REPAIR, CONSEQUENCE, AND PROFILE OF RIBONUCLEOTIDES IN DNA

A Dissertation
Presented to
The Academic Faculty

By

Kyung Duk Koh

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Biology

Georgia Institute of Technology

May 2015

Copyright © 2015 by Kyung Duk Koh

REPAIR, CONSEQUENCE, AND PROFILE OF RIBONUCLEOTIDES IN DNA

Approved by:

Dr. Francesca Storici, Advisor
School of Biology
Georgia Institute of Technology

Dr. Nicholas V. Hud
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Yury O. Chernoff
School of Biology
Georgia Institute of Technology

Dr. Kirill S. Lobachev
School of Biology
Georgia Institute of Technology

Dr. Brian K. Hammer
School of Biology
Georgia Institute of Technology

Date Approved: March 9, 2015

For my family

ACKNOWLEDGEMENTS

I wish to thank Dr. Francesca Storici for giving me an opportunity to join and work in her laboratory and providing me with all the guidance throughout my graduate study to become a better researcher and person.

I want to show my appreciation to the rest of my thesis committee, Dr. Yury Chernoff, Dr. Brian Hammer, Dr. Nicholas Hud, and Dr. Kirill Lobachev. They have been always available and open to answer any of my questions and offer insightful, helpful, and constructive feedback.

All former and current members of the Storici laboratory have supported me throughout my graduate study, and I am very grateful. From the moment I initially joined the lab, Dr. Ying Shen, Dr. Patrick Ruff, Dr. Samantha Stuckey, Dr. Kuntal Mukherjee, and Dr. Rekha Pai all were willing to give assistance, guidance, and suggestions. I also thank Havva Keskin, Sathya Balachander, and Chance Meers, who joined the lab after myself. Without them, I would not have been able to complete my Ph.D. study. I had a few opportunities to mentor undergraduate students, Alli Gombolay, Courtney Price, Lahari Shetty, and Anna Sulimirski. I appreciate their passion for research and their patience for working with me. I also thank all former lab members, including Gayathri Pratap Kurup and Taehwan Yang.

Outside of the lab, there are many people I have been fortunate enough to work and interact with. I express my sincere gratitude to all collaborators of my research. Working with Dr. Bernard Weiss at Emory University; Dr. Hsiang-Chih Chiu and Dr. Elisa Riedo in School of Physics; Annie Lesiak and Dr. Angelo Bongiorno in School of Chemistry and Biochemistry; Marina Evich and Dr. Markus Germann at Georgia State University; and Dr. Jay Hesselberth at University of Colorado and discussing science together were very enjoyable moments. I am very grateful to Dr. Natasha Degtyareva at Emory University for recommending me to Georgia Tech for admission to the Ph.D. program and supporting me throughout my graduate study. I also thank Dr. Eric Gaucher, Dr. King Jordan, Dr. Frank Stewart, and Dr. Roger Wartell in School of Biology and David Bostwick, Dr. Cameron Sullards, and Dr. Loren Williams in School of Chemistry and Biochemistry for technical support and advice. All the staff members in School of Biology, including Frank Cannella, Kendal Carey, Troy Hilley, Verene Lancasater, Jasmine Martin, Marc Pline, Kevin Roman, and Angie Wiltrout were extremely helpful. I also had opportunities to work as teaching assistant for Dr. Mirjana Brockett, Dr. Jung Choi, Dr. Linda Green, Dr. Jennifer Leavey, Dr. Kirill Lobachev, Dr. John McDonald, Dr. Chrissy Spencer, and Dr. Francesca Storici, and I appreciate all the advice.

I also would like to thank all my friends for all stress-relieving hang-outs.

Lastly, I thank my family for all the support, allowing me to take a chance to work towards a Ph.D. degree. Without my parents and my sister, I would not have made it this far.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

AFM	Atomic force microscopy
AGS	Aicardi-Goutieres syndrome
AP	Apurinic/apyrimidinic
atm	Atmosphere
AtRNL	*Arabidopsis thaliana* tRNA ligase
BER	Base excision repair
bp	base-pair
CD	Circular dichroism
cm	Centimeters
dAMP	Deoxyadenosine monophosphate
dCMP	Deoxycytidine monophosphate
dGMP	Deoxyguanosine monophosphate
DNA	Deoxyribonucleic acid
dNMP	Deoxyribonucleoside monophosphate
dNTP	Deoxyribonucleoside triphosphate
ds	Double-strand
DSB	Double-strand break
dTMP	Deoxythymidine monosphosphate
dUMP	Deoxyuridine monophosphate
fmol	Femtomole
h	hours
HU	Hydroxyurea
ins/del	Insertions and deletions

K	Kelvin
kb	Kilobases
m	Meters
M	Molar
MD	Molecular dynamics
mg	Milligrams
min	Minutes
mL	Milliliters
mM	Millimolar
MMR	Mismatch repair
N	Newtons
NER	Nucleotide excision repair
ng	Nanograms
nm	Nanometers
nM	Nanomolar
nmol	Nanomole
NMR	Nuclear magnetic resonance
nN	Nanonewtons
nt	Nucleotide
OH	Hydroxyl
oligo	Oligonucleotide
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
Pfu	*Pyrococcus furiosus*
pN	Piconewtons

Pol	Polymerase
ppm	Parts per minute
PRR	Postreplication repair
rAMP	Adenosine monophosphate
rCMP	Cytidine monophosphate
rDNA	Ribosomal DNA
RER	Ribonucleotide excision repair
rGMP	Guanosine monophosphate
RNA	Ribonucleic acid
RNase H	Ribonuclease H
rNMP	Ribonucleoside monophosphate
rNTP	Ribonucleoside triphosphate
ROS	Reactive oxygen species
rRNA	Ribosomal RNA
rUMP	Uridine monophosphate
s	Seconds
S	Stretch modulus
ss	Single-strand
Taq	*Thermus aquaticus*
Ty	Yeast transposable element
UMI	Unique molecular identifier
UV	Ultraviolet
w/v	Weight-to-volume
α	Alpha
Å	Angstroms

β	Beta
°	Degrees
℃	Celsius degrees
Δ	Deletion
δ	Delta
ε	Epthalon
γ	Gamma
μL	Microliters
μm	Micrometers
μM	Micromolar

SUMMARY

Ribonucleotides, also known as ribonucleoside monophosphates (rNMPs), are the most abundant non-canonical nucleotides incorporated into genomic DNA. Despite the relevance, information about their repair pathways, consequences, and profiles is still lacking. Exploiting the use of oligonucleotides containing rNMPs in a molecular approach to generate various RNA/DNA hybrids of chosen sequence and structure at the chromosomal level in cells, we show that mispaired rNMPs embedded into genomic DNA are not only targeted by ribonucleases H (RNases H) but also by the mismatch repair (MMR) system both in *E. coli* and *S. cerevisiae* cells. In addition, we discovered that paired rNMPs in DNA are targets of both RNase H type 2 and nucleotide excision repair (NER) in yeast. Also, we report atomic force microscopy (AFM)-based single molecule elasticity measurement, molecular dynamics simulation, and nuclear magnetic resonance spectroscopy results, showing that rNMPs in short DNA duplexes can change the elastic and structural properties of DNA. Lastly, we developed ribose-seq, a method for capturing rNMPs embedded in DNA. High-throughput sequencing of rNMP-captured molecules from the yeast *S. cerevisiae* revealed widespread but non-random rNMP distribution with preferences in base composition of rNMPs and neighboring DNA sequence context in both nuclear and mitochondrial DNA. With ribose-seq, systematic profiling of rNMP incorporation into genomic DNA is achieved, potentially allowing determination of specific signatures of rNMPs in DNA which could help to better understand the nature of rNMP repair mechanisms, effect of rNMPs on DNA mechanical properties and structure, and eventually rNMP impact on genome integrity.

CHAPTER 1

INTRODUCTION

1.1 Genome Integrity

Genome integrity is persistently tested by various exogenous agents, such as chemicals and UV radiation, and endogenous sources, including DNA replication errors, DNA single- and double-strand breaks from collapsed DNA replication forks, dysfunctional DNA repair pathways, and reactive oxygen species [1, 2]. Numerous types of mutations, such as substitutions, insertions, deletions, and chromosomal rearrangements, can result from accumulation of DNA damage. Genome instability is generally thought to be detrimental to cells, but it is also the driving force of evolution leading to genetic variation at the molecular level. Prevention of genome instability and maintenance of genome integrity are achieved by several DNA repair pathways, including direct reversal of damage and post-replicative processes of base-excision repair (BER), mismatch repair (MMR), nucleotide excision repair (NER), and double-strand break (DSB) repair [1]. In addition to well-known types of DNA damage, such as base alkylation, oxidation, and hydrolysis, ribonucleotides comprise a significant portion of DNA lesions.

1.2 Incorporation of Ribonucleotides into DNA

Ribonucleotides, also known as ribonucleoside 5′-monophosphates (rNMPs), which are normally monomers of RNA, have been found to be the most abundant non-canonical nucleotides incorporated in DNA. rNMPs have an extra 2′-hydroxyl (OH) group, differing from their corresponding deoxyribonucleotides, also known as deoxyribonucleoside 5′-monophosphates (dNMPs). Incorporation of rNMPs in DNA was initially found only in particular DNA sequences, including the mammalian

mitochondrial genome [3] and the mating type locus of fission yeast [4]. Despite the presence of conserved "steric gate" against 2′-OH of ribonucleotide, which partly prevents the incorporation of rNMP by a DNA polymerase (Pol), recent evidence shows the capacity of nearly all DNA polymerases from bacteria to yeast and human to incorporate rNMPs into DNA [5-23]. For example, *Escherichia coli* polymerases III [19] and V [12], the polymerase component of bacterial non-homologous end joining ligases [13], and the human replicative polymerases δ [14] and ε [15] all insert rNMPs. In addition, human polymerase λ can insert rNMPs with the same efficiency as deoxyribonucleotides, also known as deoxyribonucleoside 5′-monophosphates (dNMPs) [16]. All replicative polymerases of budding yeast (Pol α, δ, and ε) incorporate rNMPs into DNA with frequencies of 1 rNMP for every 625, 5,000, and 1,250 dNMPs, respectively [21]. The fact that the amount of ribonucleoside 5′-triphosphates (rNTPs) is generally 10 to 188-fold higher than that of deoxyribonucleoside 5′-triphosphates (dNTPs) in cycling cells [21, 24, 25] increases the probability of rNMP incorporation during DNA replication and repair.

Other sources of ribonucleotides in DNA include incomplete Okazaki fragment maturation during DNA replication and oxidative damage. Initial steps of DNA polymerization involve an RNA primer synthesized by an RNA primase [26]. This primer is used to initiate DNA synthesis of Okazaki fragments during lagging-strand synthesis [27]. The RNA fragment is then removed via Okazaki fragment maturation, but incomplete maturation may lead rNMPs to be remained and embedded in DNA by DNA ligase I [28]. Oxidative damage has also been shown to form rNMPs *in vitro* and *in vivo* [29]. Reactive oxygen species (ROS), among which the hydroxyl radical (·OH) is the most reactive, can cause multiple modifications to DNA [30]. Via Fenton chemistry, the

·OH radical can attach all components of DNA, including the deoxyribose that can give rise to ribose [29].

Quantitative measures of rNMPs in DNA following alkali treatment of yeast genomic DNA derived from RNase H2–deficient cells estimated ~1,500 to ~2,400 rNMPs per genome [31-33]. Similar measurement conducted for genomic DNA derived from RNase H2–deficient embryonic fibroblasts revealed the presence of more than one million rNMPs in the mouse genome, suggesting that rNMPs are the most common nucleotide base lesion in dividing mouse cells [34]. Overall, these findings show that rNMPs occur in DNA much more often than previously anticipated and much more frequently than mismatches or oxidized bases; thus, rNMPs are the most common non-canonical nucleotides found in genome of cells.

Information about the identity and the distribution of rNMPs in genomic DNA is still lacking and remains to be determined.

1.3 Repair and Tolerance of Ribonucleotides in DNA

As previously described, the number of rNMPs incorporated in DNA exceeds the number of other DNA lesions. Therefore, one would expect multiple pathways to exist to process rNMPs in DNA.

1.3.1 Ribonucleotide excision repair

Ribonucleotide excision repair (RER) is generally known as the primary mechanism of removal of rNMPs in DNA [35]. RER is initiated by ribonucleases (RNases) H, which are enzymes that have been shown *in vitro* to degrade the RNA strand of an RNA:DNA hybrid [36-38]. RNases H have been reported to be involved in DNA replication, transcription, recombination repair, and development. RNases H can be classified into

two major categories: type I/1 and type II/2. RNase HI/1 enzymes require a stretch of four or more rNMPs embedded in a DNA duplex while RNase HII/2 enzymes are able cleave even single rNMPs embedded in DNA [37, 38]. RER, for example in yeast, is initiated by the incision of the rNMP(s) by either RNase H, and the nicked DNA strand is further excised by the flap endonuclease FEN1 or Exo1 with strand displacement synthesis carried out by DNA Pol δ, the PCNA clamp, its loader RFC, and DNA ligase I (**Figure 1.1**). As mentioned earlier, all quantitative studies of rNMPs in DNA have been performed with genomic DNA from RNase H2–deficient cells, suggesting the significant role of RNase H2 in removal of rNMPs in DNA.

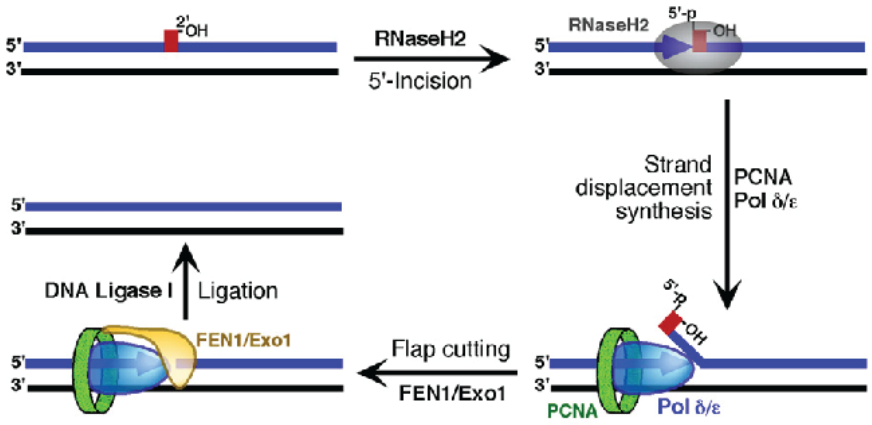


Figure 1.1 Model for ribonucleotide excision repair. rNMP is in red. RNase H2 is shown here as the substrate is a single rNMP. (from [35])

1.3.2 Topoisomerase I–mediated removal of ribonucleotides

In the absence of functional RNase H2, the accumulation of genomic rNMPs is partially suppressed by the action of DNA topoisomerase I (Top1) [32, 39, 40]. Top1 is a type IB topoisomerase whose action proceeds through the reversible formation of a DNA nick with a DNA-3′-phosphate-Top1 covalent complex and a 5′-OH. When Top1 incises at

rNMPs, it is transiently linked to 3′-phosphate of the rNMP. With the nucleophilic attack by the neighboring 2′-OH group, Top1 is released, resulting in a nick containing unligatable 5′-OH and 2′,3′-cyclic phosphate ends [41]. This 2′,3′-cyclic phosphate end-containing DNA is then unwound by a DNA helicase Srs2, which interacts with and enhances the activity of nuclease Exo1, allowing nick processing-gap filling [40]. Quantitative measurement of rNMPs removed by Top1 was only possible in the absence of RNase H2, suggesting that RNase H2 is the first repair enzyme to remove rNMPs in DNA.

1.3.3 Other removal and tolerance pathways of ribonucleotides in DNA

DNA polymerases do have the ability to proofread rNMPs but is very limited [14, 42]. For example, yeast replicative Pol ε has been shown to proofread rAMPs and rUMPs, but not rGMPs and rCMPs [42].

Incision at an rNMP embedded in DNA by RNase H2 leaves a 3′-OH end and a 5′-phosphate end, which allows the possibility of DNA ligase to attempt to re-ligate the nick. The first step of re-ligation would involve the adenylation of the rNMP–DNA junction. This adenylation is reversed by aprataxin, which restores the normal RER [43].

rNMPs in DNA, if unremoved, need to be tolerated by the cells. During the subsequent round of replication, DNA polymerases need to be able to bypass rNMPs in DNA. rNMPs in DNA template have been found to impede DNA synthesis by the yeast replicases ε and δ [44-46], with the former having much lower bypass efficiency when there was a stretch two to four rNMPs. Therefore, to properly complete replication, there must be some other pathways for tolerance. One known mechanism for tolerance is via postreplication repair (PRR). Specifically, *MMS2*-dependent template switch and Pol ζ-

dependent bypass have been found to be important for tolerating the presence of rNMPs in DNA [47].

Whether rNMPs are targeted by other repair pathways remains to be further investigated.

1.4 Consequences of rNMPs in DNA

Incorporation of rNMPs into DNA has several consequences, including both positive and negative consequences.

1.4.1 Positive consequences

There are so far two known positive consequences of rNMPs in DNA. One is observed in fission yeast *Schizosaccharomyces pombe*. The presence of two rNMPs in the *mat1* locus stalls leading-strand synthesis by Pol ε during subsequent DNA replication, initiating the replication-coupled recombination event leading to mating-type switching between M and P [4, 48].

Another positive consequence of rNMPs in DNA involves mismatch repair (MMR). MMR is responsible for correcting any kind of DNA/DNA misalignments, including base-base mismatches and insertion/deletion loops [49-51]. In order for proper repair to occur, MMR must target the nascent strand. In *E. coli*, the transiently unmethylated state of adenine in GATC sequences during DNA replication acts as the strand discrimination signal. In eukaryotes, transient 5′ ends of Okazaki fragments may allow strand discrimination during lagging-strand synthesis [52]. During leading-strand synthesis, nicks created by RNase H2 at rNMPs in DNA act as strand discrimination signals [33, 53].

1.4.2 Negative consequences

With the highly reactive extra 2′-OH group in the sugar, rNMPs embedded in DNA constitute an internal threat to DNA [54]. Despite the few instances mentioned above, overall, the presence of rNMPs in DNA is deleterious. DNA contaminated with ribose is markedly more susceptible to breakage [54]. Short rNMP tracts embedded in DNA can serve as templates for DNA synthesis and can propagate a mutation upon replication in *E. coli*, budding yeast, and human embryonic kidney cells [46, 55-58]. As mentioned earlier, while yeast replicative polymerases can incorporate dNMPs opposite to rNMPs in template, their efficiencies are lower than to DNA [44, 45]. Thus, the rNMPs embedded in DNA and/or the DNA breaks that arise from processing of rNMPs can pause DNA replication or transcription and cause nucleotide deletions during DNA synthesis, threatening genome integrity. Moreover, nucleosome assembly onto DNA duplexes containing even a single rNMP within 125 base-pairs (bp) of DNA occurs with reduced efficiency [59].

RNase H2–null *Saccharomyces cerevisiae* strains with an additional Pol ε mutant that increases rNMP incorporation show several phenotypes of replication stress, including slow growth, accumulation in S phase, and sensitivity to hydroxyurea (HU), a replication inhibitor [47, 60]. Also, increased genome instability is observed with an increase in 2–5 bp deletions in short tandem repeats [60, 61]. Pol ε mutation which decreases rNMP incorporation eliminates these effects, suggesting that the stress is dependent on the level of rNMP incorporation. Also, additional knockout of *top1* lessens these effects, suggesting that they mainly result from ends created by Top1 cleavage at rNMPs [32, 39, 62]. Deletion of either RNase H1 or H2 in mice is embryonic lethal [34, 63]. Murine RNase H2–null embryonic fibroblasts accumulate over one million rNMPs in their genomic DNA, activating a p53–dependent damage response [34]. In humans, partial-loss-of-function mutations in any of three subunits of RNase H2 are associated with the neurological syndrome of Aicardi-Goutieres (AGS) [64, 65]. Altered RNase H2 function

in AGS patients may lead to accumulation of rNMPs in DNA, which in turn is thought to induce a chronic, low-level DNA damage response signaling that stimulates innate immune pathways [66].

1.4.3 Consequences that could go both ways

The structural effects of rNMP(s) in DNA have been investigated by nuclear magnetic resonance (NMR), X-ray crystallography, and molecular dynamics (MD) simulation [67-75]. Most of these studies are analyses of DNA duplexes with a stretch of rNMPs to investigate the structural influence of RNA, which is present transiently in DNA during DNA replication as an Okazaki fragment. A few studies have investigated the structural effects of isolated rNMPs in DNA [72-75]. Crystallography studies so far indicate that the rNMP(s)-embedded DNA molecules adopt an overall A-DNA conformation, which is observed in RNA. NMR and MD simulation studies of DNA containing rNMPs in solution indicate that the duplex retains its overall B-conformation with local distortions observed in the vicinity of the isolated rNMPs. Most of these studies used self-complementary DNA sequences; for example, NMR study by DeRose *et al*. [75] used a self-complementary Dickerson dodecamer sequence, having an rNMP in each strand of DNA. Helical distortions are caused by rNMP(s) in DNA can not only be helpful but also harmful to the cells. These helical distortions can easily affect DNA-DNA and DNA-protein interactions, playing a role in DNA recognition, packaging, and modification by proteins.

Whether the local helical distortions caused by rNMP(s) in DNA affect the mechanical properties of DNA remains to be determined.

1.5 Research Goals

1.5.1 To identify the protein factors and DNA repair mechanisms that can target ribonucleotides embedded in DNA.

Given the frequency of rNMP incorporation into DNA, we propose that there are multiple repair pathways targeting rNMPs in genomic DNA, such as RNases H, mismatch repair (MMR), nucleotide excision repair (NER), and base excision repair (BER). In addition to *in vitro* cleavage assays, we have utilized *in vivo* oligonucleotide (oligo)-driven gene correction assays in *E. coli* and *S. cerevisiae* to examine gene correction efficiency in different mutant backgrounds.

1.5.2 To determine whether ribonucleotides embedded in DNA affect its mechanical and structural properties.

With the high level of rNMP incorporation into DNA and its consequences in cells, we propose that rNMPs in DNA change its elasticity and structure. We made atomic force microscopy (AFM)-based single molecule force measurements to determine the elasticity and performed molecular dynamics (MD) simulations and nuclear magnetic resonance (NMR) to study the structure of rNMP(s)-containing DNA.

1.5.3 To profile ribonucleotides incorporated into genomic DNA.

Numerous studies have shown the presence of rNMPs in DNA; however, the information about the identity and the distribution of these rNMPs in genomic DNA is unknown. We propose to develop an approach to capture rNMPs in *S. cerevisiae* genomic DNA so that, via next-generation sequencing, profiles of rNMPs in DNA could be obtained.

CHAPTER 2

RIBONUCLEOTIDES IN DNA ARE TARGETS OF RNASES H, MISMATCH REPAIR, AND NUCLEOTIDE EXCISION REPAIR.

The study in Chapter 2 consists of the work published in *Nat. Struct. Mol. Biol.* 19 (2011) 98–104.

Shen, Y.[1], Koh, K.D.[1], Weiss, B.[2], and Storici F.[1]

[1]School of Biology, Georgia Institute of Technology, Atlanta, GA 30332

[2]Department of Pathology and Laboratory Medicine, Emory University School of Medicine, Atlanta, GA 30322

and the work in preparation for publication.

Koh, K.D., Balachander, S., Gombolay, A., Shetty, L., and Storici, F.

School of Biology, Georgia Institute of Technology, Atlanta, GA 30332

2.1 Abstract

Numerous studies have shown that ribonucleoside monophosphates (rNMPs) are the most abundant among all non-standard nucleotides occurring in genomic DNA. Therefore, it is important to understand to what extent rNMPs may alter genome integrity and what factors affect their stability. We developed oligonucleotide-driven gene correction assays in the bacterium *Escherichia coli* and the yeast *Saccharomyces cerevisiae* to show that mispaired and paired rNMPs embedded into genomic DNA, if not removed, serve as templates for DNA synthesis and produce a genetic change. We discovered that isolated mispaired rNMPs in chromosomal DNA are removed by the mismatch repair system in competition with RNase H type 2. However, a mismatch within an RNA-DNA heteroduplex region requires RNase H type 1 for removal. In the absence of mismatch repair and RNases H, ribonucleotide-driven gene modification increased a factor of 47 in yeast and 77,000 in *E. coli*. Also, we found that isolated paired rNMPs in DNA are removed by RNase H type 2 and nucleotide excision repair.

2.2 Introduction

Modifications of nucleotides in DNA pose a threat to the genomic integrity of cells, often resulting in cell death or mutation. Numerous studies suggest that ribonucleoside monophosphates (rNMPs) are abundant among all non-standard nucleotides occurring in genomic DNA. Many DNA polymerases have been shown *in vitro* to incorporate rNMPs into DNA [5-8, 13, 17, 18], including all yeast replicative polymerases [21]. The abundance of rNMPs in DNA raises the possibility of multiple repair pathways to target and remove rNMPs embedded in DNA. Importantly, the combination of ribonucleotide misincorporation and mispairing *in vivo* can occur with sufficient frequency to affect mutation rates in yeast [60, 61]. The finding that rNMP incorporation during replication in yeast is coupled with genome instability in an RNase H2–defective background [60] raises the question whether mismatches generated by mispaired rNMPs or processing of

rNMPs in DNA are subject to mismatch repair (MMR) and whether paired rNMPs are targeted by other common DNA repair mechanisms, such as nucleotide excision repair (NER) and base excision repair (BER).

The concentration of ribonucleoside triphosphates in cells is generally higher than that of deoxyribonucleoside triphosphates, increasing the probability of paired and mispaired rNMP incorporation by DNA polymerases during DNA replication and repair ([21] and references therein). Regions of highly transcribed DNA were found to contain a high concentration of dUMP in yeast [76], and it would be interesting to see if these regions may also be hot spots for rNMP incorporation. In addition, DNA primases have very low fidelity, resulting in the incorporation of mispaired rNMPs, and could potentially leave one or several rNMPs embedded in chromosomal DNA by including deoxyribonucleotides in addition to ribonucleotides in the primer sequence [26, 77]. Finally, oxidative damage of DNA can convert a deoxynucleotide in DNA to an rNMP [29].

Ribonucleotides in DNA can distort the double helix [73], resulting in genomic instability, defective replication [78], or transcription and mutagenesis. Topoisomerase I can initiate a process of rNMP removal when its site of cleavage occurs 3′ to an rNMP embedded in DNA [39]. However, the enzymes that specifically cleave RNA in RNA/DNA hybrids are ribonucleases H (RNase H type 1 (I or 1) and type 2 (II or 2)) [37, 38]. *In vitro,* RNases HI or H1 and RNases HII or H2 have very distinct cleavage patterns. RNases HI or H1 require a substrate with an RNA stretch containing at least four ribonucleotides in a DNA duplex to allow cleavage, while RNases HII or H2 can cleave a single ribonucleotide embedded in DNA even when mispaired [37, 79]. Yet, up to now a detailed analysis of the *in vivo* substrate specificity of these enzymes is missing [37]. Here we utilized an approach to generate *in vivo* defined rNMP-containing DNA

duplexes and reveal the substrate preference of RNase H enzymes, and we investigated whether RNA:DNA mismatches can be recognized by the MMR system in *E. coli* and in yeast cells and whether paired rNMPs can targeted by NER and/or BER in yeast cells.

2.3 Materials and Methods

2.3.1 Bacterial strains

All *E. coli* strains used here (**Table A.1a**) have a thermoinducible λ *red* recombination system, which permits the efficient incorporation of linear DNA into the chromosome [80], We used both deletion and nonsense mutations for *rnhA* and *rnhB*. Complete deletions of the coding regions are ideal because they eliminate the possibility that defective products might still be present, although they also remove the promoters of neighboring *dnaQ* and *dnaE* genes, which then are transcribed, we assume, from alternative promoters. Therefore, we also constructed truncated forms of RNase HI and HII (verified by sequencing) by introducing two consecutive stop codons (Ochre and Opal) upstream of *dnaQ* and *dnaE* promoters and after codon 48 in *rnhA* and after codon 70 in *rnhB* in strain BW1988 and BW2037 exploiting oligo-directed transformation (**Table A.1a**). We then compared the transformation frequencies by the LacZ.R6$_{I2}$, LacZ.R1$_{S1}$ or LacZ.D oligo in the *rnh* deletion and nonsense mutants and no differences were revealed (**Table A.5a**,**b**). We concluded that both the deletion and nonsense *rnh* mutant alleles behave similarly, at least with respect to gene correction by RNA-containing oligos. Bacterial media and growth conditions were as described before [81].

2.3.2 Yeast strains

The yeast haploid strains used in this work derives from either FRO-694, which contains the GSKU cassette [82] and the I-SceI cutting site in *TRP5,* inserted between nucleotide (nt) C1001 and nt C1002, or FRO-767 [46] (**Table A.1b**). Three different oligos with homology to the *TPR5* gene containing a 2-base deletion (C-C1001-1002) plus a

nonsense mutation (nt G1017 to A), only the 2-base deletion mutation, or only the nonsense mutation were used to transform the FRO-694 to pop out the GSKU cassette in *trp5* and generate the three different *trp5* alleles in three different strains, YS-301/303, YS-316/318 and YS-320/322, respectively (**Table A.1b**), following the "*delitto perfetto*" method as described [82]. The oligos used for the construction of the different *trp5* alleles are shown here with the introduced mutations in bold:

80DeltaCC+STOP_TRP5.e:

5'-AAGAGAGTTGGAAAAGGGTTTTGATGAAGCTGTCG__GATCCCACATTCTG

AGAAGACTTCAAATCCTTGTATTCTTATATT (2-base deletion & nonsense mutation)

80DeltaCC_TRP5.e:

5'-AAGAGAGTTGGAAAAGGGTTTTGATGAAGCTGTCG__GATCCCACATTCTG

GGAAGACTTCAAATCCTTGTATTCTTATATT (2-base deletion mutation)

82STOP_TRP5.e:

5'-AAGAGAGTTGGAAAAGGGTTTTGATGAAGCTGTCGCCGATCCCACATTCTG

7 7

AGAAGACTTCAAATCCTTGTATTCTTATATT (nonsense mutation)

Null alleles for *RNH201*, *MSH2* or both were created by replacing the open reading frame of each of these genes by either the *kanMX4* or the *hygMX4* module. Isogenic y yeast haploid strains KK-158 and KK-159 were derived from FRO-767 and FRO-768 [46]. KK-158 and KK-159 were constructed from FRO-767 and FRO-768 by replacement of *UNG1* with the *hygMX4* cassette.

2.3.3 Transformation using oligos

For experiments in *E. coli*, electrocompetent cells were prepared as described [81] and used immediately. The final wash and resuspension were in RNase-free water. RNA-containing oligos retained ≥90% of their activity after incubation with fresh cells for 3

min. Electroporation was performed as described [81], immediately after adding 100 ng of oligo in 50 μl of a cell suspension. For transformation experiments to mutate the *rpsL*$^+$ gene, 10 μl of cultures after recovery in the SOC media were transferred to 1 ml of LB media and incubated overnight to reduce the proportion of the original streptomycin-sensitive ribosomes. Cells were then diluted in 10 mM $MgSO_4$ and plated in an overlay of 4 ml of soft agar on LB agar with streptomycin to select for transformants and on LB agar to measure viability. Survival after transformation was 24%-30% for BW1988, BW2037 and all the derivative strains used. All strains with *rnhA* mutations form smaller colonies on LB agar media than strains with wild-type *rnhA*. For experiments in yeast, transformation with RNA-containing or DNA-only oligos (1 nmol) were done as described [46]. Cells from each oligo transformation were either plated to selective Trp$^-$ media for *TRP5* oligos or to selective Leu$^-$ media for *LEU2* oligos and were diluted and plated on the rich YPD media. Survival after yeast transformation was 29% for WT, 26% for *msh2*, 24% for *rnh201* and 29% for *rnh201 msh2* cells, respectively. The relative transformation frequencies were calculated by dividing the number of transformants per 10^7 viable cells obtained with an RNA-containing oligo by the median of the number of transformants per 10^7 viable cells obtained with the corresponding DNA-only oligo in the same experiment. The results are each expressed as a median and 95% confidence limits (in parentheses), or alternatively the range when number of repeated experiments was <6. Random clones derived from gene correction by LacZ.$R1_{S1}$ in the *rnhB mutS* background, LacZ.$R5_{S1}$ in the *rnhA rnhB* background, RpsL.$R1_{S1}$ in the *rnhB mutS* background, or TRP5.R2_$R1_{I2_S1}$ in the *rnh201 msh2* background were sequenced at the region targeted by the oligos, and all (24/24, 19/19, 13/13 and 14/14, respectively) had the expected corrected *lacZ*$^+$, *rpsL*^{r}, or *TRP5* sequence, with no additional changes.

2.3.4 Colony PCR of Leu$^+$ transformants and StuI-digestion of their PCR products

For each transformation with *LEU2* oligos, 20 Leu^+ transformants were selected. Colony PCR was performed on those transformants, amplifying with primers LEU2.3 and LEU2.6 a 900-bp region in *LEU2* locus where a new StuI restriction site is expected. The resulting PCR products were treated with StuI (New England Biolabs) and analyzed by agarose gel electrophoresis to confirm the presence of the StuI restriction site.

2.3.5 Standard genetic and molecular biology techniques

Standard genetics and molecular biology analyses were done as described [46, 81, 82]. Samples for sequencing were submitted to Eurofins MWG Operon.

2.3.6 RNase HII cleavage assay

RNA-containing oligos and their corresponding DNA-only oligos (**Table A.2**) were 5′-end-labeled using [γ-^{32}P]ATP (PerkinElmer) and T4 Polynucleotide Kinase (NEB). Double-strand substrates (**Figure 2.2a**) were prepared by annealing appropriate complementary oligos in 1X ThermoPol Reaction Buffer (NEB), which is the RNase HII reaction buffer, at pH 8.8. RNase HII reactions were performed by incubating 40 fmol of each substrate with 1 unit of RNase HII (NEB) in the same buffer for 2 hours at 37 $^{\circ}$C, or with 0.75 units of RNase HII for 30 minutes at 37 $^{\circ}$C. 20-100 Oligo Length Standard (IDT) was used as a marker. Samples were heated to 95 $^{\circ}$C for 5 min before electrophoresis in 15% (w/v) polyacrylamide 8M Urea gel. Following electrophoresis, gels were exposed to a phosphor screen overnight. Images were taken with Typhoon $Trio^+$ (GE Healthcare) and obtained with ImageQuant (GE Healthcare). Band intensities were quantified by Multi Gauge V3.0 (Fujifilm).

2.4 Results

2.4.1 Gene correction by rNMP-containing oligos: impact of RNases H

We initially set up a gene correction assay for different RNA-containing oligonucleotides (oligos) in the *E. coli* strain BW1988, which is an MMR-deficient (*mutS* null) strain expressing the λ phage recombination system for highly efficient oligo targeting [81, 83] (**Table A.1**). In this system, lagging-strand oligos anneal to the complementary gapped single-strand region at the replication fork, are incorporated into the genome similarly to Okazaki fragments and serve as templates for gene correction in the next round of DNA replication [83]. A lagging-strand oligo containing a tract of 6 rNMPs was also much more efficient at gene correction than the corresponding complementary oligo [56]. The *E. coli* BW1988 strain had a *lacZ* marker gene altered by a two-base frameshift deletion (ΔGG1370-1) and a missense mutation at nucleotide 1384, changing codon GAG (lactose deficient, Lac^-) into AAG (lactose proficient, Lac^+) (**Table A.1a**). To correct these *lacZ* mutations, BW1988 was transformed with lagging-strand oligos having the sequence of the wild-type *lacZ* gene and containing 6 or 2 rNMPs opposite to the deletion (LacZ.R6$_{I2}$ or LacZ.R2.47$_{I2}$), 1 rNMP or 5 rNMPs opposite to the missense mutation (LacZ.R1$_{S1}$ or LacZ.R5$_{S1}$), or no rNMPs (LacZ.D or LacZ.D.47) (**Figure 2.1a** and **Table A.2**). Transformation results are shown in **Table 2.1**. In the *mutS* background, differently from the LacZ.R6$_{I2}$ and the LacZ.R2.47$_{I2}$, the LacZ.R1$_{S1}$ and the LacZ.R5$_{S1}$ oligos produced a surprisingly low frequency of Lac^+ transformants relative to the control DNA-only oligos. The experiment was repeated with mutants for RNase HI (*rnhA*) and RNase HII (*rnhB*) in the *mutS* background to see if the enzymes would affect oligo transformation by attacking the embedded rNMPs. As presented in **Table 2.1**, the *rnhA* mutation showed a strong effect only on the transformation frequency by LacZ.R5$_{S1}$ oligo (a factor of 85 ($6\times10^{-3}/7\times10^{-5}$) increase). Differently, the *rnhB* mutation enhanced the relative transformation frequency of the LacZ.R1$_{S1}$ oligo a factor of 6,800 ($3.43/5\times10^{-4}$) and that of the LacZ.R6$_{I2}$ oligo a factor of 8 (0.16/0.02). It had no effect on the LacZ.R2.47$_{I2}$ oligo, and it enhanced the frequency of the LacZ.R5$_{S1}$ oligo only in combination with the *rnhA* mutation. Clearly, our data show that *in vivo* RNase HII very efficiently targets a single

mispaired rNMP and cannot target the isolated 2-rNMP loop, whereas RNase HI is preferred to RNase HII for targeting a tract of 5 rNMPs with a mispair. The stronger resistance of LacZ.R6$_{I2}$ and in particular of LacZ.R2.47$_{I2}$ to cleavage by RNase HI and HII is probably due to the secondary structure of the insertion loop present in these oligo sequences. From *in vitro* studies it is known that RNases HI and H1 can cleave DNA substrates containing a stretch of at least four ribonucleotides, whereas RNases HII and H2 can cleave a single ribonucleotide embedded in DNA, even when mispaired, or longer embedded rNMP tracts [37, 79]. It was then surprising to find that the *rnhB* mutation did not affect the gene correction frequency by the LacZ.R5$_{S1}$ and the LacZ.R2.47$_{I2}$ oligos and it only modestly affected that by the LacZ.R6$_{I2}$ oligo compared to that by the LacZ.R1$_{S1}$ oligo (**Table 2.1**). The data also demonstrate that all the different rNMPs embedded in double-strand *E. coli* chromosomal DNA are used as templates for DNA synthesis.

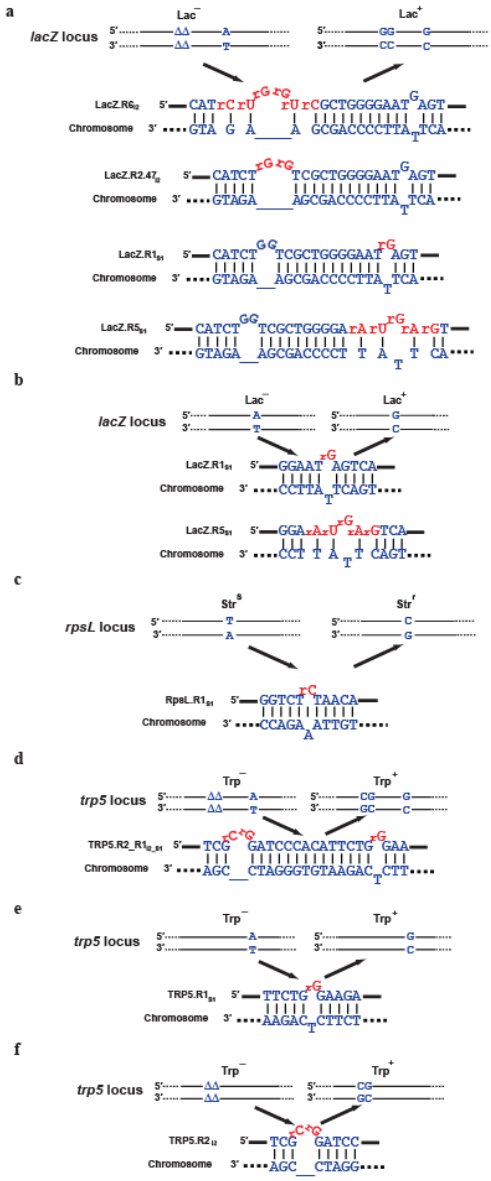


Figure 2.1 Diagrams and sequences of the loci targeted by the RNA-containing oligos. (**a**–**f**) In the name of the RNA-containing oligos, substitutions are indicated by a subscript capital "S" and insertions by a subscript capital "I". The letters "S" and "I" are followed by a subscript number indicating the number of bases that are substituted or inserted, respectively. The *lacZ* locus containing a two-base deletion and a substitution mutation targeted by the LacZ.R6$_{I2}$, LacZ.R2.47$_{I2}$, LacZ.R1$_{S1}$, or LacZ.R5$_{S1}$ oligo (**a**). The *lacZ* locus containing a substitution mutation targeted by the LacZ.R1$_{S1}$, or

LacZ.R5$_{S1}$ oligo (**b**). The *rpsL* locus targeted by the RpsL.R1$_{S1}$ oligo (**c**). The *trp5* locus containing a two-base deletion and a substitution mutations targeted by the TRP5.R2_R1$_{I2_S1}$ oligo (**d**), containing just a substitution mutation targeted by the TRP5.R1$_{S1}$ oligo (**e**), or containing only a two-base deletion mutation targeted by the TRP5.R2$_{I2}$ oligo (**f**).

Table 2.1 The effect of RNase HI (*rnhA*) and RNase HII (*rnhB*) mutations on transformation by rNMP-containing oligos in *E. coli mutS* cells.

Genotype	LacZ.R6$_{I2}$		LacZ.R2.47$_{I2}$		LacZ.R1$_{S1}$		LacZ.R5$_{S1}$	
mutS	0.02	(0.01–0.03)	0.18	(0.13–0.22)	5×10^{-4}	(4×10^{-4}–6×10^{-4})	7×10^{-5}	(0–9.7×10^{-5})
mutS rnhA	0.04	(0.02–0.05)	ND		9×10^{-4}	(7×10^{-4}–1.4×10^{-3})	6×10^{-3}	(5.5×10^{-3}–7.3×10^{-3})
mutS rnhB	0.16	(0.13–0.18)	0.16	(0.13–0.20)	3.43	(1.74–3.87)	9×10^{-5}	(4.3×10^{-5}–1.8×10^{-4})
mutS rnhA rnhB	0.38	(0.35–0.46)	ND		2.40	(1.94–2.81)	0.08	(0.06–0.09)

The values are relative frequencies of Lac$^+$ transformants (see **Materials and Methods**). The significance of all non-overlapping confidence limit values was confirmed by the Mann-Whitney *U*-test ($P < 0.05$) (**Table A.3a**). The median transformation frequencies per 10^7 cells and 95% confidence limits (in parentheses) for the strains transformed with the LacZ.D and the LacZ.D.47 oligos were as follows: *mutS*, 7,290 (4,120–10,800); *mutS rnhA*, 7,780 (5,090–11,300); *mutS rnhB* 4,840 (2,880–6,950); *mutS rnhA rnhB*, 12,908 (7,750–18,000), and *mutS*, 2,430 (1,770–3,760); *mutS rnhB* 2,720 (2,250–3,630), respectively. In the absence of a transforming oligo, the frequency of Lac$^+$ colonies per 10^7 viable cells were <0.1. The numbers of repeats for each of the *mutS*, *mutS rnhA*, *mutS rnhB*, and *mutS rnhA rnhB* strains transformed with these oligos were as follows: LacZ.R6$_{I2}$: 6 or 7; LacZ.R2.47$_{I2}$: 4; LacZ.R1$_{S1}$: 6 or 7; LacZ.R5$_{S1}$. ND, not determined. The strains used were BW1988, BW2028, BW2029, and BW2032.

