DNA-mediated Charge Transport in a Biological Context: Cooperation Among Metalloproteins to Find Lesions in the Genome

Thesis by

Pamela Alisa Sontz

In Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy in Chemistry

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2012

(Defended May 7, 2012)

ACKNOWLEDGEMENTS

I consider it a great privilege to have the opportunity to thank the people who have inspired and guided me on my journey. As John Quincy Adams once said, “If your actions inspire others to dream more, learn more, do more and become more, you are a leader.” These characteristics can be found in all of the people who have impacted my life; they are all leaders who I have been fortunate enough to know. My advisor, Professor Jacqueline K. Barton, has been a role model both in science and in life. She has a special way of tailoring her interactions to reach and inspire each person individually. Jackie has always taken the time to listen and has been extremely supportive and encouraging each step along the way. Her enthusiasm to pursue challenging topics and discuss intriguing questions has inspired me, and I have learned many lessons from her that I will never forget. She truly cares about people, and I have a great amount of respect and admiration for her. Jackie stands out not only as an incredible scientist, but also as a one-of-a kind mentor.

I am also extremely thankful for the interactions I have had with my thesis committee. The chair of my committee, Professor Harry Gray, has taught me so much along the way. Harry always had his door open to discuss science and life, and I have deeply appreciated all of the time he set aside. He is always smiling and interacting with people, and he is also one amazing tennis player! From the moment I entered Caltech, Harry showed me that he cared and always included me as though I was part of his own scientific family. Professor Dave Tirrell provided much input and was always willing to talk about ideas and my future aspirations. Dave is extremely humble and has contributed a great deal to my scientific career. Professor Bil Clemons has added much insight to help me improve biochemical experiments. Bil also taught me to really think about the road ahead…not just concerning science but also the big career picture as well. Maureen Renta has become a great friend during my time in the lab. She looks out for our entire group, gives us advice, and teaches us to look out for each other. Not to mention, the endless hours she spends helping us each day.

When I first arrived at Caltech, I did not know anyone, but that quickly changed as the Barton group members warmly reached out their hands to welcome me to the lab. Specifically, Dr. Amie Boal immediately took me under her wing and introduced me to the tools I would use to study important processes during my graduate career. Amie is extremely patient and thorough, and I am very thankful to have had the opportunity to learn from her during my beginning years in the lab and to have her as a steadfast friend and role model now. Also, Dr. Joey Genereux was always positive and supportive along the way. I learned a great deal from our discussions about science. Alison Parisian and Keshia Dykes, talented undergraduate students, contributed a great deal to the atomic force microscopy and *in vivo* studies, respectively. They made the summers a lot of fun and also even more productive! Dr. Paul Lee was a true friend inside and outside lab, and he provided much insight on experiments and late night dance parties to keep spirits high and laughter in the air. Since I began in the lab, Dr. Eric Olmon has sat to my left…and I have learned a lot from my desk neighbor through collaborations and stimulating conversations. Tim Mui and I have been on a number of adventures not just observing XPD protein but also in experiencing Pasadena, and he has continued to be a wonderful friend during my time at Caltech. Dr. Donato Ceres trained me on the atomic force microscope when I first arrived, and Dr. Bruce Brunschwig and Craig Wiggenhorn have since been very supportive whenever the instrument misbehaves. Hang, Natalie, and I have been true friends and partners in crime since we began our graduate work and have developed a friendship that will continue well into the future-the invincible trio rolls on.

"Friends are the pillars on your porch. Sometimes they hold you up, sometimes they lean on you, and sometimes it's just enough to know they are standing by." I am very grateful to my friends for being there through thick in thin. Friends from the Barton group in the past (Marisa, Brian, Ben, Wendy, Omar, Cindy, Mi Hee, Eric S.) and the current lab members (Liping, Jumpei, Shuo, Anna M., Curtis, Catrina, Anna A., Katie, Mike, Alexis, Ariel, Alyson, Helen, Phil, and Ted) have been like family. We have had so many fun moments: from camping trips to Clippers games to goofy

times in Beverly Hills and beyond. It has been an amazing experience doing science alongside people who are passionate about what they do, and through this, we will always be connected to one another. Thanks to Professor Jeff Zaleski, my undergraduate research advisor at Indiana University, for encouraging me to proceed to graduate school and continue following my passion of chemistry. My friends who I met through frisbee, hip-hop, and those in other chemistry research labs have meant so much and contributed a great amount during my time here. Silva and Agnes have become special friends within the department, and I am extremely lucky to have had the interactions with them that I did. Also, I must mention the teachers and leaders who impacted my life prior to my time at Caltech. I have always felt blessed to have teachers who not only imparted so much knowledge but also were deeply involved in shaping my life-from elementary school to high school-a huge thank you to them. To my friends who have been like siblings over the years: (Brock, Mitzi, Sarah C., Alanna, Lindsay, Carolyn, Dawn, Debra, Michelle, Harmony, Sarah T., Nyssa, and Heather), who have been there without hesitation, taught me so much, and were there to make many happy and exciting memories: thank you.

Each day my family is my source of support and inspiration. My Bubbie writes me a letter each month and is always excited to catch up on the phone or spend time with me when I visit. Though my Zadie passed away many years ago, I still recall how excited he was to talk with me, and how he smiled when we played with electric trains and cars. My Nana and Grandpa and I have a special connection-they are constantly there to encourage and guide me with their love and deep care in all that I do. Over the years, I have had the opportunity to see my grandfather, a Holocaust survivor, speak and teach the next generations to not stand by and be indifferent when injustices occur to people of any race, religion, or ethnicity-he affects and influences the lives of everyone he speaks to. My Nana is so compassionate and sincere-she is like a second mother to me, and she exudes her beautiful personality to everyone around her. I am so very proud of my grandparents and to have them in life means more than they will ever know. My dad and Jen have always been involved and

genuinely interested in my pursuits, and they always are able to provide a positive outlook. When I was younger, my dad would quiz me on my spelling words at night, even when he was tired from work, but he made it clear that he wanted me to succeed in whatever I put my mind to. I thank him for all of his love and support. My mom and Daniel have enthusiastically been there to guide me during the defining moments in my life. My best friend and hero, my mom is so selfless, kindhearted, intelligent, and is a true leader. Even working night and day in the hospital, she made time not only to talk with me and listen to me but also to find ways to get involved. She has guided me throughout my life and has always believed in me without hesitation. Just by looking at me, she tells me through emotion more than she could ever say with words. I thank her for not only being the mother that she has been but also for being a role model who I will always look up to.

I am very thankful for the time I have spent at Caltech and grateful for the people I met, the experiences I shared, and the scientific projects I had the opportunity to contribute to. I am excited to take the life lessons and scientific tools I have learned thus far and apply them to all of my future endeavors.

ABSTRACT

Not only does the DNA sequence encode proteins and drive cellular processes but the base-pair stack also exhibits a unique and exquisite property: it facilitates charge transport on ultrafast time scales over long distances. A molecular wire, DNA conducts charge with very shallow distance dependence, yet mismatches and lesions significantly attenuate this process. It is this characteristic that has inspired the research described herein: how might nature exploit DNA CT and perhaps more importantly, how can we gauge which biological pathways are governed by DNA-mediated CT?

Here, we have focused on understanding the role of DNA CT in the search for lesions by repair proteins. All of the proteins we have examined contain a redox-active [4Fe4S] cluster, thus these metalloproteins can "tap into" DNA-mediated CT. We have proposed a model where repair proteins send and receive signal over long distances via the DNA base pair stack to efficiently scan the genome for damage. As glycosylases bind to DNA, they are activated toward oxidation, triggering DNA-mediated CT to a distally bound repair protein that is then reduced, loses its affinity for DNA, and dissociates. DNA CT is remarkably sensitive to intervening lesions and mismatches; thus the signal is only transmitted between proteins if the DNA is intact. Based on this model, repair proteins will reduce the population of sites to search through a relocalization of the proteins in the vicinity of lesions that inhibit CT. To test this model, we have employed single-molecule atomic force microscopy (AFM) to investigate Endonuclease III (EndoIII), a base excision repair (BER) protein that contains a redox-active [4Fe4S] cluster. Notably, there is no simple protocol that has been established to generate long DNA strands required for AFM studies, > 3000 bp, containing site-specific modifications. To circumvent this, we have designed modular vectors that we synthesize by PCR with primers containing a 2′-*O*-methyl base to generate ~ 2000 bp duplexes with 14 base single-strand overhangs. The vectors are then pieced together to yield a site-specific lesion

in the middle of a ~ 4000 bp DNA duplex. With these DNA samples, we find a redistribution of EndoIII onto the long DNA strands containing a C:A mismatch, which is not a specific substrate of EndoIII but inhibits CT. When both long and short strands are matched, accumulation of EndoIII on longer strands is not observed. These results demonstrate the link between DNA CT and the ability of a BER protein that contains a redox-active [4Fe4S] cluster to find damage in DNA. Importantly, we recognize that interrogation times for BER proteins depend on DNA CT over long distances as well as the percentage of proteins oxidized. AFM measurements taken as a function of oxidation of proteins bound to DNA, using peroxide as an oxidant, reveal an additional increase in the ratio of EndoIII bound to strands containing a mismatch.

We then identified residues in EndoIII, specifically aromatic; that we thought may be involved in the CT pathway between the [4Fe4S] cluster and DNA. Subsequently, we have prepared and examined a variety of EndoIII mutants using electrochemical, AFM, and *in vivo* methods. The residues targeted for mutation are either conserved aromatic amino acids near the [4Fe4S] cluster or are located close to the DNA binding site. All of the EndoIII mutants display glycolytic activity within error of wild-type protein. Y75A EndoIII, protein with a single mutation making it deficient in DNA-mediated CT, does not redistribute onto strands containing a C:A mismatch. Y75F and W178A EndoIII, proteins shown to perform charge transfer within error or even more efficiently than wild-type, do show redistribution onto the mismatched DNA strands. Thus, consistently, we find a correlation between the ability of mutants to carry out DNA/protein CT and their ability to relocate near a single base mismatch.

As another test of the search model, we employed genetic methods to investigate the interactions of proteins that may cooperate with each other *in vivo* to find damage. MutY, a BER protein also containing a redox-active [4Fe4S] cluster, displays a DNA-bound potential similar to that of

EndoIII. We have performed assays to monitor MutY, a protein that removes adenine from 8-oxo-G:A mismatches, in *E. coli* and asked whether changes in EndoIII might affect MutY activity. We find that inactivating EndoIII *in vivo* leads to a decrease in MutY activity, evidenced by an increase in unrepaired MutY lesions. Additionally, we introduced oxidative stress into *in vivo* experiments to further increase the number of lesions that are specific substrates for MutY. By increasing the overall signal of MutY activity, we ultimately reduce the standard error of the measurements. Previously, we found that a mutant of EndoIII that is glycolytically inactive yet contains an intact [4Fe4S] cluster restores MutY activity; however, Y82A, a mutant of EndoIII that is deficient in its ability to perform CT, cannot aid MutY in fixing damage. Notably, mutations at the analogous 82 position in the human homologue of MutY, *MutYH*, have been shown to lead to a predisposal toward colorectal cancer. Significantly, AFM studies with Y82A provide further evidence that this mutant cannot perform DNA-mediated CT, as the protein does not redistribute in the vicinity of a lesion.

Various DNA-binding proteins, such as those involved in repair and other pathways that maintain the integrity of DNA, have already been found to contain FeS domains and other redox cofactors. We are discovering proteins from alternate repair pathways that may also utilize DNA CT to find damage in the genome. XPD, a 5′-3′ helicase involved in nucleotide excision repair from archaea, contains a conserved [4Fe4S] cluster and exhibits a DNA-bound redox potential that indicates it is able to carry out DNA CT. In AFM studies, we observe the redistribution of XPD onto long DNA strands containing a single base mismatch. Importantly, in XPD from archaea, L325V aligns with mutated residues in human XPD and *Schizosaccharomyces pombe* Rad 15 that are associated with diseases including Trichothiodystrophy and Xeroderma Pigmentosum. On the basis of electrochemical and AFM experiments, we have demonstrated this mutant is not only defective in carrying out DNA CT but also does not redistribute to find lesions in DNA strands.

Further, we recognize that we can apply AFM techniques to probe DNA-mediated signaling *between* two distinct repair proteins bound to DNA. When XPD and EndoIII are mixed together at equal concentrations, they coordinate in relocalizing onto the mismatched strand. However, when a CT-deficient mutant of either repair protein (Y82A or L325V) is combined with the CT-proficient repair partner, no relocalization occurs. When wild-type protein is titrated into the DNA/mutant protein solution at increasing concentrations, it substitutes for the protein that could not perform CT; thereby restoring redistribution. These data not only indicate a general link between the ability of a repair protein to carry out DNA CT and its propensity to redistribute onto DNA strands near lesions, but also provide evidence for coordinated DNA CT between repair proteins in their search for damage in the genome.

Extending the methodology to examine related systems, we are now employing AFM to "visualize" the helicase activity of XPD. Electrochemical studies in our lab have revealed that XPD performs DNA CT efficiently in an ATP-dependent manner. Thus, we have added ATP to DNA/protein solutions, initiating XPD helicase activity, in an effort to understand whether there is a change in redistribution that reflects the ability of XPD to perform DNA CT. Interestingly, at high concentrations of ATP, we see an increase in proteins that are bound to the short strands of DNA. Though we have introduced complementary strands to block the short-strand overhangs, over time, rather than determining the redistribution of XPD, we have instead directly observed the helicase activity of the protein. Now, we are using this knowledge to our advantage in developing new protocols to explore links between the biochemical activity of XPD and its ability to perform DNA-mediated CT.

We have also probed DNA CT using time-resolved IR (TRIR) spectroscopy. In the first step of the search model, a reduced metalloprotein binds DNA, and consequently, gets oxidized. Since the

guanine base is the most susceptible to oxidation, we have considered that guanine oxidation may indeed trigger the metalloprotein to bind DNA, initiating the search process. To investigate this, we have developed a metal-based photooxidant: $[Re(CO)_3(dppz)(py'\text{-}OR)]^+$ (dppz=dipyrido[3,2-*a*:2′,3′-*c*]phenazine; py′-OR=4-functionalized pyridine] that can be tethered to DNA duplexes via the pyridine moiety to examine the photophysical and electronic properties of DNA CT. Importantly, the (C≡O) stretching vibrations allow us to utilize TRIR to monitor DNA CT on pico and microsecond time scales. Charge injection, the motions of DNA bases, and back electron transfer rates dictate the efficiency of CT and, thereby, the response we detect. In conjunction with these studies, we have employed biochemical methods to directly observe long-range guanine oxidation of DNA by the excited state complex.

The results presented here establish a link between the ability of a metalloprotein to perform CT and its ability to find lesions in the genome. Atomic force microscopy studies have provided us with snapshots of proteins bound to DNA containing a single mismatch, unveiling the first step in the search for damage. In these experiments, we observe on average only two proteins bound to long (~ 4000 bp) strands of DNA, which supports the model that these metalloproteins may be mediating charge over very long molecular distances. By employing electrochemical, *in vivo*, and AFM techniques, we have identified residues that may be integral to the CT pathway between the [4Fe4S] cluster and the DNA. Excitingly, we have found that, regardless of the organism or the repair pathway from which they originate, repair proteins that are able to mediate charge transport can redistribute to find lesions. A potential trigger of this search process, guanine oxidation has been probed with a metal-based photooxidant. Here, we have identified and explored the properties that are required for a protein to participate in a DNA-mediated search and gained valuable insight as to how this might occur in a biological context.

TABLE OF CONTENTS

Chapter 3. Mutants of the Base Excision Repair Glycosylase, Endonuclease III: DNA Charge Transport as a First Step in Lesion Detection

Chapter 4. Monitoring MutY Activity: Interactions Among Base Excision Repair Proteins *In Vivo* in the Presence of Oxidative Stress

LIST OF FIGURES AND SCHEMES

Chapter 4 Figures

Chapter 5 Figures

Chapter 6 Figures

Chapter 7 Figures

LIST OF TABLES

CHAPTER 1

DNA-mediated Charge Transport in a Biological Context

Adapted from Barton, J. K., Olmon, E. D., Sontz, P. S. (2011) *Coordination Chemistry Reviews* **255**, 619–634, and Sontz, P. S., Muren, N. B., Barton, J. K. (2012) *Accounts of Chemical Research*, submitted.

INTRODUCTION

The DNA π-stack facilitates charge transport (CT) over long distances, and these reactions are sensitive to perturbations in base pair stacking, such as single base mismatches and lesions (*1-3*). In vitro and electrochemical studies have established that remarkably, DNA CT proceeds over molecular distances as large as 340 Å (*4*), but how might DNA CT be exploited within a cellular context? We have considered several ways in which DNA-mediated CT could act as a conduit to funnel damage to distant sites, trigger transcription of genes to activate cellular defense pathways, or even provide a platform for bound repair proteins to signal one another in their search for lesions. Indeed, we are now discovering that within the cell, redox-active proteins may utilize the same chemistry as that of synthetic metal complexes and electrochemical redox probes *in vitro*, and these proteins may harness DNA-mediated CT to reduce damage to the genome and regulate cellular processes.

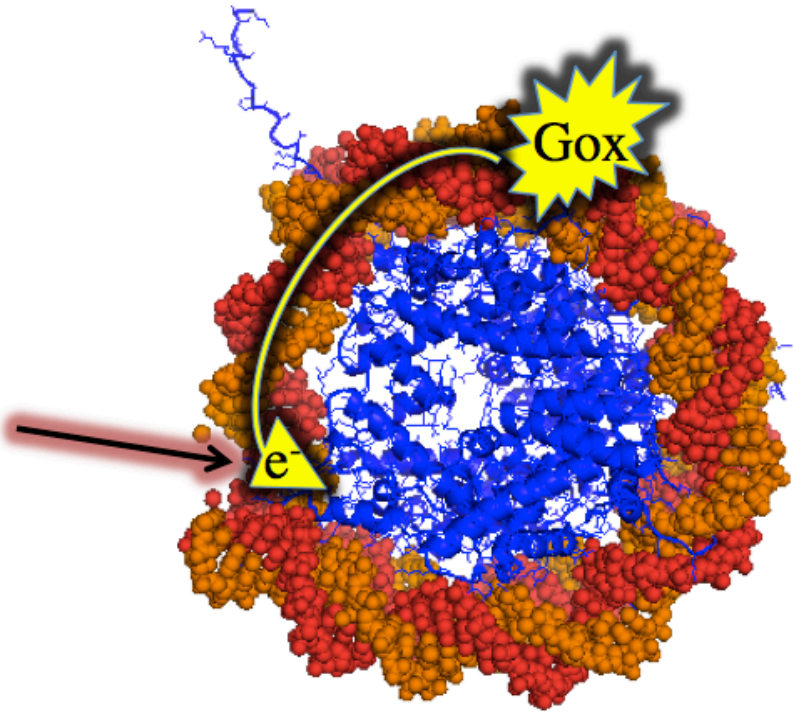
DNA Damage Over Long Ranges in the Cell

We have employed flash-quench and electrochemical techniques to characterize DNA CT (i.e., rate, distance dependence, sensitivity to structural changes), though an outstanding question still remains: could this process occur in a physiological context? In an effort to examine this question, we have probed hole migration in the nucleus of the cell (*5*). HeLa cell nuclei were isolated and mixed with a metal-based photooxidant, $Rh(phi)_2DMB^{3+}$ (DMB: 4,4′-dimethyl-2,2′-bipyridine). We found that, upon irradiation, the complex generated strand breaks at intercalation sites that were distinct from sites of guanine oxidation. Moreover, even when constitutively bound proteins prevented $Rh(phi)_2DMB^{3+}$from intercalating in the DNA, guanine doublet oxidation patterns (5′-G of 5′-GG-3′) were still observed at these sites (*5*). These results indicate that long-range CT,

mediated via π-stacking of the base pairs, can proceed across regions of DNA that would otherwise be inaccessible, generating damage at remote points.

What, then, are elements in the cell that modulate DNA CT? Methylases (i.e., M. Hha1), for example, extrude cytosines from their target sites to expose them for methylation (*6, 7*). Since this activity disrupts base pair stacking, methylases inhibit DNA CT. Similarly; TATA-binding proteins kink DNA duplexes by 90° at their recognition sites, thus attenuating CT (*7, 8*). On the other hand, there are examples of proteins that do not hinder CT when they bind; rather they facilitate electron transfer by stabilizing the duplex (*9*). Importantly, in eukaryotic cells, DNA is packed in nucleosome core particles (*10*). Strands of DNA (~ 150 bp), which have negative backbones, are wrapped around positively charged octamer histone proteins. We thought this DNA structure might confer resistance to reactive oxygen species (ROS) (*10*). To determine the effects of packing on long-range CT in DNA, we utilized rhodium photooxidants to generate guanine oxidation products in reconstituted nucleosomes (*10*). Although the complex could not access the DNA as efficiently, studies with noncovalently bound $Rh(phi)_2DMB^{3+}$ revealed damage at the 5′-G of 5′-GG-3′, similar patterns found within HeLa cells (*5, 10*). Remarkably, when $Rh(phi)_2(bpy)^{3+}$ was tethered to the 5′ end of the DNA, the complex both intercalated and produced damage at distant sites up to 24 bp away (*10*) (Figure 1.1). Long-range CT within DNA in the nucleosome was confirmed in a similar study that employed a different yet comparable photooxidant, anthraquinone (*11*). We then considered whether DNA CT over long ranges could transmit damage to distant sites that either absorb and mitigate damage or, alternatively, propagate further downstream mutations. Notably, analysis of genomic DNA has revealed that introns and exons contain different numbers of sites with low oxidation potential (*12*). It is possible then, that DNA-mediated CT funnels damage away from specific protein coding regions, clustering it elsewhere, and thereby protecting these sites from oxidation.

Figure 1.1. DNA in cells is wrapped around a histone core. Photooxidation of a metal complex tethered to DNA generates oxidative damage (G_{ox}) at distant guanine sites in the nucleosome core particle (pdb ID: 1aoi).

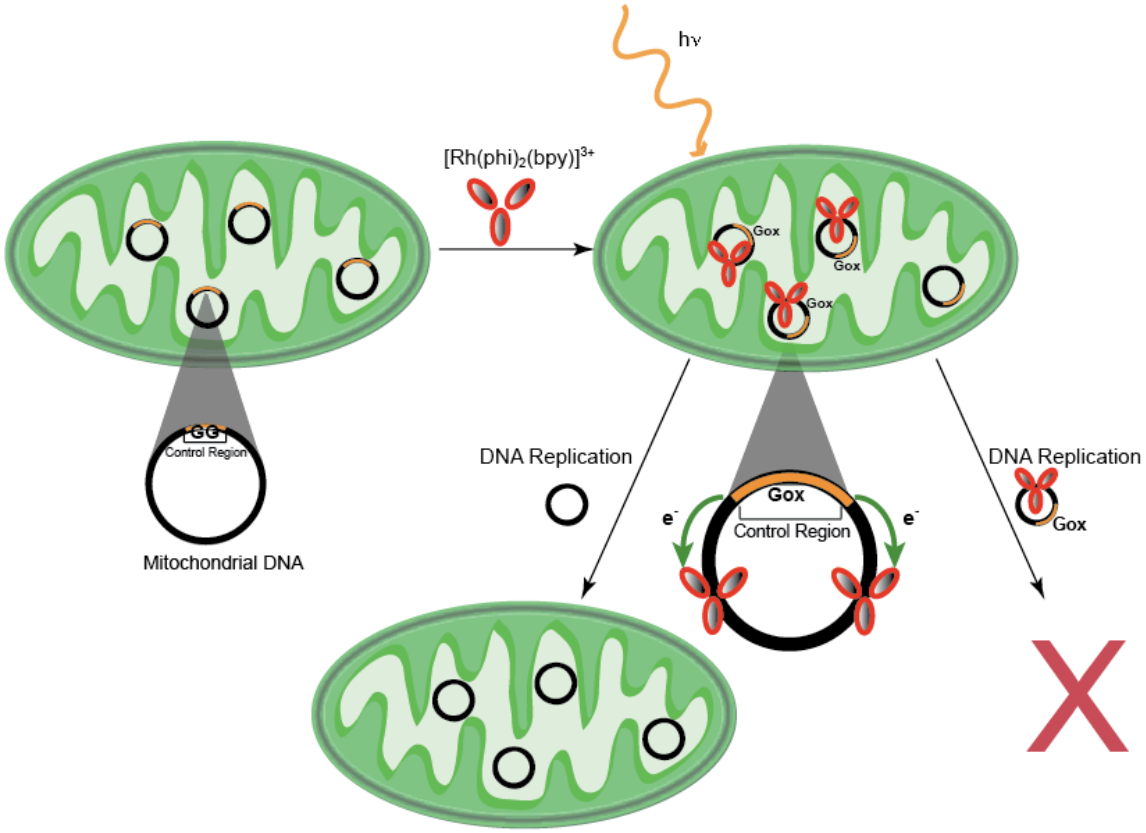


Funneling Damage to Specific Regions of DNA

Mitochondria, cellular sites of respiration, accumulate ROS resulting from oxidative phosphorylation (*13, 14*). These organelles also contain their own DNA, making them particularly interesting systems to examine with respect to DNA CT. Photoactivation of $[Rh(phi)_2(bpy)]^{3+}$ at long wavelengths (365 nm) in the presence of DNA generates base mutations, such as 8-oxodeoxyguanosine (8-oxodG), identical to naturally occurring oxidative lesions. Irradiation at shorter wavelengths (313 nm) results in direct strand cleavage, marking the specific site of complex intercalation (*15, 16*). HeLa cells incubated with $[Rh(phi)_2(bpy)]^{3+}$, which is readily absorbed, and irradiated, can also be analyzed for DNA mutations and oxidative damage (*17*). Rh complex binding sites can be directly compared to sites of guanine oxidation with a primer extension assay (*15*). Base oxidation, resulting from DNA CT hole migration *in vitro* or in HeLa cells, has been shown to occur primarily in conserved sequence block II, nucleotides 303-315, of mtDNA (*16*). Furthermore, this sequence is located within the control region of mitochondria, which regulates primer formation for mtDNA replication (*18, 19*). Mutations within this region of mtDNA are associated with tumor formation and other diseases (*20*).

Remarkably, sites of DNA oxidation via CT within the mitochondria occur as far as 70 bases from the intercalated metal complex, suggesting that DNA CT plays a role in mtDNA oxidative damage (*16*). Conserved sequence block II, which contains seven consecutive guanines, is a site of very low oxidation potential, and could, therefore, act as a sink for lesions such as 8-oxodG (*21*). Extensive damage to this region through DNA-mediated funneling would save the cell from further error propagation by preventing replication of the damaged mitochondrial genome (Figure 1.2). Nevertheless, replication of DNA containing lesions ultimately leads to mutations such as dG to dT transversions, and these mutations have been shown also to occur within conserved sequence block II. This ultimately reduces the amount of guanine in the regulatory region, eliminating the checkpoint and funneling site for oxidative stress, limiting mutant mitochondrial function.

Figure 1.2. Multiple copies of mitochondrial DNA (black) are found in mitochondria (green) organelles within the cell. Irradiation of $Rh(phi)_2(bpy)^{3+}$ photooxidant results in oxidation of sites with low oxidation potential (G_{ox}). Damage in the genome is funneled (green arrows), via DNA-mediated electron transfer (e^-) to the control region (orange), preventing replication of the lesion-filled plasmid (bottom right). DNA replication of undamaged DNA occurs, aiding in the survival of functional error-free mitochondria (bottom left).

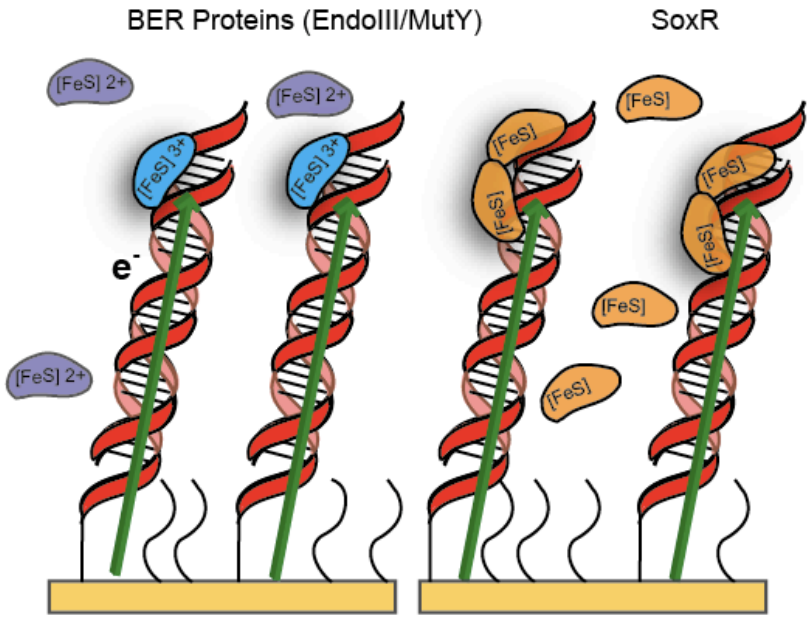


Tumor cells, which do not necessarily depend on respiration, could then survive utilizing alternative energy pathways independent of mitochondria (*17*).

DNA CT with Metalloproteins

To further our understanding of DNA CT in biology, we have extended our studies to examine metalloproteins from repair pathways. Specifically, the proteins from the base excision repair (BER) network recognize and repair DNA bases that have been damaged by oxidative stress (*22*). Left unrepaired, these lesions would threaten the integrity of the genome. We first employed electrochemical techniques to examine whether DNA damage products found in nature influence DNA-mediated signaling (*1-3*). Just as we established that single base mismatches inhibit DNA CT, as they disrupt base pair stacking, common lesions, including O4-methyl-thymine, O6-methyl-guanine, 8-oxo-guanine, and 5-hydroxy-cytosine have this effect as well (*3*). Notably, a subset of base excision repair (BER) proteins contain $[4Fe4S]^{2+}$ clusters within a Cys-X_6-Cys-X_2-Cys-X_5-Cys loop near the C terminus of the protein that are not required for the protein to fold into a native conformation (*23*). A vast number of DNA repair proteins are known to contain this loop, defined by the first two ligating cysteines, a proposed structural element (*24-26*). Having no clear structural role, we wondered: why are the iron-sulfur cofactors present? Crystal structures (with DNA) of two BER proteins from *E. coli*, MutY and Endonuclease III (EndoIII), have revealed the proximity of the [4Fe4S] cluster to the DNA backbone (*27-40*). EndoIII is resistant to reduction in solution, having an estimated $[4Fe4S]^{2+/1+}$ couple midpoint potential of < -600 mV (*26, 28*). Notably, initial measurements of the redox potentials of these proteins were performed in the absence of DNA. We then wondered whether the DNA environment would alter the redox properties of the [4Fe4S] cluster. Measurements of DNA-bound potentials were then made using DNA-modified electrodes to examine this question (Figure 1.3).

Figure 1.3. Examples of surfaces used for DNA/protein electrochemistry of BER proteins (left) or SoxR (right). DNA duplexes are attached to the gold via a 5′ thiol linker. Mercaptohexanol (curved lines) is used to backfill so that the duplexes are in an upright position. Electrons travel (arrow) from the gold surface to the bound protein. EndoIII and MutY are bound when the [4Fe4S] cluster is in the 3+ oxidation state and SoxR binds as a dimer.



MutY was probed electrochemically on loosely packed DNA gold electrode films (*34*). A thiol-terminated duplex with a non-specific DNA binding site was attached to a gold surface and passivated with mercaptohexanol. Cyclic voltammetry of MutY at the DNA-modified surface yielded a midpoint potential +90 mV vs. NHE, an appropriate potential for a physiologically active redox switch. In the absence of DNA, examination of a surface coated with mercaptohexanol yielded no electrochemical signal. Therefore, DNA was required to observe reduction of the [4Fe4S] cluster-containing protein. When the electrode was modified with duplex DNA containing an abasic site, no signal was observed, establishing that electron transfer occurs through the DNA to the bound MutY protein. The redox reaction of MutY is DNA-mediated.

Redox potentials of +100 mV vs. NHE are common for high-potential iron proteins, and in the case of MutY (+90 mV vs. NHE) indicated a shift in the oxidation state of the [4Fe4S] cluster from 2+ to 3+ as the protein binds to the DNA (*35, 36*). EndoIII was also examined bound to a DNA-modified electrode. As with MutY, for EndoIII when the potential was held at +50 mV, promoting oxidation of the cluster, the signal increased. Quite remarkably, both proteins display signals with midpoint potentials in the presence of DNA of ~ 100 mV vs. NHE. Both also display a much smaller signal when the intervening duplex contains an abasic site. We can then attribute oxidation of the cluster to DNA-mediated CT from the electrode through the base pair stack to the bound protein.

Building upon these studies, EndoIII was then examined bound to DNA versus free using highly oriented pyrolytic graphite (HOPG) electrodes (*37*). In these experiments, graphite electrodes were modified with pyrenated DNA, and both cyclic voltammetry and square wave voltammetry data were collected. A midpoint potential of ~ 20 mV vs. NHE was observed in the presence of well-matched DNA and EndoIII. In contrast, in the absence of DNA, oxidative scans of EndoIII yielded an irreversible anodic peak at ~ 250 mV and loss of the yellow solution color, an indicator of [4Fe4S] cluster degradation (*37*). Consecutive positive scans showed broad, irregular signals at -80

and -710 mV vs. NHE, representative of clusters that have lost an iron to become [3Fe4S]. The 2+/3+ EndoIII [4Fe4S] cluster redox couple exhibited a positive potential (~ 250 mV). These data taken together indicate that EndoIII in the 3+ oxidation state binds DNA at least a thousand fold more tightly than the reduced, 2+, form. This characteristic shift in binding affinity between the reduced and oxidized forms of MutY and EndoIII has been an important element in our proposed mechanism for how these repair proteins scan the genome using DNA CT to find their lesions (*35, 37-39*).

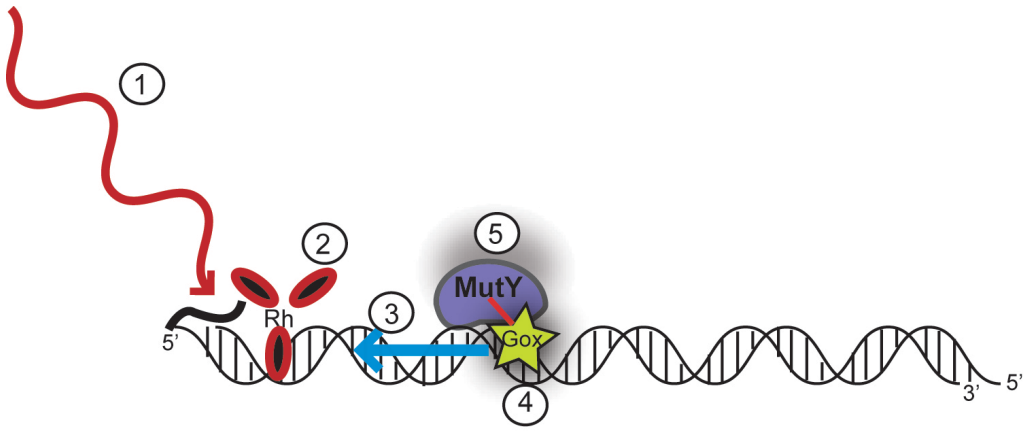
Another metalloprotein, the SoxR transcription factor, regulates the response to superoxide within the cell. SoxR contains two [2Fe2S] clusters (one in each monomer) that are not required for protein folding (*40-43*). The oxidation of SoxR leads to the expression of the SoxS transcription factor to control expression of genes involved in protecting the cell from stresses, but interestingly the affinities of SoxR for its promoter are equal for the oxidized and reduced form of the protein (*44*). Redox potentials for SoxR without DNA, however, were found to be about -290 mV vs. NHE, providing somewhat of a conundrum since that would suggest that at cellular potentials, SoxR was already oxidized, not an optimum situation for a sensor of oxidation [*42, 45, 46*]. Determination of the DNA-bound redox potential of SoxR was thought to provide insight to the activation mechanism of the protein. Well-matched DNA duplexes, attached to an HOPG electrode via pyrene modification, were self-assembled to form loosely packed DNA monolayers. Redmond red was used as the redox-active probe in these studies, with a midpoint potential of -160 mV vs. NHE (*47*). A quasi-reversible stable electrochemical signal was found for the [2Fe2S] cluster of SoxR at +200 mV vs. NHE (*44*). Differing from MutY and EndoIII signals, SoxR showed asymmetries in reduction and oxidation waves, displaying a better resolved anodic peak. Measurements of SoxR extracted from different organisms yielded similar potentials. Significantly, then, there was ~ 500 mV positive shift in redox potential of SoxR upon DNA binding. It is likely this shift in potential that provides the energetics for the torqueing of the DNA by oxidized SoxR to activate

transcription. Critically, we see the importance of performing redox measurements in the presence of DNA, as the DNA polyanion alters the protein environment and potential. MutY, EndoIII, and SoxR show redox activity in a physiologically relevant regime only when bound to DNA. These redox-active proteins, then, may use DNA CT as a sensor for oxidative damage and to locate lesions awaiting repair.

