

**Characterization of the Searching Mechanism for a DNA Repair Enzyme**

**By**

**Mark Hedglin**

A dissertation submitted in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
(Chemical Biology)  
in the University of Michigan  
2010

Doctoral Committee:

Assistant Professor Patrick J. O'Brien, Chair  
Professor Nils G. Walter  
Associate Professor George A. Garcia  
Associate Professor Bruce A. Palfey

To my amazing wife, Kathleen

## **Acknowledgements**

Attending graduate school at the University of Michigan has been one of the greatest experiences of my life and afforded me the opportunity to interact with so many interesting people and learn amazing science. There are so many people I would like to acknowledge for making my time in Ann Arbor so enjoyable. First, and foremost, I would like to thank my advisor, Patrick J. O'Brien. His passion for science is inspiring and I appreciate his support and encouragement throughout my time in his laboratory. He has provided me the greatest foundation to build a successful scientific career. I also would like to thank the members of my committee for their suggestions on my project over the years and their advice and support for continuing my scientific training. I would like to acknowledge the Biological Chemistry department where I carried out the entirety of my graduate work. All of the staff and faculty helped provide an energetic, nurturing, and collaborative environment to work in. Even though I was not a student in the department, I always felt at home, and I am extremely thankful. I would also like to thank all the members of the O'Brien lab, past and present, particularly Yaru Zhang who I worked with to collect data presented in Chapter 4.

During graduate school, my family was the greatest support system. With a large amount of my time spent away from home, my wife Kathleen has been the most understanding, loving, and supportive spouse anyway could ask for. I'm truly appreciative of everything she's endured. My parents, Frank and Helen, since the day I was born, have been unwavering in their encouragement to pursue anything I ever wanted. Everything was possible in their eyes, as long as you worked for it. Their work ethic was the greatest example and their undying support and encouragement helped carry me through all the grueling times. My father-in-law, Patrick Lavelle, Sr. has also been an amazing source of guidance and support for my wife and I during our time in Ann Arbor. We are both eternally grateful for everything he's done for us.

## Table of Contents

<b>Dedication</b> .....	ii
<b>Acknowledgements</b> .....	iii
<b>List of Figures</b> .....	vi
<b>List of Tables</b> .....	ix
<b>List of Abbreviations</b> .....	x
<b>Chapter</b>	
1. Introduction.....	1
References.....	14
2. Human Alkyladenine DNA Glycosylase Employs a Processive Search for DNA Damage.....	18
Materials and Methods.....	21
Results.....	25
Discussion.....	38
References.....	43
Appendix A.....	48
References.....	58

3. Hopping Enables a DNA Repair Glycosylase to Search Both Strands and Bypass a Bound Protein.....	59
Materials and Methods.....	60
Results and Discussion.....	63
Implications.....	72
References.....	74
Appendix B.....	78
References.....	90
4. Probing the DNA Structural Requirements for Translocation by Alkyladenine DNA Glycosylase.....	92
Materials and Methods.....	93
Results.....	96
Discussion.....	109
References.....	113
Appendix C.....	116
References.....	119
5. The Efficiency of the Search for DNA Damage by Alkyladenine DNA Glycosylase Is Dependent Upon the Nature of the Damage.....	120
Materials and Methods.....	123
Results.....	127
Discussion.....	143
References.....	147
Appendix D.....	151

## List of Figures

Figure 1-1: The four major DNA repair pathways.....	2
Figure 1-2: Two common base lesions resulting from modification of adenine.....	6
Figure 1-3: The base excision repair pathway initiated by a monofunctional glycosylase.....	8
Figure 1-4: Crystal structure of the AAG catalytic domain.....	10
Figure 2-1: Design of a simple substrate to monitor processivity of a DNA repair glycosylase.....	26
Figure 2-2: Characterization of the processivity substrate under single turnover conditions.....	29
Figure 2-3: Single turnover excision of $\epsilon$ A is independent of ionic strength and the two sites are equivalent.....	31
Figure 2-4: The multiple turnover processivity assay.....	32
Figure 2-5: Ionic strength affects multiple turnover excision of $\epsilon$ A at pH 6.1, but not at pH 7.5.....	35
Figure 2-6: Ionic strength affects the processivity of AAG.....	37
Figure 2-7: Electrostatic surface potential of the AAG catalytic domain reveals a positively charged DNA binding surface.....	41
Figure A-1: A small fraction of $\epsilon$ A lesions undergo ring opening during synthesis, deprotection, and/or purification.....	49
Figure A-2: Active site titration of AAG.....	50
Figure A-3: Stability of AAG in the absence of DNA.....	51
Figure A-4: Steady state reaction progress curve for hydrolysis of the dual lesion substrate.....	52
Figure A-5: Multiple sequence alignment of vertebrate AAG amino acid sequences.....	53

Figure A-6: Processivity substrate and definition of DNA fragments produced by base excision, alkaline hydrolysis, and denaturing PAGE.....	54
Figure 3-1: Processivity assays to determine the mechanism of linear diffusion by a DNA repair glycosylase.....	63
Figure 3-2: The ionic strength dependence for the processivity of AAG is inconsistent with a purely sliding model.....	65
Figure 3-3: AAG searches both strands of DNA.....	67
Figure 3-4: Testing the effect of a protein roadblock on linear diffusion by AAG.....	69
Figure 3-5: Effect of bound EcoRI on the processivity of AAG.....	70
Figure 3-6: Pulse-chase processivity assays indicate that AAG can bypass a bound EcoRI dimer.....	71
Figure B-1: Diagram of oligonucleotide substrates and products.....	78
Figure B-2: Native gel electrophoresis of oligonucleotides used in this study.....	81
Figure B-3: Ionic strength dependence for the multiple turnover excision of $\epsilon$ A.....	82
Figure B-4: AAG hops frequently under a wide range of conditions.....	84
Figure B-5: Active site titration of EcoRI.....	85
Figure B-6: Endonuclease activity of EcoRI on the processivity substrate in the presence and absence of unlabeled chase.....	86
Figure B-7. The complete burst of EcoRI cleavage confirms that EcoRI is saturating....	87
Figure B-8: Turnover of $72\epsilon A^2F^2$ by full-length AAG in the pulse-chase processivity assay.....	88
Figure B-9: EcoRI remains bound during the pulse-chase assay.....	89
Figure 4-1: Processivity assays to probe the DNA structural requirements for linear diffusion by a DNA repair glycosylase.....	96
Figure 4-2: Substrates to monitor the ability of AAG to bypass DNA structural perturbations.....	98
Figure 4-3: AAG is efficient at bypassing kinks in DNA.....	100
Figure 4-4: AAG bypasses flexible regions within DNA.....	104

Figure 4-5: AAG is not processive on single-stranded DNA.....	107
Figure 4-6: Diffusion along single-stranded DNA is not required for transfer between adjacent Duplex Regions.....	109
Figure 4-7: AAG bypasses structural perturbations to the B-form, double helix by hopping between the adjacent duplex regions.....	110
Figure 4-8: Transfer between duplex arms is promoted at higher ionic strength.....	111
Figure C-1: Ionic strength dependence for the multiple turnover excision of $\epsilon$ A within kinked substrates.....	117
Figure C-2: Ionic strength dependence for the multiple turnover excision of $\epsilon$ A within substrates containing points of flexibility.....	118
Figure C-3: Single-turnover glycosylase activity of AAG on single- and double-stranded DNA.....	118
Figure 5-1: Assays to monitor the processive behavior of AAG on substrates containing various lesions.....	127
Figure 5-2: The processive behavior of AAG is dependent upon the lesion.....	128
Figure 5-3: Minimal mechanism for AAG-catalyzed base excision.....	130
Figure 5-4: Pulse-chase experiment for measuring the efficiency of lesion recognition ( $E$ ) for $\epsilon$ A.....	131
Figure 5-5: Comparison of $k_{\max}$ and $k_{\text{burst}}$ for Hx excision reveals that AAG is not efficient at excising Hx.....	134
Figure 5-6: Under pre-steady conditions at low ionic strength, AAG removes $\epsilon$ A prior to Hx.....	136
Figure 5-7: Transfer to a new DNA molecule is promoted at higher DNA concentration.....	139
Figure 5-8: Effect of APE1 on multiple-turnover excision and processivity of AAG...	141
Figure D-1: Competitive inhibition of Hx excision by AAG.....	153



## List of Tables

Table A-1: Evaluation of the directional processivity of alkyladenine DNA glycosylase.....	57
Table B-1: Comparison of theoretical and measured processivity values to evaluate whether AAG employs a sliding mechanism.....	83
Table 4-1: Fraction processive of FL AAG.....	99
Table 4-2: Fraction processive of $\Delta 80$ AAG.....	102
Table C-1: Multiple turnover rate constant ( $k_{\text{cat}}$ ) of FL AAG.....	116
Table C-2: Multiple turnover rate constant ( $k_{\text{cat}}$ ) of $\Delta 80$ AAG.....	117
Table 5-1: Comparison of AAG-catalyzed excision of $\epsilon$ A and Hx under single-turnover conditions.....	122
Table 5-2: Relative rates for excision of Hx and $\epsilon$ A on dual lesion substrates.....	137

## List of Abbreviations

A: Adenine (nucleobase)

AAG: Alkyladenine DNA glycosylase, also known as methylpurine DNA glycosylase (MPG) and 3-methyladenine DNA glycosylase

BER: base excision repair

BSA: bovine serum albumin

C: Cytosine (nucleobase)

$\Delta 80$  AAG: N-terminal truncation mutant of AAG in which the N-terminal 79 amino acids have been deleted.

DTT: dithiothreitol

DRP: deoxyribose phosphate

*E*: efficiency of lesion recognition

$\epsilon$ A: 1,*N*<sup>6</sup>-ethenoadenine

EcoRI: *E. coli* restriction endonuclease I

EDTA: ethylenediaminetetraacetic acid

FAM: 6-aminofluorescein

FL AAG: full-length alkyladenine DNA glycosylase

F<sub>p</sub>: fraction processive

FPG: formamidopyrimidine DNA glycosylase

G: Guanine (nucleobase)

Hx: Hypoxanthine

I: ionic strength

NaHEPES: sodium *N*-(2-hydroxyethyl)piperazine-*N'*-(2-ethanesulfonate) NaMES:  
sodium 2-(*N*-morpholino)ethanesulfonate

PAGE: polyacrylamide gel electrophoresis

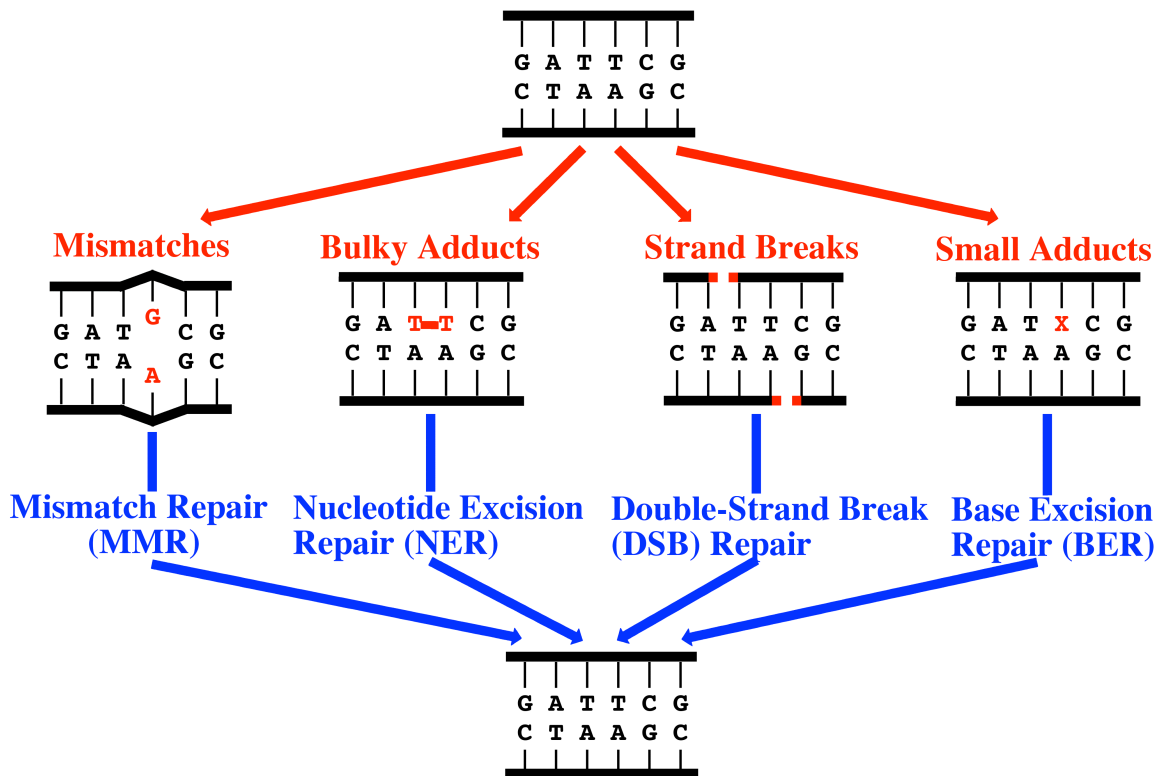
U: Uracil (nucleobase)

# Chapter 1

## Introduction

Genomic DNA is constantly subjected to spontaneous damage inflicted by reactive metabolites as well as environmental mutagens. Due to the abundance of reactive sites within the nucleobases comprising DNA and the large magnitude of damaging agents, various types of DNA damage are possible and each is a potential threat to the integrity of the genome. In order to combat the constant threat of DNA damage, cells have evolved numerous repair pathways, each responsible for the detection and efficient repair of a specific type of DNA damage. Most damage is repaired by one or a combination of four major repair pathways that are highly conserved throughout evolution (Figure 1-1). Double-strand break (DSB) repair, which encompasses homologous recombination (HR) and non-homologous end-joining (NHEJ), restores the continuity of DNA upon formation of single- or double-strand breaks induced by ionizing radiation or reactive oxygen species. Mismatch repair (MMR) repairs mismatches arising from replication and recombination errors as well as damage to nitrogenous bases. Bulky DNA adducts which distort the DNA helix, such as intrastrand crosslinks, are removed by the nucleotide excision repair (NER) pathway. Such adducts often result from exposure to UV radiation and environmental mutagens such as cisplatin. Other exogenous sources, such as oxidizing or alkylating agents, can chemically modify the base moiety of a nucleobase. Often, these modifications are very small and the resultant base lesions do not grossly change the double helical structure of DNA. Such damage is repaired by the base excision repair (BER) pathway (1-3). Together, these pathways continuously search the genome to locate and efficiently repair various types of damage, and failure to do so may lead to cell death, genomic instability, and cancer.

In eukaryotic cells, genomic DNA is packaged into the small confines of the nucleus in a hierarchical fashion. At the lowest level is the nucleosome, the repeating unit



**Figure 1-1: The four major DNA repair pathways**

Native DNA (black) is shown in cartoon form with letters indicating the form normal bases. DNA damage is depicted in red and the repair pathways are depicted in blue. Mismatch repair (MMR) corrects DNA mismatches resulting from replication and recombination errors. Nucleotide excision repair (NER) corrects bulky adducts which distort the DNA helix, such as intrastrand crosslinks. Double-strand break repair, carried out by homologous recombination (HR) or non-homologous end-joining (NHEJ), repairs breaks affecting both strands of DNA. Base excision repair (BER) corrects base lesions (designated as a red X) resulting from relatively small modifications to the base moiety of a nucleobase. Unlike that for NER, the base lesions corrected by BER do not significantly distort the double-helical structure of DNA significantly, if at all.

of chromatin consisting of 147 bp of DNA wrapped around a histone octamer.

Nucleosomes are reiterated at short intervals along the entire length of the genome and are separated by short stretches (~10 – 50 bp) of unwrapped linker DNA. Thus, 75 – 95% of the eukaryotic genome is wrapped in nucleosomes (4). Nucleosomes are further compacted into 30-nm fibers that are stabilized by linker histones. These fibers are then further condensed into higher-order chromatin structures. The level of chromatin condensation of the genome as a whole varies from lightly (referred to as euchromatin) to fully condensed (referred to as heterochromatin) with stages of the cell cycle. For instance, DNA is less condensed during S-phase to allow for replication of DNA whereas all DNA is fully condensed during M-phase for mitotic separation of chromosomes.

Outside of M-phase, chromatin condensation varies along the genome with gene activity. Heterochromatin typically occurs at gene-poor, transcriptionally-silent regions of the genome whereas euchromatin typically occurs at gene-rich, transcriptionally-active regions (5). Together, this compact structure of eukaryotic DNA along with its highly-regulated accessibility can pose a challenge to the repair machinery. In order to carry out repair within the context of the cell, these various DNA repair pathways utilize indirect and direct methods for gaining access to the genome and locating sites of DNA damage, as discussed below.

Nucleosomes can be altered by at least three distinct processes. Their composition may be modified by the incorporation of histone variants that have specialized functions. In addition, coordinated post-translational modification on histone residues can promote chromatin relaxation, i.e. euchromatin, increasing accessibility. Finally, nucleosomes can be replaced, repositioned or removed by ATP-dependent chromatin remodeling complexes, which typically contain DNA-dependent ATPases of the Swi2/Snf2 subfamily (2). Thus, an indirect method for DNA repair pathways to locate sites of damage and initiate repair would be to couple lesion site location to chromatin modification. Indeed, mounting experimental evidence suggests that repair of helix-distorting damage by NER as well as DSB repair (specifically HR) may be coupled to hyperacetylation of histones by histone acetyltransferases (HATs) and chromatin remodeling by Swi2/Snf2 (3, 6). Although the exact nature of the signal for chromatin modification upon DNA damage has yet to be described, these studies suggest a network linking changes in chromatin structure to DNA repair. Coupling repair of helix-distorting DNA damage to other DNA-templated activities, such as transcription and replication, has also been documented.

During transcription and replication, one or both strands of a given segment of DNA are thoroughly interrogated by a respective polymerase in order to ensure accurate duplication or transcription of the template strand. Certain types of helix-distorting damage, such as double-strand breaks (DSBs) and interstrand crosslinks (ICLs) can arrest RNA polymerase (RNAP) and DNA polymerase in stable complexes, which may represent a signal for damage. Thus, another indirect method for DNA repair pathways to detect damage would be to couple it to the DNA-templated activities of replication or

transcription. Over 20 years ago, a specialized NER pathway, called transcription-coupled DNA repair (TCR), was discovered which specifically removes transcription-blocking lesions from the transcribed strands of expressed genes. Although much of the details are yet to be revealed, experimental evidence suggests a model whereby an arrested RNAP complex serves as signal to recruit NER factors to the site of damage, initiating repair. Once the template is repaired, the transcription can resume (7). Likewise, DNA replication can also be coupled to the repair of helix-distorting damage. An arrested DNA polymerase complex during S phase of the cell cycle along with the consequent stalled replication fork can trigger a DNA-damage response (DDR). For instance, double-strand breaks encountered during DNA polymerization can elicit a replication checkpoint which, in addition to stabilizing the replication fork and arresting the cell cycle, recruits factors involved in homologous recombination to the double-strand break, initiating repair. Covalent crosslinks between complementary DNA strands (interstrand crosslinks, ICLs) can completely block replication fork progression by preventing strand separation. Such an event can serve as a signal to initiate the Fanconi anaemia pathway and translesion synthesis (TLS) polymerization, which together resolve the interstrand crosslink into appropriate substrates for nucleotide excision repair and homologous recombination (8). In addition to the repair of bulky, helix-distorting damage, DNA replication can also be coupled to the repair of DNA mismatches, as discussed below.

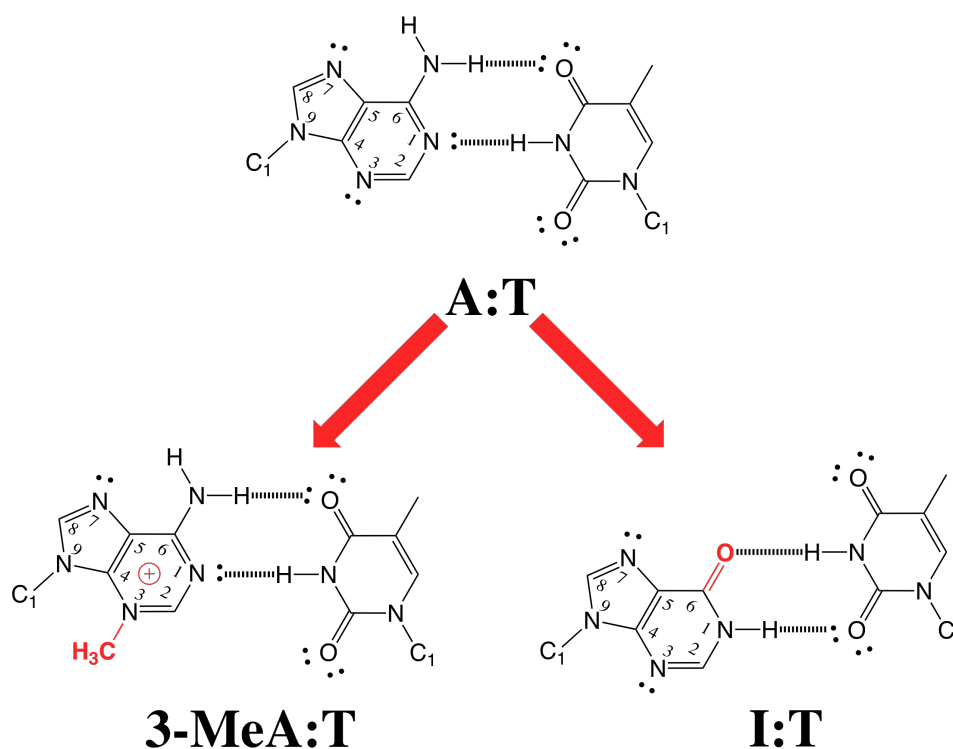
The fidelity of DNA replication within cells is estimated to be around one error per  $10^{10}$  nucleotides synthesized. Replicative polymerases are extremely precise enzymes and have error rates of approximately one erroneously inserted nucleotide per  $10^5$  nucleotides. In the unlikely event that a non-complementary nucleotide is incorporated at the end of a primer, the mispaired primer terminus is translocated into the active site of the proofreading  $3' \rightarrow 5'$  exonuclease activity associated with all replicative polymerases, where the terminal non-complementary nucleotide is excised along with 2-3 additional nucleotides and replication then continues. This proofreading provides a further 2 orders of magnitude to the fidelity of the DNA replication. However, on rare occasions a non-complementary nucleotide eludes the proofreading process and is extended by DNA polymerase, which may lead to a mismatch. Additional errors may also occur when the primer and template strands of DNA transiently dissociate from each

other during replication and reassociate out of register. Such a phenomenon, referred to as “polymerase slipping,” may give rise to insertion/deletion loops (IDLs) containing extrahelical nucleotides. Failure to correct these replication errors may lead to mutation. Thwarting this threat and providing the remaining factor of  $10^3$  to the fidelity of DNA replication is the mismatch repair (MMR) pathway, which recognizes these replication errors and degrades a large segment of the nascent strand (>1 kb) containing the error (9). The replication machinery then fills in the resultant gap, and DNA ligase seals the phosphodiester backbone. However, in order to correct these replication errors, the parent and daughter strands of DNA must be discriminated such that the newly synthesized strand is specifically targeted for repair. Although much work lies ahead in experimentally defining the details, substantial genetic and biochemical evidence suggest that strand-discrimination and strand-specific repair may occur by coupling the mismatch repair pathway to DNA replication. A current theory is that the MMR proteins responsible for recognizing the damage and initiating the pathway remain in the vicinity of the replication fork during replication via interactions with the replication machinery, and strand discontinuities (nicks) in the lagging strand and the replication machinery itself can direct MMR to the leading and lagging nascent strands (1, 10-12).

The model described above pertains to correction of DNA damage resulting from errors made during DNA replication, such as base-base mismatches. By coupling to DNA replication, MMR ensures that erroneously inserted nucleotides which escape proofreading are corrected immediately after they are created, perhaps as the mismatch emerges from the polymerase. However, not all DNA mismatches are generated during DNA replication and, consequently, their repair is not coupled to DNA replication. Within cells, the exocyclic amino groups of adenine (A), cytosine (C), and guanine (G) are subject to hydrolytic attack by oxidizing agents. Because the exocyclic amino group mediates base pairing, its loss leads to mispairing. Deamination of cytosine (C) to uracil (U) is the most common spontaneous deamination event in DNA and forms a U:G mismatch. Upon subsequent replication, U pairs with A, leading to a C:G → T:A transition mutation. These mutations may also result from deamination of 5-



methylcytosine<sup>1</sup> (5-MeC) to thymine (T), forming a T:G mismatch (13, 14). A similar situation may also arise when unnatural nucleobases are formed as a result of modifications to the base moiety of nucleotide by oxidizing and alkylating agents. Often the modifications are so minor that the resultant base lesion can still maintain normal Watson-Crick base pairing or adopt alternative base pairing conformations. Shown in Figure 1-2 are two common modifications to the normal base adenine (A) resulting from alkylation and oxidation. Methyl methanesulfonate (MMS) methylates bases in DNA, particularly at position N-3, away from the base-pairing “face” of the nitrogenous base. Consequently, the resultant lesion, referred to as 3-methyladenine (3-MeA), still



**Figure 1-2: Two common base lesions resulting from modification of adenine**

(Top) Watson-Crick base pair between adenine (A) and thymine (T). (Bottom) Two common base modifications (shown in red) of adenine resulting from exposure to alkylating or oxidizing agents and the consequent base-pair conformation. Exposure to alkylating agents, such as Methyl methanesulfonate (MMS) can methylate adenine at position N-3, away from the base-pairing “face” of the nitrogenous base such that Watson-Crick base pairing is uncompromised. Oxidizing agents, such as dinitrogen trioxide and nitrosamines, can oxidatively deaminate the 6-amino group of adenine, forming inosine (I). Because the exocyclic group mediates base-pairing with thymine (T), Watson-Crick base-pairing is prevented but alternative base-pairing conformations are permitted.

<sup>1</sup> In animals, 70% - 80% of CpG dinucleotides are methylated at the 5-position of cytosine. This modification, catalyzed by a family of conserved DNA methyltransferases (MTases), serves as an epigenetic mark and can inhibit transcription initiation and arrest transcription elongation (13).

maintains normal Watson-Crick base-pairing. Although the 3-methyl group blocks replication by replicative polymerases by projecting into the minor groove, replication may proceed through 3-MeA by error-prone polymerases that may insert the wrong base across from the damage (15). Exposure to nitrosating agents, such as dinitrogen trioxide and nitrosamines, can oxidatively deaminate the 6-amino group of adenine, forming inosine (I). Although this modification prevents Watson-Crick base-pairing, alternative base-pair conformations may be adopted with thymine (T) as well as cytosine (C). Consequently, replication of inosine (I) may lead to insertion of C across from the damaged base, resulting in A:T→G:C transition mutations (16, 17). Collectively, these modifications to the nucleobases are relatively small and the consequent effect on base-pairing (i.e. altered or abolished) causes little distortion in the DNA double helix and does not inhibit DNA-templated processes such as extension by the replication machinery or elongation by RNA polymerase (18). However, as alluded to, failure to correct these small damages may lead to mutation, and, ultimately, cancer.

The base excision repair (BER) pathway (summarized in Figure 1-3) is responsible for the repair of most of the DNA modifications mentioned above, which affect single bases (19). In humans, this pathway is initiated by one of almost a dozen DNA repair glycosylases<sup>2</sup> that continuously and independently search the genome to locate sites of damage (20). Once a damaged base is located, the glycosylase flips the damaged nucleotide out of the DNA duplex and catalyzes the hydrolysis of the *N*-glycosidic bond to release the lesioned base. The resultant abasic site is then processed by AP endonuclease, which cleaves the phosphodiester backbone 5' to the abasic site, leaving a 3' OH and a 5' deoxyribosephosphate (DRP) group<sup>3</sup>. The DRP group is then released by the dRP lyase activity of DNA polymerase β, which then uses the intact template strand to insert the correct nucleotide back into the DNA. Finally, DNA ligase completes restoration of the native DNA sequence by sealing the nick.

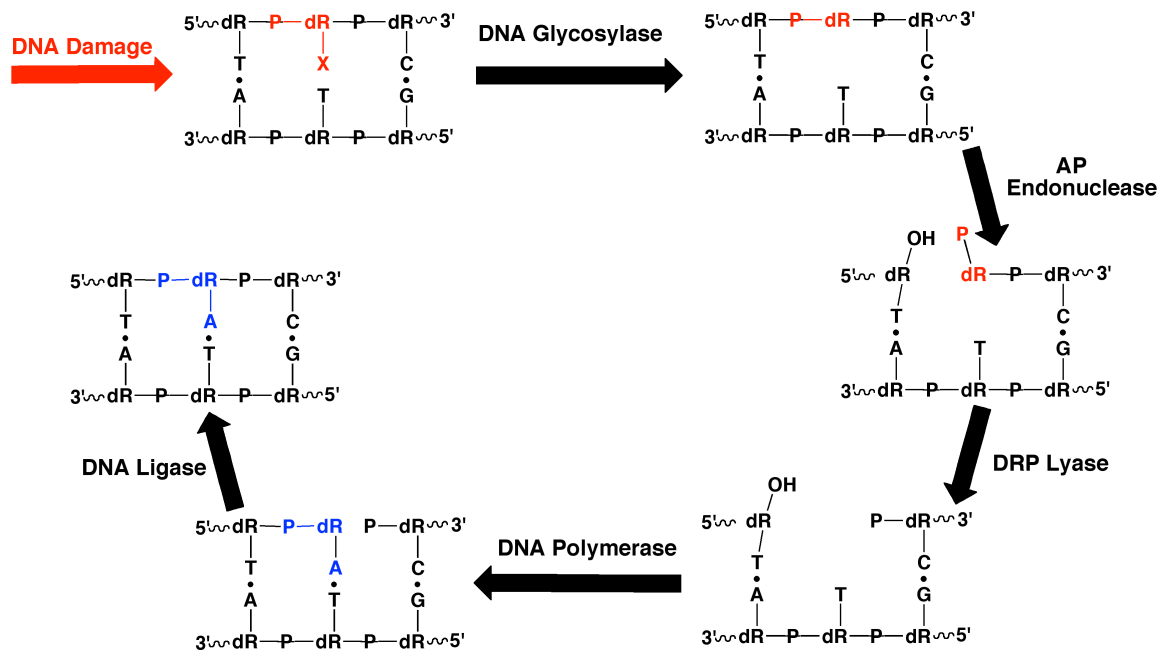
It is estimated that the BER pathway repairs ~10,000 base lesions in a typical human cell everyday (21, 22). DNA glycosylases have the daunting task of locating

---

<sup>2</sup> For an updated, comprehensive table of known human DNA glycosylases along with their DNA substrates, refer to Svilar, D. et al. *Antioxid Redox Signal*, Epub ahead of print, July 22, 2010.

<sup>3</sup> The pathway outlined in Figure 1-3 is for monofunctional glycosylases. The pathway differs for bifunctional glycosylases that also catalyze β-elimination in addition to *N*-glycosidic bond hydrolysis.

these relatively rare base lesions from among ~12,000,000,000 undamaged nucleotides in the human diploid genome. As previously discussed, the base lesions corrected by this pathway are relatively small and do not significantly distort the double helical structure of DNA nor inhibit DNA-templated processes. Thus, uncoupled from a signal for DNA damage or DNA-templated activities that expose genomic DNA, DNA glycosylases must instead scan the genome continuously throughout the cell cycle to locate sites of damage. Such a direct search for target sites is limited by accessibility to chromatic DNA, which is controlled by the dynamic properties intrinsic to nucleosomes (spontaneous unwrapping and re-wrapping), active mechanisms modulating DNA-histone interactions (histone modification, chromatin remodeling, etc.), and DNA processing events (i.e., replication and transcription) (18).



**Figure 1-3: The base excision repair pathway initiated by a monofunctional glycosylase**  
 DNA damage chemically modifies a single nucleobase, forming a base lesion (designated in red). A monofunctional DNA glycosylase initiates the BER pathway by releasing the lesioned base from the deoxyribose sugar via *N*-glycosidic bond hydrolysis. AP endonuclease recognizes the resultant abasic site and cleaves the phosphodiester backbone 5' to the abasic site, leaving a free 3' hydroxyl and a 5' DRP group. The lyase domain of DNA polymerase  $\beta$  releases the DRP group and then the polymerase domain inserts the correct nucleotide (designated in blue) back into the DNA using the intact strand as a template. The pathway is completed upon sealing of the nick by DNA ligase.

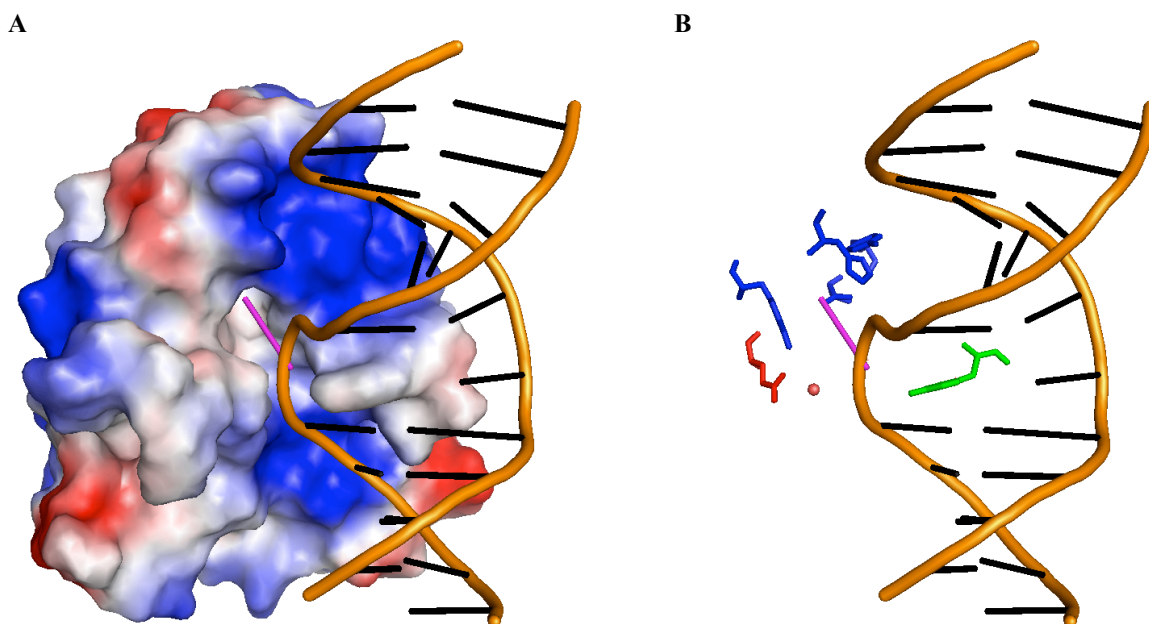
In addition to DNA repair, a direct search for target sites similar to that described above for DNA glycosylases is also central to other nuclear processes, such as DNA replication and transcription. A large body of work on restriction endonucleases and

transcription factors has demonstrated that these proteins utilize a correlated search for target sites whereby each binding encounter with a given segment of exposed DNA involves a search of multiple adjacent sites (23-28). Such a search is mediated by nonspecific binding interactions that allow linear (one-dimensional) diffusion along the DNA and has been historically referred to as a processive search because it affords a given protein the ability to catalyze multiple enzymatic events per binding encounter. Shortly after the discovery of the first DNA glycosylase in 1974 (uracil DNA glycosylase from *E. coli*) efforts began to focus on elucidating the manner in which these repair enzymes locate sites of damage and a large body of *in vitro* studies on several DNA glycosylases from various organisms have found evidence of processive action at adjacent sites on DNA (29-39). However, a common criticism of many of these *in vitro* findings is the biological relevance of the experimental conditions. For instance, several DNA glycosylases were found to be processive only at ionic strengths far below physiological levels ( $\ll 150$  mM) (31, 34, 40-43). Furthermore, *in vitro* studies of linear diffusion of proteins are often carried out on naked DNA, yet DNA is expected to be only transiently accessible within the cell. Other DNA binding proteins bind densely to chromosomal DNA and those with high affinity can have a lifetime of hours (20). In order to ensure a complete and efficient search of the genome, DNA glycosylases must be able to overcome such obstacles. Thus, despite its importance, there is still much to learn about the mechanisms DNA glycosylases utilize in order to perform genome-wide searches for rare target sites. In order to gain insight into this, I focused my studies on understanding the mechanism by which human alkyladenine DNA glycosylase searches for sites of damage.

Human alkyladenine DNA glycosylase (AAG) is a 33 kDa, monofunctional glycosylase responsible for initiating repair of a diverse group of alkylated and deaminated purine nucleotides including 3-methyladenine, 7-methylguanine, 1,  $N^6$ -ethenoadenine, inosine, and oxanine (16, 44). AAG is comprised of a poorly conserved N-terminus and a carboxy-terminal glycosylase domain that is highly conserved. It is expressed and functions as a monomer and at least two splice variants are present that differ slightly at their amino termini. The larger splice variant is 298 amino acids in length (45). The N-terminus of AAG (N-terminal 79 amino acids) is not required for *N*-

glycosidic bond hydrolysis and is suggested to be solvent exposed and flexible based on its sensitivity to proteolytic degradation and its required truncation for obtaining the crystal structure shown in Figure 1-4.

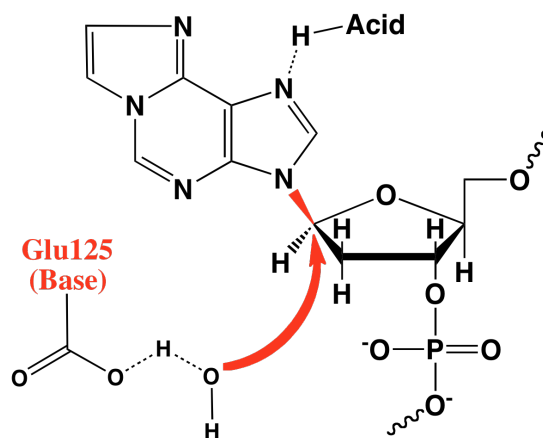
In its active complex, AAG interacts with ~12 nucleotides surrounding the damage base in a positively charged groove (Figure 1-4A) and accesses the *N*-glycosidic bond by stabilizing a kinked conformation whereby the DNA is bent 22° and the damaged base is flipped completely out of the duplex. To stabilize the void left by the flipped-out base, AAG inserts a residue an aromatic residue (Y162) into the duplex. This general mechanism, referred to as base flipping, is common to all DNA glycosylases and



**Figure 1-4: Crystal structure of the AAG catalytic domain**

The crystal structure of the AAG•εA-DNA complex was used to generate these pictures [PDB entry 1F4R(46)]. The DNA structure is in cartoon form with the phosphodiester backbone in orange, undamaged (nonspecific) nucleobases in black, and damaged nucleobase (1,*N*<sup>6</sup>-ethenoadenine) in magenta (A) Electrostatics calculations were performed on the protein alone with Pymol. A continuum from negatively charged (red) to positively charged (blue) is shown. (B) Stripped down structure of AAG highlighting key residues. The intercalating residue (Y162) is shown in green, residues comprising active site pocket (Y127, H136, Y159) in blue, and the active site general base (E125) and water molecule in red.

other enzymes that need access to a DNA base to perform chemistry on it, such as methyltransferases (46-48). Once flipped out, AAG holds the damaged base in an accommodating active site, aligning the *N*-glycosidic bond (shown in red in Scheme 1) for attack. Biochemical and genetic evidence along with that observed in the crystal structure shown in Figure 1-4 suggests that a glutamate residue (E125) activates an active site water molecule for nucleophilic attack at C1' of the ribose sugar, hydrolyzing the



Scheme 1

*N*-glycosidic bond and releasing the damaged base (Scheme 1). To ensure that AAG excises only damaged purine bases, protonation of the nucleobase leaving group inhibits excision of pyrimidines while unfavorable steric clashes with the exocyclic amino groups of adenine and guanine inhibits excision of normal purines (46, 49).

In the studies described herein, we focus on understanding the mechanism by which AAG searches for sites of DNA damage. Numerous structural and biochemical characterizations of AAG discussed above have provided great insight into the manner in which AAG interacts with and hydrolyzes distinct base lesions, particularly the lesion shown in the crystal structure, 1, *N*<sup>6</sup>-ethenoadenine ( $\epsilon$ A). However, such experiments report only on steps occurring subsequent to base flipping, i.e., after a lesion has already been located by AAG, and remarks on the manner in which AAG locates sites of damage have been purely speculative. Nonetheless, such characterization allowed us to rationally design an assay to characterize the ability of AAG to diffuse along DNA. In chapter 2, our studies focus on the repair of  $\epsilon$ A lesions and we describe a simple *in vitro* assay to monitor processive action of AAG on an oligonucleotide substrate containing two sites of damage. We show that AAG locates base lesions by thermally-driven, linear diffusion whereby the enzyme diffuses along nonspecific DNA in a non-directional manner, sampling multiple adjacent sites per binding encounter. Further analysis revealed that the amino terminus of AAG contributes to the processive behavior of AAG by increasing nonspecific DNA binding.

Linear diffusion along nonspecific DNA is common to many proteins that must perform genome-wide searches for specific target sites and two distinct mechanisms are

recognized for such movement. Sliding involves continuous contact between the protein and the DNA backbone so that transfer occurs between linearly contiguous sites on the same DNA strand. In contrast, the alternative mode of translocation, referred to as hopping, a *macroscopically* bound protein diffuses along DNA via *microscopic* dissociation-reassociation events (23). In chapter 3 we investigated the mechanism by which AAG diffuses along DNA. Our results indicate that AAG makes significant excursions from the surface of DNA, allowing it to reorient between strands, effectively searching both strands of DNA simultaneously within a single binding encounter. Such hopping events allow AAG to bypass a tightly-bound protein. This could be important for a search for DNA damage *in vivo*.

In chapter 4, we probe further into the structural requirements of DNA that permit linear diffusion by AAG. By testing oligonucleotide substrates that contained various structural perturbations to the B-form, double helix, we show that hopping between the duplex segments on either side of a structural barrier allow AAG to bypass various structures such as kinks, bubbles, and gaps. This provides further support that hopping is an important mode of translocation in the search for DNA damage in the context of the genome.

Chapters 2 – 4 focused on the manner in which AAG locates sites of damage and provided substantial evidence that AAG diffuses along nonspecific DNA by a mechanism comprised of frequent hopping events that allow the simultaneous search of both strands of DNA and bypass of structural perturbations. In chapter 5, we explore how efficient this search for damage is. By comparing the behavior of AAG on oligonucleotides that contained either hypoxanthine or 1,  $N^6$ -ethenoadenine, our results show that the efficiency of the search for DNA damage depends on the identity of the base lesion. AAG is very efficient at recognizing at 1,  $N^6$ -ethenoadenine and excises it upon the initial binding encounter. Thus, in the search for DNA damage, AAG only requires one encounter with 1,  $N^6$ -ethenoadenine in order to excise it. In contrast, our results suggest that hypoxanthine is inefficiently recognized during the initial binding encounter with AAG, even though it is an excellent substrate (16). Thus, in order to excise a hypoxanthine lesion from a given segment of DNA, AAG must encounter the lesion multiple times. As discussed above, the search for DNA damage is thermally-driven and

completely random. Thus, a highly redundant search permitting multiple encounters with a given base pair may ensure that each lesion is repaired.

Collectively, these studies provide a foundation for understanding the molecular mechanism by which a DNA repair enzyme efficiently searches the genome for rare sites of damage. Paramount to this search is the ability of such proteins to bind specific target sites with high affinity but also retain nonspecific binding affinity that allows diffusion along nonspecific DNA. Future work will be needed to explore the structural features of these enzymes that allow for both specific and nonspecific DNA binding.



## References

1. Yang, W. (2008) Structure and mechanism for DNA lesion recognition, *Cell Res* 18, 184-197.
2. Huertas, D., Sendra, R., and Munoz, P. (2009) Chromatin dynamics coupled to DNA repair, *Epigenetics* 4, 31-42.
3. Osley, M. A., Tsukuda, T., and Nickoloff, J. A. (2007) ATP-dependent chromatin remodeling factors and DNA damage repair, *Mutat Res* 618, 65-80.
4. Garcia, H. G., Grayson, P., Han, L., Inamdar, M., Kondev, J., Nelson, P. C., Phillips, R., Widom, J., and Wiggins, P. A. (2006) Biological Consequences of Tightly Bent DNA: The Other Life of a Macromolecular Assembly, *Biopolymers* 85.
5. Tamaru, H. (2010) Confining euchromatin/heterochromatin territory: jumonji crosses the line, *Genes Dev* 24, 1465-1478.
6. Misteli, T., and Soutoglou, E. (2009) The emerging role of nuclear architecture in DNA repair and genome maintenance, *Nat Rev Mol Cell Biol* 10, 243-254.
7. Tornaletti, S. (2009) DNA repair in mammalian cells: Transcription-coupled DNA repair: directing your effort where it's most needed, *Cell Mol Life Sci* 66, 1010-1020.
8. Branzei, D., and Foiani, M. (2010) Maintaining genome stability at the replication fork, *Nat Rev Mol Cell Biol* 11, 208-219.
9. Modrich, P. (2006) Mechanisms in eukaryotic mismatch repair, *J Biol Chem* 281, 30305-30309.
10. Jiricny, J. (1998) Replication errors: cha(lle)nging the genome, *EMBO J* 17, 6427-6436.
11. Kunkel, T. A., and Erie, D. A. (2005) DNA mismatch repair, *Annu Rev Biochem* 74, 681-710.
12. Scharer, O. D. (2003) Chemistry and biology of DNA repair, *Angew Chem Int Ed Engl* 42, 2946-2974.
13. Colot, V., and Rossignol, J. L. (1999) Eukaryotic DNA methylation as an evolutionary device, *Bioessays* 21, 402-411.

14. Lindahl, T., and Nyberg, B. (1974) Heat-induced deamination of cytosine residues in deoxyribonucleic acid, *Biochemistry* 13, 3405-3410.
15. Johnson, R. E., Yu, S. L., Prakash, S., and Prakash, L. (2007) A role for yeast and human translesion synthesis DNA polymerases in promoting replication through 3-methyl adenine, *Mol Cell Biol* 27, 7198-7205.
16. O'Brien, P. J., and Ellenberger, T. (2004) Dissecting the broad substrate specificity of human 3-methyladenine-DNA glycosylase, *J Biol Chem* 279, 9750-9757.
17. Williams, D. L. H. (1988) *Nitrosation*, Cambridge University Press, Cambridge ; New York.
18. Jagannathan, I., Cole, H. A., and Hayes, J. J. (2006) Base excision repair in nucleosome substrates, *Chromosome Res* 14, 27-37.
19. Robertson, A. B., Klungland, A., Rognes, T., and Leiros, I. (2009) DNA repair in mammalian cells: Base excision repair: the long and short of it, *Cell Mol Life Sci* 66, 981-993.
20. Hedglin, M., and O'Brien, P. J. (2010) Hopping enables a DNA repair glycosylase to search both strands and bypass a bound protein, *ACS Chem Biol* 5, 427-436.
21. Lindahl, T., and Wood, R. D. (1999) Quality control by DNA repair, *Science* 286, 1897-1905.
22. Lindahl, T. (1993) Instability and decay of the primary structure of DNA, *Nature* 362, 709-715.
23. Berg, O. G., Winter, R. B., and von Hippel, P. H. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory, *Biochemistry* 20, 6929-6948.
24. Winter, R. B., Berg, O. G., and von Hippel, P. H. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The Escherichia coli lac repressor--operator interaction: kinetic measurements and conclusions, *Biochemistry* 20, 6961-6977.
25. Winter, R. B., and von Hippel, P. H. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 2. The Escherichia coli repressor--operator interaction: equilibrium measurements, *Biochemistry* 20, 6948-6960.

26. Terry, B. J., Jack, W. E., and Modrich, P. (1985) Facilitated diffusion during catalysis by EcoRI endonuclease. Nonspecific interactions in EcoRI catalysis, *J Biol Chem* 260, 13130-13137.
27. Halford, S. E., and Marko, J. F. (2004) How do site-specific DNA-binding proteins find their targets?, *Nucleic Acids Res* 32, 3040-3052.
28. Stanford, N. P., Szczelkun, M. D., Marko, J. F., and Halford, S. E. (2000) One- and three-dimensional pathways for proteins to reach specific DNA sites, *EMBO J* 19, 6546-6557.
29. Sidorenko, V. S., Mechetin, G. V., Nevinsky, G. A., and Zharkov, D. O. (2008) Correlated cleavage of single- and double-stranded substrates by uracil-DNA glycosylase, *FEBS Lett* 582, 410-414.
30. Leblanc, J. P., Martin, B., Cadet, J., and Laval, J. (1982) Uracil-DNA glycosylase. Purification and properties of uracil-DNA glycosylase from *Micrococcus luteus*, *J Biol Chem* 257, 3477-3483.
31. Francis, A. W., and David, S. S. (2003) *Escherichia coli* MutY and Fpg utilize a processive mechanism for target location, *Biochemistry* 42, 801-810.
32. Lindahl, T. (1974) An N-glycosidase from *Escherichia coli* that releases free uracil from DNA containing deaminated cytosine residues, *Proc Natl Acad Sci U S A* 71, 3649-3653.
33. Sidorenko, V. S., and Zharkov, D. O. (2008) Correlated cleavage of damaged DNA by bacterial and human 8-oxoguanine-DNA glycosylases, *Biochemistry* 47, 8970-8976.
34. Bennett, S. E., Sanderson, R. J., and Mosbaugh, D. W. (1995) Processivity of *Escherichia coli* and rat liver mitochondrial uracil-DNA glycosylase is affected by NaCl concentration, *Biochemistry* 34, 6109-6119.
35. Purmal, A. A., Lampman, G. W., Pourmal, E. I., Melamede, R. J., Wallace, S. S., and Kow, Y. W. (1994) Uracil DNA N-glycosylase distributively interacts with duplex polynucleotides containing repeating units of either TGGCCAAGCU or TGGCCAAGCTTGGCCAAGCU, *J Biol Chem* 269, 22046-22053.
36. Colson, P., and Verly, W. G. (1983) Intracellular localization of rat-liver uracil-DNA glycosylase. Purification and properties of the chromatin enzyme, *Eur J Biochem* 134, 415-420.
37. Porecha, R. H., and Stivers, J. T. (2008) Uracil DNA glycosylase uses DNA hopping and short-range sliding to trap extrahelical uracils, *Proc Natl Acad Sci U S A* 105, 10791-10796.

38. Blainey, P. C., Luo, G., Kou, S. C., Mangel, W. F., Verdine, G. L., Bagchi, B., and Xie, X. S. (2009) Nonspecifically bound proteins spin while diffusing along DNA, *Nat Struct Mol Biol* 16, 1224-1229.
39. Blainey, P. C., van Oijen, A. M., Banerjee, A., Verdine, G. L., and Xie, X. S. (2006) A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA, *Proc Natl Acad Sci U S A* 103, 5752-5757.
40. Higley, M., and Lloyd, R. S. (1993) Processivity of uracil DNA glycosylase, *Mutat Res* 294, 109-116.
41. Gruskin, E. A., and Lloyd, R. S. (1986) The DNA scanning mechanism of T4 endonuclease V. Effect of NaCl concentration on processive nicking activity, *J Biol Chem* 261, 9607-9613.
42. Hamilton, R. W., and Lloyd, R. S. (1989) Modulation of the DNA scanning activity of the *Micrococcus luteus* UV endonuclease, *J Biol Chem* 264, 17422-17427.
43. Nickell, C., and Lloyd, R. S. (1991) Mutations in endonuclease V that affect both protein-protein association and target site location, *Biochemistry* 30, 8638-8648.
44. Hitchcock, T. M., Dong, L., Connor, E. E., Meira, L. B., Samson, L. D., Wyatt, M. D., and Cao, W. (2004) Oxanine DNA glycosylase activity from Mammalian alkyladenine glycosylase, *J Biol Chem* 279, 38177-38183.
45. Hedglin, M., and O'Brien, P. J. (2008) Human alkyladenine DNA glycosylase employs a processive search for DNA damage, *Biochemistry* 47, 11434-11445.
46. Lau, A. Y., Wyatt, M. D., Glassner, B. J., Samson, L. D., and Ellenberger, T. (2000) Molecular basis for discriminating between normal and damaged bases by the human alkyladenine glycosylase, AAG, *Proc Natl Acad Sci U S A* 97, 13573-13578.
47. Lau, A. Y., Scharer, O. D., Samson, L., Verdine, G. L., and Ellenberger, T. (1998) Crystal structure of a human alkylbase-DNA repair enzyme complexed to DNA: mechanisms for nucleotide flipping and base excision, *Cell* 95, 249-258.
48. Roberts, R. J., and Cheng, X. (1998) Base flipping, *Annu Rev Biochem* 67, 181-198.
49. O'Brien, P. J., and Ellenberger, T. (2003) Human alkyladenine DNA glycosylase uses acid-base catalysis for selective excision of damaged purines, *Biochemistry* 42, 12418-12429.

