6-16-2010

Biochemical Characterization of DNA Glycosylases from Mycobacterium Tuberculosis

Yin Guo
University of Vermont

BIOCHEMICAL CHARACTERIZATION OF DNA GLYCOSYLASES FROM *MYCOBACTERIUM TUBERCULOSIS*

A Dissertation Presented

by

Yin Guo

to

The Faculty of the Graduate College

of

The University of Vermont

In Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
Specializing in Microbiology and Molecular Genetics

January, 2010

Accepted by the Faculty of the Graduate College, The University of Vermont, in partial fulfillment of the requirements for the degree of Doctor of Philosophy, specializing in Microbiology and Molecular Genetics.

Thesis Examination Committee:

Susan S. Wallace, Ph.D. **Advisor**

Gregory M. Gilmartin, Ph.D.

Jeffrey P. Bond, Ph.D.

Sylvie Doublié, Ph.D.

Scott W. Morrical, Ph. D. **Chairperson**

Patricia A. Stokowski, Ph.D. **Associate Dean, Graduate College**

Date: November 20, 2009

ABSTRACT

The DNA glycosylases function in the first step of the base excision repair (BER) process, that is responsible for removing base lesions resulting from oxidation, alkylation or deamination. The DNA glycosylases that recognize oxidative base damage fall into two general families: the Fpg/Nei family and the Nth superfamily. Based on protein sequence alignments, we identified four putative Fpg/Nei family members as well as a putative Nth protein in *Mycobacterium tuberculosis* H37Rv, the causative agent of tuberculosis. While Fpg proteins are widely distributed among the bacteria and plants, Nei homologs are sparsely distributed across phyla, and are only found in γ-proteobacteria, actinobacteria and metazoans. Interestingly, *M. tuberculosis* H37Rv harbors two proteins (*Rv2464c* and *Rv3297*) from the Nei clade and two (*Rv2924c* and *Rv0944*) from the Fpg clade. All four Fpg/Nei proteins were successfully overexpressed by using a novel bicistronic vector, which theoretically prevented stable mRNA secondary structure(s) surrounding the translation initiation region (TIR) thereby improving translation efficiency. Additionally, MtuNth (*Rv3674c*) was also overexpressed in soluble form. The substrate specificities of the purified enzymes were characterized *in vitro* with oligonucleotide substrates containing single lesions. Some were further characterized by gas chromatography/mass spectrometry (GC/MS) analysis of products released from γ-irradiated DNA. MtuFpg1 (*Rv2924c*) has a substrate specificity similar to that of EcoFpg and recognizes oxidized purines. Both EcoFpg and MtuFpg1 are more efficient at removing spiroiminodihydantoin (Sp) than 7,8-dihydro-8-oxoguanine (8-oxoG); however, MtuFpg1 has a substantially increased opposite base discrimination compared to EcoFpg. The *Rv0944* gene encodes MtuFpg2, which contains only the C-terminal domain of an Fpg protein and has no detectable DNA binding activity or DNA glycosylase/lyase activity and thus appears to be a pseudogene. MtuNei1 (*Rv2464c*) recognizes oxidized pyrimidines not only on double-stranded DNA but also on single-stranded DNA. It also exhibits uracil DNA glycosylase activity as well as weak activity on FapyA and FapyG. MtuNth recognizes a variety of oxidized bases, such as urea, 5,6-dihydrouracil (DHU), 5-hydroxyuracil (5-OHU), 5-hydroxycytosine (5-OHC) and methylhydantoin (MeHyd) as well as FapyA, FapyG and 8-oxoadenine (8-oxoA). Both MtuNei1 and MtuNth excise thymine glycol (Tg); however, MtuNei1 strongly prefers the (5*R*) isomers of Tg, whereas MtuNth recognizes only the (5*S*) isomers. The other Nei paralog, MtuNei2 (*Rv3297*), did not demonstrate activity *in vitro* as a recombinant protein, but when expressed in *Escherichia coli*, the protein decreased the spontaneous mutation frequency of both the *fpg mutY nei* triple and *nei nth* double mutants, suggesting that MtuNei2 is functionally active *in vivo* recognizing both guanine and cytosine oxidation products. The kinetic parameters of the MtuFpg1, MtuNei1 and MtuNth proteins on selected substrates were also determined and compared to those of their *E. coli* homologs. Since pathogenic bacteria are often exposed to an oxidative environment, such as in macrophages, our data, together with previous observations, support the idea that the BER pathway is of importance in protecting *M. tuberculosis* against oxidative stress, as has been observed with other pathogens .

CITATION

Material from this dissertation has been published in the following form:

Guo, Y., Wallace, S. S. and Bandaru, V.. (2009). A novel bicistronic vector for overexpressing *Mycobacterium tuberculosis* proteins in *Escherichia coli.* Protein Expression and Purification, 65 (2): 230-237.

Material from this dissertation has been accepted for publication to DNA Repair on November 7th, 2009 in the following form:

Guo, Y., Bandaru V., Jaruga P., Zhao X., Burrows C. J., Iwai S., Dizdaroglu M., Bond J. P. and Wallace, S. S.. The oxidative DNA glycosylases of *Mycobacterium tuberculosis* exhibit different substrate preferences from their *Escherichia coli* counterparts. DNA Repair.

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to my advisor, Dr. Susan Wallace, for her constructive suggestions, critical comments and stimulating discussions regarding my research. I appreciate her for being supportive, positive and considerate over the years as a trusted mentor. I admire and respect her for being creative, assiduous and enthusiastic as a successful scientist in the field of DNA repair.

I would like to thank the other committee members: Drs. Gregory Gilmartin, Scott Morrical, Sylvie Doublié, and especially Dr. Jeffrey Bond for their tremendous help and suggestions. I also would like to thank Drs. Joyce Heckman and David Pederson for helping me with my courses and job hunting. I here acknowledge all the former and current members of the Wallace lab, especially Jeffrey Blaisdell, Drs. Viswanath Bandaru, Scott Kathe, Takashi Watanabe, Stephanie Duclos and Matthew Hogg for offering me generous help. They are not only my labmates but also my good friends.

I am deeply grateful to my parents. Although they are not with me in the United States, their continuous support and encouragement make me strong, confident and optimistic all the time. I am also thankful to all my best friends, especially Sandhya and Fabiola for helping me through the last period of my graduate study. Last of all, I feel so lucky to have my husband Songtao with me. His love is the most precious achievement of my lifetime.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

ARTICLE 1

ARTICLE 2

INTRODUCTION

DNA is under constant challenge by endogenous and exogenous stress, which results in both transient and accumulated DNA damages, mutations, cancer, and/or cell death. Fortunately, cells are equipped with numerous mechanisms that trigger DNA repair and/or DNA damage tolerance, arrest the cell cycle, and if necessary, initiate apoptosis to eliminate damaged cells. The proper functions of these mechanisms are critical for maintaining the integrity of the genome.

I. DNA Damage

1. Endogenous DNA Damage

1.1 Hydrolytic Damage

Endogenous DNA lesions are mainly hydrolytic and oxidative damages since DNA is surrounded by water and oxygen *in vivo*. Uracil (U) is normally confined to RNA, but it can arise in DNA by replication error and by spontaneous deamination of cytosine (Duncan & Weiss, 1982; Mosbaugh, 1988; Tye *et al.*, 1978). The deamination of cytosine can also be enhanced by UV radiation-induced cyclobutane pyrimidine dimers (CPD), certain intercalating agents or by the positioning of a mismatched or alkylated base opposite cytosine (Friedberg *et al.*, 2006). *Escherichia coli* defective in the removal of uracil from DNA have an increased spontaneous mutation rate, mainly G:C to A:T transitions (Duncan & Weiss, 1982). Uracil in DNA is subject to removal by uracil DNA glycosylase (UDG or Ung)-initiated base excision repair (BER). Spontaneous deamination of 5-methylcytosine (5-meC) generates thymine (T) leading

to T:G mispairs, which is a major cause of mutations in eukaryotes (Ehrlich *et al.*, 1990). Adenine (A) and guanine (G) can also deaminate albeit with a lower rate creating hypoxanthine and xanthine, respectively. Hypoxanthine is potentially mutagenic mispairing with C during DNA replication [for a review see (Lindahl, 1979)]. Xanthine, on the other hand, is only able to pair with C or T unstably, which may arrest DNA synthesis (Greer & Zamenhof, 1962).

Furthermore, the N-glycosylic bond of DNA, particularly with purine residues, is susceptible to spontaneous hydrolysis resulting in an apurinic/apyrimidinic site (AP site) [for a review see (Lindahl & Barnes, 2000)]. Alternatively, during BER, DNA glycosylases remove damaged bases from DNA also giving rise to AP sites (Goodman *et al.*, 1993; Loeb & Preston, 1986). AP sites can also be generated by oxidation, UV exposure, ionizing radiation, as well as exposure to some chemicals. AP sites strongly block DNA polymerases and are therefore cytotoxic (Boiteux & Laval, 1982; Evans *et al.*, 1993; Hevroni & Livneh, 1988; Kunkel *et al.*, 1983; Laspia & Wallace, 1988; Moran & Wallace, 1985; Sagher & Strauss, 1983; Schaaper & Loeb, 1981; Schaaper *et al.*, 1983). AP sites can also be mutagenic due to their propensity to mispair A during translesion synthesis (Cai *et al.*, 1993; Kunkel, 1984; Lawrence *et al.*, 1990; Neto *et al.*, 1992; Randall *et al.*, 1987; Schaaper & Loeb, 1981; Schaaper *et al.*, 1983).

1.2 Oxidative Damage

Reactive oxygen species (ROS), such as superoxide, hydrogen peroxide (H_2O_2), and hydroxyl radicals, which are generated during normal cellular metabolism and by

external sources, such as UV light, ionizing radiation and some chemicals, can lead to various types of DNA damage, including oxidized or ring-opened base lesions, abasic sites and strand breaks [for reviews see (Boiteux & Radicella, 1999; Dizdaroglu, 1991; Wallace, 1998; Wallace, 2002)]. Common oxidative base lesions are shown in Figure 1. Oxidative damage is generally repaired by BER.

Guanine (G) has low redox potential making it particularly vulnerable to oxidation [for reviews see (David *et al.*, 2007; Neeley & Essigmann, 2006)]. One of the most abundant oxidized base lesions is 7, 8-dihydro-8-oxo-2'-deoxyguanosine (8-oxoG) [Figure 1]. 8-oxoG is a major form of mutagenic base damage, which can mispair with noncognate adenine (A) during replication, giving rise to G to thymine (T) transversions (Cheng *et al.*, 1992; Demple & Harrison, 1994; Kouchakdjian *et al.*, 1991; McAuley-Hecht *et al.*, 1994; Moriya *et al.*, 1991; Moriya, 1993; Moriya & Grollman, 1993; Oda *et al.*, 1991; Shibutani *et al.*, 1991; Wood *et al.*, 1990). Further oxidization of 8-oxoG generates two hydantoin lesions: spiroiminodihydantoin (Sp) and guanidinohydantoin (Gh), as well as urea [Figure 1] (Hailer *et al.*, 2005a; Hailer *et al.*, 2005b; Luo *et al.*, 2001; Misiaszek *et al.*, 2004; Slade *et al.*, 2005; Sugden *et al.*, 2001), which pair with A or G *in vitro* and are highly mutagenic *in vivo* causing G to T and G to C transversions (Burrows *et al.*, 2002; Delaney *et al.*, 2007; Henderson *et al.*, 2003; Kornyushyna *et al.*, 2002; Kornyushyna & Burrows, 2003). A is also susceptible to oxidation giving rise to 7,8-dihydro-8-oxoadenine (8-oxoA) (Maccabee *et al.*, 1994; Malins & Haimanot, 1990; Stillwell *et al.*, 1989; Van Hemmen & Bleichrodt, 1971). Although 8-oxoA pairs primarily with its cognate base T (Guschlbauer *et al.*, 1991;

Wood *et al.*, 1992), it is weakly mutagenic mispairing with G or A (Kamiya *et al.*, 1995a; Shibutani *et al.*, 1993). 2-Hydroxyadenine (2-oxoA) [shown in Figure 1] is also mutagenic (Kamiya & Kasai, 1995; Kamiya *et al.*, 1995b; Kamiya & Kasai, 1996; Kamiya & Kasai, 1997a; Kamiya & Kasai, 1997b; Mori *et al.*, 1993; Nackerdien *et al.*, 1991; Olinski *et al.*, 1992). The ring opened purine derivatives, 2,6-diamino-4-hydroxy-5-formamidopyrimidine (FapyG) and 4,6-diamino-5-formamidopyrimidine (FapyA) [Figure 1] are generated by H_2O_2 treatment, as well as ionizing radiation, UV radiation or photosensitization [for a review see (Dizdaroglu *et al.*, 2008)]. Both damages are potentially mutagenic. The alkylated form of FapyG, 2,6-diamino-4-hydroxy-5*N*-methylformamidopyrimidine (Me-FapyG), blocks replication hence is lethal to cells (Boiteux & Laval, 1983; O'Connor *et al.*, 1988; Tudek *et al.*, 1992). 4,6-Diamino-5*N*-methylformamidopyrimidine (Me-FapyA), however, can be potentially mutagenic mispairing with C (Tudek *et al.*, 1999). Of the purine lesions studied to date, 8-oxoG appears to be responsible for the majority of the oxidative mutational events.

One of the most prevalent damages to DNA is 5,6 dihydroxy-5,6 dihydrothymine or thymine glycol (Tg) [Figure 1] resulting mainly from chemical oxidation or ionizing radiation (Breimer & Lindahl, 1985; Frenkel *et al.*, 1981; Teoule, 1987). Cellular Tg can also be generated by oxidation of deaminated 5-meC. Tg strongly blocks replicative DNA polymerases and is thus a lethal lesion (Clark & Beardsley, 1986; Hayes & LeClerc, 1986; Ide *et al.*, 1985; Laspia & Wallace, 1988; Moran & Wallace, 1985; Rouet & Essigmann, 1985). Depending on which DNA polymerases it deals with, Tg can also be weakly mutagenic with a low level of

mispairing with G (Basu *et al.*, 1989; Clark & Beardsley, 1989; Hayes & LeClerc, 1986). A ring-fragmentation product of oxidized T, urea, is also a strong block to DNA polymerases *in vitro* (Ide *et al.*, 1985). Other radiolysis products of T, 5-hydroxymethyluracil (HMU) (Frenkel *et al.*, 1985; Hagen, 1986; Teoule, 1987) and 5-formyluracil (FU) (Kasai *et al.*, 1990) can be readily bypassed during DNA synthesis (Ide *et al.*, 1991; Levy & Teebor, 1991; Masaoka *et al.*, 1999; Miyabe *et al.*, 2001; Zhang *et al.*, 1997). Misincorporation of G and C opposite FU has also been observed (Masaoka *et al.*, 1999; Miyabe *et al.*, 2001; Zhang *et al.*, 1997). Another ring saturation product of T, 5,6-dihydrothymine (DHT) can be bypassed by DNA polymerases and always pairs with cognate A thus is neither lethal nor mutagenic (Evans *et al.*, 1993; Ide *et al.*, 1991; Maccabee *et al.*, 1994). Taken together, oxidized products of T are not major sources of mutations. The ring saturation and oxidation derivatives of cytosine glycol: uracil glycol (Ug) and 5,6-dihydrouracil (DHU), as well as 5-hydroxycytosine (5-OHC) and 5-hydroxyuracil (5-OHU) are not cytotoxic (Purmal *et al.*, 1994a; Wagner *et al.*, 1992). However, they are all potentially mutagenic since they can mispair with A causing C to T transitions (Feig *et al.*, 1994; Kreutzer & Essigmann, 1998; Purmal *et al.*, 1994a; Purmal *et al.*, 1994b; Purmal *et al.*, 1998). C to T transitions are one of the major mutational events found in the oxidative mutational spectra.

1.3 Mismatches

DNA mismatches are most frequently generated by DNA polymerase errors during normal DNA replication as well as during DNA synthesis associated with DNA

repair [for a review see (Kunz *et al.*, 2009)]. Failure to correct mismatches generally gives rise to mutations. The major pathway involved in rectifying mismatches formed during replication is mismatch repair (MMR).

2. Exogenous DNA Damage

2.1 Alkylation Damage

Alkylating agents are electrophilic compounds which can attack nucleophilic centers of DNA resulting in N-alkylated or O-alkylated purines and pyrimidines as well as phosphotriesters [for a review see (Mishina *et al.*, 2006)]. The most abundant alkylation products in double-stranded DNA are 7-methylguanine and 3-methyladenine, both of which block DNA replication and thus are cytotoxic. Other alkylation products, such as O^6-methylguanine and O^4-methylthymine can mispair during DNA replication resulting in mutagenesis. The O^6-alkylated guanine lesions can be directly reversed to guanine; the other alkylated bases are repaired by BER.

2.2 UV Damage

UV-induced DNA damage can result from the direct absorption of photons by DNA. The most frequent DNA lesions formed by UV radiation are cyclobutane pyrimidine dimers (CPD) and pyrimidine-pyrimidone (6-4) photoproducts ((6-4)PP) in which two adjacent pyrimidines are covalently linked. Moreover, UV radiation results in DNA-protein cross-links and occasionally interchain DNA-DNA cross-links. These photoproducts, as well as other minor photoproducts in DNA, can interfere with DNA

replication and transcription. CPD can be directly reversed to the original pyrimidine monomers. Both CPD and 6-4 PP can be excised from DNA by the nucleotide excision repair pathway (NER). UV radiation can also form reactive oxygen species causing oxidative damage [I.1.2].

2.3 Ionizing Radiation-induced Damage

Ionizing radiation, including α-particles, β-particles, X rays and γ-rays, damages DNA by direct desposition of radiation energy (direct effect) and more commonly by ionization of water molecules to produce hydroxyl radicals (indirect effect), thus inducing multiple forms of DNA damage, including base lesions, AP sites, single-strand breaks (SSBs) as well as double-strand breaks (DSBs) [for reviews see (Wallace, 1998; Wallace, 2002)]. The damages generated by radiation-induced hydroxyl radicals overlap considerably with those generated during normal oxidative metabolism [I.1.2], which is efficiently removed by BER. A special feature of ionizing radiation is the formation of clustered damages or multiple damage sites in close proximity on the DNA [for a review see (Leadon, 1996)]. Clustered damages on the opposite strands of DNA can readily cause DSBs, which are among the most cytotoxic and mutagenic lesions. DSBs are repaired by recombinational repair.

II. DNA Repair

DNA repair can occur by one of two fundamental mechanisms that involves either the reversal of DNA damage or the excision of the damaged elements. With regard to excision repair, three basic repair pathways have evolved to prevent the cytotoxic or mutagenic outcomes of DNA damage thus maintaining genomic integrity: the base excision repair pathway (BER) that predominantly copes with small lesions resulting from oxidation, alkylation or deamination, mismatch repair (MMR) that corrects mismatches as well as insertion/deletion loops formed during replication, and nucleotide excision repair (NER) that removes bulky helix-distorting lesions. In addition, recombinational repair is crucial for repairing DSBs. Moreover, in some bacteria, the physiological responses to DNA damage are under the control of an SOS regulatory network, which leads to not only DNA repair but also to damage tolerance and mutagenesis.

1. Direct Reversal of Base Damage

Direct reversal is the simplest and the most efficient way for repairing damaged DNA. The adaptive response of *Escherichia coli* to alkylation damage involves four genes: *ada*, *alkA*, *alkB* and *aidB* [for reviews see (Nieminuszczy & Grzesiuk, 2007; Sedgwick & Vaughan, 1991; Sedgwick & Lindahl, 2002)]. Among them, Ada and AlkB remove alkylation damage in a single step. *E. coli* Ada protein catalyzes the transfer of the alkyl group from O^6-alkylguanine and O^4-alkylthymine as well as methylphosphotriesters to the active cysteine [for a review see((Mishina *et al.*, 2006)].

Alkylated Ada then becomes a transcriptional activator and induces the transcription of *ada-alkB* operon, *alkA* and *aidB* genes. AlkB has been shown to be an α-ketoglutarate-Fe(II)-dependent dioxygenase, and repairs 1-methyladenine and 3-methylcytosine in DNA by oxidative demethylation (Dinglay *et al.*, 2000; Trewick *et al.*, 2002). AlkA is a DNA glycosylase that repairs alkylated base damages through the BER pathway, which will be discussed later. Moreover, *E. coli* has a second O^6-meG-DNA methyltransferase (Ogt) and a second 3-meA-DNA glycosylase (Tag) which are expressed constitutively for repairing alkylation lesions [for a review see (Samson, 1992)].

UV-induced CPD can also be directly reversed to the original pyrimidine monomers by a CPD-photolyase in a light-dependent manner, a so-called photoreactivation [for reviews see (Goosen & Moolenaar, 2008; Kim & Sancar, 1993)]. Similarly, another redundant UV-induced DNA lesion, (6-4)PP, can be restored to its native monomers by (6-4)PP-photolyase, which has been identified in the eukaryotes [for a review see (Goosen & Moolenaar, 2008)].

2. Base Excision Repair

The BER pathway is highly conserved from bacteria to humans and is predominantly responsible for removing endogenous DNA damage, including oxidative damage [for reviews see (Wallace, 1998; Wallace, 2002; Wallace *et al.*, 2003)]. During BER [Figure 2], oxidative base damages are first recognized by DNA glycosylases, which excise the damaged base by cleaving the N-glycosylic bond between the damaged base and its deoxyribose sugar resulting in an AP site. Oxidative DNA

glycosylases are bifunctional and exhibit lyase activity in addition to DNA glycosylase activity, which further incises the phosphodiester backbone of DNA on the 3' side of the AP site leaving either an α, β-unsaturated aldehyde via β-elimination or a phosphate via β,δ-elimination. The phosphodiesterase activity of an AP endonuclease, Xth or Nfo in *E. coli*, either removes the 3' blocking group at the strand break created by a bifunctional DNA glycosylase or nicks an AP site generated by a monofunctional glycosylase both leaving a 3' hydroxyl group for DNA polymerization. *E. coli* Pol I extends the 3' hydroxyl end by one nucleotide and the resulting nick is sealed by bacterial LigI. *E. coli* cells that are totally devoid of BER have been shown to be extremely sensitive to ionizing radiation (Chen *et al.*, 1993; Cunningham *et al.*, 1986).

3. Mismatch Repair

DNA MMR is highly conserved from prokaryotes to eukaryotes. This pathway removes mispaired bases and insertion/deletion loops (IDL) that arise during DNA replication, as well as heteroduplexes created during a recombinational process or by DNA damage. In *E. coli,* MMR is highly biased to the unmethylated, nascent DNA strand with the methylated strand serving as the template for correction (Kramer *et al.*, 1984; Lu *et al.*, 1983). Deficiencies in MMR result in high spontaneous mutation rates and microsatellite repeat instability (Tiraby & Fox, 1973).

MutS, as a homodimer, initiates MMR by recognizing and binding to DNA containing a mismatch or IDL (Junop *et al.*, 2001; Wu & Marinus, 1994). The binding of MutS is necessary to recruit MutL to DNA in an ATP-dependent fashion (Ban &

Yang, 1998b; Haber & Walker, 1991; Spampinato & Modrich, 2000; Wu & Marinus, 1994). MutL interacts physically with MutS, enhances mismatch recognition, and recruits and activates the endonuclease activity of MutH (Ban & Yang, 1998b; Hall & Matson, 1999; Junop *et al.*, 2001; Spampinato & Modrich, 2000). The endonuclease activity of MutH selectively nicks the unmethylated, nascent DNA 5' to a GATC segment which is located on either side of the mismatch (Ban & Yang, 1998a; Lee *et al.*, 2005). Subsequently, MutS and MutL help load UvrD (DNA helicase II) and single-strand binding protein (SSB) to unwind the DNA towards the mismatch (Dao & Modrich, 1998; Guarne *et al.*, 2004). Depending on the polarity of the unmodified strand, either a 3' to 5' exonuclease (ExoI or ExoX) or a 5' to 3' exonuclease (RecJ or ExoVII) removes the stretch of DNA containing the mismatch (Burdett *et al.*, 2001; Viswanathan & Lovett, 1999). Finally, repair synthesis by Pol III followed by ligation completes the repair process (Li *et al.*, 2008).

4. Nucleotide Excision Repair

NER is one of the most versatile DNA repair systems and is primarily responsible for repairing a wide variety of bulky, helix-distorting lesions and crosslinks induced by UV radiation and chemicals, such as: CPD and (6-4)PP [for a review see (Pfeifer, 1997)].

NER in cells is classified into two sub pathways: Global excision repair refers to NER that incises on either side of the damaged nucleotides from transcriptionally silent DNA or from the non-transcribed strand [for a review see (Truglio *et al.*, 2006)],

Transcription-coupled NER (TC-NER) is specifically associated with arrested or stalled transcription removing the damages in the transcribed strand that block the RNA polymerase [for reviews see (Mellon & Hanawalt, 1989; Savery, 2007)]. UvrA, UvrB, UvrC and UvrD are the key players during NER in *E. coli* and interact in a sequential fashion [for reviews see (Howard-Flanders *et al.*, 1966; van der Kemp *et al.*, 1996; Van Houten, 1990)]. UvrA homodimer is a DNA-independent ATPase and DNA binding protein (Mazur & Grossman, 1991; Sancar & Rupp, 1983), which helps UvrB load onto and track along DNA and locate the damage in an ATP-dependent manner (DellaVecchia *et al.*, 2004; Goosen & Moolenaar, 2001; Moolenaar *et al.*, 2000; Van Houten & Snowden, 1993). Once the damage is encountered, UvrA dissociates leaving a stable UvrB-DNA pre-incision complex (Orren & Sancar, 1990). Therefore, the specific interaction between UvrA homodimers and UvrB as well as their interaction with damaged DNA is important for recognizing and locating the damage site during NER. UvrB-DNA complex facilitates the DNA unwinding and the distortion (Theis *et al.*, 2000) which then guides the endonuclease UvrC to incise the damaged DNA strand both 5' and 3' to the damage (Verhoeven *et al.*, 2000). UvrD, also known as DNA helicase II, functions as a dimer in order to unwind the DNA, release the damage-containing DNA fragment and dissociate the UvrC protein. The repair is completed by DNA polymerase I and DNA ligase. During TC-NER, transcription repair coupling factor (TRCF, also known as Mfd) plays an important role in binding to the stalled RNA polymerase and displacing the enzyme and the truncated transcript, which results in the association of the $(UvrA)_2UvrB$ complex to the damage site initiating NER

(Assenmacher *et al.*, 2006; Deaconescu *et al.*, 2006; Selby *et al.*, 1991; Selby & Sancar, 1993; Selby & Sancar, 1994; Svejstrup, 2002).

5. Recombinational Repair

DSBs can be generated during replication and meiosis or caused directly by ionizing radiation. They are one of the most cytotoxic damages. There are two principal mechanisms responsible for repairing DSBs: homologous recombination (HR) in which the DNA ends are aligned and joined using sequence homology [for a review see (Cromie *et al.*, 2001)] and nonhomologous end joining (NHEJ) in which the broken ends are brought together and rejoined in the absence of long tracts of sequence homology [for reviews see (Krejci *et al.*, 2003; Lees-Miller & Meek, 2003)]. In prokaryotes and lower eukaryotes, there are two HR pathways primarily responsible for repairing DSBs: RecBCD and RecFOR (Kuzminov, 1999). At the beginning, either RecBCD or RecQ (helicase), RecJ (nuclease) and RecFOR (RecA loading) initiates DSBs repair by generating a 3' single-stranded tail at the broken site. The initiation step also involves SSB which helps in creating the single-stranded tail and loading the catalytically competent RecA for the next step. RecA as well as other accessory proteins function in mediating DNA strand exchange and establishing Holliday junctions (Bedale & Cox, 1996; Bianco *et al.*, 1998; Brenner *et al.*, 1988). ATP-dependent branch migration occurs after the Holliday junction is formed and involves a heterodimer of RuvAB. Resolution of the Holliday junction is eventually accomplished by RuvC. The DNA strand exchange that occurs during HR requires the presence of

another DNA duplex sharing extensive sequence homology with the broken DNA. The proteins involved in HR are highly conserved from *E. coli* to humans.

NHEJ occurs in organisms ranging from bacteria to mammals, although HR predominates in prokaryotes and lower eukaryotes (Dudasova *et al.*, 2004). During NHEJ, DSBs can also be resealed by rejoining the free ends without the participation of a homologous partner. It involves processing the two broken ends that are usually noncompatible/nonligatable and joining them directly to each other. Therefore, while HR is intrinsically error-free, NHEJ may readily introduce deleterious sequence alterations, such as, deletions, duplications, insertions and inversions at the junctions where the two incompatible DNA ends are joined. Eukaryotic organisms employ a large number of factors, including Ku70/80 heterodimer and the LXX complex (Ligase IV/XRCC4/XLF), to repair breaks by NHEJ (Bliss & Lane, 1997; Chen *et al.*, 2000; Grawunder *et al.*, 1997; Nick McElhinny *et al.*, 2000; Ramsden & Gellert, 1998), whereas recently-identified bacterial NHEJ is generally a two-component system consisting of Ku and LigD [for reviews see (Aravind & Koonin, 2001; Bowater & Doherty, 2006; Doherty *et al.*, 2001; Weller *et al.*, 2002)]. Bacterial Ku homologs generally function as homodimers preferentially binding to DSBs sites (Aravind & Koonin, 2001; Doherty *et al.*, 2001; Weller *et al.*, 2002) and further recruiting NHEJ repair ligase LigD to the termini of DSBs thereby stimulating its multiple enzymatic activities (Della *et al.*, 2004; Pitcher *et al.*, 2005a; Weller *et al.*, 2002). LigD turns out to be a member of the ATP-dependent DNA ligase family, which contains polymerase activity (Aravind & Koonin, 2001; Pitcher *et al.*, 2007b; Weller *et al.*, 2002),

exonuclease activity (Della *et al.*, 2004) as well as ligase activity (Gong *et al.*, 2004; Pitcher *et al.*, 2005a). LigD can recognize the 5' phosphate and directly mediate the synapsis via a specific loop. After microhomology pairing, the nonextendable 3' termini can be resected by the exonuclease activity of LigD. Resynthesis and resealing of the nicks completes the repair (Della *et al.*, 2004; Pitcher *et al.*, 2005b).

6. SOS Response

RecA and LexA are central components of the bacterial response to DNA damage, termed the SOS response [for a review see (Little & Mount, 1982)]. During the uninduced state, LexA acts as a repressor protein by binding as a homodimer to a specific sequence referred to as the SOS box and inhibiting the transcription of controlled genes, including *recA* and *lexA* itself. Once DNA damage occurs, RecA binds to the region of single-stranded DNA generated after the DNA damage forming RecA-ssDNA nucleoprotein filaments, which in conjunction with LexA stimulates autocatalytic cleavage of LexA, inactivates the repressor thereby inducing the transcription of controlled genes. RecA not only directly participates in recombinational DNA repair, but also regulates the expression of other genes involved in NER and translesion synthesis through SOS response.

III. DNA Glycosylases that Recognize Oxidative Base Damage

1. Substrate Specificities

DNA glycosylases fall into two general families based on structural and sequence homology [for a review see (David & Williams, 1998)]. The Nth superfamily (also called HhH-GPD superfamily) includes endonuclease III (Nth), which removes both oxidized pyrimidines, such as Tg, urea, DHU and formamidopyrimidines: FapyA and FapyG; MutY which removes A mispaired with 8-oxoG and FapyG; Ogg1 whose primary substrate is 8-oxoG and FapyG opposite C; and AlkA which removes alkylated bases. The Fpg/Nei family is defined by formamidopyrimidine DNA glycosylase (Fpg), also called MutM, and endonuclease VIII (Nei). Fpg has a substrate specificity similar to Ogg and removes 8-oxoG and FapyG paired with C as well as FapyA paired with T. The hydantoin lesions, Gh and Sp, are also substrates for Fpg (Hazra *et al.*, 2001; Krishnamurthy *et al.*, 2007; Leipold *et al.*, 2000). In addition to purine lesions, Fpg also recognizes oxidized pyrimidines such as 5-OHC, 5-hydrouracil, and 5,6-dihydroxy-5,6-dihydroxythymine but to a much less extent (D'Ham *et al.*, 1999; Hatahet *et al.*, 1994; Jurado *et al.*, 1998). *E. coli* Nei has a substrate specificity similar to that of Nth, removing oxidized pyrimidines and FapyA. Gh and Sp lesions are also removed by endonuclease VIII (Nei) (Hailer *et al.*, 2005a; Hazra *et al.*, 2001). With the exception of MutY and AlkA, which are monofunctional, the other glycosylases listed above are bifunctional and exhibit lyase activity in addition to glycosylase activity. These DNA glycosylases exhibit redundancy with respect to substrate specificities, therefore the

loss of a single DNA glycosylase usually does not result in a complete loss of repair for a specific oxidative lesion.

2. Mutant Phenotypes of *E. coli* deficient in *fpg*, *nei*, and/or *nth*

Nearly all organisms possess an elaborate defense system, called the 'GO system', that confers protection against the mutagenic effects of 8-oxoG. In *E. coli*, the 'GO system' comprises three components (Michaels & Miller, 1992). Fpg effects the removal of 8-oxoG from 8-oxoG:C mispairs in the DNA duplex (Castaing *et al.*, 1993; Chung *et al.*, 1991; Tchou *et al.*, 1991; Tchou *et al.*, 1994). MutY catalyzes the excision of A when mispaired with 8-oxoG, thereby recreating the 8-oxoG:C substrate for Fpg (Au *et al.*, 1989; Lu & Chang, 1988; Tsai-Wu *et al.*, 1992). MutT, an 8-oxodGTPase, sanitizes the nucleotide pool by hydrolyzing 8-oxo-2'- deoxyguanosine-5'triphosphate (8-oxodGTP) to 8-oxodGMP which can no longer be incorporated into DNA (Akiyama *et al.*, 1989; Bhatnagar & Bessman, 1988; Bhatnagar *et al.*, 1991; Maki & Sekiguchi, 1992). While bacteria lacking either the *fpg* gene or the *mutY* gene show a modest increase in spontaneous mutation frequency, *E. coli fpg mutY* double mutants exhibit mutation rates dramatically higher than the sum of the rates for each single mutant (Cabrera *et al.*, 1988; Michaels *et al.*, 1991; Michaels *et al.*, 1992; Nghiem *et al.*, 1988; Radicella *et al.*, 1988). The increased mutations are mainly G to T transversions (Cabrera *et al.*, 1988; Michaels *et al.*, 1991; Michaels *et al.*, 1992; Nghiem *et al.*, 1988; Radicella *et al.*, 1988). *E. coli mutT* mutants exhibit a strong mutator phenotype leading to A to C transversions (Akiyama *et al.*, 1987; Akiyama *et*

al., 1989; Schaaper & Dunn, 1987). *E. coli mutM mutY mutT* triple knockouts display a huge increase in mutation rate compared to wild type (Michaels & Miller, 1992). In eukaryotes, the GO system is constructed along similar lines, having orthologs of MutT and MutY (MTH1 and MYH); however, MutM is replaced by a functionally analogous but structurally unrelated enzyme designated Ogg1 (Boiteux & Radicella, 2000; Boiteux *et al.*, 2002; David & Williams, 1998; Girard *et al.*, 1997; Nash *et al.*, 1996; van der Kemp *et al.*, 1996).

E. coli nei mutants do not exhibit a mutator phenotype (Jiang *et al.*, 1997a; Saito *et al.*, 1997). *E. coli nth* mutants are weak mutators and are slightly sensitive to the lethal effects of H_2O_2 and ionizing radiation (Cunningham & Weiss, 1985; Jiang *et al.*, 1997a; Saito *et al.*, 1997). *E. coli nei nth* double mutants are hypersensitive to the lethal effects of ionizing radiation and H_2O_2 and exhibit a strong mutator phenotype, all C to T transitions (Blaisdell *et al.*, 1999; Jiang *et al.*, 1997a; Saito *et al.*, 1997; Wallace *et al.*, 2003). Therefore, the main biological function of the two *E. coli* enzymes is to remove the products of oxidized pyrimidines, which pair like T. In addition, *E. coli nei fpg mutY* triple mutants and *nth nei fpg mutY* quadruple mutants were shown to display significant synergistic effects in the production of spontaneous mutations all being G to T transversions suggesting that Nei also has activity against oxidized guanine products, such as Sp and Gh (Blaisdell *et al.*, 1999; Hailer *et al.*, 2005b; Hazra *et al.*, 2001; Jiang *et al.*, 1997a).

3. Distribution of Nth, Fpg and Nei

Homologs of Nth are widely distributed across the three major kingdoms, bacteria, archaea and eukaryotes. Two EcoNth orthologs: Ntg1, a mitochondrial enzyme, and Ntg2, present in the nucleus, have been found in *Saccharomyces cerevisiae* (Augeri *et al.*, 1997; Eide *et al.*, 1996; Girard & Boiteux, 1997; You *et al.*, 1999). Nth homologs have also been identified in *Schizosaccharomyces pombe* (Nth-Spo) (Karahalil *et al.*, 1998; Roldan-Arjona *et al.*, 1996) and mammals (NTH1) (Asagoshi *et al.*, 2000a; Asagoshi *et al.*, 2000b; Aspinwall *et al.*, 1997; Dizdaroglu *et al.*, 1999; Eide *et al.*, 2001; Hilbert *et al.*, 1996; Hilbert *et al.*, 1997; Ikeda *et al.*, 1998; Sarker *et al.*, 1998). The Nth orthologs are primarily specific for pyrimidine-derived lesions.

Fpg is distributed across the bacterial kingdom and absent in the archaea and most eukaryotes; however, Fpg orthologs have been characterized in the yeast, *Candida albicans* (CalFpg) and the plant, *Arabidopsis thaliana* (AthFpg) (Gao & Murphy, 2001; Kathe *et al.*, 2009; Ohtsubo *et al.*, 1998) [Figure 3]. Interestingly, both CalFpg and AthFpg strongly prefer FapyG, FapyA, Gh and Sp but not 8-oxoG (Kathe *et al.*, 2009). Interestingly, both organisms also retain Ogg homologs (Murphy & George, 2005). Ogg, as mentioned above, removes 8-oxoG and FapyG and is present mainly in eukaryotes and archaea, although a number of bacterial Oggs have been identified recently (Denver *et al.*, 2003). Unlike human OGG which only removes 8-oxoG when paired with C, *Clostridium acetobutylicum* Ogg showed little opposite specificity when removing 8-oxoG (Robey-Bond *et al.*, 2008).

In contrast to Fpg, Nei is found only in γ-proteobacteria, actinobacteria, and metazoans [Figure 3]. Recently, homologs of EcoNei were discovered in human and mouse, and designated NEIL1, NEIL2 and NEIL3 (Nei-like) (Bandaru *et al.*, 2002; Hazra *et al.*, 2000; Hazra *et al.*, 2002b; Morland *et al.*, 2002; Takao *et al.*, 2002). All three NEIL paralogs have also been identified in the sequenced genomes of other eukaryotic vertebrates. The preferred substrates for NEIL1 and NEIL2 are oxidized pyrimidines as well as Fapy, Gh and Sp with 8-oxoG being a relatively weak substrate (Bandaru *et al.*, 2002; Hailer *et al.*, 2005a; Hazra *et al.*, 2002a; Hazra *et al.*, 2002b; Krishnamurthy *et al.*, 2008; Morland *et al.*, 2002; Rosenquist *et al.*, 2003; Takao *et al.*, 2002; Zhang *et al.*, 2005). Both enzymes showed activity on single-stranded DNA substrates (Bandaru *et al.*, 2002; Dou *et al.*, 2003; Dou *et al.*, 2008; Hailer *et al.*, 2005a; Takao *et al.*, 2002; Zhang *et al.*, 2005). Human NEIL3, however, is much less characterized due to difficulties in expression and purification (Krokeide *et al.*, 2009). NEIL3 was shown to have DNA binding activity and AP lyase activity on single-stranded substrates only (Takao *et al.*, 2009). It also recognizes Gh and Sp on single-stranded substrates demonstrating its role in repairing oxidized damage (Liu, *et al.*, unpublished data). Two human NEIL homologs have been characterized in Mimivirus (MvNei1 and MvNei2), a giant DNA virus that infects Acanthamoeba (Bandaru *et al.*, 2007). Both enzymes share not only sequence homology but also substrate specificity with the human NEILs. MvNei2 preferentially cleaves oxidized pyrimidines in single-stranded DNA forming products with only β-elimination.

NEIL1 is expressed mainly in the liver, pancreas and thymus (Hazra *et al.*, 2002a) and NEIL1 expression is upregulated during S-phase (Hazra *et al.*, 2002b). In contrast, NEIL2 expression is highest in the skeletal muscle and testes and in a cell-cycle independent manner (Hazra *et al.*, 2002b), High expression of NEIL3 was observed in hematopoietic tissue and testes (Morland *et al.*, 2002; Takao *et al.*, 2009).

4. Catalytic Mechanism of DNA Glycosylases

As shown in Figure 4, monofunctional DNA glycosylases catalyze the displacement of the damaged base using an activated water molecule as nucleophile to attack C1' of the target nucleotide thereby cleaving the glycosylic bond between N and C1' and generating an AP site. Bifunctional DNA glycosylases, however, use the amine moiety of the catalytic amino acid (Pro2 for Fpg and Nei proteins (Zharkov *et al.*, 1997); Lys120 for EcoNth and its orthologs (Dodson *et al.*, 1994; Kuo *et al.*, 1992)) to attack the C1', thereby forming an intermediate Schiff base, which can be irreversibly trapped by sodium borohydride reduction (Dodson *et al.*, 1994; Sun *et al.*, 1995). The iminal intermediate further undergoes sequential β-elimination or β,δ-elimination, resulting in complete removal of the damaged base (O'Brien, 2006).

5. Crystal Structures of Fpg, Nei and Nth

Members of the Nth and Fpg/Nei families all consist of two domains, with the active site located in the junction between the domains. A common feature of all these DNA glycosylases is that they bind primarily to the lesion-containing strand and

extrude or flip the damaged base into an extrahelical configuration and into the enzyme active site (Fromme *et al.*, 2004).

Although Fpg and Nei have different substrate specificities, they exhibit significantly similar structures [Figure 5a] (Fromme & Verdine, 2002; Fromme & Verdine, 2003a; Gilboa *et al.*, 2002; Serre *et al.*, 2002; Sugahara *et al.*, 2000; Zharkov *et al.*, 2002; Zharkov *et al.*, 2003). These studies reveal that the Fpg/Nei family members consist of two distinct domains flanking the N-terminal catalytic proline. The N-terminal domain contains an antiparallel β-sandwich, while the C-terminal domain comprises two DNA binding motifs, a helix-two turn-helix (H2TH) motif and a zinc-finger motif. The protein binds mainly the lesion-containing strand through interactions with the phosphate backbone, whereas no direct contacts are made with the backbone of the complementary strand. The binding results in a sharp kink of the DNA helix, which facilitates the damaged base being extruded. The arginine in the conserved insertion loop of Fpg/Nei family [Met74, Arg109 and Phe111; except in EcoNei, the insertion loop is Gln69, Leu70 and Tyr71] has been shown to contribute to the opposite base specificity (Fromme & Verdine, 2003b; Zharkov *et al.*, 2002). In addition, eukaryotic homologs of Fpg/Nei family proteins were identified and exhibit some structural variations. Although the overall structure of the human NEIL1 is similar to that of EcoFpg and EcoNei, it lacks the cysteine residues coordinating Zn^{2+} but shows an unusual "zinc-less finger" motif that functions as a zinc finger (Doublié *et al.*, 2004). Some other Fpg homologs, such as AthFpg, also contain a "zinc-less finger", according to sequence alignments (Kathe *et al.*, 2009) and the recent crystal structure (Duclos *et*

al., unpublished data). NEIL3 has a C-terminal extension and in addition to the two conserved glycosylase domains harbors a RanBP-type zinc finger and a topoisomerase IIIα homology domain (Zharkov, 2008).

The crystal structure of a prokaryotic Nth homolog [Figure 5b] consists entirely of α-helices and connecting loops organized into two well-defined domains: the six-helix barrel domain containing a helix-hairpin-helix (HhH) motif followed by a loop rich in glycine and proline residues with a highly conserved aspartic acid (GPD motif) and the [4Fe-4S] cluster domain (Fromme & Verdine, 2003b). Between the two domains there is a deep groove and a solvent filled pocket with the key catalytic moiety, Lys120, lying at the mouth of the deep solvent-filled pocket. Two residues, Lys120 and Asp138 within the interdomain cleft, which are important for activity, accommodate the extrahelical base and facilitate the catalytic reaction. The HhH motif is located in the interdomain interface and serves to bind the lesion-containing DNA strand on the 3' side of the lesion and presents the catalytic nucleophile to the substrate nucleoside inserted in the extrahelical active site. The [4Fe-4S] cluster of Nth was earlier described as being catalytically inactive and instead stabilizing the protein fold and interacting with the DNA phosphate backbone. However, recently, several studies have shown that binding to DNA by Nth activates the [4Fe-4S] cluster towards oxidation, possibly providing a mechanism for signaling oxidative damage to the repair enzyme at a distance through charge transport along DNA (Rogers *et al.*, 2003; Yavin *et al.*, 2005; Yavin *et al.*, 2006). Unlike the other structurally characterized HhH-GPD

proteins, such as hOGG1 and AlkA, which contact the backbone of only the lesion-containing DNA strand, Nth contacts the backbone of both DNA strands.

IV. DNA Repair in *Mycobacterium tuberculosis*

M. tuberculosis is the causative agent of tuberculosis leading to detrimental morbidity and mortality on a global scale. The G/C content of the *M. tuberculosis* genome is very high, making the pathogen particularly susceptible to guanine and cytosine oxidation and cytosine deamination (O'Sullivan *et al.*, 2005). Oxidative stress plays an important role in the host's innate immune response. Macrophages generate ROS and reactive nitrogen species (RNS) resulting in lethal damage to the DNA of pathogenic microorganisms. However, *M. tuberculosis* manages to survive and replicate in its host's macrophages. Therefore, the ability to repair DNA damage caused by exposure of the microbe to macrophages is likely to play a particularly important role in pathogen proliferation or colonization conferring a virulence advantage for mycobacteria, as has been observed with other pathogens (O'Rourke *et al.*, 2003; Suvarnapunya *et al.*, 2003).

1. Base Excision Repair in *M. tuberculosis*

M. tuberculosis is well equipped to handle the potentially high level of 8-oxoG harboring four *fpg/nei* genes, four *mutT* genes and one *mutY* gene (Cole *et al.*, 2001). *Mycobacterium smegmatis* Fpg was shown to exhibit a substrate specificity similar to that of EcoFpg (Jain *et al.*, 2007). Recently, Olsen *et al.* partially characterized the activity of purified MtuFpg1 and demonstrated its activity as an Fpg homolog (Olsen *et al.*, 2009). But the *fpg* mutant has no apparent mutator phenotype (Dos Vultos *et al.*, 2009), which suggests that there is a redundant activity for repairing 8-oxoG. In

addition, Dos Vultos *et al.* showed that both MutT1 and MutT2 of *M. tuberculosis* displayed 8-oxoGTPase activity, however, only the deficiency of MutT1 in *M. tuberculosis* resulted in a slightly increased mutation frequency. (Dos Vultos *et al.*, 2006). Similarly, mycobacterial *mutY* mutants exhibit only a mildly increased mutation frequency upon H_2O_2 treatment (Dos Vultos *et al.*, 2009). Mycobacterial MutY, unlike EcoMutY which only sees A:8-oxoG, has recently been shown to recognize A, G and T but not C opposite 8-oxoG (Kurthkoti *et al.*, 2009). Moreover, mycobacterial DNA polymerases tend to insert G instead of C or A opposite 8-oxoG, which then leads to G to C but not G to T transversions (Jain *et al.*, 2007). Sidorenko *et al.* have biochemically characterized the partially purified MtuNei1 (*Mtu*-Nei2 in their work) and showed that the enzyme preferred DHU:G but recognized 8-oxoG much less efficiently (Sidorenko *et al.*, 2008).

Like its *E. coli* counterpart, mycobacterial uracil DNA glycosylase (MtuUdg) efficiently excises uracil from DNA (Purnapatre & Varshney, 1998). The enzymatic activity can be reversibly inhibited by its inhibitor Ugi (Acharya *et al.*, 2003; Purnapatre & Varshney, 1998). Recently, another thermo-tolerant Udg (MtuUdgB) was identified in *M. tuberculosis*, which has a broader substrate specificity than MtuUdg and can rescue the C to T mutator phenotype of *E. coli ung* mutants (Sartori *et al.*, 2002; Srinath *et al.*, 2007).

In *M. tuberculosis, ada* and *alkA* are predicted to encode a fusion protein and are part of an operon that includes *ogt* (Cole *et al.*, 1998). Mutants lacking this alkylation damage repair operon are highly sensitive to certain alkylating agents

(Boshoff *et al.*, 2003). Transcription of this operon can be upregulated by UV exposure, mitomycin C and H_2O_2 treatment (Boshoff *et al.*, 2003).

AP sites are repaired by the ubiquitous AP endonucleases, such as XthA and Nfo, which are both found in *M. tuberculosis* (Cole *et al.*, 1998). The biological importance of XthA in *M. tuberculosis* is highlighted by the fact that no variations in *xthA* have been observed in clinical strains (Dos Vultos *et al.*, 2008). Also, using transposon site hybridization (TraSH), endonuclease IV (*nfo*) and exonuclease III (*xthA*) were shown to be important for the *in vivo* growth of *M. tuberculosis* during initial infection (Sassetti & Rubin, 2003). Taken together, the data strongly suggest that efficient repair of abasic sites is physiologically important to *M. tuberculosis*. However, no DNA glycosylases that repair oxidative damage were shown to be necessary for its survival *in vivo* probably due to the redundancy of these enzymes (Rengarajan *et al.*, 2005).

2. Nucleotide Excision Repair in *M. tuberculosis*

M. tuberculosis processes all the genes involved in the NER pathway. The expression of *uvr* genes is enhanced in intracellular *M. tuberculosis* demonstrating the importance of NER for the pathogen's survival upon infection (Graham & Clark-Curtiss, 1999). In addition, these genes are also involved in the response to H_2O_2 (Cabusora *et al.*, 2005) and are upregulated upon UV exposure (Boshoff *et al.*, 2003). The mutants lacking *uvrB* were highly susceptible to acidified nitrite and UV light and were markedly attenuated for survival in mice and bone marrow-derived macrophages

(Darwin & Nathan, 2005). Moreover, *M. tuberculosis* has two putative genes encoding DNA helicase II: *uvrD1* and *uvrD2* (Cole *et al.*, 1998). The Ku-dependent UvrD1 seems to be essential for mycobacterial repair of DNA damage caused by UV and ionizing radiation and is upregulated upon H_2O_2 treatment (Boshoff *et al.*, 2003; Sinha *et al.*, 2007). The absence of UvrD1 attenuates *M. tuberculosis* during infection (Curti *et al.*, 2007). On the other hand, UvrD2 is a Ku-independent helicase, which is also upregulated upon UV radiation (Boshoff *et al.*, 2003; Sinha *et al.*, 2007). The potentially redundant functions of UvrD1 and UvrD2 strongly suggest the important role of NER in *M. tuberculosis*.

3. Recombination Repair in *M. tuberculosis*

The two HR pathways, RecBCD and RecFOR, are present in *M. tuberculosis*. Both provide single-stranded DNA for RecA binding. The expression of *ruvA* and *ruvC* are upregulated upon DNA damage in *M. tuberculosis* (Boshoff *et al.*, 2003). Taken together with previous observations that *ruvC* is important for the survival of *Helicobacter pylori* in macrophages (Loughlin *et al.*, 2003), the data suggest that HR involving *ruvC* may also play an important role in repairing damages in *M. tuberculosis* (Brooks *et al.*, 2001). Although RecJ is missing in *M. tuberculosis*, the expression of mycobacterial *recR* is shown to be upregulated upon capreomycin treatment (Fu & Shinnick, 2007), and both *recR* and *recF* are upregulated upon translational inibition (Boshoff *et al.*, 2004) suggesting that the RecFOR pathway is active in *M. tuberculosis*.

The NHEJ pathway was once thought to be restricted to eukaryotes but was

recently identified in prokaryotes. The process in *M. tuberculosis* is aided by the DNA end binding protein Ku homodimer (Weller *et al.*, 2002) and DNA ligase D (LigD) (Della *et al.*, 2004). The physical interaction between mycobacterial Ku and LigD stimulates the ligase activity of LigD on DSBs (Gong *et al.*, 2005; Shuman & Glickman, 2007; Weller *et al.*, 2002). Deleting Ku or LigD significantly lowers the efficiency of NHEJ-directed DSBs repair in mycobacteria (Gong *et al.*, 2005; Shuman & Glickman, 2007). Moreover, *M. tuberculosis* contains a LigD-independent NHEJ pathway (Gong *et al.*, 2005) in which LigC may function redundantly in NHEJ. The absence of NHEJ sensitizes mycobacteria to desiccation and ionizing radiation (Pitcher *et al.*, 2007c). NHEJ may be the only pathway available to repair DSBs during latency of *M. tuberculosis*, assuming that no daughter chromatid is present for HR. NHEJ may also be important for repairing DSBs induced by clustered oxidative damages within macrophages and promoting the survival of the pathogen (Dos Vultos *et al.*, 2009). Since this pathway can sometimes be mutagenic, it could promote adaptive mutagenesis, serving as a potential selective advantage under certain conditions (Grosset, 2003; Pitcher *et al.*, 2007a; Shuman & Glickman, 2007).