2.4.2 RNase HII cleavage of heteroduplexes containing paired or mispaired rNMPs

The inability of the *rnhB* mutation to increase gene correction frequencies by the LacZ.R5$_{S1}$ and the LacZ.R2.47$_{I2}$ oligos and its capacity to increase gene correction by the LacZ.R6$_{I2}$ oligo to a much lesser extent than by the LacZ.R1$_{S1}$ oligo may be a reflection of the specificity of RNase HII. To test this hypothesis, we analyzed the ability of the purified enzyme to cleave heteroduplexes containing the same sequences and RNA:DNA mismatches that we examined in the transformation experiments (**Table A.2** and **Figure 2.2a**). RNase HII cleavage of these different substrates was determined by polyacrylamide gel electrophoresis (PAGE) following two-hour incubation at 37 °C (see Online Methods section). Under these experimental conditions, substrates with a single rNMP mispair, whether alone or in a tract of 5 rNMPs (substrates S2 and S6, respectively), were cleaved to the same extent as the rNMP-containing substrates with no mismatches (S1 and S5, respectively) (**Figure 2.2a**,**b**). RNase HII cleaved 5′ to the last ribonucleotide, or the junction ribonucleotide, in the 5 and 6-ribonucleotide tracts. The faint bands appearing below the major cleavage band for each substrate containing more than 1 rNMP are products of RNA degradation, as they appear also in the absence of RNase HII (**Figure A.1**). The substrate containing a loop of two unpaired rNMPs alone (S11) was not cleaved. The substrate containing the same rNMP loop within a stretch of 6 rNMPs was cleaved to a lesser extent (53%) than those containing a single mispair (S2 and S6). These data, except those obtained with the S6 substrate, are entirely consistent with our *in vivo* results (**Table 2.1**), in which gene correction by LacZ.R2.47$_{I2}$ or the LacZ.R6$_{I2}$ oligo is significantly (P = 0.0106 or P = 0.0021, respectively, **Table A.3a**) more efficient than correction by the LacZ.R1$_{S1}$ oligo in the presence of RNase HII, and deletion of *rnhB* does not increase at all gene correction by the LacZ.R2.47$_{I2}$ oligo. When we performed a similar experiment in conditions in which RNase HII cleaved the substrate with a single rNMP mispair (S2) at 77%, the substrate with the same mispair in a tract of 5 rNMPs (S6) was cleaved at 50%, thus less efficiently than S2 though much more efficiently than S4 (5.4%) and S11 (<1%) (**Figure 2.2c**,**d**). Overall, the RNase HII

cleavage results do not explain why *in vivo* the *rnhB* mutation did not increase gene correction by the LacZ.R5$_{S1}$ oligo while it increased gene correction by the LacZ.R6$_{I2}$ (**Table 2.1**). Therefore, in this case, there is no correlation between the preference for the substrate containing a single rNMP mispair in a tract of 5 rNMPs *in vitro* and *in vivo.*

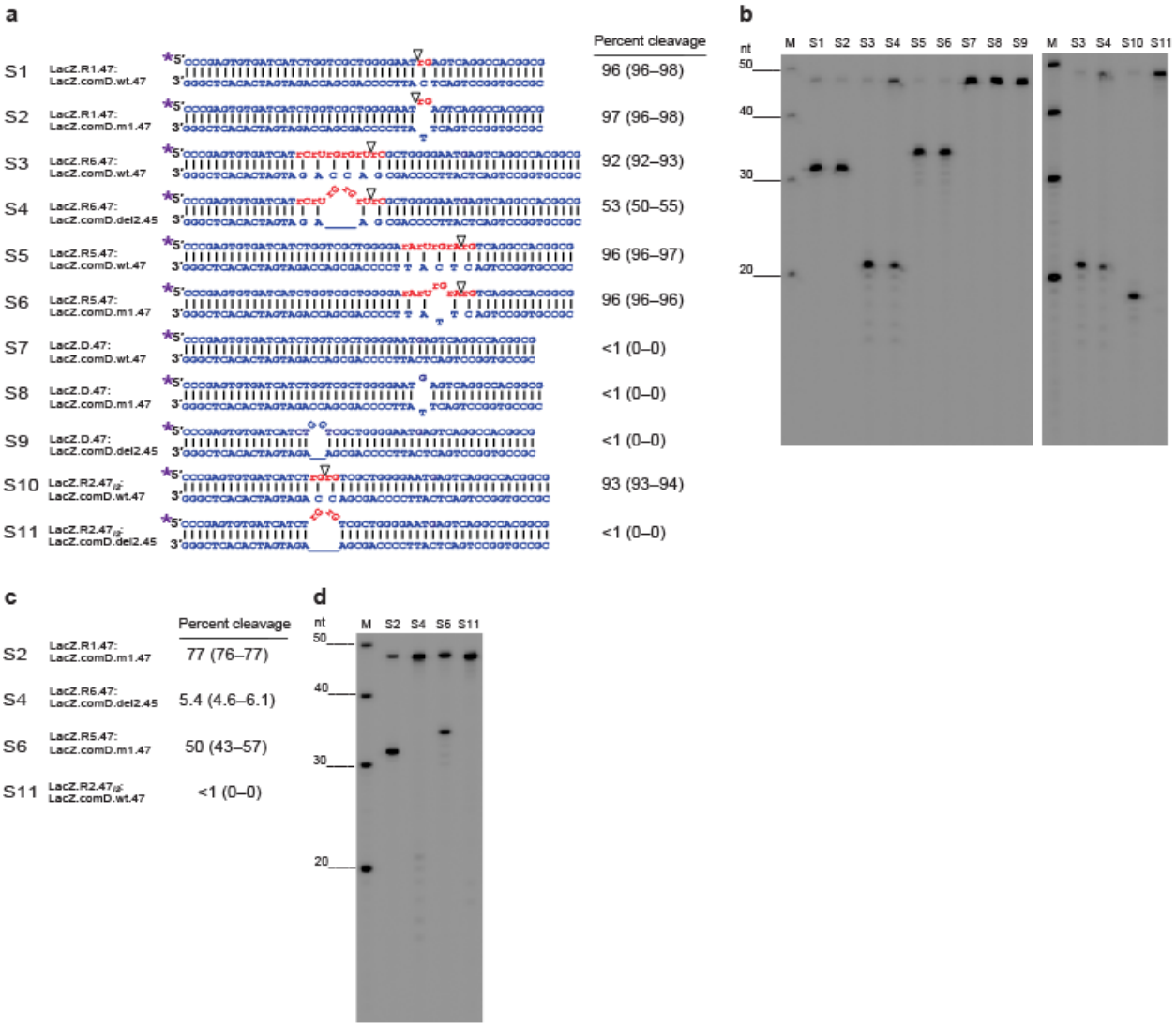


Figure 2.2 RNase HII cleavage specificity. (**a**) Structural presentation of 5′-radiolabeled (^{32}P, indicated by a purple asterisk) substrates (S1-S11) and cleavage percentage for each substrate, expressed as median and range (in parentheses) from 3 independent samples. Inverted triangles indicate the cleavage sites. (**b**) Denaturing polyacrylamide gels

showing fragments resulting from cleavage using RNase HII. M: 20-100 nt oligonucleotide marker. The gel images were cropped above the 50-nt band of the marker. S1-S11: substrates used. (**c**) Substrates (S2, S4, S6, S11) used in the experiment shown in panel **d** and their cleavage percentage, expressed as mean and range (in parentheses) from 2 independent samples. (**d**) Denaturing polyacrylamide gel showing fragments resulting from cleavage using reduced amount of RNase HII and shorter incubation time. M: 20-100 nt oligonucleotide marker. The gel image was cropped as in **b**. S2, S4, S6, S11: substrates used.

2.4.3 Functional redundancy of MMR and RNase HII at an rG:dT site

We next examined the capacity of the MMR system to recognize RNA:DNA mismatches in *E. coli*. The MMR system is highly conserved from prokaryotes to eukaryotes. It recognizes DNA:DNA mispairs, such as base-base mismatches as well as small insertions and deletions (ins/dels) occurring both during DNA replication [84] and recombination [85]. We modified the bacterial strain BW1947, containing just one missense mutation in *lacZ* (nucleotide 1384 G→A) [81], to generate a set of strains having wild-type and mutant alleles of *mutS*, *rnhA* and *rnhB* in all combinations (**Table A.1a**). We then transformed these strains with the LacZ.R1$_{S1}$ (**Figure 2.1b**) or the LacZ.D control oligo. Transformation results are shown in **Table 2.2**. As expected, *lacZ* correction by the DNA-only oligo in *mutS* cells was over two orders of magnitude higher than in all *mutS*$^{+}$ cells, in agreement with previous findings [81, 83]. MutS was also able to recognize the mismatched rG:dT mispair, as gene correction frequency by the LacZ.R1$_{S1}$ oligo in *mutS* was a factor of 5 higher than in *mutS*$^{+}$ cells. The relative transformation frequency of LacZ.R1$_{S1}$ increased only to a factor of 2 in the absence of RNase HI both in *mutS*$^{+}$ and in *mutS* cells, in line with the observation that RNase H type 1 cannot cleave at rNMPs in stretches of less than four [37, 79]. In *mutS*$^{+}$ cells an *rnhB* mutation enhanced the frequency of transformation by LacZ.R1$_{S1}$ only a factor of 11 (21.8/1.88) (**Table 2.2**), as

opposed to the factor of 6,800 enhancement seen in a *mutS* mutant (**Table 2.1**). The *rnhB mutS* double mutant showed correction frequency for the LacZ.R1$_{S1}$ oligo a factor of 17,500 (33,000/1.88) higher than the wild type and a factor of 1,500–3,400 higher than either of the single mutants (**Table 2.2**). The gene correction frequency by LacZ.R1$_{S1}$ increased a factor of 77,000 (145,000/1.88) in the *rnhA rnhB mutS* triple mutant compared to the frequency in wild-type cells, and became similar to that of the corresponding DNA-only oligo in the *mutS* backgrounds used (**Table 2.2**). Further purification of the LacZ.R1$_{S1}$ oligo by PAGE produced no observable changes in gene correction frequencies (not shown). These data prove not only that the *E. coli* MMR system can recognize and remove an rG:dT mismatch in DNA, but also that MutS and RNase HII act redundantly, via independent pathways, to remove an rG:dT mispair.

Table 2.2 Tolerance of an rG:dT mispair in *E. coli*.

	LacZ.R1$_{S1}$		**LacZ.R5$_{S1}$**		**LacZ.D**	**No oligo**
Genotype[a]	Lac$^+$ freq.[b]	Rel. tr. freq.[c]	Lac$^+$ freq.	Rel. tr. freq.	Lac$^+$ freq.	Lac$^+$ freq.
WT	1.88 (0.82–5.97)	0.02	0.36 (0–0.48)	4×10^{-3}	90.7 (38.5–177)	< 0.1
rnhA	12.1 (8.75–23.2)	0.04	256 (176–803)	0.76	336 (162–485)	< 0.1
rnhB	21.8 (20.8–26.1)	0.13	0.715 (0.62–0.89)	4.3×10^{-3}	165 (140–242)	< 0.1
rnhA rnhB	174 (154–196)	0.37	22,600 (19,700–27,300)	48.2	470 (371–2,110)	< 0.1
mutS	17.4 (14–19.6)	1.2×10^{-4}	7.53 (5.48–14.2)	10^{-5}	82,300 (50,000–104,950)	7.76 (4.6–9.51)
rnhA mutS	70 (34.8–137)	4×10^{-4}	485 (213–820)	5×10^{-3}	93,908 (37,900–145,000)	32.1 (10–57.6)
rnhB mutS	33,000 (30,400–79,500)	0.27	5.23 (4.34–5.66)	10^{-5}	124,200 (84,600–163,000)	3.77 (3.51–5.06)
rnhA rnhB mutS	145,000 (33,300–320,000)	0.69	22,400 (11,600–32,600)	0.11	212,000 (100,000–267,000)	24 (20.2–40.2)

[a]The strains used were BW2038A,B, YSB-21A,B, YSB-22A,B, and YSB-23A,B, which were *mutS*$^+$ and BW2037A,B, YSB-19A,B, YSB-17,18, YSB-20A,B, which were *mutS*.

[b]The values are median and the range (in parentheses) of *E. coli* Lac^+ transformant colonies reverting the missense mutation per 10^7 viable cells for each oligo. The numbers of repeats for each of the strains transformed with these oligos were 4. The significance of all non-overlapping confidence limit values was confirmed by the Mann-Whitney *U*-test ($P < 0.05$) (**Table A.3b**).

[c]The relative frequencies of Lac^+ transformants (see **Materials and Methods**) are used for comparisons within *mutS*$^+$ or *mutS* strains.

2.4.4 Inefficient MMR of an rC:dA mispair in *E. coli*

We then investigated whether MMR could recognize an rC:dA mismatch. We used an oligo of the lagging strand containing an rC ribonucleotide (RpsL.R1$_{S1}$) that would replace a dT in the *rpsL* wild-type gene of *E. coli* BW2037 (streptomycin sensitive cells, Str^S) and derivative strains (**Table A.1a**), generating the K87R mutation (*rpsL40*) that confers streptomycin resistance (Str^r) [81] (**Figure 2.1c**). We found that the dC:dA mismatch was efficiently repaired by the MMR system in *E. coli*, consistent with previous findings [86]. Whereas the ribonucleotide in the rC:dA mismatch in *rpsL* was removed very efficiently by RNase HII, there was only a modest, but still detectable, correction by MMR in *rnhB* cells (**Table 2.3**). These data differ from our results with an rG:dT mismatch in the *lacZ* gene, where MMR had a much stronger effect. We conclude that MMR can recognize an rC:dA mismatch, but with less efficiency than dC:dA or the complementary rG:dT mismatches.

Table 2.3 Strong effect of RNase HII and minor effect of MutS in preventing gene correction by an rC:dA mispair in *E. coli*.

	RpsL.R1$_{S1}$		RpsL.D	No oligo
Genotype[a]	Str^r freq.[b]	Rel. tr. freq.[c]	Str^r freq.	Str^r freq.

WT	8	0.8	10	<0.1
	(1.46–8.60)		(5.41–17.7)	(0–0)
rnhB	1,850	316	5.86	<0.1
	(1,410–2,460)		(4.37–8.27)	(0–0.03)
mutS	11.6	0.001	10,500	0.76
	(10.5–13.1)		(8,070–12,900)	(0.03–2.70)
rnhB mutS	4,310	0.826	5,220	0.32
	(3,000–5,320)		(2,660–8,930)	(0.13–2.39)

[a]The strains used were BW2038A,B, BW2040A,B, BW2037A,B and BW2039A,B.

[b]The frequency of streptomycin resistant transformant colonies per 10^7 viable cells for wild-type, *mutS*, *rnhB* single mutant and *rnhB mutS* double mutant after transformation with *RpsL.R1*$_{S1}$ or *RpsL.D* oligo is shown as median and 95% confidence limits (in parentheses), or alternatively range when number of repeats was <6. The numbers of repeats for each of the strains transformed with these oligos were 4 or 6. The significance of all non-overlapping CI or range values was confirmed by Mann-Whitney *U*-test ($P < 0.05$) after subtraction of the background values (**Table A.3c**).

[c]Relative frequency of Str^r transformants (see **Materials and Methods**).

2.4.5 Mismatch repair of RNA:DNA mispairs in yeast cells

To test if the MMR system could recognize RNA:DNA mismatches in eukaryotic cells, we created a yeast *S. cerevisiae* strain in which we inactivated the genomic *TRP5* gene by introducing a two-base deletion and one-base substitution to cause a frameshift and a nonsense mutation 15 bp apart (YS-301 and YS-303, (tryptophan auxotrophic cells, Trp$^-$) **Table A.1b**). The combination of these two mutations reduced the spontaneous reversion of *trp5* to the tryptophan prototrophic (Trp$^+$) phenotype to $<10^{-9}$, allowing the detection

of very low frequencies of gene correction despite the low efficiency of gene targeting in yeast [87, 88]. MMR was inactivated by deletion of *MSH2*, a yeast *mutS* homolog [84]. RNase H2 was inactivated by deleting the gene of the catalytic subunit, *RNH201* [37]. Unlike the *E. coli* λ recombination system [89], gene correction at the yeast *trp5* locus displayed no strand bias (**Figure A.2**). All *TRP5* oligos used here had a sequence corresponding to the *TRP5* sense strand. Wild-type and mutant cells were then transformed with an oligo containing two rNMPs to repair the deletion and another rNMP to correct the nonsense mutation in *trp5* (TRP5.R2_$R1_{I2_S1}$) (**Figure 2.1d** and **Table A.2**). A DNA-only oligo (TRP5.D) provided a reference. The results (**Table 2.4a**) were similar to those obtained in *E. coli* with the LacZ.$R1_{S1}$ oligo (**Table 2.2**). Gene correction in the *rnh201* single mutant was only a factor of 2.5 more than that in wild-type cells. As expected , the *msh2* deletion enhanced gene correction by the DNA-only oligo TRP5.D. With the TRP5.R2_$R1_{I2_S1}$ oligo, in which both the 2-base insertion and the substitution were rNMPs, we observed a factor of 5 (3.71/0.72) increase in gene correction efficiency in the *msh2* mutant and more than a factor of 45 (33.8/0.72) increase in the *rnh201msh2* double mutant, making the RNA-containing oligo as efficient as the corresponding DNA-only oligo. We concluded that the yeast MMR system also recognizes RNA:DNA mismatches and can act redundantly with RNase H2 to remove mispaired ribonucleotides embedded in DNA.

To determine whether the RNA:DNA ins/del and the rG:dT mispairs could be independently recognized by the MMR system in yeast, we constructed strains carrying either the nonsense mutation or the 2-base deletion in *trp5* and having *msh2* and/or *rnh201* mutant alleles. We transformed these strains with oligos containing either a 1-rNMP substitution (TRP5.$R1_{S1}$) or a 2-rNMP insertion (TRP5.$R2_{I2}$) (**Figure 2.1e**,**f** and **Table A.2**), respectively. Transformation results revealed that the rG:dT mispair was recognized by MMR in yeast (**Table A.4a**). To determine the effect of MMR on the 2-

rNMP insertion we sequenced all the Trp^+ transformant clones obtained in the experiment with the TRP5.$R2_{I2}$ oligo in order to distinguish Trp^+ clones targeted by the oligos (TRP5.$R2_{I2}$ or the TRP5.D control) from Trp^+ revertant clones (**Table A.4b–d**). Sequencing of Trp^+ transformants revealed that the ins/del RNA:DNA mismatch was clearly targeted both by Msh2 and RNase H2 (**Table 2.4b**).

Table 2.4 Competition between mismatch repair and RNase H type 2 functions in the removal of RNA:DNA mispairs in yeast.

a

	TRP5.R2_ $R1_{I2\_S1}$			**TRP5.D**		**No oligo**		**Non-specific oligo[a]**	
Genotype	Trp^+ freq.		Rel. tr. freq.[b]	Trp^+ freq.		Trp^+ freq.		Trp^+ freq.	
WT	0.72	(0.40–1.32)	0.30	2.39	(1.50–3.51)	< 0.1	(0–0)	<0.1	(0–0)
msh2	3.71	(2.07–4.31)	0.12	31.6	(24.5–37.2)	< 0.1	(0–0)	ND	ND
rnh201	1.84[c]	(1.15–2.88)	0.67	2.76	(1.83–3.29)	< 0.1	(0–0.03)	ND	ND
rnh201 msh2	33.8	(26.3–35.6)	0.76	44.6	(32.9–48.8)	< 0.1	(0–0.57)[d]	0.51	(0–1.36)

b

	TRP5.$R2_{I2}$			**TRP5.D**		**No oligo**	
Genotype	Trp^+ freq.		Rel. tr. freq.	Trp^+ freq.		Trp^+ freq.	
WT	1.65	(0.58–2.50)	0.04	39.1	(35.9–43.6)	< 0.1	(0–0)
msh2	7.64	(4.96–9.82)	0.17	46.1	(24.3–76.8)	< 0.1	(0–0)
rnh201	9.64	(7.51–15.0)	0.35	27.7	(12.3–36.9)	< 0.1	(0–0)
rnh201 msh2	26.9	(20.5–35.3)	0.38	70.0	(63.3–86.9)	< 0.1	(0–0)

(**a**) Trp^+ transformant frequencies for strains that contained both a two-base pair deletion and a nonsense mutation. Each value represents the median and range (in parentheses) for 10^7 viable cells, based on 8 experiments. The significance of all non-overlapping confidence limits or range values was confirmed by Mann-Whitney *U*-test ($P < 0.05$) after subtraction of the no-oligo background values (**Table A.3d**). (**b**) Trp^+ transformant frequencies obtained by precise correction of a two-base deletion calculated by

multiplying the number of Trp^+ transformant colonies per 10^7 viable cells (shown in **Table A.4b**) by the percentage of clones with precise correction of the "CG" deletion (pattern R in **Table A.4c,d**). Each value represents the median and range (in parentheses) for 10^7 viable cells, based on 4 experiments. The significance of all non-overlapping confidence limits or range values was confirmed by Mann-Whitney *U*-test ($P < 0.05$) after subtraction of the no-oligo background values (**Table A.3e**).

2.4.6 Removal of paired rNMPs in DNA by RNase H2 and NER in yeast cells

NER machinery has the ability to recognize a vast array of backbone distortions in DNA [90]. With the ability of rNMPs to induce helical distortions in DNA [75, 91], one could predict that NER recognizes rNMPs in DNA as damage. Since BER, specifically uracil-DNA glycosylase (Ung1)-initiated BER, recognizes uracil of dUMPs as damage [92], one could also predict that uracil of rUMPs could be targeted by Ung1. To test whether RNase H2, NER, or BER targets paired rNMPs in DNA, we implemented a yeast assay of chromosomal double-strand break (DSB) repair, in which DNA oligos carrying embedded rGMP, rUMP, or deoxyribonucleotides only are templates for DSB repair (**Figure 2.3a**). rGMP-containing DNA oligo LEU2.rG was used to test targeting of RNase H2 and NER while rUMP-contaning DNA oligo LEU2.rU was used to test targeting of RNase H2 and Ung1. Oligos were designed to repair the DSB in the *leu2* gene and to create a point mutation (A:T→G:C) four to five nt downstream of the rNMP position, allowing the creation of a new StuI restriction enzyme recognition site in LEU2 locus. RNase H2-initated ribonucleotide excision repair (RER), NER, and BER all remove a short single-strand DNA region downstream of the damage during the repair [35, 90, 92]. Therefore, when transformed with rNMP-containing oligos after induction of DSB, an increase in the percentage of the new StuI site-containing Leu^+ transformants when a repair pathway of interest is deficient would indicate that the repair pathway targeted the rNMP for removal. RER, NER, and Ung1-initiated BER were inactivated by

deletion of *rnh201*, *rad14* (*S. cerevisiae* homolog of the human XPA, which is required in all types of NER [90]), and *ung1*, respectively.

Our results show that, among the repaired Leu+ transformants, 63–65% of clones from all genotypes transformed with the DNA-only oligo control (LEU2.D) contained the StuI site (**Figure 2.3b**,**c** and **Table A.6**). We found that the single paired rGMP of the LEU2.rG oligo was targeted by both RER and NER as the StuI site was present only in 30% of the wild-type cells, but significantly up to 90% (P = 0.0286) and 63% (P = 0.0286) in *rnh201* and *rad14* cells, respectively (**Figure 2.3b** and **Table A.6a**,**b**). RER also targeted the single paired rUMP of the LEU2.rU oligo while Ung1 targeted uracil from a dUMP but not an rUMP (**Figure 2.3c** and **Table A.6c**,**d**).

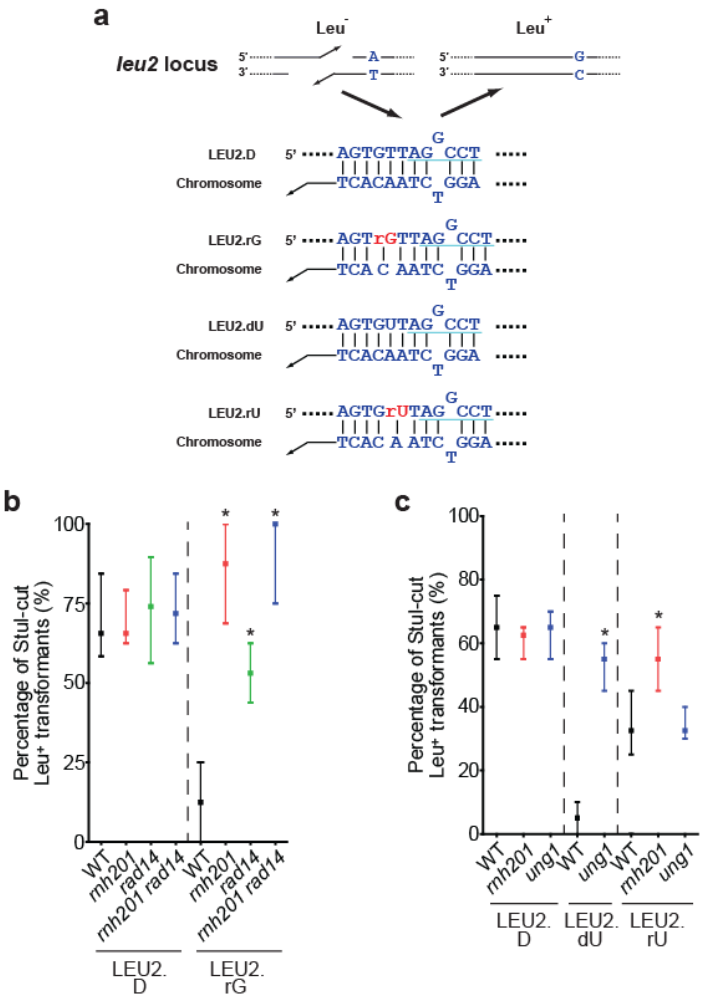


Figure 2.3 Targeting of paired rNMP by RNase H2 and NER in yeast.

(**a**) Diagram and sequence of the chromosomal *leu2* region targeted by DNA-control LEU2.D, rGMP-containing LEU2.rG, dUMP-containing LEU2.dU, and rUMP-containing LEU2.rU oligos (**Table A.2**). StuI recognition sequence is underlined in turquoise. (**b**,**c**) The oligos were transformed to either RNase H2–, NER–, and Ung1–proficient wild-type (WT; FRO-767,768), RNase H2–deficient (*rnh201*; FRO-984,985), NER–deficient (*rad14*; YS-388,389), *rnh201 rad14* (YS-390,391), or Ung1–deficient (*ung1*; KK-158,159) *S. cerevisiae* cells (see **Table A.1b**). Median percentages of StuI-cut Leu^+ transformants from four independent transformations are shown with ranges as bars. For each transformation, 20 Leu^+ transformants were selected for analysis. Mann-Whitney *U*-test was implemented for statistical analysis against the WT. *P* values of less than 0.05 are marked as asterisk. See **Table A.6** for more statistics.

2.5 Discussion

Previously, in experiments of DSB repair in yeast using oligos containing tracts of four ribonucleotides or longer, we showed that oligos containing the shortest RNA tracts were the most efficient at chromosomal gene modification [46]. Also an oligo containing a 2-base loop within a 6-rNMP tract was a factor of 25 to 50 less efficient at gene correction than the corresponding DNA-only oligo in *E. coli* [56]. Contrary to expectations, it was then remarkable to find in the current work that gene correction by an oligo containing just a single rNMP is a factor of 40 less efficient than that obtained with an oligo containing 6 rNMPs and a factor of 2,000 less than that obtained with the corresponding DNA-only oligo in *E. coli* (**Table 2.1**). Here we show that the capacity of rNMPs embedded in DNA to directly transfer genetic information to the genome is not only affected by the length but also by the structure of the embedded RNA tracts, which can be targets of specific proteins and be removed before serving as templates for DNA synthesis (**Table 2.1** and **2.2**).

While it is known that RNase H type 1 and type 2 have distinct cleavage specificities, mostly from biochemical studies [37, 38, 79], our *in vivo* gene correction results shed light on the *in vivo* substrate specificity of these enzymes and offer an opportunity to investigate the mechanism underlying cellular tolerance of ribonucleotides misincorporated into genomic DNA.

In addition to demonstrations of ribonucleotide incorporation by DNA polymerases *in vitro* [5-7, 17, 18, 21], recent work from the Kunkel group provides indirect evidence that ribonucleotides are incorporated into yeast DNA by low-fidelity (*pol2-M644G*) and wild-type Pol ε alleles and can destabilize the yeast genome in Δ*rnh201* cells [60, 61], strengthening the knowledge that yeast *rnh201* mutant cells display an increased rate of genome instability [93]. Up to a factor of 25 increase in the rate of spontaneous mutagenesis was observed in cells with *rnh201* null [60]. Interestingly, the frequency of base substitutions (G to A transitions) increased a factor of 6 in *pol2-M644G* Δ*rnh201* when the MMR gene *MSH6* was also deleted [61]. MMR could recognize RNA:DNA mismatches generated by polymerase ε and/or by other replicative polymerases [21]. Moreover, also repair polymerases, such as Pol μ, Pol β, and LigD polymerase can add rNMPs in DNA [13, 17, 18], and especially error-prone polymerases might be more promiscuous at rNMP incorporation in DNA. We suggest that RNA:DNA mismatches can arise by mispairing of an rNMP during its incorporation, or by a correct pairing during initial incorporation followed by the mispairing of a dNMP during a subsequent round of DNA synthesis. In the current work, we have found that MMR can remove the region containing a mispaired rNMP.

Among the types of RNA:DNA mismatches examined, we found that the rG:dT mispair is well recognized by the MMR system both in *E. coli* and in yeast cells. A preliminary

binding experiment of yeast Msh2–Msh6 to an rG:dT mismatch [61] supports our finding that the MMR system can target this mismatch *in vivo*. The rC:dA mispair is also targeted by the MMR system in *E. coli*, although less efficiently, since gene correction by the oligo generating the rC:dA mispair is increased only a factor of 2 in the absence of MutS in the *rnhB* mutant cells (**Table 2.3**). The sequence context surrounding the RNA:DNA mispair could certainly affect RNA:DNA mismatch recognition and/or removal. A more rigorous analysis of eukaryotic and prokaryotic MMR factor binding and ATPase functions in the context of various RNA:DNA mismatches will be an important next step to better characterize the capacity of MMR factors to process RNA:DNA mismatches. In yeast cells, gene correction by the TRP5.R2_R1$_{I2_S1}$ and by the TRP5.R2$_{I2}$ oligos was a factor of 5 and a factor of 4 more efficient in *msh2* than in wild-type cells, respectively (**Table 2.4**). Thus, we conclude that the MMR system can target, with different specificity, RNA:DNA mismatches (in *E. coli* and yeast) or small insertions/deletions (in yeast, not tested in *E. coli*). Moreover, our work demonstrates not only that RNA:DNA mismatches are susceptible to MMR but also that MMR and RNase H type 2 compete for RNA:DNA mismatches both in the *E. coli* and the yeast *S. cerevisiae* systems. Absence of both RNase H type 2 and MMR functions has a synergistic effect on gene correction frequency by the TRP5.R2_R1$_{I2_S1}$ and the LacZ.R1$_{S1}$ oligos (**Table 2.4a** and **Table A.4a**).

The MMR system of *E. coli* can recognize RNA:DNA mismatches only when these consist of one or two isolated ribonucleotides embedded in DNA, but not when the RNA:DNA mismatch is within an RNA:DNA duplex region. Indeed, while the single rG:dT mispair was very efficiently removed by the MMR system, there was no detectable effect of MMR on the same rG:dT mispair in the same sequence context, when the mispair was surrounded by two rNMPs on each side (**Figure 2.1b** and **Table 2.2**). Differently from the misincorporated rNMPs scattered in DNA, which are preferred

substrates for RNase H type 2 and are not targeted by RNase H type 1, mispairs in longer RNA-tracts, such as those that could be generated by error-prone primases that can misincorporate dNMPs, or those that could be present in RNA:DNA hybrid tracts of R-loops, are primarily targets of RNase H type 1, which can only be partially backed up by RNase H type 2 function.

NER system has the capacity to recognize a variety of DNA backbone distortions [90, 94]. Even a single rNMP embedded in DNA is sufficient enough to cause local helical distortions in DNA [75, 91] and decrease the efficiency of nucleosome assembly [59]. In this study, we found that a single rG/dC pair is recognized by NER machinery (**Figure 2.3** and **Table A.6a**,**c**). The level of targeting by NER compared to RNase H2 was lower. Whether this is due to the dominance of RNase H2 when it comes to removing rNMPs in DNA and/or the substrate specificity of NER in terms of sequence context needs further investigation.

In summary, we have demonstrated that the MMR system recognizes and targets several types of RNA:DNA mismatches present in DNA in *E. coli* and *S. cerevisiae* cells. In addition, an RNA:DNA heteroduplex region that contains a mismatch is preferentially and efficiently targeted by RNase H type 1 *in vivo*, whereas mispaired rNMPs in *E. coli* and yeast and small rNMP insertions in yeast are specific substrates for RNase H type 2. Thus, RNase H type 2 and MMR have overlapping activity in contributing to the removal of mispaired rNMPs embedded in DNA both in a prokaryotic and a eukaryotic cell system. Also, we have demonstrated that RNase H2 and NER recognize and target paired rNMPs in yeast. Our findings open up the possibility that other DNA repair mechanisms could tackle rNMPs embedded into DNA.

2.6 Acknowledgments

The authors thank P. W. Doetsch with his group and Y. W. Kow for discussions and comments; we are grateful to G. F. Crouse and R. Pai for suggestions on the paper, L. D. Williams for technical support for the gels, C. Flood for technical assistance and all the members of the Storici laboratory for advices in the course of the study. This research was supported by the Georgia Cancer Coalition grant R9028 (F.S.), the National Science Foundation grant MCB-1021763 (F.S.) and the Integrative Biosystems Institute grant IBSI-4 (F.S.).

CHAPTER 3

RNA INTRUSIONS CHANGE DNA ELASTIC PROPERTIES AND STRUCTURE

The study in Chapter 3 consists of the work published in *Nanoscale* 6 (2014) 10009-10017.

Chiu, H.C.[1*], Koh, K.D.[2*], Evich, M.[3], Lesiak, A.L.[4], Germann, M.W.[3], Bongiorno, A.[4], Riedo, E.[2], and Storici, F[1].

[1]School of Physics, Georgia Institute of Technology, Atlanta, GA 30332

[2]School of Biology, Georgia Institute of Technology, Atlanta, GA 30332

[3]Department of Chemistry, Georgia State University, Atlanta, GA 30302

[4]School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA 30332

*Equal contribution

3.1 Abstract

The units of RNA, termed ribonucleoside monophosphates (rNMPs), have been recently found as the most abundant defects present in DNA. Despite the relevance, it is largely unknown if and how rNMPs embedded in DNA can change the DNA structure and mechanical properties. Here, we report that rNMPs incorporated in DNA can change the elastic properties of DNA. Atomic force microscopy (AFM)-based single molecule elasticity measurements show that rNMP intrusions in short DNA duplexes can decrease – by 32% – or slightly increase the stretch modulus of DNA molecules for two sequences reported in this study. Molecular dynamics simulations and nuclear magnetic resonance spectroscopy identify a series of significant local structural alterations of DNA containing embedded rNMPs, especially at the rNMPs and nucleotide 30 to the rNMP sites. The demonstrated ability of rNMPs to locally alter DNA mechanical properties and structure may help in understanding how such intrusions impact DNA biological functions and find applications in structural DNA and RNA nanotechnology.

3.2 Introduction

DNA has unique mechanical properties that are crucial in many natural biochemical processes, such as specific DNA-binding to proteins, DNA replication, repair and recombination, and chromosome organization [95-101]. Comprehending the dynamics of many cellular functions requires the understanding of the physical behavior of DNA as many of the mechanisms by which genetic information is stored and used involve deformation of the DNA. Every system which binds, cleaves, or reads DNA is able to exploit and/or alter the structural and mechanical properties of DNA, which is affected by the nucleotide sequence [102] for recognition, packaging, and modification [97, 103]. These sequence-dependent effects are involved in modulating biological functions of DNA [102, 104]. DNA mechanical properties also play important roles in DNA-based nanotechnology applications, such as DNA origami, molecular scale electronics, and

nanomedicine [105-108]. It remains largely unknown how the presence of distortions and defects in DNA impacts its elasticity.

RNA is a polymer of units called ribonucleoside monophosphates (rNMPs), which differ from DNA units by an additional hydroxyl (OH) group in the sugar moiety (**Figure 3.1a**,**b**). Recent studies have revealed that rNMPs are unexpectedly the most abundant non-standard nucleotides present in DNA [12, 14, 18, 21, 34] (and references therein). Furthermore, rNMPs can be replicated during DNA synthesis and can transfer a genetic change to genomic DNA [46, 56, 58]. With the highly reactive extra OH group of the ribose sugar, accumulation of rNMPs in the DNA genome might distort the double helix, alter the elasticity, and increase the fragility of DNA. Their presence in DNA can be a threat for the genomic integrity of cells [34, 39, 54, 58] (and references therein) and could be a useful mean for manipulating DNA physical properties.

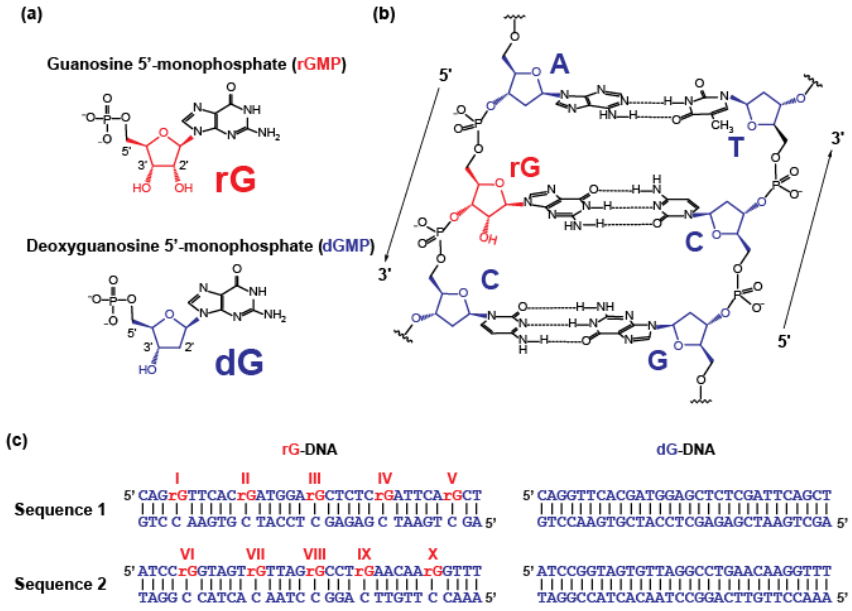


Figure 3.1 Structure and sequences of rNMP(s)-embedded DNAs analyzed in this study. (**a**) Chemical structures of an rNMP and a dNMP at the base G (rGMP and dGMP,

respectively). The ribose and deoxyribose are colored in red and blue, respectively. The 2′, 3′, and 5′ carbon atoms of the sugars are shown. (**b**) Scheme of an rGMP-embedded DNA. Hydrogen bonds are indicated by dashed lines. The 5′ to 3′ direction for each DNA strand is indicated. **(c**) Sequences of rGMP-embedded DNA, rG-DNA, and their DNA-control, dG-DNA, used in AFM experiments. The dNMPs are indicated in blue while the rGMPs are indicated in red, preceded by letter 'r'.

Despite demonstrations of their abundance and importance, very few reports address how scattered rNMPs present in DNA (**Figure 3.1b**) affect the structure and properties of DNA [67, 71, 72, 74, 75] (and references therein). In particular, only a few reports have examined the structural effects of isolated single rNMPs in DNA, and, in addition, only self-complementary DNA sequences have been used, in which an rNMP is present in both strands of DNA [72-75]. To the best of our knowledge, no data exist in the literature regarding elastic measurements and sequence-dependent structural distortions of double-strand (ds) DNA with isolated single rNMP intrusions. Here, we present an innovative, combined experimental and theoretical study, in which we designed two short ds DNA molecules containing isolated rNMP intrusions at specific bases in only one of the two strands (**Figure 3.1c**). We examined and identified how the elasticity and structure of the two DNA molecules are altered by these rNMP intrusions. Atomic force microscopy (AFM)-based single molecule force spectroscopy demonstrated that rNMP intrusions decrease – by 32% in one short DNA duplex – or slightly increase in the second duplex the stretch modulus of DNA. Molecular dynamics (MD) simulations and nuclear magnetic resonance (NMR) experiments indicated that rNMP inclusions locally introduce a torsional distortion of the sugar-phosphate backbone in DNA. The type of alteration and its degree are different for the specific rNMP sites we studied.

3.3 Materials and Methods

3.3.1 Sample preparation for AFM measurements

All oligonucleotides used in the AFM experiments were 30 nt long and were purchased from Dharmacon (**Table B.1**). For both Sequence 1 and Sequence 2, either single-strand (ss) dG or rG oligonucleotide was annealed to the complementary DNA oligonucleotides, compl_DNA, to produce ds dG-DNA and ds rG-DNA. The annealing was performed in 100 mM NaCl, 10 mM phosphate, and 0.1 mM EDTA at pH 7.4 by heating at 95 ℃ for 5 min and cooling slowly to room temperature. Each DNA was then immobilized on gold-coated substrates (Platypus Technologies, LLC), by putting a drop of DNA solution (0.1 mM of DNA molecules in 100 mM Na^+) on the substrate for 3 hours. Next, the substrate was immersed in 1 mM MCH (6-mercapto-1-hexanol) solution for 60 s to reduce non-specific binding of DNA and avoid molecular aggregation on the surface. The gold substrate was then rinsed with DEPC-treated water and is ready for use. Different concentrations of ds DNA have been tested to obtain optimal conditions that prevent formation of aggregates on the surface. For all measurements, the spring constants of gold-coated, COOH-modified silicon nitride cantilevers were individually calibrated using reference beam methods (see **Figure B.1** and **Table B.2**). To functionalize these cantilevers with streptavidin, they were first immersed in a PBS (phosphate buffered saline, pH 7.4) buffer solution of 5 mM EDC (1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride) and 10 mM NHS (N-hydroxysuccinimide) for 1 hour to activate the carboxyl (–COOH) group on the tip. Next, the cantilevers were immersed in a 100 mg/mL streptavidin solution for 2 hours. The streptavidin-coated cantilevers were then washed with PBS 10 times followed by DEPC-treated water to reduce nonspecific binding and stabilize the biomolecules. These cantilevers were then ready for AFM measurements. Next, the AFM liquid cell, which is used to hold the functionalized cantilever, was cleaned with RNaseZAP (Ambion), rinsed with copious ultra-filtered deionized water, and dried with compressed nitrogen gas, followed by the exposure to

UV light at a wavelength of 254 nm for 10 min, to remove any residual organic contamination.

3.3.1.1 Gold surface used in AFM experiments

The template stripped gold substrates were purchased from Platypus Technologies (http://www.platypustech.com/templatestrippedgold.html). The thickness of gold coating is 100 nm. The roughness of gold surface is 0.4 nm within 1x1 μm area. The AFM image of a freshly stripped gold surface is shown in **Figure B.2a**.

3.3.1.2 Calibration of AFM cantilevers

The spring constants of all used cantilevers (Novascan) are calibrated using the "reference beam method" with the same reference cantilever [109, 110]. This is to minimize the possible experimental error from the calibration procedure propagating into the final results. A reference cantilever with a known spring constant k_{ref} = 0.08 N/m was used. As shown in **Figure B.1**, a cantilever with unknown spring constant is pressed onto the reference cantilever. Then, the unknown spring constant can be determined by

$$k = k_{ref}\left(\frac{S_{hard}}{S_{ref}} - 1\right)\left(\frac{L_{ref}}{L_{ref} - \Delta L}\right)^3$$

where S_{ref} and S_{hard} are the slopes of force curves when the tip is in contact with the reference cantilever and a hard surface such as silicon, respectively. L_{ref} is the length of the reference cantilever, and ΔL is the offset between the AFM tips due to possible misalignment. For the triangular cantilevers used in this work, extra care was needed to position the AFM tip near the middle line of the reference cantilever to avoid errors owing to the torsional bending [111]. The L_{ref} of our reference cantilever is purposely chosen to be 480 μm; thus the error coming from the tip alignment is negligible. Furthermore, to properly calibrate the unknown stiffness k of a cantilever using this method, the following condition needs to be satisfied:

$$0.3k_{ref} < k < 3k_{ref}$$

According to the manufacturer's data, the spring constants of our cantilevers are approximately 0.06 N/m, which satisfies the condition above. The measured force can be calculated by $F = k \cdot m \cdot D_{\text{lever}}$, where *m* is the optical sensitivity of the AFM cantilever and D_{lever} is the deflection signal of the cantilever recorded by the photodiodes of AFM. The optical lever sensitivities (*m*) are calibrated in solution on gold substrate during the force measurements. All calibrated *w* and *k* of the used cantilevers are listed in **Table B.2**.

