A.3: RNase HIII cleaves R-loops. (A) Schematic of the R-loop structure used in these experiments. (B) EMSA showing the size of different annealed oligonucleotides. (C) Mung bean nuclease (MNB) digestion of the mRNA strand, DNA strand, and R-loop structure. NaOH digestion is used as a ladder. (D) Digestion of the R-loop with 50 nM RNase HII 50 nM RNase HIII DE/AA or 4 nM RNase HIII over time in a 20% denaturing polyacrylamide gel. Again, NaOH digestion of the R-loop is used as a ladder. (E) The sequence of the mRNA strand. Capitalized letters are those annealed in the RNA-DNA hybrid. The locations of cleavage with RNase HIII are shown with arrows. The size of the arrow corresponds to cleavage preference.

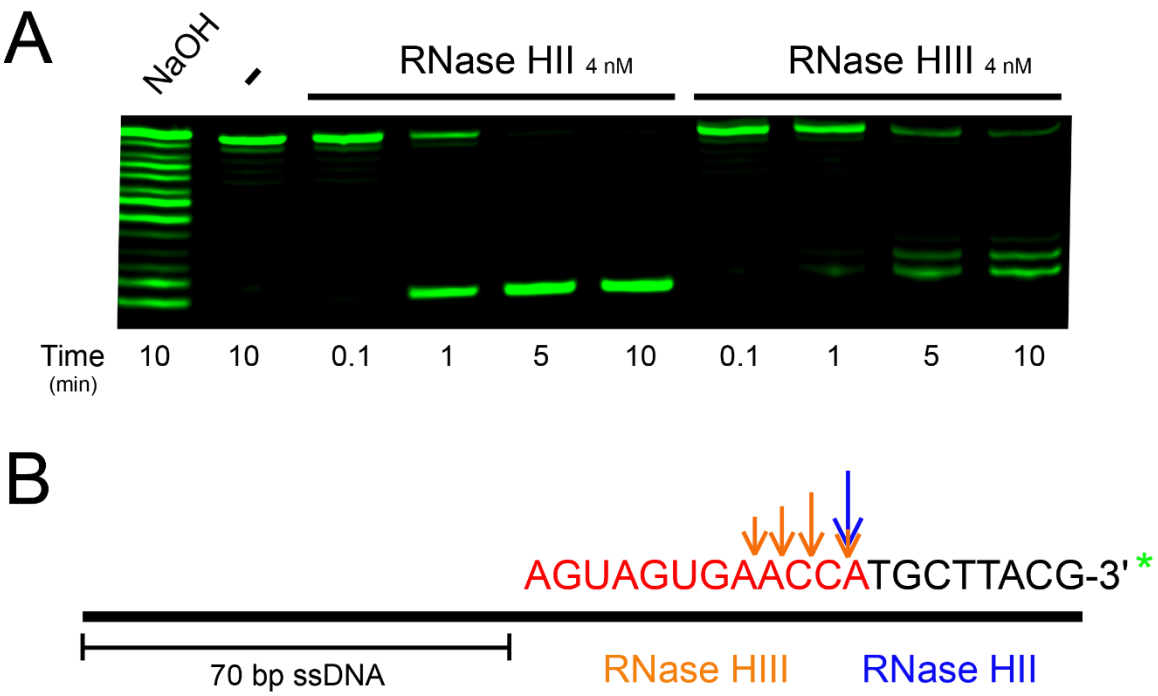

Figure A.4: RNase HII and HIII cleave Okazaki fragments. (A) Cleavage of an Okazaki fragment overtime with 4 μM RNase HII and RNase HIII. NaOH hydrolysis is used to create a ladder. (B) Structure and RNA containing sequence of the Okazaki fragment. Arrows delineate site(s) of cleavage for RNase HII (blue) or RNase HIII (orange). Length of arrow specifies cleavage preference.

Table A.1 Strains used

Name	Genotype	Source
JRR63	Wild type	Youngman et al., 1984
JRR25	Δ*rnhB*	Yao et al, 2013
JRR48	Δ*rnhC*	Yao et al., 2013
WGH28	Δ*rnhC, amyE::*p_{spnk}*rnhC*	Randall et al., 2017
WGH29	Δ*rnhC, amyE::*p_{spnk}*rnhB*	Randall et al., 2017
JRR108	Δ*rnhC, amyE::*p_{spnk}*polA*	this work
JWS266	*lexA[G92D]*	this work
JWS267	Δ*lexA*	this work
JWS268	Δ*rnhC, lexA[G92D]*	this work