Methods to probe long-range DNA-mediated oxidation of 5′-GG-3′ have aided in the study of DNA/protein interactions so that we can gain insight into how these metalloproteins might take advantage of the unique property of DNA to conduct charge. Experiments harnessing DNA-mediated CT to generate oxidized guanine, leading to DNA-protein cross-links, have provided a more detailed look at the amino acid/DNA interface (*48-56*). Noncovalent ruthenium intercalators, upon photoactivation, have been used to generate guanine radicals, and these radicals can promote adducts with histone proteins.

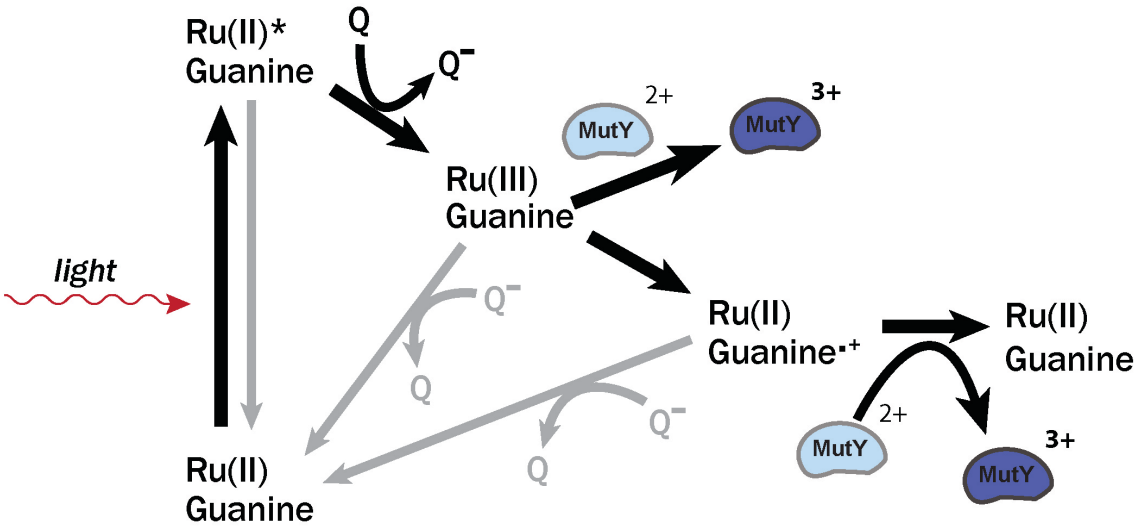
DNA/MutY cross-linking experiments were performed using tethered $[Rh(phi)_2(bpy')]^{3+}$ to generate oxidative damage (*56*). MutY was chosen for initial studies, since base flipping was thought to be involved in lesion detection. Worth nothing, lysine 142 in MutY was thought to play a large mechanistic role during the repair of 8-oxo-G:A lesions based on early cross-linking and NMR investigations (*49, 57, 58*). In our experiments, interestingly, there was no apparent change in distal/proximal guanine ratios upon DNA-mediated guanine oxidation in the presence of up to 200 nM of protein. This result indicated that MutY binding does not perturb long-range DNA CT, and thus might use a mechanism other than base flipping to search for damage. Oligonucleotides with an 8-oxo-G:7-deaza-A MutY binding site 20 Å away from the tethered Rh photooxidant were then used to probe DNA CT-mediated cross-link formation between MutY and 8-oxo-G (Figure 1.4). A covalent cross-link of 8-oxo-G with MutY was identified involving lysine 142. This provided an example of long-range DNA-mediated CT providing a route to a protein/DNA redox reaction (*56*).

Figure 1.4. Upon irradiation at 365 nm (1) of $[Rh(phi)_2(bpy')]^{3+}$ intercalator (2), electrons are pulled from the site of lowest oxidation potential, guanine (3). Guanine is then oxidized (4) to form 8-oxo-G, forming a cross-link with bound MutY through lysine 142 contacts (5).



Having reliable methods to generate oxidative damage and knowing that MutY is redox-active from electrochemical studies, we have also been able to show that MutY itself can be activated in a DNA-mediated redox dependent manner (*38*). $[Ru(phen)_2dppz]^{2+}$ binds to DNA by intercalation and is excited in the presence of visible light. This excited state is then quenched by an electron acceptor to form a Ru(III) species in situ that can then oxidize guanines from a distance, and then the guanine radical, once formed can then oxidize MutY [Scheme 1.1] (*48, 59*). To test this, poly(dGC) and poly(dAT) duplexes were irradiated in the presence of $[Ru(phen)_2dppz]^{2+}$ and quencher, $[Co(NH_3)_5Cl]^{2+}$, with or without MutY. EPR spectra at 10K revealed a primary g value of 2.02 which was attributed to the $[3Fe4S]^{1+}$ cluster, an oxidative degradation product of the $[4Fe4S]^{2+}$ cluster (*38*). Traces showed an additional signal at g values of 2.08 and 2.06 for both poly(dGC) and poly(dAT) duplexes in the presence of MutY, assigned to the oxidized $[4Fe4S]^{3+}$ cluster (*38, 60*). Notably, with poly(dAT), these signals were significantly lower in intensity than with poly(dGC). Thus the [4Fe4S] cluster of MutY can become oxidized to the 3+ state with or without intervening guanine radicals. The guanine radical, owing to its relatively long lifetime, can, however, increase the yield of protein oxidation.

Scheme 1.1. The flash-quench technique used to generate Ru(III) and subsequently oxidize DNA-bound MutY (dark blue). Following photoexcitation, Ru(II)* is quenched forming powerful ground-state oxidant Ru(III) which can then proceed to oxidize guanine within the base stack. The guanine radical not only forms oxidative products but also oxidizes proteins, such as MutY. Back electron transfer reactions are shown in gray.



Investigating oxidation of MutY in a DNA-mediated reaction on a faster time scale was carried out using transient absorption measurements in the presence of $[Ru(phen)_2dppz]^{2+}$, poly(dGC) DNA, and a $[Ru(NH_3)_6]^{3+}$ quencher (*38*). As the DNA-bound metal complex was photoactivated, emission decays were observed at 610 nm. Transient absorption signals fit to a biexponential indicated a fast phase guanine radical formation in duplex DNA that decreased in the presence of MutY. Also, a new species appeared in the poly(dGC) slow phase spectrum with an absorption maximum characteristic of the $[4Fe4S]^{3+/2+}$ difference spectrum. Importantly, these experiments established that guanine radical formation could directly lead to oxidation of bound protein. The intervening guanine radicals may allow MutY more time to undergo oxidation.

Biochemical studies examining 42-mer DNA duplexes incorporating a 5′-GG-3′ doublet covalently linked to a ruthenium intercalator also showed inhibition of 5′-G damage in the presence of MutY. This system differs from covalently tethered $[Rh(phi)_2(bpy')]^{3+}$ used to generate oxidative damage in previous experiments, owing to its ability to generate guanine radicals through the flash/quench pathway. Guanine has an oxidation potential of -1.29 V vs. NHE, making it easily oxidized in the presence of Ru(III) complex, which has a reduction potential of 1.5 V vs. NHE (*61, 62*). Importantly, the midpoint potential of DNA-bound MutY $[4Fe4S]^{2+/3+}$ is +0.1 V vs. NHE. Thus, instead of generating oxidative DNA damage, MutY can intercept the guanine radicals, becoming oxidized itself (*34*). This explains the inhibition of damage in the denaturing PAGE gel experiments tracking 5′-G damage. EPR studies further confirmed this hypothesis, displaying strong signals attributed to the oxidized $[4Fe4S]^{3+}$ cluster in samples of Ru complex tethered to a 5′-GG-3′-containing 42-mer duplex in the presence of MutY. Collectively, transient absorption, EPR, and biochemical experiments indicate that MutY can be oxidized in a DNA-mediated fashion. Moreover, the implications of these findings for DNA repair by redox-active proteins are critical, since within the cell, oxidative stress may lead to this chemistry as a means to activate repair.

CT to Trigger Transcription from a Distance

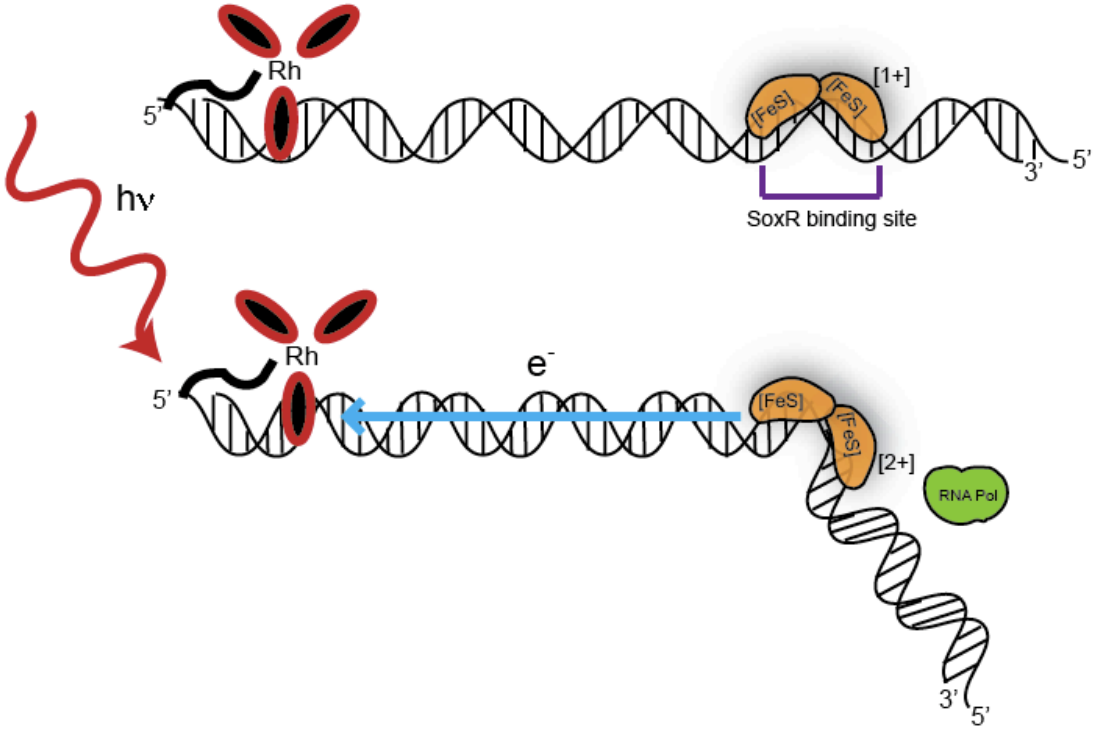
Inspired by our findings that DNA-mediated CT can trigger MutY, we have examined yet another system that includes a metallo transcription factor, SoxR. As mentioned earlier, SoxR is activated in the presence of oxidative stress; however, the specific source for oxidation in the cell is unknown. The protein is reversibly inactivated *in vitro* by reduction of its [2Fe2S] clusters using dithionite (*42*). *In vivo* studies, using redox-cyclers such as paraquat to induce oxidative stress, show that superoxide is not the direct activator of SoxR. Rather, there is depletion of cellular NADPH, which is normally required to keep SoxR in a reduced form. The redox-cyclers then undergo autooxidation and produce superoxide by losing an electron to dioxygen. Plumbagin and phenazine methosulfate, notably, have yielded reversible oxidation of the [2Fe2S] clusters (*42*). Electrochemistry of SoxR in our lab showed a redox-active signal for the [2Fe2S] cluster of SoxR at +200 mV vs. NHE, indicating the protein undergoes one electron oxidation bound to DNA (*44*). Due to the shift seen upon binding to DNA, we questioned whether the DNA-bound form of SoxR might be the missing oxidative switch. As DNA is subjected to oxidative damage, and guanine radicals, capable of migrating through the base pair stack, are generated, these guanine radicals can then serve as the oxidant for SoxR. Once oxidized, SoxR can promote transcription under conditions when the cell needs it the most. The oxidative signal then would be the guanine radical.

We have learned that the coupling properties of photooxidants used to probe DNA-mediated CT are important factors in observing this process. Therefore, SoxR studies were recently conducted in our lab using a strong DNA intercalator, $[Ru(phen)(dppz)(bpy')]^{2+}$, where dppz is the moiety interacting with the DNA, to generate guanine radicals resulting in irreversible oxidative damage (*63*). $[Co(NH_3)_5Cl]^{2+}$ was used as the quencher in these studies, eliminating back electron transfer, so that the major pathway in the reaction yields oxidative damage. Similar to DNA-bound MutY experiments described earlier where MutY oxidation reduced guanine damage, 5′-GG-3′ oxidation

products were measured with and without oxidized or reduced SoxR. Upon reduction of SoxR with sodium dithionite in the presence of DNA-bound Ru, quencher, and light, damage at guanine sites was attenuated. This indicated that SoxR was able to donate an electron to the oxidized guanine, going from a reduced to an oxidized form, filling the radical hole with its own lost electron. Conversely, there was no attenuation in damage when oxidized SoxR was used in the experiment. These results clearly showed that SoxR is able to interact with the DNA base stack, participating in DNA-mediated CT. The SoxR response to guanine radicals generated in DNA suggests that oxidative damage can be an initiator of protein activation.

Examining the SoxR response to DNA damage in its native cellular environment, *E. coli* cultures were treated with $[Rh(phi)_2bpy]^{3+}$. This intercalator, with photoactivation, was previously shown to generate guanine damage in mitochondria and HeLa cells (*15-17*). Transcription of the *soxS* product, RNA, observed using reverse transcription PCR, reflects activation of SoxR. Bacteria irradiated in the presence of the Rh photooxidant showed much greater *soxS* levels than those without. Since the complex strictly induces DNA damage and cannot oxidize SoxR directly without DNA, SoxR must be activated by guanine radicals in these experiments. Expression levels of *soxS* in the presence of $[Rh(phi)_2bpy]^{3+}$ were similar to those occurring when methyl viologen, a redox-cycler, was used, both systems activating SoxR to the same extent. But is SoxR oxidized from a distance through DNA-mediated CT (*63, 64*)? To test this idea, $[Rh(phi)_2bpy]^{3+}$ was tethered to a 180-mer duplex of DNA containing the SoxR binding site and the -10 and -35 promoter regions of *soxS* (Figure 1.5), irradiated, and gene products were monitored with an anaerobic abortive transcription assay (*41*). The abortive transcription product, a 4-mer, was only observed in samples containing reduced SoxR and tethered complex/DNA. Importantly, there was no direct photooxidation of SoxR by $[Rh(phi)_2bpy]^{3+}$ without DNA. Remarkably, then, DNA-mediated transcriptional activation of SoxR was indeed demonstrated over a 270 Å distance. These data establish that guanine radicals are capable of oxidizing SoxR from a distance.

Figure 1.5. $[Rh(phi)_2(bpy')]^{3+}$ is tethered to a duplex of DNA containing the SoxR binding site. SoxR is bound in the reduced (+1) state (yellow protein). Photoactivation of the metal complex triggers electron transfer, oxidizing SoxR (2+), which may then kink the DNA and recruit transcription machinery such as RNA polymerase (green).



Thus, CT over long ranges may play some role in a number of cellular pathways including: funneling damage in mitochondria, enhancing fidelity during replication, and activating transcription to defend the cell from oxidative stress. It is clear that in a biological context, DNA CT may "wear many hats".

DNA-mediated Signaling for Lesion Detection by Repair Proteins

DNA is the main source of genetic information in cells, allowing them to function, store information, and evolve. The specific sequence of the bases in DNA ultimately translates into the structure and function of proteins (*65*). Thus, the integrity of the DNA sequence is vital to cells, though there are many mutagens that can produce lesions that are difficult to repair, potentially blocking DNA replication or producing defective proteins. For example, thymine dimers caused by UV light, base modifications such as 7, 8-dihydro-8-oxo-2-deoxyguanosine (8-oxoG) due to oxidative damage, or double-strand breaks can all lead to insertions, deletions, and point mutations in the DNA sequence, subsequently destroying the genetic code within the cell (*66*).

Damaged bases in DNA are known to lead to the disruption of DNA replication as well as act as possible indicators of cancer due to the high frequency of mutations that arise from these lesions (*22, 35*). Thus, the cell must quickly detect and repair lesions. Examples of endogenous DNA repair pathways include direct repair, base-excision repair (BER), and nucleotide-excision repair (*19*). There is much known about the mechanism of base removal by enzymes from these pathways; however, precisely how these bases are located by glycosylases remains unknown.

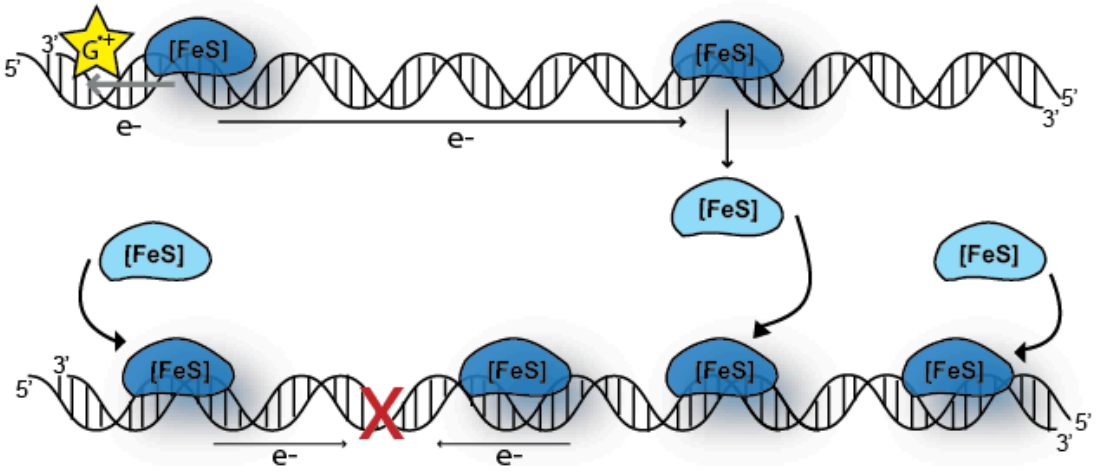
We now know that BER proteins, similar to SoxR, are capable of 'absorbing' guanine radical damage by donating one of their own electrons into the DNA π-stack. We have performed many additional *in vitro* studies to establish that BER proteins can be oxidized and reduced via DNA-mediated CT (*38, 67*). The results from these studies led us to wonder whether BER proteins might

exploit DNA CT to recognize their lesions. We have thus proposed a mechanism that may be activated within cells for repair utilizing DNA CT. Importantly, binding to DNA shifts the potential of the EndoIII [4Fe4S] cluster toward oxidation, which could activate it when repair is required (*37*). The BER proteins all have similar DNA-bound redox potentials and contain a reduced $[4Fe4S]^{2+}$ in solution in the absence of oxidative stress (*38*). Guanine radicals, forming endogenously under conditions of oxidative stress, could initiate binding of protein to DNA, oxidizing the [4Fe4S] cluster from 2+ to 3+, increasing protein affinity for the duplex. As this occurs, an electron is transferred through the DNA to a distally bound BER protein. This distant protein is then reduced, loses its affinity for DNA, and is free to relocate onto another site. The process continues as a scan of the region of DNA between the proteins as long as CT occurs. In the presence of a lesion such as a mismatch or oxidized base, however, known to attenuate charge transfer, DNA CT does not proceed between bound proteins and thus the proteins remain in the vicinity of the damage, and slowly proceed to the site of damage (Figure 1.6). In this model, proteins are expected to redistribute onto regions of DNA that contain lesions. Additionally, the rate of this process depends on the distance over which CT can occur (hundreds of base pairs) and the percentage of oxidized protein (*39*). Assuming only facilitated diffusion and instantaneous interrogation, the amount of time required for a protein that is in low copy number to search for damage within the cell is reduced by at least an order of magnitude when it is able to cooperate with other proteins (*39*).

Summary

Significantly, *in vivo* studies depend on generation of DNA damage by photooxidants to elucidate the role of CT within cells. Utilizing metal complexes as photooxidants, the funneling of damage to specific regulatory sites in the mitochondrial genome has been identified. Moreover CT can occur not just to generate guanine damage but also to activate repair proteins and transcription. Here too we have seen DNA-mediated redox chemistry, but now with metalloproteins.

Figure 1.6. Model for a DNA-mediated search by repair proteins. Guanine radicals (yellow) can initiate binding of redox-active protein to DNA, oxidizing the [4Fe4S] cluster from 2+ to 3+ (dark blue), increasing protein affinity for the duplex. An electron travels through the DNA to a distally bound BER protein. This distant protein reduced, loses its affinity for DNA (light blue) and relocates to another site. In the presence of a lesion such as a mismatch or oxidized base (bottom), DNA CT does not occur, and the proteins remain bound and process to the lesion.



Using reactions with simple coordination complexes to guide our questions, our goal now shifts to discovery of various endogenous "metal complex" systems in which a metalloprotein couples with DNA using DNA-mediated CT to activate signaling across the genome.

The search for damage in the genome is not a process limited to BER glycosylases. Proteins from a wide variety of organisms such as archaea are being identified as DNA-binding proteins that contain [4Fe4S] clusters with no obvious functional role (*68, 69*). For example, XPD, an archaeal protein, is a 5′-3′ helicase that is part of the TFIIH machinery required for nucleotide excision repair (NER). XPD contains a conserved [4Fe4S] cluster speculated to be conformationally controlled by ATP binding and hydrolysis (*68, 70, 71*). Mutations in the iron-sulfur cluster domain of XPD can lead to TFIIH instability, which results in serious diseases such as trichothiodystrophy (*71-73*). It is then interesting to consider whether XPD is yet another protein with a DNA-bound midpoint potential and CT signal conducive to assisting other proteins in their search for damage. We have demonstrated that the [4Fe4S] cluster in *Sulfolobus acidocaldarious* (Sa)XPD is redox-active (*70*), exhibiting a midpoint potential identical to that of BER proteins (~ 80 mV vs. NHE). We have developed the methods needed to investigate whether proteins that contain redox-active moieties and bind DNA, from distinct repair pathways (primases, polymerases, etc.), are responsive to DNA-mediated CT (*74*). It is then possible that there are many more proteins we have yet to uncover that participate in the DNA-mediated search for damage *in vivo*.

REFERENCES

1. Kelley, S. O., Boon, E. M., Barton, J.K., Jackson, N. M., and Hill, M. G. (1999) *Nucleic Acids Res. 27*, 4830.

2. Boon, E. M., Ceres, D. M., Drummond, T. G., Hill, M. G., and Barton, J. K. (2000) *Nat. Biotechnol. 18*, 1096.

3. Boal, A. K., and Barton, J. K. (2005) *Bioconj. Chem*. *16*, 312.

4. Slinker, J. D., Muren, N. B., Renfrew, S. E., and Barton, J. K. (2011) *Nat. Chem. 3*, 228.

5. Núñez, M. E., Holmquist, G. P., and Barton, J. K. (2001) *Biochemistry 40*, 12465.

6. Boon, E. M., Salas, J. E., and Barton, J. K. (2002) *Nat. Biotechnol. 20,* 282.

7. Rajski, S. R., and Barton, J. K. (2001) *Biochemistry 40*, 5556.

8. Gorodetsky, A. A., Ebrahim, A., and Barton, J. K. (2008) *J. Am. Chem. Soc. 130*, 2924.

9. Boon, E. M., and Barton, J. K. (2002) *Curr. Opin. Struct. Biol. 12*, 320.

10. Núñez, M. E., Noyes, K. T., and Barton, J. K. (2002) *Chem. Biol. 9*, 403.

11. Bjorklund, C. C., and David, W. B. (2006) *Nucleic. Acids Res. 34*,1836.

12. Friedman, K. A., and Heller, A. (2001) *J. Phys. Chem. B 105*, 11859.

13. Wallace, D.C. (2005) *Annu. Rev. Genet. 39*, 359.

14. Cadenas, E., and Davies, K. J. A. (2000) *Free Radical Biol. Med*. *29*, 222.

15. Merino, E. J., and Barton, J. K. (2007) *Biochemistry 46*, 2805.

16. Merino, E. J., and Barton, J. K. (2008) *Biochemistry 47*, 1511.

17. Merino, E. J., Davis, M. L., and Barton, J. K. (2009) *Biochemistry 48*, 660.

18. Xu, B., and Clayton, D. A. (1996) *EMBO J. 15*, 3135.

19. Pham, X. H., Farge, G., Shi, Y. H., Gaspari, M., Gustafsson, C. M., and Falkenberg, M. (2006) *J. Biol. Chem. 281*, 24647.

20. Tan, D. J., Bai, R. K., and Wong, L. J. C. (2002) *Cancer Res. 62*, 972.

21. Merino, E. J., Boal, A. K., and Barton, J. K. (2008) *Curr. Opin. Chem. Biol. 12*, 229.

22. David, S. S., and Williams, S. D. (1998) *Chem. Rev. 98,* 1221.

23. Porello, S. L., Cannon, M. J., and David, S. S. (1998) *Biochemistry 37,* 6465.

24. Kou, C.F., McRee, D. E., Fischer, C. L., O'Handley, S. F., Cunningham, R. P., and Tainer, J.A. (1992) *Science 258*, 434.

25. Thayer, M. M., Ahern, H., Xing, D., Cunningham, R. P., and Tainer, J. A. (1995) *EMBO J. 14*, 4108.

26. Fu, W., O'Handley, S., Cunningham, R. P., and Johnson, M. K. (1992) *J. Biol. Chem. 267*, 16135.

27. Asahara, H., Wistort, P. M., Bank, J. F., Bakerian, R. H., and Cunningham, R. P. (1989) *Biochemistry 28,* 4444.

28. Cunningham, R. P., Asahara, H., Bank, J. F., Scholes, C. P., Salerno, J. C., Surerus, K., Munck, E., McCracken, J., Peisach, J., and Emptage, M. H. (1989) *Biochemistry 28,* 4450.

29. O'Handley, S., Scholes, C. P., and Cunningham, R. P. (1995) *Biochemistry 34,* 2528.

30. Xing, D., Dorr, R., Cunningham, R. P., and Scholes, C. P. (1995) *Biochemistry 34,* 2537.

31. Michaels, M. L., Pham, L., Nghiem, Y., Cruz, C., and Miller, J. H. (1990) *Nucleic Acids Res. 18,* 3841.

32. Tsai-Wu, J. J., Liu, H. F., and Lu, A. L. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89,* 8779.

33. Lu, A. L., Tsai- Wu, J. J., and Cillo, J. (1995) *J. Biol. Chem. 270,* 23582.

34. Bulychev, N. V., Varaprasad, C. V., Dorman, G., Miller, J. H., Eisenberg, M., Grollman, A. P., and Johnson, F. (1996) *Biochemistry. 35,* 13147.

32. Porello, S. L., Williams, S. D., Kuhn, H., Michaels, M. L., and David, S. S. (1996) *J. Am. Chem. Soc. 118,* 10684.

33. Fromme, J. C., and Verdine, G. L. (2003) *EMBO J. 22,* 3461.

34. Boon, E. M., Livingston, A. L., Chmiel, N. H., David, S. S., and Barton, J. K. (2003) *Proc. Natl. Acad. Sci. 100*, 12543.

35. Boal, A. K., Yavin, E., Lukianova, O. A., O'Shea, V. L., David, S. S., and Barton, J. K. (2005) *Biochemistry 44*, 8397.

36. Cowan, J. A., and Lui, S. M. (1998) *Adv. Inorg. Chem. 45*, 313.

37. Gorodetsky, A. A., Boal, A. K., and Barton, J. K. (2006) *J. Am. Chem. Soc. 128*, 12082.

38. Yavin, E., Boal, A. K., Stemp, E. D. A., Boon, E. M., Livingston, A. L., O'Shea, V. L., David, S. S., and Barton, J. K (2005) *Proc. Natl. Acad. Sci. 102*, 3546.

39. Boal, A. K., Genereux, J. C., Sontz, P. A., Gralnick, J. A., Newman, D.K., and Barton, J. K. (2009) *Proc. Natl. Acad. Sci. 106,* 15237.

40. Pomposiello, P. J., and Demple, B. (2001) *Trends Biotechnol. 19*, 109.

41. Hidalgo, E., and Demple, B. (1994) *EMBO J. 13*, 138.

42. Gaudu, P., and Weiss, B. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*,10094.

43. Ding, H., Hidalgo, E., and Demple, B. (1996) *J. Biol. Chem. 271*, 33173.

44. Gorodetsky, A. A., Dietrich, L. E. P., Lee, P. E., Demple, B., Newman, D. K., and Barton, J. K. (2008) *Proc. Natl. Acad. Sci. U.S.A. 105*, 3684.

45. Ding, H., and Demple, B. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 8445.

46. Kobayashi, K., and Tagawa, S. (2004) *J. Biochem. 136*, 607.

47. Bard, A. J., and Faulkner, L. R. (2001) Electrochemical Methods Wiley, New York, 2nd ed.

48. Burrows, C. J., and Muller, J. G. (1998) *Chem. Rev. 98*, 1109.

49. Hickerson, R. P., Chepanoske, C. L., Williams, S. D., and David, S. S. (1999) *J. Am. Chem. Soc. 121*, 9901.

50. Chakrabarti, S. K., Bai, C. J., and Subramanian, K. S. (1999) *Toxicol. Appl. Pharmacol. 154*, 245.

51. Gavin, I. M., Melnick, S. M., Yurina, N. P., Khabarova, M. I., and Bavykin, S. G. (1998) *Anal. Biochem. 263*, 26.

52. Zitkovich, A., Lukanova, A., Popov, T., Taioli, E., Cohen, H., Costa, M., and Honiolo, P. (1996) *Biomarkers 1*, 86.

53. Izzotti, A., Bagnasco, M., Camoirano, A., Orlando, M., and De Flora, S. (1998) *Mut. Res. 400*, 233.

54. Villanueva, A., Canete, M., Trigueros, C., Rodriguez-Blorlado, L., and Juarranz, A. (1993) *Biopolymers 33*, 239.

55. Nguyen, K. L., Steryo, M., Kurbanyan, K., Nowizki, K. M., Butterfield, S. M., Ward, S.R., and Stemp, E.D.A. (2000) *J. Am. Chem. Soc. 122*, 3585.

56. Boon, E. M., Pope, M.A, Williams, S. D., David, S. S., and Barton, J. K. (2002) *Biochemistry 41*, 8464.

57. House, P. G., Volk, D. E., Thiviyanathan, V., Raymond, R. C., Luxon, B. A., Gorenstein, D. G., and Lloyd, S. R. (2001) *Prog. Nucleic Acid Res. Mol. Biol. 68*, 349.

58. Murphy, C. J., Arkin, M. R., Ghatlia, N. D., Bossmann, S., Turro, N.J., and Barton, J. K. (1994) *Proc. Natl. Acad. Sci. USA 91*, 5315.

59. Friedman, A. E., Chambron, J. C., Sauvage, J. P., Turro, N. J., and Barton, J. K. (1990) *J. Am. Chem. Soc. 112*, 4960.

60. Dilg, A. W. E., Mincione, G., Achterhold, K., Iakovleva, O., Mentler, M., Luchinat, C., Bertini, I., and Parak, F. G. (1999) *J. Biol. Inorg. Chem. 4*, 727.

61. Steenken, S., and Jovanovic, S.V. (1997) *J. Am. Chem. Soc. 119*, 617.

62. Stemp, E. D. A., Arkin, M. R., and Barton, J. K. (1997) *J. Am. Chem. Soc. 119*, 2921.

63. Lee, P. E., Demple, B., and Barton, J. K. (2009) *Proc. Natl. Acad. Sci. 106*, 13164.

64. Hall, D. B., Holmlin, R. E., and Barton, J. K. (1996) *Nature 382*, 731.

65. Malacinski, G., *Essentials of Molecular Biology*; Jones and Burtlett Publishers, Inc.: Massachusetts, 2003; 33-45.

66. Bert, J. M., Tymoczko, J. L., and Stryer, L. (2002) *Biochemistry*; W.H. Freeman and Co.: New York, 769-770.

67. Yavin, E., Stemp, E. D. A., O'Shea, V. L., David, S. S., and Barton, J. K. (2006) *Proc. Natl. Acad. Sci. USA 103*, 3610.

68. Fan, L., Fuss, J. O., Cheng, Q. J., Arvai, A. A., Hammel, M., Roberts, V. A., Cooper, P. K., and Tainer, J. A. (2008) *Cell 133*, 789.

69. Singleton, M. R., Dillingham, M. S., and Wigley, D. B. (2007) *Annu. Rev. Biochem. 76*, 23.

70. Mui, T. P., Fuss, J. O., Ishida, J. P., Tainer, J. A., and Barton, J. K. (2011) *J. Am. Chem. Soc. 133*, 16378.

71. Sontz, P. A., Mui, T. P., Fuss, J. O., Tainer, J. A., and Barton, J. K. (2012) *Proc. Natl. Acad. Sci. USA 109*, 1856.

72. Lehmann, A. R. (2008) *DNA Repair 7*, 1912.

73. Takayama, K., Salazar, E. P., Broughton, B. C., Lehmann, A. R., Sarasin, A., Thompson, L. H., and Weber, C. A. (1996) *Am. J. Hum. Genet. 58*, 263.

74. Netz, D. J. A., Stith, C. M., Stümpfig, M., Köpf, G., Vogel, D., Genau, H. M., Stodola, J. L., Lill, R., Burgers, P. M. J., and Pierik, A. J. (2012) *Nature Chem. Bio. 8*, 125.

CHAPTER 2

Redox Signaling between DNA Repair Proteins for Efficient Lesion Detection

Adapted from Boal, A. K., Genereux, J. C., Sontz, P. A., Gralnick, J. A., Newman, D. K., Barton, J. K. (2009) *Proc. Natl. Acad. Sci.* **106**, 15237-15242

P.A.S. constructed DNA substrates and performed AFM studies. A.K.B. performed electrochemistry experiments, reversion assays, and characterized wild-type and mutant proteins. J.C.G. performed the genome scanning calculations.

