DETECTION OF PRODUCTS OF GUANINE OXIDATION IN RNA

by

Anton Alenko

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Chemistry

The University of Utah

May 2018

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Anton Alenko**

has been approved by the following supervisory committee members:

Cynthia J. Burrows	, Chair	**12/14/2017** Date Approved
Vahe Bandarian	, Member	**12/14/2017** Date Approved
Matthew T. Kieber-Emmons	, Member	**12/14/2017** Date Approved
Marc D. Porter	, Member	**12/15/2017** Date Approved
Timothy G. Formosa	, Member	**12/15/2017** Date Approved

and by **Cynthia J. Burrows** , Chair/Dean of

the Department/College/School of **Chemistry**

and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

Reactive oxygen species can target nucleobases of DNA and RNA. One of the major products of oxidation of nucleic acids by ROS is 8-oxo-7,8-dihydroguanine (OG) which is prone to further oxidation to yield the two hydantoin lesions, 5-guanidinohydantoin (Gh) and spiroiminodihydantoin (Sp). These three lesions are mutagenic and their occurrence can significantly affect normal cell functions or cell survival. The OG, Gh, and Sp are thus removed from DNA and RNA and then are excreted from the cells in the form of free nucleobases. Concentrations of these species increase during oxidative stress and can be used as an indicator of this condition. The first part of this dissertation is devoted to development of aptamer-based sensors that can be used for detection of Gh in the free base form. We described synthesis of Gh and used an attachment-free SELEX method for *in vitro* selection of aptamers for the hydantoin.

Recent studies have shown that OG can play a regulatory role in both DNA and RNA. However, if oxidation of guanine in nucleic acids plays a regulatory role, it has to be directed toward specific sites. Testing whether this is true can be greatly facilitated by mapping positions of oxidation sites. The second part of this dissertation is focused on investigating whether sequencing of cDNA after reverse transcription (RT) of a target RNA can be used for mapping oxidation sites. In order to do so, we first studied the composition of canonical bases that are inserted opposite OG, Gh, and Sp in RNA by RT enzymes. We have found that SuperScript III is capable of efficiently bypassing the

studied lesions resulting in insertion of a mixture of A and C opposite OG and mixture of A and G opposite Gh and Sp. These data indicated that reverse transcription of oxidized bases is a viable approach for mapping their positions in RNA. Additionally, we have completed the work required to quantify the mutation signature of oxidized lesions using Illumina sequencing and studied the dependence of the efficiency of deletion bypass of hydantoin lesions on the sequence context.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SCHEMES

LIST OF ABBREVIATIONS

2Ih	5-carboxamido-5-formamido-2-iminohydantoin
aRP	antisense reverse primer
A	adenine
AMV	avian myeloblastosis virus
AP	affinity purification
ATP	adenosine triphosphate
bp	base pair
BDL	below detection limit
BER	base excision repair
cDNA	complementary DNA
ca^{5}C	5-carboxymethyl cytosine
C	cytosine
CE	capillary electrophoresis
ChIP-seq	chromatin immunoprecipitation
cpm	counts per minute
dA	2'-deoxyadenosine
dATP	2'-deoxyadenosine triphosphate
dC	2'-deoxycytidine
dCTP	2'-deoxycytidine triphosphate

dG	2'-deoxyguanosine
dGMP	2'-deoxyguanosine monophosphate
dGDP	2'-deoxyguanosine diphosphate
dGTP	2'-deoxyguanosine triphosphate
dGh	2'-deoxy-5-guanidinohydantoin
dNTP	2'-deoxyribonucleoside triphosphate
dOG	8-oxo-2'-deoxyguanosine
dOGMP	8-oxo-2'-deoxyguanosine monophosphate
dOGDP	8-oxo-2'-deoxyguanosine diphosphate
dOGTP	8-oxo-2'-deoxyguanosine triphosphate
ds	double-stranded
dSp	2'-deoxyspiroiminodidydantoin
dT	2'-deoxythymidine
dTTP	2'-deoxythymidine triphosphate
DAZAP1	DAZ-associated protein 1
DMS	dimethyl sulfate
DNA	2'-deoxyribonucleic acid
e^-	electron
E. coli	Escherichia coli
ECD	electronic circular dichroism
ELAVL1	ELAV-like protein 1
ELISA	enzyme-linked immunosorbent assay
ESI-MS	electrospray ionization mass spectrometry

f^5C	5-formylcytosine
gDNA	genomic DNA
G	guanine
GDP	guanosine diphosphate
Gh	5-guanidinohydantoin
GK	guanylate kinase
GMP	guanosine monophosphate
GTP	guanosine triphosphate
hm^5C	5-hydroxymethylcytosine
HeLa	Henrietta Lacks cervix cancer cell line
HIV1	human immunodeficiency virus type 1
HNRNPC	heterogeneous nuclear ribonucleoprotein C
HNRNPD	heterogeneous nuclear ribonucleoprotein D
HPLC	high-performance liquid chromatography
I	inosine
Ia	iminoallantoin
Iz	2-amino-imidazolone
LC-MS	liquid chromatography mass spectrometry
LC-MS/MS	liquid chromatography tandem mass spectrometry
m^1A	1-methyladenine
m^6A	N^6-methyladenine
$m^6{}_2A$	N^6,N^6-dimethyladenine
m^3C	3-methylcytosine

m^5C	5-methylcytosine
m^1G	1-methylguanine
$m^2{}_2G$	N^2,N^2-dimethylguanine
m^6G	O^6-methylguanine
m^7GTP	7-methylguanosine triphosphate
miRNA	micro RNA
mRNA	messenger RNA
MB	methylene blue
MMLV	ribosomal RNA
MTH1	MutT homolog 1
MTH2	MutT homolog 2
nt	nucleotide
NDPK	nucleoside-diphosphate kinase
NEIL1	Nei-like DNA glycosylase 1
NEIL3	Nei-like DNA glycosylase 3
NER	nucleotide excision repair
NGS	next-generation sequencing
NHE	normal hydrogen electrode
NTC	no template control
NUDT5	NUDIX hydrolyze 5
***O^6*-BnG**	O^6-benzylguanine
OG	8-oxo-7,8-dihydroguanine, 8-oxoguanine
OGMP	8-oxoguanosine monophosphate

OGDP	8-oxoguanosine diphosphate
OGTP	8-oxoguanosine triphosphate
OGG1	8-oxoguanine DNA glycosylase
OG-seq	8-oxoguanine sequencing
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	protein data bank
PNP	purine nucleoside phosphorylase
PNPase	polynucleotide phosphorylase
Pyr	pyridine
rG	guanosine
rGh	5-guanidinohydantoin nucleoside
rOG	8-oxoguanosine
rRNA	ribosomal RNA
rSp	spiroiminodihydantoin nucleoside
rt	room temperature
R	purine base (A or G)
RNA	ribonucleic acid
RNR	ribonucleotide reductase
ROS	reactive oxygen species
RSD	relative standard deviation
RT	reverse transcription
RT-PCR	reverse transcription polymerase chain reaction

ss	single-stranded
SD	standard deviation
SELEX	systematic evolution of ligands by exponential enrichment
SF3B4	splicing factor 3B subunit 4
SHAPE	selective 2'-hydroxyl acylation analyzed by primer extension
SMRT	single molecule real-time
Sp	spiroiminodihydantoin
SSB	strand separation buffer
tRNA	transfer RNA
tiRNA	tRNA halves
tRFs	tRNA fragments
T	thymine
TBE	Tris-borate-EDTA
TGIRT	thermostable group II intron reverse transcriptase
U	uracil
UV	ultraviolet
UV-vis	ultraviolet visible
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
XMRV	xenotropic murine leukemia virus-related virus
XRN1	5'-3' exoribonuclease 1
YB-1	Y box-binding protein 1
Z	2,2-diamino-oxazolone
Ψ	pseudouracil

ACKNOWLEDGEMENTS

First, I would like to thank my advisor Professor Cynthia J. Burrows for her understanding, generosity, continuous support, and patient guidance throughout the past five years. One could not wish for a better mentor than Cindy as she has always encouraged me to pursue every idea no matter how out there it was, has never failed to intellectually challenge me during our discussions, and has provided a fantastic environment for performing research. I would also like to express my sincerest gratitude to Professor Aaron M. Fleming for his invaluable help, patience, and friendship. His knowledgeable advice has greatly aided me on multiple occasions and made going through the challenges of graduate school much easier. Both Cindy and Aaron have been wonderful role models for me during my time at the University of Utah.

I am very thankful to the past and present members of the Burrows group for their help, friendship, and amazing work atmosphere that they helped create. Especially, I would like to thank Dr. Yun Ding, Judy Zhu, Dr. Jan Riedl, Dr. Anna H. Wolna, Dr. Omar R. Alshykhly, Dr. Pranjali Ghude, Dr. Na An, Dr. Tao Zeng, Dr. Xibo Li, Tony Hanelli, Megan E. Browning, Samuel C. J. Redstone, and Jingwei Ma.

I would also like to thank my graduate committee members, Professor Matthew T. Kieber-Emmons, Professor Vahe Bandarian, Professor Marc D. Porter, Professor Timothy G. Formosa, and Professor Jennifer M. Heemstra for their helpful feedback and the challenging questions during my oral exams.

I am eternally grateful to my incredibly intelligent and talented wife Dr. Misha Rai for her unwavering support and immense help, my mother Tatiana V. Alenko for constantly encouraging me to follow my aspirations and achieve my goals, my father Yuriy N. Alenko for inciting my interest in science from an early age, my uncle Dr. Mikhail A. Eronyan for inspiring me to pursue the field of chemistry, as well as to my other family members and friends. I would not be where I am right now without them. I would also like to acknowledge the comfort and emotional support provided during these years by my special feline companion, Dara.

CHAPTER 1

INTRODUCTION

Oxidative stress

Multiple different reactive oxygen species (ROS), such as singlet oxygen ($^1O_2$), hydroxyl radical ($^{\cdot}OH$), or hydrogen peroxide (H_2O_2), are produced in human cells every day. ROS are formed during metabolism as a byproduct or induced by exogenous sources, such as UV radiation and environmental toxins (Figure 1.1).[1-4] These unstable molecules may react with proteins or DNA and RNA leading to modifications of nucleobases, strand breaks, and inter-, or intramolecular cross-links.[5-7] Occurrence of oxidative damage presents a significant challenge for cell survival, and over the course of evolution multiple mechanisms of repair (DNA and RNA) or selective degradation (RNA and proteins) of damaged biopolymers have been selected.[8-11] Under normal conditions, there is a balance between damage and repair; disruption of this balance caused by overproduction of ROS or deficiency in repair leads to a condition called oxidative stress.[12] Under oxidative stress, cells do not have sufficient capacity to repair oxidative damage, which results in accumulation of the products of oxidative damage. This in turn leads to arrest of multiple cell functions,[13-15] and, ultimately, to apoptosis or irreversible mutations in the genetic code during replication due to improper readout from damaged DNA sites.[15, 18] Possible long-term consequences of oxidative stress include aging, cancer, and neurodegenerative

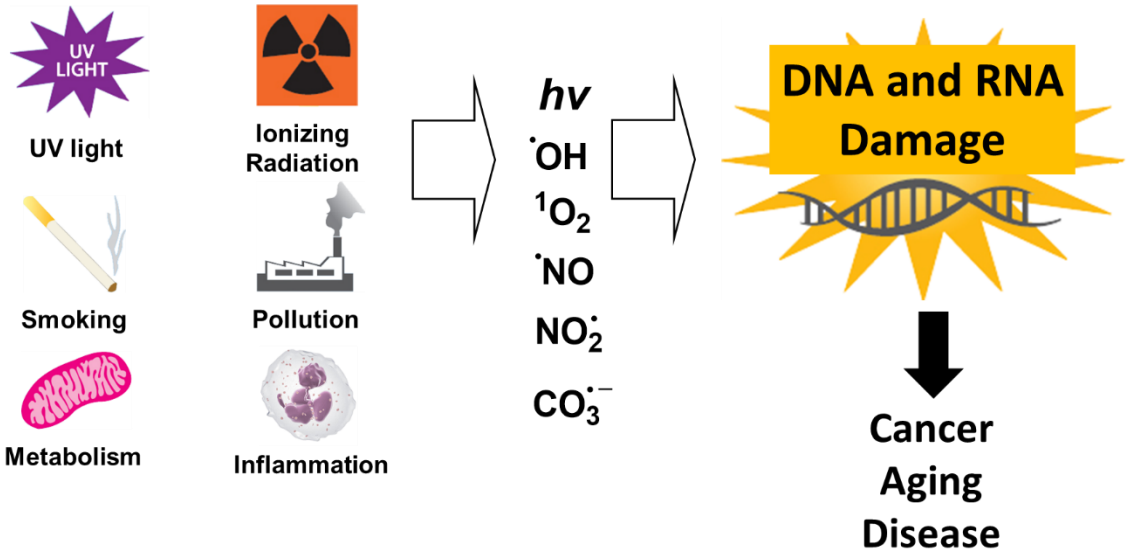


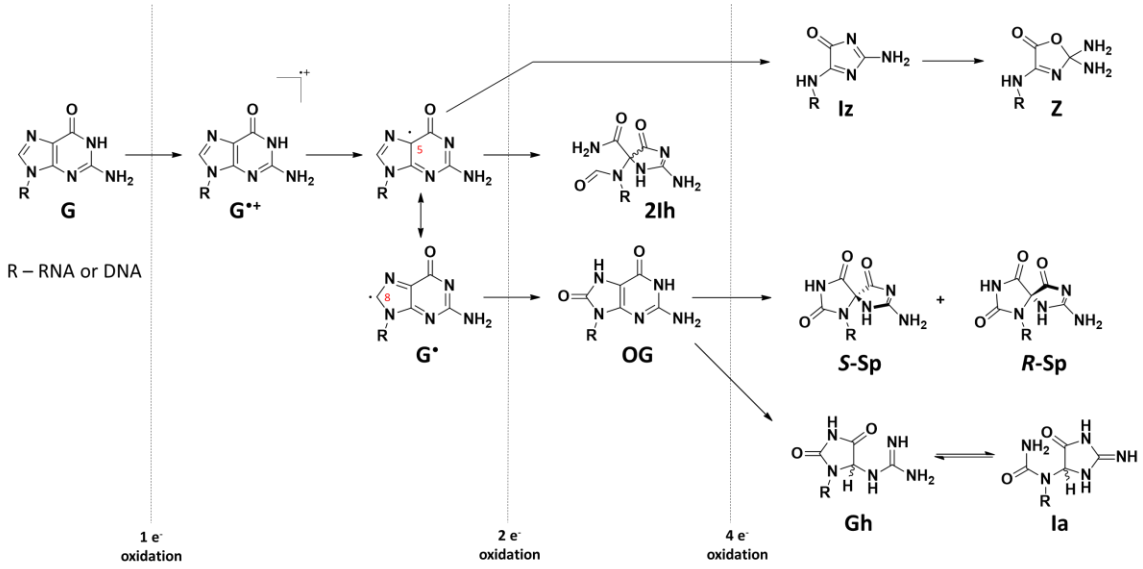
Figure 1.1. DNA and RNA damage from endogenous and exogenous reactive oxygen species.

diseases.[12, 15-17]

Guanine oxidation

Among the nucleobases of DNA and RNA, guanine has the lowest redox potential at 1.29 V vs NHE, which makes it the major target for oxidizing agents.[19] One-electron oxidation of guanine leads to formation of guanine radical cation ($G^{\cdot+}$) that can act as an oxidant for the neighboring guanines leading to transfer of the electron hole to G with the highest occupied molecular orbital (typically the 5' guanine in contiguous guanine sequences).[20-21] The acidity of the N^1-H bond in $G^{\cdot+}$ is significantly higher than in guanine, and $G^{\cdot+}$ quickly undergoes deprotonation leading to the neutral guanine radical ($G^{\cdot}$) that can react with a number of free radicals or electrophiles present in its vicinity such as carbonate radical ($CO_3^{\cdot-}$), superoxide ($O_2^{\cdot-}$), water, or primary amines yielding C5 or C8 guanine adducts (Scheme 1.1).[22-27]

One of the most abundant DNA and RNA lesions is a product of guanine two-electron oxidation along the C8 pathway, 8-oxo-7,8-dihydroguanine (OG).[28] Each human cell produces several hundreds of OG lesions per day under normal physiological conditions, but the number can significantly increase due to disturbance in the intracellular redox state (oxidative stress).[29] OG is a mutagenic lesion that has been found in RNA and DNA that can form stable base pairs with both adenine and cytosine (Figure 1.2).[30-31] Formation of an OG-A base pair is believed to be responsible for G-to-T transversions during DNA replication and a change in the target recognition by the seed sequence of miRNA.[32-33] The redox potential of OG is even lower than that of guanine (0.74 V vs NHE) making it predisposed to further oxidation along the C5 pathway.[19]



Scheme 1.1. Major pathways of guanine oxidation.

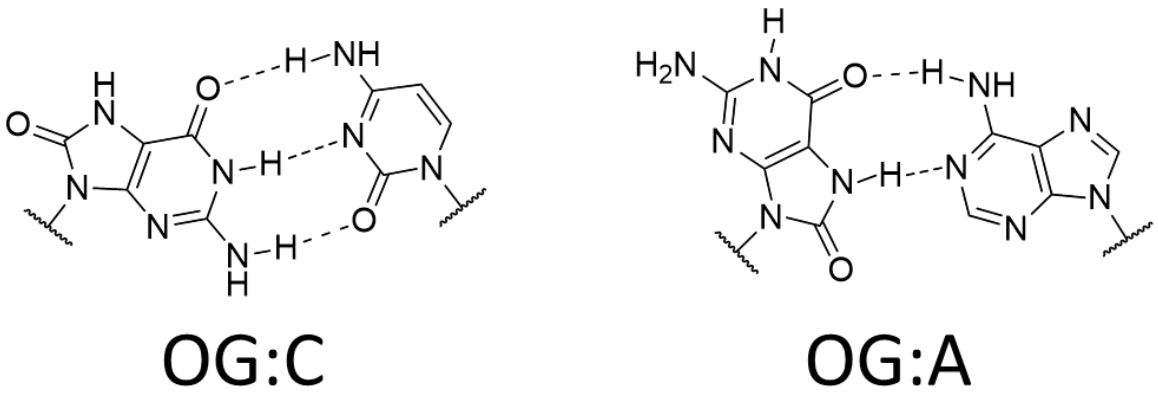


Figure 1.2. Structures of OG-C and OG-A base pairs.

Major products of OG two-electron oxidation are two hydantoin lesions, 5-guanidinohydantoin (Gh) and spiroiminodihydantoin (Sp).[23, 26-27, 34-35] Formation of these lesions is highly dependent on the pH and conformation, where Gh is a major oxidation product in double-stranded (ds) context and at pH below 5.8 in single-stranded (ss) conformations, while Sp is a major oxidation product in nucleosides, single-stranded, and G-quadruplex conformations and at pH above 5.8.[21, 36-39] Formation of both hydantoin lesions leads to creation of a new stereocenter, and thus, they are synthesized as a mixture of enantiomers (free base) or diastereomers (nucleoside/tide). However, only the stereoisomers of Sp are conformationally stable, while Gh stereoisomers exist in an equilibrium state and easily interconvert. Absolute *R* and *S* configurations for Sp diastereomers have been recently assigned by the Burrows laboratory.[40] Another interesting aspect is that Gh exists in equilibrium with iminoallantoin (Ia) with the ratio of the two being dependent on the pH.[41] Gh and Sp have been observed in cellular DNA and RNA oligonucleotides and are mutagenic due to their base pairing preference for adenine and guanine leading to G-to-T and G-to-C transversions.[34, 42-49]

RNA oxidation and its consequences

Oxidative damage to RNA has not been studied as extensively as oxidative damage to DNA. Thus, when talking about oxidative damage to RNA it is essential to draw parallels between DNA and RNA. To do that, we first have to discuss the relative quantities of DNA and RNA in cells. About 90% of cellular RNA is directly involved in protein synthesis [ribosomal RNA (rRNA), transfer RNA (tRNA), and messenger RNA (mRNA)]. Thus, the ratio of RNA to DNA in cells is highly dependent on cell type, the rate of cell division, and

the cell cycle phase.[50-52] Among normal human cells the lowest RNA:DNA ratio is observed in blood cells (0.11-0.36)[53-54] with RNA:DNA ratio in other cell types being within the range of 1-2 (Table 1.1).[55-57] Cancerous cells have higher RNA:DNA ratios that can be as high as 4 in malignant tumors.[52, 57]

There are three studies of the amount of OG in RNA that can be used to assess relative quantities of OG in RNA and DNA. Based on the data reported, OG in RNA is 2-25 times more abundant than in DNA (up to 5 OG molecules per 10^5 guanine residues in RNA).[45, 58-59] Such a wide range in OG concentration is caused by different cell types (normal lung,[58] colon,[45] and liver[45] tissue and HeLa cells[59]) and the detection methodology used. While in all these studies, DNA and RNA are digested down to nucleosides followed by their analysis via liquid chromatography-mass spectrometry (LC-MS), only in the study by Mangerich et al., has the background level of OG in DNA and RNA been measured; in the other two studies, formation of OG has been induced by treating cells with hydrogen peroxide.[45, 58-59] Based on results from these reports and considering that there can be as much as 4 times more RNA than DNA in cells, the amount of OG in RNA could be 100-times higher than in DNA.

One of the reasons why for a long time oxidation of RNA has been overlooked is the transient nature of RNA molecules. Genomic DNA (gDNA), the sole carrier of genetic information in cells, is replicated only in the cells that enter the DNA synthesis phase (S-phase) during cell division, while under normal circumstances in eukaryotes, there is only one copy of each chromosome and a significant amount of cellular resources are devoted to maintaining the integrity of gDNA because accumulation of gDNA damage can be detrimental for cell survival.[15, 60-61] At the same time, cells have multiple copies of RNA

Table 1.1. Ratio of total cellular RNA to total cellular DNA in different types of human cells.

Cell type	RNA:DNA
Breast cancer[57]	4
Noncancerous breast tumor[57]	2
Normal breast[57]	1.1
Thyroid cancer[57]	2.5
Noncancerous thyroid tumor[57]	1.3
Lung cancer[57]	2
Normal lung[57]	1
Spleen cancer[57]	1.4
Normal spleen[57]	1
Normal brain (frontal cortex) [55]	2.1
Normal liver[56]	1.4
Normal lymphocytes[53]	0.26

molecules, and RNA is continuously produced by transcribing DNA, as well as degraded by exo- and endonucleases via multiple pathways.[62-66] However, RNA becomes less expendable when moving from lower to higher organisms. The typical range of mRNA half-life in bacteria is 3-8 min,[67-69] while in human cells the average half-life of mRNA is close to 10 h,[66] and half-life values for stable RNA species such as tRNA and rRNA exceed several days.[70-72] Considering longer turnover of RNA in human cells and disruption of a number of cell functions caused by RNA damage, it is not surprising that recently several mechanisms of coping with said damage have been discovered.[9, 11, 59, 73-79] The first mechanism is repair of damaged bases, currently only one DNA repair-associated enzyme, the dioxygenase AlkB that catalyzes oxidative demethylation of 3-methylcytosine (m^3C), 1-methylguanine (m^1G), 3-methylthymine (m^3T), and 1-methyladenine (m^1A), is known to act on RNA.[74-75, 80] The second mechanism is sequestration of damaged RNA molecules followed by their selective degradation. This mechanism appears to be the predominant pathway for RNA.[11, 73] The protein that is believed to play a central role in degradation of OG-containing RNA is polynucleotide phosphorylase (PNPase).[59, 76-78] PNPase is a bifunctional enzyme with 3'-5' phosphorolytic exoribonuclease activity digesting RNA to diphosphates and is also capable of 3'-terminal polyadenylation of RNA.[81] This protein is an integral part of a multiprotein complex responsible for processing and degrading RNA (degradosome in bacteria or exosome in archaea and eukaryotes).[82-83] PNPase has been shown to have an increased affinity for OG-containing RNA,[76] and cell viability under oxidative stress, as well as the amount of OG in RNA, are heavily affected by PNPase overexpression or knockdown.[59, 77-78] Several other proteins have been identified to have affinity for OG in RNA: Y box-binding protein 1 (YB-1),[79] splicing factor 3B subunit 4

(SF3B4),[84] heterogeneous nuclear ribonucleoprotein D0 (HNRNPD),[84] heterogeneous nuclear ribonucleoprotein C (HNRNPC),[84] DAZ-associated protein 1 (DAZAP1),[84] and ELAV-like protein 1 (ELAVL1).[84] However, as of now, not enough work has been done to claim that they can play a role in RNA oxidative damage surveillance. Another two peculiar details about cellular response to RNA oxidation are that (1) levels of some OG-binding proteins (PNPase and HNRNPD) rapidly decrease in cells under oxidative stress[77, 84] and that (2) efficient degradation of OG-containing mRNA requires removal of the 5' m^7GTP cap,[18] meaning that it predominantly follows the 5'-3' XRN1 digestion pathway.[85]

Degradation of OG-containing RNA on its own would not be sufficient for removal of OG from cellular RNA because it leads to formation of 8-oxoguanosine diphosphate (OGDP) that could be used in RNA synthesis. Three enzymes are known to be responsible for removal of 8-oxo-2'-deoxyguanosine triphosphate (dOGTP) and 8-oxo-2'-deoxyguanosine diphosphate (dOGDP) from the nucleotide pool in mammalian cells, reducing the probability of incorporation of OG in DNA: MutT homolog 1 (MTH1), MutT homolog 2 (MTH2), and NUDIX hydrolyze 5 (NUDT5).[86-91] Two of them, MTH1 and NUDT5, have demonstrated the ability to recognize ribonucleotide triphosphates as substrates.[91-93] MTH1 is capable of hydrolyzing 8-oxoguanosine triphosphate (OGTP) and OGDP with ~10-times lower catalytic efficiency than dOGTP and dOGDP.[91-93] NUDT5 has been shown to act on OGDP with ~100-times lower catalytic efficiency than on dOGDP.[92] It is not known whether these enzymes can act on tri- or diphosphates of dGh, rGh, dSp, or rSp. The preference of these enzymes for di- and triphosphates of 2'-deoxyribonucleotides can probably be justified by the fact that their concentration in mammalian cells is 10-100 times lower compared to their ribo counterparts.[94-96] The

scheme showing the fate of different OG-containing species is shown on Figures 1.3 and 1.4.[9, 92, 97-105] It should be noted that degradation of 8- oxoguanosine (rOG) di- and triphosphates to monophosphates is sufficient for permanent removal of OG from the triphosphate pool used for RNA synthesis due to the fact that guanylate kinase (GK) does not act on 8-oxoguanosine monophosphate (OGMP) or 8-oxo-2'-deoxyguanosine monophosphate (dOGMP),[97-98] and that ribo- and 2'-deoxyribonucleotide pools of OG do not intermix due to OGDP not being a substrate for ribonucleotide reductase (RNR).[92]

Oxidative damage to RNA and its consequences have been predominantly studied in patients suffering from neurodegenerative diseases or severe mental illness. These studies revealed that these disorders cause a dramatic increase in RNA oxidation in brain cells without a significant increase in DNA oxidation.[14, 16, 106-111] This phenomenon can be in part attributed to the high affinity of RNA for multiply charged metal cations combined with an increased concentration of redox-active iron associated with neurodegenerative disease.[14, 112-114] It has been shown that accumulation of OG in rRNA and mRNA results in severe disruption of protein synthesis (translation stalling and production of truncated or erroneous proteins),[14, 18, 108-111] while OG in microRNA (miRNA) could change their target preference due to the ability of OG to form a stable OG-A base pair.[32, 115]

Detection of OG expelled from RNA and DNA

Expulsion of OG from DNA and RNA yields four different species, OG free base from base excision repair (BER) of DNA,[103-105] dOGMP from digestion of oligonucleotides produced during nucleotide excision repair (NER),[105, 116] OGMP from hydrolytic cleavage of RNA, and OGDP from phosphorolytic digestion of RNA (Figures

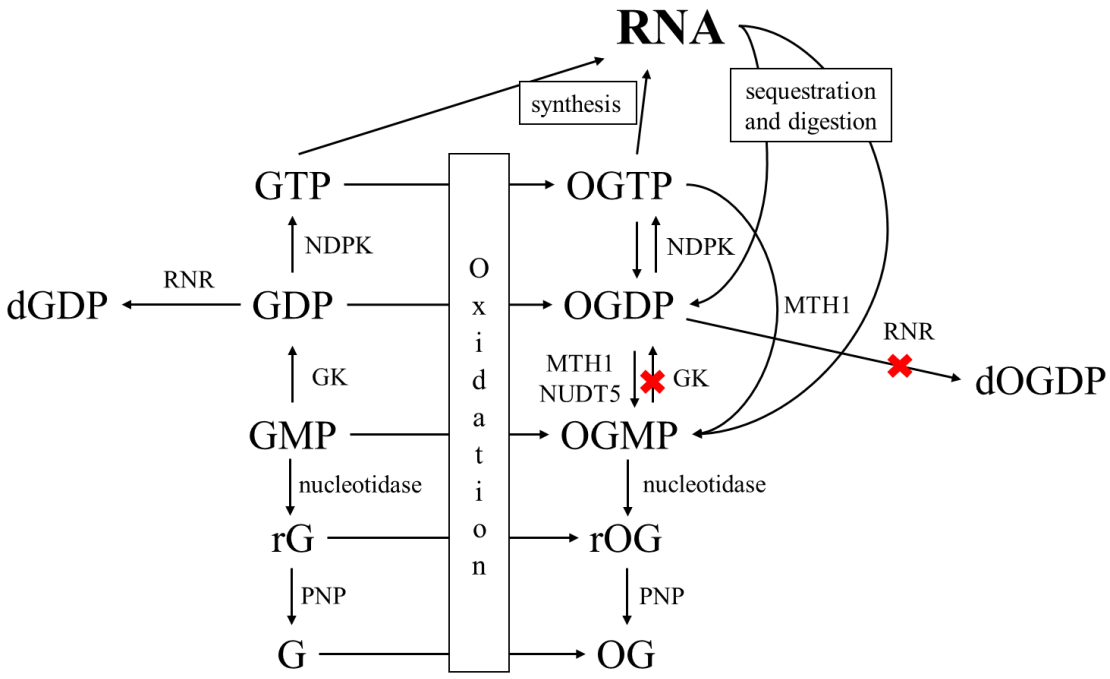


Figure 1.3. Origin and fate of OG in RNA. On the figure PNP – purine nucleoside phosphorylase, NDPK – nucleoside-diphosphate kinase.

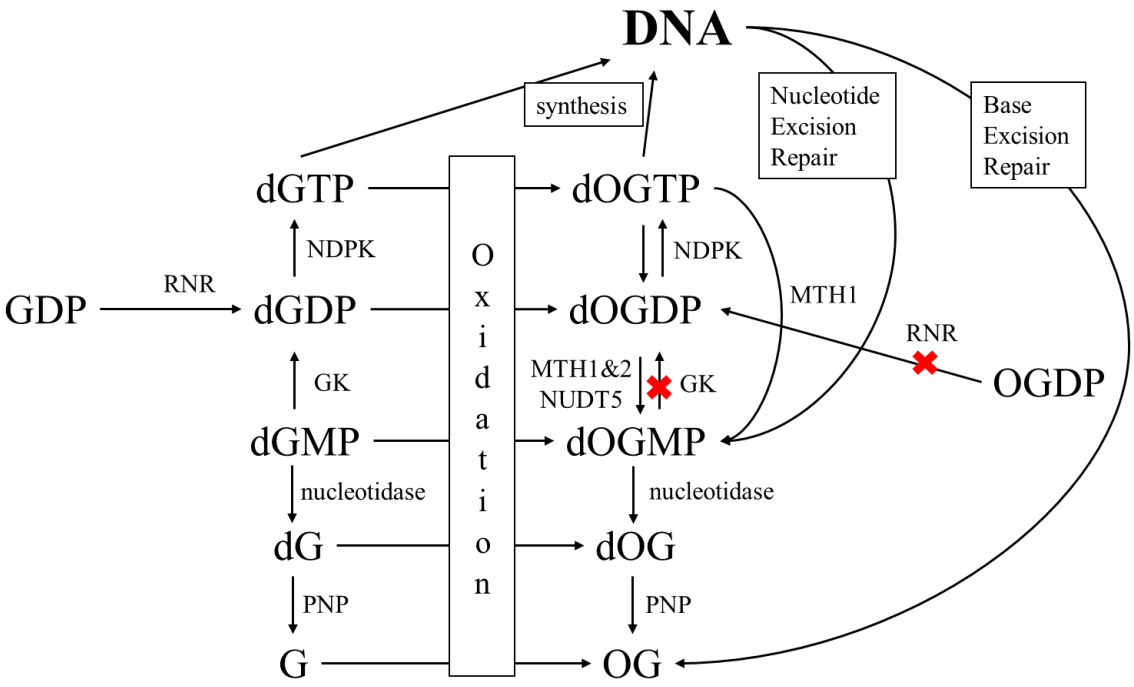


Figure 1.4. Origin and fate of OG in DNA. On the figure PNP – purine nucleoside phosphorylase, NDPK – nucleoside-diphosphate kinase.

1.3 and 1.4).[77, 117] In addition to that, OG is also produced from oxidation of guanine in the nucleotide pool.[118-119] Mono-, di-, and triphosphates containing OG are then digested to rOG and dOG that together with the OG free base can be excreted from cells.[92, 99, 101, 104-105] Currently, there is no information available on excretion of Sp or Gh from cells. OG-containing species excreted from cells have been used as a biomarker of oxidative stress for over two decades.[120-123] Typically OG, dOG, and rOG are quantified in a blood or urine sample using either chromatographic methods or enzyme-linked immunosorbent assay (ELISA) with antibodies specific for OG.[120, 122-129] Alternatively to antibodies, aptamers, DNA or RNA oligonucleotides that can specifically bind to a target molecule can be used for detection of OG or products of its oxidation. Aptamers have a number of advantages when compared to antibodies such as a longer shelf-life, lower cost, and significant structural changes induced by target binding that allow designing of complex sensors that can be used *in vivo* (Figure 1.5).[130-137] Most importantly, antibodies can only be developed for the immunogenic target molecules, which is one of the reasons why there are no known antibodies for Gh or Sp nucleobases or nucleotides. This topic is covered in more detail in Chapter 2.

Possible regulatory role of RNA oxidation

Recent studies have shown that RNA oxidation does not occur randomly and could potentially be directed toward specific RNA molecules or sites in these molecules.[108-111, 138-143] This suggests that oxidation of RNA may not only result in arrest of normal cell functions, but also change the existing regulatory role of the oxidized RNA or provide it with a new function. This could occur via two mechanisms, fragmentation of

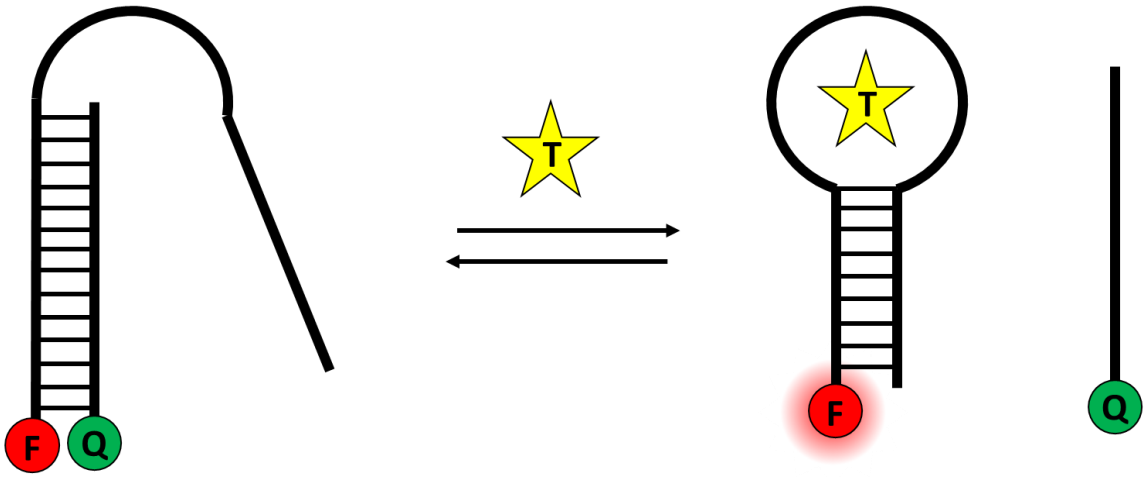


Figure 1.5. Example of a simplistic aptamer-based sensor. In the figure T = target molecule, F = fluorophore, Q = quencher.

existing mature RNA into shorter functional RNAs,[138-143] or a change in conformation or base-pairing preference of RNA upon oxidation (Figure 1.6).[32, 144] The first mechanism has been observed for selective fragmentation of tRNA[141-143] and mRNA.[139] Oxidative damage is one of the factors that ultimately induces enzymatic cleavage in the anticodon loop and division of the original tRNA into 5' and 3' halves recently named tiRNA.[138, 141] Such molecules with several smaller fragments of tRNA (tRFs) have been recently shown to play a role in proliferation of cancer cells and the cellular stress response.[138, 140-143] The second mechanism has been observed for changes in tRNA conformation upon methylation of adenine[144] and for a change in the target preference upon oxidation of guanine to OG in the seed region of miRNA.[32] The question of whether these cases are unique or there are systematic trends of RNA oxidation being directed toward specific sites and leading to subsequent gain of function still remains open. Our ability to answer it can be greatly facilitated by finding a way to map oxidation hot-spots in RNA.

Detection of modification sites in RNA and DNA

For a long time after development of Maxam-Gilbert sequencing, a method with a similar approach was used for mapping positions of modifications in RNA and DNA. In these sequencing experiments, the DNA or RNA strand is either cleaved at the modified nucleotides showing different susceptibility to the cleavage when compared to their unmodified counterparts or used as a template for the synthesis of the complement strand, producing truncated products because of polymerase stalling at the modification sites.[20, 145-147] Radiolabeled products of these reactions can typically be analyzed on polyacrylamide gels to identify the position of the modification based on the increased or

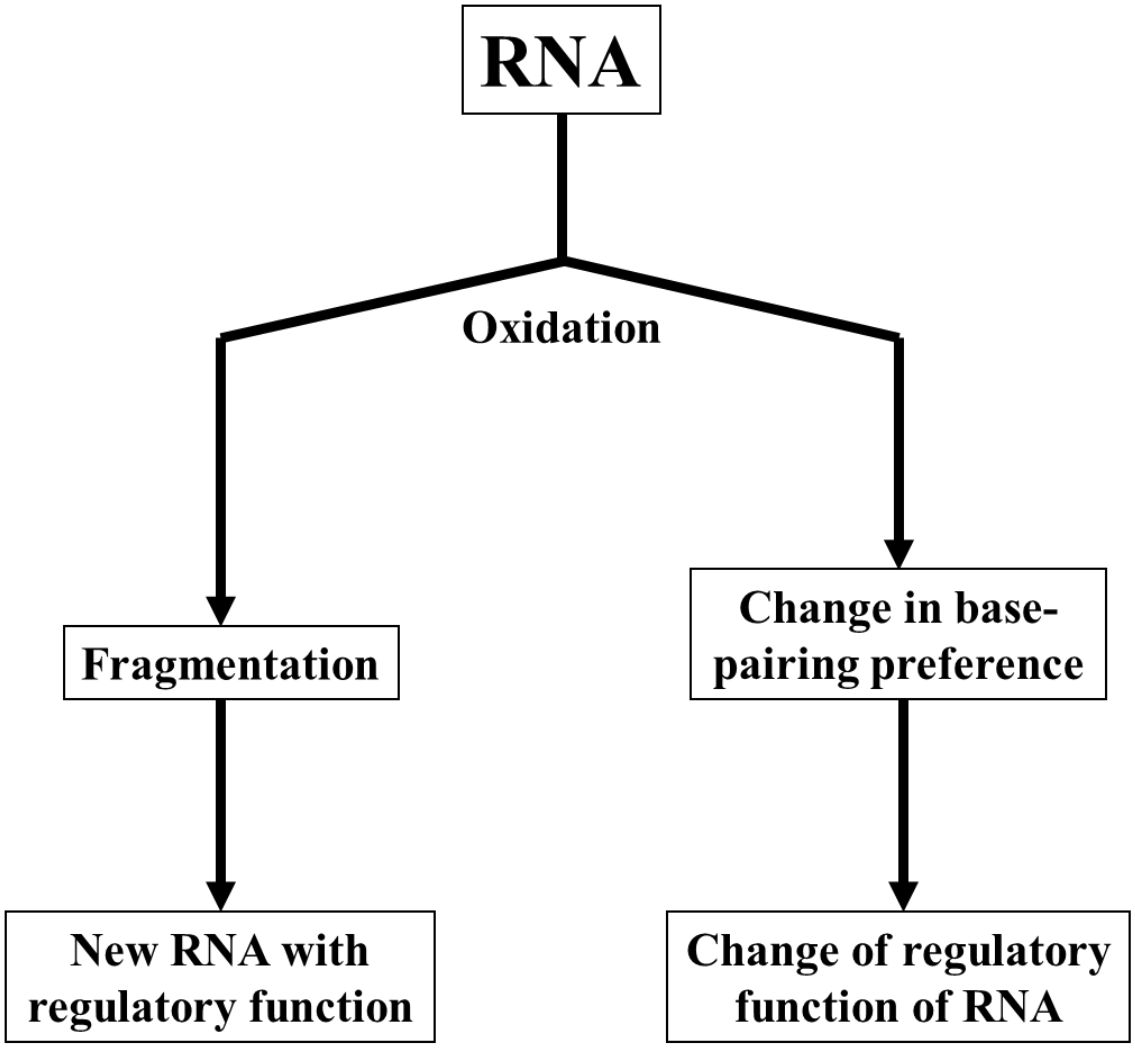


Figure 1.6. Two possible ways in which oxidation of RNA can lead to a change or gain of regulatory function.

decreased intensity of a given band. However, these sequencing methods are very labor-intensive, low-throughput, and are incapable of detection of infrequent modifications. Development of high-throughput or next-generation sequencing (NGS) methods has revolutionized the field of genomics by dramatically increasing the amount of information extracted from each sequencing experiment.[148] Since NGS became available as a research tool almost 20 years ago, multiple protocols have been developed for mapping epigenetic marks in DNA and post-transcriptional modifications in RNA (Figure 1.7).[149-154] This drastically improved our understanding of how structure and function of RNA and DNA are fine-tuned by these modifications and led to a dramatic expansion of the fields of epigenetics and epitranscriptomics.[155-156]

Common steps required for detection of modification positions in RNA or DNA include fragmentation, enrichment, library preparation, and sequencing (Figure 1.8). The goal of fragmentation is to induce random strand breaks producing DNA or RNA fragments that are short enough to be sequenced. Typically, this is done by sonication in the case of DNA and incomplete hydrolysis in the case of RNA.[157-159] The size of the fragments used for sequencing depends on the method. Currently the most common method for mapping modifications is Illumina sequencing that requires fragments shorter than 100-150 nt; another viable alternative is single-molecule real-time (SMRT) sequencing developed by Pacific Biosciences that can work on fragments longer than 20,000 nt.[160-162] Multiple RNA species that are shorter than the threshold for the sequencing method used, such as ~22 nt-long miRNA or ~76 nt-long tRNA, or even 5,000 nt 28S rRNA in case of SMRT sequencing, do not have to be cleaved. The next step after fragmentation is enrichment of the DNA or RNA strands that contain the modification of interest. Most commonly

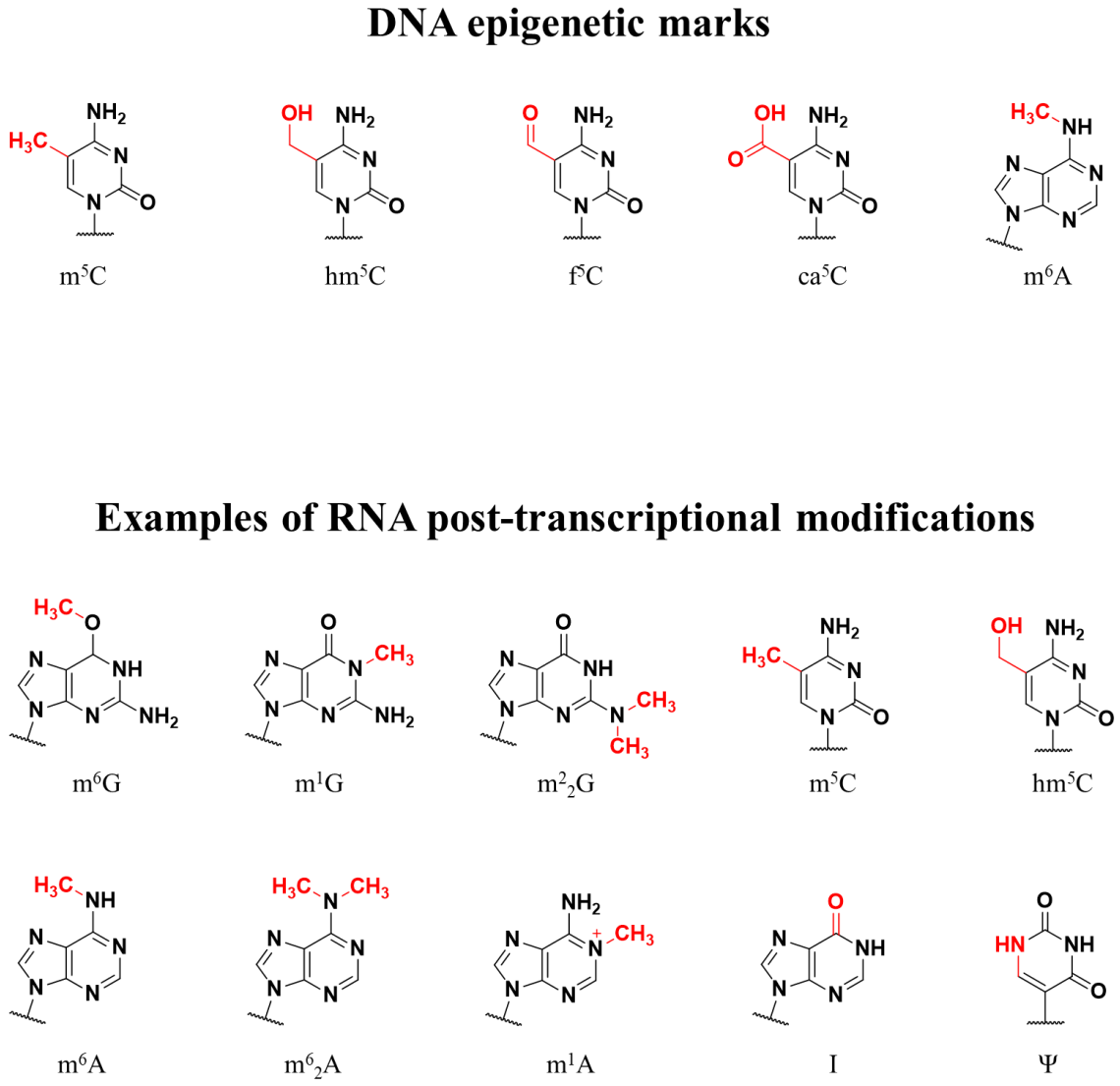


Figure 1.7. Examples of nucleotide modifications in DNA and RNA.

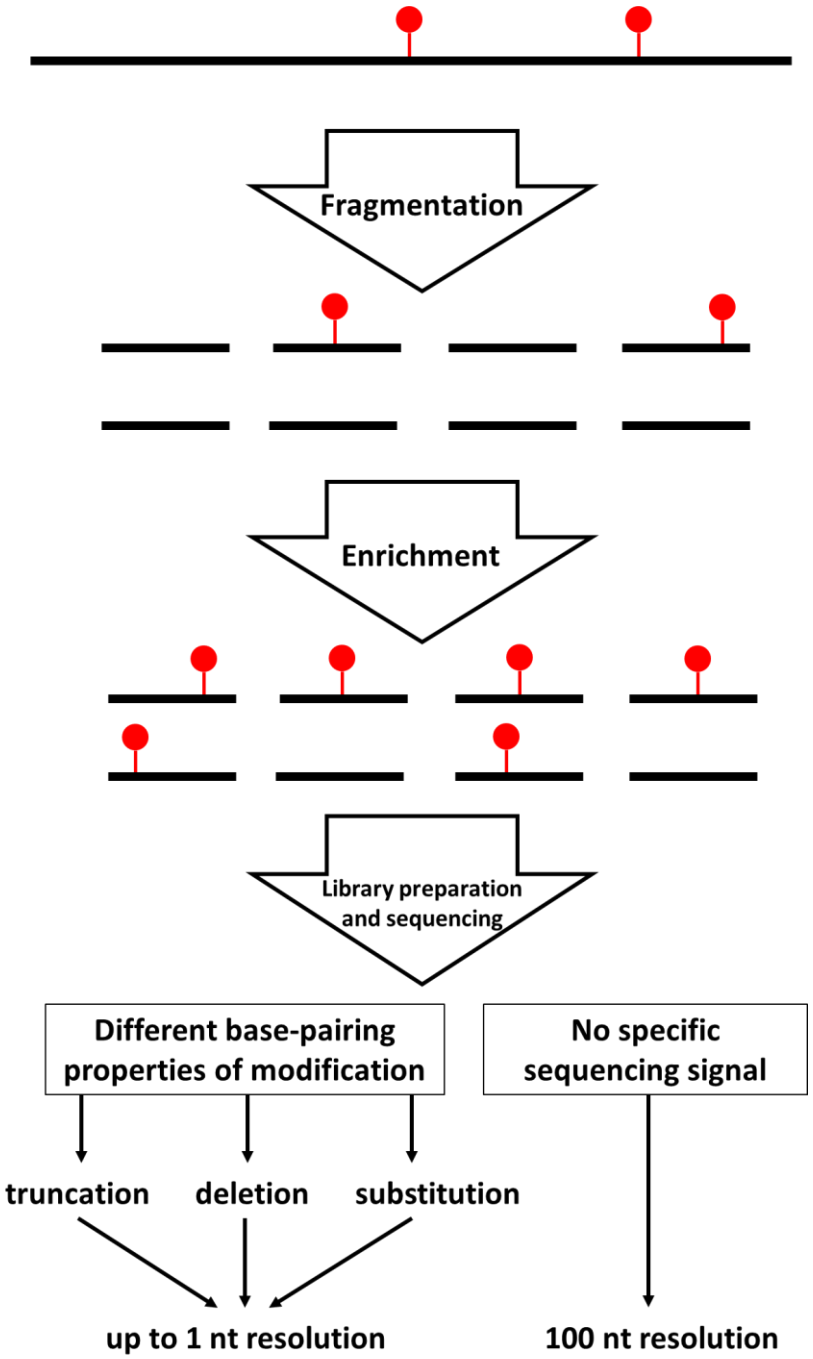


Figure 1.8. Schematic representation of generic workflow used for mapping positions of modified nucleotides in DNA or RNA.