Table A.2 Oligonucleotides used

Name	Sequence
oJR339	rArGrUrArGrUrGrArArCrCrATGCTTACG/3IR800CWN/
oJR340	CGTAAGCATGGTTCACTACTTTTTTTTTTTTTTTTTTTTTTTTTTTTT TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
oJR332	/5IRD700/GCGATCCAGAGGTTCACCTGTTTTTTTTTTTTTTTTTTTT GCCGTTGACCACTAACCGCG
oJR335	CGCGGTTAGTGGTCAACGGCCTCACATCTCCACCATCCTCCAGG TGAACCTCTGGATCGC
oJR336	/5IRD800CWN/rCrUrGrGrArGrGrArUrGrgrargrgrarurgrgrurgrgrargrar urgrurgrargrGrGrArArCrGrCrUrGrU

*Lowercase letters is RNA involved in the annealed RNA-DNA hybrid.

BIBLIOGRAPHY

BIBLIOGRAPHY

Aguilera, A., and García-Muse, T. (2012). R loops: from transcription byproducts to threats to genome stability. Mol. Cell *46*, 115–124.

Astatke, M., Ng, K., Grindley, N.D., and Joyce, C.M. (1998). A single side chain prevents Escherichia coli DNA polymerase I (Klenow fragment) from incorporating ribonucleotides. Proc. Natl. Acad. Sci. U. S. A. *95*, 3402–3407.

Bailey, T.L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc. Int. Conf. Intell. Syst. Mol. Biol. *2*, 28–36.

Balakrishnan, L., and Bambara, R.A. (2013). Okazaki fragment metabolism. Cold Spring Harb. Perspect. Biol. *5*.

Bambara, R.A., Murante, R.S., and Henricksen, L.A. (1997). Enzymes and reactions at the eukaryotic DNA replication fork. J. Biol. Chem. *272*, 4647–4650.

Baym, M., Kryazhimskiy, S., Lieberman, T.D., Chung, H., Desai, M.M., and Kishony, R. (2015). Inexpensive multiplexed library preparation for megabase-sized genomes. PloS One *10*, e0128036.

Beck, J., Vogel, M., and Nassal, M. (2002). dNTP versus NTP discrimination by phenylalanine 451 in duck hepatitis B virus P protein indicates a common structure of the dNTP-binding pocket with other reverse transcriptases. Nucleic Acids Res. *30*, 1679–1687.

Berkower, I., Leis, J., and Hurwitz, J. (1973). Isolation and characterization of an endonuclease from Escherichia coli specific for ribonucleic acid in ribonucleic acid-deoxyribonucleic acid hybrid structures. J. Biol. Chem. *248*, 5914–5921.

Bonnin, A., Lázaro, J.M., Blanco, L., and Salas, M. (1999). A single tyrosine prevents insertion of ribonucleotides in the eukaryotic-type phi29 DNA polymerase. J. Mol. Biol. *290*, 241–251.

Braithwaite, D.K., and Ito, J. (1993). Compilation, alignment, and phylogenetic relationships of DNA polymerases. Nucleic Acids Res. *21*, 787–802.

Brown, J.A., and Suo, Z. (2011). Unlocking the sugar “steric gate” of DNA polymerases. Biochemistry (Mosc.) *50*, 1135–1142.

Brown, J.A., Fiala, K.A., Fowler, J.D., Sherrer, S.M., Newmister, S.A., Duym, W.W., and Suo, Z. (2010). A novel mechanism of sugar selection utilized by a human X-family DNA polymerase. J. Mol. Biol. *395*, 282–290.

Buckstein, M.H., He, J., and Rubin, H. (2008). Characterization of nucleotide pools as a function of physiological state in Escherichia coli. J. Bacteriol. *190*, 718–726.

Cai, Y., Geacintov, N.E., and Broyde, S. (2014). Ribonucleotides as nucleotide excision repair substrates. DNA Repair *13*, 55–60.

Cases-Gonzalez, C.E., Gutierrez-Rivas, M., and Ménendez-Arias, L. (2000). Coupling ribose selection to fidelity of DNA synthesis. The role of Tyr-115 of human immunodeficiency virus type 1 reverse transcriptase. J. Biol. Chem. *275*, 19759–19767.

Cerritelli, S.M., and Crouch, R.J. (2009). Ribonuclease H: the enzymes in eukaryotes. FEBS J. *276*, 1494–1505.

Chédin, F. (2016). Nascent Connections: R-Loops and Chromatin Patterning. Trends Genet. TIG *32*, 828–838.

Chon, H., Matsumura, H., Koga, Y., Takano, K., and Kanaya, S. (2006). Crystal structure and structure-based mutational analyses of RNase HIII from Bacillus stearothermophilus: a new type 2 RNase H with TBP-like substrate-binding domain at the N terminus. J. Mol. Biol. *356*, 165–178.