INTRODUCTION

Base excision repair (BER) proteins, from bacteria to man, are challenged with combing the genome for DNA base lesions to maintain the integrity of our genetic material (*1, 2*). This challenge is remarkable given the low copy number of these proteins and that they must discriminate among small differences between modified and natural bases. For MutY, a BER repair protein in *E. coli* with a human homolog, there are $\leq$ 30 proteins in the *E. coli* cell (*1*) to interrogate 4.6 million bases; the ratio of binding affinities for the target lesion, an 8-oxoguanine:adenine mismatch, versus well-matched native base pairs is $\leq$ 1000 (*3*). Endonuclease III (EndoIII) recognizes a less prevalent lesion, hydroxylated pyrimidines, with equally low specificity; the copy number of EndoIII within *E. coli* is ~ 500 (*1*). How these glycosylases fix their substrate lesions, once found, has been well characterized (*1, 2*) as are the structures of MutY and EndoIII bound to DNA (*4, 5*). Yet how these lesions are efficiently detected before excision is not established.

Location of damaged bases in the genome is likely the rate limiting step in BER within the cell and, hence, a critical step in maintaining genomic integrity (*6*). Current models for genome scanning to detect lesions involve protein sliding along the DNA, squeezing the backbone, slipping bases out to allow for interrogation, or finding transiently opened sites (*7, 8*). However, given the low copy number of these proteins and their need to sift through the genome to find often subtle base lesions, the time required for this search is long.

Ubiquitous to a subset of these BER proteins are [4Fe4S] clusters, common redox cofactors in proteins (*1, 2*). Increasingly, iron-sulfur clusters are found associated with varied DNA-binding proteins and located far from the enzymatic active site with no apparent function. For BER proteins, [4Fe4S] clusters were first thought to play a structural role. When not bound to DNA, these proteins are found in the $[4Fe4S]^{2+}$ state and are not easily oxidized or reduced under physiological conditions (*9*). However, for MutY and EndoIII, we have demonstrated using DNA-modified electrodes that DNA binding shifts

the 3+/2+ cluster potential into a physiological range, ~ 100 mV vs. NHE for each BER enzyme (*10, 11*); DNA binding stabilizes the protein in the +3 form.

Given the sensitivity of DNA-mediated charge transport (CT) to mismatched and damaged bases, we have proposed that DNA repair glycosylases containing a redox-active [4Fe4S] cluster including EndoIII and MutY, use DNA CT as the first step in substrate detection by signaling one another to search cooperatively for damage in the genome (*10, 11*). DNA-mediated CT can proceed over long molecular distances on a short time scale (*12*). Oxidative damage to DNA has been demonstrated with oxidants covalently tethered and spatially separated from damage sites at distances of > 200 Å with negligible loss in efficiency (*13*). Reductive CT has been shown to have an equally shallow distance dependence both in electrochemical studies (*14*) and in assemblies in solution (*15*). Previous studies establish that CT through DNA is possible in biological environments that include nucleosomes (*16*) and cell nuclei (*17*). DNA CT is, however, extremely sensitive to perturbations in the intervening base pair stack, such as DNA mismatches and lesions (*18, 19*). As an example, a single molecule of DNA covalently attached within a nanotube device can conduct charge perpendicular to the π-stack similarly to graphite, but the resistance increases 300-fold with a single base mismatch (*20*). DNA-mediated electrochemistry has therefore been utilized in the development of sensors for mutational analysis (*19*) and protein binding (*21*).

Given that this chemistry occurs at a distance and is modulated by the structural integrity of the base pair stack, these reactions may be useful within the cell for long-range signaling to proteins. In that context, we have previously established the long-range oxidation of the DNA-bound BER enzymes in spectroscopic studies monitoring oxidation of the [4Fe4S] clusters by guanine radicals in the duplex (*22*). Importantly, we have also shown the injection of an electron into the base pair stack from the DNA-bound BER enzymes, with the electron trapped by a well-coupled modified base in the duplex (*23*). Both with respect to hole injection into the DNA-bound proteins and

electron injection into the DNA from the DNA-bound proteins, EndoIII and MutY behave equivalently, as expected given their similar DNA-bound redox potentials and structures. Here we now explore whether DNA-mediated CT may provide a means to facilitate the detection of damage *in vivo*.

Materials and Methods All chemicals were purchased from Sigma Aldrich. All enzymes were purchased from New England Biolabs unless otherwise specified. All buffers were freshly prepared and filtered prior to use. Mica surfaces were purchased from SPI supplies. Silicon AFM probes were purchased from Nanoscience Instruments. Oligonucleotides were purchased from IDT or synthesized on a 3400 DNA synthesizer (Applied Biosystems). All strains used were derivatives of CC104 or CC102 (*24*) and generated as described below. Luria-Bertani (LB) broth was used as the rich medium while NCE (*25*) medium supplemented with $MgSO_4$ (100 μM) and glucose (11 mM) or lactose (6 mM) was used as the minimal medium.

Genome scanning calculations The time for sliding to scan the genome is calculated from a one-dimensional random walk (*26*). The one-dimensional diffusion (sliding) constant has been measured (*27*) *in vitro* for the DNA repair proteins hOGG1 and *Bacillus stearothermophilus* MutM as $5x10^6$ bp^2/sec and $3.5x10^5$ bp^2/sec, respectively. Taking the more generous (faster) value for the diffusion constant, a genome size of $5x10^6$ base pairs, and the MutY copy number of 30 yields a scanning time of 46 minutes. Short hops have been found (*26*) to have an insubstantial effect on the total search time. This result is consistent with the analysis of Mirny et al. (*28*). The tight 10^{-7} K_d of MutY for undamaged DNA 30mers corresponds to greater than 99.99% bound protein *in vivo*, inconsistent with facile three-dimensional translocation (*29*).

We can calculate the step time for three-dimensional diffusion of the reduced protein to the DNA as

$$t = \frac{V}{k_a C_p (1 - \Theta)}$$

where V is the cell volume, C_p(1-Θ) is the number of reduced repair proteins in the volume (Θ = fraction oxidized), and k_a is the bimolecular rate constant for protein association with the DNA target within the cellular volume. The bimolecular rate constant can be determined using a modified Smoluchowski equation for protein collision with a rod of DNA within the cell volume, where the length of the rod is twice the number of bases, *N*, over which DNA-mediated CT can proceed, since reduced protein can transfer an electron from either side. Any contact of the reduced protein within the DNA rod allows electron transfer to the DNA-bound oxidized protein. We assume that DNA is a rod over a region 2N base pairs long, but clearly the organization of the bacterial nucleoid is more complex (*30*). The Smoluchowski equation is constructed with two terms: one describes the ballistic 3-dimensional diffusion of the reduced protein to the DNA and the second (*31*) considers the gyrations of a rod with persistence length of 150 base pairs and the ends fixed as part of the chromosome.

$$k_a = 4\pi\kappa f D_p (r_{DNA} + r_p) + D_a a \sqrt[3]{\frac{r_{DNA} + r_p}{a}}$$

The protein diffusion constant (D_p) is determined from the Stokes-Einstein equation using the 10 cP viscosity of *E. coli* cytoplasm (*32*) and the measured Stokes radius of EndoIII (r_p) (*33*). DNA diffusion is considered to be negligible. The persistence length of DNA (150 bp) is defined as *a*. The electrostatic (*f*) and orientational (κ) constants are taken as unity (*34*), in keeping with the high ionic strength *in vivo* environment. The dissociation rate of the protein is not included in our model because charge equilibration should occur on a much faster time scale than dissociation of the reduced protein. Note that the reduced protein is not allowed here to exploit facilitated search to colocalize with the oxidized protein. We also make no distinction between 5′ to 3′ versus 3′ to 5′

transport, although subtle differences have been observed (*35*). We can calculate the overall time to search the genome of *Z* bases through the random walk of the electron as

$$T_{ox} = t\left(\frac{4Z}{N}\right)^2$$

Since there is equal probability of reduced protein associating with each base in the CT-active target region, and N represents that maximum distance of interprotein charge transfer, the average distance of interprotein CT is N/2. Self-exchange decreases the average step distance by a further factor of 2. This represents the scanning time for a single oxidized protein. Each oxidized protein provides a separate nucleation site for CT scanning, but draws from the same reservoir of reduced protein to scan different portions of the genome,

$$T_{CT} = t\left(\frac{4Z}{\Theta C_p N}\right)^2$$

Scanning through sliding without a CT search represents a boundary condition, so that the total time is:

$$T = (T_{CT}^{-1} + T_D^{-1})^{-1}$$

where T_D is the diffusion scanning time.

Designing and constructing a labeled 100-mer DNA duplex with a single mismatch

Using the Glen Research Cy-3 phosphoramidite to label the 5′ end of a single strand, standard phosphoramidite chemistry was used to make the oligonucleotide, and this was then detritylated. The oligonucleotide was HPLC-purified on a semi-preparative C18 column after deprotection.

Separately, the Alexa Fluor 555 succinimidyl ester amine-reactive dye from Invitrogen (Molecular Probes) was used in a solution-phase coupling procedure. Synthetic DNA was attached at the 3′ end to a resin and bears a free hydroxyl group at the 5′ end. The 5′-OH was allowed to

react with 63 mg (389 μM) carbonyl diimidazole in 1 mL of dioxane for 50 minutes. A methylene linker was attached via the addition of 45 mg (425 μM) diaminohexane in 1 mL of dioxane for 25 minutes (*36*). The amino-modified DNA was removed from the resin by treatment with NH_4OH at 60°C for 12 hours and HPLC-purified on a semi-preparative C18 column. The amino-modified DNA (427 μM) was dissolved in 0.1 M sodium tetraborate pH 8.9 buffer producing a final buffer concentration of 8 mM. 1 mg of the Alexa Fluor 555 dye was dissolved in DMSO, and the DNA solution was added in a total reaction volume of 10 μL. This was stirred at 37°C for 1 hour and stopped via ethanol precipitation (10% 3 M NaCl, 3x volume EtOH). The modified oligonucleotide was HPLC-purified on a semi-preparative C18 column.

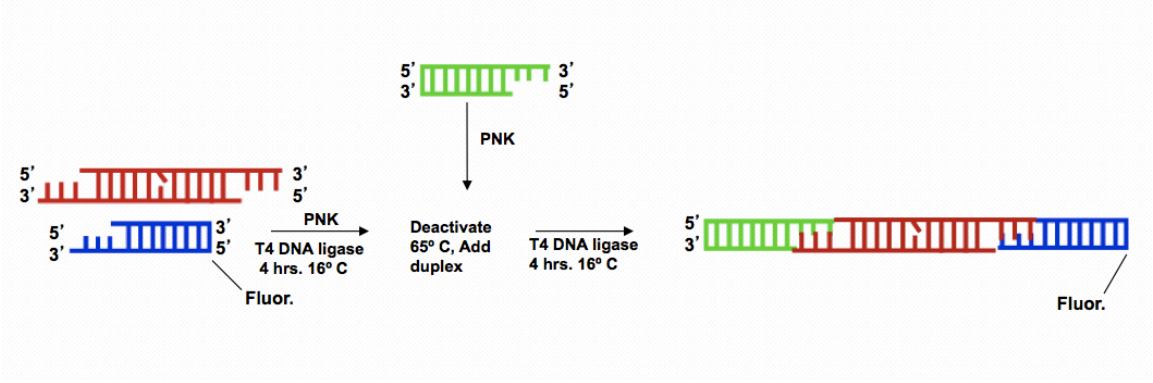
For tethering, a biotin phosphoramidite was used to label the 5′ end of a single strand, and standard phosphoramidite synthesis was used to make an oligonucleotide. The strand was subsequently HPLC-purified on a semi-preparative C18 column before and after trityl removal and deprotection.

DNA Stepwise Ligation/Purification: Two previously annealed duplexes (5 μg each) each containing a biotin and dye modification as described above were treated with 50 U polynucleotide kinase (PNK, New England Biolabs) in 10% T4 ligase buffer (New England Biolabs) in a 20 μl total volume and heated at 37°C for 1 hour. 10 U of T4 DNA ligase (Roche) was added to the reaction and this was held at 16°C for 4 hours and deactivated at 65°C for 10 minutes upon completion. 5 μg of a third duplex was treated with 50 U PNK in 10% ligation buffer for 1 hour at 37°C, and this was then added to the initial reaction mixture for a total volume of 50 μL after the addition of 10 U T4 DNA ligase and 10% T4 DNA ligase buffer. This was then held at 16°C for 4 hours, ethanol precipitated, and resuspended in 100 μL 50mM NaCl/ 5 mM PBS buffer (Scheme 2.1). The resulting duplexes were then visualized and purified on a 12.5% native acrylamide gel.

Scheme 2.1. (Top) Schematic of six single oligonucleotide strands (Top to Bottom: A1, A2, B1, B2, M1, and M2) that are then annealed to form three duplexes (A, B, M) with 4 restriction sites incorporated (BamHI, HindIII, EcoRI, and SalI) that can then be ligated to form the 100mer (with or without the C:A mismatch) shown at the bottom. Stepwise strategy for ligation of three duplexes (represented in red with/without mismatch, blue, and green). (Bottom) Two duplexes, one containing the fluorescent label, are phosphorylated and ligated followed by deactivation and addition of the third (green) phosphorylated duplex. More T4 DNA ligase is then added and the reaction is run for 4 hours at 16°C to form a 100mer with a site-specific mismatch.

```
A: 5'- GCA TCC TGT ACC GGA CTC GAC GAG GAT CCG CAT CCT GTA C-3'
       CGT AGG ACA TGG CCT GAG CTG CTC CTA GGC GTA GGA CAT GCG CTT A-5'

B:  5'- TCG ACG GCA TCC TGT AGG CCA GTG CAA GCT TGC A-3'
            3'-CGT AGG ACA TCC GGT CAC GTT CGA ACG T-5'

A: BamHI (40)mer
B: HindIII (28)mer


M: 5'-GCG AAT GCG AGT ACA GTC ATC GCG AG-3'
          3'-CGC TCA TGT CAA TAG CGC TCA GCT GC-5'
M: EcoRI
   SalI

20+12=32mer

32+40+28=100mer
```

```
5'-GCA TCC TGT ACC GGA CTC GAC GAG GAT CCG CAT CCT GTA CGC GAA TGC GAG TAC AGT CAT CGC GAG TCG ACG GCA TCC TGT AGG CCA GTG CAA GCT TGC A-3'
3'-CGT AGG ACA TGG CCT GAG CTG CTC CTA GGC GTA GGA CAT GCG CTT ACG CTC ATG TCA ATA GCG CTC AGC TGC CGT AGG ACA TCC GGT CAC GTT CGA ACG T-5'
```



Generation of 3.8 kbp DNA strands with a C:A mismatch for AFM studies

Four primers were synthesized with the following sequences,

5′-GTACAGAGTTCAGTCGGCATCCGCTTACAGACAAGC-3′ (forward),

5′-CCGGTAACTATCGTCTTGAGTCC-3′ (reverse),

5′-GACTGAACTCTGTACCTGGCACGACAGGTTTCCCG-3′ (forward), and

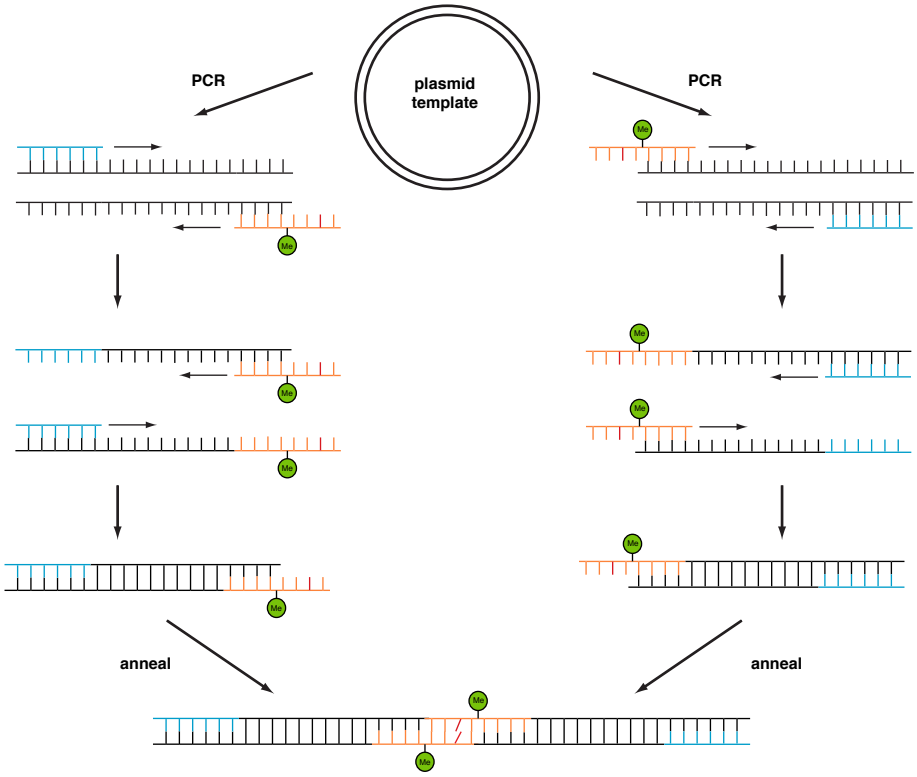
5′-GACTGAACTCTATACCTGGCACGACAGGTTTCCCG-3′ (forward). The underlined bases highlight the location of a 2′-*O*-methyl residue (*37*). These primers were used in separate PCR reactions using pUC19 as a template to generate two duplexes 1610 bp and 2157 bp (matched), each containing one 14 nt single-strand overhang. These PCR products were purified by ethanol precipitation and resuspended in 50 mM NaCl, 5 mM sodium phosphate, pH 7 buffer and quantitated by OD_{260}. Duplexes were phosphorylated using 100 U PNK in 10% T4 DNA ligase buffer for 1 hour at 37°C and deactivated for 10 minutes at 65°C. Separate duplexes were annealed at 65°C for 8 minutes then cooled to 20°C over 2 hours to form a 3767 bp (mismatched) long strand. The resulting larger duplexes were ethanol precipitated and resuspended in 100 μL 50 mM NaCl/ 5 mM phosphate buffer. 15 U T4 DNA ligase and 10% T4 ligase buffer were added (total reaction volume ~ 150 μL) and incubated overnight at 16°C followed by deactivation for 10 minutes at 65°C. We did not bring the ligation reaction to completion, so as to obtain a mixture of DNA samples that were equivalent other than the presence of the mismatch at the ligation site. The DNA duplexes (ligated and unligated) were then analyzed by 0.6% agarose gel electrophoresis. The strategy for preparation of samples is given in Scheme 2.2.

AFM deposition conditions To optimize deposition, pUC19 plasmid DNA was digested by *EcoRI*, *ScaI*, (New England Bioloabs) and *SalI* (Roche) to produce 903 bp and 1750 bp fragments. 5 μL of a plasmid DNA (2 μg/mL)/6 mM $MgCl_2$ solution was deposited onto a cleaved mica surface for 1

minute and rinsed at a 45 degree angle with 2 mL water. This was then dried with argon. For redistribution data, mica surfaces were freshly cleaved with scotch tape. EndoIII (0.4 μM) was added to the stock DNA solution containing 50-100 ng total DNA composed of the mixture of ligated 3.8 kb duplexes and the two unligated duplexes (1.6 and 2.2 kb) in 6 mM $MgCl_2$ /Tris-EDTA buffer. This protein-DNA solution was incubated at 4°C overnight and deposited (5 μL) on the mica surface for 2 minutes, rinsed with 2 mL water, and dried under argon. In order to study the effects of protein oxidation, protein-DNA solutions were treated with 1 or 5 mM H_2O_2, incubated for 2 minutes, and deposited onto the mica surface as described. Two of the five wild-type EndoIII samples analyzed were deposited onto the mica surface in 5 ng quantities followed by immediate addition of 3.7 μM EndoIII (1 μL volume). This difference in deposition conditions did not result in any significant differences in the trends observed. Mutant protein (Y82A) was added to a stock solution of 50 ng DNA for a final protein concentration of 0.4 μM. After incubation at 4°C overnight, deposition conditions were identical to that for wt EndoIII-DNA samples. For each AFM experiment, at least 6 images and > 200 strands were counted using several preparations of protein/DNA samples.

AFM instrument setup Silicon AFM Probes purchased from Nanoscience Instruments (BudgetSensors), with a spring constant of 3 N/m and a resonance frequency of 75 kHz, were used in a Digital Instruments Multimode SPM. Images with scan areas of 2 x 2 μm^2 or 1 x 1 μm^2 were acquired in tapping mode, using an amplitude of 0.54-2.00 V at a scan rate of 3.05 Hz. Scan rates of 3.05 Hz were used to obtain images of higher quality. Data analysis was performed using the WSxM program (*38*).

Scheme 2.2. Two synthesized primers, one incorporating a 2′-*O*-methyl ribonucleotide (blue and orange, respectively), are utilized to amplify plasmid sequences via PCR. A nucleotide is incorporated within the synthesized primer containing the 2′-*O*-methyl ribonucleotide (orange) that will result in a mismatched or matched final product. As the vector is amplified, polymerase stops at the ribonucleotide (green), resulting in a large single-strand overhang. Upon performing two separate PCR reactions with two primers each (orange and blue), the resulting duplexes with 14 bp single-strand overhangs can be annealed/ligated together in good yield to produce long duplexes with or without a mismatch.



Strain construction CC104 and CC102 strains were generously donated (*24*), as was CC104 *mutY*- (CC104 *muty::*mini-tn 10) (*39*). *Nth* was replaced by a chloramphenicol resistance cassette (*cm*) in CC104 and CC104 *mutY*- using a previously described in-frame deletion method (*40*). See Table 2.1 for primer sequences; *nth* homology regions are shown in regular text and *cm* priming regions are highlighted in boldface. CC102 strains were constructed using P1 transduction (*41*). Inactivation in all strains was verified with colony PCR.

Lac$^+$ reversion assays Strains were streaked to LB medium and incubated overnight at 37°C. For *nth* knockouts, strains were streaked to LB+chloramphenicol (17 μg/mL), and for *mutY* knockouts, strains were streaked to LB+tetracycline (50 μg/mL). 1 mL LB cultures were started from single colonies and grown overnight in a shaking incubator at 37°C, 220 rpm. 20 μL of each starter culture was used to inoculate a 10 mL NCE+glucose culture which was then grown to a density of 10^9 cells/mL at 37°C, 250 rpm. Cell density was determined by dilution plating a 10 μL aliquot of the NCE+glucose culture onto NCE+glucose solid medium followed by incubation at 37°C for 36 hours. 5 mLs of this culture was centrifuged in a clinical tabletop centrifuge at 4°C and plated on NCE+lactose solid medium and then incubated at 37°C for 36 hours. Colonies arising are reported as *lac*$^+$ revertants/mL cells plated. In experiments incorporating plasmid vectors, CC104 and CC102 strains were made electrocompetent (*41*) and transformed via electroporation at 1.8 kV. Transformants were selected on LB+ampicillin (amp) (100 μg/mL) solid medium after incubation overnight at 37°C. Single colonies were restreaked to LB+amp (100 μg/mL) solid media and incubated 12 hours at 37°C. 1 mL LB+amp (100 μg/mL) cultures were started from these colonies and grown overnight at 37°C, 220 rpm. 20 μL of this starter culture was used to inoculate a 10 mL NCE+glucose+amp (40 μg/mL) culture which was grown to a density of 10^9 cells/mL at 37°C, 250 rpm. 5 mLs of this culture was plated onto NCE+lactose+amp (40 μg/mL) and incubated at 37°C for 36 hours. *Lac*$^+$ revertants were reported as described above.

Table 2.1. Primer Sequences Used for nth Inactivation, Cloning, and Mutagenesis

Function	Primer Sequence
Wanner inactivation of *nth*	5'-gaagcagctgcagaacgtgcattgccaaacggtgaaacagggaatgtctg**gtgtaggctggagctgcttc**-3'
	5'-agaggataaagaaaggttatcaatggggtaatcggtgttacccctttttct**catatgaatatcctccttag**-3'
Cloning *nth* into pBBR1MCS-4	5'-**ggaattc**gcaatggcacattgtttgac-3'
	5'-**aggttcc**tcaatggggtaatcggtgtt-3'
D138A mutant generation	5'-cggctggccgactattgctgtc**gcc**acgcacattttccgcg-3' + complement
Y82A mutant generation	5'-ggggtgaaaacctatatcaaaactattgggcttgctaacagcaaagc-3' + complement

Mutagenesis *Nth* was cloned into pBBR1MCS-4 (p) using standard techniques (*42*). Primer sequences are given in Table 2.1; restriction sites are highlighted in boldface. Gene incorporation was verified by PCR and sequencing (Laragen) using KS/SK (pBBR1MCS-4 derived vectors) or M13 (pNTH10 derived vectors) sequencing primers. Mutants were generated using a QuikChange site-directed mutagenesis kit (Stratagene) and verified by 3 independent sequencing reactions (Laragen). Primers used are shown in Table 2.1; the altered codon site is highlighted in boldface.

Protein expression and purification EndoIII and Y82A EndoIII were expressed from the pNTH10 expression vector and purified as described previously (*43*). Protein concentrations were determined using the UV-vis absorbance of the [4Fe4S] cluster (410 nm, ε = 17,000 $M^{-1}cm^{-1}$) (*9*).

Glycosylase assays Oligonucleotides were synthesized containing a 5-OH-dU lesion site (purchased as a phosphoramidite from Glen Research), HPLC purified and verified with MALDI-TOF mass spectrometry. The following sequences were used:

5′-TGTCAATAGCAAGXGGAGAAGTCAATCGTGAGTCT-3′ + complementary strand where X = 5-OH-dU base-paired with G. The strand containing the lesion was 5′-^{32}P endlabelled as previously described (*42*) and annealed to its complement. Reactions were run at 37°C for 15 minutes with 10 nM DNA and 100 or 10 nM protein and quenched with 1M NaOH (*44*). Samples were analyzed by denaturing 20% PAGE and imaged by autoradiography. Band intensity was quantified using ImageQuant software.

Electrochemistry at DNA-modified electrodes DNA-modified electrodes for protein electrochemistry experiments were prepared as described (*11*) using the following sequences: thiol-modified strand, 5′- AGTACAGTCATCGCG-3′, TA complementary strand, 5′-

CGCGATGACTGTACT-3′, and abasic site (Ab) complementary strand, 5′-CGCGATGACTGTXCT-3′, where X = dSpacer (Glen Research). Surfaces were backfilled with 1 mM mercaptohexanol for 30 minutes and rinsed at least 3 times with protein storage buffer (20 mM sodium phosphate, 100 mM NaCl, 0.5 mM EDTA, 20% glycerol, pH 7.5). Protein solution was introduced to the electrode surface and allowed to incubate for ~ 20 minutes until signal reached full intensity. Cyclic voltammetry experiments were performed on a CH Instruments 760 potentiostat using a 50 mV/s scan rate, Ag/AgCl reference electrode, and Pt wire auxiliary electrode in an electrochemical cell modified for protein experiments (*11*).

Results and Discussion

A CT model to detect DNA damage We have proposed that BER proteins bearing [4Fe4S] clusters exploit DNA-mediated CT as a fast, sensitive method to detect damage (Figure 2.1). This redox signaling model is initiated when one 2+ protein (donor) binds DNA (b, e), promoting electron transfer from the donor protein to a distal protein (acceptor) (c, f), already bound in the 3+ state. The newly oxidized donor protein remains DNA-bound while the reduced acceptor diffuses away (d, f). Integral to this model is a differential DNA affinity for the $[4Fe4S]^{3+}$ and $[4Fe4S]^{2+}$ forms of the protein. We have demonstrated this differential affinity by measuring a -200 mV potential shift associated with DNA binding that corresponds thermodynamically to ≥1000-fold difference in DNA affinity between the oxidized and reduced proteins (*45*).

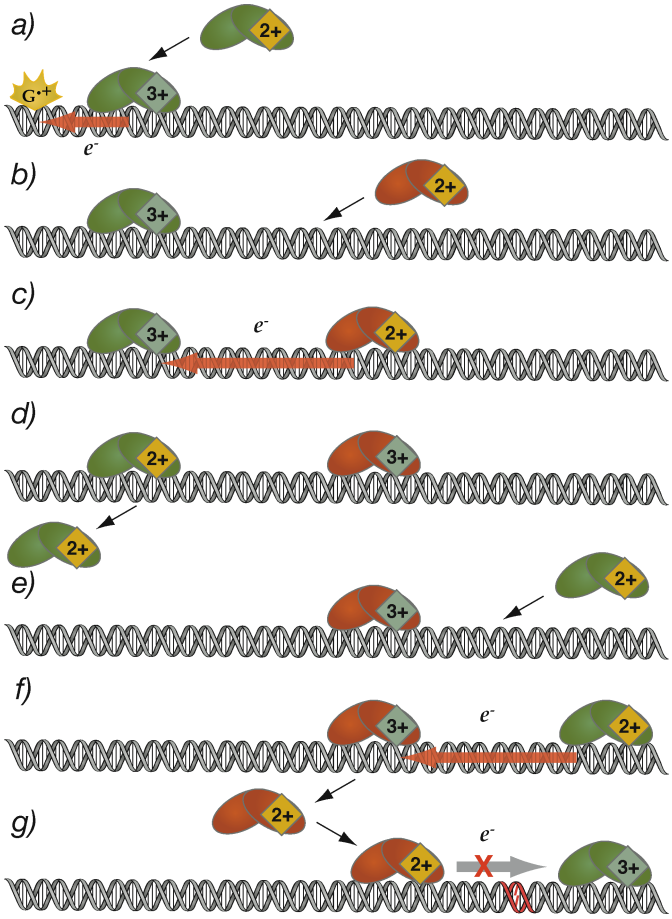
Importantly, the DNA-mediated CT reaction between two repair proteins can be considered a scan of the integrity of the intervening DNA, since DNA-mediated CT can only proceed through a well-stacked duplex. As illustrated in Figure 2.1 (g), when the repair protein, already oxidized, is bound near a base lesion, DNA-mediated CT does not provide a pathway for reduction and subsequent protein dissociation. The protein instead remains bound to the duplex so that, on a slower time scale,

the protein can processively diffuse to the target site; now, however, sliding is needed only across a small region and the low target specificity of the protein is sufficient for recognition (*3, 46, 47*). Essentially, then, our proposal for base lesion detection utilizing DNA CT yields a redistribution of the BER enzymes onto local regions of the genome that contain lesions. Critical to this mechanism is DNA-mediated signaling among proteins bound at long range so that the proteins, despite their low abundance, cooperate with one another in localizing onto target sites.

To exploit DNA-mediated CT, some proteins must exist in the oxidized state. There are many oxidants in the cellular milieu, and the level of oxidative stress will govern the proportion of oxidized protein. Indeed, we have shown that these proteins (*22*) and others (*48*) can be oxidized by guanine radicals, the first genomic signal of oxidative stress (*49*), via DNA-mediated CT. A basic model of genome scanning involving only facilitated diffusion, without CT, but where interrogation is assumed to be instantaneous, yields a genome scanning time of at least 46 minutes for MutY (Methods), wholly insufficient given the doubling time in *E. coli* of 20 minutes (*41*). This calculated scanning time is based upon measured values of protein diffusion constants (*7*), intracellular protein concentration (*1*), and genome size using a model of one-dimensional diffusion with short localized hops (*7*). Recent calculations of global searches by low copy number proteins suggest a still slower search time (*28*), and the nanomolar dissociation constant excludes three-dimensional mechanisms. This estimate significantly understates the problem, moreover, since the actual interrogation time is not instantaneous and protein traffic on the DNA necessarily interferes with sliding.

In our model, the DNA is essentially scanned by the electron with the repair proteins facilitating electron migration. Thus we calculate a genome scanning time for MutY in *E. coli* that is significantly more efficient through DNA CT. Since an injected charge equilibrates on the nanosecond timescale (*12*), and protein diffusion occurs in micro- to milliseconds (*7*), the rate-limiting step in this process is the 3D diffusion of this reduced protein within CT range of the oxidized DNA-bound protein.

Figure 2.1. A model for DNA-mediated CT in DNA repair. In this model, DNA repair proteins, containing $[4Fe4S]^{2+}$ clusters, for example, EndoIII (green) and MutY (orange), bind DNA, activating them towards oxidation to the $[4Fe4S]^{3+}$ state. The sequence of events is as follows: Guanine radical formation can oxidize a repair protein in a DNA-mediated reaction, stabilizing the oxidized protein bound to DNA (*a*). A second protein binds in the vicinity of the first protein (*b, e*). CT to a distally bound protein can occur through the DNA p-stack if the intervening DNA is undamaged (*c, f*). The newly reduced protein has a diminished affinity for DNA and diffuses away (*d*). If, instead, a lesion site is present between the proteins (*g*), the DNA-mediated CT step is inhibited and the oxidized protein remains bound to DNA. In this search mechanism the sum of the DNA-mediated electron transfer steps between proteins constitutes a full search of the genome yielding a redistribution of low abundance DNA repair proteins in the vicinity of lesions.



Hence scanning can be modeled as a random walk of the electron (or hole) on the DNA, where the step time for the walk is the average time for a reduced protein to approach within range to carry out DNA-mediated CT to the oxidized protein. In our calculation, we conservatively utilize 3D diffusion of the reduced protein within the cellular volume to approach its target site on the DNA. We also use experimental values for protein concentrations (*1*), genome size, and protein binding affinities (*2*).

Importantly, since this model involves cooperation among the repair proteins, we can utilize the *total* concentration of these proteins within the cell, rather than copy numbers for MutY or EndoIII individually. Thus MutY, present in $\leq$ 30 copies, benefits from 500 copies of EndoIII (*1*). We do, however, neglect contributions from any other proteins that might participate in DNA-mediated signaling; other DNA-bound proteins containing iron-sulfur clusters exhibit similar potentials and CT reactions involving these proteins too would substantially speed the search process.

Our model relies on the fact that DNA-mediated interprotein CT is much faster than protein diffusion, and that the oxidized repair proteins have higher nonspecific DNA affinity than the reduced proteins; both assumptions have experimental support (*12*, *45*). As we did for facilitated diffusion, we assume here also that intervening DNA-binding proteins do not inhibit scanning. In fact, one advantage of DNA CT over other search mechanisms is that the electron travels *through* the DNA base pairs and no proteins need to be displaced (*16*).

Figure 2.2 shows how the interrogation time varies as a function of N, the maximum distance over which DNA-mediated CT proceeds, and ox, the percentage of proteins oxidized. Remarkably, permitting DNA CT over 500 bp with 10% oxidized protein yields a conservative interrogation time of 5 minutes, while DNA CT over only 200 bp with 20% oxidized protein yields an interrogation time of 8 minutes; permitting DNA CT over several hundred bp yields scan times less than a minute. These values are well within the 20 minute doubling time of *E. coli* (Figure 2.3).

Figure 2.2. Scanning time as a function of maximum distance of DNA-mediated interprotein CT (N) and the fraction of repair proteins that are in the 3+ state (% ox) using the CT scanning model. At 10% oxidized protein with a maximum CT distance of 500 bp, the time required to interrogate the genome is ~ 5 minutes.

