

DISSECTING THE SEARCH PATHWAY OF DNA GLYCOSYLASES

By  
Joseph Daniel Schonhoff

A dissertation submitted to the Johns Hopkins University in conformity with the  
requirements for the degree of Doctor of Philosophy

Baltimore, Maryland  
July 2014

## **Abstract**

Glycosylase enzymes illustrate one of the most remarkable examples of molecular recognition known as they are able to find and remove rare mutagenic DNA bases present within the vast background of nonspecific DNA in the genome. In order to accelerate the search process and efficiently find base lesions, glycosylases and other site specific DNA binding proteins are thought to use a reduced dimensionality search process and are able to stochastically slide and hop along DNA.

Although many enzymes exhibit these properties, due to a lack of spatial and temporal resolution in current experimental approaches, mechanistic interpretations are often murky and inconsistent with other kinetic requirements in lesion recognition and catalysis. Therefore, in Chapter 2, using human Uracil DNA Glycosylase (hUNG), I have established a new approach that utilizes a small molecule to trap and time enzyme molecules that have ‘hopped’ off of the DNA providing novel quantitative insight into the lifetime and distance of hopping events as well as the speed and length of sliding on DNA.

In Chapter 3, using DNA constructs containing neutrally charged methylphosphonate linkages as well as engineered hUNG variants with enhanced electrostatic properties, a model emerges that goes against the current dogma that facilitated diffusion involves isoenergetic movement along a smooth free energy landscape allowed by electrostatic interactions with the DNA backbone. Rather, sliding is surprisingly independent of the latter perturbations and combined with previous NMR measurements suggests that movement on DNA is achieved by dynamic motions of the protein and that the sliding form of the enzyme is similar to the transition state for DNA dissociation.

In the next part of my thesis (Chapter 4), I investigate the effects of uracils present within densely spaced clusters and present within single stranded DNA. These two situations are relevant in the context of hUNG’s involvement in the generation of antibody diversity, where the processive single strand specific enzyme, Activation Induced Cytosine Deaminase (AID), converts cytosines to uracils within the Ig locus. Notably I find that hUNG is more processive on single stranded DNA and shows a previously unobserved directional preference in the presence of neighboring abasic sites.

Finally in Chapter 5, I incorporate experimental constraints for hUNG and another DNA glycosylase (hOGG1) into a complete model of facilitated diffusion using novel numerical simulations. Using this method, a data driven model consistent with the entire reaction coordinate is determined at unprecedented quantitative resolution. Further, these results lead to the surprising finding that despite these two enzymes divergence early in evolution, the search mechanism is nearly identical.

Thesis Advisor: James Stivers

Thesis Committee: Albert Lau, Greg Bowman, James Berger, Paul Miller

## **Acknowledgements**

I first want to thank my parents, who I am near positive have never told me what I should do with my life or given any opinions what-so-ever to the matter. Rather they focused on being good people themselves, imparting good values on me and fostered a sense a curiosity about the world - I have not yet met two nicer people. Also my siblings, Brad, Christy, and Jenny, for always treating their little brother well – it's comforting to know at the drop of a hat that they will always be there for me. I thank my wife and best friend, Elizabeth, who has been supportive in every possible way, and probably deserves a PhD or some type of financial compensation (don't hold your breath) for sticking with me for so long. I can honestly say my time in Baltimore has been the best time of my life so far, and she has been at the root of it all. I thank Jim for being one of the best boss's I have had, who has provided constant guidance, consistency and who has 'trained' me, in all sense of the word, to be a better scientist. I thank my lab mates for being great friends and colleagues and for always giving me a hard time about my forgetfulness, my eating habits, and whatever else obvious there is to make fun of me about. Thanks for keeping it real. I thank all of the great friends in Baltimore that I have made (I won't name any in the case I forget someone – see above) – it's been great and I'm truly sad to leave. Finally, I thank the PMB, Ranice, Kathie, and the great faculty here for everything they do. No regrets here.

## Table of Contents

### Chapter 1. Introduction....1

- 1.1 Glycosylases and the general biology of DNA repair....2
- 1.2 A brief history of the DNA search problem....5
- 1.3 Associative Transfer by Sliding....8
- 1.4 Dissociative Transfer by Hopping....10
- 1.5 Brief comment of modeling and theoretical studies of facilitated diffusion model....11
- 1.6 Ensemble biochemical approaches to decipher facilitated diffusion mechanisms....12
- 1.7 Single molecule approaches....14
- 1.8 Structural approaches....16
- 1.9 Experimental approach in this thesis....17

### Chapter 2. Timing Facilitated Site Transfer of an Enzyme on DNA: Development of the Molecular Clock Approach....34

- 2.1 Introduction....35
- 2.2 Results....38
  - 2.2.1 Timing DNA hopping and sliding pathways....38
  - 2.2.2 Intramolecular Site Transfer by hUNG....41
  - 2.2.3 The mean sliding length is four base pairs....48
  - 2.2.4 Strand preference and salt dependence of sliding....48
  - 2.2.5 hUNG slides using the DNA phosphate backbone....53
- 2.3 Discussion....54
- 2.4 Methods....61
  - 2.4.1 Oligonucleotide and protein reagents....61
  - 2.4.2 Oligonucleotide Sequences....62
  - 2.4.3 Intramolecular transfer assay....64
  - 2.4.4 Efficiency of Uracil Excision by hUNG....67
  - 2.4.5 Nonspecific DNA Binding to hUNG....68
  - 2.4.6 The molecular clock: theory....69
  - 2.4.7 Calculation of the trapping time ( $\tau_{\text{trap},0.5}$ ) and the mean hopping distance ( $\langle r_{\text{hop}} \rangle$ )....70
  - 2.4.8 Kinetic description of the intramolecular transfer probability ( $P_{\text{trans}}$ )....72
  - 2.4.9 Calculation of the energy barrier for hUNG sliding....74
- 2.5 References....76

### Chapter 3. DNA Translocation by Human Uracil DNA Glycosylase: Role of DNA Phosphate Charge....384

- 3.1 Introduction....85
- 3.2 Results....88
  - 3.2.1 Experimental Approach....88
  - 3.2.2 Calculating Site Transfer Probabilities....92

3.2.3	Effects of Neutral Methylphosphonate (M) Substitutions on Nonspecific DNA Binding....	95
3.2.4	Sliding of hUNG Does Not Require a Continuous Polyanion DNA Strand....	98
3.2.5	hUNG site transfer using physiological ion concentrations....	103
3.3	Discussion....	107
3.3.1	Different Effects of M Substitution on Nonspecific DNA Binding, DNA Translocation and Uracil Excision....	107
3.3.2	Boundary Estimates for 1D Translocation on Duplex DNA....	109
3.3.3	Search and Recognition in the Cell Nucleus....	111
3.4	Methods....	115
3.4.1	Protein and Oligonucleotide Reagents....	115
3.4.2	Oligonucleotide Sequences....	115
3.4.3	Experimental conditions....	117
3.4.4	Determination of DNA Dissociation Constants by Fluorescence Anisotropy....	121
3.4.5	Intramolecular Site Transfer Assay....	121
3.4.6	Analysis of the Site Transfer Data....	124
3.5	References....	125

#### **Chapter 4. DNA Translocation by Human Uracil DNA Glycosylase; The Case of ssDNA and Clustered Uracils....132**

4.1	Introduction....	133
4.2	Results....	136
4.2.1	Site Transfer Mechanism on single stranded DNA....	136
4.2.2	Intervening Abasic Sites Extend the Sliding Length of hUNG....	143
4.3	Discussion....	153
4.3.1	Does hUNG “Slide” on ssDNA? ....	153
4.3.2	Mechanism of Directional Bias During Reaction at Clustered Uracils....	156
4.3.3	Conclusion....	157
4.4	Methods....	160
4.4.1	Protein and Oligonucleotide Sequences....	160
4.4.2	Experimental conditions....	161
4.4.3	Intramolecular Site Transfer Assay....	161
4.4.4	Determination of the Efficiency of Uracil Excision....	164
4.4.5	Michaelis-Menten kinetics of hUNG2 reaction on ssDNA....	167
4.4.6	Calculation of the mean square distance between target sites using the worm like chain model for ssDNA....	168
4.4.7	Kinetic Modeling of the data for S19F....	168
4.5	References....	172

#### **Chapter 5. Integration of Biochemical Site Transfer Data Into A Complete Model Using Stochastic Simulations....177**

5.1	Introduction....	178
5.2	Results and Discussion....	184

5.2.1 Basic Description....	184
5.2.2 Major Findings for hOGG1 simulations....	186
5.2.3 Reconciling the ensemble and the single molecule observations....	187
5.2.4 Probability of long time scale translocation events....	188
5.2.5 Modeling the effects of salt on $D_1^{app}$ ....	189
5.2.6 Results for simulations for hUNG and comparison to hOGG1....	191
5.3 Methods....	195
5.3.1 Modeling sliding transfers: a random walk on DNA....	195
5.3.2 Modeling dissociative transfers: 3D random walk....	196
5.3.3 Fitting the experimental site spacing dependences of $P_{trans}$ and $P_{slide}$ ....	200
5.3.4 Calculation of microscopic parameters....	201
5.3.5 Calculation of $D_1^{app}$ ....	204
5.3.6 Why the ensemble binding lifetime includes all dissociative and associative transfer events....	206
5.4 References....	210

## **Chapter 6. Concluding Remarks....212**

## **Chapter 7. Appendix....214**

7.1 Preliminary data for hUNG disordered tail project....	217
7.2 Preliminary Data for Molecular Crowding Experiments....	223

## List of Figures

- Figure 1.1. The Base Excision Repair pathway.
- Figure 1.2. Facilitated diffusion mechanism.
- Figure 1.3. 1D Brownian motion model for enzyme sliding.
- Figure 1.4. Simulation of a hopping protein.
- Figure 1.5. The site transfer assay.
- Figure 1.6. Experimental designs used by Halford and coworkers to test for the use of hopping.
- Figure 2.1. The molecular clock approach for timing pathways for facilitated diffusion on DNA.
- Figure 2.2. hUNG binding to non-specific DNA
- Figure 2.3. Facilitated site transfer by human uracil DNA glycosylase (hUNG).
- Figure 2.4. Determination of the efficiency of uracil excision (E)
- Figure 2.5. Transfer efficiencies for hUNG at increasing site spacings.
- Figure 2.6. Probability of sliding ( $P_{\text{slide}}$ ) and hopping ( $P_{\text{hop}}$ ) as a function of site spacing and strand positioning of uracils.
- Figure 2.7. Site transfer dependence with increasing salt and in the context of single stranded DNA.
- Figure 2.8. Steady-state kinetic parameters for hUNG reaction with a 90 mer duplex DNA substrate containing a single uracil.
- Figure 2.9. Calculated targeting radius
- Figure 3.1. Molecular clock method
- Figure 3.2. Site transfer assay
- Figure 3.3. Substrate design and effects of methylphosphonate (M) substitutions on nonspecific DNA binding as measured using fluorescence anisotropy
- Figure 3.4. hUNG sliding and hopping is unaffected on double stranded DNA substrates containing intervening neutral methylphosphonate (M) linkages
- Figure 3.5. hUNG sliding on ssDNA is unaffected by methylphosphonate substitutions
- Figure 3.6. Steady-state kinetics and non-specific DNA binding of hUNG under physiological salt conditions
- Figure 3.7. Site transfer measurements of hUNG under approximated physiological ionic strength conditions
- Figure 3.8. DNA interactions of hUNG non-specifically bound to a destabilized thymine basepair
- Figure 3.9. Two-state model for hUNG sliding on nonspecific DNA
- Figure 3.10. Splint ligation strategy for synthesis and purification of methylphosphonate (M) containing oligonucleotides

Figure 4.1. Facilitated transfer of hUNG on ssDNA as determined by the 'molecular clock' approach

Figure 4.2. Determination of the excision efficiency ( $E$ ) for uracil cleavage by hUNG on ssDNA

Figure 4.3. Possible outcomes for site transfer measurements

Figure 4.4. Structure of the tetrahydrofuran abasic site mimic (F) and design of the uracil substrates containing intervening F residues

Figure 4.5. Site transfer measurements as determined using the method of initial rates for substrates S5F, S11F, and S19F

Figure 4.6. Single-turnover kinetic measurements of uracil cleavage from each site in the F containing DNA substrate S19F

Figure 4.7. Determination of the efficiency of uracil excision ( $E$ ) for single uracil containing substrates analogous to S19F

Figure 4.8. Facilitated site transfer properties of hUNG on uracil DNA constructs containing intervening F sites

Figure 4.9. Non-specific binding of hUNG2 to ssDNA

Figure 4.10. Steady-state reaction kinetics for a ssDNA substrate containing a single uracil

Figure 4.11. Gel images for substrates S5F, S11F, and S19F

Figure 4.12. Control experiments for determination of the excision efficiency ( $E$ ) for ssDNA

Mechanism 4.1. Site transfer mechanism and simulations used in numerical simulations.

Figure 5.1. Summary of Molecular Clock data for hOGG1.

Figure 5.2. Comparison of hOGG1 (IWXX) and hUNG (2OXM) DNA complexes.

Figure 5.3. Comparison of the searchable footprint of hUNG and hOGG1.

Figure 5.4. Facilitated diffusion model and pseudocode flowchart.

Figure 5.5. Simulated microscopic properties of site transfer for hOGG1

Figure 5.6. Quantitative Model for Facilitated DNA search of hUNG and hOGG1 derived from numerical simulations.

Figure 5.6. Modeling of associative transfer.

Figure 5.7. Effects of varying the  $r_{\text{trap}}$  and  $r_{\text{escape}}$  parameters in the Monte Carlo simulations for hOGG1.

Figure 5.8. Procedure for fitting the raw output from the MC simulations to the experimental data.

Figure 6.1. Quantitative Model for Facilitated DNA search of hUNG and hOGG1

Figure 7.1. Basic design of hUNG disordered tail constructs.

Figure 7.2. Modular expression plasmid design.

Figure 7.3. pET19b plasmid map for hUNG-mlef1 (C-terminus) and AntP-hUNG (N-terminus) expression plasmids.



Figure 7.4. Nonspecific DNA binding of hUNG-tail proteins to a 30mer fluorescein labeled nonspecific DNA duplex.

Figure 7.5. Multiple turnover AUPA reaction rate for hUNG tail variants.

Figure 7.6. Site transfer measurements for hUNG tail variants.

Figure 7.7. Effects of molecular crowding by polymers on hUNG site transfer.

Figure 7.8. Effects on the rate of association of hUNG with a specific site in the presence or absence of crowder.

# Chapter 1: Introduction

## 1.1 Glycosylases and the general biology of DNA repair

Genomic DNA of all organisms is under constant assault from a variety of chemical and environmental factors, resulting in modified and/or mismatched bases (1, 2). Although DNA as a polymer is ideal for information storage because of its extreme stability (3, 4), this stability is not sufficient enough to fully maintain the information content from generation to generation. Consequently, nearly all forms of life have evolved DNA repair machinery that actively scan the genome and remove chemically modified and/or mismatched bases (1, 2).

A central part of DNA repair is the Base Excision Repair (BER) pathway (1, 2, 5). The first step and arguably the most difficult is mediated by the DNA glycosylases, which have the exceedingly difficult job of locating and removing modified non Watson-Crick bases within the genome (6, 7). After removal, several downstream enzymes repair the DNA back to normal. These include endonucleases that cleave the abasic sites generated by glycosylases and polymerases and ligases that fill in and religate the DNA strand break (1, 2, 5) (**Figure 1.1**).

Uracil, while canonically associated with RNA, is often incorporated into DNA (8, 9). Depending on the condition and cell type, typically a single human cell contains from  $10^3$  to  $10^6$  uracils per genome (8, 10-14). While the presence of uracil is relevant from the standpoint of DNA repair, in DNA, uracil plays a critical role in a number of biological processes including class switch recombination (CSR) (15, 16), somatic hypermutation (SHM) (17-19) and viral restriction in HIV1 infection as part of the innate immune response (20).

There are two ways uracil can become incorporated into the genome. The first is by DNA polymerase, which results in a U/A base pair, as the enzyme cannot distinguish between dUTP and dTTP. The ratio of dUTP:dTTP therefore controls the levels of uracil incorporated by polymerase. Even though U/A base pairs are not mutagenic they still need to be repaired as they directly cause genomic instability due to strand breaks *via* BER removal (21-24). Alternatively, a U/G base pair can result from cytosine conversion to uracil by spontaneous or enzymatic deamination of the primary amine at the C4 position (8, 9). A great example of the latter is in somatic hypermutation, where the enzyme, Activation Induced Cytosine Deaminase or AID, converts cytosines to uracils, in a process of controlled mutagenesis for the purpose of generating sequence diversity within the Ig locus (17-19).

Several enzymes are encoded in the human genome that are able to locate and remove uracil, including hUNG, SMUG1, TDG, and MBD4, however hUNG (human Uracil DNA Glycosylase) is  $\sim 10^3$  times more active and by far the biggest contributor to uracil removal in the cell (25). The other enzymes are proposed to have more varied purposes and specificities. For example Thymine DNA Glycosylase (TDG) has a more relaxed specificity and is able to excise a number of pyrimidine bases including, 5-hydroxycytosine and 5-carboxylcytosine, and is thought to play an important role in removal of TET oxidized 5-methylcytosine bases (26). In contrast, hUNG's active site is extremely rigid and specific for uracil compared with other base modifications (6, 27).

Another highly biologically relevant DNA base modification is 8-oxoguanine (8oxoG or G<sup>o</sup>), that forms *via* reaction of guanine with reactive oxygen species (ROS) (28). Similar to uracil, a specific Glycosylase has evolved to search for and remove 8oxoG in humans (8-oxo-guanine DNA Glycosylase, hOGG1) (29). Both hUNG and hOGG1 activities are highly conserved across the 3 kingdoms of life, obviating the necessity to remove uracil and 8oxoG (30). Specifically, the sequence and structure of UNG is extremely well conserved dating all the way back to bacteria, suggesting that this enzyme's structure and mechanism diverged early in evolution. OGGs are similarly well conserved although not to the extent of UNG.

In the Stivers' lab we have embarked on mechanistic studies of hUNG (the primary focus of this thesis) and hOGG1 as these enzymes are the most well studied human representatives from the two largest glycosylase superfamilies – the UDG superfamily (hUNG) and the Helix-hairpin-Helix (HhH) superfamily (hOGG1) respectively. Comparatively, both enzymes occupy entirely different structural folds, yet have independently evolved to solve the problem of searching for and removing a rare base from the genome (7).

Locating a rare site that exists within an exceedingly low ratio to that of genomic DNA is a problem that almost all site specific DNA binding proteins and DNA modifying enzymes, including glycosylases, face. Further DNA glycosylases have the additional unique requirement of not only locating a single site but also must continuously scan and remove lesion sites from the entirety of genomic space. This problem has been termed the '*DNA search problem*', and the

generally accepted notion is that these proteins use a reduced dimensionality search process (facilitated diffusion) and are able to ‘slide’ and ‘hop’ along the DNA chain to find their target sites (**Figure 1.2**) (31, 32). However, despite nearly fifty years of experiment and theory the microscopic details of these mechanisms are not well understood.

In the case of hUNG, the enzyme must recognize and remove uracil, differing by only a single methyl group from thymine, within the diploid human genome of ~7 billion base pairs. If considering the typical concentration or copy number within the cell and the average density of U/A or U/G base pairs, for accurate replication fidelity, one molecule of hUNG must locate and remove a single uracil from approximately  $10^4$ - $10^5$  non-cognate sites. Similar considerations lead to approximately the same requirement for hOGG1<sup>1</sup>. Thus, the goal of this thesis and also of several of the ongoing projects in the Stivers’ lab is to decipher these mechanisms for not only the purpose of understanding DNA glycosylases, but also with the hope that a more general understanding of site specific DNA binding proteins and DNA modifying enzymes might be gained.

## **1.2 A brief history of the DNA search problem**

The diffusion limited rate of a bimolecular reaction is given by the Smoluchowski equation (eq. 1), where  $D$  is the diffusion coefficient of the macromolecules modeled as spheres,  $r$  is the radius and  $a$  is the fractional

---

<sup>1</sup>The typical copy number of DNA glycosylases is high,  $\sim 10^5$  copies/cell. The average number of uracil or 8oxoG sites per cell is  $\sim 10^4$ . For  $10^9$  base pair size genome, one copy of a DNA glycosylase is required to scan  $10^4$  bp where on average ~one 8oxoG or uracil might exist.

surface area available for binding (31-33). Using numbers for a typical proteins and/or oligonucleotides, the maximum rate calculated for bimolecular association is  $\sim 10^8 \text{ M}^{-1} \text{ s}^{-1}$ .

$$[1] \quad k_{a+b \rightarrow ab} = \frac{a4\pi(D_a + D_b)(r_a + r_b)}{1000}$$

Yet site specific DNA binding proteins are able to locate their target bases at remarkably fast rates that in some cases exceed this limit. Riggs *et al.* were the first to observe such phenomena in 1968, where in filter binding experiments, they observed the lac repressor and operator DNA association rate ( $7 \times 10^9 \text{ M}^{-1} \text{ s}^{-1}$ ) to be 100-1000 times greater than any previously measured protein-protein association rate ( $\sim 10^6 - 10^7 \text{ M}^{-1} \text{ s}^{-1}$ ) (34).

To account for these observations, Riggs and others initially proposed that these proteins are able to translocate along the DNA chain (34-36). The idea is that, instead of searching purely in three dimensions the search is limited to the much smaller linear space of the DNA chain resulting in the observed rate acceleration. Instead of acting as a decoy then, the adjacent non-cognate DNA is therefore a conduit or antenna leading to fast association with the specific site.

Collectively, this model of reduced dimensionality search is known as 'facilitated diffusion'. After its initial proposal by Riggs<sup>2</sup>, this model was later fully

---

<sup>2</sup>Notably Riggs first proposed this model and then a few years later recanted. Nonetheless this idea was picked up and developed by others, most notably by Peter von Hippel and his group at Oregon.

developed by Berg *et al.* in which they described several mechanisms of transfer on DNA (37). These include: i) Sliding in which the protein using thermal energy undergoes a 1D random walk or Brownian motion on DNA and ii) Hopping in which the protein microscopically dissociates and re-associates nearby on the same stretch of DNA. Further, Berg and others concluded that the most efficient mechanism is a combination of 3D diffusive hopping and 1D sliding, in which sliding serves to closely inspect local DNA segments, leading to recognition, while intermittent intramolecular hopping events limit the redundancy of sliding by translocating the protein farther along the DNA chain (**Figure 1.2**).

While this model is an elegant solution to the search problem, it is debatable whether 'hopping' and 'sliding' physically accelerate the association rate. One way to think of these effects is in terms of 'diffusion to capture' (38). Given a spherical target of radius  $a$  and protein at distance  $r$  away from the target, the probability of capture versus being lost to bulk solution reduces to the probability equal to  $a/r$ . In the case of glycosylases, the target size is a single base pair. Given the need to locate a single damaged base within  $\sim 10^4$  base pairs, if the mechanism were simply diffusion/collision the probability of locating the correct site would be  $1/10^4$  upon a single collision, or  $10^4$  fold lower than the diffusion limit. This assumes that each base is not spatially correlated, which of course is not the case as DNA is a linear polymer. Nonetheless association with the specific site for nearly all glycosylases and DNA binding proteins is limited by



the initial diffusive encounter<sup>3</sup>. Thus the more apt question of facilitated diffusion mechanisms is how is it that the transition from a non-specific 'decoy' site to the specific site is almost never rate limiting (32, 39)?

In essence, dissecting the role of facilitated diffusion is a classical kinetic problem in that the transitions of interest occur after the rate-limiting step. Despite the current acceptance of the facilitated diffusion model, the microscopic details and experimental approaches to probe the fast transitions from non-specific DNA to specific DNA are severely lacking aside from vague qualitative interpretations (e.g. 'sliding' and 'hopping'). In the next section I describe the basis of these mechanisms.

### 1.3 Associative Transfer by Sliding

'Sliding' on DNA by proteins is often viewed as one-dimensional Brownian motion along the DNA chain. Analytical forms of this model are derived based on a 1-dimensional random walk. In the model, a protein moves linearly along the DNA backbone and at a given site has an equal probability of stepping left or right and a small probability of dissociation in solution at each step (eq. 2). Working through this model results in (eq. 2) and predicts a squared dependence with the linear distance translocated, where  $n$  is the distance in bp. For example if a protein moves 10 bp on average within the model, the protein has taken 100 individual bp steps (**Figure 1.3**) (32, 39).

---

<sup>3</sup> The large increase in association rate observed for the *lac* repressor, is somewhat of an isolated case and could be due simply to acceleration of the encounter step by long range electrostatic attraction with the DNA chain and not directly facilitated diffusion (32, 39).

$$[2] \quad P(\text{step}) = \frac{k_{\text{step}}}{k_{\text{step}} + k_{\text{off}}}$$

$$[3] \quad P(\text{slide } n \text{ bp}) = \left( \frac{k_{\text{step}}}{k_{\text{step}} + k_{\text{off}}} \right)^{n^2}$$

This requires that the energy landscape along the DNA chain be almost completely flat and that the transition between bp sites have no greater barrier than  $\sim 1\text{-}2 k_bT$  (31, 43). Thus, a very delocalized type of binding is required for this model. To account for this, it was proposed by von Hippel and coworkers that sliding is mediated by nonspecific electrostatic attraction between the positively charged protein binding interface and the negatively charged DNA backbone where the DNA is viewed as an isopotential surface for which the protein can slide without friction (37, 40). This view despite direct experimental validation has propagated in the literature to this day and by and large is the accepted model.

With these considerations Schurr (41, 42) calculated the maximum possible diffusion coefficient for a protein diffusing along the DNA backbone. In his treatment, if the protein is sliding along the DNA backbone, every 10 bp the protein rotates 360 degrees around the helical pitch of the DNA, imparting hydrodynamic drag due to rotation. The effects of 'rotation coupled' sliding was shown to be significant and for a typical protein the maximum possible 'one-dimensional' diffusion coefficient by sliding is approximately  $10^7 \text{ bp}^2/\text{sec}$  (41, 42),

which differs from the translational diffusion rate in solution by approximately ~100 fold.

However, the experimentally measured values of the non-specific binding energy of a typical DNA-protein interaction ranges from 10-15  $k_bT$  (31, 43) and it is hard to imagine how DNA with numerous hydrogen bond donors and acceptors could have such a flat energy landscape. Moreover it is not clear from this model how a protein would transition to a specific site where numerous interactions are made. Indeed the formation of one or a few hydrogen bonds would preclude or drastically slow any movement by thermal energy.

To account for this, several authors have proposed a two-state conformational model (31, 43, 44), consisting of a loosely bound 'search' conformation and a more tightly bound 'recognition' conformation. Specific site recognition then is thought to occur akin to kinetic pre-selection (45). Yet, despite nearly 50 years of rich theory and experiment, experimental evidence regarding the nature of a sliding enzyme or protein remains elusive.

#### **1.4 Dissociative Transfer by Hopping**

In contrast to sliding, transfer on DNA by hopping is less complex as it can be described by the well-known properties of diffusion in solution. Once a protein has departed a bound state on the DNA, on only geometric considerations, more often than not, the protein will reassociate with the same DNA. As an example, simulations that I have performed of a spherical protein 'hopping' on DNA are shown in **Figure 1.4**, and illustrate several important features. First, most hops

would be expected to be short and cover a distance up to approximately 30 bp in linear DNA space. Once a protein moves beyond 10-20 nm from the DNA, most likely it will not come back within a reasonable amount of time. Second the free state of the enzyme only exists for a short period of time (nanoseconds to microseconds) before re-associating. Thus, by simple geometry hopping is expected to be fast and short ranging covering a distance of only a few nanometers<sup>4</sup>.

In my view, hopping can be thought of in the same way as the encounter complex between an enzyme and substrate, where upon initial collision, the two parties repeatedly bump into each other at short distances. Despite the acceptance of the sliding model, by the above considerations, I argue that hopping is an obligate kinetic pathway upon the initial encounter between an enzyme/protein and DNA purely based on shape.

### **1.5 Brief comment of modeling and theoretical studies of facilitated diffusion model**

Without doubt the formulism and development of the facilitated diffusion model of Berg, Winter and von Hippel published in 1981 (37) has been the major driver for many ideas and accepted notions in the field as described above. As an example, one particular idea that came of the analytical expressions derived in this work is that there exists an optimal sliding length. A sliding length that is too short negates the effects of facilitated diffusion and approaches that of

---

<sup>4</sup>Further details of these simulations and full stochastic simulations of the facilitated diffusion model are presented in Chapter 5.

diffusion-collision while a sliding length that is too long performs a redundant search and is slow for that reason (32, 37). Since this work, a number of modeling studies have been published (43, 46-51). Most operate on the assumption that there should exist some optimal set of parameters for facilitated diffusion, such as sliding length, sliding diffusion coefficient, optimal number of cycles of 1D and 3D diffusion or optimal time spent in 3D or 1D. A major barrier to these modeling studies is the lack of quantitative experimental methods that can measure these processes. While, optimums may well exist, it may not be necessary for them to be the same for all proteins, as the only real kinetic limitation set by the current data is that transfer from nonspecific DNA to specific DNA not be rate limiting.

## **1.6 Ensemble biochemical approaches to decipher facilitated diffusion mechanisms**

The original observations with the *lac* repressor have been complemented by a number of elegant biochemical studies. I aim to present only the most salient findings highlighting the major evidence for these mechanisms with a focus towards the work on DNA modifying enzymes, as glycosylases are the sole focus of my thesis work.

For DNA modifying enzymes, some of the first evidence came from experiments with EcoRI by Paul Modrich's group. In Jack *et al.* (52) they showed that increasing flanks (from 30 to 6000 bp) of non-specific DNA adjacent to the target sequence, increased the rate of association as well as the rate of

dissociation from the target sequence without affecting the thermodynamic stability of the specific complex. Perhaps the best evidence though for EcoRI came later, in which the authors developed a processivity assay (53). This assay utilizes a substrate DNA containing two recognition sites. By reacting the substrate with EcoRI they showed in the very first encounters of the enzyme with DNA, that more often the DNA was cleaved in two places, directly indicating the occurrence of intramolecular transfer of the enzyme between the two sites. This processivity or 'site transfer' assay with various modifications has been used frequently since it's development (including frequently in this thesis) in elucidating the translocation mechanisms of DNA modifying enzymes (54-59) (**Figure 1.5**).

In Stanford et al. the authors examined processivity of EcoRV by the site transfer assay using substrates with increasing site spacings and found that processivity did not follow the expected  $\text{bp}^2$  dependence expected for a sliding enzyme. Instead the data was more consistent with the  $a/r$  dependence for target location by three-dimensional diffusion (54). From these observations the authors concluded that EcoRV uses 'hopping' and ~50bp 'sliding' to efficiently locate its target sequence. In Porecha *et al.*, similar observations were observed for the E.Coli UNG (eUNG) and a ~10 bp target size was extracted from the data (59). One short coming of both of these studies is the indirect nature and the use of the diffusion to capture equation. The  $a/r$  relation was originally derived (38) for a spherical absorber, and it's use in the above studies negates the more complex possibility in using multiple cycles of hopping and sliding. Nonetheless, this aspect of the data strongly argued against the use of a pure sliding mechanism.

Other direct evidence for the use of a hopping mechanism came in Gowers *et al.* in which the authors used a substrate minicircle for reaction with EcoRV, that was concatenated with a larger plasmid (60). The rate of cleavage of the minicircle alone was slower than the minicircle linked to the plasmid, indicating the use of 3D mechanism on pathway to the recognition site. In another work from Halford and coworkers (55), the authors placed the target sequence for BBvCI on the opposite DNA strand and compared the processivity to identical substrates with the sites positioned on the same DNA strand. In the latter case the enzyme is required to perform at least one dissociation and reorientation to reach the target on the opposite strand facing in the reverse direction. The enzyme was shown to still be processive, indicating the use of a 3D mechanism. Similar results for the opposite DNA strand were obtained for eUNG as well and provided strong direct evidence for the use of a 3D kinetic pathway (59) (**Figure 1.6**).

## **1.7 Single molecule approaches**

In the last decade, single molecule approaches have added significantly to our understanding of DNA protein interactions. For interrogating the mechanisms of facilitated diffusion the primary experimental design has been based on tracking fluorescently labeled proteins by total internal reflection (TIRF) microscopy. In a typical experiment, DNA is stretched across an imaging platform, and labeled proteins are visualized by TIRF (61-63). A number of proteins have been subjected to these methods and include the *lac* repressor

(62, 64), RNA polymerase (65), p53 (66-68), Msh2-Msh6 (69) involved in mismatch repair, EcoRV(70), and a number of DNA glycosylases including hOGG1(71, 72), MutM (72), FpG (73) and Nei (73).

The first single molecule observation of a DNA glycosylase translocating along DNA came from Sunney Xie's group in 2006 in Blainey *et al* (72). In this report, the authors observed hOGG1 and MutM to translocate along flow stretched  $\lambda$ -DNA over micron lengths. Based on the independence of the observed diffusion coefficient with increasing ionic strength, the authors concluded that hOGG1 slid along the DNA without losing contact. Following this a number of other reports made similar observations with other glycosylases such as with FpG and Nei glycosylases (73).

While these studies have the major advantage of directly visualizing protein diffusion on DNA, numerous spatial and temporal limitations exist that make interpretations of the data difficult. For instance, most DNA binding proteins have dissociation constants with non-specific DNA in the low micromolar range and diffusion controlled association rates. By these measures, the lifetime ( $\tau_{\text{bind}} = K_D \times k_{\text{on}}$ ) is generally in the one to ten millisecond range. Even with the best CCD cameras, the resolution or frame rate of these imaging techniques is in the tens of milliseconds. Further, in order to track a single particle, more than one centroid/frame is required, therefore the real resolution of these methods is even lower and more realistically is in the range of tens-hundreds of milliseconds (62). Even with ideal imaging conditions the temporal range of the single molecule technique likely misses most binding events for these proteins. This aspect is



explored thoroughly in the context of bulk and single molecule observations for hOGG1 in Chapter 5.

### **1.8 Structural approaches**

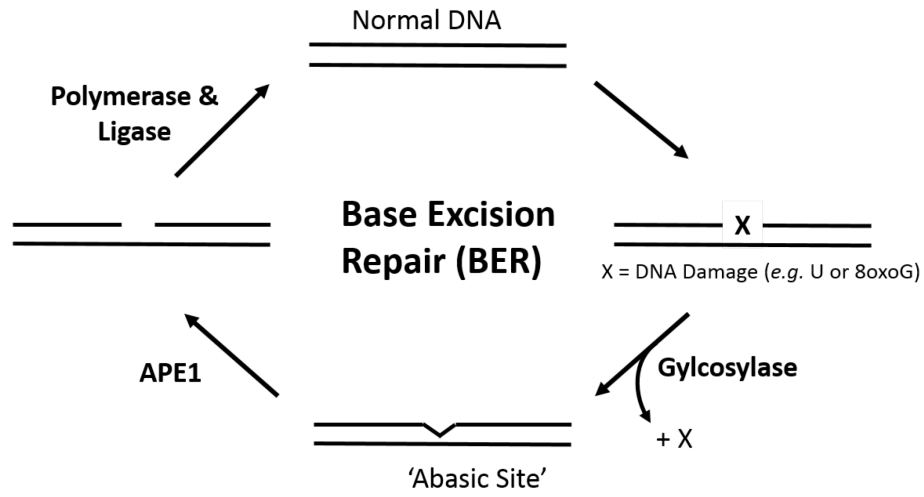
Several clever structural approaches have attempted to glean information about the 'search' complex. One of the first examples came from the Verdine group where they essentially forced hOGG1 to crystallize with a non-specific G residue in a DNA duplex by crosslinking the protein and DNA (74). In this structure the G residue is trapped in a partially flipped state. Similar structures have been solved for other proteins including hUNG and AlkA (74-77). However, in nearly all of these structures the shape of the DNA is distorted from B-DNA, suggesting that the conformation of the protein is closer to the recognition mode. Concordantly, it is not clear from these structures how a protein would transition from one base pair to the next.

NMR has offered several important insights. Friedman *et al.* showed that the free hUNG in solution has no intrinsic dynamics from the nanosecond to millisecond time scales, however intriguingly, upon the addition of nonspecific DNA, dynamic motions on the millisecond timescale are observed. These results were interpreted in a model where the enzyme transitions between an open and closed conformation (78). PRE-NMR (para magnetic relaxation) has also been applied to DNA protein interactions for this purpose. Due to the distance dependence relaxation of the PRE probes, unlike most NMR methods, PRE-NMR has the advantage of observing small high energy populations. In several

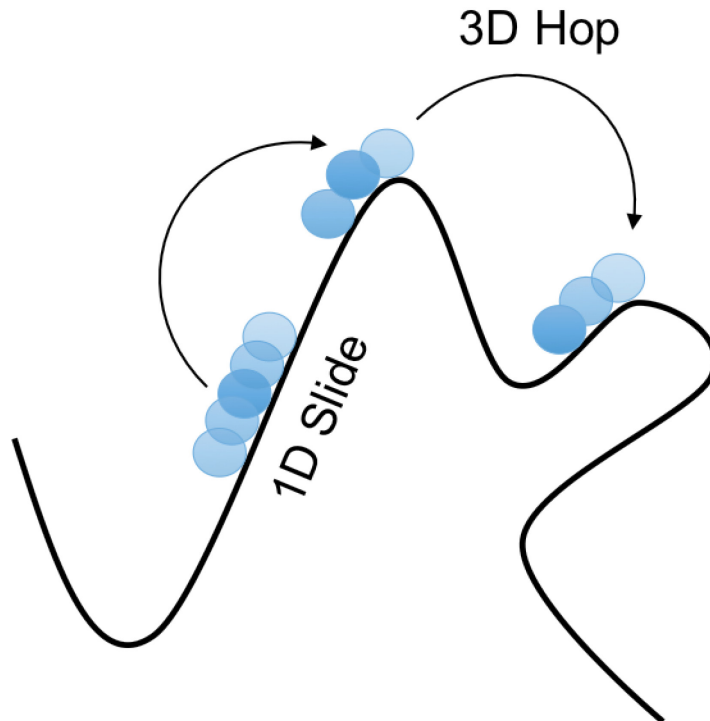
examples, mainly from Iwahara and coworkers, transcription factors have been shown to bind in a number of different conformations on non-specific DNA (79-81). These NMR methods have provided much insight, although it is not yet clear how these dynamics translate to motion on the DNA.

## **1.9 Experimental approach in this thesis**

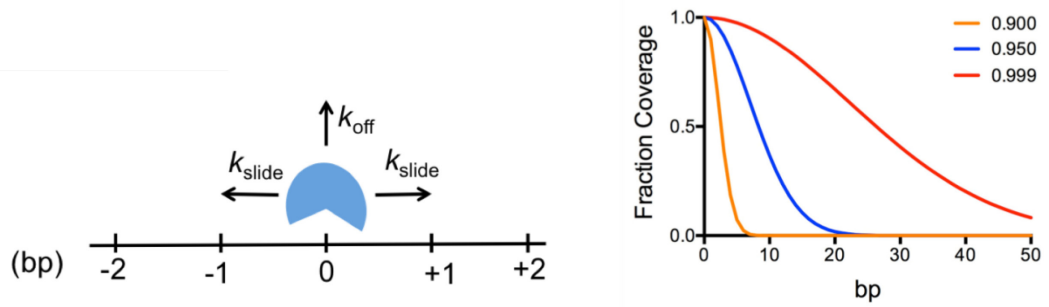
The majority of my thesis work is focused on the search mechanism of hUNG and in Chapter 2 of my thesis, I present a novel ensemble biochemical method that is able to distinguish between enzyme sliding and hopping. Notably, this method provides important quantitative insights into these mechanisms that are likely of general importance. Chapter 3 explores the role of the DNA backbone and electrostatics in enzyme processivity and then in Chapter 4 the effects of densely spaced uracil sites and translocation on single stranded DNA are investigated. In both works several unforeseen observations are made. Then finally in Chapter 5, I integrate the biochemical data collected by myself and within the Stivers' lab for hUNG and hOGG1, along with other published data including single molecule studies for hOGG1, into a complete quantitative model for damage search. In summary, despite differences in protein fold and evolutionary path, the evolved mechanism of these enzymes seems to have arrived at a near identical solution to the search problem.



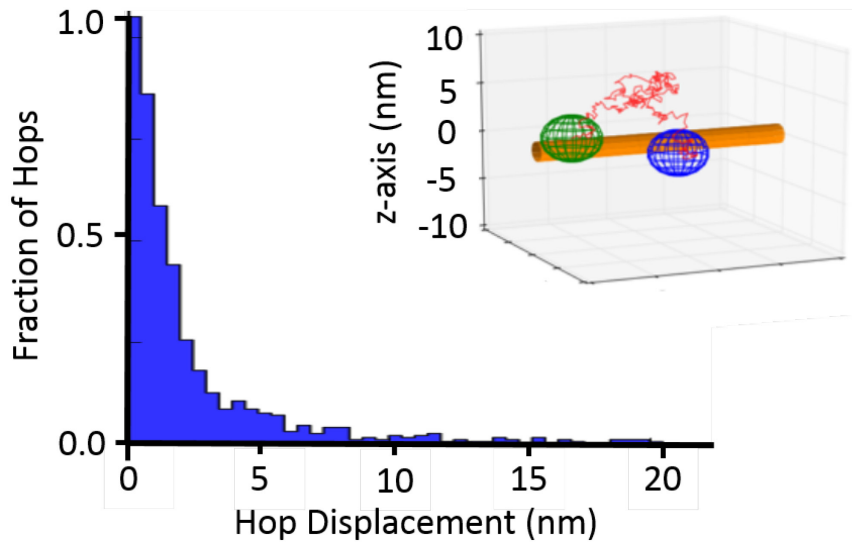
**Figure 1.1: The Base Excision Repair pathway.** Upon incorporation, a specific glycosylase finds and removes the modified base (X), resulting in an 'abasic' site. This abasic site is later processed by other enzymes to eventually repair the DNA back to its original state.



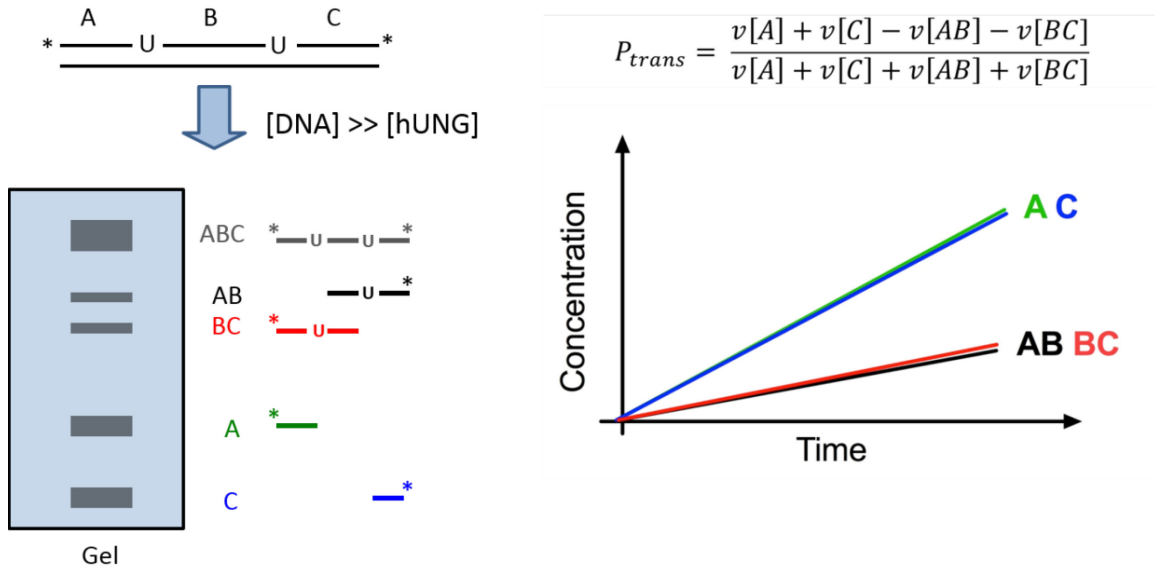
**Figure 1.2. Facilitated diffusion mechanism.** To aid in location of the target sequence, site specific DNA binding proteins have evolved to slide and hop along the DNA. Sliding is defined as movement on the DNA where the protein and DNA maintain contact. In the Hopping mechanism, the protein microscopically associates and reassociates along the DNA.



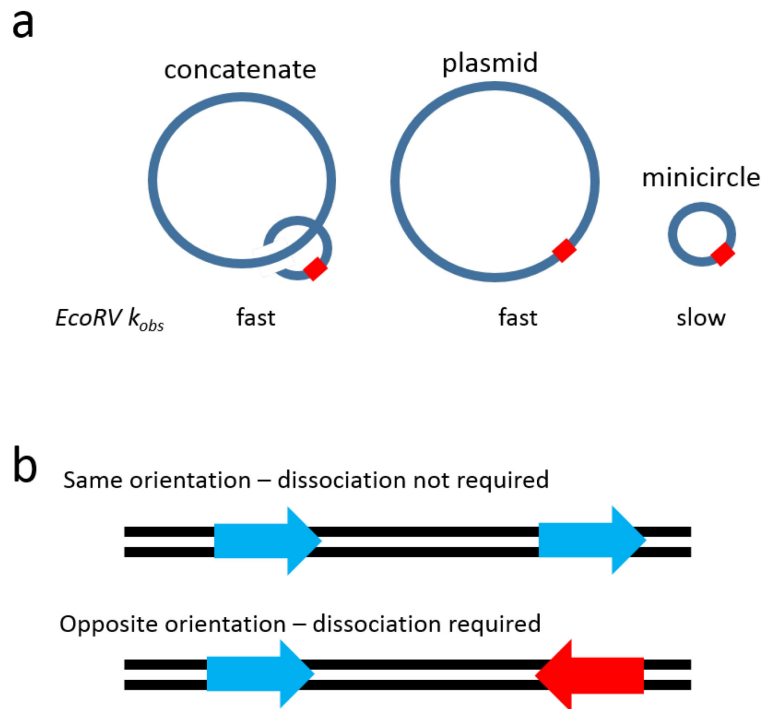
**Figure 1.3. 1D Brownian motion model for enzyme sliding.** The protein has an equal probability of stepping left or right and at each step has a finite probability of staying on the DNA. The graph on the right shows the sharp dependence with spacing for this model (see eq. 3), where the numbers and corresponding distributions indicate the probability of staying on the DNA at each step (e.g.  $k_{\text{slide}}/(k_{\text{off}} + k_{\text{slide}})$ ).



**Figure 1.4: Simulation of a hopping protein.** Protein and DNA were modeled as a sphere and cylinder, and the protein was allowed to random walk in solution upon departing the DNA cylinder. The displacement for 1000 trials is plotted as a histogram. The inset shows a single hop where the blue protein was the starting point. Based on only geometry, most hops occur on the nanosecond time scale.



**Figure 1.5 The site transfer assay.** This assay is used to measure processivity ( $P_{trans}$ ) of DNA modifying enzymes. The enzyme is reacted with a DNA substrate containing two sites where the initial rates of the single excision and double excision products are measured.



**Figure 1.6: Experimental designs used by Halford and coworkers to test for the use of hopping.** **(a)** In this experiment, the cleavage rates of a large concatenate, plasmid and minicircle for the specific site (red) were compared. The minicircle alone without the flanking non-specific DNA was the slowest, while the plasmid and concatenate were both equally fast, indicating that a 3D pathway is involved in site location. **(b)** In this experiment the recognition sites were placed on the same and opposite strand. To recognize a site on the opposite strand, the protein must dissociate and reorient at least one time.