## Chapter 2

### Human Alkyladenine DNA Glycosylase Employs a Processive Search for DNA Damage<sup>1</sup>

Although DNA is remarkably stable, it is nevertheless susceptible to spontaneous damage via reactions with cellular metabolites and environmental mutagens. Chemical reactions that alter the structure of the nucleobases within DNA are most commonly recognized and repaired by the base excision repair (BER)<sup>1</sup> pathway. Some base lesions can block DNA replication and transcription with cytotoxic effects, and many more alter the base pairing properties so that replication leads to mispairing and mutation. The base excision repair pathway is initiated by a DNA repair glycosylase that must locate the site of damage within the genome. Once a damaged base is located, the glycosylase flips out the damaged nucleotide and catalyzes the hydrolysis of the *N*-glycosidic bond to release the lesioned base. The subsequent actions of an endonuclease, abasic site lyase, DNA polymerase and DNA ligase are required to complete the repair pathway.

It is estimated that  $\sim 10^4$  base lesions are formed in a typical human cell every day and that the vast majority of these are correctly repaired by BER or other DNA repair pathways (4). On the one hand, this is a large number of potential mutagenic events that must be corrected. On the other hand, these lesions are very rare considering the size of the human diploid genome ( $\sim 10^{10}$  nucleotides), with only one out of every million nucleotides sustaining damage on any given day. To underscore the magnitude of the damage recognition problem, this level of DNA damage requires a search of  $\sim 100,000$  nucleotides each day per enzyme molecule for an abundant protein of  $\sim 10^5$  copies per

---

<sup>1</sup>Reproduced with permission from Hedglin, M., and O'Brien, P.J. (2008) Human Alkyladenine DNA Glycosylase Employs A Processive Search for DNA Damage, *Biochemistry* 47, 11434 – 11445. Copyright 2008 American Chemical Society

cell, and less abundant proteins would need to search a larger number of nucleotides<sup>2</sup>. Despite its importance, there is still much to learn about the initial recognition of DNA damage and about the mechanisms that ensure a complete and continuous search of the genome.

Human alkyladenine DNA glycosylase (AAG) is a 33 kDa monomeric protein that initiates repair of a diverse group of alkylated and deaminated purine nucleotides. These lesions include 3-methyladenine, 7-methylguanine, and 1,*N*<sup>6</sup>-ethenoadenine ( $\epsilon$ A), as well as the deaminated purines hypoxanthine and oxanine [refs 5 and 6 and refs cited therein]. A consequence of this broad specificity is that AAG removes normal bases at a low level (5, 7, 8). Substrate selection appears to be governed by a combination of selectivity filters. The first selectivity filter occurs at the nucleotide flipping step, since AAG preferentially selects lesions that are presented in unstable base pairs. The catalytic mechanism constitutes a second selectivity filter. Once bound, the use of general acid catalysis ensures that AAG only excises purine bases, even though smaller pyrimidines can fit into the active site (9, 10). The third selectivity filter consists of unfavorable steric clashes with the exocyclic amino groups of guanine and adenine, so that purine lesions lacking these functional groups are preferentially recognized. Finally, since alkylation of N3 and N7 of the purine ring leads to destabilization of the *N*-glycosidic bond, AAG is able to effectively excise *N*-alkyl lesions with a relatively modest rate enhancement.

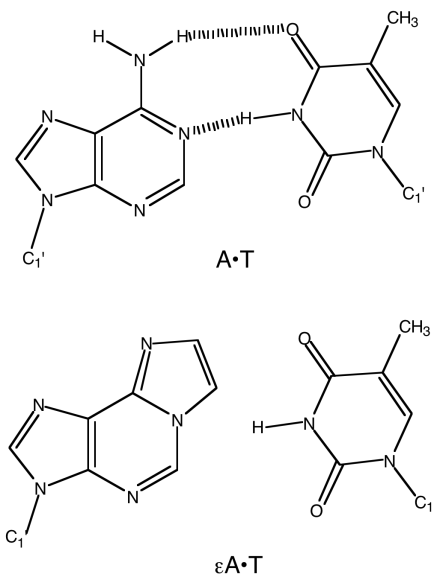
In this study, we focus on understanding the mechanism by which AAG searches for sites of DNA damage. It is widely accepted that genome-wide searches for specific sites will be most efficient if a correlated search is used whereby each binding encounter with the DNA involves the search of multiple adjacent sites (11-14). This can be accomplished by diffusion along the DNA, which can be mediated by nonspecific binding interactions. A large body of work on restriction endonucleases and transcription factors has demonstrated the ability of these proteins to slide along DNA in search of their recognition sites [e.g., (15-17)]. Studies of several BER enzymes have also found evidence for processive action at adjacent sites on DNA, albeit with decreased

---

<sup>2</sup>The abundance of AAG in human fibroblasts was determined to be  $2 \times 10^5$  molecules per cell from glycosylase activity in cell extracts (1). This is similar to the abundance of other base excision repair enzymes, which have been reported to be between  $0.5-3 \times 10^5$  molecules per cell (2, 3).

processivity relative to the restriction endonucleases (18-24). A common criticism of many of these *in vitro* findings is that processive action occurs primarily at salt concentrations that are below physiological levels. Nevertheless, these studies provide compelling evidence that these proteins are capable of diffusion along DNA. In several cases, mutants have been used to directly correlate reductions in processivity with reduced biological function (25-27). This correlation between *in vivo* function and the ability to conduct a correlated search *in vitro* suggests that diffusion along DNA is important for lesion recognition *in vivo*.

Our studies have focused on the repair of  $\epsilon$ A lesions. This lesion is thought to be the result of lipid peroxidation and it is found at low levels in human cells under normal growth conditions (28, 29). Since  $\epsilon$ A cannot hydrogen bond with any of the normal bases (Scheme 1), it is expected to present a relatively low barrier to nucleotide flipping. Indeed, it appears to bind more tightly than other substrates (5, 30). The crystal structure of AAG bound to an extrahelical  $\epsilon$ A lesion shows that it is readily accommodated in the active site pocket and the backbone amide of His136 donates a hydrogen bond to the N6 atom (31).



We have developed a simple *in vitro* processivity assay and used it to characterize the ability of AAG to diffuse along DNA. We find that AAG is able to search at least 25 base pairs of DNA prior to dissociation at physiological ionic strength and pH. The ability of AAG to diffuse along DNA is eliminated at higher ionic strength, consistent

with the importance of electrostatic interactions for DNA scanning. Deletion of the poorly conserved amino terminus of AAG results in decreased processivity, presumably by increasing the rate of dissociation from DNA. These observations demonstrate that AAG is capable of performing a correlated search of DNA *in vitro* and that this ability is expected to increase the efficiency of DNA damage recognition *in vivo*.

## Materials and Methods

### Proteins

*E. coli* formamidopyrimidine DNA glycosylase (FPG) was obtained from New England Biolabs. Full-length and truncated recombinant human AAG was expressed in *E. coli* and purified as previously described (5, 9). We refer to the amino-terminally truncated protein as  $\Delta 80$ , but residues K<sup>80</sup>G<sup>81</sup>H<sup>82</sup>L<sup>83</sup> have been replaced by the residues G<sup>80</sup>P<sup>81</sup>H<sup>82</sup>M<sup>83</sup> that remain after proteolytic cleavage by human rhinovirus 3C protease. The concentrations of AAG proteins were determined from the absorbance at 280 nm and the calculated extinction coefficients. Under conditions of low ionic strength, the excision of hypoxanthine shows burst kinetics (see Appendix A), with a rapid initial turnover followed by a slower steady state rate. We used this burst amplitude to calculate the concentration of active protein. The results from the burst analyses were in excellent agreement with the calculated concentration of AAG, indicating that greater than 90% of the recombinant proteins are active (see Appendix A).

### Synthesis and purification of Oligonucleotides

DNA substrates were synthesized by Integrated DNA Technologies or the Keck Center at Yale University. The  $\epsilon$ A-containing oligonucleotides were synthesized using ultra mild protecting groups and all other oligonucleotides were synthesized with standard protecting groups and deprotected according to the manufacturer's recommendations (Glen Research). After desalting using Sephadex G-25, oligonucleotides were purified on denaturing polyacrylamide gels. DNA was extracted and desalted by binding to a C18 reverse phase column (Sep-pak, Waters) and eluting with 30% (v/v) acetonitrile. Concentrations were determined from the absorbance at 260 nm using the calculated extinction coefficients. For  $\epsilon$ A-containing strands we calculated the extinction coefficient for the identical sequence containing A in place of  $\epsilon$ A, and then subtracted a value of 9,400 M<sup>-1</sup>cm<sup>-1</sup> per  $\epsilon$ A residue to correct for the weaker absorbance of  $\epsilon$ A relative to A.



For 5', 3', or dual fluorescein-labeled oligonucleotides we assessed the labeling efficiency by comparing the absorbance at 260 nm with that at 495 nm and the calculated labeling efficiency was greater than 85% in all cases. For routine burst analysis to measure glycosylase activity and fraction of active enzyme, the deoxyinosine (I)-containing oligonucleotide 5'-(6-fam)-CGATAGCATCCTICCTTCTCTCCAT oligonucleotide was annealed to the complementary 5'-ATGGAGAGAAGGTAGGATGCTATCG with a 2-fold excess of the unlabeled strand. The sequence for the doubly-labeled  $\epsilon$ A-containing oligonucleotide duplex is given in Figure 2-1.

### **Glycosylase activity assay**

Reactions were carried out at 37 °C. Unless otherwise indicated, the buffer consisted of 50 mM NaMES pH 6.1, 10% (v/v) glycerol, 0.1 mg/mL BSA, and sufficient NaCl to obtain the desired ionic strength. At pH 7.5, the NaMES was replaced with 50 mM NaHEPES. The enzyme concentration ranged from 20 nM to 10  $\mu$ M. Control experiments demonstrated that both full-length and  $\Delta$ 80 AAG retained greater than 60% of their activity when incubated under these conditions without DNA for up to 15 hours, and greater than 90% of their activity when incubated with DNA substrate under turnover conditions (see Appendix A). The reactions were initiated by adding a small volume of enzyme to a reaction volume of 20-60  $\mu$ L that contained between 20 nM and 5  $\mu$ M fluorescein-labeled DNA. Aliquots were withdrawn at various times and quenched with 2 volumes of 0.3 M sodium hydroxide to obtain 0.2 M final concentration. The quenched samples were stored at 4 °C until further processed to prevent base-catalyzed ring-opening and subsequent depurination of the  $\epsilon$ A lesions (32). Samples were heated at 70 °C for 15 minutes to quantitatively cleave abasic sites. Control reactions demonstrated that negligible  $\epsilon$ A sites were cleaved by this treatment ( $\leq$ 1 %, data not shown). After hydroxide-catalyzed cleavage of abasic sites, the samples were diluted with 3 volumes of formamide, containing 10 mM EDTA, and resolved on 20% (w/v) polyacrylamide sequencing gels containing 6.6 M Urea. Gels were scanned with a Typhoon fluorescence imager (GE Healthcare) to detect fluorescein (excitation at 488 nm and emission with a 520BP40 filter). The resulting fluorescence signal was quantified using ImageQuant and corrected for the amount of background signal. A standard curve was constructed by loading different amounts of fluorescently labeled oligo and this established the linearity

of the assay up to 2 pmol of DNA per band (data not shown). The intensity of each DNA band was converted into a fraction by dividing its intensity by the sum of the intensities for all of the DNA species present.

### **Single turnover kinetics**

The 47-mer processivity substrate (Figure 2-1) was incubated with excess enzyme to ensure single turnover conditions (this required >2-fold molar excess of enzyme). The reaction progress curve was plotted as the fraction of product versus time and was fit by a single exponential [ $F = e^{-(k_{\text{obs}} \times t)}$ ], in which  $F$  is the fraction of reaction,  $k_{\text{obs}}$  is the observed rate constant, and  $t$  is the time. In all cases the nonlinear least squares fit was excellent ( $R^2 > 0.95$ ). At saturating concentration of enzyme the observed single turnover rate constant reaches a maximum that we will refer to as  $k_{\text{max}}$  [(33); we have previously referred to this value as  $k_{\text{st}}$ , the single turnover rate constant (5, 9)]. The concentration of enzyme was varied by at least 3-fold to establish that the observed rate was independent of the concentration of enzyme, indicating that enzyme was in excess and at a saturating concentration (i.e.,  $k_{\text{obs}} = k_{\text{max}}$ ).

### **Multiple turnover kinetics**

Steady state kinetics for the dual lesion substrate were measured with a range of 10-1000-fold excess of substrate over enzyme. Unless otherwise indicated, the concentrations used were 2  $\mu\text{M}$  DNA and 20 nM AAG (100-fold excess of substrate). Although reaction rates were linear to greater than 50% consumption of substrate, we used only the first 15-20% of the reactions to calculate initial rates. The initial rates for the disappearance of the substrate were calculated as a fractional change per unit of time. Values of  $k_{\text{cat}}$  were calculated by multiplying this rate by the concentration of DNA and dividing by the concentration of enzyme ( $V_{\text{init}} (\text{nM}/\text{min}) = V_{\text{init}}(\text{fract}/\text{min}) \times [\text{DNA}] (\text{nM}) = k_{\text{cat}}[\text{E}]$ ). At pH 6.1 and low ionic strength, the rate-limiting step for multiple turnover is dissociation of the abasic product. We have fit the ionic strength dependence of  $k_{\text{cat}}$  with a cooperative model [ $k_{\text{cat}} = k_{\text{max}}(I^n / ((K_a)^n + I^n))$ ], in which  $I$  is the ionic strength,  $n$  is the number of cation binding sites,  $K_a$  is the average affinity constant for binding of the cation, and  $k_{\text{max}}$  is the maximal single turnover rate constant.

### Determination of the fraction processive ( $F_p$ )

The ratio of processive events in which both damaged bases are excised in a single binding encounter divided by the total number of enzymatic events is defined as the fraction processive. Since only the ends of the DNA are labeled (Figure 2-1), an expression must be derived that depends only on the initial rates of formation of products and intermediates (15, 34). Accordingly,  $F_p = (V_A + V_C - V_{AB} - V_{BC}) / (V_A + V_B + V_{AB} + V_{BC})$ ; in which  $V_A$  and  $V_B$  are the initial velocities for the formation of the two products and  $V_{AB}$  and  $V_{BC}$  are the initial velocities for the formation of the two intermediates (Figure 2-1; see Appendix A for the derivation). Averages and standard deviations are reported for 3-10 independent determinations. The fraction processive decreases as the rate of dissociation increases. We have fit this to a cooperative model in which cations can bind to multiple sites on the DNA and thereby affect the rate of AAG dissociation. This model is analogous to the Hill equation that takes the following form.  $F_p = F_{p,max} - \Delta F_p I^n / (K_a^n + I^n)$ ; in which  $F_p$  is the fraction processive,  $F_{p,max}$  is the maximal processivity observed,  $\Delta F_p$  is the difference between the maximal and minimal processivity observed,  $n$  is the number of cation binding sites, and the  $K_a$  is the average association rate constant for cation binding.

We found that a small percentage (~4%) of the substrate contains a ring-opened form of  $\epsilon A$ , such that only a single  $\epsilon A$  lesion is available for AAG-catalyzed excision (see Appendix A). This heterogeneity is expected to decrease the observed fraction processive. Assuming that AAG does not distinguish between oligonucleotides containing one or two lesions, it is predicted that single  $\epsilon A$  substrates will constitute ~4% of the binding encounters during initial rate conditions (This has the effect of decreasing the initial rate of product formation by 4% and increasing the initial rate of intermediate formation by 4%). AAG-catalyzed excision of  $\epsilon A$  from a substrate that contains only a single site of damage will generate an intermediate length oligonucleotide of 37 or 34 nucleotides that cannot be distinguished from an intermediate that arose from distributive action (i.e., dissociation by AAG prior to engagement of the second  $\epsilon A$  site). Thus, even if AAG were 100% processive, our preparation of substrate sets an upper limit to the fraction processive of 0.92 [ $F_p = (V_p - V_{int}) / (V_p + V_{int}) = (V_p - 0.04V_p - 0.04V_p) / (V_p -$

$0.04V_p+0.04V_p) = 0.92$ ; in which  $V_p$  and  $V_{int}$  are the sums of the initial rates for products and intermediates;  $V_p=V_A+V_C$ ;  $V_{int}=V_{AB}+V_{BC}$ ].

The processivity function predicts a minimum value of 0 for purely distributive action. In this case, the initial rates for formation of products and intermediates will be identical ( $V_p=V_{int}$ ). However, for initial rates that proceed up to 10% reaction of substrate the probability of rebinding a released intermediate is not infinitely low. For example, at 10% reaction the probability of rebinding an intermediate is 10% that of binding a new substrate, if one assumes that the majority of binding encounters involve nonspecific interactions with DNA. The gradual accumulation of intermediates is expected to cause downward curvature in the reaction progress curve for the concentration of the intermediates and upward curvature in the reaction progress curve for the concentration of products. The effect of rebinding intermediate in a purely distributive mechanism can be roughly estimated by evaluating the processivity at both 10 and 20% reaction. When 10% product is formed, the predicted velocity for formation of product will not be affected because binding to either substrate or intermediates gives rise to the same products ( $V_{p,corrected} = 0.9V_p+0.1V_p = V_p$ ). However, the corrected velocity for formation of the intermediates will be decreased because fewer enzymatic events form intermediates from substrate and because action on existing intermediates decreases the amount of intermediates ( $V_{int,corrected} = 0.9V_p-0.1V_p = 0.8V_p$ ). This gives a predicted processivity value of 0.11 [ $F_p = (V_p-V_{int})/(V_p+V_{int}) = (V_p-0.8V_p)/(V_p+0.8V_p) = 0.11$ ]. At 20% reaction this increases to  $F_p = 0.25$ . Therefore, the plateau of  $\sim 0.1$  that we observe for the fraction processive at high ionic strength is likely to reflect a fully distributive mechanism rather than a residual ionic strength independent processive mechanism.

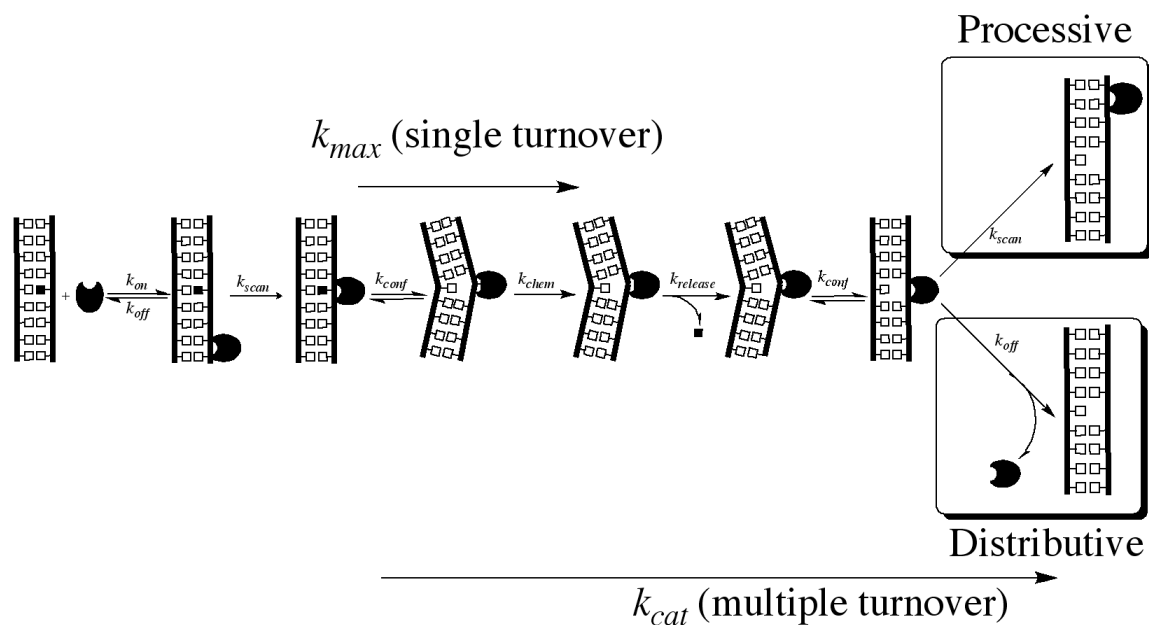
## Results

### **Design of a quantitative processivity assay to characterize the ability of AAG to diffuse along DNA**

We were interested in examining to what extent AAG can use linear diffusion to search multiple nucleotides during a single binding encounter. Previous studies have shown that other base excision repair enzymes are able to act processively at multiple sites on concatameric substrates (18-21). However, the previously employed assays have the limitation that most individual intermediates cannot be resolved. Therefore, we



faster than translocation to the other site of damage), then intermediates corresponding to action at only a single site will accumulate at the same rate as the terminal fragments (enzymatic events  $E_1$  and  $E_2$ ; Figure 2-1). This extreme is referred to as distributive action. Depending on the distance between the target sites and the solution conditions, the behavior of the enzyme is expected to lie somewhere between these two extremes. We use the fraction processive ( $F_p$ ) to define the fraction of enzymatic binding events that are processive versus the total number of processive and distributive events (Scheme 2; see Methods and Appendix A). The average processivity can be determined by following many DNA binding events under steady state conditions (15).



Scheme 2

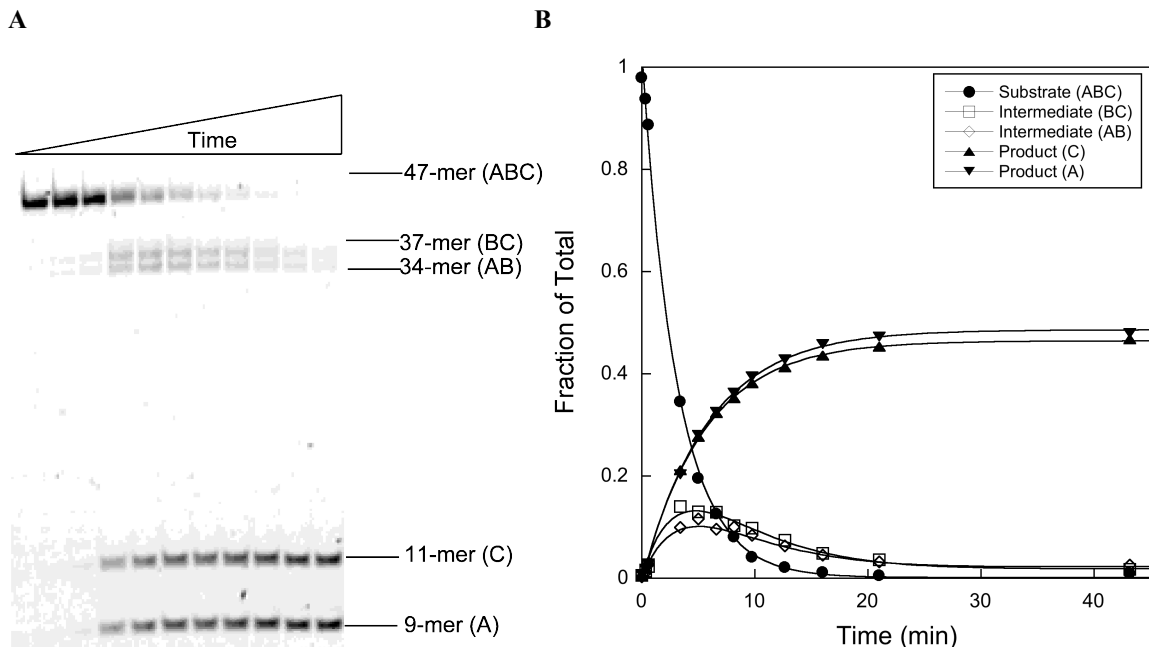
AAG is a monofunctional DNA glycosylase and its products are an abasic site and a free nucleobase. To monitor creation of the abasic site, enzymatic reactions were quenched in sodium hydroxide and heated to quantitatively convert the abasic sites into single strand breaks. Substrates, products and intermediates (resulting from base excision at one of the two sites of damage) were separated on a denaturing polyacrylamide gel and their relative intensities quantified with a fluorescence scanner (see Figures 2-1 & 2-2). The internal fragment (B; Figure 2-1) is unique to processive enzymatic events, but it is not labeled and cannot be detected directly in our experiments. Nevertheless, we can calculate the fraction processive because all of the other DNA species are observed and independently quantified (see Materials and Methods and Appendix A).

Human AAG exists in at least two splice forms that differ slightly at their amino terminus and the larger splice variant is 298 amino acids in length (36-38). The amino terminus of AAG is poorly conserved even among mammals (see Appendix A). In contrast, the carboxy-terminal glycosylase domain is highly conserved across vertebrates and sequence conservation can be detected in some prokaryotic DNA glycosylases (39, 40). The amino-terminal portion of human AAG is sensitive to proteolytic degradation and it was not included in the crystal structures of AAG bound to DNA substrate and inhibitor (31, 41). Our initial experiments used a truncated form of AAG ( $\Delta 80$ ) that lacks the first 79 amino acids. This protein appears to be fully functional for *N*-glycosylase activity in a variety of *in vitro* assays (9, 36, 42). Other studies have implicated the amino terminus in interactions with other proteins, hHR23a/b and MBD1 (43, 44). Therefore, we have also examined the processivity of the full-length recombinant protein (longest splice variant).

### **Characterization of the processivity substrate**

We performed single turnover reactions with excess  $\Delta 80$  AAG over DNA as an initial characterization of the substrate (Figure 2-2). Under these conditions, both sites of damage can be simultaneously saturated (two AAG molecules per substrate oligonucleotide), so that any differences in reaction rate at the two sites can be monitored. As the two  $\epsilon$ A sites have very similar sequence contexts, it was expected that the two sites would have similar reactivity (Figure 2-1A). Indeed, the single turnover rate constant was essentially identical for both sites [ $k_{\max} = 0.20 \pm 0.01 \text{ min}^{-1}$ ; Figures 2-2 and 2-3]. This rate constant is identical to the previously reported single turnover excision of a 25-mer  $\epsilon$ A-containing oligonucleotide that shares the same sequence context [ $k_{\max} = 0.2 \text{ min}^{-1}$ ; (5)]. The observed disappearance of substrate occurs at twice the rate ( $k_{\text{obs}} = 0.4 \text{ min}^{-1}$ ) since excision at either of the two sites depletes the concentration of substrate (Figure 2-2B). For two independent excision events, the intermediates, in which only a single  $\epsilon$ A is excised, build up and then decay as a function of the rate constants for excision at both sites, which are identical in this case. We let the AAG-catalyzed reaction go to completion in order to directly compare the fluorescence of the 5'- and 3'- labeled fragments (Figure 2-2). The almost identical fluorescence of the two bands indicates that the 5'-(6-amino)fluorescein and 3'-fluorescein labels that were used have very similar

quantum yields. Therefore, no corrections need to be made to the raw fluorescence values obtained from scans of the gel.



**Figure 2-2: Characterization of the processivity substrate under single turnover conditions**

A representative time course for AAG-catalyzed excision of  $\epsilon$ A with 35 nM oligonucleotide duplex and 350 nM  $\Delta$ 80 AAG at pH 6.1 and an ionic strength of 300 mM (see Materials and Methods for details). (A) Fluorescent scan of a 20% denaturing polyacrylamide gel showing increasing incubation times from left to right. The positions of the 47-mer substrate (ABC), 37-mer (BC), 34-mer (AB), 11-mer (C), and 9-mer (A) products are shown on the right (see Figure 2-1 for a schematic of the substrate and expected products resulting from *N*-glycosidic bond cleavage and hydroxide-catalyzed abasic site hydrolysis). (B) The amount of each labeled DNA is expressed as a fraction of the total fluorescence and the symbol legend is inset into the plot. Since the substrate contains two labels and the other species contain only a single label, the maximum for the product and intermediate bands is 0.5. The very similar maximal fractions observed for products and intermediates demonstrate that no correction is needed for either labeling efficiency or quantum yield of the 5' and 3'-fluorescein labels. Furthermore, the almost identical rate constants indicate that both sites are recognized by AAG with equal efficiency.

Since  $\epsilon$ A is susceptible to ring-opening, especially at alkaline pH (32), we were concerned that some of the  $\epsilon$ A sites might be damaged during solid phase synthesis, deprotection, and purification. Indeed, this could explain the persistence of the intermediate fragments AB and BC over long reaction times (e.g., Figure 2-2). Since AAG has little or no activity against the ring-opened form of  $\epsilon$ A (45), such damage would lead to an underestimate of the degree of processivity because the enzyme could only act on these substrate molecules once, regardless of residence time. We allowed the single turnover reaction with AAG to proceed for more than 10 half-lives and determined that greater than 99% of the fluorescein-labeled substrate contains at least one  $\epsilon$ A that



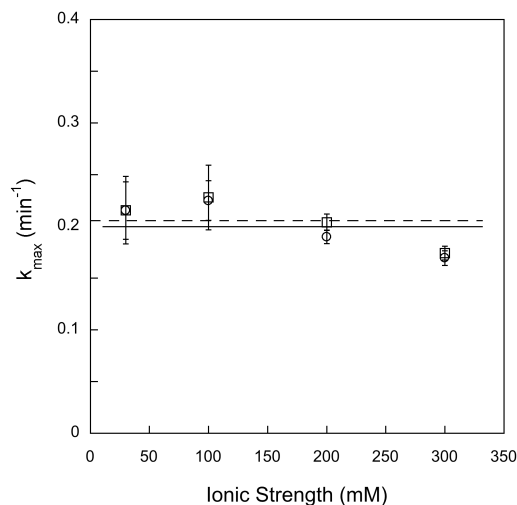
could be recognized by AAG and ~95% contains two  $\epsilon$ A lesions. This is consistent with a ~2% chance that a given  $\epsilon$ A nucleotide undergoes spontaneous degradation during synthesis, deprotection, and purification. Control reactions with *E. coli* FPG, an enzyme known to be active on the ring-opened form of  $\epsilon$ A (45), confirmed that the majority of the AAG-resistant lesions could be excised by FPG and are likely to be ring-opened  $\epsilon$ A bases (see Appendix A). This small percentage of substrate that is refractory to dual excision by AAG (~4%) leads to a slight underestimate of the processivity. We calculate a theoretical maximum of  $F_p = 0.92$  for a fully processive enzyme acting on this substrate (see Materials and Methods). Since none of our conclusions rely on the exact value of the fraction processive, we have not corrected any of the observed values of  $F_p$  for the heterogeneity existing in the DNA substrate.

To ensure that the AAG•DNA complex is fully saturated at high ionic strength, we investigated the ionic strength dependence of the single turnover reaction at several concentrations of AAG. At saturating concentrations, the single turnover rate constant reports on steps that occur subsequent to DNA binding up to and including *N*-glycosidic bond cleavage (Scheme 2). Since base flipping is expected to be fast, *N*-glycosidic bond cleavage is likely to be rate-limiting under these conditions (5). In the range from 30 to 300 mM ionic strength, the single turnover rate constant is independent of ionic strength and identical for both sites ( $k_{\max} = 0.20 \pm 0.01$ ; Figure 2-3). These data are for  $\Delta 80$  AAG, but the single turnover rate constant for full-length AAG under these conditions was the same within error ( $k_{\max} = 0.23 \pm 0.04$ ; data not shown). These data confirm that the AAG•DNA complex is saturated at both sites even at high salt concentrations, and suggest that AAG-catalyzed *N*-glycosidic bond cleavage is insensitive to ionic strength in this range.

### **Processivity of AAG at low ionic strength**

Previous studies of processive action of enzymes on DNA have consistently found that processivity is greatest at low ionic strength and that it decreases with increasing ionic strength, presumably because dissociation from DNA is accelerated at increased ionic strength (15, 18, 46). Therefore, we performed multiple turnover experiments under low ionic strength conditions ( $I = 50$  mM) to address whether AAG exhibits a processive searching mechanism. AAG has a slightly acidic pH-rate optimum

for excision of neutral lesions, including  $\epsilon$ A, so reactions were initially carried out at pH 6.1 (5, 9). Under these conditions, the multiple turnover reaction was very slow, requiring incubation times of up to 12 hours. This corresponds to multiple turnover rate constants

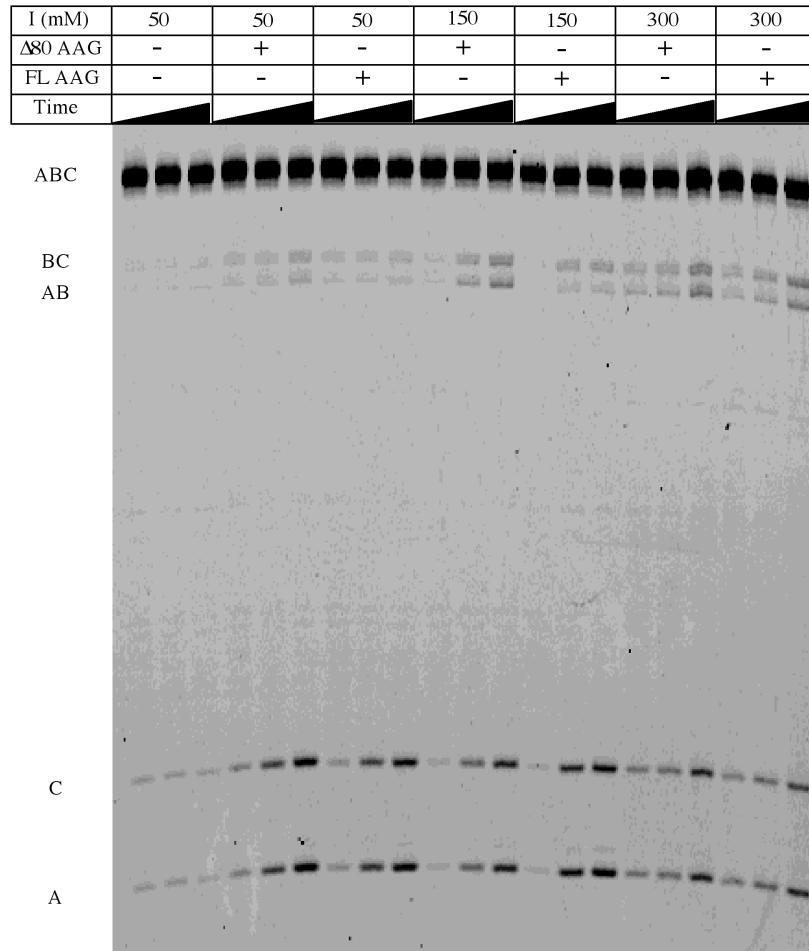


**Figure 2-3: Single turnover excision of  $\epsilon$ A is independent of ionic strength and the two sites are equivalent**

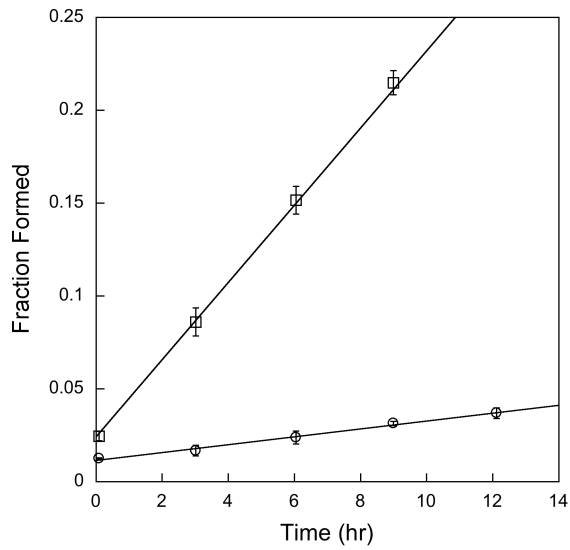
Single turnover excision of  $\epsilon$ A at site 1 (open circles) and site 2 (open squares) were measured with 1  $\mu$ M DNA and 3, 6, and 9  $\mu$ M  $\Delta$ 80 AAG. The observed rate constants were independent of the concentration of AAG, so the average and standard deviation are shown (each data point represents at least 9 independent determinations of the rate constant). The rates of excision at both sites are the same within error and independent of ionic strength ( $k_{\max} = 0.2 \text{ min}^{-1}$ ). These results are the same as was previously reported for excision of a single  $\epsilon$ A lesion from a similar sequence context using a  $^{32}\text{P}$ -based glycosylase assay, suggesting that the activity of AAG is not affected by either of the fluorescein labels (5).

that are 10–20-fold lower than the single turnover rate constant. Stability controls indicated that both full-length and  $\Delta$ 80 AAG retain at least 50% activity over 12 hours when incubated without DNA, but are considerably more stable when incubated with DNA substrate so that no loss of AAG activity was detected during the course of the assay (see Appendix A). Figure 2-4 shows a representative gel from a multiple turnover processivity experiment. The left-most lanes show time courses for reactions with no enzyme,  $\Delta$ 80 AAG, or full-length AAG at 50 mM ionic strength (Figure 2-4A). For both enzymes it is apparent that the products resulting from processive action build up much more quickly than the intermediates that result from distributive action, as expected for a processive mode of action (Figure 2-4B & 2-4C). Surprisingly,  $\Delta$ 80 AAG was  $\sim$ 2-fold faster than full-length AAG for multiple-turnover excision of  $\epsilon$ A under these conditions (see below). The fraction processive was calculated as described in the Methods and gave values of  $0.76 \pm 0.09$  and  $0.88 \pm 0.04$  for  $\Delta$ 80 and full-length AAG, respectively,

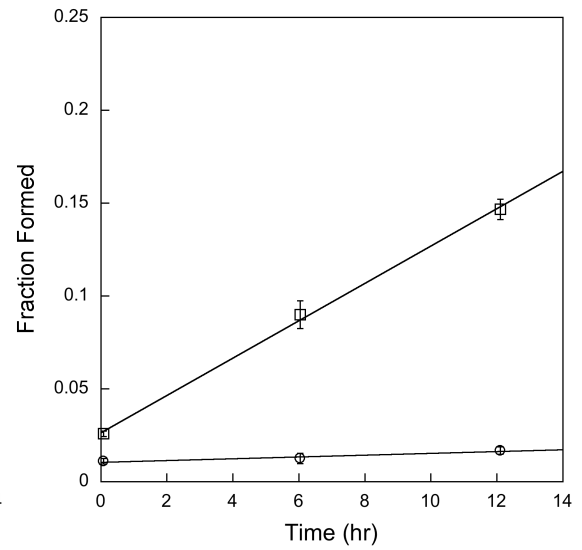
**A**

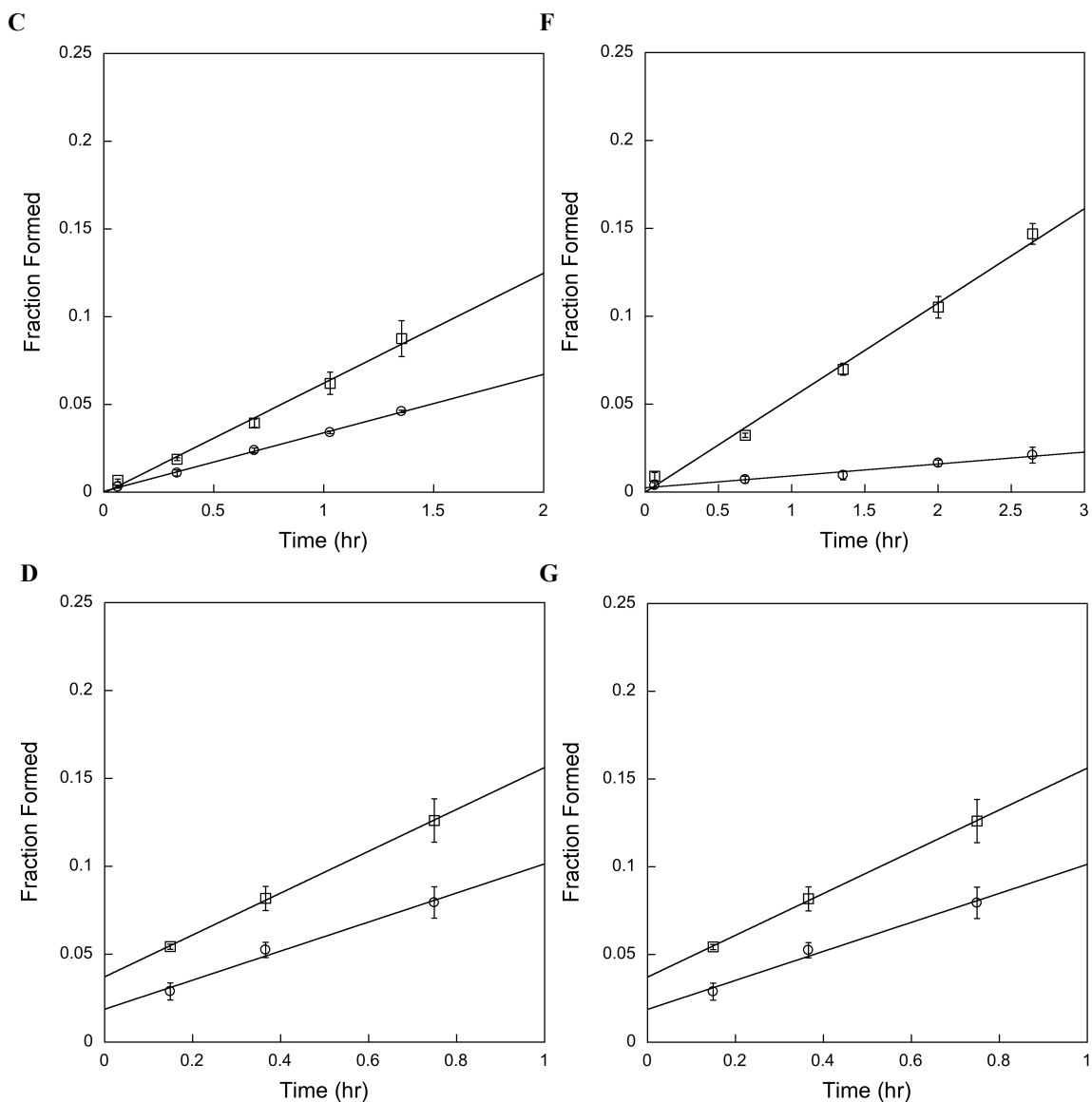


**B**



**E**





**Figure 2-4: The multiple turnover processivity assay**

(A) This Representative gel compares reactions containing full-length or  $\Delta 80$  AAG at pH 6.1 with 50–150 mM ionic strength. All reactions contained 2  $\mu\text{M}$  oligonucleotide substrate and 20 nM enzyme. Three time points between 5 and 20% reaction were chosen for each reaction condition (the time varies between 0.1 and 12 hours, since the steady state rate is dependent upon ionic strength). (B-G) Reaction progress curves for products (fragments A & C) and intermediates (fragments AB & BC) are shown for each set of reactions. In this experiment each reaction was performed in triplicate and additional time points analyzed on additional gels are included. The average values for each condition are plotted and the error bars indicate the standard deviation. Panels B - D show results obtained with  $\Delta 80$  AAG and Panels E - G show results obtained with full-length AAG. The ionic strength was 50 mM (Panels B & E), 150 mM (Panels C & F), and 300 mM (Panels D & G). The fraction processive was calculated from these data and from additional experiments to obtain the average  $F_p$  values that are shown in Figure 2-6A.

indicating that both enzymes are highly processive at low ionic strength.

Under multiple turnover conditions, there is very low probability of multiple proteins binding to the same DNA molecule so any intrinsic difference in binding or base

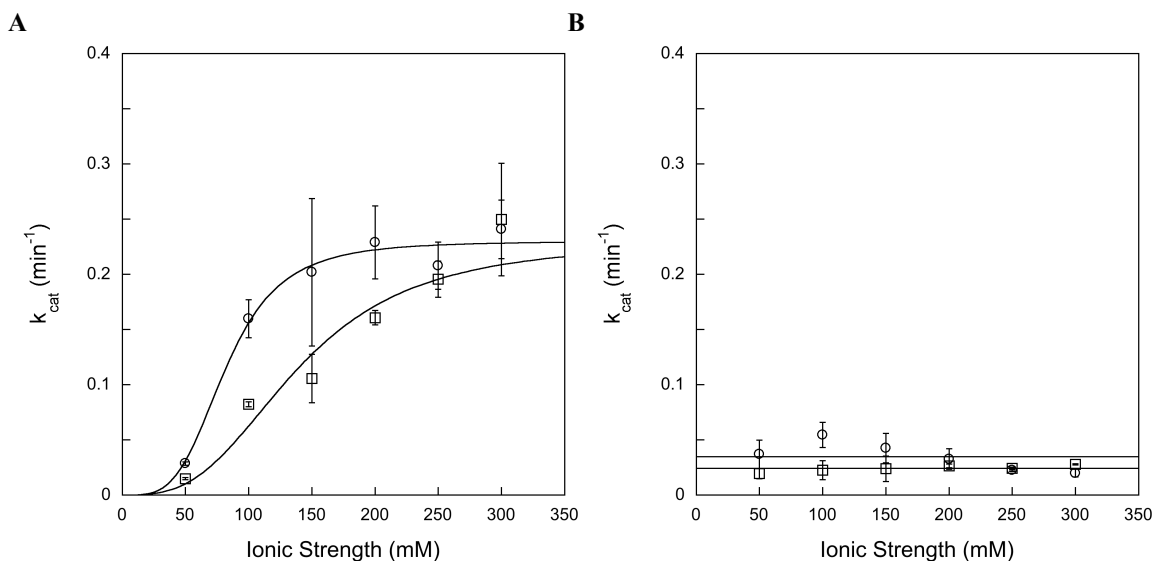
excision at the two sites can be detected. Both products (fragment A and C) are formed at essentially the same rate and therefore AAG does not have a preference for either the 5' or 3'-εA lesion (Figure 2-4A and data not shown). Although much lower amounts of the two intermediate fragments (AB and BC) were formed under these conditions, the rates for their formation were also the same within error (Figure 2-4A and data not shown). In principle, it is possible for an enzyme to exhibit different processivity depending upon which site it acts at first (i.e.,  $E_{1,2} \neq E_{2,1}$ ; Figure 1). We considered this possibility, but did not find a significant difference between the two possible pathways [(15, 17); See Appendix A]. This suggests that AAG recognizes and excises the lesions from the two sites with identical efficiency despite the different polarity and distance from the two DNA ends. Therefore, we routinely determined initial rates for the sum of the two products and for the sum of the two intermediates (e.g., Figure 2-4B & C). This facilitated our ability to measure extremely slow rates of formation of intermediates.

#### **Effect of ionic strength on multiple turnover reaction and processivity**

The ionic strength dependence of the steady state reaction was characterized in order to better understand the slow steady state rate at low ionic strength and the modest, but reproducible, increase in the reaction rate for the truncated form of AAG relative to the full-length protein. Since the DNA binding surface of AAG contains many charged groups and because linear diffusion along DNA by other enzymes is very sensitive to ionic strength, we anticipated that increased ionic strength would cause a switch to a distributive mechanism. Comparison of processivity of full-length and truncated proteins at increased ionic strength would allow even a modest change in processivity to be detected.

First, we determined the effect of added sodium chloride on the steady state reaction rate. The  $k_{cat}$  values were obtained from linear fits to the initial rates for disappearance of substrate for reactions in which the ionic strength was adjusted between 50 mM and 300 mM by the addition of sodium chloride (see Appendix A for a representative plot). The results for both full-length and Δ80 AAG are summarized in Figure 2-5A. For Δ80 AAG, the multiple turnover rate constant increases with increasing ionic strength until ~200 mM at which point it reaches the single turnover rate constant. The full-length AAG follows a very similar ionic strength dependence, but does not reach

the single turnover rate constant until  $\sim 250$  mM ionic strength. Across the range between 50 and 200 mM ionic strength the truncated enzyme maintains  $\sim 2$ -fold faster rate of reaction. Although this difference between full-length and  $\Delta 80$  AAG is modest, there are several reasons to believe that it is real. To control for possible differences in protein concentration, we performed active site titrations of both enzymes on the same substrate to determine the active concentration of enzyme (see Appendix A). To control for possible differences in DNA substrate concentration, since the  $k_{\text{cat}}$  value in a gel-based assay is also dependent upon this concentration, the steady state kinetics were carried out side-by-side with the same stock of DNA. Furthermore, both enzymes gave the same  $k_{\text{cat}}$  value at high ionic strength and this rate constant was in very close agreement with the  $k_{\text{max}}$  values for single turnover excision.



**Figure 2-5: Ionic strength affects multiple turnover excision of  $\epsilon A$  at pH 6.1, but not at pH 7.5**  
 The multiple turnover rate constants for  $\Delta 80$  (open circles) and full-length AAG (open squares) were measured at the indicated ionic strength. The average of 3-8 replicates is shown and the error bars indicate one standard deviation from the mean. The solid lines shows the best fits to a cooperative model in which multiple sodium ions cause an increased rate of dissociation up to the threshold at which the rate of dissociation is greater than the rate constant for  $N$ -glycosidic bond cleavage (see Methods for details).

The ionic strength dependence of the steady state excision by AAG strongly suggests that a different step is rate-limiting at low and high ionic strength. The rate-limiting step at low ionic strength increases in response to added salt up to the point at which an ionic strength independent step becomes rate limiting. At high ionic strength the excellent agreement between the single turnover and multiple turnover rate constants indicates that the rate-limiting step is the same; hydrolysis of the  $N$ -glycosidic bond. The

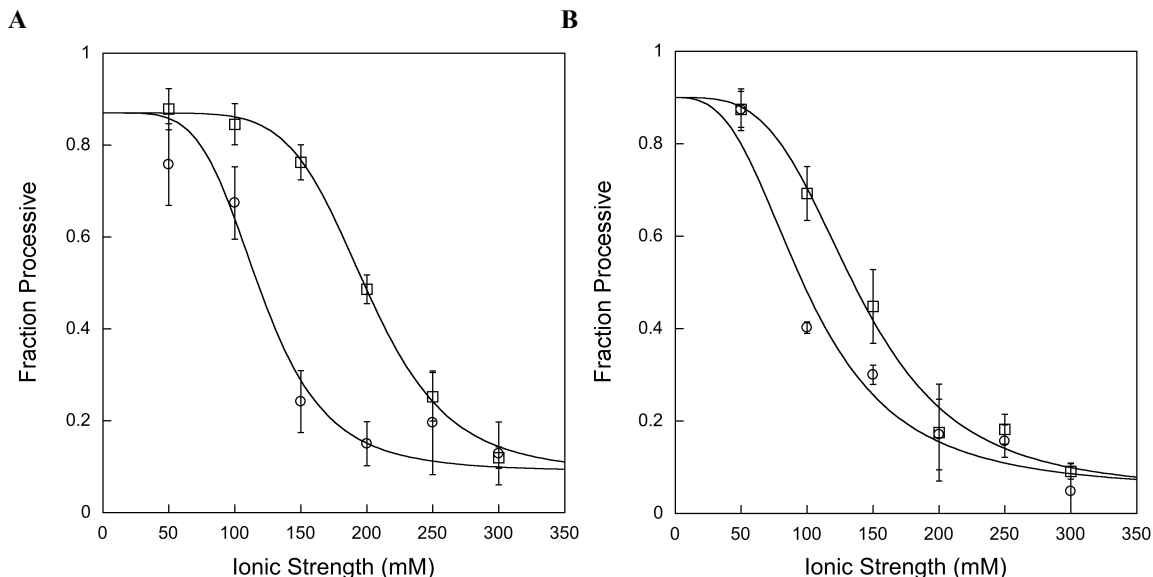
rate-limiting step at low ionic strength for multiple turnover excision must occur subsequent to hydrolysis of the *N*-glycosidic bond and is most likely to be dissociation of the DNA product. An increased dissociation rate constant at higher concentration of salt is consistent with weakened electrostatic interactions between the positively charged protein and the negatively charged DNA.

