Influence of DNA Sequence Context on O^6-Methylguanine Mutagenicity in *Escherichia coli*

by

James Cullen Delaney

B.S. and M.S. Chemistry
Yale University, 1992

SUBMITTED TO THE DEPARTMENT OF CHEMISTRY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN BIOLOGICAL CHEMISTRY
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 1999

Signature of Author: Signature redacted
Department of Chemistry
May 21, 1999

Certified by: Signature redacted
John M. Essigmann
Professor of Chemistry and Toxicology
Thesis Supervisor

Accepted by: Signature redacted
Dietmar Seyferth
Chairman, Departmental Committee on Graduate Studies

Thesis Committee

Professor Steven R. Tannenbaum Signature redacted
Chairperson

Professor John M. Essigmann Signature redacted
Thesis Supervisor

Professor Gerald N. Wogan Signature redacted

Influence of DNA Sequence Context on O^6-Methylguanine Mutagenicity in *Escherichia coli*

by

James Cullen Delaney

Submitted to the Department of Chemistry on May 21, 1999
in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy in Biological Chemistry

ABSTRACT

The study of how mutational hotspots induced by alkylating agents arise is complicated by the intertwining of several variables. The selective formation or repair of an alkyl DNA lesion such as O^6-methylguanine (m^6G) can, in principle, be influenced by surrounding nucleotide sequence context. Similarly, the infidelity of a DNA polymerase may be influenced by the sequence context surrounding the adduct. Thus, the creation of a non-uniform mutational landscape could be due to any one or a combination of a number of variables. The work described in this thesis introduces new methodology to deconvolute mutational spectra.

In the first part of this thesis research, a novel method is described whereby the mutation frequency (MF) of any site-specific DNA lesion exhibiting targeted point mutations can be determined rapidly and accurately, regardless of the local surrounding base sequence. Many simulated MFs were used to validate the assay (named REAP for Restriction Endonuclease And Postlabeling determination of MF) and it was demonstrated that both the REAP assay and a more traditional method of individually counting different colors of plaques (for one sequence context) yielded statistically identical MFs. Furthermore, triplicate measurements of each simulated MF using the REAP assay were strikingly reproducible. Since the REAP assay does not require phenotypic selection, it is therefore free of biases that might be so introduced, and should work in both single- and double-stranded vector genomes in *Escherichia coli* (*E. coli*) as well as in mammalian cells.

In the second part of this work, a nearest-neighbor, site-specific approach was used to evaluate sequence context effects on the repair of m^6G by the *E. coli* methyltransferases Ada and Ogt, and the infidelity of the DNA polymerase holoenzyme in copying past this modified base *in vivo*. A novel technique was used to construct sixteen individual single-stranded M13 genomes with m^6G immediately flanked by each permutation from the set of G, A, T, and C bases. Each member of the array was electroporated into wild-type and isogenic nucleotide excision repair-deficient *E. coli* that lacked the Ogt, Ada, or both methyltransferases. The biological processing of the lesion in each sequence context was determined by measuring the MFs using the validated REAP assay. When an *ada*, *ogt*, *uvrB* cell line was used, m^6G in all sequence contexts was nearly 100% mutagenic, suggesting that the DNA polymerase III holoenzyme always placed a T opposite m^6G. Consequently, it was concluded that the sequence context surrounding m^6G does not cause the DNA polymerase to create mutational hot- or coldspots. In an *ogt*, *uvrB* cell line, the MFs ranged from 35% to 90%. Within this range, attributed to the effect of sequence context on m^6G repair by Ada, a 5′ sequence context influence was evident. The order of increasing MF attributed to Ada was, in general, GX < CX < TX < AX (X = m^6G). Ada also showed a 3′ influence of XPyrimidine < XPurine. By contrast, in an *ada*, *uvrB* cell line, the MFs ranged from 10% to 25%. Ogt demonstrated a 5′ influence of GX < AX, and the contexts 5′-AXN-3′ (N = any base) provided the highest MFs. The MFs with respect to sequence context in the wild-type cell line (Ada, Ogt, and UvrB all operative) revealed a pattern strikingly similar to that obtained using *ada*, *uvrB E. coli*, suggesting that Ogt is the dominant repair element for m^6G in uninduced *E. coli*.

In the third part of this work, two original proposals are presented that will hopefully contribute further to the deconvolution of mutational spectra. The first describes a unified method whereby the effects of sequence context on lesion repair, adduct formation, or adduct degradation can be assessed *in vitro*. The second describes an assay based on mass spectrometry (MS) whereby the repair of any DNA lesion, whose mutational specificity is known, can be studied in 16 to 256 sequence contexts simultaneously *in vivo*. This mass spectrometry approach may also be used to obtain consensus sequences surrounding a DNA lesion that are blocks to a DNA polymerase *in vivo*, and may be additionally used to determine in what sequence contexts a lesion is more apt to miscode by the replicative DNA polymerase *in vivo*.

Thesis Supervisor: Professor John M. Essigmann

Title: Professor of Chemistry and Toxicology

ACKNOWLEDGMENTS

This thesis is dedicated to my parents, Thomas and Adeline Delaney, whose resources of love, support, food, and shelter have given me the freedom to follow my academic interests. I would like to thank foremost my thesis advisor, John Essigmann, for giving me the freedom to, in a sense, carry out my own research program and also for planting the seeds of some good ideas in my head. My gratitude goes to him and the members of my thesis committee, Professor Steven Tannenbaum and Professor Gerald Wogan, whose interest in this work is greatly appreciated. I also thank Dr. John "Pete" Wishnok for a critical reading of Chapter 4C (concerning MS) and Dr. Paul Henderson for a critical reading of Chapter 4B. I am indebted to Professor Peter Lobel, who introduced me to the wonderful world of molecular biology when I worked in his lab at the Center for Advanced Biotechnology and Medicine during the summer vacations of my bright college years. Thanks goes to my undergraduate advisor, Professor Donald M. Crothers, for giving me the freedom to use my developing organic synthesis skills for solving biochemical problems. I also express my gratitude to Professor Harry H. Wasserman, who had the extraordinary ability of making organic chemistry fun and easy to understand. I thank Drs. Edward A. Ryan and J. Freilich for reattaching my retina so that this work could be seen to completion. Finally, I'd thank you to stop reading this section and move on.

"*It's not fun if it's already been done.*"
Jim Delaney

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

5′ dAMP	2′-deoxyadenosine 5′-monophosphate
5′ dCMP	2′-deoxycytidine 5′-monophosphate
5′ dGMP	2′-deoxyguanosine 5′-monophosphate
5′ dm^3GMP	2′-deoxy-3-methylguanine 5′-monophosphate
5′ dm^6GMP	2′-deoxy-*O*^{6}-methylguanine 5′-monophosphate
5′ dm^7GMP	2′-deoxy-7-methylguanine 5′-monophosphate
5′ dNMP	2′-deoxyribonucleotide 5′-monophosphate
5′ dRiboseMP	2′-deoxyribose 5′-monophosphate (and other abasic 5′-monophosphate derivatives)
5′ dTMP	thymidine 5′-monophosphate
8MOP	8-methoxypsoralen
8oxoG	7,8-dihydro-8-oxoguanine
α-2′-dA	α-2′-deoxyadenosine
A	adenine or adenosine
AAF	2-(*N*-acetoxy-*N*-acetylamino)fluorene
Ada	*O*^{6}-methylguanine-DNA methyltransferase
AF	2-acetylaminofluorene
$AFB_1{}^7G$	8,9-dihydro-8-(N^7-guanyl)-9-hydroxyaflatoxin B_1
AQO	4-acetoxy-aminoquinoline-1-oxide
ATP	adenosine 5′-triphosphate
BPDE	benzo[*a*]pyrene diol-epoxide
BrCN	cyanogen bromide
bz^6G	*O*^{6}-benzylguanine
C	cytosine or cytidine
CCNU	1-(2-chloroethyl)-3-cyclohexyl-1-nitrosourea
ce^6G	*O*^{6}-2-chloroethylguanine
CE	capillary gel electrophoresis
cis-DDP	*cis*-diamminedichloroplatinum(II)
CPG	controlled pore glass
D	deuterated sample
dATP	2′-deoxyadenosine 5′-triphosphate
dCTP	2′-deoxycytidine 5′-triphosphate
dm^4TTP	2′-deoxy-*O*^{4}-methylthymine 5′-triphosphate
dm^6G	2′-deoxy-*O*^{6}-methylguanine
dm^6GTP	2′-deoxy-*O*^{6}-methylguanine 5′-triphosphate
DMT	4,4′-dimethoxytrityl
DNA	2′-deoxyribonucleic acid
dNTP	2′-deoxyribonucleotide 5′-triphosphate

DTT	dithiothreitol
ϵA	1,N^6-ethenodeoxyadenosine
e^2T	O^2-ethylthymine
e^4T	O^4-ethylthymine
e^6G	O^6-ethylguanine
E. coli	*Escherichia coli*
EDC	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride
EDTA	ethylenediaminetetraacetic acid
EMS	ethyl methanesulfonate
ENNG	*N*-ethyl-*N*′-nitro-*N*-nitrosoguanidine
ENU	*N*-ethyl-*N*-nitrosourea
G	guanine or guanosine
H	non-deuterated sample
he^4T	O^4-2-hydroxyethylthymine
he^6G	O^6-2-hydroxyethylguanine
HPLC	high pressure liquid chromatography
I	inosine
Im	imidazole
ipr^6G	O^6-isopropylguanine
IPTG	isopropyl β-D-thiogalactopyranoside
kb	kilobases
m^4T	O^4-methylthymine
m^5C	5-methylcytosine
m^6G	O^6-methylguanine
m^7G	7-methylguanine
MF	mutation frequency
MMR	mismatch repair
MMS	methyl methanesulfonate
MNNG	*N*-methyl-*N*′-nitro-*N*-nitrosoguanidine
MNU	*N*-methyl-*N*-nitrosourea
MS	mass spectrometry
N	G, A, T, and C simultaneously present at position N
NER	nucleotide excision repair
NMR	nuclear magnetic resonance
NONP	1-nitroso-8-nitropyrene
O	abasic site
Ogt	DNA methyltransferase II
oh^2A	2-hydroxyadenine
oh^5C	2′-deoxy-5-hydroxycytidine
oh^5U	2′-deoxy-5-hydroxyuridine
oh^8G	7,8-dihydro-8-oxoguanine
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction

PdG	1,N^2-propanodeoxyguanosine
PEI	polyethyleneimine
P-gal	phenyl-β-D-galactopyranoside
PhIP	2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine
Pol I, II, or III	DNA polymerase I, II or III
pol	DNA polymerase
PNNG	*N*-propyl-*N'*-nitro-*N*-nitrosoguanidine
Pu	purine (G or A)
Py	pyrimidine (T or C)
RE	restriction endonuclease
REAP	Restriction Endonuclease And Postlabeling determination of mutation frequency
RNA	ribonucleic acid
RNase A	ribonuclease A
SA	streptavidin
SAM	*S*-adenosylmethionine
SSM	site-specific methodology
SVPD	snake venom phosphodiesterase I
T	thymine or thymidine
TE	Tris(hydroxymethyl)aminomethane ethylene-diaminetetraacetic acid buffer
TLC	thin layer chromatography
Tm	melting temperature
U	uracil or uridine
UV	ultraviolet
X	lesion or base to be monitored
X-gal	5-bromo-4-chloro-3-indolyl β-D-galactopyranoside

CHAPTER 1:

O^6-Methylguanine Mutagenesis

1A. Background

The improvements made by the medical profession over the past two hundred years, combined with increased public awareness of how disease is spread and hence, how to prevent it, has prolonged human life. Instead of dying from smallpox and other infectious diseases, we can now expect to die from heart failure or cancer, the "grim reapers" at the end of the road for modern day humans. The heart, while an essential part of the body, is just a single organ and one can envision biomedical engineers fusing their knowledge of biomimetic tissue with the electrical patterns that make for proper cardiac function to provide a synthetic heart, should the original fail. However, the eradication of cancer is seemingly more problematic, since every cell in the body has the capability of undergoing uncontrolled cell growth. Instead of dealing with the treatments and possible systemic cure for the many different types of cancer in the various organs of the body, this thesis explores the impact of DNA sequence context within the realm of the somatic theory of mutation as a prelude to cancer. More precisely, to determine if the immediate sequence context flanking a carcinogenic DNA adduct can impact its repair and replication fidelity and thus, the mutation frequency at the lesion site.

The early 1950s was a great time of discovery in which the landscape of DNA was surveyed and mapped. A high quality X-ray diffraction pattern of B form DNA from Rosalind Franklin (Franklin and Gosling, 1953) and a somewhat more mediocre pattern from Wilkins (Wilkins et al., 1953) were produced. Watson and Crick pieced together these X-ray diffraction

patterns of B form DNA and, utilizing the observation by Chargaff that "in all desoxypentose nucleic acids examined thus far the molar ratios of total purines to total pyrimidines, and also of adenine to thymine and of guanine to cytosine, were not far from 1" (Chargaff, 1950), a structural model of DNA was born (Watson and Crick, 1953b). Indeed, their model proposed in 1953 led Watson and Crick to speculate that DNA is replicated semi-conservatively and allowed them to hypothesize that base-pairing infidelity during DNA synthesis, such as that exhibited by a minor tautomeric form, could lead to a heritable mutation (Watson and Crick, 1953a). The Watson-Crick model paved the way for Loveless in 1969 to postulate, after finding that O^6-methylguanine (m^6G) can be formed by reaction of deoxyguanosine with *N*-methyl-*N*-nitrosourea (MNU) under physiological conditions *in vitro*, that m^6G may be mutagenic (Loveless, 1969). It should be noted that Friedman *et al.* had also detected the formation of m^6G *in vitro* six years prior to the revelation by Loveless; however, deoxyguanosine was treated with diazomethane in an organic methanol-ether system (Friedman et al., 1963).

Following the *in vitro* formation and detection of m^6G, further analytical evidence supporting m^6G formation *in vivo* by alkylating agents was provided by Frei, who revealed that DNA extracted from mice treated with [^{14}C]-MNU contains [^{14}C]-m^6G (Frei, 1971). The first direct evidence supporting the mispairing idea of Loveless came from Gerchman & Ludlum in 1973 when it was shown that an RNA polymerase can misincorporate uracil into RNA from a DNA template containing randomly situated m^6G residues (Gerchman and Ludlum, 1973). Further work using poly(dC-dG) DNA templates that were globally modified with MNU showed that *E. coli* DNA polymerase I can misincorporate thymine (Abbott and Saffhill, 1979). The

proof that m^6G is mutagenic resulted from site-specific mutagenesis experiments in which m^6G was placed at a specific location in a viral genome (Green et al., 1984) and was found to elicit G to A mutations at that site (Loechler et al., 1984). Taken together, the experiments stated above provide evidence that a DNA modifying agent may lead to a heritable mutation. If this mutation activates an oncogene or inactivates a tumor suppressor gene, then cancer tumor growth may occur.

There is substantial evidence that cellular DNA becomes modified after treatment with a chemical mutagen, such as MNU. The modification, if not repaired prior to DNA replication, can elicit a mutation. Although studies such as those involving the Ha-*ras* oncogene have shown that the DNA sequence of an activated oncogene has a point mutation after being treated with an alkylating agent (Sukumar et al., 1983; Zarbl et al., 1985), one should be cautious in abandoning the epigenetic theory of carcinogenesis altogether, since preexisting mutants may have been selected for in the aforementioned studies (Cha et al., 1994). Even so, preexisting mutants must have been created somehow. It has been well established that an O^6-alkylguanine DNA lesion, such as m^6G, can form by exposure of cells to exogeneous agents, such as MNU and *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (MNNG), and the presence of this lesion correlates with the observed increase in G:C to A:T transitions that occurs in *E. coli* following treatment with these and other alkylating agents (Coulondre and Miller, 1977). These types of mutations also occur spontaneously and the endogenously formed *S*-adenosylmethionine (SAM) has been shown to form alkylated guanine residues (Barrows and Magee, 1982; Rydberg and Lindahl, 1982). SAM is a cofactor used in the maintenance methylation pattern of cellular DNA. Rydberg and Lindahl

found that SAM has an alkylation pattern similar to that of the S_N2 alkylator, methyl methanesulfonate (MMS) forming mostly 7-methylguanine and some 3-methyladenine, but the formation of m^6G was less than 1% of 7-methylguanine and was probably around the background level of the assay. Since DNA is being constantly exposed to this alkylator, a low rate of m^6G formation undetectable by *in vitro* analytical means may give rise to *in vivo* mutations. Although SAM has been touted as a good candidate for endogenous methylation of the O^6 position of guanine, a better source is the nitrosation of amides and amines (Taverna and Sedgwick, 1996; Sedgwick, 1997). The Taverna and Sedgwick experiment used *E. coli* that lacked the Ada and Ogt methyltransferases, which demonstrated an increased spontaneous mutation rate with respect to wild-type. Cells that additionally lack Moa, which makes the cells defective in the molybdopterin guanine dinucleotide cofactor required by nitrate reductases, lowers both the spontaneous mutation rate and methylamine-plus-nitrite induced mutagenesis, thus implicating the cofactor-using enzymes as both the nitrosating and methylating agent.

It is thus not surprising that *E. coli* and mammals have evolved DNA repair proteins to combat chemical insults to their genomes. The proposed mechanism for the removal of O^6-alkylation damage to guanine or O^4-alkylation damage to thymine by the *E. coli* methyltransferases Ada and Ogt is shown in Figure 1. During the direct reversal of base damage, the alkyl group is transferred stoichiometrically to an active site cysteine residue, thus forming *S*-alkylcysteine and normal guanine or thymine (Lindahl et al., 1982). The N-terminal domain can also remove the S (but not R) stereochemical configuration of a methylphosphotriester lesion from the DNA backbone (Hamblin and Potter, 1985). This also is the trigger for the adaptive

response discovered in 1977 by Leona Samson and John Cairns when they challenged *E. coli* with a low dose of MNNG followed by a higher dose, and found the cells to have acquired resistance to both mutagenesis and death (Samson and Cairns, 1977). In this process, Ada becomes a transcriptional activator of itself as well as other DNA repair proteins in the Ada regulon (the *alkA*, *alkB*, and *aidB* genes). Since the active site cysteine is not regenerated after alkyltransfer from the base (Lindahl et al., 1982) or from the DNA backbone (McCarthy and Lindahl, 1985), this class of repair systems has been given the moniker "suicide enzyme." This type of repair is obviously non-enzymatic and Ada and Ogt therefore shall be referred to as repair proteins. Since the X-ray structure of Ada has been elucidated (Moore et al., 1994), it would be interesting to see if these structural data can explain the set of nearest-neighbor rules generated for the repair of m^6G by Ada in *E. coli* as described in this dissertation.

Human-made compounds that form m^6G, such as MNNG and MNU, have been placed into the S_N1 category of chemical reactivity. They are expected to be less sensitive to the substrates on DNA with which they react, giving the less nucleophilic sites on DNA, namely the O^6 and O^4 positions of guanine and thymine, respectively, a chance to be alkylated. For example, MNU demonstrates a unimolecular reaction mechanism with its substrate and has been placed into the S_N1 category of alkylators having the low Swain-Scott value of $s = 0.42$ (Velemínský et al., 1970). Salmon sperm DNA treated with MNU at neutral pH *in vitro* forms the alkylated bases 7-methylguanine, 3-methyladenine, and O^6-methylguanine in the ratios of 82:10:8, respectively (results from two experiments were normalized; the similar ratios were averaged and renormalized) (Lawley and Shah, 1973b; Lawley et al., 1973a). However, dimethyl sulfate, with

a higher Swain-Scott value of $s = 0.86$ (Osterman-Golkar et al., 1970) demonstrates more bimolecular S_N2 character and thus, reacts with DNA more selectively forming adducts at only the most nucleophilic sites. In contrast with MNU treatment, salmon sperm DNA treated with dimethyl sulfate at neutral pH *in vitro* forms the alkylated bases 7-methylguanine, 3-methyladenine, and O^6-methylguanine in the ratios of 83:16:< 1, respectively (Lawley and Thatcher, 1970). One may reasonably speculate that nearest-neighbor bases surrounding a particular guanine to be alkylated may affect the extent of alkylation if the alkylating species were large enough to sense the presence of the flanking bases. This may occur if an S_N2 alkylator or the precursor of an S_N1 alkylating species is targeted to a "hotspot" on DNA, where it then reacts directly with DNA or becomes activated, respectively. It is tempting to consider this possibility for MNNG, since it is activated by thiols. If glutathione or a protein with an exposed cysteine residue has a particular affinity for a DNA sequence, then that DNA sequence may be a hotspot for MNNG induced mutagenesis. While thiol activation has been demonstrated for MNNG (Lawley and Thatcher, 1970), it has not been shown for MNU (Wheeler and Bowdon, 1972). This speculation does not seem to hold true however for MNNG and MNU, since the same hotspots and coldspots are observed in an *in vivo* system (Koch et al., 1994). A precovalent mode of binding exhibited by bulky intercalating DNA alkylators, such as aflatoxin (Benasutti et al., 1988) is probably more apt to show nearest-neighbor sequence selectivity based on steric grounds.

Bases flanking a reactive guanine may have, in addition to a steric component, an electronic impact on reactivity toward alkylating agents. There is evidence that the 5′-most

guanine in a 5′-GG-3′ sequence is the most readily oxidized (Saito et al., 1995), correlating with theoretical studies demonstrating that the 5′-most guanine has the lower ionization potential and encompasses most of the highest occupied molecular orbital (Saito et al., 1995; Sugiyama and Saito, 1996). While it is tempting to speculate that a guanine that has a lower affinity for its electrons may be more reactive with an electropositive "methyl cation" into which its electrons are donated, this hypothesis does not correlate with the mutagenicity of S_N1 alkylating agents. The mutational hotspots are usually created in the second G of a 5′-PuG-3′ sequence context, not at the 5′-most guanine. A possible melding of oxidation potential and second guanine alkylation may have been provided unwittingly by Buckley in 1987. He proposed that the O^6 of the 5′ guanine may first attack the imidourea (and not the carbonyl (Buckley, 1987b)) form of an alkylnitrosourea (incidentally, the tetrahedral addition product is the same for both cases). Following this initial attack, the O^6 of the 3′ guanine receives the alkyl group, and water and nitrogen gas are released. Hydrolysis of the carbamate at the O^6 of the 5′ guanine finalizes the conversion of 5′-GG-3′ to 5′-Gm^{6}G -3′ (Buckley, 1987a).

Structural studies of m^{6}G paired opposite T or C demonstrate the miscoding nature of the lesion. A mutagenic base pairing can occur between m^{6}G and thymine (Figure 2A and 2B), as well as a non-mutagenic base pairing between m^{6}G and cytosine (Figure 2C and Figure 2D). The structures found in Figure 2 were obtained using X-ray crystallography (Figure 2A) (Leonard et al., 1990), NMR (Figure 2B) (Kalnik et al., 1989b), NMR (Figure 2C) (Kalnik et al., 1989a), and X-ray crystallography (Figure 2D) (Ginell et al., 1990). In Figure 2D, it is possible that the N1 of m^{6}G and N3 of C were close to each other, but not hydrogen bonded (Ginell et al., 1990).

Although one may think that the protonated G:C base pair depicted in Figure 2D would not be seen at physiological pH, arguments for its existence have been proposed (Williams and Shaw, 1987). Since the pKa of an amino acid in a protein can be modulated by its environment so that the most thermodynamically stable system will result, the same phenomena may be seen for DNA; that is, the pKa of the N3 position of cytosine may be increased to allow for better hydrogen-bonding and stacking of the protonated pair. This in turn may lead to cross-strand deamination in which the original m^6G:C or m^6G:m^5C pair becomes an m^6G:U or m^6G:T pair, respectively (Williams and Shaw, 1987).

The relative orientation of the base pairs and the orientation of the methyl group in Figure 2A and Figure 2C deserve discussion. The m^6G:T pair shown in Figure 2A retains the normal Watson-Crick pairing with the methyl group in an *anti* conformation with respect to N1 of m^6G, while the m^6G:C pair in Figure 2C exists as a wobble with the methyl group *syn* to N1 of m^6G. Using molecular modeling, Loechler showed that the *syn* conformation of an m^6G:T pair is slightly more favored (<~1 kcal/mol) (Loechler, 1991). Although the structures depicted in Figure 2A and Figure 2C have only two hydrogen bonds each, the extra stacking afforded by the *syn* orientation in Figure 2C should make the m^6G:C pair more thermodynamically stable than the m^6G:T pair. This indeed has been seen in melting temperature studies (Gaffney et al., 1984; Gaffney and Jones, 1989). The question that follows is, if an m^6G:C pair is more thermodynamically stable than an m^6G:T pair, why is m^6G in DNA mutagenic? A hypothesis was presented by Swann, who suggests that the more Watson-Crick like geometry of an m^6G:T pair over that of an m^6G:C pair might facilitate in both the formation of the 5′ and 3′

phosphodiester bonds of the incoming base opposite the lesion and in the escape of exonucleic proofreading (Swann, 1990). This hypothesis was supported by Tan *et al.* who demonstrated that the rate limiting step for nucleotide incorporation opposite m^6G by Klenow was the formation of the phosphodiester bond, and that bond formation was faster for T than C (Tan et al., 1994).

O^6-Methylguanine is thought to elicit mutations by providing a surface complementary to that of an incoming thymine during DNA synthesis (Loveless, 1969), and this type of mutation has been demonstrated in *E. coli* (Loechler et al., 1984) using plasmids that contain only m^6G at a specific position within the vector (Green et al., 1984). If m^6G in a DNA template strand always mispairs with T during DNA synthesis, a G to A transition mutation will be generated wherever an m^6G lesion is formed. Assuming 100% miscoding by the replicative DNA polymerase, the frequency of this mutation would range from 100% in the absence of repair, to 50% if repair occurs after one round of DNA synthesis, to 0% if the lesion is repaired before the next round of semi-conservative DNA synthesis.

The aforementioned argument begs the question: where does the m^6G lesion come from? Treatment of *E. coli* with MNNG produces predominantly G:C to A:T transition mutations in the *lacI* gene (Coulondre and Miller, 1977). This and other experiments that exhibit a similar mutational specificity for S_N1 alkylators are mentioned countless times as evidence that the transitions are created by alkylation of guanine in DNA strands rather than alkylation of the DNA precursor pool. The argument against nucleotide pool alkylation as the cause to mutations is usually stated as follows. During replication, an dm^6GTP can pair opposite T (Hall and Saffhill,

1983). If the proofreading exonuclease of the DNA polymerase is eluded and a methyltransferase converts the T:m^{6}G pair to a T:G pair that eludes mismatch repair, an overall T:A to C:G transition would result instead of the observed G:C to A:T transition of Coulondre and Miller. However, a G:C to A:T transition could arise from the nucleotide precursor pool since dm^{4}TTP can pair opposite a template G during DNA synthesis (Hall and Saffhill, 1983). The G:C to A:T transition may result if the G:m^{4}T pair is acted upon by a methyltransferase and the subsequent G:T pair eludes mismatch repair. Indeed, the nucleotide precursor pool would be expected to be more susceptible to alkylation, since the nucleophilic heteroatoms are not tied up in interstrand hydrogen-bonding. The estimation of the pool/DNA methylation ratio for dm^{6}GTP/m^{6}G in a DNA polynucleotide has been calculated to be ~400 (Topal and Baker, 1982); however, to my knowledge, it has not been calculated for dm^{4}TTP/m^{4}T in a DNA polynucleotide. A strong argument against mutagenesis from the nucleotide precursor pool can be made if the probability of formation and incorporation of dm^{6}GTP, and methyltransferase repair of the resulting m^{6}G exceeds that of dm^{4}TTP and m^{4}T (which would provide the unseen T:A to C:G transitions). It has been estimated by the Lawley group that *in vitro* reaction of salmon sperm DNA with the S_N1 alkylator MNU produces 70-fold more m^{6}G than m^{4}T (Lawley and Shah, 1973b). Since the O^4 of T and O^6 of G are both in the major grove and are hydrogen-bonded to an exocyclic amine, one may expect that the 70-fold bias in DNA polynucleotides would also exist in the free nucleotides. Theoretical studies support this showing that m^{6}G as a nucleotide monomer is more likely than m^{4}T to be alkylated due to both the more favorable electrostatic potential and steric accessibility (Pullman and Pullman, 1981). Although the assumption of a minimum of 70 dm^{6}GTP per dm^{4}TTP is probably therefore true, dm^{4}TTP is 40 times more likely

than dm^{6}GTP to be inserted opposite poly (dC-dG) or poly (dA-dT), respectively by DNA polymerase I (Pol I) (Hall and Saffhill, 1983). An approximate 2 to 1 bias in favor of insertion of m^{6}G would then exist favoring nucleotide precursor pool mutagenesis, which is not entirely convincing for ruling out G:C to A:T transitions *via* dm^{4}TTP. However, m^{4}T is not well repaired with respect to m^{6}G by the eukaryotic methyltransferases of human and yeast (Sassanfar et al., 1991) or human and rat (Zak et al., 1994), of which G:C to A:T transitions are observed with alkylating agents; and when similar site-specifically modified plasmids harboring an m^{6}G and m^{4}T lesion are introduced into wild-type *E. coli*, m^{6}G is repaired at least 10-fold better than m^{4}T, thus further diminishing the importance of nucleotide pool alkylation. Furthermore, Brennand *et al.* showed that treatment of Chinese hamster V79 cells with O^4-methylthymidine results in the incorporation of m^{4}T into the DNA; however, the lesion is not mutagenic, further suggesting poor repair (Brennand et al., 1982). Finally, one could make the simple argument that alkylation damage in DNA strands has more of a chance to miscode, since the hurdle of incorporation of an alkylated nucleotide triphosphate by a DNA polymerase and the escape of 3′ exonuclease proofreading does not have to be overcome. This thesis will therefore explore only the effect of sequence context on DNA strand alkylation and not the possibility of dm^{4}TTP being incorporated with a probability that varies with flanking template bases, as was suggested and shown for dm^{6}GTP (Toorchen and Topal, 1983; Topal et al., 1986).

1B. Mutations Tend to be Distributed Non-randomly

While the gross structure of DNA was surveyed and mapped in the early 1950s, a lot of work emanating from Seymour Benzer's lab in the late 1950s and early 1960s did much to reveal how the molecular landscape within genes changes when introduced to different chemical agents. Benzer compiled data from the same genetic target treated with different chemicals. His data clearly demonstrated that a different mutational spectrum is seen for each type of ultimate chemical species that reacts covalently with DNA (Benzer, 1961). A corollary to this conclusion is that, for a given mutagen, induced mutations are not randomly distributed. Assuming the hierarchy of adduct formation, adduct repair, replication past unrepaired adduct yields a mutation, Occam's razor (i.e., the simplest explanation is the correct one) predicts that non-uniform adduct formation would result in mutational spectra being non-uniform. Of course, an adduct must be formed to create a lesion that can be mis-replicated. Indeed, a few studies looked at this phenomenon at the adduct formation level without the complexities of repair and replication. In 1985, Briscoe and Cotter (Briscoe and Cotter, 1985) had treated duplex DNA polynucleotides with tritiated MNU, depurinated the DNA using mild acid hydrolysis, and separated the methylated products by HPLC. The m^6G/G ratio increased in the order 5′-TGT-3′ < 5′-AGA-3′ < 5′-CGC-3′ < 5′-GGG-3′, with a two-fold variation in range. A study by Dolan *et al.* in which a few self-complementary dodecamer oligonucleotides were made and allowed to react with tritiated MNU, and the depurinated methylated bases analyzed by HPLC, revealed that the m^6G/G ratio increased in the order 5′-ACGCGT-3′ < 5′-ACCGGT-3′ < 5′-AGGCCT-3′

(Dolan et al., 1988). O^6-Methylguanine is formed about four times higher in the sequence 5′-AGGCCT-3′ than in the sequence 5′-ACGCGT-3′. This suggests that a guanine in the sequence 5′-CGC-3′ and/or 5′-CGT-3′ is four-fold less likely to be methylated at the O^6 of guanine than in the sequence 5′-AGG-3′ and/or 5′-GGC-3′. Furthermore, m^6G is formed in the sequence 5′-ACTAGT-3′ twice as much as in the sequence 5′-ACATGT-3′, suggesting that a guanine is better *O*-alkylated when flanked 5′ by A rather than T (the 3′ base was T in both cases). In 1989, Richardson *et al.* (Richardson et al., 1989) made four oligonucleotides of the same sequence 5′ $CCG_1TG_2G_3G_4ATATGGGCGT$ 3′ using solid-phase DNA synthesis employing a guanine phosphoramidite (tritiated at the 1′ and 2′ positions of the deoxyribose ring) at the subscripted positions. By allowing each oligonucleotide containing a single tritiated deoxyguanosine at a different guanine position to react with the methylating agent, MNU, and analyzing the enzymatically released deoxynucleosides by HPLC radiochromatography, the order of increasing m^6G formation in duplex DNA (m^6G:C) was found to be: $G_1 < G_2 < G_4 \sim G_3$, with a five-fold variation in range. The corresponding nearest-neighbor translation is: 5′-CGT-3′ < 5′-TGG-3′ < 5′-GGA-3′ ~ 5′-GGG-3′. In 1991, Sendowski and Rajewsky (Sendowski and Rajewsky, 1991) used polynucleotides and 24-mers containing guanines in triplet sequences repeated eight times to provide all possible permutations flanking guanine. Duplexes were treated with MNU and m^6G formation was monitored by a competitive radioimmunoassay against the enzymatically liberated m^6G nucleoside (Müller and Rajewsky, 1980). The analysis for triplets containing two guanines and for triplets that have a guanine in each non-self-complementary strand is complicated. For example, m^6G formation for a 5′-GGAGGA-3′ sequence may be the result of 5′-GGA-3′, 5′-AGG-3′, or both. A 5′-CGCCGC-3′ sequence

could promote m^6G formation by flanking cytosines, or by the triplets 5′-CGG-3′ and/or 5′-GGC-3′ in the complementary strand 5′-GCGGCG-3′. Therefore, only those sequences that contain triplets harboring a single guanine and no cytosines in the "target" strand are presented. The order of increasing formation of m^6G (the m^6G/G ratio) is 5′-TGT-3′ ~ 5′-TGA-3′ < 5′-AGT-3′ ~ 5′-AGA-3′ < 5′-GGG-3′ (from poly(dG):poly(dC)), with an eight-fold variation in range. Although the trends of alkylation hotspots for m^6G formation seem to correlate well for both the Briscoe and Sendowski experiments (5′-TGT-3′ < 5′-AGA-3′ < 5′-GGG-3′), a clear interpretation of isolated nearest-neighbor effects for all contexts on m^6G formation is precluded because of the ways in which the experiments were performed (Dolan et al., 1988; Sendowski and Rajewsky, 1991). This complication is due to more than one guanine in the "target" strand flanked by a different permutation of bases in an oligonucleotide, and/or by a guanine in the strand complementary to the "target," which could also be methylated. While the method of Richardson *et al.* can be adapted so that 16 oligonucleotides with a central tritiated guanine flanked by all permutations of G, A, T, and C could be individually made, treated with MNU, and rank ordered for tritiated m^6G content, the process seems costly and lengthy. Perhaps the method I propose in Chapter 4B of this thesis will be applicable.

Technology that came out of the Essigmann lab in the early 1980's allowed placement of the putatively carcinogenic lesion, m^6G, into a defined position in an oligonucleotide (Fowler et al., 1982), placement of the oligonucleotide into a defined position within a vector (Green et al., 1984), and observation of the ultimate fate of the lesion by scoring the mutation frequency of progeny at the lesion site (Loechler et al., 1984). This site-specific approach has been adopted by

hundreds of laboratories for the purposes of identifying the coding properties of candidate mutagenic DNA lesions *in vivo* and *in vitro*, determining the kinetics of primer extension at and past DNA lesions, evaluating the repair of DNA lesions by purified DNA repair enzymes, and performing structural studies such as X-ray, NMR, and Tm. Sequence context experiments using this methodology, by definition, remove differential adduct formation as a cause of mutational hotspots, thus focusing on differential adduct repair and/or DNA polymerase fidelity.

The effect of sequence context on the repair of m^6G without the complication of preferential adduct formation or coding ambiguity by a DNA polymerase has also been studied for a few isolated sequences. In 1988, Dolan *et al.* found that self-complementary oligonucleotides of sequence 5′-Cm^6GC-3′ were repaired better than 5′-Gm^6GC-3′ by Ada and human HT 29 colon carcinoma cell extracts (Dolan et al., 1988). In 1991 Georgiadis *et al.* used purified Ada to measure rates of repair of m^6G derived from the Ha-*ras* sequence 5′ GGCGC TGGAGGCGTG 3′ *via* an immunoprecipitation assay (Georgiadis et al., 1991). A direct correlation was made between the rate of m^6G repair and the avidity constant from m^6G oligonucleotides reacting with an antiserum to dm^6G raised in rabbits, suggesting that the more accessible m^6G in DNA is to an alkyltransferase, the better it is repaired. The order of decreasing rates of duplex repair is 5′-GCm^6GAG-3′ > 5′-TTm^6GAG-3′ > 5′-CTm^6GGA-3′ > 5′-CTm^6GAA-3′ > 5′-TGm^6GAG-3′. This study also reveals that next-to-nearest-neighbors can influence repair, since the 5′-TTm^6GAG-3′ was ~4 times better repaired than 5′-CTm^6GAA-3′. In 1996, a marginally comprehensive study of sequence context effects on repair of an O^6-alkylguanine lesion was done by Bender *et al.*, in which duplexes containing O^6-ethylguanine

were allowed to react with human O^6-alkylguanine-DNA-alkyltransferase (Bender et al., 1996). Although as emphasized in the title of their paper, that the rate of repair did not exhibit "significant dependence on sequence context," the error of their analysis was rather large, thus precluding a rank order of repair rates.

The effect of sequence context on replication past m^6G was explored in two *in vitro* primer extension studies emanating from the Singer lab. One study reveals that DNA polymerase I (Klenow) has a small, but reproducible two- to three-fold preference of inserting T instead of C opposite m^6G in a 5′-Cm^6GC-3′ context, but equal preference in a 5′-Tm^6GT-3′ context (Dosanjh et al., 1991). The other study shows Klenow to have a six- to seven-fold preference of inserting T instead of C opposite m^6G in a 5′-Am^6GC-3′ context, while no preference exists in a 5′-Am^6GT-3′ context (Singer et al., 1989), implying that a single base change 3′ to m^6G can modulate the mutagenic response of Klenow. Taken together, these experiments demonstrate that the DNA polymerase itself can contribute to the non-uniformity of mutational spectra.

It has been suggested that DNA polymerase III (Pol III) is the DNA polymerase responsible for m^6G lesion bypass, since the MFs of gapped plasmids containing a site-specific m^6G in the single-stranded region in *E. coli* cells that were (Pol I^-), (Pol II^-), and (Pol I^-, Pol II^-), were all equal (~70%) (Pauly et al., 1995). This is interesting, since Pol I is known to fill in small gaps. As pointed out by Pauly *et al.*, this result could also arise if each of the three DNA polymerases used the "correct" dCTP 30% of the time. It is known, however, that different DNA

polymerases can yield different frequencies of dNTP incorporation opposite a DNA lesion in the same sequence context (Chary and Lloyd, 1995b; Shibutani et al., 1991). If this also holds true for m^6G, Pol III would be a more attractive candidate as the DNA polymerase that copies past m^6G in *E. coli*. One should always be cautious in extrapolating primer extension studies to the generation of mutational spectra *in vivo*, since the DNA polymerase used may not be what the cell uses to copy past the lesion. Even if the "correct" DNA polymerase is used, such as the DNA polymerase III holoenzyme in most cases, one can never be sure that all the accessory proteins are present, and in the correct ratios.

When a DNA lesion is placed site-specifically into a vector, the possibility of differential adduct formation as contributing to mutational hotspots is eliminated. Certainly, this does not mean that differential adduct formation should be dismissed. This approach does however, allow one to determine the combined effects of DNA repair and replication on lesion processing. The intertwining of repair and replication contributions can further be partitioned if parallel experiments are performed in repair-deficient cell lines. In 1986, Topal *et al.* used a 61-mer containing six m^6G residues to prime replication in the ampicillinase gene of a single-stranded phage genome (Topal et al., 1986). After ligation and transformation into *E. coli*, the order of increasing mutation frequency was 5′-Tm^{6}GA-3′ < 5′-Gm^{6}GG-3′ < 5′-Gm^{6}GA-3′ ~ 5′-Tm^{6}GA-3′ < 5′-Cm^{6}GT-3′ < 5′-Am^{6}GC-3′. The same scheme was employed with singly-modified oligonucleotides in which the order 5′-Gm^{6}GG-3′ < 5′-Cm^{6}GT-3′ was again established. In 1992, Pletsa *et al.* found that plasmids substituted with m^6G at G_1 or G_2 of codon 12 of the human Ha-*ras* gene 5′-CG_1G_2C-3′, when replicated in simian COS7 cells proficient in

alkyltransferase, have a higher mutation frequency at G_2 than G_1 (Pletsa et al., 1992). In 1994, they performed a similar experiment in *E. coli* using the Ha-*ras* gene and found reduced repair at G_2 (Pletsa et al., 1994).

Many more studies have been performed in which *E. coli* and plasmids (already in cells or later to be processed in cells) are treated with alkylating agents. While it is harder to extract the individual contribution of sequence context on adduct formation, repair, polymerase fidelity, and possible incorporation of alkylated dNTPs from these experiments with globally modified DNA, these studies reveal what the mutational hotspots and coldspots actually are, since all systems are operative in reality. In 1987, Richardson *et al.* treated *E. coli* harboring a plasmid containing the xanthine guanine phosphoribosyltransferase gene as the genetic target with MNU (Richardson et al., 1987a). It was found that ~80% of mutations were at the second G in 5′-GG(A/T)-3′ sequences. They performed a similar experiment that same year using MNNG and unadapted or adapted *E. coli* and found a similar bias for second guanine alkylation in the 5′-GG(A/T)-3′ sequence (Richardson et al., 1987b). In that same year, Burns *et al.* also treated *E. coli* with MNNG, and selected colonies that grew on phenyl-β-D-galactopyranoside (P-gal) as the only carbon source (*lacI*$^-$ mutants in the F′ episome) (Burns et al., 1987). The *lacI*$^-$ mutants were cloned *in vivo* into an M13 strain (*lacI*$^+$, *lacZ*$^-$) *via* phage infection and recombination, and recombinant *lacI*$^-$, *lacZ*$^+$ single-stranded M13 progeny that gave blue plaques on an indicator bacterial strain were sequenced. The order of increasing mutation frequency was 5′-CG-3′ ~ 5′-TG-3′ < 5′-AG-3′ < 5′-GG-3′. Although they mention that under their treatment conditions, the methyltransferase repair system would probably be fully saturated and inactivated (Schendel

and Robins, 1978), differential adduct formation and polymerase miscoding were still factors that could have contributed to differential mutation frequencies at different sites of the gene. In 1990, the same group restated their support of initial damage as the cause of hotspots when their experiment was repeated in three and analyzed in five strains of isogenic *E. coli* (wild-type (Burns et al., 1987), *recA* (Gordon et al., 1988), *umuC*, *uvrB*, and *polA*), and the same 5′ influence was obtained (Gordon et al., 1990); however, polymerase infidelity (and methyltransferase repair) were still formal variables. In 1993, Mironov *et al.* treated rat liver DNA with MNU, PCR amplified the Ha-*ras* region, and cloned the region into the replicative form of M13 from which sequencing reactions were performed (Mironov et al., 1993). They also found that the 3′ guanine in a 5′-GG-3′ sequence was most likely to be mutated. Approximately 70% of the mutations were found in 5′-GGA-3′ and 5′-GGT-3′ sequences, whereas none were found at 5′-AGG-3′, 5′-GGG-3′, 5′-TGT-3′, 5′-TGC-3′, 5′-CGA-3′, 5′-CGC-3′, and interestingly codon 12 of Ha-*ras* (5′-GGA-3′). While differential repair is not an issue, the hotspots could have arisen not only from adduct formation, but Taq polymerase infidelity with respect to bases flanking m^6G. This possibility also exists in a study by Hatahet *et al.* whereby a site-specific lesion within a randomized flanking sequence is amplified by PCR, subcloned, and sequenced (Hatahet et al., 1998).