Chon, H., Sparks, J.L., Rychlik, M., Nowotny, M., Burgers, P.M., Crouch, R.J., and Cerritelli, S.M. (2013). RNase H2 roles in genome integrity revealed by unlinking its activities. Nucleic Acids Res. *41*, 3130–3143.

Corn, J.E., and Berger, J.M. (2006). Regulation of bacterial priming and daughter strand synthesis through helicase-primase interactions. Nucleic Acids Res. *34*, 4082–4088.

Crow, Y.J. (2013). Aicardi-Goutières syndrome. Handb. Clin. Neurol. *113*, 1629–1635.

Crow, Y.J., Leitch, A., Hayward, B.E., Garner, A., Parmar, R., Griffith, E., Ali, M., Semple, C., Aicardi, J., Babul-Hirji, R., et al. (2006). Mutations in genes encoding ribonuclease H2 subunits cause Aicardi-Goutières syndrome and mimic congenital viral brain infection. Nat. Genet. *38*, 910–916.

Das, U., and Shuman, S. (2013). Mechanism of RNA 2',3'-cyclic phosphate end healing by T4 polynucleotide kinase-phosphatase. Nucleic Acids Res. *41*, 355–365.

Das, U., Chauleau, M., Ordonez, H., and Shuman, S. (2014). Impact of DNA3'pp5'G capping on repair reactions at DNA 3' ends. Proc. Natl. Acad. Sci. U. S. A. *111*, 11317–11322.

DeAngelis, M.M., Wang, D.G., and Hawkins, T.L. (1995). Solid-phase reversible immobilization for the isolation of PCR products. Nucleic Acids Res. *23*, 4742–4743.

DeLucia, A.M., Grindley, N.D.F., and Joyce, C.M. (2003). An error-prone family Y DNA polymerase (DinB homolog from Sulfolobus solfataricus) uses a "steric gate" residue for discrimination against ribonucleotides. Nucleic Acids Res. *31*, 4129–4137.

Dervyn, E., Suski, C., Daniel, R., Bruand, C., Chapuis, J., Errington, J., Jannière, L., and Ehrlich, S.D. (2001). Two essential DNA polymerases at the bacterial replication fork. Science *294*, 1716–1719.

Drolet, M., Bi, X., and Liu, L.F. (1994). Hypernegative supercoiling of the DNA template during transcription elongation in vitro. J. Biol. Chem. *269*, 2068–2074.

Duigou, S., Ehrlich, S.D., Noirot, P., and Noirot-Gros, M.-F. (2005). DNA polymerase I acts in translesion synthesis mediated by the Y-polymerases in Bacillus subtilis. Mol. Microbiol. *57*, 678–690.

Duquette, M.L., Handa, P., Vincent, J.A., Taylor, A.F., and Maizels, N. (2004). Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. Genes Dev. *18*, 1618–1629.

Dutra, B.E., and Lovett, S.T. (2006). Cis and trans-acting effects on a mutational hotspot involving a replication template switch. J. Mol. Biol. *356*, 300–311.

Eder, P.S., Walder, R.Y., and Walder, J.A. (1993). Substrate specificity of human RNase H1 and its role in excision repair of ribose residues misincorporated in DNA. Biochimie *75*, 123–126.

van der Ende, A., Baker, T.A., Ogawa, T., and Kornberg, A. (1985). Initiation of enzymatic replication at the origin of the Escherichia coli chromosome: primase as the sole priming enzyme. Proc. Natl. Acad. Sci. U. S. A. *82*, 3954–3958.

Fukushima, S., Itaya, M., Kato, H., Ogasawara, N., and Yoshikawa, H. (2007). Reassessment of the in vivo functions of DNA polymerase I and RNase H in bacterial cell growth. J. Bacteriol. *189*, 8575–8583.

Gao, G., Orlova, M., Georgiadis, M.M., Hendrickson, W.A., and Goff, S.P. (1997). Conferring RNA polymerase activity to a DNA polymerase: a single residue in reverse transcriptase controls substrate selection. Proc. Natl. Acad. Sci. U. S. A. *94*, 407–411.

Gardner, A.F., and Jack, W.E. (1999). Determinants of nucleotide sugar recognition in an archaeon DNA polymerase. Nucleic Acids Res. *27*, 2545–2553.

Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. ArXiv12073907 Q-Bio.

Genschel, J., and Modrich, P. (2003). Mechanism of 5'-directed excision in human mismatch repair. Mol. Cell *12*, 1077–1086.

