Specificity and Searching Mechanism of Alkyladenine DNA Glycosylase

by

Yaru Zhang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemical Biology)
in the University of Michigan
2014

Doctoral Committee:

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

To my husband, Zhaoqiang

Acknowledgements

First and foremost, I would like to thank my advisor, Prof. Patrick J. O'Brien. His passion for science is always inspiring, and his patience, kindness, support and encouragement really helped me enormously throughout my whole graduate study. I am also very grateful that he understood and helped me overcome the obstacles in research and academic writing, as well as life in general as an international student. I could not have imagined a PhD journey without him for guidance. I also would like to thank the members of my committee, Prof. Bruce Palfey, Nils Walter and George Garcia, for their helpful discussions and insightful comments and suggestions on my projects over the years, as well as Prof. Jason Gestwicki, who was on my committee until his relocation to San Francisco. I would also like to thank all the members of the O'Brien lab, past and present, for helpful discussions in science and constant help in language. Special thanks to Dr. Mark Hedglin, for mentoring me when I first joined the lab and for pioneering in most of the projects presented in this thesis; to Dr. Daniel Eyler, for his great help in the yeast assay presented in Chapter 5; to Jenna Hendershot and Erin Taylor, for all the helpful discussions of science and life in general; to Michael Baldwin, for placing orders, keeping the lab well-organized and being a great help all the time. I would like to acknowledge the Chemical Biology Program and the Biological Chemistry Department, especially Laura Howe and Beth Goodwin for guiding me through the whole PhD process, and all the faculty, staff and peer graduate students for providing an energetic, nurturing, and collaborative environment to work in.

I have made so many amazing friends during my time in Ann Arbor, and I will always remember the fun time we spent together.

I want to thank my parents and siblings, whose unchanged trust and love made me a better self. Sincere thanks to my amazing husband, who made the tough decision to quit his job and accompany me in the States to support my study. I could not have made it without his understanding, encouragement and sacrifice.

Table of Contents

Chapter

List of Figures

List of Tables

List of Schemes

List of Abbreviations

3MeA: 3-methyladenine

7MeG: 7-methylguanine

A: adenine (nucleobase)

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

AP: abasic site

BER: base excision repair

BSA: bovine serum albumin

C: cytosine (nucleobase)

Δ80 AAG: N-terminal truncation mutant of AAG in which the N-terminal 79 amino acids have been deleted.

DTT: dithiothreitol

DRP: deoxyribose phosphate

E or εA: 1,N^6-ethenoadenine

E_{ex}: efficiency of excision

EDTA: ethylenediaminetetraacetic acid

FAM: 6-aminofluorescein

FL AAG: full-length alkyladenine DNA glycosylase

F_p: fraction processive

G: guanine (nucleobase)

Hx: hypoxanthine

I: deoxyinosine

MAG1: yeast 3-methyladenine DNA glycosylase

MMS: methyl methanesulfonate

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

NaMES: sodium 2-(N-morpholino)ethanesulfonate

OGG1: 8-oxoguanine DNA glycosylase

PAGE: polyacrylamide gel electrophoresis

PEG: polyethylene glycol

P_{trans}: probability of site transfer

T: thymine (nucleobase)

THF: tetrahydrofuran

U: uracil (nucleobase)

UDG: uracil DNA glycosylase

Abstract

Human alkyladenine DNA glycosylase (AAG) initiates the base excision repair pathway, and it is responsible for locating and excising alkylated and deaminated base lesions such as 1,N^6-ethenoadenine (εA) and hypoxanthine (Hx). AAG uses facilitated diffusion to search for target lesion sites via nonspecific electrostatic interactions. Previous studies have shown that AAG is capable of hopping over tightly bound roadblock proteins on DNA. This enables AAG to perform redundant searching on short DNA segments, but it has the pitfall of potentially trapping the enzyme in local regions. We show that AAG uses intersegmental transfer to balance local and global searching. Intersegmental transfer is more prominent at high DNA concentrations, suggesting an important contribution of this searching mode in the crowded nucleus. We next investigated excision efficiency of AAG on the εA substrate using pulse-chase assays to separate searching and catalysis. Results show that the commitment of AAG for εA excision is salt concentration dependent, and increasing salt concentration results in decreased excision efficiency. Furthermore, excision efficiency was found to be dependent on the chase concentration at low salt conditions, which is consistent with the aforementioned intersegmental transfer model at high chase DNA concentrations. We then compared the catalytic specificity values of AAG on εA opposing different bases and on εA•T and Hx•T lesions. At low salt, all substrates tested had similar k_{cat}/K_M values due to facilitated diffusion. At high salt, AAG excises εA opposing different bases with different k_{cat}/K_M values: εA•C ≈ εA•T > εA•A > εA•G. It was also shown that the εA•T lesion is a preferred substrate over the Hx•T lesion by a factor of over 100-fold. Finally, we tested how the *in vitro* kinetic parameters of AAG are related to its *in vivo* function. Eight single point mutants at the DNA-AAG binding interface were tested, and we showed that catalytic specificity and searching efficiency are correlated with each other, and also show a rough correlation with the *in vivo* function of AAG. This provides evidence that the search for DNA damage can limit the rate of repair in cells. Taken together, these studies shed light on the searching and catalytic mechanisms of AAG.

Chapter 1

Introduction

DNA Damage Repair Pathways

DNA is the hereditary material for all organisms from bacteria to human beings, yet its integrity suffers constant assaults by exogenous carcinogens and reactive endogenous metabolites (1). These different assaults result in various types of DNA damage, which have been shown to be involved in cell death, aging, and cancer development. Therefore, DNA repair pathways are required to find and correct damaged DNA to prevent these diseases from happening (2,3).

Damage to DNA comes in many forms, including alkylation, oxidation and hydrolysis of DNA bases, inter- and intra-strand crosslinks, and single-strand and double-strand breaks of the DNA backbone (4). Correspondingly, different DNA repair mechanisms are evolved through evolution to counteract these deleterious damages, including direct reversal repair (DRR), mismatch excision repair (MMR), base excision repair (BER), nucleotide excision repair (NER), homologous recombination (HR) and non-homologous end-joining repair (NHEJ) (Figure 1-1) (2-4). Different repair pathways have evolved to specifically act on different types of DNA damage; occasionally, certain types of damage are recognized and repaired by more than one DNA repair pathway. Together, these different repair pathways maintain the integrity of the genome.

Direct reversal repair is the direct elimination of the damaging adduct. A typical example is O^6-methylguanine DNA methyltransferase, which cleaves methyl, ethyl and benzyl adducts from guanine bases on the DNA structure (5). This protein acts on double-stranded DNA by flipping the O^6-alkylguanine out of the DNA helix into a binding pocket containing the cysteine acceptor site. This reaction is unusual and expensive in that it is stoichiometric rather than catalytic, as one protein molecule is consumed for each adduct that is removed. Overexpression of the enzyme in mice significantly decreases the rate of tumorigenesis after exposure to

alkylating agents (6), whereas loss of the gene encoding the protein increases the carcinogenic risk (7).

Mismatch repair enzymes excise incorrectly incorporated (normal) nucleotides that arise from replication error in the daughter DNA, and repair it using the parent DNA strand as the template (8,9). These enzymes can also correct a variety of base pair anomalies resulting from DNA oxidation or alkylation. Mutations in the human mismatch repair enzymes can result in microsatellite instability, and inactivation of the repair pathway can lead to hereditary nonpolyposis colorectal cancers (10).

Base excision repair involves excision and replacement of bases damaged by oxidation, alkylation or deamination. The damaged bases excised by this pathway usually have minimal impact on the structure of the DNA double helix (11,12). The deletion of downstream genes of the BER pathway in mice leads to embryonic lethality (13,14).

In contrast, the nucleotide excision repair pathway has a strong preference for bulky, helix-distorting lesions formed by UV light, environmental mutagens and chemotherapeutic agents, such as pyrimidine dimers (15). Deficiency in proteins of the nucleotide excision repair pathway leads to rare recessive syndromes such as Xeroderma pigmentosum, and patients with the disease have increased risk for skin cancer due to their deficiency in UV DNA damage repair (16).

Double-strand breaks are believed to be particularly hazardous to cells because they can lead to large-scale genome rearrangements (17,18). Homologous recombination requires the presence of an identical or nearly identical sequence to be used as a template for repair of the break; therefore, this pathway is activated when the cell is in late S/G2 phase after the DNA has been duplicated. Non-homologous end joining is used at other points of the cell cycle when sister chromatids are not available as HR templates; this repair can be accurate or it can introduce mutations during repair. Patients deficient in normal double-strand break repair are not only sensitive to ionizing radiation but are also severely immunodeficient as a result of defects in V(D)J recombination.
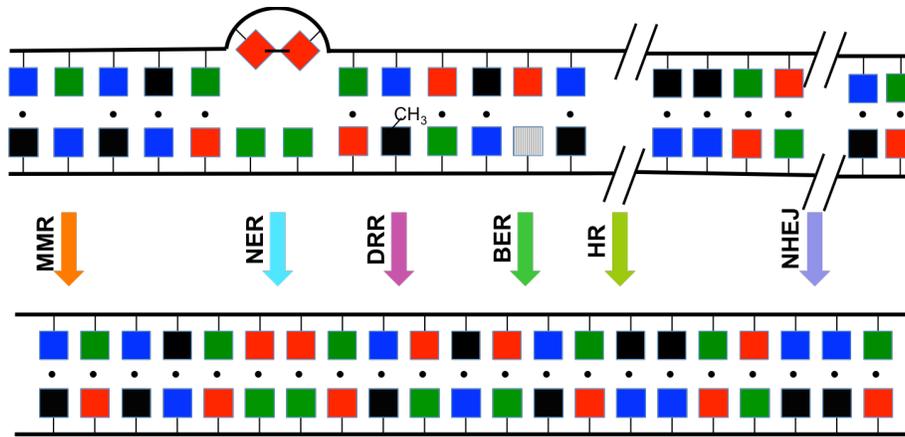
Figure 1-1. DNA damage repair pathways.
Different repair pathways coexist in cells to deal with various types of DNA damage. Repair pathways, their cognate substrates and the repair outcomes are shown. DNA bases are shown as filled boxes of different colors: thymine is shown in red, adenine in green, guanine in black and cytosine in blue. The substrates shown schematically for the different pathways are: A/C mismatch for mismatch excision repair (MMR), T-T pyrimidine dimer for nucleotide excision repair (NER), O^6-methylguanine for direct reversal repair (DRR), damaged adenine base (in grey) for base excision repair (BER), and double-strand breaks for homologous recombination (HR) and non-homologous end joining (NHEJ).

The Base Excision Repair Pathway and Alkyladenine DNA Glycosylase

It is thought that single-base lesions are the most prevalent form of DNA damage, and it is estimated that about ten thousand base lesions are formed in a typical human cell each day (19), in contrast to around ten double-strand breaks per cell per day (20). These base lesions result primarily from deamination, alkylation and oxidation, and the majority of these are repaired by the BER pathway (11). Unrepaired base lesions can lead to deleterious consequences, such as gene mutations and cell death.

The BER pathway is initiated by a DNA glycosylase, which locates the damaged bases in the context of the whole genome and removes the lesion via subsequent hydrolysis of the N-glycosidic bond (11,12). In the case of a monofunctional DNA glycosylase, the reaction products are the lesioned base and an abasic site in the DNA. The abasic site is further processed by AP endonuclease I and DNA polymerase β to nick the DNA backbone, remove the abasic sugar residue and trim the DNA ends for proper nucleotide incorporation and ligation. Repair synthesis by polymerase β uses the intact strand as a template to ensure incorporation of the correct nucleotide, and the nick is ultimately sealed by a DNA ligase (Figure 1-2) (4,12). The specificity of the BER pathway comes at the very first step, and the subsequent reactions following the abasic site formation are the same for all base lesions.

Human cells have approximately a dozen DNA glycosyases identified so far. These enzymes belong to different structural families and they catalyze the removal of different base lesions (Table 1-1) (21-23). These DNA glycosylases can be organized into two classes based on their catalytic mechanisms: monofunctional DNA glycosylase and bifunctional DNA glycosylase. Monofunctional DNA glycosylases use water as the nucleophile to attack the anomeric carbon of the damaged nucleotide, whereas bifunctional DNA glycosylases use an active site amine group (from the protein N-terminus or an active site lysine residue) to displace the damaged base and generate a Schiff base covalent enzyme intermediate. Monofunctional DNA glycosylases always generate the AP-site-containing product, while bifunctional DNA glycosylases can produce the AP-site-containing product or nicked DNA intermediates.

Table 1-1: DNA glycosylases in human cells

Enzyme	Altered base removed from DNA	Associated AP lyase
UNG	U	No
SMUG1	U, 5-hydroxymethyluracil	No
MBD4	U or T opposite G at CpG sequences, T opposite O^6-methylguanine	No
TDG	U, T or etheno-C opposite G (preferably CpG sites)	No
OGG1	Oxidized and ring-opened purines including 8-oxoG and formamidopyrimidine	Yes
MYH	A opposite 8-oxoG, 2-OH-A opposite G	Yes
NTH1	Ring-saturated or fragmented pyrimidines	Yes
AAG	3-methyladenine, 7-methylguanine, hypoxanthine, and ethenoadenine	No
NEIL1	Ring-saturated or fragmented pyrimidines; also formamidopyrimidine and 8-oxoG	Yes
NEIL2	Oxidized or fragmented pyrimidines	Yes
NEIL3	Ring-saturated or oxidized pyrimidines	Yes

Alkyladenine DNA glycosylase (AAG; also known as MPG, methylpurine DNA glycosylase) is a small monomeric protein of 298 amino acids, and it is comprised of a well-conserved catalytic domain at the C-terminus (~220 amino acids) and a poorly conserved N-terminus (the first 80 amino acids) (Figure 1-3), which contributes little to substrate recognition or to N-glycosylase activity (24-26). AAG is one of the monofunctional DNA glycosylases in humans (Table 1-1), and it is also one of the best-studied glycosylases structurally and kinetically (24,27-30). Although initially identified as a 3-methyladenine DNA glycosylase, AAG has a broad substrate range, and it also catalyzes the excision of other alkylated and deaminated purines, such as 1,N^6-ethenoadenine (εA), 7-methylguanine (7MeG) and

hypoxanthine (Hx) (Figure 1-4) (25,31-35).
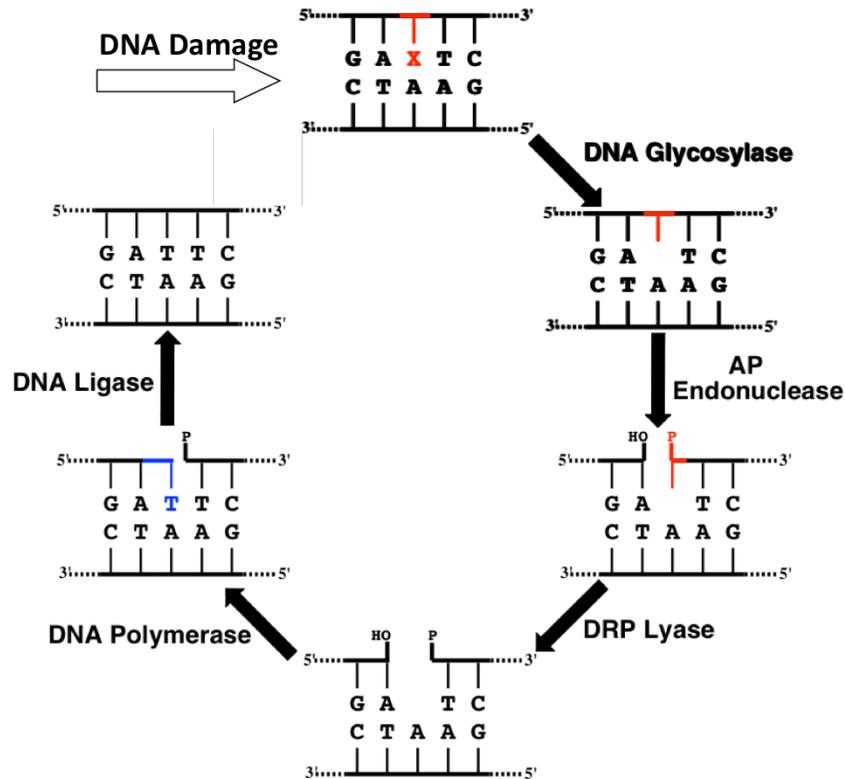
Figure 1-2. The base excision repair pathway.
A DNA glycosylase recognizes and excises the lesioned base that is denoted by X, resulting in the formation of the abasic product. The DNA backbone is subsequently nicked by AP endonuclease I 5' to the abasic site, yielding a 3'-OH and a 5'-dRP termini. The dRP lyase activity of DNA polymerase β further processes 5'-dRP to 5'-P, and its polymerase activity incorporates the right nucleotide using the complement strand as a template. Finally, DNA ligase I seals the nick and completes the BER pathway.

Consistent with this substrate specificity, the loss of AAG renders cells significantly more sensitive to methyl methanesulfonate-induced DNA damage, and to cell killing induced by alkylating agents compared to the wild type cells (31,36). Somewhat unexpectedly, the overexpression of AAG also leads to increased sensitivity to alkylating agent-induced cytotoxicity in certain contexts, possibly due to the rapid conversion of nontoxic lesions into highly toxic repair intermediates and imbalanced coordination of the downstream enzymes (37,38). Thus, it seems important for cells to keep a proper level of AAG for normal cellular functions.

AAG knockout mice were viable, fertile, and had a normal life span (31,39). Primary fibroblasts derived from these mice were only moderately sensitive to alkylating agents. This

moderate sensitivity suggested the presence of a redundant DNA glycosylase that could compensate for this activity, or the existence of another repair pathway that would repair these alkyl lesions. Indeed, it was demonstrated recently that a human dioxygenase, ABH2, could directly reverse some of the base lesions repaired by AAG (40).

The crystal structure of the AAG catalytic domain has been solved in complex with its substrate DNA containing the εA lesion (27,28). The εA base lesion is shown to be formed *in vivo* when the adenine base reacts with lipid peroxidation products (41-43). The structures reveal a positively charged DNA-AAG binding interface, and they also provide evidence that AAG uses base flipping to access and interrogate the damaged base (Figure 1-5).

AAG is proposed to use a general acid to stabilize the purine leaving group, and it activates a nucleophilic water molecule with the active site amino acid Glu125. The activity of AAG is pH dependent, and both a general acid and a general base are needed for the excision of neutral lesions, such as εA and Hx (optimal pH at 6.1) (24). In contrast, only a general base is required for the excision of positively charged substrates, such as 7-methyguanine (24). AAG also binds pyrimidine bases, but does not catalyze their excision (24,44). AAG has different levels of activity towards its different substrates, and it was proposed that AAG has evolved as a hypoxanthine glycosylase, based on its $\sim 10^{17}$ catalytic proficiency for Hx excision (29).

From extensive kinetic characterizations, the minimal kinetic mechanism of AAG on its εA substrate DNA has been proposed (Figure 1-6) (30,45). The enzyme first binds to the substrate DNA nonspecifically to form a nonspecific complex, and then rapidly scans the DNA duplex for the target lesion sites. The initial binding step has been shown to be diffusion limited and the scanning is fast and efficient. Initial recognition of the target site results in partial disruption of the base stacking, and the protein subsequently flips out the lesion base and nicely accommodates the lesion base into the protein active site. The void left by the flipped-out base is occupied by the intercalating residue Y162, which acts as a wedge to further stabilize the flipped-out conformation. The base flipping equilibrium greatly favors the flipped-out state (K_{flip} = 1300), as is also observed in the crystal structure of AAG in complex with its substrate DNA (28). The N-glycosidic bond cleavage step is much slower than the preceding steps, providing enough time and multiple opportunities for the enzyme to discriminate between damaged and undamaged nucleotides. AAG quickly releases the excised lesion base and more slowly dissociates from the abasic DNA product.

```
Human      ..............................................................
Cow        MPLGEVRSPGLPFTPAAYSPSHC.KRSSSRGLRPRWEAPAPPPNRAVLPGGLLGGLRAGV
Rat        ............................SKEPVSVVLPDAEHPA.FPGRTRRPGNA....RAGS
Mouse      MP..........ARGGSARPGRGALKPVSVTLLPDTEQPP.FLGRARRPGNA....RAGS
Dog        ..............................................................
Monkey     ..............................................................
Zebrafish  ..............................................................
```

```
                 α1           α2              β1
                ℓℓℓℓ      ℓℓℓℓℓℓℓℓ          ➙         TT      ─
               1        10        20        30        40        50
               .         .         .         .         .         .
Human      ...MVTPALQMKKPKQFCRRMGQKKQRPARAGQPHSSSDAAQAPAEQPHSSSDAAQAPCP
Cow        SGAAADRWKPTRRGAQLSCTMGQKKQRALEAK..............RHQNPSDGTQTPSP
Rat        Q..VTGSREVGQMPAPLSRKIGQKKQQLAQSEQ.......................QQTP
Mouse      L..VTGYHEVGQMPAPLSRKIGQKKQRLADSEQ.......................QQTP
Dog        .....MTVDFTPGPETLSCRMGQKRQRVLEAD..............RE.SPRGRAQTPSP
Monkey     ........MPARSGAQFSRRMGQKKQRLAKA...........EQPHSPSDTAQAPCP
Zebrafish  ..............................MTTRKRN..KPSLPESDSDKTECEAAK......RLEQCQS
```

```
               β2              TT    η1        α3       TT    β3          β4          TT
              ➙              ➙      ℓℓℓ       ℓℓℓℓ            ➙          ➙
               60        70        80        90        100       110
               .         .         .         .         .         .
Human      RERCLGPPTTPGPYRSIYFSSPKGHLTRLGLEFFDQPAVPLARAFLGQVLVRRLPNGTEL
Cow        KELHLGPPATLGPTRSIYFSSPEGRPARLGSAFFDQPAVSLARAFLGQILVRRLDDGTEL
Rat        KEKL...SSTPGLLRSIYFSSPEDRPARLGPEYFDQPAVTLARAFLGQVLVRRLADGTEL
Mouse      KERL...LSTPGLRRSIYFSSPEDHSGRLGPEFFDQPAVTLARAFLGQVLVRRLADGTEL
Dog        GEPCLGPPATLGPRRSIYFSSPKSHSAQLGSDFFDQPAVALARAFLGQVLVRRLDDGTEL
Monkey     KEPCLGPPTTLGPYRSIYFSSPKGHLTRLGLEFFDQPAVPLARAFLGQVLVRRLPNGTEL
Zebrafish  DMNTVGHNDLLHHHLSPYFSMR..DSPRLTYSFFNQPCVELAKAFLGKVLVRKLTDGTEL
```

```
               β2              TT    η1        α3       TT    β3          β4          TT
              ➙              ➙      ℓℓℓ       ℓℓℓℓ            ➙          ➙
               120       130       140       150       160       170
               .         .         .         .         .         .
Human      RGRIVETEAYLGPEDEAAHSRGGRQTPRNRGMFMKPGTLYVYIIYGMYFCMNISSQGDGA
Cow        RGRIVETEAYLGPEDEAAHSRGGRQTPRNRGMFMKPGTLYVYLIYGMYFCMNVSSRGDGA
Rat        RGRIVETEAYLGPEDEAAHSRGGRQTPRNRGMFMKPGTLYVYLIYGMYFCLNVSSQGAGA
Mouse      RGRIVETEAYLGPEDEAAHSRGGRQTPRNRGMFMKPGTLYVYLIYGMYFCLNVSSQGAGA
Dog        RGRIVETEAYLGPEDEAAHSRGGRQTPRNRGMFMKPGTLYVYIIYGMYFCLNVSSQGDGA
Monkey     RGRIVETEAYLGPEDEAAHSRGGRQTPRNRGMFMKPGTLYVYIIYGMYFCMNISSQGDGA
Zebrafish  RGKIVETEAYLGGEDKASHSAGGKRTERNTAMFMKPGTIYVYPIYGIYLCMNISSQGEGA
```

```
               β5        α4                    η2        α5          η3         TT
              ➙        ℓℓℓℓℓℓ                ℓℓℓ       ℓℓℓℓℓℓℓ      ℓℓℓ
               180       190       200       210       220       230
               .         .         .         .         .         .
Human      CVLLRALEPLEGLETMRQLRSTLRKGTASRVLKDRELCSGPSKLCQALAINKSFDQRDLA
Cow        CVLLRALEPLGGLEAMRQLRHALRKGAAGRALKDRELCNGPSKLCQALAIDRSFDQRDLA
Rat        CVLLRALEPLEGLETMRQLRNSLRKSTVGRSLKDRELCNGPSKLCQALARSKSFDQRDLA
Mouse      CVLLRALEPLEGLETMRQLRNSLRKSTVGRSLKDRELCSGPSKLCQALAIDKSFDQRDLA
Dog        CVLLRALEPLGGLETMRQLRSTLRKGTTGRALKDRELCSGPSKLCQALAIDKSFDQRDLA
Monkey     CVLLRALEPLEGLETMRQLRSTLRKGTASRVLKDRELCSGPSKLCQALAINKSFDQRDLA
Zebrafish  AVLLRSLEPLSGQDVMRGLRAAKRKP.GAKSLKDKELCNGPSKLCQALDIRRSFDRRDLA
```

```
               β6           β7          TT    α6        β8    TT  TT      α7
              ➙            ➙ .➙               ℓℓℓℓ     ➙               ℓℓℓℓ
               240       250       260       270       280       290
               .         .         .         .         .         .
Human      QDEAVWLERGPLEPS..EPAVVA.AARVCVGHAGEWARKPLRFYVRGSPWVSVVDRVAEQ
Cow        RDESVWLEQGPPEPS..EPAVVA.AARVCIGQAGEWVQKPLRFYIRGSPWVSVVDRAAER
Rat        QDEAVWLEHGPLESS..SPAVVAA..AAGIGHAGEWTQKPLRFYVQGSPWVSVVDRVAEQ
Mouse      QDDAVWLEHGPLESS..SPAVVVAAARIGIGHAGEWTQKPLRFYVQGSPWVSVVDRVAEQ
Dog        KDRAVWLERSPPGPS..EPAVVA.AARVGIGHAGEWAQKPLRFYLRGSPWVSVVDRAAEQ
Monkey     QDEAIWLERGPLEPS..EPAVVT.AARVGIGYAGEWARKPLRFYVRGSPWVSVVDRVAEQ
Zebrafish  TDVEVWLEMDPEKEAIVDAGEVVMAPRICVDSHREWATKPLRFYLRGHPCVSVLNKDAER
```

```
Human      DTQA..................
Cow        DTQTG....ARACSHKDF....
Rat        MYQPQ....QTACSDXALIVQK
Mouse      MDQPQ....QTACSEGLLIVQK
Dog        NAQA..................
Monkey     DTQV..................
Zebrafish  RMDSQSDTERTVC.........
```

Figure 1-3. Multiple sequence alignment of AAG from different species.
The amino-terminus exhibits very weak similarity both in length and sequence; in contrast, the catalytic domain is highly conserved. The alignment was performed using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/), and the final representation was generated by using ESPript 3 (espript.ibcp.fr/ESPript/cgi-bin/ESPript.cgi) (46). The crystal structure of human AAG in complex with its substrate DNA (PDB: 1F4R) (28) was used for secondary structure depiction on top of the sequence alignment.

3-Methyladenine   7-Methylguanine   1,N⁶-Ethenoadenine Hypoxanthine

Figure 1-4. A subset of base lesions excised by AAG.
AAG excises alkylated and deaminated purine bases. These base lesions can be classified as positively charged (3-methyladeine and 7-methylguanine) and neutral (1,N^6-ethenoadenine and hypoxanthine) bases.

**Figure 1-5. Crystal structure of AAG in complex with its substrate DNA.**
Electrostatic surface potential of the human AAG catalytic domain that reveals a positively charged DNA binding surface. The crystal structure of the AAG·εA-DNA complex was used to generate this figure (PDB: 1F4R) (28). Electrostatic calculations were performed with Pymol on the protein alone. A continuum from -2 (red) to +2 (blue) is shown. A view of the active site and bound DNA is shown, illustrating the positively charged DNA binding surface. Normal DNA base pairs in the DNA duplex are shown in black, and the flipped-out lesioned base in the protein active site is shown in magenta.

3-Methyladenine DNA Glycosylases in Model Organisms

DNA alkylation is a common form of damage to all organisms, and DNA glycosylases that remove 3-methyladenine DNA lesions are found in bacteria, yeast, plants, rodents, and humans (47). In *E. coli* cells, two DNA glycosylases evolved independently from AAG, Tag and AlkA (3-methyladenine DNA glycosylase I and II), are identified (23,48-50). Tag has a surprisingly high specificity for the 3MeA lesion, and is constitutively expressed, while AlkA

8

has a very broad substrate range and is inducible upon exposure of cells to low levels of DNA alkylation damage (51,52). In *S. cerevisiea*, the functional counterpart of AAG is Mag1 (an AlkA homologue), which belongs to a different structural family from AAG (53). Although these glycosylases may have evolved from different ancestors, these proteins are, at least in part, functionally complemented. Expression of the AAG protein in Tag and AlkA double deletion *E. coli* cells (54), or in Mag1 knockout yeast cells (55), protects these cells from killing by alkylating agents.

The *E. coli* Tag does not remove any normal bases, while the *E. coli* AlkA can remove all four normal bases, and the yeast MAG1 and human AAG can remove normal guanines and adenines (29,56). The unexpected activity of these glycosylases to remove innocuous bases from the genome may, under circumstances such as unregulated protein overexpression, have detrimental consequences.
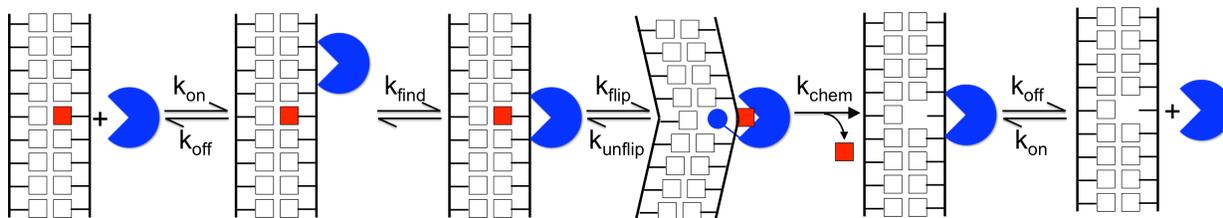
Figure 1-6. Minimal kinetic mechanism of AAG on its substrate DNA.
AAG is shown as a blue Pacman. Black open boxes denote normal DNA bases, and the red filled box represents the lesioned base. The intercalating residue Y162 is shown as a blue ball inserted in the void left by the flipped-out base. This reaction scheme is adapted from Ref (30).

Modes of Facilitated Diffusion and *in vivo* Significance

As the initiating enzyme of the BER pathway, AAG is responsible for finding the rare lesion sites in a vast majority of nonspecific DNA in the genome. This searching problem is also common to many DNA-interacting proteins, such as repressor proteins (57), DNA methyltransferases (58), restriction endonucleases (59,60) and various DNA repair proteins (61). Theoretical and experimental work has shown that many of these proteins use facilitated diffusion to efficiently locate specific sites, which accounts for the observation that some proteins find their target sites faster than allowed by 3D diffusion (62). Facilitated diffusion is mediated by nonspecific DNA-protein interactions, which are electrostatic in nature. Facilitated diffusion can be achieved by different modes of searching, including sliding, hopping and intersegmental transfer. In sliding, the protein keeps in persistent contact with the phosphate backbone and moves bi-directionally to search for target sites; in hopping, the protein

9

microscopically dissociates from the initial binding site and then rapidly reassociates with a nearby site on the same or opposing DNA strand; in intersegmental transfer, the protein is capable of long distance translocation from one DNA segment to another when they are close by in space. These searching modes contribute to varying extents to the searching process of different proteins. For example, human uracil DNA glycosylase was reported to use frequent hopping ($\geq$1 hop/3 ms) as the primary searching mode (63), while human oxoguanine DNA glycosylase 1 was suggested to employ nearly barrierless Brownian sliding for searching (64).

Previous work in our lab shows that AAG uses nonspecific binding interactions and facilitated diffusion to conduct a highly redundant search of adjacent sites (65,66). AAG is capable of searching both DNA strands and hopping over tightly bound proteins on DNA (66). This ensures that every site is searched, but could be detrimental if the protein is trapped in a local segment of DNA. Intersegmental transfer between DNA segments that are transiently in close proximity provides an elegant solution that balances global and local searching processes. Previous intersegmental transfer studies mostly used dimeric proteins or proteins with two DNA binding sites (67-72), but have not tested whether monomeric proteins can also use intersegmental transfer for efficient searching. In the crowded nucleus, the accessible DNA nucleotide concentration is estimated to be as high as 10 mM, and the presence of intersegmental transfer would undoubtedly greatly accelerate the searching process for many monomeric proteins.

Facilitated diffusion has been suggested to be essential for the *in vivo* cellular function of the proteins under study by two groups (73-77). One Group used T4 endonuclease V, which repairs pyrimidine dimers from UV radiation, and showed a direct correlation between the level of processivity of wild type and mutant endonuclease V and the degree of enhanced UV resistance which is conferred to the repair-deficient *E. coli* cells (73-75). The other group used *E. coli* restriction endonuclease EcoRV, and observed a strong correlation between the ability of EcoRV mutants to slide along the DNA *in vitro* and to protect *E. coli* cells from phage infection *in vivo*, demonstrating the essentiality of linear diffusion for effective phage restriction (76,77). These two studies highlighted the important role of facilitated diffusion on the survival of prokaryotic cells. However, it is not clear whether such a correlation applies to eukaryotic cells, since the eukaryotic genome has higher structural complexity compared to the prokaryotic counterpart.

Eukaryotic Nucleus

In the eukaryotic nucleus, DNA molecules are organized with nucleoproteins into chromatin, and the human genome is compacted ~400,000-fold through this organization to fit within a nuclear volume of ~1000 μm^3. The basic repeating unit of chromatin is the nucleosome, which comprises a core particle with 146 bp of DNA wrapped 1.67 left-handed superhelical turns around an octamer of histones (H2A, H2B, H3 and H4; 2 copies of each) (Figure 1-7) (78). Nucleosome core particles are connected by stretches of linker DNA, which can be 10-80 bp in length. Free linker DNA results in non-condensed nucleosome, which is often depicted as beads-on-a-string structure, whereas binding of linker DNA by linker histones, such as H1, is involved in chromatin compaction. Nucleosome arrays can be further packaged into 30 nm fibers and ultimately metaphase chromosome during mitosis and meiosis. This hierarchical DNA packing is an extremely efficient way of storing the DNA within the nucleus; however, this packaging also has the potential to limit the accessibility of the genetic information for DNA-dependent processes such as DNA replication, transcription and repair (79).
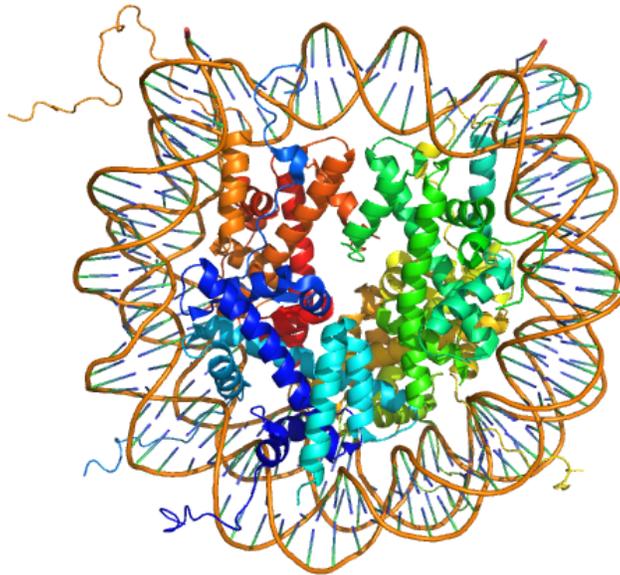
Figure 1-7. Nucleosome: the basic repeating unit of chromatin in eukaryotic cells.
The crystal structure of the nucleosome core particle (PDB: 1EQZ) consisting of H2A, H2B, H3 and H4 core histones and DNA (78). Histones are colored in rainbow, and DNA is colored by elements.

Indeed, studies show that the assembly of DNA into nucleosomes greatly restricts the accessibility of many DNA-binding factors when tested *in vitro*, including those involved in DNA repair. For example, the efficiency of nucleotide excision repair of UV-induced DNA

lesions is reduced in chromatin substrates (80,81), and evidence indicates that the repair activity is stimulated by chromatin remodeling factors (82). The ability of several different DNA glycosylases to act on target lesion sites buried in nucleosomes has also been investigated, with somewhat conflicting results. For instance, oxoguanine DNA glycosylase activity was reported to be inhibited ~100-fold at a site 10 bp from the nucleosome dyad (83), whereas uracil DNA glycosylase was found to exhibit marginally reduced activities on nucleosomal substrates (84,85). These reported differences could be reconciled by sequence context and position-specific effects of repair hotspots (85), the presence of which could influence critical biological processes.

Work Described in this Thesis

In Chapter 2, we developed biochemical assays that allowed us to observe and measure the rates of intersegmental transfer by AAG. AAG has a flexible amino terminus that tunes its affinity for nonspecific DNA, but we find that it is not required for intersegmental transfer. As AAG has only a single DNA binding site, this argues against the bridging model for intersegmental transfer. The rates of intersegmental transfer are strongly dependent on the salt concentration, supporting a jumping mechanism that involves microscopic dissociation and capture by a proximal DNA site. As many DNA-binding proteins have only a single binding site, jumping may be a common mechanism for intersegmental transfer.

In chapter 3, we used pulse-chase assays, which monitor the partition of chemical catalysis and dissociation of the enzyme from the enzyme-DNA complex, to measure the excision efficiency of AAG. Results show that the commitment to catalysis by AAG decreases as the salt concentration is increased. Under low salt conditions, the observed excision efficiency is dependent on the chase concentration used, which reveals at least one AAG-DNA species that is capable of intersegmental transfer. The non-zero values of excision efficiency at very high chase concentrations suggest that an AAG-DNA complex species resistant to intersegmental transfer is present on the pathway towards catalysis. The separation of excision efficiency and site transfer probability from F_P enables the direct comparison of searching efficiency of AAG with other enzymes.

In chapter 4, we have examined the ability of AAG to excise 1,N^6-ethenoadenine (εA) opposing different bases as well as that on the εA•T and Hx•T lesion substrates. We found that catalytic specificity (k_{cat}/K_M) is dependent on the assay condition. At low salt conditions all

lesions were excised with similar k_{cat}/K_M as a result of irreversible DNA binding, while at high salt AAG excises εA in the order of εA•C ≈ εA•T > εA•A > εA•G, which is inversely correlated to the stability of the corresponding duplex DNA. AAG has little discrimination for the εA•T and Hx•T site at low salt, while the preference for the εA•T site to the Hx•T site steadily increased to approximately 100-fold at high salt. The absence of discrimination for different target sites at low salt can be attributed to efficient facilitated diffusion and irreversible binding of AAG on its substrate DNA. Similar preference as observed at high salt was revealed when pre-steady state analysis was performed on the hybrid DNA substrates containing the εA•T and Hx•T sites. Finally, we showed that competitor DNA molecules have similar effects as salt in accentuating the ability of AAG to discriminate between different lesion sites.

Crystal structures of AAG in complex with damaged DNA reveal a positively charged DNA binding surface that nicely accommodates the bend of the extrahelical recognition complex (27,28), but it is not known whether nonspecific DNA interactions would take advantage of the same set of electrostatic interactions. In chapter 5, we have individually mutated the 5 arginine and 3 lysine residues that are near to the DNA binding interface and evaluated their contributions to both *in vitro* and *in vivo* DNA repair. The results establish that catalytic specificity, k_{cat}/K_M, is positively correlated with processivity, suggesting that most residues contribute to both specific binding and nonspecific binding. The mutants were then tested for their ability to complement the MMS sensitivity of a Mag1 yeast deletion strain, revealing a positive correlation between catalytic specificity and cell survival. This is consistent with the model that cell survival requires efficient capture of cytotoxic lesions. Survival of cells with searching deficient mutant proteins could be rescued by overexpressing the mutant proteins. Thus, it appears that chromosomal access is not restricted and sites of damage are readily accessible to a searching protein.

Collectively, these studies provide a framework for understanding the molecular mechanism by which a DNA glycosylase efficiently searches the genome for rare sites of damage and recognizes and excises different base lesions with varying specificity.