The work coming out of the Pueyo lab has done much to advance the effect of sequence context on mutagenesis by alkylating agents and is therefore presented here separately. In 1993, Ariza *et al.* treated *uvrB*, *ada*, *ogt E. coli* with MNU and scored mutants using the *supF* target, finding that guanines flanked 5′ by a guanine were preferred sites for mutations (Ariza et al.,

1993). In 1994, Roldán-Arjona *et al.* again found most mutations in the second G of a 5′-GG-3′ site in *uvrB E. coli* carrying the *supF* plasmid after alkylation (Roldán-Arjona et al., 1994). This was true for four derivative strains that were (*uvrB*), (*uvrB*, *ada*), (*uvrB*, *ogt*), and (*uvrB*, *ada*, *ogt*) when treated with MNU, (*uvrB*, *ada*) and (*uvrB*, *ada*, *ogt*) when treated with *N*-ethyl-*N*-nitrosourea (ENU), and the (*uvrB*, *ada*, *ogt*) derivative when treated with ethyl methanesulfonate (EMS). In 1995, Vidal *et al.* also found that when *uvrB E. coli* harboring an F′ episome carrying the *lacI* gene was mutated by 100 mM EMS, the average number of mutations per site were, in increasing order, 5′-CG-3′ (1), 5′-GG-3′ (2), 5′-TG-3′ (3), and 5′-AG-3′ (11) (Vidal et al., 1995). It is interesting to note that when the Ogt^- strain was compared with the Ogt^+ strain after EMS treatment, G:C to A:T transition mutations doubled for the 5′-AGA-3′ context, tripled for 5′-TGC-3′, but dropped six-fold for the 5′-CGG-3′ sequence context. Since EMS demonstrates borderline S_N1/S_N2 character (Hudson and Withey, 1964) (s = 0.64 (Osterman-Golkar et al., 1970)), ethylation at O^6 of guanine can occur. It can be inferred from the data that Ogt works poorly on 5′-Ae^{6}GA-3′ and 5′-Te^{6}GC-3′ contexts, but well on 5′-Ce^{6}GG-3′ contexts. While Ada may also be a factor, its concentration is ~30 times lower than Ogt, and although Ada can transfer a chloroethyl group from one of the two chloroethylphosphotriester stereoisomers (Carter et al., 1988), induction of the adaptive response by ethylation rather than methylation of the N-terminal Ada domain is unknown. In 1995, Jurado *et al.* also used the *E. coli lacI* system to show the sequence specificity of 1-(2-chloroethyl)-3-cyclohexyl-1-nitrosourea (Jurado et al., 1995). When a *uvrB*, *ogt*, *ada* strain was compared with an isogenic *uvrB* strain (no methyltransferase vs. Ogt^+, Ada^+), a statistically significant decrease in mutagenesis at the central G in 5′-GGG-3′, 5′-GGA-3′, and 5′-TGC-3′ sequences was observed, suggesting that Ogt repairs

O^6-chloroethylguanine better when it is flanked 5′ by a guanine. It is also interesting that the two sites of highest mutation in *uvrB E. coli* possessing Ada and Ogt both had the 5′-AGA-3′ sequence context, while lower mutagenesis at these sites was observed in cells lacking Ada and Ogt, suggesting poor repair of this site by methyltransferases. In 1997, Vidal *et al.* used the same *lacI* system to see how mutations are distributed in Ogt^- and Ogt^+ *E. coli* when treated with ENU and MNU (Vidal et al., 1997). Upon treatment of *E. coli* with ENU or MNU, they found a statistically significant increase in MF in a 5′-AGA-3′ sequence context, going from a *uvrB*, *ogt* to a *uvrB* (Ada^+, Ogt^- to Ada^+, Ogt^+), or a *uvrB*, *ogt*, *ada* to a *uvrB*, *ada* line (Ogt^- to Ogt^+). They also found a statistically significant decrease in mutagenesis in a 5′-GGT-3′ sequence context. This result correlates with their findings in which the 5′-AGA-3′ context was refractory to repair, while a 5′ guanine enhanced Ogt repair of O^6-chloroethylguanine lesions (Jurado et al., 1995). It should also be noted that, of the sixteen possible permutations of bases flanking guanine, three cannot be assayed by the methodology used by the Pueyo group; two of these three sequences have a 5′ adenine.

The information in Table 1 demonstrates that a variety of chemicals and site-specifically introduced DNA lesions can exhibit a variety of mutational specificities. Moreover, a specific DNA lesion can be formed and/or repaired and replicated with varying fidelity with respect to DNA topology and/or sequence context. Indeed, even a single stereoisomer of a covalently bound DNA adduct may exhibit multiple conformations (possibly due to the surrounding sequence context) that may lead to different biological endpoints (Eckel and Krugh, 1994; Dosanjh et al., 1993; Rodriguez and Loechler, 1993a; Jelinsky et al., 1995). Although not

tabulated, there also exist sequences, such as triplet, doublet, and singlet repeats, where the repetitive nature of "normal" undamaged DNA can cause a DNA polymerase to generate frameshifts *via* primer-template slippage (Streisinger et al., 1966); both mutations (or microheterogeneity) in a DNA polymerase (Fowler et al., 1986; Strauss et al., 1997) and the lack of cellular mismatch-repair (targeted to small loops) (Schaaper and Dunn, 1987) may exacerbate frameshift mutagenesis. The reader is also directed to an excellent review that covers the effect of sequence context and DNA structure on adduct deposition from aflatoxin, alkylating agents, benzo[*a*]pyrene diol-epoxide (BPDE), mitomycin C, anthramycin, transition metal complexes, *N*-(bromoacetyl)distamycin, and CC-1065 (Warpehoski and Hurley, 1988). There are other lesions that experience mutational hotspots (Maccabee et al., 1994), but were omitted because nearest-neighbor rules could not be defined for adduct formation and/or mutational specificity. This demonstrates the caveat that for many lesions, more distant sequences flanking a lesion may influence its biological processing.

The information in Table 1 is a distillation from a broad survey of the effects of DNA sequence and topology on lesion formation, repair, and replication. The data are grouped by (putative) lesion and are listed in chronological order (alphabetically with respect to referenced author within each year). In many cases, placement of the data into the headline columns "formation," "repair," and "replication," is arbitrary, especially when non-site-specific studies employing repair-proficient genomes are performed. For non-site-specific studies, the "lesion" is that most likely to be responsible for the observed mutagenic effects listed in the "exhibits" column. In some instances, only the major event is indicated. Some entries also use the

headlines to describe the experiment concisely, giving such information as cellular repair and SOS translesion bypass status, but this will be obvious. Unless stated otherwise, each entry describes conditions (sequence contexts) that favor "formation," "repair," or "replication" and sequences are implied to be read 5′ to 3′. If an experiment did not obtain a MF, the "exhibits" column was left blank (for example, no G → T is entered for the Benasutti *et al.* reference in which the effect of sequence context on aflatoxin binding (covalent modification of the N7 position of G) was analyzed). Likewise, other columns were left blank if they did not impact the experiment described in the "reference" column. Descriptions separated by semicolons are linked to information in adjacent columns in the same semicolon register. Colons between bases represent a base pair. The superscript number refers to the position of base covalent modification, the symbol (^) represents a crosslink, and the position of modification and/or mutation is underlined. Treatment *via* SSM means that site-specific methodology was used.

In none of the studies listed in Table 1, or for that matter, anywhere, were the bases flanking a site-specific normal base or lesion systematically varied to study the effect of all permutations of nearest-neighbors on the formation, repair, or replication bypass of a DNA lesion *in vitro* or *in vivo*. The studies we performed do this at the level of m^6G repair by the Ada and Ogt methyltransferases and at the level of m^6G replication bypass, when vectors bearing the lesion are placed inside isogenic repair-deficient *E. coli* hosts. These studies necessitated the development of a method that would rapidly and accurately measure the MF of m^6G in any surrounding sequence context. The development of this assay and its validation are discussed in Chapter 2.

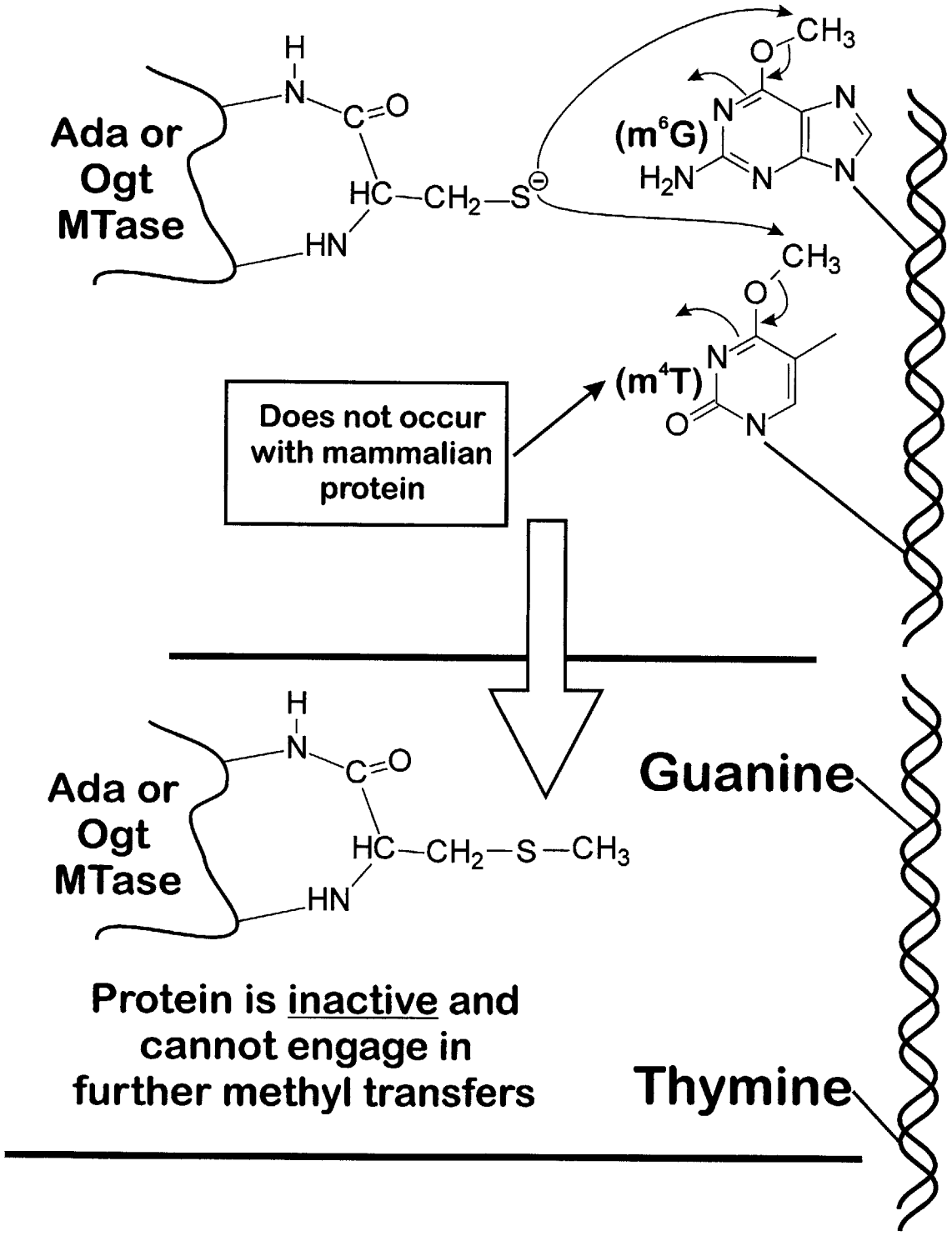


Figure 1. Proposed mechanism for direct reversal of m^6G or m^4T base damage by DNA alkyltransferase repair proteins. The UVR nucleotide excision repair system also can repair m^6G.

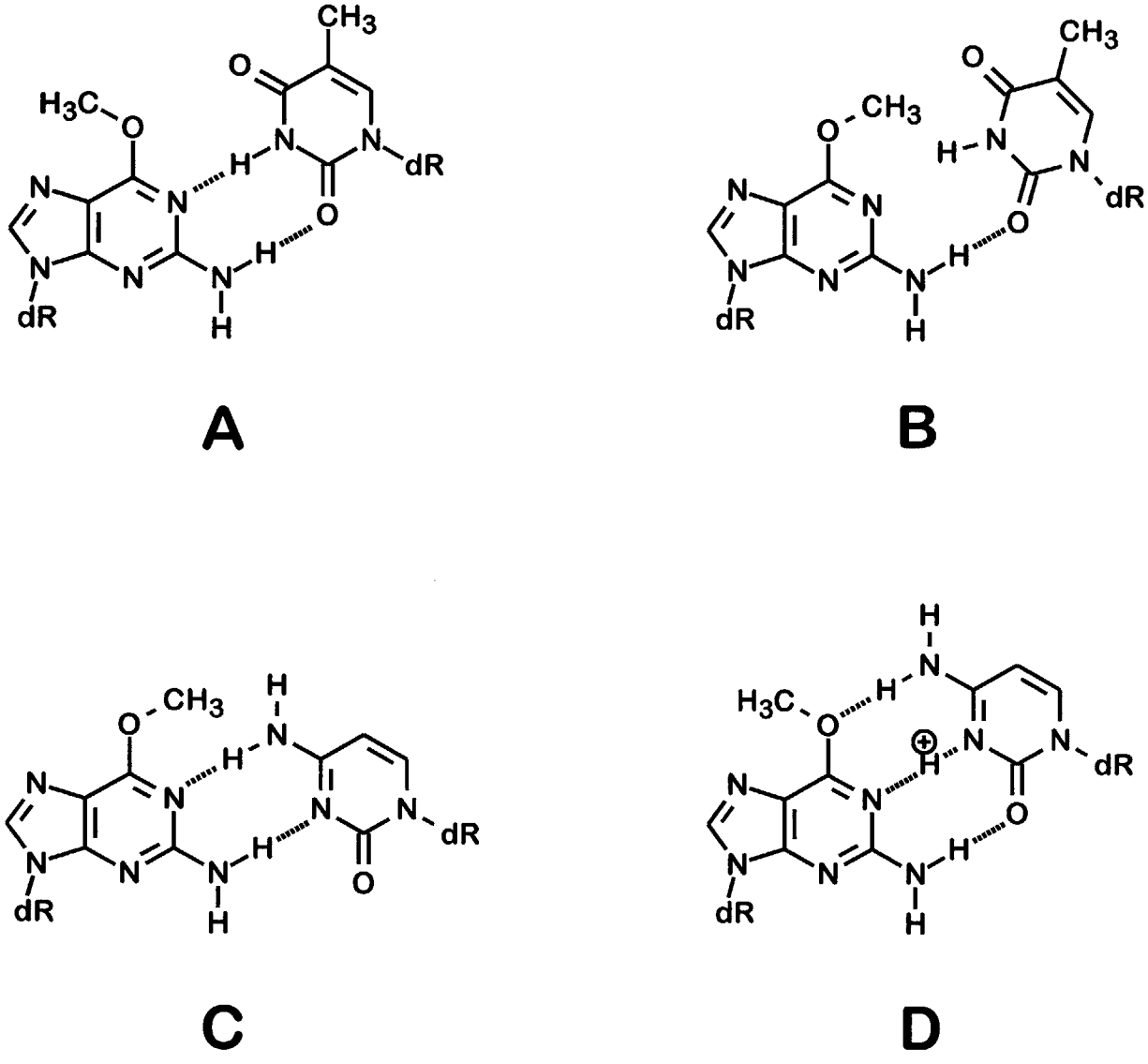


Figure 2. Proposed base pairing schemes from structural studies of m^6G:T pairs (A and B) and m^6G:C pairs (C and D).

Table 1: Experiments in which DNA sequence context and topology have been shown to influence DNA adduct formation, lesion repair, and polymerase miscoding. See pages 36-37 for key.

Lesion	Formed from	Treatment *via*	Exhibits	Influence of DNA structure or flanking base on:			Reference
				Formation	Repair	Replication	
m^6G	G	MNU poly(dG-dm^5C)	G → A		Z-form DNA, refractory by Ada		(Boiteux et al., 1985)
e^6G	G	ENU	G → A	low ionic strength, extended chromatin			(Nehls and Rajewsky, 1985)
e^6G	G	EMS	G → A		nucleotide excision repair (NER), flanking A:T pairs		(Burns et al., 1986)
m^6G	G	MNNG	G → A	G<u>G</u> & A<u>G</u> (5′ G > 5′ A)			(Burns et al., 1987)
m^6G	G	MNNG	G → A		adapted *E. coli*, transcribed strand		(Reed and Hutchinson, 1987)
m^6G; e^6G	G	MNU; ENU	G → A; G → A > A → G	MNU & ENU at G<u>G</u>(A/T)			(Richardson et al., 1987a)
m^6G	G	MNNG	G → A	G<u>G</u>(A/T) ± *E. coli* adaptation, non-transcribed strand			(Richardson et al., 1987b)
m^6G	G	MNU	G → A	Pu<u>G</u>			(Burns et al., 1988a)
e^6G; e^4T	G; T	ENU	G → A; > A → G	Pu<u>G</u>; or Pu<u>T</u>	both refractory to NER repair if flanking G:C pairs		(Burns et al., 1988b)
m^6G	G	MNU & SSM		A<u>G</u> > T<u>G</u> (MNU)	Cm^6G > Gm^6G by Ada (SSM)		(Dolan et al., 1988)
m^6G	G	MNNG	G → A	Pu<u>G</u> > Py<u>G</u>	± RecA		(Gordon et al., 1988)
m^6G	G	*N*-nitroso-*N*-methyl-*N*-alpha-acetoxybenzylamine	G → A	Pu<u>G</u> > Py<u>G</u>			(Horsfall and Glickman, 1988)

Lesion	Formed from	Treatment *via*	Exhibits	Influence of DNA structure or flanking base on:			Reference
				Formation	Repair	Replication	
e^6G	G	ENU		non-random (electron microscopy)			(Nehls et al., 1988)
e^6G; e^4T	G; T	*N*-ethyl-*N'*-nitro-*N*-nitrosoguanidine (ENNG)	G → A; > T → C ~ T → A	GG(A/T) non-transcribed strand; NTC (T → C)			(Richardson et al., 1988)
he^6G; he^4T	G; T	1-(2-hydroxyethyl)-1-nitrosourea	G → A; > T → C	GG(A/T) non-transcribed strand; NTC (T → C)			(Richardson et al., 1988)
e^6G	G	ENU	G → A ~ A → C			see A → C with SOS^+	(Eckert et al., 1989)
m^6G, e^6G, m^4T		SSM			Ada $m^6G >> e^6G > e^4T$ (ipr^6G not repaired)		(Graves et al., 1989)
m^6G	G	*N*-nitroso-*N*-methyl-*N*-alpha-acetoxymethylamine	G → A	PuG > PyG, some A → G, transversions, & frameshifts (dose dependent)			(Horsfall and Glickman, 1989a)
m^6G	G	*N*-nitroso-*N*,*N*-dimethylamine	G → A	PuG > PyG, some A → G & A → T			(Horsfall et al., 1989b)
m^6G; bz^6G		SSM	G → A (m^6G); G → A, C, T (bz^6G)	in ras codon 12 G_1G_2A, m^6G & bz^6G = mutagenicity at G_1 & G_2	implies not heavily influenced		(Mitra et al., 1989)
m^6G	G	MNU (specific residues could be monitored)		in $CCG_1TG_2G_3G_4AT$... duplex, G_3 & G_4 > G_1 & G_2			(Richardson et al., 1989)
m^6G		SSM				for m^6GC, m^6G:T > m^6G:C for m^6GT, m^6G:T ~ m^6G:C by Klenow and pol α	(Singer et al., 1989)
ipr^{6}G	G	*N*-propyl-*N'*-nitro-*N*-nitrosoguanidine (PNNG)	G → A > G → T ~ A → C	unlike MNNG, no neighboring base influence			(van der Vliet et al., 1989)

Lesion	Formed from	Treatment *via*	Exhibits	Influence of DNA structure or flanking base on:			Reference
				Formation	Repair	Replication	
*i*pr^{6}G	G	PNNG	G → A > G → T		± UvrB		(van der Vliet et al., 1990)
m^{6}G:C, m^{6}G:T		SSM		TG_1G_2A stability $G_2 > G_1$ & m^6G_2:C > m^6G_2:T, m^6G_1:C = m^6G_1:T			(Voigt and Topal, 1990)
m^{6}G		SSM		$CGCTG_1G_2AGGCG$ hairpin and duplex more stable at G_2			(Bishop and Moschel, 1991)
m^{6}G		SSM		Cm^{6}GC & Tm^{6}GT, same frequency of dNTP incorporation		Klenow extension past m^{6}G favored in Cm^{6}GC	(Dosanjh et al., 1991)
m^{6}G		SSM			G_1G_2A $G_1 > G_2$ by Ada, antibody affinity correlates positively with repair		(Georgiadis et al., 1991)
m^{6}G	G	MNU	G → A, some G → T	PuGG	± methyltransferase		(Moriwaki et al., 1991)
m^{6}G		SSM	G → A	at 5′ end of G run			(Rossi and Topal, 1991)
m^{6}G; e^{6}G	G	MNU; ENU		N_1GN_2 at N_1, G > A > C > m^{5}C ~ T at N_2, C + T			(Sendowski and Rajewsky, 1991)
m^{6}G; m^{4}T	G; T	MNU	G → A T → C transversions	dose dependent, all G → A on non-transcribed strand; transitions at G<u>G</u>; T<u>T</u>			(Yang et al., 1991)
		MNU		influenced by 5′ (not 3′) base		strand fidelity, lagging = leading	(Basic-Zaninovic et al., 1992)
m^{6}G	G	MMS, EMS, MNU, ENU	G → A	G<u>G</u>			(Lee et al., 1992)

Lesion	Formed from	Treatment *via*	Exhibits	Influence of DNA structure or flanking base on:			Reference
				Formation	Repair	Replication	
m^6G		SSM	G → A		m^6G_1GC & Gm^6G_2C, $m^6G_1 > m^6G_2$		(Pletsa et al., 1992)
m^6G	G	MNU	G → A	PuGN in non-transcribed strand			(Akagi et al., 1993)
m^6G	G	MNU	G → A	GG			(Ariza et al., 1993)
m^6G		SSM	G → A		methyltransferase m^6G_1GC & Gm^6G_2C, $m^6G_1 > m^6G_2$		(Bishop et al., 1993)
	T	ENU	$UmuC^+$ transitions; transversions	PuT; CT		$UmuC^-$, no T → A or T → C	(Fix, 1993)
m^6G	G	MNU	G → A	GGT & GGA (not in *ras* codon 12), NPuG, (GPuPu inhibitory)			(Mironov et al., 1993)
m^6G	G	EMS	G → A		G → A in ± UvrB, NER flanking A:T		(Pienkowska et al., 1993)
e^6G	G	ENU	G → A	GGC (*ras* codon 12)			(Pourzand and Cerutti, 1993)
m^5Cm^6G		SSM			human methyltransferase, $Cm^6G > m^5Cm^6G$		(Bentivegna and Bresnick, 1994)
e^6G	G	ENU	G → A	p53 codon 248 CG_1G	G → A mutations in non-transcribed strand, transitions at $G_1 > C$, prob. better NER of $C:e^6G$ in transcribed strand		(Hussain et al., 1994)
m^6G	G	MNNG	G → A	transcribed strand TG, for rpsL target in wt & (Ada^-, Ogt^-)	if lowered transcription of rpsL, GG(A/C) & 66% mutation on non-transcribed strand		(Ito et al., 1994)

Lesion	Formed from	Treatment *via*	Exhibits	Influence of DNA structure or flanking base on:			Reference
				Formation	Repair	Replication	
m^6G	G	SSM	G → A		$G_1G_2CG_3GT$ (*ras* codons 12 & 13), G_2 exhibits reduced methyltransferase repair		(Pletsa et al., 1994)
m^6G; e^6G	G	MNU, MMS; ENU	G → A	GG in non-transcribed strand (± Ada & Ogt)			(Roldán-Arjona et al., 1994)
e^6G	G	ENU		in transcribed gene	in transcribed gene		(Thomale et al., 1994)
		ENU	T → A	at PuT in non-transcribed strand			(Yang et al., 1994a)
m^6G	G	MNNG	G → A	GGPu	human methyltransferase, transcribed strand and at GGG		(Yang et al., 1994b)
ce^6G	G	1-(2-chloroethyl)-3-cyclohexyl-1-nitrosourea (CCNU)	G → A	no bias; flanking (A/T) on both sides > flanking (G/C) on at least one side	Ada^- & Ogt^-; Ada^+ & Ogt^+,	methyltransferase inhibited by (A/T)ce^6G(A/T)	(Jurado et al., 1995)
e^6G	G	ENU			methyltransferase, *Not* I - GCGGCCGC > *Eco*R I - GAATTC, *Bam*H I - GGATCC		(Musarrat et al., 1995)
e^6G	G	EMS	G → A	(A/T)G > (G/C)G; no bias	Ogt^+; Ogt		(Vidal et al., 1995)
m^6G		SSM				bypass depends on position, polymerase, and metal ion	(Voigt and Topal, 1995)
m^6G e^6G; bz^6G		SSM	G → A; (bz^6G also G → T G → C)	G_1G_2A (*ras* codon 12) at G_1, m^6G,e^6G, bz^6G all G → A and semi-targeted at G_2 for bz^6G & e^6G at G_2, e^6G, bz^6G also give semi-targeted at G_1; bz^6G			(Bishop et al., 1996)

Lesion	Formed from	Treatment *via*	Exhibits	Influence of DNA structure or flanking base on:			Reference
				Formation	Repair	Replication	
					human methyl-transferase binds DNA G/C rich somewhat better than A/T, and single-stranded > double-stranded		(Fried et al., 1996)
e^6G; also m^4T	G	ENNG; *N*-nitroso-*N*,*N*-diethylamine	G → A; also A → G, deletions, frameshifts	G → A at PuG; PuC			(Jiao et al., 1996)
ce^6G	G	CCNU	G → A	PuG, GGT & GGG non-transcribed strand wt, GGA transcribed strand NER^-			(Iannone et al., 1997)
ce^6G	G	CCNU	G → A; deletions	GG on non-transcribed strand; (>3 bp deletions at repetitive sequences)			(Luque-Romero et al., 1997)
e^6G	G	ENU	G → A		Ogt prefers 5′ G or C		(Vidal et al., 1997)
m^6G; e^2T	G T	MNU; ENU	G → A; T → A T → G	all formed in non-transcribed strand			(Jansen et al., 1994a)
e^2T	T	ENU	T → A in non-transcribed strand				(Jansen et al., 1994b)
e^2T	T	ENU	T → A		NER on transcribed strand		(Op het Veld et al., 1997)
alkyl^{7}G	G	nitrogen mustards, chloro-ENU		runs of Gs			(Hartley et al., 1988)
alkali-labile adducts		MNU			better when gene is actively transcribed		(LeDoux et al., 1990)

Lesion	Formed from	Treatment *via*	Exhibits	Influence of DNA structure or flanking base on: Formation	Repair	Replication	Reference
m^7G	G	dimethyl sulfate			equal for silent & both strands of active genes		(Scicchitano and Hanawalt, 1990)
N-methylpurines	G & A	MMS		mitochondria > transcriptionally active gene > inactive			(Piršel and Bohr, 1993)
glutathione-S-ethyl^{7}G	G	1,2 dibromoethane	G → A	on non-transcribed strand, PyPuG			(Ballering et al., 1998)
AFB$_1$^{7}G		aflatoxin B_1 (AFB_1)		GG > CG >> AG > TG, GG > GT >> GC > GA			(Benasutti et al., 1988)
AAF-G	G	2-(*N*-acetoxy-*N*-acetylamino)fluorene (AAF)	base substitutions (assay cannot detect frameshifts)			at A:T pairs adjacent to G:C	(Gentil et al., 1986)
AF^{8}G	G	SSM with 2-acetylaminofluorene (AF) adduct	-2 frameshifts only when adduct is at G_3			"Nar I site" $G_1G_2CG_3CC$	(Burnouf et al., 1989)
AF^{8}G		SSM with AF adduct		using oligonucleotides containing Nar I site $G_1G_2CG_3CC$, only G_3 becomes left-handed			(Koehl et al., 1989)
AF^{8}G		SSM with AF adduct			using oligonucleotides containing Nar I site $G_1G_2CG_3CC$, NER excision $G_1 > G_3 > G_2$, but NER recognition all equal		(Seeberg and Fuchs, 1990)
AAF^{8}G; AF^{8}G	G	AAF; AF	-G deletion; +1 addition	from GGGGG		SOS$^+$	(Gupta et al., 1991)
AF^{8}G	G	SSM with AF adduct		in $CCCG_1G_2G_3$ -1 at $G_3 > G_2 > G_1$; in $CACG_1G_2G_3G_4T$ -1 at $G_4 >> G_1$	UvrB$^-$	SOS$^+$, targeted > semi-targeted (in the 5′ repetitive sequences); all targeted deletions	(Lambert et al., 1992b)

Lesion	Formed from	Treatment *via*	Exhibits	Influence of DNA structure or flanking base on:			Reference
				Formation	Repair	Replication	
AF^8G	G	SSM with AF adduct				±SOS, lagging > leading strand (mutations)	(Veaute and Fuchs, 1993)
AF-G	G	SSM with AF adduct			$G_1G_2CG_3CC$ repair ratio, $G_1/G_2/G_3$ *E. coli* NER, 100/18/66 human NER, 38/100/68		(Mu et al., 1994)
AF-G	G	SSM with AF adduct	-1 frameshifts	at contiguous Gs, 3′ > 5′ end		At contiguous Gs, 3′ > 5′ end. In CGGG-(non-transcribed strand), A ~ G > C > T	(Napolitano et al., 1994)
AAF-G; AF-G		SSM with AAF; AF adduct	GC deletions; G → T, G → A		$UvrA^-$	SOS^+	(Tebbs and Romano, 1994)
AF-G; AAF-G		SSM with AF; AAF adduct		$G_1G_2CG_3CC$, deformation at G_1 & G_2 (AAF > AF), G_3 equal.		bypass, AF, (G_1 & G_2 > G_3); AAF, G_1, G_2, G_3 all blocked	(Belguise-Valladier and Fuchs, 1995)
AF-G	G	AF	-GC deletion			N_1GCGCN_2, N_2 > N_1 influence	(Koffel-Schwartz and Fuchs, 1995)
					human NER better for helix-destabilizers AF > UV ~ benzo[*a*]pyrene diol-epoxide (BPDE) > 8-methoxypsoralen (8MOP) > anthramycin > CC-1065		(Gunz et al., 1996)
AAF^8G and AF^8G	G	SSM			NER complex formation and incision GGCGCC > GATGATA (G = AAF or AF)		(Mekhovich et al., 1998)

Lesion	Formed from	Treatment *via*	Exhibits	Influence of DNA structure or flanking base on: Formation	Repair	Replication	Reference
adducts at A; and G	A; G	(−)-benzo[*c*]phenanthrene (4*R*,3*S*)-dihydrodiol (2*S*,1*R*)-epoxide	A → T; ~ G → T			AAC, GAG; AGA	(Bigger et al., 1989)
		± *anti* BPDE	frameshifts; G → T	frameshifts in GC runs; G → T at PyG	Uvr⁻ *E. coli* used	G → T at PyG prob. *via* abasic site (labile BPDE⁷G)	(Bernelot-Moens et al., 1990)
BPDE-G	G	(+)-7,8-dihydroxy-9,10-epoxy-7,8,9,10-tetrahydrobenzo[*a*] pyrene	mostly G → T	target AGG > AG	NER on transcribed strand		(Andersson et al., 1992)
(+)-*anti*-B[*a*]PDE-G	G	(+)-*anti*-benzo[*a*]pyrene-7,8-dihydrodiol-9,10-epoxide	G → T G → A G → C			at TG, only G → T	(Rodriguez and Loechler, 1993a)
(+)-*anti*-B[*a*]PDE-G	G	(+)-*anti*-benzo[*a*]pyrene-7,8-dihydrodiol-9,10-epoxide	G → T; G → A; G → C;			SOS^+, G → T, (A/T)G > (G/C)G; G(C/G) > G(A/T); at TG, only G → T	(Rodriguez and Loechler, 1993b)
Four isomeric benzo[*a*]-pyrene-A adducts		SSM				pol α & HIV-1 RT C10 "S" block 3′ to adduct, C10 "R" incorporates T	(Christner et al., 1994)
(+)-*anti*-BPDE²G; (−)-*anti*-BPDE²G	G	SSM		in $TG_1G_2G_3T$, bends if at G_2 & G_3 (GX preferred); not seen in TG_1G_2C			(Liu et al., 1996)
(+)- and (−)-*trans-anti*-BPDE²G	G	SSM	targeted G → T, but at G_2 also G → A & G → C in COS cells	$CCTG_1G_2CCT$, in E. coli (+)-BPDE²G $G_1 > G_2$ ± SOS, (−) adduct $G_1 = G_2$ in COS (+ & −)-BPDE²G $G_2 > G_1$			(Moriya et al., 1996)

Lesion	Formed from	Treatment *via*	Exhibits	Influence of DNA structure or flanking base on:			Reference
				Formation	Repair	Replication	
B[*a*]P-DE 1-nitro-B[*a*]P-DE 3-nitro-B[*a*]PDE			G → T (esp. in GG, GGA, TGGA, lower for 1-nitro-B[*a*]P-DE)	GPu (no preference for 1-nitro-B[*a*]P-DE)	mutated G on non-transcribed strand		(Zhan et al., 1996)
(+)-*anti*-B[*a*]PDE(+)-*trans-anti*-B[*a*]P^{2}G	G	SSM	in SOS$^+$ mostly G → T at $G_1 > G_2$			$GCG_1G_2CCAAAG$, V_{max}/K_m for dATP $G_1 > G_2$ only BPDE^{2}G:A is extended when terminus is programmed at G_1	(Hanrahan et al., 1997)
B[*a*]PDE-^{2}G (*trans* addition to C10)	G	SSM with (+)-*anti*-B[*a*]PDE	G → A; G → T			CGT (82%); TGC (97%)	(Shukla et al., 1997)
many BP[*a*]DE stereoisomers of G; or A adducts		SSM	-G at 3′ end of G run (not in CGA); MF in TAG > GAT				(Page et al., 1998)
four stereoisomers of BPDE^{6}A		SSM	A → T			MF, TTTAGAG > CAGATTT	(Page et al., 1999)
NONP^{8}G	G	1-nitroso-8-nitropyrene (NONP)	-1 deletions	contiguous GC runs	UvrB$^-$	-1/-2 ratio differs from AAF spectra	(Lambert et al., 1991)
NONP^{8}G	G	NONP	frameshifts; G → T	contiguous GC runs; no flanking base or strand preference	± NER different MF, but similar distribution of mutational classes and strand specificity	Muc(A/B)$^+$, ↑G → T > ↑ -G on transcribed strand	(Lambert et al., 1992a)
AQO^{2}G AQO^{8}G	G	4-acetoxy-aminoquinoline-1-oxide (AQO)				N2 (not C8) adduct influenced by 5′ base	(Daubersies et al., 1992)
4-aminobiphenyl-G	G	4-aminobiphenyl	deletions; additional complex frameshifts	CCGCGCGCGG, CG & GC deletions; CC & GG deletions		± uvrB, SOS$^-$; -uvrB, SOS$^+$	(Levine et al., 1994)

Lesion	Formed from	Treatment *via*	Exhibits	Influence of DNA structure or flanking base on:			Reference
				Formation	Repair	Replication	
1,N^2-propanodeoxy-guanosine (PdG)		SSM	−1, −2 deletions G → T G → C	TPdGT CPdGC		Klenow, PdG:A, V_{max}/K_m TPdGT > CPdGC (PdG:G same), G → T in TPdGT context only	(Hashim and Marnett, 1996)
PhIP[8]G	G	2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhIP)	G → T; −1 del	GGA; GGGA	guanine mutations on non-transcribed strand		(Yadollahi-Farsani et al., 1996)
m^5C	C	SAM methylase	gene regulation	does not form if adjacent m^6G present			(Hepburn et al., 1991)
cis-DDP adducts	GG AG	*cis*-diamminedichloro-platinum(II) (*cis*-DDP)			best in active gene		(Bohr, 1991)
cis-DDP $^7G^{\wedge}{}^7G$	GG	SSM *cis*-DDP			TCN_1GGN_2TC, HMG1domA recognizes A > T > C at N_2		(Dunham and Lippard, 1997)
1,N^6-ethenodeoxy-adenosine (ϵA)		SSM	all four bases can pair opposite lesion, polymerase dependent, ϵA:A & ϵA:C > ϵA:G (ϵA:T not seen) pol I			TTϵATT, TTϵAGG, TTϵATA extended by pol I while AAϵAAA & AAϵATT blocked	(Litinski et al., 1997)
Py^Py	TC CC	UV	C → T	hairpin loops			(Todd and Glickman, 1982)
Py^Py	pyrimidines	UV		3′ ends of pyrimidine tracts			(Sauerbier, 1986)
Py^Py	pyrimidines	UV			transcribed strand		(Mellon et al., 1987)
Py^Py	pyrimidines	UV			photolyase repairs TT & TC, but not CC		(Myles et al., 1987)

Lesion	Formed from	Treatment *via*	Exhibits	Influence of DNA structure or flanking base on:			Reference
				Formation	Repair	Replication	
8MOP^T; T^(8MOP)^T (interstrand cross-links)	T; TA	8MOP + UV	mono-adducts and crosslinks form at TA	TTT >> TT > T in GC environment; ATATAA > ATAT > ATA >> TAT ~ TA >> AT			(Sage and Moustacchi, 1987)
cis-syn T^T cyclobutane (intrastrand)		SSM	T → A T → C; T → C	T^T; T^T (if T^T...hairpin, then all T → A)		blocked in SOS^-, transversions and transitions in SOS^+	(Banerjee et al., 1988)
trans-syn T^T cyclobutane (intrastrand)		SSM	T deletions; T deletions, substitutions, additions			partial block SOS^-; lesser block SOS^+	(Banerjee et al., 1990)
psoralen interstrand cross-link	TA	SSM psoralen + UV			NER acts on furan-T strand if GC rich 6-12 bases on 5′ side, if GC rich on 3′ or not GC rich then no preference		(Jones and Yeung, 1990)
pyrimidine (6-4) pyrimidones (assay cannot detect cyclobutane dimers)	PyPy	UV	T^C forms 3′ to Py runs TT^C strong hotspot > C^C (T^T not seen)	forms at TC & CC inhibited at (T/C)m^5CG			(Pfeifer et al., 1991)
photoproducts	adjacent pyrimidines	UV			NER repairs transcribed strand		(Sockett et al., 1991)
cis-syn T^T cyclobutane (intrastrand)		SSM			NER shows no context influence		(Svoboda et al., 1993)
photoproducts	PyPy	UV			transcribed > non-transcribed strand		(Chandrasekhar and Van Houten, 1994)
oh^{8}G	G	SSM	targeted G → T			± SOS	(Moriya et al., 1991)

Lesion	Formed from	Treatment *via*	Exhibits	Influence of DNA structure or flanking base on:			Reference
				Formation	Repair	Replication	
oh^8G ?	G ?	$Fe^{3+}/EDTA/H_2O_2$	G → T G → C	PuGA > PyGN		SOS^+	(Akasaka and Yamamoto, 1994)
oh^8G:C,T,G,A; abasic:C,T,G,A		SSM			hOGH1, oh^8G:C best and β-elimination occurs, oh^8G:T > G > A but no strand cleavage; only abasic:C cleaved		(Bjørås et al., 1997)
oh^8G:C,T,G,A		SSM			OGG1 oh^8G excision and strand cleavage, oh^8G:C > T >> G & A		(Girard et al., 1997)
oh^8G:C (Fpg repair study), oh^8G (polymerase replication study)		SSM	G → T		poor removal of oh^8G by Fpg in oh^8GA ($oh^8GAPuAPu$ < $PuPuPuPuoh^8GA$ < $PuPuPuPuoh^8G$), well-cleaved in oh^8G(C/T)C ($PyPyoh^8G$ ~ $Pyoh^8G$(G/T) > oh^8GPyC	Goh^8GA good sequence for G → T	(Hatahet et al., 1998)
2-hydroxy-adenine (oh^2A)		SSM	−1 in (+) strand, A → G, A → T in (−) strand			Goh^2AC > Aoh^2AG	(Kamiya and Kasai, 1997)
2′-deoxy-5-hydroxy-cytidine (oh^5C), 2′-deoxy-5-hydroxy-uridine(oh^5U)		SSM	C → T; C → G	context 1 - CXG; context 2 - GXT (surrounding sequence different)		by Klenow (exo^-) oh^5C:G > oh^5C:A, oh^5U:A; oh^5C:C & oh^5U:C	(Purmal et al., 1994)
Thymine glycol	T	OsO_4				Pol I bypass only in CTG or CTA	(Hayes and LeClerc, 1986)
α-2′-deoxy-adenosine (α-2′-dA)	A	SSM	-1 deletion at α-2′-dA			in (α-2′-dA)T, 26%, in (α-2′-dA)G, 1%	(Shimizu et al., 1997)

Lesion	Formed from	Treatment *via*	Exhibits	Influence of DNA structure or flanking base on:			Reference
				Formation	Repair	Replication	
strand brakes		Cu^{2+}/H_2O_2		positively > negatively supercoiled DNA			(LaMarr et al., 1997)
strand brakes		calicheamicin γ_1^I & glutathione		negatively > positively supercoiled DNA (better access of glutathione to DNA-bound drug)			(LaMarr et al., 1998)
		hydroxylamine, nitrous acid		GC			(Brown et al., 1993)
abasic site	U & uracil DNA glycosylase	SSM with abasic site (O)	O → T > A & C	GOTG (T > A ~ C) GTOG (T >> C > A)		poor but better bypass in SOS^+	(Lawrence et al., 1990)
abasic site		SSM abasic site "O"	O → G; O → T	in yeast; in *E. coli*		O:C; O:A	(Gibbs and Lawrence, 1995)
Cr^{3+}-S-glutathione-phosphate-DNA	DNA phosphate backbone	Cr^{3+}-S-glutathione	G → T	GA or GG (spontaneous at GA)			(Voitkun et al., 1998)
G:T, A:C					mismatch repair (MMR), helical (not looped) structures		(Fazakerley et al., 1986)
T; U	m^5C; C	spontaneous	C → T	at C, flanking A:T pairs			(Fix and Glickman, 1986)
U	C	spontaneous	C → T	non-transcribed strand, 3′ to three A:T pairs			(Fix and Glickman, 1987)
mismatches	bases	deamination, recombination, etc.			mismatch repair, G:G > C:C > A:A > T:T, G:T > A:C > C:T > A:G, favored G & C retention		(Brown and Jiricny, 1989)

Lesion	Formed from	Treatment *via*	Exhibits	Influence of DNA structure or flanking base on: Formation	Repair	Replication	Reference
U:G; U:T		SSM, mimics cytosine deamination; or 2′-deoxyuridine 5′-triphosphate incorporation			human, viral, and bacterial uracil-DNA glycosylase, U:G > U:A		(Verri et al., 1992)
G:T		SSM			G:T mismatch binding proteins, equal binding, T incision only if CpG context CG & m^5CG		(Griffin and Karran, 1993)
U		SSM			*E. coli* uracil-DNA glycosylase repairs U in hairpin loops poorly		(Kumar and Varshney, 1994)
mismatches; N:inosine (I)		SSM			deoxyinosine 3′-endonuclease, strand-specific (terminus dependent?), mismatch cleavage inhibited by flanking (G/C); specific for I strand (N:I)		(Yao and Kow, 1994)
G:T		SSM			Vsr, CTAGG best, but also repair if 5′C or 3′G changed		(Lieb and Rehmat, 1995)
U:A, U:G					uracil-DNA glycosylase, (A/T)UA(A/T) >> (G/C)U(T/G/C), neighbor influence > mispair type		(Nilsen et al., 1995)
hetero-mismatches		SSM			bias toward retention of G or C bases		(Bill et al., 1998)

CHAPTER 2:

Description and Validation of a Method for Determining Context Dependent Mutagenesis of DNA Lesions

2A. Introduction

As a prelude to addressing the contribution of all possible nearest-neighbor contexts on the replication past and repair of O^6-methylguanine (m^6G) *in vivo*, we have devised a mutation frequency (MF) detection strategy based on the properties of type IIs restriction enzymes. The method quickly and accurately reported the mutation frequency of a DNA lesion, which could have been placed in any varied local sequence context.