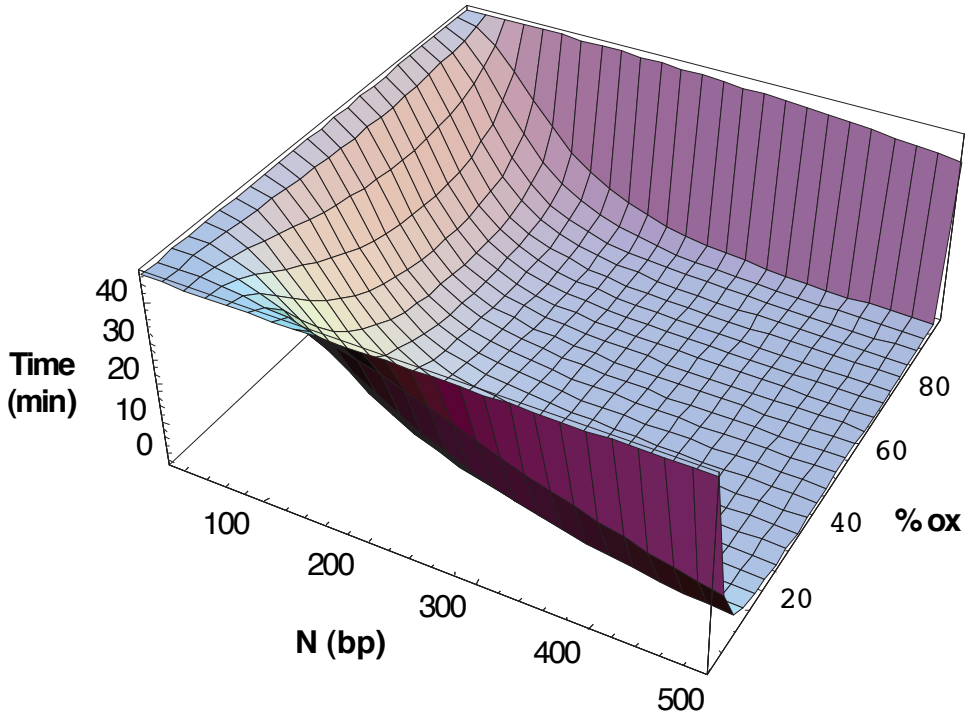
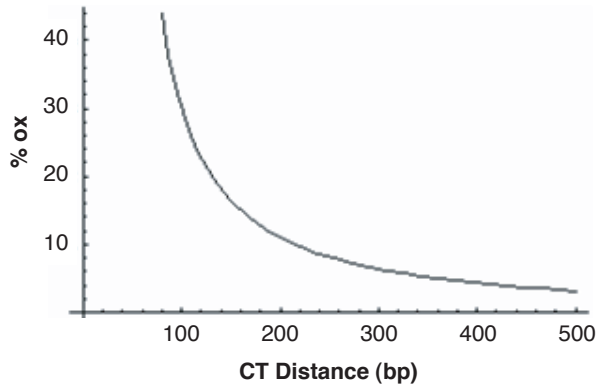


Figure 2.3. Plot showing the contour from Figure 2.2 that corresponds to a genome scanning time (T) of 20 minutes, the doubling time of an *E. coli* cell.



While we have not yet established the distance limits for DNA CT, we have demonstrated substantial oxidative damage in tethered DNA assemblies *in vitro* over 60 bp and in DNA within mitochondria over ~ 100 bp (*13*, *50*). The dependence of interrogation time on the percentage of proteins oxidized is also noteworthy (Figure 2.2). The scanning efficiency resembles a switch that is turned on at low levels of oxidation, when DNA repair is needed. Activation of this switch depends upon the redox buffering capacity of the cell and the level of oxidative stress.

An AFM assay to measure protein redistribution onto mismatched DNA While we have earlier carried out studies establishing hole and electron injection across the protein/DNA interface (*23*, *45*), our model also predicts that DNA/protein CT would promote the redistribution of repair proteins in the vicinity of base lesions or mismatches. We can assay for this redistribution by AFM. There are many challenges associated with making long strands of DNA containing a site-specific mismatch, and initially methods were chosen (ligation, PCR modular vectors) such that a variety of modifications could easily be made for further experiments. Strategically, it is important to find a consistent way of forming long duplexes. The duplexes that are ligated together must also contain single-stranded overhangs long enough to overcome blunt end or non-specific ligation. In preparation for AFM measurements, mismatched DNA has been synthesized and ligated with matched duplexes to yield a 100 bp helix with a C:A mismatch, which is not the primary substrate for MutY or EndoIII, but does significantly attenuate charge transport. The strand has been designed to incorporate at least four restriction sites that could later be used to introduce additional lesions (Scheme 2.1). As Scheme 2.1 shows, six strands of DNA were synthesized: A1, A2, B1, B2, M1, and M2 that were then annealed to form a 40mer (A), 28mer (B), and a 32mer (M), where M contains the C:A mismatch as well as restriction sites at the overhangs which were then ligated to duplexes A and B. A was constructed with a blunt end as well as a sticky end, and similarly, B was constructed the same way. M incorporates two sticky ends that, upon ligation, yield an A:M:B

100mer duplex configuration. Functionalizing the blunt ends of the strands with modifications such as biotin or dyes for use in AFM and other surface experiments.

Importantly, a stepwise strategy was developed to produce a higher yield of 100mer by ligating two of the duplexes, followed by the addition and ligation of the last one. Scheme 2.1 shows the method employed to construct the 100mer and isolate it in high yield. The final duplex was obtained by phosphorylating two of the duplexes and ligating them under previously described ligation conditions. Concurrently, the last duplex was treated with PNK and added to each of the reactions that had been used to assemble 60mers. Figure 2.4 shows the 100mer product, formed using the stepwise ligation strategy to attach one duplex containing a 5′ biotin tag, one containing a mismatch/match, and the other labeled with Cy-3 to be used for surface experiments.

AFM experiments, however, require very long strands of DNA (> 1,000 bp) to accurately observe protein distribution. As a control, restricted pUC19 plasmid pieces (1750 and 900 bp) were deposited onto a cleaved mica surface (0.01 μg) in deposition buffer and visualized on a 1 x 1μm^2 scale. Samples were diluted to yield uniform surface coverage. In Figure 2.5, strands are identifiable based on length and height. Notably, though some of the longer strands are curled, most are linear. Histograms for the images were made using a WSxM (*38*) program to quantitate the amount of DNA on the surface of different lengths. 900 bp strands are 300 nm long, and the 1750 bp strands are 578 nm.

Following the protocol described in the methods section, a mixture of DNAs, both long (3.8 kilobase) DNA duplexes containing a single C:A mismatch and short (2.2 and 1.6 kilobases) well-matched duplexes of the same total sequence were prepared (*37*) (Figure 2.6); the longer sequence was obtained by ligation of the two shorter sequences. This mixture of matched and mismatched DNA strands was incubated with EndoIII and examined using established AFM techniques (*51*) (Figure 2.7). Only clearly identifiable long or short strands were counted.

Figure 2.4. 12.5% acrylamide gel run at 12W for 1.5 hours showing in Lanes 1–6: mismatched 100mer (13 μM), matched 100mer (11 μM), three duplexes without ligase (A has biotin tag, B has Cy-3 tag, and M with/without mismatch), 5 μg of 10 bp marker, mismatched 60mer made from two duplexes, and matched 60mer made from two duplexes. Bands with Cy-3 appear pink over UV-light. Cy-3 absorbs at 546 nm and emits at 563 nm.

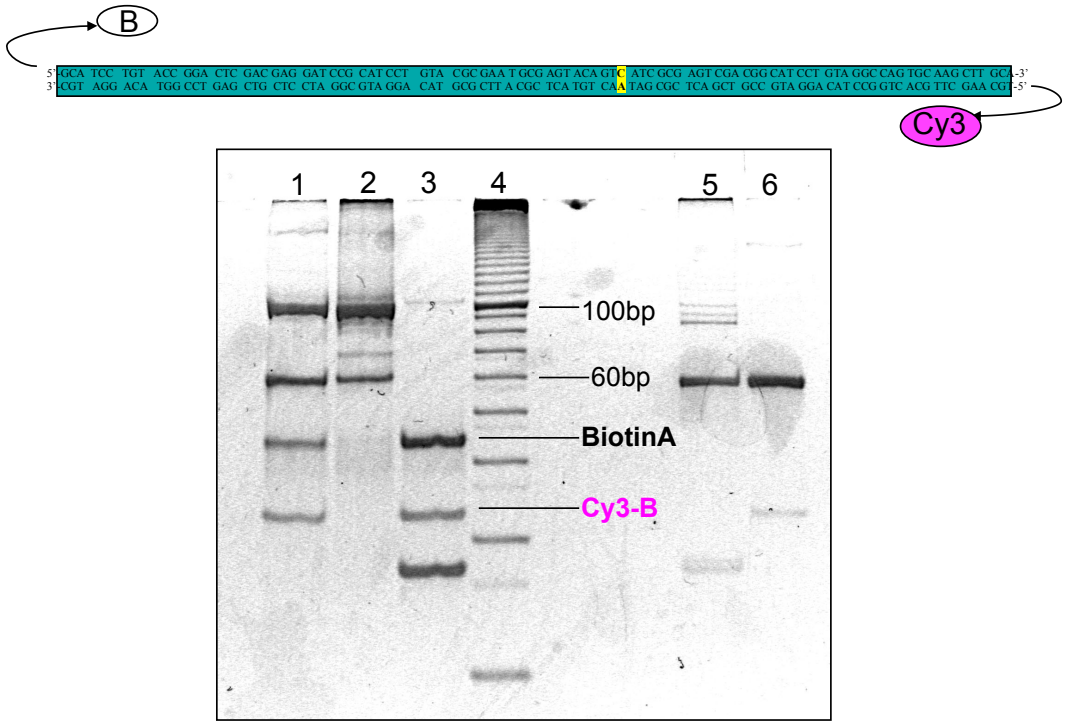


Figure 2.5. 0.6% agarose gel run at 12W for 1 hour. Lanes 1, 2: ~ 600 ng total of two duplexes amplified post-PCR and ligation to form matched strands ~ 3800 bp. Lane 3: 5 μg of Marker III (Roche), Lanes 4, 5: ~ 600 ng total of two duplexes amplified post-PCR and ligation to form mismatched strands

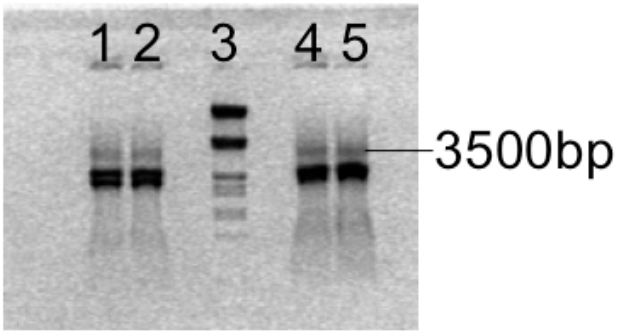


Figure 2.6. (Top left) A 2 x 2 μm^2 image using tapping mode AFM on a cleaved mica surface in air. DNA is a mixture of 1750 and 900 bp fragments from the restriction of pUC19 (0.01 μg) in 6 mM $MgCl_2$, Tris-EDTA deposition buffer. The inset shows a zoomed-in image to differentiate sizes of DNA. (Top right) Image of DNA at higher concentration (5 μL of a 2 μg/mL solution of DNA plasmid restriction pieces in 5 mM $MgCl_2$). (Bottom) DNA (1 μg/mL) of different lengths on ~ 3 x 3 μm^2 scale. Large blotches (white) are salt. The average height of DNA strands is 2 nm.

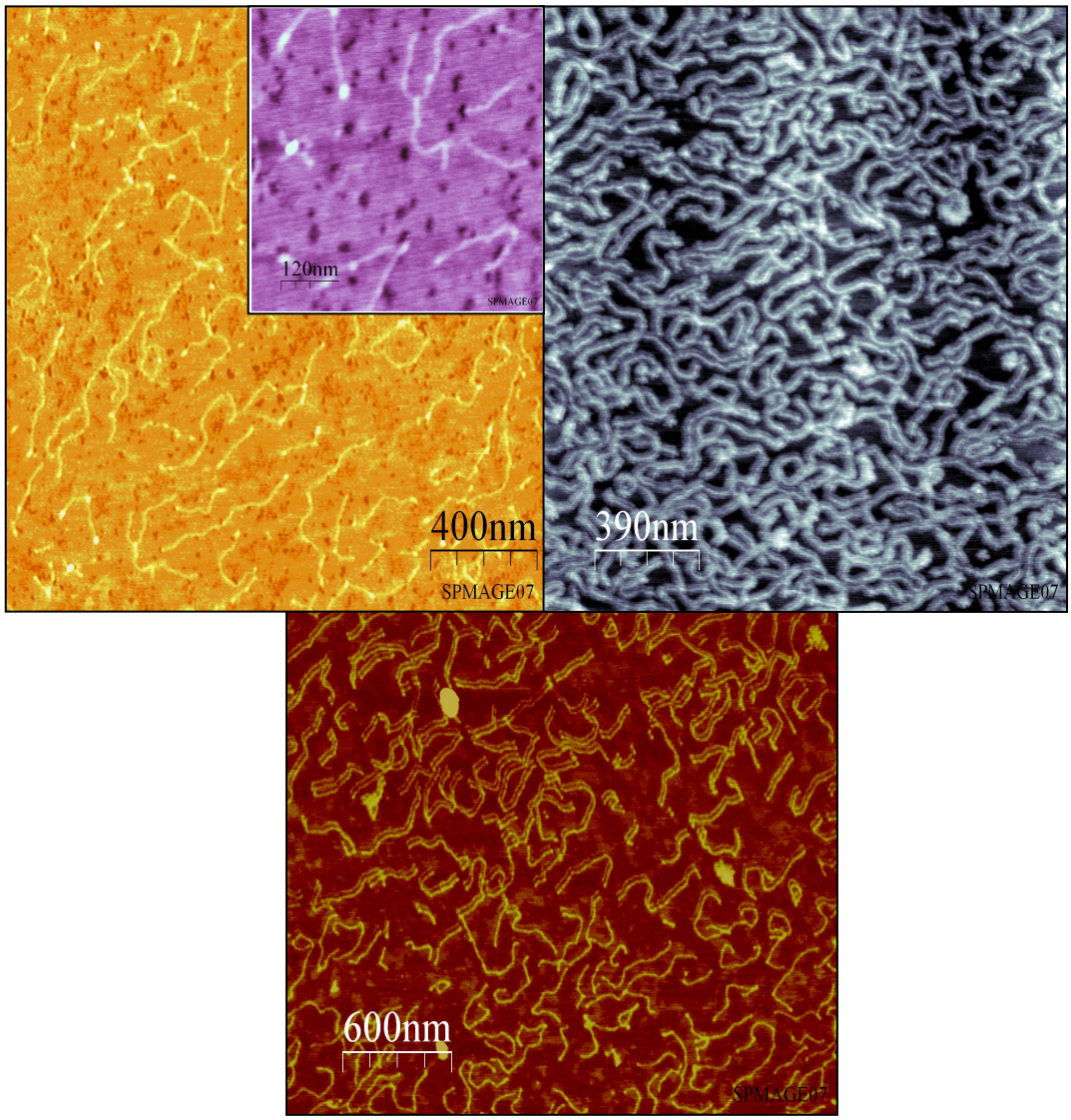
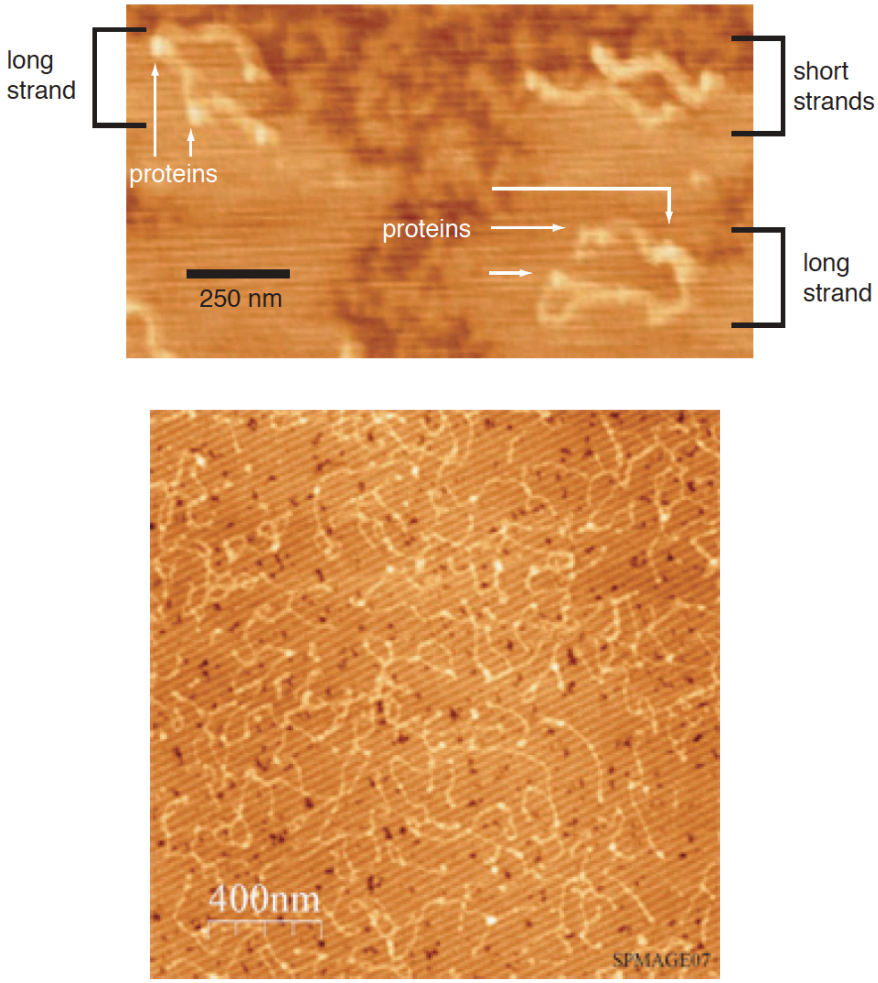


Figure 2.7. Measurements of repair protein distributions on DNA by AFM. A zoomed-in view (top) and a zoomed-out view (bottom) of representative AFM images of DNA strands incubated overnight with wild-type EndoIII. A higher density of proteins is apparent on the longer DNA strands containing the single base C:A mismatch.



Protein assignments were verified through analysis of their 4 nm heights in the images; without protein, features of this dimension are not observed and still larger heights indicate salt precipitates. Although a C:A mismatch effectively inhibits DNA CT (*12*), it is not a lesion that is preferentially bound by EndoIII; a gel shift assay on 21-mers with and without a central C:A mismatch shows no detectable difference in EndoIII binding. Thus without DNA CT between bound EndoIII molecules, one might expect an equal density of proteins on the short and long strands.

We find that EndoIII shows a significant preference for the longer strands containing the C:A mismatch (Figure 2.8). Examination of the number of proteins bound to 300 long strands and 465 short strands reveals a greater density of proteins bound to the long strand; r(long/short) is 1.6. If instead we examine the distribution of EndoIII on long versus short strands, where all strands are matched, we see a small preference for the short strands; the ratio of protein densities, r(long/short), is 0.9. When we calculate the strand preference based on DNA CT, this protein density ratio depends upon the DNA CT length and/or the length of the DNA over which protein can diffuse before dissociating. Using a signaling/sliding length of 90 base pairs, we calculate a protein density ratio of 1.6, that which we find, where half of the protein population is near the mismatch (Figure 2.9).

AFM measurements as a function of oxidation of proteins bound to DNA, using H_2O_2 as oxidant, reveal an additional increase in the ratio of EndoIII bound to mismatch-containing strands (Figure 2.10). Examination of more than 250 long C:A mismatch-containing strands and 300 shorter matched strands incubated with EndoIII and treated with 5 μM peroxide reveals a ratio of bound protein densities, r (long/short), of 2.4; when both long and short strands are matched, the ratio is 0.83.

Figure. 2.8. Tapping mode AFM images of DNA and EndoIII protein on mica imaged in air. DNA (strands) and protein (dots) are visible on the surface. Images were acquired with a scan size of 3 x 3 μm^2, at a rate of 3.05 Hz with a data scale of 10 nm. (Bottom) Zoomed-in images of a DNA strand with a bound protein subjected to height profiling to distinguish protein. The profile of the strand, to the right of the zoomed-in image, shows the height of DNA ranges from 1-2 nm, with a much higher (> 2 nm) protein peak.

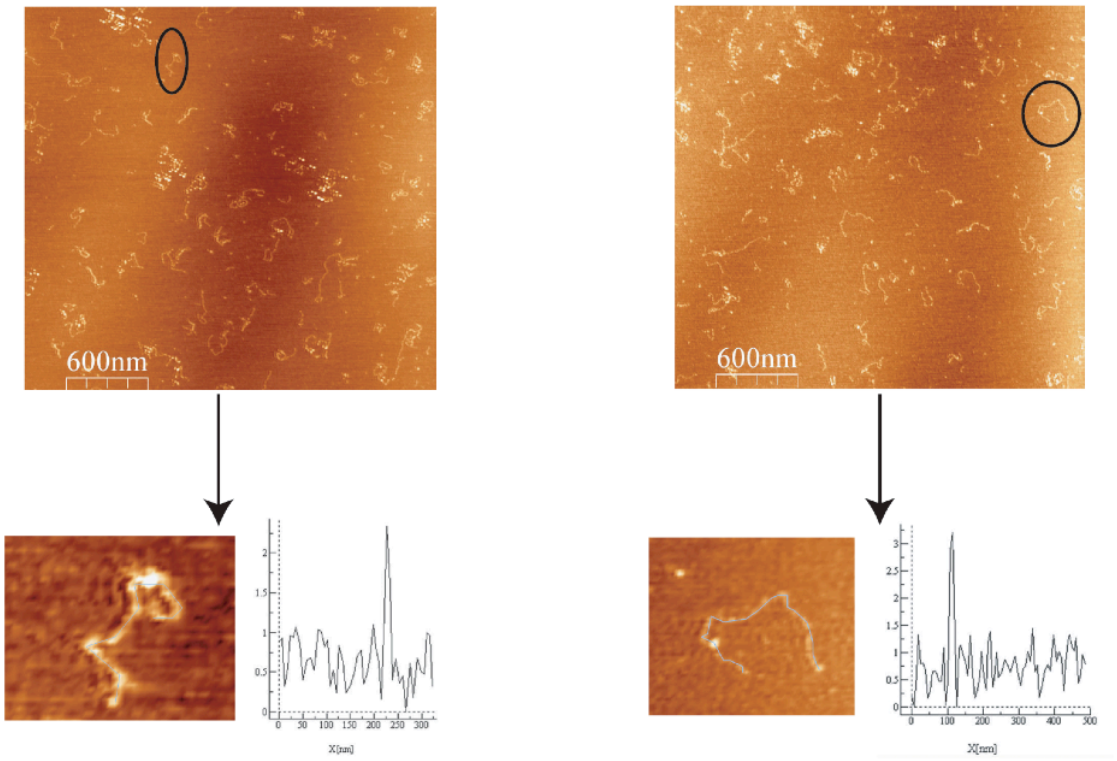


Figure 2.9. Quantitation of protein density ratios. A C:A mismatch is contained on the long strand except for the sample indicated by matched DNA where both the long and the short strands are fully matched. EndoIII redistributes onto the strand with the C:A mismatch.

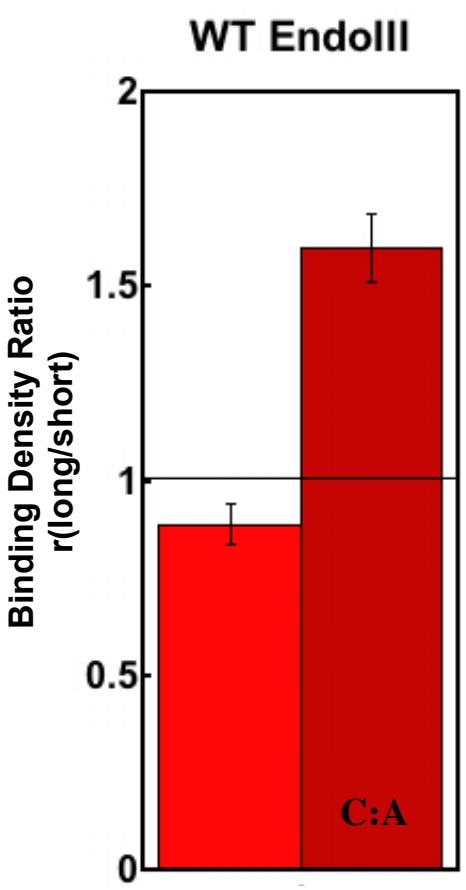
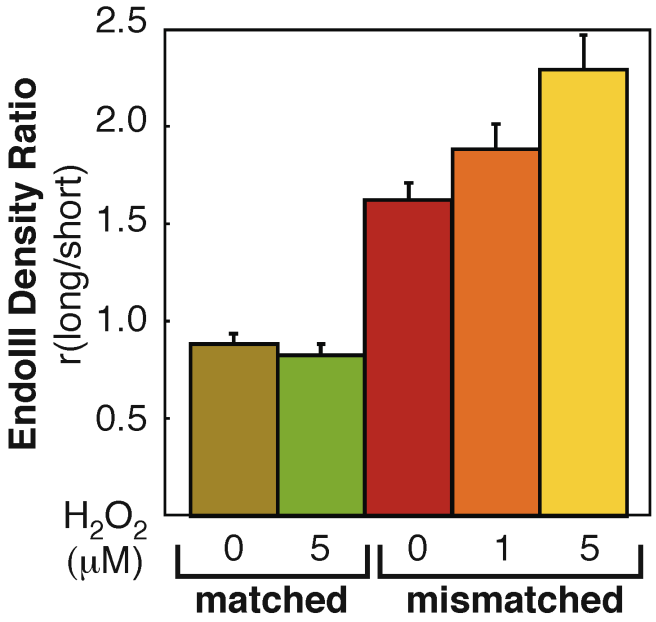


Figure 2.10. Quantitation of protein density ratios. A C:A mismatch is contained on the long strand except for the sample indicated by matched DNA where both the long and the short strands are fully matched. EndoIII redistributes onto the strand with the C:A mismatch and that preference is increased with increasing concentrations of peroxide.



These results are consistent with our model. DNA-mediated CT will drive the redistribution of repair proteins away from undamaged regions such that the proteins will cluster near damaged sites. As a result, we see the proteins redistribute preferentially onto the DNA strands containing the mismatch even though a C:A mismatch is not a substrate for EndoIII. Moreover, as predicted by the model, the redistribution of EndoIII is more pronounced in the presence of oxidative stress.

Cooperation between EndoIII and MutY inside the cell This CT scanning model can also be tested *in vivo* by assaying for the cooperation among repair proteins facilitated by DNA-mediated signaling. If these proteins are able to help each other in their search for damage using DNA CT, knocking out the gene for EndoIII or reducing its capability to carry out CT should lead to a decrease in MutY activity *in vivo*. Assays for MutY and EndoIII activity inside *E. coli* cells have already been developed (*24*). The assay for "helper function" used here employs engineered mutations in the *lacZ* gene to report the frequency of a particular base pair substitution. The strain that serves as an assay for MutY activity, CC104, substitutes a cytosine for an adenine in the *lacZ* Glu 461 codon, which is essential for β-galactosidase activity. Since MutY prevents GC to TA transversions (*52*), reversion of this original mutation back to wild-type (wt) *lacZ* reflects a deficiency in MutY activity. Analogously, the CC102 strain (*24*) serves as an assay for EndoIII activity by monitoring GC to AT transitions (*53, 54*).

In the CC104 MutY activity reporter strain (Table 2.2), 20 ± 9 *lac*+ revertants are observed per 10^9 cells, while inactivation of *mutY* in CC104 (CC104 *mutY*-) causes the number of *lac*+ revertants to increase 15x (300 ± 33) as expected (*24*, *52*). When the gene encoding EndoIII (*nth*) is inactivated in CC104 (CC104 *nth*-), the *lac*+ reversion frequency observed is 54 ± 5, more than a factor of two increase over CC104. Thus, loss of EndoIII does have a small but significant effect on MutY activity *in vivo*. This loss in activity is consistent with a loss in helper function by EndoIII, as predicted; the lower activity of MutY without EndoIII could reflect the lack of cooperative

searching via DNA CT. An alternative explanation, however, is that MutY and EndoIII share some overlapping ability to repair lesions. In this case, the *lac*+ reversion frequency of the CC104 *mutY-/nth-* strain (270 ± 29) should be greater than that of CC104 *mutY*, but they are, within error, equivalent.

This *in vivo* relationship between EndoIII and MutY has been observed previously, although in different experimental contexts. Small increases in mutational frequency have been detected when *mutY* is inactivated in CC102 (*24*) (Table 2.3) or when *nth* is inactivated in CC104 (*54*). In the latter case, it was proposed that this could be due to some intrinsic ability of EndoIII to repair oxidatively damaged guanine residues. Reported EndoIII repair activities do not prevent GC to TA transversion mutations (*55*) and, thus, are not relevant to the CC104 assay.

We can furthermore test directly whether the loss of MutY activity in the CC104 assay is the result of overlapping glycosylase activities by determining whether the number of *lac*+ revertants is still suppressed by an EndoIII mutant that is biochemically incompetent to carry out the glycosylase reaction. A mutant of EndoIII (D138A) that is known to be deficient in glycosylase activity (*56*) was introduced on a plasmid into both the CC102 and CC104 strains along with appropriate vector controls. Because this mutant cannot perform the base excision reaction, D138A fails to reduce the high reversion frequency observed with CC102 *nth-*. However, D138A is able to complement the CC104 *nth-* strain (Table 2.4). Thus, the glycosylase activity of EndoIII is not required for its helper function to aid MutY in repairing lesions inside the cell. Nonetheless, it appears that EndoIII lacking D138 can bind DNA and contains an intact [4Fe4S] cluster (39). Based upon our model, D138A should be competent to carry out DNA-mediated CT and thus serve as a helper to MutY, as we observe.

A mutant defective in DNA/protein CT In our model, it is the ability to carry out DNA-mediated CT not the glycosylase activity of EndoIII that is critical to its helper function.

Table 2.2. Assay for DNA Repair in *E. coli* by MutY (CC104).

Strain	***Lac*+ Revertants**[a,b]	**Increase**
	(*lac*$^+$ colonies/ 10^9 cells)	(x/CC104)
CC104[c]	20 ± 9	---
CC104 *nth-*	54 ± 5	2.7
CC104 *mutY-*	300 ± 33	15
CC104 *mutY-/nth-*	270 ± 29	13.5

a. *Lac*$^+$ revertants are reported as the average number of *lac*$^+$ colonies that arise per 10^9 cells plated on minimal lactose media.

b. These data represent a single set of experiments with 10 replicates per strain assayed concurrently. Values reported as the mean ± s.d.

c. CC104 strains reflect the rate of GC to TA transversion mutations and serve as a reporter for MutY activity in *E. coli*.

Table 2.3. Assay for DNA Repair in *E. Coli* by EndoIII (CC102)

Strain	***Lac*+ Revertants**[a,b]	**Increase**
	(*lac*$^+$ colonies/ 10^9 cells)	(x/CC102)
CC102[c]	14 ± 4	---
CC102 *nth-*	34 ± 8	2.4
CC102 *mutY-*	27 ± 9	1.9
CC102 *mutY-/nth-*	48 ± 16	3.4

a. *Lac*$^+$ revertants are reported as the average number of *lac*$^+$ colonies that arise per 10^9 cells plated on minimal lactose media.

b. These data represent a single set of experiments with 10 replicates per strain assayed concurrently.

c. CC102 strains reflect the rate of GC to TA transversion mutations and serve as a reporter for MutY activity in *E. coli*.

Table 2.4. (Top) MutY Activity Assay (CC104) with an Enzymatic EndoIII Mutant (D138A). (Bottom) EndoIII Activity Assay (CC102) with an Enzymatic EndoIII Mutant (D138A) and a DNA CT Mutant (Y82A)

Strain	***Lac*+ Revertants**[a,b]	**Increase**
	(*lac*$^+$ colonies/ 10^9 cells)	(x/CC104/p)
CC104/p[c]	33 ± 2.0	---
CC104 *nth-*/p	64 ± 7.4	1.9
CC104 *nth-*/pnth	36 ± 3.3	1.1
CC104 *nth-*/D138A	32 ± 3.2	1.0

Strain	***Lac*+ Revertants**[a,b]	**Increase**
	(*lac*$^+$ colonies/ 10^9 cells)	(x/CC102/p)
CC102/p[c]	10 ± 4.2	---
CC102 *nth-*/p	68 ± 26	6.8
CC102 *nth-*/pnth	18 ± 5.4	1.8
CC102 *nth-*/D138A	62 ± 31	6.2
CC102 *nth-*/Y82A	16 ± 7.1	1.6

a. *Lac*$^+$ revertants are reported as the average number of *lac*$^+$ colonies that arise per 10^9 cells plated on minimal lactose media with ampicillin (40 μg/mL).

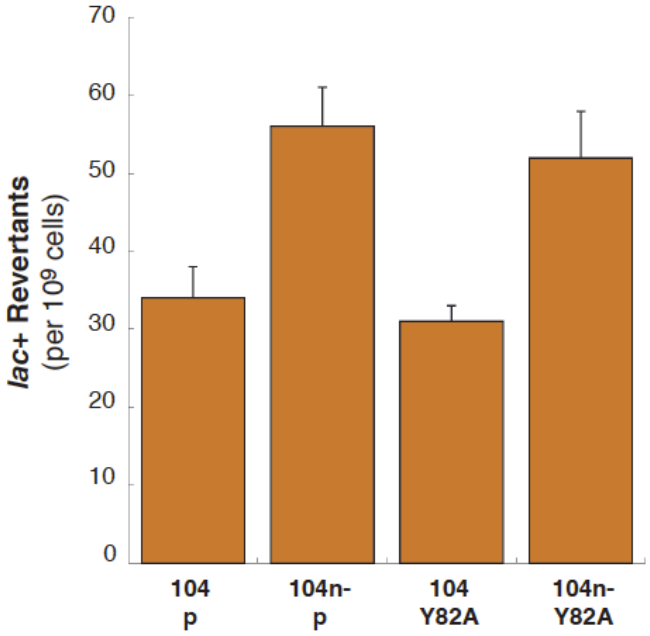
b. These data represent a single set of experiments with 10 replicates per strain assayed concurrently.

c. CC102 strains reflect the rate of GC to AT transition mutations and serve as a reporter for EndoIII activity in *E. coli*.

Thus, perturbing the path for electron transfer to the DNA would interfere with this helper function. Aromatic tyrosine and tryptophan residues often facilitate long-range electron transfers in proteins (*57, 58*), and EndoIII contains many of these residues. In particular, Y82 is conserved in most EndoIII and MutY homologs (*44*), and an analogous mutation (Y166S) in the human homolog of MutY is associated with cancer (*59*). In the crystal structure, Y82 is located very close to the DNA backbone (*4*). Y82A EndoIII was thus introduced on a plasmid into both reporter strains (CC102 and CC104) and their *nth* knockouts to explore whether this mutation attenuates helper function (Table 2.5). Significantly, Y82A in the CC104 *nth*- strain shows an increase in mutation rate versus the CC104/Y82A and CC104/p controls (Figure 2.11). The number of *lac*+ revertants is found to increase by 53±16% when comparing CC104 *nth*-/ Y82A to CC104/p. When comparing CC104 *nth*-/Y82A to CC104/Y82A, the number of *lac*+ revertants increases by 68 ± 13%. Similarly, for these trials, the ratio of the number of *lac*+ revertants for CC104 *nth*-/ p versus CC104/p is 165 ± 13%. These results clearly indicate that Y82A does not restore helper function.

It is noteworthy that inclusion of Y82A EndoIII in CC102 *nth*- leads to a diminished mutation rate indicating that this mutant is competent for EndoIII activity inside the cell (Table 2.5). Interestingly, the observation that Y82A complements CC102 *nth*-, but not CC104 *nth*-, is consistent with the conclusion that the glycosylase activity of EndoIII is not a source of helper function. Moreover the fact that Y82A complements CC102 *nth*- is understandable in the context of our model, because of the higher copy number of EndoIII in *E. coli* cells than MutY. In our model, without oxidative stress, we would predict that DNA CT is not essential for EndoIII repair activity inside the cell. We would therefore anticipate that the role of EndoIII in helping MutY search for lesions may be more important than the ability of EndoIII to find its own lesions. This distinction becomes more complex when considering that other DNA-binding proteins with iron-sulfur clusters might also participate in the signaling process.