We compared the ionic strength dependence of the fraction processive for both full-length and  $\Delta 80$  AAG and the results are shown in Figure 2-6A. Both proteins show a steep decrease in processivity with increasing ionic strength, however the full-length protein shows significantly higher processivity at intermediate ionic strength (100–200 mM). We believe that the lower limit of the processivity of  $\sim 0.1$  that both proteins approach at high ionic strength is reflective of a fully distributive mechanism, since this value is expected for distributive mechanisms at 10-20% reaction due to the low, but finite, probability of rebinding an intermediate containing a single  $\epsilon A$  lesion (see Materials and Methods). The observation of decreased fraction processivity under certain conditions provides an important validation of the processivity assay by addressing a trivial alternative interpretation of the low levels of intermediates observed in the steady state assay. For example, if AAG preferentially rebinds a substrate with an abasic site, then the pattern of products would appear to be processive even though the mechanism of base excision was distributive. Further evidence against this alternative model is that linear initial rates are observed for greater than 40% of the reaction, indicating that even when abasic product and  $\epsilon A$ -containing substrate are present in roughly equal amounts, AAG preferentially binds to the substrate (see Appendix A for a representative time course).

### **Multiple turnover and Processivity of AAG at physiological pH**

The initial processivity experiments were carried out at pH 6.1 because this is the optimal pH for AAG-catalyzed excision of  $\epsilon A$  *in vitro*. However, since  $\epsilon A$  base excision is slower at higher pH, we also performed processivity experiments at pH 7.5. The  $k_{\text{cat}}$  values for both enzymes are significantly lower at pH 7.5 than at pH 6.1 (Figure 2-5). However, in contrast to the ionic strength-dependent multiple turnover reaction observed at lower pH, the multiple turnover reaction at higher pH was independent of ionic strength between 50 and 300 mM (Figure 2-5B). The change in the ionic strength

dependence suggests that multiple turnover excision is limited by *N*-glycosidic bond cleavage at pH 7.5, and that dissociation of the abasic product is relatively fast.



**Figure 2-6: Ionic strength affects the processivity of AAG**

Processivity at the optimal pH of 6.1 (A) and at pH 7.5 (B) was determined at increasing ionic strength, as described in the methods. Both  $\Delta 80$  (open symbols) and full-length AAG (closed symbols) were examined. The average value of 3-8 independent determinations is shown and the error bars indicate the standard deviation for each condition.

As described above for the pH 6.1 conditions, the fraction processive was calculated for full-length and truncated AAG as a function of ionic strength at pH 7.5 (Figure 2-6B). The results closely resemble the results at the lower pH, with maximal processivity at low ionic strength ( $F_p = 0.87 \pm 0.04$  for both full-length and  $\Delta 80$  AAG) and with decreasing processivity at higher ionic strength. The midpoint of the ionic strength dependence is shifted to lower ionic strength for both full-length and truncated proteins, relative to the midpoint observed at pH 6.1. For intermediate ionic strength at both pH 6.1 and pH 7.5, the full-length protein exhibits higher processivity than the  $\Delta 80$  truncated protein. Although the contribution of the amino terminus to a processive search is modest, it is interesting to note that this effect is greatest at physiological ionic strength. This observation is consistent with the idea that residues in the amino terminus fine-tune the nonspecific DNA binding activity of AAG to allow for short-range correlated searching of adjacent sites.



## Discussion

We have investigated the mechanism by which human AAG, a DNA repair glycosylase that recognizes a wide variety of alkylated and deaminated purines, locates sites of DNA damage. We describe a processivity assay that allows quantitative measurement of the ability of AAG to remove multiple base lesions from a simple oligonucleotide substrate in a single binding encounter. This assay has provided evidence that AAG employs a processive searching mechanism that makes use of nonspecific DNA binding interactions to carry out a highly redundant search of adjacent sites. By comparing an amino-terminally truncated enzyme to the full-length enzyme, we found that the amino terminus plays a role in nonspecific DNA binding and increases the probability of a correlated search at physiological ionic strength. These results are similar to those obtained for a variety of DNA binding enzymes and further supports the idea that nonspecific DNA binding is an important feature of enzymes that must carry out genome-wide searches for specific sites [e.g., (11, 14, 15, 23, 24, 47-49)].

The relatively short, dual-lesion substrate that we have utilized allows a simple and quantitative measure of the ability of a DNA repair glycosylase to translocate between nearby sites. This assay has several advantages over more commonly employed assays involving concatameric substrates. Most importantly, the small substrate size allows each possible product to be quantified and any inherent directionality of the scanning process to be detected. In addition, these substrates can be directly synthesized to allow for a wide-range of site-specific modifications, such as the incorporation of fluorescent labels. However, the dual-lesion assay has the disadvantage that long sliding distances cannot be measured, because solid phase synthesis is limited to relatively short oligonucleotides. We have overcome this limitation by increasing the ionic strength to weaken the protein-DNA interaction. Beyond allowing quantitative comparison of mutants or alternative substrates, the ionic strength dependence provides mechanistic insight into the DNA damage recognition process.

At low ionic strength, both full-length and  $\Delta 80$  AAG exhibit a high degree of processivity at both pH 6.1 and pH 7.5 (Figure 2-6). This demonstrates that nonspecific DNA binding by AAG enables a correlated search over a distance of at least several turns of the DNA helix. Since the  $\epsilon$ A lesions are separated by 25 base pairs and the pitch of B-

form DNA is 10.4 base pairs per turn, these lesions are predicted to be on opposite sides of the DNA duplex. This rules out a hand-over-hand model for diffusion along one face of the duplex that would be analogous to the models that have been proposed for the movement of motor proteins such as myosin and kinesin along either actin or microtubule filaments (50, 51). These data are consistent with either a one-dimensional mode of diffusion along one strand of the duplex (52, 53), or a two-dimensional mode of diffusion in which both strands can be simultaneously searched (34, 54). We have fit the ionic strength dependence of the fraction processive with a cooperative model whereby multiple cations affect the probability of finding a second site of damage. In principle, ionic strength dependent changes in dissociation rate, scanning rate, or excision rate could be responsible for the ionic strength dependence of the processivity. However, the base excision step is insensitive to ionic strength because the single turnover reaction does not change between 50 and 300 mM ionic strength (Figure 2-3). Therefore, the simplest interpretation of the ionic strength dependence of the processivity is that the dissociation rate is dependent on ionic strength and that the rate of scanning is not. Consistent with this, the ionic strength at which half of the maximal  $k_{cat}$  is obtained is similar to the ionic strength at which half of the maximal processivity is observed. Nevertheless, these experimental observations cannot rule out the possibility that the rate of scanning is also dependent upon the ionic strength.

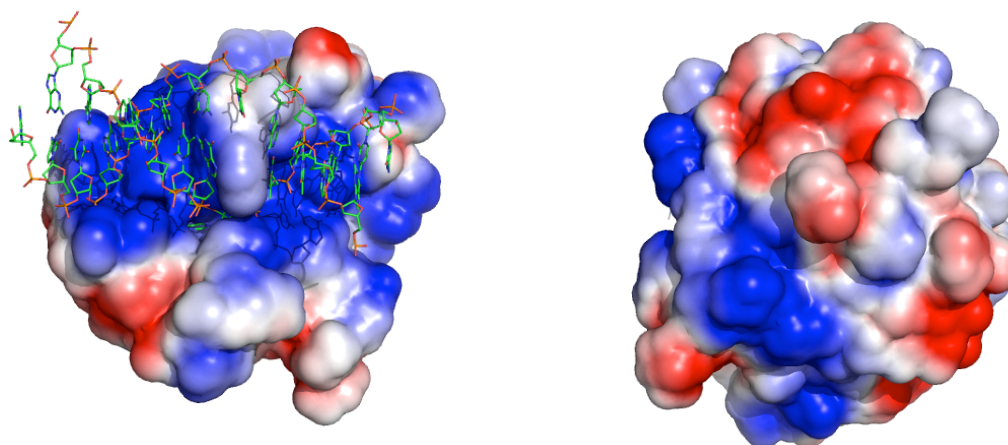
At physiological pH and ionic strength (pH 7.5, I=150 mM), AAG exhibits a processivity of 0.45, which is near the midpoint of the range of processivity values that we observe (0.1–0.9; Figure 2-6B). Under these conditions, AAG has a roughly equal probability of finding the second site of damage or dissociating. It is possible that interactions with other proteins or covalent modifications of AAG increase its processivity *in vivo*. However, the observed *in vitro* processivity is consistent with the requirements for a genome-wide search *in vivo*. The dissociative extreme in which only a single base is sampled per binding encounter would be inefficient because the search for damage would involve the entire three-dimensional space of the nucleus. The associative extreme would be inefficient because the repair protein would be restricted to distinct domains of DNA and movement between accessible regions of DNA would be limited. Therefore, an intermediate level of processivity, in which short sections of DNA are

exhaustively searched prior to dissociation, is expected to balance the requirements of covering every single base of the genome (14, 16).

At the optimal pH of 6.1, the  $k_{\text{cat}}$  value for both  $\Delta 80$  and full-length AAG is dependent upon the ionic strength. Since the single turnover rate constant is independent of ionic strength, this suggests that dissociation from the abasic DNA product is rate-limiting for multiple turnover base excision. The AAG DNA binding surface has a high density of positive charge, so it is not surprising that high ionic strength weakens binding (Figure 2-7). We fit this ionic strength dependence with a simple model in which the rate of dissociation is dependent upon the number of sodium ions bound. At high ionic strength, dissociation becomes sufficiently fast that an earlier step, presumably *N*-glycosidic bond hydrolysis, becomes rate limiting. Consistent with this interpretation, the multiple turnover rate constant reaches the same value as the  $k_{\text{max}}$  value for single turnover (Figure 2-5A). In contrast, at pH 7.5 the value of  $k_{\text{cat}}$  is independent of ionic strength, suggesting that *N*-glycosidic bond hydrolysis step is rate-limiting for multiple turnover at even the lowest ionic strength tested. This is consistent with the observation that the single turnover rate constant for excision of  $\epsilon A$  is  $\sim 8$ -fold lower at pH 7.5 relative to pH 6.1 (9).

The slow rate constants that are observed for dissociation of the abasic product, reflected by slow multiple turnover excision at low ionic strength, suggest a relatively long-lived AAG•DNA complex. For example, at pH 6.1 and 150 mM ionic strength the full-length protein has a half-life of 7 minutes [ $t_{1/2} = \ln 2/k_{\text{dissociation}} = \ln (2/0.1 \text{ min}^{-1}) = 7 \text{ min}$ ] for its dissociation from DNA containing an abasic DNA product. This can be compared to the half-life of 3 minutes for the  $\Delta 80$  truncated protein. At 50 mM ionic strength, the half-life of the AAG•DNA complexes increases to 46 and 23 minutes for the full-length and truncated proteins, respectively. Presumably dissociation would be faster from undamaged DNA, but even 100-fold weaker binding to undamaged DNA would imply a dissociation half-time of many seconds. A single molecule study of 8-oxoguanosine DNA glycosylase provided an estimate of as many as 3000 base pairs sampled per second on undamaged DNA (55). If AAG exhibits similar fast sliding, then this implies a massively redundant search of adjacent sites in this long-lived AAG•DNA

complex, in which case our lower estimate of >25 base pairs is an underestimate of the distance traveled during a binding encounter.



**Figure 2-7: Electrostatic surface potential of the AAG catalytic domain reveals a positively charged DNA binding surface.**

The crystal structure of the AAG• $\epsilon$ A-DNA complex was used to generate this figure (1F4R; (31)). Electrostatic calculations were performed with Pymol on the protein alone (W.L. DeLano; <http://www.pymol.org>) and APBS ((58); using a plug-in written by M. Lerner; <http://www-personal.umich.edu/~mlerner/PyMOL/>). A continuum from -2 (red) to +2 (blue) is shown. A view of the active site and bound DNA is on the left, illustrating the positively charged DNA binding surface, and on the right the molecule is rotated horizontally by 180° to show that the positively charged surface continues around the protein.

The importance of electrostatic interactions for the stabilization of the AAG•DNA complex is apparent from the increasing rate of dissociation from DNA and from the decreasing processivity that are observed upon increasing the ionic strength. These observations are consistent with the positively charged DNA binding groove that is observed in crystal structures of the AAG• $\epsilon$ A-DNA complex [Figure 7; (31, 41)]. The DNA binding groove contains 4 positively charged residues that directly contact the phosphate backbone (Arg 141, Arg197, Arg182, K229). It is not known whether there are additional contacts between the amino terminus of AAG, which was not present in the crystal structure, and the DNA. However, there are 13 arginine and lysine residues present in the first 80 amino acids, one or more of which could provide a positive electrostatic interaction. The increase in the multiple turnover rate constant and the decreased processivity at higher ionic strength that was observed upon deletion of the amino terminal 80 amino acids suggests that the amino terminus either contacts DNA directly or alters the conformation of AAG to slow dissociation. However, the amino

terminus does not appear to have any effect on the rate of base excision since the rate constant for the full-length protein reaches the rate constant of the  $\Delta 80$  truncated protein at high ionic strength (Figure 2-5).

Although it is difficult to quantitatively compare the processivity results from AAG with the two-lesion substrate with the results from concatameric multiple-lesion substrates that have been reported for other base excision repair enzymes, there are some obvious qualitative similarities. Whereas several enzymes exhibit processive excision of adjacent lesions at low ionic strength ( $\leq 70$  mM), including *E. coli* and human UNG (19, 20), *M. luteus* (56) and T4 endoV (46, 57), human APE1 (18), *E. coli* MutY (21), and *E. coli* FPG (21), their behavior becomes distributive when the ionic strength is greater than 70 mM. Only *E. coli* FPG acting on an 8-oxoguanosine•cytosine-containing concatameric DNA showed processive behavior with ionic strength greater than 100 mM (21). It appears that AAG is more processive than these other enzymes that have been previously studied, because it shows a bias towards processive action up to 200 mM ionic strength at optimal pH and up to 150 mM at physiological pH. Taken together, the results from a variety of eukaryotic and prokaryotic base excision repair enzymes are consistent with the idea that a coordinated search is important over relatively short distances along DNA.

In summary, our results reveal that AAG searches many adjacent sites on a DNA molecule in a single binding event prior to dissociation. This observation suggests that the majority of lesion recognition events involve initial nonspecific binding to undamaged sites followed by diffusion along the DNA. This searching process is expected to be highly redundant given the long lifetime of the AAG•DNA complex, providing ample opportunity for the enzyme to recognize and excise lesions that minimally perturb the structure of DNA. As ionic strength is increased above physiological levels, the rate of dissociation from DNA increases and AAG switches to a distributive searching mechanism. In addition, deletion of the amino terminal 80 amino acids, a region dispensable for catalytic activity, results in significantly decreased processivity at physiological ionic strength. These observations suggest that the nonspecific binding affinity of AAG is tuned to allow for correlated searches of a local DNA domain, while still allowing freedom for a long-range three-dimensional search.

## References

- (1) Ye, N., Holmquist, G. P., and O'Connor, T. R. (1998) Heterogeneous repair of N-methylpurines at the nucleotide level in normal human cells. *J Mol Biol* 284, 269-85.
- (2) Chen, D. S., Herman, T., and Demple, B. (1991) Two distinct human DNA diesterases that hydrolyze 3'-blocking deoxyribose fragments from oxidized DNA. *Nucleic Acids Res* 19, 5907-14.
- (3) Cappelli, E., Hazra, T., Hill, J. W., Slupphaug, G., Bogliolo, M., and Frosina, G. (2001) Rates of base excision repair are not solely dependent on levels of initiating enzymes. *Carcinogenesis* 22, 387-93.
- (4) Lindahl, T. (1993) Instability and decay of the primary structure of DNA. *Nature* 362, 709-15.
- (5) O'Brien, P. J., and Ellenberger, T. (2004) Dissecting the broad substrate specificity of human 3-methyladenine-DNA glycosylase. *J Biol Chem* 279, 9750-7.
- (6) Hitchcock, T. M., Dong, L., Connor, E. E., Meira, L. B., Samson, L. D., Wyatt, M. D., and Cao, W. (2004) Oxanine DNA glycosylase activity from Mammalian alkyladenine glycosylase. *J Biol Chem* 279, 38177-83.
- (7) Berdal, K. G., Johansen, R. F., and Seeberg, E. (1998) Release of normal bases from intact DNA by a native DNA repair enzyme. *Embo J* 17, 363-7.
- (8) Connor, E. E., and Wyatt, M. D. (2002) Active-site clashes prevent the human 3-methyladenine DNA glycosylase from improperly removing bases. *Chem Biol* 9, 1033-41.
- (9) O'Brien, P. J., and Ellenberger, T. (2003) Human alkyladenine DNA glycosylase uses acid-base catalysis for selective excision of damaged purines. *Biochemistry* 42, 12418-29.
- (10) Biswas, T., Clos, L. J., 2nd, SantaLucia, J., Jr., Mitra, S., and Roy, R. (2002) Binding of specific DNA base-pair mismatches by N-methylpurine-DNA glycosylase and its implication in initial damage recognition. *J Mol Biol* 320, 503-13.
- (11) Berg, O. G., Winter, R. B., and von Hippel, P. H. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* 20, 6929-48.
- (12) Winter, R. B., Berg, O. G., and von Hippel, P. H. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The Escherichia coli lac repressor--operator interaction: kinetic measurements and conclusions. *Biochemistry* 20, 6961-77.

- (13) Winter, R. B., and von Hippel, P. H. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 2. The Escherichia coli repressor--operator interaction: equilibrium measurements. *Biochemistry* 20, 6948-60.
- (14) Zharkov, D. O., and Grollman, A. P. (2005) The DNA trackwalkers: principles of lesion search and recognition by DNA glycosylases. *Mutat Res* 577, 24-54.
- (15) Terry, B. J., Jack, W. E., and Modrich, P. (1985) Facilitated diffusion during catalysis by EcoRI endonuclease. Nonspecific interactions in EcoRI catalysis. *J Biol Chem* 260, 13130-7.
- (16) Halford, S. E., and Marko, J. F. (2004) How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res* 32, 3040-52.
- (17) Stanford, N. P., Szczelkun, M. D., Marko, J. F., and Halford, S. E. (2000) One- and three-dimensional pathways for proteins to reach specific DNA sites. *Embo J* 19, 6546-57.
- (18) Carey, D. C., and Strauss, P. R. (1999) Human apurinic/aprimidinic endonuclease is processive. *Biochemistry* 38, 16553-60.
- (19) Higley, M., and Lloyd, R. S. (1993) Processivity of uracil DNA glycosylase. *Mutat Res* 294, 109-16.
- (20) Bennett, S. E., Sanderson, R. J., and Mosbaugh, D. W. (1995) Processivity of Escherichia coli and rat liver mitochondrial uracil-DNA glycosylase is affected by NaCl concentration. *Biochemistry* 34, 6109-19.
- (21) Francis, A. W., and David, S. S. (2003) Escherichia coli MutY and Fpg utilize a processive mechanism for target location. *Biochemistry* 42, 801-10.
- (22) Zharkov, D. O., Shoham, G., and Grollman, A. P. (2003) Structural characterization of the Fpg family of DNA glycosylases. *DNA Repair (Amst)* 2, 839-62.
- (23) Lloyd, R. S., Hanawalt, P. C., and Dodson, M. L. (1980) Processive action of T4 endonuclease V on ultraviolet-irradiated DNA. *Nucleic Acids Res* 8, 5113-27.
- (24) Ganesan, A. K., Seawell, P. C., Lewis, R. J., and Hanawalt, P. C. (1986) Processivity of T4 endonuclease V is sensitive to NaCl concentration. *Biochemistry* 25, 5751-5.
- (25) Dowd, D. R., and Lloyd, R. S. (1990) Biological significance of facilitated diffusion in protein-DNA interactions. Applications to T4 endonuclease V-initiated DNA repair. *J Biol Chem* 265, 3424-31.

- (26) Jeltsch, A., Wenz, C., Stahl, F., and Pingoud, A. (1996) Linear diffusion of the restriction endonuclease EcoRV on DNA is essential for the in vivo function of the enzyme. *Embo J* 15, 5104-11.
- (27) McKinney, K., Mattia, M., Gottifredi, V., and Prives, C. (2004) p53 linear diffusion along DNA requires its C terminus. *Mol Cell* 16, 413-24.
- (28) Barbin, A. (2000) Etheno-adduct-forming chemicals: from mutagenicity testing to tumor mutation spectra. *Mutat Res* 462, 55-69.
- (29) Gros, L., Ishchenko, A. A., and Sapparbaev, M. (2003) Enzymology of repair of etheno-adducts. *Mutat Res* 531, 219-29.
- (30) Abner, C. W., Lau, A. Y., Ellenberger, T., and Bloom, L. B. (2001) Base excision and DNA binding activities of human alkyladenine DNA glycosylase are sensitive to the base paired with a lesion. *J Biol Chem* 276, 13379-87.
- (31) Lau, A. Y., Wyatt, M. D., Glassner, B. J., Samson, L. D., and Ellenberger, T. (2000) Molecular basis for discriminating between normal and damaged bases by the human alkyladenine glycosylase, AAG. *Proc Natl Acad Sci U S A* 97, 13573-8.
- (32) Speina, E., Ciesla, J. M., Wojcik, J., Bajek, M., Kusmierk, J. T., and Tudek, B. (2001) The pyrimidine ring-opened derivative of 1,N6-ethenoadenine is excised from DNA by the Escherichia coli Fpg and Nth proteins. *J Biol Chem* 276, 21821-7.
- (33) Bennett, M. T., Rodgers, M. T., Hebert, A. S., Ruslander, L. E., Eisele, L., and Drohat, A. C. (2006) Specificity of human thymine DNA glycosylase depends on N-glycosidic bond stability. *J Am Chem Soc* 128, 12510-9.
- (34) Kampmann, M. (2004) Obstacle bypass in protein motion along DNA by two-dimensional rather than one-dimensional sliding. *J Biol Chem* 279, 38715-20.
- (35) Sidorenko, V. S., Mechetin, G. V., Nevinsky, G. A., and Zharkov, D. O. (2008) Correlated cleavage of single- and double-stranded substrates by uracil-DNA glycosylase. *FEBS Lett* 582, 410-4.
- (36) O'Connor, T. R. (1993) Purification and characterization of human 3-methyladenine-DNA glycosylase. *Nucleic Acids Res* 21, 5561-9.
- (37) Vickers, M. A., Vyas, P., Harris, P. C., Simmons, D. L., and Higgs, D. R. (1993) Structure of the human 3-methyladenine DNA glycosylase gene and localization close to the 16p telomere. *Proc Natl Acad Sci U S A* 90, 3437-41.



- (38) Bonanno, K., Wyrzykowski, J., Chong, W., Matijasevic, Z., and Volkert, M. R. (2002) Alkylation resistance of *E. coli* cells expressing different isoforms of human alkyladenine DNA glycosylase (hAAG). *DNA Repair (Amst)* 1, 507-16.
- (39) Alseth, I., Rognes, T., Lindback, T., Solberg, I., Robertsen, K., Kristiansen, K. I., Mainieri, D., Lillehagen, L., Kolsto, A. B., and Bjoras, M. (2006) A new protein superfamily includes two novel 3-methyladenine DNA glycosylases from *Bacillus cereus*, AlkC and AlkD. *Mol Microbiol* 59, 1602-9.
- (40) Aamodt, R. M., Falnes, P. O., Johansen, R. F., Seeberg, E., and Bjoras, M. (2004) The *Bacillus subtilis* counterpart of the mammalian 3-methyladenine DNA glycosylase has hypoxanthine and 1,N6-ethenoadenine as preferred substrates. *J Biol Chem* 279, 13601-6.
- (41) Lau, A. Y., Scharer, O. D., Samson, L., Verdine, G. L., and Ellenberger, T. (1998) Crystal structure of a human alkylbase-DNA repair enzyme complexed to DNA: mechanisms for nucleotide flipping and base excision. *Cell* 95, 249-58.
- (42) Roy, R., Biswas, T., Hazra, T. K., Roy, G., Grabowski, D. T., Izumi, T., Srinivasan, G., and Mitra, S. (1998) Specific interaction of wild-type and truncated mouse N-methylpurine-DNA glycosylase with ethenoadenine-containing DNA. *Biochemistry* 37, 580-9.
- (43) Miao, F., Bouziane, M., Dammann, R., Masutani, C., Hanaoka, F., Pfeifer, G., and O'Connor, T. R. (2000) 3-Methyladenine-DNA glycosylase (MPG protein) interacts with human RAD23 proteins. *J Biol Chem* 275, 28433-8.
- (44) Watanabe, S., Ichimura, T., Fujita, N., Tsuruzoe, S., Ohki, I., Shirakawa, M., Kawasuji, M., and Nakao, M. (2003) Methylated DNA-binding domain 1 and methylpurine-DNA glycosylase link transcriptional repression and DNA repair in chromatin. *Proc Natl Acad Sci U S A* 100, 12859-64.
- (45) Speina, E., Zielinska, M., Barbin, A., Gackowski, D., Kowalewski, J., Graziewicz, M. A., Siedlecki, J. A., Olinski, R., and Tudek, B. (2003) Decreased repair activities of 1,N(6)-ethenoadenine and 3,N(4)-ethenocytosine in lung adenocarcinoma patients. *Cancer Res* 63, 4351-7.
- (46) Gruskin, E. A., and Lloyd, R. S. (1986) The DNA scanning mechanism of T4 endonuclease V. Effect of NaCl concentration on processive nicking activity. *J Biol Chem* 261, 9607-13.
- (47) von Hippel, P. H., and Berg, O. G. (1989) Facilitated target location in biological systems. *J Biol Chem* 264, 675-8.

- (48) Jack, W. E., Terry, B. J., and Modrich, P. (1982) Involvement of outside DNA sequences in the major kinetic path by which EcoRI endonuclease locates and leaves its recognition sequence. *Proc Natl Acad Sci U S A* 79, 4010-4.
- (49) Gowers, D. M., and Halford, S. E. (2003) Protein motion from non-specific to specific DNA by three-dimensional routes aided by supercoiling. *Embo J* 22, 1410-8.
- (50) Yildiz, A., Tomishige, M., Vale, R. D., and Selvin, P. R. (2004) Kinesin walks hand-over-hand. *Science* 303, 676-8.
- (51) Yildiz, A., Forkey, J. N., McKinney, S. A., Ha, T., Goldman, Y. E., and Selvin, P. R. (2003) Myosin V walks hand-over-hand: single fluorophore imaging with 1.5-nm localization. *Science* 300, 2061-5.
- (52) Verdine, G. L., and Bruner, S. D. (1997) How do DNA repair proteins locate damaged bases in the genome? *Chem Biol* 4, 329-34.
- (53) Sun, J., Viadiu, H., Aggarwal, A. K., and Weinstein, H. (2003) Energetic and structural considerations for the mechanism of protein sliding along DNA in the nonspecific BamHI-DNA complex. *Biophys J* 84, 3317-25.
- (54) Gowers, D. M., Wilson, G. G., and Halford, S. E. (2005) Measurement of the contributions of 1D and 3D pathways to the translocation of a protein along DNA. *Proc Natl Acad Sci U S A* 102, 15883-8.
- (55) Blainey, P. C., van Oijen, A. M., Banerjee, A., Verdine, G. L., and Xie, X. S. (2006) A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA. *Proc Natl Acad Sci U S A* 103, 5752-7.
- (56) Hamilton, R. W., and Lloyd, R. S. (1989) Modulation of the DNA scanning activity of the *Micrococcus luteus* UV endonuclease. *J Biol Chem* 264, 17422-7.
- (57) Nickell, C., and Lloyd, R. S. (1991) Mutations in endonuclease V that affect both protein-protein association and target site location. *Biochemistry* 30, 8638-48.
- (58) Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* 98, 10037-41.

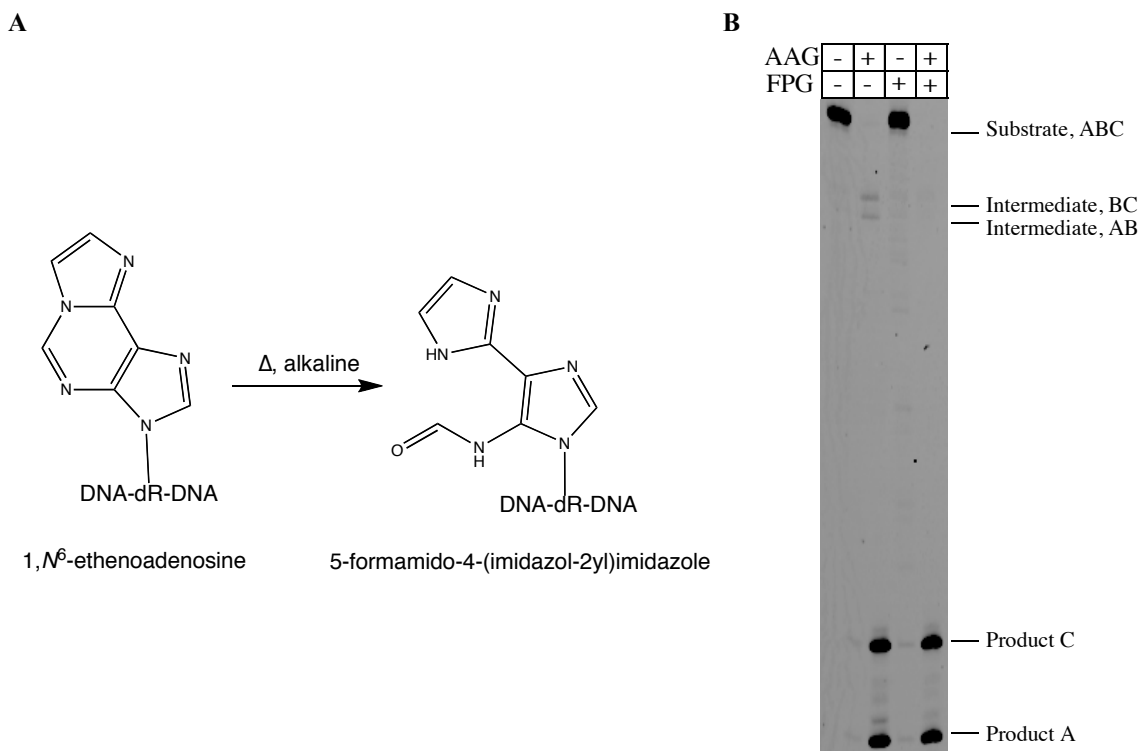
## Appendix A

Additional data figures and accompanying discussion to support Chapter 2

### Supporting Results and Discussion

#### Evaluation of $\epsilon$ A sites in the processivity substrate

It has been shown that  $\epsilon$ A nucleotides are susceptible to ring-opening reactions, especially under basic conditions, and that the resulting ring-opened products are substrates for formamidopyrimidine DNA glycosylase (FPG), but not for AAG (*I*). Therefore, we incubated the 47-mer processivity substrate with either AAG, FPG, or both enzymes and allowed sufficient time for the AAG reaction to reach completion (>10 half-lives). Samples were processed in sodium hydroxide according to our standard protocol and run out on a 20% polyacrylamide denaturing gel (Figure A-1B). Even at very long incubation times some of the intermediates AB and BC remain resistant to AAG-catalyzed excision. However, when FPG is present both sites are completely processed to products A and C. This suggests that ~2% of each  $\epsilon$ A site has undergone a ring-opening reaction to produce a lesion that can no longer be recognized by AAG, but is a substrate of FPG.

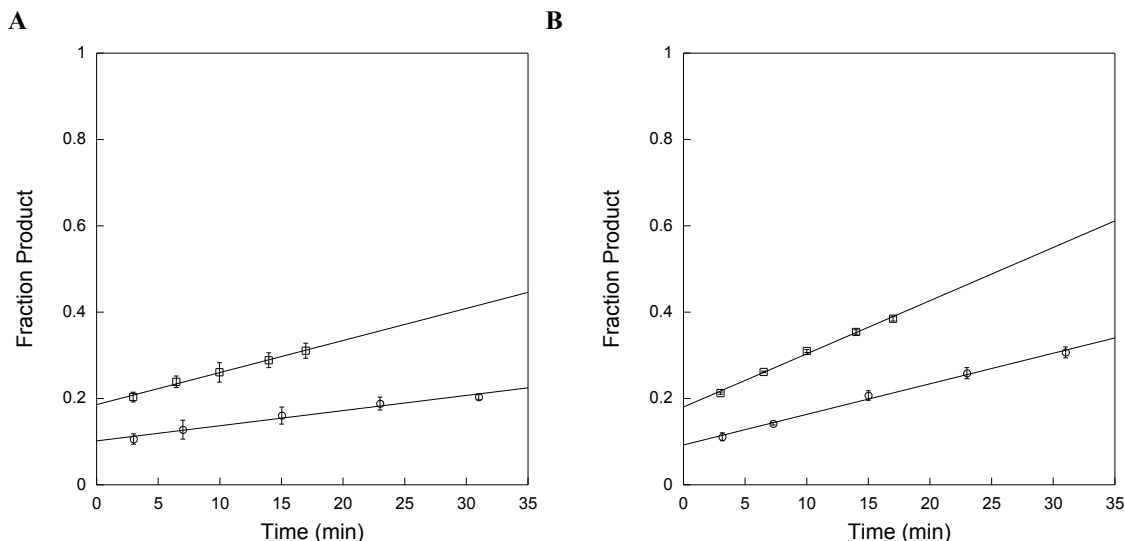


**Figure A-1: A small fraction of  $\epsilon$ A lesions undergo ring opening during synthesis, deprotection, and/or purification**

(A) Proposed ring opening reaction that  $\epsilon$ A undergoes at alkaline pH (1). (B) Denaturing PAGE analysis of 47-mer dual-lesion oligonucleotide duplex incubated for 10 hours with either AAG or FPG or both enzymes.

### Determination of the concentration of active AAG

To determine steady-state kinetic parameters, it is important to know the concentration of active enzyme. Therefore, we performed active site titrations as described in the Methods section of the text and representative data is shown in Figure A-2. We observe that most of the recombinant full-length (91%) and truncated (98%) AAG molecules are active so that only minor corrections needed to be made to the concentrations that were determined by absorbance at 280 nm.



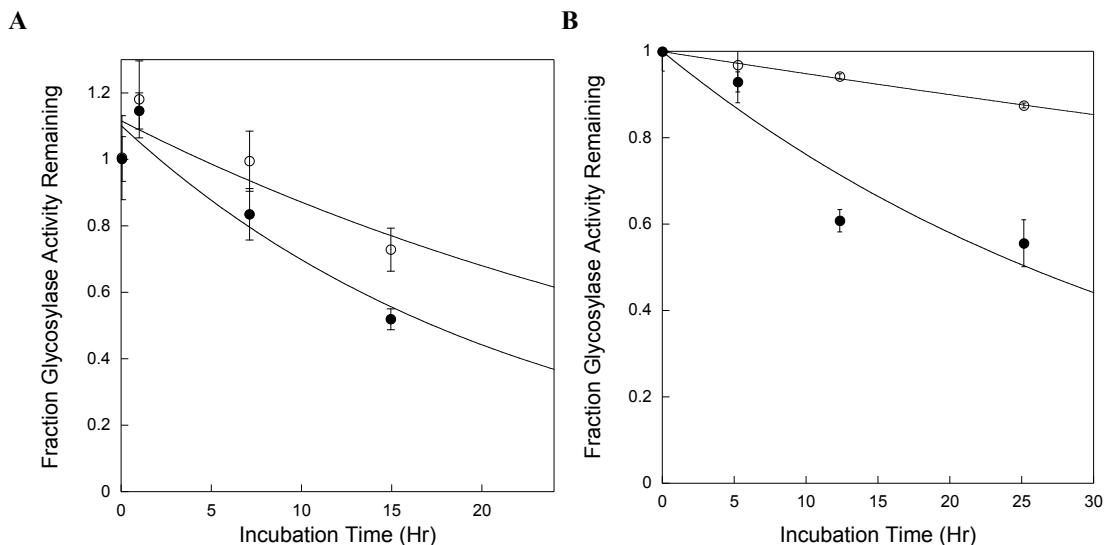
**Figure A-2: Active site titration of AAG**

Since excision of hypoxanthine from deoxyinosine-containing DNA shows burst kinetics, we were able to quantify the amount of active enzyme by extrapolating the steady state rate back to the origin. The results for both truncated (A) and full-length (B) AAG are shown. The concentration of DNA substrate was 1  $\mu$ M and the concentration of glycosylase was either 100 nM (open circles) or 200 nM (open squares). Reactions were performed in triplicate, as described in the Methods section, and the average and standard deviation are shown). In this experiment, the fraction of active truncated protein was  $0.98 \pm 0.04$  and the fraction of active full-length protein was  $0.91 \pm 0.01$ . In the experiments described in the text, the  $k_{cat}$  values were calculated from the concentration of active enzyme.

### Stability of AAG under steady-state assay conditions

We performed pre-incubation controls in which either full-length or truncated AAG was incubated under the standard reaction conditions for long amounts of time prior to the addition of substrate (Figure A-3). Since we found that protein stability is dependent upon the concentration of protein, we were careful to carry out these incubations at the same concentration as was used in the processivity assays (20 nM). Protein incubated at higher concentrations showed much greater stability under the same conditions. The results show that both enzymes slowly lose activity on the hour timescale

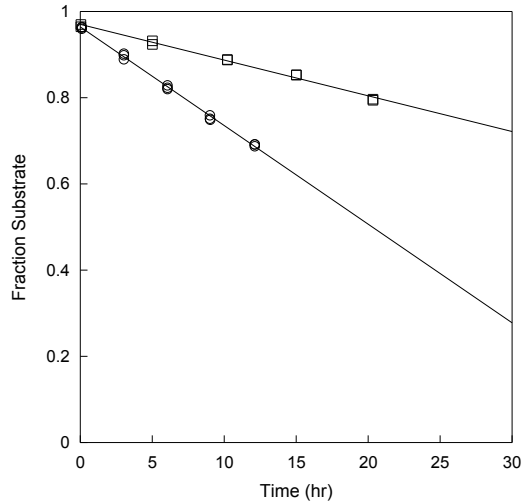
(half-lives ranged from 15-140 hours) and that they lose activity more slowly when the ionic strength is increased from 50 to 300 mM.



**Figure A-3: Stability of AAG in the absence of DNA**

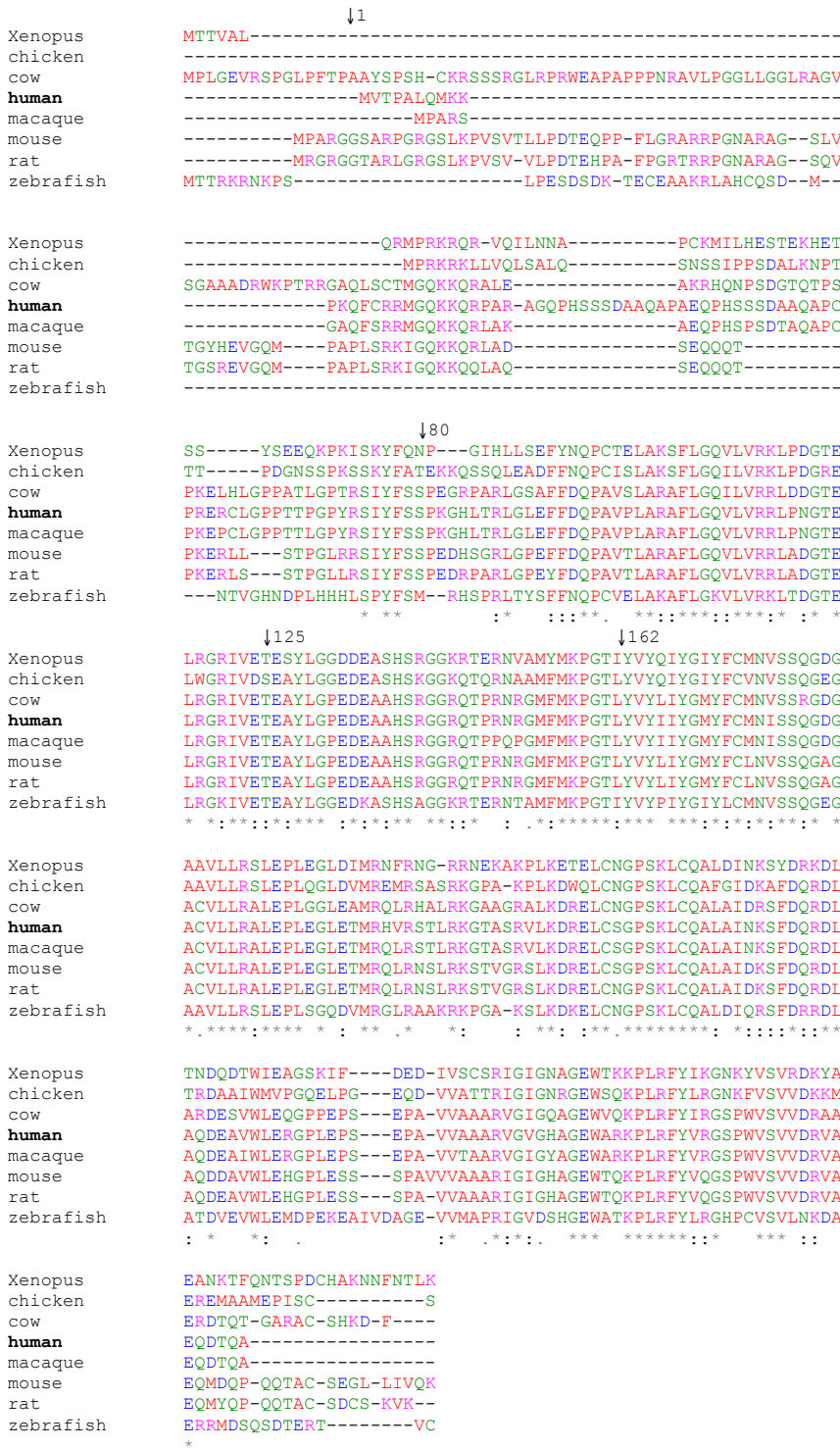
Stability of Full-length (A) or truncated (B) AAG in the absence of DNA. 20 nM AAG was incubated under our standard reaction conditions for periods of up to 24 hours (50 mM NaMES, pH 6.1, 10% glycerol, 0.1 mg/mL BSA, 1 mM DTT, 1 mM EDTA with additional NaCl to obtain the desired ionic strength). Glycosylase activity was measured under multiple-turnover conditions with saturating deoxyinosine-containing DNA and the resulting velocity was normalized by dividing by the rate constant without preincubation. AAG showed slightly greater stability at higher ionic strength (300 mM; open circles) than at lower ionic strength (50 mM; closed circles). These data were fit by single exponential curves with inactivation rate constants of 0.05 and 0.03  $\text{hr}^{-1}$  for full length protein at 50 and 300 mM ionic strength and of 0.03 and 0.005  $\text{hr}^{-1}$  for  $\Delta 80$  AAG at 50 and 300 mM ionic strength.

We expected that saturation with DNA substrate would similarly increase the stability of AAG under multiple turnover conditions. If AAG were to lose activity over time during steady-state kinetic assays, then we would expect the slope of the reaction progress curve to decrease since the velocity at any point in time is proportional to the amount of active enzyme. Even for very long time courses, there was no evidence of curvature in the reaction progress curve, suggesting that both truncated and full-length AAG are significantly more stable when bound to DNA (Figure A-4).



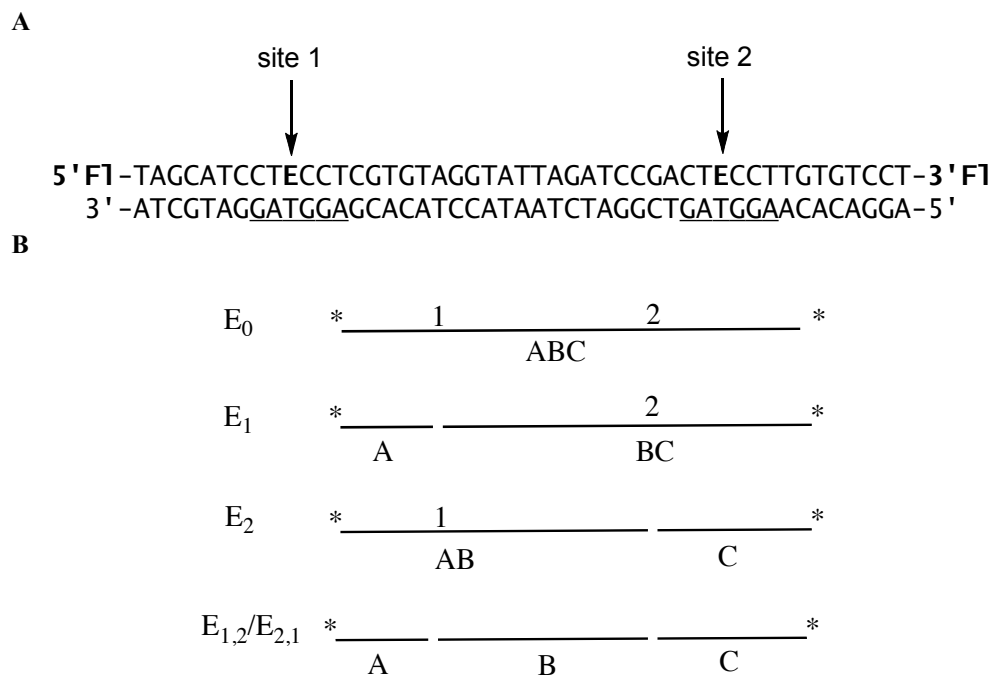
**Figure A-4: Steady state reaction progress curve for hydrolysis of the dual lesion substrate**

Steady state reaction progress curve for hydrolysis of the dual lesion substrate by full-length (open squares) and truncated AAG (open circles) demonstrates that these enzymes are stable over the time course of our assays. Reaction conditions were 37 °C in a buffer composed of 50 mM NaMES pH 6.1, 1 mM DTT, 1 mM EDTA, 0.1 mg/mL BSA, with ionic strength adjusted to 50 mM. Substrate concentration was 2  $\mu$ M and AAG concentration was 20 nM. Under these conditions the reaction rates remain linear for greater than 20% of the reaction and for 12-24 hours, indicating little or no loss of activity on this time scale. This result suggests that the presence of saturating DNA substrate provides significant stabilization of both full-length and truncated AAG. Comparable results were obtained at higher ionic strength and at pH 7.5 (data not shown).



**Figure A-5: Multiple sequence alignment of vertebrate AAG amino acid sequences**  
Generated using T-coffee; (2). Similar (:) and identical (\*) amino acids are indicated below the alignment. The numbers and arrows above the alignment indicate the numbering for the human amino acid sequence (start of the human sequence, position 80, the conserved active site base glutamate 125, and the conserved DNA intercalating tyrosine 162).





**Figure A-6: Processivity substrate and definition of DNA fragments produced by base excision, alkaline hydrolysis, and denaturing PAGE.**  
 Only the lesion-containing strand is shown and the fluorescein label is indicated (\*).

**Derivation of processivity equations**

The analysis of processivity using an externally labeled, two-lesion-containing oligonucleotide substrate is essentially the same as has been previously described for restriction endonucleases (3-6). For endonuclease activity, the theoretical maximum fraction processive has been assumed to be 0.5, because the double-strand break creates two DNA molecules and the enzyme can only remain bound to one. In the more general case of enzymes that do not create double strand breaks, a theoretical limit of 1.0 is expected.

Under multiple turnover conditions ( $[S] \gg [E]$ ), the small fraction of substrate that is bound by an enzyme can be identified once a base lesion is excised. There are 4 possible outcomes for AAG•DNA complex that result in at least one excision event (see Figure A-6). These are denoted in equation 1, in which subscript 1 and 2 correspond to action at site 1 or 2, respectively.



The concentration of each DNA fragment resulting from AAG-catalyzed base excision and alkaline cleavage is simply the sum of the different enzymatic events giving rise to that particular fragment (Equations 2-6).

$$A = E_1 + E_{1,2} + E_{2,1} \tag{Eq. 2}$$

$$B = E_{1,2} + E_{2,1} \tag{Eq. 3}$$

$$C = E_2 + E_{1,2} + E_{2,1} \tag{Eq. 4}$$

$$BC = E_1 \tag{Eq. 5}$$

$$AB = E_2 \tag{Eq. 6}$$

The fraction of processive events ( $F_p$ ) is given by the ratio of the processive events ( $E_{1,2}$  and  $E_{2,1}$ ) divided by the total number of events (Equation 7). Substitution using equations 2-6 yields an expression based upon the concentration of the different DNA products formed (Equation 8).

$$F_p = (E_{1,2} + E_{2,1}) / (E_1 + E_2 + E_{1,2} + E_{2,1}) \tag{Eq. 7}$$

$$F_p = [B] / ([B] + [AB] + [BC]) \tag{Eq. 8}$$

Fragment B is the key species in determining the fraction processive, because it is only formed by processive events (Equation 3). However, with the external labeling strategy that we have employed this species is not directly observed and it is necessary to express the concentration of this species as a function of the labeled species. Accordingly, Equations 2-6 can be rearranged to yield two solutions for the concentration of fragment B (Equations 9 & 10). These two independent determinations of the concentration of fragment B can be averaged, as depicted in Equation 11.

$$[B] = [A] - [BC] \tag{Eq. 9}$$

$$[B] = [C] - [AB] \tag{Eq. 10}$$

$$[B] = \frac{1}{2} ([A] + [C] - [AB] - [BC]) \tag{Eq. 11}$$

Substitution of Equation 9 and 10 into Equation 8 gives two different expressions that involve the concentrations of labeled species. The two expressions reflect the two possible pathways for processive excision. Equation 12 describes the event  $E_{1,2}$ , whereby AAG acts first at site 1 and subsequently at site 2 ( $F_p^{(1-2)} = E_{1,2} / (E_{1,2} + E_1)$ ). Equation 13

describes the converse event  $E_{2,1}$ , in which AAG acts first at site 2 and subsequently at site 1 ( $F_p^{(2 \rightarrow 1)} = E_{2,1}/(E_{2,1} + E_2)$ ).

$$F_p^{(1 \rightarrow 2)} = ([A]-[BC])/([A]+[AB]) \quad (\text{Eq. 12})$$

$$F_p^{(2 \rightarrow 1)} = ([C]-[AB])/([C]+[BC]) \quad (\text{Eq. 13})$$

These expressions are true at any time under the initial rates conditions (i.e., after any burst is complete, but before excessive product has built up). However, at early times the signal to noise is expected to be poor and at later times the product distribution could be affected by rebinding events. Therefore, it is preferable to use the initial velocities for formation of each species, rather than the discrete concentrations at a single time point (e.g.,  $V_A$  = initial rate for formation of product A). The corresponding directional processivity equations are given below (Equations 14 & 15).

$$F_p^{(1 \rightarrow 2)} = (V_A - V_{BC})/(V_A + V_{AB}) \quad (\text{Eq. 14})$$

$$F_p^{(2 \rightarrow 1)} = (V_C - V_{AB})/(V_C + V_{BC}) \quad (\text{Eq. 15})$$

For AAG, both equations give identical values for  $F_p$  under a wide range of conditions, indicating that the processivity of AAG is the same regardless of which lesion is excised first (See Table A1). Therefore, it is convenient to substitute Equation 11 into Equation 8 to obtain the average processivity that is independent of pathway (Equation 16). This reduces the error when comparing independent replicates of given experimental conditions by minimizing the impact of small errors in fluorescence intensity calculations and background subtractions.

$$F_p = (V_A + V_C - V_{AB} - V_{BC})/(V_A + V_C + V_{AB} + V_{BC}) \quad (\text{Eq. 16})$$

### **Evaluation of the directional processivity**

Since both ends of the DNA substrate are labeled, it is possible to distinguish the two possible pathways for processive action (i.e., evaluate whether there is a different outcome depending upon which site is encountered first). Representative data for multiple turnover excision by both full-length and truncated AAG at pH 7.5 are summarized in Table A1. The average and standard deviation for 5-8 independent determinations is given. The directional processivity was calculated from Equations 14 and 15, using the initial rates calculated for each of the product and intermediate fragments. The average processivity was calculated as described in the main text using

Equation 16. The results show that AAG exhibits very similar processivity regardless of which lesion it excises first.