The analysis of the mutational properties of specific types of DNA damage has advanced through the past decade to the point where it is possible to determine the mutational specificity of nearly all types of DNA base damage (Basu and Essigmann, 1988; Loechler, 1996). By chemically synthesizing and placing a modified base at a specific site in a genome, and propagating that genome within cells, one can obtain feedback on how the lesion is processed during its replication and repair. The degree to which a DNA lesion is repaired by a particular system prior to replication can be determined by comparing the mutant to total progeny ratio (the MF) between cells that possess and lack that repair system. Many methods have been developed for determining the MF of a DNA lesion. If the local sequence context is not an issue, the lesion can be flanked by base combinations whereby the mutation dependent alteration of a start (Pauly et al., 1994) or stop (Basu et al., 1993) codon is easily visualized from plated progeny. To analyze the genetic effects of a DNA adduct in any sequence context, one could use differential DNA hybridization (Chary et al., 1995a) or simply sequence individual progeny one at a time

(Lawrence et al., 1990). These systems work well but, because they rely on the screening of a subset from the progeny, large numbers of individual clones must be counted or manipulated per transformation to determine the MF for a single sequence context scenario. Thus it would be an arduous task to study all nearest-neighbor contributions to DNA lesion processing within multiple cell lines. Here, we describe and validate a method whereby the facile detection of mutagenesis in all sequence contexts is accomplished.

2B. Results

Determination of mutation frequency of DNA lesions in any sequence context

The newly developed assay (Figure 3) was named REAP (Restriction Endonuclease And Postlabeling determination of mutation frequency). The method is based upon the properties of type IIs restriction endonucleases, which cleave a specified phosphodiester bond a defined number of bases away from the sequence specific enzyme recognition sequence. As shown in Figure 3, an M13 derivative bacteriophage genome was constructed in which a unique site for a type IIs enzyme, *Bbs* I, was positioned such that the phosphodiester bond to be cleaved was immediately 5′ to the interrogation site (the site that contained the DNA lesion, m^6G, before it was biologically processed). Following replication *in vivo*, double-stranded DNA progeny was cleaved with *Bbs* I, dephosphorylated with shrimp alkaline phosphatase, and treated with RNase A. The linearized, dephosphorylated vector was purified by size exclusion chromatography and 5′ ^{32}P-labeled with T4 polynucleotide kinase, resulting in the incorporation of radiolabel into both DNA strands. Complete digestion of this substrate with *Hae* III liberated a radiolabeled 18-mer from the plus strand and an undesired 207-mer from the minus strand, which were resolved by 20% denaturing polyacrylamide gel electrophoresis (Figure 4). The base composition at the 5′ end of the 18-mer provided the MF. After excision, elution, and desalting, the 5′ ^{32}P-labeled 18-mer was digested to 2′-deoxyribonucleotide 5′-monophosphates (5′ dNMPs) with snake venom phosphodiesterase. At this stage, the sample was applied to a thin layer chromatography (TLC) plate and the 5′ dNMPs were resolved (Figure 5). The relative amounts of radioactivity in the

partitioned spots gave the fractional base composition at the lesion site after mutation creation, which ultimately provided the MF.

The mutation analysis system was validated by performing a reconstruction experiment in which various ratios of a model wild-type and mutant were passaged through cells, and the output "MFs" were determined by two independent methods. The mutational specificity (G → A mutations) (Loechler et al., 1984) and various MFs for m^6G in the sequence 5′-Tm^6GG-3′ were simulated by mixing varying amounts of phage containing genomes of sequence 5′-TGG-3′ and 5′-TAG-3′ (lesion site underlined) from two pure M13 pools. For this particular nearest-neighbor sequence context, the simulated MFs could be analyzed using both the novel assay described and an α-complementation assay in which plaques of different shades of blue were counted. In this assay, lawns of amber suppressing *E. coli* which contain an F′ episome for M13 infection and express the β-galactosidase omega peptide were infected with the mixed populations of M13 phage containing 5′-TGG-3′ and 5′-TAG-3′ sequences within the M13 *lacZ* coding region. Phage with the 5′-TGG-3′ sequence formed dark blue plaques on IPTG/X-gal indicator plates, since β-galactosidase gene activity was restored through α-complementation (Sambrook et al., 1989), thus allowing hydrolysis of the chromogenic substrate X-gal; however, phage that harbored the 5′-TAG-3′ sequence inside the suppressing host formed light blue plaques, since a substantial amount of *lacZ* mRNA was not fully translated. The phage with the 5′-TAG-3′ sequence modeled m^6G induced mutants.

Success of the REAP assay relied on the mutant to wild-type ratio remaining the same for

both phage (generated from an electroporation of genome construct or a contrived mixture) and the replicative form DNA generated from said phage (Figure 3). The contrived phage mixtures, designated as input (Table 2), were used to infect, in triplicate, cultures of *dam*, *dcm E. coli* containing an F′ episome so that double-stranded DNA could be harvested for MF analysis (Figure 3). The composition of the phage and replicative form DNA populations after infection and growth are designated as output (Table 2). The input and output (for one set) mutation frequencies were obtained by plating dilutions of the appropriate phage on indicator plates and counting over 1500 total dark and light blue plaques per mixture (Table 2). The difference between each output vs. input mutation frequency revealed that no bias occurred during growth, with 95% confidence using the difference of proportions statistical test. Furthermore, each output mutation frequency determined by the REAP assay fell within the respective 95% confidence interval obtained from plaque counting (as analyzed using the double-stranded DNA and phage progeny from the same culture tube), suggesting that the REAP and plaque counting assays provided the same results. As shown in Tables 2 and 3, the mutation frequencies analyzed in triplicate using the REAP assay were reproducible. Figure 6 further illustrates that (A) the composition of phage population remains the same and (B) the REAP assay gives MF values identical to the method of counting plaques.

A TLC plate used to generate the data in Table 2 (Set 1) by the REAP assay is shown in Figure 5. For all samples, the amount of radioactivity in the C and T region constituted less than one percent of the total radiation from all 5′ dNMP spots. For the simulated 0% and 100% mutation frequency determinations, the amount of radioactivity in the expected spot represented

greater than 99% of the total radiation emanating from all four 5′ dNMPs. The low background and small standard deviation in the mutation frequencies obtained by the REAP assay instils confidence that a subtle difference (~2%) in the mutation frequency of a biologically processed DNA lesion in different sequence contexts can be deemed statistically significant.

2C. Discussion

DNA damaging agents form a plethora of DNA adducts upon interaction with the genome. A key challenge has been to identify the relative involvement of each of these lesions to the spectrum of mutations by a given mutagen. The ability to place specific adducts at pre-selected sites in genomes has allowed clear statements to be made on the qualitative features of mutagenesis by a host of DNA damaging agents. While the types of genetic changes induced by specific agents are rarely the cause of debate, the frequencies of these specific genetic events vary considerably from laboratory to laboratory. The notion that sequence context effects on repair or replication could be the explanation for these differences has been advanced but, to date, few data are available that either support or refute that view. There has been no generally applicable system by which the genetic effects of lesions in all sequence contexts could be conveniently determined. We developed the system described here in order to address that need.

A key component of the methodology for mutation detection is the use of a type IIs restriction endonuclease. Many parameters were taken into account in choosing the *Bbs* I recognition sequence to provide scission at the interrogation site. The site was introduced as a unique site within the oligonucleotide used to construct the genome, so only genomes that had contained the DNA lesion incorporated radioactivity during postlabeling. Cleavage by *Bbs* I generated a 5′ overhang of 4 bases, which gave unhindered access to alkaline phosphatase and polynucleotide kinase. We assumed that there was low exonuclease contamination since, after

10-fold overdigestion of lambda DNA, greater than 95% of the fragments generated from a *Bbs* I digest could be ligated and recut (New England Biolabs, 1998). The recognition sequence was not too far away from the cleavage site, which may be important if a greater distance between the two had promoted DNA slippage with respect to the type IIs catalytic site, resulting in some cleavage that deviated from the canonical distance. Furthermore, since the genomes were constructed so that two intervening bases lay between the type IIs recognition sequence and the lesion, next-to-nearest neighbor base influences on lesion processing could have also been analyzed using the same assay (i.e., NNm^{6}GNN contexts). An advantage of the experimental design was that double-stranded DNA could be generated on demand by simply mixing the progeny phage, which we found to be stable for several years at 4 °C, with an *E. coli* strain that has an F′ episome allowing for bacteriophage infection. The SCS110 *E. coli* strain was chosen for this purpose because it lacked the site-specific methylases Dam and Dcm, thus eliminating potential cutting bias by *Bbs* I due to host methylation of certain vector sequence contexts.

The REAP assay would have provided inaccurate MFs if the rates of enzymatic activity for *Bbs* I, alkaline phosphatase, and polynucleotide kinase varied with respect to the base at the interrogation site, and the enzymatic reactions at the lesion site had not gone to completion. To address this possibility, separate experiments were performed, under the conditions described in the Methods section of this chapter (2D), revealing that the *Bbs* I and alkaline phosphatase reactions went to completion (data not shown). The 4 hr *Bbs* I and alkaline phosphatase reactions were performed simultaneously, because the *Bbs* I digestion was complete within 2 hr, and the alkaline phosphatase reaction was complete within 1 hr. It is also noteworthy that excess

RNase A was used during the double-stranded vector isolation and cleavage steps prior to size exclusion chromatography, which ensured efficient degradation of large molecular weight RNA to small fragments that were later trapped by the resin as shown in Figure 4. Since small phosphorylatable oligonucleotide contaminants were removed by size exclusion chromatography prior to labeling, a greater than 10-fold molar excess of ATP to 5′ phosphorylatable vector ends should have been maintained throughout the entire reaction. This was necessary to assure that different substrates (e.g., 5′ GG... 3′ and 5′ AG... 3′) would be labeled equally, regardless of differences in the rate constants for substrate phosphorylation. Separate experiments were also performed revealing that the shrimp alkaline phosphatase and polynucleotide kinase were completely inactivated before the next step in the protocol (data not shown). This result assured that the ATP was not depleted by futile cycling during labeling (*vide supra*) and also eliminated the possibility of ^{32}P incorporation into the 5′ ends of the newly exposed *Hae* III fragments *via* the minor 5′ phosphate exchange reaction of polynucleotide kinase. It is reassuring that the intensities of the 18-mer and 207-mer bands were equal to each other, and that these bands dominated the labeled species (Figure 4, lane 5), again indicating that excess ATP was maintained.

Size exclusion chromatography effectively trapped all small phosphorylatable oligonucleotide contaminants generated during double-stranded M13 DNA isolation (< 3% of a radiolabeled double-stranded 24-mer co-eluted with the M13 genome (data not shown)); however, there appeared to be a specific degradation band generated after size exclusion chromatography that ran slightly above the desired 18-mer (Figure 4, visible in lanes 1-3). Due

to the close proximity, this contaminant was most likely excised with the 18-mer band and carried through the assay; however, the intensity of this band was always < 3% of the desired 18-mer. It has yet to be determined where partitioning of the radiation from the contaminant would occur on the TLC plate. We also noticed a well-resolved contaminant that migrated slightly above the 5′ dGMP spot on the TLC plate (Figure 5) that was absent when a "marker" oligonucleotide with a degenerate 5′ terminus was treated exactly as described for the MF samples, omitting the size exclusion chromatography step. Therefore, under the conditions of the REAP assay, the material within the spot was not created by an enzymatic impurity (e.g., adenosine deaminase) or thermal degradation and was tentatively a 5′ ribonucleotide monophosphate, which would not impact MF assessment.

Another method does exist for locating lesion-induced hotspots in a site-specific fashion (Hatahet et al., 1998). While this method may work well for obtaining consensus sequences flanking a lesion that are simultaneously most favorable for miscoding and least favorable for base excision repair, the *in vitro* system may not accurately reflect the realities of lesion processing *in vivo*. Furthermore, many sequencing reactions would have to be performed to see a subtle difference in the contribution of flanking bases to mutations, as would any *in vivo* system whereby a lesion-containing oligonucleotide with flanking degenerate nearest-neighbors is placed in a vector and replicated in cells.

In conclusion, a system has been developed by which the mutagenic potential of any DNA lesion in any and all sequence contexts can be rapidly determined. The low background,

high throughput, and lack of requirement of phenotypic selection will make it possible to discern sequence context effects on the processing of a DNA lesion, m^6G, that exhibits targeted point mutations. The REAP assay can be used to study lesions other than m^6G that exhibit point mutations, in single- or double-stranded vectors; and, with the use of the appropriate shuttle-vector, the assay should be able to determine the effect of sequence context on DNA lesion processing in mammalian cells. In addition to being able to detect the mutation frequency of a DNA lesion placed in any sequence context in single- or double-stranded vectors, the obvious utility of the REAP assay is that it is fast and accurate with a low background. We note that the mutation frequency is generated from the entire population of isolated progeny, which may account for the small standard deviations shown in Tables 2 & 3. The high throughput afforded by one-dimensional TLC made possible the processing of 100 samples by one person in about a week, starting with the progeny phage populations (Figure 3).

Furthermore, since there are no restrictions on the bases flanking the lesion, the same experimental system can be used for determining in what sequence contexts a lesion is more likely to miscode or be poorly repaired, as reflected by their mutation frequencies when placed in the appropriate repair-deficient or repair-proficient cell lines. The *in vivo* nearest-neighbor rules generated from such an experiment may provide insight for determining the important contacts made between the region surrounding a DNA adduct and either a DNA polymerase for effecting nucleotide triphosphate misincorporation, or a DNA repair protein for providing poor binding and/or catalytic activity.

2D. Methods

Simulation of mutation frequency

Mutation frequencies were simulated by mixing varying amounts of phage from two homogeneous pools, whose DNA was identical except at the interrogation site. The pools were derived from the single-stranded parent vector M13mp7L2 (C. Lawrence, University of Rochester), whose hairpin at the *Eco*R I site was replaced with the sequence 5′ GAAGACCTXG GCGTCC 3′, with X being G or A representing pure wild-type or mutant progeny, respectively. Double-stranded DNA was obtained by infecting an overnight culture of SCS110 (JM110, *end A1*; Stratagene) cells grown in unsupplemented LB media (Ausubel et al., 1993) with the phage mixtures. Approximately 10^{10} phage were mixed with 10^8 cells in 10 mL LB and grew on a culture wheel at 37°C for 12 hr. The entire culture was transfered to a 15 mL polypropylene tube, pelleted at 16000 × g for 10 min and drained. Some supernatant was saved to assess the output mutation frequency by plaque color screening as will later be described.

Analysis of MF using the REAP assay

Double-stranded DNA was prepared as follows, using a modification of the technique described by Lee & Rasheed (Lee and Rasheed, 1990). The cell pellets obtained as described above were resuspended in 500 µL R1: 50 mM glucose, 25 mM Tris-HCl buffer (pH 8.0), 10 mM EDTA, supplemented with 5 mg/mL lysozyme powder (grade I from chicken egg white, Sigma no. L6876) and 400 µg/mL of boiled RNase A (type II-A from bovine pancreas, Sigma no. R5000) solution prepared as described (Sambrook et al., 1989). In sets of ten, to 2 mL polypropylene tubes containing the above suspension was added 250 µL of R2: 2% SDS, 0.4 N NaOH obtained from 10% and 2 N stock solutions, respectively. The tubes were agitated gently at room temperature on an Orbitron (Boekel) for 5 min, after which was added 400 µL ice cold R3: 7.5 M NH_4OAc (pH 7.6). Tubes were inverted gently 15 times and placed on ice for 5 min, after which they were centrifuged at 4°C for 10 min. The supernatant was transfered to 1.5 mL tubes, placed on ice for 5 min, and centrifuged at 4°C for 10 min. The supernatant was then decanted into 2 mL tubes containing 650 µL isopropanol, and the contents were mixed and kept at room temperature for 10 min. Tubes were centrifuged at room temperature for 10 min and the supernatant was discarded. A 750 µL solution of 2 M NH_4OAc (pH 7.4) was added and the tubes were agitated for 5 min at high speed on a Vortex Genie 2™ (VWR) equipped to hold thirty 1.5 mL tubes. The contents were pulsed at 1500 × g for 5 sec, transfered to 1.5 mL tubes and placed at 4°C overnight. Tubes were centrifuged at 4°C for 10 min and the supernatant was decanted into 1.5 mL tubes containing 750 µL isopropanol. All following manipulations were performed at room temperature. The contents were mixed, and after 10 min, centrifuged for 10 min. The supernatant was discarded and 1 mL 70% ethanol was added, after which the tubes

were shaken, centrifuged for 10 min, then drained. After another 70% ethanol rinsing, the samples were dried under vacuum in a Speed Vac® (Savant) for 20 min and stored at −20°C. The yield was ~15 μg (1 A_{260} = 50 μg/mL).

Double-stranded DNA was linearized at the lesion site and radiolabeled as follows. The samples obtained above were resuspended in 50 μL of 1 × buffer 2 (New England Biolabs: 10 mM Tris-HCl, 10 mM $MgCl_2$, 50 mM NaCl, 1 mM dithiothreitol (DTT) (pH 7.9)) containing 10 U *Bbs* I (New England Biolabs), 3 U shrimp alkaline phosphatase (Boehringer-Mannheim), and 30 μg RNase A prepared as previously described. After incubation at 37°C for 4 hr, the samples were heated at 80°C for 5 min to inactivate the alkaline phosphatase and stored at −20°C until further use.

The linearized vector was purified away from small contaminating oligonucleotides by size exclusion chromatography as follows. Disposable home-made stationary columns were made from polyethylene transfer pipets (Corning Samco no. 202) and empty spin columns equipped with a 30 micron frit (Americal Bioanalytical). The tapered and closed ends of the pipet were snipped off and the smaller end was forced into the top of the spin column that was filled with water. After the assembly was primed and the water allowed to drain, approximately 4.6 mL of well suspended Sephracryl S-400 resin (Pharmacia) was added. The open end of the pipet bulb acted as a solvent reservoir as the drained column (91 × 7 mm) was washed with 3 mL of 10 mM Tris-HCl buffer (pH 8.5). For ease of processing multiple samples, the columns were inserted into holes made in a styrofoam rack (Sarstedt no. 95.064.249) designed to hold

microfuge tubes (16 holes were made on one side of each rack), and an identical rack was aligned underneath for fraction collection. The sample volume was adjusted to 100 μL with 10 mM Tris-HCl buffer (pH 8.5) and applied to the drained column. Adsorption was allowed to occur and after 10 min, 0.18 g sand, dried and washed (Mallinckrodt), was sprinkled on top. Test runs in which 100 μL of Tris buffer were added and collected revealed that M13 DNA elution consistently started after fraction 13 and ended before fraction 20. Therefore, 1.3 mL of 10 mM Tris-HCl buffer (pH 8.5) was applied to the sample columns that were allowed to drain, after which 600 μL of Tris buffer was added and collected into tubes containing 150 μL 10 M NH_4OAc (pH 7.6) and 750 μL isopropanol. All following manipulations were also performed at room temperature. Samples were mixed, kept overnight, centrifuged in sets of 10 for 10 min, and the supernatant discarded. Approximately 1 mL 70% ethanol was added, after which the tubes were shaken, centrifuged for 10 min, then drained. After another 70% ethanol wash, the samples were dried in a Speed Vac® for 20 min and stored at −20°C. The yield of M13 was approximately 1 pmol (1 pmol = 4.72 μg).

The vector ends were ^{32}P-labeled as follows. Samples were resuspended in 50 μL 10 mM Tris-HCl buffer (pH 8.0) and 40% was dried down (45 min in a Speed Vac® at room temperature) and used for labeling. The DNA was then resuspended in 7.5 μL 1 × buffer 2 containing the following: supplemental DTT for a final concentration of 10 mM, [γ-^{32}P]ATP (New England Nuclear, 6000 Ci/mmol) diluted with ATP (Pharmacia) to provide 10 pmol ATP at specific activity of 1000 Ci/mmol, and 5 U T4 polynucleotide kinase (New England Biolabs). This provided a greater than 10-fold molar excess of ATP to 5′ phosphorylatable ends. After

incubation at 37°C for 1 hr, the samples were heated at 65°C for 20 min to inactivate the T4 polynucleotide kinase.

The 18-mer with variable 5′ base composition (Figure 3) was created and isolated as follows. To the above solution was added 2.5 μL 1 × buffer 2 containing 30 U *Hae* III (New England Biolabs, 50 U/μL stock). After this solution was incubated at 37°C for 2 hr, 10 μL of denaturing dye (98% formamide, 10 mM EDTA (pH 8.0), 0.025% bromophenol blue, 0.025% xylene cyanol FF) was added and the samples were stored at −20°C until further use. Small, thin 17 cm × 13.5 cm × 0.8 mm 20% denaturing polyacrylamide gels (7 M urea, 19:1 acrylamide:*N*-*N*′-methylene-bis-acrylamide, 1 × TBE (0.09 M Tris-HCl, 0.09 M boric acid, 0.002 M EDTA (pH 8.3))) were prerun at 550 volts for 15 min. Every other lane was loaded with sample, and 10 samples per gel were run at 550 volts for 70 min. After the free ATP was cut out of the gel and discarded, the gel was placed between SaranWrap™ and set under a storage phosphor screen for 30 min; the output was generated on a transparency using a PhosphorImager (Molecular Dynamics). The gel was sliced vertically down the unloaded lanes to prevent cross-contamination and for each sample, the well and 18-mer from the transparency placed underneath the gel strip was aligned with the well and the bromophenol blue and xylene cyanol dye markers flanking the 18-mer from the gel. The 18-mers were excised in 1 cm × 0.5 cm slices that were crushed and had 200 μL water added to them in 0.5 mL flip-top tubes (Sarstedt no. 72.699), whose caps and lips had been removed with a utility knife. After sitting at room temperature overnight, the gel bits were agitated by drawing and expelling 100 μL of solution, and 20 min passed before the liquid was separated from the gel. Alternatively, the crushed gel can be soaked

for 1 hr, agitated, and soaked for an extra hour to allow for complete equilibration between the gel and the liquid. A home-made device was made that quickly forced the liquid out of the gel solution, leaving most of the gel bits behind. Two 1.5 mL flip-top tubes (Sarstedt no. 72.690) were stacked; the top one had the cap pulled off, a small hole at the bottom made by a 26 gauge needle, and contained 0.09 g sand. The 0.5 mL tube containing the gel solution was placed into the assembly inverted and spun at high speed in a microcentrifuge for 1 min (with every other chamber loaded). Greater than 60% of the radioactivity came out in the eluant. Home-made desalting spin columns were made using Sephadex G-25 Fine resin (Pharmacia) that had been swelled and washed with 3 volumes of water. Resin was added to primed empty columns (American Bioanalytical) and the liquid was allowed to drain. Each column was placed into two 1.5 mL Sarstedt flip-top tubes that were stacked, had both lids pulled off, and a hole made in the top supporting tube. The assembly was placed in a swinging bucket centrifuge (RC-3B Sorvall) and spun at 700 × g (1500 rpm) for 2 min. The packed bed height was 42 mm. All of the gel eluant (approximately 160 μL) was layered onto the desalting column, which was placed inside the supporting tube and stacked on a new 1.5 mL Sarstedt tube, which was spun at 700 × g for 4 min. Greater than 75% of the radioactivity loaded on the column came out in the eluant, with a volume change of less than 10%. The sample to sample recovery from both gel elution and Sephadex desalting were consistent. The eluant was dried at room temperature under vacuum in a Speed Vac® overnight.

The 18-mers were digested to monomers and analyzed by TLC as follows. The 5′ ^{32}P labeled 18-mers were digested to 5′ deoxynucleotide monophosphates using snake venom

phosphodiesterase I (ICN Biomedical no. 100978, 40 U/mg solid). A 20% glycerol stock at 50 mU/μL had been made from the lyophilized enzyme and stored at −20°C. A 5 μL solution of 100 mM Tris-HCl buffer (pH 8.8), 15 mM $MgCl_2$, and 50 mU snake venom phosphodiesterase I was used to resuspend the 18-mers and digestion was carried out at 37°C for 1 hr. Approximately 0.5 μL of sample was drawn into a 1 μL glass micro capillary tube (Drummond) attached to an adjustable positive pressure dispenser and spotted on a 20 × 20 cm polyethyleneimine (PEI) cellulose TLC plate (J. T. Baker). The spots were spaced 2.5 cm from the bottom and 1 cm apart. After the last spot was dry (5 min) the plates were placed in TLC tanks (27 L × 7.5 W × 26 H cm) containing 200 mL of saturated $(NH_4)_2SO_4$. The plates were developed until the solvent almost reached the top (~5 hr), dried at room temperature for 1 hr, and placed under storage phosphor screens for 36 hr. Mutation frequencies were calculated using ImageQuant 5.0 software (Molecular Dynamics) by drawing ellipses (local average background correction) around the A (5′ dAMP) and G (5′ dGMP) spots and reporting the percentage value of $100 \times [A/(A + G)]$.

Determination of mutation frequency by counting individual plaques

Mutation frequencies were obtained by counting dark and light blue plaques for the simulated mutation frequency experiment, as well as for the *in vivo* m^6G processing experiment in which a 5′-Tm^{6}GG-3′ sequence context was used (Chapter 3). An overnight culture of NR9050 *E. coli* (*suB*, *F′ prolacIZΔM15* (relevant genotype); R. M. Schaaper, NIEHS) was diluted 1:5 in 2 × YT and grew on a culture wheel at 37°C for 90 min before being used as plating bacteria. In triplicate or greater, 300 μL plating bacteria, 25 μL 1% thiamine, 10 μL isopropyl β-D-thiogalactopyranoside (IPTG, 24 mg/mL stock in water), 40 μL 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside (X-gal, 40 mg/mL stock in *N,N*-dimethylformamide), and phage to yield about 500 plaques per 100 × 15 mm plate were added to 2 mL B-broth soft agar (10 g/L tryptone, 8 g/L NaCl, 10 mg/L thiamine, and 0.6% (w/v) Bacto agar) kept at 47°C and spread immediately on 25 mL B-broth plates (2% (w/v) Bacto agar). After 10 min, the plates were inverted and incubated at 37°C for 12-16 hr. This technique gave excellent discrimination between dark and light blue plaques.

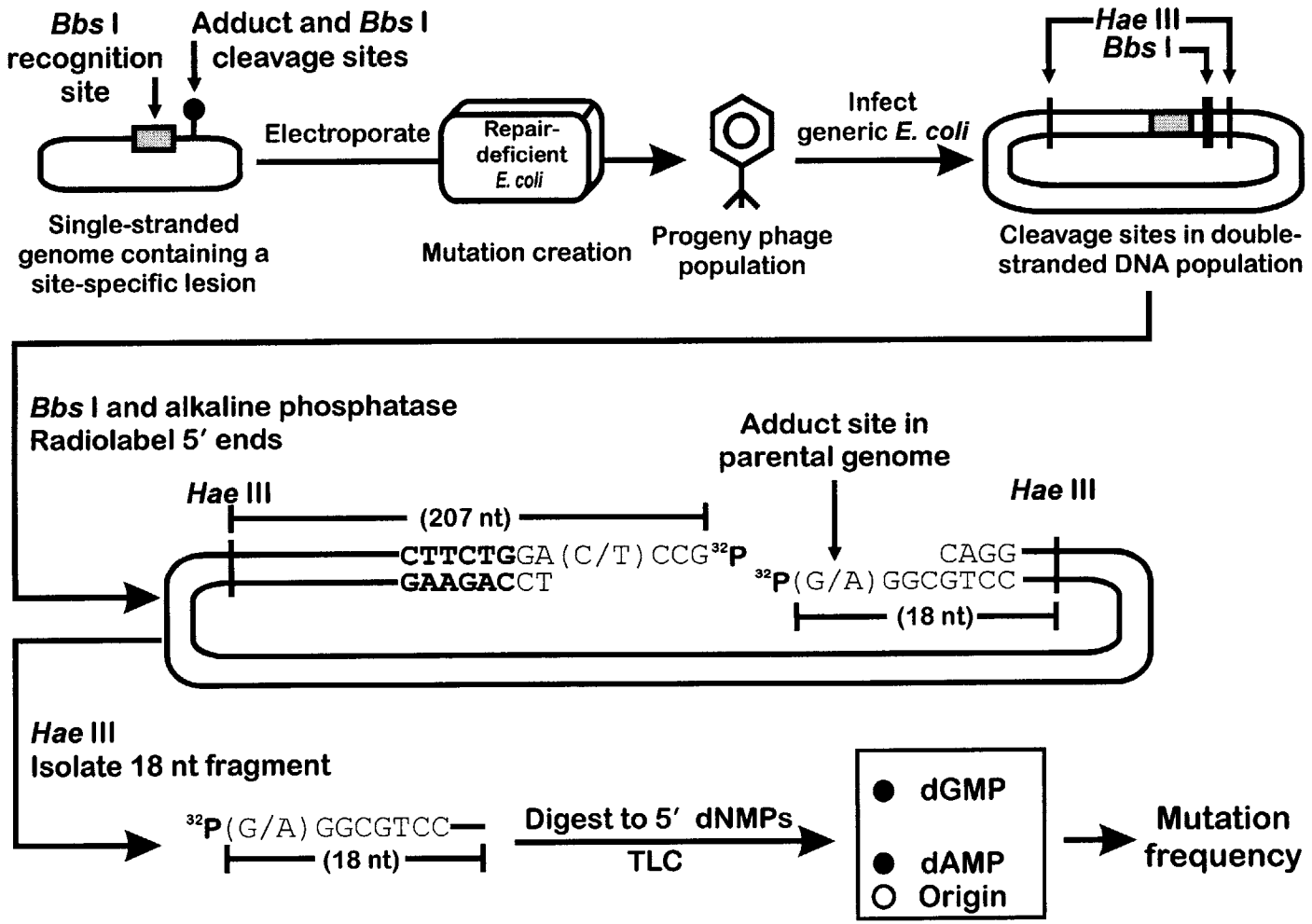


Figure 3. Detection of mutations in any sequence context using the REAP assay. A site-specifically adducted genome was introduced into a repair-deficient (or proficient) cell line, and the progeny phage from the biological processing of the lesion was used to infect a generic strain of *E. coli*, thus creating a double-stranded DNA population. DNA was cleaved at the position that had contained the lesion in the parental vector with the type IIs restriction endonuclease *Bbs* I (recognition sequence in bold). The 5′ ends of the linearized duplex were radiolabeled at the lesion site which, in the case of m^6G, contained a mixture of G and A after passage through the cell. Treatment with *Hae* III liberated an 18-mer, whose base composition at the 5′ end provided the MF. To determine the MF quantitatively, the purified 18-mer was digested to 5′ deoxynucleotide monophosphates (5′ dNMPs) with snake venom phosphodiesterase. Partitioning of the radioactive 5′ dNMPs on a thin-layer chromatography (TLC) plate, followed by PhosphorImager analysis, provided the fractional base composition at the lesion site, from which the mutational specificity and mutation frequency were determined.

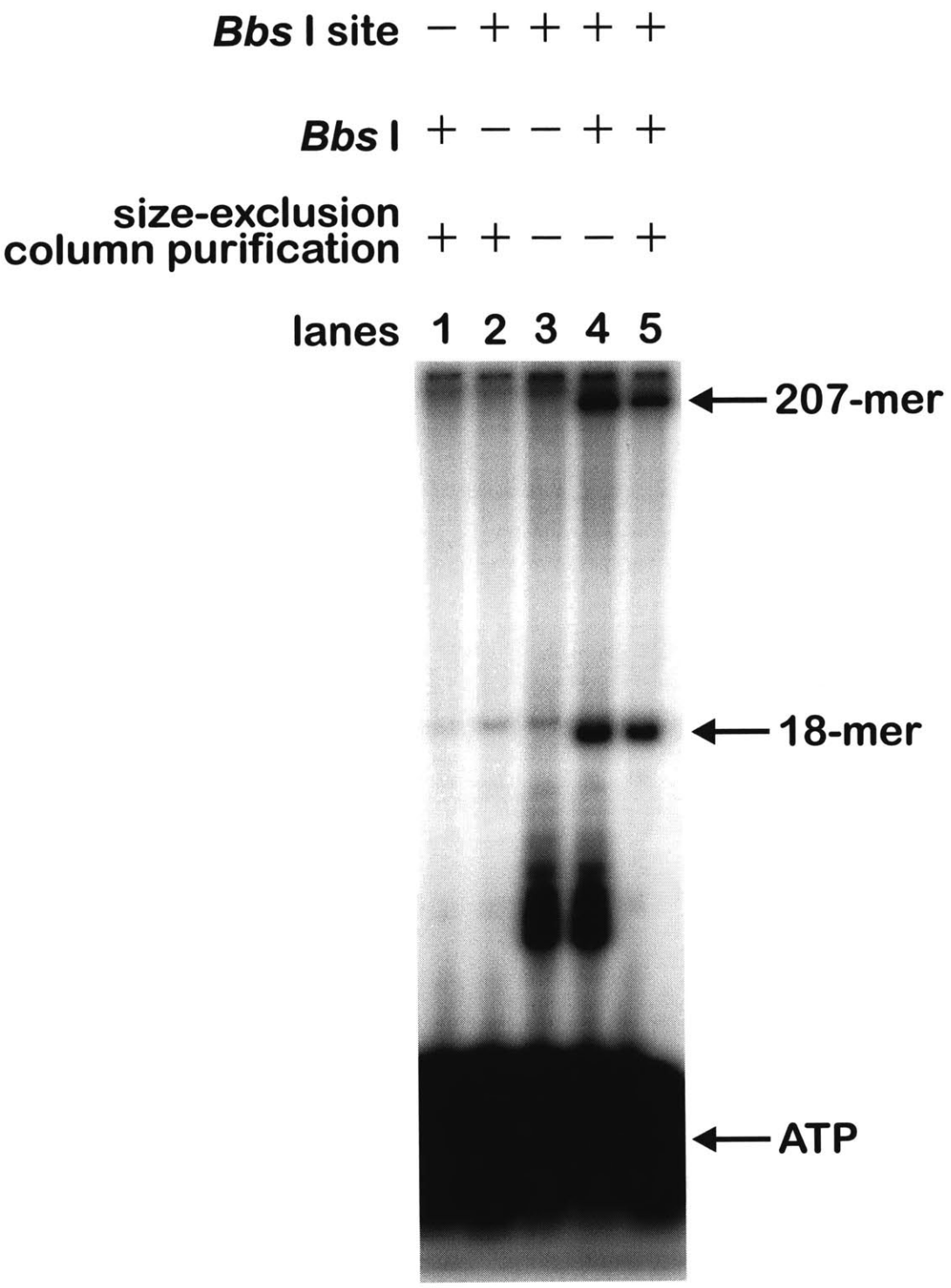


Figure 4. Gel purification of 18-mer containing a 5′ ^{32}P phosphate at the interrogation site. Lane 5 is the result of the REAP assay exactly as described, whereas the omission of certain elements in other lanes emphasizes the importance of size exclusion chromatography prior to radiolabeling.

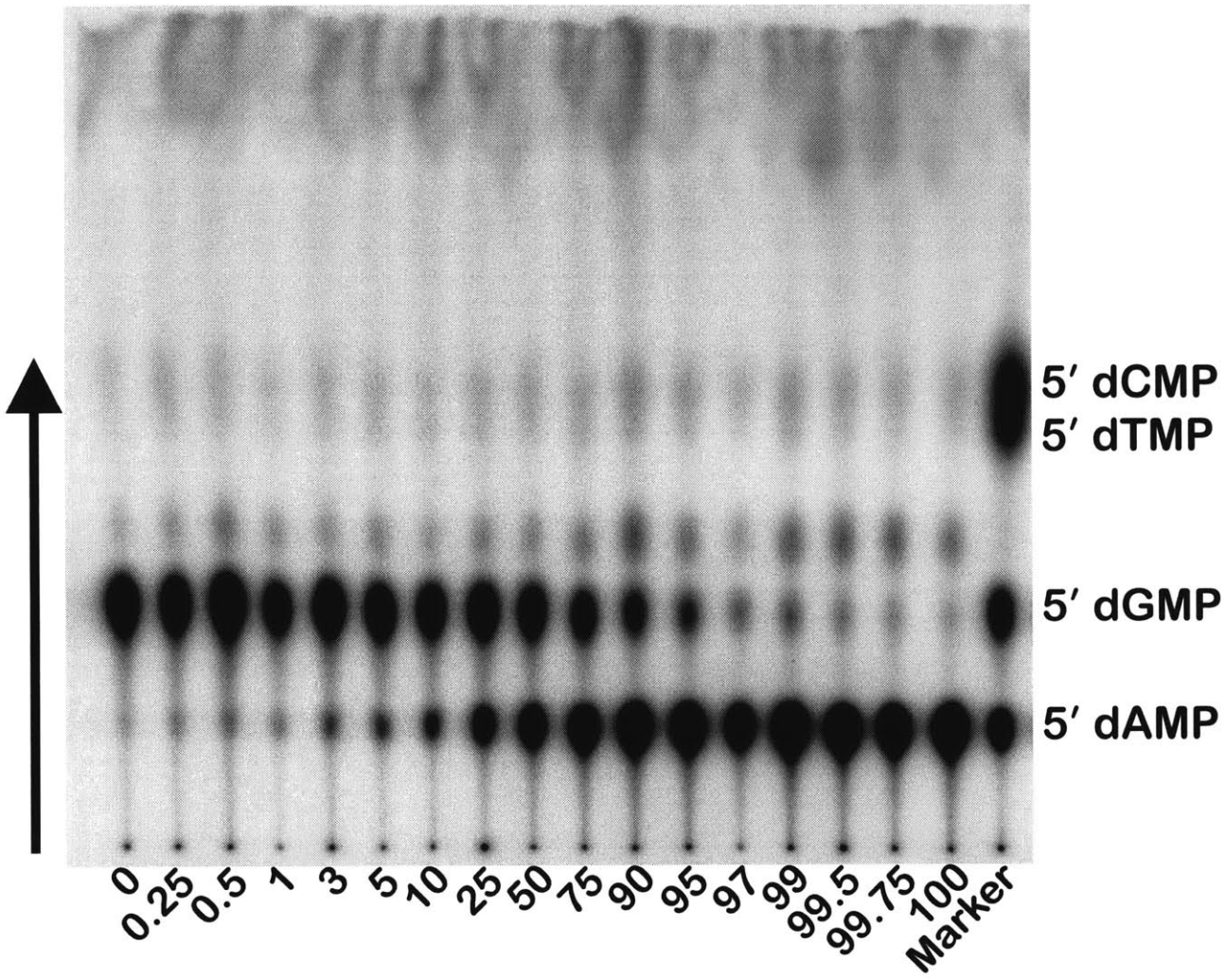


Figure 5. Thin-layer chromatography showing the MF analysis by the REAP assay for a range of simulated MFs at the "lesion" site. Model mutant (A at "lesion" site) and wild-type (G at "lesion" site) genomes were mixed at known input ratios and taken through the REAP assay. Following isolation of ^{32}P-labeled 18-mer (Figure 4), and enzymatic hydrolysis, 5′ nucleotides were resolved by TLC and visualized by PhosphorImager analysis. The numbers from 0 to 100 are simulated target input MFs. The marker consists of a hydrolysate of a 5′ ^{32}P-labeled oligonucleotide whose 5′ end was degenerate.

Table 2: Validation of the REAP assay.

Target Input MF (%)	Input (Plaque Counting)			Output Set 1 (Plaque Counting)			Output MF (REAP)			
	Dark	Light	MF	Dark	Light	MF	Set 1	Set 2	Set 3	Average*
0	1740	2	0.1	5409	9	0.2	0.2	0.2	0.2	0.2 ± 0.0
0.25	ND	ND	ND	ND	ND	ND	0.4	0.5	0.2	0.4 ± 0.1
0.50	ND	ND	ND	ND	ND	ND	0.6	0.5	0.5	0.5 ± 0.1
1	1747	18	1.0	1696	16	0.9	1.0	0.7	0.6	0.8 ± 0.2
3	1564	44	2.7	2162	81	3.6	3.5	2.4	2.6	2.8 ± 0.6
5	1801	86	4.6	1690	87	4.9	5.6	4.5	4.8	5.0 ± 0.6
10	1502	154	9.3	1445	156	9.7	10.7	10.0	9.7	10.1 ± 0.5
25	1303	391	23.1	1353	437	24.4	26.0	25.6	27.3	26.3 ± 0.9
50	850	816	49.0	1389	1425	50.6	52.3	52.8	51.1	52.1 ± 0.9
75	531	1528	74.2	569	1678	74.7	75.7	76.8	76.0	76.1 ± 0.6
90	173	1531	89.8	163	1365	89.3	90.2	90.6	90.9	90.6 ± 0.4
95	120	1984	94.3	100	1668	94.3	95.2	95.3	96.0	95.5 ± 0.4
97	47	1588	97.1	36	1494	97.6	98.4	97.8	98.1	98.1 ± 0.3
99	37	3647	99.0	25	1797	98.6	99.1	99.0	99.3	99.1 ± 0.1
99.5	ND	ND	ND	ND	ND	ND	99.5	99.6	99.7	99.6 ± 0.1
99.75	ND	ND	ND	ND	ND	ND	99.7	99.7	99.8	99.7 ± 0.1
100	0	1613	100	1	1621	99.9	99.9	99.9	99.9	99.9 ± 0.0

Simulated MFs were determined from contrived mixtures of phage from two pure phage pools (5′-TGG-3′ and 5′-TAG-3′) in proportions to yield the target input MFs for m^6G at the underlined site. Dilutions of these mixtures were plated on indicator plates and wild-type (Dark) and mutant (Light) blue plaques were counted to provide the true input MFs. The mixtures were propagated in a bacterial cell line and progeny phage from the supernatant was used to determine the output MFs, again by plaque counting, whereas progeny double-stranded DNA inside the cell pellets was used to determine the output MFs by the REAP assay.