References

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

2.   Hoeijmakers, J. H. (2001) Genome maintenance mechanisms for preventing cancer. *Nature* **411**, 366-374

3.   Hoeijmakers, J. H. (2007) Genome maintenance mechanisms are critical for preventing cancer as well as other aging-associated diseases. *Mechanisms of ageing and development* **128**, 460-462

4.   Scharer, O. D. (2003) Chemistry and biology of DNA repair. *Angewandte Chemie* **42**, 2946-2974

5.   Pegg, A. E. (2000) Repair of O(6)-alkylguanine by alkyltransferases. *Mutation research* **462**, 83-100

6.   Dumenco, L. L., Allay, E., Norton, K., and Gerson, S. L. (1993) The prevention of thymic lymphomas in transgenic mice by human O6-alkylguanine-DNA alkyltransferase. *Science* **259**, 219-222

7.   Shiraishi, A., Sakumi, K., and Sekiguchi, M. (2000) Increased susceptibility to chemotherapeutic alkylating agents of mice deficient in DNA repair methyltransferase. *Carcinogenesis* **21**, 1879-1883

8.   Iaccarino, I., Marra, G., Palombo, F., and Jiricny, J. (1998) hMSH2 and hMSH6 play distinct roles in mismatch binding and contribute differently to the ATPase activity of hMutSalpha. *The EMBO journal* **17**, 2677-2686

9.   Jiricny, J. (1998) Replication errors: cha(lle)nging the genome. *The EMBO journal* **17**, 6427-6436

10.   Kolodner, R. D. (1995) Mismatch repair: mechanisms and relationship to cancer susceptibility. *Trends in biochemical sciences* **20**, 397-401

11.   Krokan, H. E., Standal, R., and Slupphaug, G. (1997) DNA glycosylases in the base excision repair of DNA. *The Biochemical journal* **325 (Pt 1)**, 1-16

12.   Lindahl, T., and Wood, R. D. (1999) Quality control by DNA repair. *Science* **286**, 1897-1905

13.   Friedberg, E. C., and Meira, L. B. (2004) Database of mouse strains carrying targeted mutations in genes affecting biological responses to DNA damage (Version 6). *DNA repair* **3**, 1617-1638

14.   Larsen, E., Meza, T. J., Kleppa, L., and Klungland, A. (2007) Organ and cell specificity of base excision repair mutants in mice. *Mutation research* **614**, 56-68

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

16.   de Boer, J., and Hoeijmakers, J. H. (2000) Nucleotide excision repair and human syndromes. *Carcinogenesis* **21**, 453-460

17.   Khanna, K. K., and Jackson, S. P. (2001) DNA double-strand breaks: signaling, repair and the cancer connection. *Nature genetics* **27**, 247-254

18.   van Gent, D. C., Hoeijmakers, J. H., and Kanaar, R. (2001) Chromosomal stability and the DNA double-stranded break connection. *Nature reviews. Genetics* **2**, 196-206

19.   Ye, N., Holmquist, G. P., and O'Connor, T. R. (1998) Heterogeneous repair of N-methylpurines at the nucleotide level in normal human cells. *Journal of molecular biology* **284**, 269-285

20.   Lieber, M. R. (2010) The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annual review of biochemistry* **79**, 181-211

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

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

23.   Stivers, J. T., and Jiang, Y. L. (2003) A mechanistic perspective on the chemistry of DNA repair glycosylases. *Chemical reviews* **103**, 2729-2759

24.   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

25.   O'Connor, T. R. (1993) Purification and characterization of human 3-methyladenine-DNA glycosylase. *Nucleic acids research* **21**, 5561-5569

26.   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**, 580-589

27.   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

28.   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. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 13573-13578

29.   O'Brien, P. J., and Ellenberger, T. (2004) Dissecting the broad substrate specificity of human 3-methyladenine-DNA glycosylase. *The Journal of biological chemistry* **279**, 9750-9757

30.   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

31.   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. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 13087-13092

32.   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. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 12869-12874

33.   Miao, F., Bouziane, M., and O'Connor, T. R. (1998) Interaction of the recombinant human methylpurine-DNA glycosylase (MPG protein) with oligodeoxyribonucleotides containing either hypoxanthine or abasic sites. *Nucleic acids research* **26**, 4034-4041

34.   Saparbaev, M., and Laval, J. (1994) Excision of hypoxanthine from DNA containing dIMP residues by the Escherichia coli, yeast, rat, and human alkylpurine DNA glycosylases. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 5873-5877

35.   Singer, B., Antoccia, A., Basu, A. K., Dosanjh, M. K., Fraenkel-Conrat, H., Gallagher, P. E., Kusmierek, J. T., Qiu, Z. H., and Rydberg, B. (1992) Both purified human 1,N6-ethenoadenine-binding protein and purified human 3-methyladenine-DNA glycosylase act on 1,N6-ethenoadenine and 3-methyladenine. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 9386-9390

36.   Engelward, B. P., Dreslin, A., Christensen, J., Huszar, D., Kurahara, C., and Samson, L. (1996) Repair-deficient 3-methyladenine DNA glycosylase homozygous mutant mouse cells have increased sensitivity to alkylation-induced chromosome damage and cell killing. *The EMBO journal* **15**, 945-952

37.   Rinne, M. L., He, Y., Pachkowski, B. F., Nakamura, J., and Kelley, M. R. (2005) N-methylpurine DNA glycosylase overexpression increases alkylation sensitivity by rapidly removing non-toxic 7-methylguanine adducts. *Nucleic acids research* **33**, 2859-2867

38.   Trivedi, R. N., Wang, X. H., Jelezcova, E., Goellner, E. M., Tang, J. B., and Sobol, R. W. (2008) Human methyl purine DNA glycosylase and DNA polymerase beta expression collectively predict sensitivity to temozolomide. *Molecular pharmacology* **74**, 505-516

39. Elder, R. H., Jansen, J. G., Weeks, R. J., Willington, M. A., Deans, B., Watson, A. J., Mynett, K. J., Bailey, J. A., Cooper, D. P., Rafferty, J. A., Heeran, M. C., Wijnhoven, S. W., van Zeeland, A. A., and Margison, G. P. (1998) Alkylpurine-DNA-N-glycosylase knockout mice show increased susceptibility to induction of mutations by methyl methanesulfonate. *Molecular and cellular biology* **18**, 5828-5837

40. Ringvoll, J., Moen, M. N., Nordstrand, L. M., Meira, L. B., Pang, B., Bekkelund, A., Dedon, P. C., Bjelland, S., Samson, L. D., Falnes, P. O., and Klungland, A. (2008) AlkB homologue 2-mediated repair of ethenoadenine lesions in mammalian DNA. *Cancer research* **68**, 4142-4149

41. Chung, F. L., Chen, H. J., and Nath, R. G. (1996) Lipid peroxidation as a potential endogenous source for the formation of exocyclic DNA adducts. *Carcinogenesis* **17**, 2105-2111

42. el Ghissassi, F., Barbin, A., Nair, J., and Bartsch, H. (1995) Formation of 1,N6-ethenoadenine and 3,N4-ethenocytosine by lipid peroxidation products and nucleic acid bases. *Chemical research in toxicology* **8**, 278-283

43. Gros, L., Ishchenko, A. A., and Saparbaev, M. (2003) Enzymology of repair of etheno-adducts. *Mutation research* **531**, 219-229

44. 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. *Journal of molecular biology* **320**, 503-513

45. Hendershot, J. M., Wolfe, A. E., and O'Brien, P. J. (2011) Substitution of active site tyrosines with tryptophan alters the free energy for nucleotide flipping by human alkyladenine DNA glycosylase. *Biochemistry* **50**, 1864-1874

46. Robert, X., and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic acids research* **42**, W320-324

47. Wyatt, M. D., Allan, J. M., Lau, A. Y., Ellenberger, T. E., and Samson, L. D. (1999) 3-methyladenine DNA glycosylases: structure, function, and biological importance. *BioEssays : news and reviews in molecular, cellular and developmental biology* **21**, 668-676

48. Nakabeppu, Y., Kondo, H., and Sekiguchi, M. (1984) Cloning and characterization of the alkA gene of Escherichia coli that encodes 3-methyladenine DNA glycosylase II. *The Journal of biological chemistry* **259**, 13723-13729

49. Nakabeppu, Y., Miyata, T., Kondo, H., Iwanaga, S., and Sekiguchi, M. (1984) Structure and expression of the alkA gene of Escherichia coli involved in adaptive response to alkylating agents. *The Journal of biological chemistry* **259**, 13730-13736

50.	Sakumi, K., Nakabeppu, Y., Yamamoto, Y., Kawabata, S., Iwanaga, S., and Sekiguchi, M. (1986) Purification and structure of 3-methyladenine-DNA glycosylase I of Escherichia coli. *The Journal of biological chemistry* **261**, 15761-15766

51.	Lindahl, T., Sedgwick, B., Sekiguchi, M., and Nakabeppu, Y. (1988) Regulation and expression of the adaptive response to alkylating agents. *Annual review of biochemistry* **57**, 133-157

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

53.	Chen, J., Derfler, B., and Samson, L. (1990) Saccharomyces cerevisiae 3-methyladenine DNA glycosylase has homology to the AlkA glycosylase of E. coli and is induced in response to DNA alkylation damage. *The EMBO journal* **9**, 4569-4575

54.	Chakravarti, D., Ibeanu, G. C., Tano, K., and Mitra, S. (1991) Cloning and expression in Escherichia coli of a human cDNA encoding the DNA repair protein N-methylpurine-DNA glycosylase. *The Journal of biological chemistry* **266**, 15710-15715

55.	Glassner, B. J., Rasmussen, L. J., Najarian, M. T., Posnick, L. M., and Samson, L. D. (1998) Generation of a strong mutator phenotype in yeast by imbalanced base excision repair. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 9997-10002

56.	Berdal, K. G., Johansen, R. F., and Seeberg, E. (1998) Release of normal bases from intact DNA by a native DNA repair enzyme. *The EMBO journal* **17**, 363-367

57.	Riggs, A. D., Bourgeois, S., and Cohn, M. (1970) The lac repressor-operator interaction. 3. Kinetic studies. *Journal of molecular biology* **53**, 401-417

58.	Surby, M. A., and Reich, N. O. (1996) Facilitated diffusion of the EcoRI DNA methyltransferase is described by a novel mechanism. *Biochemistry* **35**, 2209-2217

59.	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. *Proceedings of the National Academy of Sciences of the United States of America* **79**, 4010-4014

60.	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. *The EMBO journal* **19**, 6546-6557

61.	Verdine, G. L., and Bruner, S. D. (1997) How do DNA repair proteins locate damaged bases in the genome? *Chemistry & biology* **4**, 329-334

62.	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

63.   Schonhoft, J. D., and Stivers, J. T. (2012) Timing facilitated site transfer of an enzyme on DNA. *Nature chemical biology* **8**, 205-210

64.   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. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 5752-5757

65.   Hedglin, M., and O'Brien, P. J. (2008) Human alkyladenine DNA glycosylase employs a processive search for DNA damage. *Biochemistry* **47**, 11434-11445

66.   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

67.   Clapier, C. R., and Cairns, B. R. (2009) The biology of chromatin remodeling complexes. *Annual review of biochemistry* **78**, 273-304

68.   Doucleff, M., and Clore, G. M. (2008) Global jumping and domain-specific intersegment transfer between DNA cognate sites of the multidomain transcription factor Oct-1. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 13871-13876

69.   Fried, M. G., and Crothers, D. M. (1984) Kinetics and mechanism in the reaction of gene regulatory proteins with DNA. *Journal of molecular biology* **172**, 263-282

70.   Iwahara, J., and Clore, G. M. (2006) Direct observation of enhanced translocation of a homeodomain between DNA cognate sites by NMR exchange spectroscopy. *Journal of the American Chemical Society* **128**, 404-405

71.   Kozlov, A. G., and Lohman, T. M. (2002) Kinetic mechanism of direct transfer of Escherichia coli SSB tetramers between single-stranded DNA molecules. *Biochemistry* **41**, 11611-11627

72.   Ruusala, T., and Crothers, D. M. (1992) Sliding and intermolecular transfer of the lac repressor: kinetic perturbation of a reaction intermediate by a distant DNA sequence. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 4903-4907

73.   Dowd, D. R., and Lloyd, R. S. (1989) Site-directed mutagenesis of the T4 endonuclease V gene: the role of arginine-3 in the target search. *Biochemistry* **28**, 8699-8705

74.   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. *Journal of molecular biology* **208**, 701-707

75.   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. *The Journal of biological chemistry* **265**, 3424-3431

76.   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. *The EMBO journal* **15**, 5104-5111

77.   Wenz, C., Jeltsch, A., and Pingoud, A. (1996) Probing the indirect readout of the restriction enzyme EcoRV. Mutational analysis of contacts to the DNA backbone. *The Journal of biological chemistry* **271**, 5565-5573

78.   Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251-260

79.   Groth, A., Rocha, W., Verreault, A., and Almouzni, G. (2007) Chromatin challenges during DNA replication and repair. *Cell* **128**, 721-733

80.   Hara, R., Mo, J., and Sancar, A. (2000) DNA damage in the nucleosome core is refractory to repair by human excision nuclease. *Molecular and cellular biology* **20**, 9173-9181

81.   Powell, N. G., Ferreiro, J., Karabetsou, N., Mellor, J., and Waters, R. (2003) Transcription, nucleosome positioning and protein binding modulate nucleotide excision repair of the Saccharomyces cerevisiae MET17 promoter. *DNA repair* **2**, 375-386

82.   Hara, R., and Sancar, A. (2002) The SWI/SNF chromatin-remodeling factor stimulates repair by human excision nuclease in the mononucleosome core particle. *Molecular and cellular biology* **22**, 6779-6787

83.   Menoni, H., Gasparutto, D., Hamiche, A., Cadet, J., Dimitrov, S., Bouvet, P., and Angelov, D. (2007) ATP-dependent chromatin remodeling is required for base excision repair in conventional but not in variant H2A.Bbd nucleosomes. *Molecular and cellular biology* **27**, 5949-5956

84.   Beard, B. C., Wilson, S. H., and Smerdon, M. J. (2003) Suppressed catalytic activity of base excision repair enzymes on rotationally positioned uracil in nucleosomes. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 7465-7470

85.   Cole, H. A., Tabor-Godwin, J. M., and Hayes, J. J. (2010) Uracil DNA glycosylase activity on nucleosomal DNA depends on rotational orientation of targets. *The Journal of biological chemistry* **285**, 2876-2885

Chapter 2

Isolating Contributions from Intersegmental Transfer to DNA Searching by Alkyladenine DNA Glycosylase[1,2]

DNA repair proteins must search the entire genome to locate and repair rare sites of damage. Many of these proteins use nonspecific DNA binding that is mediated by electrostatic interactions to diffuse along the surface of the DNA (Figure 2-1). This process, which is often referred to as facilitated diffusion, can occur by either hopping and/or sliding (1-3). In sliding, a protein maintains continuous contact with the phosphate backbone, effectively conducting a one-dimensional search. In hopping, a protein microscopically dissociates but rapidly reassociates on either the same or opposing strand at a nearby site. These local sliding or hopping steps provide a highly redundant search that enables a DNA repair enzyme to detect sites of damage, but they are inefficient for covering large distances because diffusion is not directional (2). Therefore, it is advantageous to balance local search with long-range transfer steps. Intersegmental transfer (i.e., direct transfer between DNA segments) is expected to be more efficient than random 3-D diffusion, because it maximizes the time that the protein is actively searching DNA and minimizes the time that it is free in solution. Intersegmental transfer mechanisms have been invoked for a variety of different transcription factors (4-9), but few studies of DNA repair proteins have been reported. Recently single molecule approaches were used to observe intersegmental transfer by a mismatch DNA repair protein, MutL α (10). The current work investigates whether an enzyme that initiates the base excision DNA repair pathway uses an intersegmental transfer mechanism to search DNA.

---

[1]This research was originally published in The Journal of Biological Chemistry. Hedglin, M., Zhang, Y. and O'Brien, P.J. Isolating Contributions from Intersegmental Transfer to DNA Searching by Alkyladenine DNA Glycosylase. *J. Biol. Chem,* 2013; 288:24550-24559. © 2013 by The American Society for Biochemistry and Molecular Biology

[2]Author contributions: M.H., Y.Z., and P.J.O. designed research; M.H. pioneered the project and performed the processivity assay on the PEG and reference substrates (Figure 2-4A), and Y.Z. performed all the follow-up experiments; Y.Z., M.H., and P.J.O. analyzed data; P.J.O., M.H., and Y.Z. wrote the paper.

We examined the searching mechanism of human alkyladenine DNA glycosylase (AAG), because it has previously been shown to search DNA via facilitated diffusion (11). AAG is a small (33 kD) monomer that is responsible for finding a wide variety of alkylated and deaminated purines in DNA, including 3-methyladenine, 7-methylguanine, $1,N^6$-ethenoadenine (εA), hypoxanthine and xanthine (12-15). This enzyme initiates the base excision repair pathway by catalyzing hydrolysis of the *N*-glycosidic bond to release the damaged base and create an abasic site. The structure of AAG in complex with εA-DNA has been determined (16), and stopped-flow experiments found that AAG has pM affinity for this lesion (17). This tight binding makes εA an excellent substrate for studying diffusion, because every recognition event results in *N*-glycosidic bond cleavage. The formation of an alkaline-labile abasic site provides an unambiguous chemical readout for encounter by a searching protein. Previously AAG was found to make frequent hops to search both strands of DNA and it is capable of diffusing past a tightly bound protein (18).
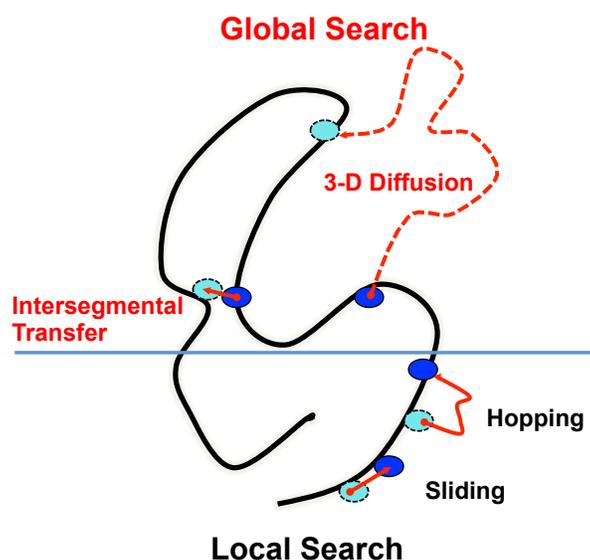
Figure 2-1. Mechanisms of facilitated diffusion for searching DNA.
The local search involves sliding and hopping, and this is highly redundant. To prevent becoming trapped in a small region, the local search must be balanced by long-range events (three-dimensional diffusion or intersegmental transfer).

The AAG primary structure is highly conserved in vertebrates, however the region of conservation is limited to the carboxy-terminal region of approximately 220 amino acids. The amino-terminal region (80 amino acids in human AAG) is poorly conserved both in length and in amino acid sequence. This region of human AAG is proteolytically sensitive and appears to be

flexible (19). Many eukaryotic DNA binding and DNA repair proteins also have disordered amino- or carboxy-termini (20-22) and it is proposed that one role of these flexible regions is to enable bridging interactions between two segments of DNA (23,24).

We present a new biochemical assay to detect intersegmental transfer that is based upon tethering of DNA substrates with polyethylene glycol (PEG) linkers and measuring the frequency of correlated enzymatic action on the linked substrates. These assays and supporting kinetic characterization of intermolecular transfer rates demonstrate that AAG is capable of rapid intersegmental transfer. Extrapolation of the rates measured at low concentrations of DNA to the expected concentration of DNA in the nucleus indicates that both intermolecular and intramolecular diffusion occur on the same time scale. This suggests that AAG is able to balance an efficient local search with the global search for DNA damage. Although the amino terminus of AAG enhances DNA binding affinity and decreases the rate constant for dissociation, it is not required for intersegmental transfer. We conclude that this monomeric DNA binding protein is capable of intersegmental transfer via microscopic dissociation and capture by a new DNA strand during a transient encounter, without the requirement for a bridging intermediate.

MATERIALS AND METHODS

Protein and Oligonucleotides

Full-length and N-terminal truncated (Δ80) AAG were purified and the concentration of active enzyme determined by burst analysis as described previously (25). Oligonucleotides were obtained from the Keck Center at Yale University or Integrated DNA Technologies, purified by denaturing PAGE, and the concentrations determined by absorbance at 260 nm.

Glycosylase Assays

Assays were performed at 37 °C in 50 mM NaMES, pH 6.1, 10% (v/v) glycerol, 1 mM DTT, 1 mM EDTA and 0.1 mg/ml BSA. The sodium concentration was adjusted to the desired level with the addition of NaCl. Reactions were quenched at the desired time in 0.2 M NaOH, heated at 70 °C for 15 min to cleave the abasic sites and supplemented with formamide/EDTA loading buffer. DNA fragments were resolved by PAGE on 14% gels, scanned with a Typhoon Trio[+] fluorescence imager and quantified with ImageQuant TL as previously described (18). Initial rates were determined from the first 10% of the reaction progress curves.

Multiple-turnover Processivity Assay

The lesion-containing strand (fluorescein-labeled) and the complement strand(s) were annealed at a ratio of 1:2. Unless otherwise indicated, reactions contained 200 nM substrate and 2 nM full-length or $\triangle$80 AAG. The initial rates for intermediate (V_{int}) and product formation (V_{prod}) were determined from linear fits. In all cases the two products and two intermediates were formed at identical rates indicating that AAG encounters either lesion with equal probability. The fraction processive values were calculated by $F_{p, obs} = (V_{int} - V_{prod})/(V_{int} + V_{prod})$, and corrected for the small amount of ring-opened εA as previously described by $F_p = (F_p - 0.05)/(0.92 - 0.05)$ (18). At low ionic strength it was necessary to account for the nonenzymatic rate of εA depurination, which was independently determined from control reactions without AAG.

Determination of Relative k_{cat}/K_M Values

Multiple turnover kinetics were performed with equal concentration of the 25εA•T reference and the indicated substrate (200 nM each) and 4 nM AAG. All substrates and products were resolved by denaturing PAGE and the relative k_{cat}/K_M value is given by the ratio of the initial rates for reaction of the indicated substrate and that of the reference duplex (26).

Pre-steady State Burst Analysis

To minimize the contribution of two proteins bound to the same DNA, reactions contained 100 nM AAG and 1 μM processivity substrate. The fraction of substrate remaining was fit by an exponential and a steady-state linear phase. To obtain 20 mM Na$^+$ the concentration of all buffer components was decreased to 40% of the standard concentrations (0.4× buffer). The fraction intermediate was fit by the theoretical equation for the scheme shown in Figure 2-4A (see Appendix Scheme A-2).

DNA-stimulated Dissociation

The fluorescein-labeled lesion-containing strand and the complement strand were annealed at a 1:1 ratio, as control experiments showed that single-stranded DNA also facilitates the enzyme dissociation from the product. The homogeneity of the duplexes was confirmed by native polyacrylamide gel electrophoresis. The substrate concentration used in this experiment ranged from 0.2 μM to 4 μM, with a fixed substrate to enzyme ratio of 100:1. The initial rates were plotted versus the substrate concentration and were fit to the linear equation ($V/E = k_{off} + k_{trans}[DNA]$). The Y-intercept of the fit represents the intrinsic dissociation rate of the enzyme from the abasic product and the slope is the second order rate constant for direct transfer from

the abasic product to a new substrate. Stimulation by competitor DNA was performed with 2 nM AAG and 200 nM processivity substrate (annealed 1:2). The competitor DNA (25A•T and 25D•T) was annealed at a 1:1.2 ratio.

RESULTS

Processivity assays provide a powerful tool to quantify the ability of an enzyme to diffuse along DNA and to locate a specific site (27-29). With this approach, an initial binding event is monitored by enzymatic activity at one of two sites on a DNA molecule (Figure 2-2A).

Figure 2-2. Oligonucleotides used to monitor the diffusion of AAG.
A. Multiple-turnover processivity assays monitor the partitioning between dissociation (k_{off}) and finding the second lesion ($k_{capture}$). One possible pathway is shown, but binding is random, and so there are two pathways for processive excision to form abasic sites (Ab) from the ethenoadenosine lesions (E). B. Two-lesion and one-lesion oligonucleotide sequences (asterisk indicates the position of fluorescein labels).

Subsequently, partitioning between dissociation (which produces intermediate oligonucleotides containing a single site of damage) and between capture and excision of the second site (which produces product oligonucleotides containing no sites of damage) can be determined. The excess of substrate over enzyme in the multiple-turnover assays ensures that the probability of rebinding

25

the same substrate molecule is negligible. The fraction processive (F_P) provides a quantitative measure of the macroscopic rate constants for dissociation and capture at the second site [$F_P = k_{capt}/(k_{capt} + k_{off})$]. This assay was previously used to establish that AAG conducts an efficient and highly redundant search of DNA (11,18). In the current work, we covalently linked two duplexes using a flexible tether to test whether AAG is capable of direct transfer between duplexes. The purpose of the tether is two-fold. First, the covalent linkage allows processivity assays to be used to correlate the enzymatic action on two tethered molecules. Second, the flexible linkage provides a high effective concentration of the two duplexes relative to each other that ensures intersegmental transfer between the linked duplexes is favored over transfer between separate DNA molecules. For this assay to accurately report on intersegmental transfer, it is critical that the protein is not able to diffuse along the tether. In the following section we investigate the possibility of using single strand DNA or artificial PEG as a flexible tether. These studies reveal that single strand DNA is not suitable, because AAG is capable of facilitated diffusion on single-strand DNA. However, AAG is incapable of diffusing along PEG polymers and these tethers could then be used to characterize the intersegmental transfer mechanism of AAG.

Evaluate the Diffusion of AAG on PEG and Single-strand DNA Polymers Using Direct Competition Assays

It is known that AAG is strongly dependent on duplex DNA for its glycosylase activity (11,30), and therefore it is not informative to examine the processivity on single-strand substrates. As an alternative approach, we measured the relative k_{cat}/K_M values for εA-containing duplexes that were extended on the 3' end by either a single strand DNA region or a PEG chain (Figure 2-3A). If the 3' extension provides a pathway for productively binding to the site of damage (i.e., allows facilitated diffusion), then the k_{cat}/K_M value will be increased by the larger number of binding sites relative to the duplex region on its own. By competing two DNA substrates in the same reaction, one of which has a 3' extension, the relative k_{cat}/K_M values can be directly obtained from the relative initial rates (25). Under conditions conducive to efficient facilitated diffusion (100 mM NaCl), any contribution of the 3' extension to searching can be detected (Figure 2-3B). This is in contrast to conditions of very high salt (1000 mM NaCl) in which facilitated diffusion is not observed (Figure 2-3C). All substrates were designed to have
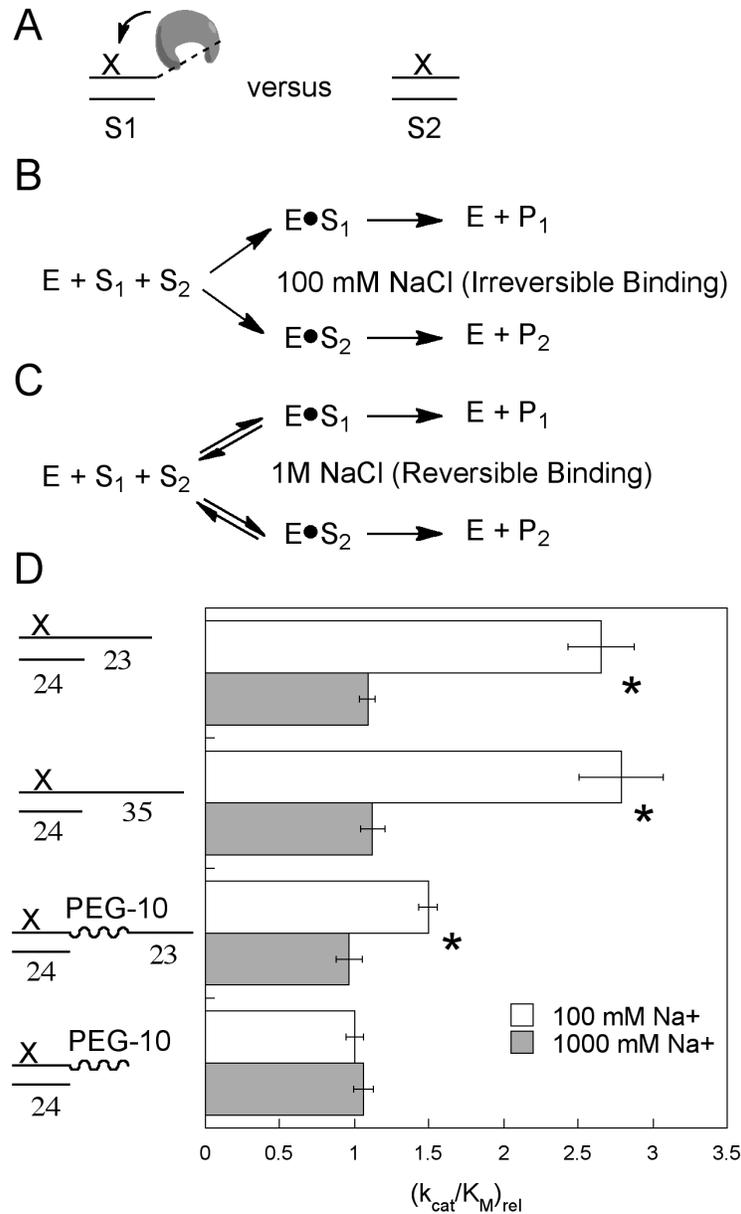
Figure 2-3. Steady state competition experiments to evaluate binding to PEG and single-strand DNA.
A. If AAG is capable of binding to, and diffusing from, a 3' extension, then there will be a larger number of productive encounters relative to DNA substrate lacking the extension and a larger value of k_{cat}/K_M will be observed. B. At low salt concentration AAG binding is irreversible, and the protein uses facilitated diffusion to search every site of a DNA molecule. The k_{cat}/K_M values simply reflect the association rate constant. C. At high salt concentration, AAG binding is fully reversible, and its glycosylase activity is distributive. D. The indicated substrates were competed against an equal concentration of 25-mer duplex with a central $\varepsilon A \bullet T$ lesion, and the relative k_{cat}/K_M values are shown (mean ± S.D., $n > 3$). The 100 mM data (white bars) indicate that AAG can productively transfer from single-strand DNA to duplex DNA but is unable to transfer from PEG to duplex DNA. The 1 M data (gray bars) provide an important control that confirms that the intrinsic reactivity is the same for all $\varepsilon A \bullet T$ sites that were tested. Statistical significance for comparison of 1 M and 100 mM NaCl conditions was evaluated by Student's t test using GraphPad Prism (* denotes $p < 0.001$).

27

equivalent εA sites, but with different sizes to allow them to be analyzed simultaneously on the same denaturing acrylamide gels.

In each case the indicated substrate was competed against the same 25-mer duplex substrate that served as a reference (Figure 2-3A). Under conditions of 100 mM NaCl the k_{cat}/K_M values for the DNA with single strand overhangs are approximately 2.5-fold larger than for the blunt end duplex, demonstrating that single stranded DNA serves as a conduit for AAG to diffuse to the εA•T lesion site (Figure 2-3D, open bars). In contrast, the PEG overhang has the same k_{cat}/K_M value as the duplex indicating that any binding of AAG to the PEG region does not result in productive transfer to the εA site. The substrate containing the PEG-10 tether to a single strand extension shows a modest increase in k_{cat}/K_M relative to the duplex with no extension, suggesting that the PEG linker acts as an insulator to AAG diffusion, but that AAG can transfer from the single strand region to the duplex region directly (i.e., intersegmental transfer). As a control, we carried out the same set of competition experiments at 1M NaCl. Under these conditions the binding of AAG is reversible and there is no facilitated diffusion (Figure 2-3C). All of the substrates showed the same k_{cat}/K_M values under these conditions (Figure 2-3D, filled bars), confirming that the concentrations of DNA were correct. We conclude that AAG is capable of diffusing along single-strand DNA, but cannot diffuse on PEG. Therefore we used PEG linkers in assays to detect intersegmental transfer.

Multiple-turnover Processivity Assays to Monitor Intersegmental Transfer

To determine if intersegmental transfer contributes to the search for DNA damage by AAG, two oligonucleotide duplexes that each contained an εA lesion site were tethered by a PEG linker (Figure 2-2B). Correlated excision at the two sites of damage prior to macroscopic diffusion provides evidence for intersegmental transfer between duplexes. The processivity for these tethered oligonucleotides was compared to that of a continuous duplex substrate with two εA sites separated by 25 base pairs (47E2F2; Figure 2-2B), that was previously characterized (11).

Multiple turnover processivity assays were performed under conditions of partial processivity (200 mM Na^+), because this condition is most sensitive to any changes in processivity (11). The processivity values of AAG with PEG tethers of 1, 4 and 10 units are the same within error of the intact duplex substrate (Figure 2-4A; $F_P = 0.6$). These results provide compelling evidence that AAG is capable of rapid intersegmental transfer, with the rate of

28

transfer far exceeding the rate of macroscopic dissociation. The lack of a length dependence for the different length PEG tethers is not surprising, because PEG is extremely flexible and has a short persistence length (31). Previous studies of the properties of PEG tethers provide estimates for the effective concentration of tethered duplex that a protein bound to one duplex would experience (32,33). For the PEG-10 substrate (tethered by 60 ethylene glycol units) the effective concentration is estimated to be approximately 3 mM, which appears to be sufficiently high to support rapid transfer.

To test whether intersegmental transfer is sensitive to the concentration of cations, we measured the steady-state processivity of AAG with the PEG-10 substrate across a wide range of sodium concentration and compared it to the processivity of the continuous duplex substrate. Between 50 mM and 300 mM Na^+, AAG efficiently transfers between the tethered duplexes and has the same probability of excising the lesion as if the lesions were separated by a continuous duplex (Figure 2-4B). Furthermore, the steady-state rates are indistinguishable for the two substrates (Figure 2-4C). These results establish that intersegmental transfer is efficient across a wide range of salt concentration and it is fast relative to dissociation from the abasic product. However, it is not possible to infer from these data whether the intersegmental transfer step might be sensitive to the salt concentration. This is because the rate of macroscopic dissociation is strongly dependent on the salt concentration and the intersegmental transfer rate could change significantly without becoming the rate-limiting step.

Transient Kinetic Analysis of Transfer on Processivity Substrates

To gain deeper insight into the rate of intersegmental transfer, we performed pre-steady state glycosylase experiments with the continuous duplex and PEG-tethered substrates. Under burst conditions, with a 10-fold excess of substrate over enzyme, the individual steps can be monitored to gain insight into steps that are faster than the overall steady-state rate. For these processivity substrates, the formation and disappearance of the intermediate with a single εA lesion follows a branched, two step irreversible pathway, in which the first step is base excision and the second step is either capture at the second site or dissociation (Figure 2-5A). Under conditions for which DNA searching and/or intersegmental transfer is slower than the rate of base excision, the transient accumulation of the intermediate with a single εA lesion excised will be increased and the subsequent decrease to steady state levels will be slowed.
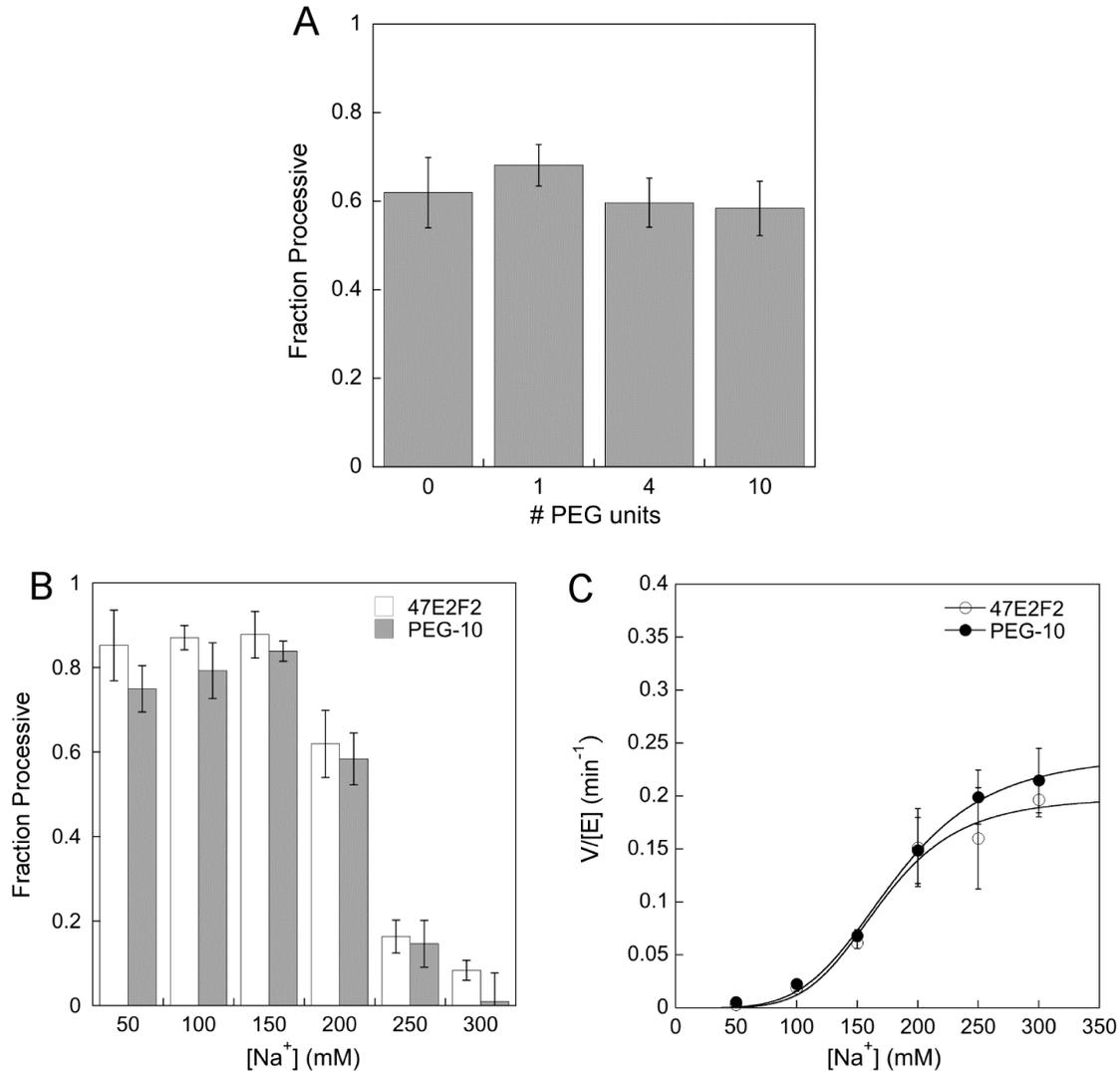
Figure 2-4. Processivity assays with PEG-tethered duplexes demonstrate that AAG is capable of intersegmental transfer.
A. The steady state processivity of the previously characterized substrate (47E2F2) (11), in which two εA lesions are located 25 bp apart, is compared with the processivity of substrates in which the two εA lesions are on separate DNA molecules that are tethered with PEG linkers. The Na⁺ concentration was 200 mM. B. Comparison of the salt dependence for the processivity of AAG with the PEG-10 substrate. C. The values of V/E for the experiments shown in panel B. Each value is the average ± S.D. (n>3).

At 50 mM NaCl, a clean burst in substrate disappearance is observed for both the continuous duplex and the PEG-tethered substrates (Figure 2-5B, squares). This confirms that the concentration of enzyme and DNA were correctly determined. From the same time points the concentration of the product bands and intermediate bands (containing one abasic site and one εA lesion site) were also calculated. It should be noted that AAG does not prefer either of the lesion sites and both of the intermediate and product bands are observed in very similar intensity

(18). Therefore, the sum of the two intermediate species and of the two product species were calculated and plotted. The reaction progress curve of the intermediate species is the most informative (Figure 2-5B, circles). The equations governing the formation of the intermediate
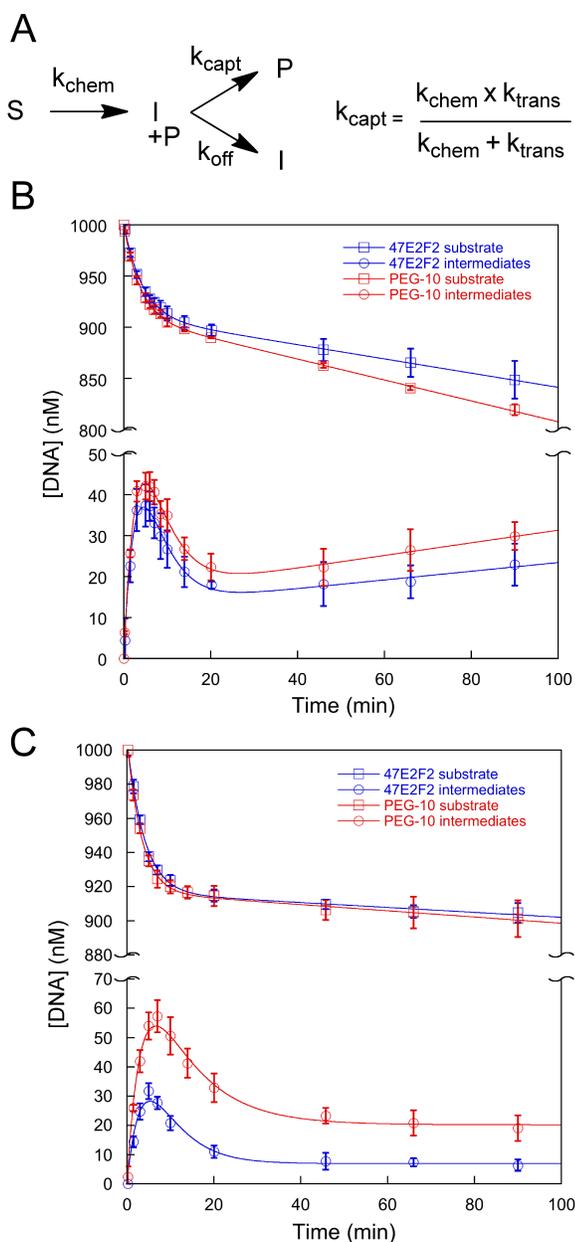
Figure 2-5. Pre-steady state kinetics with two-lesion substrates.
A. Minimal steps involved in the AAG-catalyzed reaction for oligonucleotides with two εA lesion sites. The apparent rate constant for capture (k_{capt}) is dictated by the rate constant for transfer to the second site (k_{trans}) and εA excision (k_{chem}). B and C. Burst experiments were performed with a 1:10 ratio of AAG to DNA substrate (47E2F2 or PEG-10), and the concentration of each species was calculated from the fraction of the total fluorescence determined by denaturing PAGE. Reactions were performed in triplicate, and the mean ± S.D. is plotted. Lines indicate the fits according to the irreversible model shown (see Appendix A for the derivation of the equations; rate constants are compiled in Table 2-1). The total concentration of Na^+ was either 50 mM (B) or 20 mM (C).

were derived according to Figure 2-5A (see the Appendix A for the derivation of these equations). Fitting this equation to the data for the continuous duplex substrate (Figure 2-5B, blue circles) reveals that the rate constants for formation and breakdown of the intermediate species are the same within error (Table 2-1). For the PEG-10 substrate (Figure 2-5B, red circles), the rate constant for the formation of the intermediate was also identical within error (Table 2-1). However, a reproducible increase in the level of intermediate was observed that was accompanied by a slightly slower rate of breakdown of the intermediate (Figure 2-5B; Table 2-1). This suggests that a new step, presumably intersegmental transfer, is partially rate-limiting and contributing to the rate of capture of the second site. Although identical results were obtained in multiple independent experiments, the magnitude of this effect is small and it is difficult to exclude experimental error.

Table 2-1. Rate constants from pre-steady state kinetic analysis of two-lesion substrates

	$k_{chem}(min^{-1})^a$	k_{capt} $(min^{-1})^a$	k_{trans} $(min^{-1})^b$
50 mM Na$^+$			
47E2F2	0.23 ± 0.01	0.26 ± 0.02	$>3^c$
PEG-10	0.27 ± 0.04	0.19 ± 0.03	0.7 ± 0.5
20 mM Na$^+$			
47E2F2	0.27 ± 0.04	0.14 ± 0.01	0.3 ± 0.1
PEG-10	0.31 ± 0.02	0.07 ± 0.01	0.09 ± 0.01

[a]The k_{chem} values were determined from the pre-steady state burst for the disappearance of substrate (Figure 2-5). Values for k_{capt} were determined by fitting the burst of intermediate formation and breakdown with equation S9 in Appendix A. The values reflect the mean ± S.D. (≥ 3 independent determinations).
[b]The rate constant for transfer was calculated from the values of k_{chem} and k_{capt} according to the scheme in Figure 2-5A and the error was estimated by propagation of the S.D. of the individual rate constants.
[c]No detectable delay was observed, indicating that the transfer step is at least 10-fold faster than the base excision step (k_{chem}).