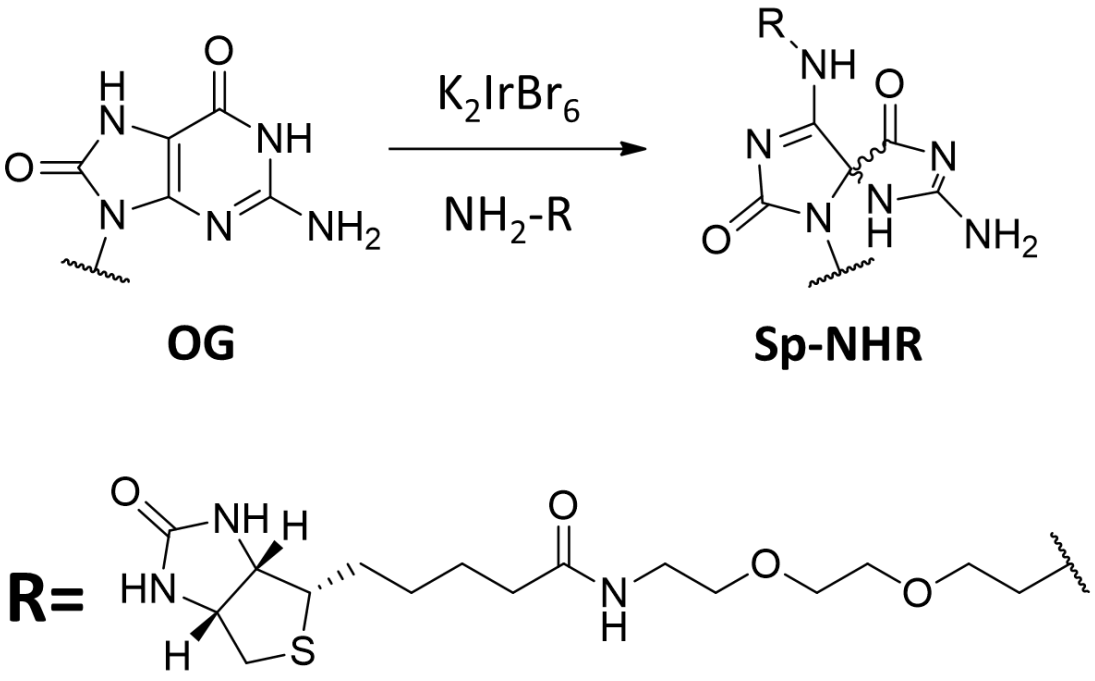
enrichment is achieved by using antibodies specific for a given modified nucleotide immobilized on solid support.[154] Just as with the fragmentation, in certain cases this step can be omitted. Enrichment is not required when mapping positions of high-frequency modifications such as m^5C or hm^5C in DNA; however, it is absolutely crucial when trying to map positions of low-abundance modifications such as OG due to the fact that the frequency of their occurrence is comparable to the error rate of polymerases used for library preparation and sequencing.[150, 154, 163-166] After the enrichment step, sequencing the dsDNA library is prepared using standard protocols that in the case of RNA sequencing includes reverse-transcription of the original RNA into complementary DNA (cDNA).[167]

Upon sequencing of the dsDNA library, there are two major ways of defining positions of modification sites. The first one is a high-resolution method requiring changes in base-pairing capabilities of the modified nucleotide compared to the unmodified strand that allows determination of the positions of post-transcriptional modifications or epigenetic marks at single-nucleotide resolution. The shift in base-pairing properties can be either native (insertion of thymine opposite m^6G or $m_2{}^2G$)[151, 168] or induced by chemical treatment (bisulfite sequencing of m^5C or cross-linking with antibodies used in the enrichment step).[154, 165-166, 169] In this approach, the presence of the modifications leads to generation of a characteristic signal when it is encountered by the polymerase. This signal could be either formation of truncated products, deletion of one or more bases, base substitution, or a combination of all signal types. Truncated products and deletions are formed when the polymerase is incapable of inserting one of the four standard dNTPs opposite the modified nucleotide that leads to either arrest of polymerization or formation of an extrahelical loop containing the modification site.[46-47, 154, 170] Most commonly these

sequencing signatures are used for mapping the positions of modified nucleotides upon formation of their bulky adducts with antibodies.[154, 169, 171-172] Products containing base substitutions result from different base-pairing preferences of modified and unmodified nucleotides leading to a specific mutation profile at the site of the modification.[150-153, 168-169, 173-174] The second approach for detecting positions of modified nucleotides is a low-resolution method that closely resembles chromatin immunoprecipitation and sequencing (ChIP-seq). This method is more convenient because it does not require a specific sequencing signal coming from a modified nucleotide. Instead the modification sites are mapped at the resolution of fragmentation (~100 nt for Illumina sequencing) as spikes in abundance of RNA or DNA fragments coming from a specific locus.[154, 175-177] Alternatively, SMRT or nanopore sequencing can be used for mapping positions of nucleotide modifications in unamplified DNA (SMRT and nanopore) or RNA (only nanopore).[161-162, 178] These methods can be used to detect modification sites in unprocessed DNA or RNA due to the fact that they can recognize signals other than insertion of a specific base by polymerase during sequencing (polymerase kinetics in case of SMRT sequencing and nucleotide size in case of nanopore sequencing). However, due to relatively high error rates of both these methods, they can only be used for locating modifications that occur with high frequency.

Mapping the position of oxidation sites and OG in DNA has been accomplished using multiple methods such as site-specific piperidine cleavage followed by PAGE analysis of reaction products,[42] Sanger sequencing as a single-nucleotide deletion product upon excision of OG with the repair enzyme and ligation of the gap in DNA[179-180] or as unnatural hydrophobic base inserted instead of OG after repair,[179, 181] and, recently, low-

resolution NGS (OG-seq) using selective conversion of OG to a biotin-adduct of Sp (Scheme 1.2). [175, 179] Prior to the work shown here mapping of OG in RNA has only been attempted with the gel-based method analogous to piperidine cleavage of DNA.[42] This topic is covered in more detail in Chapter 3 and Chapter 4. One of the goals of the research presented in this dissertation is to facilitate development of a method that can be used for sequencing OG in RNA.



Scheme 1.2. Selective conversion of OG into biotin-adduct of Sp used for enrichment of OG in the sequencing library in OG-Seq method.

References

1. De Bont, R.; van Larebeke, N., Endogenous DNA damage in humans: a review of quantitative data. *Mutagenesis* **2004,** *19* (3), 169-185.

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

3. Wogan, G. N.; Hecht, S. S.; Felton, J. S.; Conney, A. H.; Loeb, L. A., Environmental and chemical carcinogenesis. *Semin. Cancer Biol.* **2004,** *14* (6), 473-486.

4. Loft, S.; Poulsen, H. E., Cancer risk and oxidative DNA damage in man. *J. Mol. Med.* **1996,** *74* (6), 297-312.

5. Noll, D. M.; Mason, T. M.; Miller, P. S., Formation and repair of interstrand cross-links in DNA. *Chem. Rev.* **2006,** *106* (2), 277-301.

6. Johansen, M. E.; Muller, J. G.; Xu, X.; Burrows, C. J., Oxidatively induced DNA-protein cross-linking between single-stranded binding protein and oligodeoxynucleotides containing 8-oxo-7,8-dihydro-2'-deoxyguanosine. *Biochemistry* **2005,** *44* (15), 5660-5671.

7. Oleinick, N. L.; Chiu, S. M.; Ramakrishnan, N.; Xue, L. Y., The formation, identification, and significance of DNA-protein cross-links in mammalian cells. *Br. J. Cancer Suppl.* **1987,** *8*, 135-140.

8. Milowska, K.; Gabryelak, T., Reactive oxygen species and DNA damage after ultrasound exposure. *Biomol. Eng.* **2007,** *24* (2), 263-267.

9. Li, Z.; Wu, J.; Deleo, C. J., RNA damage and surveillance under oxidative stress. *IUBMB Life* **2006,** *58* (10), 581-588.

10. Chatterjee, N.; Walker, G. C., Mechanisms of DNA damage, repair, and mutagenesis. *Environ. Mol. Mutag.* **2017,** *58* (5), 235-263.

11. Wurtmann, E. J.; Wolin, S. L., RNA under attack: cellular handling of RNA damage. *Crit. Rev. Biochem. Mol. Biol.* **2009,** *44* (1), 34-49.

12. Finkel, T.; Holbrook, N. J., Oxidants, oxidative stress and the biology of ageing. *Nature* **2000,** *408* (6809), 239-247.

13. Dunlop, R. A.; Brunk, U. T.; Rodgers, K. J., Oxidized proteins: mechanisms of removal and consequences of accumulation. *IUBMB Life* **2009,** *61* (5), 522-527.

14. Honda, K.; Smith, M. A.; Zhu, X.; Baus, D.; Merrick, W. C.; Tartakoff, A. M.; Hattier, T.; Harris, P. L.; Siedlak, S. L.; Fujioka, H.; Liu, Q.; Moreira, P. I.; Miller, F. P.; Nunomura, A.; Shimohama, S.; Perry, G., Ribosomal RNA in Alzheimer

disease is oxidized by bound redox-active iron. *J. Biol. Chem.* **2005,** *280* (22), 20978-20986.

15. Cooke, M. S.; Evans, M. D.; Dizdaroglu, M.; Lunec, J., Oxidative DNA damage: mechanisms, mutation, and disease. *FASEB J.* **2003,** *17* (10), 1195-1214.

16. Kong, Q.; Lin, C. L., Oxidative damage to RNA: mechanisms, consequences, and diseases. *Cell. Mol. Life Sci.* **2010,** *67* (11), 1817-1829.

17. Ames, B. N.; Shigenaga, M. K.; Hagen, T. M., Oxidants, antioxidants, and the degenerative diseases of aging. *Proc. Natl. Acad. Sci. U. S. A.* **1993,** *90* (17), 7915-7922.

18. Stirpe, M.; Palermo, V.; Ferrari, M.; Mroczek, S.; Kufel, J.; Falcone, C.; Mazzoni, C., Increased levels of RNA oxidation enhance the reversion frequency in aging pro-apoptotic yeast mutants. *Apoptosis* **2017,** *22* (2), 200-206.

19. Steenken, S.; Jovanovic, S. V.; Bietti, M.; Bernhard, K., The trap depth (in DNA) of 8-oxo-7,8-dihydro-2'deoxyguanosine as derived from electron-transfer equilibria in aqueous solution. *J. Am. Chem. Soc.* **2000,** *122* (10), 2373-2374.

20. Saito, I.; Nakamura, T.; Nakatani, K.; Yoshioka, Y.; Yamaguchi, K.; Sugiyama, H., Mapping of the hot spots for DNA damage by one-electron oxidation: efficacy of GG doublets and GGG triplets as a trap in long-range hole migration. *J. Am. Chem. Soc.* **1998,** *120* (48), 12686-12687.

21. Morikawa, M.; Kino, K.; Oyoshi, T.; Suzuki, M.; Kobayashi, T.; Miyazawa, H., Analysis of guanine oxidation products in double-stranded DNA and proposed guanine oxidation pathways in single-stranded, double-stranded or quadruplex DNA. *Biomolecules* **2014,** *4* (1), 140-159.

22. Ghosh, A. K.; Schuster, G. B., Role of the guanine N1 imino proton in the migration and reaction of radical cations in DNA oligomers. *J. Am. Chem. Soc.* **2006,** *128* (13), 4172-4173.

23. Xu, X. Y.; Muller, J. G.; Ye, Y.; Burrows, C. J., DNA-protein cross-links between guanine and lysine depend on the mechanism of oxidation for formation of C5 vs C8 guanosine adducts. *J. Am. Chem. Soc.* **2008,** *130* (2), 703-709.

24. Yun, B. H.; Geacintov, N. E.; Shafirovich, V., Generation of guanine-thymidine cross-links in DNA by peroxynitrite/carbon dioxide. *Chem. Res. Toxicol.* **2011,** *24* (7), 1144-1152.

25. Alshykhly, O. R.; Fleming, A. M.; Burrows, C. J., 5-Carboxamido-5-formamido-2-iminohydantoin, in addition to 8-oxo-7,8-dihydroguanine, Is the major product of the iron-Fenton or X-ray radiation-induced oxidation of guanine under aerobic reducing conditions in nucleoside and DNA contexts. *J. Org. Chem.* **2015,** *80* (14), 6996-7007.

26. Thapa, B.; Munk, B. H.; Burrows, C. J.; Schlegel, H. B., Computational study of the radical mediated mechanism of the formation of C8, C5, and C4 guanine:lysine adducts in the presence of the benzophenone photosensitizer. *Chem. Res. Toxicol.* **2016,** *29* (9), 1396-1409.

27. Fleming, A. M.; Kannan, A.; Muller, J. G.; Liao, Y.; Burrows, C. J., Copper/H_2O_2-mediated oxidation of 2 '-deoxyguanosine in the presence of 2-naphthol leads to the formation of two distinct isomeric adducts. *J. Org. Chem.* **2011,** *76* (19), 7953-7963.

28. Neeley, W. L.; Essigmann, J. M., Mechanisms of formation, genotoxicity, and mutation of guanine oxidation products. *Chem. Res. Toxicol.* **2006,** *19* (4), 491-505.

29. Cheng, K. C.; Cahill, D. S.; Kasai, H.; Nishimura, S.; Loeb, L. A., 8-Hydroxyguanine, an abundant form of oxidative DNA damage, causes G----T and A----C substitutions. *J. Biol. Chem.* **1992,** *267* (1), 166-172.

30. Lowe, L. G.; Guengerich, F. P., Steady-state and pre-steady-state kinetic analysis of dNTP insertion opposite 8-oxo-7,8-dihydroguanine by *Escherichia coli* polymerases I exo- and II exo. *Biochemistry* **1996,** *35* (30), 9840-9849.

31. Furge, L. L.; Guengerich, F. P., Analysis of nucleotide insertion and extension at 8-oxo-7,8-dihydroguanine by replicative T7 polymerase exo(-) and human immunodeficiency virus-1 reverse transcriptase using steady-state and pre-steady-state kinetics. *Biochemistry* **1997,** *36* (21), 6475-6487.

32. Wang, J. X.; Gao, J.; Ding, S. L.; Wang, K.; Jiao, J. Q.; Wang, Y.; Sun, T.; Zhou, L. Y.; Long, B.; Zhang, X. J.; Li, Q.; Liu, J. P.; Feng, C.; Liu, J.; Gong, Y.; Zhou, Z.; Li, P. F., Oxidative modification of miR-184 enables it to target Bcl-xL and Bcl-w. *Mol. Cell* **2015,** *59* (1), 50-61.

33. Xie, Y.; Yang, H.; Cunanan, C.; Okamoto, K.; Shibata, D.; Pan, J.; Barnes, D. E.; Lindahl, T.; McIlhatton, M.; Fishel, R.; Miller, J. H., Deficiencies in mouse Myh and Ogg1 result in tumor predisposition and G to T mutations in codon 12 of the K-ras oncogene in lung tumors. *Cancer Res.* **2004,** *64* (9), 3096-3102.

34. Henderson, P. T.; Delaney, J. C.; Muller, J. G.; Neeley, W. L.; Tannenbaum, S. R.; Burrows, C. J.; Essigmann, J. M., The hydantoin lesions formed from oxidation of 7,8-dihydro-8-oxoguanine are potent sources of replication errors *in vivo*. *Biochemistry* **2003,** *42* (31), 9257-9262.

35. Fleming, A. M.; Muller, J. G.; Ji, I.; Burrows, C. J., Characterization of 2'-deoxyguanosine oxidation products observed in the Fenton-like system Cu(II)/H_2O_2/reductant in nucleoside and oligodeoxynucleotide contexts. *Org. Biomol. Chem.* **2011,** *9* (9), 3338-3348.

36. Niles, J. C.; Wishnok, J. S.; Tannenbaum, S. R., Spiroiminodihydantoin and guanidinohydantoin are the dominant products of 8-oxoguanosine oxidation at low

fluxes of peroxynitrite: mechanistic studies with ^{18}O. *Chem. Res. Toxicol.* **2004,** *17* (11), 1510-1519.

37. Gremaud, J. N.; Martin, B. D.; Sugden, K. D., Influence of substrate complexity on the diastereoselective formation of spiroiminodihydantoin and guanidinohydantoin from chromate oxidation. *Chem. Res. Toxicol.* **2010,** *23* (2), 379-385.

38. Fleming, A. M.; Muller, J. G.; Dlouhy, A. C.; Burrows, C. J., Structural context effects in the oxidation of 8-oxo-7,8-dihydro-2'-deoxyguanosine to hydantoin products: electrostatics, base stacking, and base pairing. *J. Am. Chem. Soc.* **2012,** *134* (36), 15091-15102.

39. Fleming, A. M.; Burrows, C. J., G-quadruplex folds of the human telomere sequence alter the site reactivity and reaction pathway of guanine oxidation compared to duplex DNA. *Chem. Res. Toxicol.* **2013,** *26* (4), 593-607.

40. Fleming, A. M.; Orendt, A. M.; He, Y.; Zhu, J.; Dukor, R. K.; Burrows, C. J., Reconciliation of chemical, enzymatic, spectroscopic and computational data to assign the absolute configuration of the DNA base lesion spiroiminodihydantoin. *J. Am. Chem. Soc.* **2013,** *135* (48), 18191-18204.

41. Zhu, J.; Fleming, A. M.; Orendt, A. M.; Burrows, C. J., pH-dependent equilibrium between 5-guanidinohydantoin and iminoallantoin affects nucleotide insertion opposite the DNA lesion. *J. Org. Chem.* **2016,** *81* (2), 351-359.

42. Fleming, A. M.; Alshykhly, O.; Zhu, J.; Muller, J. G.; Burrows, C. J., Rates of chemical cleavage of DNA and RNA oligomers containing guanine oxidation products. *Chem. Res. Toxicol.* **2015,** *28* (6), 1292-1300.

43. Tomaszewska-Antczak, A.; Guga, P.; Nawrot, B.; Pratviel, G., Guanosine in a single stranded region of anticodon stem-loop tRNA models is prone to oxidatively generated damage resulting in dehydroguanidinohydantoin and spiroiminodihydantoin lesions. *Chem. Eur. J.* **2015,** *21* (17), 6381-6385.

44. Hailer, M. K.; Slade, P. G.; Martin, B. D.; Sugden, K. D., Nei deficient *Escherichia coli* are sensitive to chromate and accumulate the oxidized guanine lesion spiroiminodihydantoin. *Chem. Res. Toxicol.* **2005,** *18* (9), 1378-1383.

45. Mangerich, A.; Knutson, C. G.; Parry, N. M.; Muthupalani, S.; Ye, W.; Prestwich, E.; Cui, L.; McFaline, J. L.; Mobley, M.; Ge, Z.; Taghizadeh, K.; Wishnok, J. S.; Wogan, G. N.; Fox, J. G.; Tannenbaum, S. R.; Dedon, P. C., Infection-induced colitis in mice causes dynamic and tissue-specific changes in stress response and DNA damage leading to colon cancer. *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109* (27), E1820-E1829.

46. Kornyushyna, O.; Berges, A. M.; Muller, J. G.; Burrows, C. J., *In vitro* nucleotide misinsertion opposite the oxidized guanosine lesions spiroiminodihydantoin and

guanidinohydantoin and DNA synthesis past the lesions using *Escherichia coli* DNA polymerase I (Klenow fragment). *Biochemistry* **2002,** *41* (51), 15304-15314.

47. Kornyushyna, O.; Burrows, C. J., Effect of the oxidized guanosine lesions spiroiminodihydantoin and guanidinohydantoin on proofreading by *Escherichia coli* DNA polymerase I (Klenow fragment) in different sequence contexts. *Biochemistry* **2003,** *42* (44), 13008-13018.

48. Jia, L.; Shafirovich, V.; Shapiro, R.; Geacintov, N. E.; Broyde, S., Structural and thermodynamic features of spiroiminodihydantoin damaged DNA duplexes. *Biochemistry* **2005,** *44* (40), 13342-13353.

49. Zhao, X.; Muller, J. G.; Halasyam, M.; David, S. S.; Burrows, C. J., *In vitro* ligation of oligodeoxynucleotides containing C8-oxidized purine lesions using bacteriophage T4 DNA ligase. *Biochemistry* **2007,** *46* (12), 3734-3744.

50. Moore, P. B.; Steitz, T. A., The roles of RNA in the synthesis of protein. *Cold Spring Harb. Perspect. Biol.* **2011,** *3* (11), a003780.

51. Leick, V., Ratios between contents of DNA RNA and protein in different micro-organisms as a function of maximal growth rate. *Nature* **1968,** *217* (5134), 1153-1155.

52. Traganos, F.; Darzynkiewicz, Z.; Melamed, M. R., The ratio of RNA to total nucleic-acid content as a quantitative measure of un-balanced cell-growth. *Cytometry* **1982,** *2* (4), 212-218.

53. Glen, A. C., Measurement of DNA and RNA in human peripheral blood lymphocytes. *Clin. Chem.* **1967,** *13* (4), 299-313.

54. Chomczynski, P.; Wilfinger, W. W.; Eghbalnia, H. R.; Kennedy, A.; Rymaszewski, M.; Mackey, K., Inter-individual differences in RNA levels in human peripheral blood. *PLoS One* **2016,** *11* (2), e0148260.

55. Lajtha, A.; Rappoport, D. A.; Fritz, R. R.; Myers, J. L., *Handbook of Neurochemistry*. Plenum Press: New York, NY, 1969; Vol. 1, p 101-115.

56. Olmsted, P. S.; Villee, C. A., Nucleic acid composition of human liver cell fractions. *J. Biol. Chem.* **1955,** *212* (1), 179-186.

57. Kravetz de Srulijes, L.; Israeli, E.; Barzilai, D., RNA:DNA ratios in human tumors. *J. Natl. Cancer Inst.* **1977,** *58* (3), 769-770.

58. Hofer, T.; Badouard, C.; Bajak, E.; Ravanat, J. L.; Mattsson, A.; Cotgreave, I. A., Hydrogen peroxide causes greater oxidation in cellular RNA than in DNA. *Biol. Chem.* **2005,** *386* (4), 333-337.

59. Wu, J.; Li, Z., Human polynucleotide phosphorylase reduces oxidative RNA damage and protects HeLa cell against oxidative stress. *Biochem. Biophys. Res. Commun.* **2008,** *372* (2), 288-292.

60. Elledge, S. J., Cell cycle checkpoints: preventing an identity crisis. *Science* **1996,** *274* (5293), 1664-1672.

61. Zhou, B. B.; Elledge, S. J., The DNA damage response: putting checkpoints in perspective. *Nature* **2000,** *408* (6811), 433-439.

62. Houseley, J.; Tollervey, D., The many pathways of RNA degradation. *Cell* **2009,** *136* (4), 763-776.

63. Dong, H.; Nilsson, L.; Kurland, C. G., Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. *J. Mol. Biol.* **1996,** *260* (5), 649-663.

64. Gorlich, D.; Mattaj, I. W., Nucleocytoplasmic transport. *Science* **1996,** *271* (5255), 1513-1518.

65. Wagatsuma, A.; Sadamoto, H.; Kitahashi, T.; Lukowiak, K.; Urano, A.; Ito, E., Determination of the exact copy numbers of particular mRNAs in a single cell by quantitative real-time RT-PCR. *J. Exp. Biol.* **2005,** *208* (12), 2389-2398.

66. Schwanhausser, B.; Busse, D.; Li, N.; Dittmar, G.; Schuchhardt, J.; Wolf, J.; Chen, W.; Selbach, M., Global quantification of mammalian gene expression control. *Nature* **2011,** *473* (7347), 337-342.

67. Bernstein, J. A.; Khodursky, A. B.; Lin, P. H.; Lin-Chao, S.; Cohen, S. N., Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proc. Natl. Acad. Sci. U. S. A.* **2002,** *99* (15), 9697-9702.

68. Pedersen, S.; Reeh, S., Functional mRNA half lives in *E. coli*. *Mol. Genet. Genomics* **1978,** *166* (3), 329-336.

69. Selinger, D. W.; Saxena, R. M.; Cheung, K. J.; Church, G. M.; Rosenow, C., Global RNA half-life analysis in *Escherichia coli* reveals positional patterns of transcript degradation. *Genome Res.* **2003,** *13* (2), 216-223.

70. Defoiche, J.; Zhang, Y.; Lagneaux, L.; Pettengell, R.; Hegedus, A.; Willems, L.; Macallan, D. C., Measurement of ribosomal RNA turnover *in vivo* by use of deuterium-labeled glucose. *Clin. Chem.* **2009,** *55* (10), 1824-1833.

71. Miller, B. G., The biological half-lives of ribosomal and transfer RNA in the mouse uterus. *J. Endocrinol.* **1973,** *59* (1), 81-85.

72. Engelke, D. R.; Hopper, A. K., Modified view of tRNA: stability amid sequence diversity. *Mol. Cell* **2006,** *21* (2), 144-145.

73. Bellacosa, A.; Moss, E. G., RNA repair: damage control. *Curr. Biol.* **2003,** *13* (12), R482-R484.

74. Aravind, L.; Koonin, E. V., The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. *Genome Biol.* **2001,** *2* (3), research0007.

75. Aas, P. A.; Otterlei, M.; Falnes, P. O.; Vagbo, C. B.; Skorpen, F.; Akbari, M.; Sundheim, O.; Bjoras, M.; Slupphaug, G.; Seeberg, E.; Krokan, H. E., Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. *Nature* **2003,** *421* (6925), 859-863.

76. Hayakawa, H.; Kuwano, M.; Sekiguchi, M., Specific binding of 8-oxoguanine-containing RNA to polynucleotide phosphorylase protein. *Biochemistry* **2001,** *40* (33), 9977-9982.

77. Hayakawa, H.; Sekiguchi, M., Human polynucleotide phosphorylase protein in response to oxidative stress. *Biochemistry* **2006,** *45* (21), 6749-6755.

78. Wu, J.; Jiang, Z.; Liu, M.; Gong, X.; Wu, S.; Burns, C. M.; Li, Z., Polynucleotide phosphorylase protects *Escherichia coli* against oxidative stress. *Biochemistry* **2009,** *48* (9), 2012-2020.

79. Hayakawa, H.; Uchiumi, T.; Fukuda, T.; Ashizuka, M.; Kohno, K.; Kuwano, M.; Sekiguchi, M., Binding capacity of human YB-1 protein for RNA containing 8-oxoguanine. *Biochemistry* **2002,** *41* (42), 12739-12744.

80. Delaney, J. C.; Essigmann, J. M., Mutagenesis, genotoxicity, and repair of 1-methyladenine, 3-alkylcytosines, 1-methylguanine, and 3-methylthymine in AlkB *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **2004,** *101* (39), 14051-14056.

81. Yehudai-Resheff, S.; Hirsh, M.; Schuster, G., Polynucleotide phosphorylase functions as both an exonuclease and a poly(A) polymerase in spinach chloroplasts. *Mol. Cell. Biol.* **2001,** *21* (16), 5408-5416.

82. Carpousis, A. J., The *Escherichia coli* RNA degradosome: structure, function and relationship in other ribonucleolytic multienzyme complexes. *Biochem. Soc. Trans.* **2002,** *30* (2), 150-155.

83. Schilders, G.; van Dijk, E.; Raijmakers, R.; Pruijn, G. J., Cell and molecular biology of the exosome: how to make or break an RNA. *Int. Rev. Cytol.* **2006,** *251*, 159-208.

84. Hayakawa, H.; Fujikane, A.; Ito, R.; Matsumoto, M.; Nakayama, K. I.; Sekiguchi, M., Human proteins that specifically bind to 8-oxoguanine-containing RNA and

their responses to oxidative stress. *Biochem. Biophys. Res. Commun.* **2010,** *403* (2), 220-224.

85. Garneau, N. L.; Wilusz, J.; Wilusz, C. J., The highways and byways of mRNA decay. *Nat. Rev. Mol. Cell Biol.* **2007,** *8* (2), 113-126.

86. Mo, J. Y.; Maki, H.; Sekiguchi, M., Hydrolytic elimination of a mutagenic nucleotide, 8-oxodGTP, by human 18-kilodalton protein: sanitization of nucleotide pool. *Proc. Natl. Acad. Sci. U. S. A.* **1992,** *89* (22), 11021-11025.

87. Sakumi, K.; Furuichi, M.; Tsuzuki, T.; Kakuma, T.; Kawabata, S.; Maki, H.; Sekiguchi, M., Cloning and expression of cDNA for a human enzyme that hydrolyzes 8-oxo-dGTP, a mutagenic substrate for DNA-synthesis. *J. Biol. Chem.* **1993,** *268* (31), 23524-23530.

88. Cai, J. P.; Ishibashi, T.; Takagi, Y.; Hayakawa, H.; Sekiguchi, M., Mouse MTH2 protein which prevents mutations caused by 8-oxoguanine nucleotides. *Biochem. Biophys. Res. Commun.* **2003,** *305* (4), 1073-1077.

89. Ishibashi, T.; Hayakawa, H.; Sekiguchi, M., A novel mechanism for preventing mutations caused by oxidation of guanine nucleotides. *EMBO Rep.* **2003,** *4* (5), 479-483.

90. Petrocchi, A.; Leo, E.; Reyna, N. J.; Hamilton, M. M.; Shi, X.; Parker, C. A.; Mseeh, F.; Bardenhagen, J. P.; Leonard, P.; Cross, J. B.; Huang, S.; Jiang, Y.; Cardozo, M.; Draetta, G.; Marszalek, J. R.; Toniatti, C.; Jones, P.; Lewis, R. T., Identification of potent and selective MTH1 inhibitors. *Bioorg. Med. Chem. Lett.* **2016,** *26* (6), 1503-1507.

91. Speina, E.; Arczewska, K. D.; Gackowski, D.; Zielinska, M.; Siomek, A.; Kowalewski, J.; Olinski, R.; Tudek, B.; Kusmierek, J. T., Contribution of hMTH1 to the maintenance of 8-oxoguanine levels in lung DNA of non-small-cell lung cancer patients. *J. Natl. Cancer Inst.* **2005,** *97* (5), 384-395.

92. Hayakawa, H.; Hofer, A.; Thelander, L.; Kitajima, S.; Cai, Y.; Oshiro, S.; Yakushiji, H.; Nakabeppu, Y.; Kuwano, M.; Sekiguchi, M., Metabolic fate of oxidized guanine ribonucleotides in mammalian cells. *Biochemistry* **1999,** *38* (12), 3610-3614.

93. Fujikawa, K.; Kamiya, H.; Yakushiji, H.; Nakabeppu, Y.; Kasai, H., Human MTH1 protein hydrolyzes the oxidized ribonucleotide, 2-hydroxy-ATP. *Nucleic Acids Res.* **2001,** *29* (2), 449-454.

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

95. Ferraro, P.; Franzolin, E.; Pontarin, G.; Reichard, P.; Bianchi, V., Quantitation of cellular deoxynucleoside triphosphates. *Nucleic Acids Res.* **2010,** *38* (6), e85.

96. Kennedy, E. M.; Gavegnano, C.; Nguyen, L.; Slater, R.; Lucas, A.; Fromentin, E.; Schinazi, R. F.; Kim, B., Ribonucleoside triphosphates as substrate of human immunodeficiency virus type 1 reverse transcriptase in human macrophages. *J. Biol. Chem.* **2010,** *285* (50), 39380-39391.

97. Hayakawa, H.; Taketomi, A.; Sakumi, K.; Kuwano, M.; Sekiguchi, M., Generation and elimination of 8-oxo-7,8-dihydro-2'-deoxyguanosine 5'-triphosphate, a mutagenic substrate for DNA synthesis, in human cells. *Biochemistry* **1995,** *34* (1), 89-95.

98. Zhu, J.; Shang, Y.; Xia, C.; Wang, W.; Wen, W.; Zhang, M., Guanylate kinase domains of the MAGUK family scaffold proteins as specific phospho-protein-binding modules. *EMBO J.* **2011,** *30* (24), 4986-4997.

99. Sekiguchi, T.; Ito, R.; Hayakawa, H.; Sekiguchi, M., Elimination and utilization of oxidized guanine nucleotides in the synthesis of RNA and its precursors. *J. Biol. Chem.* **2013,** *288* (12), 8128-8135.

100. Brosh, S.; Sperling, O.; Dantziger, E.; Sidi, Y., Metabolism of guanine and guanine nucleotides in primary rat neuronal cultures. *J. Neurochem.* **1992,** *58* (4), 1485-1490.

101. Obtulowicz, T.; Swoboda, M.; Speina, E.; Gackowski, D.; Rozalski, R.; Siomek, A.; Janik, J.; Janowska, B.; Ciesla, J. M.; Jawien, A.; Banaszkiewicz, Z.; Guz, J.; Dziaman, T.; Szpila, A.; Olinski, R.; Tudek, B., Oxidative stress and 8-oxoguanine repair are enhanced in colon adenoma and carcinoma patients. *Mutagenesis* **2010,** *25* (5), 463-471.

102. Mundt, J. M.; Hah, S. S.; Sumbad, R. A.; Schramm, V.; Henderson, P. T., Incorporation of extracellular 8-oxodG into DNA and RNA requires purine nucleoside phosphorylase in MCF-7 cells. *Nucleic Acids Res.* **2008,** *36* (1), 228-236.

103. de Souza-Pinto, N. C.; Eide, L.; Hogue, B. A.; Thybo, T.; Stevnsner, T.; Seeberg, E.; Klungland, A.; Bohr, V. A., Repair of 8-oxodeoxyguanosine lesions in mitochondrial dna depends on the oxoguanine DNA glycosylase (OGG1) gene and 8-oxoguanine accumulates in the mitochondrial DNA of OGG1-defective mice. *Cancer Res.* **2001,** *61* (14), 5378-5381.

104. Alseth, I.; Korvald, H.; Osman, F.; Seeberg, E.; Bjoras, M., A general role of the DNA glycosylase Nth1 in the abasic sites cleavage step of base excision repair in *Schizosaccharomyces pombe*. *Nucleic Acids Res.* **2004,** *32* (17), 5119-5125.

105. Hazra, T. K.; Das, A.; Das, S.; Choudhury, S.; Kow, Y. W.; Roy, R., Oxidative DNA damage repair in mammalian cells: a new perspective. *DNA Repair* **2007,** *6* (4), 470-480.

106. Che, Y.; Wang, J. F.; Shao, L.; Young, T., Oxidative damage to RNA but not DNA in the hippocampus of patients with major mental illness. *J. Psychiatry Neurosci.* **2010,** *35* (5), 296-302.

107. Poulsen, H. E.; Specht, E.; Broedbaek, K.; Henriksen, T.; Ellervik, C.; Mandrup-Poulsen, T.; Tonnesen, M.; Nielsen, P. E.; Andersen, H. U.; Weimann, A., RNA modifications by oxidation: a novel disease mechanism? *Free Radic. Biol. Med.* **2012,** *52* (8), 1353-1361.

108. Shan, X.; Tashiro, H.; Lin, C. L., The identification and characterization of oxidized RNAs in Alzheimer's disease. *J. Neurosci.* **2003,** *23* (12), 4913-4921.

109. Shan, X.; Lin, C. L., Quantification of oxidized RNAs in Alzheimer's disease. *Neurobiol. Aging* **2006,** *27* (5), 657-662.

110. Shan, X.; Chang, Y.; Lin, C. L., Messenger RNA oxidation is an early event preceding cell death and causes reduced protein expression. *FASEB J.* **2007,** *21* (11), 2753-2764.

111. Chang, Y.; Kong, Q.; Shan, X.; Tian, G.; Ilieva, H.; Cleveland, D. W.; Rothstein, J. D.; Borchelt, D. R.; Wong, P. C.; Lin, C. L., Messenger RNA oxidation occurs early in disease pathogenesis and promotes motor neuron degeneration in ALS. *PLoS One* **2008,** *3* (8), e2849.

112. Ke, Y.; Ming Qian, Z., Iron misregulation in the brain: a primary cause of neurodegenerative disorders. *Lancet Neurol.* **2003,** *2* (4), 246-253.

113. Honda, K.; Casadesus, G.; Petersen, R. B.; Perry, G.; Smith, M. A., Oxidative stress and redox-active iron in Alzheimer's disease. *Ann. N.Y. Acad. Sci.* **2004,** *1012*, 179-182.

114. Lipfert, J.; Doniach, S.; Das, R.; Herschlag, D., Understanding nucleic acid-ion interactions. *Annu. Rev. Biochem.* **2014,** *83*, 813-841.

115. Guan, L.; Luo, Y.; Ja, W. W.; Disney, M. D., Small molecule alteration of RNA sequence in cells and animals. *Bioorg. Med. Chem. Lett.* **2017**.

116. Fuss, J. O.; Cooper, P. K., DNA repair: dynamic defenders against cancer and aging. *PLoS Biol.* **2006,** *4* (6), e203.

117. Deutscher, M. P.; Reuven, N. B., Enzymatic basis for hydrolytic versus phosphorolytic mRNA degradation in *Escherichia coli* and *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U. S. A.* **1991,** *88* (8), 3277-3280.

118. Foti, J. J.; Devadoss, B.; Winkler, J. A.; Collins, J. J.; Walker, G. C., Oxidation of the guanine nucleotide pool underlies cell death by bactericidal antibiotics. *Science* **2012,** *336* (6079), 315-319.

119. Nakabeppu, Y., Cellular levels of 8-oxoguanine in either DNA or the nucleotide pool play pivotal roles in carcinogenesis and survival of cancer cells. *Int. J. Mol. Sci.* **2014,** *15* (7), 12543-12557.

120. Shigenaga, M. K.; Gimeno, C. J.; Ames, B. N., Urinary 8-hydroxy-2'-deoxyguanosine as a biological marker of *in vivo* oxidative DNA damage. *Proc. Natl. Acad. Sci. U. S. A.* **1989,** *86* (24), 9697-9701.

121. Park, E. M.; Shigenaga, M. K.; Degan, P.; Korn, T. S.; Kitzler, J. W.; Wehr, C. M.; Kolachana, P.; Ames, B. N., Assay of excised oxidative DNA lesions - isolation of 8-oxoguanine and its nucleoside derivatives from biological-fluids with a monoclonal-antibody column. *Proc. Natl. Acad. Sci. U. S. A.* **1992,** *89* (8), 3375-3379.

122. Farinati, F.; Cardin, R.; Degan, P.; Rugge, M.; Di Mario, F.; Bonvicini, P.; Naccarato, R., Oxidative DNA damage accumulation in gastric carcinogenesis. *Gut* **1998,** *42* (3), 351-356.

123. Collins, A. R.; Brown, J.; Bogdanov, M.; Cadet, J.; Cooke, M.; Douki, T.; Dunster, C.; Eakins, J.; Epe, B.; Evans, M.; Farmer, P.; Gedik, C. M.; Halliwell, B.; Herbert, K.; Hofer, T.; Hutchinson, R.; Jenner, A.; Jones, G. D. D.; Kasai, H.; Kelly, F.; Lloret, A.; Loft, S.; Lunec, J.; McEwan, M.; Moller, L.; Olinski, R.; Podmore, I.; Poulsen, H.; Ravanat, J. L.; Rees, J. F.; Reetz, F.; Shertzer, H.; Spiegelhalder, B.; Turesky, R.; Tyrrell, R.; Vina, J.; Vinicombe, D.; Weimann, A.; de Wergifosse, B.; Wood, S. G.; Escodd, Comparison of different methods of measuring 8-oxoguanine as a marker of oxidative DNA damage. *Free Radic. Res.* **2000,** *32* (4), 333-341.

124. Bruskov, V. I.; Gaziev, A. I.; Mantsygin, Y. A.; Malakhova, L. V.; Morenkov, O. S., Monoclonal antibodies to 8-oxo-2'-deoxyguanosine (8-hydroxyguanosine): characterization and use in determining damage to DNA by reactive oxygen species. *Biokhimiya* **1996,** *61* (4), 535-540.

125. Toyokuni, S.; Tanaka, T.; Hatton, Y.; Nishiyama, Y.; Yoshida, A.; Uchida, K.; Hiai, H.; Ochi, H.; Osawa, T., Quantitative immunohistochemical determination of 8-hydroxy-2'-deoxyguanosine by a monoclonal antibody N45.1: its application to ferric nitrilotriacetate-induced renal carcinogenesis model. *Lab. Invest.* **1997,** *76* (3), 365-374.

126. Shimoi, K.; Kasai, H.; Yokota, N.; Toyokuni, S.; Kinae, N., Comparison between high-performance liquid chromatography and enzyme-linked immunosorbent assay for the determination of 8-hydroxy-2'-deoxyguanosine in human urine. *Cancer Epidemiol. Biomarkers Prev.* **2002,** *11* (8), 767-770.

127. Degan, P.; Shigenaga, M. K.; Park, E. M.; Alperin, P. E.; Ames, B. N., Immunoaffinity isolation of urinary 8-hydroxy-2'-deoxyguanosine and 8-hydroxyguanine and quantitation of 8-hydroxy-2'-deoxyguanosine in DNA by polyclonal antibodies. *Carcinogenesis* **1991,** *12* (5), 865-871.

128. Sliwinska, A.; Kwiatkowski, D.; Czarny, P.; Toma, M.; Wigner, P.; Drzewoski, J.; Fabianowska-Majewska, K.; Szemraj, J.; Maes, M.; Galecki, P.; Sliwinski, T., The levels of 7,8-dihydrodeoxyguanosine (8-oxoG) and 8-oxoguanine DNA glycosylase 1 (OGG1) - a potential diagnostic biomarkers of Alzheimer's disease. *J. Neurol. Sci.* **2016,** *368*, 155-159.

129. Santella, R. M., Immunological methods for detection of carcinogen-DNA damage in humans. *Cancer Epidemiol. Biomarkers Prev.* **1999,** *8* (9), 733-739.

130. Navani, N. K.; Li, Y., Nucleic acid aptamers and enzymes as sensors. *Curr. Opin. Chem. Biol.* **2006,** *10* (3), 272-281.

131. Porchetta, A.; Vallée-Bélisle, A.; Plaxco, K. W.; Ricci, F., Using distal-site mutations and allosteric inhibition to tune, extend, and narrow the useful dynamic range of aptamer-based sensors. *J. Am. Chem. Soc.* **2012,** *134* (51), 20601-20604.

132. Xing, H.; Hwang, K.; Li, J.; Torabi, S. F.; Lu, Y., DNA aptamer technology for personalized medicine. *Curr. Opin. Chem. Eng.* **2014,** *4*, 79-87.

133. Zheng, D.; Seferos, D. S.; Giljohann, D. A.; Patel, P. C.; Mirkin, C. A., Aptamer nano-flares for molecular detection in living cells. *Nano Lett.* **2009,** *9* (9), 3258-3261.

134. Ferguson, B. S.; Hoggarth, D. A.; Maliniak, D.; Ploense, K.; White, R. J.; Woodward, N.; Hsieh, K.; Bonham, A. J.; Eisenstein, M.; Kippin, T. E.; Plaxco, K. W.; Soh, H. T., Real-time, aptamer-based tracking of circulating therapeutic agents in living animals. *Sci. Transl. Med.* **2013,** *5* (213), 213ra165.

135. Torabi, S. F.; Lu, Y., Functional DNA nanomaterials for sensing and imaging in living cells. *Curr. Opin. Biotechnol.* **2014,** *28*, 88-95.

136. Jayasena, S. D., Aptamers: an emerging class of molecules that rival antibodies in diagnostics. *Clin. Chem.* **1999,** *45* (9), 1628-1650.

137. Nutiu, R.; Mei, S.; Liu, Z.; Li, Y., Engineering DNA aptamers and DNA enzymes with fluorescence-signaling properties. *Pure Appl. Chem.* **2004,** *76* (7-8), 1547-1561.

138. Saikia, M.; Jobava, R.; Parisien, M.; Putnam, A.; Krokowski, D.; Gao, X. H.; Guan, B. J.; Yuan, Y.; Jankowsky, E.; Feng, Z.; Hu, G. F.; Pusztai-Carey, M.; Gorla, M.; Sepuri, N. B.; Pan, T.; Hatzoglou, M., Angiogenin-cleaved tRNA halves interact with cytochrome c, protecting cells from apoptosis during osmotic stress. *Mol. Cell. Biol.* **2014,** *34* (13), 2450-2463.

139. Bazin, J.; Langlade, N.; Vincourt, P.; Arribat, S.; Balzergue, S.; El-Maarouf-Bouteau, H.; Bailly, C., Targeted mRNA oxidation regulates sunflower seed dormancy alleviation during dry after-ripening. *Plant Cell* **2011,** *23* (6), 2196-2208.

140. Mleczko, A. M.; Celichowski, P.; Bakowska-Zywicka, K., Ex-translational function of tRNAs and their fragments in cancer. *Acta Biochim. Pol.* **2014,** *61* (2), 211-216.

141. Nawrot, B.; Sochacka, E.; Duchler, M., tRNA structural and functional changes induced by oxidative stress. *Cell. Mol. Life Sci.* **2011,** *68* (24), 4023-4032.

142. Thompson, D. M.; Lu, C.; Green, P. J.; Parker, R., tRNA cleavage is a conserved response to oxidative stress in eukaryotes. *RNA* **2008,** *14* (10), 2095-2103.

143. Lee, Y. S.; Shibata, Y.; Malhotra, A.; Dutta, A., A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). *Genes Dev.* **2009,** *23* (22), 2639-2649.

144. Mishima, E.; Inoue, C.; Saigusa, D.; Inoue, R.; Ito, K.; Suzuki, Y.; Jinno, D.; Tsukui, Y.; Akamatsu, Y.; Araki, M.; Araki, K.; Shimizu, R.; Shinke, H.; Suzuki, T.; Takeuchi, Y.; Shima, H.; Akiyama, Y.; Toyohara, T.; Suzuki, C.; Saiki, Y.; Tominaga, T.; Miyagi, S.; Kawagisihi, N.; Soga, T.; Ohkubo, T.; Yamamura, K.; Imai, Y.; Masuda, S.; Sabbisetti, V.; Ichimura, T.; Mount, D. B.; Bonventre, J. V.; Ito, S.; Tomioka, Y.; Itoh, K.; Abe, T., Conformational change in transfer RNA is an early indicator of acute cellular damage. *J. Am. Soc. Nephrol.* **2014**, 2316-2326.

145. Schibler, U.; Kelley, D. E.; Perry, R. P., Comparison of methylated sequences in messenger RNA and heterogeneous nuclear RNA from mouse L cells. *J. Mol. Biol.* **1977,** *115* (4), 695-714.

146. Kane, S. E.; Beemon, K., Precise localization of m^6A in Rous sarcoma virus RNA reveals clustering of methylation sites: implications for RNA processing. *Mol. Cell. Biol.* **1985,** *5* (9), 2298-2306.

147. Bakin, A.; Ofengand, J., Four newly located pseudouridylate residues in *Escherichia coli* 23S ribosomal RNA are all at the peptidyltransferase center: analysis by the application of a new sequencing technique. *Biochemistry* **1993,** *32* (37), 9754-9762.

148. Koboldt, D. C.; Steinberg, K. M.; Larson, D. E.; Wilson, R. K.; Mardis, E. R., The next-generation sequencing revolution and its impact on genomics. *Cell* **2013,** *155* (1), 27-38.

149. Brenner, S.; Johnson, M.; Bridgham, J.; Golda, G.; Lloyd, D. H.; Johnson, D.; Luo, S.; McCurdy, S.; Foy, M.; Ewan, M.; Roth, R.; George, D.; Eletr, S.; Albrecht, G.; Vermaas, E.; Williams, S. R.; Moon, K.; Burcham, T.; Pallas, M.; DuBridge, R. B.; Kirchner, J.; Fearon, K.; Mao, J.; Corcoran, K., Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat. Biotechnol.* **2000,** *18* (6), 630-634.

150. Yu, M.; Hon, G. C.; Szulwach, K. E.; Song, C. X.; Jin, P.; Ren, B.; He, C., Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine. *Nat. Protoc.* **2012,** *7* (12), 2159-2170.

151. Ryvkin, P.; Leung, Y. Y.; Silverman, I. M.; Childress, M.; Valladares, O.; Dragomir, I.; Gregory, B. D.; Wang, L. S., HAMR: high-throughput annotation of modified ribonucleotides. *RNA* **2013,** *19* (12), 1684-1692.

152. Booth, M. J.; Marsico, G.; Bachman, M.; Beraldi, D.; Balasubramanian, S., Quantitative sequencing of 5-formylcytosine in DNA at single-base resolution. *Nat. Chem.* **2014,** *6* (5), 435-440.

153. Pais, J. E.; Dai, N.; Tamanaha, E.; Vaisvila, R.; Fomenkov, A. I.; Bitinaite, J.; Sun, Z.; Guan, S.; Correa, I. R., Jr.; Noren, C. J.; Cheng, X.; Roberts, R. J.; Zheng, Y.; Saleh, L., Biochemical characterization of a *Naegleria* TET-like oxygenase and its application in single molecule sequencing of 5-methylcytosine. *Proc. Natl. Acad. Sci. U. S. A.* **2015,** *112* (14), 4316-4321.

154. Li, X.; Xiong, X.; Yi, C., Epitranscriptome sequencing technologies: decoding RNA modifications. *Nat. Methods* **2016,** *14* (1), 23-31.

155. Sarda, S.; Hannenhalli, S., Next-generation sequencing and epigenomics research: a hammer in search of nails. *Genomics Inform.* **2014,** *12* (1), 2-11.

156. Marzluff, W. B., Twenty years of RNA: reflections on post-transcriptional regulation. *RNA* **2015,** *21* (4), 687-689.

157. Poptsova, M. S.; Il'icheva, I. A.; Nechipurenko, D. Y.; Panchenko, L. A.; Khodikov, M. V.; Oparina, N. Y.; Polozov, R. V.; Nechipurenko, Y. D.; Grokhovsky, S. L., Non-random DNA fragmentation in next-generation sequencing. *Sci. Rep.* **2014,** *4*, 4532.

158. Cloonan, N.; Forrest, A. R.; Kolle, G.; Gardiner, B. B.; Faulkner, G. J.; Brown, M. K.; Taylor, D. F.; Steptoe, A. L.; Wani, S.; Bethel, G.; Robertson, A. J.; Perkins, A. C.; Bruce, S. J.; Lee, C. C.; Ranade, S. S.; Peckham, H. E.; Manning, J. M.; McKernan, K. J.; Grimmond, S. M., Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat. Methods* **2008,** *5* (7), 613-619.

159. Lee, C.; Harris, R. A.; Wall, J. K.; Mayfield, R. D.; Wilke, C. O., RNaseIII and T4 polynucleotide kinase sequence biases and solutions during RNA-seq library construction. *Biol. Direct* **2013,** *8*, 16.

160. Quail, M. A.; Smith, M.; Coupland, P.; Otto, T. D.; Harris, S. R.; Connor, T. R.; Bertoni, A.; Swerdlow, H. P.; Gu, Y., A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* **2012,** *13*, 341.

161. Roberts, R. J.; Carneiro, M. O.; Schatz, M. C., The advantages of SMRT sequencing. *Genome Biol.* **2013,** *14* (7), 405.

162. Nakano, K.; Shiroma, A.; Shimoji, M.; Tamotsu, H.; Ashimine, N.; Ohki, S.; Shinzato, M.; Minami, M.; Nakanishi, T.; Teruya, K.; Satou, K.; Hirano, T.,

Advantages of genome sequencing by long-read sequencer using SMRT technology in medical area. *Hum. Cell* **2017,** *30* (3), 149-161.

163. Potter, J.; Zheng, W.; Lee, J., Thermal stability and cDNA sythesis capability of SuperScript III reverse transcriptase. *Focus* **2003,** *25* (1), 19-24.

164. Potapov, V.; Ong, J. L., Examining sources of error in PCR by single-molecule sequencing. *PLoS One* **2017,** *12* (1), e0169774.

165. Schaefer, M.; Pollex, T.; Hanna, K.; Lyko, F., RNA cytosine methylation analysis by bisulfite sequencing. *Nucleic Acids Res.* **2009,** *37* (2), e12.

166. Squires, J. E.; Patel, H. R.; Nousch, M.; Sibbritt, T.; Humphreys, D. T.; Parker, B. J.; Suter, C. M.; Preiss, T., Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic Acids Res.* **2012,** *40* (11), 5023-5033.

167. Hou, Z.; Jiang, P.; Swanson, S. A.; Elwell, A. L.; Nguyen, B. K.; Bolin, J. M.; Stewart, R.; Thomson, J. A., A cost-effective RNA sequencing protocol for large-scale gene expression studies. *Sci. Rep.* **2015,** *5*, 9570.

168. Kietrys, A. M.; Velema, W. A.; Kool, E. T., Fingerprints of modified RNA bases from deep sequencing profiles. *J. Am. Chem. Soc.* **2017**.

169. Linder, B.; Grozhik, A. V.; Olarerin-George, A. O.; Meydan, C.; Mason, C. E.; Jaffrey, S. R., Single-nucleotide-resolution mapping of m^6A and m^6Am throughout the transcriptome. *Nat. Methods* **2015,** *12* (8), 767-772.

170. Tserovski, L.; Marchand, V.; Hauenschild, R.; Blanloeil-Oillo, F.; Helm, M.; Motorin, Y., High-throughput sequencing for 1-methyladenosine (m^1A) mapping in RNA. *Methods* **2016,** *107*, 110-121.

171. Gilham, P. T., Addition reaction specific for uridine and guanosine nucleotides and its application to modification of ribonuclease action. *J. Am. Chem. Soc.* **1962,** *84* (4), 687-688.

172. Suzuki, T.; Ueda, H.; Okada, S.; Sakurai, M., Transcriptome-wide identification of adenosine-to-inosine editing using the ICE-seq method. *Nat. Protoc.* **2015,** *10* (5), 715-732.

173. Marx, A.; Aschenbrenner, J.; Werner, S.; Marchand, V.; Adam, M.; Motorin, Y.; Helm, M., Engineering of a DNA polymerase for direct m^6A sequencing. *Angew. Chem. Int. Ed.* **2017**.

174. Aschenbrenner, J.; Marx, A., Direct and site-specific quantification of RNA 2'-O-methylation by PCR with an engineered DNA polymerase. *Nucleic Acids Res.* **2016,** *44* (8), 3495-3502.

175. Ding, Y.; Fleming, A. M.; Burrows, C. J., Sequencing the mouse genome for the oxidatively modified base 8-oxo-7,8-dihydroguanine by OG-Seq. *J. Am. Chem. Soc.* **2017,** *139* (7), 2569-2572.