*The average output MF determined by the REAP assay was based on MFs obtained using the double-stranded DNA output from three cultures of *E. coli* that were independently infected with a common phage input mixture, and is reported as ± one standard deviation. ND, not determined.

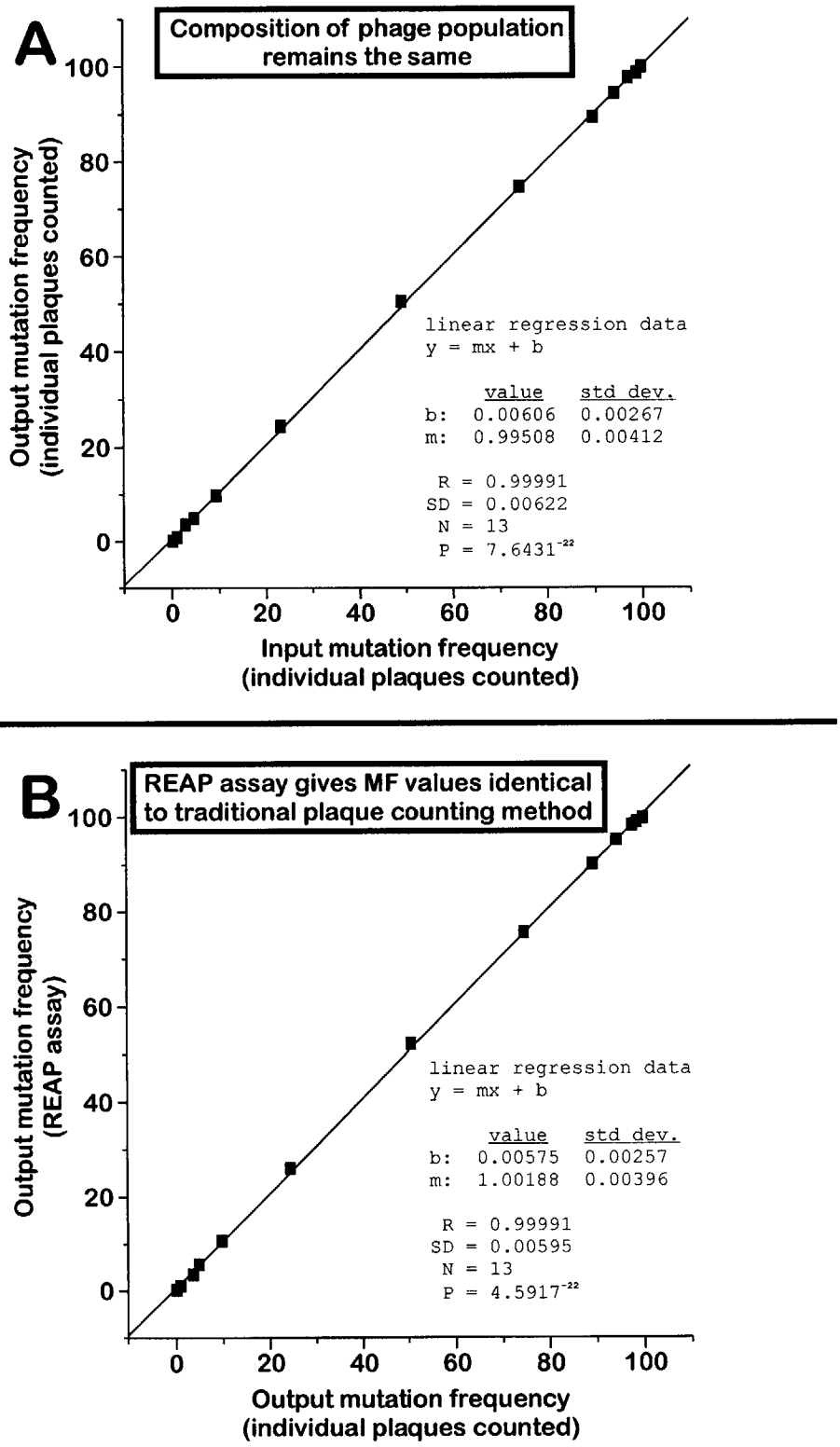


Figure 6. Linear regression analysis of (A) output MF by plaque counting vs. input MF by plaque counting and (B) output MF by the REAP assay vs. output MF by plaque counting.

CHAPTER 3:

Influence of DNA Sequence Context on O^6-Methylguanine Mutagenicity in *Escherichia coli*

3A. Introduction

Detailed analysis of mutational hotspots following DNA damage provides an understanding of oncogene activation and tumor suppressor gene inactivation and, hence, provides insight into the earliest steps in the induction of cancer. A mutational hotspot may be created by preferential lesion formation, by decreased lesion repair, or by increased misinsertion past the lesion during DNA replication. The respective contributions of these factors may be influenced by the DNA sequence context of the hotspot.

We chose m^6G (Figure 7) as the inaugural lesion for the REAP assay described in Chapter 2, since the persistence of O^6-alkylguanine DNA adducts correlates well with the mutagenic and carcinogenic effects of alkylating agents (Goth and Rajewsky, 1974; Kleihues and Margison, 1974), which can be formed endogenously by nitrosation of amides and amines (Taverna and Sedgwick, 1996; Sedgwick, 1997). The mutagenic consequence of unrepaired m^6G is the creation of exclusively G to A transition mutations, as demonstrated in *E. coli* (Loechler et al., 1984). *E. coli* has proteins that remove O^6-methylguanine, thus preventing mutations. The inducible Ada and constitutive Ogt methyltransferases have active site cysteines that receive the alkyl group directly, converting O^6-alkylguanine to guanine, but the active site is not regenerated (Samson, 1992). Both the UvrABC nucleotide excision repair (Samson et al., 1988; Voigt et al., 1989) and mismatch repair (Rasmussen and Samson, 1996) pathways have also been implicated in m^6G repair.

The experimental design used to measure the impact of DNA repair and replication on m^6G mutagenesis with respect to nearest-neighbor sequence context is outlined in Figure 8. Sixteen individual single-stranded viral M13 genomes were made, containing m^6G flanked immediately 5′ and 3′ by each permutation of G, A, T, and C within the same region of the M13 genome. Since site-specific methodology was used, the effect of differential adduct formation with respect to sequence context, while still a formal possibility for generating non-uniform mutational spectra, was removed. Electroporation of each member of the viral genome array into *E. coli* that lacked all known repair elements that could directly remove m^6G assessed the effect of sequence context on dNTP misincorporation by the replicative DNA polymerase holoenzyme (pol). If differences in the MFs among genomes in *ada*, *ogt*, *uvrB E. coli* existed, the effects would most likely be attributed to the DNA polymerase. Likewise, differences in repair-proficient isogenic cell lines would be due to a combination of the DNA polymerase and the respective repair element. If no differences in MFs among genomes were seen in *ada*, *ogt*, *uvrB E. coli*, then differences in MFs among genomes in cells containing one element of repair would be attributed solely to that repair element.

In order to achieve our goal of constructing the viral genome array (Figure 8), we also report a method for constructing site-specific single-stranded viral DNA genomes that should yield identical ligation efficiencies, regardless of the lesion or its surrounding sequence context.

Although the REAP assay had already been validated by using simulated mutation frequencies in which the ratios of pure mutant and wild-type phage were mixed, as explained in

Chapter 2, we wanted to validate the assay using a few genomes containing *bona fide* m^6G lesions prior to execution of the full scheme outlined in Figure 8. As shown in Figure 7, the three sequences chosen were 5′-Gm^{6}GA-3′, 5′-Am^{6}GA-3′, and 5′-Tm^{6}GG-3′. The first two differed from each other only in that they had a different 5′ base. The MF of the 5′-Tm^{6}GG-3′ sequence could additionally be obtained by counting dark and light blue plaques as described in Chapter 2 and thus, was used as an independent method of further validating the REAP assay. The information in the Results (3B) and the Discussion (3C) sections will be structured such that the three sequence sampling (*vide supra*) analyzed in the two *E. coli* strains (*ada*, *ogt*, *uvrB*) and (*ogt*, *uvrB*) will precede the full sixteen nearest-neighbor sampling in (*ada*, *ogt*, *uvrB*), (*ogt*, *uvrB*), (*ada*, *uvrB*), and wild-type *E. coli* as outlined in Figure 8.

3B. Results

Construction of site-specific genomes

After the REAP assay was validated using simulated MFs, three viral single-stranded genomes were constructed that contained m^6G in the contexts 5′-Gm^{6}GA-3′, 5′-Am^{6}GA-3′, and 5′-Tm^{6}GG-3′. A procedure was developed whereby a DNA lesion would easily be placed in all possible contexts (Figure 9). Initially, an M13 circular single-stranded genome was linearized at a unique *Eco*R I site within a hairpin. A previous study (Lawrence et al., 1990) has used one long scaffold to bridge the resulting vector termini, thus creating a gap into which the insert containing the DNA lesion could be placed prior to ligation. Instead of using a different long scaffold to achieve a high ligation efficiency for each sequence to be tested (16 scaffolds would be needed to test all Nm^{6}GN contexts), we used two shorter scaffolds from which the variant genomes were made. It is noteworthy that the same scaffolds and M13 genome preparations are used in all samples; only the adduct-containing 16-mer varies from sample to sample. An equimolar mixture of components was created, with each scaffold annealing to 20 bases of the vector and 6 or 7 bases of the 5′ phosphorylated insert. The lesion-containing oligonucleotide was ligated into the vector and the scaffolds were degraded with exonuclease III until they no longer annealed to the vector (data not shown). Since the technique described uses two scaffolds that do not span the variable positions that flank the lesion, although not essential for this project, identical ligation efficiencies should be achieved for the construction of all nearest-neighbor, lesion-containing genomes, regardless of lesion or bases flanking the lesion. Furthermore, this

method should improve the ligation efficiency of bulky DNA adducts, since there is no destabilizing base placed opposite the lesion. One could adapt this method for constructing genomes containing thermally unstable adducts by pre-annealing the scaffolds to the vector and adding the adduct-containing insert at low temperature. We also found that passing the M13mp7L2 vector source through a hydroxylapatite column made M13 genomes viable after being exposed to high temperature prior to electroporation.

Influence of sequence context on replication past m^6G by DNA polymerase III and repair of m^6G by Ada in vivo using a three nearest-neighbor analysis - Further validation of the REAP assay

Viral genomes containing m^6G in the 5′-Gm^{6}GA-3′, 5′-Am^{6}GA-3′, and 5′-Tm^{6}GG-3′ contexts were made in duplicate and each construct was electroporated into repair-deficient *E. coli* as described in the Methods section of this chapter (3D). The mutation frequency from each transformation was analyzed in triplicate using harvested progeny phage to infect *E. coli* for isolation of double-stranded DNA employed in the REAP assay (Figure 3). As shown in Table 3, the mutation frequency in an *ada*, *ogt*, *uvrB* cell line was nearly 100% for all three sequence contexts. This result demonstrates the nearly complete miscoding property of m^6G *in vivo* in the absence of repair. The isogenic cell line that expressed uninduced Ada demonstrated a sequence context dependence by this protein in the repair of m^6G, as reflected in the difference in the mutation frequencies. For the 5′-Gm^{6}GA-3′, 5′-Am^{6}GA-3′, and 5′-Tm^{6}GG-3′ contexts, the respective mutation frequencies obtained from the REAP assay were 53%, 90%, and 82% for the first independent trial and 45%, 90%, and 81%, for the second. The 5′-Tm^{6}GG-3′ context was also analyzed by counting dark and light blue plaques, providing MFs of 81% and 79% for the first and second trials, respectively (data not shown). The mutation frequencies obtained by the REAP and plaque counting assays were statistically identical. Furthermore, there was an insignificant variation in the MF obtained with the REAP assay from each m^6G biologically processed genome when performed in triplicate, thus further validating the REAP assay.

Influence of sequence context on replication past m^6G by DNA polymerase III and repair of m^6G by Ada or Ogt or (Ada, Ogt, and UvrB simultaneously) in vivo using a sixteen nearest-neighbor analysis

The full sixteen nearest-neighbor sampling was analyzed in *E. coli* containing the genotypes (*ada*, *ogt*, *uvrB* = Pol III influence), (*ogt*, *uvrB* = Ada influence), (*ada*, *uvrB* = Ogt influence), and (wild-type = Pol III, Ada, Ogt, and UvrB all operative) as outlined in Figure 8. Sixteen nearest-neighbor viral genomes containing m^6G in were made and each construct was electroporated into the appropriate repair-deficient *E. coli* as described in the Methods section of this chapter (3D). The results plotted as bar graphs are shown in Figures 10-15. While the data in Figure 10 (Pol III influence) and in Figure 15 (all repair systems operative) were generated once using an older, unoptimized version (not shown) of the REAP assay, the data in Figure 11 and Figure 12 (Ada influence, same data but grouped to show a 5′ or 3′ influence, respectively) and the data in Figure 13 and Figure 14 (Ogt influence, again, same data but grouped to reveal a 5′ or 3′ influence, respectively) were performed in duplicate starting with the genome construction stage and using the optimized version of the REAP assay as described in Chapter 2D of this work. Each independent genome construction and electroporation is designated as A or B in Figures 11-14.

As shown in Figure 10, when an *ada*, *ogt*, *uvrB* cell line was used, m^6G in all sequence contexts was nearly 100% mutagenic, suggesting that the DNA polymerase III holoenzyme always placed a T opposite m^6G; and consequently, that the sequence context surrounding m^6G does not cause the DNA polymerase to create mutational hot- or coldspots. As shown in Figure

11 and Figure 12, in an *ogt*, *uvrB* cell line, the MFs ranged from 35% to 90%. Within this range, attributed to the effect of sequence context on m^6G repair by Ada, a 5′ sequence context influence was evident (Figure 11). The order of increasing MF attributed to Ada was, in general, GX < CX < TX < AX (X = m^6G). Ada also showed a 3′ influence of XPyrimidine < XPurine (Figure 12). By contrast, as shown in Figure 13 and Figure 14, in an *ada*, *uvrB* cell line, the MFs ranged from 10% to 25%. Ogt demonstrated a 5′ influence of GX < AX (Figure 13), and the contexts 5′-AXN-3′ (N = any base) provided the highest MFs (Figure 14). The MFs with respect to sequence context in the wild-type cell line (Ada, Ogt, and UvrB all operative) (Figure 15) revealed a pattern strikingly similar to that obtained using *ada*, *uvrB E. coli* (Figure 13), suggesting that Ogt is the dominant repair element for m^6G in uninduced *E. coli*.

3C. Discussion

Influence of sequence context on replication past m^6G by DNA polymerase III and repair of m^6G by Ada in vivo using a three nearest-neighbor analysis - Further validation of the REAP assay

As can be seen in Table 3, the mutation frequencies for m^6G in the *ada*, *ogt*, *uvrB* cell line were nearly 100% for all three contexts tested. Our values were thus very similar to results of Pauly et al. (Pauly et al., 1998; Pauly et al., 1995) who, by using a single context opposite a gap, show a 94% MF for m^6G in an *ada*, *ogt*, *mutS E. coli* cell line. Our results showed this high MF holds for multiple contexts. An *in vitro* primer extension study using a template containing m^6G has shown that the Klenow fragment of *E. coli* DNA polymerase I will place a T opposite the lesion ~50% or ~85% of the time depending on the flanking base (Singer et al., 1989); our data suggest that the DNA polymerase holoenzyme almost always places a T opposite m^6G during replication *in vivo*, at least for the three contexts tested. Our data further suggest that differential DNA polymerase fidelity with respect to bases flanking m^6G does not play an important role in the creation of mutational hot- or coldspots.

When genomes containing m^6G were passed through the isogenic *ogt*, *uvrB* cell line (i.e., cells with a constitutive level of Ada), the MFs dropped for all three sequence contexts tested (Table 3). Although the mutation frequencies attributed to Ada were similar for identical genomes containing m^6G in the context 5′-Am^6GA-3′ (~90%) or 5′-Tm^6GG-3′ (~82%), they differed significantly for the 5′-Gm^6GA-3′ context (53% vs. 45%). This difference is most likely

attributed to the higher repair efficiency of Ada on the 5′-Gm^{6}GA-3′ context and cannot be explained by experimental error, which was maximally about 2% (with 95% confidence) for a 50% MF (Table 2). It has been shown that the concentration of uninduced Ada in *E. coli* can vary with respect to when the culture is harvested (Taverna and Sedgwick, 1996). Since m^{6}G was relatively well repaired in the 5′-Gm^{6}GA-3′ context, a slight variation in methyltransferase concentration during different batch preparations of electrocompetent cells may have had a greater impact on the absolute mutation frequency; however, within each batch preparation, the sought after rank order of mutation frequency with respect to sequence context (*vide infra*) will remain the same.

The Ada protein is one of the few DNA repair proteins that can work on single-stranded DNA. Although single-stranded viral m^{6}G genomes were introduced into the repair-deficient cell lines, the observed effects may have very well arisen from the reaction of Ada with m^{6}G in duplex form. The reaction with m^{6}G in single-stranded DNA occurs at only 0.1% of the rate for duplex DNA (Lindahl et al., 1982) and it has been estimated that there are only one or two molecules of Ada in uninduced *E. coli* (Rebeck et al., 1989). Indeed, if the protein were acting on m^{6}G in duplex form, the opposing base would be T, since the work presented here demonstrates that the DNA polymerase always placed a T opposite the lesion. The fact that the mutation frequency dropped by half when a single base 5′ to m^{6}G was changed from an A to a G suggested that the activity of the Ada repair protein, either directly or indirectly, is strongly influenced by the 5′ flanking base. One would expect, therefore, that genomes of alkyltransferase-proficient cells treated with alkylating agents would show an excess of mutations

in 5′-AG-3′ sites as compared to 5′-GG-3′ sites. There are indirect data supporting this prediction. Specifically, alkylation treatment of alkyltransferase-proficient *E. coli*, with respect to their isogenic alkyltransferase-deficient counterpart, results in increased guanine mutagenesis in the 5′-AG-3′ sequence (Vidal et al., 1995; Vidal et al., 1997), and decreased mutagenesis in the 5′-GG-3′ sequence (Jurado et al., 1995). One caveat is that in most of these published studies on alkyltransferase-deficient cells, Ogt was the variable modulating the site of mutation, whereas our conclusion was based upon Ada.

We are cautious in attributing the lower MF for the 5′-Gm^{6}GA-3′ context directly to Ada alone, since our cells had mismatch repair capability. Three formal scenarios can explain the differences in the MF in the Ada-proficient cell line without Ada being affected by sequence context. First, sequence context may affect binding of the mismatch repair recognition protein MutS to an m^{6}G:T pair, making some sequences better shielded by MutS from repair by Ada. Second, the m^{6}G:T mispairs may be converted equally to G:T mispairs by Ada, but initiation of the methyl-directed mismatch repair process by MutS, that would ultimately lead to a G:C pair, may be context dependent. Third, m^{6}G may be a stronger block to the DNA polymerase in some sequence contexts, thus giving Ada more time to convert m^{6}G to G in single-stranded DNA before the mutation is created.

To conclude, using the novel methodology of two scaffold single-stranded genome construction and repair-deficient *Escherichia coli*, we discovered that m^{6}G in three sequence contexts tested was nearly 100% mutagenic *in vivo*, showing that the DNA polymerase

holoenzyme always placed T opposite m^6G during replication. In partially repair-proficient Ada^+ cells (*ogt*, *uvrB*), it was demonstrated that the Ada O^6-methylguanine-DNA methyltransferase repair protein (Ada) acted with twice the efficiency on m^6G when a G rather than an A was 5′ to the lesion.

Influence of sequence context on replication past m^6G by DNA polymerase III and repair of m^6G by Ada or Ogt or (Ada, Ogt, and UvrB simultaneously) in vivo using a sixteen nearest-neighbor analysis

As shown in Figure 10, there was no influence of sequence context on the DNA polymerase, as T was inserted ~100% of the time. It should be noted that this data set, as well as the wild-type set (Figure 15) were obtained using an older, unoptimized version of the REAP assay, which may explain the relatively higher background (as reflected in the non-m^6G containing genomes of sequence 5′-TGG-3′ and 5′-TAG-3′ in Figure 10 and Figure 15 vs. Figures 11-14, and also as reflected in the 5′-Gm^{6}GA-3′, 5′-Am^{6}GA-3′, and 5′-Tm^{6}GG-3′ genomes in Figure 10 vs. Table 3 (*ada*, *ogt*, *uvrB*) *E. coli*).

As shown in Figure 11 and Figure 12, in an *ogt*, *uvrB* cell line, the MFs ranged from 35% to 90%. Within this range, attributed to the effect of sequence context on m^6G repair by Ada, a 5′ sequence context influence was evident (Figure 11). The order of increasing MF attributed to Ada was, in general, GX < CX < TX < AX (X = m^6G). Ada also showed a 3′ influence of XPyrimidine < XPurine (Figure 12).

By contrast, as shown in Figure 13 and Figure 14, in an *ada*, *uvrB* cell line, the MFs ranged from 10% to 25%. Ogt demonstrated a 5′ influence of GX < AX (Figure 13), and the contexts 5′-AXN-3′ (N = any base) provided the highest MFs (Figure 14).

As shown in Figure 15, the MFs with respect to sequence context in the wild-type cell

line (Ada, Ogt, and UvrB all operative) revealed a pattern strikingly similar to that obtained using *ada*, *uvrB E. coli*, suggesting that Ogt is the dominant repair element for m^6G in uninduced E. coli.

The data in Figure 11 and Figure 12 were generated such that an identical, but independently constructed genome of sequence 5′-Tm^6GG-3′ was electroporated ~2 hours after the previous one for each independent set of experiments (A and B). This was a control. As shown for both independently constructed and electroporated genomes in sets A and B, there was practically no variance in MFs within each set for the control which, incidentally, further validated the REAP assay. This is an extremely important observation, since ~3 hours elapse between the time it takes for one person to electroporate and grow in suspension the first and the twentieth genome (sixteen nearest-neighbors and four controls). This instils confidence that the rather long duration of keeping electrocompetent cells on ice does not promote a change in the activity or concentration of the methyltransferases inside the cell. Had this occurred, the entire approach as outlined in Figure 8 would have to be abandoned, since the MFs of genomes within a set would vary as a function of when the genomes were electroporated relative to one another.

It is noteworthy that all MFs in Figures 11-14 set A were higher than those in set B. While the A and B designation is not important, it is important to observe that each MF within one set in a given cell line was higher than the corresponding MF within the other set in the same cell line. As explained in this section for the three nearest-neighbor analysis, this discrepancy was probably the result of a difference in methyltransferase concentration between the two batch

preparations of cells used. On a related note, a scatter plot of the average MFs between genomes from sets A and B (Figure 11 or Figure 12 (they are the same data set)) vs. the difference between the larger and smaller MFs (A - B) reveals a negative correlation (data not shown), which means that one may not simply subtract a constant from all MFs in one set to normalize it to the other set. The most likely explanation for this negative correlation is that m^6G in a sequence context that is relatively refractory to repair (5′-Am^{6}GA-3′) is less sensitive to slight changes of methyltransferase concentration that may occur for different preparations of electrocompetent cells. One can reach the conclusion that data sets for these types of sequence context experiments should not be combined. These observations lead to an interesting hypothesis concerning site-specific sequence context experiments in general, such as outlined in Figure 8: one should not combine the MF data sets from different cell preparations for any lesion when its repair is to be studied. This caveat notwithstanding, the sought after rank order of mutation frequency with respect to sequence context was the same in sets A and B. It is also apparent from the graphs (Figures 10-15) that no error bars are included. While data sets should not be combined (*vide supra*), one could obtain error bars by analyzing the MF from the progeny within each data set in triplicate. However, error analysis was deemed unnecessary for two reasons. First, this triplicate analysis was already done for simulated MFs (Chapter 2, Table 2) and for three m^6G genomes (Chapter 3, Table 3) and the variance was negligible. Second, the 5′-Tm^{6}GG-3′ duplicate control provided a strikingly similar MF.

A further caveat in studying lesions in single-stranded viral genomes is that the effect of repair systems that do not directly act upon the lesion, such as nucleotide excision repair (NER)

and mismatch repair (MMR), cannot be assayed *via* MF measurement. To explain further, consider the following scenario. If m^6G in a single-stranded genome is copied past, and the m^6G is excised in the form of an oligonucleotide, then no matter what the sequence context preference is for NER, the mutation fixation created by the DNA polymerase will dictate the MF. A corollary to this is that if the DNA polymerase always miscodes past a lesion in all sequence contexts, no matter what the sequence preference of NER is, a 100% MF would be seen in all contexts. This may be why MFs from genomes processed in (*ada*, *ogt* = NER influence) *E. coli* (using an older, unoptimized version of the REAP assay, (data not shown)) were similar to those obtained in (*ada*, *ogt*, *uvrB* = Pol III influence) (Figure 10). The same scenario is applicable to MMR, since futile exonucleic processing and strand resynthesis past the lesion will provide MFs as a function of the DNA polymerase, regardless of the repair event (such as the initiation of MMR by MutS binding to a lesion:base mismatch). However, again assuming the DNA polymerase always miscodes, one can perform sequence context experiments *in vivo* using a double-stranded vector if the initial lesion:base pair is of the type that would not be formed by a DNA polymerase (such as an m^6G:C pair as demonstrated in this work), since the repair event can occur prior to mutation fixation by the DNA polymerase.

In conclusion, we report a method for constructing site-specific single-stranded viral DNA genomes that should yield identical ligation efficiencies, regardless of the lesion or its surrounding sequence context. Using this methodology and repair-deficient *E. coli*, we discovered that m^6G in sixteen sequence contexts tested was nearly 100% mutagenic *in vivo*, showing that the DNA polymerase holoenzyme always placed T opposite m^6G during replication.

This suggests that the replicative DNA polymerase in *E. coli* does not contribute to the non-uniform distribution of m^6G derived mutational hot- or coldspots when cells are treated with simple alkylating agents. In partially repair-proficient cells, it was demonstrated that the Ada and Ogt O^6-methylguanine-DNA methyltransferase repair proteins acted on m^6G in a context dependent manner as shown in the Results section (3B). Since the mutational spectra from the m^6G genomes in wild-type and (*ada*, *uvrB*) *E. coli* were similar, Ogt appears to be the dominant repair element in uninduced *E. coli*. The REAP assay has been validated to provide rapid and accurate measurements of the MFs of m^6G in all sixteen nearest-neighbor sequence contexts.

3D. Methods

Synthesis of m^6G oligonucleotides

All oligonucleotides were made on an Applied Biosystems 391 DNA Synthesizer using the 0.2 μmol scale. Sixteen individual homogeneous 16-mers of sequence 5′ GAAGACC (G, A, T, or C)(m^6G)(G, A, T, or C)GCGTCC 3′ and a control oligonucleotide lacking guanine residues, 5′ TAATACCTm^6GTACA-TCC 3′ were made using the *N*-isobutyryl m^6G protected phosphoramidite (Glen Research). Deprotection was carried out by placing the controlled pore glass resin in dry glass vials equipped with septa to which were added, under argon, 1 mL 10% 1,8-diazabicyclo[5.4.0]undec-7-ene in anhydrous methanol (Aldrich). The solution stirred at room temperature for three weeks in a dark desiccator, after which the contents were transferred to microfuge tubes and lyophilized to an oil. After resuspension in 1 mL 10 mM NaOH, the DNA was ethanol precipitated, mixed with 50% formamide, and purified by migration through ~ 25 cm of a 20% denaturing polyacrylamide gel containing 7 M urea and 1 × TBE. Bands were excised, crushed and soaked with agitation in 10 mL water, and desalted on Sep-Pak® C-18 cartridges (Waters). Oligonucleotides were characterized as follows. In a 40 μL solution of 100 mM Tris buffer (pH 8.8) and 15 mM $MgCl_2$ was digested approximately 2 nmol of each oligonucleotide with 300 mU snake venom phosphodiesterase I (ICN Biomedical) and 20 U alkaline phosphatase (type VII-T from bovine intestinal mucosa, Sigma no. P6774) at 37°C for 1 hr. All protected and deprotected nucleosides were separated by gradient reverse phase HPLC using an analytical Beckman Ultrasphere™ 5 μ C-18 column (250 × 4.6 mm) and a flow of 1

mL/min. The linear gradient was 0 to 8.2% B over 35 min, then 8.2 to 30% B over 25 min (solvent A, 0.1 M NH_4OAc; solvent B, CH_3CN (neat)). Detection and quantitation were performed with an in-line diode array detector (Hewlett Packard 1040A). A comparison of the retention times and ultraviolet (UV) absorbance peak profiles between a mixture of authentic standards and the nucleoside digests from the synthesized m^6G oligonucleotides revealed the following. The isobutyryl group had been completely removed from m^6G, no guanine was formed at the lesion site (as judged by the identically deprotected control oligonucleotide), and no O^6-ethylguanine was created during the ethanol precipitation step (data not shown). The limit of detection was at least > 0.5% for each event monitored.

Construction of site-specific, single-stranded viral genomes

M13mp7L2 was prepared as follows. To a 2 L baffled flask was added 1 L 2 × YT media (Ausubel et al., 1993) and 2 mL of a saturated culture of GW5100 *E. coli* (JM103, $P1^-$; G. C. Walker, Massachusetts Institute of Technology). The flask was shaken (275 RPM) at 37°C for 2.5 hr, after which was added 2 mL of saturated M13mp7L2 progeny phage supernatant (~6×10^9 plaque forming units (pfu)) and growth continued for 9 hr. Cells were pelleted and a solution that was 20% polyethylene glycol (MW 8000, Sigma no. P2139) and 2.5 M NaCl was added to the supernatant (1:4 v/v). The phage in the supernatant were precipitated at 4°C overnight, pelleted, combined and resuspended in 19 mL of TE (10 mM Tris-HCl, 1 mM EDTA (pH 8.0)), split into thirds, and extracted with 5 × 3 mL phenol:chloroform:isoamyl alcohol (25:24:1 v/v/v) until the aqueous-organic interface remained clear. The aqueous phase (~15 mL) was applied to a drained column (35 mm × 10 mm) containing 1 g DNA-Grade Bio-Gel® HTP hydroxylapatite resin (BioRad), which had been conditioned with 5 mL TE (an acceptable flow-rate was achieved by applying moderate air pressure over the solvent bed). The column was then washed with 5 mL TE and the DNA was eluted with a solution of phosphate buffer (0.07 M KH_2PO_4, 0.07 M K_2HPO_4) at the molarity suggested by the manufacturer that would elute single-stranded DNA. Fractions containing the most concentrated DNA were pooled (~5 mL), split in half, and concentrated/dialyzed through two Centricon-100 spin-dialysis devices (Amicon no. 4212) using three, 2 mL TE rinses. The final yield of single-stranded M13mp7L2 was ~4.5 mg.

Genomes containing m^6G were constructed as follows. *Eco*R I, T4 polynucleotide kinase, and T4 DNA ligase were from New England Biolabs. ATP and exonuclease III were from

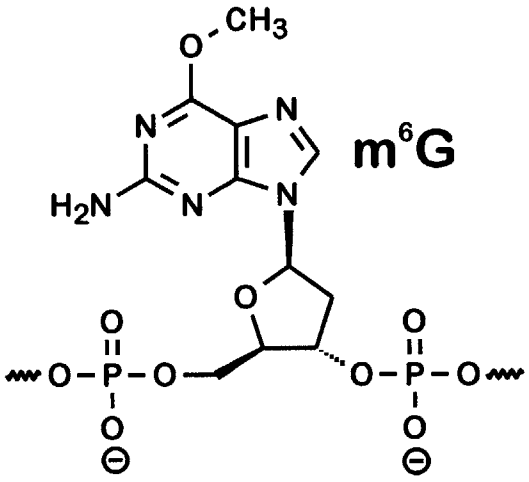
Pharmacia. Oligonucleotides containing m^6G (10 pmol) were 5′-phosphorylated in a 30 μL volume with T4 polynucleotide kinase (0.5 U/μL) in 70 mM Tris-HCl buffer (pH 7.5), 10 mM $MgCl_2$, 5 mM DTT, and 1 mM ATP at 37°C for 1 hr. Single-stranded M13mp7L2 (0.15 pmol/μL, (2.36 μg/pmol)) was linearized with *Eco*R I (0.37 U/μL) in 100 mM Tris-HCl buffer (pH 7.5), 50 mM NaCl, 10 mM $MgCl_2$, and 100 μg/mL bovine serum albumin at 23°C for 4 hr, after which, an equimolar amount of the scaffolds 5′ GGTCTTCCACTGAATCATGGTCAT AGC 3′ and 5′ AAAACGACGGCCAGTGAATTGGACGC 3′ were added, with negligible volume increase. Approximately 100 μL of this solution (10 pmol in each component) was added to the tube containing the phosphorylated oligonucleotide. The components were annealed in a PCR machine using the following program: 80°C for 5 min, 80°C to 50°C at 1°C /min, 50°C to 0°C at 0.33°C /min. The solution was made 10 mM DTT, 1 mM ATP, and 5.4 U/μL T4 DNA ligase, with negligible volume increase, after which it was incubated at 16°C for 2 hr. Scaffolds were degraded to approximately 4-mers (data not shown) by incubation with exonuclease III (1 U/μL) at 37°C for 2 hr. The constructed genomes were desalted by spin-dialysis through Centricon-100 devices using two, 2 mL rinses with TE, after which each sample was divided equally into eight tubes (~8 μL/tube) and stored at −20°C.

Electroporation of genomes into repair-deficient isogenic E. coli

The drug-resistant alleles Δ*ada*-25::Cam^r (Shevell et al., 1988), *ogt*-1::Kan^r (Rebeck and Samson, 1991), and *uvr*B5::Tn*10*Tet^r (Backendorf et al., 1986) had been previously used to construct the triple mutant *E. coli* strain CJM2 (*ada*, *ogt*, *uvrB*) (Samson et al., 1997) by P1 *vir* transduction into the wild-type strain FC215 (Mackay et al., 1994). We also used P1 *vir*

transduction (Miller, 1992) to transfer the *ogt*-1::Kan^r and *uvr*B5::Tn*10*Tet^r alleles from CJM2 into FC215 to create the isogenic repair-deficient *E. coli* strain C216/uvrB$^-$ (*ogt*, *uvrB*); the same procedure was performed to transfer the Δ*ada*-25::Cam^r and *uvr*B5::Tn*10*Tet^r alleles from CJM2 into FC215 to create the isogenic repair-deficient *E. coli* strain C217/uvrB$^-$ (*ada*, *uvrB*) Repair-deficient *E. coli* cells were made electrocompetent as follows. A 5 mL saturated culture of cells was used to innoculate 500 mL LB (supplemented with 2.5 g/L $MgSO_4$, 0.2% (w/v) maltose, and the appropriate drugs at concentrations of 10 μg/mL for chloramphenicol, 50 μg/mL for kanamycin, and 15 μg/mL for tetracycline). After growing with aeration in a baffled flask at 37°C, the cells were harvested at mid-log phase (OD_{600} = 0.5). At 0°C, cells from 350 mL of culture were washed thoroughly with 350 mL, then 175 mL of ice-cold sterile water. Cells were finally resuspended in 4 mL 10% glycerol, providing ~2 × 10^9 cells per 100 μL, as judged by counting plated dilutions of cells. Approximately 100 μL of electrocompetent cells were mixed with a thawed aliquot of desalted genome construct (*vide supra*). To a chilled electroporation cuvette (0.2 cm gap) containing the above solution was applied 2.5 kV at 129 ohms from an Electro Cell Manipulator 600® electroporation system (BTX). The voltage delivered (2.36 kV) and time constant (5.60 msec) were consistent between electroporations for all genomes and cell lines, and rendered a reduction in cell survival (electroporation of cells alone) by no more than 60%. Immediately, 1 mL SOC media (Sambrook et al., 1989) was added to the cuvette and the contents transferred to 9 mL LB. The number of initial independent events was determined by immediately plating a dilution of the mixture onto a lawn of NR9050 cells (*vide infra*). One million electroporated cells secreted phage and were thus independently transformed with the M13 construct. The 10 mL culture was immediately placed on a roller drum at 37°C for 2.5 hr,

after which the supernatant containing progeny phage at ~ 10^{10} pfu/mL to be used for MF analysis was stored at 4 °C.



5′ GAAGACCG**m**^{6}**G**AGCGTCC 3′
5′ GAAGACCA**m**^{6}**G**AGCGTCC 3′
5′ GAAGACCT**m**^{6}**G**GGCGTCC 3′

Figure 7. O^6-Methylguanine in three nearest-neighbor sequence contexts used in further validating the REAP assay.

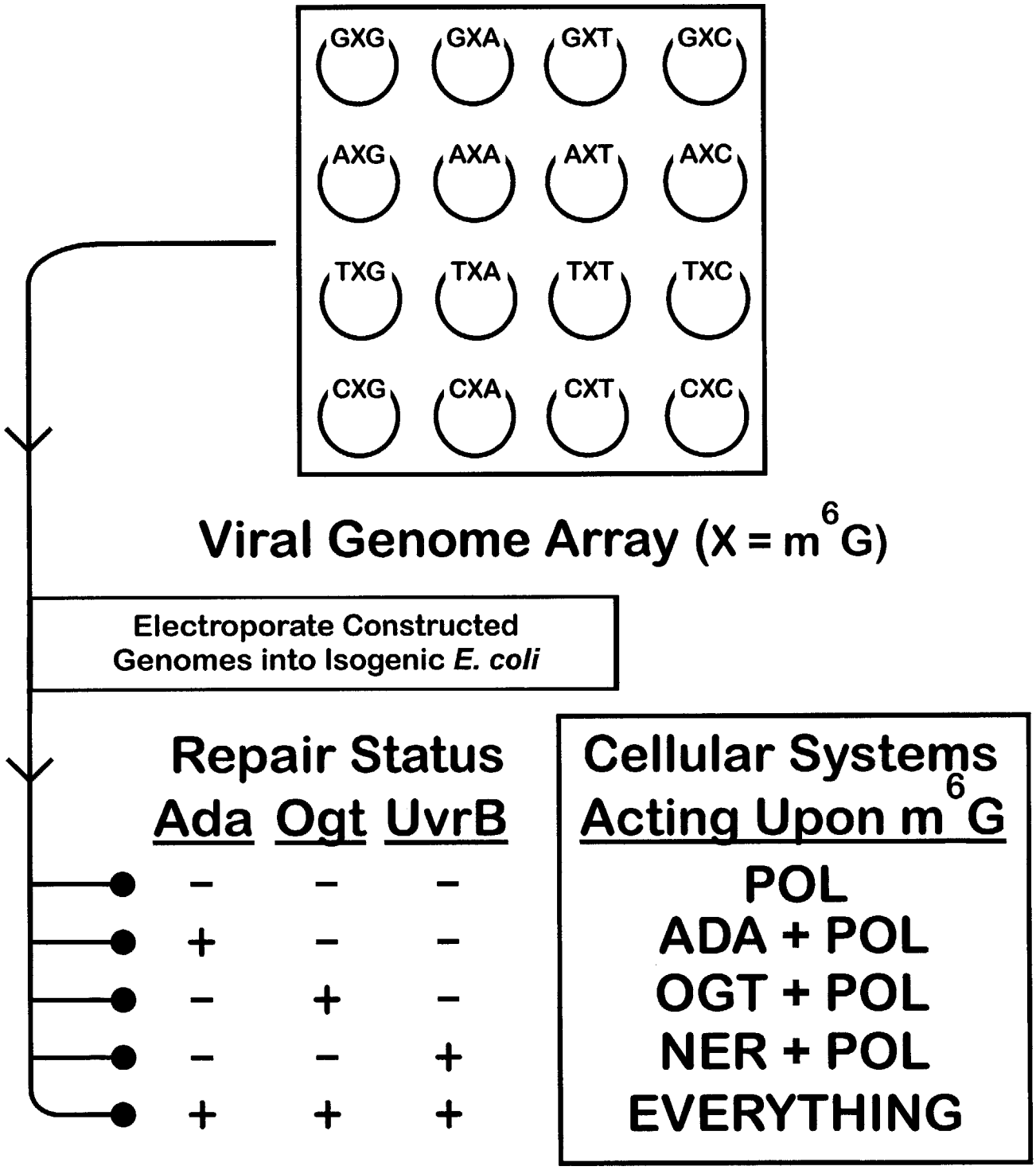


Figure 8. Overall scheme for determining if bases immediately flanking m^6G influence its biological processing *in vivo*. Differences in MFs among m^6G genomes for a given cell type will be attributed to sequence selectivity of the DNA polymerase and operative repair protein.

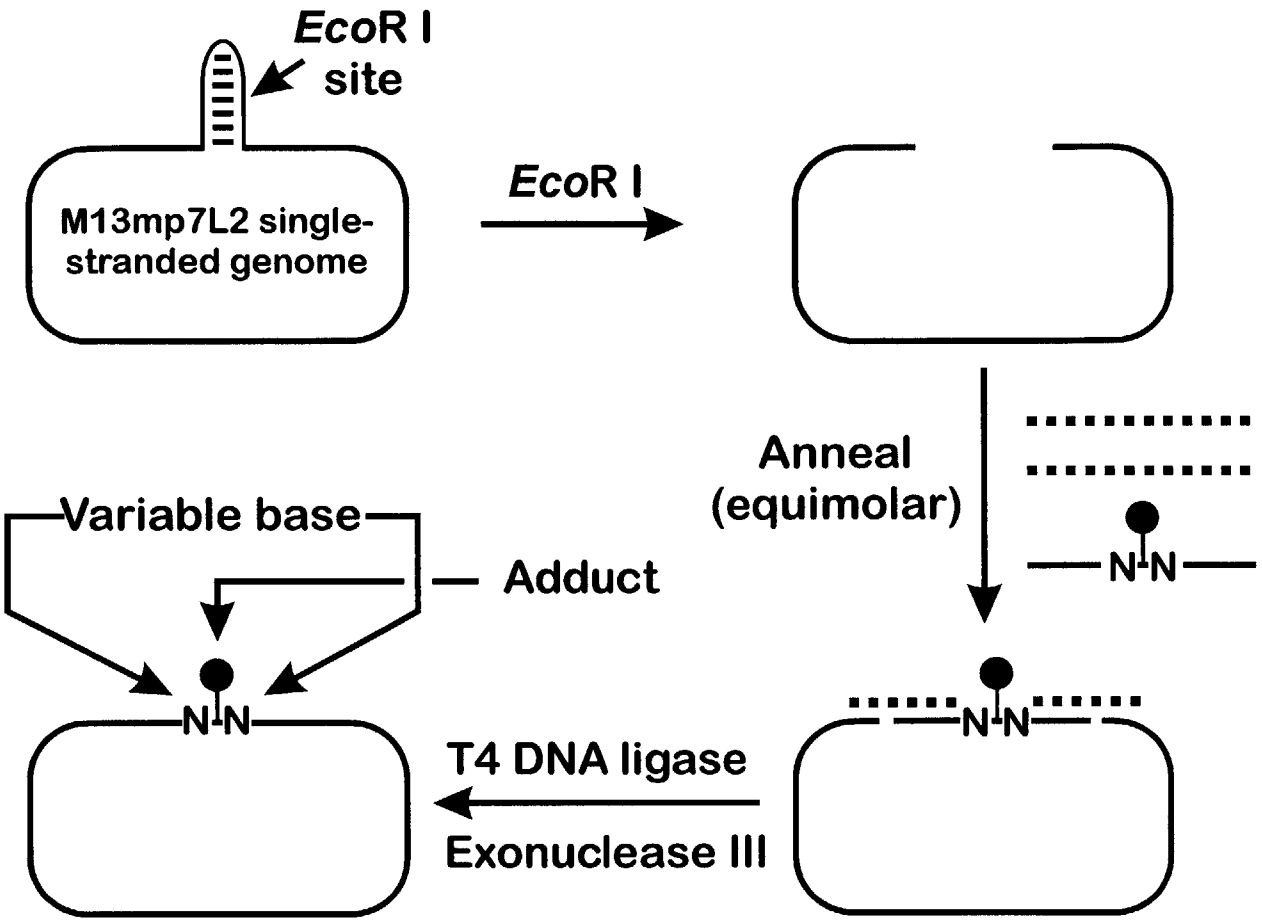


Figure 9. Scheme for constructing single-stranded genomes containing an adduct in any variable surrounding sequence context. The M13 single-stranded circular genome contains an *Eco*R I site that is in a hairpin. Cleavage with *Eco*R I yielded a linear product. Scaffolds (indicated by broken lines) of 27 and 26 nucleotides were annealed to the 3′ and 5′ ends, respectively, of the M13 genome. The same mixture included a 16-mer in which m^6G was present in a defined sequence context (indicated by -N(lollipop)N-). Ligation covalently inserted the adduct-containing oligonucleotide into the M13 genome, and exonuclease III removed the scaffolds, affording a single-stranded, adduct-containing genome. It is noteworthy that the scaffold did not overlap with the -N(lollipop)N- trinucleotide; this feature is believed to promote a high ligation efficiency.