Genschel, J., Kadyrova, L.Y., Iyer, R.R., Dahal, B.K., Kadyrov, F.A., and Modrich, P. (2017). Interaction of proliferating cell nuclear antigen with PMS2 is required for MutLα activation and function in mismatch repair. Proc. Natl. Acad. Sci. U. S. A. *114*, 4930–4935.

Ghodgaonkar, M.M., Lazzaro, F., Olivera-Pimentel, M., Artola-Borán, M., Cejka, P., Reijns, M.A., Jackson, A.P., Plevani, P., Muzi-Falconi, M., and Jiricny, J. (2013). Ribonucleotides misincorporated into DNA act as strand-discrimination signals in eukaryotic mismatch repair. Mol. Cell *50*, 323–332.

Hall, B.M., Ma, C.-X., Liang, P., and Singh, K.K. (2009). Fluctuation analysis CalculatOR: a web tool for the determination of mutation rate using Luria-Delbruck fluctuation analysis. Bioinforma. Oxf. Engl. *25*, 1564–1565.

Haruki, M., Tsunaka, Y., Morikawa, M., and Kanaya, S. (2002). Cleavage of a DNA-RNA-DNA/DNA chimeric substrate containing a single ribonucleotide at the DNA-RNA junction with prokaryotic RNases HII. FEBS Lett. *531*, 204–208.

Helmann, J.D. (2014). Specificity of metal sensing: iron and manganese homeostasis in Bacillus subtilis. J. Biol. Chem. *289*, 28112–28120.

Hiller, B., Achleitner, M., Glage, S., Naumann, R., Behrendt, R., and Roers, A. (2012). Mammalian RNase H2 removes ribonucleotides from DNA to maintain genome integrity. J. Exp. Med. *209*, 1419–1426.

Hogrefe, H.H., Hogrefe, R.I., Walder, R.Y., and Walder, J.A. (1990). Kinetic analysis of Escherichia coli RNase H using DNA-RNA-DNA/DNA substrates. J. Biol. Chem. *265*, 5561–5566.

Hong, X., Cadwell, G.W., and Kogoma, T. (1995). Escherichia coli RecG and RecA proteins in R-loop formation. EMBO J. *14*, 2385–2392.

Ide, H., Akamatsu, K., Kimura, Y., Michiue, K., Makino, K., Asaeda, A., Takamori, Y., and Kubo, K. (1993). Synthesis and damage specificity of a novel probe for the detection of abasic sites in DNA. Biochemistry (Mosc.) *32*, 8276–8283.

Itaya, M. (1990). Isolation and characterization of a second RNase H (RNase HII) of Escherichia coli K-12 encoded by the rnhB gene. Proc. Natl. Acad. Sci. U. S. A. *87*, 8587–8591.

Itaya, M., Omori, A., Kanaya, S., Crouch, R.J., Tanaka, T., and Kondo, K. (1999). Isolation of RNase H genes that are essential for growth of Bacillus subtilis 168. J. Bacteriol. *181*, 2118–2123.

Johansson, E., and Dixon, N. (2013). Replicative DNA polymerases. Cold Spring Harb. Perspect. Biol. *5*.

Johnson, A., and O'Donnell, M. (2005). Cellular DNA replicases: components and dynamics at the replication fork. Annu. Rev. Biochem. *74*, 283–315.

Jongruja, N., You, D.-J., Angkawidjaja, C., Kanaya, E., Koga, Y., and Kanaya, S. (2012). Structure and characterization of RNase H3 from Aquifex aeolicus. FEBS J. *279*, 2737–2753.

Joyce, C.M. (1997). Choosing the right sugar: how polymerases select a nucleotide substrate. Proc. Natl. Acad. Sci. U. S. A. *94*, 1619–1622.

Kao, H.-I., and Bambara, R.A. (2003). The protein components and mechanism of eukaryotic Okazaki fragment maturation. Crit. Rev. Biochem. Mol. Biol. *38*, 433–452.

Kasiviswanathan, R., and Copeland, W.C. (2011). Ribonucleotide discrimination and reverse transcription by the human mitochondrial DNA polymerase. J. Biol. Chem. *286*, 31490–31500.

Kim, N., Huang, S.N., Williams, J.S., Li, Y.C., Clark, A.B., Cho, J.-E., Kunkel, T.A., Pommier, Y., and Jinks-Robertson, S. (2011). Mutagenic processing of ribonucleotides in DNA by yeast topoisomerase I. Science *332*, 1561–1564.

Kirkpatrick, D.P., and Radding, C.M. (1992). RecA protein promotes rapid RNA-DNA hybridization in heterogeneous RNA mixtures. Nucleic Acids Res. *20*, 4347–4353.