176. Dominissini, D.; Moshitch-Moshkovitz, S.; Salmon-Divon, M.; Amariglio, N.; Rechavi, G., Transcriptome-wide mapping of N-6-methyladenosine by m^6A-seq based on immunocapturing and massively parallel sequencing. *Nat. Protoc.* **2013,** *8* (1), 176-189.

177. Pang, K. H.; Catto, J.; Wilson, S.; Cumberbatch, M.; Lesbriel, S., Identification of N^6-adenosine methylated RNAs by m^6A-seq within prostate cancer. *Int. J. Urol.* **2016,** *23*, 131-131.

178. Ayub, M.; Hardwick, S. W.; Luisi, B. F.; Bayley, H., Nanopore-based identification of individual nucleotides for direct RNA sequencing. *Nano Lett.* **2013,** *13* (12), 6144-6150.

179. Fleming, A. M.; Ding, Y.; Burrows, C. J., Sequencing DNA for the oxidatively modified base 8-oxo-7,8-dihydroguanine. *Methods Enzymol.* **2017,** *591*, 187-210.

180. Riedl, J.; Fleming, A. M.; Burrows, C. J., Sequencing of DNA lesions facilitated by site-specific excision via base excision repair DNA glycosylases yielding ligatable gaps. *J. Am. Chem. Soc.* **2016,** *138* (2), 491-494.

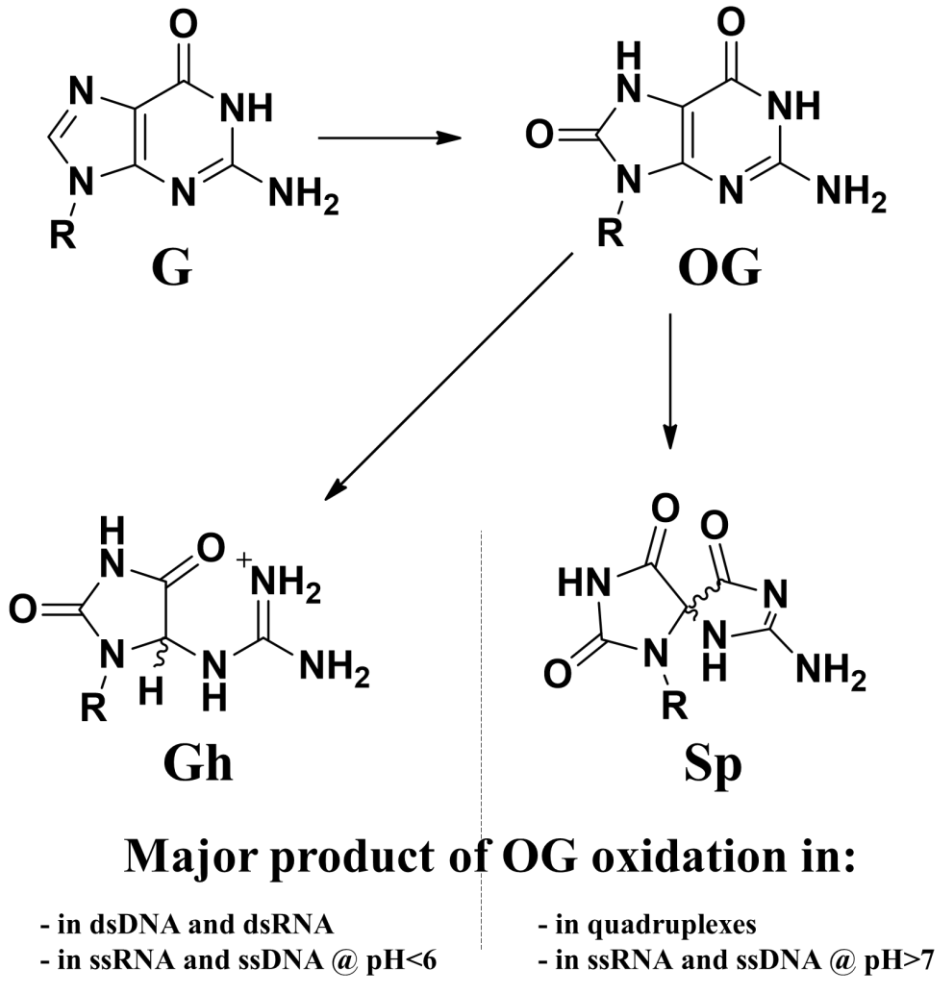
181. Riedl, J.; Ding, Y.; Fleming, A. M.; Burrows, C. J., Identification of DNA lesions using a third base pair for amplification and nanopore sequencing. *Nat. Commun.* **2015,** *6*, 8807.

CHAPTER 2

IN VITRO SELECTION OF APTAMERS FOR 5-GUANIDINOHYDANTOIN

Introduction

Exposure of DNA or RNA to reactive oxygen species (ROS) may result in modifications of DNA bases, strand breaks, or cross-links between two nucleobases or between nucleobases and other molecules present in the cell.[1-9] Such damage can significantly disrupt normal cell functions such as protein synthesis and ultimately trigger cell apoptosis or lead to irreversible mutations in the genetic code during replication due to insertion of incorrect nucleotides opposite damaged DNA bases.[10-13] One of the most abundant lesions is a two-electron oxidation product of guanine, 8-oxo-7,8-dihydroguanine (OG).[14-15] This lesion is highly susceptible to further oxidation due to its low redox potential.[14, 16] Oxidation of OG can yield two hydantoin lesions, 5-guanidinohydantoin (Gh) and spiroiminodihydantoin (Sp) in a ratio highly dependent on reaction conditions and sequence contexts (Scheme 2.1).[17-19] OG is recognized and expelled from both DNA [by 8-oxoguanine glycosylase (OGG1)][20-23] and RNA [by polynucleotide phosphorylase (PNPase)].[24-27] Both hydantoin lesions in DNA are recognized by specific DNA glycosylases, NEIL1 (Nei-like DNA glycosylase 1) and NEIL3 (Nei-like DNA glycosylase 3), while the fate of these lesions in RNA remains unknown.[17, 28-33]



Scheme 2.1. Pathway of guanine oxidation leading to Sp and Gh.

Mechanisms of repairing damaged DNA bases include two major pathways: Nucleotide Excision Repair (NER) (Figure 2.1, A) and Base Excision Repair (BER) (Figure 2.1, B).[34] NER is generally triggered by the presence of bulky adducts or cross-links that significantly perturb the helical structure of DNA.[35] This repair mechanism leads to excision of a short single-stranded DNA fragment containing the lesion that is then followed by restoration of the damaged fragment based on the undamaged strand and ligation of the nick.[35] Small lesions like OG, Gh or Sp do not induce significant changes in structure of the DNA helix, and thus are substrates for BER enzymes.[36] The BER mechanism requires a set of glycosylases (each specific for a certain type of lesion) that can recognize the damaged base and excise it from the DNA strand creating an abasic site and a free damaged base as products.[36]

Considering that nothing is known about the fate of Gh in RNA, the goal of the research presented in this chapter was the development of a system for detection and quantification of free Gh base produced in the process of BER based on DNA aptamers. The same aptamer-based sensors could then theoretically be used to detect Gh in ribo- and 2'-deoxyribonucleotide pools. Aptamers are oligonucleotides capable of selectively and tightly binding to a target molecule. Free OG, Sp and Gh can serve as biomarkers of oxidative stress and thus can be used for cancer risk assessment or early stage detection.[37-39] Currently, specific detection of damaged nucleobases is limited to antibodies developed for OG.[40-41] Application of aptamers, instead of antibodies, for bioanalytical measurements has certain advantages. Antibodies can only be produced *in vivo* and only for immunogenic and nontoxic targets, making the process quite costly, nonapplicable for some molecules of interest, and discrepancies between two batches of antibodies can be created.

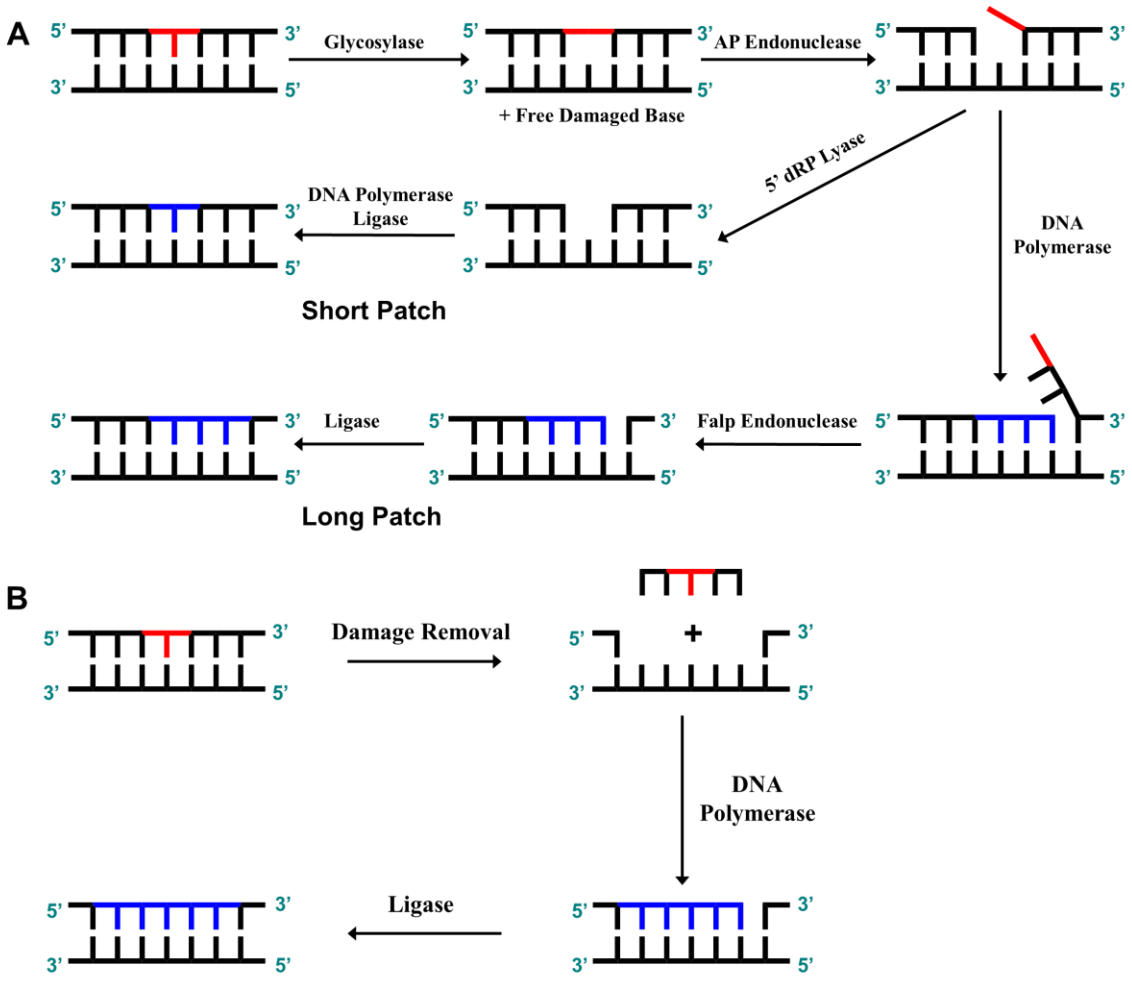


Figure 2.1. Schematic representation of mechanism of base excision repair (A) and nucleotide excision repair (B) of DNA. Damage site is marked in red, nucleotides inserted during repair marked are in blue.

Selection and production of aptamers on the other hand is done *in vitro*, which lifts the limitations and makes the process significantly less expensive.[42] In addition to that, like all other proteins, antibodies can aggregate and have limited stability as they are stable only in a narrow range of pH, temperature, and salt concentration with typical shelf lives of about one year.[42] On the other hand, DNA aptamers can endure a much wider range of environmental conditions and their storage time is practically unlimited compared to proteins.[42] Finally, unlike antibodies, aptamers undergo drastic changes in their secondary and tertiary structure upon binding to the target, which allows for building much more flexible biosensors[43-46] and provides the potential ability to use such sensors *in vivo.*[47-49]

The process of *in vitro* selection of aptamers is called Systematic Evolution of Ligands by EXponential enrichment (SELEX). Invention of this method was independently reported in 1990 by different research groups: Turek and Gold, who described selection of RNA aptamers for bacteriophage T4 DNA polymerase[50] and Ellington and Szostak with the selection of RNA aptamers for several organic dyes.[51] Twelve years later, in 2002, the Breaker lab reported discovery of the first riboswitch – a natural RNA aptamer sensor.[52-53]

Conventional procedures for *in vitro* selection of DNA aptamers include the following steps[54] (Figure 2.2, A):

1) Design the DNA library. In this case, the DNA library is a set of ssDNA with two constant primer regions (typically 20-30 nts in length) on 5' and 3' ends and a randomized region (with a typical length of 20-50 nts or 10^{12}-10^{30} possible different sequences) in between.
2) Immobilization of the target molecule on streptavidin agarose resin.

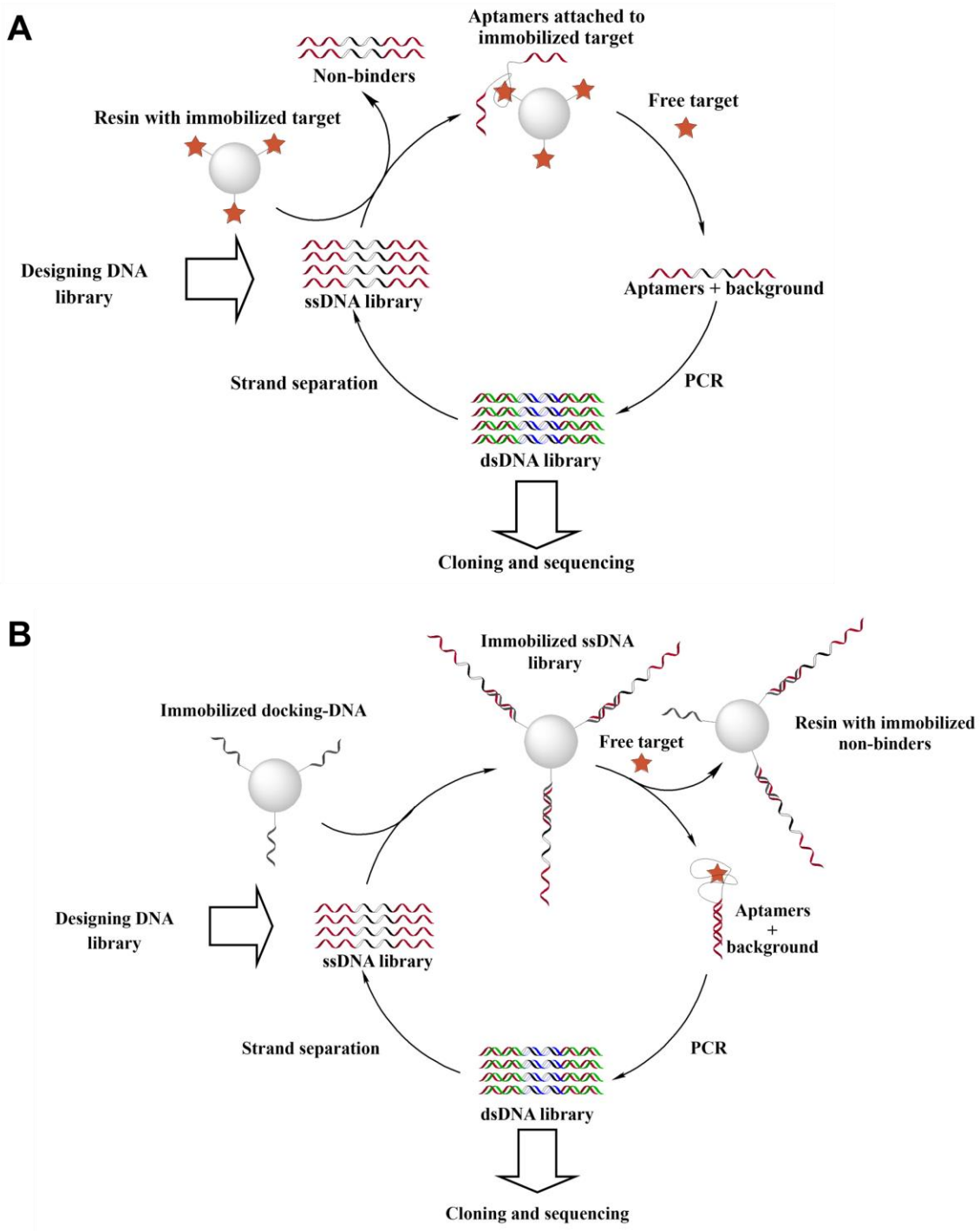


Figure 2.2. Schematic representation of steps in normal SELEX cycle **(A)** and attachment-free SELEX cycle **(B)**.

3) Incubation of the DNA library with resin containing the target followed by removal of strands that do not bind to the target.
4) Separation of DNA strands that bind to the target by rinsing the resin with solution containing the free target molecule.
5) PCR amplification of the enriched ssDNA library.
6) Transformation of the amplified dsDNA library into a ssDNA library by digestion or separation of the complementary strand.

Steps 3-6 are then repeated until an increase in affinity of the random DNA pool to the target is no longer observed. Since 1990, many new variations of the standard SELEX method have emerged, such as CE SELEX, tailored SELEX, or microfluidic SELEX.[54-55] However, most of these novel selection methodologies are either specific or preferable for development of protein-binding aptamers, while selection of aptamers for small molecules can still present a significant challenge.[56] Immobilization of the target molecule requires its modification, and in the case of small molecules it can present a nontrivial synthetic problem or significantly alter its structure and thus interfere with the target-aptamer binding pattern. Attachment-free/beacon/structure-switching SELEX (Figure 2.2, B) allows these drawbacks to be overcome by immobilizing the ssDNA random pool instead of the target. In this method, during the selection, the ssDNA library is hybridized to a biotinylated docking-DNA strand attached to the agarose resin. Upon binding to the target DNA, strands should switch their structure from a complex with docking-DNA to a stem loop with a shorter duplex region compensating for the energy difference by interactions with the target.[57] Although this approach requires more careful design of the primer regions for the DNA library and cannot be used for selection of aptamers that do not switch their

structure upon binding to the target, it has been successfully applied for selection of several aptamers and currently presents the best alternative to SELEX methods using immobilization of the target small molecules.[57-59]

Materials and methods

Synthesis of Gh free base by direct oxidation of G or OG. Oxidation of OG has been carried out by adding either 2 equivalents of Na_2IrCl_6 or 50 equivalents of $K_2S_2O_8$ to a solution containing 1 mM OG and one of the following buffer solutions: 75 mM pH 4 acetate buffer, 75 mM pH 4 citrate buffer, or 0.1% formic/acetic acid. Next, the reaction mixture containing $K_2S_2O_8$ was irradiated with UV-light (350 nm) for 1 h, and the reaction mixture containing Na_2IrCl_6 was incubated for 10 min at room temperature, and then oxidation was quenched by addition of 6-fold excess of EDTA over Na_2IrCl_6. Na_2IrCl_6 was then removed by passing the reaction mixture through Dowex 1x8 anion-exchange resin (Sigma-Aldrich, St. Louis, MO). Oxidation of G has been carried out by irradiating the solution containing 1 mM G, 200 μM methylene blue, and one of the buffer solutions listed above for 1 h with 350 nm UV-light. Methylene blue was then removed by passing the reaction mixture through a NAP column packaged with G-25 Sephadex resin (GE Healthcare, Chicago, IL). Reaction products were analyzed by HPLC using a Hypercarb (150 × 4.6 mm, 5 μm) column (Thermo Fisher Scientific, Waltham, MA). The major product formed during the reaction in all cases had significantly lower molecular weight than Gh (MH^+ m/z 102.07 instead of 158.07).

Synthesis of Gh free base from dG. Oxidation of dG was carried out by following a literature protocol.[60-61] In short, solution containing 1 mM dG, 10 μM methylene blue, and

75 mM pH 4 acetate buffer was irradiated for 1 h with 350 nm UV-light. Methylene blue was then removed by passing the reaction mixture through an NAP column packaged with G-25 Sephadex resin (GE Healthcare). Excess of unreacted dG was then removed by passing reaction mixture through Microsorb-MV 100-5 C18 (250x4.6 mm) HPLC column (Varian, Palo Alto, CA) and collecting the void volume containing dGh and dSp. dGh was then purified on Hypercarb (150 × 4.6 mm, 5 μm) HPLC column (Thermo Fisher Scientific) using isocratic elution with 0.1% acetic acid. Identity of dGh was then confirmed via electrospray ionization-mass spectrometry (ESI-MS) with MH^+ m/z of 274.10 (calc. 274.11).

Conversion of dGh to Gh free base was then accomplished by incubating 270 nmol dGh in 50 μl HF/pyridine mixture at 37°C for 30 min followed by neutralizing the reaction mixture with Na_2CO_3.[62-63] Gh free base was then purified on a Hypercarb (150 × 4.6 mm, 5 μm) HPLC column (Thermo Fisher Scientific) using isocratic elution with 0.1% acetic acid. Identity of dGh was then confirmed via electrospray ionization-mass spectrometry (ESI-MS) with MH^+ m/z of 157.96 (calc. 158.07).

Oligonucleotide synthesis. SELEX library, capture strand, and PCR primers were synthesized by the University of Utah core facilities following the standard solid-phase synthesis protocols.

PCR amplification of the initial library. For the initial PCR amplification of the SELEX library, 20 pmol of the randomized DNA template was added to the freshly made stock solution containing 1 μM forward primer, 1 μM reverse primer, 400 μM of each dNTP, 1x ThermoPol reaction buffer (20 mM Tris-HCl, 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 2mM $MgSO_4$, 0.1% Triton X-100, pH 8.8, New England Biolabs, Ipswich, MA), and 2.5

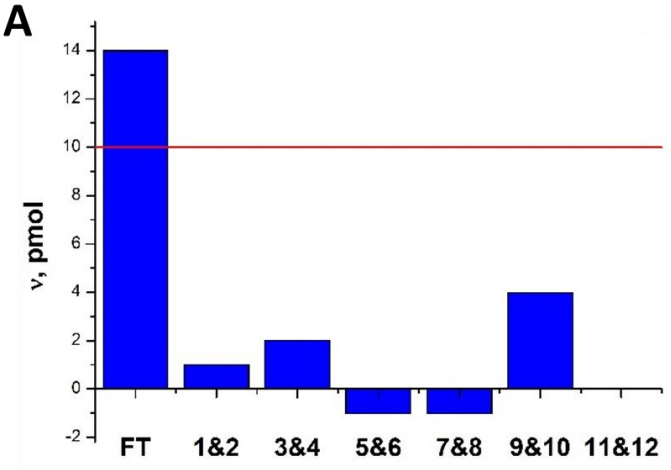
U/50 µL Taq DNA polymerase (New England Biolabs) to give the final volume of 10 mL. Reaction mixture was divided between 100 reaction tubes (100 µL per tube) and 20 cycles of PCR were carried out using the following program: 15 s denaturation at 95 °C, 30 s annealing at 60 °C, and 45 s extension at 72 °C. The PCR products were resolved on 4% agarose gel, gel bands corresponding to the amplified SELEX library were excised, and PCR amplicons were purified using Zymoclean gel DNA recovery kit (Zymo Research, Irvine, CA) following protocol suggested by the manufacturer. Double-stranded SELEX library was converted to the single-stranded library according to the protocol described in the next section. All PCR reactions have been performed using C1000 Touch thermal cycler (BioRad, Hercules, CA).

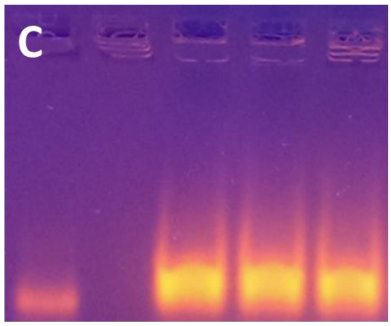
General SELEX protocol. Steps described below were followed for each cycle of SELEX.

Immobilization: 100 pmol of ssDNA pool were incubated for 1 h with 250 µL of Pierce streptavidin agarose resin (Thermo Fisher Scientific) containing 10-fold excess (1 nmol) of bound biotinylated docking-DNA. After incubation resin was consecutively washed 12 times with 250 µl of SELEX buffer (200 mM NaCl, 20 mM $MgCl_2$, and 20 mM Tris-HCl, pH 7.4). Absorbance of Cy3 was then measured for the flow through and washes (Figure 2.3, A).

Elution: Agarose resin from the previous step was incubated 3 times with 250 µL of Gh solution (100-10 µM) for 30 min each time, all eluates were combined together, and starting from round 3, 10% of the combined elute was taken for measurement of Cy3 absorbance (Figure 2.3, B). Negative selection against 250 µl of 100 µM (dA), 2'-deoxycytidine (dC), 2'-deoxyguanosine (dG), 2'-deoxythymidine (dT), uric acid, guanine,

A





B

	Round	3-7 (100 µM)	8-10 (50 µM)	11-12 (25 µM)	13-14 (10 µM)
Gh(b)	bound DNA	BDL	BDL	BDL	BDL
	pool K_D	>323 µM	>162 µM	>81 µM	>32 µM

Figure 2.3. Data collected during the aptamer selection process. Example of a chart with the amount of DNA in flow through and 12 washes combined in pairs; red line indicates detection limit of 10 pmols (A). Elution profiles obtained for 3-14 rounds of Gh SELEX (B). BDL stands for Below Detection Limit (3.3 pmol per wash). Example of preparative agarose gel (C), leftmost band is a positive PCR control; 3 bands on the right are the product of PCR amplification.

OG and arginine was performed in round 6.

PCR amplification: The combined elute was PCR amplified for 10-12 cycles using protocol described in the previous section. PCR products were analyzed and purified on a 4% agarose gel as described in the previous section (Figure 2.3, C). Negative PCR no template control (NTC) and positive control (amplification of initial DNA library) were done with each PCR mixture. All reactions have been performed using C1000 Touch thermal cycler (BioRad).

Strand separation: Biotin attached to the 5'-end of the PCR reverse primer allowed for immobilization of PCR products on streptavidin agarose resin. After PCR amplification 250 pmol of dsDNA SELEX library dissolved in strand-separation buffer (SSB: 1 M NaCl and 20 mM Tris-HCl pH 7.4) were loaded on 200 µL of the Pierce streptavidin agarose resin (Thermo Fisher Scientific) by incubating them for 15 min at room temperature. Unbound DNA was then reapplied to the column and incubated for another 15 min. The resin was then washed 5 times with 200 µL of SSB followed by two 10 min incubations at the room temperature with 200 µL of 0.2 M NaOH solution. NaOH eluates containing ssDNA SELEX library were neutralized with 3M NaOAc.

Cloning and sequencing. For separation and sequencing of individual aptamers upon completion of the selection process, TOPO TA Cloning Kit (Thermo Fisher Scientific) was used following the protocol suggested by the manufacturer. Prior to insertion into a TOPO vector supplied with the kit, dsDNA libraries from several SELEX rounds were PCR amplified with nonmodified primers and purified on an agarose gel as described in the previous section. Then to add 3' A overhangs ("sticky ends") required for plasmid insertion, 1 µL of 1 ng/µL solution of ssDNA SELEX library was combined with

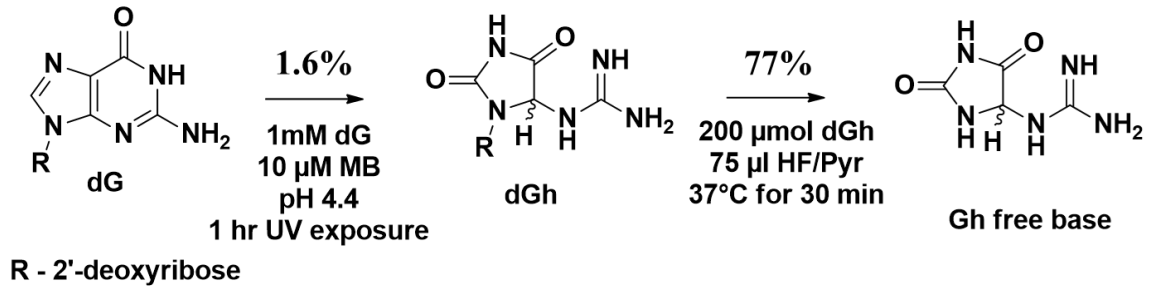
2 µL of solution containing 10 equivalents of ATP (0.21 pmol), 1 unit of Taq polymerase, and 1.7x ThermoPol PCR buffer. The reaction mixture was incubated at 72 °C for 10 min. After that 1 µL of ddH_2O, 1 µL of salt solution (1.2 M NaCl and 60 mM $MgCl_2$), and 1 µL of TOPO vector stock solution (10 ng/µL) were added to the mixture, and the final reaction mixture was incubated for 5 min at room temperature. After each DNA library was individually inserted into plasmids, they were transformed into One Shot TOP10 chemically competent *E. coli* cells via induced heat shock. Then bacteria cells were spread onto LB-agar plate containing 100 µg/ml ampicillin and 0.8 mg X-Gal. After 20-24 h of incubation at 37 °C individual cultures were replanted on another set of LB-agar plates with ampicillin, grown for 12 h at 37 °C, transferred into 4 mL of liquid LB media and incubated at 37 °C for 15 h with horizontal shaking. An aliquot of each bacterial colony was frozen at -80 °C, and the remaining cells were used for plasmid extraction with PureLink Quick Plasmid Extraction Kit (Thermo Fisher Scientific) according to the provided instructions. All plasmids were submitted for Sanger sequencing using standard M13 reverse sequencing primer.

Results and discussion

Synthesis of Gh free base. The first goal of this project was to derive a method for synthesis of the free Gh base as a target for *in vitro* selection. Synthesis of Gh nucleoside has been previously reported by our lab with use of various single-electron and singlet oxygen oxidants, such as Na_2IrCl_6, transition metal ions with peroxide, riboflavin, and methylene blue.[9, 64-66] At the same time, synthesis of Gh free base on a scale required for use in SELEX has not yet been reported.

The first synthetic approach was a direct oxidation of OG in a free base form. Synthesis was attempted with use of three different oxidants, Na_2IrCl_6, $K_2S_2O_8$ and methylene blue, in phosphate, acetate, citrate and formate buffers at pH 4-5. Reaction mixtures were then analyzed on a Hypercarb HPLC column. Only trace amounts of Gh were detected by HPLC and LC-MS with $K_2S_2O_8$ as the oxidant, while all reactions in acetate buffer with all three oxidants produced a product with absorbance peak ~204 nm and MH^+ m/z of 102, which likely corresponds to N-acetylguanidine. At the same time, oxidation of OG with methylene blue in water without any buffer (pH~6.5) produced significant quantities of Sp free base. Acquired data suggested that one of the intermediates of OG to Gh oxidation or Gh itself is unstable in the form of free base. Thus, a 2-step of synthesis starting from 2'-deoxyguanosine was chosen as an alternative (Scheme 2.2). During the first step, dG was oxidized with methylene blue to form dGh that was characterized and purified by HPLC. The second step was treatment of purified dGh with HF/pyridine, which led to base elimination and decomposition of the sugar.[62-63] The Gh free base was then purified by HPLC and characterized by UV absorbance and MS data. The final yield was estimated from UV absorbance at 240 nm based on the assumption that extinction coefficients are the same for dGh (2412 dm^3 mol^{-1} cm^{-1}) and Gh.[65]

Aptamer selection. As previously described, the attachment-free SELEX method was then used for selection of aptamers for Gh. The DNA library (Figure 2.4) similar to the one previously described by the Stojanovic lab was used for the selection process.[57] Each strand had an N_{30} random region that provides up to 10^{18} possible different sequences. The forward primer region was modified with Cy3 dye on its 5' end. Cy3 modification facilitated detection of DNA bound to the target molecule. Prior to the first round of



Scheme 2.2. Pathway used for synthesizing Gh free base.

DNA library

(Cy3) 5'-GGA GGC TCT CGG GAC GAC N_{30} GTC GTC CCG ATG CTG CAA TCG TAA-3'

PCR Forward primer

(Cy3) 5'-GGA GGC TCT CGG GAC GAC - 3'

PCR Reverse primer

3'-G CAG GGC TAC GAC GTT AGC ATT-5'(Btn)

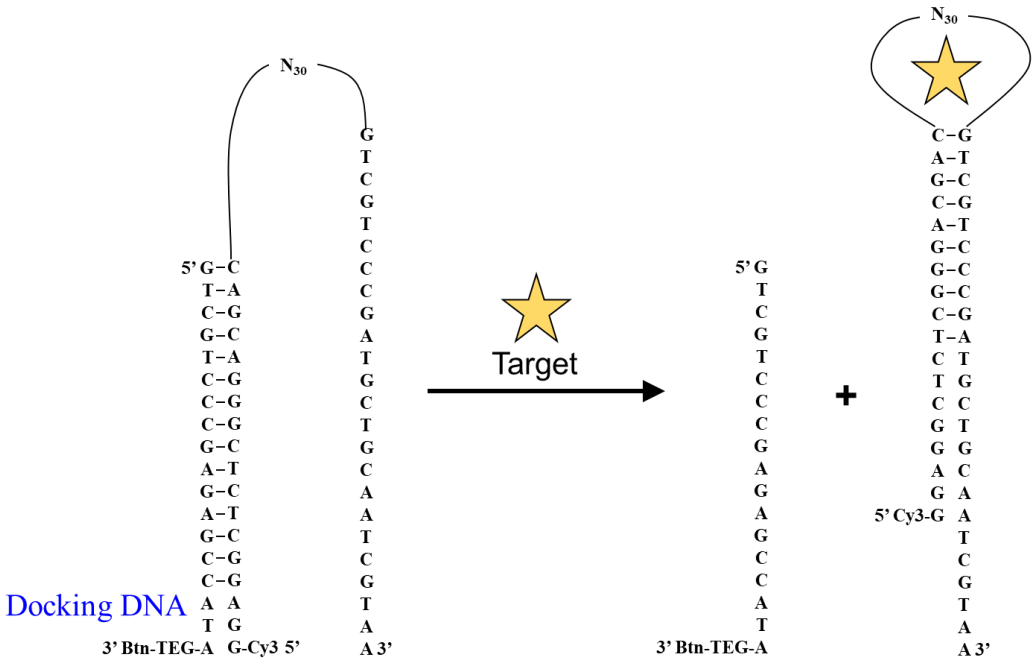


Figure 2.4. Principle of structure-switching SELEX and sequences of DNA library, primers, and docking-DNA used for the SELEX.

SELEX 100 pmols of randomized DNA pool were PCR amplified to have a stock solution of the same DNA sequences that were to be used for *in vitro* selection. A total of 14 rounds of SELEX were completed according to the protocol described in the materials and methods section of this chapter. Concentration of a target molecule (Gh) and number of PCR cycles used for amplification of the strands eluted by the target in each SELEX cycle can be found in Table 2.1.

It should be noted that in all of the 14 SELEX rounds performed for Gh, the amount of DNA pool eluted with the target solution was never above the detection limit (3 pmol). This can either be attributed to the low affinity of the selected DNA library to Gh, or to a low rate of DNA pool enrichment with aptamers having low K_D values which can be caused by a large amount of DNA strands that can weakly bind to the target containing the positively charged guanidinium group without change in aptamer conformation required for elution. To test whether this was specific for Gh, we conducted several rounds of *in vitro* aptamer selection for dG and G free base (Table 2.2). Selection of aptamers for these two targets did not encounter similar problems.

Analysis of sequencing data. The first set of sequencing data was obtained for the aptamer library after the 14^{th} round of SELEX. A total of 29 sequences of sufficient quality were retrieved from 37 submitted plasmids. Out of these 29 sequences (Table 2.3), only 13 had different N_{30} random regions with 3 sequences (C14S1, C14S2, C14S3) being represented 4 or more times. All 29 sequences demonstrated significant mutations in the antisense reverse primer (aRP) region, and only 2 sequences (34 and 36) demonstrated no changes in the original aRP sequence (marked in red, Table 2.3), while other sequences had at least one base substitution, deletion or insertion. Also, all analyzed sequences had

Table 2.1. Amount of target and number of PCR cycles used for each round of SELEX.

Cycle #	Elution conditions	# of PCR cycles
1-5	3x250 ul elutions with 100 uM Gh base	12
6	1x250 ul elution with 100 uM dA, dC, dT, dG, guanine, 8-oxoguanine, uric acid, arginine (negative selection)	10
	2x250 ul elutions with 100 uM Gh base	
7	3x250 ul elutions with 100 uM Gh base	10
8-10	3x250 ul elutions with 50 uM Gh base	10
11-12	3x250 ul elutions with 25 uM Gh base	10
13-14	3x250 ul elutions with 10 uM Gh base	10

Table 2.2. Elution profiles obtained for 3-14 rounds aptamer selection for dG and G base.

	Round	3-7(100 µM)	8 (50 µM)	9 (50 µM)	10 (25 µM)	11 (10 µM)
dG	bound DNA	BDL	10.8 pmol	12.3 pmol	BDL	5.3 pmol
	pool K_D	>323 µM	41.2 µM	35.5 µM	>81 µM	18.5 µM
G(b)	bound DNA	BDL	5.1 pmol	6.9 pmol	8.4 pmol	4.2 pmol
	pool K_D	>323 µM	92.9 µM	67.3 µM	27.1 µM	22.7 µM

Table 2.3. Aptamer sequences obtained for the DNA library after 14th round of SELEX. On the figure sequences of RCR primer regions are marked in red, sequences of aRP repeats are marked in green.

#	ID	Forward primer	N30 random region	Antisence reverse primer
13	C14S1	GGAGGCTCTCGGGACGAC	AACCAGAGAGCCGACAGATTGGAACTGAGG	GTCGTTCCGGTGCCGTAATCGTAACGTCCCGATGCTGCAATCGTAAACGGTTGTCATCGTCCCGATGCTGCAATCGTAA
14		GGAGGCTCTCGGGACGAC	AACCAGAGAGCCGACAGATTGGAACTGAGG	GTCGTTCCGGTGCTGTAATCGTAACGTCCCGATGCTGCAATCGTAAACGGTTGTCATCGTCCCGATGCTGCAATCGTAA
15		GGAGGCTCTCGGGACGAC	AACCAGAGAGCCGACAGATTGGAACTGAGG	GTCGTCTCGGTGCTGTAATCGTGACGTCCCGATGCTGCAATCGTAA
19		GGAGGCTCTCGGGACGAC	AACCAGAGAGCCGACAGATTGGAACTGAGG	GTCGTTCCGGTGCTGTAATCGTAACGTCCCGATGCTGCAATCGTAAACGATAATGCGACGTCCCGATGCTGCAATCGTAA
42		GGAGGCTCTCGGGACGAC	AACCAGAGAGCCGACAGATTGGAACTGAGG	GTCGTCTCGGTGCTGTAATCGTGACGTCCCGATGCTGCAATCGTAAGCACGATAAAGCGTCCCGATGCTGCAATCGTAA
50		GGAGGCTCTCGGGACGAC	AACCAGAGAGCCGACAGATTGGAACTGAGG	GTCGTTCCGGTGCTGTAACCGTAACGTCCCGATGCTGCAATCGTAA
12	C14S2	GGAGGCTCTCGGGACGAC	AGGCCCGGAACCCGGGGTAACCCGTAGAGC	GTCCCGATGCTACA TCGCAAACGTCCCGATGCTGCAATCGTAAACACGATAAAGCGTCCCGATGCTGCAATCGTAA
33		GGAGGCTCTCGGGACGAC	AGGCCCGGAACCCGGGGTAACCCGTAGAGC	ATCCCGATGCTACA TCGCAAACGTCCCGATGCTGCAATCGTAAAGCAACGGCCCGTCCCGATGCTGCAATCGTAA
39		G GGCTCTCGGGACGAC	AGGCCCGGAACCCGGGGTAACCCGTAGAGC	ATCCCGATGCTACA TCGTAAACGTCCCGATGCTGCAATCGTAAGGCATCGTCCCGATGCTGCAATCGTAA
53		GGAGGCTCTCGGGACGAC	AGGCCCGGAACCCGGGGTAACCCGTAGAGC	ATCCCGATGCTACA TCGTAAACGTCCCGATGCTGCAATCGTAAACGGTTGAGCGTCCCGATGCTGCAATCGTAA
3	C14S3	GGAGGCTCTCGGGACGAC	AGGCCAGCCTGCTGTGGCATAACACCGAGA	GTCATCCCGTTGCTGCGATCGTAACGTCCCGATGCTGCAATCGTAAGACGATGCTGTATCGTCCCGATGCTGCAATCGTAA
11		GGAGGCTCTCGGGACGAC	AGGCCAGCCTGCTGTGGCATAACACCGAGA	GTCATCCCGTTGCTGCGATCGTAACGTCCCGATGCTGCAATCGTAA
18		GGAGGCTCTCGGGACGAC	AGGCCAGCCTGCTGTGGCATAACACCGAGA	GTCATCCCGTTGCTGCGATCGTAACGTCCCGATGCTGCAATCGTAA
22		GGAGGCTCTCGGGACGAC	AGGCCAGCCTGCTGTGGCATAACACCGAGA	GTCATCCCGTTGCTGCGATCGTAACGTCCCGATGCTGCAATCGTAA
23		GGAGGCTCTCGGGACGAC	AGGCCAGCCTGCTGTGGCATAACACCGAGA	GTCATCCCGTTGCTGCGATCGTAACGTCCCGATGCTGCAATCGTAAGGCATCGTCCCGATGCTGCAATCGTAA
34		GGAGGCTCTCGGGACGAC	AGGCCAGCCTGCTGTGGCATAACACCGAGA	GTCGTCCCGATGCTGCAATCGTAAAAACGTACTGCGCGTCCCGATGCTGCAATCGTAAGCACGATAAAGCGTCCCGATGCTGCAATCGTAA
36		GGAGGCTCTCGGGACGAC	AGGCCAGCCTGCTGTGGCATAACACCGAGA	GTCGTCCCGATGCTGCAATCGTAAAAACGTACTGCGCGTCCCGATGCTGCAATCGTAACACGATAAAGCGTCCCGATGCTGCAATCGTAA
37		GGAGGCTCTCGGGACGAC	AGGCCAGCCTGCTGTGGCATAACACCGAGA	GTCATCCCGTTGCTGCGATCGTAACGTCCCGATGCTGCAATCGTAAACGGTTGAGCGTCCCGATGCTGCAATCGTAA
43	C14S4	GGAGGCTCTCGGGACGAC	ACCCGTAGAGACTCCTACGACTGCTTAACA	GTCGTCCTGATGCTGCGATTGTAACGTCCCGATGCTGCAATCGTAA
46		GGAGGCTCTCGGGACGAC	ACCCGTAGAGACTCCTACGACTGCTTAACA	GTCGTCCTGATGCTGCGATCGTAACGTCCCGATGCTGCAATCGTAAACGGTTGTCATCGTCCCGATGCTGCAATCGTAA
20	C14S4'	GGAGGCTCTCGGGACGAC	ACCCGACGAGACTCCTAACGGATGATGCAA	TTCGTCCCGATGCTGCAATCGTAAGCACGATAAAGCGTCCCGATGCTGCAATCGTAA
25	C14S5	GGAGGCTCTCGGGACGAC	AGAGGGCCTCCTCGATAGTGTTTAAACGTG	GTCGTCCTGATGCTGCAATCGTAAAGATAGCCGTCCCGATGCTGCAATCGTAAACGGTTGTCATCGTCCCGATGCTGCAATCGTAA
49	C14S6	GGAGGCTCTCGGGACGAC	AGAGAGGCTCCAAGTTTTCTGTCGAGGGTG	GTCGTCCC ATGCCGCAATCGTAAGGCATCGTCCCGATGCTGCAATCGTAAACGGTGCTAGTGTCGTCCCGATGCTGCAATCGTAA
24	C14S7	GGAGGCTCTCGGGACGAC	ACAACACGAGAGACTCCAGATCTACAGCA	GACGTCCCGATGCTGCAATCGTAAAGCGAGGCACATCGTCCCGATGCTGCAATCGTAAACACGATAAAGCGTCCCGATGCTGCAATCGTAA
17	C14S8	GGAGGCTCTCGGGACGAC	ATGGGCAATGAGCGCATAACAACTCCGAGA	G CGTCCTGATGCTACA TCGTAAACGTCCCGATGCTGCAATCGTAA
28	C14S9	GGAGGCTCTCGGGACGAC	GGGAACCGGCGGACCAGGCCTCCCGATCGG	GTCGTCCCGATGCAGCAGTCATAACGTCCCGATGCTGCAATCGTAAGCGGTTGTCATCGTCCCGATGCTGCAATCGTAA
38	C14S10	GGAGGCTCTCGGGACGAC	ACGAGAGACTCCCACGAATAAAGCAGAACG	GTCGTCCAGATGCTGCGCCCGTAAAGCGTCCCGATGCTGCAATCGTAA
44	C14S11	GGAGGCTCTCGGGACGAC	AATGCCCGAGTGGTTGCCTCCTACCGAATG	GTCGTCCTGATGCTGCAGTCGTAACGTCCCGATGCTGCAATCGTAA
35	C14S12	GGAGGCTCTCGGGACGAC	AGGAACACGAGAGACTCCAGAAGTGCTATC	GCATCGTCCCGATGCTGCAATCGTAAGCAGCCGTCCCGATGCTGCAATCGTAA

at least one incomplete repeat of aRP region (marked in green, Table 2.3) with 3 nucleotides missing on the 5' end and 1-14 nt insertions in between the aRP regions.

Problems with primer repeats sometimes arise for PCR amplification products in the case of poor primer design (including the possibility of substantial self-complementarity). Although it generally results in high molecular weight smears on the gel, this was not the case in our situation. Alignment of N_{30} random regions (Figure 2.5) of all 13 sequences showed substantial similarities in the first 13-20 nucleotides, while the 3' ends of all sequences were significantly more specific for all the strands and coincided with different mutation patterns in the first aRP region. This suggested that repeats of the aRP region are possibly caused by its hybridization with the end of the N_{30} region during PCR, and mutations in the first aRP sequence accommodate such folding (Figure 2.6, A). Strands with primer repeats unexpectedly emerged as a major PCR product after the 11^{th} round of SELEX (Figure 2.6, B). At that point, a slight shift upwards in the position of a band on the agarose gel was attributed to possible insertions in the random region or gel defects. Thus, SELEX rounds 12-14 were performed exclusively for an elongated DNA library. The fact that strands with primer repeats became the almost exclusive PCR product in round 11 from not being observed at all in round 10 suggests that the elongated library may have improved binding affinity towards the selection target. Such a sudden change should have been preceded by sufficient accumulation of either DNA strands with similar randomized regions or observed mutations in the aRP primer region. Another possible explanation could be more efficient PCR amplification of elongated strands, although it is highly unlikely that it can be applicable to the strands that lack full complementarity to reverse primer.

```
    C14S1      AACCAGA GAGCCGACAG ATTGGAACTG AGG
   C14S10       ACGAGA GACTCCCACG AATAAAGCAG AACG
    C14S7    ACAACACGA GAGACTCCAG ATCTACAGCA
   C14S12   AGGAACACGA GAGACTCCAG AAGTGCTATC
    C14S5          AGA GGGCCTCCTC GATAGTGTTT AAACGTG
    C14S6          AGA GAGGCTCCAA GTTTTCTGTC GAGGGTG
   C14S4'      ACCCGAC GAGACTCCTA ACGGATGATG CAA
    C14S4      ACCCGTA GAGACTCCTA CGACTGCTTA ACA
    C14S2      AGGCCCG GAACCCGGGG TAACCCGTAG AGC
    C14S3      AGGCCAG CCTGCTGTGG CATAACACCG AGA
    C14S8    ATGGGCAAT GAGCGCATAA CAACTCCGAG A
    C14S9     GGGAACCG GCGGACCAGG CCTCCCGATC GG
   C14S11    AATGCCCGA GTGGTTGCCT CCTACCGAAT G
Consensus    ..a.cc.ga Gag.ctcc.g ..t..c..t. a......
```

Figure 2.5. Alignment of the randomized region of set of sequenced aptamers from round 14 of SELEX. Nucleotides marked in blue are conserved among 50% of sequences, marked in red – among 90%.

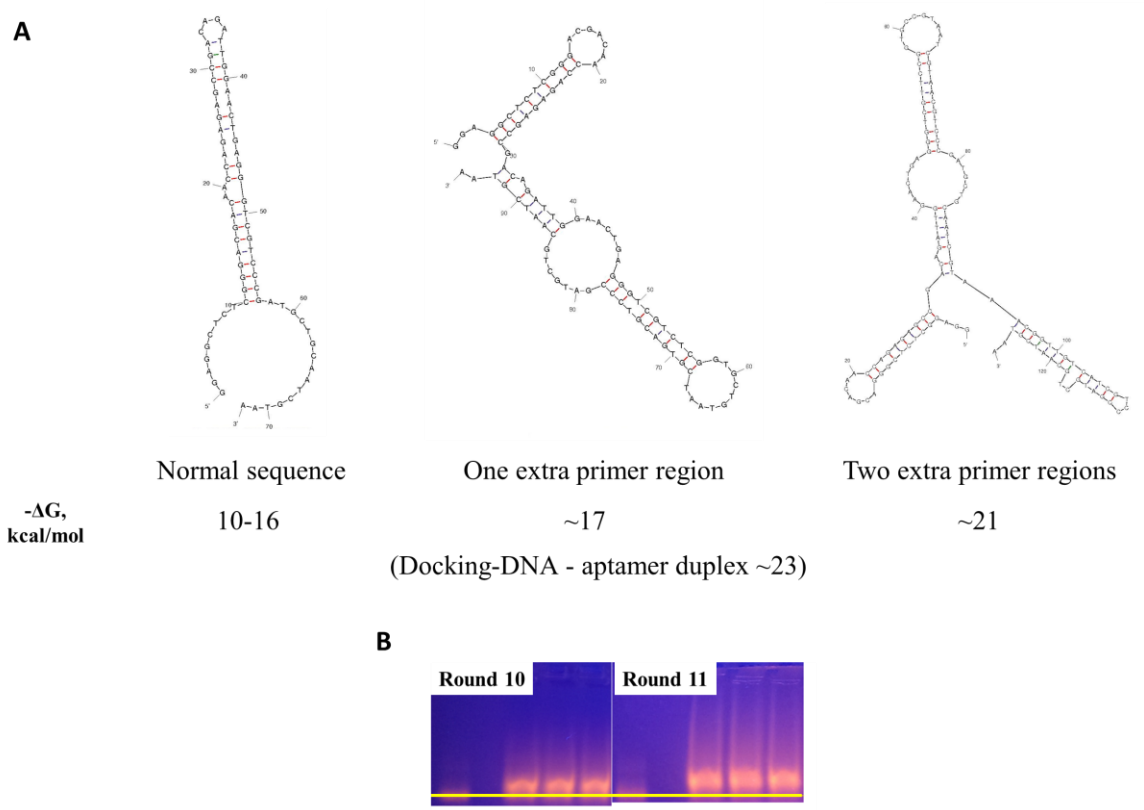


Figure 2.6. Origin and characteristics of strands with multiple primer repeats. **(A)** Structures obtained using Mfold software for C14S1 aptamer. Structure on the left – with normal primers, structure in the middle – with primers taken from sequence #15 (Figure 2.5), structure on the right – with primers taken from sequences #13 and #14. Below the structures are given –ΔG values for all aptamer sequences calculated with Mfold. **(B)** Preparative agarose gels with PCR products after SELEX rounds 10 and 11. The yellow line indicates the band position of a normal-length PCR product in a lane with the positive control.

To test for the aforementioned similarities in sequences of N_{30} region or aRP mutation patterns between elongated and normal length libraries, an additional round of SELEX was performed starting with the dsDNA library from the 10^{th} selection round. This round was marked 11'. Libraries from both rounds 10 and 11' were then sequenced (Figure 2.7), aligned and compared with the sequence set from round 14. Despite the fact that before round 11' of selection, the dsDNA pool was carefully purified, elongated strands accounted for a significant portion of PCR product after selection (~50%). As a result, even after two rounds of agarose gel purification prior to vector insertion, 3 out of 22 obtained sequences for round 11' had two aRP repeats (11-37, 11-38, 11-13). Two of these sequences had identical sequences of N_{30} region with C14S1 and C14S2, respectively, while the third had one nucleotide difference with C14S9. Also, among 31 sequences (22 from round 11' and 9 from round 10), three pairs of identical sequences were observed (10-28 and 11-50, 11-32 and 11-42, 11-1 and 11-26). When aligned, N_{30} regions from normal length sequences and sequences containing primer repeats (Figure 2.8) on most occasions did not have much in common and form separate clusters, except several partially merged together groups containing 7 sequences (Figure 2.9). Another feature that all analyzed sequences share is high enrichment in A (>36% on average) with at least one run of 3-8 As in a row being present in most sequences.