Table 3: Influence of sequence context on m^6G replication and repair in three sequence contexts analyzed in triplicate by the REAP assay.

	Sequence (5′ to 3′)	Cellular Repair Status			Output MF (REAP)			
		Ada	Ogt	UvrB	Set 1	Set 2	Set 3	Average*
	Gm^6GA	no	no	no	99.1	99.2	98.8	99.1 ± 0.2
	Am^6GA	no	no	no	99.8	99.6	99.6	99.6 ± 0.1
	Tm^6GG	no	no	no	99.2	99.3	99.3	99.3 ± 0.1
Experiment 1								
	Gm^6GA	yes	no	no	ND	53.5	53.1	53.3 ± 0.3
	Am^6GA	yes	no	no	90.5	90.6	90.2	90.4 ± 0.2
	Tm^6GG	yes	no	no	82.1	82.1	81.9	82.0 ± 0.1
	Gm^6GA	no	no	no	97.9	97.5	97.9	97.8 ± 0.2
	Am^6GA	no	no	no	99.8	99.7	99.7	99.7 ± 0.1
	Tm^6GG	no	no	no	98.2	97.8	98.1	98.0 ± 0.2
Experiment 2								
	Gm^6GA	yes	no	no	44.8	45.0	44.4	44.7 ± 0.3
	Am^6GA	yes	no	no	90.6	90.3	88.1	89.7 ± 1.4
	Tm^6GG	yes	no	no	81.3	81.7	80.8	81.2 ± 0.5

Single-stranded genomes containing m^6G in three different sequence contexts were constructed, electroporated into two strains of repair-deficient *E. coli*, and their subsequent mutation frequencies were determined using the REAP assay. Each experiment refers to an independent genome construction and subsequent electroporation, whereas genomes containing identical sequence contexts within the same experiment were from the same genome construction.

*The average output MF determined by the REAP assay was based on MFs obtained using the double-stranded DNA output from three cultures of *E. coli* that were independently infected with the same progeny phage pool obtained after electroporation, and is reported as ± one standard deviation. ND, not determined.

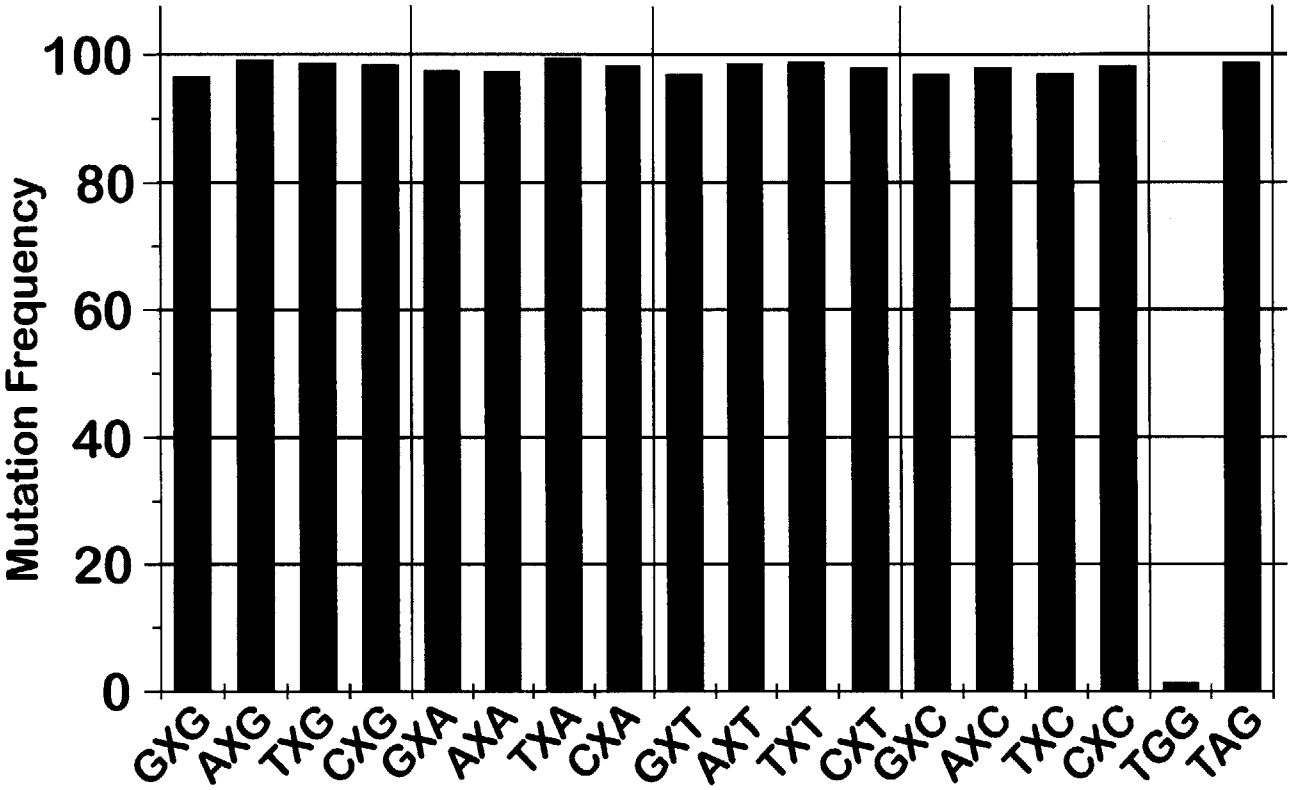


Figure 10. Context preference of DNA polymerase on m^6G in (*ada*, *ogt*, *uvrB*) *E. coli*. Sequences are read 5′ to 3′ and data are grouped such that the 5′ end is varied in the order G, A, T, and C as read from left to right. This data set was generated from constructed viral genomes once, using an older, non-optimized version of the REAP assay. X was m^6G; the pure wild-type (5′-TGG-3′) and mutant (5′-TAG-3′) genomes were analyzed in parallel.

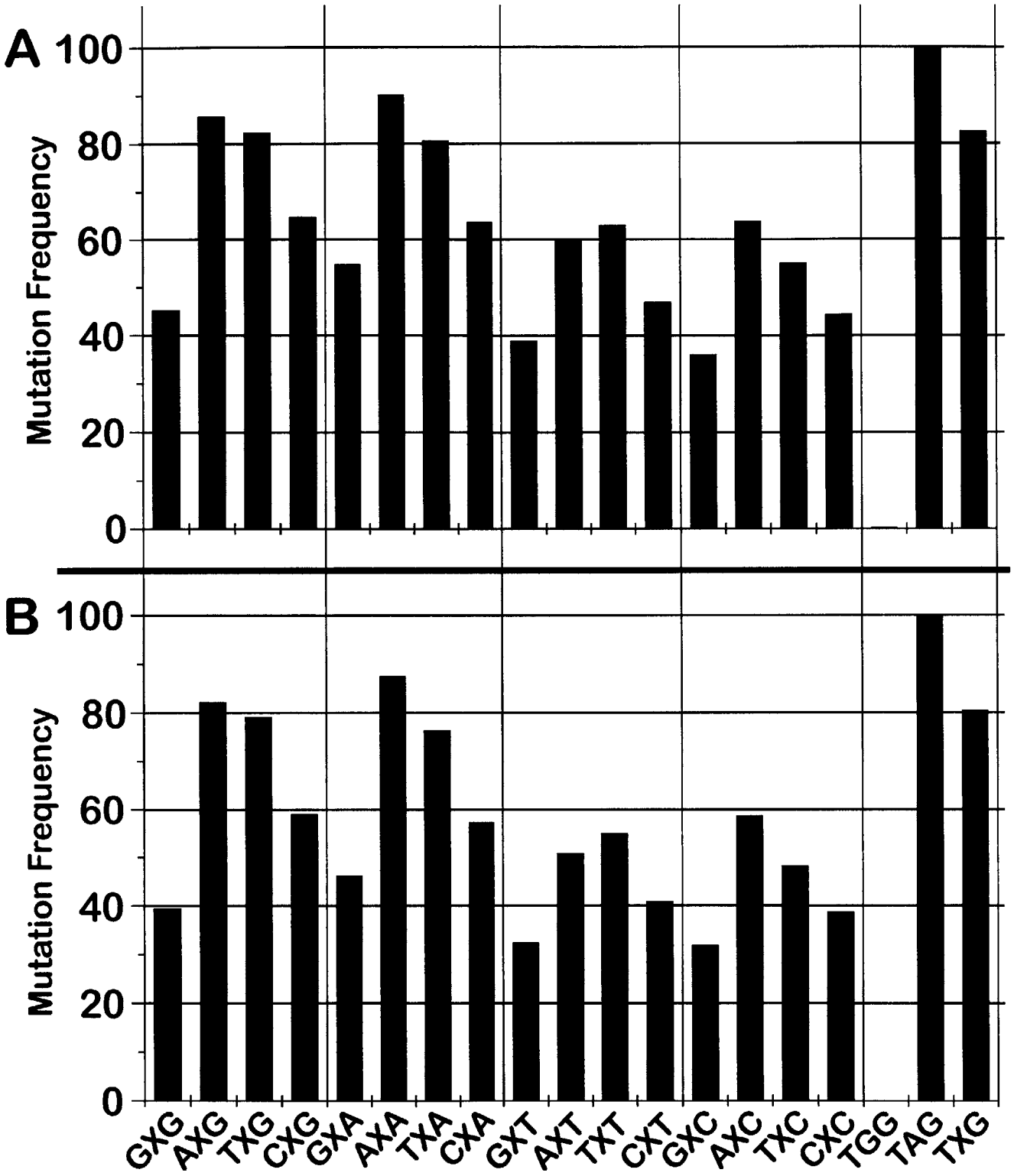


Figure 11. Context preference of Ada on m^6G in (*ogt*, *uvrB*) *E. coli*. Sequences are read 5′ to 3′ and data are grouped such that the 5′ end is varied in the order G, A, T, and C as read from left to right. The data sets (A and B) were generated from two independently constructed and electroporated sets of viral genomes, using the optimized REAP assay described in Chapter 2. X was m^6G; the pure wild-type (5′-TGG-3′), mutant (5′-TAG-3′), and a duplicate 5′-TXG-3′ genome electroporated ~2 hr after the original were analyzed in parallel.

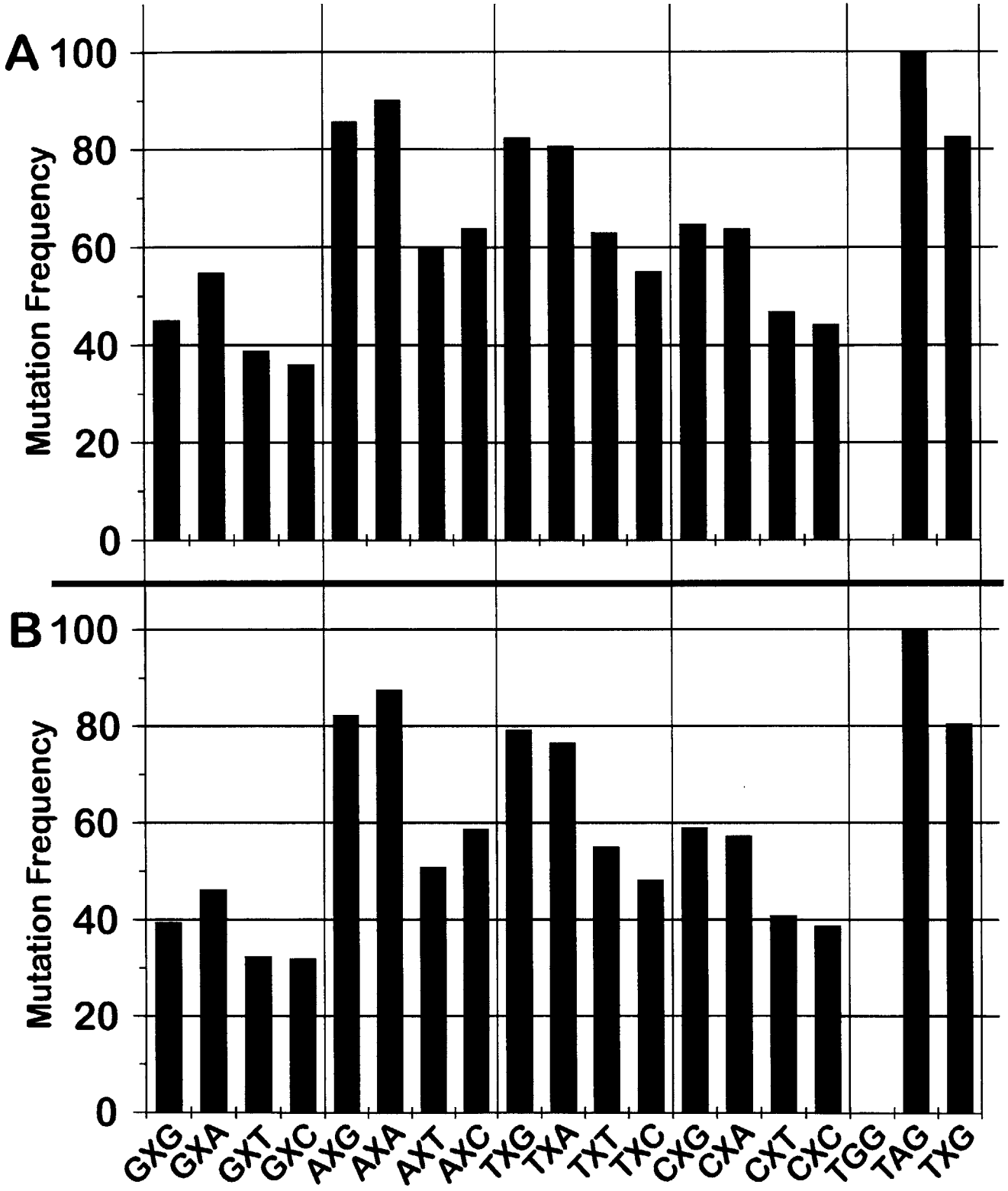


Figure 12. Context preference of Ada on m^6G in (*ogt*, *uvrB*) *E. coli*. This is the same data set as in Figure 11; however, the data are grouped such that the 3′ end is varied in the order G, A, T, and C as read from left to right.

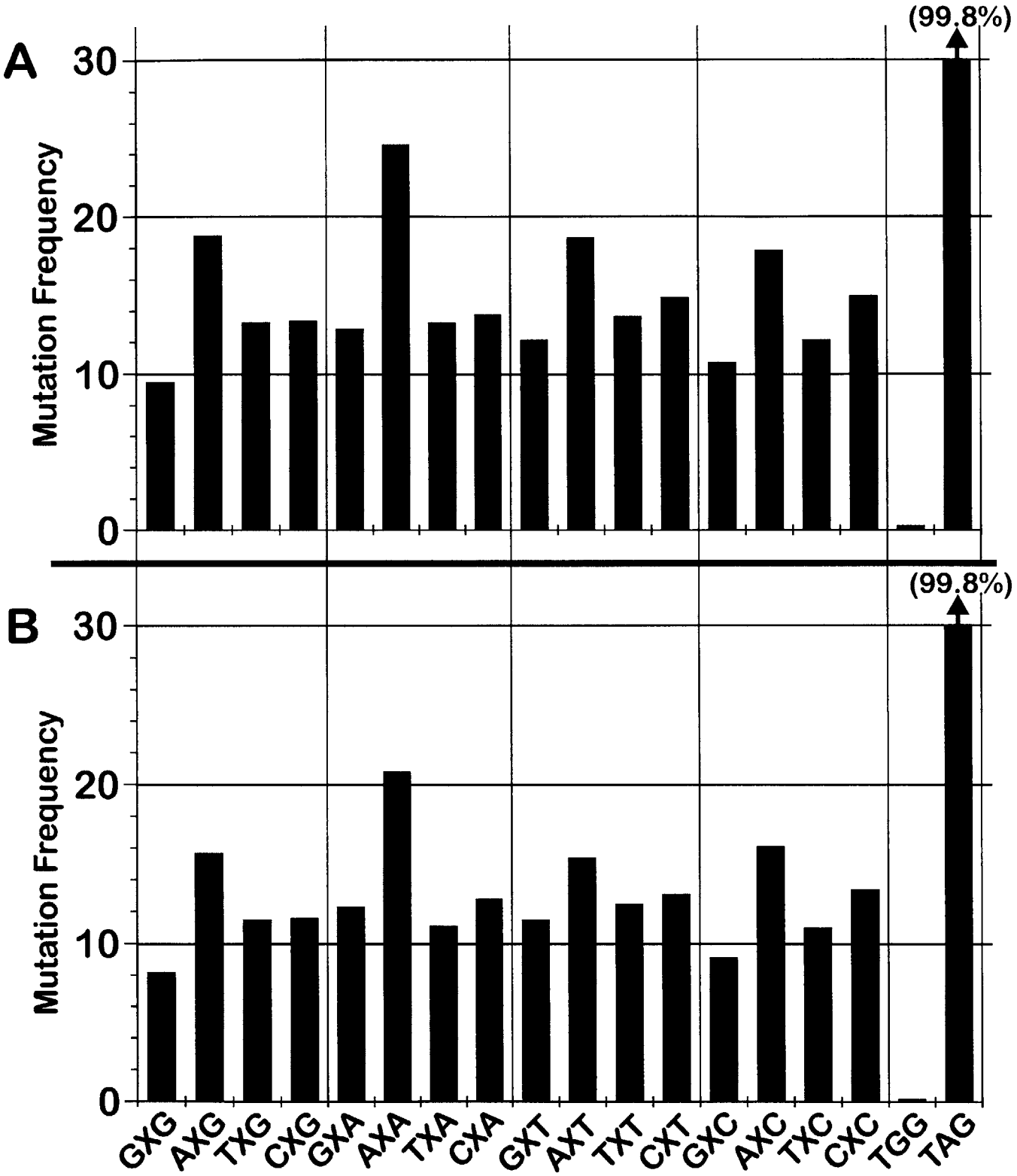


Figure 13. Context preference of Ogt on m^6G in (*ada*, *uvrB*) *E. coli*. Sequences are read 5′ to 3′ and data are grouped such that the 5′ end is varied in the order G, A, T, and C as read from left to right. The data sets (A and B) were generated from two independently constructed and electroporated sets of viral genomes, using the optimized REAP assay described in Chapter 2. X was m^6G; the pure wild-type (5′-TGG-3′) and mutant (5′-TAG-3′) genomes were analyzed in parallel.

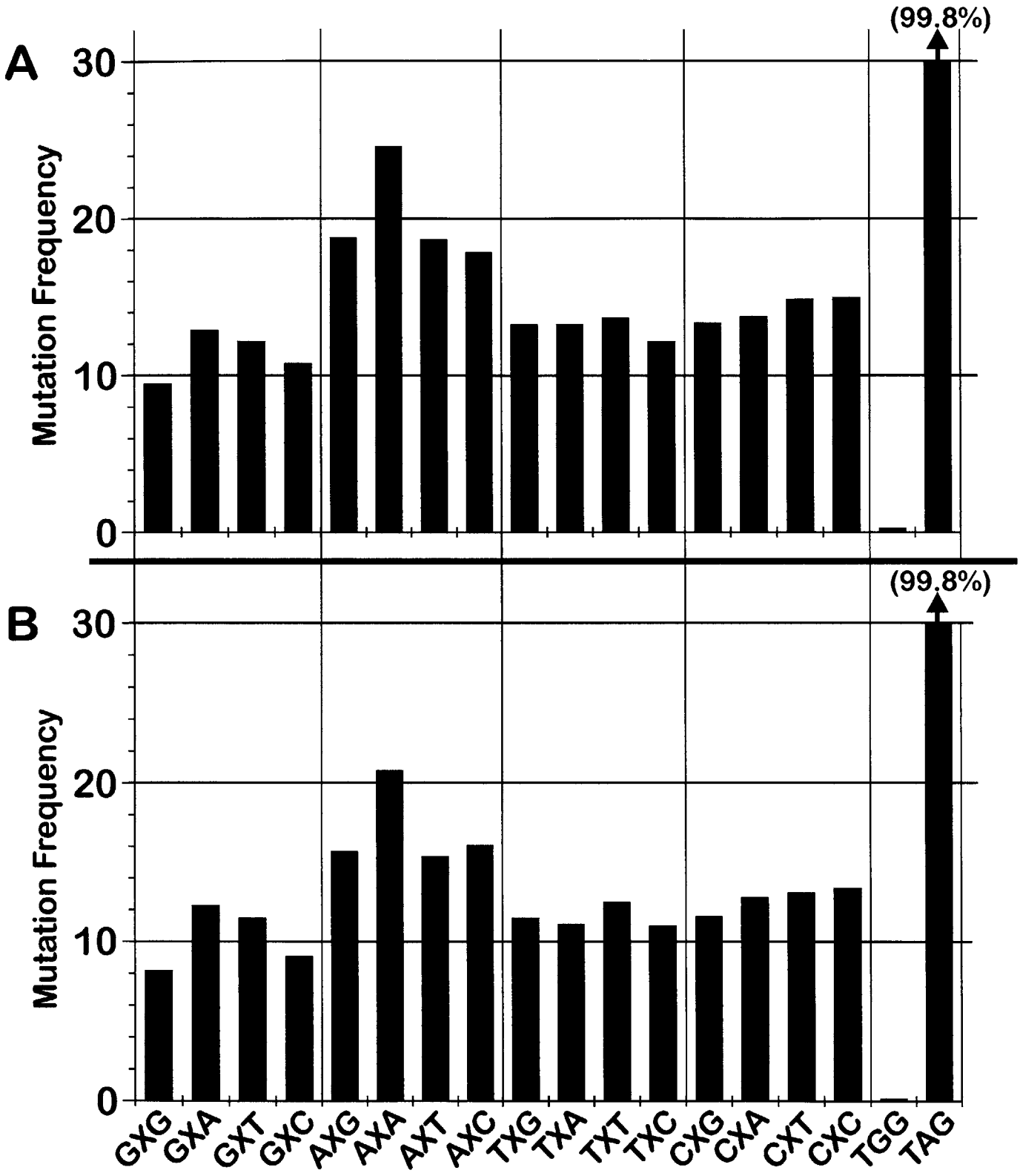


Figure 14. Context preference of Ogt on m^6G in (ada, *uvrB*) *E. coli*. This is the same data set as in Figure 13; however, the data are grouped such that the 3′ end is varied in the order G, A, T, and C as read from left to right.

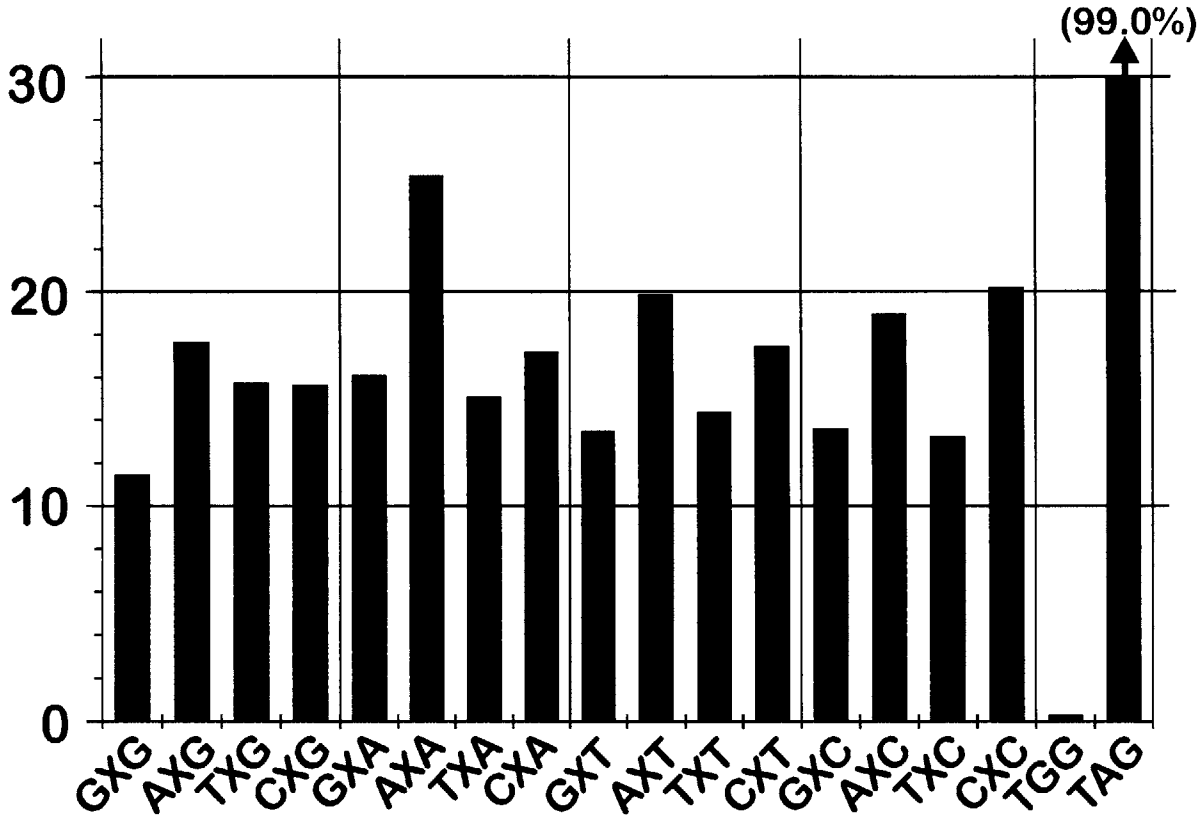


Figure 15. Mutation frequency of m^6G with respect to sequence context in wild-type *E. coli*. Sequences are read 5′ to 3′ and data are grouped such that the 5′ end is varied in the order G, A, T, and C as read from left to right. This data set was generated from constructed viral genomes once, using an older, non-optimized version of the REAP assay. X was m^6G; the pure wild-type (5′-TGG-3′) and mutant (5′-TAG-3′) genomes were analyzed in parallel.

CHAPTER 4:

Proposals to Deconvolute Mutational Spectra

4A. Introduction

This section of a dissertation usually contains a brief description of the next steps that would occur during a research program. Instead, I have elected to write two detailed plans that I hope will help future workers achieve the long-range objective of our program - specifically, to understand how context affects the mutagenic activity of a DNA lesion. The first part will describe a single assay whereby the effects of sequence context on lesion repair, adduct formation, or adduct degradation can be assessed *in vitro*; the second part will describe an assay based on mass spectrometry (MS) whereby the repair of any DNA lesion whose mutational specificity is known can be studied in 16 to 256 sequence contexts simultaneously *in vivo*. This mass spectrometry approach may also be used to obtain consensus sequences surrounding a DNA lesion that are blocks to a DNA polymerase *in vivo*, and may be additionally used to determine in what sequence contexts a lesion is more apt to miscode by the replicative DNA polymerase *in vivo*. The ideas presented in this chapter are all original.

4B. Proposed Assay for Determining the Influence of Sequence Context on Lesion Repair, Adduct Formation, or Adduct Degradation *in vitro*

I have designed an assay that may help provide a complete mechanistic rationale for mutational hotspots. The *in vitro* assay, based on "prelabeling," will rapidly assess the effect of sequence context on the repair of any DNA lesion by any repair protein. The same assay can also be used to determine hotspots for initial DNA damage in any form, such as adduct formation, ring fragmentation, depurination, deamination, etc., by any source, such as alkylating agents, UV light, reactive oxygen species, gamma irradiation, etc., in a dose-dependent manner. One can also use the method to address how sequence context affects secondary damage events by analyzing the fate of a site-specific lesion when treated with sources of damage (e.g., 7,8-dihydro-8-oxoguanine (8oxoG) with type I and type II photosensitizers or peroxynitrite) (Figure 16). Furthermore, an efficient method for constructing sixteen individual oligonucleotides with a centrally located base flanked by all nearest-neighbor base permutations to achieve this goal is presented (Figure 17). The assay can be used on single- and double-stranded DNA, and, in addition to looking at programmed nearest-neighbor effects in a site-specific fashion as described in this dissertation, one can take an alternative route and program the "viewing window" along several positions of a reporter gene that one wishes to study. I intend to use this assay to determine the relative rates of m^6G repair in sixteen nearest-neighbor sequence contexts in single- and double-stranded (m^6G opposite T or C) oligonucleotides by the Ada and Ogt methyltransferases using the same sequence context that flanked the variable region

described previously in this thesis. If the consensus sequences attributed to differential repair by Ada and Ogt in *E. coli* presented in this dissertation are similar to those found using this proposed assay with m^6G paired opposite T, but not with single-stranded DNA, then it is very likely that the effects seen in *E. coli* were due to direct action of the methyltransferase on duplex DNA (an m^6G:T pair) and, if shielding by a repair protein such as MutS did occur and hinder access of the methyltransferase to m^6G repair, it did so irrespective of sequence context.

One can easily study the *in vitro* repair of end-labeled oligonucleotides that contain site-specific DNA lesions, if the lesions are susceptible to base excision repair, since initiation of the repair event by the glycosylase results in the formation of an abasic site that can be cleaved with piperidine and many samples can be analyzed quickly on a denaturing polyacrylamide gel. Below are listed many of these glycosylases, the gene from which they are coded (italicized in parentheses), and their substrates (Wilson III et al., 1998; Friedberg et al., 1995). The formamidopyrimidine-DNA-glycosylase (*fpg*/*mutM*) removes 8oxoG, 2,6-diamino-4-hydroxy-5-*N*-methylformamidopyrimidine and 4,6-diamino-5-*N*-methylformamidopyrimidine (the imidazole ring-opened forms of 7-methylguanine and 7-methyladenine, respectively), 5-hydroxycytosine, and 5-hydroxyuracil. The 3-methyladenine-DNA-glycosylase I (*tag*) removes 3-alkylated purines. The 3-methyladenine-DNA-glycosylase II (*alkA*) removes 3- and 7-alkylated purines, O^2-methylpyrimidines, $1,N^6$-ethenoadenine, 5-formyluracil, and 5-hydroxymethyluracil. Thymine glycol-DNA-glycosylase (*nth*) removes thymine glycol, uracil glycol, 5-hydroxycytosine, 5-hydroxyuracil, urea, and β-ureidoisobutyric acid. Uracil-DNA-glycosylase (*ung*) removes uracil and 5-hydroxyuracil. MutY (*mutY*) removes adenine of an

A:8oxoG, G or C pair.

By contrast to the lesions described above, the repair of *O*-alkylated bases by methyltransferases is harder to monitor, since repair results in a normal G or T from O^6-alkylguanine or O^4-alkylthymine, respectively. Methods that have been developed to measure *O*-alkylated base repair include separating repaired from unrepaired end-labeled oligonucleotides using reversed-phase HPLC (Graves et al., 1987; Dolan et al., 1988), immunoprecipitating oligonucleotides containing m^6G from those containing G using an antibody raised against the lesion (Souliotis et al., 1989; Mironov et al., 1989), and using an m^6G substrate containing ^{32}P immediately 3′ to the lesion with subsequent degradation of the substrate to 3′ deoxynucleotide monophosphates which are separated by HPLC (Klein and Oesch, 1990).

The proposed assay outlined in Figure 16 has several advantages over the aforementioned methods concerning alkylation repair and can also be used to study the many base excision repair systems listed above. Since both oligonucleotide retention time on an HPLC column and antibody binding to m^6G are probably dependent on sequence context, optimization for each sequence context to be tested is not needed. The method described here is versatile and amenable to high-throughput. Moreover, unlike the technique of Klein and Oesch (Klein and Oesch, 1990) which shows a high background for HPLC separation of 3′ radiolabeled monophosphates, substrate creation does not rely on incorporation *via* dm^6GTP at restriction site overhangs by Klenow. While labeling and ligating an oligonucleotide containing a modified base at the 5′ end may not be efficient, only a small amount of ligated ^{32}P material is needed.

Equal radioactive counts will be used assuring equal molar amounts of each sequence context to be studied so that the sought after relativistic rank order of base change at the interrogation position can be found. If a substantial amount of material is desired for these studies, then all one has to do is add equal molar amounts of cold, full-length oligonucleotides to the appropriate tubes, each containing equal molar amounts of internally radiolabeled sample.

This versatile technique also has advantages and disadvantages with respect to mass spectrometry, whereby the fate of a site-specific base or lesion (labeled with ^{2}H, ^{15}N, and/or ^{13}C) in one defined sequence context is monitored. For mass spectrometry to faithfully monitor, at a specific site, lesion formation from a normal base, or lesion repair and/or lesion transformation with respect to sequence context, a stable isotope should be incorporated, which can be expensive. This is mandatory for situations where the monomerized product is identical to any component of monomerized substrate (e.g., G from $m^{6}G$ *via* Ada or Ogt repair (other Gs exist) or abasic sites and ring-fragmentation products from 8oxoG/oxidant treatment (these would form also from the other bases)).

While mass spectrometry can accurately detect, at a specific site, the relative amount of a modified base product with respect to its unmodified base if that base had been isotopically labeled (e.g., deuterated $m^{6}G$ from deuterated G *via* MNU, or deuterated 8oxoG from deuterated G *via* oxidants), the absolute value of products (and unreacted base) may need to be determined by calibrating the system (co-injection) with yet another stable isotope, different in mass from the one already employed, since unlabeled sites containing guanine outside the site-specific

"monomer window" will be converted to similar products non-uniformly (thus barring the possibility of using a non-isotopically labeled standard). The same argument holds for site-specific adduct degradation studies if changes in the deuterated site-specific adduct do not overwhelm the formation of said lesion or degradation products from bases outside the "monomer window;" if this is not the case, then one can simply co-inject a known amount of non-deuterated standard. The absolute value of products, however, can easily be determined using the methodology proposed in Figure 16, since the specific activity of ^{32}P will be known. While providing a lower signal, ^{33}P can likewise be used (e.g., with 8oxoG experiments) to abate the possibility of radiolysis due to ionizing radiation.

Some disadvantages of the proposed technique are that unstable products may not be seen due to the time it takes to degrade the DNA to 5′ dNMPs and resolve these either by TLC or HPLC. A corollary to this possible limitation is that the product analysis should be delayed until relatively stable compounds are formed. This will prevent smearing on a TLC plate or infinite band broadening on an HPLC column that may result if product conversion occurs during separation. Also, as the number of possible (stable) products in a given experiment increases, resolution will probably be disproportionately harder to achieve than when MS is used (although the same sample can be run in different TLC or HPLC solvent systems and the relative amounts of all products can be determined by comparing the data sets).

A possible complication of the proposed assay concerns modification or reaction of the DNA backbone. For example, treatment of DNA with MNU will give not only m^6G, but also

~25% of methylphosphotriesters (as demonstrated with RNA (Singer et al., 1975)), which may prevent the snake venom phosphodiesterase from cleaving the phosphate linkage. Products other than monomers containing a 5′ ^{32}P-phosphate and a 3′ hydroxyl are defined as "background" for DNA lesion formation from both alkylating agents and ionizing radiation (*vide infra*). The multiplicity of background products from incomplete SVPD cleavage may complicate the TLC analysis of radiolabeled products within the "monomer window." The radiolabeled monomers resulting from cleavage of phosphodiesters with SVPD will be liberated as a mixture of 5′ dGMP, 5′ dm^{7}GMP, 5′ dm^{6}GMP, 5′ dm^{3}GMP, and possibly 5′ dRiboseMP from depurination of 5′ dm^{7}GMP; however, 16 radiolabeled dimers containing a phosphotriester linkage (which would probably prevent cleavage by SVPD) for each of these products will also exist (each can have a G, A, T, or C as the 5′ or 3′ base and be in either the R or S stereoisomeric configuration), as will 32 radiolabeled dimers from the phosphotriesters of normal bases (both stereoisomers). Although these dimers will contain the same -2 charge as for the 5′ dNMPs, the doubling of mass may make all dimers migrate very differently on a PEI-TLC plate than the desired monomers. Treatment of phosphotriesters with the N-terminal domain of Ada prior to snake venom phosphodiesterase will, at best, only reduce the number of dimers by half, since dimers containing the R stereochemical configuration will not be repaired (Hamblin and Potter, 1985).

While the overall assay may have to be modified for experiments designed to study the effect of sequence context on DNA adduct formation by alkylating agents (*vide supra*), it may also have to be modified when lesion formation or degradation by ionizing radiation or chemicals that generate free radicals is examined. These damaging agents can cleave the phosphodiester

backbone and, while the majority of newly formed 5′ ends retain their 5′ phosphate, the majority of 3′ ends contain either 3′ phosphate (~70%) or 3′ phosphoglycolate (~30%) residues (Henner et al., 1983b). This implies that the sugar linking the two phosphates is released, and this is seen in the form of base propenals, which do not contain phosphate and will therefore not have a negative impact on the assay. Further ramifications of these observations with respect to the proposed assay are shown in Figure 18. While the radiolabeled monomer formed from release of a base propenal 5′ to the "monomer window" (Figure 18, from the deoxyribose labeled as 1) will have no untoward effects regarding the analysis as outlined in Figure 16, possible formation of base propenals at the monomer window or 3′ to it (Figure 18, from the deoxyribose labeled 2 or 3, respectively) will give radiolabeled monomers containing not only 5′ phosphate, but also 3′ phosphate or 3′ phosphoglycolate residues. As with the alkylation scenario, this multiplicity of radiolabeled background products may make the TLC analysis of products within the "monomer window" more complex. It is possible that the two extra negative charges in these "background" products will cause retention of monomers containing 3′ phosphate or 3′ phosphoglycolate near the origin when spotted on TLC plates containing a weak anion-exchange matrix, such as PEI-cellulose; thus, they may well be separated from the desired 5′ radiolabeled monophosphates of the "monomer window" without further interference. If this is not the case, one can create a free 3′ hydroxyl group at the 3′ end of a single-stranded break using DNA repair enzymes prior to monomerization by SVPD. Exonuclease III, while requiring a double-stranded substrate, can remove both 3′ phosphate and 3′ phosphoglycolate residues (Henner et al., 1983a). Most experiments will probably be performed in duplex DNA anyway (or one could anneal an excess of cold complementary strand prior to exonuclease III treatment). One may also be able to

perform the entire analysis on single-stranded DNA by treating samples with a mixture of T4 polynucleotide kinase to remove 3′ phosphates and the newly discovered exonuclease IX of *E. coli* (Sandigursky and Franklin, 1998) to remove 3′ phosphoglycolates prior to SVPD digestion.

If a DNA base or lesion is damaged by a source of ionizing radiation, although far lower in concentration than base propenals, different types of abasic sites, such as C1′-aldehyde, C1′ keto (lactone), C4′ keto, etc. may be formed. These abasic moieties will probably all co-migrate differently from monomers that contain a base, and will thus be quantitated as the sum of all abasic sites (5′ dRiboseMP (abasic site) in Figure 16) formed in the "monomer window."

One final complication of the proposed assay, which may easily be overcome for most studies, is that certain DNA base modifications may impede the snake venom phosphodiesterase; however, other nucleases may also be used, separately or simultaneously, as long as the 5′ ^{32}P-phosphate is not removed.

In conclusion, the proposed assay should work well for studying the effect of sequence context on DNA lesion repair *in vitro* using purified DNA repair proteins. The system should also work well for monitoring the effect of sequence context on enzymatic methylation (e.g., of the *Eco*R I restriction site by the *Eco*R I methylase, or methylation by the Dam, Dcm, or CpG methylases). Although studies whereby the effect of sequence context on adduct formation or adduct degradation may be somewhat more challenging, the possible complications do not seem to be insurmountable.

4C. Proposed Assay for Determining the Influence of Sequence Context on Lesion Mutagenicity *in vivo* Stochastically by Mass Spectrometry

The work described in this dissertation allows one to determine the mutation frequencies of DNA lesions placed in any context in an accurate and timely manner. While one can now assay the mutation frequencies of m^6G in sixteen nearest-neighbor sequence contexts processed in four separate cell lines in under one month, can one additionally test next-to-nearest neighbor contexts in an even shorter period of time? That is, can one probe context effects in NNXNN sequences? Using the methodology presented in this thesis, 256 oligonucleotides containing m^6G would have to be individually made, incorporated into genomes, and 1,024 electroporations would have to be performed. Such an approach would be impractical. I describe below a strategy based on mass spectrometry that could allow the mutation frequencies of <u>any</u> DNA lesion whose mutational specificity is known, in 2 to 4 neighboring base sequence contexts (16 to 256 individual analyses per cell line as would be done in the work described above) to be analyzed using a single degenerate oligonucleotide incorporated into a genome, followed by a single electroporation for each cell line. Both single-stranded and double-stranded genome scenarios using m^6G will be presented with consideration of using the data also to obtain consensus sequences for lesion bypass, as well as consensus sequences that modulate the specificity of dNTP incorporation opposite the lesion.