## 1.10 References

1. Sancar A, Lindsey-Boltz LA, Ünsal-Kaçmaz K, Linn S. 2004. Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annual review of biochemistry* 73:39–85.
2. Hoeijmakers JHJ. 2009. DNA Damage, Aging, and Cancer. *N Engl J Med* 361:1475–1485.
3. Allentoft ME, Collins M, Harker D, Haile J, Oskam CL, Hale ML, Campos PF, Samaniego JA, Gilbert MTP, Willerslev E, Zhang G, Scofield RP, Holdaway RN, Bunce M. 2012. The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proceedings of the Royal Society B: Biological Sciences* 279:4724–4733.
4. Goldman N, Bertone P, Chen S, Dessimoz C, Leproust EM, Sipos B, Birney E. 2013. nature11875. *Nature* 494:77–80.
5. Seeberg E, Eide L, Bjørås M. 1995. The base excision repair pathway. *Trends Biochem. Sci.* 20:391–397.
6. Stivers JT, Jiang YL. 2003. A mechanistic perspective on the chemistry of DNA repair glycosylases. *Chem. Rev.* 103:2729–2759.
7. Friedman JI, Stivers JT. 2010. Detection of Damaged DNA Bases by DNA Glycosylase Enzymes. *Biochemistry* 49:4957–4967.
8. Olinski R, Jurgowiak M, Zaremba T. 2010. Uracil in DNA—Its biological significance. *Mutation Research-Reviews in Mutation Research* 705:239–245.
9. Krokan HE, Drabløs F, Slupphaug G. 2002. Uracil in DNA—occurrence,

- consequences and repair. *Oncogene* 21:8935–8948.
10. Mashiyama ST, Hansen CM, Roitman E, Sarmiento S, Leklem JE, Shultz TD, Ames BN. 2008. An assay for uracil in human DNA at baseline: effect of marginal vitamin B6 deficiency. *Anal. Biochem.* 372:21–31.
  11. Mosbaugh DW, Bennett SE. 1994. Uracil-excision DNA repair. [Prog Nucleic Acid Res Mol Biol. 1994] - PubMed - NCBI. Progress in nucleic acid research and ....
  12. Cross DR, Miller BJ, James SJ. 1993. A simplified HPLC method for simultaneously quantifying ribonucleotides and deoxyribonucleotides in cell extracts or frozen tissues. *Cell proliferation.*
  13. Horowitz RW, Zhang H, Schwartz EL, Ladner RD, Wadler S. 1997. Measurement of deoxyuridine triphosphate and thymidine triphosphate in the extracts of thymidylate synthase-inhibited cells using a modified DNA polymerase assay. *Biochem. Pharmacol.* 54:635–638.
  14. Aquaro S, Caliò R, Balzarini J, Bellocchi MC, Garaci E, Perno CF. 2002. Macrophages and HIV infection: therapeutical approaches toward this strategic virus reservoir. *Antiviral Res.* 55:209–225.
  15. Stavnezer J, Guikema JEJ, Schrader CE. 2008. Mechanism and regulation of class switch recombination. *Annu. Rev. Immunol.* 26:261–292.
  16. Chaudhuri J, Alt FW. 2004. Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. *Nat. Rev. Immunol.* 4:541–552.
  17. Martomo SA, Gearhart PJ. 2006. Somatic hypermutation: subverted DNA

- repair. *Curr. Opin. Immunol.* 18:243–248.
18. Rada C, Williams GT, Nilsen H, Barnes DE, Lindahl T, Neuberger MS. 2002. Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. *Curr. Biol.* 12:1748–1755.
  19. Li Z, Woo CJ, Iglesias-Ussel MD, Ronai D, Scharff MD. 2004. The generation of antibody diversity through somatic hypermutation and class switch recombination. *Genes Dev* 18:1–11.
  20. Weil AF, Ghosh D, Zhou Y, Seiple L, McMahon MA, Spivak AM, Siliciano RF, Stivers JT. 2013. Uracil DNA glycosylase initiates degradation of HIV-1 cDNA containing misincorporated dUTP and prevents viral integration. *Proc. Natl. Acad. Sci. U.S.A.* 110:E448–57.
  21. el-Hajj HH, Zhang H, Weiss B. 1988. Lethality of a dut (deoxyuridine triphosphatase) mutation in *Escherichia coli*. *Journal of Bacteriology* 170:1069–1075.
  22. Gadsden MH, McIntosh EM, Game JC, Wilson PJ, Haynes RH. 1993. dUTP pyrophosphatase is an essential enzyme in *Saccharomyces cerevisiae*. *EMBO J* 12:4425–4431.
  23. Seiple L, Jaruga P, Dizdaroglu M, Stivers JT. 2006. Linking uracil base excision repair and 5-fluorouracil toxicity in yeast. *Nucleic Acids Res.* 34:140–151.
  24. Kouzminova EA, Kuzminov A. 2004. Chromosomal fragmentation in dUTPase-deficient mutants of *Escherichia coli* and its recombinational repair. *Mol Microbiol* 51:1279–1295.

25. Parker JB, Stivers JT. 2011. Dynamics of Uracil and 5-Fluorouracil in DNA. *Biochemistry* 50:612–617.
26. Kohli RM, Zhang Y. 2013. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* 502:472–479.
27. Stivers JT. 2008. Extrahelical Damaged Base Recognition by DNA Glycosylase Enzymes. *Chem. Eur. J.* 14:786–793.
28. Neeley WL, Essigmann JM. 2006. Mechanisms of Formation, Genotoxicity, and Mutation of Guanine Oxidation Products. *Chem. Res. Toxicol.* 19:491–505.
29. David SS, O'Shea VL, Kundu S. 2007. Base-excision repair of oxidative DNA damage. *Nature* 447:941–950.
30. Denver DR. 2003. An Evolutionary Analysis of the Helix-Hairpin-Helix Superfamily of DNA Repair Glycosylases. *Molecular Biology and Evolution* 20:1603–1611.
31. Mirny L, Slutsky M, Wunderlich Z, Tafvizi A, Leith J, Kosmrlj A. 2009. How a protein searches for its site on DNA: the mechanism of facilitated diffusion. *J. Phys. A: Math. Theor.* 42:434013.
32. Halford SE, Marko JF. 2004. How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.* 32:3040–3052.
33. Smoluchowski MV. 1917. Mathematical theory of the kinetics of the coagulation of colloidal solutions. *Z Phys Chem* 92:129–168.
34. Riggs AD, Bourgeois S, Cohn M. 1970. The lac represser-operator interaction: III. Kinetic studies. *J. Mol. Biol.* 53:401–417.

35. Richter PH, Eigen M. 1974. Diffusion controlled reaction rates in spheroidal geometry. Application to repressor--operator association and membrane bound enzymes. *Biophysical chemistry* 2:255–263.
36. Berg OG, Blomberg C. 1976. Association kinetics with coupled diffusional flows. Special application to the lac repressor--operator system. *Biophysical chemistry* 4:367–381.
37. Berg OG, Winter RB, Hippel Von PH. 1981. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* 20:6929–6948.
38. Berg HC. 1993. *Random walks in biology* Princeton University Press.
39. Halford SE. 2009. An end to 40 years of mistakes in DNA-protein association kinetics? *Biochemical Society Transactions*.
40. Hippel von PH, Berg OG. 1989. Facilitated target location in biological systems. *J. Biol. Chem.* 264:675–678.
41. Schurr JM. 1979. The one-dimensional diffusion coefficient of proteins absorbed on DNA: Hydrodynamic considerations. *Biophysical chemistry* 9:413–414.
42. Bagchi B, Blainey PC, Xie XS. 2008. Diffusion Constant of a Nonspecifically Bound Protein Undergoing Curvilinear Motion along DNA. *J. Phys. Chem. B* 112:6282–6284.
43. Slutsky M, Mirny LA. 2004. Kinetics of protein-DNA interaction: facilitated target location in sequence-dependent potential. *Biophys J* 87:4021–4035.

44. Zhou H-X. 2011. Rapid search for specific sites on DNA through conformational switch of nonspecifically bound proteins. *Proc. Natl. Acad. Sci. U.S.A.* 108:8651–8656.
45. Hopfield JJ. 1974. Kinetic Proofreading: A New Mechanism for Reducing Errors in Biosynthetic Processes Requiring High Specificity.
46. Coppey M, Benichou O, Voituriez R, Moreau M. 2004. Kinetics of target site localization of a protein on DNA: a stochastic approach. *Biophys J* 87:1640–1649.
47. Kolomeisky AB, Veksler A. 2012. How to accelerate protein search on DNA: location and dissociation. *J Chem Phys* 136:125101.
48. Veksler A, Kolomeisky AB. 2013. Speed-Selectivity Paradox in the Protein Search for Targets on DNA: Is It Real or Not? *J. Phys. Chem. B.*
49. Zabet NR, Adryan B. 2012. Computational models for large-scale simulations of facilitated diffusion. *Mol BioSyst* 8:2815.
50. DeSantis MC, Li J-L, Wang YM. 2011. Protein sliding and hopping kinetics on DNA. *Phys. Rev. E* 83:021907.
51. Hu T, Shklovskii BI. 2006. How does a protein search for the specific site on DNA: The role of disorder. *Phys Rev E Stat Nonlin Soft Matter Phys* 74:021903.
52. Jack WE, Terry BJ, 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 ....*
53. Terry BJ, Jack WE, Modrich P. 1985. Facilitated diffusion during catalysis

- by EcoRI endonuclease. Nonspecific interactions in EcoRI catalysis. *J. Biol. Chem.* 260:13130–13137.
54. Stanford NP, Szczelkun MD, Marko JF, Halford SE. 2000. One- and three-dimensional pathways for proteins to reach specific DNA sites. *EMBO J.* 19:6546–6557.
  55. Gowers DM, Wilson GG, Halford SE. 2005. Measurement of the contributions of 1D and 3D pathways to the translocation of a protein along DNA. *Proc. Natl. Acad. Sci. U.S.A.* 102:15883–15888.
  56. Hedglin M, O'Brien PJ. 2008. Human Alkyladenine DNA Glycosylase Employs a Processive Search for DNA Damage. *Biochemistry* 47:11434–11445.
  57. Hedglin M, O'Brien PJ. 2010. Hopping Enables a DNA Repair Glycosylase To Search Both Strands and Bypass a Bound Protein. *ACS Chemical Biology* 5:427–436.
  58. Sidorenko VS, Zharkov DO. 2008. Correlated Cleavage of Damaged DNA by Bacterial and Human 8-Oxoguanine-DNA Glycosylases. *Biochemistry* 47:8970–8976.
  59. Porecha RH, Stivers JT. 2008. Uracil DNA glycosylase uses DNA hopping and short-range sliding to trap extrahelical uracils. *Proc. Natl. Acad. Sci. U.S.A.* 105:10791–10796.
  60. Gowers DM, Halford SE. 2003. Protein motion from non-specific to specific DNA by three-dimensional routes aided by supercoiling. *EMBO J.* 22:1410–1418.

61. Gorman J, Greene EC. 2008. Visualizing one-dimensional diffusion of proteins along DNA. *Nat. Struct. Mol. Biol.* 15:768–774.
62. Wang YM, Austin RH. 2011. Single-Molecule Imaging of LacI Diffusing Along Nonspecific DNA, pp. 9–37. *In* Williams, MC, Maher, JL (eds.), *Biophysics of DNA-Protein Interactions From Single Molecules to Biological Systems*. Springer.
63. Lee AJ, Warshaw DM, Wallace SS. 2014. Insights into the glycosylase search for damage from single-molecule fluorescence microscopy. *DNA Repair*.
64. Wang YM, Austin RH, Cox EC. 2006. Single Molecule Measurements of Repressor Protein 1D Diffusion on DNA. *Physical Review Letters* 97:048302.
65. Kabata H, Kurosawa O, Arai I, Washizu M, Margaron SA, Glass RE, Shimamoto N. 1993. Visualization of single molecules of RNA polymerase sliding along DNA. *Science* 262:1561–1563.
66. Tafvizi A, Huang F, Fersht AR, Mirny LA, van Oijen AM. 2011. A single-molecule characterization of p53 search on DNA. *Proc. Natl. Acad. Sci. U.S.A.* 108:563–568.
67. Tafvizi A, Huang F, Leith JS, Fersht AR, Mirny LA, van Oijen AM. 2008. Tumor suppressor p53 slides on DNA with low friction and high stability. *Biophys J* 95:L01–3.
68. Leith JS, Tafvizi A, Huang F, Uspal WE, Doyle PS, Fersht AR, Mirny LA, van Oijen AM. 2012. Sequence-dependent sliding kinetics of p53. *Proc.*



Natl. Acad. Sci. U.S.A.

69. Gorman J, Chowdhury A, Surtees JA, Shimada J, Reichman DR, Alani E, Greene EC. 2010. Collaborative Dynamic DNA Scanning by Nucleotide Excision Repair Proteins Investigated by Single- Molecule Imaging of Quantum-Dot-Labeled ProteinsMolecular cell.
70. Biebricher A, Wende W, Escude C, Pingoud A. 2009. Tracking of Single Quantum Dot Labeled EcoRV Sliding along DNA Manipulated by Double Optical Tweezers. *Biophys. J.*
71. Blainey PC, Luo G, Kou SC, Mangel WF, Verdine GL, Bagchi B, Xie XS. 2009. Nonspecifically bound proteins spin while diffusing along DNA. *Nat Struct Mol Biol* 16:1224–1229.
72. Blainey PC, van Oijen AM, Banerjee A, Verdine GL, Xie XS. 2006. A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA. *Proc. Natl. Acad. Sci. U.S.A.* 103:5752–5757.
73. Dunn AR, Kad NM, Nelson SR, Warshaw DM, Wallace SS. 2011. Single Qdot-labeled glycosylase molecules use a wedge amino acid to probe for lesions while scanning along DNA. *Nucleic Acids Res.*
74. Banerjee A, Yang W, Karplus M, Verdine GL. 2005. Structure of a repair enzyme interrogating undamaged DNA elucidates recognition of damaged DNA. *Nature* 434:612–618.
75. Parker JB, Bianchet MA, Krosky DJ, Friedman JI, Amzel LM, Stivers JT. 2007. Enzymatic capture of an extrahelical thymine in the search for uracil in DNA. *Nature* 449:433–437.

76. Banerjee A, Verdine GL. 2006. A nucleobase lesion remodels the interaction of its normal neighbor in a DNA glycosylase complex. *Proc. Natl. Acad. Sci. U.S.A.* 103:15020–15025.
77. Bowman BR, Lee S, Wang S, Verdine GL. 2010. Structure of *Escherichia coli* AlkA in Complex with Undamaged DNA. *J. Biol. Chem.* 285:35783–35791.
78. Friedman JI, Majumdar A, Stivers JT. 2009. Nontarget DNA binding shapes the dynamic landscape for enzymatic recognition of DNA damage. *Nucleic Acids Res.* 37:3493–3500.
79. Iwahara J, Clore GM. 2006. Detecting transient intermediates in macromolecular binding by paramagnetic NMR. *Nature* 440:1227–1230.
80. Iwahara J, Schwieters CD, Clore GM. 2004. Characterization of Nonspecific Protein–DNA Interactions by <sup>1</sup>H Paramagnetic Relaxation Enhancement. *J. Am. Chem. Soc.* 126:12800–12808.
81. Zandarashvili L, Vuzman D, Esadze A, Takayama Y, Sahu D, Levy Y, Iwahara J. 2012. Asymmetrical roles of zinc fingers in dynamic DNA-scanning process by the inducible transcription factor Egr-1. *Proc. Natl. Acad. Sci. U.S.A.* 109:E1724–32.

## Chapter 2

# Timing Facilitated Site Transfer of an Enzyme on DNA: Development of the Molecular Clock

## Approach

Reproduced in part from:

Schonhofs JD and Stivers JT (2012) Timing facilitated site transfer of an enzyme on DNA. *Nature Chem. Biol.* 8 (2), 205-210.

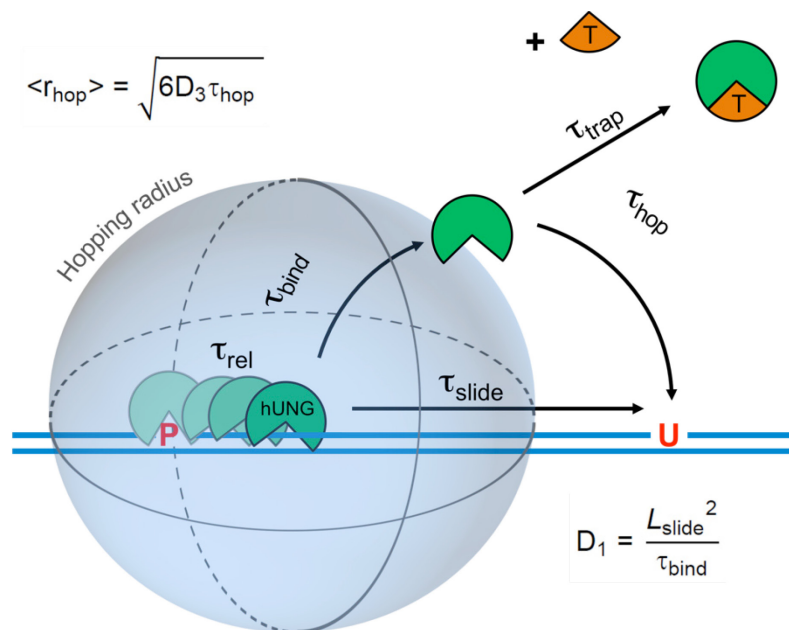
## 2.1 Introduction

Many enzymes that react with specific sites in DNA exhibit the property of facilitated diffusion, where the DNA chain is used as a conduit to accelerate site location. Despite the importance of such mechanisms in gene regulation and DNA repair, there have been few viable approaches to elucidate the microscopic process of facilitated diffusion. Here I describe a new method where a small molecule trap (uracil) is used to clock a DNA repair enzyme as it hops and slides between damaged sites in DNA. The “molecular clock” provides unprecedented information: the mean length for DNA sliding, the 1D sliding constant, the maximum hopping radius and time frame for DNA hopping events. In addition, the data establish that the DNA phosphate backbone is a sufficient requirement for DNA sliding.

With the original observation that the lac-repressor was able to locate its DNA target site faster than the Smoluchowski diffusion limit<sup>1</sup>, it has since been proposed that DNA binding proteins and enzymes enhance the efficiency of locating their individual DNA target sites by reducing the dimensionality of the search through a process known as facilitated diffusion on the DNA chain<sup>2</sup>. In recent years, numerous bulk solution and single molecule measurements have thoroughly established that facilitated diffusion is a general property of protein-DNA interactions, both *in vitro*(1-8) and *in vivo*(9, 10). Although facilitated diffusion mechanisms are now generally accepted to involve both 1D sliding and 3D hopping events on the DNA chain(1, 2, 4, 11-14), these individual pathways have never been simultaneously detected and mechanistically defined for any

protein-DNA interaction. There is good reason for this deficiency. While single molecule studies have measured the speed of various proteins sliding on DNA, and also their mean sliding lengths(12-20), limits in resolution of the method can lead to uncertainty in the mechanistic interpretation of these measurements. In particular, what may be construed as sliding may in fact involve microscopic 3D excursions of the protein from DNA, where the residence time and distances are too short to be resolved by current single molecule approaches(1, 21, 22) (see Chapter 5). Different limitations apply to ensemble kinetic approaches, where the kinetic pathways of interest are very rapid, and often occur after the slower steps of catalysis or product release, obscuring the time scales for hopping and sliding events that may be occurring.

In the initial part of my work in the Stivers lab, I developed an entirely new “molecular clock” approach to dissect sliding and hopping events of the human DNA repair enzyme uracil DNA glycosylase (hUNG) on DNA, and to determine microscopic details of its target search process that have been previously inaccessible (**Figure 2.1**). The highly optimized target search by hUNG allows it to locate and excise damaged uracil bases in either single stranded or duplex DNA, making this enzyme a master catalyst in uracil damage repair, as well as adaptive immunity(23).



**Figure 2.1. The molecular clock approach for timing pathways for facilitated diffusion on DNA.** The molecular clock begins when hUNG is released from the abasic site product (P) produced from its excision of uracil from one site in DNA and then begins its journey to the second site (U) by either hopping or sliding. The rate-limiting product release step occurs with a characteristic release time ( $\tau_{\text{rel}}$ ), after which hUNG is bound non-specifically to DNA with a lifetime  $\tau_{\text{bind}} = 1/k_{\text{off}}$ . The hopping pathway involves one or more dissociation events from non-specific DNA to generate a free enzyme molecule that is still very close to the original DNA chain from which it dissociated. The free lifetime ( $\tau_{\text{hop}}$ ) of a dissociated enzyme molecule following the hopping pathway will depend on its 3D diffusion constant ( $D_3$ ) and its distance from the DNA chain ( $r$ ). In contrast, the sliding pathway involves direct transfer of an enzyme molecule to the second site without dissociation. The sliding length ( $L_{\text{slide}}$ ) will depend on how long the enzyme remains bound to non-specific DNA ( $\tau_{\text{bind}}$ ) and its 1D diffusion constant ( $D_1$ ). The timing mechanism of the clock is provided by the concentration and diffusion constant of a small molecule trap (T) that can capture hopping, but not sliding enzyme molecules. The molecular clock can be used to calculate  $D_1$  and the hopping radius ( $\langle r_{\text{hop}} \rangle$ ), which is the distance at which half of the hopping enzyme molecules are trapped and half find the second target site successfully by hopping (see text).

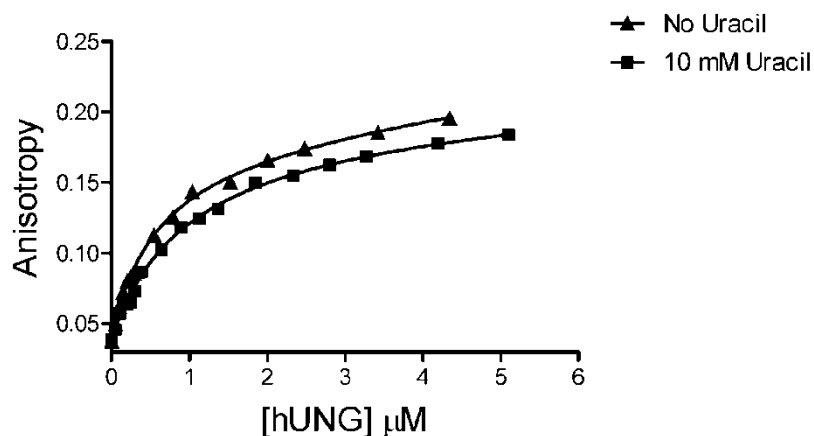
## 2.2 Results

### 2.2.1 Timing DNA hopping and sliding pathways

Molecular clocks are molecules that react with a known time constant, and therefore, allow the clocking of other molecular events that occur on a similar time scale. A classic example are the radical clocks (24), but other molecular clocks have been used successfully to measure lifetimes of reactive intermediates that occur during a variety of chemical transformations (25, 26). In the present case we wish to clock the DNA hopping and sliding pathways used by the human DNA repair enzyme, uracil DNA glycosylase (hUNG), as it departs one target site in DNA and transfers to a second nearby site in the same DNA molecule (**Figure 2.1**). To separately clock the hopping and sliding pathways for transfer, we envisioned that a small molecule active site-directed inhibitor of the enzyme would selectively trap enzyme molecules that had dissociated from the DNA during a hopping event, while leaving sliding enzyme molecules unperturbed because their active sites would be shielded from the trap by DNA. The ideal trap (T) should have the following properties: (i) The trap should not bind too tightly to the enzyme, otherwise all of the enzymes molecules will be in the inhibited ET form and the rate of reaction will be prohibitively slow. Thus, a weak binding trap is essential ( $K_i \sim 1$  mM). (ii) The trap must be sufficiently soluble such that high millimolar concentrations can be achieved. This is an essential requirement because the trap must densely populate the solution volume surrounding the dissociated enzyme such that the trap has an opportunity to diffuse to the enzyme active site during the lifetime of the hopping

event. In other words,  $1/k_{\text{trap}}[T] = \tau_{\text{trap}}$  must be comparable to the lifetime of the dissociated hopping enzyme ( $\tau_{\text{hop}}$ ) (**Figure 2.1**). (iii) Finally, the trap must not bind to DNA and interfere with the sliding pathway. In the case of hUNG, these trap criteria were met by the previously characterized active site inhibitor uracil (U) (27), which has a  $K_i = 0.3$  mM (28), solubility in water of at least 20 mM at 37 °C, and it does not interfere with binding of hUNG to nonspecific DNA (**Figure 2.2**).





**Figure 2.2: hUNG binding to non-specific DNA.** hUNG was titrated into a solution containing 50 nM nonspecific DNA duplex (nsDNA, see Methods). Dilution was considered to be negligible and the resulting anisotropy data was fitted to a one site binding model by the non-linear least squares method as described in the Methods. For the binding experiment with and without the addition of uracil  $K_D$  values were indistinguishable within error of the measurements ( $0.82 \pm 0.26 \mu\text{M}$  and  $1.06 \pm 0.13 \mu\text{M}$ , respectively). The maximal anisotropy values ( $B_{\text{max}}$ ) were  $0.181 \pm 0.044$  and  $0.160 \pm 0.012$  anisotropy units with and without the addition of uracil. Both data sets were repeated in triplicate and the values listed are the average plus or minus the standard deviation of three independent experiments.

## 2.2.2 Intramolecular Site Transfer by hUNG

To measure facilitated transfer of hUNG between two uracil sites spaced 5 to 55 bp apart on the same strand in duplex DNA (designated as S5, S6, S10, S20 and S55) in the presence and absence of the uracil trap, we employed an assay that had been previously developed to study facilitated diffusion of *Escherichia coli* UNG (eUNG) (6). This initial rate, steady-state assay quantifies the fraction of enzyme molecules that react at one uracil site and then transfer and excise uracil at the second site without dissociating to bulk solution (**Figure 2.3a, b**) (7). After appropriate sample processing (see **Methods Section 2.4**), the uracil excision events produce discrete fragments of the product DNA. Distributive single site excision events lead to the larger fragments AB and BC, while processive double excision events lead to an excess of the smaller fragments A and C (**Figure 2.3a, b**). Thus, intramolecular transfer is qualitatively indicated by an excess of the double excision products A and C relative to the single excision products AB and BC. The overall probability for intramolecular site excision ( $P_{\text{trans}}$ ) is precisely calculated by linearly extrapolating to zero time the transfer efficiency at each time point in the progress of the reaction ( $P_{\text{trans}}^{\text{obs}}$ ) using eq 2.1 (**Figure 2.3c**).

$$[1] \quad P_{\text{trans}}^{\text{obs}} = \frac{[A] + [C] - [BC] - [AB]}{[A] + [C] + [BC] + [AB]}$$

$P_{\text{trans}}$  consists of two distinct and measurable components ( $P_{\text{trans}} = E \times P_{\text{trans}}'$ ): the site transfer probability ( $P_{\text{trans}}'$ ) and the efficiency ( $E$ ) of excising the second uracil

as opposed to falling off the DNA once the second site is reached (7). Accordingly, we divide  $P_{\text{trans}}$  by the measured  $E$  value for the cleavage reaction to isolate the site transfer probability  $P_{\text{trans}}'$  (see **Methods Section 2.4** and **Figure 2.4**).

In the absence of the uracil trap,  $P_{\text{trans}}'$  for hUNG decreased with increasing site separation from  $0.75 \pm 0.05$  at 5 bp separation to  $0.27 \pm 0.08$  at 55 bp separation (**Figure 2.5**). For site spacings between 20 and 55 base pairs, human hUNG showed indistinguishable transfer probabilities as compared to previous measurements of eUNG (**Figure 2.5**)(6). Notably, for spacings greater than 20 bp,  $P_{\text{trans}}'$  for both enzymes followed a  $1/r$  dependence (where  $r$  is the site spacing in nm), as expected for a transfer mechanism involving hopping(29). Thus, it was remarkable that there was no change in the site search mechanism of these enzymes, even though the human and bacterial genomes differ in size by over 1000-fold, and the former is condensed into chromatin structures that differ significantly from bacterial DNA architectures. Apparently, the search mechanism was optimized early during evolution, and further increases in genome complexity have not provided a selection pressure for further variation.

For hUNG, we made measurements of  $P_{\text{trans}}'$  at very short site spacings from 5 to 10 bp (**Figure 2.3a, 2.3b**), and deviations from the  $1/r$  dependence expected from hopping were observed (**Figure 2.5**). This result suggested that a change to a sliding transfer pathway might be occurring at short site spacings. To explore this possibility further, we measured  $P_{\text{trans}}'$  for S5, S10 and S20 in the presence of increasing concentrations of the uracil trap, with the goal of

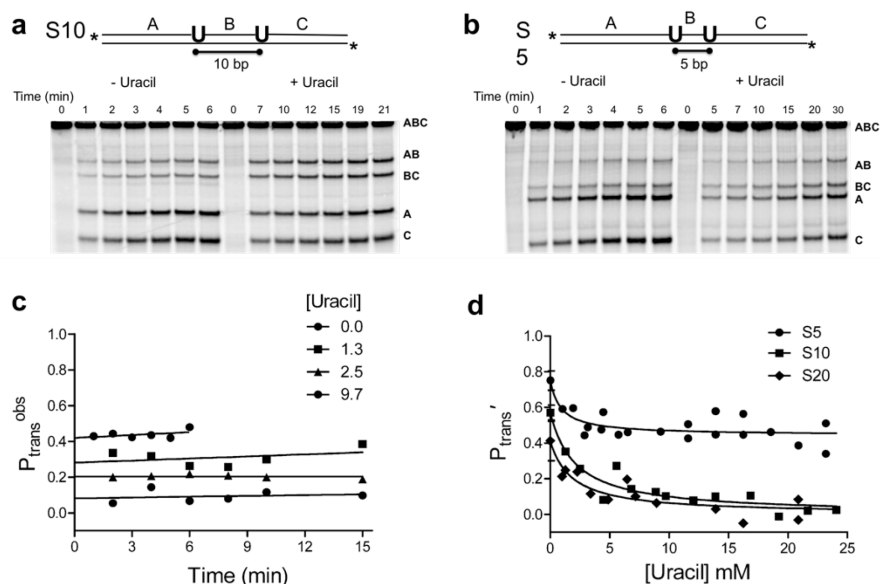
dissecting  $P_{\text{trans}}'$  into separate sliding and hopping transfer probabilities ( $P_{\text{slide}}'$  and  $P_{\text{hop}}'$ , respectively). For substrates S10 and S20, transfer was completely ablated at uracil concentrations greater than 10 mM (**Figure 2.3a, c, d**), indicating that at least one hopping event from the DNA had occurred, even with a short 10 bp spacing. In contrast, for S5 the transfer probability plateaued at  $P_{\text{trans}}' = 0.47 \pm 0.08$ , even at the highest uracil concentrations achievable (**Figure 2.3b, d**), indicating that a large fraction of the enzyme molecules that depart the first site make it to the second site 5 bp away without hopping. The curves in **Figure 2.3d** are fitted to eq 2 (derived in Section 2.4 Materials and Methods), which describes the probability of transfer via two parallel pathways (sliding or hopping) as a function of trap concentration (**Figure 2.1**).

$$[2] \quad P_{\text{trans}}' = P_{\text{slide}}' + P_{\text{off}} \left( \frac{k'_{\text{hop}}}{k'_{\text{hop}} + k_{\text{trap}}[\text{trap}]} \right)$$

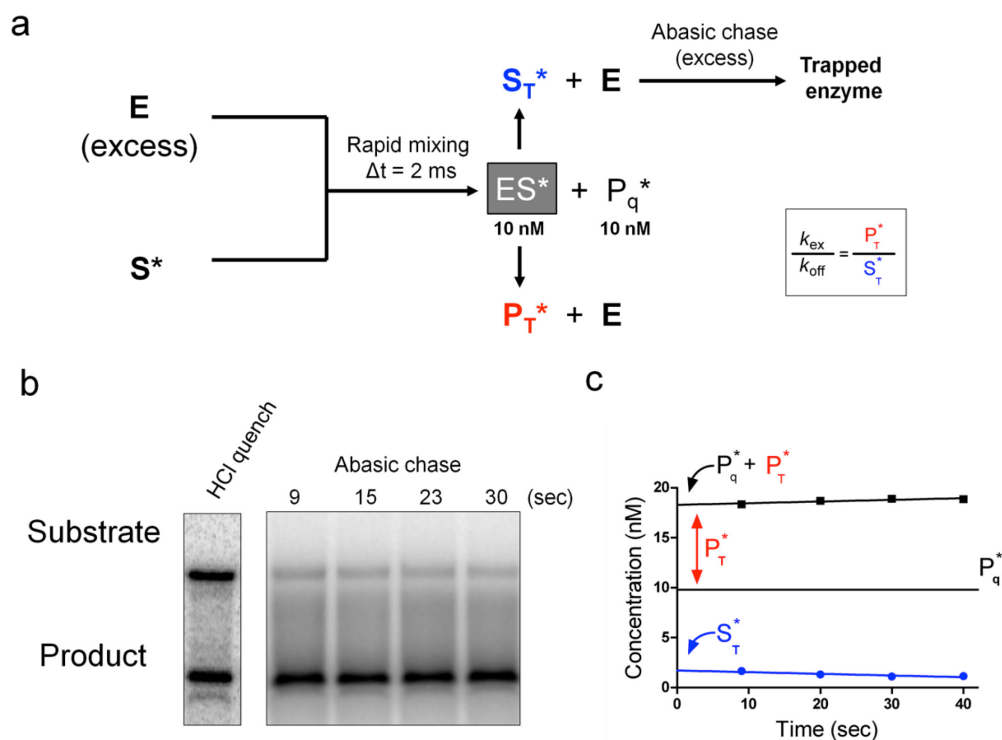
It is significant that the experimental findings match the two major expectations from this mechanism: the contribution of the hopping pathway decreases in a hyperbolic manner according to the trap concentration, and a second pathway that is refractory to trap persists at short, but not longer, site spacings.

We excluded other explanations for these results. The decreased intramolecular transfer observed in the presence of uracil does not involve uracil binding to the DNA and subsequent disruption of DNA sliding by the enzyme, because binding of hUNG to nonspecific DNA is unaffected by the presence of

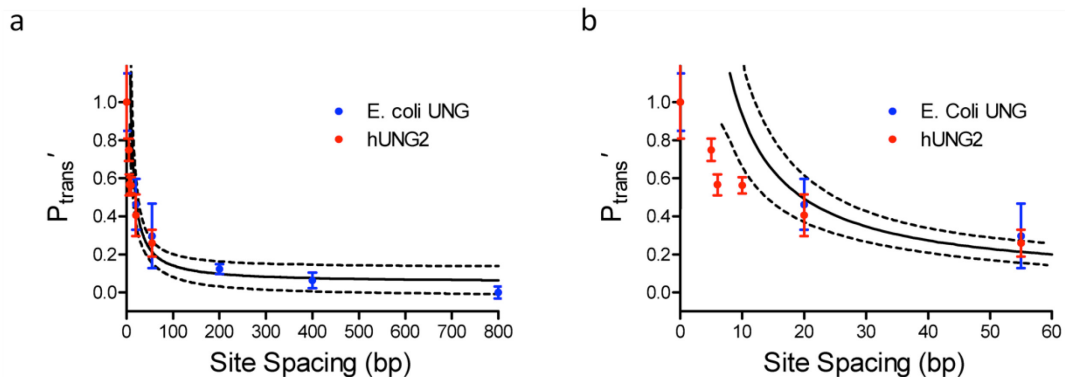
uracil (**Figure 2.2**). In addition, if uracil binding to DNA were important, this binding should disrupt sliding in all the DNA constructs, yet a non-zero plateau value for  $P_{\text{trans}}$  persists for S5 between 5 and 20 mM [uracil] (**Figure 2.3d**). In addition, we show below that transfer between two uracils placed five base pairs apart on opposite strands of the DNA is largely negated by the uracil trap, indicating that the plateau value observed with S5 is due to a distinct transfer pathway (likely strand specific sliding, see below). In addition, we excluded the possibility that Group I or II salts in the uracil stock might have disrupted transfer by ICP-MS analysis [all ions < 1:1000 (mol/mol) compared to uracil], and  $^1\text{H}$  NMR confirmed the purity of the uracil stock with respect to organic contaminants.



**Figure 2.3. Facilitated site transfer by human uracil DNA glycosylase (hUNG).** In all facilitated transfer assays the hUNG concentration is 5-20 pM and the DNA substrate concentration is 40 nM. **(a)** Facilitated site transfer of hUNG between two uracil sites on the same DNA strand separated by 10 bps (S10). Reactions in the absence and presence of 10 mM uracil are shown. Facilitated transfer is qualitatively indicated by an excess of double excision fragments (A and C) relative single site excision products (AB and BC). **(b)** Facilitated site transfer by hUNG between sites separated by 5 bp (S5) in the absence and presence of 10 mM uracil. **(c)** Observed probability of site transfer ( $P_{trans}^{obs}$ ) as a function of time and uracil concentration for the substrate with a 10 bp site separation. Linear extrapolation to the y axis provides the true transfer probability at zero time ( $P_{trans}$ ). **(d)**  $P_{trans}'$  as a function of increasing uracil for substrates with 5, 10 and 20 bp site spacings. Each data point represents an individual experiment as in panels a and b and the prime notation indicates correction for the efficiency of excision (see text). The non-linear least squares fits in (d) use a kinetic partitioning model that relates the dependence of the total transfer probability ( $P_{trans}' = P_{slide}' + P_{hop}'$ ) to the uracil trap concentration (see Section 2.4 Materials and Methods). Error bars represent the mean  $\pm$  1 s.d. of at least three independent trials.



**Figure 2.4. Determination of the efficiency of uracil excision (E)** (for details see Methods). **(a)** Briefly, hUNG was rapidly mixed with DNA under single turnover conditions (20 nM DNA, 1  $\mu$ M hUNG). After an aging time of 2 ms the reaction was quenched directly either by HCl, or alternatively, chased with 60  $\mu$ M of the abasic containing chDNA, followed by manual quenching with acid over times ranging between 9 and 40 s post mixing. **(b)** Denaturing gels showing separation of the 5'  $^{32}$ P-labeled product and substrate after direct acid quench at 2 ms, or during the chase period. **(c)** The amount of product ( $P_q^* + P_T^*$ ) and substrate ( $S_T^*$ ) remaining after the chase period is depicted by squares and circles respectively. In order to calculate the ratio of  $k_{ex}$  to  $k_{off}$  and in turn the excision efficiency ( $E = k_{ex} / (k_{off} + k_{ex}) = P_T^* / (P_T^* + S_T^*)$ ), it is necessary to correct for the amount of product already formed within the 2 ms aging time ( $P_q^*$ ) as determined from the acid quenched samples. Note that since the abasic chase DNA is not 100% efficient, and the possibility exists for the slow unbinding and rebinding of substrate DNA, the value of  $k_{ex}$  and  $k_{off}$  was calculated using the linearly extrapolated values at zero time of product ( $P_q^* + P_T^*$ ) and substrate ( $S_T^*$ ). The experiment was repeated five times and E was determined to be  $81 \pm 16\%$ , where the error represents one standard deviation.



**Figure 2.5. Transfer efficiencies for hUNG at increasing site spacings.** For comparison are shown the previously measured transfer efficiencies of *E. coli* UNG (blue) (6). The fitted curve uses a hopping model for intramolecular transfer, where the probability of encountering the second site scales with the size of the site (a) and the inverse of the distance ( $r$ ) to the site ( $P_{hop}' = E \times a/r$ ) (29). The transfer efficiency at zero spacing ( $E$ ) is determined by the efficiency of uracil excision by hUNG once it encounters a uracil site (**Figure 2.4**). The dotted lines display the 95% confidence interval of the fit. For both *E. coli* UNG and hUNG error bars represent one standard deviation and experiments were independently repeated at least three times. Data and fits presented in panels **a** and **b** are the same although the deviation from the  $1/r$  hopping model at short spacings is highlighted in panel **b**, suggesting a change in pathway from hopping to sliding at small site spacings.



### 2.2.3 The mean sliding length is four base pairs

If hUNG uses 1D sliding over distances less than 10 bp, a plot of  $P_{\text{slide}'}$  against uracil spacing (in bp) will follow an  $S^N$  dependence, where  $N = \text{bp}^2$  is the number of steps taken in the 1D walk (29, 30). To test these expectations, we plotted the sliding contribution ( $P_{\text{slide}'}$ ) to the overall transfer probability against bp spacing for S5 through S55 (**Figure 2.6a**). The data corresponded well to the 1D random walk model, as the probability of transfer to the second site decreased sharply with increased spacing, with a mean sliding length of  $4.2 \pm 0.8$  bp. A similar plot of the hopping probability ( $P_{\text{hop}'}$ ) against bp spacing for S5 through S55 showed a broad site spacing dependence as expected for a pathway involving hopping events (**Figure 2.6b**). Notably, this plot shows a downward deviation at spacings less than 10 bp (dotted line, **Figure 2.6b**), indicating that a significant fraction of the transfer events occur by the sliding pathway for spacings less than 10 bp.

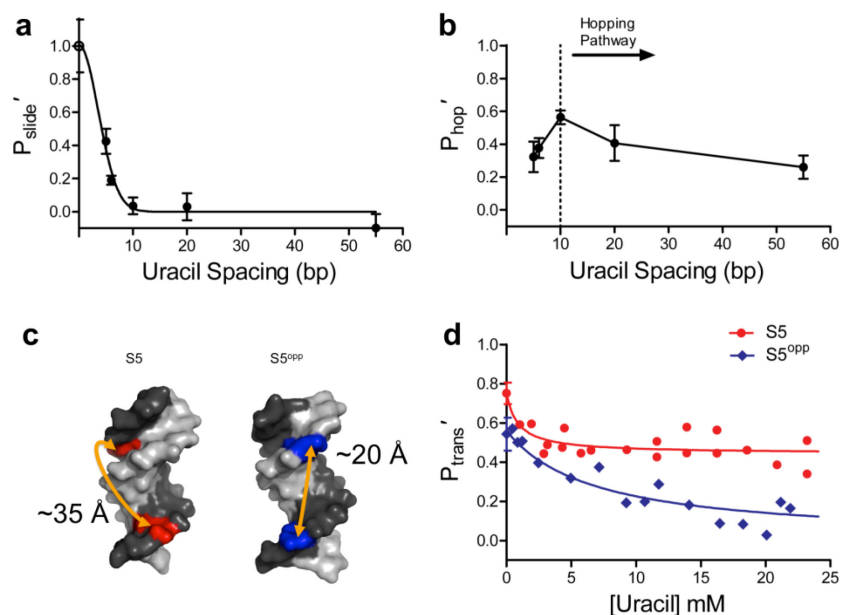
### 2.2.4 Strand preference and salt dependence of sliding

We reasoned that if the untrappable component to  $P_{\text{trans}'}$  was truly a result of hUNG sliding on a DNA strand or groove, this component ( $P_{\text{slide}'}$ ) would disappear if the two uracil sites were placed on opposite strands of the DNA(1). To test this prediction, transfer efficiencies were measured at increasing concentrations of uracil using a substrate with uracil sites spaced 5 bp apart, but located on opposite strands of the duplex ( $S5^{\text{opp}}$ , **Figure 2.6c**). Without the addition of uracil,  $P_{\text{trans}'}$  for  $S5^{\text{opp}}$  ( $0.54 \pm 0.08$ ) was found to be lower than S5

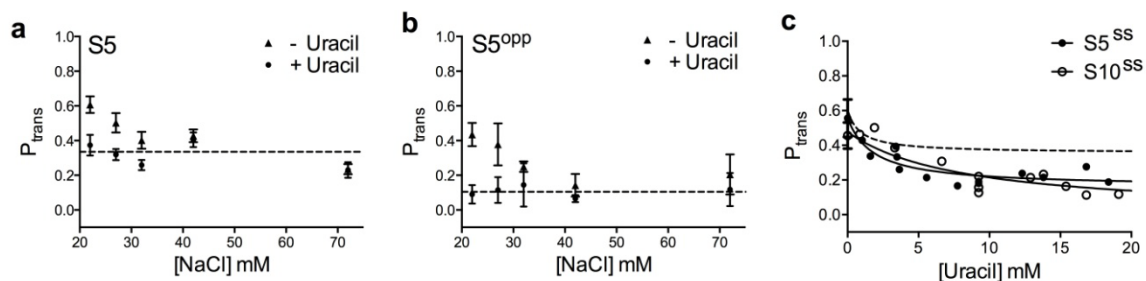
( $0.75 \pm 0.05$ ). Moreover,  $S5^{opp}$  differed markedly in its response to uracil, with  $P_{trans}$  falling to nearly zero as the uracil concentration was increased (**Figure 2.6d**). The distinct responses of these substrates to the uracil trap establishes that hUNG reaches a uracil target on the same strand using both hopping and sliding pathways, but that the enzyme must always complete at least one dissociation/reorientation event in order to transfer to a uracil site on the opposite strand even when they are only 5 bp apart(1).

Another expectation is that a sliding enzyme would be insensitive to the ionic strength of the bulk solution<sup>2</sup>, while a hopping enzyme would allow ions to recondense around the DNA chain after dissociation, thereby reducing the transfer efficiency of hopping. We tested the salt concentration dependence of  $P_{trans}$  for both S5 and  $S5^{opp}$  in the presence and absence of 10 mM uracil (**Figure 2.7a, b**). For S5 in the absence of uracil,  $P_{trans}$  diminished by about 50% over the NaCl concentration range ~20-35 mM, and then plateaus at  $P_{trans} \sim 0.3$  over the range 35 – 75 mM NaCl. (We do not correct  $P_{trans}$  for the uracil excision efficiency (E) in these experiments because E has not been measured at each salt concentration.) For S5 in the presence of uracil,  $P_{trans} \sim 0.3$  over the entire range of salt concentrations investigated, which is equivalent to the plateau value that is reached in the absence of uracil. For  $S5^{opp}$ ,  $P_{trans}$  approaches zero ( $0.13 \pm 0.09$ , dotted line **Figure 2.7b**) as the salt concentration increases, which is the expected transfer efficiency at high concentrations of uracil (see **Figure 2.6d**). The distinct behaviors of S5 and  $S5^{opp}$  towards added salt and uracil, support the

proposal that a sliding (salt resistant) pathway operates with S5, but in the case of S5<sup>opp</sup> a hopping event must occur to reach the opposite strand.



**Figure 2.6. Probability of sliding ( $P_{slide}'$ ) and hopping ( $P_{hop}'$ ) as a function of site spacing and strand positioning of uracils.** The individual contributions of the sliding and hopping pathways at each site spacing were determined by measuring the limiting  $P_{trans}'$  at high concentrations of uracil (sliding only), and the difference between the  $P_{trans}$  at zero and high concentrations of uracil (hopping only). **(a)** Site spacing dependence of  $P_{slide}'$ . The non-linear least squares fit to the function  $P_{slide}' = S^N$  is shown, where  $S$  is a constant and  $N$  is the site spacing in bp squared (29, 30). The maximal efficiency of a sliding enzyme at zero spacing is 1.0 and is equivalent to the efficiency of excision (open circle). The mean sliding length ( $L_{slide}$ ) was calculated from the spacing at  $P_{slide}' = 0.5$  and was determined to be  $4.2 \pm 0.8$  bp where the error represents the 95% confidence interval of the least squares fit. **(b)** Site spacing dependence of  $P_{hop}'$ . The dotted line indicates the approximate base pair separation where a change from a predominantly hopping to sliding pathway for facilitated transfer occurs. **(c)** Structural models of B-DNA helices illustrating the position and approximate distances of two uracils positioned on the same (S5) or opposite DNA strand (S5<sup>opp</sup>). Images were made in Pymol™ using PDB I.D. 2L8Q. **(d)** Transfer efficiencies for S5 and S5<sup>opp</sup> as a function of uracil concentration. Error bars represent the mean  $\pm$  1 s.d. of at least three independent trials.