Conclusions

In this chapter we have described a way to synthesize Gh free base, performed 14 rounds of SELEX using Gh as a target molecule, and sequenced aptamer libraries after SELEX rounds 10, 11, 11', and 14. Significant similarities were observed between the

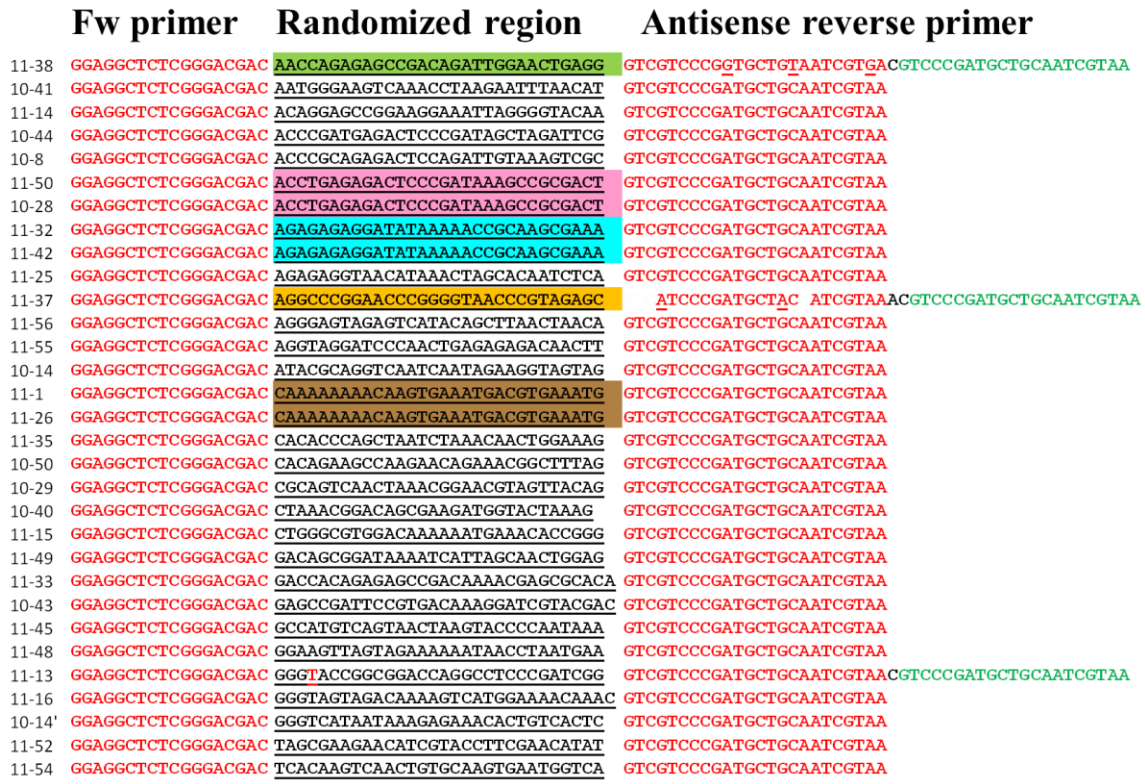


Figure 2.7. Aptamer sequences obtained for the DNA library after SELEX rounds 10 and 11'. Sequences of PCR primer regions are marked in red, sequences of aRP repeats are marked in green.

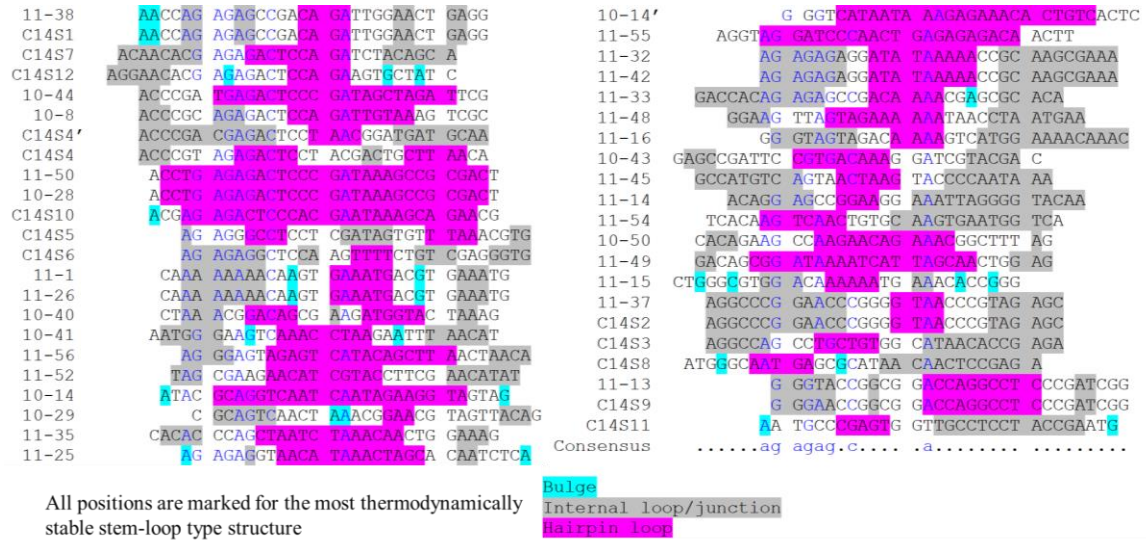


Figure 2.8. Alignment of full set of sequenced aptamers from rounds 14, 10 and 11'. Color coding is based on the structure with lowest ΔG obtained from Mfold.

```
10-44       ACCCGA TGAGACTCCC GATAGCTAGA TTCG
 10-8       ACCCGC AGAGACTCCA GATTGTAAAG TCGC
C14S4'      ACCCGA CGAGACTCCT AACGGATGAT GCAA
C14S4       ACCCGT AGAGACTCCT ACGACTGCTT AACA
11-50        ACCTG AGAGACTCCC GATAAAGCCG CGACT
10-28        ACCTG AGAGACTCCC GATAAAGCCG CGACT
```

Figure 2.9. Cluster of similar normal and elongated sequences. Nucleotides marked in blue are conserved among 50% of sequences, marked in red – among 90%.

sequenced aptamers as well as enrichment of the DNA pool with the specific strands. Both these parameters can serve as a characteristic of successful selection of aptamers.[67-68] However, due to the fact that the overall aptamer library was highly mutated and that average library affinity for the selection target (Gh) remained low, it is not likely that the sequenced aptamers retained the structure-switching properties characteristic for the initial DNA library. Thus, measurement of affinity of the sequenced DNA strands for Gh would be more challenging and costly than initially expected. In parallel with development of aptamers for Gh, the same SELEX protocol was used by another member of Burrows lab to obtain aptamers for OG, dOG, Sp, and dSp.[69] Although aptamers selected for those targets were not as mutated as the ones selected for Gh, they were found not to have sufficient selectivity for their target molecules. The most likely reason for that is that during the attachment-free SELEX apart from the aptamers that efficiently bind the target molecule strands that are efficient at switching structure without binding or weakly binding the target molecule are also selected. Because of that we decided not to pursue further studies of the aptamers selected for Gh. The only solution for this issue is using a different approach for aptamer selection. Potentially, conventional SELEX may be done for the products of guanine oxidation immobilized on the solid support. This immobilization can be achieved by using nucleosides or 2-deoxynucleosides modified with biotinylated polyethylene glycol tail attached to 3'-OH or 5'-OH, or short randomized biotinylated oligonucleotides containing one of the oxidized lesions. Alternatively, sol-gel SELEX can be attempted for the products of guanine oxidation, where target molecule is immobilized within a sol-gel matrix without covalent attachment to the solid support.[68, 70]

References

1. Noll, D. M.; Mason, T. M.; Miller, P. S., Formation and repair of interstrand cross-links in DNA. *Chem. Rev.* **2006,** *106* (2), 277-301.

2. Oleinick, N. L.; Chiu, S. M.; Ramakrishnan, N.; Xue, L. Y., The formation, identification, and significance of DNA-protein cross-links in mammalian cells. *Br. J. Cancer Suppl.* **1987,** *8*, 135-140.

3. Johansen, M. E.; Muller, J. G.; Xu, X.; Burrows, C. J., Oxidatively induced DNA-protein cross-linking between single-stranded binding protein and oligodeoxynucleotides containing 8-oxo-7,8-dihydro-2'-deoxyguanosine. *Biochemistry* **2005,** *44* (15), 5660-5671.

4. Ghosh, A. K.; Schuster, G. B., Role of the guanine N1 imino proton in the migration and reaction of radical cations in DNA oligomers. *J. Am. Chem. Soc.* **2006,** *128* (13), 4172-4173.

5. Xu, X. Y.; Muller, J. G.; Ye, Y.; Burrows, C. J., DNA-protein cross-links between guanine and lysine depend on the mechanism of oxidation for formation of C5 vs C8 guanosine adducts. *J. Am. Chem. Soc.* **2008,** *130* (2), 703-709.

6. Yun, B. H.; Geacintov, N. E.; Shafirovich, V., Generation of guanine-thymidine cross-links in DNA by peroxynitrite/carbon dioxide. *Chem. Res. Toxicol.* **2011,** *24* (7), 1144-1152.

7. Alshykhly, O. R.; Fleming, A. M.; Burrows, C. J., 5-Carboxamido-5-formamido-2-iminohydantoin, in addition to 8-oxo-7,8-dihydroguanine, Is the major product of the iron-Fenton or X-ray radiation-induced oxidation of guanine under aerobic reducing conditions in nucleoside and DNA contexts. *J. Org. Chem.* **2015,** *80* (14), 6996-7007.

8. Thapa, B.; Munk, B. H.; Burrows, C. J.; Schlegel, H. B., Computational study of the radical mediated mechanism of the formation of C8, C5, and C4 guanine:lysine adducts in the presence of the benzophenone photosensitizer. *Chem. Res. Toxicol.* **2016,** *29* (9), 1396-1409.

9. Fleming, A. M.; Kannan, A.; Muller, J. G.; Liao, Y.; Burrows, C. J., Copper/H_2O_2-mediated oxidation of 2 '-deoxyguanosine in the presence of 2-naphthol leads to the formation of two distinct isomeric adducts. *J. Org. Chem.* **2011,** *76* (19), 7953-7963.

10. Cooke, M. S.; Evans, M. D.; Dizdaroglu, M.; Lunec, J., Oxidative DNA damage: mechanisms, mutation, and disease. *FASEB J.* **2003,** *17* (10), 1195-1214.

11. Stirpe, M.; Palermo, V.; Ferrari, M.; Mroczek, S.; Kufel, J.; Falcone, C.; Mazzoni, C., Increased levels of RNA oxidation enhance the reversion frequency in aging pro-apoptotic yeast mutants. *Apoptosis* **2017,** *22* (2), 200-206.

12. Dunlop, R. A.; Brunk, U. T.; Rodgers, K. J., Oxidized proteins: mechanisms of removal and consequences of accumulation. *IUBMB Life* **2009,** *61* (5), 522-527.

13. Honda, K.; Smith, M. A.; Zhu, X.; Baus, D.; Merrick, W. C.; Tartakoff, A. M.; Hattier, T.; Harris, P. L.; Siedlak, S. L.; Fujioka, H.; Liu, Q.; Moreira, P. I.; Miller, F. P.; Nunomura, A.; Shimohama, S.; Perry, G., Ribosomal RNA in Alzheimer disease is oxidized by bound redox-active iron. *J. Biol. Chem.* **2005,** *280* (22), 20978-20986.

14. Steenken, S.; Jovanovic, S. V.; Bietti, M.; Bernhard, K., The trap depth (in DNA) of 8-oxo-7,8-dihydro-2'deoxyguanosine as derived from electron-transfer equilibria in aqueous solution. *J. Am. Chem. Soc.* **2000,** *122* (10), 2373-2374.

15. Neeley, W. L.; Essigmann, J. M., Mechanisms of formation, genotoxicity, and mutation of guanine oxidation products. *Chem. Res. Toxicol.* **2006,** *19* (4), 491-505.

16. Fortini, P.; Pascucci, B.; Parlanti, E.; D'Errico, M.; Simonelli, V.; Dogliotti, E., 8-Oxoguanine DNA damage: at the crossroad of alternative repair pathways. *Mutat. Res.* **2003,** *531* (1-2), 127-139.

17. Zhao, X.; Krishnamurthy, N.; Burrows, C. J.; David, S. S., Mutation versus repair: NEIL1 removal of hydantoin lesions in single-stranded, bulge, bubble, and duplex DNA contexts. *Biochemistry* **2010,** *49* (8), 1658-1666.

18. McKibbin, P. L.; Fleming, A. M.; Towheed, M. A.; Van Houten, B.; Burrows, C. J.; David, S. S., Repair of hydantoin lesions and their amine adducts in DNA by base and nucleotide excision repair. *J. Am. Chem. Soc.* **2013,** *135* (37), 13851-13861.

19. Henderson, P. T.; Delaney, J. C.; Muller, J. G.; Neeley, W. L.; Tannenbaum, S. R.; Burrows, C. J.; Essigmann, J. M., The hydantoin lesions formed from oxidation of 7,8-dihydro-8-oxoguanine are potent sources of replication errors *in vivo*. *Biochemistry* **2003,** *42* (31), 9257-9262.

20. de Souza-Pinto, N. C.; Eide, L.; Hogue, B. A.; Thybo, T.; Stevnsner, T.; Seeberg, E.; Klungland, A.; Bohr, V. A., Repair of 8-oxodeoxyguanosine lesions in mitochondrial dna depends on the oxoguanine DNA glycosylase (OGG1) gene and 8-oxoguanine accumulates in the mitochondrial DNA of OGG1-defective mice. *Cancer Res.* **2001,** *61* (14), 5378-5381.

21. Xie, Y.; Yang, H.; Cunanan, C.; Okamoto, K.; Shibata, D.; Pan, J.; Barnes, D. E.; Lindahl, T.; McIlhatton, M.; Fishel, R.; Miller, J. H., Deficiencies in mouse Myh and Ogg1 result in tumor predisposition and G to T mutations in codon 12 of the K-ras oncogene in lung tumors. *Cancer Res.* **2004,** *64* (9), 3096-3102.

22. Kitsera, N.; Stathis, D.; Luhnsdorf, B.; Muller, H.; Carell, T.; Epe, B.; Khobta, A., 8-Oxo-7,8-dihydroguanine in DNA does not constitute a barrier to transcription,

but is converted into transcription-blocking damage by OGG1. *Nucleic Acids Res.* **2011,** *39* (14), 5926-5934.

23. Sliwinska, A.; Kwiatkowski, D.; Czarny, P.; Toma, M.; Wigner, P.; Drzewoski, J.; Fabianowska-Majewska, K.; Szemraj, J.; Maes, M.; Galecki, P.; Sliwinski, T., The levels of 7,8-dihydrodeoxyguanosine (8-oxoG) and 8-oxoguanine DNA glycosylase 1 (OGG1) - a potential diagnostic biomarkers of Alzheimer's disease. *J. Neurol. Sci.* **2016,** *368*, 155-159.

24. Hayakawa, H.; Kuwano, M.; Sekiguchi, M., Specific binding of 8-oxoguanine-containing RNA to polynucleotide phosphorylase protein. *Biochemistry* **2001,** *40* (33), 9977-9982.

25. Hayakawa, H.; Sekiguchi, M., Human polynucleotide phosphorylase protein in response to oxidative stress. *Biochemistry* **2006,** *45* (21), 6749-6755.

26. Wu, J.; Li, Z., Human polynucleotide phosphorylase reduces oxidative RNA damage and protects HeLa cell against oxidative stress. *Biochem. Biophys. Res. Commun.* **2008,** *372* (2), 288-292.

27. Wu, J.; Jiang, Z.; Liu, M.; Gong, X.; Wu, S.; Burns, C. M.; Li, Z., Polynucleotide phosphorylase protects *Escherichia coli* against oxidative stress. *Biochemistry* **2009,** *48* (9), 2012-2020.

28. Fleming, A. M.; Zhou, J.; Wallace, S. S.; Burrows, C. J., A role for the fifth G-track in G-quadruplex forming oncogene promoter sequences during oxidative stress: do these "spare tires" have an evolved function? *ACS Cent. Sci.* **2015,** *1* (5), 226-233.

29. Liu, M. M.; Bandaru, V.; Bond, J. P.; Jaruga, P.; Zhao, X. B.; Christov, P. P.; Burrows, C. J.; Rizzo, C. J.; Dizdaroglu, M.; Wallace, S. S., The mouse ortholog of NEIL3 is a functional DNA glycosylase *in vitro* and *in vivo*. *Proc. Natl. Acad. Sci. U. S. A.* **2010,** *107* (11), 4925-4930.

30. Krokeide, S. Z.; Laerdahl, J. K.; Salah, M.; Luna, L.; Cederkvist, F. H.; Fleming, A. M.; Burrows, C. J.; Dalhus, B.; Bjoras, M., Human NEIL3 is mainly a monofunctional DNA glycosylase removing spiroimindiohydantoin and guanidinohydantoin. *DNA Repair* **2013,** *12* (12), 1159-1164.

31. Liu, M. M.; Doublie, S.; Wallace, S. S., Neil3, the final frontier for the DNA glycosylases that recognize oxidative damage. *Mutat. Res.-Fundam. Mol. Mech. Mutag.* **2013,** *743*, 4-11.

32. Rolseth, V.; Krokeide, S. Z.; Kunke, D.; Neurauter, C. G.; Suganthan, R.; Sejersted, Y.; Hildrestrand, G. A.; Bjoras, M.; Luna, L., Loss of Neil3, the major DNA glycosylase activity for removal of hydantoins in single stranded DNA, reduces cellular proliferation and sensitizes cells to genotoxic stress. *Biochim. Biophys. Acta* **2013,** *1833* (5), 1157-1164.

33. Zhou, J.; Liu, M. M.; Fleming, A. M.; Burrows, C. J.; Wallace, S. S., Neil3 and NEIL1 DNA glycosylases remove oxidative damages from quadruplex DNA and exhibit preferences for lesions in the telomeric sequence context. *J. Biol. Chem.* **2013,** *288* (38), 27263-27272.

34. Plosky, B.; Samson, L.; Engelward, B. P.; Gold, B.; Schlaen, B.; Millas, T.; Magnotti, M.; Schor, J.; Scicchitano, D. A., Base excision repair and nucleotide excision repair contribute to the removal of N-methylpurines from active genes. *DNA Repair* **2002,** *1* (8), 683-696.

35. Fuss, J. O.; Cooper, P. K., DNA repair: dynamic defenders against cancer and aging. *PLoS Biol.* **2006,** *4* (6), e203.

36. Liu, Y.; Prasad, R.; Beard, W. A.; Kedar, P. S.; Hou, E. W.; Shock, D. D.; Wilson, S. H., Coordination of steps in single-nucleotide base excision repair mediated by apurinic/apyrimidinic endonuclease 1 and DNA polymerase beta. *J. Biol. Chem.* **2007,** *282* (18), 13532-13541.

37. Fraga, C. G.; Shigenaga, M. K.; Park, J. W.; Degan, P.; Ames, B. N., Oxidative damage to DNA during aging: 8-hydroxy-2'-deoxyguanosine in rat organ DNA and urine. *Proc. Natl. Acad. Sci. U. S. A.* **1990,** *87* (12), 4533-4537.

38. Dziaman, T.; Banaszkiewicz, Z.; Roszkowski, K.; Gackowski, D.; Wisniewska, E.; Rozalski, R.; Foksinski, M.; Siomek, A.; Speina, E.; Winczura, A.; Marszalek, A.; Tudek, B.; Olinski, R., 8-Oxo-7,8-dihydroguanine and uric acid as efficient predictors of survival in colon cancer patients. *Int. J. Cancer* **2014,** *134* (2), 376-383.

39. Poulsen, H. E.; Nadal, L. L.; Broedbaek, K.; Nielsen, P. E.; Weimann, A., Detection and interpretation of 8-oxodG and 8-oxoGua in urine, plasma and cerebrospinal fluid. *Biochim. Biophys. Acta Biochim Biophys Acta* **2014,** *1840* (2), 801-808.

40. Schur, P. H.; DeAngelis, D.; Jackson, J. M., Immunological detection of nucleic acids and antibodies to nucleic acids and nuclear antigens by counterimmunoelectrophoresis. *Clin. Exp. Immunol.* **1974,** *17* (1), 209-218.

41. Santella, R. M., Immunological methods for detection of carcinogen-DNA damage in humans. *Cancer Epidemiol. Biomarkers Prev.* **1999,** *8* (9), 733-739.

42. Jayasena, S. D., Aptamers: an emerging class of molecules that rival antibodies in diagnostics. *Clin. Chem.* **1999,** *45* (9), 1628-1650.

43. Nutiu, R.; Mei, S.; Liu, Z.; Li, Y., Engineering DNA aptamers and DNA enzymes with fluorescence-signaling properties. *Pure Appl. Chem.* **2004,** *76* (7-8), 1547-1561.

44. Navani, N. K.; Li, Y., Nucleic acid aptamers and enzymes as sensors. *Curr. Opin. Chem. Biol.* **2006,** *10* (3), 272-281.

45. Porchetta, A.; Vallée-Bélisle, A.; Plaxco, K. W.; Ricci, F., Using distal-site mutations and allosteric inhibition to tune, extend, and narrow the useful dynamic range of aptamer-based sensors. *J. Am. Chem. Soc.* **2012,** *134* (51), 20601-20604.

46. Xing, H.; Hwang, K.; Li, J.; Torabi, S. F.; Lu, Y., DNA aptamer technology for personalized medicine. *Curr. Opin. Chem. Eng.* **2014,** *4*, 79-87.

47. Zheng, D.; Seferos, D. S.; Giljohann, D. A.; Patel, P. C.; Mirkin, C. A., Aptamer nano-flares for molecular detection in living cells. *Nano Lett.* **2009,** *9* (9), 3258-3261.

48. Ferguson, B. S.; Hoggarth, D. A.; Maliniak, D.; Ploense, K.; White, R. J.; Woodward, N.; Hsieh, K.; Bonham, A. J.; Eisenstein, M.; Kippin, T. E.; Plaxco, K. W.; Soh, H. T., Real-time, aptamer-based tracking of circulating therapeutic agents in living animals. *Sci. Transl. Med.* **2013,** *5* (213), 213ra165.

49. Torabi, S. F.; Lu, Y., Functional DNA nanomaterials for sensing and imaging in living cells. *Curr. Opin. Biotechnol.* **2014,** *28*, 88-95.

50. Tuerk, C.; Gold, L., Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **1990,** *249* (4968), 505-510.

51. Ellington, A. D.; Szostak, J. W., *In vitro* selection of RNA molecules that bind specific ligands. *Nature* **1990,** *346* (6287), 818-822.

52. Nahvi, A.; Sudarsan, N.; Ebert, M. S.; Zou, X.; Brown, K. L.; Breaker, R. R., Genetic control by a metabolite binding mRNA. *Chem. Biol.* **2002,** *9* (9), 1043.

53. Breaker, R. R., Riboswitches and the RNA world. *Cold Spring Harb. Perspect. Biol.* **2012,** *4* (2).

54. Yang, Y.; Yang, D.; Schluesener, H. J.; Zhang, Z., Advances in SELEX and application of aptamers in the central nervous system. *Biomol. Eng.* **2007,** *24* (6), 583-592.

55. Kong, H. Y.; Byun, J., Nucleic acid aptamers: new methods for selection, stabilization, and application in biomedical science. *Biomol. Ther. (Seoul)* **2013,** *21* (6), 423-434.

56. McKeague, M.; Derosa, M. C., Challenges and opportunities for small molecule aptamer development. *J. Nucleic Acids* **2012,** *2012*, 748913.

57. Yang, K. A.; Pei, R.; Stefanovic, D.; Stojanovic, M. N., Optimizing cross-reactivity with evolutionary search for sensors. *J. Am. Chem. Soc.* **2012,** *134* (3), 1642-1647.

58. Nutiu, R.; Li, Y., Structure-switching signaling aptamers. *J. Am. Chem. Soc.* **2003,** *125* (16), 4771-4778.

59. Nutiu, R.; Li, Y., *In vitro* selection of structure-switching signaling aptamers. *Angew. Chem. Int. Ed.* **2005,** *44* (7), 1061-1065.

60. Ye, Y.; Muller, J. G.; Luo, W.; Mayne, C. L.; Shallop, A. J.; Jones, R. A.; Burrows, C. J., Formation of ^{13}C-, ^{15}N-, and ^{18}O-labeled guanidinohydantoin from guanosine oxidation with singlet oxygen. Implications for structure and mechanism. *J. Am. Ceram. Soc.* **2003,** *125* (46), 13926-13927.

61. Zhu, J.; Fleming, A. M.; Orendt, A. M.; Burrows, C. J., pH-dependent equilibrium between 5-guanidinohydantoin and iminoallantoin affects nucleotide insertion opposite the DNA lesion. *J. Org. Chem.* **2016,** *81* (2), 351-359.

62. Polverelli, M.; Berger, M.; Mouret, J. F.; Odin, F.; Cadet, J., Acidic hydrolysis of the N-glycosidic bonds of deoxyribo-nucleic acid by hydrogen fluoride stabilized in pyridine. *Nucleosides and Nucleotides* **1990,** *9* (3), 451-452.

63. Helbock, H. J.; Thompson, J.; Yeo, H.; Ames, B. N., N^2-methyl-8-oxoguanine: a tRNA urinary metabolite role of xanthine oxidase. *Free Radic. Biol. Med.* **1996,** *20* (3), 475-481.

64. Luo, W.; Muller, J. G.; Burrows, C. J., The pH-dependent role of superoxide in riboflavin-catalyzed photooxidation of 8-oxo-7,8-dihydroguanosine. *Org. Lett.* **2001,** *3* (18), 2801-2804.

65. Luo, W.; Muller, J. G.; Rachlin, E. M.; Burrows, C. J., Characterization of hydantoin products from one-electron oxidation of 8-oxo-7,8-dihydroguanosine in a nucleoside model. *Chem. Res. Toxicol.* **2001,** *14* (7), 927-938.

66. Ghude, P.; Schallenberger, M. A.; Fleming, A. M.; Muller, J. G.; Burrows, C. J., Comparison of transition metal-mediated oxidation reactions of guanine in nucleoside and single-stranded oligodeoxynucleotide sontexts. *Inorg. Chim. Acta* **2011,** *369* (1), 240-246.

67. Irvine, D.; Tuerk, C.; Gold, L., SELEXION. Systematic evolution of ligands by exponential enrichment with integrated optimization by non-linear analysis. *J. Mol. Biol.* **1991,** *222* (3), 739-761.

68. Bae, H.; Ren, S.; Kang, J.; Kim, M.; Jiang, Y.; Jin, M. M.; Min, I. M.; Kim, S., Sol-gel SELEX circumventing chemical conjugation of low molecular weight metabolites discovers aptamers selective to xanthine. *Nucleic Acid Ther.* **2013,** *23* (6), 443-449.

69. Ghude, P. Structure-switching SELEX for selection of aptamers of damaged nucleosides and nucleobases 8-oxoguanine and spiroiminodihydantoin. Ph.D. dissertation. University of Utah, Salt Lake City, UT, 2015.

70. Ahn, J. Y.; Lee, S.; Jo, M.; Kang, J.; Kim, E.; Jeong, O. C.; Laurell, T.; Kim, S., Sol-gel derived nanoporous compositions for entrapping small molecules and their outlook toward aptamer screening. *Anal. Chem.* **2012,** *84* (6), 2647-2653.

CHAPTER 3

REVERSE TRANSCRIPTION PAST PRODUCTS OF GUANINE OXIDATION IN RNA, 8-OXO-7,8-DIHYDROGUANINE, 5-GUANIDINOHYDANTOIN, AND SPIROIMINODIHYDANTOIN DIASTEREOMERS*

Introduction

The nucleic acids DNA and RNA are prone to oxidative damage from reactive oxygen species (ROS) that are formed during metabolism or induced by exogenous sources, such as UV radiation, ionizing radiation, or environmental toxins.[1-4] Exposure to these damaging agents may result in modifications of nucleobases, strand breaks, or cross-links with other molecules present in the cell.[5-9] Oxidation of DNA has been an area of intense study for over two decades because it can lead to irreversible mutations in the genetic code that result in cancer and numerous genetic diseases.[10-12]

Oxidative damage to RNA has received much less attention. The likely reasons for this include the many challenges of working with the inherently less chemically stable

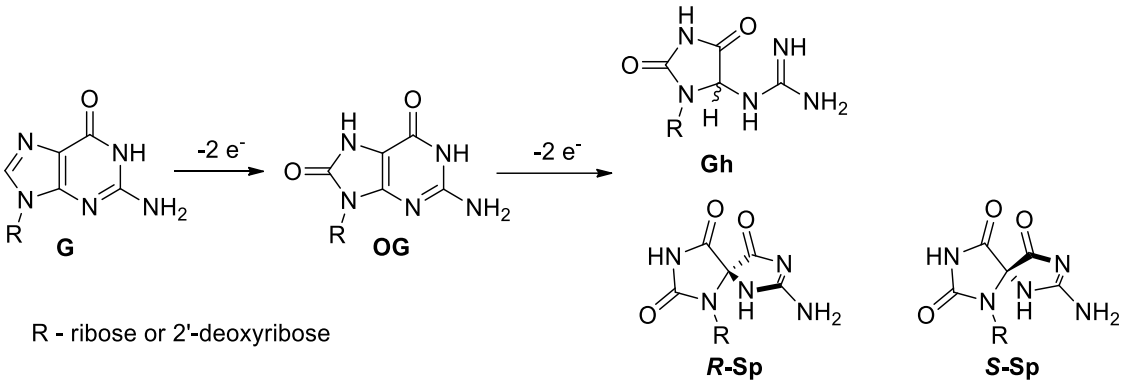
---

*Reproduced in part with permission from "Alenko, A.; Fleming, A. M.; Burrows, C. J., Reverse Transcription Past Products of Guanine Oxidation in RNA Leads to Insertion of A and C opposite 8-Oxo-7,8-dihydroguanine and A and G opposite 5-Guanidinohydantoin and Spiroiminodihydantoin Diastereomers. *Biochemistry* **2017**, *56*(38), 5053-5064." DOI: 10.1021/acs.biochem.7b00730

RNA and the assumption that oxidation of RNA does not significantly disturb normal cell functions due to turnover of RNA molecules in the cell.[13] While the latter may be true for lower organisms with predominantly short-lived RNA,[14] the average half-life of mRNA in human cells is ~10 h while for the long-lived tRNA and rRNA this value can reach several days.[15] On average mammalian cells contain at least as much RNA as DNA[13, 16] and RNA itself is 2-25 times[17-19] more susceptible to oxidation by ROS *in vivo* than DNA. Therefore, RNA oxidation can present a substantial challenge for cell survival, and multiple studies have linked it to development and progression of cancer and neurodegenerative diseases.[20-29] Recent work has discovered pathways of surveillance, sequestration, and, in some cases, repair of RNA damage.[13, 30-34]

Among the nucleobases, guanine has the lowest redox potential (1.29 V vs. NHE) that makes it the major target for oxidizing agents.[35] Thus, one of the most abundant lesions is a product of its 2-electron oxidation, 8-oxo-7,8-dihydroguanine (OG).[36] The redox potential of OG is even lower than that of guanine (0.74 V vs. NHE) making these sites predisposed to further oxidation[35] to yield two hydantoin lesions, 5-guanidinohydantoin (Gh) and spiroiminodihydantoin (Sp) (Scheme 3.1).[37-38] OG has been found in both cellular DNA and RNA, while Gh and Sp have thus far been characterized only in cellular DNA and in RNA oligomers mimicking the tRNA anticodon loop or short single strands of RNA.[18, 39-40] The latter lesion, Sp, exists as a pair of enantiomers (*R*-Sp and *S*-Sp) as the free base, which form a pair of diastereomers once attached to the chiral ribose or 2`-deoxyribose components in nucleic acids.[41]

Mapping positions of the oxidation sites in DNA has been used for achieving a deeper understanding of guanine susceptibility to oxidation and composition of the



Scheme 3.1. Pathways of guanine oxidation.

products on the sequence and structural context.[36, 42-46] However, methods that have been successfully used for finding where oxidation events occur in DNA cannot be directly applied to RNA.[40] Techniques developed for mapping nucleotide modifications in DNA or RNA could be broken into 3 major categories.

Category one includes methods that induce a strand break directly in the studied sequence by an enzyme or a chemical agent (e.g., Maxam-Gilbert chemistry, RNA digestion with specific RNases, or base excision from the DNA strand by repair enzymes) with further analysis of products by polyacrylamide gel electrophoresis (PAGE) or capillary electrophoresis (CE). We have previously tested the possibility of using simple cleavage of oxidized bases in RNA by hot aniline treatment and discovered that most products of guanine oxidation in RNA are more resistant to cleavage under standard conditions than the same lesions in DNA when treated with hot piperidine.[40] In addition, RNA itself is more susceptible than DNA to position-independent backbone cleavage,[47] which results in a high background resulting in a limited practical use of chemical methods for locating lesion sites.

Category two comprises mass spectroscopic techniques developed for characterization of post-transcriptional modifications in RNA (e.g., simple analysis of MS/MS fragmentation of oligonucleotides[48] or more complex LC-MS/MS methods developed by the McCloskey laboratory[49-51]). Although these methods can possibly be highly useful for mapping guanine oxidation products, they have never been used for that purpose. Optimization of this type of analysis for detection of oxidized lesions could be quite challenging considering that oxidized lesions are likely to be more randomly distributed than localized post-transcriptional modifications.

The third category consists of primer extension assays utilizing either incorporation of natural (e.g., bisulfite sequencing of m^5C, or SHAPE-MaP[52]) or unnatural bases (e.g., insertion of artificial nucleotides opposite m^5C or O^6-BnG[53-56]) opposite modified nucleotides or arrest of the polymerase activity, if it cannot efficiently insert a base opposite the modification site (e.g., adenosine methylation by DMS or older versions of SHAPE probing[57-58]). Currently, there are no known artificial bases that can be inserted opposite OG or the hydantoin lesions with the required specificity to exclusively map these sites.[59] At the same time, multiple studies have described insertion of canonical nucleobases opposite these lesions in DNA.[60-65] The polymerases studied insert A or C opposite OG[62-63, 65] and A or G opposite Gh and Sp,[60-61] with the ratio being highly dependent on the polymerase. Among the aforementioned papers, only one was focused on polymerase insertion fidelity opposite OG by RNA-dependent DNA polymerase[62] called reverse transcriptase (RT). In that study, insertion opposite OG in a DNA (not RNA) template strand by HIV1 RT was interrogated finding insertion of A and C in a 14:1 ratio.[62]

Thus, there are so far no conclusive reports on how reverse transcriptase enzymes behave when they encounter OG, Gh, or Sp in a RNA template. This chapter is focused on testing whether commercially available reverse transcriptase enzymes can insert canonical nucleobases opposite OG, Gh, or Sp when located in a RNA template. For the potential use of reverse transcription as a method for detection of oxidized guanine lesions, reverse transcription should result in either termination of polymerization at the lesion sites or insertion of any base other than C to allow discrimination from unoxidized G. Considering that several types of RNA (18S rRNA and tRNA in eukaryotes[66]) contain post-transcriptional modifications, such as m^1G or $m^2{}_2G$, that also result in polymerase arrest,[67]

the latter option would be preferable for application of this lesion sequencing approach to biological samples. Overall this study provides a foundation for development of a method for mapping OG, Gh, or Sp in RNA templates using reverse transcription to induce mutations that can be tracked after next-generation sequencing.

Materials and methods

Oligomer synthesis. DNA and RNA oligomers were synthesized by the core facilities at the University of Utah using solid-phase synthesis following standard protocols. The RNA templates containing OG were synthesized using the commercially available rOG phosphoramidite (ChemGenes, Wilmington, MA). The RNA strands were synthesized with a 3` dT to maximize the solid-phase synthesis yield of the modified RNA strands; the added dT will not impact the polymerase extension studies. All oligonucleotides were purified via analytical ion-exchange HPLC and dialyzed against ddH_2O.

Synthesis of the Sp and Gh hydantoins into the RNA strands was accomplished by Aaron Fleming. This was achieved utilizing the RNA strand with rOG incorporated via its phosphoramidite at the site of modification. Synthesis of Gh was conducted by dissolving 1 nmol of rOG-containing RNA into 50 µL of ddH_2O. The sample was placed on ice for 10 min followed by a bolus addition of 12 equivalents of Na_2IrCl_6, and the reaction was allowed to progress for 30 min. The rGh-containing strands were purified from the reaction mixture by ion-exchange HPLC on a DNAPac PA100 column (250 x 4.6 mm). The HPLC mobile phases consisted of A = 9:1 ddH_2O:MeCN and B = 1.5 M NaOAc (pH 7.0) in 9:1 ddH_2O:MeCN while running a linear gradient from 25% B to 100% B over 30 min with a

flow rate of 1 mL/min while monitoring the absorbance at 260 nm. The RNA strands containing diastereomers of rGh were purified and studied as a mixture because the epimers readily interconvert.[68] The purified strands were dialyzed against ddH_2O for 24 h while changing the ddH_2O every 6 h to remove the purification salts. The rSp-containing strands were synthesized by placing 1 nmol of purified rOG-containing RNA into 50 µL of 20 mM NaP_i (pH 7.4) buffer with 100 mM NaCl at 22 °C followed by a bolus addition of 12 equivalents of Na_2IrCl_6. The reaction was allowed to progress for 30 min followed by workup via the same method outlined for the rGh-containing RNA strands. The diastereomers of rSp were individually purified for the polymerase studies. The absolute configurations for the Sp diastereomers have been determined in DNA strands and nucleosides but not in RNA strands or nucleosides[41]; therefore, we validated the absolute configurations for the Sp diastereomers in RNA strands and nucleosides to find identical results between the two polymers (Figure 3.1). The purity of the hydantoins in the RNA oligomers studied was determined by ion-exchange HPLC (Figure 3.2) and the product identities were identified by ESI-MS (OG-1 calcd = 6608.0, expt = 6608.1; Gh-1 calcd = 6598.0, expt = 6598.5; *S*-Sp-1 calcd = 6624.0, expt = 6624.8; *R*-Sp-1 calcd = 6624.0, expt = 6624.5).

Confirmation of the absolute configurations for the rSp diastereomers in a single-stranded RNA. Previously our laboratory determined the absolute configurations for the Sp diastereomers in short DNA oligomers and 2`-deoxyribonucleosides.[69] Our work noted that the elution order of the dSp diastereomers changed when going from analysis in a short, single strand of DNA analyzed by ion-exchange HPLC vs. dSp in the nucleoside context analyzed by a Hypercarb column. In the present work, we verified the same elution

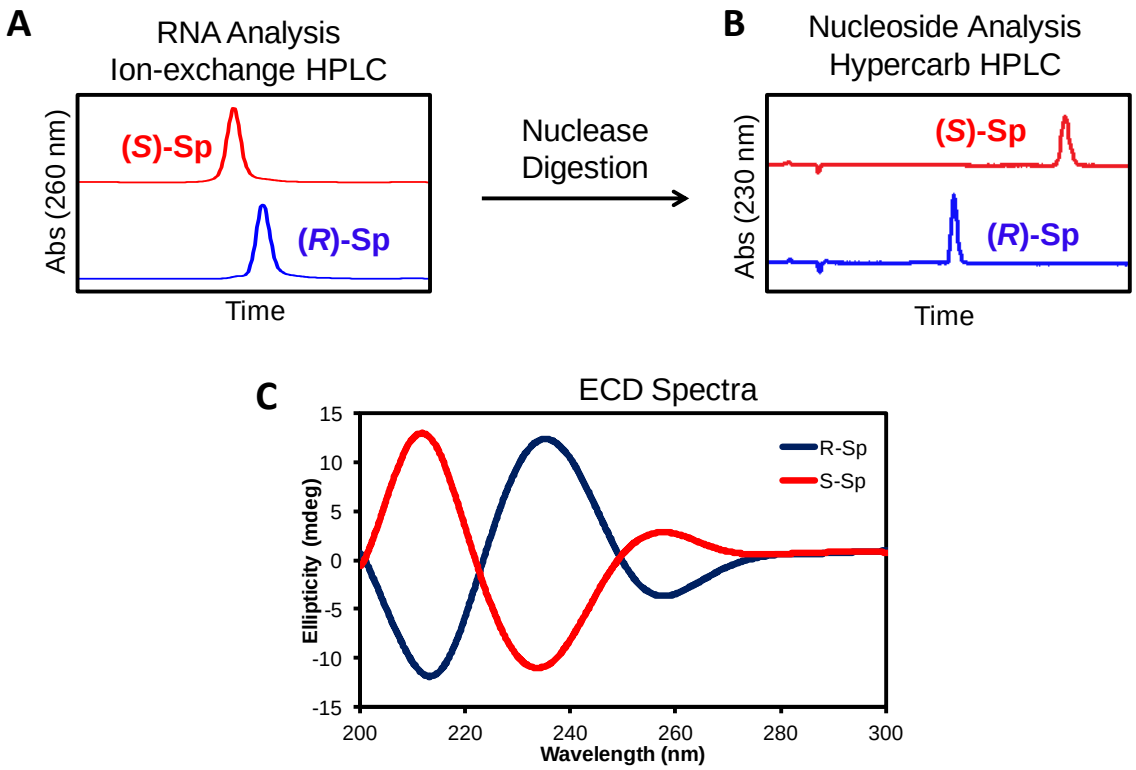


Figure 3.1. Confirmation of the absolute configurations for the rSp diastereomers in a single-stranded RNA.

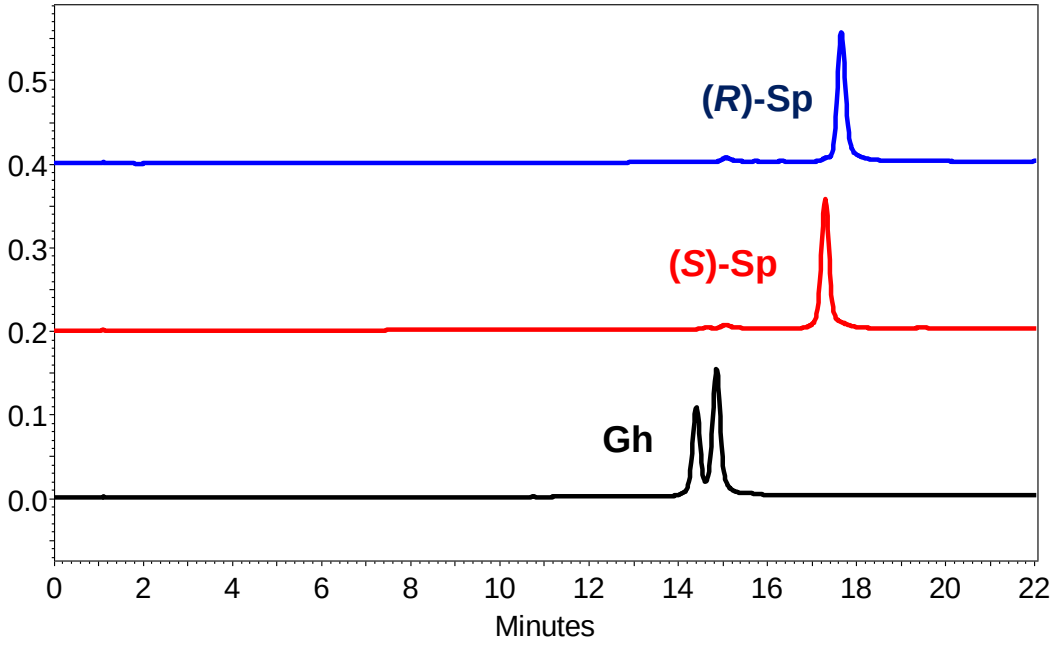


Figure 3.2 . Example ion-exchange HPLC traces to demonstrate the purity of the Gh-, (S)-Sp, and (R)-Sp-containing RNA strands.

order switching phenomenon occurs in short RNA strands. Confirmation of the absolute configurations for the rSp diastereomers in the RNA templates studied was achieved as follows. First, the RNA strands were purified to contain only one rSp diastereomer using a Dionex DNAPac PA100 (250 x 4.6 mm) ion-exchange HPLC column following a method previously described by our laboratory for dSp in a DNA strand (Figure 3.1 A).[69] Next, the purified strands were digested to nucleosides using nuclease P1, snake venom phosphodiesterase, and calf intestinal phosphatase following a method previously described by our laboratory.[70] The nucleosides were then analyzed on a Hypercarb HPLC (150 x 4.6 mm, 5 µm; Thermo Scientific) column to verify the elution order of the diastereomers switched when compared to the ion-exchange HPLC column (Figure 3.1 B). Lastly, to confirm the absolute configurations, the individual rSp diastereomeric nucleosides were HPLC purified and then analyzed by electronic circular dichroism (ECD) spectroscopy. The ECD spectra obtained for each rSp diastereomer were identical to those previously obtained for the dSp diastereomers (Figure 3.1 C).[69] Finally, this confirms that in the RNA templates studied, the first eluting strand has the *S* diastereomer of Sp and the second eluting strand has the *R* diastereomer of Sp. Characterization of rSp diastereomers has been accomplished by Aaron Fleming.

Labeling of the DNA primer. To monitor progression of primer extension by reverse transcriptases via PAGE, the DNA primer was 5' end-labeled with ^{32}P following a procedure adopted from the literature[71] using T4 polynucleotide kinase (New England Biolabs, Ipswich, MA.) and [γ-^{32}P] ATP (PerkinElmer, Waltham, MA.).

Polymerase nucleotide insertion and extension efficiency assays. The following enzymes were used: SuperScript III (200 U/µL, Invitrogen, Carlsbad, CA), AMV RT (25

U/µL, New England Biolabs, Ipswich, MA.), MMLV RT (200 U/µL, New England Biolabs, Ipswich, MA.), ProtoScript II (200 U/µL, New England Biolabs, Ipswich, MA.), and Omniscript (4 U/µL, Qiagen, Hilden, Germany). Before all primer extension assays, samples were annealed by heating 10-14 µL of an aqueous solution containing 0.44 pmol (22 nM in a final volume of 20 µL) of RNA template and 0.4 pmol (20 nM in 20 µL) DNA primer including ~30,000 cpm of ^{32}P-labeled strand to 95 °C for 5 min, then incubating them at 55 °C for another 5 min followed by cooling at 15 °C for 10 min. Upon completion of annealing, stock solutions of reaction buffer (50 mM Tris-HCl, 75 mM KCl (75 mM KAc for AMV RT), 3 mM $MgCl_2$ (8 mM $MgAc_2$ for AMV RT), pH 8.3 in 20 µL, 1x commercial reaction buffer for Omniscript), DTT (10 mM in 20 µL and the commercially defined concentration of DTT in the buffer for Omniscript), dNTPs were added to the reaction mixture to bring the volume to 18 µL. Then 2 µL of stock solution containing one of the reverse transcriptases in 50% glycerol was added to give a final volume of 20 µL (final concentrations of dNTPs and enzymes are provided in the next paragraph). Reaction mixtures were incubated for 30 min at 37 °C, and then to quench the reaction, the mixture was diluted with an equal volume of 2x gel loading buffer (8 M urea, 0.01% xylene cyanol, 0.01% bromphenol blue, x1 TBE buffer) and heated to 95 °C for 10 min. About 15 µL of the resulting solution were analyzed via 20% denaturing PAGE. Gels were stored with a storage phosphor screen for 12-18 h that was then scanned using a phosphoimager. The resulting images were analyzed using ImageJ2 software.[72-73] For alignment of the gel lane pixel density plots, they were rescaled along the y-axis to normalize intensities and translated along the x-axis without rescaling to align the peaks corresponding to the unextended primer.

Final concentrations of enzymes and triphosphates for polymerase nucleotide insertion studies were as follows: for templates containing OG or G – 50 μM dATP, dCTP, dGTP, or dTTP, 3U SuperScript III, 0.4 U AMV RT, 2 U MMLV RT, 5U ProtoScript II, 0.3 U Omniscript in 20 μL; for Gh-1 template – 100 μM individual dNTPs, 4U SuperScript III in 20 μL; for *S*-Sp-1 template - 200 μM individual dNTPs, 40U SuperScript III in 20 μL; for *R*-Sp-1 template - 200 μM individual dNTPs, 20U SuperScript III in 20 μL. Final concentrations for full extension efficiency study were 100U of SuperScript III in 20 μL and 500 or 200 μM of each dNTP (dATP, dCTP, dGTP, and dTTP). Final concentrations for comparing extension efficiency past OG-A and OG-C were 6U SuperScript III in 20 μL and 100 μM dATP for lanes 1 (A), 2 (AT), 7 (AC), and 8 (ACT), 200 μM dATP for lane 5 (2A), 100 μM dCTP for lanes 3 (C), 4 (CT), 7 (AC), and 8 (ACT), 200 μM dCTP for lane 6 (2C), and 100 μM dTTP for lanes 2 (AT), 4 (CT), and 8 (ACT). Final concentrations for comparing extension efficiency with Gh-1 template were 8U SuperScript III in 20 μL and 200 μM dATP for lanes 1 (A), 2 (AT), 7 (AG), and 8 (AGT), 400 μM dATP for lane 5 (2A), 200 μM dGTP for lanes 3 (G), 4 (GT), 7 (AG), and 8 (AGT), 400 μM dGTP for lane 6 (2G), and 200 μM dTTP for lanes 2 (AT), 4 (GT), and 8 (AGT). Final enzyme concentrations for reactions using *S*-Sp-1 and *R*-Sp-1 templates were 80U and 40U SuperScript III in 20 μL correspondingly; concentrations of triphosphates were the same as described for the Gh-1 template.

Steady-state kinetics. A procedure analogous to that described in the previous section was used with the following changes. Instead of 30,000 cpm of ^{32}P-labeled strand, 100,000 cpm was added. After annealing, only reaction buffer and DTT stock solutions were added, and then the samples were preincubated for 1 min at 37 °C, followed by

addition of 2 µL of SuperScript III stock solution to give 1U (OG), 4U (Gh), 10U (*R*-Sp), or 20U (*S*-Sp) in 20 µL followed by preincubation for another 1 min at 37 °C and then addition of 2 µL of a stock solution of dATP, dCTP, or dGTP of varied concentrations to initiate the reaction. Next, 5-µL aliquots were taken in regular time intervals for the first 2-2.5 min of reaction, rapidly mixed into an equal volume of the loading buffer and heated to 95 °C for 10 min. The diluted stock solutions of triphosphates and enzymes were prepared fresh and used the same day. Samples were analyzed on a denaturing PAGE gel the same way as described above. Intensities of the bands corresponding to the unextended primer (P) and the primer extended by one base (P+1) were quantified using ImageJ2 software[72-73] and converted into the concentration of P+1 strand using eq 1,[61, 74] where C_{P+1} is the concentration of P+1 strand, C_t is the total concentration of the radiolabeled primer, and I_P and I_{P+1} are intensities of P and P+1 bands on the gel. No significant accumulation of higher-order bands was observed. Initial reaction velocities were extracted from fitting the data in coordinates of C_{P+1} (nM) vs. time (min) to a linear regression curve as a slope of the fitted line (Figure 3.3). Each experiment was repeated at least 3 times. Conversion of V_{max} to k_{cat} was achieved using specific activity and the molecular weight of SuperScript III provided by Invitrogen (410,000 U/mg, 78 kDa).

$$C_{P+1} = C_t \frac{I_{P+1}}{I_P + I_{P+1}} \quad (3.1)$$

Results and discussion

Reverse transcriptases. Reverse transcriptases are polymerase type enzymes that synthesize a complementary DNA (cDNA) based on an RNA template. While enzymes of

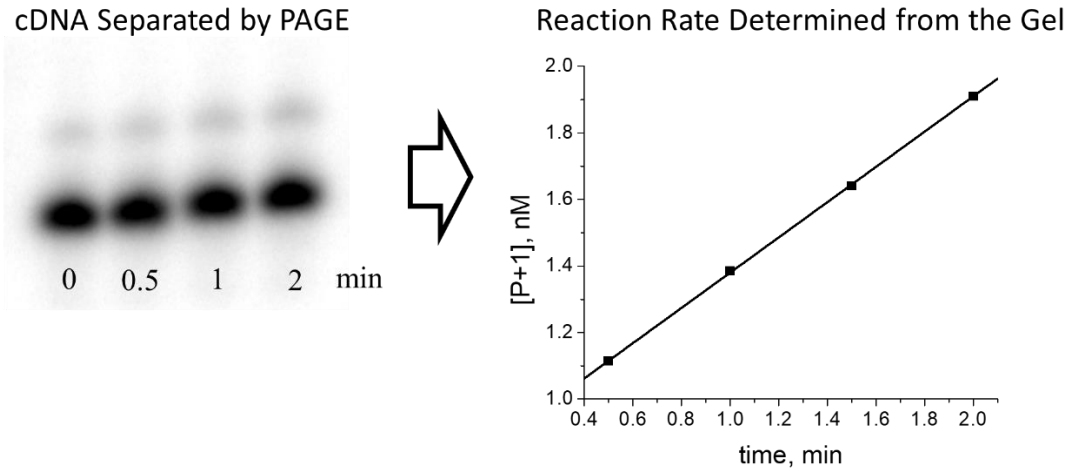


Figure 3.3. Example of data processing for determining initial nucleotide incorporation rates.

this type have been discovered in prokaryotes and yeast,[75-76] the majority of them have retroviral origins. Most commercially available reverse transcriptase enzymes originate either from Avian Myeloblastosis Virus (AMV) RT or Moloney Murine Leukemia Virus (MMLV) RT. The primary purpose of reverse transcription in contemporary research is sequencing and quantification of mRNA via RT-PCR and RT-qPCR. Engineered versions of these enzymes have been created by multiple vendors to improve their stability, sensitivity, and reverse transcription yield (Table 3.1).[77-78] In the current study, we screened insertion profiles of 5 RT enzymes (underlined in Table 3.1) opposite OG and G in a RNA template. Because similar insertion profiles have been observed for all reverse transcriptases used, only one of them, SuperScript III, was picked for further experiments. SuperScript III was chosen over other enzymes due to it being one of the polymerases of choice for next-generation sequencing of RNA.[58, 77-80]

Polymerase nucleotide insertion studies. To study the behavior of RT enzymes encountering products of guanine oxidation in RNA, we designed two RNA-DNA hybrid duplexes (Figure 3.4). Both of them employ the same DNA primer and are suited for studying reverse transcription under standing start conditions with the only difference between the two being two nucleotides on the 5' end following guanine or its oxidation products. To perform an initial evaluation of reverse transcription as a way of detecting oxidized lesions, we studied insertion profiles opposite G and OG in both RNA templates for 5 RT enzymes: SuperScript III, AMV RT, MMLV RT, Omniscript, and ProtoScript II. To do so, we ran reverse transcription reaction with A, G, C, or T triphosphates in separate tubes for a fixed amount of time and resolved reaction products on a polyacrylamide gel. Although all these RTs have identical unit definitions (amount of an enzyme incorporating

Table 3.1. Selection of commercially available reverse transcriptases.