3.3.2 UV melting and circular dichroism for AFM substrates

Duplexes used in AFM measurements were analyzed by UV melting and circular dichroism (CD). The samples were prepared by annealing appropriate complementary oligos at concentrations of 1.38 μM in 100 mM NaCl, 10 mM phosphate, and 0.1 mM EDTA at pH 7.4. UV absorption changes at 260 nm were acquired by Cary 1E UV-Vis Spectrophotometer, between 25 ℃ to 91 ℃. The temperature was raised at a rate of 0.5 ℃/min. Absorbance values were normalized at the value at 25 ℃. Only the T_m was calculated from the UV melting curves since the substrates were all 30 bp long. CD spectra were acquired on Jasco J-810 Circular Dichroism Spectrometer between 210 nm to 320 nm, with a scanning rate of 200 nm/min and a band width of 1 nm. The samples were placed in 1-cm path length cells at 25 ℃.

3.3.3 AFM measurements

A Veeco Multimode Nanoscope IV AFM was used to perform single molecule force spectroscopy. During the measurements, the approaching/retracting tip velocity was kept at 29.1 nm/s. A total of 4,096 data points were acquired for each approaching–retracting cycle. To avoid multiple pick-ups during the experiment, we intentionally reduced the density of DNA distribution on the surface; thus, the successful DNA pick-up rate by the

tip was less than 10% from approximately 6,000 force–distance curves. Occasionally, multiple pick-ups of DNA did occur. Only force–distance curves clearly showing one DNA stretching were analyzed and reported. The same measurements were performed with ss DNA molecules as experimental controls (**Figure B.3** and **Table B.3**–**5**). For each sequence (with and without rGMP intrusions) the experiments were repeated for two to three different samples, and for each sample between 50 and 130 force curves were acquired pulling different DNA molecules present on the sample surface. We also performed AFM imaging of DNA molecules deposited on the gold surface (**Figure B.2**).

3.3.3.1 Selection and calibration of AFM force–distance curves

We only consider force-distance curve showing only one DNA pick-up by the tip. We do not use data showing multiple DNA pick-ups. Typical force curves showing multiple DNA pick-ups are shown in **Figure B.4**. During the data acquisition, the force-distance curve is recorded as cantilever deflection vs. distance moved by the piezo scanner, i.e. Δz_{lever} vs. z_{piezo}, as shown in **Figure B.5a**. This is not the real separation distance, *d*, between the AFM tip apex and the gold surface. To accurately determine the tip-surface distance, *d*, the cantilever deflection Δz_{lever} has to be subtracted from z_{piezo}, i.e., $d = z_{piezo} - \Delta z_{lever}$ [112]. In addition, the cantilever deflection is assumed to be zero when the tip is far away from the surface and is used to offset the whole force curve. The calibrated force-distance curve is shown in **Figure B.5b**.

3.3.3.2 Determination of L_0, δ, and F_{st}

The calibrated force-distance curve during tip retracting shown in **Figure B.5b** is now presented in **Figure B.6**. As indicated by two red solid lines, we performed linear fitting to two linear sections of the retracting curve to determine the position when the DNA begins to be stretched. The intersected point of the fitted lines is determined to be the position where the stretching of DNA begins. The sudden jump in the force curve

indicates when the DNA is suddenly detached from tip, i.e. when the bonding between streptavidin and biotin is broken. Then, the cantilever goes back to its initial position when there is no force, and the deflection of becomes zero again. The initial contour length L_0, extension δ of DNA, and the stretching force F_{st} can then be determined from the force curve and are used to calculate the stretch modulus by using $S = F_{st} \cdot L_0 / \delta$.

3.3.3.3 Removal of outliers

We used Peirce's criterion to perform the outlier test [113, 114]. Using Peirce's criterion, multiple outliers can be removed. **Table B.6** summarizes the number of outliers that was excluded to obtain the final data sets for statistical consideration. Only as much as 6% of total data were removed.

3.3.3.4 Gaussian fitting

The best Gaussian fit to the data was obtained using the following equation:

$$y = y_0 + Ae^{-\frac{(x-x_c)^2}{2w^2}}$$

where the offset y_0 is set to 0 during the fitting procedure. The value of x_c is the peak position of the Gaussian distribution.

3.3.4 MD simulations

MD simulations were performed by using an in-house Fortran code. Energy and atomic forces were calculated by using the potential energy and parameters of the Amber force fields parmbsc0. The in-house MD code implements periodic boundary conditions, the Verlet algorithm to integrate the equations of motion, the Ewald method to calculate Coulomb interactions and forces, the Nose–Hoover thermostat and Parrinello–Rahman barostat methods to control temperature and pressure of the system, standard routines to calculate short-range energy and force contributions, and Message Passing Interface

instructions to run simulations on parallel computer clusters [115-117]. Initial structures consisted of 10-bp duplexes with the standard B-DNA geometry immersed in a tetrahedral box containing 1,535 water molecules and 20 sodium cations to neutralize the whole system. A duplex is oriented and periodic along the z-axis, and the dimensions of the simulation box are about 38 Å × 38 Å × 34 Å. The systems were first optimized and then equilibrated for about 1 ns at a temperature and pressure of 300 K and 1 atm, respectively, by using an isothermic, isobaric ensemble. Simulations were then extended for about 20 ns in the microcanonical ensemble. In this last step, temperature and pressure remained close to 300 K and 1 atm, respectively, and the MD trajectories were used for the structural analyses. Further technical details and applications of our Fortran MD code can be found in our previous work [115, 118].

3.3.5 NMR

NMR experiments were performed on a Bruker Avance 600 spectrometer, equipped with a 5mmQXI ^{1}H, ^{31}P, ^{13}C, ^{15}N probe (Bruker). Acquisition and processing parameters are similar to those described in our earlier studies [119] with the following variables.
For experiments in D2O: NOESY spectra (2k × 600) were collected with mixing times of 75 ms, 125 ms, and 250 ms and a relaxation delay of 4 s. COSY experiments (2k × 1200) were run with ^{31}P decoupling, a 2 s relaxation delay, and zero filled to (4k × 4k). ^{1}H–^{31}P correlation (HPCOR) [120] experiments (2k × 400) with a sweep width of 9 and 12 ppm for ^{1}H and ^{31}P, respectively, used a relaxation delay of 2 s. For water experiments, a 1-1 jump and return and a 1-1 jump and return NOESY (2k × 400) with a 150 ms mixing time were used with a 1 s delay. Assignment and integration of 2D spectra were carried out using SPARKY 3.33 [121]. The phosphodiester signals were assigned based on their correlation to assigned H3′ and H4′ protons. ^{1}H and ^{31}P were referenced to internal DSS and external 85% H_3PO_4 (capillary in D_2O), respectively.

3.3.5.1 UV melting for NMR substrates

Thermal denaturation curves for an rGMP-containing 9-bp duplex, ATGGArGCTC (with rGMP III) and its DNA-control were obtained on a Cary UV-Vis Spectrophotometer. The duplexes were prepared in 100 mM NaCl, 10 mM phosphate, 0.5 mM EDTA at pH 6.6. Melting temperatures (T_m) were derived from a six-parameter fit of the melting curves for a series of duplex concentrations ranging from 2 to 20 μM [122]. Enthalpy and entropy values were then calculated from a linear fit of the van't Hoff plot.

3.4 Results and Discussion

In order to investigate the effect of rNMP intrusions in DNA, we have performed elastic measurements on 30-bp ds DNA molecules with two different sequences, Sequence 1 and Sequence 2, as shown in **Figure 3.1c**, where the rNMP intrusions were always introduced at the bases of guanosine, dG of the corresponding DNA molecule. These intrusions are therefore called riboguanosine (rGMP) or rG. Riboguanosine is the most frequently incorporated rNMP by DNA polymerases *in vitro* [14, 42], and it is well recognized by ribonuclease H type 2 (RNase HII/2) and mismatch repair if mispaired in DNA or by RNase HII/2 and nucleotide excision repair if paired both in *E. coli* and *S. cerevisiae* cells (Koh *et al*., unpublished) [58]. For this reason, we incorporated rGMP into our sequences. So far, no data exist in the literature to the best of our knowledge on the effect of any rNMP intrusions on the elasticity of DNA. Therefore, to start our studies we have chosen two sequences with the only constraint that they did not present any self-complementarity, which could produce hairpin loops. In particular, Sequence 2 derives from the yeast *S. cerevisiae* genome and is a sequence we utilize to study the impact of rNMPs *in vivo* [55]. With the concern that the mechanical alterations caused by a single rGMP embedded in a 30-bp ds DNA molecule could be below AFM detection capacity, in both sequences, an rGMP was introduced every four to six nucleotides and in only one

of the two strands (**Figure 3.1c**). MD simulations and NMR have instead been performed on segments of these two sequences because the diverse techniques dictate different lengths of the investigated DNA molecules.

The elastic properties of micrometer-long ds DNA have been studied extensively in the last two decades using AFM [123], magnetic [124], and optical tweezers [125]. The stretch modulus of long ds DNA with a few thousand bp is found to be approximately 1000 pN [123-125]. Recently, ds RNA with a similar length has also been investigated using these techniques; experiments using AFM and magnetic tweezers have found that elasticity of micrometer-long ds RNA can be 10% to 20% larger than that of ds DNA [126, 127]. AFM, in particular, is a powerful technique to study the elasticity of nano-systems [112, 128], and ds DNA shorter than hundred nanometers has been investigated using AFM [129-132]. Interestingly, recent studies about the elastic properties of ds DNA using AFM [130, 132], X-ray diffraction [133, 134], and fluorescence resonance energy transfer (FRET) [134] techniques have consistently found that ds DNAs are much more elastic than the micrometer-long ones on the nanoscale. For ds DNAs shorter than 150 bp, their stretch moduli are found to be about 100 pN, an order of magnitude smaller compared to that of few thousand-bp-long ds DNA. The difference in elastic properties of ds DNA of different lengths cannot be explained by the classical Worm-Like-Chain model [130, 133, 134], which was successfully applied to describe the mechanical properties of micrometerlong ds DNA under strain [124, 125, 135]. However, all the DNA and rG-DNA molecules used here are about 10 nm long, and for DNA molecules shorter than about 50 nm, the WLC model is not appropriate, as already discussed in previous studies [130, 133]. Although the origin of ds DNA softening on the nanoscale is not clear, it is hypothesized that the base-pair breathing of the DNA chain is one possible cause [134, 136]. Further discussion about this phenomenon is beyond the scope of this paper, interested readers are referred to the aforementioned references for further details.

To study how the embedded rNMPs alter the mechanical properties of DNA, we used AFM-based single molecule force spectroscopy to stretch two individual short DNA duplexes attached between an AFM tip and a gold surface. Thermodynamic data of duplex formation showed that the rNMPs are tolerated well in a DNA duplex and do not result in a marked alteration of the duplex stability (**Table B.7**), demonstrating that the duplexes are suitable for AFM studies. Circular dichroism (CD) spectra were collected for the investigated two 30-bp sequences of DNA with and without rNMP intrusions. All these DNA molecules showed a typical conservative spectrum with a positive and a negative peak at 280 nm and 250 nm, respectively, indicative of a B-form helical structure (**Figure B.7**). During the AFM measurements, as illustrated in **Figure 3.2a**, at one end, both the DNA strands were covalently anchored on a gold substrate through thiol–gold chemistry, while the strands at the other end were attached to the AFM tip via streptavidin–biotin bonding [132]. The elastic properties were then investigated by stretching DNA using the AFM tip (**Figure 3.2a**). For each sequence, we compared the stretch modulus of DNA containing rNMPs with the modulus of the corresponding DNA sequence without rNMPs. **Figure 3.2b** shows an AFM image of anchored rG-DNA molecules of Sequence 1 on the gold surface (see **Figure B.2** for images of other DNA molecules). The height profile shown in **Figure 3.2c** indicates that DNA molecules on the surface have heights of a few nanometers, a value which corresponds to the length of these molecules, proving that they are standing up and not lying on the surface [132]. To avoid multiple DNA pick-ups during the pulling experiment, we have purposely reduced the distribution of DNA density on the surface for each measurement. In addition, the measurements were performed in a buffer solution of 100 mM Na^+, which gave rise to a Debye length of 1 nm from the surface and occasionally resulted in repulsive force between the tip and the gold substrate at short separation distance. During the pulling measurements, the AFM tip is first brought into gentle contact with the gold substrate to

pick up a DNA molecule through streptavidin–biotin interaction. When the AFM tip is retracted from the surface, the picked DNA molecule is extended to its natural contour length L_0, and then further stretched to $L_0 + \delta$ until the streptavidin–biotin bonding is broken (**Figure 3.2a**,**d**). The typical binding force for a streptavidin–biotin bond is about 100 pN [137] for the tip velocity used in this experiment, which is an order of magnitude weaker than the covalent thiol–gold binding force, approximately 1.4 nN [138]. This significant difference in the binding force magnitude ensured that the DNA molecules could be repeatedly stretched by the AFM tip and were not plucked away from the surface during the experiment. Sometimes multiple DNA pick-ups did occur (**Figure B.4**); thus, we discarded such events and used only force–distance curves showing a single DNA pick-up for data analysis. Typical force vs. tip-substrate distance curves during the tip approaching and retraction are shown in **Figure 3.2d** and schematically illustrated in **Figure 3.2e**. Details about force curve calibration can be found in **Materials and Methods**, **Figure B.5**, and the literature [139-143]. In the retracting force curve (red in **Figure 3.2d**), it is possible to observe that the tip has to overcome an initial adhesion force F_{adh} to detach from the substrate. Once the tip is out-of-contact from the substrate, further retraction of the tip extends the DNA to its natural contour length L_0. During this elongation, no force is detected. However, when the tip moves further up, the DNA is stretched to $L_0 + \delta$ and simultaneously a sudden increase of the force that pulls the tip downward towards the substrate is detected (**Figure 3.2d**,**e**). After the bond between streptavidin and biotin abruptly breaks, the cantilever jumps back to its zero-force position, corresponding to zero-cantilever-bending. The difference in force magnitude between the point where the DNA stretching is at a maximum and the point where the DNA detaches from the AFM tip is defined as the stretching force F_{st} exerted on the DNA. Details about determination of all the parameters are shown in **Materials and Methods** and **Figure B.5**. Finally, by the definition of stretch modulus S, we obtain $S = F_{st} \cdot L_0 / \delta$ [123, 132].

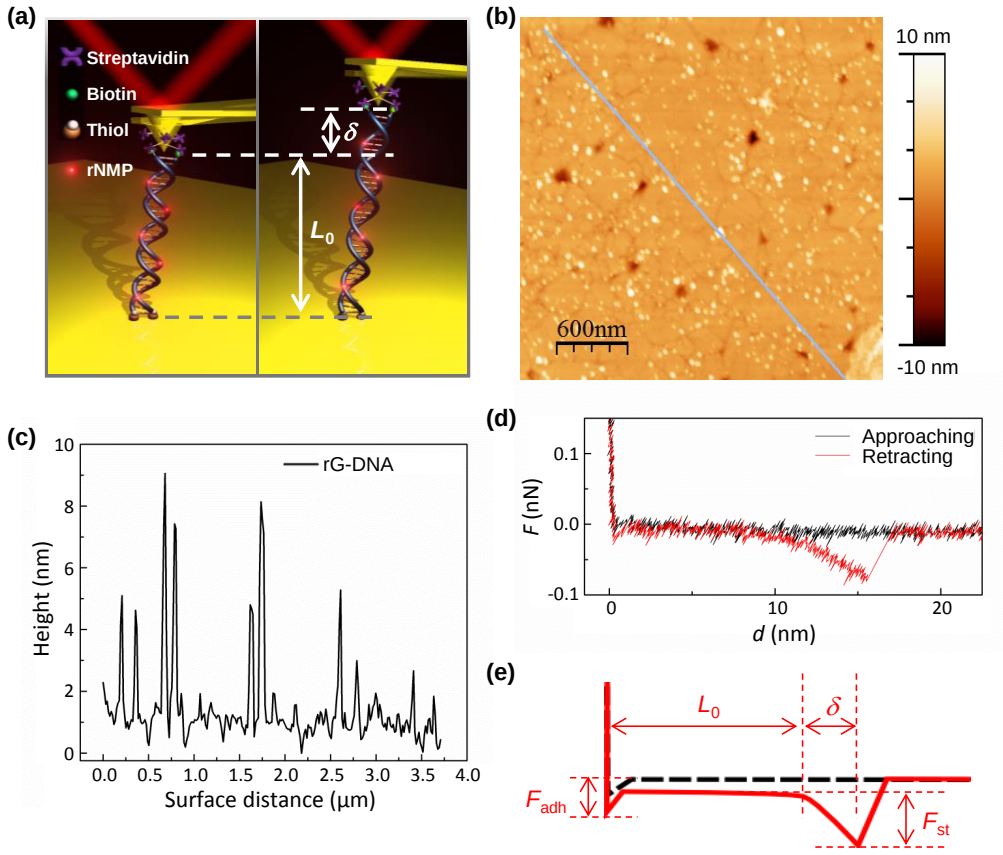
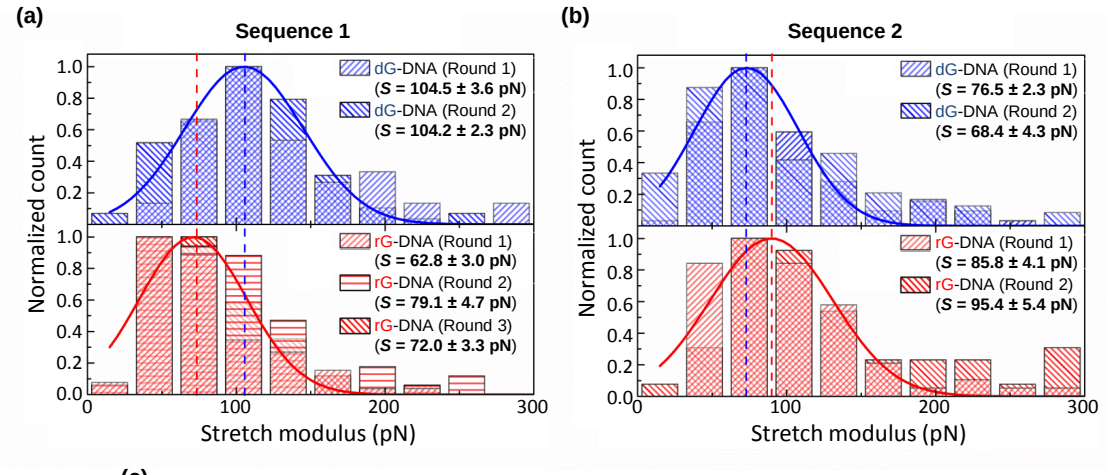


Figure 3.2 DNA stretching experiments using AFM. (**a**) Schematic view of a DNA molecule with rNMP intrusions during AFM stretching. The DNA biotinylated ends interact with the streptavidin-modified tip, while the DNA thiolated ends are attached to the Au surface. (**b**) AFM topographic image of Sequence 1 ds DNA molecules containing rGMP intrusions deposited on an Au surface. (**c**) Height profile of the cross section (blue line in **b**) of the topographic image shown in **b**. (**d**) Typical force–distance curves acquired during AFM force measurements after calibration. (**e**) Schematic of force–distance curves for the determination of L_0, δ, and F_{st}.

The histograms of the measured stretch moduli of ds DNAs with Sequences 1 and 2 (with and without rGMPs) are presented in **Figure 3.3a** and **b**, respectively. In all histograms, the magnitude of the peak (the maximum number of elastic measurements performed for

a given sample) was normalized to 1 for clarity when comparing different measurements. The exact number of performed measurements is reported in **Table B.6**, and it ranges between 50 and 130 per sample. For each sequence, we plotted the histograms of the stretch modulus obtained for the ds DNA molecules with (in red) and without (in blue) rGMP intrusions, shown in the top and the bottom panels, respectively. Individual histograms for each ds DNA molecule are shown in **Figure B.8**–**10**. All the histograms show typical Gaussian distributions in which the peak position can be obtained directly from the Gaussian fit. For each sequence, the solid lines are the best Gaussian fit to all the combined data obtained from different measurements and samples. The parameters L_0, δ, and F_{st} used for the calculation of S are summarized in the tables presented in **Figure 3.3c**,**d**. See **Table B.8**–**10** for mean and median values and **Table B.11**–**14** for summary of detailed statistical analysis. Interestingly, for Sequence 1 the peak position of all the combined data indicates that the stretch modulus in the presence of rGMPs (71.1 ± 3.5 pN) is 32% lower than the modulus in the absence of intrusions (105.3 ± 2.0 pN) (**Figure 3.3a**,**c**). The presence of rGMPs in DNA is thus softening the DNA for Sequence 1. On the other hand, for Sequence 2 (**Figure 3.3b**,**d**) the Gaussian distribution of the data corresponding to ds DNA in the presence of rGMPs is very similar to the distribution of the data in the absence of rGMPs, and the peak position of the modulus is even slightly shifting towards larger values (stiffening) in the presence of rGMP intrusions, precisely from S = 73.2 ± 2.8 pN without rGMPs to 89.9 ± 3.8 pN with rGMPs. These results demonstrate that rGMP intrusions in DNA can substantially decrease the stretch modulus of DNA, as in the case of Sequence 1, and that this effect is not a general alteration caused by rGMPs in DNA. In fact, rGMPs in Sequence 2 cause only a minor perturbation of DNA elasticity, and even in the opposite direction, inducing a mild increase of the modulus. Since both sequences have the same number of rNMPs of the same base, rG, the measurements suggest that very different effects are likely depending on the position of the rNMPs and the sequence context. We point out that the

stretch moduli of the ds DNA without rGMP intrusions are both around 100 pN, which is consistent with those obtained using different techniques [133, 134], for both Sequences 1 and 2. Moreover, the stretch modulus of ds DNA without rGMP intrusions is larger for Sequence 1 than for Sequence 2, owing to the previously shown sequence-dependent effect of DNA elasticity [144, 145].



(c)

Sequence 1	Gaussian Peak				Median	
	L_0 (nm)	δ (nm)	F_{st} (pN)	S (pN)	S (pN)	
dG-DNA (n=157)	9.0 ± 0.2	2.1 ± 0.4	27.1 ± 0.8	105.3 ± 2.0	109.3 (99.3 – 121.4)	p < 0.0001
rG-DNA (n=251)	11.2 ± 0.2	3.8 ± 0.1	28.6 ± 0.1	71.1 ± 3.5	76.5 (67.2 – 86.1)	

(d)

Sequence 2	Gaussian Peak				Median	
	L_0 (nm)	δ (nm)	F_{st} (pN)	S (pN)	S (pN)	
dG-DNA (n=186)	9.9 ± 0.3	2.8 ± 0.1	22.5 ± 0.3	73.2 ± 2.8	84.1 (75.0 – 95.2)	p < 0.0001
rG-DNA (n=139)	10.3 ± 0.3	1.6 ± 0.4	31.3 ± 0.2	89.9 ± 3.8	107 (92.4 – 126.6)	

Figure 3.3 Stretch moduli of investigated DNA without and with rGMPs. (**a**) Histograms of the stretch modulus distributions of Sequence 1 DNA without rGMPs (dG-DNA, blue) (2 repeats) and with rGMPs (rG-DNA, red) (3 repeats). (**b**) Histograms of the stretch modulus distributions of Sequence 2 DNA without (2 repeats) and with (3 repeats) rGMPs. Solid lines are the best Gaussian fit for all the combined data. (**c** and **d**) Summary of values L_0, δ, and F_{st} extracted from the force–distance curves and used to obtain the stretch modulus S for Sequences 1 and 2, respectively. The values are obtained from a Gaussian fit and represent the peak position ± the standard error from the Gaussian fit.

The stretch modulus is also shown as the median value with 99% confidence interval of the median in parentheses. The value n denotes the total number of force measurements used for fitting. *P* values, comparing the stretch moduli of dG-DNA and rG-DNA for each sequence, are shown. The Mann–Whitney *U*-test was utilized to obtain P values.

In order to gain some molecular insight into the origin of the elastic properties of modified and control DNA oligomers, we performed MD simulations using all the different rGMPs (I–X) of the DNA sequences used in the AFM measurements (**Figure 3.1c**). Our simulations showed that three of the five rGMP intrusions in DNA of Sequence 1 are able to induce local structural distortions involving the rGMP and/or the following nucleotide in the 3′ direction (**Figure 3.4a**,**b**). In the particular case of sequence CrGATGGArGCT (**Figure 3.4c**), the two rGMPs (II and III) are both sandwiched by C and A nucleotides, and MD simulations show that the decamer undergoes significant local distortions (see alpha and gamma torsional angles of the sugar-phosphate backbone relative to average values in control DNA in **Figure 3.4a**,**b**) in correspondence of the nucleotide on the 3′ side of each rGMP. Inspection of the MD trajectories suggests that these distortions arise from the formation of a hydrogen bond between the hydroxyl group of an rGMP and neighboring electronegative sites of either the backbone or the vicinal base in the 3′ direction (**Figure 3.4d**–**g**). A similar local distortion was found also in the case of the decamer with sequence GrGTTCArGGTT for rGMP I (**Figure 3.4b**). Although deriving conclusive results from our MD simulations is challenging, the MD runs show nonetheless that rNMPs are capable of triggering the occurrence of local distortions having lifetimes of the order of nanoseconds (**Figure B.11**), thereby suggesting that these local distortions at rGMP intrusions might be at the origin of different elastic properties of modified and control DNA oligomers. In the past, it has been shown that the structural distortion of the sugar-phosphate backbone of nucleic acids can substantially influence DNA flexibility [146, 147]. Thus, the significant local

structural distortions due to rNMP intrusions in the DNA chain found by MD simulations likely alter the elasticity of DNA molecules in the presence of rNMPs.

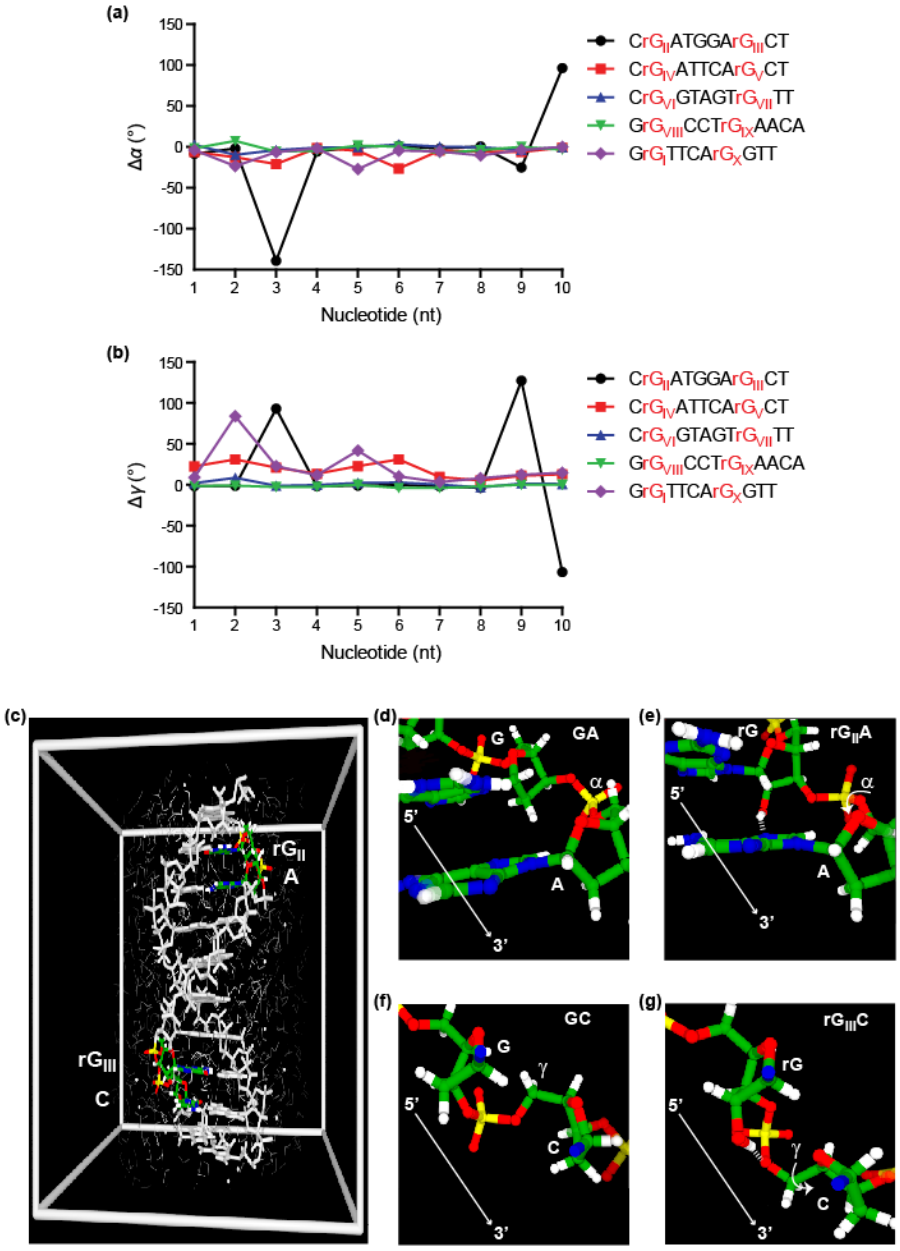


Figure 3.4 Molecular dynamics simulation of DNA sites with an rGMP. Mean deviation of the (**a**) alpha (α) torsional angles and (**b**) gamma (γ) torsional angles from expected values of corresponding DNA-only sequences across all rGMP segments (I to X) of Sequences 1 and 2 modeled via MD has been calculated; averages have been

performed by taking instantaneous values between the 5th and the 15th ns of each MD run. Black circles, red squares, blue triangles, green inverted triangles, and purple rhombus present data for the shown sequences with rGMP II and III, IV and V, VI and VII, VIII and IX, and I and X, respectively. rGMPs I–V and VI–X are from Sequence 1 and Sequence 2, respectively. (**c**) Illustration of the CrGATGGArGCT duplex with rGMP II and III, showing in colors the (top) 5′-rGA-3′ and (bottom) 5′-rGC-3′ dinucleotides, respectively. O, C, N, H, and P atoms are shown in red, green, blue, white, and yellow colors, respectively; the rest of the duplex is displayed in grey. The white box indicates a periodic unit of an infinite oligomer used in simulations. (**d**) Zoom-in of a 5′-GA-3′ dinucleotide exhibiting regular values of α and γ. (**e**) Illustration of the 5′-rGA-3′ dinucleotide (with rGMP II) in a distorted conformation with α and γ deviating by about 200 ° and 100 °, respectively; this conformation appears to be stabilized by the formation of a hydrogen bond (white dashed line) between the 2′-hydroxyl group of rG and the N7 site of adenine in the 5′-rGA-3′ dinucleotide. (**f**) Zoom-in of a 5′-GC-3′ dinucleotide showing regular values of α and γ. (**g**) Illustration of the backbone region of the 5′-rGC-3′ dinucleotide (with rGMP III) showing the occurrence of a metastable local distortion involving significant deviations of both α and γ. Also in this case, the local distortion is accompanied by the formation of a hydrogen bond between the 2′-OH group of rG and an O atom of the phosphate group. Water molecules and Na^+ ions are not shown for clarity in **d**–**g** while bases are also not shown in **f** and **g**.

To further probe the structural impact of a single rGMP embedded in DNA, NMR spectroscopy was performed on three selected segments from Sequence 1 and Sequence 2 of the AFM study. All three segments chosen for NMR studies contain a single rGMP embedded at the 6th or the 5th position of a 9-bp DNA duplex (**Figure 3.5**). The rG is tolerated well in the DNA duplex and does not appreciably affect duplex stability (**Figure 3.5a**,**c**,**e** and **Table B.15**). All three rGMP-containing 9-bp duplexes and their DNA-

controls exhibit a characteristic B-form helical structure. Additionally, imino ^{1}H NMR spectra (**Figure 3.5** and **Table B.16**) reveal that all base-pairs are formed, including the rG:C base-pair, with chemical shift perturbation localized to the rG:C and neighboring base-pairs (**Figure 3.5a**,**c**,**e**).

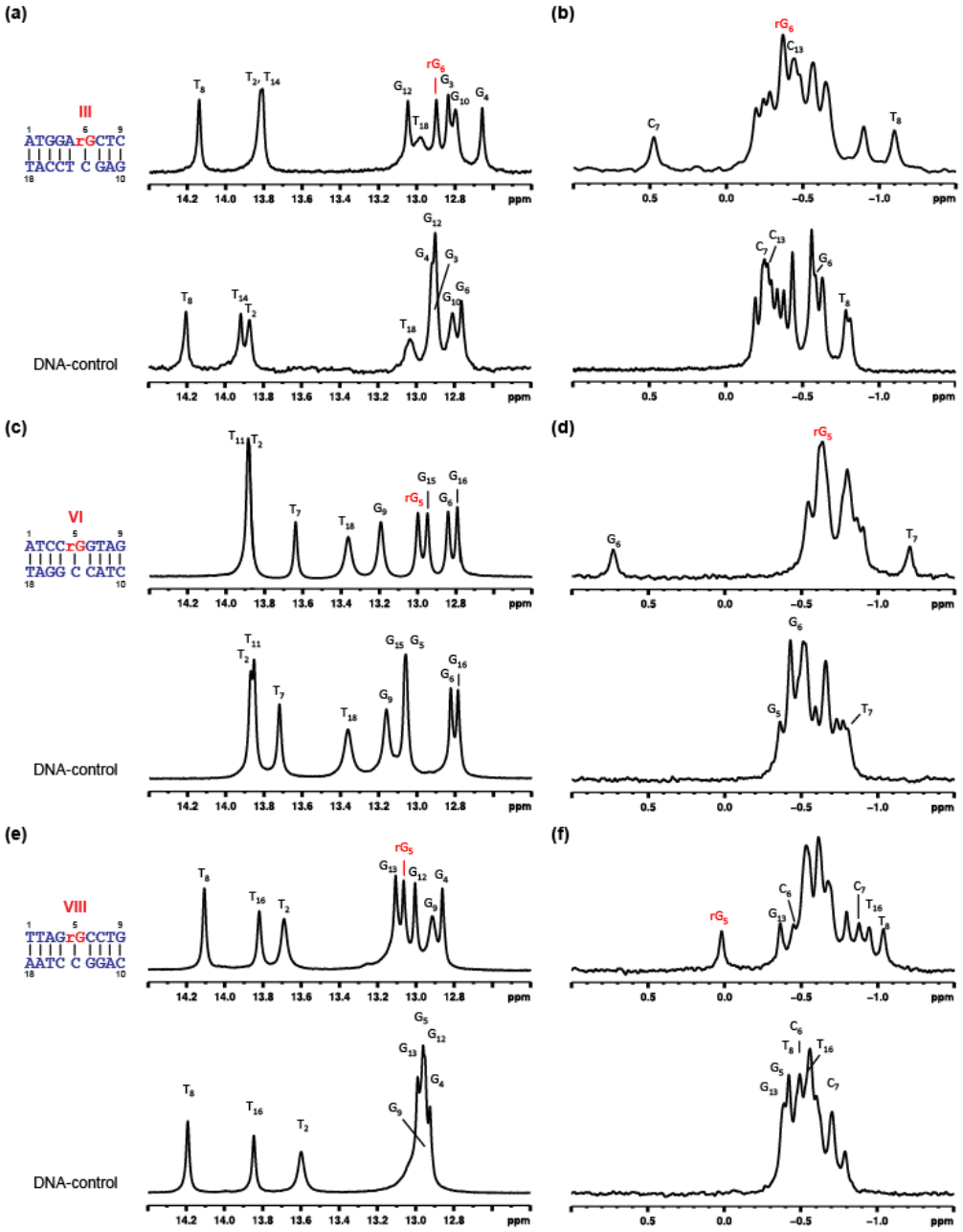


Figure 3.5 Structural perturbation caused by an rGMP embedded in a DNA duplex as observed by ^{1}H and ^{31}P NMR. (**a**) Imino ^{1}H NMR spectra for the 5′-ATGGArGCTC-3′ duplex containing rGMP III of Sequence 1 (top) and its DNA-control (bottom) at 280 K in 100 mM NaCl, 10 mM phosphate, 10% D2O buffer at pH 6.4. (**b**) ^{31}P NMR spectra of the 5′-ATGGArGCTC-3′ duplex (top) and its DNA-control (bottom) recorded at 294 K. (**c**) Imino ^{1}H NMR spectra for the 5′-ATCCrGGTAG-3′ duplex containing rGMP VI of Sequence 2 (top) and its DNA-control (bottom) at 280 K under the same buffer conditions. (**d**) ^{31}P NMR spectra for the 5′-ATCCrGGTAG-3′ duplex (top) and its DNA-control (bottom) recorded at 294 K. (**e**) Imino ^{1}H NMR spectra for the 5′-TTAGrGCCTG-3′ duplex containing rGMP VIII of Sequence 2 (top) and its DNA-control (bottom) at 280 K under the same buffer conditions. (**f**) ^{31}P NMR spectra of the 5′-TTAGrGCCTG-3′ duplex (top) and its DNA-control recorded (bottom) at 294 K.

The NMR study focused on the phosphodiester backbone of the duplexes. The phosphorous chemical shifts depend on the environment; the major determinants of the ^{31}P chemical shift are the alpha and zeta torsion angles [147]. Typically, B-form DNA phosphorous resonances are confined to a narrow shift window (~0.6–0.8 ppm) as seen for the DNA-controls. The presence of a single rGMP results in local perturbations to the duplex backbone. This localized perturbation is limited to the nearest and next nearest base-pairs and is consistent with a recent study of an rGMP-containing dodecamer sequence [75]. In contrast to this study, however, due to the non-self-complementary design of our duplexes with a single rGMP on only one strand, we observed an asymmetric 3′ perturbation of the duplex primarily on the rGMP-containing strand (**Figure 3.5b**,**d**,**f** and **Table B.17**).

In two of the three rGMP-containing duplexes, 5′-ATGGArGCTC-3′ (rGMP III) and 5′-ATCCrGGTAG-3′ (rGMP VI), the rG phosphorous resonance showed relatively little

deviation <0.25 ppm from the DNA-control, while the phosphorous peaks following rG were shifted downfield by 0.80 and 1.28 ppm, respectively (**Figure 5b**,**d** and **Table B.17**). Additionally, the next to nearest neighboring phosphorous resonances 3′ of the rG experienced an upfield shift of 0.27–0.29 ppm. This is indicative of a distortion of the backbone, localized 3′ of the rG base in these sequence contexts. The other NMR sequence, 5′-TTAGrGCCTG-3′ (rGMP VIII), exhibited a different trend; in this duplex, the ^{31}P NMR spectrum of the rGMP-containing oligonucleotide appears less perturbed, the furthest downfield shifted phosphorous resonance corresponded to rG_5 while the phosphodiester on the 3′ side of rG is essentially unaffected (**Figure 3.5f**). Taken together this means that the resulting backbone distortions are not the same, highlighting the importance of the sequence context previously mentioned in the AFM study.
Although two of the three selected NMR duplexes contain a purine-rG-pyrimidine motif (5′-ArGC-3′, rGMP III and 5′-GrGC-3′, rGMP VIII) and would be expected to have similar stacking interactions, they exhibit strikingly different ^{31}P spectra (**Figure 3.5b**,**f**). Interestingly, in the MD trajectory of the 5′-ArGC-3′ (rGMP III) sequence, we observe a hydrogen bond between the 2′-OH and the phosphate group (**Figure 3.4g**). Such a hydrogen bond may simultaneously dampen the dynamics and change the local environment and rationalize the different behavior of the 5′-ArGC-3′ (rGMP III) sequence context.

3.5 Conclusions

In summary, we have studied the elastic properties and structure of two short ds DNAs with and without rGMP intrusions, employing a combined experimental and theoretical approach on different segments of these two sequences. AFM-based single molecule force measurements showed that, depending on the DNA sequence and/or the specific positions of the rGMP intrusions in the sequence, rGMPs can dramatically decrease (up to 32% for the sequences used here) or slightly increase thestretch modulus of ds DNA.

Snapshots of MDsimulations reveal sequence-dependent local structure alterations of the torsion of the DNA backbone caused by the intrusion of rGMPs in the DNA chain. The major alterations identified by MD simulation involve the rGMP and the nucleotide 3′ from the rGMP. Consistent with MD simulations, NMR spectra demonstrate that even a single rGMP can substantially alter the local sugar-phosphate backbone, and major alterations also involve the rGMP and nucleotide 3′ from the rGMP. Our findings point towards a marked effect in the elastic properties and structure of ds DNA possibly played by the sequence context in the immediate vicinity of the embedded rNMPs, at the nucleotide 3′ to the rNMP sites. It is reasonable to think that the nucleotide to the 3′ side of the rNMP is the most altered in the structure because it is the closest nucleotide to the 2′-OH group of the rNMP (**Figure 3.1b**, **3.4**, and **3.5**). The combined theoretical and experimental approach accomplished here opens a new route to understand how rNMP intrusions, at which sites and densities, can modify the structural, physical, and mechanical properties of DNA, and ultimately change its chemical and biological functions. Overall, our results reveal a complex effect of rNMPs on DNA elastic properties, the direction and the impact of which can be determined for each specific sequence via AFM. Only a high throughput and systematic analysis of multiple DNA sequence contexts with rNMPs can help to elucidate the rules on how rNMPs alter DNA mechanical properties. Furthermore, this study shows that DNA elasticity could be modulated by means of rNMP inclusions for a variety of applications in nanobiotechnology.

3.6 Acknowledgments

We thank N. Hud for assistance of CD spectra acquisition, A.Spring for providing the sugar puckering data and assignments for the control DNA sequence, R. Wartell for assistance with UV-melting experiment and suggestions, and L. Williams and D. Dunlap for detailed conversations. H.C.C. and E.R. acknowledge the support of the Office of

Basic Energy Sciences of the US Department of Energy (DE-FG02-06ER46293). A.L.L., E.R., and A.B. thank the support of the National Science Foundation (NSF) (CMMI-1100290 and DMR-0820382). A.L.L. and A.B. acknowledge the support of the Samsung Advanced Institute of Technology and the NSF grant CHE-0946869. K.D.K., F.S., and E.R. are grateful for the support by the Integrative Biosystems Institute grant IBSI-4; K.D.K. and F.S. also thank the support from the Georgia Cancer Coalition grant R9028 and the NSF grant MCB-1021763.

CHAPTER 4

RIBOSE-SEQ: GLOBAL MAPPING OF RIBONUCLEOTIDES EMBEDDED IN GENOMIC DNA

The study in Chapter 4 consists of the work published in *Nat. Methods* 12 (2015) 251–257.

Koh, K.D.[1], Balachander, S.[1], Hesselberth, J.R.[2], and Storici F.[1]

[1]School of Biology, Georgia Institute of Technology, Atlanta, GA 30332

[2]Department of Biochemistry and Molecular Genetics, University of Colorado, Anschutz Medical School, Aurora, CO 80045

4.1 Abstract

Abundant ribonucleotide incorporation in DNA during replication and repair has profound consequences for genome stability, but the global distribution of ribonucleotide incorporation is unknown. We developed ribose-seq, a method for capturing unique products generated by alkaline cleavage of DNA at embedded ribonucleotides. High-throughput sequencing of these fragments in DNA from the yeast *Saccharomyces cerevisiae* revealed widespread ribonucleotide distribution, with a strong preference for cytidine and guanosine, and identified hotspots of ribonucleotide incorporation in nuclear and mitochondrial DNA. Ribonucleotides were primarily incorporated on the newly synthesized leading strand of nuclear DNA and were present upstream of (G+C)-rich tracts in the mitochondrial genome. Ribose-seq is a powerful tool for the systematic profiling of ribonucleotide incorporation in genomic DNA.