**Table A-1: Evaluation of the directional processivity of alkyladenine DNA glycosylase**

Enzyme	50 mM ionic strength			150 mM ionic strength		
	$F_p^{1 \rightarrow 2}$	$F_p^{2 \rightarrow 1}$	$F_p^{avg}$	$F_p^{1 \rightarrow 2}$	$F_p^{2 \rightarrow 1}$	$F_p^{avg}$
<b><math>\Delta 80</math> AAG</b>	$0.91 \pm 0.06$	$0.90 \pm 0.06$	$0.90 \pm 0.06$	$0.29 \pm 0.02$	$0.30 \pm 0.05$	$0.30 \pm 0.03$
<b>FL AAG</b>	$0.86 \pm 0.04$	$0.90 \pm 0.04$	$0.88 \pm 0.04$	$0.42 \pm 0.10$	$0.48 \pm 0.12$	$0.44 \pm 0.09$

## References

- (1) Speina, E., Ciesla, J. M., Wojcik, J., Bajek, M., Kusmierk, J. T., and Tudek, B. (2001) The pyrimidine ring-opened derivative of 1,N6-ethenoadenine is excised from DNA by the Escherichia coli Fpg and Nth proteins. *J Biol Chem* 276, 21821-7.
- (2) Poirot, O., O'Toole, E., and Notredame, C. (2003) Tcoffee@igs: A web server for computing, evaluating and combining multiple sequence alignments. *Nucleic Acids Res* 31, 3503-6.
- (3) Terry, B. J., Jack, W. E., and Modrich, P. (1985) Facilitated diffusion during catalysis by EcoRI endonuclease. Nonspecific interactions in EcoRI catalysis. *J Biol Chem* 260, 13130-7.
- (4) Stanford, N. P., Szczelkun, M. D., Marko, J. F., and Halford, S. E. (2000) One- and three-dimensional pathways for proteins to reach specific DNA sites. *Embo J* 19, 6546-57.
- (5) Kampmann, M. (2004) Obstacle bypass in protein motion along DNA by two-dimensional rather than one-dimensional sliding. *J Biol Chem* 279, 38715-20.
- (6) Gowers, D. M., Wilson, G. G., and Halford, S. E. (2005) Measurement of the contributions of 1D and 3D pathways to the translocation of a protein along DNA. *Proc Natl Acad Sci U S A* 102, 15883-8.

## Chapter 3

### **Hopping Enables a DNA Repair Glycosylase to Search Both Strands and Bypass a Bound Protein<sup>1</sup>**

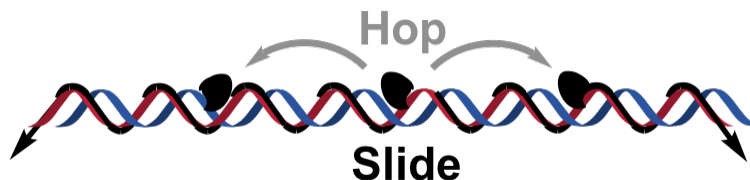
The human base excision DNA repair pathway repairs ~10,000 lesions per cell per day (1). This is a daunting task because these relatively rare lesions must be located from among ~12,000,000,000 normal nucleotides in the genome. Almost a dozen different human DNA repair glycosylases continuously and independently search the genome for a wide variety of oxidized or alkylated bases. Once a damaged nucleotide has been located, the glycosylase catalyzes the hydrolysis of the N-glycosidic bond to release the damaged base and create an abasic site. This abasic site is further processed to restore the correct DNA sequence using the opposing nucleotide as a template. There is considerable in vitro evidence that glycosylases use thermally-driven linear diffusion to efficiently search for sites of damage, whereby the enzyme diffuses along DNA in a non-directional manner, searching many adjacent sites within a single binding event (2-8). The biological importance of linear diffusion has been confirmed by the findings that mutants of T4 pyrimidine dimer glycosylase and EcoRV endonuclease that are deficient in linear diffusion have decreased activity in vivo (9-11).

The task of locating specific sites within the genome is central to DNA repair and to many other nuclear processes such as DNA replication and transcription (12-14). Two distinct mechanisms are recognized for diffusion along DNA and they are commonly referred to as sliding and hopping (13-22). As illustrated in Scheme 1, sliding involves continuous contact between the protein and the DNA backbone so that transfer occurs between linearly contiguous sites on the same strand. This implies that sliding follows a helical path, and there is experimental evidence of this for several proteins, including

---

<sup>1</sup>Reproduced with permission from Hedglin, M., and O'Brien, P.J. (2010) Hopping Enables a DNA Repair Glycosylase to Search Both Strands and Bypass a Bound Protein, *ACS Chemical Biology* 5, 427 – 436. Copyright 2010 American Chemical Society

DNA glycosylases (3, 15, 23-25). In the alternative mode of translocation, referred to as hopping, a bound protein microscopically dissociates to a point at which it is still very close to the originally bound site and will with high probability re-associate to the same or nearby site on either DNA strand (12, 22). This pathway has been experimentally observed for several proteins (16-22), including *E. coli* uracil DNA glycosylase (7).



**Scheme 1:** Sliding and hopping are distinct mechanisms of linear diffusion

We investigated the mechanism by which alkyladenine DNA glycosylase (AAG), a human DNA glycosylase responsible for the repair of a diverse set of alkylated and deaminated purines, locates damaged nucleotides. The results provide strong evidence for hopping by this single domain protein and demonstrate that AAG simultaneously searches both strands of DNA. Remarkably, a tightly bound protein serves as only a partial block of AAG diffusion. We suggest that hopping plays important roles in the search for DNA damage, allowing greater distances to be covered in a given time, both strands to be sampled, and bound proteins to be circumvented.

## Materials and Methods

### Proteins

Full-length and truncated recombinant human AAG were purified and the concentration of active AAG was determined by burst analysis as previously described (5). EcoRI was expressed and purified as previously described (40) and the concentration of active dimer was determined by burst analysis (See Appendix B). T4 DNA Ligase was from New England Biolabs.

### Oligonucleotides

DNA substrates were synthesized by Integrated DNA Technologies or the Keck Center at Yale University and purified by denaturing PAGE as previously described (5). For the GNRAIF and PolyT hairpin substrates, the DNA was synthesized in fragments and each DNA fragment was purified. DNA fragments for each respective substrate were then

annealed in annealing buffer (10 mM NaMES, pH 6.5, 50 mM NaCl) and ligated for 24 hrs at 16 °C using T4 DNA Ligase in the recommended NEB Ligase Buffer. Ligation reactions were quenched after 24 hrs with 20 mM EDTA and complete ligation products were purified. The concentration of single-stranded DNA was determined from the absorbance at 260 nm using the calculated extinction coefficients, and the concentration of duplex (hairpin) DNA was determined from the extinction coefficient at 495 nm for fluorescein ( $\epsilon_{495} = 7.5 \times 10^5 \text{ M}^{-1} \text{ cm}^{-1}$ ).

### **Glycosylase Activity Assay**

Reactions were carried out at 37 °C in 50 mM NaMES, pH 6.1, 1 mM EDTA, 1 mM DTT, 10% glycerol, 0.1 mg mL<sup>-1</sup> BSA and the ionic strength was adjusted with NaCl. Reactions were initiated by adding enzyme (2 – 7.5 nM final concentration) to obtain a reaction volume of 50 – 100  $\mu\text{L}$  that contained 75 - 200 nM fluorescein-labeled DNA. Aliquots were withdrawn at various times and quenched with 2 volumes of 0.3 M NaOH or 0.5 volumes of 0.6 M NaOH to yield a final concentration of 0.2 M. Samples were heated to 70 °C for 15 minutes, formamide was added to 65%, and the DNA fragments were resolved on 14% (w/v) polyacrylamide gels containing 8 M urea as previously described (5). Gels were scanned with a Typhoon Trio<sup>+</sup> fluorescence imager (GE Healthcare) to detect fluorescein (excitation at 488 nm and emission with 520BP40 filter). The resulting fluorescent signal was quantified using ImageQuant TL and corrected for background signal. The intensity of each DNA band was converted into a fraction by dividing its intensity by the sum of the intensities for all of the DNA species present.

### **Multiple Turnover Kinetics**

Steady state kinetics for the dual lesion substrates were measured with 100-fold excess of substrate (200 nM) over enzyme (2 nM) as previously described (5). The initial rates were calculated from the first 10-15% of the reaction and were linear in all cases. Values for both  $k_{\text{cat}}$  and  $F_p$  were calculated and the ionic strength dependence of both were fit with a cooperative model (5). The value of  $F_p$  was calculated according to equation 3, in which  $V_i$  and  $V_p$  are the initial rates for the formation of intermediates (retaining an  $\epsilon\text{A}$  lesion) and products (both  $\epsilon\text{A}$  excised), respectively. The processivity equation takes into account the fact that a single excision event gives rise to one product and one



intermediate. Approximately 2% of  $\epsilon$ A nucleotides are damaged during synthesis, deprotection and gel purification, and therefore the maximal processivity value that is expected is 0.92 (5). The minimal value that could be observed is  $\sim$ 0.05 for a completely distributive mechanism, that is attributed to rebinding of AAG. Therefore the effective range of experimental  $F_p$  values is 0.87 (0.92-0.05). To facilitate the comparison with the theoretical  $F_p$  values based upon the sliding model, we have corrected all of the observed processivity values (eq 4).

$$F_{p, obs} = (V_p - V_i) / (V_i + V_p) \quad (3)$$

$$F_p = (F_{p, obs} - 0.05) / 0.87 \quad (4)$$

### **Pulse-Chase Processivity Assay**

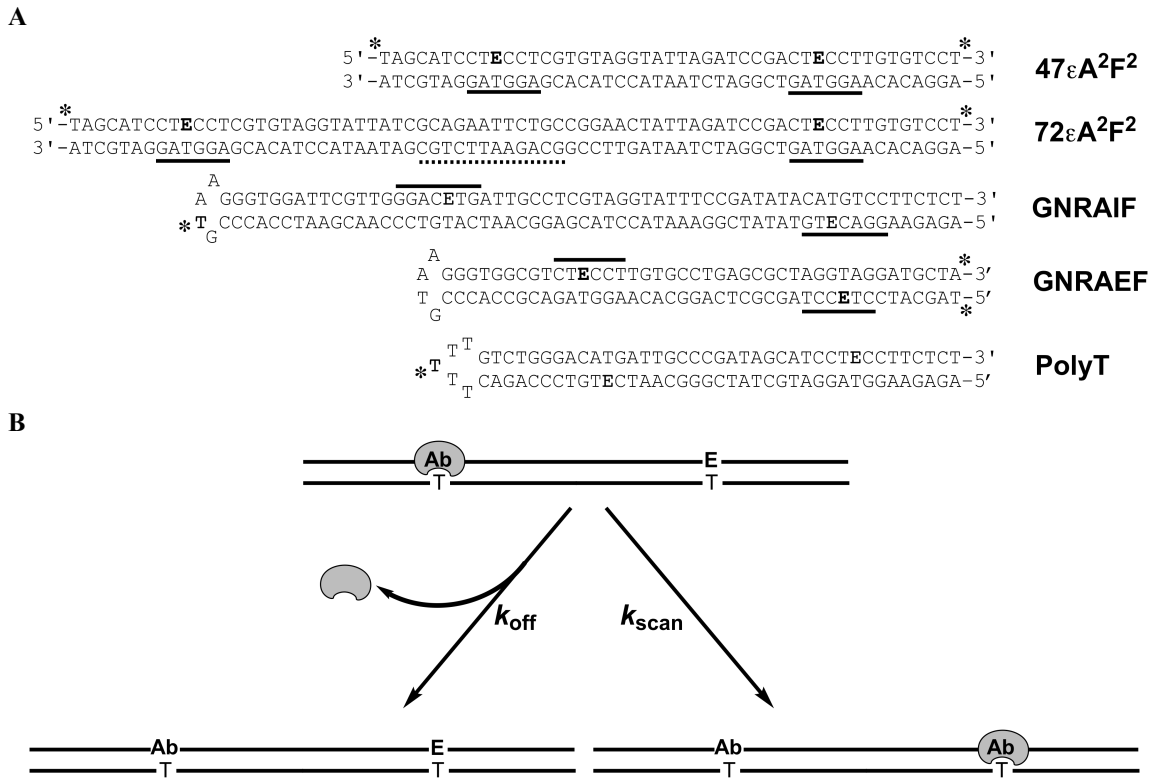
To test whether AAG can bypass a protein roadblock, we incubated labeled substrate ( $72\epsilon A^2 F^2$ ) with either 0 or 1.5-fold excess EcoRI endonuclease as indicated in Figure 3-6. The final concentration of  $72\epsilon A^2 F^2$  was 75nM, AAG was 7.5 nM, EcoRI dimer was 110 nM, and  $72\epsilon A^2$  chase DNA was 2  $\mu$ M. As a control to establish that occupancy of the recognition sequence by EcoRI was responsible for the decrease in processivity, a reaction was carried out in which  $72\epsilon A^2 F^2$  was incubated in the absence of EcoRI. After dilution and the addition of AAG to the labeled substrate, unlabeled substrate ( $72\epsilon A^2$ ) that had been pre-incubated with EcoRI was added. Glycosylase activity was measured as described above.

The amplitude of the burst phase for each DNA fragment was determined by the Y-intercept of a linear fit (slope = 0) to the data after the burst phase was complete (25-50 min). The burst amplitude of the substrate matched the expected burst size, and the burst rate constant was the same within error as the previously determined single turnover rate constant for excision of  $\epsilon$ A (5). The value of  $F_p$  was calculated from the concentration of products and intermediates that were formed during the burst phase, by substituting the initial rates in equation 1 with the corresponding concentrations of products ( $p$ ) and intermediates ( $i$ ) to give equation 5. The range of  $F_p$  values in the pulse-chase assay with 10-fold excess of DNA are identical to the range in the multiple turnover assay with 100-fold excess of DNA. The upper limit is set to 0.92 due to the damaged  $\epsilon$ A sites, whereas the lower limit of 0.05 is due to the 1% probability that two AAG molecules will bind to the same DNA molecule.

$$F_{p, obs} = ([p] - [i]) / ([p] + [i]) \quad (5)$$

## Results and Discussion

We previously described a processivity assay that allows the transfer of a glycosylase between two sites on a DNA molecule to be monitored by measuring the correlated cleavage events for a substrate containing two sites of damage (5). Base excision catalyzed by AAG at one of the two sites results in an abasic product and AAG subsequently partitions between action at the second site and dissociation into solution (Figure 3-1). The fraction of processive events ( $F_p$ ) is determined by alkaline hydrolysis of abasic sites and initial rates for single and double excision events. These experiments demonstrated that AAG locates sites of damage via linear diffusion and that electrostatic



**Figure 3-1: Processivity assays to determine the mechanism of linear diffusion by a DNA repair glycosylase**

(A) Sequences of oligonucleotides that were employed in this study. All DNA duplexes contained two  $\epsilon$ A lesions (E) on the same or opposing strands and one or two fluorescein labels (asterisk). The local sequence context for the  $\epsilon$ A lesions are marked by a solid line if they are identical for a given substrate and the 10 bp palindrome that contains a central EcoRI recognition sequence (GAATTC) is marked by a dashed line. (B) Processivity assays follow events subsequent to an initial base excision event, effectively measuring partitioning between dissociation and correlated excision at the nearby lesion site. Substrates were designed so that AAG randomly binds to and excises either of the two  $\epsilon$ A lesions to create an abasic site (Ab). AAG release is irreversible under both multiple turnover and pulse-chase conditions, because the excess substrate prevents rebinding to a released intermediate.

interactions are critical to maintaining contact with DNA. Although it is not absolutely required, the amino terminus of AAG contributes to the processivity of the enzyme (5). These experiments provided insight into the mechanism by which AAG locates sites of damage, but they did not distinguish whether AAG uses hopping, sliding, or a combination of these two modes of diffusion (Scheme 1). The physical and chemical mechanism of linear diffusion is critical to understanding biological processes in the nucleus, such as transcription and DNA repair. In the present work we have varied the distance between the two lesions to investigate the effective searching distance and to gain additional insight into the mechanism of diffusion along DNA.

### **Evaluating the Sliding-Only Model**

There is evidence that many proteins use a sliding mode to diffuse along DNA (3, 15, 16, 24). Sliding relies on simple diffusion and does not have a directional bias. Therefore, a bound protein has equal probability of moving to either adjacent position and the mean position of the sliding protein will not vary from its starting point (14, 19, 20, 26). However, the distribution around the mean will broaden with time, with a translocation of  $n$  base pairs from the mean requiring  $n^2$  single base pair steps (designated as  $N$ ). The probability of a protein reaching a position  $n$  base pairs away ( $P_n$ ) is given by equation 1, in which  $P_1$  is the probability that the enzyme will move one step along the DNA without dissociating,  $k_s$  is the rate constant for sliding by one base pair and  $k_{off}$  is the rate constant for dissociation into solution (13, 19, 20). If protein translocation is monitored by enzymatic activity at the target site, the efficiency of recognition ( $E$ ) must also be considered (eq 2), because some binding encounters might be unproductive (7).

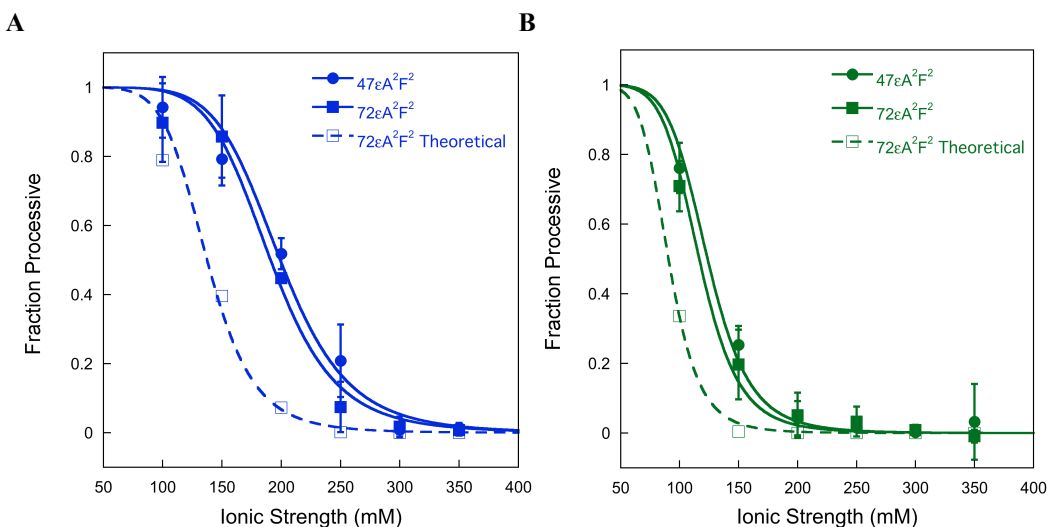
$$P_n = P_1^N = [k_s/(k_s + k_{off})]^N \quad (1)$$

$$F_p = E \times P_A^N = E[k_s/(k_s + k_{off})]^N \quad (2)$$

The processivity of AAG is decreased by increasing the ionic strength or by truncation of the poorly conserved amino terminus (5). The theoretical model derived for a sliding-only model (eq 2) applies at any ionic strength, and changes in  $F_p$  must be attributed to the ionic strength dependence of one or more of these terms. Experimental and theoretical work suggests that  $k_s$  is insensitive to changes in ionic strength (21, 27). Therefore, the observed ionic strength dependence of  $F_p$  could be due to changes in  $E$  or  $k_{off}$ . It is known that the efficiency of recognition of  $\epsilon A$  is relatively insensitive to ionic

strength (28), whereas the rate of dissociation of AAG from DNA is strongly dependent upon the ionic strength (5, 29). These considerations suggest that the rate of dissociation controls the processivity of AAG across a wide range of salt concentrations. Truncation of the amino terminus ( $\Delta 80$ ) appears to decrease the processivity by increasing the rate of dissociation (5).

To test whether AAG moves predominantly by sliding, we compared the relative processivity on substrates that have a different number of base pairs (bp) between the two sites (Figure 3-1). The microscopic rate constants  $k_s$  and  $k_{off}$  and the value of  $E$  are identical for the two substrates. Therefore, the partition function ( $k_s/(k_s+k_{off})$ ) is a constant for a given condition and can be used to predict how the  $F_p$  value would change with sliding distance according to the sliding-only model. Multiple turnover processivity assays for substrates with lesions separated by 25 bp ( $47\epsilon A^2F^2$ ) and 50 bp ( $72\epsilon A^2F^2$ ) were performed for both full-length and  $\Delta 80$  AAG and the results are presented in Figure 3-2. The data for the 25 bp separation are in excellent agreement with previous data obtained at 10-fold higher substrate concentration (5). We used the purely sliding model to predict the theoretical effect of increasing the searching distance from 25 to 50 bp (Figure 3-2,



**Figure 3-2: Ionic strength dependence for the processivity of AAG is inconsistent with a purely sliding model**

Multiple turnover processivity assays were performed with either full-length (A) or  $\Delta 80$  AAG (B), using substrates that contained two  $\epsilon A$  lesions separated by 25 bp ( $47\epsilon A^2F^2$ , ●) or 50 bp ( $72\epsilon A^2F^2$ , ■). The fraction processive was calculated as described in the Methods and the ionic strength dependence was fit by a cooperative model with 7 inhibitory sodium ions (5). Each data point reflects the mean value from at least two independent experiments and the error bars indicate one standard deviation ( $n \geq 4$ ). The theoretical processivity for the 72mer (□, dashed lines) was calculated from the data for the 47mer, using the sliding model described in the text (See Appendix B).

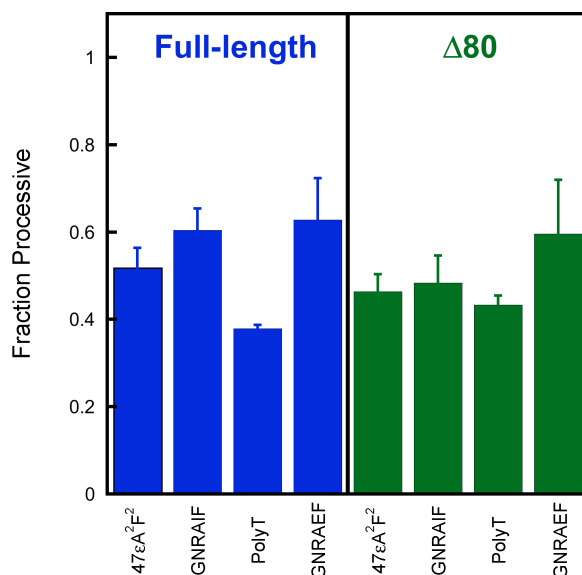
dashed lines; see Appendix B). It is notable that for both enzymes and all ionic strength conditions there is no significant difference between the two substrates, in contrast to the large difference that is predicted from a sliding-only model (22). This suggests that AAG employs a mode of diffusion other than/or in addition to sliding, such as hopping. Below we describe assays designed to directly detect hopping.

### **Direct Evidence for Hopping**

If the search for DNA damage is restricted to sliding, then only one strand of DNA could be searched during an individual DNA binding encounter. However, if the protein were capable of microscopic dissociation and re-association (i.e., hopping), then the protein could switch between searching one or the other strand. Therefore, we tested to what extent AAG acts processively on substrates that contained lesions on opposing strands. A key feature of this experimental design is that the two strands are connected via a hairpin so that action at both sites is intramolecular and strand switching can be monitored. To evaluate the possible effects of the DNA ends, we varied the position of the fluorescein labels (internal or external) and compared T<sub>5</sub> (polyT) and GNRA hairpin structures (Figure 3-1A). GNRA hairpins are known to form compact three dimensional structures in DNA or RNA (30). In contrast, a polyT loop is expected to be more flexible.

Under conditions of low ionic strength, full-length AAG was highly processive on the GNRAEF substrate that had lesions on opposing strands and the processivity was indistinguishable from that observed for the substrates with lesions on the same strand (see Appendix B). This demonstrates that AAG is able to hop between strands at least once so that both lesions are excised prior to macroscopic dissociation and rebinding to another DNA molecule. However, a decrease in the processivity would be difficult to detect under these conditions because the residence time of AAG on the DNA is very long. The maximum sensitivity would be observed when 50% of the binding events are processive (i.e.,  $F_p = 0.5$ ). These conditions are obtained at an ionic strength of 200 mM for full-length and 115 mM for the truncated form of AAG (Figure 3-2). Therefore, multiple turnover processivity assays were carried out under these conditions for each of the hairpin substrates with both forms of AAG and the results are summarized in Figure 3-3.

In all cases, the processivity of AAG on a substrate with lesions on opposing strands was similar to the processivity on a substrate with lesions on the same strand. Similar values were obtained with different hairpins, GNRA versus polyT, and with different positions of the fluorescein label, suggesting that strand switching occurs at internal sites rather than at the hairpin or blunt ends. These results are indicative of frequent hopping events that allow AAG to simultaneously search both strands of an exposed DNA substrate. Furthermore, it is clear that the amino terminus of AAG is not required for strand switching (Figure 3-3).



**Figure 3-3: AAG searches both strands of DNA**

To test whether hopping contributes to the searching mechanism of AAG, we measured the processivity for substrates in which lesions are on the opposing strands and compared this to a substrate in which the lesions are on the same strand (47εA<sup>2</sup>F<sup>2</sup>). See Figure 3-1 for the DNA sequences. Multiple-turnover processivity assays were performed at an ionic strength of 200 mM for full-length AAG (blue) or 115 mM for Δ80 AAG (green). Each column represents the average of at least two independent experiments with error bars indicating one standard deviation from the mean ( $n \geq 4$ ).

Although our results seem to be at odds with recent reports that several glycosylases and other single domain enzymes rotate as they diffuse along DNA (15, 24), it should be noted that rotational diffusion (due to sliding along one strand) does not exclude hopping. Indeed, a recent report showed that *E. coli* uracil DNA glycosylase is capable of frequent hops (7). These differences between experiments conducted at the ensemble and single molecule level can be explained by a majority of steps being classified as sliding, with less frequent hopping steps. As glycosylases can only sample one nucleotide of a base pair, but lesions can occur in either strand, strand switching via

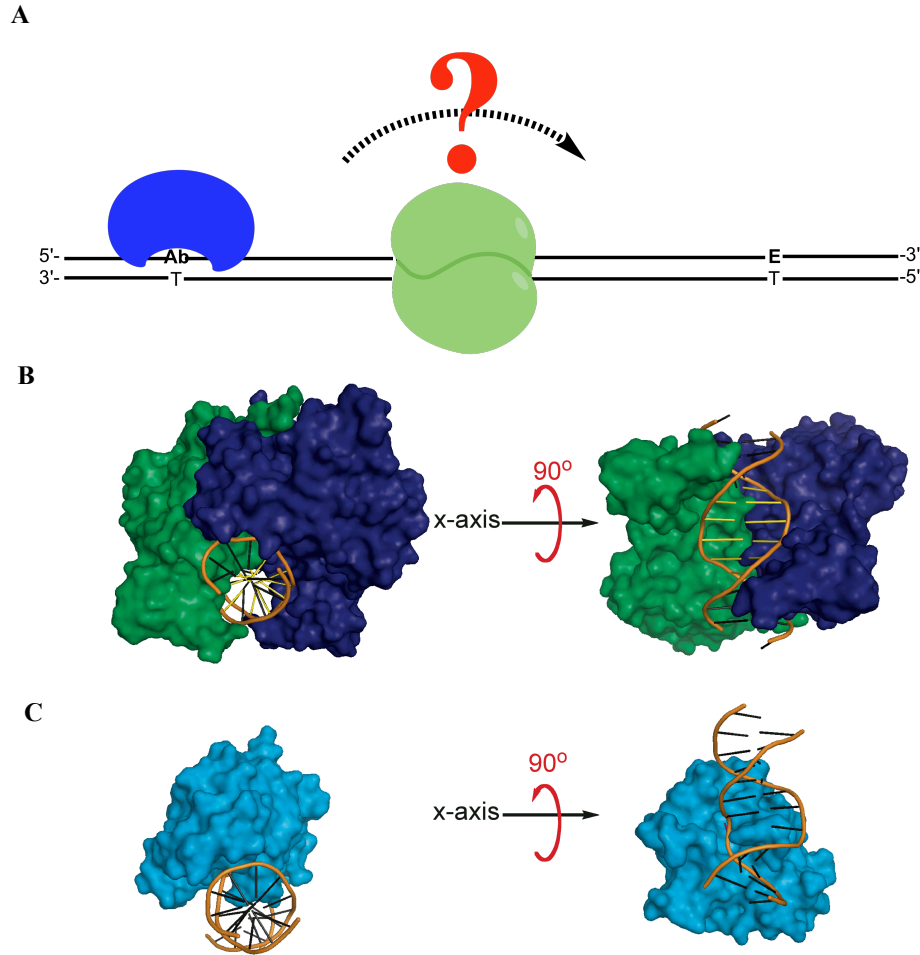
hopping events is expected to greatly increase the efficiency of the search for DNA damage.

### **AAG Can Bypass a Protein Roadblock**

A criticism of in vitro studies of linear diffusion of proteins on naked DNA is that DNA is expected to be only transiently accessible in the cell. DNA binding proteins bind densely to chromosomal DNA and those with high affinity can have lifetimes of hours. Previous studies have revealed that tightly bound proteins pose a barrier to sliding by endonuclease EcoRI (25), to translocation by mismatch repair proteins (31), and to transcription by RNA polymerase (32). In contrast, we hypothesized that microscopic dissociation followed by re-association might allow a protein such as AAG to hop past a tightly bound protein.

Several previous studies have employed EcoRI as a block, because it binds tightly as a homodimer and makes intimate contacts with both strands of the DNA (25, 31, 32). Therefore, we tested whether EcoRI blocks the diffusion of AAG (Figure 3-4). The 72 bp oligonucleotide that was employed is shorter than the persistence length of DNA, rendering intersegmental transfer unlikely. EcoRI has been reported to bend DNA by  $\sim 50^\circ$  (33), but this is insufficient to juxtapose the two lesions. AAG does not require  $Mg^{2+}$ , and this allowed us to use wildtype EcoRI since it binds with high affinity to its recognition site in the absence of  $Mg^{2+}$  (34). The use of wildtype EcoRI has an advantage over the use of an inactive mutant, because its presence at its recognition site can be directly quantified by the rapid DNA cleavage that occurs upon the addition of  $Mg^{2+}$ .

Control experiments confirmed that the 72εA<sup>2</sup>F<sup>2</sup> substrate, which contains a single EcoRI recognition site, could be bound completely at a 1:1 stoichiometry of EcoRI dimer to DNA under the glycosylase assay conditions (see Appendix B). Multiple turnover processivity experiments were performed in the presence of EcoRI with full-length AAG at both 100 and 200 mM ionic strength (Figure 3-5). The addition of one equivalent of EcoRI dimer decreased the processivity of AAG by  $\sim 50\%$  under both conditions. Addition of another equivalent of EcoRI did not have any further effect on the processivity of AAG, confirming that the DNA was saturated with EcoRI. Thus, a bound dimer of EcoRI appears to be only a partial block of the processive action of AAG.

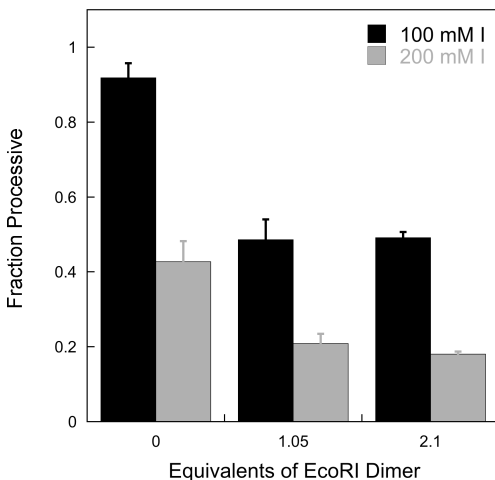


**Figure 3-4: Testing the effect of a protein roadblock on linear diffusion by AAG**

(A) The 72mer substrate is depicted with an EcoRI dimer (green) bound to the central recognition site, and with an AAG monomer (blue) bound to the abasic product from the first excision reaction. If AAG is able to bypass the tightly bound protein (dashed arrow), then processive excision of the second  $\epsilon$ A will be observed. (B) Structure of the EcoRI•DNA complex (41) is from the pdb (1ERI). The surfaces of the two EcoRI monomers are shown in blue and green and the DNA is depicted as a cartoon with the backbone in orange and the central EcoRI recognition sequence in yellow. (C) The structure of the complex of the catalytic domain of AAG bound to  $\epsilon$ A-DNA (42) is from the pdb (1F4R). Images were rendered with Pymol (<http://www.pymol.org>).

However, the long assay time required for multiple turnover assays poses a problem for the roadblock assay because EcoRI might diffuse away from its recognition site or dissociate into solution. This limitation was overcome with pulse-chase experiments that limit the processivity assay to a single encounter of AAG with DNA. In this approach AAG is incubated with labeled DNA substrate, then immediately chased with an excess of unlabeled substrate. AAG can dissociate directly, or excise either one





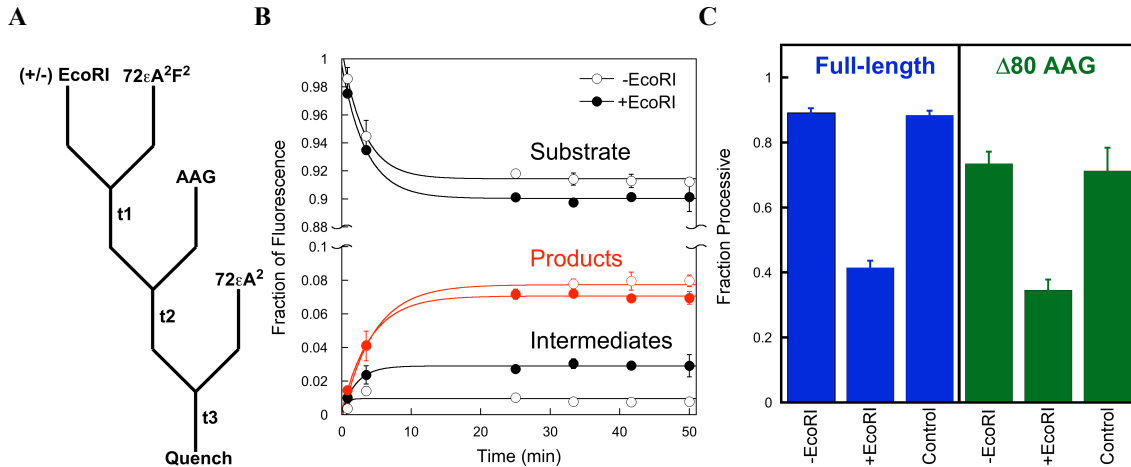
**Figure 3-5: Effect of bound EcoRI on the processivity of AAG**

Multiple turnover processivity assays were performed with 200 nM  $72\epsilon A^2F^2$ , 2 nM full-length AAG, and 0, 210, and 420 nM EcoRI dimer at both 100 and 200 mM ionic strength and the calculated processivity values are shown. The presence of the tightly bound EcoRI dimer reduces the processivity of AAG by ~50% at ionic strengths of 100 mM (black bars) and 200 mM (gray bars).

or two  $\epsilon A$  lesions before dissociating. Due to the presence of chase DNA, once AAG has dissociated it has a very low probability of rebinding a labeled substrate. We used an unlabeled DNA chase that is otherwise identical to the labeled substrate, and therefore it captures any AAG or EcoRI molecules that dissociate during the experiment.

To confirm the validity of this approach, we first performed the pulse chase experiment in the absence of EcoRI (ionic strength = 100 mM; Figure 3-6). A single turnover of substrate equal to the amount of AAG was completed within 25 minutes (Figure 3-6b), in agreement with the single-turnover rate constant of  $0.20 \text{ min}^{-1}$  (5). Quantification of the intermediates (single  $\epsilon A$  excised) and products (two  $\epsilon A$  excised) allowed the processivity to be calculated (eq 5). The resulting processivity from the pulse chase method is identical within error to the value determined for the multiple turnover processivity assay for both full-length ( $0.89 \pm 0.01$  versus  $0.90 \pm 0.1$ ) and  $\Delta 80$  AAG ( $0.74 \pm 0.04$  versus  $0.71 \pm 0.07$ ). Thus, the pulse-chase assay measures correlated events occurring within a single binding encounter.

We next evaluated the effectiveness of the EcoRI block with the pulse-chase assay. A greater amount of intermediates were released by AAG when EcoRI was bound, indicating a modest decrease in processivity (Figure 3-6b). This was quantified and the results for both full-length and  $\Delta 80$  AAG are shown in Figure 3-6c. EcoRI decreased the processivity of both full-length and  $\Delta 80$  AAG by approximately 50%. This is identical



**Figure 3-6: Pulse-chase processivity assays indicate that AAG can bypass a bound EcoRI dimer**  
 (A) The experimental design is depicted. Fluorescein-labeled substrate ( $72\epsilon A^2 F^2$ ) was incubated with or without EcoRI for 1 hour (t1), after which AAG was added. AAG was incubated for 40 seconds (t2), before mixing with excess unlabeled substrate ( $72\epsilon A^2$ ). Incubations continued for 50 minutes (t3) and aliquots were removed and analyzed by the gel-based glycosylase assay. The ratio of AAG to labeled substrate to unlabeled chase was 1:10:260. (B) A representative time course for full-length AAG was performed in duplicate at an ionic strength of 100 mM. Substrate depletion is fit to a single exponential ( $k_{obs} = k_{chem}$ ). The amplitude of  $\sim 10\%$  disappearance of substrate confirms that AAG was bound and excised at least one  $\epsilon A$  lesion prior to dissociation. No further glycosylase activity was observed up to 50 minutes, confirming that adequate chase was used. The build-up of products (red) and intermediates (black) were fit to single-exponentials solely to show trends. (C) The fraction processive was calculated from the burst amplitudes in panel B and from additional experiments for both full-length (blue) and  $\Delta 80$  AAG (green). Each column represents the average of two independent experiments with error bars representing one standard deviation from the mean ( $n \geq 4$ ). The column labeled “Control” is from reactions in which EcoRI was first bound to the unlabeled substrate instead of the labeled substrate. The final reaction conditions are identical to the +EcoRI reactions, but the endonuclease and glycosylase are bound on different DNA molecules. Therefore, the decrease in processivity is due to a direct block of AAG diffusion as opposed to an artifact of some other component of the EcoRI sample.

within error to the results from multiple turnover assays (Figure 3-5). We confirmed that the decrease in processivity requires EcoRI to be pre-bound to the labeled substrate, because control reactions in which EcoRI was first bound to the unlabeled chase did not show any decrease in processivity (Figure 3-6c). Additional controls ruled out the trivial possibility that partial blockage was due to incomplete saturation by EcoRI or dissociation of EcoRI on the time scale of the experiment. To address both points, we added  $Mg^{2+}$  after the AAG burst was over and measured cleavage by EcoRI. The results demonstrated that less than 2% of the EcoRI had dissociated (see Appendix B, Figure B9).

These experiments establish that EcoRI remains bound to the same DNA molecule for the entire assay. However, because EcoRI locates and leaves its recognition sequence predominantly by sliding, these data don't address transient excursions of

EcoRI away from its specific recognition site (31, 35-38). Such excursions have been observed on the same time scale ( $t_{1/2} \sim 40$  minutes) as our processivity assay (31). However, the short substrate that we employed is constrained by the ends of the DNA so that EcoRI cannot diffuse past the  $\epsilon$ A lesion to allow access by AAG. Although EcoRI spends most of its time bound to the specific recognition site, and bypass by AAG most likely occurs at this site, we cannot rule out the possibility that AAG and EcoRI transiently pass each other at an adjacent nonspecific site. Several pathways can be envisioned for the bypass of EcoRI by AAG (Figure 3-4). (i) Hopping by AAG may occur over a sufficient distance to allow AAG to dissociate and re-associate on the other side of the bound EcoRI. (ii) AAG may be able to diffuse over the surface of EcoRI, possibly facilitated by electrostatic interactions between the positively charged AAG and the negatively charged solution-exposed face of EcoRI. (iii) AAG may be able to navigate past EcoRI by multiple strand switching events, perhaps facilitated by breathing of EcoRI. Additional experiments will be required to distinguish these possibilities. Regardless of the exact pathway(s), it is notable that a tightly bound protein hinders, but does not completely block linear diffusion by AAG.

### **Implications**

Recent ensemble and single-molecule studies have focused on elucidating the contributions of each mode of translocation to the searching mechanism of DNA glycosylases to gain insight into how these enzymes rapidly and efficiently locate rare sites of DNA damage. Theoretical analysis suggests that the rate of target site location is optimized by a combination of sliding and hopping (13, 39). However, a mechanism comprised almost exclusively of rotation-coupled sliding over hundreds of base pairs of nonspecific DNA has been suggested for human OGG1, *E. coli* MutM, and *B. stearothermophilus* MutY (3, 15). By maintaining constant contact with the DNA backbone, redundant sliding over relatively short stretches of nonspecific DNA allows each nucleotide of the bound strand to be encountered multiple times and increases the probability that a lesion is recognized. However, reliance on sliding would slow down searches over longer distances and would allow only a single strand to be searched. Hopping can optimize the rate of target site location by allowing a sliding enzyme to escape from redundantly scanned stretches of DNA and access new sites (3, 13, 19). Our

results indicate that diffusion by AAG involves a significant contribution from hopping and are in good agreement with the conclusion that diffusion by *E. coli* uracil DNA glycosylase is dominated by hopping over long distances with local sliding contributing to damage recognition (7). The ability to hop to the opposing strand allows the rapid and essentially simultaneous search of both strands of a given segment of DNA. We suggest that this ability also enables AAG to bypass a tightly bound protein, which has important implications for the biological search for DNA damage.

### **Acknowledgment**

We thank P. Modrich for providing the expression plasmid for *E. coli* EcoRI endonuclease.

## References

1. Lindahl, T. (1993) Instability and decay of the primary structure of DNA, *Nature* 362, 709-715.
2. Bennett, S. E., Sanderson, R. J., and Mosbaugh, D. W. (1995) Processivity of Escherichia coli and rat liver mitochondrial uracil-DNA glycosylase is affected by NaCl concentration, *Biochemistry* 34, 6109-6119.
3. Blainey, P. C., van Oijen, A. M., Banerjee, A., Verdine, G. L., and Xie, X. S. (2006) A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA, *Proc. Natl. Acad. Sci. U. S. A.* 103, 5752-5757.
4. Francis, A. W., and David, S. S. (2003) Escherichia coli MutY and Fpg utilize a processive mechanism for target location, *Biochemistry* 42, 801-810.
5. Hedglin, M., and O'Brien, P. J. (2008) Human alkyladenine DNA glycosylase employs a processive search for DNA damage, *Biochemistry* 47, 11434-11445.
6. Higley, M., and Lloyd, R. S. (1993) Processivity of uracil DNA glycosylase, *Mutat. Res.* 294, 109-116.
7. Porecha, R. H., and Stivers, J. T. (2008) Uracil DNA glycosylase uses DNA hopping and short-range sliding to trap extrahelical uracils, *Proc. Natl. Acad. Sci. U. S. A.* 105, 10791-10796.
8. Sidorenko, V. S., and Zharkov, D. O. (2008) Correlated cleavage of damaged DNA by bacterial and human 8-oxoguanine-DNA glycosylases, *Biochemistry* 47, 8970-8976.
9. Dowd, D. R., and Lloyd, R. S. (1989) Biological consequences of a reduction in the non-target DNA scanning capacity of a DNA repair enzyme, *J. Mol. Biol.* 208, 701-707.
10. Dowd, D. R., and Lloyd, R. S. (1990) Biological significance of facilitated diffusion in protein-DNA interactions. Applications to T4 endonuclease V-initiated DNA repair, *J. Biol. Chem.* 265, 3424-3431.
11. Jeltsch, A., Wenz, C., Stahl, F., and Pingoud, A. (1996) Linear diffusion of the restriction endonuclease EcoRV on DNA is essential for the in vivo function of the enzyme, *EMBO J.* 15, 5104-5111.
12. Berg, O. G., Winter, R. B., and von Hippel, P. H. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory, *Biochemistry* 20, 6929-6948.

13. Halford, S. E., and Marko, J. F. (2004) How do site-specific DNA-binding proteins find their targets?, *Nucleic Acids Res.* *32*, 3040-3052.
14. Gorman, J., and Greene, E. C. (2008) Visualizing one-dimensional diffusion of proteins along DNA, *Nat. Struct. Mol. Biol.* *15*, 768-774.
15. Blainey, P. C., Luo, G., Kou, S. C., Mangel, W. F., Verdine, G. L., Bagchi, B., and Xie, X. S. (2009) Nonspecifically bound proteins spin while diffusing along DNA, *Nat. Struct. Mol. Biol.* *16*, 1224-1229.
16. Bonnet, I., Biebricher, A., Porte, P. L., Loverdo, C., Benichou, O., Voituriez, R., Escude, C., Wende, W., Pingoud, A., and Desbiolles, P. (2008) Sliding and jumping of single EcoRV restriction enzymes on non-cognate DNA, *Nucleic Acids Res.* *36*, 4118-4127.
17. Gowers, D. M., and Halford, S. E. (2003) Protein motion from non-specific to specific DNA by three-dimensional routes aided by supercoiling, *EMBO J.* *22*, 1410-1418.
18. Gowers, D. M., Wilson, G. G., and Halford, S. E. (2005) Measurement of the contributions of 1D and 3D pathways to the translocation of a protein along DNA, *Proc. Natl. Acad. Sci. U. S. A.* *102*, 15883-15888.
19. Halford, S. E. (2009) An end to 40 years of mistakes in DNA-protein association kinetics?, *Biochem. Soc. Trans.* *37*, 343-348.
20. Halford, S. E., and Szczelkun, M. D. (2002) How to get from A to B: strategies for analysing protein motion on DNA, *Eur. Biophys. J.* *31*, 257-267.
21. Komazin-Meredith, G., Mirchev, R., Golan, D. E., van Oijen, A. M., and Coen, D. M. (2008) Hopping of a processivity factor on DNA revealed by single-molecule assays of diffusion, *Proc. Natl. Acad. Sci. U. S. A.* *105*, 10721-10726.
22. Stanford, N. P., Szczelkun, M. D., Marko, J. F., and Halford, S. E. (2000) One- and three-dimensional pathways for proteins to reach specific DNA sites, *EMBO J.* *19*, 6546-6557.
23. Sakata-Sogawa, K., and Shimamoto, N. (2004) RNA polymerase can track a DNA groove during promoter search, *Proc. Natl. Acad. Sci. U. S. A.* *101*, 14731-14735.
24. Lin, Y., Zhao, T., Jian, X., Farooqui, Z., Qu, X., He, C., Dinner, A. R., and Scherer, N. F. (2009) Using the bias from flow to elucidate single DNA repair protein sliding and interactions with DNA, *Biophys. J.* *96*, 1911-1917.

25. Jeltsch, A., Alves, J., Wolfes, H., Maass, G., and Pingoud, A. (1994) Pausing of the restriction endonuclease EcoRI during linear diffusion on DNA, *Biochemistry* 33, 10215-10219.
26. Zharkov, D. O., and Grollman, A. P. (2005) The DNA trackwalkers: principles of lesion search and recognition by DNA glycosylases, *Mutat. Res.* 577, 24-54.
27. Kochaniak, A. B., Habuchi, S., Loparo, J. J., Chang, D. J., Cimprich, K. A., Walter, J. C., and van Oijen, A. M. (2009) Proliferating cell nuclear antigen uses two distinct modes to move along DNA, *J. Biol. Chem.* 284, 17700-17710.
28. Wolfe, A. E., and O'Brien, P. J. (2009) Kinetic mechanism for the flipping and excision of 1,N(6)-ethenoadenine by human alkyladenine DNA glycosylase, *Biochemistry* 48, 11357-11369.
29. Baldwin, M. R., and O'Brien, P. J. (2009) Human AP Endonuclease I Stimulates Multiple-Turnover Base Excision by Alkyladenine DNA Glycosylase, *Biochemistry* 48, 6022-6033.
30. Varani, G. (1995) Exceptionally stable nucleic acid hairpins, *Annu. Rev. Biophys. Biomol. Struct.* 24, 379-404.
31. Pluciennik, A., and Modrich, P. (2007) Protein roadblocks and helix discontinuities are barriers to the initiation of mismatch repair, *Proc. Natl. Acad. Sci. U. S. A.* 104, 12709-12713.
32. Pavco, P. A., and Steege, D. A. (1990) Elongation by Escherichia coli RNA polymerase is blocked in vitro by a site-specific DNA binding protein, *J. Biol. Chem.* 265, 9960-9969.
33. Thompson, J. F., and Landy, A. (1988) Empirical estimation of protein-induced DNA bending angles: applications to lambda site-specific recombination complexes, *Nucleic Acids Res.* 16, 9687-9705.
34. Terry, B. J., Jack, W. E., Rubin, R. A., and Modrich, P. (1983) Thermodynamic parameters governing interaction of EcoRI endonuclease with specific and nonspecific DNA sequences, *J. Biol. Chem.* 258, 9820-9825.
35. Jack, W. E., Terry, B. J., and Modrich, P. (1982) Involvement of outside DNA sequences in the major kinetic path by which EcoRI endonuclease locates and leaves its recognition sequence, *Proc. Natl. Acad. Sci. U. S. A.* 79, 4010-4014.
36. Pingoud, A., and Jeltsch, A. (2001) Structure and function of type II restriction endonucleases, *Nucleic Acids Res.* 29, 3705-3727.

37. Rau, D. C., and Sidorova, N. Y. (2009) Diffusion of the Restriction Nuclease EcoRI along DNA, *J. Mol. Biol.*
38. Wright, D. J., Jack, W. E., and Modrich, P. (1999) The kinetic mechanism of EcoRI endonuclease, *J. Biol. Chem.* 274, 31896-31902.
39. Givaty, O., and Levy, Y. (2009) Protein sliding along DNA: dynamics and structural characterization, *J. Mol. Biol.* 385, 1087-1097.
40. Cheng, S. C., Kim, R., King, K., Kim, S. H., and Modrich, P. (1984) Isolation of gram quantities of EcoRI restriction and modification enzymes from an overproducing strain, *J. Biol. Chem.* 259, 11571-11575.
41. Kim, Y. C., Grable, J. C., Love, R., Greene, P. J., and Rosenberg, J. M. (1990) Refinement of Eco RI endonuclease crystal structure: a revised protein chain tracing, *Science* 249, 1307-1309.
42. Lau, A. Y., Wyatt, M. D., Glassner, B. J., Samson, L. D., and Ellenberger, T. (2000) Molecular basis for discriminating between normal and damaged bases by the human alkyladenine glycosylase, AAG, *Proc. Natl. Acad. Sci. U. S. A.* 97, 13573-13578.



## Appendix B

Additional experimental methods, figures and accompanying discussion to support Chapter 3

### Supplemental Methods

#### Steady State Processivity Assay

A schematic of the substrates and expected products resulting from AAG-catalyzed base excision and alkaline abasic site hydrolysis is provided in Figure B-1. Only the labeled strands are shown, but the AAG substrates were all duplexes. Each labeled band was quantified separately, except in the case of  $72eA^2F^2$ , for which enzymatic action at site 1 and 2 gave rise to similar sized intermediate bands that were quantified together.

Enzymatic Action	DNA Substrate				
	$47eA^2F^2$	$72eA^2F^2$	GNRAEF	GNRAIF	PolyT
$E_0$	*—1—2*	*—1—2*	—1—* —2—*	—1—* —2—*	*—1— *—2—
$E_1$	*—2*	*—2*	—1—* —2—*	—1—* —2—*	*—2—
$E_2$	*—1—*	*—1—*	—1—* —2—*	—1—* —2—*	*—1—
$E_{1,2}/E_{2,1}$	*—* *—*	*—* *—*	—1—* —2—*	—1—* —2—*	*—* *—*

**Figure B-1: Diagram of oligonucleotide substrates and products**

The expected products after AAG-catalyzed base excision and alkaline cleavage of the abasic product for each substrate shown in Figure 3-1A. Only the lesion-containing strand is shown, and the asterisk indicates a fluorescein label.

We have fit the ionic strength dependence of  $k_{cat}$  with a cooperative model as previously described (7). The ionic strength dependent value of  $k_{cat}$  is given by Eq B1, in which  $I$  is the ionic strength,  $n$  is the number of cation binding sites,  $K_a$  is the average affinity constant for binding of the cation, and  $k_{max}$  is the maximal single turnover rate constant. The fits shown correspond to a value of  $n=7$ . Similarly, the ionic strength dependence of the processivity factor is given by eq B2 (7). The observed values of  $F_p$  were corrected as described in the manuscript, so that they range from 0–1.

$$k_{cat} = k_{max}(I^n / ((K_a)^n + I^n)) \quad (B1)$$

$$F_p = 1 - (I^n / (K_a^n + I^n)) \quad (B2)$$

### **Native Gel Electrophoresis**

Oligonucleotide substrates were annealed in 10 mM NaMES pH 6.5 with 50 mM NaCl by heating to 95 °C and slowly cooling to room temperature. Annealed substrates were subsequently diluted into standard pH 6.1 glycosylase reaction buffer with an ionic strength of either 115 or 200 mM. Samples were run on 10% native PAGE (19:1 acrylamide:bis-acrylamide, 0.5 X TBE) at constant voltage of 10 v/cm. Fluorescein-labeled DNA was visualized as described for sequencing gels.