	Original RT	Enzyme concentration
AMV RT	AMV RT	25 U/µL
MMLV RT	MMLV RT	200 U/µL
ProtoScript II®	MMLV RT	200 U/µL
Omniscript®	undisclosed	4 U/µL
Sensiscript®	undisclosed	undisclosed
ImProm-II®	undisclosed	undisclosed
ThermoScript®	AMV RT	15 U/µL
SuperScript II®	MMLV RT	200 U/µL
SuperScript III®	MMLV RT	200 U/µL
SuperScript IV®	MMLV RT	200 U/µL
TGIRT®	Group II intron RT	200 U/µL

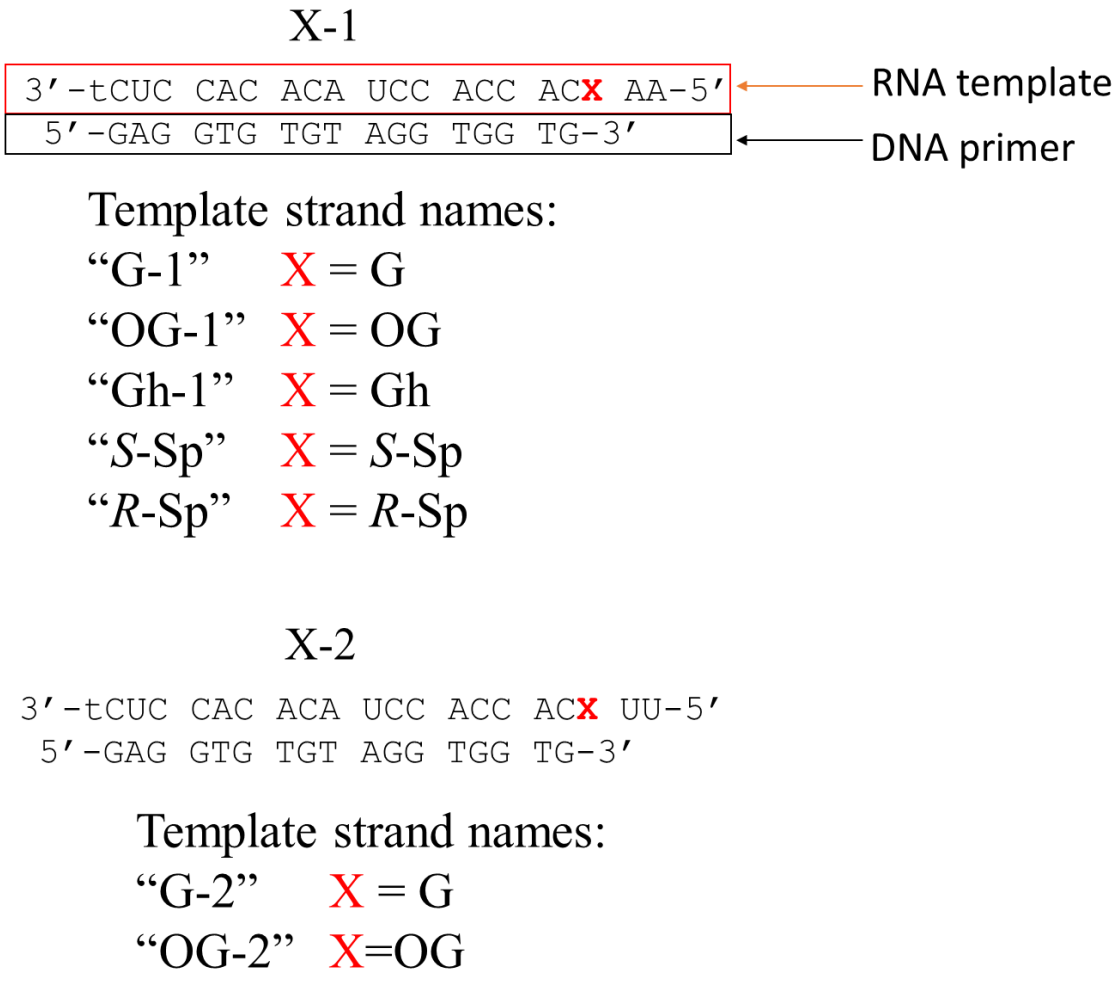


Figure 3.4. Studied RNA-DNA hybrid duplexes.

1 nmol dTTP in 50 µL in 10 min at 37 ºC using poly(rA)18 • poly(dT)12-18 duplex and 500 µM dTTP), their activities are standardized under different conditions (0.1-0.4 mM primer-template duplex, 3-6 mM $MgCl_2$, 40-75 mM KCl, 1-10 mM DTT, 0-0.1 mg/mL BSA; no data are available on Omniscript). Additionally, there are slight variations in the reaction buffers supplied with AMV and the other enzymes that were used for the reactions (75 mM KOAc instead of 75 mM KCl and 8mM $MgOAc_2$ instead of 3 mM $MgCl_2$; no data are available on Omniscript reaction buffer composition). Thus, we found it necessary to individually adjust the concentration of each enzyme to achieve ~50% insertion of dCTP (50 µM) opposite OG in the OG-1 strand. Results of the interrogation of nucleotide insertion profile of the reverse transcriptases studied are presented in Figure 3.5, Figure 3.6, and in Table 3.2. On the gels, the lowest band corresponds to the unextended primer (P) and all bands above it correspond to a primer extended by 1 (P+1), 2 (P+2), or 3 (P+3) nucleotides.

Although some reverse transcriptases showed a preference for one sequence context over another (e.g., SuperScript III gave more primer extension product for the OG-1 template than for the OG-2 template), all examined enzymes demonstrated matching nucleotide insertion profiles inserting C or A opposite OG and C or T opposite G when only one of the dNTPs was present. Considering that the assayed reverse transcriptases demonstrated similar patterns of nucleotide insertion for both RNA templates, we limited further studies to using only SuperScript III and template number 1 (Figure 3.4). Before examining which nucleotides are inserted opposite Gh and diastereomers of Sp, we again made adjustments to concentrations of the triphosphates and the enzyme to achieve similar reactivities. Based on the adjusted parameters, the rate of polymerization for the RNA

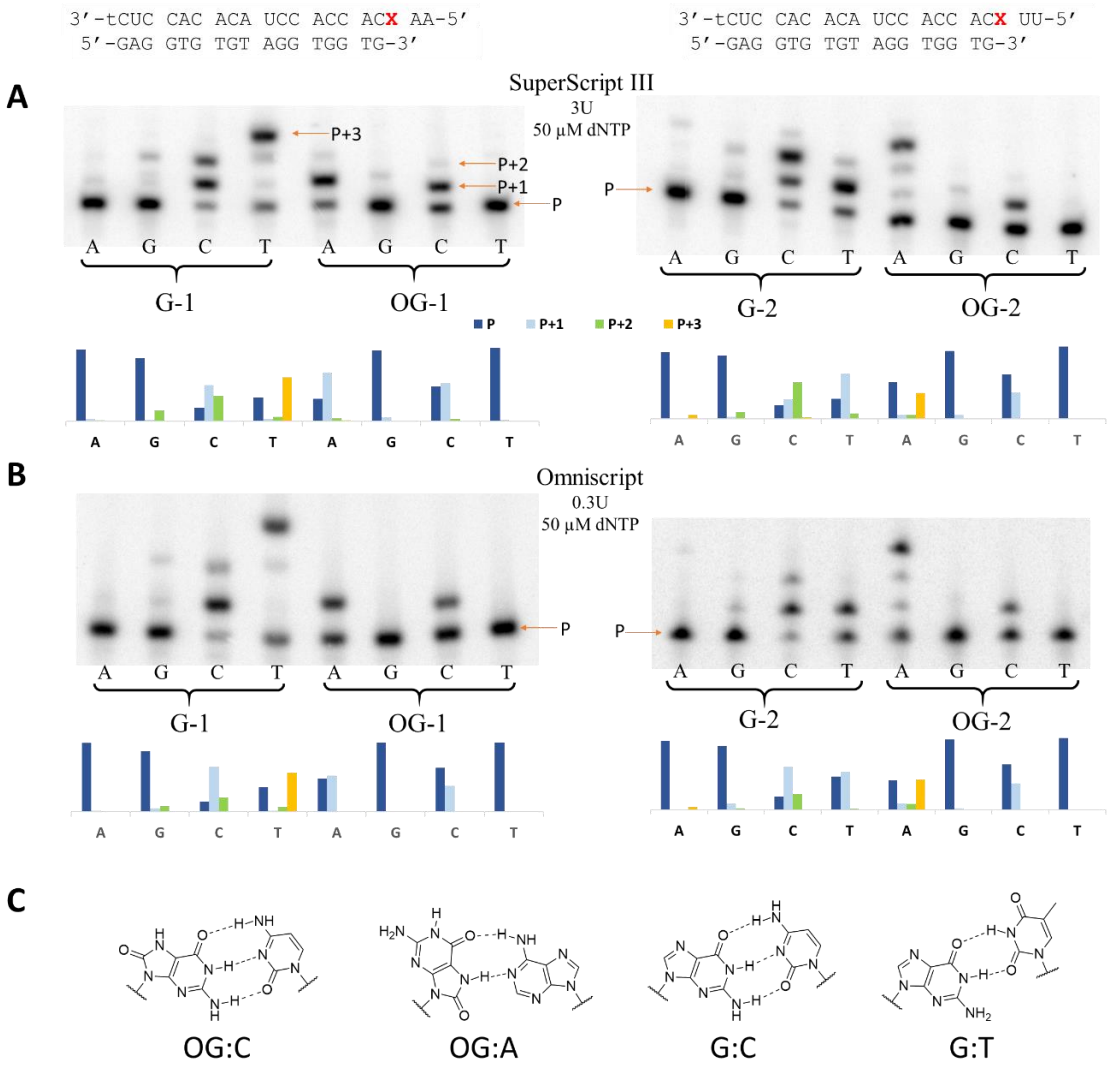


Figure 3.5. Nucleotide insertion profiles opposite G or OG in the templates. Insertion of A, G, C, or T in the reaction mixture with only one of the dNTPs present by SuperScript III **(A)** and Omniscript **(B)** reverse transcriptases for G-1, OG-1, G-2, and OG-2 template.

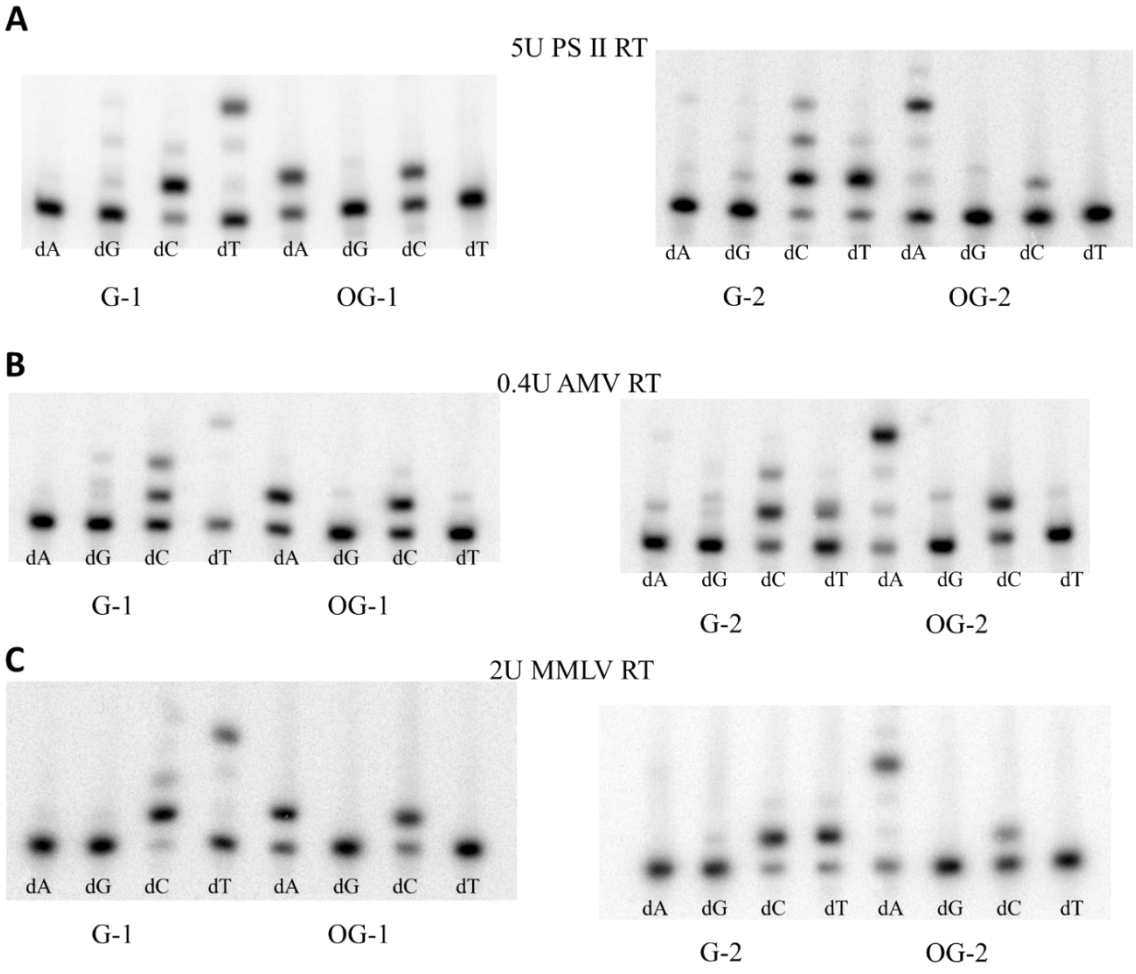


Figure 3.6 . Insertion profiles opposite G and 8-OG in the templates for ProtoScript II **(A)**, AMV RT **(B)**, and MMLV RT **(C)**.

Table 3.2. Results of nucleotide insertion assays for AMV RT, MMLV RT, and ProtoScript II RT.

AMV RT

G-2	C	33%	47%	19%	1%	
	T	60%	37%	2%		
OG-2	A	25%	10%	5%	60%	
	C	42%	58%			
G-1	C	44%	33%	23%		
	T	76%	2%	2%	19%	
OG-2	A	41%	58%	1%		
	C	42%	55%	2%		

MMLV RT

G-2	C	21%	75%	4%		
	T	25%	72%	4%		
OG-2	A	37%	4%	3%	53%	3%
	C	70%	30%			
G-1	C	11%	71%	15%	3	
	T	53%	2%	5%	39%	
OG-2	A	36%	64%			
	C	33%	67%			

ProtoScript II

G-2	C	18%	50%	20%	12%	
	T	25%	68%	7%		
OG-2	A	50%	9%	4%	35%	2%
	C	78%	22%			
G-1	C	26%	67%	7%		
	T	51%	6%	6%	38%	
OG-1	A	43%	57%			
	C	54%	46%			
		P	P+1	P+2	P+3	P+4

templates containing products of guanine oxidation decreased in the following order: OG>Gh>>*R*-Sp>*S*-Sp. All three hydantoin lesions showed similar base pairing preferences leading to insertion of A or G (Figure 3.7). Insertion of A or G opposite Gh, *S*-Sp, and *R*-Sp when only one of the triphosphates was present was similar to insertion of A or C opposite OG, in perfect agreement with what was previously reported for insertion opposite these lesions in DNA.[60-63, 65] These results show that reverse transcriptases are capable of inserting canonical A, G, or C bases opposite OG, Gh, *S*-Sp, and *R*-Sp in an RNA template.

Steady-state kinetics. To examine SuperScript III behavior when it encounters oxidized lesions in more detail, we studied the kinetics of insertion of A or C opposite OG and A or G opposite Gh, *S*-Sp, or *R*-Sp in the template RNA strand. Additionally, insertion of C opposite G in the template was studied for a comparison that served as a control for SuperScript III kinetics when encountering canonical bases. Initial velocities of first-base insertions were derived in units of nM/min and plotted against concentration of triphosphate used in each case. The data were then fitted to the Michaelis-Menten equation (eq 2) to determine the steady-state kinetics parameters V_{max} and K_M. V_{max} was then divided by the total enzyme concentration to calculate k_{cat} according to the eq 3. Michaelis-Menten curves and calculated kinetics parameters are shown on Figure 3.8, Figure 3.9, and Table 3.3. The error bars on the graphs and in the table are representative of 68% confidence intervals.

$$V = \frac{V_{\max}[\mathrm{dNTP}]}{K_{\mathrm{M}} + [\mathrm{dNTP}]} \tag{3.2}$$

$$k_{\mathrm{cat}} = {V_{\max}}/{\mathrm{E_t}} \tag{3.3}$$

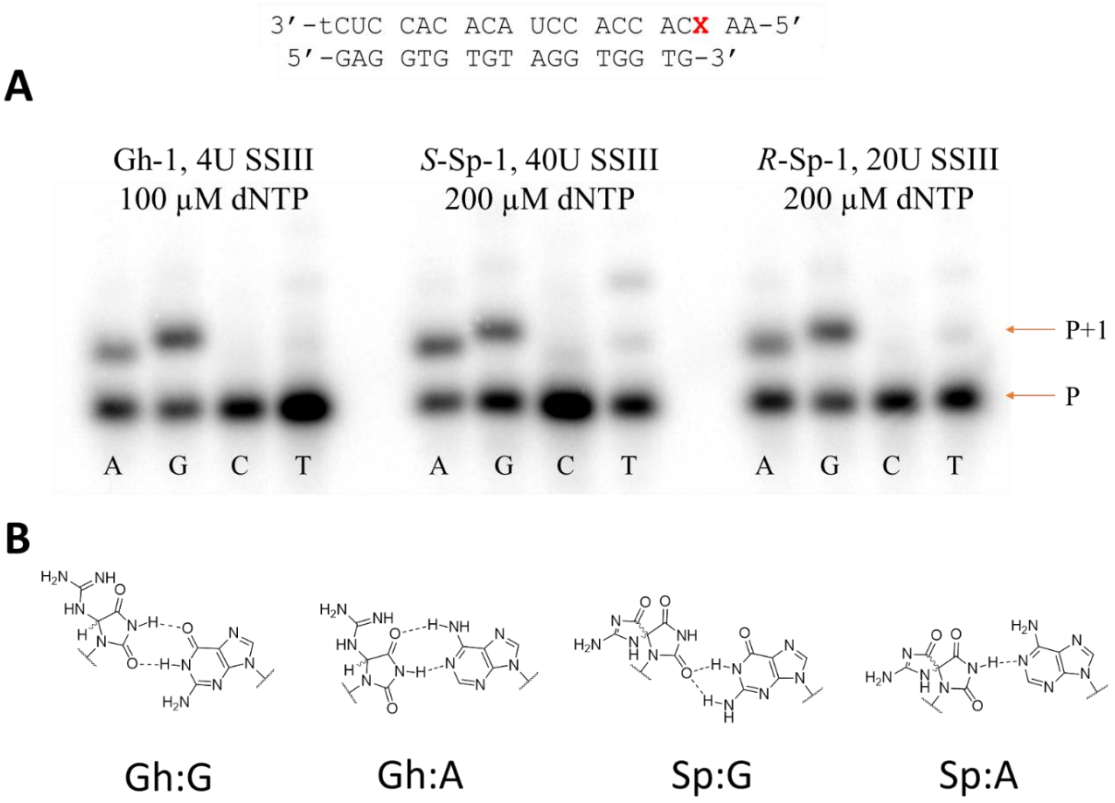


Figure 3.7. Nucleotide insertion assays for Gh and Sp. **(A)** Insertion of A, G, C, or T in the reaction mixture with only one of the dNTPs present by SuperScript III reverse transcriptase for Gh-1, *S*-Sp-1, and *R*-Sp-1 templates. (**B)** Structures of the base pairs between G or A and Gh or Sp are based on previous literature reports.[81-82]

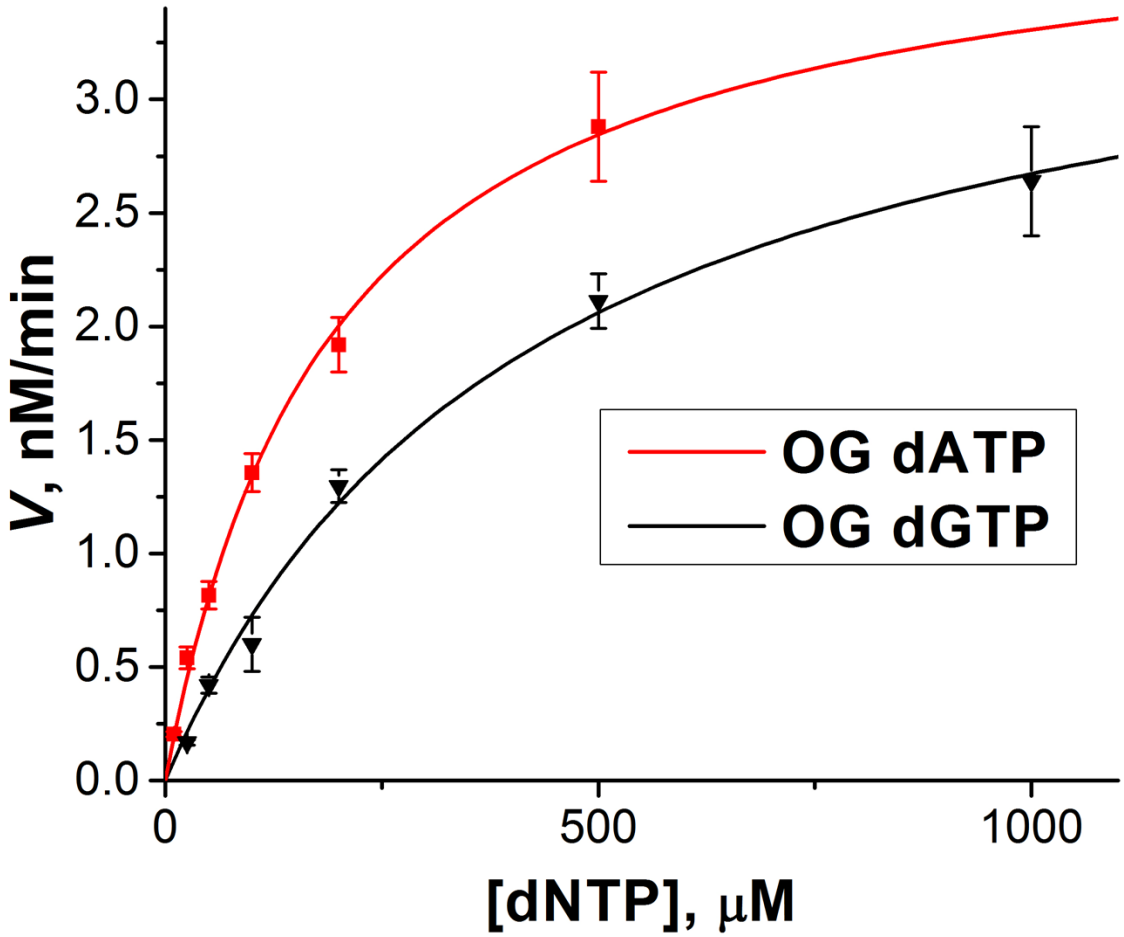


Figure 3.8. Michaelis-Menten plot for insertion of A or G opposite OG.

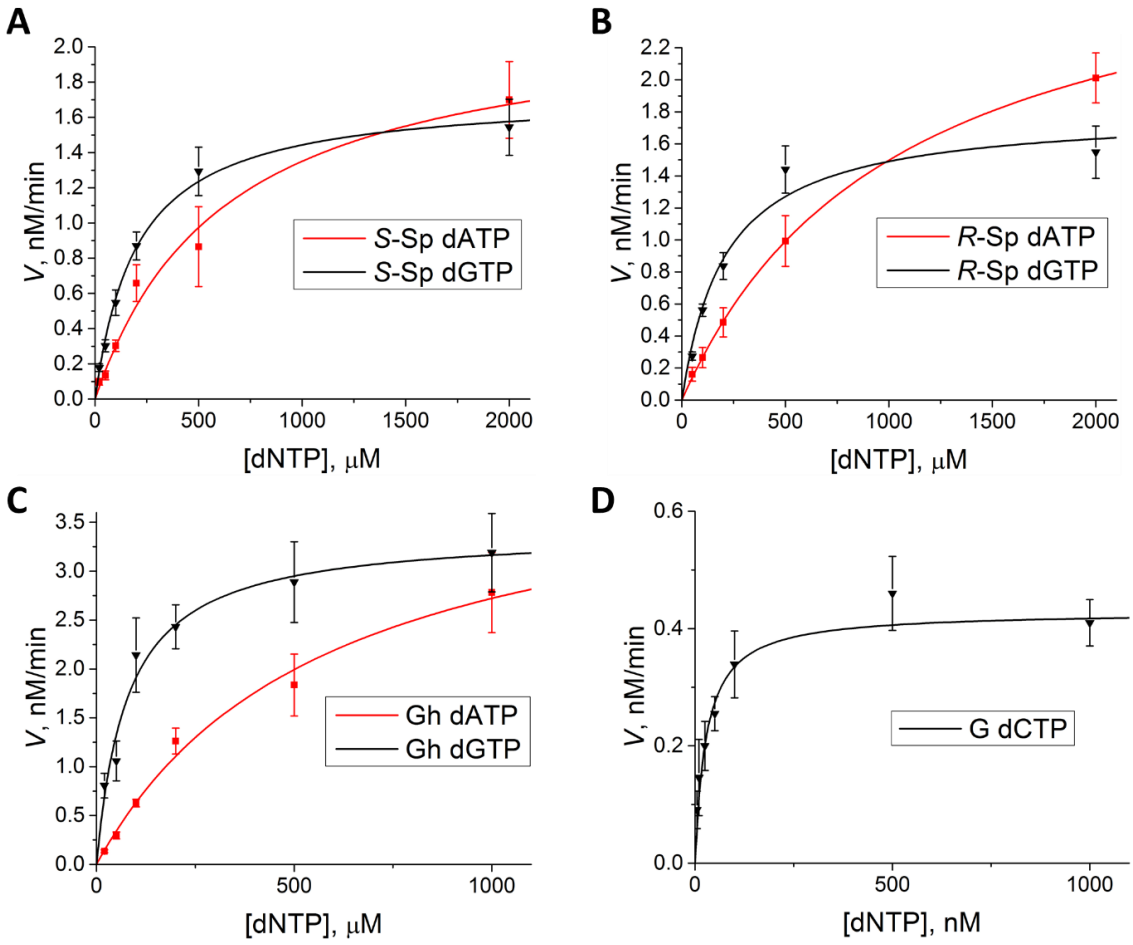


Figure 3.9. Michaelis-Menten plots for insertion of A or G opposite *S*-Sp (**A**), A or G opposite *R*-Sp (**B**), A or G opposite Gh (**C**), and C opposite G (**D**).

Table 3.3. Steady-state kinetic parameters.

		Michaelis-Menten parameters			V/E_t at 500 μM
		k_{cat}, min^{-1}	K_M, μM	k_{cat}/K_M*10^3	dNTP, min^{-1}
OG-1	**dCTP**	2.4 ± 0.2	450 ± 60	5.3	1.3
	dATP	2.5 ± 0.1	190 ± 20	13	1.8
Gh-1	**dGTP**	0.55 ± 0.03	80 ± 15	7	0.47
	dATP	0.69 ± 0.08	600 ± 100	1	0.31
***S*-Sp-1**	**dGTP**	0.056 ± 0.002	200 ± 20	0.29	0.041
	dATP	0.071 ± 0.006	600 ± 100	0.12	0.032
***R*-Sp-1**	**dGTP**	0.12 ± 0.01	210 ± 50	0.57	0.085
	dATP	0.196 ± 0.003	1040 ± 30	0.19	0.064
G	**dCTP**	2.7±0.2	0.028±0.005	96,000	2.7

all error bars represent 68% confidence interval

Three main parameters derived from the Michaelis-Menten curve are k_{cat}, K_M, and their ratio k_{cat}/K_M. The catalytic rate constant or turnover number k_{cat} indicates the maximum number of reactions a single enzyme can catalyze per unit of time. In the case of reverse transcription, it is a combination of the rate of catalysis (condensation between 3' end of the primer strand and a triphosphate being inserted opposite a studied base) and the rate of dissociation between the primer-template complex and the enzyme after the primer extension.

For polymerases, k_{cat} is normally defined by the rate of the slower dissociation step;[62-64, 83-84] thus, it was not surprising that we observed very close values with overlapping confidence intervals for formation of OG-A (2.5 ± 0.1 min^{-1}), OG-C (2.4 ± 0.2 min^{-1}), and G-C (2.7 ± 0.2 min^{-1}) base pairs. Similarities between turnover numbers for incorporation of a base opposite G and OG have been previously reported for translesion DNA polymerase η and HIV-1 RT by the Guengerich laboratory.[62, 64] Unlike that of OG, the rate of phosphodiester bond formation apparently was affected by the presence of hydantoin lesions in the template strong enough to result in lower k_{cat} values for Gh-A (0.69 ± 0.08 min^{-1}), Gh-G (0.55 ± 0.03 min^{-1}), *S*-Sp-A (0.071 ± 0.006 min^{-1}), *S*-Sp-G (0.056 ± 0.002 min^{-1}), *R*-Sp-A (0.196 ± 0.003 min^{-1}), and *R*-Sp-G (0.12 ± 0.01 min^{-1}) base pairs. The k_{cat} values for templates containing different products of guanine oxidation follow the same rule as reactivities determined for the nucleotide insertion assays: OG>Gh>*R*-Sp>*S*-Sp (Figure 3.5 and Table 3.2). Interestingly, in all cases, the k_{cat} values for insertion of A were higher than for the insertion of C.

For polymerases, the Michaelis-Menten constant K_M indicates how well an enzyme utilizes the substrate, but it cannot be directly linked to the dNTP dissociation constant.[64,

[84] The calculated K_M values for insertion of C opposite G were more than 3 orders of magnitude higher than for the oxidized lesions showing that SuperScript III has a strong preference for this conventional Watson-Crick base pair. When we compared K_M values for insertion of A, C, or G opposite each product of guanine oxidation, in the case of OG we observed a significantly higher (2-fold) value for the OG-C base pair than for the OG-A base pair, while for Gh, *S*-Sp, and *R*-Sp, substantially higher values for Gh-A (6-fold), *S*-Sp-A (3-fold), and *R*-Sp-A (5-fold) were observed compared to those of the base pairs with G. From the k_{cat} and K_M values measured, we could calculate their ratio, k_{cat}/K_M, the catalytic efficiency of the enzyme that is proportional to a rate of dNTP insertion when the concentration of dNTP approaches 0. Catalytic efficiency is commonly used to determine enzyme preference for one dNTP over another as a substrate. Based on the determined k_{cat}/K_M values, SuperScript III has a 2.5-fold preference for inserting A over C opposite OG, a 6-fold preference for inserting G over A opposite Gh, a 2.4-fold preference for inserting G over A opposite *S*-Sp, and a 3-fold preference for inserting G over A opposite *R*-Sp.

Having the sequencing for oxidative damage to G in RNA as an ultimate goal, we used the steady-state kinetic parameters to estimate the ratio of nucleotides inserted opposite each studied base. This information would allow us to know if during cDNA synthesis SuperScript III would provide characteristic mutations that could be analyzed in order to locate sites of G oxidation. The results are shown in Table 3.3. The estimated insertion ratios were: for OG 1:1.4 C to A; for Gh 1:1.5 A to G; for both Sp diastereomers 1:1.3 A to G. Overall these data suggest that OG and the hydantoin lesions in RNA can be detected by sequencing cDNA created upon reverse transcription of RNA containing these

products of guanine oxidation. While the three hydantoin lesions Gh, *S*-Sp, and *R*-Sp are likely not to be distinguishable by this method due to the similar sequencing signals observed in these studies, they should be easily separable from G and OG. Furthermore, knowing that a hydantoin is present at a given site in RNA would be a significant advancement in our knowledge of oxidative modification of RNA in cell. For the purpose of sequencing, a higher fraction of A insertion than C opposite OG is preferable because it allows an easier differentiation between unoxidized G and OG that is essential to increase the method sensitivity. Overall, based on the determined kinetic parameters, 3 different sequencing signals are expected in which G provides ~100% C insertion, OG yields A insertion at the modified site with ~60% efficiency, and the hydantoins yield insertion of a mixture of A and G with a modest preference for G.

Extension efficiency studies. After discovering which bases are inserted opposite OG, Gh, *S*-Sp, and *R*-Sp, we wanted to investigate whether the extension efficiency is affected by what nucleotide is inserted opposite the lesions. To test this, we performed primer extension assays using different combinations of dATP, dTTP, and dCTP (for OG) or dGTP (for the hydantoins). The reactions were conducted at twice higher concentrations of enzyme (6 U for OG, 8 U for Gh, 80 U for *S*-Sp, and 40 U for *R*-Sp in 20 µL) and increased concentrations of the triphosphates (100-200 µM for OG and 200-400 µM for Gh and Sp), compared to the nucleotide insertion studies; the reason for these changes was to drive the reaction closer to complete base insertion opposite the lesion site and be able to analyze products of primer extension past the site. The results of these studies are presented in Figure 3.10, Table 3.4, and Figure 3.11. On the gels, lanes 2 (AT) and 4 (CT or GT) correspond to reactions containing one of the triphosphates that can be efficiently

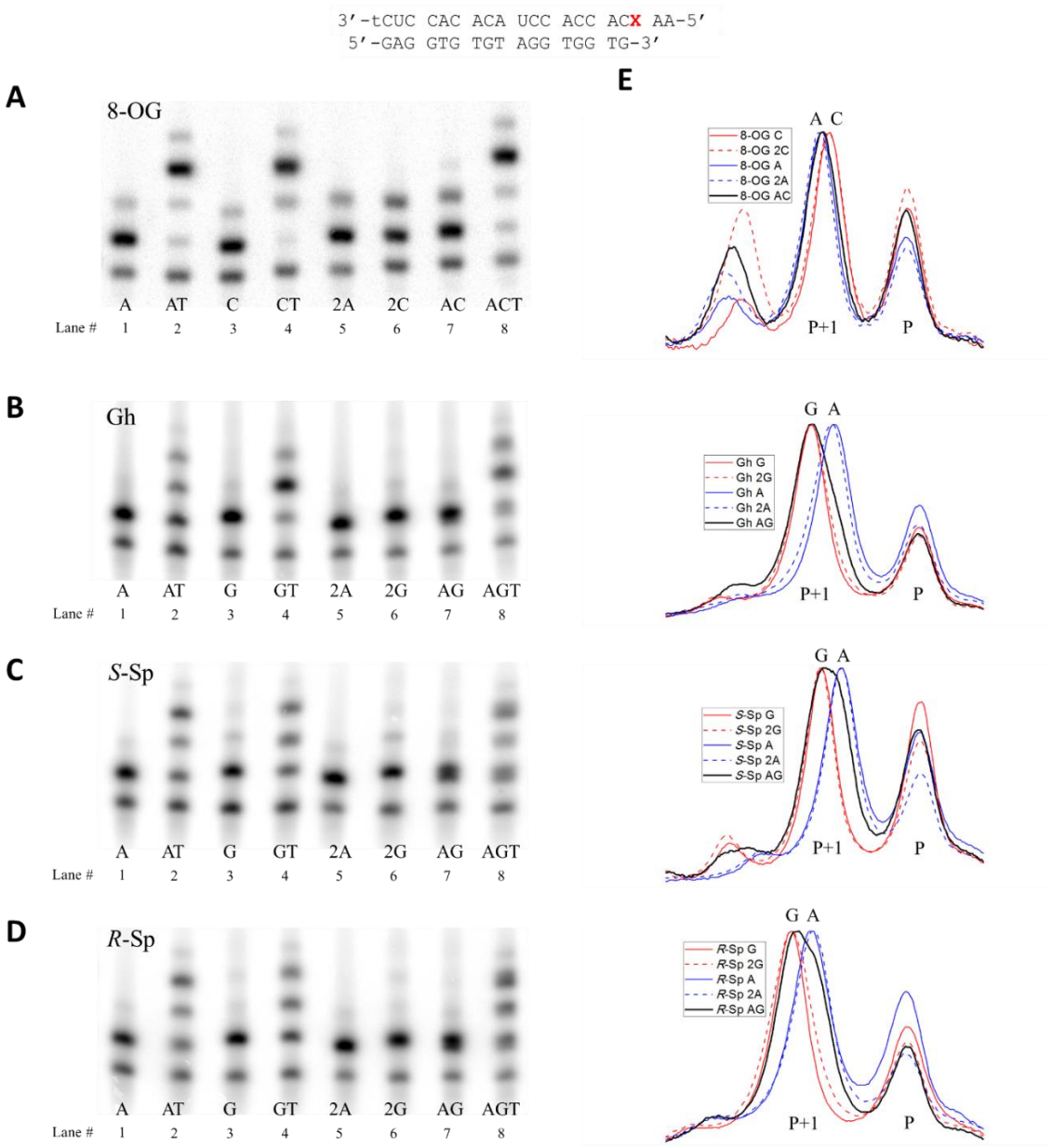


Figure 3.10. Efficiency of extension past different base pairs. **(A-D)** Polyacrylamide gels showing results of primer extension in the presence of one (lanes 1, 3, 5, and 6), two (lanes 2, 4, and 7), or three (lane 8) different triphosphates aimed at comparing bypass efficiencies for base pairs between OG, Gh, *S*-Sp, and *R*-Sp and A, C or G. Samples in lanes 5 and 6 marked 2N contained doubled concentration of corresponding dNTP compared to those in lanes 1 and 3. **(E)** Aligned gel lane plots showing separation between cDNA strands containing different bases inserted opposite OG, Gh, *S*-Sp, and *R*-Sp.

Table 3.4. Efficiency of extension past different base pairs formed by OG (**A**), Gh (**B**), *S*-Sp (**C**), and *R*-Sp (**D**) (analysis of gels from Figure 3.10). Tables to the left show fractions of all bands within one lane, tables to the left show fractions of all bands within one lane not accounting for the unextended primer.

A 8-OG

	A	AT	C	CT	2A	2C	AC	ACT
P	27%	27%	33%	27%	23%	28%	26%	23%
P+1	61%	7%	57%	3%	59%	43%	50%	9%
P+2	12%	7%	11%	14%	18%	29%	21%	10%
P+3		52%		53%			3%	52%
P+4		7%		4%				7%

	A	AT	C	CT	2A	2C	AC	ACT
P	-	-	-	-	-	-	-	-
P+1	84%	9%	84%	4%	76%	60%	67%	11%
P+2	16%	10%	16%	19%	24%	40%	29%	12%
P+3		72%		72%			3%	67%
P+4		9%		5%				9%

B Gh

	A	AT	G	GT	2A	2G	AG	AGT
P	28%	31%	26%	22%	23%	21%	21%	19%
P+1	71%	32%	72%	16%	76%	75%	75%	21%
P+2	1%	19%	2%	48%	1%	3%	3%	39%
P+3		16%		14%		1%	1%	19%
P+4		2%		1%				2%

	A	AT	G	GT	2A	2G	AG	AGT
P	-	-	-	-	-	-	-	-
P+1	98%	46%	97%	20%	98%	95%	95%	26%
P+2	2%	28%	3%	61%	2%	4%	3%	48%
P+3		23%		17%		2%	1%	23%
P+4		3%		1%				3%

C *S*-Sp

	A	AT	G	GT	2A	2G	AG	AGT
P	32%	29%	38%	33%	24%	32%	28%	24%
P+1	65%	22%	52%	22%	73%	56%	66%	32%
P+2	3%	16%	6%	21%	3%	8%	4%	14%
P+3		30%	3%	23%		3%	2%	29%
P+4		3%		1%				1%

	A	AT	G	GT	2A	2G	AG	AGT
P	-	-	-	-	-	-	-	-
P+1	96%	32%	84%	33%	96%	83%	91%	42%
P+2	4%	22%	10%	31%	4%	12%	6%	18%
P+3		43%	6%	34%		5%	3%	38%
P+4		4%		1%				2%

D *R*-Sp

	A	AT	G	GT	2A	2G	AG	AGT
P	34%	24%	31%	26%	19%	22%	19%	16%
P+1	64%	26%	64%	29%	78%	72%	77%	30%
P+2	2%	16%	2%	22%	3%	2%	2%	22%
P+3		31%	3%	21%		5%	2%	31%
P+4		4%		2%				3%

	A	AT	G	GT	2A	2G	AG	AGT
P	-	-	-	-	-	-	-	-
P+1	97%	34%	92%	39%	97%	92%	95%	35%
P+2	3%	21%	3%	30%	3%	2%	3%	26%
P+3		40%	5%	28%		6%	3%	36%
P+4		5%		3%				3%

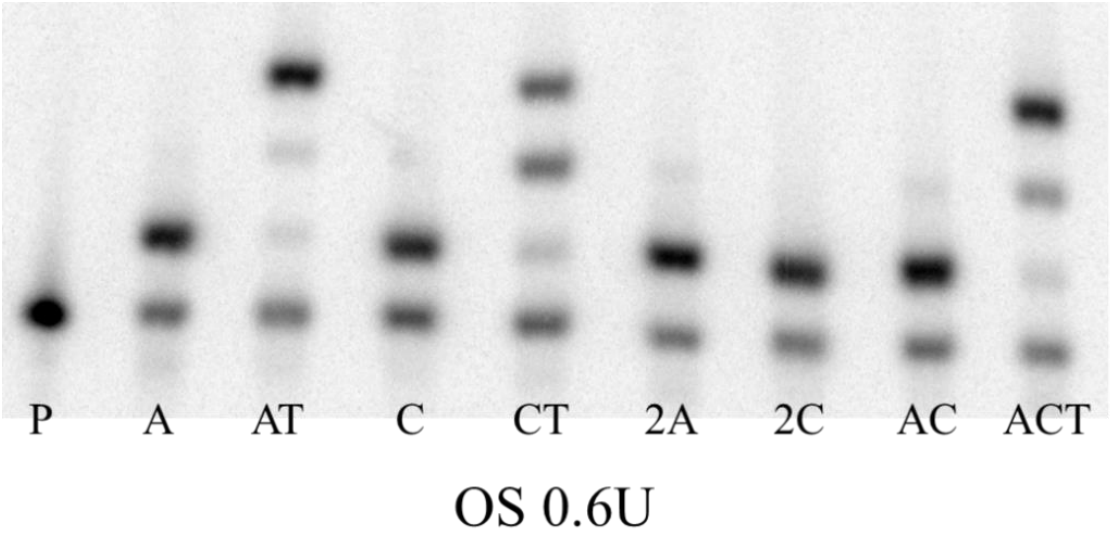


Figure 3.11 . Comparison of extension efficiency by Omniscript RT past OG-A and OG-C base pairs for the OG-1 template.

inserted opposite an oxidized lesion (A and G or C) and dTTP sufficient for full extension of the primer, respectively. Thus, these studies can be used to discover which of the base pairs is more disruptive to the duplex structure and causes more polymerase inhibition. Lanes 1 (A) and 3 (C or G) serve as a control of the basal level of insertion of the first base under the same reaction conditions. Lane 8 (ACT or AGT) corresponds to reaction mixtures containing both triphosphates of the nucleotides that can be inserted opposite OG, Gh, or Sp and dTTP shows how the efficiency of extension past the studied base is affected by the presence of both possible base pairs in the ratio at which they are inserted by reverse transcriptase. Lane 7 serves as a control of the extent to which the first base is inserted under the same conditions. Lanes 5 and 6 serve as additional controls for insertion levels of the first base under twice higher concentrations of triphosphates corresponding to a sum of concentrations of dATP and dCTP or dGTP for lane 7. If the efficiency of insertion of the first base in lane 7 is higher than in lanes 1 and 3, lanes 5 and 6 should show whether it is achievable simply by matching total dNTP concentrations.

From the gels, it is evident that there is almost no difference in the level of polymerase blocking by the OG-A and OG-C base pairs (Figure 3.10 A). Although the presence of an OG-C base pair results in slightly more efficient insertion of the first T (CT lane) directly following C, insertion of the second T is less efficient than in presence of an OG-A base pair (AT lane) leading to a similar amount of full extension product for both base pairs. Expectedly, a mixture of OG-A and OG-C base pairs leads to a very similar level of polymerase blocking (lane 8). Thus, there is practically no difference between OG-A and OG-C base pairs when it comes to how efficiently reverse transcription proceeds after them, meaning that the ratio at which they are inserted opposite OG should match the

A:C ratio in a full-length cDNA.

For the template containing Gh, the presence of a Gh-G base pair (GT lane) leads to significantly more efficient insertion of the first T than of the Gh-A base pair, but the polymerase struggles to insert the next base (Figure 3.10 B and Table 3.4 B). In the case of a Gh-A base pair (AT lane), the most polymerase blockage is observed right after insertion of A. The presence of both dATP and dGTP in the reaction mixture (AGT lane) leads to polymerase arrest primarily after insertion of the T directly following the position of the Gh-N base pair. This hints at predominant insertion of G that aligns with the kinetic data for A and G insertion. Overall, insertion of G and A led to formation of comparable amounts of the fully extended primer, indicating that extension efficiency should not affect the A:G ratio in the fully extended cDNA.

Templates with *R*-Sp or *S*-Sp showed similar behavior with more efficient bypass of Sp-A base pairs in the AT lane than Sp-G base pairs in the GT lane (Figure 3.10 C, D and Table 3.4 C, D). An intermediate amount of polymerase blockage was observed for the case when both dATP and dGTP were present (AGT lane) for both Sp diastereomers, indicating that both bases are inserted in this case. Due to a lower amount of polymerase blockage caused by the Sp-A bypass, it is likely that the fraction of this base pair in the fully extended cDNA would be higher than the ~45% estimate from the insertion kinetic studies.

There are two other noteworthy details about the gels in Figure 3.10. First, in all cases apart from the extension product of the expected length (P+3 for lanes 2, 4, and 8; P+1 for lanes 1, 3, 5, 6, and 7), a product overextended by one base was also observed. This is especially prominent in the case of the OG-1 template and can be attributed to strong

template-independent polymerase activity of reverse transcriptase enzymes.[85-86] Second, the P+1 and P+3 bands are significantly broader in lanes 7 and 8 when both triphosphates are present. This is especially prominent for the Sp diastereomers and is caused by the dependence of electrophoretic mobility on nucleotide composition [C(fastest)>A>T>G].[65, 87] Thus, insertion of a mixture of two nucleotides opposite OG, Gh, or Sp leads to a mixture of two cDNA strands with slightly different electrophoretic mobilities resulting in either broadened or split bands on the polyacrylamide gel. This can be used to obtain a rough estimate of the composition of bases inserted opposite each studied lesion when both dNTPs are present. To do so, we generated pixel density plots for lanes 1, 3, 5, 6, and 7 and aligned them to match the position of the peak corresponding to the primer (P) (Figure 3.10 E). It is possible to characterize insertion of A and C or G only qualitatively from these data due to the presence of the overextended peak P+2 and overlap between peaks corresponding to insertion of different bases. What could be said based on these plots is that A is predominantly inserted opposite OG, G is predominantly inserted opposite Gh, an almost 1:1 ratio of A:G is inserted opposite *S*-Sp, and a slightly larger fraction of G than A is inserted opposite the *R*-Sp isomer. Considering that Sp does not form stable base pairs with any natural nucleotides,[61] insertion of comparable amounts of A and G opposite the hydantoin most likely can be attributed to fitting in the SuperScript III active site due to a similar size and minor groove binding pattern of Sp to C or T,[88-90] rather than to efficient hydrogen-bonding with A or G. The ratio of A to C incorporated opposite OG by the polymerase can significantly deviate from the one calculated based on kinetic parameters in the presence of multiple triphosphates.[64] Thus, we tested whether this was true for SuperScript III using a gel assay (Figures 3.12 and 3.13). Since only qualitative

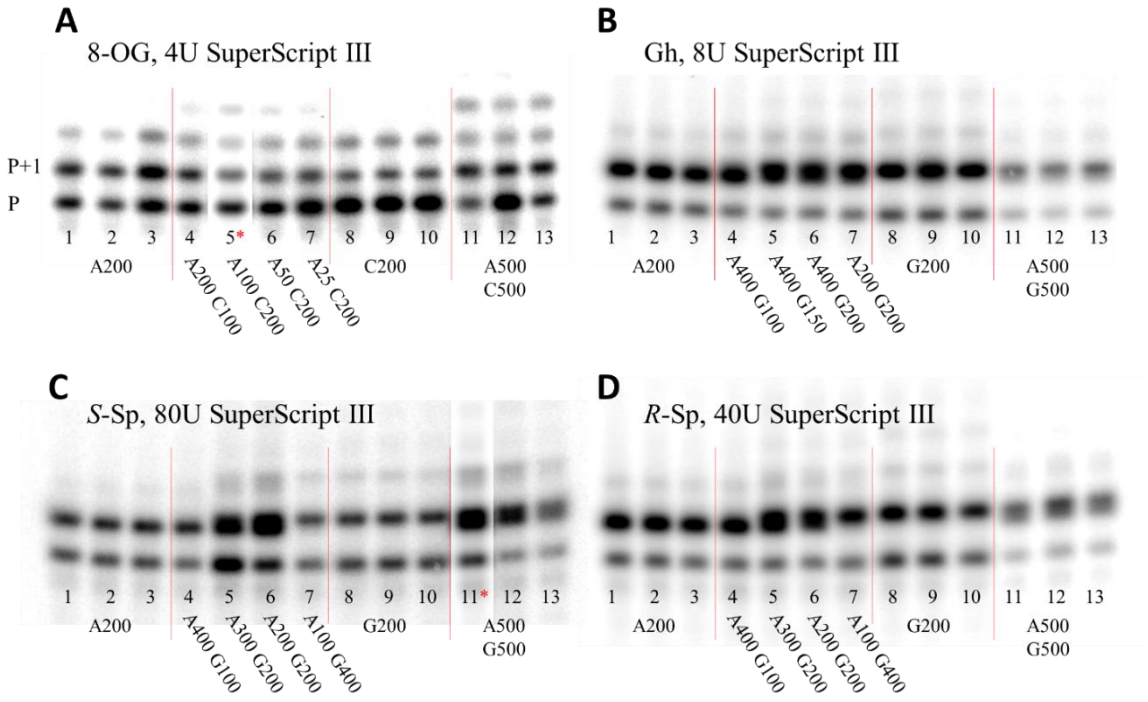


Figure 3.12. Estimation of nucleotide insertion ratio by PAGE opposite 8-OG (**A**), Gh (**B**), *S*-Sp (**C**), or *R*-Sp (**D**) in the template. Brightness for the lanes marked with * was adjusted separately from the rest of the gel. Primer extension was performed as described in materials and methods using indicated amount of SuperScript III reverse transcriptase in 20 µL. On the gels lanes 1-3 correspond to reactions containing 200 µM dATP, lanes 8-10 – 200 µM dCTP or dGTP, lanes 11-13 – 500 µM mixture of dATP and dCTP or dGTP, lanes 4-7 – different ratios of dATP and dCTP or dGTP (indicated on the gel in µM).