Intersegmental transfer is expected to be sensitive to the salt concentration (4,34) and therefore we lowered the salt concentration to 20 mM and repeated the transient kinetic experiment that was described above (Figure 2-5C). Both substrates showed the expected burst of substrate disappearance and the multiple-turnover rate was greatly decreased due to slower dissociation from the abasic product (11,35). Nevertheless, the rate constant for removal of the first εA lesion is the same as observed at 50 mM Na$^+$. For the 47E2F2 substrate a noticeable delay was observed corresponding to a rate constant for finding the second εA lesion of 0.14 min^{-1}. In contrast, the PEG-10 substrate exhibited a significantly slower rate constant for capturing the second εA lesion of 0.07 min^{-1} (Table 2-1), resulting in greater transient

accumulation of the intermediate with a single εA lesion (Figure 2-5C). These data allow the rate constants for intramolecular transfer across 25 bp of duplex and intersegmental transfer between tethered duplexes to be calculated (Table 2-1). Remarkably, the intersegmental transfer between duplexes tethered by 60 ethylene glycol units is only 3-fold slower than the intramolecular transfer between sites 25 bp apart on a continuous duplex (20 mM salt condition; Table 2-1). It is not possible to measure rates of transfer at physiological salt concentration using this assay, because the rates are much faster than base excision, but it is apparent that AAG transfer is accelerated by increased concentration of sodium ions. For the PEG-10 substrate the observed transfer rate constant increased by 7-fold when the sodium concentration was increased from 20 to 50 mM (Table 2-1).

Evidence of Intersegmental Transfer Obtained by Examining the DNA Concentration Dependence

The observation of burst kinetics (Figure 2-5) demonstrates that dissociation from the abasic product is rate-limiting under low salt conditions. Intersegmental transfer to a new DNA molecule would provide an alternative pathway to dissociation and result in an increased reaction velocity (Figure 2-6A). We therefore measured multiple turnover glycosylase activity on the standard 47mer processivity substrate at a range of DNA concentration (0.2 - 4 μM) that is far above the K_M value for εA binding, and the results are shown in Figure 2-6B. The reaction velocity is linearly dependent on the substrate concentration, consistent with the predictions of the intersegmental transfer pathway. The data were fit by the theoretical model ($V/[E] = k_{off} + k_{transfer}[DNA]$), that yields the off-rate of AAG at infinite dilution (y-intercept) and the bimolecular transfer rate constant (slope). The processivity was also determined as a function of DNA concentration under the same conditions (Figure 2-6C). At 50 and 100 mM NaCl, the processivity is unchanged over the range of DNA concentration tested indicating that AAG always finds and removes the second lesion prior to dissociation or intersegmental transfer. In contrast, at 150 mM NaCl the processivity decreases with increasing concentration of DNA. This demonstrates that intersegmental transfer begins to compete with the pathway for finding the second lesion at this higher salt concentration and mheconcentration of DNA. The unimolecular and bimolecular rate constants determined from the data in Figure 2-6B are both strongly dependent upon the salt concentration (Table 2-2).

Intermolecular transfer can also be monitored by addition of competitor DNA that does not contain sites of damage. Transfer to undamaged DNA allows for more rapid dissociation or transfer to a new substrate molecule. As predicted by this model, multiple turnover reactions of AAG were linearly accelerated by the addition of undamaged duplex (Figure 2-7A, open symbols). The observation of a significant rate increase, without a decrease in processivity, shows that intersegmental transfer to the competitor DNA is slower than intramolecular

$$V/E = k_{off} + k_{trans}[DNA]$$
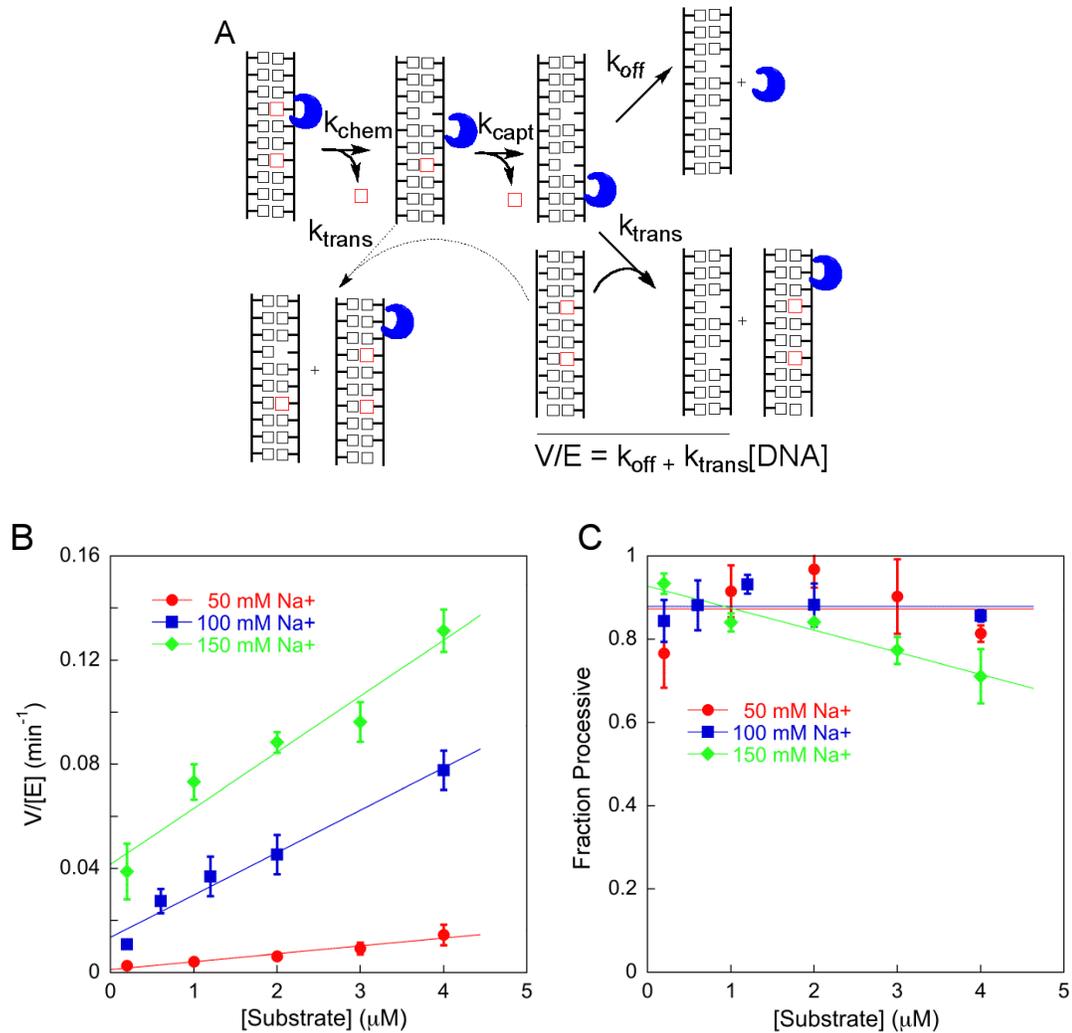
Figure 2-6. Transfer to a new DNA molecule is promoted at higher DNA concentration.
A. Minimal kinetic mechanism for the AAG-catalyzed reaction on an oligonucleotide with two sites of damage. At dilute concentrations of DNA, the rate-limiting step is dissociation from the abasic product (k_{off}), but at higher concentrations of DNA, an intermolecular transfer step (k_{trans}) accelerates the overall rate of reaction. Multiple-turnover processivity assays were performed at 50 -150 mM Na$^+$ using the 47E2F2 substrate. B. Reaction velocities (mean ± S.D., n ≥ 3) were fit by linear regression (R^2 ≥ 0.94). C. The fraction processive was calculated as described under "Materials and Methods" and analyzed by linear regression. The 50 and 100 mM NaCl conditions did not show a significant slope (p > 0.01; GraphPad Prism). At 150 mM NaCl, a modest, yet statistically significant, slope was observed (p < 0.0001), indicating that intersegmental transfer begins to compete with finding the second lesion. This additional pathway for transfer is illustrated by the dashed lines in panel A.

34

Table 2-2. Rate constants for dissociation and intersegmental transfer of AAG

[Na$^+$] (mM)	k_{off} (s^{-1})^a	k_{trans} (M^{-1}s^{-1})^b	Extrapolated k_{trans} (s^{-1})^c
50	2.1×10^{-5}	50	$1\times10^{-2\ d}$
100	2.3×10^{-4}	270	6×10^{-2}
150	7.0×10^{-4}	360	8×10^{-2}

^{a}Rate constant for dissociation of the abasic product from the intercept in Figure 2-6B. The rate constant for dissociation from nonspecific DNA is substantially faster, because nonspecific DNA stimulates multiple-turnover by AAG (Figure 2-7A).
^{b}Bimolecular rate constants for intersegmental transfer are from the slopes in Figure 2-6B.
^{c}The rate constant for intersegmental transfer *in vivo* was estimated by assuming that 10% of the genome is accessible (~10 mM DNA bp); the observed bimolecular rate constant was divided by the length of the substrate (47 bp) and multiplied by 0.01M bp.
^{d}Coincidentally, the rate constant for transfer between the duplex arms of the PEG-10 substrate is the same under this condition (Table 1; $k_{trans} = 0.012$ s^{-1}). Transient kinetic approaches could not be used to measure transfer on the PEG substrates at higher cation concentration, because transfer is much faster than base excision.

searching, but faster than AAG dissociation when the concentration of NaCl is 50 or 100 mM. However, at 150 mM NaCl and high concentration of competitor DNA the rate of transfer becomes competitive with the macroscopic rate constant for finding the site of damage and there is a small, but detectable decrease in the processivity of AAG (Figure 2-7B, open diamonds).

As a control, we compared the kinetic effects of nonspecific 25-mer competitor to those of a DNA that contains a synthetic tetrahydrofuran (THF) abasic site. This site is structural mimic of the abasic DNA product (Figure 2-2B) and the rate of AAG dissociation is expected to be similar. We observe essentially no effect of the added THF-DNA on the rate (Figure 2-7A, filled diamonds), providing further support for the model that the rate effect observed with substrate or nonspecific competitor DNA is due to intermolecular transfer. The rate of intermolecular transfer is the same for the THF-DNA as for undamaged DNA, because the decreased processivity is identical within error for the experiments with the two different DNA molecules (Figure 2-7B, diamonds). The use of competitor DNA provides important controls that rule out alternative models, such as effects due to added DNA, or incomplete saturation of the enzyme. However, it is not desirable to obtain quantitative values for the intermolecular transfer frequency (k_{trans}) from these data. This is because the DNA that is added will also act as a competitive inhibitor with respect to substrate binding and the overall reaction rates are a combination of inhibition and enhanced turnover via the intermolecular transfer pathway. Therefore, it is preferable to use the experiments that vary substrate DNA to obtain values of k_{trans} (Figure 2-6).
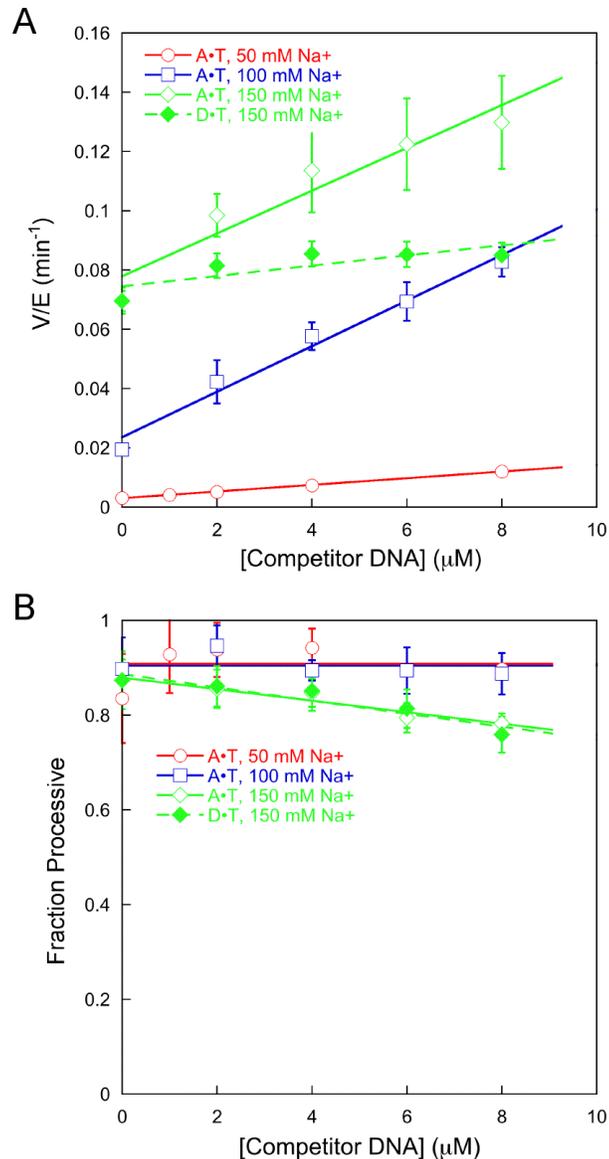
Figure 2-7. Stimulation of AAG by nonspecific DNA.
Multiple turnover reactions of the processivity substrate (47E2F2) were performed in the presence of increasing concentrations of 25-mer competitor DNA at different concentrations of NaCl. The open symbols correspond to a nonspecific 25-mer (25A•T), and the closed symbols correspond to a specific inhibitor (25D•T, D = THF) DNA. A. The velocity increases with increasing concentration of nonspecific DNA, indicating that intersegmental transfer provides an alternative pathway to dissociation from the abasic product and that subsequent transfer from the nonspecific DNA is rapid. In contrast, the specific inhibitor has no effect on the reaction velocity. This can be explained by the fact that dissociation from the abasic analog is slow. B. The fraction processive is unchanged by increasing concentration of competitor at low salt (no significant slope, $p > 0.01$; GraphPad Prism). At 150 mM NaCl, a modest, yet statistically significant ($p < 0.003$), dependence is observed for both the nonspecific competitor and the specific inhibitor DNA. Values are the mean ± S.D. ($n \geq 3$, except for the 8 μM 25A•T point for which n = 2).

Probing the Contribution of the Amino Terminus of AAG to Intersegmental Transfer

Eukaryotic DNA-binding proteins commonly have disordered tails that have net positive charge. It has been proposed that these basic tails enable efficient intersegmental transfer by allowing for transient formation of a bridging complex (23,24,36). AAG similarly has a positively charged amino terminus that decreases the rate of dissociation and thereby contributes to processive searching (11). Whereas the catalytic domain (80-298 in human AAG) is highly conserved among AAG homologs, the amino terminus is poorly conserved and varies widely in length (Figure 2-8A). The rate of N-glycosidic bond cleavage for the truncated protein ($\Delta80$) is identical to that of the full-length protein (37). Therefore, we tested to what extent the truncated protein that lacks the amino terminus is capable of intersegmental transfer. Processivity assays with the PEG-tethered substrates clearly show that $\Delta80$ AAG is capable of efficient intersegmental transfer (Figure 2-8B). We also examined the effect of DNA concentration on the observed rate of dissociation, as described for full-length AAG. The intersegmental transfer rate, which is given by the slope in the DNA concentration dependence, is almost identical for the full-length and truncated protein (Figure 2-8C). Although truncation does not alter the rate constant for intersegmental transfer, the intercept is ~30-fold higher for $\Delta80$ AAG than for full-length AAG, confirming that the amino terminus increases the binding affinity for the abasic DNA product (11). These results demonstrate that the positively charged amino terminus of AAG is not required for intersegmental transfer.

DISCUSSION

It has been suggested that both short-range intramolecular and long-range intersegmental searching steps are required for the efficient search of genomic DNA (1,38-40), but there is little experimental information about the relative contributions of the two pathways, particularly the intersegmental pathway(s). In studying the searching mechanism of AAG, we have developed simple kinetic assays that allow for the quantitative dissection of intersegmental transfer pathways. Previous studies have examined DNA binding proteins that are multimeric, or have multiple putative DNA binding domains (4,5,10). In the current work we focused on the ability of a small, monomeric enzyme to search DNA. Our results strongly favor efficient intermolecular transfer at a biologically relevant DNA concentration.
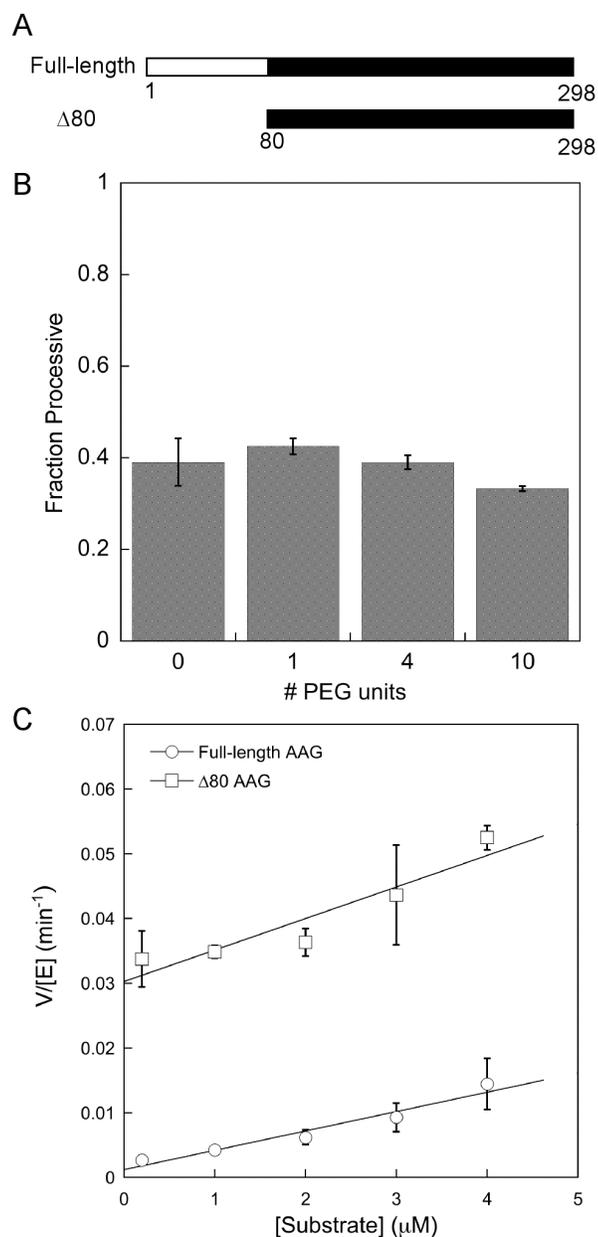
**Figure 2-8. The amino terminus of AAG is not required for intersegmental transfer.**
A. AAG has a poorly conserved amino-terminal region (32) and a highly conserved carboxyl-terminal catalytic domain (black). B. The processivity of Δ80 AAG is not decreased by a PEG linker. Processivity measurements were performed at 115 mM NaCl as described for full-length AAG. C. Multiple-turnover glycosylase activity was measured for 47E2F2 with 50 mM Na[+]. The stimulation of multiple-turnover glycosylase activity by high concentrations of DNA is similar for full-length and Δ80 AAG.

Two mechanisms for intersegmental transfer have been suggested that are distinct from simple three-dimensional diffusion (1,41). In the first mechanism, there is an intermediate state in which the protein is simultaneously bound to two DNA sites. This mechanism has historically been called simply intersegmental transfer (1), but we refer to it as the bridging mechanism to

38

avoid confusion with the fact that there is more than one mechanism for intermolecular/ intersegmental transfer (Figure 2-9). In the second mechanism, the protein microscopically dissociates (hops) and encounters a second site on a separate DNA molecule (or distant site on the same DNA molecule) that happens to be close by. Microscopic dissociation events that result in rebinding far away along the contour of the DNA, or on a different DNA molecule altogether, have been referred to as a jump (4,41). We will use this term here for consistency with existing literature. Both hops and jumps involve microscopic dissociation and reassociation, but a hop is defined as an intramolecular transfer step and a jump is an intersegmental transfer step. In some cases, the observation that the observed dissociation rates are dependent upon the concentration of the acceptor DNA has been taken as evidence for the bridging model of intersegmental transfer (4-6). However, it is important to recognize that this observation can also be explained by a jumping mechanism, whereby a hop to a new DNA segment occurs much more quickly than macroscopic dissociation into solution. In a few cases there is direct biochemical evidence for bridging intermediates formed by multimeric proteins (7,9,42), but the distinction between these two mechanisms of intermolecular transfer has been largely ignored because most studies of facilitated diffusion have examined proteins with more than one DNA binding site.

Of the two models for intermolecular transfer (Figure 2-9), our results are best explained by the jumping mechanism, which involves microscopic dissociation and re-association. We consider it unlikely that AAG could employ a bridging mechanism, because the positively charged amino terminus is not required for intersegmental transfer (Figure 2-8) and there is not an obvious secondary binding site. Furthermore, it is expected that a bridging mechanism would be saturable with respect to the second DNA species. Saturation is most likely to be observed at low salt concentration, which stabilizes electrostatic DNA binding interactions. Even at low salt concentration we observe no sign of saturation up to 4 μM DNA (Figure 2-6B). We cannot rule out that saturation might occur under some other conditions that we did not test, but we note that the linear dependence on DNA concentration is fully consistent with the jumping mechanism, because the rate of microscopic dissociation is expected to be fast. The jumping mechanism also predicts that the frequency of jumps is increased by higher concentrations of cations, and we observe faster rates of intersegmental transfer by AAG at higher concentration of sodium cations (Table 2-2). These studies, and the previous study identifying the importance of hopping in the facilitated diffusion of AAG (18), suggest that AAG searches DNA with a combination of very

fast short-range intramolecular hops that are interspersed with occasional intersegmental jumps. The simplest model to explain these two types of translocations is that they both involve a common intermediate that arises from microscopic dissociation. If another DNA segment happens to form a collision complex during the lifetime of the intermediate, then association with the new DNA segment results in a jump. If no other segments of DNA are nearby, then these microscopic dissociation events are usually resolved by returning to the same DNA molecule and would be classified as a hop.
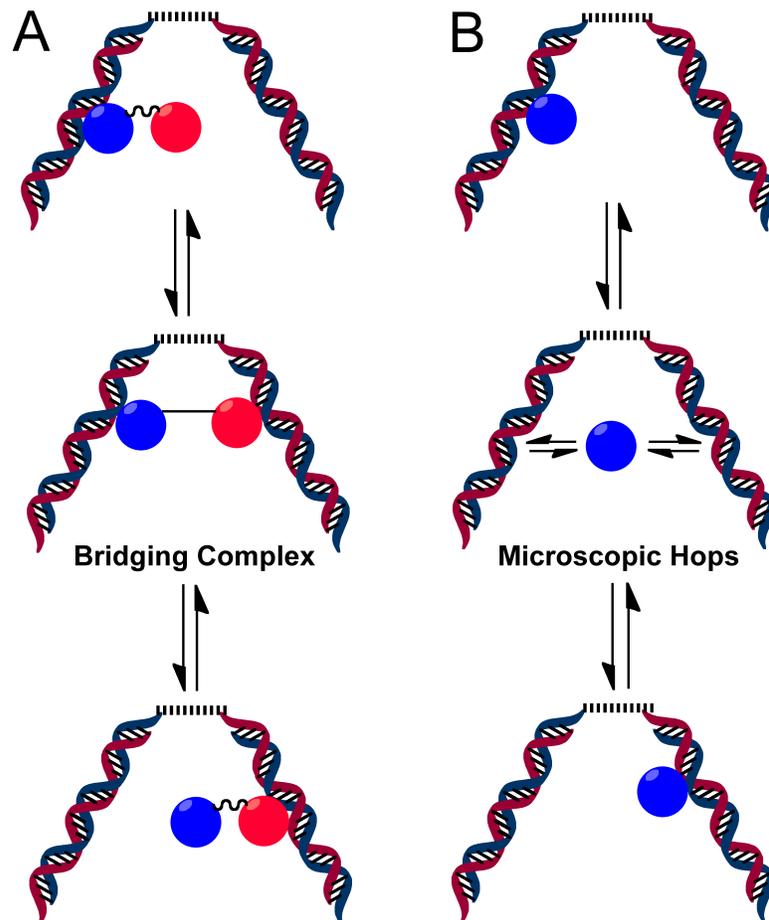
Figure 2-9. Models for intermolecular/intersegmental transfer by DNA-binding proteins.
Direct transfer between sites that are distant on the same DNA molecule is equivalent to direct transfer between two DNA molecules. A. The presence of two DNA binding domains (red and blue spheres) allows for a bridging intermediate in which two segments of DNA are simultaneously bound. This mechanism has traditionally been referred to as intersegmental transfer (1). Here we refer to this as the bridging mechanism. B. Proteins with a single DNA binding site may exhibit intersegmental transfer without macroscopic release into bulk solvent provided that the probability of recapture is sufficiently high. This has been called "jumping" (28). The results for AAG are best explained by the jumping mechanism of intersegmental transfer.

We find that the positively charged amino terminus of AAG does not contribute to intersegmental transfer, however it clearly tunes the DNA binding affinity to balance the partitioning between facilitated diffusion and three-dimensional macroscopic diffusion. It is intriguing that this region is poorly conserved even among closely related mammals, and raises the possibility that cellular parameters such as protein abundance and genome size might require a different level of processivity. Alternatively, it may be that animals differ in the efficiency with which they are able to repair deaminated and alkylated damage.

The rate of intersegmental transfer that we measured at μM concentration of DNA is slow, relative to intramolecular searching, but intersegmental transfer and intramolecular transfer are predicted to occur on similar time scales at the mM concentration of DNA that is in the nucleus. This conclusion is supported by the observation that AAG shows very rapid transfer rates between PEG-tethered duplexes for which the tether increases the effective concentration into the mM range (Figure 2-5). The ability of AAG to engage in intersegmental transfer optimizes the efficiency of the search for DNA damage by allowing a searching protein to escape a local DNA site. Electrostatic interactions are the dominant factor influencing the ability of AAG to perform both intramolecular hops (11,18) and intersegmental transfer. As most DNA binding proteins employ positively charged DNA binding sites, these searching mechanisms may apply more broadly to other proteins that search the genome, including other DNA repair and replication proteins, transcription factors, and DNA modifying enzymes.

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. Halford, S. E., and Marko, J. F. (2004) How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.* **32**, 3040-3052

3. Friedman, J. I., and Stivers, J. T. (2010) Detection of damaged DNA bases by DNA glycosylase enzymes. *Biochemistry* **49**, 4957-4967

4. Doucleff, M., and Clore, G. M. (2008) Global jumping and domain-specific intersegment transfer between DNA cognate sites of the multidomain transcription factor Oct-1. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 13871-13876

5. Iwahara, J., and Clore, G. M. (2006) Direct observation of enhanced translocation of a homeodomain between DNA cognate sites by NMR exchange spectroscopy. *Journal of the American Chemical Society* **128**, 404-405

6. Lieberman, B. A., and Nordeen, S. K. (1997) DNA intersegment transfer, how steroid receptors search for a target site. *J. Biol. Chem.* **272**, 1061-1068

7. Fried, M. G., and Crothers, D. M. (1984) Kinetics and mechanism in the reaction of gene regulatory proteins with DNA. *Journal of molecular biology* **172**, 263-282

8. Ruusala, T., and Crothers, D. M. (1992) Sliding and intermolecular transfer of the lac repressor: kinetic perturbation of a reaction intermediate by a distant DNA sequence. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 4903-4907

9. Kozlov, A. G., and Lohman, T. M. (2002) Kinetic mechanism of direct transfer of Escherichia coli SSB tetramers between single-stranded DNA molecules. *Biochemistry* **41**, 11611-11627

10. Gorman, J., Wang, F., Redding, S., Plys, A. J., Fazio, T., Wind, S., Alani, E. E., and Greene, E. C. (2012) Single-molecule imaging reveals target-search mechanisms during DNA mismatch repair. *Proc. Natl. Acad. Sci. U. S. A.*

11. Hedglin, M., and O'Brien, P. J. (2008) Human alkyladenine DNA glycosylase employs a processive search for DNA damage. *Biochemistry* **47**, 11434-11445

12. 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

13. O'Brien, P. J., and Ellenberger, T. (2004) Dissecting the broad substrate specificity of human 3-methyladenine-DNA glycosylase. *The Journal of biological chemistry* **279**, 9750-9757

14.	O'Connor, T. R. (1993) Purification and characterization of human 3-methyladenine-DNA glycosylase. *Nucleic acids research* **21**, 5561-5569

15.	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. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 13087-13092

16.	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. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 13573-13578

17.	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

18.	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

19.	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

20.	Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F., and Jones, D. T. (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.* **337**, 635-645

21.	Dyson, H. J., and Wright, P. E. (2005) Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **6**, 197-208

22.	Hegde, M. L., Hazra, T. K., and Mitra, S. (2010) Functions of disordered regions in mammalian early base excision repair proteins. *Cell. Mol. Life Sci.* **67**, 3573-3587

23.	Vuzman, D., and Levy, Y. (2012) Intrinsically disordered regions as affinity tuners in protein-DNA interactions. *Mol. BioSyst.* **8**, 47-57

24.	Vuzman, D., Polonsky, M., and Levy, Y. (2010) Facilitated DNA search by multidomain transcription factors: cross talk via a flexible linker. *Biophys. J.* **99**, 1202-1211

25.	Baldwin, M. R., and O'Brien, P. J. (2010) Nonspecific DNA binding and coordination of the first two steps of base excision repair. *Biochemistry* **49**, 7879-7891

26.	Baldwin, M. R., and O'Brien, P. J. (2012) Defining the functional footprint for recognition and repair of deaminated DNA. *Nucleic acids research* **40**, 11638-11647

27.   Terry, B. J., Jack, W. E., and Modrich, P. (1985) Facilitated diffusion during catalysis by EcoRI endonuclease. Nonspecific interactions in EcoRI catalysis. *The Journal of biological chemistry* **260**, 13130-13137

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. *The EMBO journal* **19**, 6546-6557

29.   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

30.   Lyons, D. M., and O'Brien, P. J. (2009) Efficient recognition of an unpaired lesion by a DNA repair glycosylase. *Journal of the American Chemical Society* **131**, 17742-17743

31.   Kienberger, F., Pastushenko, C., Kada, G., Gruber, H., Riener, C., Schindler, H., and Hinterdorfer, P. (2000) Static and dynamical properties of single poly(ethylene glycol) molecules investigated by force spectroscopy. *Single Mol* **1**, 123-128

32.   Krishnamurthy, V. M., Semetey, V., Bracher, P. J., Shen, N., and Whitesides, G. M. (2007) Dependence of effective molarity on linker length for an intramolecular protein-ligand system. *J. Am. Chem. Soc.* **129**, 1312-1320

33.   Tian, L., and Heyduk, T. (2009) Bivalent ligands with long nanometer-scale flexible linkers. *Biochemistry* **48**, 264-275

34.   Takayama, Y., and Clore, G. M. (2011) Intra- and intermolecular translocation of the bi-domain transcription factor Oct1 characterized by liquid crystal and paramagnetic NMR. *Proc. Natl. Acad. Sci. U. S. A.* **108**, E169-176

35.   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

36.   Vuzman, D., Azia, A., and Levy, Y. (2010) Searching DNA via a "Monkey Bar" mechanism: the significance of disordered tails. *J. Mol. Biol.* **396**, 674-684

37.   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

38.   von Hippel, P. H., and Berg, O. G. (1989) Facilitated target location in biological systems. *The Journal of biological chemistry* **264**, 675-678

39.   Sheinman, M., and Kafri, Y. (2009) The effects of intersegmental transfers on target location by proteins. *Phys. Biol.* **6**, 016003

40.   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.

Proceedings of the National Academy of Sciences of the United States of America **102**, 15883-15888

41. Halford, S. E. (2001) Hopping, jumping and looping by restriction enzymes. *Biochem. Soc. Trans.* **29**, 363-374

42. Wentzell, L. M., and Halford, S. E. (1998) DNA looping by the Sfi I restriction endonuclease. *J. Mol. Biol.* **281**, 433-444

Appendix A

Additional equation derivation to support Chapter 2

Scheme A-1. Fragments formed from base excision and NaOH cleavage of two-lesion substrates.

$$
\begin{array}{llll}
E_0 & * \underline{12} * \\
& ABC \\[8pt]
E_1 & * \underline{} \; \underline{2} * \\
& A BC \\[8pt]
E_2 & * \underline{1} \; \underline{} * \\
& AB C \\[8pt]
E_{1,2}/E_{2,1} & * \underline{} \; \underline{} \; \underline{} * \\
& A B C
\end{array}
$$
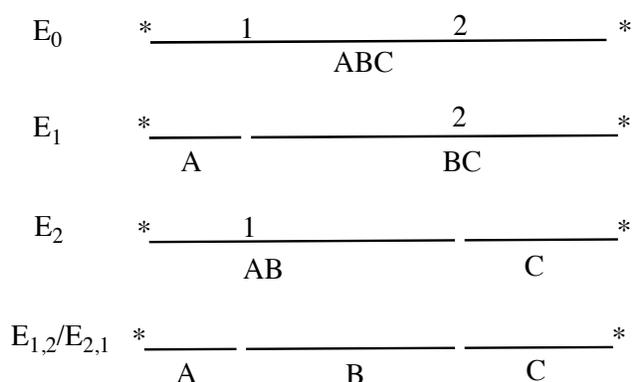
The analysis of two-lesion processivity substrates under steady-state conditions has been previously discussed (1). We refer to the fragments AB and BC (scheme A-1) as intermediates, because they contain a site of damage that could be acted on by AAG. Under processive conditions (low to intermediate salt) these intermediates do not accumulate, because AAG excises both lesions before dissociating. However, in the pre-steady state phase the two excision steps must be sequential because only one AAG molecule is bound to any given DNA. Below we present a simplified kinetic scheme and derive the expression for transient build-up and breakdown of the intermediate.

$$
x \xrightarrow{k_1} z_1 + y_1 \begin{array}{c} \nearrow^{k_2} z_2 \\ \searrow_{k_3} y_2 \end{array}
$$
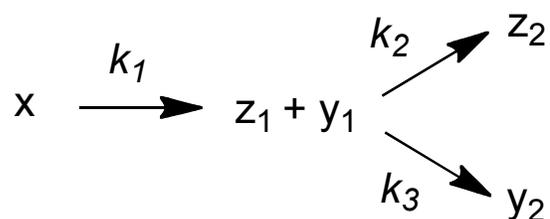
Scheme A-2. Reaction steps for activity of AAG on a two-lesion, two-label substrate (Figure 2-2).

Processing of substrate (x) gives two labeled DNA fragments according to scheme A-2 (the lesion-containing strand is labeled at both 5' and 3' end with fluorescein; we can ignore the unlabeled central fragment corresponding to processive cleavage at both lesion sites). The first

cleavage event gives rise to one product (z_1) and one intermediate (y_1). Subsequently the intermediate can be released to form y_2, or the second lesion can be excised to give z_2. Under the experimental conditions (10:1 substrate:AAG), AAG is more likely to bind to the excess of substrate that is present rather than to rebind a released intermediate. We previously showed that AAG has identical activity toward the two lesion sites, such that 50% of the events are initiated at site 1 (1). Therefore, the initial product (z_1) has equal concentrations of the two fragments and C (Scheme A-1). Equations for the conservation of mass can be written (Eq S1 & S2) and combined to give Eq S3 (x_0 = initial concentration of x).

$$z_1 = x_0 - x \tag{S1}$$
$$x_0 = x + (y_1 + y_2 + z_1 + z_2)/2 \tag{S2}$$
$$x_0 = y_1 + y_2 + z_2 + x \tag{S3}$$

The burst phase for the disappearance of x is given by a single exponential function (Eq S4). The burst kinetics for the intermediate is identical to that of a two step-irreversible pathway (2), except that the rate constant for the second step is the sum of the rate constants to the two pathways shown (Eq S5). The ratio of the products formed during the burst phase is determined by the ratio of the rate constants according to Eq S6. Substitution of Eq S6 into Eq S3 and rearrangement gives Eq S7. The total intermediates (Eq S8) can be solved by adding together Eq S5 and Eq S7, which yields Eq S9 after rearrangement. Both Eq S4 and Eq S9 were multiplied by the burst amplitude (~10% for a 1:10 ratio of enzyme to DNA), and fit to the pre-steady state data in Figure. 2-5 of the text. The steady-state reaction phase that occurs after the burst phase was fit by adding a linear term ($+Vt$;, in which V is the steady state velocity).

$$x = x_o \exp(-k_1 t) \tag{S4}$$
$$y_1 = (k_1 x_0)(\exp(-k_1 t) - \exp(-(k_2+k_3)t)/(k_2+k_3-k_1)) \tag{S5}$$
$$z_2/y_2 = k_2/k_3 \tag{S6}$$
$$y_2 = (x_0 - x - y_1)(k_3/(k_2+k_3)) \tag{S7}$$
$$y_{total} = y_1 + y_2 \tag{S8}$$
$$y_{total} = x_0[(k_3/(k_2+k_3)) + (1/(k_2+k_3))((k_1 k_2/(k_2+k_3-k_1)) - k_3) \times \exp(-k_1 t) - (k_1 k_2/(k_2+k_3)(k_2+k_3-k_1)) \exp(-(k_2+k_3)t)] \tag{S9}$$

Reference

1.   Hedglin M & O'Brien PJ (2008) Human alkyladenine DNA glycosylase employs a processive search for DNA damage. *Biochemistry* 47(44):11434-11445.

2.   Fersht A (1999) *Structure and Mechanism in Protein Science* (W.H. Freeman, New York).

Chapter 3

Excision Efficiency of AAG on Substrate DNA and Its Dependence on Salt Concentration[1]

Many DNA-binding enzymes interact with rare specific sites in a genome comprised of mostly nonspecific DNA, and this presents a daunting task for these enzymes, as they must find these rare target sites promptly during each cell cycle to perform their proper function. An example is the human DNA glycosylase, a family of initiating enzymes in the base excision repair pathway, which need to locate and repair ~10,000 base lesions per human cell per day in a genome of approximately 10^{10} nucleotides (1). Almost a dozen human DNA glycosylases continuously and independently search the genome for their cognate lesion bases, which usually have minimal impact on the structure of the DNA double helix (2,3). Previous studies have shown that many of these enzymes use facilitated diffusion (4-6), which enhances the efficiency of locating the target sites by reducing the dimensionality of the searching space. Facilitated diffusion can be measured experimentally by processivity assays, which measure the fraction of correlated catalytic events on a defined piece of DNA during a single encounter with the enzyme. Subsequently, the searching mechanism can be further dissected by varying the reaction conditions of the processivity assay and/or the relative positions of the target sites on the substrate DNA.

Human alkyladenine DNA glycosylase (AAG) has a broad substrate range, catalyzing the removal of alkylated and deaminated purines such as 1, N^6-ethenoadenine (εA) and hypoxanthine (Hx) (7-10). It has been shown in multiple-turnover processivity assays that AAG uses facilitated diffusion to search for its target lesion sites (11,12). The processivity assay monitors intramolecular transfer between target sites by using catalytic excision of the sites as the readout. Experimentally, the fraction processive is the observed parameter to describe the fraction of correlated actions of an enzyme on its substrate, and is equivalent to the fraction of

---

processive events in all the events (F_P = processive events/total events). Mechanistically, the fraction processive value (F_P) depends on two kinetically determined probabilities, namely the intramolecular transfer probability (P_{trans}) between the target sites and the excision efficiency (E_{ex}) of the target lesion, by the equation $F_P = P_{trans} \times E_{ex}$. The excision efficiency of the protein on the target site can be measured using a pulse-chase assay (13,14). This pulse-chase assay approach has been previously adopted by Stivers and co-workers in their studies on uracil DNA glycosylase (4,15-17), and more recently on 8-oxoguanine DNA glycosylase (18). The site transfer probability can subsequently be isolated by dividing F_P by the measured excision efficiency value.

In this study, we used the pulse-chase assay, which monitors the partition of chemical catalysis and dissociation of the enzyme from the enzyme-DNA complex, to measure the excision efficiency of AAG. Results show that the commitment to catalysis by AAG decreases as the salt concentration is increased. Under low salt conditions, the observed excision efficiency is dependent on the chase concentration used, which reveals at least one AAG-DNA species that is susceptible to DNA-mediated intersegmental transfer. The non-zero values of excision efficiency at very high chase concentrations indicate that an AAG-DNA complex resistant to intersegmental transfer is present on the pathway towards catalysis. The separation of excision efficiency and site transfer probability from F_P enables the direct comparison of the searching ability of AAG with that of other enzymes.

MATERIALS AND METHODS

Proteins

Full-length human AAG was cloned into the pET SUMO expression vector (Invitrogen) between the N-terminal BamHI and C-terminal XhoI site, encoding an N-terminal 6× His-Smt3 fusion protein that can be cleaved with the SUMO protease ULP1. The AAG fusion protein was expressed in *E. coli* BL21 Star (DE3) pRare2 cells using autoinduction media (19). Typically, cells were grown at 37 °C until they reached an optical density of 1.0 at 600 nm. Cultures were then transferred to a 16 °C shaker, and protein was expressed for 24 hours with final OD_{600} values at ~20 before cells were harvested and frozen at -80 °C. Cell pellets were suspended in lysis buffer (25 mM potassium phosphate, pH 7.0, 250 mM NaCl, 10% (v/v) Glycerol, 0.1% NP-40) and disrupted by a high pressure homogenizer (EmulsiFlex, Avestin). Cell lysates were

slowly precipitated with polyethylenimine to remove nucleic acids, and the recombinant AAG protein was purified by metal affinity chromatography using a His tag that was subsequently cleaved with the SUMO protease ULP1 to produce full-length AAG with an N-terminal serine in place of the initiator methionine of the native protein. Subsequent ion exchange chromatography (HiTrap SP, GE Healthcare) and dialysis (MWCO 12,000-14,000, Fisherbrand) yielded recombinant protein that was homogeneous as judged by SDS-PAGE with Coomassie staining.

The N-terminally truncated protein Δ80 AAG was expressed and purified as previously described (20). The starting residues $K^{80}G^{81}H^{82}L^{83}$ have been substituted by $G^{80}P^{81}H^{82}M^{83}$ that remain after proteolytic cleavage by human rhinovirus 3C protease.