4. SOS Response in *M. tuberculosis*

RecA is highly conserved throughout all bacteria. The mycobacterial RecA was demonstrated to be functional in DNA repair and HR by its ability to fully restore the phenotype of a *M. smegmatis recA* mutant (Frischkorn *et al.*, 1998), but it showed only partial complementation of an *E. coli recA* mutant (Davis *et al.*, 1991). Unlike *E. coli* or

B. subtilis, whose *recA* is expressed from a single promoter (Cheo *et al.*, 1993; Weisemann & Weinstock, 1991), there are two promoters upstream to the mycobacterial *recA* gene (Movahedzadeh *et al.*, 1997b), both of which are inducible by DNA damage, although via different mechanisms. One promoter is regulated classically by the repressor LexA (Davis *et al.*, 2002; Movahedzadeh *et al.*, 1997a), while the other promoter is LexA and RecA independent (RecA-NDp) (Gamulin *et al.*, 2004; Rand *et al.*, 2003). Surprisingly, only a few of the genes under the control of LexA and RecA in *M. tuberculosis* are known to be involved in DNA repair including: *recA*, *lexA*, *ruvA*, *ruvB*, *ruvC* and *dnaE2*. On the other hand, most of the inducible genes with a known or predicted function in DNA repair were fully independent of *recA*, such as *uvrA*, *uvrB*, *nei2*, *xthA*, *uvrD*, *dnaE1* and *ligB* etc (Rand *et al.*, 2003). Some genes appear to be subject to regulation by both recA-dependent and –independent mechanisms, including *recA*, *radA*, *ruvC* and *ssb*. These findings suggest that a novel LexA and RecA independent mechanism regulates the majority of inducible DNA repair genes in *M. tuberculosis* (Gamulin *et al.*, 2004; Rand *et al.*, 2003), even though in most bacterial systems studied so far a RecA-dependent mechanism is the major DNA damage response mechanism.

5. Mismatch Repair in *M. tuberculosis*

The DNA repair mechanisms of *M. tuberculosis*, as well as its response to DNA damages has distinguishing features compared to some other bacteria. Most strikingly, the genome sequences of *M. tuberculosis*, *Mycobacterium leprae*, and *M. smegmatis*

indicate that these microbes lack the MMR homologs (Dos Vultos *et al.*, 2009; Mizrahi & Andersen, 1998). Although MutS1 and MutL were present in the last common ancestor of these species, several parallel losses of these genes have been observed (Eisen & Hanawalt, 1999). The deficiency of MMR may provide some selective advantages and allow adaptation during specific periods of the life cycle (Eisen & Hanawalt, 1999; Mizrahi & Andersen, 1998). This turned out to be true for some other pathogens, such as *H. pylori*, suggesting that the absence of MMR allows for increased mutation frequencies, emergence of antibiotic resistance, fitness for survival and pathogenicity under certain stressful situations (Kang & Blaser, 2006). In contrast, this high genetic diversity is not observed in *M. tuberculosis*. Therefore, the potential advantage of losing MMR in *M. tuberculosis* is still unclear. One possibility is that slow DNA synthesis and high fidelity DNA polymerases may promote DNA stability, thereby reducing genetic diversity that would occur in an MMR-deficient background (Radman, 1998). Alternatively, novel proteins may be involved in MMR in *M. tuberculosis*, which have no sequence homology to known MMR components.

Taken together, among the major DNA repair pathways, the BER and the NER pathways contribute significantly to maintain the genomic integrity of *M. tuberculosis.*

V. Research Aims and Experimental Design

A large number of Fpg/Nei homologs were identified in *Actinobacteria*, including *M. tuberculosis*, which harbors four Fpg/Nei family members and one Nth ortholog designated as MtuFpg1 (*Rv2924c*), MtuFpg2 (*Rv0944*), MtuNei1 (*Rv2464c*), MtuNei2 (*Rv3297*) and MtuNth (*Rv3674c*). It is widely known that the ability to repair DNA damages is of importance in the survival of facultative intracellular parasites in oxidative environments, such as in macrophages. We therefore hypothesize that these mycobacterial DNA glycosylases display different substrate specificities and function in protecting *M. tuberculosis* against oxidative stress, as has been observed with other pathogens. To gain insight into the biochemical functions of these enzymes and to ultimately understand the evolution of the Fpg/Nei DNA glycosylases, we decided to overexpress and characterize all four Fpg/Nei enzymes and the Nth protein.

The high G/C content of the mycobacterial genes as well as the catalytic properties of Fpg/Nei family members make the heterologous expression of these proteins in *E. coli* challenging. To achieve high levels of soluble and active mycobacterial Fpg/Nei homologs in *E. coli*, we designed a bicistronic expression vector, which is able to facilitate translation. The details are described in Article 1. All four Fpg/Nei proteins were successfully overexpressed by using the bicistronic vector. MtuNth was overexpressed in soluble form using a pET28a vector. The purified enzymes were then extensively characterized by *in vitro* and *in vivo* assays as described in Article 2. Briefly, MtuFpg1 has a substrate specificity similar to that of EcoFpg, but MtuFpg1 does not recognize 8-oxoG:A or 8-oxoA:T as well as does EcoFpg. MtuFpg2

contains only the C-terminal domain of an Fpg protein and has no detectable DNA binding activity or DNA glycosylase/lyase activity and thus appears to be a pseudogene. MtuNei1 recognizes oxidized pyrimidines not only in double-stranded but also in single-stranded DNA suggesting that it may function during DNA replication. MtuNth recognizes a variety of oxidized damages, similar to EcoNth. Strikingly, both MtuNei1 and MtuNth excise thymine glycol but with stringent stereoselectivity; MtuNei1 strongly prefers the (5*R*) isomers ((5*R*,6*S*) and (5*R*,6*R*) *cis*-*trans* pair) of Tg, whereas MtuNth recognizes only the (5*S*) isomers ((5*S*,6*R*) and (5*S*,6*S*) *cis*-*trans* pair). Moreover, MtuNei2 is functionally active *in vivo* decreasing the spontaneous mutation frequency of both *E. coli fpg mutY nei* triple and *nei nth* double mutants, as do EcoNei and MtuNei1. Therefore, our data suggest that *M. tuberculosis* Fpg, Nei and Nth orthologs display more divergent substrate specificities than their *E. coli* counterparts. Our data also demonstrate that the BER pathway is of potential importance in protecting *M. tuberculosis* against oxidative stress.

References

Acharya, N., Kumar, P. & Varshney, U. (2003). Complexes of the uracil-DNA glycosylase inhibitor protein, Ugi, with *Mycobacterium smegmatis* and *Mycobacterium tuberculosis* uracil-DNA glycosylases. *Microbiology (Reading, England)* 149, 1647-1658.

Akiyama, M., Horiuchi, T. & Sekiguchi, M. (1987). Molecular cloning and nucleotide sequence of the mutT mutator of *Escherichia coli* that causes A:T to C:G transversion. *Mol Gen Genet* 206, 9-16.

Akiyama, M., Maki, H., Sekiguchi, M. & Horiuchi, T. (1989). A specific role of MutT protein: to prevent dG.dA mispairing in DNA replication. *Proceedings of the National Academy of Sciences of the United States of America* 86, 3949-3952.

Aravind, L. & Koonin, E. V. (2001). Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system. *Genome Res* 11, 1365-1374.

Asagoshi, K., Odawara, H., Nakano, H., Miyano, T., Terato, H., Ohyama, Y., Seki, S. & Ide, H. (2000a). Comparison of substrate specificities of *Escherichia coli* endonuclease III and its mouse homologue (mNTH1) using defined oligonucleotide substrates. *Biochemistry* 39, 11389-11398.

Asagoshi, K., Yamada, T., Okada, Y., Terato, H., Ohyama, Y., Seki, S. & Ide, H. (2000b). Recognition of formamidopyrimidine by *Escherichia coli* and mammalian thymine glycol glycosylases. Distinctive paired base effects and biological and mechanistic implications. *The Journal of biological chemistry* 275, 24781-24786.

Aspinwall, R., Rothwell, D. G., Roldan-Arjona, T. & other authors (1997). Cloning and characterization of a functional human homolog of *Escherichia coli* endonuclease III. *Proceedings of the National Academy of Sciences of the United States of America* 94, 109-114.

Assenmacher, N., Wenig, K., Lammens, A. & Hopfner, K. P. (2006). Structural basis for transcription-coupled repair: the N terminus of Mfd resembles UvrB with degenerate ATPase motifs. *Journal of molecular biology* 355, 675-683.

Au, K. G., Clark, S., Miller, J. H. & Modrich, P. (1989). *Escherichia coli* mutY gene encodes an adenine glycosylase active on G-A mispairs. *Proceedings of the National Academy of Sciences of the United States of America* 86, 8877-8881.

Augeri, L., Lee, Y. M., Barton, A. B. & Doetsch, P. W. (1997). Purification, characterization, gene cloning, and expression of *Saccharomyces cerevisiae*

redoxyendonuclease, a homolog of *Escherichia coli* endonuclease III. *Biochemistry* 36, 721-729.

Ban, C. & Yang, W. (1998a). Crystal structure and ATPase activity of MutL: implications for DNA repair and mutagenesis. *Cell* 95, 541-552.

Ban, C. & Yang, W. (1998b). Structural basis for MutH activation in *E.coli* mismatch repair and relationship of MutH to restriction endonucleases. *The EMBO journal* 17, 1526-1534.

Bandaru, V., Sunkara, S., Wallace, S. S. & Bond, J. P. (2002). A novel human DNA glycosylase that removes oxidative DNA damage and is homologous to *Escherichia coli* endonuclease VIII. *DNA repair* 1, 517-529.

Bandaru, V., Zhao, X., Newton, M. R., Burrows, C. J. & Wallace, S. S. (2007). Human endonuclease VIII-like (NEIL) proteins in the giant DNA Mimivirus. *DNA repair* 6, 1629-1641.

Basu, A. K., Loechler, E. L., Leadon, S. A. & Essigmann, J. M. (1989). Genetic effects of thymine glycol: site-specific mutagenesis and molecular modeling studies. *Proceedings of the National Academy of Sciences of the United States of America* 86, 7677-7681.

Bedale, W. A. & Cox, M. (1996). Evidence for the coupling of ATP hydrolysis to the final (extension) phase of RecA protein-mediated DNA strand exchange. *The Journal of biological chemistry* 271, 5725-5732.

Bhatnagar, S. K. & Bessman, M. J. (1988). Studies on the mutator gene, mutT of *Escherichia coli*. Molecular cloning of the gene, purification of the gene product, and identification of a novel nucleoside triphosphatase. *The Journal of biological chemistry* 263, 8953-8957.

Bhatnagar, S. K., Bullions, L. C. & Bessman, M. J. (1991). Characterization of the mutT nucleoside triphosphatase of *Escherichia coli*. *The Journal of biological chemistry* 266, 9050-9054.

Bianco, P. R., Tracy, R. B. & Kowalczykowski, S. C. (1998). DNA strand exchange proteins: a biochemical and physical comparison. *Front Biosci* 3, D570-603.

Blaisdell, J. O., Hatahet, Z. & Wallace, S. S. (1999). A novel role for *Escherichia coli* endonuclease VIII in prevention of spontaneous G-->T transversions. *Journal of bacteriology* 181, 6396-6402.

Bliss, T. M. & Lane, D. P. (1997). Ku selectively transfers between DNA molecules with homologous ends. *The Journal of biological chemistry* 272, 5765-5773.

Boiteux, S. & Laval, J. (1982). Coding properties of poly(deoxycytidylic acid) templates containing uracil or apyrimidinic sites: *in vitro* modulation of mutagenesis by deoxyribonucleic acid repair enzymes. *Biochemistry* 21, 6746-6751.

Boiteux, S. & Laval, J. (1983). Imidazole open ring 7-methylguanine: an inhibitor of DNA synthesis. *Biochemical and biophysical research communications* 110, 552-558.

Boiteux, S. & Radicella, J. P. (1999). Base excision repair of 8-hydroxyguanine protects DNA from endogenous oxidative stress. *Biochimie* 81, 59-67.

Boiteux, S. & Radicella, J. P. (2000). The human OGG1 gene: structure, functions, and its implication in the process of carcinogenesis. *Archives of biochemistry and biophysics* 377, 1-8.

Boiteux, S., Gellon, L. & Guibourt, N. (2002). Repair of 8-oxoguanine in *Saccharomyces cerevisiae*: interplay of DNA repair and replication mechanisms. *Free radical biology & medicine* 32, 1244-1253.

Boshoff, H. I., Reed, M. B., Barry, C. E., 3rd & Mizrahi, V. (2003). DnaE2 polymerase contributes to *in vivo* survival and the emergence of drug resistance in *Mycobacterium tuberculosis*. *Cell* 113, 183-193.

Boshoff, H. I., Myers, T. G., Copp, B. R., McNeil, M. R., Wilson, M. A. & Barry, C. E., 3rd (2004). The transcriptional responses of *Mycobacterium tuberculosis* to inhibitors of metabolism: novel insights into drug mechanisms of action. *The Journal of biological chemistry* 279, 40174-40184.

Bowater, R. & Doherty, A. J. (2006). Making ends meet: repairing breaks in bacterial DNA by non-homologous end-joining. *PLoS genetics* 2, e8.

Breimer, L. H. & Lindahl, T. (1985). Enzymatic excision of DNA bases damaged by exposure to ionizing radiation or oxidizing agents. *Mutation research* 150, 85-89.

Brenner, S. L., Zlotnick, A. & Griffith, J. D. (1988). RecA protein self-assembly. Multiple discrete aggregation states. *Journal of molecular biology* 204, 959-972.

Brooks, P. C., Movahedzadeh, F. & Davis, E. O. (2001). Identification of some DNA damage-inducible genes of *Mycobacterium tuberculosis*: apparent lack of correlation with LexA binding. *Journal of bacteriology* 183, 4459-4467.

Burdett, V., Baitinger, C., Viswanathan, M., Lovett, S. T. & Modrich, P. (2001). *In vivo* requirement for RecJ, ExoVII, ExoI, and ExoX in methyl-directed mismatch repair. *Proceedings of the National Academy of Sciences of the United States of America* 98, 6765-6770.

Burrows, C. J., Muller, J. G., Kornyushyna, O., Luo, W., Duarte, V., Leipold, M. D. & David, S. S. (2002). Structure and potential mutagenicity of new hydantoin products from guanosine and 8-oxo-7,8-dihydroguanine oxidation by transition metals. *Environmental health perspectives* 110 Suppl 5, 713-717.

Cabrera, M., Nghiem, Y. & Miller, J. H. (1988). mutM, a second mutator locus in *Escherichia coli* that generates G.C----T.A transversions. *Journal of bacteriology* 170, 5405-5407.

Cabusora, L., Sutton, E., Fulmer, A. & Forst, C. V. (2005). Differential network expression during drug and stress response. *Bioinformatics* 21, 2898-2905.

Cai, H., Bloom, L. B., Eritja, R. & Goodman, M. F. (1993). Kinetics of deoxyribonucleotide insertion and extension at abasic template lesions in different sequence contexts using HIV-1 reverse transcriptase. *The Journal of biological chemistry* 268, 23567-23572.

Castaing, B., Geiger, A., Seliger, H., Nehls, P., Laval, J., Zelwer, C. & Boiteux, S. (1993). Cleavage and binding of a DNA fragment containing a single 8-oxoguanine by wild type and mutant FPG proteins. *Nucleic acids research* 21, 2899-2905.

Chen, D. S., Law, C. & Keng, P. (1993). Reduction of radiation cytotoxicity by human apurinic endonuclease in a radiation-sensitive *Escherichia coli* mutant. *Radiat Res* 135, 405-410.

Chen, L., Trujillo, K., Sung, P. & Tomkinson, A. E. (2000). Interactions of the DNA ligase IV-XRCC4 complex with DNA ends and the DNA-dependent protein kinase. *The Journal of biological chemistry* 275, 26196-26205.

Cheng, K. C., Cahill, D. S., Kasai, H., Nishimura, S. & Loeb, L. A. (1992). 8-Hydroxyguanine, an abundant form of oxidative DNA damage, causes G----T and A----C substitutions. *The Journal of biological chemistry* 267, 166-172.

Cheo, D. L., Bayles, K. W. & Yasbin, R. E. (1993). Elucidation of regulatory elements that control damage induction and competence induction of the *Bacillus subtilis* SOS system. *Journal of bacteriology* 175, 5907-5915.

Chung, M. H., Kasai, H., Jones, D. S., Inoue, H., Ishikawa, H., Ohtsuka, E. & Nishimura, S. (1991). An endonuclease activity of *Escherichia coli* that specifically removes 8-hydroxyguanine residues from DNA. *Mutation research* 254, 1-12.

Clark, J. M. & Beardsley, G. P. (1986). Thymine glycol lesions terminate chain elongation by DNA polymerase I *in vitro*. *Nucleic acids research* 14, 737-749.

Clark, J. M. & Beardsley, G. P. (1989). Template length, sequence context, and 3'-5' exonuclease activity modulate replicative bypass of thymine glycol lesions *in vitro*. *Biochemistry* 28, 775-779.

Cole, S. T., Brosch, R., Parkhill, J. & other authors (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393, 537-544.

Cole, S. T., Eiglmeier, K., Parkhill, J. & other authors (2001). Massive gene decay in the leprosy bacillus. *Nature* 409, 1007-1011.

Cromie, G. A., Connelly, J. C. & Leach, D. R. (2001). Recombination at double-strand breaks and DNA ends: conserved mechanisms from phage to humans. *Molecular cell* 8, 1163-1174.

Cunningham, R. P. & Weiss, B. (1985). Endonuclease III (nth) mutants of *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* 82, 474-478.

Cunningham, R. P., Saporito, S. M., Spitzer, S. G. & Weiss, B. (1986). Endonuclease IV (nfo) mutant of *Escherichia coli*. *Journal of bacteriology* 168, 1120-1127.

Curti, E., Smerdon, S. J. & Davis, E. O. (2007). Characterization of the helicase activity and substrate specificity of *Mycobacterium tuberculosis* UvrD. *Journal of bacteriology* 189, 1542-1555.

D'Ham, C., Romieu, A., Jaquinod, M., Gasparutto, D. & Cadet, J. (1999). Excision of 5,6-dihydroxy-5,6-dihydrothymine, 5,6-dihydrothymine, and 5-hydroxycytosine from defined sequence oligonucleotides by *Escherichia coli* endonuclease III and Fpg proteins: kinetic and mechanistic aspects. *Biochemistry* 38, 3335-3344.

Dao, V. & Modrich, P. (1998). Mismatch-, MutS-, MutL-, and helicase II-dependent unwinding from the single-strand break of an incised heteroduplex. *The Journal of biological chemistry* 273, 9202-9207.

Darwin, K. H. & Nathan, C. F. (2005). Role for nucleotide excision repair in virulence of *Mycobacterium tuberculosis*. *Infection and immunity* 73, 4581-4587.

David, S. S. & Williams, S. D. (1998). Chemistry of Glycosylases and Endonucleases Involved in Base-Excision Repair. *Chem Rev* 98, 1221-1262.

David, S. S., O'Shea, V. L. & Kundu, S. (2007). Base-excision repair of oxidative DNA damage. *Nature* 447, 941-950.

Davis, E. O., Sedgwick, S. G. & Colston, M. J. (1991). Novel structure of the recA locus of *Mycobacterium tuberculosis* implies processing of the gene product. *Journal of bacteriology* 173, 5653-5662.

Davis, E. O., Dullaghan, E. M. & Rand, L. (2002). Definition of the mycobacterial SOS box and use to identify LexA-regulated genes in *Mycobacterium tuberculosis*. *Journal of bacteriology* 184, 3287-3295.

Deaconescu, A. M., Chambers, A. L., Smith, A. J., Nickels, B. E., Hochschild, A., Savery, N. J. & Darst, S. A. (2006). Structural basis for bacterial transcription-coupled DNA repair. *Cell* 124, 507-520.

Delaney, S., Neeley, W. L., Delaney, J. C. & Essigmann, J. M. (2007). The substrate specificity of MutY for hyperoxidized guanine lesions *in vivo*. *Biochemistry* 46, 1448-1455.

Della, M., Palmbos, P. L., Tseng, H. M. & other authors (2004). Mycobacterial Ku and ligase proteins constitute a two-component NHEJ repair machine. *Science* (New York, NY 306, 683-685.

DellaVecchia, M. J., Croteau, D. L., Skorvaga, M., Dezhurov, S. V., Lavrik, O. I. & Van Houten, B. (2004). Analyzing the handoff of DNA from UvrA to UvrB utilizing DNA-protein photoaffinity labeling. *The Journal of biological chemistry* 279, 45245-45256.

Demple, B. & Harrison, L. (1994). Repair of oxidative damage to DNA: enzymology and biology. *Annu Rev Biochem* 63, 915-948.

Denver, D. R., Swenson, S. L. & Lynch, M. (2003). An Evolutionary Analysis of the Helix-Hairpin-Helix Superfamily of DNA Repair Glycosylases. *Mol Biol Evol* 20, 1603-1611.

Dinglay, S., Trewick, S. C., Lindahl, T. & Sedgwick, B. (2000). Defective processing of methylated single-stranded DNA by E. coli AlkB mutants. *Genes & development* 14, 2097-2105.

Dizdaroglu, M. (1991). Chemical determination of free radical-induced damage to DNA. *Free radical biology & medicine* 10, 225-242.

Dizdaroglu, M., Karahalil, B., Senturker, S., Buckley, T. J. & Roldan-Arjona, T. (1999). Excision of products of oxidative DNA base damage by human NTH1 protein. *Biochemistry* 38, 243-246.

Dizdaroglu, M., Kirkali, G. & Jaruga, P. (2008). Formamidopyrimidines in DNA: mechanisms of formation, repair, and biological effects. *Free radical biology & medicine* 45, 1610-1621.

Dodson, M. L., Michaels, M. L. & Lloyd, R. S. (1994). Unified catalytic mechanism for DNA glycosylases. *The Journal of biological chemistry* 269, 32709-32712.

Doherty, A. J., Jackson, S. P. & Weller, G. R. (2001). Identification of bacterial homologues of the Ku DNA repair proteins. *FEBS letters* 500, 186-188.

Dos Vultos, T., Blazquez, J., Rauzier, J., Matic, I. & Gicquel, B. (2006). Identification of Nudix hydrolase family members with an antimutator role in *Mycobacterium tuberculosis* and *Mycobacterium smegmatis*. *Journal of bacteriology* 188, 3159-3161.

Dos Vultos, T., Mestre, O., Rauzier, J. & other authors (2008). Evolution and diversity of clonal bacteria: the paradigm of *Mycobacterium tuberculosis*. *PloS one* 3, e1538.

Dos Vultos, T., Mestre, O., Tonjum, T. & Gicquel, B. (2009). DNA repair in *Mycobacterium tuberculosis* revisited. *FEMS microbiology reviews* 33, 471-487.

Dou, H., Mitra, S. & Hazra, T. K. (2003). Repair of oxidized bases in DNA bubble structures by human DNA glycosylases NEIL1 and NEIL2. *The Journal of biological chemistry* 278, 49679-49684.

Dou, H., Theriot, C. A., Das, A., Hegde, M. L., Matsumoto, Y., Boldogh, I., Hazra, T. K., Bhakat, K. K. & Mitra, S. (2008). Interaction of the human DNA glycosylase NEIL1 with proliferating cell nuclear antigen. The potential for replication-associated repair of oxidized bases in mammalian genomes. *The Journal of biological chemistry* 283, 3130-3140.

Doublie, S., Bandaru, V., Bond, J. P. & Wallace, S. S. (2004). The crystal structure of human endonuclease VIII-like 1 (NEIL1) reveals a zincless finger motif required for glycosylase activity. *Proceedings of the National Academy of Sciences of the United States of America* 101, 10284-10289.

Dudasova, Z., Dudas, A. & Chovanec, M. (2004). Non-homologous end-joining factors of *Saccharomyces cerevisiae*. *FEMS microbiology reviews* 28, 581-601.

Duncan, B. K. & Weiss, B. (1982). Specific mutator effects of ung (uracil-DNA glycosylase) mutations in *Escherichia coli*. *Journal of bacteriology* 151, 750-755.

Ehrlich, M., Zhang, X. Y. & Inamdar, N. M. (1990). Spontaneous deamination of cytosine and 5-methylcytosine residues in DNA and replacement of 5-methylcytosine residues with cytosine residues. *Mutation research* 238, 277-286.

Eide, L., Bjoras, M., Pirovano, M., Alseth, I., Berdal, K. G. & Seeberg, E. (1996). Base excision of oxidative purine and pyrimidine DNA damage in *Saccharomyces cerevisiae* by a DNA glycosylase with sequence similarity to endonuclease III from *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* 93, 10735-10740.

Eide, L., Luna, L., Gustad, E. C., Henderson, P. T., Essigmann, J. M., Demple, B. & Seeberg, E. (2001). Human endonuclease III acts preferentially on DNA damage opposite guanine residues in DNA. *Biochemistry* 40, 6653-6659.

Eisen, J. A. & Hanawalt, P. C. (1999). A phylogenomic study of DNA repair genes, proteins, and processes. *Mutation research* 435, 171-213.

Evans, J., Maccabee, M., Hatahet, Z., Courcelle, J., Bockrath, R., Ide, H. & Wallace, S. (1993). Thymine ring saturation and fragmentation products: lesion bypass, misinsertion and implications for mutagenesis. *Mutation research* 299, 147-156.

Feig, D. I., Sowers, L. C. & Loeb, L. A. (1994). Reverse chemical mutagenesis: identification of the mutagenic lesions resulting from reactive oxygen species-mediated damage to DNA. *Proceedings of the National Academy of Sciences of the United States of America* 91, 6609-6613.

Frenkel, K., Goldstein, M. S. & Teebor, G. W. (1981). Identification of the cis-thymine glycol moiety in chemically oxidized and gamma-irradiated deoxyribonucleic acid by high-pressure liquid chromatography analysis. *Biochemistry* 20, 7566-7571.

Frenkel, K., Cummings, A., Solomon, J., Cadet, J., Steinberg, J. J. & Teebor, G. W. (1985). Quantitative determination of the 5-(hydroxymethyl)uracil moiety in the DNA of gamma-irradiated cells. *Biochemistry* 24, 4527-4533.

Friedberg, E. C., Walker, G. C., Siede, W., Wood, R. D., Schultz, R. A. & Ellenberger, T. (2006). *DNA repair and mutagenesis* (second edition): ASM Press.

Frischkorn, K., Sander, P., Scholz, M., Teschner, K., Prammananan, T. & Bottger, E. C. (1998). Investigation of mycobacterial recA function: protein introns in the RecA of

pathogenic mycobacteria do not affect competency for homologous recombination. *Molecular microbiology* 29, 1203-1214.

Fromme, J. C. & Verdine, G. L. (2002). Structural insights into lesion recognition and repair by the bacterial 8-oxoguanine DNA glycosylase MutM. *Nature structural biology* 9, 544-552.

Fromme, J. C. & Verdine, G. L. (2003a). Structure of a trapped endonuclease III-DNA covalent intermediate. *The EMBO journal* 22, 3461-3471.

Fromme, J. C. & Verdine, G. L. (2003b). DNA lesion recognition by the bacterial repair enzyme MutM. *The Journal of biological chemistry* 278, 51543-51548.

Fromme, J. C., Banerjee, A. & Verdine, G. L. (2004). DNA glycosylase recognition and catalysis. *Current opinion in structural biology* 14, 43-49.

Fu, L. M. & Shinnick, T. M. (2007). Genome-wide exploration of the drug action of capreomycin on *Mycobacterium tuberculosis* using Affymetrix oligonucleotide GeneChips. *The Journal of infection* 54, 277-284.

Gamulin, V., Cetkovic, H. & Ahel, I. (2004). Identification of a promoter motif regulating the major DNA damage response mechanism of *Mycobacterium tuberculosis*. *FEMS microbiology letters* 238, 57-63.

Gao, M. J. & Murphy, T. M. (2001). Alternative forms of formamidopyrimidine-DNA glycosylase from *Arabidopsis thaliana*. *Photochemistry and photobiology* 73, 128-134.

Gilboa, R., Zharkov, D. O., Golan, G., Fernandes, A. S., Gerchman, S. E., Matz, E., Kycia, J. H., Grollman, A. P. & Shoham, G. (2002). Structure of formamidopyrimidine-DNA glycosylase covalently complexed to DNA. *The Journal of biological chemistry* 277, 19811-19816.

Girard, P. M. & Boiteux, S. (1997). Repair of oxidized DNA bases in the yeast *Saccharomyces cerevisiae*. *Biochimie* 79, 559-566.

Girard, P. M., Guibourt, N. & 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 research* 25, 3204-3211.

Gong, C., Martins, A., Bongiorno, P., Glickman, M. & Shuman, S. (2004). Biochemical and genetic analysis of the four DNA ligases of mycobacteria. *The Journal of biological chemistry* 279, 20594-20606.

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

Goodman, M. F., Creighton, S., Bloom, L. B. & Petruska, J. (1993). Biochemical basis of DNA replication fidelity. *Critical reviews in biochemistry and molecular biology* 28, 83-126.

Goosen, N. & Moolenaar, G. F. (2001). Role of ATP hydrolysis by UvrA and UvrB during nucleotide excision repair. *Research in microbiology* 152, 401-409.

Goosen, N. & Moolenaar, G. F. (2008). Repair of UV damage in bacteria. *DNA repair* 7, 353-379.

Graham, J. E. & Clark-Curtiss, J. E. (1999). Identification of *Mycobacterium tuberculosis* RNAs synthesized in response to phagocytosis by human macrophages by selective capture of transcribed sequences (SCOTS). *Proceedings of the National Academy of Sciences of the United States of America* 96, 11554-11559.

Grawunder, U., Wilm, M., Wu, X., Kulesza, P., Wilson, T. E., Mann, M. & Lieber, M. R. (1997). Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. *Nature* 388, 492-495.

Greer, S. & Zamenhof, S. (1962). Studies on depurination of DNA by heat. *Journal of molecular biology* 4, 123-141.

Grosset, J. (2003). *Mycobacterium tuberculosis* in the extracellular compartment: an underestimated adversary. *Antimicrobial agents and chemotherapy* 47, 833-836.

Guarne, A., Ramon-Maiques, S., Wolff, E. M., Ghirlando, R., Hu, X., Miller, J. H. & Yang, W. (2004). Structure of the MutL C-terminal domain: a model of intact MutL and its roles in mismatch repair. *The EMBO journal* 23, 4134-4145.

Guschlbauer, W., Duplaa, A. M., Guy, A., Teoule, R. & Fazakerley, G. V. (1991). Structure and *in vitro* replication of DNA templates containing 7,8-dihydro-8-oxoadenine. *Nucleic acids research* 19, 1753-1758.

Haber, L. T. & Walker, G. C. (1991). Altering the conserved nucleotide binding motif in the *Salmonella typhimurium* MutS mismatch repair protein affects both its ATPase and mismatch binding activities. *The EMBO journal* 10, 2707-2715.

Hagen, U. (1986). Current aspects on the radiation induced base damage in DNA. *Radiation and environmental biophysics* 25, 261-271.

Hailer, M. K., Slade, P. G., Martin, B. D., Rosenquist, T. A. & Sugden, K. D. (2005a). Recognition of the oxidized lesions spiroiminodihydantoin and guanidinohydantoin in DNA by the mammalian base excision repair glycosylases NEIL1 and NEIL2. *DNA repair* 4, 41-50.

Hailer, M. K., Slade, P. G., Martin, B. D. & Sugden, K. D. (2005b). Nei deficient *Escherichia coli* are sensitive to chromate and accumulate the oxidized guanine lesion spiroiminodihydantoin. *Chemical research in toxicology* 18, 1378-1383.

Hall, M. C. & Matson, S. W. (1999). The *Escherichia coli* MutL protein physically interacts with MutH and stimulates the MutH-associated endonuclease activity. *The Journal of biological chemistry* 274, 1306-1312.

Hatahet, Z., Kow, Y. W., Purmal, A. A., Cunningham, R. P. & Wallace, S. S. (1994). New substrates for old enzymes. 5-Hydroxy-2'-deoxycytidine and 5-hydroxy-2'-deoxyuridine are substrates for *Escherichia coli* endonuclease III and formamidopyrimidine DNA N-glycosylase, while 5-hydroxy-2'-deoxyuridine is a substrate for uracil DNA N-glycosylase. *The Journal of biological chemistry* 269, 18814-18820.

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

Hazra, T. K., Izumi, T., Venkataraman, R., Kow, Y. W., Dizdaroglu, M. & Mitra, S. (2000). Characterization of a novel 8-oxoguanine-DNA glycosylase activity in *Escherichia coli* and identification of the enzyme as endonuclease VIII. *The Journal of biological chemistry* 275, 27762-27767.

Hazra, T. K., Muller, J. G., Manuel, R. C., Burrows, C. J., Lloyd, R. S. & Mitra, S. (2001). Repair of hydantoins, one electron oxidation product of 8-oxoguanine, by DNA glycosylases of *Escherichia coli*. *Nucleic acids research* 29, 1967-1974.

Hazra, T. K., Izumi, T., Boldogh, I., Imhoff, B., Kow, Y. W., Jaruga, P., Dizdaroglu, M. & Mitra, S. (2002a). Identification and characterization of a human DNA glycosylase for repair of modified bases in oxidatively damaged DNA. *Proceedings of the National Academy of Sciences of the United States of America* 99, 3523-3528.

Hazra, T. K., Kow, Y. W., Hatahet, Z., Imhoff, B., Boldogh, I., Mokkapati, S. K., Mitra, S. & Izumi, T. (2002b). Identification and characterization of a novel human DNA glycosylase for repair of cytosine-derived lesions. *The Journal of biological chemistry* 277, 30417-30420.

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

Hevroni, D. & Livneh, Z. (1988). Bypass and termination at apurinic sites during replication of single-stranded DNA *in vitro*: a model for apurinic site mutagenesis. *Proceedings of the National Academy of Sciences of the United States of America* 85, 5046-5050.

Hilbert, T. P., Boorstein, R. J., Kung, H. C., Bolton, P. H., Xing, D., Cunningham, R. P. & Teebor, G. W. (1996). Purification of a mammalian homologue of *Escherichia coli* endonuclease III: identification of a bovine pyrimidine hydrate-thymine glycol DNAse/AP lyase by irreversible cross linking to a thymine glycol-containing oligoxynucleotide. *Biochemistry* 35, 2505-2511.

Hilbert, T. P., Chaung, W., Boorstein, R. J., Cunningham, R. P. & Teebor, G. W. (1997). Cloning and expression of the cDNA encoding the human homologue of the DNA repair enzyme, *Escherichia coli* endonuclease III. *The Journal of biological chemistry* 272, 6733-6740.

Howard-Flanders, P., Boyce, R. P. & Theriot, L. (1966). Three loci in *Escherichia coli* K-12 that control the excision of pyrimidine dimers and certain other mutagen products from DNA. *Genetics* 53, 1119-1136.

Ide, H., Kow, Y. W. & Wallace, S. S. (1985). Thymine glycols and urea residues in M13 DNA constitute replicative blocks *in vitro*. *Nucleic acids research* 13, 8035-8052.

Ide, H., Petrullo, L. A., Hatahet, Z. & Wallace, S. S. (1991). Processing of DNA base damage by DNA polymerases. Dihydrothymine and beta-ureidoisobutyric acid as models for instructive and noninstructive lesions. *The Journal of biological chemistry* 266, 1469-1477.

Ikeda, S., Biswas, T., Roy, R., Izumi, T., Boldogh, I., Kurosky, A., Sarker, A. H., Seki, S. & Mitra, S. (1998). Purification and characterization of human NTH1, a homolog of *Escherichia coli* endonuclease III. Direct identification of Lys-212 as the active nucleophilic residue. *The Journal of biological chemistry* 273, 21585-21593.

Jain, R., Kumar, P. & Varshney, U. (2007). A distinct role of formamidopyrimidine DNA glycosylase (MutM) in down-regulation of accumulation of G, C mutations and protection against oxidative stress in mycobacteria. *DNA repair* 6, 1774-1785.

Jiang, D., Hatahet, Z., Blaisdell, J. O., Melamede, R. J. & Wallace, S. S. (1997). *Escherichia coli* endonuclease VIII: cloning, sequencing, and overexpression of the nei

structural gene and characterization of nei and nei nth mutants. *Journal of bacteriology* 179, 3773-3782.

Junop, M. S., Obmolova, G., Rausch, K., Hsieh, P. & Yang, W. (2001). Composite active site of an ABC ATPase: MutS uses ATP to verify mismatch recognition and authorize DNA repair. *Molecular cell* 7, 1-12.

Jurado, J., Saparbaev, M., Matray, T. J., Greenberg, M. M. & Laval, J. (1998). The ring fragmentation product of thymidine C5-hydrate when present in DNA is repaired by the *Escherichia coli* Fpg and Nth proteins. *Biochemistry* 37, 7757-7763.

Kamiya, H. & Kasai, H. (1995). Formation of 2-hydroxydeoxyadenosine triphosphate, an oxidatively damaged nucleotide, and its incorporation by DNA polymerases. Steady-state kinetics of the incorporation. *The Journal of biological chemistry* 270, 19446-19450.

Kamiya, H., Miura, H., Murata-Kamiya, N. & other authors (1995a). 8-Hydroxyadenine (7,8-dihydro-8-oxoadenine) induces misincorporation in *in vitro* DNA synthesis and mutations in NIH 3T3 cells. *Nucleic acids research* 23, 2893-2899.

Kamiya, H., Ueda, T., Ohgi, T., Matsukage, A. & Kasai, H. (1995b). Misincorporation of dAMP opposite 2-hydroxyadenine, an oxidative form of adenine. *Nucleic acids research* 23, 761-766.

Kamiya, H. & Kasai, H. (1996). Effect of sequence contexts on misincorporation of nucleotides opposite 2-hydroxyadenine. *FEBS letters* 391, 113-116.

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

Kamiya, H. & Kasai, H. (1997b). Mutations induced by 2-hydroxyadenine on a shuttle vector during leading and lagging strand syntheses in mammalian cells. *Biochemistry* 36, 11125-11130.

Kang, J. & Blaser, M. J. (2006). Bacterial populations as perfect gases: genomic integrity and diversification tensions in *Helicobacter pylori*. *Nat Rev Microbiol* 4, 826-836.

Karahalil, B., Roldan-Arjona, T. & Dizdaroglu, M. (1998). Substrate specificity of *Schizosaccharomyces pombe* Nth protein for products of oxidative DNA damage. *Biochemistry* 37, 590-595.

Kasai, H., Iida, A., Yamaizumi, Z., Nishimura, S. & Tanooka, H. (1990). 5-Formyldeoxyuridine: a new type of DNA damage induced by ionizing radiation and its mutagenicity to salmonella strain TA102. *Mutation research* 243, 249-253.

Kathe, S. D., Barrantes-Reynolds, R., Jaruga, P., Newton, M. R., Burrows, C. J., Bandaru, V., Dizdaroglu, M., Bond, J. P. & Wallace, S. S. (2009). Plant and fungal Fpg homologs are formamidopyrimidine DNA glycosylases but not 8-oxoguanine DNA glycosylases. *DNA repair* 8, 643-653.

Kim, S. T. & Sancar, A. (1993). Photochemistry, photophysics, and mechanism of pyrimidine dimer repair by DNA photolyase. *Photochemistry and photobiology* 57, 895-904.

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

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

Kouchakdjian, M., Bodepudi, V., Shibutani, S., Eisenberg, M., Johnson, F., Grollman, A. P. & Patel, D. J. (1991). NMR structural studies of the ionizing radiation adduct 7-hydro-8-oxodeoxyguanosine (8-oxo-7H-dG) opposite deoxyadenosine in a DNA duplex. 8-Oxo-7H-dG(syn).dA(anti) alignment at lesion site. *Biochemistry* 30, 1403-1412.

Kramer, B., Kramer, W. & Fritz, H. J. (1984). Different base/base mismatches are corrected with different efficiencies by the methyl-directed DNA mismatch-repair system of E. coli. *Cell* 38, 879-887.

Krejci, L., Chen, L., Van Komen, S., Sung, P. & Tomkinson, A. (2003). Mending the break: two DNA double-strand break repair machines in eukaryotes. *Progress in nucleic acid research and molecular biology* 74, 159-201.

Kreutzer, D. A. & Essigmann, J. M. (1998). Oxidized, deaminated cytosines are a source of C --> T transitions *in vivo*. *Proceedings of the National Academy of Sciences of the United States of America* 95, 3578-3582.

Krishnamurthy, N., Muller, J. G., Burrows, C. J. & David, S. S. (2007). Unusual structural features of hydantoin lesions translate into efficient recognition by *Escherichia coli* Fpg. *Biochemistry* 46, 9355-9365.

Krishnamurthy, N., Zhao, X., Burrows, C. J. & David, S. S. (2008). Superior removal of hydantoin lesions relative to other oxidized bases by the human DNA glycosylase hNEIL1. *Biochemistry* 47, 7137-7146.

Krokeide, S. Z., Bolstad, N., Laerdahl, J. K., Bjoras, M. & Luna, L. (2009). Expression and purification of NEIL3, a human DNA glycosylase homolog. *Protein expression and purification* 65, 160-164.

Kunkel, T. A., Schaaper, R. M. & Loeb, L. A. (1983). Depurination-induced infidelity of deoxyribonucleic acid synthesis with purified deoxyribonucleic acid replication proteins *in vitro*. *Biochemistry* 22, 2378-2384.

Kunkel, T. A. (1984). Mutational specificity of depurination. *Proceedings of the National Academy of Sciences of the United States of America* 81, 1494-1498.

Kunz, C., Saito, Y. & Schar, P. (2009). DNA Repair in mammalian cells: Mismatched repair: variations on a theme. *Cell Mol Life Sci* 66, 1021-1038.

Kuo, C. F., McRee, D. E., Cunningham, R. P. & Tainer, J. A. (1992). Crystallization and crystallographic characterization of the iron-sulfur-containing DNA-repair enzyme endonuclease III from *Escherichia coli*. *Journal of molecular biology* 227, 347-351.

Kurthkoti, K., Srinath, T., Kumar, P., Malshetty, V. S., Sang, P. B., Jain, R., Manjunath, R. & Varshney, U. (2009). A distinct physiological role of MutY in mutation prevention in mycobacteria. *Microbiology (Reading, England)* in press.

Kuzminov, A. (1999). Recombinational repair of DNA damage in *Escherichia coli* and bacteriophage lambda. *Microbiol Mol Biol Rev* 63, 751-813, table of contents.

Laspia, M. F. & Wallace, S. S. (1988). Excision repair of thymine glycols, urea residues, and apurinic sites in *Escherichia coli*. *Journal of bacteriology* 170, 3359-3366.

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

Leadon, S. A. (1996). Repair of DNA Damage Produced by Ionizing Radiation: A Minireview. *Semin Radiat Oncol* 6, 295-305.

Lee, J. Y., Chang, J., Joseph, N., Ghirlando, R., Rao, D. N. & Yang, W. (2005). MutH complexed with hemi- and unmethylated DNAs: coupling base recognition and DNA cleavage. *Molecular cell* 20, 155-166.

Lees-Miller, S. P. & Meek, K. (2003). Repair of DNA double strand breaks by non-homologous end joining. *Biochimie* 85, 1161-1173.

Leipold, M. D., Muller, J. G., Burrows, C. J. & David, S. S. (2000). Removal of hydantoin products of 8-oxoguanine oxidation by the *Escherichia coli* DNA repair enzyme, FPG. *Biochemistry* 39, 14984-14992.

Levy, D. D. & Teebor, G. W. (1991). Site directed substitution of 5-hydroxymethyluracil for thymine in replicating phi X-174am3 DNA via synthesis of 5-hydroxymethyl-2'-deoxyuridine-5'-triphosphate. *Nucleic acids research* 19, 3337-3343.

Li, F., Liu, Q., Chen, Y. Y., Yu, Z. N., Zhang, Z. P., Zhou, Y. F., Deng, J. Y., Bi, L. J. & Zhang, X. E. (2008). *Escherichia coli* mismatch repair protein MutL interacts with the clamp loader subunits of DNA polymerase III. *Mutation research* 637, 101-110.

Lindahl, T. (1979). DNA glycosylases, endonucleases for apurinic/apyrimidinic sites, and base excision-repair. *Progress in nucleic acid research and molecular biology* 22, 135-192.

Lindahl, T. & Barnes, D. E. (2000). *Repair of endogenous DNA damage.* Cold Spring Harbor symposia on quantitative biology 65, 127-133.

Little, J. W. & Mount, D. W. (1982). The SOS regulatory system of *Escherichia coli*. *Cell* 29, 11-22.

Loeb, L. A. & Preston, B. D. (1986). Mutagenesis by apurinic/apyrimidinic sites. *Annual review of genetics* 20, 201-230.

Loughlin, M. F., Barnard, F. M., Jenkins, D., Sharples, G. J. & Jenks, P. J. (2003). *Helicobacter pylori* mutants defective in RuvC Holliday junction resolvase display reduced macrophage survival and spontaneous clearance from the murine gastric mucosa. *Infection and immunity* 71, 2022-2031.

Lu, A. L., Clark, S. & Modrich, P. (1983). Methyl-directed repair of DNA base-pair mismatches *in vitro*. *Proceedings of the National Academy of Sciences of the United States of America* 80, 4639-4643.

Lu, A. L. & Chang, D. Y. (1988). A novel nucleotide excision repair for the conversion of an A/G mismatch to C/G base pair in E. coli. *Cell* 54, 805-812.

Luo, W., Muller, J. G. & Burrows, C. J. (2001). The pH-dependent role of superoxide in riboflavin-catalyzed photooxidation of 8-oxo-7,8-dihydroguanosine. *Organic letters* 3, 2801-2804.

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

Maki, H. & Sekiguchi, M. (1992). MutT protein specifically hydrolyses a potent mutagenic substrate for DNA synthesis. *Nature* 355, 273-275.

Malins, D. C. & Haimanot, R. (1990). 4,6-Diamino-5-formamidopyrimidine, 8-hydroxyguanine and 8-hydroxyadenine in DNA from neoplastic liver of English sole exposed to carcinogens. *Biochemical and biophysical research communications* 173, 614-619.

Masaoka, A., Terato, H., Kobayashi, M., Honsho, A., Ohyama, Y. & Ide, H. (1999). Enzymatic repair of 5-formyluracil. I. Excision of 5-formyluracil site-specifically incorporated into oligonucleotide substrates by alka protein (*Escherichia coli* 3-methyladenine DNA glycosylase II). *The Journal of biological chemistry* 274, 25136-25143.

Mazur, S. J. & Grossman, L. (1991). Dimerization of *Escherichia coli* UvrA and its binding to undamaged and ultraviolet light damaged DNA. *Biochemistry* 30, 4432-4443.

McAuley-Hecht, K. E., Leonard, G. A., Gibson, N. J., Thomson, J. B., Watson, W. P., Hunter, W. N. & Brown, T. (1994). Crystal structure of a DNA duplex containing 8-hydroxydeoxyguanine-adenine base pairs. *Biochemistry* 33, 10266-10270.

Mellon, I. & Hanawalt, P. C. (1989). Induction of the *Escherichia coli* lactose operon selectively increases repair of its transcribed DNA strand. *Nature* 342, 95-98.

Michaels, M. L., Pham, L., Cruz, C. & Miller, J. H. (1991). MutM, a protein that prevents G.C----T.A transversions, is formamidopyrimidine-DNA glycosylase. *Nucleic acids research* 19, 3629-3632.

Michaels, M. L., Cruz, C., Grollman, A. P. & Miller, J. H. (1992). Evidence that MutY and MutM combine to prevent mutations by an oxidatively damaged form of guanine in DNA. *Proceedings of the National Academy of Sciences of the United States of America* 89, 7022-7025.

Michaels, M. L. & Miller, J. H. (1992). The GO system protects organisms from the mutagenic effect of the spontaneous lesion 8-hydroxyguanine (7,8-dihydro-8-oxoguanine). *Journal of bacteriology* 174, 6321-6325.

Mishina, Y., Duguid, E. M. & He, C. (2006). Direct reversal of DNA alkylation damage. *Chem Rev* 106, 215-232.

Misiaszek, R., Crean, C., Joffe, A., Geacintov, N. E. & Shafirovich, V. (2004). Oxidative DNA damage associated with combination of guanine and superoxide radicals and repair mechanisms via radical trapping. *The Journal of biological chemistry* 279, 32106-32115.

Miyabe, I., Zhang, Q. M., Sugiyama, H., Kino, K. & Yonei, S. (2001). Mutagenic effects of 5-formyluracil on a plasmid vector during replication in *Escherichia coli*. *International journal of radiation biology* 77, 53-58.

Mizrahi, V. & Andersen, S. J. (1998). DNA repair in *Mycobacterium tuberculosis*. What have we learnt from the genome sequence? *Molecular microbiology* 29, 1331-1339.

Moolenaar, G. F., Herron, M. F., Monaco, V., van der Marel, G. A., van Boom, J. H., Visse, R. & Goosen, N. (2000). The role of ATP binding and hydrolysis by UvrB during nucleotide excision repair. *The Journal of biological chemistry* 275, 8044-8050.

Moran, E. & Wallace, S. S. (1985). The role of specific DNA base damages in the X-ray-induced inactivation of bacteriophage PM2. *Mutation research* 146, 229-241.

Mori, T., Hori, Y. & Dizdaroglu, M. (1993). DNA base damage generated *in vivo* in hepatic chromatin of mice upon whole body gamma-irradiation. *International journal of radiation biology* 64, 645-650.

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

Moriya, M. (1993). Single-stranded shuttle phagemid for mutagenesis studies in mammalian cells: 8-oxoguanine in DNA induces targeted G.C-->T.A transversions in simian kidney cells. *Proceedings of the National Academy of Sciences of the United States of America* 90, 1122-1126.

Moriya, M. & Grollman, A. P. (1993). Mutations in the mutY gene of *Escherichia coli* enhance the frequency of targeted G:C-->T:a transversions induced by a single 8-oxoguanine residue in single-stranded DNA. *Mol Gen Genet* 239, 72-76.

Morland, I., Rolseth, V., Luna, L., Rognes, T., Bjoras, M. & Seeberg, E. (2002). Human DNA glycosylases of the bacterial Fpg/MutM superfamily: an alternative pathway for the repair of 8-oxoguanine and other oxidation products in DNA. *Nucleic acids research* 30, 4926-4936.

Mosbaugh, D. W. (1988). Purification and characterization of porcine liver DNA polymerase gamma: utilization of dUTP and dTTP during *in vitro* DNA synthesis. *Nucleic acids research* 16, 5645-5659.

Movahedzadeh, F., Colston, M. J. & Davis, E. O. (1997a). Characterization of *Mycobacterium tuberculosis* LexA: recognition of a Cheo (Bacillus-type SOS) box. *Microbiology (Reading, England)* 143 (Pt 3), 929-936.

Movahedzadeh, F., Colston, M. J. & Davis, E. O. (1997b). Determination of DNA sequences required for regulated *Mycobacterium tuberculosis* RecA expression in response to DNA-damaging agents suggests that two modes of regulation exist. *Journal of bacteriology* 179, 3509-3518.

Murphy, T. M. & George, A. (2005). A comparison of two DNA base excision repair glycosylases from *Arabidopsis thaliana*. *Biochemical and biophysical research communications* 329, 869-872.

Nackerdien, Z., Kasprzak, K. S., Rao, G., Halliwell, B. & Dizdaroglu, M. (1991). Nickel(II)- and cobalt(II)-dependent damage by hydrogen peroxide to the DNA bases in isolated human chromatin. *Cancer research* 51, 5837-5842.

Nash, H. M., Bruner, S. D., Scharer, O. D., Kawate, T., Addona, T. A., Spooner, E., Lane, W. S. & Verdine, G. L. (1996). Cloning of a yeast 8-oxoguanine DNA glycosylase reveals the existence of a base-excision DNA-repair protein superfamily. *Curr Biol* 6, 968-980.

Neeley, W. L. & Essigmann, J. M. (2006). Mechanisms of formation, genotoxicity, and mutation of guanine oxidation products. *Chemical research in toxicology* 19, 491-505.

Neto, J. B., Gentil, A., Cabral, R. E. & Sarasin, A. (1992). Mutation spectrum of heat-induced abasic sites on a single-stranded shuttle vector replicated in mammalian cells. *The Journal of biological chemistry* 267, 19718-19723.

Nghiem, Y., Cabrera, M., Cupples, C. G. & Miller, J. H. (1988). The mutY gene: a mutator locus in *Escherichia coli* that generates G.C----T.A transversions. *Proceedings of the National Academy of Sciences of the United States of America* 85, 2709-2713.