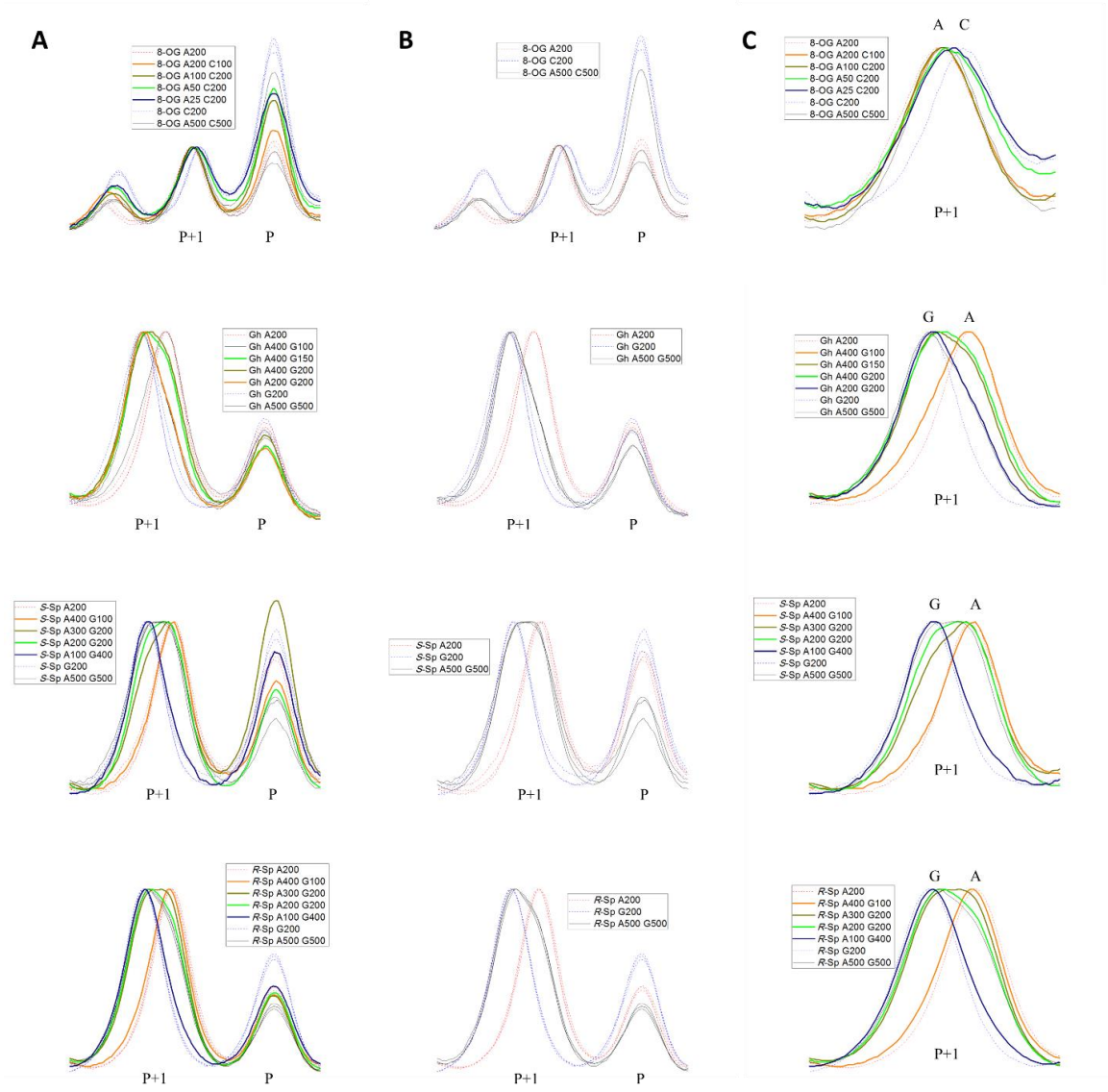


Figure 3.13. Estimation of nucleotide insertion ratio by PAGE, gel lane plots. **(A)** Combined plots of all lanes showing alignment; **(B)** Plots of lanes done in triplicate (1-3, 8-10, and 11-13) showing reproducibility of relative peak positions; **(C)** Plots of P+1 peak showing ratios of insertion of A and C or G at different dNTPs concentrations (only one representative peak is shown for the samples done in triplicate).

characterization of the insertion rate was possible we attempted to adjust concentrations to achieve an ~1:1 insertion ratio and compared it to the one expected from the reaction rates calculated at those concentrations (Table 3.5). In all four cases, insertion of a larger fraction of A over C or G than expected from the k_{cat} and K_M was observed. Expected insertion ratios were within one standard deviation (SD) for Gh and *R*-Sp, two SDs for *S*-Sp, and 4 SDs for OG. Additionally, insertion ratios at 500 μM concentration of competing triphosphates were studied showing ~1:1 insertion of A and G opposite both Sp diastereomers, almost exclusive insertion of A opposite OG, and predominant insertion of G opposite Gh (Figures 3.12 and 3.13).

The other question that we wanted to ask was how efficiently reverse transcriptases can fully extend the primer when a G oxidation product is located in the template strand. For this purpose, we ran RT reactions with RNA templates containing OG, Gh, the Sp diastereomers, or G, using the latter study as a control, under conditions close to those recommended for SuperScript III by the manufacturer of the polymerase (500 μM and 200 μM mixture of all four dNTPs and 100 U of enzyme in 20 μL). Products of the extension reactions were resolved on polyacrylamide gels (Figure 3.14). As expected, RNA templates containing G gave the largest amount of the full extension product (~80% with 500 μM dNTPs), while the presence of OG, Gh, or *R*-Sp in the RNA template resulted in a decreased amount of the fully extended primer (~60%), and the *S*-Sp-1 template gave the smallest amount of the full extension product (~40%). Although all tested RNA templates gave the full extension product, it is apparent that OG, Gh, and Sp cause some level of polymerase inhibition even after the primer is extended past their position, most likely due to either the more dynamic nature of the base pairs they form or negative steric interactions

Table 3.5. Comparison of insertion ratios derived from kinetics and gel-mobility assay. "Calculated insertion ratio" – insertion ratio calculated from steady-state kinetic parameters for the given concentrations of dNTPs. "RSD" – relative standard deviation, "SD" – standard deviation, "# of SD to 1:1" – how many standard deviation intervals it would take to make calculated insertion ratio 1:1.

		[dNTP], μM	Actual insertion ratio	Calculated insertion ratio	RSD*	SD	# of SD to 1:1
OG	dCTP	200	1	2.5	19%	0.48	4
	dATP	25	1	1			
Gh	dGTP	200	1	1.4	32%	0.46	1
	dATP	400	1	1			
S-Sp	dGTP	200	1	1.6	26%	0.42	2
	dATP	200	1	1			
R-Sp	dGTP	200	1	1.3	27%	0.5	1
	dATP	300	1	1			

*-calculated based on propagation of error for $k_{cat}*K_M'/(k_{cat}'*K_M)$

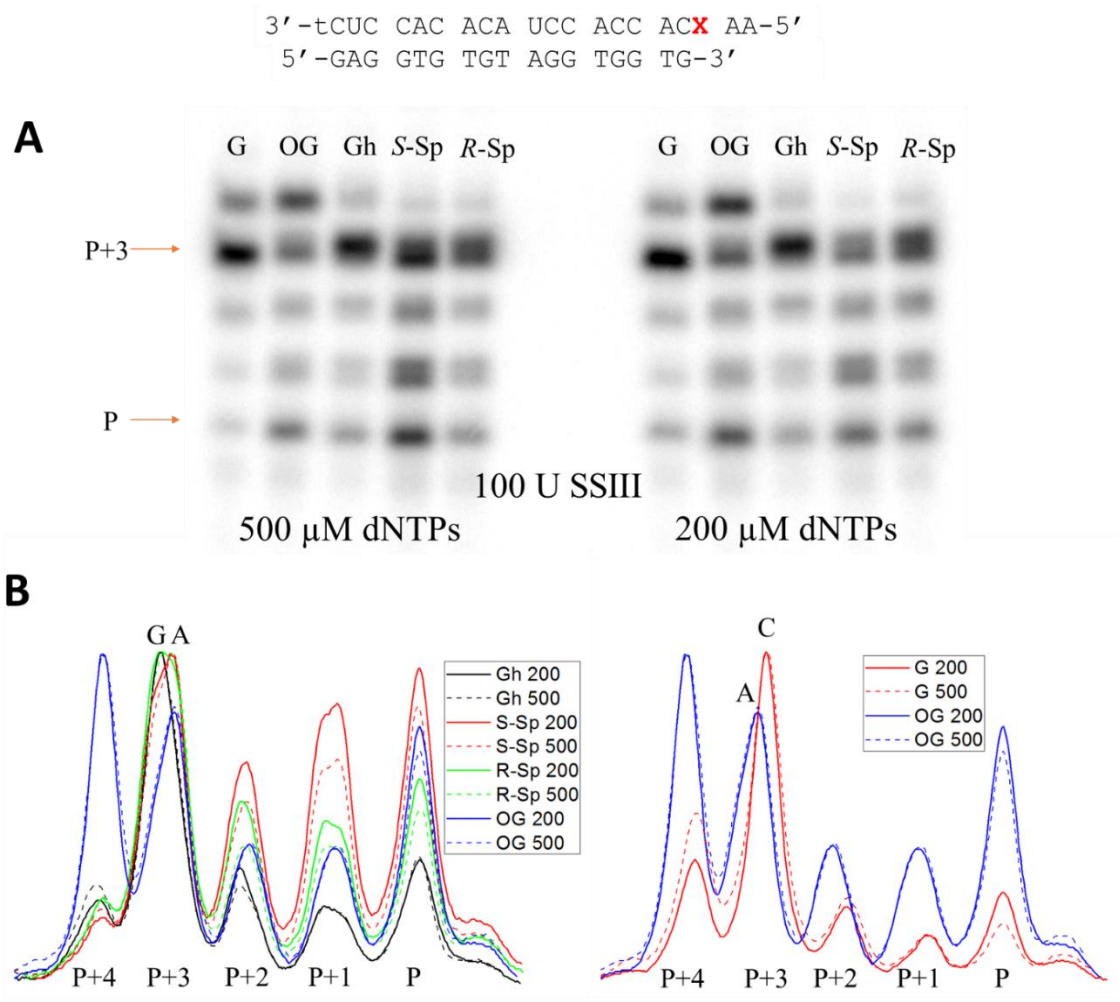


Figure 3.14. Efficiency of complete primer extension. **(A)** Polyacrylamide gel showing the efficiency of synthesis of full-length cDNA based on the templates containing G, OG, Gh, *S*-Sp, or *R*-Sp at two different concentrations of the dNTP mixture. (**B)** Aligned gel lane plots for different templates for reaction mixtures containing 500 and 200 µM of each dNTP.

with the flanking bases.[91-92] Interestingly, the highest level of inhibition was caused by *S*-Sp which was also found to be the most blocking in the polymerase nucleotide insertion studies (Figure 3.5). Analysis similar to the one presented in Figure 3.10 E was performed on fully extended primers (Figure 3.14 B). Results for OG-1 (predominant insertion of A) and Gh-1 (predominant insertion of G) templates were comparable to those obtained for the insertion ratios of the first nucleotides in the presence of multiple dNTPs (Figures 3.10, 3.12, and 3.13). For both diastereomers of Sp, we observed a minor increase in the fraction of fully extended primer containing A inserted opposite Sp compared to the primer extended by just one nucleotide. This led to full-length cDNA containing an ~1:1 ratio of A to G for the *R*-Sp-1 template and a larger fraction of A for the *S*-Sp template. Most likely this result was caused by a higher efficiency of bypassing the Sp-A base pair than the Sp-G base pair by the SuperScript III, as highlighted above.

Conclusions

In this study, we investigated the behavior of reverse transcriptases when they encounter products of guanine oxidation, OG, Gh, *S*-Sp, and *R*-Sp, in an RNA template under standing start conditions. Insertion of A or C opposite OG and A or G opposite the hydantoin lesions was observed that correlated with the previous reports for DNA containing these bases.[37, 61, 64] We also reported steady-state kinetic parameters for single nucleotide insertion opposite the studied lesions in the 5'-AXC-3' sequence context by the SuperScript III reverse transcriptase. Results of this assay indicated a preference for insertion of A opposite OG and G opposite the Gh and Sp diastereomers. We discovered that all these lesions can be bypassed by the SuperScript III under the standard conditions

used for RNA sequencing with sufficient efficiency to yield full-length cDNAs containing A or C inserted opposite OG with predominance of A, A or G inserted opposite Gh with predominance of G, A or G inserted opposite *S*-Sp with minor preference for A, and an equal amount of A or G inserted opposite *R*-Sp. These data indicate that mapping positions of OG, Gh, and Sp in RNA is potentially achievable with the use of reverse transcription to induce characteristic mutations upon cDNA synthesis. Sequencing of the cDNA strand should allow separate detection of unoxidized G as a nonmutagenic C, OG as a C to A mutation (G to T in the complementary stand), Gh as a C to G mutation (G to C in the complementary stand), and the Sp diastereomers as an ~1:1 mixture of C to A and C to G mutations (G to T and G to C in the complementary strand). Post-transcriptional modifications of G and I that disrupt their base pairing preference for C (m^1G, m^2G, $m^2{}_2G$, and m^1I) give different sequencing signals (preferential insertion of T[93-94]) and could be removed with dealkylating enzymes.[67, 95] Therefore, they should not interfere with the reverse transcription results for the oxidized guanine lesions. Relatively efficient bypass of all lesions studied also permits amplification of cDNA via PCR (RT-PCR) which allows interrogation of much smaller quantities of RNA.

References

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

2. Loft, S.; Poulsen, H. E., Cancer risk and oxidative DNA damage in man. *J. Mol. Med.* **1996,** *74* (6), 297-312.

3. De Bont, R.; van Larebeke, N., Endogenous DNA damage in humans: a review of quantitative data. *Mutagenesis* **2004,** *19* (3), 169-185.

4. Wogan, G. N.; Hecht, S. S.; Felton, J. S.; Conney, A. H.; Loeb, L. A., Environmental and chemical carcinogenesis. *Semin. Cancer Biol.* **2004,** *14* (6), 473-486.

5. Noll, D. M.; Mason, T. M.; Miller, P. S., Formation and repair of interstrand cross-links in DNA. *Chem. Rev.* **2006,** *106* (2), 277-301.

6. Oleinick, N. L.; Chiu, S. M.; Ramakrishnan, N.; Xue, L. Y., The formation, identification, and significance of DNA-protein cross-links in mammalian cells. *Br. J. Cancer Suppl.* **1987,** *8*, 135-140.

7. Johansen, M. E.; Muller, J. G.; Xu, X.; Burrows, C. J., Oxidatively induced DNA-protein cross-linking between single-stranded binding protein and oligodeoxynucleotides containing 8-oxo-7,8-dihydro-2'-deoxyguanosine. *Biochemistry* **2005,** *44* (15), 5660-5671.

8. Kundu, L. M.; Linne, U.; Marahiel, M.; Carell, T., RNA is more UV resistant than DNA: the formation of UV-induced DNA lesions is strongly sequence and conformation dependent. *Chem. Eur. J.* **2004,** *10* (22), 5697-5705.

9. Kong, Q.; Lin, C. L., Oxidative damage to RNA: mechanisms, consequences, and diseases. *Cell. Mol. Life Sci.* **2010,** *67* (11), 1817-1829.

10. Ames, B. N.; Shigenaga, M. K.; Hagen, T. M., Oxidants, antioxidants, and the degenerative diseases of aging. *Proc. Natl. Acad. Sci. U. S. A.* **1993,** *90* (17), 7915-7922.

11. Finkel, T.; Holbrook, N. J., Oxidants, oxidative stress and the biology of ageing. *Nature* **2000,** *408* (6809), 239-247.

12. Cooke, M. S.; Evans, M. D.; Dizdaroglu, M.; Lunec, J., Oxidative DNA damage: mechanisms, mutation, and disease. *FASEB J.* **2003,** *17* (10), 1195-1214.

13. Li, Z.; Wu, J.; Deleo, C. J., RNA damage and surveillance under oxidative stress. *IUBMB Life* **2006,** *58* (10), 581-588.

14. Bernstein, J. A.; Khodursky, A. B.; Lin, P. H.; Lin-Chao, S.; Cohen, S. N., Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proc. Natl. Acad. Sci. U. S. A.* **2002,** *99* (15), 9697-9702.

15. Defoiche, J.; Zhang, Y.; Lagneaux, L.; Pettengell, R.; Hegedus, A.; Willems, L.; Macallan, D. C., Measurement of ribosomal RNA turnover *in vivo* by use of deuterium-labeled glucose. *Clin. Chem.* **2009,** *55* (10), 1824-1833.

16. Schmidt, E. E.; Schibler, U., Cell size regulation, a mechanism that controls cellular RNA accumulation: consequences on regulation of the ubiquitous transcription factors Oct1 and NF-Y and the liver-enriched transcription factor DBP. *J. Cell Biol.* **1995,** *128* (4), 467-483.

17. Hofer, T.; Badouard, C.; Bajak, E.; Ravanat, J. L.; Mattsson, A.; Cotgreave, I. A., Hydrogen peroxide causes greater oxidation in cellular RNA than in DNA. *Biol. Chem.* **2005,** *386* (4), 333-337.

18. Mangerich, A.; Knutson, C. G.; Parry, N. M.; Muthupalani, S.; Ye, W.; Prestwich, E.; Cui, L.; McFaline, J. L.; Mobley, M.; Ge, Z.; Taghizadeh, K.; Wishnok, J. S.; Wogan, G. N.; Fox, J. G.; Tannenbaum, S. R.; Dedon, P. C., Infection-induced colitis in mice causes dynamic and tissue-specific changes in stress response and DNA damage leading to colon cancer. *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109* (27), E1820-E1829.

19. Liu, M.; Gong, X.; Alluri, R. K.; Wu, J.; Sablo, T.; Li, Z., Characterization of RNA damage under oxidative stress in *Escherichia coli*. *Biol. Chem.* **2012,** *393* (3), 123-132.

20. Saikia, M.; Jobava, R.; Parisien, M.; Putnam, A.; Krokowski, D.; Gao, X. H.; Guan, B. J.; Yuan, Y.; Jankowsky, E.; Feng, Z.; Hu, G. F.; Pusztai-Carey, M.; Gorla, M.; Sepuri, N. B.; Pan, T.; Hatzoglou, M., Angiogenin-cleaved tRNA halves interact with cytochrome c, protecting cells from apoptosis during osmotic stress. *Mol. Cell. Biol.* **2014,** *34* (13), 2450-2463.

21. Mishima, E.; Inoue, C.; Saigusa, D.; Inoue, R.; Ito, K.; Suzuki, Y.; Jinno, D.; Tsukui, Y.; Akamatsu, Y.; Araki, M.; Araki, K.; Shimizu, R.; Shinke, H.; Suzuki, T.; Takeuchi, Y.; Shima, H.; Akiyama, Y.; Toyohara, T.; Suzuki, C.; Saiki, Y.; Tominaga, T.; Miyagi, S.; Kawagisihi, N.; Soga, T.; Ohkubo, T.; Yamamura, K.; Imai, Y.; Masuda, S.; Sabbisetti, V.; Ichimura, T.; Mount, D. B.; Bonventre, J. V.; Ito, S.; Tomioka, Y.; Itoh, K.; Abe, T., Conformational change in transfer RNA is an early indicator of acute cellular damage. *J. Am. Soc. Nephrol.* **2014**, 2316-2326.

22. Mleczko, A. M.; Celichowski, P.; Bakowska-Zywicka, K., Ex-translational function of tRNAs and their fragments in cancer. *Acta Biochim. Pol.* **2014,** *61* (2), 211-216.

23. Lee, Y. S.; Shibata, Y.; Malhotra, A.; Dutta, A., A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). *Genes Dev.* **2009,** *23* (22), 2639-2649.

24. Thompson, D. M.; Lu, C.; Green, P. J.; Parker, R., tRNA cleavage is a conserved response to oxidative stress in eukaryotes. *RNA* **2008,** *14* (10), 2095-2103.

25. Nawrot, B.; Sochacka, E.; Duchler, M., tRNA structural and functional changes induced by oxidative stress. *Cell. Mol. Life Sci.* **2011,** *68* (24), 4023-4032.

26. Moreira, P. I.; Nunomura, A.; Nakamura, M.; Takeda, A.; Shenk, J. C.; Aliev, G.; Smith, M. A.; Perry, G., Nucleic acid oxidation in Alzheimer disease. *Free Radic. Biol. Med.* **2008,** *44* (8), 1493-1505.

27. Poulsen, H. E.; Specht, E.; Broedbaek, K.; Henriksen, T.; Ellervik, C.; Mandrup-Poulsen, T.; Tonnesen, M.; Nielsen, P. E.; Andersen, H. U.; Weimann, A., RNA modifications by oxidation: a novel disease mechanism? *Free Radic. Biol. Med.* **2012,** *52* (8), 1353-1361.

28. Castellani, R. J.; Nunomura, A.; Rolston, R. K.; Moreira, P. I.; Takeda, A.; Perry, G.; Smith, M. A., Sublethal RNA oxidation as a mechanism for neurodegenerative disease. *Int. J. Mol. Sci.* **2008,** *9* (5), 789-806.

29. Che, Y.; Wang, J. F.; Shao, L.; Young, T., Oxidative damage to RNA but not DNA in the hippocampus of patients with major mental illness. *J. Psychiatry Neurosci.* **2010,** *35* (5), 296-302.

30. Fujikawa, K.; Kamiya, H.; Yakushiji, H.; Nakabeppu, Y.; Kasai, H., Human MTH1 protein hydrolyzes the oxidized ribonucleotide, 2-hydroxy-ATP. *Nucleic Acids Res.* **2001,** *29* (2), 449-454.

31. Wu, J.; Li, Z., Human polynucleotide phosphorylase reduces oxidative RNA damage and protects HeLa cell against oxidative stress. *Biochem. Biophys. Res. Commun.* **2008,** *372* (2), 288-292.

32. Ishibashi, T.; Hayakawa, H.; Ito, R.; Miyazawa, M.; Yamagata, Y.; Sekiguchi, M., Mammalian enzymes for preventing transcriptional errors caused by oxidative damage. *Nucleic Acids Res.* **2005,** *33* (12), 3779-3784.

33. Bellacosa, A.; Moss, E. G., RNA repair: damage control. *Curr. Biol.* **2003,** *13* (12), R482-R484.

34. Wurtmann, E. J.; Wolin, S. L., RNA under attack: cellular handling of RNA damage. *Crit. Rev. Biochem. Mol. Biol.* **2009,** *44* (1), 34-49.

35. Steenken, S.; Jovanovic, S. V.; Bietti, M.; Bernhard, K., The trap depth (in DNA) of 8-oxo-7,8-dihydro-2'deoxyguanosine as derived from electron-transfer equilibria in aqueous solution. *J. Am. Chem. Soc.* **2000,** *122* (10), 2373-2374.

36. Neeley, W. L.; Essigmann, J. M., Mechanisms of formation, genotoxicity, and mutation of guanine oxidation products. *Chem. Res. Toxicol.* **2006,** *19* (4), 491-505.

37. Henderson, P. T.; Delaney, J. C.; Muller, J. G.; Neeley, W. L.; Tannenbaum, S. R.; Burrows, C. J.; Essigmann, J. M., The hydantoin lesions formed from oxidation of 7,8-dihydro-8-oxoguanine are potent sources of replication errors *in vivo*. *Biochemistry* **2003,** *42* (31), 9257-9262.

38. McKibbin, P. L.; Fleming, A. M.; Towheed, M. A.; Van Houten, B.; Burrows, C. J.; David, S. S., Repair of hydantoin lesions and their amine adducts in DNA by base and nucleotide excision repair. *J. Am. Chem. Soc.* **2013,** *135* (37), 13851-13861.

39. Tomaszewska-Antczak, A.; Guga, P.; Nawrot, B.; Pratviel, G., Guanosine in a single stranded region of anticodon stem-loop tRNA models is prone to oxidatively generated damage resulting in dehydroguanidinohydantoin and spiroiminodihydantoin lesions. *Chem. Eur. J.* **2015,** *21* (17), 6381-6385.

40. Fleming, A. M.; Alshykhly, O.; Zhu, J.; Muller, J. G.; Burrows, C. J., Rates of chemical cleavage of DNA and RNA oligomers containing guanine oxidation products. *Chem. Res. Toxicol.* **2015,** *28* (6), 1292-1300.

41. Fleming, A. M.; Orendt, A. M.; He, Y.; Zhu, J.; Dukor, R. K.; Burrows, C. J., Reconciliation of chemical, enzymatic, spectroscopic and computational data to assign the absolute configuration of the DNA base lesion spiroiminodihydantoin. *J. Am. Chem. Soc.* **2013,** *135* (48), 18191-18204.

42. Delaney, S.; Jarem, D. A.; Volle, C. B.; Yennie, C. J., Chemical and biological consequences of oxidatively damaged guanine in DNA. *Free Radic. Res.* **2012,** *46* (4), 420-441.

43. Pratviel, G.; Meunier, B., Guanine oxidation: one- and two-electron reactions. *Chem. Eur. J.* **2006,** *12* (23), 6018-6030.

44. Gimisis, T.; Cismas, C., Isolation, characterization, and independent synthesis of guanine oxidation products. *Eur. J. Org. Chem.* **2006,** (6), 1351-1378.

45. Cadet, J.; Wagner, J. R.; Shafirovich, V.; Geacintov, N. E., One-electron oxidation reactions of purine and pyrimidine bases in cellular DNA. *Int. J. Radiat. Biol.* **2014,** *90* (6), 423-432.

46. Burrows, C. J.; Muller, J. G., Oxidative nucleobase modifications leading to strand scission. *Chem. Rev.* **1998,** *98* (3), 1109-1152.

47. Wells, S. E.; Hughes, J. M. X.; Igel, A. H.; Ares, M., Use of dimethyl sulfate to probe RNA structure *in vivo*. *Methods Enzymol.* **2000,** *318*, 479-493.

48. Ross, R. L.; Cao, X.; Limbach, P. A., Mapping post-transcriptional modifications onto transfer ribonucleic acid sequences by liquid chromatography tandem mass spectrometry. *Biomolecules* **2017,** *7* (1), E21.

49. Wagner, T. M.; Nair, V.; Guymon, R.; Pomerantz, S. C.; Crain, P. F.; Davis, D. R.; McCloskey, J. A., A novel method for sequence placement of modified nucleotides in mixtures of transfer RNA. *Nucleic Acids Symp. Ser. (Oxf)* **2004,** (48), 263-264.

50. Kowalak, J. A.; Pomerantz, S. C.; Crain, P. F.; McCloskey, J. A., A novel method for the determination of post-transcriptional modification in RNA by mass spectrometry. *Nucleic Acids Res.* **1993,** *21* (19), 4577-4585.

51. Guymon, R.; Pomerantz, S. C.; Ison, J. N.; Crain, P. F.; McCloskey, J. A., Post-transcriptional modifications in the small subunit ribosomal RNA from *Thermotoga maritima*, including presence of a novel modified cytidine. *RNA* **2007,** *13* (3), 396-403.

52. Siegfried, N. A.; Busan, S.; Rice, G. M.; Nelson, J. A.; Weeks, K. M., RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat. Methods* **2014,** *11* (9), 959-965.

53. von Watzdorf, J.; Marx, A., 6-Substituted 2-aminopurine-2'-deoxyribonucleoside 5'-triphosphates that trace cytosine methylation. *ChemBioChem* **2016,** *17* (16), 1532-1540.

54. von Watzdorf, J.; Leitner, K.; Marx, A., Modified nucleotides for discrimination between cytosine and the epigenetic marker 5-methylcytosine. *Angew. Chem. Int. Ed.* **2016,** *55* (9), 3229-3232.

55. Wyss, L. A.; Nilforoushan, A.; Eichenseher, F.; Suter, U.; Blatter, N.; Marx, A.; Sturla, S. J., Specific incorporation of an artificial nucleotide opposite a mutagenic DNA adduct by a DNA polymerase. *J. Am. Chem. Soc.* **2015,** *137* (1), 30-33.

56. Gahlon, H. L.; Schweizer, W. B.; Sturla, S. J., Tolerance of base pair size and shape in postlesion DNA synthesis. *J. Am. Chem. Soc.* **2013,** *135* (17), 6384-6387.

57. Merino, E. J.; Wilkinson, K. A.; Coughlan, J. L.; Weeks, K. M., RNA structure analysis at single nucleotide resolution by selective 2 '-hydroxyl acylation and primer extension (SHAPE). *J. Am. Chem. Soc.* **2005,** *127* (12), 4223-4231.

58. Loughrey, D.; Watters, K. E.; Settle, A. H.; Lucks, J. B., SHAPE-Seq 2.0: systematic optimization and extension of high-throughput chemical probing of RNA secondary structure with next generation sequencing. *Nucleic Acids Res.* **2014,** *42* (21), e165.

59. Dahlmann, H. A.; Vaidyanathan, V. G.; Sturla, S. J., Investigating the biochemical impact of DNA damage with structure-based probes: abasic sites, photodimers, alkylation adducts, and oxidative lesions. *Biochemistry* **2009,** *48* (40), 9347-9359.

60. Kornyushyna, O.; Burrows, C. J., Effect of the oxidized guanosine lesions spiroiminodihydantoin and guanidinohydantoin on proofreading by *Escherichia coli* DNA polymerase I (Klenow fragment) in different sequence contexts. *Biochemistry* **2003,** *42* (44), 13008-13018.

61. Kornyushyna, O.; Berges, A. M.; Muller, J. G.; Burrows, C. J., *In vitro* nucleotide misinsertion opposite the oxidized guanosine lesions spiroiminodihydantoin and guanidinohydantoin and DNA synthesis past the lesions using *Escherichia coli* DNA polymerase I (Klenow fragment). *Biochemistry* **2002,** *41* (51), 15304-15314.

62. Furge, L. L.; Guengerich, F. P., Analysis of nucleotide insertion and extension at 8-oxo-7,8-dihydroguanine by replicative T7 polymerase exo(-) and human immunodeficiency virus-1 reverse transcriptase using steady-state and pre-steady-state kinetics. *Biochemistry* **1997,** *36* (21), 6475-6487.

63. Lowe, L. G.; Guengerich, F. P., Steady-state and pre-steady-state kinetic analysis of dNTP insertion opposite 8-oxo-7,8-dihydroguanine by *Escherichia coli* polymerases I exo- and II exo. *Biochemistry* **1996,** *35* (30), 9840-9849.

64. Xue, Q.; Zhong, M.; Liu, B.; Tang, Y.; Wei, Z.; Guengerich, F. P.; Zhang, H., Kinetic analysis of bypass of 7,8-dihydro-8-oxo-2'-deoxyguanosine by the catalytic core of yeast DNA polymerase eta. *Biochimie* **2016,** *121*, 161-169.

65. Hsu, G. W.; Ober, M.; Carell, T.; Beese, L. S., Error-prone replication of oxidatively damaged DNA by a high-fidelity DNA polymerase. *Nature* **2004,** *431* (7005), 217-221.

66. Cantara, W. A.; Crain, P. F.; Rozenski, J.; McCloskey, J. A.; Harris, K. A.; Zhang, X.; Vendeix, F. A.; Fabris, D.; Agris, P. F., The RNA modification database, RNAMDB: 2011 update. *Nucleic Acids Res.* **2011,** *39* (Database issue), D195-201.

67. Zheng, G.; Qin, Y.; Clark, W. C.; Dai, Q.; Yi, C.; He, C.; Lambowitz, A. M.; Pan, T., Efficient and quantitative high-throughput tRNA sequencing. *Nat. Methods* **2015,** *12* (9), 835-837.

68. Ye, Y.; Muller, J. G.; Luo, W.; Mayne, C. L.; Shallop, A. J.; Jones, R. A.; Burrows, C. J., Formation of ^{13}C-, ^{15}N-, and ^{18}O-labeled guanidinohydantoin from guanosine oxidation with singlet oxygen. Implications for structure and mechanism. *J. Am. Ceram. Soc.* **2003,** *125* (46), 13926-13927.

69. Fleming, A. M.; Orendt, A. M.; He, Y.; Zhu, J.; Dukor, R. K.; Burrows, C. J., Reconciliation of chemical, enzymatic, spectroscopic and computational data to assign the absolute configuration of the DNA base lesion spiroiminodihydantoin. *J. Am. Chem. Soc.* **2013,** *135* (48), 18191-18204.

70. Chen, X.; Fleming, A. M.; Muller, J. G.; Burrows, C. J., Endonuclease and exonuclease activities on oligodeoxynucleotides containing spiroiminodihydantoin depend on the sequence context and the lesion stereochemistry. *New J. Chem.* **2013,** *37* (11), 3440-3449.

71. Fleming, A. M.; Muller, J. G.; Dlouhy, A. C.; Burrows, C. J., Structural context effects in the oxidation of 8-oxo-7,8-dihydro-2'-deoxyguanosine to hydantoin products: electrostatics, base stacking, and base pairing. *J. Am. Chem. Soc.* **2012,** *134* (36), 15091-15102.

72. Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W., NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **2012,** *9* (7), 671-675.

73. Schindelin, J.; Rueden, C. T.; Hiner, M. C.; Eliceiri, K. W., The ImageJ ecosystem: an open platform for biomedical image analysis. *Mol. Reprod. Dev.* **2015,** *82* (7-8), 518-529.

74. Morales, J. C.; Kool, E. T., Importance of terminal base pair hydrogen-bonding in 3'-end proofreading by the Klenow fragment of DNA polymerase I. *Biochemistry* **2000,** *39* (10), 2626-2632.

75. Boutabout, M.; Wilhelm, M.; Wilhelm, F. X., DNA synthesis fidelity by the reverse transcriptase of the yeast retrotransposon Ty1. *Nucleic Acids Res.* **2001,** *29* (11), 2217-2222.

76. Mohr, S.; Ghanem, E.; Smith, W.; Sheeter, D.; Qin, Y.; King, O.; Polioudakis, D.; Iyer, V. R.; Hunicke-Smith, S.; Swamy, S.; Kuersten, S.; Lambowitz, A. M., Thermostable group II intron reverse transcriptase fusion proteins and their use in cDNA synthesis and next-generation RNA sequencing. *RNA* **2013,** *19* (7), 958-970.

77. Stahlberg, A.; Kubista, M.; Pfaffl, M., Comparison of reverse transcriptases in gene expression analysis. *Clin. Chem.* **2004,** *50* (9), 1678-1680.

78. Okello, J. B. A.; Rodriguez, L.; Poinar, D.; Bos, K.; Okwi, A. L.; Bimenya, G. S.; Sewankambo, N. K.; Henry, K. R.; Kuch, M.; Poinar, H. N., Quantitative assessment of the sensitivity of various commercial reverse transcriptases based on armored HIV RNA. *PLoS One* **2010,** *5* (11), e13931.

79. Di Giallonardo, F.; Zagordi, O.; Duport, Y.; Leemann, C.; Joos, B.; Kunzli-Gontarczyk, M.; Bruggmann, R.; Beerenwinkel, N.; Gunthard, H. F.; Metzner, K. J., Next-generation sequencing of HIV-1 RNA genomes: determination of error rates and minimizing artificial recombination. *PLoS One* **2013,** *8* (9), e74249.

80. He, L.; Sok, D.; Azadnia, P.; Hsueh, J.; Landais, E.; Simek, M.; Koff, W. C.; Poignard, P.; Burton, D. R.; Zhu, J., Toward a more accurate view of human B-cell repertoire by next-generation sequencing, unbiased repertoire capture and single-molecule barcoding. *Sci. Rep.* **2014,** *4*, 6778.

81. Jia, L.; Shafirovich, V.; Shapiro, R.; Geacintov, N. E.; Broyde, S., Structural and thermodynamic features of spiroiminodihydantoin damaged DNA duplexes. *Biochemistry* **2005,** *44* (40), 13342-13353.

82. Zhao, X.; Muller, J. G.; Halasyam, M.; David, S. S.; Burrows, C. J., *In vitro* ligation of oligodeoxynucleotides containing C8-oxidized purine lesions using bacteriophage T4 DNA ligase. *Biochemistry* **2007,** *46* (12), 3734-3744.

83. Gardner, A. F.; Joyce, C. M.; Jack, W. E., Comparative kinetics of nucleotide analog incorporation by vent DNA polymerase. *J. Biol. Chem.* **2004,** *279* (12), 11834-11842.

84. Einolf, H. J.; Guengerich, F. P., Kinetic analysis of nucleotide incorporation by mammalian DNA polymerase delta. *J. Biol. Chem.* **2000,** *275* (21), 16316-16322.

85. Ohtsubo, Y.; Nagata, Y.; Tsuda, M., Efficient N-tailing of blunt DNA ends by moloney murine leukemia virus reverse transcriptase. *Sci. Rep.* **2017,** *7*, 41769.

86. Bibillo, A.; Eickbush, T. H., End-to-end template jumping by the reverse transcriptase encoded by the R2 retrotransposon. *J. Biol. Chem.* **2004,** *279* (15), 14945-14953.

87. Frank, R.; Koster, H., DNA chain length markers and the influence of base composition on electrophoretic mobility of oligodeoxyribonucleotides in polyacrylamide-gels. *Nucleic Acids Res.* **1979,** *6* (6), 2069-2087.

88. Hottin, A.; Marx, A., Structural insights into the processing of nucleobase-modified nucleotides by DNA polymerases. *Acc. Chem. Res.* **2016,** *49* (3), 418-427.

89. Ludmann, S.; Marx, A., Getting it right: how DNA polymerases select the right nucleotide. *Chimia* **2016,** *70* (3), 203-206.

90. Marx, A.; Summerer, D., Molecular insights into error-prone DNA replication and error-free lesion bypass. *ChemBioChem* **2002,** *3* (5), 405-407.

91. Ye, Y. From nucleosides and nucleotides to DNA: synthesis, enzymatic study and structural analysis of oxidized lesions beyond 8-oxo-purine. Ph.D. dissertation. University of Utah, Salt Lake City, UT, 2007.

92. Khutsishvili, I.; Zhang, N.; Marky, L. A.; Crean, C.; Patel, D. J.; Geacintov, N. E.; Shafirovich, V., Thermodynamic profiles and nuclear magnetic resonance studies of oligonucleotide duplexes containing single diastereomeric spiroiminodihydantoin lesions. *Biochemistry* **2013,** *52* (8), 1354-1363.

93. Delaney, J. C.; Essigmann, J. M., Mutagenesis, genotoxicity, and repair of 1-methyladenine, 3-alkylcytosines, 1-methylguanine, and 3-methylthymine in AlkB *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **2004,** *101* (39), 14051-14056.

94. Ryvkin, P.; Leung, Y. Y.; Silverman, I. M.; Childress, M.; Valladares, O.; Dragomir, I.; Gregory, B. D.; Wang, L. S., HAMR: high-throughput annotation of modified ribonucleotides. *RNA* **2013,** *19* (12), 1684-1692.

95. Dai, Q.; Zheng, G.; Schwartz, M. H.; Clark, W. C.; Pan, T., Selective enzymatic demethylation of N^2,N^2-dimethylguanosine in RNA and its application in high-throughput tRNA sequencing. *Angew. Chem. Int. Ed.* **2017,** *56* (18), 5017-5020.

CHAPTER 4

SEQUENCING OF RNA FOR PRODUCTS OF GUANINE OXIDATION

Introduction

In Chapter 3 we studied which nucleotides are inserted opposite OG and two products of its further oxidation, Gh and Sp, and whether these three lesions can be efficiently bypassed by RT polymerases. Having in mind that the long-term goal of this project is mapping products of guanine oxidation in RNA, in this chapter we focused on identifying the sequencing signatures for OG, Gh, and Sp. A more detailed introduction into this topic was given in Chapter 3.

Materials and methods

Oligomer synthesis. The RNA and DNA oligonucleotides were synthesized by the core facilities at the University of Utah using standard solid-phase synthesis protocols. The RNA templates containing OG were synthesized using the commercially available rOG phosphoramidite (ChemGenes, Wilmington, MA). The RNA strands were synthesized with a 3` dT to maximize the solid-phase synthesis yield of the modified RNA strands; the added dT will not impact cDNA synthesized during reverse transcription. All oligonucleotides were purified via an analytical anion-exchange HPLC and dialyzed against ddH_2O. Oligonucleotides containing Gh and Sp were synthesized by Aaron

Fleming as previously described in Chapter 3. Oligomers containing Sp were purified and studied in this chapter as a mixture of diastereomers because the lengths of the RNA templates were too long resulting in the isomers eluting as a single peak from the HPLC column.

Labeling of the DNA primer. Monitoring the efficiency of reverse transcription and deletion bypass was achieved via polyacrylamide gel electrophoresis (PAGE) wherein the DNA primer was 5' end-labeled with ^{32}P following a procedure adopted from the literature[1] using T4 polynucleotide kinase (New England Biolabs, Ipswich, MA.) and [γ-^{32}P] ATP (PerkinElmer, Waltham, MA.).

Reverse transcription. Before reverse transcription, all samples were annealed by heating 7 µL of aqueous solution containing 1.32 pmol (132 nM in a final volume of 10 µL) of RNA template and 1.2 pmol (120 nM in 10 µL) DNA primer including ~30,000 cpm ^{32}P-labeled strand, 750 nmol KCl (75 mM in 10 µL), and 500 nmol Tris-HCl pH 8.3 buffer (50 mM in 10 µL) to 95 °C for 5 min, and then incubating them at 55 °C for 5 min followed by cooling at 15 °C for another 10 min. Then 3 µL of freshly made stock solution containing 30 nmol $MgCl_2$, 100 nmol DTT, 5 nmol of each dNTP, and 100 units SuperScript III was added to the annealed RNA-DNA duplex to give final concentrations of 3 mM $MgCl_2$, 10 mM DTT, 500 µM of each dNTP (1 mM dATP and 500 µM other dNTPs for OG-A, and Sp-A), and 10 U/µL of SuperScript III (or MMLV RT of the same concentration) in the final volume of 10 µL. Reaction mixtures were incubated for 40 min at room temperature (rt, ~25 °C), 35 °C, 37 °C, 40 °C, 45 °C, or 50 °C. Then to quench the reaction, an equal volume of 2x gel loading buffer (8 M urea, 0.01% xylene cyanole, 0.01% bromphenol blue, x1 TBE buffer) were added and the mixture was heated to 95 °C for 10

min. About 10 μL of the resulting solution were analyzed on 20% denaturing PAGE. Gels were stored for 12-18 h with a storage phosphor screen that was then scanned using a phosphoimager. The resulting images were analyzed using ImageJ2 software.[2-3] All reactions have been performed using C1000 Touch thermal cycler (BioRad, Hercules, CA).

Reverse transcription-polymerase chain reaction. Samples for the reverse transcription have been prepared as described in the previous section on the scale from 10 to 40 μL without addition of ^{32}P-labeled primer. Reverse transcription has been performed at 45 °C for 40 min. Upon completion of the reaction samples were incubated at 93 °C for 10 min to facilitate degradation of RNA template.

For PCR amplification of cDNA, 2 μL of RT reaction mixture were mixed with 48 μL of freshly made stock solution containing the remaining components required for PCR reaction to give final concentrations of 700 nM forward primer (including ~100,000 cpm ^{32}P labeled primer for the samples analyzed by PAGE), 700 nM reverse primer, 400 μM of each dNTP, 1x ThermoPol reaction buffer (20 mM Tris-HCl, 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 2mM $MgSO_4$, 0.1% Triton X-100, pH 8.8, New England Biolabs, Ipswich, MA), and 2.5 U/50 μL Taq DNA polymerase (New England Biolabs, Ipswich, MA) in 50 μL. Then samples containing ^{32}P labeled forward primer were denatured by heating with gel loading buffer and 15 μL of this mixture were analyzed on acrylamide gel as described in the previous section. Nonradioactive samples were analyzed via agarose gel electrophoresis. PCR conditions are described in results and discussion section of this chapter, all reactions have been performed using C1000 Touch thermal cycler (BioRad). Negative PCR no template control (NTC) was done with each PCR mixture.

Sanger and Illumina sequencing. Samples for Sanger sequencing were prepared by

purifying nonradioactive PCR products using QIAquick PCR purification kit (Qiagen, Hilden, Germany) following protocol suggested by manufacturer. Then 5-22 ng of purified PCR product and 3 pmol of sequencing primer were submitted for sequencing at the DNA Sequencing Core Facility, University of Utah. Concentration of dsDNA was calculated in ng/μL based on absorbance at 260 nm.

Samples for Illumina sequencing were prepared by purifying nonradioactive PCR products via reverse-phase HPLC on Microsorb 100-5 C18 column (150 x 2 mm) while running a linear gradient 5-60% B over 80 min at 0.7 mL/min flow rate (solvent A = 20 mM NH_4AcO pH 7, solvent B = MeCN). Then fractions containing PCR product were combined and submitted to High Throughput Genomics Core Facility, University of Utah.

Results and discussion

Oligonucleotide design. The next step after studying which nucleotides are inserted opposite OG, Gh, and Sp is investigating in more detail the sequencing signals from these lesions. For this purpose, five longer 35-37 nucleotide (nt) RNA templates containing one of the studied lesions in the middle were used (Figure 4.1). These templates were designed by choosing a sequence of 3-5 nts containing a lesion in the second position from the 3' end and adding sequences of PCR primers to the 3' and 5' end of it. Recommended length for primers for PCR is 18-30 nts with a GC-content of 40-60%, but the primers used for amplification in this case were just 16 nt-long and significantly more GC-rich (75%) to increase their melting temperature. Solid-phase synthesis of RNA still remains quite a challenging task, especially when nonstandard phosphoramidites like OG are used, and there is a significant drop in yield and purity of the final product when its length approaches

1L-X 3′-GCGUCACGGUACUGCCU**X**CACUGGACCUCGCGGAG
2L-X 3′-GCGUCACGGUACUGCCU**X**ACACUGGACCUCGCGGAG
2LG-X 3′-GCGUCACGGUACUGCCU**X**GCACUGGACCUCGCGGAG
2LU-X 3′-GCGUCACGGUACUGCCU**X**AUACUGGACCUCGCGGAG
3L-X 3′-GCGUCACGGUACUGCCU**X**AACACUGGACCUCGCGGAG

X = OG, Gh, or Sp

Figure 4.1. Sequences of RNA templates. Underlined nucleotides correspond to variable region in final RT-PCR product; nucleotides that are not underlined belong to primer regions. Blue markings indicate changes compared to 2L template.

40 nts.[4] Thus, the rationale behind using shorter PCR primers was to shorten RNA templates as much as possible. Design of these templates allowed amplification of cDNA via PCR upon reverse transcription (RT-PCR). Two different sets of primers were used for RT-PCR amplification of these templates. The first set was designed for next-generation sequencing (NGS) and consisted of two 16-mer DNA primers for PCR, P16 and P16r, former has also been used for RT (Figure 4.2 A). The second set was designed for Sanger sequencing and consisted of two PCR primers, 20-mer forward primer (P20S) and 16-mer reverse primer (P16r), and 80-mer DNA primer (P80) used for RT (Figure 4.2 B). The long RT primer hybridized to the RNA template via the same 16 bases on its 3'-end as the one used for NGS samples, and the only difference between them was a 64 nt-long 5' overhang designed not to have secondary structure. During Sanger sequencing information from 50-60 nucleotides just downstream of the primer binding site is lost. The reason for using such a long RT primer was to move the region of interest containing OG, Gh, or Sp outside of this sequencing dead-zone. The PCR forward primer has been designed to satisfy all requirements for Sanger sequencing primer (18-24 nts, 3' GC-lock, 45-55% GC content, and 50-65 °C T_m) and has been used in this role. The PCR reverse primer was identical to the one used for NGS samples. While studying deletion bypass of Sp and Gh elongated versions of RT primers were used to imitate insertion of either A or G opposite these lesions (P18A, P18G, P82A, and P82G).

RT-PCR condition optimization. Prior to submitting products of RT-PCR amplification of RNA templates for sequencing conditions for both reverse transcription and PCR were optimized. For this purpose, reverse transcription has been conducted at rt, 37 °C, and 50 °C using 2L RNA template and 16-mer RT primer. Products of the reaction

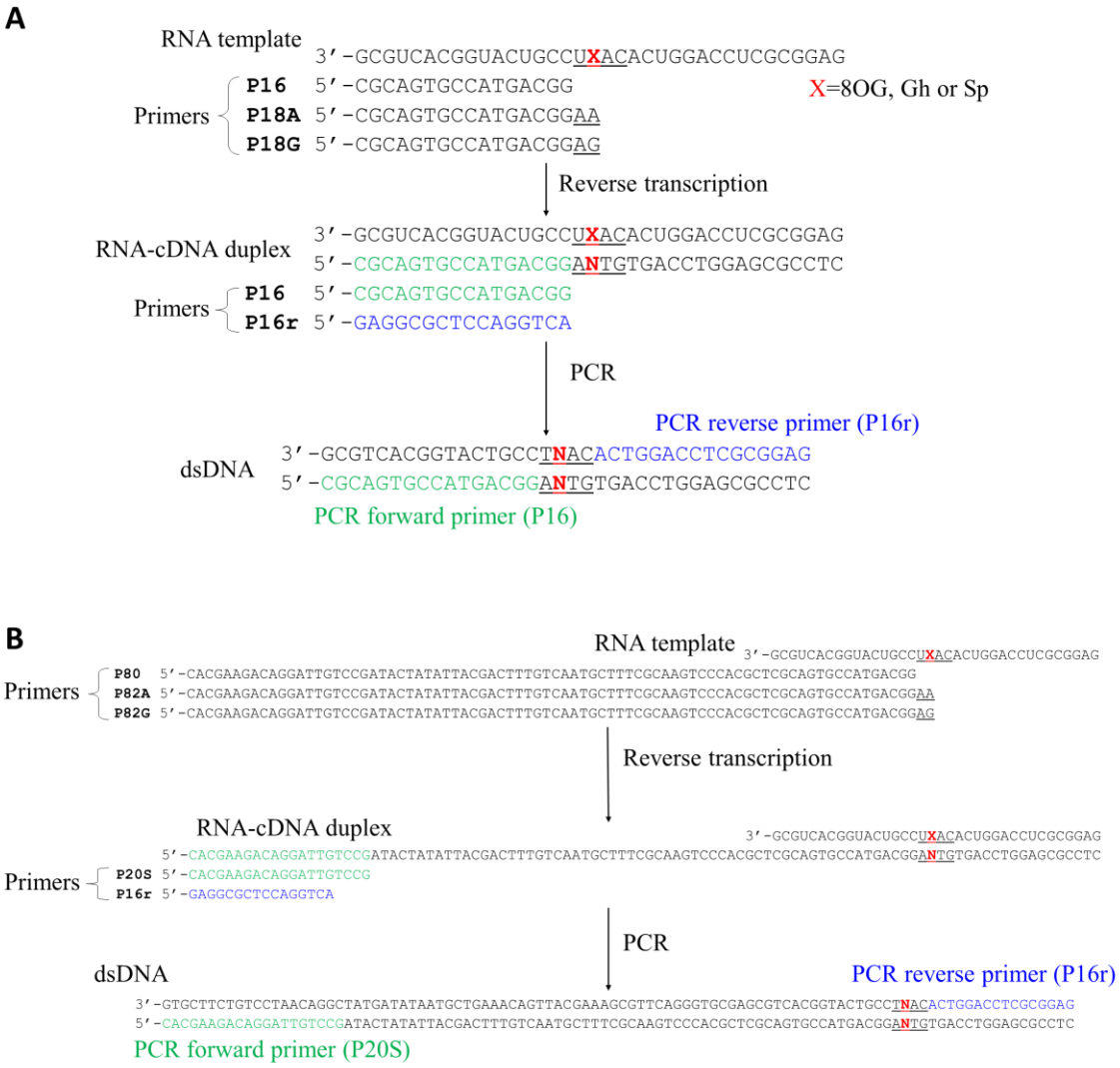


Figure 4.2. Schematic representation of RT-PCR with two different sets of primers. RT-PCR products amplified using a short RT primer **(A)** were used for gel analysis and Illumina sequencing, while RT-PCR that originated from RT with longer primer **(B)** were used for Sanger sequencing. Nucleotides marked in green are identical to the PCR forward primer, the ones marked in blue – to the PCR reverse primer.

were then resolved via PAGE and the amount of fully extended cDNA was quantified (Figure 4.3). Unsurprisingly the RNA template containing OG gave the greatest amount of full extension product followed by Gh- and Sp-containing templates. The presence of hydantoin lesions in the template caused a significant amount of polymerization arrest at the stage of insertion of a base opposite the lesion, as well as two bases following it (bands P+1, P+2, and P+3 correspondingly). These results were in perfect agreement with the results reported in Chapter 3. A minor trend of increasing yield of fully extended cDNA with the increase in reaction temperature was observed for Gh- and Sp-containing templates, while for OG-containing template the highest yield was observed at 37 °C and it decreased slightly from 37 °C to 50 °C. Hence, we chose a temperature of 45 °C for conducting all future RT reactions.