### **Determining the Concentration of Active EcoRI Endonuclease**

A pulse-chase procedure was used to quantify the amount of EcoRI (8). Specific EcoRI•DNA complexes were formed by incubating EcoRI and 72εA<sup>2</sup>F<sup>2</sup> at 37°C in cleavage buffer without Mg<sup>2+</sup> (100 mM Tris, pH 8.5, 0.05 mg/mL BSA), with 1 mM EDTA and ionic strength adjusted to 100 mM I (by the addition of NaCl) for one hour. DNA cleavage was initiated by mixing samples (20 μL) with an equal volume of the same buffer containing 11 mM MgCl<sub>2</sub> and 7.6 mM unlabeled substrate (72εA<sup>2</sup>). The final concentration of labeled substrate (72εA<sup>2</sup>F<sup>2</sup>) was 200 nM and that of the unlabeled substrate (72εA<sup>2</sup>) was 3.8 μM (19-fold excess), yielding a total DNA concentration of 4 μM. The final concentration of EcoRI monomer was either 40 or 80 nM. Aliquots were withdrawn at various times and quenched with 3 volumes of a quench solution containing 23 mM EDTA in 95% Formamide. Quenching prior to the addition of MgCl<sub>2</sub> demonstrated that no cleavage occurred during the pre-incubation. Labeled DNA fragments were resolved on 14 % (w/v) polyacrylamide sequencing gels containing 8 M urea. Gels were imaged and DNA bands were quantified as discussed in the main text.

Under these conditions, cleavage by EcoRI exhibits burst kinetics with rapid phosphodiester bond cleavage followed by a slower rate of product dissociation. The steady state portion of the curve was extrapolated to obtain the burst amplitude (Y-intercept).

### **Confirmation That EcoRI Remains Bound For the Duration of the Processivity Assay.**

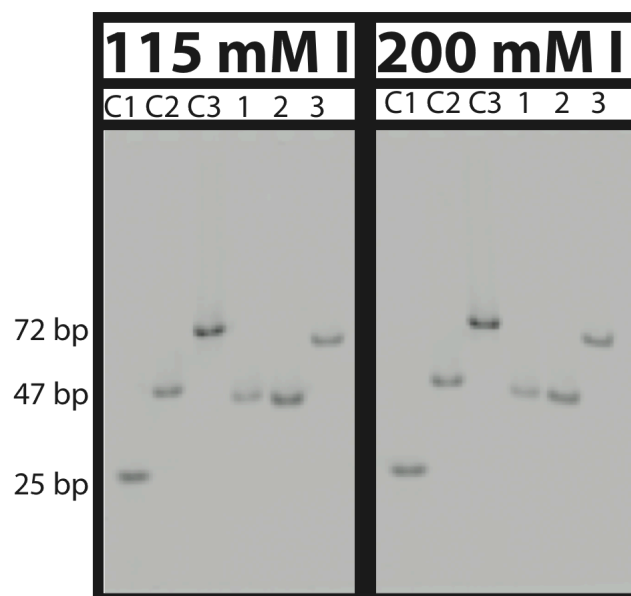
The extent of EcoRI dissociation during the pulse-chase processivity assays was determined using the same samples that were used for measuring glycosylase activity. After completion of the pulse-chase assay (50 minutes), an aliquot of each reaction

solution was removed and diluted 4-fold into 1.33X EcoRI cleavage buffer containing additional unlabeled substrate ( $72\epsilon A^2$ ) and 6.9 mM  $MgCl_2$  (5.2 mM final concentration). The final concentration of EcoRI and labeled substrate ( $72\epsilon A^2 F^2$ ) was 28 and 19 nM respectively and the total concentration of unlabeled substrate was 1.4  $\mu M$  (50-fold excess over EcoRI). Aliquots were withdrawn at various times and quenched with 0.5 volumes of 100 mM EDTA (50 mM final concentration) allowing the steady-state cleavage of the  $72\epsilon A^2 F^2$  substrate on the labeled strand to be monitored over time. Under these conditions, any labeled substrate that was still bound by EcoRI after completion of the pulse-chase assay will be rapidly cleaved in the first turnover of EcoRI. After the first turnover of EcoRI, labeled substrate will be cleaved at a much slower steady-state rate due to the high excess of unlabeled substrate. Extrapolation of the steady state rate for the cleavage of labeled substrate back to the origin yields the fraction of labeled substrate cleaved during the burst phase. The fraction of bound EcoRI bound was obtained by dividing the observed burst amplitude by the maximal amount of DNA that could be cleaved by EcoRI ( $0.90 \pm 0.01$ ; Figure B-7).

## **Supplemental Results and Discussion**

### **Characterization of Processivity Substrates**

The purity of all of the fluorescein-labeled substrates was evaluated by denaturing gel electrophoresis as previously described (7). In all cases, the purity was greater than 98% (data not shown). We used native gel electrophoresis to confirm that the designed hairpin structures formed unique, homogenous duplexes of the expected size (Figure B-2).

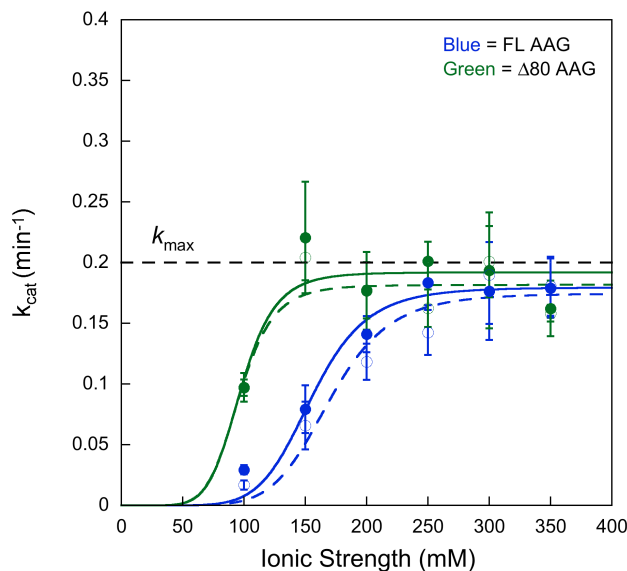


**Figure B-2: Native gel electrophoresis of oligonucleotides used in this study**

Annealed substrates were diluted into glycosylase reaction buffer in the absence of AAG at either 115 or 200 mM ionic strength as indicated. Oligonucleotide duplexes of 25 (C1), 47 (C2), and 72 (C3) base pairs were run on the same gel as size standards. The hairpins (1 = GNRAEF, 2 = PolyT, 3 = GNRAIF) migrate at approximately the expected size of 42, 39, and 60 bp. This confirms that each exists exclusively as a hairpin, rather than a homodimer that would be approximately twice the size.

### Comparison of $k_{cat}$ with Processivity Substrates

The steady state rate constant ( $k_{cat}$ ) was calculated for the experiments described in Figure 3-2 of Chapter 3 and the results are shown in Figure B-3. Under conditions of low ionic strength, the rate-limiting step is dissociation of AAG from the DNA (7, 9). The ionic strength dependence for  $72\epsilon A^2 F^2$  is essentially identical to that of  $47\epsilon A^2 F^2$ , indicating that the addition of 25 bp of nonspecific DNA does not change the rate at which AAG dissociates from DNA.



**Figure B-3: Ionic strength dependence for the multiple turnover excision of  $\epsilon A$**

The data for  $47\epsilon A^2F^2$  (open circles, dashed line) and  $72\epsilon A^2F^2$  (closed circles, solid line) are essentially the same for  $\Delta 80$  (green) and full-length (blue) AAG. The average of two to five independent measurements is shown, and the error bars indicate one standard deviation from the mean. A cooperative model was fit to the data, whereby multiple sodium ions accelerate the rate of reaction (See Supporting Methods). The plateau indicates a change in rate-limiting step from DNA dissociation at low ionic strength to *N*-glycosidic bond cleavage at high ionic strength (7). The black dashed line indicates the single turnover rate constant for excision of  $\epsilon A$  that was measured independently under these conditions.

### Calculating the Theoretical Processivity Predicted by a Sliding Model

If AAG moves along DNA by sliding, then the probability of reaching a site  $n$  base pairs away ( $P_n$ ) is described by eq B3.  $F_p$  is the fraction of processive events, and  $N$  is equal to  $n^2$ . The microscopic rate constants for sliding by one nucleotide ( $k_s$ ) and for dissociation into solution ( $k_{off}$ ) are expected to be the same for AAG bound to different DNA molecules. Similarly, the efficiency with which AAG recognizes a lesion ( $E$ ) is expected to be the same for substrates with identical sequence context. To evaluate the purely sliding model, the processivity observed for the shorter substrate ( $n=25$  bp,  $47\epsilon A^2F^2$ ) was used to predict the processivity for the longer substrate ( $n=50$  bp,  $72\epsilon A^2F^2$ ), assuming an ionic strength independent value of  $E = 1$ .

$$F_p = EP_n = E[k_s/(k_s + k_{off})]^N \quad (B3)$$

$$F_{p,25} = [k_s/(k_s + k_{off})]^{625} \quad (B4)$$

$$[k_s/(k_s + k_{off})] = (F_{p,25})^{1/625} \quad (B5)$$

$$F_{p,50} \text{ (theoretical)} = [k_s/(k_s+k_{off})]^{2500} \quad (B6)$$

$$F_{p,50} \text{ (theoretical)} = (F_{p,25})^4 \quad (B7)$$

The calculated values for full-length AAG are given in Table B-1. Analogous calculations gave the predicted processivity of  $\Delta 80$  AAG (see Figure 3-2 in the Chapter 3).

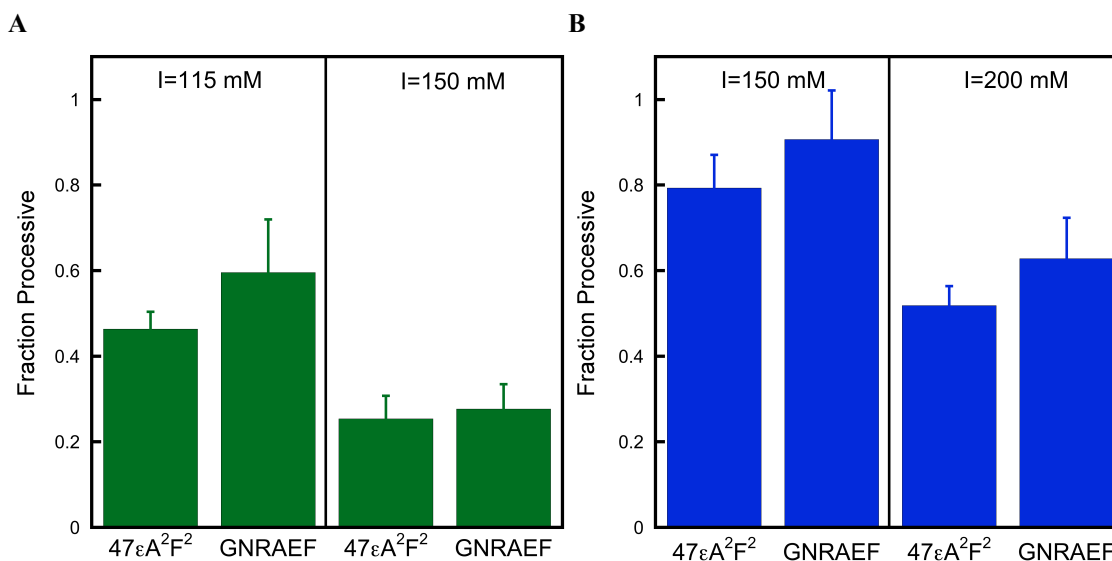
**Table B-1: Comparison of theoretical and measured processivity values to evaluate whether AAG employs a sliding mechanism<sup>1</sup>**

l (mM)	F <sub>P,25</sub> Experimental	F <sub>P,50</sub> Theoretical	F <sub>P,50</sub> Experimental
100	0.94	0.79	0.90
150	0.79	0.40	0.86
200	0.52	0.07	0.45
250	0.21	0	0.07
300	0	0	0.02
350	0	0	0

<sup>1</sup>The experimental processivity values for full-length AAG were taken from Figure 3-2 in the Chapter 3 for  $47\epsilon A^2 F^2$  (F<sub>P,25</sub>). An efficiency of recognition of 1 was assumed. Values close to 1 have been measured previously by pulse chase (10), and the data in Figure B-8 confirm that  $E$  is close to 1 under these experimental conditions. The theoretical value of F<sub>P</sub> for the 50 bp distance (F<sub>P,50</sub>) was calculated using equation B7 and juxtaposed against the experimentally measured values from Figure 3-2 in the Chapter 3.

### **Hopping Between Strands is Efficient Across a Range of Ionic Strength Conditions.**

Sliding and hopping are expected to have different ionic strength dependencies, with hopping becoming more frequent at higher ionic strength. Therefore, it was of interest to evaluate the ability of AAG to hop between strands as a function of ionic strength. We compared processivity when the lesions were on the same strand ( $47\epsilon A^2 F^2$ ) or opposing strands (GNRAEF) for both full-length and  $\Delta 80$  AAG under different ionic strength conditions to achieve a range of processivity values from 0.3-1 and the results are summarized in Figure B-4. Under the experimentally accessible range, AAG has equal probability of locating a lesion located on the same or opposing strand.

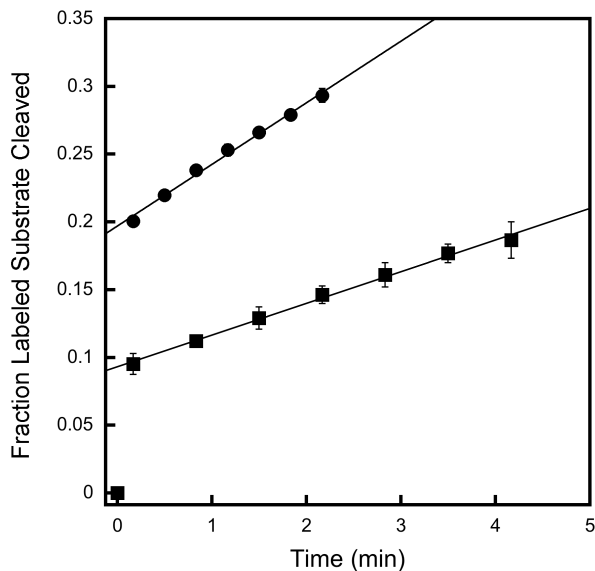


**Figure B-4: AAG hops frequently under a wide range of conditions**

Multiple turnover processivity assays were performed for full-length (A) and  $\Delta 80$  AAG (B) for the substrate with the lesions on the same strand ( $47\epsilon A^2 F^2$ ) and for the substrate with the lesions located on the opposing strands (GNRAEF). The processivity value has been corrected for the small fraction of damaged substrate as described in the Methods of Chapter 3, so that it will have a maximum possible value of 1 and a minimum value of 0.

### Active Site Titration of EcoRI Endonuclease to Determine the Concentration of Enzyme

EcoRI-catalyzed phosphodiester bond hydrolysis is extremely rapid under optimal conditions. Therefore, we have used a pulse-chase procedure similar to one that has been described previously by Wright and coworkers (11). This procedure takes advantage of the extremely slow dissociation of EcoRI from its recognition site. First a stoichiometric complex is formed of EcoRI on the labeled DNA containing a single EcoRI site ( $72\epsilon A^2 F^2$ ) in the absence of  $Mg^{2+}$ . Then,  $Mg^{2+}$  is added to initiate the reaction along with excess unlabeled DNA ( $72\epsilon A^2$ ) to slow down the steady state cleavage of labeled DNA that is not initially bound by EcoRI and the burst of cleaved product can be determined. A representative experiment is shown in Figure B-5. The concentration of the EcoRI stock was calculated from the stoichiometry of the burst, assuming that EcoRI is active as a dimer and the efficiency of cleavage is 1.0. The actual efficiency of cleavage that we measure for this substrate under the conditions of this assay is 0.9, so this titration underestimates the concentration of EcoRI by  $\sim 10\%$  (See Figures B-6 and B-7).



**Figure B-5: Active site titration of EcoRI**

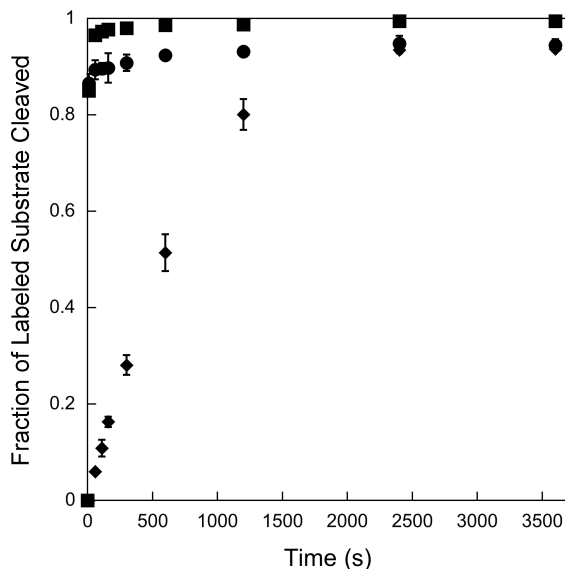
Rapid burst of EcoRI cleavage were monitored for 40 (●) or 80 nM (■) *EcoRI* monomer with 200 nM labeled substrate ( $72\epsilon A^2F^2$ ). To slow the steady state EcoRI cleavage reaction excess unlabeled substrate was included (See Supporting Methods for details). Linear fits to the steady state portion of the curve were used to obtain the  $k_{cat}$  value for EcoRI ( $k_{cat} = 4.7 \pm 0.2 \text{ min}^{-1}$ ) and the burst amplitude indicates the ratio of active dimer to labeled substrate.

### Saturation of $72\epsilon A^2F^2$ by EcoRI Endonuclease.

We used the endonuclease activity of EcoRI to test whether the EcoRI site was completely saturated. Figure B-6 shows endonuclease activity in the presence and absence of unlabeled chaser. Essentially all of the substrate is cleaved in the absence of chaser. However, when excess unlabeled DNA was included as a chaser, only 90% of the substrate is cleaved. Incomplete cleavage cannot be due to insufficient EcoRI, because the same limit of 90% cleaved is observed at several concentrations of EcoRI (Figure B-7). The incomplete cleavage of this substrate is most likely explained by the fact that the two cleavage events of EcoRI are not absolutely coupled and EcoRI occasionally dissociates after cleaving one of the two strands (11). In the presence of excess unlabeled DNA, EcoRI is unlikely to rebind to the nicked DNA. This property of EcoRI has been previously described and varies between 75 and 100 % for EcoRI sites in different sequence contexts (12-14). Furthermore, the release of singly nicked DNA is altered by reaction conditions and is increased by mutations in EcoRI (11-16). Our data suggest that EcoRI cleaves both strands prior to dissociation 80% of the time under the conditions that we have employed. We infer that the events that involve single strand cleavage (20% of total) will be evenly distributed between top strand and bottom strand cleavage. Thus,



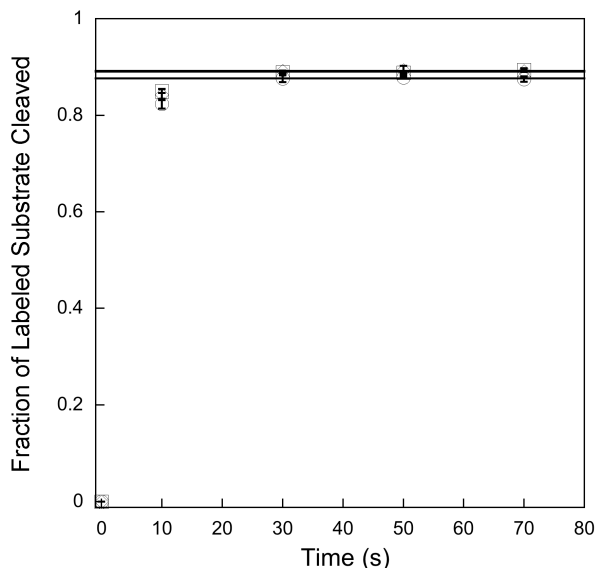
10% of the single strand cleavage events will release a nicked DNA molecule in which the fluorescein-labeled strand remains intact.



**Figure B-6: Endonuclease activity of EcoRI on the processivity substrate in the presence and absence of unlabeled chase**

Stoichiometric amounts of EcoRI was incubated with  $72\epsilon A^2 F^2$  at pH 6.1, 100 mM I in the absence of  $Mg^{2+}$ . Reactions were divided in half and to one half a 26-fold excess of unlabeled DNA  $72\epsilon A^2$  was added and the incubation was continued for 25 minutes. Samples were diluted into the EcoRI cleavage buffer (containing 5 mM  $MgCl_2$ ) to initiate strand cleavage. The fraction of substrate cleaved by EcoRI is indicated for reactions in the presence (●) and absence (■) of chase DNA. As a control, EcoRI was first incubated with the unlabeled DNA prior to the addition of the labeled DNA (◆). Reactions were performed in duplicate and the average is shown. Controls in which aliquots were quenched prior to the addition of  $MgCl_2$  confirmed that the substrate was intact and there was no endonuclease activity in the absence of  $Mg^{2+}$  (time = 0).

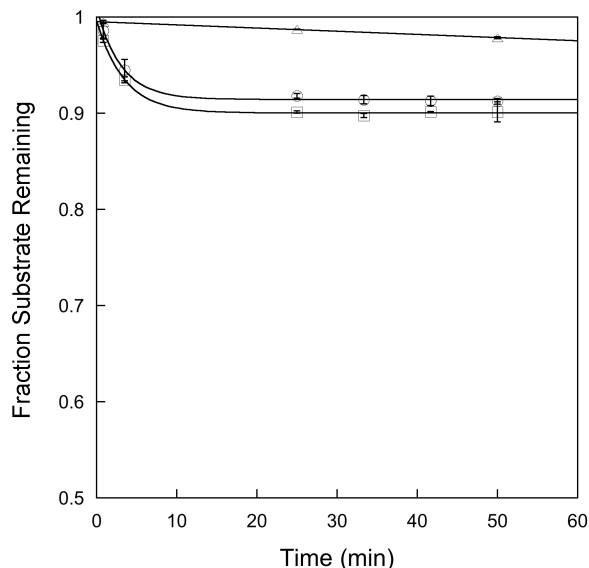
To confirm that the EcoRI site is fully saturated, we performed this experiment with up to 2-fold excess EcoRI. It is apparent that addition of extra EcoRI has no effect on the burst amplitude, strongly suggesting that EcoRI is fully saturating at the lowest concentration employed (Figure B-7). Therefore, we interpret a burst of EcoRI cleavage of 90% to indicate full occupancy of the recognition site of this oligonucleotide.



**Figure B-7. The complete burst of EcoRI cleavage confirms that EcoRI is saturating**  
 Complexes of EcoRI on fluorescein-labeled substrate ( $72\epsilon A^2F^2$ ) were formed in the absence of  $Mg^{2+}$  and single-turnover strand cleavage was monitored in the presence of a high excess of unlabeled substrate and  $Mg^{2+}$ . The extent of labeled strand cleavage during a single turnover for 210 ( $\circ$ ), 315 ( $\square$ ) and 420 ( $\diamond$ ) nM EcoRI dimer on 210 nM labeled substrate ( $72\epsilon A^2F^2$ ) were determined by extrapolating data points in the linear phase to zero time by linear least squares analysis. Reactions were performed in duplicate as described above and the average for each time point for each concentration of EcoRI is shown. Error bars indicate one standard deviation from the mean. Data points at zero time are from controls in which aliquots were quenched prior to the addition of  $MgCl_2$  to demonstrate that no cleavage occurs during EcoRI•DNA pre-incubation.

### Controls for the Pulse-Chase Processivity Assay.

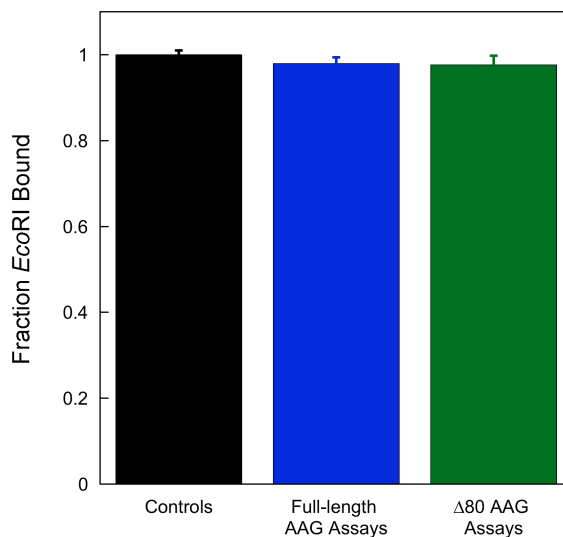
The pulse-chase processivity assay relies on an excess of unlabeled substrate to prevent rebinding of AAG to labeled processivity substrate. Therefore, we performed a simple control to evaluate whether sufficient unlabeled DNA was employed as a chase. AAG was initially bound to the unlabeled DNA, after which the labeled DNA was added (Figure B-8). In this case a slow steady state glycosylase activity was observed, with less than 1% of the labeled substrate being turned over in 25 minutes. Thus, the glycosylase activity that was observed in the pulse-chase experiment is attributed to molecules of AAG that were pre-bound to labeled DNA prior to the addition of unlabeled chase. This conclusion is further supported by the observation of the expected 10% burst indicates that every molecule of AAG excised at least a single  $\epsilon A$  lesion prior to dissociation (i.e.,  $E = 1$  under these conditions).



**Figure B-8: Turnover of  $72\epsilon A^2F^2$  by full-length AAG in the pulse-chase processivity assay**

This experiment was performed as described in Figure 3-5 of Chapter 3. Reactions in which AAG was added to labeled substrate in either the absence ( $\circ$ ) or presence ( $\square$ ) of 1.5-fold excess of EcoRI dimer gave bursts of 8-10% substrate consumed, in good agreement with the expected burst of 10%. Fitting the reaction progress curve with a single exponential gave an observed rate constant of  $0.3 \text{ min}^{-1}$ , in reasonable agreement with the expected single turnover rate constant of  $0.2 \text{ min}^{-1}$  for excision of  $\epsilon A$  by AAG (7). As a control, AAG was added to the unlabeled substrate prior to the addition of the labeled substrate ( $\triangle$ ).

The observation that EcoRI partially, but not completely, prevents diffusion of AAG raised the concern that EcoRI might be dissociating from DNA in the presence of AAG and that partial dissociation of EcoRI could be allowing apparent bypass by AAG. Therefore, we designed an experiment to directly measure the occupancy of EcoRI in the samples that were used to determine the processivity of AAG (Figure B-9). This was made possible by the fact the unlabeled DNA that we used as a chase is able to sequester both AAG and EcoRI that dissociates from the labeled processivity substrate. After the burst of AAG-catalyzed  $\epsilon A$  excision was complete, aliquots were removed and diluted into EcoRI cleavage buffer, and the initial burst of EcoRI cleavage was measured. The burst amplitude reveals the amount of EcoRI that was bound at the time that  $Mg^{2+}$  was added. These data indicate that EcoRI dissociates very slowly from DNA under the conditions of the processivity assay, and therefore the processivity of AAG was monitored in the presence of bound EcoRI.



**Figure B-9: EcoRI remains bound during the pulse-chase assay**

The occupancy of EcoRI after the burst of AAG-catalyzed excision of  $\epsilon$ A was determined by diluting samples into EcoRI cleavage buffer and monitoring the amount of labeled substrate cleaved by EcoRI in the initial burst. The fraction of EcoRI bound was corrected by dividing the observed fraction cleaved by the maximal amount of DNA that could be cleaved by EcoRI ( $0.90 \pm 0.01$ ; Figure B-7). The first column is the reference of the amount dissociated in the absence of AAG (Control). The next columns are taken from the processivity assays in which either full-length or  $\Delta 80$  AAG was present (Figure 6 in the text). Each column represents the average of two independent determinations each carried out in duplicate. Error bars indicate one standard deviation from the mean. AAG has no significant effect on the dissociation of EcoRI, and more than 98% of the EcoRI remains bound after 50 minutes incubation in the presence of chase DNA.

## References

1. Speina, E., Ciesla, J. M., Wojcik, J., Bajek, M., Kusmierk, J. T., and Tudek, B. (2001) The pyrimidine ring-opened derivative of 1,N6-ethenoadenine is excised from DNA by the Escherichia coli Fpg and Nth proteins, *J Biol Chem* 276, 21821-21827.
2. Poirot, O., O'Toole, E., and Notredame, C. (2003) Tcoffee@igs: A web server for computing, evaluating and combining multiple sequence alignments, *Nucleic Acids Res* 31, 3503-3506.
3. Terry, B. J., Jack, W. E., and Modrich, P. (1985) Facilitated diffusion during catalysis by EcoRI endonuclease. Nonspecific interactions in EcoRI catalysis, *J Biol Chem* 260, 13130-13137.
4. Stanford, N. P., Szczelkun, M. D., Marko, J. F., and Halford, S. E. (2000) One- and three-dimensional pathways for proteins to reach specific DNA sites, *Embo J* 19, 6546-6557.
5. Kampmann, M. (2004) Obstacle bypass in protein motion along DNA by two-dimensional rather than one-dimensional sliding, *J Biol Chem* 279, 38715-38720.
6. Gowers, D. M., Wilson, G. G., and Halford, S. E. (2005) Measurement of the contributions of 1D and 3D pathways to the translocation of a protein along DNA, *Proc Natl Acad Sci U S A* 102, 15883-15888.
7. Hedglin, M., and O'Brien, P. J. (2008) Human alkyladenine DNA glycosylase employs a processive search for DNA damage, *Biochemistry* 47, 11434-11445.
8. Jack, W. E., Terry, B. J., and Modrich, P. (1982) Involvement of outside DNA sequences in the major kinetic path by which EcoRI endonuclease locates and leaves its recognition sequence, *Proc. Natl. Acad. Sci. U. S. A.* 79, 4010-4014.
9. Baldwin, M. R., and O'Brien, P. J. (2009) Human AP Endonuclease I Stimulates Multiple-Turnover Base Excision by Alkyladenine DNA Glycosylase, *Biochemistry* 48, 6022-6033.
10. Wolfe, A. E., and O'Brien, P. J. (2009) Kinetic mechanism for the flipping and excision of 1,N(6)-ethenoadenine by human alkyladenine DNA glycosylase, *Biochemistry* 48, 11357-11369.
11. Wright, D. J., Jack, W. E., and Modrich, P. (1999) The kinetic mechanism of EcoRI endonuclease, *J. Biol. Chem.* 274, 31896-31902.
12. Halford, S. E., and Johnson, N. P. (1983) Single turnovers of the EcoRI restriction endonuclease, *Biochem. J.* 211, 405-415.

13. Ruben, G., Spielman, P., Tu, C. D., Jay, E., Siegel, B., and Wu, R. (1977) Relaxed circular SV40 DNA as cleavage intermediate of two restriction endonucleases, *Nucleic Acids Res.* *4*, 1803-1813.
14. Rubin, R. A., and Modrich, P. (1978) Substrate dependence of the mechanism of EcoRI endonuclease, *Nucleic Acids Res.* *5*, 2991-2997.
15. Langowski, J., Urbanke, C., Pingoud, A., and Maass, G. (1981) Transient cleavage kinetics of the Eco RI restriction endonuclease measured in a pulsed quench-flow apparatus: enzyme concentration-dependent activity change, *Nucleic Acids Res.* *9*, 3483-3490.
16. Modrich, P., and Zabel, D. (1976) EcoRI endonuclease. Physical and catalytic properties of the homogenous enzyme, *J. Biol. Chem.* *251*, 5866-5874.
17. Hedglin, M., and O'Brien, P. J. (2010) Hopping enables a DNA repair glycosylase to search both strands and bypass a bound protein, *ACS Chem Biol* *5*, 427-436.

## Chapter 4

### Probing the DNA Structural Requirements for Translocation by Alkyladenine DNA Glycosylase<sup>1</sup>

In order to conduct genome-wide searches for rare target sites, many proteins employ linear diffusion along nonspecific DNA, whereby each binding encounter with a given segment of DNA involves a search of multiple adjacent sites (1-7). Such a search is mediated by nonspecific, electrostatic binding interactions and may occur by hopping, sliding, or a combination of both modes. Sliding involves continuous contact between the protein and the DNA backbone so that transfer occurs between linearly continuous sites on the same strand. Although such characteristics afford a sliding protein multiple encounters with each nucleotide of the bound strand, thus increasing the probability that a lesion is recognized, reliance on sliding would slow down searches over long distances and render a protein susceptible to “roadblocks” which cannot be circumvented by sliding, such as tightly-bound DNA-binding proteins and distortions to the equipotential binding surface of the B-form DNA helix (8, 9). Such blockades would limit the breadth of nonspecific sites searched and decrease the likelihood of target site recognition. Indeed it has been observed for *EcoRI*, a protein suggested to locate its recognition sequence by an exclusive sliding mechanism, that proteins which bind firmly to DNA as well as relatively minor alterations in DNA conformation, such as an 18° bend, significantly reduce the ability of *EcoRI* to pass over a particular region of DNA (10, 11). However, if a given protein utilizes hopping movements during linear diffusion, then successive *microscopic* dissociation-reassociation events may enable the protein to bypass the “roadblock” and continue in its search. By monitoring the processive action of AAG in the presence of a tightly-bound protein, we observed that a tightly-bound protein only decreased the processivity of AAG by ~50%. This suggests that hopping by AAG may occur over a sufficient distance to allow AAG to *microscopically* dissociate and

---

<sup>1</sup> Special thanks to Yaru Zhang who helped perform the experiments and analyze the data in this chapter.

reassociate on the other side of a tightly-bound DNA-binding protein and that unbent, naked DNA is not required for processive action by AAG (8).

In the present work, we wish to better define the structural requirements of DNA that permit linear diffusion by AAG and gain greater insight into how processive proteins can overcome the structural obstacles to a genome-wide search for specific target sites. By utilizing our previously described *in vitro* processivity assay, we examined the extent to which various structural perturbations to the B-form helix impeded the processive action of AAG and found that kinks, bubbles, polyethylene glycol, and single-stranded regions do not pose a barrier to processive action by AAG. This further demonstrates that AAG does not maintain continuous contact with the DNA while diffusing along it and suggests that such excursions from the surface of DNA via hopping allow bypass of various DNA structures by transfer to adjacent sites on the other side of a structural barrier. In conjunction with previous reports, this provides ample support for hopping as an important mode of translocation in the search for DNA damage in the context of the genome (8, 9).

## **Materials and Methods**

### **Proteins**

Full-length and truncated recombinant human AAG were purified and the concentration of active AAG was determined by burst analysis as previously described (12)

### **Oligonucleotides**

DNA substrates were synthesized by Integrated DNA Technologies or the Keck Center at Yale University and purified by denaturing PAGE as previously described(12). The concentration of single-stranded DNA was determined from the absorbance at 260 nm using the calculated extinction coefficients. For 5', 3', or dual fluorescein-labeled oligonucleotides, we assessed the labeling efficiency by comparing the absorbance at 260 nm with that at 495 nm and the calculated labeling efficiency was greater than 95% in all cases. Unless otherwise stated, for a given duplex substrate, the labeled strand was annealed to a 2-fold excess of unlabeled complementary strand in annealing buffer (10 mM NaMES, pH 6.5, 50 mM NaCl). The sequences for all doubly-labeled  $\epsilon$ A-containing oligonucleotide duplexes are given in Figure 4-2A.



### **Native Gel Electrophoresis**

Oligonucleotide substrates were annealed in 10 mM NaMES pH 6.5 with 50 mM NaCl by heating to 95 °C and slowly cooling to room temperature. Annealed substrates were subsequently diluted into standard pH 6.1 glycosylase reaction buffer with an ionic strength of 200 mM. Samples were run on 10% native PAGE (19:1 acrylamide:bis-acrylamide, 0.5 X TBE) at constant voltage of 10 v/cm. Fluorescein-labeled DNA was visualized as described for sequencing gels.

### **Glycosylase Activity Assay**

Reactions were carried out as previously described at 37 °C in 50 mM NaMES, pH 6.1, 1 mM EDTA, 1 mM DTT, 10% glycerol, 0.1 mg mL<sup>-1</sup> BSA and the ionic strength was adjusted with NaCl(8, 12). Reactions were initiated by adding enzyme (2 - 1000 nM final concentration) to obtain a reaction volume of 50 – 100 µL that contained 100 - 200 nM fluorescein-labeled DNA. Aliquots were withdrawn at various times and quenched with NaOH (0.2 M final concentration). Samples were heated to 70 °C for 15 minutes, formamide was added to 65%, and the DNA fragments were resolved on 14% (v/v) polyacrylamide gels containing 8 M urea as previously described(8, 12). Gels were scanned with a Typhoon Trio<sup>+</sup> fluorescence imager (GE Healthcare) to detect fluorescein (excitation at 488 nM and emission with 520BP40 filter). The resulting fluorescent signal was quantified using ImageQuant TL and corrected for background signal. The intensity of each DNA band was converted into a fraction by dividing its intensity by the sum of the intensities for all of the DNA species present.

### **Multiple Turnover Kinetics**

Steady state kinetics for the dual lesion substrates were measured with 100-fold excess of substrate (200 nM) over enzyme (2 nM) as previously described (8, 12). The initial rates were calculated from the first 10-15% of the reaction and were linear in all cases. Values for both  $k_{cat}$  and  $F_p$  were calculated from initial rates for substrate depletion ( $k_{cat}$ ) or intermediates and products ( $F_p$ ). The value of  $F_p$  was calculated according to equation 1, in which  $V_i$  and  $V_p$  are the initial rates for the formation of intermediates (retaining an εA lesion) and products (both εA excised), respectively. The processivity equation takes into account the fact that a single excision event gives rise to one product and one intermediate. Approximately 2% of εA nucleotides are damaged during synthesis,

deprotection and gel purification, and therefore the maximal processivity value that is expected is 0.92 (12). The minimal value that could be observed is ~0.05 for a completely distributive mechanism, that is attributed to rebinding of AAG. Therefore, the effective range of experimental  $F_p$  values is 0.87 (0.92-0.05). To facilitate the comparison with previously published results, we have corrected all of the observed processivity values (eq 2).

$$F_{p, obs} = (V_p - V_i) / (V_i + V_p) \quad (1)$$

$$F_p = (F_{p, obs} - 0.05) / 0.87 \quad (2)$$

### Single-Turnover Kinetics

For single-turnover analysis of 1, $N^6$ -ethenoadenine removal from single- and double-stranded substrates with a single lesion, a 25-mer 1, $N^6$ -ethenoadenine-containing oligonucleotide 5'-(6-fam)-CGATAGCATCCTECCTTCTCTCCAT was annealed to either 0 or 1.5 equivalents of the unlabeled the complementary 5'-ATGGAGAGAAGGTAGGATGCTATCG oligonucleotide. The 25-mer substrate was then incubated with excess full-length AAG enzyme (referred to herein as FL AAG) to ensure single-turnover conditions. For single-stranded 25-mer substrate, this required >3-fold molar excess of enzyme. Due to  $\epsilon$ A damage during synthesis, deprotection and gel-purification, the maximal values expected for the fraction of product cleaved and the amplitude (A) is 0.97. To facilitate comparisons, we have corrected the fraction product by dividing all values by 0.97. The reaction progress curve was plotted as the fraction of product versus time and was fit by a single exponential (eq 3), where F is the fraction of reaction, A the amplitude,  $k_{obs}$  the observed rate constant, and  $t$  time.

$$F = A(1 - e^{-(k_{obs}t)}) \quad (3)$$

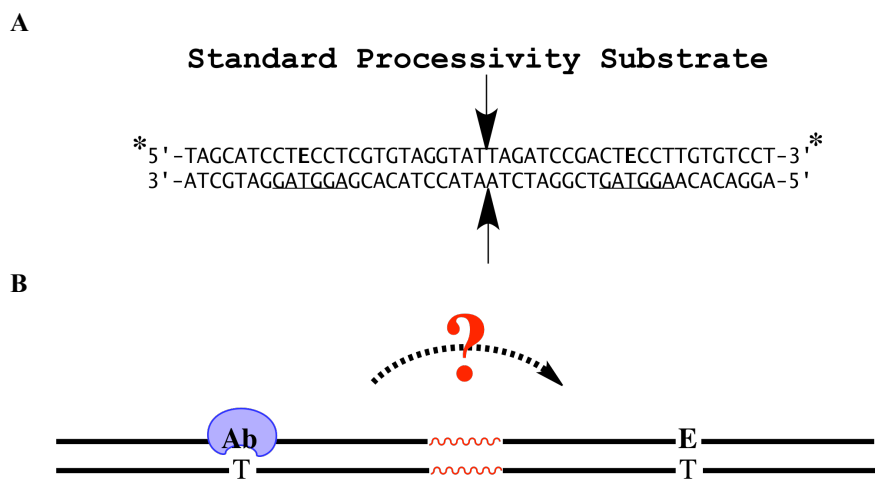
In all cases, the nonlinear least-squares fit was excellent ( $R^2 > 0.99$ ). At a saturating concentration of enzyme, the observed single-turnover rate constant reaches a maximum termed  $k_{max}$ . The concentration of enzyme (FL AAG) was varied by 5-fold for the single-stranded  $\epsilon$ A experiments to establish that the observed rate for single-stranded  $\epsilon$ A was independent of the concentration of enzyme, indicating that enzyme was in excess and at a saturating concentration (i.e.,  $k_{obs} = k_{max}$ , see Appendix C, Figure C-3A). For double-stranded  $\epsilon$ A, only 200 nM FL AAG was used (2-fold molar excess). It is known that enzyme is saturating at this concentration (12, 13). Indeed,  $k_{obs} = 0.23 \text{ min}^{-1} \pm 0.01$

(Appendix C, Figure C-3B), in excellent agreement with previous reports of  $k_{\max}$  for this substrate (13).

## Results

### Design of Substrates for Characterizing the Ability of AAG to Bypass Structural Alterations in DNA

Our multiple turnover processivity assay, described previously in Chapter 2, allows the transfer of a glycosylase between two sites on a DNA molecule to be monitored by measuring correlated cleavage for a substrate containing two sites of damage (Figure 4-1). Base excision catalyzed by AAG at one of the two sites results in an abasic product, and AAG subsequently partitions between action at the second site and dissociation into solution. Due to the large excess of intact substrate, AAG release is irreversible under these conditions. The fraction of processive events ( $F_p$ ) is determined by alkaline hydrolysis of abasic sites and initial rates for single and double excision events (8, 12). It was previously demonstrated that insertion of 25 extra base pairs of nonspecific duplex DNA between the two  $\epsilon$ A lesions did not have any effect on the



**Figure 4-1: Processivity assays to probe the DNA structural requirements for linear diffusion by A DNA repair glycosylase**

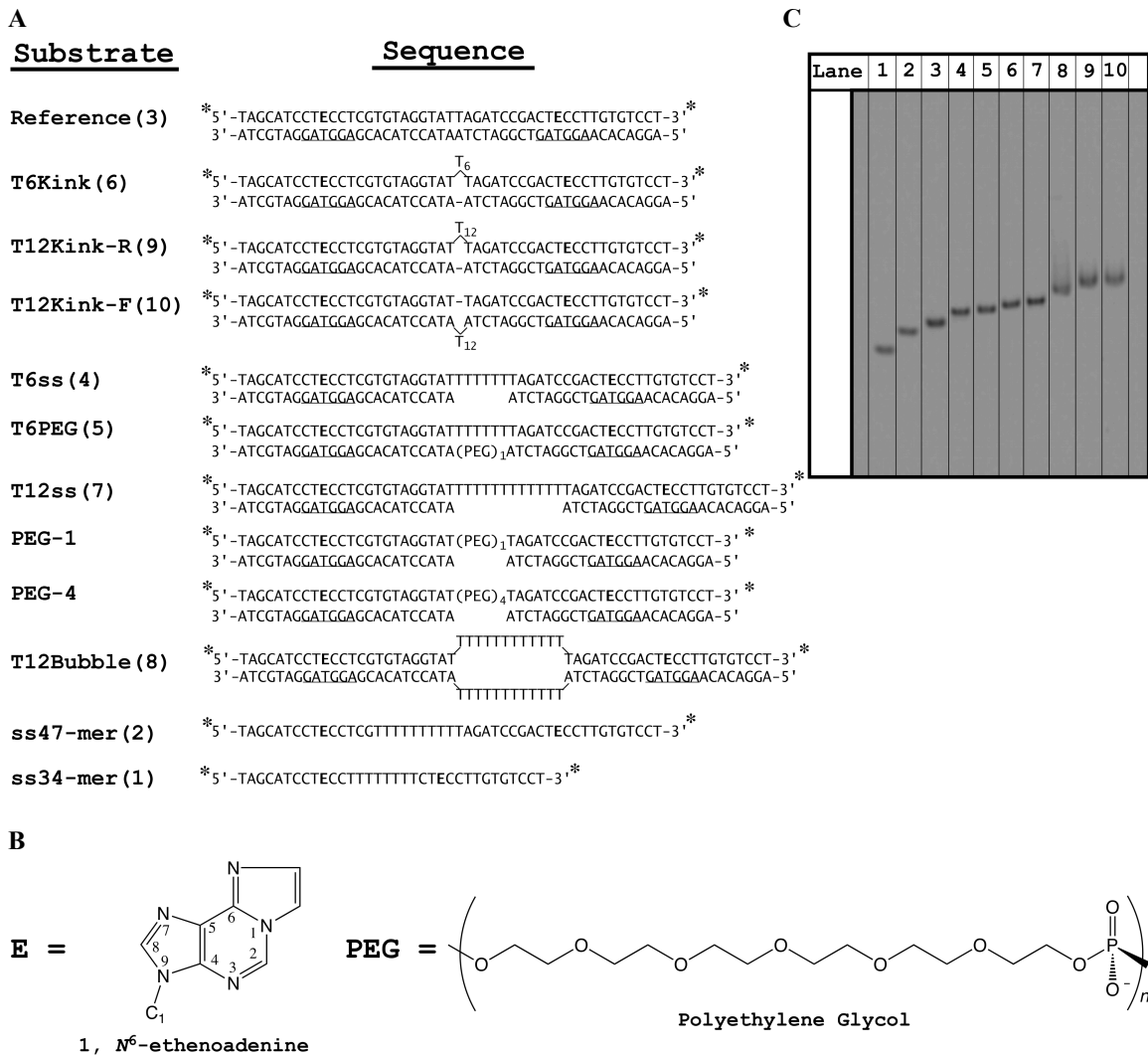
(A) Standard processivity assay substrate (referred to herein as “Reference”). A 47-mer oligonucleotide duplex that contains two  $\epsilon$ A lesions separated by 25 base pairs. The lesion-containing strand is labeled at both 3' and 5' ends with fluorescein (designated as asterisk). The immediate sequence contexts of the two lesions are identical (underlined). Arrows designate location between the two  $\epsilon$ A lesions where structural elements were introduced into the substrate. (B) Multiple turnover processivity assays follow events subsequent to an initial base excision event, effectively measuring partitioning between dissociation and correlated excision at the remaining lesion site. Substrates were designed so that AAG (blue) randomly binds to and excises either of the two  $\epsilon$ A lesions to create an abasic site (Ab). AAG release is irreversible under these conditions because substrate is in excess. Red wavy lines designate a given structural perturbation to the DNA duplex. If AAG is able to bypass a structural perturbation to the duplex (dashed arrow), then processive excision of the second  $\epsilon$ A lesion will be observed.

fraction processive of AAG, demonstrating that AAG searches much more than 50 bp per binding encounter (8). In order to examine the extent to which AAG is capable of bypassing structural perturbations to the B-form double helix, we inserted various DNA structures between the two sites of damage within our standard processivity substrate (Figure 4-1), and monitored the fraction processive ( $F_p$ ) under multiple-turnover conditions as previously described (12). An effect on the fraction processive of AAG will thus reflect on AAG's ability to bypass the inserted DNA.

### **AAG Is Efficient At Bypassing Kinks in DNA**

Insertion of unpaired nucleotides at internal positions in duplex DNA causes the unopposed bases to “bulge” outward, kinking the DNA. Such anomalies may arise naturally in heteroduplex DNA as a result of recombination between imperfectly homologous sequences or from replication errors. As a result of DNA kinking, the electrophoretic mobility of the DNA is reduced (Figure 4-2), the magnitude of which depends on the number and sequence of unpaired bases ( $n$ ), and the flanking base pairs (14-20). In order to test whether AAG is capable of bypassing a kink, we introduced a kink between the two  $\epsilon$ A lesions of the standard 47-mer processivity substrate (reference) by inserting 6 consecutive, unpaired thymines into the standard 47-mer labeled strand and annealing to the 47-mer, unlabeled complementary strand. The resultant substrate (T6Kink) is expected to contain a  $\sim 90^\circ$  kink approximately in the middle of the two  $\epsilon$ A lesions (Figure 4-3A). As expected, introduction of such a bulge into the standard processivity substrate (reference) retarded the mobility of the DNA, as shown in Figure 4-2C (compare lanes 3 and 6). Larger bulges ( $n = 12$ ) were also constructed by inserting 12 consecutive, unpaired thymines into the labeled (T12Kink-R) or unlabeled strand (T12Kink-F) of the reference substrate and annealing each to the appropriate opposing strand. A defined bend angle has yet to be determined for a 12-nucleotide bulge due to conflicting results from different experimental techniques. Nonetheless, the end-to-end distance of a DNA duplex is significantly reduced by insertion of a 12 nt bulge and suggests that this is the result of a kink with a bend angle near  $90^\circ$  (18).

To test whether AAG is capable of bypassing a DNA kink, we compared the relative processivity of full-length AAG on substrates in which a kink was present



**Figure 4-2: Substrates to monitor the ability of AAG to bypass DNA structural perturbations**  
 (A) Sequences of oligonucleotides that were employed in this study. All DNA duplexes are derived from the standard processivity substrate (47-mer, reference) and thus each contained two  $\epsilon$ A lesions (E) on the same strand within the same local sequence context as well as two fluorescein labels (asterisk). The T6Kink, T6PEG, and T6ss substrates all contain the same labeled strand. Likewise, the T12Kink-R, T12ss, and T12Bubble substrates all contain the same labeled strand. Numbers in parentheses next to a given substrate designate the lane a given substrate was run in the native PAGE shown in Figure 4-2C. (B) Structures of 1, N<sup>6</sup>-ethenoadenine (E) and polyethylene glycol (PEG) linker. (C) Native gel electrophoresis of oligonucleotides used in this study. Annealed substrates were diluted into glycosylase reaction buffer in the absence of AAG at 200 mM ionic strength.