4.2 Introduction

Genomic DNA contains embedded ribonucleotides (rNMPs) that are incorporated during DNA replication and repair or formed during DNA damage (reviewed in [148]). The modifications have been linked to genome instability and disease, but no method currently exists to profile their locations genome wide.

rNMPs were initially found at specific DNA loci in mouse and human mitochondrial DNA [149] and the mating type locus of fission yeast [4], but they have since been detected in a variety of cell types [150]. Many DNA polymerases can incorporate rNMPs into DNA, including the human replicative DNA polymerase (Pol) δ [14] and mitochondrial Pol γ [19], budding yeast nuclear replicative Pol α, δ, and ε [21], *Escherichia coli* polymerase V [12], and the polymerase components of bacterial nonhomologous end joining ligases [151]. rNMP incorporation could also be a consequence of incomplete maturation of Okazaki fragments during lagging strand

synthesis in DNA replication [28]. Moreover, generation of hydroxyl radicals during oxidative stress can modify DNA deoxyribose sugars to ribose, forming rNMPs in DNA both *in vitro* and *in vivo* [29].

RNase H type 2 (RNase H2 or HII) cleaves single rNMPs or longer rNMP tracts incorporated in DNA [37] and initiates ribonucleotide excision repair, the main rNMP repair mechanism in bacterial DNA and in eukaryotic nuclear DNA ([35] and references therein). By contrast, RNase H1 (or HI) recognizes only rNMP tracts longer than four nucleotides. Inactivation of RNase H2 leads to the accumulation of high amounts of rNMPs in genomic DNA, enabling >1 million rNMPs to be quantified per mouse embryonic fibroblast genome and suggesting that rNMPs are the most common noncanonical nucleotides in dividing mouse cells [34]. Similar measurements on genomic DNA derived from RNase H2–deficient (*rnh201*Δ) budding yeast estimated a few thousand rNMPs incorporated per genome per cell cycle [32, 33], and RNase HII–null *Bacillus subtilis* cells have high levels of incorporated rNMPs [11]. Embedded rNMPs in DNA have highly reactive 2′-hydroxyl groups, altering its properties, structure, and function [54, 91] and leading to genome instability [32, 39, 40, 62]. In humans, mutations in any of the three subunits of RNase H2 are associated with Aicardi-Goutieres syndrome (AGS), a neurological disorder [64].

Despite abundant evidence for the frequent incorporation of rNMPs in DNA, a comprehensive and detailed picture of rNMP incorporation throughout a genome is lacking. Here we introduce ribose-seq: a technique for mapping rNMPs in genomic DNA.

4.3 Materials and Methods

4.3.1 Yeast strain construction

Yeast strains used in this study are presented in **Table C.1**. Isogenic yeast haploid strains KK-100, KK-174, KK-107 and KK-120 were derived from E134 (*MATα ade5-1 lys2-14A trp1-289 his7-2 leu2-3,112 ura3-52*) [152]. KK-100 was made from E134 by deletion and replacement of *RNH201* via transformation with a PCR product containing the *hygMX4* cassette flanked by 50 nt of sequence homologous to regions upstream and downstream of the *RNH201* ORF. KK-174 was constructed from KK-100 by deletion and replacement of *RNH1* via transformation with a PCR product containing the *kanMX4*. KK-107 was generated by introducing the *pol2-4* mutation into KK-100 via integration-excision using plasmid YIpJB1 [153]. KK-120 was made by introducing the *pol3-5DV* mutation into KK-100 via integration-excision using plasmid p170-5DV [154].

Isogenic yeast haploid strains KK-30, KK-125, KK-164 and KK-170 were derived from FRO-767,768 (*ho*Δ *hml*Δ::*ADE1 MATa-inc hmr*Δ::*ADE1 ade1 leu2-3,112 lys5 trp1*::*hisG ura3-52 ade3*::*GAL*::*HO leu2*::*HOcs mata*Δ::*hisG*) [46]. KK-30 was made from FRO-768 by reversion of *ade3*::*GAL*::*HO* to intact *ADE3* via transformation with a PCR product containing *ADE3*, followed by replacement of *RNH201* with the *hygMX4*. KK-125 was constructed from KK-30 by replacement of *RNH1* with the *kanMX4* cassette. KK-164 was generated from KK-125 by replacement of *UNG1* with the *natMX4* cassette. KK-170 was made by introducing the *pol2-M644G* mutation into KK-30 via integration-excision using plasmid p173-M644G [155].

Isogenic yeast haploid strains KK-158 and KK-159 were derived from FRO-767 and FRO-768 [46]. KK-158 and KK-159 were constructed from FRO-767 and FRO-768 by replacement of *UNG1* with the *hygMX4* cassette.

4.3.2 AtRNL ligation assay

rNMP-containing DNA oligonucleotide (oligo) Lig.47.R (see **Table C.2**) and its DNA-only control, Lig.47.D, were 5′ end-labeled with [γ-^{32}P]ATP (PerkinElmer) by T4 polynucleotide kinase (New England Biolabs). Alkali treatment was carried out in 0.3 M NaOH for 2 h at 55 ℃. The resulting solution was neutralized and diluted. 100 nM of alkali-treated 5′-radiolabeled products were incubated in 50 mM Tris-HCl, pH 7.5, 40 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 30 μM ATP (Sigma-Aldrich), and 1 μM AtRNL [156] for 1 h at 30 ℃. After dilution, the ligated products and remaining substrates were treated with T5 exonuclease (NEB) for 2 h at 37 ℃. Aliquots were withdrawn after appropriate steps and quenched with 90% formamide. The products were analyzed by 15% (w/v) polyacrylamide, 8 M urea gel electrophoresis (urea-PAGE). 20–100 Oligonucleotide Length Standard (Integrated Device Technology) was used as a ladder. After electrophoresis, gels were exposed to a phosphor screen overnight. Images were taken with Typhoon Trio$^+$ (GE Healthcare) and obtained with ImageQuant (GE Healthcare). Band intensities were quantified by Multi Gauge V3.0 (Fujifilm).

4.3.3 3′ base bias for AtRNL ligation assay

rAMP, rGMP, rUMP, and rCMP-containing DNA oligos (Lig.30.rA, Lig.30.rG, Lig.30.rU, and Lig.30.rC, respectively; see **Table C.2**) were 5′ end-labeled with either hot [γ-^{32}P]ATP (PerkinElmer) or cold ATP (Sigma-Aldrich) by T4 polynucleotide kinase (NEB). Each of the hot rNMP-containing 5′-radiolabeled DNA oligonucleotides was mixed with the other three cold DNA oligonucleotides at equimolar ratios. The mixtures were treated with 0.3 M NaOH for 2 h at 55 ℃, neutralized, and diluted. 100 nM alkali-cleaved products (25 nM of each base) was then incubated in 50 mM Tris-HCl, pH 7.5, 40 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 30 μM ATP (Sigma-Aldrich), and either 1 μM or 200 nM AtRNL [156] for 1 h at 30 ℃. After dilution, the resulting products were treated with T5 exonuclease for 2 h at 37 ℃. Aliquots were withdrawn after appropriate steps, quenched, and analyzed by urea-PAGE.

4.3.4 rNMP bypass assay

A DNA primer oligo, ByPrim (see **Table C.2**), was 5′ end-labeled with [γ-^{32}P]ATP (PerkinElmer) by T4 polynucleotide kinase (NEB). The 5′-radiolabeled primer was annealed to either rCMP- or rUMP-containing template oligonucleotide (ByTemp.rC or ByTemp.rU, respectively). 100 nM annealed substrates was incubated in HF Buffer, 2 mM dNTPs and 0.2 units of Phusion High-Fidelity DNA Polymerase (NEB) for 30 s at 72 ℃. The reactions were quenched and analyzed by urea-PAGE. Bypass probability was calculated as the band intensity at the +1 position plus all longer products divided by the intensity at the −1 position (preceding the rNMP) plus all longer products, as described by Kokoska *et al.* [157].

4.3.5 Double-strand break repair assay with rNMP-containing oligos

Transformations with rNMP-containing DNA oligos LEU2.rG and LEU2.rU (see **Table C.2**) and DNA-only oligos LEU2.D and LEU2.dU were done as described by Storici *et al.* [46]. Cells from each oligonucleotide transformation were plated to selective Leu$^-$ medium. For each transformation, 20 Leu$^+$ transformants were selected. Colony PCR was performed on those transformants, amplifying with primers LEU2.3 and LEU2.6 a 900-bp region in *LEU2* locus where a new StuI restriction site is expected. The resulting PCR products were treated with StuI (NEB) and analyzed by agarose gel electrophoresis to confirm the presence of the StuI restriction site.

4.3.6 Ribose-seq library construction to map rNMPs in DNA

Genomic DNA from *S. cerevisiae* cells grown in liquid rich medium containing yeast extract, peptone and 2% (w/v) dextrose (YPD) for 2 days to stationary phase was extracted following the protocol “Preparation of Yeast Samples” in the Qiagen Genomic DNA Handbook. Genomic-tip 500/G (Qiagen), Genomic DNA Buffer Set (Qiagen),

proteinase K (Qiagen), RNase A (Qiagen) and lyticase (Sigma-Aldrich) were used to extract genomic DNA from *S. cerevisiae* cells. Extracted genomic DNA was digested with SspI, DraI, and EcoRV (NEB) overnight at 37 ℃ to create a population of 500- to 3,000-bp genomic fragments with an average size of ~1.5 kb. Assuming that rNMPs, if present, could be located in any position of each genomic fragment, an average of 1.5 kb allows a reasonable window for rNMP capture. Following confirmation of digestion by Experion Automated Electrophoresis System (Bio-Rad), the fragments were tailed with dATP (Sigma-Aldrich) by exo$^-$ Klenow fragment (NEB) for 30 min at 37 ℃. The resulting products were purified by spin column (Qiagen) and then ligated to preannealed double-strand adaptors (Adaptor.L:Adaptor.S; see **Table C.2**) that contain single dT overhangs and a randomized 8-base unique molecular identifier (UMI) by T4 DNA ligase (NEB) overnight at 15 ℃. The products were purified using AMPure XP beads (Beckman Coulter). All subsequent purifications were done using AMPure XP beads. The adaptor-ligated DNA fragments were incubated in 0.3 M NaOH for 2 h at 55 ℃ to expose 2′,3′-cyclic phosphate and 2′-phosphate termini of DNA at rNMP sites, followed by neutralization and purification. The resulting single-strand (ss) fragments were incubated in 50 mM Tris-HCl, pH 7.5, 40 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 30 μM ATP (Sigma-Aldrich), and 1 μM AtRNL for 1 h at 30 ℃, followed by purification. The products and remaining fragments of DNA were treated with T5 exonuclease (NEB) for 2 h at 37 ℃ to degrade the background of unligated, linear ss DNA, leaving self-ligated ss DNA circles intact. Treatment with 1 μM Tpt1 [156] in 20 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$, 0.1 mM DTT, 0.4% Triton X-100 and 10 mM NAD^+ (Sigma-Aldrich) for 1 h at 30 °C was used to remove the 2′ phosphate remaining at the ligation junction. After purification and resuspension, the libraries were PCR-amplified with one of the barcoded primers, PCR.1.Index1-4, and PCR.2 (see **Table C.2**) using either Phusion High-Fidelity DNA Polymerase (NEB) or EconoTaq DNA Polymerase (Lucigen), confirmed by 6% PAGE, purified, and pooled for analysis by Illumina sequencing. 100-bp DNA Ladder

(NEB) was used as a size standard. SYBR Gold Nucleic Acid Gel Stain (Life Technologies) was used to stain PAGE gels for visualization under UV light.

4.3.7 DNA sequencing

Indexed sequencing libraries were mixed at equimolar concentrations and normalized to 10 nM. Libraries were sequenced on an Illumina MiSeq and 50-cycle single-end reads were collected. Raw sequencing reads are available at NCBI GEO [158] under accession code GSE61464.

4.3.8 Sequence alignment and processing

Reads were aligned to the *S. cerevisiae* genome (sacCer2) with bowtie using two different settings to report uniquely aligning and multiple aligning reads ("-m 1" and "--all", respectively). Aligned reads in BAM format were processed to remove PCR duplicates using umitools (https://github.com/brwnj/umitools/), which filters reads that contain duplicate UMIs and reports reads with unique UMIs. Reads in this study had an eight-base UMI incorporated during ligation, corresponding to the first eight cycles of raw FASTQ sequence. Following UMI removal, read depths at each 5′ position were calculated with BEDTools [159, 160].

4.3.9 Nucleotide frequencies

Nucleotide frequencies for mapped rNMP positions (that is, the 5′ position of each aligned read) were calculated and normalized to genome frequencies (nuclear and mitochondrial genomes in sacCer2). The identity of the rNMP base is the reverse complement of the 5′ base of each read. Nucleotide frequencies of downstream sequences of incorporated rNMPs, including the +1 position, cannot be affected by our approach of capturing rNMPs in DNA because the rNMPs and their upstream sequences are captured, sequenced and aligned to the reference genome.

4.3.10 Replication correlations

The density and identity of rNMPs present on leading and lagging strands were calculated relative to annotated origins of replication [161] and were further categorized by replication timing [162]. Data were filtered for specified replication timings (for example, 25 min after release into S phase) and distances relative to the middle of each ARS annotation (for example, 5.0 kb upstream and downstream of each ARS).

4.3.11 Determination of hotspots of rNMP incorporation in genomic DNA

Two different analyses were conducted to identify hotspots of rNMP incorporation in genomic DNA. Peak calling was performed with macs2 (version 2.1.0.20140616) [163] with specific parameters (--keep-dup all --nomodel -s 25 --extsize 5 --call-summits). Peaks of length greater than 1,000 were filtered from further analysis, and remaining peaks with a *q*-value less than 0.001 were selected. A second analysis involved finding positions of rNMPs within the locus of interest with ribose-seq signal greater than the mean plus three standard deviations for each library from *rnh201*Δ (KK-100), *rnh201*Δ (KK-100, EconoTaq), *rnh201*Δ (KK-30), *rnh1*Δ *rnh201*Δ (KK-174) and *rnh1*Δ *rnh201*Δ (KK-125) cells.

4.3.12 Data presentation and statistics

Graphs were made using GraphPad Prism 5 (GraphPad Software). A nonparametric two-tailed Mann-Whitney *U*-test [164] was implemented for statistical analysis of AtRNL ligation efficiencies, rNMP bypass probabilities and the percentages of StuI-cut Leu^+ transformants in the DSB repair assay. A chi-squared goodness-of-fit test [164] was used for statistical comparison of the distribution of rNMP reads to the expected Poisson distribution.

4.4 Results

4.4.1 Ribose-seq strategy to capture rNMPs in DNA

Ribose-seq captures rNMP-terminated single-strand (ss) DNA fragments generated by alkaline cleavage of rNMPs in DNA (**Figure 4.1** and **Figure C.1**). We exploited the distinctive ligation mechanism of *Arabidopsis thaliana* tRNA ligase (AtRNL), normally involved in tRNA maturation. AtRNL converts 2′,3′-cyclic phosphate ends of RNA to 2′-phosphate and ligates these to 5′-phosphate ends of RNA [156, 165] or DNA [165]. We demonstrated that AtRNL captures 2′,3′-cyclic phosphate or 2′-phosphate termini of DNA derived from alkaline cleavage of a DNA oligo at an embedded rNMP, ligating the 2′-phosphate end to the 5′-phosphate terminus of the same DNA molecule and producing a ss DNA circle containing an embedded rNMP. Self-ligation was strongly preferred over dimerization, as linear dimers were not detected (**Figure 4.1a**). Further, these ss DNA circles are resistant to T5 exonuclease, enabling their enrichment relative to unligated linear DNA upon exonuclease treatment (**Figure 4.1a**). We did not observe any bias for the 3′ rNMP substrate of AtRNL: the ligase captured an embedded rAMP, rCMP, rGMP, or rUMP with equal efficiency ($0.49 \leq P \leq 1.0$) (**Figure C.2** and **Table C.3**), nor was any bias observed in a previous study [166]. These data indicate that self-ligation is favored for AtRNL on 2′-phosphate-terminated ss DNA fragments as small as 22 nt (**Figure 4.1a** and **Figure C.2**), thus facilitating library construction and high-throughput DNA sequencing.

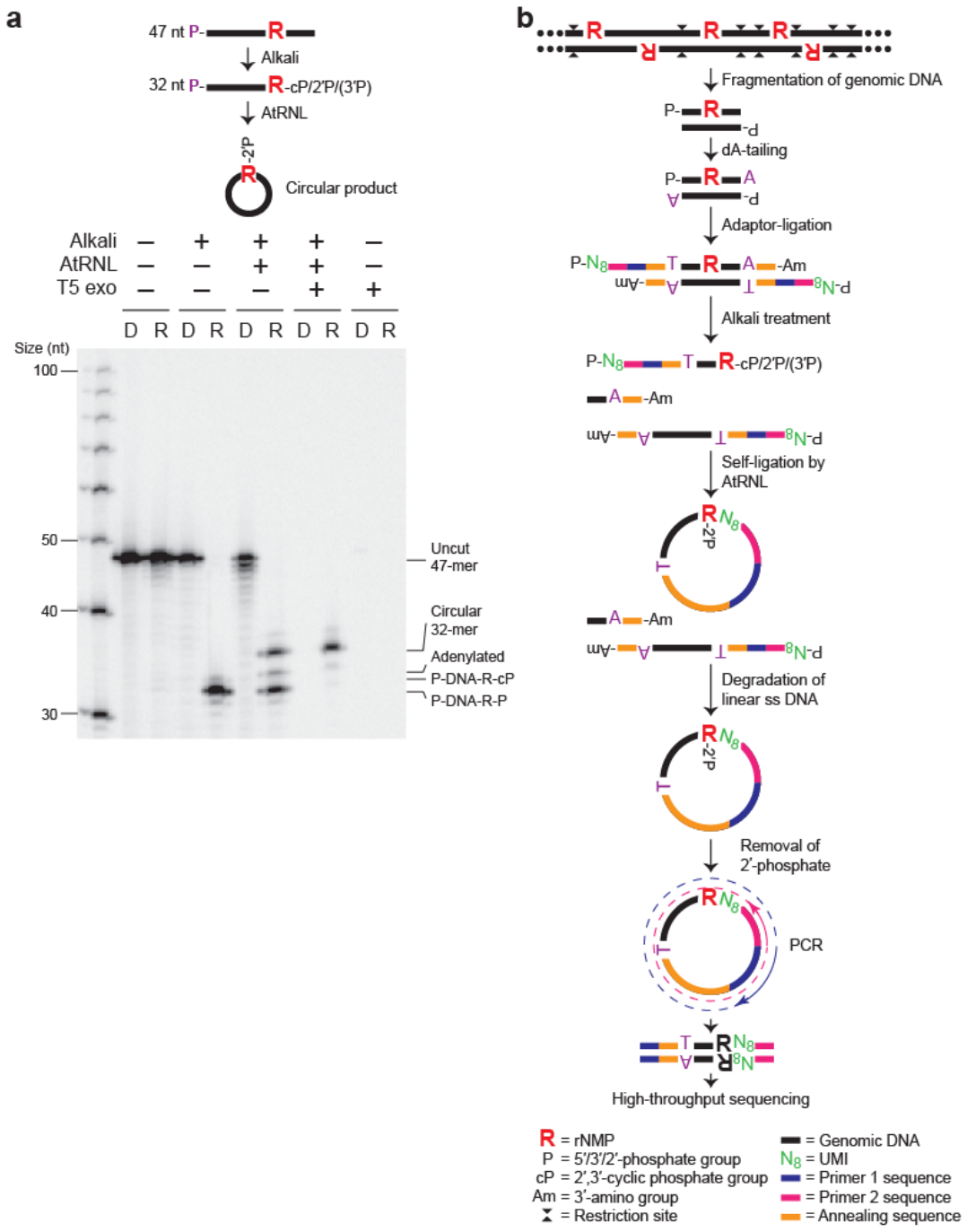


Figure 4.1 Ribose-seq method for mapping rNMPs in genomic DNA. (**a**) AtRNL captures 2′,3′-cyclic phosphate (cP) or 2′-phosphate (2′P) DNA termini, and does not capture 3′-phosphate (3′P) DNA termini (indicated in parentheses), generated by alkaline cleavage of a single rGMP in a 5′-radiolabeled 47-nt ss DNA oligo (see **Table C.2**). R, rGMP, or rGMP-bearing oligo. D, DNA-only control oligo. P in purple indicates the 5′

radiolabel. T5 exonuclease treatment confirms the presence of circular ligation product. Left lane, ss DNA ladder. Ligation efficiency was about 50%, as expected owing to the mixture of 2′-phosphate and 3′-phosphate ends generated upon alkaline cleavage. (**b**) Schematic of the ribose-seq approach. Genomic DNA is fragmented, dA-tailed and ligated to a molecular barcode-containing sequencing adaptor. Alkali treatment denatures the DNA and cleaves at rNMP sites, exposing 2′,3′-cyclic phosphate and 2′-phosphate termini, which are self-ligated to 5′-phosphate ends by AtRNL. Linear, unligated fragments are degraded by T5 exonuclease and the remaining rNMP-captured, circular DNA molecules, upon removal of the 2′-phosphate at the ligation junction by the 2′-phosphotransferase Tpt1, are PCR-amplified and sequenced. UMI, unique molecular identifier. R in black indicates the rNMP converted to a dNMP during PCR. Ribose-seq does not capture RNA primers of Okazaki fragments because the 5′-most rNMP is a 5′-triphosphate [26], and the T4 DNA ligase used to attach the sequencing adaptors absolutely requires a 5′-monophosphate [167]. Moreover, the rest of the primers are reduced to single nucleotides upon alkali treatment, and they will have no adaptor sequence ligated on. Ribose-seq also does not detect rNMP positions derived from residual RNA molecules or RNA:DNA hybrids not embedded in DNA (such as cDNA) nor DNA abasic sites, which could have been ligated to the adaptor sequence by T4 DNA ligase. Following alkali treatment, RNA stretches are reduced to single nucleotides that are removed in subsequent purification steps; even if the 5′-most rNMP is captured, the rNMP-containing single-stranded circle would not have any sequence to be aligned to the reference genome. Abasic sites undergo both β- and δ-eliminations to yield 5′-phosphate and 3′-phosphate ends [168, 169], which cannot be ligated by AtRNL. Because of the nature of alkaline hydrolysis within a stretch of rNMPs embedded in DNA, our ribose-seq captures only the 5′-most rNMP of the stretch of two or more rNMPs. Moreover, ribose-seq does not require rNMPs to be present at the same location from cell to cell, and it can identify incorporated rNMPs with single-base precision.

We applied ribose-seq to identify rNMPs embedded in nuclear and mitochondrial DNA of RNase H2–deficient budding yeast (strain KK-100, **Table C.1**). Genomic DNA was extracted from cells grown to stationary phase, and a mixture of three blunt-end restriction enzymes was used to fragment the DNA. Application of our rNMP-capture scheme (**Figure 4.1b**) yielded a library of DNA molecules (**Figure C.3a**) with an average size of ~350 bp, each of which maps to a single site of rNMP incorporation and its upstream sequence. In control experiments, we found that exclusion of either AtRNL (**Figure C.3a**) or alkali treatment (**Figure C.3b**) prevented library formation, validating that captured molecules derive from rNMPs embedded in DNA.

4.4.2 Spectrum of rNMPs in *S. cerevisiae* genome

A ribose-seq library prepared from *rnh201*Δ cells (KK-100) was sequenced to a depth of ~2 million reads, which were mapped to the yeast *S. cerevisiae* genome, allowing us to define rNMP locations along yeast nuclear and mitochondrial DNA with single-nucleotide resolution. This analysis uncovered widespread rNMP incorporation with a coverage of 0.449 and 19.5 rNMP reads per kb in the nuclear and mitochondrial genome, respectively (**Figure 4.2a** and **Table C.4**). While broadly scattered, the rNMP sites in the nuclear and mitochondrial DNA were not randomly distributed (**Figure 4.2b**,**c**). We found no major Watson/Crick strand bias in ribonucleotide distribution throughout the genome (**Figure 4.2a**), and the number of rNMPs identified per nuclear chromosome was proportional to chromosome size (**Figure 4.2d**).

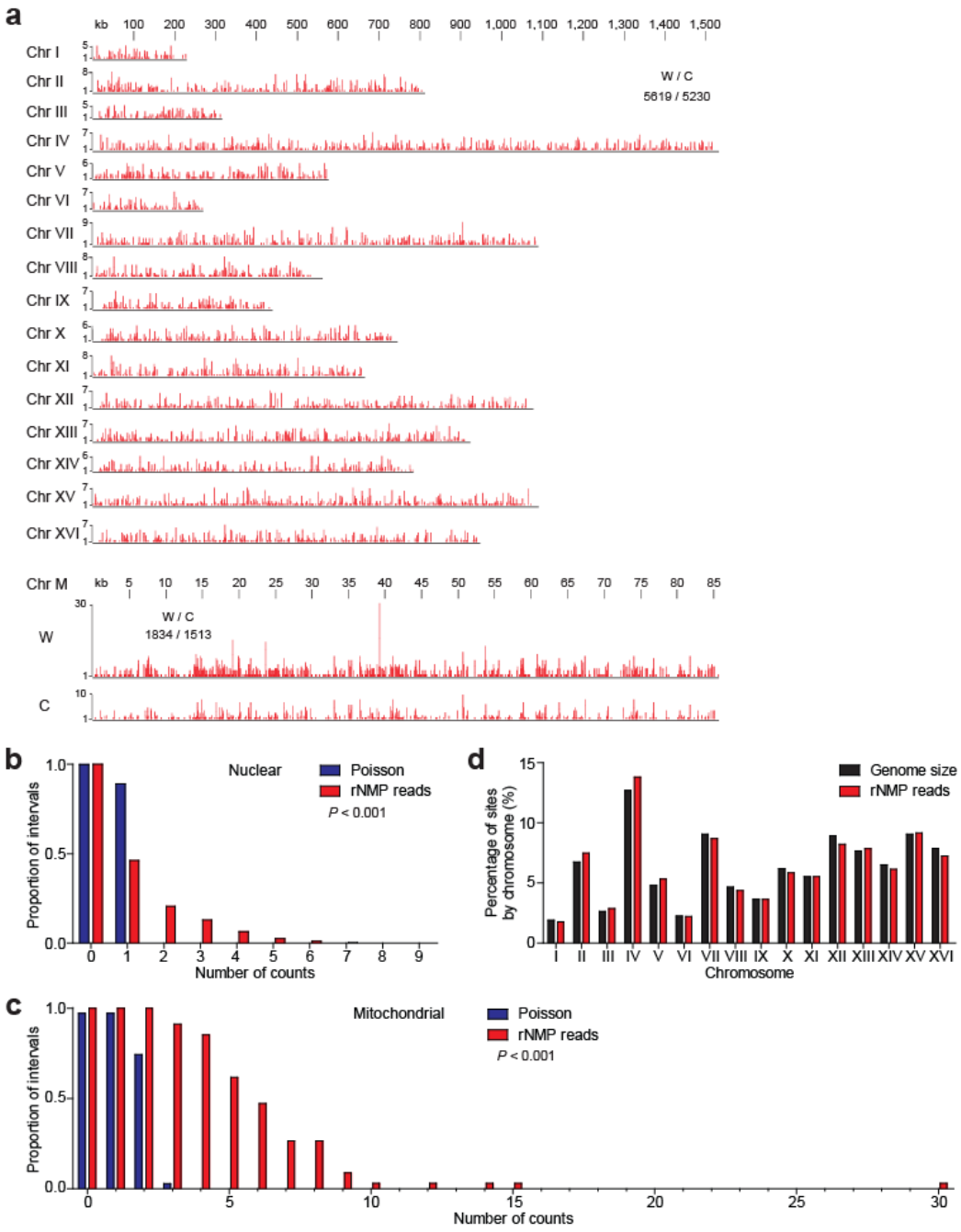


Figure 4.2 Distribution of rNMP incorporation in the *S. cerevisiae* genome. (**a**) Ribose-seq map of rNMPs in genomic DNA from *rnh201*Δ (KK-100) cells. The data, as peaks of rNMP reads, are shown for the individual nuclear chromosomes (Chr I–XVI) and the two strands of mitochondrial DNA (Chr M). The height of each peak corresponds to the number of reads. A comparison of nuclear and mitochondrial rNMP reads for

Watson (W) and Crick (C) strands is also displayed. (**b**,**c**) The proportion of 2.5-kb windows containing an observed number of rNMPs was calculated for (**b**) nuclear and (**c**) mitochondrial genomes and compared to random expectation based on Poisson frequencies. The *P* values calculated from a chi-squared goodness-of-fit test are shown (*n* = 10,847 and 3,347 aligned rNMP sites for nuclear and mitochondrial, respectively). (**d**) Chromosomal distribution of rNMPs compared to the size of each nuclear chromosome.

We determined the identity and relative frequencies of incorporated rNMPs, the reverse complement of the 5′ base of each read, as well as flanking bases for the nuclear and mitochondrial genomes in *rnh201*Δ cells. At the site of rNMP incorporation, we found that rCMP and rGMP were incorporated more frequently than expected from the G+C content, while rAMP and in particular rUMP were incorporated less frequently than expected from the A+T content in both nuclear and mitochondrial genomes, indicating a strong bias in the rNMP spectrum considering the (A+T)-rich nature of these genomes in yeast (62% and 83%, respectively) (**Figure 4.3a**,**b**). Examining the absolute composition of the genomic rNMPs, we found 44% rC, 28.1% rG, 15.4% rA, and 12.5% rU in the nuclear genome and 36.8% rC, 25.6% rA, 19% rG, and 18.7% rU in the mitochondrial genome (**Table C.4**). The difference in the base composition between nuclear and mitochondrial rNMPs is likely to be due to the higher A+T content of the mitochondrial genome.

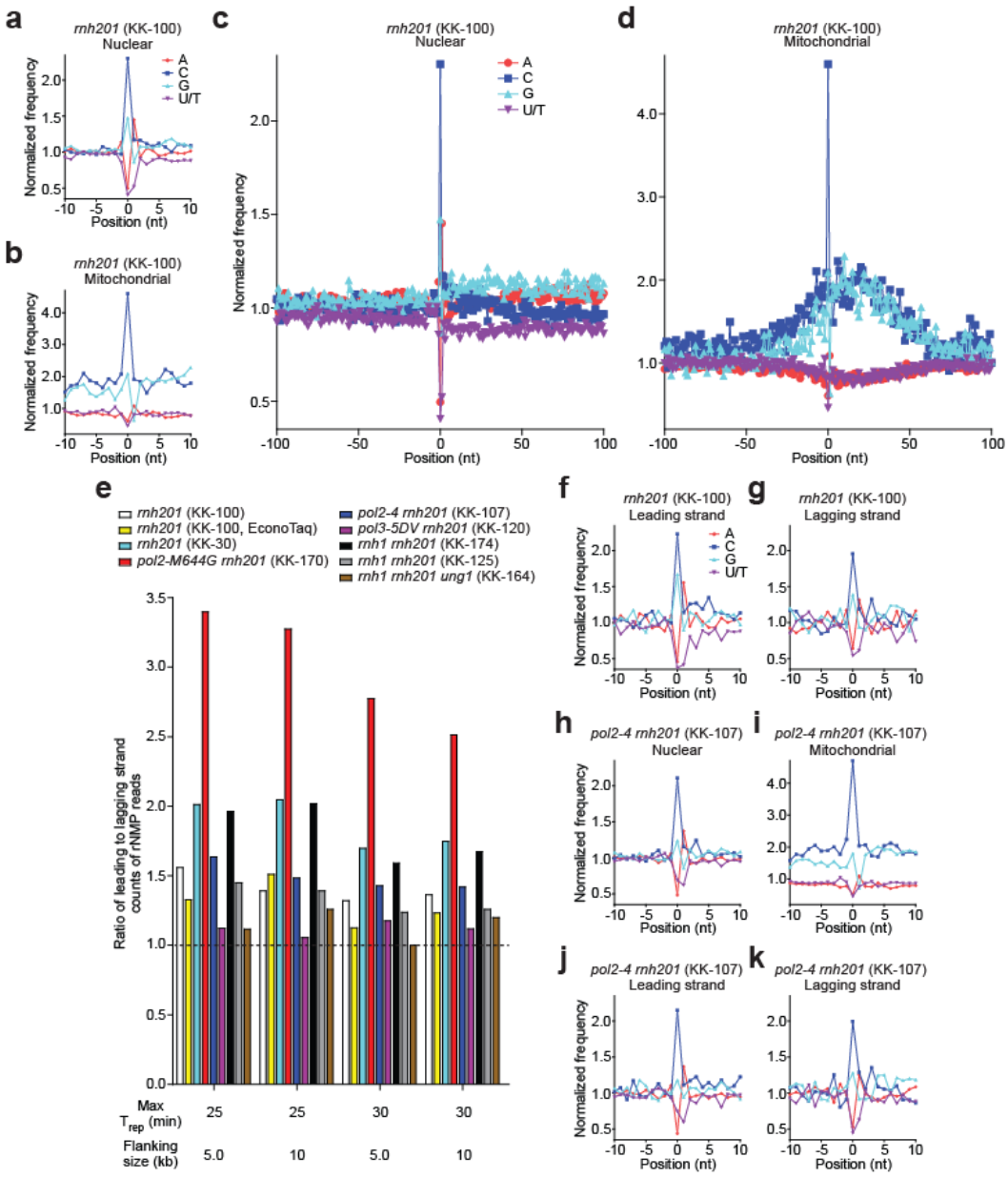


Figure 4.3 Identity and sequence contexts of rNMP incorporation in *S. cerevisiae* genome. (**a**,**b**) Normalized nucleotide frequencies relative to mapped positions of sequences from the ribose-seq library of *rnh201*Δ (KK-100) cells. Position 0 is the rNMP. (**c**,**d**) Zoom-out of frequencies. (**e**) Ratios of rNMPs on newly synthesized leading to lagging strand for all ribose-seq libraries. Early-firing ARSs selected by their replication timing (T_{rep}) were investigated for two different flanking sizes. EconoTaq indicates the library constructed using a Taq-based DNA polymerase. All other libraries were

constructed with a Pfu-based DNA polymerase. (**f**,**g**) Normalized nucleotide frequencies relative to mapped sequence positions in leading (**f**) and lagging (**g**) strands from the *rnh201*Δ (KK-100) library. ARSs with a T_{rep} of no longer than 25 min were selected with flanking size of 10 kb. (**h**–**k**) Normalized nucleotide frequencies relative to mapped sequence positions from a *pol2-4 rnh201*Δ (KK-107) library. Reads were mapped to the (**a**,**c**,**h**) nuclear genome, (**b**,**d**,**i**) mitochondrial genome, (**f**,**j**) leading strand, or (**g**,**k**) lagging strand.

The high level of rCMP and low level of rUMP observed both for nuclear and mitochondrial DNA in the *rnh201*Δ library are not attributable to differential bypass by the Pfu-based DNA polymerase used for PCR (**Figure C.4** and **Table C.6**), as we also observed similar rNMP patterns using a Taq-based DNA polymerase (**Figure C.5a**,**b** and **C.6a**,**b**). Similarly, the nucleotide frequency derived from a ribose-seq library constructed from another *rnh201*Δ strain (KK-30, **Table C.1**) was comparable to that obtained from strain KK-100 both for nuclear and mitochondrial sites (**Figure C.5c**,**d** and **C.6c**,**d**). Additional deletion of the gene encoding RNase H1 (*rnh1*Δ), generating *rnh1*Δ *rnh201*Δ strains KK-174 and KK-125, did not affect the nucleotide frequency of rNMP incorporation (**Figure C.5e**–**h** and **C.6e**–**h**). While some variation in the absolute rNMP counts were found among these different libraries in the mitochondrial DNA (**Table C.5**), the high level of rCMP and low level of rUMP remained constant, as well as a preferred rNMP incorporation in (G+C)-rich regions of the mitochondrial DNA. These data support a model in which rNMPs in yeast genomic DNA are present as single, di- or tri-nucleotides, which are not substrates of RNase H1 [34], and indicate that RNase H1 has only a minor impact on the distribution of genomic rNMPs.

To test whether the low frequency of rUMP incorporation was a consequence of removal by the uracil N-glycosylase, Ung1, we deleted the *UNG1* gene in the RNases H–defective background (*rnh201*Δ *rnh1*Δ *ung1*Δ, strain KK-164) and mapped rNMP sites in these cells. Ung1 repairs dUMP from nuclear and mitochondrial DNA [92]. Although Ung1 does not act on uracil in RNA (for example, ribosomal RNA) [170], it is not known whether Ung1 can act on rUMP embedded in a DNA duplex. We found that the level of rUMP incorporation in the chromosomal and mitochondrial genomes of an *rnh1*Δ *rnh201*Δ *ung1*Δ strain was similar to that in an *rnh1*Δ *rnh201*Δ strain, demonstrating that Ung1 does not target rUMP in DNA (**Figure C.5i**,**j** and **C.6i**,**j** and **Table C.5**).

Using a yeast assay of chromosomal double-strand break repair (DSB), in which DNA oligonucleotides carrying embedded rGMP, rUMP, or deoxyribonucleotides only are templates for DSB repair (**Figure C.7a**), we demonstrated that Ung1 targets uracil from a dUMP but not an rUMP embedded in DNA, while RNase H2 targets only rNMPs (rGMP and rUMP in this experiment) but not dNMPs (**Figure C.7b** and **Table C.7**). We attribute rNMP incorporation frequencies to the levels of corresponding dNTPs. dCTP and dGTP are typically the least abundant dNTPs [21, 25] and therefore might be depleted faster than dTTP and dATP, increasing the probability of rCMP and rGMP incorporation over rUMP and rAMP. These results are also consistent with the finding that rCMP and rGMP are the most frequently incorporated rNMPs by DNA polymerases *in vitro* under physiological dNTP and rNTP concentrations [14, 42]. This ability of DNA polymerases to incorporate rNMPs into genomic DNA could serve as a mechanism for continuing replication under conditions in which one or more dNTP pools are depleted. In the presence of hydroxyurea, a known ribonucleotide reductase inhibitor, higher levels of rNMPs are found incorporated in genomic DNA [34]. However, extensive rNMP incorporation would also result in increased breaks and genomic instability.

4.4.3 Pattern of sequences flanking rNMPs in *S. cerevisiae* DNA

Downstream of incorporated rNMPs, we found that the +1 position was most frequently dA and least frequently dG both in the nuclear and in the mitochondrial genomes, with 42–52% dA and 6–16% dG among all four deoxyribonucleotides (dA, dC, dG, and dT) (**Table C.5**). At the +1 position, dT was also frequent (31–40%) in the mitochondrial genome. In mitochondrial DNA, the high level of dA or dT at the +1 position 3′ from the rNMP could reflect the high A+T content in the mitochondrial genome. It is also possible that the dA in +1 position influences rNMP incorporation by DNA polymerases. Alternatively, we speculate that dA in the +1 position might stabilize incorporated rNMPs, possibly by affecting base stacking and preventing its repair by mechanisms other than ribonucleotide excision repair. It will be useful to determine the nearest-neighbor thermodynamic parameters for single rNMPs in DNA duplex and, in particular, the stability trend for the base pair 3′ of the rNMP sites. We recently showed that single rGMPs embedded in a short DNA duplex have a marked effect on the elastic properties of DNA by altering the DNA structure at the site encompassing the rNMP and the nucleotide 3′ to it [91]. Thus, it is reasonable to think that the +1 position 3′ from the rNMP is prone to altered structure, and it is the most critical site for signaling the presence of an rNMP in DNA because it is the closest nucleotide to the 2′-OH group of the rNMP.

Sites of rNMP incorporation were flanked by sequence contexts that differed between the nuclear and mitochondrial DNA genomes. While nucleotide frequencies up- and downstream of rNMP sites in the nuclear genome were largely similar to background frequencies (**Figure 4.3c**), rNMP sites in mitochondrial DNA were primarily upstream of (G+C)-rich regions, concentrated in areas in which G+C content was 1.7 to 1.8 times that of the background (**Figure 4.3d**). Notably, mitochondrial G+C tracts have been shown to have recombinogenic properties [171], and mitochondrial DNA recombination has been

suggested to initiate mitochondrial DNA replication in yeast ([172] and references therein). Thus, it is possible that the presence of rNMP sites in yeast mitochondrial G+C clusters influences these recombination events in mitochondrial DNA.

4.4.4 rNMP incorporation by replicative DNA polymerases

We next analyzed rNMP incorporation in the newly synthesized leading and lagging strands of yeast nuclear DNA. We selected 154 to 271 early-firing yeast autonomously replicating sequences (ARSs) (activated in the first 25 or 30 min, respectively) on the basis of replication timing [162]. We examined the type and abundance of rNMPs incorporated in regions 5 or 10 kb upstream and downstream from selected ARSs. This analysis was conducted using all our ribose-seq libraries, including a library derived from yeast RNase H2–deficient cells containing the low-fidelity Pol ε mutant (*rnh201*Δ *pol2-M644G*, **Table C.1**). Because yeast Pol ε is mainly responsible for leading strand synthesis during DNA replication, yeast cells containing the *pol2-M644G* mutation, which leads to increased rNMP incorporation, would be predicted to contain more rNMPs on the newly synthesized leading strand than on the lagging strand [33].

We found higher rNMP incorporation on the newly synthesized leading strand of DNA replication (**Figure 4.3e**), consistent with previous observations that the leading strand DNA Pol ε incorporates more rNMPs than the lagging strand Pol δ [21]. As expected, analysis of rNMPs from *rnh201*Δ *pol2-M644G* cells revealed a stronger bias toward rNMP incorporation on the newly synthesized leading strand as compared to that in all other libraries (**Figure 4.3e**). However, the increase in rNMP incorporation by the low fidelity Pol ε mutant did not change the overall rNMP spectrum, which had similar patterns of rNMP incorporation to those of libraries derived from wild-type Pol ε strains (**Figure C.5k**,**l** and **C.6k**,**l**). Furthermore, we examined whether the spectrum of rNMP

incorporation was different between the newly synthesized leading and the lagging strand upon DNA replication. Cells containing either *rnh201*Δ or *rnh1*Δ *rnh201*Δ mutations had similar spectra of rNMP incorporation on the leading and the lagging strand (**Figure 4.3f**,**g** and **C.8a**,**b**). In contrast, mapping of rNMPs in yeast cells carrying a mutant allele of DNA Pol ε that is defective in proofreading activity (*pol2-4*, **Table C.1**) showed a lower frequency of rA versus rU in the nuclear but not in the mitochondrial genome (**Figure 4.3h**,**i** and **C.6m**,**n** and **Table C.5**), and it showed a bias for lower rA than rU only on the newly synthesized leading strand (**Figure 4.3j**,**k**), which was not observed in libraries derived from wild-type nor low-fidelity mutant Pol ε strains. The rNMP spectra for a strain with proofreading-defective DNA Pol δ (*pol3-5DV*, **Table C.1**) were not different from those containing the wild-type Pol δ (**Figure C.5m**,**n**, **C.6o**,**p**, and **C.8c**,**d** and **Table C.5**). These results suggest that DNA Pol ε can proofread rNMPs, particularly rUMP, in DNA and that this activity is superior to that of DNA Pol δ, as is consistent with previous biochemical studies [14, 42].

4.4.5 Hotspots of rNMP incorporation in the *S. cerevisiae* genome

We performed two types of analysis to determine potential hotspots of rNMP incorporation in the *S. cerevisiae* genome. We identified enriched regions of rNMP incorporation in genomic DNA from ribose-seq data (see **Materials and Methods**). We found several regions of notable rNMP incorporation in mitochondrial DNA for each ribose-seq library in this study (**Figure 4.4a** displays a few regions). Because this analysis excludes all reads aligning to more than one position in the genome, we performed a second analysis with respect to specific loci to identify single-nucleotide hotspots that were reproducibly present in multiple ribose-seq libraries (see **Materials and Methods**). We identified hotspots of rNMP incorporation in sequences present in multiple copies per yeast cell: the mitochondrial genome (~80 copies) [173], the

ribosomal DNA (rDNA) repeats (~140) clustered on chromosome XII [174] and the yeast retrotransposon (Ty), of which there are ~30 copies encoded on multiple chromosomes [175]. In mitochondrial DNA, we found a marked hotspot at an rAMP on the Watson strand in the cytochrome oxidase B gene (*COB*) and the overlapping maturase *BI3* and *BI4* genes (**Figure 4.4b** and **Table C.8**), in addition to several other hotspots (**Table C.8**). In the rDNA locus, the strongest hotspot was found in gene *RDN37-1* and the overlapping *RDN25-1* at an rGMP (**Figure 4.4c** and **Table C.8**). In the yeast Ty1 sequence, we found a hotspot at an rAMP in the coding sequence of *TY1A-1* (**Figure 4.4d** and **Table C.8**).