**Figure 2.7. Site transfer dependence with increasing salt and in the context of single stranded DNA.** No correction for the cleavage efficiency ( $E$ ) was made in these transfer measurements. **(a & b)** Transfer probabilities for substrate S5 and S5<sup>opp</sup> with and without the addition of 10 mM uracil as a function of increasing [NaCl]. For both substrates with increasing NaCl, the transfer probabilities plateau at the level observed in the presence of uracil. The average salt-independent values for  $P_{trans}$  in the presence of uracil are indicated by the dotted lines in panels (a) & (b). For both (a) & (b) in the presence of uracil, the  $P_{trans}$  value at 22 mM NaCl is the plateau average with increasing uracil (Figure 2d and Figure 4b). **(c)** hUNG slides on single stranded DNA. Transfer probabilities were measured for a 90mer substrate with uracils positioned 5 and 10 bp apart (S5<sup>SS</sup>, S10<sup>SS</sup>). The plateau in  $P_{trans}$  for S5<sup>SS</sup> and S10<sup>SS</sup> equals  $0.22 \pm 0.04$  and  $0.16 \pm 0.05$  respectively. For comparison, the dashed line is the theoretical fit to the data for the duplex form S5 (also with no correction for the cleavage efficiency). Error bars represent the mean  $\pm 1$  s.d. of at least three independent trials.

### 2.2.5 hUNG slides using the DNA phosphate backbone

To gain insight into whether DNA grooves or the phosphate backbone were essential for hUNG sliding, transfer efficiencies were measured on a single stranded DNA substrate. We designed 90 mer sequences containing uracil sites positioned 5 and 10 bp apart ( $S5^{ss}$  and  $S10^{ss}$ ) that had no potential for secondary structure formation that might give rise to differential catalytic efficiency at the uracil sites (see Methods Section 5.4).

For  $S5^{ss}$  and  $S10^{ss}$  in the absence of uracil, the measured values for  $P_{trans}$  ( $0.56 \pm 0.11$  and  $0.46 \pm 0.08$ ) were near that of the respective duplex DNA forms (**Figure 2.7c**). In contrast to the behavior of duplex DNA,  $S5^{ss}$  and  $S10^{ss}$  both showed a plateau in  $P_{trans}$  ( $0.22 \pm 0.04$  and  $0.16 \pm 0.05$ ) in response to increasing uracil, indicating that the mean sliding length is longer for single stranded DNA. It is possible that the additional flexibility of single stranded DNA allows enhanced tracking of the phosphate backbone compared to the more rigid duplex DNA, although other mechanisms are certainly possible. The observation of sliding on this single stranded substrate establishes that the major or minor groove of duplex DNA are not essential for sliding and supports a mechanism involving primarily phosphate backbone tracking. Such a mechanism is entirely consistent with the observation that hUNG preferentially locates uracils by sliding on the same DNA strand in duplex DNA (**Figure 2.6d**).

## 2.3: Discussion

The molecular clock allows measurement of several key microscopic aspects of DNA sliding and hopping that have been previously inaccessible. First, the 1D diffusion constant ( $D_1$ ) may be calculated using eq 3(30-32), using the mean sliding length of 4 bp (**Figure 2.6a**), and the binding lifetime of hUNG to non-target DNA ( $\tau_{bind} = 1/K_D k_{on} = 3$  ms) (see Methods, and Figure S1 and S5).

$$[3] \quad D_1 = \frac{L_{slide}^2}{\tau_{bind}}$$

This approach yields  $D_1 = 6.0 \times 10^3 \text{ bp}^2/\text{s} = 6.9 \times 10^{-4} \text{ } \mu\text{m}^2/\text{s}$  which is several orders of magnitude less than the theoretical maximum for sliding ( $\sim 1 \times 10^7 \text{ bp}^2/\text{s}$  or  $\sim 1 \text{ } \mu\text{m}^2/\text{s}$ , Section 2.4 Methods)(31, 33, 34). As compared to previous single molecule measurements of the human 8-oxoG glycosylase enzyme (hOGG1)(16), an enzyme that recognizes 8-oxoguanine in DNA(35), the 1D diffusion process of hUNG has a higher activation barrier ( $\sim 7 k_B T$  as compared to  $\sim k_B T$  for hOgg1, Section 2.4 Materials and Methods), and a much shorter mean sliding length (4 bp versus versus  $\sim 400$  bp for hOGG1). A more detailed analysis between hOGG1 and hUNG is presented in Chapter 5. Nevertheless, the short lifetime of hUNG on DNA allows it to dissociate often and then rapidly diffuse short distances by 3D diffusion, which efficiently leads to reassociation with the nearby DNA chain. This iterative use of slower short range sliding and rapid 3D hopping is highly desirable because it allows rapid and comprehensive searching of both DNA strands, and this mechanism can operate even in the presence of bound

proteins. Accordingly, these results nicely account for the observation that human alkyladenine DNA glycosylase (AAG) can efficiently bypass bound protein obstacles during transfer between two target sites separated by only 50 base pairs(4).

A perplexing aspect of barrierless Brownian sliding (such as observed with hOGG1)(16, 17) is that it is difficult to envision how such an extraordinarily rapid sliding process is arrested at a damaged site such that enough time is spent at the site to initiate the first steps of site specific recognition. For both hOGG1 and hUNG, these recognition events involve DNA bending, damaged base extrusion and conformational changes in these enzymes that occur on the ms time scale (1-20 ms for hOGG1(36, 37)), time frames that are much greater than the calculated base pair residence times of  $\sim 10^{-7}$  s per bp for barrierless sliding(16, 17). In contrast, short range sliding over 4 bp within a bound lifetime of 3 ms gives a mean base pair residence time of about 0.8 ms for hUNG at 37°C, which is compatible with the time scale of uracil base pair opening motions that have been measured, as well as the dynamic motions of the enzyme that have been observed using NMR relaxation methods(38-40).

These results are the first to directly observe a strand specificity for sliding, and a requirement for phosphate backbone tracking. Although single molecule studies have suggested that proteins use rotation coupled movement around the DNA helix consistent with the Schurr model for DNA sliding(33), current single-molecule methods lack the temporal and spatial resolution to exclude other explanations for these observations, such as microscopic hops in combination



with strand sliding(1, 21, 22). Indeed, even with the high resolution of the studies presented here (~5 bp or 1.7 nm), we cannot exclude the possibility that the efficient location of a uracil site on the opposite strand by hUNG might involve a more complicated trajectory such as tumbling or rolling on the DNA helix(41).

Although hopping is frequently invoked in discussions of target site location mechanisms of enzymes, there have been no direct measurements that define the distances or time scales involved in hopping events. The data present two independent methods to estimate these parameters. The first qualitative approach we call the “DNA ruler”. If the mean hopping distance is defined as the characteristic site spacing where there is an equal probability that a dissociated enzyme molecule will be lost to bulk solution, or alternatively, hop back onto the DNA and locate the second site, this condition is approximately met at a 10 bp site spacing for hUNG (**Figure 2.6b**)(30). It is important to point out that all of the transfer events that occur over a 10 bp spacing involve at least one hopping event and *no direct sliding transfers* (i.e 10 bp is the shortest spacing where all transfer events are trappable by uracil). Thus, 10 bp (3.4 nm) provides a reasonable boundary to the hopping distance. Using this value, and the calculated diffusion constant for hUNG (Section 2.4 Materials and Methods), the upper limit hopping time  $\tau_{\text{hop}} \sim 20$  ns can be estimated from eq 4).

[4] 
$$\tau_{\text{hop}} = \frac{\langle r_{\text{hop}} \rangle^2}{6D_3}$$

Such short range hopping events are entirely expected because hUNG shows a preference in locating a site 10 bp away over a site located 20 bp away. If the hopping distance away from the DNA were very large compared to these site spacings, then no preference for transfer between these sites would be observed (i.e. for very long hops each site would be roughly equidistant from the dissociated enzyme). Therefore, the data are only consistent with short, fast microscopic hops which would serve to accelerate movement of the enzyme to new regions of the DNA chain. This qualitative view of short, fast microscopic hopping is quantitatively supported by the molecular clock approach that follows.

What is the characteristic distance that a hopping hUNG molecule diffuses before being trapped by uracil? This parameter, which we call the hopping radius (**Figure 2.1**), may be estimated using the concentration of uracil trap at which 50% of the hopping transfers are eliminated ( $[\text{uracil}]_{0.5} = 2.1 \text{ mM}$ , **Figure 2.3d**). Under this condition, the time constant for trapping ( $\tau_{\text{trap}, 0.5} = 1/[k_{\text{trap}}]_{0.5}$ ) is equal to the characteristic lifetime of the free enzyme before it hops back onto DNA ( $\tau_{\text{hop}}$ ), thereby providing a molecular clock for timing hopping events (**Figure 2.1**). The value  $\tau_{\text{trap}, 0.5} = 60 \text{ ns}$  may be directly calculated from the Smoluchowski diffusion equation, using the measured diffusion constant of uracil(42). Given that  $\tau_{\text{trap}, 0.5} = \tau_{\text{hop}}$  under the condition where 50% of the hopping events are trapped, the Einstein equation may then be used to calculate the distance ( $\langle r_{\text{hop}} \rangle = 7 \text{ nm}$ ) traveled by hUNG in a hopping event of duration  $\tau_{\text{hop}}$  using a calculated  $D_3 = 1.4 \times 10^8 \text{ nm}^2/\text{s}$  (Section 2.4 Materials and Methods) for hUNG (eq 5).

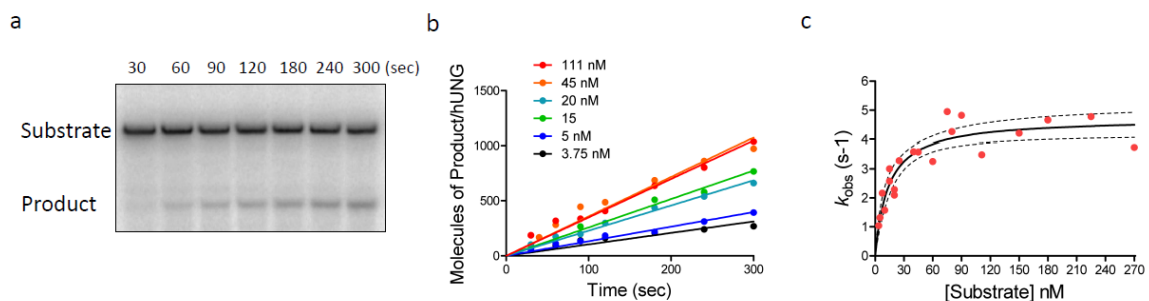
[5] 
$$\langle r_{\text{hop}} \rangle = \sqrt{6D_3\tau_{\text{hop}}}$$

Importantly, while the value for  $\langle r_{\text{hop}} \rangle$  is based on an estimate of the bimolecular diffusion rate of uracil and hUNG, uncertainty in these values does not significantly affect the calculated value for  $\langle r_{\text{hop}} \rangle$  because of its square root dependence on  $\tau_{\text{hop}}$ . This estimated time frame and distance for hopping agrees well with the aforementioned approach using the DNA ruler.

The above analysis uses the simplest hopping model with only a single enzyme dissociation event, which is a reasonable assumption given the short spacing between sites (**Figure 2.1**). With this model, the total distance traveled by hUNG during  $\tau_{\text{hop}}$  is the maximum distance possible for a single hop. If a more complex model employing multiple microscopic hopping steps is used, with each hopping step leading to a possible uracil trapping event, the hopping distance would be reduced from that of the single hop model. However, since hopping is sensitive to increasing ionic strength (**Figure 2.7a, b**), hUNG must escape the ion atmosphere of the DNA (~1 nm)(43) during hops. Therefore, the data brackets the boundary for hopping in the range of 1 to 7 nm. Comprehensive numerical simulations presented in Chapter 5 further support these observations.

The general picture that emerges from these measurements is that frequent hopping events ( $\geq 1$  hop/3 ms) result in dissociated hUNG molecules that escape the ion atmosphere around the DNA, but these enzyme molecules

remain in close proximity to the DNA chain (within 7 nm) allowing very rapid reassociation by hopping ( $k_{\text{hop}} \sim 20 \times 10^6 \text{ s}^{-1}$ ). Thus, DNA hopping through solution is rapid compared with scanning, and the two pathways work synergistically along with the dynamic time scale of enzyme and DNA motions to result in efficient recognition of damaged uracil bases.



**Figure 2.8: Steady-state kinetic parameters for hUNG reaction with a 90 mer duplex DNA substrate containing a single uracil** (see Materials and Methods for further details). **(a)** Representative gel analysis showing a reaction time course and resolution of product and substrate species. **(b)** Initial rates of hUNG cleavage at various substrate concentrations. **(c)** Kinetic parameters  $K_M$  and  $k_{cat}$  were determined to be  $13 \pm 3$  nM and  $4.7 \pm 0.3$  s<sup>-1</sup>, and  $k_{cat}/K_M = 3.4 \times 10^8$  M<sup>-1</sup> s<sup>-1</sup>. The errors are the standard errors of the non-linear regression fit to the Michaelis-Menten equation.

## 2.4 Methods

### 2.4.1 Oligonucleotide and protein reagents

Wild type human Uracil DNA Glycosylase (hUNG) was purified as the catalytic N-terminal truncated form (amino acids 82-304). Specifically the DNA encoding the hUNG amino acid sequence was cloned into a pET-21a vector and expressed in BL21-DE3 E.Coli cells. Briefly, cells containing the hUNG encoding vector were grown in 1 Liter of Luria Broth at 37 °C to an OD<sub>600</sub> of 0.4 and then to an OD<sub>600</sub> = 0.6 at 16 °C. hUNG expression was then induced by the addition of IPTG and the cells were grown at 16 °C overnight. The cells were then harvested by centrifugation and then resuspended in Buffer A (50 mM Tris-Acetate pH 7.0, 10 mM NaCl, 1mM DTT) followed by cell lysis using a microfluidizer. The supernatant was then clarified by centrifugation at 40,000 xg for 45 minutes at 4 °C and directly loaded onto an anion exchange column (UNO-Q12 purchased from BioRad™). The flow through fractions containing hUNG were then loaded onto a Mono-S cation exchange column (GE) that had been preequilibrated with buffer A. hUNG was then purified by gradient elution with Buffer A containing 800 mM NaCl. The final step was gel filtration using BioRad™ P-40 resin. The purified protein was then dialyzed and concentrated into 10 mM sodium phosphate pH 8.0, 150 mM NaCl, 1 mM DTT, 50% glycerol and then stored at -20 °C. The concentrations of hUNG stock solutions were determined using the absorbance at 280 nm and an extinction coefficient of 33.68 mM<sup>-1</sup> cm<sup>-1</sup>.

Oligonucleotides were purchased from Integrated DNA technologies ([www.idtdna.com](http://www.idtdna.com)) and purified in house by denaturing gel. Concentrations of

stock solutions were determined by the absorbance at 260 nm using extinction coefficients calculated from nearest neighbor parameters.

## **2.4.2 Oligonucleotide Sequences**

### *125mer-55 bp U separation (S55)*

GGT ATC CGC TCA CAA TTC CAC ACA ATG CTG AGG AAT CGA U AGC  
TAA GTG AAT CTC TCA CGT CAC ATC GTC CGC ACT AGC ACA TGG AAT  
GAA TCG A U AGC TAA GCT GAG GCA TAC AGT GTC GAG CC

### *90mer-20 bp U separation (S20)*

5' - GGT ATC CGC TCA CAA TTC CAC ACA ATG CTG AGG AAT CGA U AG  
CTA AGT AGG ATG AAT CGA U AG CTA AGC TGA GGC ATA CAG TGT CGA  
GCC

### *90mer-10 bp U separation (S10)*

5' - GGT ATC CGCT AGT CAC AAT TCC ACA CAATGC TGA GGA ATC GA U  
AG CTA AT CGA U AGC TAA GCT GAG GCATAC AGG ATC AAT TGT CGA  
GCC

### *90mer-6 bp U separation (S6)*

5' - GGT ATC CGC TGA AGT AGT CAC AAT CCA CAC AAT GCT GAG GAA  
TCG AUA GCC GAU AGC TAA GCT GAG GCA TAC AGG ATC AAT TGT CGA  
GCC

*90mer-5 bp U separation (S5)*

5' - GGT ATC CGC TGAA GTA GTC ACA ATT CCA CAC AAT GCT GAG GAA  
TCG AUA GCG AUA GCT AAG CTG AGG CAT ACA GGA TCA ATT GTC GAG  
CC

*90mer-5 bp U separation on opposite strands (S5<sup>opp</sup>)*

5' - GGC TCG ACA AGC TCT CAT CTG TGC TGA GGA TGT TAG CTU AGG  
CTA TCG ATT CCT CAG CAT TGT GTG GAG ACA CCT TTG TGA GCG GAT  
ACC

(underlined A indicates position of the uracil on the complementary strand)

*90mer-5bp U single strand (S5<sup>ss</sup>)*

5' - ACC ATA ATA ATA ACA CAT ACA CCA TAC TAC ATA CAT CAA CTA AAA  
CAU ACA CAU ACA AAA TCA ACT AAT AAC AAC ACATAC ACC ATA ACA

*90mer-10bp U single strand (S10<sup>ss</sup>)*

CAC AAT AAC ACA TAC ACC ATA CTA CAT ACA TCA ACT AAA ACA UAC  
ACA ACA CAU ACA AAA TCA ACT AAT AAC AAC ACA TAC ACC ATA ACA

*Non-specific duplex (nsDNA)*

3' - C GCG TGT GCC



5' - G CGC ACA CGG - FAM

*90mer- 1 uracil site (1XU)*

5' - GTT ATC CGC TCA CAA TTC CAC ACA ATG CTG AGG AAT CGA UAG  
CTA AGT AGG ATG TTA GCT ATC GAT TCA TCC TCA GCA CAG TGT CGA  
GCC

*Chase duplex (chDNA)*

5' - GCG GCC AAA  $\phi$  AA AAA GCG C

3' - GCG GCC AAA A TT TTT CGC G

( $\phi$  - tetrahydrofuran abasic site mimic)

### **2.4.3 Intramolecular transfer assay**

Each duplex strand was labeled on the 5' end with  $P^{33}$  by incubation with [ $\gamma^{33}P$ ] ATP (Perkin-Elmer) and T4 polynucleotide kinase (New England Biolabs). The 5'-labeled strands were then hybridized by heating to 95 °C in a heat block and after 10 minutes the block was allowed to cool to room temperature. Unincorporated [ $\gamma^{33}P$ ] ATP was then removed by gel filtration.

For the intramolecular transfer assay each reaction contained 40 nM  $^{33}P$ -labeled duplex DNA substrate in 20 mM HEPES pH 7.5, 0.002% Brij 35, 3 mM EDTA (added from stock at pH 8.0), and 1 mM DTT in a total volume of 28  $\mu$ l total. The pH of all buffer stocks were adjusted with sodium hydroxide giving a final [ $Na^+$ ] of 22 mM. The reaction was then initiated by the addition of 2  $\mu$ l of

human UNG (hUNG) to a final concentration of 5-20 pM and incubated at 37 °C. At each time point 4 µl of the reaction mix was taken out and quenched with Uracil DNA Glycosylase Inhibitor (UGI) at a final concentration of 0.1 Units (New England Biolabs) or 50 nM final concentration of a highly potent duplex DNA inhibitor (2'-fluoro-2'-deoxyuridine paired with 4-methylindole in duplex DNA)(6, 44), both of which rapidly and efficiently quenched hUNG activity. Following reaction quenching, each aliquot was treated with human abasic endonuclease (APE1) and the nicking enzyme Nt.BbvCI to generate discrete double stranded fragments as previously described(6). The fragments were then separated on a 10% non-denaturing polyacrylamide gel (19:1 bis-acrylamide ratio, 0.5 mm thickness, run at 20 watts with 1X TBE running buffer in a model S2 sequencing gel apparatus). The gel was dried and exposed to a storage phosphor screen and imaged with a Typhoon 8600 phosphorimager (GE Healthcare).

For experiments in which uracil was present, an appropriate volume according to the final reaction concentration of a 10 mM uracil stock solution in water was dried and then reconstituted in reaction buffer. The intramolecular transfer assay was then performed exactly as above. Control reactions confirmed that the nicking and abasic site cleavage activities of Nt.BbvCI and APE1 proceeded to completion in the presence of high uracil concentrations. Importantly, the uracil solutions were prepared from uracil of the highest available purity (>99%, Sigma-Aldrich catalog# U0750). This purity was confirmed by <sup>1</sup>H NMR. In addition, the uracil stock was shown to be free of group I and II metals

by Inductively Coupled Plasma Mass Spectrometry (ICP-MS) at the University of Georgia Center for Applied Isotope Studies (<http://www.uga.edu/cais>).

For S5<sup>ss</sup> and S10<sup>ss</sup>, the sequence was designed to eliminate potential secondary structure formation based on Watson-Crick and uracil wobble pairings. Using the hybridization prediction program mFold(45), it was determined that only a maximum of two neighboring bases could potentially pair making any secondary structure formation unlikely. For the assay, both the 5' and 3' ends of the DNA were labeled with [ $\gamma$ -<sup>32</sup>P] ATP and 3'-deoxyadenosine 5'-triphosphate (Cordycepin 5'-triphosphate),-[ $\alpha$ -<sup>32</sup>P] respectively by incubation with T4 polynucleotide kinase and terminal transferase (New England Biolabs). The reactions were performed as described above for duplex DNA except that the abasic sites were cleaved by the addition of a final concentration of 1M piperidine followed by heating to 90° C for 20 min. The reaction mix was then dried to completion in a vacuum centrifuge and resuspended in 50 % formamide containing trace amounts of bromophenol blue and xylene cyanol dyes. The samples were then separated on a 0.5 or 0.4 mm thick 10% denaturing (7M urea) polyacrylamide gel run in 1X TBE at 75 watts in a model S2 sequencing gel apparatus. The gels were then exposed to a storage phosphor screen.

All gel images were quantified using QuantityOne<sup>TM</sup> (Bio-Rad) by the box method. The percentage of intramolecular transfer at each time point ( $P_{\text{trans}}^{\text{obs}}$ ) was then calculated using eq 1(7). The percent transfer at zero time ( $P_{\text{trans}}$ ) was then calculated by linear extrapolation and then divided by the site excision efficiency ( $E = 0.81$ ) to give  $P_{\text{trans}}'$  (**Figure 2c, d**).

#### 2.4.4 Efficiency of Uracil Excision by hUNG

The efficiency of uracil excision by hUNG when it lands on a uracil site is determined by the ratio of the excision rate ( $k_{\text{ex}}$ ) to the off-rate ( $k_{\text{off}}$ )<sup>2</sup>. This efficiency was determined by a pulse-chase kinetic partitioning experiment<sup>2</sup>. Using a three syringe rapid mixing apparatus (Kintek RQF3), 2  $\mu\text{M}$  of hUNG solution was rapidly mixed with 40 nM of a uracil-containing 90mer duplex substrate (1XU) labeled with <sup>32</sup>P at the 5' terminus. After 2 ms aging time the reaction was quenched with either 0.5 M HCl or 60  $\mu\text{M}$  of a 19mer chase DNA duplex (chDNA) containing a tetrahydrofuran abasic site product analogue. Under these high chase DNA concentrations, the trapping of the enzyme is independent of the chDNA concentration. For the acid quenched samples an equal volume of phenol/chloroform/isoamylalcohol (25:24:1, Invitrogen) was added and followed by vortexing. The sample was then centrifuged to separate aqueous and organic phases and 50  $\mu\text{l}$  of the aqueous layer was transferred to a clean tube. 5  $\mu\text{l}$  of 10 M piperidine was then added and the solution was heated to 90 °C for 20 minutes to cleave abasic sites. For the reactions that were quenched with the chDNA duplex, after rapid mixing, 20  $\mu\text{l}$  aliquots were taken and manually quenched using 20  $\mu\text{l}$  of 0.5 M HCl at several time points (**Figure 2.4**). An equal volume of phenol/chloroform/isoamylalcohol was then added and the samples were vortexed. The aqueous and organic layers were then allowed to separate by gravity and 80  $\mu\text{l}$  of 2M piperidine was added. Following centrifugation the aqueous layer was transferred to a clean tube and heated for

20 minutes at 90 °C. After heating, both acid- and chase-quenched samples were dried in a vacuum centrifuge and resuspended in 50% formamide containing trace amounts of bromophenol blue and xylene cyanol dyes. The samples were heated and loaded directly onto a 10% (19:1 bis-acrylamide ratio) denaturing polyacrylamide gel and electrophoresed at 25 watts for 30 min. The gel was dried and exposed to a phosphorimager screen and imaged with a Typhoon 8600 phosphorimager. Gel images were quantified using QuantityOne™ software.

The excision efficiency (E) was calculated using eq 6 (**Figure 2.4**). For a more detailed explanation see Porecha *et al.* and corresponding supplementary methods(6).

$$[6] \quad E = \frac{k_{\text{ex}}}{k_{\text{ex}} + k_{\text{off}}}$$

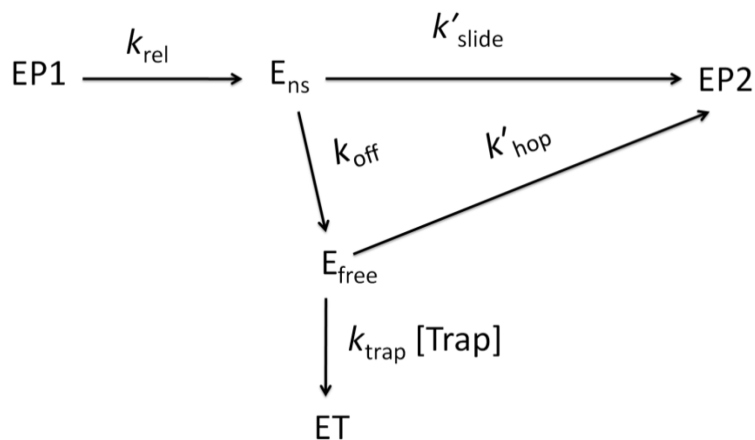
#### 2.4.5 Nonspecific DNA Binding to hUNG

hUNG was titrated into a cuvette containing 50 nM of a 10 bp non-specific DNA duplex (nsDNA) using the same buffer conditions as the intramolecular transfer experiments. The fluorescence anisotropy increase upon hUNG addition was measured using a Spex Fluoromax 3. The dissociation constant ( $K_D$ ) was obtained by fitting to a single site binding isotherm ( $B_{\text{max}} \times [\text{ligand}]_{\text{free}} / (K_D^{\text{ns}} + [\text{hUNG}]_{\text{free}}) + \text{offset}$ ) in Graphpad Prism 5. In this analysis, the free hUNG concentration was assumed to be equal to the total hUNG concentration due to the greater than 10-fold higher  $K_D$  value relative to the concentration of the

labeled nsDNA. The binding lifetime to non-target DNA was calculated as  $\tau_{\text{bind}} = 1/K_D^{\text{ns}} k_{\text{on}}$ . An estimate for the on-rate ( $k_{\text{on}}$ ) of UNG to DNA can be obtained from the diffusion-controlled  $k_{\text{cat}}/K_m = 3.4 \times 10^8 \text{ M}^{-1} \text{ s}^{-1}$  (**Figure 2.8**), and previous stopped-flow on-rate measurements using various DNA constructs ( $2$  to  $7 \times 10^8 \text{ M}^{-1} \text{ s}^{-1}$ )(44, 46). We employ the value obtained from Figure 2.8 in the above text and calculations.

#### 2.4.6 The molecular clock: theory

The principle of a “molecular clock” has been used previously to measure the lifetimes of unstable reaction intermediates that develop during organic reactions(24-26). In the present example, we use a small molecule inhibitor (uracil)(27, 28) to capture free hUNG intermediates that have transiently dissociated from DNA during the process of intramolecular site transfer (**Scheme 1**). In the presence of the uracil trap, these



**Scheme S1**

dissociated enzyme molecules can kinetically partition between two pathways: hopping back onto the DNA chain to complete the site transfer and uracil excision process ( $k'_{\text{hop}} = k_{\text{hop}}k_{\text{ex}}$ ), or they will be captured by the trap, which is a second-order process equal to  $k_{\text{trap}}[\text{trap}]$ . Accordingly, when  $k_{\text{trap}}[\text{trap}] \gg k'_{\text{hop}}$ , trapping is 100 % efficient and no dissociated enzyme molecules proceed to the second uracil site by the hopping pathway. It is important to note that even though the enzyme has dissociated from the DNA in order to be trapped,  $k'_{\text{hop}}$  is *effectively an intramolecular process* because the *same* enzyme molecule hops back onto the *same* DNA chain. Thus,  $k'_{\text{hop}}$  has the units of  $\text{s}^{-1}$ .

A useful condition is when the concentration of trap is sufficient to shut down 50% of the hopping events. When this condition is met,  $k'_{\text{hop}} = k_{\text{trap}}[\text{trap}]$ . A physical interpretation of this particular condition is that uracil diffuses to and traps the free enzyme at the same rate at which the free enzyme diffuses back onto the DNA chain and becomes refractile to trapping. Thus, under this condition the time constant for trapping ( $\tau_{\text{trap}} = 1/k_{\text{trap}}[\text{trap}]$ ) is equal to the time constant for hopping ( $\tau_{\text{hop}} = 1/k'_{\text{hop}}$ ). As described in the next section, the equivalence of these time constants provides the basis for estimating the characteristic distance traveled by the enzyme before it reencounters the DNA chain during a hopping event of time duration  $\tau_{\text{hop}}$ .

#### **2.4.7 Calculation of the trapping time ( $\tau_{\text{trap},0.5}$ ) and the mean hopping distance ( $\langle r_{\text{hop}} \rangle$ )**

To calculate the trapping time where 50% of the hopping enzyme molecules are trapped by uracil ( $\tau_{\text{trap}, 0.5} = 1/k_{\text{trap}}[\text{uracil}]_{0.5}$ ), we employ the Smoluchowski equation for bimolecular association (eq 7)(47). This calculation uses the measured stokes radius of uracil ( $r_U = 0.227 \times 10^{-7}$  cm)(42) to calculate a temperature-corrected diffusion constant of uracil ( $D_3^U = 1.45 \times 10^{-5}$  cm<sup>2</sup> s<sup>-1</sup> at 37 °C in water), the diffusion constant of hUNG using its stokes radius of  $2.3 \times 10^{-7}$  cm ( $r_E$ ) based on small angle x-ray scattering (SAXS) data ( $D_3^E = 1.4 \times 10^{-6}$  cm<sup>2</sup> s<sup>-1</sup>)(48), and Avogadro's number ( $N_A$ ). A value for the fractional binding surface contributed by the active site of the enzyme ( $a = 0.3$ ) was employed, and an electrostatic factor was omitted since uracil is uncharged.

$$[7] \quad k_{\text{trap}} = \frac{4\pi a(D_3^E + D_3^U)(r_E + r_U)N_A}{1000}$$

Although uncertainty in 'a' contributes to uncertainty in the calculated value for  $k_{\text{trap}}$  (and also  $\langle r_{\text{hop}} \rangle$ , see below),  $\langle r_{\text{hop}} \rangle$  is only weakly dependent on 'a' because of its square root dependence on  $\tau_{\text{hop}} = \tau_{\text{trap}, 0.5}$  (**Figure 2.9**). Once  $k_{\text{trap}}$  is calculated from eq 7 ( $8 \times 10^9$  M<sup>-1</sup> s<sup>-1</sup>),  $\tau_{\text{trap}, 0.5} = \{[\text{uracil}]_{0.5} k_{\text{trap}}\}^{-1}$  may be calculated using the concentration of uracil where 50% of the hopping events are trapped (2.1 mM determined from nonlinear least-squares fitting to eq S10). From this approach, a value for  $\tau_{\text{trap}, 0.5} = 60$  ns was obtained.



Given that  $\tau_{\text{trap}, 0.5} = \tau_{\text{hop}}$  under the condition where 50% of the hopping molecules are trapped, the Einstein equation may be used to estimate the characteristic distance traveled by hUNG in a hopping event of duration  $\tau_{\text{hop}}$  using  $D_3^E = 1.4 \times 10^{-6} \text{ cm}^2 \text{ s}^{-1} = 1.4 \times 10^8 \text{ nm}^2 \text{ s}^{-1}$  (eq 8).

$$[8] \quad \langle r_{\text{hop}} \rangle = \sqrt{6D_3^E \tau_{\text{hop}}}$$

#### 2.4.8 Kinetic description of the intramolecular transfer probability ( $P_{\text{trans}}$ )

Here we derive the kinetic expressions to show how  $P_{\text{trans}}'$ ,  $P_{\text{slide}}'$  and  $P_{\text{hop}}'$  are related to the lifetime of hUNG on nonspecific DNA ( $1/k_{\text{off}}$ ), the rate constants for sliding ( $k_{\text{slide}}$ ), hopping ( $k_{\text{hop}}$ ), and the trapping efficiency ( $k_{\text{trap}}[\text{trap}]$ ) (see Scheme 1). Note that in Scheme 1,  $k'_{\text{hop}} = k_{\text{hop}}k_{\text{ex}}$  and  $k'_{\text{slide}} = k_{\text{slide}}k_{\text{ex}}$ , where  $k_{\text{ex}}$  is the uracil excision step. The clock begins after hUNG is released from the abasic site (EP1) generated from excision of uracil from the first site that is encountered ( $k_{\text{rel}}$ ). Since abasic site dissociation is the rate limiting step during steady-state turnover(46), none of the subsequent fast steps can be directly measured. However, the partitioning of the enzyme down different kinetic pathways can be ascertained by varying the concentration of the uracil trap, and also from knowledge of the lifetime of hUNG on non-specific DNA. After release from the first site to generate EP1, the enzyme becomes bound to non-specific DNA ( $E_{\text{ns}}$ , Scheme 1). From this state hUNG may slide to the next site with a probability ( $P_{\text{slide}}'$ ) that is determined by the rate constants for sliding ( $k'_{\text{slide}}$ ) and

for falling off the DNA chain ( $k_{\text{off}}$ ) (eq 9). A similar expression describes the probability for falling off the DNA chain ( $P_{\text{off}}$ , eq 10).

$$[9] \quad P_{\text{slide}}' = \frac{k'_{\text{slide}}}{k_{\text{off}} + k'_{\text{slide}}}$$

$$[10] \quad P_{\text{off}} = \frac{k_{\text{off}}}{k_{\text{off}} + k'_{\text{slide}}}$$

If the site spacing is too large, hUNG will ultimately dissociate to the free enzyme state ( $E_{\text{free}}$ , Scheme 1). In the presence of trap, the free enzyme can partition between hopping back on the DNA chain with a probability  $P_{\text{hop}}'$  (eq 11), or being trapped with a probability  $P_{\text{trap}}$  (eq 12)

$$[11] \quad P_{\text{hop}}' = \frac{k'_{\text{hop}}}{k'_{\text{hop}} + k_{\text{trap}}[\text{trap}]}$$

$$[12] \quad P_{\text{trap}} = \frac{k_{\text{trap}}[\text{trap}]}{k'_{\text{hop}} + k_{\text{trap}}[\text{trap}]}$$

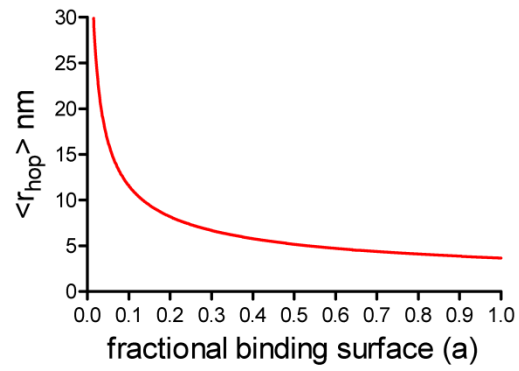
The total transfer probability ( $P_{\text{trans}}'$ ), as measured in the facilitated transfer assay, is the sum of all pathways leading to EP2, which can be defined in terms of the above probabilities:

$$[13] \quad P_{\text{trans}}' = P_{\text{slide}}' + P_{\text{off}}P_{\text{hop}}$$

$$[14] \quad P_{\text{trans}}' = P_{\text{slide}}' + P_{\text{off}} \left( \frac{k'_{\text{hop}}}{k'_{\text{hop}} + k_{\text{trap}}[\text{trap}]} \right) = \frac{k'_{\text{slide}}}{k_{\text{off}} + k'_{\text{slide}}} + \left( \frac{k_{\text{off}}}{k_{\text{off}} + k'_{\text{slide}}} \right) \left( \frac{k'_{\text{hop}}}{k'_{\text{hop}} + k_{\text{trap}}[\text{trap}]} \right)$$

#### 2.4.9 Calculation of the energy barrier for hUNG sliding

Given the experimentally determined one dimensional diffusion coefficient for hUNG sliding to be  $6.0 \times 10^3 \text{ bp}^2/\text{s}$  or  $6.9 \times 10^{-4} \text{ }\mu\text{m}^2/\text{s}$ , the energetic barrier of sliding can be calculated by setting the reference state to that of 'barrierless' sliding as calculated for rotation coupled diffusion around the DNA helix ( $\sim 10^7 \text{ bp}^2/\text{s}$ ,  $\sim 1 \text{ }\mu\text{m}^2/\text{s}$ )(31, 33, 34). For hUNG sliding we calculate this energy to be approximately  $7 k_b T$  (e.g.  $7 k_b T = k_b T \times \ln(D_{1\text{-experimental}}/D_{1\text{-theoretical}})$ ).



**Figure 2.9: Calculated targeting radius ( $\langle r_{hop} \rangle$ ) based on the Smoluchowski equation (eq 7) and the Einstein equation for a diffusing particle (eq 8) as a function of the fractional binding surface.**

## 2.5: References

1. Gowers DM, Wilson GG, Halford SE. 2005. Measurement of the contributions of 1D and 3D pathways to the translocation of a protein along DNA. *Proc. Natl. Acad. Sci. U.S.A.* 102:15883–15888.
2. Gowers DM, Halford SE. 2003. Protein motion from non-specific to specific DNA by three-dimensional routes aided by supercoiling. *EMBO J.* 22:1410–1418.
3. Stanford NP, Szczelkun MD, Marko JF, Halford SE. 2000. One- and three-dimensional pathways for proteins to reach specific DNA sites. *EMBO J.* 19:6546–6557.
4. Hedglin M, O'Brien PJ. 2010. Hopping Enables a DNA Repair Glycosylase To Search Both Strands and Bypass a Bound Protein. *ACS Chemical Biology* 5:427–436.
5. Sidorenko VS, Zharkov DO. 2008. Correlated Cleavage of Damaged DNA by Bacterial and Human 8-Oxoguanine-DNA Glycosylases. *Biochemistry* 47:8970–8976.
6. Porecha RH, Stivers JT. 2008. Uracil DNA glycosylase uses DNA hopping and short-range sliding to trap extrahelical uracils. *Proc. Natl. Acad. Sci. U.S.A.* 105:10791–10796.
7. Terry BJ, Jack WE, Modrich P. 1985. Facilitated diffusion during catalysis by EcoRI endonuclease. Nonspecific interactions in EcoRI catalysis. *J. Biol. Chem.* 260:13130–13137.
8. Hedglin M, O'Brien PJ. 2008. Human Alkyladenine DNA Glycosylase

- Employs a Processive Search for DNA Damage. *Biochemistry* 47:11434–11445.
9. Elf J, Li G-W, Xie XS. 2007. Probing Transcription Factor Dynamics at the Single-Molecule Level in a Living Cell. *Science* 316:1191–1194.
  10. Dowd DR, Lloyd RS. 1989. Site-directed mutagenesis of the T4 endonuclease V gene: the role of arginine-3 in the target search.
  11. Tafvizi A, Huang F, Fersht AR, Mirny LA, van Oijen AM. 2011. A single-molecule characterization of p53 search on DNA. *Proc. Natl. Acad. Sci. U.S.A.* 108:563–568.
  12. Komazin-Meredith G, Mirchev R, Golan DE, van Oijen AM, Coen DM. 2008. Hopping of a processivity factor on DNA revealed by single-molecule assays of diffusion. *Proc. Natl. Acad. Sci. U.S.A.* 105:10721–10726.
  13. Gorman J, Chowdhury A, Surtees JA, Shimada J, Reichman DR, Alani E, Greene EC. 2010. Collaborative Dynamic DNA Scanning by Nucleotide Excision Repair Proteins Investigated by Single- Molecule Imaging of Quantum-Dot-Labeled Proteins *Molecular cell*.
  14. Kad NM, Wang H, Kennedy GG, Warshaw DM, Van Houten B. 2010. Collaborative Dynamic DNA Scanning by Nucleotide Excision Repair Proteins Investigated by Single- Molecule Imaging of Quantum-Dot-Labeled Proteins. *Mol. Cell* 37:702–713.
  15. Tafvizi A, Mirny LA, van Oijen AM. 2011. Dancing on DNA: kinetic aspects of search processes on DNA. *Chemphyschem* 12:1481–1489.
  16. Blainey PC, van Oijen AM, Banerjee A, Verdine GL, Xie XS. 2006. A base-

- excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA. *Proc. Natl. Acad. Sci. U.S.A.* 103:5752–5757.
17. Blainey PC, Luo G, Kou SC, Mangel WF, Verdine GL, Bagchi B, Xie XS. 2009. Nonspecifically bound proteins spin while diffusing along DNA. *Nat Struct Mol Biol* 16:1224–1229.
  18. Wang YM, Austin RH, Cox EC. 2006. Single Molecule Measurements of Repressor Protein 1D Diffusion on DNA. *Physical Review Letters* 97:048302.
  19. Gorman J, Chowdhury A, Surtees JA, Shimada J, Reichman DR, Alani E, Greene EC. 2007. Dynamic Basis for One-Dimensional DNA Scanning by the Mismatch Repair Complex Msh2-Msh6. *Mol. Cell* 28:359–370.
  20. Dunn AR, Kad NM, Nelson SR, Warshaw DM, Wallace SS. 2011. Single Qdot-labeled glycosylase molecules use a wedge amino acid to probe for lesions while scanning along DNA. *Nucleic Acids Res.*
  21. Gorman J, Plys AJ, Visnapuu M-L, Alani E, Greene EC. 2010. Visualizing one-dimensional diffusion of eukaryotic DNA repair factors along a chromatin lattice. *Nat. Struct. Mol. Biol.* 17:932–938.
  22. Wang YM, Austin RH. 2011. Single-Molecule Imaging of LacI Diffusing Along Nonspecific DNA, pp. 9–37. *In* Williams, MC, Maher, JL (eds.), *Biophysics of DNA-Protein Interactions From Single Molecules to Biological Systems*. Springer.
  23. Maul RW, Gearhart PJ. 2010. AID and Somatic Hypermutation, pp. 159–191. *In* Alt, FW (ed.), *Advances in Immunology*. Academic Press.

24. Griller D, Ingold KU. 1980. Free-radical clocks. *Acc. Chem. Res.* 13:317–323.
25. Fishbein JC, Jencks WP. 1988. Elimination reactions of beta-cyano thioethers: evidence for a carbanion intermediate and a change in rate-limiting step. *J. Am. Chem. Soc.* 110:5075–5086.
26. Amyes TL, Jencks WP. 1989. Lifetimes of oxocarbenium ions in aqueous solution from common ion inhibition of the solvolysis of alpha-azido ethers by added azide ion. *J. Am. Chem. Soc.* 111:7888–7900.
27. Xiao G, Tordova M, Jagadeesh J, Drohat AC, Stivers JT, Gilliland GL. 1999. Crystal structure of *Escherichia coli* uracil DNA glycosylase and its complexes with uracil and glycerol: structure and glycosylase mechanism revisited. *Proteins* 35:13–24.
28. Jiang YL, Krosky DJ, Seiple L, Stivers JT. 2005. Uracil-Directed Ligand Tethering: An Efficient Strategy for Uracil DNA Glycosylase (UNG) Inhibitor Development. *J. Am. Chem. Soc.* 127:17412–17420.
29. Berg HC. 1993. *Random walks in biology* Princeton University Press.
30. Halford SE, Marko JF. 2004. How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.* 32:3040–3052.
31. Berg OG, Winter RB, Hippel Von PH. 1981. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* 20:6929–6948.
32. Belotserkovskii BP, Zarling DA. 2004. Analysis of a one-dimensional



- random walk with irreversible losses at each step: applications for protein movement on DNA. *Journal of theoretical biology* 226:195–203.
33. Schurr JM. 1979. The one-dimensional diffusion coefficient of proteins absorbed on DNA: Hydrodynamic considerations. *Biophysical chemistry* 9:413–414.
  34. Bagchi B, Blainey PC, Xie XS. 2008. Diffusion Constant of a Nonspecifically Bound Protein Undergoing Curvilinear Motion along DNA. *J. Phys. Chem. B* 112:6282–6284.
  35. Bruner SD, Norman DPG, Verdine GL. 2000. Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature* 403:859–866.
  36. Kuznetsov NA, Koval VV, Nevinsky GA, Douglas KT, Zharkov DO, Fedorova OS. 2007. Kinetic conformational analysis of human 8-oxoguanine-DNA glycosylase. *J. Biol. Chem.* 282:1029–1038.
  37. Kuznetsov N, Koval V, Fedorova O. 2011. Mechanism of recognition and repair of damaged DNA by human 8-oxoguanine DNA glycosylase hOGG1. *Biochemistry (Moscow)* 76:118–130.
  38. Cao C, Jiang YL, Stivers JT, Song F. 2004. Dynamic opening of DNA during the enzymatic search for a damaged base. *Nat. Struct. Mol. Biol.* 11:1230–1236.
  39. Parker JB, Bianchet MA, Krosky DJ, Friedman JI, Amzel LM, Stivers JT. 2007. Enzymatic capture of an extrahelical thymine in the search for uracil in DNA. *Nature* 449:433–437.

40. Friedman JI, Majumdar A, Stivers JT. 2009. Nontarget DNA binding shapes the dynamic landscape for enzymatic recognition of DNA damage. *Nucleic Acids Res.* 37:3493–3500.
41. Kampmann M. 2004. Obstacle bypass in protein motion along DNA by two-dimensional rather than one-dimensional sliding. *J. Biol. Chem.* 279:38715–38720.
42. Nishida K, Ando Y, Kawamura H. 1983. Diffusion coefficients of anticancer drugs and compounds having a similar structure at 30 Å°C. *Colloid Polym. Sci.* 261:70–73.
43. Young MA, Ravishanker G, Beveridge DL. 1997. A 5-nanosecond molecular dynamics trajectory for B-DNA: analysis of structure, motions, and solvation.
44. Krosky DJ, Song F, Stivers JT. 2005. The Origins of High-Affinity Enzyme Binding to an Extrahelical DNA Base †. *Biochemistry* 44:5949–5959.
45. Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31:3406–3415.
46. Stivers JT, Pankiewicz KW, Watanabe KA. 1999. Kinetic Mechanism of Damage Site Recognition and Uracil Flipping by Escherichia coli Uracil DNA Glycosylase. *Biochemistry* 38:952–963.
47. Smoluchowski MV. 1917. Mathematical theory of the kinetics of the coagulation of colloidal solutions. *Z Phys Chem* 92:129–168.
48. Timchenko SK, Kubareva EA, Volkov EM, Voronina OL, Lunin VG, Gonchar DA, Degtiarev SK, Timchenko MA, Kihara H, Kimura K. 2006.