Kirkpatrick, D.P., Rao, B.J., and Radding, C.M. (1992). RNA-DNA hybridization promoted by E. coli RecA protein. Nucleic Acids Res. *20*, 4339–4346.

Kochiwa, H., Tomita, M., and Kanai, A. (2007). Evolution of ribonuclease H genes in prokaryotes to avoid inheritance of redundant genes. BMC Evol. Biol. *7*, 128.

Kogoma, T. (1997). Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription. Microbiol. Mol. Biol. Rev. MMBR *61*, 212–238.

Kogoma, T., and von Meyenburg, K. (1983). The origin of replication, oriC, and the dnaA protein are dispensable in stable DNA replication (sdrA) mutants of Escherichia coli K-12. EMBO J. *2*, 463–468.

Kogoma, T., Hong, X., Cadwell, G.W., Barnard, K.G., and Asai, T. (1993). Requirement of homologous recombination functions for viability of the Escherichia coli cell that lacks RNase HI and exonuclease V activities. Biochimie *75*, 89–99.

Kornberg, A., and Baker, T. (1992). DNA Replication (W. H. Freeman and Company).

Kunkel, T.A. (2004). DNA Replication Fidelity. J. Biol. Chem. *279*, 16895–16898.

Kunkel, T.A., and Bebenek, K. (2000). DNA replication fidelity. Annu. Rev. Biochem. *69*, 497–529.

Kuong, K.J., and Kuzminov, A. (2009). Cyanide, peroxide and nitric oxide formation in solutions of hydroxyurea causes cellular toxicity and may contribute to its therapeutic potency. J. Mol. Biol. *390*, 845–862.

Lahue, R.S., Au, K.G., and Modrich, P. (1989). DNA mismatch correction in a defined system. Science *245*, 160–164.

Lang, K.S., Hall, A.N., Merrikh, C.N., Ragheb, M., Tabakh, H., Pollock, A.J., Woodward, J.J., Dreifus, J.E., and Merrikh, H. (2017). Replication-Transcription Conflicts Generate R-Loops that Orchestrate Bacterial Stress Survival and Pathogenesis. Cell *170*, 787–799.e18.

Le Chatelier, E., Bécherel, O.J., d'Alençon, E., Canceill, D., Ehrlich, S.D., Fuchs, R.P.P., and Jannière, L. (2004). Involvement of DnaE, the second replicative DNA polymerase from Bacillus subtilis, in DNA mutagenesis. J. Biol. Chem. *279*, 1757–1767.

Lenhart, J.S., Pillon, M.C., Guarné, A., Biteen, J.S., and Simmons, L.A. (2016). Mismatch repair in Gram-positive bacteria. Res. Microbiol. *167*, 4–12.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinforma. Oxf. Engl. *25*, 1754–1760.

Li, Y., and Breaker, R.R. (1999). Kinetics of RNA Degradation by Specific Base Catalysis of Transesterification Involving the 2'-Hydroxyl Group. J. Am. Chem. Soc. *121*, 5364–5372.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinforma. Oxf. Engl. *25*, 2078–2079.

Liao, Y., Schroeder, J.W., Gao, B., Simmons, L.A., and Biteen, J.S. (2015). Single-molecule motions and interactions in live cells reveal target search dynamics in mismatch repair. Proc. Natl. Acad. Sci. U. S. A. *112*, E6898-6906.

Lu, Z., Liang, R., Liu, X., Hou, J., and Liu, J. (2012a). RNase HIII from Chlamydophila pneumoniae can efficiently cleave double-stranded DNA carrying a chimeric ribonucleotide in the presence of manganese. Mol. Microbiol. *83*, 1080–1093.

Lu, Z., Hou, J., Wang, Y., and Liu, J. (2012b). Involvement of Ser94 in RNase HIII from Chlamydophila pneumoniae in the recognition of a single ribonucleotide misincorporated into double-stranded DNA. Biochim. Biophys. Acta *1824*, 859–865.

Lujan, S.A., Williams, J.S., Clausen, A.R., Clark, A.B., and Kunkel, T.A. (2013). Ribonucleotides are signals for mismatch repair of leading-strand replication errors. Mol. Cell *50*, 437–443.

McDonald, J.P., Vaisman, A., Kuban, W., Goodman, M.F., and Woodgate, R. (2012). Mechanisms employed by Escherichia coli to prevent ribonucleotide incorporation into genomic DNA by Pol V. PLoS Genet. *8*, e1003030.

Miyashita, S., Tadokoro, T., Angkawidjaja, C., You, D.-J., Koga, Y., Takano, K., and Kanaya, S. (2011). Identification of the substrate binding site in the N-terminal TBP-like domain of RNase H3. FEBS Lett. *585*, 2313–2317.