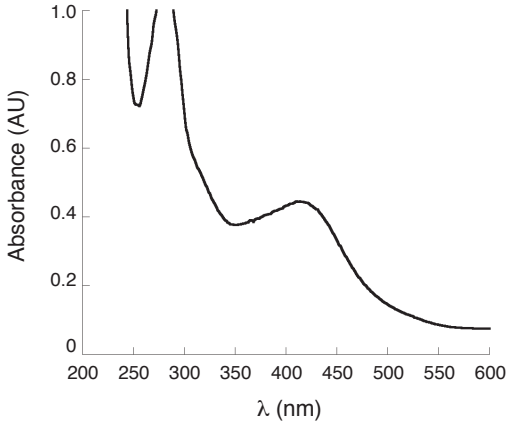
Figure 2.11. Bar graph showing *lac*+ revertants for CC104/p, CC104 *nth*-/p, CC104/Y82A and CC104 *nth*-/Y82A strains, where p denotes inclusion of an empty vector. *Lac*+ revertants are reported as the average number *lac*+ colonies that arise per 10^9 cells plated on minimal lactose media containing ampicillin. Data for the CC104 strains are shown based upon five sets of independent experiments, each containing 10 replicates per strain.



To establish the biochemical characteristics of Y82A EndoIII, the protein was purified and its redox and glycosylase activities examined. Importantly, the mutant enzyme does contain the [4Fe4S] cluster, characterized by its distinctive absorbance spectrum (Figure 2.12). Y82A EndoIII also maintains glycosylase activity against a 5-OH-dU lesion in a ^{32}P-5′-endlabelled 35-mer duplex (Figure 2.12); the activity of the mutant in this assay is equal to that of wild-type. Note that this experiment on a 35-mer duplex measures only the base excision reaction, not the search process. Similarly, in the *E. coli* EndoIII activity assay, where we expect that the search process is not rate-limiting, Y82A EndoIII activity is comparable to that of wt EndoIII. In contrast, D138A EndoIII, which instead inhibits the base excision reaction, fails to complement the *nth* knockout in the EndoIII activity reporter strain but does complement the *nth* knockout in the MutY activity reporter strain, where lesion detection is limiting.

To test for DNA-bound redox activity, Y82A was examined on a Au electrode modified with thiol-terminated DNA duplexes. Significantly, in the cyclic voltammogram, the potential for the DNA-bound mutant resembles that of the wild-type (*11*), but the signal intensity is diminished (Figure 2.13). The protein concentrations are determined based on the 410 nm absorbance of the [4Fe4S] cluster; the smaller electrochemical signal observed with Y82A does not reflect a lower concentration of [4Fe4S] clusters. Over several trials, Y82A EndoIII exhibits a signal that is 50 ± 13% smaller than that for wt EndoIII (per [4Fe4S] cluster). This signal intensity provides a reliable measurement of reduction/oxidation of the DNA-bound protein. Since the glycosylase activity on the 35-mer is equal for the mutant and wild-type, this diminished signal cannot reflect diminished binding of the mutant to the DNA. Instead this lowered signal intensity would be expected with an attenuated efficiency of CT from the cluster to DNA and reflects poor electronic coupling of the mutant with the DNA duplex. These results therefore indicate that Y82A EndoIII is defective in DNA-mediated signaling.

Figure 2.12. Characterization of Y82A protein. (Left) The UV-visible spectrum of Y82A EndoIII. A peak in the visible region is observed at 410 nm characteristic of a [4Fe4S] cluster. (Right) Autoradiogram after denaturing PAGE of ^{32}P-5′-TGTCAATAGCAAG**X**GGAGAAGT-CAATCGTGAGTCT-3′ + complementary strand where X = 5-OH-dU base-paired with G. Protein samples (100 or 10 nM) were incubated with duplexes for 15 min at 37°C and quenched with 1 M NaOH. Cleavage of the ^{32}P-labelled strand at the lesion site (**X**) by EndoIII results in formation of a 14mer. No significant difference in glycosylase activity (10% uncertainty) is observed between Y82A and wt EndoIII.



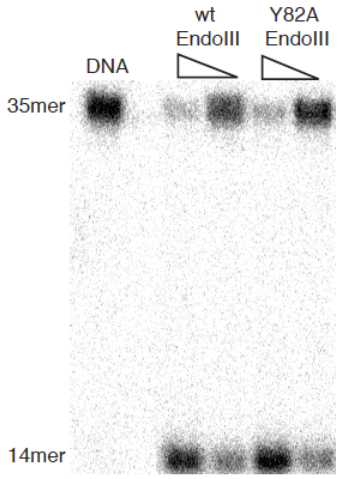
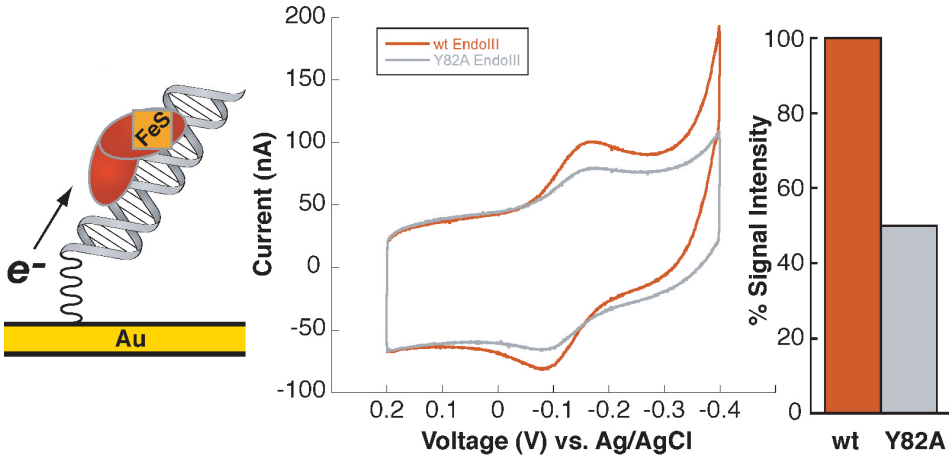


Figure 2.13. Cyclic voltammetry of Y82A EndoIII at a Au electrode modified with $SH(CH_2)_2CONH(CH_2)_6NHOCO$-5′-AGTACAGTCATCGCG-3′ + complementary strand showing the reduction and reoxidation of the DNA-bound protein. DNA-modified surfaces were prepared, backfilled with mercaptohexanol, and wt or Y82A EndoIII was tested. Surfaces were then rinsed and the other protein analyzed on the same surface. Over several trials, the electrochemical signal associated with Y82A is 50 ± 13% smaller per [4Fe4S] cluster compared with wt EndoIII, reflecting poor electronic coupling of the mutant to the DNA-modified electrode.

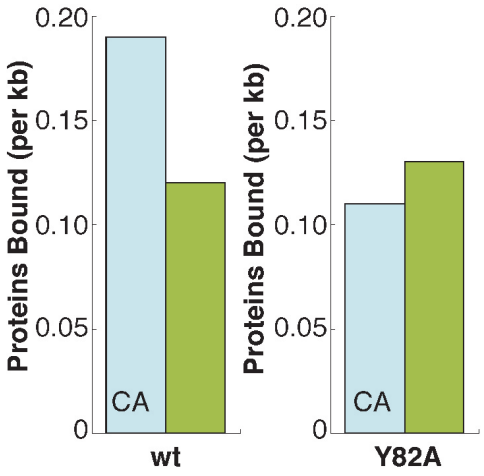


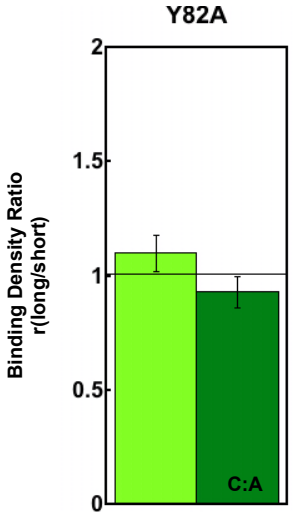
Significantly, and consistent with these results, examination of the distribution of Y82A on mismatched and matched strands by AFM shows no preference for the mismatched strand; we observe 0.11 proteins per kilobase long strand and 0.13 proteins per kilobase short strand (Figure 2.14). In fact, the ratio of protein densities on mismatched versus matched strands with Y82A, r(long/short) is 0.9, essentially equal to that of wild-type EndoIII bound to fully matched long versus short strands (Figure 2.14). Since the Y82A mutant, biochemically defective only in DNA CT, cannot redistribute to the vicinity of the lesion, DNA CT must play a role in finding the lesion both in the AFM experiment and in the helper function assay. These results together demonstrate a distinct connection between DNA-mediated CT to the [4Fe4S] cluster, the detection of DNA defects, and the *in vivo* relationship observed between MutY and EndoIII.

Discussion

These experiments indicate that MutY and EndoIII cooperate in their search for damage in the genome and redistribute in the vicinity of lesions consistent with CT scanning. This cooperation, or helper function, does not involve the glycosylase reaction. Based on their chromosomal arrangement, the expression of MutY and EndoIII, furthermore, do not appear to be linked (*60*). There is also no chemical evidence that the proteins physically bind to one another, and their low abundance within the cell makes random associations improbable. This cooperation thus arises from a distance. Importantly, what does appear to be required for helper function is an intact [4Fe4S] cluster as well as an electroactive protein-DNA interface. Mutation of an aromatic amino acid residue near the DNA binding site, Y82A, leads to a decrease in CT efficiency *in vitro*, the inability of the protein to redistribute near lesions by AFM, and diminished helper function *in vivo*. These experiments thus establish a link between DNA-mediated CT and the cooperative search for damage by these repair proteins both *in vitro* and *in vivo*.

Figure 2.14. (Top) Comparative densities for WT and Y82A EndoIII bound to matched versus mismatched (C:A) strands measured by AFM. Although wt EndoIII preferentially redistributes onto the mismatched strand, (bottom) Y82A shows no preference.





BER glycosylases are known to prevent mutations inside the cell, yet in most organisms, these enzymes are not required for normal growth and development (*2*). Recently it was discovered that germline mutations in human BER homologs result in genetic predisposition to cancer (*59*). Specifically, the human homolog of *mutY* (*MUTYH*) is found mutated in a subset of patients predisposed to colorectal cancer. Many of the cancer-associated mutations in *MUTYH* are missense, or single amino acid, mutations. Though several of the most common mutants have been characterized biochemically, it remains unclear exactly how these variants lead to disease. Given that initial detection of lesions is likely the rate-limiting step in BER (*6*), it is possible that that mutants with defects in protein-DNA CT would be associated with cancer. Indeed, many of these *MUTYH* missense mutations found in colorectal cancer patients result in loss or gain of aromatic residues near predicted protein-DNA interfaces (*59*). Significantly, *MUTYH* contains two adjacent tyrosine residues (Y165 and Y166) that closely align with Y82 in *E. coli* EndoIII and inherited mutations in these *MUTYH* residues (Y165C and, less commonly, Y166S) are clinically relevant in cancer. These results thus provide tantalizing evidence for association between defects in lesion detection via DNA-mediated CT by BER enzymes and human disease.

Iron-sulfur clusters are becoming increasingly ubiquitous to proteins that repair, replicate, and transcribe DNA (*61, 62*). Recent characterizations of archaeal DNA primase, RNA polymerase, and nucleotide excision repair helicase (XPD) homologs reveal an iron-sulfur cluster required for normal enzyme function. Though the precise role of the cluster in these proteins is unclear, the cysteine residues ligating the cluster are conserved in eukaryotic homologs of these proteins. In archaeal XPD, moreover, the iron-sulfur cluster occupies a site far from the ATP hydrolysis domain but implicated in DNA binding (*61*). It is interesting to consider whether in these proteins, as in BER enzymes, the iron-sulfur cluster is poised to send and receive redox signals mediated by the DNA helix, which may, in turn, modulate DNA binding affinity, enzyme activity, or protein structure. Such long-range signaling among proteins bound to DNA would make searching for

lesions much more efficient and may generally provide a means of genome-wide communication to monitor cellular stresses.

DNA-mediated CT serves as a fast and efficient reaction that is exquisitely sensitive to lesions in the base pair stack. This chemistry helps to explain how these repair glycosylases locate their lesions efficiently in the cell, a key function since mutations in these enzymes in humans are implicated in colorectal cancer (*59*). This mechanism furthermore provides a rationale for iron-sulfur clusters in DNA repair proteins. More generally, these experiments illustrate the importance of DNA-mediated CT in long-range signaling among proteins in low abundance that are bound to DNA. Other roles for DNA-mediated CT in biological signaling must now be considered.

REFERENCES

1. Demple, B., and Harrison, L. (1994) *Annu. Rev. Biochem. 63*, 915.

2. David S. S., O'Shea V. L., and Kundu, S. (2007) *Nature 447*, 941.

3. Francis, A. W., Helquist, S. A., Kool, E. T., and David, S. S. (2003) *J. Am. Chem. Soc. 125*, 16235.

4. Fromme, J. C., and Verdine, G. L. (2003) *EMBO J. 22*, 461.

5. Fromme, J. C., Banerjee, A., Huang, S. J., and Verdine, G. L. (2004) *Nature 427*, 652.

6. Livingston, A. L., O'Shea, V. L., Kim, T., Kool, E. T., and David, S. S. (2008) *Nat. Chem. Biol. 4*, 51.

7. Berg, O. G., Winter, R. B., and von Hippel, P. H. (1981) *Biochemistry 20*, 6929.

8. Parker J. B., Bianchet, M. A., Krosky, D. J., Friedman, J. I., Amzel, L. M., and Stiver, J. T. (2007) *Nature 449*, 433.

9. Cunningham, R. P., Asahara, H., Bank, J. F., Scholes, C. P., Salerno, J. C., Surerus, K., Munck, E., McCracken, J., Peisach, J., and Emptage, M. H. (1989) *Biochemistry 28,* 4450.

10. Boon, E. M., Livingston, A. L., Chmiel, N. H., David, S. S., and Barton, J. K. (2003) *Proc. Natl. Acad. Sci. USA 100*, 12543.

11. Boal, A. K., Yavin, E., Lukianova, O. A., O'Shea, V. L., David, S. S., and Barton, J. K. (2005) *Biochemistry 44*, 8397.

12. *Charge Transfer in DNA: From Mechanism to Application* (2005) ed Wagenknecht HA (Wiley-VCH, Weinheim).

13. Nunez, M. E., Hall, D. B., and Barton, J. K. (1999) *Chem. Biol. 6*, 85.

14. Kelley, S. O., Jackson, N. M., Hill, M. G., and Barton, J. K. (1999) *Angew. Chem. Int. Ed. 38*, 941.

15. Elias, B., Shao, F., and Barton, J. K. (2008) *J. Am. Chem. Soc. 130*, 1152.

16. Nunez, M. E., Noyes, K. T., and Barton, J. K. (2002) *Chem Biol. 9*, 403.

17. Nunez, M. E., Holmquist, G. P., and Barton, J. K. (2001) *Biochemistry 40*, 12465.

18. Boal, A. K., and Barton, J. K. (2005) *Bioconj. Chem. 16*, 312.

19. Boon, E. M., Ceres, D. M., Drummond, T. G., Hill, M. G., and Barton, J. K. (2000) *Nat. Biotechnol. 18*, 1096.

20. Guo, X. F., Gorodetsky, A. A., Hone, J., Barton, J. K., and Nuckolls, C. (2008) *Nat. Nanotechnol. 3*, 163.

21. Boon, E. M., Salas, J. E., and Barton, J. K. (2002) *Nat. Biotechnol. 20,* 282.

22. Yavin, E., Boal, A. K., Stemp, E. D. A., Boon, E. M., Livingston, A. L., O'Shea, V. L., David, S. S., and Barton, J. K (2005) *Proc. Natl. Acad. Sci. 102*, 3546.

23. Yavin, E., Stemp, E. D. A., O'Shea, V. L., David, S. S., and Barton, J. K. (2006) *Proc. Natl. Acad. Sci. USA 103*, 3610.

24. Cupples, C. G., and Miller, J. H. (1989) *Proc. Natl. Acad. Sci. USA 86*, 5345.

25. Davis, R. W., Botstein, D., and Roth, J. R. (1980) *Advanced Bacterial Genetics: A Manual for Genetic Engineering* (Cold Spring Harbor Laboratory Press).

26. Berg, O. G., Winter, R. B., and von Hippel, P. H. (1981) *Biochemistry 20*, 6929.

27. Blainey, P. C., van Oijen, A. M., Banerjee, A., Verdine, G. L., and Xie, X.S. (2006) *Proc. Natl. Acad. Sci. USA 103*, 5752.

28. Wunderlich, Z., and Mirny, L. A. (2008) *Nucleic Acids Res. 36*, 3570.

29. Porello, S. L., Williams, S. D., Kuhn, H., Michaels, M.L., and David, S. S. (1996) *J. Am. Chem. Soc. 118*, 10684.

30. Thanbichler, M., Wang, S. C., and Shapiro, L. (2005) *J. Cell Biochem. 96*, 506.

31. Berg, O. G. (1986) *Biopolymers 25*, 811.

32. Mullineaux, C. W., Nenninger, A., Ray, N., and Robinson, C. (2006) *J. Bacteriol. 188*, 3442.

33. Asahara, H., Wistort, P. M., Bank, J. F., Bakerian, R. H., and Cunningham, R. P. (1989) *Biochemistry 28*, 4444.

34. Berg, O. G., and von Hippel, P. H. (1985) *Annu. Rev. Biophys. Chem. 14*, 131.

35. O'Neill, M. A., and Barton, J. K. (2002) *Proc. Natl. Acad. Sci. USA 99*, 16543.

36. Kelley, S. O., Boon, E. M., Barton, J. K., Jackson, N. M., and Hill, M. G. (1999) *Nucleic Acids Res. 27*, 4830.

37. Donahue, W. F., Turczyk, B. M., and Jarrell, K. A. (2002) *Nucleic Acids Res. 30*, e95.

38. Horcas I., Fernández, R., Gómez-Rodríguez, J. M., Colchero, J., Gómez-Herrero, J., and Baro, A. M. (2007) *Rev. Sci. Instrum. 78*, 013705.

39. Golinelli, M. P., Chmiel, N. H., and David, S. S. (1999) *Biochemistry 38*, 6997.

40. Datsenko, K. A., and Wanner, B. L. (2000) *Proc. Natl. Acad. Sci. USA 97*, 6640.

41. Miller, J. H. (1992) *A Short Course in Bacterial Genetics: A Laboratory Manual and Handbook for Escherichia coli and Related Bacteria.* (Cold Spring Harbor Laboratory Press).

42. Sambrook, J., Russell, D. W. (2001) Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory Press).

43. Boiteux, S., O'Connor, T. R., and Laval, J. (1987) *EMBO J. 6*, 3177.

44. Watanabe, T., Blaisdell, J. O., Wallace, S. S., and Bond, J. P. (2005) *J. Biol. Chem. 280*, 34378.

45. Gorodetsky, A. A., Boal, A. K., and Barton J. K. (2006) *J. Am. Chem. Soc.* 128, 12082.

46. O'Handley, S., Scholes, C. P., and Cunningham, R. P. (1995) *Biochemistry 34*, 2528.

47. Slutsky, M., and Mirny, L. A. (2004) *Biophys. J. 87*, 4021.

48. Augustyn, K. E., Merino, E. J., and Barton, J. K. (2007) *Proc. Natl. Acad. Sci. USA 104*, 18907.

49. Burrows, C. J., and Muller, J. G. (1998) *Chem. Rev. 98*, 1109.

50. Merino, E. J., and Barton, J. K. (2007) *Biochemistry 46*, 2805.

51. Chen, L., Haushalter, K. A., Lieber, C. M., and Verdine, G. L. (2002) *Chem. Biol. 9*, 345.

52. Nghiem, Y., Cabrera, M., Cupples, C. G., and Miller, J. H. (1988) *Proc. Natl. Acad. Sci. USA 85*, 2709.

53. Kreutzer, D. A., and Essigmann, J. M. (1998) *Proc. Natl. Acad. Sci. USA 95*, 3578.

54. Tano, K., Iwamatsu, Y., Yasuhira, S., Utsumi, H., and Takimoto, K. (2001) *J Radiat. Res. 42*, 409.

55. Thayer, M. M., Ahern, H., Xing, D., Cunningham, R. P., and Tainer, J. A. (1995) *EMBO J. 14*, 4108.

56. Kuo, C. F., McRee, D. E., Fisher, C. L., O'Handley, S. F., Cunningham, R. P., and Tainer, J. A. (1995) *Science 258*, 434.

57. Stubbe, J., Nocera, D. G., Yee, C. S., Chang, M. C. (2003) *Chem. Rev. 103*, 2167.

58. Shih, C., Museth, A. K., Abrahamsson, M., Blanco-Rodríguez, A. M., Di Bilio, A. J., Sudhamsu, J., Crane, B. R., Ronayne, K. L., Towrie, M., Vlček, A., Richards, J. H., Winkler, J. R., and Gray, H. B. (2008) *Science 320*, 1760.

59. Cheadle, J. P., Sampson, J. R. (2007) *DNA Repair 6*, 274.

60. Gifford, C. M., and Wallace, S. S. (2000) *Nucleic Acids Res 28*, 762.

61. Fan, L., Fuss, J. O., Cheng, Q. J., Arvai, A. A., Hammel, M., Roberts, V. A., Cooper, P. K., and Tainer, J. A. (2008) *Cell 133*, 789.

62. Hirata, A., Klein, B. J., and Murakami, K. S. (2008) *Nature 451*, 851.

CHAPTER 3

Mutants of the Base Excision Repair Glycosylase, Endonuclease III: DNA Charge Transport as a First Step in Lesion Detection

Adapted from Romano, C. A.*, Sontz, P. A.*, Barton, J. K. (2011) *Biochemistry* **50**, 6133–6145

P.A.S. carried out AFM experiments and binding assays. C.A.R. isolated and characterized mutants. P.A.S. and C.A.R. performed *in vivo* studies, but results reported here were obtained by P.A.S.

C.A.R. and P.A.S. gratefully acknowledge H.B. Gray, D.K. Newman, A.K. Boal, and J.C. Genereux for helpful discussions. T.J. Ge and A. Parisian provided technical assistance, and E.D. Olmon helped prepare figures. P.A.S. is also grateful to the Beckman Institute MMRC for AFM instrumentation.

INTRODUCTION

With millions of DNA bases encountering more than 10,000 oxidative lesions per day (*1*), DNA repair proteins are charged with the daunting task of searching an entire genome to detect and repair these multiple sites of damage (*2, 3*). Our laboratory has proposed that DNA repair proteins search for DNA lesions by exploiting DNA/protein charge transport (CT). DNA CT occurs over long molecular distances but is remarkably sensitive to DNA lesions (*4, 5, 6*). Endonuclease III (EndoIII) and MutY are base excision repair (BER) proteins in *Escherichia coli* that contain [4Fe4S] clusters which, when bound to DNA, are activated toward one electron oxidation (*7,8*). EndoIII and MutY homologues are found in many species, ranging from bacteria to humans, and their structural biology and enzymatic properties have been studied in great detail by both crystallography (*9-14*) and mutagenesis-based enzymology (*7, 13, 15-22*). Both enzymes are present in low copy number *in vivo* (~ 30 copies per cell of MutY and ~ 500 copies per cell of EndoIII) (*3*), making their ability to find and repair DNA lesions all the more remarkable. MutY has also drawn attention from the biomedical community because patients with mutations in its human homologue, MUTYH, are predisposed towards developing colorectal cancer (*23-27*).

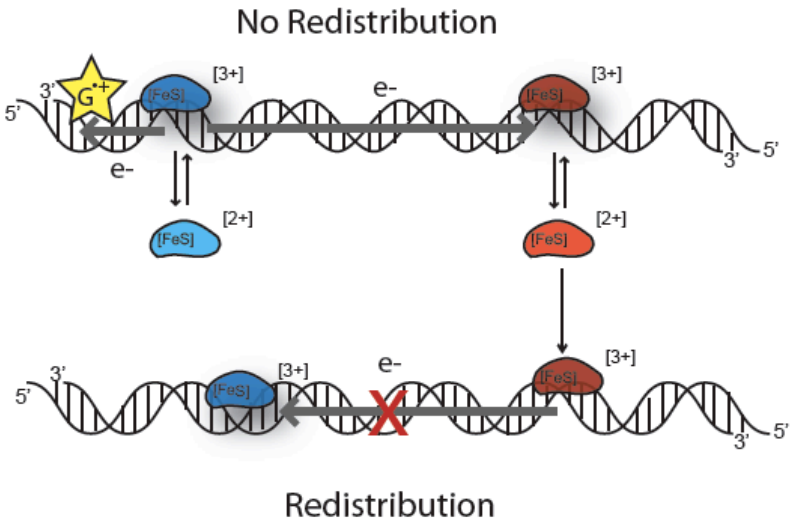
Despite these extensive studies, two features of the BER family of enzymes remain intriguing to consider. First, what is the role of the [4Fe4S] cluster? The [4Fe4S] cluster does not participate in base excision in either enzyme, nor is it necessary for MutY to fold properly (*7*), although it is essential for MutY to bind DNA. Secondly, how are these proteins able to scan the vast amount of DNA present in a cell in order to detect their substrate lesions? Although the excision step in repair by these BER glycosylases has been well characterized, the answer to these questions has not been resolved.

Our laboratory has proposed a model in which MutY and EndoIII cooperatively scan the genome for DNA lesions using their [4Fe4S] clusters to participate in DNA-mediated redox signaling (*28-31*). DNA mediates CT very efficiently owing to the π-stacking of the aromatic, heterocyclic base

pairs within the helix (*5, 32, 33*). CT through proteins occurs through a variety of mechanisms (*34-39*). These CT mechanisms could permit charge to travel from the DNA into a bound protein, and then through the protein to a redox-active cofactor, such as a [4Fe4S] cluster. In this model, under oxidative stress, oxidation of some DNA-bound protein occurs by a nearby guanine radical, taking the [4Fe4S] cluster from the 2+ form to 3+. DNA-mediated CT to another distally DNA-bound protein would lead to its reduction and dissociation. This CT process relies on proper π-stacking of the bases between the donor protein and the distal acceptor protein, however (*28*); if the π-stack is perturbed by a lesion, DNA-mediated CT will not proceed. Thus this signaling serves as a scan for the integrity of the intervening DNA as the first step in damage detection, promoting the dissociation of repair proteins away from unmodified DNA and leading to their relocalization in the vicinity of damage (Figure 3.1).

One question that remains is how charge is transported from the DNA, through the protein, to the [4Fe4S] cluster of MutY or EndoIII. Specifically, which amino acids in these proteins facilitate CT between the DNA helix and the protein metal cluster? One EndoIII mutant, Y82A, has already been shown to be deficient in CT activity, implicating Y82 as a crucial residue in the CT pathway of EndoIII (*28*). The current study builds on this result by characterizing several novel mutants of EndoIII to discover other residues that mediate protein/DNA CT. Several categories of mutants were prepared, including five additional mutants at or near the Y82 position, to better assess the types of amino acid side chains that are CT proficient in EndoIII. Mutations were also made targeting other aromatic residues in the protein because aromatic amino acids have been shown to mediate CT in other systems (*35-39*). A third category of mutations were made near the [4Fe4S] cluster, assuming that amino acids close to this redox-active moiety would likely lie along the CT pathway. All of the mutants were characterized by cyclic voltammetry on DNA-modified electrodes, glycosylase assays, and some by circular dichroism (CD) spectroscopy.

Figure 3.1. Model for the first step in damage detection by BER proteins. Given two repair proteins of similar potential (red and blue), CT can proceed through well-matched unmodified DNA (top). Stimulated through oxidation by guanine radicals (yellow) formed under oxidative stress, DNA CT occurs between DNA-bound repair proteins, promoting the dissociation of reduced protein. DNA CT does not occur (bottom), however, in the presence of an intervening mismatch or lesion (red X); repair proteins are not reduced, do not dissociate, and instead remain bound in the vicinity of the lesion. On a slower time scale, the proteins can then processively move to the lesion site for repair. This process leads to the redistribution of proteins in the vicinity of the lesion.



These experiments helped us identify several amino acids that facilitate CT between the DNA helix and the [4Fe4S] cluster.

Upon characterization of EndoIII mutants, we then applied them in atomic force microscopy (AFM) studies to examine this first step of lesion detection. AFM is a valuable technique that allows for visualization of single-molecule protein/DNA interactions (*40-42*). AFM requires very low concentrations of DNA and protein for measurements and relies on cationic interactions for DNA/surface contact rather than chemical modifications. AFM images can be used to pinpoint the position of proteins on DNA, and hence the statistical distribution of proteins along the strand (*43, 44*). Here, then AFM may be used to test our model of protein redistribution on DNA in the vicinity of a lesion that interferes with CT. Indeed we find a correlation between the ability of mutants to carry out DNA/protein CT and their ability to relocate near a single base mismatch.

Materials and Methods

Preparation of DNA and protein samples for electrochemistry experiments Redmond Red (RR) is a redox active probe used to quantify the amount of DNA on the electrode surface (*45*). The sequences used for electrochemistry experiments were SH-(C_6H_{12})-(RR)-5′-GA GAT ATA AAG CAC GCA-3′ and complement, and SH-(C_6H_{12})-5′-AGT ACA GTC ATC GCG-3′ annealed to 5′-CGC GAT GAC TGT ACT-3′-RR. Redmond Red was connected to the oligonucleotide via a pyrrolidinyloxy linker. RR-labeled DNA was prepared according to established methods (*4*). Briefly, all DNA sequences were prepared using standard phosphoramidite chemistry on an Applied Biosystems 3400 DNA synthesizer. Phosphoramidites were purchased from Glen Research. For thiolated strands, the 5′ end was modified with the Thiol Modifier C6 S-S phosphoramidite using standard protocols from Glen Research, Inc. DNA modified with RR on the 3′ terminus was prepared on Epoch Redmond Red CPG columns from Glen Research with ultramild phosphoramidites and reagents. DNA modified with Redmond Red attached to the thiol was

prepared with Epoch Redmond Red phosphoramidite from Glen Research with ultramild phosphoramidites and reagents.

Deprotection, purification, chemical modification of DNA, annealing, and preparation of the DNA-modified Au electrode were performed as described previously (*6, 46-48*). Au substrates were purchased from Agilent Technologies. Each experiment used 50 μL of 50-100 μM EndoIII in protein buffer (20 mM NaH_2PO_4 pH 7.5, 100 mM NaCl, 5% glycerol, 1 mM EDTA). The protein concentration was quantified using ε_{410}= 17,000 $M^{-1}cm^{-1}$ (*49*).

Cyclic voltammetry All cyclic voltammetry experiments were performed as described previously (*28-30*). A Au on mica (Molecular Imaging) electrode was assembled and incubated with thiol-modified DNA duplex for 24-36 hours in the absence of Mg^{2+}; under these conditions, the DNA is loosely packed on the surface and protein can diffuse through the monolayer. The electrode was then backfilled with 1 mM mercaptohexanol, rinsed in DNA buffer (50 mM NaCl, 5 mM NaH_2PO_4 pH 7.5), and then rinsed in protein buffer before protein samples were added to the surface. The DNA-modified Au electrode served as the working electrode, a Pt wire served as the auxiliary electrode, and the reference was either a Ag/AgCl electrode modified with an agarose tip or a 66-EE009 Ag/AgCl reference electrode (ESA Biosciences). All scans were taken at a rate of 50 mV/s on a CH Instruments 760 potentiostat. To begin an experiment, a protein sample was placed on an electrode surface and successive scans were taken every five minutes until the signal stabilized. Protein concentrations used were those required to saturate binding to the surface as determined by titration (*29, 30*). It should be noted that interfacial electron transfer rates redox moieties bound to DNA on these DNA-modified surfaces are limited by tunneling through the alkane-thiol linker based on measurements of peak widths as a function of scan rate (*4*).

To determine the relative CT efficiencies of EndoIII mutants, the signal intensity of the proteins was normalized to DNA concentration using the intensity of the RR signal and compared one to

another on the same surface. For each comparison, the protein samples were measured consecutively on the same electrode surface. Thus, to compare the CT signal strength of each mutant relative to wt EndoIII, the peak under the reductive and/or oxidative cyclic voltammogram curve (consistently for different proteins on a given surface) was measured using the linear baseline correction of CH Instruments' Electrochemical Workstation software program. After each measurement of a protein sample, the electrode surface was rinsed, and the measurement was repeated for the RR signal. The intensity measured for each mutant was divided by that of the RR signal measured on the rinsed electrode immediately following the mutant scan. This ratio was also tabulated for wild-type EndoIII on each electrode. The final ratios reported are those of the RR corrected mutant value divided by the RR corrected wild-type value. This procedure provides a correction for the variability in surface coverage associated with individual DNA-modified surfaces. Surface coverages were typically 5 pmole/$Å^2$ but variations of ~ 50% were apparent for different electrodes on different days.

Preparation of EndoIII overexpression construct The nth gene that encodes EndoIII was cloned using the Failsafe System with Buffer G (Epicentre Biotechnologies) and plasmid pBBR1MCS-4 (*50*), which contains the nth gene as a template. (This construct was originally prepared by cloning the nth gene from Escherichia coli chromosomal DNA). Primers used were 5′-CGC CCGCG GTGGT ATG AAT AAA GCA AAA GCG CTG– 3′ and 5′- CGC GGATCC TCA GAT GTC AAC TTT CTC TTT- 3′. PCR products were purified on a 1% agarose gel and then excised using the QIAquick gel extraction kit (Qiagen). These fragments were ligated into the BamHI and SacII sites of the pET11-ubiquitin-His vector, which expresses hexahistidine and ubiquitin tags at the N-terminus of the insertion site. This vector was derived from the pET11 vector (Novagen), but the ubiquitin gene and hexahistidine tag were engineered into the vector by the

laboratory of S. Mayo, from whom this vector was a kind donation. The ligation reaction was used to transform both DH5α *E. coli* (Invitrogen) for sub-cloning, and BL21star(DE3)pLysS *E. coli* (Invitrogen) for overexpression. Transformants resistant to ampicillin and chloramphenicol were selected and the plasmids were isolated using the QIAPrep MiniPrep kit (Qiagen). After the size of the insert was verified by restriction digestion, plasmids from positive transformants were submitted for sequencing (Laragen). Freezer stocks of transformants containing the correct EndoIII sequence were prepared and stored at -80 °C.

Purification of EndoIII and mutants WT EndoIII and Y82A samples used for AFM experiments were prepared as described previously (*28*). Protein samples for other experiments and all other mutants were prepared as described here. Freezer stocks of BL21star(DE3)pLysS containing pET11-ubiquitin-His with the nth gene were used to inoculate 10 mL of LB media containing 100 μg/mL ampicillin and 30 μg/mL chloramphenicol. Each 10 mL culture was incubated at 37°C overnight and then used to inoculate 1 L of LB/ampicillin/chloramphenicol. Each 1 L flask was grown to OD_{600} = 0.6 at 37°C and then isopropyl β-D-1-thiogalactopyranoside was added to a total concentration of 0.3 mM. Cells were incubated at 30°C for 3.5 hours and harvested by centrifugation (5000 rpm, 10 minutes). Pellets were stored at -80 °C. For lysis, pellets were dissolved in 25 mL lysis buffer (50 mM Tris-HCl, pH 8, 5% glycerol, 250 mM NaCl, 5 mM DTT, 1 mM phenylmethylsulfonyl fluoride) and lysed by sonication. Cell lysate was fractionated by centrifugation (7000 rpm, 10 minutes). The supernatant was filtered, and then loaded onto a 5 mL HisTrap HP column (GE Healthcare) pre-equilibrated with binding buffer (20 mM sodium phosphate, 0.5 M NaCl, 20 mM imidazole, 1 mM DTT, pH 7.4) at a flow rate of ~ 1 mL/min The column was washed with binding buffer until the UV baseline was stable. The fusion protein was eluted using a gradient of 0–100% elution buffer (20 mM sodium phosphate, 0.5 M NaCl, 500 mM imidazole, 1 mM DTT, pH 7.4) over 10–20 column volumes on an AKTA FPLC (GE Healthcare).