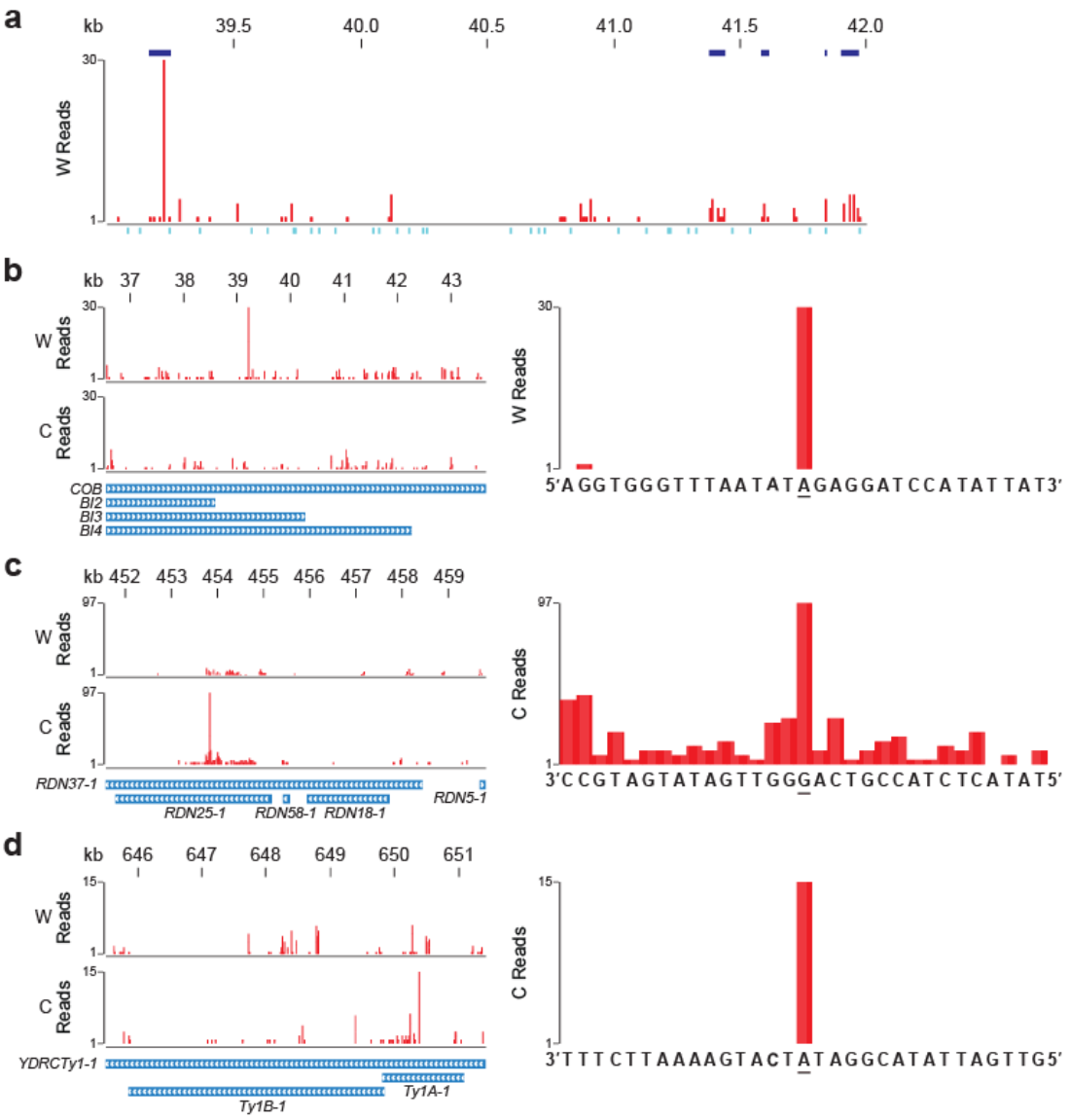


Figure 4.4 Hotspots of rNMP incorporation in *S. cerevisiae* mitochondrial DNA, rDNA repeat and Ty1. (**a**) Ribose-seq map of rNMPs in a 3-kb window (39,001–42,000) of mitochondrial DNA showing enriched regions of rNMP incorporation. Enriched regions with *q*-value < 0.001 are shown in blue above the plot. Positions of restriction sites used for genomic fragmentation are displayed below the plot in turquoise. (**b**) Map of rNMPs in the *COB* mitochondrial locus (left). Zoom-in map (right) with sequence at the hotspot site (underlined). (**c**) Map of rNMPs in the first of two rDNA repeat loci on Chr XII, based on alignment data from the two loci of the reference genome (left). Zoom-in map (right) of the rDNA hotspot. (**d**) Map of rNMPs in the Ty1

locus *YDRCTy1-1* on Chr IV based on multiple-alignment data from several Ty1 loci (left). Zoom-in map (right) of the Ty1 hotspot. Results are shown for *rnh201*Δ (KK-100) cells. W, Watson strand; C, Crick strand.

The occurrence of such hotspots indicates that there are preferred sites for rNMP incorporation in the mitochondrial genome, rDNA and Ty1 sequences. In addition to the recombinogenic properties of mitochondrial G+C clusters discussed above, yeast rDNA and Ty are also active in recombination [174, 176]. Frequent rNMP incorporation could trigger recombination, as do rNMPs embedded in the mating type locus of *Schizosaccharomyces pombe* [4]. The rNMPs detected in Ty DNA could originate from cDNA rather than genomic DNA. Because *rnh201*Δ *rnh1*Δ cells have abundant Ty cDNA [177], if rNMPs are incorporated in Ty1 during the process of reverse transcription, which forms the cDNA, we would expect a different rNMP pattern in Ty1 DNA in *rnh201*Δ *rnh1*Δ than in *rnh201*Δ cells. Although we did not observe major differences in the rNMP spectra derived from *rnh201*Δ single mutant versus *rnh201*Δ *rnh1*Δ cells at the Ty1 locus or in general (**Figure 4.3e**, **C.5**, and **C.6** and **Table C.8**), it would be of interest to conduct *in vitro* tests to determine whether Ty reverse transcriptase incorporates rNMPs frequently opposite RNA and/or DNA, and whether it has a particular bias for rNMP incorporation.

4.5 Discussion

rNMP incorporation has been extensively studied in recent years; however, locating sites of rNMP incorporation in genomic DNA has not yet been possible. Alkaline cleavage of rNMPs and AtRNL ligation exclude Okazaki fragments and DNA abasic sites, allowing the construction of ribose-seq libraries containing stably incorporated rNMP sites. Ribose-seq enabled us to determine the widespread but nonrandom distribution of rNMPs in budding yeast genomic nuclear and mitochondrial DNA, with several hotspots.

Our findings both validated the approach and uncovered new aspects of rNMP incorporation in the yeast genome. The observed strand bias incorporation on the newly synthesized leading strand in wild-type and low-fidelity Pol ε strains, and the specific rNMP pattern in yeast containing proofreading deficient Pol ε, provide strong support for the *in vitro* results obtained for these forms of Pol ε. rCMP and rGMP were more abundant than rAMP and rUMP, and there was frequently a dA downstream of the rNMPs. We also found that RNase H1 did not contribute substantially to rNMP incorporation, and Ung1 did not remove genomic uracil. It is possible that the paucity of rUMP in DNA reflects inherent cleavage bias in other rNMP removal pathways, such as topoisomerase-mediated rNMP cleavage [62]. It would be of interest to determine the rNMP spectrum in cells with defects in alternative rNMP removal pathways, either in RNase H2 wild-type or null cells growing under normal and/or stressed conditions. Ribose-seq, together with HydEn-seq [178], emRiboSeq [31], and Pu-seq [179] which were developed in parallel and capture the nucleotides downstream or upstream of rNMP positions, should allow us to better understand the impact of rNMPs on the structure and function of DNA and chromatin, and specific rNMP signatures may represent new biomarkers for human diseases such as AGS, cancer, and other degenerative disorders.

4.6 Acknowledgments

We thank N.V. Hud and L.D. Williams for support with urea-PAGE gels and for advice on this study and the manuscript; M. Goodman, B. Weiss, and I.K. Jordan for suggestions on this study and the manuscript and assistance with data analysis; S. Garrey for AtRNL and Tpt1 protein purification; C. Cox for sequencing; Y. Shen for assistance with statistical analysis; A. Gombolay and L. Shetty for technical help; and all members of the Storici laboratory for advice in the course of the study. This research was supported by US National Science Foundation award number MCB-1021763 (to F.S.), Georgia

Research Alliance award number R9028 (to F.S.), an American Cancer Society Research Scholar Grant (to J.R.H.), a Damon Runyon-Rachleff Innovation Award from the Damon Runyon Cancer Research Foundation (to J.R.H.), and the University of Colorado Golfers Against Cancer (to J.R.H.).

CHAPTER 5

CONCLUSIONS

rNMPs are now known to be the most abundant non-canonical nucleotides incorporated into DNA [148]. Given the high frequency of rNMP incorporation into DNA, repair pathways to remove and tolerate rNMPs in DNA, their consequences to genomic integrity, and their profiles, including their identities and locations, need to be determined.

RNases H type 1 and type 2, involved in RER, have been well-characterized *in vitro* as enzymes to target rNMPs embedded in DNA [37]. Most *in vivo* studies so far performed to investigate the role of RNases H have been based on the ability of DNA polymerases to incorporate rNMPs into DNA [32, 33, 39, 60-62], preventing the studies to focus on a single-nt position in a specific locus. As demonstrated in bacteria *E. coli*, yeast *S. cerevisiae*, and human HEK-293 cells using synthetic RNA-containing oligos, RNA can serve as direct template for DNA synthesis at the chromosomal level [46, 55-57]. Following this demonstration, we used an approach to generate desired RNA:DNA hybrids at the chromosomal level by transformation with rNMP(s)-containing oligos with homology to a specific chromosomal locus. This procedure allowed us to select specific single nucleotide positions in any locus as rNMPs and determine whether those specific rNMPs could transfer genetic information and/or could be targeted by different DNA repair mechanisms. We showed that isolated mispaired rNMPs in chromosomal DNA are recognized and removed not only by RNases H but also by MMR in both *E. coli* and *S. cerevisiae* cells. In addition, isolated paired rNMPs in DNA are targeted by RNase H2 and NER in yeast, similar to NER's activity in *E. coli* [180, 181], indicating that sugar-phosphate backbone distortions caused by rNMP embedded in DNA are sufficient for

recognition by NER. Our results suggest that RNases H are DNA repair proteins with ability to remove rNMPs incorporated into DNA while MMR and NER also play a role in their repair.

Ung1-initiated BER did not target uracil from an rUMP embedded in DNA, indicating that Ung1 is a true DNA glycosylase recognizing also the sugar of the deoxyribonucleotide, not just the uracil base. However, this result does not rule out the possibility that BER could still be involved in repair of rNMPs in DNA. rNMPs, if not removed by RNase H2, could be incised by Top1, transiently linking to 3′-phosphate of the rNMP [32, 39, 41, 62]. Top1 is released upon nucleophilic attack by the 2′-OH, resulting in a nick containing 2′,3′-cyclic phosphate, which is hypothesized by many as a substrate for endonucleases, exonucleases, and/or helicases [150]. Srs2 and Exo1 have been reported to be involved in Top1-rNMP-resulting nick processing [40]. Other possible protein factors are apurinic/apyrimidinic (AP) endonucleases (Apn1 and Apn2) of the BER pathway, which have the capability to remove 3′-phosphates [182]. Whether AP endonucleases and/or lyases (Ntg1 and Ntg2) involved in BER could remove rNMPs in DNA still needs to be determined.

Consequences of rNMPs have been shown to be both positive and negative for cells [48, 54, 148, 150]. With only a few instances where rNMPs in DNA are helpful, replication stress and genomic instability have been generally considered as consequences of rNMPs in DNA. The characteristics of rNMPs in DNA that lead to these consequences remain largely under-studied. With the highly reactive extra 2′-OH group in the sugar, rNMPs in DNA have the potential to affect its mechanical and structural properties, which ultimately affect DNA-DNA and DNA-protein interactions involved in variety of cellular processes. Most structural studies done on rNMPs in DNA focused on stretches of two or more rNMPs, likely to be found in Okazaki fragments during DNA replication [72, 73],

whether isolated single rNMPs change DNA elasticity and structure needed to be determined. Using AFM-base single molecule force spectroscopy, MD simulations, and NMR, we showed that single rNMPs do change the elastic and structural properties of DNA. We observed that the direction and the degree of the change in elasticity was different in two sequences tested, suggesting sequence context-dependence. Also, MD simulations and NMR both showed local helical distortions, mostly at and downstream of the rNMPs. Different sequences did lead to different backbone distortions, also suggesting sequence context-dependence.

The evidence of high frequency of rNMP incorporation into DNA, existence of repair pathways to remove them, and their helpful and harmful roles in cells begs the question of which rNMPs are and where are they incorporated into genomic DNA. We developed ribose-seq, a unique, innovative approach to profile and map rNMPs in DNA. By generating rNMP-specific phosphate ends by alkaline cleavage of rNMPs, ribose-seq allows capture and next-generation sequencing of rNMPs embedded in genomic DNA. rNMP incorporation spectrum in *S. cerevisiae* genome revealed widespread but nonrandom distribution, with preferences for base composition of rNMPs and their neighboring DNA sequence context and several hotspots in both nuclear and mitochondrial DNA. Ribose-seq allows us to explore rNMP incorporation into DNA potentially in any cell type of any organism.

Overall, our ribose-seq approach opens up a new direction to better understand the impact of rNMPs on the structure and function of DNA and chromatin. With ribose-seq, specific signatures of rNMP incorporation into DNA could be determined and then be applied to our *in vivo* assay with rNMP(s)-containing oligos and physical experiment employing AFM, MD simulations, and NMR. For example, we can examine if certain hotspots/coldspots of rNMP incorporation from the ribose-seq data are more/less

recognized by DNA repair enzymes or alter more/less the elasticity and structure of DNA. Moreover, as we observed different signatures of rNMP incorporation in yeast nuclear and mitochondrial genomes, it is possible that rNMP incorporation may vary in DNA of different cell types of different organisms. The variability could be related to the complexity of organisms, such as prokaryotes vs. eukaryotes, unicellular vs. multicellular, and invertebrate vs. vertebrata, or the cell type, like stem cell vs. somatic cell and muscular cell vs. neuronal cell, and therefore could have evolutionary and developmental implications.

Also of interest are the effects of different growth/stress conditions on rNMP incorporation into genomic DNA. Unlike other known rNMP-mapping techniques, such as HydEn-seq [178], emRiboSeq [31], and Pu-seq [179], the unique nature of direct capture of rNMPs embedded in DNA by ribose-seq allows us to investigate rNMP incorporation into DNA under stress. HU, a ribonucleotide reductase inhibitor, has been shown to increase the level of rNMP incorporation into DNA of mouse embryonic fibroblasts [34] while oxidative damage can convert dNMPs into rNMPs [29]. Whether rNMP spectrum is affected by HU, oxidative damage, or even growth phase and which rNMP repair pathways might be specific to those effects could now be determined.

Several studies on rNMP-induced genome instability have reported rNMP-Top1-dependent hotspots of 2–5 bp deletions [32, 39, 60, 62]. Nick McElhinny *et al*. defined 5′-CACA-3′ as a strong hotspot of rNMP-induced genome instability [60]. This is in agreement with our ribose-seq results that rCMP was highly favored for incorporation into DNA and that dAMP was most common dNMP immediately downstream of rNMPs. Ribose-seq could be applied to genomic DNA of cells with defects in alternative rNMP removal pathways, such as Top1–null background. Furthermore, the specific rNMP signature could be implemented in our *in vivo* assay and/or physical experiment via

rNMP-containing oligos so that more information about the specificity of repair pathways and its physical effects can be obtained.

Lastly, ribose-seq allows the possibility of finding specific rNMP signatures which could represent novel biomarkers for human diseases, such as AGS, cancer, and other degenerative disorders. For example, it is known that partial loss-of-function mutations in subunits of RNase H2 are associated with AGS [64]. However, it still remains unknown what is resulted from these RNase H2 mutations which trigger AGS. An increased level of general rNMP incorporation is a possibility while a change in rNMP incorporation spectrum is also plausible. With combination of our ribose-seq, *in vivo* repair assay, and physical experiment, the true effects of AGS-associated RNase H2 mutations could potentially be determined.

APPENDIX A

SUPPLEMENTARY MATERIALS FOR CHAPTER 2

Table A.1 Bacterial and yeast strains used in this study.

a

Strain[a]	Relevant genotype	Source
BW1892	Δ*mutS*::*gm λcI857* Δ(*cro-bio*) *lacZ*(ΔGG1370-1; G1384→A)	[82]
BW1947	Δ*mutS*:: *gm λcI857* Δ(*cro-bio*) *lacZ*(G1384→A)	[82]
BW1947 Δ(*N-gam*)	BW1947 Δ(*N-gam*)[b]	Transformation of BW1947 with an oligo [80]
BW1988	BW1892 Δ(*N-gam*)	Transformation of BW1892 with an oligo [80]
BW2028	BW1988 Δ*rnhA733*::*kan*	P1(JW0204-2) × BW1988[c]
BW2029	BW1988 Δ*rnhB782*::*kan*	P1(JW0178-1) × BW1988
BW2031	BW1988 Δ*rnhB*::*FRT*	Excision of *kan* cassette from BW2029
BW2032	BW1988 Δ*rnhB*::*FRT* Δ*rnhA733*::*kan*	P1(JW0204-2) × BW2031
BW2037A,B	Δ*mutS*:: *gm srlD3131*::Tn*10* λ*cI857* Δ(*cro-bio*) Δ(*N-gam*) *lacZ*(G1384→A)	P1(CAG18642) × BW1947 Δ(*N-gam*)[d]
BW2038A,B	BW2037 *mutS*$^+$	P1(CAG18642) × BW1947 Δ(*N-gam*)[d]
BW2039A,B	BW2037 Δ*rnhB782*::*kan*	P1(JW0178-1) × BW2037
BW2040A,B	BW2037 *mutS*$^+$ Δ*rnhB782*::*kan*	P1(JW0178-1) × BW2038
CAG18642	*srlD3131*::Tn*10*	[183, 184]
JW0178-1	Δ*rnhB782*::*kan*	[185]
JW0204-2	Δ*rnhA733*::*kan*	[185]
YSB-13,14	BW1988 *rnhA*(48Oc,Op)	Transformation of BW1988 with an oligo
YSB-15,16	BW1988 *rnhB*(70Oc,Op)	Transformation of BW1988 with an oligo
YSB-17,18	BW2037 *rnhB*(70Oc,Op)	Transformation of BW2037 with an oligo
YSB-19A,B	BW2037 *rnhA*(48Oc,Op)	Transformation of BW2037 with an oligo
YSB-20A,B	BW2037 *rnhA*(48Oc,Op) *rnhB*(70Oc,Op)	Transformation of YSB-18 with an oligo
YSB-21A,B	YSB-19 *mutS*$^+$	P1(JW2674-1) × YSB-19
YSB-22A,B	YSB-18 *mutS*$^+$	P1(JW2674-1) × YSB-18
YSB-23A,B	YSB-20 *mutS*$^+$	P1(JW2674-1) × YSB-20
JW2674-1	Δ*srlD*::(FRT-*kan*-FRT)	[185]

b

BY4742	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*	[87, 186]
FRO-694	BY4742 *trp5*::GSKU	Insertion of GSKU cassette into *TRP5*
YS-301,303	BY4742 *trp5*(ΔCC1001-2; G1017→A)	Replacement of GSKU in FRO-694 with an oligo to introduce the desired mutations
YS-305,306	YS-301,303 Δ*rnh201*::*kanMX4*	Disruption of *RNH201* with *kanMX4* PCR product
YS-307,308	YS-301,303 Δ*msh2*::*kanMX4*	Disruption of *MSH2* with *kanMX4* PCR product
YS-313,314	YS-301,303 Δ*rnh201*::*hygMX4* Δ*msh2*::*kanMX4*	Disruption of *RNH201* with *hygMX4* PCR product and of *MSH2* with *kanMX4* PCR product
YS-316,318	BY4742 *trp5*(ΔCC1001-2)	Replacement of GSKU in FRO-694 with an oligo to introduce the desired mutation
YS-331,332	YS-316,318 Δ*rnh201*::*hygMX4*	Disruption of *RNH201* with *hygMX4* PCR product
YS-327,328	YS-316,318 Δ*msh2*::*kanMX4*	Disruption of *MSH2* with *kanMX4* PCR product
YS-323,324	YS-316,318 Δ*rnh201*::*hygMX4* Δ*msh2*::*kanMX4*	Disruption of *RNH201* with *hygMX4* PCR product and of *MSH2* with *kanMX4* PCR product
YS-320,322	BY4742 *trp5*(G1017→A)	Replacement of GSKU in FRO-694 with an oligo to introduce the desired mutation
YS-333,334	YS-320,322 Δ*rnh201*::*hygMX4*	Disruption of *RNH201* with *hygMX4* PCR product
YS-329,330	YS-320,322 Δ*msh2*::*kanMX4*	Disruption of *MSH2* with *kanMX4* PCR product
YS-325,326	YS-320,322 Δ*rnh201*::*hygMX4* Δ*msh2*::*kanMX4*	Disruption of *RNH201* with *hygMX4* PCR product and of *MSH2* with *kanMX4* PCR product
FRO-767,768	*hoΔ hmlΔ::ADE1 MATa-inc hmrΔ::ADE1 ade1 leu2-3,112 lys5 trp1::hisG ura3-52 ade3::GAL::HO leu2::HOcs mataΔ::hisG*	[46]
FRO-984,985	FRO-767,768 *rnh201Δ::kanMX4*	[187]
YS-388,389	FRO-767,768 *rad14Δ::kanMX4*	Disruption of *RAD14* with *kanMX4* PCR product
YS-390,391	YS-388,389 *rnh201Δ::hygMX4*	Disruption of *RNH201* with *hygMX4* PCR product
KK-158,159	FRO-767,768 *ung1Δ::hygMX4*	[187]

(**a**) *E. coli* and (**b**) *S. cerevisiae* strains.

[a]All *E. coli* strains are derivatives of *E. coli* K-12 F^- λ^-. Abbreviations: *gm*, gentamycin resistance gene; *kan*, kanamycin resistance gene; FRT, Flp recombinase target sequence; GSKU, cassette containing the I-SceI endonuclease gene under the inducible *GAL1* promoter, the *kanMX4* kanamycin resistance gene, the counterselectable *KlURA3* marker and the I-SceI cutting site.

[b]Because both RecB and RecC are required for the viability of an *rnhA* mutant [188], we deleted the prophage *N-gam* segment [80], the product of which is an inhibitor of the

RecBCD enzyme [189]. The partially deleted prophage retained functional *exo* and *bet* genes of the λ *red* region, of which only the *bet* gene is necessary for oligo transformation [80].

[c]Transductions were performed with bacteriophage P1 *dam rev6* [190]. Transductions with bacteriophage P1 are described as follow s: P1(donor) × recipient. Selections were for resistance to kanamycin, tetracycline (Tn*10*), or gentamycin.

[d]Selection was for tetracycline resistance (*srlD*::Tn*10*). The presence of a cotransduced *mutS*$^{+}$ allele was determined by gentamycin sensitivity and by loss of the mutator phenotype as evidenced by over a 90% reduction in spontaneous streptomycin-resistant mutants.

Table A.2 Oligos used in this study.

Name	Size	Sequence
LacZ.R6$_{I2}$	65	5′-GTAATCACCCGAGTGTGATCATrCrUrGrGrUrCGCTGGGGAATGAGTCAGGCCACGGCGCTAATCACGAC
LacZ.R1$_{S1}$	65	5′-GTAATCACCCGAGTGTGATCATCTGGTCGCTGGGGAATrGAGTCAGGCCACGGCGCTAATCACGAC
LacZ.R5$_{S1}$	65	5′-GTAATCACCCGAGTGTGATCATCTGGTCGCTGGGGArArUrGrArGTCAGGCCACGGCGCTAATCACGAC
LacZ.D	65	5′-GTAATCACCCGAGTGTGATCATCTGGTCGCTGGGGAATGAGTCAGGCCACGGCGCTAATCACGAC
RpsL.R1$_{S1}$	70	5′- ACGTACGGTGTGGTAACGAACACCCGGGAGGTCTrCTAACACGACCGCCACGGATCAGGATCACGGAGTGC
RpsL.D	70	5′- ACGTACGGTGTGGTAACGAACACCCGGGAGGTCTCTAACACGACCGCCACGGATCAGGATCACGGAGTGC
TRP5.R2_R1$_{I2_S1}$	65	5′- AAAAGGGTTTTGATGAAGCTGTCGrCrGGATCCCACATTCTGrGGAAGACTTCAAATCCTTGTATTCT
TRP5.D	65	5′- AAAAGGGTTTTGATGAAGCTGTCGCGGATCCCACATTCTGGGAAGACTTCAAATCCTTGTATTCT
TRP5.R1$_{S1}$	65	5′- AAAAGGGTTTTGATGAAGCTGTCGCCGATCCCACATTCTGrGGAAGACTTCAAATCCTTGTATTCT
TRP5.R2$_{I2}$	65	5′- AAAAGGGTTTTGATGAAGCTGTCGrCrGGATCCCACATTCTGGGAAGACTTCAAATCCTTGTATTCT
TRP5.Dcc	65	5′- AAAAGGGTTTTGATGAAGCTGTCGCCGATCCCACATTCTGGGAAGACTTCAAATCCTTGTATTCT
TRP5.72D	72	5′-AAGAGAGTTGGAAAAGGGTTTTGATGAAGCTGTCGCGGATCCCACATTCTGGGAAGACTTCAAATCCTTGTA
TRP5.72Dcom	72	5′-TACAAGGATTTGAAGTCTTCCCAGAATGTGGGATCCGCGACAGCTTCATCAAAACCCTTTTCCAACTCTCTT
LEU2.R1dw	60	5′-TTAGGTGCTGTGGGTGGTCCTAAATGGGGATCCGGTAGTGTTAGrGCCTGAACAAGGTTTA
LacZ.R1.47	47	5′-CCCGAGTGTGATCATCTGGTCGCTGGGGAATrGAGTCAGGCCACGGCG
LacZ.R6.47	47	5′-CCCGAGTGTGATCATrCrUrGrGrUrCGCTGGGGAATGAGTCAGGCCACGGCG
LacZ.R2.47$_{I2}$	47	5′-CCCGAGTGTGATCATCTrGrGTCGCTGGGGAATGAGTCAGGCCACGGCG
LacZ.R5.47	47	5′-CCCGAGTGTGATCATCTGGTCGCTGGGGArArUrGrArGTCAGGCCACGGCG
LacZ.D.47	47	5′-CCCGAGTGTGATCATCTGGTCGCTGGGGAATGAGTCAGGCCACGGCG
LacZ.comD.wt.47	47	5′-CGCCGTGGCCTGACTCATTCCCCAGCGACCAGATGATCACACTCGGG
LacZ.comD.m1.47	47	5′-CGCCGTGGCCTGACTTATTCCCCAGCGACCAGATGATCACACTCGGG

LacZ.comD.del2.45	45	5′-CGCCGTGGCCTGACTCATTCCCCAGCGAAGATGATCACACTCGGG
LEU2.D	60	5’-TTAGGTGCTGTGGGTGGTCCTAAATGGGGATCCGGTAGTGTTAGGCCTGAACAAGGTTTA
LEU2.rG	60	5’-TTAGGTGCTGTGGGTGGTCCTAAATGGGGATCCGGTAGTrGTTAGGCCTGAACAAGGTTTA
LEU2.dU	60	5’-TTAGGTGCTGTGGGTGGTCCTAAATGGGGATCCGGTAGTGUTAGGCCTGAACAAGGTTTA
LEU2.rU	60	5’-TTAGGTGCTGTGGGTGGTCCTAAATGGGGATCCGGTAGTGrUTAGGCCTGAACAAGGTTTA
LEU2.3	20	5’-ATGTCTGCCCCTAAGAAGAT
LEU2.3	20	5’-TGCCAAAGAATAAGGTCAAC

In the name of the RNA-containing oligos, substitutions are indicated by a subscript capital “S” and insertions by a subscript capital “I”. The letters “S” and “I” are followed by a subscript number indicating the number of bases that are substituted or inserted, respectively. The structures of the oligos used in this study are described from the 5′end, with DNA sequences (D) shown in blue and RNA sequences (R) in red. Bases with homology to the chromosomal DNA are shown in brackets and underlined. Insertions are indicated as “ins::”. Base substitutions are indicated as “s”. In the oligo sequences, the base changes introduced by the oligos are in bold and shown in blue, if these consist of DNA bases and in red, with lower case “r” on the left side, if these consist of RNA bases. The RNA tracts of the non-specific LEU2.R1dw oligo and of the *lacZ* oligos used only in the RNase HII cleavage assay are shown in regular red type. The restriction sites that are introduced by the oligos are underlined (thin underline for *Van*91I, thick line for BamHI). The DNA oligos were desalted (synthesized by Invitrogen or Alpha DNA); the RNA-containing oligos were desalted and deprotected (by Dharmacon). The LacZ.R1$_{S1}$ and all the lacZ 45- and 47-mers were also PAGE purified (by Dharmacon or Invitrogen).

Table A.3 Statistical comparisons (*P* values) between gene correction frequencies obtained for different oligos in different genetic backgrounds.

a

I)

Oligo	*mutS vs. mutS rnhA*	*mutS vs. mutS rnhB*	*mutS vs. mutS rnhA rnhB*	*mutS rnhA vs. mutS rnhA rnhB*	*mutS rnhB vs. mutS rnhA rnhB*
LacZ.R6$_{I2}$	0.0960	0.0012	0.0021	0.0021	0.0034
LacZ.R2.47$_{I2}$	ND	0.6857	ND	ND	ND

LacZ.R1$_{S1}$	0.0032	0.0021	0.0034	0.0034	0.6166
LacZ.R5$_{S1}$	0.0050	0.7000	0.0022	0.0050	0.0022

II)

Oligo	*mutS*	*mutS rnhA*	*mutS rnhB*	*mutS rnhA rnhB*
LacZ.R6$_{I2}$ *vs*. LacZ.R2.47$_{I2}$	0.0061	ND	0.7619	ND
LacZ.R6$_{I2}$ *vs*. LacZ.R1$_{S1}$	0.0021	0.0021	0.0034	0.0034
LacZ.R6$_{I2}$ *vs*. LacZ.R5$_{S1}$	0.0167	0.0214	0.0238	0.0034
LacZ.R1$_{S1}$ *vs*. LacZ.R5$_{S1}$	0.0222	0.0218	0.0222	0.0022
LacZ.R1$_{S1}$ *vs*. LacZ.R2.47$_{I2}$	0.0106	ND	0.0106	ND
LacZ.R5$_{S1}$ *vs*. LacZ.R2.47$_{I2}$	0.0095	ND	0.0095	ND

b

I)

Oligo	WT *vs. rnhA*	WT *vs. rnhB*	WT *vs. rnhA rnhB*	*rnhA vs. rnhA rnhB*	*rnhB vs. rnhA rnhB*
LacZ.R1$_{S1}$	0.0286	0.0286	0.0286	0.0286	0.0286
LacZ.R5$_{S1}$	0.0294	0.0294	0.0294	0.0286	0.0286
LacZ.D	0.0571	0.1143	0.0286	0.3429	0.0286

Oligo	*mutS vs. rnhA mutS*	*mutS vs. rnhB mutS*	*mutS vs. rnhA rnhB mutS*	*rnhA mutS vs. rnhA rnhB mutS*	*rnhB mutS vs. rnhA rnhB mutS*
LacZ.R1$_{S1}$	0.0286	0.0286	0.0286	0.0286	0.2
LacZ.R5$_{S1}$	0.0286	0.6857	0.0286	0.0286	0.0286
LacZ.D	1.0000	0.1143	0.1143	0.1143	0.2

Oligo	WT *vs. mutS*	*rnhA vs. rnhA mutS*	*rnhB vs. rnhB mutS*	*rnhA rnhB vs. rnhA rnhB mutS*
LacZ.R1$_{S1}$	0.0286	0.0286	0.0286	0.0286
LacZ.R5$_{S1}$	1.0000	1.0000	0.3429	1.0000
LacZ.D	0.0286	0.0286	0.0286	0.0286

II)

Oligo	WT	*rnhA*	*rnhB*	*rnhA rnhB*

No oligo *vs.* LacZ.R1$_{S1}$	0.0286	0.0286	0.0286	0.0286
No oligo *vs.* LacZ.R5$_{S1}$	1	0.0286	0.1143	0.0286

c

I)

Oligo	WT *vs. rnhB*	WT *vs. mutS*	WT *vs. rnhB mutS*	*rnhB vs. rnhB mutS*	*mutS vs. rnhB mutS*
RpsL.R1$_{S1}$	0.029	0.029	0.029	0.029	0.029
RpsL.D	0. 171	0.010	0.029	0.010	0. 038

II)

Oligo	WT	*rnhB*	*mutS*	*rnhB mutS*
No oligo *vs*. RpsL.R1$_{S1}$	0.021	0.027	0.029	0.029
No oligo *vs.* RpsL.D	0.021	0.013	0.010	0.029
RspL.R1$_{S1}$ *vs*. RpsL.D	0.486	0.010	0.010	0.486

d

I)

Oligo	WT *vs. msh2*	WT *vs. rnh201*	WT *vs. rnh201 msh2*	*msh2 vs. rnh201 msh2*	*rnh201vs. rnh201 msh2*
TRP5.R2_R1$_{I2_S2}$	0.008	0.082	0.001	0.001	0.001
TRP5.D	0.001	0. 875	0.001	0.083	0.001
Non-specific oligo	ND	ND	0.197	ND	ND

II)

Oligo	WT	*msh2*	*rnh201*	*rnh201 msh2*
No oligo *vs*.TRP5.R2_R1$_{I2_S2}$	0.004	0.001	<0.001	0.001
No oligo *vs*. TRP5.D	0.001	<0.001	<0.001	0.001
TRP5.R2_R1$_{I2_S2}$ *vs*. TRP5.D	0.023	0.002	0. 267	0.052
No oligo *vs.* non-specific oligo	ND	ND	ND	0.814

e

I)

Oligo	WT *vs. msh2*	WT *vs. rnh201*	WT *vs. rnh201 msh2*	*msh2 vs. rnh201 msh2*	*rnh201 vs. rnh201 msh2*
TRP5.R2$_{I2}$	0. 029	0.029	0.029	0.029	0.029
TRP5.D	1	0.057	0.029	0.342	0.029

II)

Oligo	WT	*msh2*	*rnh201*	*rnh201 msh2*
TRP5.R2$_{I2}$ *vs.* TRP5.D	0.029	0.029	0.057	0.029

f

I)

Oligo	WT *vs. msh2*
TRP5.72D	0.029
TRP5.72Dcomp	0.029
No oligo	0.027

II)

Oligo	WT	*msh2*
No oligo *vs*. TRP5.72D	0.027	0.029
No oligo *vs.* TRP5.72Dcomp	0.027	0.029
TRP5.72D *vs.* TRP5.72Dcomp	0.029	0.886

g

I)

Oligo	WT *vs*. *msh2*	WT *vs*. *rnh201*	WT *vs*. *rnh201 msh2*	*msh2 vs*. *rnh201 msh2*	*rnh201 vs*. *rnh201 msh2*
TRP5.R1$_{S1}$	0.057	0. 486	0.029	0. 114	0.029
TRP5.Dcc	0.029	0. 486	0.029	0. 343	0.029

II)

Oligo	WT	*msh2*	*rnh201*	*rnh201 msh2*
No oligo *vs*.TRP5.R1$_{S1}$	0.028	0.029	0.028	0.029
No oligo *vs*. TRP5.Dcc	0.029	0.029	0.029	0.029
TRP5.R1$_{S1}$ *vs*. TRP5.Dcc	0.307	0.029	0.029	1.000

h

I)

Oligo	WT *vs. msh2*	WT *vs. rnh201*	WT *vs*. *rnh201 msh2*	*msh2 vs*. *rnh201 msh2*	*rnh201 vs*. *rnh201 msh2*
TRP5.R2$_{I2}$	0. 114	0.029	0.029	0.029	0.029
TRP5.D	0.029	0.057	0.029	0.2	0.029

II)

Oligo	WT	*msh2*	*rnh201*	*rnh201 msh2*
No oligo *vs*.TRP5.R2$_{I2}$	0.041	0.029	0.029	0.029
No oligo *vs.* TRP5.D	0.029	0.029	0.029	0.029
TRP5.R2$_{I2}$ *vs*. TRP5.D	0.029	0.029	0.057	0.029

i

Oligo	BW-2028 *vs.* YSB-13, 14	BW-2029 *vs.* YSB-15, 16
LacZ.R6$_{I2}$	1	0.548
LacZ.R1$_{S1}$	0.109	0.647

j

Oligo	BW2039 *vs*. YSB-18
LacZ.R1$_{S1}$	0.710
LacZ.D	0.111
No oligo	0.114

Mann-Whitney *U*-test was applied to determine whether a difference exists between various pairs of gene correction frequencies. Comparison of frequencies presented in (**a**) **Table 2.1**, (**b**) **Table 2.2**, (**c**) **Table 2.3**, (**d**) **Table 2.4a**, (**e**) **Table 2.4b**, (**f**) **Figure A.2**, (**g**) **Table A.4a**, (**h**) **Table A.4b**, (**i**) **Table A.5a**, and (**j**) **Table A.5b**. Two groups in a pair were considered to be significantly different when *P* values were less than 0.05. I) Comparisons were between relative frequencies obtained for each oligo in different backgrounds, and II) between relative frequencies obtained for different oligos in the same backgrounds. ND, not determined.

Table A.4 Reversion frequency of a nonsense mutation or a two-base deletion in the yeast *trp5* gene following transformation by rNMP-containing oligos in MMR and RNase H mutant cells

a

Oligo	WT	*msh2*	*rnh201*	*rnh201 msh2*

TRP5.R1$_{S1}$	9.01	(8.23–11.9)	31.4[a]	(20.6–43.7)	11.2	(9.23–12.6)	84.7	(35.1–130)
TRP5.Dcc	7.79	(5.46–11.9)	56.4	(51.4–69.1)	7.52	(5.93–7.84)	78.6	(54.6–90.9)
No oligo	0.68	(0–0.99)	10.7	(8.6–19.3)	0.68	(0.37–1.65)	14.8	(8.2–20.7)

b

Oligo	WT		*msh2*		*rnh201*		*rnh201 msh2*	
TRP5.R2$_{I2}$	3.54 **46.7%**	(1.24–5.35)	12.6[b] **60.6%**	(8.18–16.2)	9.64 **100%**	(7.51–15.0)	35.5 **75.7%**	(27.1–46.6)
TRP5.D	39.1 **100%**	(35.9–43.6)	52.3 **88.2%**	(27.6–87.0)	27.7 **100%**	(12.3–36.9)	84.6 **82.8%**	(76.5–105)
No oligo	0.38 **0.0%**	(0–1.24)	5.27 **0.0%**	(2.01–7.24)	0.28 **0.0%**	(0–1.70)	6.53 **0.0%**	(3.73–9.80)

c

Pattern	Sequence
Chr	AAAAGGGTTTTGATGAAGCTGTCG--GATCCCACATTCTGGGAAGACTTCAAATCCTTGTATTCT
R	AAAAGGGTTTTGATGAAGCTGTCG**CG**GATCCCACATTCTGGGAAGACTTCAAATCCTTGTATTCT
R1	AAAAGGGTTTTGATGAAGCTGTCG**CGA**ATCCCACATTCTGGGAAGACTTCAAATCCTTGTATTCT
R2	AAAAGGGTTTTGATGAAGCTGTCG**CG**GA**C**CCCACATTCTGGGAAGACTTCAAATCCTTGTATTCT
A	AAAAGGGTTT-GATGAAGCTGTCG--GATCCCACATTCTGGGAAGACTTCAAATCCTTGTATTCT
B	AAAAGGGTTTTGATGAAGCTGTC---GATCCCACATTCTGGGAAGACTTCAAATCCTTGTATTCT
C	AAAAGGGTTTTGATGAAGCTGTCG--GATCC-ACATTCTGGGAAGACTTCAAATCCTTGTATTCT
D	AAAAGGGTTTTGATGAAGCTGTCG--GATCCCACATTCTGG-AAGACTTCAAATCCTTGTATTCT
E	AAAAGGGTTTTGATGAAGCTGTCG--GATCCCA-ATTCTGGGAAGACTTCAAATCCTTGTATTCT
F	AAAAGGGTTTTGATGAAGCTGTC--GGATCCCACATTCTGGGAAGACTTCAAATCCTTG-ATTCT
G	AAAAGGGTTTTGATGAA----TC**T**--GATCCCACATTCTGGGAAGACTTCAAATCCTTGTATTCT
H	AAAAGGGTTTTGATGAAGCTGTCG--G-TCCCACATTCTGGGAAGACTTCAAATCCTTGTATTCT
I	AAAAGGGTTTT**TG**GATGAAGCTGTCG--GATCCCACATTCTGGGAAGACTTCAAATCCTTGTATTCT
J	AAAAGGGTTTTGATGAAGCTGTCG--GATCCCACATA-TGGGAAGACTTCAAATCCTTGTATTCT

d

Oligo	WT		*msh2*		*rnh201*		*rnh201 msh2*	
No oligo	R	0/3	R	0/14	R	0/4	R	0/18
	A	1/3	A	10/14	A	2/4	A	11/18
	C	1/3	D	3/14	B	1/4	C	2/18
	D	1/3	E	1/14	C	1/4	E	5/18
TRP5.R2$_{I2}$	R	7/15	R	20/33	R	65/65	R	115/152
	B	7/15	A	9/33			R1	1/152
	H	1/15	B	1/33			R2	1/152
			C	1/33			A	28/152
			E	1/33			C	1/152
			I	1/33			D	2/152
							F	1/152
							G	2/152

							J	1/152
TRP5.D	R	28/28	R	15/17	R	30/30	R	24/29
			A	2/17			A	5/29

(**a**) Median and range (in parentheses) of Trp^+ transformant colonies per 10^7 viable wild-type, MMR, and RNase H mutant cells, obtained to revert the nonsense mutation, or (**b**) the two-base deletion in *trp5*. The number of repeats for each of the strains transformed with these oligos was 4. Results obtained for oligos TRP5.R1$_{S1}$ and TRP5.R2$_{I2}$ were not due to contamination with DNA oligos. The significance of all non-overlapping CI or range values was confirmed by Mann-Whitney test ($P < 0.05$) after subtraction of the no oligo background values. Comparison of frequencies is presented in **Tables A.3g**,**h**. In panel **b**, the percentage of clones with precise correction of the "CG" deletion (pattern R in panel **c**) is shown in bold under the median. The median and these percentages were used to calculate the frequency of Trp^+ transformant colonies with precise correction of the two-base deletion mutation presented in (**Table 2.4b**). (**c**) Sequence patterns of the *TRP5* region in Trp^+ colonies transformed with no oligo, TRP5.R2$_{I2}$, or TRP5.D oligos. Differences from the chromosomal sequence are in bold. Chr, chromosomal sequence of *trp5* with the 2-base deletion mutation; R, sequence of the *TRP5* region targeted with oligo TRP5.R2$_{I2}$ or TRP5.D, containing the "CG" insertion; R1, R2, the *TRP5* region corrected by oligos presenting additional mutations; A-J, patterns of the *TRP5* region from spontaneous Trp^+ revertants. (**d**) The frequency of each sequence pattern obtained in WT, *msh2*, *rnh201*, and *rnh201 msh2* colonies following sequence analysis of 408 Trp^+ clones from the experiment shown in panel **b** above. These data were used to generate results presented in **Table 2.4b**.