For optimization of PCR conditions cDNA was amplified using a different number of PCR cycles (15-20), annealing temperatures (50-57 °C), and extension conditions (0.5-1 min at 68-72 °C). The main criteria for evaluation of PCR conditions were amount of PCR product and absence of amplification in negative NTC (data not shown). The final conditions used for PCR were: 18 cycles, 30 s denaturation at 95 °C, 30 s annealing at 57 °C, and 1 min extension at 72 °C.

Sample preparation for Illumina sequencing. To quantify the insertion ratio of A and C opposite OG and G and A opposite Sp and Gh we performed RT-PCR amplification of the 2L template using short (P16) RT primer. Apart from quantifying the insertion ratio under the standard RT conditions, I also wanted to test to what extent sensitivity of detection of these bases can be improved by increasing the concentration of the dNTP corresponding to one of two nucleotides inserted opposite them. This approach can allow

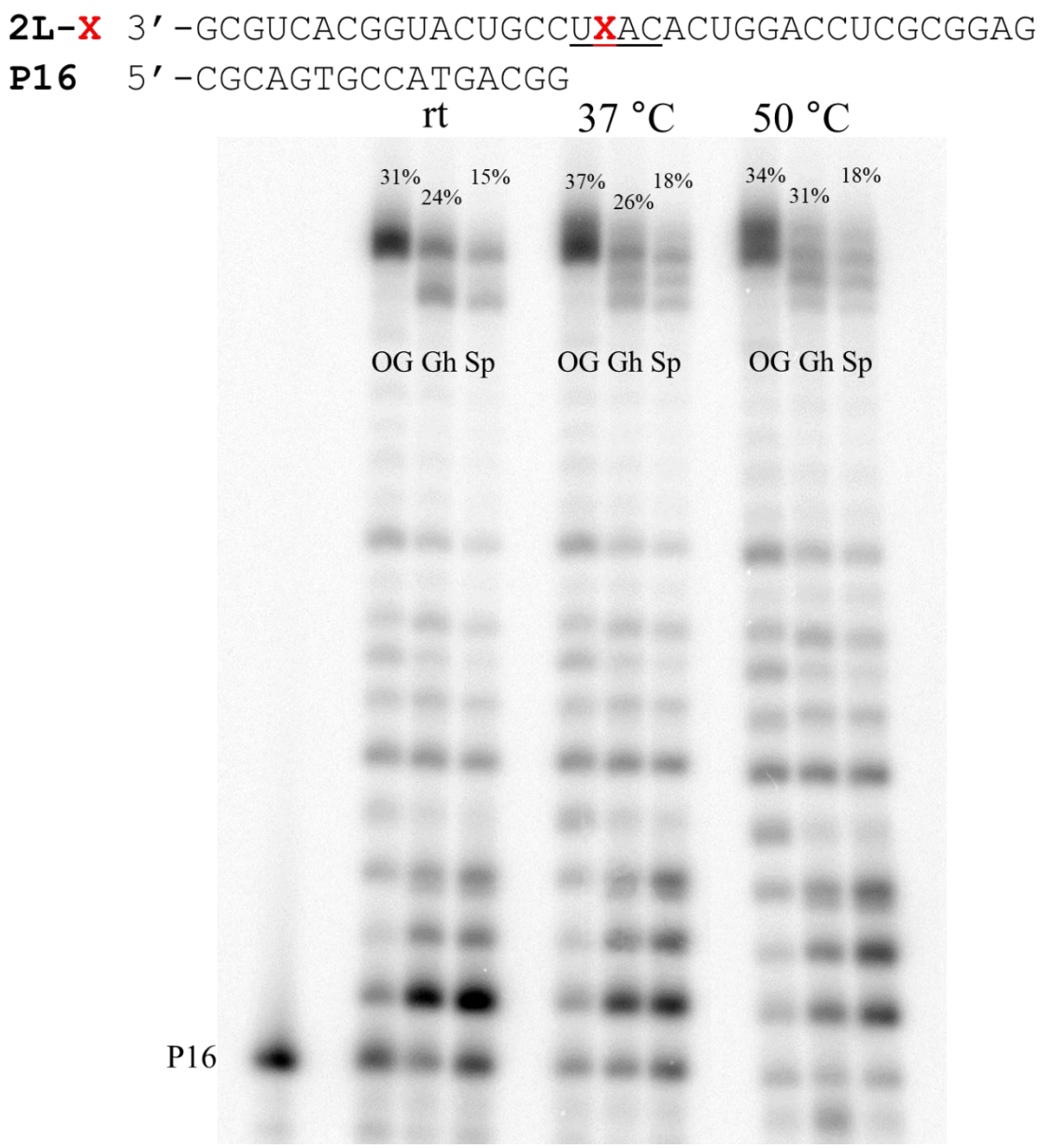


Figure 4.3. PAGE analysis of reverse transcription products at different reaction temperatures.

shifting insertion of a mixture of two bases toward just one of them and by doing so improve the noise to signal ratio that will be crucial once this sequencing method is applied to biological systems. For this purpose, there were two additional samples submitted for Illumina sequencing, “OG-A” and “Sp-A,” in case of these samples the RT reaction was conducted with a twice higher concentration of dATP. We also considered trying to channel insertion opposite Gh or Sp towards G, but increasing the dGTP concentration in the reaction mixture would likely have led to an increase in the amount of G insertion opposite T. An increase in reverse transcription errors upon an increase in single dNTP concentration has to be separately studied to evaluate whether this approach can be useful. RT-PCR products synthesized for Illumina sequencing were tested via PAGE (Figure 4.4) and purified via reverse-phase HPLC (Figure 4.5 A). Based on the gel, the PCR yielded close to 50% of amplification product for OG and Sp and ~70% for Gh. Interestingly, RT-PCR of the templates containing both hydantoin lesions produced about 1:1 ratio of the full-length and 2 nts shorter amplification products. This phenomenon is further investigated in the next section of the current chapter. The main reason for using HPLC was insufficient length of PCR amplicons for purification with standard PCR cleanup kits. All fractions coming between 2 min and 81 min were collected and then their absorbance spectra were analyzed. Three peaks on the chromatogram in Figure 4.5 were found to have absorbance maximum at 260 nm: 2 min, 10 min, and 11 min. These fractions were further tested by performing their PCR amplification (Figure 4.5 B). The peak at 2 min most likely contained dNTPs and unextended primers and gave smeared bands on the agarose gel upon amplification due to their concentrations exceeding the optimal ones for PCR. The peaks at 10 and 11 min gave clear PCR product bands indicating the presence of the appropriate

2L-X 3′-GCGUCACGGUACUGCCU**X**ACACUGGACCUCGCGGAG

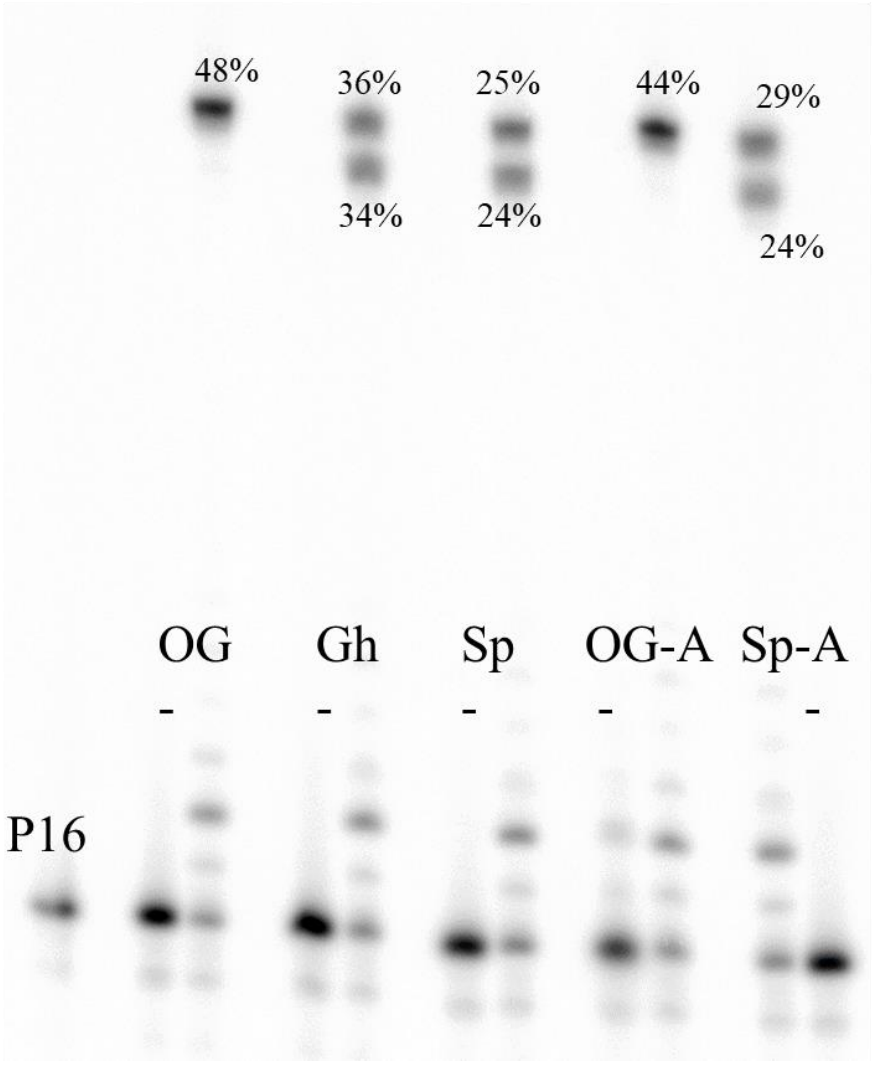
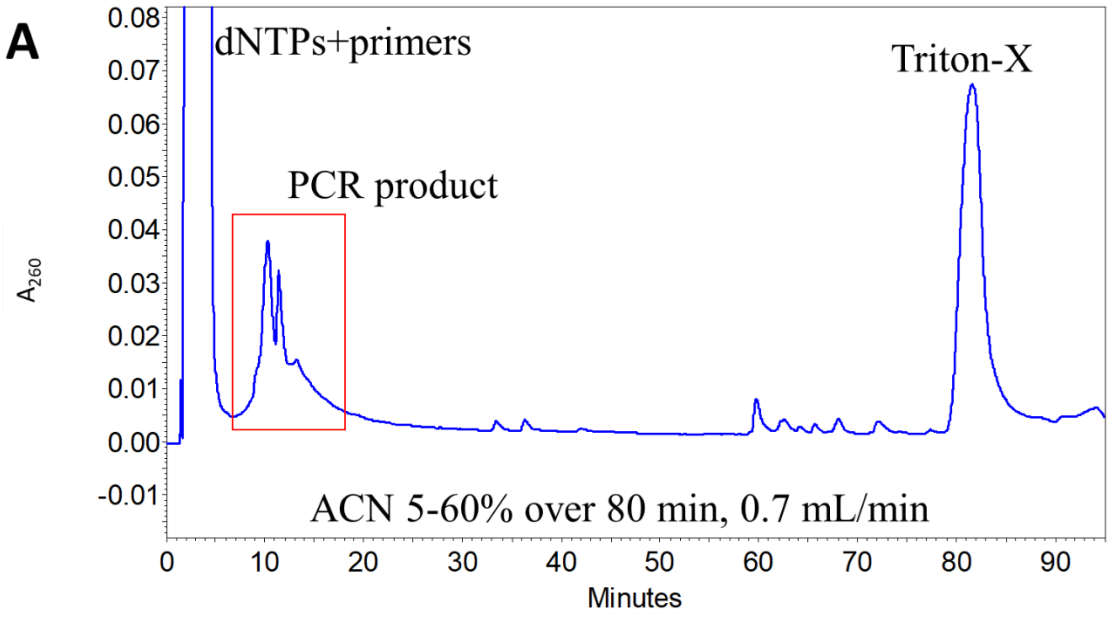


Figure 4.4. PAGE analysis of RT-PCR products originating from 2L templates. Lanes marked with "–" are negative PCR NTC.



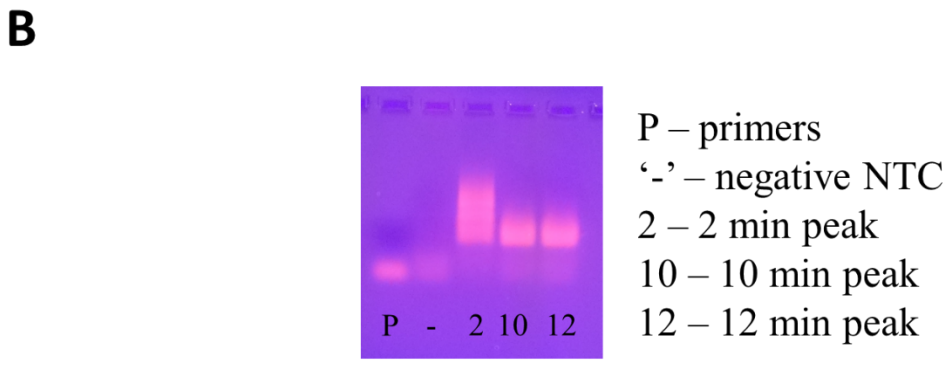


Figure 4.5. Purification of RT-PCR products by HPLC and fraction analysis. **(A)** Typical HPLC trace, red box marks peaks corresponding to PCR product. **(B)** Agarose gel showing products of PCR amplification of different fractions collected from HPLC.

template for PCR. Hence, these two fractions for each of the five samples (OG, Gh, Sp, OG-A, and Sp-A) were combined together and submitted for Illumina sequencing.

Deletion bypass of Sp and Gh. Previously our lab reported that primers extended just until the position occupied by Gh in the template show improved resistance to proofreading by *E. coli* DNA polymerase I Klenow fragment when base positioned opposite Gh can form a stable base pair with the next nucleotide.[5] This together with the previous observation that polymerization based on Gh-containing templates produces a significant amount of n-1 primer extension product[6] led to a hypothesis that after insertion of A or G opposite of Gh this lesion can be twisted out of the DNA helix if this leads to formation of a stable A-T or G-C base pair (Figure 4.6). To test whether this phenomenon was responsible for the appearance of shorter RT-PCR product when RNA templates contained Gh or Sp (Figure 4.4), we studied the dependence of the ratio between full-length and truncated PCR products on two factors:

a) identity of the base inserted opposite the hydantoins
b) nucleotide sequence 5' downstream from the lesion position in the template.

For this purpose, we chose to use Sanger sequencing of long RT-PCR products in combination with PAGE analysis of the short RT-PCR products because of its lower cost and faster sample turnover. Apart from the RNA template 2L used in the previous studies, we synthesized OG-, Gh-, and Sp-containing versions of another four template strands (Figure 4.1). These templates had different amounts of purine bases between the site occupied by one of the lesions and the next pyrimidine base that can form a stable base pair with A or G inserted opposite Sp or Gh. This design allowed study of the efficiency of deletion of 1-3 bases via the mechanism displayed on Figure 4.6. In addition to the 16

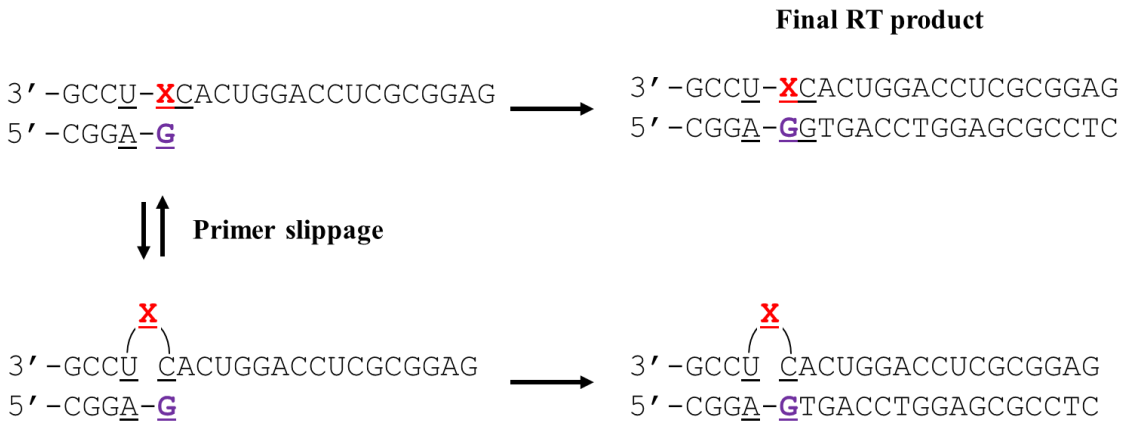


Figure 4.6. Mechanism of deletion bypass of Gh and Sp.

nt- and 80 nt-long RT primers (P16 and P80 on Figure 4.1) for the reverse transcription, we used their longer versions extended by 2 nucleotides to imitate 100% insertion of either A or G (P18A, P18G, P82A, and P82G of Figure 4.1) and to study efficiency of deletion bypass separately for A and G insertion.

Prior to studying deletion bypass of Sp and Gh, we tested whether the system designed can be used for Sanger sequencing. To do so, we reverse transcribed 2L-OG template using the P80 primer, PCR-amplified cDNA using P20S and P16r primers, and submitted the purified PCR product for sequencing with the P20S primer. Sequencing chromatograms, as well as gel data showing efficiency of RT and subsequent PCR, are presented on Figure 4.7. Based on the PCR product gel analysis and Sanger sequencing trace, the designed system can successfully be used to obtain an ensemble sequencing information on RT-PCR products. However, it should be noted that errors introduced during solid-phase synthesis of the 80 nt-long primer cause appearance of a noticeable amount n-1 bands on sequencing chromatograms and make PAGE analysis of RT products impossible due to very diffuse bands.

For comprehensive investigation of deletion bypass of Sp and Gh, we employed the strategy schematically represented on Figure 4.8. First, we conducted RT with one of the short primers (P16, P18A, or P18G) and subsequently PCR-amplified the cDNA with P16 and P16r primers. For the gel analysis of the RT products, we used ^{32}P-labeled primers during the RT reaction, while for gel analysis of the RT-PCR products, RT was done with nonradioactive primers and only during PCR ^{32}P-labeled P16 primer was added. Next, we conducted RT-PCR amplification of all samples using long RT primers (P80, P82A, or P82G) and submitted the products for Sanger sequencing. This workflow allowed

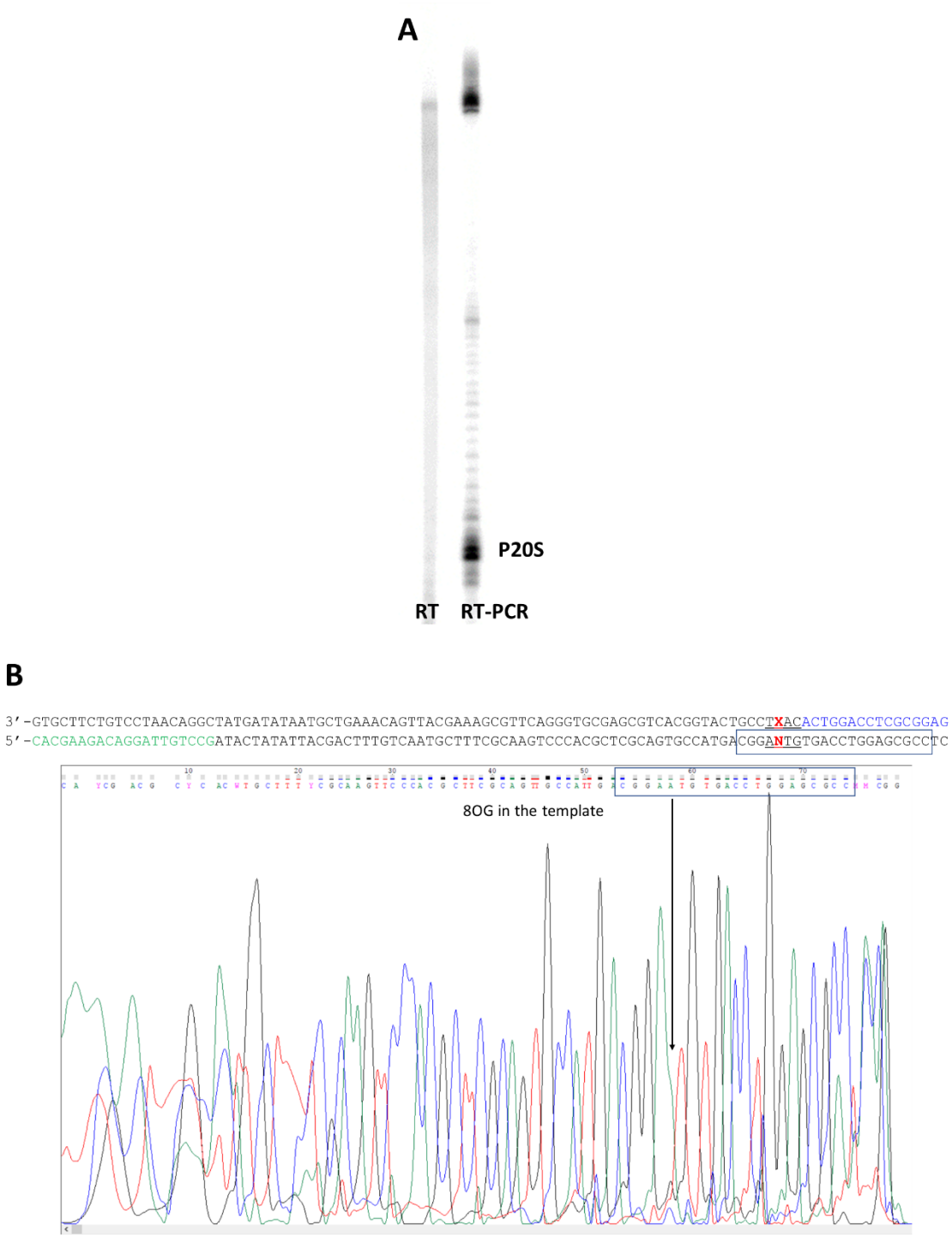


Figure 4.7. PAGE **(A)** and Sanger sequencing **(B)** data for products of RT or RT-PCR originating from 2L-OG template. On the gel P20S marks position of PCR primer. On the sequencing chromatogram black arrow indicates position of OG in the template.

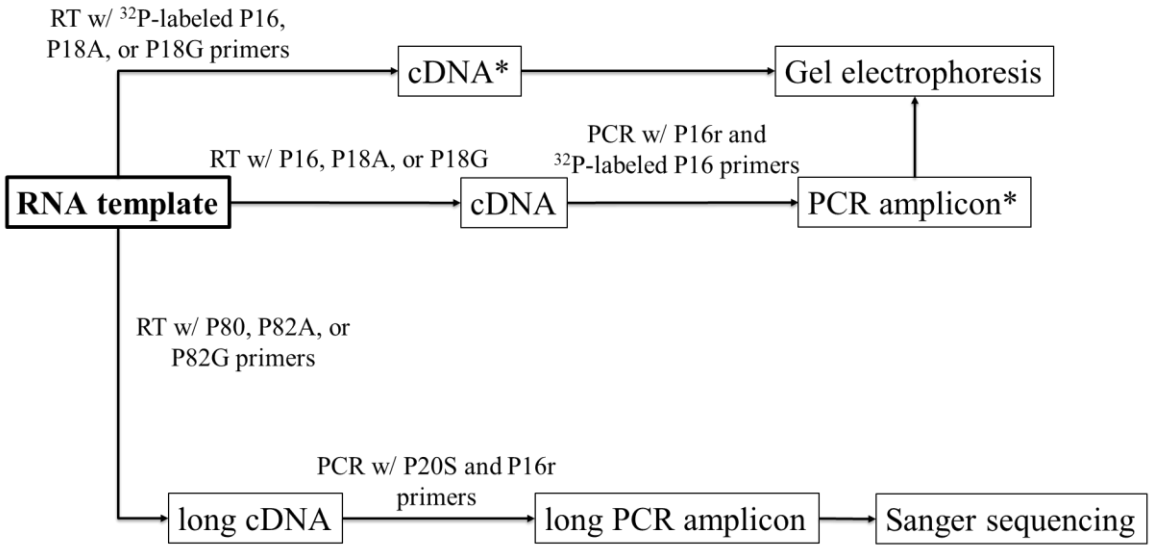


Figure 4.8. Schematic representation of the workflow. * indicates radiolabeled species.

quantification of the amount of full-length and 1-3 nts shorter PCR products based on PAGE analysis and to test whether formation of shorter PCR products is caused by expulsion of hydantoin bases from the RNA-DNA helix using Sanger sequencing data. The RT and PCR reactions were conducted under conditions previously optimized for samples prepared for Illumina sequencing. Results of this study are presented on Figures 4.9, 4.10, 4.12-18 and Table 4.1. On the figures the samples are labeled as NRRR-BB, where RRR-BB is a previously used identifier for RNA template used for RT and oxidized lesion that this template contains and N, when it is added, indicates that reverse transcription was done using 2 nts longer primer with N base at the 3'-end of the primer positioned opposite OG, Gh, or Sp.

Gels on Figures 4.9 and 4.10 show the efficiency of primer extension for all five templates with three different primers. Samples extended using P16 show efficiency of insertion and extension under running start conditions, while samples extended using P18N primers show efficiency of extension under standing-start conditions. The presence of both Sp and Gh causes a dramatic decrease polymerase activity when extension is conducted under the standing-start vs. the running-start conditions. This behavior is typical for mismatches[7] and is likely caused by the fact that neither Sp nor Gh form stable base pairs leading to end fraying at the 3'-end of the primer and affecting the ability for the polymerase to recognize the substrate and initiate primer extension, the same effect is thought to be responsible for DNA polymerase proofreading.[8] To test whether this effect is specific for Gh- and Sp-containing templates, we ran RT of 1L-OG and 2L-OG templates (Figure 4.11). Indeed, when primer extension under standing-start conditions is conducted with the P16A primer that can form a stable OG-A base pair, this effect is significantly less

1L-X 3′-GCGUCACGGUACUGCCU**X**CACUGGACCUCGCGGAG
2L-X 3′-GCGUCACGGUACUGCCU**X**ACACUGGACCUCGCGGAG
2LG-X 3′-GCGUCACGGUACUGCCU**X**GCACUGGACCUCGCGGAG
2LU-X 3′-GCGUCACGGUACUGCCU**X**AUACUGGACCUCGCGGAG
3L-X 3′-GCGUCACGGUACUGCCU**X**AACACUGGACCUCGCGGAG

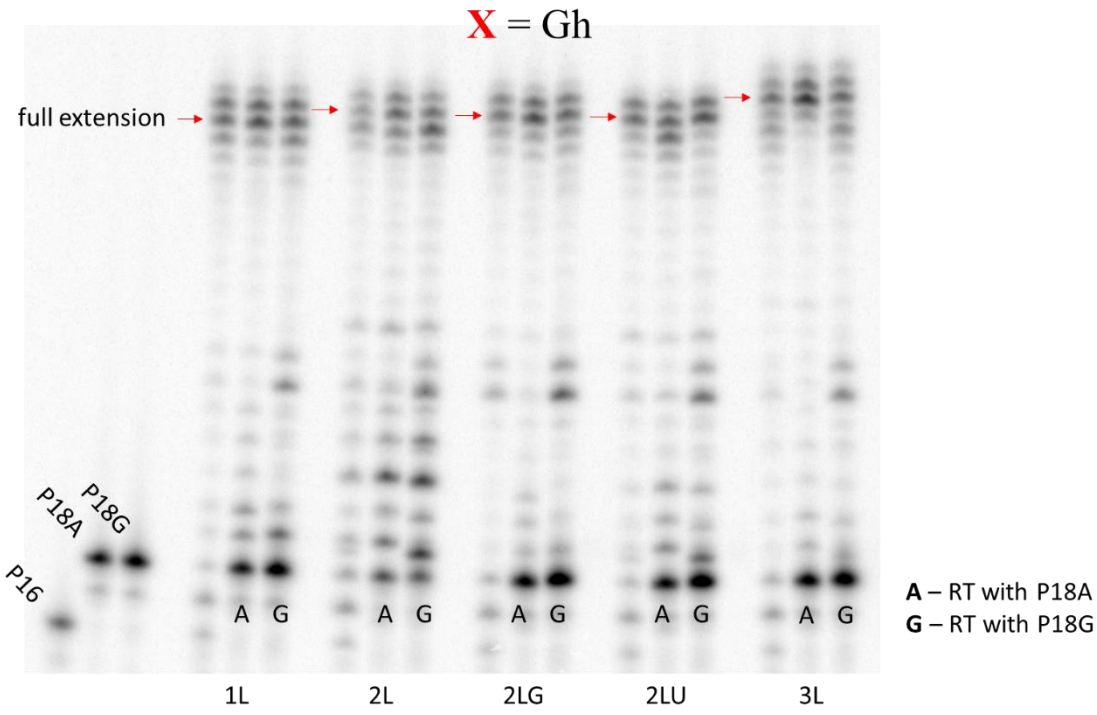


Figure 4.9. PAGE analysis of cDNA obtained by reverse transcribing Gh-containing templates. Red arrows indicate position of full extension product on the gel.

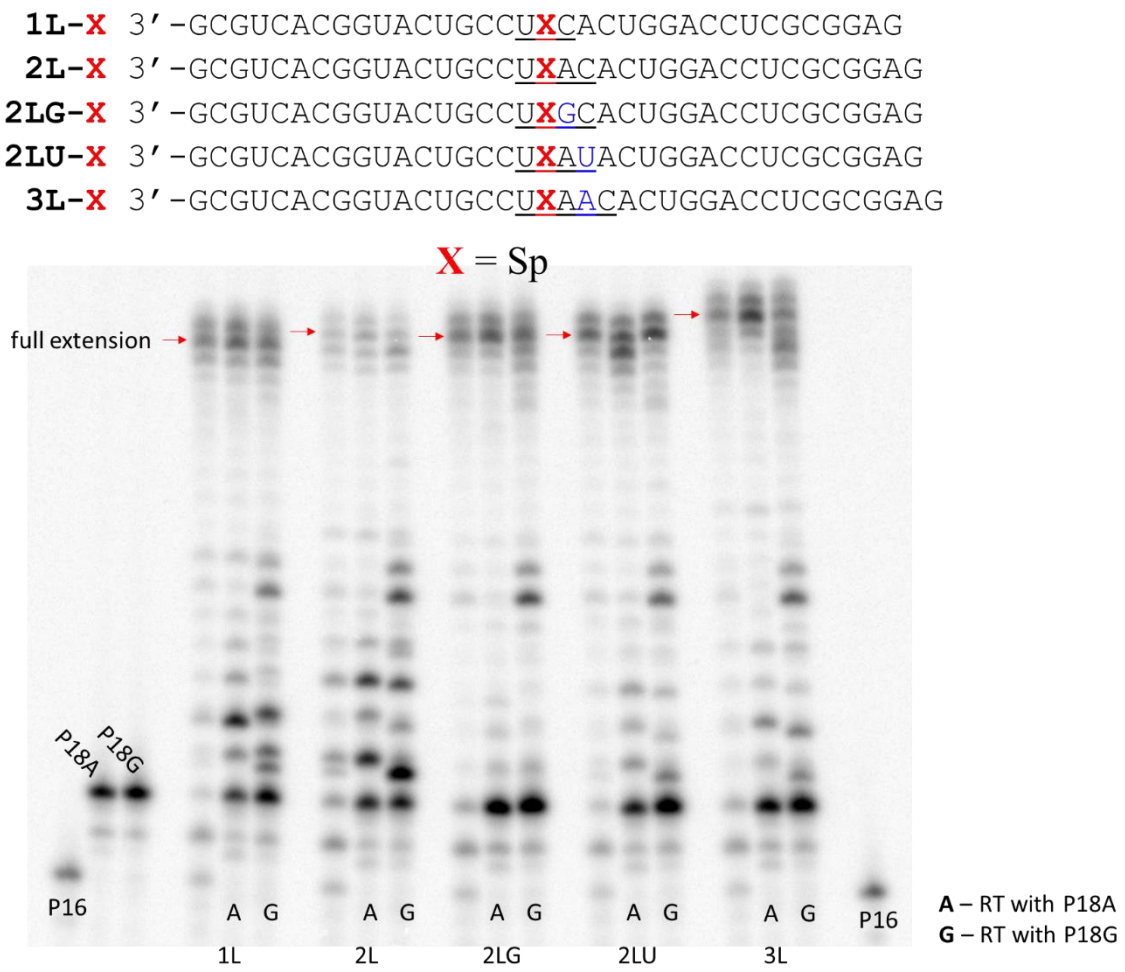


Figure 4.10. PAGE analysis of cDNA obtained by reverse transcribing Sp-containing templates. Red arrows indicate position of full extension product on the gel.

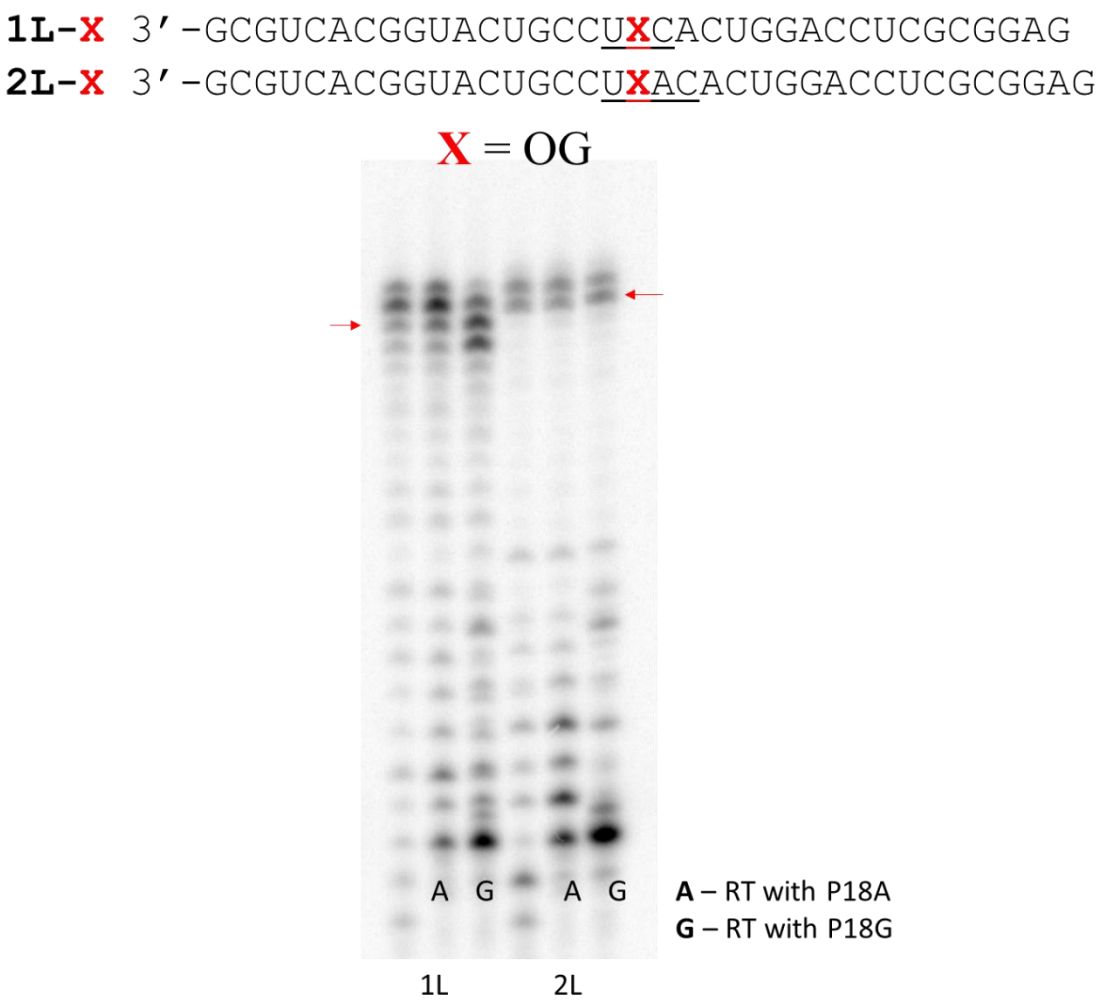


Figure 4.11. PAGE analysis of cDNA obtained by reverse transcribing OG-containing templates. Red arrows indicate position of full extension product on the gel.

prominent when compared to extension with the P16G primer or any templates having hydantoin lesions. Notably, reverse transcription of both 2L-Gh and 2L-Sp templates under the standing-start conditions leads to the largest amount of truncated products and the smallest amount of unextended primer. In case of the 2L-Sp template, reverse transcription also yields the lowest amount of full-length cDNA. Improved efficiency of RT initiation is indicative of better recognition of primer-template duplexes involving 2L template by polymerase, while the abundance of produced truncated products shows that for some reason polymerase-substrate interactions get disrupted once primer growth is initiated. Due to the fact that RT leads to formation of numerous truncated and overextended products it is impossible to estimate efficiency of deletion bypass based on these gels.

Measurement of a ratio between full-length and deletion cDNA was done based on PAGE analysis of RT-PCR products (Figures 4.12 and 4.13). These measurements were repeated 3 times for all templates with Sp and 5 times for templates with Gh; the results are summarized in Table 4.1. Major discrepancies in efficiency of deletion bypass between templates with Sp and Gh are highlighted in red. Additional RNA degradation control was added to the gel on Figure 4.12. Taq polymerase is known to be able to use RNA as a template[9] and the purpose of this control was to check whether any of the observed RT-PCR products originated from this. To test this, we conducted the RT reaction without the enzyme followed by a standard heating step that we used to degrade the RNA template and attempted to amplify degraded RNA template via PCR. This control was indistinguishable from the negative NTC meaning that none of the bands on the gel originate from direct RNA amplification. For templates 1L and 2L, the presence of Gh led to formation of different amounts of deletion product in the case of either G or A insertion when compared

1L-X 3′-GCGUCACGGUACUGCCU**X**CACUGGACCUCGCGGAG
2L-X 3′-GCGUCACGGUACUGCCU**X**ACACUGGACCUCGCGGAG
2LG-X 3′-GCGUCACGGUACUGCCU**X**GCACUGGACCUCGCGGAG
2LU-X 3′-GCGUCACGGUACUGCCU**X**AUACUGGACCUCGCGGAG
3L-X 3′-GCGUCACGGUACUGCCU**X**AACACUGGACCUCGCGGAG

X = Gh

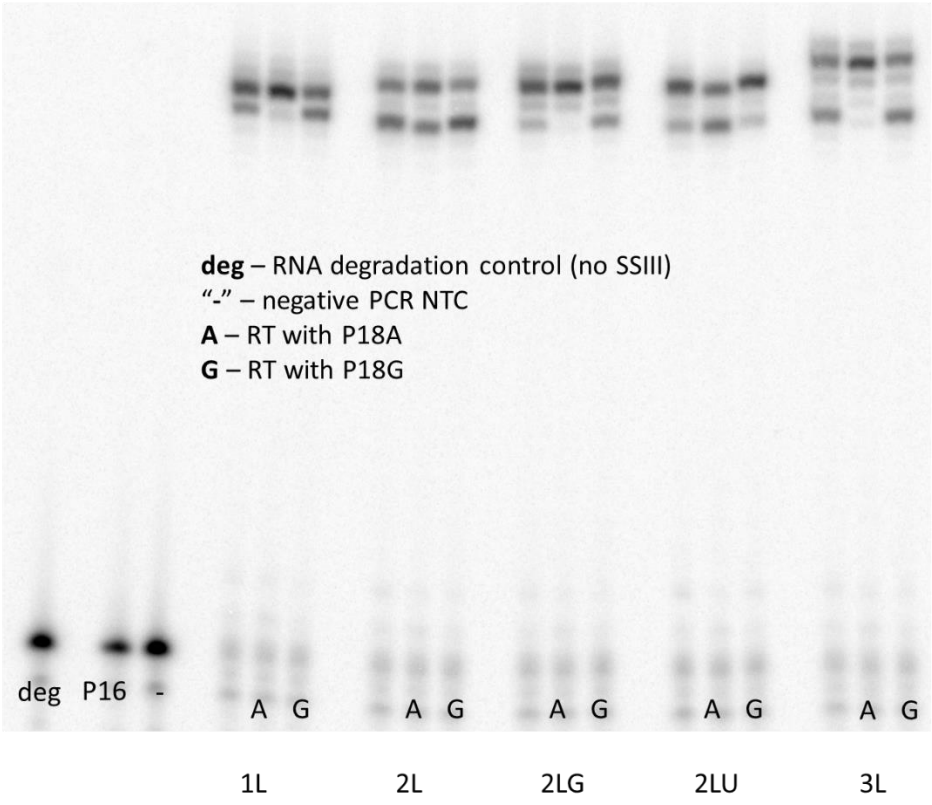


Figure 4.12. PAGE analysis of RT-PCR products originating from Gh-containing templates.

1L-X 3′-GCGUCACGGUACUGCCU**X**CACUGGACCUCGCGGAG
2L-X 3′-GCGUCACGGUACUGCCU**X**ACACUGGACCUCGCGGAG
2LG-X 3′-GCGUCACGGUACUGCCU**X**GCACUGGACCUCGCGGAG
2LU-X 3′-GCGUCACGGUACUGCCU**X**AUACUGGACCUCGCGGAG
3L-X 3′-GCGUCACGGUACUGCCU**X**AACACUGGACCUCGCGGAG

X = Sp

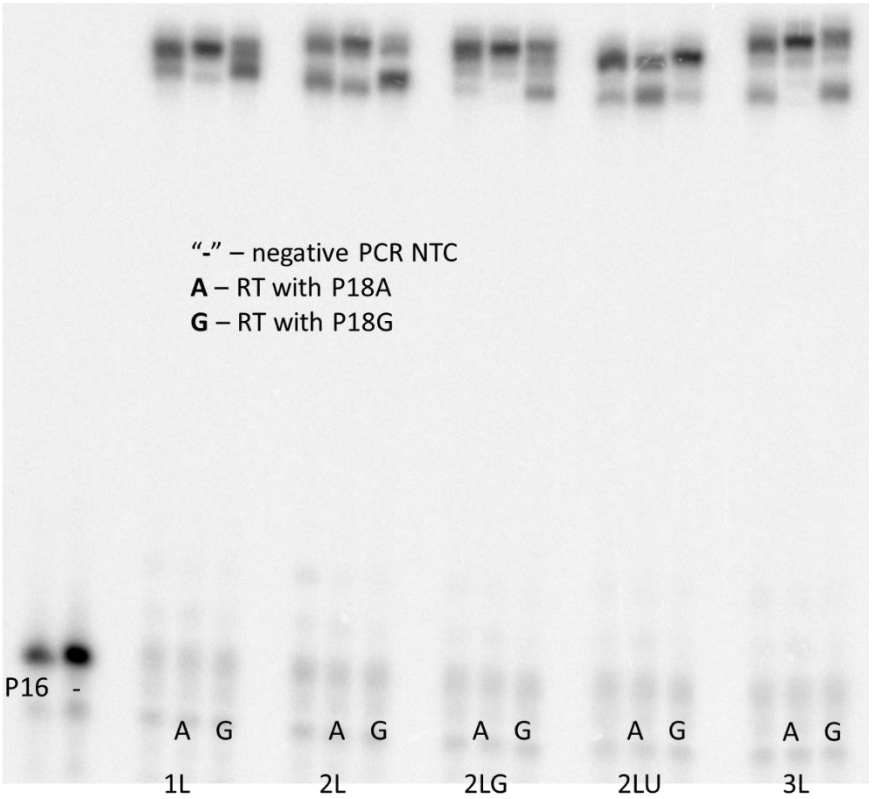


Figure 4.13. PAGE analysis of RT-PCR products originating from Sp-containing templates.

Table 4.1. Results of quantification of fraction of deletion bypass products on Figures 4.12 and 4.13.

	Gh		Sp
1L	30±2%*	**1L**	28±1%
A1L	11±4%	**A1L**	15±2%
G1L	39.8±0.5%	**G1L**	48±1%
2L	51±2%	**2L**	47.3±0.6%
A2L	43±2%	**A2L**	35.5±0.2%
G2L	57±1%	**G2L**	60.2±0.4%
2LG	19±1%	**2LG**	11±1%
A2LG	4.6±0.9%	**A2LG**	3.2±0.4%
G2LG	28±1%	**G2LG**	28.5±0.5%
2LU	29±1%	**2LU**	27±1%
A2LU	47.2±0.6%	**A2LU**	47±1%
G2LU	21±2%	**G2LU**	17±2%
3L	38±1%	**3L**	29.1±0.3%
A3L	12±1%	**A3L**	10.1±0.6%
G3L	41±1%	**G3L**	37.7±0.5%

*- 97% confidence intervals from 5 measurements for Gh and 3 samples for Sp

to Sp, but because of a higher ratio of G inserted opposite Gh this did not result in change in efficiency of deletion bypass under standing-start conditions when both nucleotides can be inserted. For the same reason, 2LG and 3L templates containing Gh showed formation of a larger amount of deletion product under running-start conditions while individual insertion of A and G yielded a similar amount of deletion product. Interestingly, RT-PCR of 2LU template with all three used primers yielded similar results for both Sp and Gh. The data received from Sanger sequencing were in good alignment with the data from the gels (Figures 4.14-18) with a few exceptions (2LG-Gh on Figure 4.16, 2LU-Gh, G2LU-Gh, 2LU-Sp, and G2LU-Sp on Figure 4.17) where the amount of deletion bypass product was underestimated by Sanger sequencing.

Overall, the efficiency of deletion bypass based on PAGE and Sanger sequencing data appeared to be highly dependent on the nature of bases 5' downstream from the lesion position. The highest amount of deletion RT-PCR product was observed for the 2L template. However, the formation of this product does not appear to be driven just by formation of a stable G-C base pair (G2L samples) because large amounts of deletion product were observed even when A was inserted opposite Gh or Sp (A2L samples). Deletion bypass in this case occurs through formation of an A-C base pair that becomes stable only at low pH when N1 of A is protonated.[10] Apart from this anomaly, deletion bypass becomes prominent only when a stable locking base pair can be formed (G-C for 1L, 2LG, and 3L and A-U or G-U for 2LU). Interestingly, the efficiency of deletion bypass also depends on the nature of the purine base that is expelled from the duplex together with Gh or Sp. When this base is an A (2L samples), the deletion product amount is significantly higher than when it is G (G2L sample). In addition to that, 2LG templates produced a

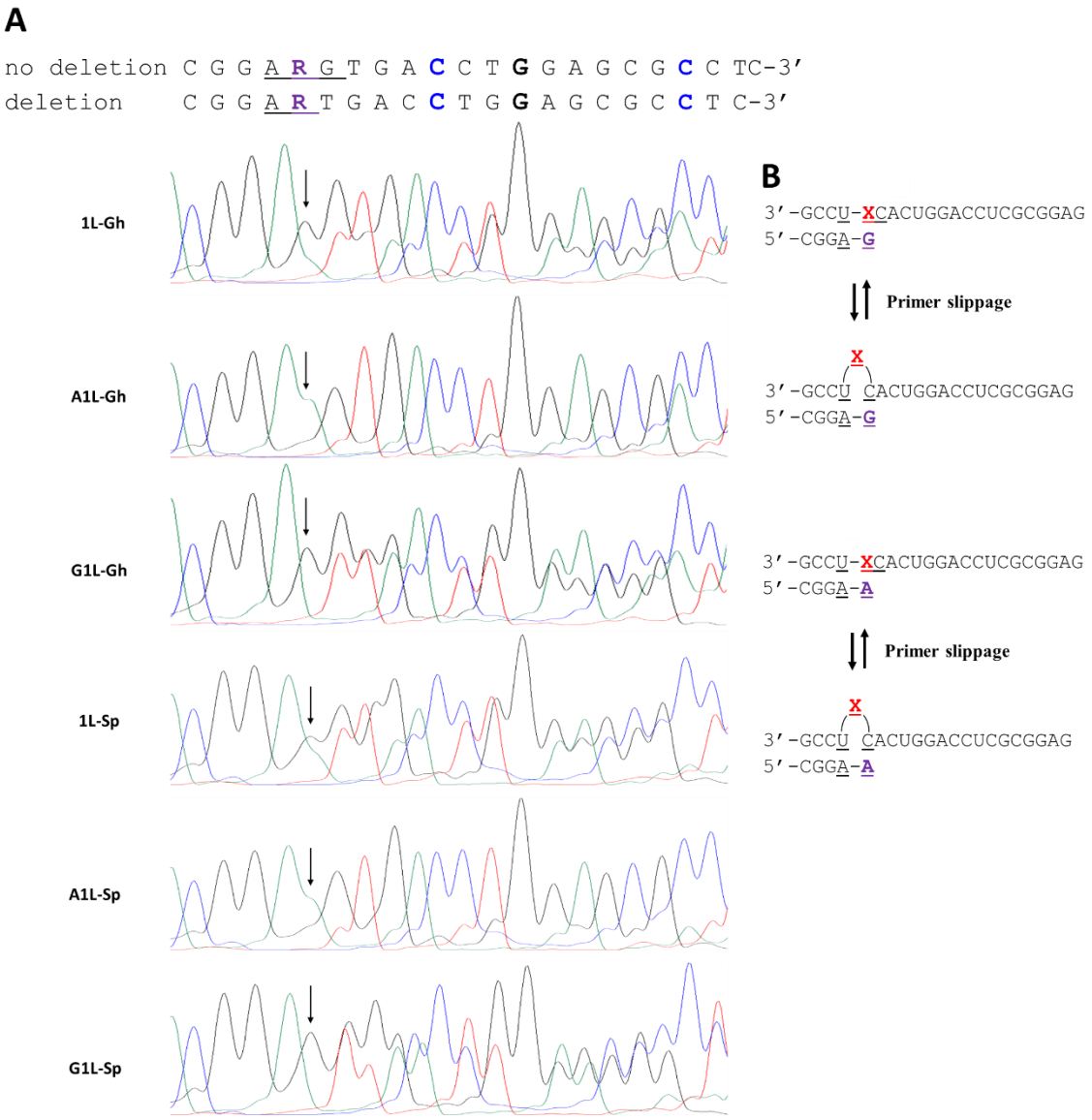


Figure 4.14. Deletion bypass of hydantoin lesions in 1L template. **(A)** Sanger sequencing data for RT-PCR products originating from 1L template containing Gh or Sp. **(B)** Mechanism of deletion product formation. Black arrow indicates position occupied by lesion in the template.