Structure of *Escherichia coli* uracil DNA glycosylase and its complexes with nonhydrolyzable substrate analogues in solution observed by synchrotron small-angle X-ray scattering. *Biofizika* 51:5–12.

## Chapter 3

# DNA Translocation by Human Uracil DNA Glycosylase: Role of DNA Phosphate Charge

Reproduced in part from:

Schonhoft JD and Stivers JT. (2013) DNA Translocation by Human Uracil DNA  
Glycosylase: The Case of Single-Stranded DNA and Clustered Uracils  
*Biochemistry* 52 (15), 2536-2544

### 3.1 Introduction

Human DNA repair glycosylases must encounter and inspect each DNA base in the genome in order to discover damaged bases that may be present at a density of less than one in ten million normal base pairs. This remarkable example of specific molecular recognition requires a reduced dimensionality search process (facilitated diffusion) that involves both hopping and sliding along the DNA chain. Despite the widely accepted importance of facilitated diffusion in protein-DNA interactions, the molecular features of DNA that influence hopping and sliding are poorly understood. Here we explore the role of the charged DNA phosphate backbone in sliding and hopping by human uracil DNA glycosylase (hUNG), which is an exemplar that efficiently locates rare uracil bases in both dsDNA and ssDNA. Substitution of neutral methylphosphonate groups for anionic DNA phosphate groups weakened nonspecific DNA binding affinity by 0.4-0.5 kcal/mole per substitution. In contrast, sliding of hUNG between uracil sites embedded in duplex and single stranded DNA substrates persisted unabated when multiple methylphosphonate linkages were inserted between the sites. Thus a continuous phosphodiester backbone negative charge is not essential for sliding over nonspecific DNA binding sites. We consider several alternative mechanisms for these results. A model consistent with previous structural and NMR dynamic results invokes the presence of open and closed conformational states of hUNG. The open state is short-lived and has weak or nonexistent interactions with the DNA backbone that are conducive for sliding, and the populated closed state has stronger interactions with the phosphate

backbone. These data suggest that the fleeting sliding form of hUNG is a distinct weakly interacting state that facilitates rapid movement along the DNA chain and resembles the transition state for DNA dissociation.

The integrity of the information content of genomic DNA depends on efficient and accurate repair of damaged DNA bases. In many cases, this task is initiated by base excision repair DNA glycosylases, which locate and cleave the glycosidic bond of rare mutagenic bases in DNA (1, 2). Unlike transcription factors or other DNA binding proteins, these unique repair glycosylases must rapidly encounter and inspect each base in the genome in the process of efficiently locating their damage targets. This unique search requirement, which is driven by the evolutionary necessity to patrol the genome, places stringent restraints on the thermodynamic and kinetic aspects of the enzyme-nucleic acid interaction that almost certainly differ from typical DNA binding proteins. If the glycosylase interacts too strongly with nonspecific DNA, then it spends too much time at non-target sites, if it interacts too weakly or moves too fast, then its residence time is not long enough to allow detection of DNA damage when it is encountered. These properties of an efficient damage search are one example of what has been called the “search-speed/stability” paradox (3, 4).

To resolve the paradox, DNA glycosylases have harnessed the most favorable mechanistic features of two distinct modes of facilitated diffusion: DNA hopping and sliding (2, 3, 5, 6). Frequent dissociation from the DNA chain most often results in reassociation at a nearby DNA segment (hopping), keeping the enzyme from wasting time unproductively searching regions where there is no

DNA and allowing it to bypass bound proteins (7, 8). Once the enzyme has encountered a new DNA segment, it then has an opportunity to remain in contact with the chain and move along it in a one-dimensional sliding mode (3, 5, 6). An upper limit on the length of DNA over which sliding can occur is determined by the residence time of the enzyme on nonspecific DNA and the 1D diffusion constant (6). The importance of sliding, even over short segments of the DNA chain, is that the enzyme remains in contact with its substrate, thereby expanding the number of bases that can be inspected during each binding event. These two general modes of the search have been observed (or inferred) for many DNA glycosylases and other site-specific DNA binding proteins (7-20).

Although the fundamental importance of hopping and sliding in the damage search is well appreciated, a quantitative mechanistic understanding of the molecular features of the DNA chain that influence an enzyme's ability to hop and slide are poorly understood. In this regard, it is widely believed that the polyanion character of the DNA phosphate backbone provides an important nonspecific electrostatic handle allowing engagement of positively charged side chains on the enzyme. Such interactions may play a role in both hopping and sliding along nonspecific DNA, but also in other steps of the reaction such as specific recognition, making it challenging to sort out these individual effects (21-23). Specifically, electrostatic tracking along the phosphate backbone is often invoked as the primary translocation mode for DNA sliding, but a direct test of this mechanism has been absent. Here we investigate the role of charged DNA phosphate groups in the ability of human uracil DNA glycosylase (hUNG) to hop

and slide along DNA during its search for uracil bases. The results show that a continuous backbone charge is not required for hUNG to track efficiently along a DNA strand, and that the transient sliding state has features that resemble the transition state for DNA dissociation.



## 3.2 Results

### 3.2.1 Experimental Approach

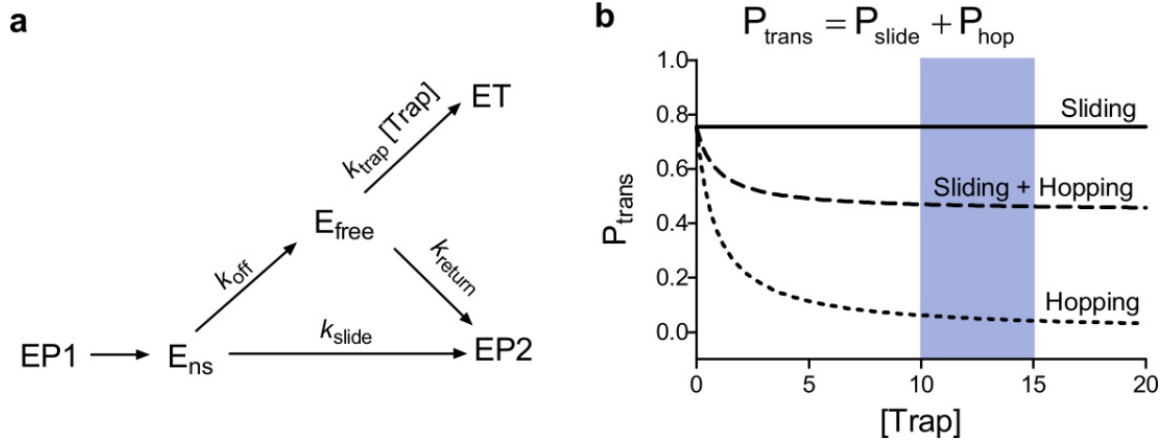
A method was recently described that allows measurement of the probability ( $P_{\text{trans}}$ ) that hUNG will successfully transfer between two uracil sites embedded in a single DNA chain separated by a known distance using a sliding or hopping pathway (**Figure 3.1a**) (9). This is the first approach that allows dissection of the total transfer probability into the individual contributions from hopping ( $P_{\text{hop}}$ ) and sliding ( $P_{\text{slide}}$ ), where  $P_{\text{trans}} = P_{\text{hop}} + P_{\text{slide}}$  (**Figure 3.1**).

The method requires quantitative site transfer probability measurements (see below) in the absence and presence of a small molecule trap of the enzyme. Inclusion of the trap (the free uracil base) serves to capture all enzyme molecules that have dissociated from the DNA during the process of transferring between the two uracil sites (i.e. the enzyme molecules that have hopped off the DNA). The trap has no effect on enzyme transfers that follow the sliding pathway because the binding site for the trap is blocked when hUNG is bound to nonspecific DNA. Separation of the two pathways ( $P_{\text{hop}}$  and  $P_{\text{slide}}$ ) is possible because at zero concentration of trap transfer can occur by both hopping and sliding, but as the trap concentration increases, the hopping contribution diminishes in a hyperbolic fashion, ultimately approaching a limiting asymptote equal to  $P_{\text{slide}}$  (long dashed line, **Figure 3.1b**). If the site spacing is large enough, no enzyme molecules will reach the second site without departing the DNA at least once, and transfer will be entirely ablated at high concentrations of trap (short dashed line, **Figure 3.1b**). Conversely, at short site spacings all transfers

may occur by sliding and therefore will be impervious to the trap (solid line, **Figure 3.1b**).

The reader is referred to reference (9) for a detailed description of the method including control experiments that establish its utility for hUNG. The experimental observations that support the conclusion that uracil serves as a trap of a dissociated state of hUNG without disrupting DNA sliding are: (i) Two pathways for transfer between substrate sites are observed (uracil insensitive and sensitive). (ii) The uracil concentration dependence of  $P_{\text{trans}}$  follows the expected hyperbolic kinetic behavior (**Figure 3.1b**), including the non-zero plateau value at short site spacings and high uracil concentrations, as would be expected for concurrent hopping and sliding. (iii) The site spacing dependences of  $P_{\text{slide}}$  and  $P_{\text{hop}}$  are consistent with those expected for hopping and sliding pathways. That is, the probability for hopping (uracil sensitive) follows a  $1/r$  dependence on site spacing, while the probability for sliding (uracil insensitive) shows a  $\text{bp}^2$  dependence on site spacing. (iii) Transfer was completely eliminated at high uracil concentrations when the substrate sites were positioned on opposite DNA strands where an obligate dissociation/reassociation step is required for transfer. This was observed even though the opposite strand sites were closer in space than when positioned on the same strand. Additionally, at site spacings exceeding the sliding length, identical values of  $P_{\text{trans}}$  were previously observed for sites positioned on the same and opposite strands, consistent with hopping (11). (iv) The hopping pathway was highly sensitive to increases in ionic strength while the sliding pathway was not (9). (v) High

concentrations of uracil have no effect on the dissociation constant for nonspecific binding of UNG to DNA, but uracil blocks the catalytic activity of the enzyme (9). These observations are consistent with the trap having no effect on sliding and acting solely by trapping the active site of the dissociated enzyme.



**Figure 3.1. Molecular clock method.** a) The ‘molecular clock’ approach uses a small molecule inhibitor of hUNG (uracil) to trap enzyme molecules that have hopped off the DNA chain during transfer between two uracil target sites, while leaving sliding enzymes unperturbed (9). Thus, this method allows quantitative determination of the individual contributions of hopping and sliding transfers (where the total transfer probability is  $P_{\text{trans}} = P_{\text{hop}} + P_{\text{slide}}$ ). b) Simulations showing the dependence of facilitated transfer ( $P_{\text{trans}}$ ) on the concentration of the trap [based on the mechanism in (a)]. As previously noted (9), the probability of locating a site by hopping includes the probability that the enzyme initially falls off the DNA ( $P_{\text{off}}$ ) as well as the probability that it returns to the same DNA chain ( $P_{\text{return}}$ ) without getting lost to bulk solution ( $P_{\text{hop}} = P_{\text{off}}P_{\text{return}}$ ). The trap allows selective disruption of the hopping pathway because the probability that a hopping enzyme returns to the DNA chain decreases according to  $P_{\text{return}} = k_{\text{return}}/(k_{\text{return}} + k_{\text{Trap}}[\text{Trap}])$ . The utility of this approach and the numerous control experiments that confirm its utility have been previously published (9).

### 3.2.2 Calculating Site Transfer Probabilities

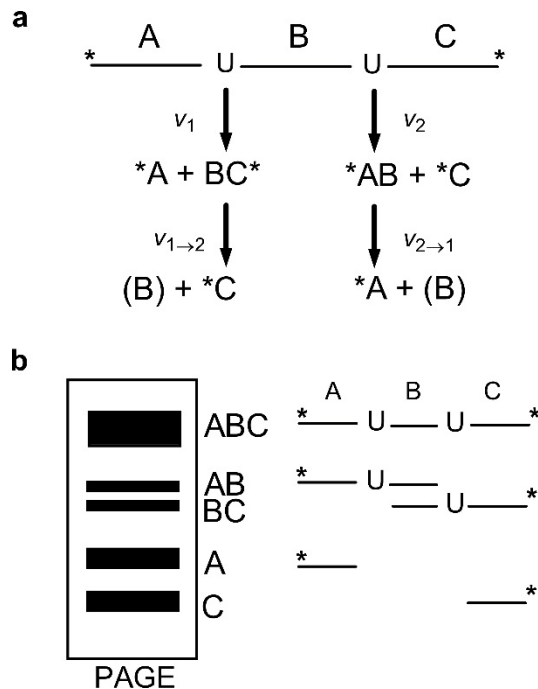
To determine the probabilities for facilitated site transfer by hUNG, we use an initial-rate, steady-state assay that quantifies the fraction of enzyme molecules that excise one uracil site (primary excision events) and then successfully transfer and excise the other uracil in the same DNA molecule (secondary excision events) (**Figure 3.2a**) (9, 24, 25). Primary or secondary uracil excision events will lead to discrete fragments of the double end-labeled DNA which may be resolved by polyacrylamide gel electrophoresis after post-reaction sample processing (see Materials and Methods) (**Figure 3.2b**). If only primary excision events occur at site 1 or 2, the DNA fragments A + BC or AB + C will be produced in equal amounts with apparent velocities  $v_1 = v_2$  if each site reacts identically. However, if intramolecular transfer occurs, the larger AB and BC fragments will be efficiently converted into the smaller fragments A and C (as well as the unobserved B fragment) with velocities  $v_{2 \rightarrow 1}$  (reflecting 2→1 transfers) and  $v_{1 \rightarrow 2}$  (reflecting 1→2 transfers). It is worth noting that the initial rates for formation of fragments A and C depend on both primary and secondary events, and therefore,  $v_{2 \rightarrow 1}$  and  $v_{1 \rightarrow 2}$  are not necessarily equivalent to the initial rates of appearance of fragments A and C. In general, the qualitative hallmark of intramolecular transfer is the production of greater amounts of the secondary excision products A and C at the expense of the single excision products AB and BC (20, 24).

The overall transfer probability ( $P_{\text{trans}}$ ) may be precisely calculated from the time dependent fragment concentrations. These concentrations are inserted

into eq 3.1, which requires extrapolation to zero time to obtain the true transfer probability (24). The basis for this equation can be easily understood: the denominator counts all excision events and the numerator counts only secondary excision events.

$$[3.1] \quad P_{\text{trans}}^{\text{obs}} = \frac{[A] + [C] - [BC] - [AB]}{[A] + [C] + [BC] + [AB]}$$

Thus, the ratio reveals the fraction of all excision events that lead to successful transfers to the second site. (The term  $-[AB] - [BC]$  in the numerator corrects for the fact that fragments A and C can result from both primary excision events  $ABC \rightarrow A + BC$  and  $ABC \rightarrow AB + C$ , or secondary excision events  $AB \rightarrow A + B$  and  $BC \rightarrow B + C$ .) We find that this is a useful and straightforward analytical approach when there is no site preference for excision of the individual sites or no directional bias to transfer (9, 11). In the following chapter we use a modified analytical approach that is useful when a site excision or transfer bias is present (Chapter 4).



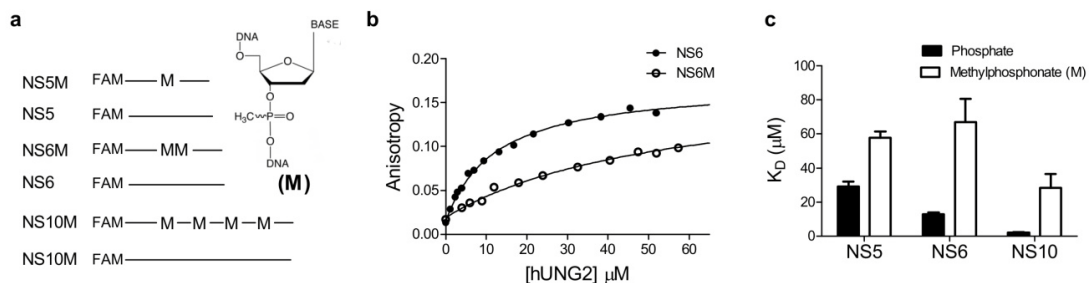
**Figure 3.2. Site transfer assay.** a) Schematic of the site transfer assay where a single strand of a duplex DNA substrate contains two uracil sites is reacted with hUNG ( $[\text{hUNG}] \ll [\text{DNA}]$ ). After quenching, cleavage of the product abasic sites results in single site (AB and BC) and double site (A and C) product fragments (produced from intramolecular translocation events). Qualitatively, intramolecular translocation of hUNG is indicated by an excess of the A and C fragments under initial rate conditions (see text). b) Schematic of the reaction products as analyzed by gel electrophoresis. Bands are quantified by imaging and the intramolecular transfer efficiency is calculated from eq 1 (see text).

### 3.2.3 Effects of Neutral Methylphosphonate (M) Substitutions on Nonspecific DNA Binding

Previous work suggested that a continuous DNA phosphate backbone was necessary and sufficient for DNA sliding because (i) hUNG sliding only occurred between uracil sites that were positioned on the same strand in duplex DNA, and (ii) sliding between uracil sites was observed on ssDNA substrates (reference (9) and Chapter 4). To begin to explore the role of phosphate backbone charge on DNA sliding, the effects of neutral M substitutions on nonspecific DNA binding were first determined using fluorescence anisotropy measurements (**Figure 3.3**). Three 5' fluorescein-labeled DNA constructs were investigated that contained mixed diastereomer M linkages at one or more positions, which were then compared with the corresponding all phosphodiester versions (Figure 3a). The M-DNA constructs shown in **Figure 3.3a** were chosen to match the intervening DNA strand segments used in the site transfer measurements described below (NS5M, NS6M), and also to evaluate the effect of removing as many as four phosphate charges (NS10M). Using single DNA strands as models for the intervening sequences in duplex DNA is justified because (i) structural studies indicate that hUNG primarily interacts with the phosphate backbone on the single strand of DNA that connects the two target bases (26), (ii) hUNG is known to slide along a single strand in duplex DNA (reference (9) and Chapter 4), and (iii) M substitutions in a single strand of duplex DNA still allow binding to the other strand and would complicate the interpretation of equilibrium dissociation constants.



Binding measurements revealed that single or multiple M substitutions in single stranded DNA decreased the binding affinity ( $K_d$ ) of hUNG compared to matched all phosphodiester controls (**Figure 3.3b & c**). The 2 fold-effect of a single M substitution in the center of a 5 mer strand (NS5M and NS5,  $\Delta\Delta G = 0.42$  kcal/mol) was increased 5-fold in the 6 mer strand containing two M substitutions (NS6M,  $\Delta\Delta G = 1.01$  kcal/mol, or  $\sim 0.5$  kcal/mol per M linkage). Similarly, the 10mer strand containing four methylphosphonates (NS10M) showed a 14-fold deficit in binding ( $\Delta\Delta G = 1.52$  kcal/mol, or  $\sim 0.4$  kcal/mol per M linkage). Qualitatively, these results demonstrate that the removal of phosphate charge has a damaging effect on nonspecific DNA binding, and raise the expectation that if site transfer by hUNG requires interaction with the phosphate backbone, the removal of these interactions should diminish successful sliding between two uracil sites. We defer to the Discussion possible further physical interpretations of the effects of M substitution on nonspecific DNA binding.



**Figure 3.3. Substrate design and effects of methylphosphonate (M) substitutions on nonspecific DNA binding as measured using fluorescence anisotropy.** (a) Design of 5' fluorescein labeled oligonucleotides containing all phosphodiester or methylphosphonate (M) linkages. NS5 and NS6 and their corresponding M containing oligonucleotide sequences were chosen to match the intervening sequences in the uracil containing substrate used in the site transfer assays. (b) Equilibrium hUNG binding to oligonucleotides NS6 and NS6M. (c) Summary of determined dissociation constants for hUNG and non-specific DNA. Error bars represent mean  $\pm$  1 s.d. of at least three independent trials.

### 3.2.4 Sliding of hUNG Does Not Require a Continuous Polyanion DNA Strand

To address the question of whether a continuous backbone charge is a requirement for sliding along duplex DNA, we synthesized two 90mer M-substituted DNA substrates containing M linkages on the DNA strand connecting the two uracils (**Figure 3.4a**). The intervening nonspecific DNA strand that connects the two uracil sites in these substrates corresponds exactly to the sequences of NS5M or NS6M described above. Importantly, the substrates in **Figure 3.4a** were constructed with uracil spacings of five and six bp (S5M, S6M) because it has been previously shown that 40% and 20% of the hUNG site transfers occur by a sliding pathway at these spacings when uninterrupted phosphodiester linkages are present (9). In addition, we took care to position the M linkages far enough away from the uracil sites such that the footprint of the specific hUNG catalytic complex does not overlap these positions. This aspect of the substrate design is critical because it has been shown that specific M substitutions within two nucleotides of the uracil site can have a large damaging effect on catalysis ( $\Delta\Delta G$  up to 10 kcal/mole) (22, 23).

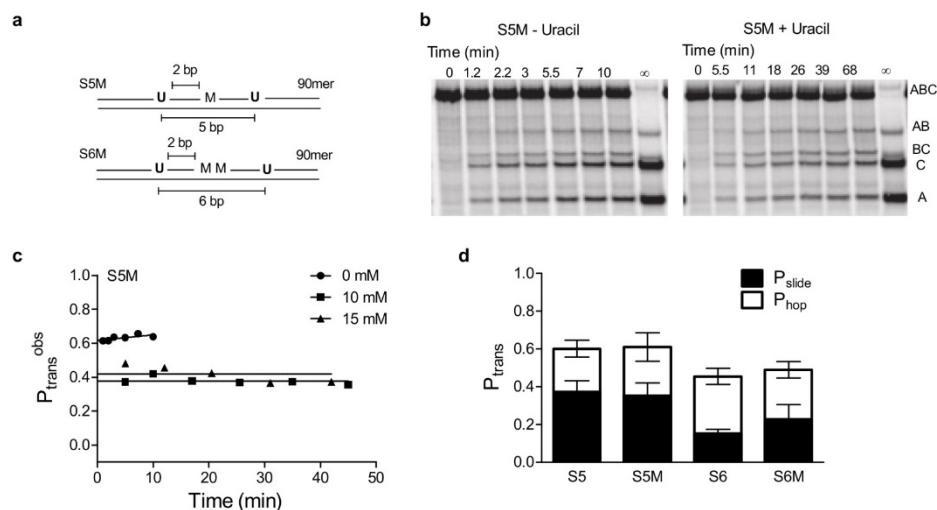
We measured  $P_{\text{trans}}$ ,  $P_{\text{slide}}$ , and  $P_{\text{hop}}$  for S5M and S6M, and compared these values to those previously measured for the analogous phosphodiester substrates S5 and S6 (**Figure 4**) (9). As described above, measurement of  $P_{\text{trans}}$  was obtained in the absence of the uracil trap, and measurements of  $P_{\text{slide}}$  were obtained in the presence of 10 and 15 mM trap (the two values were identical, confirming that the transfer measurements were in the plateau region depicted in

Figure 1b). For all of the data,  $P_{\text{slide}}$  is reported as the average value obtained at 10 and 15 mM uracil concentration ( $n = 3$  for each concentration) and  $P_{\text{hop}}$  is the difference between  $P_{\text{trans}}$  and  $P_{\text{slide}}$ . Representative transfer data for S5M in the absence and presence of the uracil trap shows a significant degree of intramolecular transfer as revealed by excess A and C fragments (**Figure 3.4b**). Addition of 10 or 15 mM uracil trap leads to a reduction in the successful transfer events, but transfer is not entirely ablated indicating that a sliding pathway is present (**Figure 3.4b**). Extrapolation to zero time using eq 1 shows that  $P_{\text{trans}} = 0.61 \pm 0.08$  and  $P_{\text{slide}} = 0.35 \pm 0.07$  for S5M, which are values indistinguishable from those previously reported for S5 (**Figure 3.4c and d**). In addition, there was no difference between the transfer parameters of S6M containing two intervening M linkages and the all phosphodiester analog S6 (**Figure 3.4d**). The transfer parameters for these phosphodiester and M-substituted substrates are summarized in Figure 4d from which we conclude that ablating as many as one-third of the intervening negative charges connecting the two uracil sites has no measurable effect on  $P_{\text{trans}}$ ,  $P_{\text{slide}}$ , or  $P_{\text{hop}}$ .

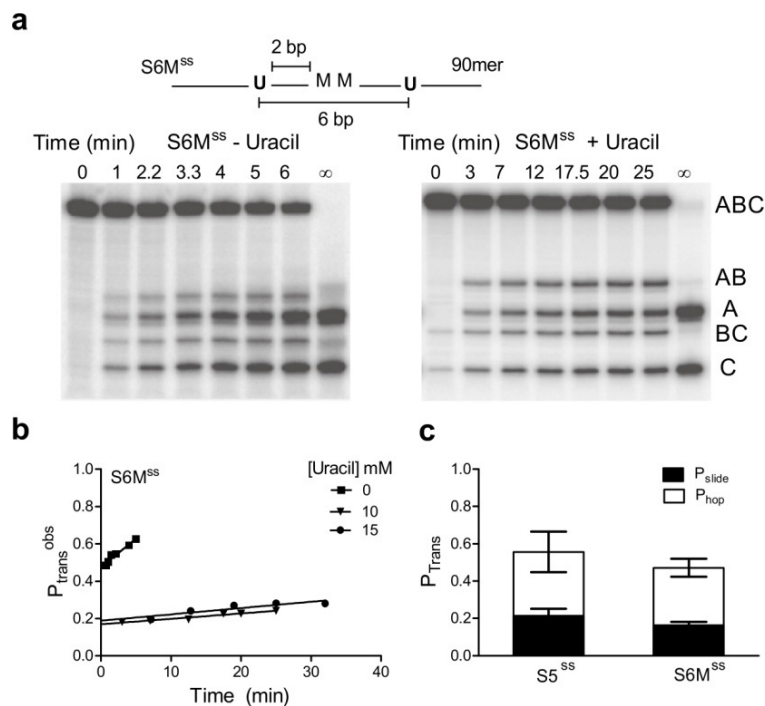
We next examined M linkages in the context of transfer of hUNG on single stranded DNA using a ssDNA substrate that contained two M linkages analogous to the duplex S6M (S6M<sup>ss</sup>). S6M<sup>ss</sup> was designed to have minimal secondary structure and to have no more than two adjacent Watson-Crick pairings to eliminate potential secondary structure. The data for S6M<sup>ss</sup> is summarized in **Figure 3.5** and show that M linkages have no effect on site transfer compared to the all phosphodiester ssDNA substrate S5<sup>ss</sup>. Comparing S6<sup>ss</sup> to S5<sup>ss</sup> is justified

because  $P_{\text{slide}}$  for ssDNA has a flat dependence with site spacing between 5 and 10 ntds (9) (see also Chapter 4).

The absence of a requirement for a continuous phosphate charge in sliding or hopping between two closely spaced sites in dsDNA or ssDNA is striking in comparison with the damaging effect of M substitutions on the  $K_m$  for a uracil-containing substrate ( $\Delta\Delta G = \sim 1-4$  kcal/mole depending on position) (22), the large and highly stereospecific 5-10 kcal/mol effects of single M substitutions on the activation barriers for uracil excision in single strand or dsDNA (22, 23), and the significant effects of M substitution on nonspecific DNA binding reported above. These differences suggest that the rapid kinetic process of nonspecific sliding does not involve the same phosphate backbone interactions observed in crystal structures of nonspecific and specific complexes between hUNG and DNA (26, 27).



**Figure 3.4. hUNG sliding and hopping is unaffected on double stranded DNA substrates containing intervening neutral methylphosphonate (M) linkages.** a) Schematic of the substrates used (S5M, S6M). Methylphosphonate positioning was chosen so that the catalytic excision of uracil by hUNG is unaffected (22, 23). b) Gel images of the site transfer products derived from S5M in the presence and absence of uracil. c) Determination of  $P_{trans}$ , where the observed site transfer ( $P_{trans}^{obs}$ , eq 1) is calculated at each time point and linearly extrapolated to zero time to determine the true site transfer value ( $P_{trans}$ ). Values at 10 and 15 mM uracil were identical indicating measurements were made within the plateau region depicted in Figure 1b. d) Summary of the site transfer properties of hUNG for double stranded methylphosphonate containing DNA substrates compared to the all phosphodiester versions. Data for S5 and S6 were reported previously and are shown for comparison (9). Values are equal to the mean  $\pm$  1 s.d. of at least 3 trials at 0 mM Uracil and 6 at high uracil (3 each at 10 and 15 mM uracil).



**Figure 3.5. hUNG sliding on ssDNA is unaffected by methylphosphonate substitutions.** (a) Gel images of the site transfer assay in the presence and absence of uracil for S6M<sup>ss</sup> which contains two intervening M substitutions. We note the presence of a small amount of cleavage in the zero time lane (<1%). This was found to be the result of the commercially available terminal deoxynucleotidyl transferase used in the 3' end labeling having a very small amount of uracil DNA glycosylase activity over the >2 hr incubation at 37 °C, likely from copurification. These background bands were determined to have no effect on the site transfer calculations. (b) Linear extrapolation of  $P_{trans}^{obs}$  for S6M<sup>ss</sup> in the presence and absence of uracil to determine  $P_{trans}$ . (c) Comparison of the site transfer measurements of S6M<sup>ss</sup> to that of the all phosphodiester substrate S5<sup>ss</sup>. Comparison with S5<sup>ss</sup> is reasonable because site transfer by sliding on ssDNA is flat for site spacings between 5 and 10 ntds (see text).

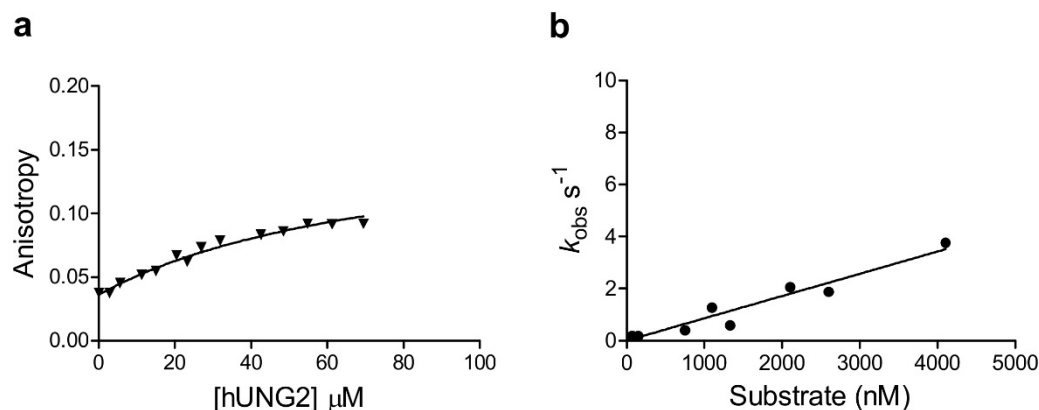
### 3.2.5 hUNG site transfer using physiological ion concentrations

Site transfer measurements published previously have been studied under conditions where [NaCl] ranged from 22 to 72 mM. In this range, sliding was found to be insensitive to salt, but hopping was fully ablated at salt concentrations exceeding 42 mM (9). However, it is desirable to evaluate these parameters under conditions that more closely mimic the intracellular ion concentrations. For this purpose we use a buffer consisting of 140 mM potassium glutamate, 10 mM Na-HEPES pH 7.5 and 200  $\mu$ M MgSO<sub>4</sub>.

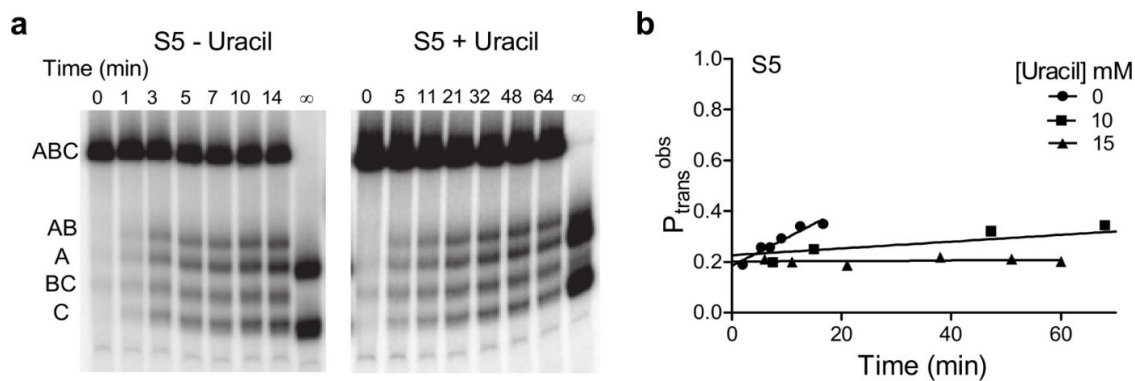
Measurements of equilibrium nonspecific DNA binding and hUNG catalytic activity under these conditions were first made. Compared to low ionic strength conditions, the equilibrium dissociation constant for nonspecific binding was increased  $\sim$ 100 fold ( $0.82 \pm 0.26 \mu$ M to  $85 \pm 19 \mu$ M) (**Figure 3.6a**). Similarly, the catalytic activity of hUNG for a 90mer DNA duplex substrate containing a single uracil site was reduced  $\sim$ 300 fold under these conditions with a measured  $k_{\text{cat}}/K_m = 1 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$  (**Figure 3.6b**). This value may be compared with the previously measured  $k_{\text{cat}}/K_m = 3.4 \times 10^8 \text{ M}^{-1}\text{s}^{-1}$  for the identical substrate at low ionic strength. We note that accurate determinations of  $k_{\text{cat}}$  and  $K_m$  were not possible under physiological salt conditions, but a good estimate of  $k_{\text{cat}}/K_m$  could be obtained from the linear increase in rate using 0 - 4  $\mu$ M substrate (**Figure 3.6b**). Most of the effect on activity is assignable to  $K_m$  because the maximal observed rates under physiological salt concentrations approached the  $k_{\text{cat}}$  value of  $5 \text{ s}^{-1}$  under low salt conditions (9).



Site transfer measurements were then made using the physiological buffer with duplex substrate S5, which contains two uracils positioned 5 bp apart (**Figure 3.7**). At this spacing, intramolecular site transfer was still observed ( $P_{\text{trans}} = 0.20 \pm 0.04$ ). The transfer probability was similar at high uracil ( $P_{\text{slide}} = 0.16 \pm 0.06$ ), indicating that within error of the measurements all transfers occur by sliding, although  $P_{\text{slide}}$  was reduced compared to the value at low ionic strength ( $P_{\text{slide}} = 0.37 \pm 0.06$ ). This result matches the previous finding that the hopping pathway is eliminated with a salt concentration of 42 mM while sliding persists unabated at the same concentration (9). Thus despite the 100-fold decrease in nonspecific binding affinity under mock physiological conditions, short-range sliding of hUNG on DNA can still occur. These findings imply that the binding interface of the sliding form of hUNG is immune to invasion by salt ions. This observation is consistent with the inability of the uracil trap to access the active site of hUNG during the process of sliding.



**Figure 3.6. Steady-state kinetics and non-specific DNA binding of hUNG under physiological salt conditions** (140 mM potassium glutamate, 200 μM MgSO<sub>4</sub>, 10 mM Na-HEPES pH 7.5). (a) Binding of hUNG to a 10mer FAM-labeled duplex DNA (NS10, 50 nM) performed as described in the Methods section (main text). The dissociation constant from least squares fitting to a one site binding isotherm was  $85 \pm 19$  μM. (b) Steady-state reaction kinetics for uracil excision from a 90mer DNA duplex substrate (1XU\_90mer) by hUNG. Rates were measured as previously described (2), by resolving the <sup>32</sup>P labeled DNA substrate from product using electrophoresis. Due to product inhibition, rates were not able to be measured accurately at substrate concentrations above 4 μM.  $k_{cat}/K_m$  was determined to be  $1 \times 10^6$  M<sup>-1</sup> s<sup>-1</sup> from the slope of the linear increase in rate. The hUNG concentration ranged from 0.2 nM to 1.6 nM.



**Figure 3.7. Site transfer measurements of hUNG under approximated physiological ionic strength conditions** (140 mM potassium glutamate, 10 mM Na-HEPES pH 7.5, 200  $\mu$ M MgSO<sub>4</sub>). (a) Gel images of the raw site transfer measurements in the presence and absence of uracil. (b) Determination of  $P_{trans}$  by linear extrapolation. The values are the same in the presence and absence of uracil indicating that site transfer occurs entirely by sliding under these high salt conditions.

### 3.3 Discussion

#### 3.3.1 Different Effects of M Substitution on Nonspecific DNA Binding, DNA Translocation and Uracil Excision

Comparison of the divergent effects of methylphosphonate (M) substitutions on nonspecific DNA binding, translocation between uracil sites, and catalysis by UNG leads to the conclusion that the requirements for a charged phosphate group in these processes are very different. Previous studies where stereospecific M substitutions were made in single stranded and duplex substrates of UNG have revealed that substitutions at the +1, -1, and -2 phosphates surrounding the uracil site (5'p<sup>+1</sup>Up<sup>-1</sup>Np<sup>-2</sup>N3') result in stereospecific 10<sup>1</sup>-10<sup>8</sup> fold damaging effects on catalysis (22, 23). Most of these large effects were attributed to the beneficial energetic effects of the anionic phosphate groups towards stabilization of the glycosyl cation transition state. The previously measured damaging effects of single M substitutions on the ground state Michaelis complex were not stereospecific and were less than the effects on the activation barrier (i.e.  $K_m$  effects were in the range 10-100 fold) (22, 23). The even smaller damaging effect of a single M substitution on nonspecific DNA binding (~2-fold, **Figure 3.3c**), would suggest that the nonspecific complex differs in its interactions with the phosphate backbone as compared to the Michaelis complex. Despite the apparent differences between these complexes revealed by M substitution, the high resolution crystal structures of the specific and nonspecific hUNG-DNA complexes show that the same phosphate groups form hydrogen bonds with neutral serine or histidine side chains, or backbone amide

groups, and that there are few cationic groups  $\leq 3.3$  Å from phosphate oxygens (**Figure 3.8**) (26). Thus, taken together, these energetic measurements suggest that as the enzyme moves forward along the reaction coordinate it forms increasingly important electrostatic interactions with the phosphate backbone. The interesting exception, as shown in this work, is the transient state for DNA sliding which apparently has no requirement for an uninterrupted charged phosphate chain.

What is the physical basis for the different effects of M substitution on nonspecific binding and DNA translocation? Although M substitution has only a minor effect on B DNA structural parameters (28-30), this substitution can change duplex hydration patterns (29, 31) and quite possibly reduce the ion count in the cloud loosely associated with the DNA (32, 33). Thus, these indirect outcomes of M substitution can make unique mechanistic interpretations of the observed effects elusive. In the present case, the small damaging effects of M substitution on nonspecific DNA binding (0.5 kcal/mol per substitution) could reflect direct disruption of the backbone hydrogen bonding in the complex (**Figure 3.8**), or a reduction in minor groove hydration waters or ions around the neutral patch (29, 31). If these indirect effects prevail, then the reduction in binding affinity upon M substitution could arise from a smaller favorable entropy change resulting from fewer water molecules or ions being released to bulk solution upon complexation (34). Although such indirect effects might provide viable explanations for the reduced binding affinity of hUNG for M substituted DNA, they do not reasonably account for the absence of an effect of M

substitution on DNA sliding because sliding occurs in a kinetic event after the ion cloud has been dispersed.

The absence of a functional requirement for a continuous negatively charged backbone in site translocation strongly suggests that the sliding form of hUNG cannot simply involve translocation of the crystallographic conformation of hUNG along DNA (26, 27, 35, 36). Rather, the data would suggest that the sliding conformation of hUNG is an open state that interacts loosely with the DNA backbone, with perhaps intervening water molecules (but not solute ions) that would serve to shield charge. This view of a loose, transiently bound conformation is consistent with CPMG NMR dynamic measurements indicating that UNG oscillates between an open and closed form on the millisecond time scale when bound to nonspecific DNA (37). The open form was proposed to function in stochastic sliding along the DNA chain, and the closed form resembles the crystallographic conformation, allowing hUNG to interrogate the integrity of base pairs. Indeed, a two state conformational change has been postulated as a general mechanism for site specific DNA binding proteins to overcome the “search-speed/stability” paradox (3, 4, 38), and recent structural evidence obtained with other DNA glycosylases suggests evidence for more than one conformation involved in search and recognition by these enzymes (12, 13). The findings reported here provide a first glimpse at the electrostatic properties of this transient state of hUNG.

### ***3.3.2 Boundary Estimates for 1D Translocation on Duplex DNA***

Employing the measured values for the average lifetime of hUNG on nonspecific DNA ( $\tau_{\text{bind}} = 3 \text{ ms}$ ), and its mean sliding length ( $L_{\text{slide}} = 4.2 \text{ bp}$ ), we previously used eq 3.2 to estimate the 1-dimensional diffusion constant ( $D_1$ ) of hUNG on nonspecific duplex DNA ( $D_1 = 6 \times 10^3 \text{ bp}^2 \text{ s}^{-1} = 7 \times 10^{-4} \text{ } \mu\text{m}^2 \text{ s}^{-1}$ ) (9). This value was several orders of magnitude below the theoretical upper limit ( $\sim 10^7 \text{ bp}^2 \text{ s}^{-1}$  or  $\sim 1 \text{ } \mu\text{m}^2 \text{ s}^{-1}$ ) (5, 39, 40).

$$[3.2] \quad D_1 = \frac{L_{\text{slide}}^2}{\tau_{\text{bind}}}$$

This calculation assumes that the entire bound lifetime of hUNG is available for DNA sliding. However the current data, which requires the presence of at least two nonspecific states of hUNG, also requires that only a fraction of the bound lifetime is available for sliding (i.e the time spent in the open state). An estimated lower limit for the population of the transient sliding state may be estimated based on the sensitivity of the NMR-relaxation dispersion dynamic measurements previously performed on the hUNG-nonspecific DNA complex (37). This methodology would not be able to detect a transient sliding state with a population of less than  $\sim 5\%$  of the total, setting a lower limit for the time spent sliding of  $0.05 \times 3 \text{ ms} \geq 0.15 \text{ ms}$ . It is difficult to set an upper boundary, but it must be considerably less than  $\tau_{\text{bind}} = 3 \text{ ms}$ . Using eq 2 and this lower limit for the sliding time, we calculate an upper limit for  $D_1 \leq 10^5 \text{ bp}^2 \text{ s}^{-1}$ . Thus, the previous and current estimates place  $D_1$  in the range  $\sim 10^4$  to  $10^5 \text{ bp}^2 \text{ s}^{-1}$ . Given this refined two-state view of sliding by hUNG, we suggest that the sliding state

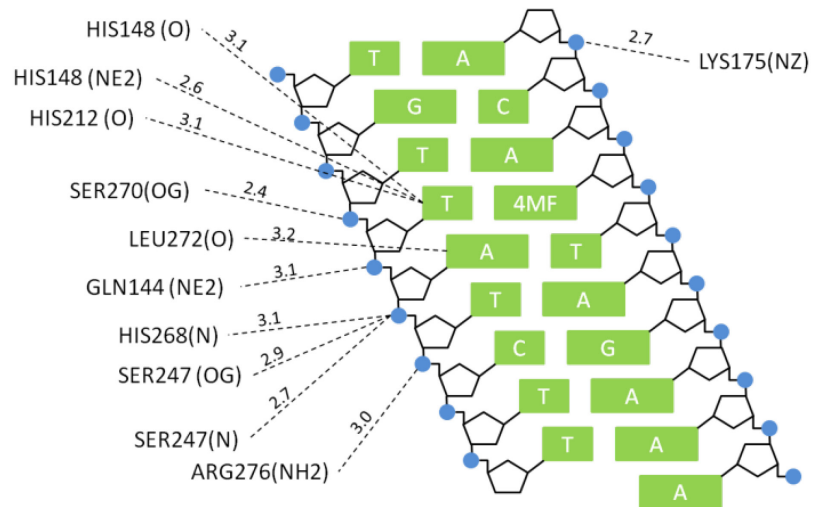
resembles the transition state for DNA dissociation. However, instead of falling off the DNA chain the enzyme closes on the DNA and completes a sliding transfer. This viewpoint of short range sliding as an aborted transition state for DNA dissociation differs considerably from other characterizations of protein sliding whereby the protein moves isoenergetically along the surface of the DNA (41-43). These aspects of the hUNG search mechanism are depicted in the model presented in **Figure 3.9**.

### **3.3.3 Search and Recognition in the Cell Nucleus**

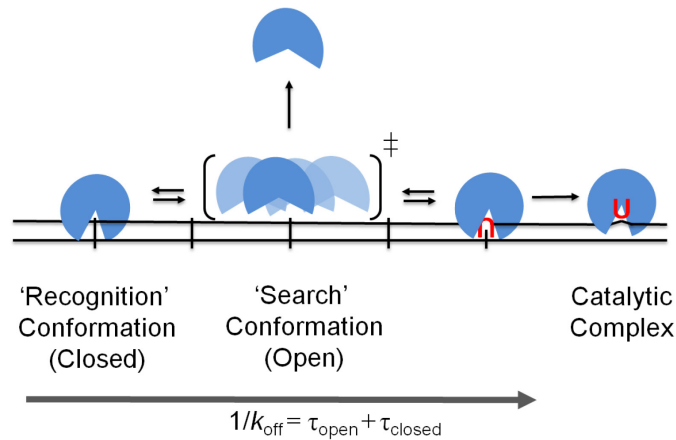
Under optimal low salt reaction conditions hUNG is an evolutionarily optimized enzyme with a catalytic power that vastly exceeds any other DNA glycosylase (1). However, under conditions that more closely mimic the intracellular ionic environment, its ability to bind nonspecific DNA is severely hampered by a factor of around 100-fold, which exerts a profound effect on the mechanism of site location. One major ramification of the ionic strength effect on nonspecific DNA binding is that hopping becomes a less productive pathway. Each time hUNG dissociates from the DNA chain under high salt conditions, there will be a reduced probability that a reassociation attempt will result in a productive binding event. Thus many more attempts will have to be made, which will result in an increase in the search time contributed by hopping. In contrast, DNA sliding is largely refractory to increases in ionic strength, and the search time resulting from sliding will remain largely unchanged. This important property of sliding, even over the short ranges traveled by hUNG, is essential for



increasing coverage of the genome and for the ultimate detection of damage (**Figure 3.9**). An additional consideration within the nuclear environment is the effect of crowding, as well as excluded volume effects (44, 45). Such factors could favor compact sliding states and also increase the contribution of hopping because of the high local concentration of DNA chains. Consideration of such effects requires improved experimental models for search and recognition.



**Figure 3.8. DNA interactions of hUNG non-specifically bound to a destabilized thymine basepair** (PDB ID: 2OXM (27), 4MF = 4-methylindole). Residues shown have a nitrogen or oxygen atom  $< 3.3 \text{ \AA}$  of a nitrogen atom of a DNA base or a phosphate oxygen of the DNA backbone.