The method of analysis for all cases listed above is outlined in Figure 19. Analysis relies

on excision of the stochastic[1] region as a 5-mer, which will be described later. The idea is that the relative intensities of, for example, 5′-CTAGG-3′ and 5′-CTGGG-3′ will give the mutation frequency of m^6G in the context 5′-CTm^6GGG-3′. A 50% mixture of mutant and wild-type deuterated standards containing all possible flanking sequences, the formulation of which will be described later, could be co-injected into the mass spectrometer with the non-deuterated sample. The oligonucleotide 5-mers representing all possible outcomes (32 for the NXN or 512 for the NNXNN scenarios) could be separated and identified using in-line capillary electrophoresis-nanospray mass spectrometry-tandem collision mass spectrometry (CE-MS-MS). All oligonucleotides, including permuted[2] oligonucleotides that have the same mass but represent different flanking sequences would be separated during the capillary electrophoresis stage. The first mass spectrometer will provide the quantitation and base composition for each oligonucleotide, but not the order of bases within the oligonucleotide. The identity of each oligonucleotide is then assigned using the fragmentation patterns given by collision MS in the second mass spectrometer.

---

[1]The term "stochastic" refers to the region containing the lesion site and its possible flanking bases.

[2]Each member in the set of "permuted" oligonucleotides contains the same number of each type of base, but in a different position.

The order in which the topics are presented here is the order in which they should be attempted (it makes little sense to develop the probably more difficult method of excising and purifying the stochastic 5-mers if MS cannot determine accurately the mutant to wild-type ratio of predetermined mixtures injected into the MS). To my knowledge, each topic listed below either has not been explored in its own right, or has not been used for the goal stated in this section and hence, few references are provided. The order of topics that will be discussed is as follows:

1. MS resolution and identification of oligonucleotides.
2. Creation of deuterated standards for quantitation (Figure 20).
3. MS quantitation of simulated MFs.
4. Lesion bypass studies (Figure 21).
5. Method for excision of stochastic 5-mers (Figure 22).
6. Construction of perfectly paired stochastic positions in duplex DNA (Figure 23).
7. Construction of 5′, 3′ bisphosphate from any commercially available phosphoramidite (Figure 24).
8. Construction of a double-stranded stochastic genome in which the signal from the base placed opposite the lesion is eliminated (Figure 25).

4C - 1. MS resolution and identification of oligonucleotides

The first step in this project must be to address the power of MS to resolved the 5-mer oligonucleotides. The following oligonucleotides could be made, with N being a degenerate base, to test the resolution strategy for either just the nearest-neighbor, 5′-CNANG-3′, 5′-CNGNG-3′, or both nearest- and next-to-nearest neighbor 5′-NNANN-3′, 5′-NNGNN-3′ scenarios. All oligonucleotides are modified at the 5′ end by an ethyl phosphodiester linkage for quantitation purposes that will be described later. Qualitatively, an equal amount of 5′-CNANG-3′ and 5′-CNGNG-3′ oligonucleotide will be mixed and it will be determined if CE-MS-MS can identify the 32 distinct species that would have resulted from an immediate nearest-neighbor experiment. The 32 species can be broken down as follows: number of species having a unique mass (6), number of sets containing two oligonucleotides having the same mass (6), number of sets containing three oligonucleotides having the same mass (2), number of sets containing four oligonucleotides having the same mass (2). It is a possibility that CE does not have to separate each and every oligonucleotide, just the few species within each set that have the same mass. If this is successful, the same qualitative experiment will be done using the 5′-NNANN-3′, 5′-NNGNN-3′ degenerate oligonucleotides to see if 512 species can be deconvoluted. Perhaps the greatest challenge for the 512 mixture scenario would occur when a G, A, T and C each fill the degenerate positions as shown in columns 1 and 2 (mutant has central A) or 4 and 5 (wild-type has central G) in Table 4. This yields the highest diversity of oligonucleotides of the same mass. The oligonucleotides in columns 3 or 6 also have,

respectively, the same mass as those in columns 1 and 2, or 4 and 5. There then exists 2 sets that contain 36 oligonucleotides of the same mass, which CE must separate.

If the separation cannot be done by CE alone, then an orthogonal separation technique, such as reversed-phase HPLC, could be performed prior to CE. If the technology to join an HPLC to CE for nanospray-MS cannot be developed, then all species (mutant and wild-type non-deuterated sample and mutant and wild-type deuterated standard) can be separate into 5 to 10 groups by conventional reversed-phase HPLC (or something else). Each group will be injected into the CE-MS-MS setup. Any sample loss for a particular species due to handling will not affect an accurate report of the mutation frequency, since the intensity of mutant and wild-type sample oligonucleotides are normalized to the known mixture of deuterated standards, which are carried through the entire process. The use of HPLC may, however, give poorer resolution due to band broadening.

4C - 2. Creation of deuterated standards for quantitation (Figure 20)

The creation of standards will be an essential element of this project. The response of the instrument will change over time and, since a mutant and wild-type oligonucleotide will be coming off the CE and entering the mass spectrometer at different times, an equimolar mixture of, for example, 5′-CTAGG-3′ and 5′-CTGGG-3′ will probably have different intensities. These intensities can be normalized to give the proper ratio and hence, the mutation frequency, if co-injected with a mixture of their deuterated counterparts, whose mutant deuterated to wild-type deuterated ratio is absolutely known. These standards must have the same chemical and physical (chromatographic) properties as the sample to be analyzed, but a different mass. The formulation of deuterated standards of known mutant to wild-type ratios poses a challenging problem in this combinatorial-type project and must be explained in detail.

One could make 32 individual 5-mer deuterated oligonucleotides for the 16 sequence context scenario, purify, quantitate by UV and mix the 16 mutant and wild-type oligonucleotides to formulate a standard for co-injection with the sample to be analyzed. While this is impractical for the 256 context scenario, it is doable for 16 contexts. A method for formulating deuterated mutation frequency standards will be presented that is quick and applicable to both scenarios.

Instead of incorporating deuterium into the bases, which can be expensive and lead to complication in interpreting the results if the amount of label depends on the base composition, a

postlabeling procedure will be used to label all species uniformly with 5 deuterium atoms quickly and cheaply (Figure 20). ^{13}C is the most abundant naturally occurring isotope, with a $^{13}C/^{12}C$ ratio of 1/100. A 5-mer will contain no more than 50 carbon atoms. Using the binomial formula and a 1% probability of ^{13}C at any carbon, one can expect that 60.5%, 30.6%, 7.6%, 1.2%, 0.1%, and 0.01% of a 5-mer would contain 0, 1, 2, 3, 4 or 5 ^{13}C atoms, respectively. Increasing the mass of a standard by 5 atomic mass units will therefore suffice so that a the mass spectrum of a deuterated 5-mer will not overlap with its non-deuterated 5-mer counterpart so that accurate quantitation may be achieved. Labeling will be performed using 5′-phosphorylated oligonucleotides (which are obtained after a restriction enzyme digest), the water-soluble carbodiimide EDC and ethanol (Gottikh et al., 1985). The ethanol is deuterated for standards or non-deuterated for samples as shown in Figure 20. The ethylphosphodiester tag is small enough so that the contributions of base composition and base position to resolution should not be diminished. If this is not the case, an increase of 3 atomic mass units by addition of methanol may remedy this; however, a higher background would result (e.g., the peak showing the mass to charge ratio of 60.5% of a particular deuterated standard will also consist of 1.2% of its non-deuterated sample counterpart).

At first glance, it may seem impossible to make standards of known MF composition using a degenerate approach. When an oligonucleotide is chemically made on a DNA synthesizer, an equimolar mixture of phosphoramidites will not yield 25% of each base in a degenerate position due to differences in coupling efficiencies. One can diminish bias by supplying a doped mixture of phosphoramidites during an addition to a degenerate position;

however, there may still be some bias. As long as the bias is equal, an equimolar mixture of 5′-$N_4N_3\underline{A}N_2N_1$-3′ and 5′-$N_4N_3\underline{G}N_2N_1$-3′ will generate exact mutation frequencies of 50% to act as standards for all 256 sequence contexts (Figure 20). By using the same doped phosphoramidite bottle and making the oligonucleotides at nearly the same time, the 3′ to 5′ chemical synthesis should yield the same fractional composition of N_2N_1 for two syntheses. A greater than 10-fold molar excess of phosphoramidite is used per coupling, so coupling of N_2 to either a G or A at the middle position can probably be driven to completion; one may even use a 100-fold molar excess of phosphoramidite to make the 5-mers to ensure couplings go to completion. If differential bias occurs, it would happen at N_3, which would propagate to N_4. This would occur if the primary hydroxyl of 5′-AN_2N_1-3′ reacts differently than that of 5′-GN_2N_1-3′ with the doped phosphoramidite mixture. There is no evidence that this will happen, but it is a formal possibility. The 5′-phosphorylation will be driven to completion on the DNA synthesizer using a 10- or 100-fold molar excess of Chemical Phosphorylation Reagent (Glen Research, 1998).

4C - 3. MS quantitation of simulated MFs

The method of creating standards using degenerate oligonucleotide pools of mutant and wild-type species will be validated by formulating exact 50% mutation frequencies for all 16 sequence contexts from the deuterated 16 mutant and wild-type oligonucleotides that were individually made, quantitated, mixed, and ethylated. Equal amounts (as judged by UV absorbance) of the ethylated, non-deuterated degenerate mutant and degenerate wild-type oligonucleotides "sample" will be mixed and added to the deuterated mixture of known composition. After MS, normalization factors for the deuterated oligonucleotides will be generated and used to quantitate the relative intensity of each non-deuterated mutant/wild-type pair from the "sample" mixture. If a 50% mutation frequency is found for all non-deuterated sample sequence contexts, then the assumption that a doped phosphoramidite mixture will give identical positional bias for two oligonucleotides made at approximately the same time, regardless of there being a centrally located A or G, is probably true, and deuterated standards for the 256 sequence context scenario can easily be made and used with confidence.

The next step must be to determine the accuracy of quantitating the relative amounts of the mutant and wild-type species. Oligonucleotide 5-mers of highest sequence diversity, but identical mass for a 256 nearest- and next-to-nearest experiment are listed in Table 4. The 24 oligonucleotides from columns 1 and 2 represent mutagenic outcomes from *in vivo* processing of m^6G (central A), while the 24 oligonucleotides from columns 4 and 5 represent wild-type (central

G). All oligonucleotides in Table 4 can be synthesized and chemically modified (5′-tagged with non-deuterated ethanol). These can be used to simplify the optimization of their separation during the early phase of the experiment. After separation has been achieved, the 24 mutant oligonucleotides are mixed with the 24 wild-type oligonucleotides in ratios that will simulate different mutation frequencies of the following percentages: 0, 0.25, 0.5, 1, 3, 5, 10, 25, 50, 75, 90, 95, 97, 99, 99.25, 99.75, 100. It can be determined if the output values from the optimized MS conditions will accurately reflect the input ratios for the seventeen trials shown above, after they are normalized using a mixture of co-injected deuterated standards of known composition (50% MF).

4C - 4. Lesion bypass studies (Figure 21)

An important corollary to the MS approach for finding mutation frequencies of lesions is that it may also be used to determine in what sequence contexts a particular DNA adduct is the most blocking to a DNA polymerase *in vivo*, as outlined in Figure 21. This can provide a clue as to how mutational hotspots arise since the contexts that provide the least hindrance to a DNA polymerase should have a selective growth advantage. This may be determined under the assumption that if the fraction of a certain sequence context within the entire mixture before electroporation is much greater than that obtained from the corresponding mutant + wild-type fractions after biological processing, then that context is a detriment to copying past m^6G by the DNA polymerase. Since the molar fraction of each specie must be known, a 256 context (next-to-nearest-neighbor) analysis in single- or double-stranded genomes is impractical. A nearest-neighbor analysis in double-stranded genomes should also be avoided due to possible bias in the enzymatic steps during genome construction, the method of which will be described later. The single-stranded nearest-neighbor experiment must be done in a cell line that lacks all known repair systems for m^6G, since those sequences in which m^6G is preferentially repaired may have a selective growth advantage.

The fractional composition of input contexts can be determined by individually making, purifying, and quantitating 16 m^6G oligonucleotides and then ligating an equimolar mixture of these into a genome in an unbiased manner. This can be done using the two scaffold method

described in this dissertation, if a constant base 5′ and 3′ to the stochastic region is additionally not hybridized to the scaffolds, to avoid possible stacking bias (as shown in Figure 21). The molar fraction of each of the 32 possible sample specie outcomes will be calculated using a co-injection of an exact equimolar mixture of 32 deuterated standards into the mass spectrometer. The normalization constant for each specie, *Cn*, will be determined from the deuterated equimolar mixture by the formula $\frac{Dn \times Cn}{\sum_{n=1}^{32} Dn} = \frac{1}{32}$, where *D* is the intensity of the deuterated oligonucleotide specie designated as *n*. The normalized fraction of each sample output specie is given by $\frac{Hn \times Cn}{\sum_{n=1}^{32} Hn \times Cn}$, where *H* is the intensity of the non-deuterated sample oligonucleotide specie designated as *n*, and *Cn* is the normalization factor for a particular specie obtained from the previous equation. If this fractional output composition for the sum of a particular mutant and wild-type specie is statistically smaller than 1/16, then m^6G in this particular sequence context may be a strong block to the DNA polymerase, thus creating the bias in growth.

4C - 5. Method for excision of stochastic 5-mers (Figure 22)

If mass spectrometry can measure the mutation frequency of all sequence contexts to within ± 5%, the biological system should be developed. Many DNA repair enzymes can repair DNA damage only if the damage is a double stranded form. For these types of lesions, a single-stranded vector may not be a detriment, since repair would not occur until a double-stranded form occurs during DNA replication past the lesion. The overall strategy for analyzing 16 or 256 sequence contexts simultaneously in single-stranded genomes is presented. Briefly, an oligonucleotide 34 nucleotides long containing a centrally located m^6G lesion flanked by degenerate bases of the sequence 5′ CNXNG 3′ or 5′ CNNXNNG 3′ will be made and ligated into a linearized single-stranded vector as described previously in this dissertation. The scaffolds will not overlap with this 5 or 7 base stretch to avoid ligation bias due to scaffold stacking for the lesion bypass experiments. Following a single electroporation of this stochastic genome, the single-stranded viral DNA from the population of progeny will be purified. As shown in Figure 22, a 50 bp region containing the stochastic region will be amplified using high fidelity preparative scale PCR and biotinylated primers. After biological processing of the 16 context scenario, the region that contained m^6G will have the sequence 5′ CN(G/A)NG 3′; while the 256 context scenario will have 5′ NN(G/A)NN 3′. The small 5-mer oligonucleotide will be released from the PCR product using two type IIs restriction enzymes. One will cut at the 5′ end and the other at the 3′ end of the desired oligonucleotide, to provide the 5-mer, with the hydrogen bonded complementary strand being an 11-mer. Passage of the digest through a denaturing streptavidin

column will eliminate approximately 50% of undesired bulk oligonucleotide. The desired 5-mer is then purified from the 11, 20 and 23-mers on a denaturing gel. Consensus sequences obtained after MS analysis can be individually tested using another type IIs restriction enzyme that will cleave the phosphodiester bond immediately 5′ to where the lesion occurred in harvested RF DNA. The type IIs restriction enzyme and radiolabeling TLC technique previously described in this thesis can then be employed.

4C - 6. Construction of perfectly paired stochastic positions in duplex DNA (Figure 23)

Construction of a covalently closed circular, double-stranded virus or plasmid genome containing a DNA adduct assures one that the DNA repair enzymes inside the cell will act on the lesion in its most natural state. New techniques additionally have to be invented for the construction of double-stranded stochastic genomes. One cannot simply anneal two complementary oligonucleotides that contain a central stochastic region and expect perfect pairing, since thermodynamic discrimination would be too small for the length of duplex needed for site-specific experiments. Moreover, if a DNA polymerase is used to create perfect pairing of the stochastic region surrounding a DNA lesion (Hatahet et al., 1998), control over what base is placed opposite the lesion is lost. Using this method, one is forced to study the type of lesion:base pair that the DNA polymerase prefers to create; the m^6G:C base pair could not be created. Even if the predominant insertion event is to be studied, such as the *in vivo* processing of a pure m^6G:T base pair for each sequence context, the there is a strong possibility that the percent of m^6G:T formation for different sequence context substrates may depend on nearest-neighbor bases flanking m^6G (Singer et al., 1989). This *in vitro* manipulation would introduce bias into the experiment before the genome containing m^6G is biologically processed.

The strategy shown in Figure 23 will create perfect pairing of each degenerate position, with a centrally located lesion paired opposite any desired base. A self-priming hairpin containing half of the degenerate region at the 5′ terminus will be constructed using

5′-phosphorylated oligonucleotides. The numbers in parenthesis are the approximate nucleotide length and the circle containing "B" represents a biotin incorporated into a four thymine loop *via* the biotin-dT-phosphoramidite. The ligated "half-product" is about 100 nucleotides long, and is extended by two nucleotides using DNA polymerase and dNTPs. Taq polymerase followed by Klenow fragment will be used to copy the degenerate position to avoid the extension of blunt end product by one nucleotide (Masse et al., 1998). Since gel purification cannot be used to purify the large desired extended polymerization product, the reaction will be optimized to give 100% extended product. Full extension can be monitored by incorporation of ^{32}P into the 5′ hydroxyl of the 3′ terminal oligonucleotide during hairpin construction. After polymerization, the substrate is digested with a restriction enzyme to liberate short ^{32}P-labeled oligonucleotides, approximately 20, 21, 22 or 23 nt in length symbolizing extension by 0, 1, 2 or 3 nucleotides, that are separate by polyacrylamide gel electrophoresis (PAGE) and quantitated using radiography. After full extension (a 22-mer in the above diagnostic) is verified, a 5′-pXp-3′ bisphosphate of the lesion, the construction of which will be described shortly, will be ligated to the 3′ hydroxyl of the stochastic region using T4 RNA ligase. Substrate that does not incorporate the bisphosphate will be extended by multiple additions of dATP using terminal transferase to allow for product separation by PAGE. The entire process is performed in parallel using a different set of oligonucleotides and the bisphosphate of the base to be paired opposite the lesion. The strategy of having a 5′ and 3′ phosphate on one hairpin, and no terminal phosphates on the other will ensure ligation of two different hairpins. After the two halves are joined, the ligation product is gel purified. The remaining nick contains a 3′ phosphate and a 5′ hydroxyl. Enzymatic ligation would require phosphorylation in the opposite sense. Although the

3′ phosphate can be removed by alkaline phosphatase, T4 polynucleotide kinase cannot phosphorylate the 5′ hydroxyl in the nick. Template-directed chemical ligation has been used successfully to join intramolecularly a 3′ phosphate with a 5′ hydroxyl group using cyanogen bromide (BrCN), imidazole (Im) and nickel ion (Rubin et al., 1995), and will therefore be utilized. After ligation, denaturing PAGE is able to separate closed from nicked species (unpublished data from the Essigmann Lab). The hairpin ends are clipped off using a type IIs restriction enzyme that will leave 3′ four base overhangs of sequence 5′ GGGG 3′ and 5′ GCCG 3′. At this point, the oligonucleotides are long enough to prevent strand exchange, as long as the oligonucleotide solution is not heated. Passage of the digested solution through a streptavidin column will eliminate the biotinylated oligonucleotides that would compete during annealing with a genome for the lesion containing duplex. Ligation of an equimolar mixture of prepared M13 genome and lesion containing insert having the extensions 5′ GGGG 3′ and 5′ GCCG 3′ will ensure that the desired linkage of insert to vector is formed. The GC rich overhangs will aid in the ligation efficiency and covalently closed species can be purified by agarose gel electrophoresis containing an intercalation molecule, such as ethidium bromide, in the running buffer. After isolation of the covalently closed circular supercoiled genome, the dye is removed, and the duplex can be used for electroporation.

4C - 7. Construction of 5′, 3′ bisphosphate from any commercially available phosphoramidite (Figure 24)

The strategy presented for making bisphosphate monomers from any phosphoramidite is outlined in Figure 24. This method will also allow researchers that are not equipped to do organic synthesis to make the nucleoside, 5′ monophosphate, or 3′ monophosphate from a commercially available phosphoramidite. To make a nucleoside, one would use a "Universal Support" (Glen Research, 1998) to couple the phosphoramidite, followed by detritylation and deprotection (the 3′ phosphate that came from the phosphoramidite is removed during deprotection). To make a 5′ monophosphate, one would use the "Universal Support", couple the phosphoramidite, then couple the "Chemical Phosphorylation Reagent" (Glen Research, 1998). A 3′ monophosphate would be made using a "3′-Phosphate CPG" (Glen Research, 1998) to couple the phosphoramidite.

4C - 8. Construction of a double-stranded stochastic genome in which the signal from the base placed opposite the lesion is eliminated (Figure 25)

A problem with using double-stranded vectors is that a signal will come form the base placed opposite the lesion to be studied, regardless of how the lesion is processed. This background, which is 50% if both strands are replicated with equal efficiency, will be eliminated using the following construction strategy depicted in Figure 25. The key to this strategy is that the M13 gene III minor phage coat product is essential for infection of *E. coli* harboring an F′ episome. The vector is constructed so that the (−) strand contains the lesion and DNA coding for functional gene III protein (Figure 25 "go"), while the (+) strand contains the base opposing the lesion and a DNA sequence that will create a stop codon in gene III mRNA (Figure 25 "stop"). All viral M13 mRNAs are transcribed from the (−) strand (Denhardt et al., 1978). Once transfected into hosts for replication, populations of phage from both the plus and minus strands in the parental cell replicate, are packaged as (+) strands, and infect normally. However, previously uninfected cells that subsequently become infected by progeny phage harboring the stop codon (from the non-adduct containing strand) will not produce additional infectious phage particles. The net result is that only the progeny from the adducted (−) strand are amplified. This strategy may be an improvement over methods whereby background signal from the strand opposite the lesion is eliminated by treating the (+) strand with UV light or allowing the (+) strand to grow in the presence of uracil prior to harvesting (followed by abasic site formation *in vivo*, which would block DNA replication). Those methods are not always reliable and

furthermore, one would expect a Poisson distribution of damage events with the UV photodimer or abasic site formation; the net result being that progeny would be biased in favor of genomes containing the least amount of damage. This will not occur using the proposed site-specific approach as outlined in Figure 25.

A concern exists if the initial progeny from the (−) and (+) strands co-infect cells, thus propagating the undesired signal due to the expression of functional gene III. This can be overcome, if necessary, by either harvesting the single-stranded DNA from the viral progeny and performing another electroporation or by infecting F′ cells with a low phage progeny to cell ratio. Before the (+) strand source containing a stop codon in gene III can be used as a reagent (Figure 25A), gene III DNA is manipulated *in vitro* to contain a premature stop codon. The manipulated genome must then be electroporated and propagated in *E. coli* that contains a plasmid expressing functional gene III, but lacking the M13 origin of replication. Ultimately, the genome prepared as outlined in Figure 25 (from parts A, B, and C) will contain a C:T mismatch that is greater than 2 kb from the lesion and should be a poor substrate for cellular mismatch repair, although mismatch repair-deficient cell lines can be used if necessary. PCR can be performed on the isolated progeny and the mutation frequency of all possible sequences can be determined as outlined in Figure 19.

Finally, all the techniques described and developed in this thesis should be applicable to the analysis of the effects of sequence context on m^6G mutagenesis in mammalian cells. Transfection of Chinese Hamster Ovary cells that either have or lack the m^6G-DNA-

methyltransferase with vectors that contain both mammalian and M13 origins of replication will allow biological processing of the lesion in mammalian cells, and subsequent analysis in *E. coli*.

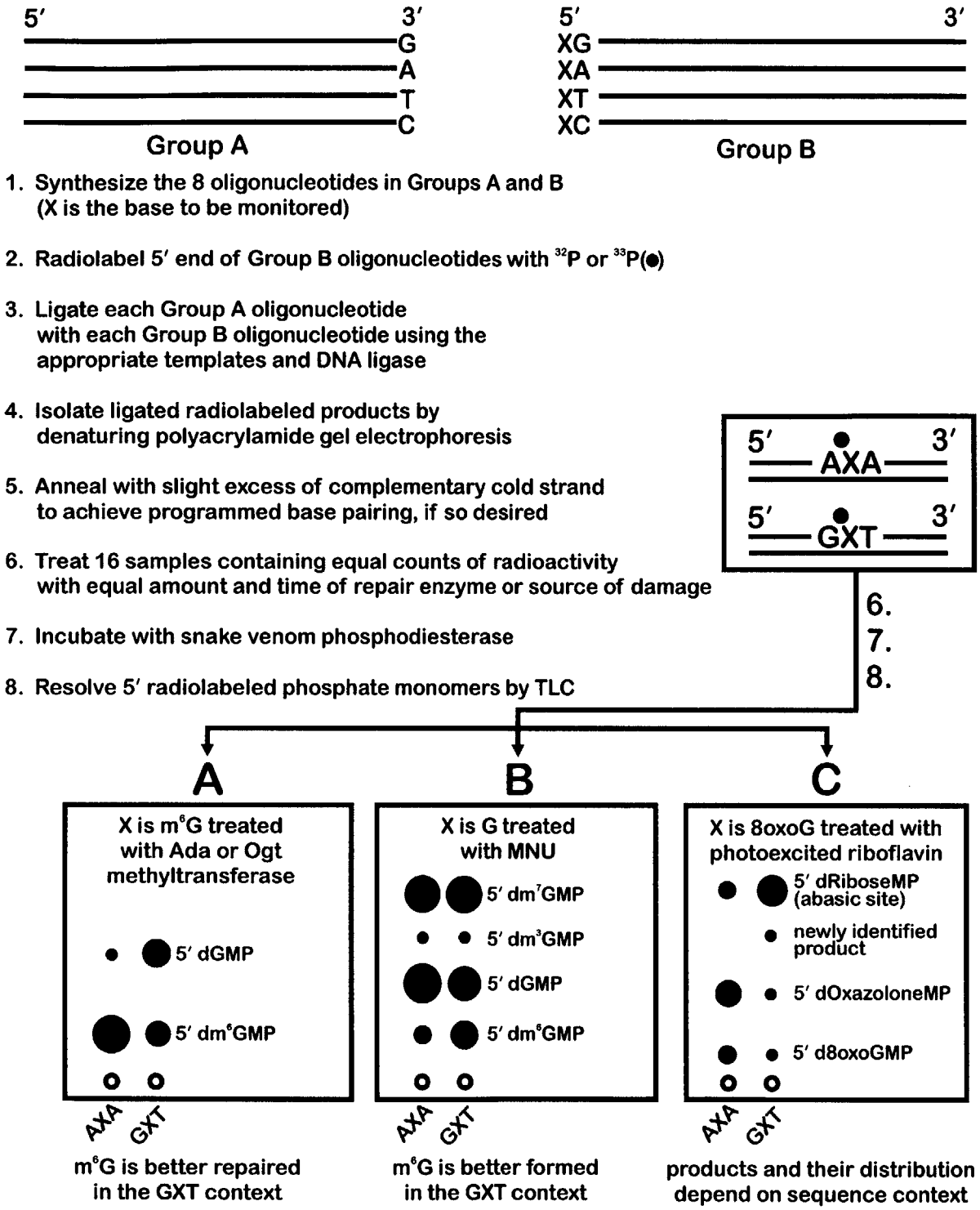


Figure 16. Proposed assay for studying the effects of sequence context on lesion repair, adduct formation, or adduct degradation *in vitro*. Pre-labeling the DNA substrate internally creates a site-specifically radiolabeled "monomer window" that will report changes in a DNA adduct or base within a defined sequence context, thus eliminating signal from identical products generated outside the window. Order and intensity of radiolabeled monophosphates depicted was arbitrary chosen.

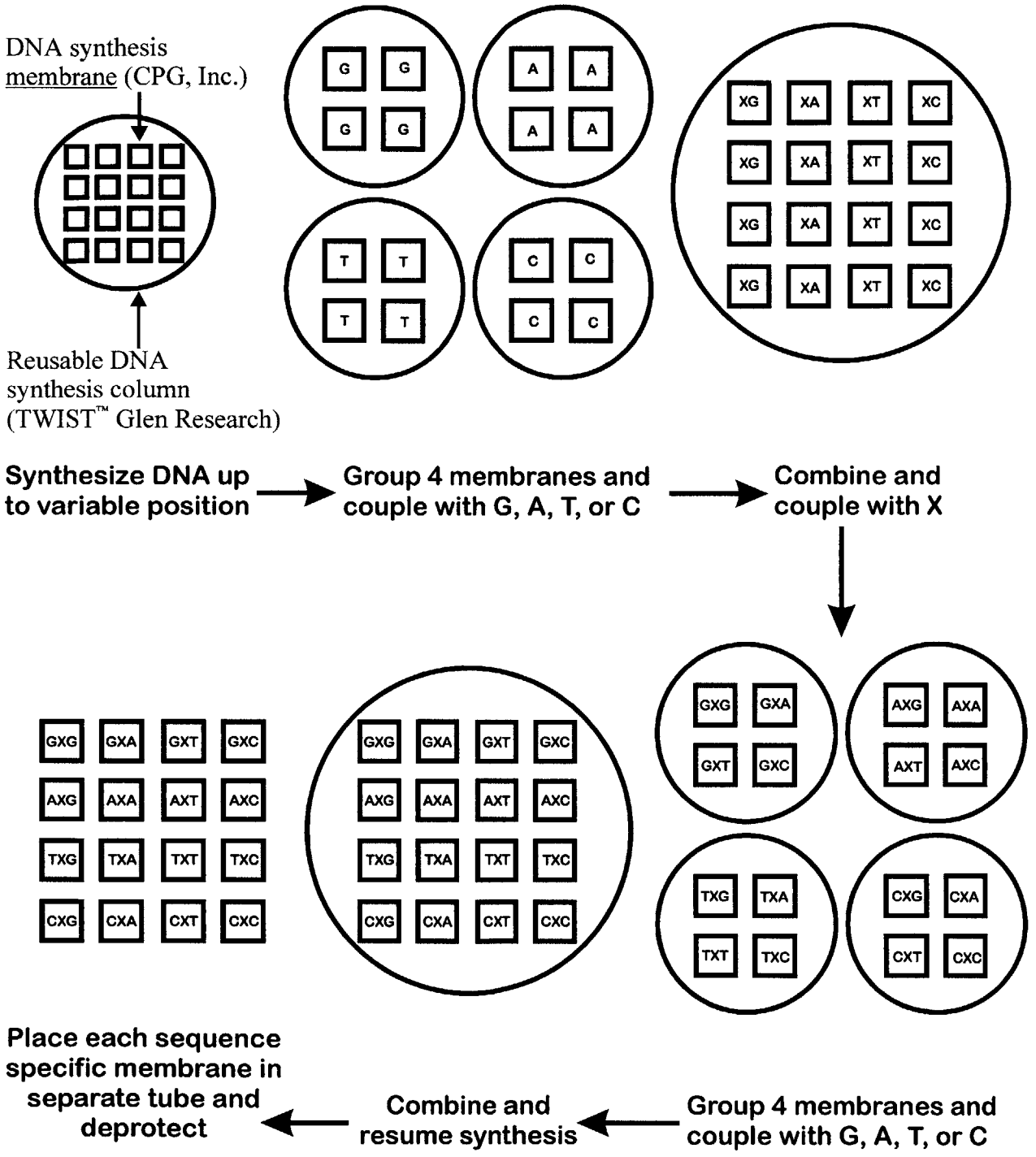


Figure 17. Proposed scheme for synthesizing sixteen individual nearest-neighbor oligonucleotides. This method can be adapted to synthesize efficiently all oligonucleotides needed for the experiments presented in Figure 16. X can be a modified or normal base.

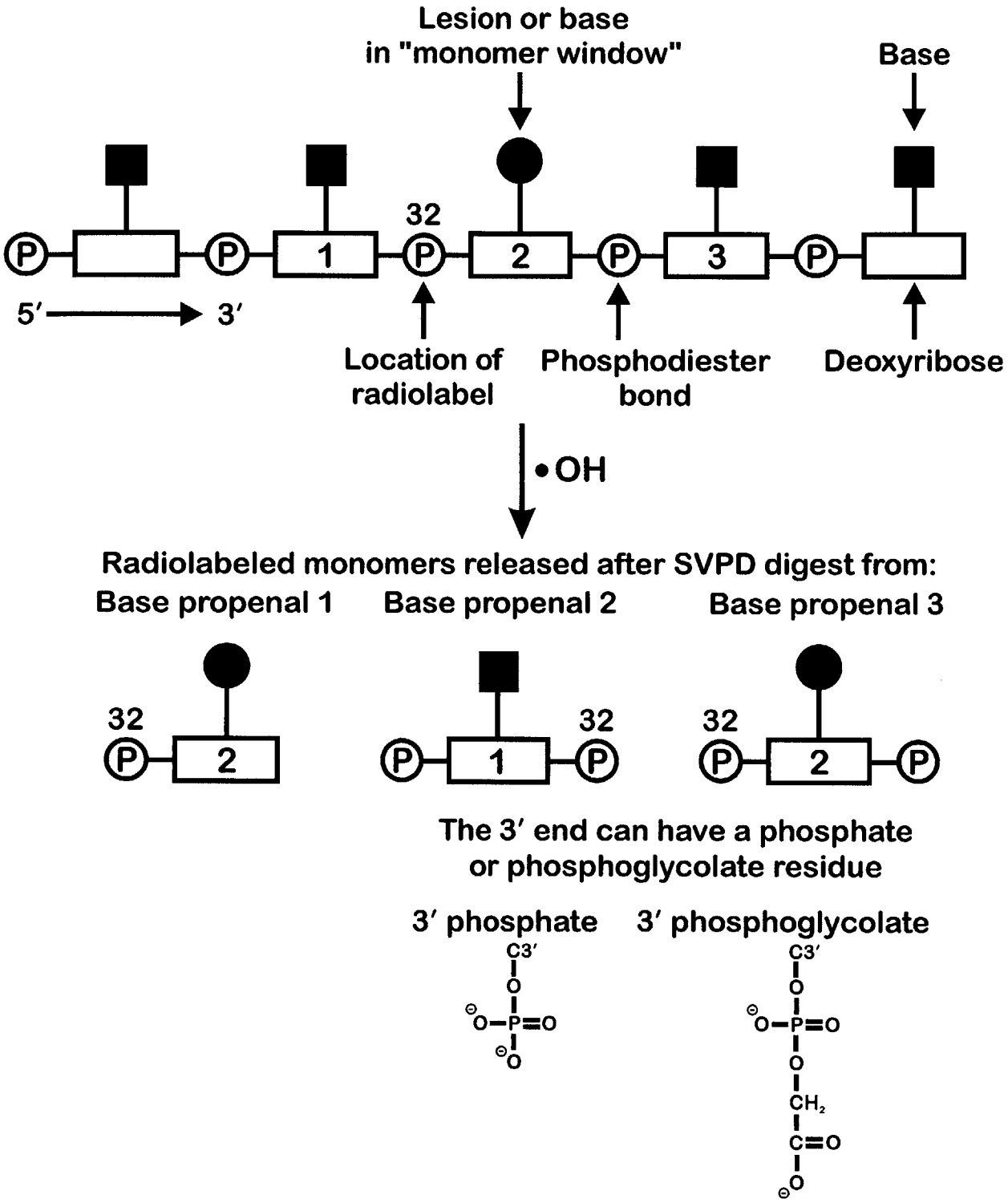


Figure 18. Proposed effect of DNA strand breaks created by free radicals on analysis of adduct formation or adduct degradation by the assay outlined in Figure 16.

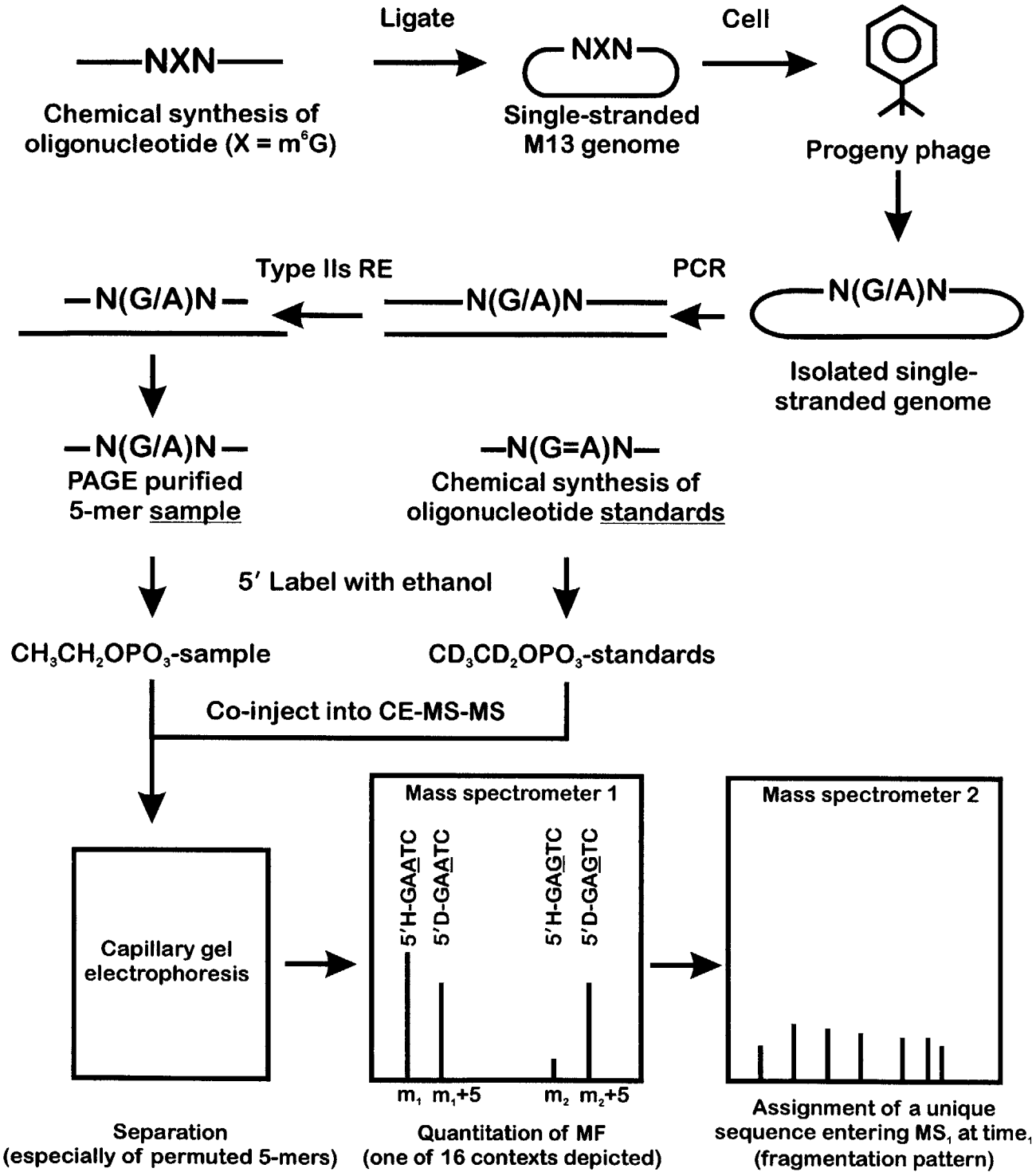


Figure 19. Proposed overall scheme for determining the mutation frequencies of a DNA lesion in all sequence contexts simultaneously by mass spectrometry as demonstrated using an m^6G stochastic single-stranded genome.

Table 4: Highest number of oligonucleotide 5-mers (permuted) of equal mass.

Column 1	Column 2	Column 3	Column 4	Column 5	Column 6
GAATC	CAATG	AAGTC	GAGTC	CAGTG	GGATC
GAACT	TAACG	AAGCT	GAGCT	TAGCG	GGACT
GTAAC	CTAAG	ATGAC	GTGAC	CTGAG	GTAGC
GCAAT	TCAAG	ATGCA	GCGAT	TCGAG	GTACG
GTACA	ATACG	ACGTA	GTGCA	ATGCG	GCAGT
GCATA	ACATG	ACGAT	GCGTA	ACGTG	GCATG
AGATC	ATAGC	TAGAC	AGGTC	ATGGC	TGAGC
AGACT	ACAGT	TAGCA	AGGCT	ACGGT	TGACG
TGAAC	TAAGC	TCGAA	TGGAC	TAGGC	TCAGG
TGACA	TCAGA	CAGAT	TGGCA	TCGGA	CGAGT
CGAAT	CAAGT	CAGTA	CGGAT	CAGGT	CGATG
CGATA	CTAGA	CTGAA	CGGTA	CTGGA	CTAGG

Each oligonucleotide in columns 1, 2, and 3 or columns 4, 5, and 6 has an identical mass; hence, a separation step is necessary prior to MS analysis of the outcome for a 256 nearest- and next-to-nearest-neighbor sequence context experiment. The register of columns 1 and 2 (mutant) corresponds respectively with that of columns 4 and 5 (wild-type).

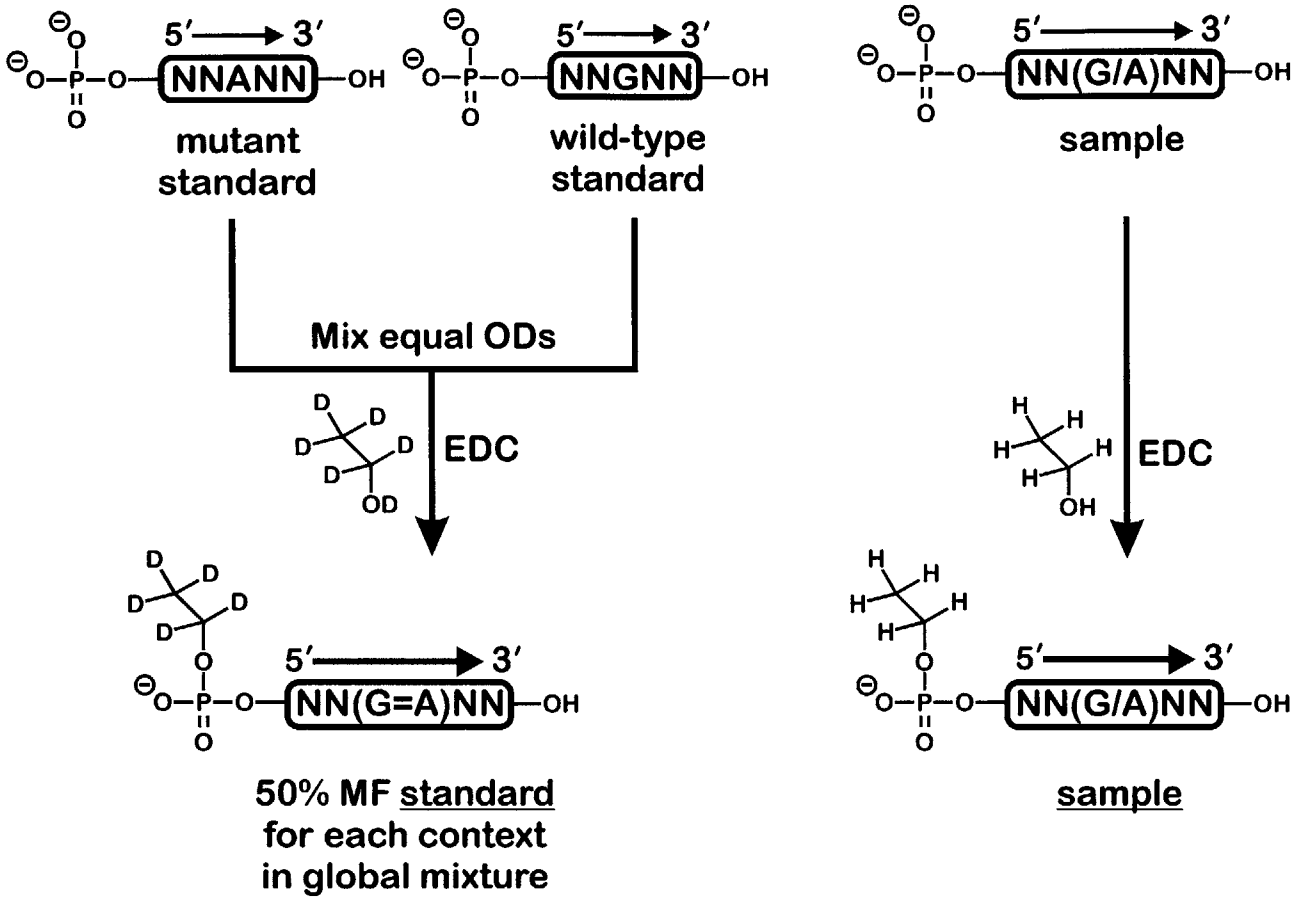


Figure 20. Proposed scheme for creating mass spectrometry standards. Chemically synthesized 5-mers of stochastic mutant (central A) and stochastic wild-type (central G) are mixed equally, and derivatized at the 5′ phosphate by chemical condensation with deuterated ethanol *via* EDC coupling, thus generating a 50% MF for each pair of mutant and wild-type oligonucleotide within the possible set of 256. After biological processing of a set of stochastic genomes (that contained m^6G), the stochastic 5′ phosphorylated 5-mer region is excised from the sample (as outlined in Figure 22), derivatized with non-deuterated ethanol, and mixed with 50% global MF standard. The A/(G + A) ratio for each individual sequence context within the stochastic set used in the biological experiment can be obtained accurately by comparison with the appropriate 50% global MF standard pair.