[a]The gene correction frequency value obtained in the *msh2* mutant cells was different from that obtained in the wild-type cells at the significant level 0.1 ($P = 0.057$) after subtraction of the no-oligo background value.

[b]The gene correction frequency value obtained in the *msh2* mutant cells was not significantly different from that obtained in the wild-type cells ($P = 0.114$) after

subtraction of the no oligo background value. However, after sequencing all the Trp^+ clones of this experiment, the frequency value of precise gene correction obtained in the *msh2* mutant cells was significantly different from that obtained in the wild-type cells ($P = 0.029$, see **Table 2.4b** and **A.3e**).

Table A.5 Deletion and nonsense mutations of *rnhA* and *rnhB* genes similarly affect gene correction by oligos in the BW1988 and the BW2037 backgrounds.

a

	LacZ.R6$_{I2}$		LacZ.R1$_{S1}$	
***rnhA* (BW2028)**	0.04	(0.02–0.05)	9×10^{-4}	(7×10^{-4}–1.4×10^{-3})
***rnhA* (YSB-13, 14)**	0.04	(0.02–0.05)	7×10^{-4}	(5.5×10^{-4}–9.1×10^{-4})
***rnhB* (BW2029)**	0.16	(0.08–0.19)	3.43	(1.74–3.87)
***rnhB* (YSB-15, 16)**	0.12	(0.09–0.17)	3.07	(2.67–3.97)

b

	LacZ. R1$_{S1}$		LacZ.D	No oligo
Genotype	Lac^+ freq.	Rel. tr. freq.[a]	Lac^+ freq.	Lac^+ freq.
***rnhB mutS* (BW2039)**	17,500	0.2	86,700	3.98
	(6,250–30,000)		(45,000–103,000)	(2.23–25.4)
***rnhB mutS* (YSB-18)**	33,000	0.27	124,200	3.77
	(30,400–79,500)		(84,600–163,000)	(3.51–5.06)

(**a**) The strains used are BW2028, YSB-13, 14, BW2029, and YSB-15, 16, all of which are *mutS* mutants. All the strains were transformed with the LacZ.R6$_{I2}$, LacZ.R1$_{S1}$, or LacZ.D oligo. The values are relative transformation frequencies (see **Materials and Methods**). The numbers of repeats for each of strains transformed with these oligos were as follows: LacZ.R6$_{I2}$: 7,3,7,3; LacZ.R1$_{S1}$:6,3,7,3; and LacZ.D: 6,3,6,3. The significance of all non-overlapping confidence limit values was confirmed by the Mann-Whitney *U*-test ($P < 0.05$) (**Table A.3i**). (**b**) The strains used were BW2039 and YSB-18. The strains were transformed with LacZ.R1$_{S1}$ and LacZ.D oligos. The values are median and the range (in parentheses) of Lac^+ transformant colonies reverting the missense mutation per

10^7 cells. The numbers of repeats for each of the strains transformed with these oligos were 4. The significance of all non-overlapping confidence limit values was confirmed by the Mann-Whitney *U*-test ($P < 0.05$) (**Table A.3j**).

[a]Relative frequency of Lac^+ transformants (see **Materials and Methods**). In the absence of a transforming oligo, the number of Lac^+ transformants per 10^7 viable cells were <0.1.

Table A.6 Results of *leu2* DSB repair assay with rNMP-containing oligos.

a

Oligo	WT	*rnh201*	*rad14*	*rnh201 rad14*
LEU2.D	65% (55–75)	63% (55–65)	65% (50–75)	65% (55–70)
LEU2.rG	30% (20–40)	90% (75–100)	63% (55–70)	100% (80–100)

b

Oligo	WT	*rnh201*	*ung1*
LEU2.D	65% (55–75)	63% (55–65)	65% (55–70)
LEU2.dU	5.0% (0–10)	N/A	55% (45–60)
LEU2.rU	33% (25–45)	55% (45–65)	33% (30–40)

c

Oligo	*rnh201*	*rad14*	*rnh201 rad14*
LEU2.D	0.5357	1.0000	1.0000
LEU2.rG	0.0286	0.0286	0.0286

d

Oligo	*rnh201*	*ung1*
LEU2.D	0.5357	1.0000
LEU2.dU	N/A	0.0294
LEU2.rU	0.0421	1.0000

(**a** and **b**) Data shown in **Figure 2.3** are presented here as median percentages of StuI-cut Leu^+ transformants from four independent transformations and ranges in parentheses. For each transformation, 20 Leu^+ transformants were selected for analysis. (**c** and **d**) Mann-Whitney *U*-test was implemented for statistical analysis against the WT, and *P* values are displayed. N/A, not applicable because data are not available for comparison.

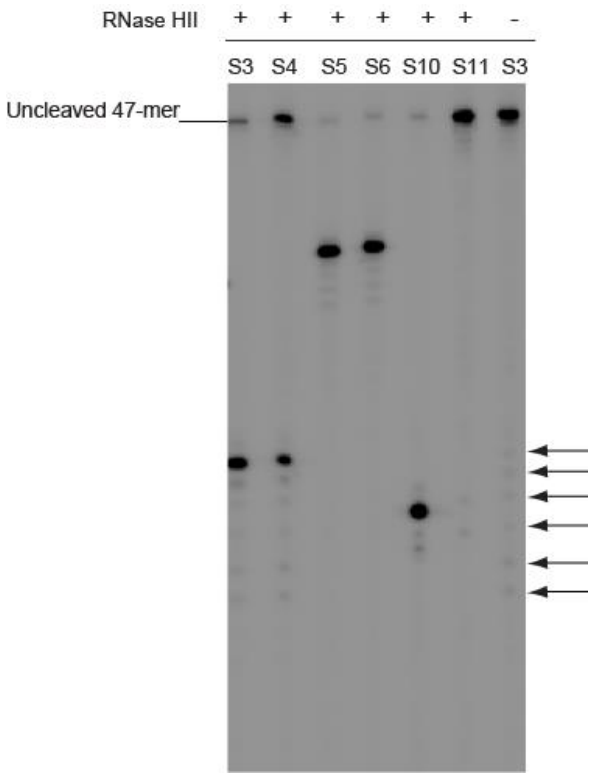
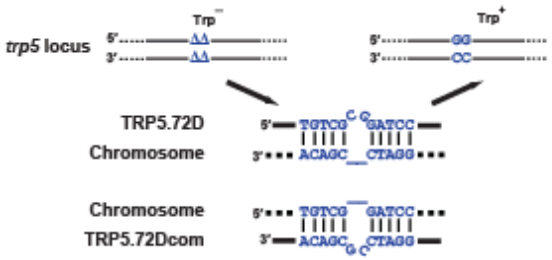


Figure A.1 PAGE gel showing fragments resulting from RNase HII cleavage and random RNA degradation in the absence of RNase HII. Structural presentation of the substrates is shown in **Figure 2.2a**. S3, the substrate that contains a stretch of 6 fully complementary rNMPs, was used to show fragments resulting from random degradation without RNase HII. The position of the uncleaved 47-mers is shown to the left. The positions of the random cleavage of the S3 substrate in the absence of RNase HII are indicated by arrows to the right. These barely detectable bands are similar to those seen for the other substrates containing more than one rNMP.

a



b

Genotype	TRP5.72D		TRP5.72Dcom		No oligo	
WT	28.2	(26.6–29.6)	10.8	(8.98–11.5)	<0.1	(0–0.17)
msh2	76	(62.4–99.0)	67.9	(60.5–89.9)	5.27	(2.01–7.24)

Figure A.2 Strand-bias targeting effect of two complementary oligos, TRP5.72D and TRP5.72com. The effect is mild in wild-type and cannot be observed in *msh2* mutant strains. (a) Diagram and sequence of the chromosomal *trp5* region targeted by the oligos. (b) Median and range of yeast Trp^+ transformant colonies reverting the two-base deletion mutation per 10^7 viable wild-type or MMR mutant cells. The significance of all non-overlapping range value was confirmed by Mann-Whitney *U*-test ($P < 0.05$) after subtraction of the no oligo background values. Comparison of frequencies is presented in **Table A.3f**. The number of repeats for each of the strains transformed with the oligos was 4.

APPENDIX B

SUPPLEMENTARY MATERIALS FOR CHAPTER 3

Table B.1 Sequences of synthetic oligos used in this study.

Name	Sequence
Sequence 1 dG	5'-Bi-CAGGTTCACGATGGAGCTCTCGATTCAGCT-SH-3'
Sequence 1 rG	5'-Bi-CAGrGTTCACrGATGGArGCTCTCrGATTCArGCT-SH-3'
Sequence 1 compl_DNA	5'-SH-AGCTGAATCGAGAGCTCCATCGTGAACCTG-Bi-3'
Sequence 2 dG	5'-Bi-ATCCGGTAGTGTTAGGCCTGAACAAGGTTT-SH-3'
Sequence 2 rG	5'-Bi-ATCCrGGTAGTrGTTAGrGCCTrGAACAArGGTTT-SH-3'
Sequence 2 compl_DNA	5'-SH-AAACCTTGTTCAGGCCTAACACTACCGGAT-Bi-3'
dG_III	5'-ATGGAGCTC-3'
rG_III	5'-ATGGArGCTC-3'
compl_DNA_III	5'-GAGCTCCAT-3'
dG_VI	5'-ATCCGGTAG-3'
rG_VI	5'-ATCCrGGTAG-3'
compl_DNA_VI	5'-CTACCGGAT-3'
dG_VIII	5'-TTAGGCCTG-3'
rG_VIII	5'-TTAGrGCCTG-3'
compl_DNA_VIII	5'-CAGGCCTAA-3'

Biotin and thiol group modifications are indicated by "Bi" and "SH," respectively. dNMPs are shown in blue while rNMPs are shown in red, preceded by letter "r." 30-mers were used in AFM experiments while 9-mers were used in NMR experiments. All oligos used in AFM experiments were PAGE-purified.

Table B.2 List of the optical lever sensitivity (w) and spring constant (k_N) of cantilevers used in the measurements.

Tips used for	Samples	w (nm/V)	k (N/m)
Sequence 1	dG-DNA Round 1	59.9±1.0	0.068±0.002
	rG-DNA Round 1	62.3±1.0	0.057±0.002
	dG-DNA Round 2 / rG-DNA Round 2 (Measured with the same tip)	50.0±1.2	0.053±0.002
	rG-DNA Round 3	53.5±1.7	0.415±0.002

	dG-DNA Round 1	53.1±1.3	0.045±0.002
Sequence 2	rG-DNA Round 1	48.0±1.0	0.051±0.004
	dG-DNA Round 2 / rG-DNA Round 2 (Measured with the same tip)	50.3±1.0	0.049±0.002

Table B.3 Mean values of all the parameters and stretch modulus of ss substrates with *Sequence 1*.

Substrate	Mean			
	L_0 (nm)	δ (nm)	F_{st} (pN)	S (pN)
dG	9.7 ± 3.6	3.0 ± 1.9	20.0 ± 11.1	84.9 ± 61.1
rG	13.7 ± 4.5	4.3 ± 2.0	26.2 ± 12.1	106.0 ± 84.7

Mean values are presented with standard deviation of the mean.

Table B.4 Mean values of all the parameters and stretch modulus of ss substrates with *Sequence 2*.

Substrate	Mean			
	L_0 (nm)	δ (nm)	F_{st} (pN)	S (pN)
dG	12.8 ± 3.4	5.2 ± 2.2	25.8 ± 9.3	69.7 ± 33.3
rG	12.0 ± 3.0	5.2 ± 1.9	22.7 ± 8.6	60.4 ± 38.8

Mean values are presented with standard deviation of the mean.

Table B.5 Comparison of Gaussian peak values and median values of stretch modulus of ss substrates.

Substrate	Gaussian Peak				Median
	L_0 (nm)	δ (nm)	F_{st} (pN)	S (pN)	S (pN)
Sequence 1 dG (n=108)	9.1±0.2	2.2±0.1	20.2±0.1	67.2 ± 4.3	68.1 (59.4 – 86.9)
Sequence 1 rG (n=75)	13.5±0.7	3.7±0.2	23.4±0.1	57.7 ± 4.8	72.4 (55.9 – 102.3)
Sequence 2 dG (n=52)	12.0±0.2	5.0±0.2	24.6±0.1	54.0 ± 1.7	59.0 (51.6 – 73.5)

Sequence 2 rG (n=132)	12.1±0.2	5.1±0.2	23.0±0.2	51.0 ± 0.6	50.2 (44.4 – 60.1)

Gaussian peak values of all the parameters are also listed. Gaussian fitted values are presented with standard error of fit while median values are presented with 99% confidence interval of the median.

Table B.6 Summary of the number of data population before the removal of outliers and the values of stretch modulus that are considered as outliers and removed.

Substrate		Data population before the removal of outliers	Number of outliers found	Stretch modulus removed as outliers (pN)
Double-strand (ds)	*Sequence 1* dG-DNA Round 1	52	2	437.0, 455.6
	Sequence 1 dG-DNA Round 2	112	5	2086.7,1943.5, 903.2, 955.9, 1391.1
	Sequence 1 rG-DNA Round 1	76	3	416.7, 500.63, 407.9
	Sequence 1 rG-DNA Round 2	72	3	551.6, 830.2, 615.2
	Sequence 1 rG-DNA Round 3	112	3	1243.3, 709.4, 761.8
	Sequence 2 dG-DNA Round 1	99	4	590.6, 786.5, 1068.0, 1035.9
	Sequence 2 dG-DNA Round 2	94	3	1037.6, 1679.3, 2258.0
	Sequence 2 rG-DNA Round 1	84	5	662.1, 550.1, 989.5, 600.2, 697.8
	Sequence 2 rG-DNA Round 2	64	4	537.9, 520.1, 552.4, 647.8
Single-strand (ss)	*Sequence 1* dG	113	5	662.1, 550.1, 898.5, 600.2, 697.8
	Sequence 1 rG	78	3	399.8, 384.2, 408.4
	Sequence 2 dG	55	3	352.2, 457.6, 338.2
	Sequence 2 rG	140	8	444.7, 491.9, 432.1, 406.6, 297.4, 298.2, 295.5, 280.2

See **Materials and Methods** for details of the outlier test.

Table B.7 Thermal stability of 30-bp duplexes used in AFM experiments.

	T_m (°C)
Sequence 1 dG-DNA	73.6

Sequence 1 rG-DNA	70.2
Sequence 2 dG-DNA	69.6
Sequence 2 rG-DNA	66.2

Table B.8 Mean values of all the parameters and stretch modulus of ds substrates with *Sequence 1*.

Substrate	Mean			
	L_0 (nm)	δ (nm)	F_{st} (pN)	S (pN)
dG-DNA (n=157)	10.0 ± 3.0	3.0 ± 1.8	34.2 ± 21.3	125.6 ± 83.3
rG-DNA (n=251)	11.4 ± 3.7	4.2 ± 2.1	30.9 ± 14.2	94.4 ± 67.7

Mean values are presented with standard deviation of the mean.

Table B.9 Mean values of all the parameters and stretch modulus of ds substrates with *Sequence 2*.

Substrate	Mean			
	L_0 (nm)	δ (nm)	F_{st} (pN)	S (pN)
dG-DNA (n=186)	10.5 ± 4.0	3.3 ± 1.6	28.9 ± 20.1	106.4 ± 97.0
rG-DNA (n=139)	11.4 ± 3.9	3.2 ± 2.1	31.2 ± 14.5	147.5 ± 112.1

Mean values are presented with standard deviation of the mean.

Table B.10 Comparison of Gaussian peak values and median values of stretch modulus of ds substrates.

Substrate	S (pN)	
	Gaussian Peak	Median
Sequence 1 dG-DNA Round 1	104.5 ± 3.6	111.6 (99.3 – 147.4)
Sequence 1 dG-DNA Round 2	104.2 ± 2.3	107.4 (91.3 – 121.3)
Sequence 1 rG-DNA Round 1	62.8 ± 3.0	70.4 (56.3 – 84.6)
Sequence 1 rG-DNA Round 2	79.1 ± 4.7	94.5 (67.2 – 108.7)
Sequence 1 rG-DNA Round 3	72.0 ± 3.3	73.6 (65.7 – 88.4)

Sequence 2 dG-DNA Round 1	76.5 ± 2.3	84.4 (65.2 – 98.8)
Sequence 2 dG-DNA Round 2	68.4 ± 4.3	83.2 (71.8 – 99.1)
Sequence 2 rG-DNA Round 1	85.8 ± 4.1	102.5 (73.4 – 120.1)
Sequence 2 rG-DNA Round 2	95.4 ± 5.4	119.9 (93.5 – 181.6)

Gaussian fitted values are presented with standard error of fit while median values are presented with 99% confidence interval of the median.

Table B.11 Summary of *P* values of all combined data for *Sequence 1*.

Substrate	dG-DNA	rG-DNA	dG	rG
dG-DNA	–	< 0.0001	< 0.0001	0.0008
rG-DNA	< 0.0001	–	0.1357	0.9233
dG	< 0.0001	0.1357	–	0.4098
rG	0.0008	0.9233	0.4098	–

Mann-Whitney *U*-test was performed to obtain the *P* values.

Table B.12 Summary of *P* values of all combined data for *Sequence 2*.

Substrate	dG-DNA	rG-DNA	dG	rG
dG-DNA	–	< 0.0001	0.0008	< 0.0001
rG-DNA	< 0.0001	–	< 0.0001	< 0.0001
dG	0.0008	< 0.0001	–	0.0138
rG	< 0.0001	< 0.0001	0.0138	–

Mann-Whitney *U*-test was performed to obtain the *P* values.

Table B.13 Summary of *P* values of each individual round of measurements of ds substrates with *Sequence 1*.

Substrate	dG-DNA Round 1	dG-DNA Round 2	rG-DNA Round 1	rG-DNA Round 2	rG-DNA Round 3
dG-DNA Round 1	–	0.0425	< 0.0001	0.0041	< 0.0001
dG-DNA Round 2	0.0425	–	< 0.0001	0.1352	< 0.0001
rG-DNA Round 1	< 0.0001	< 0.0001	–	0.0143	0.2921
rG-DNA Round 2	0.0041	0.1352	0.0143	–	0.0855
rG-DNA	< 0.0001	< 0.0001	0.2921	0.0855	–

Round 3				

Mann-Whitney *U*-test was performed to obtain the *P* values.

Table B.14 Summary of *P* values of each individual round of measurements of ds substrates with *Sequence 2*.

Substrate	dG-DNA Round 1	dG-DNA Round 2	rG-DNA Round 1	rG-DNA Round 2
dG-DNA Round 1	–	0.5045	0.0326	0.0001
dG-DNA Round 2	0.5045	–	0.0726	< 0.0001
rG-DNA Round 1	0.0326	0.0726	–	0.0370
rG-DNA Round 2	0.0001	< 0.0001	0.0370	–

Mann-Whitney *U*-test was performed to obtain the *P* values.

Table B.15 Thermal stability of an rGMP-containing 9-bp duplex and its DNA-control.

	Enthalpy (kJ/mol)	Entropy (kJ/mol)	T_m (K)
ATGGArGCTC	241 ± 6	0.661	317.6
DNA-control	261 ± 9	0.733	314.9

The buffer condition was 100 mM NaCl, 10 mM phosphate, 0.5 mM EDTA at pH 6.6. The duplex concentration was 30 μM.

Table B.16 Imino proton NMR chemical shift data for three rGMP-containing 9-bp duplexes, ATGGArGCTC (with rGMP III), ATCCrGGTAG (with rGMP VI), and TTAGrGCCTG (with rGMP VIII), and their DNA-controls.

Base	δ (ppm)		Δδ (ppm)
	DNA-control	ATGGArGCTC	
T18	13.03	12.98	0.05
T2	13.87	13.82	0.05
G3	12.90	12.84	0.06
G4	12.92	12.66	0.26
T14	13.92	13.81	0.11
G6/rG6	12.76	12.90	-0.14
G12	12.90	13.05	-0.15
T8	14.20	14.14	0.06

G10	12.81	12.80	0.01

Base	δ (ppm)		Δδ (ppm)
	DNA-control	ATCCrGGTAG	
T18	13.36	13.36	0.00
T2	13.87	13.88	-0.01
G16	12.79	12.79	0.00
G15	13.06	12.95	0.11
G5/rG5	13.06	13.00	0.06
G6	12.82	12.84	-0.02
T7	13.72	13.64	0.08
T11	13.85	13.88	-0.03
G9	13.16	13.19	-0.03

Base	δ (ppm)		Δδ (ppm)
	DNA-control	TTAGrGCCTG	
T1	-	-	-
T2	13.59	13.69	-0.10
T16	13.84	13.82	0.02
G4	12.93	12.86	0.07
G5/rG5	12.96	13.06	-0.10
G13	12.99	13.11	-0.12
G12	12.95	13.00	-0.05
T8	14.19	14.12	0.07
G9	12.98	12.92	0.05

Spectra were recorded of 1.1 mM duplexes in 100 mM NaCl, 10 mM phosphate, 10% D_2O buffer (pH 6.4) at 280K using the solvent suppression jump and return pulse program.

Table B.17 ^{31}P NMR chemical shift data for three rGMP-containing 9-bp duplexes, ATGGArGCTC (with rGMP III), ATCCrGGTAG (with rGMP VI), and TTAGrGCCTG (with rGMP VIII), and their DNA-controls at 294K.

Nucleotide	^{31}P δ (ppm)		Δδ (ppm)
	DNA-control	ATGGArGCTC	
A1	–	–	–
T2	-0.68	-0.68	0.00
G3	-0.24	-0.19	0.05
G4	-0.32	-0.31	0.01
A5	-0.49	-0.39	0.10
G6/rG6	-0.64	-0.40	0.24
C7	-0.30	0.50	0.80
T8	-0.83	-1.12	-0.29
C9	-0.42	-0.47	-0.05
G10	–	–	–
A11	-0.48	-0.53	-0.05
G12	-0.61	-0.57	0.04

C13	-0.34	-0.45	-0.11
T14	-0.86	-0.92	-0.05
C15	-0.68	-0.70	-0.02
C16	-0.38	-0.36	0.02
A17	-0.28	-0.23	0.05
T18	-0.61	-0.60	0.01

Nucleotide	^{31}P δ (ppm)		Δδ (ppm)
	DNA-control	ATCCrGGTAG	
A1	-	-	-
T2	-0.77	-0.77	0.00
C3	-0.65	-0.63	0.02
C4	-0.51	-0.42	0.09
G5/rG5	-0.36	-0.50	-0.14
G6	-0.43	0.85	1.28
T7	-0.80	-1.07	-0.27
A8	-0.59	-0.69	-0.10
G9	-0.53	-0.54	-0.01
C10	-	-	-
T11	-0.73	-0.73	0.00
A12	-0.51	-0.52	-0.01
C13	-0.66	-0.66	0.00
C14	-0.54	-0.50	0.04
G15	-0.43	-0.40	0.03
G16	-0.43	-0.48	-0.05
A17	-0.43	-0.48	-0.05
T18	-0.67	-0.67	0.00

Nucleotide	^{31}P δ (ppm)		Δδ (ppm)
	DNA-control	TTAGrGCCTG	
T1	-	-	-
T2	-0.72	-0.79	-0.07
A3	-0.47	-0.54	-0.07
G4	-0.50	-0.50	0.00
G5/rG5	-0.38	0.02	0.40
C6	*	-0.44	*
C7	*	-0.87	*
T8	-0.70	-1.02	-0.32
G9	-0.43	-0.54	-0.11
C10	-	-	-
A11	-0.43	-0.53	-0.10
G12	-0.50	-0.61	-0.11
G13	-0.40	-0.36	0.04
C14	-0.57	-0.67	-0.10
C15	-0.55	-0.69	-0.14
T16	-0.79	-0.94	-0.15
A17	-0.61	-0.62	-0.01
A18	-0.57	-0.60	-0.03

The phosphorous resonances are a good indicator on the status of the nucleic acid backbone. Shown here, large deviations in chemical shift between rGMP-containing duplexes and their DNA-controls suggest localized perturbations in the backbone 3′ of the damage site on both the top and bottom strand.

Note: 5′ nucleotides A1 and G10 do not have phosphate groups.

*Denotes uncertainty; ^{31}P resonances for nucleotides C6 and C7 were in the range of -0.47 to -0.74 ppm.

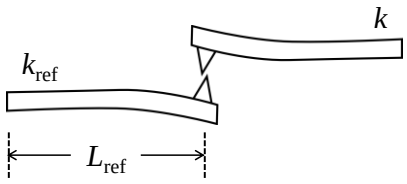


Figure B.1 Schematic of the reference beam method for the calibration of cantilever spring constant.

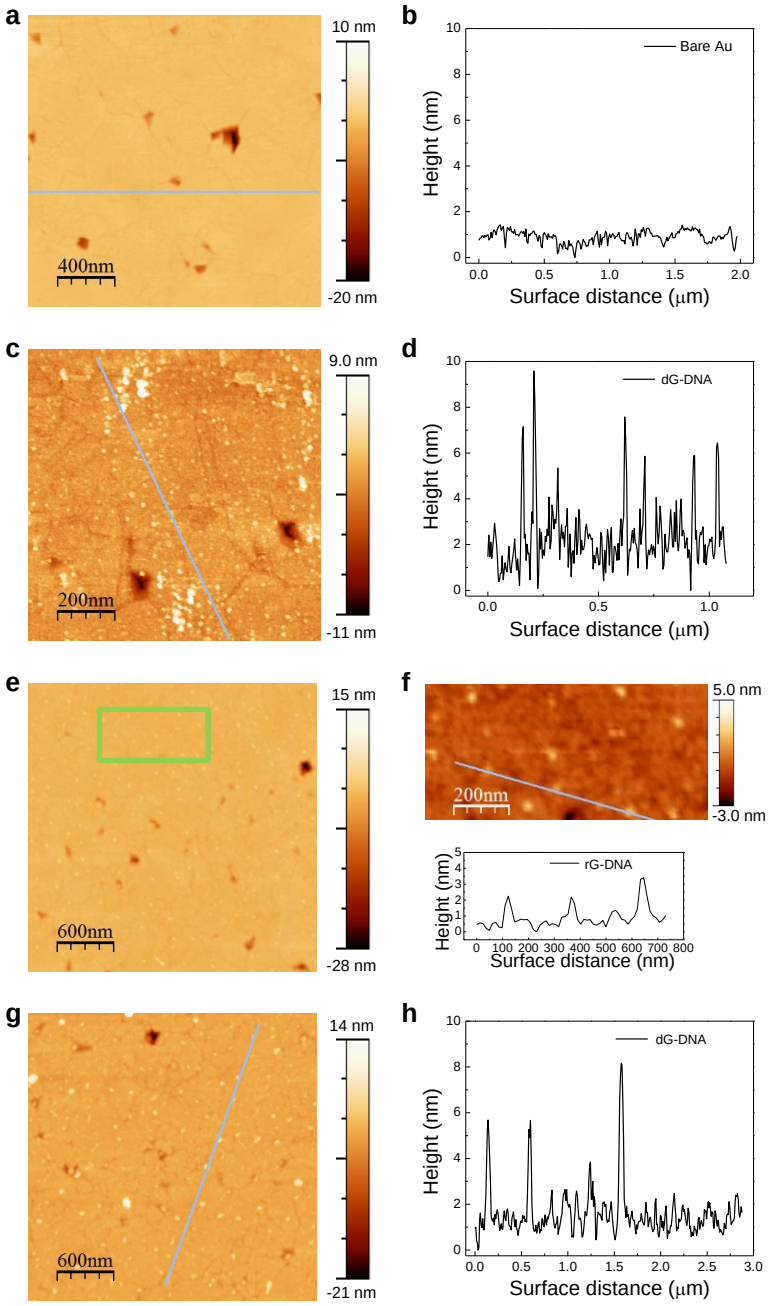


Figure B.2 Image and profile of gold surfaces. (**a**) AFM image of a fresh gold surface. (**b**) Surface profile of the gold surface indicated by the blue line in **a**. (**c**) A typical image of dG-DNA molecules of *Sequence 1* attached on gold surface in liquid. (**d**) Height profile of the blue line indicated in **c** shows that DNAs are also standing up on the surface. (**e**) A typical image of rG-DNA molecules of *Sequence 2* attached on gold surface in air. (**f**) Top: Zoom-in image in **e** indicated by the green box; Bottom: Height profile of the

blue line indicated in the top panel. (**g**) A typical image of dG-DNA molecules of *Sequence 2* attached on gold surface in air. (**h**) Height profile of the blue line indicated in **g**.

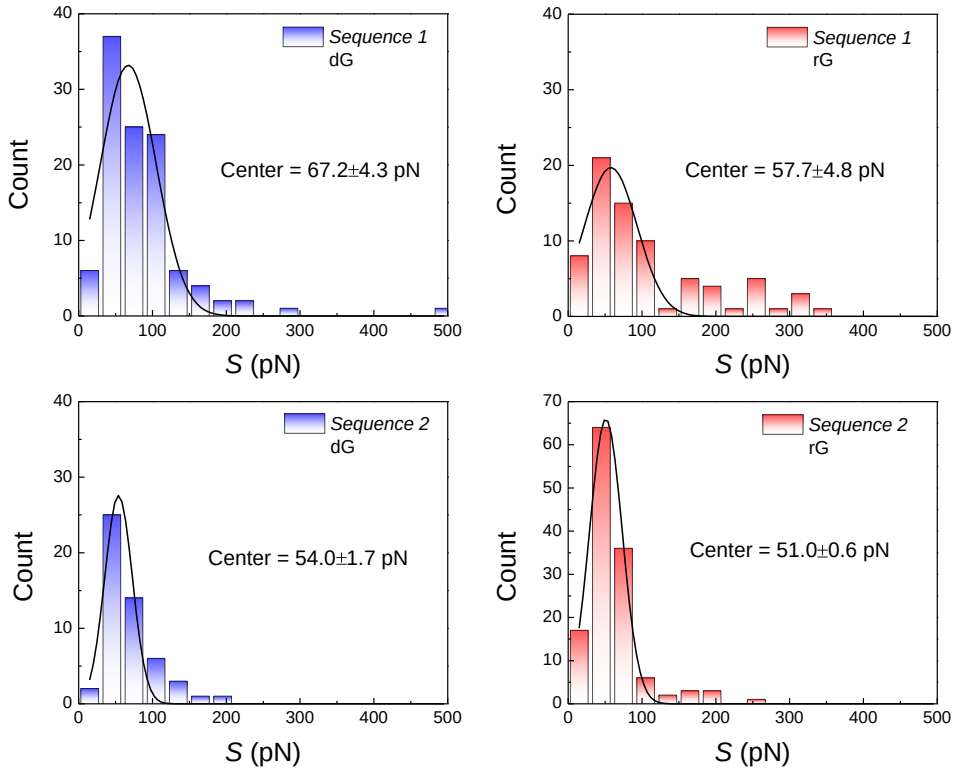


Figure B.3 Histograms of stretch moduli of ss substrates with *Sequence 1* and *Sequence 2*. Peak position is presented as the fitted value ± standard error of the fit.

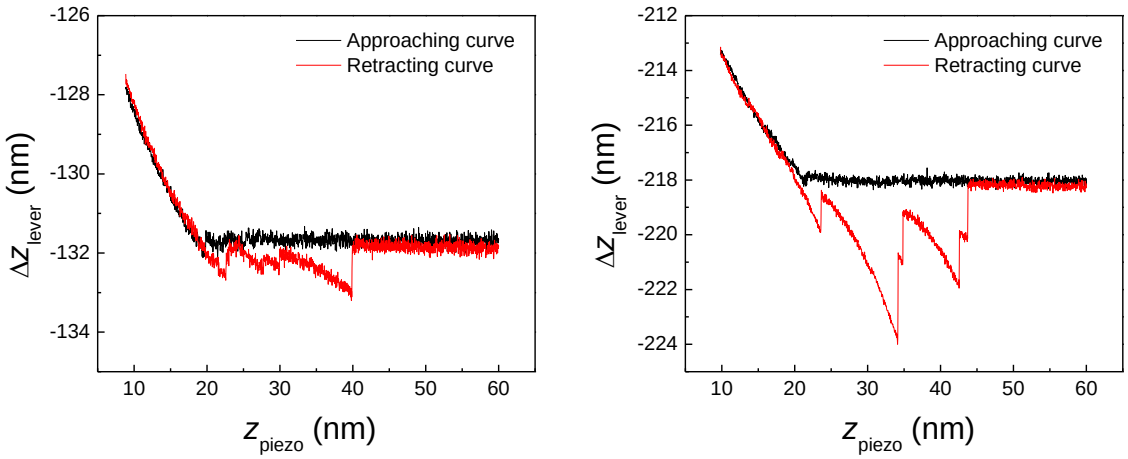


Figure B.4 Typical force-distance curves when the AFM tip picks up multiple DNAs.

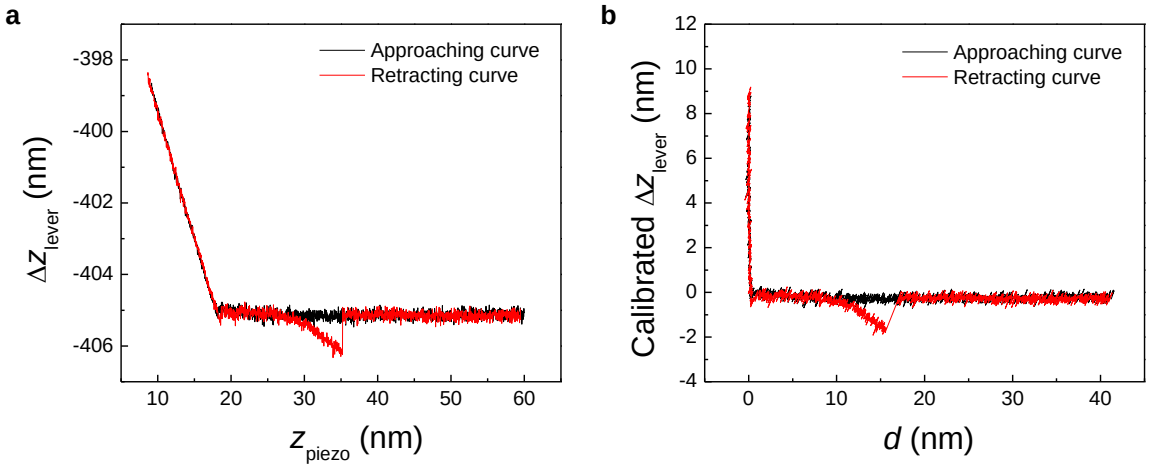


Figure B.5 Force–distance curves. (**a**) Before and (**b**) after the calibration procedure.

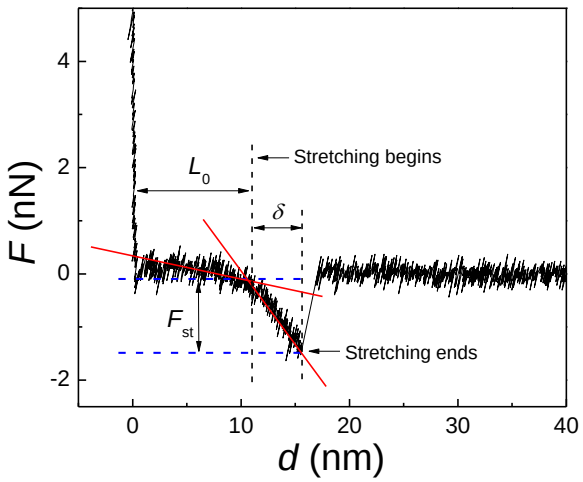


Figure B.6 Procedure to determine L_0, δ, and F_{st}.

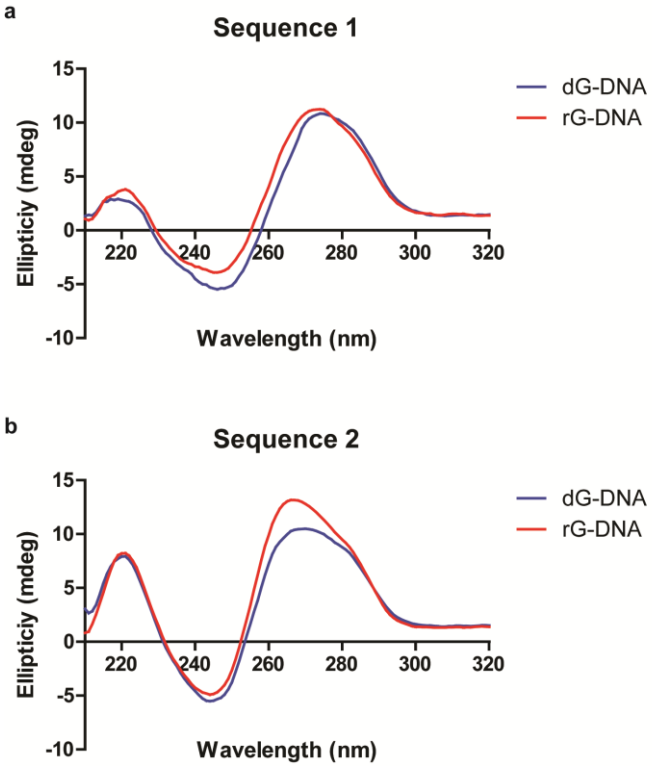


Figure B.7 CD spectra of dG-DNA and rG-DNA used in AFM experiments. (**a**) CD spectra of duplexes with *Sequence 1*; (**b**) CD spectra of duplexes with *Sequence 2*.

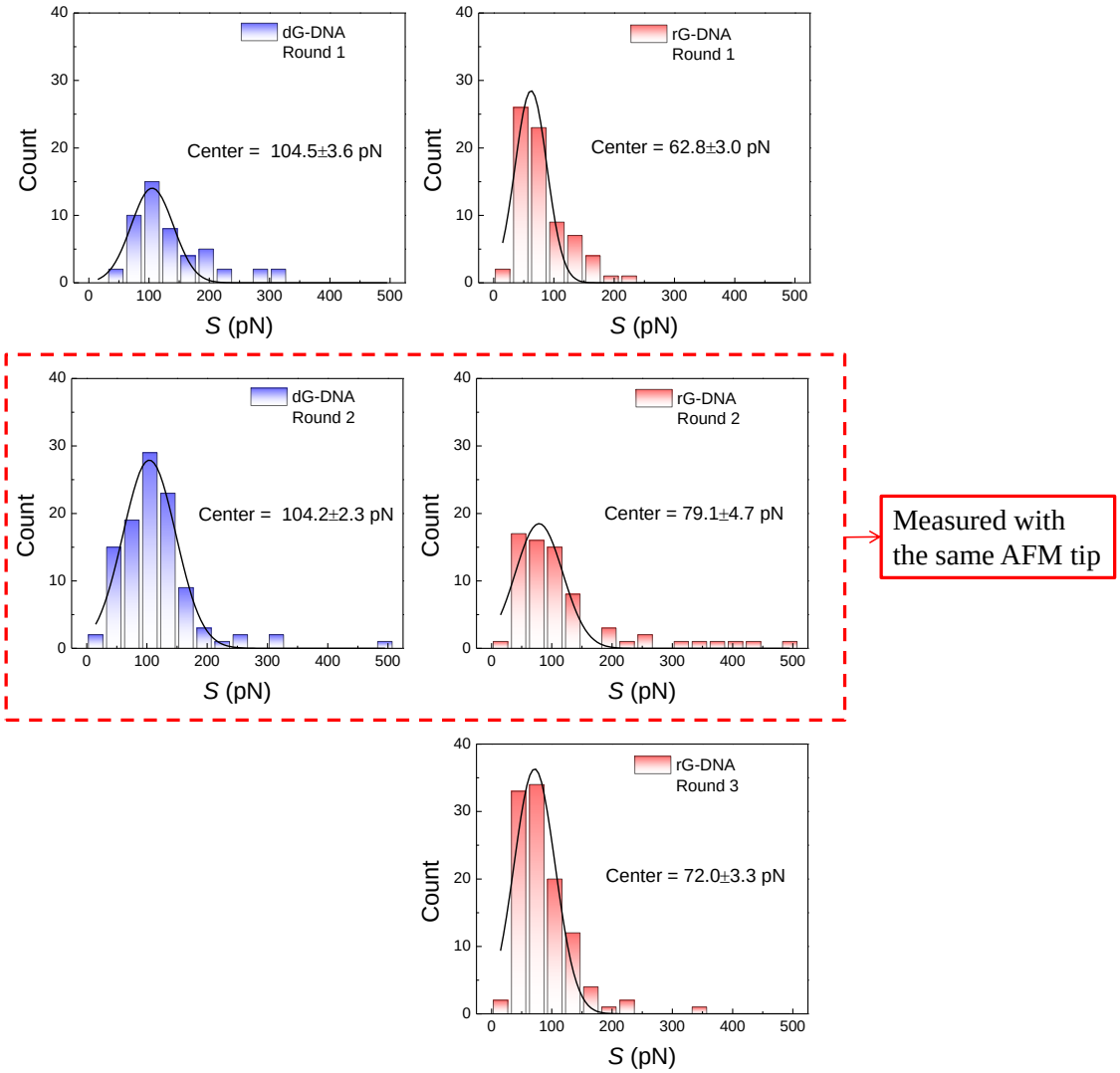


Figure B.8 Histograms of stretch moduli of ds substrates with *Sequence 1*. Peak position is presented as the fitted value ± standard error of the fit.

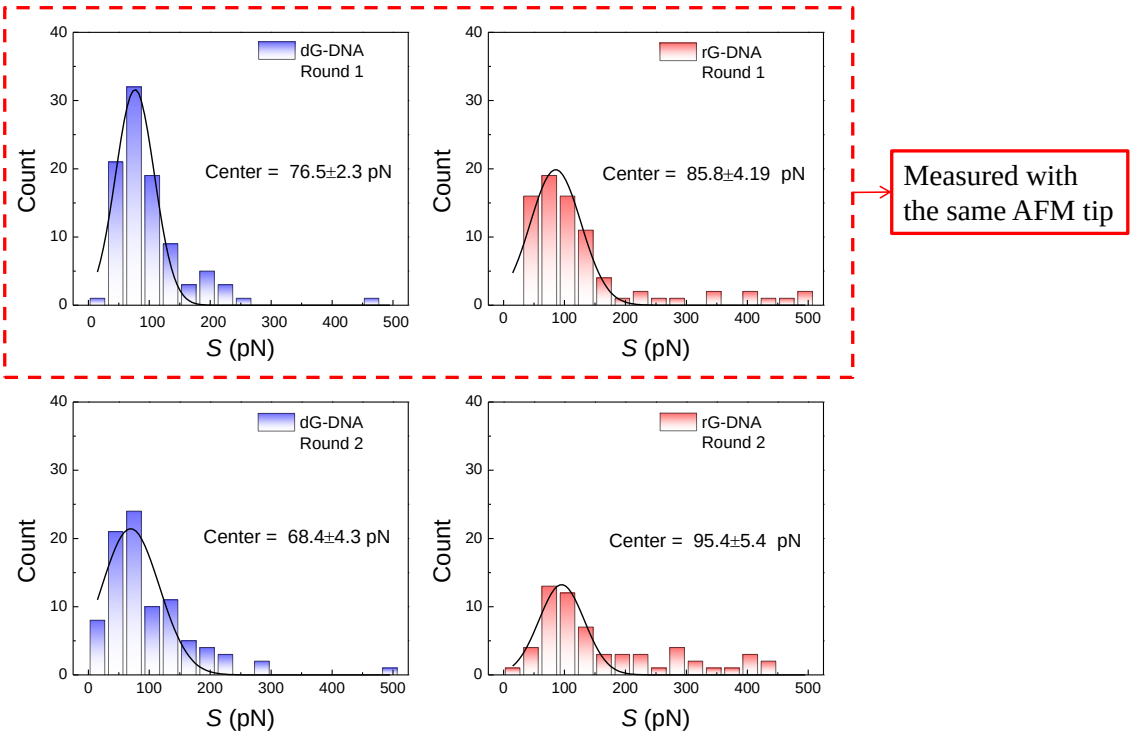


Figure B.9 Histograms of stretch moduli of ds substrates with *Sequence 2*. Peak position is presented as the fitted value ± standard error of the fit.