**Figure 3.9. Two-state model for hUNG sliding on nonspecific DNA.** In this model hUNG exists in a highly populated “closed” recognition conformation that makes multiple interactions with the phosphate backbone as observed in crystallographic studies, and also a transient mobile “open” sliding conformation that makes little or no interactions with the phosphate backbone (this work). It is reasonable to view the open state as the aborted transition state preceding DNA dissociation. The open state, which must be present at least 5% of the total bound lifetime (see text), allows for fast movement on the DNA while also allowing time for recognition of uracil bases when they are encountered.

## 3.4 Methods

### 3.4.1 Protein and Oligonucleotide Reagents

hUNG was purified as previously described (9). Protein concentrations were determined by absorbance measurements at 280 nm using an extinction coefficient of  $33.68 \text{ mM}^{-1} \text{ cm}^{-1}$ . Oligonucleotides except for those containing methylphosphonate linkages were ordered from Integrated DNA Technologies ([www.IDTDNA.com](http://www.IDTDNA.com)) in the crude desalted form and purified by denaturing PAGE. All oligonucleotide sequences are reported in the Supplemental Methods and concentrations were determined by UV absorption at 260 nm using extinction coefficients calculated from nearest neighbor parameters.

### 3.4.2 Oligonucleotide Sequences

Oligonucleotide substrates are named as follows, with the number indicating the spacing of the uracil sites and the superscript “ss” indicating single stranded DNA, while otherwise the substrates were used in the duplex form. The lowercase m indicates the presence of a methylphosphonate (M) linkage. Importantly, the single stranded substrates were designed to have no more than two possible adjacent Watson-Crick pairings as determined by the mFold hybridization prediction (1) (Supplementary Figure S2).

Oligonucleotides for non-specific binding

*NS10:*

FAM – GGC ACA CGC G

*NS10M:*

FAM – GGCm ACmA CmGCm G

*NS5:*

FAM AGCGA

*NS5M:*

FAM-AGCmGA

*NS6:*

FAM-AGC CGA

*NS6M:*

FAM-AGCmCmGA

Oligonucleotides used for site transfer measurements

*S5M\_90mer:*

5' - GGT ATC CGC TGAA GTA GTC ACA ATT CCA CAC AAT GCT GAG GA A  
TCG AUA G (Cme) G AUA GCT A AG CTG AGG CAT ACA GGA TCA ATT  
GTC GAG CC

(Cme = Cytosine methylphosphonate linkage)

*Precursor oligonucleotides:*

*S5M\_piece2:*

A TCG AUA G Cme G AUA GCT A

*S5M\_piece1:*

GGT ATC CGC TGAA GTA GTC ACA ATT CCA CAC AAT GCT GAG GA

*S5M\_piece3:*

AG CTG AGG CAT ACA GGA TCA ATT GTC GAG CC

*S5M\_ligationsplint:*

TGA TCC TGT ATG CCT CAG CTT AGC TAT CGC TAT CGA TTC CTC  
AGC ATT GTG TGG AAT

*S6M\_90mer:*

GGT ATC CGC TGAA GTA GTC AC A AT CCA CAC AAT GCT GAG GAA TCG  
AUA G Cme Cme G AUA GCT AAG CTG AGG CAT ACA GGA TCA ATT GTC  
GAG CC

(Cme = Cytosine methylphosphonate linkage)

*Precursor oligonucleotides:*

*S6\_piece2:*

A TCG AUA G Cme Cme G AUA GCT A

(Cme = Cytosine methylphosphonate linkage)

S6\_piece1:

GGT ATC CGC TGAA GTA GTC ACA AT CCA CAC AAT GCT GAG GA

S6\_piece3:

AG CTG AGG CAT ACA GGA TCA ATT GTC GAG CC

S6\_ligationsplint:

TGA TCC TGT ATG CCT CAG CTT AGC TAT CGG CTA TCG ATT CCT  
CAG CAT TGT GTG GAT T

*S6M<sup>ss</sup>\_90mer:*

ACC ATA ATA ATA ACA CAT ACA CCA TAC TAC ATA CAT CAA CTA AAA  
CAU AC Ame Ame CAU ACA AAA TCA ACT AAT AAC AAC ACATAC ACC TA  
ACA

(Ame = Adenine methylphosphonate linkage)

*Precursor oligonucleotides:*

*S6M<sup>ss</sup>\_piece2:*

A AA CA U A C Ame Ame C A U ACA AA

*S6M<sup>ss</sup>\_piece1:*

ACC ATA ATA ATA ACA CAT ACA CCA TAC TAC ATA CAT CAA CTA

*S6ss\_piece3:*

A TCA ACT AAT AAC AAC ACATAC ACC TA ACA

*S6ss\_spline:*

TGT TGT TAT TAG TTG ATT TTG TAT GTT GTA TGT TTT AGT TGA  
TGT ATG TAG TAT G

Oligonucleotides used for steady-state kinetic and non-specific DNA binding measurements:

*1XU\_90mer:*

GTT ATC CGC TCA CAA TTC CAC ACA ATG CTG AGG AAT CGA UAG CTA  
AGT AGG ATG TTA GCT ATC GAT TCA TCC TCA GCA CAG TGT CGA GCC

### **3.4.3 Experimental conditions**

All measurements were made at 37 °C in a standard reaction buffer consisting of 20 mM HEPES pH 7.5, 0.002% Brij 35 detergent (Sigma Aldrich),

3 mM EDTA (added from a 0.5 M pH 8.0 stock), and 1 mM DTT unless otherwise noted.

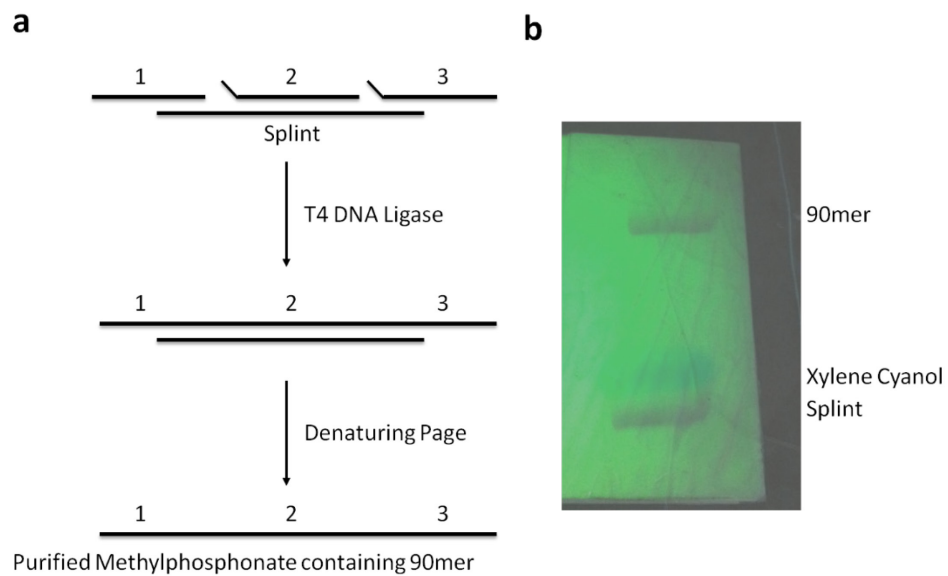
*Synthesis of oligonucleotides containing methylphosphonate linkages.*

Oligonucleotides containing methylphosphonate linkages were synthesized using standard phosphoramidite synthesis procedures on an Applied Biosystems 390 DNA/RNA synthesizer. Nucleoside phosphoramidites and methylphosphoramidites were purchased from Glenn Research (Sterling, VA). After synthesis, the DNA was deprotected and cleaved from the silica support by the addition of 0.5 mL 45:45:10 acetonitrile/ethanol/ammonium hydroxide and allowed to incubate at room temperature for 30 minutes. 0.5 mL of ethylenediamine was then added and the DNA containing solution was allowed to sit overnight at room temperature. The DNA containing solution was separated from the silica support and dried under vacuum. After resuspension in 25 mM Tris-HCl pH 7.5 (Buffer A) the DNA was then purified from the failure products by HPLC by injection onto a Dionex<sup>TM</sup> DNA Pac anion exchange column and eluted with a linear gradient from 10% Buffer A to 90% Buffer B (25 mM Tris-HCl pH 7.5, 1M NaCl).

90mer oligonucleotide substrates used in the site transfer assays containing methylphosphonate linkages were first synthesized as smaller precursors and then a 3-piece ligation was performed to create the final product (Supplemental Figure S1). Sequences of the precursor oligonucleotides are listed in the Supplemental Methods. For ligation, piece 1 (1.5 nanomoles) and piece 2 containing methylphosphonate linkages (2 nanomoles) were first phosphorylated

at the 5' end by incubation with T4 PNK (New England Biolabs) at 37 °C in a single reaction mixture (~200 µL volume in 1X DNA ligase buffer, New England Biolabs™). After inactivation of T4 PNK, Piece 3 (2 nmoles) and the Splint (2 nanomoles) were then added to the reaction mix and hybridized by heating to 95 °C for 5 minutes and allowed to cool slowly to room temperature by placing the heat block on the bench top. Fresh ATP was then added to 1 mM final concentration along with T4 DNA ligase (New England Biolabs) and the reaction mixture was incubated at 37 °C overnight. The reaction was then mixed with 50% formamide (final concentration) and the ligated product was purified by denaturing PAGE (**Figure 3.10**).





**Figure 3.10 Splint ligation strategy for synthesis and purification of methylphosphonate (M) containing oligonucleotides. UV shadowing of 10% denaturing gel of purified 90mer DNA containing methylphosphonate linkage.**

### **3.4.4 Determination of DNA Dissociation Constants by Fluorescence Anisotropy**

Binding of hUNG to non-specific DNA was determined using fluorescence anisotropy in a Spex Fluormax 3 fluorimeter at 37 °C. Concentrated hUNG in the standard reaction buffer containing 50 nM labeled DNA was titrated into a cuvette containing 50 nM labeled DNA in reaction buffer in order to avoid dilution of the DNA during the titration. After each addition the cuvette was placed in the fluorimeter and allowed to equilibrate for 2 minutes as the reading was found to stabilize after 60-90 seconds. For dissociation constants greater than 5  $\mu$ M,  $K_D$  values were determined by diluting a concentrated solution of hUNG in reaction buffer and 50 nM labeled DNA with a solution of labeled DNA only. The forward titration was found to overlay titrations performed by dilution indicating that anisotropy values were determined at equilibrium. Data were then fitted to a single site binding isotherm (anisotropy =  $B_{\max} \times [\text{hUNG}]_{\text{free}} / (K_D + [\text{hUNG}]_{\text{free}}) + B_{\min}$ ), where  $B_{\max}$  and  $B_{\min}$  are the maximal and minimum anisotropies, and it was assumed that the free DNA concentration equals the total (which is a valid assumption given that the  $K_D \gg [\text{DNA}]_{\text{total}}$ ).

### **3.4.5 Intramolecular Site Transfer Assay**

Site transfer measurements were performed identically as before (9, 11) with some modifications in the steps after reaction quenching to account for substrate and buffer differences. The DNA concentration in all site transfer measurements was 40 nM and the hUNG concentration ranged from 10 – 20  $\mu$ M

under the standard reaction conditions, and for the data presented in Figure 6 from 300 pM to 1.5 nM.

For methylphosphonate containing duplexes (S5M and S6M), 30 picomoles of the top and bottom DNA strands were 5'-end labeled with  $^{33}\text{P}$  by incubation with T4 polynucleotide kinase and  $[\gamma^{33}\text{P}]$  ATP in separate reactions. The reactions were then mixed and the strands hybridized by heating to 95 °C for 10 minutes in a dry heat block followed by slow cooling to room temperature by placing the heat block on the bench top. The hybridized DNA was then separated from the unincorporated  $[\gamma^{33}\text{P}]$  ATP by gel filtration using P30 resin (BioRad<sup>TM</sup>) and then desalted using P6 resin (BioRad<sup>TM</sup>). Samples obtained before and after gel filtration were analyzed by native gel electrophoresis, where percent recovery was calculated from imaging of the band densities, and completeness of hybridization was confirmed. In general, the percent recovery was at least 80%. After reaction with hUNG and quenching by uracil DNA glycosylase inhibitor protein (UGI, New England Biolabs<sup>TM</sup>) each individual reaction was treated with the nicking enzyme Nt.BbvCI and APE1 endonuclease as previously described (9, 11) resulting in discrete double-stranded fragments corresponding to the hUNG reaction products. Each sample was then separated by electrophoresis on a 0.5 millimeter thick 10% native gel (19:1 bis:acrylamide) run in 1X TBE buffer at 20 Watts in a model S2 sequencing gel for 1 hour and 40 minutes without pre-running the gel.

For S6M<sup>SS</sup> the 5' and 3' ends were labeled by incubation with  $[\gamma^{32}\text{P}]$  ATP and 3'-deoxyadenosine 5'-triphosphate (cordycepin 5'-triphosphate)- $[\alpha\text{-}^{32}\text{P}]$  using

polynucleotide kinase and terminal transferase (New England Biolabs), respectively. Similarly as above for the duplex substrates, after radiolabeling the unincorporated nucleotides and excess salts were removed by gel filtration using P30 and P6 resins (Biorad™). After reaction with hUNG and quenching, the resulting abasic sites were cleaved by the addition of 0.25 M ethylenediamine pH 8.0 (final concentration) followed by immediate heating to 95 °C for 5 minutes. Formamide containing both xylene cyanol and bromphenol blue was then added to 65% final concentration and the samples were loaded onto a 10% denaturing gel (19:1 bis:acrylamide).

For the duplex substrate S5 under physiological salt conditions (140 mM potassium glutamate, 200 µM MgSO<sub>4</sub>, 10 mM Na-HEPES, pH 7.5) the uracil containing strand was first labeled with <sup>32</sup>P at the 5' and 3' ends as described for S6M<sup>ss</sup> above. The labeled strand was then hybridized to the complementary strand and the unincorporated radiolabel was removed using P30 resin (Biorad™). Forty nanomolar of the duplex substrate was then reacted with hUNG and quenched at various time points using UGI as described above. To each aliquot, 3 µl of 0.25 M ethylenediamine pH 8.0 was added and the reaction was immediately heated to 95 °C for 5 minutes to cleave the DNA at the abasic sites. Formamide gel loading buffer was then added to 65% final concentration and the samples were heated at 95 °C for an additional 3 minutes. The samples were immediately loaded onto a pre-heated 10% (19:1 bis:acrylamide) denaturing gel in order to fully denature any residual structure.

### **3.4.6 Analysis of the Site Transfer Data**

All gels were exposed to a storage phosphor screen and digitized using a phosphorimager. For each reaction time course, product band densities were quantified in QuantityOne™ using the box method. More details concerning the data analysis are presented in the Results. All errors presented in the text are standard deviations derived from at least three independent measurements.

### 3.5: References

1. Stivers, J. T., and Jiang, Y. L. (2003) A mechanistic perspective on the chemistry of DNA repair glycosylases., *Chem. Rev.* *103*, 2729–2759.
2. Friedman, J. I., and Stivers, J. T. (2010) Detection of Damaged DNA Bases by DNA Glycosylase Enzymes, *Biochemistry* *49*, 4957–4967.
3. Mirny, L., Slutsky, M., Wunderlich, Z., Tafvizi, A., Leith, J., and Kosmrlj, A. (2009) How a protein searches for its site on DNA: the mechanism of facilitated diffusion, *J. Phys. A: Math. Theor.* *42*, 434013.
4. Slutsky, M., and Mirny, L. A. (2004) Kinetics of protein-DNA interaction: facilitated target location in sequence-dependent potential., *Biophys J* *87*, 4021–4035.
5. Berg, O. G., Winter, R. B., and Hippel, Von, P. H. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory, *Biochemistry* *20*, 6929–6948.
6. Halford, S. E., and Marko, J. F. (2004) How do site-specific DNA-binding proteins find their targets?, *Nucleic Acids Res.* *32*, 3040–3052.
7. Gowers, D. M., Wilson, G. G., and Halford, S. E. (2005) Measurement of the contributions of 1D and 3D pathways to the translocation of a protein along DNA., *Proc. Natl. Acad. Sci. U.S.A.* *102*, 15883–15888.
8. Hedglin, M., and O'Brien, P. J. (2010) Hopping Enables a DNA Repair Glycosylase To Search Both Strands and Bypass a Bound Protein, *ACS Chemical Biology* *5*, 427–436.

9. Schonhofs, J. D., and Stivers, J. T. (2012) Timing facilitated site transfer of an enzyme on DNA., *Nat. Chem. Biol.* 8, 205–210.
10. Blainey, P. C., van Oijen, A. M., Banerjee, A., Verdine, G. L., and Xie, X. S. (2006) A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA., *Proc. Natl. Acad. Sci. U.S.A.* 103, 5752–5757.
11. Porecha, R. H., and Stivers, J. T. (2008) Uracil DNA glycosylase uses DNA hopping and short-range sliding to trap extrahelical uracils., *Proc. Natl. Acad. Sci. U.S.A.* 105, 10791–10796.
12. 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., *Proc. Natl. Acad. Sci. U.S.A.* 109, 1086–1091.
13. Setser, J. W., Lingaraju, G. M., Davis, C. A., Samson, L. D., and Drennan, C. L. (2012) Searching for DNA lesions: structural evidence for lower- and higher-affinity DNA binding conformations of human alkyladenine DNA glycosylase., *Biochemistry* 51, 382–390.
14. Dunn, A. R., Kad, N. M., Nelson, S. R., Warshaw, D. M., and Wallace, S. S. (2011) Single Qdot-labeled glycosylase molecules use a wedge amino acid to probe for lesions while scanning along DNA, *Nucleic Acids Res.*
15. Terakawa, T., Kenzaki, H., and Takada, S. (2012) p53 Searches on DNA by Rotation-Uncoupled Sliding at C-Terminal Tails and Restricted Hopping of Core Domains., *J. Am. Chem. Soc.* 134, 14555–14562.

16. Leith, J. S., Tafvizi, A., Huang, F., Uspal, W. E., Doyle, P. S., Fersht, A. R., Mirny, L. A., and van Oijen, A. M. (2012) Sequence-dependent sliding kinetics of p53., *Proc. Natl. Acad. Sci. U.S.A.*
17. li, A. G., Yeykal, C. C., Robertson, R. B., and Greene, E. C. (2006) Long-distance lateral diffusion of human Rad51 on double-stranded DNA, *Proc. Natl. Acad. Sci. U.S.A.* 103, 1221–1226.
18. Senavirathne, G., Jaszczur, M., Auerbach, P. A., Upton, T. G., Chelico, L., Goodman, M. F., and Rueda, D. (2012) Single-stranded DNA scanning and deamination by APOBEC3G cytidine deaminase at single molecule resolution., *J. Biol. Chem.* 287, 15826–15835.
19. Hammar, P., Leroy, P., Mahmutovic, A., Marklund, E. G., and al, E. (2012) The lac repressor displays facilitated diffusion in living cells, *Science* 336, 1595–1598.
20. Stanford, N. P., Szczelkun, M. D., Marko, J. F., and Halford, S. E. (2000) One- and three-dimensional pathways for proteins to reach specific DNA sites., *EMBO J.* 19, 6546–6557.
21. Werner, R. M., Jiang, Y. L., Gordley, R. G., Jagadeesh, G. J., Ladner, J. E., Xiao, G., Tordova, M., Gilliland, G. L., and Stivers, J. T. (2000) Stressing-Out DNA? The Contribution of Serine–Phosphodiester Interactions in Catalysis by Uracil DNA Glycosylase, *Biochemistry* 39, 12585–12594.



22. Jiang, Y. L., Ichikawa, Y., Song, F., and Stivers, J. T. (2003) Powering DNA Repair through Substrate Electrostatic Interactions, *Biochemistry* 42, 1922–1929.
23. Parker, J. B., and Stivers, J. T. (2008) Uracil DNA glycosylase: revisiting substrate-assisted catalysis by DNA phosphate anions., *Biochemistry* 47, 8614–8622.
24. Terry, B. J., Jack, W. E., and Modrich, P. (1985) Facilitated diffusion during catalysis by EcoRI endonuclease. Nonspecific interactions in EcoRI catalysis., *J. Biol. Chem.* 260, 13130–13137.
25. Porecha, R. H., and Stivers, J. T. (2008) Uracil DNA glycosylase uses DNA hopping and short-range sliding to trap extrahelical uracils, *Proc. Natl. Acad. Sci. U.S.A.* 105, 10791–10796.
26. Slupphaug, G., Mol, C. D., Kavli, B., Arvai, A. S., Krokan, H. E., and Tainer, J. A. (1996) A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA., *Nature* 384, 87–92.
27. Parker, J. B., Bianchet, M. A., Krosky, D. J., Friedman, J. I., Amzel, L. M., and Stivers, J. T. (2007) Enzymatic capture of an extrahelical thymine in the search for uracil in DNA, *Nature* 449, 433–437.
28. Thiviyanathan, V., Vyazovkina, K. V., Gozansky, E. K., Bichenchova, E., Abramova, T. V., Luxon, B. A., Lebedev, A. V., and Gorenstein, D. G. (2002) Structure of Hybrid Backbone Methylphosphonate DNA Heteroduplexes: Effect of R and S Stereochemistry †,‡, *Biochemistry* 41, 827–838.

29. Hamelberg, D., Williams, L. D., and Wilson, W. D. (2002) Effect of a neutralized phosphate backbone on the minor groove of B-DNA: molecular dynamics simulation studies., *Nucleic Acids Res* 30, 3615–3623.
30. Maher, L. D. W. A. L. J., III, Williams, L. D., and Maher, L. J., III. (2013) Electrostatic mechanisms of DNA deformation., *Annu Rev Biophys Biomol Struct* 29, 497–521.
31. Hausheer, F. H., Rao, B. G., Saxe, J. D., and Singh, U. C. (1992) Physicochemical properties of (R)- vs (S)-methylphosphonate substitution on antisense DNA hybridization determined by free energy perturbation and molecular dynamics, *J. Am. Chem. Soc.* 114, 3201–3206.
32. Bai, Y., Chu, V. B., Lipfert, J., Pande, V. S., Herschlag, D., and Doniach, S. (2008) Critical Assessment of Nucleic Acid Electrostatics via Experimental and Computational Investigation of an Unfolded State Ensemble, *J. Am. Chem. Soc.* 130, 12334–12341.
33. Das, R., Mills, T., Kwok, L., Maskel, G., Millett, I., Doniach, S., Finkelstein, K., Herschlag, D., and Pollack, L. (2003) Counterion Distribution around DNA Probed by Solution X-Ray Scattering, *Physical Review Letters* 90, 188103.
34. Privalov, P. L., Dragan, A. I., and Crane-Robinson, C. (2011) Interpreting protein/DNA interactions: distinguishing specific from non-specific and electrostatic from non-electrostatic components., *Nucleic Acids Res.* 39, 2483–2491.

35. Mol, C. D., Arvai, A. S., Slupphaug, G., Kavli, B., Alseth, I., Krokan, H. E., and Tainer, J. A. (1995) Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis., *Cell* 80, 869–878.
36. Parikh, S. S., Mol, C. D., Slupphaug, G., Bharati, S., Krokan, H. E., and Tainer, J. A. (1998) Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA, *EMBO J.* 17, 5214–5226.
37. 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 Res.* 37, 3493–3500.
38. Zhou, H.-X. (2011) Rapid search for specific sites on DNA through conformational switch of nonspecifically bound proteins., *Proc. Natl. Acad. Sci. U.S.A.* 108, 8651–8656.
39. Schurr, J. M. (1979) The one-dimensional diffusion coefficient of proteins absorbed on DNA: Hydrodynamic considerations, *Biophysical chemistry* 9, 413–414.
40. Bagchi, B., Blainey, P. C., and Xie, X. S. (2008) Diffusion Constant of a Nonspecifically Bound Protein Undergoing Curvilinear Motion along DNA, *J. Phys. Chem. B* 112, 6282–6284.
41. Berg, O. G., and Hippel, von, P. H. (1985) Diffusion-controlled macromolecular interactions., *Annu Rev Biophys Biophys Chem* 14, 131–160.

42. Winter, R. B., Berg, O. G., and Hoppel, P. H. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The Escherichia coli lac repressor-operator interaction: kinetic measurements and conclusions, *Biochemistry* 20, 6961–6977.
43. Winter, R. B., and Hoppel, P. H. (1982) How do genome-regulatory proteins locate their DNA target sites?, *Trends Biochem. Sci.* 7, 52–55.
44. Zhou, H.-X., Rivas, G., and Minton, A. P. (2008) Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences., *Annu Rev Biophys* 37, 375–397.
45. Schreiber, G., Haran, G., and Zhou, H.-X. (2009) Fundamental aspects of protein-protein association kinetics., *Chem. Rev.* 109, 839–860.

## Chapter 4

# DNA Translocation by Human Uracil DNA Glycosylase; The Case of ssDNA and Clustered Uracils

Reproduced in part from:

Schonhoff JD, Kosowicz JG, and Stivers JT (2013) DNA Translocation by Human Uracil DNA Glycosylase: Role of DNA Phosphate Charge. *Biochemistry* 52 (15), 2526-2535

## 4.1 Introduction

Human uracil DNA glycosylase (hUNG) plays a central role in DNA repair and programmed mutagenesis of Ig genes, requiring it to act on sparsely or densely spaced uracil bases located in a variety of contexts, including U/A and U/G base pairs, and potentially uracils within single stranded DNA. An interesting question is whether the facilitated search mode of hUNG, which includes both DNA sliding and hopping, changes in these different contexts. Here we find that hUNG uses an enhanced local search mode when it acts on uracils in ssDNA, and also, in a context where uracils are densely clustered in duplex DNA. In the context of ssDNA hUNG performs an enhanced local search by sliding with a larger mean sliding length as compared to dsDNA. In the context of duplex DNA, insertion of high-affinity abasic product sites between two uracil lesions serves to significantly extend the apparent sliding length on dsDNA from 4 to 20 base pairs, and in some cases, leads to directionally biased 3'→5' sliding. The presence of intervening abasic product sites mimics the situation where hUNG acts iteratively on densely spaced uracils. The findings suggest that intervening product sites serve to increase the amount of time the enzyme remains associated with DNA as compared to nonspecific DNA, which in turn increases the likelihood of sliding as opposed to falling off the DNA. These findings illustrate how the search mechanism of hUNG is not predetermined, but instead, depends on the context in which the uracils are located.

Human uracil DNA glycosylase (hUNG) is an extremely versatile catalyst that excises uracils in a wide variety of genomic DNA contexts (1). For example,

during chemotherapy with antifolate and fluoropyrimidine drugs dUTP levels rise and replicative DNA polymerases frequently incorporate dUTP opposite to adenine (2, 3). Within this framework hUNG is likely confronted with densely spaced uracils in the context of U/A base pairs. The enzyme must also act on uracils that are generated by the enzyme activation induced cytosine deaminase (AID) during the process of adaptive immunity, which includes the two distinct processes of somatic hypermutation (SHM) (4) and class switch recombination (CSR) (5). Somatic hypermutation involves the iterative action of AID on multiple cytosines localized in the hypervariable regions of Ig genes. These cytosines are generated during the transient period where these regions are present as single stranded R loops during active transcription (4, 6). Thus, depending on the timing of hUNG with respect to transcription-coupled hypermutation, the enzyme might encounter either single-stranded uracils or uracils that are paired with guanine. Finally, to initiate CSR, AID must deaminate closely spaced cytosines on opposite strands of duplex DNA (generating U/G mismatches) such that recombinogenic double-stranded breaks are introduced after hUNG acts at such sites (5). Given the diverse contexts of these genomic uracils we wondered whether the facilitated search mechanism of hUNG might be altered from that observed with sparsely spaced uracils in duplex DNA (7, 8).

What specific aspects of a uracil's environment might influence the search mechanism of hUNG? In the case of DNA sliding to a uracil site, the most critical factors are the bound lifetime ( $\tau_{\text{bind}}$ ), which provides the upper limit time frame for sliding, and the 1D diffusion constant ( $D_1$ ), that sets the speed limit for sliding (9).

Together, these parameters define the sliding length,  $L = \sqrt{D_1 \times \tau_{bind}}$ . Thus, if  $D_1$  or  $\tau_{bind}$  are increased in a given DNA context, hUNG will search longer stretches of DNA by a sliding mode. In the case of DNA hopping between uracil sites (i.e. short range dissociation and reassociation events), such events will become more likely if the persistence length of DNA is decreased (as in ssDNA), because the probability ( $P$ ) for successful hopping between sites is inversely related to the distance ( $r$ ) between the sites ( $P \sim 1/r$ ) (10). In addition, hopping is a required pathway to locate clustered uracils that are located on opposite strands of DNA (7, 8). From these considerations, it is reasonable to envisage that features such as single strand DNA bubbles, U/G base mismatches, and clustered uracils or abasic sites could change the hopping or sliding efficiency.

Here we explore how the search mechanism of hUNG is affected by the context in which the uracil sites are found. These studies show that the enzyme has an enhanced ability to slide along linear ssDNA substrates as compared to helical duplex DNA. Additionally, the sliding length of hUNG2 can also be significantly extended in duplex DNA when high affinity product abasic sites are inserted between two uracil target sites. These findings provide a window into the flexibility of the search mechanism of hUNG, which can be tuned to optimally locate densely spaced uracils that occur during adaptive immunity as well as sparse uracils that arise during spontaneous cytosine deamination or infrequent incorporations of dUTP.



## 4.2 Results

### 4.2.1 Site Transfer Mechanism on single stranded DNA

Previous site transfer measurements were made on ssDNA substrates with closely spaced uracils at 5 and 10 ntds (8). In order to further understand hUNG transfer on ssDNA we made site transfer measurements at lengths out to 40 ntds (S5<sup>ss</sup>, S10<sup>ss</sup>, S20<sup>ss</sup> and S40<sup>ss</sup>). These ssDNA sequences were designed to minimize any propensity for intramolecular hydrogen bonding that might give rise to secondary structures and anomalous site transfer results. Representative data for S20<sup>ss</sup> in the absence and presence of the uracil trap shows a significant degree of intramolecular transfer (revealed by excess A and C fragments) that is diminished, but not entirely removed, by the presence of uracil (**Figure 4.1a**). Furthermore, the plateau region of trapping has been reached because identical data were obtained in the presence of 10 and 15 mM uracil. Extrapolation of  $P_{\text{trans}}^{\text{obs}}$  (eq 1) to zero time shows that  $P_{\text{trans}} = 0.44 \pm 0.03$  and  $P_{\text{slide}} = 0.14 \pm 0.03$  for S20<sup>ss</sup> (**Figure 4.1b**). Similar measurements were made on the 40 ntd ssDNA substrates and the results are summarized in **Figure 4.1c-e**.

$$[1] \quad P_{\text{trans}}^{\text{obs}} = \frac{[A] + [C] - [AB] - [BC]}{[A] + [C] + [AB] + [BC]}$$

A unique feature of the transfer data for ssDNA substrates as compared to previous results with dsDNA is the flat dependence of  $P_{\text{trans}}$  on uracil spacing (**Figure 4.1c**). In fact, extrapolation of the data in **Figure 4.1c** to zero site

spacing suggests that the maximal transfer efficiency is only around 50% for single stranded DNA. One possible explanation for this limiting value is that once a uracil site is encountered, hUNG then partitions evenly between falling off the DNA and moving forward along the reaction coordinate to excise the base (i.e.  $k_{\text{ex}}/k_{\text{off}} \sim 1$ ). This ratio will serve to limit excision events at the second site even if intramolecular transfer is very efficient (7, 8). Previous measurements of this partitioning ratio for cleavage of uracil sites in dsDNA by both the human and E. coli UNG enzymes established that uracil sites were processed efficiently when they were encountered ( $k_{\text{ex}}/k_{\text{off}} \sim 5/1$ ) (7, 8). Here a similar pulse-chase rapid kinetic approach was used to measure a much lower  $k_{\text{ex}}/k_{\text{off}} = 0.64$  for uracil within a ssDNA context (**Figure 4.2**). This ratio indicates that the efficiency ( $E$ , see eq 2) of excising a uracil site once it is encountered in ssDNA is only  $0.39 \pm 0.14$  (**Figure 4.2** and denoted by the half-filled circle in **Figure 4.1c-e**). Correcting the  $P_{\text{trans}}$  values in **Figure 4.1c** for this efficiency (i.e.  $P_{\text{trans}}/E$ ) yields the true site transfer probability for ssDNA, which is in the remarkably high range of  $\sim 0.6$  to  $1.0$  for spacings from 40 to 5 ntds.

[2] 
$$E = \frac{k_{\text{ex}}}{k_{\text{ex}} + k_{\text{off}}}$$

Do ssDNA site transfers in the presence of uracil correspond to chain sliding? It would be anticipated that the probability of successful intramolecular transfer by sliding would follow a site spacing dependence according to eq 3, where  $E$  is the site excision efficiency at zero site spacing and  $S$  is the kinetic

partitioning ratio ( $S = k_{sl}/(k_{sl} + k_{off})$ ) that describes the likelihood that the enzyme will slide along the DNA strand as opposed to dissociating after each sliding step ( $n$ ) during transfer (9, 12).

$$[3] \quad P_{slide} = E \times S^{n^2}$$

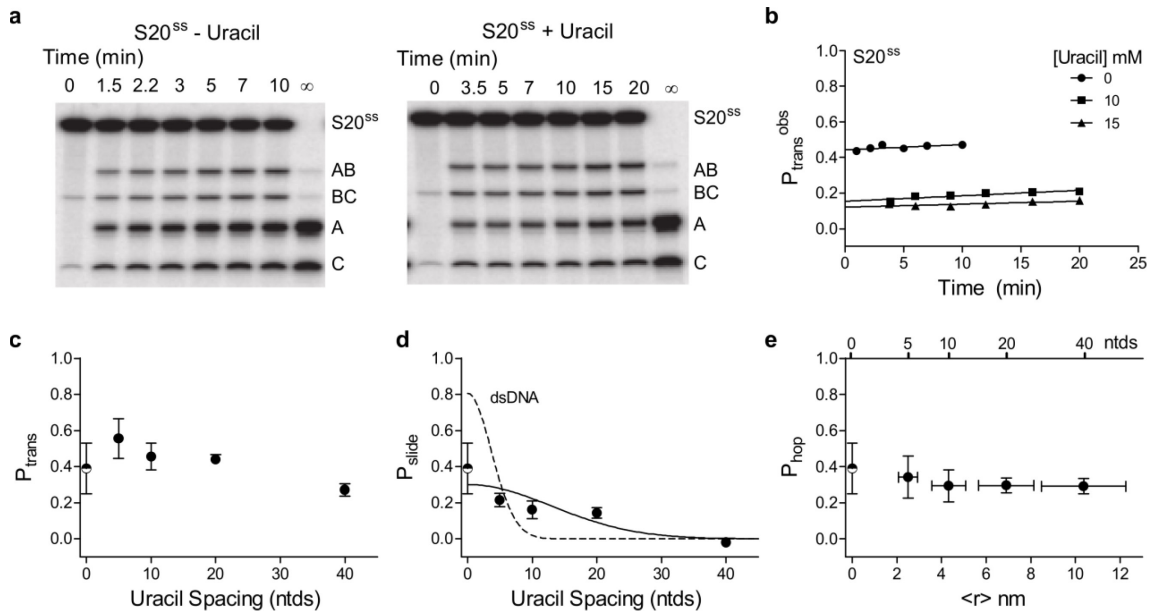
The square in eq 3 results from the fundamental stochastic property of diffusion where the total number of steps taken to traverse a given distance varies with the square of the separation in step length units (i.e. traversing a site separation of 10 ntd step lengths would require an average of 100 total steps) (9, 12). We used eq 3 to fit the observed transfer probabilities in the presence of uracil as a function of nucleotide spacing between the two sites, and compared the results obtained with ssDNA to that of duplex DNA (**Figure 1d**, dashed line). Within experimental error the site transfer probabilities on ssDNA decrease with increasing site spacing according to the expectations of a sliding mechanism. Using this model, the mean sliding length (defined as the uracil spacing where  $P_{slide}$  is diminished by 50%) is calculated as 19 ntds, compared to only 4-5 bp for duplex DNA (**Figure 1d**, dashed line) (8). A discussion of whether the term “sliding” is an appropriate descriptor for a 1-dimensional random walk on a flexible polymer like ssDNA is deferred to the Discussion.

The probability of successfully hopping between two sites ( $P_{hop} = P_{off}P_{return}$ , see **Figure 3.1**) separated by a linear distance  $r$  should follow the relationship of eq 4, where  $a$  is the diameter of an idealized spherical target (9,

12). For single stranded DNA, the average distances between target sites ( $\langle r \rangle$ ) must be obtained using the worm like chain model and experimental estimates for the persistence and contour lengths of ssDNA at the salt concentration used in these experiments (Section 4.4 Methods) (13, 14).

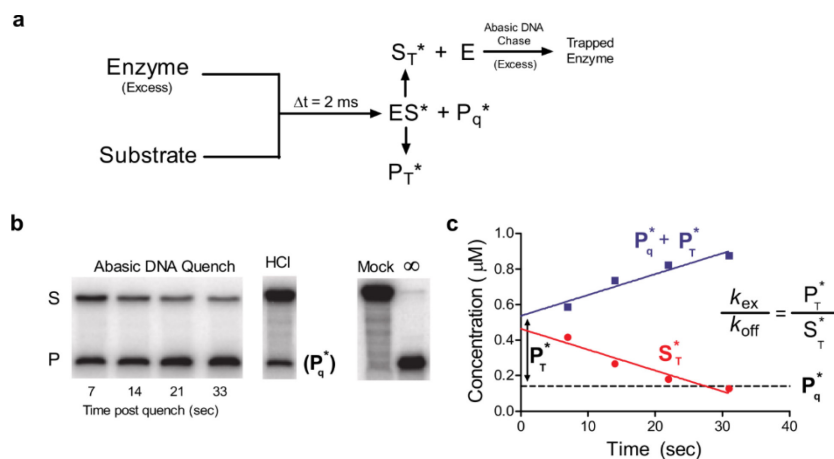
$$[4] \quad P_{\text{hop}} = E \times \frac{a}{r}$$

Although eq 4 describes very well the transfer probability of human and *E. coli* UNG for site spacings in duplex DNA between 20 bp and 800 bp (7, 8), this relationship fails to account for the flat distance dependence of  $P_{\text{hop}}$  in ssDNA (**Figure 4.1e**). This result is not unexpected because the persistence length of ssDNA is very short compared to dsDNA (1-3 nm versus 50 nm) (13, 14). Thus, the largest uracil spacing of 40 ntds only results in an average target site separation of about 10 nanometers (**Figure 4.1e**). Moreover, eq 4 breaks down when  $a \sim r$  because the enzyme engages a length of DNA that is similar to the site spacing (hUNG contacts at least 5 ntds of single stranded DNA) (15).



**Figure 4.1. Facilitated transfer of hUNG on ssDNA as determined by the ‘molecular clock’ approach (8).** (a) Gel images of the site transfer assay performed using S20<sup>SS</sup> in the presence and absence of the uracil trap. We note the presence of a very small percentage (<1%) of the 3’ labeled BC and C bands in the zero time lane. These bands result from a very small amount of uracil DNA glycosylase activity (<1% cleavage in >2 hrs reaction at 37 °C) in the commercially available 3’ terminal deoxynucleotidyl transferase enzyme used in the 3’ <sup>32</sup>P labeling of the DNA substrates. These background bands were determined to have a negligible effect on the calculations of site transfer probabilities and initial rates. (b) Determination of P<sub>trans</sub> for S20<sup>SS</sup> by linear extrapolation of the P<sub>trans</sub><sup>obs</sup> (eq 1) to zero time. (c) Total transfer (P<sub>trans</sub>) as determined by the site transfer assay without the addition of the trap (uracil). (d) Transfer by sliding (P<sub>slide</sub>) by determination of site transfer at high trap concentrations. The solid line is a least squares fit to a random walk sliding model ( $P_{slide} = E \times [k_{slide}/(k_{slide} + k_{off})]^N$  where  $N = ntds^2$  is the number of stochastic steps taken during sliding transfers and E is the efficiency of uracil excision (see **Figure 4.2 and eq 2**). The dotted line is the same fit previously obtained for dsDNA (8). (e) Spacing dependence of the hopping probability as determined by the difference between P<sub>trans</sub> and P<sub>slide</sub>. The mean distance between uracil sites was determined from the worm like chain model (33). The x-axis error bars for the mean square distance (<r>) are maximum and minimum values calculated using various experimental estimates of the persistence and contour lengths for ssDNA (13, 14) (see Methods). The shown data for ssDNA at 5 and 10 bp spacing were previously reported (8). P<sub>slide</sub> was calculated as the average plateau value for all data points performed at 10 or 15 mM uracil. For data

obtained at 20 and 40 bp uracil spacings,  $P_{\text{slide}}$  was determined from the average of 3 replicate trials at each uracil concentration. The half-filled circle at zero site spacing in all panels is the efficiency of uracil excision when hUNG has found a target site as determined in Figure 4.2 ( $E = 0.39 \pm 0.14$  for ssDNA). All transfer probability errors represent the mean plus or minus one standard deviation obtained from at least three replicate measurements at 0 mM uracil and six replicates at high uracil concentrations (three replicates each at 10 and 15 mM uracil).



**Figure 4.2. Determination of the excision efficiency ( $E$ ) for uracil cleavage by hUNG on ssDNA.** (a) Rapid mixing scheme where hUNG ( $4 \mu\text{M}$ ) was reacted with  $1.0$  or  $0.5 \mu\text{M}$  90mer DNA substrate containing a single uracil site ( $1\text{XU}^{\text{SS}}$  90mer, Methods. Both DNA concentrations gave identical results). After mixing, the reaction was either quenched with  $0.5 \text{ M HCl}$  or chased with  $60 \mu\text{M}$  duplex DNA containing a tetrahydrofuran abasic site mimic (chDNA). Following the chase, the reaction mix was then manually quenched with HCl at the indicated times in panels b and c. (b) Separation of product and substrate bands by denaturing gel electrophoresis in the quenched and chased samples. Mock is a control reaction where the DNA was subjected to the exact processing procedure without the addition of enzyme. Note that the amount of trapped product at the time of chase mixing ( $P_T^*$ ) must be obtained from  $P_{\text{tot}}$  by subtracting the amount of product already present within the  $2 \text{ ms}$  aging time as determined by the acid quenched samples. Thus, when  $2 \mu\text{M}$  UNG was mixed with  $1 \mu\text{M}$  substrate for  $2 \text{ ms}$  and the reaction was quenched with  $0.5 \text{ M HCl}$ ,  $0.13 \pm 0.01 \mu\text{M}$  excision product was formed ( $P_q^*$ ) and  $0.86 \pm 0.01 \mu\text{M}$  bound substrate was left unreacted ( $ES^*$ ). Linear extrapolation to zero time is used to determine the amount of total product ( $P_{\text{tot}} = P_q^* + P_T^*$ ) formed and substrate ( $S_T^*$ ) remaining at the time of addition of the chase DNA. (c) When the acid quench was replaced with  $60 \mu\text{M}$  F containing DNA duplex (chDNA) to serve as a trap for hUNG after it dissociated from the  $ES^*$  complex, the  $0.86 \mu\text{M}$   $ES^*$  present at  $2 \text{ ms}$  was converted to  $0.53 \pm 0.08 \mu\text{M}$  free substrate ( $S_T^*$ ), and  $0.47 \pm 0.08 \mu\text{M}$  was excised to form product ( $P_T^*$ ). Because  $P_T^*/S_T^* = k_{\text{ex}}/k_{\text{off}} = 0.34/0.53 = 0.64$ , then the average excision efficiency  $E = k_{\text{ex}}/(k_{\text{off}} + k_{\text{ex}}) = P_T^*/(P_T^* + S_T^*)$  may be directly calculated as  $0.39 \pm 0.14$  from nine independent trials. Values are reported as the average  $\pm 1 \text{ s.d.}$  Control experiments varying the enzyme/substrate ratio as well as the chase DNA concentration gave identical results and are shown in Figure 4.12.

#### 4.2.2 Intervening Abasic Sites Extend the Sliding Length of hUNG

We used a modified method to analyze the transfer data for the DNA constructs that contained abasic sites due to a large apparent site preference with some of these constructs (**Figure 4.3**). This method (initial rates method) differs in that the initial rates for formation of the individual fragments are calculated first ( $v_{AB}$ ,  $v_{BC}$ ,  $v_A$ ,  $v_C$ ) and  $P_{\text{trans}}$  is calculated using eq 5 (11), whereas previously we had calculated the site transfer using eq 1 at each time point and linearly extrapolated back to zero time (extrapolation method).

$$[5] \quad P_{\text{trans}} = \frac{v_A + v_C - v_{AB} - v_{BC}}{v_A + v_C + v_{AB} + v_{BC}}$$

Both methods are equivalent, however we find that using the initial rates in the site transfer equation was more reliable and intuitive when a site excision or directional biases to transfer were present ( $v_{AB} \neq v_{BC}$  or  $v_C \neq v_A$ ). As previously described (11), the initial rates for formation of the individual fragments describe four possible mechanistic scenarios as follows (**Figure 4.3**): (Case 1) when  $v_{AB} = v_{BC} = v_C = v_A$ , there is no site preference and only primary excision events occur, (Case 2) when  $v_A = v_C > v_{AB} = v_{BC}$ , there is no site preference, but directionally unbiased intramolecular transfers lead to consumption of the AB and BC intermediates, (Case 3) when  $v_{BC} = v_A$  and  $v_{AB} = v_C$ , a site preference exists, but only primary excision events occur, and (Case 4) when  $v_A \neq v_C > v_{AB} \neq v_{BC}$ , a site preference *or* directionally biased transfer is indicated, which cannot be distinguished unless other information is available.



Substrates containing one or more intervening F residues were designed with five, eleven and nineteen bp uracil site spacings (**Figure 4.4**). These uracil spacings are equal to or greater than the previously determined sliding length of ~4 bp for dsDNA and allow investigation of the effects of intervening abasic sites on both hopping and sliding efficiencies. The substrates S5F, S11F and S19F were designed to have identical two or three bp sequences surrounding the uracil target sites and the F residues were located no closer than two base pairs from the uracil sites to minimize possible direct effects on the catalytic complex. In addition, with the aim of increasing the probability of successful sliding between F sites, each intervening site was separated by three bp, which is less than the sliding length on duplex DNA. Thus the substrate with the five bp uracil spacing contained one F site, the substrate with an 11 bp spacing contained two intervening F sites, and the 19 bp site spacing contained four F sites (S5F, S11F and S19F, **Figure 4.4**). The effect of multiple abasic site substitutions on the structure of the local intervening DNA is not obtainable, but it is reasonable to expect that a dynamic equilibrium between structures that resemble locally unpaired and paired strands might exist. Although it is certainly desirable to understand the structural effects of these pseudo-abasic site constructs, measurement of the transfer effects does not require knowledge of structure. Of course, interpretation of the observed effects must be made with this uncertainty in mind.