Nick McElhinny, S. A., Snowden, C. M., McCarville, J. & Ramsden, D. A. (2000). Ku recruits the XRCC4-ligase IV complex to DNA ends. *Molecular and cellular biology* 20, 2996-3003.

Nieminuszczy, J. & Grzesiuk, E. (2007). Bacterial DNA repair genes and their eukaryotic homologues: 3. AlkB dioxygenase and Ada methyltransferase in the direct repair of alkylated DNA. *Acta biochimica Polonica* 54, 459-468.

O'Brien, P. J. (2006). Catalytic promiscuity and the divergent evolution of DNA repair enzymes. *Chem Rev* 106, 720-752.

O'Connor, T. R., Boiteux, S. & Laval, J. (1988). Ring-opened 7-methylguanine residues in DNA are a block to *in vitro* DNA synthesis. *Nucleic acids research* 16, 5879-5894.

O'Rourke, E. J., Chevalier, C., Pinto, A. V., Thiberge, J. M., Ielpi, L., Labigne, A. & Radicella, J. P. (2003). Pathogen DNA as target for host-generated oxidative stress: role for repair of bacterial DNA damage in *Helicobacter pylori* colonization. *Proceedings of the National Academy of Sciences of the United States of America* 100, 2789-2794.

O'Sullivan, D. M., McHugh, T. D. & Gillespie, S. H. (2005). Analysis of rpoB and pncA mutations in the published literature: an insight into the role of oxidative stress in *Mycobacterium tuberculosis* evolution? *The Journal of antimicrobial chemotherapy* 55, 674-679.

Oda, Y., Uesugi, S., Ikehara, M., Nishimura, S., Kawase, Y., Ishikawa, H., Inoue, H. & Ohtsuka, E. (1991). NMR studies of a DNA containing 8-hydroxydeoxyguanosine. *Nucleic acids research* 19, 1407-1412.

Ohtsubo, T., Matsuda, O., Iba, K., Terashima, I., Sekiguchi, M. & Nakabeppu, Y. (1998). Molecular cloning of AtMMH, an *Arabidopsis thaliana* ortholog of the *Escherichia coli* mutM gene, and analysis of functional domains of its product. *Mol Gen Genet* 259, 577-590.

Olinski, R., Zastawny, T., Budzbon, J., Skokowski, J., Zegarski, W. & Dizdaroglu, M. (1992). DNA base modifications in chromatin of human cancerous tissues. *FEBS letters 309, 193-198.*

Olsen, I., Balasingham, S. V., Davidsen, T., Debebe, E., Rodland, E. A., van Soolingen, D., Kremer, K., Alseth, I. & Tonjum, T. (2009). Characterization of the major formamidopyrimidine-DNA glycosylase homolog in *Mycobacterium tuberculosis* and its linkage to variable tandem repeats. *FEMS immunology and medical microbiology* 56, 151-161.

Orren, D. K. & Sancar, A. (1990). Formation and enzymatic properties of the UvrB.DNA complex. *The Journal of biological chemistry* 265, 15796-15803.

Pfeifer, G. P. (1997). Formation and processing of UV photoproducts: effects of DNA sequence and chromatin environment. *Photochemistry and photobiology* 65, 270-283.

Pitcher, R. S., Tonkin, L. M., Green, A. J. & Doherty, A. J. (2005a). Domain structure of a NHEJ DNA repair ligase from *Mycobacterium tuberculosis*. *Journal of molecular biology* 351, 531-544.

Pitcher, R. S., Wilson, T. E. & Doherty, A. J. (2005b). New insights into NHEJ repair processes in prokaryotes. *Cell cycle* (Georgetown, Tex) 4, 675-678.

Pitcher, R. S., Brissett, N. C. & Doherty, A. J. (2007a). Nonhomologous end-joining in bacteria: a microbial perspective. *Annual review of microbiology* 61, 259-282.

Pitcher, R. S., Brissett, N. C., Picher, A. J., Andrade, P., Juarez, R., Thompson, D., Fox, G. C., Blanco, L. & Doherty, A. J. (2007b). Structure and function of a mycobacterial NHEJ DNA repair polymerase. *Journal of molecular biology* 366, 391-405.

Pitcher, R. S., Green, A. J., Brzostek, A., Korycka-Machala, M., Dziadek, J. & Doherty, A. J. (2007c). NHEJ protects mycobacteria in stationary phase against the harmful effects of desiccation. *DNA repair* 6, 1271-1276.

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

Purmal, A. A., Lampman, G. W., Kow, Y. W. & Wallace, S. S. (1994b). The sequence context-dependent mispairing of 5-hydroxycytosine and 5-hydroxyuridine *in vitro*. *Annals of the New York Academy of Sciences* 726, 361-363.

Purmal, A. A., Bond, J. P., Lyons, B. A., Kow, Y. W. & Wallace, S. S. (1998). Uracil glycol deoxynucleoside triphosphate is a better substrate for DNA polymerase I Klenow fragment than thymine glycol deoxynucleoside triphosphate. *Biochemistry* 37, 330-338.

Purnapatre, K. & Varshney, U. (1998). Uracil DNA glycosylase from *Mycobacterium smegmatis* and its distinct biochemical properties. *European journal of biochemistry / FEBS* 256, 580-588.

Radicella, J. P., Clark, E. A. & Fox, M. S. (1988). Some mismatch repair activities in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* 85, 9674-9678.

Radman, M. (1998). DNA replication: one strand may be more equal. *Proceedings of the National Academy of Sciences of the United States of America* 95, 9718-9719.

Ramsden, D. A. & Gellert, M. (1998). Ku protein stimulates DNA end joining by mammalian DNA ligases: a direct role for Ku in repair of DNA double-strand breaks. *The EMBO journal* 17, 609-614.

Rand, L., Hinds, J., Springer, B., Sander, P., Buxton, R. S. & Davis, E. O. (2003). The majority of inducible DNA repair genes in *Mycobacterium tuberculosis* are induced independently of RecA. *Molecular microbiology* 50, 1031-1042.

Randall, S. K., Eritja, R., Kaplan, B. E., Petruska, J. & Goodman, M. F. (1987). Nucleotide insertion kinetics opposite abasic lesions in DNA. *The Journal of biological chemistry* 262, 6864-6870.

Rengarajan, J., Bloom, B. R. & Rubin, E. J. (2005). Genome-wide requirements for *Mycobacterium tuberculosis* adaptation and survival in macrophages. *Proceedings of the National Academy of Sciences of the United States of America* 102, 8327-8332.

Robey-Bond, S. M., Barrantes-Reynolds, R., Bond, J. P., Wallace, S. S. & Bandaru, V. (2008). Clostridium acetobutylicum 8-Oxoguanine DNA Glycosylase (Ogg) Differs from Eukaryotic Oggs with Respect to Opposite Base Discrimination. *Biochemistry* 47, 7626-7636.

Rogers, P. A., Eide, L., Klungland, A. & Ding, H. (2003). Reversible inactivation of E. coli endonuclease III via modification of its [4Fe-4S] cluster by nitric oxide. *DNA repair* 2, 809-817.

Roldan-Arjona, T., Anselmino, C. & Lindahl, T. (1996). Molecular cloning and functional analysis of a *Schizosaccharomyces pombe* homologue of *Escherichia coli* endonuclease III. *Nucleic acids research* 24, 3307-3312.

Rosenquist, T. A., Zaika, E., Fernandes, A. S., Zharkov, D. O., Miller, H. & Grollman, A. P. (2003). The novel DNA glycosylase, NEIL1, protects mammalian cells from radiation-mediated cell death. *DNA repair* 2, 581-591.

Rouet, P. & Essigmann, J. M. (1985). Possible role for thymine glycol in the selective inhibition of DNA synthesis on oxidized DNA templates. *Cancer research* 45, 6113-6118.

Sagher, D. & Strauss, B. (1983). Insertion of nucleotides opposite apurinic/apyrimidinic sites in deoxyribonucleic acid during *in vitro* synthesis: uniqueness of adenine nucleotides. *Biochemistry* 22, 4518-4526.

Saito, Y., Uraki, F., Nakajima, S., Asaeda, A., Ono, K., Kubo, K. & Yamamoto, K. (1997). Characterization of endonuclease III (nth) and endonuclease VIII (nei) mutants of *Escherichia coli* K-12. *Journal of bacteriology* 179, 3783-3785.

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

Sancar, A. & Rupp, W. D. (1983). A novel repair enzyme: UVRABC excision nuclease of *Escherichia coli* cuts a DNA strand on both sides of the damaged region. *Cell* 33, 249-260.

Sarker, A. H., Ikeda, S., Nakano, H. & other authors (1998). Cloning and characterization of a mouse homologue (mNthl1) of *Escherichia coli* endonuclease III. *Journal of molecular biology* 282, 761-774.

Sartori, A. A., Fitz-Gibbon, S., Yang, H., Miller, J. H. & Jiricny, J. (2002). A novel uracil-DNA glycosylase with broad substrate specificity and an unusual active site. *The EMBO journal* 21, 3182-3191.

Sassetti, C. M. & Rubin, E. J. (2003). Genetic requirements for mycobacterial survival during infection. *Proceedings of the National Academy of Sciences of the United States of America* 100, 12989-12994.

Savery, N. J. (2007). The molecular mechanism of transcription-coupled DNA repair. *Trends in microbiology* 15, 326-333.

Schaaper, R. M. & Loeb, L. A. (1981). Depurination causes mutations in SOS-induced cells. *Proceedings of the National Academy of Sciences of the United States of America* 78, 1773-1777.

Schaaper, R. M., Kunkel, T. A. & Loeb, L. A. (1983). Infidelity of DNA synthesis associated with bypass of apurinic sites. *Proceedings of the National Academy of Sciences of the United States of America* 80, 487-491.

Schaaper, R. M. & Dunn, R. L. (1987). *Escherichia coli* mutT mutator effect during *in vitro* DNA synthesis. Enhanced A.G replicational errors. *The Journal of biological chemistry* 262, 16267-16270.

Sedgwick, B. & Vaughan, P. (1991). Widespread adaptive response against environmental methylating agents in microorganisms. *Mutation research* 250, 211-221.

Sedgwick, B. & Lindahl, T. (2002). Recent progress on the Ada response for inducible repair of DNA alkylation damage. *Oncogene* 21, 8886-8894.

Selby, C. P., Witkin, E. M. & Sancar, A. (1991). *Escherichia coli* mfd mutant deficient in "mutation frequency decline" lacks strand-specific repair: *in vitro* complementation

with purified coupling factor. *Proceedings of the National Academy of Sciences of the United States of America* 88, 11574-11578.

Selby, C. P. & Sancar, A. (1993). Molecular mechanism of transcription-repair coupling. *Science* (New York, NY) 260, 53-58.

Selby, C. P. & Sancar, A. (1994). Mechanisms of transcription-repair coupling and mutation frequency decline. *Microbiological reviews* 58, 317-329.

Serre, L., Pereira de Jesus, K., Boiteux, S., Zelwer, C. & Castaing, B. (2002). Crystal structure of the Lactococcus lactis formamidopyrimidine-DNA glycosylase bound to an abasic site analogue-containing DNA. *The EMBO journal* 21, 2854-2865.

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

Shibutani, S., Bodepudi, V., Johnson, F. & Grollman, A. P. (1993). Translesional synthesis on DNA templates containing 8-oxo-7,8-dihydrodeoxyadenosine. *Biochemistry* 32, 4615-4621.

Shuman, S. & Glickman, M. S. (2007). Bacterial DNA repair by non-homologous end joining. *Nature reviews* 5, 852-861.

Sidorenko, V. S., Rot, M. A., Filipenko, M. L., Nevinsky, G. A. & Zharkov, D. O. (2008). Novel DNA glycosylases from *Mycobacterium tuberculosis*. *Biochemistry* (Mosc) 73, 442-450.

Sinha, K. M., Stephanou, N. C., Gao, F., Glickman, M. S. & Shuman, S. (2007). Mycobacterial UvrD1 is a Ku-dependent DNA helicase that plays a role in multiple DNA repair events, including double-strand break repair. *The Journal of biological chemistry* 282, 15114-15125.

Slade, P. G., Hailer, M. K., Martin, B. D. & Sugden, K. D. (2005). Guanine-specific oxidation of double-stranded DNA by Cr(VI) and ascorbic acid forms spiroiminodihydantoin and 8-oxo-2'-deoxyguanosine. *Chemical research in toxicology* 18, 1140-1149.

Spampinato, C. & Modrich, P. (2000). The MutL ATPase is required for mismatch repair. *The Journal of biological chemistry* 275, 9863-9869.

Srinath, T., Bharti, S. K. & Varshney, U. (2007). Substrate specificities and functional characterization of a thermo-tolerant uracil DNA glycosylase (UdgB) from *Mycobacterium tuberculosis*. *DNA repair* 6, 1517-1528.

Stillwell, W. G., Xu, H. X., Adkins, J. A., Wishnok, J. S. & Tannenbaum, S. R. (1989). Analysis of methylated and oxidized purines in urine by capillary gas chromatography-mass spectrometry. *Chemical research in toxicology* 2, 94-99.

Sugahara, M., Mikawa, T., Kumasaka, T., Yamamoto, M., Kato, R., Fukuyama, K., Inoue, Y. & Kuramitsu, S. (2000). Crystal structure of a repair enzyme of oxidatively damaged DNA, MutM (Fpg), from an extreme thermophile, *Thermus thermophilus* HB8. *The EMBO journal* 19, 3857-3869.

Sugden, K. D., Campo, C. K. & Martin, B. D. (2001). Direct oxidation of guanine and 7,8-dihydro-8-oxoguanine in DNA by a high-valent chromium complex: a possible mechanism for chromate genotoxicity. *Chemical research in toxicology* 14, 1315-1322.

Sun, B., Latham, K. A., Dodson, M. L. & Lloyd, R. S. (1995). Studies on the catalytic mechanism of five DNA glycosylases. Probing for enzyme-DNA imino intermediates. *The Journal of biological chemistry* 270, 19501-19508.

Suvarnapunya, A. E., Lagasse, H. A. & Stein, M. A. (2003). The role of DNA base excision repair in the pathogenesis of *Salmonella* enterica serovar *Typhimurium*. *Molecular microbiology* 48, 549-559.

Svejstrup, J. Q. (2002). Mechanisms of transcription-coupled DNA repair. *Nat Rev Mol Cell Biol* 3, 21-29.

Takao, M., Kanno, S., Kobayashi, K., Zhang, Q. M., Yonei, S., van der Horst, G. T. & Yasui, A. (2002). A back-up glycosylase in Nth1 knock-out mice is a functional Nei (endonuclease VIII) homologue. *The Journal of biological chemistry* 277, 42205-42213.

Takao, M., Oohata, Y., Kitadokoro, K., Kobayashi, K., Iwai, S., Yasui, A., Yonei, S. & Zhang, Q. M. (2009). Human Nei-like protein NEIL3 has AP lyase activity specific for single-stranded DNA and confers oxidative stress resistance in *Escherichia coli* mutant. *Genes Cells* 14, 261-270.

Tchou, J., Kasai, H., Shibutani, S., Chung, M. H., Laval, J., Grollman, A. P. & Nishimura, S. (1991). 8-oxoguanine (8-hydroxyguanine) DNA glycosylase and its substrate specificity. *Proceedings of the National Academy of Sciences of the United States of America* 88, 4690-4694.

Tchou, J., Bodepudi, V., Shibutani, S., Antoshechkin, I., Miller, J., Grollman, A. P. & Johnson, F. (1994). Substrate specificity of Fpg protein. Recognition and cleavage of oxidatively damaged DNA. *J Biol Chem* 269, 15318-15324.

Teoule, R. (1987). Radiation-induced DNA damage and its repair. *International journal of radiation biology and related studies in physics, chemistry, and medicine* 51, 573-589.

Theis, K., Skorvaga, M., Machius, M., Nakagawa, N., Van Houten, B. & Kisker, C. (2000). The nucleotide excision repair protein UvrB, a helicase-like enzyme with a catch. *Mutation research* 460, 277-300.

Tiraby, J. G. & Fox, M. S. (1973). Marker discrimination in transformation and mutation of pneumococcus. *Proceedings of the National Academy of Sciences of the United States of America* 70, 3541-3545.

Trewick, S. C., Henshaw, T. F., Hausinger, R. P., Lindahl, T. & Sedgwick, B. (2002). Oxidative demethylation by *Escherichia coli* AlkB directly reverts DNA base damage. *Nature* 419, 174-178.

Truglio, J. J., Croteau, D. L., Van Houten, B. & Kisker, C. (2006). Prokaryotic nucleotide excision repair: the UvrABC system. *Chem Rev* 106, 233-252.

Tsai-Wu, J. J., Liu, H. F. & Lu, A. L. (1992). *Escherichia coli* MutY protein has both N-glycosylase and apurinic/apyrimidinic endonuclease activities on A.C and A.G mispairs. *Proceedings of the National Academy of Sciences of the United States of America* 89, 8779-8783.

Tudek, B., Boiteux, S. & Laval, J. (1992). Biological properties of imidazole ring-opened N7-methylguanine in M13mp18 phage DNA. *Nucleic acids research* 20, 3079-3084.

Tudek, B., Graziewicz, M., Kazanova, O., Zastawny, T. H., Obtulowicz, T. & Laval, J. (1999). Mutagenic specificity of imidazole ring-opened 7-methylpurines in M13mp18 phage DNA. *Acta biochimica Polonica* 46, 785-799.

Tye, B. K., Chien, J., Lehman, I. R., Duncan, B. K. & Warner, H. R. (1978). Uracil incorporation: a source of pulse-labeled DNA fragments in the replication of the *Escherichia coli* chromosome. *Proceedings of the National Academy of Sciences of the United States of America* 75, 233-237.

van der Kemp, P. A., Thomas, D., Barbey, R., de Oliveira, R. & Boiteux, S. (1996). Cloning and expression in *Escherichia coli* of the OGG1 gene of *Saccharomyces cerevisiae*, which codes for a DNA glycosylase that excises 7,8-dihydro-8-oxoguanine and 2,6-diamino-4-hydroxy-5-N-methylformamidopyrimidine. *Proceedings of the National Academy of Sciences of the United States of America* 93, 5197-5202.

Van Hemmen, J. J. & Bleichrodt, J. F. (1971). The decomposition of adenine by ionizing radiation. *Radiat Res* 46, 444-456.

Van Houten, B. (1990). Nucleotide excision repair in *Escherichia coli*. *Microbiological reviews* 54, 18-51.

Van Houten, B. & Snowden, A. (1993). Mechanism of action of the *Escherichia coli* UvrABC nuclease: clues to the damage recognition problem. *Bioessays* 15, 51-59.

Verhoeven, E. E., van Kesteren, M., Moolenaar, G. F., Visse, R. & Goosen, N. (2000). Catalytic sites for 3' and 5' incision of *Escherichia coli* nucleotide excision repair are both located in UvrC. *The Journal of biological chemistry* 275, 5120-5123.

Viswanathan, M. & Lovett, S. T. (1999). Exonuclease X of *Escherichia coli*. A novel 3'-5' DNase and Dnaq superfamily member involved in DNA repair. *The Journal of biological chemistry* 274, 30094-30100.

Wagner, J. R., Hu, C. C. & Ames, B. N. (1992). Endogenous oxidative damage of deoxycytidine in DNA. *Proceedings of the National Academy of Sciences of the United States of America* 89, 3380-3384.

Wallace, S. S. (1998). Enzymatic processing of radiation-induced free radical damage in DNA. *Radiat Res* 150, S60-79.

Wallace, S. S. (2002). Biological consequences of free radical-damaged DNA bases. *Free radical biology & medicine* 33, 1-14.

Wallace, S. S., Bandaru, V., Kathe, S. D. & Bond, J. P. (2003). The enigma of endonuclease VIII. *DNA repair* 2, 441-453.

Weisemann, J. M. & Weinstock, G. M. (1991). The promoter of the recA gene of *Escherichia coli*. *Biochimie* 73, 457-470.

Weller, G. R., Kysela, B., Roy, R. & other authors (2002). Identification of a DNA nonhomologous end-joining complex in bacteria. *Science* (New York, NY 297, 1686-1689.

Wood, M. L., Dizdaroglu, M., Gajewski, E. & Essigmann, J. M. (1990). Mechanistic studies of ionizing radiation and oxidative mutagenesis: genetic effects of a single 8-hydroxyguanine (7-hydro-8-oxoguanine) residue inserted at a unique site in a viral genome. *Biochemistry* 29, 7024-7032.

Wood, M. L., Esteve, A., Morningstar, M. L., Kuziemko, G. M. & Essigmann, J. M. (1992). Genetic effects of oxidative DNA damage: comparative mutagenesis of 7,8-

dihydro-8-oxoguanine and 7,8-dihydro-8-oxoadenine in *Escherichia coli*. *Nucleic acids research* 20, 6023-6032.

Wu, T. H. & Marinus, M. G. (1994). Dominant negative mutator mutations in the mutS gene of *Escherichia coli*. *Journal of bacteriology* 176, 5393-5400.

Yavin, E., Boal, A. K., Stemp, E. D., Boon, E. M., Livingston, A. L., O'Shea, V. L., David, S. S. & Barton, J. K. (2005). Protein-DNA charge transport: redox activation of a DNA repair protein by guanine radical. *Proceedings of the National Academy of Sciences of the United States of America* 102, 3546-3551.

Yavin, E., Stemp, E. D., O'Shea V, L., David, S. S. & Barton, J. K. (2006). Electron trap for DNA-bound repair enzymes: a strategy for DNA-mediated signaling. *Proceedings of the National Academy of Sciences of the United States of America* 103, 3610-3614.

You, H. J., Swanson, R. L., Harrington, C. & other authors (1999). *Saccharomyces cerevisiae* Ntg1p and Ntg2p: broad specificity N-glycosylases for the repair of oxidative DNA damage in the nucleus and mitochondria. *Biochemistry* 38, 11298-11306.

Zhang, Q. M., Sugiyama, H., Miyabe, I., Matsuda, S., Saito, I. & Yonei, S. (1997). Replication of DNA templates containing 5-formyluracil, a major oxidative lesion of thymine in DNA. *Nucleic acids research* 25, 3969-3973.

Zhang, Q. M., Yonekura, S., Takao, M., Yasui, A., Sugiyama, H. & Yonei, S. (2005). DNA glycosylase activities for thymine residues oxidized in the methyl group are functions of the hNEIL1 and hNTH1 enzymes in human cells. *DNA repair* 4, 71-79.

Zharkov, D. O., Rieger, R. A., Iden, C. R. & Grollman, A. P. (1997). NH2-terminal proline acts as a nucleophile in the glycosylase/AP-lyase reaction catalyzed by *Escherichia coli* formamidopyrimidine-DNA glycosylase (Fpg) protein. *The Journal of biological chemistry* 272, 5335-5341.

Zharkov, D. O., Golan, G., Gilboa, R., Fernandes, A. S., Gerchman, S. E., Kycia, J. H., Rieger, R. A., Grollman, A. P. & Shoham, G. (2002). Structural analysis of an *Escherichia coli* endonuclease VIII covalent reaction intermediate. *The EMBO journal* 21, 789-800.

Zharkov, D. O., Shoham, G. & Grollman, A. P. (2003). Structural characterization of the Fpg family of DNA glycosylases. *DNA repair* 2, 839-862.

Zharkov, D. O. (2008). Base excision DNA repair. *Cell Mol Life Sci* 65, 1544-1565.

Figure 1 Typical stable pyrimidine and purine damages [modified from (Wallace, 1998)].

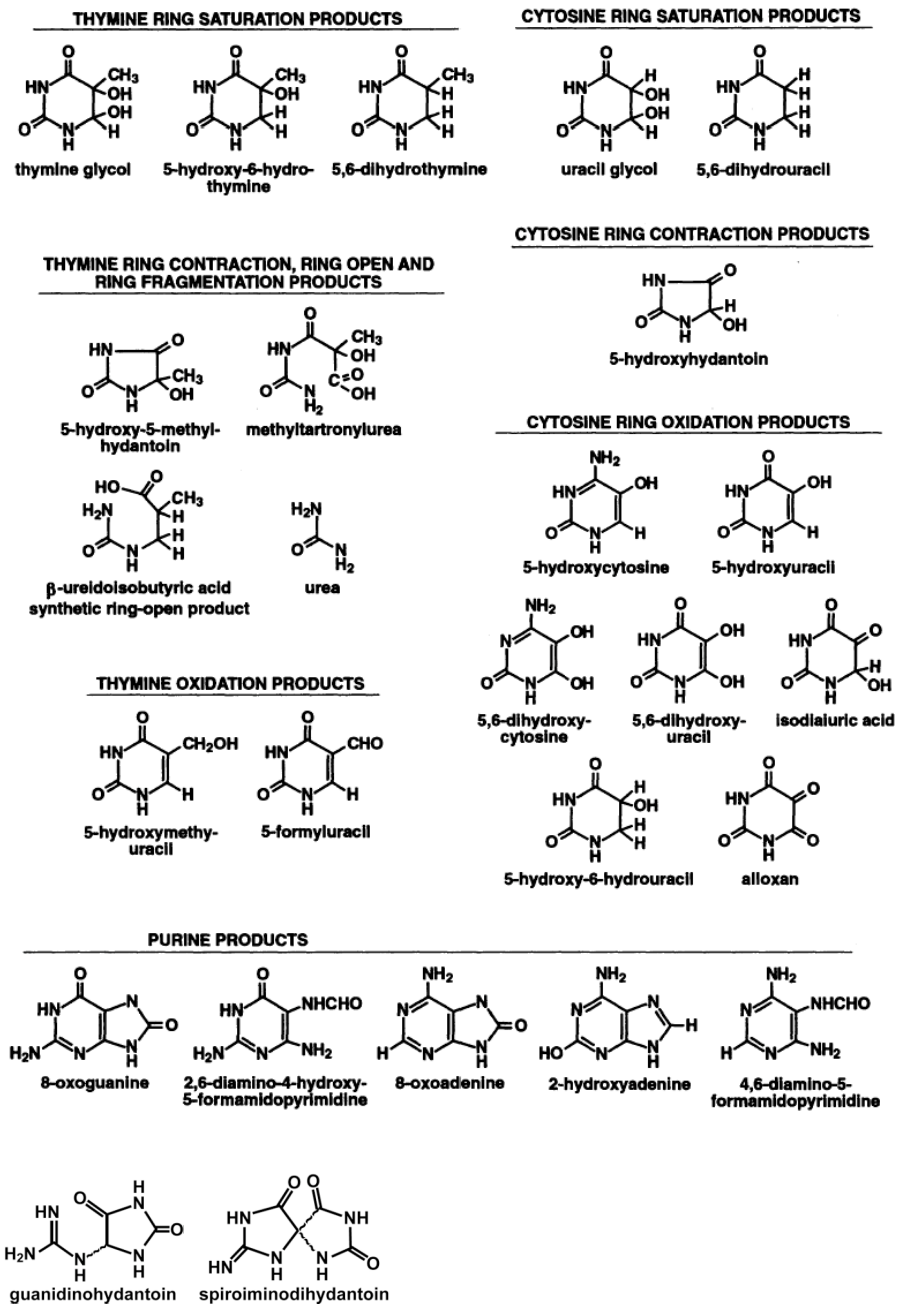


Figure 2 The base excision repair pathway (Wallace, 1998).

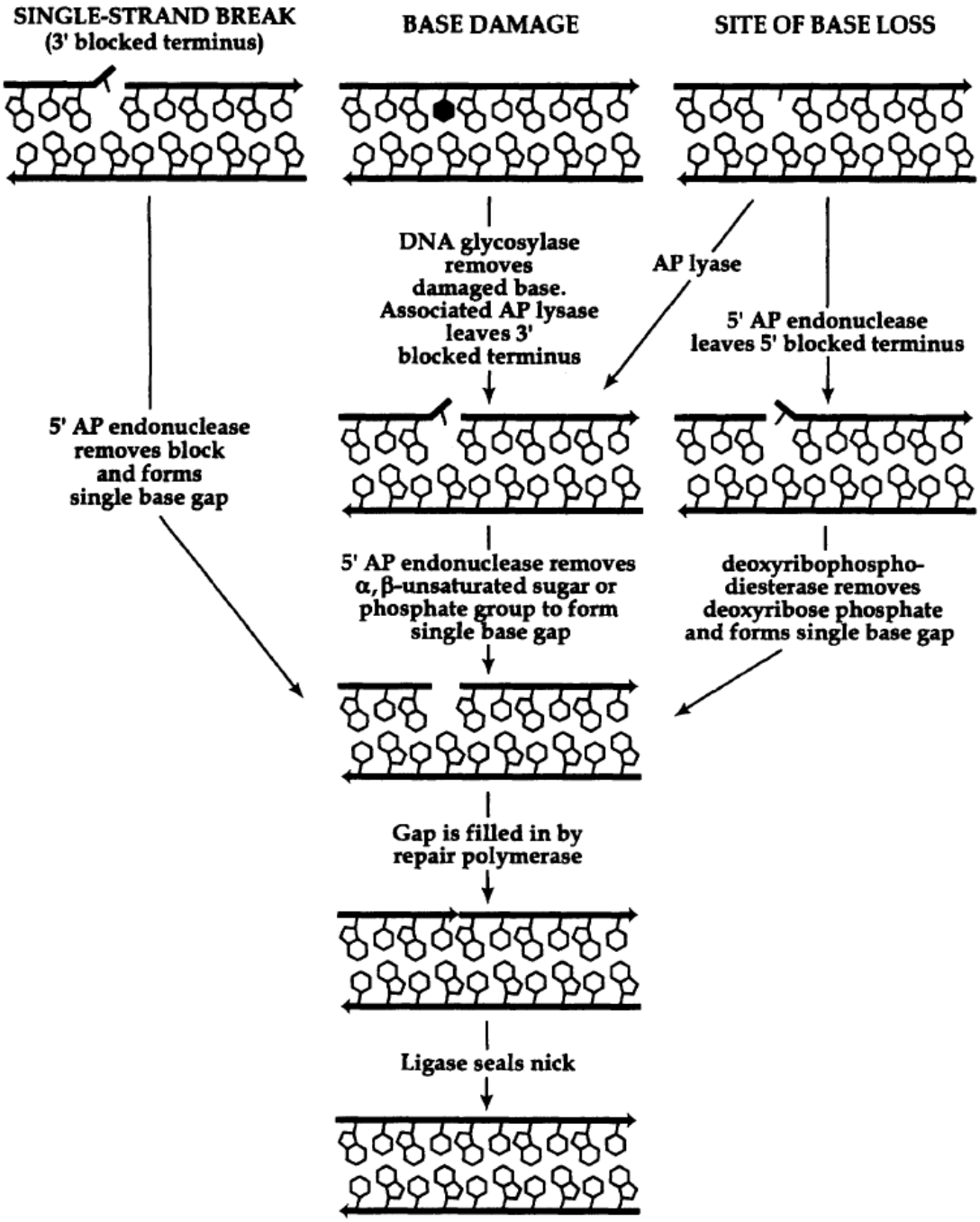


Figure 3 Phylogeny of Fpg/Nei family DNA glycosylases [modified from (Kathe *et al.*, 2009)].

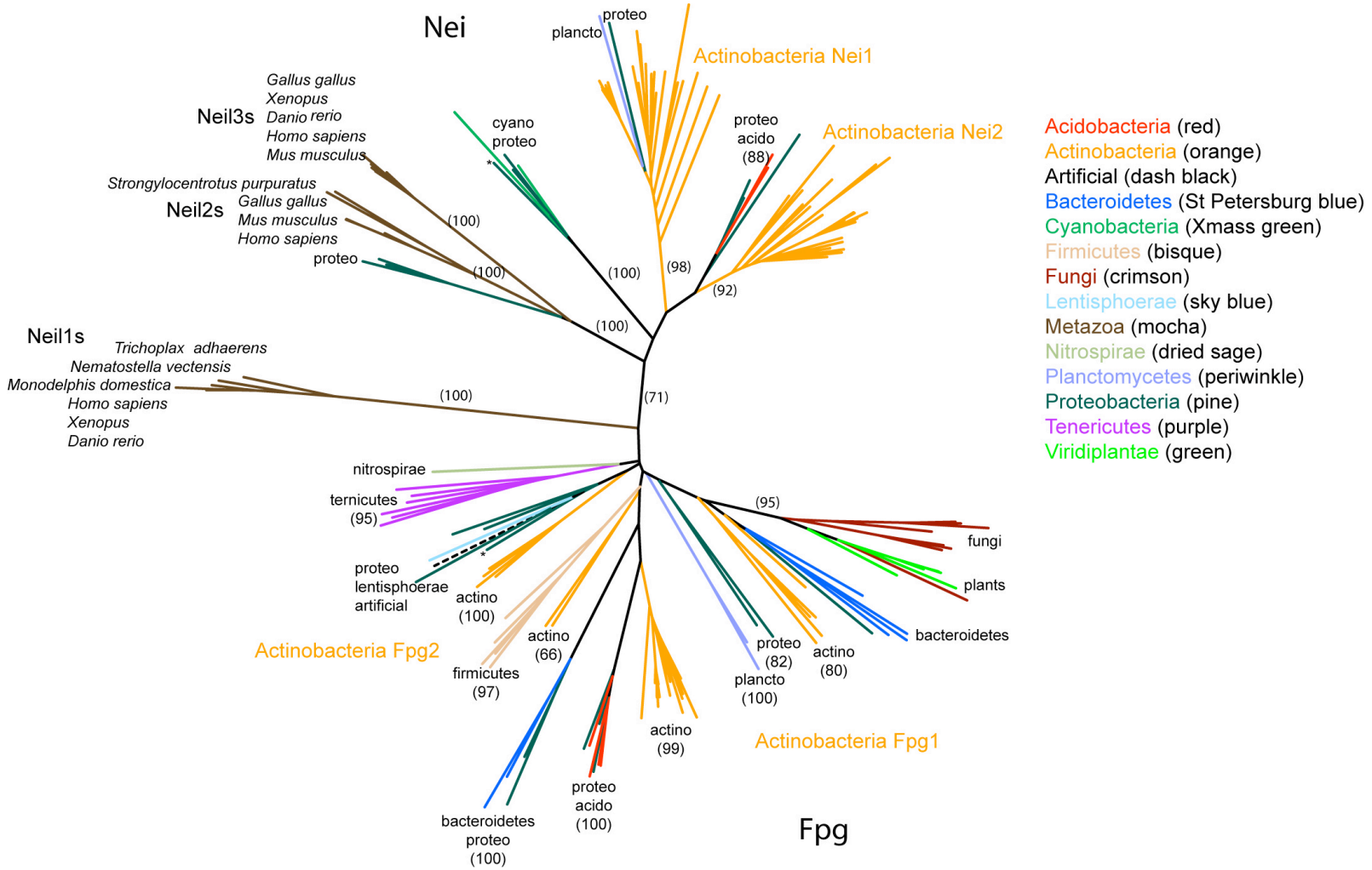

Nei
proteo
plancto
Actinobacteria Nei1
Actinobacteria Nei2
proteo
acido
(88)
Gallus gallus
Xenopus
Neil3s
Danio rerio
Homo sapiens
Mus musculus
cyano
proteo
*
Strongylocentrotus purpuratus
Neil2s
Gallus gallus
Mus musculus
Homo sapiens
proteo
(100)
(100)
(100)
(98)
(92)
(100)
Neil1s
Trichoplax adhaerens
Nematostella vectensis
Monodelphis domestica
Homo sapiens
Xenopus
Danio rerio
(100)
(71)
nitrospirae
ternicutes
(95)
proteo
lentisphoerae
artificial
actino
(100)
Actinobacteria Fpg2
actino
(66)
firmicutes
(97)
(95)
fungi
plants
bacteroidetes
proteo
(82)
actino
(80)
plancto
(100)
actino
(99)
Actinobacteria Fpg1
proteo
acido
(100)
bacteroidetes
proteo
(100)
Fpg
Acidobacteria (red)
Actinobacteria (orange)
Artificial (dash black)
Bacteroidetes (St Petersburg blue)
Cyanobacteria (Xmass green)
Firmicutes (bisque)
Fungi (crimson)
Lentisphoerae (sky blue)
Metazoa (mocha)
Nitrospirae (dried sage)
Planctomycetes (periwinkle)
Proteobacteria (pine)
Tenericutes (purple)
Viridiplantae (green)

Figure 4 Catalytic mechanism of DNA glycosylases (Fromme *et al.*, 2004).

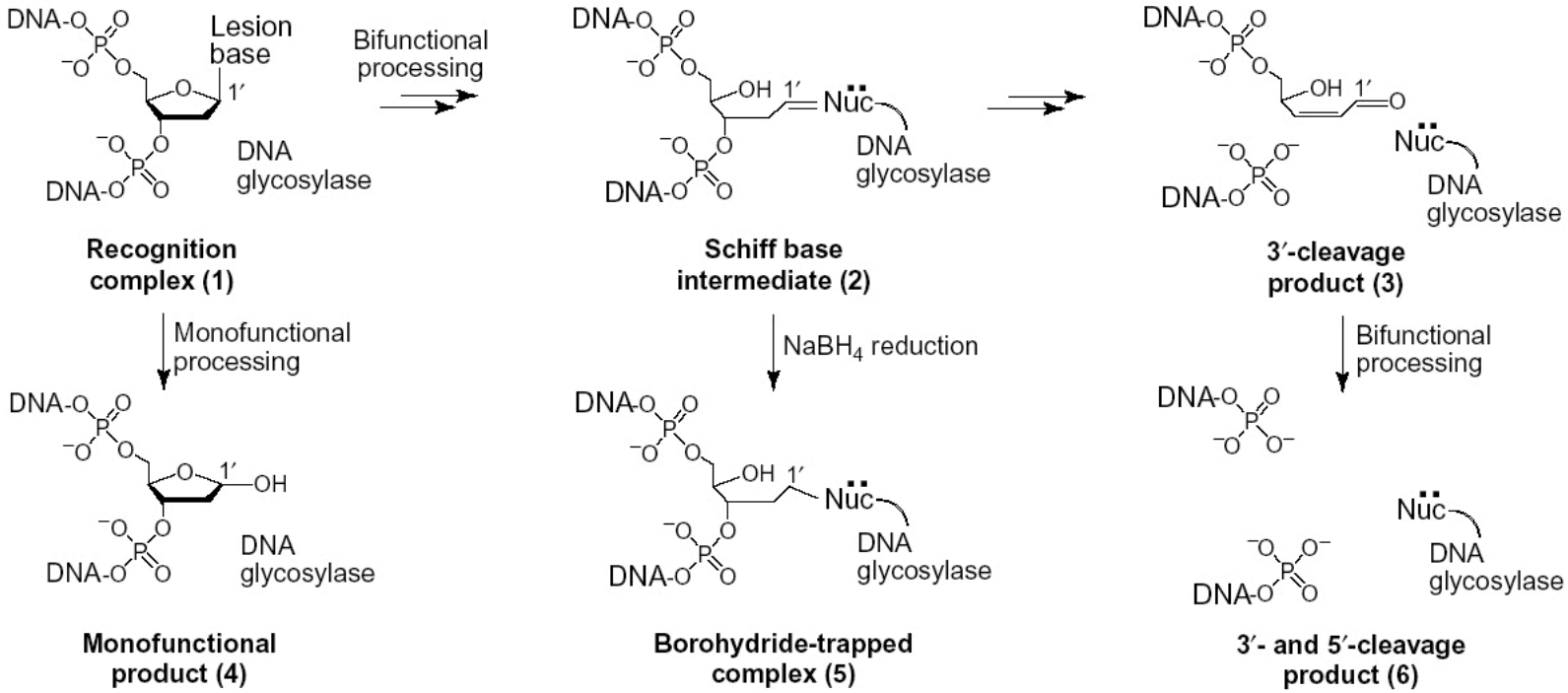
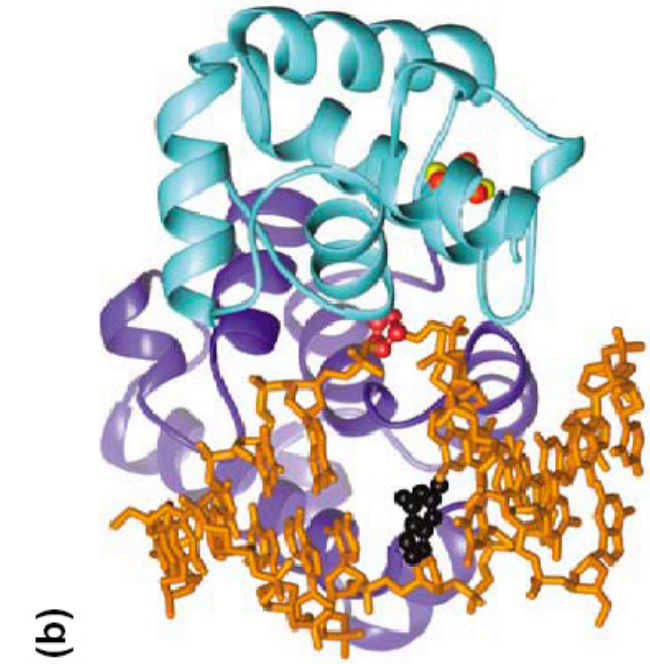
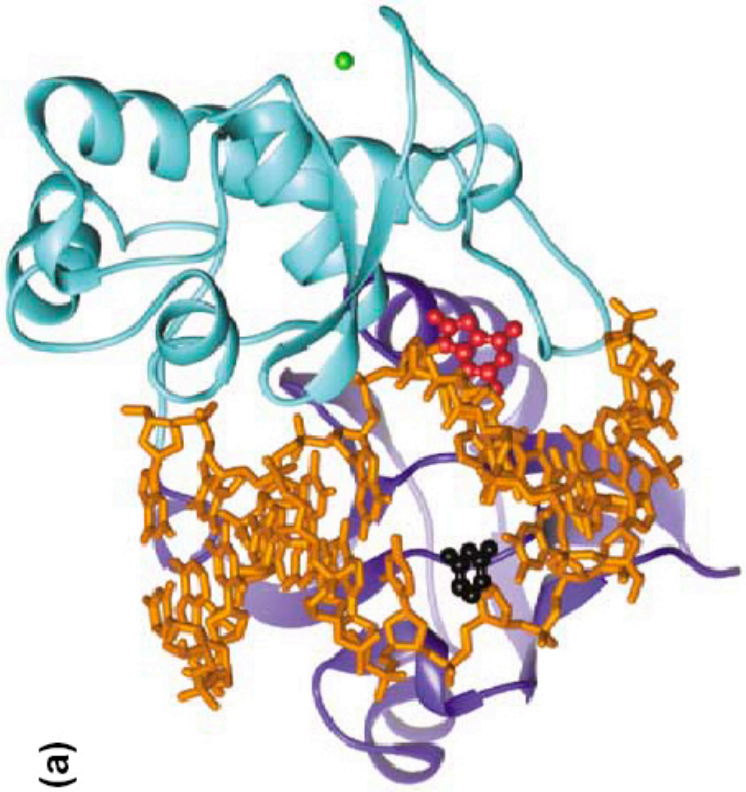

DNA-O
Lesion base
1′
Bifunctional processing
DNA glycosylase
Recognition complex (1)
Monofunctional processing
OH
Monofunctional product (4)
NaBH4 reduction
Nuc
Schiff base intermediate (2)
Borohydride-trapped complex (5)
3′-cleavage product (3)
Bifunctional processing
3′- and 5′-cleavage product (6)

Figure 5 Crystal structures of Fpg (a) and Nth (b) bound to DNA (Fromme *et al.*, 2004).



(b)



(a)

A Novel Bicistronic Vector for Overexpressing *Mycobacterium tuberculosis* Proteins in *Escherichia coli*

Yin Guo[a], Susan S. Wallace[a*], Viswanath Bandaru[b]

[a]Department of Microbiology and Molecular Genetics

The Markey Center for Molecular Genetics

University of Vermont

Stafford Hall, 95 Carrigan Dr.

Burlington, VT 05405-0068

Tel.: 802-656-2164

Fax: 802-656-8749

E-mail: susan.wallace@uvm.edu

* To whom correspondence should be addressed.

[b] Current address: Mylan Pharmaceuticals, 3711 Collins Ferry Rd, Morgantown WV 26505

Abstract

A putative DNA glycosylase encoded by the *Rv3297* gene (MtuNei2) has been identified in *Mycobacterium tuberculosis*. Our efforts to express this gene in *Escherichia coli* either by supplementing tRNAs for rare codons or optimizing the gene with preferred codons for *E. coli* resulted in little or no expression. On the other hand, high-level expression was observed using a bicistronic expression vector in which the target gene was translationally coupled to an upstream leader sequence. Further comparison of the predicted mRNA secondary structures supported the hypothesis that mRNA secondary structure(s) surrounding the translation initiation region (TIR), rather than codon usage, played the dominant role in influencing translation efficiency, although manipulation of codon usage or tRNA supplementation did further enhance expression in the bicistronic vector. Addition of a cleavable N-terminal tag also facilitated gene expression in *E. coli*, possibly through a similar mechanism. However, since cleavage of N-terminal tags is determined by the amino acid at the P_1' position downstream of the protease recognition sequence and results in the addition of an extra amino acid in front of the N-terminus of the protein, this strategy is not particularly amenable to Fpg/Nei family DNA glycosylases which carry the catalytic proline residue at the P_1' position and require a free N-terminus. On the other hand, the bicistronic vector constructed here is potentially valuable particularly when expressing proteins from G/C rich organisms and when the proteins carry proline residues at the N-terminus in their native form. Thus the bicistronic expression system can be used to

improve translation efficiency of mRNAs and achieve high-level expression of mycobacterial genes in *E. coli*.

Keywords: Codon usage, mRNA secondary structure, Translation initiation, Bicistronic vector

Introduction

Escherichia coli remains a common host for high-level expression of heterologous genes, however, this often depends on the source of the target genes. A number of factors that significantly influence gene expression at the translation level have been identified with codon usage and mRNA secondary structures being the major concerns.

There are marked differences in codon usage from one organism to another. Significant variation in codon usage patterns among genes in one organism appears to be associated with their expression levels. Genes with a high proportion of optimal codons are highly expressed, whereas those with rare codons are poorly expressed (Andersson & Sharp, 1996; Makoff *et al.*, 1989). Moreover, the presence of rare codons can cause ribosome stalling, slow translation, pre-mature translation termination and translation errors and therefore inhibit proper protein synthesis and even cell growth (Hannig & Makrides, 1998; Kane, 1995; Kurland & Gallant, 1996; Sorensen *et al.*, 1989; Zahn, 1996). To avoid the potential expression problems resulting from rare codons, one can either optimize codon usage in the target gene by silent mutations for expression in *E. coli* or expand the intracellular tRNA pool of rare codons by introducing a plasmid which encodes these tRNAs (Hannig & Makrides, 1998). Both strategies have been successfully used to enhance the expression of heterologous genes with rare codons in *E. coli* (Makoff *et al.*, 1989; Zhou *et al.*, 2004). However, negative results have been reported indicating that factors other than codon usage can affect protein expression (Griswold *et al.*, 2003; Zhou *et al.*, 2004).

It has been widely accepted that secondary structure of mRNA in the translation initiation region (TIR) plays a crucial role in controlling translation efficiency. A quantitative analysis revealed a strict correlation between the translational efficiency and the stability of the local secondary structure (de Smit & van Duin, 1990). Since the 30S ribosomal subunit most likely binds to single-stranded regions of mRNA and slides into place while unfolding the TIR, the secondary structure(s) by sequestering either the ribosome binding site (RBS) and/or the initiation codon from ribosome binding will block translation initiation thereby inhibiting translation (Hall *et al.*, 1982; Iserentant & Fiers, 1980; Studer & Joseph, 2006).

A series of two-cistron plasmids (bicistronic vectors) have been successfully used to overcome translational inhibition of mRNAs with stable secondary structures (Ishida & Oshima, 1994; Schoner *et al.*, 1984; Schoner *et al.*, 1986; Schoner *et al.*, 1990; Spanjaard *et al.*, 1989). In such a system, the first/upstream cistron is generally an A/T-rich sequence, which can minimize local secondary structure(s) thereby allowing efficient translation initiation (Schoner *et al.*, 1986; Schoner *et al.*, 1990), while the second/downstream cistron containing the coding sequence of the target gene is translationally coupled to the first/upstream cistron. Therefore, protein production from a two-cistron system is theoretically dependent upon the efficiency of translation of the first/upstream cistron (Makoff & Smallwood, 1990). Two models have been proposed in terms of the mechanism of translational coupling (Baughman & Nomura, 1983; Makoff & Smallwood, 1990; Schumperli *et al.*, 1982). One is that a ribosome translating the upstream cistron disrupts the inhibitory secondary structure thereby

making the RBS of the downstream cistron accessible to other ribosomes which can initiate translation of the coupled downstream gene. The other is that the same translating ribosome can re-initiate and continue to translate the downstream cistron.

Mycobacterium tuberculosis, the causative agent of tuberculosis, is a high-G/C gram-positive bacterium with a genomic G/C content of approximately 65%. The high G/C content results in a markedly different pattern of codon usage from that of *Escherichia coli*. A number of rare codons for *E. coli*, widely used in the genome of *M. tuberculosis*, might be one reason for the poor expression of mycobacterial genes in *E. coli*, even in the presence of strong *E. coli* promoters (Alldread *et al.*, 1992; Andersson & Sharp, 1996; de Miranda *et al.*, 2000). Moreover, stable secondary structures of mRNA surrounding the TIR can be formed due to the high G/C content and may be responsible for the poor expression. A number of expression systems have been developed for expressing mycobacterial proteins in bacteria that are phylogenetically closer to *M. tuberculosis,* particularly *Mycobacterium smegmatis* (Daugelat *et al.*, 2003; Ehrt *et al.*, 2005; Harth *et al.*, 1997; Mahenthiralingam *et al.*, 1993; Triccas *et al.*, 1998). However, the success of these expression systems is still limited to a few genes (Daugelat *et al.*, 2003; Ehrt *et al.*, 2005; Harth *et al.*, 1997; Mahenthiralingam *et al.*, 1993; Triccas *et al.*, 1998). Thus a versatile efficient expression system is still an urgent need.

The sequence analysis of the genome of *M. tuberculosis* (Cole *et al.*, 1998) allowed us to identify three putative DNA glycosylase genes of the Fpg/Nei family, *Rv2464c*, *Rv2924c* and *Rv3297*. The *Rv2924c* gene encodes a 32.0 kDa

formamidopyrimidine (Fpg) DNA glycosylase (MtuFpg1). The *Rv2464c* and *Rv3297* genes encode a 29.7 kDa endonuclease VIII (MtuNei1) and a 28.5 kDa endonuclease VIII (MtuNei2) respecitively. The DNA glycosylases of the Fpg/Nei family recognize and remove oxidized DNA bases as a first step in the base excision repair (BER) process responsible for removing endogenous oxidative damages from the DNA (David & Williams, 1998). As shown in *Salmonella typhimurium* (Suvarnapunya *et al.*, 2003) and *Helicobactor pylori* (O'Rourke *et al.*, 2003), the BER pathway may be involved in pathogen proliferation or colonization conferring a virulence advantage for the microorganisms, and if this is true for *M. tuberculosis*, it could provide a target for future therapy.

The potential role of codon usage and mRNA secondary structure(s) in regulating gene expression was tested using the mycobacterial *Rv3297* gene as an example. Although both are regarded as potential causes of poor expression of *M. tuberculosis* genes in *E. coli*, our results suggest that mRNA secondary structure(s), rather than codon usage, is the primary determinant in influencing translation efficiency. To our knowledge, this is the first report where a mycobacterial gene has been overexpressed in high levels in *E. coli*, and the bicistronic vector designed here should be generally applicable to improve translation efficiency of mRNAs and to achieve high-level gene expression of heterologous genes in *E. coli*.

Materials and Methods

Materials

The genomic DNA of *M. tuberculosis* H37Rv was kindly provided by Dr. Karin Eiglmeier (*Unité de Génétique Moléculaire Bactérienne*, Paris, France). The DNA sequences were retrieved from GenBank for *M. tuberculosis* H37Rv (*Rv2924c*: gi|15610061; *Rv2464c*: gi|15609601; *Rv3297*: gi|1877352). The primers were chemically synthesized by Operon Biotechnologies, Inc (Germantown, MD) and Midland Certified Reagent Company, Inc (Midland, TX). Restriction enzymes were purchased from New England Biolabs (NEB, Beverly, MA). Cloned Pfu DNA polymerase was purchased from Stratagene (Cedar Creek, TX). Expression vectors pET30a (Novagen, Madison, WI) and the bicistronic pET vector (pET30a/ORF6) constructed in this paper were used to overexpress the three mycobacterial genes in *E. coli*. Recombinant plasmids were amplified in *E. coli* TOP10 cells (Invitrogen, Carlsbad, CA) and plasmid DNA was purified using Wizard Plus Midiprep DNA Purification System (Promega, Madison, WI). BL21-Gold (DE3) (Stratagene) was used as the host strain for protein expression in *E. coli*.