The active concentrations of FL and Δ80 AAG were determined by burst analysis as described previously (11), and the enzyme concentrations used in the experiments described here refer to the corrected active concentration.

Oligonucleotides

DNA substrates were synthesized by Integrated DNA Technologies or the Keck Center at Yale University and purified by denaturing PAGE as previous described (11). Oligo concentrations were determined from the absorbance at 260 nm using the calculated extinction coefficients. For 5' fluorescein-labeled oligonucleotides we assessed the labeling efficiency by comparing the absorbance at 260 nm with that at 495 nm. DNA duplexes were annealed at 1:1 ratio of the lesion containing fluorescein-labeled and the complement strand. DNA substrates used in this study are listed in Figure 3-1.

Gel-Based Glycosylase Assay

Reactions were carried out at 37 °C in a buffer system consisted of 50 mM NaMES pH 6.1, 10% (v/v) glycerol, 0.1 mg/mL BSA, 1 mM DTT, 1 mM EDTA and varying concentrations of NaCl to obtain the desired ionic strength. Reactions were initiated by the addition of enzyme to a final reaction volume of 20-60 μL. Aliquots were withdrawn at various times and quenched with NaOH (0.2 M final concentration). Samples were heated at 70 °C for 15 min, loading buffer consisting of 10 mM EDTA and 98% formamide was added, and the DNA species were resolved on 14% (w/v) denaturing polyacrylamide gels with 8 M urea. Gels were scanned with a Typhoon Trio+ fluorescence imager (GE Healthcare), and fluorescein signal was detected using excitation wavelength at 532 nm and emission with a 520BP40 filter. The resulting fluorescence signal was quantified with ImageQuant TL and corrected for background signal. The intensity of each DNA

band was converted into a fraction of the total DNA by dividing its intensity by the sum of the intensities for all of the DNA species in the reaction.

Pulse-Chase Assay

To measure the excision efficiency of AAG on its substrate DNA, we conducted pulse-chase assays at 37 °C in the standard buffer at different salt concentrations by the addition of NaCl. The pulse-chase measurements were performed as previously described (12). Briefly, 50 nM fluorescein-labeled 25εA•T substrate DNA was mixed with 100 nM FL or Δ80 AAG for 20 sec (t_1), and then chase solutions containing 1-100 μM of unlabeled 25εA•T or 25Y•T DNA were added (Figure 3-2A). The effectiveness of the DNA chase species was tested in control reactions conducted by pre-incubating chase DNA and enzyme together before the substrate was added to initiate the reaction (Appendix Figure B-1). At various time points (t_2), a sample from the reaction was removed, quenched and analyzed as described for Glycosylase Activity Assays.

The committed base excision pathway results in fluorescein-labeled product formation following a single-exponential, whereas dissociation releases unreacted fluorescein-labeled substrate. Depending on the chase concentration used, a steady-state product formation phase may be observed. The partitioning between hydrolysis and dissociation can be determined from either the exponential rate constant or by the change in burst amplitude in comparison with the no chase control reaction. 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. The data were converted to fraction product (fraction product = product/(product + substrate)) and fit by a single exponential followed by a steady-state phase.

During the 20 sec enzyme-substrate pre-incubation, the substrate is converted to product at a rate constant of approximately 0.2 min^{-1} (the rate constant is obtained from the single-turnover reaction), and this corresponds to a product formation of 6.4% of the substrate in principle, which is not negligible. Therefore, we took time points at 20 sec in the single-turnover assay and used the measured result as the baseline level of product formation when the chase species was added (A_{20s}). This term was incorporated into curve fitting for the pulse-chase assay so that the pre-existing product was not included in the burst phase. After obtaining the burst amplitude (A_{PC}) values from the data fitting, the efficiency of excision can be determined as compared to the same experiment without chase, i.e., the single-turnover reaction (A_{max}), by Eq

1:

$$\text{Efficiency of Excision } (E_{ex}) = A_{PC}/(A_{max} - A_{20s}) \tag{1}$$

Alternatively, the excision efficiency can be determined from $k_{off,obs}$ and k_{max} values (Eq 2), where $k_{off,obs}$ is the observed rate constant for dissociation from the bound complex and k_{max} is the maximal single turnover rate constant for formation of product. For branched pathways, the observed rate constant (k_{obs}) for the burst phase of the pulse-chase experiment is given by the sum of the rate constants for the competing pathways (Eq 3). Therefore, the efficiency of excision can be calculated using Eq 4.

$$\text{Efficiency of Excision } (E_{ex}) = k_{max}/(k_{off,obs} + k_{max}) \tag{2}$$

$$k_{obs} = k_{off,obs} + k_{max} \tag{3}$$

$$\text{Efficiency of Excision } (E_{ex}) = k_{max}/k_{obs} \tag{4}$$

The two methods give very similar results, while the calculations using burst amplitudes are more reproducible among different measurements. Therefore, we used the amplitude values for excision efficiency calculation throughout this study.

The excision efficiency measurements were performed at different salt concentrations ranging from 100 mM to 1000 mM Na^+. We have fit the salt dependent results to a cooperative model in which cations can bind to multiple sites on the DNA and thereby affect the efficiency of excision. This model is analogous to the Hill equation, which takes the following form. $E_{ex} = E_{ex,max} - \Delta E_{ex}*I^n/(K_a^n + I^n)$, where E_{ex} is the efficiency of excision, $E_{ex,max}$ the maximal efficiency observed, ΔE_{ex} the difference between the maximal and minimal observed efficiency, n the number of cation binding sites, K_a the average association rate constant for cation binding, and I the cation concentration. All the curve fits performed in this study use KaleidaGraph (Synergy Software).

RESULTS

Efficiency of Excision Measured by Pulse-chase Assays

Pulse-chase assays were performed to measure the commitment of AAG on the εA duplex DNA, and the extent of commitment is represented by the excision efficiency of catalyzed εA removal. The assays were set up under single-turnover conditions with 50 nM fluorescein-labeled DNA substrate and 100 nM AAG, and then excess unlabeled chase DNA was added to bind the dissociated free enzyme after pre-incubation of the enzyme and substrate for

20 sec (Figure 3-2A). Two DNA species were used as chase, unlabeled 25εA•T substrate DNA

and 25Y•T competitor DNA (Y is short for pyrrolidine, which is a transition state ribose analog)

(Figure 3-1) (21). In the presence of the chase DNA, there should be a decrease in the amplitude

and an increase in the rate constant of the exponential phase compared to the single-turnover

chase-free control reaction for any reaction that is not 100% committed. The two chase species

were tested in the pulse-chase assay under the same reaction conditions, and they gave identical

burst amplitudes (~0.5 at 250 mM Na$^+$) and different steady-state rates, which originates from

the different properties of the two chase species (Figure 3-2B). When the unlabeled 25εA•T

DNA is used as chase, the trapped enzyme is capable of reacting with the chase species,

dissociating from the resulting unlabeled product DNA, and then enters another catalytic cycle

```
17ET   *5'-AGCATCCTECCTTCTCT-3'
        3'-TCGTAGGATGGAAGAGA-5'

19ET   *5'-TAGCATCCTECCTTCTCTC-3'
        3'-ATCGTAGGATGGAAGAGAG-5'

25ET   *5'-CGATAGCATCCTECCTTCTCTCCAT-3'
        3'-GCTATCGTAGGATGGAAGAGAGGTA-5'

49ET   *5'-GACATGATTGCCCGATAGCATCCTECCTTCTCTCCATGCGTCAATTGTC-3'
        3'-CTGTACTAACGGGCTATCGTAGGATGGAAGAGAGGTACGCAGTTAACAG-5'

25XT    5'-CGATAGCATCCTXCCTTCTCTCCAT-3'
        3'-GCTATCGTAGGATGGAAGAGAGGTA-5', X=E or Y
```


Figure 3-1. DNA oligo sequences used in the pulse-chase assays.
Substrate DNA oligos are labeled with fluorescein, which is denoted by an asterisk; chase DNA oligos are not labeled. The target lesion site 1, N[6]-ethenoadenine (E, or εA in the main text) and the transition state ribose mimic pyrrolidine (Y) were underlined.

with the substrate and chase DNA. This results in a chase concentration dependent steady-state

formation of the labeled product that is inversely proportional to the concentration of the chase.

Pyrrolidine-containing chase DNA binds tightly to the enzyme and does not undergo catalytic

turnover (22), leading to a slower steady-state rate as compared to the 25εA•T chase at the same

chase concentration. Consistent with this analysis, increasing the concentration of the chase

DNA from 1 μM to 5 μM, and then to 10 μM, decreased the steady-state rates significantly, especially for the 25εA•T chase, while the exponential phase remained unchanged. Therefore, the two chase species gave the same excision efficiency under the same reaction condition.
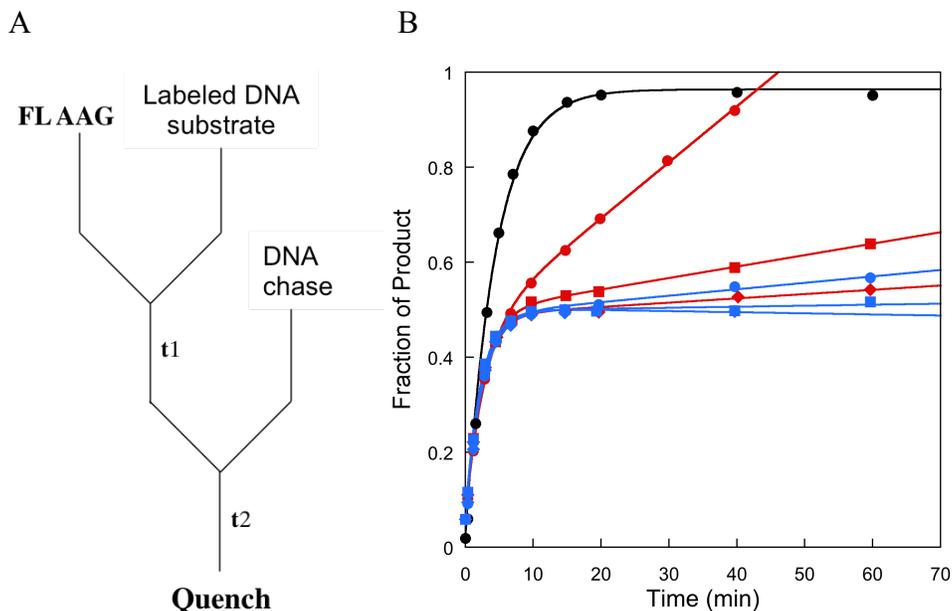
Figure 3-2. Pulse-chase reaction measures the partition of product formation and dissociation of the enzyme from the enzyme-DNA complex.
A. Pulse-chase reaction scheme. AAG and fluorescein-labeled substrate DNA are incubated for 20 sec (t_1), and time points were taken from 20 sec to 60 min (t_2) after chase addition. Reactions were quenched with 0.2 M NaOH (final concentration). B. A representative reaction time course for single-turnover (filled black circle) and pulse-chase assays using 25εA•T (red symbols) or 25Y•T chase (blue symbols) at 250 mM Na$^+$. The reactions include 50 nM 25εA•T labeled substrate, 100 nM FL AAG and different concentrations of chase at 1 μM (circle), 5 μM (square) and 10 μM (diamond). Different steady-state phases were observed depending on the identity and concentration of the chase species, but all reactions had the same exponential phase.

We next investigated the dependence of the excision efficiency on substrate length, and DNA oligos with a central εA lesion in the same sequence context at different lengths ranging from 17 bp to 49 bp were tested (Figure 3-3). Experiments were performed with 5 μM 25εA•T chase at 250 mM Na$^+$, and results clearly show that the excision efficiency of εA removal by AAG is dependent on the length of the substrate below 25 bp (0.32 for 17εA•T and 0.35 for 19εA•T), and then approaches a plateau at around 25 bp and above (~0.5 for 25εA•T and 49εA•T). Based on this result, all the subsequent assays used the 25εA•T DNA, unless otherwise specified, as the model substrate for measurements.
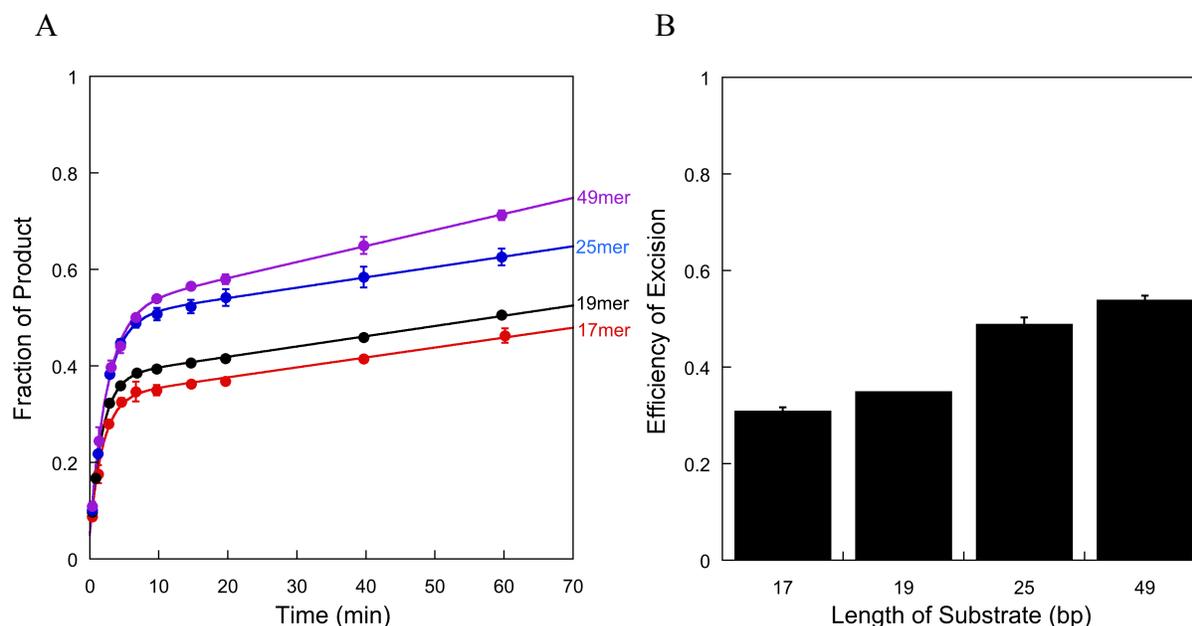
Figure 3-3. Length dependence of the efficiency of excision measurement.
A. Plots of product formation of the pulse-chase assay at 250 mM Na[+] using DNA substrates of different lengths. These reactions contain 50 nM labeled DNA substrate, 100 nM FL AAG and 5 μM unlabeled 25εA•T chase. B. Bar graph presentation of the efficiency of excision versus the length of substrates. Reactions were performed in duplicate and the mean ± S.D. is plotted, except for the reaction on the19εA•T substrate (single measurement).

The catalytic mechanism of AAG on the εA substrate DNA has been proposed to include fast, two-step binding to specific DNA, thermodynamically favorable nucleotide flipping and rate-limiting N-glycosidic bond cleavage (Appendix Figure B-4) (23,24). Some of the steps, such as AAG-DNA association and dissociation, are salt concentration dependent, while the chemical catalysis step is not sensitive to salt concentration changes (11,24). Because the excision efficiency measurement monitors the partitioning of chemical catalysis and macroscopic dissociation of the enzyme from the bound complex, it is expected to be salt dependent. Therefore, we performed the excision efficiency assay in reaction solutions containing 100 mM to 1000 mM Na[+], and a fixed chase concentration of 1 μM. At 100-200 mM Na[+], concentrated chase stocks were added directly into the reaction solution. At 250-1000 mM Na[+], substrate DNA and AAG were pre-incubated in the standard glycosylase reaction buffer with a salt concentration of 100 mM to make sure that all the substrates started in the bound complex, and chase solutions were prepared at different salt concentrations in order to achieve the specified final salt concentrations. Under the experimental setting, all the substrate DNA is in the specific complex with AAG for initial lesion recognition when the chase DNA is added, since searching has been shown to be very fast and efficient. Representative reaction curves are shown in Figure

55

3-4A, and the resultant excision efficiency values are plotted as a function of salt concentration in Figure 3-4B based on Eq 1 presented in Materials and Methods. It is shown clearly that the excision efficiency of AAG on its substrate DNA is dependent on the salt concentration. At low salt, AAG is very committed to the εA lesion, and the excision efficiency is close to unity; at high salt, however, AAG is less committed and the excision efficiency drops steeply. Interestingly, the excision efficiency of AAG did not decrease to 0 even at 1000 mM Na$^+$.

A   B
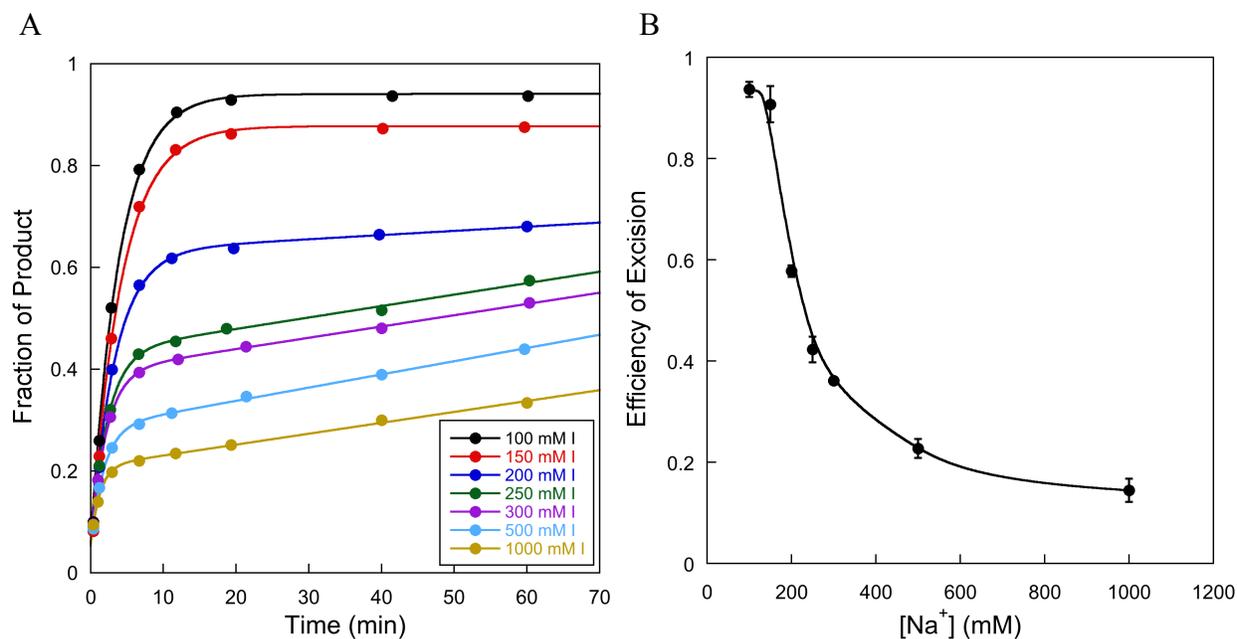
**Figure 3-4. Excision efficiency of AAG is salt concentration dependent.**
A. Representative time course of product formation using 1 μM chase at 100-1000 mM Na$^+$. B. Summary plot of the excision efficiency as a function of salt conditions. Values are the mean ± S.D. (n = 3). A line is drawn for the dependence of excision efficiency on salt concentration to guide the eye.

We then used higher chase concentrations to validate the results observed in Figure 3-4, and representative reaction curves are shown in Figure 3-5A using 40 (for the reaction at 1000 mM Na$^+$) or 80 (for reactions at 100-500 mM Na$^+$) μM chase. Most of the excision efficiency values observed here are the same within error to those obtained at 1 μM chase, but unexpectedly the excision efficiency values at low salt are lower here than observed before (Figure 3-5B). Therefore, we decided to measure a complete dependence profile of the excision efficiency on the chase concentration, and we saw persistent chase concentration dependence at low salt concentrations (≤ 200 mM Na$^+$), but not at high salt (≥ 250 mM Na$^+$) (Figure 3-6).
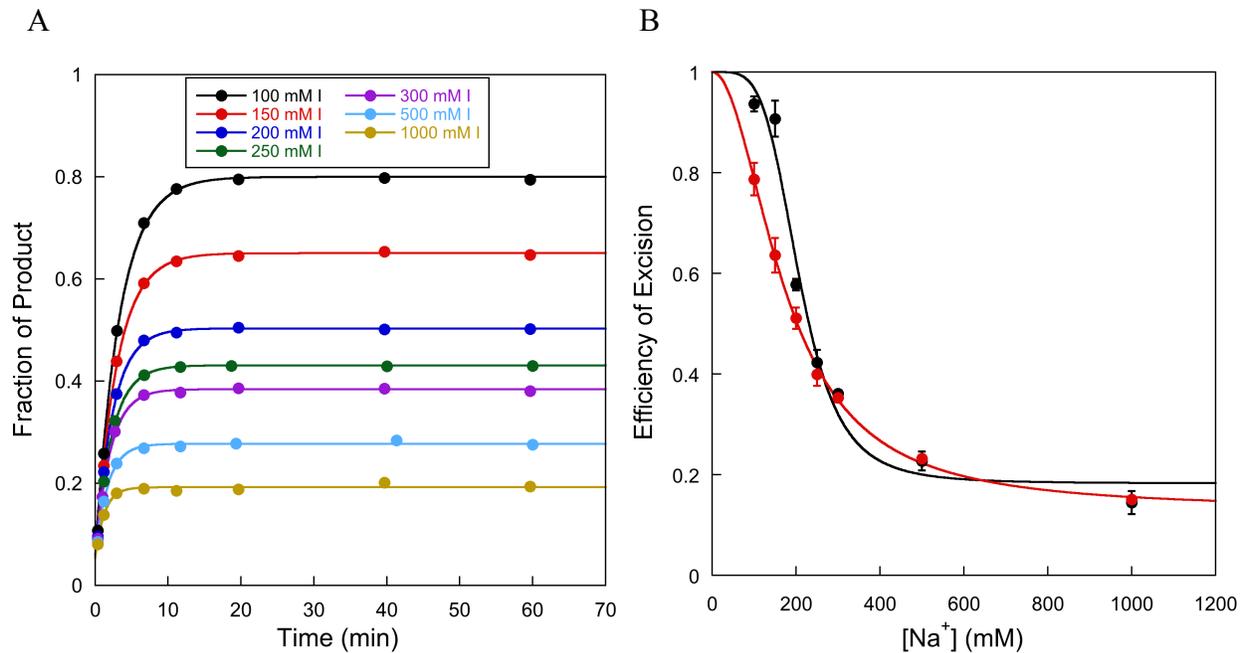
Figure 3-5. Excision efficiency of AAG is chase and salt concentration dependent.
A. Representative time course of product formation using 80 μM chase at 100-500 mM Na$^+$ and 40 μM chase at 1000 mM Na$^+$. B. Summary plot of the excision efficiency under different salt conditions. Values obtained at low chase concentration (1 μM) are shown in black and those obtained at high chase concentrations (40 or 80 μM) are in red. Values are the mean ± S.D. (n = 3). The two sets of results obtained at low and high chase concentrations are fit by the cation cooperative binding equation described in Materials and Methods.
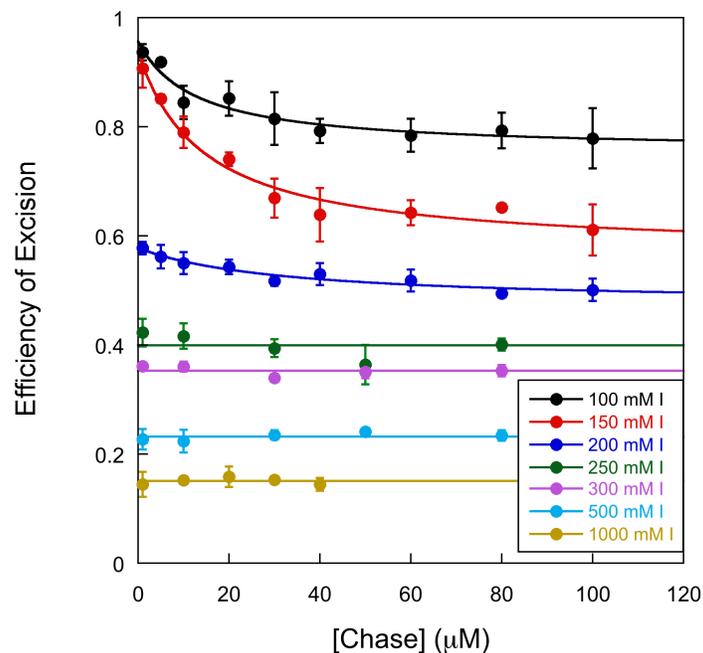
Figure 3-6. Excision efficiency of AAG is dependent on the concentration of salt and DNA chase.
Pulse-chase assays were performed with 50 nM 25εA•T substrate, 100 nM FL AAG and 1-100 μM unlabeled 25Y•T chase at 100-1000 mM Na$^+$. The excision efficiency is dependent on the chase concentration at 100-200 mM Na$^+$, but not at 250 mM Na$^+$ and above. All reactions were performed at least in triplicate, and the mean and standard deviation are shown. Lines are shown to guide the eye.

A dependence on chase concentration could have indicated that insufficient chase was used to trap all dissociated enzyme. Therefore, we tested the possibility that the observed chase concentration dependence is due to experimental artifact at low chase concentrations, although we do not deem this very likely since the dependence was only observed at low salt concentrations and both chase species gave the same result. Pulse-chase control experiments were performed using both chase species to test their abilities to trap the enzyme at low chase concentrations (Appendix Figure B-1). Although the two chases behaved differently in the control assays, they were both shown to be effective even at 1 μM, which is the lowest chase concentration used in all the measurements. As a further control, we conducted the pulse-chase assay at lower enzyme and substrate concentrations (still ensuring the single-turnover condition), while still keeping the chase concentration at 1 μM. The rationale is that if the observed high excision efficiency is due to insufficient chase, we would be able to see a decrease in the value under the new reaction condition; otherwise, the observed excision efficiency would be the same for the two reactions. The pulse-chase assay was performed at 150 mM Na$^+$, and we observed no difference in the excision efficiency for the two pulse-chase assays (Appendix Figure B-2). Together, these two sets of control experiments suggest that the observed chase concentration dependence is not an artifact cased by insufficient chase concentrations.

The reaction conditions under which the chase concentration dependence is observed also support intersegmental transfer of AAG from the substrate molecule to a competitor or a different substrate DNA molecule (25). Therefore, we concluded that AAG is capable of intersegmental transfer from certain bound complex(es) to the excess chase DNA on the pathway to catalysis (Appendix Figure B-4). The excision efficiency at very high chase concentrations suggests that an AAG-DNA complex exists that is resistant to intersegmental transfer and is present on the pathway towards catalysis.

Consistent with the intersegmental transfer model, we observed similar but steeper concentration dependence on the 17εA•T substrate compare to the 25εA•T substrate at 150 mM Na$^+$ (Appendix Figure B-3). Steeper chase concentration dependence is expected on the shorter DNA substrate because the shorter DNA molecule has fewer binding sites than the longer one for a microscopically dissociated protein to bind to during intersegmental transfer.

The chase concentration dependence of 25εA•T results in two excision efficiency values at the two concentration extremes at low salt, producing 0.94 and 0.79, 0.91 and 0.64, 0.58 and

0.51 at 100, 150 and 200 mM Na$^+$ respectively (Figure 3-6). At low chase concentrations, AAG has a high probability of reassociating with the substrate DNA after it microscopically dissociates from the complex. These microscopic events can be intercepted by collision with a high concentration of chase molecules.

AAG has a poorly conserved N-terminus comprised of 80 residues, which is not essential for its glycosylase activity but has been suggested to fine tune its binding to DNA. The truncated protein, Δ80 AAG, has been shown to exhibit decreased processivity compared to the full-length enzyme (11). Therefore, we tested whether the truncation of the protein N-terminus also affects the excision efficiency. Pulse-chase assays were performed at 100, 150 and 200 mM Na$^+$ (Figure 3-7), and the excision efficiency values were lower than those for the full-length enzyme under the same reaction conditions. Similar to that has been observed for the full-length protein, the truncated AAG also shows chase concentration dependence of the excision efficiency. However, the dependence was only clearly seen at 100 mM Na$^+$ but not at 150 and 200 mM Na$^+$, which is similar to the salt dependence for the fraction processivity assay (very processive at 100 mM Na$^+$ but much less so at 150 and 200 mM Na$^+$, see Ref (11))
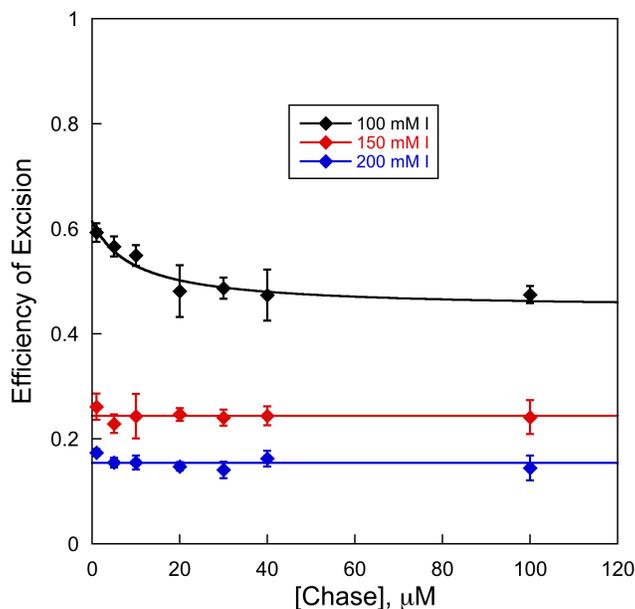

Figure 3-7. Excision efficiency of Δ80 AAG is chase and salt concentration dependent.
Pulse-chase assays were performed with 50 nM 25εA•T substrate, 100 nM Δ80 AAG and 1-100 μM unlabeled 25Y•T chase at 100-200 mM Na$^+$. All reactions were performed at least in triplicate, and the mean and standard deviation are shown. Lines are shown to guide the eye.

Intramolecular Site Transfer by AAG

The previously reported processivity assays measure the fraction of the enzyme molecules that react with one εA site and then transfer and excise εA at the second site without dissociating into bulk solution. The obtained fraction processive (F_P) consists of two distinct and measurable components ($F_P = P_{trans} \times E_{ex}$): the probability of site transfer (P_{trans}) and the efficiency of εA excision (E_{ex}) once the second site is reached. Therefore, the probability of site transfer (P_{trans}) can be isolated by dividing fraction processive (F_P) by the excision efficiency (E_{ex}). P_{trans} values were calculated using F_P values measured in Chapter 2 (Figure 2-4B on 47εA2F2 substrate) and E_{ex} values were measured in this study. At 100-200 mM Na$^+$, the E_{ex} values used for the calculation were extrapolated from the linear fits of 1-10 μM chase data to 0 μM chase; at 250-300 mM Na$^+$, the E_{ex} values were taken from the average results of all the data. As can be seen from Figure 3-8, the probability of site transfer by AAG (red bars) is very high at low salt conditions, and it slowly decreases as salt concentration increases.
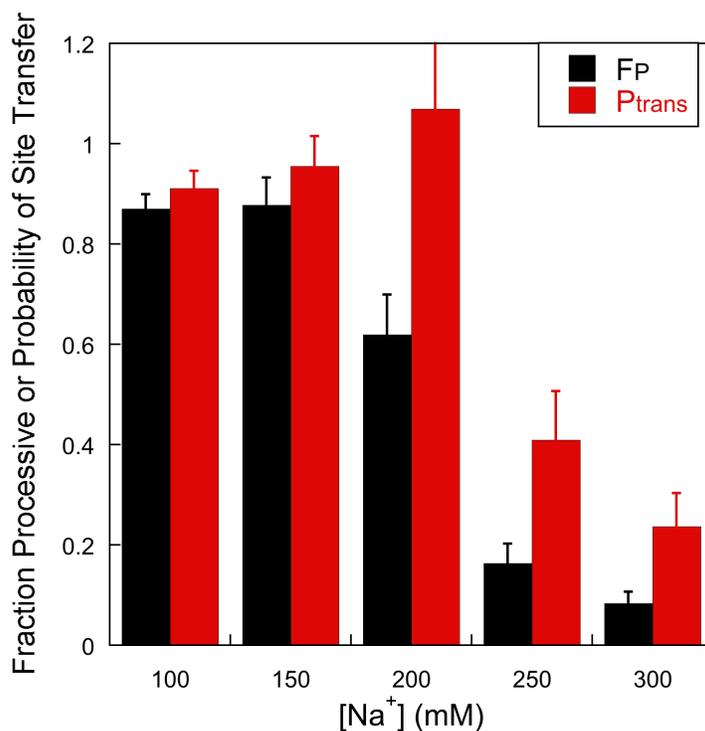
Figure 3-8. Probability of site transfer is salt dependent.
Fraction processivity results (F_P; black bars) were taken from Figure 4B of Ref (25) (see also Chapter 2 of this thesis). The P_{trans} values (red bars) were calculated based on the equation $P_{trans} = F_P/E_{ex}$, where F_P values were as shown in the plot and the E_{ex} values (extrapolated results at [chase] = 0 μM) were taken from this study under the corresponding salt conditions. The error bars for P_{trans} were calculated by error propagation from F_P and E_{ex}.

DISCUSSION

It has been widely accepted that many DNA-binding proteins use facilitated diffusion to search for their specific target sites more efficiently (6,26-29). Different methods have been developed to study this searching process using single-molecule and bulk solution measurements (4,5,11,28,30,31). One of the most frequently employed approaches to characterize the searching ability of a variety of enzymes is the activity-based processivity assay, such as restriction endonucleases EcoRI (32) and EcoRV (28), and DNA repair enzymes uracil DNA glycosylase (UDG) (4) and 8-oxoguanine DNA glycosylase (OGG1) (5). A processive enzyme can catalyze multiple reactions occurred on a defined piece of DNA (synthetic oligo DNA or plasmid DNA) prior to the macroscopic dissociation of the enzyme from the DNA to which it is originally bound. Fraction processive is the observed parameter for quantifying correlated enzymatic actions, but it is not equivalent to the transfer probability. To compare transfer probability among different DNA-binding proteins, we need to separate the two components in the processivity assay: the probability of site transfer and the excision efficiency of the target site. In this current work, we measured the excision efficiency of AAG on its εA substrate DNA as a function of salt concentration. These results also make it possible to directly compare the probability of site transfer of AAG to other proteins.

The excision efficiency of AAG on the εA substrate is DNA length dependent (Figure 3-3), and the minimal oligo length (25 bp) at which optimal excision efficiency is realized is longer than the footprint observed in the AAG-DNA crystal structure (8 bp around the damaged nucleotide) (21,33). Previously, DNase I protection was used to probe the DNA binding footprint on the εA-containing DNA, and a footprint of 11 bp upstream and 5 bp downstream of the lesion site was observed (34,35). However, an excellent agreement with the crystal contacts was observed using a hairpin DNA substrate in a more recent functional footprint assay (36). Based on these results, we postulate that the length effect on the excision efficiency is due to end fraying of short duplex DNA, in which hydrogen bonds break and eventually base pairs become free at the termini of the helix (37-40).

We found that the excision efficiency of AAG on εA DNA was very sensitive to the salt concentration (Figure 3-4; Figure 3-5). This observation emphasizes the necessity of performing a complete salt dependent profile in order to interpret and use the excision efficiency results in a meaningful way. AAG binds to its substrate DNA via a positively charged binding interface (33),

and therefore it was not surprising that salt can affect the partitioning of the enzyme between catalysis and falling off from the DNA substrate. At low salt AAG exhibited high excision efficiency on εA DNA, while at high salt the excision efficiency was drastically decreased. We have fit the salt concentration dependence of the excision efficiency with a cooperative model, in which multiple cations affect the partitioning between catalysis and dissociation of the enzyme (Figure 3-5B). In principle, salt-dependent changes in dissociation rate or excision rate could be responsible for the observed salt dependence, while the base excision step has been shown to be insensitive to salt changes (11). Therefore, the most straightforward interpretation of the salt dependence is that the apparent dissociation rate of AAG from the catalytically competent complex is dependent on salt concentration.

Unexpectedly, we also observed that the excision efficiency was dependent on the chase concentration at low salt (100-200 mM Na^+ for FL AAG) (Figure 3-6). The chase concentration dependence is best explained by the intersegmental transfer model via microscopic dissociation and reassociation of AAG from the substrate DNA to the chase DNA (Appendix Figure B-4). At low chase concentrations, a microscopically dissociated protein has a high probability to return to the substrate DNA to which it was initially bound, resulting in high excision efficiency values. At high chase concentrations, a microscopically dissociated protein has increasing chances to be intercepted by a nearby chase DNA molecule, leading to lower excision efficiency values.

The non-zero excision efficiency values observed at very high chase concentrations could be attributed to the presence of a bound complex species that is impregnable to intersegmental transfer. According to the proposed minimal reaction mechanism of AAG on its substrate DNA (23,24), this bound species could be the flipped-out complex. This model could be tested by performing additional experiments, such as measuring the rate constants for nucleotide flipping and unflipping (k_{flip} and k_{unflip}) directly under these different conditions used for the pulse-chase assays. The model predicts that changing the base pair stability could alter the commitment for target excision, which could be tested easily by altering the identity of the opposing base of the lesion site.

Since AAG has a broad substrate range, and catalyzes the removal of many other lesions such as hypoxanthine (Hx), 3-methyladenine and 7-methylguanine, besides the εA lesion tested here (7-9,34,41,42), it is of interest to test if AAG also excise other lesion bases efficiently in the pulse-chase assay. In a pilot experiment, we measured the excision efficiency of AAG on the Hx

substrate (Appendix Figure B-5) and found the efficiency value much lower than that on the εA substrate under the same reaction condition. The observed differences in the excision efficiency on the two different substrate lesions suggest that it might provide an opportunity for AAG to discriminate between different bases.

To evaluate the effect of the positively charged amino terminus of AAG on the excision efficiency, we performed pulse-chase assays using the truncated enzyme and found that Δ80 AAG had similar chase concentration dependence as the full length protein at low salt, suggesting that the poorly conserved N-terminus is not required for intersegmental transfer, which is consistent with the previous results (25). However, Δ80 AAG clearly has decreased excision efficiency when assayed under the same salt condition as the FL AAG (Figure 3-6; Figure 3-7). This difference could be explained by weaker DNA binding affinity.

The excision efficiency values for AAG on εA DNA are very high at low salt concentrations (Figure 3-6), and essentially every encounter with the target lesion site results in catalytic excision. This high excision efficiency is similar to those reported for UDG on the uracil site (0.73 in Ref (4), ~0.9 in Ref (17)) and for OGG1 on the 8-oxoguanine site (0.85 ± 0.03 in Ref (18)) at low salt conditions. At high salt concentrations, the excision efficiency of AAG on the εA DNA substrate is greatly decreased (Figure 3-6). It is expected that the excision efficiency values of UDG and OGG1 on their respective cognate lesion sites would show similar salt concentration dependence as that observed for AAG, based on the electrostatic interactions in their protein-DNA complex structures (43,44). However, such measurements have yet to be performed for these proteins at high salt.

We calculated intramolecular transfer probability (P_{trans}) of two sites separated by 25 bp (47εA2F2) for AAG, and found that the probability of site transfer values were all close to 1 at 200 mM Na^+ and below (Figure 3-8). Previous study showed that increasing the distance of the two lesion sites from 25 bp (47εA2F2) to 50 bp (72εA2F2) did not change the fraction processivity values (12), and thus the intramolecular transfer probability of AAG. In contrast to the extremely efficient site transfer of AAG, UDG is less efficient for lesion searching: at 10 mM NaCl, the intramolecular site transfer probability was ~0.4 for two sites separated by 20 bp, and ~0.3 for those separated by 56 bases; while at 150 mM NaCl, the transfer probability decreased to around 0 (4). Therefore, it is likely that the less efficient site transfer probability of UDG may have been compensated by its extremely fast reaction rate constant for target excision (45). In

contrast, AAG seems to have evolved to optimize the site transfer ability for its different lesion sites under physiological salt conditions.