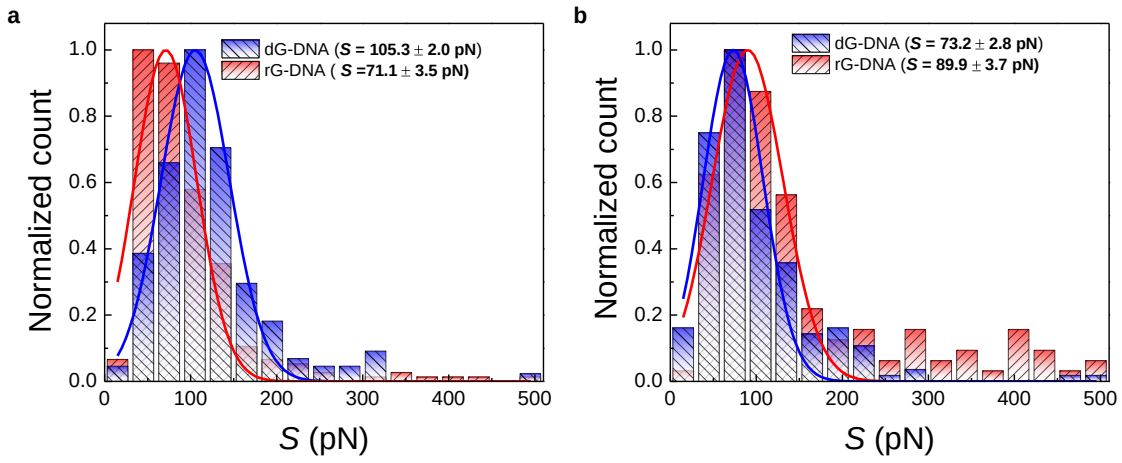


Figure B.10 Gaussian fitting for all combined data. (**a**) *Sequence 1*. (**b**) *Sequence 2*.

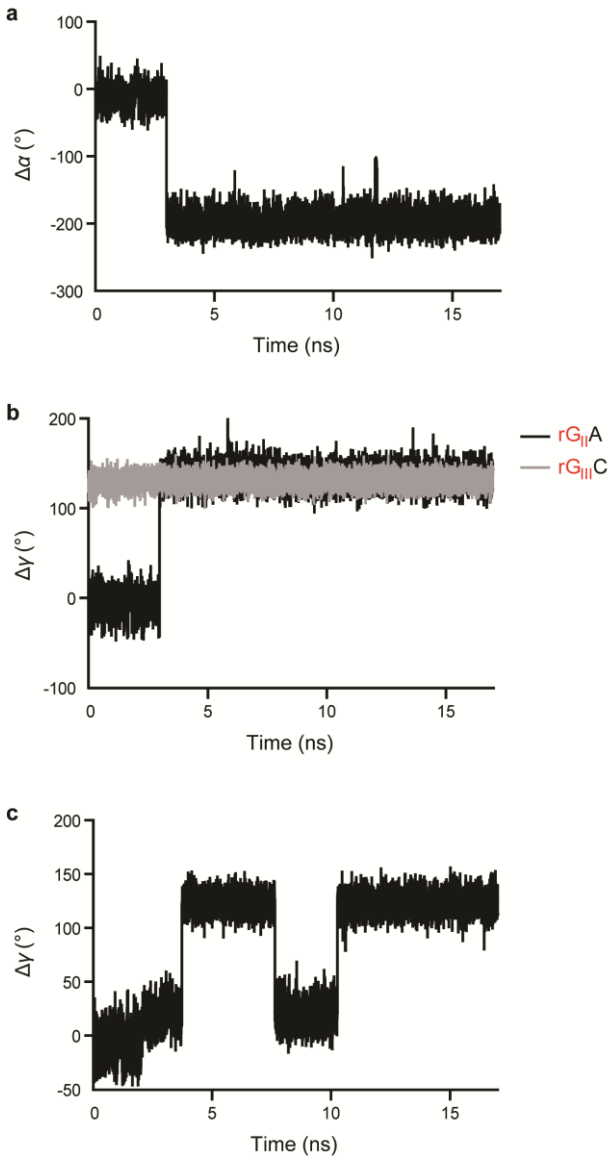


Figure B.11 Instantaneous deviations of α and γ torsional angles of rNMPs in DNA. (**a**) Instantaneous deviation of α torsional angle of the dAMP following rGMP in the 5′ to 3′ direction in CrGATGGArGCT for rGMP II. (**b**) Instantaneous deviations of γ torsional angles of the dAMP following rGMP and dCMP following rGMP in the 5′ to 3′ direction in CrGATGGArGCT for rGMPs II and III. (**c**) Instantaneous deviation of γ torsional angle of the rGMP I in GrGTTCArGGTT.

APPENDIX C

SUPPLEMENTARY MATERIALS FOR CHAPTER 4

Table C.1 *S. cerevisiae* strains used in this study.

a

Strain	Relevant genotype	Source
KK-100	*MATα ade5-1 lys2-14A trp1-289 his7-2 leu2-3,112 ura3-52 rnh201*Δ::*hygMX4*	this study
KK-30	*ho*Δ *hml*Δ::*ADE1 MATa-inc hmr*Δ::*ADE1 ade1 leu2-3,112 lys5 trp1*::*hisG ura3-52 leu2*::*HOcs mata*Δ::*hisG rnh201*Δ::*hygMX4*	this study
KK-174	KK-100 *rnh1*Δ::*kanMX4*	this study
KK-125	KK-30 *rnh1*Δ::*kanMX4*	this study
KK-164	KK-125 *ung1*Δ::*natMX4*	this study
KK-170	KK-30 *pol2-M644G*	this study
KK-107	KK-100 *pol2-4*	this study
KK-120	KK-100 *pol3-5DV*	this study

b

Strain	Relevant genotype	Source
FRO-767,768	*ho*Δ *hml*Δ::*ADE1 MATa-inc hmr*Δ::*ADE1 ade1 leu2-3,112 lys5 trp1*::*hisG ura3-52 ade3*::*GAL*::*HO leu2*::*HOcs mata*Δ::*hisG*	[46]
FRO-984,985	FRO-767,768 *rnh201*Δ::*kanMX4*	this study
KK-158,159	FRO-767,768 *ung1*Δ::*hygMX4*	this study

Yeast strains used in (**a**) ribose-seq library construction and (**b**) DSB repair assay with rNMP-containing oligos.

Table C.2 Oligos used in this study.

Name	Length (nt)	Sequence (5'-3') with end modifications	Purification	Experiment

Lig.47.D	47	CCCGAGTGTGATCATCTGGTCGCTGGGGAATGAGTCAGGCCACGGCG	PAGE	AtRNL ligation assay
Lig.47.R	47	CCCGAGTGTGATCATCTGGTCGCTGGGGAATrGAGTCAGGCCACGGCG	PAGE	AtRNL ligation assay
Lig.30.rA	30	NNNNNNNNNNNNNNNNNNNNrANNNNNNNN	PAGE	AtRNL 3' base bias assay
Lig.30.rG	30	NNNNNNNNNNNNNNNNNNNNrGNNNNNNNN	PAGE	AtRNL 3' base bias assay
Lig.30.rU	30	NNNNNNNNNNNNNNNNNNNNrUNNNNNNNN	PAGE	AtRNL 3' base bias assay
Lig.30.rC	30	NNNNNNNNNNNNNNNNNNNNrCNNNNNNNN	PAGE	AtRNL 3' base bias assay
Adaptor.L	87	P-NNNNNNNNAGATCGGAAGAGCGTCGTGTAGGGAAAG AGGGAGTTCAGACGTGTGCTCTTCCGATCTAGC CAGCGCAGACCGTGAGGT	PAGE	Ribose-seq library construction
Adaptor.S	20	P-CCTCACGGTCTGCGCTGGCT-Am	Desalted	Ribose-seq library construction
PCR.1.Index1	63	CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC	Desalted	Ribose-seq library construction
PCR.1.Index2	63	CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC	Desalted	Ribose-seq library construction
PCR.1.Index3	63	CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC	Desalted	Ribose-seq library construction
PCR.1.Index4	63	CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC	Desalted	Ribose-seq library construction
PCR.2	58	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT	Desalted	Ribose-seq library construction
ByTemp.rC	46	NNNNNNNrCNNNNNNNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAG	PAGE	Polymerase bypass assay
ByTemp.rU	46	NNNNNNNrUNNNNNNNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAG	PAGE	Polymerase bypass assay
ByPrim	30	CTCTTTCCCTACACGACGCTCTTCCGATCT	PAGE	Polymerase bypass assay
LEU2.D	60	TTAGGTGCTGTGGGTGGTCCTAAATGGGGATCCGGTAGTGTTAGGCCTGAACAAGGTTTA	Desalted	*leu2* DSB repair assay
LEU2.rG	60	TTAGGTGCTGTGGGTGGTCCTAAATGGGGATCCGGTAGTrGTTAGGCCTGAACAAGGTTTA	Desalted	*leu2* DSB repair assay
LEU2.dU	60	TTAGGTGCTGTGGGTGGTCCTAAATGGGGATCCGGTAGTGUTAGGCCTGAACAAGGTTTA	Desalted	*leu2* DSB repair assay
LEU2.rU	60	TTAGGTGCTGTGGGTGGTCCTAAATGGGGATCCGGTAGTGrUTAGGCCTGAACAAGGTTTA	Desalted	*leu2* DSB repair assay
LEU2.3	20	ATGTCTGCCCCTAAGAAGAT	Desalted	*leu2* DSB repair assay
LEU2.6	20	TGCCAAAGAATAAGGTCAAC	Desalted	*leu2* DSB repair assay

Name, length, and sequence of oligos used in this study are described. The purification type and the specific experiments in which the oligos were used are indicated. rNMPs are in red, preceded by 'r'. End modifications of phosphate and amino groups are indicated by 'P' and 'Am', respectively. All PAGE-purified oligos were synthesized by Thermo Scientific Dharmacon with exceptions for Lig.47.D and Adaptor.L, which were

synthesized by Life Technologies and IDT, respectively. All desalted oligos were synthesized by Eurofins Genomics.

Table C.3 Results of 3′ base bias for AtRNL ligation.

a

Base	Circular	Dimer	Circular dimer
A	48% (44–49)	4.3% (1.8–7.1)	1.5% (0.71–2.1)
G	47% (44–48)	4.0% (2.6–5.8)	1.5% (0.88–2.3)
U	47% (45–49)	4.4% (2.1–5.1)	1.5% (0.73–2.0)
C	47% (44–49)	4.5% (1.9–5.0)	1.4% (0.59–1.8)

b

P value	G	U	C
A	0.4857	0.8857	1.0000
G	–	0.8857	0.6857
U	–	–	1.0000

c

Base	Circular	Dimer	Circular dimer
A	27% (23–31)	1.7% (1.5–2.3)	0.53% (0.45–0.74)
G	27% (24–28)	1.9% (1.5–2.5)	0.54% (0.39–1.1)
U	29% (24–30)	2.3% (1.3–2.7)	0.64% (0.48–1.2)
C	29% (25–32)	1.8% (1.7–2.0)	0.53% (0.47–0.80)

d

P value	G	U	C
A	1.0000	1.0000	0.6857
G	–	0.3429	0.2000
U	–	–	0.8857

(**a**) Levels of AtRNL ligation in reaction conditions described in **Figure C.1** are expressed as median percentage and range (in parentheses) from four independent reactions. (**b**) Mann-Whitney *U*-test was performed for statistical analysis, and *P* values are displayed, all greater than 0.05. (**c**) AtRNL ligation was performed with reduced 200 nM AtRNL, instead of 1 μM, to compare the levels of ligation when the reactions were incomplete. Median percentages and ranges (in parentheses) from four independent reactions are displayed. (**d**) Mann-Whitney *U*-test was performed for statistical analysis, and *P* values are displayed, all greater than 0.05. No 3′ base bias was observed for AtRNL ligation.

Table C.4 Ribose-seq coverage for each library in this study.

Ribose-seq library	Coverage (aligned reads/kb)	
	Nuclear	Mitochondrial
rnh201 (KK-100)	0.449	19.5
rnh201 (KK-100, EconoTaq)	0.883	47.8
rnh201 (KK-30)	0.149	8.42
rnh1 rnh201 (KK-174)	0.149	9.92
rnh1 rnh201 (KK-125)	0.239	13.2
rnh1 rnh201 ung1 (KK-164)	0.269	42.2
pol2-M644G rnh201 (KK-170)	0.254	7.89
pol2-4 rnh201 (KK-107)	0.528	34.2
pol3-5DV rnh201 (KK-120)	0.510	33.9

Coverage is expressed as aligned reads per kb and does not reflect the relative abundance of rNMPs among different strains.

Table C.5 Absolute nucleotide frequencies of rNMPs and 3′ flanking nucleotide.

	Base	Position 0		Position +1	
		Nuclear	Mitochondrial	Nuclear	Mitochondrial
rnh201 (KK-100)	A	15.4%	25.6%	45.0%	45.8%
	C	44.0%	36.8%	22.4%	15.3%
	G	28.1%	19.0%	16.5%	5.8%
	U/T	12.5%	18.7%	16.1%	33.1%
rnh201 (KK-100, EconoTaq)	A	23.2%	38.2%	43.3%	43.5%
	C	35.4%	25.6%	19.3%	10.0%
	G	22.7%	14.5%	13.5%	6.2%
	U/T	18.7%	21.7%	23.9%	40.3%
rnh201 (KK-30)	A	20.4%	35.7%	47.5%	47.8%
	C	39.2%	28.3%	19.6%	11.1%
	G	27.5%	24.1%	14.4%	7.2%
	U/T	12.8%	11.9%	18.5%	33.8%
rnh1 rnh201 (KK-174)	A	17.1%	33.6%	44.9%	46.5%
	C	40.2%	27.0%	22.0%	11.8%
	G	27.7%	23.7%	15.0%	7.4%
	U/T	15.1%	15.7%	18.2%	34.2%
rnh1 rnh201 (KK-125)	A	20.1%	35.4%	45.2%	44.3%
	C	36.8%	28.6%	19.4%	12.1%
	G	29.7%	20.9%	15.0%	5.9%
	U/T	13.4%	15.1%	20.4%	37.7%
rnh1 rnh201 ung1 (KK-164)	A	24.3%	35.8%	44.3%	47.1%
	C	35.3%	30.2%	19.4%	13.2%
	G	26.5%	22.7%	15.3%	6.5%
	U/T	14.0%	11.3%	21.0%	33.2%
pol2-M644G rnh201	A	19.5%	38.9%	52.2%	47.2%
	C	40.3%	28.6%	18.2%	11.5%

(KK-170)	G	26.5%	21.5%	13.0%	6.6%
	U/T	13.7%	11.0%	16.6%	34.7%
pol2-4 rnh201 (KK-107)	A	14.9%	21.9%	42.5%	46.2%
	C	40.2%	43.1%	22.0%	16.3%
	G	23.6%	16.3%	16.4%	6.4%
	U/T	21.2%	18.6%	19.2%	31.1%
pol3-5DV rnh201 (KK-120)	A	20.4%	30.0%	44.3%	45.3%
	C	37.1%	33.1%	19.4%	14.6%
	G	25.0%	16.7%	15.3%	6.1%
	U/T	17.5%	20.3%	21.0%	34.0%

Absolute nucleotide frequencies of nuclear and mitochondrial rNMPs and the nucleotide immediately downstream (position +1) from each ribose-seq library.

Table C.6 Results of rNMP bypass by Phusion DNA Polymerase.

a

Base	Bypass probability
C	93% (93–93)
U	93% (92–94)

b

P value	U
C	0.6857

(**a**) Bypass probabilities in reaction conditions described in **Figure C.4** are expressed as median percentage and range (in parentheses) from four independent reactions. (**b**) Mann-Whitney *U*-test was performed for statistical analysis, and *P* value is displayed.

Table C.7 Results of DSB repair assay with rNMP-containing oligos.

a

Oligo	WT	*rnh201*	*ung1*
LEU2.D	65% (55–75)	63% (55–65)	65% (55–70)
LEU2.rG	30% (20–40)	90% (75–100)	N/A
LEU2.dU	5.0% (0–10)	N/A	55% (45–60)
LEU2.rU	33% (25–45)	55% (45–65)	33% (30–40)

b

Oligo	*rnh201*	*ung1*
LEU2.D	0.5357	1.0000
LEU2.rG	0.0286	N/A
LEU2.dU	N/A	0.0294
LEU2.rU	0.0421	1.0000

(**a**) Data shown in **Figure C.7** are presented here as median percentages of StuI-cut Leu^+ transformants from four independent transformations and ranges in parentheses. For each

transformation, 20 Leu^{+} transformants were selected for analysis. (**b**) Mann-Whitney *U*-test was implemented for statistical analysis against the WT, and *P* values are displayed. N/A, not applicable because data are not available for comparison.

Table C.8 List of hotspots of rNMP incorporation within *S. cerevisiae* mitochondrial DNA, rDNA repeat, and *Ty1*.

Position	Strand	Gene	Base	Number of rNMP reads								
				rnh201 **(KK-100)**	***rnh201*** **(KK-100, EconoTaq)**	***rnh201*** **(KK-30)**	***rnh1 rnh201*** **(KK-174)**	***rnh1 rnh201*** **(KK-125)**	*rnh1 rnh201 ung1* (KK-164)	*pol2-M644G rnh201* (KK-170)	*pol2-4 rnh201* (KK-107)	*pol3-5DV rnh201* (KK-120)
Chr M 39,224	W	*COB*	A	**30**	**136**	**17**	**24**	**34**	2	20	10	22
Chr XII[a] 453,839	C	*RDN25*	G	**97**	**45**	**18**	**30**	**22**	1	25	12	10
Chr IV[b] 650,383	C[b]	*Ty1*	A	**15**	**42**	**10**	**11**	**19**	46	20	19	110
Chr M 14,688	W	*COX1*	A	**7**	**49**	**5**	**10**	**14**	33	12	8	38
Chr M 14,739	W	*COX1*	A	**8**	**46**	**7**	**14**	**12**	0	5	0	13
Chr M 19,157	W	*COX1*	A	**15**	**73**	**4**	**5**	**22**	28	5	7	27

Hotspots of rNMP incorporation were determined by finding positions of rNMPs within the locus of interest with ribose-seq signal greater than the mean plus three standard deviations for each library from *rnh201*Δ (KK-100), *rnh201*Δ (KK-100, EconoTaq), *rnh201*Δ (KK-30), *rnh1*Δ *rnh201*Δ (KK-174), and *rnh1*Δ *rnh201*Δ (KK-125) cells (in bold). Ribose-seq signal counts found in all other libraries are also shown.

[a]There are two rDNA repeats on Chr XII in the reference genome (sacCer2). Only the first repeat unit is shown as an example.

[b]Because of the presence of multiple copies of *Ty1* in the genome, *YDRCTy1-1* on Chr IV is shown as an example.

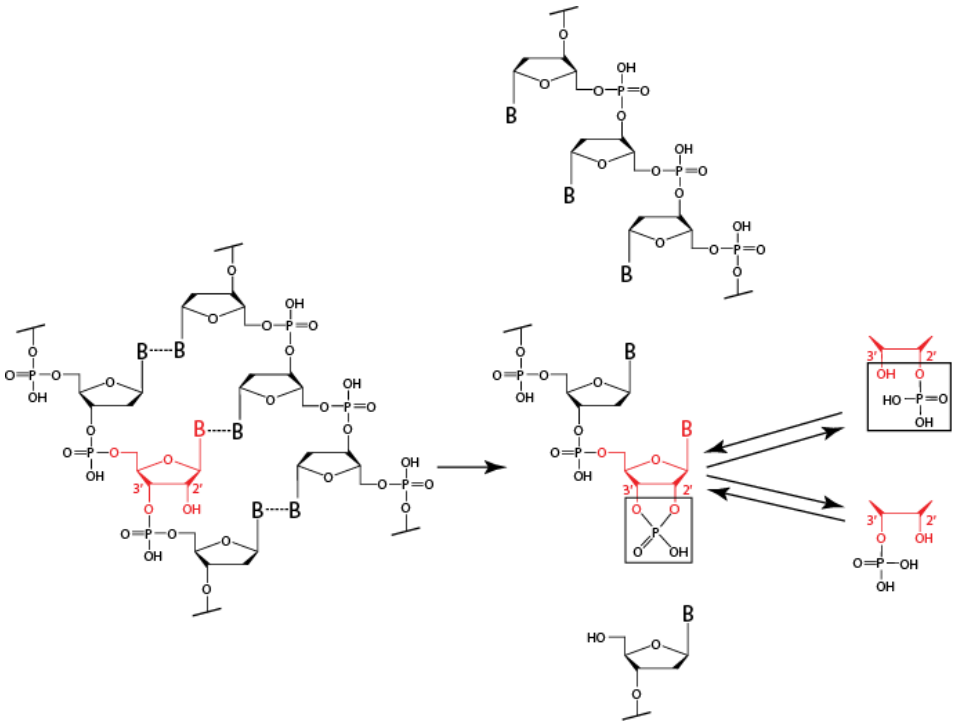


Figure C.1 Mechanism of alkaline cleavage of ribonucleotides in DNA. The ribonucleoside embedded in double-stranded DNA is in red. During alkaline treatment, DNA strands are denatured, and cleavage occurs at the rNMP site, generating a 2′,3′-cyclic phosphate end and an opposite 5′-hydroxyl end. The 2′,3′-cyclic phosphate is in equilibrium with 2′-phosphate and 3′-phosphate forms. Boxes in black indicate the 2′,3′-cyclic phosphate and 2′-phosphate DNA termini, which are substrates of AtRNL.

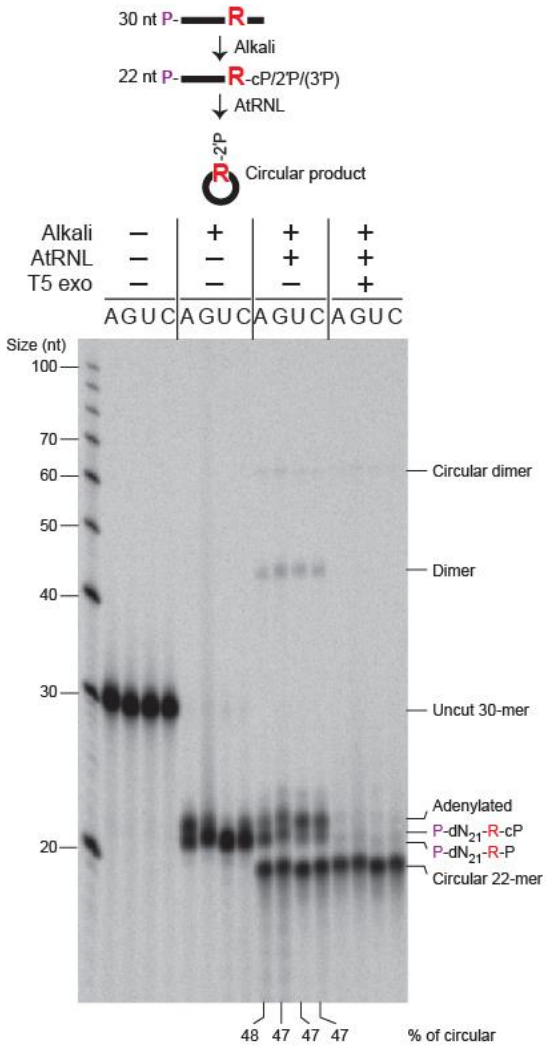


Figure C.2 3′ base bias for AtRNL ligation. Hot 5′-radiolabeled 30-nt DNA oligo with a single rNMP (either A, G, U, or C) in the 22nd position was mixed with cold equimolar 30-nt DNA oligos with rNMPs of 3 other bases in the 22nd positions. 5′-radiolabel is indicated by 'P' in purple. The mixture was treated with 0.3M NaOH for 2 hr at 55 °C and neutralized. 100 nM of alkali-cleaved products (25 nM of each base) were then incubated with 1 μM AtRNL in appropriate buffer (see **Materials and Methods**) for 1 hr at 30 ℃. The resulting products were treated with T5 exonuclease for 2 hr at 37 ℃. Aliquots were withdrawn after appropriate steps and quenched. The products were analyzed by urea-PAGE. The circular 22-mer migrates faster than the unligated, linear 22-mer. Only circular products were resistant to T5 exonuclease while all linear

substrates/products were degraded. Median percentages of circular 22-mer formation from four independent reactions are displayed. See Supplementary Table 1 for more statistics. First left lane, ss DNA ladder. No 3′ base bias was observed for AtRNL ligation (see **Table C.3**). Self-ligation was preferred to dimerization with a shorter 22-nt substrate; however, with the shorter substrate, lower levels of linear dimers, which are not resistant to T5 exonuclease, and circular dimers were observed. Increasing the length of the ss DNA substrate from 22 nt to 32 nt eliminated dimerization (**Figure 4.1a**).

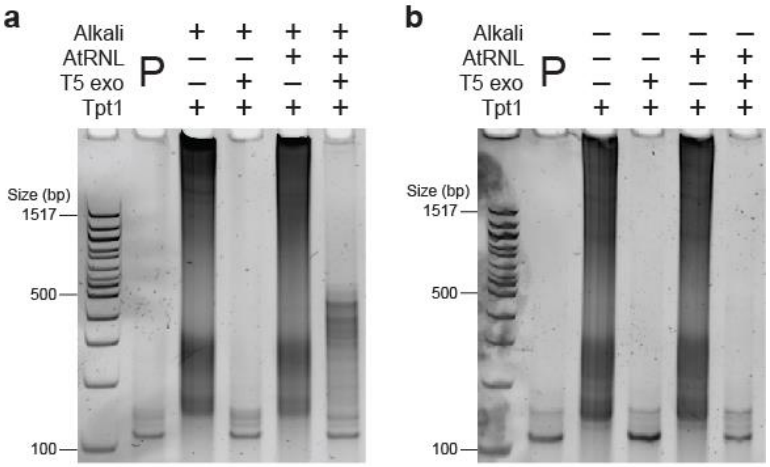


Figure C.3 Ribose-seq library from genomic DNA of *S. cerevisiae rnh201*Δ (KK-100) cells. Appropriate PCR products were analyzed by PAGE. 'P' indicates primers-only. No amplification product was observed when either (**a**) AtRNL ligation step or (**b**) alkali treatment was omitted. Tpt1 denotes the step of 2'-phosphate removal at the ligation junction in **Figure 4.1a**. First left lane, ds DNA ladder.

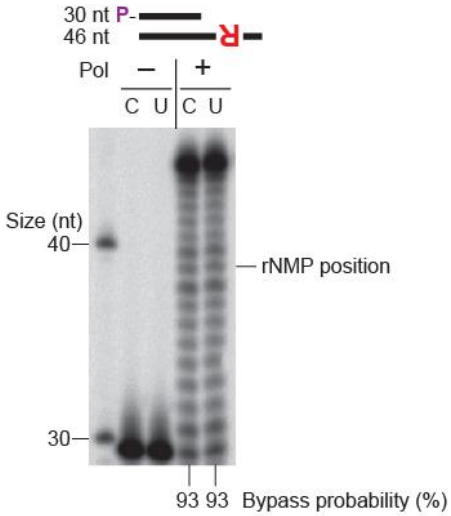


Figure C.4 Bypass of a single rNMP by Phusion DNA Polymerase. 5′-radiolabeled 30-nt primer, ByPrim (**Table C.2**), was annealed to the 46-nt template oligo containing either rCMP (ByTemp.rC) or rUMP (ByTemp.rU) in the 8th position. 100 nM of annealed substrate was incubated with 0.2 units of Phusion High-Fidelity DNA Polymerase (NEB) and 2 mM dNTPs in appropriate buffer (see **Materials and Methods**) for 30 sec at 72 ℃. The reactions were quenched and analyzed by urea-PAGE. Median bypass probabilities from four independent reactions are shown. See **Table C.6** for more statistics. First left lane, ss DNA ladder. The primer extension assay showed no significant difference between bypass efficiency over rUMP and rCMP by Phusion DNA Polymerase (**Table C.6**).

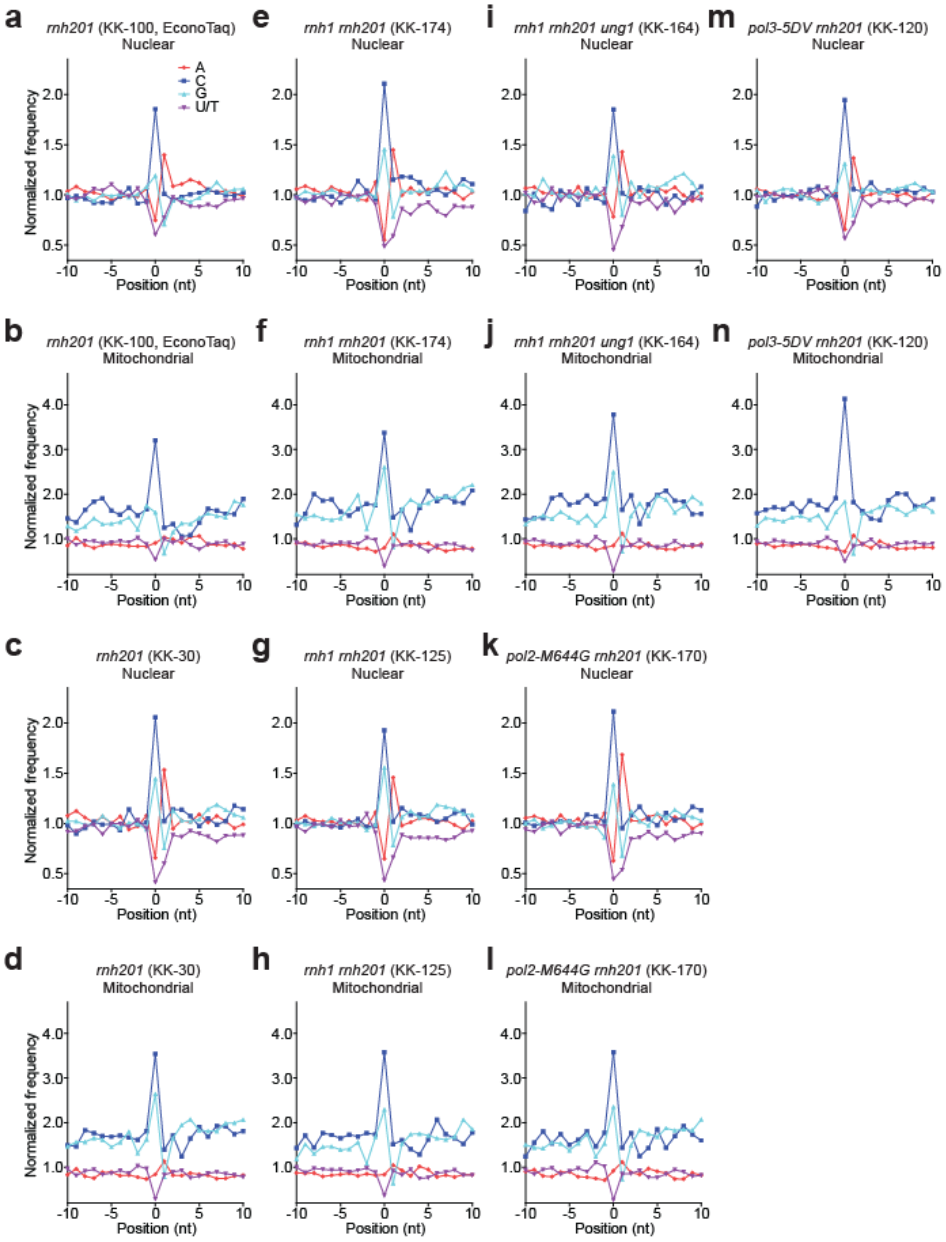


Figure C.5 Normalized frequency of nucleotides surrounding the rNMP sites. Normalized frequency of nucleotides relative to (**a**) nuclear and (**b**) mitochondrial mapped positions of sequences from ribose-seq library, PCR-amplified with EconoTaq DNA Polymerase (Lucigen), of genomic DNA from *S. cerevisiae rnh201*Δ (KK-100) cells. Position 0 corresponds to the rNMP. Negative and positive numbers (from -10 to -1 and 1 to 10) correspond to upstream and downstream positions from the rNMP, respectively. Frequencies were normalized to either nuclear or mitochondrial genomic

mononucleotide frequencies. Normalized frequency of nucleotides relative to (**c**) nuclear and (**d**) mitochondrial mapped positions of sequences from ribose-seq library of genomic DNA from *S. cerevisiae rnh201*Δ (KK-30) cells. Normalized frequency of nucleotides relative to (**e**) nuclear and (**f**) mitochondrial mapped positions of sequences from ribose-seq library of genomic DNA from *S. cerevisiae rnh1*Δ *rnh201*Δ (KK-174) cells. Normalized frequency of nucleotides relative to (**g**) nuclear and (**h**) mitochondrial mapped positions of sequences from ribose-seq library of genomic DNA from *S. cerevisiae rnh1*Δ *rnh201*Δ (KK-125) cells. Normalized frequency of nucleotides relative to (**i**) nuclear and (**j**) mitochondrial mapped positions of sequences from ribose-seq library of genomic DNA from *S. cerevisiae rnh1*Δ *rnh201*Δ *ung1*Δ (KK-164) cells. Normalized frequency of nucleotides relative to (**k**) nuclear and (**l**) mitochondrial mapped positions of sequences from ribose-seq library of genomic DNA from *S. cerevisiae pol2-M644G rnh201*Δ (KK-170) cells. Normalized frequency of nucleotides relative to (**m**) nuclear and (**n**) mitochondrial mapped positions of sequences from ribose-seq library of genomic DNA from *S. cerevisiae pol3-5DV rnh201*Δ (KK-120) cells.

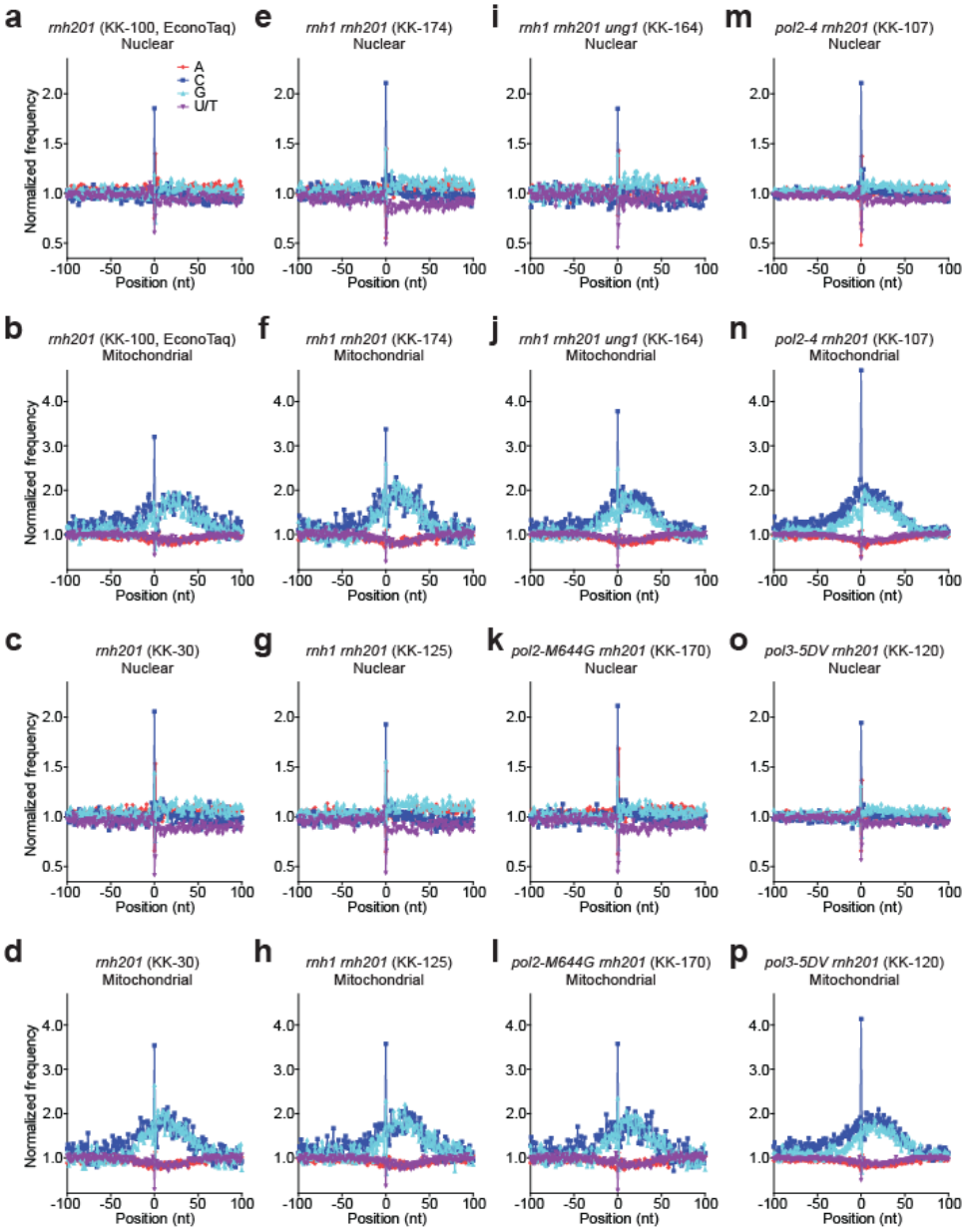


Figure C.6 Zoom-out of normalized frequency of nucleotides surrounding the rNMP sites. Normalized frequency of nucleotides relative to (**a**) nuclear and (**b**) mitochondrial mapped positions of sequences from ribose-seq library, PCR-amplified with EconoTaq DNA Polymerase (Lucigen), of genomic DNA from *S. cerevisiae rnh201*Δ (KK-100) cells. Position 0 corresponds to the rNMP. Negative and positive numbers (from -100 to -1 and 1 to 100) correspond to upstream and downstream positions from the rNMP, respectively. Frequencies were normalized to either nuclear or mitochondrial genomic

mononucleotide frequencies. Normalized frequency of nucleotides relative to (**c**) nuclear and (**d**) mitochondrial mapped positions of sequences from ribose-seq library of genomic DNA from *S. cerevisiae rnh201*Δ (KK-30) cells. Normalized frequency of nucleotides relative to (**e**) nuclear and (**f**) mitochondrial mapped positions of sequences from ribose-seq library of genomic DNA from *S. cerevisiae rnh1*Δ *rnh201*Δ (KK-174) cells. Normalized frequency of nucleotides relative to (**g**) nuclear and (**h**) mitochondrial mapped positions of sequences from ribose-seq library of genomic DNA from *S. cerevisiae rnh1*Δ *rnh201*Δ (KK-125) cells. Normalized frequency of nucleotides relative to (**i**) nuclear and (**j**) mitochondrial mapped positions of sequences from ribose-seq library of genomic DNA from *S. cerevisiae rnh1*Δ *rnh201*Δ *ung1*Δ (KK-164) cells. Normalized frequency of nucleotides relative to (**k**) nuclear and (**l**) mitochondrial mapped positions of sequences from ribose-seq library of genomic DNA from *S. cerevisiae pol2-M644G rnh201*Δ (KK-170) cells. Normalized frequency of nucleotides relative to (**m**) nuclear and (**n**) mitochondrial mapped positions of sequences from ribose-seq library of genomic DNA from *S. cerevisiae pol2-4 rnh201*Δ (KK-107) cells. Normalized frequency of nucleotides relative to (**o**) nuclear and (**p**) mitochondrial mapped positions of sequences from ribose-seq library of genomic DNA from *S. cerevisiae pol3-5DV rnh201*Δ (KK-120) cells.

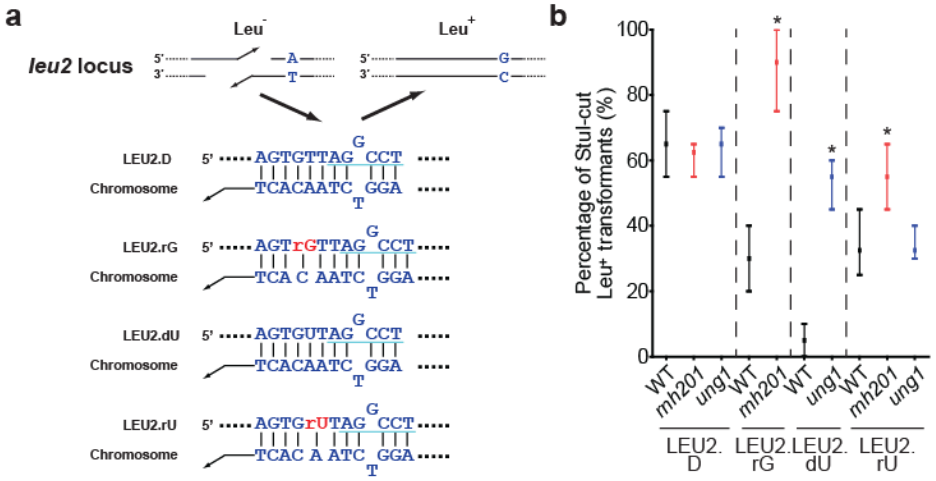


Figure C.7 Targeting of rGMP and rUMP by RNase H2 and uracil DNA N-glycosylase during DSB repair in *S. cerevisiae* cells. (**a**) Diagram and sequence of the chromosomal *leu2* region targeted by DNA-control LEU2.D, rGMP-containing LEU2.rG, dUMP-containing LEU2.dU, and rUMP-containing LEU2.rU oligos (**Table C.2**). StuI recognition sequence is underlined in turquoise. Position of either rGMP, dUMP, or rUMP was selected so that it is about 4–5 nt upstream of the G-T mispair. Both RNase H2-initiated excision repair (RER) and base excision repair (BER) remove a short ss DNA region downstream of the damage during the repair [35, 92]. (**b**) The oligos were transformed to either RNase H2– and uracil DNA N-glycosylase–proficient wild-type (WT; FRO-767,768), RNase H2–deficient (*rnh201*; FRO-984,985), or DNA N-glycosylase–deficient (*ung1*; KK-158,159) *S. cerevisiae* cells (see **Table C.1**). Median percentages of StuI-cut Leu$^+$ transformants from four independent transformations are shown with ranges as bars. For each transformation, 20 Leu$^+$ transformants were selected for analysis. Mann-Whitney *U*-test was implemented for statistical analysis against the WT. *P* values of less than 0.05 are marked as asterisk. See **Table C.7** for more statistics.

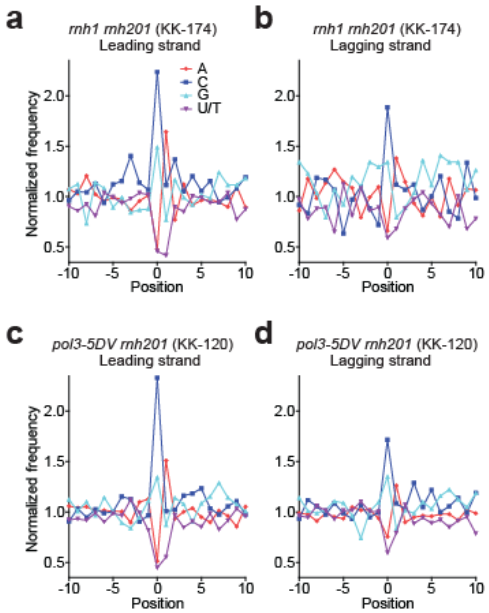


Figure C.8 Normalized frequency of nucleotides surrounding the rNMP sites on leading and lagging strands. Normalized frequency of nucleotides relative to mapped positions of sequences in (**a**) leading and (**b**) lagging strands from ribose-seq library of genomic DNA from *S. cerevisiae rnh1*Δ *rnh201*Δ (KK-174) cells. Position 0 corresponds to the rNMP. Negative and positive numbers (from -10 to -1 and 1 to 10) correspond to upstream and downstream positions from the rNMP, respectively. ARSs with T_{rep} of no longer than 25 min were selected with flanking size of 10 kb. Frequencies were normalized to genomic mononucleotide frequencies of either leading or lagging strand of the selected ARSs and flanking size. Normalized frequency of nucleotides relative to mapped positions of sequences in (**c**) leading and (**d**) lagging strands from ribose-seq library of genomic DNA from *S. cerevisiae pol3-5DV rnh201*Δ (KK-120) cells.

REFERENCES

[1] J.H. Hoeijmakers, Genome maintenance mechanisms for preventing cancer, Nature, 411 (2001) 366-374.