Navarra, P., and Preziosi, P. (1999). Hydroxyurea: new insights on an old drug. Crit. Rev. Oncol. Hematol. *29*, 249–255.

Neuhard, J., Price, A.R., Schack, L., and Thomassen, E. (1978). Two thymidylate synthetases in Bacillus subtilis. Proc. Natl. Acad. Sci. U. S. A. *75*, 1194–1198.

Nick McElhinny, S.A., Watts, B.E., Kumar, D., Watt, D.L., Lundström, E.-B., Burgers, P.M.J., Johansson, E., Chabes, A., and Kunkel, T.A. (2010a). Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases. Proc. Natl. Acad. Sci. U. S. A. *107*, 4949–4954.

Nick McElhinny, S.A., Kumar, D., Clark, A.B., Watt, D.L., Watts, B.E., Lundström, E.-B., Johansson, E., Chabes, A., and Kunkel, T.A. (2010b). Genome instability due to ribonucleotide incorporation into DNA. Nat. Chem. Biol. *6*, 774–781.

Ogawa, T., and Okazaki, T. (1984). Function of RNase H in DNA replication revealed by RNase H defective mutants of Escherichia coli. Mol. Gen. Genet. MGG *193*, 231–237.

Ohtani, N., Haruki, M., Morikawa, M., and Kanaya, S. (1999a). Molecular diversities of RNases H. J. Biosci. Bioeng. *88*, 12–19.

Ohtani, N., Haruki, M., Morikawa, M., Crouch, R.J., Itaya, M., and Kanaya, S. (1999b). Identification of the genes encoding Mn2+-dependent RNase HII and Mg2+-dependent RNase HIII from Bacillus subtilis: classification of RNases H into three families. Biochemistry (Mosc.) *38*, 605–618.

Ohtani, N., Haruki, M., Muroya, A., Morikawa, M., and Kanaya, S. (2000). Characterization of ribonuclease HII from Escherichia coli overproduced in a soluble form. J. Biochem. (Tokyo) *127*, 895–899.

Oivanen, M., Kuusela, S., and Lönnberg, H. (1998). Kinetics and Mechanisms for the Cleavage and Isomerization of the Phosphodiester Bonds of RNA by Brønsted Acids and Bases. Chem. Rev. *98*, 961–990.

Ordonez, H., Uson, M.L., and Shuman, S. (2014). Characterization of three mycobacterial DinB (DNA polymerase IV) paralogs highlights DinB2 as naturally adept at ribonucleotide incorporation. Nucleic Acids Res. *42*, 11056–11070.

Patel, P.H., and Loeb, L.A. (2000). Multiple amino acid substitutions allow DNA polymerases to synthesize RNA. J. Biol. Chem. *275*, 40266–40272.

Paul, S., Million-Weaver, S., Chattopadhyay, S., Sokurenko, E., and Merrikh, H. (2013). Accelerated gene evolution through replication-transcription conflicts. Nature *495*, 512–515.

Petruska, J., and Goodman, M.F. (1985). Influence of neighboring bases on DNA polymerase insertion and proofreading fidelity. J. Biol. Chem. *260*, 7533–7539.

Pillon, M.C., Babu, V.M.P., Randall, J.R., Cai, J., Simmons, L.A., Sutton, M.D., and Guarné, A. (2015). The sliding clamp tethers the endonuclease domain of MutL to DNA. Nucleic Acids Res.

Pizzi, S., Sertic, S., Orcesi, S., Cereda, C., Bianchi, M., Jackson, A.P., Lazzaro, F., Plevani, P., and Muzi-Falconi, M. (2015). Reduction of hRNase H2 activity in Aicardi-Goutières syndrome cells leads to replication stress and genome instability. Hum. Mol. Genet. *24*, 649–658.

Pluciennik, A., Dzantiev, L., Iyer, R.R., Constantin, N., Kadyrov, F.A., and Modrich, P. (2010). PCNA function in the activation and strand direction of MutLα endonuclease in mismatch repair. Proc. Natl. Acad. Sci. U. S. A. *107*, 16066–16071.

Potenski, C.J., Niu, H., Sung, P., and Klein, H.L. (2014). Avoidance of ribonucleotide-induced mutations by RNase H2 and Srs2-Exo1 mechanisms. Nature *511*, 251–254.

Pursell, Z.F., Isoz, I., Lundström, E.-B., Johansson, E., and Kunkel, T.A. (2007). Yeast DNA polymerase epsilon participates in leading-strand DNA replication. Science *317*, 127–130.