Construction of the bicistronic vector

A 74-mer oligonucleotide (5' GGAATTCCATATGAAAATCGAAGCAGGTA AACTGGTACAG**aaggag**ATTAACT*GATATC*GGATCCCTCGAGCGG) and its complementary strand were designed and chemically synthesized by Midland Certified Reagent Company, Inc that included NdeI and XhoI restriction sites (underlined) at the

N and C-termini, an EcoRV restriction site (italics) for cloning the target gene and a RBS (lowercase/bold) for the target gene. An equal amount of the oligonucleotide and its complementary strand were annealed in NEBuffer 4 (50 mM potassium acetate, 20 mM Tris acetate, 10 mM magnesium acetate, 1 mM DTT, pH 7.9) and then cleaved with NdeI and XhoI. The resulting fragment was purified from a 1% agarose gel with ß-Agarase (NEB) and then cloned between the NdeI and XhoI sites of pET30a vector to create the pET30a-ORF6 vector.

Cloning of the target Rv3297 gene into the pET30a vector

The nucleotide sequence of the *Rv3297* gene was optimized with *E. coli* preferred codons and chemically synthesized by GeneScript Corporation (Piscataway, NJ) resulting in the synthesized *Rv3297* (*sRv3297*) gene. Using two primers: sRv3297-Fwd1 (5' GATCCAATCATATGCCGGAAGGTGATACCGTGTGGC) and sRv3297-Rev1 (5' TGGCTCGAGACGCTGGCACGCCGGGCACCAATAGC), the gene was amplified and cloned between NdeI-XhoI sites in pET30a vector as described below resulting in pET30a/sRv3297-His expression vector.

The *Rv3297* gene was also cloned from the genomic DNA using two primers: Rv3297-F (5' AGATATACATATGCCGGAGGGCGACACCGTCTGGCAC) and pETRv3297Rev-2 (5' CCGCTCGAGGCGCTGGCAGGCCGGGCACCAATACC) to yield the cloned *Rv3297* (*cRv3297*) gene. The PCR reaction was carried out with cloned Pfu DNA polymerase in an Air Thermo-Cycler (Idaho Technology, Salt Lake City, UT) for an initial denaturation for 5 minutes (min) at 94ºC, 40 cycles with 94ºC

for 15 second (sec), 60ºC for 45 sec and 72ºC for 2.5 min and an additional extension for 5 min at 72ºC. The PCR product was cleaved with NdeI and XhoI restriction enzymes, purified from a 1% agarose gel with ß-Agarase and then inserted between the NdeI and XhoI sites of pET30a vector to yield pET30a/cRv3297-His vector.

Additionally, in order to express MtuNei2 fused to a 37 amino acid N-terminal His/thrombin/S tag, the *cRv3297* gene was sub-cloned into pET30a vector between the KpnI and XhoI sites using a forward primer Fwd-5 (5' GAAGGAGATGGTACCATGCCGGAGGGCGACACCG) and pETRv3297Rev-2 to yield pET30a/His-cRv3297.

Cloning of the genes into the bicistronic pET30a-ORF6 vector

To improve expression of the *Rv3297* in *E. coli*, the target gene was amplified and cloned into the biscistronic pET30a-ORF6 vector as follows. Forward primers, Rv3297-F (5' GCCGGAGGGCGACACCGTCTGGCAC) and sRv3297-Fwd2 (5' GCCGGAAGGTGATACCGTGTGGC) missing the 'AT' residues in the initiation codon were used with pETRv3297Rev-2, sRv3297-Rev1, respectively, to amplify the target gene as described above. The purified and XhoI cleaved PCR products were subsequently cloned into EcoRV-XhoI digested pET30a-ORF6 resulting in pET30a-ORF6/cRv3297-His and pET30a-ORF6/sRv3297-His expression vectors.

Similarly, the cloned *Rv2924c* (*cRv2924*) gene was amplified and sub-cloned into the pET30a-ORF6 vector using two primers: Rv2924F (5' pGCCGGAGCTGCCTGAAGTCGAGG) and pETRv2924cR-2 (5'

CCGCTCGAGTTTACGTGGACGTGGCTGGCAACGC) to yield pET30a-ORF6/cRv2924. Also the synthesized *Rv2464c* (*sRv2464*) gene with *E. coli* preferred codons was also sub-cloned into the pET30a-ORF6 vector using two primers: Rv2464c Fwd-4 (5' GCCGGAGGGTCATACGCTGCATCGG) and sRv2464-Rev1 (5' CCGCTCGAGGGTCTGGCACACCGGGCACCAAAACACG) resulting in pET30a-ORF6/sRv2464.

The nucleotide sequences of all the vectors constructed above were verified by DNA sequencing (Vermont Cancer Center DNA core facility, Burlington, VT) and confirmed using Sequencher 4.2.2 (Gene Codes Corporation, Ann Arbor, MI).

mRNA secondary structures

All the mRNA secondary structures were predicted by using Mfold 3.2 (Zuker, 2003). The structures with the lowest energy were chosen for further comparison and analysis.

Protein expression

The pRARE2 plasmid, which carries seven rare-codon tRNA genes for overcoming codon usage bias, was isolated from Rosetta 2(DE3) competent cells (Novagen) and co-transformed with each of the expression plasmids when necessary. The parental vector of the pRARE2 plasmid pACYC184, which does not carry any tRNA genes, was used as a control.

Expression vectors carrying the target gene and tRNA plasmids were co-transformed into BL21-Gold (DE3) and selected on LB agar plates containing kanamycin (50 μg/mL) transformed colonies were inoculated into 60 mL of Luria Broth medium (LB) containing kanamycin (50 μg/mL) and chloramphenicol (34 μg/mL) and grown at 37°C. Since Fpg/Nei family DNA glycosylases generally contain a zinc-finger motif, in order to facilitate the proper folding of the target protein, cultures were supplemented with 10 μM of $ZnSO_4$ after reaching an OD_{600} of 0.2. To further increase the probability of correct folding, the cultures were then incubated at 22°C with shaking until an OD_{600} of 0.5 (Buchko *et al.*, 2000). After reaching the desired cell density, 30 mL of the cell culture was aliquoted and induced in a 250 mL flask with 1 mM IPTG at 22°C for 18-20 hrs.

To test for protein induction, cells were harvested by centrifugation and re-suspended in Buffer A (50 mM sodium phosphate buffer (pH 8.0), 100 mM NaCl, 10% (v/v) glycerol, 5 mM ß-mercaptoethanol) supplemented with 1 mM PMSF, 10 mM Benzamidine and 1% deoxyribonuclease I (Invitrogen). After sonication, the whole cell lysates were collected, boiled with 1x SDS-loading dye (50 mM Tris buffer (pH 8.0), 2% (w/v) SDS, 0.1% (w/v) Bromophenol Blue, 10% (v/v) glycerol, 100 mM ß-mercaptoethanol) for 5 min and analysed on a 12% SDS-PAGE. The protein bands of interest were then visualized by staining the gel with GelCode Blue Stain Reagent (Pierce, Rockford, IL) and quantitated with Quantity One (Bio-Rad, Hercules, CA). N-terminal protein sequencing (Biomolecular Resource Facility, The University of Texas Medical Branch) was used to confirm identity of the expressed proteins.

Protein purification

His-tagged proteins from 1 liter of induced cultures were purified on a 5 mL chelating HP column (GE Healthcare, Piscataway, NJ) using ÄKTA purifier (GE Healthcare) as previously described (Bandaru *et al.*, 2006). Briefly, cell lysates were prepared as described above and the soluble fraction harvested after centrifugation was loaded onto the column using Buffer A. The target protein was eluted with a linear gradient of 0-100% Buffer B (50 mM sodium phosphate buffer (pH 8.0), 150 mM NaCl, 500 mM Imidazole (pH 8.0), 10% (v/v) glycerol, 5 mM ß-mercaptoethanol) in 20 column volumes. Fractions containing target protein were identified on an SDS-PAGE, pooled and dialyzed into the storage buffer containing 25% glycerol (20 mM HEPES-NaOH (pH 7.6), 150 mM NaCl, 5 mM DTT, 25% (v/v) glycerol). After subsequent dialysis into the storage buffer containing 50% glycerol, the protein preparations were quantitated using the Bradford Assay (Bio-Rad), aliquoted and stored at -20ºC until use.

Results

Overcoming codon bias does not enhance the expression of MtuNei2 in E. coli.

A number of *E. coli* rare codons, such as GGA (Gly), CUA (Leu), CCC (Pro), AGG (Arg), CGG (Arg), and CGA (Arg) are widely used in the genome of *M. tuberculosis*. The total usage of rare codons is 12.9%, and the CGG codon accounts for 6.6% (Table 1) in the *cRv3297* gene. Tandem repeats of CGG codons which can inhibit translation are also present in the coding sequence of the *cRv3297* gene (McNulty *et al.*, 2003). We initially attempted to express *cRv3297* fused to a C-terminal hexa-his tag using pET30a vector in *E. coli*. No detectable expression was observed upon induction with IPTG for 18 h (Figure 1, compare lanes 1 and 2). Similar observations were also made when expressing the cloned *Rv2924c* and the cloned *Rv2464c* genes in the pET30a vector (data not shown).

If translation of the *cRv3297* gene was limited by the rare codons, supplementing the tRNA pool with rare codon tRNAs during expression in *E. coli* or optimizing codon usage should enhance MtuNei2 expression. To test if rare codon usage is the reason for the poor translation of the *cRv3297* gene in *E. coli*, we then co-expressed pET30a/cRv3297-His in the presence of pRARE2 tRNA plasmid**,** which carries seven rare-codon tRNA genes (Table 1). No difference in expression was observed in the presence of the pRARE2 plasmid (Figure 1, compare lanes 3 and 4 to lanes 1 and 2), Alternatively, *Rv3297* gene optimized with *E. coli* preferred codons (*sRv3297*) was synthesized, cloned into pET30a and expressed in *E. coli*. As a control,

since no rare codons are present in the *sRv3297* gene, the pACYC184 vector missing the tRNA genes was co-transformed with pET30a/sRv3297-His into *E. coli*. Analysis of the whole cell lysates on a SDS-PAGE gel showed little or no expression of MtuNei2 even after optimizing codon usage in the synthetic MtuNei2 gene (Figure 1, compare lanes 5 and 6). These results indicate that rare codon usage is not the principal factor contributing to the poor expression of MtuNei2 in *E. coli*.

The bicistronic vector enhances the expression of MtuNei2.

Another reason for the poor expression of heterologous genes in *E. coli* is that mRNA secondary structure(s) surrounding the TIR sequesters the accessibility of the RBS and/or the translation initiation codon thereby inhibiting efficient translation. Several two-cistron expression systems have been successfully applied to enhance translation efficiency thus achieving high-level expression of mammalian and archaeal genes (Birikh *et al.*, 1995; Ishida & Oshima, 1994; Makoff & Smallwood, 1990; Saito *et al.*, 1987; Schoner *et al.*, 1986; Schoner *et al.*, 1990; Schumperli *et al.*, 1982; Spanjaard *et al.*, 1989; Suzuki *et al.*, 1997). To test if secondary structure at the TIR was responsible for the poor expression of MtuNei2 in *E. coli*, we designed a bicistronic vector in which a leader sequence or an open reading frame (ORF) preceded the target gene and included the RBS for the downstream gene (Figure 2a). Also, the last nucleotide in the stop codon for the ORF overlaps with the initiation codon of the downstream gene thereby coupling the translation of the target gene with the upstream ORF sequence (Figure 2b). Further, inclusion of the RBS for the downstream gene in

the upstream ORF sequence should prevent any secondary structure(s) thereby facilitating efficient translation initiation of the target gene. The sequence for the ORF used here was derived from the first 15 amino acids of the maltose binding protein (MBP) gene and was optimized to prevent translational frameshifts especially near the RBS in the ORF sequence (data not shown). Finally, two restriction sites (EcoRV and XhoI) were used to aid in cloning the target genes readily into the biscistronic vector. Restriction digestion with EcoRV leaves a blunt end and provides the 'AT' nucleotides in the initiation codon (Figure 2a). Amplification of the target gene with an N-terminal primer missing the start codon and with a 'G' overhang restores the initiation codon for the target gene after cloning into the bicistronic vector (Figure 2b). The cloned and synthesized *Rv3297* genes were amplified and sub-cloned into pET30a and the bicistronic vector as described above resulting in plasmids, pET30a-ORF6/cRv3297-His and pET30a-ORF6/sRv3297-His, respectively.

To compare their translational efficiency, *E. coli* cells were transformed with the bicistronic expression vectors carrying the cloned or synthesized *Rv3297* genes in the presence of pRARE2 tRNA plasmid or pACYC184 plasmid missing the tRNA genes, respectively. Analysis of induced cultures by SDS-PAGE showed greater than thirty-fold increased expression from the bicistronic vectors only carrying the cloned (Figure 3, lanes 4 and 6) or the synthesized *Rv3297* gene (Figure 3, lane 10). Interestingly, the bicistronic vector carrying the *cRv3297* gene showed expression upon IPTG induction both in the absence (Figure 3, lane 4) and in the presence (Figure 3, lane 6) of the tRNA plasmid. These data support the hypothesis that disrupting the

mRNA secondary structure at the TIR improves the translation efficiency of the target gene as evidenced by expression of the *cRv3297* gene even in the absence of the tRNA plasmid (Figure 3, lane 4) and the *sRv3297* with no rare codons using bicistronic vector only (Figure 3, compare lanes 8 and 10). Also, these data suggest that the tRNA for the rare arginine CGG codon is not the rate limiting step for the poor expression of *cRv3297* in *E. coli* (Figure 3, compare lanes 4 and 6 with and without the tRNA plasmid, respectively). Finally, the enhanced expression observed with the bicistronic vector carrying the *sRv3297* gene (Figure 3, lane 10) also suggests that overcoming rare codon usage significantly improves protein expression even though the mRNA secondary structure determines the translation efficiency at the TIR. The solubility of the MtuNei2 protein was further improved by using pET30a-ORF6/sRv3297-His in Arctic Express (DE3) cells (Stratagene), which express cold-adapted chaperonins (data not shown). Apparently, the enhancement of the overall protein expression is a prerequisite for the further improvement of soluble expression.

Translational coupling using the bicistronic vector inhibits mRNA secondary structure(s) surrounding the TIR of the target gene.

In principle, optimization of codon usage also alters the G/C content of a gene which could influence accessibility and mRNA secondary structures at or near the translation initiation site (TIR). To investigate this possibility, we analyzed the G/C content of the nucleotide sequences of both the cloned and the synthesized *Rv3297* genes. Whereas the overall G/C content was similar for both, the G/C content of the N-

terminal first 50 nucleotides in the synthesized gene was 63% compared to 74% for the cloned *Rv3297* gene. While this difference in the N-terminal G/C content might explain the enhanced expression of the synthesized *Rv3297* gene in the bicistronic vector (Figure 3, lane 10), this does not account for lack of expression of the same gene when present in the pET30a vector (Figure 3, lane 8).

To further address this question, we analyzed the potential mRNA conformations of the first 165 nucleotides transcribed for the cloned and synthesized *Rv3297* in pET30a and bicistronic vectors using Mfold 3.2 (Zuker, 2003). For the bicistronic vector constructs, the 165 nucleotide sequence included not only the RBS and initiation codon for the ORF6 sequence but also the first 50 nucleotides of the target *Rv3297* where the G/C content of the cloned and synthesized *Rv3297* genes differed by more than 10%. Figure 4 illustrates the most energetically favored potential stem-loop structures involving the RBS or the AUG codons within the first 165 nucleotides. The average free energy values (ΔG) of the secondary structures in the two pET30a constructs are -56.3 kcal/mol and -53.8 kcal/mol, respectively (Figure 4a and 4b), whereas for the bicistronic vector constructs the values are -42.23 kcal/mol and -42.03 kcal/mol (Figure 4c and 4d). This decrease in the ΔG values by more than -10 kcal/mol is consistent with the observed expression of both the cloned and synthesized *Rv3297* genes in bicistronic vector constructs only. Accessibility of the RBS (shaded) in the original pET30a vector (Figure 4a and 4b) is potentially weak since the RBS in both constructs are involved in a long-range stem-loop structure with ΔG values of -1.8 kcal/mol. The existence of stable secondary structures can potentially block the

efficient binding of ribosomes and further melting of the TIR. In contrast, the mRNA secondary structures at the RBS for the ORF6 sequence in pET30a-ORF6 constructs (Figures 4c and 4d) are involved in short stem-loop structures and are thermodynamically less favored with positive ΔG values of 1.2 kcal/mol. Although the free energy values of the mRNA secondary structures in all four constructs surrounding the RBS and initiation codon of the target *Rv3297* gene are similar (Figure 4, ΔG= -9.3 to 10.6 kcal/mol), our data suggest that expression of the target gene in the bicistronic constructs is improved by the effective translation initiation of the leader ORF6 sequence and by the translational coupling mechanism (Baughman & Nomura, 1983; Makoff & Smallwood, 1990; Schumperli *et al.*, 1982).

Furthermore, the difference in the expression levels of the cloned and synthesized *Rv3297* genes in the bicistronic vector constructs (Figure 3, compare lanes 6 and 10) is somewhat surprising since the constructs did not show distinct differences in the overall ΔG values of the secondary structures in the first 165 nt of the target genes (Figure 4c and 4d). However, the ΔG values for the stem-loop structures near the initiation codon of the downstream target genes in the bicistronic vector constructs are -10.4 kcal/mol and -9.3 kcal/mol respectively (Figures 4c and 4d). This difference in the free energy values near the initiation codon of the target gene along with efficient translational initiation of the upstream ORF6 sequence may explain enhanced expression of the synthesized gene in the bicistronic vector (Figure 4d). Also, these data suggest that ribosome binding may be very sensitive to small changes in the stability of secondary structures resulting from decreased G/C content of the N-

terminus of the target gene. Additionally, it is likely that codon usage is also a contributing factor to enhanced expression of the synthesized *Rv3297* expression in the bicistronic vector.

The bicistronic vector is convenient for expressing proteins in their native forms

To test the hypothesis that secondary structure(s) at or near TIR is influencing expression, the *cRv3297* gene was sub-cloned between the KpnI and XhoI sites of the pET30a vector resulting in pET30a/His-cRv3297 for expression of MtuNei2 fused to a long 37 amino acid N-terminal His/thrombin/S tag (His-MtuNei2, 32.7 kDa). If mRNA conformation surrounding the TIR is important in controlling translational efficiency, fusion of the target gene to the long N-terminal tag should lead to efficient expression of MtuNei2. As shown in Figure 5, the N-terminal tagged *cRv3297* was expressed at high levels in the presence of the tRNA plasmid (Figure 5, lane 10). Expression of the *cRv3297* was also observed in the absence of tRNA (Figure 5, lane 8) consistent with our earlier observation (Figure 3, lane 4) that disrupting mRNA secondary structure at the TIR is a pre-requisite for translational efficiency. However, MtuNei2 is non-functional with an N-terminal tag since the N-terminal proline of Fpg/Nei family DNA glycosylases is the catalytic residue that initiates excision of oxidized bases from DNA (Zharkov *et al.*, 1997). On the other hand, the cloned and synthesized *Rv3297* genes were efficiently expressed (Figure 3, lanes 6 and 10) by using the pET30a-ORF6 bicistronic vector and the soluble proteins catalyze excision of 5-hydroxyuracil from single-stranded DNA containing this lesion (data not shown).

The bicistronic vector was also successfully used to overexpress the other two mycobacterial DNA glycosylases in this family, MtuNei1 (encoded by *sRv2464*) and MtuFpg1 (encoded by *cRv2924*) (Figure 6). Both proteins catalyze excision of oxidized bases from DNA (data not shown). This further demonstrates that the bicistronic vector described here should be generally applicable to achieve high-level expression of heterologous genes in *E. coli*.

Discussion

Codon usage vs. efficiency of translational initiation

Theoretically, gene expression can be improved by overcoming codon bias. However, to our knowledge, there is only one group who showed improved expression of *M. tuberculosis* antigen 85B by replacing rare codons with *E. coli* preferred codons through silent mutations (Lakey *et al.*, 2000). A greater than 50-fold improvement of 85B expression was achieved after two pairs of substitutions, one of which was located at nucleotides 7 to 15 within the TIR. Numerous examples of increased protein expression through altered codon usage have been reported. In many of these examples, codon usage at the N-terminal region was found to be of particular importance (Goldman *et al.*, 1995; Griswold *et al.*, 2003). Taken together, these observations suggest that the codon optimization associated alteration in mRNA conformations should not be ignored.

As was also reported by other groups (Zhou *et al.*, 2004), we found that either supplementing the rare tRNAs or using the codon-optimized gene failed to enhance the expression of the target protein in the pET30a vector (Figure 1), although expression was enhanced in the bicistronic vector (Figures 3 and 5). These results strongly suggest that factors, other than codon usage, play a dominant role in controlling translation efficiency.

mRNA secondary structure vs. efficiency of translational initiation

A prevalent mechanism for translational control of gene expression is through modulation of mRNA secondary structure(s) at the TIR. The stable mRNA secondary structures involving the RBS or the AUG codons were not only observed in the pET30a construct containing MtuNei2 gene (*cRv3297*) but also in the pET30a constructs containing the MtuFpg1 (*cRv2924*) and MtuNei1 (*sRv2464*) genes (data not shown), which we believe to be responsible for the poor expression of these heterologous genes in *E. coli* as well. The stem-loop structure surrounding the translational start site as well as other potential stable mRNA secondary structures due to the high G/C content should also be present in *M. tuberculosis* and may function as translational regulators. However, to our knowledge, there are no reports in the literature describing how *M. tuberculosis* may utilize these structures.

Many reports suggest that the possibility of inhibitory base-pairing in mRNAs surrounding the TIR could be decreased by designing a proper leader sequence as the first cistron in a bicistronic vector (Birikh *et al.*, 1995; Ishida & Oshima, 1994; Ishida & Oshima, 2002; Schoner *et al.*, 1984; Schoner *et al.*, 1990; Spanjaard & van Duin, 1989; Suzuki *et al.*, 1997). However, whether or not mRNA secondary structure(s) is necessarily involved in the translational coupling of bicistronic vectors remains unclear. Our results comparing the predicted mRNA secondary structures between the bicistronic vector and pET30a vector suggest that the decrease in the stability of the secondary structure surrounding the TIR using the bicistronic constructs is consistent with the enhancement of expression levels. Therefore, our results strongly support the

hypothesis that mRNA secondary structure(s) surrounding the TIR, rather than codon usage, is the critical first step in effecting translation efficiency.

Based on our observations, we conclude that there are two rate-limiting steps during translation: the initiation step (ribosome binding) and the elongation step (protein synthesis). Translation efficiency first depends on the rate of translation initiation, where the limiting factor is the stability of mRNA secondary structure(s) surrounding the TIR. After efficient translational initiation, codon usage controls the elongation kinetics due to different tRNA availabilities.

Summary

Our results demonstrate that rare codon usage is not the primary barrier to efficient translation of the *Rv3297* gene. Instead, mRNA secondary structures appear to play a significant role in controlling translational efficiency. The bicistronic vector, pET30a-ORF6, greatly enhanced the expression of MtuNei2 most likely by destabilizing the mRNA secondary structures surrounding the TIR thereby increasing translation initiation.

Although the versatility of bicistronic systems has been questioned (Makoff & Smallwood, 1990), a number of bicistronic vectors have been successfully used to express foreign genes whose expression was originally poor in *E. coli* (Birikh *et al.*, 1995; Ishida & Oshima, 1994; Makoff & Smallwood, 1990; Spanjaard *et al.*, 1989; Suzuki *et al.*, 1997). We show here that the bicistronic system is particularly useful to achieve expression of soluble and active proteins, which carry the proline residue at the

P_1' position and require a free N-terminus. To our knowledge, this is the first report to show high level expression of mycobacterial genes using a bicistronic vector in *E. coli*.

Acknowledgements

We thank Dr. Stewart Cole and Dr. Karin Eiglmeier of *Unité de Génétique Moléculaire Bactérienne*, Paris, France, for generously providing the genomic DNA of *M. tuberculosis* H37Rv and Melissa Batonick for technical assistance. We are grateful to Dr. Jeffery Bond for the helpful comments on the manuscript. The study was supported by National Institutes of Health grant PO1 CA098993 awarded by the National Cancer Institute. Yin Guo was supported by a research fellowship from DoE EPSCoR DE-F602-00ER45828.

References

Alldread, R. M., Nicholls, D. J., Sundaram, T. K., Scawen, M. D. & Atkinson, T. (1992). Overexpression of the *Thermus aquaticus* B malate dehydrogenase-encoding gene in *Escherichia coli*. *Gene* 114, 139-143.

Andersson, G. E. & Sharp, P. M. (1996). Codon usage in the *Mycobacterium tuberculosis* complex. *Microbiology (Reading, England)* 142 (Pt 4), 915-925.

Bandaru, V., Blaisdell, J. O. & Wallace, S. S. (2006). Oxidative DNA glycosylases: recipes from cloning to characterization. *Methods in enzymology* 408, 15-33.

Baughman, G. & Nomura, M. (1983). Localization of the target site for translational regulation of the L11 operon and direct evidence for translational coupling in *Escherichia coli*. *Cell* 34, 979-988.

Birikh, K. R., Lebedenko, E. N., Boni, I. V. & Berlin, Y. A. (1995). A high-level prokaryotic expression system: synthesis of human interleukin 1 alpha and its receptor antagonist. *Gene* 164, 341-345.

Buchko, G. W., Hess, N. J., Bandaru, V., Wallace, S. S. & Kennedy, M. A. (2000). Spectroscopic studies of zinc(II)- and cobalt(II)-associated *Escherichia coli* formamidopyrimidine-DNA glycosylase: extended X-ray absorption fine structure evidence for a metal-binding domain. *Biochemistry* 39, 12441-12449.

Cole, S. T., Brosch, R., Parkhill, J. & other authors (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393, 537-544.

Daugelat, S., Kowall, J., Mattow, J., Bumann, D., Winter, R., Hurwitz, R. & Kaufmann, S. H. (2003). The RD1 proteins of *Mycobacterium tuberculosis*: expression in *Mycobacterium smegmatis* and biochemical characterization. *Microbes and infection / Institut Pasteur* 5, 1082-1095.

David, S. S. & Williams, S. D. (1998). Chemistry of Glycosylases and Endonucleases Involved in Base-Excision Repair. *Chem Rev* 98, 1221-1262.

de Miranda, A. B., Alvarez-Valin, F., Jabbari, K., Degrave, W. M. & Bernardi, G. (2000). Gene expression, amino acid conservation, and hydrophobicity are the main factors shaping codon preferences in *Mycobacterium tuberculosis* and *Mycobacterium leprae*. *Journal of molecular evolution* 50, 45-55.

de Smit, M. H. & van Duin, J. (1990). Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. *Proceedings of the National Academy of Sciences of the United States of America* 87, 7668-7672.

Ehrt, S., Guo, X. V., Hickey, C. M., Ryou, M., Monteleone, M., Riley, L. W. & Schnappinger, D. (2005). Controlling gene expression in mycobacteria with anhydrotetracycline and Tet repressor. *Nucleic acids research* 33, e21.

Goldman, E., Rosenberg, A. H., Zubay, G. & Studier, F. W. (1995). Consecutive low-usage leucine codons block translation only when near the 5' end of a message in *Escherichia coli*. *Journal of molecular biology* 245, 467-473.

Griswold, K. E., Mahmood, N. A., Iverson, B. L. & Georgiou, G. (2003). Effects of codon usage versus putative 5'-mRNA structure on the expression of *Fusarium solani* cutinase in the *Escherichia coli* cytoplasm. *Protein expression and purification* 27, 134-142.

Hall, M. N., Gabay, J., Debarbouille, M. & Schwartz, M. (1982). A role for mRNA secondary structure in the control of translation initiation. *Nature* 295, 616-618.

Hannig, G. & Makrides, S. C. (1998). Strategies for optimizing heterologous protein expression in *Escherichia coli*. *Trends in biotechnology* 16, 54-60.

Harth, G., Lee, B. Y. & Horwitz, M. A. (1997). High-level heterologous expression and secretion in rapidly growing nonpathogenic mycobacteria of four major *Mycobacterium tuberculosis* extracellular proteins considered to be leading vaccine candidates and drug targets. *Infection and immunity* 65, 2321-2328.

Iserentant, D. & Fiers, W. (1980). Secondary structure of mRNA and efficiency of translation initiation. *Gene* 9, 1-12.

Ishida, M. & Oshima, T. (1994). Overexpression of genes of an extreme thermophile *Thermus thermophilus*, in *Escherichia coli* cells. *Journal of bacteriology* 176, 2767-2770.

Ishida, M. & Oshima, T. (2002). Effective structure of a leader open reading frame for enhancing the expression of GC-rich genes. *Journal of biochemistry* 132, 63-70.

Kane, J. F. (1995). Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Current opinion in biotechnology* 6, 494-500.

Kurland, C. & Gallant, J. (1996). Errors of heterologous protein expression. *Current opinion in biotechnology* 7, 489-493.

Lakey, D. L., Voladri, R. K., Edwards, K. M., Hager, C., Samten, B., Wallis, R. S., Barnes, P. F. & Kernodle, D. S. (2000). Enhanced production of recombinant *Mycobacterium tuberculosis* antigens in *Escherichia coli* by replacement of low-usage codons. *Infection and immunity* 68, 233-238.

Mahenthiralingam, E., Draper, P., Davis, E. O. & Colston, M. J. (1993). Cloning and sequencing of the gene which encodes the highly inducible acetamidase of *Mycobacterium smegmatis*. *Journal of general microbiology* 139, 575-583.

Makoff, A. J., Oxer, M. D., Romanos, M. A., Fairweather, N. F. & Ballantine, S. (1989). Expression of tetanus toxin fragment C in *E. coli*: high level expression by removing rare codons. *Nucleic acids research* 17, 10191-10202.

Makoff, A. J. & Smallwood, A. E. (1990). The use of two-cistron constructions in improving the expression of a heterologous gene in E. coli. *Nucleic acids research* 18, 1711-1718.

McNulty, D. E., Claffee, B. A., Huddleston, M. J., Porter, M. L., Cavnar, K. M. & Kane, J. F. (2003). Mistranslational errors associated with the rare arginine codon CGG in *Escherichia coli*. *Protein expression and purification* 27, 365-374.

Merchant, S. S., Prochnik, S. E., Vallon, O. & other authors (2007). The Chlamydomonas genome reveals the evolution of key animal and plant functions. *Science (New York, NY* 318, 245-250.

O'Rourke, E. J., Chevalier, C., Pinto, A. V., Thiberge, J. M., Ielpi, L., Labigne, A. & Radicella, J. P. (2003). Pathogen DNA as target for host-generated oxidative stress: role for repair of bacterial DNA damage in *Helicobacter pylori* colonization. *Proceedings of the National Academy of Sciences of the United States of America* 100, 2789-2794.

Saito, Y., Ishii, Y., Niwa, M. & Ueda, I. (1987). Direct expression of a synthetic somatomedin C gene in *Escherichia coli* by use of a two-cistron system. *Journal of biochemistry* 101, 1281-1288.

Schoner, B. E., Hsiung, H. M., Belagaje, R. M., Mayne, N. G. & Schoner, R. G. (1984). Role of mRNA translational efficiency in bovine growth hormone expression in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* 81, 5403-5407.

Schoner, B. E., Belagaje, R. M. & Schoner, R. G. (1986). Translation of a synthetic two-cistron mRNA in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* 83, 8506-8510.

Schoner, B. E., Belagaje, R. M. & Schoner, R. G. (1990). Enhanced translational efficiency with two-cistron expression system. *Methods in enzymology* 185, 94-103.

Schumperli, D., McKenney, K., Sobieski, D. A. & Rosenberg, M. (1982). Translational coupling at an intercistronic boundary of the *Escherichia coli* galactose operon. *Cell* 30, 865-871.

Sorensen, M. A., Kurland, C. G. & Pedersen, S. (1989). Codon usage determines translation rate in *Escherichia coli*. *Journal of molecular biology* 207, 365-377.

Spanjaard, R. A., van Dijk, M. C., Turion, A. J. & van Duin, J. (1989). Expression of the rat interferon-alpha 1 gene in *Escherichia coli* controlled by the secondary structure of the translation-initiation region. *Gene* 80, 345-351.

Spanjaard, R. A. & van Duin, J. (1989). Translational reinitiation in the presence and absence of a Shine and Dalgarno sequence. *Nucleic acids research* 17, 5501-5507.

Studer, S. M. & Joseph, S. (2006). Unfolding of mRNA secondary structure by the bacterial translation initiation complex. *Molecular cell* 22, 105-115.

Suvarnapunya, A. E., Lagasse, H. A. & Stein, M. A. (2003). The role of DNA base excision repair in the pathogenesis of *Salmonella* enterica serovar *Typhimurium*. *Molecular microbiology* 48, 549-559.

Suzuki, T., Tanaka, Y., Ishida, M., Ishizuka, M., Yamagishi, A. & Oshima, T. (1997). Screening of a mutant plasmid with high expression efficiency of GC-rich leuB gene of an extreme thermophile, *Thermus thermophilus*, in *Escherichia coli*. *Journal of biochemistry* 121, 1031-1034.

Triccas, J. A., Parish, T., Britton, W. J. & Gicquel, B. (1998). An inducible expression system permitting the efficient purification of a recombinant antigen from *Mycobacterium smegmatis*. *FEMS microbiology letters* 167, 151-156.

Zahn, K. (1996). Overexpression of an mRNA dependent on rare codons inhibits protein synthesis and cell growth. *Journal of bacteriology* 178, 2926-2933.

Zharkov, D. O., Rieger, R. A., Iden, C. R. & Grollman, A. P. (1997). NH_2-terminal proline acts as a nucleophile in the glycosylase/AP-lyase reaction catalyzed by *Escherichia coli* formamidopyrimidine-DNA glycosylase (Fpg) protein. *The Journal of biological chemistry* 272, 5335-5341.

Zhou, Z., Schnake, P., Xiao, L. & Lal, A. A. (2004). Enhanced expression of a recombinant malaria candidate vaccine in *Escherichia coli* by codon optimization. *Protein expression and purification* 34, 87-94.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic acids research* 31, 3406-3415.

Table 1 Comparison of rare codon usage in the cloned and synthesized *Rv3297* (MtuNei2) genes.

AminoAcid	RareCodon	pRARE2	Percentage of Total Codons in the Genes	
			Cloned	Synthesized
Gly	GGA	+	0.78%	0.00%
Leu	CUA	+	0.78%	0.00%
Ile	AUA	+	0.00%	0.00%
Pro	CCC	+	0.78%	0.00%
Arg	AGG	+	0.39%	0.00%
	AGA	+	0.00%	0.00%
	CGG	+	6.64%	0.00%
	CGA	-	1.95%	0.00%
Total			12.89%	0.00%

Figure 1 Expression of *Rv3297* (MtuNei2) in the pET30a vector in the presence of pRARE2 tRNA plasmid. Uninduced (Lanes 1, 3 and 5) and induced (Lanes 2, 4 and 6) whole cell lysates of transformed BL21-Gold (DE3) cells were analyzed on a 12% SDS-PAGE gel. Lane M, Novagen Perfect Protein™ Marker (15-150 kDa). Lane P, purified MtuNei2-His (29.6 kDa).

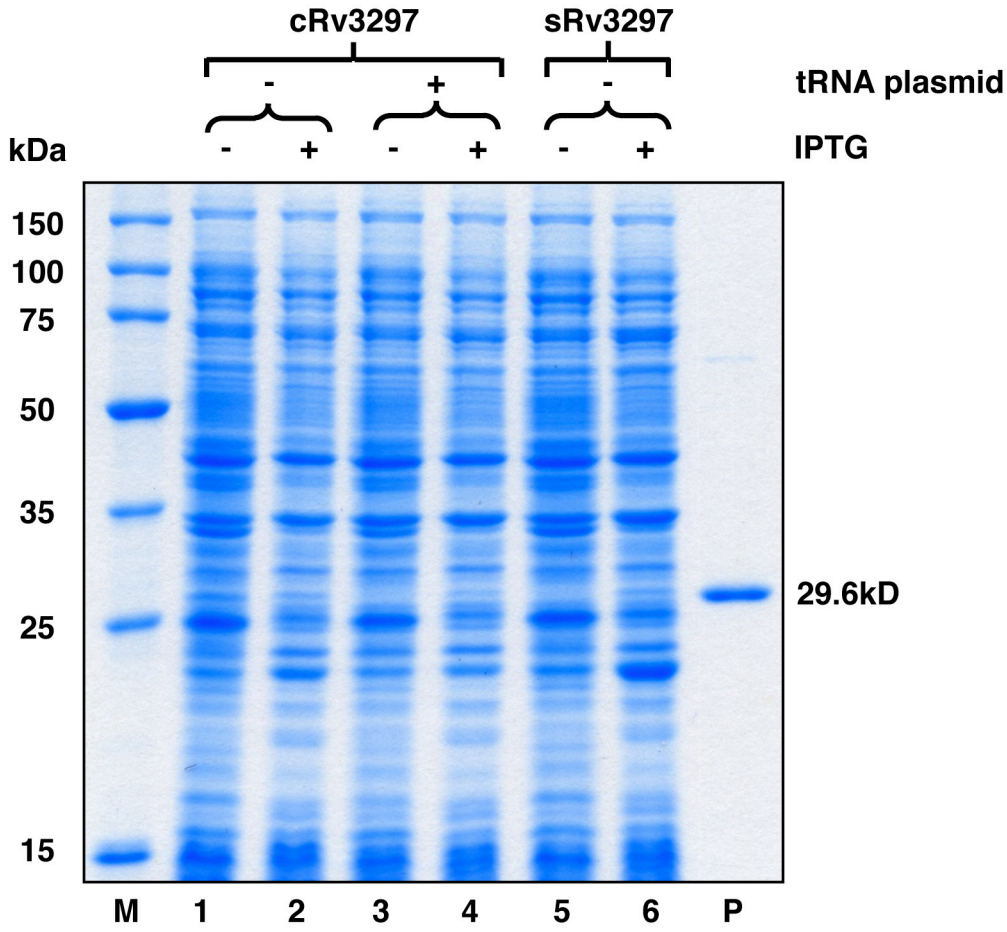


Figure 2 Schematic of the bicistronic vectors used in this study. a. Parental pET30a-ORF6 vector, b. *Rv3297* cloned into the pET30a-ORF6. RBS for the upstream leader sequence and the downstream target gene are boxed. The cloning sites for the *Rv3297* (MtuNei2) gene are underlined in the parental vector. Nucleotides 'AT' of the EcoRV recognition sequence overlap with the stop codon for the upstream ORF6 and the initiation codon for the downstream target gene.

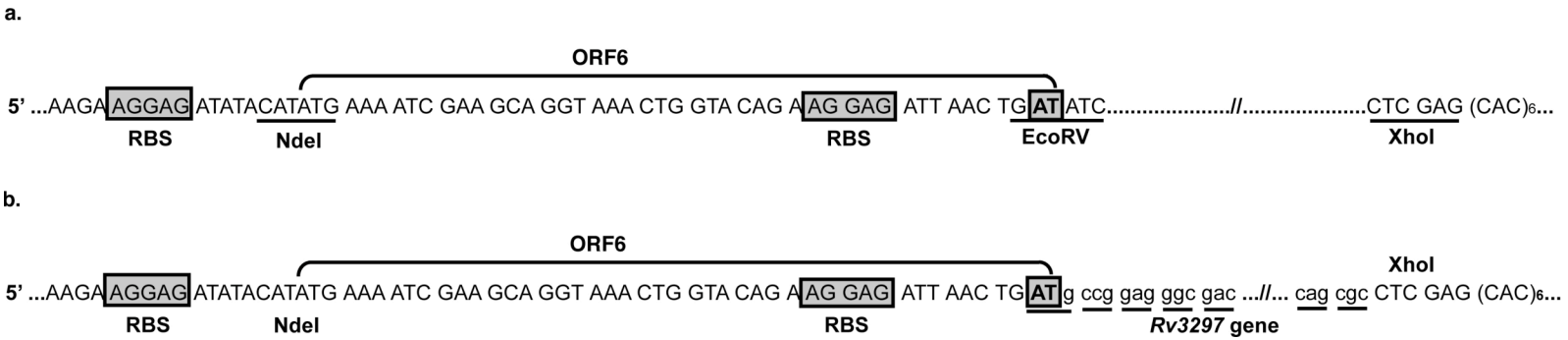

a.
ORF6
5' ...AAGA AGGAG ATATACATATG AAA ATC GAA GCA GGT AAA CTG GTA CAG A AG GAG ATT AAC TG AT ATC.....//.....CTC GAG (CAC)6...
RBS
NdeI
RBS
EcoRV
XhoI
b.
ORF6
XhoI
5' ...AAGA AGGAG ATATACATATG AAA ATC GAA GCA GGT AAA CTG GTA CAG A AG GAG ATT AAC TG ATg ccg gag ggc gac ...//... cag cgc CTC GAG (CAC)6...
RBS
NdeI
RBS
Rv3297 gene

Figure 3 Induction of *Rv3297* (MtuNei2) expression in pET30a and the pET30a-ORF6 bicistronic vector in the presence of pRARE tRNA plasmid or pACYC184 missing the tRNA genes. Uninduced (Lanes 1, 3, 5, 7 and 9) and induced (Lanes 2, 4, 6, 8 and 10) whole cell lysates of transformed BL21-Gold (DE3) cells were analyzed on a 12% SDS-PAGE. Lane M, Novagen Perfect Protein™ Marker (15-150 kDa). Lane P, purified MtuNei2-His (29.6 kDa).

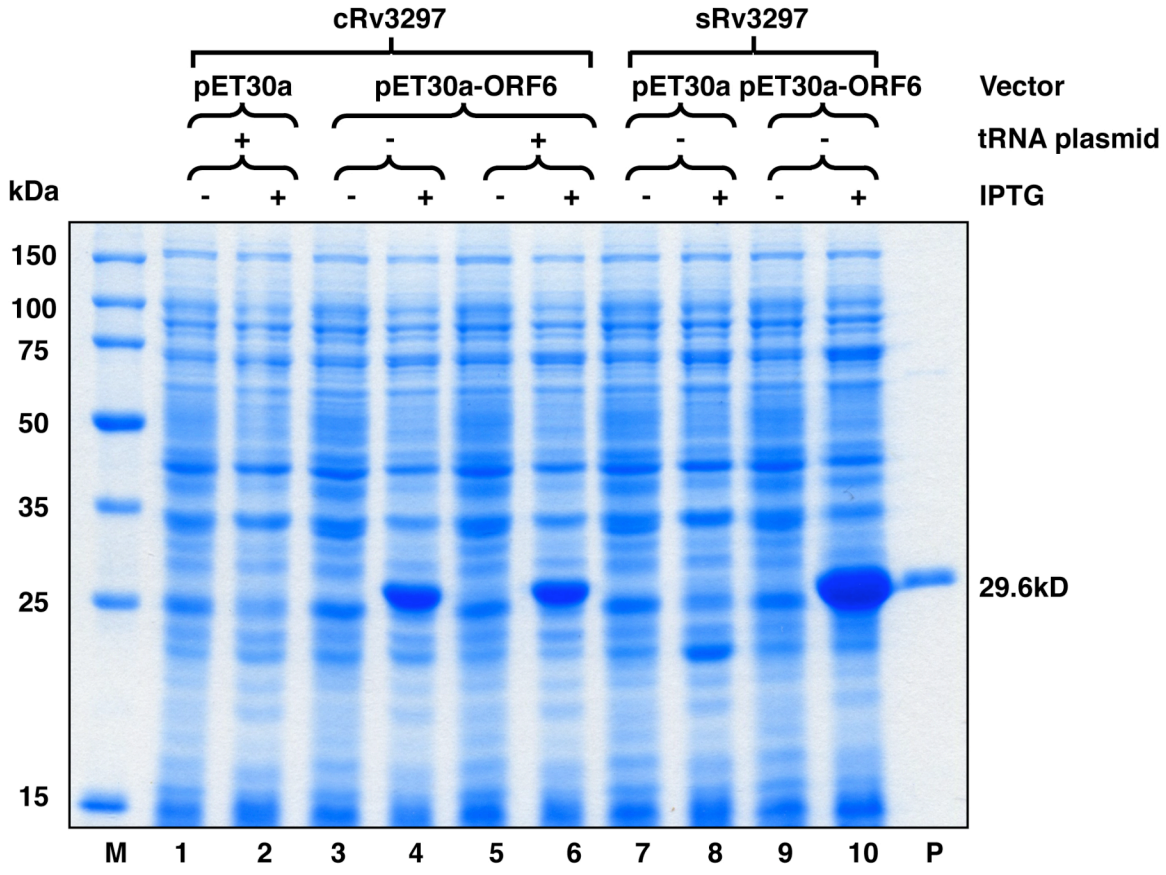


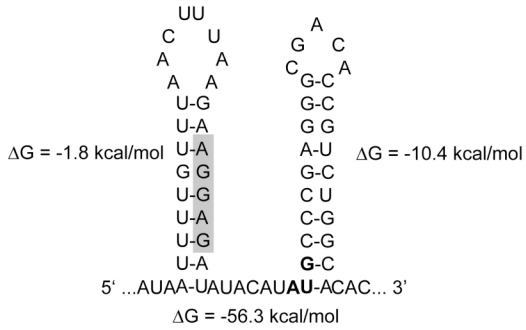
Figure 4 Partial mRNA conformations of the first 165 nucleotides transcribed in pET30a and the pET30a-ORF6 constructs. a) pET30a/cRv3297-His, b) pET30a/sRv3297-His, c) pET30a-ORF6/cRv3297-His, d) pET30a-ORF6/sRv3297-His. Only secondary structures near the RBS (shaded) and AUG start codons (bold) are shown. The free energy (ΔG) values for the individual stem-loop structures (aside) and for the folded 165 nt mRNA (below) are shown.

pET30a

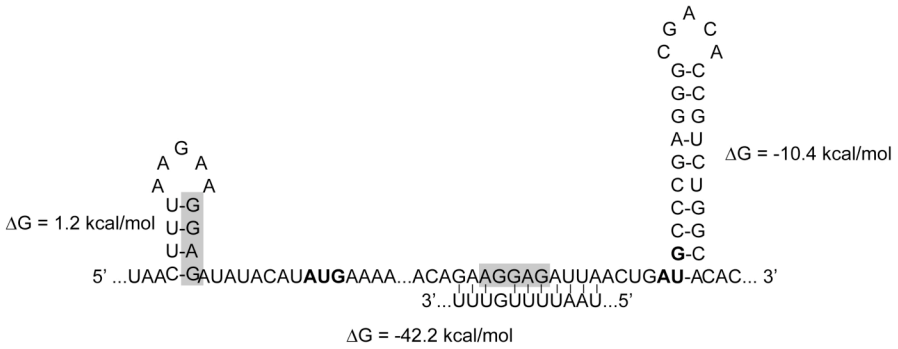
pET30a-ORF6 (Bicistronic)

Cloned

a.

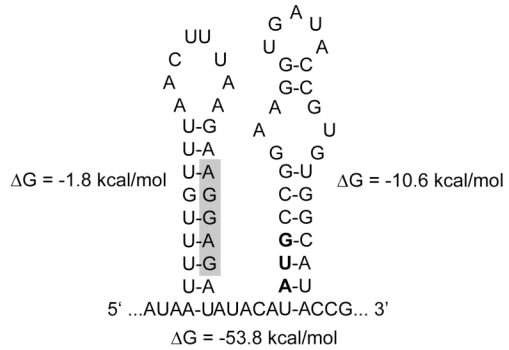


c.



Synthesized

b.



d.

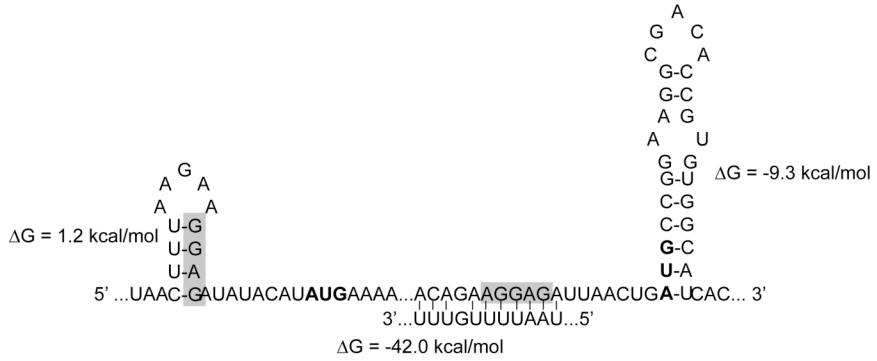


Figure 5 Expression of cloned *Rv3297* (MtuNei2) as N- or C-terminal hexa his tagged fusion proteins. Uninduced (Lanes 1, 3, 5, 7 and 9) and induced (Lanes 2, 4, 6, 8 and 10) whole cell lysates of transformed BL21-Gold (DE3) cells were analyzed on a 12% SDS-PAGE. Lane M, Novagen Perfect Protein™ Marker (15-150κDa). Lane P, purified MtuNei2-His (29.6 kDa) .

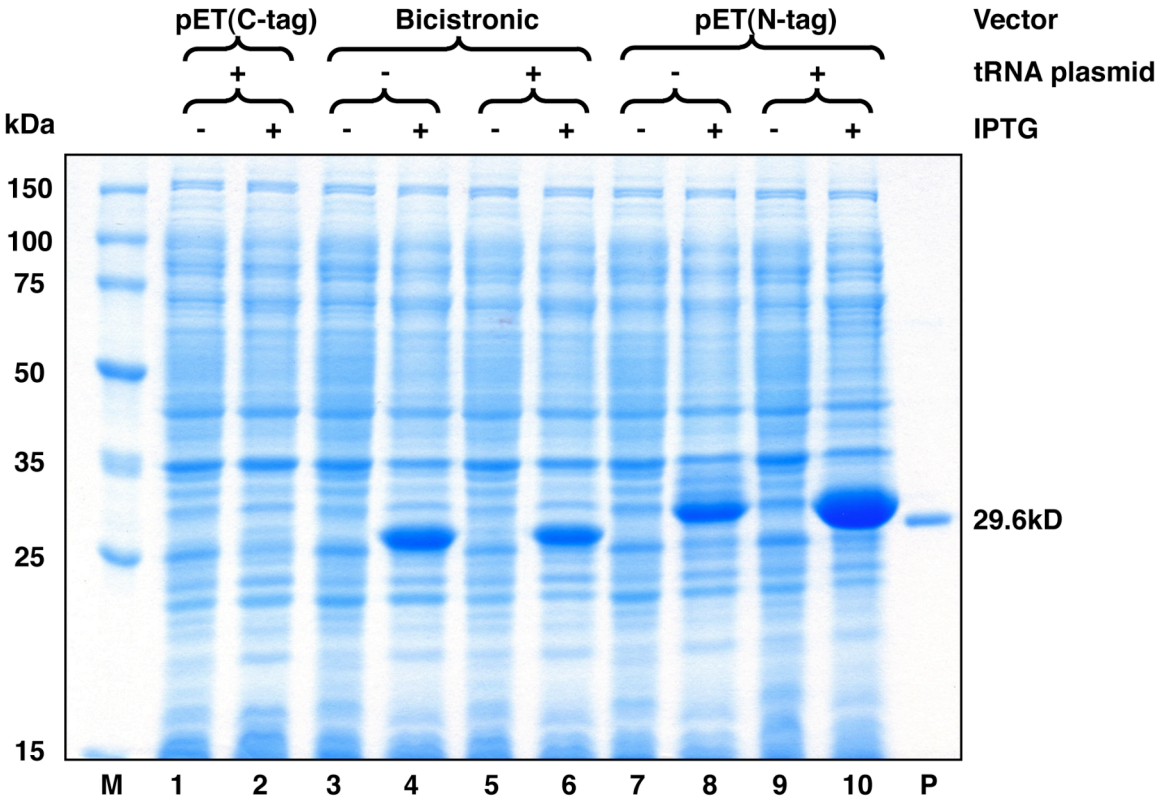
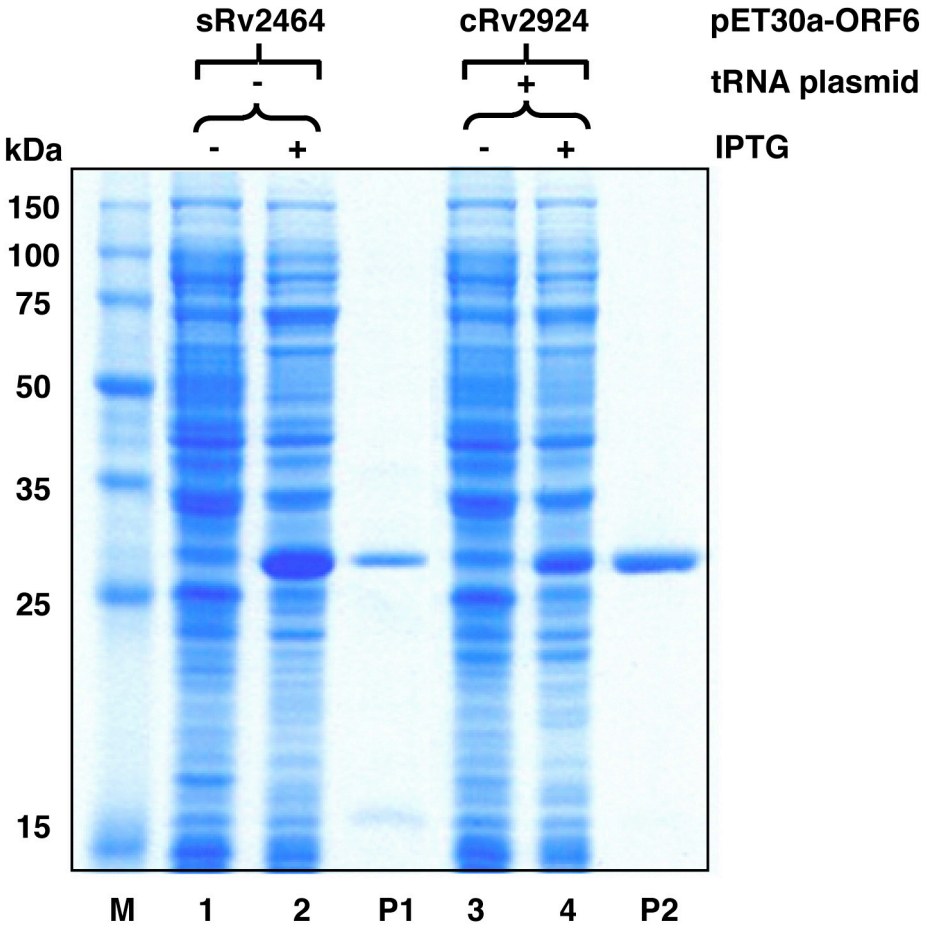


Figure 6 Induction of *sRv2464* (MtuNei1) and *cRv2924* (MtuFpg1) expression in the pET30a-ORF6 bicistronic vector. Uninduced (Lanes 1 and 3) and induced (Lanes 2 and 4) whole cell lysates of transformed BL21-Gold (DE3) cells were analyzed on a 12% SDS-PAGE. Lane M, Novagen Perfect Protein™ Marker (15-150 kDa). Lane P1, purified MtuNei1-His (30.8 kDa). Lane P2, purified MtuFpg1-His (33.0 kDa).