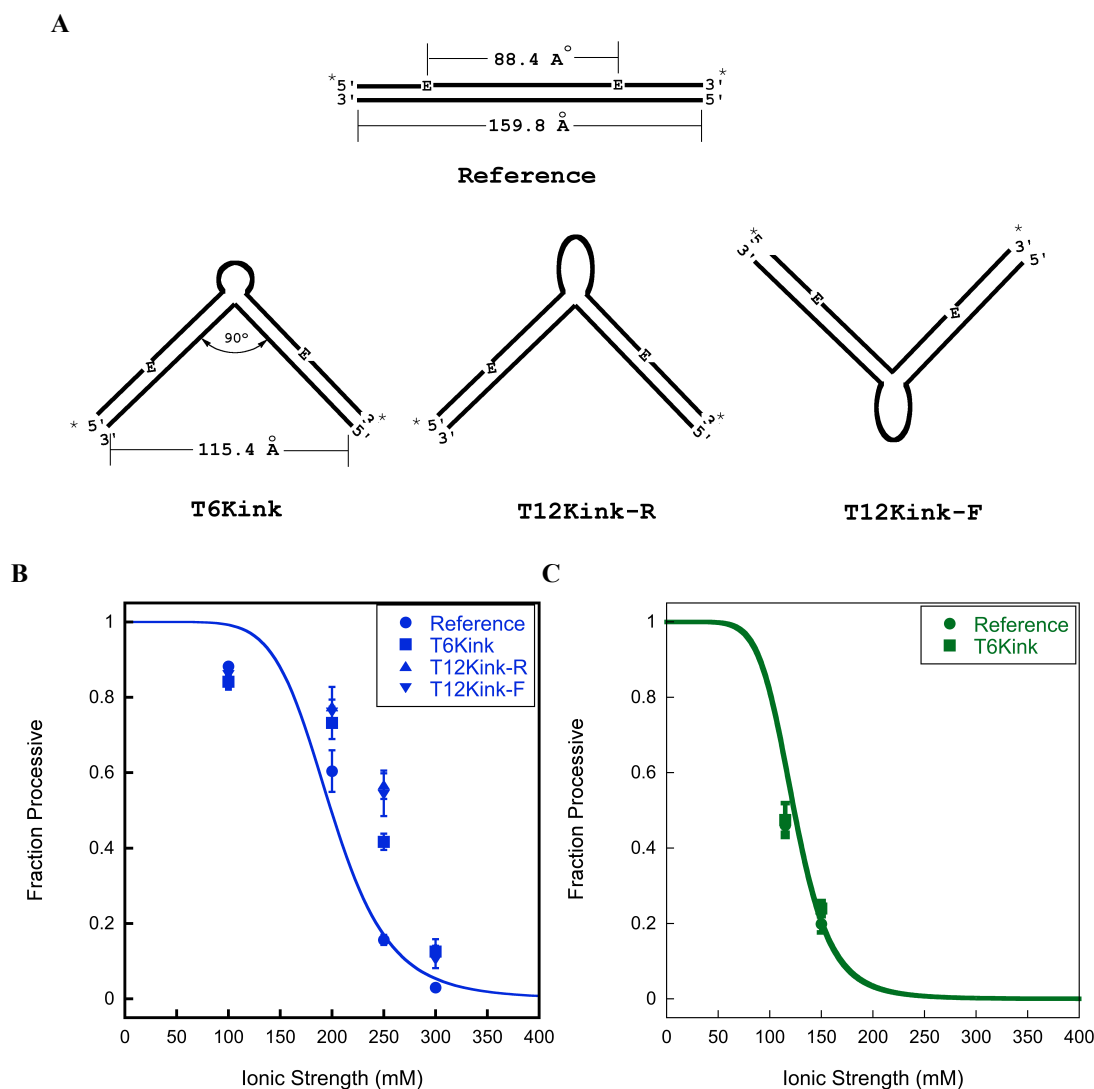
between the two  $\epsilon$ A lesions (T6Kink, T12Kink-R, T12Kink-F) to that for the reference substrate in which the lesions are separated by 25 base pairs of unbent, B-form DNA. Under conditions of low ionic strength (100 mM), full-length AAG (referred to herein as FL AAG) is highly processive on the reference substrate, with  $F_p = 0.88 \pm 0.01$  (Table 4-1), in good agreement with previously published data (8, 12). Under the same

**Table 4-1: Fraction processive of FL AAG**

Substrate	I = 100 mM		I = 200 mM		I = 250 mM		I = 300 mM	
	$F_p$	$F_p(\text{rel})$	$F_p$	$F_p(\text{rel})$	$F_p$	$F_p(\text{rel})$	$F_p$	$F_p(\text{rel})$
Reference	$0.88 \pm 0.01$	(1.0)	$0.60 \pm 0.05$	(1.0)	$0.16 \pm 0.014$	(1.0)	$0.030 \pm 0.008$	(1.0)
T6Kink	$0.84 \pm 0.02$	(0.95)	$0.73 \pm 0.04$	(1.2)	$0.42 \pm 0.022$	(2.7)	$0.13 \pm 0.02$	(4.2)
T12Kink-R	$0.86 \pm 0.04$	(0.97)	$0.78 \pm 0.05$	(1.3)	$0.57 \pm 0.038$	(3.6)	$0.14 \pm 0.02$	(4.6)
T12Kink-F	$0.86 \pm 0.01$	(0.97)	$0.76 \pm 0.03$	(1.3)	$0.54 \pm 0.057$	(3.5)	$0.10 \pm 0.02$	(3.5)
T6ss	$0.87 \pm 0.01$	(0.99)	$0.67 \pm 0.04$	(1.1)	$0.23 \pm 0.034$	(1.5)	$0.061 \pm 0.047$	(2.1)
T6PEG	$0.85 \pm 0.02$	(0.96)	$0.64 \pm 0.05$	(1.1)	$0.32 \pm 0.010$	(2.0)	$0.11 \pm 0.01$	(3.6)
T12ss	$0.89 \pm 0.02$	(1.0)	$0.70 \pm 0.06$	(1.2)	$0.32 \pm 0.030$	(2.0)	$0.026 \pm 0.034$	(1.7)
T12Bubble	$0.91 \pm 0.04$	(1.0)	$0.78 \pm 0.01$	(1.3)	$0.61 \pm 0.001$	(3.9)	$0.11 \pm 0.01$	(3.8)
PEG-1	N.D.		$0.68 \pm 0.03$	(1.1)	N.D.		N.D.	
PEG-4	N.D.		$0.64 \pm 0.04$	(1.1)	N.D.		N.D.	
ss47-mer	$-0.0061 \pm 0.04$		$0.056 \pm 0.04$		N.D.		$0.0089 \pm 0.0120$	
ss34-mer	$-0.024 \pm 0.03$		$0.018 \pm 0.02$		N.D.		$-0.031 \pm 0.018$	

Multiple turnover processivity experiments were performed with full-length AAG using substrates that contained two  $\epsilon$ A lesions separated by either unbent, B-form DNA (reference) or a structural obstacle. The fraction processive was calculated as described in Materials and Methods. Values in parentheses are the relative values obtained by dividing each respective  $F_p$  value by that for the reference substrate for a given ionic strength (I) condition. Corresponding multiple-turnover rate constants ( $k_{\text{cat}}$ ) calculated from the same experiments are shown in Appendix C, Table C-1.

conditions, FL AAG is also highly processive on all substrates in which a kink was present between the two lesions (T6Kink, T12Kink-R, T12Kink-F) and all  $F_p$  values were both non-directional (data not shown) and indistinguishable from that observed for the reference (Table 4-1 & Figure 4-3B). This demonstrates that FL AAG is able to bypass kinked DNA from either direction, presumably by hopping between the duplex arms at least once, so that both lesions are excised prior to *macroscopic* dissociation and rebinding to another DNA molecule. Under low ionic strength conditions, turnover is limited by dissociation from the abasic product ( $k_{\text{cat}} = k_{\text{dissociation}}$ ). At 100 mM ionic strength, the full-length protein has a half-life of  $\sim 40$  min [ $t_{1/2} = (\ln 2)/k_{\text{dissociation}} = (\ln 2)/0.017 \text{ min}^{-1} = 41 \text{ min}$ ] for its dissociation from DNA containing an abasic DNA product ((12) Appendix C, Table C-1 & Figure C-1). A difference in the processivity would be difficult to detect under these conditions because the residence time of FL AAG on the DNA is very long (12). The maximum sensitivity is observed when  $\sim 50\%$  of the binding events are processive (*i.e.*,  $F_p \sim 0.5$ ). These conditions are obtained at an ionic strength of 200 mM for the full-length protein (8, 12). Therefore, multiple turnover processivity assays were carried out on these substrates at 200 mM ionic strength and the results were directly compared.



**Figure 4-3: AAG is efficient at bypassing kinks in DNA**

(A) Cartoon depiction of reference and kinked substrates. Based on the persistence length ( $\sim 150$  bp) and the pitch for B-form DNA ( $3.4 \text{ \AA/bp}$ ), the distance between the DNA termini and between the two  $\epsilon A$  lesions in the reference substrate are estimated to be  $159.8$  and  $88.4 \text{ \AA}$ , respectively. Introduction of six consecutive, unpaired thymines into the lesion-containing strand of the reference substrate produces  $\sim 90^\circ$  kink between the two  $\epsilon A$  lesions (T6Kink), decreasing the end-to-end distance from  $159.8$  to  $115.4 \text{ \AA}$ . The resultant bend angle from increasing the number of unpaired thymines from 6 to 12 (T12Kink-R or T12Kink-F) has yet to be determined. Multiple turnover processivity experiments were performed with either full-length (B) or  $\Delta 80$  AAG (C) using substrates that contained two  $\epsilon A$  lesions separated by either unbent, B-form DNA (reference,  $\bullet$ ) or a DNA kink (T6Kink =  $\blacksquare$ , T12Kink-R =  $\blacktriangle$ , T12Kink-F =  $\blacktriangledown$ ). The fraction processive was calculated as described in Materials and Methods. Data points are overlaid on top of the ionic strength dependence of the fraction processive previously determined for each enzyme form on the reference substrate (data points removed for clarity) and fit to a cooperative model with 7 inhibitory sodium ions (8, 12). Corresponding multiple-turnover rate constants ( $k_{\text{cat}}$ ) calculated from the same experiments are shown in Appendix C, Figure C-1.

For the reference substrate at  $200 \text{ mM}$  ionic strength, we obtained values of  $0.60 \pm 0.05$  for  $F_p$  (Table 4-1 & Figure 4-3B) and  $0.13 \text{ min}^{-1} \pm 0.02$  for  $k_{\text{cat}}$  (Appendix C,

Table C-1 & Figure C-1A), both in good agreement with that observed previously (8, 12). When the processivity of FL AAG on the reference substrate was directly compared to substrates in which lesions were separated by a kink (T6Kink, T12Kink-R and -F),  $F_p$  values were consistently observed which were 1.2- to 1.3-fold greater than that for the reference (Table 4-1). Similar values were obtained with different size bulges,  $n = 6$  versus  $n = 12$ , and with different directions of tilt of the kink; away from the complementary strand (T12Kink-R) versus toward the complementary strand (T12Kink-F). Taken together, these data clearly demonstrates that a DNA kink is not an impediment to processive action by FL AAG to any appreciable extent and that after removing either of the two  $\epsilon$ A lesions, FL AAG is able to traverse a DNA segment containing a kink just as efficiently as B-form, double helical DNA in order to remove the remaining  $\epsilon$ A lesion prior to dissociation. However, it is interesting to note that at higher ionic strength a greater difference in processivity is observed between the linear and bent substrates. When directly compared to the reference substrate ( $F_p = 0.16 \pm 0.01$ ) at 250 mM ionic strength,  $F_p$  increased 2.7-, 3.6-, and 3.5-fold for the T6Kink, T12Kink-R, and T12Kink-F substrates, respectively (Figure 4-3B). For a hopping protein, the probability of encountering a given site after departing a defined position on the DNA is expected to decrease with the inverse of the straight-line distance ( $r$ ) between the two sites ( $1/r$ ) (21). This suggests that, under these conditions, a kink increases the probability of locating a second lesion, presumably by bringing the two lesion sites closer together. This will be discussed further in later sections.

In Chapter 2, we demonstrated that, although not required for *N*-glycosylase activity, the N-terminal 79 amino acids of AAG plays a role in nonspecific DNA binding, possibly by directly contacting the DNA, leading to enhanced processive behavior (8, 12). However, as shown in Chapter 3, the N-terminus is not required for strand transfer between opposing strands of a given duplex DNA or bypass of a tightly-bound DNA binding protein (8, 12). From the data shown in Table 4-1 & Figure 4-3B, it is clear that full-length AAG is able to bypass various DNA kinks with high efficiency. We next sought to address whether this behavior was dependent upon the N-terminus by evaluating the processive behavior of a truncation mutant in which the N-terminal 79 amino acids of AAG have been truncated (referred to herein as  $\Delta 80$  AAG). Just as with the full-length



**Table 4-2: Fraction processive of  $\Delta 80$  AAG**

Substrate	I = 115 mM		I = 150 mM	
	$F_p$	$F_p(\text{rel})$	$F_p$	$F_p(\text{rel})$
Reference	$0.45 \pm 0.01$	(1.0)	$0.20 \pm 0.02$	(1.0)
T6Kink	$0.47 \pm 0.05$	(1.1)	$0.24 \pm 0.02$	(1.2)
T6ss	$0.42 \pm 0.02$	(0.94)	$0.20 \pm 0.04$	(1.0)
T6PEG	$0.49 \pm 0.04$	(1.1)	$0.16 \pm 0.02$	(0.82)
PEG-1	$0.42 \pm 0.02$	(0.93)	N.D.	
PEG-4	$0.38 \pm 0.03$	(0.84)	N.D.	
ss34-mer	$-0.075 \pm 0.046$		N.D.	

Multiple turnover processivity experiments were performed with  $\Delta 80$  AAG using substrates that contained two  $\epsilon$ A lesions separated by either unbent, B-form DNA (reference) or a structural obstacle. The fraction processive was calculated as described in Materials and Methods. Values in parentheses are the relative values obtained by dividing each respective  $F_p$  value by that for the reference substrate for a given ionic strength (I) condition. Corresponding multiple-turnover rate constants ( $k_{\text{cat}}$ ) calculated from the same experiments are shown in Appendix C, Table C-2.

protein, we carried out multiple turnover processivity experiments with  $\Delta 80$  AAG under conditions where  $\sim 50\%$  of the binding events are processive (*i.e.*,  $F_p \sim 0.5$ ). These conditions are obtained at an ionic strength of 115 mM for the truncation mutant (8, 12). As shown in Table 4-2 and Figure 4-3C, when the processive behavior of  $\Delta 80$  AAG on the reference and T6Kink substrates was directly compared, the observed  $F_p$  values were indistinguishable ( $F_p = 0.45 \pm 0.01$  and  $0.48 \pm 0.05$  for the reference and T6Kink substrates, respectively). At an increased ionic strength (150 mM I) the processive behavior of  $\Delta 80$  AAG on each substrate remains indistinguishable ( $F_p = 0.20 \pm 0.02$  and  $0.24 \pm 0.02$  for the reference and T6Kink substrates, respectively). These results clearly demonstrate that the N-terminus is not required for bypass of a DNA kink.

#### **AAG is Efficient at Bypassing Single-Stranded Gaps and Bubbles in DNA**

Bulges in DNA create a rigid kink such that the relative positions of the duplex arms are statically defined (14-20). The results discussed above suggest that such static orientation may permit hopping between the duplex arms, allowing AAG to remove both  $\epsilon$ A lesions within a single binding encounter. To examine this further, we evaluated the extent to which AAG can bypass flexible regions, which are expected to promote a dynamic ensemble of the relative orientations of the duplex arms.

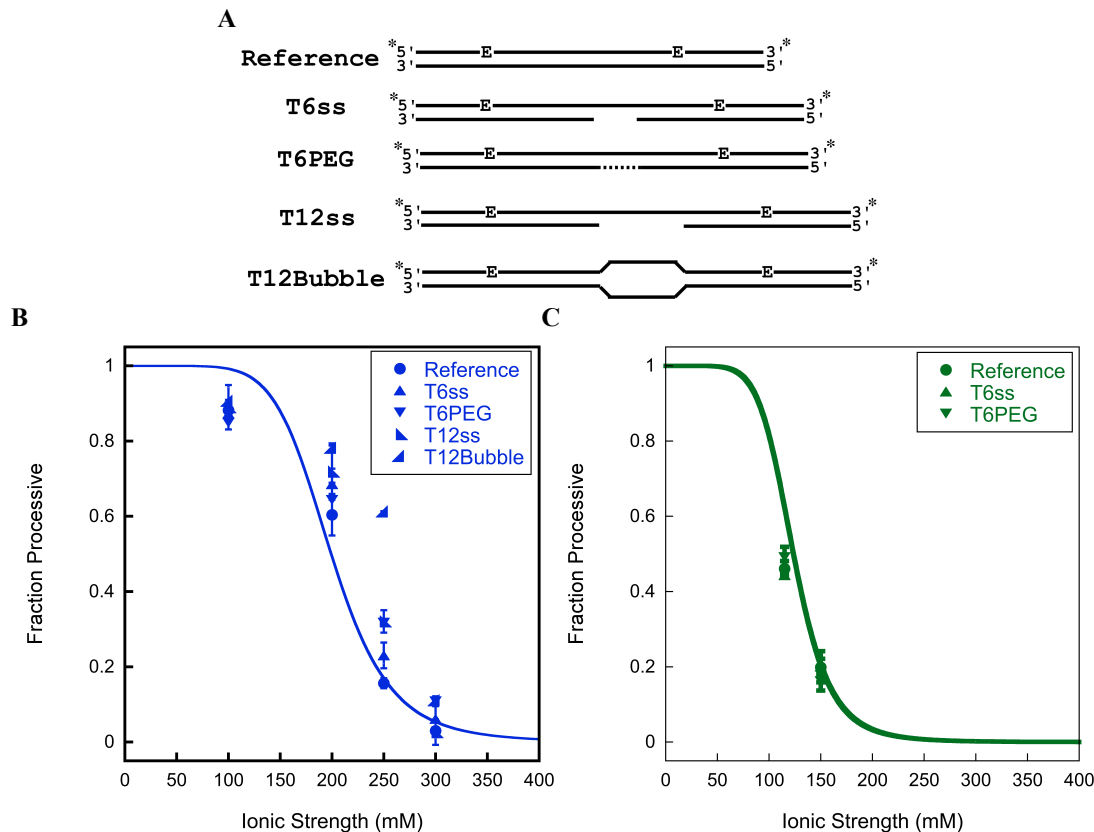
Due to the lack of base-pairing, single-stranded DNA is intrinsically flexible with measured persistence lengths of only 2 – 6 nt, in stark contrast to the  $\sim 150$  bps for double-stranded B-form DNA (22-25). However, the flexibility of single-stranded DNA,

just like that for double-stranded DNA, is dependent on a multitude of parameters such as base sequence, temperature, counterion concentration, and length (23).

Oligodeoxythymidylates (oligo-dT) do not have base pairing capabilities, significant intrastrand interactions (e.g. base stacking), or base ordering and, thus, are widely used for studying single-stranded DNA (23, 25). Therefore, we inserted oligo-dTs (6 or 12) between the two lesions of the standard 47-mer labeled strand and annealed the flanking 3'- and 5' sequences each to discrete complementary strands such that the opposing strand was discontinuous. Thus, a single-stranded DNA gap was present between B-form duplex regions, which each contain an  $\epsilon$ A lesion (Figures 4-2A and 4-4A). We then measured the ability of AAG to bypass a single-stranded DNA gap by directly comparing the processivity of AAG on the single-stranded gap substrates (T6ss or T12ss) to that for the reference substrate. The results are summarized in Table 4-1 and Figure 4-4.

As discussed above, at 200 mM ionic strength,  $F_p = 0.60 \pm 0.049$  for the full-length protein on the reference substrate. Under the same conditions, the processive behavior of FL AAG is similar on substrates in which a single-stranded DNA gap was present between the two  $\epsilon$ A lesions (T6ss or T12ss). This behavior was observed at 250 mM ionic strength as well. Thus, it is evident that a single-stranded DNA gap ( $\geq 12$  nt) is not an impediment to diffusion by the full-length protein. When experiments were carried out with  $\Delta 80$  AAG (Table 4-2 & Figure 4-C), the processive behavior on the T6ss substrate was indistinguishable from that for the reference substrate at both 115 mM and 150 mM ionic strength. Together, this clearly demonstrates that single-stranded DNA gaps do not pose a threat to the processive behavior of AAG and that the N-terminus of AAG is not required for bypass of single-stranded DNA gap. This behavior was also observed when the discontinuous segments of the complementary strands were connected in a manner where the consecutive oligo-dTs of the labeled strand were not conformed to base pairs, as discussed below.

For the T6PEG substrate, the discontinuous segments of the complementary strand were connected by a non-DNA linker comprised of 6 consecutive ethylene glycols (referred to as polyethylene glycol or PEG) such that the single-stranded oligo-dTs of the labeled strand were not opposed by any nucleotides but the complementary strand was now continuous (Figures 4-2A and 4-4A). The unit length within a PEG linker



**Figure 4-4: AAG bypasses flexible regions within DNA**

(A) Cartoon depiction of the reference substrate and substrates with regions of flexibility. (B - C) Multiple turnover processivity experiments were performed with either full-length (B) or  $\Delta 80$  AAG (C) using substrates that contained two  $\epsilon A$  lesions separated by either unbent, B-form DNA (reference, ●) or a flexible region (T6ss = ▲, T6PEG = ▼, T12ss = ▴, T12Bubble = ▾). Shown is the fraction processive for each substrate calculated at the indicated ionic strengths as described in Materials and Methods. As in Figure 4-3B – C, data points are overlaid on top of the ionic strength dependence of the fraction processive previously determined for each enzyme form on the reference substrate (data points removed for clarity) and fit to a cooperative model with 7 inhibitory sodium ions (8, 12). Corresponding multiple-turnover rate constants ( $k_{cat}$ ) calculated from the same experiments are shown in Appendix C, Figure C-2.

(3.50 Å/ethylene glycol) is less than that for single-stranded DNA (6.3 Å/base) and is very similar to that for double-stranded DNA (3.40 Å/bp). This may constrain the movement of the non-base-paired region of the substrate. When directly compared, the processive behavior of AAG on the T6PEG substrate was indistinguishable from that observed for the T6ss substrate under all conditions for both enzyme forms (Tables 4-1 and 4-2 and Figures 4-4B and 4-4C). Thus, connecting the discontinuous segments of the complementary strand and opposing the single-stranded DNA gap with non-DNA did not have any effect on the processive behavior of AAG. To probe further, we next connected the discontinuous segments of the complementary strand for the T12ss substrate with 12 consecutive oligo-dTs, thus converting the single-stranded DNA gap into a region of

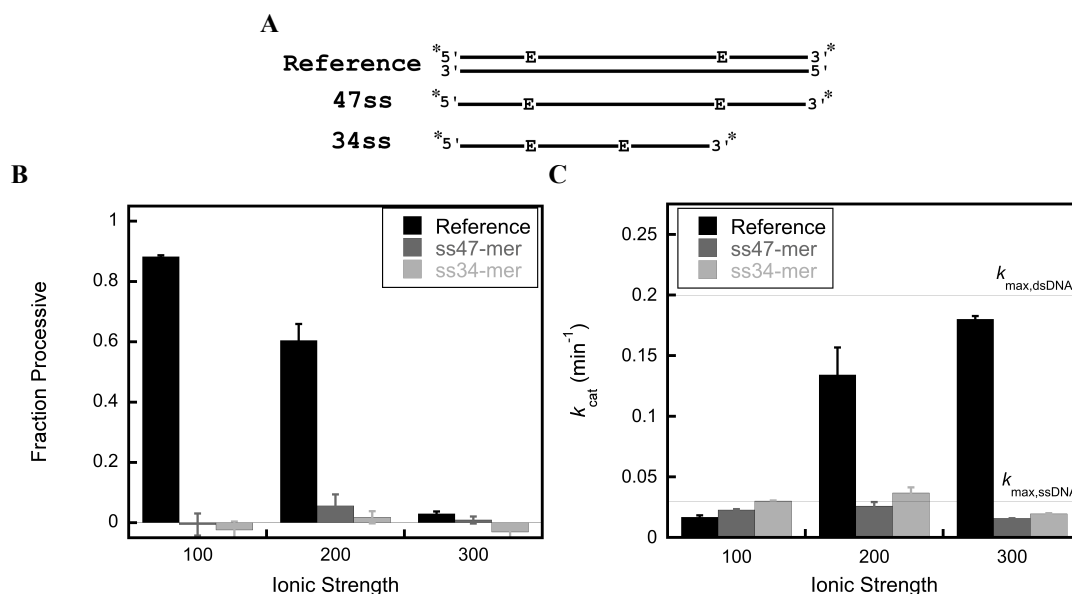
consecutive mismatches incapable of forming Watson-Crick base pairs, commonly referred to as a bubble (T12Bubble, Figures 4-2A and 4-4A). Gel electrophoresis suggests that bubbles only have a minor effect on the helical trajectory of duplex DNA and are suggested to represent a more poorly defined, flexible structure (19). When the processive behavior of FL AAG on the T12Bubble, T12ss, and reference substrates was directly compared similar values were obtained under most ionic strength conditions (Tables 4-1 and 4-2 and Figures 4-4B and 4-4C). It should be noted that under a single experimental condition (250 mM ionic strength), a considerably higher  $F_P$  value ( $0.61 \pm 0.01$ ) was consistently observed for the T12Bubble substrate that was 1.9- and 3.9-fold greater than that observed for the reference and T12ss substrates, respectively. This will be discussed further below. Taken together, these results clearly demonstrate that single-stranded gaps ( $\geq 12$  nt), PEG, and bubbles are not barriers to processive action by AAG and that the N-terminus of AAG is not required for circumventing such obstacles.

These results are consistent with a model for bypass of a flexible region where random, thermally-excited bending of the flexible region allows the adjacent B-form, duplex segments to frequently collide and AAG to hop between them (6, 26). However, in each substrate discussed above, the lesion-containing strand is continuous and the duplex regions are connected by single-stranded DNA. Even for a DNA bubble, the DNA strands maintain significant single-stranded character, as demonstrated by single-strand specific nuclease sensitivity (19). Thus, a simpler model that cannot be ruled out is that AAG diffuses along the single-stranded DNA connecting the two duplex arms. Indeed, it is known that AAG has considerable binding affinity to both specific (lesion-containing) and nonspecific (undamaged) single-stranded DNA. Previous reports have shown that AAG binds with decreased but moderate affinity to single-stranded, lesion-containing DNA as demonstrated by its ability to remove 1, $N^6$ -ethenoadenine ( $\epsilon$ A), among other lesions, from single-stranded DNA, albeit with a decreased activity (27, 28). As shown in Appendix Figure C-3A, the single-turnover rate constant ( $k_{\max}$ ) is  $0.029 \text{ min}^{-1} \pm 0.001$  for FL AAG on single-stranded  $\epsilon$ A, an  $\sim 8$ -fold decrease compared to that observed for a double-stranded substrate of the same sequence context (Appendix C, Figure C-3B, and reference 8). Based on EMSA as well as competition experiments, it has been observed that AAG also binds to undamaged, single-stranded DNA (personal

communication from P.O. on unpublished data). Thus, it is conceivable that AAG may be able to diffuse along single-stranded DNA and that such diffusion contributes to the results discussed above. As an initial attempt to address this, we sought to test whether AAG is capable of removing multiple  $\epsilon$ A lesions from single-stranded DNA within a single-binding encounter. Such processive behavior would demonstrate that AAG is capable of diffusing along single-stranded DNA. Therefore, we measured the fraction processive ( $F_p$ ) on single-stranded DNA substrates containing two  $\epsilon$ A lesions separated by either 12 or 25 nt. The results are summarized in Tables 4-1 and 4-2 and Figure 4-5.

### **AAG Is Not Processive on Single-Stranded DNA**

The lesion-containing, labeled strand of the reference substrate described above (Figure 4-2A) is a 47-mer oligonucleotide that contains two terminal fluorescein labels and two  $\epsilon$ A lesions separated from each other by 25 nt and from the respective termini by a unique distance. As an initial attempt to test whether AAG is processive on single-stranded DNA, experiments were carried out by directly comparing the fraction processive ( $F_p$ ) of AAG on the 47-mer labeled strand when it was single-stranded or annealed to 2-fold excess complement. However, results indicated that the 5'-end of this labeled oligonucleotide displays considerable double-stranded behavior, even at low concentrations of DNA such that AAG displayed a considerable preference (>5-fold) for the 5' lesion under multiple-turnover conditions and displayed significant processive behavior only in the 3'  $\rightarrow$  5' direction. Further experiments revealed that, under single-turnover conditions, the 5' lesion is removed with a  $k_{\max}$  ( $\sim 0.16 - 0.28 \text{ min}^{-1}$ ) similar to that for double-stranded DNA ( $0.20 \text{ min}^{-1}$ ), whereas the 3' lesion was removed with a much lower  $k_{\max}$  value of  $\sim 0.007 - 0.04 \text{ min}^{-1}$  (data not shown). We infer that these values correspond to the rate constant for excision of  $\epsilon$ A from single-stranded DNA (see below). Sequence analysis revealed that a particular internal sequence of nucleotides between the two  $\epsilon$ A lesions was possibly contributing to this behavior. Nucleotides 17 – 24 (sequence 5'-GTAGGTAT-3') of the standard 47-mer labeled strand were changed to oligo-dT, keeping both the 5'- and 3' termini and, thus, the local sequence context surrounding each  $\epsilon$ A lesion, identical to that for the standard 47-labeled strand. We then carried out multiple turnover processivity assays on the resultant oligonucleotide, referred to as ss47-mer (Figure 4-2A); the results are summarized in Figure 4-5.



**Figure 4-5: AAG is not processive on single-stranded DNA**

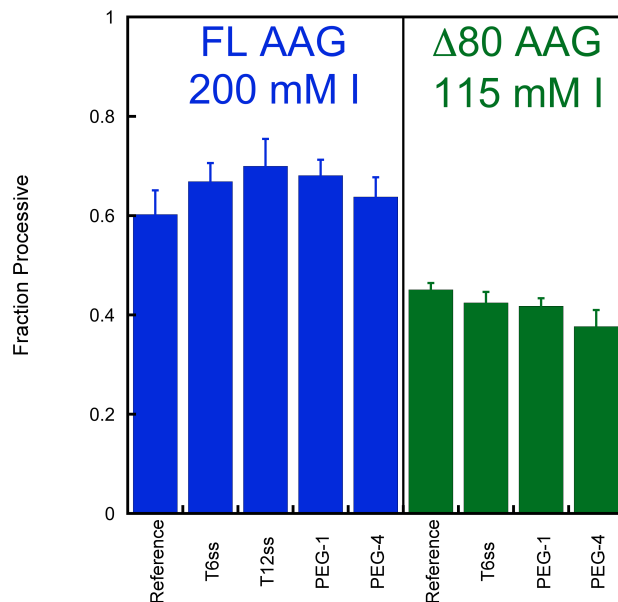
(A) Cartoon depiction of double-stranded (reference) and single-stranded substrates (ss47-mer, ss34-mer). Multiple turnover processivity assays were performed at optimal pH (6.1) with 200 nM substrate (reference, ss47-mer, or ss34-mer), and 2 nM FL AAG at 100, 200, and 300 mM ionic strength. Shown is the fraction processive (B) and multiple turnover rate constant,  $k_{cat}$ , (C) for each substrate calculated at the indicated ionic strengths as described in Materials and Methods. Each column represents the average of two to six independent determinations with error bars indicating one standard deviation from the mean ( $n \geq 2$ ).

Even at the lowest ionic strength tested (100 mM) where  $F_p = 0.88 \pm 0.01$  for the reference substrate, FL AAG did not display any measurable processivity ( $F_p \sim 0.0$ ) on single-stranded DNA. This behavior was also observed when the distance between the two  $\epsilon A$  lesions was reduced from 25 nt to 12 nt (ss34-mer)<sup>2</sup>. It is also worth noting that the multiple-turnover rate constants ( $k_{cat}$ ) for removal of  $\epsilon A$  from single- and double-stranded DNA display different dependencies on ionic strength. As shown in Figure 4-5C, for removal of  $\epsilon A$  from double-stranded DNA (reference), the multiple turnover rate constant ( $k_{cat}$ ) increases with increasing ionic strength until it reaches the single-turnover rate constant ( $k_{max,dsDNA}$ ). This is because dissociation from abasic DNA product is rate-limiting for multiple turnover base excision at low ionic strength. Thus, increasing the ionic strength promotes dissociation, resulting in decreased processivity at higher ionic strengths (12). However, for removal of  $\epsilon A$  from single-stranded DNA (ss47-mer or ss34-mer),  $k_{cat}$  is independent of ionic strength and comparable to the single-turnover rate

<sup>2</sup>  $\Delta 80$  AAG did not display any processive behavior on the ss34-mer substrate at 115 mM ionic strength (Table 4-2).

constant for removal of  $\epsilon$ A from single-stranded DNA ( $k_{\max,ssDNA}$ ). This suggests that, unlike that observed for double-stranded DNA, *N*-glycosidic bond hydrolysis is rate-limiting for removal of  $\epsilon$ A from single-stranded DNA under all conditions and dissociation from the abasic product is faster from single-stranded DNA, in agreement with processivity data (Figure 4-5B). However, from the data in Figure 4-5, it cannot be discerned whether AAG is incapable of diffusing along single-stranded DNA or whether the remaining  $\epsilon$ A lesion was located prior to dissociation but AAG but failed to excise it due to unproductive binding encounters. In order to distinguish between the two, the efficiency of recognition ( $E$ ) for  $\epsilon$ A on single-stranded DNA will be examined, which measures the extent to which enzyme-bound substrate is committed to base excision as opposed to being released into solution (8, 9). Preliminary experiments suggest that AAG is not efficient at excising  $\epsilon$ A from single-stranded DNA (personal communication from Y.Z.). Thus, further analysis is required in order to ascertain whether or not AAG is capable of diffusing along single-stranded DNA and, based on these experiments, it cannot be determined whether such diffusion is responsible for bypass of single-stranded DNA gaps as well as other flexible regions and kinks. An alternative and more direct method to discern between the two models for bypass would be to remove all nucleic acid between the outer duplex arms, leaving a gap comprised only of polyethylene glycol (PEG), such that transfer between the duplex arms could only occur by hopping. Therefore, we inserted PEG spacers ( $n = 1$  or  $4$ ) between the two lesions of the standard 47-mer labeled strand and annealed the flanking 3'- and 5' sequences each to discrete complementary strands such that the opposing strand was discontinuous. Thus, a non-DNA gap was present between B-form duplex regions, which each contain an  $\epsilon$ A lesion (Figure 4-2A). We then measured the ability of AAG to bypass a PEG gap by directly comparing the processivity of AAG on the PEG gap substrates (PEG-1 or PEG-4) to that for the single-stranded DNA gap substrates (T6ss and T12ss) as well as the reference substrate. The results are summarized in Tables 4-1 and 4-2 and Figure 4-6.

In all cases, the processivity of AAG (FL and  $\Delta 80$ ) on substrates with lesions separated by either B-form, duplex DNA or a single-stranded DNA gap was similar to the processivity on a substrate with lesions separated by a PEG linker. This demonstrates



**Figure 4-6: Diffusion along single-stranded DNA is not required for transfer between adjacent duplex regions**

To test whether diffusion along single-stranded is responsible for transfer between the adjacent duplex regions within the single-stranded DNA gap substrates, we measured the processivity for substrates in which lesions were separated by B-form duplex DNA (reference), a single-stranded DNA gap (T6ss or T12ss) or a PEG linker (PEG-6 or PEG-24). Multiple turnover processivity assays were performed at optimal pH (6.1) an ionic strength of 200 mM for full-length AAG (blue, FL AAG) or 115 mM for the N-terminal truncation mutant (green, Δ80 AAG). Each column represents the average of at least 3 independent determinations with error bars indicating one the standard condition from the mean ( $n \geq 3$ ).

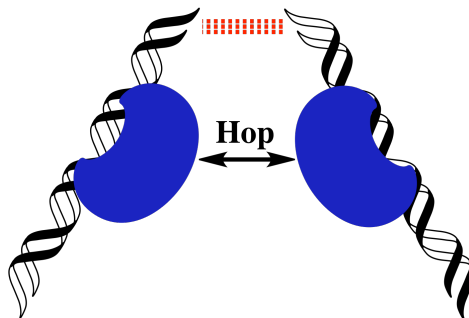
that diffusion along single-stranded DNA is not required for and does not contribute significantly to bypass of various structural obstacles. Furthermore, when the PEG linker is increased from  $n = 1$  (PEG-1) to  $n = 4$  (PEG-4), which is expected to decrease the frequency of collisions between the adjacent duplex regions, the fraction processive remained constant for both enzyme forms. This suggests that the duplex arms in the PEG-4 substrate collide successfully, with sufficient frequency, to allow transfer of AAG between the duplex regions and infer that a much greater increase ( $\gg 4$ -fold) in the length of the PEG linker is required in order to observe any effect on the fraction processive.

## Discussion

We have investigated the DNA structural requirements that permit translocation by alkyladenine DNA glycosylase. By inserting various DNA structures in between two sites of damage, multiple turnover processivity assays have provided evidence that various structural perturbations to the B-form double helix do not impede processive



behavior of AAG. By comparing an amino-terminally truncated enzyme to the full-length enzyme, we also found that the amino terminus is not required for processive behavior on these substrates. Although it still remains in question whether AAG is capable of diffusing along single-stranded DNA, we have demonstrated that such diffusion is not required for transfer of AAG past flexible regions and kinks present in DNA. Taken together, the data presented in this chapter is most consistent with a model whereby introduction of a flexible region or a kink into DNA decreases the distance ( $r$ ) between the duplex regions and may also allow the duplex regions to frequently collide with each other. Consequently, AAG can bypass such obstacles by transferring between the adjacent duplex regions via hopping (Figure 4-7).

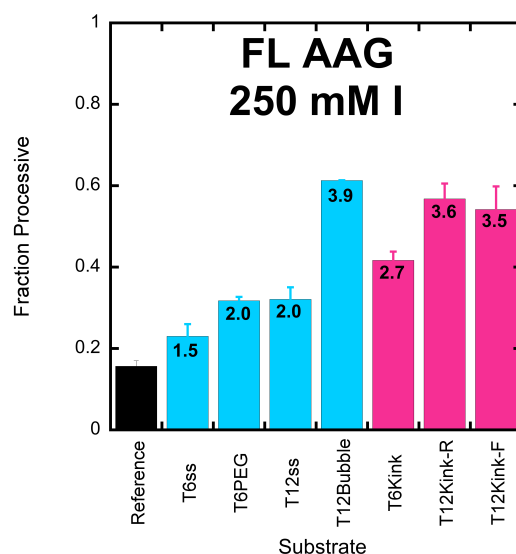


**Figure 4-7: AAG bypasses structural perturbations to the B-form, double helix by hopping between the adjacent duplex regions**

Introduction of structural perturbations (red dashed lines) such as flexible regions or kinks into B-form DNA decreases the straight-line ( $r$ ) between the adjacent duplex regions or allows the duplex regions to frequently collide. Consequently, AAG (blue) is permitted to hop between the duplex arms, bypassing the structural obstacle and continuing its search for DNA damage.

During a hopping event, the probability of a protein encountering a given site after departing a defined position on the DNA decreases with the inverse of the straight-line distance ( $r$ ) between the two sites (21). Hence, transfer will be most efficient when  $r$  is minimal. We recently showed that doubling the distance between the two lesion sites of the reference substrate by inserting nonspecific, duplex DNA had no effect on the processivity of AAG across a wide range of ionic strengths (8). This suggested that, upon removal of either of the two  $\epsilon$ A lesions, the probability of transferring to the remaining lesion site prior to dissociation is insensitive to relatively small changes ( $\leq 2$ -fold) in the distance ( $r$ ) between the lesion sites. When a  $90^\circ$  kink is inserted between the two  $\epsilon$ A lesion sites of the reference substrate, the end-to-end distance as well as the distance between the lesion sites is expected to decrease  $\sim 1.4$ -fold (Figure 4-3). Thus, an

increase in the fraction processive as compared to that for the reference substrate was not anticipated. However, as summarized in Figure 4-8, a significant increase (>2.7-fold) in the fraction processive was consistently observed at 250 mM ionic strength for all kinked substrates (magenta). It has been suggested that relief of phosphate-phosphate repulsion by counterion condensation may promote bending and so perhaps the bend angle is increased at higher ionic strengths such that two lesion sites are essentially juxtaposed, minimizing  $r$  (29). Further experiments are required to support this hypothesis.



**Figure 4-8: Transfer between duplex arms is promoted at higher ionic strength**

Comparison of the fraction processive for FL AAG at 250 mM ionic strength on the reference substrate (black) to substrates which each contained various structural obstacles between the two  $\epsilon$ A lesions. Data in cyan represent substrates in which a flexible region was present between the two  $\epsilon$ A lesions. Data in magenta represent substrates in which a kink ( $n = 6$  or  $12$ ) was present between the two  $\epsilon$ A lesions. See legend for specific labels. Values in each column are the relative values obtained by dividing each respective  $F_P$  value by that for the reference substrate.

For substrates containing more flexible regions, random thermally-excited bending of the flexible region should allow the outer duplex arms to collide with each other, presumably very often (6, 26). Although such movement is expected to be highly dynamic and the relative orientations of the duplex arms only transient, a significant increase (1.5- to 3.9-fold) in the fraction processive was consistently observed at 250 mM ionic strength for all substrates containing a flexible region in between the two  $\epsilon$ A lesions (cyan). This suggests that the outer duplex arms of substrates containing a flexible region collide successfully, with sufficient frequency, to allow transfer of AAG between the

outer duplex arms and removal of both  $\epsilon$ A lesions within a single binding encounter. Consistent with this model, it has also been observed that transfer between DNA molecules is promoted at higher concentrations of DNA substrate, where collisions between DNA molecules will be increased (discussed further in Chapter 5). It is also plausible that accordion-like motion brings the sites immediately adjacent to the flexible regions in close proximity to each other, allowing hopping between the duplex arms to occur (30).

It is interesting to note that at both ionic strengths tested for  $\Delta$ 80 AAG, the  $F_p$  values for the reference substrate were indistinguishable from those for substrates in either a kink or a flexible region was present between the two  $\epsilon$ A lesions. These observations entice one to speculate that the N-terminus directly contributes to transfer between the duplex arms, perhaps by a mechanism other than hopping, such as intersegmental transfer. However, as discussed above, nucleic acid structure and dynamics are highly dependent upon counterion concentration. Perhaps at an ionic strength of 250 mM, the distance between the lesion sites ( $r$ ) is minimized for substrates containing a flexible region or a kink such that the fraction processive is increased relative to the reference substrate. Such an effect would be hard if not impossible to observe for  $\Delta$ 80 AAG because  $F_p$  reaches  $\sim 0.0$  at 200 mM ionic strength for the reference substrate. Thus, these discrepancies likely reflect the different ionic strength dependencies of the fraction processive of FL and  $\Delta$ 80 AAG (8, 12). Nonetheless, additional experiments will be carried out to ascertain whether the N-terminus directly contributes to transfer between duplex segments of DNA.

In summary, our results reveal that kinks, bubbles, and single-stranded gaps ( $\geq 12$  nt) are not barriers to diffusion by AAG and that the N-terminus of AAG is not required for circumventing such obstacles. Although it remains to be seen whether AAG is capable of diffusing along single-stranded DNA, AAG is able to bypass single-stranded DNA and other flexible regions as well as kinks by transferring between the adjacent duplex regions via hopping. These observations are consistent with the previous conclusion that AAG frequently hops while diffusing along DNA and that such events enable AAG to bypass structural obstacles, which may have important implications for the search for DNA damage in a biological context (8, 9).

## References

1. Berg, O. G., Winter, R. B., and von Hippel, P. H. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory, *Biochemistry* 20, 6929-6948.
2. Winter, R. B., Berg, O. G., and von Hippel, P. H. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The Escherichia coli lac repressor--operator interaction: kinetic measurements and conclusions, *Biochemistry* 20, 6961-6977.
3. Winter, R. B., and von Hippel, P. H. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 2. The Escherichia coli repressor--operator interaction: equilibrium measurements, *Biochemistry* 20, 6948-6960.
4. Zharkov, D. O., and Grollman, A. P. (2005) The DNA trackwalkers: principles of lesion search and recognition by DNA glycosylases, *Mutat Res* 577, 24-54.
5. Terry, B. J., Jack, W. E., and Modrich, P. (1985) Facilitated diffusion during catalysis by EcoRI endonuclease. Nonspecific interactions in EcoRI catalysis, *J Biol Chem* 260, 13130-13137.
6. Halford, S. E., and Marko, J. F. (2004) How do site-specific DNA-binding proteins find their targets?, *Nucleic Acids Res* 32, 3040-3052.
7. Stanford, N. P., Szczelkun, M. D., Marko, J. F., and Halford, S. E. (2000) One- and three-dimensional pathways for proteins to reach specific DNA sites, *EMBO J* 19, 6546-6557.
8. Hedglin, M., and O'Brien, P. J. (2010) Hopping enables a DNA repair glycosylase to search both strands and bypass a bound protein, *ACS Chem Biol* 5, 427-436.
9. Blainey, P. C., Luo, G., Kou, S. C., Mangel, W. F., Verdine, G. L., Bagchi, B., and Xie, X. S. (2009) Nonspecifically bound proteins spin while diffusing along DNA, *Nat Struct Mol Biol* 16, 1224-1229.
10. Rau, D. C., and Sidorova, N. Y. (2010) Diffusion of the restriction nuclease EcoRI along DNA, *J Mol Biol* 395, 408-416.
11. Jeltsch, A., Alves, J., Wolfes, H., Maass, G., and Pingoud, A. (1994) Pausing of the restriction endonuclease EcoRI during linear diffusion on DNA, *Biochemistry* 33, 10215-10219.

12. Hedglin, M., and O'Brien, P. J. (2008) Human alkyladenine DNA glycosylase employs a processive search for DNA damage, *Biochemistry* 47, 11434-11445.
13. O'Brien, P. J., and Ellenberger, T. (2004) Dissecting the broad substrate specificity of human 3-methyladenine-DNA glycosylase, *J Biol Chem* 279, 9750-9757.
14. Lilley, D. M. (1995) Kinking of DNA and RNA by base bulges, *Proc Natl Acad Sci U S A* 92, 7140-7142.
15. Wang, Y. H., and Griffith, J. (1991) Effects of bulge composition and flanking sequence on the kinking of DNA by bulged bases, *Biochemistry* 30, 1358-1363.
16. Zacharias, M., and Hagerman, P. J. (1995) Bulge-induced bends in RNA: quantification by transient electric birefringence, *J Mol Biol* 247, 486-500.
17. Rice, J. A., and Crothers, D. M. (1989) DNA bending by the bulge defect, *Biochemistry* 28, 4512-4516.
18. Gohlke, C., Murchie, A. I., Lilley, D. M., and Clegg, R. M. (1994) Kinking of DNA and RNA helices by bulged nucleotides observed by fluorescence resonance energy transfer, *Proc Natl Acad Sci U S A* 91, 11660-11664.
19. Bhattacharyya, A., and Lilley, D. M. (1989) The contrasting structures of mismatched DNA sequences containing looped-out bases (bulges) and multiple mismatches (bubbles), *Nucleic Acids Res* 17, 6821-6840.
20. Hsieh, C. H., and Griffith, J. D. (1989) Deletions of bases in one strand of duplex DNA, in contrast to single-base mismatches, produce highly kinked molecules: possible relevance to the folding of single-stranded nucleic acids, *Proc Natl Acad Sci U S A* 86, 4833-4837.
21. Porecha, R. H., and Stivers, J. T. (2008) Uracil DNA glycosylase uses DNA hopping and short-range sliding to trap extrahelical uracils, *Proc Natl Acad Sci U S A* 105, 10791-10796.
22. Kuznetsov, S. V., Shen, Y., Benight, A. S., and Ansari, A. (2001) A semiflexible polymer model applied to loop formation in DNA hairpins, *Biophys J* 81, 2864-2875.
23. Mills, J. B., Vacano, E., and Hagerman, P. J. (1999) Flexibility of single-stranded DNA: use of gapped duplex helices to determine the persistence lengths of poly(dT) and poly(dA), *J Mol Biol* 285, 245-257.

24. Rivetti, C., Walker, C., and Bustamante, C. (1998) Polymer chain statistics and conformational analysis of DNA molecules with bends or sections of different flexibility, *J Mol Biol* 280, 41-59.
25. Murphy, M. C., Rasnik, I., Cheng, W., Lohman, T. M., and Ha, T. (2004) Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy, *Biophys J* 86, 2530-2537.
26. Halford, S. E., and Szczelkun, M. D. (2002) How to get from A to B: strategies for analysing protein motion on DNA, *Eur Biophys J* 31, 257-267.
27. Lee, C. Y., Delaney, J. C., Kartalou, M., Lingaraju, G. M., Maor-Shoshani, A., Essigmann, J. M., and Samson, L. D. (2009) Recognition and processing of a new repertoire of DNA substrates by human 3-methyladenine DNA glycosylase (AAG), *Biochemistry* 48, 1850-1861.
28. Hitchcock, T. M., Dong, L., Connor, E. E., Meira, L. B., Samson, L. D., Wyatt, M. D., and Cao, W. (2004) Oxanine DNA glycosylase activity from Mammalian alkyladenine glycosylase, *J Biol Chem* 279, 38177-38183.
29. Range, K., Mayaan, E., Maher, L. J., 3rd, and York, D. M. (2005) The contribution of phosphate-phosphate repulsions to the free energy of DNA bending, *Nucleic Acids Res* 33, 1257-1268.
30. Mazur, A. K. (2009) Analysis of accordion DNA stretching revealed by the gold cluster ruler, *Phys Rev E Stat Nonlin Soft Matter Phys* 80, 010901.

## Appendix C

Additional figures and accompanying discussion to support Chapter 4

### Comparison of $k_{\text{cat}}$ for Various Substrates

The steady state rate constant ( $k_{\text{cat}}$ ) was calculated for the experiments described in Figures 4-3 and 4-4 in Chapter 4 and the results are shown in Tables C1 and C2 as well as Figures C1 and C2. Under conditions of low ionic strength, the rate-limiting step is dissociation of AAG from the DNA. At high ionic strength, the rate-limiting step is *N*-glycosidic bond cleavage (1, 2). At each ionic strength tested,  $k_{\text{cat}}$  for all substrates containing a structural perturbation between the two  $\epsilon$ A lesions is essentially identical to that of the reference substrate for both enzyme forms. This indicates that insertion of the structural perturbations into the reference substrate does not change the rate at which AAG dissociates from DNA or excises  $\epsilon$ A.

**Table C-1: Multiple turnover rate constant ( $k_{\text{cat}}$ ) of FL AAG**

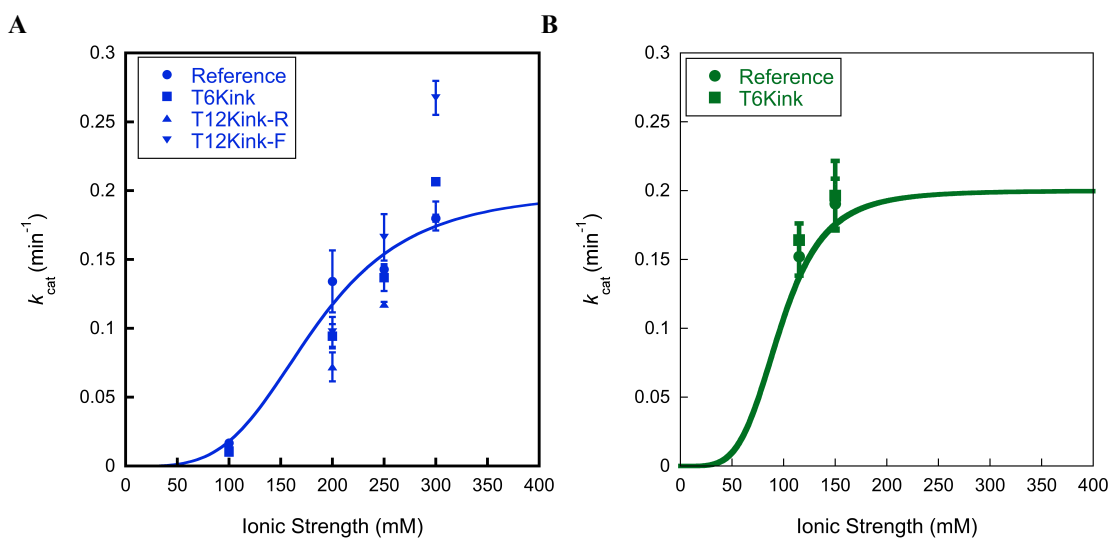
Substrate	I = 100 mM		I = 200 mM		I = 250 mM		I = 300 mM	
	$k_{\text{cat}}$	$k_{\text{cat}}(\text{rel})$	$k_{\text{cat}}$	$k_{\text{cat}}(\text{rel})$	$k_{\text{cat}}$	$k_{\text{cat}}(\text{rel})$	$k_{\text{cat}}$	$k_{\text{cat}}(\text{rel})$
Reference	0.017 ± 0.002	(1.0)	0.13 ± 0.02	(1.0)	0.14 ± 0.01	(1.0)	0.18 ± 0.01	(1.0)
T6Kink	0.011 ± 0.001	(0.63)	0.094 ± 0.009	(0.71)	0.14 ± 0.01	(0.96)	0.21 ± 0.01	(1.1)
T12Kink-R	0.0093 ± 0.0010	(0.56)	0.072 ± 0.011	(0.54)	0.12 ± 0.01	(0.82)	0.18 ± 0.01	(1.0)
T12Kink-F	0.012 ± 0.001	(0.71)	0.097 ± 0.011	(0.73)	0.17 ± 0.02	(1.2)	0.27 ± 0.01	(1.5)
T6ss	0.016 ± 0.001	(0.98)	0.12 ± 0.04	(0.89)	0.16 ± 0.01	(1.15)	0.21 ± 0.01	(1.1)
T6PEG	0.013 ± 0.001	(0.79)	0.097 ± 0.007	(0.73)	0.14 ± 0.01	(0.96)	0.20 ± 0.01	(1.1)
T12ss	0.018 ± 0.004	(1.1)	0.12 ± 0.03	(0.88)	0.14 ± 0.01	(1.0)	0.19 ± 0.01	(1.1)
T12Bubble	0.015 ± 0.004	(0.92)	0.11 ± 0.02	(0.86)	0.14 ± 0.02	(1.0)	0.19 ± 0.01	(1.0)
PEG-6		N.D.	0.12 ± 0.01	(0.92)		N.D.		N.D.
PEG-24		N.D.	0.14 ± 0.01	(1.0)		N.D.		N.D.
ss47-mer	0.023 ± 0.001		0.026 ± 0.003			N.D.	0.016 ± 0.001	
ss34-mer	0.030 ± 0.001		0.037 ± 0.005			N.D.	0.019 ± 0.001	

Multiple turnover processivity experiments were performed with full-length AAG using substrates that contained two  $\epsilon$ A lesions separated by either unbent, B-form DNA (reference) or a structural obstacle. The  $k_{\text{cat}}$  (in units of  $\text{min}^{-1}$ ) was calculated as described in Materials and Methods. Values in parentheses are reference values determined by dividing each respective  $k_{\text{cat}}$  value by that for the reference substrate for a given ionic strength condition.