Que, Q., and Helmann, J.D. (2000). Manganese homeostasis in Bacillus subtilis is regulated by MntR, a bifunctional regulator related to the diphtheria toxin repressor family of proteins. Mol. Microbiol. *35*, 1454–1468.

Randall, J.R., Hirst, W.G., and Simmons, L.A. (2017). Substrate specificity for bacterial RNase HII and HIII is influenced by metal availability. J. Bacteriol.

Reijns, M.A.M., and Jackson, A.P. (2014). Ribonuclease H2 in health and disease. Biochem. Soc. Trans. *42*, 717–725.

Reijns, M.A.M., Rabe, B., Rigby, R.E., Mill, P., Astell, K.R., Lettice, L.A., Boyle, S., Leitch, A., Keighren, M., Kilanowski, F., et al. (2012). Enzymatic removal of ribonucleotides from DNA is essential for mammalian genome integrity and development. Cell *149*, 1008–1022.

Roberts, R.W., and Crothers, D.M. (1992). Stability and properties of double and triple helices: dramatic effects of RNA or DNA backbone composition. Science *258*, 1463–1466.

Rohland, N., and Reich, D. (2012). Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. Genome Res. *22*, 939–946.

Rowen, L., and Kornberg, A. (1978a). A ribo-deoxyribonucleotide primer synthesized by primase. J. Biol. Chem. *253*, 770–774.

Rowen, L., and Kornberg, A. (1978b). Primase, the dnaG protein of Escherichia coli. An enzyme which starts DNA chains. J. Biol. Chem. *253*, 758–764.

Roy, D., and Lieber, M.R. (2009). G clustering is important for the initiation of transcription-induced R-loops in vitro, whereas high G density without clustering is sufficient thereafter. Mol. Cell. Biol. *29*, 3124–3133.

Roy, D., Zhang, Z., Lu, Z., Hsieh, C.-L., and Lieber, M.R. (2010). Competition between the RNA transcript and the nontemplate DNA strand during R-loop formation in vitro: a nick can serve as a strong R-loop initiation site. Mol. Cell. Biol. *30*, 146–159.

Rychlik, M.P., Chon, H., Cerritelli, S.M., Klimek, P., Crouch, R.J., and Nowotny, M. (2010). Crystal structures of RNase H2 in complex with nucleic acid reveal the mechanism of RNA-DNA junction recognition and cleavage. Mol. Cell *40*, 658–670.

Sale, J.E., Lehmann, A.R., and Woodgate, R. (2012). Y-family DNA polymerases and their role in tolerance of cellular DNA damage. Nat. Rev. Mol. Cell Biol. *13*, 141–152.

Sanders, G.M., Dallmann, H.G., and McHenry, C.S. (2010). Reconstitution of the B. subtilis replisome with 13 proteins including two distinct replicases. Mol. Cell *37*, 273–281.

Sanjanwala, B., and Ganesan, A.T. (1991). Genetic structure and domains of DNA polymerase III of Bacillus subtilis. Mol. Gen. Genet. MGG *226*, 467–472.

Santos-Pereira, J.M., and Aguilera, A. (2015). R loops: new modulators of genome dynamics and function. Nat. Rev. Genet. *16*, 583–597.

Schroeder, J.W., Randall, J.R., Matthews, L.A., and Simmons, L.A. (2014). Ribonucleotides in bacterial DNA. Crit. Rev. Biochem. Mol. Biol. 1–13.

Schroeder, J.W., Hirst, W.G., Szewczyk, G.A., and Simmons, L.A. (2016). The Effect of Local Sequence Context on Mutational Bias of Genes Encoded on the Leading and Lagging Strands. Curr. Biol. CB *26*, 692–697.

Schroeder, J.W., Randall, J.R., Hirst, W.G., O'Donnell, M.E., and Simmons, L.A. (2017). Mutagenic cost of ribonucleotides in bacterial DNA. Proc. Natl. Acad. Sci. U. S. A. *114*, 11733–11738.

Sekiguchi, J., and Shuman, S. (1997). Site-specific ribonuclease activity of eukaryotic DNA topoisomerase I. Mol. Cell *1*, 89–97.

Setlow, P., Brutlag, D., and Kornberg, A. (1972). Deoxyribonucleic acid polymerase: two distinct enzymes in one polypeptide. I. A proteolytic fragment containing the polymerase and 3' leads to 5' exonuclease functions. J. Biol. Chem. *247*, 224–231.

Shaw, N.N., and Arya, D.P. (2008). Recognition of the unique structure of DNA:RNA hybrids. Biochimie *90*, 1026–1039.