References

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

2.   Krokan, H. E., Standal, R., and Slupphaug, G. (1997) DNA glycosylases in the base excision repair of DNA. *The Biochemical journal* **325 (Pt 1)**, 1-16

3.   Lindahl, T., and Wood, R. D. (1999) Quality control by DNA repair. *Science* **286**, 1897-1905

4.   Porecha, R. H., and Stivers, J. T. (2008) Uracil DNA glycosylase uses DNA hopping and short-range sliding to trap extrahelical uracils. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 10791-10796

5.   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

6.   Verdine, G. L., and Bruner, S. D. (1997) How do DNA repair proteins locate damaged bases in the genome? *Chemistry & biology* **4**, 329-334

7.   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. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 12869-12874

8.   O'Connor, T. R. (1993) Purification and characterization of human 3-methyladenine-DNA glycosylase. *Nucleic acids research* **21**, 5561-5569

9.   Saparbaev, M., and Laval, J. (1994) Excision of hypoxanthine from DNA containing dIMP residues by the Escherichia coli, yeast, rat, and human alkylpurine DNA glycosylases. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 5873-5877

10.   Singer, B., Antoccia, A., Basu, A. K., Dosanjh, M. K., Fraenkel-Conrat, H., Gallagher, P. E., Kusmierek, J. T., Qiu, Z. H., and Rydberg, B. (1992) Both purified human 1,N6-ethenoadenine-binding protein and purified human 3-methyladenine-DNA glycosylase act on 1,N6-ethenoadenine and 3-methyladenine. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 9386-9390

11.   Hedglin, M., and O'Brien, P. J. (2008) Human alkyladenine DNA glycosylase employs a processive search for DNA damage. *Biochemistry* **47**, 11434-11445

12.   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

13.   Rose, I. A. (1980) The isotope trapping method: desorption rates of productive E.S complexes. *Methods in enzymology* **64**, 47-59

14.   Rose, I. A., O'Connell, E. L., and Litwin, S. (1974) Determination of the rate of hexokinase-glucose dissociation by the isotope-trapping method. *The Journal of biological chemistry* **249**, 5163-5168

15.   Schonhoft, J. D., and Stivers, J. T. (2012) Timing facilitated site transfer of an enzyme on DNA. *Nature chemical biology* **8**, 205-210

16.   Schonhoft, J. D., Kosowicz, J. G., and Stivers, J. T. (2013) DNA translocation by human uracil DNA glycosylase: role of DNA phosphate charge. *Biochemistry* **52**, 2526-2535

17.   Schonhoft, J. D., and Stivers, J. T. (2013) DNA translocation by human uracil DNA glycosylase: the case of single-stranded DNA and clustered uracils. *Biochemistry* **52**, 2536-2544

18.   Rowland, M. M., Schonhoft, J. D., McKibbin, P. L., David, S. S., and Stivers, J. T. (2014) Microscopic mechanism of DNA damage searching by hOGG1. *Nucleic acids research*

19.   Studier, F. W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein expression and purification* **41**, 207-234

20.   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

21.   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

22.   Scharer, O. D., Nash, H. M., Jiricny, J., Laval, J., and Verdine, G. L. (1998) Specific binding of a designed pyrrolidine abasic site analog to multiple DNA glycosylases. *The Journal of biological chemistry* **273**, 8592-8597

23.   Hendershot, J. M., Wolfe, A. E., and O'Brien, P. J. (2011) Substitution of active site tyrosines with tryptophan alters the free energy for nucleotide flipping by human alkyladenine DNA glycosylase. *Biochemistry* **50**, 1864-1874

24.   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

25.   Hedglin, M., Zhang, Y., and O'Brien, P. J. (2013) Isolating contributions from intersegmental transfer to DNA searching by alkyladenine DNA glycosylase. *The Journal of biological chemistry* **288**, 24550-24559

26.   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. *Proceedings of the National Academy of Sciences of the United States of America* **79**, 4010-4014

27. Riggs, A. D., Bourgeois, S., and Cohn, M. (1970) The lac repressor-operator interaction. 3. Kinetic studies. *Journal of molecular biology* **53**, 401-417

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. *The EMBO journal* **19**, 6546-6557

29. Surby, M. A., and Reich, N. O. (1996) Facilitated diffusion of the EcoRI DNA methyltransferase is described by a novel mechanism. *Biochemistry* **35**, 2209-2217

30. Elf, J., Li, G. W., and Xie, X. S. (2007) Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* **316**, 1191-1194

31. 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. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 15883-15888

32. Terry, B. J., Jack, W. E., and Modrich, P. (1985) Facilitated diffusion during catalysis by EcoRI endonuclease. Nonspecific interactions in EcoRI catalysis. *The Journal of biological chemistry* **260**, 13130-13137

33. 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. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 13573-13578

34. Miao, F., Bouziane, M., and O'Connor, T. R. (1998) Interaction of the recombinant human methylpurine-DNA glycosylase (MPG protein) with oligodeoxyribonucleotides containing either hypoxanthine or abasic sites. *Nucleic acids research* **26**, 4034-4041

35. 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**, 580-589

36. Baldwin, M. R., and O'Brien, P. J. (2012) Defining the functional footprint for recognition and repair of deaminated DNA. *Nucleic acids research* **40**, 11638-11647

37. Andreatta, D., Sen, S., Perez Lustres, J. L., Kovalenko, S. A., Ernsting, N. P., Murphy, C. J., Coleman, R. S., and Berg, M. A. (2006) Ultrafast dynamics in DNA: "fraying" at the end of the helix. *Journal of the American Chemical Society* **128**, 6885-6892

38.	Holbrook, S. R., and Kim, S. H. (1984) Local mobility of nucleic acids as determined from crystallographic data. I. RNA and B form DNA. *Journal of molecular biology* **173**, 361-388

39.	Leroy, J. L., Kochoyan, M., Huynh-Dinh, T., and Gueron, M. (1988) Characterization of base-pair opening in deoxynucleotide duplexes using catalyzed exchange of the imino proton. *Journal of molecular biology* **200**, 223-238

40.	Nonin, S., Leroy, J. L., and Gueron, M. (1995) Terminal base pairs of oligodeoxynucleotides: imino proton exchange and fraying. *Biochemistry* **34**, 10652-10659

41.	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. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 13087-13092

42.	Singer, B., Abbott, L. G., and Spengler, S. J. (1984) Assessment of mutagenic efficiency of two carcinogen-modified nucleosides, 1,N6-ethenodeoxyadenosine and O4-methyldeoxythymidine, using polymerases of varying fidelity. *Carcinogenesis* **5**, 1165-1171

43.	Bjoras, M., Seeberg, E., Luna, L., Pearl, L. H., and Barrett, T. E. (2002) Reciprocal "flipping" underlies substrate recognition and catalytic activation by the human 8-oxo-guanine DNA glycosylase. *Journal of molecular biology* **317**, 171-177

44.	Parikh, S. S., Walcher, G., Jones, G. D., Slupphaug, G., Krokan, H. E., Blackburn, G. M., and Tainer, J. A. (2000) Uracil-DNA glycosylase-DNA substrate and product structures: conformational strain promotes catalytic efficiency by coupled stereoelectronic effects. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 5083-5088

45.	Kavli, B., Slupphaug, G., Mol, C. D., Arvai, A. S., Peterson, S. B., Tainer, J. A., and Krokan, H. E. (1996) Excision of cytosine and thymine from DNA by mutants of human uracil-DNA glycosylase. *The EMBO journal* **15**, 3442-3447

Appendix B

Additional information to support Chapter 3

Evaluation of the chase species in the pulse-chase control assays

Control experiments were performed by pre-incubating the chase species with the enzyme for 20 sec before adding the substrate. Both the 25εA•T and 25Y•T chase species were tested. When the 25εA•T chase was pre-incubated with the enzyme first, a lag phase was observed after substrate addition, followed by a steady-state formation of the labeled products (Figure B-1A). The lag phase is present because AAG needs to react with the unlabeled DNA and dissociate from the abasic product before it can bind and react with the labeled DNA substrate. This phase shortens as the salt concentration is increased due to faster release of the protein from the abasic product. When the 25Y•T chase was pre-incubated with the enzyme first, however, a small but discernible burst phase was seen at low salt, followed by a very slow steady-state formation of the labeled products (Figure B-1B). This burst phase disappeared when the salt concentration was increased due to the change of the rate-limiting step from product release to chemical catalysis at high salt (1). The observed rates at low and high salt are consistent with ~1 nM free enzyme in the reaction system. Therefore, the calculated excision efficiency values may be overestimated with an upper limit of 0.02 (1 nM AAG, 50 nM substrate DNA) at 1 µM 25Y•T chase at low salt.

The effectiveness of 25Y•T as a chase was also tested using 10 nM 25εA•T substrate DNA, 20 nM AAG and 1 µM chase in the standard pulse-chase assay at 150 mM Na^+ (Figure B-2), since the excision efficiency shows the most significant chase concentration dependence at this salt concentration. In the new assay system, both the enzyme (20 nM versus 100 nM) and substrate (10 nM versus 50 nM) concentrations decreased by 5-fold compared to the original pulse-chase assay, and therefore the ratio of chase to enzyme and chase to substrate increased by 5-fold fold. If 1 µM chase is enough to trap all the dissociate enzyme in the original pulse-chase assay, this new reaction should give the same excision efficiency value as measured in Figure 3-

4. If 1 µM chase is not sufficient for the original assay, then decreased excision efficiency will be observed in this new assay. The two conditions gave indistinguishable burst amplitudes, validating that 1 µM chase is a sufficiently high concentration to trap all dissociated AAG.

A

B
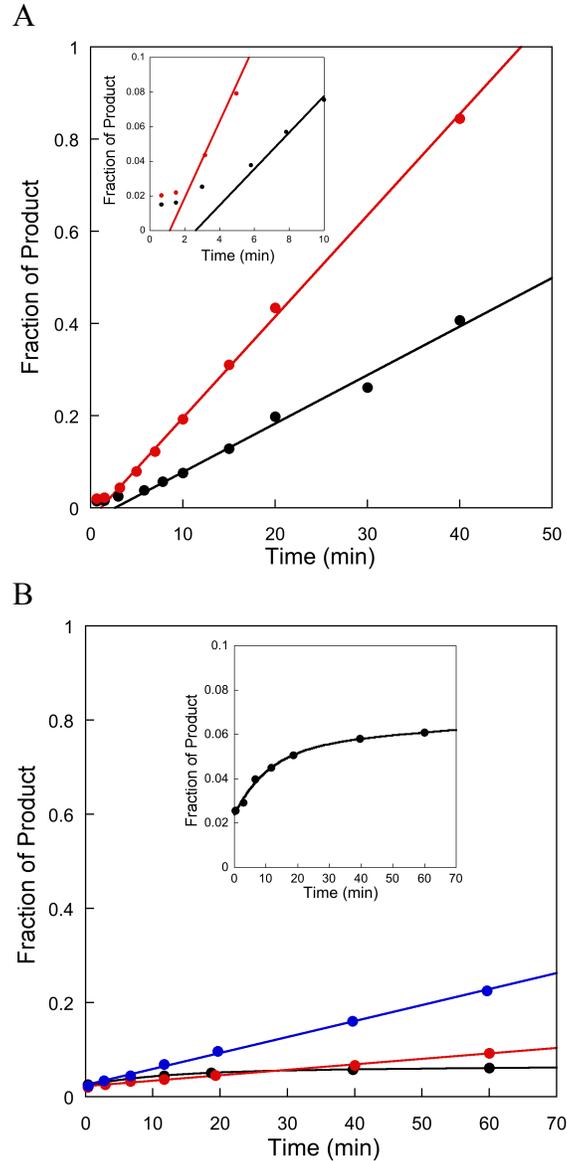
**Figure B-1. Pulse-chase control reactions to test the efficiency of the chase species.**
1 µM chase was incubated with 100 nM AAG for 20 sec, and then 50 nM labeled 25εA•T substrate DNA was added to initiate reaction. A. Pulse-chase control assays using unlabeled 25εA•T chase DNA. The reactions were performed at 100 mM Na^+ (black symbol) and 200 mM Na^+ (red symbol). The inset shows the same reactions with a different scale for the first 10 min, and lag phases can be seen at the initial time points deviating from the linear fits (linear fits were performed without the first three time points). A product fraction of 0.02 is the baseline level of product formation for sample preparation in the absence of the enzyme. B. Pulse-chase control assays using unlabeled 25Y•T chase DNA. The reactions were performed at 100 mM Na^+ (black symbol), 200 mM Na^+ (red symbol) and 300 mM Na^+ (blue symbol). The inset shows the reaction at 100 mM Na^+ with a different Y-axis scale. A burst amplitude of ~0.03 at an observed rate constant of 0.15 min^{-1} is seen at 100 mM Na^+, while no burst phases are seen at 200 and 300 Na^+.
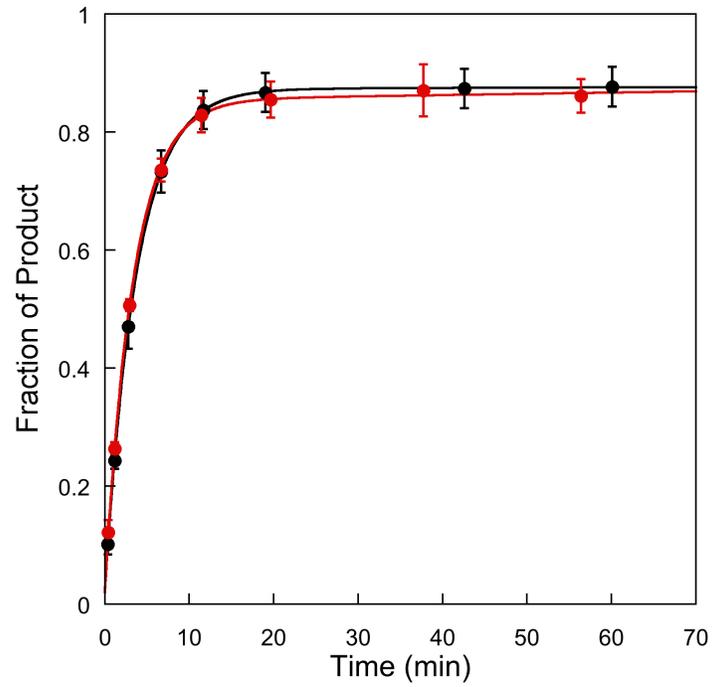
Figure B-2. Pulse-chase assay to ascertain the effectiveness of the 25Y•T chase.
Pulse-chase assays were performed using 1 µM chase at 150 mM Na^+ with 10 nM substrate and 20 nM enzyme (red symbols) or 50 nM substrate and 100 nM enzyme (black symbols). The two reactions gave excision of efficiency results identical within error. The reactions were performed in triplicate. The average values for each data point are plotted and the error bars indicate the standard deviation.
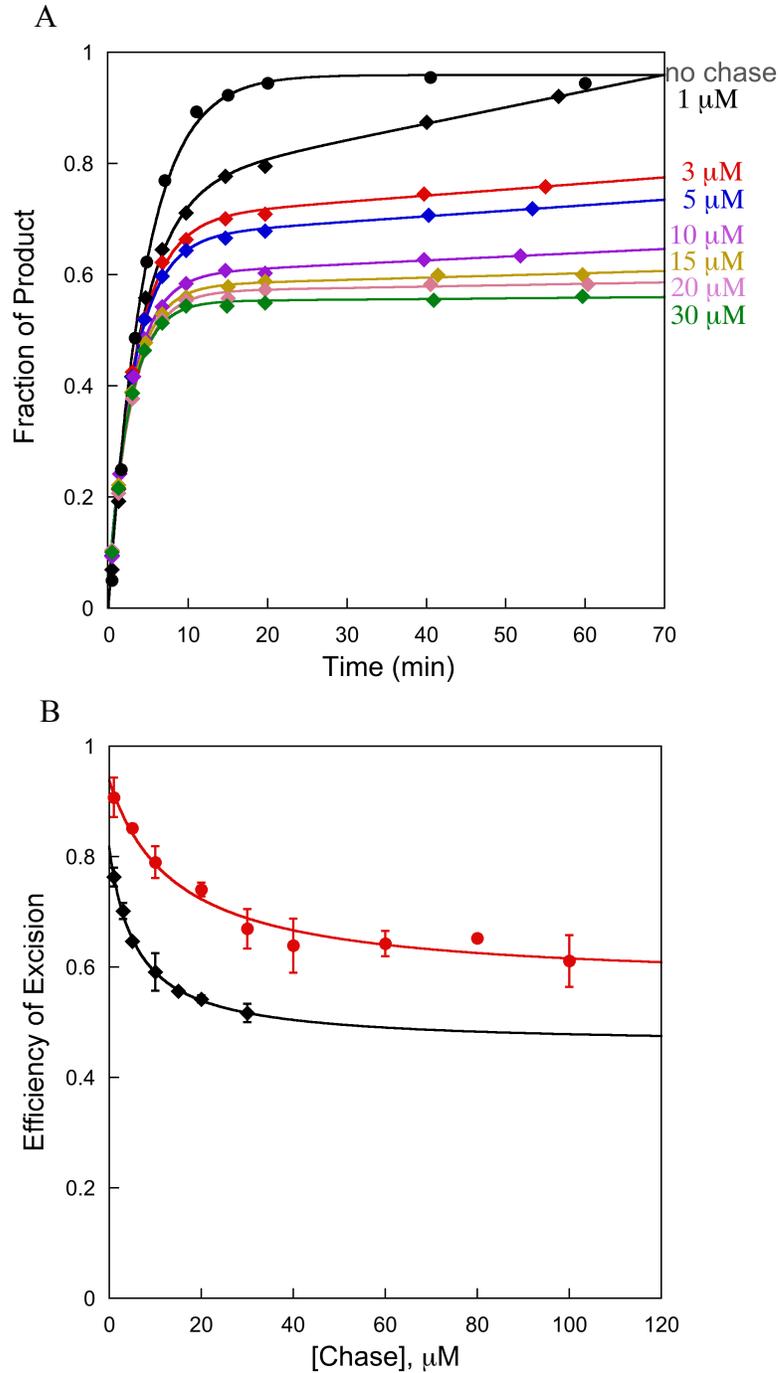
Figure B-3. Efficiency of excision on 17εA•T DNA is chase concentration dependent.
A. Standard pulse-chase assays were performed using 50 nM 17εA•T substrate, 100 nM AAG, 1-30 μM unlabeled 25εA•T chase DNA at 150 mM Na[+]. B. The chase concentration dependence of AAG on the 17εA•T substrate (black diamond) is steeper than that on the 25εA•T substrate (red circle; from Figure 3-6 at 150 mM Na[+]) and lower chase concentrations are required to achieve the plateaued value. The 25εA•T chase was used for the assay on the 17εA•T substrate, and the 25Y•T chase was used for the assay on the 25εA•T substrate. Values shown were mean ± S.D., n = 2 for the 17εA•T substrate, and n = 3 for the 25εA•T substrate.
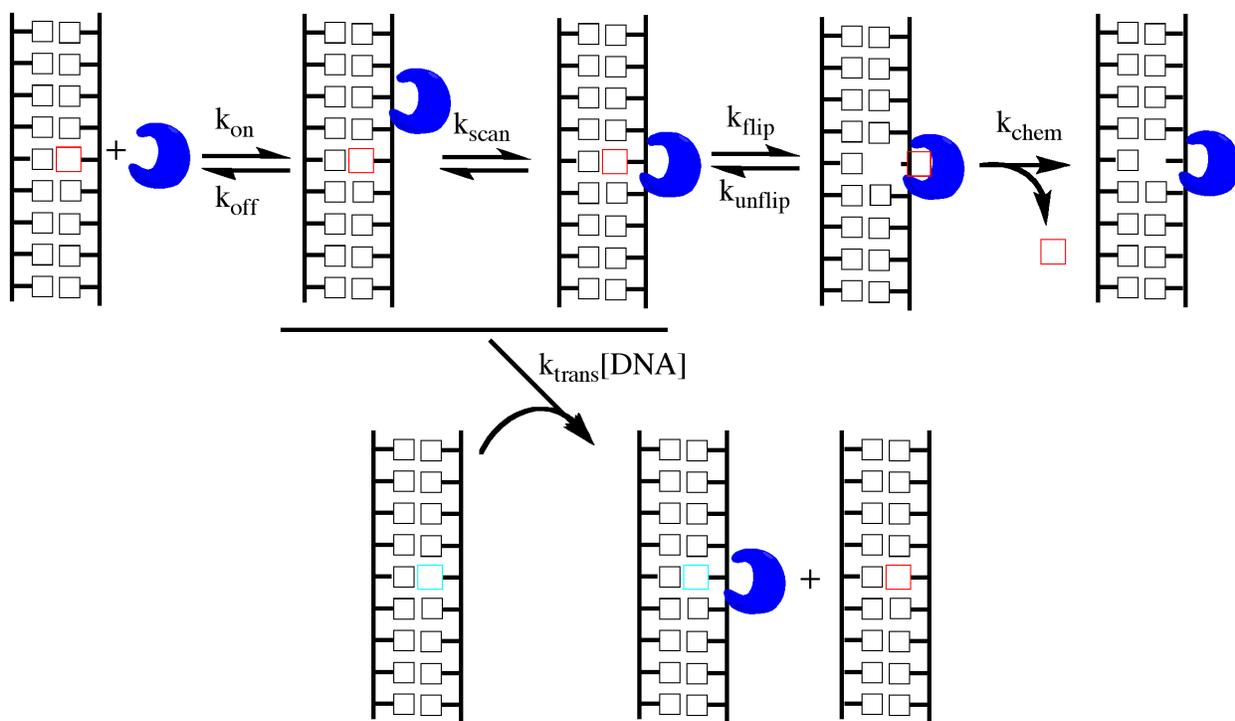
Figure B-4. Proposed mechanism for the chase concentration dependence of excision efficiency.
A. Minimal kinetic mechanism for the AAG-catalyzed reaction in the presence of high concentrations of chase. We propose that AAG is capable of intersegmental transfer from certain bound complex(es) to the excess chase DNA on the pathway to catalysis. The blue crescent denotes AAG, the red box is the lesion base of the substrate DNA, and the cyan box presents the pyrrolidine site of the chase DNA. This scheme is adapted from Ref (2).
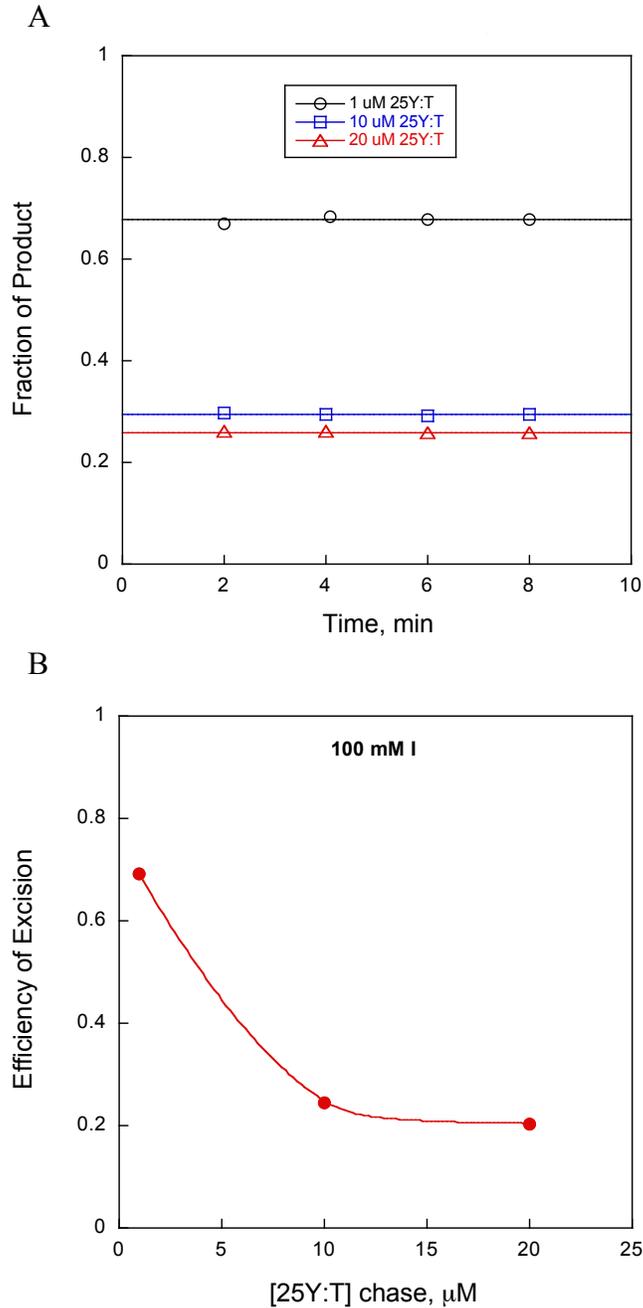
Figure B-5. Excision efficiency measurement of AAG on the Hx substrate.
A. Pulse-chase assays were performed using 1-20 µM chase at 100 mM Na$^+$ with 50 nM 25Hx•T substrate and 200 nM enzyme. The reactions were started at pH 8.0 to slow down the enzyme-substrate DNA reaction during pre-incubation (20 sec) (3). The final reaction condition after chase addition was at pH 6.1, and only the steady-state time points were taken. B. The excision efficiency of AAG on the Hx substrate is chase concentration dependent. The results are taken from single measurements.

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., Zhang, Y., and O'Brien, P. J. (2013) Isolating contributions from intersegmental transfer to DNA searching by alkyladenine DNA glycosylase. *The Journal of biological chemistry* **288**, 24550-24559

3.   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 4

Exploring the Impact of Facilitated Diffusion on the Measurement of Catalytic Specificity of Human Alkyladenine DNA Glycosylase

Human alkyladenine DNA glycosylase (AAG) is a monofunctional DNA glycosylase in the base excision repair pathway responsible for the recognition and excision of alkylated and deaminated purines, such as 1, N^6-ethenoadenine (εA) and hypoxanthine (Hx) (1-6). Like many other DNA glycosylases with broad substrate specificity, AAG recognizes its different cognate lesions with varying affinity and catalyzes their removal at different rates (7). AAG has been shown to use facilitated diffusion to search the genome, which enables the protein to find its target lesion sites more efficiently (8,9).

εA is one of the exocyclic DNA adducts produced by chloroacetalaldehyde, which is a reactive metabolite of an environmental carcinogen vinyl chloride (10-12). This base lesion has also been detected in DNA of the liver and other organs of humans and other animals (13,14), suggesting its formation from endogenous sources. Indeed, it was shown to occur during lipid peroxidation by the interaction of the resulting reactive aldehydes and hydroxy-alkenals with DNA (15-17). Even though εA is removed from DNA by AAG in human cells, it is suspected to have a significant impact on carcinogenesis because of its miscoding capability. εA does not form stable Waston-Crick base pairs with any of the four normal bases since the etheno group eliminates the hydrogen donor at N^6 position of the lesion base. In principle, misincorporation of all nucleotides opposing εA is possible if it is replicated prior to being repaired. Although only marginally mutagenic in *E. coli (18)*, εA has great mutagenic potential in cultured mammalian cells (19,20). For instance, εA was shown to primarily undergo εA to G transition in simian kidney cells, indicating that the insertion of C opposite the εA adduct is predominant (20).

In contrast, Hx is a relatively small lesion that is formed as a result of oxidative deamination of adenine and may occur spontaneously by hydrolysis (21,22). The Hx lesion alters base pairing and converts a Watson-Crick base pair with T to a wobble base pair, and it can also

stably pair with C. AAG is very adept at excising Hx from deoxyinosine-containing DNA[1], and the rate constant for N-glycosidic bond cleavage is ~50-fold faster than the εA substrate (7). However, AAG has weaker binding affinity to the Hx•T site than to the εA•T site (7), and a previously study has established that other mismatches to the Hx lesion were recognized less efficiently than the Hx•T mismatch (23).

In this study, we compared the catalytic specificity of AAG on the εA lesion opposing different bases as well as that on the εA•T and Hx•T lesion substrates. We found that the catalytic specificity of AAG on its different lesion sites is dependent on the assay condition. At low salt conditions all lesions were excised with similar efficiency, while at high salt AAG excises εA in the order of εA•C ≈ εA•T > εA•A > εA•G, which is reversely correlated to the stability of the corresponding duplex DNA. AAG has little discrimination for the εA•T and Hx•T site at low salt, while the preference for the εA•T site to the Hx•T site steadily increased to approximately 100-fold at high salt. The absence of discrimination for different target sites at low salt can be attributed to efficient facilitated diffusion and irreversible DNA binding of AAG on its substrate DNA. Similar preference as observed at high salt was revealed when pre-steady state analysis was performed on hybrid DNA substrates containing both εA•T and Hx•T sites. Finally, we showed that competitor DNA molecules have a similar effect as salt in accentuating the ability of AAG to discriminate between different lesion sites.

MATERIALS AND METHODS

Proteins

Full-length human AAG was cloned into the pET SUMO expression vector (Invitrogen) between the N-terminal BamHI and C-terminal XhoI site, encoding an N-terminal 6× His-Smt3 fusion protein that can be cleaved with the SUMO protease ULP1. The AAG fusion protein was expressed in *E. coli* BL21 Star (DE3) pRare2 cells using autoinduction media (24). Typically, cells were grown at 37 °C until they reached an optical density of 1.0 at 600 nm. Cultures were then transferred to a 16 °C shaker, and protein was expressed for 24 hours with final OD_{600} values at ~20 before cells were harvested and frozen at -80 °C. Cell pellets were suspended in

---

[1] Inosine (I) is a nucleoside that contains hypoxanthine (Hx) 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 deoxyinosine. Upon excision and release from DNA by AAG, the free base is referred to as hypoxanthine.

lysis buffer (25 mM potassium phosphate, pH 7.0, 250 mM NaCl, 10% (v/v) Glycerol, 0.1% NP-40) and disrupted by a high pressure homogenizer (EmulsiFlex, Avestin). Cell lysates were slowly precipitated with polyethylenimine to remove nucleic acids, and the recombinant AAG protein was purified by metal affinity chromatography using a His tag that was subsequently cleaved with the SUMO protease ULP1 to produce full-length AAG with an N-terminal serine in place of the initiator methionine of the native protein. Subsequent ion exchange chromatography (HiTrap SP, GE Healthcare) and dialysis (MWCO 12,000-14,000, Fisherbrand) yielded recombinant protein that was homogeneous as judged by SDS-PAGE with Coomassie staining. The active concentration of AAG was determined by burst analysis as described previously (8), and the enzyme concentration used in the experiments described here refers to the corrected active concentration.

Oligonucleotides

DNA substrates were synthesized by Integrated DNA Technologies or the Keck Center at Yale University and purified by denaturing PAGE as previous described (8). Oligo concentrations were determined from the absorbance at 260 nm using the calculated extinction coefficients. For 5' fluorescein-labeled oligonucleotides we assessed the labeling efficiency by comparing the absorbance at 260 nm with that at 495 nm. DNA duplexes were annealed at 1:1.2 ratio of the lesion containing fluorescein-labeled strand and the complement. The sequences for all the oligonucleotides used in this study are shown in Appendix Figure C-1.

Gel-Based Glycosylase Assay

Reactions were carried out at 37 °C in a buffer system consisted of 50 mM NaMES pH 6.1, 10% (v/v) glycerol, 0.1 mg/mL BSA, 1 mM DTT, 1 mM EDTA and varying concentrations of NaCl to obtain the desired ionic strength. Reactions were initiated by the addition of enzyme to a final reaction volume of 20-60 μL. Aliquots were withdrawn at various times and quenched with NaOH (0.2 M final concentration). Samples were heated at 70 °C for 15 min, loading buffer consisting of 10 mM EDTA and 98% formamide was added, and the DNA species were resolved on 14% (w/v) denaturing polyacrylamide gels with 8 M urea. Gels were scanned with a Typhoon Trio+ fluorescence imager (GE Healthcare), and fluorescein signal was detected using excitation wavelength at 532 nm and emission with a 520BP40 filter. The resulting fluorescence signal was quantified with ImageQuant TL and corrected for background signal. The intensity of each DNA

band was converted into a fraction of the total DNA by dividing its intensity by the sum of the intensities for all of the DNA species in the reaction.

Single-turnover Assay

Single-turnover kinetics was performed on 25εA•T, 25εA•C, 25εA•A and 25εA•G DNA substrates. The concentration of AAG (100 and 500 nM) was varied in excess over the DNA substrates (50 nM) to ensure single-turnover conditions. In all cases the reactions followed single exponential kinetics F = A(1-exp(-kt)), in which A is the reaction amplitude and k is the observed single-turnover rate constant.

Catalytic Specificity (k_{cat}/K_M) Measurement

k_{cat}/K_M was measured on 25εA•T, 25εA•C, 25εA•A and 25εA•G DNA substrates. Different salt ranges were used for different substrates: 600-1000 mM Na^+ for 25εA•T and 25εA•C, 400-800 mM Na^+ for 25εA•A, and 300-700 mM Na^+ for 25εA•G. Multiple-turnover reactions were performed with DNA substrate concentrations at 25-100 nM, and fixed enzyme concentration at 2 nM. The k_{cat}/K_M values were calculated from the slopes of the linear fits or the Michaelis-Menten equation V_{init}/[Enz]=(k_{cat}/K_M)*[S]/(1+[S]/K_M). The k_{cat}/K_M versus salt concentration data were fit by assuming a log-linear relationship $\log(k_{cat}/K_M) = a + b*\log([Na^+])$, according to the counterion condensation theory (25,26).

Determination of Relative k_{cat}/K_M Values

The direct competition assay was performed under multiple-turnover conditions with equal concentration of the 19εA•T reference DNA and the competing DNA oligos 25εA•T, 25εA•C, 25εA•A or 25εA•G (200 nM each), and 2 nM AAG over a wide range of salt concentrations. All substrates and products were resolved by denaturing PAGE, and the relative k_{cat}/K_M values are given by the ratio of the concentration and initial rates of the competing DNA oligos: $(k_{cat}/K_M)_{rel} = ([A]*(k_{cat}/K_M)_A)/ ([B]*(k_{cat}/K_M)_B)$, in which A and B are the two competing DNA substrates. Direct competition assays of 19εA•T and 25εA•A were also performed in the presence of 0-20 μM 25A•T nonspecific DNA at 150 mM Na^+ in order to test the effect of excess nonspecific DNA on catalytic specificity.

The relative k_{cat}/K_M values of εA•T and Hx•T were calculated from direct competition of 25εA•T versus 19Hx•T under the same multiple-turnover conditions as described above. The assay was also conducted using the 25εA•A DNA as the common bridging substrate for the

competition of 19εA•T vs 19Hx•T. The relative k_{cat}/K_M of εA•T vs Hx•T was given by the ratio of the two direct competition results $(k_{cat}/K_M)_{εA•T/Hx•T} = (k_{cat}/K_M)_{εA•T/εA•A}/ (k_{cat}/K_M)_{Hx•T/εA•A}$.

Steady State Analysis on Hybrid Substrate

Steady state analysis on the hybrid substrate containing both an εA lesion and a Hx lesion was performed using 200 nM DNA and 2 nM AAG at salt concentrations ranging from 50 mM to 300 mM. The initial rates were determined from the first 10% of the reaction and were linear in all cases. The intermediates refer to those species retaining an εA or a Hx lesion site and the products refer to those retaining no lesion sites.

**Pre-steady State Burst Analysis on Hybrid Substrate**

Under low salt conditions, AAG catalyzes a rapid burst of lesion excision, followed by slow rate-limiting product release. To minimize the contribution of two proteins bound to the same DNA, pre-steady state burst analysis was performed containing 1 μM 47mer hybrid DNA substrate containing both an εA lesion and a Hx lesion and 100 nM FL AAG at 50 mM Na^+. Time points were taken spanning both the initial burst phase and the steady state phase. The concentration of product remaining was fit by an exponential and a steady state linear phase. The intermediate was fit by the theoretical equation as derived in Ref (27).

RESULTS

Catalytic Specificity Measurement on εA Opposing Different Bases

In an effort to evaluate the effect of the opposing bases to εA lesion recognition and catalytic activity, we determined the catalytic specificity (k_{cat}/K_M) of AAG on the 25mer εA-containing DNA with A, T, G or C as the opposing bases. Due to the extremely high binding affinity of AAG towards its substrate DNA, k_{cat}/K_M values could not be determined at low salt concentrations. Plots of the initial rate constants of product formation as a function of DNA concentration are shown in Figure 4-1A, and the reactions on the 25εA•T substrate were used as an example. Reaction curves were fit by the Michaelis-Menten equation, unless they can be fit by linear regression. The measured k_{cat}/K_M values are plotted as a function of salt concentration in Figure 4-1B. All the k_{cat}/K_M values measured here showed similar salt concentration dependence, and all the linear fits (log-log scale) were approximately parallel to each other. These results demonstrate that AAG had the highest catalytic specificity on εA opposing C or T

under all the salt concentrations tested here. Placing A opposite εA reduced k_{cat}/K_M to some extent, while G opposite εA decreased the value even further. These large differences are not due to potential changes in the rate of N-glycosidic bond cleavage, because single-turnover experiments showed that k_{chem} for N-glycosidic bond cleavage of the εA lesion was minimally affected by the opposing bases (Appendix Figure C-1).

A   B
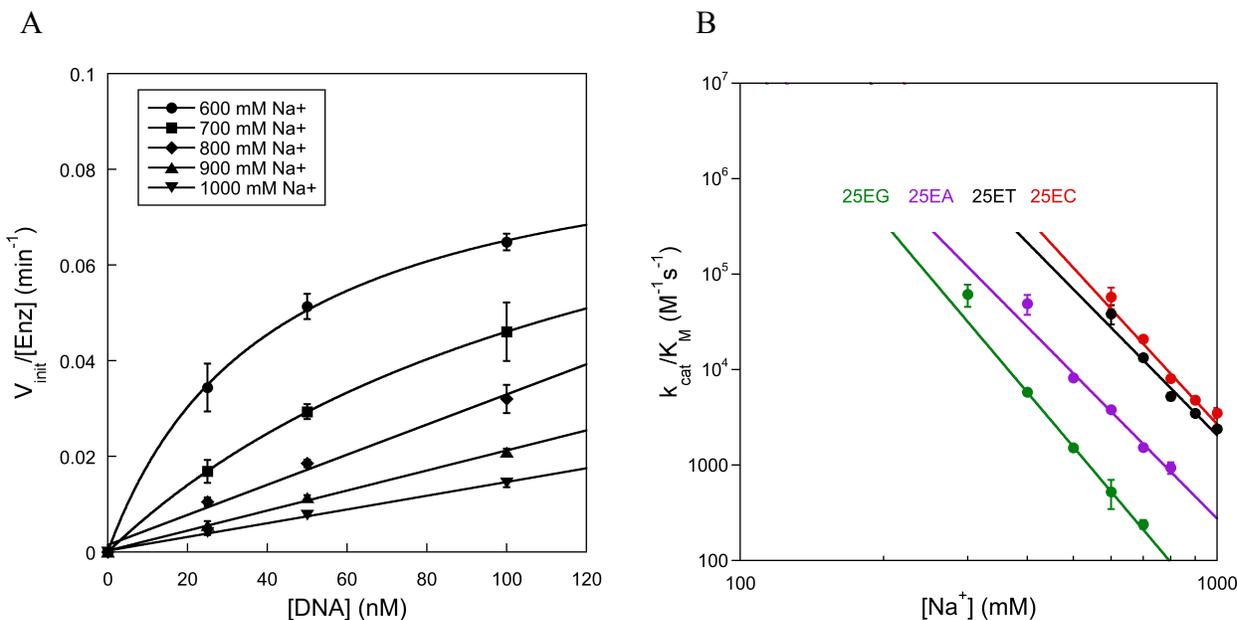
Figure 4-1. Catalytic specificity k_{cat}/K_M of AAG is dependent on salt concentration.
A. A representative k_{cat}/K_M measurement plot of AAG on 25εA•T substrate at 600-1000 mM Na$^+$. The k_{cat}/K_M values were taken from the Michaelis-Menten fits or from the slopes of the linear fits. B. Catalytic specificity of AAG on its different substrates was measured under high salt conditions, and the continuous line represents a linear extrapolation assuming a linear relationship between $\log(k_{cat}/K_M)$ and $\log[Na^+]$, according to the counterion condensation theory (25,26). At low salt conditions, the k_{cat}/K_M values can deviate from those predicted from the linear extrapolation as they reflect the association rate for binding of AAG to DNA. The lines are from a weighted linear regression analysis using the relative errors in the k_{cat}/K_M values. Values shown are mean ± S.D. (n=3).

Catalytic specificity of AAG on εA opposing different bases can also be determined in direct competition assays (28), which were conducted over a wide range of salt concentrations. At low salt, AAG binding is essentially irreversible and the protein uses facilitated diffusion to search the target lesion site on a DNA molecule, and therefore the k_{cat}/K_M values simply reflect the association rate constant. At high salt, AAG binding is reversible and k_{cat}/K_M reflects both the binding and catalysis steps (27). Consistent with these theoretical analyses, the relative k_{cat}/K_M values were close to 1 for all the competing substrate pairs at low salt, while they showed different changes at high salt (Figure 4-2).

We found that the relative k_{cat}/K_M values reached different plateaus at high salt

81

concentrations for different competing substrate pairs. To test if and how the length differences of the DNA molecules affect the competition results, we competed 19εA•T DNA against
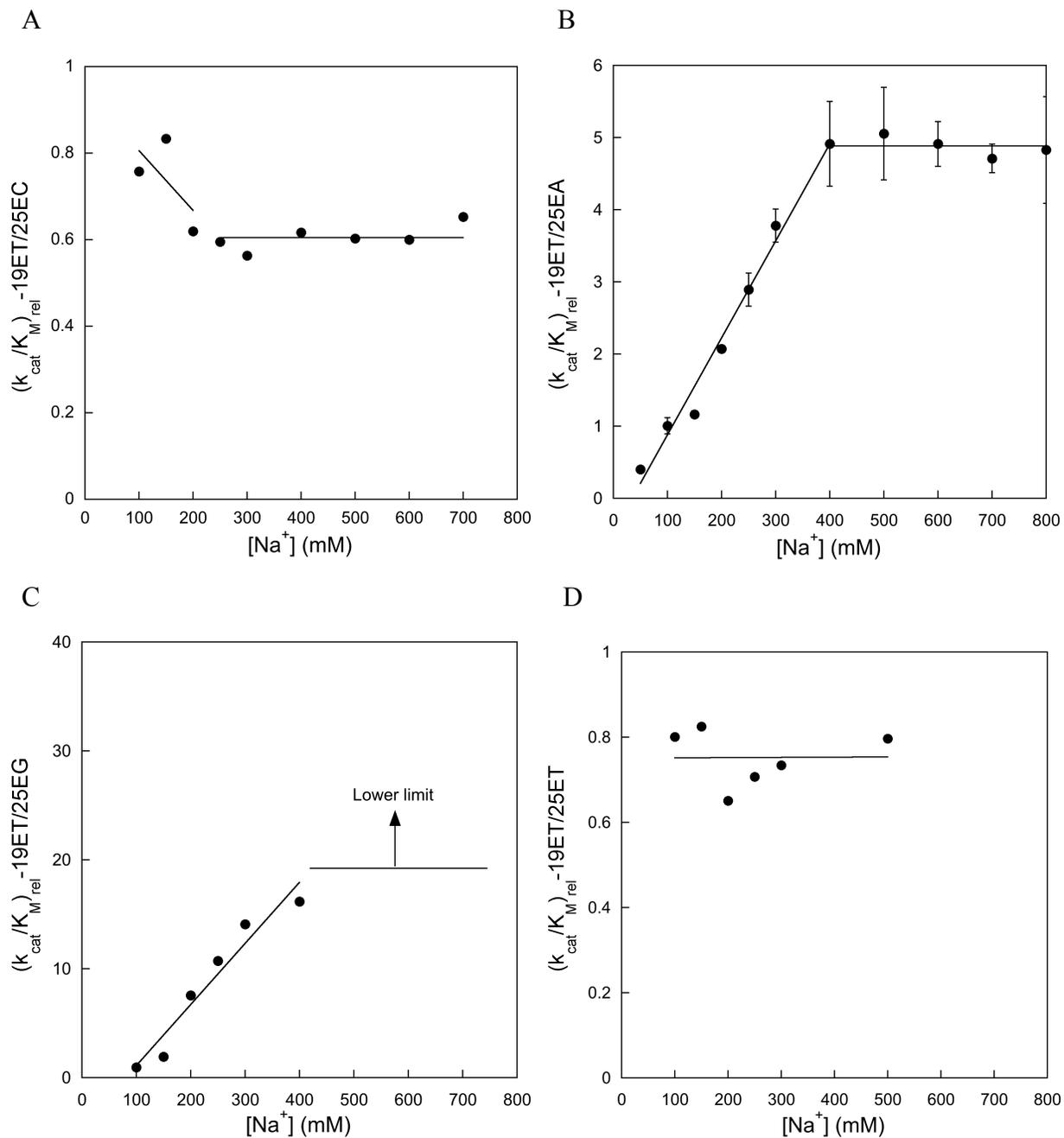
Figure 4-2. Relative catalytic specificity of AAG on εA is dependent on opposing bases.
Direct competition measurements were performed with 2 nM AAG, 200 nM 19εA•T reference DNA and 200 nM 25εA•C (A), 25εA•A (B), 25εA•G (C) or 25εA•T (D) DNA at a variety of salt concentrations. Curves fits were shown to guide the eye for changes of $(k_{cat}/K_M)_{rel}$ as a function of salt concentration. Linear regression was used for data points spanning (A) 100-200 mM Na$^+$ (B) 50-400 mM Na$^+$ and (C) 100-400 mM Na$^+$. The rest of the data segments were fit with flat lines. In panel B average results and standard deviation of at least three independent measurements are shown, whereas panel A, C and D reflect single measurements.