Fractions containing protein (determined by their yellow color) were pooled and loaded onto a Superose 12 column (GE Healthcare) that had been pre-equilibrated with Superose 12 Buffer (50 mM NaH_2PO_4, 0.15 M NaCl, pH 7.0). The tagged EndoIII, His6-ubiquitin- EndoIII, eluted after ~ 12 mL. Fractions containing EndoIII were pooled and dialyzed overnight into protein storage buffer (20 mM sodium phosphate pH 7.5, 0.5 mM EDTA, 100 mM NaCl, 20% glycerol). Aliquots of 50-70 μL EndoIII were prepared, frozen in dry ice, and stored at -80 °C. Protein purity was assessed by SDS-PAGE, and concentration was measured using ε_{410}= 17,000 $M^{-1}cm^{-1}$ (*49*).

Site-directed mutagenesis Mutations were encoded on the pET-11 based EndoIII overexpression plasmid using the Quikchange Site-Directed Mutagenesis Kit (Stratagene). Primers used are listed in Table 3.1, where bold letters indicate the encoded mutation. DNA primers were purchased from Integrated DNA Technologies. All mutagenized plasmids were sequenced (Laragen) to verify accurate mutagenesis before being used for protein overexpression.

Instrumentation All protein samples were examined using a Beckman DU 7400 spectrophotometer. Protein concentration was measured using ε_{410} = 17,000 $M^{-1}cm^{-1}$ (*49*).

Circular dichroism Samples were measured at a concentration of ~ 5 μM at ambient temperature on a Model 62A DS Circular Dichroism Spectrometer (AVIV, Lakewood, NJ). For thermal denaturation experiments, the temperature was varied from 20°C to 60°C. Each trace shown is the average of at least three independent experiments.

DNA glycosylase assay This protocol was adapted from methods described previously (*15*). DNA strands synthesized were 5′- TGT CAA TAG CAA GXG GAG AAG TCA ATC GTG AGT CT-3′ where X = 5-hydroxyuracil, and the complement with G opposite X.

Table 3.1. Primer Sequences Used for EndoIII Mutagenesis.

Mutant	Primer Sequence
H140A	5’- ccgactattgctgtcgacacg**gcc**attttccgcgtttgtaatcg-3’
	5’- cgattacaaacgcggaaaat**ggc**cgtgtcgacagcaatagtcgg-3’
L81C	5’- gggtgaaaacctatatcaaaacgattggg**tgt**tataacagcaaagc-3’
	5’- gctttgctgttata**aca**cccaatcgttttgatataggttttcaccc- 3’
Y82S	5’- ggggtgaaaacctatatcaaaacgattgggctt**tct**aacagcaaagc -3’
	5’- gctttgctgtt**aga**aagcccaatcgttttgatataggttttcacccc-3’
Y82W	5’ - ggggtgaaaacctatatcaaaacgattgggctt**tgg**aacagcaaagc-3’
	5’-gctttgctgtt**cca**aagcccaatcgttttgatataggttttcacccc-3’
W178A	5’-gtcgactgccaccat**gcg**ttgatcctgcacgggcg-3’
	5’- cgcccgtgcaggatcaa**cgc**atggtggcagtcgac-3’
Y82F	5’-ggtgaaaacctatatcaaaacgattgggctt**ttt**aacagcaaagc-3’
	5’-gctttgctgtt**aaa**aagcccaatcgttttgatataggttttcacc-3’
Y82C	5’-ggtgaaaacctatatcaaaacgattgggctt**tgt**aacagcaaagc-3’
	5’-gctttgctgtt**aca**aagcccaatcgttttgatataggttttcacc-3’
Y185A	5’-cctgcacgggcgt**gct**acctgcattgcccgcaagccccgc-3’
	5’-gcggggcttgcgggcaatgcaggt**agc**acgcccgtgcagg-3’
F30A	5’- ccgagcttaatttcagttcgcct**gct**gaattgctgattgccgtactgc-3’
	5’- gcagtacggcaatcagcaattc**agc**aggcgaactgaaattaagctcgg-3’
Y55A	5’- gcgacggcgaaactc**gcc**ccggtggcgaatacgcctgcagc-3’
	5’- gctgcaggcgtattcgccaccgg**ccg**gagtttcgccgtcgc-3’
Y75A	5’-gaaggggtgaaaacc**gct**atcaaaacgattgggctttataacagc-3’
	5’-gctgttataaagcccaatcgttttgat**agc**ggttttcacccttc-3’

DNA was prepared using standard phosphoramidite chemistry with reagents purchased from Glen Research. Prior to annealing, single-strand DNA was purified using reversed-phased HPLC and the substrate-containing strand was 5′- radiolabeled using ^{32}P-ATP (*51*) with polynucleotide kinase (Roche).

Glycosylase activity was determined by monitoring nick formation in the hydroxyuracil-containing duplex using denaturing gel electrophoresis (*16*). For this assay, 100 nM radiolabeled duplex was incubated with EndoIII or a variant at a range of concentrations (10 nM, 100 nM or 1 μM) in reaction buffer (10 mM Tris pH = 7.6, 1 mM EDTA, 50 mM NaCl) for 15 minutes at 37°C. The reactions were quenched upon addition of NaOH to a final concentration of 100 nM. Samples were dried, counted by scintillation, and diluted with loading buffer (80% formamide, 10 mM NaOH, 0.025% xylene cyanol, 0.025% bromophenol blue in Tris-Borate-EDTA buffer) to normalize the radioactivity. Samples were then heated at 90 °C for five minutes prior to loading and then separated by denaturing PAGE. Glycosylase activity was determined by quantifying the amount of 14-mer product visualized in the gel relative to the total amount of DNA present.

Lac$^+$ reversion assays This protocol was adapted from methods described previously (*28*). Strains were streaked to LB medium and incubated overnight at 37°C. For *nth* knockouts, strains were streaked to LB+chloramphenicol (17 μg/mL). 1 mL LB cultures were started from single colonies and grown overnight in a shaking incubator at 37°C, 220 rpm. 20 μL of each starter culture was used to inoculate a 10 mL NCE+glucose culture which was then grown to a density of 10^9 cells/mL at 37°C, 250 rpm. 5 mLs of this culture was centrifuged in a clinical tabletop centrifuge at 4°C and plated on NCE+lactose solid medium and then incubated at 37°C for 36 hours. Colonies arising are reported as *lac*$^+$ revertants/mL cells plated. In experiments incorporating plasmid vectors, CC104 were made electrocompetent and transformed via electroporation at 1.8 kV. Transformants were selected on LB+ampicillin (amp) (100 μg/mL) solid medium after incubation overnight at 37°C.

Single colonies were restreaked to LB+amp (100 μg/mL) solid media and incubated 12 hours at 37°C. 1 mL LB+amp (100 μg/mL) cultures were started from these colonies and grown overnight at 37°C, 220 rpm. 20 μL of this starter culture was used to inoculate a 10 mL NCE+glucose+amp (40 μg/mL) culture which was grown to a density of 10^9 cells/mL at 37°C, 250 rpm. 5 mLs of this culture was plated onto NCE+lactose+amp (40 μg/mL) and incubated at 37°C for 36 hours.

Mismatched (C:A) strand synthesis Four primers with the following sequences were synthesized using standard phosphoramidite chemistry:

5′-GTACAGAGTTCAGTCGGCATCCGCTTACAGACAAGC-3′ (forward),

5′-CCGGTAACTATCGTCTTGAGTCC-3′ (reverse),

5′-GACTGAACTCTGTACCTGGCACGACAGGTTTCCCG-3′ (forward),

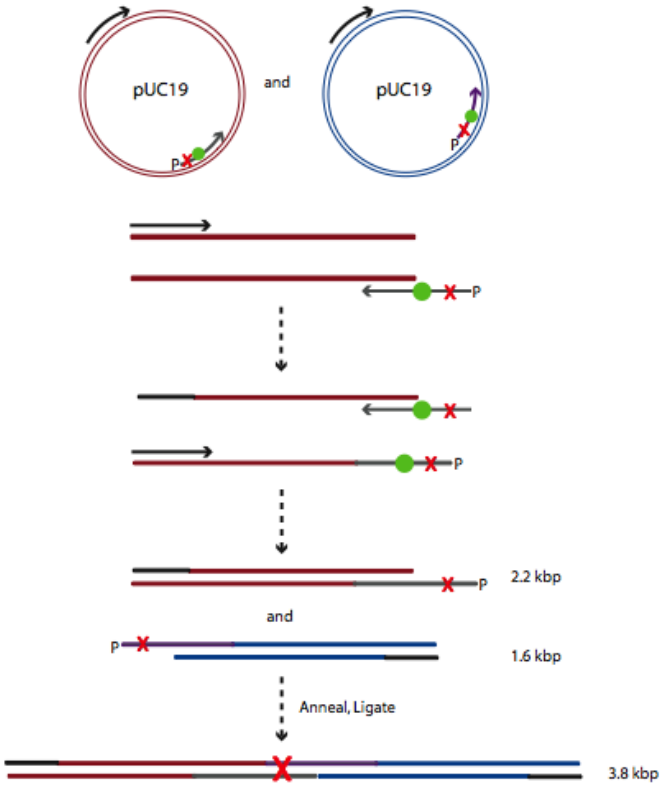
5′-GACTGAACTCTATACCTGGCACGACAGGTTTCCCG-3′ (forward)

The underlined bases highlight the location of a 2′-*O*-methyl residue (Glen Research). Primers were phosphorylated using 5 U PNK, 10% PNK buffer, 0.5 mM ATP for 5 hrs at 37 °C. Primers were purified using phenol-chloroform extraction followed by ethanol precipitation (*52*). After being dried in vacuo, primers were redissolved in 20 μL water and used in separate PCR reactions (*53*) using pUC19 as a template to generate three duplexes 1610 bp, 2157 bp (matched), or 3767 bp (mismatched) long and each containing one 14-nucleotide single-strand overhang. Each 100 μL PCR reaction contained 50 pmol of each of two primers, 1X Taq buffer (100 mM Tris-HCl, 15 mM $MgCl_2$, 500 mM KCl, pH 8.3, 0.2 mM each dNTP, 1 ng plasmid template pUC19 and 3 U Taq polymerase (Roche). A typical step program for PCR was as follows: After incubation at 94°C for 10 minutes, 34 cycles were performed as follows: 94°C for 1 minute, (52°C for primers 1+2, 54°C for primers 2+3 or 2+4), for 1 min, and 72°C for 3 minutes. The PCR product was then suspended in 50 mM NaCl/ 5 mM phosphate buffer and quantitated. Separate duplexes were annealed at 65 °C

for 8 min in 10 mM Tris buffer, then cooled to 20 °C during 2 h. A total of 15 units of T4 DNA ligase (New England Biolabs) and 10% T4 ligase buffer were added (total reaction volume ~ 20 μL) and incubated overnight at 16 °C, followed by deactivation for 10 min at 65 °C. We did not bring the ligation reaction to completion, so as to obtain a mixture of DNA samples that were equivalent other than the presence of the mismatch at the ligation site (Figure 3.2). The DNA duplexes (ligated and unligated) were then analyzed by 0.6% agarose gel electrophoresis.

AFM experiments The protocol used was similar to that reported previously (*28*). Briefly, mica surfaces were freshly cleaved with tape. EndoIII (0.4 μM) (*28*), stored in 20 mM phosphate, 20 mM NaCl, 20% glycerol, 1 mM EDTA was added to the stock DNA solution containing 50–100 ng of total DNA composed of the mixture of ligated 3.8 kbp duplexes and the two unligated duplexes (1.6 and 2.2 kb) in 6 mM $MgCl_2$/Tris-EDTA buffer. This protein/DNA solution was incubated at 4 °C overnight and deposited (5–10 μL) on the mica surface for 1–2 min, rinsed with 2 mL of water, and dried under argon. Two of the five wt EndoIII samples analyzed were deposited onto the mica surface in 5 ng quantities, followed by immediate addition of 3.7 μM EndoIII (1 μL volume). This difference in deposition conditions did not result in any significant differences in the trends observed. Mutant protein (Y82A) was added to a stock solution of 50 ng DNA for a final protein concentration of 0.4 μM. After incubation at 4 °C overnight, deposition conditions were identical to that for wt EndoIII-DNA samples.

Figure 3.2. Synthesis of DNA with site-specific C:A mismatch. Two synthesized phosphorylated primers, one incorporating a 2′-*O*-methyl ribonucleotide (black and gray, respectively), are utilized to amplify plasmid sequences via PCR. A nucleotide is incorporated within the synthesized primer containing the 2′-*O*-methyl ribonucleotide (green) that will result in a mismatched or matched final product. As the vector is amplified, polymerase stops at the ribonucleotide, resulting in a large single-strand overhang. Upon performing two separate PCR reactions (red and blue) with two primers each, the resulting duplexes (2.2 or 1.6 kbp) with 14 base single-strand overhangs can be annealed/ligated together in good yield to produce long duplexes with or without a mismatch.



For Y82F EndoIII, W178A EndoIII, or Y75A EndoIII mutants (0.6 μM), stored in 20 mM sodium phosphate, pH 7.5, 100 mM NaCl, 20% glycerol, and 0.5 mM EDTA, protein/DNA complexes were formed with DNA solution containing ~ 200 ng of the mixture of ligated PCR product as well as the two non-ligated duplexes with overhangs (~ 6 μM DNA) in 6 mM $MgCl_2$/Tris-EDTA buffer at 4°C overnight. The reaction mixture was then deposited (5–10 μL) on the mica surface for 1–2 minutes, rinsed with 2 mL water and dried under argon.

AFM instrumentation Silicon AFM Probes purchased from Nanoscience Instruments (BudgetSensors), with a spring constant of 3 N/m and a resonance frequency of 75 kHz, were used in a Digital Instruments Multimode SPM. Images were captured in air with scan areas of 2 x 2 μm^2 or 3 x 3 μm^2 in tapping mode, at an amplitude of 0.54-2.00 V, and at a scan rate of 3.05 Hz. Scan rates of 3.05 Hz were used in order to obtain images of higher quality.

Binding density ratio calculations WSxM software was used to measure general DNA contour lengths and height profiles of the proteins as described previously (*28, 54*). For each data set, images from at least three independent samples were analyzed, compared, and pooled (> 200 long or short strands). Distinguishable strands and proteins were counted by hand. The binding density ratio, r, is defined as the ratio of the proteins bound on long strands divided by proteins bound on short strands. The ratio is normalized for length by dividing by 1.9 kbp, which is the average length of the short strands. The uncertainty was determined through the total number of proteins observed.

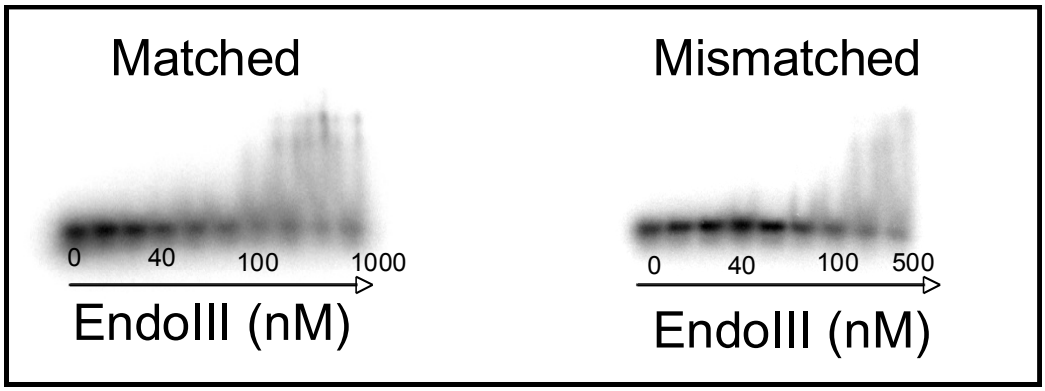
Gel shift assays of EndoIII and C:A DNA strands In order to examine whether EndoIII preferentially binds to duplex DNA containing a C:A mismatch, strands of DNA were synthesized to anneal forming matched and mismatched duplexes. The sequences were chosen to examine

nonspecific binding of Endo to shorter duplexes (21 mer) (Figure 3.3). Strands were ordered from IDT and purified via HPLC DMT off methods. Maldi-TOF was performed and strands yielded the correct masses within 9 mass units. Upon quantitating strands with UV-Vis. 400 pmol of strands were 5′ labeled with ^{32}P according to standard labeling procedures, and excess ATP was removed (*51*). Titration was used to quantitate the amount of labeled hot strand by titrating cold strand (0–128 nM) to constant (100 nM) hot strand. Samples were annealed in NaCl/PBS buffer for 1 hour using a standard program on the thermocycler. Scintillation counting was performed to ensure an equal amount of radioactivity was used for each sample. Cold and hot strands were annealed on the thermocycler to yield 40 nM stock duplexes. 116 μM wt EndoIII stock was placed on ice and diluted. Sample volumes were 20 μL in protein storage buffer with 3 nM duplex. EndoIII concentration ranged from 0 nM-1 μM with 3 nM matched or mismatched duplexes. Upon addition of protein to the DNA samples for 25 minutes, scintillation counting was performed. Mismatched samples were loaded into the gel 5 minutes prior to matched samples (10% TBE), run for 1.5 hours at 50 V, blotted one hour, and exposed overnight. Utilizing ImageQuant, where: DNA=1, Protein binding area=2, Total lane area=3, ratios 1/3 and 2/3 were used to determine percentage of protein bound. 2/3 divided by 2/3+1/3 where 2/3 has been corrected for 0 binding x 100 yields percent bound. K_d values were determined using Origin nonlinear curve fit, pharmacology function, one site bind basic fit for 10 iterations where y=Bx/k+x where k=K_d (Figure 3.3).

Results

Electrochemistry on DNA-modified electrodes The CT capabilities of several mutants of EndoIII were investigated on DNA-modified electrode surfaces passivated with mercaptohexanol (Figure 3.4). This strategy has been used previously to measure the DNA-bound electrochemical properties of proteins containing [4Fe4S] clusters (*28, 30*).

Figure 3.3. Gel shift assay of EndoIII with 3 nM matched (left) and C:A mismatched (right) DNA duplex. EndoIII is added to the DNA and incubated for 25 minutes. 10% TBE gel run 2 hours at 50 V. EndoIII concentration is increasing from 0–100 nM. Nonspecific binding occurs around 100 nM for both. EndoIII/matched K_d: 105 nM and EndoIII/mismatched K_d: 183 nM



5′^{32}P-ACTTGCGGATCCGGAGCGTCA-3′
3′ -TGAACGCCTAGGCCTCGCAGT-5′

5′^{32}P-ACTTGCGGATCCGGAGCGTCA-3′
3′ -TGAACGCCCAGGCCTCGCAGT-5′

In this study, the thiolated DNA duplex was also modified with Redmond Red, a redox probe whose signal intensity can be used to measure the amount of DNA on a given electrode surface (*45*). The Redmond Red was positioned so that it could contact the Au surface and be directly reduced without requiring DNA-mediated reaction.

For each EndoIII sample, the electrochemical signal grew in over 30-45 minutes. Scan rate dependence measurements on the protein signal showed a linear relationship between the peak current and the square root of scan rate, indicative of a diffusion-limited process (*55*). These same experiments showed a linear relationship between the peak current of Redmond Red and scan rate, indicative of a surface-bound species. Thus, while the modified DNA is surface bound, the protein freely diffuses to the DNA. Figures 3.4B and 3.4C show typical cyclic voltammograms of two separate surfaces, with two peaks evident, one from the [4Fe4S] cluster of the DNA-bound protein and one from the Redmond Red probe. When compared on these surfaces, all mutants exhibited midpoint potentials of 80 (±30) mV vs. NHE, and all were within error of that of WT (Table 3.2). While the DNA-bound potential for these mutants was seen to be invariant, significant variations in redox signal intensity were observed (Table 3.2 and Figure 3.4D). The signal strength of each mutant was quantified and normalized to the amount of DNA covalently bound on the electrode. To compare the signal strength of each mutant relative to that of WT, these signals were then normalized to the Redmond Red-corrected signal strength of wt EndoIII measured on the same surface. This normalization accounts for the variability in surface coverage associated with different electrodes.

Among the first samples to be examined were mutants relevant to colorectal cancer research, those at positions L81 and Y82. Y82F and Y82W exhibit signal strengths within error of that of WT, suggesting that other aromatic residues at position 82 confer CT capabilities equivalent to those of the native tyrosine. By contrast, Y82S exhibits a very weak electrochemical signal relative to WT (Figure 3.4B, 3.4D).

Figure 3.4. Quantitative cyclic voltammetry of EndoIII variants on DNA-modified electrodes. (**A**) Schematic of DNA-modified electrochemistry. A Au surface is treated with thiol-modified DNA and then backfilled with mercaptohexanol. The covalent Redmond Red redox probe, (red oval), is used to quantify the amount of DNA on the surface. Each protein mutant was allowed to bind DNA on such a surface, and the [4Fe4S] cluster was measured by cyclic voltammetry. The protein was then rinsed from surface and WT was measured on the same surface. (**B**) Representative cyclic voltammogram of Y82S (brown) compared to wt EndoIII (blue). Y82S exhibits a weaker electrochemical signal. For reference, a scan taken in the absence of protein (buffer only) is also shown (gray). (**C**) Representative cyclic voltammogram of W178A (orange) compared to wt EndoIII (blue) and a buffer-only scan (gray). (**D**) The CT capability of each mutant was quantified based on the area under its redox peak. Each mutant's signal intensity was normalized to the intensity of the signal for WT measured on the same surface.

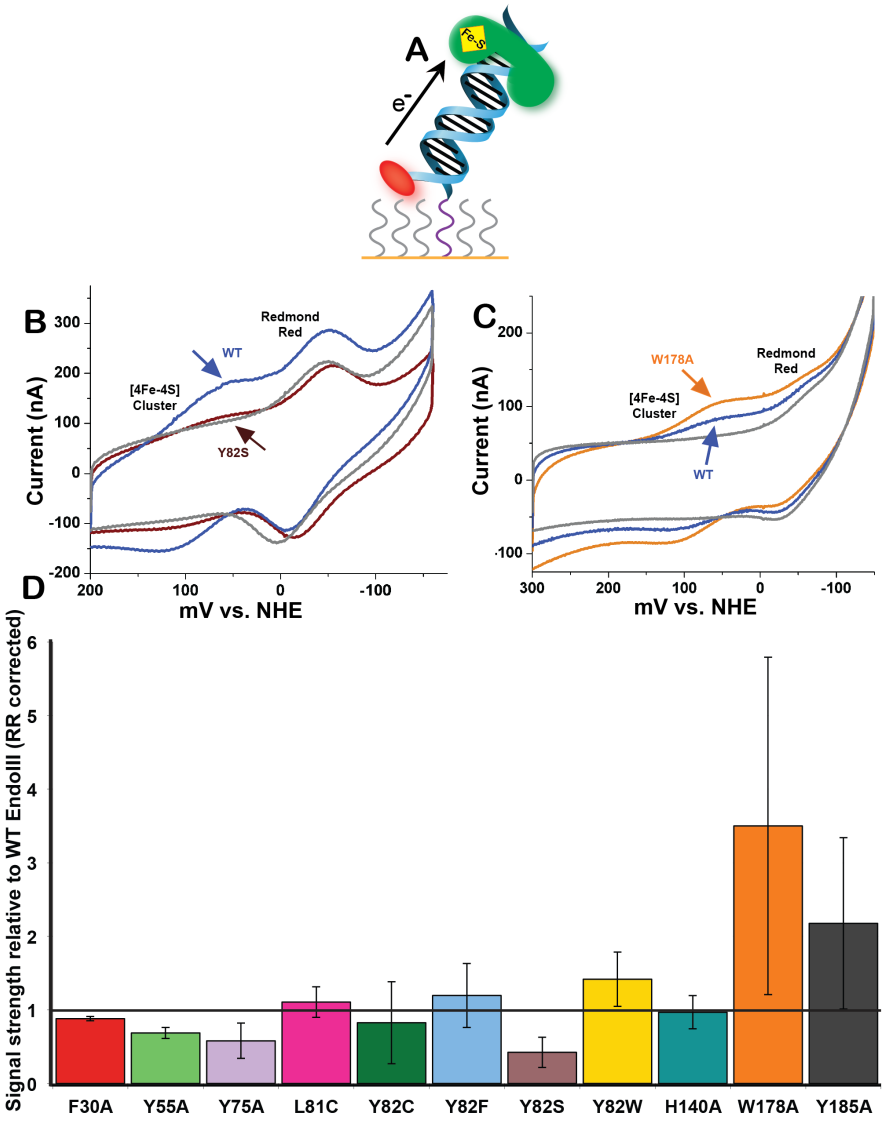


Table 3.2. DNA Electrochemistry of EndoIII Mutants. [a]

EndoIII	Midpoint potential (mV vs. NHE)[b]	Mutant / WT Ratio (RR corrected)
WT	78	
F30A	97	0.88 ± 0.03
Y55A	97	0.7 ± 0.1
Y75A	88	0.6 ± 0.2
L81C	84	1.1 ± 0.2
Y82C	92	0.8 ± 0.6
Y82F	91	1.2 ± 0.4
Y82S	91	0.4 ± 0.2
Y82W	92	1.4 ± 0.4
H140A[c]	71	1.0 ± 0.2
W178A	83	3.5 ± 2.3
Y185A	85	2.2 ± 1.2

a) Experimental conditions are described in materials and methods. Each experiment used 50 μL of 50-100 μM EndoIII in protein buffer (20 mM NaH_2PO_4 pH 7.5, 100 mM NaCl, 5% glycerol, 1 mM EDTA).

b) Measurements have an uncertainty of ± 30 mV.

c) This sample was not measured using Redmond Red-modified DNA, although extra trials were performed to verify the midpoint potential and signal strength ratio that are shown.

The signal intensity from Y82C is on average the same as that of WT, although highly variable, possibly due to the ability of cysteine to facilitate electron transfer in certain contexts (*34, 56, 57*), albeit through the formation of an unstable radical.

After the experiments with Y82 mutants established the importance of aromatic residues for DNA/protein CT, other aromatic amino acids in EndoIII were targeted for mutagenesis studies. F30A, Y55A, Y75A, and H140A were examined. The first three of these mutants all displayed a CT deficiency relative to wt EndoIII (Figure 3.4D). In the protein structure, Y75, Y55, and F30 are relatively close to Y82, and form a line along one side of the protein (*10, 12, 14*). These residues may comprise a pathway of aromatic amino acids through which electrons can travel efficiently. Such aromatic "π-ways" have been found in other peptides (*38, 39*). The H140A mutant, by contrast, proved CT-proficient in the experiments performed here so no conclusions can be made about its electrochemical properties. However, this residue was found to be enzymatically important (Table 3.3).

The final category of EndoIII mutants examined were W178A and Y185A, substitutions involving aromatic residues close to the [4Fe4S] cluster. These mutants were expected to produce weak electrochemical signals because of their proximity to the metallocluster, but they instead produced signals that were large and highly variable relative to WT (Figure 3.4C, 3.4D). One explanation for this phenomenon is protein aggregation, as denatured samples can produce large and erratic signals on DNA-modified electrodes (data not shown). However, because the midpoint potentials of these mutants are within error of those produced by other EndoIII samples, it is unlikely that the [4Fe4S] cluster degraded in W178A and Y185A, given that [3Fe4S] clusters have different potentials (*58-60*) than the [4Fe4S] cluster. These results with W178A and Y185A prompted further experiments, described below, that suggest these residues may help protect the [4Fe4S] cluster from exposure to solvent.

Table 3.3. Summary of Glycosylase Assay Results with EndoIII Mutants.[a]

EndoIII	% Activity relative to WT at 1 μM concentration
F30A	93.7 ± 2.6
Y55A	96.7 ± 1.0
Y75A	99.8 ± 4.0
L81C	99.2 ± 0.6
Y82C	99.2 ± 0.5
Y82F	99.4 ± 1.1
Y82S	99.4 ± 1.6
Y82W	98.1 ± 3.6
H140A	39.1 ± 6.8
W178A	98.2 ± 1.7
Y185A	97.3 ± 0.6

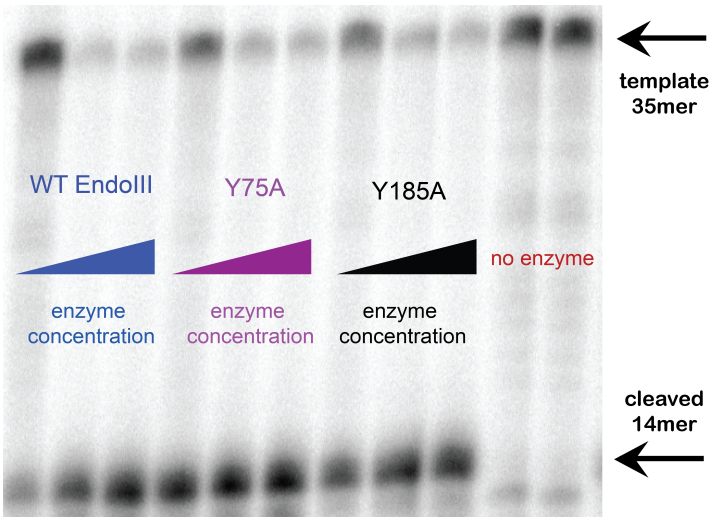
a) Experimental conditions are described in materials and methods. Experiments were conducted using 1 μM protein and 100 nM annealed duplex in reaction buffer (10 mM Tris pH = 7.6, 1 mM EDTA, 50 mM NaCl) for 15 minutes at 37°C. The reactions were quenched upon addition of NaOH to a final concentration of 100 nM. Samples were dried, counted by scintillation, and diluted with loading buffer (80% formamide, 10 mM NaOH, 0.025% xylene cyanol, 0.025% bromophenol blue in Tris-Borate-EDTA buffer) to normalize the radioactivity. Samples were then heated at 90°C for five minutes prior to loading and then separated by denaturing PAGE Glycosylase activity was determined by comparing the amount of 14-mer produced to the total amount of DNA. The glycosylase activity of each mutant as a percentage of WT is shown. Most mutants have activity within error of that of WT.

Enzymatic assays of mutant glycosylase activity The glycosylase activity of each mutant was measured according to established methods (*15, 28*). Briefly, 35-mer strands of DNA were synthesized with 5-hydroxyuracil (5-OH-dU), an EndoIII substrate analogue, incorporated into the sequence. The sequence was 5′-radiolabeled with ^{32}P-ATP and annealed to its complement. Solutions of 100 nM DNA were incubated with 1 μM enzyme. Active enzymes will remove 5-OH-dU from the DNA backbone, leaving an abasic site whose phosphodiester bonds are cleaved with 1 M NaOH. This cleavage produces a 14-mer strand, which can be visualized by denaturing PAGE and autoradiography (Figure 3.5). The amount of 14-mer relative to the total quantity of DNA was used to assess enzymatic activity. Mutants that are glycolytically active bind DNA comparably to WT, while mutants that bind DNA poorly show diminished glycolytic activity (*15-17*). Consequently, this experiment is an important complement to the electrochemical studies, as it verifies that any weak cyclic voltammetry signals produced by mutants stem from a CT deficiency and not an inability to bind DNA.

Of the eleven mutants examined, most were found to exhibit glycosylase activity within error of that of wt EndoIII (Table 3.3). The exceptions were F30A and H140A. F30A was slightly impaired in its glycolytic activity, possibly because F30 is located near the substrate binding pocket of EndoIII (*10, 12, 14*), so mutations at this position may impede DNA binding. H140A is deficient in glycosylase activity relative to WT. In earlier crystallographic studies, the H140 residue was identified as possibly increasing DNA binding by providing a positively charged histidine (*12*). The enzymatic activity results presented here are consistent with that hypothesis.

Circular dichroism to examine the structural stability of EndoIII mutants While most of the mutants examined gave electrochemical signals with intensities weaker than or within error of those of wt EndoIII, the mutants W178A and Y185A were exceptions to this trend.

Figure 3.5. Glycosylase assay results of EndoIII variants. For this assay, 1 μM, 100 nM, or 10 nM protein sample was incubated with 100 nM of 5′-radiolabeled 35-mer duplex DNA containing 5-hydroxy-uracil. Reactions were incubated for 15 minutes at 37°C and quenched with 1 M NaOH. Reactions were then examined by denaturing gel electrophoresis. Cleavage of the DNA results in a 14-mer. The representative gel below shows the results of wt EndoIII (blue) compared to Y75A (purple) and Y185A (black). Enzyme-free control reactions were loaded into the final two lanes. The reactions with 1 μM protein were used to compare the glycosylase activities of EndoIII variants (Table 3.3).



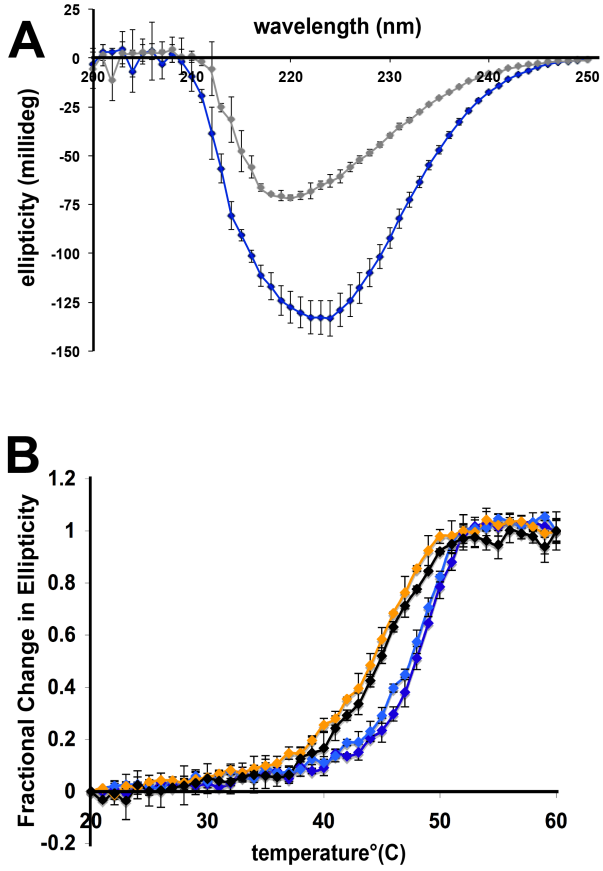
Consequently, CD experiments were performed to examine the secondary structure of W178A and Y185A. The CD spectra of wt EndoIII and several mutants were examined for both fully folded and thermally denatured protein (Figure 3.6). All folded samples exhibited similar spectral shapes.

In order to examine the stability of W178A and Y185A relative to wt EndoIII, thermal denaturation experiments were performed. In these experiments, the protein sample was heated gradually and its ellipticity was measured as a function of temperature. The ellipticity of all variants measured decreased with increasing temperature, but W178A and Y185A denatured at a lower temperature than wt EndoIII and a CT proficient mutant, Y82F (Figure 3.6). These data indicate that W178A and Y185A are less stable than wt EndoIII.