[2] C.L. Peterson, J. Cote, Cellular machineries for chromosomal DNA repair, Genes & development, 18 (2004) 602-616.

[3] M.Y. Yang, M. Bowmaker, A. Reyes, L. Vergani, P. Angeli, E. Gringeri, H.T. Jacobs, I.J. Holt, Biased incorporation of ribonucleotides on the mitochondrial L-strand accounts for apparent strand-asymmetric DNA replication, Cell, 111 (2002) 495-505.

[4] S. Vengrova, J.Z. Dalgaard, The wild-type Schizosaccharomyces pombe mat1 imprint consists of two ribonucleotides, EMBO Rep, 7 (2006) 59-65.

[5] P.H. Patel, L.A. Loeb, Multiple amino acid substitutions allow DNA polymerases to synthesize RNA, The Journal of biological chemistry, 275 (2000) 40266-40272.

[6] M. Astatke, K. Ng, N.D. Grindley, C.M. Joyce, A single side chain prevents Escherichia coli DNA polymerase I (Klenow fragment) from incorporating ribonucleotides, Proceedings of the National Academy of Sciences of the United States of America, 95 (1998) 3402-3407.

[7] A. Bonnin, J.M. Lazaro, L. Blanco, M. Salas, A single tyrosine prevents insertion of ribonucleotides in the eukaryotic-type phi29 DNA polymerase, Journal of molecular biology, 290 (1999) 241-251.

[8] N.A. Cavanaugh, W.A. Beard, V.K. Batra, L. Perera, L.G. Pedersen, S.H. Wilson, Molecular insights into DNA polymerase deterrents for ribonucleotide insertion, The Journal of biological chemistry, 286 (2011) 31650-31660.

[9] J.A. Brown, Z. Suo, Unlocking the sugar "steric gate" of DNA polymerases, Biochemistry, 50 (2011) 1135-1142.

[10] W. Wang, E.Y. Wu, H.W. Hellinga, L.S. Beese, Structural factors that determine selectivity of a high fidelity DNA polymerase for deoxy-, dideoxy-, and ribonucleotides, The Journal of biological chemistry, 287 (2012) 28215-28226.

[11] N.Y. Yao, J.W. Schroeder, O. Yurieva, L.A. Simmons, M.E. O'Donnell, Cost of rNTP/dNTP pool imbalance at the replication fork, Proceedings of the National Academy of Sciences of the United States of America, 110 (2013) 12942-12947.

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

[13] C. Gong, P. Bongiorno, A. Martins, N.C. Stephanou, H. Zhu, S. Shuman, M.S. Glickman, Mechanism of nonhomologous end-joining in mycobacteria: a low-fidelity repair system driven by Ku, ligase D and ligase C, Nature structural & molecular biology, 12 (2005) 304-312.

[14] A.R. Clausen, S. Zhang, P.M. Burgers, M.Y. Lee, T.A. Kunkel, Ribonucleotide incorporation, proofreading and bypass by human DNA polymerase delta, DNA repair, 12 (2013) 121-127.

[15] A.Y. Goksenin, W. Zahurancik, K.G. LeCompte, D.J. Taggart, Z. Suo, Z.F. Pursell, Human DNA polymerase epsilon is able to efficiently extend from multiple consecutive ribonucleotides, The Journal of biological chemistry, 287 (2012) 42675-42684.

[16] R.A. Gosavi, A.F. Moon, T.A. Kunkel, L.C. Pedersen, K. Bebenek, The catalytic cycle for ribonucleotide incorporation by human DNA Pol lambda, Nucleic acids research, 40 (2012) 7518-7527.

[17] S.A. Nick McElhinny, D.A. Ramsden, Polymerase mu is a DNA-directed DNA/RNA polymerase, Molecular and cellular biology, 23 (2003) 2309-2315.

[18] N.A. Cavanaugh, W.A. Beard, S.H. Wilson, DNA polymerase beta ribonucleotide discrimination: insertion, misinsertion, extension, and coding, The Journal of biological chemistry, 285 (2010) 24457-24465.

[19] R. Kasiviswanathan, W.C. Copeland, Ribonucleotide discrimination and reverse transcription by the human mitochondrial DNA polymerase, The Journal of biological chemistry, 286 (2011) 31490-31500.

[20] E.M. Kennedy, S.M. Amie, R.A. Bambara, B. Kim, Frequent incorporation of ribonucleotides during HIV-1 reverse transcription and their attenuated repair in macrophages, The Journal of biological chemistry, 287 (2012) 14280-14288.

[21] S.A. Nick McElhinny, B.E. Watts, D. Kumar, D.L. Watt, E.B. Lundstrom, P.M. Burgers, E. Johansson, A. Chabes, T.A. Kunkel, Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases, Proceedings of the National Academy of Sciences of the United States of America, 107 (2010) 4949-4954.

[22] K. Collins, C.W. Greider, Utilization of ribonucleotides and RNA primers by Tetrahymena telomerase, The EMBO journal, 14 (1995) 5422-5432.

[23] J.B. Boule, F. Rougeon, C. Papanicolaou, Terminal deoxynucleotidyl transferase indiscriminately incorporates ribonucleotides and deoxyribonucleotides, The Journal of biological chemistry, 276 (2001) 31388-31393.

[24] T.W. Traut, Physiological concentrations of purines and pyrimidines, Molecular and cellular biochemistry, 140 (1994) 1-22.

[25] M. Fasullo, O. Tsaponina, M. Sun, A. Chabes, Elevated dNTP levels suppress hyper-recombination in Saccharomyces cerevisiae S-phase checkpoint mutants, Nucleic acids research, 38 (2010) 1195-1203.

[26] R.D. Kuchta, G. Stengel, Mechanism and evolution of DNA primases, Biochimica et biophysica acta, 1804 (2010) 1180-1189.

[27] P.M. Burgers, Polymerase dynamics at the eukaryotic DNA replication fork, The Journal of biological chemistry, 284 (2009) 4041-4045.

[28] J.A. Rumbaugh, R.S. Murante, S. Shi, R.A. Bambara, Creation and removal of embedded ribonucleotides in chromosomal DNA during mammalian Okazaki fragment processing, The Journal of biological chemistry, 272 (1997) 22591-22599.

[29] K. Randerath, R. Reddy, T.F. Danna, W.P. Watson, A.E. Crane, E. Randerath, Formation of ribonucleotides in DNA modified by oxidative damage in vitro and in vivo. Characterization by 32P-postlabeling, Mutation research, 275 (1992) 355-366.

[30] H. Wiseman, B. Halliwell, Damage to DNA by reactive oxygen and nitrogen species: role in inflammatory disease and progression to cancer, Biochem J, 313 (Pt 1) (1996) 17-29.

[31] M.A. Reijns, H. Kemp, J. Ding, S.M. de Proce, A.P. Jackson, M.S. Taylor, Lagging-strand replication shapes the mutational landscape of the genome, Nature, (2015).

[32] J.S. Williams, D.J. Smith, L. Marjavaara, S.A. Lujan, A. Chabes, T.A. Kunkel, Topoisomerase 1-mediated removal of ribonucleotides from nascent leading-strand DNA, Molecular cell, 49 (2013) 1010-1015.

[33] S.A. Lujan, J.S. Williams, A.R. Clausen, A.B. Clark, T.A. Kunkel, Ribonucleotides are signals for mismatch repair of leading-strand replication errors, Molecular cell, 50 (2013) 437-443.

[34] M.A. Reijns, B. Rabe, R.E. Rigby, P. Mill, K.R. Astell, L.A. Lettice, S. Boyle, A. Leitch, M. Keighren, F. Kilanowski, P.S. Devenney, D. Sexton, G. Grimes, I.J. Holt, R.E. Hill, M.S. Taylor, K.A. Lawson, J.R. Dorin, A.P. Jackson, Enzymatic

removal of ribonucleotides from DNA is essential for mammalian genome integrity and development, Cell, 149 (2012) 1008-1022.

[35] J.L. Sparks, H. Chon, S.M. Cerritelli, T.A. Kunkel, E. Johansson, R.J. Crouch, P.M. Burgers, RNase H2-initiated ribonucleotide excision repair, Molecular cell, 47 (2012) 980-986.

[36] N. Ohtani, M. Haruki, M. Morikawa, S. Kanaya, Molecular diversities of RNases H, Journal of bioscience and bioengineering, 88 (1999) 12-19.

[37] S.M. Cerritelli, R.J. Crouch, Ribonuclease H: the enzymes in eukaryotes, FEBS J, 276 (2009) 1494-1505.

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

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

[40] C.J. Potenski, H. Niu, P. Sung, H.L. Klein, Avoidance of ribonucleotide-induced mutations by RNase H2 and Srs2-Exo1 mechanisms, Nature, (2014).

[41] J. Sekiguchi, S. Shuman, Site-specific ribonuclease activity of eukaryotic DNA topoisomerase I, Molecular cell, 1 (1997) 89-97.

[42] J.S. Williams, A.R. Clausen, S.A. Nick McElhinny, B.E. Watts, E. Johansson, T.A. Kunkel, Proofreading of ribonucleotides inserted into DNA by yeast DNA polymerase varepsilon, DNA repair, 11 (2012) 649-656.

[43] P. Tumbale, J.S. Williams, M.J. Schellenberg, T.A. Kunkel, R.S. Williams, Aprataxin resolves adenylated RNA-DNA junctions to maintain genome integrity, Nature, 506 (2014) 111-115.

[44] D.L. Watt, E. Johansson, P.M. Burgers, T.A. Kunkel, Replication of ribonucleotide-containing DNA templates by yeast replicative polymerases, DNA repair, 10 (2011) 897-902.

[45] A.R. Clausen, M.S. Murray, A.R. Passer, L.C. Pedersen, T.A. Kunkel, Structure-function analysis of ribonucleotide bypass by B family DNA replicases, Proceedings of the National Academy of Sciences of the United States of America, 110 (2013) 16802-16807.

[46] F. Storici, K. Bebenek, T.A. Kunkel, D.A. Gordenin, M.A. Resnick, RNA-templated DNA repair, Nature, 447 (2007) 338-341.

[47] F. Lazzaro, D. Novarina, F. Amara, D.L. Watt, J.E. Stone, V. Costanzo, P.M. Burgers, T.A. Kunkel, P. Plevani, M. Muzi-Falconi, RNase H and postreplication repair protect cells from ribonucleotides incorporated in DNA, Molecular cell, 45 (2012) 99-110.

[48] J.Z. Dalgaard, Causes and consequences of ribonucleotide incorporation into nuclear DNA, Trends in genetics : TIG, 28 (2012) 592-597.

[49] T.A. Kunkel, D.A. Erie, DNA mismatch repair, Annual review of biochemistry, 74 (2005) 681-710.

[50] J. Jiricny, The multifaceted mismatch-repair system, Nature reviews. Molecular cell biology, 7 (2006) 335-346.

[51] K. Fukui, DNA mismatch repair in eukaryotes and bacteria, Journal of nucleic acids, 2010 (2010).

[52] S.A. Nick McElhinny, G.E. Kissling, T.A. Kunkel, Differential correction of lagging-strand replication errors made by DNA polymerases {alpha} and {delta}, Proceedings of the National Academy of Sciences of the United States of America, 107 (2010) 21070-21075.

[53] M.M. Ghodgaonkar, F. Lazzaro, M. Olivera-Pimentel, M. Artola-Boran, P. Cejka, M.A. Reijns, A.P. Jackson, P. Plevani, M. Muzi-Falconi, J. Jiricny, Ribonucleotides misincorporated into DNA act as strand-discrimination signals in eukaryotic mismatch repair, Molecular cell, 50 (2013) 323-332.

[54] K.W. Caldecott, Molecular biology. Ribose--an internal threat to DNA, Science, 343 (2014) 260-261.

[55] Y. Shen, F. Storici, Generation of RNA/DNA hybrids in genomic DNA by transformation using RNA-containing oligonucleotides, Journal of visualized experiments : JoVE, (2010).

[56] Y. Shen, P. Nandi, M.B. Taylor, S. Stuckey, H.P. Bhadsavle, B. Weiss, F. Storici, RNA-driven genetic changes in bacteria and in human cells, Mutation research, 717 (2011) 91-98.

[57] Y. Shen, F. Storici, Detection of RNA-templated double-strand break repair in yeast, Methods in molecular biology, 745 (2011) 193-204.

[58] Y. Shen, K.D. Koh, B. Weiss, F. Storici, Mispaired rNMPs in DNA are mutagenic and are targets of mismatch repair and RNases H, Nature structural & molecular biology, 19 (2012) 98-104.

[59] K.R. Hovatter, H.G. Martinson, Ribonucleotide-induced helical alteration in DNA prevents nucleosome formation, Proceedings of the National Academy of Sciences of the United States of America, 84 (1987) 1162-1166.

[60] S.A. Nick McElhinny, D. Kumar, A.B. Clark, D.L. Watt, B.E. Watts, E.B. Lundstrom, E. Johansson, A. Chabes, T.A. Kunkel, Genome instability due to ribonucleotide incorporation into DNA, Nature chemical biology, 6 (2010) 774-781.

[61] A.B. Clark, S.A. Lujan, G.E. Kissling, T.A. Kunkel, Mismatch repair-independent tandem repeat sequence instability resulting from ribonucleotide incorporation by DNA polymerase epsilon, DNA repair, 10 (2011) 476-482.

[62] J.E. Cho, N. Kim, Y.C. Li, S. Jinks-Robertson, Two distinct mechanisms of Topoisomerase 1-dependent mutagenesis in yeast, DNA repair, 12 (2013) 205-211.

[63] J.J. Champoux, S.J. Schultz, Ribonuclease H: properties, substrate specificity and roles in retroviral reverse transcription, FEBS J, 276 (2009) 1506-1516.

[64] Y.J. Crow, A. Leitch, B.E. Hayward, A. Garner, R. Parmar, E. Griffith, M. Ali, C. Semple, J. Aicardi, R. Babul-Hirji, C. Baumann, P. Baxter, E. Bertini, K.E. Chandler, D. Chitayat, D. Cau, C. Dery, E. Fazzi, C. Goizet, M.D. King, J. Klepper, D. Lacombe, G. Lanzi, H. Lyall, M.L. Martinez-Frias, M. Mathieu, C. McKeown, A. Monier, Y. Oade, O.W. Quarrell, C.D. Rittey, R.C. Rogers, A. Sanchis, J.B. Stephenson, U. Tacke, M. Till, J.L. Tolmie, P. Tomlin, T. Voit, B. Weschke, C.G. Woods, P. Lebon, D.T. Bonthron, C.P. Ponting, A.P. Jackson, Mutations in genes encoding ribonuclease H2 subunits cause Aicardi-Goutieres syndrome and mimic congenital viral brain infection, Nat Genet, 38 (2006) 910-916.

[65] G. Rice, T. Patrick, R. Parmar, C.F. Taylor, A. Aeby, J. Aicardi, R. Artuch, S.A. Montalto, C.A. Bacino, B. Barroso, P. Baxter, W.S. Benko, C. Bergmann, E. Bertini, R. Biancheri, E.M. Blair, N. Blau, D.T. Bonthron, T. Briggs, L.A. Brueton, H.G. Brunner, C.J. Burke, I.M. Carr, D.R. Carvalho, K.E. Chandler, H.J. Christen, P.C. Corry, F.M. Cowan, H. Cox, S. D'Arrigo, J. Dean, C. De Laet, C. De Praeter, C. Dery, C.D. Ferrie, K. Flintoff, S.G. Frints, A. Garcia-Cazorla, B. Gener, C. Goizet, F. Goutieres, A.J. Green, A. Guet, B.C. Hamel, B.E. Hayward, A. Heiberg, R.C. Hennekam, M. Husson, A.P. Jackson, R. Jayatunga, Y.H. Jiang, S.G. Kant, A. Kao, M.D. King, H.M. Kingston, J. Klepper, M.S. van der Knaap, A.J. Kornberg, D. Kotzot, W. Kratzer, D. Lacombe, L. Lagae, P.G. Landrieu, G. Lanzi, A. Leitch, M.J. Lim, J.H. Livingston, C.M. Lourenco, E.G. Lyall, S.A. Lynch, M.J. Lyons, D. Marom, J.P. McClure, R. McWilliam, S.B. Melancon, L.D. Mewasingh, M.L. Moutard, K.K. Nischal, J.R. Ostergaard, J. Prendiville, M. Rasmussen, R.C. Rogers, D. Roland, E.M. Rosser, K. Rostasy, A. Roubertie, A. Sanchis, R. Schiffmann, S. Scholl-Burgi, S. Seal, S.A. Shalev, C.S. Corcoles, G.P. Sinha, D. Soler, R. Spiegel, J.B. Stephenson, U. Tacke, T.Y. Tan, M. Till, J.L.

Tolmie, P. Tomlin, F. Vagnarelli, E.M. Valente, R.N. Van Coster, N. Van der Aa, A. Vanderver, J.S. Vles, T. Voit, E. Wassmer, B. Weschke, M.L. Whiteford, M.A. Willemsen, A. Zankl, S.M. Zuberi, S. Orcesi, E. Fazzi, P. Lebon, Y.J. Crow, Clinical and molecular phenotype of Aicardi-Goutieres syndrome, American journal of human genetics, 81 (2007) 713-725.

[66] S. Brzostek-Racine, C. Gordon, S. Van Scoy, N.C. Reich, The DNA damage response induces IFN, Journal of immunology, 187 (2011) 5336-5345.

[67] A.H. Wang, S. Fujii, J.H. van Boom, G.A. van der Marel, S.A. van Boeckel, A. Rich, Molecular structure of r(GCG)d(TATACGC): a DNA--RNA hybrid helix joined to double helical DNA, Nature, 299 (1982) 601-604.

[68] J.R. Mellema, C.A. Haasnoot, G.A. van der Marel, G. Wille, C.A. van Boeckel, J.H. van Boom, C. Altona, Proton NMR studies on the covalently linked RNA-DNA hybrid r(GCG)d(TATACGC). Assignment of proton resonances by application of the nuclear Overhauser effect, Nucleic acids research, 11 (1983) 5717-5738.

[69] M. Egli, N. Usman, S.G. Zhang, A. Rich, Crystal structure of an Okazaki fragment at 2-A resolution, Proceedings of the National Academy of Sciences of the United States of America, 89 (1992) 534-538.

[70] C.A. Haasnoot, H.P. Westerink, G.A. van der Marel, J.H. van Boom, Conformational analysis of a hybrid DNA-RNA double helical oligonucleotide in aqueous solution: d(CG)r(CG)d(CG) studied by 1D- and 2D-1H NMR spectroscopy, Journal of biomolecular structure & dynamics, 1 (1983) 131-149.

[71] S.H. Chou, P. Flynn, A. Wang, B. Reid, High-resolution NMR studies of chimeric DNA-RNA-DNA duplexes, heteronomous base pairing, and continuous base stacking at junctions, Biochemistry, 30 (1991) 5248-5257.

[72] T.N. Jaishree, G.A. van der Marel, J.H. van Boom, A.H. Wang, Structural influence of RNA incorporation in DNA: quantitative nuclear magnetic resonance refinement of d(CG)r(CG)d(CG) and d(CG)r(C)d(TAGCG), Biochemistry, 32 (1993) 4903-4911.

[73] M. Egli, N. Usman, A. Rich, Conformational influence of the ribose 2'-hydroxyl group: crystal structures of DNA-RNA chimeric duplexes, Biochemistry, 32 (1993) 3221-3237.

[74] C. Ban, B. Ramakrishnan, M. Sundaralingam, Crystal structure of the highly distorted chimeric decamer r(C)d(CGGCGCCG)r(G).spermine complex--spermine binding to phosphate only and minor groove tertiary base-pairing, Nucleic acids research, 22 (1994) 5466-5476.

[75] E.F. DeRose, L. Perera, M.S. Murray, T.A. Kunkel, R.E. London, Solution structure of the Dickerson DNA dodecamer containing a single ribonucleotide, Biochemistry, 51 (2012) 2407-2416.

[76] N. Kim, S. Jinks-Robertson, dUTP incorporation into genomic DNA is linked to transcription in yeast, Nature, 459 (2009) 1150-1153.

[77] L. Rowen, A. Kornberg, A ribo-deoxyribonucleotide primer synthesized by primase, The Journal of biological chemistry, 253 (1978) 770-774.

[78] B. Rydberg, J. Game, Excision of misincorporated ribonucleotides in DNA by RNase H (type 2) and FEN-1 in cell-free extracts, Proceedings of the National Academy of Sciences of the United States of America, 99 (2002) 16654-16659.

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

[80] H.M. Ellis, D. Yu, T. DiTizio, D.L. Court, High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides, Proceedings of the National Academy of Sciences of the United States of America, 98 (2001) 6742-6746.

[81] B. Weiss, Removal of deoxyinosine from the Escherichia coli chromosome as studied by oligonucleotide transformation, DNA repair, 7 (2008) 205-212.

[82] F. Storici, M.A. Resnick, The delitto perfetto approach to in vivo site-directed mutagenesis and chromosome rearrangements with synthetic oligonucleotides in yeast, Methods in enzymology, 409 (2006) 329-345.

[83] X.T. Li, N. Costantino, L.Y. Lu, D.P. Liu, R.M. Watt, K.S. Cheah, D.L. Court, J.D. Huang, Identification of factors influencing strand bias in oligonucleotide-mediated recombination in Escherichia coli, Nucleic acids research, 31 (2003) 6674-6687.

[84] R.R. Iyer, A. Pluciennik, V. Burdett, P.L. Modrich, DNA mismatch repair: functions and mechanisms, Chemical reviews, 106 (2006) 302-323.

[85] B.D. Harfe, S. Jinks-Robertson, DNA mismatch repair and genetic instability, Annual review of genetics, 34 (2000) 359-399.

[86] C. Dohet, R. Wagner, M. Radman, Repair of defined single base-pair mismatches in Escherichia coli, Proceedings of the National Academy of Sciences of the United States of America, 82 (1985) 503-505.

[87] F. Storici, L.K. Lewis, M.A. Resnick, In vivo site-directed mutagenesis using oligonucleotides, Nature biotechnology, 19 (2001) 773-776.

[88] F. Storici, C.L. Durham, D.A. Gordenin, M.A. Resnick, Chromosomal site-specific double-strand breaks are efficiently targeted for repair by oligonucleotides in yeast, Proceedings of the National Academy of Sciences of the United States of America, 100 (2003) 14994-14999.

[89] Y.W. Kow, G. Bao, J.W. Reeves, S. Jinks-Robertson, G.F. Crouse, Oligonucleotide transformation of yeast reveals mismatch repair complexes to be differentially active on DNA replication strands, Proceedings of the National Academy of Sciences of the United States of America, 104 (2007) 11352-11357.

[90] S. Prakash, L. Prakash, Nucleotide excision repair in yeast, Mutation research, 451 (2000) 13-24.

[91] H.C. Chiu, K.D. Koh, M. Evich, A.L. Lesiak, M.W. Germann, A. Bongiorno, E. Riedo, F. Storici, RNA intrusions change DNA elastic properties and structure, Nanoscale, 6 (2014) 10009-10017.

[92] H.E. Krokan, F. Drablos, G. Slupphaug, Uracil in DNA--occurrence, consequences and repair, Oncogene, 21 (2002) 8935-8948.

[93] A. Arudchandran, S. Cerritelli, S. Narimatsu, M. Itaya, D.Y. Shin, Y. Shimada, R.J. Crouch, The absence of ribonuclease H1 or H2 alters the sensitivity of Saccharomyces cerevisiae to hydroxyurea, caffeine and ethyl methanesulphonate: implications for roles of RNases H in DNA replication and repair, Genes Cells, 5 (2000) 789-802.

[94] L.C. Gillet, O.D. Scharer, Molecular mechanisms of mammalian global genome nucleotide excision repair, Chemical reviews, 106 (2006) 253-276.

[95] K.E. Duderstadt, K. Chuang, J.M. Berger, DNA stretching by bacterial initiators promotes replication origin opening, Nature, 478 (2011) 209-213.

[96] K.S. Bloom, Beyond the code: the mechanical properties of DNA as they relate to mitosis, Chromosoma, 117 (2008) 103-110.

[97] C. Bustamante, S.B. Smith, J. Liphardt, D. Smith, Single-molecule studies of DNA mechanics, Current opinion in structural biology, 10 (2000) 279-285.

[98] T. Nishinaka, Y. Ito, S. Yokoyama, T. Shibata, An extended DNA structure through deoxyribose-base stacking induced by RecA protein, Proceedings of the National Academy of Sciences of the United States of America, 94 (1997) 6623-6628.

[99] A. Mazurek, C.N. Johnson, M.W. Germann, R. Fishel, Sequence context effect for hMSH2-hMSH6 mismatch-dependent activation, Proceedings of the National Academy of Sciences of the United States of America, 106 (2009) 4177-4182.

[100] P. Gross, N. Laurens, L.B. Oddershede, U. Bockelmann, E.J.G. Peterman, G.J.L. Wuite, Quantifying how DNA stretches, melts and changes twist under tension, Nat Phys, 7 (2011) 731-736.

[101] G. Weber, J.W. Essex, C. Neylon, Probing the microscopic flexibility of DNA from melting temperatures, Nat Phys, 5 (2009) 769-773.

[102] A.G. Pedersen, L.J. Jensen, S. Brunak, H.H. Staerfeldt, D.W. Ussery, A DNA structural atlas for Escherichia coli, Journal of molecular biology, 299 (2000) 907-930.

[103] H.G. Garcia, P. Grayson, L. Han, M. Inamdar, J. Kondev, P.C. Nelson, R. Phillips, J. Widom, P.A. Wiggins, Biological consequences of tightly bent DNA: the other life of a macromolecular celebrity, Biopolymers, 85 (2007) 115-130.

[104] C.A. Hunter, Sequence-dependent DNA structure. The role of base stacking interactions, Journal of molecular biology, 230 (1993) 1025-1054.

[105] P.W. Rothemund, Folding DNA to create nanoscale shapes and patterns, Nature, 440 (2006) 297-302.

[106] D. Schiffels, T. Liedl, D.K. Fygenson, Nanoscale structure and microscale stiffness of DNA nanotubes, ACS nano, 7 (2013) 6700-6710.

[107] H.T. Maune, S.P. Han, R.D. Barish, M. Bockrath, W.A. Goddard, P.W.K. Rothemund, E. Winfree, Self-assembly of carbon nanotubes into two-dimensional geometries using DNA origami templates, Nat Nanotechnol, 5 (2010) 61-66.

[108] M. Chang, C.S. Yang, D.M. Huang, Aptamer-Conjugated DNA Icosahedral Nanoparticles As a Carrier of Doxorubicin for Cancer Therapy, ACS nano, 5 (2011) 6156-6163.

[109] C.T. Gibson, G.S. Watson, S. Myhra, Determination of the spring constants of probes for force microscopy/spectroscopy, Nanotechnology, 7 (1996) 259-262.

[110] A.D. Slattery, A.J. Blanch, J.S. Quinton, C.T. Gibson, Calibration of atomic force microscope cantilevers using standard and inverted static methods assisted by FIB-milled spatial markers, Nanotechnology, 24 (2013).

[111] P.J. Cumpson, C.A. Clifford, J. Hedley, Quantitative analytical atomic force microscopy: a cantilever reference device for easy and accurate AFM spring-constant calibration, Meas Sci Technol, 15 (2004) 1337-1346.

[112] H.J. Butt, B. Cappella, M. Kappl, Force measurements with the atomic force microscope: Technique, interpretation and applications, Surf Sci Rep, 59 (2005) 1-152.

[113] B. Peirce, Criterion for the Rejection of Doubtful Observations, Astron. J., 2 (1852) 161-163.

[114] B.A. Gould, On Peirce's Criterion for the Rejection of Doubtful Observations, with Tables for Facilitating its Application, Astron. J., 4 (1855) 81-87.

[115] A. Bongiorno, Energy landscape of an electron hole in hydrated DNA, The journal of physical chemistry. B, 112 (2008) 13945-13950.

[116] C.L. Cleveland, R.N. Barnett, A. Bongiorno, J. Joseph, C. Liu, G.B. Schuster, U. Landman, Steric effects on water accessability control sequence-selectivity of radical cation reactions in DNA, Journal of the American Chemical Society, 129 (2007) 8408-8409.

[117] R.N. Barnett, A. Bongiorno, C.L. Cleveland, A. Joy, U. Landman, G.B. Schuster, Oxidative damage to DNA: counterion-assisted addition of water to ionized DNA, Journal of the American Chemical Society, 128 (2006) 10795-10800.

[118] L.C. Gallington, A. Bongiorno, Thermodynamic stability limits of simple monoatomic materials, The Journal of chemical physics, 132 (2010) 174707.

[119] J.M. Aramini, S.H. Cleaver, R.T. Pon, R.P. Cunningham, M.W. Germann, Solution structure of a DNA duplex containing an alpha-anomeric adenosine: Insights into substrate recognition by endonuclease IV, Journal of molecular biology, 338 (2004) 77-91.

[120] V. Sklenar, H. Miyashiro, G. Zon, H.T. Miles, A. Bax, Assignment of the 31P and 1H resonances in oligonucleotides by two-dimensional NMR spectroscopy, FEBS letters, 208 (1986) 94-98.

[121] T.D. Goddard, D.G. Kneller, SPARKY, in, 3.33 ed., University of California, San Francisco, CA, U.S.A., 2008.

[122] J.M. Aramini, M.W. Germann, Solution structure of a DNA.RNA hybrid containing an alpha-anomeric thymidine and polarity reversals: d(ATGG-3'-3'-alphaT-5'-5'-GCTC). r(gagcaccau), Biochemistry, 38 (1999) 15448-15458.

[123] T. Morii, R. Mizuno, H. Haruta, T. Okada, An AFM study of the elasticity of DNA molecules, Thin Solid Films, 464 (2004) 456-458.

[124] S.B. Smith, L. Finzi, C. Bustamante, Direct Mechanical Measurements of the Elasticity of Single DNA-Molecules by Using Magnetic Beads, Science, 258 (1992) 1122-1126.

[125] C.G. Baumann, S.B. Smith, V.A. Bloomfield, C. Bustamante, Ionic effects on the elasticity of single DNA molecules, Proceedings of the National Academy of Sciences of the United States of America, 94 (1997) 6185-6190.

[126] J.A. Abels, F. Moreno-Herrero, T. van der Heijden, C. Dekker, N.H. Dekker, Single-molecule measurements of the persistence length of double-stranded RNA, Biophysical journal, 88 (2005) 2737-2744.

[127] E. Herrero-Galan, M.E. Fuentes-Perez, C. Carrasco, J.M. Valpuesta, J.L. Carrascosa, F. Moreno-Herrero, J.R. Arias-Gonzalez, Mechanical Identities of RNA and DNA Double Helices Unveiled at the Single-Molecule Level, Journal of the American Chemical Society, 135 (2013) 122-131.

[128] J. Zlatanova, S.M. Lindsay, S.H. Leuba, Single molecule force spectroscopy in biology using the atomic force microscope, Prog Biophys Mol Bio, 74 (2000) 37-61.

[129] A. Noy, D.V. Vezenov, J.F. Kayyem, T.J. Meade, C.M. Lieber, Stretching and breaking duplex DNA by chemical force microscopy, Chemistry & biology, 4 (1997) 519-527.

[130] P.A. Wiggins, T. van der Heijden, F. Moreno-Herrero, A. Spakowitz, R. Phillips, J. Widom, C. Dekker, P.C. Nelson, High flexibility of DNA on short length scales probed by atomic force microscopy, Nat Nanotechnol, 1 (2006) 137-141.

[131] J. Morfill, F. Kuhner, K. Blank, R.A. Lugmaier, J. Sedlmair, H.E. Gaub, B-S transition in short oligonucleotides, Biophysical journal, 93 (2007) 2400-2409.

[132] T.H. Nguyen, S.M. Lee, K. Na, S. Yang, J. Kim, E.S. Yoon, An improved measurement of dsDNA elasticity using AFM, Nanotechnology, 21 (2010).

[133] R.S. Mathew-Fenn, R. Das, P.A. Harbury, Remeasuring the double helix, Science, 322 (2008) 446-449.

[134] C. Yuan, H. Chen, X.W. Lou, L.A. Archer, DNA bending stiffness on small length scales, Physical review letters, 100 (2008) 018102.

[135] C. Bustamante, J.F. Marko, E.D. Siggia, S. Smith, Entropic elasticity of lambda-phage DNA, Science, 265 (1994) 1599-1600.

[136] O.C. Lee, J.H. Jeon, W. Sung, How double-stranded DNA breathing enhances its flexibility and instability on short length scales, Phys Rev E, 81 (2010).

[137] J. Wong, A. Chilkoti, V.T. Moy, Direct force measurements of the streptavidin-biotin interaction, Biomolecular engineering, 16 (1999) 45-55.

[138] M. Grandbois, M. Beyer, M. Rief, H. Clausen-Schaumann, H.E. Gaub, How strong is a covalent bond?, Science, 283 (1999) 1727-1730.

[139] H.C. Chiu, S. Kim, C. Klinke, E. Riedo, Morphology dependence of radial elasticity in multiwalled boron nitride nanotubes, Appl Phys Lett, 101 (2012).

[140] H.C. Chiu, B. Ritz, S. Kim, E. Tosatti, C. Klinke, E. Riedo, Sliding on a nanotube: interplay of friction, deformations and structure, Advanced materials, 24 (2012) 2879-2884.

[141] M. Lucas, X. Zhang, I. Palaci, C. Klinke, E. Tosatti, E. Riedo, Hindered rolling and friction anisotropy in supported carbon nanotubes, Nature materials, 8 (2009) 876-881.

[142] T.D. Li, J.P. Gao, R. Szoszkiewicz, U. Landman, E. Riedo, Structured and viscous water in subnanometer gaps, Phys Rev B, 75 (2007).

[143] D. Ortiz-Young, H.C. Chiu, S. Kim, K. Voitchovsky, E. Riedo, The interplay between apparent viscosity and wettability in nanoconfined water, Nat Commun, 4 (2013).

[144] F. Lankas, J. Sponer, P. Hobza, J. Langowski, Sequence-dependent elastic properties of DNA, Journal of molecular biology, 299 (2000) 695-709.

[145] S. Hormeno, B. Ibarra, J.L. Carrascosa, J.M. Valpuesta, F. Moreno-Herrero, J.R. Arias-Gonzalez, Mechanical properties of high-G.C content DNA with a-type base-stacking, Biophysical journal, 100 (2011) 1996-2005.

[146] G.G. Prive, U. Heinemann, S. Chandrasegaran, L.S. Kan, M.L. Kopka, R.E. Dickerson, Helix geometry, hydration, and G.A mismatch in a B-DNA decamer, Science, 238 (1987) 498-504.

[147] D.G. Gorenstein, Conformation and Dynamics of DNA and Protein-DNA Complexes by P-31 Nmr, Chemical reviews, 94 (1994) 1315-1338.

[148] J.S. Williams, T.A. Kunkel, Ribonucleotides in DNA: origins, repair and consequences, DNA repair, 19 (2014) 27-37.

[149] L.I. Grossman, R. Watson, J. Vinograd, The presence of ribonucleotides in mature closed-circular mitochondrial DNA, Proceedings of the National Academy of Sciences of the United States of America, 70 (1973) 3339-3343.

[150] C.J. Potenski, H.L. Klein, How the misincorporation of ribonucleotides into genomic DNA can be both harmful and helpful to cells, Nucleic acids research, (2014).

[151] H. Zhu, S. Shuman, Bacterial nonhomologous end joining ligases preferentially seal breaks with a 3'-OH monoribonucleotide, The Journal of biological chemistry, 283 (2008) 8331-8339.

[152] A.B. Clark, M.E. Cook, H.T. Tran, D.A. Gordenin, M.A. Resnick, T.A. Kunkel, Functional analysis of human MutSalpha and MutSbeta complexes in yeast, Nucleic acids research, 27 (1999) 736-742.

[153] A. Morrison, J.B. Bell, T.A. Kunkel, A. Sugino, Eukaryotic DNA polymerase amino acid sequence required for 3'----5' exonuclease activity, Proceedings of the National Academy of Sciences of the United States of America, 88 (1991) 9473-9477.

[154] Y.H. Jin, R. Obert, P.M. Burgers, T.A. Kunkel, M.A. Resnick, D.A. Gordenin, The 3'-->5' exonuclease of DNA polymerase delta can substitute for the 5' flap endonuclease Rad27/Fen1 in processing Okazaki fragments and preventing genome instability, Proceedings of the National Academy of Sciences of the United States of America, 98 (2001) 5122-5127.

[155] Z.F. Pursell, I. Isoz, E.B. Lundstrom, E. Johansson, T.A. Kunkel, Yeast DNA polymerase epsilon participates in leading-strand DNA replication, Science, 317 (2007) 127-130.

[156] K. Schutz, J.R. Hesselberth, S. Fields, Capture and sequence analysis of RNAs with terminal 2',3'-cyclic phosphates, Rna, 16 (2010) 621-631.

[157] R.J. Kokoska, S.D. McCulloch, T.A. Kunkel, The efficiency and specificity of apurinic/apyrimidinic site bypass by human DNA polymerase eta and Sulfolobus solfataricus Dpo4, The Journal of biological chemistry, 278 (2003) 50537-50545.

[158] T. Barrett, S.E. Wilhite, P. Ledoux, C. Evangelista, I.F. Kim, M. Tomashevsky, K.A. Marshall, K.H. Phillippy, P.M. Sherman, M. Holko, A. Yefanov, H. Lee, N. Zhang, C.L. Robertson, N. Serova, S. Davis, A. Soboleva, NCBI GEO: archive for functional genomics data sets--update, Nucleic acids research, 41 (2013) D991-995.

[159] A.R. Quinlan, I.M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features, Bioinformatics, 26 (2010) 841-842.

[160] R.K. Dale, B.S. Pedersen, A.R. Quinlan, Pybedtools: a flexible Python library for manipulating genomic datasets and annotations, Bioinformatics, 27 (2011) 3423-3424.

[161] C.C. Siow, S.R. Nieduszynska, C.A. Muller, C.A. Nieduszynski, OriDB, the DNA replication origin database updated and extended, Nucleic acids research, 40 (2012) D682-686.

[162] N. Yabuki, H. Terashima, K. Kitada, Mapping of early firing origins on a replication profile of budding yeast, Genes Cells, 7 (2002) 781-789.

[163] Y. Zhang, T. Liu, C.A. Meyer, J. Eeckhoute, D.S. Johnson, B.E. Bernstein, C. Nusbaum, R.M. Myers, M. Brown, W. Li, X.S. Liu, Model-based analysis of ChIP-Seq (MACS), Genome biology, 9 (2008) R137.

[164] R.R. Sokal, F.J. Rohlf, Biometry : the principles and practice of statistics in biological research, 3rd ed., W.H. Freeman, New York, 1995.

[165] B.S. Remus, S. Shuman, Distinctive kinetics and substrate specificities of plant and fungal tRNA ligases, Rna, 20 (2014) 462-473.

[166] D.A. Cooper, B.K. Jha, R.H. Silverman, J.R. Hesselberth, D.J. Barton, Ribonuclease L and metal-ion-independent endoribonuclease cleavage sites in host and viral RNAs, Nucleic acids research, 42 (2014) 5202-5216.

[167] C.W.J. Smith, RNA-protein interactions : a practical approach, Oxford University Press, Oxford ; New York, 1998.

[168] J. Lhomme, J.F. Constant, M. Demeunynck, Abasic DNA structure, reactivity, and recognition, Biopolymers, 52 (1999) 65-83.

[169] V. Bailly, W.G. Verly, Possible roles of beta-elimination and delta-elimination reactions in the repair of DNA containing AP (apurinic/apyrimidinic) sites in mammalian cells, Biochem J, 253 (1988) 553-559.

[170] T. Lindahl, S. Ljungquist, W. Siegert, B. Nyberg, B. Sperens, DNA N-glycosidases: properties of uracil-DNA glycosidase from Escherichia coli, The Journal of biological chemistry, 252 (1977) 3286-3294.

[171] F. Foury, T. Roganti, N. Lecrenier, B. Purnelle, The complete sequence of the mitochondrial genome of Saccharomyces cerevisiae, FEBS letters, 440 (1998) 325-331.

[172] J.M. Gerhold, A. Aun, T. Sedman, P. Joers, J. Sedman, Strand invasion structures in the inverted repeat of Candida albicans mitochondrial DNA reveal a role for homologous recombination in replication, Molecular cell, 39 (2010) 851-861.

[173] C.T. Moraes, What regulates mitochondrial DNA copy number in animal cells?, Trends in genetics : TIG, 17 (2001) 199-205.

[174] J.W. Szostak, R. Wu, Unequal crossing over in the ribosomal DNA of Saccharomyces cerevisiae, Nature, 284 (1980) 426-430.

[175] J. Hani, H. Feldmann, tRNA genes and retroelements in the yeast genome, Nucleic acids research, 26 (1998) 689-696.

[176] P.A. Mieczkowski, F.J. Lemoine, T.D. Petes, Recombination between retrotransposons as a source of chromosome rearrangements in the yeast Saccharomyces cerevisiae, DNA repair, 5 (2006) 1010-1020.

[177] A. El Hage, S. Webb, A. Kerr, D. Tollervey, Genome-Wide Distribution of RNA-DNA Hybrids Identifies RNase H Targets in tRNA Genes, Retrotransposons and Mitochondria, PLoS genetics, 10 (2014) e1004716.

[178] A.R. Clausen, S.A. Lujan, A.B. Burkholder, C.D. Orebaugh, J.S. Williams, M.F. Clausen, E.P. Malc, P.A. Mieczkowski, D.C. Fargo, D.J. Smith, T.A. Kunkel, Tracking replication enzymology in vivo by genome-wide mapping of ribonucleotide incorporation, Nature structural & molecular biology, (2015).

[179] Y. Daigaku, A. Keszthelyi, C.A. Muller, I. Miyabe, T. Brooks, R. Retkute, M. Hubank, C.A. Nieduszynski, A.M. Carr, A global profile of replicative polymerase usage, Nature structural & molecular biology, (2015).

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

[181] Y. Cai, N.E. Geacintov, S. Broyde, Ribonucleotides as nucleotide excision repair substrates, DNA repair, 13 (2014) 55-60.

[182] J.R. Vance, T.E. Wilson, Repair of DNA strand breaks by the overlapping functions of lesion-specific and non-lesion-specific DNA 3' phosphatases, Molecular and cellular biology, 21 (2001) 7191-7198.

[183] B.P. Nichols, O. Shafiq, V. Meiners, Sequence analysis of Tn10 insertion sites in a collection of Escherichia coli strains used for genetic mapping and strain construction, Journal of bacteriology, 180 (1998) 6408-6411.

[184] M. Singer, T.A. Baker, G. Schnitzler, S.M. Deischel, M. Goel, W. Dove, K.J. Jaacks, A.D. Grossman, J.W. Erickson, C.A. Gross, A collection of strains containing genetically linked alternating antibiotic resistance elements for genetic mapping of Escherichia coli, Microbiological reviews, 53 (1989) 1-24.

[185] T. Baba, T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K.A. Datsenko, M. Tomita, B.L. Wanner, H. Mori, Construction of Escherichia coli K-12 in-

frame, single-gene knockout mutants: the Keio collection, Molecular systems biology, 2 (2006) 2006 0008.

[186] C.B. Brachmann, A. Davies, G.J. Cost, E. Caputo, J. Li, P. Hieter, J.D. Boeke, Designer deletion strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications, Yeast, 14 (1998) 115-132.

[187] K.D. Koh, S. Balachander, J.R. Hesselberth, F. Storici, Ribose-seq: global mapping of ribonucleotides embedded in genomic DNA, Nature methods, (2015).

[188] M. Itaya, R.J. Crouch, Correlation of activity with phenotypes of Escherichia coli partial function mutants of rnh, the gene encoding RNase H, Molecular & general genetics : MGG, 227 (1991) 433-437.

[189] A.E. Karu, Y. Sakaki, H. Echols, S. Linn, The gamma protein specified by bacteriophage gamma. Structure and inhibitory activity for the recBC enzyme of Escherichia coli, The Journal of biological chemistry, 250 (1975) 7377-7387.

[190] N.L. Sternberg, R. Maurer, Bacteriophage-mediated generalized transduction in Escherichia coli and Salmonella typhimurium, Methods in enzymology, 204 (1991) 18-43.