**Table C-2: Multiple turnover rate constant ( $k_{cat}$ ) of  $\Delta 80$  AAG**

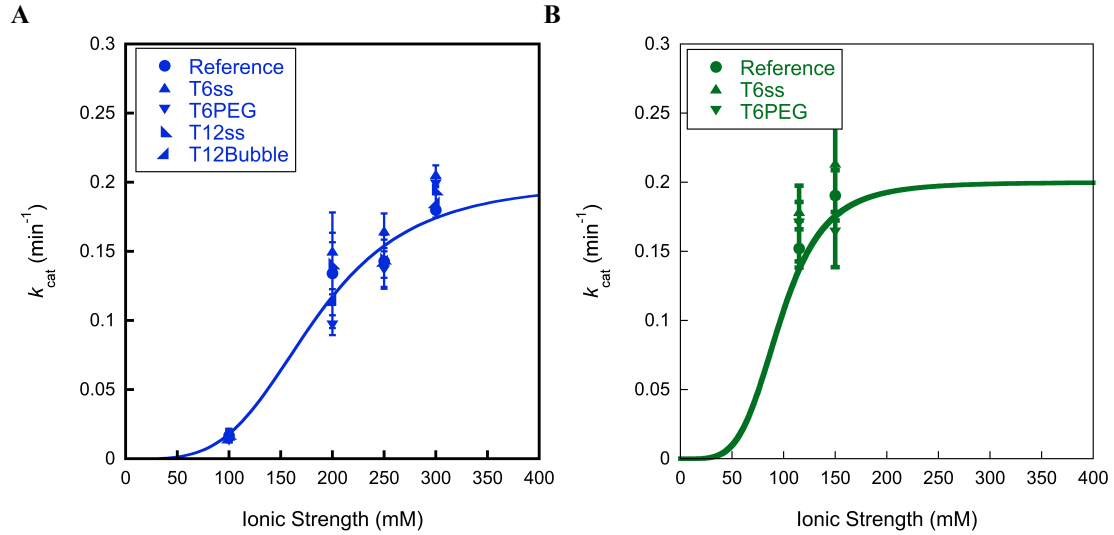
Substrate	I = 115 mM		I = 150 mM	
	$k_{cat}$	$k_{cat}(\text{rel})$	$k_{cat}$	$k_{cat}(\text{rel})$
Reference	$0.17 \pm 0.02$	(1.0)	$0.19 \pm 0.02$	(1.0)
T6Kink	$0.16 \pm 0.01$	(0.94)	$0.20 \pm 0.03$	(1.0)
T6ss	$0.15 \pm 0.03$	(0.86)	$0.21 \pm 0.04$	(1.1)
T6PEG	$0.17 \pm 0.03$	(0.98)	$0.16 \pm 0.03$	(0.86)
PEG-6	$0.18 \pm 0.01$	(1.0)	N.D.	
PEG-24	$0.19 \pm 0.01$	(1.1)	N.D.	
ss34-mer	$0.026 \pm 0.001$		N.D.	

Multiple turnover processivity experiments were performed with  $\Delta 80$  AAG using substrates that contained two  $\epsilon A$  lesions separated by either unbent, B-form DNA (reference) or a structural obstacle. The  $k_{cat}$  (in units of  $\text{min}^{-1}$ ) was calculated as described in Materials and Methods. Values in parentheses are reference values determined by dividing each respective  $k_{cat}$  value by that for the reference substrate for a given ionic strength condition.

**Figure C-1: Ionic strength dependence for multiple turnover excision of  $\epsilon A$  within kinked substrates**

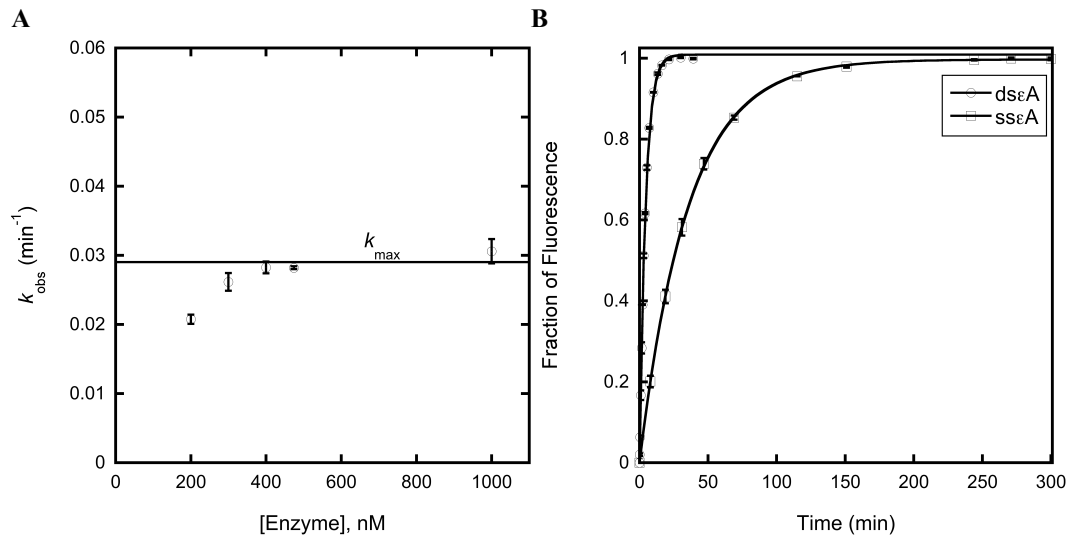
The multiple-turnover rate constants ( $k_{cat}$ ) for  $\Delta 80$  (Green) and full-length AAG (Blue) were measured at the indicated ionic strength for the reference (●), T6Kink (■), T12Kink-R (▲), and T12Kink-F (▼) substrates. The data for all substrates are comparable. The average value of two to six independent determinations is shown, and the error bars indicate the standard deviation for each substrate. Data points are overlaid on top of the ionic strength dependence of  $k_{cat}$  previously determined for each enzyme form on the reference substrate (data points removed for clarity) and fit to a cooperative model with 7 inhibitory sodium ions (1, 2).





**Figure C-2: Ionic strength dependence for multiple turnover excision of  $\epsilon$ A within substrates containing points of flexibility**

The multiple-turnover rate constants ( $k_{\text{cat}}$ ) for full-length (A) and  $\Delta 80$  AAG (B) were measured at the indicated ionic strength for the reference ( $\bullet$ ), T6ss ( $\blacktriangle$ ), T6PEG ( $\blacktriangledown$ ), T12ss ( $\blacktriangleleft$ ), and T12Bubble ( $\blacktriangleright$ ) substrates. The data for all substrates is approximately equal at each ionic strength for both enzyme forms. The average value of two to six independent determinations is shown, and the error bars indicate the standard deviation for each substrate. Data points are overlaid on top of the ionic strength dependence of  $k_{\text{cat}}$  previously determined for each enzyme form on the reference substrate (data points removed for clarity) and fit to a cooperative model with 7 inhibitory sodium ions (1, 2).



**Figure C-3: Single-turnover glycosylase activity of AAG on single- and double-stranded DNA**

(A) Concentration dependence for single-turnover glycosylase activity of AAG on  $\epsilon$ A from single-stranded DNA. Single-turnover excision of  $\epsilon$ A from the single-stranded 25-mer substrate was assessed with 100 nM DNA, and 200 – 1000 nM FL AAG. Each data point corresponds to the average and standard deviation from 2 – 4 individual reactions. Above 400 nM FL AAG, the observed rate constants ( $k_{\text{obs}}$ ) were independent of the concentration of FL AAG indicating that  $k_{\text{max}} = 0.029 \text{ min}^{-1} \pm 0.001$ . (B) Representative time course for AAG-catalyzed excision of  $\epsilon$ A with 100 nM of either single-stranded ( $\circ$ ) or double-stranded ( $\square$ ) 25-mer substrate and saturating concentration of FL AAG (200 nM for dsDNA, 400 nM for ssDNA).  $k_{\text{max}}$  for  $\epsilon$ A cleavage from double-stranded DNA (dsDNA) is  $0.227 \text{ min}^{-1} \pm 0.003$ , 7.82-fold greater than that observed for single-stranded DNA (ssDNA).

## References

1. Hedglin, M., and O'Brien, P. J. (2008) Human alkyladenine DNA glycosylase employs a processive search for DNA damage, *Biochemistry* 47, 11434-11445.
2. Hedglin, M., and O'Brien, P. J. (2010) Hopping enables a DNA repair glycosylase to search both strands and bypass a bound protein, *ACS Chem Biol* 5, 427-436.

## Chapter 5

### **The Efficiency of the Search for DNA Damage by Alkyladenine DNA Glycosylase Is Dependent Upon the Nature of the Damage**

The intrinsic reactivity of nucleobases within DNA renders them susceptible to spontaneous modification via reactions with cellular metabolites and environmental carcinogens (1). Consequently, modifications to single bases are the most frequently occurring type of DNA damage (2). If left uncorrected, some can block DNA replication and transcription and many more can alter the base pairing properties, all of which could lead to cell death or cancer. In cells, the majority of single-base modifications, commonly referred to as base lesions, are repaired by the base excision repair (BER) pathway, which is initiated by DNA glycosylases that catalyze hydrolysis of the *N*-glycosidic bond (3, 4).

In humans, the BER pathway repairs ~10,000 lesions per cell per day. Thus, DNA glycosylases have the daunting task of locating these relatively rare sites of damage among a vast excess of ~12,000,000,000 undamaged nucleotides. Almost a dozen different DNA glycosylases continuously and independently search the genome for a wide variety of oxidized or alkylated bases (5). In recent years, much effort has been focused on elucidating the mechanism by which DNA glycosylases efficiently scan the genome. Considerable *in vitro* evidence suggests that glycosylases use thermally driven linear diffusion to diffuse along DNA in a nondirectional manner, searching many adjacent sites within a single binding event. Such behavior has been demonstrated for UDG, hOGG1, MutY, and Fpg (6-11). We recently investigated the mechanism by which alkyladenine DNA glycosylase (AAG), a human DNA glycosylase responsible for the repair of a diverse set of alkylated and deaminated purines, locates damaged nucleotides. Utilizing oligonucleotide substrates containing two lesion sites, processivity experiments demonstrated that AAG locates sites of damage by diffusing along nonspecific DNA. Such studies provided valuable insight into the manner in which AAG

diffuses along DNA to locate sites of damage and showed that frequent hopping events contribute to the searching mechanism and allow AAG to simultaneously search both strands of duplex DNA and bypass tightly-bound proteins as well structural perturbations to the B-form double helix (refer to Chapter 4, and references (1, 5)). However, conclusions pertaining to how well AAG locates sites of DNA damage were limited because analysis was restricted to a single lesion. Unlike most glycosylases, which are specific for a single lesions, it is known that AAG catalyzes the excision of a broad range of modified bases with varying degrees of activity (3). In the present work, we examine the efficiency of the search for DNA damage by AAG by directly comparing the behavior of AAG on substrates containing two different lesions; hypoxanthine and 1,*N*<sup>6</sup>-ethenoadenine (Table 5-1).

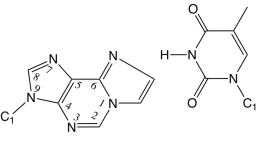
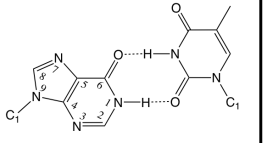
1,*N*<sup>6</sup>-ethenoadenine ( $\epsilon$ A) is a bulky, uncharged lesion that can be formed from lipid peroxidation products and exposure to exogenous agents such as urethane and vinyl chloride (12, 13).  $\epsilon$ A does not form stable base pairs with any of the normal bases because the etheno group eliminates a hydrogen bond donor at N-6 and (14, 15). Consequently, misincorporation of all possible nucleotides occurs if  $\epsilon$ A is replicated prior to repair, leading a high frequency of mutations in mammalian cells (16, 17). AAG is very adept at excising  $\epsilon$ A, with a rate enhancement and catalytic proficiency on the order of  $10^5$  and  $10^{15}$ , respectively (Table 5-1, adapted from reference 3). Our results indicate that AAG is very efficient at recognizing  $\epsilon$ A and excises it upon the initial binding encounter. Thus, in the search for DNA damage, AAG only requires one encounter with  $\epsilon$ A.

Unlike  $\epsilon$ A, hypoxanthine<sup>1</sup> (Hx) is a relatively small lesion that is formed by deamination of adenine and may occur spontaneously (hydrolytic deamination) or from exposure to nitrosating agents (18, 19). Rather than abolishing the ability to base-pair, as observed for  $\epsilon$ A, this chemical modification alters the base-pairing properties of the damaged nucleobase, converting an A:T Watson-Crick base pair to a wobble base pair (3). However, Hx can also pair with cytosine and, thus, failure to repair

---

<sup>1</sup> Inosine is a nucleoside that contains hypoxanthine as the nitrogenous base attached to the ribose ring via a *N*-glycosidic bond. Thus, when within DNA, the correct manner to refer to the damaged nucleotide is inosine. Upon release from DNA by AAG, the free base is referred to as hypoxanthine.

**Table 5-1: Comparison AAG-catalyzed excision of  $\epsilon$ A and Hx under single-turnover conditions**

	 <b><math>\epsilon</math>A T</b>	 <b>I:T</b>	<b>X-Fold Difference<sup>a</sup></b>
<b><math>K_{1/2}</math><sup>b</sup></b>	<b>2 nM</b>	<b>18 nM</b>	<b>0.11</b>
<b><math>k_{\max}</math><sup>c</sup></b>	<b>0.2 min<sup>-1</sup></b>	<b>10.8 min<sup>-1</sup></b>	<b>54</b>
<b>Catalytic Specificity<sup>d</sup></b> $k_{\max}/K_{1/2}$	<b>1.7 x 10<sup>6</sup> M<sup>-1</sup>s<sup>-1</sup></b>	<b>1.0 x 10<sup>7</sup> M<sup>-1</sup>s<sup>-1</sup></b>	<b>6</b>
<b>Rate Enhancement<sup>e</sup></b> $k_{\max}/k_{\text{non}}$	<b>2.9 x 10<sup>5</sup></b>	<b>9.0 x 10<sup>7</sup></b>	<b>320</b>
<b>Catalytic Proficiency<sup>f</sup></b> $(k_{\max}/K_{1/2})/k_w$	<b>8.0 x 10<sup>15</sup></b>	<b>2.8 x 10<sup>17</sup></b>	<b>35</b>

Adapted from reference 3. Enzymatic reactions are in standard reaction buffer at 100 mM ionic strength.

<sup>a</sup> ratio of kinetic parameter for I:T to  $\epsilon$ A:T.

<sup>b</sup>  $K_{1/2}$  is the concentration of enzyme at which  $k_{\text{st}}$  is one-half the maximal value ( $k_{\max}$ ). The  $K_{1/2}$  is analogous to  $K_M$  but this value may differ from the  $K_M$  value for multiple turnover because this can be affected by product release.

<sup>c</sup>  $k_{\max}$  is the maximal single-turnover rate constant ( $k_{\text{st}}$ ) with saturating enzyme and reports on flipping and *N*-glycosidic bond hydrolysis steps.

<sup>d</sup> The catalytic efficiency ( $k_{\max}/K_{1/2}$ ) is analogous to the specificity constant ( $k_{\text{cat}}/K_M$ ) in steady-state kinetics and reports on specificity between competing substrates.

<sup>e</sup>  $k_{\text{non}}$  is the nonenzymatic rate constant for *N*-glycosidic bond hydrolysis of a particular lesion.  $k_{\max}/k_{\text{non}}$  represents the rate enhancement and is a measurement of how well the active site environment stabilizes the transition state relative to that of the nonenzymatic reaction.

<sup>f</sup>  $k_w$  is the rate of the nonenzymatic reaction ( $k_w = k_{\text{non}}/55 \text{ M}$ ).  $(k_{\max}/K_{1/2})/k_w$  represents the catalytic proficiency, which reports on the rate enhancement to substrate binding and positioning as well as the rate acceleration of the bond cleavage step.

a Hx lesion prior to replication may lead to A:T  $\rightarrow$  G:C transition mutations (20).

Extensive biochemical characterization of various alkylated and deaminated purines revealed that AAG is most adept at excising hypoxanthine (Hx) from inosine-containing DNA, based on catalytic proficiency and rate enhancement measurements (Table 5-1). However, in contrast to that observed for  $\epsilon$ A, our results suggest that Hx is inefficiently recognized during the initial binding encounter with AAG, even though it is an excellent substrate. Thus, in order to excise a Hx lesion from a given segment of DNA, AAG must encounter the lesion multiple times. Together, these results demonstrate the efficiency of the search for DNA damage by AAG is dependent upon the identity of the lesion and

suggest that a highly redundant search may enable a broader substrate tolerance. Furthermore, preliminary results suggest that coordination with downstream BER proteins may increase the efficiency of the search for DNA damage and shelter harmful BER pathway intermediates.

## **Materials and Methods**

### **Proteins**

Full-length and truncated recombinant human AAG were purified and the concentration of active AAG was determined by burst analysis as previously described(1). Full-length human APE1 was purified and the concentration was determined as previously described(21).

### **Oligonucleotides**

DNA substrates were synthesized by Integrated DNA Technologies or the Keck Center at Yale University. Purification and characterization was carried out as described previously(1). For a given duplex substrate, the labeled strand was annealed to a 1.5- to 2-fold excess of unlabeled complementary strand in annealing buffer (10 mM NaMES, pH 6.5, 50 mM NaCl).

### **Glycosylase Activity Assay**

Reactions were carried out as previously described at 37 °C in 50 mM NaMES, pH 6.1, 1 mM EDTA, 1 mM DTT, 10% glycerol, 0.1 mg mL<sup>-1</sup> BSA and the ionic strength was adjusted with NaCl(1, 5). Reactions were initiated by adding enzyme (2 - 300 nM final concentration) to obtain a reaction volume of 50 – 225 µL that contained 100 - 2000 nM fluorescein-labeled DNA. Aliquots were withdrawn at various times and quenched with NaOH (0.2 M final concentration). Samples were heated to 70 °C for 15 minutes, formamide was added to 65%, and the DNA fragments were resolved on 14% (v/v) polyacrylamide gels containing 8 M urea as previously described(1, 5). Gels were scanned with a Typhoon Trio<sup>+</sup> fluorescence imager (GE Healthcare) to detect fluorescein (excitation at 488 nM and emission with 520BP40 filter). The resulting fluorescent signal was quantified using ImageQuant TL and corrected for background signal. The intensity of each DNA band was converted into a fraction by dividing its intensity by the sum of the intensities for all of the DNA species present. Where applicable, fractions were converted to concentrations by multiplying by the concentration of

### Multiple Turnover Kinetics

Multiple-turnover processivity assays on dual lesion substrates were carried out as previously described(1, 5). A constant ratio of 100:1 [DNA]:[FL AAG] was maintained and a 10-fold range of substrate concentration (200 - 2000 nM) was tested. For reactions containing hAPE1, DNA substrate (200 nM) was pre-incubated with hAPE1 (0 – 292 nM) for 20 minutes prior to the addition of FL AAG. To confirm that hAPE1 was inactive under the conditions employed, we performed a control in which reactions were quenched in formamide and 20 mM EDTA and mixtures directly analyzed by denaturing PAGE, as previously described (21). This established that the rate of strand nicking is at least as fast as the rate of multiple-turnover excision of Hx at 50 mM ionic strength (data not shown). At 150 mM ionic strength, no strand nicking activity was observed within 10% reaction. Initial rates were calculated from the first 10-15% of the reaction and were linear in all cases. Values for  $F_p$  were calculated from initial rates (units of fraction/time) for intermediates ( $V_i$ ) and products ( $V_p$ ) according to equation 1. For processivity experiments on  $\epsilon$ A-containing substrates,  $F_p$  was normalized to take into account substrate impurity, as previously described(5).

$$F_{p, obs} = (V_p - V_i)/(V_i + V_p) \quad (1)$$

Values for the observed rate constant for turnover ( $k_{obs}$ ,  $\text{min}^{-1}$ ) were calculated by multiplying the initial rates for the disappearance of substrate (fraction/min) by the concentration of DNA (nM) and dividing by the concentration of enzyme (nM). Values for the observed rate constants for either 5' or 3' excision ( $k_{obs}$ ,  $\text{min}^{-1}$ ) were calculated by summing the initial rates for the appearance of intermediate and product for excision at either lesion site and multiplying this sum by the concentration of DNA (nM) and dividing by the concentration of enzyme (nM).

### Pulse-Chase Assay

For measuring the efficiency of base excision ( $E$ ) for  $\epsilon$ A, a 25-mer 1, $N^6$ -etheno adenine-containing oligonucleotide 5'-(6-fam)-CGATAGCATCCTECCTTCTCTCCAT was annealed to the complementary 5'-ATGGAGAGAAGGTAGGATGCTATCG oligonucleotide with a 1.5-fold excess of the unlabeled strand. Pulse-chase experiments were then performed as previously described(22). Assays were conducted at 37°C in the standard buffer and ionic strength was held constant at 70 mM by the addition of NaCl.

In 60  $\mu\text{L}$  reaction mixtures, 100 nM fluorescein-labeled 25-mer  $\epsilon\text{A}$  substrate was mixed with 200 nM AAG for 20 s ( $t_1$ ). A chase of 5 – 20  $\mu\text{M}$  unlabeled 25-mer  $\epsilon\text{A}$  substrate was then added. If AAG dissociates from the labeled DNA before the chemical cleavage step and then binds to the unlabeled DNA, less of the reaction will occur during the single-turnover part of the curve as compared to the same experiment without chase. Samples were taken at the specified times and reactions were quenched as described above. The samples were run on sequencing gels, and the fraction product was calculated as described above. Due to  $\epsilon\text{A}$  damage during synthesis, deprotection and gel purification, the maximal values expected for the fraction of product cleaved and the amplitude ( $A$ ) is 0.97. To facilitate comparisons, we have corrected the fraction product by dividing all values by 0.97. The production of product follows a simple exponential (eq 1)

$$\text{Fraction product} = A(1 - e^{-k_{\text{obs}}t}) \quad (1)$$

Since all labeled substrate is initially bound, the efficiency of excision is the fraction of product that goes on to react rather than dissociate, which is simply the amplitude ( $A$ ) given by eq 2 where  $k_{\text{max}}$  is the maximal single turnover rate constant for formation of product, and  $k_{\text{off,obs}}$  is the macroscopic rate constant for dissociation from the stable flipped-out complex.

$$A = k_{\text{max}} / (k_{\text{off,obs}} + k_{\text{max}}) = E \quad (2)$$

Alternatively,  $k_{\text{off,obs}}$  and  $k_{\text{max}}$  can be determined directly to calculate the efficiency of base excision. For branched pathways, the observed rate constant ( $k_{\text{obs}}$ ) for the burst phase of the pulse-chase experiment is given by the sum of the rate constants for the competing pathways. Formation of product is given by  $k_{\text{max}}$ , and the macroscopic dissociation of substrate is designated  $k_{\text{off,obs}}$  (eq 3). Solving for  $k_{\text{off,obs}}$  gives eq 4.

$$k_{\text{obs}} = k_{\text{off,obs}} + k_{\text{max}} \quad (3)$$

$$k_{\text{off,obs}} = k_{\text{obs}} - k_{\text{max}} \quad (4)$$

Control reactions in which no chase was added provided the single-turnover rate constant,  $k_{\text{max}}$ , and confirmed that these concentrations of AAG were saturating. From these values, the efficiency of base excision ( $E$ ) was calculated as well. Both methods gave essentially identical values for  $E$ , and we report the average of the results obtained with both methods.



### Single-Turnover Kinetics

A 47-mer substrate containing a single 3' Hx lesion (annealed to 1.5-fold of the unlabeled complement) was incubated with excess enzyme (FL AAG) to ensure single-turnover conditions (2- to 3-fold fold molar excess of enzyme). The reaction progress curve was plotted as the fraction of product versus time and was fit by a single exponential  $F = A(1 - e^{-k_{\text{obs}}t})$ , where  $F$  is the fraction of reaction,  $A$  the amplitude,  $k_{\text{obs}}$  the observed rate constant, and  $t$  time. In all cases, the nonlinear least-squares fit was excellent ( $R^2 > 0.99$ ). At a saturating concentration of enzyme, the observed single-turnover rate constant reaches a maximum termed  $k_{\text{max}}$ . The concentration of enzyme (FL AAG) was varied to establish that the observed rate was independent of the concentration of enzyme, indicating that enzyme was in excess and at a saturating concentration (i.e.,  $k_{\text{obs}} = k_{\text{max}}$ ).

### Burst Analysis

Under low ionic strength conditions, AAG catalyzes a rapid burst of excision, followed by slow rate-limiting product release. In order to determine rate constants for the burst phase, 200 nM labeled DNA substrate was incubated with 20 nM full-length protein (10:1 labeled DNA:Enzyme) to obtain a burst amplitude of 10%. Time points covering both the initial burst and the steady state phase were taken and enzymatic activity of AAG was monitored, as described above. All reactions for a given substrate were corrected for nonenzymatic activity by normalization to control reactions in which no enzyme was added. For hybrid substrates containing both an  $\epsilon$ A lesion and a Hx lesion, the reaction progress curve was plotted as the fraction of substrate versus time and fit by eq 5 where the burst phase gives the amplitude ( $A$ ) and an observed rate constant ( $k_{\text{obs}} = k_{\text{burst}}$ ). The steady-state phase yields the steady-state rate ( $v_{\text{ss}}$ ) and is dependent upon the observed rate constant for turnover ( $k_{\text{obs}}$ ), the concentration of enzyme ( $E$ ), and the concentration of the labeled substrate.

$$\text{Fraction Substrate} = 1 - A(1 - e^{-k_{\text{obs}}t}) - (v_{\text{ss}}t) \quad (5)$$

For substrates containing only a Hx lesion, the reaction progress curve was plotted either as the fraction of substrate versus time (eq 5) or as the fraction of product versus time and fit by eq 6.

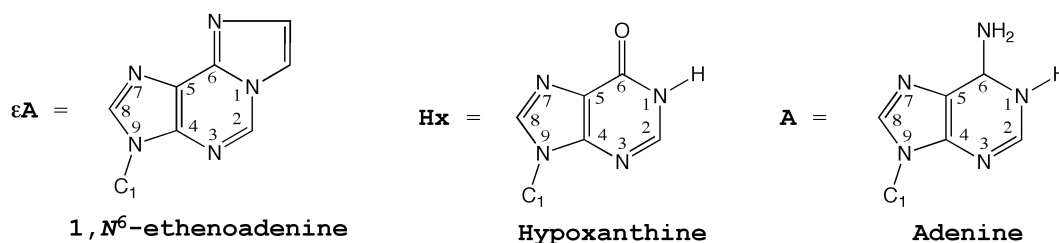
$$\text{Fraction Product} = A(1 - e^{-k_{\text{obs}}t}) + (v_{\text{ss}}t) \quad (6)$$

## Results

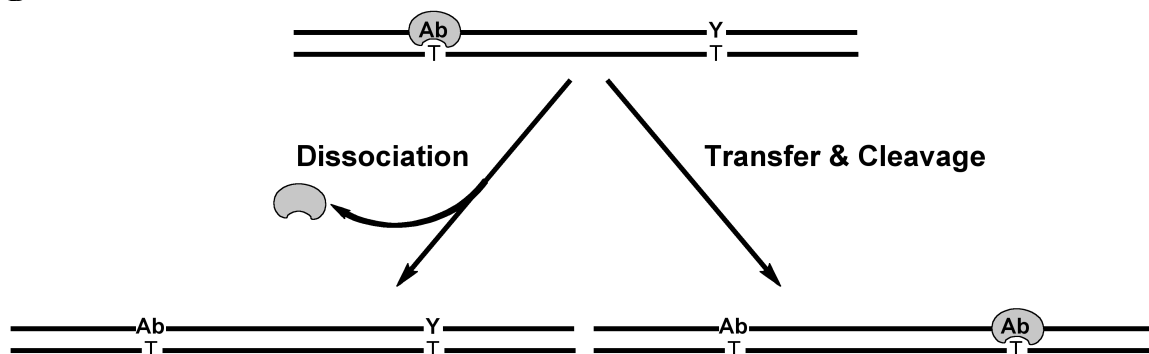
### The Processivity of AAG on Inosine-containing Substrates is Less Than That on $\epsilon$ A-containing Substrates

Multiple-turnover experiments on substrates containing two lesions allow enzymatic actions subsequent to the initial cleavage event to be monitored (Figure 5-1). Under conditions of excess substrate, a single AAG molecule will randomly bind to

A



B

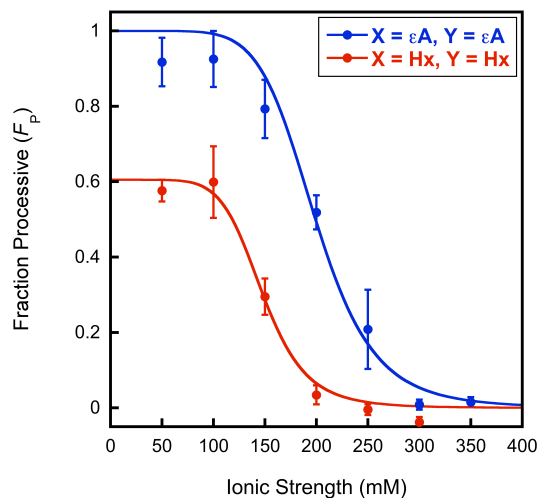


**Figure 5-1: Assays to monitor the processive behavior of AAG on substrates containing various lesions**

(A) Standard processivity assay substrate. A 47-mer oligonucleotide duplex that contains two lesions (designated as X and Y) separated by 25 base pairs. The lesion-containing strand is labeled at both 3' and 5' ends with fluorescein (designated as asterisk). The immediate sequence contexts of the two lesions are identical (underlined). (B) Multiple turnover processivity assays follow events subsequent to an initial base excision event, effectively measuring partitioning between dissociation and correlated excision at the remaining lesion site. Substrates were designed so that AAG (grey) randomly binds to and excises either of the two lesions to create an abasic site (Ab). AAG release is irreversible under these conditions due to excess substrate.

either of the two lesions. Upon *N*-glycosidic bond cleavage, an abasic site is formed, establishing AAG's occupancy at that particular site. AAG subsequently partitions between dissociation into solution and action at the second lesion site. The probability that AAG will remove the remaining lesion prior to dissociation, referred to as the

fraction of processive events ( $F_p$ ), is determined by alkaline hydrolysis of abasic sites and initial rates for single and double excision events (1, 5). Utilizing an oligonucleotide substrate containing two 1, $N^6$ -ethenoadenine ( $\epsilon$ A) lesion sites separated by 25 base pairs ( $X = \epsilon$ A,  $Y = \epsilon$ A) it was previously shown that AAG is highly processive on  $\epsilon$ A-containing DNA. As shown in Figure 5-2, at optimal pH and under low ionic strength conditions, after removing either of the two  $\epsilon$ A lesions, AAG always removes the remaining lesion prior to dissociation with  $F_p \sim 1.0$ . As the ionic strength is increased, the fraction processive decreases. Further experiments suggest this is the result of an accelerated rate of dissociation from DNA at increased ionic strengths (1, 5). Together, these results demonstrate that, after removal of an  $\epsilon$ A lesion, AAG diffuses away from the abasic product, samples other sites on the same DNA substrate, recognizes and removes the remaining  $\epsilon$ A lesion prior to dissociating. We next addressed whether this processive behavior extends to hypoxanthine excision by incorporating Hx into both lesion sites of the oligonucleotide substrate described in Figure 5-1A ( $X = \text{Hx}$ ,  $Y = \text{Hx}$ ) and monitored the fraction processive ( $F_p$ ) as a function of ionic strength; the results are summarized in Figure 5-2.



**Figure 5-2: The processive behavior of AAG is dependent upon the lesion**

Multiple turnover processivity assays were performed at optimal pH (6.1) with full-length AAG, using processivity substrates in which both lesions were either Hx (●) or  $\epsilon$ A (●). The fraction processive ( $F_p$ ) was calculated as described in Materials and Methods, and the ionic strength dependence for both substrates was fit by a cooperative model with 7 inhibitory sodium ions. Each data point reflects the mean value from two to five independent experiments, and error bars indicate one standard deviation.

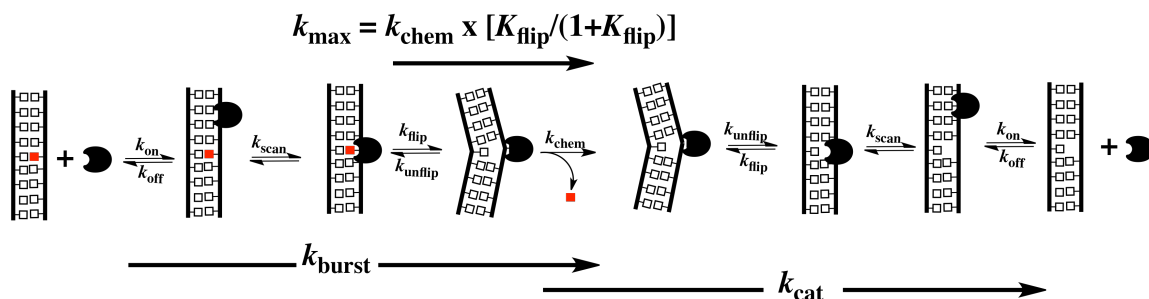
At each condition between 50 – 250 mM ionic strength, the  $F_p$  values for the substrate with two Hx lesions is dramatically lower than the  $F_p$  value on the substrate

with two  $\epsilon$ A lesions. Furthermore, in the range of 50 – 100 mM ionic strength, the fraction processive for Hx seems to be independent of ionic strength, reaching a plateau at a value significantly less than 1.0. This clearly demonstrates that the ability to remove multiple lesions per binding encounter (processivity) is decreased when both lesions are Hx as compared to that for  $\epsilon$ A. As mentioned previously, the fraction processive ( $F_p$ ) calculates the probability that AAG will remove the remaining lesion prior to dissociation and reports on all steps in between the two cleavage events; diffusion away from the abasic product, transfer to the remaining lesion site, and lesion site recognition (base flipping). A negative impact on any intermediate step(s) between the two cleavage events could then decrease the processive behavior of AAG. Regardless of the identity of the initial lesion removed within the processivity substrate ( $\epsilon$ A or Hx), the same product is formed upon *N*-glycosidic bond hydrolysis; an abasic site across from a thymine (Figure 5-1B). Thus, it is unlikely that diffusion of AAG away from an abasic site or translocation along the intervening nonspecific DNA would be influenced by the identity of the remaining lesion site 25 base pairs away. This suggests that, compared to  $\epsilon$ A, AAG is not efficient at recognizing Hx. Thus, multiple encounters with a Hx lesion are required in order for excision to occur. Such nonproductive encounters lead to a decreased processivity, as compared to that observed for  $\epsilon$ A. Below we describe assays designed to report on the efficiency of recognition by AAG.

### **The Efficiency of Recognition for $\epsilon$ A and Hx are Different**

Structural and biochemical studies have provided considerable insight into the mechanism by which AAG recognizes and excises damaged bases and have suggested the minimal mechanism illustrated in Figure 5-3 [adapted from (22)]. Due to the low frequency with which sites of DNA damage occur, initial binding will most likely occur at an undamaged site. AAG then scans DNA in search of sites of damage via non-directional, diffusion along nonspecific DNA (1, 5). Once a damaged nucleotide is encountered, it must be flipped out of the duplex by 180° into the enzyme active site where lesion recognition and *N*-glycosidic bond cleavage can occur (23-25). The excised base is released and the abasic nucleotide disengages from the active site in the reverse of the flipping step. Nonspecific binding interactions then allow AAG to diffuse away from the abasic product in search of additional sites of damage, eventually being released into

solution. The efficiency of recognition ( $E$ ) represents the extent to which enzyme-bound substrate is committed to base excision as opposed to being released into solution (10). In order to measure  $E$  for  $\epsilon$ A, we carried out a pulse-chase assay, as described previously and diagrammed in Figure 5-4A (10, 22). In these experiments, labeled substrate is briefly incubated with a small excess of AAG to allow for binding and flipping to equilibrate ( $t_1$ ). Under these conditions, the maximal single turnover rate constant

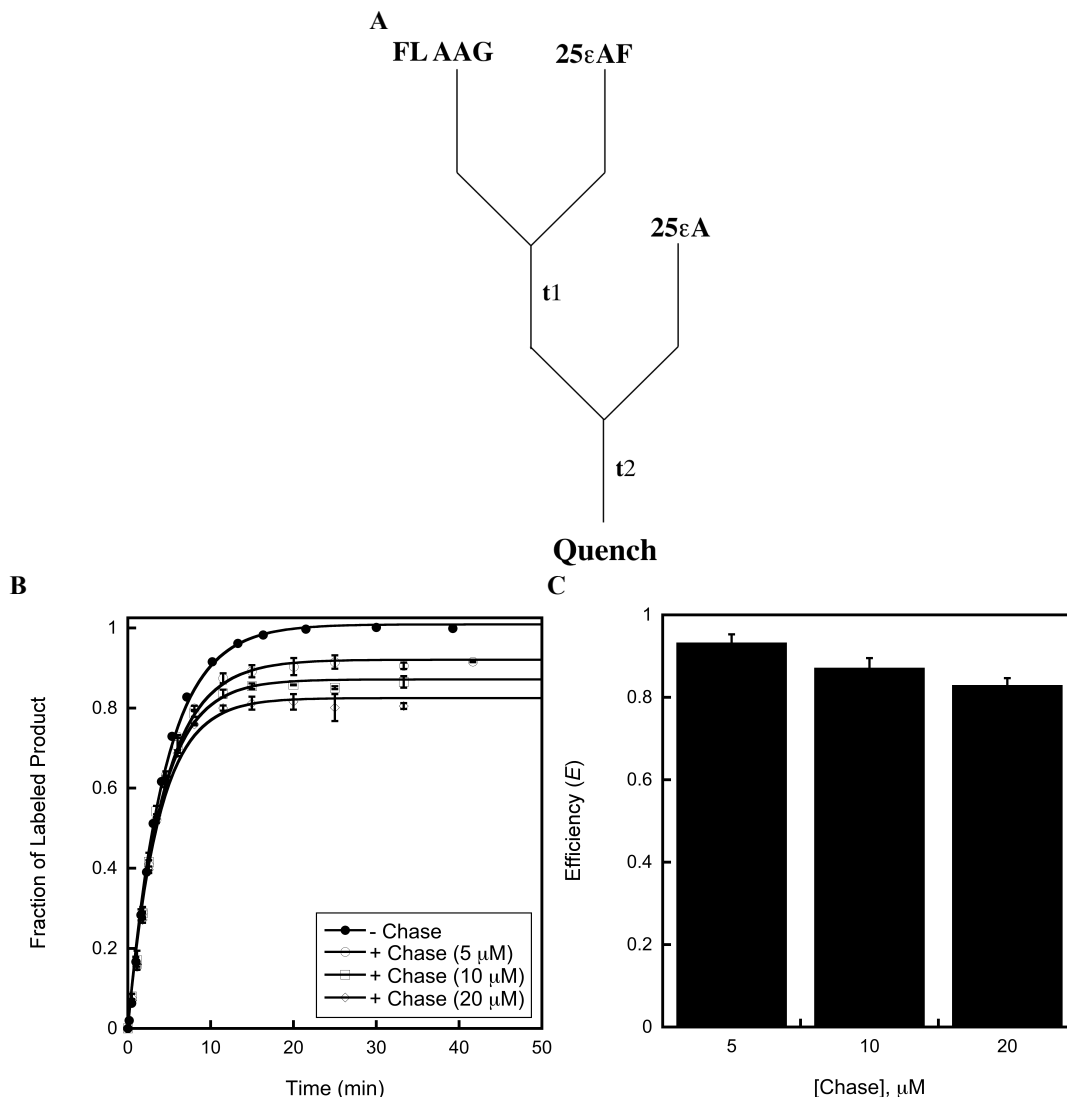


**Figure 5-3: Minimal mechanism for AAG-catalyzed base excision**

Adopted from reference 19. AAG (black crescent) binds nonspecifically to DNA and rapidly scans in search of DNA damage. Once a lesion (red rectangle) is encountered in an initial damage recognition complex, it can be flipped out into the active to form the specific recognition complex. From this specific complex, AAG catalyzes hydrolysis of the  $N$ -glycosidic bond ( $k_{\text{chem}}$ ). The lesion base is subsequently released ( $k_{\text{release}}$ ). Diffusion away from the abasic product requires rotation of the abasic sugar back into the duplex ( $k_{\text{unflip}}$ ). From the nonspecifically and less tightly bound complex, AAG can translocate along the DNA in search of additional sites of damage, eventually dissociating into solution ( $k_{\text{off}}$ ).

( $k_{\max}$ ) is  $0.2 \text{ min}^{-1}$  (3, 4). Thus, minimal  $\epsilon$ A excision occurs during  $t_1$  ( $<0.065$ ). Excess unlabeled substrate is then added as chase and samples are taken over the expected time course for single-turnover excision to evaluate the partitioning of bound complex between  $N$ -glycosidic bond hydrolysis and dissociation. Figure 5-4B shows the  $\epsilon$ A excision reaction in the presence and absence of chase. In the absence of chase,  $k_{\text{obs}} = 0.23 \text{ min}^{-1} \pm 0.002$ , in excellent agreement with  $k_{\max}$  for this substrate, indicating that the concentration of AAG was saturating (3). The addition of chase caused a minor decrease in the fraction of substrate hydrolyzed and a small increase in the observed rate constant, as expected if the forward rate constant for excision is greater than the rate constant for dissociation. Under these conditions, the efficiency of recognition for  $\epsilon$ A ranged from  $0.93 \pm 0.02$  to  $0.83 \pm 0.02$ , depending on the concentration of chase (Figure 5-4C), demonstrating that AAG is highly efficient at recognizing and excising  $\epsilon$ A, as observed previously (5). It is interesting to note that as the concentration of chase is increased, a minor decrease in the efficiency of excision was consistently observed. This

could indicate that insufficient chase was used. However, under all conditions, a single-turnover of substrate was completed within 25 minutes and no further glycosylase activity was observed up to >40 minutes. This raises the alternative possibility that AAG is capable of transferring between DNA substrates and that, as the concentration of DNA



**Figure 5-4: Pulse-chase experiment for measuring the efficiency of lesion recognition ( $E$ ) for  $\epsilon A$**   
 (A) Experimental design. Fluorescein (F)-labeled DNA (100 nM 25-mer  $\epsilon A$  substrate) was mixed with excess AAG (200 nM) for 20 s (incubation time,  $t_1$ ), and then 5 – 20  $\mu M$  unlabeled 25-mer  $\epsilon A$  substrate was added as chase. The reactions were quenched at the indicated time points ( $t_2$ ), and the fraction of abasic DNA product was determined by alkaline hydrolysis and gel electrophoresis (see Materials and Methods for additional details). (B) Single-turnover excision of  $\epsilon A$  by AAG in the absence (●) or presence of chase (○ = 5  $\mu M$ , □ = 10  $\mu M$ , ◇ = 20  $\mu M$ ) under the standard reaction conditions and 70 mM ionic strength. Shown are the averages and standard deviations of reactions carried out in duplicate with full-length AAG for each concentration of chase. Lines indicate the best fit to a single exponential (eq 1). (C) The efficiency of excision ( $E$ ) at each concentration of chase was calculated from either the end point (amplitude) or the observed rate constant ( $k_{obs}$ ) in panel B. Analysis by both methods gave essentially identical values for the efficiency of excision the average of both methods is reported.

is increased, transfer to a new DNA substrate is promoted, causing a decrease in the efficiency of recognition. Such behavior is expected for a protein capable of hopping, such as AAG, and will be interrogated further in later sections (see Figure 5-7). Nonetheless, the effect is very small and it is clear that AAG is very efficient at recognizing an  $\epsilon$ A lesion.

The relatively slow excision of  $\epsilon$ A ( $k_{\max} = 0.23 \text{ min}^{-1} \pm 0.01$ ) allowed the pulse-chase experiments described above to be carried out by hand. However,  $k_{\max}$  for excision of Hx is  $\sim 54$ -fold greater than  $k_{\max}$  for excision of  $\epsilon$ A (Table 5-1). Thus, rapid mixing experiments must be performed with a quenched-flow apparatus. Such efforts are currently underway. As an alternative method for gaining insight into the efficiency of recognition for Hx, we can measure burst kinetics to directly compare the rate constant for the burst ( $k_{\text{burst}}$ ) to the single turnover rate constant ( $k_{\max}$ ).

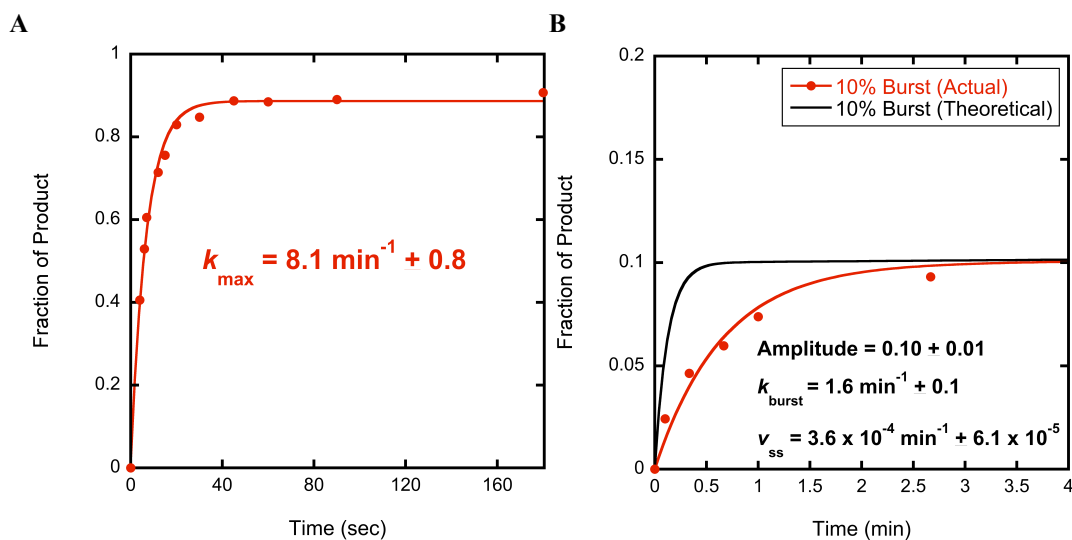
A single turnover of AAG includes all of the steps subsequent to the formation of the nonspecific complex up to an including *N*-glycosidic bond cleavage. Once AAG locates a lesion via linear diffusion, forming the initial damage recognition complex, the lesion can be flipped out into the active to form the specific recognition complex. However, due to the reversibility of the flipping step, it is likely that a given AAG molecule will also sample other base pairs or even dissociate from the DNA substrate completely upon collapse of the specific recognition complex. Under conditions of excess enzyme ( $k_{\text{obs}} = k_{\max}$ ), if a specific recognition complex collapses prior to *N*-glycosidic bond cleavage (nonproductive complex), another complex is immediately reformed with another AAG molecule such that the AAG•lesion complexes are saturated. Therefore, the complex in which the base is not yet flipped out is considered to be at equilibrium ( $K_{\text{flip}}$ ) with the complex in which the base is flipped out into the active site pocket and  $k_{\max} = k_{\text{chem}}[K_{\text{flip}} / (1 + K_{\text{flip}})]$ . However, under conditions of limiting enzyme, where a single AAG molecule interacts with a single DNA substrate, collapse of a specific recognition complex prior to *N*-glycosidic bond cleavage allows for sampling of other binding sites and dissociation into solution to occur prior to excision such that *N*-glycosidic bond hydrolysis may be limited by such futile events. Thus, if AAG is not efficient at recognizing the lesion and steps up to *N*-glycosidic bond hydrolysis are repeated multiple times prior to excision (i.e. a redundant search),  $k_{\text{burst}}$  may be less than

$k_{\max}$ . Under multiple turnover conditions at low ionic strength, the rate-limiting step for AAG-catalyzed excision of Hx is dissociation of the abasic DNA product (21). Such behavior has been exploited for determining the concentration of active AAG enzyme via pre-steady state analysis whereby a rapid pre-steady state burst of Hx excision is observed, followed by slow steady-state phase of rate-limiting product release. Therefore, we carried out experiments at low ionic strength (50 mM) and compared  $k_{\max}$  determined with excess enzyme to  $k_{\text{burst}}$  determined with limited enzyme; the results are summarized in Figure 5-5.

Under conditions of excess enzyme,  $k_{\max}$  for excision of Hx from a 47-mer substrate containing a single 3' Hx lesion (X = A, Y = Hx in Figure 5-1A) is  $8.1 \text{ min}^{-1} \pm 0.8$  (Figure 5-5A). Under conditions of limited enzyme (10:1 ratio of DNA:AAG), if AAG is efficient at recognizing Hx, a rapid burst of excision equivalent to 0.10 will be observed in the first 30 seconds of reaction, followed by a slow steady-state phase. However, as shown in Figure 5-5B, an extended burst phase is observed up to ~3 minutes, yielding  $k_{\text{burst}} = 1.6 \text{ min}^{-1} \pm 0.1$ , ~5-fold slower than  $k_{\max}$ . Distinct burst and steady-state phases are still observed because  $v_{\text{ss}}$  is > 4200-fold slower than  $k_{\text{burst}}$  and a burst amplitude (0.10) equal to that expected for 10:1 stoichiometry of [Enz]:[DNA] is observed. Furthermore, when burst analysis was carried out on  $\epsilon$ A-containing DNA,  $k_{\text{burst}} = k_{\max}$  (see Figure 5-6). An alternative possibility could also be that the concentration of DNA (200 nM) used in burst analysis was not saturating under these low ionic strength conditions (50 mM). This seems unlikely because 200 nM DNA is ~10-fold higher than both  $K_{1/2}$  and  $K_M$  determined at higher ionic strength conditions (see Table 5-1, and Appendix D, Figure D-1). Together, this suggests that AAG is not efficient at recognizing Hx and multiple encounters with this particular lesion are required in order for excision to occur. Such a search consisting of nonproductive encounters may decrease the processive behavior of AAG for Hx-containing substrates (Figure 5-2).

The data shown in Figure 5-5 could be simulated using simplified kinetic mechanisms (see Appendix D). Such simulations revealed that the equilibrium for flipping Hx lies on the unflipped side ( $K_{\text{flip}} = 0.44 \pm 0.03$ ) and that AAG is very inefficient at recognizing Hx ( $E = 0.017 \pm 0.004$ ), both in stark contrast to that observed





**Figure 5-5: Comparison of  $k_{\max}$  and  $k_{\text{burst}}$  for Hx excision reveals that AAG is not efficient at excising Hx**

(A) Determination of  $k_{\max}$  for Hx excision. Fluorescein (F)-labeled DNA containing a single 3' Hx lesion (50 nM) was mixed with excess AAG (150 - 200 nM) at optimal pH (6.1) and low ionic strength (50 mM). Shown is a representative time course for AAG-catalyzed excision of Hx with 150 nM full-length AAG.  $k_{\max}$  was calculated from single-exponential fits to the data in panel a and from additional experiments for both 150 and 200 nM full-length enzyme, as described in Materials and Methods. Values determined at both concentrations of full-length AAG agreed very well and the average and standard deviation is reported in panel a. (B) Determination of  $k_{\text{burst}}$  for Hx excision. Fluorescein (F)-labeled DNA containing a single 3' Hx lesion (200 nM) was mixed with limiting AAG (20 nM) at optimal pH (6.1) and low ionic strength (50 mM). Shown is a representative time course for the experimentally observed burst phase of AAG-catalyzed excision of Hx (red) along with the theoretical burst phase (black) observed if  $k_{\text{burst}} = k_{\max} \cdot k_{\text{burst}}$ ,  $A$ , and  $k_{\text{cat}}$  were calculated from nonlinear least-squares first to the data using eq 6 (see Materials and Methods for details). The average and standard deviation from 4 independent experiments is reported.