We collected site transfer data for S5F, S11F, and S19F in the presence and absence of uracil and the individual analyses of the velocity data are

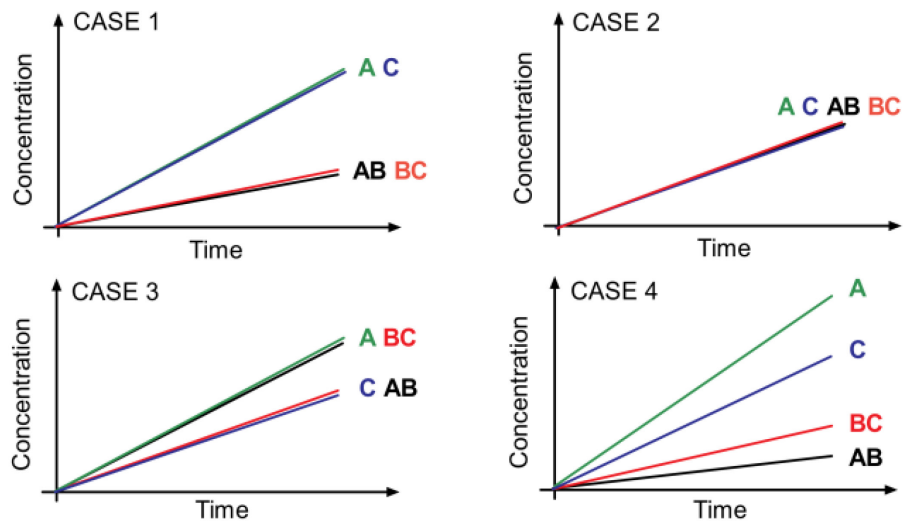
shown in **Figure 4.5**. An interesting aspect of the transfer reactions using the F substrates was that the apparent initial velocities for primary excision events at site 1 ( $v_{BC}$ ) and 2 ( $v_{AB}$ ) became increasingly divergent as the site spacing increased, indicating a site preference. Indeed, a plot of  $v_{BC}/v_{AB}$  against site spacing shows that the ratio is essentially unity at the five bp uracil spacing and increases to almost 10 at the 19 bp spacing (**Figure 4.5a**). For substrate SF5 (**Figure 5b and e**), the transfer measurements in the absence and presence of uracil correspond to Case 1 in **Figure 4.3** (i.e. no site preference with facilitated transfers by hopping and sliding). For substrates S11F (**Figure 4.5c and f**) and S19F (**Figure 4.5d and g**), the measurements correspond to Case 4 (a small or large site preference, with facilitated transfers). It is notable that no previously investigated duplex or single stranded substrates for hUNG have displayed a site preference, and that the large preference for site 1 appears only as the site spacing is increased.

Numerical simulations were used to explore possible interpretations for the large preference for excision at site 1 in S19F (Section 4.4 Methods). These simulations confirmed that the data cannot distinguish between three scenarios (i) a greater rate of cleavage at site 1 as compared to site 2, resulting in low levels of fragment AB compared to BC, (ii) a preference for transfer in the site 2→1 direction (which would consume AB efficiently to make A), or (iii) a combination of both mechanisms.

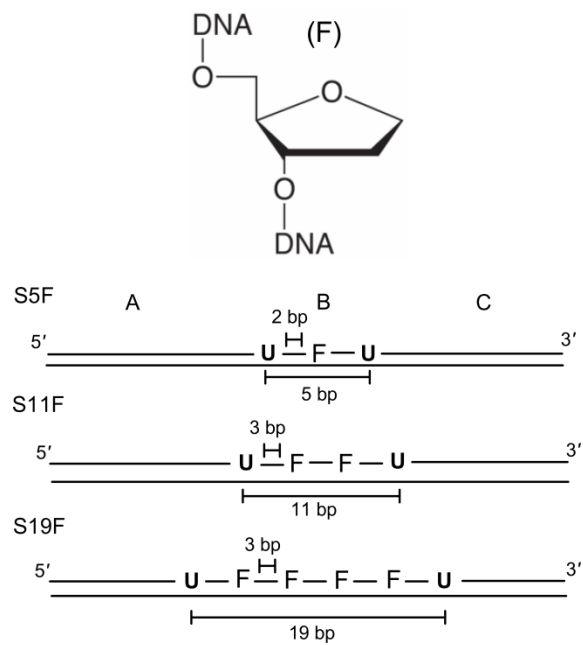
To distinguish between these possibilities, substrates were designed that were identical to S19F but contain only a single uracil site (**Figure 4.6a**). Single

turnover measurements confirmed that the rates of uracil cleavage ( $k_{\text{ex}}$ ) at each uracil site (5' or 3') to the intervening F residues were identical (**Figure 4.6**). Another possibility for these results would be a difference in the excision efficiency between the two sites as described above for single stranded DNA. Using an identical trapping approach as described above for the ssDNA substrate, we determined that the efficiency (E) of cleaving a uracil once the site is located is identical for each site ( $0.92 \pm 0.12$  for the 5' site and  $0.86 \pm 0.04$ ) (**Figure 4.7**). These results establish that there is no off-rate difference once hUNG has landed on either uracil site. Therefore, the only reasonable explanation for the observed site preference is preferential transfer in the site 2 → site 1 direction.

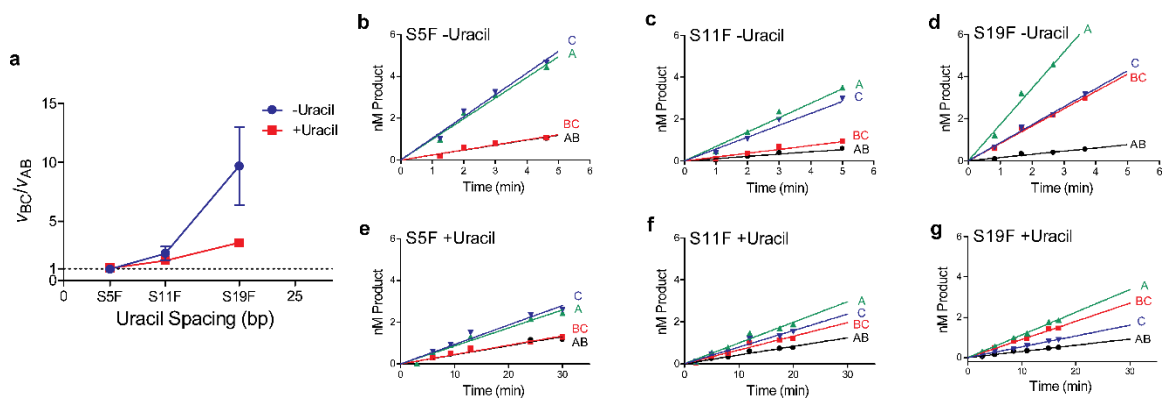
A summary of the overall site transfer properties for these substrates is presented in **Figure 4.8**. Notably the uracil insensitive or sliding pathway ( $P_{\text{slide}}$ ) persists at eleven and nineteen base pairs which is considerably longer than that of duplex DNA where sliding persists over only 4-5 bp (8). Although, the molecular origin of the increased site transfer in the 5→3' direction is not fully discernible, it is clear that the site transfer properties, including the apparent sliding length of hUNG, are very much context dependent.



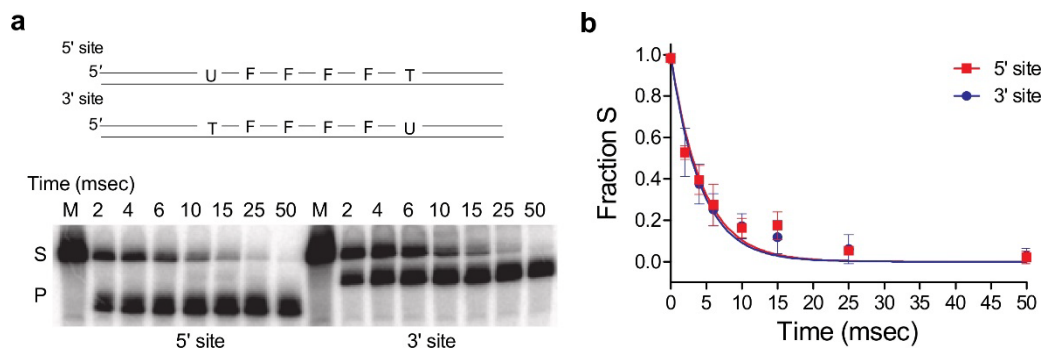
**Figure 4.3. Possible outcomes for site transfer measurements** as determined using the method of initial rates [adapted from (11)]. The cartoon panels depict the initial time courses for formation of the cleavage products AB, BC, A and C. Case I: hUNG is processive with an equal preference for both sites resulting in a larger initial velocity for the products A and C compared to AB and BC. Case II: hUNG is fully distributive in its reaction and the enzyme dissociates from the DNA after each single excision event, after which it becomes in equilibrium with all DNA substrate molecules. Case III: hUNG is fully distributive, but reacts preferentially at site 1 as compared to site 2, generating more of fragments A and BC. Case IV: hUNG is processive but also has a site preference. As shown in the Results and Materials and Methods, an apparent site preference may result from either an excision preference or biased transfer in one direction.



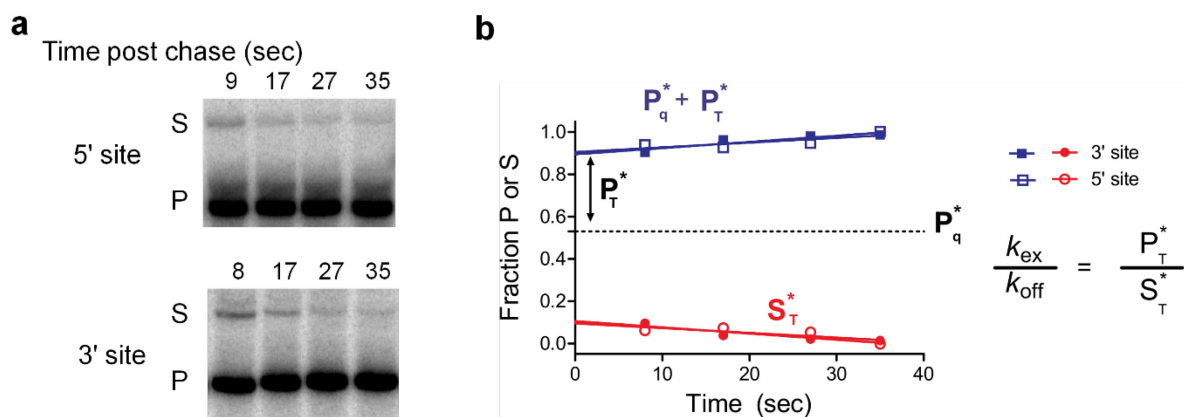
**Figure 4.4. Structure of the tetrahydrofuran abasic site mimic (F) and design of the uracil substrates containing intervening F residues.**



**Figure 4.5. Site transfer measurements as determined using the method of initial rates for substrates S5F, S11F, and S19F.** (a) The site preference ( $v_{BC}/v_{AB}$ ) as a function of uracil site spacing in the presence and absence of uracil. Panels (b) and (e); velocities for formation of individual fragments derived from substrate S5F in the presence (b) and absence (e) of uracil. Panels (c) and (f); velocities for formation of individual fragments derived from substrate S11F in the presence (c) and absence (f) of uracil. Panels (d) and (g); velocities for formation of individual fragments derived from substrate S19F in the presence (d) and absence (f) of uracil. Reported errors are 1 SD as determined from at least 3 trials at 0 mM uracil trap and 6 trials at high uracil trap concentrations (3 trials each at 10 and 15 mM).

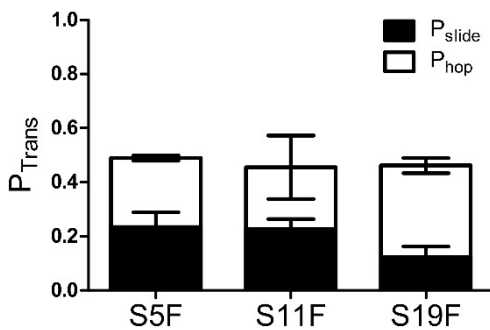


**Figure 4.6. Single-turnover kinetic measurements of uracil cleavage from each site in the F containing DNA substrate S19F.** (a) Substrate design and representative gel showing separation of substrate and product as a function of aging time in the rapid mixer using a final concentration of 4  $\mu\text{M}$  hUNG and 280 nM substrate DNA (5' or 3' site). (b) Fraction of substrate as function of aging time. The least squares fit is to a single exponential that provides the cleavage rate of uracil from DNA. The determined rates were independent of enzyme concentration and identical within error,  $k_{\text{ex}}$  (5' site) =  $239 \pm 19 \text{ s}^{-1}$  and  $k_{\text{ex}}$  (3' site) =  $226 \pm 17 \text{ s}^{-1}$  where the error is the standard error of the least squares fit as determined using Graphpad Prism. The data points shown are the average  $\pm$  SD ( $n = 3$ ) of data points determined using 4 and 8  $\mu\text{M}$  hUNG and 280 nM DNA substrate.



**Figure 4.7: Determination of the efficiency of uracil excision (E) for single uracil containing substrates analogous to S19F.** The experiment is identical to the scheme depicted in Figure 4.2a where hUNG was rapidly mixed with DNA substrate and chased at 2 milliseconds with 20  $\mu$ M F site containing DNA (chDNA). See Figure 4.2 legend and main text for detailed description. The data depicted here were determined using a final concentration of hUNG at 4  $\mu$ M, and the substrate concentrations were 280 nM (shown above) or 140 nM. Identical results were achieved with each Enzyme/DNA ratio. The HCl quench to determine  $P_q^*$  is shown in Figure 4.6 as part of the single turnover measurements. The excision efficiency,  $E = k_{ex}/(k_{off} + k_{ex}) = P_T^*/(P_T^* + S_T^*)$  for the 5' site was  $E = 0.92 \pm 0.12$  and for the 3' site  $E = 0.86 \pm 0.04$ . The reported values are the average  $\pm$  1 SD.





**Figure 4.8. Facilitated site transfer properties of hUNG on uracil DNA constructs containing intervening F sites.** The total transfer ( $P_{trans}$ ) is the sum of the trappable pathway or uracil sensitive ( $P_{hop}$ ) and the untrappable or uracil insensitive pathway ( $P_{slide}$ ). Errors represent the mean plus or minus 1 s.d. determined from at least 3 trials at 0 mM uracil and 6 at high uracil (3 replicates each at 10 and 15 mM uracil).

### 4.3 Discussion

hUNG is unique among DNA glycosylases in that it has the catalytic flexibility to remove uracils from both duplex and single stranded DNA with almost equivalent efficiency, and thus provides a valuable system to understand intramolecular site transfer within a variety of DNA contexts. Although the action of hUNG on uracils in ssDNA has not been directly established *in vivo*, the process of somatic hypermutation of Ig genes in B cells involves enzymatic deamination of cytosines by AID in single stranded DNA that forms transiently during active transcription of these genes (16). Indeed AID has been shown to be highly processive in deaminating neighboring cytosines leaving behind clusters of closely but not uniformly spaced uracils, similar to the spacing in our assays here (17-19). Thus, it seems hUNG likely acts on uracils positioned within a variety of contexts including ssDNA and uracils positioned among neighboring abasic sites.

#### 4.3.1 Does hUNG “Slide” on ssDNA?

The present data with ssDNA substrates unambiguously show that hUNG can efficiently transfer between uracil sites in ssDNA and that transfers persist even in the presence of the uracil trap, which is at least phenomenologically consistent with “sliding” (see Results and **Figure 4.1d**). Despite this apparent sliding behavior on the ssDNA platform, it should be noted that rigorous interpretation of the observed transfer behavior with ssDNA is inherently more complicated than dsDNA because of several intrinsic properties of ssDNA.

Boundary estimates for one-dimensional sliding of hUNG on duplex DNA have been previously determined (8)(Chapter 3) and it is of interest to perform a similar analysis for ssDNA to gain insights into the possible nature of sliding on ssDNA as compared to dsDNA. Calculation of a lower limit 1D diffusion constant for sliding transfers requires knowledge of the mean sliding length on ssDNA ( $L_{\text{slide}}$ ) and the binding lifetime to nontarget ssDNA ( $\tau_{\text{bind}} = 1/k_{\text{off}}$ ). The required value for  $L_{\text{slide}} = 19$  ntds may be obtained from **Figure 1d**, and a value for  $\tau_{\text{bind}} = 1/K_{\text{D}}^{\text{ns}} \times k_{\text{on}} = 4$  ms may be estimated from (i) measurements of the nonspecific ssDNA binding affinity of hUNG as measured by fluorescence anisotropy ( $K_{\text{D}}^{\text{ns}} = 2.0 \pm 0.3 \mu\text{M}$ , **Figure 4.9**) and, (ii) the diffusion-controlled on-rate for reaction of ssDNA substrate DNA ( $k_{\text{on}} = 1.1 \times 10^8 \text{ M}^{-1} \text{ s}^{-1}$ , **Figure 4.10**). Insertion of these parameters into eq 6 (9), gives a value for  $D_1 = 8 \times 10^4 \text{ ntd}^2 \text{ s}^{-1}$ .

$$[6] \quad D_1 = \frac{L_{\text{slide}}^2}{\tau_{\text{bind}}}$$

This value may be converted to standard distance units using a contour length for ssDNA of 0.6 nm under the low salt conditions employed here (13, 14), which gives  $D_1$  (ssDNA) =  $3 \times 10^{-2} \mu\text{m}^2 \text{ s}^{-1}$  (Methods). This value may be compared with the boundary limits for duplex DNA which were calculated using two limiting cases (i) that hUNG during its entire bound lifetime is in a conformation that is competent for sliding ( $D_1 = 0.07 \times 10^{-2} \mu\text{m}^2 \text{ s}^{-1}$ ), and (ii) that hUNG is in a conformational state that is competent for sliding only 5% of its total bound

lifetime ( $D_1 = 1.4 \times 10^{-2} \mu\text{m}^2 \text{s}^{-1}$ ) (8). As discussed in Chapter 3, these boundary conditions were estimated based on NMR studies of the conformational dynamics of hUNG bound to nonspecific DNA (20)(Chapter 3).

The above calculations indicate that the apparent  $D_1$  for ssDNA (calculated using a sliding time equal to the total bound lifetime) is 40 times larger than the corresponding value for duplex DNA. Thus, substantial differences in the interactions or mechanism of sliding between ssDNA and duplex DNA are clearly apparent. The smaller diffusion constant for sliding on duplex DNA is not likely to arise from its higher charge density as compared to ssDNA because methylphosphonate substitution does not reveal any evidence for a strong electrostatic component to sliding (Chapter 3). Since there are no structures of long ssDNA molecules bound to hUNG (21-24), it is quite possible that the flexible polymer nature of ssDNA may allow interactions over a more extended binding surface of hUNG than the more rigid duplex DNA polymer. Such an extended surface for ssDNA, and even “scrunching” of the polymer, could lead to longer apparent sliding lengths and correspondingly larger calculated diffusion constants (see above). Another possible explanation for the smaller diffusion constant for duplex DNA is that sliding on duplex DNA involves the increased frictional resistance arising from rotation-coupled diffusion along the helical DNA chain while sliding on single stranded DNA does not (25, 26). Additionally, we note that facilitated diffusion on ssDNA has been previously observed in bulk solution and single-molecule measurements of the AID family member cytidine deaminase APOBEC3G (27, 28).

### 4.3.2 Mechanism of Directional Bias During Reaction at Clustered Uracils

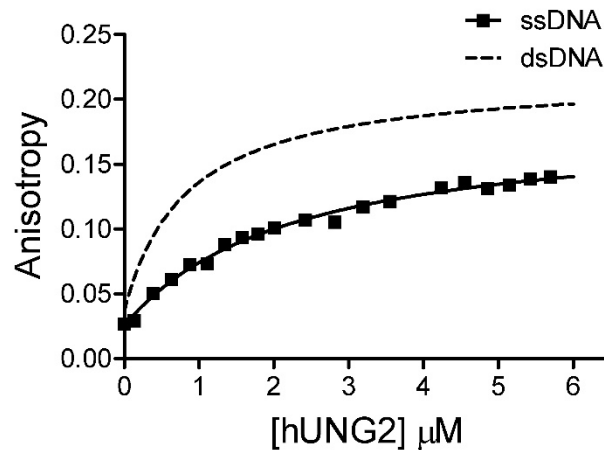
Although some enzymes such as helicases and DNA polymerases can use the free energy of nucleotide hydrolysis to move directionally along DNA, the movement of DNA glycosylases is driven only by thermal energy and thus would not be expected to have directional bias. Indeed, we have always found that there is no 5' or 3' directional bias for transfer of hUNG between uracil sites in duplex or single stranded DNA (7, 8). However, the expectation of no directional bias for a thermally driven transfer process might be negated if the intervening DNA chain connecting the sites contained high affinity regions that served as thermodynamic sinks to pull the enzyme in a biased direction.

Here we have shown that insertion of high affinity and flexible abasic site mimics between two uracil target sites can increase the average apparent sliding length of hUNG by 5-fold. This finding suggests that when hUNG acts on clustered uracils (resulting in clustered abasic sites) its search strategy is modified to increase the local search efficiency by sliding. An unexpected result in these studies was the observed 3'→5' directional bias observed for S19F (**Figure 4.5d**). While the exact mechanism of this directionally biased transfer is not easily discernible, the data would seem to require asymmetric interactions of hUNG with the DNA, because the only difference between the two uracil target sites is whether the F sites lie 3' or 5' distal to the uracil. Recent H/D exchange mass spectrometry experiments have provided evidence for an asymmetry in the interaction of hUNG with a 30 mer duplex DNA containing a F site (29). This

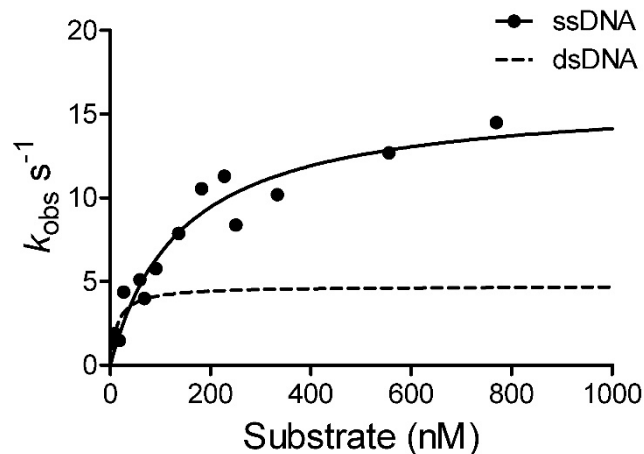
data suggested a previously undetected DNA binding surface of the enzyme that could interact with the DNA 3' to the product site. Thus one reason that directionality may appear in these abasic constructs and not normal DNA is that binding to the newly detected DNA binding region is favored by the introduction of duplex destabilizing lesions which allowing more facile DNA bending. This is also consistent with previous studies showing that hUNG favors binding to destabilized base pairs (30, 31). It is not known whether the 3'→5' bias is merely related to an inherent asymmetry in the hUNG DNA binding mode, or if there exists a larger functional significance for this behavior in a biological context.

#### **4.3.3 Conclusion**

This report establishes that hUNG can change its search mode in the direction of longer DNA sliding events when it is confronted with ssDNA and clustered lesions such as abasic sites. This property is likely to be relevant during adaptive immunity, as selective patterning of uracil cleavage events has been shown to be important in the controlled mutagenesis of immunoglobulin hypervariable sequences (32). Additionally, it is envisaged that facilitated sliding will be important in other regions of the genome where destabilized or ssDNA persists such as replication foci or regions of high negative supercoiling.



**Figure 4.9: Non-specific binding of hUNG2 to ssDNA.** hUNG2 was titrated into a cuvette containing 50 nM of a 10mer FAM labeled oligo and the anisotropy was recorded. Using a one-site binding isotherm, the  $K_D$  was determined as  $2.0 \pm 0.3 \mu\text{M}$  with a maximal anisotropy of  $0.16 \pm 0.01$ . The least-squares fit for non-specific hUNG2 binding for the identical DNA in the duplex form is depicted by the dotted line as determined in (7).



**Figure 4.10: Steady-state reaction kinetics for a ssDNA substrate containing a single uracil (1XU\_90mer).** The  $K_m$  and  $V_{max}/[E_{tot}]$  were determined by fitting to the Michaelis-Menten equation:  $K_m = 141 \pm 34$  nM and  $V_{max}/[E_{tot}] = 16 \pm 2$   $s^{-1}$ . The errors are the standard error from the least squares fit to the data in Graphpad Prism 5. The fit for dsDNA under identical conditions is shown by the dotted line (7).



## 4.4 Methods

### 4.4.1 Protein and Oligonucleotide Sequences

Oligonucleotide substrates are named as follows, with the number indicating the spacing of the uracil sites and the superscript “ss” indicating single-stranded DNA; all other substrates were used in the duplex form. The letters F indicate a tetrahydrofuran abasic site mimic. Importantly, the single stranded substrates were designed to have no more than two possible adjacent Watson-Crick pairings as determined from the hybridization prediction program mFold (1).

*S20<sup>ss</sup>\_90mer:*

CAC AAT AAC ACA TAC A CTA A TCAT ACA TCA CAC AAA ACA U ACA AAA  
CAC AA CAC AAA ACAUACA AAA CAC ACT AATC C ACAC AC ATA ACA

*S40<sup>ss</sup>\_90mer:*

CAC AAT AAC ACA TAC A CTA A TCAT ACA TCA CAC AAA ACA U ACA AAA  
CAC AA CAC AAA ACA ACA AAA CAC AA CAC AAA ACA U ACA AAA CAC  
ACT AATC C ACAC AC ATA ACA

*S5F\_90mer:*

GGT ATC CGC TGA AGT AGT CAC AAT TCC ACA CAA TGC TGA GGA ATC  
GA U AG F GA U AGC TAA GCT GAG GCA TAC AGG ATC AAT TGT CGA  
GCC

*S11F\_90mer:*

GGT ATC CGCT AGT CAC AAT TC ACA CAATGC TGA GG A AT CGA U AG C  
F ATA F CGA U AGC TAA GCT GAG GCATAC AGG ATC AAT TGT CGA GCC

(F = Tetrahydrofuran abasic mimic)

*S19F\_90mer:*

GGT ATC CGCT AGT CAC AAGT TC AATGC TGA GG A AT CGA U AG C F  
ATA F TGT F ATA F CGA U AGC TAA GCT GAG GCATC AGG AT TGT CGA  
GCC

(F = Tetrahydrofuran abasic mimic)

*1XU<sup>ss</sup>\_90mer:*

CAC AAT AAC ACA TAC ACT AAT CAT ACA TCA CAC AAA ACA U ACA AAA  
CAC AAC ACA AAA CAT ACA AAA CAC ACT AAT CCA CAC ACA TAA CA

Chase duplex (chDNA)

5' - GCG GCC AAA F AA AAA GCG C

3' - CGC CGG TTT A TT TTT CGC G

(F = Tetrahydrofuran abasic mimic)

Non-specific ssDNA (nsDNA<sup>ss</sup>)

CGC GTG TGC C – FAM

hUNG was purified as previously described (8). Protein concentrations were determined by absorbance measurements at 280 nm using an extinction coefficient of 33.68 mM<sup>-1</sup> cm<sup>-1</sup>. All 90 mer uracil containing oligonucleotides were ordered from Integrated DNA Technologies ([www.IDTDNA.com](http://www.IDTDNA.com)) in the crude desalted form and purified by denaturing PAGE.

#### **4.4.2 Experimental conditions**

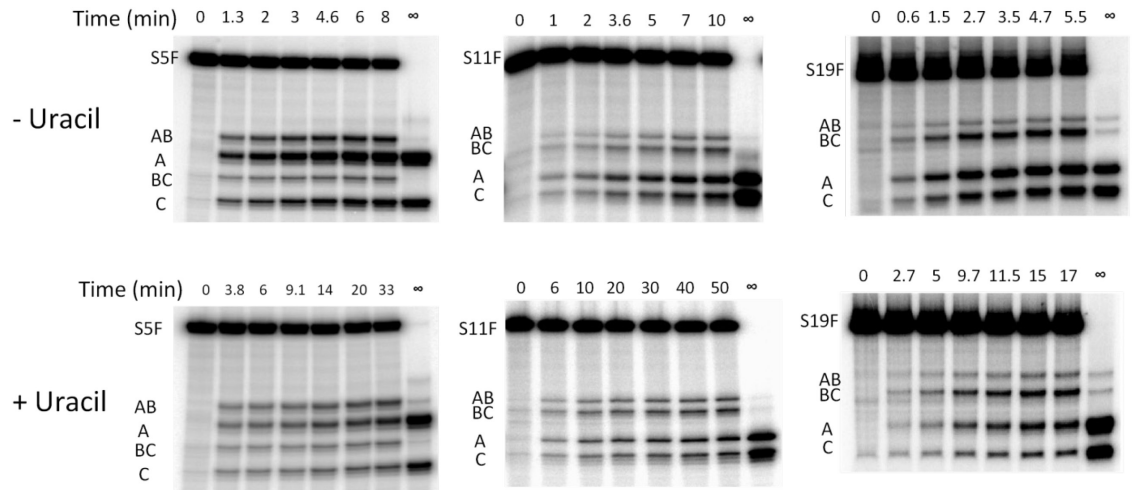
All measurements were made at 37 °C in a standard reaction buffer consisting of 20 mM HEPES pH 7.5, 0.002% Brij 35 detergent (Sigma Aldrich), 3 mM EDTA (added from a 0.5 M pH 8.0 stock), and 1 mM DTT unless otherwise noted.

#### **4.4.3 Intramolecular Site Transfer Assay**

For all substrates used here, oligonucleotides were labeled at the 5' and 3' ends by incubation with [ $\gamma$ - $^{32}\text{P}$ ] ATP and 3'-deoxyadenosine 5'-triphosphate (cordycepin 5'-triphosphate),-[ $\alpha$ - $^{32}\text{P}$ ] and polynucleotide kinase and terminal transferase (New England Biolabs) respectively and excess radioactivity and salt was removed by gel filtration as described in Chapter 3. Duplex substrates containing tetrahydrofuran (F) abasic site mimics (S5F, S11F, and S19F) were hybridized to the complement oligonucleotide by heating to 95 °C for 10 min in a dry heat block and allowed to slowly cool to room temperature. Gel images for each substrate with and without uracil are presented in **Figure 4.11**.

For post-reaction processing of the single and double stranded uracil substrates, the abasic sites generated by hUNG were cleaved with either hot piperidine (20 minutes at 90 °C), or by the addition of 0.25 M ethylenediamine (pH 8.0) (Chapter 3). Samples were then loaded onto a 10% denaturing gel (19:1 bis:acrylamide) that was preheated in order to denature any residual structure.

For ssDNA substrates where no site preference was observed, the data was analyzed as previously using eq 1 (Chapter 3). However in the case of substrates containing F sites, the data was analyzed using the method of Stanford *et al.* (see above Results section) (11).



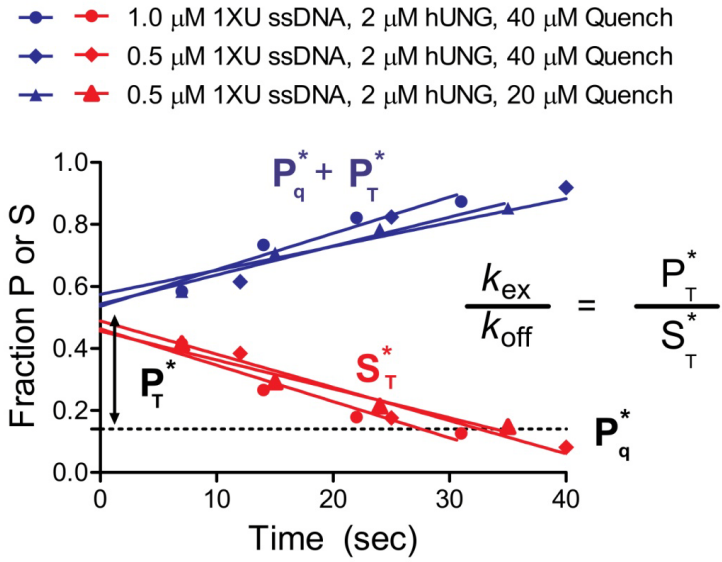
**Figure 4.11: Gel images for substrates S5F, S11F, and S19F (from left to right) in the absence (top row) and presence (bottom row) of uracil.**

#### 4.4.4 Determination of the Efficiency of Uracil Excision

The efficiency ( $E$ ) with which hUNG excises a uracil when it encounters a site as opposed to falling off the site, is defined by eq 2. The efficiency for uracil in a ssDNA and in the context of the F containing substrate S19F was determined as previously described for duplex DNA using a pulse-chase kinetic partitioning experiment (7, 8). Briefly for a substrate uracil within ssDNA, using a three syringe rapid mixing device (Kintek RQF3), 20  $\mu$ L of a 4  $\mu$ M solution of hUNG was rapidly mixed with 20  $\mu$ L of 5'  $^{32}$ P labeled 1XU<sup>ss</sup> substrate at a concentration of 0.5 or 1  $\mu$ M. The reaction was quenched at a 2 ms aging time by the addition of either 0.5 M HCl, or chased with a duplex DNA (60  $\mu$ M or 30  $\mu$ M) containing a high affinity F site (chDNA). The concentrations of the quench listed are that in the quench syringe of the rapid mixer resulting in approximately a 2/3 dilution in the final quenched solution. Identical results were observed in experiments varying the DNA/Enzyme ratio and when varying chase DNA concentration (**Figure 4.12, Figure 4.2**). For the samples machine-quenched by the chDNA, subsequent time points were taken between 5 and 30 seconds and manually quenched with an equal volume of 0.5 M HCl. To all samples an equal volume of phenol/chloroform/isoamyl alcohol (25:24:1, Invitrogen) was added and the samples were vortexed. The layers were allowed to separate by gravity and an equal volume of 2 M piperidine followed by centrifugation for 10 minutes at 13,000 x g. The aqueous layer was then transferred to another tube and heated to 90 °C for 20 minutes to cleave the abasic sites and then dried to completion to remove the piperidine. The samples were resuspended in 50%

formamide gel loading buffer and substrate and product were separated by electrophoresis on a 10% denaturing gel. The gels were dried and imaged as describe above. Detailed explanation and kinetic simulations validating the approach are described in Porecha *et al.* and the corresponding Supplemental Methods (7).

For determining the efficiency of cleavage and the single turnover rate of single uracil containing substrates analogous to S19F (S19F 5' site and S19F 3' site) an identical procedure was employed, however after reaction with hUNG, quenching and phenol-chloroform extraction, the DNA containing solution was instead neutralized with an appropriate volume of 3M Tris base. Formamide was then added to 65% final concentration and the sample was subsequently heated to 90 °C for 3 hours to cleave the abasic sites and immediately run on a 10% denaturing polyacrylamide gel as described above.



**Figure 4.12: Control experiments for determination of the excision efficiency (E) for ssDNA** showing the independence of the DNA/hUNG ratio and also independence of the concentration of the quencher (chDNA) indicating zero order trapping of the enzyme after dissociation from the substrate DNA. See Figure 4.2 for a more detailed description.

#### 4.4.5 Michaelis-Menten kinetics of hUNG2 reaction on ssDNA

1XU<sup>ss</sup> ssDNA was labeled at the 5' end with <sup>32</sup>P using  $\gamma$ -<sup>32</sup>P ATP and polynucleotide kinase. Concentrations of 1XU<sup>ss</sup> from 0-800 nM were reacted with hUNG (4-15 pM) and at time points in the initial rate regime 4  $\mu$ L aliquots were quenched with 5  $\mu$ L 0.5 M NaOH. The quenched samples were then heated to 95 °C for 10 minutes to cleave abasic sites. Formamide loading buffer was added to 50% final concentration and substrate and product fragments resolved by denaturing gel electrophoresis. Band intensities were quantified and initial rates determined by linear regression. The resulting rates were normalized to enzyme concentration and plotted against substrate concentration. The resulting data was fitted to the Michaelis-Menten equation to determine  $V_{max}$  and the  $K_m$  (Figure 4.10).



#### 4.4.6 Calculation of the mean square distance between target sites using the worm like chain model for ssDNA

For analysis of hUNG hopping,  $P_{\text{trans}}$  values were plotted against the mean square distance  $\langle r \rangle$  of the uracil target sites as determined from the worm like chain model (2)(eq 7).

$$[7] \quad \langle r \rangle = \left\{ 2PL \left[ 1 - \left( \frac{P}{L} \right) \left( 1 - e^{-\frac{L}{P}} \right) \right] \right\}^{-1/2}$$

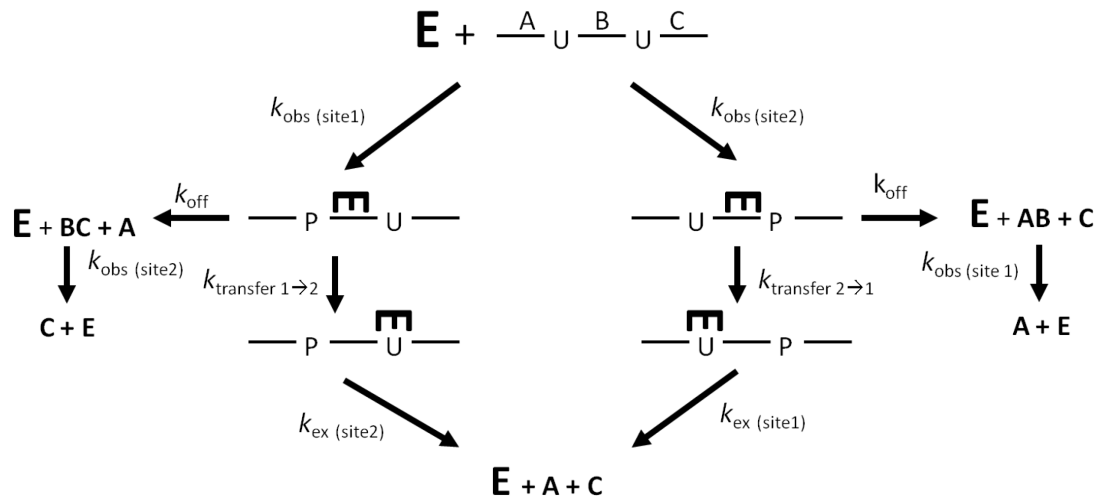
Where the  $P$  is the persistence length and  $L$  is the contour length of the chain. To estimate  $\langle r \rangle$  for ssDNA we employed experimental parameters obtained from single molecule FRET experiments on ssDNA (3, 4). At the salt concentrations employed here experimental estimates of  $P$  were between 2 and 3 nm, while the contour length for ssDNA is estimated as 0.5 – 0.7 nm. Error bars for the  $x$ -axis in **Figure 4.1e** represent the maximum and minimum values for  $\langle r \rangle$  using the above ranges for  $P$  and  $L$ .

#### 4.4.7 Kinetic Modeling of the data for S19F

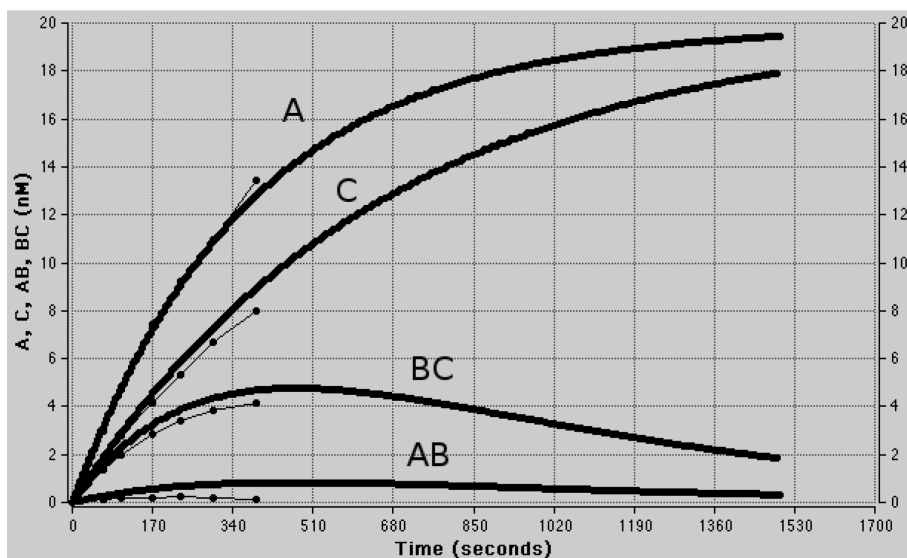
Mechanism 1 (below) was used to simulate the reaction timecourse for S19F. The concentrations of the DNA and enzyme were set to match those of the site transfer assay ( $[\text{hUNG}] = 50 \text{ pM}$ ,  $[\text{DNA}] = 40 \text{ nM}$ ). At these concentrations the rate-limiting step is release of the enzyme from the first site after uracil excision (5). These constants are all encapsulated within the rate

constant  $k_{\text{obs}}$ . Once cleavage and release from the first product site has occurred the enzyme can transfer and find the second site, or it can fall off of the DNA. The exact magnitudes of the rate constants following the initial rate-limiting step are not critical in this analysis, because it is the partitioning of the enzyme-F DNA complex between dissociation and transfer to the second site that determines the outcome of the simulation (i.e.  $P_{\text{transfer}} = k_{\text{transfer}}/(k_{\text{transfer}} + k_{\text{off}})$ ). For the simulation, we chose realistic values for the reaction steps following the initial cleavage at the first site: (i) the off-rate from the intervening DNA containing F sites ( $k_{\text{off}}$ ) was set at  $20 \text{ s}^{-1}$  obtained from stopped flow measurements of enzyme dissociation from F site containing DNA (5), (ii) the rate of cleavage at the second step is set to the single turnover rate for uracil excision ( $k_{\text{ex}} = 100 \text{ s}^{-1}$  at  $25 \text{ }^{\circ}\text{C}$  (6) and estimated to be  $200 \text{ s}^{-1}$  at  $37 \text{ }^{\circ}\text{C}$ ).

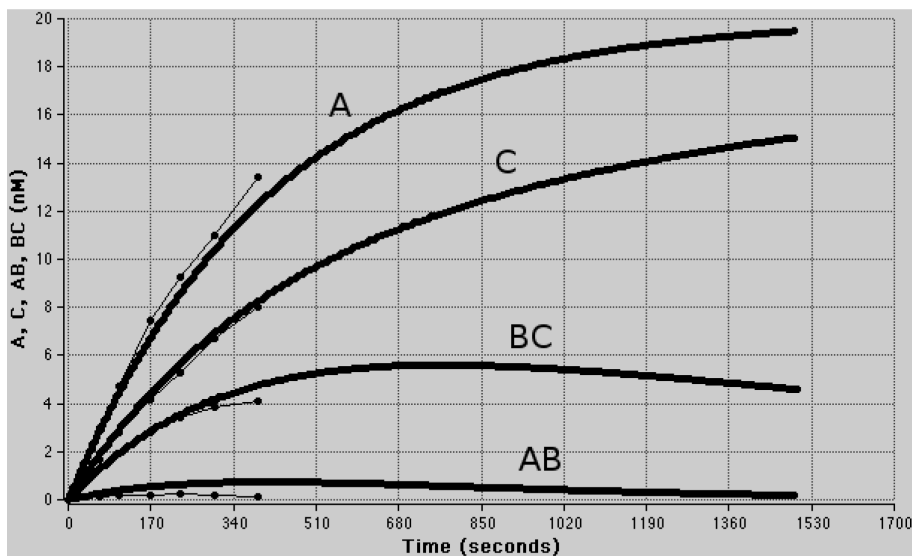
The results of the simulation I and II below show that the site preference observed for S19F can be attributed to either differences in cleavage rates at each site, differences in transfer rates, or a combination of both effects. In the simulations below, the raw data for S19F is plotted along with the simulated progress curves.



**Mechanism 4.1: Site transfer mechanism used in numerical simulations.**  $k_{\text{obs}}$  at both sites is the initial binding and cleavage steps of the hUNG with a uracil site.  $k_{\text{off}}$  is the off-rate for the intervening DNA between uracil sites and  $k_{\text{transfer}}$  is the rate at which the enzyme transfers between uracil sites.  $k_{\text{ex}}$  is the rate of uracil excision once the enzyme has found the uracil site.



**Simulation I: Differential transfer rates.** We first tested a model involving differential rates of transfer between each site. In this simulation, all cleavage rates at sites 1 and 2 were set to be equal ratio ( $k_{\text{obs}}$  site 1 and 2 =  $0.03 \text{ nM}^{-1} \text{ s}^{-1}$ ), while the ratio of  $k_{\text{transfer}2 \rightarrow 1} = 100 \text{ s}^{-1}$  was set to be 5 times faster than  $k_{\text{transfer}1 \rightarrow 2} = 20 \text{ s}^{-1}$ .



**Simulation II: Differential rates of cleavage.** In the second simulation we tested a model involving differential rates of cleavage between each uracil site.  $k_{\text{obs}}(\text{site } 1)$  was set to be  $0.05 \text{ nM}^{-1} \text{ s}^{-1}$ ,  $k_{\text{obs}}(\text{site } 2)$  was set to be  $0.01 \text{ nM}^{-1} \text{ s}^{-1}$  and the transfer rates were set to be equal ( $20 \text{ s}^{-1}$ ).

## 4.5: References

1. Stivers, J. T. (2008) Extrahelical damaged base recognition by DNA glycosylase enzymes., *Chemistry* 14, 786–793.
2. Longley, D. B., Harkin, D. P., and Johnston, P. G. (2003) 5-fluorouracil: mechanisms of action and clinical strategies., *Nat. Rev. Cancer* 3, 330–338.
3. Van Triest, B., Pinedo, H. M., Giaccone, G., and Peters, G. J. (2000) Downstream molecular determinants of response to 5-fluorouracil and antifolate thymidylate synthase inhibitors., *Ann. Oncol.* 11, 385–391.
4. Di Noia, J. M., and Neuberger, M. S. (2007) Molecular mechanisms of antibody somatic hypermutation., *Annu. Rev. Biochem.* 76, 1–22.
5. Stavnezer, J., Guikema, J. E. J., and Schrader, C. E. (2008) Mechanism and regulation of class switch recombination., *Annu. Rev. Immunol.* 26, 261–292.
6. Li, Z., Woo, C. J., Iglesias-Ussel, M. D., Ronai, D., and Scharff, M. D. (2004) The generation of antibody diversity through somatic hypermutation and class switch recombination., *Genes Dev.* 18, 1–11.
7. Porecha, R. H., and Stivers, J. T. (2008) Uracil DNA glycosylase uses DNA hopping and short-range sliding to trap extrahelical uracils., *Proc. Natl. Acad. Sci. U.S.A.* 105, 10791–10796.

8. Schonhofs, J. D., and Stivers, J. T. (2012) Timing facilitated site transfer of an enzyme on DNA., *Nat. Chem. Biol.* 8, 205–210.
9. Halford, S. E., and Marko, J. F. (2004) How do site-specific DNA-binding proteins find their targets?, *Nucleic Acids Res.* 32, 3040–3052.
10. Berg, H. C. (1993) Random walks in biology. Princeton University Press.
11. Stanford, N. P., Szczelkun, M. D., Marko, J. F., and Halford, S. E. (2000) One- and three-dimensional pathways for proteins to reach specific DNA sites., *EMBO J.* 19, 6546–6557.
12. Berg, O. G., Winter, R. B., and Hippel, Von, P. H. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory, *Biochemistry* 20, 6929–6948.
13. Murphy, M. C., Rasnik, I., Cheng, W., Lohman, T. M., and Ha, T. (2004) Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy., *Biophys. J.* 86, 2530–2537.
14. Chen, H., Meisburger, S. P., Pabit, S. A., Sutton, J. L., Webb, W. W., and Pollack, L. (2012) Ionic strength-dependent persistence lengths of single-stranded RNA and DNA., *Proc. Natl. Acad. Sci. U.S.A.* 109, 799–804.
15. Berg, H. C. (1993) Random walks in biology. Princeton University Press.
16. Maul, R. W., Maul, R. W., Gearhart, P. J., and Gearhart, P. J. (2010) AID and Somatic Hypermutation, in *Advances in Immunology* (Alt, F. W., and Alt, F. W., Eds.), pp 159–191. Academic Press.