The oxidative DNA glycosylases of *Mycobacterium tuberculosis* exhibit different substrate preferences from their *Escherichia coli* counterparts

Yin Guo[a], Viswanath Bandaru[b], Pawel Jaruga[c,d], Xiaobei Zhao[e], Cynthia J. Burrows[e], Shigenori Iwai[f], Miral Dizdaroglu[c], Jeffrey P. Bond[a], Susan S. Wallace[a]

[a]Department of Microbiology and Molecular Genetics

The Markey Center for Molecular Genetics

University of Vermont

Stafford Hall, 95 Carrigan Dr.

Burlington, VT 05405-0068

Tel.: 802-656-2164

Fax: 802-656-8749

[b] Current address: Mylan Pharmaceuticals, 3711 Collins Ferry Rd, Morgantown, WV 26505

[c] Chemical Science and Technology Laboratory, National Institute of Standards and Technology, Gaithersburg, MD 20899

[d] Department of Clinical Biochemistry, Collegium Medicum, Nicolaus Copernicus University, Bydgoszcz, Poland

[e] Department of Chemistry, University of Utah, Salt Lake City, UT 84112-0850

[f] Division of Chemistry, Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan

Abstract

The DNA glycosylases that remove oxidized DNA bases fall into two general families: the Fpg/Nei family and the Nth superfamily. Based on protein sequence alignments, we identified four putative Fpg/Nei family members, as well as a putative Nth protein in *Mycobacterium tuberculosis* H37Rv. All four Fpg/Nei proteins were successfully overexpressed using a bicistronic vector created in our laboratory. The MtuNth protein was also overexpressed in soluble form. The substrate specificities of the purified enzymes were characterized *in vitro* with oligodeoxynucleotide substrates containing single lesions. Some were further characterized by gas chromatography/mass spectrometry (GC/MS) analysis of products released from γ-irradiated DNA. MtuFpg1 has a substrate specificity similar to that of EcoFpg. Both EcoFpg and MtuFpg1 are more efficient at removing spiroiminodihydantoin (Sp) than 7,8-dihydro-8-oxoguanine (8-oxoG). However, MtuFpg1 shows a substantially increased opposite base discrimination compared to EcoFpg. MtuFpg2 contains only the C-terminal domain of an Fpg protein and has no detectable DNA binding activity or DNA glycosylase/lyase activity and thus appears to be a pseudogene. MtuNei1 recognizes oxidized pyrimidines on both double-stranded and single-stranded DNA and exhibits uracil DNA glycosylase activity. MtuNth recognizes a variety of oxidized bases, including urea, 5,6-dihydrouracil (DHU), 5-hydroxyuracil (5-OHU), 5-hydroxycytosine (5-OHC) and methylhydantoin (MeHyd). Both MtuNei1 and MtuNth excise thymine glycol (Tg); however, MtuNei1 strongly prefers the (5*R*) isomers,

whereas MtuNth recognizes only the (5*S*) isomers. MtuNei2 did not demonstrate activity *in vitro* as a recombinant protein, but like MtuNei1 when expressed in *Escherichia coli*, it decreased the spontaneous mutation frequency of both the *fpg mutY nei* triple and *nei nth* double mutants, suggesting that MtuNei2 is functionally active *in vivo* recognizing both guanine and cytosine oxidation products. The kinetic parameters of the MtuFpg1, MtuNei1 and MtuNth proteins on selected substrates were also determined and compared to those of their *E. coli* homologs.

Keyword (Complete): Base excision repair; Oxidative DNA glycosylases; *Mycobacterium tuberculosis*

Introduction

Oxidative damage to DNA is produced by endogenous reactive oxygen species (ROS) generated during normal cellular metabolism and by exposing cells to ionizing radiation and other free radical-generating systems. These lesions include single strand breaks, abasic sites and a plethora of oxidative base damages that can be potentially mutagenic and/or cytotoxic [for a review see (Wallace, 2002)] . The DNA glycosylases function in the first step of the base excision repair (BER) pathway responsible for removing oxidatively induced base damages from DNA [for reviews see (David *et al.*, 2007; Lindahl & Barnes, 2000; Wallace *et al.*, 2003; Wilson *et al.*, 2003)]. A DNA glycosylase cleaves the N-glycosyl bond between the sugar and the damaged base creating an abasic (AP) site. Most DNA glycosylases that recognize oxidized bases are bifunctional and contain lyase activity which then cleaves the phosphodiester backbone at the AP site leaving either an α, β-unsaturated aldehyde (β-elimination) or a phosphate (β, δ-elimination) group [for reviews see (Krokan *et al.*, 2000; Lindahl & Barnes, 2000)]. The DNA glycosylases that remove oxidative DNA damages fall into two general families based on structural and sequence homology, the Fpg/Nei family and the Nth superfamily [for reviews see (Aravind *et al.*, 1999; Eisen & Hanawalt, 1999)]. The Fpg/Nei family contains formamidopyrimidine DNA glycosylase (Fpg, also called MutM), which mainly recognizes 7,8-dihydro-8-oxoguanine (8-oxoG), 4,6-diamino-5-formamidopyrimidine (FapyA) and 2,6-diamino-4-hydroxy-5-formamidopyrimidine (FapyG) and endonuclease VIII (Nei), which removes oxidized pyrimidines and FapyA [for reviews see (David *et al.*, 2007; Dizdaroglu, 2005; Lindahl

& Barnes, 2000; Wallace *et al.*, 2003; Wilson *et al.*, 2003)]. The Nth superfamily includes endonuclease III (Nth), which has overlapping substrate specificity with Nei and is primarily responsible for removing oxidized pyrimidines, Ogg, whose primary substrates are 8-oxoG and FapyG paired with C; MutY, a monofunctional glycosylase that removes A when paired with 8-oxoG, and AlkA, also a monofunctional glycosylase that removes alkylated bases [for reviews see (David *et al.*, 2007; Lindahl & Barnes, 2000; Wallace *et al.*, 2003; Wilson *et al.*, 2003)]. Nth is found in the three major kingdoms, bacteria, eukaryotes and archaea. While Fpg proteins are widely distributed among the bacteria and plants, Nei homologs are sparsely distributed across phyla, and are only found in γ-proteobacteria, actinobacteria and metazoans (Wallace *et al.*, 2003).

Oxidative stress also plays an important role in the host's innate immune response. Macrophages generate ROS and reactive nitrogen species (RNS) resulting in lethal damage to the DNA of pathogenic microorganisms. *Mycobacterium tuberculosis*, the causative agent of tuberculosis, which is responsible for 3 million deaths annually and causes latent infection in one-third of the world's population, manages to survive and replicate in its host's macrophages. Therefore, the ability to repair DNA damages caused by exposure to ROS in macrophages is likely to play a particularly important role in pathogen proliferation or colonization conferring a virulence advantage for mycobacteria, as has been observed with other pathogens (O'Rourke *et al.*, 2003; Suvarnapunya *et al.*, 2003).

The DNA repair mechanisms of *M. tuberculosis*, as well as its response to DNA

damage has distinguishing features compared to some other bacteria. First, the genome sequences of *M. tuberculosis*, *Mycobacterium leprae*, and *Mycobacterium smegmatis* indicate that these microbes lack the mismatch repair pathway (Dos Vultos *et al.*, 2009; Mizrahi & Andersen, 1998). Furthermore, *M. tuberculosis* regulates most of its inducible DNA repair genes using a novel RecA-independent mechanism, even though in most bacterial systems studied so far a RecA-dependent mechanism is the major DNA damage response mechanism (Gamulin *et al.*, 2004; Rand *et al.*, 2003). *M. tuberculosis* possesses genes for nucleotide excision repair (NER), BER, recombination and the SOS response (Mizrahi & Andersen, 1998). Among the major DNA repair pathways, the BER and the NER pathways may contribute significantly to maintaining the genomic integrity of mycobacteria. The role of the NER pathway in *M. tuberculosis* has been examined by using *uvrB* mutants, which were highly susceptible to acidified nitrite and UV light but not to several sources of reactive oxygen intermediates (ROI) (Darwin & Nathan, 2005). They were also markedly attenuated for survival in mice. This attenuation was reversed in mice lacking either nitric oxide synthase 2 (iNOS) or both iNOS and phagocyte oxidase, further indicating the important role that the NER pathway plays in resisting both NOS and ROS. *Ung* deficiency led to an enhanced sensitivity to reactive nitrogen intermediates (RNI) and a decrease in survival of *M. smegmatis* and *Pseudomonas aeruginosa* in macrophages, which underscores the importance of the BER pathway for these two high G+C gram-positive organisms (Venkatesh *et al.*, 2003). Disrupting the *fpg* gene in *M. smegmatis* enhanced the spontaneous mutation frequency and made the microbe more susceptible to hydrogen

peroxide (Jain *et al.*, 2007). Using transposon site hybridization (TraSH), members of the BER pathway: endonuclease IV (*nfo*), uracil glycosylase (*ung*) and exonuclease III (*xthA*) were shown to be important for the *in vivo* growth of *M. tuberculosis* during initial infection (Sassetti & Rubin, 2003); however, no DNA glycosylases that remove oxidatively modified DNA bases were shown to be necessary for its survival *in vivo* probably due to the redundant activities of these enzymes (Rengarajan *et al.*, 2005).

Four putative Fpg/Nei family members, as well as one putative Nth protein were identified in *M. tuberculosis* H37Rv, MtuFpg1, MtuFpg2, MtuNei1, MtuNei2 and MtuNth (Cole *et al.*, 1998). Some progress has already been made towards defining the biochemical properties of these DNA glycosylases. Sidorenko *et al.* cloned two of the four Fpg/Nei enzymes: MtuFpg2 and MtuNei1 (*Mtu*-Nei2 in their study) and biochemically characterized the partially purified MtuNei1 (Sidorenko *et al.*, 2008). More recently, Olsen *et al.* partially characterized the activity of purified MtuFpg1 and pointed out that the *in vivo* transcription of the gene encoding MtuFpg1 decreased with decreasing the length of the intergenic repeat upstream to the gene (Olsen *et al.*, 2009). However, no extensive biochemical characterization of the mycobacterial DNA glycosylases was done.

We have successfully overexpressed and purified all four MtuFpg/Nei proteins and the MtuNth protein. The substrate specificities of MtuFpg1, MtuNei1 and MtuNth were characterized and compared to their *E. coli* counterparts *in vitro* with oligodeoxynucleotide substrates containing single lesions. They were further characterized by gas chromatography/mass spectrometry (GC/MS) analysis of products

released from γ-irradiated DNA. The kinetic parameters of the MtuFpg1, MtuNei1 and MtuNth proteins on selected substrates as well as those of their *E. coli* homologs were also determined. Both Nei paralogs, MtuNei1 and MtuNei2, were also shown to be functional *in vivo*. Taken together, our data indicate that the DNA glycosylases of *M. tuberculosis* that recognize oxidatively induced DNA damages have substrate specificities that overlap but are distinct from their *E. coli* counterparts. The exception is MtuFpg2 that appears to be a pseudogene.

Materials and Methods

Materials

The genomic DNA of *M. tuberculosis* H37Rv was kindly provided by Dr. Karin Eiglmeier and Dr. Nadine Homoré (Institute Pasteur, Paris, France). The gene sequences were retrieved from GenBank for *M. tuberculosis* H37Rv (*Rv0944*: gi|15608084, *Rv2464c*: gi|15609601, *Rv2924c*: gi|15610061, *Rv3297*: gi|1877352, *Rv3674c*: gi|57117142). Based on the sequence alignment, two of the five were identified as putative Fpg homologs: MtuFpg1 (*Rv2924c*) and MtuFpg2 (*Rv0944*), two fell into the Nei clade: MtuNei1 (*Rv2464c*) and MtuNei2 (*Rv3297*), and one is an Nth homolog: MtuNth (*Rv3674c*).

Protein expression and purification

The genes encoding MtuFpg1 (*Rv2924c*) and MtuFpg2 (*Rv0944*) were individually cloned between the EcoRV and XhoI sites of the bicistronic vector pET30a-ORF6, as previously described to allow active expression of the C-terminal hexa-his tagged proteins (Guo *et al.*, 2008). In order to avoid the potential inhibitory effect of rare codons during translation, the pRARE2 plasmid (Novagen, Madison, WI) carrying seven rare-codon tRNA genes (GGA, CUA, AUA, CCC, AGG, AGA, CGG) was co-transformed with each expression plasmid into *E. coli* ER2566 (Fpg-). The expression of MtuFpg1 and MtuFpg2 were respectively induced with 0.5 mM IPTG in Luria Broth (LB) medium (50 μg/ml kanamycin, 34 μg/ml chloramphenicol and 10 μM

$ZnSO_4$) at an OD_{600} of 0.5. The cultures were further grown at 20 °C for overnight. To purify these target proteins, cells from 2 liters of induced cultures were harvested by centrifugation and re-suspended in chelating column Buffer A (50 mM sodium phosphate buffer (pH 8.0), 100 mM NaCl, 10% (v/v) glycerol, 5 mM ß-mercaptoethanol) supplemented with 1 mM PMSF, 10 mM benzamidine and 1% deoxyribonuclease I (Invitrogen). The clear cell lysates were harvested after sonication and centrifugation then loaded onto a 5 mL HiTrap chelating HP column (GE Healthcare, Piscataway, NJ) using ÄKTA purifier (GE Healthcare). The target protein was eluted with a linear gradient of 0-100% chelating column Buffer B (50 mM sodium phosphate buffer (pH 8.0), 150 mM NaCl, 500 mM imidazole (pH 8.0), 10% (v/v) glycerol, 5 mM ß-mercaptoethanol) in 20 column volumes. Fractions containing the protein were identified by SDS-PAGE, pooled and loaded onto a 1 mL HiTrap SP FF column (GE Healthcare) with SP column buffer A (20 mM HEPES (pH 7.6), 50 mM NaCl. 10% (v/v) glycerol, 5 mM ß-mercaptoethanol). The protein was further eluted with a linear gradient of 0-100% SP column Buffer B (20 mM HEPES (pH 7.6), 1 M NaCl. 10% (v/v) glycerol, 5 mM ß-mercaptoethanol). The purified target protein was pooled and stored in the MtuFpg storage buffer (20 mM HEPES·KOH (pH 7.6), 250 mM NaCl, 2 mM TCEP, 50% glycerol) at -20 °C.

The genes for MtuNei1 (*Rv2464c*) and MtuNei2 (*Rv3297*) were optimized with *E. coli* preferred codons and chemically synthesized by GeneScript Corporation (Piscataway, NJ). The two synthesized genes were also cloned between the EcoRV and XhoI sites of pET30a-ORF6 vector. Expression was induced in *E. coli* Arctic Express

(DE3) cells (Stratagene, Cedar Creek, TX) with 1 mM IPTG in the presence of 10 μM $ZnSO_4$. The target protein was first purified with a 5 mL HiTrap chelating HP column (Buffer A: 50 mM sodium phosphate buffer (pH 8.0), 500 mM NaCl, 10% (v/v) glycerol, 15 mM ß-mercaptoethanol; Buffer B: 50 mM sodium phosphate buffer (pH 8.0), 500 mM NaCl, 500 mM imidazole (pH 8.0), 10% (v/v) glycerol, 15 mM ß-mercaptoethanol) followed by a 1 mL HiTrap phenyl HP column (GE Healthcare) (Buffer A: 50 mM sodium phosphate buffer (pH 8.0), 1 M $(NH_4)_2SO_4$, 10% (v/v) glycerol, 15 mM ß-mercaptoethanol; Buffer B: 50 mM sodium phosphate buffer (pH 8.0), 10% (v/v) glycerol, 15 mM ß-mercaptoethanol). The purified protein was then stored in the MtuNei storage buffer (20 mM HEPES·NaOH (pH 7.6), 150 mM NaCl, 2 mM TCEP, 50% glycerol) at -20 °C.

The MtuNth (*Rv3674c*) gene was cloned between the NdeI and SalI sites of the pET28a vector in order to express the protein fused with an N-terminal hexa-his tag. To avoid the potential inhibitory effect of rare codons during translation, the pRARE2 plasmid (Novagen) was co-transformed with the pET28a containing MtuNth. The expression of MtuNth was induced in *E. coli* ER2566 (Fpg-) with 0.1 mM IPTG at 20 °C overnight. Similar to MtuNei1 and MtuNei2, the MtuNth protein was also purified with a 5 mL HiTrap chelating HP column (GE Healthcare) followed by a 1 mL HiTrap SP FF column. The purified MtuNth was stored in Nth storage buffer (50 mM HEPES·NaOH (pH 7.6), 50 mM NaCl, 2 mM TCEP, 50% glycerol).

The identity of each purified protein was confirmed by N-terminal protein sequencing (Biomolecular Resource Facility, The University of Texas Medical

Branch). The control enzymes (EcoFpg, EcoNei, NEIL1 and EcoNth) were prepared in our laboratory as previously described (Bandaru *et al.*, 2006). The protein concentrations were determined with Coomassie Plus Protein Assay Reagent (Pierce, Rockford, IL).

Substrates

The 35-mer and 51-mer oligodeoxynucleotides carrying various oxidatively induced lesions and the complementary strands were purchased from Midland Certified Reagent Co. (Midland, TX). The sequence of the 35-mer damage-containing oligodeoxynucleotides is: 5' - TGTCAATAGCAAGXGGAGAAGTCAATCGTGAGTCT - 3', where X = 7,8-dihydro-8-oxoguanine (8-oxoG), thymine glycol (Tg), 5,6-dihydrouracil (DHU), 5,6-dihydrothymine (DHT), 5-hydroxyuracil (5-OHU), 5-hydroxycytosine (5-OHC), uracil (U) and furan (F). Another 35-mer complementary strand was designed based on the work of Dou *et al.* (Dou *et al.*, 2008): 5' - AGACTCACGATTGACTTCTCCGGAACGATAACTGT - 3'. This oligodeoxynucleotide contained a noncomplementary sequence for annealing to the 5-OHU-containing 35-mer to generate the 3' forked structure. A 30-mer oligodeoxynucleotide carrying either of the two *cis* isomers of Tg (5' - CTCGTCAGCATCTZCATCATACAGTCAGTG - 3', where Z = 5*R*-Tg ((5*R*,6*S*) plus (5*R*,6*R*)) or 5*S*-Tg ((5*S*,6*R*) plus (5*S*,6*S*)) as well as the complementary strand was synthesized (Iwai, 2001; Shimizu *et al.*, 2006). Three 51-mer oligodeoxynucleotides

carrying 8-oxoG, U or 5-OHU at position 26 from the 5'-end and a series of complementary strands were synthesized as previously described (Dou *et al.*, 2003). These complementary strands contained the sequences for either creating base-paired duplex substrates or producing bubble structures with 5, 11 or 19 unpaired bases flanking the damage. A 30-mer oligodeoxynucleotide was synthesized carrying either guanidinohydantoin (Gh) or spiroiminodihydantoin (Sp) (5' - TGTTCATCATGCGTCYTCGGTATATCCCAT - 3', where Y = Gh or Sp) (Kornyushyna *et al.*, 2002).

1 pmol of each damage-containing strand was 5' end-labeled with [γ-^{32}P] ATP by T4 polynucleotide kinase (New England Biolabs, Beverly, MA) at 37 °C for 20 min. The reaction was terminated by addition of EDTA followed by heat inactivation. The labeled oligodeoxynucleotide was then ethanol precipitated and diluted with 9 pmol of the unlabeled lesion-containing oligodeoxynucleotide in the substrate buffer (20 mM Tris·HCl (pH 8.0), 50 mM NaCl and 1 mM EDTA) and used as the single-stranded substrate. To create the duplex substrate, the lesion-containing strand was annealed to an equal amount of its complementary strand by heating to 95 °C then slowly cooling down to room temperature. To create a urea substrate, double-stranded oligodeoxynucleotides containing Tg were subjected to alkali hydrolysis by dialyzing against 40 mM sodium phosphate buffer (pH 12.0) and 1 mM EDTA for overnight then dialyzed into the substrate buffer for a further 24 hours (Ide *et al.*, 1985). To create an abasic (AP) site, duplex or single-stranded oligodeoxynucleotides containing U were treated with EcoUdg (New England Biolabs) at room temperature for 15 min.

Determining the active fraction of the enzyme preparations

The Schiff base assay was used to determine the active fraction of each mycobacterial enzyme preparation. An increasing amount of each enzyme preparation was incubated with 1 μM of its optimal substrate in the presence of 50 mM of sodium borohydride ($NaBH_4$). The concentration of NaCl used in the Schiff base reactions was adjusted correspondingly to keep the final sodium concentration the same as in the optimal reaction buffer for each enzyme. After 30 min of incubation at 37 °C, the reactions were terminated with SDS loading buffer (50 mM Tris buffer (pH 8.0), 2% (w/v) SDS, 0.1% (w/v) bromophenol blue, 10% (v/v) glycerol, 100 mM ß-mercaptoethanol) and boiled for 5 min before loading on 12% SDS-PAGE. The gel bands were detected by using a phosphoimage screen (Bio-Rad, Hercules, CA) and quantified with Quantity-one (Bio-Rad). We plotted the quantity of trapped substrate formed in the Schiff base assay versus the total enzyme concentration used in the assay. The slope of the linear region before the plateau was reached indicates the active fraction of the enzyme preparation (Blaisdell & Wallace, 2007). Similarly, the active fraction of each control enzyme was determined using the molecular accessibility assay (Blaisdell & Wallace, 2007). A table of the enzyme preparations used is shown in the Supplementary Material (Table S1). Notably, MtuFpg2 (*Rv0944*) showed no detectable DNA glycosylase/lyase activity hence could not form a DNA-enzyme complex in the presence of $NaBH_4$ (data not shown). Except for MtuFpg2 (*Rv0944*), the amount of

each enzyme used in the following assays was corrected for the percent active fraction determined for each preparation.

DNA glycoslyase/lyase assays

For duplex substrates, 15 nM labeled substrate was incubated with either 15 nM control enzymes or an increasing amount of the mycobacterial enzymes (1.5 nM, 15 nM or 150 nM) in their optimal reaction buffers supplemented with 0.1 mg/mL of BSA. The reaction buffers are as follows: MtuFpg1, EcoFpg, EcoNei and NEIL1 (20 mM Tris·HCl (pH 8.0), 75 mM NaCl and 1 mM EDTA); MtuNei1, MtuNth and EcoNth (20 mM Tris·HCl (pH 8.0), 50 mM NaCl and 1 mM EDTA). The reactions for EcoFpg, MtuFpg1, EcoNth and MtuNth were carried out at 37 °C for 30 min. Since MtuNei1 is temperature-sensitive and unstable at 37 °C, the reactions for EcoNei, NEIL1 and MtuNei1 were carried out at room temperature for 30 min. For single-stranded substrates, 15 nM of each labeled substrate was incubated with 15 nM of each enzyme at the corresponding optimal reaction condition as described above. All the reactions were terminated with an equal volume of formamide stop loading buffer (98% formamide, 5 mM EDTA, 0.1% xylene cyanol and 0.1% bromophenol blue). The cleaved and uncleaved substrates were separated on a 12% denaturing polyacrylamide gel and quantified as above.

Electrophoretic mobility shift assays

The ^{32}P-labeled 35-mer duplex substrates (25 nM) containing either 8-oxoG, furan or the normal base G at the corresponding position were incubated with either 25 nM of MtuFpg1 (*Rv2924c*) or 5 μM of MtuFpg2 (*Rv0944*) in the MtuFpg1 optimal reaction buffer supplemented with 5% glycerol and 0.1 mg/mL BSA at 25 °C for 30 min. The reactions were electrophoresed on a 4% native polyacrylamide gel (4% acrylamide, 0.5×TBE, 1% glycerol) at 4 °C.

Enzyme kinetics

Kinetic parameters were determined for MtuFpg1, MtuNei1, MtuNth and their *E. coli* homologs, EcoFpg, EcoNei and EcoNth, on their representative duplex substrates. The reactions were terminated at appropriate time points with formamide stop loading buffer. Kin-Tek Rapid Quench was used when the reactions were too fast to be terminated manually. The k_{obs} and $(k_{cat}/K_m)^{app}$ were solved under single-turnover conditions and multiple-turnover conditions respectively. Under single-turnover conditions, enzymes with concentrations higher than the dissociation constants (K_d) were incubated with either 0.5 nM or 5 nM labeled substrate in their optimal reaction buffers supplemented with 0.1 mg/mL of BSA. Reactions were terminated at the appropriate time points ranging from 0.1 seconds to 30 minutes with formamaide stop loading buffer. The data were analyzed with either the one phase or two phase exponential association model of Prism 4 (GraphPad Software, Inc. La Jolla, CA). Similarly, under multiple turnover conditions, either 5 nM or 10 nM enzymes were incubated with increasing amounts of substrate ([enzyme]:[substrate] ratios were 1:2,

1:5, 1:10 and 1:20 respectively in each group of data). Initial velocities (v_0) were determined and plotted versus different substrate concentrations ([S]) with Prism 4 (GraphPad Software, Inc. La Jolla, CA). When the substrate concentration was very low, the $(k_{cat}/K_m)^{app}$ values could be calculated using velocity (v), substrate concentration [S] and enzyme concentration [E_T]:

$$v = (\frac{k_{cat}}{K_m})^{app}[S][E_T] \quad \text{(Eq. 1)}$$

Gas chromatography/mass spectrometry (GC/MS) analysis

Calf thymus DNA was γ-irradiated with 40 Gray in N_2O saturated phosphate buffer solution to create the DNA damages as previously described (Dizdaroglu *et al.*, 2000; Kathe *et al.*, 2009). Stable isotope-labeled analogs of the damaged DNA bases as internal standards were added to 50 μg of γ-irradiated DNA. The samples were then incubated with 2 μg of active protein (MtuFpg1, EcoFpg, MtuNth or EcoNth) in 50 μl buffer (50 mM phosphate buffer (pH7.4), 100 mM KCl, 1 mM EDTA and 0.1 mM DTT) at 37 °C for 1 h. The reactions for EcoNei and MtuNei1 were performed at room temperature. For each enzyme, the concentration used was in the linear portion of the progress curve. The subsequent GC/MS analysis was explained previously (Dizdaroglu *et al.*, 2000; Jaruga *et al.*, 2004) and briefly described below. The GC/MS data were collected from three independent experiments using three independently prepared DNA substrates.

GC/MS data consist of quantities, Q_{epr}, each associated with an enzyme, *e*, a product, *p*, and a replicate index, *r*. The quantity of each product released by each enzyme, S_{ep}, was calculated using

$$S_{ep} \equiv \frac{1}{|R_{ep}|}\sum_{r\in R_{ep}} Q_{epr} - \frac{1}{|R_{0p}|}\sum_{r\in R_{0p}} Q_{0pr} \qquad \text{(Eq. 2)}$$

where no enzyme is indicted by e=0, R_{ep} is the set of replicate indices, and $|R_{ep}|$ is the number of replicates (Kathe *et al.*, 2009).

Spontaneous mutation frequency assays

E. coli KL16 wild type strain was generously provided by Dr. Bernard Weiss, Department of Pathology, Emory University. KL16 Δ*fpg*::*Amp* Δ*nei*::*Cam mutY*::*Tet* and KL16 Δ*nei*::*Cam nth*::*Kan* were created in our laboratory as described earlier (Blaisdell *et al.*, 1999). The mutant strains carrying the λDE3 lysogen were generated using the λDE3 lysogenization kit (Novagen) to allow heterologous expression with pET vectors. The pET21a-ORF6 vector containing either MtuNei1 (*Rv2464c*) or MtuNei2 (*Rv3297*) and the pET28a containing the EcoNei gene were respectively electroporated into KL16 Δ*fpg*::*Amp* Δ*nei*::*Cam mutY*::*Tet* (DE3) and KL16 Δ*nei*::*Cam nth*::*Kan* (DE3). A low-level of heterologous expression was maintained with 0.1 mM IPTG and confirmed by SDS-PAGE. The activities of EcoNei and MtuNei1 were confirmed by cell extract assays as described earlier (Blaisdell *et al.*, 1999; Jiang *et al.*, 1997a). Overnight cultures of each strain containing 0.1 mM IPTG and appropriate

antibiotics were diluted 50-fold into fresh LB medium containing only 0.1 mM IPTG and further grown at 37 °C until an OD_{600} of 0.5. The mid-log-phase cultures were plated onto LB-agar plates with and without 100 μg/ml of rifampin (Rif) and 0.1 mM IPTG. The Rif-resistant (Rif^R) mutant fraction for each strain was calculated.

Results

MtuFpg1 prefers double-stranded DNA containing oxidized purines.

To determine the substrate specificity of MtuFpg1, we tested its activity on double-stranded oligodeoxynucleotides containing a single base lesion or an AP site with EcoFpg as a control. Like other Fpg proteins, MtuFpg1 is bifunctional with glycosylase and lyase activities. As shown in Fig. 1 and supplementary Fig. S1, MtuFpg1 recognizes an AP site opposite G, T and C equally well but is slightly less active on AP:A. MtuFpg1 also cleaved the phosphodiester backbone at the AP site via β, δ-elimination leaving a phosphate at the 3' end. Besides an AP site, 8-oxoG:C is the best substrate for MtuFpg1 as expected. MtuFpg1 also discriminates among the bases opposite with 8-oxoG: C > G ~ T >> A (Fig. 1 and supplementary Fig. S2), consistent with a previous report on *M. smegmatis* Fpg (Jain *et al.*, 2007) and similar to EcoFpg (Tchou *et al.*, 1994b). Notably, while EcoFpg exhibited slower enzymatic activity on 8-oxoG:A than 8-oxoG:C (Leipold *et al.*, 2000a), MtuFpg1 showed barely detectable activity on 8-oxoG:A (Fig. 1 and (Olsen *et al.*, 2009)). This observation was confirmed as shown in Fig. 2. Complete cleavage of 8-oxoG:A was achieved by increasing the amount of EcoFpg, but cleavage of 8-oxoG:A by MtuFpg1 was barely detectable even at an enzyme to substrate ratio of 10:1.

8-oxoG can be further oxidized to yield the hydantoin lesions, spiroiminodihydantoin (Sp) and guanidinohydantoin (Gh) (Burrows *et al.*, 2002; Henderson *et al.*, 2003). Unlike EcoFpg which recognizes Gh and Sp equally well

(Hazra *et al.*, 2001; Krishnamurthy *et al.*, 2007a; Leipold *et al.*, 2000a), MtuFpg1 prefers Sp:C over Gh:C (Fig. 1). On the other hand, similar to EcoFpg, MtuFpg1 was less active on pyrimidine substrates including DHU, 5-OHU, 5-OHC and urea. MtuFpg1 did not recognize Tg:A or DHT:A. Taken together, our biochemical data indicate that the substrate preference of MtuFpg1 is similar to that of EcoFpg, which primarily recognizes modified purines in double-stranded oligodeoxynucleotides; however, MtuFpg1 exhibits a much stronger opposite base discrimination against 8-oxoG:A than does EcoFpg.

Rv0944 encoding MtuFpg2 appears to be a pseudogene.

MtuFpg2, a 16 kDa protein, contains only the intact C-terminal domain of an Fpg protein according to the sequence alignment. This protein lacks the N-terminal proline (Pro2), which is the catalytic residue for nearly all the Fpg/Nei proteins (Zharkov *et al.*, 1997), the Glu3 which is highly conserved and required for glycosylase activity of EcoFpg (Lavrukhin & Lloyd, 2000), as well as the Lys53 (EcoNei position) which together with the H2TH and zinc finger motifs create strong hydrogen bonding to stabilize the damaged DNA (Gilboa *et al.*, 2002; Zharkov *et al.*, 2002). The loss of these critical amino acids as well as part of the N-terminal domain in MtuFpg2 suggested that this protein had lost its original DNA glycosylase/lyase activity and, as expected, we were not able to demontrate either activity (data not shown). Considering that the C-terminal domain of an Fpg protein functions in binding to DNA (Castaing *et al.*, 1992; Fromme & Verdine, 2002; Gilboa *et al.*, 2002; O'Connor *et al.*, 1993;

Sugahara *et al.*, 2000; Tchou *et al.*, 1993; Tchou *et al.*, 1994b), we tested whether MtuFpg2 contains DNA binding ability using the electrophoretic mobility shift assay (shown in supplementary Fig. S3). While MtuFpg1 bound to the duplex substrates containing 8-oxoG, furan (F) and to undamaged DNA, MtuFpg2 did not show any DNA binding ability. Since MtuFpg2 has no detectable DNA binding activity or DNA glycosylase/lyase activity, the gene encoding MtuFpg2 appears to be a pseudogene.

MtuNei1 recognizes oxidized pyrimidines in both double-stranded and single-stranded DNA.

We first tested the substrate specificity of MtuNei1 on double-stranded substrates and compared its activity to EcoNei and human NEIL1 (Fig. 3A). As expected, the best substrates for MtuNei1 were Tg, DHU, urea, and an AP site. MtuNei1 cleaved the phosphodiester backbone via β, δ-elimination, as do other Fpg/Nei family members. Similar to EcoNei (Jiang *et al.*, 1997b; Melamede *et al.*, 1994) and NEIL1 (Bandaru *et al.*, 2002; Hazra *et al.*, 2002a; Hazra *et al.*, 2002b), MtuNei1 exhibited weak activity on DHT, 5-OHU and 5-OHC. We did not detect any activity of MtuNei1 against 8-oxoG at the enzyme concentrations used; however, NEIL1 did show weak activity against 8-oxoG, which is consistent with previous observations (Bandaru *et al.*, 2002; Hazra *et al.*, 2002a; Morland *et al.*, 2002). Like EcoNei (Hailer *et al.*, 2005b; Hazra *et al.*, 2001) and NEIL1 (Hailer *et al.*, 2005a; Krishnamurthy *et al.*, 2008), MtuNei1 could also excise Gh and Sp, which suggests a backup role for repairing oxidized purines.

Interestingly, MtuNei1 contains uracil DNA glycosylase activity (Fig. 3B); weak uracil DNA glycosylase activity was also observed for EcoNei and NEIL1 when the reactions were incubated at 37 °C instead of room temperature (data not shown). MtuNei1 was able to excise uracil (U) from both double-stranded and single-stranded substrates (Fig. 3B and 4) with higher efficiency than EcoNei and NEIL1 at both temperatures. It recognizes uracil opposite T, G and C equally well and slightly better than U:A. Uracil arises in DNA by deamination of cytosine or by incorporation of dUMP, so among these substrates, U:A and U:G are the biological substrates. This redundant uracil DNA glycosylase activity may be particularly important for maintaining the integrity of the G/C-rich genome of *M. tuberculosis*.

MtuNei1 was also capable of incising the oxidized pyrimidines, Tg, DHU, DHT, 5-OHU and 5-OHC from single-stranded oligodeoxynucleotides (Fig. 4A), but MtuNei1 was less active on single-stranded substrates than on double-stranded substrates. We quantified the activity of MtuNei1 on single-stranded oligodeoxynucleotides containing single lesions in order to compare it to EcoNei and NEIL1 (Fig. 4B). NEIL1 protein excised the single-stranded substrates significantly better than either EcoNei or MtuNei1; however, MtuNei1 was significantly more active on Tg in single-stranded DNA compared to the other lesion-containing single-stranded substrates.

Since the activity of MtuNei1 on single-stranded substrates suggests a possible function in replication-associated DNA repair, we tested the activity of MtuNei1 on bubble and 3’ forked structures, which mimic the transient structures generated during

DNA replication and transcription, and compared its activity to EcoNei and NEIL1. As stated above, the three enzymes are more active on double-stranded DNA than single-stranded DNA (Fig. 5). Moreover, duplex DNA containing 5-OHU is not as good a substrate for MtuNei1 as for EcoNei or NEIL1. Each enzyme had comparable activity on the three bubble structures (Fig. 5A), although their preferences toward differerent-sized bubbles are different (Fig. 5A and (Dou *et al.*, 2003)). MtuNei1 showed comparable activity on duplex DNA and the 3' forked structure, as did NEIL1 (Dou *et al.*, 2008); on the other hand, EcoNei preferred duplex DNA (Fig. 5B). The different cleavage levels observed with the 35-mer duplex substrates (Fig. 5A) compared to the 51-mer duplex substrates (Fig. 5B) may due to the different lengths and sequence contexts. Regardless, MtuNei1 was significantly more active on lesions in duplex DNA, bubble and 3' forked structures than in single-stranded DNA. Notably MtuFpg1 and MtuNth showed no activity against any lesion in single-stranded DNA, unlike EcoFpg which recognizes 8-oxoG in single-stranded DNA (Fig. 4A and (Ishchenko *et al.*, 1999a; Ishchenko *et al.*, 1999b)); essentially complete cleavage was observed for all enzymes tested on the single-stranded oligodeoxynucleotide containing an AP site (Fig. 4A).

Comparison of the kinetic parameters of the M. tuberculosis and E. coli oxidative DNA glycosylases

The kinetic parameters, k_{obs} (min^{-1}) and $(k_{cat}/K_m)^{app}$ ($min^{-1}nM^{-1}$), of MtuFpg1, MtuNei1 and MtuNth as well as their *E. coli* homologs were determined for selected

double-stranded substrates under their optimal reaction conditions and shown in Table 1. Overall, the rate constants of the *M. tuberculosis* enzymes are lower than those of the *E. coli* enzymes.

MtuFpg1 showed specificity similar to EcoFpg but with the k_{obs} values about 10-fold lower than those of EcoFpg. 8-OxoG and its oxidized product Sp are much better substrates for the two Fpg enzymes than Tg and DHU. Consistent with previous observations (Krishnamurthy *et al.*, 2007a), we found that EcoFpg was at least 100-fold more efficient at removing Sp than 8-oxoG probably because of the tighter binding to Sp. Under these same conditions, MtuFpg1 was only 10-fold more efficient at removing Sp than 8-oxoG. Moreover, MtuFpg1 repaired Tg and DHU much less efficiently than EcoFpg. Therefore, MtuFpg1 appears to be more purine-specific than EcoFpg.

The kinetic rates of MtuNei1 are 100-fold lower than those of EcoNei. In terms of enzymatic specificity, 5-OHC is the worst substrate for MtuNei1 among those tested, similar to EcoNei. Although EcoNei did not show measurable activity against 8-oxoG under these conditions, the activity of EcoNei against Sp was similar to that of EcoFpg (Bandaru *et al.*, 2007; Hazra *et al.*, 2001).

The kinetic parameters of MtuNth are comparable to those of EcoNth. The k_{obs} of MtuNth for 5-OHC is even higher than that of EcoNth, although the overall efficiency of MtuNth repairing 5-OHC is much lower, probably because of low binding affinity. The best substrate for MtuNth is DHU. The analysis of the single turnover kinetics of the Nei and Nth enzymes on Tg fits a double exponential model much better

than a single exponential model, which suggested that these enzymes were stereospecific (see below).

GC/MS analysis of products released by the M. tuberculosis proteins compared to their E. coli counterparts

The product spectrum of damages released from γ-irradiated DNA after incubation with the mycobacterial DNA glycosylases as well as their *E. coli* homologs were determined by GC/MS (Table 2). The GC/MS data showed some new substrates for the mycobacterial DNA glycosylases and confirmed our observations with the *in vitro* biochemical and kinetics assays. Not only 8-oxoG but also FapyG and FapyA are good substrates for MtuFpg1; however, MtuFpg1 does not recognize 8-oxoA as well as EcoFpg. 8-OxoA, however, is a good substrate for MtuNth. The activity of MtuNei1 on MeHyd, 5-OHU or FapyG is comparable to that of EcoNei. While EcoNei prefers FapyA and 5-OHC, MtuNei1 is not efficient at excising these two lesions. 5-OHC, 5-OHU and FapyA turned out to be good substrates for both MtuNth and EcoNth. The low efficiency of MtuNei1 and MtuNth in repairing Tg suggested by GC/MS analysis is probably due to the heterogeneity of Tg produced by γ-irradiation (Teebor *et al.*, 1987).

MtuNth strongly prefers the (5S) isomers of Tg, whereas MtuNei1 recognizes the (5R) isomers significantly better.

Nth proteins are oxidized pyrimidine-specific DNA glycosylases. Similar to EcoNth, the good substrates for MtuNth were DHU, urea, AP sites, 5-OHU and 5-OHC

(Fig. 3A). DHT was a poor substrate for both MtuNth and EcoNth. Surprisingly, unlike EcoNth, MtuNth did not excise Tg with high efficiency (Fig. 3A). Considering that in aqueous solution Tg is a mixture of two *cis-trans* pairs of isomers ((5*R*,6*S*;5*R*,6*R*) *cis-trans* pair and (5*S*,6*R*;5*S*,6*S*) *cis-trans* pair) and the Tg presented in our 35-mer oligodeoxynucleotide consists predominantly of the (5*R*) *cis-trans* pair, this result may be due to the stereospecificity of MtuNth toward the (5*S*) *cis-trans pair*. Moreover, the single turnover kinetics with Tg suggested that both MtuNth and EcoNth were stereospecific (Table 1). The smaller k_{obs} numbers are associated with the predominant isomer, which suggests that both Nth enzymes prefer the (5*S*) isomers. Both Nei proteins were also stereospecific. In contrast to the Nth proteins, the large k_{obs} values obtained with MtuNei1 and EcoNei correspond to the predominant (5*R*) isomers. This is consistent with what we observed in Fig. 3 as well as with previous observations of EcoNei and EcoNth (Katafuchi *et al.*, 2004; Miller *et al.*, 2004). We then further tested the stereospecificities of the four enzymes with the 30-mer oligodeoxynucleotide containing either the (5*R*) or the (5*S*) *cis-trans* pairs. As shown in Fig. 6, MtuNth strongly preferred the (5*S*) isomers and did not efficiently cleave the (5*R*) isomers, whereas MtuNei1 excised the (5*R*) isomers much better than the (5*S*) isomers. The enzymatic activities of MtuNth and MtuNei1 in excising the two *cis-trans* pairs of Tg are perfectly complementary, and even though MtuNth and MtuNei1 exhibit the stereospecific tendencies of their *E. coli* homologs (Katafuchi *et al.*, 2004; Miller *et al.*, 2004), they are much more stereospecific (Fig. 6).

In order to further explore which amino acids of MtuNth are responsible for its stringent stereospecificity, Thr47 and Ser187 in the potential lesion recognition pocket of MtuNth were mutated to Val and His respectively. However, none of the mutant enzymes showed enhanced activity on (5*R*)-Tg (data not shown).

MtuNei2 can decrease the spontaneous mutation frequency of E. coli mutants.

The *M. tuberculosis Rv3297* gene encodes a 28.5 kDa endonuclease VIII (MtuNei2). We successfully overexpressed and purified MtuNei2 using the bicistronic vector. Unfortunately, the active portion of each MtuNei2 preparation was low, 0.5% - 1% based on our Schiff base assays (data not shown). The cell extract assays with the cell lysates expressing MtuNei2 did not give us promising results either (data not shown). We then decided to determine if MtuNei2 was able to complement the spontaneous mutation frequencies of *E. coli* mutants. Fig. 7 and supplementary Table S2 show the spontaneous mutation frequencies to Rifampin resistance (Rif^R) of wild type *E. coli*, *E. coli fpg mutY nei* triple mutants, *E. coli nei nth* double mutants, as well as these same mutants expressing either EcoNei, MtuNei1 or MtuNei2. IPTG, 0.1mM, was used to maintain a low-level of heterologous expression. The activity of MtuNei1 and EcoNei was confirmed using cell extracts with double-stranded oligodeoxynucleotides containing Tg (data not shown). Consistent with what we have previously observed (Blaisdell *et al.*, 1999), the *fpg mutY nei* triple mutants exhibit an extremely high spontaneous mutation frequency, about 500-fold higher than the wild type (Fig. 7A). This mutation frequency can be significantly decreased by expressing

either EcoNei, MtuNei1 or MtuNei2 ($p < 0.05$) in the mutant cells. Since the triple mutant accumulates G to T transversions due to its failure to repair guanine damage, which is the biological function of Fpg and MutY (Blaisdell *et al.*, 1999), our data suggest that both MtuNei1 and MtuNei2 can prevent G to T transversions as does EcoNei probably by removing oxidized guanine products, such as Sp and urea. Furthermore, *nei nth* double mutants also exhibit an elevated spontaneous mutation frequency although not as high as the *fpg mutY nei* triple mutants (Fig. 7, supplementary Table S2 and (Blaisdell *et al.*, 1999; Jiang *et al.*, 1997a)). The spontaneous mutations observed in the double mutant are solely C to T transitions due to the inability to repair oxidized cytosines (Blaisdell *et al.*, 1999), which can be efficiently prevented by EcoNei, MtuNei1 and MtuNei2 (Fig. 7B). Because Nei proteins play a major role in repairing pyrimidine damage and a backup role in repairing guanine damage, their effects on decreasing mutation frequencies are greater with the *nei nth* double mutants (Fig. 7B) than the *fpg mutY nei* triple mutants (Fig. 7A). Taken together, our data suggest that MtuNei2 can function as an active DNA glycosylase removing oxidized G and C products *in vivo* therefore preventing mutagenesis.

Discussion

In contrast to EcoFpg, MtuFpg1does not remove 8-oxoG when paired with A.

The most notable difference between EcoFpg and MtuFpg1 is their opposite base specificities. Both enzymes recognize 8-oxoG paired with C, T and G, however, MtuFpg1 exhibits strong opposite base specificity having almost no activity on 8-oxoG:A (Fig. 1, 2 and (Olsen *et al.*, 2009)), whereas EcoFpg retains its ability to repair 8-oxoG:A albeit with significantly lower efficiency than 8-oxoG:C (Castaing *et al.*, 1993; Krishnamurthy *et al.*, 2007b; Leipold *et al.*, 2000b; Tchou *et al.*, 1994a). The basis for this difference is still unknown, as MtuFpg1 exhibits significant sequence similarity to EcoFpg and contains the conserved Fpg family insertion loop (Met74, Arg109 and Phe111), of which Arg109 has been shown to contribute to the opposite base specificity (Fromme & Verdine, 2002). Interestingly, the replicative DNA polymerase of *M. tuberculosis* tends to insert G instead of A or C opposite 8-oxoG resulting in 8-oxoG:G mispair (Jain *et al.*, 2007). In this case, the activity of MtuFpg1 on 8-oxoG:G would lead to a G to C transition. Although the 8-oxoG:A mispair may occur less frequently in *M. tuberculosis* than in *E. coli,* when A is inserted opposite 8-oxoG, its mutagenic effect could be efficiently prevented by MutY. Recently, mycobacterial MutY has been shown to recognize A, G and T but not C opposite 8-oxoG, unlike EcoMutY which only sees 8-oxoG:A (Kurthkoti *et al.*, 2009). Thus MtuMutY can remove G from the 8-oxoG mispair formed during *M. tuberculosis*

replication past 8-oxoG. Taken together, the mycobacterial proteins have adapted to the high G-C content of their genome.

Pseudogenization may facilitate the pathogenesis of M. tuberculosis.

Although Sidorenko *et al.* have shown that MtuFpg2 slightly increased (about 1.5 fold) the mutation frequency of *E. coli* CC104 *fpg mutY* strain (Sidorenko *et al.*, 2008), our biochemical data suggest that the gene encoding MtuFpg2 is most likely a pseudogene. During prolonged intracellular survival, mycobacteria tend to eliminate their important pathogenic genes, thereby reducing virulence and facilitating endosymbiosis (Gericke, 2006). Some other nonpathogenic mycobacteria, such as *M. smegmatis* and *Mycobacterium avium,* harbor a full-length Fpg2 homolog. We thus propose that the redundant Fpg enzymes may have at some point contributed to the virulence of the pathogen in the nonhost environment. DNA glycosylase assays on these full-length Fpg2 homologs should elucidate the original biochemical activity of MtuFpg2. The most parsimonious assumption for the activity of Fpg2 is that it is similar to those of other Fpg proteins. Nevertheless, the presence of redundant Fpg proteins raises the possibility that Fpg2 has assumed some new function or mode of operation in at least a few species (Dou *et al.*, 2003; Dou *et al.*, 2008; Pumo *et al.*, 2009). An extreme case of reductive evolution is *M. leprae* (Cole *et al.*, 2001). Surprisingly, although more than half of its genes exist as pseudogenes, *M. leprae* maintains the full-length orthologs of MtuFpg1 and MtuNth, which again indicates the important physiological role of these DNA glycosylases.

Recognizing single-stranded DNA is a property of the Fpg/Nei family members.

EcoNei, MtuNei1 and NEIL1 are all capable of removing oxidized pyrimidines not only from double-stranded oligodeoxynucleotides but also from single-stranded oligodeoxynucleotides (Fig. 3 and 4 and (Dou *et al.*, 2003; Dou *et al.*, 2008; Hailer *et al.*, 2005a; Zhang *et al.*, 2005)). The two prokaryotic Nei proteins studied are also functional on bubble structures and 3' forked structure, as are their eukaryotic homologs (Fig. 5 and (Dou *et al.*, 2003; Dou *et al.*, 2008)). In addition, both MtuFpg1 and EcoFpg (Fig. 4A and (Castaing *et al.*, 1992; Neto *et al.*, 1992)) exhibit lyase activity on single-stranded substrates and EcoFpg recognizes a single-stranded substrate containing 8-oxoG (Fig. 4A and (Ishchenko *et al.*, 1999a; Ishchenko *et al.*, 1999b)). Taken together with the recent observations on Nei orthologs from Mimivirus (Bandaru *et al.*, 2007) as well as Fpg homologs derived from *Candida albicans* and *Arabidopsis thaliana* (Kathe *et al.*, 2009), the data suggest that recognizing lesions in single-stranded DNA as well as other DNA structures containing single-stranded regions is not exclusively a function of the human/mammalian NEILs (Dou *et al.*, 2003; Dou *et al.*, 2008; Hailer *et al.*, 2005a; Takao *et al.*, 2002; Zhang *et al.*, 2005), but rather a characteristic of the Fpg/Nei family members in general. Since transient single-stranded DNA is formed during DNA replication and transcription, the replication-associated DNA repair functions proposed for mammalian cells (Dou *et al.*, 2003) may also occur in prokaryotes, lower eukaryotes and plants.

MtuNei1 and MtuNth exhibit stringent stereospecificities and are catalytically complementary in repairing Tg.

We observed that MtuNth strongly prefers the (5*S*) *cis-trans* pair of Tg, while MtuNei1 removes the (5*R*) *cis-trans* pair much more efficiently (Fig. 6). The two enzymes are catalytically complementary in repairing Tg and more stereoselective compared to their *E. coli* homologs. Our observation on the stereospecificity of EcoNth and EcoNei is consistent with what has been reported (Katafuchi *et al.*, 2004; Miller *et al.*, 2004). Interestingly, mammalian NTH proteins exhibit opposite stereospecificities compared to their prokaryotic homologs and prefer the (5*R*) *cis-trans* pair of Tg (Katafuchi *et al.*, 2004; McTigue *et al.*, 2004; Miller *et al.*, 2004; Ocampo-Hafalla *et al.*, 2006). Moreover, although human NEIL1 prefers the (5*R*) isomers (Katafuchi *et al.*, 2004; Ocampo-Hafalla *et al.*, 2006) as EcoNei does, mouse NEIL1 prefers the (5*S*) isomers (Miller *et al.*, 2004).