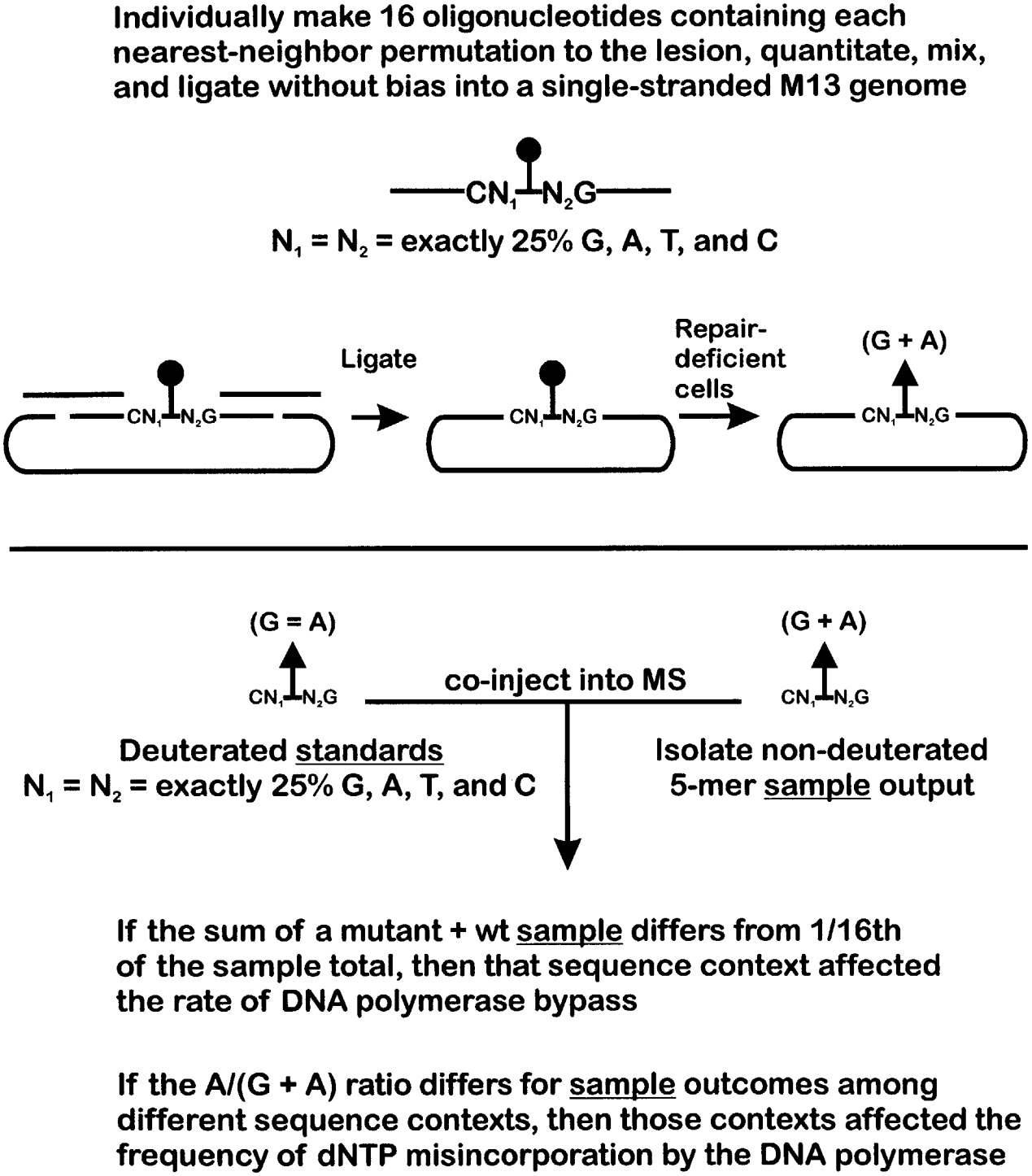


Figure 21. Proposed assay for determining the effect of sequence context on lesion bypass and polymerase miscoding as determined by mass spectrometry.

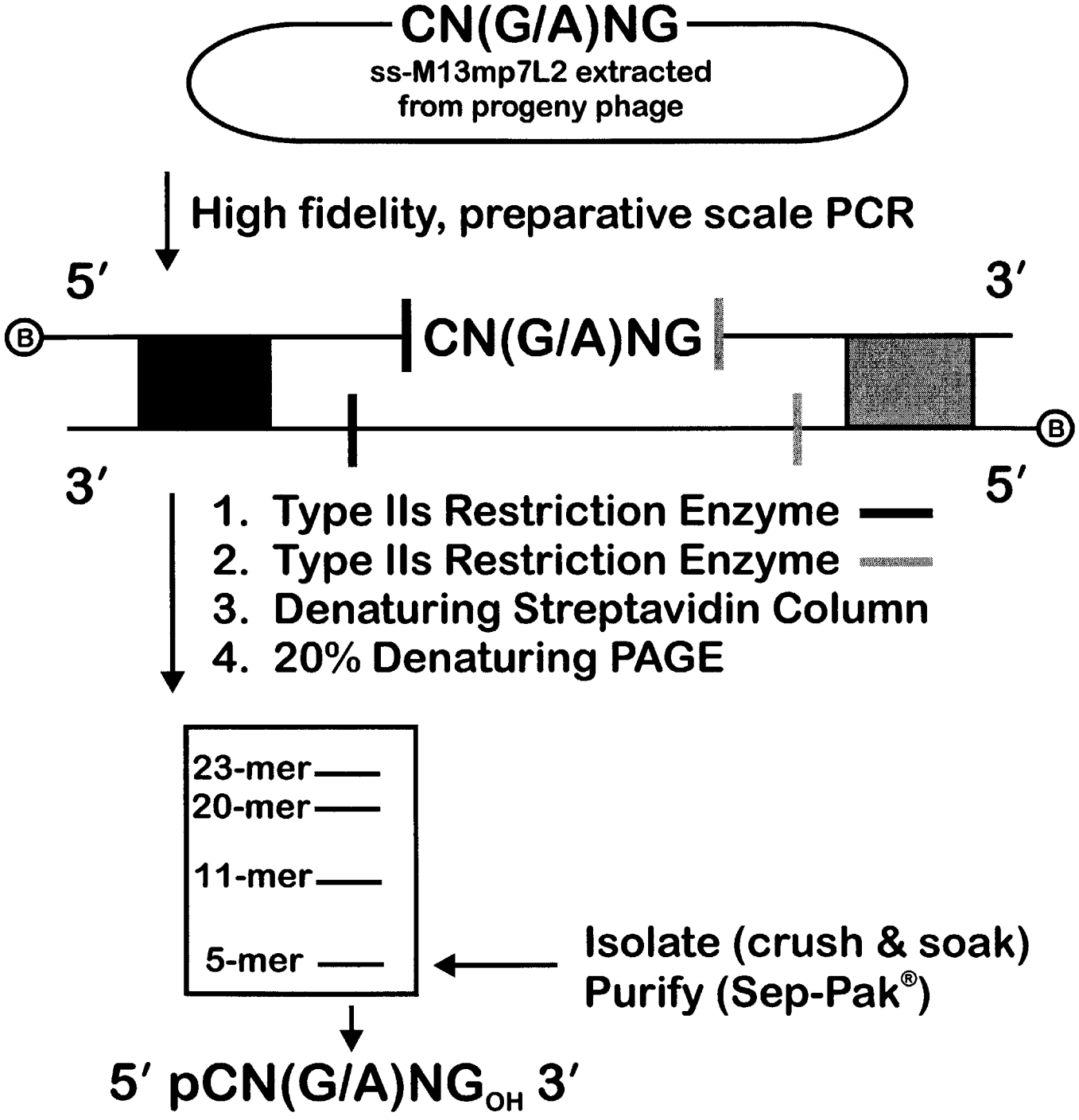


Figure 22. Proposed scheme for liberating all possible 5-mers from a stochastic genome after cellular processing of a DNA lesion. The 5-mers will be used for the mutation frequency analysis by mass spectrometry as outlined in Figures 19 and 20.

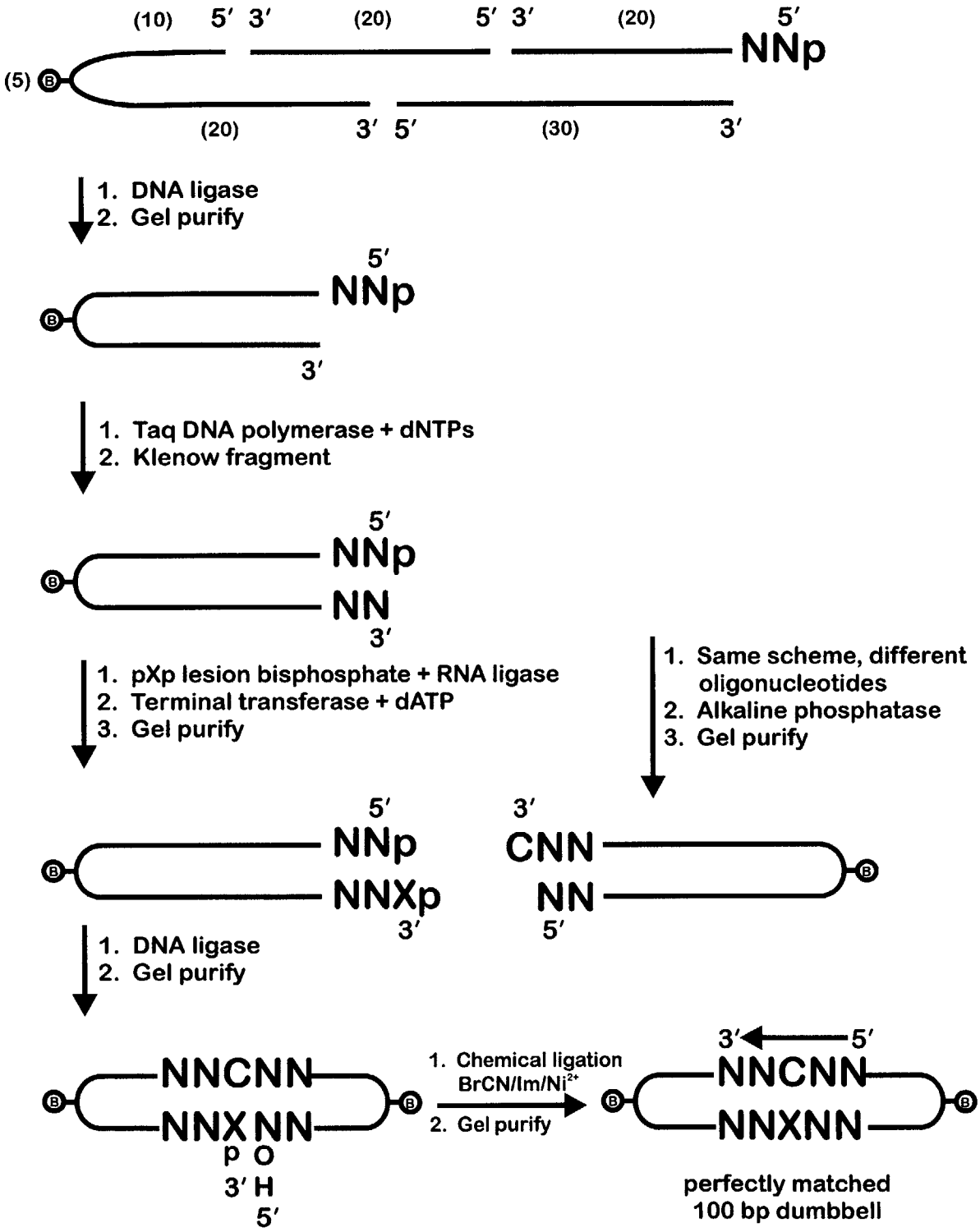


Figure 23. Proposed scheme for constructing a perfectly matched 100 bp duplex containing a centrally located programmed lesion:base pair within a stochastic sequence context. Numbers in parentheses represent the approximate number of bases of each pre-assembled oligonucleotide. Circles containing an inscribed "B" symbolize biotin conjugated to the oligonucleotide during chemical synthesis. This duplex will be used in the construction of stochastic duplex genomes as outlined in Figure 25.

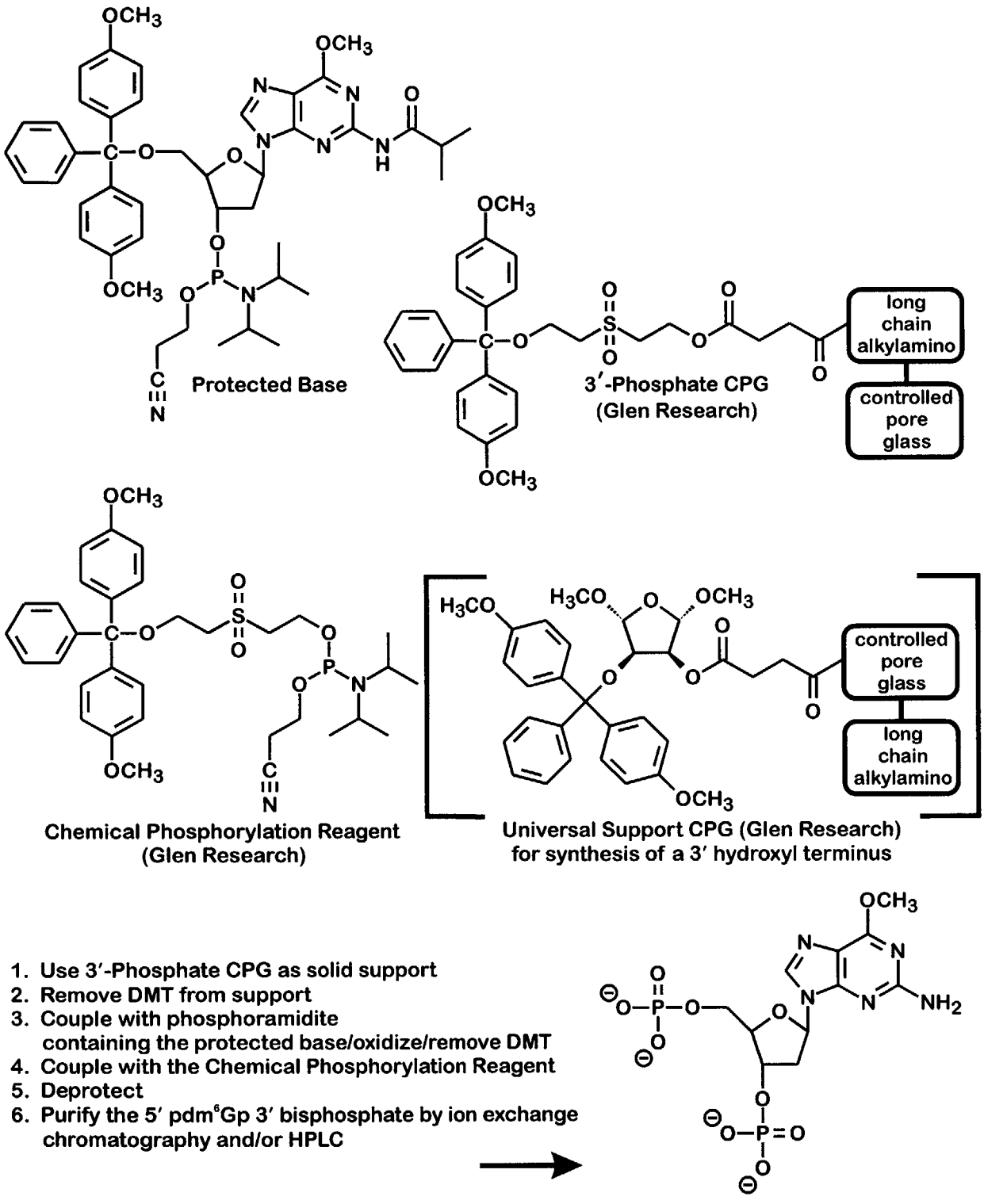


Figure 24. Proposed scheme for solid-phase synthesis of 3′, 5′ bisphosphate from the commercially available materials shown above. The bisphosphate will be used for constructing a perfectly matched 100 bp duplex containing a centrally located programmed lesion:base pair within a stochastic sequence context as outlined in Figure 23.

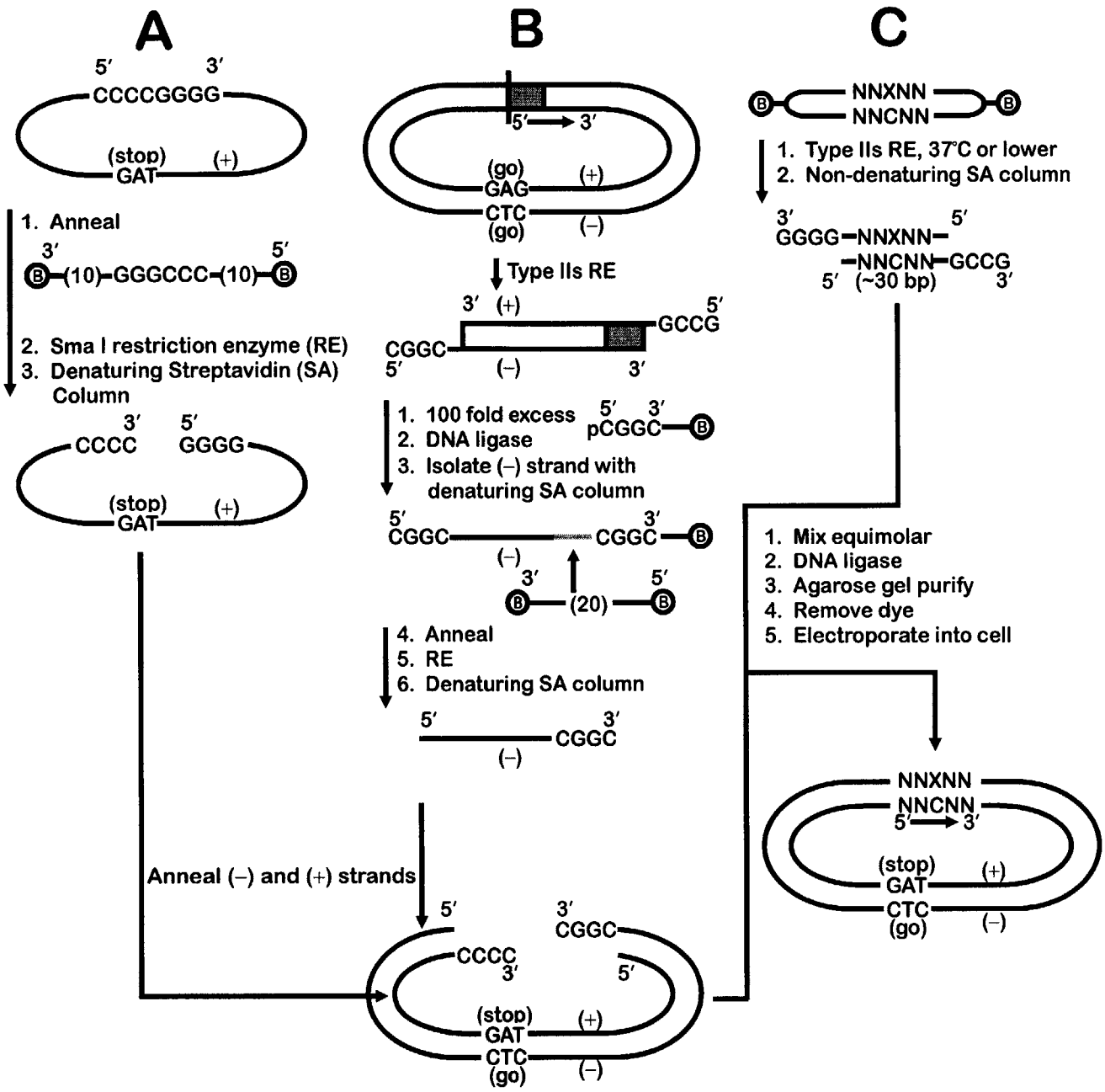


Figure 25. Proposed scheme for constructing a double-stranded genome containing a programmed lesion:base pair within a stochastic sequence context using substrates shown in A, B, and C. Amplification should occur from only the lesion-bearing (−) strand. A. The viral (+) strand contains a stop codon. B. The viral (−) strand contains the coding sequence for the full-length infectivity protein, gene III. C. The dumbbell containing the lesion within a stochastic duplex is prepared as outlined in Figure 23. Numbers in parentheses represent an approximate number of bases, while circles containing an inscribed "B" symbolize biotin conjugated to the oligonucleotide during chemical synthesis. Use of non-cohesive GC rich overhanging termini during construction is expected to yield efficient, unidirectional ligation.

REFERENCES

Abbott, P. J. and Saffhill, R. (1979). DNA synthesis with methylated poly(dC-dG) templates. Evidence for a competitive nature to miscoding by O^6-methylguanine. Biochim. Biophys. Acta *562*, 51-61.

Akagi, T., Hiromatsu, K., Iyehara-Ogawa, H., Kimura, H. and Kato, T. (1993). Specificity of mutations induced by *N*-methyl-*N*-nitrosourea in a cDNA of the *hprt* gene. Carcinogenesis *14*, 725-729.

Akasaka, S. and Yamamoto, K. (1994). Hydrogen peroxide induces G:C to T:A and G:C to C:G transversions in the *supF* gene of *Escherichia coli*. Mol. Gen. Genet. *243*, 500-505.

Andersson, B., Fält, S. and Lambert, B. (1992). Strand specificity for mutations induced by (+)-anti BPDE in the hprt gene in human T-lymphocytes. Mutat. Res. *269*, 129-140.

Ariza, R. R., Roldán-Arjona, T., Hera, C. and Pueyo, C. (1993). A method for selection of forward mutations in *supF* gene carried by shuttle-vector plasmids. Carcinogenesis *14*, 303-305.

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K. (1993). Current protocols in molecular biology (New York, NY: John Wiley & Sons, Inc.).

Backendorf, C., Spaink, H., Barbeiro, A. P. and van de Putte, P. (1986). Structure of the *uvrB* gene of *Escherichia coli*. Homology with other DNA repair enzymes and characterization of the uvrB5 mutation. Nucleic Acids Res. *14*, 2877-2890.

Ballering, L. A., Vogel, E. W., Vrieling, H. and Nivard, M. J. (1998). Strand-specific mutation induction by 1,2-dibromomethane at the hypoxanthine-guanine phosphoribosyltransferase locus of Chinese hamster ovary cells. Mutagenesis *13*, 61-65.

Banerjee, S. K., Borden, A., Christensen, R. B., LeClerc, J. E. and Lawrence, C. W. (1990). SOS-dependent replication past a single *trans-syn* T-T cyclobutane dimer gives a different mutation spectrum and increased error rate compared with replication past this lesion in uninduced cells. J. Bacteriol. *172*, 2105-2112.

Banerjee, S. K., Christensen, R. B., Lawrence, C. W. and LeClerc, J. E. (1988). Frequency and spectrum of mutations produced by a single cis-syn thymine-thymine cyclobutane dimer in a single-stranded vector. Proc. Natl. Acad. Sci. USA *85*, 8141-8145.

Barrows, L. R. and Magee, P. N. (1982). Nonenzymatic methylation of DNA by S-adenosylmethionine *in vitro*. Carcinogenesis *3*, 349-351.

Basic-Zaninovic, T., Palombo, F., Bignami, M. and Dogliotti, E. (1992). Fidelity of replication of the leading and the lagging DNA strands opposite N-methyl-N-nitrosourea-induced DNA damage in human cells. Nucleic Acids Res. *20*, 6543-6548.

Basu, A. K. and Essigmann, J. M. (1988). Site-specifically modified oligodeoxynucleotides as probes for the structural and biological effects of DNA-damaging agents. Chem. Res. Toxicol. *1*, 1-18.

Basu, A. K., Wood, M. L., Niedernhofer, L. J., Ramos, L. A. and Essigmann, J. M. (1993). Mutagenic and genotoxic effects of three vinyl chloride-induced DNA lesions: 1,N^6-ethenoadenine, 3,N^4-ethenocytosine, and 4-amino-5-(imidazol-2-yl)imidazole. Biochemistry *32*, 12793-12801.

Belguise-Valladier, P. and Fuchs, R. P. (1995). *N*-2-aminofluorene and *N*-2 acetylaminofluorene adducts: the local sequence context of an adduct and its chemical structure determine its replication properties. J. Mol. Biol. *249*, 903-913.

Benasutti, M., Ejadi, S., Whitlow, M. D. and Loechler, E. L. (1988). Mapping the binding site of aflatoxin B_1 in DNA: systematic analysis of the reactivity of aflatoxin B_1 with guanines in different DNA sequences. Biochemistry *27*, 472-481.

Bender, K., Federwisch, M., Loggen, U., Nehls, P. and Rajewsky, M. F. (1996). Binding and repair of O^6-ethylguanine in double-stranded oligodeoxynucleotides by recombinant human O^6-alkylguanine-DNA alkyltransferase do not exhibit significant dependence on sequence context. Nucleic Acids Res. *24*, 2087-2094.

Bentivegna, S. S. and Bresnick, E. (1994). Inhibition of human O^6-methylguanine-DNA methyltransferase by 5-methylcytosine. Cancer Res. *54*, 327-329.

Benzer, S. (1961). On the topography of the genetic fine structure. Proc. Natl. Acad. Sci. USA *47*, 403-415.

Bernelot-Moens, C., Glickman, B. W. and Gordon, A. J. (1990). Induction of specific frameshift and base substitution events by benzo[*a*]pyrene diol epoxide in excision-repair-deficient *Escherichia coli*. Carcinogenesis *11*, 781-785.

Bigger, C. A., Strandberg, J., Yagi, H., Jerina, D. M. and Dipple, A. (1989). Mutagenic specificity of a potent carcinogen, benzo[*c*]phenanthrene (4*R*,3*S*)-dihydrodiol (2*S*,1*R*)-epoxide, which reacts with adenine and guanine in DNA. Proc. Natl. Acad. Sci. USA *86*, 2291-2295.

Bill, C. A., Duran, W. A., Miselis, N. R. and Nickoloff, J. A. (1998). Efficient repair of all types of single-base mismatches in recombination intermediates in Chinese hamster ovary cells. Competition between long-patch and G-T glycosylase-mediated repair of G-T mismatches. Genetics *149*, 1935-1943.

Bishop, R. E., Dunn, L. L., Pauly, G. T., Dolan, M. E. and Moschel, R. C. (1993). The role of O^6-alkylguanine-DNA alkyltransferase in protecting Rat4 cells against the mutagenic effects of O^6-substituted guanine residues incorporated in codon 12 of the H-*ras* gene. Carcinogenesis *14*, 593-598.

Bishop, R. E. and Moschel, R. C. (1991). Positional effects on the structure and stability of abbreviated H-*ras* DNA sequences containing O^6-methylguanine residues at codon 12. Chem. Res. Toxicol. *4*, 647-654.

Bishop, R. E., Pauly, G. T. and Moschel, R. C. (1996). O^6-Ethylguanine and O^6-benzylguanine incorporated site-specifically in codon 12 of the rat H-*ras* gene induce semi-targeted as well as targeted mutations in Rat4 cells. Carcinogenesis *17*, 849-856.

Bjørås, M., Luna, L., Johnsen, B., Hoff, E., Haug, T., Rognes, T. and Seeberg, E. (1997). Opposite base-dependent reactions of a human base excision repair enzyme on DNA containing 7,8-dihydro-8-oxoguanine and abasic sites. EMBO J. *16*, 6314-6322.

Bohr, V. A. (1991). Gene specific damage and repair after treatment of cells with UV and chemotherapeutical agents. Adv. Exp. Med. Biol. *283*, 225-233.

Boiteux, S., de Oliveira, R. C. and Laval, J. (1985). The *Escherichia coli* O^6-methylguanine-DNA methyltransferase does not repair promutagenic O^6-methylguanine residues when present in Z-DNA. J. Biol. Chem. *260*, 8711-8715.

Brennand, J., Saffhill, R. and Fox, M. (1982). The effects of methylated thymidines upon cultures of V79 cells and the mechanism of incorporation of O^4-methylthymidine into their DNA. Carcinogenesis *3*, 219-222.

Briscoe, W. T. and Cotter, L. E. (1985). DNA sequence has an effect on the extent and kinds of alkylation of DNA by a potent carcinogen. Chem. Biol. Interact. *56*, 321-331.

Brown, M. D., Povinelli, C. M. and Hall, D. H. (1993). Distribution and characterization of mutations induced by nitrous acid or hydroxylamine in the intron-containing thymidylate synthase gene of bacteriophage T4. Biochem. Genet. *31*, 507-520.

Brown, T. C. and Jiricny, J. (1989). Repair of base-base mismatches in simian and human cells. Genome *31*, 578-583.

Buckley, N. (1987a). A regioselective mechanism for mutagenesis and oncogenesis caused by alkylnitrosourea sequence-specific DNA alkylation. J. Am. Chem. Soc. *109*, 7918-7920.

Buckley, N. (1987b). Structure-activity relations of (2-chloroethyl)nitrosoureas. 1. Deuterium isotope effects in the hydrolysis of 1-(2-chloroethyl)-1-nitrosoureas: evidence for the rate-limiting step. J. Org. Chem. *52*, 484-488.

Burnouf, D., Koehl, P. and Fuchs, R. P. (1989). Single adduct mutagenesis: strong effect of the position of a single acetylaminofluorene adduct within a mutation hot spot. Proc. Natl. Acad. Sci. USA *86*, 4147-4151.

Burns, P. A., Allen, F. L. and Glickman, B. W. (1986). DNA sequence analysis of mutagenicity and site specificity of ethyl methanesulfonate in Uvr^+ and $UvrB^-$ strains of *Escherichia coli*. Genetics *113*, 811-819.

Burns, P. A., Gordon, A. J. and Glickman, B. W. (1987). Influence of neighbouring base sequence on *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine mutagenesis in the *lac*I gene of *Escherichia coli*. J. Mol. Biol. *194*, 385-390.

Burns, P. A., Gordon, A. J. and Glickman, B. W. (1988a). Mutational specificity of *N*-methyl-*N*-nitrosourea in the *lacI* gene of *Escherichia coli*. Carcinogenesis *9*, 1607-1610.

Burns, P. A., Gordon, A. J., Kunsmann, K. and Glickman, B. W. (1988b). Influence of neighboring base sequence on the distribution and repair of *N*-ethyl-*N*-nitrosourea-induced lesions in *Escherichia coli*. Cancer Res. *48*, 4455-4458.

Carter, C. A., Kirk, M. C. and Ludlum, D. B. (1988). Phosphotriester formation by the haloethylnitrosoureas and repair of these lesions by *E. coli* BS21 extracts. Nucleic Acids Res. *16*, 5661-5672.

Cha, R. S., Thilly, W. G. and Zarbl, H. (1994). *N*-Nitroso-*N*-methylurea-induced rat mammary tumors arise from cells with preexisting oncogenic *Hras1* gene mutations. Proc. Natl. Acad. Sci. USA *91*, 3749-3753.

Chandrasekhar, D. and Van Houten, B. (1994). High resolution mapping of UV-induced photoproducts in the *Escherichia coli lacI* gene. Inefficient repair of the non-transcribed strand correlates with high mutation frequency. J. Mol. Biol. *238*, 319-332.

Chargaff, E. (1950). Chemical specificity of nucleic acids and mechanism of their enzymatic degradation. Experientia *6*, 201-209.

Chary, P., Latham, G. J., Robberson, D. L., Kim, S. J., Han, S., Harris, C. M., Harris, T. M. and Lloyd, R. S. (1995a). *In vivo* and *in vitro* replication consequences of stereoisomeric benzo[*a*]pyrene-7,8-dihydrodiol 9,10-epoxide adducts on adenine N^6 at the second position of N- *ras* codon 61. J. Biol. Chem. *270*, 4990-5000.

Chary, P. and Lloyd, R. S. (1995b). *In vitro* replication by prokaryotic and eukaryotic polymerases on DNA templates containing site-specific and stereospecific benzo[*a*]pyrene-7,8-dihydrodiol-9,10-epoxide adducts. Nucleic Acids Res. *23*, 1398-1405.

Christner, D. F., Lakshman, M. K., Sayer, J. M., Jerina, D. M. and Dipple, A. (1994). Primer extension by various polymerases using oligonucleotide templates containing stereoisomeric benzo[*a*]pyrenedeoxyadenosine adducts. Biochemistry *33*, 14297-14305.

Coulondre, C. and Miller, J. H. (1977). Genetic studies of the *lac* repressor. IV. Mutagenic specificity in the *lacI* gene of *Escherichia coli*. J. Mol. Biol. *117*, 577-606.

Daubersies, P., Galiègue-Zouitina, S., Koffel-Schwartz, N., Fuchs, R. P., Loucheux-Lefebvre, M. H. and Bailleul, B. (1992). Mutation spectra of the two guanine adducts of the carcinogen 4-nitroquinoline 1-oxide in *Escherichia coli*. Influence of neighbouring base sequence on mutagenesis. Carcinogenesis *13*, 349-354.

Denhardt, D. T., Dressler, D. and Ray, D. S. (1978). The single-stranded DNA phages (Cold Spring Harbor, New York, NY: Cold Spring Harbor Laboratory Press).

Dolan, M. E., Oplinger, M. and Pegg, A. E. (1988). Sequence specificity of guanine alkylation and repair. Carcinogenesis *9*, 2139-2143.

Dosanjh, M. K., Galeros, G., Goodman, M. F. and Singer, B. (1991). Kinetics of extension of O^6-methylguanine paired with cytosine or thymine in defined oligonucleotide sequences. Biochemistry *30*, 11595-11599.

Dosanjh, M. K., Loechler, E. L. and Singer, B. (1993). Evidence from *in vitro* replication that O^6-methylguanine can adopt multiple conformations. Proc. Natl. Acad. Sci. USA *90*, 3983-3987.

Dunham, S. U. and Lippard, S. J. (1997). DNA sequence context and protein composition modulate HMG-domain protein recognition of cisplatin-modified DNA. Biochemistry *36*, 11428-11436.

Eckel, L. M. and Krugh, T. R. (1994). 2-Aminofluorene modified DNA duplex exists in two interchangeable conformations. Nat. Struct. Biol. *1*, 89-94.

Eckert, K. A., Ingle, C. A. and Drinkwater, N. R. (1989). *N*-Ethyl-*N*-nitrosourea induces A:T to C:G transversion mutations as well as transition mutations in SOS-induced *Escherichia coli*. Carcinogenesis *10*, 2261-2267.

Fazakerley, G. V., Quignard, E., Woisard, A., Guschlbauer, W., van der Marel, G. A., van Boom, J. H., Jones, M. and Radman, M. (1986). Structures of mismatched base pairs in DNA and their recognition by the *Escherichia coli* mismatch repair system. EMBO J. *5*, 3697-3703.

Fix, D. (1993). *N*-Ethyl-*N*-nitrosourea-induced mutagenesis in *Escherichia coli*: multiple roles for UmuC protein. Mutat. Res. *294*, 127-138.

Fix, D. F. and Glickman, B. W. (1986). Differential enhancement of spontaneous transition mutations in the *lacI* gene of an Ung$^-$ strain of *Escherichia coli*. Mutat. Res. *175*, 41-45.

Fix, D. F. and Glickman, B. W. (1987). Asymmetric cytosine deamination revealed by spontaneous mutational specificity in an Ung$^-$ strain of *Escherichia coli*. Mol. Gen. Genet. *209*, 78-82.

Fowler, K. W., Büchi, G. and Essigmann, J. M. (1982). Synthesis and characterization of an oligonucleotide containing a carcinogen-modified base: O^6-methylguanine. J. Am. Chem. Soc. *104*, 1050-1054.

Fowler, R. G., Schaaper, R. M. and Glickman, B. W. (1986). Characterization of mutational specificity within the *lacI* gene for a *mutD5* mutator strain of *Escherichia coli* defective in 3′→5′ exonuclease (proofreading) activity. J. Bacteriol. *167*, 130-137.

Franklin, R. E. and Gosling, R. G. (1953). Molecular configuration in sodium thymonucleate. Nature *171*, 740-741.

Frei, J. V. (1971). Tissue-dependent differences in DNA methylation products of mice treated with methyl-labelled methylnitrosourea. Int. J. Cancer *7*, 436-442.

Fried, M. G., Kanugula, S., Bromberg, J. L. and Pegg, A. E. (1996). DNA binding mechanism of O^6-alkylguanine-DNA alkyltransferase: stoichiometry and effects of DNA base composition and secondary structure on complex stability. Biochemistry *35*, 15295-15301.

Friedberg, E. C., Walker, G. C. and Siede, W. (1995). DNA repair and mutagenesis (Washington, DC: American Society for Microbiology Press).

Friedman, O., Mahapatra, G. N. and Stevenson, R. (1963). The methylation of deoxyribonucleosides by diazomethane. Biochim. Biophys. Acta *68*, 144-146.

Gaffney, B. L. and Jones, R. A. (1989). Thermodynamic comparison of the base pairs formed by the carcinogenic lesion O^6-methylguanine with reference both to Watson-Crick pairs and to mismatched pairs. Biochemistry *28*, 5881-5889.

Gaffney, B. L., Marky, L. A. and Jones, R. A. (1984). Synthesis and characterization of a set of four dodecadeoxyribonucleoside undecaphosphates containing O^6-methylguanine opposite adenine, cytosine, guanine, and thymine. Biochemistry *23*, 5686-5691.

Gentil, A., Margot, A. and Sarasin, A. (1986). 2-(*N*-acetoxy-*N*-acetylamino)fluorene mutagenesis in mammalian cells: sequence-specific hot spot. Proc. Natl. Acad. Sci. USA *83*, 9556-9560.

Georgiadis, P., Smith, C. A. and Swann, P. F. (1991). Nitrosamine-induced cancer: selective repair and conformational differences between O^6-methylguanine residues in different positions in and around codon 12 of rat H-*ras*. Cancer Res. *51*, 5843-5850.

Gerchman, L. L. and Ludlum, D. B. (1973). The properties of O^6-methylguanine in templates for RNA polymerase. Biochim. Biophys. Acta *308*, 310-316.

Gibbs, P. E. and Lawrence, C. W. (1995). Novel mutagenic properties of abasic sites in *Saccharomyces cerevisiae*. J. Mol. Biol. *251*, 229-236.

Ginell, S. L., Kuzmich, S., Jones, R. A. and Berman, H. M. (1990). Crystal and molecular structure of a DNA duplex containing the carcinogenic lesion O^6-methylguanine. Biochemistry *29*, 10461-10465.

Girard, P. M., Guibourt, N. and Boiteux, S. (1997). The Ogg1 protein of *Saccharomyces cerevisiae*: a 7,8-dihydro-8-oxoguanine DNA glycosylase/AP lyase whose lysine 241 is a critical residue for catalytic activity. Nucleic Acids Res. *25*, 3204-3211.

Glen Research catalog (1998), (Sterling, VA: Glen Research).

Gordon, A. J., Burns, P. A. and Glickman, B. W. (1988). *N*-Methyl-*N'*-nitro-*N*-nitrosoguanidine-induced mutation in a RecA strain of *Escherichia coli*. Mutat. Res. *201*, 219-228.

Gordon, A. J., Burns, P. A. and Glickman, B. W. (1990). *N*-Methyl-*N'*-nitro-*N*-nitrosoguanidine induced DNA sequence alteration; non-random components in alkylation mutagenesis. Mutat. Res. *233*, 95-103.

Goth, R. and Rajewsky, M. F. (1974). Persistence of *O*6-ethylguanine in rat-brain DNA: correlation with nervous system-specific carcinogenesis by ethylnitrosourea. Proc. Natl. Acad. Sci. USA *71*, 639-643.

Gottikh, M. B., Ivanovskaya, M. G., Yacobi, L. V. and Shabarova, Z. A. (1985). An effective method of synthesis of phosphodiester derivatives of mono- and oligonucleotides in aqueous medium. Bioorg. Khim. *11*, 613-620.

Graves, R. J., Li, B. F. and Swann, P. F. (1987). Repair of synthetic oligonucleotides containing O^6-methylguanine, O^6-ethylguanine and O^4-methylthymine, by O^6-alkylguanine-DNA alkyl-transferase. IARC Sci. Publ. *84*, 41-43.

Graves, R. J., Li, B. F. and Swann, P. F. (1989). Repair of O^6-methylguanine, O^6-ethylguanine, O^6-isopropylguanine and O^4- methylthymine in synthetic oligodeoxynucleotides by *Escherichia coli ada* gene O^6-alkylguanine-DNA-alkyltransferase. Carcinogenesis *10*, 661-666.

Green, C. L., Loechler, E. L., Fowler, K. W. and Essigmann, J. M. (1984). Construction and characterization of extrachromosomal probes for mutagenesis by carcinogens: site-specific incorporation of O^6-methylguanine into viral and plasmid genomes. Proc. Natl. Acad. Sci. USA *81*, 13-17.

Griffin, S. and Karran, P. (1993). Incision at DNA G·T mispairs by extracts of mammalian cells occurs preferentially at cytosine methylation sites and is not targeted by a separate G·T binding reaction. Biochemistry *32*, 13032-13039.

Gunz, D., Hess, M. T. and Naegeli, H. (1996). Recognition of DNA adducts by human nucleotide excision repair. Evidence for a thermodynamic probing mechanism. J. Biol. Chem. *271*, 25089-25098.

Gupta, P. K., Pandrangi, R. G., Lee, M. S. and King, C. M. (1991). Induction of mutations by *N*-acetoxy-*N*-acetyl-2-aminofluorene modified M13 viral DNA. Carcinogenesis *12*, 819-824.

Hall, J. A. and Saffhill, R. (1983). The incorporation of O^6-methyldeoxyguanosine and O^4-methyldeoxythymidine monophosphates into DNA by DNA polymerases I and α. Nucleic Acids Res. *11*, 4185-4193.