25εA•T DNA. Results showed that AAG has slightly higher k_{cat}/K_M on the 25mer substrate, and this preference seems to be independent of the salt concentration (Figure 4-2D). Competition of 19εA•T with 25εA•C demonstrated that εA•C is marginally preferred over εA•T (Figure 4-2A). In contrast, 19εA•T is preferred over 25εA•A by a factor of approximately 5-fold (Figure 4-2B). Most significantly, 19εA•T DNA is overwhelmingly preferred over 25εA•G. The exact fold of preference at high salt concentrations above 400 mM Na$^+$ could not be determined accurately, but a conservative estimation put a lower limit of 20-fold (Figure 4-2C). The relative k_{cat}/K_M values at high salt are in good agreement with those calculated from the individually measured k_{cat}/K_M values in Figure 4-1.

Activity Comparison of AAG on εA and Hx Lesions

εA and Hx are two of the base lesions caused by damage to the adenine base (16,21), and they are both opposite with T in the natural context. Kinetic parameters have been measured on both substrates previously, revealing different strategies of lesion recognition and catalysis: tight binding and slow catalysis for εA (7,29) versus weak binding and fast catalysis for Hx (7,30). Despite of these extensive studies, it is still under dispute whether AAG prefers one lesion to the other or whether they are equally good. In this study, we use direct competition assays to compare the relative catalytic specificity of AAG on the two lesion sites.

Direct competition of 25εA•T with 19Hx•T revealed a big preference for εA•T, while the exact fold of preference at high salt could not be determined accurately (Appendix Figure C-2). Therefore, we used 25εA•A DNA as the bridging substrate to compete with 19εA•T (Figure 4-2B) and 19Hx•T (Figure 4-3A) DNA individually and obtained the relative catalytic specificity of 19εA•T and 19Hx•T indirectly (Figure 4-3B). At low salt AAG has similar k_{cat}/K_M on both lesion sites, whereas at high salt εA is preferred over Hx by a factor of around 100-fold.

Multiple Turnover Kinetics on Hybrid Substrates

All the above measurements have the competing lesion sites on different DNA molecules, and we next investigated the preference for recognition and catalysis when the εA and Hx lesions are located on the same piece of DNA. The two lesion sites are located 25 bp apart on the 47mer hybrid substrate and the reaction scheme is illustrated in Figure 4-4A. Both lesion orientations (47EI and 47IE) were tested in case of position-specific biases, but for simplicity only one orientation (47EI) is shown in the scheme. Under multiple turnover

conditions, the concentration of the DNA substrate is in excess of AAG, and the reaction is initiated by first excising one lesion site through pathway 1 or 2. The removal of the first lesion site releases the corresponding product species and an accompanying intermediate species that can be further processed to product by removing the remaining lesion through pathway 3 and 5 under processive conditions at low salt, or it can be released as an intermediate into the bulk solution through pathway 4 and 6 (Figure 4-4). Formation of each intermediate species is through one pathway, while generation of each product species can be achieved by two different pathways (Figure 4-4). The preference for the lesion sites can therefore be determined from production of these different species.

A

B
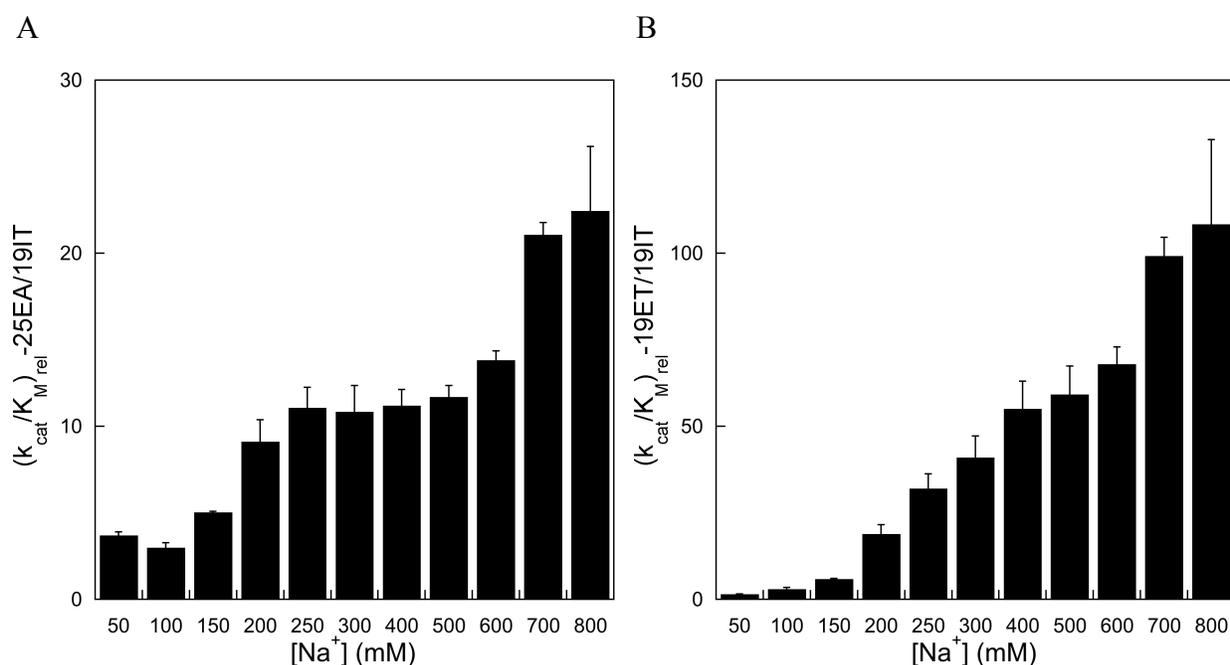
Figure 4-3. εA•T is preferred over Hx•T at high salt.
A. Direct competition measurements were performed with 2 nM AAG, 200 nM 25εA•A and 200 nM 19I•T DNA at a variety of salt concentrations. B. Relative k_{cat}/K_M of 19εA•T and 19I•T DNA calculated indirectly from panel A of this figure and Figure 4-2B as described in Materials and Methods.

Under high salt conditions (300 mM Na$^+$), searching is distributive (8,27) and the preference for the two lesion sites can be determined from the formation of the respective product species. Initial velocity of Prod-E was faster than Prod-I for around 15-fold and 30-fold for the EI (Figure 4-4B and C) and IE (data not shown) substrate respectively, indicating that AAG greatly favors the εA site to the Hx site. It has previously been demonstrated that the εA site is susceptible to a ring-opening reaction, which produces a lesion that can no longer be

recognized by AAG (31). Therefore, this measured preference for εA over Hx presents a lower limit for the difference, since some of the Prod-I species will be generated in the absence of the competing εA site on the same DNA molecule. This result is roughly consistent with the direct competition assay using two different DNA molecules (Figure 4-3B).
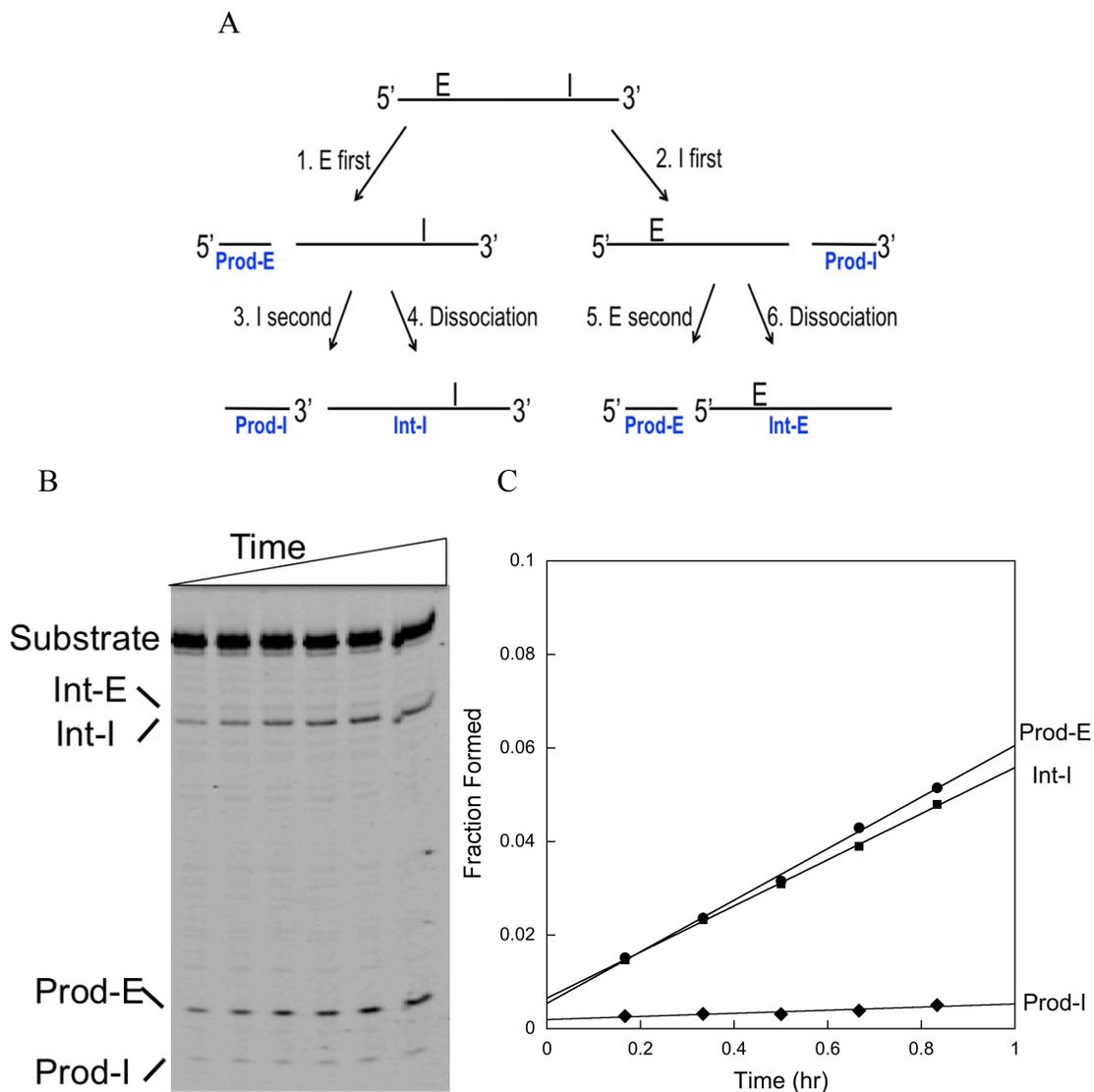
Figure 4-4. Multiple-turnover assay on the hybrid substrate.
A. Steady-state reaction scheme on the hybrid substrate. The substrates used here are duplex DNA, and only the lesion-containing strand is shown for clarity. The substrate DNA is labeled at 5' and 3' end with fluorescein. Two substrates differ only in the orientations of the lesion sites were tested, and only one substrate orientation is shown in the scheme for simplicity. The two lesion bases are 1, N^6-ethenoadenine (E) and hypoxanthine/deoxyinosine (I). B. Representative denaturing PAGE of the steady-state assay performed with 200 nM 47EI substrate, 2 nM AAG at 300 mM Na$^+$. C. Reaction progress curves for products and intermediate for the reaction in panel B. The Int-E species was formed at the baseline level and did not show any time dependent changes, and was not plotted here. The slopes of the linear fits for Prod-E and Prod-I formation were at a ratio of 15:1.

When the salt concentration is decreased, facilitated diffusion enables a second excision on the intermediate species, and product can be generated from both primary excision of the substrate and secondary excision of the intermediate. As a result, the preference for a lesion site cannot be established in the steady-state assay at low salt, but can be distinguished using pre-steady state analysis to monitor these steps individually.

Transient Kinetic Analysis of Lesion Preference on Hybrid Substrates

To gain a deeper insight into the preference for different lesion sites at low salt, we performed pre-steady state glycosylase experiments with the hybrid DNA substrates containing both the εA and Hx sites. Under burst conditions, with 10-fold excess of substrate over enzyme, the individual steps can be monitored to allow dissection of the sequence of multiple reactions. For these hybrid substrates, the formation and disappearance of the intermediate with a single lesion (εA or Hx) follow a branched, two-step irreversible pathway, in which the first step is base excision and the second step is either excision at the second lesion site or dissociation (Figure 4-4A). This transient phase is followed by a steady state formation of the intermediate that is revealed via distributive action.
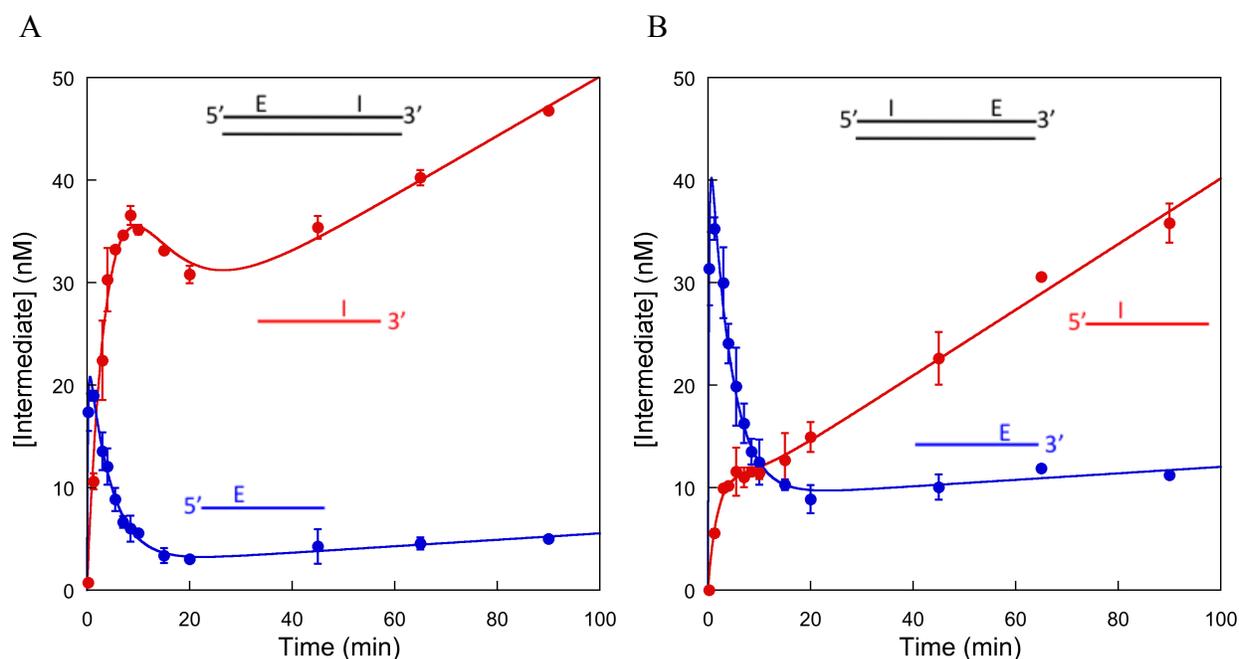
Figure 4-5. Pre-steady state analysis on hybrid substrate.
Burst experiments were performed with a 1:10 ratio of AAG to DNA substrate 47EI (A) or 47IE (B) at 50 mM Na$^+$, and the concentration of each intermediate species was calculated from the fraction of the total fluorescence determined by denaturing PAGE. Reactions were performed in duplicate, and the mean ± S.D. is plotted. Lines indicate the fits according to the irreversible model as derived in reference (27), and the equation derivation is shown in Appendix Scheme C-1. Slightly preference of the 5' site was observed for both substrates.

86

Under these conditions, AAG catalyzed the excision of different lesion sites with similar probability in the first catalytic turnover judging from their similar amplitudes (Figure 4-5). After the burst phase, however, the intermediate species retaining the εA site was quickly converted into product and extremely slow accumulation of the intermediate species was observed in the steady-state phase. In contrast, the intermediate species retaining the Hx site was only partially converted to the corresponding product and intermediate species was formed at a much faster rate in the steady-state phase (Figure 4-5). The Int-I was formed approximately 10-fold faster than the Int-E species from both hybrid substrates, and this preference for the εA site is likely to be a lower limit due to the presence of the damaged εA sites (8). This result suggests that εA and Hx are removed with similar efficiency during initial encounter at low salt, while subsequent excision of the εA lesion is much preferred over excision of the Hx lesion.

Direct Competition Assay with Excess DNA

The above measurements clearly show that AAG distinguishes εA opposite different bases best at high salt concentrations. This salt dependence raises the question as to how the direct competition results obtained at high salt are related to the physiological condition in cells, whose ionic strength is generally accepted to be around 150 mM (32,33). Considering that rare lesion sites are buried in a sea of nonspecific DNA in the genome for the enzyme to locate, we performed direct competition assays with excess nonspecific DNA in the reaction system to
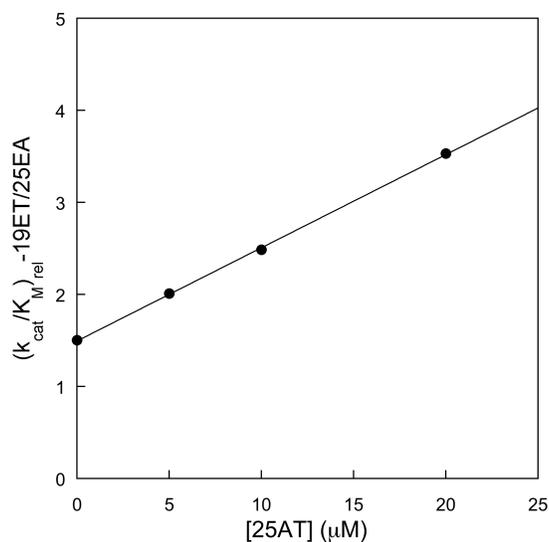
Figure 4-6. Nonspecific DNA has similar effect on relative catalytic specificity to salt.
Direct competition measurements were conducted with 2 nM AAG, 200 nM 19εA•T and 200 nM 25εA•A DNA at 150 mM Na$^+$. Excess 25A•T oligo DNA at 0-20 μM was added into the reaction, and linear increase in $(k_{cat}/K_M)_{rel}$ was observed. The measurement was performed once.

mimic the cellular environment. Competition assays were conducted using 19εA•T and 25εA•A DNA at 150 mM Na$^+$ and results are shown in Figure 4-6. In the absence of nonspecific DNA, AAG shows very little preference for the εA•T site; however, with the addition of nonspecific DNA in the reaction, AAG exhibits a greater preference for the εA•T site. The relative catalytic specificity of 19εA•T versus 25εA•A is linearly dependent on the concentration of the nonspecific DNA up to 20 μM. At 20 μM nonspecific DNA, AAG has a preference for 19εA•T over 25εA•A for ~4-fold, very close to the maximal preference of ~5-fold at high salt concentrations. Therefore, excess DNA has similar effect as salt in greater discrimination of different sites, which may better reflect the reaction situation *in vivo*.

DISCUSSION

AAG catalyzes the excision of a wide range of modified bases, among them hypoxanthine (Hx) and 1,N^6-ethenoadenine (εA). In mammalian cells, excision of Hx and εA by AAG and subsequent processing by downstream BER enzymes is the predominant pathway for the repair of these base lesions (1,2). In this study, we used different kinetic assays to compare the catalytic specificity of AAG on εA opposing different bases and between the εA and the Hx lesion. Direct measurement of k_{cat}/K_M (Figure 4-1) and competition assay of two substrates of εA opposing different bases (Figure 4-2) gave essentially the same results, while the competition assay is less tedious and more reproducible as competing substrates are assayed in the same reactions. We found that the relative activity of AAG on different substrates was dependent on the salt concentration. At low salt, all substrates were removed with similar k_{cat}/K_M values in direct competition assays; at high salt, maximal discrimination in k_{cat}/K_M for different substrates was realized (Figure 4-2). The lack of substrate discrimination at low salt can be ascribed to facilitated diffusion. Facilitated diffusion is so efficient at very low salt that almost every nonspecific encounter with the DNA molecule results in binding of the specific target site (i.e., irreversible binding) and excision of the lesion base. As the salt concentration increases, searching and excision efficiency decrease continually, and the preference for certain lesion sites can start to be revealed. At very high salt concentrations when searching is all distributive, the discrimination for different lesion substrates is maximized.

Although the maximal discrimination for substrates on different DNA molecules was

observed only at high salt concentrations, we showed in a modified direct competition assay that AAG is capable of discriminating between different substrate lesions at physiological salt concentrations when nonspecific DNA is present (Figure 4-6). This change in relative k_{cat}/K_M is consistent with the model that AAG is capable of intersegmental transfer from one DNA molecule to another in a DNA concentration dependent manner. In the human nucleus, the accessible DNA base pair concentration is estimated to be approximately 10 mM, therefore we expect that the *in vivo* discrimination will be similar to the maximal discrimination that was observed at high salt in the *in vitro* assays.

Previous studies show that εA has great mutagenic potential in cultured mammalian cells (19), even more so than 8-oxodeoxyguanosine, which induces G to T transversions in HeLa cells (34). The presence of εA primarily undergoes A to G transition in simian kidney cells (20), and A to T, A to G and A to C mutations in human cells, especially A to T transversions (19). Consistent with this, these mutations were also observed in the rat p53 gene in one tumorigenicity study with vinyl chloride (35). Therefore, kinetic characterization of AAG on the εA lesion opposite to different bases may shed light on the contribution of this repair pathway to the mutagenic profiles. In this study, the catalytic specificity of AAG on εA opposite to different bases was shown to be εA•C ≈ εA•T > εA•A > εA•G, with approximately two orders of magnitude differences between the best and worst substrates (Figure 4-1 and 4-2). The εA lesion formed by spontaneous damage is paired with T in the natural context, and the observed high activity of AAG on the εA•T site suggests that the enzyme is likely to repair these lesion sites to the correct base prior to DNA replication. If εA is not repaired, DNA replication can result in the incorporation of C, G or A to form different εA pairs. Our current study shows that AAG also removes εA from the εA•C pair very efficiently, and this would lead to A to G transition after the base excision repair pathway is completed. The observed lower activity on the εA•A pair by AAG makes it unlikely that AAG would be responsible for the A to T transversions. However, direct repair of εA by the human dioxygenase ABH2 or ABH3 could result in the A to T transversion (36), or it could simply occur after a second round of DNA replication. Therefore, it is possible that AAG is at least partially responsible for the observed mutational pattern of εA in human cells due to preferential excision of some of the mismatched lesion sites.

AAG does not appear to make specific contacts with the base opposite to the cognate lesion site (37), and therefore the relationship between base-pair stability and k_{cat}/K_M for different mismatches can be examined. A previous study using the Hx lesion opposing different bases suggests that the stability of base pairing affects the catalytic specificity on the substrate (23). εA has an exocyclic adduct, which prevents it from forming Watson-Crick base pairs with the opposing base; nonetheless, the εA lesion can still form Hoogsteen base pair with G (38-40). Indeed, a previous thermodynamic characterization demonstrated that the stability of εA pairs shows the trend $G > A > T \approx C$ (29). In this study, we found an inverse correlation between duplex stability and catalytic specificity, which is consistent with previous results on the Hx lesion (23). The k_{cat}/K_M values are smaller when εA is opposing with A and G than with C and T, suggesting that εA is in a more stable context when opposite to purines than with pyrimidines.

Studies have shown that many DNA repair enzymes use facilitated diffusion via electrostatic interactions to search for target sites, therefore these findings made above on AAG may extend to other repair enzymes. This study demonstrates that measurements made under low salt conditions underestimate the differences in the k_{cat}/K_M values of the enzyme on its different substrates. In order to get maximal discrimination for different substrates, assays should be performed using high salt concentrations or including competitor DNA in the assay system. Furthermore, it is shown in this study that direct competition assays can be a valuable tool to quantify catalytic specificity.

References

1.   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. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 13087-13092

2.   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. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 12869-12874

3.   Miao, F., Bouziane, M., and O'Connor, T. R. (1998) Interaction of the recombinant human methylpurine-DNA glycosylase (MPG protein) with oligodeoxyribonucleotides containing either hypoxanthine or abasic sites. *Nucleic acids research* **26**, 4034-4041

4.   O'Connor, T. R. (1993) Purification and characterization of human 3-methyladenine-DNA glycosylase. *Nucleic acids research* **21**, 5561-5569

5.   Saparbaev, M., and Laval, J. (1994) Excision of hypoxanthine from DNA containing dIMP residues by the Escherichia coli, yeast, rat, and human alkylpurine DNA glycosylases. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 5873-5877

6.   Singer, B., Antoccia, A., Basu, A. K., Dosanjh, M. K., Fraenkel-Conrat, H., Gallagher, P. E., Kusmierek, J. T., Qiu, Z. H., and Rydberg, B. (1992) Both purified human 1,N6-ethenoadenine-binding protein and purified human 3-methyladenine-DNA glycosylase act on 1,N6-ethenoadenine and 3-methyladenine. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 9386-9390

7.   O'Brien, P. J., and Ellenberger, T. (2004) Dissecting the broad substrate specificity of human 3-methyladenine-DNA glycosylase. *The Journal of biological chemistry* **279**, 9750-9757

8.   Hedglin, M., and O'Brien, P. J. (2008) Human alkyladenine DNA glycosylase employs a processive search for DNA damage. *Biochemistry* **47**, 11434-11445

9.   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

10.   Jacobsen, J. S., and Humayun, M. Z. (1990) Mechanisms of mutagenesis by the vinyl chloride metabolite chloroacetaldehyde. Effect of gene-targeted in vitro adduction of M13 DNA on DNA template activity in vivo and in vitro. *Biochemistry* **29**, 496-504

11.   Jacobsen, J. S., Perkins, C. P., Callahan, J. T., Sambamurti, K., and Humayun, M. Z. (1989) Mechanisms of mutagenesis by chloroacetaldehyde. *Genetics* **121**, 213-222

12. Kusmierek, J. T., and Singer, B. (1982) Chloroacetaldehyde-treated ribo- and deoxyribopolynucleotides. 1. Reaction products. *Biochemistry* **21**, 5717-5722

13. Nair, J., Barbin, A., Guichard, Y., and Bartsch, H. (1995) 1,N6-ethenodeoxyadenosine and 3,N4-ethenodeoxycytine in liver DNA from humans and untreated rodents detected by immunoaffinity/32P-postlabeling. *Carcinogenesis* **16**, 613-617

14. Nair, J., Gal, A., Tamir, S., Tannenbaum, S. R., Wogan, G. N., and Bartsch, H. (1998) Etheno adducts in spleen DNA of SJL mice stimulated to overproduce nitric oxide. *Carcinogenesis* **19**, 2081-2084

15. Chung, F. L., Chen, H. J., and Nath, R. G. (1996) Lipid peroxidation as a potential endogenous source for the formation of exocyclic DNA adducts. *Carcinogenesis* **17**, 2105-2111

16. el Ghissassi, F., Barbin, A., Nair, J., and Bartsch, H. (1995) Formation of 1,N6-ethenoadenine and 3,N4-ethenocytosine by lipid peroxidation products and nucleic acid bases. *Chemical research in toxicology* **8**, 278-283

17. Gros, L., Ishchenko, A. A., and Saparbaev, M. (2003) Enzymology of repair of etheno-adducts. *Mutation research* **531**, 219-229

18. Basu, A. K., Wood, M. L., Niedernhofer, L. J., Ramos, L. A., and Essigmann, J. M. (1993) Mutagenic and genotoxic effects of three vinyl chloride-induced DNA lesions: 1,N6-ethenoadenine, 3,N4-ethenocytosine, and 4-amino-5-(imidazol-2-yl)imidazole. *Biochemistry* **32**, 12793-12801

19. 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 research* **60**, 4098-4104

20. Pandya, G. A., and Moriya, M. (1996) 1,N6-ethenodeoxyadenosine, a DNA adduct highly mutagenic in mammalian cells. *Biochemistry* **35**, 11487-11492

21. 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

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

23. Lyons, D. M., and O'Brien, P. J. (2009) Efficient recognition of an unpaired lesion by a DNA repair glycosylase. *Journal of the American Chemical Society* **131**, 17742-17743

24. Studier, F. W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein expression and purification* **41**, 207-234

25. Manning, G. S. (1978) The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Quarterly reviews of biophysics* **11**, 179-246

26. Record, M. T., Jr., Anderson, C. F., and Lohman, T. M. (1978) Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: the roles of ion association or release, screening, and ion effects on water activity. *Quarterly reviews of biophysics* **11**, 103-178

27. Hedglin, M., Zhang, Y., and O'Brien, P. J. (2013) Isolating contributions from intersegmental transfer to DNA searching by alkyladenine DNA glycosylase. *The Journal of biological chemistry* **288**, 24550-24559

28. Baldwin, M. R., and O'Brien, P. J. (2010) Nonspecific DNA binding and coordination of the first two steps of base excision repair. *Biochemistry* **49**, 7879-7891

29. 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. *Journal of molecular biology* **320**, 503-513

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. *The Journal of biological chemistry* **276**, 13379-13387

31. Speina, E., Ciesla, J. M., Wojcik, J., Bajek, M., Kusmierek, 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. *The Journal of biological chemistry* **276**, 21821-21827

32. Dick, D. A. (1978) The distribution of sodium, potassium and chloride in the nucleus and cytoplasm of Bufo bufo oocytes measured by electron microprobe analysis. *The Journal of physiology* **284**, 37-53

33. Paine, P. L., Pearson, T. W., Tluczek, L. J., and Horowitz, S. B. (1981) Nuclear sodium and potassium. *Nature* **291**, 258-259

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

35. Barbin, A., Froment, O., Boivin, S., Marion, M. J., Belpoggi, F., Maltoni, C., and Montesano, R. (1997) p53 gene mutation pattern in rat liver tumors induced by vinyl chloride. *Cancer research* **57**, 1695-1698

36. Ringvoll, J., Moen, M. N., Nordstrand, L. M., Meira, L. B., Pang, B., Bekkelund, A., Dedon, P. C., Bjelland, S., Samson, L. D., Falnes, P. O., and Klungland, A. (2008) AlkB homologue 2-mediated repair of ethenoadenine lesions in mammalian DNA. *Cancer research* **68**, 4142-4149

37.   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. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 13573-13578

38.   de los Santos, C., Kouchakdjian, M., Yarema, K., Basu, A., Essigmann, J., and Patel, D. J. (1991) NMR studies of the exocyclic 1,N6-ethenodeoxyadenosine adduct (epsilon dA) opposite deoxyguanosine in a DNA duplex. Epsilon dA(syn).dG(anti) pairing at the lesion site. *Biochemistry* **30**, 1828-1835

39.   Kouchakdjian, M., Eisenberg, M., Yarema, K., Basu, A., Essigmann, J., and Patel, D. J. (1991) NMR studies of the exocyclic 1,N6-ethenodeoxyadenosine adduct (epsilon dA) opposite thymidine in a DNA duplex. Nonplanar alignment of epsilon dA(anti) and dT(anti) at the lesion site. *Biochemistry* **30**, 1820-1828

40.   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

Appendix C

Additional information to support Chapter 4

```
25EX   *5'-CGATAGCATCCTECCTTCTCTCCAT-3'
        3'-GCTATCGTAGGAXGGAAGAGAGGTA-5',    X=A,T,G or C

19YT   *5'-TAGCATCCTYCCTTCTCTC-3'
        3'-ATCGTAGGATGGAAGAGAG-5',          Y=E or I

25AT   *5'-CGATAGCATCCTACCTTCTCTCCAT-3'
        3'-GCTATCGTAGGATGGAAGAGAGGTA-5'

47EI   *5'-TAGCATCCTECCTCGTGTAGGTATTAGATCCGACTICCTTGTGTCCT-3'*
        3'-ATCGTAGGATGGAGCACATCCATAATCTAGGCTGATGGAACACAGGA-5'

47IE   *5'-TAGCATCCTICCTCGTGTAGGTATTAGATCCGACTECCTTGTGTCCT-3'*
        3'-ATCGTAGGATGGAGCACATCCATAATCTAGGCTGATGGAACACAGGA-5'
```
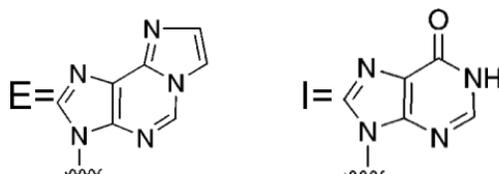

Figure C-1. Oligonucleotide sequences used for catalytic specificity measurements.
The target lesion sites are shown in bold and asterisk indicates the position of the fluorescein label.

A

B
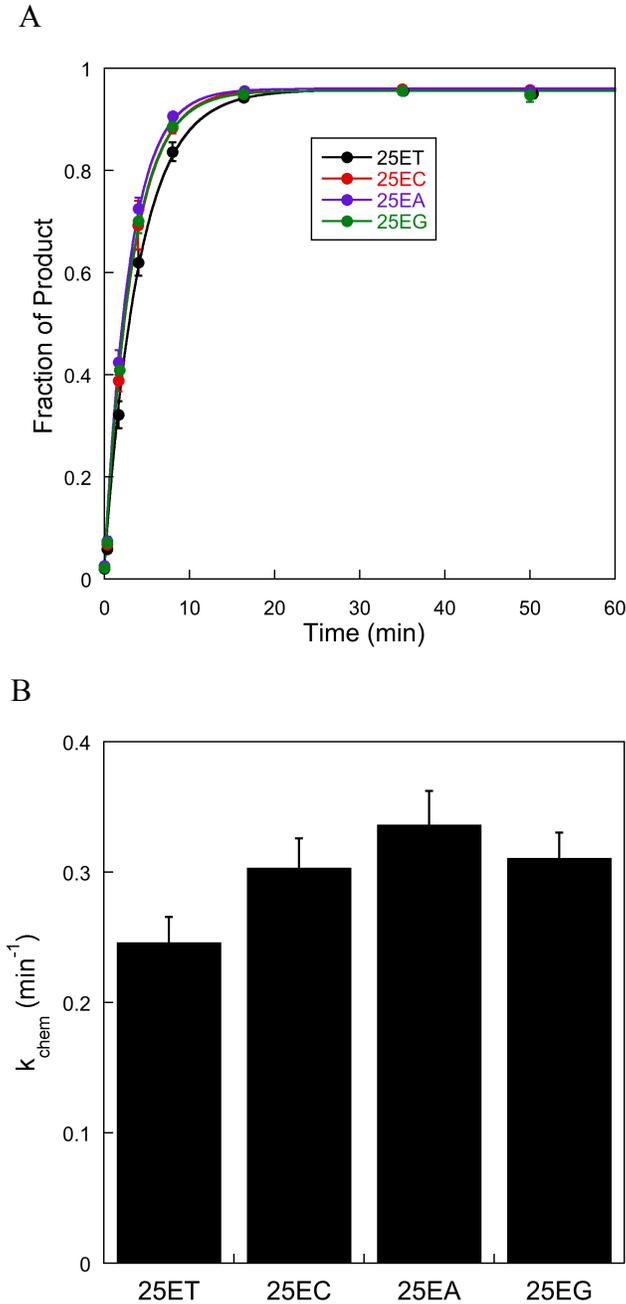
Figure C-2. Single-turnover excision rate constants of εA opposing different bases are all comparable.
A. Single-turnover measurement of AAG on εA lesion opposite different bases. At least two enzyme concentrations in excess of the DNA substrates were used for the measurements. The measured data points were fit by single exponential. B. A plot of the calculated k_{chem} values from panel A. AAG has k_{chem} values ranging from 0.25 ± 0.02 min^{-1} on 25εA•T to 0.34 ± 0.03 min^{-1} on 25εA•A substrates. Therefore, AAG excises the εA lesion opposing different bases with very similar single-turnover rate constants. All the data points and k_{chem} results were calculated from triplicate measurements and the average values and standard deviation are shown.
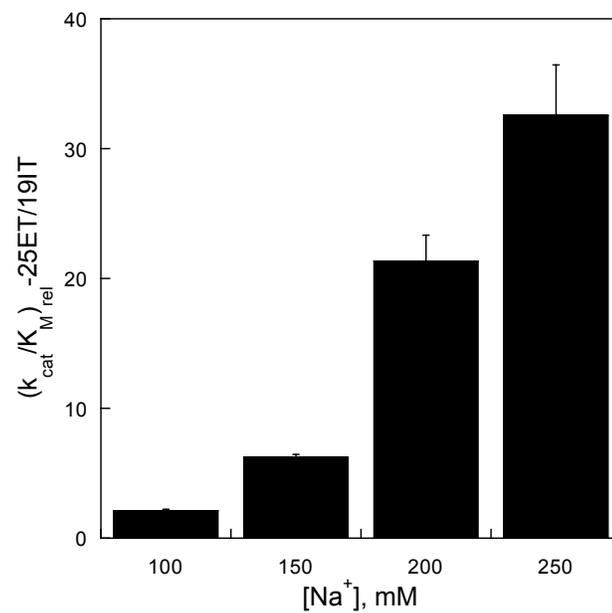
Figure C-3. Direct competition of εA•T and Hx•T DNA.
Direct competition assay of 25εA•T and 19Hx•T was performed with 200 nM DNA substrate (each), 4 nM AAG at 100-250 mM Na$^+$. Reactions were performed in duplicate, and the mean ± S.D. is plotted.

Chapter 5

Contributions of Positively Charged DNA Binding Residues to Searching and Catalysis by Human Alkyladenine DNA Glycosylase

Many metabolic enzymes with enormous catalytic power are thought to have evolved by optimizing k_{cat}/K_M (1-3), which is an apparent second order rate constant for the reaction of free enzyme and free substrate, and it has been generally accepted that the *in vivo* activity of a metabolic enzyme is dependent on its catalytic specificity. This seems to be true for most cytoplasmic enzymes when the encounter of the free enzyme and free substrate in the three-dimensional space is controlled by the diffusion limit (4,5). Rhomboid proteases, a family of intramembrane proteases restrained to the water-excluding, two-dimensional membrane environment to interact with their substrates and catalyze the proteolytic cleavage, have recently been shown to be governed by k_{cat} but not by substrate binding affinity (6). This raises the question of whether the reduction in dimension may contribute to the observed difference. DNA as a polyanionic macromolecule has distinct properties from regular small molecule substrates, and proteins that interact with specific DNA sites bind also nonspecifically to noncognate sites with appreciable affinity. Indeed, most of the DNA-interacting proteins that are not bound at their target sites *in vivo* can be found bound to nonspecific sites (7).Nonspecific DNA binding can be used to enable a more efficient search for specific sites via processes collectively referred to as facilitated diffusion (8). As an important intermediate step in the process of cognate site recognition and binding, nonspecific interactions can therefore play an important role in the physiological functions of the protein.

Previous studies using *E. coli* as a model organism suggested that the searching efficiency of an enzyme on its substrate DNA correlates with its *in vivo* cell function (9,10). In the prokaryotic *E. coli* cells, DNA exists as a naked molecule in the form of a circular chromosome, and all sites are freely accessible to a searching protein. In contrast, DNA molecules in eukaryotic cells are organized into nucleosomes and structures of higher complexity with various nucleoproteins, and the accessibility of specific sites on the chromosome to an

enzyme can be greatly limited. Indeed, it has been reported that many DNA-interacting enzymes, such as excision nucleases and DNA glycosylases, showed different levels of decrease in activity on the nucleosomal substrates (11,12). Therefore, these results raise the question whether kinetic parameters measured *in vitro* using naked DNA substrates still correlate with the *in vivo* function of a DNA-interacting enzyme in the eukaryotic nucleus.

Human alkyladenine DNA glycosylase (AAG) initiates the base lesion repair pathway, and it is the only DNA glycosylase in human cells to remove alkylated purine bases, such as 1, N^6-ethenoadenine (εA), 3-methyladenine and 7-methylguanine (13-16). Previous studies show that AAG searches for its target lesion sites on DNA via facilitated diffusion (17,18). The crystal structure of AAG in complex with its substrate DNA has been solved, revealing a positively charged binding interface and a backbone bending angle of 22° across the central eight base pairs (19,20). However, it is not clear whether the protein uses the same binding interface for nonspecific DNA binding and searching, as studies have revealed that specific binding is often coupled with extensive conformational changes in both protein and DNA (3,21,22).

In this study, we use the well-characterized AAG to test whether it uses the same electrostatic patch for specific and specific interactions, and if k_{cat}, k_{cat}/K_M or searching efficiency is correlated with its *in vivo* function in yeast cells. To this end, we constructed and purified wild type and mutant AAG, and extensively characterized these proteins using kinetic assays, including single-turnover, multiple-turnover and processivity measurements. The same proteins were subsequently introduced into a functional complementation assay system in the yeast cell, and the ability of these proteins to protect the yeast cells from the cytotoxic effects of an alkylating agent was measured. Our results show that k_{cat}/K_M and searching efficiency are positively correlated, and that both parameters correlate with yeast cell survival, suggesting that the ability of AAG to capture lesions is important for survival under conditions of exogenous DNA damaging agents.

MATERIALS AND METHODS

Proteins

Full-length human AAG was cloned into the pET SUMO expression vector (Invitrogen) between the N-terminal BamHI and C-terminal XhoI site, encoding an N-terminal 6× His-Smt3 fusion protein that can be cleaved with the SUMO protease ULP1. All the AAG mutants were

constructed using site-directed mutagenesis from the wild type AAG plasmid, and their sequences were confirmed by DNA sequencing. All the AAG fusion proteins were expressed in *E. coli* BL21 Star (DE3) pRare2 cells using autoinduction media (23). Typically, cells were grown at 37 °C until they reached an optical density of 0.5-1.0 at 600 nm. Cultures were then transferred to a 16-25 °C shaker, and proteins were expressed for 16-24 hours with final OD_{600} values at 15~25 before cells were harvested and frozen at -80 °C. Cell pellets were suspended in lysis buffer (25 mM potassium phosphate, pH 7.0, 250 mM NaCl, 10% (v/v) Glycerol, 0.1% NP-40) and disrupted by a high pressure homogenizer (EmulsiFlex, Avestin). Cell lysates were slowly precipitated with polyethylenimine to remove nucleic acids, and AAG proteins were purified by metal affinity chromatography using a His tag that was subsequently cleaved with the SUMO protease ULP1 to produce full-length AAG with an N-terminal serine in place of the initiator methionine of the native protein. Subsequent ion exchange chromatography (HiTrap SP, GE Healthcare) and gel filtration (Superdex 200, GE Healthcare) or dialysis (MWCO 12,000-14,000, Fisherbrand) yielded proteins that were homogeneous as judged by SDS-PAGE with Coomassie staining.