MtuNei2 is under the control of a RecA-independent promoter.

Nei proteins are missing from the majority of eubacterial genomes sequenced thus far except for actinobacteria. The common ancestor of all the actinomycetes had two Fpg and two Nei proteins (Pumo *et al.*, 2009). The gene encoding MtuNei2 (*Rv3297*) is predicted to be co-transcribed with a potential ATP-dependent helicase gene *lhr* (long helicase-related), which is immediately upstream to it (http://www.microbesonline.org/operons/gnc83332.html). The putative operon is under the control of a recently identified RecA-independent promoter (RecA-NDp) and could

be induced through a novel RecA-independent mechanism (Gamulin *et al.*, 2004; Rand *et al.*, 2003). Even though disrupting *lhr* in *E. coli* did not show any effect on cell growth or enhanced sensitivities toward UV irradiation and H_2O_2 treatment (Reuven *et al.*, 1995), the potential role of *lhr* during infection and in the DNA damage response has been underscored by the fact that the expression of *lhr* in *M. tuberculosis* could be upregulated by UV irradiation and mitomycin C treatment and induced in the macrophage environment (Boshoff *et al.*, 2003; Schnappinger *et al.*, 2003). Therefore, it is possible that the interaction between Lhr and MtuNei2 may be crucial for recruiting MtuNei2 onto damaged DNA and stimulating the activity of MtuNei2 during DNA repair. Furthermore, the *lhr*/*nei2* operon as well as an upstream promoter similar to *M. tuberculosis* RecA-NDp is present in most if not all the actinobacteria, which strongly suggests that the operon is under selection. Although the biological function of MtuNei2 appears to be redundant to that of MtuNei1 (Fig. 7 and Table S2), its activity is regulated differently and may be inducible upon infection or after DNA damage. Thus rather than possessing a complementary activity, MtuNei2 may act as a backup for MtuNei1 under stressful conditions or function during a different life stage of *M. tuberculosis*.

Mycobacterial enzymes are catalytically slower than their E. coli homologs.

Our kinetic data indicate that the overall rate constants for the *M. tuberculosis* enzymes, particularly MtuFpg1 and MtuNei1 are lower than those of the *E. coli* enzymes. This is possibly due to the slower growth rate of the pathogen, which might

obviate the need for catalytically effective glycosylases. Moreover, even though these mycobacterial enzymes exhibit significant sequence similarity to their *E. coli* homologs, structural differences, as yet undefined, may account not only for lower catalytic rates but also for the altered substrate preferences of these enzymes.

Conclusions

Mycobacterial DNA repair mechanisms may contribute to its survival and persistent infection. The genome of *M. tuberculosis* is highly G/C-rich, which renders the pathogen more susceptible to oxidative DNA damage. It turns out that *M. tuberculosis* is well equipped to deal with this potential problem harboring at least three Fpg/Nei members and one Nth homolog as well as other BER enzymes. Here we overexpressed, purified and characterized MtuFpg1, MtuFpg2, MtuNei1 and MtuNth *in vitro* and evaluated the biological role of MtuNei2 *in vivo*. We have shown that the substrate specificities of these DNA glycosylases are overlapping but less redundant compared to their *E. coli* counterparts. Our data provide direct biochemical evidence for the importance of these mycobacterial DNA glycosylases in DNA repair and suggest that BER is an essential process to maintain the genomic stability of *M. tuberculosis*.

Acknowledgements

We thank Dr. Nadine Honoré and Dr. Karin Eiglmeier of Institute Pasteur, Paris, France, for generously providing the genomic DNA of *M. tuberculosis* H37Rv. Moreover, we would like to thank Wendy Cooper and Alicia Holmes for their help in protein purification. We also thank Dr. Sylvie Doublié and April Averill for helpful suggestions on protein expression and purification. We are grateful to Jeffrey Blaisdell and Dr. Scott Kathe for determining the active fractions of all the control enzymes as well as helpful discussion in terms of enzyme kinetics. We also appreciate the generous help of Drs. Stephanie Duclos, Pierre Aller, Matthew Hogg, Dorothy Pumo, as well as that of Ramiro Barrantes and Minmin Liu. The study was supported by National Institutes of Health grant PO1 CA098993 awarded by the National Cancer Institute. Yin Guo was supported by a research fellowship from DOE EPSCoR DE-F602-00ER45828.

Certain commercial equipment or materials are identified in this paper in order to specify adequately the experimental procedure. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

References

Aravind, L., Walker, D. R. & Koonin, E. V. (1999). Conserved domains in DNA repair proteins and evolution of repair systems. *Nucleic Acids Res* 27, 1223-1242.

Bandaru, V., Sunkara, S., Wallace, S. S. & Bond, J. P. (2002). A novel human DNA glycosylase that removes oxidative DNA damage and is homologous to *Escherichia coli* endonuclease VIII. *DNA Repair (Amst)* 1, 517-529.

Bandaru, V., Blaisdell, J. O. & Wallace, S. S. (2006). Oxidative DNA glycosylases: recipes from cloning to characterization. *Methods Enzymol* 408, 15-33.

Bandaru, V., Zhao, X., Newton, M. R., Burrows, C. J. & Wallace, S. S. (2007). Human endonuclease VIII-like (NEIL) proteins in the giant DNA Mimivirus. *DNA Repair (Amst)* 6, 1629-1641.

Blaisdell, J. O., Hatahet, Z. & Wallace, S. S. (1999). A novel role for Escherichia coli endonuclease VIII in prevention of spontaneous G-->T transversions. *J Bacteriol* 181, 6396-6402.

Blaisdell, J. O. & Wallace, S. S. (2007). Rapid determination of the active fraction of DNA repair glycosylases: a novel fluorescence assay for trapped intermediates. *Nucleic Acids Res* 35, 1601-1611.

Boshoff, H. I., Reed, M. B., Barry, C. E., 3rd & Mizrahi, V. (2003). DnaE2 polymerase contributes to in vivo survival and the emergence of drug resistance in Mycobacterium tuberculosis. *Cell* 113, 183-193.

Burrows, C. J., Muller, J. G., Kornyushyna, O., Luo, W., Duarte, V., Leipold, M. D. & David, S. S. (2002). Structure and potential mutagenicity of new hydantoin products from guanosine and 8-oxo-7,8-dihydroguanine oxidation by transition metals. *Environmental health perspectives* 110 Suppl 5, 713-717.

Castaing, B., Boiteux, S. & Zelwer, C. (1992). DNA containing a chemically reduced apurinic site is a high affinity ligand for the E. coli formamidopyrimidine-DNA glycosylase. *Nucleic Acids Res* 20, 389-394.

Castaing, B., Geiger, A., Seliger, H., Nehls, P., Laval, J., Zelwer, C. & Boiteux, S. (1993). Cleavage and binding of a DNA fragment containing a single 8-oxoguanine by wild type and mutant FPG proteins. *Nucleic acids research* 21, 2899-2905.

Cole, S. T., Brosch, R., Parkhill, J. & other authors (1998). Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature* 393, 537-544.

Cole, S. T., Eiglmeier, K., Parkhill, J. & other authors (2001). Massive gene decay in the leprosy bacillus. *Nature* 409, 1007-1011.

Darwin, K. H. & Nathan, C. F. (2005). Role for nucleotide excision repair in virulence of Mycobacterium tuberculosis. *Infect Immun* 73, 4581-4587.

David, S. S., O'Shea, V. L. & Kundu, S. (2007). Base-excision repair of oxidative DNA damage. *Nature* 447, 941-950.

Davis, E. O., Springer, B., Gopaul, K. K., Papavinasasundaram, K. G., Sander, P. & Bottger, E. C. (2002). DNA damage induction of recA in Mycobacterium tuberculosis independently of RecA and LexA. *Molecular microbiology* 46, 791-800.

Dizdaroglu, M., Bauche, C., Rodriguez, H. & Laval, J. (2000). Novel substrates of Escherichia coli nth protein and its kinetics for excision of modified bases from DNA damaged by free radicals. *Biochemistry* 39, 5586-5592.

Dizdaroglu, M. (2005). Base-excision repair of oxidative DNA damage by DNA glycosylases. *Mutat Res* 591, 45-59.

Dos Vultos, T., Mestre, O., Tonjum, T. & Gicquel, B. (2009). DNA repair in Mycobacterium tuberculosis revisited. *FEMS Microbiol Rev* 33, 471-487.

Dou, H., Mitra, S. & Hazra, T. K. (2003). Repair of oxidized bases in DNA bubble structures by human DNA glycosylases NEIL1 and NEIL2. *J Biol Chem* 278, 49679-49684.

Dou, H., Theriot, C. A., Das, A., Hegde, M. L., Matsumoto, Y., Boldogh, I., Hazra, T. K., Bhakat, K. K. & Mitra, S. (2008). Interaction of the human DNA glycosylase NEIL1 with proliferating cell nuclear antigen. The potential for replication-associated repair of oxidized bases in mammalian genomes. *J Biol Chem* 283, 3130-3140.

Eisen, J. A. & Hanawalt, P. C. (1999). A phylogenomic study of DNA repair genes, proteins, and processes. *Mutat Res* 435, 171-213.

Fromme, J. C. & Verdine, G. L. (2002). Structural insights into lesion recognition and repair by the bacterial 8-oxoguanine DNA glycosylase MutM. *Nature structural biology* 9, 544-552.

Gamulin, V., Cetkovic, H. & Ahel, I. (2004). Identification of a promoter motif regulating the major DNA damage response mechanism of Mycobacterium tuberculosis. *FEMS Microbiol Lett* 238, 57-63.

Gericke, G. S. (2006). Is there an emerging endosymbiotic relationship between mycobacteria and the human host based on horizontal transfer of genetic sequences? *Med Hypotheses* 67, 1419-1428.

Gilboa, R., Zharkov, D. O., Golan, G., Fernandes, A. S., Gerchman, S. E., Matz, E., Kycia, J. H., Grollman, A. P. & Shoham, G. (2002). Structure of formamidopyrimidine-DNA glycosylase covalently complexed to DNA. *The Journal of biological chemistry* 277, 19811-19816.

Guo, Y., Wallace, S. S. & Bandaru, V. (2009). A novel bicistronic vector for overexpressing Mycobacterium tuberculosis proteins in Escherichia coli. *Protein Expr Purif* 65: 230-237.

Hailer, M. K., Slade, P. G., Martin, B. D., Rosenquist, T. A. & Sugden, K. D. (2005a). Recognition of the oxidized lesions spiroiminodihydantoin and guanidinohydantoin in DNA by the mammalian base excision repair glycosylases NEIL1 and NEIL2. *DNA repair* 4, 41-50.

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

Hazra, T. K., Muller, J. G., Manuel, R. C., Burrows, C. J., Lloyd, R. S. & Mitra, S. (2001). Repair of hydantoins, one electron oxidation product of 8-oxoguanine, by DNA glycosylases of Escherichia coli. *Nucleic Acids Res* 29, 1967-1974.

Hazra, T. K., Izumi, T., Boldogh, I., Imhoff, B., Kow, Y. W., Jaruga, P., Dizdaroglu, M. & Mitra, S. (2002a). Identification and characterization of a human DNA glycosylase for repair of modified bases in oxidatively damaged DNA. *Proc Natl Acad Sci U S A* 99, 3523-3528.

Hazra, T. K., Kow, Y. W., Hatahet, Z., Imhoff, B., Boldogh, I., Mokkapati, S. K., Mitra, S. & Izumi, T. (2002b). Identification and characterization of a novel human DNA glycosylase for repair of cytosine-derived lesions. *J Biol Chem* 277, 30417-30420.

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

Ide, H., Kow, Y. W. & Wallace, S. S. (1985). Thymine glycols and urea residues in M13 DNA constitute replicative blocks in vitro. *Nucleic Acids Res* 13, 8035-8052.

Ishchenko, A. A., Bulychev, N. V., Maksakova, G. A., Johnson, F. & Nevinsky, G. A. (1999a). Single-stranded oligodeoxyribonucleotides are substrates of Fpg protein from Escherichia coli. *IUBMB Life* 48, 613-618.

Ishchenko, A. A., Koval, V. V., Fedorova, O. S., Douglas, K. T. & Nevinsky, G. A. (1999b). Structural requirements of double and single stranded DNA substrates and inhibitors, including a photoaffinity label, of Fpg protein from Escherichia coli. *J Biomol Struct Dyn* 17, 301-310.

Iwai, S. (2001). Synthesis and thermodynamic studies of oligonucleotides containing the two isomers of thymine glycol. *Chemistry* 7, 4343-4351.

Jain, R., Kumar, P. & Varshney, U. (2007). A distinct role of formamidopyrimidine DNA glycosylase (MutM) in down-regulation of accumulation of G, C mutations and protection against oxidative stress in mycobacteria. *DNA Repair (Amst)* 6, 1774-1785.

Jaruga, P., Birincioglu, M., Rosenquist, T. A. & Dizdaroglu, M. (2004). Mouse NEIL1 protein is specific for excision of 2,6-diamino-4-hydroxy-5-formamidopyrimidine and 4,6-diamino-5-formamidopyrimidine from oxidatively damaged DNA. *Biochemistry* 43, 15909-15914.

Jiang, D., Hatahet, Z., Blaisdell, J. O., Melamede, R. J. & Wallace, S. S. (1997a). Escherichia coli endonuclease VIII: cloning, sequencing, and overexpression of the nei structural gene and characterization of nei and nei nth mutants. *J Bacteriol* 179, 3773-3782.

Jiang, D., Hatahet, Z., Melamede, R. J., Kow, Y. W. & Wallace, S. S. (1997b). Characterization of Escherichia coli endonuclease VIII. *J Biol Chem* 272, 32230-32239.

Katafuchi, A., Nakano, T., Masaoka, A., Terato, H., Iwai, S., Hanaoka, F. & Ide, H. (2004). Differential specificity of human and Escherichia coli endonuclease III and VIII homologues for oxidative base lesions. *The Journal of biological chemistry* 279, 14464-14471.

Kathe, S. D., Barrantes-Reynolds, R., Jaruga, P., Newton, M. R., Burrows, C. J., Bandaru, V., Dizdaroglu, M., Bond, J. P. & Wallace, S. S. (2009). Plant and fungal Fpg homologs are formamidopyrimidine DNA glycosylases but not 8-oxoguanine DNA glycosylases. *DNA Repair (Amst)* 8, 643-653.

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

Krishnamurthy, N., Muller, J. G., Burrows, C. J. & David, S. S. (2007a). Unusual structural features of hydantoin lesions translate into efficient recognition by Escherichia coli Fpg. *Biochemistry* 46, 9355-9365.

Krishnamurthy, N., Muller, J. G., Burrows, C. J. & David, S. S. (2007b). Unusual structural features of hydantoin lesions translate into efficient recognition by Escherichia coli Fpg. *Biochemistry* 46, 9355-9365.

Krishnamurthy, N., Zhao, X., Burrows, C. J. & David, S. S. (2008). Superior removal of hydantoin lesions relative to other oxidized bases by the human DNA glycosylase hNEIL1. *Biochemistry* 47, 7137-7146.

Krokan, H. E., Nilsen, H., Skorpen, F., Otterlei, M. & Slupphaug, G. (2000). Base excision repair of DNA in mammalian cells. *FEBS Lett* 476, 73-77.

Kurthkoti, K., Srinath, T., Kumar, P., Malshetty, V. S., Sang, P. B., Jain, R., Manjunath, R. & Varshney, U. (2009). A distinct physiological role of MutY in mutation prevention in mycobacteria. *Microbiology* in press.

Lavrukhin, O. V. & Lloyd, R. S. (2000). Involvement of phylogenetically conserved acidic amino acid residues in catalysis by an oxidative DNA damage enzyme formamidopyrimidine glycosylase. *Biochemistry* 39, 15266-15271.

Leipold, M. D., Muller, J. G., Burrows, C. J. & David, S. S. (2000a). Removal of hydantoin products of 8-oxoguanine oxidation by the Escherichia coli DNA repair enzyme, FPG. *Biochemistry* 39, 14984-14992.

Leipold, M. D., Muller, J. G., Burrows, C. J. & David, S. S. (2000b). Removal of hydantoin products of 8-oxoguanine oxidation by the Escherichia coli DNA repair enzyme, FPG. *Biochemistry* 39, 14984-14992.

Lindahl, T. & Barnes, D. E. (2000). Repair of endogenous DNA damage. *Cold Spring Harb Symp Quant Biol* 65, 127-133.

McTigue, M. M., Rieger, R. A., Rosenquist, T. A., Iden, C. R. & De Los Santos, C. R. (2004). Stereoselective excision of thymine glycol lesions by mammalian cell extracts. *DNA repair* 3, 313-322.

Melamede, R. J., Hatahet, Z., Kow, Y. W., Ide, H. & Wallace, S. S. (1994). Isolation and characterization of endonuclease VIII from Escherichia coli. *Biochemistry* 33, 1255-1264.

Miller, H., Fernandes, A. S., Zaika, E., McTigue, M. M., Torres, M. C., Wente, M., Iden, C. R. & Grollman, A. P. (2004). Stereoselective excision of thymine glycol from oxidatively damaged DNA. *Nucleic acids research* 32, 338-345.

Mizrahi, V. & Andersen, S. J. (1998). DNA repair in Mycobacterium tuberculosis. What have we learnt from the genome sequence? *Mol Microbiol* 29, 1331-1339.

Morland, I., Rolseth, V., Luna, L., Rognes, T., Bjoras, M. & Seeberg, E. (2002). Human DNA glycosylases of the bacterial Fpg/MutM superfamily: an alternative pathway for the repair of 8-oxoguanine and other oxidation products in DNA. *Nucleic Acids Res* 30, 4926-4936.

Neto, J. B., Gentil, A., Cabral, R. E. & Sarasin, A. (1992). Mutation spectrum of heat-induced abasic sites on a single-stranded shuttle vector replicated in mammalian cells. *J Biol Chem* 267, 19718-19723.

O'Connor, T. R., Graves, R. J., de Murcia, G., Castaing, B. & Laval, J. (1993). Fpg protein of Escherichia coli is a zinc finger protein whose cysteine residues have a structural and/or functional role. *The Journal of biological chemistry* 268, 9063-9070.

O'Rourke, E. J., Chevalier, C., Pinto, A. V., Thiberge, J. M., Ielpi, L., Labigne, A. & Radicella, J. P. (2003). Pathogen DNA as target for host-generated oxidative stress: role for repair of bacterial DNA damage in Helicobacter pylori colonization. *Proc Natl Acad Sci U S A* 100, 2789-2794.

Ocampo-Hafalla, M. T., Altamirano, A., Basu, A. K., Chan, M. K., Ocampo, J. E., Cummings, A., Jr., Boorstein, R. J., Cunningham, R. P. & Teebor, G. W. (2006). Repair of thymine glycol by hNth1 and hNeil1 is modulated by base pairing and cis-trans epimerization. *DNA repair* 5, 444-454.

Olsen, I., Balasingham, S. V., Davidsen, T., Debebe, E., Rodland, E. A., van Soolingen, D., Kremer, K., Alseth, I. & Tonjum, T. (2009). Characterization of the major formamidopyrimidine-DNA glycosylase homolog in Mycobacterium tuberculosis and its linkage to variable tandem repeats. *FEMS Immunol Med Microbiol* 56, 151-161.

Papavinasasundaram, K. G., Anderson, C., Brooks, P. C., Thomas, N. A., Movahedzadeh, F., Jenner, P. J., Colston, M. J. & Davis, E. O. (2001). Slow induction of RecA by DNA damage in Mycobacterium tuberculosis. *Microbiology* 147, 3271-3279.

Pumo, D., Barrantes-Reynolds, R., Kathe, S., Wallace, S. & Bond, J. (2009). Evolution of the Fpg/Nei family of DNA glycosylases. In *Genetic and Epigenetic Regulation of Genome Stability and Transgenerational Response to Stress*, pp. in press. Hauppauge, New York: Nova Science Publishers, Inc.

Rand, L., Hinds, J., Springer, B., Sander, P., Buxton, R. S. & Davis, E. O. (2003). The majority of inducible DNA repair genes in Mycobacterium tuberculosis are induced independently of RecA. *Mol Microbiol* 50, 1031-1042.

Rengarajan, J., Bloom, B. R. & Rubin, E. J. (2005). Genome-wide requirements for Mycobacterium tuberculosis adaptation and survival in macrophages. *Proc Natl Acad Sci U S A* 102, 8327-8332.

Reuven, N. B., Koonin, E. V., Rudd, K. E. & Deutscher, M. P. (1995). The gene for the longest known Escherichia coli protein is a member of helicase superfamily II. *J Bacteriol* 177, 5393-5400.

Sassetti, C. M. & Rubin, E. J. (2003). Genetic requirements for mycobacterial survival during infection. *Proc Natl Acad Sci U S A* 100, 12989-12994.

Schnappinger, D., Ehrt, S., Voskuil, M. I. & other authors (2003). Transcriptional Adaptation of Mycobacterium tuberculosis within Macrophages: Insights into the Phagosomal Environment. *J Exp Med* 198, 693-704.

Shimizu, T., Manabe, K., Yoshikawa, S., Kawasaki, Y. & Iwai, S. (2006). Preferential formation of (5S,6R)-thymine glycol for oligodeoxyribonucleotide synthesis and analysis of drug binding to thymine glycol-containing DNA. *Nucleic Acids Res* 34, 313-321.

Sidorenko, V. S., Rot, M. A., Filipenko, M. L., Nevinsky, G. A. & Zharkov, D. O. (2008). Novel DNA glycosylases from Mycobacterium tuberculosis. *Biochemistry (Mosc)* 73, 442-450.

Sugahara, M., Mikawa, T., Kumasaka, T., Yamamoto, M., Kato, R., Fukuyama, K., Inoue, Y. & Kuramitsu, S. (2000). Crystal structure of a repair enzyme of oxidatively damaged DNA, MutM (Fpg), from an extreme thermophile, Thermus thermophilus HB8. *The EMBO journal* 19, 3857-3869.

Suvarnapunya, A. E., Lagasse, H. A. & Stein, M. A. (2003). The role of DNA base excision repair in the pathogenesis of Salmonella enterica serovar Typhimurium. *Mol Microbiol* 48, 549-559.

Takao, M., Kanno, S., Kobayashi, K., Zhang, Q. M., Yonei, S., van der Horst, G. T. & Yasui, A. (2002). A back-up glycosylase in Nth1 knock-out mice is a functional Nei (endonuclease VIII) homologue. *J Biol Chem* 277, 42205-42213.

Tchou, J., Michaels, M. L., Miller, J. H. & Grollman, A. P. (1993). Function of the zinc finger in Escherichia coli Fpg protein. *The Journal of biological chemistry* 268, 26738-26744.

Tchou, J., Bodepudi, V., Shibutani, S., Antoshechkin, I., Miller, J., Grollman, A. P. & Johnson, F. (1994a). Substrate specificity of Fpg protein. Recognition and cleavage of oxidatively damaged DNA. *The Journal of biological chemistry* 269, 15318-15324.

Tchou, J., Bodepudi, V., Shibutani, S., Antoshechkin, I., Miller, J., Grollman, A. P. & Johnson, F. (1994b). Substrate specificity of Fpg protein. Recognition and cleavage of oxidatively damaged DNA. *J Biol Chem* 269, 15318-15324.

Teebor, G., Cummings, A., Frenkel, K., Shaw, A., Voituriez, L. & Cadet, J. (1987). Quantitative measurement of the diastereoisomers of cis thymidine glycol in gamma-irradiated DNA. *Free Radic Res Commun* 2, 303-309.

Venkatesh, J., Kumar, P., Krishna, P. S., Manjunath, R. & Varshney, U. (2003). Importance of uracil DNA glycosylase in Pseudomonas aeruginosa and Mycobacterium smegmatis, G+C-rich bacteria, in mutation prevention, tolerance to acidified nitrite, and endurance in mouse macrophages. *J Biol Chem* 278, 24350-24358.

Wallace, S. S. (2002). Biological consequences of free radical-damaged DNA bases. *Free Radic Biol Med* 33, 1-14.

Wallace, S. S., Bandaru, V., Kathe, S. D. & Bond, J. P. (2003). The enigma of endonuclease VIII. *DNA Repair (Amst)* 2, 441-453.

Wilson, D. M., 3rd, Sofinowski, T. M. & McNeill, D. R. (2003). Repair mechanisms for oxidative DNA damage. *Front Biosci* 8, d963-981.

Zhang, Q. M., Yonekura, S., Takao, M., Yasui, A., Sugiyama, H. & Yonei, S. (2005). DNA glycosylase activities for thymine residues oxidized in the methyl group are functions of the hNEIL1 and hNTH1 enzymes in human cells. *DNA Repair (Amst)* 4, 71-79.

Zharkov, D. O., Rieger, R. A., Iden, C. R. & Grollman, A. P. (1997). NH2-terminal proline acts as a nucleophile in the glycosylase/AP-lyase reaction catalyzed by Escherichia coli formamidopyrimidine-DNA glycosylase (Fpg) protein. *J Biol Chem* 272, 5335-5341.

Zharkov, D. O., Golan, G., Gilboa, R., Fernandes, A. S., Gerchman, S. E., Kycia, J. H., Rieger, R. A., Grollman, A. P. & Shoham, G. (2002). Structural analysis of an Escherichia coli endonuclease VIII covalent reaction intermediate. *Embo J* 21, 789-800.

Figure 1 DNA glycosylase/lyase activity of MtuFpg1 on double-stranded DNA substrates. Labeled substrate, 15 nM, was incubated with either no enzyme (-), 15 nM EcoFpg or increasing amounts of MtuFpg1 (1.5 nM, 15 nM or 150 nM).

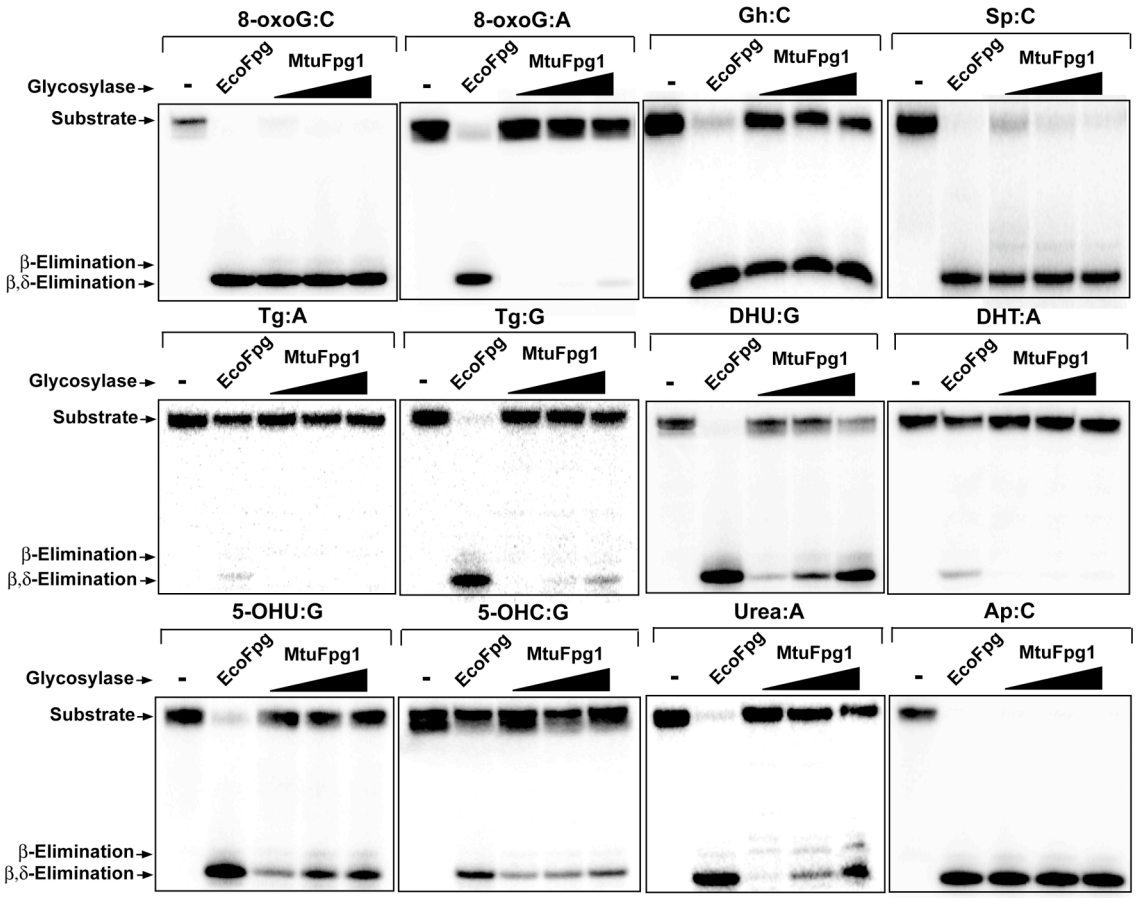


Figure 2 Enzymatic activities of MtuFpg1 (●) and EcoFpg (O) on 8-oxoG:A. Labeled 8-oxoG:A, 15 nM, was incubated with increasing amounts of either MtuFpg1 or EcoFpg at 37 °C for 30 min. The data represent the means of three independent experiments, the uncertainties are standard deviations.

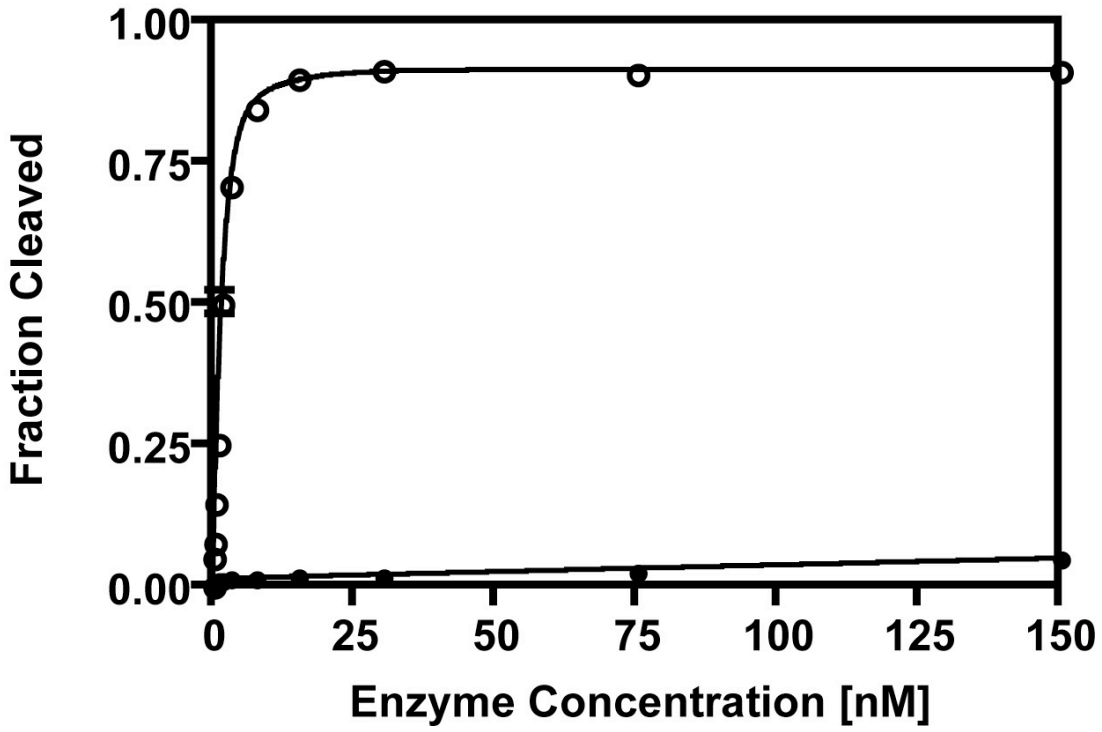
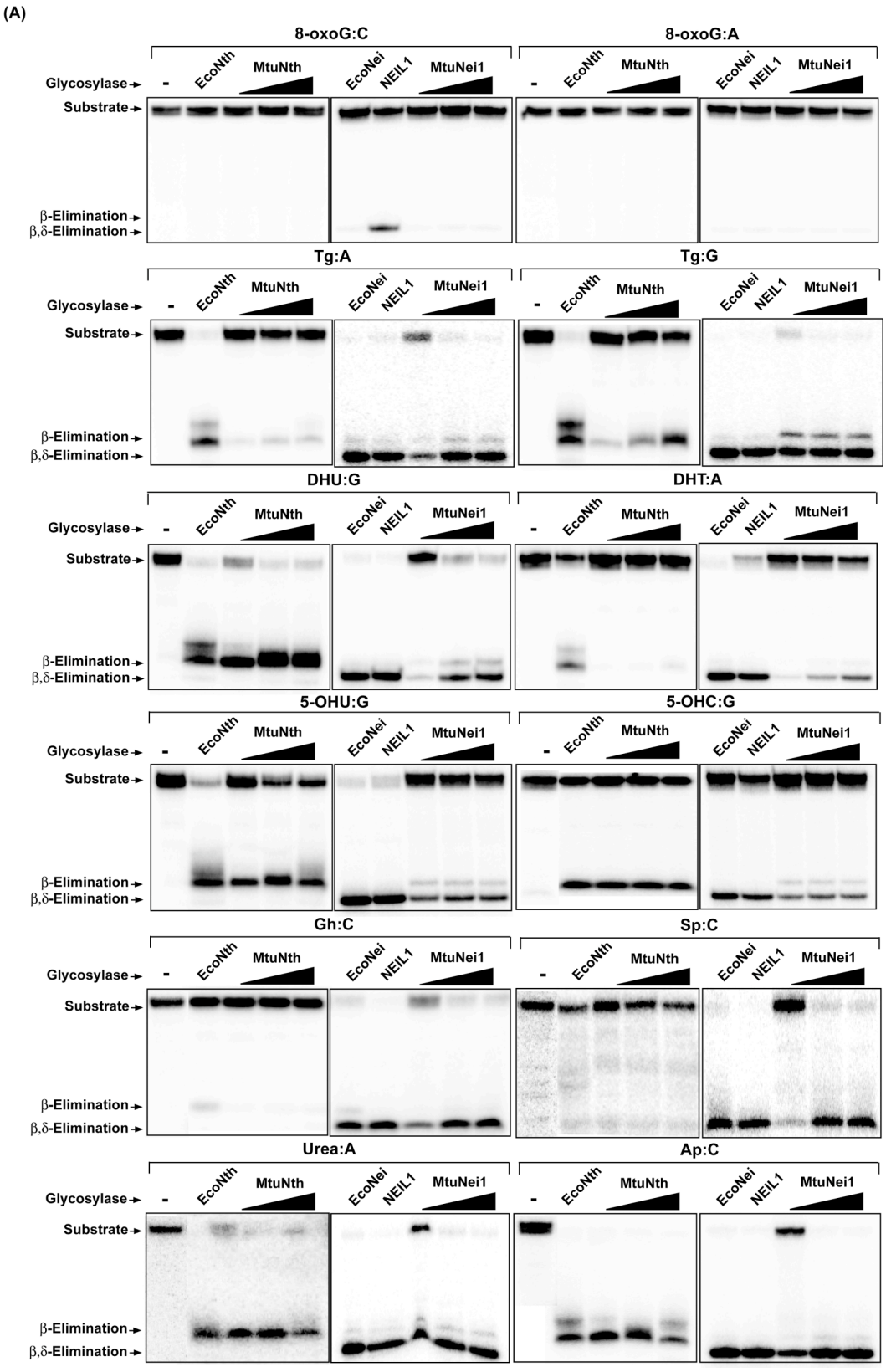


Figure 3 (A) DNA glycosylase/lyase activities of MtuNth and MtuNei1 on double-stranded DNA substrates. Labeled substrate, 15 nM, was incubated with no enzyme (-), 15 nM EcoNth, EcoNei or NEIL1, or increasing amounts of MtuNth or MtuNei1 (1.5 nM, 15 nM or 150 nM). **(B) Uracil DNA glycosylase activity of MtuNei1.** Double-stranded DNA substrate containing uracil, 15 nM, was incubated with no enzyme (-), 15 nM EcoFpg, MtuFpg1, EcoNth, MtuNth, EcoNei or NEIL1, or increasing amounts of MtuNei1 (1.5 nM, 15 nM or 150 nM).


(A)
8-oxoG:C
8-oxoG:A
Tg:A
Tg:G
DHU:G
DHT:A
5-OHU:G
5-OHC:G
Gh:C
Sp:C
Urea:A
Ap:C
Glycosylase
Substrate
β-Elimination
β,δ-Elimination
-
EcoNth
MtuNth
EcoNei
NEIL1
MtuNei1

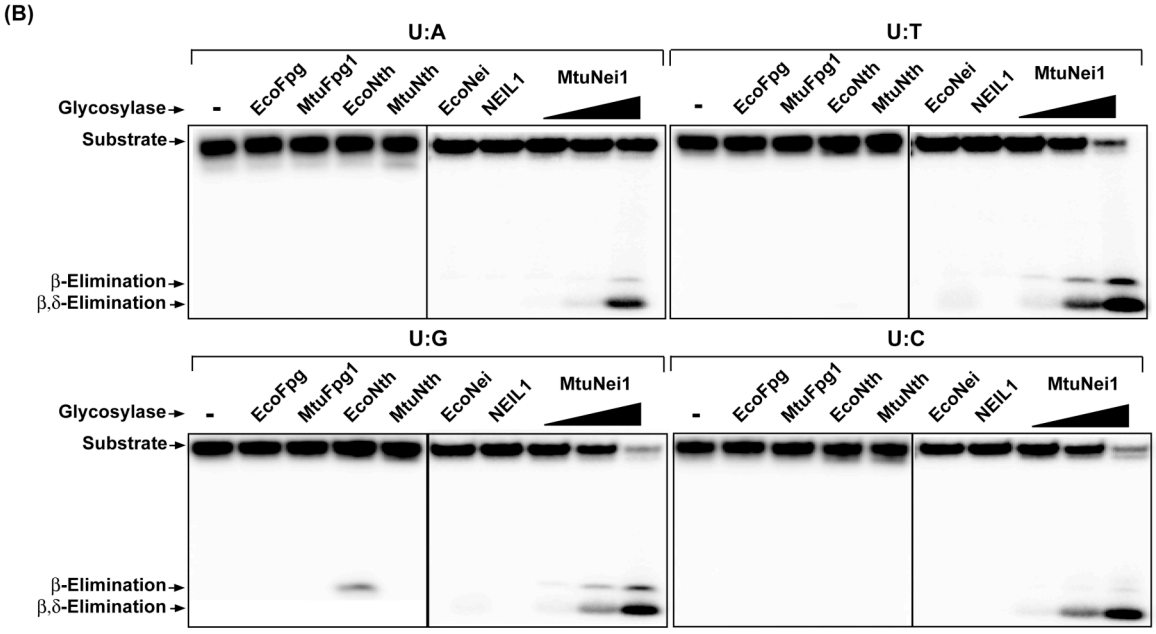
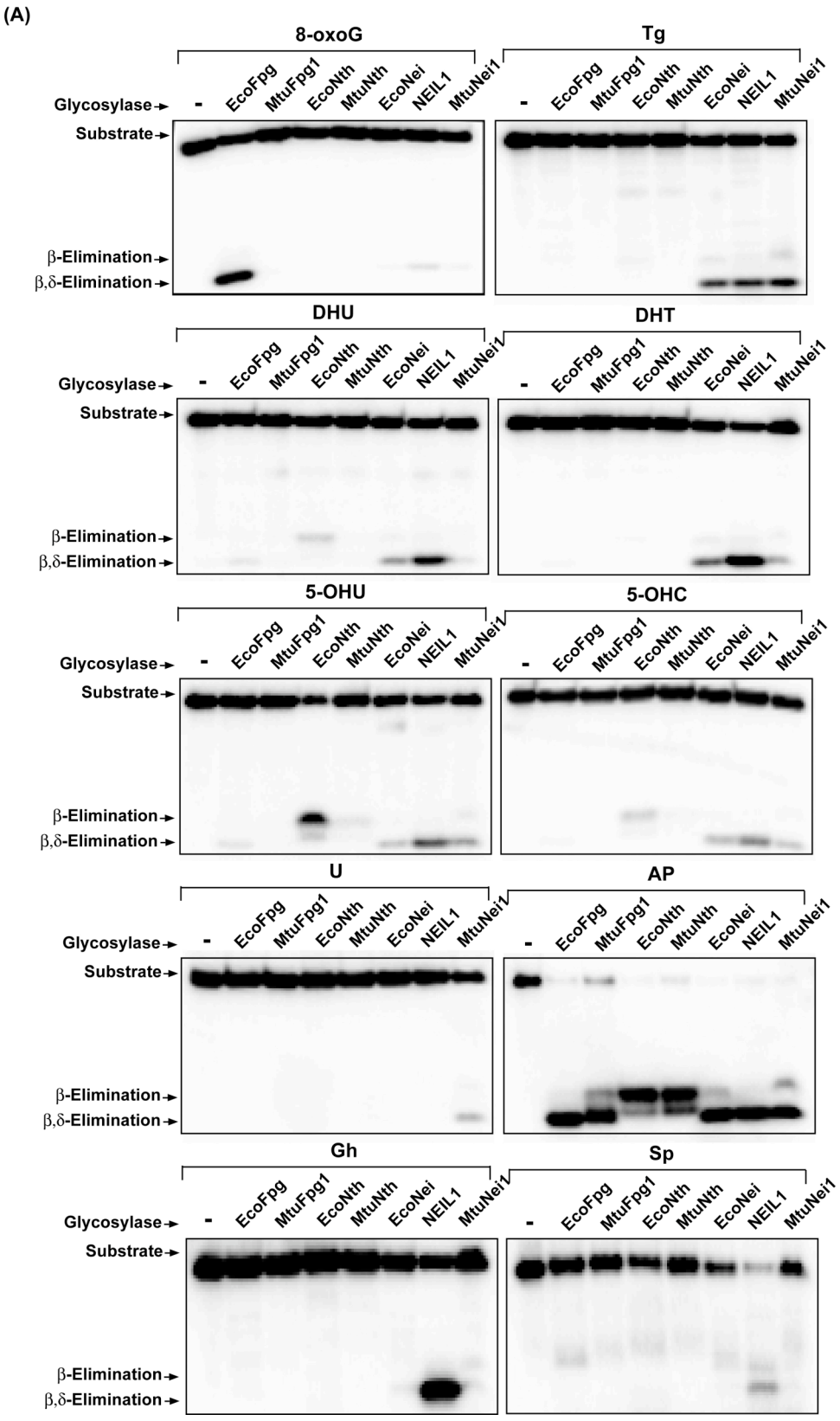

(B)
U:A
U:T
U:G
U:C
Glycosylase
Substrate
β-Elimination
β,δ-Elimination
-
EcoFpg
MtuFpg1
EcoNth
MtuNth
EcoNei
NEIL1
MtuNei1

Figure 4 DNA glycosylase/lyase activity of MtuNei1 on single-stranded DNA substrates. (A) Labeled substrate, 15 nM, was incubated with no enzyme (-), 15 nM EcoFpg, MtuFpg, EcoNth, MtuNth, EcoNei or NEIL1, or MtuNei1. (B) The fraction cleaved was calculated for the reactions with 15 nM enzyme. The data represent the means of three independent experiments and the error bars indicate the standard deviations.


(A)
8-oxoG
Tg
DHU
DHT
5-OHU
5-OHC
U
AP
Gh
Sp
Glycosylase
Substrate
β-Elimination
β,δ-Elimination
-
EcoFpg
MtuFpg1
EcoNth
MtuNth
EcoNei
NEIL1
MtuNei1

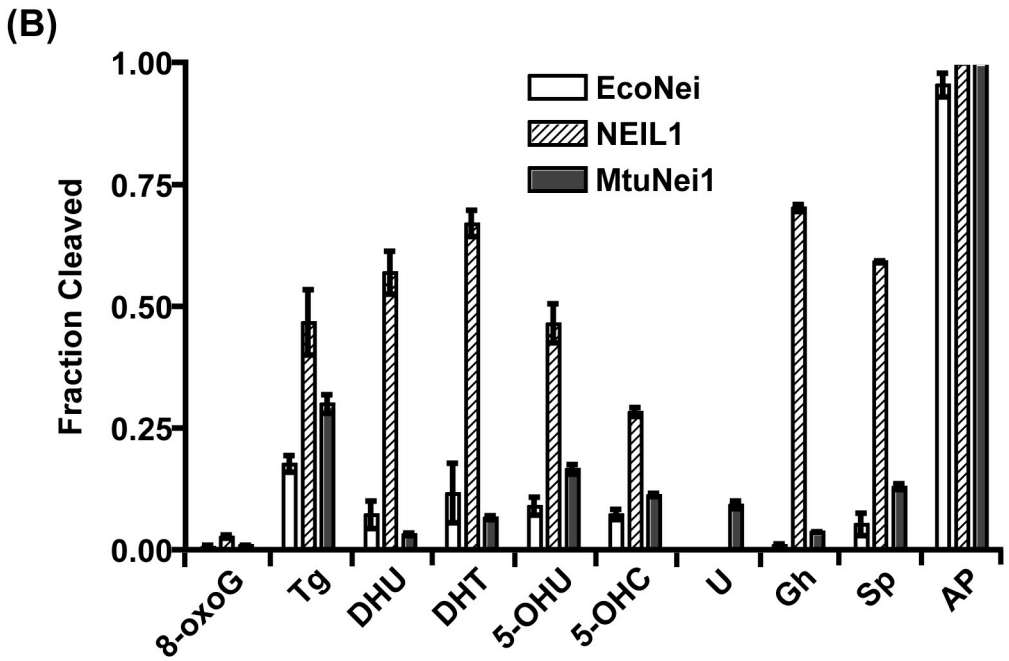

(B)
1.00
0.75
0.50
0.25
0.00
Fraction Cleaved
EcoNei
NEIL1
MtuNei1
8-oxoG
Tg
DHU
DHT
5-OHU
5-OHC
U
Gh
Sp
AP

Figure 5 Activity of MtuNei1 on 5-OHU-containing (A) DNA bubble structures with 5, 11 or 19 unpaired bases and (B) 3' forked structure. Labeled substrate, 15 nM, was incubated with 15 nM EcoNei, NEIL1 or MtuNei1. Cleaved fractions were calculated for each reaction. The double-stranded and single-stranded 51-mer substrates containing 5-OHU were used as controls for the reactions with the bubble structures. The double-stranded and single-stranded 35-mer substrates containing 5-OHU were controls for the reactions with the 3' forked structure. The data represent the means of three independent experiments, the uncertainties are standard deviations.

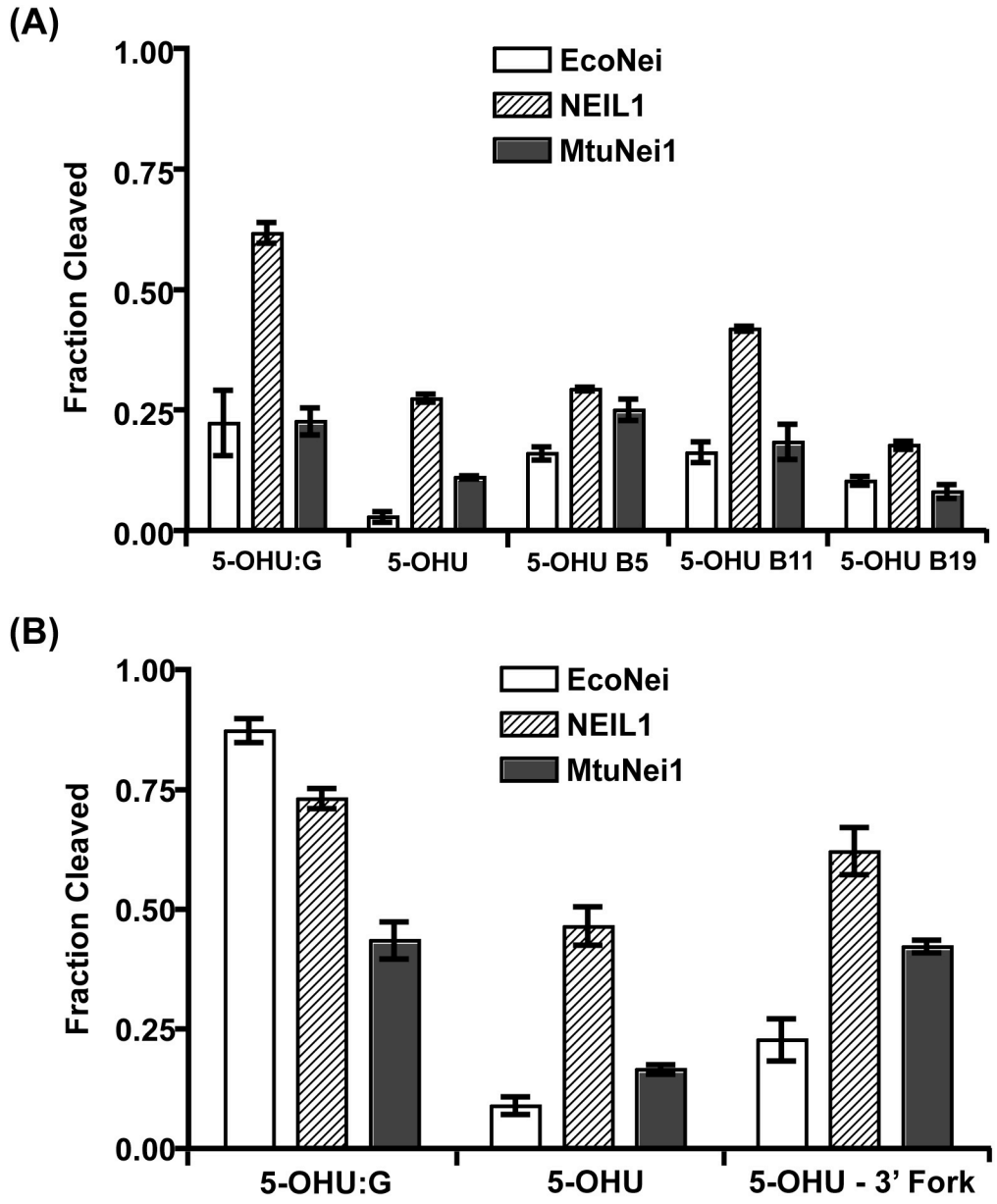

(A)
1.00
0.75
0.50
0.25
0.00
Fraction Cleaved
EcoNei
NEIL1
MtuNei1
5-OHU:G
5-OHU
5-OHU B5
5-OHU B11
5-OHU B19
(B)
1.00
0.75
0.50
0.25
0.00
Fraction Cleaved
EcoNei
NEIL1
MtuNei1
5-OHU:G
5-OHU
5-OHU - 3' Fork

Table 1 - Kinetic parameters: k_{obs} (min^{-1}) and $(k_{cat}/K_m)^{app}$ ($min^{-1}nM^{-1}$) of MtuFpg1, MtuNei1 and MtuNth as well as their *E. coli* homologs on double-stranded oligodeoxynucleotide substrates.

The amount of each enzyme was corrected for the fraction of active enzyme. The reactions were terminated at appropriate time points with formamide stop loading buffer. The k_{obs} and the $(k_{cat}/K_m)^{app}$ values were solved under single-turnover conditions and multiple-turnover conditions respectively. Data were analyzed with Prism 4 (GraphPad Software, Inc. La Jolla, CA). (ND – not detectable)

	MtuFpg1		EcoFpg		MtuNei1		EcoNei		MtuNth		EcoNth	
Substrates	k_{obs}	$(k_{cat}/K_m)^{app}$	k_{obs}	$(k_{cat}/K_m)^{app}$	k_{obs}	$(k_{cat}/K_m)^{app}$	k_{obs}	$(k_{cat}/K_m)^{app}$	k_{obs}	$(k_{cat}/K_m)^{app}$	k_{obs}	$(k_{cat}/K_m)^{app}$
8-oxoG:C	1.50	0.029	15.1	0.012	ND	ND	ND	ND				
Sp:C	4.58	0.30	34.2	5.05	0.46	0.065	13.0	1.39				
Tg:A	ND	ND	0.74	0.012	5.55/0.46	0.010	82.0/4.90	3.30	0.32/ 19.0	0.044	16.0/15.2	1.18
Tg:G	0.06	0.00092	0.69	0.16	8.01/0.59	0.025	260/23.4	8.84	0.027/22.9	0.033	22.6/283.9	0.41
DHU:G	0.14	0.0041	6.58	0.30	1.43	0.028	150	2.11	14.1	0.37	46.4	0.38
5-OHC:G					0.97	0.0090	63.3	0.30	76.7	0.083	25.8	0.18

Table 2 - The activities of mycobacterial DNA glycosylases as well as their *E. coli* homologs on γ-irradiated DNA as measured by GC/MS.