17. Bransteitter, R., Pham, P., Calabrese, P., and Goodman, M. F. (2004) Biochemical analysis of hypermutational targeting by wild type and mutant activation-induced cytidine deaminase., *J. Biol. Chem.* **279**, 51612–51621.
18. Pham, P., Chelico, L., and Goodman, M. F. (2007) DNA deaminases AID and APOBEC3G act processively on single-stranded DNA., *DNA Repair* **6**, 689–92– author reply 693–4.
19. Pham, P., Bransteitter, R., Petruska, J., and Goodman, M. F. (2003) Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation., *Nature* **424**, 103–107.
20. 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 Res.* **37**, 3493–3500.
21. Parker, J. B., Bianchet, M. A., Krosky, D. J., Friedman, J. I., Amzel, L. M., and Stivers, J. T. (2007) Enzymatic capture of an extrahelical thymine in the search for uracil in DNA, *Nature* **449**, 433–437.
22. Parker, J. B., and Stivers, J. T. (2008) Uracil DNA glycosylase: revisiting substrate-assisted catalysis by DNA phosphate anions., *Biochemistry* **47**, 8614–8622.
23. Jiang, Y. L., Ichikawa, Y., Song, F., and Stivers, J. T. (2003) Powering DNA Repair through Substrate Electrostatic Interactions, *Biochemistry* **42**, 1922–1929.

24. Xiao, G., Tordova, M., Jagadeesh, J., Drohat, A. C., Stivers, J. T., and Gilliland, G. L. (1999) Crystal structure of Escherichia coli uracil DNA glycosylase and its complexes with uracil and glycerol: structure and glycosylase mechanism revisited., *Proteins* 35, 13–24.
25. Schurr, J. M. (1979) The one-dimensional diffusion coefficient of proteins absorbed on DNA: Hydrodynamic considerations, *Biophysical chemistry* 9, 413–414.
26. Bagchi, B., Blainey, P. C., and Xie, X. S. (2008) Diffusion Constant of a Nonspecifically Bound Protein Undergoing Curvilinear Motion along DNA, *J. Phys. Chem. B* 112, 6282–6284.
27. Chelico, L., Pham, P., Calabrese, P., and Goodman, M. F. (2006) APOBEC3G DNA deaminase acts processively 3' → 5' on single-stranded DNA, *Nat. Struct. Mol. Biol.* 13, 392–399.
28. Senavirathne, G., Jaszczur, M., Auerbach, P. A., Upton, T. G., Chelico, L., Goodman, M. F., and Rueda, D. (2012) Single-stranded DNA scanning and deamination by APOBEC3G cytidine deaminase at single molecule resolution., *J. Biol. Chem.* 287, 15826–15835.
29. Roberts, V. A., Pique, M. E., Hsu, S., Li, S., Slupphaug, G., Rambo, R. P., Jamison, J. W., Liu, T., Lee, J. H., Tainer, J. A., Eyck, Ten, L. F., and Woods, V. L. (2012) Combining H/D exchange mass spectroscopy and computational docking reveals extended DNA-binding surface on uracil-DNA glycosylase., *Nucleic Acids Res.* 40, 6070–6081.



30. Krosky, D. J., Schwarz, F. P., and Stivers, J. T. (2004) Linear Free Energy Correlations for Enzymatic Base Flipping: How Do Damaged Base Pairs Facilitate Specific Recognition?, *Biochemistry* 43, 4188–4195.
31. Krosky, D. J., Song, F., and Stivers, J. T. (2005) The Origins of High-Affinity Enzyme Binding to an Extrahelical DNA Base, *Biochemistry* 44, 5949–5959.
32. Pérez-Durán, P., Belver, L., de Yébenes, V. G., Delgado, P., Pisano, D. G., and Ramiro, A. R. (2012) UNG shapes the specificity of AID-induced somatic hypermutation., *J. Exp. Med.* 209, 1379–1389.
33. Mazur, A. K. (2006) Evaluation of elastic properties of atomistic DNA models., *Biophys. J.* 91, 4507–4518.

## Chapter 5

# Integration of Biochemical Site Transfer Data Into A Complete Model Using Stochastic Simulations

Reproduced in part from:

Rowland M, Schonhoft JD, and Stivers JT. (2014) Microscopic Mechanism of DNA Damage Searching of hOGG1. *Submitted*

## 5.1: Introduction

Recently, experiments performed in collaboration with Chloe Rowland, a current post doctoral fellow, have extended the Molecular Clock approach described for hUNG in Chapter 2 to human 8-oxoguanine DNA Glycosylase (hOGG1). A summary of the data is presented in **Figure 5.1** for reference. The observations made for hOGG1 are surprisingly similar to that of hUNG, despite hOGG1 being part of a different DNA glycosylase superfamily (Helix-hairpin-Helix (HhH)) and having taken a different evolutionary route (**Figure 5.2**).

In the site transfer assay, for both proteins, high concentrations of the trap eliminate transfer at long lesions spacings indicating the use of a ‘hopping’ or dissociative search pathway. At short lesion spacings a non-zero plateau in site transfer is observed at high concentrations of the trap indicating the use of an associative or untrappable kinetic pathway. For simplicity I refer to this pathway as ‘sliding’, although as noted below the term ‘sliding’ can be misleading. The mean sliding length ( $L_{\text{slide}}$ ) for hUNG and hOGG1 is 4 bp and 9 bp respectively and both enzymes use cycles of sliding and hopping to increase their searchable footprint. The searchable footprint is similar for hOGG1 and hUNG and upon binding both enzymes are able to cover an average distance on the DNA of ~40-60 bp (**Figure 5.3**).

The term ‘sliding’ can be deceptive and many interpretations exist in the current literature. The most widely used is one where the protein is able to ‘slide’ without friction or random walk on DNA while maintaining constant contact (1). One obvious shortcoming of this interpretation is that all DNA binding proteins

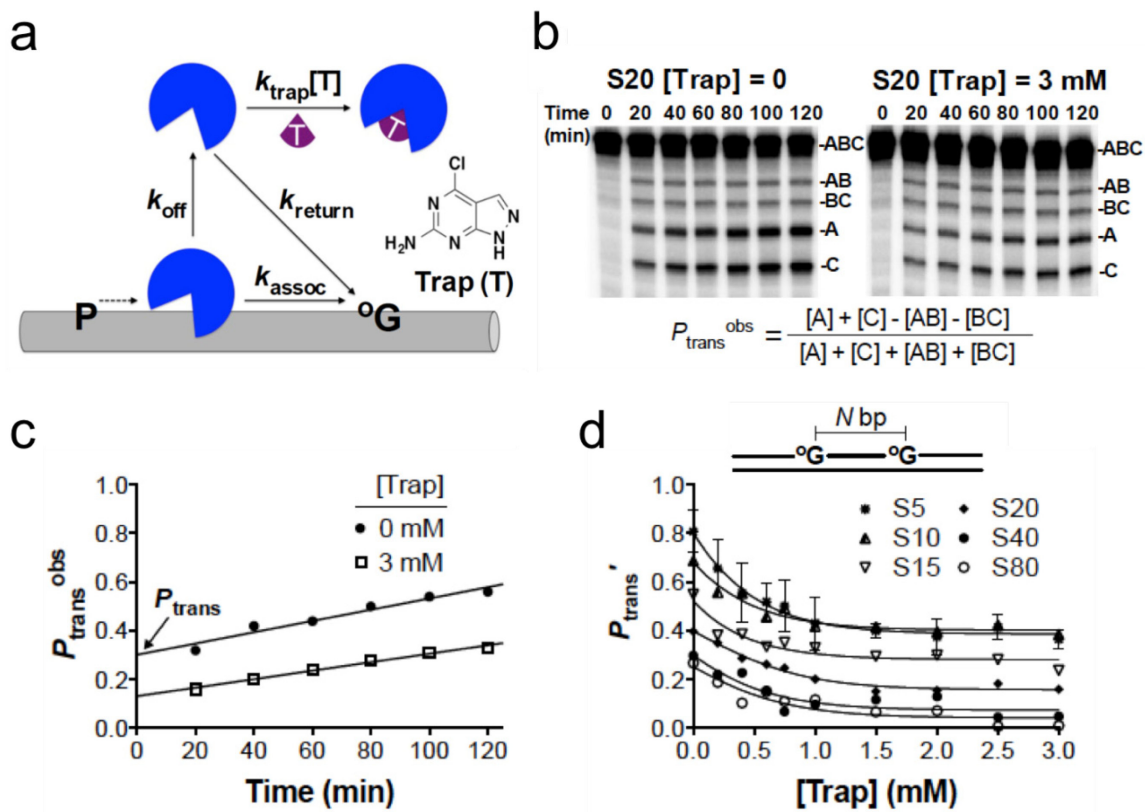
make direct interactions with the DNA and must be able to eventually recognize specific sites. It is not clear at this point how a protein balances these kinetic and thermodynamic requirements. Accordingly in my thesis 'sliding' simply refers to the untrappable kinetic pathway without applying any mechanistic constraints. At the end of this chapter I discuss the properties of a 'sliding' enzyme based on the available data.

A major advantage in studying hOGG1 is that single molecule fluorescence data from the Sunney Xie group is available for direct comparison (2, 3). In this work the authors observed hOGG1 translocation on flow stretched lambda DNA that had the following properties: i) hOGG1 was able to cover enormous lengths of DNA with an average translocation length of approximately 500-1000 bp under comparable conditions as the bulk biochemical experiments performed in our lab, ii) The speed or 1D diffusion coefficient ( $D_1^{\text{app}}$ ) on DNA was determined to be in the range of  $2-5 \times 10^6$  bp<sup>2</sup>/sec, and iii) with increasing ionic strength, the time and translocation distance decreased proportionally and the value for  $D_1^{\text{app}}$  remained unchanged. This  $D_1^{\text{app}}$  value was interpreted as 'barrierless' diffusion on DNA because it approaches  $\sim 0.5 k_bT$  of the maximum possible diffusion along DNA with hydrodynamic drag from helical rotation around the helix (The 'Schurr' Model (4)). Collectively the authors interpreted these observations in a model where hOGG1 slides frictionlessly along the helical path of the DNA without breaking contact.

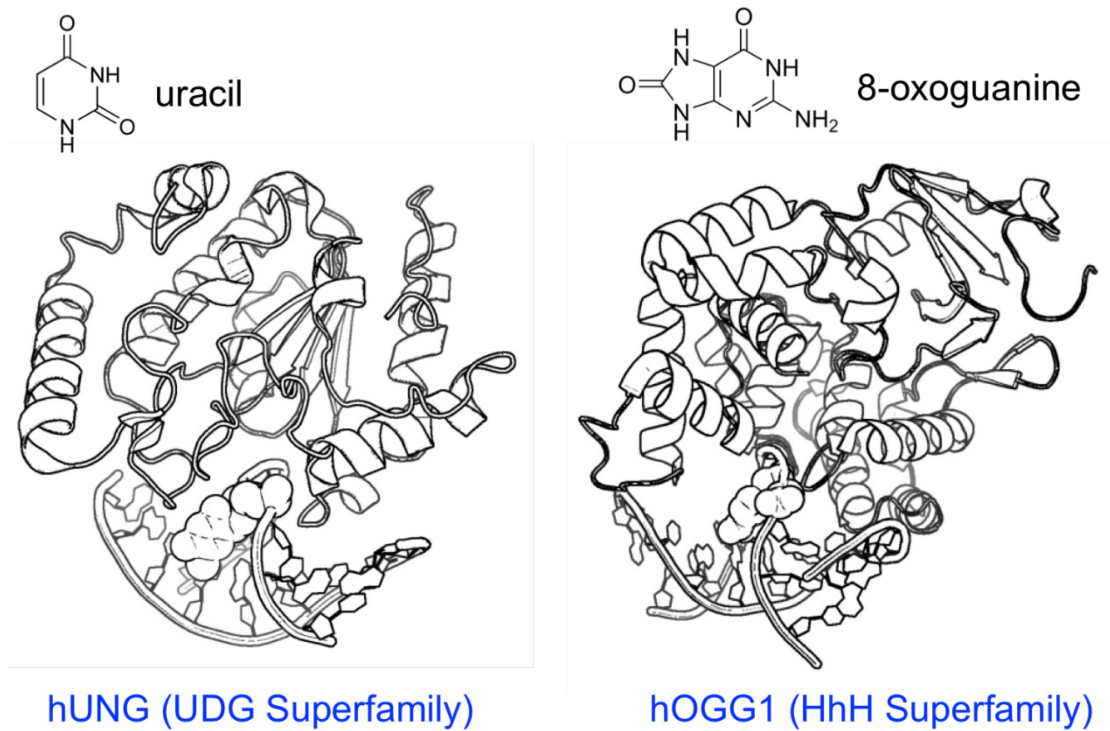
The observations and interpretation for hOGG1 are seemingly not consistent with the bulk biochemical data collected in our lab. The most obvious

differences are the mean translocation length (~60 bp versus ~500 bp) and the observed bound lifetime on non-specific DNA (3 ms versus ~100-200 ms). Therefore I sought to understand why these differences arise between the two approaches using direct Monte Carlo simulations of the facilitated diffusion model. Specifically, with these simulations in this chapter I set out to answer such questions as:

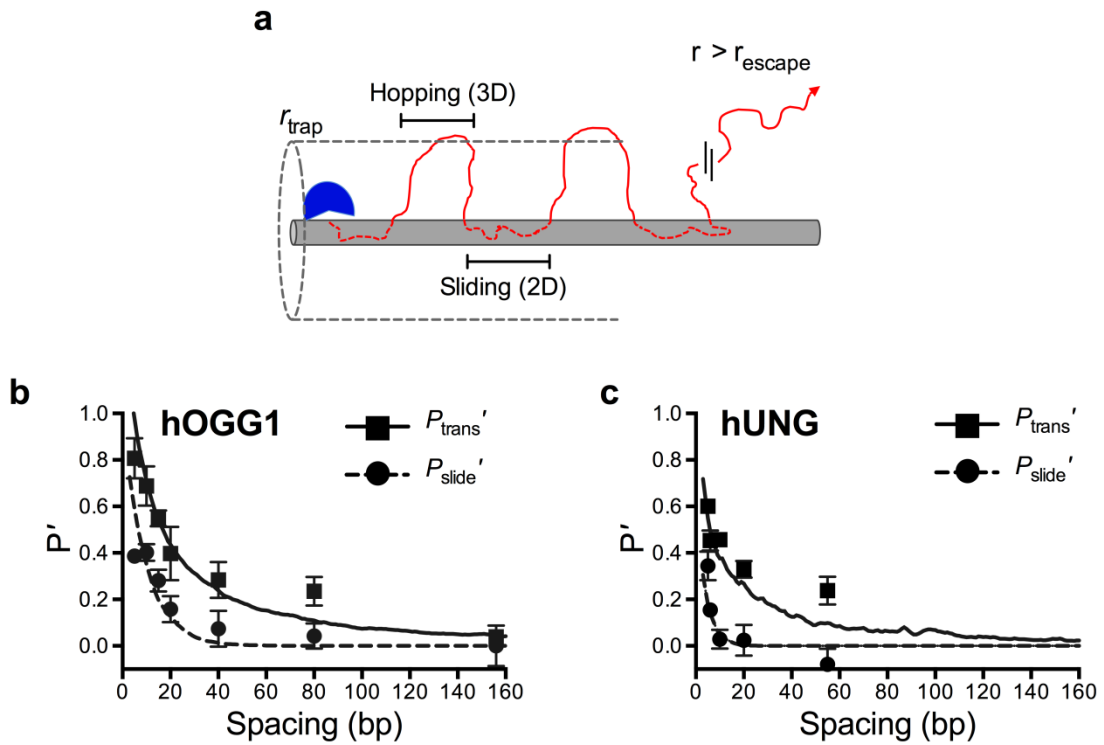
- Would a single molecule experiment be able to distinguish between 'hopping' and 'sliding'?
- Why do bulk averaged biochemical and single molecule experiments arrive at such disparate conclusions?
- How does hopping contribute to the searchable footprint?
- What is the expected distance distribution of a hopping enzyme?
- What is the effect of hopping on the apparent diffusion coefficient on DNA?
- What are the quantitative differences between hOGG1 and hUNG



**Figure 5.1. Summary of Molecular Clock data for hOGG1.** Figure and data are presented for reference and are courtesy of Chloe Rowland Ph.D (Stivers lab). (a) Molecule clock method for trapping a microscopically dissociated enzyme. The chemical structure of the Trap used for hOGG1 is depicted. (b) Representative data with and without Trap for site transfer of hOGG1 using a DNA substrate containing two 8-oxoguanine sites separated by 20 bp. (c) Determination by linear fit of  $P_{\text{trans}}^{\text{obs}}$  to determine  $P_{\text{trans}}$  for the data in panel b. (d) Dependence of  $P_{\text{trans}}'$  with the concentration of Trap. The prime indicates correction by the excision efficiency ( $P_{\text{trans}}' = P_{\text{trans}}/E$ ).



**Figure 5.2. Comparison of hOGG1 (2I5W) and hUNG (2OXM) DNA complexes.** hOGG1 and hUNG are the primary representatives of the helix-hairpin-helix (HhH) superfamily and UDG superfamily respectively.



**Figure 5.3. Comparison of the searchable footprint of hUNG and hOGG1.** The percentage of transfer corrected for the efficiency ( $E$ ) of cleavage once the enzyme has encountered a lesion site ( $P_{trans}' = P_{trans}/E$ ) determined using the site transfer assay. **(a)** Schematic of the facilitated diffusion model used for numerical simulations (see text). **(b & c)** Data and fit to the facilitated diffusion model for hOGG1 and hUNG.



## 5.2: Results and Discussion

### 5.2.1 Basic Description

Although Monte Carlo simulations have been previously used to model mechanisms of protein-DNA facilitated diffusion (5-7), the approach here is distinct because experimental restraints obtained from molecular clock and kinetic measurements are used to place boundaries on the simulations. For a single trajectory in the Monte Carlo simulation, the DNA is modeled along the  $x$ -axis and the protein starts its diffusion trajectory at  $x_{\text{pos}} = 0$  and undergoes a random walk on the DNA or 'slides' (see below) until a dissociation event occurs. The frequency of dissociation events, and thus the average time for a sliding transfer, are constrained by experimental measurements as described below. Once dissociated from the DNA, the protein undergoes a stepwise 3D random walk (*e.g. hops*) that leads to reassociation with the DNA or diffuses to a distance ( $r$ ) away from the DNA where reassociation is highly improbable (*e.g. escapes* when  $r > r_{\text{escape}}$ ). When escape occurs, the trajectory is terminated. If the protein returns to the DNA chain, it will undergo another round of sliding and the cycle repeats until the protein escapes to bulk (**Figure 5.4**).

The model does not require any force constants between hOGG1 and the DNA. Instead, the time frames for the associative and dissociative processes are obtained from the molecular clock data, the binding lifetime of hOGG1 and the Monte Carlo statistics (see below). Specifically, the likelihood of a sliding transfer over a given distance is given by the experimental site spacing dependence of  $P_{\text{slide}}$ . This probability reports on the stepwise likelihood at which hOGG1 exits

the associative sliding mode and enters the dissociative transfer mode  $p_{\text{exit}} = k_{\text{exit}}/(k_{\text{step}} + k_{\text{exit}})$ . The timing for the 3D steps in solution is simply given by the free diffusive behavior of hOGG1.

The efficiency of the trap offers an additional experimental constraint. As described in Chapter 2, the concentration of the trap at which all of the hopping pathway is eliminated can be used to estimate the resolution of the molecular clock method. Using the concentration at which all enzyme molecules are trapped and the rate of trapping, ~3 mM for hOGG1 and hUNG, the maximum resolution of the trapping method is estimated to be approximately 3 nm (see Section 5.3 Methods). That is, once hOGG1 has diffused an average of 3 nm away from the DNA during the dissociated lifetime it is likely to be trapped near 100% of the time at mM concentrations of the trap. In the simulations we introduce another parameter,  $r_{\text{trap}}$ , which is the minimum distance an untrappable hop, as short lived microscopic hops below this ~3 nm threshold are outside of the resolution limits of the method. In summary, a single trajectory comprises cycles of sliding by random walk on the DNA followed by hops that are bound between the distance of  $r_{\text{trap}}$  and  $r_{\text{escape}}$ . The effects of varying both  $r_{\text{escape}}$  and  $r_{\text{trap}}$  are described in detail in the Methods (**Section 5.3**).

In simple terms, the simulations ask the question of whether hopping is a viable mechanism to explain the difference in the searchable footprint between  $P_{\text{trans}}$  (without trap) and  $P_{\text{slide}}$  (with trap). Another important point is that sliding is referred to here as simply the untrappable kinetic pathway, and while in the simulations we model this pathway empirically as a random walk on DNA, in

reality the actual mechanism of translocation is largely unknown and we can only infer with these experiments that the protein does not diffuse far away from the DNA during this period as it is not trapped by the high concentration of the small molecule. The Monte Carlo simulation program were implemented in Python and a full description is presented in the Methods (**Section 5.3**).

### **5.2.2 Major Findings for hOGG1 simulations**

At this juncture, the molecular clock approach has been successfully applied to both hOGG1 and hUNG. The data indicate the presence of at least two separable kinetic pathways in the facilitated DNA lesion search of these enzymes. The mechanism can be generally characterized by the diffusion limited association with the DNA followed by cycles of short range hopping and associative sliding transfers eventually leading to lesion recognition.

A simple inspection of the data for hOGG1 reveals several important properties. The searchable footprint of hOGG1 upon binding is on average ~60 bp while the untrappable pathway ( $P_{\text{slide}}$ ) persists over a shorter distance of ~10 bp. The dissociative or trappable pathway contributes significantly to the apparent footprint of the enzyme. At site separations greater than ~150 bp no detectable transfer is observed. These distance dependences fit well to the predicted distance dependence from the numerical simulations (solid and dashed lines, **Figure 5.3b&c**). Thus the expanded footprint of 60 bp that is derived from the addition of the trappable kinetic pathway can be explained quantitatively by

'hopping' and the simple geometry of the protein and the DNA and the diffusive properties of the protein in solution.

This approach allows useful boundaries to be placed on key transfer events that occur during the average lifetime of the bound enzyme ( $\tau_{\text{bind}} = 3 \text{ ms} = K_D \times k_{\text{on}}$ ). This ensemble bound lifetime consists of all events that occur between the initial encounter and escape from the DNA and thus include all sliding and hopping transfers on a single DNA. Thus a 'bound' protein consists of all events that occur between the initial encounter and escape. A quantitative argument of why this is the case and its consistency with other bulk biochemical data is presented in the Methods (**Section 5.3**).

Several quantitative parameters derived from the simulations of hOGG1 include (i) the average number of sliding transfer events per binding event (four), (ii) the associative transfer time and mean transfer distance (ten to hundreds of microseconds and 9 bp), (iii) the dissociative hop time and median distance (ten to hundreds of nanoseconds and 20 bp), and (iii) the average damage search footprint that arises from the use of both transfer pathways (~60 bp). The analysis and distributions obtained from the simulations for hOGG1 are presented in **Figure 5.5** and model in **Figure 5.6**.

### **5.2.3 Reconciling the ensemble and the single molecule observations**

The major findings of single molecule studies for hOGG1 (2, 3), which differ significantly from our ensemble studies, are as follows: (i) the apparent 1D diffusion constant ( $D_1^{\text{app}}$ ) was determined to be in the range of  $2\text{-}5 \times 10^6 \text{ bp}^2/\text{sec}$

and was determined to be salt independent, (ii) Under similar ionic strength conditions the total length of DNA covered within the flow cell was much larger (~900 bp versus ~60 bp), and (iii) very long binding lifetimes of ~100-200 ms were observed compared with the ensemble lifetime of 3 ms.

The numerical simulations were used to calculate the expected  $D_1^{\text{app}}$  value using the combined sliding and hopping mechanism. To answer this we calculated the mean-squared displacements from our simulation trajectories and plotted these against the calculated time for each trajectory (See Methods, **Figure 5.6d**). The slope of this plot gave  $D_1^{\text{app}} = 3.8 \times 10^6 \text{ bp}^2/\text{sec}$ . Thus, the ensemble measurements can recapitulate the macroscopic  $D_1^{\text{app}}$  even though the microscopic process is not 1D as originally concluded in Blainey *et al.* (2, 3). In essence, the rapid microscopic pathways detected in the molecular clock method are blurred into an apparently 1D process due to the distance and time resolution limits of the single-molecule TIRF measurements<sup>33</sup>.

#### 5.2.4 Probability of long time scale translocation events

Previous single molecule studies of hOGG1 observed long time scale translocation events lasting an average of 100-200 ms under conditions similar to ours, which is difficult to reconcile with the ensemble lifetime on nonspecific DNA of 3 ms. Here I estimate the probability that hOGG1 undergoes translocations of this duration using our ensemble measurements and numerical simulations. The trajectories that were used to model  $P_{\text{trans}}$  were binned according to their individual total trajectory times ( $\tau_{\text{tot}(i)}$ , see Methods). The fractional occurrences of

the binned values were fitted to a single exponential ( $P(t) = \lambda e^{-\lambda t}$ )<sup>13</sup> to extrapolate the probability of rare events occurring at longer relative times (**Figure 5.5e**). The mean of the exponential distribution ( $1/\lambda$ ) is equal to the ensemble binding lifetime (3 ms). Extrapolation to long time scales indicated that events lasting longer than the single-molecule frame rate of 10 ms occur at a probability of  $\ll 1\%$ . Further this is the absolute minimum time for observation as tracking a single particle requires several frames and thus the real resolution is higher in the  $\sim 100$  ms range(8). We conclude that the single-molecule measurements must selectively detect rare long-lived events that are not observable in the ensemble measurements.

### 5.2.5 Modeling the effects of salt on $D_1^{\text{app}}$

Previously it has been argued for hOGG1 and other DNA binding proteins that DNA sliding involving continuous DNA contact is proven if  $D_1^{\text{app}}$  is independent of salt concentration(2, 3, 9, 10). I surmised that the major effect of increasing ionic strength is to reduce the probability of hOGG1 rebinding to the DNA after dissociation. This is supported by the observation that the excision rate constant of  $G^0$  is salt insensitive, indicating that salt does not perturb the ES complex. Thus, the decrease in nonspecific DNA binding affinity with increasing salt is most reasonably attributed to salt screening effects on the association rate ( $E + S \rightarrow ES$ ). We modeled the effect of salt on association by introducing a parameter ( $p_{\text{bind}}$ ) that serves to decrease the efficiency of rebinding during dissociative transfers.  $p_{\text{bind}}$  is the probability of entering the sliding pathway upon

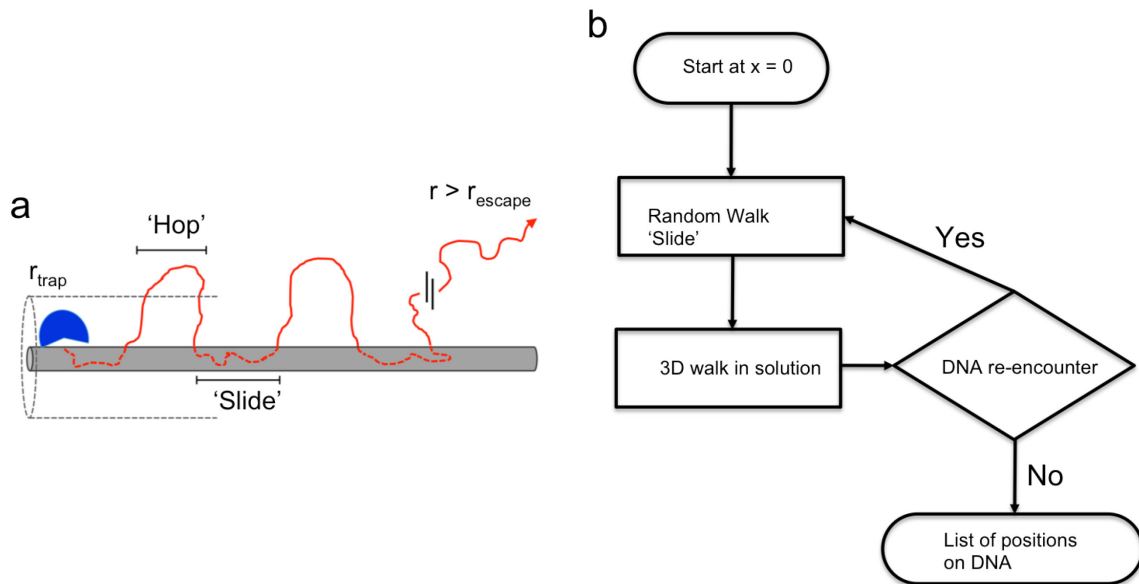
encounter with the DNA. For example if  $p_{\text{bind}} = 0.1$ , 10% of encounters are productive and result in a cycle of associative sliding, while the remainder are reflected from the DNA and continue to diffuse in 3D.

I calculated  $D_1^{\text{app}}$  and  $\tau_{\text{bind}}$  for simulations where  $p_{\text{bind}}$  was varied in the range 1 to 0.01. The simulated dependences of the calculated values for  $\tau_{\text{bind}}$ , the mean squared displacements (MSD), and the  $D_1^{\text{app}}$  as a function  $p_{\text{bind}}$  are shown in **Figure 5.5f & g**. These simulations showed that  $\tau_{\text{bind}}$  and MSD decrease as  $p_{\text{bind}}$  decreases, while  $D_1^{\text{app}}$  remains nearly constant. The calculated values for  $D_1^{\text{app}}$  were in the range  $3.1 \times 10^6$  to  $2.3 \times 10^6$  bp<sup>2</sup>/sec. This model makes the assumption that associative transfer times remain unchanged which is supported by the experimental observations with the excision rate described above and also the prior determination that the number of associated ions around the DNA remains essentially constant within the range of monovalent ion concentration relevant here (0.01 – 0.10 M)(11). Indeed recent findings in the lab with hUNG have shown that the association rate is salt dependent while the dissociation rate is insensitive (Shannen Cravens, unpublished). These findings are readily understood because free diffusion in solution is independent of salt concentration, and therefore,  $D_3$  for the dissociative transfers is not changed by salt. However, salt results in fewer productive encounter events, leading to reduced mean-squared displacements (arising from fewer opportunities for sliding transfers) and more frequent escape to bulk. In conclusion, salt effects cannot unambiguously distinguish between transfer mechanisms.

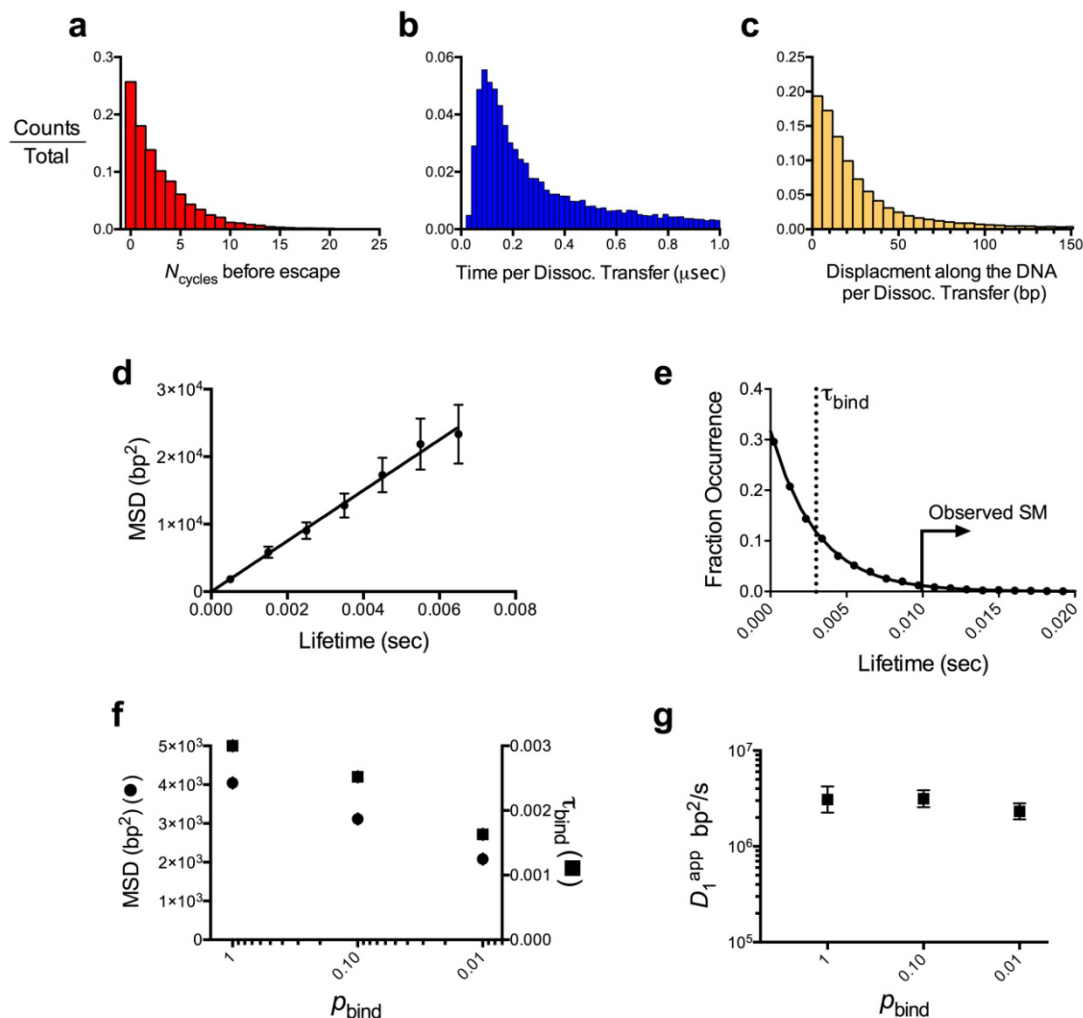
### 5.2.6 A general search mechanism for DNA glycosylases hOGG1 and hUNG

Similar results were obtained in simulations for hUNG. The median searchable footprint obtained was 52 bp compared to 64 bp for hOGG1 and  $D_1^{\text{app}}$  value was  $7.9 \times 10^6 \text{ bp}^2/\text{sec}$ . Qualitatively this is expected as both proteins have similar translocation lengths and identical binding lifetimes ( $\tau_{\text{bind}} = 3 \text{ ms}$ ). A quantitative model for both enzymes derived from the data and simulations is presented in **Figure 6.1**. Thus it seems that despite the early divergence in evolution of these two enzymes, that the search mechanism is conserved.





**Figure 5.4 Facilitated diffusion model (a) and pseudocode flowchart (b).** The protein begins on the DNA, modeled along the x-axis, and undergoes one round of random walk sliding. Then the protein undergoes a 3D walk through solution and if captured again by the DNA re-enters the 'sliding' mode. However if the protein diffuses beyond the threshold of rescape the trajectory is finished. At each round of sliding all positions on the DNA are recorded.



**Figure 5.5. Simulated microscopic properties of site transfer for hOGG1.** (a) Histogram showing the probability distribution for making  $N_{\text{cycles}}$  of associative-dissociative transfers before escape to bulk (bin size = 1). For the average trajectory  $N_{\text{cycles}} = 3$ . (b) Histogram of the time taken during a single dissociative hopping event (bin size = 0.02  $\mu\text{sec}$ ). Once outside of  $r_{\text{trap}}$ , hOGG1 reassociates with the DNA with a median time of 360 nanosec. (c) Histogram of the displacements along the DNA axis for single dissociative transfer events (bin size = 2 bp). The median displacement of a single dissociative transfer is 18 bp. (d) Histogram of the mean squared displacement for 10,000 total trajectories binned according to time (bin size = 1 ms). Calculation of the time for each trajectory is described in the Methods section.  $D_1^{\text{app}}$  value determined from the slope is  $3.8 \times 10^6 \text{ bp}^2/\text{sec}$ . Error bars represent  $\pm 1$  standard error of the mean for each 1 ms bin. (e) The data points are the fraction of trajectories (Total = 10,000) that occurred with a given lifetime and the fit is to a single exponential ( $P(t) = \lambda e^{-\lambda t}$ ), where  $1/\lambda$  corresponds to the mean or ensemble lifetime of 3 ms (see text). The solid line marks the minimum binding lifetime (10 ms) that could be observed in single-molecule experiments of hOGG1 as calculated from the reported

maximum frame rate of 100 per second<sup>29</sup>. **(f)**  $\tau_{\text{bind}}$  and the MSD were calculated for each value of the binding efficiency ( $\rho_{\text{bind}}$ )(see text). **(g)** Calculated apparent diffusion coefficients ( $D_1^{\text{app}}$ ) for decreasing values of  $\rho_{\text{bind}}$ . 500 total trajectories were run for each value of  $\rho_{\text{bind}}$ . The  $D_1^{\text{app}}$  values ranged from  $3.1 \times 10^6$  to  $2.3 \times 10^6$  bp<sup>2</sup>/s as  $\rho_{\text{bind}}$  was decreased from 1 to 0.01. Error bars represent the 95% confidence interval from the linear fit of MSD versus time using *Method II* (see Methods).

## 5.3 Methods

### 5.3.1 Modeling sliding transfers: a random walk on DNA

Several observations from the Stivers lab indicate that strand sliding along the helix is not a viable mechanism for transfer along the DNA by sliding. Most notably, in site transfer experiments with hOGG1, a covalently attached small molecule ‘roadblock’ on the lesion strand between 8-oxoguanine sites had no effect on site transfer by sliding or hopping. Also for hOGG1 transfer to the opposite DNA strand was only minimally affected by the presence of the trap. Given these and other experimental observations, I modeled sliding transfers as a 2D random walk with arbitrary bp length steps ( $\delta_{2D} = 0.34$  nm). A 1D random walk on the DNA chain (e.g. on a line) could also be used and in this model 1D and 2D walks result in identical outcomes. For hOGG1 the 2D random walk model was chosen as it is most consistent with the actual data, and clarity in publication we chose this model to abate confusion. For hUNG a 1D random walk on a line was used.

In this model the protein (3.2 nm radius of gyration) has an equal probability of stepping in any direction along the DNA cylinder (1 nm radius) with the x-coordinate oriented along the cylinder longitudinal axis. The probability of dissociation at each step in the random walk [ $p_{\text{exit}} = k_{\text{exit}} / (k_{\text{step},2D} + k_{\text{exit}})$ ] was determined empirically by simulating 2D random walks on the DNA using varying values of  $p$  in the range of 0.1 to 0.001 (where  $k_{\text{step},2D}$  is the 2D step rate and  $k_{\text{diss}}$  is the dissociation rate constant at each step). The same procedure and model was used for hUNG, however a 1D random walk model was used. From these

simulations, we recorded  $x$ -components of the random walk steps along the DNA axis. The number of protein molecules that successfully executed sliding transfers (defined as transfers that do not result in dissociation beyond  $r_{\text{trap}}$ ) to points along the  $x$ -axis were binned into histograms using a 1 bp increment with the largest bin from 0-1 bp normalized to one. These distributions, which are sensitive to the chosen  $p_{\text{exit}}$  value, were fitted to the distance dependence of the untrappable component of the experimental data ( $P_{\text{slide}}$ ) (**Figure 5.6**). The best fit to the experimental distribution was empirically obtained using a  $p_{\text{exit}} = 0.0028$  (hOGG1) and 0.0385 (hUNG). The choice of step length is arbitrary, and controls varying the step length had no effect on the outcome of the simulations.

### 5.3.2 Modeling dissociative transfers: 3D random walk

To begin a dissociative transfer (3D), the protein exits an associative step on the DNA cylinder with probability  $p_{\text{exit}}$  (see above). The protein is allowed to move in the  $x$ ,  $y$ ,  $z$  directions using a step length ( $\delta_{3D}$ ) and time ( $\zeta_{3D}$ ) as calculated below and only dissociation events that exceed  $r_{\text{trap}}$  are used. The direction of each step was determined by picking random points on a sphere<sup>5</sup>. 3D steps were continued until a step resulted in reencounter with the DNA surface or escape to bulk solution occurs ( $r > r_{\text{escape}}$ , see below). A reencounter event is defined as when the center-to-center distance of the protein and DNA are equal to the sum of their respective radii (**Figure 5.7a**).

**3D step length and time.** The law of equipartition of energy was used to calculate the length ( $\delta_{3D}$ ) and time ( $\zeta_{3D}$ ) for each discretized step in a 3D random

walk for a sphere the size of hOGG1 ( $r_{\text{gyration}} = 3.2 \text{ nm}$ ) (eq 1)<sup>4,6</sup>. Employing a temperature of 310 K and the molecular mass of a single hOGG1 molecule, the instantaneous velocity ( $v_{\text{rms}}$ ) was determined to be 14.3 m/sec using eq 1. Using this velocity, the relations in eq 2 and a  $D_3 = 1.03 \times 10^8 \text{ nm}^2/\text{sec}$ , a value  $\delta_{3D} = 0.433 \text{ \AA}$  and  $\zeta_{3D} = 3.03 \text{ ps}$  were calculated. As a control, simulations using Gaussian distributed step lengths around the discrete mean values of  $\zeta_{3D} 0.443 \text{ \AA}$  and  $\zeta_{2D} = 0.34 \text{ nm}$  also yielded identical results.

$$[1] \quad v_{rms} = \sqrt{\frac{3k_b T}{mass}} = \frac{\delta_{3D}}{\zeta_{3D}}$$

$$[2] \quad D_3 = \frac{\delta_{3D}^2}{6 \zeta_{3D}} = \frac{k_b T}{6 \pi \eta r}$$

**Calculation of  $r_{\text{trap}}$ .** In the Monte Carlo simulations, a large majority of the dissociation events do not result in displacements from the DNA cylinder greater than 1 nm (see reference<sup>4</sup>). Since these pseudo-dissociation steps immediately lead to reencounter with the DNA and are observed as part of the associative transfers in the experimental measurements, these events are already accounted for in the 2D random walk model of associative transfer described above. This is justified because (i) these events occur with exceedingly short lifetimes (<1 ns for dissociations <1 nm) and cannot be captured using millimolar concentrations of a diffusion-controlled trap (see below), (ii) they occur within the environment of the DNA ion cloud<sup>7</sup> and (iii) the enzyme is in a crowded environment (e.g. in close

proximity to the DNA) where free access to the trap is hindered. For any or all of these reasons the trap has a limiting distance over which it is effective ( $r_{\text{trap}}$ ) and any dissociation-rebinding events that do not exceed this radius are already counted as associative transfers.

Thus the simulated dissociative transfers must only include dissociation events that move far enough away from the DNA such that the protein can be effectively trapped ( $r_{\text{trap}}$ ) (**Figure 5.7a-b**). A reasonable estimate for  $r_{\text{trap}}$  can be obtained from the diffusion-controlled trapping rate determined from the Smoluchowski equation ( $k_{\text{trap}} = 3 \times 10^{10} \text{ M}^{-1} \text{ sec}^{-1}$ , **eq 3**) using a Stokes radius of hOGG1 and the trap of 3.2 nm ( $r_A$ ) and 0.29 nm ( $r_B$ ) respectively. At 3 mM trap the ensemble average trapping time is  $1/(k_{\text{trap}}[\text{trap}]) = 11 \text{ ns}$ , which we use as the minimum pseudo-dissociation lifetime for hOGG1 that can be effectively trapped. In this 11 ns trapping time a dissociated hOGG1 molecule can diffuse a root-mean-squared distance of 2.6 nm based on the Stokes-Einstein relation using  $D_3 = 1.03 \times 10^{-6} \text{ cm}^2/\text{sec}$ . This minimum trapping distance is essentially the maximum resolution of the trapping method assuming the trap is 100% efficient.

$$[3] \quad k_{\text{smol}} = \frac{4\pi(r_A + r_B)(D_A + D_B)}{1000}$$

Using simulations, the  $r_{\text{trap}}$  parameter was varied around this calculated value (0.5 – 10.0 nm) and we found that values of  $4 \leq r_{\text{trap}} \leq 6 \text{ nm}$  showed reasonable fits to the experimental  $P_{\text{trans}}$  and  $P_{\text{assoc}}$  data sets (**Figure 5.7c**, see

next section for description of fitting). Thus the simulations and calculations based on the experimental data are congruent with a maximum resolution of the trapping approach of  $\sim 2.6$  nm. The  $r_{\text{trap}}$  value was set to a best fit value of 5 nm in the simulations presented in **Fig. 3** of the main text. The agreement between the simulated effective trapping distance and the estimate derived from the assumption of diffusion-controlled trapping, provides confirmation that the trap operates within a factor of two of the diffusion controlled limit.

**Calculation of  $r_{\text{escape}}$ .** To model the dissociative pathway a distance threshold ( $r_{\text{escape}}$ ) must be chosen to define the diffusion distance where hOGG1 escapes to bulk solution and a transfer trajectory is terminated (i.e. when the probability of hOGG1 reassociating with the DNA it departed is no greater than with other DNA in the bulk). In our simulations, hOGG1 is considered lost to bulk solution when the 3D diffusion time exceeds the time ( $\tau_{\text{escape}}$ ) for an average particle to diffuse a root-mean-square distance of  $r_{\text{escape}}$  based on the Stokes-Einstein equation (**Figure 5.7a-b**). We varied the threshold for escape from 10 nm (0.16  $\mu\text{sec}$ ) to 300 nm (146  $\mu\text{sec}$ ) and found no discernible difference in our simulations for values greater than 100 nm (16  $\mu\text{sec}$ ) (**Figure 5.7c**). In other words, when hOGG1 has moved an average distance  $\geq 100$  nm away from the DNA it has a near zero probability of coming back to the same DNA molecule within a reasonable time frame. While in principle, long hops over 100 nm are possible and have been observed in single molecule experiments<sup>1,8</sup>, trajectories containing these events are exceedingly rare and are not relevant in our experiments as our DNA substrates are of finite length. Moreover, such rare



events have no effect on the ensemble average properties. In the simulations presented for hOGG1,  $r_{\text{escape}} = 150 \text{ nm}$  and  $\tau_{\text{escape}} = 37 \text{ }\mu\text{sec}$ .

### 5.3.3 Fitting the experimental site spacing dependences of $P_{\text{trans}}'$ and $P_{\text{slide}}'$

$P_{\text{trans}}'$ . Experimental data for the site spacing dependence of  $P_{\text{trans}}'$  was simulated using trajectories that contained both associative and dissociative transfers as defined above. The positions of all associative transfer steps in 10,000 total trajectories were binned into a histogram according to  $x$ -distance along the DNA cylinder (bin size = 1 bp). Thus this procedure simulates the number of times in 10,000 trajectories that hOGG1 visits each distance bin along the cylinder axis. The histogram has broad tails that arise from the dissociative transfer events and models the probability that a hOGG1 molecule will diffuse to a position a given distance away from the origin using the combined associative and dissociative pathways ( $P_{\text{trans}}'$ ).

$P_{\text{slide}}'$ . Simulations that included only associative transfers ( $r < r_{\text{trap}}$ ) were used to model the distance dependence of  $P_{\text{slide}}'$  (e.g. associative trajectories are terminated once hOGG1 has stepped off of the DNA). The simulation results from 10,000 trajectories were binned and treated as described above for  $P_{\text{trans}}'$ .