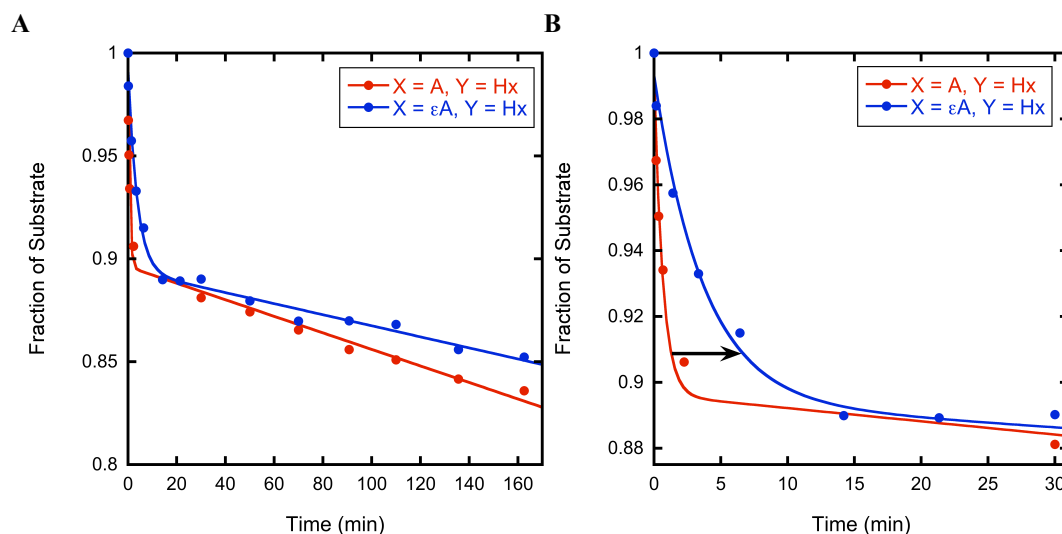
for  $\epsilon A$  (22). This simulated value for  $E$  seems low and in disagreement with the data shown in Figure 5-2 for the standard 47-mer processivity substrate, which imply that ~60% of the encounters with a Hx lesion are productive. Perhaps the efficiency of recognition for the 3' and 5' inosine lesions within this substrate are different. However, at low to moderate ionic strength (50 – 150 mM) where the fraction processive is measurable, processivity remained non-directional (data not shown). Furthermore, at high ionic strength ( $\geq 200$  mM) where  $F_p \sim 0.0$ , a minor preference (<1.4-fold) was observed for the 5' inosine lesion only at 200 and 250 mM ionic strength (data not shown). Together, this suggests that the efficiency of recognition is the same for both lesion sites within the standard 47-mer processivity substrate. Thus, additional experiments are needed to better examine the efficiency of recognition for Hx in order to provide a reliable value for  $E$ .

Collectively, the data shown in figures 5-2 to 5-5 suggest that the efficiency of recognition of  $\epsilon$ A is much greater than that for Hx. In order to directly test this, we synthesized hybrid substrates containing both an Hx and an  $\epsilon$ A lesion ( $X = \text{Hx}$ ,  $Y = \epsilon\text{A}$  or  $X = \epsilon\text{A}$ ,  $Y = \text{Hx}$ ) and measured glycosylase activity. Under low ionic strength conditions, where AAG has some probability of removing both lesions within a single binding encounter, we carried out pre-steady state analyses at a 10:1 ratio of DNA:AAG to determine which lesion AAG removes first. For hybrid substrates, there are two pathways for the depletion of substrate; excision of the  $\epsilon$ A lesion or excision of the Hx lesion. Thus, observed rate constant for the depletion of substrate ( $k_{\text{burst}}$ ) will be equivalent to the sum of the effective rate constants for excision of  $\epsilon$ A and excision of Hx. It is assumed that when a single AAG molecule initially binds to a single DNA substrate nonspecifically, there will be an equal opportunity to encounter either lesion. If the lesion initially encountered is  $\epsilon$ A, it is anticipated that AAG will catalyze *N*-glycosidic bond hydrolysis upon flipping the lesion out of the DNA duplex. Due the inefficiency of Hx excision, if the lesion initially encountered is Hx, it is expected that AAG will unproductively sample the site many times, eventually diffusing to the  $\epsilon$ A lesion and removing it prior to catalyzing excision of Hx. Thus, it is hypothesized that  $k_{\text{burst}} \sim k_{\text{max}}$  for  $\epsilon$ A excision ( $\sim 0.2 \text{ min}^{-1}$ ). The data shown in Figure 5-6 corroborates this model. When an  $\epsilon$ A lesion is present at the 5' end and a Hx lesion is present at the 3' end ( $X = \epsilon\text{A}$ ,  $Y = \text{Hx}$  in Figure 5-1), substrate is depleted 10% in the burst phase ( $A = 0.90 \pm 0.01$ ), in agreement with the ratio of [DNA]:[FL AAG] and  $k_{\text{burst}} = 0.26 \text{ min}^{-1} \pm 0.02$ . This value for  $k_{\text{burst}}$  is approximately equal to  $k_{\text{max}}$  for  $\epsilon$ A excision ( $0.2 \text{ min}^{-1}$ ) and is  $\sim 6$ -fold less than that observed for a substrate in which only a 3' Hx is present. Together, this clearly demonstrates that depletion of substrate goes thru  $\epsilon$ A excision, almost exclusively<sup>2</sup>.

As an alternative method to that described above, we also carried out multiple turnover experiments on dual lesion substrates at high ionic strength. Under these

---

<sup>2</sup> It should be noted that under these conditions, there is a 1% chance that a given DNA substrate will have two AAG molecules bound such that excision of both lesions will be independent. This may account for the small difference in  $k_{\text{burst}}$  for the hybrid substrate ( $0.26 \text{ min}^{-1} \pm 0.02$ ) and  $k_{\text{max}}$  for  $\epsilon$ A excision ( $\sim 0.2 \text{ min}^{-1}$ ).



**Figure 5-6: Under pre-steady conditions at low ionic strength, AAG removes  $\epsilon$ A prior to Hx**  
 Burst analysis was carried out by mixing 200 nM Fluorescein (F)-labeled DNA substrate with limiting AAG (20 nM) at optimal pH (6.1) and low ionic strength (50 mM). (A) Shown is a full time course encompassing the observed burst phase and steady-state phase for substrates containing either a single Hx lesion ( $\bullet$ ) or both an  $\epsilon$ A and a Hx lesion ( $\bullet$ ). (B) The burst phase. The first 10% of reaction in panel a was expanded to highlight the burst phase.  $k_{\text{burst}}$ ,  $A$ , and  $k_{\text{cat}}$  for substrate depletion were calculated from nonlinear least-squares first to the data using eq 5 (see Materials and Methods for details). The black arrow emphasizes the decrease in  $k_{\text{burst}}$  upon introduction of an  $\epsilon$ A lesion at the 5' end. The average and standard deviations for  $k_{\text{burst}}$  and  $A$  from two to four independent experiments are reported in the main text.

conditions (100:1 DNA:Enzyme, 300 mM ionic strength), AAG is completely distributive regardless of the lesion and will only excise one lesion within a single binding encounter with a given substrate (Figure 5-2). As shown in Table 5-2, on a substrate containing two Hx lesions ( $X = \text{Hx}$ ,  $Y = \text{Hx}$  in Figure 5-1), AAG removes either of the Hx lesions in a completely random fashion, having an equal probability of removing either lesion per binding encounter. However, when the 5' lesion is changed from an Hx to an  $\epsilon$ A lesion ( $X = \epsilon\text{A}$ ,  $Y = \text{Hx}$ ), the observed rate constant for depletion of substrate ( $k_{\text{obs}} = \text{rate}/[\text{E}]_0$ ) decreases  $\sim 8$ -fold from  $1.65 \pm 0.06 \text{ min}^{-1}$  to  $0.21 \pm 0.02 \text{ min}^{-1}$  and Hx excision is barely detectable. Thus, excision of  $\epsilon$ A is now almost entirely responsible for the depletion of substrate. The same behavior was observed when the location of the lesions were switched ( $X = \text{Hx}$ ,  $Y = \epsilon\text{A}$ ), collectively yielding  $\sim 30$ -fold preference for  $\epsilon$ A over Hx. It should be noted that this most likely represents a lower

limit due to substrate impurity<sup>3</sup>. Furthermore, the observed rate constants for depletion of substrate for the hybrid substrates are in excellent agreement with that observed for a substrate containing two  $\epsilon$ A lesions (X =  $\epsilon$ A, Y =  $\epsilon$ A, Table 5-2). This clearly demonstrates that under conditions where AAG can only remove one lesion per binding encounter, AAG will always remove an  $\epsilon$ A lesion rather than Hx when both are on the same substrate.

**Table 5-2: Relative rates for excision of Hx and  $\epsilon$ A on dual lesion substrates**

5' Lesion (X)	3' Lesion (Y)	$k_{\text{obs}} = \text{Rate}/[\text{E}]_{\text{o}}, \text{min}^{-1}$			<u>Rate of 5' Lesion Excision</u> Rate of 3' Lesion Excision
		Substrate Depletion	5' Lesion Excision	3' Lesion Excision	
Hx	Hx	$1.65 \pm 0.06$	$0.85 \pm 0.019$	$0.81 \pm 0.06$	1.05
$\epsilon$ A	Hx	$0.21 \pm 0.02$	$0.20 \pm 0.017$	$7.2 \times 10^{-3} \pm 6.3 \times 10^{-4}$	28.3
Hx	$\epsilon$ A	$0.21 \pm 0.01$	$8.3 \times 10^{-3} \pm 1.3 \times 10^{-3}$	$0.20 \pm 0.015$	0.042
$\epsilon$ A	$\epsilon$ A	$0.20 \pm 0.01$	$0.12 \pm 0.009$	$0.083 \pm 0.010$	1.44

Multiple turnover processivity assays were performed on dual lesion substrates with full-length AAG at optimal pH (6.1) and 300 mM ionic strength. As in Figure 5-1, “X” represents the 5’ lesion, “Y,” the 3’ lesion.  $k_{\text{obs}}$  for substrate depletion, 5’ lesion excision and 3’ lesion excision were calculated as described in Materials and Methods. Each value represents the average of at least two independent experiments with error bars indicating one standard deviation from the mean. Under these ionic strength conditions,  $k_{\text{obs}} = k_{\text{cat}}$  for  $\epsilon$ A excision because it is known that AAG is saturated (1, 5).

Collectively, the data reported in this section suggest that AAG is much more efficient at recognizing  $\epsilon$ A than Hx. When a Hx lesion is encountered, AAG has a very low probability of removing it. In stark contrast, when an  $\epsilon$ A lesion is encountered it is always removed. Thus, a Hx lesion must be presented to the active site multiple times in order to excise the lesion. This implies that the ability of AAG to excise an Hx lesion within a single binding encounter will be dependent upon the number of encounters with the lesion. In the next sections, we describe assays designed to further examine this.

### **Transfer to A New DNA Substrate Decreases Processivity of AAG on Hx-containing DNA But Has No Effect on $\epsilon$ A-containing DNA**

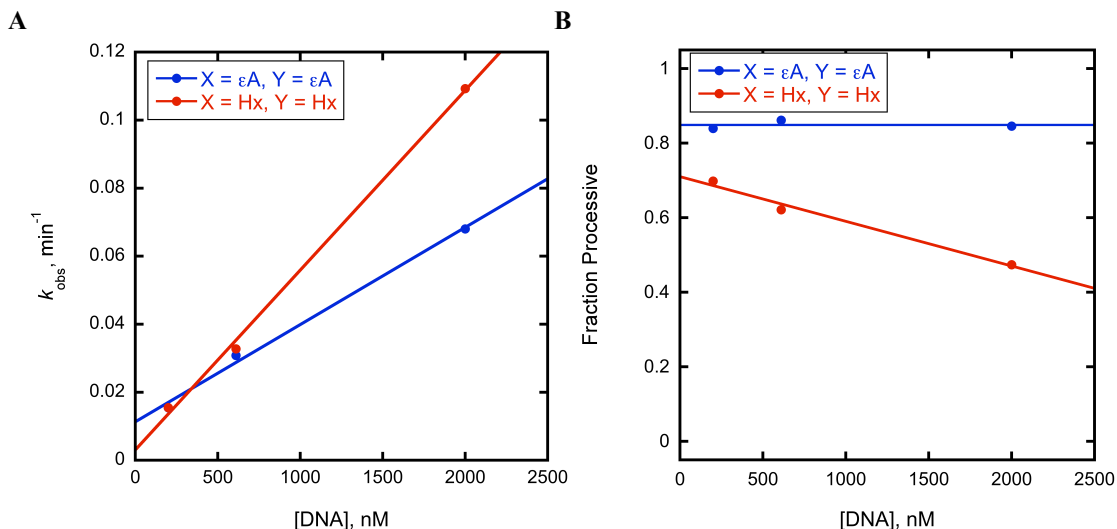
In order to locate sites of DNA damage, AAG utilizes a diffusive mechanism consisting of significant hopping events whereby successive *microscopic* dissociation-

<sup>3</sup> Characterization of the hybrid substrates revealed that ~2.3% did not contain an  $\epsilon$ A lesion due to ring-opening during synthesis and purification, in agreement with previous reports (1). For these hybrid substrates, 50 minutes was required to reach 10% reaction. Thus, for 200 nM substrate, the intermediate and product from Hx excision on substrates lacking an  $\epsilon$ A lesion will each accumulate to  $0.023 \times 0.1 \times 200 \text{ nM} = 0.46 \text{ nM}$  yielding a  $k_{\text{obs}} = 0.46 \text{ nM}/50 \text{ mins} = 9.2 \times 10^{-3} \text{ min}^{-1}$ . This value agrees very well with that observed experimentally ( $8.3 \times 10^{-3}$  and  $7.2 \times 10^{-3} \text{ min}^{-1}$ ).

association events allow the protein to traverse the DNA duplex (1, 5). For each hopping event, it is expected that the probability that the protein will reach a particular site after departing from the DNA decreases with the inverse of the straight-line distance ( $r$ ) between the initial and final positions (26). At dilute concentrations of DNA, each DNA molecule will inhabit a discrete domain that contains a high concentration of nonspecific sites. These discrete domains that contain DNA are separated by many domains that do not contain any DNA. Thus, at dilute concentrations of DNA, once a hopping protein dissociates from a DNA molecule, there is a much higher probability that it will re-associate with the same DNA molecule rather than to a different molecule (27). However, as the concentration of DNA is increased, the frequency of collisions between DNA molecules will increase and the distance separating discrete DNA molecules will decrease such that transfer to a new DNA molecule will be more likely (26, 28). Such behavior could limit the number of encounters with a lesion on a given DNA molecule and have a negative impact on the processive behavior of AAG. In order to test this, we carried out multiple turnover processivity experiments at increasing concentrations of DNA on substrates containing either two Hx lesions or two  $\epsilon$ A lesions; the results are summarized in Figure 5-7.

Under low ionic strength conditions, turnover on Hx- or  $\epsilon$ A-containing substrates is limited by dissociation from the abasic product (1, 5). Processivity experiments on dual lesion substrates at relatively low concentrations of DNA (200 nM) have demonstrated that upon formation of an abasic product at either of the two lesion sites, AAG will diffuse away from the abasic product to sample other sites (Figure 5-2). Such diffusion along nonspecific DNA provides an opportunity to remove the remaining lesion site but also to escape from the bound DNA substrate. If transfer to a new DNA is promoted at increasing concentrations of DNA, then an increase in the observed rate constant for substrate depletion ( $k_{\text{obs}} = \text{rate}/[E]_0$ ) should be observed. If such transfer to a different DNA starts to outcompete removal of the remaining lesion on the same DNA, then a decrease in the fraction processive is expected. As reported in Figure 5-7A, preliminary experiments carried out at a constant ratio of 100:1 [DNA]:[FL AAG] and 100 mM ionic strength show that  $k_{\text{obs}}$  increases linearly over a 10-fold range of DNA concentration. For a dual lesion substrate in which both lesions are Hx (X = Hx, Y =

Hx),  $k_{\text{obs}}$  increased  $\sim 7.9$ -fold. When both lesions are  $\epsilon A$  ( $X = \epsilon A$ ,  $Y = \epsilon A$ ),  $k_{\text{obs}}$  increased  $\sim 4.5$ -fold over the same range of DNA concentrations. Together, this suggests that transfer to a new DNA molecule is promoted at increasing concentrations of DNA, providing a pathway for dissociation from a DNA molecule and, thus, increasing the



**Figure 5-7: Transfer to a new DNA molecule is promoted at higher DNA concentration**

Multiple turnover processivity assays were performed at optimal pH (6.1) and 100 mM ionic strength with full-length AAG, using processivity substrates in which both lesions were either Hx (●) or  $\epsilon A$  (●). The [DNA] was varied over a 10-fold range and the ratio of [DNA]:[FL AAG] was held constant at 100:1. The fraction processive ( $F_p$ ) and observed rate constant for substrate depletion ( $k_{\text{obs}} = v_o/[E]_o$ ) were calculated as described in Materials and Methods. Each data point reflects the values obtained for a single experiment at given condition. All conditions were directly compared.

observed rate constant for substrate depletion. When the values for the fraction processive calculated from the same experiments are compared, it can be seen that such behavior has different effects on the processivity of AAG on  $\epsilon A$ - and Hx-containing substrates. When both lesions are  $\epsilon A$  ( $X = \epsilon A$ ,  $Y = \epsilon A$ ), the fraction processive is independent of the concentration of DNA in the range tested. This suggests that upon excision of either of the two  $\epsilon A$  lesions of a given DNA molecule, transfer to a new DNA molecule does not competes with removal of the remaining  $\epsilon A$  lesion on the same DNA molecule. However, the fraction processive on a substrate in which both lesions are Hx ( $X = \text{Hx}$ ,  $Y = \text{Hx}$ ) decreases in a linear fashion (1.4-fold) over the same range of [DNA]'s. Taken together, this suggests that upon removal of either of the two lesions of a given dual lesion substrate, AAG will diffuse away from the abasic product to sample other sites but transfer to new DNA molecules can occur and its frequency is proportional to the concentration of DNA. Due to a high efficiency of recognition, this has no effect

on the fraction processive for  $\epsilon$ A over this range of DNA concentrations<sup>4</sup> because AAG must transfer to the remaining  $\epsilon$ A lesion site only once in order to remove it. On the contrary, because AAG is so inefficient at recognizing Hx, multiple encounters with the remaining Hx lesion are required in order to remove it. As the concentration of DNA is increased, transfer to a new DNA molecule is promoted, limiting the number of encounters with the remaining Hx lesion and decreasing the processivity. As a result, the processivity of AAG on Hx-containing substrates is dependent upon the concentration of DNA where that for  $\epsilon$ A-containing substrates is not. In the next section we show that abasic product inhibition also has different effects on the processive behavior of AAG on  $\epsilon$ A- and Hx-containing substrates.

### **APE1 Stimulates Processive Behavior on Hx-Containing Substrates But Has No Effect on $\epsilon$ A-Containing Substrates.**

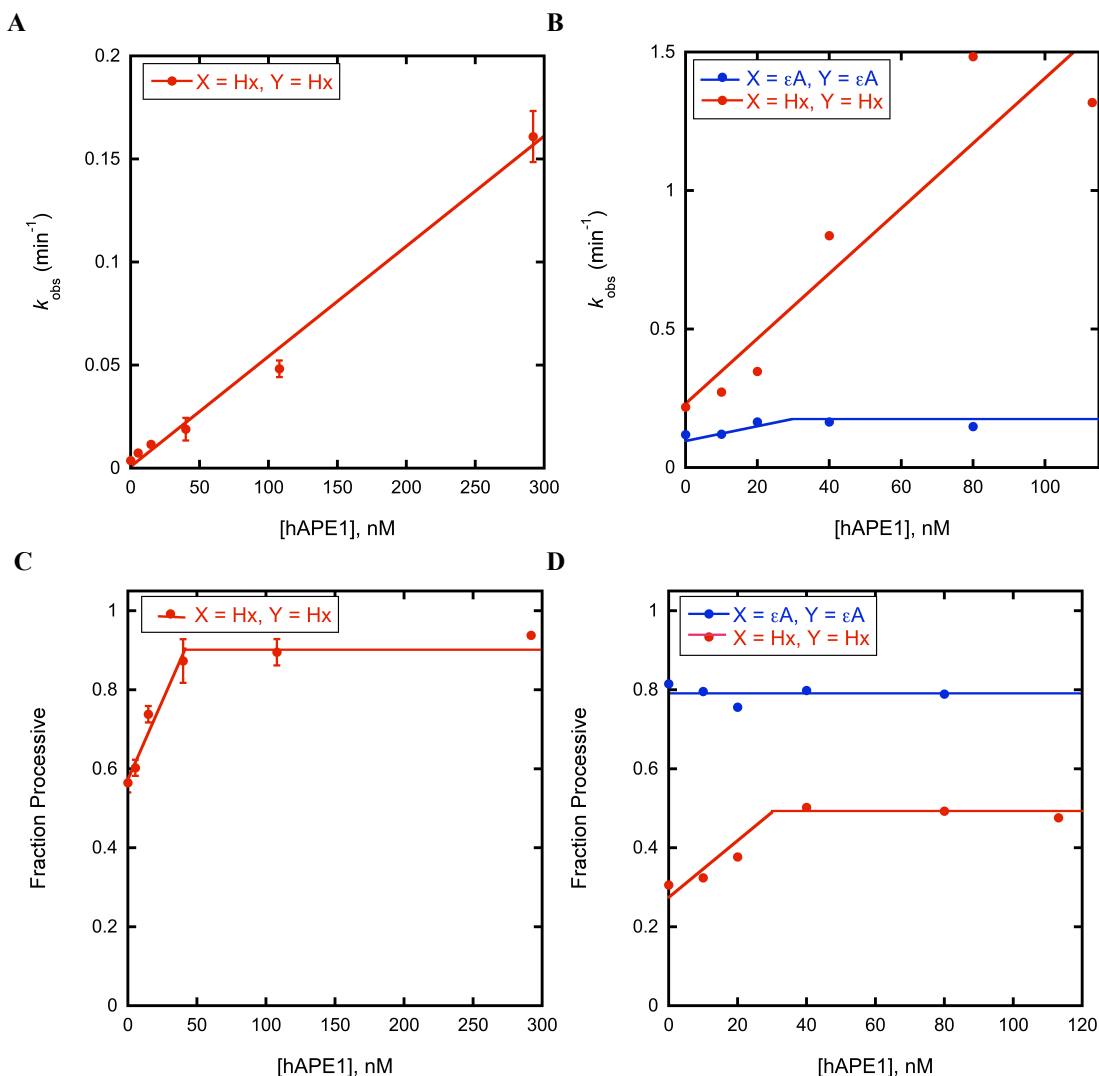
As previously mentioned, turnover on Hx- and  $\epsilon$ A-containing substrates is limited by dissociation from the abasic product under low ionic strength conditions (1, 5). Processivity experiments have clearly demonstrated that AAG will leave the abasic product, remain on the same DNA molecule, and diffuse along the duplex sampling other sites. Most often AAG will either excise another lesion or return to the abasic site. Given sufficient time, it will eventually dissociate into solution. In other reports from our lab, it has been demonstrated that when AAG diffuses away from an abasic product, the site can be captured by human AP endonuclease 1 (hAPE1), inhibiting re-binding by AAG and stimulating turnover without displacing AAG into solution (21). In this section, we wish to address whether abasic site capture by hAPE1 has any effect on the processive behavior of AAG. Below we described assays aimed to directly test this.

We monitored both the observed rate constant for substrate depletion ( $k_{\text{obs}}$ ) and the fraction processive ( $F_p$ ) of FL AAG as a function of hAPE1 concentration. As shown in Figure 5-8A, at 50 mM ionic strength, the observed rate constant for substrate depletion ( $k_{\text{obs}} = v_o/[AAG]_o$ ) increased linearly (~43-fold) over >50-fold range of [hAPE1], in agreement with that observed previously (21). Upon inspection of the

---

<sup>4</sup> It is anticipated that as the concentration of DNA is increased further (>5  $\mu$ M), transfer to a new DNA molecule will begin to compete with transfer to the remaining  $\epsilon$ A lesion and decrease the fraction processive for  $\epsilon$ A, based off the results observed in Figure 5-3

fraction processive calculated from the same experiments, it can be seen that  $F_P$  increases linearly up to 40 nM hAPE1 where it plateaus at  $\sim 0.90$  and is no longer dependent upon [hAPE1]. An  $F_P$  value of 0.90 is within error of the maximum value observed for a substrate containing two  $\epsilon A$  lesions ( $X = \epsilon A$ ,  $Y = \epsilon A$ ) at 50 mM ionic strength in the absence of hAPE1 (Figure 5-2). Furthermore, for a completely processive search ( $F_P = 1.0$ ), the concentration of abasic sites is  $\sim 40$  nM at 10% reaction ( $0.10 \times 200 \text{ nM} \times 2 = 40 \text{ nM}$ ), which is exactly the concentration of hAPE1 at which the effect on the fraction processive is saturated. It should be noted that at 50 mM ionic strength, hAPE1 strand



**Figure 5-8: Effect of APE1 on multiple-turnover excision and processivity of AAG**

Multiple turnover processivity assays on substrates containing either two Hx (●) or two  $\epsilon A$  lesions (●) were performed at optimal pH (6.1) with full-length AAG in the presence of increasing concentrations of hAPE1. Experiments were conducted either at 50 mM (A,C) or 150 mM (B,D) ionic strength. The fraction processive (B,D) and observed rate constant for substrate depletion (A,C) were calculated as described in Materials and Methods.  $k_{obs} = \text{observed rate constant for substrate depletion} = v_o/[AAG]_o$ .



nicking activity was observed that was at least as fast as the multiple-turnover reaction catalyzed by AAG even though  $Mg^{2+}$  was omitted from our reaction buffer (data not shown). However, at 150 mM ionic strength where hAPE1 strand nicking activity is absent, similar effects of hAPE1 on  $k_{obs}$  and  $F_p$  were observed for this substrate (Figures 5-8B and 5-8D)<sup>5</sup>. Also, under all conditions tested, the fraction processive remained non-directional in the presence of hAPE1 (data not shown). Together, this data supports the model for stimulation of AAG turnover by hAPE1 and demonstrates that abasic product capture by hAPE1 also increases the processive behavior of AAG on Hx-containing substrates. To address whether the same effects are observed with  $\epsilon A$ , we carried out experiments on substrates containing two  $\epsilon A$  lesions ( $X = \epsilon A$ ,  $Y = \epsilon A$ ).

Under low ionic strength conditions (50 mM), the fraction processive of AAG on substrates containing two  $\epsilon A$  lesions is maximal ( $F_p \sim 1.0$ , Figure 5-2) and, thus, any effect on the fraction processive would be hard to observe. Therefore, in preliminary experiments we only tested 150 mM ionic strength where an effect on both the fraction processive as well as  $k_{obs}$  could be observable. As shown in Figures 5-8B, preliminary experiments reveal that  $k_{obs}$  increases linearly with hAPE1 up to a value  $\sim 0.15 \text{ min}^{-1}$ , a value similar to the maximal value for  $\epsilon A$  excision under these conditions ( $0.12 \text{ min}^{-1}$ )<sup>6</sup>. This suggests that hAPE1 stimulates AAG turnover by abasic product capture, as would be expected. However, over the entire range of hAPE1 concentrations tested,  $F_p$  remained completely independent of [hAPE1], in contrast to the decreasing  $F_p$  value observed for a substrate containing two Hx lesions. Taken together, these results clearly demonstrate that relief of product inhibition by hAPE1 binding to the abasic site stimulates AAG turnover, in agreement with previous reports (21). However, these

---

<sup>5</sup> It is interesting to note that unlike that observed at 50 mM I,  $F_p$  increases to a value ( $\sim 0.5$ ) that is significantly less than that observed for a substrate containing two  $\epsilon A$  lesions in the absence of hAPE1 ( $\sim 0.8$ ). It is possible that the hAPE1 strand nicking activity observed only at 50 mM ionic strength enhances the fraction processive. Further experiments will be carried out to investigate this.

<sup>6</sup> For a completely processive enzyme ( $F_p = 1.0$ ), both lesion sites are removed prior to dissociation. Thus, the fastest multiple-turnover excision of  $\epsilon A$  can go under these conditions is  $0.1 \text{ min}^{-1}$  [ $1/k_{obs} = (1/k_{max, \text{site 1}}) + (1/k_{max, \text{site 2}}) = (1/0.2 \text{ min}^{-1}) + (1/0.2 \text{ min}^{-1}) = 10$ ]. At 150 mM ionic strength ( $F_p \sim 0.80$ ), 80% of the binding encounters are processive, and 20% are distributive. Thus,  $k_{obs} \sim 0.12 \text{ min}^{-1}$  [ $(0.8 \times 0.1 \text{ min}^{-1}) + (0.2 \times 0.2 \text{ min}^{-1})$ ].

results strongly suggest that the effect of product inhibition on the processive behavior of AAG is dependent upon the identity of the lesion.

## Discussion

Unlike most glycosylases, which are specific for a single lesions, AAG catalyzes the excision of a broad range of modified bases, among them hypoxanthine (Hx) and 1,*N*<sup>6</sup>-ethenoadenine ( $\epsilon$ A) (3). In mammalian cells, excision of  $\epsilon$ A and Hx by AAG and subsequent processing by BER is the predominant pathway for correction of these base lesions (29-32). Previous work has established that AAG locates such sites of damage by a thermally-driven diffusion mechanism consisting of frequent hopping events. In the present work, we sought to address how efficient this search is by carrying out glycosylase activity assays on substrates containing either  $\epsilon$ A or Hx. Our results indicate that the efficiency of the search for DNA damage depends on the identity of the base lesion.

Once a damaged nucleotide is encountered, it must be flipped out of the duplex by 180° into the enzyme active site where lesion recognition and *N*-glycosidic bond cleavage can occur (23-25). Pulse-chase assays demonstrated that AAG is very efficient at recognizing  $\epsilon$ A, and always excises it upon encounter (Figure 5-4). It is known that AAG has tighter binding affinity for  $\epsilon$ A compared to other lesions and crystal structures of AAG bound to  $\epsilon$ A reveal that the  $\epsilon$ A lesion fits snugly into the active site where it accepts a hydrogen bond from the backbone amide of His136 (3, 20, 33, 34). Furthermore, the inability of  $\epsilon$ A to hydrogen bond to any of the normal bases is expected to present a relatively low barrier for base flipping and, indeed, a recent report observed that flipping of  $\epsilon$ A by AAG is highly favorable, with a calculated equilibrium constant for flipping of ~1300 (22). Together this suggests that  $\epsilon$ A excision occurs through a Briggs-Haldane type mechanism where only one encounter with an  $\epsilon$ A lesion is required in order to remove it due to the high stability of the extrahelical complex (35).

In contrast to  $\epsilon$ A, AAG is very inefficient at recognizing Hx, such that steps up to *N*-glycosidic bond hydrolysis are repeated multiple times prior to excision (i.e. a redundant search), as demonstrated by burst analysis (Figure 5-5). Paradoxically, analysis of single lesion substrates has led to the conclusion that Hx is a much better

substrate than  $\epsilon A$  for AAG (see Table 5-1 for details). For example, the rate enhancement and catalytic proficiency for Hx excision are 320- and 35-fold greater than that for  $\epsilon A$ , respectively. Simulations of the data obtained from burst analysis experiments revealed that the equilibrium for flipping of Hx by AAG is unfavorable, in agreement with previous proposals (3, 34). Indeed, Hx can wobble base-pair with the opposing thymine of the characterized substrates, which is expected to be an impediment to flipping. This may contribute to the weaker binding affinity for Hx:T compared to  $\epsilon A:T$  (Table 5-1). Although an exact value for the efficiency of recognition for Hx could not be assigned, by carrying out direct competition experiments, it was demonstrated that AAG is much less efficient at recognizing Hx than an  $\epsilon A$  lesion ( $\geq 30$ -fold, Figures 5-5 and Table 5-2). Collectively, this suggests that the mechanisms for excision of Hx and  $\epsilon A$  are different.  $\epsilon A$  is excised by a Briggs-Haldane type mechanism whereas Hx is excised by a Michaelis-Menten type mechanism where the specific recognition complex is rapidly reversible due inefficient recognition (35).

It is interesting to note that the catalytic specificity of AAG ( $k_{cat}/K_M$ ) determined under single turnover conditions is 6-fold higher for Hx than  $\epsilon A$ , suggesting that, in direct competition, Hx should be excised with a 6-fold preference over  $\epsilon A$ . However, the results presented here revealed that  $\epsilon A$  is highly preferred over Hx when the lesions are on the same substrate. This intriguing discrepancy will be investigated further by directly comparing competition experiments when the lesions are either on the same or different substrates. Nonetheless, the results presented here demonstrate that, due to varying degrees of lesion recognition, the efficiency of the search for DNA damage is dependent upon the identity of the lesion. For a protein like AAG, which has exceptionally broad range of substrates, it makes sense that such a highly accommodating active site might not be able to hold on to all lesions tightly. Linear diffusion is driven by thermal fluctuations and, thus, the search for DNA damage on nonspecific DNA is completely random. Perhaps a low efficiency of recognition may enable a more broad substrate tolerance by affording multiple encounters with each lesion site. Such a redundant search would ensure that each lesion is repaired.

Processivity experiments have clearly demonstrated that, upon excision of either lesion, AAG will leave the resultant abasic product, remain on the same DNA molecule, and diffuse along the duplex sampling other sites. Most often AAG will either excise another lesion or return to the abasic site, eventually dissociating into solution. Previous and upcoming reports from our lab have demonstrated that when AAG diffuses away from an abasic product, the site can be captured by human AP endonuclease 1 (hAPE1), inhibiting re-binding by AAG and stimulating turnover without displacing AAG into solution (21). We addressed whether abasic site capture by hAPE1 influenced the processive behavior of AAG and found that the effects were dependent upon the identity of the base lesion; hAPE1 increased the processivity of AAG on Hx-containing substrates whereas it had no effect on that for  $\epsilon$ A-containing substrates (Figure 5-8). Regardless of the identity of the initial lesion removed within a dual lesion substrate, the same product is formed upon *N*-glycosidic bond hydrolysis; an abasic site opposed by a thymine (Figure 5-1B). These results suggest that the initial abasic product has different effects on the probability that AAG will remove the remaining lesion site prior to dissociation. When the remaining lesion site is Hx, the presence of the initial abasic site decreases the probability that the remaining Hx lesion will be removed. Thus, in the presence of hAPE1, the abasic site is removed and the processivity increases. However, when the remaining lesion is  $\epsilon$ A, the presence or absence of the initial abasic site has no effect on the probability of removing the remaining  $\epsilon$ A prior to dissociation. Several mechanisms can be envisioned for this effect: (i) The presence of an abasic site provides a pathway for dissociation. For Hx excision multiple encounters with the remaining lesion site are required in order to remove it. Dissociation from the initial abasic site limits the frequency of such encounters. Thus, hAPE1 binding to the initial abasic site removes a pathway for dissociation, increasing the number of encounters with the remaining lesion site, and the fraction processive. However, this pathway seems highly unlikely given the substantial evidence that *macroscopic* dissociation occurs via nonspecific DNA (21, 22). Indeed, we observe that binding to a given substrate is  $\sim$ 57-fold tighter when an abasic site analog (THF) is present (6.53 nM) than when it is absent (373 nM) (Appendix D, Figure D-1). (ii) AAG is able to discern structural abnormalities in the B-form duplex from a distance such that diffusion is directional and hierarchal, i.e.  $\epsilon$ A > Ab > Hx. Thus,

when the remaining lesion site is  $\epsilon$ A, diffusion is toward the  $\epsilon$ A lesion. When the remaining lesion site is Hx, diffusion is towards the abasic site such that transfer to the remaining Hx lesion site is impeded. Inclusion of hAPE1 shields the abasic site from AAG, promoting transfer to the remaining Hx lesion. However, reports on the propagation of structural defects in duplex DNA along with the extensive evidence for non-directional translocation of proteins utilizing linear diffusion make this pathway seem highly unlikely as well (36, 37, see chapters 3 & 4 for more refs). Nonetheless, it cannot be ruled out. (iii) AAG rarely ever diffuses away from an abasic site. The current model is that, once removed one base pair from an abasic site, AAG enters into completely non-directional diffusion and the protein has a 50% chance of diffusing back to the abasic site and a 50% chance of diffusing in the opposite direction. However, it is known that introduction of an abasic site into a particular position within DNA perturbs not only the conformation of that particular position, but also the neighboring base pairs (36). Perhaps, the abasic site has an “extended” binding landscape such that upon diffusing to the neighboring base pairs, AAG preferentially diffuses back to the abasic site and escape from the grasp of the abasic site rarely ever occurs. However, these diffusive movements located centrally around an abasic site permit transient exposure and capture of the abasic site by hAPE1, allowing AAG to escape, transfer to the remaining lesion, and excise it prior to dissociation. It makes physiological sense for glycosylases to remain bound to abasic sites until the next enzyme in the pathway (hAPE1) arrives because abasic sites themselves are potentially mutagenic. Directional diffusion back towards abasic sites should both cut down the amount of abasic sites formed per binding encounter and shelter existing abasic sites until the downstream enzyme arrives. Additional experiments will be required in order to distinguish these possibilities. Regardless of the exact pathway(s), it is clear that abasic site capture by hAPE1 both stimulates turnover and increases the processive behavior on Hx-containing substrates.

## References

1. Hedglin, M., and O'Brien, P. J. (2008) Human alkyladenine DNA glycosylase employs a processive search for DNA damage, *Biochemistry* 47, 11434-11445.
2. Lindahl, T., and Wood, R. D. (1999) Quality control by DNA repair, *Science* 286, 1897-1905.
3. O'Brien, P. J., and Ellenberger, T. (2004) Dissecting the broad substrate specificity of human 3-methyladenine-DNA glycosylase, *J Biol Chem* 279, 9750-9757.
4. O'Brien, P. J., and Ellenberger, T. (2003) Human alkyladenine DNA glycosylase uses acid-base catalysis for selective excision of damaged purines, *Biochemistry* 42, 12418-12429.
5. Hedglin, M., and O'Brien, P. J. (2010) Hopping enables a DNA repair glycosylase to search both strands and bypass a bound protein, *ACS Chem Biol* 5, 427-436.
6. Bennett, S. E., Sanderson, R. J., and Mosbaugh, D. W. (1995) Processivity of Escherichia coli and rat liver mitochondrial uracil-DNA glycosylase is affected by NaCl concentration, *Biochemistry* 34, 6109-6119.
7. Blainey, P. C., van Oijen, A. M., Banerjee, A., Verdine, G. L., and Xie, X. S. (2006) A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA, *Proc Natl Acad Sci U S A* 103, 5752-5757.
8. Francis, A. W., and David, S. S. (2003) Escherichia coli MutY and Fpg utilize a processive mechanism for target location, *Biochemistry* 42, 801-810.
9. Higley, M., and Lloyd, R. S. (1993) Processivity of uracil DNA glycosylase, *Mutat Res* 294, 109-116.
10. Porecha, R. H., and Stivers, J. T. (2008) Uracil DNA glycosylase uses DNA hopping and short-range sliding to trap extrahelical uracils, *Proc Natl Acad Sci U S A* 105, 10791-10796.
11. Sidorenko, V. S., Mechetin, G. V., Nevinsky, G. A., and Zharkov, D. O. (2008) Correlated cleavage of single- and double-stranded substrates by uracil-DNA glycosylase, *FEBS Lett* 582, 410-414.
12. Gros, L., Ishchenko, A. A., and Saparbaev, M. (2003) Enzymology of repair of etheno-adducts, *Mutat Res* 531, 219-229.

13. Speina, E., Zielinska, M., Barbin, A., Gackowski, D., Kowalewski, J., Graziewicz, M. A., Siedlecki, J. A., Olinski, R., and Tudek, B. (2003) Decreased repair activities of 1,N(6)-ethenoadenine and 3,N(4)-ethencytosine in lung adenocarcinoma patients, *Cancer Res* 63, 4351-4357.
14. Biswas, T., Clos, L. J., 2nd, SantaLucia, J., Jr., Mitra, S., and Roy, R. (2002) Binding of specific DNA base-pair mismatches by N-methylpurine-DNA glycosylase and its implication in initial damage recognition, *J Mol Biol* 320, 503-513.
15. Leonard, G. A., McAuley-Hecht, K. E., Gibson, N. J., Brown, T., Watson, W. P., and Hunter, W. N. (1994) Guanine-1,N6-ethenoadenine base pairs in the crystal structure of d(CGCGAATT(epsilon dA)GCG), *Biochemistry* 33, 4755-4761.
16. Pandya, G. A., and Moriya, M. (1996) 1,N6-ethenodeoxyadenosine, a DNA adduct highly mutagenic in mammalian cells, *Biochemistry* 35, 11487-11492.
17. Levine, R. L., Yang, I. Y., Hossain, M., Pandya, G. A., Grollman, A. P., and Moriya, M. (2000) Mutagenesis induced by a single 1,N6-ethenodeoxyadenosine adduct in human cells, *Cancer Res* 60, 4098-4104.
18. Karran, P., and Lindahl, T. (1980) Hypoxanthine in deoxyribonucleic acid: generation by heat-induced hydrolysis of adenine residues and release in free form by a deoxyribonucleic acid glycosylase from calf thymus, *Biochemistry* 19, 6005-6011.
19. Williams, D. L. H. (1988) *Nitrosation*, Cambridge University Press, Cambridge ; New York.
20. Abner, C. W., Lau, A. Y., Ellenberger, T., and Bloom, L. B. (2001) Base excision and DNA binding activities of human alkyladenine DNA glycosylase are sensitive to the base paired with a lesion, *J Biol Chem* 276, 13379-13387.
21. Baldwin, M. R., and O'Brien, P. J. (2009) Human AP endonuclease 1 stimulates multiple-turnover base excision by alkyladenine DNA glycosylase, *Biochemistry* 48, 6022-6033.
22. Wolfe, A. E., and O'Brien, P. J. (2009) Kinetic mechanism for the flipping and excision of 1,N(6)-ethenoadenine by human alkyladenine DNA glycosylase, *Biochemistry* 48, 11357-11369.
23. Roberts, R. J., and Cheng, X. (1998) Base flipping, *Annu Rev Biochem* 67, 181-198.

24. Lau, A. Y., Scharer, O. D., Samson, L., Verdine, G. L., and Ellenberger, T. (1998) Crystal structure of a human alkylbase-DNA repair enzyme complexed to DNA: mechanisms for nucleotide flipping and base excision, *Cell* 95, 249-258.
25. Lau, A. Y., Wyatt, M. D., Glassner, B. J., Samson, L. D., and Ellenberger, T. (2000) Molecular basis for discriminating between normal and damaged bases by the human alkyladenine glycosylase, AAG, *Proc Natl Acad Sci U S A* 97, 13573-13578.
26. Halford, S. E. (2001) Hopping, jumping and looping by restriction enzymes, *Biochem Soc Trans* 29, 363-374.
27. Berg, O. G., Winter, R. B., and von Hippel, P. H. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory, *Biochemistry* 20, 6929-6948.
28. Halford, S. E. (2009) An end to 40 years of mistakes in DNA-protein association kinetics?, *Biochem Soc Trans* 37, 343-348.
29. Engelward, B. P., Weeda, G., Wyatt, M. D., Broekhof, J. L., de Wit, J., Donker, I., Allan, J. M., Gold, B., Hoeijmakers, J. H., and Samson, L. D. (1997) Base excision repair deficient mice lacking the Aag alkyladenine DNA glycosylase, *Proc Natl Acad Sci U S A* 94, 13087-13092.
30. Hang, B., Singer, B., Margison, G. P., and Elder, R. H. (1997) Targeted deletion of alkylpurine-DNA-N-glycosylase in mice eliminates repair of 1,N6-ethenoadenine and hypoxanthine but not of 3,N4-ethenocytosine or 8-oxoguanine, *Proc Natl Acad Sci U S A* 94, 12869-12874.
31. Kisby, G. E., Olivas, A., Park, T., Churchwell, M., Doerge, D., Samson, L. D., Gerson, S. L., and Turker, M. S. (2009) DNA repair modulates the vulnerability of the developing brain to alkylating agents, *DNA Repair (Amst)* 8, 400-412.
32. Meira, L. B., Moroski-Erkul, C. A., Green, S. L., Calvo, J. A., Bronson, R. T., Shah, D., and Samson, L. D. (2009) Aag-initiated base excision repair drives alkylation-induced retinal degeneration in mice, *Proc Natl Acad Sci U S A* 106, 888-893.
33. Scharer, O. D., and Verdine, G. L. (1995) A Designed Inhibitor of Base-Excision DNA-Repair, *Journal of the American Chemical Society* 117, 10781-10782.
34. Vallur, A. C., Feller, J. A., Abner, C. W., Tran, R. K., and Bloom, L. B. (2002) Effects of hydrogen bonding within a damaged base pair on the activity of wild type and DNA-intercalating mutants of human alkyladenine DNA glycosylase, *J Biol Chem* 277, 31673-31678.



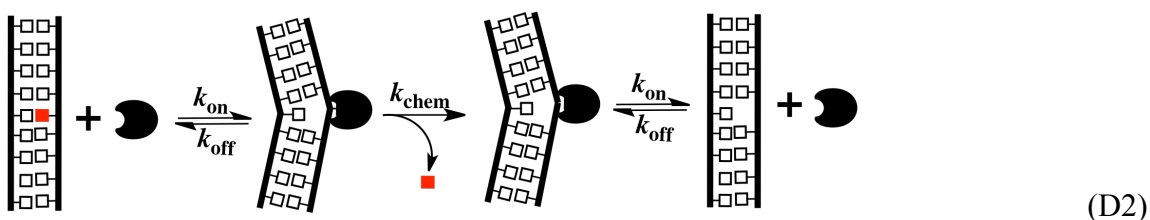
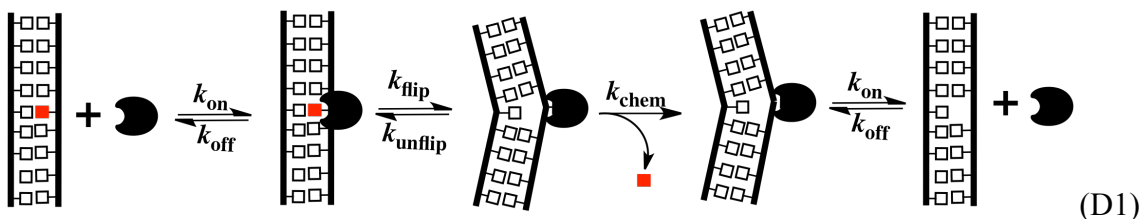
35. Fersht, A. (1999) *Structure and mechanism in protein science : a guide to enzyme catalysis and protein folding*, W.H. Freeman, New York.
36. Chen, J., Dupradeau, F. Y., Case, D. A., Turner, C. J., and Stubbe, J. (2008) DNA oligonucleotides with A, T, G or C opposite an abasic site: structure and dynamics, *Nucleic Acids Res* 36, 253-262.
37. Chen, J., Dupradeau, F. Y., Case, D. A., Turner, C. J., and Stubbe, J. (2007) Nuclear magnetic resonance structural studies and molecular modeling of duplex DNA containing normal and 4'-oxidized abasic sites, *Biochemistry* 46, 3096-3107.

## Appendix D

Additional experimental methods, figures and accompanying discussion to support Chapter 5

### Use of Berkeley Madonna to Determine $K_{\text{flip}}$ and Efficiency of Recognition ( $E$ ) for Hx

We used a simplified scheme to model the flipping equilibrium of Hx (Scheme D1). This scheme was put into Berkeley Madonna and the rate constant for *N*-glycosidic bond hydrolysis ( $k_{\text{chem}}$ ) was set at  $8.06 \text{ min}^{-1}$  for all experiments. The values of  $k_{\text{on}}$  and  $k_{\text{off}}$  for both substrate and product were allowed to vary as well as those for  $k_{\text{flip}}$  and  $k_{\text{-flip}}$ . The values that gave the best fit to the data in Figure 5-5B of the main text along with other replicates were determined.  $K_{\text{flip}}$  for Hx was determined by taking the ratio of  $k_{\text{flip}}/k_{\text{-flip}}$  and was found to be  $0.44 \pm 0.03$ . To more easily model the efficiency of lesion excision ( $E$ ) for Hx, we simplified the scheme further (Scheme D2).  $E$  for Hx was determined by taking the ratio of  $k_{\text{chem}}/(k_{\text{chem}} + k_{\text{off}})$  and was found to be  $0.014 \pm 0.004$ .



### Competitive Inhibition of Hypoxanthine Excision

The relative affinity for a substrate (Hx:T) and competitor can be readily determined by mixing together different ratios of labeled Hx-containing substrate DNA and competitor

substrates and measuring the initial rates of abasic product formation. Initial rates experiments (60  $\mu$ L reaction volume) were carried out at pH 6.1, 150 mM I on a labeled 25-mer Hx substrate in the presence of increasing concentrations of competitor (Appendix Figure D-1). 200 nM labeled substrate with 0 – 12800 nM competitor was incubated in the absence of any enzyme in 1X pH 6.1 glycosylase reaction buffer at 37°C for 10 minutes. After 10 minutes, FL AAG (2 nM final) was added to initiate the glycosylase reaction. Standard NaOH quenches were taken over time and worked up as described previously. Experiments were carried out in duplicate and data was fit to the equation for a competitive inhibitor (eq D1).  $v_o$  is the observed velocity,  $k_{cat}$  is the multiple-turnover rate constant in the absence of inhibitor,  $K_M$  is the half-maximal concentration for saturation by substrate (S), and  $K_I$  is the half-maximal concentration for binding to the inhibitor (I). Competitor substrates were unlabeled 25-mer substrates in which the central Hx:T lesion of the labeled substrate was replaced with either a THF:T, THF:G, or A:T site. THF (tetrahydrofuran, labeled as D below) is a stable abasic site analog. Below is a description of the substrate and all competitors.

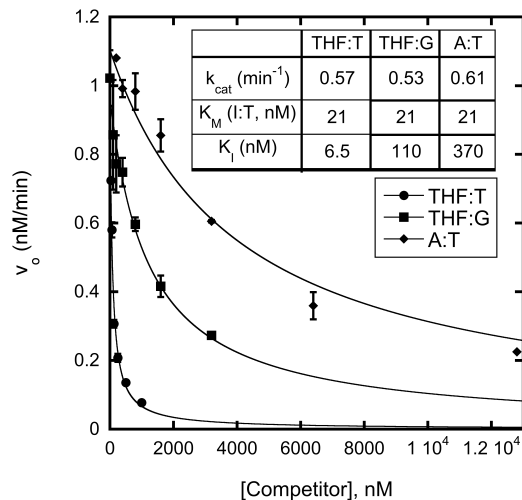
$$V_o = k_{cat}[E]_o/[K_M(1 + [I]/K_I) + [S]] \quad (D1)$$

**Substrate:** 5' -FAM-CGATAGCATCCTICCTTCTCTCCAT-3'  
3' -GCTATCGTAGGATGGAAGAGAGGTA-5'

**Competitors:** 5' -CGATAGCATCCTDCCTTCTCTCCAT-3'  
3' -GCTATCGTAGGATGGAAGAGAGGTA-5'

5' -CGATAGCATCCTDCCTTCTCTCCAT-3'  
3' -GCTATCGTAGGAGGGAAGAGAGGTA-5'

5' -CGATAGCATCCTACCTTCTCTCCAT-3'  
3' -GCTATCGTAGGATGGAAGAGAGGTA-5'



**Figure D-1: Competitive inhibition of Hx excision by AAG**

The relative affinity for the substrate and various competitors was determined for full-length AAG by measuring the initial rate of product formation for mixtures of Hx substrate and competitor and plotting the initial rate ( $v_o$ ) versus the concentration of competitor. See above for experimental details. The reactions for each condition were performed in duplicate and the error bars indicate the standard deviation from the mean. The lines indicate best fits of eq a to the data (see above) and yield  $K_i$  values of 6.5, 110, and 370 nM for THF:T, THF:G, and A:T (see inset legend).