The same amounts of active enzymes were used in each reaction. Numbers represent released damaged bases per 10^6 total DNA bases after 1 hour incubation with γ-irradiated DNA in substrate excess. The data represent the means of three independent experiments. The uncertainties are standard deviations.

	MtuFpg1	EcoFpg	MtuNei1	EcoNei	MtuNth	EcoNth
MeHyd	0.9±0.5	2.4±1.2	18±0.5	23±0.5	20±0.9	27±0.7
5-OHC			4.1±0.3	55±3.2	160±9	246±3
Tg			10±1.4	51±1.2	6.4±0.5	73±1.8
FapyA	84±1.5	202±6	20±0.8	148±0.6	66±2.6	118±1
8-oxoA	2.6±0.6	46±1.8			14±0.5	21±0.7
FapyG	387±4	683±22	16±0.9	15±1.2	18±1.1	12±0.6
8-oxoG	246±6	625±19				
5-OHU			8.9±1.7	11±1.6	108±6.8	199±15

Figure 6 Stereospecificity of MtuNth and MtuNei1 on thymine glycol. Double-stranded oligodeoxynucleotide substrates, 15 nM, containing either 5*R* ((5*R*,6*S*), (5*R*,6*R*)) or 5*S* ((5*S*,6*R*), (5*S*,6*S*)) Tg isomers were incubated with either no enzyme (-) or increasing amounts of MtuNth, EcoNth, MtuNei1 and EcoNei (1.5 nM, 15 nM or 150 nM).

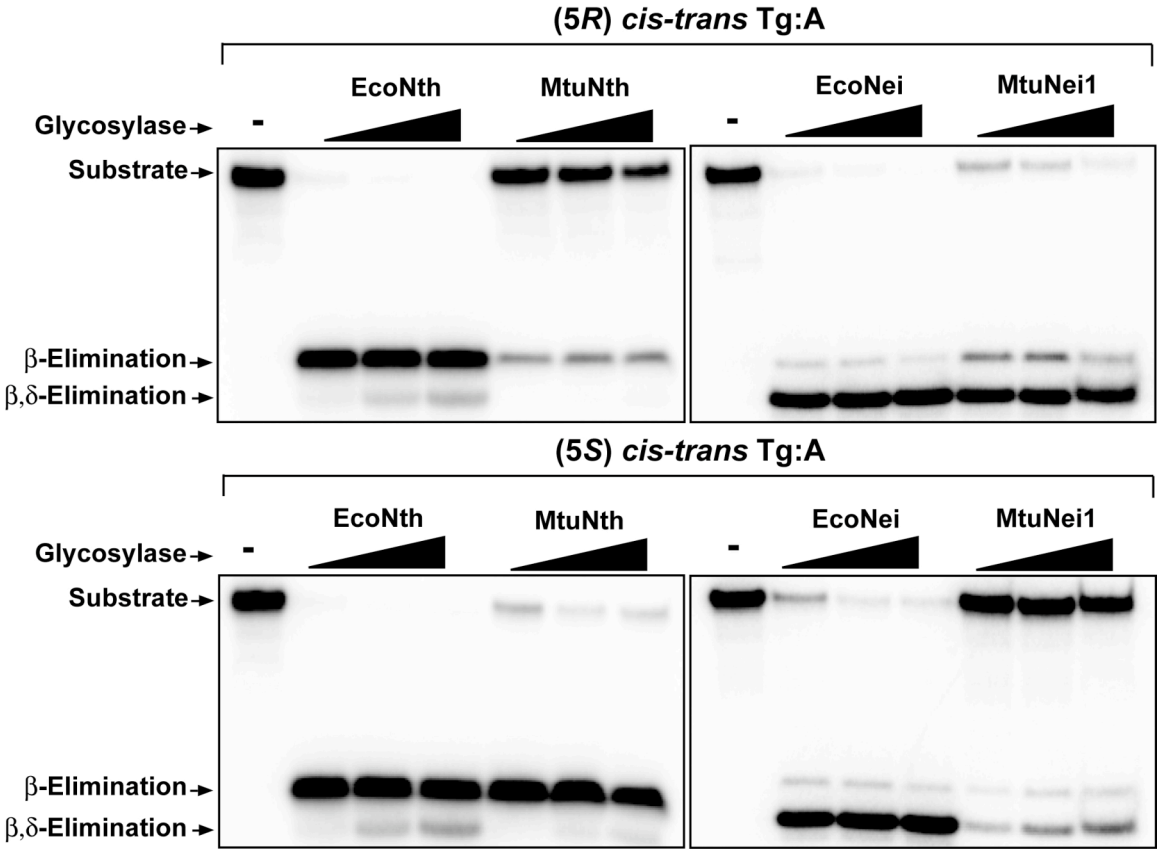
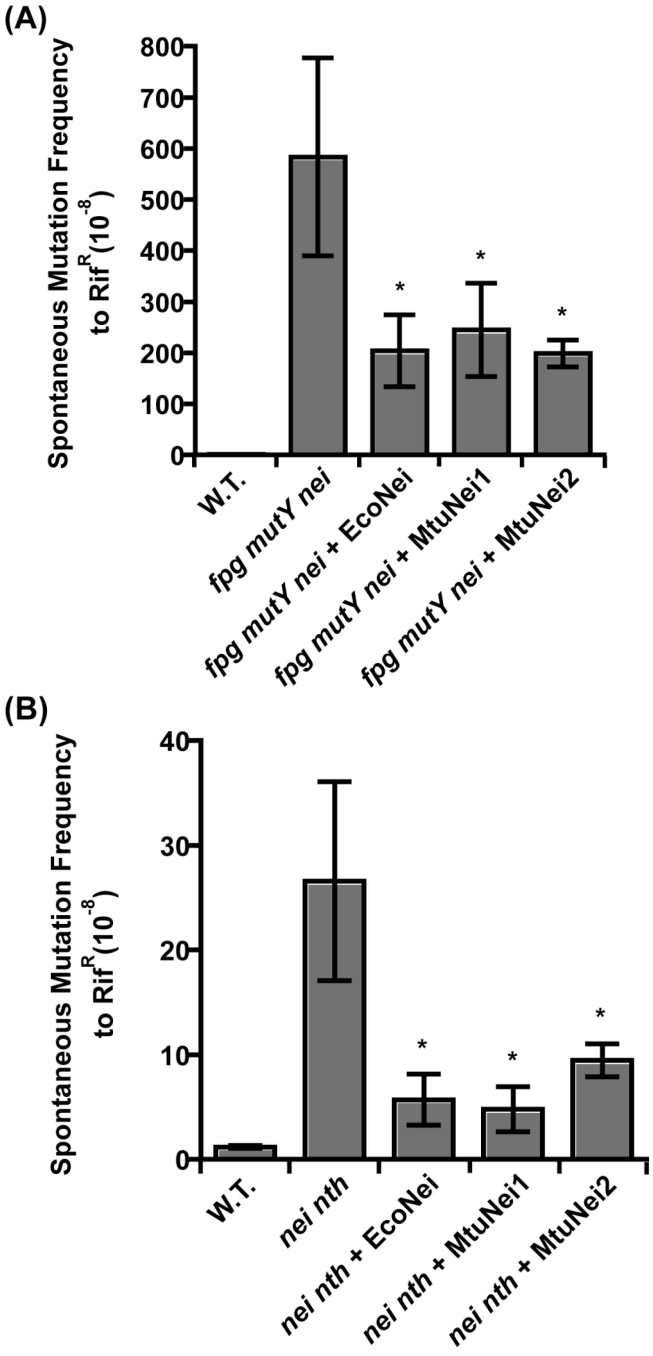


Figure 7 Spontaneous mutation frequencies to rifampin resistance in *E. coli* wild type (W.T.), *fpg mutY nei* triple mutant, *nei nth* double mutant as well as the mutant strains expressing either EcoNei, MtuNei1 or MtuNei2. Mid-log-phase cultures were plated on LB agar with or without 100 μg/ml of rifampin. Mutation frequencies per 10^8 cells were calculated. The data shown are the averages of the three or four independent experiments. The error bars indicate the standard deviations. * - significant difference ($p < 0.05$ by the student's *t* test) compared to the mutation frequency of the *fpg mutY nei* triple mutant or the *nei nth* double mutant.


(A)
Spontaneous Mutation Frequency to Rif^{R}(10^{-8})
800
700
600
500
400
300
200
100
0
*
*
*
W.T.
fpg mutY nei
fpg mutY nei + EcoNei
fpg mutY nei + MtuNei1
fpg mutY nei + MtuNei2
(B)
Spontaneous Mutation Frequency to Rif^{R}(10^{-8})
40
30
20
10
0
*
*
*
W.T.
nei nth
nei nth + EcoNei
nei nth + MtuNei1
nei nth + MtuNei2

Supplementary data

Table S1 - List of enzyme preparations used in this work.

Protein	Purification Date	Total Concentration (μM)	Active Fraction
MtuFpg1	11/22/2007	22.0	18%
MtuFpg1	11/03/2008	19.4	22%
MtuFpg2	07/14/2008	56.8	ND
MtuNei1	09/30/2007	11.4	15%
MtuNei1	10/30/2008	23.8	20%
MtuNth	09/27/2008	137	11%
MtuNth	01/08/2009	89.0	15%
MtuNth	03/21/2009	79.0	10%
EcoFpg	04/13/2006	46.0	14%
EcoFpg	02/29/2008	216	22%
EcoNei	01/17/2008	417	20%
EcoNth	07/08/2009	217	40%
NEIL1	10/01/2008	26.6	31%

The amount of each enzyme used in the experimental protocols was corrected for the percent active fraction determined for each preparation, as shown above. (ND-not detectable)

Table S2 Spontaneous mutation frequencies in wild type *E. coli*, the *E. coli* mutants lacking oxidative base excision repair enzymes, and the mutants expressing EcoNei, MtuNei1 or MtuNei2.

Genotype	Spontaneous mutation frequency to Rif^R (10^{-8})					Mutatioin frequency relative to:
	Exp1	Exp2	Exp3	Exp4	Avg±SD	wild type
wild type	1.2	1.3	1.0	1.3	1.2±0.1	1
fpg mutY nei	458	872	508	496	584±194	486
fpg mutY nei (EcoNei)	133	155	273	255	204±70	170
fpg mutY nei (MtuNei1)	164	268	364	183	245±91	204
fpg mutY nei (MtuNei2)	170	228	212	186	199±26	166
nei nth	26	36	17		26±9.5	17.1
nei nth (EcoNei)	2.8	6.1	8.7	5.2	5.7±2.4	4.8
nei nth (MtuNei1)	2.4	5.2	4.1	7.5	4.8±2.1	4.0
nei nth (MtuNei2)	9.9	10	11	7.2	9.5±1.6	7.9

Figure S1 Opposite base specificities of MtuFpg1 on an AP site. Labeled substrate, 15 nM, was incubated with no enzyme (-), 15 nM EcoFpg or EcoNei or increasing amounts of MtuFpg1 (1.5 nM, 15 nM or 150 nM).

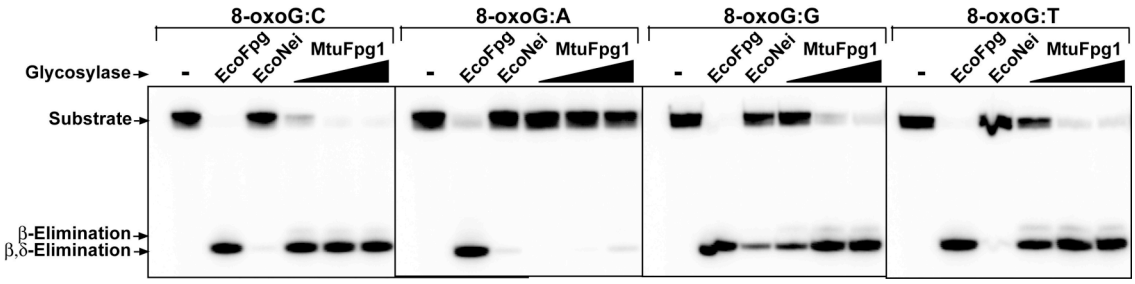


Figure S2 Opposite base specificities of MtuFpg1 on 8-oxoG. Labeled substrate, 15 nM, was incubated with no enzyme (-), 15 nM EcoFpg or EcoNei or increasing amounts of MtuFpg1 (1.5 nM, 15 nM or 150 nM).

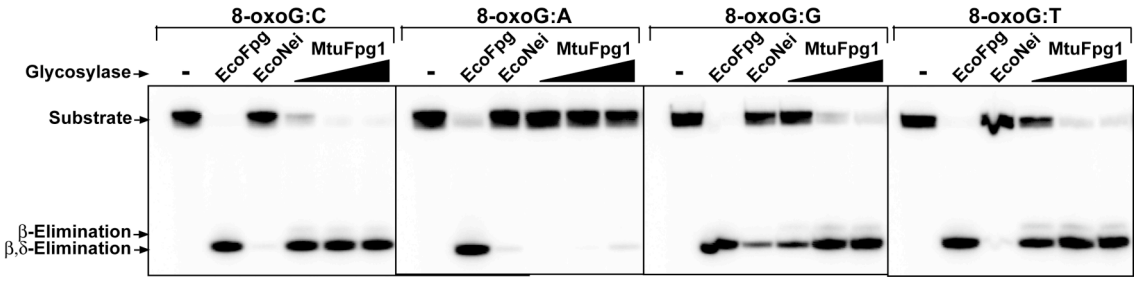
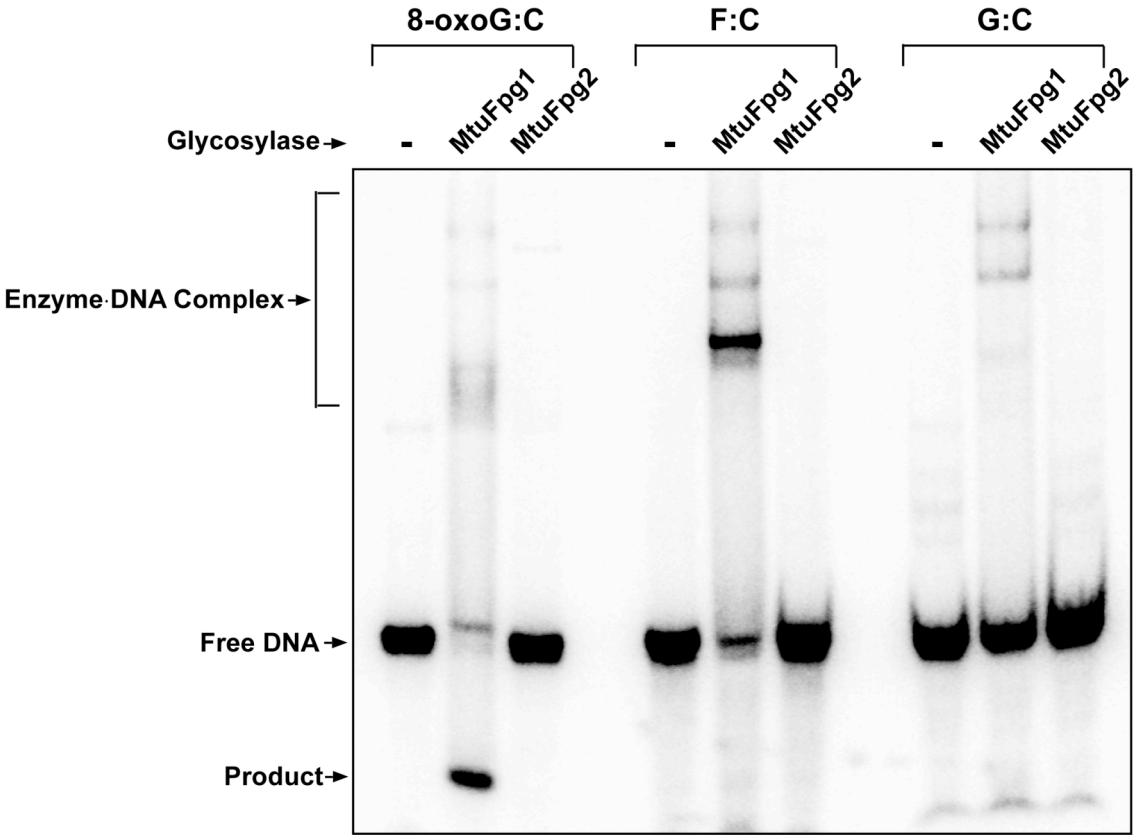


Figure S3 DNA binding by MtuFpg1 and MtuFpg2. Duplex DNA substrates, 25 nM, containing either 8-oxoG, furan (F) or guanine at the corresponding position was incubated with either no enzyme, 25 nM of MtuFpg1 or 5 μM of MtuFpg2.



SUMMARY AND FUTURE DIRECTIONS

Highly redundant Fpg/Nei DNA glycosylases were found in *Actinobacteria,* including *M. tuberculosis*. We managed to overexpress, purify and characterize *in vivo* three of all four Fpg/Nei enzymes (MtuFpg1, MtuFpg2 and MtuNei1) as well as the Nth protein (MtuNth) derived from *M. tuberculosis*. We also evaluated the biological role of the other Nei paralog, MtuNei2 *in vivo*. Our data suggest that mycobacterial DNA glycosylases exhibit more divergent substrate specificities compared to their *E. coli* counterparts. The study provides direct biochemical evidence for the importance of these DNA glycosylases in protecting *M. tuberculosis* from oxidative damage.

Similar to EcoFpg, the essential role of MtuFpg1 is to excise FapyG, 8-oxoG and FapyA but not 8-oxoA. The most notable difference between EcoFpg and MtuFpg1 is their opposite base specificities. MtuFpg1 exhibits significant opposite base specificity having almost no activity on 8-oxoG:A, whereas EcoFpg retains its ability to repair 8-oxoG:A albeit with significantly lower efficiency than 8-oxoG:C. The basis for this difference is still unknown, since MtuFpg exhibits significant sequence similarity to EcoFpg and contains the conserved Fpg family insertion loop (Met74, Arg109 and Phe111) which contributes to the opposite base specificity of EcoFpg (Fromme & Verdine, 2002). In order to determine which amino acid residues are responsible for this difference in opposite base specificity, the structure of MtuFpg1 needs to be solved.

The original biochemical activity of MtuFpg2 remains unknown. Further studies on the full-length Fpg2 homologs derived from other mycobacteria, such as *M. smegmatis* should provide insights. If the biochemical data suggest that Fpg2 has

gained some new functions, further studies on identifying the critical amino acids in charge of the new functions or on identifying the potential protein partners will be interesting and necessary.

Thus far, the substrate specificity of MtuNei2 remains to be identified, although it has been shown to decrease the spontaneous mutation frequencies of *E. coli* mutant strains caused by oxidized guanines and cytosines. The idea that the gene encoding MtuNei2 (*Rv3297*) is in the same operon as *lhr* and under the control of a RecA-NDp promoter can be easily confirmed by using RT-PCR. We have proposed that the interaction between Lhr and MtuNei2 may be critical for recruiting MtuNei2 onto damaged DNA and stimulating the activity of MtuNei2 during DNA repair. *E. coli* Lhr has been successfully overexpressed and purified. Similar methods may be applicable in overexpressing mycobacterial Lhr. The interaction between Lhr and MtuNei2 can be studied using co-immunoprecipitation. Whether or not Lhr can stimulate the activity of MtuNei2 can be further tested with biochemical assays. If Lhr does stimulate the activity of MtuNei2 through protein-protein interaction, site-directed mutagenesis can be used to identify the potential interacting regions. Additionally, obtaining crystal structures of both proteins should provide information on how the interaction occurs and whether or not conformational changes are involved. Above all, these studies will extend our knowledge about the physiological contributions of the two enzymes and more importantly, how these potential protein-protein interactions may facilitate DNA repair in prokaryotes.

One of the most interesting observations made in this work is the stringent stereospecificities of MtuNth and MtuNei1 towards Tg. The two mycobacterial proteins showed much stronger stereospecificities in recognizing and removing the isomers of Tg, compared to their *E. coli* counterparts. Moreover, their stereospecificities are perfectly complementary to each other as MtuNth prefers the (5*S*) *cis-trans* isomers of Tg, whereas MtuNei1 prefers (5*R*) *cis-trans* isomers of Tg. Further attempts at identifying the amino acids responsible for the shift in stereospecificity will be necessary. Although two crystal structures of unliganded Nth homologs have been solved, there is no structure available of Nth with a lesion in the substrate binding pocket. The crystal structure of MtuNth, particularly with DNA containing Tg, would tell us which amino acid(s) play a key role in the stereospecific recognition of Tg as well as provide insights into lesion recognition in general.

Above of all, the work presented here provides a lot of information on biochemical functions of the mycobacterial DNA glycosylases that recognize oxidized base damage. More biochemical studies as well as structural studies will further complement our knowledge on how BER facilitate the pathogenesis of *M.tuberculosis* and how DNA glycosylases maintain the functional redundancy or achieve the functional divergence in different organisms to accommodate different environments.

BIBLIOGRAPHY

Acharya, N., Kumar, P. & Varshney, U. (2003). Complexes of the uracil-DNA glycosylase inhibitor protein, Ugi, with Mycobacterium smegmatis and Mycobacterium tuberculosis uracil-DNA glycosylases. *Microbiology (Reading, England)* 149, 1647-1658.

Akiyama, M., Horiuchi, T. & Sekiguchi, M. (1987). Molecular cloning and nucleotide sequence of the mutT mutator of Escherichia coli that causes A:T to C:G transversion. *Mol Gen Genet* 206, 9-16.

Akiyama, M., Maki, H., Sekiguchi, M. & Horiuchi, T. (1989). A specific role of MutT protein: to prevent dG.dA mispairing in DNA replication. *Proceedings of the National Academy of Sciences of the United States of America* 86, 3949-3952.

Alldread, R. M., Nicholls, D. J., Sundaram, T. K., Scawen, M. D. & Atkinson, T. (1992). Overexpression of the Thermus aquaticus B malate dehydrogenase-encoding gene in Escherichia coli. *Gene* 114, 139-143.

Andersson, G. E. & Sharp, P. M. (1996). Codon usage in the Mycobacterium tuberculosis complex. *Microbiology (Reading, England)* 142 (Pt 4), 915-925.

Aravind, L., Walker, D. R. & Koonin, E. V. (1999). Conserved domains in DNA repair proteins and evolution of repair systems. *Nucleic Acids Res* 27, 1223-1242.

Aravind, L. & Koonin, E. V. (2001). Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system. *Genome Res* 11, 1365-1374.

Asagoshi, K., Odawara, H., Nakano, H., Miyano, T., Terato, H., Ohyama, Y., Seki, S. & Ide, H. (2000a). Comparison of substrate specificities of Escherichia coli endonuclease III and its mouse homologue (mNTH1) using defined oligonucleotide substrates. *Biochemistry* 39, 11389-11398.

Asagoshi, K., Yamada, T., Okada, Y., Terato, H., Ohyama, Y., Seki, S. & Ide, H. (2000b). Recognition of formamidopyrimidine by Escherichia coli and mammalian thymine glycol glycosylases. Distinctive paired base effects and biological and mechanistic implications. *The Journal of biological chemistry* 275, 24781-24786.

Aspinwall, R., Rothwell, D. G., Roldan-Arjona, T. & other authors (1997). Cloning and characterization of a functional human homolog of Escherichia coli endonuclease III. *Proceedings of the National Academy of Sciences of the United States of America* 94, 109-114.

Assenmacher, N., Wenig, K., Lammens, A. & Hopfner, K. P. (2006). Structural basis for transcription-coupled repair: the N terminus of Mfd resembles UvrB with degenerate ATPase motifs. *Journal of molecular biology* 355, 675-683.

Au, K. G., Clark, S., Miller, J. H. & Modrich, P. (1989). Escherichia coli mutY gene encodes an adenine glycosylase active on G-A mispairs. *Proceedings of the National Academy of Sciences of the United States of America* 86, 8877-8881.

Augeri, L., Lee, Y. M., Barton, A. B. & Doetsch, P. W. (1997). Purification, characterization, gene cloning, and expression of Saccharomyces cerevisiae redoxyendonuclease, a homolog of Escherichia coli endonuclease III. *Biochemistry* 36, 721-729.

Ban, C. & Yang, W. (1998a). Crystal structure and ATPase activity of MutL: implications for DNA repair and mutagenesis. *Cell* 95, 541-552.

Ban, C. & Yang, W. (1998b). Structural basis for MutH activation in E.coli mismatch repair and relationship of MutH to restriction endonucleases. *The EMBO journal* 17, 1526-1534.

Bandaru, V., Sunkara, S., Wallace, S. S. & Bond, J. P. (2002). A novel human DNA glycosylase that removes oxidative DNA damage and is homologous to Escherichia coli endonuclease VIII. *DNA repair* 1, 517-529.

Bandaru, V., Blaisdell, J. O. & Wallace, S. S. (2006). Oxidative DNA glycosylases: recipes from cloning to characterization. *Methods in enzymology* 408, 15-33.

Bandaru, V., Zhao, X., Newton, M. R., Burrows, C. J. & Wallace, S. S. (2007). Human endonuclease VIII-like (NEIL) proteins in the giant DNA Mimivirus. *DNA Repair (Amst)* 6, 1629-1641.

Basu, A. K., Loechler, E. L., Leadon, S. A. & Essigmann, J. M. (1989). Genetic effects of thymine glycol: site-specific mutagenesis and molecular modeling studies. *Proceedings of the National Academy of Sciences of the United States of America* 86, 7677-7681.

Baughman, G. & Nomura, M. (1983). Localization of the target site for translational regulation of the L11 operon and direct evidence for translational coupling in Escherichia coli. *Cell* 34, 979-988.

Bedale, W. A. & Cox, M. (1996). Evidence for the coupling of ATP hydrolysis to the final (extension) phase of RecA protein-mediated DNA strand exchange. *The Journal of biological chemistry* 271, 5725-5732.

Bhatnagar, S. K. & Bessman, M. J. (1988). Studies on the mutator gene, mutT of Escherichia coli. Molecular cloning of the gene, purification of the gene product, and identification of a novel nucleoside triphosphatase. *The Journal of biological chemistry* 263, 8953-8957.

Bhatnagar, S. K., Bullions, L. C. & Bessman, M. J. (1991). Characterization of the mutT nucleoside triphosphatase of Escherichia coli. *The Journal of biological chemistry* 266, 9050-9054.

Bianco, P. R., Tracy, R. B. & Kowalczykowski, S. C. (1998). DNA strand exchange proteins: a biochemical and physical comparison. *Front Biosci* 3, D570-603.

Birikh, K. R., Lebedenko, E. N., Boni, I. V. & Berlin, Y. A. (1995). A high-level prokaryotic expression system: synthesis of human interleukin 1 alpha and its receptor antagonist. *Gene* 164, 341-345.

Blaisdell, J. O., Hatahet, Z. & Wallace, S. S. (1999). A novel role for Escherichia coli endonuclease VIII in prevention of spontaneous G-->T transversions. *Journal of bacteriology* 181, 6396-6402.

Blaisdell, J. O. & Wallace, S. S. (2007). Rapid determination of the active fraction of DNA repair glycosylases: a novel fluorescence assay for trapped intermediates. *Nucleic Acids Res* 35, 1601-1611.

Bliss, T. M. & Lane, D. P. (1997). Ku selectively transfers between DNA molecules with homologous ends. *The Journal of biological chemistry* 272, 5765-5773.

Boiteux, S. & Laval, J. (1982). Coding properties of poly(deoxycytidylic acid) templates containing uracil or apyrimidinic sites: in vitro modulation of mutagenesis by deoxyribonucleic acid repair enzymes. *Biochemistry* 21, 6746-6751.

Boiteux, S. & Laval, J. (1983). Imidazole open ring 7-methylguanine: an inhibitor of DNA synthesis. *Biochemical and biophysical research communications* 110, 552-558.

Boiteux, S. & Radicella, J. P. (1999). Base excision repair of 8-hydroxyguanine protects DNA from endogenous oxidative stress. *Biochimie* 81, 59-67.

Boiteux, S. & Radicella, J. P. (2000). The human OGG1 gene: structure, functions, and its implication in the process of carcinogenesis. *Archives of biochemistry and biophysics* 377, 1-8.

Boiteux, S., Gellon, L. & Guibourt, N. (2002). Repair of 8-oxoguanine in Saccharomyces cerevisiae: interplay of DNA repair and replication mechanisms. *Free radical biology & medicine* 32, 1244-1253.

Boshoff, H. I., Reed, M. B., Barry, C. E., 3rd & Mizrahi, V. (2003). DnaE2 polymerase contributes to in vivo survival and the emergence of drug resistance in Mycobacterium tuberculosis. *Cell* 113, 183-193.

Boshoff, H. I., Myers, T. G., Copp, B. R., McNeil, M. R., Wilson, M. A. & Barry, C. E., 3rd (2004). The transcriptional responses of Mycobacterium tuberculosis to inhibitors of metabolism: novel insights into drug mechanisms of action. *The Journal of biological chemistry* 279, 40174-40184.

Bowater, R. & Doherty, A. J. (2006). Making ends meet: repairing breaks in bacterial DNA by non-homologous end-joining. *PLoS genetics* 2, e8.

Breimer, L. H. & Lindahl, T. (1985). Enzymatic excision of DNA bases damaged by exposure to ionizing radiation or oxidizing agents. *Mutation research* 150, 85-89.

Brenner, S. L., Zlotnick, A. & Griffith, J. D. (1988). RecA protein self-assembly. Multiple discrete aggregation states. *Journal of molecular biology* 204, 959-972.

Brooks, P. C., Movahedzadeh, F. & Davis, E. O. (2001). Identification of some DNA damage-inducible genes of Mycobacterium tuberculosis: apparent lack of correlation with LexA binding. *Journal of bacteriology* 183, 4459-4467.

Buchko, G. W., Hess, N. J., Bandaru, V., Wallace, S. S. & Kennedy, M. A. (2000). Spectroscopic studies of zinc(II)- and cobalt(II)-associated Escherichia coli formamidopyrimidine-DNA glycosylase: extended X-ray absorption fine structure evidence for a metal-binding domain. *Biochemistry* 39, 12441-12449.

Burdett, V., Baitinger, C., Viswanathan, M., Lovett, S. T. & Modrich, P. (2001). In vivo requirement for RecJ, ExoVII, ExoI, and ExoX in methyl-directed mismatch repair. *Proceedings of the National Academy of Sciences of the United States of America* 98, 6765-6770.

Burrows, C. J., Muller, J. G., Kornyushyna, O., Luo, W., Duarte, V., Leipold, M. D. & David, S. S. (2002). Structure and potential mutagenicity of new hydantoin products from guanosine and 8-oxo-7,8-dihydroguanine oxidation by transition metals. *Environmental health perspectives* 110 Suppl 5, 713-717.

Cabrera, M., Nghiem, Y. & Miller, J. H. (1988). mutM, a second mutator locus in Escherichia coli that generates G.C----T.A transversions. *Journal of bacteriology* 170, 5405-5407.

Cabusora, L., Sutton, E., Fulmer, A. & Forst, C. V. (2005). Differential network expression during drug and stress response. *Bioinformatics* 21, 2898-2905.

Cai, H., Bloom, L. B., Eritja, R. & Goodman, M. F. (1993). Kinetics of deoxyribonucleotide insertion and extension at abasic template lesions in different sequence contexts using HIV-1 reverse transcriptase. *The Journal of biological chemistry* 268, 23567-23572.

Castaing, B., Boiteux, S. & Zelwer, C. (1992). DNA containing a chemically reduced apurinic site is a high affinity ligand for the E. coli formamidopyrimidine-DNA glycosylase. *Nucleic acids research* 20, 389-394.

Castaing, B., Geiger, A., Seliger, H., Nehls, P., Laval, J., Zelwer, C. & Boiteux, S. (1993). Cleavage and binding of a DNA fragment containing a single 8-oxoguanine by wild type and mutant FPG proteins. *Nucleic acids research* 21, 2899-2905.

Chen, D. S., Law, C. & Keng, P. (1993). Reduction of radiation cytotoxicity by human apurinic endonuclease in a radiation-sensitive Escherichia coli mutant. *Radiat Res* 135, 405-410.

Chen, L., Trujillo, K., Sung, P. & Tomkinson, A. E. (2000). Interactions of the DNA ligase IV-XRCC4 complex with DNA ends and the DNA-dependent protein kinase. *The Journal of biological chemistry* 275, 26196-26205.

Cheng, K. C., Cahill, D. S., Kasai, H., Nishimura, S. & Loeb, L. A. (1992). 8-Hydroxyguanine, an abundant form of oxidative DNA damage, causes G----T and A----C substitutions. *The Journal of biological chemistry* 267, 166-172.

Cheo, D. L., Bayles, K. W. & Yasbin, R. E. (1993). Elucidation of regulatory elements that control damage induction and competence induction of the Bacillus subtilis SOS system. *Journal of bacteriology* 175, 5907-5915.

Chung, M. H., Kasai, H., Jones, D. S., Inoue, H., Ishikawa, H., Ohtsuka, E. & Nishimura, S. (1991). An endonuclease activity of Escherichia coli that specifically removes 8-hydroxyguanine residues from DNA. *Mutation research* 254, 1-12.

Clark, J. M. & Beardsley, G. P. (1986). Thymine glycol lesions terminate chain elongation by DNA polymerase I in vitro. *Nucleic acids research* 14, 737-749.

Clark, J. M. & Beardsley, G. P. (1989). Template length, sequence context, and 3'-5' exonuclease activity modulate replicative bypass of thymine glycol lesions in vitro. *Biochemistry* 28, 775-779.

Cole, S. T., Brosch, R., Parkhill, J. & other authors (1998). Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature* 393, 537-544.

Cole, S. T., Eiglmeier, K., Parkhill, J. & other authors (2001). Massive gene decay in the leprosy bacillus. *Nature* 409, 1007-1011.

Cromie, G. A., Connelly, J. C. & Leach, D. R. (2001). Recombination at double-strand breaks and DNA ends: conserved mechanisms from phage to humans. *Molecular cell* 8, 1163-1174.

Cunningham, R. P. & Weiss, B. (1985). Endonuclease III (nth) mutants of Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* 82, 474-478.

Cunningham, R. P., Saporito, S. M., Spitzer, S. G. & Weiss, B. (1986). Endonuclease IV (nfo) mutant of Escherichia coli. *Journal of bacteriology* 168, 1120-1127.

Curti, E., Smerdon, S. J. & Davis, E. O. (2007). Characterization of the helicase activity and substrate specificity of Mycobacterium tuberculosis UvrD. *Journal of bacteriology* 189, 1542-1555.

D'Ham, C., Romieu, A., Jaquinod, M., Gasparutto, D. & Cadet, J. (1999). Excision of 5,6-dihydroxy-5,6-dihydrothymine, 5,6-dihydrothymine, and 5-hydroxycytosine from defined sequence oligonucleotides by Escherichia coli endonuclease III and Fpg proteins: kinetic and mechanistic aspects. *Biochemistry* 38, 3335-3344.

Dao, V. & Modrich, P. (1998). Mismatch-, MutS-, MutL-, and helicase II-dependent unwinding from the single-strand break of an incised heteroduplex. *The Journal of biological chemistry* 273, 9202-9207.

Darwin, K. H. & Nathan, C. F. (2005). Role for nucleotide excision repair in virulence of Mycobacterium tuberculosis. *Infect Immun* 73, 4581-4587.

Daugelat, S., Kowall, J., Mattow, J., Bumann, D., Winter, R., Hurwitz, R. & Kaufmann, S. H. (2003). The RD1 proteins of Mycobacterium tuberculosis: expression in Mycobacterium smegmatis and biochemical characterization. *Microbes and infection / Institut Pasteur* 5, 1082-1095.

David, S. S. & Williams, S. D. (1998). Chemistry of Glycosylases and Endonucleases Involved in Base-Excision Repair. *Chem Rev* 98, 1221-1262.

David, S. S., O'Shea, V. L. & Kundu, S. (2007). Base-excision repair of oxidative DNA damage. *Nature* 447, 941-950.

Davis, E. O., Sedgwick, S. G. & Colston, M. J. (1991). Novel structure of the recA locus of Mycobacterium tuberculosis implies processing of the gene product. *Journal of bacteriology* 173, 5653-5662.

Davis, E. O., Dullaghan, E. M. & Rand, L. (2002a). Definition of the mycobacterial SOS box and use to identify LexA-regulated genes in Mycobacterium tuberculosis. *Journal of bacteriology* 184, 3287-3295.

Davis, E. O., Springer, B., Gopaul, K. K., Papavinasasundaram, K. G., Sander, P. & Bottger, E. C. (2002b). DNA damage induction of recA in Mycobacterium tuberculosis independently of RecA and LexA. *Molecular microbiology* 46, 791-800.

de Miranda, A. B., Alvarez-Valin, F., Jabbari, K., Degrave, W. M. & Bernardi, G. (2000). Gene expression, amino acid conservation, and hydrophobicity are the main factors shaping codon preferences in Mycobacterium tuberculosis and Mycobacterium leprae. *Journal of molecular evolution* 50, 45-55.

de Smit, M. H. & van Duin, J. (1990). Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. *Proceedings of the National Academy of Sciences of the United States of America* 87, 7668-7672.

Deaconescu, A. M., Chambers, A. L., Smith, A. J., Nickels, B. E., Hochschild, A., Savery, N. J. & Darst, S. A. (2006). Structural basis for bacterial transcription-coupled DNA repair. *Cell* 124, 507-520.

Delaney, S., Neeley, W. L., Delaney, J. C. & Essigmann, J. M. (2007). The substrate specificity of MutY for hyperoxidized guanine lesions in vivo. *Biochemistry* 46, 1448-1455.

Della, M., Palmbos, P. L., Tseng, H. M. & other authors (2004). Mycobacterial Ku and ligase proteins constitute a two-component NHEJ repair machine. *Science (New York, NY* 306, 683-685.

DellaVecchia, M. J., Croteau, D. L., Skorvaga, M., Dezhurov, S. V., Lavrik, O. I. & Van Houten, B. (2004). Analyzing the handoff of DNA from UvrA to UvrB utilizing DNA-protein photoaffinity labeling. *The Journal of biological chemistry* 279, 45245-45256.

Demple, B. & Harrison, L. (1994). Repair of oxidative damage to DNA: enzymology and biology. *Annu Rev Biochem* 63, 915-948.

Denver, D. R., Swenson, S. L. & Lynch, M. (2003). An Evolutionary Analysis of the Helix-Hairpin-Helix Superfamily of DNA Repair Glycosylases. *Mol Biol Evol* 20, 1603-1611.

Dinglay, S., Trewick, S. C., Lindahl, T. & Sedgwick, B. (2000). Defective processing of methylated single-stranded DNA by E. coli AlkB mutants. *Genes & development* 14, 2097-2105.

Dizdaroglu, M. (1991). Chemical determination of free radical-induced damage to DNA. *Free radical biology & medicine* 10, 225-242.

Dizdaroglu, M., Karahalil, B., Senturker, S., Buckley, T. J. & Roldan-Arjona, T. (1999). Excision of products of oxidative DNA base damage by human NTH1 protein. *Biochemistry* 38, 243-246.

Dizdaroglu, M., Bauche, C., Rodriguez, H. & Laval, J. (2000). Novel substrates of Escherichia coli nth protein and its kinetics for excision of modified bases from DNA damaged by free radicals. *Biochemistry* 39, 5586-5592.

Dizdaroglu, M. (2005). Base-excision repair of oxidative DNA damage by DNA glycosylases. *Mutat Res* 591, 45-59.

Dizdaroglu, M., Kirkali, G. & Jaruga, P. (2008). Formamidopyrimidines in DNA: mechanisms of formation, repair, and biological effects. *Free radical biology & medicine* 45, 1610-1621.

Dodson, M. L., Michaels, M. L. & Lloyd, R. S. (1994). Unified catalytic mechanism for DNA glycosylases. *The Journal of biological chemistry* 269, 32709-32712.

Doherty, A. J., Jackson, S. P. & Weller, G. R. (2001). Identification of bacterial homologues of the Ku DNA repair proteins. *FEBS letters* 500, 186-188.

Dos Vultos, T., Blazquez, J., Rauzier, J., Matic, I. & Gicquel, B. (2006). Identification of Nudix hydrolase family members with an antimutator role in Mycobacterium tuberculosis and Mycobacterium smegmatis. *Journal of bacteriology* 188, 3159-3161.

Dos Vultos, T., Mestre, O., Rauzier, J. & other authors (2008). Evolution and diversity of clonal bacteria: the paradigm of Mycobacterium tuberculosis. *PloS one* 3, e1538.

Dos Vultos, T., Mestre, O., Tonjum, T. & Gicquel, B. (2009). DNA repair in Mycobacterium tuberculosis revisited. *FEMS Microbiol Rev* 33, 471-487.

Dou, H., Mitra, S. & Hazra, T. K. (2003). Repair of oxidized bases in DNA bubble structures by human DNA glycosylases NEIL1 and NEIL2. *The Journal of biological chemistry* 278, 49679-49684.

Dou, H., Theriot, C. A., Das, A., Hegde, M. L., Matsumoto, Y., Boldogh, I., Hazra, T. K., Bhakat, K. K. & Mitra, S. (2008). Interaction of the human DNA glycosylase NEIL1 with proliferating cell nuclear antigen. The potential for replication-associated repair of oxidized bases in mammalian genomes. *J Biol Chem* 283, 3130-3140.

Doublie, S., Bandaru, V., Bond, J. P. & Wallace, S. S. (2004). The crystal structure of human endonuclease VIII-like 1 (NEIL1) reveals a zincless finger motif required for glycosylase activity. *Proceedings of the National Academy of Sciences of the United States of America* 101, 10284-10289.

Dudasova, Z., Dudas, A. & Chovanec, M. (2004). Non-homologous end-joining factors of Saccharomyces cerevisiae. *FEMS microbiology reviews* 28, 581-601.

Duncan, B. K. & Weiss, B. (1982). Specific mutator effects of ung (uracil-DNA glycosylase) mutations in Escherichia coli. *Journal of bacteriology* 151, 750-755.

Ehrlich, M., Zhang, X. Y. & Inamdar, N. M. (1990). Spontaneous deamination of cytosine and 5-methylcytosine residues in DNA and replacement of 5-methylcytosine residues with cytosine residues. *Mutation research* 238, 277-286.

Ehrt, S., Guo, X. V., Hickey, C. M., Ryou, M., Monteleone, M., Riley, L. W. & Schnappinger, D. (2005). Controlling gene expression in mycobacteria with anhydrotetracycline and Tet repressor. *Nucleic acids research* 33, e21.

Eide, L., Bjoras, M., Pirovano, M., Alseth, I., Berdal, K. G. & Seeberg, E. (1996). Base excision of oxidative purine and pyrimidine DNA damage in Saccharomyces cerevisiae by a DNA glycosylase with sequence similarity to endonuclease III from Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* 93, 10735-10740.

Eide, L., Luna, L., Gustad, E. C., Henderson, P. T., Essigmann, J. M., Demple, B. & Seeberg, E. (2001). Human endonuclease III acts preferentially on DNA damage opposite guanine residues in DNA. *Biochemistry* 40, 6653-6659.

Eisen, J. A. & Hanawalt, P. C. (1999). A phylogenomic study of DNA repair genes, proteins, and processes. *Mutation research* 435, 171-213.

Evans, J., Maccabee, M., Hatahet, Z., Courcelle, J., Bockrath, R., Ide, H. & Wallace, S. (1993). Thymine ring saturation and fragmentation products: lesion bypass, misinsertion and implications for mutagenesis. *Mutation research* 299, 147-156.

Feig, D. I., Sowers, L. C. & Loeb, L. A. (1994). Reverse chemical mutagenesis: identification of the mutagenic lesions resulting from reactive oxygen species-mediated

damage to DNA. *Proceedings of the National Academy of Sciences of the United States of America* 91, 6609-6613.

Frenkel, K., Goldstein, M. S. & Teebor, G. W. (1981). Identification of the cis-thymine glycol moiety in chemically oxidized and gamma-irradiated deoxyribonucleic acid by high-pressure liquid chromatography analysis. *Biochemistry* 20, 7566-7571.

Frenkel, K., Cummings, A., Solomon, J., Cadet, J., Steinberg, J. J. & Teebor, G. W. (1985). Quantitative determination of the 5-(hydroxymethyl)uracil moiety in the DNA of gamma-irradiated cells. *Biochemistry* 24, 4527-4533.

Friedberg, E. C., Walker, G. C., Siede, W., Wood, R. D., Schultz, R. A. & Ellenberger, T. (2006). *DNA repair and mutagenesis (second edition)*: ASM Press.

Frischkorn, K., Sander, P., Scholz, M., Teschner, K., Prammananan, T. & Bottger, E. C. (1998). Investigation of mycobacterial recA function: protein introns in the RecA of pathogenic mycobacteria do not affect competency for homologous recombination. *Molecular microbiology* 29, 1203-1214.

Fromme, J. C. & Verdine, G. L. (2002). Structural insights into lesion recognition and repair by the bacterial 8-oxoguanine DNA glycosylase MutM. *Nature structural biology* 9, 544-552.

Fromme, J. C. & Verdine, G. L. (2003a). DNA lesion recognition by the bacterial repair enzyme MutM. *The Journal of biological chemistry* 278, 51543-51548.

Fromme, J. C. & Verdine, G. L. (2003b). Structure of a trapped endonuclease III-DNA covalent intermediate. *The EMBO journal* 22, 3461-3471.

Fromme, J. C., Banerjee, A. & Verdine, G. L. (2004). DNA glycosylase recognition and catalysis. *Current opinion in structural biology* 14, 43-49.

Fu, L. M. & Shinnick, T. M. (2007). Genome-wide exploration of the drug action of capreomycin on Mycobacterium tuberculosis using Affymetrix oligonucleotide GeneChips. *The Journal of infection* 54, 277-284.

Gamulin, V., Cetkovic, H. & Ahel, I. (2004). Identification of a promoter motif regulating the major DNA damage response mechanism of Mycobacterium tuberculosis. *FEMS microbiology letters* 238, 57-63.

Gao, M. J. & Murphy, T. M. (2001). Alternative forms of formamidopyrimidine-DNA glycosylase from Arabidopsis thaliana. *Photochemistry and photobiology* 73, 128-134.

Gericke, G. S. (2006). Is there an emerging endosymbiotic relationship between mycobacteria and the human host based on horizontal transfer of genetic sequences? *Med Hypotheses* 67, 1419-1428.

Gilboa, R., Zharkov, D. O., Golan, G., Fernandes, A. S., Gerchman, S. E., Matz, E., Kycia, J. H., Grollman, A. P. & Shoham, G. (2002). Structure of formamidopyrimidine-DNA glycosylase covalently complexed to DNA. *The Journal of biological chemistry* 277, 19811-19816.

Girard, P. M. & Boiteux, S. (1997). Repair of oxidized DNA bases in the yeast Saccharomyces cerevisiae. *Biochimie* 79, 559-566.

Girard, P. M., Guibourt, N. & 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 research* 25, 3204-3211.

Goldman, E., Rosenberg, A. H., Zubay, G. & Studier, F. W. (1995). Consecutive low-usage leucine codons block translation only when near the 5' end of a message in Escherichia coli. *Journal of molecular biology* 245, 467-473.

Gong, C., Martins, A., Bongiorno, P., Glickman, M. & Shuman, S. (2004). Biochemical and genetic analysis of the four DNA ligases of mycobacteria. *The Journal of biological chemistry* 279, 20594-20606.

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

Goodman, M. F., Creighton, S., Bloom, L. B. & Petruska, J. (1993). Biochemical basis of DNA replication fidelity. *Critical reviews in biochemistry and molecular biology* 28, 83-126.

Goosen, N. & Moolenaar, G. F. (2001). Role of ATP hydrolysis by UvrA and UvrB during nucleotide excision repair. *Research in microbiology* 152, 401-409.

Goosen, N. & Moolenaar, G. F. (2008). Repair of UV damage in bacteria. *DNA repair* 7, 353-379.

Graham, J. E. & Clark-Curtiss, J. E. (1999). Identification of Mycobacterium tuberculosis RNAs synthesized in response to phagocytosis by human macrophages by selective capture of transcribed sequences (SCOTS). *Proceedings of the National Academy of Sciences of the United States of America* 96, 11554-11559.

Grawunder, U., Wilm, M., Wu, X., Kulesza, P., Wilson, T. E., Mann, M. & Lieber, M. R. (1997). Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. *Nature* 388, 492-495.

Greer, S. & Zamenhof, S. (1962). Studies on depurination of DNA by heat. *Journal of molecular biology* 4, 123-141.

Griswold, K. E., Mahmood, N. A., Iverson, B. L. & Georgiou, G. (2003). Effects of codon usage versus putative 5'-mRNA structure on the expression of Fusarium solani cutinase in the Escherichia coli cytoplasm. *Protein expression and purification* 27, 134-142.

Grosset, J. (2003). Mycobacterium tuberculosis in the extracellular compartment: an underestimated adversary. *Antimicrobial agents and chemotherapy* 47, 833-836.

Guarne, A., Ramon-Maiques, S., Wolff, E. M., Ghirlando, R., Hu, X., Miller, J. H. & Yang, W. (2004). Structure of the MutL C-terminal domain: a model of intact MutL and its roles in mismatch repair. *The EMBO journal* 23, 4134-4145.

Guo, Y., Wallace, S. S. & Bandaru, V. (2009). A novel bicistronic vector for overexpressing Mycobacterium tuberculosis proteins in Escherichia coli. *Protein Expr Purif* 65, 230-237.

Guschlbauer, W., Duplaa, A. M., Guy, A., Teoule, R. & Fazakerley, G. V. (1991). Structure and in vitro replication of DNA templates containing 7,8-dihydro-8-oxoadenine. *Nucleic acids research* 19, 1753-1758.

Haber, L. T. & Walker, G. C. (1991). Altering the conserved nucleotide binding motif in the Salmonella typhimurium MutS mismatch repair protein affects both its ATPase and mismatch binding activities. *The EMBO journal* 10, 2707-2715.

Hagen, U. (1986). Current aspects on the radiation induced base damage in DNA. *Radiation and environmental biophysics* 25, 261-271.

Hailer, M. K., Slade, P. G., Martin, B. D., Rosenquist, T. A. & Sugden, K. D. (2005a). Recognition of the oxidized lesions spiroiminodihydantoin and guanidinohydantoin in DNA by the mammalian base excision repair glycosylases NEIL1 and NEIL2. *DNA Repair (Amst)* 4, 41-50.

Hailer, M. K., Slade, P. G., Martin, B. D. & Sugden, K. D. (2005b). Nei deficient Escherichia coli are sensitive to chromate and accumulate the oxidized guanine lesion spiroiminodihydantoin. *Chemical research in toxicology* 18, 1378-1383.

Hall, M. C. & Matson, S. W. (1999). The Escherichia coli MutL protein physically interacts with MutH and stimulates the MutH-associated endonuclease activity. *The Journal of biological chemistry* 274, 1306-1312.

Hall, M. N., Gabay, J., Debarbouille, M. & Schwartz, M. (1982). A role for mRNA secondary structure in the control of translation initiation. *Nature* 295, 616-618.

Hannig, G. & Makrides, S. C. (1998). Strategies for optimizing heterologous protein expression in Escherichia coli. *Trends in biotechnology* 16, 54-60.

Harth, G., Lee, B. Y. & Horwitz, M. A. (1997). High-level heterologous expression and secretion in rapidly growing nonpathogenic mycobacteria of four major Mycobacterium tuberculosis extracellular proteins considered to be leading vaccine candidates and drug targets. *Infection and immunity* 65, 2321-2328.

Hatahet, Z., Kow, Y. W., Purmal, A. A., Cunningham, R. P. & Wallace, S. S. (1994). New substrates for old enzymes. 5-Hydroxy-2'-deoxycytidine and 5-hydroxy-2'-deoxyuridine are substrates for Escherichia coli endonuclease III and formamidopyrimidine DNA N-glycosylase, while 5-hydroxy-2'-deoxyuridine is a substrate for uracil DNA N-glycosylase. *The Journal of biological chemistry* 269, 18814-18820.

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

Hazra, T. K., Izumi, T., Venkataraman, R., Kow, Y. W., Dizdaroglu, M. & Mitra, S. (2000). Characterization of a novel 8-oxoguanine-DNA glycosylase activity in Escherichia coli and identification of the enzyme as endonuclease VIII. *The Journal of biological chemistry* 275, 27762-27767.

Hazra, T. K., Muller, J. G., Manuel, R. C., Burrows, C. J., Lloyd, R. S. & Mitra, S. (2001). Repair of hydantoins, one electron oxidation product of 8-oxoguanine, by DNA glycosylases of Escherichia coli. *Nucleic Acids Res* 29, 1967-1974.

Hazra, T. K., Izumi, T., Boldogh, I., Imhoff, B., Kow, Y. W., Jaruga, P., Dizdaroglu, M. & Mitra, S. (2002a). Identification and characterization of a human DNA glycosylase for repair of modified bases in oxidatively damaged DNA. *Proceedings of the National Academy of Sciences of the United States of America* 99, 3523-3528.