Cooperation between MutY and EndoIII Variants inside the cell This CT scanning model can also be tested *in vivo* by assaying for the cooperation among repair proteins facilitated by DNA-mediated signaling. If these proteins are able to help each other in their search for damage using DNA CT, variants of EndoIII with decreased or increased capability to carry out CT will alter MutY activity *in vivo*. Assays for MutY (CC104) and EndoIII activity inside *E. coli* cells have already been developed (*28*). The assay for "helper function" used is described in reference 28**.**

EndoIII variants W178A, Y82F, and Y82S EndoIII were introduced into the reporter strain CC104 *nth* via independent plasmids to explore whether these mutations attenuate or enhance helper function (Table 3.4). Significantly, W178A, a CT-proficient mutant, in the CC104 *nth*-strain shows a decrease in mutation rate versus the CC104 *nth*-/p controls (Figure 3.7). The number of *lac*+ revertants with W178A decreases to a count within error of CC104 *nth*-/pnth measurements, indicating that W178A is able to help MutY repair damage in the assay.

Figure 3.6. Circular dichroism of EndoIII. (**A**) Comparison of fully folded and denatured spectra of EndoIII. Samples were measured at room temperature at a concentration of ~ 5 μM. WT spectra are shown, although all mutants displayed a similar trend with denatured samples (gray) consistently displaying less ellipticity than fully folded samples (blue). The spectra shown are the average of three trials. (**B**) Circular dichroism thermal denaturation of select EndoIII variants. Samples were measured at concentrations of ~ 5 μM. Each trace shown is the average of at least three independent experiments. The graphs show the fractional change in ellipticity for each variant measured, where the fully folded protein was given a value of 0 and the denatured form was given a value of 1, consistent with previous secondary structure studies (*7*). The variants WT (dark blue) and Y82F (pale blue) display more structural stability than W178A (orange) and Y185A (black).



Alternatively, when comparing CC104 *nth*-/p to CC104 *nth*-/Y82F and CC104 *nth*-/Y82S, the number of *lac*+ revertants increases by a ratio of 1.5 for both mutants. These results clearly indicate that Y82F and Y82S are unable to restore helper function. It is noteworthy that inclusion of Y82F and Y82S EndoIII in CC104 *nth*- strains may not couple to the DNA as well as other mutants, including Y82W, which will be examined in this assay in future studies.

Experimental strategy using AFM to probe protein redistribution We tested our model for protein redistribution stimulated by DNA CT by preparing DNA strands containing a single base mismatch, a modification that inhibits CT, and DNA strands containing no mismatches. Our model would predict the redistribution of proteins onto the mismatched strands. To distinguish the strands by AFM, we made mismatched strands > 1000 base pairs longer than the matched strands. We employed AFM to obtain a snapshot of repair proteins bound to DNA as they search for damage (Figure 3.8).

Long (3.8 kilobase) DNA duplexes containing a site-specific C:A mismatch and short (2.2 and 1.6 kilobases) well-matched duplexes of the same total sequence were therefore constructed. To prepare these DNA samples, pUC19 plasmid was amplified with primers incorporating a 2′-*O*-methyl residue to yield two short duplexes containing 14-nucleotide single-strand overhangs. Upon ligation of these two duplexes, 3.8 kbp strands contain only one C:A mismatch in the middle of the strand (Figure 3.2). The ligation reaction was not taken to completion so as to have a mixture of well-matched short and mismatched long strands for protein distribution studies.

Although a C:A mismatch effectively inhibits DNA CT (*61*), it is not a lesion that is preferentially bound by EndoIII. Gel shift assays on 21-mers with and without a central C:A mismatch show no detectable differences in EndoIII binding (Figure 3.3). Therefore, any differences in protein distribution can be attributed to nonspecific association.

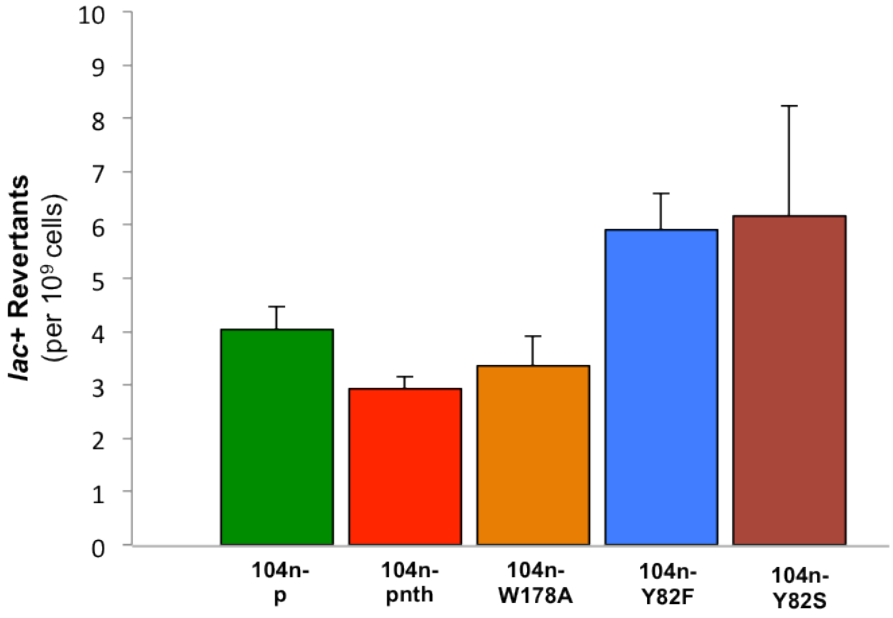
Table 3.4. MutY Activity Assay (CC104) with EndoIII Mutants W178A, Y82F, and Y82S.

Strain	***Lac*+ Revertants**[a,b]	**Increase**
CC104 *nth*-/p	4 ± 0.4	
CC104 *nth*-/pnth	2.9 ± 0.2	0.7
CC104 *nth*-/W178A	3.4 ± 0.6	0.9
CC104 *nth*-/Y82F	5.9 ± 0.7	1.5
CC104 *nth*-/Y82S	6.2 ± 2.1	1.5

a) Lac^+ revertants are reported as the average number of lac^+ colonies that arise per 10^9 cells plated on minimal lactose media.

b) These data represent a triplicate set of experiments (p, pnth, Y82F), a duplicate set (W178A), and a single set of experiments (Y82S) with 10–12 replicates per strain assayed concurrently.

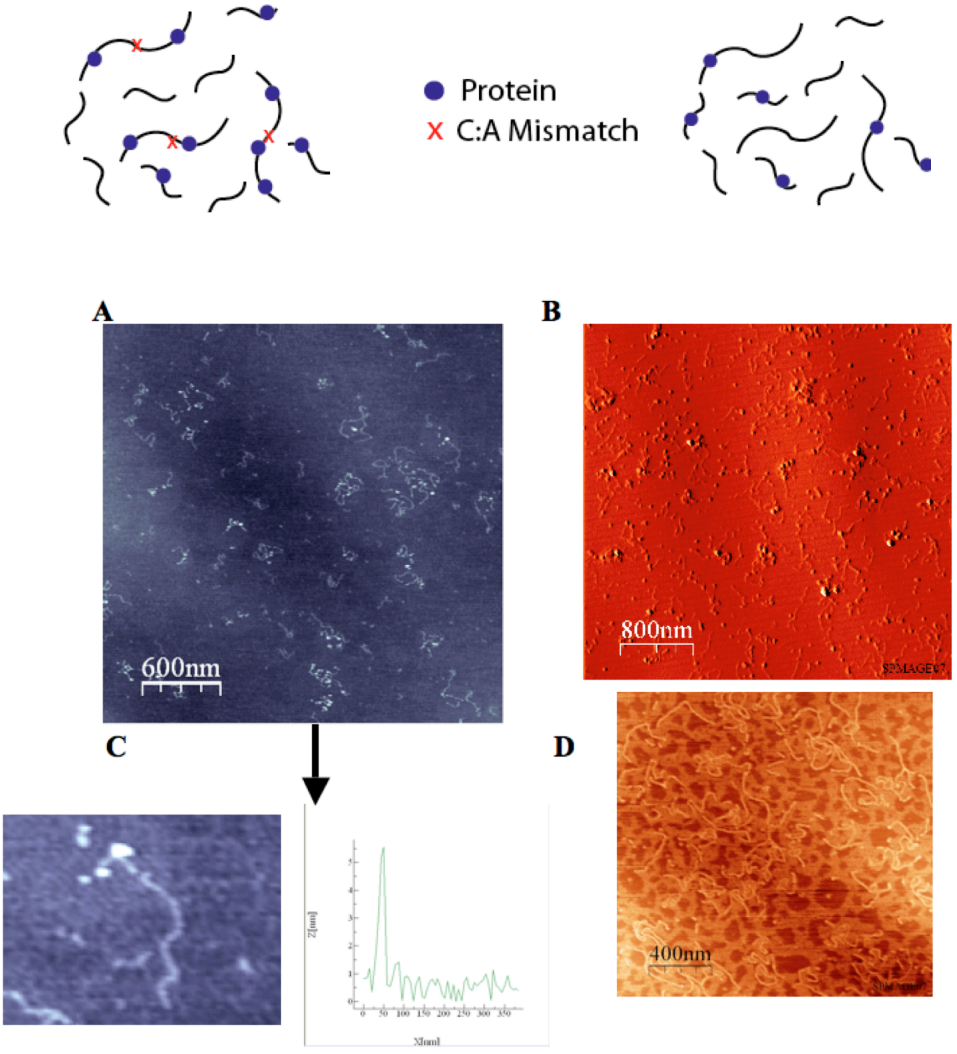
Figure 3.7. Bar graph showing *lac*+ revertants for CC104 *nth*-/p, CC104 *nth*-/pnth, CC104 *nth*-/W178A, CC104 *nth*-/Y82F, and CC104 *nth*-/Y82S strains, where p denotes inclusion of an empty vector. *Lac*+ revertants are reported as the average number of *lac*+ colonies that arise per 10^9 cells plated on minimal lactose media containing ampicillin. Data for the CC104 strains are shown based upon three sets (p, pnth, Y82F), two sets (W178A), and one set (Y82S) of independent experiments, each containing 10–12 replicates per strain.



We determined the proper biochemical conditions for protein/DNA deposition by varying the volume of reaction mixture deposited onto the mica surface (5–10 μL) and the concentration of Mg^{2+} cation in the adsorption buffer (5–10 mM $MgCl_2$). To insure that proteins were not attracted to the mica surface, DNA/protein complexes were formed in solution one day prior to deposition. Some unbound proteins were observed on the mica surface although the concentration of DNA (6 μM bp) was much higher than that of the protein (0.4–0.6 μM) so as to reach binding saturation. Protein assignments were verified through analysis of their 3–5 nm heights in the images. In samples prepared without protein, features of this dimension were not observed, although some larger heights (> 7 nm) were noted and attributed to salt precipitates. Only clearly identifiable long or short strands and bound proteins were counted.

Atomic force microscopy of EndoIII mutants AFM images of DNA-bound protein complexes are represented in Figure 3.8. As observed, protein bound to DNA (Figure 3.8A/B) can be easily distinguished based on strand height profiles of tapping mode images (Figure 3.8C). Images of DNA with/without (Figure 3.8D) protein were acquired with a scan size of 2 x 2 μm^2 or 3 x 3 μm^2 at a scan rate of 3.05 Hz. Since biomolecule surface density varied between samples, > 200 long and short strands were counted in at least three independent trials. Error was calculated based on the total number of proteins counted. Previously, we found that wt EndoIII showed a significant preference for the longer strands containing a C:A mismatch (*28*). The ratio of protein binding densities, r, for (long/short) was 1.6 with 5.6% error (total proteins counted=318). If instead we examined the distribution of EndoIII on long versus short strands, where all strands were matched, we saw a small preference for the short strands; with r (long/short)=0.9. Based on the DNA CT length and/or the length of the DNA over which proteins can diffuse, we would expect a ratio of 1 if there were an equal distribution of proteins on matched and mismatched strands.

Figure 3.8. Atomic force microscopy to visualize protein. Tapping mode AFM images of DNA and EndoIII protein on mica imaged in air. The mixture of mismatched and matched strands (**A**) or fully matched long and short strands (**B**) and protein (dots) are visible on the surface. Images were acquired with a scan size of 2 x 2 μm^2 or 3 x 3 μm^2, at a rate of 3.05 Hz with a data scale of 10 nm/ 0.5 V. In **C** is shown a zoomed-in image of a DNA strand with a bound protein subjected to height profiling to distinguish protein; the profile of the strand, to the right of the zoomed-in image, shows the height of DNA ranges from 1–2 nm, with a much higher protein peak. In **D** is an AFM image taken of the DNA in the absence of protein.



Note that these micrographs also provide a measure of relative binding affinity of the mutants for DNA. All mutants examined in these trials showed similar total amounts of proteins bound per base pair, although variations were found in relative distributions on the long and short strands.

In order to probe the CT role of aromatic residues at the 82 position, Y82F (0.6 μM) and DNA complexes were examined with AFM. Y82F EndoIII, a mutant with CT activity comparable to that of WT (Table 3.2), shows a redistribution onto long mismatched strands with a binding density ratio, r (long/short) of 1.4 ± 0.09 (total proteins counted=212), similar to that of wt protein. We observed 0.15 proteins per kilobase of long strand and 0.10 proteins per kilobase of short strand upon counting 260 long mismatched strands and 346 shorter matched strands. Matched controls were found from counting 250 long matched strands and 313 shorter matched strands. We had earlier tested Y82A, which shows diminished CT efficiency (50% ± 13%) relative to wt EndoIII (*28*). For Y82A, r (long/short) is 0.9 ± 0.1, within error of wild-type EndoIII bound to fully matched long versus short strands. These results led us to mutate other aromatic residues potentially involved in the DNA-protein CT pathway.

Y75A EndoIII mutant exhibits a signal that is 58 ± 20% smaller than that of wt EndoIII (Table 3.2). Significantly, similar to Y82A, examination of the distribution of Y75A on mismatched and matched strands by AFM shows no preference for the mismatched strand. Upon counting 303 long strands and 497 short strands with 5% error, there is no preference for the mismatched strands; Y75A exhibits a protein binding density ratio, r (long/short) of 1.1 ± 0.05 (total proteins counted=495) in mismatched samples, which is within error of fully matched (221 long strands and 296 short strands) density ratio r (long/short) of 0.94 ± 0.06 for protein binding for fully matched strands (221 long strands and 296 short strands). Thus this mutant that is also deficient in DNA CT also does not redistribute in our AFM assay. Though the [4Fe4S] cluster is intact and the protein maintains glycolytic activity, it cannot search for damage efficiently.

In order to probe residues near the [4Fe4S] cluster, residues W178 and Y185 were mutated to alanines. Both mutants displayed DNA-bound electrochemical signals significantly greater than that of wt EndoIII (Table 3.2). When W178A/DNA complexes were studied by AFM, interestingly, they exhibit a very high protein binding density ratio r (long/short) of 1.9 ± 0.09 (total proteins counted=393). More strands were counted in W178A samples than with other proteins: 303 long mismatched strands and 395 short matched strands. 499 long matched strands and 505 short matched strands were counted for controls with < 5% error. We observed 0.25 proteins per kilobase long strand and 0.14 proteins per kilobase short strand in mismatched samples. There are ~ twofold more proteins bound to the longer strands containing one single base mismatch in 3.8kb compared to the well-matched strands. Upon performing comparisons for the three individual samples from both long matched and long mismatched strand images, all measurements are within error of one another. W178A redistributes to an even greater extent than wt EndoIII. This redistribution clearly cannot be explained by binding specificity of EndoIII for the mismatch. Note that all mutant/DNA complexes display heights and strand lengths within error of wt EndoIII/DNA complexes (Figure 3.9A). Figure 3.9B shows binding density ratio comparisons for wt EndoIII and mutants: Y82A, Y75A, Y82F, and W178A. These data indicate a strong correlation between the ability of EndoIII to carry out DNA CT and its localization in the vicinity of mismatches in AFM studies.

Discussion

Biochemical characterization of EndoIII mutants Eleven mutants of EndoIII were prepared and characterized biochemically in order to identify residues important for protein/DNA CT. They were selected based on several criteria including sequence conservation, aromatic character, and possible association with cancer. Additionally, mutations were made in several different regions of the protein to ascertain CT characteristics of different domains (Figure 3.10). When assayed by

cyclic voltammetry on DNA-modified electrodes, all of the mutants display midpoint potentials within error of that of wt EndoIII, indicating an intact [4Fe4S] cluster, irrespective of the mutation. Thus our focus shifted to monitoring changes in signal intensity as a function of mutation, reflecting differences in coupling along the path for DNA/protein CT. The electrochemical studies coupled with AFM experiments provide a measure of CT function: the ability to redistribute repair proteins in the vicinity of a base lesion.

It is therefore remarkable to see the clear correlation between electrochemical signal and the ability of the repair proteins to redistribute in the vicinity of lesions (Figure 3.11). First this correlation provides some confidence that electrochemical signal strength on the DNA-modified electrodes is a useful measure of protein/DNA coupling. But more importantly, this correlation provides a direct link between DNA CT activity and lesion detection, between protein/DNA coupling and function. We can therefore consider various mutants we have prepared in the context of protein/DNA CT (Table 3.5).

Variations at the protein-DNA interface L81 and Y82 are both residues whose MUTYH equivalents exhibit mutations in cancer patients (*25*), and Y82 is well conserved. Additionally, L81 and Y82 are extremely close to the DNA interface in bound EndoIII. This proximity is important when the CT properties of DNA are considered. DNA is able to mediate CT by means of π-stacking, or the overlap of aromatic nucleotides whose many π-bonds provide an efficient conduit for electron flow (*5, 31, 32, 62-66*). L81 and Y82 may be able to intercalate into the DNA helix (*9*). L81 and Y82 would thus join the network of π-bonds in DNA that facilitate electron flow. Consequently, L81 and Y82 were seen as potentially important residues for mediating CT. Earlier research from our laboratory confirmed this hypothesis by establishing that Y82A is CT deficient relative to wt EndoIII (*28*).

Figure 3.9. EndoIII mutant binding density ratio comparisons. (**A**) Representative AFM images of DNA strands incubated overnight with Y82F EndoIII. (**B**) Quantitation of protein density ratios (< 10% uncertainty) where C:A indicates a mismatch is contained in long strands. The unmarked bars show the control measurements of for fully matched long and short strands. Y82A and Y75A, both CT deficient, do not show a redistribution onto the mismatched strand. Y82F and W178A, both CT proficient, redistribute similarly to WT.

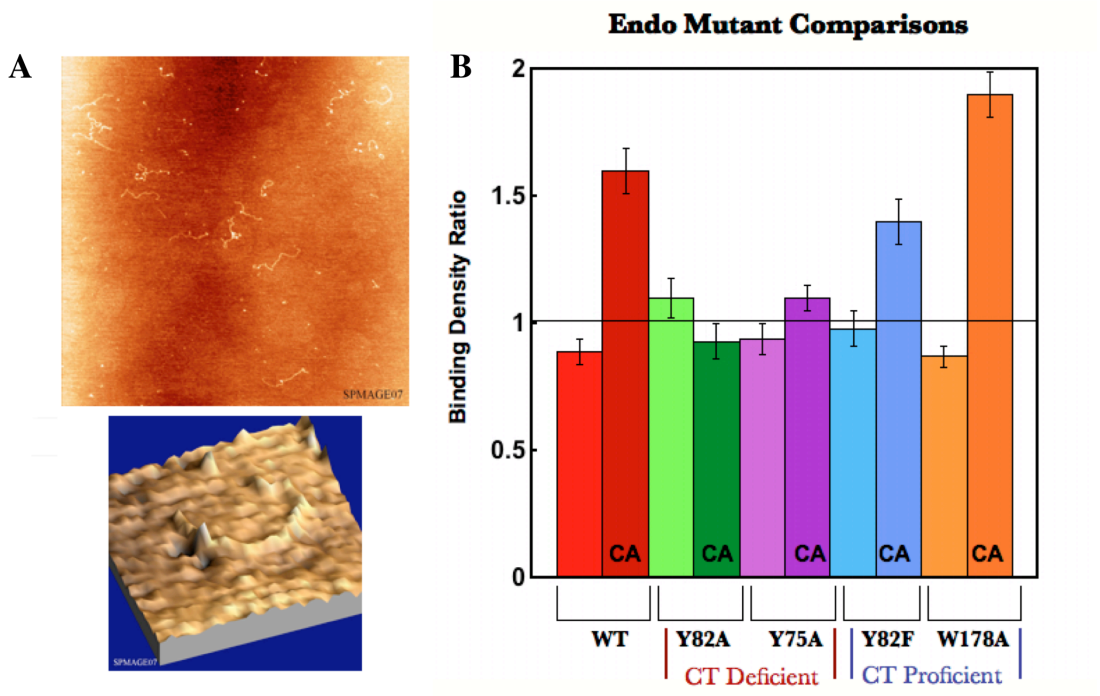


Figure 3.10. Crystal structure of DNA-bound Endonuclease III. The structure shown was adapted from references (*10, 14*) and formatted using PyMol (*75*). Residues targeted for mutagenesis studies are emphasized in color.

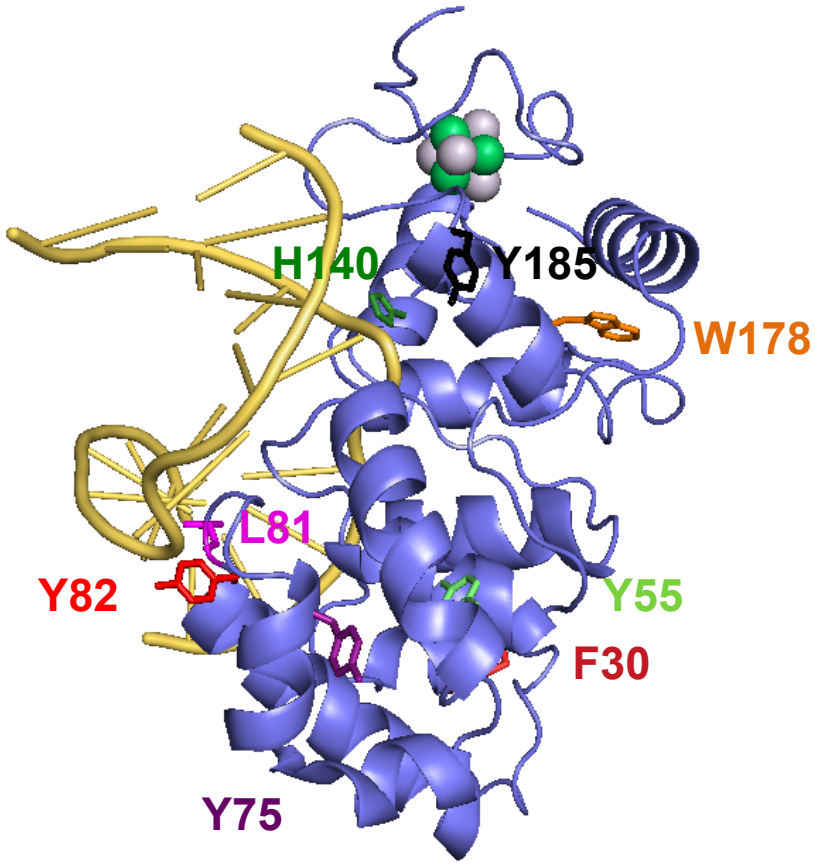


Figure 3.11. Plot of protein binding density ratio versus electrochemical signal. EndoIII mutants (Y82A (green), Y75A (purple), WT (red), Y82F (blue), W178A (orange)) binding density ratios obtained for the mixture of mismatched and matched strands are plotted against electrochemical signals. Mutants that are CT-deficient show low binding density ratios and do not redistribute onto the mismatched strand, while those with particularly high coupling to the DNA electrochemically (albeit with a high uncertainty in the measurement) show high binding density ratios.

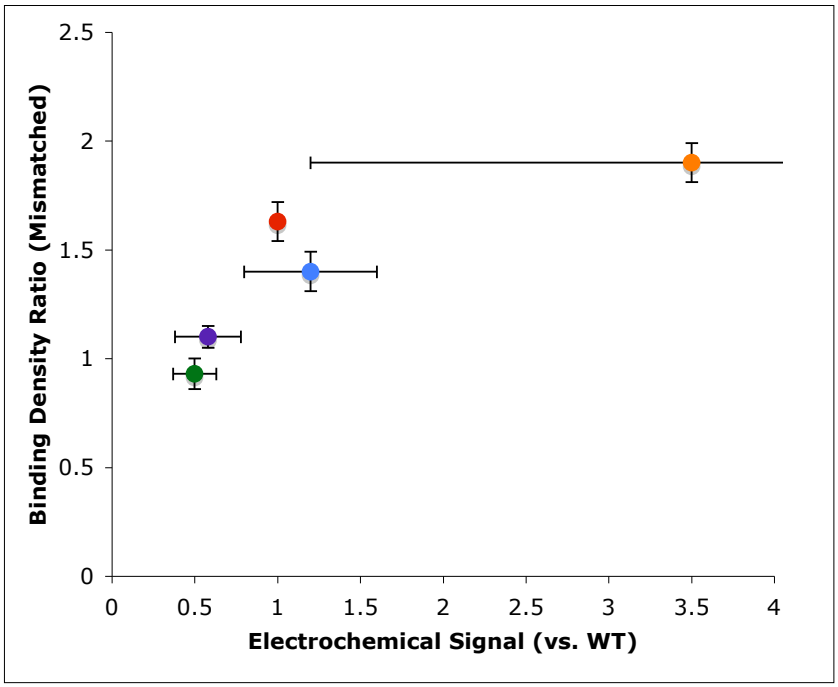


Table 3.5. EndoIII Variant Characterization Comparisons. BDR=Binding density ratio

EndoIII Variant	4Fe4S Cluster?	Glycolytic Activity?	Electrochem. Signal vs. WT	C:A BDR	Total Proteins/bp
WT	+	Yes		1.6 ± 0.09	1.6×10^{-4}
Y82A	+	Yes	0.5 ± 0.07	0.9 ± 0.1	1.2×10^{-4}
Y75A	+	Yes	0.6 ± 0.2	1.1 ± 0.05	2.2×10^{-4}
Y82F	+	Yes	1.2 ± 0.4	1.4 ± 0.09	1.3×10^{-4}
W178A	+	Yes	3.5 ± 2.3	1.9 ± 0.09	2.1×10^{-4}

To follow up on these experiments, additional mutants were made at L81 and Y82 to assess, in more detail, the characteristics of CT-proficient amino acids in EndoIII. L81C is CT-proficient, so no definite conclusions can be made about the electrochemical properties of L81. The mutant Y82S is electrochemically deficient relative to wt EndoIII, whereas Y82F and Y82W are CT proficient, indicating that aromatic amino acids are important mediators of CT at the Y82 position. This CT proficiency with an aromatic residue at position 82 is reinforced by our AFM study where Y82F shows an ability to relocalize onto the mismatched strand that is comparable to WT.

It is interesting to consider mutations at these positions in the context of colorectal cancer and DNA repair. EndoIII bears structural similarity MUTYH. When MUTYH acquires certain mutations, it is no longer able to repair damage in the adenomatous polyposis coli (APC) gene, which regulates the proliferation of colonic cells (*25, 67*). APC is thought to be particularly susceptible to MUTYH activity because it contains 216 "GAA" codons in which a G:C → T:A transversion would result in a stop codon. Genes associated with other cancers do not contain as many such sites (*25*). Several MUTYH mutations have been detected in colorectal cancer patients, including those at positions Y114 and Y166 (*25*). These residues align with F30 and Y82 in EndoIII, respectively. Since certain mutations at Y82 are deficient in CT activity but not enzymatic activity, perhaps the MUTYH mutants are similarly deficient in CT. According to our BER search model, these mutants would be less effective at detecting DNA damage, allowing mutations to accumulate in the APC gene that could lead to colorectal cancer.

Consideration of pathways for DNA-mediated CT Crystal structures of DNA-bound EndoIII place the [4Fe4S] cluster and DNA roughly 14 Å apart at their closest (*10, 12, 14*). At this distance, CT could proceed by either a single-step tunneling process or a multi-step tunneling process, also called hopping, in which amino acids act as "stepping stones" for the electron as it

travels through the protein (*35, 36, 68-70*). In hopping systems, many of the amino acids through which electrons hop are aromatic residues (*36, 68*). With this information in mind, additional mutations were made at aromatic residues closer to the interior of EndoIII to find residues that could be part of a CT pathway. Emphasis was placed on aromatic residues near Y82, as this tyrosine has already been established as CT-active. F30, Y55, and Y75 all display weaker electrochemical signals than wt EndoIII, suggesting they participate in a CT pathway. The distances covered are 9 Å from Y82 to Y75, 8 Å from Y75 to Y55, and 6 Å from Y55 to F30, which are reasonable distances to consider in the context of hopping (*35*). It is interesting to see the effects of mutations at Y75 both electrochemically and by AFM. Y75 is clearly biochemically active as a glycosylase. However, functionally, both in carrying out DNA/protein CT and in redistributing in the vicinity of a mismatch, Y75 is defective. These results support the idea that Y75 is part of a hopping pathway from the cluster to DNA interface. Time-resolved spectroscopy will be necessary to distinguish these pathways. Tunneling efficiency is exponential with distance (*69*), so even a slight reduction in tunneling distance can increase CT rates significantly.

Beyond F30, the remainder of the CT pathway was not probed for several reasons. First, many residues in the region between F30 and the [4Fe4S] cluster are part of the helix-hairpin-helix motif of EndoIII, a well conserved motif in DNA-binding proteins (*14, 71*) that helps these proteins interact with their substrates. Mutations in this region could render the protein unstable and/or unable to bind DNA. Second, mutations in this region of the protein may not exhibit a detectable CT deficiency even if the targeted residues are able to facilitate CT, because multiple CT pathways could be present, as is often the case in redox-active proteins (*70*). In ribonucleotide reductase, for example, electrons are known to find alternative pathways between two redox centers if a primary pathway is impeded by mutagenesis or incorporation of a non-natural amino acid (*36, 37*). Similar phenomena could occur in EndoIII, especially given that all of the mutants examined display at least a partial signal, and that the degree of attenuation lessens as the

mutations are made further from the DNA. The signal from Y82A is 50 % ± 13 % that of WT (*28*), whereas the signal from F30A is 88 % ± 3 % that of WT. These values suggest that protein/DNA CT may involve Y82, but that it could circumvent F30. Third, mutations very close to the [4Fe4S] cluster could compromise the stability of this cluster, as may have been the case with the mutants W178A and Y185A.

Characterization of EndoIII mutants with large electrochemical signals After establishing that mutations in aromatic residues are more likely to impede CT in EndoIII, the mutants W178A and Y185A were expected to produce weak cyclic voltammetry signals. However, their CT activity is unexpectedly high, larger even than that of wt EndoIII. It is more remarkable still that AFM results support this higher coupling of the protein with DNA for CT. This increase in protein/DNA coupling may be attributed to structural changes near the [4Fe4S] cluster of the protein. The loss of the bulky, aromatic residues could increase the conformational flexibility of EndoIII, allowing water molecules to access the protein and accelerate CT to the [4Fe4S] cluster (*72, 73*).

One question that could be raised by this result is why alanine is not present at positions 178 and 185 in the native protein since it permits more efficient CT than tryptophan and tyrosine at these positions. A possible answer to this question is presented by the circular dichroism data. W178A and Y185A are found to be less stable structurally than WT, and the resulting instability or flexibility could affect the nearby [4Fe4S] cluster. Aromatic residues could play a protective role as well. Given the proximity of W178 and Y185 to the [4Fe4S] cluster of EndoIII, it is also possible that these aromatic residues shield the cluster from solvent exposure under physiological conditions. However, these mutants still exhibit WT level enzymatic activity, indicating that the structural perturbations are probably localized.

DNA-mediated search for damage within the genome DNA CT provides a solution to the problem of how these BER proteins can effectively scan the *E. coli* genome despite the low copy number and poor specificity of the repair proteins. In fact, CT through DNA with transfer over only a few hundred base pairs is sufficient to decrease the genome scanning time significantly (*28*). Here, AFM is utilized to probe this model of genome scanning by following the redistribution of proteins near lesions. It is interesting to note that in our AFM experiments only 1–4 proteins are bound to the long (3.8 kbp) strands, suggesting that DNA CT is occurring over distances much greater (> 500 bp) than those we have measured electrochemically (*4*).

Moreover, here, using mutagenesis, electrochemistry and AFM, we have probed possible electronic pathways between DNA and the [4Fe4S] cluster of the repair protein. Significantly, Y82 has already been established as a cancer-relevant aromatic residue located at the protein/DNA binding surface (*28*). Is Y82 the moiety that captures electrons from the DNA: the protein's "toe in the water?" Mutation of W178, interestingly, results in both an increase in DNA CT and efficient redistribution in AFM assays. Likely here, as in other proteins, W178 serves to shield the EndoIII [4Fe4S] cluster, protecting the cluster from solvent and reaction with water and oxygen (*74*). These results certainly support the idea that mutations in repair proteins that alter DNA CT influence the search for damage.

Importantly, the requirement for a rapid, comprehensive search for damage is not a process limited to BER glycosylases. DNA-binding proteins that contain [4Fe4S] clusters with no obvious functional role have been increasingly ubiquitous (*31*). The results presented here elucidate a possible pathway through which charge can travel between the DNA and the iron sulfur cluster of EndoIII. Moreover these results demonstrate that the ability of the protein to perform DNA CT directly correlates with its function as a first step in damage detection. Consequently, these results add further support to the possibility that DNA-bound proteins communicate *in vivo* by means of DNA-mediated long-range signaling.

REFERENCES

1. Ames, B. N., Shigenaga, M. K., and Hagen, T. M. (1993) *Proc. Natl. Acad. Sci. USA 90*, 7915.

2. David, S. S., O'Shea, V. L., and Kundu, S. (2007) *Nature 447*, 941.

3. Demple, B., and Harrison, L. (1994) *Annu. Rev. Biochem. 63*, 915.

4. Slinker, J. D., Muren, N. B., Renfrew, S. E., and Barton, J. K. (2011) *Nat. Chem*. *3*, 230.

5. Nunez, M. E., Hall, D. B., and Barton, J. K. (1999) *Chem Biol. 6*, 85.

6. Boon, E. M., Livingston, A. L., Chmiel, N. H., David, S. S., and Barton, J. K. (2003) *Proc. Natl. Acad. Sci. USA 100*, 12543.

7. Porello, S. L., Cannon, M. J., and David, S. S. (1998) *Biochemistry 37*, 6465.

8. Lukianova, O. A., and David, S. S. (2005) *Curr. Opin. Chem. Biol. 9*, 145.

9. Fromme, J. C., Banerjee, A., Huang, S. J., and Verdine, G. L. (2004) *Nature 427*, 652.

10. Fromme, J. C., and Verdine, G. L. (2003) *EMBO J. 22*, 3461.

11. Guan, Y., Manuel, R. C., Arvai, A. S., Parikh, S. S., Mol, C. D., Miller, J. H., Lloyd, R. S., and Tainer, J. A. (1998) *Nat. Struct. Biol. 5*, 1058.

12. Kuo, C. F., McRee, D. E., Fisher, C. L., O'Handley, S. F., Cunningham, R. P., and Tainer, J. A. (1992) *Science 258*, 434.

13. Manuel, R. C., Hitomi, K., Arvai, A. S., House, P. G., Kurtz, A. J., Dodson, M. L., McCullough, A. K., Tainer, J. A., and Lloyd, R. S. (2004) *J. Biol. Chem. 279*, 46930.