Hamblin, M. R. and Potter, B. V. (1985). *E. coli* Ada regulatory protein repairs the *Sp* diastereoisomer of alkylated DNA. FEBS Lett. *189*, 315-317.

Hanrahan, C. J., Bacolod, M. D., Vyas, R. R., Liu, T., Geacintov, N. E., Loechler, E. L. and Basu, A. K. (1997). Sequence specific mutagenesis of the major (+)-*anti*-benzo[*a*]pyrene diol epoxide-DNA adduct at a mutational hot spot *in vitro* and in *Escherichia coli* cells. Chem. Res. Toxicol. *10*, 369-377.

Hartley, J. A., Lown, J. W., Mattes, W. B. and Kohn, K. W. (1988). DNA sequence specificity of antitumor agents. Oncogenes as possible targets for cancer therapy. Acta Oncol. *27*, 503-510.

Hashim, M. F. and Marnett, L. J. (1996). Sequence-dependent induction of base pair substitutions and frameshifts by propanodeoxyguanosine during *in vitro* DNA replication. J. Biol. Chem. *271*, 9160-9165.

Hatahet, Z., Zhou, M., Reha-Krantz, L. J., Morrical, S. W. and Wallace, S. S. (1998). In search of a mutational hotspot. Proc. Natl. Acad. Sci. USA *95*, 8556-8561.

Hayes, R. C. and LeClerc, J. E. (1986). Sequence dependence for bypass of thymine glycols in DNA by DNA polymerase I. Nucleic Acids Res. *14*, 1045-1061.

Henner, W. D., Grunberg, S. M. and Haseltine, W. A. (1983a). Enzyme action at 3′ termini of ionizing radiation-induced DNA strand breaks. J. Biol. Chem. *258*, 15198-15205.

Henner, W. D., Rodriguez, L. O., Hecht, S. M. and Haseltine, W. A. (1983b). γ-Ray induced deoxyribonucleic acid strand breaks. 3′ Glycolate termini. J. Biol. Chem. *258*, 711-713.

Hepburn, P. A., Margison, G. P. and Tisdale, M. J. (1991). Enzymatic methylation of cytosine in DNA is prevented by adjacent O^6-methylguanine residues. J. Biol. Chem. *266*, 7985-7987.

Horsfall, M. J. and Glickman, B. W. (1988). Mutation site specificity of *N*-nitroso-*N*-methyl-*N*-α-acetoxybenzylamine: a model derivative of an esophageal carcinogen. Carcinogenesis *9*, 1529-1532.

Horsfall, M. J. and Glickman, B. W. (1989a). Mutational specificities of environmental carcinogens in the *lac*I gene of *Escherichia coli*. I. The direct-acting analogue *N*-nitroso-*N*-methyl-*N*-α-acetoxymethylamine. Carcinogenesis *10*, 817-822.

Horsfall, M. J., Zeilmaker, M. J., Mohn, G. R. and Glickman, B. W. (1989b). Mutational specificities of environmental carcinogens in the *lac*I gene of *Escherichia coli*. II: A host-mediated approach to *N*-nitroso-*N*,*N*-dimethylamine and endogenous mutagenesis *in vivo*. Mol. Carcinog. *2*, 107-115.

Hudson, R. F. and Withey, J. (1964). Nucleophilic reactivity. Part V. The reaction between amines and ethyl methanesulphonate. J. Chem. Soc. *678*, 3513-3517.

Hussain, S. P., Aguilar, F. and Cerutti, P. (1994). Mutagenesis of codon 248 of the human p53 tumor suppressor gene by *N*-ethyl-*N*-nitrosourea. Oncogene *9*, 13-18.

Iannone, R., Inga, A., Luque-Romero, F. L., Menichini, P., Abbondandolo, A., Abril, N., Pueyo, C. and Fronza, G. (1997). Mutation spectra analysis suggests that *N*-(2-chloroethyl)-*N*′-cyclohexyl-*N*-nitrosourea-induced lesions are subject to transcription-coupled repair in *Escherichia coli*. Mol. Carcinog. *19*, 39-45.

Ito, T., Nakamura, T., Maki, H. and Sekiguchi, M. (1994). Roles of transcription and repair in alkylation mutagenesis. Mutat. Res. *314*, 273-285.

Jansen, J. G., Mohn, G. R., Vrieling, H., van Teijlingen, C. M., Lohman, P. H. and van Zeeland, A. A. (1994a). Molecular analysis of *hprt* gene mutations in skin fibroblasts of rats exposed *in vivo* to *N*-methyl-*N*-nitrosourea or *N*-ethyl-*N*-nitrosourea. Cancer Res. *54*, 2478-2485.

Jansen, J. G., van Teijlingen, C. M., Mohn, G. R., van Zeeland, A. A. and Vrieling, H. (1994b). AT base pairs are the main target for mutations at the hprt locus of rat skin fibroblasts exposed *in vitro* to the monofunctional alkylating agent *N*-ethyl- *N*-nitrosourea. Mutagenesis *9*, 417-421.

Jelinsky, S. A., Liu, T., Geacintov, N. E. and Loechler, E. L. (1995). The major, N^2-Gua adduct of the (+)-*anti*-benzo[*a*]pyrene diol epoxide is capable of inducing G→A and G→C, in addition to G→T, mutations. Biochemistry *34*, 13545-13553.

Jiao, J., Pienkowska, M., Glickman, B. W. and Zielenska, M. (1996). Molecular analysis of mutations induced by ethylating *N*-nitroso compounds in the *lacI* gene of *Escherichia coli*. Mutat. Res. *352*, 39-45.

Jones, B. K. and Yeung, A. T. (1990). DNA base composition determines the specificity of UvrABC endonuclease incision of a psoralen cross-link. J. Biol. Chem. *265*, 3489-3496.

Jurado, J., Ferrezuelo, F. and Pueyo, C. (1995). Mutational specificity of 1-(2-chloroethyl)-3-cyclohexyl-1-nitrosourea in the *Escherichia coli lacI* gene of O^6-alkylguanine-DNA alkyltransferase-proficient and -deficient strains. Mol. Carcinog. *14*, 233-239.

Kalnik, M. W., Li, B. F., Swann, P. F. and Patel, D. J. (1989a). O^6-Ethylguanine carcinogenic lesions in DNA: an NMR study of O^6etG.C pairing in dodecanucleotide duplexes. Biochemistry *28*, 6182-6192.

Kalnik, M. W., Li, B. F., Swann, P. F. and Patel, D. J. (1989b). O^6-Ethylguanine carcinogenic lesions in DNA: an NMR study of O^6etG.T pairing in dodecanucleotide duplexes. Biochemistry *28*, 6170-6181.

Kamiya, H. and Kasai, H. (1997). Substitution and deletion mutations induced by 2-hydroxyadenine in *Escherichia coli*: effects of sequence contexts in leading and lagging strands. Nucleic Acids Res. *25*, 304-311.

Kleihues, P. and Margison, G. P. (1974). Carcinogenicity of *N*-methyl-*N*-nitrosourea: possible role of excision repair of O^6-methylguanine from DNA. J. Natl. Cancer Inst. *53*, 1839-1841.

Klein, S. and Oesch, F. (1990). A new assay for O^6-alkylguanine-DNA-alkyltransferase to determine DNA repair capacities using lambda-phage DNA as substrate. Carcinogenesis *11*, 1771-1774.

Koch, W. H., Henrikson, E. N., Kupchella, E. and Cebula, T. A. (1994). *Salmonella typhimurium* strain TA100 differentiates several classes of carcinogens and mutagens by base substitution specificity. Carcinogenesis *15*, 79-88.

Koehl, P., Valladier, P., Lefèvre, J. F. and Fuchs, R. P. (1989). Strong structural effect of the position of a single acetylaminofluorene adduct within a mutation hot spot. Nucleic Acids Res. *17*, 9531-9541.

Koffel-Schwartz, N. and Fuchs, R. P. (1995). Sequence determinants for -2 frameshift mutagenesis at *Nar*I-derived hot spots. J. Mol. Biol. *252*, 507-513.

Kumar, N. V. and Varshney, U. (1994). Inefficient excision of uracil from loop regions of DNA oligomers by *E. coli* uracil DNA glycosylase. Nucleic Acids Res. *22*, 3737-3741.

LaMarr, W. A., Sandman, K. M., Reeve, J. N. and Dedon, P. C. (1997). Differential effects of DNA supercoiling on radical-mediated DNA strand breaks. Chem. Res. Toxicol. *10*, 1118-1122.

LaMarr, W. A., Yu, L., Nicolaou, K. C. and Dedon, P. C. (1998). Supercoiling affects the accessibility of glutathione to DNA-bound molecules: positive supercoiling inhibits calicheamicin-induced DNA damage. Proc. Natl. Acad. Sci. USA *95*, 102-107.

Lambert, I. B., Gordon, A. J., Bryant, D. W., Glickman, B. W. and McCalla, D. R. (1991). The action of 1-nitroso-8-nitropyrene in *Escherichia coli*: DNA adduct formation and mutational consequences in the absence of nucleotide excision-repair. Carcinogenesis *12*, 879-884.

Lambert, I. B., Gordon, A. J., Glickman, B. W. and McCalla, D. R. (1992a). The influence of local DNA sequence and DNA repair background on the mutational specificity of 1-nitroso-8-nitropyrene in *Escherichia coli*: inferences for mutagenic mechanisms. Genetics *132*, 911-927.

Lambert, I. B., Napolitano, R. L. and Fuchs, R. P. (1992b). Carcinogen-induced frameshift mutagenesis in repetitive sequences. Proc. Natl. Acad. Sci. USA *89*, 1310-1314.

Lawley, P. D., Orr, D. J., Shah, S. A., Farmer, P. B. and Jarman, M. (1973a). Reaction products from *N*-methyl-*N*-nitrosourea and deoxyribonucleic acid containing thymidine residues. Synthesis and identification of a new methylation product, O^4-methylthymidine. Biochem. J. *135*, 193-201.

Lawley, P. D. and Shah, S. A. (1973b). Methylation of DNA by ^{3}H-^{14}C-*methyl*-labelled *N*-methyl-*N*-nitrosourea - evidence for transfer of the intact methyl group. Chem. Biol. Interact. *7*, 115-120.

Lawley, P. D. and Thatcher, C. J. (1970). Methylation of deoxyribonucleic acid in cultured mammalian cells by *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine. The influence of cellular thiol concentrations on the extent of methylation and the 6-oxygen atom of guanine as a site of methylation. Biochem. J. *116*, 693-707.

Lawrence, C. W., Borden, A., Banerjee, S. K. and LeClerc, J. E. (1990). Mutation frequency and spectrum resulting from a single abasic site in a single-stranded vector. Nucleic Acids Res. *18*, 2153-2157.

LeDoux, S. P., Patton, N. J., Nelson, J. W., Bohr, W. A. and Wilson, G. L. (1990). Preferential DNA repair of alkali-labile sites within the active insulin gene. J. Biol. Chem. *265*, 14875-14880.

Lee, G. S., Blonsky, K. S., van On, D. L., Savage, E. A., Morgan, A. R. and von Borstel, R. C. (1992). Base alterations in yeast induced by alkylating agents with differing Swain-Scott substrate constants. J. Mol. Biol. *223*, 617-626.

Lee, S. Y. and Rasheed, S. (1990). A simple procedure for maximum yield of high-quality plasmid DNA. BioTechniques *9*, 676-679.

Leonard, G. A., Thomson, J., Watson, W. P. and Brown, T. (1990). High-resolution structure of a mutagenic lesion in DNA. Proc. Natl. Acad. Sci. USA *87*, 9573-9576.

Levine, J. G., Schaaper, R. M. and DeMarini, D. M. (1994). Complex frameshift mutations mediated by plasmid pKM101: mutational mechanisms deduced from 4-aminobiphenyl-induced mutation spectra in Salmonella. Genetics *136*, 731-746.

Lieb, M. and Rehmat, S. (1995). Very short patch repair of T:G mismatches *in vivo*: importance of context and accessory proteins. J. Bacteriol. *177*, 660-666.

Lindahl, T., Demple, B. and Robins, P. (1982). Suicide inactivation of the *E. coli* O^6-methylguanine-DNA methyltransferase. EMBO J. *1*, 1359-1363.

Litinski, V., Chenna, A., Sagi, J. and Singer, B. (1997). Sequence context is an important determinant in the mutagenic potential of 1,N^6-ethenodeoxyadenosine (εA): formation of εA basepairs and elongation in defined templates. Carcinogenesis *18*, 1609-1615.

Liu, T., Xu, J., Tsao, H., Li, B., Xu, R., Yang, C., Amin, S., Moriya, M. and Geacintov, N. E. (1996). Base sequence-dependent bends in site-specific benzo[*a*]pyrene diol epoxide-modified oligonucleotide duplexes. Chem. Res. Toxicol. *9*, 255-261.

Loechler, E. L. (1991). Rotation about the C6-O^6 bond in O^6-methylguanine: the *syn* and *anti* conformers can be of similar energies in duplex DNA as estimated by molecular modeling techniques. Carcinogenesis *12*, 1693-1699.

Loechler, E. L. (1996). The role of adduct site-specific mutagenesis in understanding how carcinogen-DNA adducts cause mutations: perspective, prospects and problems. Carcinogenesis *17*, 895-902.

Loechler, E. L., Green, C. L. and Essigmann, J. M. (1984). *In vivo* mutagenesis by O^6-methylguanine built into a unique site in a viral genome. Proc. Natl. Acad. Sci. USA *81*, 6271-6275.

Loveless, A. (1969). Possible relevance of O-6 alkylation of deoxyguanosine to the mutagenicity and carcinogenicity of nitrosamines and nitrosamides. Nature *223*, 206-207.

Luque-Romero, F. L., Iannone, R., Inga, A., Fronza, G. and Pueyo, C. (1997). Mutational specificity of 1-(2-chloroethyl)-3-cyclohexyl-1-nitrosourea in *Escherichia coli*: comparison of *in vivo* with *in vitro* exposure of the *supF* gene. Environ. Mol. Mutagen. *30*, 65-71.

Maccabee, M., Evans, J. S., Glackin, M. P., Hatahet, Z. and Wallace, S. S. (1994). Pyrimidine ring fragmentation products. Effects of lesion structure and sequence context on mutagenesis. J. Mol. Biol. *236*, 514-530.

Mackay, W. J., Han, S. and Samson, L. D. (1994). DNA alkylation repair limits spontaneous base substitution mutations in *Escherichia coli*. J. Bacteriol. *176*, 3224-3230.

Masse, J. E., Bortmann, P., Dieckmann, T. and Feigon, J. (1998). Simple, efficient protocol for enzymatic synthesis of uniformly ^{13}C, ^{15}N-labeled DNA for heteronuclear NMR studies. Nucleic Acids Res. *26*, 2618-2624.

McCarthy, T. V. and Lindahl, T. (1985). Methyl phosphotriesters in alkylated DNA are repaired by the Ada regulatory protein of *E. coli*. Nucleic Acids Res. *13*, 2683-2698.

Mekhovich, O., Tang, M. S. and Romano, L. J. (1998). Rate of incision of *N*-acetyl-2-aminofluorene and *N* -2-aminofluorene adducts by UvrABC nuclease is adduct- and sequence-specific: comparison of the rates of UvrABC nuclease incision and protein-DNA complex formation. Biochemistry *37*, 571-579.

Mellon, I., Spivak, G. and Hanawalt, P. C. (1987). Selective removal of transcription-blocking DNA damage from the transcribed strand of the mammalian DHFR gene. Cell *51*, 241-249.

Miller, J. H. (1992). A short course in bacterial genetics: a laboratory manual and handbook for *Escherichia coli* and related bacteria (Cold Spring Harbor, New York, NY: Cold Spring Harbor Laboratory Press).

Mironov, N. M., Bleicher, F., Martel-Planche, G. and Montesano, R. (1993). Nonrandom distribution of O^6-methylguanine in H-ras gene sequence from DNA modified with *N*-methyl-*N*-nitrosourea. Mutat. Res. *288*, 197-205.

Mironov, N. M., Wild, C. P., Martel-Planche, G., Swann, P. F. and Montesano, R. (1989). Measurement of the removal of O^6-methylguanine and O^4-methylthymine from oligodeoxynucleotides using an immunoprecipitation technique. Anal. Biochem. *183*, 74-79.

Mitra, G., Pauly, G. T., Kumar, R., Pei, G. K., Hughes, S. H., Moschel, R. C. and Barbacid, M. (1989). Molecular analysis of O^6-substituted guanine-induced mutagenesis of *ras* oncogenes. Proc. Natl. Acad. Sci. USA *86*, 8650-8654.

Moore, M. H., Gulbis, J. M., Dodson, E. J., Demple, B. and Moody, P. C. (1994). Crystal structure of a suicidal DNA repair protein: the Ada O^6-methylguanine-DNA methyltransferase from *E. coli*. EMBO J. *13*, 1495-1501.

Moriwaki, S. I., Yagi, T., Nishigori, C., Imamura, S. and Takebe, H. (1991). Analysis of *N*-methyl-*N*-nitrosourea-induced mutations in a shuttle vector plasmid propagated in mouse O^6-methylguanine-DNA methyltransferase-deficient cells in comparison with proficient cells. Cancer Res. *51*, 6219-6223.

Moriya, M., Ou, C., Bodepudi, V., Johnson, F., Takeshita, M. and Grollman, A. P. (1991). Site-specific mutagenesis using a gapped duplex vector: a study of translesion synthesis past 8-oxodeoxyguanosine in *E. coli*. Mutat. Res. *254*, 281-288.

Moriya, M., Spiegel, S., Fernandes, A., Amin, S., Liu, T., Geacintov, N. and Grollman, A. P. (1996). Fidelity of translesional synthesis past benzo[*a*]pyrene diol epoxide-2′-deoxyguanosine DNA adducts: marked effects of host cell, sequence context, and chirality. Biochemistry *35*, 16646-16651.

Mu, D., Bertrand-Burggraf, E., Huang, J. C., Fuchs, R. P. and Sancar, A. (1994). Human and *E.coli* excinucleases are affected differently by the sequence context of acetylaminofluorene-guanine adduct. Nucleic Acids Res. *22*, 4869-4871.

Musarrat, J., Arezina-Wilson, J. and Wani, A. A. (1995). Repair of base alkylation damage in targeted restriction endonuclease sequences of plasmid DNA. Biochim. Biophys. Acta *1263*, 201-211.

Müller, R. and Rajewsky, M. F. (1980). Immunological quantification by high-affinity antibodies of O^6-ethyldeoxyguanosine in DNA exposed to *N*-ethyl-*N*-nitrosourea. Cancer Res. *40*, 887-896.

Myles, G. M., Van Houten, B. and Sancar, A. (1987). Utilization of DNA photolyase, pyrimidine dimer endonucleases, and alkali hydrolysis in the analysis of aberrant ABC excinuclease incisions adjacent to UV-induced DNA photoproducts. Nucleic Acids Res. *15*, 1227-1243.

Napolitano, R. L., Lambert, I. B. and Fuchs, R. P. (1994). DNA sequence determinants of carcinogen-induced frameshift mutagenesis. Biochemistry *33*, 1311-1315.

Nehls, P. and Rajewsky, M. F. (1985). Ethylation of nucleophilic sites in DNA by *N*-ethyl-*N*-nitrosourea depends on chromatin structure and ionic strength. Mutat. Res. *150*, 13-21.

Nehls, P., Spiess, E., Weber, E., Berger, J. and Rajewsky, M. F. (1988). Distribution of O^6-ethylguanine in DNA exposed to ethylnitrosourea *in vitro* as visualized by electron microscopy using a monoclonal antibody. Mutat. Res. *198*, 179-189.

New England Biolabs catalog (1998), (Beverly, MA: New England Biolabs).

Nilsen, H., Yazdankhah, S. P., Eftedal, I. and Krokan, H. E. (1995). Sequence specificity for removal of uracil from U·A pairs and U·G mismatches by uracil-DNA glycosylase from *Escherichia coli*, and correlation with mutational hotspots. FEBS Lett. *362*, 205-209.

Op het Veld, C. W., van Hees-Stuivenberg, S., van Zeeland, A. A. and Jansen, J. G. (1997). Effect of nucleotide excision repair on *hprt* gene mutations in rodent cells exposed to DNA ethylating agents. Mutagenesis *12*, 417-424.

Osterman-Golkar, S., Ehrenberg, L. and Wachtmeister, C. A. (1970). Reaction kinetics and biological action in barley of mono-functional methanesulfonic esters. Radiation Botany *10*, 303-327.

Page, J. E., Pilcher, A. S., Yagi, H., Sayer, J. M., Jerina, D. M. and Dipple, A. (1999). Mutational consequences of replication of M13mp7L2 constructs containing cis-opened Benzo[*a*]pyrene 7,8-diol 9,10-epoxide-deoxyadenosine adducts. Chem. Res. Toxicol. *12*, 258-263.

Page, J. E., Zajc, B., Oh-hara, T., Lakshman, M. K., Sayer, J. M., Jerina, D. M. and Dipple, A. (1998). Sequence context profoundly influences the mutagenic potency of trans-opened benzo[*a*]pyrene 7,8-diol 9,10-epoxide-purine nucleoside adducts in site-specific mutation studies. Biochemistry *37*, 9127-9137.

Pauly, G. T., Hughes, S. H. and Moschel, R. C. (1994). Response of repair-competent and repair-deficient *Escherichia coli* to three O^6-substituted guanines and involvement of methyl-directed mismatch repair in the processing of O^6-methylguanine residues. Biochemistry *33*, 9169-9177.

Pauly, G. T., Hughes, S. H. and Moschel, R. C. (1995). Mutagenesis in *Escherichia coli* by three O^6-substituted guanines in double-stranded or gapped plasmids. Biochemistry *34*, 8924-8930.

Pauly, G. T., Hughes, S. H. and Moschel, R. C. (1998). Comparison of mutagenesis by O^6-methyl- and O^6-ethylguanine and O^4-methylthymine in *Escherichia coli* using double-stranded and gapped plasmids. Carcinogenesis *19*, 457-461.

Pfeifer, G. P., Drouin, R., Riggs, A. D. and Holmquist, G. P. (1991). *In vivo* mapping of a DNA adduct at nucleotide resolution: detection of pyrimidine (6-4) pyrimidone photoproducts by ligation-mediated polymerase chain reaction. Proc. Natl. Acad. Sci. USA *88*, 1374-1378.

Pienkowska, M., Glickman, B. W., Ferreira, A., Anderson, M. and Zielenska, M. (1993). Large-scale mutational analysis of EMS-induced mutation in the *lacI* gene of *Escherichia coli*. Mutat. Res. *288*, 123-131.

Piršel, M. and Bohr, V. A. (1993). Methyl methanesulfonate adduct formation and repair in the DHFR gene and in mitochondrial DNA in hamster cells. Carcinogenesis *14*, 2105-2108.

Pletsa, V., Gentil, A., Margot, A., Armier, J., Kyrtopoulos, S. A. and Sarasin, A. (1992). Mutagenesis by O^6 meG residues within codon 12 of the human Ha-ras proto-oncogene in monkey cells. Nucleic Acids Res. *20*, 4897-4901.

Pletsa, V., Troungos, C., Souliotis, V. L. and Kyrtopoulos, S. A. (1994). Comparative study of mutagenesis by O^6-methylguanine in the human Ha-ras oncogene in *E. coli* and *in vitro*. Nucleic Acids Res. *22*, 3846-3853.

Pourzand, C. and Cerutti, P. (1993). Mutagenesis of H-*ras* codons 11 and 12 in human fibroblasts by *N*-ethyl-*N*-nitrosourea. Carcinogenesis *14*, 2193-2196.

Pullman, A. and Pullman, B. (1981). Molecular electrostatic potential of the nucleic acids. Q. Rev. Biophys. *14*, 289-380.

Purmal, A. A., Kow, Y. W. and Wallace, S. S. (1994). Major oxidative products of cytosine, 5-hydroxycytosine and 5-hydroxyuracil, exhibit sequence context-dependent mispairing *in vitro*. Nucleic Acids Res. *22*, 72-78.

Rasmussen, L. J. and Samson, L. (1996). The *Escherichia coli* MutS DNA mismatch binding protein specifically binds *O*^{6}-methylguanine DNA lesions. Carcinogenesis *17*, 2085-2088.

Rebeck, G. W. and Samson, L. (1991). Increased spontaneous mutation and alkylation sensitivity of *Escherichia coli* strains lacking the *ogt* *O*^{6}-methylguanine DNA repair methyltransferase. J. Bacteriol. *173*, 2068-2076.

Rebeck, G. W., Smith, C. M., Goad, D. L. and Samson, L. (1989). Characterization of the major DNA repair methyltransferase activity in unadapted *Escherichia coli* and identification of a similar activity in *Salmonella typhimurium*. J. Bacteriol. *171*, 4563-4568.

Reed, J. and Hutchinson, F. (1987). Effect of the direction of DNA replication on mutagenesis by *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine in adapted cells of *Escherichia coli*. Mol. Gen. Genet. *208*, 446-449.

Richardson, F. C., Boucheron, J. A., Skopek, T. R. and Swenberg, J. A. (1989). Formation of *O*^{6}-methyldeoxyguanosine at specific sites in a synthetic oligonucleotide designed to resemble a known mutagenic hotspot. J. Biol. Chem. *264*, 838-841.

Richardson, K. K., Crosby, R. M., Richardson, F. C. and Skopek, T. R. (1987b). DNA base changes induced following *in vivo* exposure of unadapted, adapted or Ada$^-$ *Escherichia coli* to N-methyl-N'-nitro-N-nitrosoguanidine. Mol. Gen. Genet. *209*, 526-532.

Richardson, K. K., Crosby, R. M. and Skopek, T. R. (1988). Mutation spectra of *N*-ethyl-*N'*-nitro-*N*-nitrosoguanidine and 1-(2-hydroxyethyl)-1-nitrosourea in *Escherichia coli*. Mol. Gen. Genet. *214*, 460-466.

Richardson, K. K., Richardson, F. C., Crosby, R. M., Swenberg, J. A. and Skopek, T. R. (1987a). DNA base changes and alkylation following *in vivo* exposure of *Escherichia coli* to *N*-methyl-*N*-nitrosourea or *N*-ethyl-*N*-nitrosourea. Proc. Natl. Acad. Sci. USA *84*, 344-348.

Rodriguez, H. and Loechler, E. L. (1993a). Mutagenesis by the (+)-*anti*-diol epoxide of benzo[*a*]pyrene: what controls mutagenic specificity? Biochemistry *32*, 1759-1769.

Rodriguez, H. and Loechler, E. L. (1993b). Mutational specificity of the (+)-*anti*-diol epoxide of benzo[*a*]pyrene in a *supF* gene of an *Escherichia coli* plasmid: DNA sequence context influences hotspots, mutagenic specificity and the extent of SOS enhancement of mutagenesis. Carcinogenesis *14*, 373-383.

Roldán-Arjona, T., Luque-Romero, F. L., Ariza, R. R., Jurado, J. and Pueyo, C. (1994). Influence of DNA repair by *ada* and *ogt* alkyltransferases on the mutational specificity of alkylating agents. Mol. Carcinog. *9*, 200-209.

Rossi, S. C. and Topal, M. D. (1991). Mutagenic frequencies of site-specifically located O^6-methylguanine in wild-type *Escherichia coli* and in a strain deficient in Ada-methyltransferase. J. Bacteriol. *173*, 1201-1207.

Rubin, E., Rumney IV, S., Wang, S. and Kool, E. T. (1995). Convergent DNA synthesis: a non-enzymatic dimerization approach to circular oligodeoxynucleotides. Nucleic Acids Res. *23*, 3547-3553.

Rydberg, B. and Lindahl, T. (1982). Nonenzymatic methylation of DNA by the intracellular methyl group donor S-adenosyl-L-methionine is a potentially mutagenic reaction. EMBO J. *1*, 211-216.

Sage, E. and Moustacchi, E. (1987). Sequence context effects on 8-methoxypsoralen photobinding to defined DNA fragments. Biochemistry *26*, 3307-3314.

Saito, I., Takayama, M., Sugiyama, H. and Nakatani, K. (1995). Photoinduced DNA cleavage via electron transfer: demonstration that guanine residues located 5′ to guanine are the most electron-donating sites. J. Am. Chem. Soc. *117*, 6406-6407.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). Molecular cloning: a laboratory manual (2nd edn), (Cold Spring Harbor, New York, NY: Cold Spring Harbor Laboratory Press).

Samson, L. (1992). The suicidal DNA repair methyltransferases of microbes. Molecular Microbiology *6*, 825-831.

Samson, L. and Cairns, J. (1977). A new pathway for DNA repair in *Escherichia coli*. Nature *267*, 281-283.

Samson, L., Han, S., Marquis, J. C. and Rasmussen, L. J. (1997). Mammalian DNA repair methyltransferases shield O^4MeT from nucleotide excision repair. Carcinogenesis *18*, 919-924.

Samson, L., Thomale, J. and Rajewsky, M. F. (1988). Alternative pathways for the *in vivo* repair of O^6-alkylguanine and O^4-alkylthymine in *Escherichia coli*: the adaptive response and nucleotide excision repair. EMBO J. *7*, 2261-2267.

Sandigursky, M. and Franklin, W. A. (1998). Exonuclease IX of *Escherichia coli* removes 3′ phosphoglycolate end groups from DNA. Radiat. Res. *150*, 609-611.

Sassanfar, M., Dosanjh, M. K., Essigmann, J. M. and Samson, L. (1991). Relative efficiencies of the bacterial, yeast, and human DNA methyltransferases for the repair of O^6-methylguanine and O^4-methylthymine. Suggestive evidence for O^4-methylthymine repair by eukaryotic methyltransferases. J. Biol. Chem. *266*, 2767-2771.

Sauerbier, W. (1986). Pyrimidine dimer dependent cleavage of single-stranded DNA by T4 UV endonuclease. Biochem. Biophys. Res. Commun. *141*, 204-212.

Schaaper, R. M. and Dunn, R. L. (1987). Spectra of spontaneous mutations in *Escherichia coli* strains defective in mismatch correction: the nature of *in vivo* DNA replication errors. Proc. Natl. Acad. Sci. USA *84*, 6220-6224.

Schendel, P. F. and Robins, P. E. (1978). Repair of O^6-methylguanine in adapted *Escherichia coli*. Proc. Natl. Acad. Sci. USA *75*, 6017-6020.

Scicchitano, D. A. and Hanawalt, P. C. (1990). Lack of sequence-specific removal of *N*-methylpurines from cellular DNA. Mutat. Res. *233*, 31-37.

Sedgwick, B. (1997). Nitrosated peptides and polyamines as endogenous mutagens in O^6-alkylguanine-DNA alkyltransferase deficient cells. Carcinogenesis *18*, 1561-1567.

Seeberg, E. and Fuchs, R. P. (1990). Acetylaminofluorene bound to different guanines of the sequence -GGCGCC- is excised with different efficiencies by the UvrABC excision nuclease in a pattern not correlated to the potency of mutation induction. Proc. Natl. Acad. Sci. USA *87*, 191-194.

Sendowski, K. and Rajewsky, M. F. (1991). DNA sequence dependence of guanine-O^6 alkylation by the *N*-nitroso carcinogens *N*-methyl- and *N*-ethyl-*N*-nitrosourea. Mutat. Res. *250*, 153-160.

Shevell, D. E., Abou-Zamzam, A. M., Demple, B. and Walker, G. C. (1988). Construction of an *Escherichia coli* K-12 *ada* deletion by gene replacement in a *recD* strain reveals a second methyltransferase that repairs alkylated DNA. J. Bacteriol. *170*, 3294-3296.

Shibutani, S., Takeshita, M. and Grollman, A. P. (1991). Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG. Nature *349*, 431-434.

Shimizu, H., Yagi, R., Kimura, Y., Makino, K., Terato, H., Ohyama, Y. and Ide, H. (1997). Replication bypass and mutagenic effect of α-deoxyadenosine site-specifically incorporated into single-stranded vectors. Nucleic Acids Res. *25*, 597-603.

Shukla, R., Liu, T., Geacintov, N. E. and Loechler, E. L. (1997). The major, N^2-dG adduct of (+)-*anti*-B[*a*]PDE shows a dramatically different mutagenic specificity (predominantly, G→A) in a 5′-CGT-3′ sequence context. Biochemistry *36*, 10256-10261.

Singer, B., Chavez, F., Goodman, M. F., Essigmann, J. M. and Dosanjh, M. K. (1989). Effect of 3′ flanking neighbors on kinetics of pairing of dCTP or dTTP opposite O^6-methylguanine in a defined primed oligonucleotide when *Escherichia coli* DNA polymerase I is used. Proc. Natl. Acad. Sci. USA *86*, 8271-8274.

Singer, B., Sun, L. and Fraenkel-Conrat, H. (1975). Effects of alkylation of phosphodiesters and of bases of infectivity and stability of tobacco mosaic virus RNA. Proc. Natl. Acad. Sci. USA *72*, 2232-2236.

Sockett, H., Romac, S. and Hutchinson, F. (1991). DNA sequence changes in mutations induced by ultraviolet light in the *gpt* gene on the chromosome of *Escherichia coli* uvr^+ and *urvA* cells. Mol. Gen. Genet. *230*, 295-301.

Souliotis, V. L., Giannopoulos, A., Koufakis, I., Kaila, S., Dimopoulos, C. and Kyrtopoulos, S. A. (1989). Development and validation of a new assay for O^6-alkylguanine-DNA-alkyltransferase based on the use of an oligonucleotide substrate, and its application to the measurement of DNA repair activity in extracts of biopsy samples of human urinary bladder mucosa. Carcinogenesis *10*, 1203-1208.

Strauss, B. S., Sagher, D. and Acharya, S. (1997). Role of proofreading and mismatch repair in maintaining the stability of nucleotide repeats in DNA. Nucleic Acids Res. *25*, 806-813.

Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugita, A., Terzaghi, E. and Inouye, M. (1966). Frameshift mutations and the genetic code. Cold Spring Harbor Symposium of Quantitative Biology *31*, 77-84.

Sugiyama, H and Saito, I. (1996). Theoretical studies of GG-specific photocleavage of DNA via electron transfer: significant lowering of ionization potential and 5′-localization of HOMO of stacked GG bases in B-form DNA. J. Am. Chem. Soc. *118*, 7063-7068.

Sukumar, S., Notario, V., Martin-Zanca, D. and Barbacid, M. (1983). Induction of mammary carcinomas in rats by nitroso-methylurea involves malignant activation of H-*ras*-1 locus by single point mutations. Nature *306*, 658-661.

Svoboda, D. L., Smith, C. A., Taylor, J. S. and Sancar, A. (1993). Effect of sequence, adduct type, and opposing lesions on the binding and repair of ultraviolet photodamage by DNA photolyase and (A)BC excinuclease. J. Biol. Chem. *268*, 10694-10700.

Swann, P. F. (1990). Why do O^6-alkylguanine and O^4-alkylthymine miscode? The relationship between the structure of DNA containing O^6-alkylguanine and O^4-alkylthymine and the mutagenic properties of these bases. Mutat. Res. *233*, 81-94.

Tan, H. B., Swann, P. F. and Chance, E. M. (1994). Kinetic analysis of the coding properties of O^6-methylguanine in DNA: the crucial role of the conformation of the phosphodiester bond. Biochemistry *33*, 5335-5346.

Taverna, P. and Sedgwick, B. (1996). Generation of an endogenous DNA-methylating agent by nitrosation in *Escherichia coli*. J. Bacteriol. *178*, 5105-5111.

Tebbs, R. S. and Romano, L. J. (1994). Mutagenesis at a site-specifically modified *Nar*I sequence by acetylated and deacetylated aminofluorene adducts. Biochemistry *33*, 8998-9006.

Thomale, J., Hochleitner, K. and Rajewsky, M. F. (1994). Differential formation and repair of the mutagenic DNA alkylation product O^6-ethylguanine in transcribed and nontranscribed genes of the rat. J. Biol. Chem. *269*, 1681-1686.

Todd, P. A. and Glickman, B. W. (1982). Mutational specificity of UV light in *Escherichia coli*: indications for a role of DNA secondary structure. Proc. Natl. Acad. Sci. USA *79*, 4123-4127.

Toorchen, D. and Topal, M. D. (1983). Mechanisms of chemical mutagenesis and carcinogenesis: effects on DNA replication of methylation at the O^6-guanine position of dGTP. Carcinogenesis *4*, 1591-1597.

Topal, M. D. and Baker, M. S. (1982). DNA precursor pool: a significant target for *N*-methyl-*N*-nitrosourea in C3H/10T½ clone 8 cells. Proc. Natl. Acad. Sci. USA *79*, 2211-2215.

Topal, M. D., Eadie, J. S. and Conrad, M. (1986). O^6-methylguanine mutation and repair is nonuniform. Selection for DNA most interactive with O^6-methylguanine. J. Biol. Chem. *261*, 9879-9885.

van der Vliet, G. M., Zielenska, M., Anderson, M. W. and Glickman, B. W. (1989). The complexity of nitrosoguanidine mutagenesis increases with size: observations of the mutational specificity of *N*-propyl-*N'*-nitro-*N*-nitrosoguanidine. Carcinogenesis *10*, 949-952.

van der Vliet, G. M., Zielenska, M., Anderson, M. W. and Glickman, B. W. (1990). The influence of excision repair on the distribution of *N*-propyl-*N'*-nitro-*N*-nitrosoguanidine-induced mutation in *Escherichia coli*. Mutagenesis *5*, 127-130.

Veaute, X. and Fuchs, R. P. (1993). Greater susceptibility to mutations in lagging strand of DNA replication in *Escherichia coli* than in leading strand. Science *261*, 598-600.

Velemínský, J., Osterman-Golkar, S. and Ehrenberg, L. (1970). Reaction rates and biological action of *N*-methyl- and *N*-ethyl-*N*-nitrosourea. Mutat. Res. *10*, 169-174.

Verri, A., Mazzarello, P., Spadari, S. and Focher, F. (1992). Uracil-DNA glycosylases preferentially excise mispaired uracil. Biochem. J. *287*, 1007-1010.

Vidal, A., Abril, N. and Pueyo, C. (1995). DNA repair by Ogt alkyltransferase influences EMS mutational specificity. Carcinogenesis *16*, 817-821.

Vidal, A., Abril, N. and Pueyo, C. (1997). The influence of DNA repair by Ogt alkyltransferase on the distribution of alkylnitrosourea-induced mutations in *Escherichia coli*. Environ. Mol. Mutagen. *29*, 180-188.

Voigt, J. M. and Topal, M. D. (1990). O^6-Methylguanine and A·C and G·T mismatches cause asymmetric structural defects in DNA that are affected by DNA sequence. Biochemistry *29*, 5012-5018.

Voigt, J. M. and Topal, M. D. (1995). O^6-Methylguanine-induced replication blocks. Carcinogenesis *16*, 1775-1782.

Voigt, J. M., Van Houten, B., Sancar, A. and Topal, M. D. (1989). Repair of O^6-methylguanine by ABC excinuclease of *Escherichia coli in vitro*. J. Biol. Chem. *264*, 5172-5176.

Voitkun, V., Zhitkovich, A. and Costa, M. (1998). Cr(III)-mediated crosslinks of glutathione or amino acids to the DNA phosphate backbone are mutagenic in human cells. Nucleic Acids Res. *26*, 2024-2030.

Warpehoski, M. A. and Hurley, L. H. (1988). Sequence selectivity of DNA covalent modification. Chem. Res. Toxicol. *1*, 315-333.

Watson, J. D. and Crick, F. H. (1953a). Genetical implications of the structure of deoxyribonucleic acid. Nature *171*, 964-967.

Watson, J. D. and Crick, F. H. (1953b). Molecular structure of nucleic acids. A structure for deoxyribose nucleic acid. Nature *171*, 737-738.

Wheeler, G. P. and Bowdon, B. J. (1972). Comparison of the effects of cysteine upon the decomposition of nitrosoureas and of 1-methyl-3-nitro-1-nitrosoguanidine. Biochem. Pharmacol. *21*, 265-267.

Wilkins, M. H., Stokes, A. R. and Wilson, H. R. (1953). Molecular structure of deoxypentose nucleic acids. Nature *171*, 738-740.

Williams, L. D. and Shaw, B. R. (1987). Protonated base pairs explain the ambiguous pairing properties of O^6-methylguanine. Proc. Natl. Acad. Sci. USA *84*, 1779-1783.

Wilson III, D. M., Engelward, B. P., and Samson, L (1998). Prokaryotic base excision repair. *In* DNA damage and repair, Vol. I: DNA repair in prokaryotes and lower eukaryotes. (Nickoloff, J. A. and Hoekstra, M. F., ed.), pp 29-64, (Totowa, NJ: Humana Press Inc.).

Yadollahi-Farsani, M., Gooderham, N. J., Davies, D. S. and Boobis, A. R. (1996). Mutational spectra of the dietary carcinogen 2-amino-1-methyl-6-phenylimidazo[4,5-*b*]pyridine (PhIP) at the Chinese hamster *hprt* locus. Carcinogenesis *17*, 617-624.

Yang, J. L., Hsieh, F. P., Lee, P. C. and Tseng, H. J. (1994b). Strand- and sequence-specific attenuation of *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine-induced G·C to A·T transitions by expression of human O^6-methylguanine-DNA methyltransferase in Chinese hamster ovary cells. Cancer Res. *54*, 3857-3863.

Yang, J. L., Hu, M. C. and Wu, C. W. (1991). Novel mutational spectrum induced by *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine in the coding region of the hypoxanthine (guanine) phosphoribosyltransferase gene in diploid human fibroblasts. J. Mol. Biol. *221*, 421-430.

Yang, J. L., Lee, P. C., Lin, S. R. and Lin, J. G. (1994a). Comparison of mutation spectra induced by *N*-ethyl-*N*-nitrosourea in the *hprt* gene of Mer^+ and Mer^- diploid human fibroblasts. Carcinogenesis *15*, 939-945.

Yao, M. and Kow, Y. W. (1994). Strand-specific cleavage of mismatch-containing DNA by deoxyinosine 3'-endonuclease from *Escherichia coli*. J. Biol. Chem. *269*, 31390-31396.

Zak, P., Kleibl, K. and Laval, F. (1994). Repair of O^6-methylguanine and O^4-methylthymine by the human and rat O^6-methylguanine-DNA methyltransferases. J. Biol. Chem. *269*, 730-733.

Zarbl, H., Sukumar, S., Arthur, A. V., Martin-Zanca, D. and Barbacid, M. (1985). Direct mutagenesis of Ha-*ras*-1 oncogenes by *N*-nitroso-*N*-methylurea during initiation of mammary carcinogenesis in rats. Nature *315*, 382-385.

Zhan, D. J., Heflich, R. H. and Fu, P. P. (1996). Molecular characterization of mutation and comparison of mutation profiles in the *hprt* gene of Chinese hamster ovary cells treated with benzo[*a*]pyrene *trans*-7,8-diol-*anti*-9,10-epoxide, 1-nitrobenzo[*a*]pyrene *trans*-7,8-diol-*anti*-9,10-epoxide, and 3-nitrobenzo[*a*]pyrene *trans*-7,8-diol-*anti*-9,10-epoxide. Environ. Mol. Mutagen. *27*, 19-29.

DISCLAIMER NOTICE

Page 178 removed by request of the author and with the approval of the Office of Graduate Education.