**Fitting the histograms to the experimental data.** After averaging over all trajectories, the distance dependent probability distributions for the dissociative and associative transfers were used to fit the experimental site spacing dependence of  $P_{\text{trans}}'$  and  $P_{\text{slide}}'$  by multiplying each histogram by a *single*

normalization factor determined by globally minimizing the sum-of-squares between the simulated curves and the experimental data. Because a single normalization factor was used, the distinct shape of the curves, as well as the relative differences between the two curves, are taken into account during the fitting (**Figure 5.8**). For both  $P_{\text{trans}}$  and  $P_{\text{slide}}$ , identical results were achieved for trajectories  $n = 500$  through 10,000 indicating that the simulations achieved convergence.

### 5.3.4 Calculation of microscopic parameters

The Monte Carlo simulations combined with time constraints obtained from ensemble measurements allow us to make estimates of several microscopic parameters that describe the associative and dissociative kinetic pathways. These parameters and their definitions are described below and correspond to the model in **Figure 6.1**. The parameters used in the final simulations were  $r_{\text{trap}} = 5$  nm,  $r_{\text{escape}} = 150$  nm,  $p_{\text{exit}} = 0.0028$ ,  $\delta_{3D} = 0.443$  Å, and  $\delta_{2D} = 0.34$  nm for hOGG1:

$\tau_{\text{bind}}$ : The binding lifetime of hOGG1 for nonspecific DNA from ensemble measurements.  $\tau_{\text{bind}}$  consists of the sum of times spent in all stationary states and the associative and dissociative transfer states (i.e.  $\tau_{\text{bind}} = 3$  ms =  $\tau_{\text{stat,tot}} + \tau_{\text{assoc,tot}} + \tau_{\text{diss,tot}}$ ). As described in the text, the stationary state and associative state are interconverted by a conformational change in

hOGG1.

$N_{\text{cycles}}$ : The number of associative-dissociative transfer cycles that occur in single binding event of duration  $\tau_{\text{bind}}$ .  $N_{\text{cycles}}$  is obtained from the numerical simulations and is determined in part by the exit probability ( $p_{\text{exit}} = 0.0028$ , see above) and the probability for enzyme rebinding or escaping after entering a dissociative transfer step, which is determined by the 3D part of the numerical simulations. Both probabilities are constrained by the experimental transfer probabilities (**Figure 5.3**). Associative transfers begin at a DNA encounter event, or alternatively, from the stationary state after the required conformational change.

$\tau_{\text{diss}}$ : The time for a **single** dissociative transfer as determined in numerical simulations.

$\tau_{\text{assoc}}$ : The time for a **single** associative transfer.  $\tau_{\text{assoc}} \ll \tau_{\text{stat}}$  because weak interactions are required for movement and the associative state is a high-energy state on the pathway for dissociation from the DNA.

$\tau_{\text{stat}}$ : The time spent in a **single** stationary state before initiating a cycle of associative and dissociative transfers.  $\tau_{\text{stat}}$  is likely to consume the major portion of  $\tau_{\text{bind}}$ . No movement is possible in the stationary state and a conformational change is required to generate the weakly bound associative state.

$D_1^{\text{app}}$ : The macroscopically observed apparent diffusion coefficient along the longitudinal DNA axis.

$\tau_{\text{bind}}$ :  $\tau_{\text{bind}} = K_D^* k_{\text{on}} = 1/k_{\text{off}} = 3 \text{ ms}$  may be calculated using the temperature adjusted diffusion controlled on-rate for hOGG1 ( $k_{\text{on}} = 5 \times 10^8 \text{ M}^{-1} \text{ s}^{-1}$ )<sup>9,10</sup> and its non-specific DNA dissociation constant ( $K_D = 0.7 \text{ }\mu\text{M}$ ). We note that during this macroscopic bound lifetime a single hOGG1 molecule executes many microscopic dissociation-reassociation cycles on a single DNA chain, but its time-averaged occupancy on the DNA is  $\sim 100\%$ . During the molecular clock measurements, individual enzyme molecules in the process of site transfer are not in equilibrium with bulk DNA, but instead, are positionally correlated with a single DNA molecule until a rare dissociation trajectory takes the enzyme past the escape radius. Thus, the molecular clock and numerical simulations can report on this dynamic aspect of binding.

$N_{\text{cycles}}$ : Numerical simulations of the site spacing dependent decay of  $P_{\text{trans}}$  provides the number of associative-dissociative cycles per binding event (**Figure 5.3**). From this analysis, hOGG1 performs three cycles on average (i.e. four associative transfers) before an exit event finally leads to bulk escape (**Figure 5.5a-c**). The range for  $N_{\text{cycles}}$  was one to four (25<sup>th</sup> to 75<sup>th</sup> percentile) for  $r_{\text{trap}} = 5 \text{ nm}$ . Although  $N_{\text{cycles}}$  is also dependent on the value for  $r_{\text{trap}}$ , the uncertainty in  $r_{\text{trap}}$  introduces negligible error in  $N_{\text{cycles}}$  (**Figure 5.5a-c**).

$\tau_{\text{diss}}$ : From statistical analysis of 10,000 trajectories, the mean value for  $\tau_{\text{diss}} = \langle n^{3D} \rangle \zeta^{3D} = 2.3 \text{ }\mu\text{s}$ , where  $\langle n^{3D} \rangle$  is average number of 3D steps taken in a dissociative transfer. Due to the skewed distribution (**Figure 5.5c**), the mode for

$\tau_{\text{diss}}$  was 0.09  $\mu\text{s}$  and ranged from 0.15 - 1.43  $\mu\text{s}$  (25<sup>th</sup> to 75<sup>th</sup> percentile, median = 0.36  $\mu\text{s}$ ). The distribution of the displacements along the DNA axis arising from the initial and final points of dissociative transfers generated in 10,000 trajectories is shown in **Figure 5.5c**.

**$\tau_{\text{stat}}$  and  $\tau_{\text{assoc}}$ .** Since  $\tau_{\text{bind}} \sim \tau_{\text{stat, tot}} + \tau_{\text{assoc, tot}}$  and there are four cycles of associative transfer in an average bound lifetime, then the average time for  $\tau_{\text{stat}} + \tau_{\text{assoc}} = 3 \text{ ms}/4 = 750 \mu\text{s}$ . The individual values for  $\tau_{\text{stat}}$  and  $\tau_{\text{assoc}}$  are not known. However, based on the requirement that  $\tau_{\text{stat}} \gg \tau_{\text{assoc}}$  (as demanded by the nonspecific DNA binding affinity of hOGG1), and NMR dynamic experiments with hUNG where the associative mobile transfer state was estimated to comprise only ~5% of the bound lifetime(12, 13), we estimate that  $\tau_{\text{assoc}}$  is in the range of ~50  $\mu\text{s}$ . These estimates were used to generate the model in **Figure 6.1**.

### 5.3.5 Calculation of $D_1^{\text{app}}$

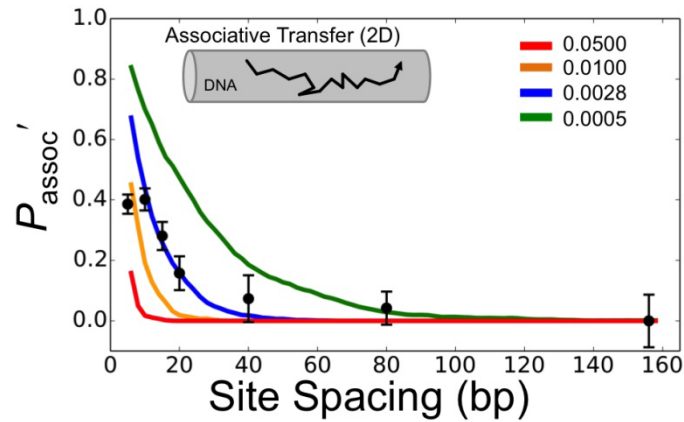
Numerical simulations were used to connect our ensemble measurements with previous single-molecule diffusion measurements of hOGG1. A key parameter is the diffusion coefficient ( $D_1^{\text{app}}$ ) along the longitudinal DNA axis and whether the ensemble measurements and simulations are consistent with the previously measured values of  $D_1^{\text{app}} = 2\text{-}5 \times 10^6 \text{ bp}^2/\text{sec}$  (2). We refer to this diffusion coefficient as ‘apparent’, because in our model it contains cycles of dissociative and associative transfer that are ‘blurred’ together. Two methods for calculation of  $D_1^{\text{app}}$  are presented and both agree well with the previously measured value (2).

**Method I:** The simplest method is presented first and uses the mean displacement for the histogram in **Figure 5.3** (64 bp for hOGG1) and the equilibrium binding lifetime ( $\tau_{\text{bind}} = 3$  ms). The calculated average is then,  $D_1^{\text{app}} = \langle \Delta x \rangle^2 / \tau_{\text{bind}} = (64 \text{ bp})^2 / 0.003 \text{ sec} = 1.3 \times 10^6 \text{ bp}^2/\text{sec}$ .

**Method II:** Here, directly analogous to a single molecule experiment, we calculate the displacement and time for each trajectory, which consists of associative and dissociative steps. The  $D_1^{\text{app}}$  value is calculated from a plot of the mean-squared displacement versus time. First, the ensemble average step time ( $\langle \zeta^{2D} \rangle$ ) for associative transfer was calculated by dividing the average time spent in associative transfers (3 ms) by the average number of 2D steps per trajectory (i.e.  $\langle \zeta^{2D} \rangle = 2.1 \text{ } \mu\text{sec} = 3 \text{ ms} / 1433 \text{ steps}$ ). Then, for each trajectory  $i$  the total time spent in 2D and 3D transfers ( $\tau_{\text{tot}(i)}$ ) was calculated by multiplying the step time ( $\zeta$ ) by the number of 2D or 3D steps ( $\tau_{\text{tot}(i)} = n^{2D, \text{total}} \times \zeta^{2D} + n^{3D, \text{total}} \times \zeta^{3D}$ ). The displacement for each trajectory ( $\Delta x_i$ ) was calculated as the difference between the initial and final  $x$  positions on the DNA cylinder ( $\Delta x_i = x_{\text{start}(i)} - x_{\text{finish}(i)}$ ). Each  $\Delta x_i$  was binned into a histogram according to  $\tau_{\text{tot}(i)}$  with a 1 ms bin size and the mean squared displacement (MSD) was calculated for each bin time ( $\langle \Delta x_i \rangle^2$ ).  $D_1^{\text{app}}$  was determined from the linear fit of  $\langle \Delta x_i \rangle^2$  versus time as is typically done in analyses of single molecule diffusion data (**Figure 5.5d**). From this analysis,  $D_1^{\text{app}} = 3.8 \times 10^6 \text{ bp}^2/\text{sec}$ .

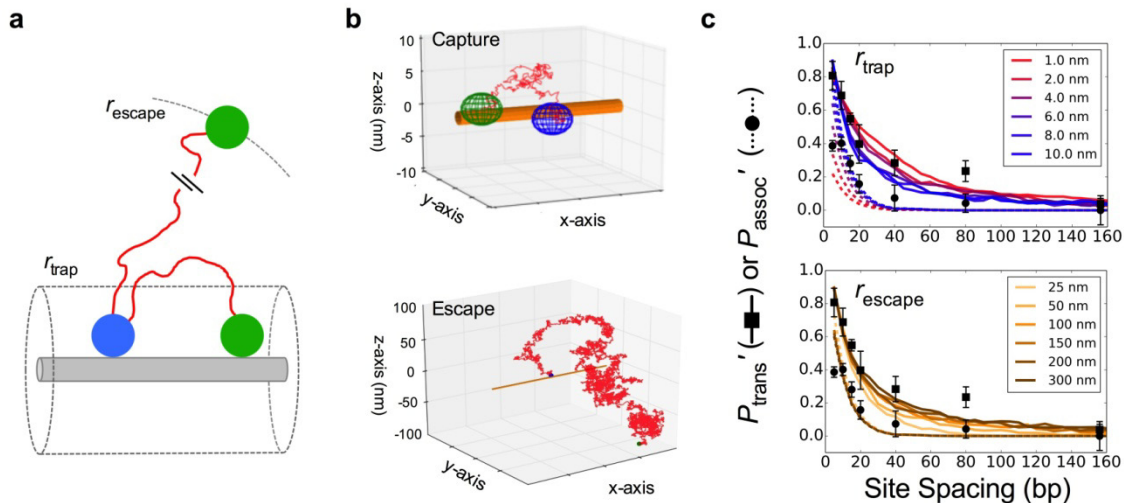
### **5.3.6 Why the ensemble binding lifetime includes all dissociative and associative transfer events**

The majority of hops off of the DNA are exceedingly short relative to the total binding lifetime. For instance even the top 1% of long time-scale hops occur only on the microsecond scale. Even though this may seem like a short time, a hop that takes 10 microseconds ( $<1\%$  of  $\tau_{\text{bind}}$ ) transports the protein on average a linear distance of  $\sim 200$  bp. With the mechanisms described here the two kinetic pathways operate on timescales that differ by at least a thousand fold. Even if each binding event involves 10-100 hops, the time taken in 3D would still be negligible compared to the total binding lifetime. Therefore typical equilibrium techniques that measure binding cannot detect these dissociations. For one example, in the case of anisotropy binding experiments, where the signal comes from the slowed rotational motion of the DNA when bound to protein, microscopic dissociations would have no effect on the population-averaged signal.

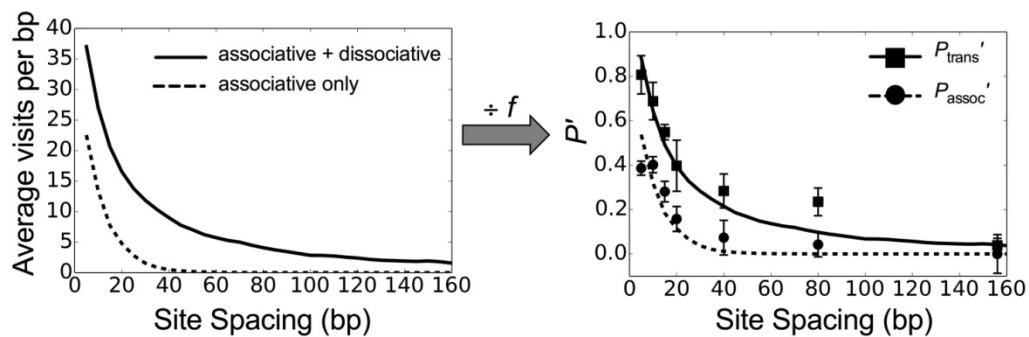


**Figure 5.6: Modeling of associative transfer.** Associative transfer is modeled by a 2D random walk on the DNA surface using 0.34 nm steps (note that the step length is arbitrary; see above text). At each step in the random walk the protein has a probability ( $p_{\text{exit}}$ ) of exiting the walk DNA and entering the dissociative pathway [ $p_{\text{exit}} = k_{\text{exit}}/(k_{\text{step},2\text{D}} + k_{\text{exit}})$ ]. The optimal value of  $p_{\text{exit}} = 0.0028$  was determined empirically by performing 1000 simulations. Four simulations are shown where  $p_{\text{exit}}$  was varied in the range 0.05 – 0.005.





**Figure 5.7: Effects of varying the  $r_{trap}$  and  $r_{escape}$  parameters in the Monte Carlo simulations for hOGG1.** (a) Dissociative 3D transfers are modeled as a 3D random walk in solution. The dissociated protein is recaptured when the center-to-center distance between the DNA and protein is equal to the sum of their radii. To eliminate the effect of very short transfers within the atmosphere of bound ions, only dissociations that have escaped a distance greater than  $r_{trap}$  are used. Escape to bulk occurs when the protein exceeds a distance of  $r_{escape}$  from the initial starting point (see text). (b) An example of a single dissociative transfer trajectory involving diffusion outside of  $r_{trap}$  followed by return to the DNA chain and an example of a dissociative trajectory that leads to escape. (c) Effects of varying  $r_{trap}$  and  $r_{escape}$  on fitting the experimental data. (Top panel)  $r_{escape}$  was set to equal 150 nm and values of  $r_{trap}$  were varied in the range 0.5 to 10 nm. Values from 4-6 nm were found to fit optimally. (Bottom panel)  $r_{trap}$  was set to 5 nm and values of  $r_{escape}$  were varied in the range 25 to 300 nm. Values greater than 100 nm had no effect on the fitted curves indicating that at this distance the probability of hOGG1 returning to the DNA approaches zero. The procedure used to fit the simulated curves to the experimental data is described in **Figure 5.8** and **Section 5.3 Methods**.



**Figure 5.8: Procedure for fitting the raw output from the MC simulations to the experimental data.** For 10,000 total trajectories, the average number of times the protein visited a particular 1 bp slice along the DNA axis is shown on the left. To fit the experimental data, the curves on the left were simply divided by a single normalization factor ( $f$ ) that was determined by least-squares minimization. In essence, this procedure tested whether the relative shapes and vertical displacements of the two simulation curves simultaneously fit the experimental data.

## 5.4 References

1. Hipfel von PH, Berg OG. 1989. Facilitated target location in biological systems. *J. Biol. Chem.* 264:675–678.
2. Blainey PC, van Oijen AM, Banerjee A, Verdine GL, Xie XS. 2006. A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA. *Proc. Natl. Acad. Sci. U.S.A.* 103:5752–5757.
3. Blainey PC, Luo G, Kou SC, Mangel WF, Verdine GL, Bagchi B, Xie XS. 2009. Nonspecifically bound proteins spin while diffusing along DNA. *Nat Struct Mol Biol* 16:1224–1229.
4. Schurr JM. 1979. The one-dimensional diffusion coefficient of proteins absorbed on DNA: Hydrodynamic considerations. *Biophysical chemistry* 9:413–414.
5. Das RK, Kolomeisky AB. 2010. Facilitated search of proteins on DNA: correlations are important. *Phys Chem Chem Phys* 12:2999–3004.
6. Kampmann M. 2004. Obstacle bypass in protein motion along DNA by two-dimensional rather than one-dimensional sliding. *J. Biol. Chem.* 279:38715–38720.
7. DeSantis MC, Li J-L, Wang YM. 2011. Protein sliding and hopping kinetics on DNA. *Phys. Rev. E* 83:021907.
8. Wang YM, Austin RH. 2011. Single-Molecule Imaging of LacI Diffusing Along Nonspecific DNA, pp. 9–37. *In* Williams, MC, Maher, JL (eds.), *Biophysics of DNA-Protein Interactions From Single Molecules to Biological Systems*. Springer.

9. Tafvizi A, Huang F, Fersht AR, Mirny LA, van Oijen AM. 2011. A single-molecule characterization of p53 search on DNA. *Proc. Natl. Acad. Sci. U.S.A.* 108:563–568.
10. Leith JS, Tafvizi A, Huang F, Uspal WE, Doyle PS, Fersht AR, Mirny LA, van Oijen AM. 2012. Sequence-dependent sliding kinetics of p53. *Proc. Natl. Acad. Sci. U.S.A.* 109:16552–16557.
11. Bai Y, Greenfeld M, Travers KJ, Chu VB, Lipfert J, Doniach S, Herschlag D. 2007. Quantitative and comprehensive decomposition of the ion atmosphere around nucleic acids. *J. Am. Chem. Soc.* 129:14981–14988.
12. Friedman JI, Majumdar A, Stivers JT. 2009. Nontarget DNA binding shapes the dynamic landscape for enzymatic recognition of DNA damage. *Nucleic Acids Res.* 37:3493–3500.
13. Schonhoft JD, Kosowicz JG, Stivers JT. 2013. DNA translocation by human uracil DNA glycosylase: role of DNA phosphate charge. *Biochemistry* 52:2526–2535.

## Chapter 6: Concluding Remarks

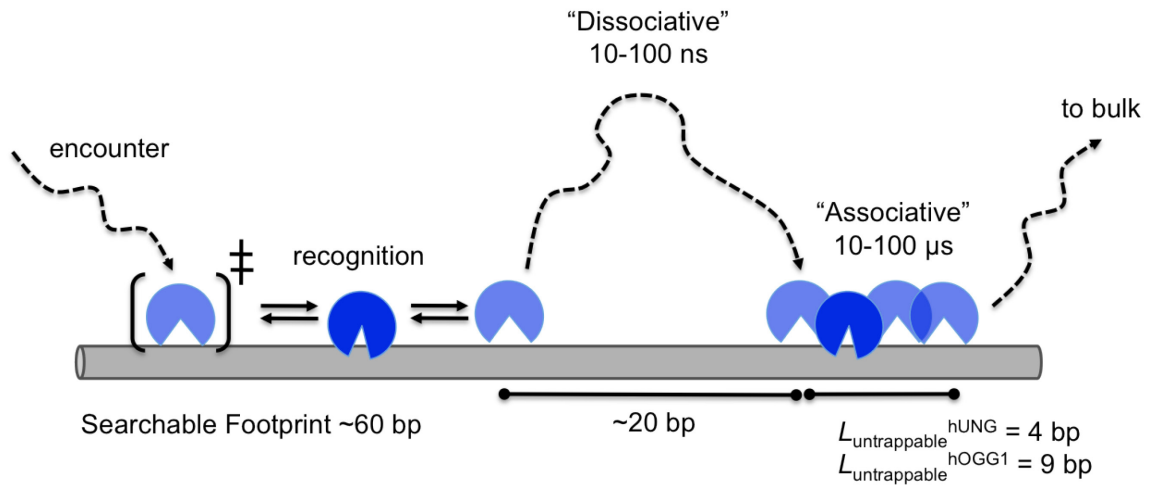
The data overwhelmingly support a model in which glycosylases search for lesions on the DNA by using a combination of 3-dimensional dissociative transfer combined with short range 1-dimensional associative transfer (**Figure 6.1**). Most surprisingly, the search mechanism for hOGG1 and hUNG are near identical. Given the early evolutionary divergence, the search mechanism for these proteins seems to have evolved early and independently of the genome size, and suggests that the actual search process may not be the limiting factor and instead some other requirement may be the real reason behind the evolution of these mechanisms. Possibly, the cycles of short range sliding do not necessarily function to increase genome coverage, but instead function to allow sufficient sampling of the DNA bases so that kinetic pre-selection of the correct lesion base can be efficient. Other factors such as compartmentalization of these enzymes near sites of common damage, increasing the copy number and condensing the nucleus itself into condensed and open chromatin could instead serve to limit the demands of locating a rare base. It could be argued that search may be a limiting factor for low copy proteins such as transcription factors. But in this case the search problem could still be overcome by the above factors and in the case of prokaryotes, the genomes are often organized in such a way that transcription factors are encoded nearby to their target sequence. Therefore I propose that the *search problem* is really more of a *recognition problem*, and these mechanisms instead function to allow highly selective binding to the target site over non-specific DNA.

In this thesis, a major advance has been in defining experimentally the process of hopping. Using the methods developed in Chapter 2 and the simulations in Chapter 5, a model where the protein undergoes hops on the nanosecond timescale and nanometer distance scale (on average), has provided unprecedented quantitative experimental insight into this process.

How does a protein in actuality 'slide' on DNA? This process may not be so different than hopping. For hUNG the protein is largely absent of charge within the DNA binding interface, and experiments in Chapter 3 show that ablating charge on the DNA between uracil sites has no effect on transfer. Combined with previous NMR studies, where hUNG was shown to have millisecond timescale dynamics upon binding to nonspecific DNA, a model was proposed where the protein undergoes a conformational change between an open and closed state, and in transferring between sites, the protein opens and releases away from the DNA and diffuses to the next site. Thus, sliding is not necessarily mediated by any particular interaction, such as electrostatics, but rather diffusion itself, and in transferring between sites the protein enters a transition state highly similar to that of dissociation. Thus, the data most supports a model where sliding and hopping are not necessarily two separate processes, but more of a continuum.

Also, several unique observations were made in Chapter 4, where hUNG was shown to slide longer on single stranded DNA and to have a directional preference in the presence of abasic sites. We explained this observation in terms of an extended binding interface of the enzyme. This 'change' of search

mode most certainly has an effect on processes where hUNG must remove uracils from single stranded DNA and within clusters of uracil (*i.e.* somatic hypermutation), although the exact biological outcome has yet to be defined. Finally, I believe there still exists a major gap in extending current mechanistic understanding to the context of the cellular environment which differs on nearly every level to that of the test tube, and it will be interesting to see how these types of mechanistic studies translate to this environment.



**Figure 6.1. Quantitative Model for Facilitated DNA search of hUNG and hOGG1.** Based on the simulations and data the search process begins by the diffusion limited encounter of the enzyme and DNA in which the protein then finds the specific site by a 1D/3D search mechanism. In 1D, the protein uses a conformational change to switch between search recognition modes (see Chapter 6 and Chapter 3), where the search mode can be viewed as a loosely bound and resembling the transition state for dissociation. In 3D the protein microscopically dissociates and reassociates with the DNA chain. The parameters determined from the simulations and ensemble data are shown for both proteins.



## Chapter 7: Appendix

## 7.1 Preliminary data for hUNG disordered tail project

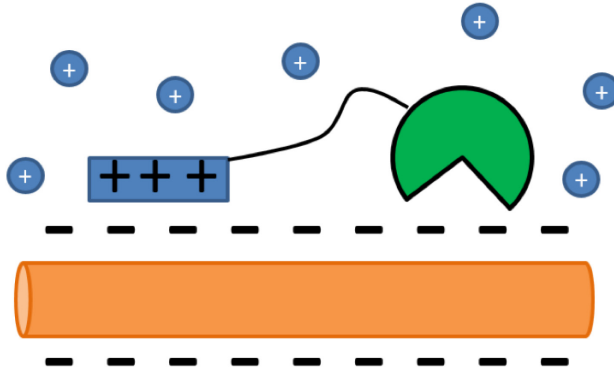
The goal of the project is to enhance nonspecific DNA binding of hUNG by attaching disordered peptide tails from homeodomain transcription factors, in order to directly test the electrostatic hypothesis for DNA translocation.

### Disordered Tails from Homeodomain TF

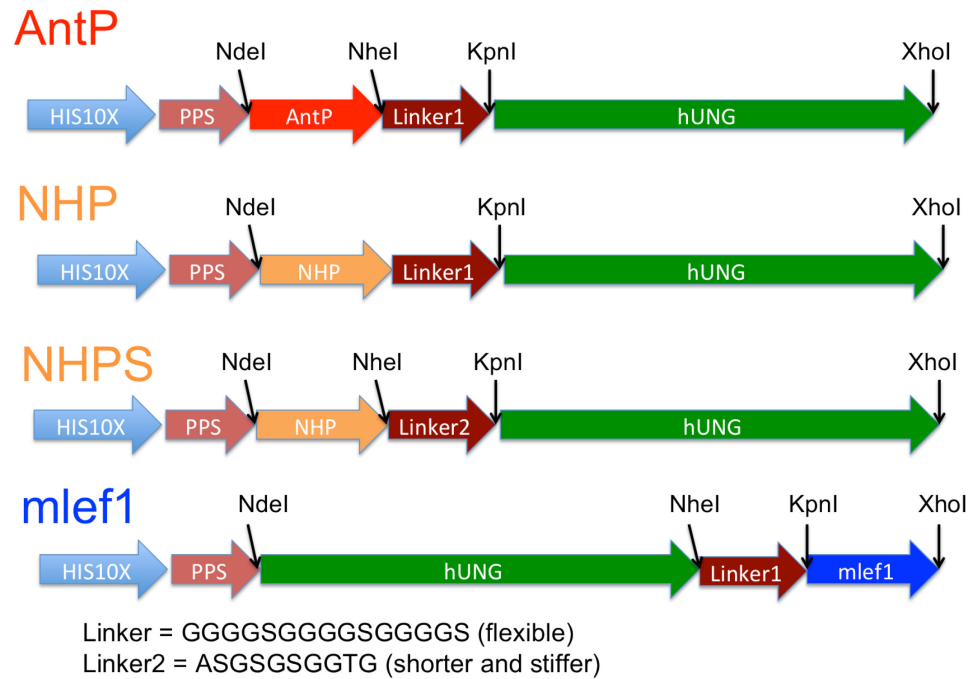
AntP - MRKRGRQTYT (N-tail)

NHP - MVTPREPKKRTTRKKKD  
(N-tail)

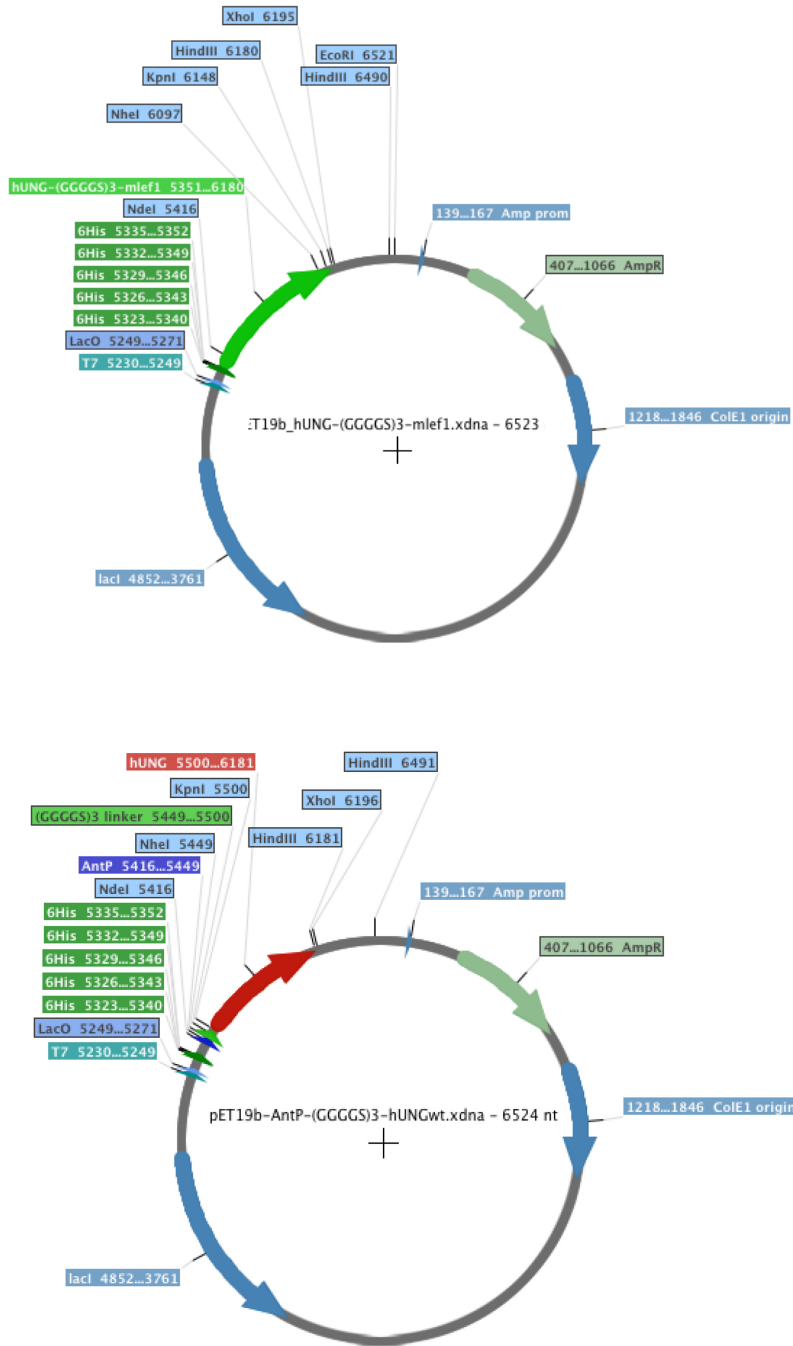
mlef1 - KKKRKRER (C-tail)



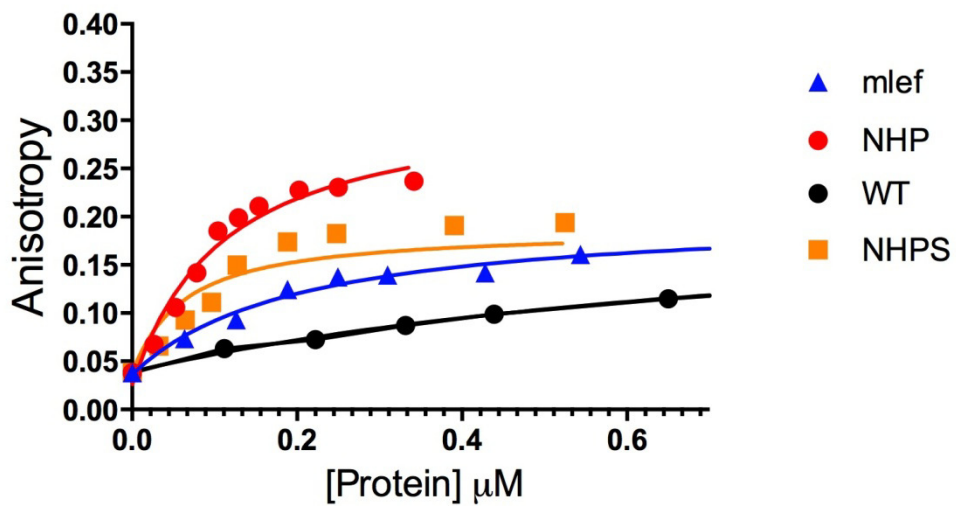
**Figure 7.1:** Basic design of hUNG disordered tail constructs.



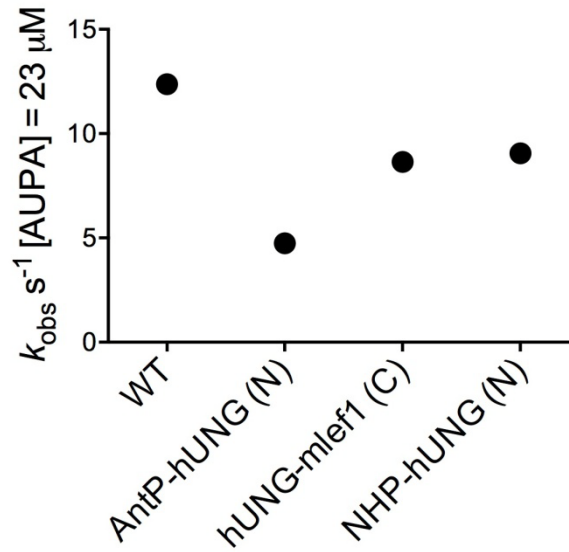
**Figure 7.2. Modular expression plasmid design.** Plasmids were constructed from pET19b. Restriction sites allow facile change of linkers and disordered tails. These four plasmids have been cloned in the lab so far.



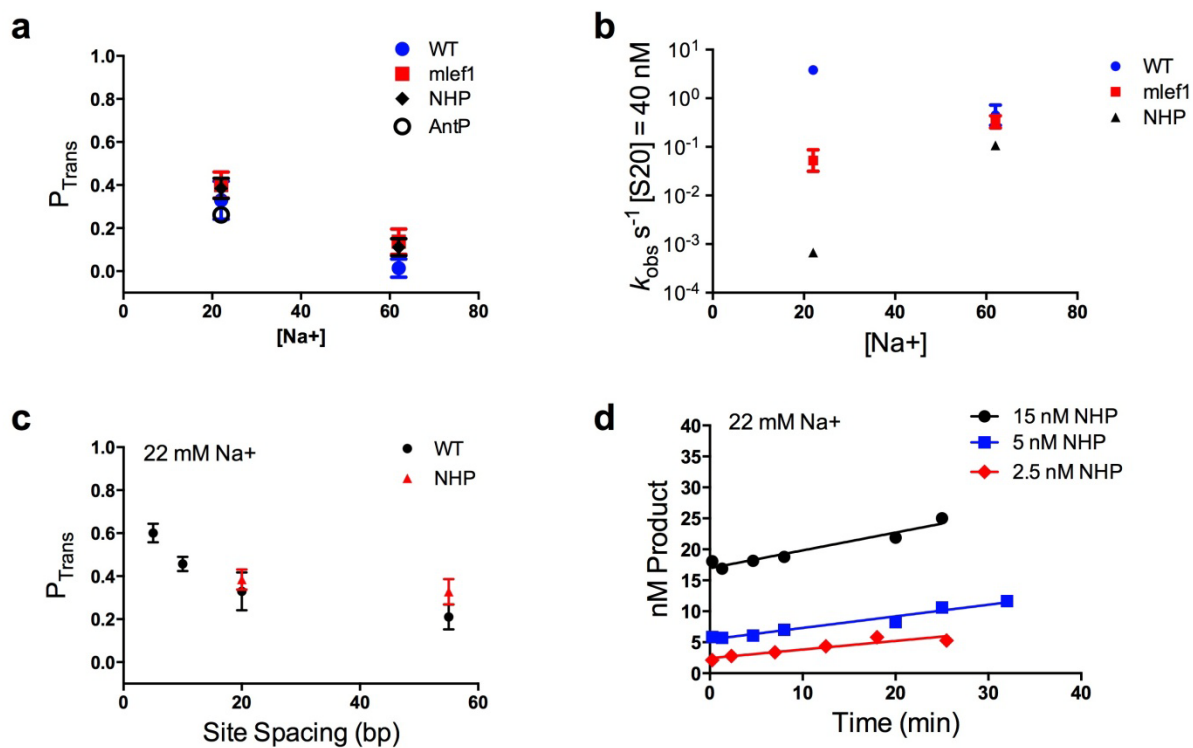
**Figure 7.3. pET19b plasmid map for hUNG-mlef1 (C-terminus) and AntP-hUNG (N-terminus) expression plasmids.**



**Figure 7.4. Nonspecific DNA binding of hUNG-tail proteins to a 30mer fluorescein labeled nonspecific DNA duplex.** The DNA concentration was 10 nM and the fits are to the quadratic binding equation.



**Figure 7.5. Multiple turnover AUPA reaction rate for hUNG tail variants.** Minimal change in rate was observed for a single concentration of AUPA ( $[\text{AUPA}] = 23 \mu\text{M}$ ).

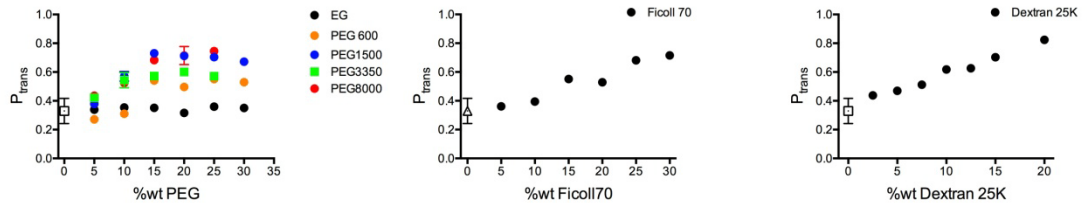


**Figure 7.6. Site transfer measurements for hUNG tail variants.** (a) Processivity measurements for S20 (20 bp site spacing) at 22 and 62 mM Na<sup>+</sup>. (b) Rates of product formation. (c) Spacing dependence for NHP. (d) Product formation with time for NHP-hUNG showing a burst equal to the enzyme concentration.

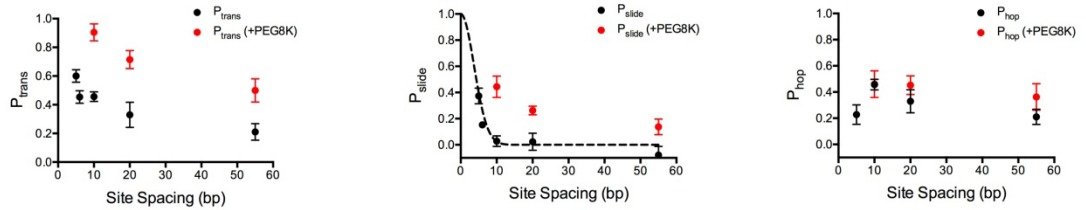
## 7.2 Preliminary Data for Molecular Crowding Experiments

The goal of this project is probe the affects of molecular crowding on DNA site transfer. To mimic cellular crowding conditions synthetic polymers are used such as Ficoll 70, Dextran 25K and polyethylene glycol.

### Crowder Screen

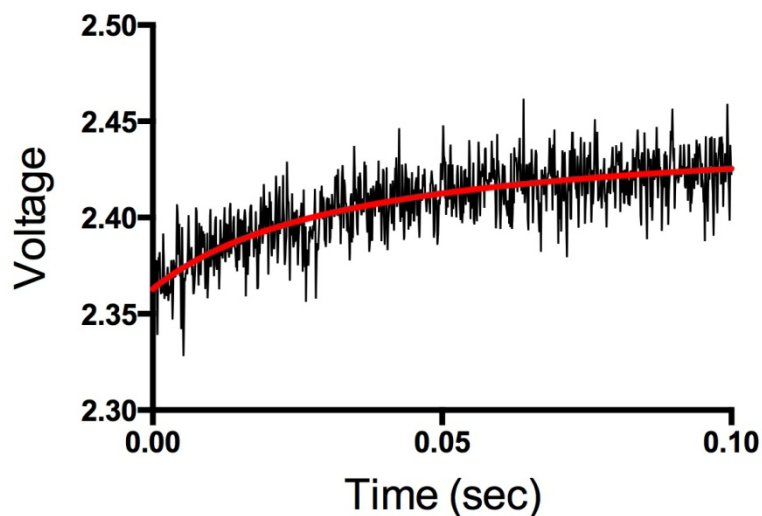


### Site Spacing Dependence (with and without 20% PEG 8K)



**Figure 7.7. Effects of molecular crowding by polymers on hUNG site transfer.** (Top Panel) hUNG processivity with the addition of polyethylene glycols, Ficoll 70 and Dextran 25K. (Bottom Panel) Effects of 20% PEG 8000 on  $P_{trans}$ ,  $P_{slide}$  and  $P_{hop}$ .





Conditions	[DNA U $\beta$ ] <sub>final</sub>	[hUNG] <sub>final</sub>	Signal $\Delta$ (V)	Rate (M <sup>-1</sup> sec <sup>-1</sup> )
1x buffer	100 nM	100 nM	0.0941	3.49E+09
1x buffer	200 nM	200 nM	0.1833	2.36E+09
1x buffer + 20% PEG8K	200 nM	200 nM	0.1992	2.28E+08
1x buffer + 20% PEG8K	100 nM	100 nM	0.08418	2.86E+08

**Figure 7.8. Effects of the hUNG rate of association with a specific site in the presence or absence of 20% PEG 8000.** The DNA sequence used was GCGGCCAAPU $\beta$ AAAAGCGCGC, where U $\beta$  is a no-hydrolyzable uracil analog. In the complement a G was placed across from the U $\beta$  as G results in a greater signal change compared to A. The fluorescent signal change arises from 2-aminopurine (P) fluorescence's. (Top) Representative curve from a stopped-flow experiment. The average of at least 10 transients was used. Association rates were calculated by fitting the data to a model for second order association where [hUNG] = [DNA] (Bottom Panel).

## **Curriculum Vitae, Joseph D. Schonhoft**

---

7 E Eager, Apt 2R  
Baltimore, MD 21202

330-685-6115  
[jschonh1@jhu.edu](mailto:jschonh1@jhu.edu)  
Citizenship: U.S.A.

### **Educational History**

---

**Ph.D. 2014 Program in Molecular Biophysics Johns Hopkins SOM**  
**MS/BS\* 2009 Chemistry & Biochemistry Kent State University**  
*\*combined bachelors/masters program*

### **Research Experience**

---

2009-current **Johns Hopkins University SOM, Department of Pharmacology and Molecular Sciences**  
Ph.D. thesis, Advisor: James T. Stivers  
*Thesis Title: Dissecting the search pathway for DNA glycosylases.*

2007-2009 **Kent State University, Department of Chemistry**  
MS/BS Thesis research, Advisor: Soumitra Basu  
*Thesis Title: Biophysical and biochemical studies of G-Quadruplex nucleic acid structures.*

2006-2007 **NSF – REU Program and Undergraduate Research, Kent State University**  
Advisor: Soumitra Basu, Department of Chemistry  
*Topic: Proteomic analysis of serotonin deficient mice brain tissue*

2004 **SURP Undergraduate Research Program, Case Western Reserve University**  
Advisor: John Mieyal, Department of Pharmacology SOM  
*Topic: Reactive oxygen species involved in cell signaling*

### **Peer Reviewed Publications**

---

- 8) **Schonhoft, J.D.**, and Stivers, J.T. DNA Translocation by Human Uracil DNA Glycosylase: The Case of ssDNA and Clustered Uracils. *Biochemistry* (2013), 52 (15), 2536–2544.
- 7) **Schonhoft, J.D.**, Kocsowiz, J.G., and Stivers, J.T. DNA Translocation by Human Uracil DNA Glycosylase: The Role of the Phosphate Backbone. *Biochemistry* (2013), 52 (15), 2526–2535.
- 6) **Schonhoft, J.D.**, and Stivers, J.T. Timing Facilitated Site Transfer of an Enzyme on DNA. *Nature Chem. Biol.* (2012) 8 (2), 205-210.
- 5) Morris, M., Negishi, Y., Pazsint, C., **Schonhoft, J. D.** and Basu, S. RNA G-Quadruplex Is Essential for Cap-Independent Translation Initiation in Human VEGF IRES. *J. Am. Chem. Soc.* (2010) 132 (50), 17831–17839.

- 4) Dhakal, S., **Schonhoft, J.D.**, Koirala, D., Yu, Z., Basu, S., Mao, H. Coexistence of an ILPR i-motif and a partially folded structure with comparable mechanical stability revealed at the single-molecule level. **J. Am. Chem. Soc.** (2010) 132 (26), 8991-8997.
- 3) **Schonhoft, J. D.**, Das, A., Achamyeleh, F., Samdani, S., Mao, H., and Basu, S. ILPR repeats adopt diverse G-quadruplex conformations that determine insulin binding. **Biopolymers** (2010) 93, 21-31.
- 2) **Schonhoft, J. D.**, Bajracharya, R., Dhakal, S., Yu, Z., Mao, H., and Basu, S. Direct experimental evidence for quadruplex-quadruplex interaction within the human ILPR. **Nucleic Acids Res.** (2009) 37, 3310-3320.
- 1) Yu, Z., **Schonhoft, J. D.**, Dhakal, S., Bajracharya, R., Hegde, R., Basu, S., and Mao, H. ILPR G-Quadruplexes Formed in Seconds Demonstrate High Mechanical Stabilities. **J. Am. Chem. Soc.** (2009) 131, 1876-1882.

## Awards & Fellowships

---

2013	1/75 U.S. graduate students sponsored to attend the 63 <sup>rd</sup> Lindau Meeting of Nobel Laureates
2012-2014	American Heart Association pre-doctoral fellowship
2009	Nomination for the Bush Prize in teaching, Kent State University
2009	Todd Prize for Research, Kent State University
2007	NSF-REU Travel Award, used to attend 2007 national ACS conference
2005	Lillian A. Korenke Scholarship Nomination, Kent State University
2003-2005	Student-Athlete Mid-American Conference Commissioner's Award
2003	CRC Press Freshman Award
2002-2006	Knauss Scholarship, Kent State University
2002-2006	Trustee Scholarship, Kent State University

## Conferences & Presentations

---

2013	63 <sup>rd</sup> Lindau Meeting of Nobel Laureates in Chemistry
2011	Gordon Conference – 'Enzymes, Coenzymes and Metabolic Pathways' Title: <i>Timing Facilitated Site Transfer of an Enzyme on DNA</i>
2009	2 <sup>nd</sup> International Meeting on Quadruplex DNA Title: <i>Putative Loop Complementarity Affects Quadruplex Conformation and Insulin Binding Ability in G-Quadruplexes Formed by ILPR Repeats</i>
2007	233 <sup>rd</sup> American Chemical Society National Meeting Title: <i>Proteomic Analysis of the Mid/Hindbrain of Pet-1 Knockout Mice.</i>