The active concentrations of the AAG variants were determined by burst analysis as described previously (17), and the enzyme concentrations used in the experiments described here all refer to the corrected active concentrations.

Oligonucleotides

DNA substrates were synthesized by Integrated DNA Technologies or the Keck Center at Yale University and purified by denaturing PAGE as previously described (17). Oligo concentrations were determined from the absorbance at 260 nm using the calculated extinction coefficients. For 5' or dual fluorescein-labeled oligonucleotides we assessed the labeling efficiency by comparing the absorbance at 260 nm with that at 495 nm. DNA duplexes were annealed at 1:1 ratio of the lesion containing fluorescein-labeled strand and the complement for the k_{cat}/K_M assay and 1:2 for the rest. The sequences of the DNA oligonucleotides used in this study are listed in Appendix Figure D-1.

Glycosylase Activity Assays

Reactions were carried out at 37 °C in a buffer system consisted of 50 mM NaMES pH 6.1, 10% (v/v) glycerol, 0.1 mg/mL BSA, 1 mM DTT, 1 mM EDTA and varying concentrations of NaCl to obtain the desired ionic strength. Reactions were initiated by the addition of enzyme

to a final reaction volume of 20-60 μL. Aliquots were withdrawn at various times and quenched with NaOH (0.2 M final concentration). Samples were heated at 70 $^{\circ}$C for 15 min, loading buffer consisting of 10 mM EDTA and 98% formamide was added, and the DNA species were resolved on 14% (w/v) denaturing polyacrylamide gels with 8 M urea. Gels were scanned with a Typhoon Trio+ fluorescence imager (GE Healthcare), and fluorescein signal was detected using excitation wavelength at 532 nm and emission with a 520BP40 filter. The resulting fluorescence signal was quantified with ImageQuant TL and corrected for background signal. The intensity of each DNA band was converted into a fraction of the total DNA by dividing its intensity by the sum of the intensities for all of the DNA species in the reaction.

Single-turnover Kinetics

The single-turnover assay was performed on the εA-containing 25mer DNA substrate as previously described (24-26). The concentration of AAG was varied in excess over the DNA substrates (100 nM) to ensure single-turnover conditions. In all cases the reactions followed single exponential kinetics $F = A(1-exp(-kt))$, in which A is the reaction amplitude and k is the observed single-turnover rate constant.

Multiple-turnover Processivity Assay

The multiple-turnover assay was performed on the previously characterized processivity substrate with 100-fold excess of substrate (200 nM) over enzyme (2 nM) as described (17,18). The initial rates for intermediate (V_{int}) and product formation (V_{prod}) were determined from linear fits of the first 15% reactions. In all cases, the two products and two intermediates were formed at identical rates, indicating that AAG encounters either lesion with equal probability. The fraction processive values were calculated by $F_{P,obs} = (V_{int} - V_{prod})/(V_{int} + V_{prod})$ and corrected for the small amount of ring-opened εA as described previously by $F_P = (F_{P,obs} - 0.05)/(0.92 - 0.05)$ (18).

Catalytic Specificity Measurement

Catalytic specificity k_{cat}/K_M was measured on the εA-containing 25-mer DNA substrate. We used high salt concentration at 1 M Na$^+$ in the reaction to eliminate the contribution of facilitated diffusion so that potential differences in k_{cat}/K_M among these AAG mutants can be revealed. Under these reaction conditions, the Michaelis constant K_M was at μM level even for the wild type enzyme. Therefore, DNA substrate concentrations at 25-300 nM were used for all AAG variants with fixed enzyme concentration at 2 or 3 nM, except for R197S. 1000-4000 nM

DNA and 40 nM enzyme were used for R197S due to its extremely low reaction rates and much weakened substrate binding affinity. The k_{cat}/K_M values were calculated from the slopes of the linear fits.

Yeast Survival Assay

pYES2-N169S AAG plasmid was a kind gift from Dr. Michael Wyatt (University of South Carolina) (27), and all the AAG wild type and mutants were constructed by site-directed mutagenesis using this plasmid. The BY4741-Mag1Δ yeast strain (MATa his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 mag1::KanMX), which is highly sensitive to alkylating agent treatment (28), was a kind gift from Dr. Dave Engelke (University of Michigan, Ann Arbor). Single colonies from re-streaked plates following plasmid transformation were inoculated into 5 mL SD-URA drop-out broth containing glucose or raffinose as the only carbon source for 3 days until yeast growth reached saturation. The saturated cell cultures were then inoculated into 25 mL fresh SD-URA drop-out broth containing glucose (from the glucose culture) or galactose (from the raffinose culture) with a starting OD_{600} of 0.05. Cells were grown for ~12 hours and reached an OD_{600} of ~1 before they were collected for the MMS sensitivity assay and the measurement of the AAG expression level (see the section below). Samples were incubated at room temperature on bench with slow rotation (LabQuake®, Thermo Scientific) with or without 0.3% (v/v) MMS for 1 hr, and they were then immediately diluted and plated on YPD rich media (Figure 5-6A). Plates were incubated at 30 °C for 2-3 days before colonies were scored. Efforts were made to have 20-300 colonies per plate for accurate counting. Survival rates were calculated by dividing the number of colonies after MMS treatment by the total number of colonies in the control samples without MMS treatment.

AAG Concentration Determination in Yeast Cells

Yeast cells used for the AAG expression level measurement were from the same cultures as used for the yeast survival assay. Cells were split at OD_{600} of ~1 after overnight cultivation, and 1 mL of the cell cultures was used for the MMS survival assay and the rest were harvested by centrifugation and stored at -80 °C. Cell pellets were resuspended with 50 μL disruption buffer for the glucose cultures and 150 μL buffer for the galactose cultures. The cell suspensions were mixed with equal volumes of ice-cold glass beads (acid washed 425-600 μm, Sigma-Aldrich) and disrupted at 6.5 m/s using MP FastPrep®-24 (MP Bio) for 1 min to make yeast whole cell extract. Cell lysates were centrifuged at 13.2k rpm for 60 min to remove cell debris

and pellets, and the supernatants were flash frozen in liquid nitrogen and kept at -80 $^{\circ}$C for the glycosylase assays.

The prepared yeast whole cell extracts were assayed for AAG glycosylase activity using the 25Hx-bulge DNA (Appendix Figure D-1) as the substrate at 250 mM Na$^+$ under the standard glycosylase activity assay condition, except that 3 mM EDTA was used to eliminate unwanted nonspecific DNA degradation. As an internal quality control of the cell extract preparation protocol, all the extracts were also assayed on the 25U•G DNA substrate (Appendix Figure D-1) for their intrinsic yeast uracil DNA glycosylase (UDG) activity under the same reaction conditions. Concentrations of the AAG variants in the yeast cells were calculated using the measured activity divided by the measured rate constants using the recombinant AAG variants under the same reaction conditions (Appendix Table D-1).

RESULTS

Glycosylase Activity and Searching Efficiency of AAG Mutants Towards εA DNA

To examine the contribution of the positively charged binding interface in the εA DNA-AAG crystal structure to nonspecific binding and searching, we individually mutated three lysine and five arginine residues at the DNA binding interface into serine or methionine residues (Figure 5-1A). Four of the amino acids, R141, R182, R197 and K229, interact directly with the DNA backbone via hydrogen bonds and salt bridges (Figure 5-1A, blue residues), while the other residues (R138, R145, K210 and K220) are further away from the DNA backbone, and interact with the molecule indirectly (Figure 5-1A, red residues) (19,20). R182, R197 and K220 interact with the lesion-containing DNA strand in the structure, while the rest of the residues interact with the opposing strand. These positively charged residues spread evenly throughout the whole binding interface, with four residues located upstream of the lesion (R138, R141, R145 and K229) and the remaining four downstream of the site. The electron density of the side chain of K210 was not observed in the crystal diffraction data (20), and an alanine residue was modeled in the structure as a result. All of these positively charged residues are conserved throughout evolution in mammalian species, if not in all organisms (Table 5-1). We sought minimally perturbing mutations that would remove the positive charge without globally affecting AAG structure. Therefore, the choice of mutations were guided by the results of a random mutation assay by Loeb and co-workers, which identified AAG variants that could be tolerated through a

high-throughput selection of *E. coli* cell survival after methyl methanesulfonate (MMS) treatment (Table 5-1) (29).
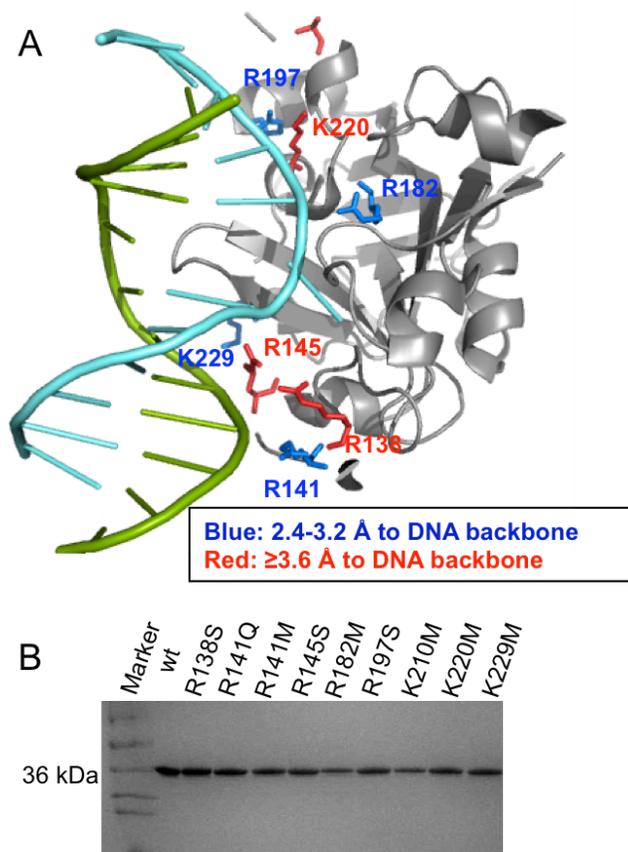
Figure 5-1. AAG mutants tested in this study.
A. Cartoon representation of the AAG-εA DNA crystal structure (PDB: 1f4r), with the mutated residues shown in sticks (20). The lesion-containing strand of the DNA molecule is colored in cyan and the opposing strand is in green. The protein molecule is shown in grey. All the mutated residues are located at the protein-DNA binding interface, yet far away from the protein active site. Residues were colored differently based on their distances to the DNA backbone: residues colored in blue interact directly with the DNA molecule, while those in red do not make direct interactions. B. SDS-PAGE of the purified recombinant AAG variants.

AAG expression plasmids were constructed and proteins were expressed and purified as described in Materials and Methods. All the AAG variants were purified to homogeneity as determined by SDS-PAGE with coomassie staining (Figure 5-1B). The active enzyme concentrations of the AAG variants were determined by active site titration as described previously (17), and all the following *in vitro* characterizations used the corrected protein concentrations (Appendix Figure D-2).

Table 5-1. Known AAG variants and mutations chosen for removal of positively charged residues at the DNA binding interface.

Residue	Natural Variation	Loeb Mutant[&]	Mut. Const.
R138	R*	W,S	S
R141	R/K	L	M
R145	R	H,C	S
R182	R	none	M
R197	R	S,C	S
K210	K*	M,W,E	M
K220	K/R	M,T	M
K229	K*	M,T	M

*In mammalian only
[&]Loeb L.A. and co-workers, PNAS (2004) 101:9205-9210

To evaluate the glycosylase activity of the AAG variants, single-turnover activity was measured with the enzyme concentrations in excess over the DNA substrate. The single-turnover reaction includes all the steps following the formation of the initial AAG•DNA complex up to and including N-glycosidic bond cleavage (Figure 5-2A), and the measured single-turnover rate constants reflect the chemical catalysis step in this assay since the searching step is fast and efficient. The reaction time courses were fit by a single exponential equation (Figure 5-2B). As the mutated amino acids are all far away from the enzyme active site (Figure 5-1A) and most of the mutants were identified in the *in vivo* screening assay (29), it was expected that their glycosylase activities would be retained. Indeed, all the mutants have very similar single-turnover rate constants to the wild type enzyme (Figure 5-2C). The wild type AAG showed a k_{max} of 0.22 min^{-1} in this assay (Figure 5-2C), consistent with the previously reported results (17,25). The R145S mutant and the K220M mutant had the lowest and highest k_{max} values at 0.14 min^{-1} and 0.32 min^{-1} respectively, still well within 50% that of the wild type protein. These results demonstrate that the chemical step of N-glycosidic bond cleavage is not compromised by these mutations, indicating that the structure is not affected by the mutations.
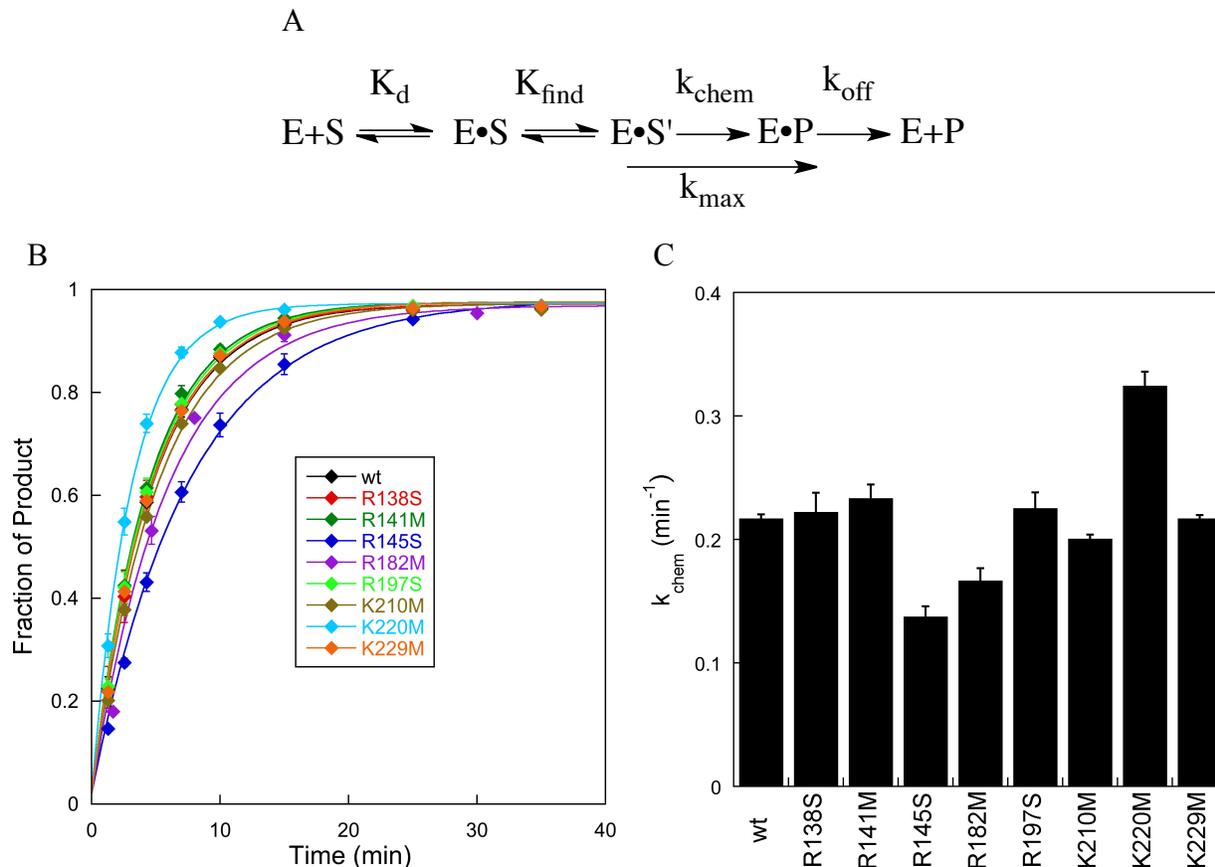
Figure 5-2. All AAG mutants retain glycosylase activity on the substrate DNA.
A. A minimal reaction scheme of AAG on its substrate DNA. The single-turnover assay monitors the chemical step. E•S denotes the nonspecific binding complex, while E•S' is the specific complex. Because the finding step is much faster than the chemical step under the reaction condition used here, the measured single-turnover rate constant k_{max} is simply the chemical step of N-glycosidic bond cleavage. B. Single-turnover product formation as a function of time by the wild type and mutant AAG. Two concentrations of AAG (250 and 500 nM) were used for each variant and they gave identical reaction curves (50 nM 25εA•T substrate DNA), indicating that AAG is saturating under these conditions. The lines indicate the best fits to a single exponential. C. A summary bar graph of the single-turnover rate constants of the AAG variants (mean ± S.D., n = 3). All the AAG mutants have very similar single-turnover rate constants to the wild type.

To more fully explore the impact of the individual mutations on εA catalysis, we measured the second order rate constant, k_{cat}/K_M, under multiple-turnover conditions. This parameter measures both binding and turnover. At low salt concentration the binding of AAG to its substrate is irreversible, and the protein uses facilitated diffusion to search every site on the DNA molecule. As a result, the k_{cat}/K_M values simply reflect the association rate constants. At high salt concentration, AAG binding is fully reversible, and catalytic specificity measurement includes all the steps in the enzymatic reaction from the binding of free enzyme and free

substrate up to and including the first irreversible step, which corresponds to N-glycosidic bond cleavage in this case (Figure 5-3A). Therefore, k_{cat}/K_M results were obtained by performing

A

B   C
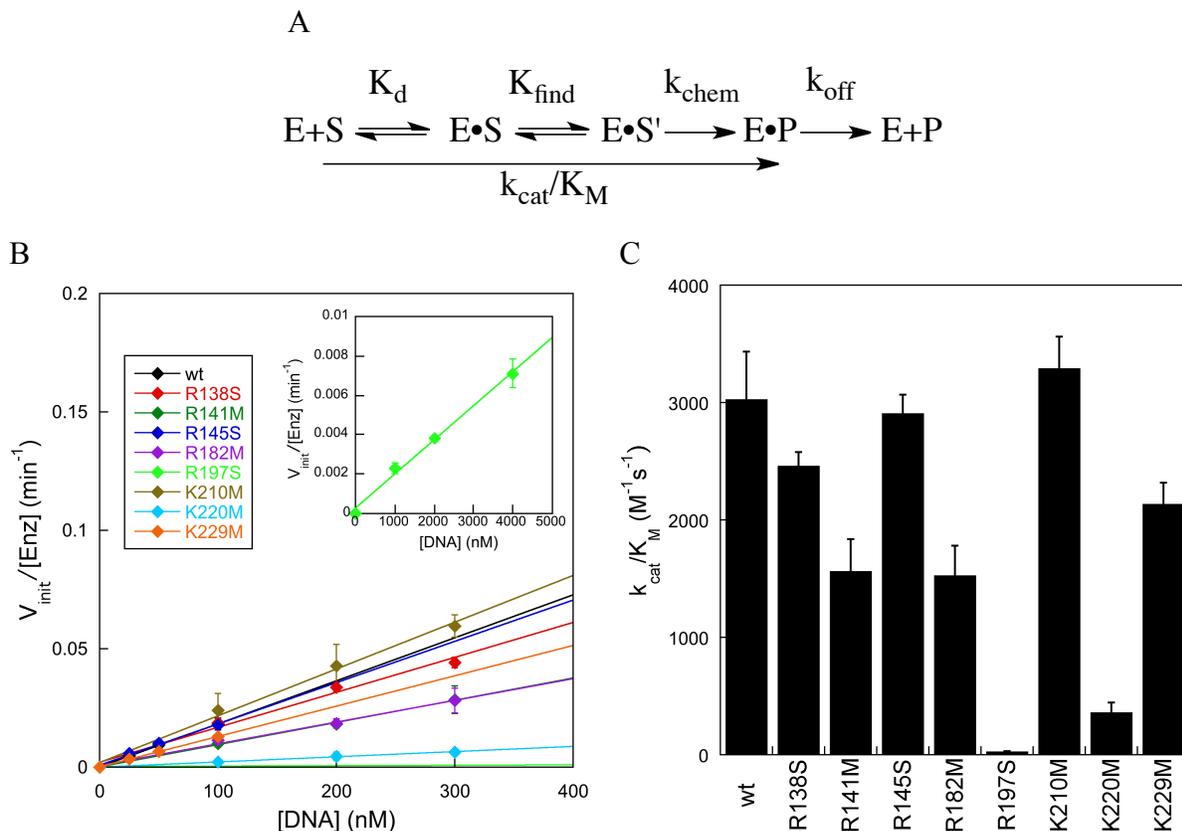
Figure 5-3. AAG mutants have different catalytic specificity values.
A. The k_{cat}/K_M measurement monitors both the enzyme-substrate binding and the chemical catalysis step. B. Initial rate constants V_{init}/[Enz] are plotted as a function of substrate concentration. Reactions were performed under multiple-turnover conditions, and the initial rate constants were calculated from the first 15% of reactions. Catalytic specificity values were calculated from the slopes of the linear fits. The inset shows the reaction of the R197S mutant for which higher substrate concentrations were used. C. A summary bar graph of the catalytic specificity of the AAG variants (mean ± S.D., n ≥ 3). The tested mutants show different degrees of decrease in k_{cat}/K_M compared to the wild type protein.

multiple-turnover reactions at 1000 mM Na$^+$ at substrate concentrations below the K_M values, which are in the range of low μM to tens of μM for different AAG variants (data not shown). The initial rate constants are plotted as a function of substrate concentration in Figure 5-3B, and catalytic specificity values were calculated from the slopes of the linear fits. The wild type enzyme had a k_{cat}/K_M value of 3000 ± 400 M^{-1}s^{-1}, and different mutants showed varying degrees of decrease compared to the wild type enzyme (Figure 5-3C). Among these, two mutants R197S and K220M, both of which interact with the lesion-containing strand, had the most dramatic decreases in catalytic specificity at 30 ± 3 M^{-1}s^{-1} and 360 ± 80 M^{-1}s^{-1}, which correspond to 100-

fold and 8-fold decreases compared to the wild type enzyme respectively. In contrast, R182M, which also interacts with the lesion-containing strand, only had a 2-fold decrease in k_{cat}/K_M compared to the wild type enzyme. All the mutants that interact with the opposing DNA strand showed k_{cat}/K_M values from no decrease to approximately 2-fold decreases compared to the wild type enzyme.

Catalytic specificity as measured above reflects specific binding and catalysis of AAG towards its substrate DNA, while a processivity assay can provide information about the nonspecific binding and searching process. Therefore, we performed the multiple-turnover fraction processivity assay, which allows the translocation of AAG between two sites on a DNA molecule to be monitored by measuring the correlated cleavage events on a substrate DNA containing two sites of cognate damage (47εA2F2 DNA substrate) (Figure 5-4A). Consistent
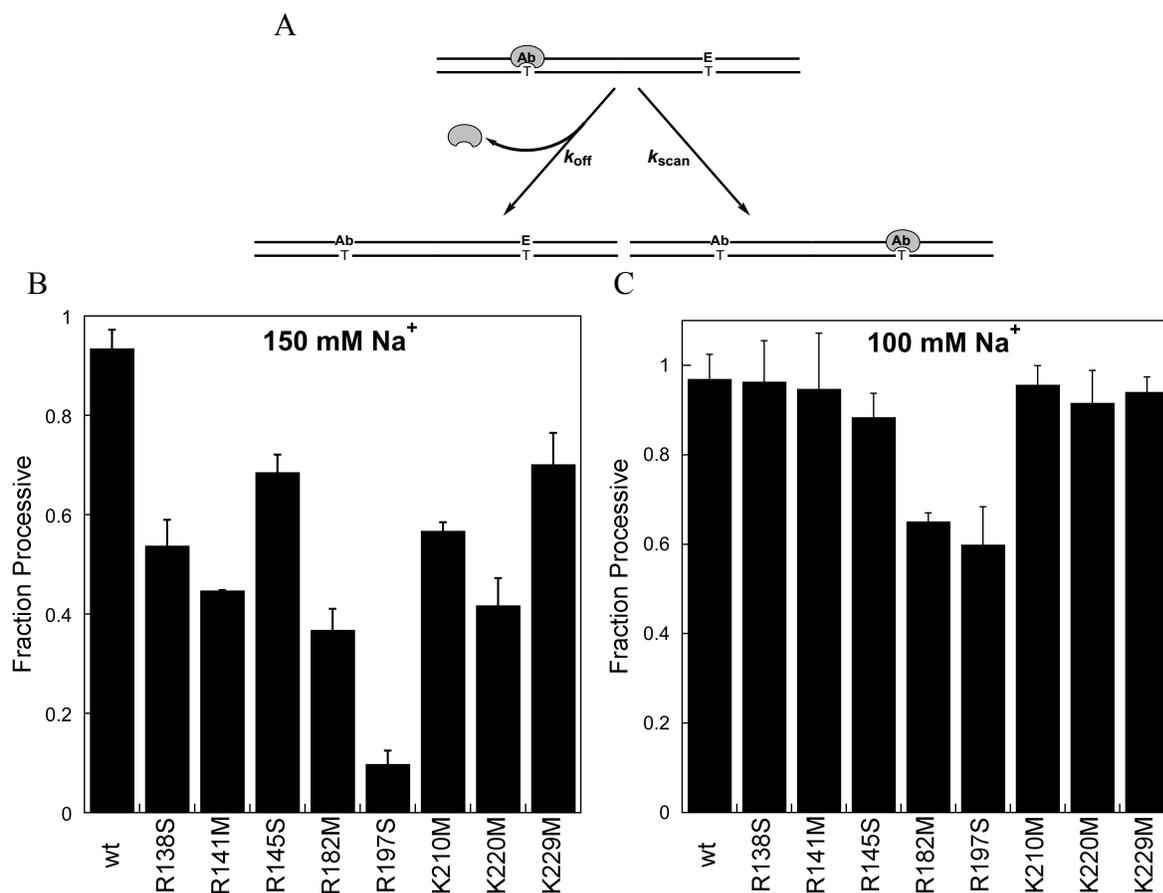
Figure 5-4. AAG mutants show decreased searching efficiency at physiological salt condition.
A. A scheme of the processivity assay, which monitors the fraction of the correlated excision events of the two lesion sites on the same DNA molecule. B. AAG mutants show different degrees of searching deficiency compared to the wild type enzyme at 150 mM Na^+. C. All AAG mutants are capable of efficient searching at low salt condition at 100 mM Na^+. All reactions were performed at least in triplicate, and the average and standard deviation are shown. To be consistent with the previously published results, the F_P values shown in here were normalized according to the previously derived equation (18).

with previous results, the wild type enzyme searches the model substrate very efficiently at 150 mM Na$^+$, and the fraction processivity value (F_P) is close to unity (17,18). In contrast, all the AAG mutants showed decreased searching efficiency compared to the wild type enzyme. The R197S mutant had the most dramatic decrease in processivity ($F_P = 0.10$), indicating that searching is almost completely distributive for this mutant (Figure 5-4B). Previous assays showed that processivity is mediated by electrostatic interactions and is salt concentration dependent (17). Therefore, we performed the processivity assay on all the AAG mutants at decreased salt to test if they are still capable of efficient searching. All the mutants had F_P values close to the wild type protein ($F_P \sim 1$) at 100 mM Na$^+$, except for R182M and R197S, which showed a slight decrease in their fraction processive values at around 0.6 (Figure 5-4C).

We noticed similar trends of changes in the two measured kinetic parameters k_{cat}/K_M and F_P for the AAG mutants. Therefore, we plotted k_{cat}/K_M versus F_P, and observed a positive linear correlation of the two parameters with a R^2 value of 0.82 (Figure 5-5). The observed positive correlation between the catalytic specificity and the searching ability suggests that most electrostatic contacts at the DNA-AAG binding interface are important for both specific and nonspecific DNA binding.
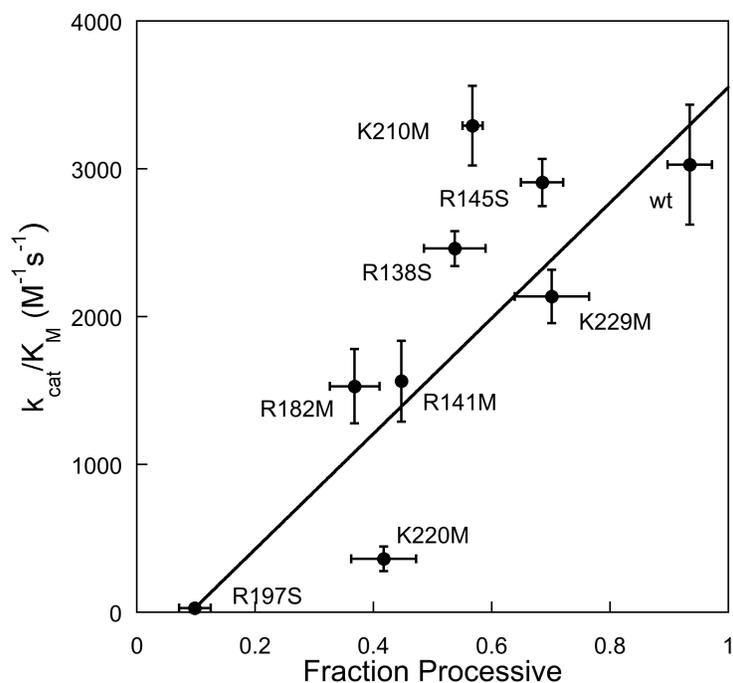

Figure 5-5. Searching efficiency correlates with catalytic specificity.
Catalytic specificity k_{cat}/K_M and searching efficiency F_P of AAG variants shows a positive correlation with a R^2 value of 0.82. The line is from a weighted linear regression analysis using the relative errors in the k_{cat}/K_M values.

AAG Mutants Protect Yeast Cells to Different Extents

To test the *in vivo* function of the constructed AAG mutants, we performed functional complementation assays in the baker's yeast *S. cerevisiae*. Previous studies show that yeast cells lacking Mag1, which is the counterpart of AAG in human cells, are highly sensitive to alkylating agents (28). As a control, we compared the sensitivity of Mag1 knockout BY4741 yeast cells and the parent cells to MMS treatment, and confirmed that the Mag1 deletion strain has dramatically decreased survival rates (data not shown). All the AAG variants characterized in the previous *in vitro* assays, together with a catalytically inactive mutant E125Q (20,30), were subcloned into the pYES2 plasmid vector under the control of the Gal promoter for the yeast functional complementation assay. Yeast cells expressing different AAG variants were harvested at log phase and treated with 0.3% (v/v) MMS or water as a control for 1 hr at room temperature (Figure 5-6A). Under the glucose repressive expression condition, yeast cells harboring wild type AAG showed ~10% survival compared to its no MMS treatment control, while those harboring the catalytically inactive mutant E125Q and empty pYES2 vector had less than 0.1% survival. This wide range of separation in survival rates between the wild type AAG and the negative controls makes it possible to distinguish the functional contributions of the different AAG mutants. Indeed, AAG mutants showed varying degrees of protection to their hosting yeast cells under this low expression condition (Figure 5-6B, black bars), and yeast cells expressing the R182M and R197S mutants had more than ten-fold lower survival rates than those expressing the wild type enzyme.

Overexpression of these AAG mutants in the yeast cells is expected to compensate for their poor searching ability and restore their protection to the wild type AAG level. Therefore, we cultured the same yeast cells in galactose media, which induces the overexpression of the AAG variants. We observed elevated yeast survival rates to the wild type protein level for all the functional AAG mutants, except for R182M, which conferred approximately ten-fold less protection compared to other AAG constructs (discussed later) (Figure 5-6B, red bars). As expected, the galactose condition did not affect MMS sensitivity of the yeast cells, because the negative control strains that harbored empty vector or the inactive E125Q mutant showed the same cell survival in either glucose or galactose media.
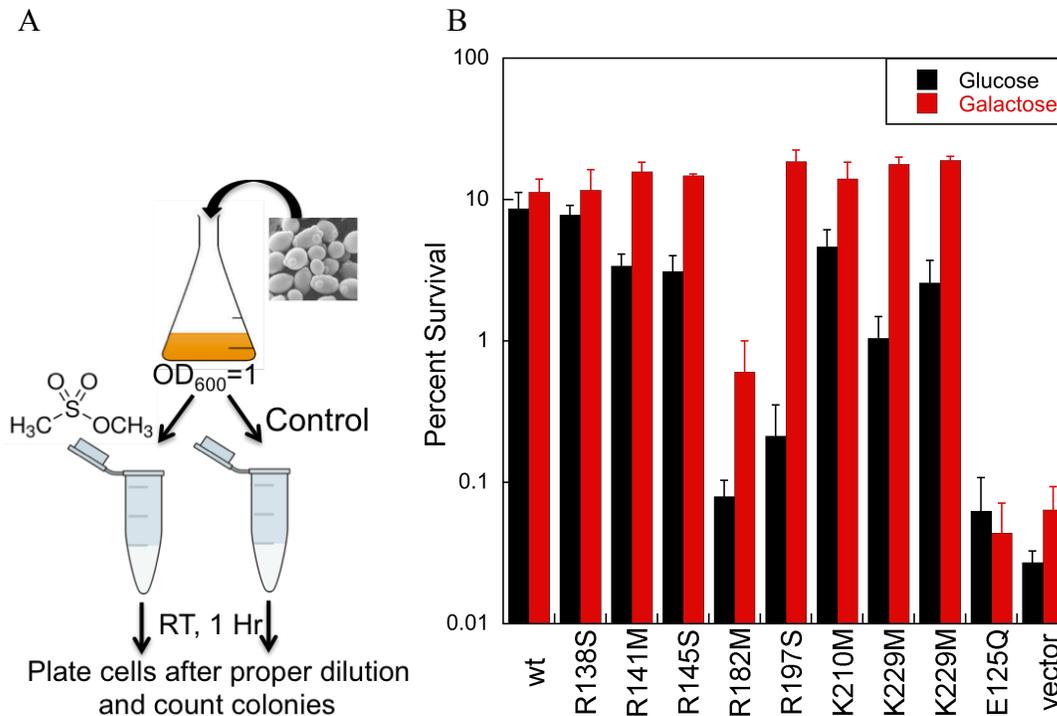
Figure 5-6. Survival rates of yeast cells expressing different AAG variants upon MMS treatment.
A. A scheme of the MMS liquid killing yeast survival assay. 0.3% (v/v) methyl methanesulfonate (MMS) was used to treat log phase yeast cells for 1 hour at room temperature. Survival rates were calculated by comparing the colony forming units of the MMS treated samples and the control samples. B. AAG variants protect yeast cells to different extents under glucose low expression level condition, and almost all of the functional mutants (except for R182S) were restored to the wild type AAG protection level when overexpressed in the galactose broth. E125Q is a catalytically inactive mutant of AAG, and it served as a negative control together with the empty pYES2 vector. All the results were calculated from at least three independent measurements, and the average and standard deviation are shown.

An explicit assumption for the comparison of yeast survival rates among different AAG variants is that they all express at the same level in the yeast cells under the given growth condition. It is possible that point mutants would affect protein stability. Therefore, we used activity assays to quantify the expression level of these AAG proteins from the same log-phase cells as used for the survival assays. Cells were disrupted and whole cell extracts were prepared as described in Materials and Methods. Cell extracts were first assayed for their intrinsic UDG activity on a uracil-containing DNA, which showed that all the cell extracts were comparable to each other (Figure 5-7B). We then measured the AAG glycosylase activity in the cell extracts, and obtained the corresponding enzyme concentrations based on their calculated activity values using known concentrations of the recombinantly purified AAG mutants (Appendix Figure D-3). Most of the AAG variants, including the wild type protein and the majority of the active mutants
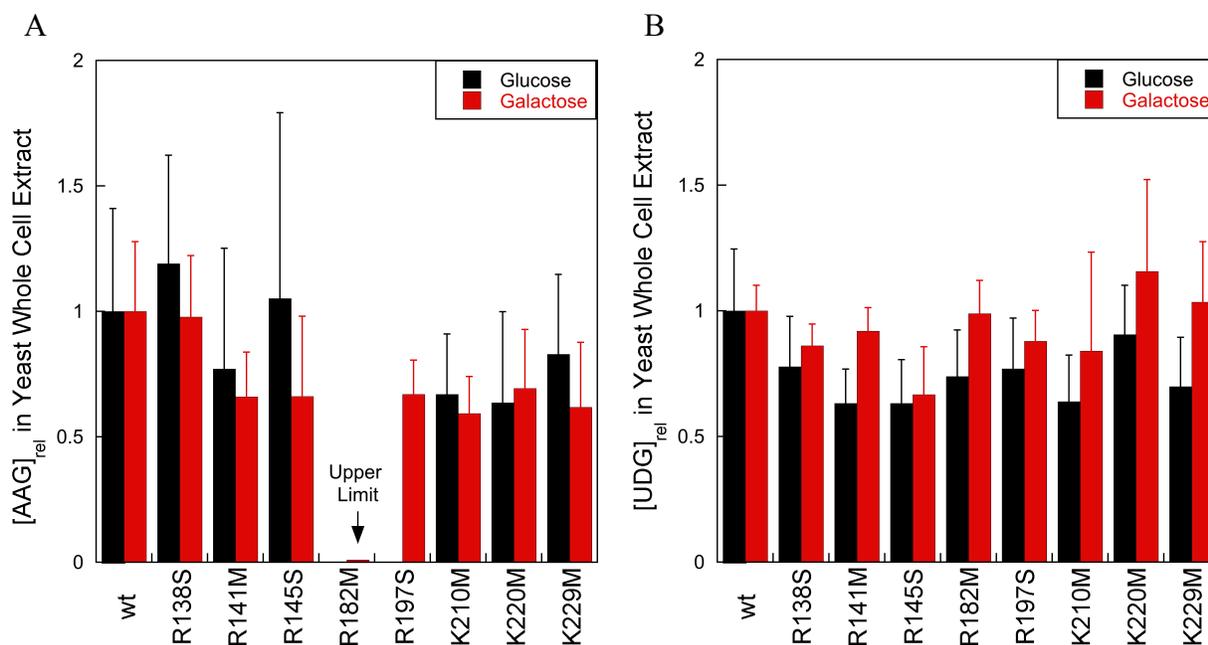
Figure 5-7. Almost all AAG variants were expressed at the same level in the yeast cells.
A. The relative expression levels of AAG variants in yeast cells under glucose and galactose conditions. The AAG expression levels were normalized by dividing that of the wild type protein under glucose or galactose expression conditions. The expression level of the R197S mutant under the glucose condition could not be determined with accuracy due to its extremely low glycosylase activity. It can be postulated that this mutant was expressed at a similar level to the wild type enzyme from its overexpression level under the galactose condition. The expression level of R182M could not be determined under both conditions, and an upper limit was calculated based on the glycosylase activity of the purified recombinant enzyme on the model substrate. Reactions were performed in quadruplicate, and the average and standard deviation are shown. B. All the yeast cell extracts have similar UDG activity. To check for the preparation consistency of yeast cell extracts, the endogenous expressed nuclear protein yeast uracil DNA glycosylase was assayed on its substrate DNA 25U•G under the same reaction condition as described for the AAG glycosylase assay. The UDG levels were normalized by dividing that from the yeast cells expressing the wild type AAG. Reactions were performed in quadruplicate and duplicate under glucose and galactose conditions respectively, and the average and standard deviation are shown.

have the same level of expression (Figure 5-7A; Appendix Table D-1). The expression level of the R197S mutant could not be determined with confidence under the glucose condition, as this mutant has ~100-fold decreased activity compared to the wild-type protein on the model substrate. However, it can be extrapolated from the galactose condition that the R197S mutant was most likely to be expressed at the same level as other variants. In contrast, the R182M mutant was below the detection limit even when it was overexpressed under the galactose condition. A conservative estimation of the expression upper limit of the R182M mutant is 1% of the wild type enzyme. This low expression level also explains the observed abnormality in the yeast survival rate compared to other functional mutants in galactose media. Due to its extremely low level of expression, the R182M mutant was omitted from the correlation of *in vitro* kinetic

parameters and *in vivo* functional outcomes.

In vitro Catalytic Specificity and Processivity Correlate with *in vivo* Cell Survival

To determine which of the *in vitro* kinetic parameters of the AAG mutants best predict the functional complementation results *in vivo*, we analyzed to what extent the k_{chem}, k_{cat}/K_M and F_P values were correlated to the yeast cell survival data. The plot of k_{chem} versus survival rate did not show any clear correlation (Appendix Figure D-4). However, rough correlations of k_{cat}/K_M and F_P values versus survival rate were observed, with the correlation coefficient R^2 values at 0.80 and 0.66, respectively (Figure 5-8). We conclude from this finding that the ability of AAG to capture lesions is important for cell survival.
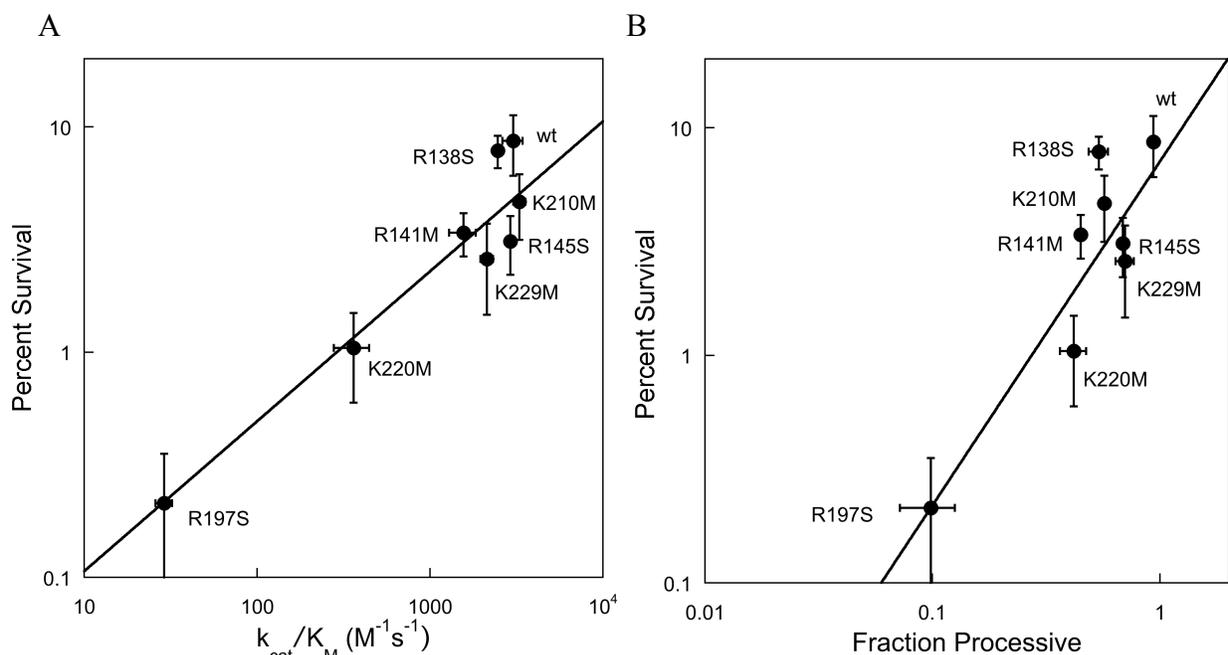
Figure 5-8. *In vitro* kinetic parameters are correlated with *in vivo* cell survival rates.
A. Catalytic specificity k_{cat}/K_M of the AAG variants is positively correlated to the yeast survival rate with a R^2 value of 0.80. B. The searching efficiency F_P of the variants is positively correlated to the yeast survival rate with a R^2 value of 0.66. The lines are from weighted linear regression analyses using the relative errors in the percent survival.