Sollier, J., and Cimprich, K.A. (2015). Breaking bad: R-loops and genome integrity. Trends Cell Biol. *25*, 514–522.

Sparks, J.L., Chon, H., Cerritelli, S.M., Kunkel, T.A., Johansson, E., Crouch, R.J., and Burgers, P.M. (2012). RNase H2-initiated ribonucleotide excision repair. Mol. Cell *47*, 980–986.

Sung, H.-M., Yeamans, G., Ross, C.A., and Yasbin, R.E. (2003). Roles of YqjH and YqjW, homologs of the Escherichia coli UmuC/DinB or Y superfamily of DNA polymerases, in stationary-phase mutagenesis and UV-induced mutagenesis of Bacillus subtilis. J. Bacteriol. *185*, 2153–2160.

Sung, W., Ackerman, M.S., Gout, J.-F., Miller, S.F., Williams, E., Foster, P.L., and Lynch, M. (2015). Asymmetric Context-Dependent Mutation Patterns Revealed through Mutation-Accumulation Experiments. Mol. Biol. Evol. *32*, 1672–1683.

Tabor, S., and Richardson, C.C. (1989). Effect of manganese ions on the incorporation of dideoxynucleotides by bacteriophage T7 DNA polymerase and Escherichia coli DNA polymerase I. Proc. Natl. Acad. Sci. U. S. A. *86*, 4076–4080.

Tadokoro, T., and Kanaya, S. (2009). Ribonuclease H: molecular diversities, substrate binding domains, and catalytic mechanism of the prokaryotic enzymes. FEBS J. *276*, 1482–1493.

Tanaka, N., and Shuman, S. (2011). RtcB is the RNA ligase component of an Escherichia coli RNA repair operon. J. Biol. Chem. *286*, 7727–7731.

Tannous, E., Kanaya, E., and Kanaya, S. (2015). Role of RNase H1 in DNA repair: removal of single ribonucleotide misincorporated into DNA in collaboration with RNase H2. Sci. Rep. *5*, 9969.

Thompson, J.E., Kutateladze, T.G., Schuster, M.C., Venegas, F.D., Messmore, J.M., and Raines, R.T. (1995). Limits to Catalysis by Ribonuclease A. Bioorganic Chem. *23*, 471–481.

Torrents, E. (2014). Ribonucleotide reductases: essential enzymes for bacterial life. Front. Cell. Infect. Microbiol. *4*, 52.

Traut, T.W. (1994). Physiological concentrations of purines and pyrimidines. Mol. Cell. Biochem. *140*, 1–22.

Vaisman, A., McDonald, J.P., Huston, D., Kuban, W., Liu, L., Van Houten, B., and Woodgate, R. (2013). Removal of misincorporated ribonucleotides from prokaryotic genomes: an unexpected role for nucleotide excision repair. PLoS Genet. *9*, e1003878.

Vaisman, A., McDonald, J.P., Noll, S., Huston, D., Loeb, G., Goodman, M.F., and Woodgate, R. (2014). Investigating the mechanisms of ribonucleotide excision repair in Escherichia coli. Mutat. Res. *761*, 21–33.

Van de Sande, J.H., Loewen, P.C., and Khorana, H.G. (1972). Studies on polynucleotides. 118. A further study of ribonucleotide incorporation into deoxyribonucleic acid chains by deoxyribonucleic acid polymerase I of Escherichia coli. J. Biol. Chem. *247*, 6140–6148.

Williams, J.S., and Kunkel, T.A. (2014). Ribonucleotides in DNA: origins, repair and consequences. DNA Repair *19*, 27–37.

Williams, J.S., Smith, D.J., Marjavaara, L., Lujan, S.A., Chabes, A., and Kunkel, T.A. (2013). Topoisomerase 1-mediated removal of ribonucleotides from nascent leading-strand DNA. Mol. Cell *49*, 1010–1015.

Yang, G., Franklin, M., Li, J., Lin, T.-C., and Konigsberg, W. (2002). A conserved Tyr residue is required for sugar selectivity in a Pol alpha DNA polymerase. Biochemistry (Mosc.) *41*, 10256–10261.

Yao, N.Y., Schroeder, J.W., Yurieva, O., Simmons, L.A., and O'Donnell, M.E. (2013). Cost of rNTP/dNTP pool imbalance at the replication fork. Proc. Natl. Acad. Sci. U. S. A. *110*, 12942–12947.

Youngman, P., Perkins, J.B., and Losick, R. (1984). Construction of a cloning site near one end of Tn917 into which foreign DNA may be inserted without affecting transposition in Bacillus subtilis or expression of the transposon-borne erm gene. Plasmid *12*, 1–9.