Hazra, T. K., Kow, Y. W., Hatahet, Z., Imhoff, B., Boldogh, I., Mokkapati, S. K., Mitra, S. & Izumi, T. (2002b). Identification and characterization of a novel human DNA glycosylase for repair of cytosine-derived lesions. *The Journal of biological chemistry* 277, 30417-30420.

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

Hevroni, D. & Livneh, Z. (1988). Bypass and termination at apurinic sites during replication of single-stranded DNA in vitro: a model for apurinic site mutagenesis. *Proceedings of the National Academy of Sciences of the United States of America* 85, 5046-5050.

Hilbert, T. P., Boorstein, R. J., Kung, H. C., Bolton, P. H., Xing, D., Cunningham, R. P. & Teebor, G. W. (1996). Purification of a mammalian homologue of Escherichia coli endonuclease III: identification of a bovine pyrimidine hydrate-thymine glycol DNAse/AP lyase by irreversible cross linking to a thymine glycol-containing oligoxynucleotide. *Biochemistry* 35, 2505-2511.

Hilbert, T. P., Chaung, W., Boorstein, R. J., Cunningham, R. P. & Teebor, G. W. (1997). Cloning and expression of the cDNA encoding the human homologue of the DNA repair enzyme, Escherichia coli endonuclease III. *The Journal of biological chemistry* 272, 6733-6740.

Howard-Flanders, P., Boyce, R. P. & Theriot, L. (1966). Three loci in Escherichia coli K-12 that control the excision of pyrimidine dimers and certain other mutagen products from DNA. *Genetics* 53, 1119-1136.

Ide, H., Kow, Y. W. & Wallace, S. S. (1985). Thymine glycols and urea residues in M13 DNA constitute replicative blocks in vitro. *Nucleic acids research* 13, 8035-8052.

Ide, H., Petrullo, L. A., Hatahet, Z. & Wallace, S. S. (1991). Processing of DNA base damage by DNA polymerases. Dihydrothymine and beta-ureidoisobutyric acid as models for instructive and noninstructive lesions. *The Journal of biological chemistry* 266, 1469-1477.

Ikeda, S., Biswas, T., Roy, R., Izumi, T., Boldogh, I., Kurosky, A., Sarker, A. H., Seki, S. & Mitra, S. (1998). Purification and characterization of human NTH1, a homolog of Escherichia coli endonuclease III. Direct identification of Lys-212 as the active nucleophilic residue. *The Journal of biological chemistry* 273, 21585-21593.

Iserentant, D. & Fiers, W. (1980). Secondary structure of mRNA and efficiency of translation initiation. *Gene* 9, 1-12.

Ishchenko, A. A., Bulychev, N. V., Maksakova, G. A., Johnson, F. & Nevinsky, G. A. (1999a). Single-stranded oligodeoxyribonucleotides are substrates of Fpg protein from Escherichia coli. *IUBMB Life* 48, 613-618.

Ishchenko, A. A., Koval, V. V., Fedorova, O. S., Douglas, K. T. & Nevinsky, G. A. (1999b). Structural requirements of double and single stranded DNA substrates and inhibitors, including a photoaffinity label, of Fpg protein from Escherichia coli. *J Biomol Struct Dyn* 17, 301-310.

Ishida, M. & Oshima, T. (1994). Overexpression of genes of an extreme thermophile Thermus thermophilus, in Escherichia coli cells. *Journal of bacteriology* 176, 2767-2770.

Ishida, M. & Oshima, T. (2002). Effective structure of a leader open reading frame for enhancing the expression of GC-rich genes. *Journal of biochemistry* 132, 63-70.

Iwai, S. (2001). Synthesis and thermodynamic studies of oligonucleotides containing the two isomers of thymine glycol. *Chemistry* 7, 4343-4351.

Jain, R., Kumar, P. & Varshney, U. (2007). A distinct role of formamidopyrimidine DNA glycosylase (MutM) in down-regulation of accumulation of G, C mutations and protection against oxidative stress in mycobacteria. *DNA Repair (Amst)* 6, 1774-1785.

Jaruga, P., Birincioglu, M., Rosenquist, T. A. & Dizdaroglu, M. (2004). Mouse NEIL1 protein is specific for excision of 2,6-diamino-4-hydroxy-5-formamidopyrimidine and 4,6-diamino-5-formamidopyrimidine from oxidatively damaged DNA. *Biochemistry* 43, 15909-15914.

Jiang, D., Hatahet, Z., Blaisdell, J. O., Melamede, R. J. & Wallace, S. S. (1997a). Escherichia coli endonuclease VIII: cloning, sequencing, and overexpression of the nei structural gene and characterization of nei and nei nth mutants. *J Bacteriol* 179, 3773-3782.

Jiang, D., Hatahet, Z., Melamede, R. J., Kow, Y. W. & Wallace, S. S. (1997b). Characterization of Escherichia coli endonuclease VIII. *J Biol Chem* 272, 32230-32239.

Junop, M. S., Obmolova, G., Rausch, K., Hsieh, P. & Yang, W. (2001). Composite active site of an ABC ATPase: MutS uses ATP to verify mismatch recognition and authorize DNA repair. *Molecular cell* 7, 1-12.

Jurado, J., Saparbaev, M., Matray, T. J., Greenberg, M. M. & Laval, J. (1998). The ring fragmentation product of thymidine C5-hydrate when present in DNA is repaired by the Escherichia coli Fpg and Nth proteins. *Biochemistry* 37, 7757-7763.

Kamiya, H. & Kasai, H. (1995). Formation of 2-hydroxydeoxyadenosine triphosphate, an oxidatively damaged nucleotide, and its incorporation by DNA polymerases. Steady-

state kinetics of the incorporation. *The Journal of biological chemistry* 270, 19446-19450.

Kamiya, H., Miura, H., Murata-Kamiya, N. & other authors (1995a). 8-Hydroxyadenine (7,8-dihydro-8-oxoadenine) induces misincorporation in in vitro DNA synthesis and mutations in NIH 3T3 cells. *Nucleic acids research* 23, 2893-2899.

Kamiya, H., Ueda, T., Ohgi, T., Matsukage, A. & Kasai, H. (1995b). Misincorporation of dAMP opposite 2-hydroxyadenine, an oxidative form of adenine. *Nucleic acids research* 23, 761-766.

Kamiya, H. & Kasai, H. (1996). Effect of sequence contexts on misincorporation of nucleotides opposite 2-hydroxyadenine. *FEBS letters* 391, 113-116.

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

Kamiya, H. & Kasai, H. (1997b). Mutations induced by 2-hydroxyadenine on a shuttle vector during leading and lagging strand syntheses in mammalian cells. *Biochemistry* 36, 11125-11130.

Kane, J. F. (1995). Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli. *Current opinion in biotechnology* 6, 494-500.

Karahalil, B., Roldan-Arjona, T. & Dizdaroglu, M. (1998). Substrate specificity of Schizosaccharomyces pombe Nth protein for products of oxidative DNA damage. *Biochemistry* 37, 590-595.

Kasai, H., Iida, A., Yamaizumi, Z., Nishimura, S. & Tanooka, H. (1990). 5-Formyldeoxyuridine: a new type of DNA damage induced by ionizing radiation and its mutagenicity to salmonella strain TA102. *Mutation research* 243, 249-253.

Katafuchi, A., Nakano, T., Masaoka, A., Terato, H., Iwai, S., Hanaoka, F. & Ide, H. (2004). Differential specificity of human and Escherichia coli endonuclease III and VIII homologues for oxidative base lesions. *J Biol Chem* 279, 14464-14471.

Kathe, S. D., Barrantes-Reynolds, R., Jaruga, P., Newton, M. R., Burrows, C. J., Bandaru, V., Dizdaroglu, M., Bond, J. P. & Wallace, S. S. (2009). Plant and fungal Fpg homologs are formamidopyrimidine DNA glycosylases but not 8-oxoguanine DNA glycosylases. *DNA repair* 8, 643-653.

Kim, S. T. & Sancar, A. (1993). Photochemistry, photophysics, and mechanism of pyrimidine dimer repair by DNA photolyase. *Photochemistry and photobiology* 57, 895-904.

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

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

Kouchakdjian, M., Bodepudi, V., Shibutani, S., Eisenberg, M., Johnson, F., Grollman, A. P. & Patel, D. J. (1991). NMR structural studies of the ionizing radiation adduct 7-hydro-8-oxodeoxyguanosine (8-oxo-7H-dG) opposite deoxyadenosine in a DNA duplex. 8-Oxo-7H-dG(syn).dA(anti) alignment at lesion site. *Biochemistry* 30, 1403-1412.

Kramer, B., Kramer, W. & Fritz, H. J. (1984). Different base/base mismatches are corrected with different efficiencies by the methyl-directed DNA mismatch-repair system of E. coli. *Cell* 38, 879-887.

Krejci, L., Chen, L., Van Komen, S., Sung, P. & Tomkinson, A. (2003). Mending the break: two DNA double-strand break repair machines in eukaryotes. *Progress in nucleic acid research and molecular biology* 74, 159-201.

Kreutzer, D. A. & Essigmann, J. M. (1998). Oxidized, deaminated cytosines are a source of C --> T transitions in vivo. *Proceedings of the National Academy of Sciences of the United States of America* 95, 3578-3582.

Krishnamurthy, N., Muller, J. G., Burrows, C. J. & David, S. S. (2007a). Unusual structural features of hydantoin lesions translate into efficient recognition by Escherichia coli Fpg. *Biochemistry* 46, 9355-9365.

Krishnamurthy, N., Muller, J. G., Burrows, C. J. & David, S. S. (2007b). Unusual structural features of hydantoin lesions translate into efficient recognition by Escherichia coli Fpg. *Biochemistry* 46, 9355-9365.

Krishnamurthy, N., Zhao, X., Burrows, C. J. & David, S. S. (2008). Superior removal of hydantoin lesions relative to other oxidized bases by the human DNA glycosylase hNEIL1. *Biochemistry* 47, 7137-7146.

Krokan, H. E., Nilsen, H., Skorpen, F., Otterlei, M. & Slupphaug, G. (2000). Base excision repair of DNA in mammalian cells. *FEBS Lett* 476, 73-77.

Krokeide, S. Z., Bolstad, N., Laerdahl, J. K., Bjoras, M. & Luna, L. (2009). Expression and purification of NEIL3, a human DNA glycosylase homolog. *Protein expression and purification* 65, 160-164.

Kunkel, T. A., Schaaper, R. M. & Loeb, L. A. (1983). Depurination-induced infidelity of deoxyribonucleic acid synthesis with purified deoxyribonucleic acid replication proteins in vitro. *Biochemistry* 22, 2378-2384.

Kunkel, T. A. (1984). Mutational specificity of depurination. *Proceedings of the National Academy of Sciences of the United States of America* 81, 1494-1498.

Kunz, C., Saito, Y. & Schar, P. (2009). DNA Repair in mammalian cells: Mismatched repair: variations on a theme. *Cell Mol Life Sci* 66, 1021-1038.

Kuo, C. F., McRee, D. E., Cunningham, R. P. & Tainer, J. A. (1992). Crystallization and crystallographic characterization of the iron-sulfur-containing DNA-repair enzyme endonuclease III from Escherichia coli. *Journal of molecular biology* 227, 347-351.

Kurland, C. & Gallant, J. (1996). Errors of heterologous protein expression. *Current opinion in biotechnology* 7, 489-493.

Kurthkoti, K., Srinath, T., Kumar, P., Malshetty, V. S., Sang, P. B., Jain, R., Manjunath, R. & Varshney, U. (2009). A distinct physiological role of MutY in mutation prevention in mycobacteria. *Microbiology* in press.

Kuzminov, A. (1999). Recombinational repair of DNA damage in Escherichia coli and bacteriophage lambda. *Microbiol Mol Biol Rev* 63, 751-813, table of contents.

Lakey, D. L., Voladri, R. K., Edwards, K. M., Hager, C., Samten, B., Wallis, R. S., Barnes, P. F. & Kernodle, D. S. (2000). Enhanced production of recombinant Mycobacterium tuberculosis antigens in Escherichia coli by replacement of low-usage codons. *Infection and immunity* 68, 233-238.

Laspia, M. F. & Wallace, S. S. (1988). Excision repair of thymine glycols, urea residues, and apurinic sites in Escherichia coli. *Journal of bacteriology* 170, 3359-3366.

Lavrukhin, O. V. & Lloyd, R. S. (2000). Involvement of phylogenetically conserved acidic amino acid residues in catalysis by an oxidative DNA damage enzyme formamidopyrimidine glycosylase. *Biochemistry* 39, 15266-15271.

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

Leadon, S. A. (1996). Repair of DNA Damage Produced by Ionizing Radiation: A Minireview. *Semin Radiat Oncol* 6, 295-305.

Lee, J. Y., Chang, J., Joseph, N., Ghirlando, R., Rao, D. N. & Yang, W. (2005). MutH complexed with hemi- and unmethylated DNAs: coupling base recognition and DNA cleavage. *Molecular cell* 20, 155-166.

Lees-Miller, S. P. & Meek, K. (2003). Repair of DNA double strand breaks by non-homologous end joining. *Biochimie* 85, 1161-1173.

Leipold, M. D., Muller, J. G., Burrows, C. J. & David, S. S. (2000a). Removal of hydantoin products of 8-oxoguanine oxidation by the Escherichia coli DNA repair enzyme, FPG. *Biochemistry* 39, 14984-14992.

Leipold, M. D., Muller, J. G., Burrows, C. J. & David, S. S. (2000b). Removal of hydantoin products of 8-oxoguanine oxidation by the Escherichia coli DNA repair enzyme, FPG. *Biochemistry* 39, 14984-14992.

Levy, D. D. & Teebor, G. W. (1991). Site directed substitution of 5-hydroxymethyluracil for thymine in replicating phi X-174am3 DNA via synthesis of 5-hydroxymethyl-2'-deoxyuridine-5'-triphosphate. *Nucleic acids research* 19, 3337-3343.

Li, F., Liu, Q., Chen, Y. Y., Yu, Z. N., Zhang, Z. P., Zhou, Y. F., Deng, J. Y., Bi, L. J. & Zhang, X. E. (2008). Escherichia coli mismatch repair protein MutL interacts with the clamp loader subunits of DNA polymerase III. *Mutation research* 637, 101-110.

Lindahl, T. (1979). DNA glycosylases, endonucleases for apurinic/apyrimidinic sites, and base excision-repair. *Progress in nucleic acid research and molecular biology* 22, 135-192.

Lindahl, T. & Barnes, D. E. (2000). Repair of endogenous DNA damage. *Cold Spring Harb Symp Quant Biol* 65, 127-133.

Little, J. W. & Mount, D. W. (1982). The SOS regulatory system of Escherichia coli. *Cell* 29, 11-22.

Loeb, L. A. & Preston, B. D. (1986). Mutagenesis by apurinic/apyrimidinic sites. *Annual review of genetics* 20, 201-230.

Loughlin, M. F., Barnard, F. M., Jenkins, D., Sharples, G. J. & Jenks, P. J. (2003). Helicobacter pylori mutants defective in RuvC Holliday junction resolvase display reduced macrophage survival and spontaneous clearance from the murine gastric mucosa. *Infection and immunity* 71, 2022-2031.

Lu, A. L., Clark, S. & Modrich, P. (1983). Methyl-directed repair of DNA base-pair mismatches in vitro. *Proceedings of the National Academy of Sciences of the United States of America* 80, 4639-4643.

Lu, A. L. & Chang, D. Y. (1988). A novel nucleotide excision repair for the conversion of an A/G mismatch to C/G base pair in E. coli. *Cell* 54, 805-812.

Luo, W., Muller, J. G. & Burrows, C. J. (2001). The pH-dependent role of superoxide in riboflavin-catalyzed photooxidation of 8-oxo-7,8-dihydroguanosine. *Organic letters* 3, 2801-2804.

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

Mahenthiralingam, E., Draper, P., Davis, E. O. & Colston, M. J. (1993). Cloning and sequencing of the gene which encodes the highly inducible acetamidase of Mycobacterium smegmatis. *Journal of general microbiology* 139, 575-583.

Maki, H. & Sekiguchi, M. (1992). MutT protein specifically hydrolyses a potent mutagenic substrate for DNA synthesis. *Nature* 355, 273-275.

Makoff, A. J., Oxer, M. D., Romanos, M. A., Fairweather, N. F. & Ballantine, S. (1989). Expression of tetanus toxin fragment C in E. coli: high level expression by removing rare codons. *Nucleic acids research* 17, 10191-10202.

Makoff, A. J. & Smallwood, A. E. (1990). The use of two-cistron constructions in improving the expression of a heterologous gene in E. coli. *Nucleic acids research* 18, 1711-1718.

Malins, D. C. & Haimanot, R. (1990). 4,6-Diamino-5-formamidopyrimidine, 8-hydroxyguanine and 8-hydroxyadenine in DNA from neoplastic liver of English sole exposed to carcinogens. *Biochemical and biophysical research communications* 173, 614-619.

Masaoka, A., Terato, H., Kobayashi, M., Honsho, A., Ohyama, Y. & Ide, H. (1999). Enzymatic repair of 5-formyluracil. I. Excision of 5-formyluracil site-specifically incorporated into oligonucleotide substrates by alka protein (Escherichia coli 3-

methyladenine DNA glycosylase II). *The Journal of biological chemistry* 274, 25136-25143.

Mazur, S. J. & Grossman, L. (1991). Dimerization of Escherichia coli UvrA and its binding to undamaged and ultraviolet light damaged DNA. *Biochemistry* 30, 4432-4443.

McAuley-Hecht, K. E., Leonard, G. A., Gibson, N. J., Thomson, J. B., Watson, W. P., Hunter, W. N. & Brown, T. (1994). Crystal structure of a DNA duplex containing 8-hydroxydeoxyguanine-adenine base pairs. *Biochemistry* 33, 10266-10270.

McNulty, D. E., Claffee, B. A., Huddleston, M. J., Porter, M. L., Cavnar, K. M. & Kane, J. F. (2003). Mistranslational errors associated with the rare arginine codon CGG in Escherichia coli. *Protein expression and purification* 27, 365-374.

McTigue, M. M., Rieger, R. A., Rosenquist, T. A., Iden, C. R. & De Los Santos, C. R. (2004). Stereoselective excision of thymine glycol lesions by mammalian cell extracts. *DNA repair* 3, 313-322.

Melamede, R. J., Hatahet, Z., Kow, Y. W., Ide, H. & Wallace, S. S. (1994). Isolation and characterization of endonuclease VIII from Escherichia coli. *Biochemistry* 33, 1255-1264.

Mellon, I. & Hanawalt, P. C. (1989). Induction of the Escherichia coli lactose operon selectively increases repair of its transcribed DNA strand. *Nature* 342, 95-98.

Michaels, M. L., Pham, L., Cruz, C. & Miller, J. H. (1991). MutM, a protein that prevents G.C----T.A transversions, is formamidopyrimidine-DNA glycosylase. *Nucleic acids research* 19, 3629-3632.

Michaels, M. L., Cruz, C., Grollman, A. P. & Miller, J. H. (1992). Evidence that MutY and MutM combine to prevent mutations by an oxidatively damaged form of guanine in DNA. *Proceedings of the National Academy of Sciences of the United States of America* 89, 7022-7025.

Michaels, M. L. & Miller, J. H. (1992). The GO system protects organisms from the mutagenic effect of the spontaneous lesion 8-hydroxyguanine (7,8-dihydro-8-oxoguanine). *Journal of bacteriology* 174, 6321-6325.

Miller, H., Fernandes, A. S., Zaika, E., McTigue, M. M., Torres, M. C., Wente, M., Iden, C. R. & Grollman, A. P. (2004). Stereoselective excision of thymine glycol from oxidatively damaged DNA. *Nucleic acids research* 32, 338-345.

Mishina, Y., Duguid, E. M. & He, C. (2006). Direct reversal of DNA alkylation damage. *Chem Rev* 106, 215-232.

Misiaszek, R., Crean, C., Joffe, A., Geacintov, N. E. & Shafirovich, V. (2004). Oxidative DNA damage associated with combination of guanine and superoxide radicals and repair mechanisms via radical trapping. *The Journal of biological chemistry* 279, 32106-32115.

Miyabe, I., Zhang, Q. M., Sugiyama, H., Kino, K. & Yonei, S. (2001). Mutagenic effects of 5-formyluracil on a plasmid vector during replication in Escherichia coli. *International journal of radiation biology* 77, 53-58.

Mizrahi, V. & Andersen, S. J. (1998). DNA repair in Mycobacterium tuberculosis. What have we learnt from the genome sequence? *Molecular microbiology* 29, 1331-1339.

Moolenaar, G. F., Herron, M. F., Monaco, V., van der Marel, G. A., van Boom, J. H., Visse, R. & Goosen, N. (2000). The role of ATP binding and hydrolysis by UvrB during nucleotide excision repair. *The Journal of biological chemistry* 275, 8044-8050.

Moran, E. & Wallace, S. S. (1985). The role of specific DNA base damages in the X-ray-induced inactivation of bacteriophage PM2. *Mutation research* 146, 229-241.

Mori, T., Hori, Y. & Dizdaroglu, M. (1993). DNA base damage generated in vivo in hepatic chromatin of mice upon whole body gamma-irradiation. *International journal of radiation biology* 64, 645-650.

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

Moriya, M. (1993). Single-stranded shuttle phagemid for mutagenesis studies in mammalian cells: 8-oxoguanine in DNA induces targeted G.C-->T.A transversions in simian kidney cells. *Proceedings of the National Academy of Sciences of the United States of America* 90, 1122-1126.

Moriya, M. & Grollman, A. P. (1993). Mutations in the mutY gene of Escherichia coli enhance the frequency of targeted G:C-->T:a transversions induced by a single 8-oxoguanine residue in single-stranded DNA. *Mol Gen Genet* 239, 72-76.

Morland, I., Rolseth, V., Luna, L., Rognes, T., Bjoras, M. & Seeberg, E. (2002). Human DNA glycosylases of the bacterial Fpg/MutM superfamily: an alternative pathway for the repair of 8-oxoguanine and other oxidation products in DNA. *Nucleic acids research* 30, 4926-4936.

Mosbaugh, D. W. (1988). Purification and characterization of porcine liver DNA polymerase gamma: utilization of dUTP and dTTP during in vitro DNA synthesis. *Nucleic acids research* 16, 5645-5659.

Movahedzadeh, F., Colston, M. J. & Davis, E. O. (1997a). Determination of DNA sequences required for regulated Mycobacterium tuberculosis RecA expression in response to DNA-damaging agents suggests that two modes of regulation exist. *Journal of bacteriology* 179, 3509-3518.

Movahedzadeh, F., Colston, M. J. & Davis, E. O. (1997b). Characterization of Mycobacterium tuberculosis LexA: recognition of a Cheo (Bacillus-type SOS) box. *Microbiology (Reading, England)* 143 (Pt 3), 929-936.

Murphy, T. M. & George, A. (2005). A comparison of two DNA base excision repair glycosylases from Arabidopsis thaliana. *Biochemical and biophysical research communications* 329, 869-872.

Nackerdien, Z., Kasprzak, K. S., Rao, G., Halliwell, B. & Dizdaroglu, M. (1991). Nickel(II)- and cobalt(II)-dependent damage by hydrogen peroxide to the DNA bases in isolated human chromatin. *Cancer research* 51, 5837-5842.

Nash, H. M., Bruner, S. D., Scharer, O. D., Kawate, T., Addona, T. A., Spooner, E., Lane, W. S. & Verdine, G. L. (1996). Cloning of a yeast 8-oxoguanine DNA glycosylase reveals the existence of a base-excision DNA-repair protein superfamily. *Curr Biol* 6, 968-980.

Neeley, W. L. & Essigmann, J. M. (2006). Mechanisms of formation, genotoxicity, and mutation of guanine oxidation products. *Chemical research in toxicology* 19, 491-505.

Neto, J. B., Gentil, A., Cabral, R. E. & Sarasin, A. (1992). Mutation spectrum of heat-induced abasic sites on a single-stranded shuttle vector replicated in mammalian cells. *J Biol Chem* 267, 19718-19723.

Nghiem, Y., Cabrera, M., Cupples, C. G. & Miller, J. H. (1988). The mutY gene: a mutator locus in Escherichia coli that generates G.C----T.A transversions. *Proceedings of the National Academy of Sciences of the United States of America* 85, 2709-2713.

Nick McElhinny, S. A., Snowden, C. M., McCarville, J. & Ramsden, D. A. (2000). Ku recruits the XRCC4-ligase IV complex to DNA ends. *Molecular and cellular biology* 20, 2996-3003.

Nieminuszczy, J. & Grzesiuk, E. (2007). Bacterial DNA repair genes and their eukaryotic homologues: 3. AlkB dioxygenase and Ada methyltransferase in the direct repair of alkylated DNA. *Acta biochimica Polonica* 54, 459-468.

O'Brien, P. J. (2006). Catalytic promiscuity and the divergent evolution of DNA repair enzymes. *Chem Rev* 106, 720-752.

O'Connor, T. R., Boiteux, S. & Laval, J. (1988). Ring-opened 7-methylguanine residues in DNA are a block to in vitro DNA synthesis. *Nucleic acids research* 16, 5879-5894.

O'Connor, T. R., Graves, R. J., de Murcia, G., Castaing, B. & Laval, J. (1993). Fpg protein of Escherichia coli is a zinc finger protein whose cysteine residues have a structural and/or functional role. *The Journal of biological chemistry* 268, 9063-9070.

O'Rourke, E. J., Chevalier, C., Pinto, A. V., Thiberge, J. M., Ielpi, L., Labigne, A. & Radicella, J. P. (2003). Pathogen DNA as target for host-generated oxidative stress: role for repair of bacterial DNA damage in Helicobacter pylori colonization. *Proceedings of the National Academy of Sciences of the United States of America* 100, 2789-2794.

O'Sullivan, D. M., McHugh, T. D. & Gillespie, S. H. (2005). Analysis of rpoB and pncA mutations in the published literature: an insight into the role of oxidative stress in Mycobacterium tuberculosis evolution? *The Journal of antimicrobial chemotherapy* 55, 674-679.

Ocampo-Hafalla, M. T., Altamirano, A., Basu, A. K., Chan, M. K., Ocampo, J. E., Cummings, A., Jr., Boorstein, R. J., Cunningham, R. P. & Teebor, G. W. (2006). Repair of thymine glycol by hNth1 and hNeil1 is modulated by base pairing and cis-trans epimerization. *DNA repair* 5, 444-454.

Oda, Y., Uesugi, S., Ikehara, M., Nishimura, S., Kawase, Y., Ishikawa, H., Inoue, H. & Ohtsuka, E. (1991). NMR studies of a DNA containing 8-hydroxydeoxyguanosine. *Nucleic acids research* 19, 1407-1412.

Ohtsubo, T., Matsuda, O., Iba, K., Terashima, I., Sekiguchi, M. & Nakabeppu, Y. (1998). Molecular cloning of AtMMH, an Arabidopsis thaliana ortholog of the Escherichia coli mutM gene, and analysis of functional domains of its product. *Mol Gen Genet* 259, 577-590.

Olinski, R., Zastawny, T., Budzbon, J., Skokowski, J., Zegarski, W. & Dizdaroglu, M. (1992). DNA base modifications in chromatin of human cancerous tissues. *FEBS letters* 309, 193-198.

Olsen, I., Balasingham, S. V., Davidsen, T., Debebe, E., Rodland, E. A., van Soolingen, D., Kremer, K., Alseth, I. & Tonjum, T. (2009). Characterization of the major

formamidopyrimidine-DNA glycosylase homolog in Mycobacterium tuberculosis and its linkage to variable tandem repeats. *FEMS immunology and medical microbiology* 56, 151-161.

Orren, D. K. & Sancar, A. (1990). Formation and enzymatic properties of the UvrB.DNA complex. *The Journal of biological chemistry* 265, 15796-15803.

Pfeifer, G. P. (1997). Formation and processing of UV photoproducts: effects of DNA sequence and chromatin environment. *Photochemistry and photobiology* 65, 270-283.

Pitcher, R. S., Tonkin, L. M., Green, A. J. & Doherty, A. J. (2005a). Domain structure of a NHEJ DNA repair ligase from Mycobacterium tuberculosis. *Journal of molecular biology* 351, 531-544.

Pitcher, R. S., Wilson, T. E. & Doherty, A. J. (2005b). New insights into NHEJ repair processes in prokaryotes. *Cell cycle (Georgetown, Tex* 4, 675-678.

Pitcher, R. S., Brissett, N. C. & Doherty, A. J. (2007a). Nonhomologous end-joining in bacteria: a microbial perspective. *Annual review of microbiology* 61, 259-282.

Pitcher, R. S., Brissett, N. C., Picher, A. J., Andrade, P., Juarez, R., Thompson, D., Fox, G. C., Blanco, L. & Doherty, A. J. (2007b). Structure and function of a mycobacterial NHEJ DNA repair polymerase. *Journal of molecular biology* 366, 391-405.

Pitcher, R. S., Green, A. J., Brzostek, A., Korycka-Machala, M., Dziadek, J. & Doherty, A. J. (2007c). NHEJ protects mycobacteria in stationary phase against the harmful effects of desiccation. *DNA repair* 6, 1271-1276.

Pumo, D., Barrantes-Reynolds, R., Kathe, S., Wallace, S. & Bond, J. (2009). Evolution of the Fpg/Nei family of DNA glycosylases. In *Genetic and Epigenetic Regulation of Genome Stability and Transgenerational Response to Stress*, pp. in press. Hauppauge, New York: Nova Science Publishers, Inc.

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

Purmal, A. A., Lampman, G. W., Kow, Y. W. & Wallace, S. S. (1994b). The sequence context-dependent mispairing of 5-hydroxycytosine and 5-hydroxyuridine in vitro. *Annals of the New York Academy of Sciences* 726, 361-363.

Purmal, A. A., Bond, J. P., Lyons, B. A., Kow, Y. W. & Wallace, S. S. (1998). Uracil glycol deoxynucleoside triphosphate is a better substrate for DNA polymerase I

Klenow fragment than thymine glycol deoxynucleoside triphosphate. *Biochemistry* 37, 330-338.

Purnapatre, K. & Varshney, U. (1998). Uracil DNA glycosylase from Mycobacterium smegmatis and its distinct biochemical properties. *European journal of biochemistry / FEBS* 256, 580-588.

Radicella, J. P., Clark, E. A. & Fox, M. S. (1988). Some mismatch repair activities in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* 85, 9674-9678.

Radman, M. (1998). DNA replication: one strand may be more equal. *Proceedings of the National Academy of Sciences of the United States of America* 95, 9718-9719.

Ramsden, D. A. & Gellert, M. (1998). Ku protein stimulates DNA end joining by mammalian DNA ligases: a direct role for Ku in repair of DNA double-strand breaks. *The EMBO journal* 17, 609-614.

Rand, L., Hinds, J., Springer, B., Sander, P., Buxton, R. S. & Davis, E. O. (2003). The majority of inducible DNA repair genes in Mycobacterium tuberculosis are induced independently of RecA. *Mol Microbiol* 50, 1031-1042.

Randall, S. K., Eritja, R., Kaplan, B. E., Petruska, J. & Goodman, M. F. (1987). Nucleotide insertion kinetics opposite abasic lesions in DNA. *The Journal of biological chemistry* 262, 6864-6870.

Rengarajan, J., Bloom, B. R. & Rubin, E. J. (2005). Genome-wide requirements for Mycobacterium tuberculosis adaptation and survival in macrophages. *Proc Natl Acad Sci U S A* 102, 8327-8332.

Reuven, N. B., Koonin, E. V., Rudd, K. E. & Deutscher, M. P. (1995). The gene for the longest known Escherichia coli protein is a member of helicase superfamily II. *J Bacteriol* 177, 5393-5400.

Robey-Bond, S. M., Barrantes-Reynolds, R., Bond, J. P., Wallace, S. S. & Bandaru, V. (2008). Clostridium acetobutylicum 8-Oxoguanine DNA Glycosylase (Ogg) Differs from Eukaryotic Oggs with Respect to Opposite Base Discrimination. *Biochemistry* 47, 7626-7636.

Rogers, P. A., Eide, L., Klungland, A. & Ding, H. (2003). Reversible inactivation of E. coli endonuclease III via modification of its [4Fe-4S] cluster by nitric oxide. *DNA repair* 2, 809-817.

Roldan-Arjona, T., Anselmino, C. & Lindahl, T. (1996). Molecular cloning and functional analysis of a Schizosaccharomyces pombe homologue of Escherichia coli endonuclease III. *Nucleic acids research* 24, 3307-3312.

Rosenquist, T. A., Zaika, E., Fernandes, A. S., Zharkov, D. O., Miller, H. & Grollman, A. P. (2003). The novel DNA glycosylase, NEIL1, protects mammalian cells from radiation-mediated cell death. *DNA repair* 2, 581-591.

Rouet, P. & Essigmann, J. M. (1985). Possible role for thymine glycol in the selective inhibition of DNA synthesis on oxidized DNA templates. *Cancer research* 45, 6113-6118.

Sagher, D. & Strauss, B. (1983). Insertion of nucleotides opposite apurinic/apyrimidinic sites in deoxyribonucleic acid during in vitro synthesis: uniqueness of adenine nucleotides. *Biochemistry* 22, 4518-4526.

Saito, Y., Ishii, Y., Niwa, M. & Ueda, I. (1987). Direct expression of a synthetic somatomedin C gene in Escherichia coli by use of a two-cistron system. *Journal of biochemistry* 101, 1281-1288.

Saito, Y., Uraki, F., Nakajima, S., Asaeda, A., Ono, K., Kubo, K. & Yamamoto, K. (1997). Characterization of endonuclease III (nth) and endonuclease VIII (nei) mutants of Escherichia coli K-12. *Journal of bacteriology* 179, 3783-3785.

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

Sancar, A. & Rupp, W. D. (1983). A novel repair enzyme: UVRABC excision nuclease of Escherichia coli cuts a DNA strand on both sides of the damaged region. *Cell* 33, 249-260.

Sarker, A. H., Ikeda, S., Nakano, H. & other authors (1998). Cloning and characterization of a mouse homologue (mNthl1) of Escherichia coli endonuclease III. *Journal of molecular biology* 282, 761-774.

Sartori, A. A., Fitz-Gibbon, S., Yang, H., Miller, J. H. & Jiricny, J. (2002). A novel uracil-DNA glycosylase with broad substrate specificity and an unusual active site. *The EMBO journal* 21, 3182-3191.

Sassetti, C. M. & Rubin, E. J. (2003). Genetic requirements for mycobacterial survival during infection. *Proc Natl Acad Sci U S A* 100, 12989-12994.

Savery, N. J. (2007). The molecular mechanism of transcription-coupled DNA repair. *Trends in microbiology* 15, 326-333.

Schaaper, R. M. & Loeb, L. A. (1981). Depurination causes mutations in SOS-induced cells. *Proceedings of the National Academy of Sciences of the United States of America* 78, 1773-1777.

Schaaper, R. M., Kunkel, T. A. & Loeb, L. A. (1983). Infidelity of DNA synthesis associated with bypass of apurinic sites. *Proceedings of the National Academy of Sciences of the United States of America* 80, 487-491.

Schaaper, R. M. & Dunn, R. L. (1987). Escherichia coli mutT mutator effect during in vitro DNA synthesis. Enhanced A.G replicational errors. *The Journal of biological chemistry* 262, 16267-16270.

Schnappinger, D., Ehrt, S., Voskuil, M. I. & other authors (2003). Transcriptional Adaptation of Mycobacterium tuberculosis within Macrophages: Insights into the Phagosomal Environment. *J Exp Med* 198, 693-704.

Schoner, B. E., Hsiung, H. M., Belagaje, R. M., Mayne, N. G. & Schoner, R. G. (1984). Role of mRNA translational efficiency in bovine growth hormone expression in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* 81, 5403-5407.

Schoner, B. E., Belagaje, R. M. & Schoner, R. G. (1986). Translation of a synthetic two-cistron mRNA in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* 83, 8506-8510.

Schoner, B. E., Belagaje, R. M. & Schoner, R. G. (1990). Enhanced translational efficiency with two-cistron expression system. *Methods in enzymology* 185, 94-103.

Schumperli, D., McKenney, K., Sobieski, D. A. & Rosenberg, M. (1982). Translational coupling at an intercistronic boundary of the Escherichia coli galactose operon. *Cell* 30, 865-871.

Sedgwick, B. & Vaughan, P. (1991). Widespread adaptive response against environmental methylating agents in microorganisms. *Mutation research* 250, 211-221.

Sedgwick, B. & Lindahl, T. (2002). Recent progress on the Ada response for inducible repair of DNA alkylation damage. *Oncogene* 21, 8886-8894.

Selby, C. P., Witkin, E. M. & Sancar, A. (1991). Escherichia coli mfd mutant deficient in "mutation frequency decline" lacks strand-specific repair: in vitro complementation with purified coupling factor. *Proceedings of the National Academy of Sciences of the United States of America* 88, 11574-11578.

Selby, C. P. & Sancar, A. (1993). Molecular mechanism of transcription-repair coupling. *Science (New York, NY* 260, 53-58.

Selby, C. P. & Sancar, A. (1994). Mechanisms of transcription-repair coupling and mutation frequency decline. *Microbiological reviews* 58, 317-329.

Serre, L., Pereira de Jesus, K., Boiteux, S., Zelwer, C. & Castaing, B. (2002). Crystal structure of the Lactococcus lactis formamidopyrimidine-DNA glycosylase bound to an abasic site analogue-containing DNA. *The EMBO journal* 21, 2854-2865.

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

Shibutani, S., Bodepudi, V., Johnson, F. & Grollman, A. P. (1993). Translesional synthesis on DNA templates containing 8-oxo-7,8-dihydrodeoxyadenosine. *Biochemistry* 32, 4615-4621.

Shimizu, T., Manabe, K., Yoshikawa, S., Kawasaki, Y. & Iwai, S. (2006). Preferential formation of (5S,6R)-thymine glycol for oligodeoxyribonucleotide synthesis and analysis of drug binding to thymine glycol-containing DNA. *Nucleic Acids Res* 34, 313-321.

Shuman, S. & Glickman, M. S. (2007). Bacterial DNA repair by non-homologous end joining. *Nature reviews* 5, 852-861.

Sidorenko, V. S., Rot, M. A., Filipenko, M. L., Nevinsky, G. A. & Zharkov, D. O. (2008). Novel DNA glycosylases from Mycobacterium tuberculosis. *Biochemistry (Mosc)* 73, 442-450.

Sinha, K. M., Stephanou, N. C., Gao, F., Glickman, M. S. & Shuman, S. (2007). Mycobacterial UvrD1 is a Ku-dependent DNA helicase that plays a role in multiple DNA repair events, including double-strand break repair. *The Journal of biological chemistry* 282, 15114-15125.

Slade, P. G., Hailer, M. K., Martin, B. D. & Sugden, K. D. (2005). Guanine-specific oxidation of double-stranded DNA by Cr(VI) and ascorbic acid forms spiroiminodihydantoin and 8-oxo-2'-deoxyguanosine. *Chemical research in toxicology* 18, 1140-1149.

Sorensen, M. A., Kurland, C. G. & Pedersen, S. (1989). Codon usage determines translation rate in Escherichia coli. *Journal of molecular biology* 207, 365-377.

Spampinato, C. & Modrich, P. (2000). The MutL ATPase is required for mismatch repair. *The Journal of biological chemistry* 275, 9863-9869.

Spanjaard, R. A., van Dijk, M. C., Turion, A. J. & van Duin, J. (1989). Expression of the rat interferon-alpha 1 gene in Escherichia coli controlled by the secondary structure of the translation-initiation region. *Gene* 80, 345-351.

Spanjaard, R. A. & van Duin, J. (1989). Translational reinitiation in the presence and absence of a Shine and Dalgarno sequence. *Nucleic acids research* 17, 5501-5507.

Srinath, T., Bharti, S. K. & Varshney, U. (2007). Substrate specificities and functional characterization of a thermo-tolerant uracil DNA glycosylase (UdgB) from Mycobacterium tuberculosis. *DNA repair* 6, 1517-1528.

Stillwell, W. G., Xu, H. X., Adkins, J. A., Wishnok, J. S. & Tannenbaum, S. R. (1989). Analysis of methylated and oxidized purines in urine by capillary gas chromatography-mass spectrometry. *Chemical research in toxicology* 2, 94-99.

Studer, S. M. & Joseph, S. (2006). Unfolding of mRNA secondary structure by the bacterial translation initiation complex. *Molecular cell* 22, 105-115.

Sugahara, M., Mikawa, T., Kumasaka, T., Yamamoto, M., Kato, R., Fukuyama, K., Inoue, Y. & Kuramitsu, S. (2000). Crystal structure of a repair enzyme of oxidatively damaged DNA, MutM (Fpg), from an extreme thermophile, Thermus thermophilus HB8. *The EMBO journal* 19, 3857-3869.

Sugden, K. D., Campo, C. K. & Martin, B. D. (2001). Direct oxidation of guanine and 7,8-dihydro-8-oxoguanine in DNA by a high-valent chromium complex: a possible mechanism for chromate genotoxicity. *Chemical research in toxicology* 14, 1315-1322.

Sun, B., Latham, K. A., Dodson, M. L. & Lloyd, R. S. (1995). Studies on the catalytic mechanism of five DNA glycosylases. Probing for enzyme-DNA imino intermediates. *The Journal of biological chemistry* 270, 19501-19508.

Suvarnapunya, A. E., Lagasse, H. A. & Stein, M. A. (2003). The role of DNA base excision repair in the pathogenesis of Salmonella enterica serovar Typhimurium. *Mol Microbiol* 48, 549-559.

Suzuki, T., Tanaka, Y., Ishida, M., Ishizuka, M., Yamagishi, A. & Oshima, T. (1997). Screening of a mutant plasmid with high expression efficiency of GC-rich leuB gene of an extreme thermophile, Thermus thermophilus, in Escherichia coli. *Journal of biochemistry* 121, 1031-1034.

Svejstrup, J. Q. (2002). Mechanisms of transcription-coupled DNA repair. *Nat Rev Mol Cell Biol* 3, 21-29.

Takao, M., Kanno, S., Kobayashi, K., Zhang, Q. M., Yonei, S., van der Horst, G. T. & Yasui, A. (2002). A back-up glycosylase in Nth1 knock-out mice is a functional Nei (endonuclease VIII) homologue. *The Journal of biological chemistry* 277, 42205-42213.

Takao, M., Oohata, Y., Kitadokoro, K., Kobayashi, K., Iwai, S., Yasui, A., Yonei, S. & Zhang, Q. M. (2009). Human Nei-like protein NEIL3 has AP lyase activity specific for single-stranded DNA and confers oxidative stress resistance in Escherichia coli mutant. *Genes Cells* 14, 261-270.

Tchou, J., Kasai, H., Shibutani, S., Chung, M. H., Laval, J., Grollman, A. P. & Nishimura, S. (1991). 8-oxoguanine (8-hydroxyguanine) DNA glycosylase and its substrate specificity. *Proceedings of the National Academy of Sciences of the United States of America* 88, 4690-4694.

Tchou, J., Michaels, M. L., Miller, J. H. & Grollman, A. P. (1993). Function of the zinc finger in Escherichia coli Fpg protein. *The Journal of biological chemistry* 268, 26738-26744.

Tchou, J., Bodepudi, V., Shibutani, S., Antoshechkin, I., Miller, J., Grollman, A. P. & Johnson, F. (1994a). Substrate specificity of Fpg protein. Recognition and cleavage of oxidatively damaged DNA. *The Journal of biological chemistry* 269, 15318-15324.

Tchou, J., Bodepudi, V., Shibutani, S., Antoshechkin, I., Miller, J., Grollman, A. P. & Johnson, F. (1994b). Substrate specificity of Fpg protein. Recognition and cleavage of oxidatively damaged DNA. *J Biol Chem* 269, 15318-15324.

Teebor, G., Cummings, A., Frenkel, K., Shaw, A., Voituriez, L. & Cadet, J. (1987). Quantitative measurement of the diastereoisomers of cis thymidine glycol in gamma-irradiated DNA. *Free Radic Res Commun* 2, 303-309.

Teoule, R. (1987). Radiation-induced DNA damage and its repair. *International journal of radiation biology and related studies in physics, chemistry, and medicine* 51, 573-589.

Theis, K., Skorvaga, M., Machius, M., Nakagawa, N., Van Houten, B. & Kisker, C. (2000). The nucleotide excision repair protein UvrB, a helicase-like enzyme with a catch. *Mutation research* 460, 277-300.

Tiraby, J. G. & Fox, M. S. (1973). Marker discrimination in transformation and mutation of pneumococcus. *Proceedings of the National Academy of Sciences of the United States of America* 70, 3541-3545.

Trewick, S. C., Henshaw, T. F., Hausinger, R. P., Lindahl, T. & Sedgwick, B. (2002). Oxidative demethylation by Escherichia coli AlkB directly reverts DNA base damage. *Nature* 419, 174-178.

Triccas, J. A., Parish, T., Britton, W. J. & Gicquel, B. (1998). An inducible expression system permitting the efficient purification of a recombinant antigen from Mycobacterium smegmatis. *FEMS microbiology letters* 167, 151-156.

Truglio, J. J., Croteau, D. L., Van Houten, B. & Kisker, C. (2006). Prokaryotic nucleotide excision repair: the UvrABC system. *Chem Rev* 106, 233-252.

Tsai-Wu, J. J., Liu, H. F. & Lu, A. L. (1992). Escherichia coli MutY protein has both N-glycosylase and apurinic/apyrimidinic endonuclease activities on A.C and A.G mispairs. *Proceedings of the National Academy of Sciences of the United States of America* 89, 8779-8783.

Tudek, B., Boiteux, S. & Laval, J. (1992). Biological properties of imidazole ring-opened N7-methylguanine in M13mp18 phage DNA. *Nucleic acids research* 20, 3079-3084.

Tudek, B., Graziewicz, M., Kazanova, O., Zastawny, T. H., Obtulowicz, T. & Laval, J. (1999). Mutagenic specificity of imidazole ring-opened 7-methylpurines in M13mp18 phage DNA. *Acta biochimica Polonica* 46, 785-799.

Tye, B. K., Chien, J., Lehman, I. R., Duncan, B. K. & Warner, H. R. (1978). Uracil incorporation: a source of pulse-labeled DNA fragments in the replication of the Escherichia coli chromosome. *Proceedings of the National Academy of Sciences of the United States of America* 75, 233-237.

van der Kemp, P. A., Thomas, D., Barbey, R., de Oliveira, R. & Boiteux, S. (1996). Cloning and expression in Escherichia coli of the OGG1 gene of Saccharomyces cerevisiae, which codes for a DNA glycosylase that excises 7,8-dihydro-8-oxoguanine and 2,6-diamino-4-hydroxy-5-N-methylformamidopyrimidine. *Proceedings of the National Academy of Sciences of the United States of America* 93, 5197-5202.

Van Hemmen, J. J. & Bleichrodt, J. F. (1971). The decomposition of adenine by ionizing radiation. *Radiat Res* 46, 444-456.

Van Houten, B. (1990). Nucleotide excision repair in Escherichia coli. *Microbiological reviews* 54, 18-51.

Van Houten, B. & Snowden, A. (1993). Mechanism of action of the Escherichia coli UvrABC nuclease: clues to the damage recognition problem. *Bioessays* 15, 51-59.

Venkatesh, J., Kumar, P., Krishna, P. S., Manjunath, R. & Varshney, U. (2003). Importance of uracil DNA glycosylase in Pseudomonas aeruginosa and Mycobacterium smegmatis, G+C-rich bacteria, in mutation prevention, tolerance to acidified nitrite, and endurance in mouse macrophages. *J Biol Chem* 278, 24350-24358.

Verhoeven, E. E., van Kesteren, M., Moolenaar, G. F., Visse, R. & Goosen, N. (2000). Catalytic sites for 3' and 5' incision of Escherichia coli nucleotide excision repair are both located in UvrC. *The Journal of biological chemistry* 275, 5120-5123.

Viswanathan, M. & Lovett, S. T. (1999). Exonuclease X of Escherichia coli. A novel 3'-5' DNase and Dnaq superfamily member involved in DNA repair. *The Journal of biological chemistry* 274, 30094-30100.

Wagner, J. R., Hu, C. C. & Ames, B. N. (1992). Endogenous oxidative damage of deoxycytidine in DNA. *Proceedings of the National Academy of Sciences of the United States of America* 89, 3380-3384.

Wallace, S. S. (1998). Enzymatic processing of radiation-induced free radical damage in DNA. *Radiat Res* 150, S60-79.

Wallace, S. S. (2002). Biological consequences of free radical-damaged DNA bases. *Free radical biology & medicine* 33, 1-14.

Wallace, S. S., Bandaru, V., Kathe, S. D. & Bond, J. P. (2003). The enigma of endonuclease VIII. *DNA repair* 2, 441-453.

Weisemann, J. M. & Weinstock, G. M. (1991). The promoter of the recA gene of Escherichia coli. *Biochimie* 73, 457-470.

Weller, G. R., Kysela, B., Roy, R. & other authors (2002). Identification of a DNA nonhomologous end-joining complex in bacteria. *Science (New York, NY* 297, 1686-1689.

Wilson, D. M., 3rd, Sofinowski, T. M. & McNeill, D. R. (2003). Repair mechanisms for oxidative DNA damage. *Front Biosci* 8, d963-981.

Wood, M. L., Dizdaroglu, M., Gajewski, E. & Essigmann, J. M. (1990). Mechanistic studies of ionizing radiation and oxidative mutagenesis: genetic effects of a single 8-hydroxyguanine (7-hydro-8-oxoguanine) residue inserted at a unique site in a viral genome. *Biochemistry* 29, 7024-7032.

Wood, M. L., Esteve, A., Morningstar, M. L., Kuziemko, G. M. & Essigmann, J. M. (1992). Genetic effects of oxidative DNA damage: comparative mutagenesis of 7,8-

dihydro-8-oxoguanine and 7,8-dihydro-8-oxoadenine in Escherichia coli. *Nucleic acids research* 20, 6023-6032.

Wu, T. H. & Marinus, M. G. (1994). Dominant negative mutator mutations in the mutS gene of Escherichia coli. *Journal of bacteriology* 176, 5393-5400.

Yavin, E., Boal, A. K., Stemp, E. D., Boon, E. M., Livingston, A. L., O'Shea, V. L., David, S. S. & Barton, J. K. (2005). Protein-DNA charge transport: redox activation of a DNA repair protein by guanine radical. *Proceedings of the National Academy of Sciences of the United States of America* 102, 3546-3551.

Yavin, E., Stemp, E. D., O'Shea V, L., David, S. S. & Barton, J. K. (2006). Electron trap for DNA-bound repair enzymes: a strategy for DNA-mediated signaling. *Proceedings of the National Academy of Sciences of the United States of America* 103, 3610-3614.

You, H. J., Swanson, R. L., Harrington, C. & other authors (1999). Saccharomyces cerevisiae Ntg1p and Ntg2p: broad specificity N-glycosylases for the repair of oxidative DNA damage in the nucleus and mitochondria. *Biochemistry* 38, 11298-11306.

Zahn, K. (1996). Overexpression of an mRNA dependent on rare codons inhibits protein synthesis and cell growth. *Journal of bacteriology* 178, 2926-2933.

Zhang, Q. M., Sugiyama, H., Miyabe, I., Matsuda, S., Saito, I. & Yonei, S. (1997). Replication of DNA templates containing 5-formyluracil, a major oxidative lesion of thymine in DNA. *Nucleic acids research* 25, 3969-3973.

Zhang, Q. M., Yonekura, S., Takao, M., Yasui, A., Sugiyama, H. & Yonei, S. (2005). DNA glycosylase activities for thymine residues oxidized in the methyl group are functions of the hNEIL1 and hNTH1 enzymes in human cells. *DNA repair* 4, 71-79.

Zharkov, D. O., Rieger, R. A., Iden, C. R. & Grollman, A. P. (1997). NH2-terminal proline acts as a nucleophile in the glycosylase/AP-lyase reaction catalyzed by Escherichia coli formamidopyrimidine-DNA glycosylase (Fpg) protein. *J Biol Chem* 272, 5335-5341.

Zharkov, D. O., Golan, G., Gilboa, R., Fernandes, A. S., Gerchman, S. E., Kycia, J. H., Rieger, R. A., Grollman, A. P. & Shoham, G. (2002). Structural analysis of an Escherichia coli endonuclease VIII covalent reaction intermediate. *Embo J* 21, 789-800.

Zharkov, D. O., Shoham, G. & Grollman, A. P. (2003). Structural characterization of the Fpg family of DNA glycosylases. *DNA repair* 2, 839-862.

Zharkov, D. O. (2008). Base excision DNA repair. *Cell Mol Life Sci* 65, 1544-1565.

Zhou, Z., Schnake, P., Xiao, L. & Lal, A. A. (2004). Enhanced expression of a recombinant malaria candidate vaccine in Escherichia coli by codon optimization. *Protein expression and purification* 34, 87-94.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic acids research* 31, 3406-3415.