DISCUSSION

Catalytic Specificity and Searching Efficiency are Correlated *in vitro*

Previous studies have shown that AAG catalyzes the removal of εA lesions through nonspecific binding, initial recognition of the lesion site, base flipping and eventually lesion removal in the enzyme active site. The single-turnover assay reports k_{max}, which is equivalent to $k_{chem}*K_{flip}/(K_{flip}+1)$, in which the flipping equilibrium $K_{flip}>>1$ according to previous experiments, and k_{chem} is thus equivalent to the observed k_{max} (24,26). The k_{chem} value is

minimally affected by the mutations tested in this study, which is consistent with the observation that the mutated amino acids are all far away from the enzyme active site. In contrast, catalytic specificity measures both binding and catalysis. At the high salt concentration used for the assay, the measured k_{cat}/K_M values reflect specific binding of AAG to the lesion site and catalysis. The observed wide scattering of the catalytic specificity values for the AAG variants suggests that electrostatic interactions contribute to the stability of the specific binding complex. Searching efficiency is measured by the processivity assay, which monitors the interaction of AAG with nonspecific DNA sequences by correlated excision events. The different degrees of decreases in processivity observed for the AAG mutants demonstrate the involvement of these charged residues in nonspecific DNA binding and the searching process.

Together, interactions of the positively charged residues with the DNA backbone contribute to specific and nonspecific DNA binding, but are not essential for chemical catalysis. The k_{cat}/K_M and F_P values both decreased most dramatically in mutants that interact with the lesion-containing strand, while only mild decreases were seen for other mutants regardless of their distances to the protein active site or the DNA backbone. The linear correlation between k_{cat}/K_M and F_P (Figure 5-5) suggests that nonspecific and specific AAG-DNA complex may take advantage of the same binding surface for interaction, and that the nonspecific complex may not undergo any significant conformational changes upon specific site recognition.

AAG Expression in Yeast

Most of the AAG mutants were expressed at the same level in yeast cells under the same cultivation conditions, with the exception of R182M, which may be due to misfolding or instability of this protein in the yeast cell. Consistent with this observation, no active mutations of the R182 residue were identified in the Loeb study (Table 5-1 and (29)). Under the glucose condition, the number of AAG molecules in the hosting yeast cell is estimated to be 10-20 when we assume a cell density of 2×10^7 cells/OD$_{600}$ (calculated from the plating of control samples without MMS treatment) and a cell disruption efficiency of 90%; under galactose inducing condition, the average number increased 1000-2000 fold (Table D-1) to around 2×10^4 molecules/cell. Considering that the size of the haploid yeast is ~12 Mbp, while the diploid human genome is ~6000 Mbp, the difference of their genome sizes is ~500 fold. To maintain the same ratio of protein to DNA, this level of AAG in the yeast cells under the low expression condition corresponds to $(0.5-1) \times 10^4$ molecules per human cell. This is significantly below the

114

previous estimation of 2×10^5 AAG molecules/cell in cultured human fibroblasts (31). In contrast, the AAG level in the yeast cells under the overexpression condition, which corresponds to approximately 1×10^7 molecules per human cell, is significantly higher than that observed in cultured human fibroblasts.

Yeast Survival Rates Under Different Conditions

Although the expression level of AAG was increased 1000-2000 fold under the galactose inducing condition compared to the glucose repressive condition, the yeast survival rates after MMS treatment did not show significant increases for the wild type protein. There are several possible explanations to account for this observation. First, overexpression of AAG may break the balance with the downstream BER enzymes and oversaturate the repair pathway. Second, it is possible that the few copies of the wild type AAG in the yeast cell under the glucose condition is sufficient to handle its cognate lesion substrates, and more copies of AAG did not really help. Third, overexpression of AAG may result in the excision of innocuous normal bases, and this increased deleterious consequence offsets the beneficial effects. Similar observations to that presented above were made previously by Samson and co-workers, and similar interpretations were also suggested (28).

In contrast to the varying degrees of protection conferred to yeast cells by different AAG variants under the glucose condition, the protections by most of these proteins under the galactose condition were very similar to one another. This result suggests that searching and binding are no longer rate limiting with excess AAG present in the yeast cells.

Correlation of *in vitro* Kinetic Parameters and *in vivo* Function

Previous studies by Dowd *et al.* suggested that mutants of T4 endonuclease V (also known as T4-pdg for pyrimidine dimer glycosylase) with diminished processivity had significantly reduced survival following UV challenge as measured by colony-forming ability of *E. coli* cells (9,32,33). In another study by Jeltsch *et al.*, the authors observed a strong correlation between the ability of the restriction endonuclease EcoRV mutants to search DNA *in vitro* and to protect *E. coli* cells from phage infection *in vivo*, demonstrating the importance of linear diffusion for effective phage restriction (10,34). These two studies highlighted the important role of facilitated diffusion on the survival of prokaryotic cells. However, neither study considered whether k_{cat}/K_M was correlated with linear diffusion.

In this study, we tested the relationship of the *in vitro* kinetic parameters and *in vivo*

115

survival of a DNA repair enzyme in the eukaryotic cell. We observed positive linear correlations of catalytic specificity and searching efficiency of AAG mutants on the model substrate to yeast cell survival rates under low protein expression level conditions, which may not necessarily be the case due to the organizational complexity of the eukaryotic nucleus and the coordination of the rest of the BER pathway. For example, it has been reported that UNG has different degrees of decreases in activity towards its nucleosomal substrates depending on their specific positions in the nucleosome (11,35). The positive linear correlation of catalytic specificity and searching efficiency of the AAG mutants with their protection against MMS killing under the glucose condition suggests that the target site in the genome is accessible to the repair protein. The correlations observed here are also consistent with the previous results that AAG can hop over protein roadblocks and search protein-bound DNA molecules efficiently (18).

References

1.   Fersht, A. R. (1974) Catalysis, binding and enzyme-substrate complementarity. *Proceedings of the Royal Society of London. Series B, Containing papers of a Biological character. Royal Society* **187**, 397-407

2.   Garcia-Viloca, M., Gao, J., Karplus, M., and Truhlar, D. G. (2004) How enzymes work: analysis by modern rate theory and computer simulations. *Science* **303**, 186-195

3.   Warshel, A. (1998) Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites. *The Journal of biological chemistry* **273**, 27035-27038

4.   Samson, R., and Deutch, J. M. . (1978) Diffusion-controlled reaction rate to a buried active site. *J. Chem. Phys.* **68**, 285-290

5.   Schurr, J. M., and Schmitz, K. S. (1976) Orientation constrains and rotational diffusion in bimolecular solution kinetics. A Simplification. *J. Phys. Chem.* **80**, 1934-1936

6.   Dickey, S. W., Baker, R. P., Cho, S., and Urban, S. (2013) Proteolysis inside the membrane is a rate-governed reaction not driven by substrate affinity. *Cell* **155**, 1270-1281

7.   Kao-Huang, Y., Revzin, A., Butler, A. P., O'Conner, P., Noble, D. W., and von Hippel, P. H. (1977) Nonspecific DNA binding of genome-regulating proteins as a biological control mechanism: measurement of DNA-bound Escherichia coli lac repressor in vivo. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 4228-4232

8.   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

9.   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. *The Journal of biological chemistry* **265**, 3424-3431

10.   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. *The EMBO journal* **15**, 5104-5111

11.   Beard, B. C., Wilson, S. H., and Smerdon, M. J. (2003) Suppressed catalytic activity of base excision repair enzymes on rotationally positioned uracil in nucleosomes. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 7465-7470

12.   Hara, R., Mo, J., and Sancar, A. (2000) DNA damage in the nucleosome core is refractory to repair by human excision nuclease. *Molecular and cellular biology* **20**, 9173-9181

13. 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. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 13087-13092

14. 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. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 12869-12874

15. O'Connor, T. R. (1993) Purification and characterization of human 3-methyladenine-DNA glycosylase. *Nucleic acids research* **21**, 5561-5569

16. Singer, B., Antoccia, A., Basu, A. K., Dosanjh, M. K., Fraenkel-Conrat, H., Gallagher, P. E., Kusmierek, J. T., Qiu, Z. H., and Rydberg, B. (1992) Both purified human 1,N6-ethenoadenine-binding protein and purified human 3-methyladenine-DNA glycosylase act on 1,N6-ethenoadenine and 3-methyladenine. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 9386-9390

17. Hedglin, M., and O'Brien, P. J. (2008) Human alkyladenine DNA glycosylase employs a processive search for DNA damage. *Biochemistry* **47**, 11434-11445

18. 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

19. 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

20. 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. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 13573-13578

21. Kalodimos, C. G., Biris, N., Bonvin, A. M., Levandoski, M. M., Guennuegues, M., Boelens, R., and Kaptein, R. (2004) Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. *Science* **305**, 386-389

22. Spolar, R. S., and Record, M. T., Jr. (1994) Coupling of local folding to site-specific binding of proteins to DNA. *Science* **263**, 777-784

23. Studier, F. W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein expression and purification* **41**, 207-234

24. Hendershot, J. M., Wolfe, A. E., and O'Brien, P. J. (2011) Substitution of active site tyrosines with tryptophan alters the free energy for nucleotide flipping by human alkyladenine DNA glycosylase. *Biochemistry* **50**, 1864-1874

25. O'Brien, P. J., and Ellenberger, T. (2004) Dissecting the broad substrate specificity of human 3-methyladenine-DNA glycosylase. *The Journal of biological chemistry* **279**, 9750-9757

26. 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

27. Connor, E. E., Wilson, J. J., and Wyatt, M. D. (2005) Effects of substrate specificity on initiating the base excision repair of N-methylpurines by variant human 3-methyladenine DNA glycosylases. *Chemical research in toxicology* **18**, 87-94

28. Glassner, B. J., Rasmussen, L. J., Najarian, M. T., Posnick, L. M., and Samson, L. D. (1998) Generation of a strong mutator phenotype in yeast by imbalanced base excision repair. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 9997-10002

29. Guo, H. H., Choe, J., and Loeb, L. A. (2004) Protein tolerance to random amino acid change. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 9205-9210

30. 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

31. Ye, N., Holmquist, G. P., and O'Connor, T. R. (1998) Heterogeneous repair of N-methylpurines at the nucleotide level in normal human cells. *Journal of molecular biology* **284**, 269-285

32. Dowd, D. R., and Lloyd, R. S. (1989) Site-directed mutagenesis of the T4 endonuclease V gene: the role of arginine-3 in the target search. *Biochemistry* **28**, 8699-8705

33. 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. *Journal of molecular biology* **208**, 701-707

34. Wenz, C., Jeltsch, A., and Pingoud, A. (1996) Probing the indirect readout of the restriction enzyme EcoRV. Mutational analysis of contacts to the DNA backbone. *The Journal of biological chemistry* **271**, 5565-5573

35. Cole, H. A., Tabor-Godwin, J. M., and Hayes, J. J. (2010) Uracil DNA glycosylase activity on nucleosomal DNA depends on rotational orientation of targets. *The Journal of biological chemistry* **285**, 2876-2885

Appendix D

Additional information to support Chapter 5

```
              5'-Fam-CGATAGCATCCTICCTTCTCTCCAT
25I•T
                3'-GCTATCGTAGGATGGAAGAGAGGTA

              5'-Fam-CGATAGCATCCTECCTTCTCTCCAT
25εA•T
                3'-GCTATCGTAGGATGGAAGAGAGGTA
                                I
              5'-Fam-CGATAGCATCCTCCTTCTCTCCAT
25I-bulge
                3'-GCTATCGTAGGAGGAAGAGAGGTA

              5'-Fam-CGATAGCATCCTUCCTTCTCTCCAT
25U•G
                3'-GCTATCGTAGGAGGGAAGAGAGGTA

              5'-Fam-TAGCATCCTECCTCGTGTAGGTATTAGATCCGACTECCTTGTGTCCT-Fam-3'
47εA2F2
                3'-ATCGTAGGATGGAGCACATCCATAATCTAGGCTGATGGAACACAGGA-5'
```

Figure D-1. Sequence of oligonucleotides that were used in this study.
The lesion-containing strands had 5' or 5' and 3'-fluorescein attached by a 6-aminohexyl linker (Fam). Different substrates oligos were used in different assays. 25I•T was used in the active site titration measurement for active concentration determination of the purified AAG variants. 25εA•T was used in the single-turnover and multiple-turnover assays for k_{chem} and k_{cat}/K_M measurements. 25I-bulge was used to determine the active concentrations of the plasmid expressed AAG variants in yeast cell extracts. 25U•G was used to compare the relative concentrations of the yeast UDG enzyme in different yeast cell extracts. 47εA2F2 was used in the processivity assay to determine the searching efficiency of the recombinantly expressed AAG variants.
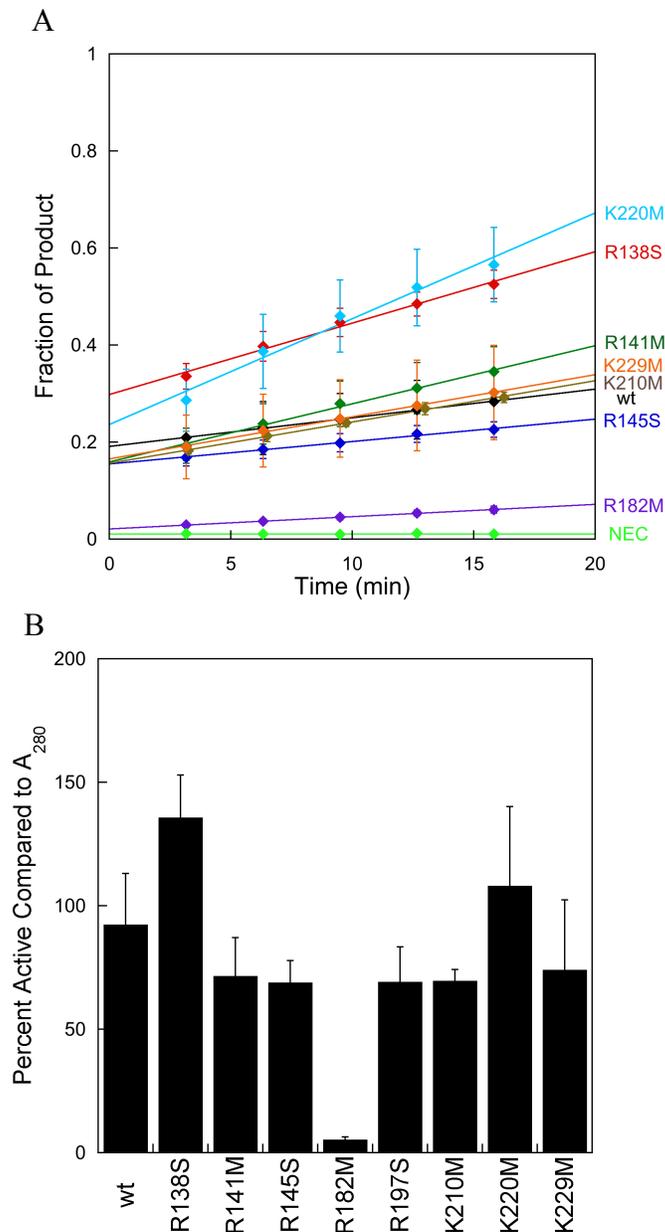
Figure D-2. Active site titration of the purified AAG variants.
A. Excision the Hx base from the 25I•T substrate DNA shows burst kinetics at low salt condition, and as a result we were able to quantify the amount of active enzyme by extrapolating the steady-state rate back to the origin. The results for the different AAG variants are shown. The concentration of the DNA substrate was 1 µM and the concentration of the AAG variants were 200 nM as determined from absorbance at 280 nm. Reactions were performed in triplicate, and the average and standard deviation are shown (mean ± S.D., n = 3). Reactions were also performed using 100 nM AAG variants and gave around half the burst phases compared to their counterpart 200 nM reactions. Those data were not shown here for the sake of clarity. The R197S mutant did not show burst kinetics due to weakened product inhibition, and its concentration was determined by comparing k_{chem} from the single-turnover reaction and V_{init}/[Enz] from the multiple-turnover reaction under conditions where the same chemical step was monitored. NEC denotes the no enzyme control reaction, which shows the nonenzymatic baseline level of product formation at approximately 1% of the substrate. The calculated active enzyme concentrations were corrected by this baseline level of product formation. B. Bar graph summary of the active percentage of the purified AAG variants (mean ± S.D., n ≥ 3). Most of the variants have active enzyme concentrations close to those determined by A_{280}, except for R182M. The corrected active concentrations of the variants were used throughout the study.
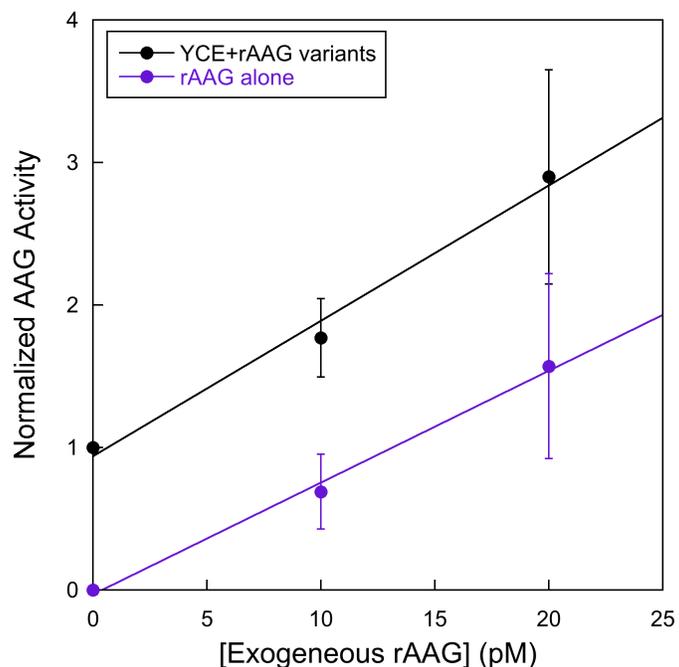
Figure D-3. Activity based AAG expression level determination in yeast whole cell extracts.
To test if glycosylase assay on the 25I-bulge substrate DNA can be used to determine the expression level of AAG variants in the yeast cells, 0, 10 and 20 pM of the purified recombinant AAG variants (rAAG) were assayed with (black circle) and without (magenta circle) their corresponding yeast cell extracts. The glycosylase activity with the yeast cell extract alone was used as 1 (the reference value) for each variant, and all other reactions were normalized to that level. Results shown are from the wt, R145S, K210M and K229M reactions (one measurement for each sample), and the average and standard deviation are shown. The activity dependence on the added recombinant AAG concentration (the slopes of the linear fits) is the same in the presence and absence of yeast cell extracts, suggesting that the AAG expression level can be calculated reliably from the activity based assay.
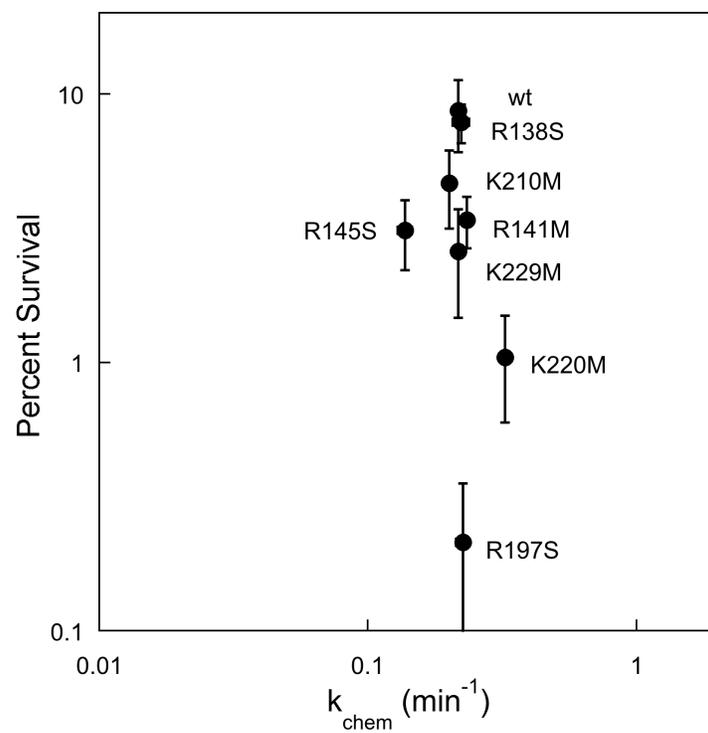
Figure D-4. Yeast cell survival versus glycosylase activity k_chem.
The yeast survival rate shows no obvious correlation with k_{chem} from the single-turnover assay of the different AAG variants.

Table D-1. AAG expression level in yeast cell extracts

All the reactions were performed at 37 °C, 50 mM sodium acetate pH 6.1, 1 mM dithiothreitol, 3 mM EDTA, 0.1 mg/ml bovine serum albumin, and 250 mM ionic strength adjusted with NaCl. The 25I-bulge DNA substrate was used in all the measurements. Values in parentheses are reference values (i.e., other values normalized to this level).

AAG	$V_{init}/[Enz]$, min^{-1} [a]	Glucose Condition			Galactose Condition		
		V_{init}, nM/Hr [b]	[AAG], pM [c]	$[AAG]_{rel}$ [d]	V_{init}, nM/min [b]	[AAG], nM [c]	$[AAG]_{rel}$ [d]
WT	2.77 ± 0.52	21 ± 5	125 ± 36	(1) ± 0.41	138 ± 8	50 ± 10	(1) ± 0.28
R138S	1.13 ± 0.06	10 ± 2	149 ± 33	1.19 ± 0.43	55 ± 8	49 ± 8	0.98 ± 0.24
R141M	0.37 ± 0.06	2.2 ± 1.2	97 ± 53	0.77 ± 0.48	12 ± 1	33 ± 6	0.66 ± 0.18
R145S	1.60 ± 0.17	13 ± 8	132 ± 85	1.05 ± 0.74	52 ± 22	33 ± 15	0.66 ± 0.32
R182M	0.51 ± 0.06		N/A			N/A	
R197S	0.03 ± 0.00		N/A		0.85 ± 0.04	33 ± 2	0.67 ± 0.14
K210M	1.96 ± 0.25	9.9 ± 1.6	84 ± 17	0.67 ± 0.24	58 ± 5	29 ± 4	0.59 ± 0.15
K220M	0.24 ± 0.07	1.2 ± 0.5	80 ± 39	0.64 ± 0.36	8 ± 1	34 ± 10	0.69 ± 0.24
K229M	1.10 ± 0.17	6.9 ± 1.3	104 ± 26	0.83 ± 0.32	34 ± 11	31 ± 11	0.62 ± 0.26

[a]$V_{init}/[Enz]$ values were measured using known concentrations of the purified AAG variants.

[b]V_{init} values were measured by using yeast cell extracts prepared from cells under the glucose or galactose growth condition (see Materials and Methods for cell extract preparation). The reported values shown in the table reflect the final results after taking into consideration of sample dilution in the glycosylase reaction system. Different units are used for the V_{init} values under the two different growth conditions out of convenience.

[c]The enzyme concentrations in yeast cell extracts under glucose and galactose growth conditions were calculated using V_{init} values of the yeast cell extracts and $V_{init}/[Enz]$ values of the recombinant proteins. Different units are used for the two growth conditions out of convenience. The yeast cells cultivated under different growth conditions were resuspended with different volumes of the disruption buffer, and these cell extracts were therefore diluted to different extents from the initial cell pellets (see Materials and Methods). When all the factors were taken into consideration, the expression level of the AAG mutants under the galactose condition could be calculated as 1000~2000-fold that of the glucose condition.

[d]The relative AAG concentrations were normalized based on the expression level of the wild type AAG under their respective growth condition.

Chapter 6
Conclusions and Future Directions

CONCLUSIONS

The goal of this work presented in this thesis was to advance our understanding of the specificity and searching mechanism of AAG for its target lesion bases and how these processes contribute to the *in vivo* function of the enzyme. My work identified the use of intersegmental transfer for efficient global searching of AAG, measured the excision efficiency and k_{cat}/K_M values of AAG on its different substrates, and highlighted the significance of efficient lesion capture for the *in vivo* function of the protein. This work emphasizes the electrostatic nature of the AAG-DNA interaction, and demonstrates how nonspecific and specific AAG-DNA binding contribute to *in vitro* and *in vivo* AAG function.

AAG is Capable of Intersegmental Transfer

AAG has been shown to search for the target lesion sites via facilitated diffusion (1), which appears to be very rapid and is characterized by frequent hops (2). However, hopping is not efficient in covering long distances of DNA as it is a thermo-driven process and has no directional bias (3). Therefore, a hopping-only protein may be trapped in a local region without sampling other regions of the DNA. My work established that AAG uses intersegmental transfer to balance local searching with global searching. The extrapolated rate of intersegmental transfer that is expected at the concentration of DNA in the nucleus is on the same order of magnitude with that of intramolecular hopping. As AAG does not appear to be capable of forming a bridging interaction between two DNA sites, our data suggest that intersegmental transfer can be achieved by microscopic dissociation and rapid capture by a new DNA strand during a transient encounter. These results have shown, for the first time, that an enzyme that initiates the base excision DNA repair pathway uses intersegmental transfer to search DNA. Furthermore, this finding has great implications for the use of intersegmental transfer as a common searching mechanism by other DNA-interacting proteins.

Excision Efficiency and Transfer Probability are Salt Dependent

The processivity assay measures correlated cleavage events of different target sites by an enzyme on a DNA molecule. The apparent parameter fraction processive from the assay includes two terms: transfer probability between two target sites and excision efficiency upon target binding. Separation of the two terms from fraction processive is needed for a better understanding of both parameters. My work showed that excision efficiency of AAG on its substrate DNA is dependent on reaction conditions, and it decreases as the salt concentration is increased. The excision efficiency measurement at low salt showed a dependence on the chase concentration, providing evidence for intersegmental transfer between the substrate and chase DNA species. The determination of transfer probability from the measured fraction processive and excision efficiency values enables a direct comparison of the searching ability of AAG with that of other enzymes.

k_{cat}/K_M of Lesion Excision by AAG is Salt Dependent

AAG has a broad substrate range, and it removes alkylated and deaminated purines such as εA and Hx (4,5). The activity of AAG on its different substrates has been measured, but it is not clear which is the preferred substrate and whether and how the preference is dependent on the reaction condition. The εA lesion opposing different bases and the Hx lesion opposing T were used for the measurements. My work demonstrated that εA is a better substrate than Hx when both are opposing T, and that AAG removes εA opposing different bases with different k_{cat}/K_M. Salt dependence studies revealed that the biggest discrimination was seen at high salt. Addition of excess competitor DNA in the direct competition assay for the measurement of the relative k_{cat}/K_M of two competing lesion sites at low salt concentration had similar effect as the increase in salt concentration in allowing for better discrimination of the competing lesion bases. These findings suggest that the relative k_{cat}/K_M results measured *in vitro* at high salt are predicative of the *in vivo* situation of lower salt and higher DNA concentration.

Correlation of *in vitro* Kinetic Parameters and *in vivo* Function

Electrostatic interactions are important for both nonspecific and specific binding of AAG to DNA, but it is not clear whether AAG uses the same binding surface for the formation of these complexes. Positively charged residues at the AAG-DNA binding interface were individually mutated and most of these were well tolerated regarding their glycosylase activity. The majority of these mutations had effects on both specific and nonspecific binding. It was demonstrated that

126

k_{cat}/K_M and fraction processive were linearly correlated, suggesting the same binding surface for specific and nonspecific interaction. These mutants were tested in a yeast functional complementation assay to measure their ability to protect cells against killing from alkylating agent exposure to DNA. When the mutants were expressed at low levels, the protection conferred by these mutants to the yeast cells was roughly correlated with their k_{cat}/K_M and fraction processive values. In contrast, overexpression of these different AAG mutants restored protection to the same level as that observed for the wild type enzyme. The observation that deficiencies in finding lesions can be compensated for by having more proteins suggests that lesions are accessible and searching limits their removal.

Collectively, these studies shed light on the searching and catalytic mechanisms of AAG on substrate DNA. These results provide a framework for understanding the molecular mechanism by which a DNA repair enzyme efficiently patrols the genome for rare sites of damage and protect cells from DNA damage.

FUTURE DIRECTIONS

AAG is one of the best-characterized human DNA glycosylases. Despite this, many questions essential for understanding the function of this enzyme remain. A brief discussion of some of the remaining questions in the field, and brief proposals on how I would approach these questions are described below.

Defining the Structure of Searching Complexes

The crystal structure of AAG has been solved in complex with the transition state mimic abasic pyrrolidine (6), the εA substrate (7) and the εC inhibitor DNA (8), but not in complex with nonspecific DNA sequences. Nonspecific protein-DNA interactions are electrostatic in nature, while specific interactions often involve more extensive bonding network (9,10). To gain a better understanding of the molecular mechanism of the searching process, it is important that the interaction of the protein with nonspecific DNA be revealed at the atomic level.

X-ray crystallographic study can be used to obtain the structure information of the nonspecific AAG-DNA complex. However, due to the relatively low binding affinity of AAG towards the nonspecific DNA and the equilibrium of multiple possible bound states of the complex, it is very difficult, if not impossible, to crystallize the complex. These problems can potentially be solved by designing specific chemical crosslinking sites on the DNA and protein

molecules. This approach has been adopted successfully for the crystallization of the MutM-nonspecific DNA complex using disulfide bond crosslinking (11). However, this strategy has the potential to introduce artifacts in the bound complex due to crosslinking and crystal packing. Alternatively, nuclear magnetic resonance can be used to investigate these interactions. This technique has already been used to study the dynamic interaction of UDG and nonspecific DNA, and it revealed millisecond time scale motions that are induced upon binding to non-target DNA (12). These studies will provide a better understanding of the similarities and differences of the specific and nonspecific bound complexes, and give insight into the searching process and the transition from the searching complex to the catalytically competent complex.

Sliding vs Hopping

Two distinct mechanisms are recognized for protein diffusion along DNA: sliding and hopping (3,13). Sliding involves continuous contact with the DNA backbone to search for contiguous sites on the same strand, while hopping uses microscopic dissociation and reassociation to search for sites on both strands. Previous studies yielded controversial results regarding the searching modes of individual proteins; some proposed that sliding is the predominant searching mode (14-16), while others concluded that sliding contributes very little to efficient searching compared to hopping (17-19). For instance, hUDG (human uracil DNA glycosylase) was suggested to use frequent hopping and short-distance sliding (4 bp) for searching (17,19), while studies on hOGG1 (human 8-oxoguanine DNA glycosylase) produced contradictory results. One study by Xie and co-workers suggested that hOGG1 uses fast sliding for searching, with the mean sliding length at 440 bp (14), whereas another study by Stivers and co-workers suggested that it uses frequent hopping and very limited strand sliding (shorter than 7 bp) for searching (18).

Single-molecule approaches can be used to further address this question. Single-molecule imaging is a rapidly developing field, and a very powerful technique for direct visualization and analysis of biological phenomenon. Improvements of the technique in recent years allow images of high spatial and temporal resolution to be collected and analyzed. By labeling the DNA substrate and the protein with donor and acceptor fluorophores, single-molecule Förster resonance energy transfer (smFRET) can be monitored in real time in a distance dependent manner. The readout of this assay system is not dependent on the catalytic activity of the protein,

but on the searching process itself and different protein-lesion binding states, and these different efficiencies will be reflective of different searching mechanisms.

Similar to the single-molecule FRET assay, the translocation of a labeled protein along a flow-stretched DNA molecule can be monitored in real time by recording the trajectory of the movement directly. By comparing the dependence of facilitated diffusion (one-dimensional diffusion constant) on salt concentration, the different searching modes can be distinguished. This technique can also be extended to the *in vivo* system of live cells, as evidenced by a recent report of the real-time observation of the searching process of a transcription factor labeled with the yellow fluorescent protein in an *E. coli* cell (20).

Kinetic assays can also be used to distinguish hopping and sliding, such as that based on the dependence of transfer probability on the distance of the different target sites (17). Kinetics can compensate for the defects of the single-molecule assays, such as the heavy modifications by fluorophores or quantum dots used for visualization and detection of the target molecules, and the unusual experimental conditions for these assays. Together, these assays will allow a better understanding of the contributions of the different searching mechanisms employed by AAG.

Searching in a Nucleosome Context

All the previous catalysis and processivity assays of AAG were performed on naked DNA substrates, but DNA molecules in the eukaryotic nucleus are organized with nucleoproteins into chromatin. The basic repeating unit of chromatin is the nucleosome, consisting of a core particle of 146 bp DNA wrapped around a histone octamer and connected by free linker DNA of 10-80 bp (21). There are two models for the function of AAG on the nucleosomal DNA: AAG may function only on transiently exposed regions of DNA, or AAG may be able to search DNA while still bound to nucleosomes. The choice of the model has important implications for the *in vivo* searching behavior of AAG. Therefore a possible future direction is to test the effect of nucleosome assembly on the searching process of AAG under the physiological condition.

Processivity assays can be performed on reconstructed lesion-containing nucleosomal substrates to measure the searching efficiency and compare the result to that on the naked DNA substrate. In nucleosome, DNA wraps around the histone proteins in 1.67 left-handed superhelical turns, therefore some of the lesion sites are brought closer together in 3-dimensional space in nucleosome than in the extended naked DNA conformation. As a result, it is possible that some of the searching processes may actually be accelerated due to frequent hopping (1,2)

and intersegmental transfer (22), assuming that AAG can go around nucleosomes without macroscopically falling off the DNA molecule. Alternatively, single-molecule experiments can be used to directly visualize the diffusion of the repair protein on the nucleosomal DNA. This latter approach has been used by Greene and co-workers to monitor two eukaryotic DNA repair proteins to diffuse along a chromatin lattice (23). These experiments will be essential for understanding the searching mechanisms of AAG in the cellular environment.

Contribution of Searching to *in vivo* Function

Facilitated diffusion has been suggested to be essential for *in vivo* cellular function based on the observed correlation of reduced *in vitro* processivity of protein mutants and their reduced biological function in *E. coli* cells (24-26). However, these studies did not measure if the mutant enzymes also had changes in other kinetic parameters such as k_{cat}/K_M, and therefore, it is not clear if the observed changes in function are due to changes in searching only. Furthermore, both studies were performed in prokaryotic cells, where DNA molecules do not adopt nucleosomal structures, and similar experiments have not been attempted in eukaryotes with hierarchical DNA organization.

An assay for testing the *in vivo* function of different AAG proteins with different defects was described in Chapter 5, but the majority of the mutants characterized there affect both nonspecific and specific interactions. If we were able to identify separation-of-function mutants, which retain the wild type k_{chem} and k_{cat}/K_M, but have different degrees of deficiency in searching efficiency, then it would be possible to investigate the role of facilitated diffusion *in vivo*. Such mutants are theoretically possible because the interface of AAG for nonspecific and specific interactions may not fully overlap due to DNA bending upon nucleotide flipping in the specific complex (6,7). The identified mutants will be tested in the *in vivo* functional complementation assays as used in Chapter 5. Alternatively, the *in vivo* function of these different separation-of-function mutants can be quantified by measuring their lesion removal activity using quantitative mass spectrometry. The HPLC-tandem MS method has been used by Eichman and co-workers to identify and quantify the alkylated base excision repair adducts released from alkylating agent challenged genomic DNA after being repaired by different 3-methyladenine DNA glycosylases, including human AAG and yeast Mag1 (27). The *in vivo* studies will establish to what extent efficient searching is required for biological function in a eukaryotic cell and allow us to extend our biochemical understanding of the enzyme to the more complex environment of the nucleus.

Searching Mechanism of Other DNA Glycosylases

      All the studies proposed here are also applicable to characterizing the searching mechanisms of other DNA glycosylases. Human cells have eleven DNA glycosylases identified so far, and they belong to different structural families. Although these different glycosylases have different substrate specificity, they all need to patrol the genome, find their respective target lesion sites, and initiate the base excision repair pathway. Therefore, efficient searching is required for the proper function of all these enzymes. Some of the DNA glycosylases have been characterized extensively for searching using different types of assays, such as UDG (17,19,28,29) and OGG1 (14,18,20), as well as AAG presented in this thesis (1,2,22), while studies on other glycosylases such as MBD4 (30) have just started. Therefore, it would be interesting to compare the searching mechanisms of proteins from different families to learn to what extent the searching processes are the same or different. Comparing the searching mechanisms employed by these different human DNA glycosylases will provide a glimpse into how nature has evolved these different enzymes for performing a very similar task.

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 chemical biology* **5**, 427-436

3.   Halford, S. E., and Marko, J. F. (2004) How do site-specific DNA-binding proteins find their targets? *Nucleic acids research* **32**, 3040-3052

4.   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. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 12869-12874

5.   Singer, B., Antoccia, A., Basu, A. K., Dosanjh, M. K., Fraenkel-Conrat, H., Gallagher, P. E., Kusmierek, J. T., Qiu, Z. H., and Rydberg, B. (1992) Both purified human 1,N6-ethenoadenine-binding protein and purified human 3-methyladenine-DNA glycosylase act on 1,N6-ethenoadenine and 3-methyladenine. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 9386-9390

6.   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

7.   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. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 13573-13578

8.   Lingaraju, G. M., Davis, C. A., Setser, J. W., Samson, L. D., and Drennan, C. L. (2011) Structural basis for the inhibition of human alkyladenine DNA glycosylase (AAG) by 3,N4-ethenocytosine-containing DNA. *The Journal of biological chemistry* **286**, 13205-13213

9.   Kaplan, N., Moore, I. K., Fondufe-Mittendorf, Y., Gossett, A. J., Tillo, D., Field, Y., LeProust, E. M., Hughes, T. R., Lieb, J. D., Widom, J., and Segal, E. (2009) The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **458**, 362-366

10.   Stawiski, E. W., Gregoret, L. M., and Mandel-Gutfreund, Y. (2003) Annotating nucleic acid-binding function based on protein structure. *Journal of molecular biology* **326**, 1065-1079

11.   Qi, Y., Nam, K., Spong, M. C., Banerjee, A., Sung, R. J., Zhang, M., Karplus, M., and Verdine, G. L. (2012) Strandwise translocation of a DNA glycosylase on undamaged DNA. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 1086-1091

132

12. Friedman, J. I., Majumdar, A., and Stivers, J. T. (2009) Nontarget DNA binding shapes the dynamic landscape for enzymatic recognition of DNA damage. *Nucleic acids research* **37**, 3493-3500

13. Halford, S. E., and Szczelkun, M. D. (2002) How to get from A to B: strategies for analysing protein motion on DNA. *European biophysics journal : EBJ* **31**, 257-267

14. 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. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 5752-5757

15. 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

16. Berkhout, B., and van Wamel, J. (1996) Accurate scanning of the BssHII endonuclease in search for its DNA cleavage site. *The Journal of biological chemistry* **271**, 1837-1840

17. Porecha, R. H., and Stivers, J. T. (2008) Uracil DNA glycosylase uses DNA hopping and short-range sliding to trap extrahelical uracils. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 10791-10796

18. Rowland, M. M., Schonhoft, J. D., McKibbin, P. L., David, S. S., and Stivers, J. T. (2014) Microscopic mechanism of DNA damage searching by hOGG1. *Nucleic acids research*

19. Schonhoft, J. D., and Stivers, J. T. (2012) Timing facilitated site transfer of an enzyme on DNA. *Nature chemical biology* **8**, 205-210

20. Elf, J., Li, G. W., and Xie, X. S. (2007) Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* **316**, 1191-1194

21. Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251-260

22. Hedglin, M., Zhang, Y., and O'Brien, P. J. (2013) Isolating contributions from intersegmental transfer to DNA searching by alkyladenine DNA glycosylase. *The Journal of biological chemistry* **288**, 24550-24559

23. Gorman, J., Plys, A. J., Visnapuu, M. L., Alani, E., and Greene, E. C. (2010) Visualizing one-dimensional diffusion of eukaryotic DNA repair factors along a chromatin lattice. *Nature structural & molecular biology* **17**, 932-938

24. 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. *Journal of molecular biology* **208**, 701-707

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. *The Journal of biological chemistry* **265**, 3424-3431

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. *The EMBO journal* **15**, 5104-5111

27.   Mullins, E. A., Rubinson, E. H., Pereira, K. N., Calcutt, M. W., Christov, P. P., and Eichman, B. F. (2013) An HPLC-tandem mass spectrometry method for simultaneous detection of alkylated base excision repair products. *Methods* **64**, 59-66

28.   Schonhoft, J. D., Kosowicz, J. G., and Stivers, J. T. (2013) DNA translocation by human uracil DNA glycosylase: role of DNA phosphate charge. *Biochemistry* **52**, 2526-2535

29.   Schonhoft, J. D., and Stivers, J. T. (2013) DNA translocation by human uracil DNA glycosylase: the case of single-stranded DNA and clustered uracils. *Biochemistry* **52**, 2536-2544

30.   Walavalkar, N. M., Cramer, J. M., Buchwald, W. A., Scarsdale, J. N., and Williams, D. C., Jr. (2014) Solution structure and intramolecular exchange of methyl-cytosine binding domain protein 4 (MBD4) on DNA suggests a mechanism to scan for mCpG/TpG mismatches. *Nucleic acids research*