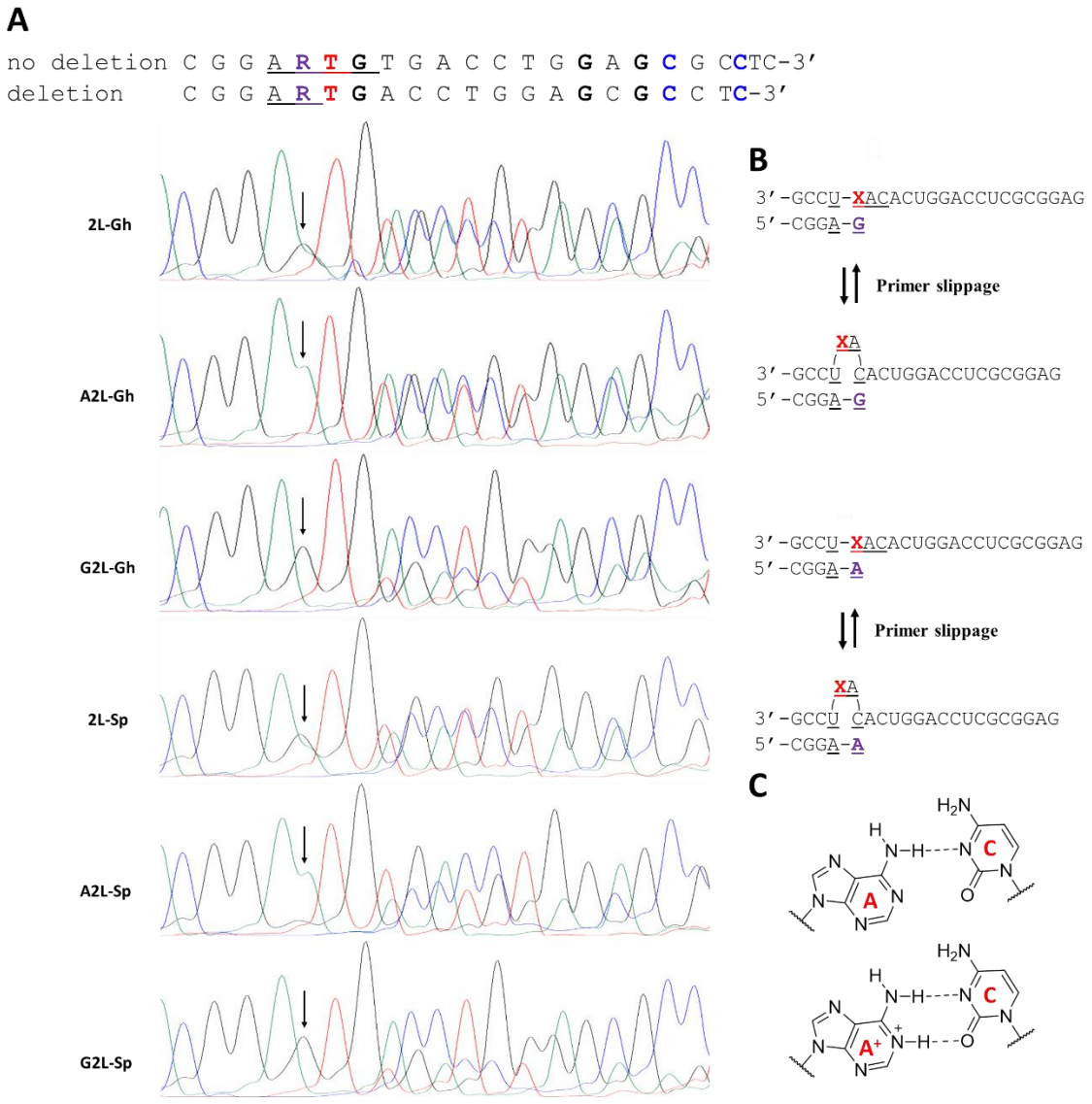


Figure 4.15. Deletion bypass of hydantoin lesions in 2L template. **(A)** Sanger sequencing data for RT-PCR products originating from 2L template containing Gh or Sp. **(B)** Mechanism of deletion product formation. **(C)** Structures of protonated and unprotonated A-C base pairs. Black arrow indicates position occupied by lesion in the template.

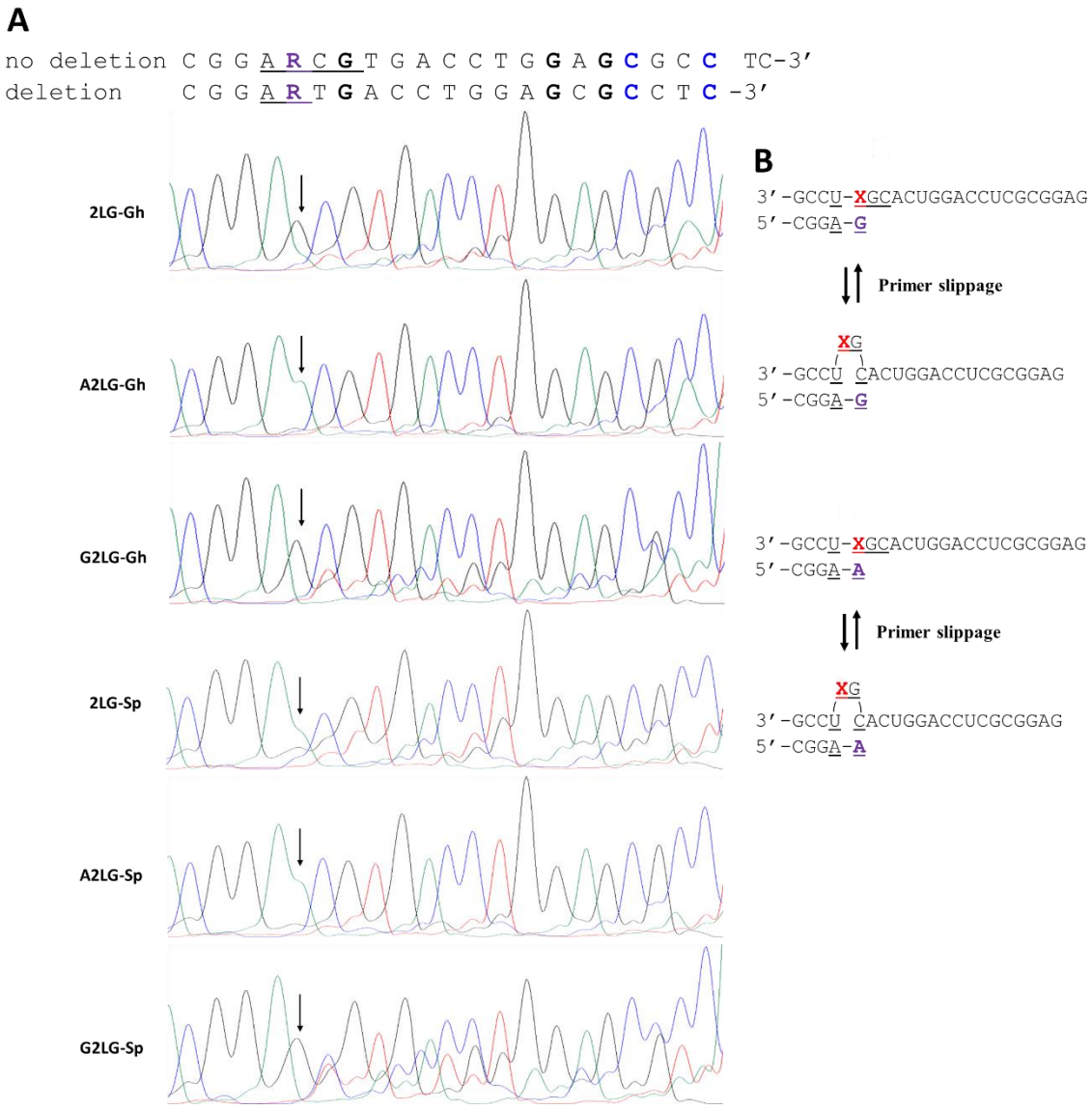


Figure 4.16. Deletion bypass of hydantoin lesions in 2LG template. **(A)** Sanger sequencing data for RT-PCR products originating from 2LG template containing Gh or Sp. **(B)** Mechanism of deletion product formation. Black arrow indicates position occupied by lesion in the template.

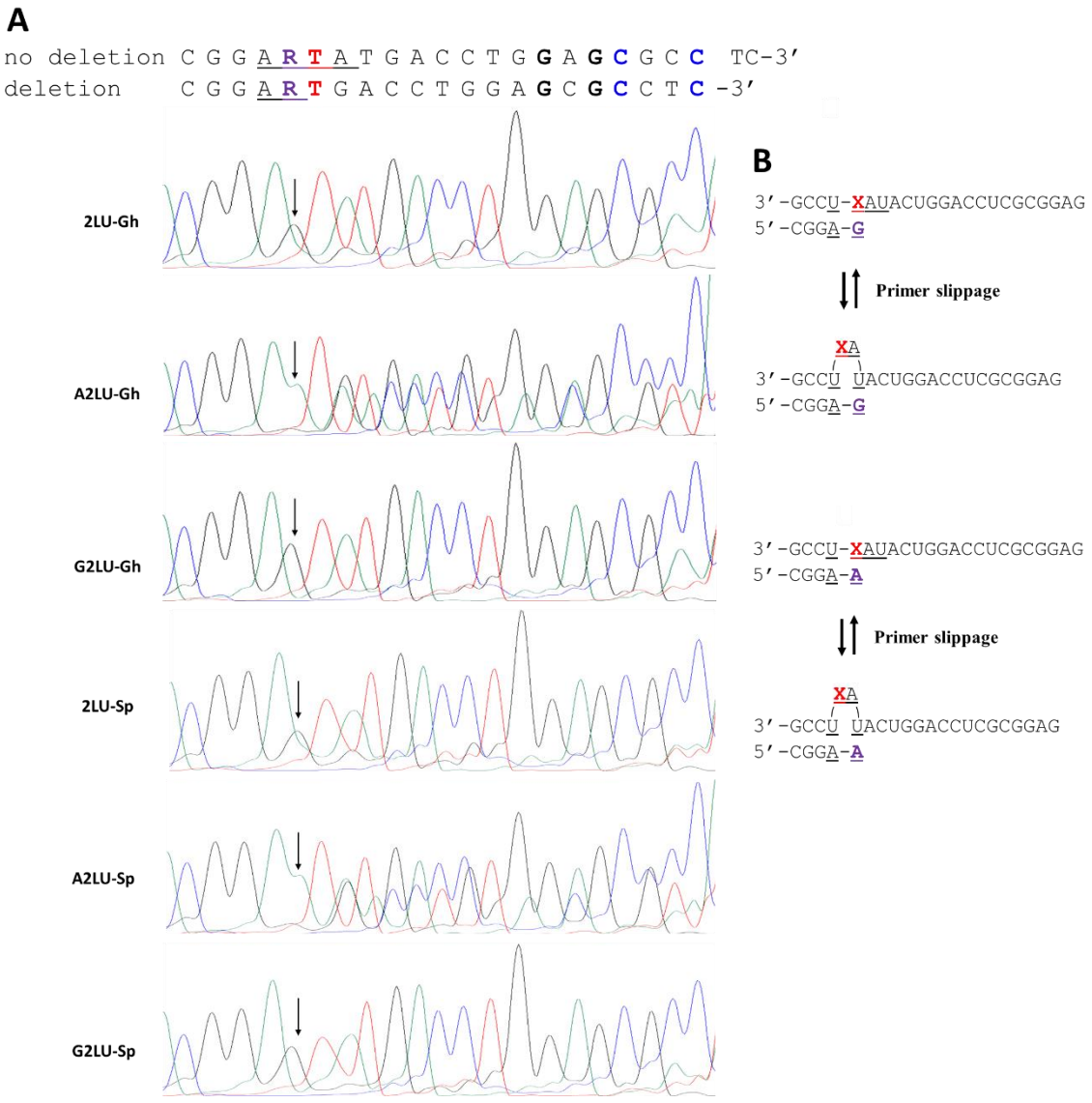


Figure 4.17. Deletion bypass of hydantoin lesions in 2LU template. **(A)** Sanger sequencing data for RT-PCR products originating from 2LU template containing Gh or Sp. **(B)** Mechanism of deletion product formation. Black arrow indicates position occupied by lesion in the template.

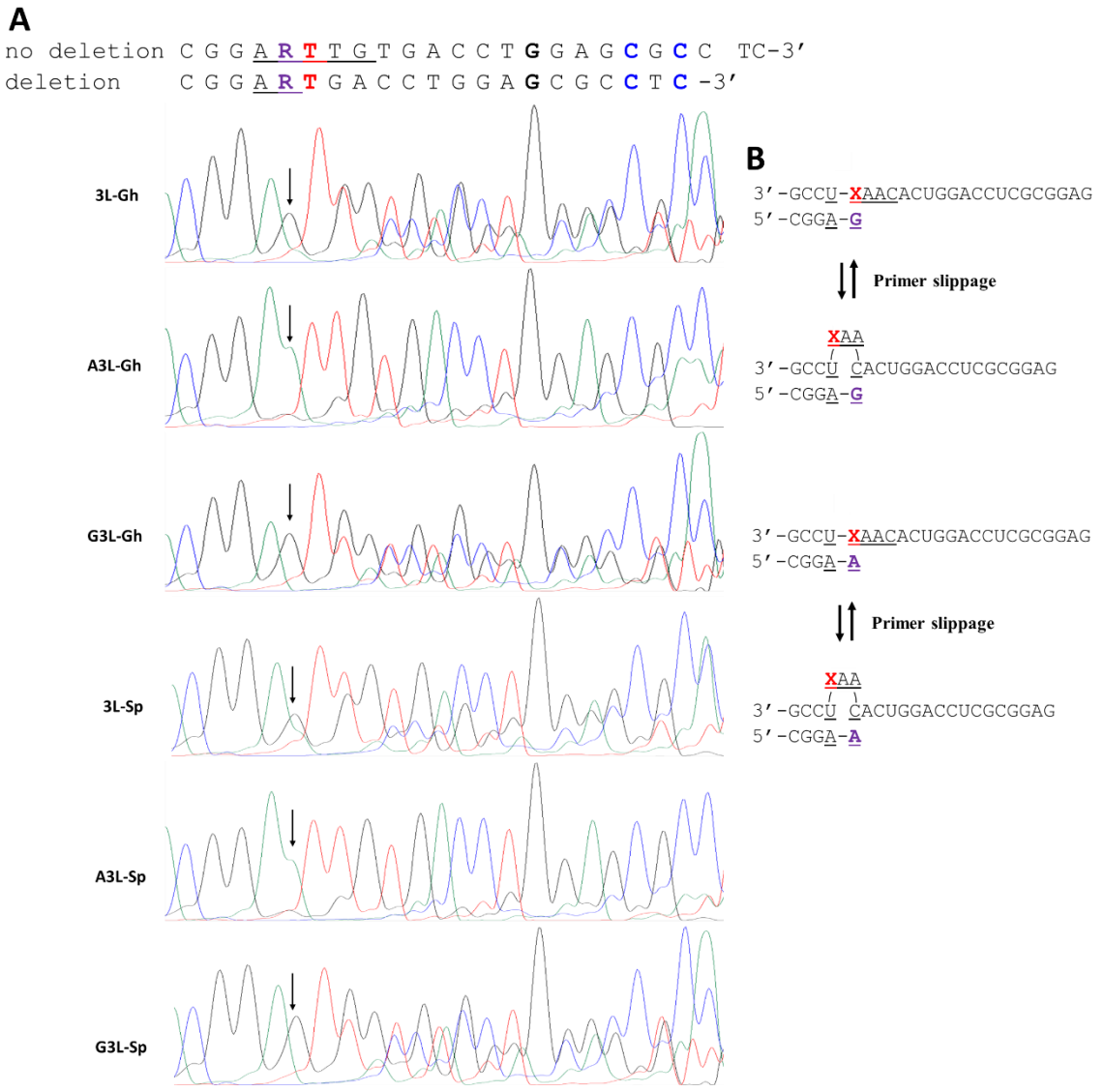


Figure 4.18. Deletion bypass of hydantoin lesions in 3L template. **(A)** Sanger sequencing data for RT-PCR products originating from 3L template containing Gh or Sp. **(B)** Mechanism of deletion product formation. Black arrow indicates position occupied by lesion in the template.

slightly increased amount of n-1 deletion product with P82G and P18G indicative of deletion bypass via formation of G-G base pair; this effect was more prominent for Sp. Another interesting detail is that efficiency of deletion bypass increases from the 1L template to the 3L template. This is highly unusual behavior because deletion bypass in the second case requires formation of extrahelical loop containing 3 nucleotides instead of just one nucleotide in the first case. To test whether the observed trend for deletion bypass efficiency of Sp-R and Gh-R (R=purine base A or G) base pairs is similar to mismatched base pairs, we used the same five templates containing OG. These templates were reverse-transcribed using P80G (enforces G-G mismatch at the growing end of the primer) and P80 (control) primers and PCR amplified. The results of Sanger sequencing of the RT-PCR products are presented on Figures 4.19 and 4.20. In this case, the only template that led to formation of a significant amount of deletion product was 1L-OG (~70% n-1 product). This confirmed our initial assessment that deletion bypass of Sp and Gh is highly unusual and is unlikely to be driven only by formation of stable base pair at the 3' end of the growing primer.

We hypothesized that unusual susceptibility of hydantoin-containing templates can be caused by the specific interactions that the extrahelical bases form with the enzyme.[11] To do so, we first checked whether SuperScript III has point mutations compared to the parent MMLV RT in proximity to the region where the extrahelical bases would be positioned. The Protein Data Bank (PDB) contains only crystal structures of MMLV RT without its substrate.[12] However, there is a crystal structure of RT-substrate complex for an RT from a Xenotropic murine leukemia virus-related virus (XMRV), a virus derived from MMLV. The crystal structure of XMRV RT[13] (4HKQ in PBB) was aligned to the

1L-X 3′-GCGUCACGGUACUGCCU**X**CACUGGACCUCGCGGAG
2L-X 3′-GCGUCACGGUACUGCCU**X**ACACUGGACCUCGCGGAG
2LG-X 3′-GCGUCACGGUACUGCCU**X**GCACUGGACCUCGCGGAG
2LU-X 3′-GCGUCACGGUACUGCCU**X**AUACUGGACCUCGCGGAG
3L-X 3′-GCGUCACGGUACUGCCU**X**AACACUGGACCUCGCGGAG

X = OG

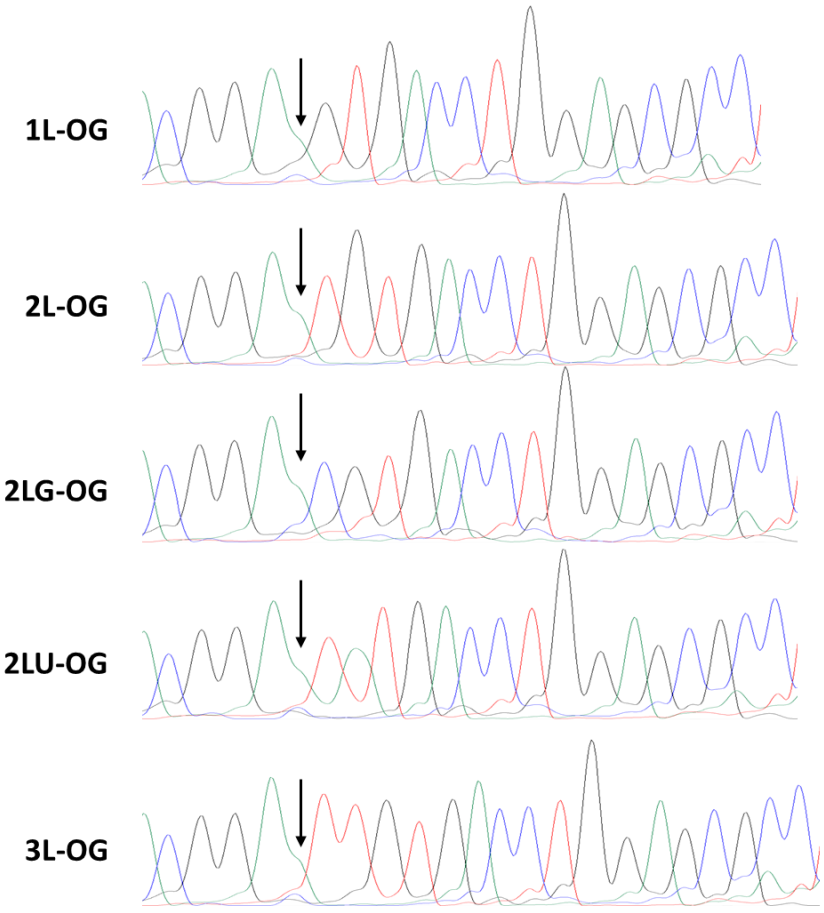


Figure 4.19. Sanger sequencing data for RT-PCR products originating from reverse transcription of OG-containing templates with P80 primer. Black arrow indicates position occupied by OG in the template.

1L-X 3′-GCGUCACGGUACUGCCU**X**CACUGGACCUCGCGGAG
2L-X 3′-GCGUCACGGUACUGCCU**X**ACACUGGACCUCGCGGAG
2LG-X 3′-GCGUCACGGUACUGCCU**X**GCACUGGACCUCGCGGAG
2LU-X 3′-GCGUCACGGUACUGCCU**X**AUACUGGACCUCGCGGAG
3L-X 3′-GCGUCACGGUACUGCCU**X**AACACUGGACCUCGCGGAG

X = OG

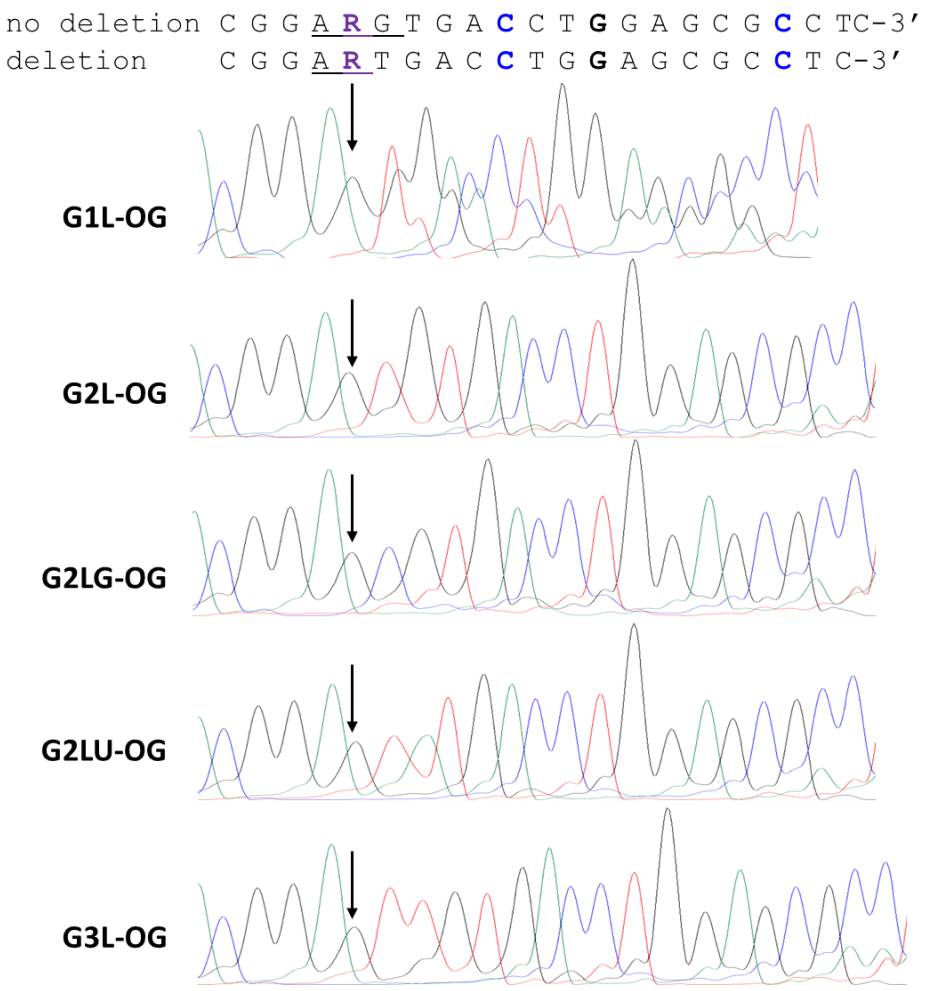


Figure 4.20. Sanger sequencing data for RT-PCR products originating from reverse transcription of OG-containing templates with P82G primer. Black arrow indicates position occupied by OG in the template.

crystal structure of MMLV RT[12] (4MH8 in PDB) and we found these two proteins to have identical structures and sequences in the region of interest (positions 116-127 and 190-197 in both proteins, Figure 4.21 A). Positions of point mutations introduced in SuperScript III RT compared to MMLV RT are known from the patent.[14] Three of these point mutations are in close proximity to the region where extrahelical bases would reside during the initial primer slippage required for deletion bypass (Figure 4.21 B). Next, we conducted RT-PCR amplification of the 2L-Gh template using MMLV RT at 37 °C for RT. When RT-PCR products were resolved on the gel we detected a significant amount of deletion bypass product only when G was inserted opposite Gh (Figure 4.21 C). Apart from the absence of the anomalous product of deletion bypass when A was inserted opposite Gh, the overall efficiency of deletion bypass was lower for MMLV RT. These data confirm our hypothesis that deletion bypass of Sp and Gh is driven in part by interaction of extrahelical bases and enzyme active site and show that point mutations introduced in SuperScript III enzyme promote deletion bypass at least in the case of templates containing hydantoin lesions. Considering that RT with MMLV RT was conducted at lower temperature (40 °C instead of 45 °C), we investigated whether temperature affects efficiency of deletion bypass. The 2L-Gh template was used for this experiment. We found that the ratio between n-2 and full-length products in the studied temperature range (35-50 °C) changes only when A is inserted and deletion bypass requires formation of A-C base pair (Figure 4.22). At the same time, the fraction of deletion product remained unchanged when G is inserted opposite Gh.

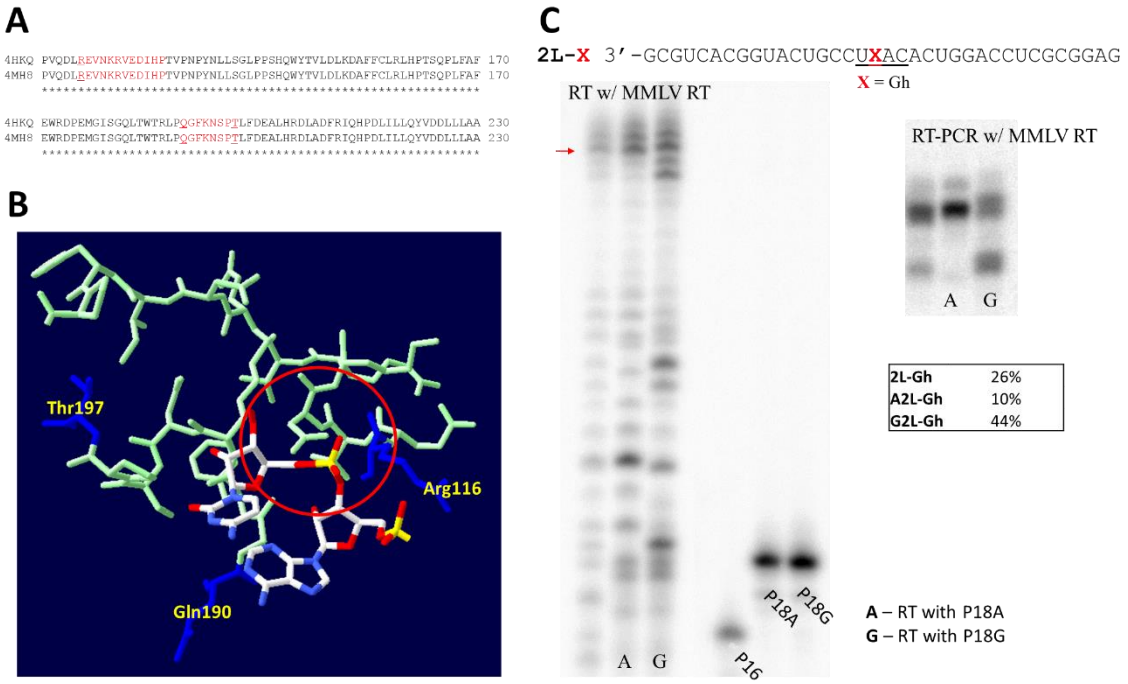
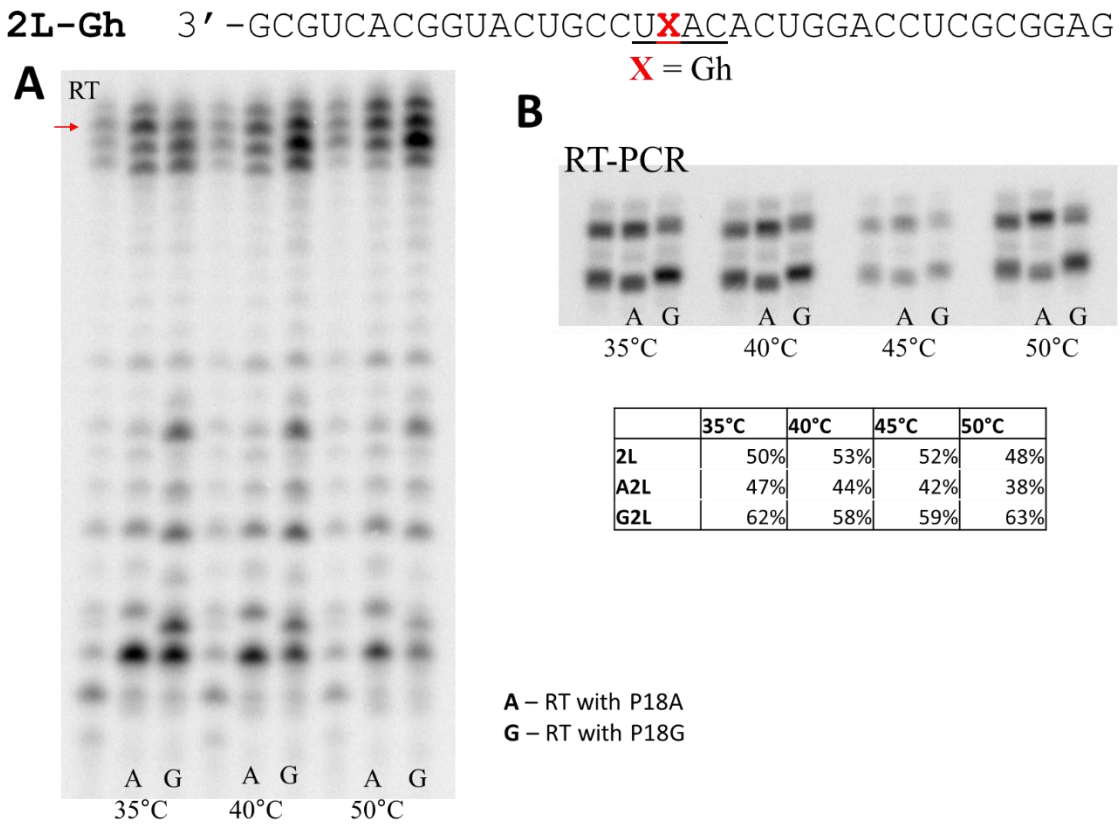


Figure 4.21. Dependence of deletion bypass on reverse transcriptase. **(A)** Fragment of aligned sequences of XMRV RT (4HKQ)[13] and MMLV RT (4MH8).[12] **(B)** Fragment of crystal structure of XMRV RT surrounding presumed position of extrahelical bases after the initial primer slippage (red circle).[13] Positions of mutated amino acids in SuperScript III are marked in blue in the crystal structure and underlined in aligned sequences. **(C)** PAGE analysis of RT and RT-PCR products obtained by using MMLV RT. Fractions of deletion products are given in the table below RT-PCR gel.



	35°C	40°C	45°C	50°C
2L	50%	53%	52%	48%
A2L	47%	44%	42%	38%
G2L	62%	58%	59%	63%

Figure 4.22. Dependence of deletion bypass efficiency on the temperature. **(A)** PAGE analysis of RT products at different reaction temperatures. Red arrow indicates position of full extension product on the gel. **(B)** PAGE analysis of RT-PCR products at different RT reaction temperatures. Fractions of deletion products are given in the table below RT-PCR gel.

Conclusions

In this chapter, we laid groundwork for mapping oxidized lesions in RNA. First, we did the required preliminary work for quantification of the amount of different nucleotides inserted opposite OG, Gh, and Sp with Illumina sequencing. These data would be important for estimation of sensitivity of this method and its further optimization. Next, we confirmed the previously proposed mechanism of deletion bypass of Gh and Sp in template strand and studied this phenomenon. The frequency of deletion bypass appeared to be highly dependent on the sequence context and on the identity of the RT enzyme, while unaffected by reaction temperature. Understanding of this phenomenon would be crucial for analysis of sequencing signature of Gh and Sp lesions. However, the frequency at which deletion bypass occurs can significantly complicate analysis of NGS data and hinder our ability to map hydantoin lesions. In addition to that, we also confirmed the identity of the bases inserted opposite OG (primarily A with some C), Gh, and Sp (G and A) using Sanger sequencing.

Future directions

The work presented in this dissertation was aimed at mapping several products of guanine oxidation in RNA. Several important steps were taken toward reaching this goal. We identified that RT enzymes recognize these lesions and can insert canonical DNA bases opposite them. This produces localized characteristic mutation signature in cDNA that can be amplified using standard RT-PCR protocols. And PCR amplification allows working with significantly smaller amount of RNA extracted from cells. The next goal of this project is attempting to map OG in RNA. All three guanine oxidation products studied here

are very low-abundance base modifications. Among them OG occurs with the highest frequency (one OG per ~10^6 nts in DNA[15-16] and 10^5-10^6 nts in RNA[15, 17-18]; one Gh or Sp per ~10^8 nts in DNA[15]) and thus presents an easier target for detection. Mapping positions of such infrequent base modification present significant challenge and can only be accomplished by using NGS methods.[19-20] Prior to sequencing, we would have to optimize the conditions for two steps preceding sequencing library preparation, fragmentation of cellular RNA, and enrichment of the strands containing the OG (Figure 4.23).

There are three main ways of cutting RNA into random fragments of ~150 nts that can be analyzed by Illumina sequencing: endonuclease digestion, mechanical shearing, and chemical cleavage. The first one uses RNAse III digestion of RNA, fragments generated by this method are 3' and 5' phosphorylated that makes them ready for adapter ligation without additional processing. The drawback of this method is that enzymatic cleavage is biased and leads to gaps and pile-ups (unequal representation of specific regions) in the final sequencing data.[21] The second approach uses an acoustic or hydrodynamic wave to break DNA or RNA apart. Mechanical shearing is typically a method of choice for DNA fragmentation; its main appeal is that it allows production of uniform fragments with minimal bias and without additional purification steps or risk of contaminating sequencing samples with the enzymes or chemicals that can negatively impact the sequencing read quality. However, mechanical shearing leads to local formation of free radicals and hydrogen peroxide that can introduce oxidative damage at the ends of RNA or DNA fragments.[22] Due to this fact, DNA repair enzymes are typically included in the enzyme mixtures aimed at production of blunt-end dsDNA fragments before adapter ligation; the extent of oxidation can also be reduced by adding antioxidants to the DNA or RNA solution

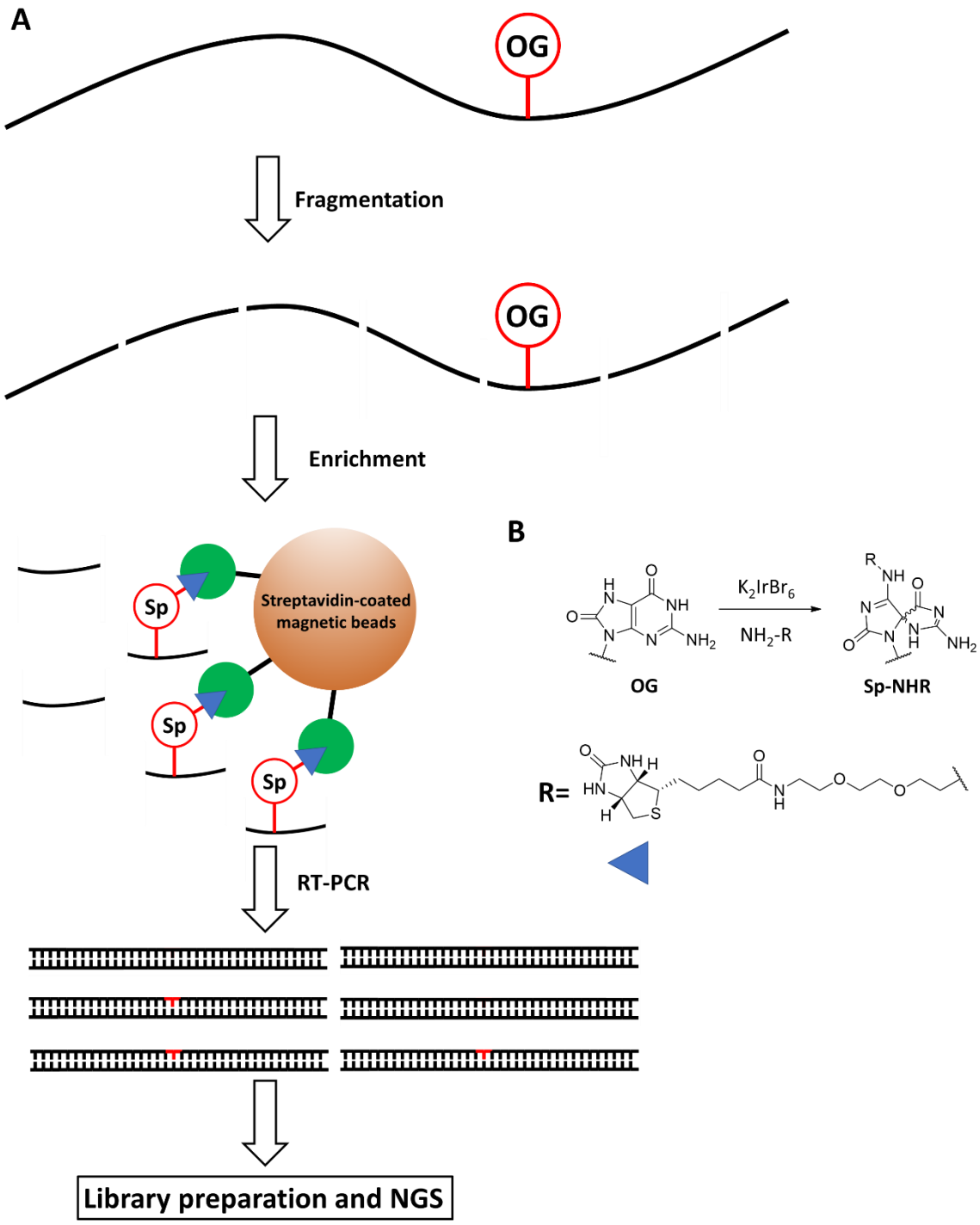


Figure 4.23. Application of OG-Seq for mapping oxidation sites in RNA. **(A)** Schematic representation of OG-Seq workflow adopted for RNA sequencing. **(B)** Selective oxidation of OG in the presence of primary amine-terminated biotin.[23]

during fragmentation.[23] In addition to that, mechanical fragmentation method is not completely unbiased; its efficiency depends on sequence and, more importantly for RNA, secondary structure.[24] The last method exploits the susceptibility of ssRNA to backbone cleavage in solutions of divalent metal ions, typically RNA fragmentation is done at 70-95 °C in the presence of magnesium or zinc.[25] Currently most RNA Illumina sequencing is done using metal ion cleavage. This approach allows cleaving RNA above its melting temperature removing potential bias for single-stranded regions. The negative aspect of this method is deamination of RNA bases (primarily C) induced at high temperature.[26] Before applying this method to OG-containing RNA we would have to optimize conditions for fragmentation (time, temperature, and concentrations) and subsequent purification to minimize sample losses and number of sequencing errors introduced in this step.

Enrichment of RNA strands containing the modification of interest is usually accomplished by immunoprecipitation with specific antibodies. However, it has been shown that antibodies that are currently available for OG do not have sufficient binding affinity or selectivity for effective enrichment.[23] An alternative enrichment approach is using nonantibody proteins that can bind OG or chemical ligation followed by affinity purification (AP). Several proteins are known to bind OG in RNA: polynucleotide phosphorylase (PNPase),[27-28] Y box-binding protein 1 (YB-1),[29] splicing factor 3B subunit 4 (SF3B4),[30] heterogeneous nuclear ribonucleoprotein D0 (HNRNPD),[30] heterogeneous nuclear ribonucleoprotein C (HNRNPC),[30] DAZ-associated protein 1 (DAZAP1),[30] and ELAV-like protein 1 (ELAVL1).[30] Affinity or selectivity of OG binding by any of these proteins is unknown and has to be tested if we decide to use them for enrichment. AP of OG-containing oligonucleotides has been described before.[23] In this approach OG is

selectively oxidized to an electrophilic intermediate that then is trapped by a primary amine nucleophile present in the reaction mixture.[31] In this case, OG in RNA is converted to the biotin-adduct of Sp that allows selective extraction by AP (Figure 4.23). Before applying this method to RNA, we would have to study the behavior of RT enzymes when they encounter bulky Sp-biotin adducts and optimize RT reaction conditions.

For interpretation of sequencing data, we would additionally need to quantify OG in RNA. This step is required because of the potential underestimation of the OG amount in RNA due to bias of RNA fragmentation (even chemical method of RNA fragmentation has been found to have some bias of unknown origin)[21] and RT (as we have shown in Chapter 3 introduction of OG, Gh, or Sp in RNA decreases efficiency of RT). We have considered two methods that potentially can be used for this purpose. The first one utilizes a similar approach to the one described for AP, where OG is selectively oxidized in the presence of a primary amine that has a fluorophore attached to it.[31] In this method, cellular RNA is immobilized on protamine-coated surface and after oxidation and subsequent elution of unbound fluorophore amount of OG in the sample is measured based on fluorescence intensity of bound fluorophore. The second method is an LC-MS approach where OG is quantified upon digestion of RNA to nucleosides.[15]

Another direction of the future work is attempting to use different RT enzymes. Two main concerns about using SuperScript III for future work is its relatively high error rate (1 in 32,000 bases)[32] and the absence of consolidated sequencing signature for Gh and Sp due to the prevalence of deletion bypass. Considering the low abundance of OG (below SuperScript III error rate) and potential necessity of mapping it as an adduct of Sp, these drawbacks can be detrimental for the success of sequencing OG in RNA. A potential

solution for both problems is using high fidelity RT enzymes like AccuScript, Transcriptor, PrimeScript, PyroScript, or TGIRT. The fidelity of RT can also be increased by addition of DNA 3'-5' exonucleases. This approach imitates proofreading during DNA replication, and its validity has to be tested for RT past oxidized lesions.[33]

The most obvious target for mapping OG is ribosomal RNA (rRNA). Unlike DNA, about 90% of cellular RNA is consolidated in just a few rRNA molecules (5S, 16S, and 23S in prokaryotes and 5S, 5.8S, 18S, and 28S in eukaryotes).[34] This means that there are up to 10 million copies of each of rRNA in mammalian cells,[35] and up to 70 thousand copies in bacteria.[36] Such high abundance would greatly simplify isolation of the amount of RNA required for sequencing. Additionally, rRNA has been found to be one of the major targets of iron-mediated oxidation in neuron cells, which makes it a promising candidate for mapping the distribution of OG.[37] Therefore, the first RNA target to sequence for OG by an NGS approach should focus on rRNA.

References

1. Fleming, A. M.; Muller, J. G.; Dlouhy, A. C.; Burrows, C. J., Structural context effects in the oxidation of 8-oxo-7,8-dihydro-2'-deoxyguanosine to hydantoin products: electrostatics, base stacking, and base pairing. *J. Am. Chem. Soc.* **2012,** *134* (36), 15091-15102.

2. Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W., NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **2012,** *9* (7), 671-675.

3. Schindelin, J.; Rueden, C. T.; Hiner, M. C.; Eliceiri, K. W., The ImageJ ecosystem: an open platform for biomedical image analysis. *Mol. Reprod. Dev.* **2015,** *82* (7-8), 518-529.

4. Beaucage, S. L.; Reese, C. B., Recent advances in the chemical synthesis of RNA. *Curr. Protoc. Nucleic Acid Chem.* **2009,** *Chapter 2*, Unit 2 16 11-31.

5. Kornyushyna, O.; Burrows, C. J., Effect of the oxidized guanosine lesions spiroiminodihydantoin and guanidinohydantoin on proofreading by *Escherichia coli* DNA polymerase I (Klenow fragment) in different sequence contexts. *Biochemistry* **2003,** *42* (44), 13008-13018.

6. Kornyushyna, O.; Berges, A. M.; Muller, J. G.; Burrows, C. J., *In vitro* nucleotide misinsertion opposite the oxidized guanosine lesions spiroiminodihydantoin and guanidinohydantoin and DNA synthesis past the lesions using *Escherichia coli* DNA polymerase I (Klenow fragment). *Biochemistry* **2002,** *41* (51), 15304-15314.

7. Singh, J.; Su, L.; Snow, E. T., Replication across O^6-methylguanine by human DNA polymerase beta *in vitro*. Insights into the futile cytotoxic repair and mutagenesis of O^6-methylguanine. *J. Biol. Chem.* **1996,** *271* (45), 28391-28398.

8. Andricioaei, I.; Goel, A.; Herschbach, D.; Karplus, M., Dependence of DNA polymerase replication rate on external forces: a model based on molecular dynamics simulations. *Biophys. J.* **2004,** *87* (3), 1478-1497.

9. Tse, W. T.; Forget, B. G., Reverse transcription and direct amplification of cellular RNA transcripts by Taq polymerase. *Gene* **1990,** *88* (2), 293-296.

10. Boulard, Y.; Cognet, J. A.; Gabarro-Arpa, J.; Le Bret, M.; Sowers, L. C.; Fazakerley, G. V., The pH dependent configurations of the C.A mispair in DNA. *Nucleic Acids Res.* **1992,** *20* (8), 1933-1941.

11. Wilson, R. C.; Pata, J. D., Structural insights into the generation of single-base deletions by the Y family DNA polymerase Dbh. *Mol. Cell* **2008,** *29* (6), 767-779.

12. Das, D.; Georgiadis, M. M., The crystal structure of the monomeric reverse transcriptase from moloney murine leukemia virus. *Structure* **2004,** *12* (5), 819-829.

13. Nowak, E.; Potrzebowski, W.; Konarev, P. V.; Rausch, J. W.; Bona, M. K.; Svergun, D. I.; Bujnicki, J. M.; Le Grice, S. F.; Nowotny, M., Structural analysis of monomeric retroviral reverse transcriptase in complex with an RNA/DNA hybrid. *Nucleic Acids Res.* **2013,** *41* (6), 3874-3887.

14. Smith, M.; Potter, R. J.; Dhariwal, G.; Gerard, G. F.; Rosenthal, K. Thermostable reverse transcriptases and uses thereof. U.S. patent 7078208 B2, July 18, 2006.

15. Mangerich, A.; Knutson, C. G.; Parry, N. M.; Muthupalani, S.; Ye, W.; Prestwich, E.; Cui, L.; McFaline, J. L.; Mobley, M.; Ge, Z.; Taghizadeh, K.; Wishnok, J. S.; Wogan, G. N.; Fox, J. G.; Tannenbaum, S. R.; Dedon, P. C., Infection-induced colitis in mice causes dynamic and tissue-specific changes in stress response and DNA damage leading to colon cancer. *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109* (27), E1820-E1829.

16. Ohno, M.; Miura, T.; Furuichi, M.; Tominaga, Y.; Tsuchimoto, D.; Sakumi, K.; Nakabeppu, Y., A genome-wide distribution of 8-oxoguanine correlates with the preferred regions for recombination and single nucleotide polymorphism in the human genome. *Genome Res.* **2006,** *16* (5), 567-575.

17. Hofer, T.; Badouard, C.; Bajak, E.; Ravanat, J. L.; Mattsson, A.; Cotgreave, I. A., Hydrogen peroxide causes greater oxidation in cellular RNA than in DNA. *Biol. Chem.* **2005,** *386* (4), 333-337.

18. Wu, J.; Li, Z., Human polynucleotide phosphorylase reduces oxidative RNA damage and protects HeLa cell against oxidative stress. *Biochem. Biophys. Res. Commun.* **2008,** *372* (2), 288-292.

19. Thomson, J. P.; Hunter, J. M.; Meehan, R. R., Deep C diving: mapping the low-abundance modifications of the DNA demethylation pathway. *Genome Biol.* **2013,** *14* (5), 118.

20. Li, X.; Xiong, X.; Yi, C., Epitranscriptome sequencing technologies: decoding RNA modifications. *Nat. Methods* **2016,** *14* (1), 23-31.

21. Lee, C.; Harris, R. A.; Wall, J. K.; Mayfield, R. D.; Wilke, C. O., RNaseIII and T4 polynucleotide kinase sequence biases and solutions during RNA-seq library construction. *Biol. Direct* **2013,** *8*, 16.

22. Milowska, K.; Gabryelak, T., Reactive oxygen species and DNA damage after ultrasound exposure. *Biomol. Eng.* **2007,** *24* (2), 263-267.

23. Ding, Y.; Fleming, A. M.; Burrows, C. J., Sequencing the mouse genome for the oxidatively modified base 8-oxo-7,8-dihydroguanine by OG-Seq. *J. Am. Chem. Soc.* **2017,** *139* (7), 2569-2572.

24. Poptsova, M. S.; Il'icheva, I. A.; Nechipurenko, D. Y.; Panchenko, L. A.; Khodikov, M. V.; Oparina, N. Y.; Polozov, R. V.; Nechipurenko, Y. D.;

Grokhovsky, S. L., Non-random DNA fragmentation in next-generation sequencing. *Sci. Rep.* **2014,** *4*, 4532.

25. Cloonan, N.; Forrest, A. R.; Kolle, G.; Gardiner, B. B.; Faulkner, G. J.; Brown, M. K.; Taylor, D. F.; Steptoe, A. L.; Wani, S.; Bethel, G.; Robertson, A. J.; Perkins, A. C.; Bruce, S. J.; Lee, C. C.; Ranade, S. S.; Peckham, H. E.; Manning, J. M.; McKernan, K. J.; Grimmond, S. M., Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat. Methods* **2008,** *5* (7), 613-619.

26. Lindahl, T.; Nyberg, B., Heat-induced deamination of cytosine residues in deoxyribonucleic acid. *Biochemistry* **1974,** *13* (16), 3405-3410.

27. Hayakawa, H.; Kuwano, M.; Sekiguchi, M., Specific binding of 8-oxoguanine-containing RNA to polynucleotide phosphorylase protein. *Biochemistry* **2001,** *40* (33), 9977-9982.

28. Wu, J.; Jiang, Z.; Liu, M.; Gong, X.; Wu, S.; Burns, C. M.; Li, Z., Polynucleotide phosphorylase protects *Escherichia coli* against oxidative stress. *Biochemistry* **2009,** *48* (9), 2012-2020.

29. Hayakawa, H.; Uchiumi, T.; Fukuda, T.; Ashizuka, M.; Kohno, K.; Kuwano, M.; Sekiguchi, M., Binding capacity of human YB-1 protein for RNA containing 8-oxoguanine. *Biochemistry* **2002,** *41* (42), 12739-12744.

30. Hayakawa, H.; Fujikane, A.; Ito, R.; Matsumoto, M.; Nakayama, K. I.; Sekiguchi, M., Human proteins that specifically bind to 8-oxoguanine-containing RNA and their responses to oxidative stress. *Biochem. Biophys. Res. Commun.* **2010,** *403* (2), 220-224.

31. Xue, L.; Greenberg, M. M., Facile quantification of lesions derived from 2'-deoxyguanosine in DNA. *J. Am. Chem. Soc.* **2007,** *129* (22), 7010-7011.

32. Potter, J.; Zheng, W.; Lee, J., Thermal stability and cDNA sythesis capability of SuperScript III reverse transcriptase. *Focus* **2003,** *25* (1), 19-24.

33. Arezi, B.; Hogrefe, H. H., Escherichia coli DNA polymerase III epsilon subunit increases moloney murine leukemia virus reverse transcriptase fidelity and accuracy of RT-PCR procedures. *Anal. Biochem.* **2007,** *360* (1), 84-91.

34. Fang, N.; Akinci-Tolun, R., Depletion of ribosomal RNA sequences from single-cell RNA-sequencing library. *Curr. Protoc. Mol. Biol.* **2016,** *115*, 7 27 21-27 27 20.

35. Gorlich, D.; Mattaj, I. W., Nucleocytoplasmic transport. *Science* **1996,** *271* (5255), 1513-1518.

36. Fegatella, F.; Lim, J.; Kjelleberg, S.; Cavicchioli, R., Implications of rRNA operon copy number and ribosome content in the marine oligotrophic ultramicrobacterium

Sphingomonas sp. strain RB2256. *Appl. Environ. Microbiol.* **1998,** *64* (11), 4433-4438.

37. Honda, K.; Smith, M. A.; Zhu, X.; Baus, D.; Merrick, W. C.; Tartakoff, A. M.; Hattier, T.; Harris, P. L.; Siedlak, S. L.; Fujioka, H.; Liu, Q.; Moreira, P. I.; Miller, F. P.; Nunomura, A.; Shimohama, S.; Perry, G., Ribosomal RNA in Alzheimer disease is oxidized by bound redox-active iron. *J. Biol. Chem.* **2005,** *280* (22), 20978-20986.