14. Thayer, M. M., Ahern, H., Xing, D., Cunningham, R. P., and Tainer, J. A. (1995) *EMBO J. 14*, 4108.

15. Watanabe, T., Blaisdel, J. O., Wallace, S. S., and Bond, J. P. (2005) *J. Biol. Chem. 280*, 34378.

16. Messick, T. E., Chmiel, N. H., Golinelli, M.-P., Langer, M. R., Joshua-Tor, L., and David, S. S. (2002) *Biochemistry 41*, 3931.

17. Livingston, A. L., Kundu, S., Pozzi, M. H., Anderson, D. W., and David, S. S. (2005) *Biochemistry 44*, 14179.

18. Golinelli, M. P., Chmiel, N. H., and David, S. S. (1999) *Biochemistry 38*, 6997.

19. Chmiel, N. H., Golinelli, M.-P., Francis, A. W., and David, S. S. (2001) *Nucleic Acids Res. 29*, 553.

20. Bai, H. B., Jones, S., Guan, X., Wilson, T. M., Sampson, J. R., Cheadle, J. P., and Lu, A. L. (2005) *Nucleic Acids Res. 33*, 597.

21. Bai, H. B., Grist, S., Gardner, J., Suthers, G., Wilson, T. M., and Lu, A. L. (2007) *Cancer Lett. 250*, 74.

22. Chepanoske, C. L., Lukianova, O. A., Lombard, M., Cohen, M. P. G.., and David, S. S. (2004) *Biochemistry 43*, 651.

23. Chmiel, N. H., Livingston, A. L., and David, S. S. (2003) *J. Mol. Biol. 327*, 431.

24. Al-Tassan, N., Chmiel, N. H., Maynard, J., Fleming, N., Livingston, A. L., Williams, G. T., Hodges, A. K., Davies, D. R., David, S. S., and Cheadle, J. P. (2002) *Nat. Genet. 30*, 227.

25. Cheadle, J. P., and Sampson, J. R. (2007) *DNA Repair 6*, 274.

26. Sampson, J. R., Jones, S., Dolwani, S., and Cheadle, J. P. (2005) *Biochem. Soc. T. 33*, 679.

27. Kundu, S., Brinkmeyer, M. K., Livingston, A. L., and David, S. S. (2009) *DNA Repair (Amst) 8,* 1400.

28. Boal, A. K., Genereux, J., Sontz, P., Gralnick, J. A., Newman, D. K., and Barton, J. K. (2009) *Proc. Natl. Acad. Sci. USA 106*, 15237.

29. Boal, A. K. (2008) DNA-mediated Charge Transport in DNA Repair, Thesis in Chemistry, California Institute of Technology, Pasadena, CA.

30. Boal, A. K., Yavin, E., Lukianova, O. A., O'Shea, V. L., David, S. S., and Barton, J. K. (2005) *Biochemistry 44*, 8397.

31. Genereux, J. C., Boal, A. K., and Barton, J. K. (2010) *J. Am. Chem. Soc. 132*, 891.

32. Genereux, J. C., and Barton, J. K. (2010) *Chem. Rev. 110*, 1642.

33. Merino, E. J., Boal, A. K., and Barton, J. K. (2008) *Curr. Opin. Chem. Biol. 12*, 229.

34. Giese, B., Wang, M., Gao, J., Stoltz, M., Muller, P., and Graber, M. (2009) *J. Org. Chem. 74*, 3621.

35. Shih, C., Museth, A. K., Abrahamsson, M., Blanco-Rodriguez, A. M., Di Bilio, A. J., Sudhamsu, J., Crane, B. R., Ronayne, K. L., Towrie, M., Vlcek, A., Richards, J. H., Winkler, J. R., and Gray, H. B. (2008) *Science 320*, 1760.

36. Stubbe, J., Nocera, D. G., Yee, C. S., and Chang, M. C. Y. (2003) *Chem. Rev. 103*, 2167.

37. Chang, M. C., Yee, C. S., Nocera, D. G., and Stubbe, J. (2004) *J. Am. Chem. Soc. 126*, 16702.

38. Giese, B., Graber, M., and Cordes, M. (2008) *Curr. Opin. Chem. Biol. 12*, 755.

39. Cordes, M., Kottgen, A., Jasper, C., Jacques, O., Boudebous, H., and Giese, B. (2008) *Angew. Chem. Int. Edit. 47*, 3461.

40. Sun, H. B., and Yokota, H. (2000) *Anal. Chem. 72*, 3138.

41. Sorel, I., Piétrement, O., Hamon, L., Baconnais, S., Le Cam, E., and Pastré, D. (2006) *Biochemistry 45*, 14675.

42. Pastré, D., Hamon, L., Sorel, I., Le Cam, E., Curmi, P. A., and Piétrement, O. (2010) *Langmuir 26*, 2618.

43. Hansma, H. G. (2001) *Annu. Rev. Phys. Chem. 52*, 71.

44. Allen, S., Rigby-Singleton, S. M., Harris, H., Davies, M. C., and O'Shea, P. (2003) *Biochem. Soc. Trans. 31*, 1052.

45. Buzzeo, M. C., and Barton, J. K. (2008) *Bioconj. Chem*. *19*, 2110.

46. Kelley, S. O., Boon, E. M., Barton, J. K., Jackson, N. M., and Hill, M. G. (1999) *Nucleic Acids Res. 27*,4830.

47. Kelley, S. O., Jackson, N. M., Hill, M. G., and Barton, J. K. (1999) *Angew. Chem. Int. Ed. 38*, 941.

48. Boon, E. M., Salas, J.E., and Barton, J.K. (2002) *Nat. Biotechnol. 20*, 282.

49. Cunningham, R. P., Asahara, H., Bank, J. F., Scholes, C. P., Salerno, J. C., Surerus, K., Münck, E., McCracken, J., Peisach, J., and Emptage, M. H. (1989) *Biochemistry 28*, 4450.

50. Kovach, M. E., Elzer, P. H., Hill, D. S., Robertson, G. T., Farris, M. A., II, R. M. R., and Peterson, K. M. (1995) *Gene 166*, 175.

51. Sambrook, J., and Russell, D. W. (2001) Molecular cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press.

52. Davis, R. W., Botstein, D., Roth, J. R. (1980) Advanced Bacterial Genetics: A Manual for Genetic Engineering (Cold Spring Harbor Laboratory Press).

53. Donahue, W. F., Turczyk, B. M., and Jarrell, K. A. (2002) *Nucleic Acids Res. 30*, e95.

54. Horcas I., Fernández, R., Gómez-Rodríguez, J. M., Colchero, J., Gómez-Herrero, J., and Baro, A. M. (2007) *Rev. Sci. Instrum. 78*, 013705.

55. Bard, A. J., and Faulkner, L. R. (2001) Electrochemical Methods Wiley, New York, 2nd ed.

56. Wang, M., Gao, J., Muller, P., and Giese, B. (2009) *Angew. Chem. Int. Edit. 48*, 4232.

57. Bhattacharjee, S., Deterding, L. J., Jiang, J., Bonini, M. G., Tomer, K. B., Ramirez, D. C., and Mason, R. P. (2007) *J. Am. Chem. Soc. 129*, 13493.

58. Camba, R., and Armstrong, F. A. (2000) *Biochemistry 39*, 10587.

59. Beinert, H. (2000) *J. Biol. Inorg. Chem. 5*, 409.

60. Beinert, H. (1997) *Science 277*, 653.

61. Boon, E. M., Ceres, D. M., Drummond, T. G., Hill, M. G., and Barton, J. K. (2000) *Nat. Biotechnol. 18*, 1096.

62. Giese, B. (2002) *Annu. Rev. Biochem. 71*, 51.

63. Kelley, S. O., and Barton, J. K. (1999) *Science 283*, 375.

64. Kelley, S. O., Holmlin, R. E., Stemp, E. D. A., and Barton, J. K. (1997) *J. Am. Chem. Soc. 119*, 9861.

65. Holmlin, R. E., Dandliker, P. J., and Barton, J. K. (1997) *Angew. Chem. Int. Edit. 36*, 2715.

66. Giese, B. (2002) *Curr. Opin. Chem. Biol. 6*, 612.

67. Fearnhead, N. S., Britton, M. P., and Bodmer, W. F. (2001) *Hum. Mol. Genet. 10*, 721.

68. Gray, H. B., and Winkler, J. R. (2005) *Proc. Natl. Acad. Sci. USA 102*, 3534.

69. Gray, H. B., and Winkler, J. R. (2003) *Q. Rev. Biophys. 36*, 341.

70. Onuchic, J. N., Beratan, D. N., Winkler, J. R., and Gray, H. B. (1992) *Annu. Rev. Bioph. Biom. 21*, 349.

71. Doherty, A. J., Serpell, L. C., and Ponting, C. P. (1996) *Nucleic Acids Res. 24*, 2488.

72. Berghuis, A. M., Guillemette, J. G., Smith, M., and Brayer, G. D. (1994) *J. Mol. Biol. 235*, 1326.

73. Casimiro, D. R., Richards, J. H., Winkler, J. R., and Gray, H. B. (1993) *J. Phys. Chem. 97*, 13073.

74. Biological Inorganic Chemistry Structure and Reactivity (2007), ed. Bertini, I., Gray, H. B., Stiefel, E. I., Valentine, J. S. (University Science Books–California).

75. The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.

CHAPTER 4

Monitoring MutY Activity: Interactions Among Base Excision Repair Proteins *In Vivo* in the Presence of Oxidative Stress

Introduction

Given that CT chemistry can occur over long molecular distances and is sensitive to intervening mismatches and lesions that perturb the base pair stack, we have begun to explore how DNA CT might be exploited within the cell (*1-12*). Specifically, we have examined proteins from the base excision repair (BER) pathway, the major network cells use to find and repair oxidative damage in the genome (*12*). Significantly, the BER enzymes contain a [4Fe4S] cluster that is not required for protein structure or folding (*13, 14*). We have established that these proteins exhibit DNA-bound redox potentials of ~ 100 mV vs. NHE, which is biologically relevant (*10-12*). Additionally, as BER proteins bind to DNA, the [4Fe4S] cluster is activated toward oxidation, and, necessarily, the protein has a > 1000-fold higher affinity for the duplex (*12*). On the basis of electrochemical, EPR, transient absorption, and genome scanning calculations, we have developed a model that describes how these proteins may cooperate to recognize lesions in an efficient manner (*12*). According to the model we have proposed, charge migrates between two redox-active proteins bound to DNA, thus any proteins with similar redox potentials could participate in the search for damage. We have used genetic techniques to assay for cooperation among repair proteins from *E. coli*, Endonuclease III and MutY, *in vivo* (Figure 4.1) (*12*). This assay for "helper function" employs a strain, CC104, which contains a single cytosine to adenine substitution in the *lacZ* Glu 461 codon (*lac*$^-$*)* (*15-18*). *E. coli* strains expressing this mutant form of β-galactosidase cannot use lactose for energy, and therefore cannot grow on lactose media (*15-18*). Since MutY can prevent GC to TA transversions by excising adenine mispaired with 8-oxo-G (Figure 4.2), reversion of the CC104 strain from *lac*$^-$ to *lac*$^+$ is indicative of a decrease in the ability of MutY to find and repair its substrate (*15-18*). When EndoIII is knocked out of the CC104 strain (*nth*$^-$) or a mutant of EndoIII (Y82A) that cannot perform DNA CT is introduced on a plasmid, MutY activity decreases (*12*). If, instead, a mutant of EndoIII (D138A) that is glycolytically inactive yet exhibits a redox potential ~ 100 mV vs. NHE is

added, MutY activity is restored (*12*). Results from the assay for helper function indicate that EndoIII and MutY help one another in the search for damage using DNA-mediated CT. We are currently extending this method to evaluate the abilities of other EndoIII mutants to cooperate with MutY *in vivo*. We have used a variety of oxidizing agents (e.g., methyl viologen, $Rh(phi)_2bpy^{3+}$) and anaerobic methods to generate additional lesions that must be repaired by MutY to enhance the overall signal in the assay. Importantly, this assay allows us to evaluate whether specific proteins, that we identify individually, cooperate with MutY *in vivo*.

Further, we have begun developing a high-throughput screen to identify proteins, like EndoIII, may exploit DNA-mediated CT to help MutY scan the genome to find sites of damage. The methodology for this screen is inspired by the helper function assay. We have designed and evaluated a 96-well microtiter plate-based screen to examine single-gene knockout strains from the KEIO knockout collection (*19*). Since β-galactosidase is not endogenously expressed in these strains, each strain is transformed with a reporter plasmid that contains *lacZ* from the CC104 strain. Then, after a period of growth, the cells are lysed and assayed with a fluorescent substrate to detect total β-galactosidase activity. Differences in cell growth between samples are taken into account by assaying for total cellular protein. Strains with high β-galactosidase activity are associated with high transversion rates, which indicates one of two possibilities: i) that the specific protein that was knocked helps MutY locate its target lesion via DNA-mediated CT signaling or ii) it interferes with expression of MutY or affects the repair of 8-oxo-guanine. Our method to identify the former case requires restoring the activity of MutY in these strains using a plasmid containing EndoIII D138A. If MutY activity is restored, this suggests DNA CT signaling rather than some other metabolic change being responsible for the increase in revertants. Guided by these principles, we have performed controls and initial studies to examine the feasibility of the high-throughput screen.

Figure 4.1 A schematic illustrating the helper function assay (*in vivo*), which is used to monitor MutY activity. MutY contains a [4Fe4S] cluster (green cube) and repairs A:8-oxo-G lesions. MutY is able to perform DNA CT to/from other proteins that have similar DNA-bound redox potentials, and could cooperate to search for lesions (1). When one of the helper proteins (i.e., EndoIII) is knocked out, MutY cannot find lesions as efficiently, and an increase in unrepaired lesions (red) is observed in the assay (2).

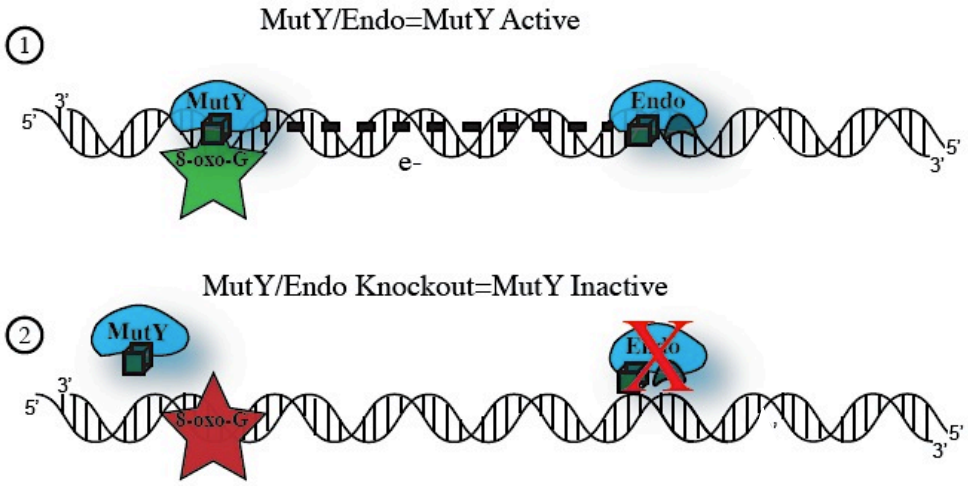
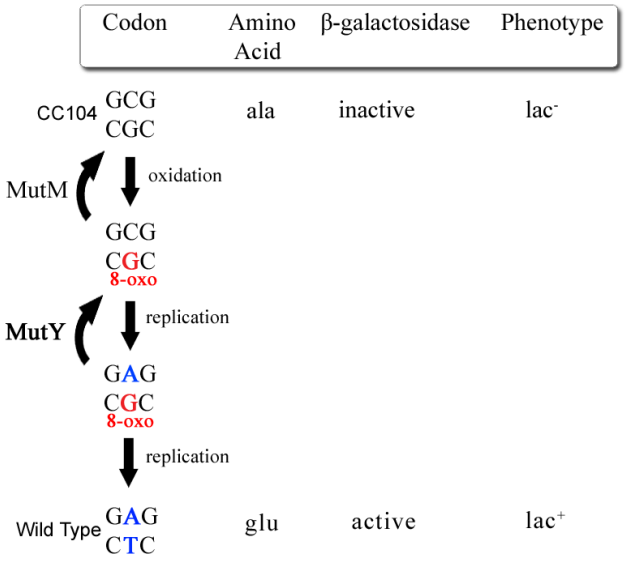


Figure 4.2. Reversion from the CC104 strain to wild-type depends on formation of an oxidized guanine product (8-oxo-G, red), misincorporation of an adenine opposite this lesion (blue), and a second round of replication to revert the point mutation. MutM excises 8-oxo-G from DNA, and MutY excises adenine mispaired to the 8-oxo-G lesion.



Materials and Methods In the experiments conducted, all strains were first streaked to selective media. Independent colonies of each strain were grown to a density of 10^9 cells/mL in minimal media (NCE) supplemented with glucose. These cells were then plated on NCE media supplemented with lactose and incubated for 36 hours at 37°C. The resulting *lac+* revertants are reported as the average ± the standard error per 10^9 cells/mL plated (if 7 or more replicates were tested, the two highest and lowest values are removed). In experiments where plasmids were utilized all media was supplemented with ampicillin (40 μg/mL in NCE, 100 μg/mL in LB). Luria-Bertani (LB) broth was used as the rich medium while NCE (*20*) medium supplemented with $MgSO_4$ (100 μM) and glucose (11 mM) or lactose (6 mM) was used as the minimal medium. Minimal Media A (MMA) (*21*) was supplemented with 46 mM K_2HPO_4, 33 mM KH_2PO_4, 7.5 mM $(NH_4)_2SO_4$, 1.7 mM sodium citrate • 2 H_2O, 5 mL 20% glucose, 800 μM $MgSO_4$, and 0.5 mL 1% B1 (thiamine hydrochloride). L-rhamnose (Fluka) was prepared such that the final concentration is 1 mM (0.5 mL of a sterile 20 mM stock).

Lac$^+$ reversion assays The protocol was adapted from methods described previously (*12*). Strains were streaked to LB medium and incubated overnight at 37°C. For *nth* knockouts, strains were streaked to LB+chloramphenicol (17 μg/mL). 1 mL LB cultures were started from single colonies and grown overnight in a shaking incubator at 37°C, 220 rpm. 20 μL of each starter culture was used to inoculate a 10 mL NCE+glucose culture which was then grown to a density of 10^9 cells/mL at 37°C, 250 rpm. 5 mLs of this culture was centrifuged in a clinical tabletop centrifuge at 4°C and plated on NCE+lactose solid medium and then incubated at 37°C for 36 hours. Colonies arising are reported as *lac$^+$* revertants/mL cells plated. In experiments incorporating plasmid vectors, CC104 were made electrocompetent and transformed via electroporation at 1.8 kV. Transformants were selected on LB+ampicillin (amp) (100 μg/mL) solid medium after incubation overnight at 37°C. Single colonies were restreaked to LB+amp (100 μg/mL) solid media and incubated 12 hours at

37°C. 1 mL LB+amp (100 μg/mL) cultures were started from these colonies and grown overnight at 37°C, 220 rpm. 20 μL of each starter culture was used to inoculate a 10 mL NCE+glucose+amp (40 μg/mL) culture which was grown to a density of 10^9 cells/mL at 37°C, 250 rpm. 5 mLs of this culture was plated onto NCE+lactose+amp (40 μg/mL) and incubated at 37°C for 36 hours.

Lac$^+$ reversion assays with Oxidative Stress: Rh(phi)$_2$bpy^{3+} Cell strains were streaked onto LB solid medium and incubated overnight at 37°C. For *nth* knockouts, strains were streaked to LB+chloramphenicol (17 μg/mL). For treated samples, 50 μM Rh(phi)$_2$bpy^{3+} complex was added to LB, and 1 mL cultures were started from single colonies. Cultures were grown overnight in a shaking incubator at 37°C. Cultures were spun down and resuspended in 2 mL NCE+glucose minimal media in a 24-well plate. Cells were irradiated at 365 nm using a solar simulator (340 nm internal filter) for 1 hour. Cells were grown to an OD at 600 nm of 0.6–0.7 (with Rh) or 2 (controls without Rh). 500 μL of the solution was spun down and plated on NCE+lactose solid medium. No cells grew on NCE+glucose solid medium. Upon irradiation, the temperature of the samples did not get above 26° C.

Lac$^+$ reversion assays with Oxidative Stress: Methyl viologen Cell strains were streaked onto LB medium and incubated overnight at 37°C. For *nth* knockouts, strains were streaked to LB+chloramphenicol (17 μg/mL). 1 mL LB cultures were started from single colonies and grown overnight in a shaking incubator at 37°C. 20 μL of each starter culture was used to inoculate a 10 mL NCE+glucose culture. After two hours, cultures were treated with methyl viologen (Sigma) (50 or 100 μM sterile-filtered in water) and grown to for 6 hours at 250 rpm. Cells grown in the presence of methyl viologen did not grow to the same OD (600 nm) as the control strains (0.1 and

0.6, respectively). 5 mLs of these cultures were plated onto NCE+lactose solid medium and incubated at 37°C for 40 hours.

Growth Curves of strains under oxidative stress (CC104, CC104 + Rh(phi)$_2$bpy^{3+}, and CC104 + methyl viologen) CC104 strain was streaked onto LB medium and incubated overnight at 37°C. Single colonies were used to start 1 mL LB cultures. A 1:100 inoculation was performed into LB and with (50 μM Rh(phi)$_2$bpy^{3+} oxidant or 50 μM methyl viologen) to track growth. Cells were grown in a shaking incubator at 220 rpm and OD measurements (600 nm) were made every 30 minutes. Samples that had been treated with methyl viologen grow at the same rate as untreated cells over ~ 6 hours.

Lac$^+$ reversion assays (CGSC CC104, CC104 nth- strains) Cell strains were streaked onto LB medium and incubated overnight at 37°C. For *nth* knockouts, strains were streaked to LB+chloramphenicol (17 μg/mL). A colony was then chosen for a 1 mL LB starter culture (grown for 12 hours in a shaking incubator). 20 μL of the starter culture was used to inoculate 10 mL NCE+glucose culture which was grown for 8 hours. OD measurements at 600 nm were recorded. 5 mL of these cultures was plated onto NCE+lactose solid medium and incubated at 37°C for 36 hours. The experiment was repeated using anhydrous $MgSO_4$; however for all of the experiments that follow, heptahydrate $MgSO_4$ was used as the standard reagent.

Lac$^+$ reversion assays under anaerobic conditions (argon) Cell strains were streaked onto LB medium and incubated overnight at 37°C. For *nth* knockouts, strains were streaked to LB+chloramphenicol (17 μg/mL). 1 mL LB cultures were started from single colonies, purged for one minute with argon gas, capped, and grown overnight in a shaking incubator at 37°C (Figure 4.3). 20 μL of each starter culture was used to inoculate a 10 mL NCE+glucose culture which was

grown to a density of 10^9 cells/mL at 37°C, 250 rpm. 5 mL of this culture was plated onto NCE+lactose and incubated at 37°C for 36 hours.

Lac$^+$ reversion assays under anaerobic conditions (glovebox + methyl viologen) Similar to the protocol described above, cell strains were streaked onto LB medium and incubated overnight at 37°C. For *nth* knockouts, strains were streaked to LB+chloramphenicol (17 μg/mL). 1 mL LB cultures were started from single colonies, and these cultures were grown for 36 hours in an anaerobic incubator glovebox (10% CO_2, 10% H_2, and 80% N_2) located in the Mazmanian lab at the California Institute of Technology (Figure 4.3). 20 μL of each starter culture was used to inoculate a 10 mL NCE+glucose culture which took much longer to grow to a density of 10^9 cells/mL. Half of the NCE+glucose cultures were treated with 50 μM methyl viologen (Sigma) that was sterilized prior to use. The cells treated with methyl viologen reached an OD at 600 nm of only 0.4, even after 23 hours, but they did not grow in the assay.

Lac$^+$ reversion assays with methyl viologen (pelleting method in NCE or MMA) Strains were streaked to LB medium and incubated overnight at 37°C. For *nth* knockouts, strains were streaked to LB+chloramphenicol (17 μg/mL). 1 mL LB cultures were started from single colonies and grown overnight in a shaking incubator at 37°C, 220 rpm. 20 μL of each starter culture was used to inoculate a 10 mL LB culture with/without the addition of methyl viologen (sterile). Cells were placed back in the shaking incubator for 1–1.5 hours. Samples were then centrifuged and resuspended in 10 mL NCE+glucose culture which was then grown to a density of 10^9 cells/mL at 37°C (~ 5 hours), 250 rpm (*22*). Cell density was determined by dilution plating a 10 mL aliquot of the NCE+glucose culture onto NCE+glucose solid medium followed by incubation at 37°C for 24 hours.

Figure 4.3. Schematic illustrating anaerobic methods employed to increase oxidative stress. (Top) Cells are bubbled with argon and capped prior to inoculating minimal media with starter culture, spinning down, and plating onto minimal media/lactose plates. (Bottom) The cells are cultured inside a glovebag with a 10% H_2, 10% CO_2, and 80% N_2 atmosphere and transferred to minimal media. Prior to spinning the cells down and plating them on minimal media, some cultures are treated with methyl viologen to induce additional stress.

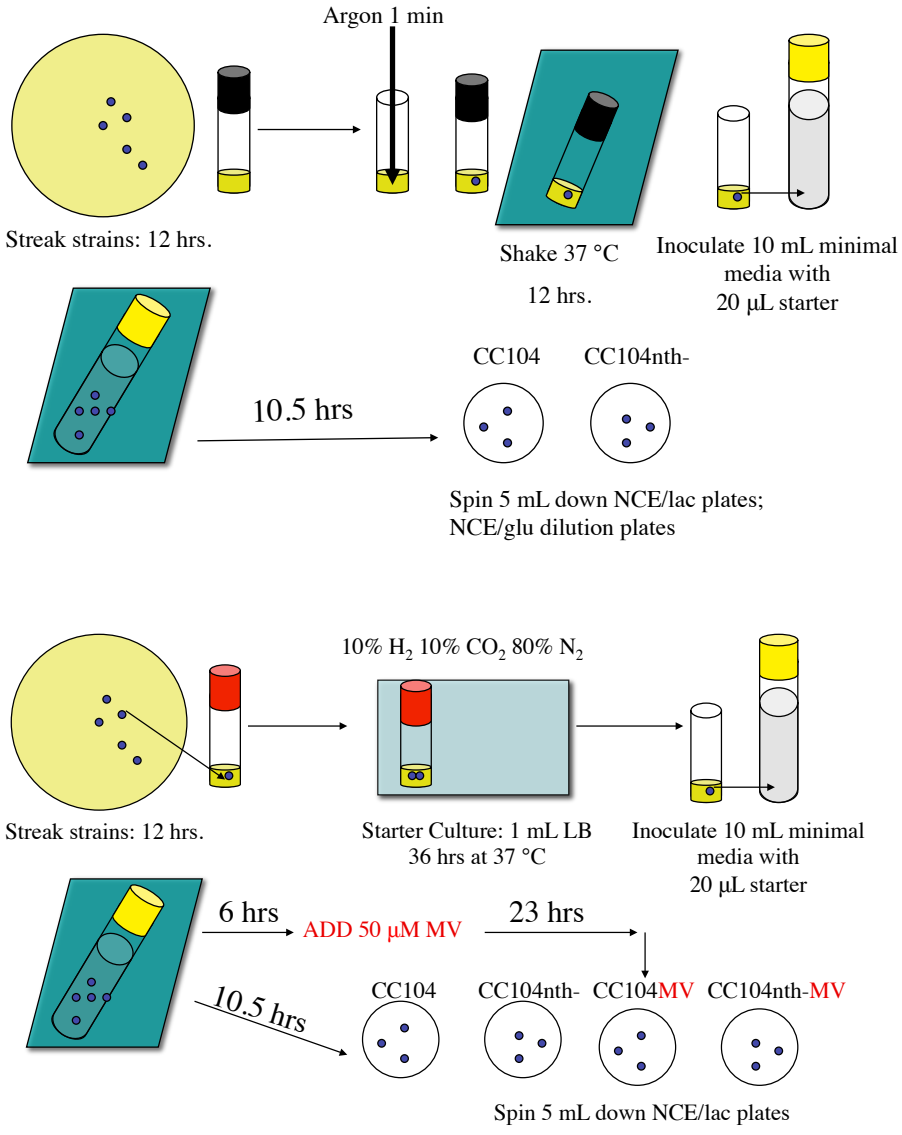
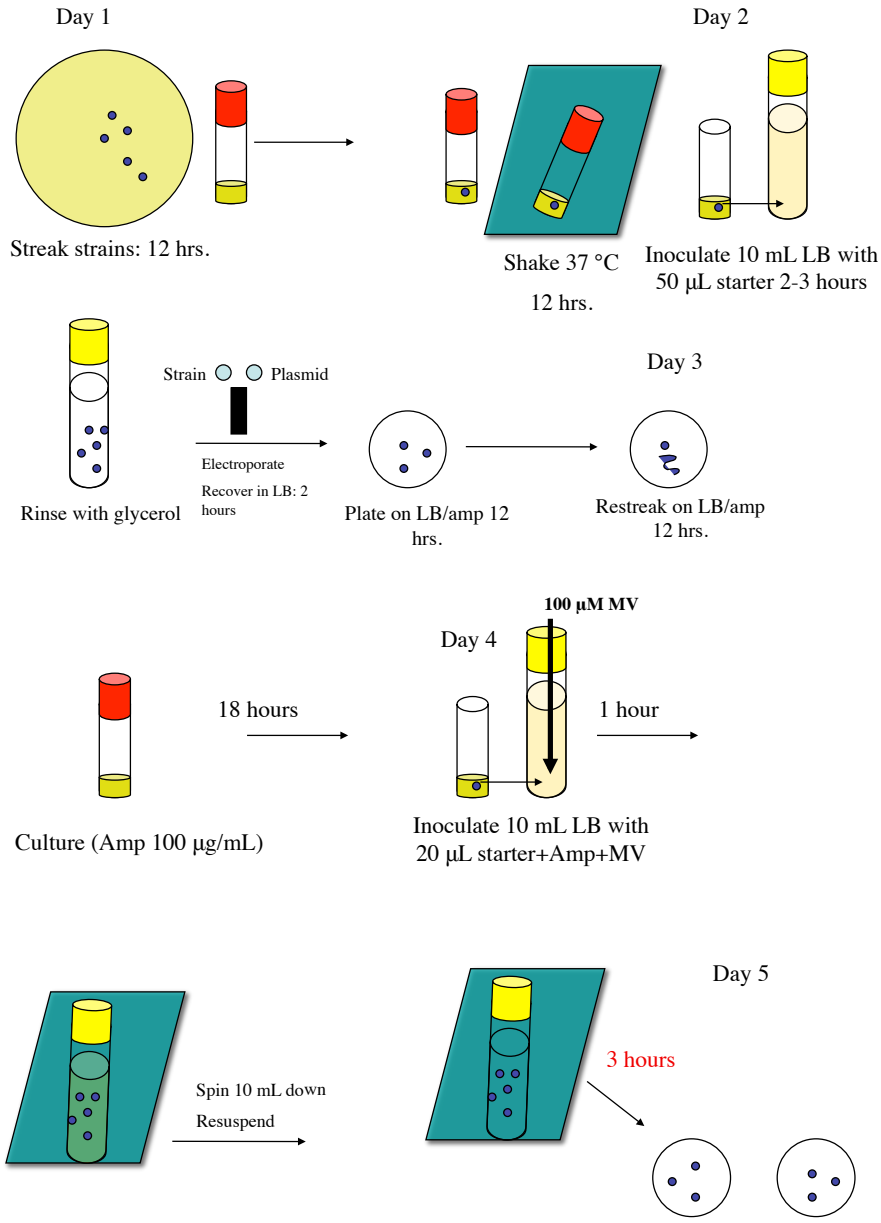


Figure 4.4. Schematic of helper function assay method with plasmid transformation. Strains are streaked onto media (Day 1). Colonies are incubated in LB for 12 hours to generate a starter culture. Following this, 10 mL LB is inoculated with the starter culture (Day 2) and set to shake for 2–3 hours. Cells are pelleted and prepared for plasmid transformation (glycerol). Upon electroporation, cells are recovered in LB for 2 hours and plated onto LB/amp. Once the cells are restreaked (Day 3), they are grown to isolate the colonies that have successfully been transformed with plasmids. Colonies are chosen and cultured in LB/amp for 18 hours. 10 mL LB/amp to which methyl viologen has been added is inoculated with the starter culture (Day 4) and shaken for 1 hour. Cells are spun down and resuspended in minimal media. 5 mL of the culture is spun down and plated onto minimal media plates for revertant counts (Day 5).



5 mL of this culture was centrifuged in a clinical tabletop centrifuge at 4°C and plated on NCE+lactose solid medium and then incubated at 37°C for 36 hours. Methyl viologen was prepared by dissolving 30 mg of the compound in 600 μL (50 μM) water and sterile-filtered into a sterile eppendorf prior to use.

In experiments incorporating plasmid vectors, CC104 strains were made electrocompetent and transformed via electroporation at 1.8 kV. Transformants were selected on LB+ampicillin (amp) (100 μg/mL) solid medium after incubation overnight at 37°C. Single colonies were restreaked to LB+amp (100 μg/mL) solid media and incubated 12 hours at 37°C. 1 mL LB+amp (100 μg/mL) cultures were started from these colonies and grown 18 hours at 37°C, 220 rpm. 20 μL of this starter culture was used to inoculate a 10 mL LB+ amp (100 μg/mL) culture with/without the addition of methyl viologen (sterile). Cells were placed back in the shaking incubator for 1–1.5 hours. Samples were then centrifuged and resuspended in 10 mL NCE or MMA+glucose+amp (40 μg/mL) culture which was then grown to a density of 10^9 cells/mL at 37°C (~ 3 hours), 250 rpm. Cell density was determined by dilution plating a 10 mL aliquot of the NCE or MMA+glucose+amp culture onto NCE or MMA+glucose+amp solid medium followed by incubation at 37°C for 24 hours. 5 mLs of this culture was plated onto NCE or MMA+lactose+amp (40 μg/mL) and incubated at 37°C for 36 hours. *Lac*$^+$ revertants were reported as described above.

Growth Curves (wild-type CC104, CC104 nth-, and CC104 Δnth) A new EndoIII knockout strain, CC104 Δ*nth* was prepared by Christine Romano (*23*). CC104, CC104 *Δnth* strains were streaked onto LB plates and CC104 *nth-* to LB+chloramphenicol (17 μg/mL). 1 colony was chosen for a 1 mL starter culture in LB (3 replicates). A 1:100 inoculation was performed into LB to track growth. Cultures were grown in a shaking incubator at 37 °C, 220 rpm, and the OD at 600 nm was measured every 30 minutes. Strains, including the EndoIII reinsertion CC104 *Δnth* WTnth that

Christine prepared, were also examined in the presence of inducer, rhamnose, and oxidative stress. Cells were streaked onto LB plates and incubated for 12 hours. Starter cultures of 1 mL LB were made and grown overnight at 37°C, 220 rpm. 19 mL LB (with 1 mM rhamnose, 50 μM methyl viologen) was inoculated with 200 μL starter culture. OD measurements were taken every 30 minutes over ~ 800 minutes. The strains grew at identical rates.

Lac$^+$ reversion assays (CC104, CC104 Δnth, CC104 ΔnthWTnth, CC104 ΔnthY82A) +/- methyl viologen, +/- 1 mM rhamnose Strains were streaked to LB medium and incubated overnight at 37°C. For nth knockouts, strains were streaked to LB+chloramphenicol (17 μg/mL). 1 mL LB cultures were started from single colonies and grown overnight in a shaking incubator at 37°C, 220 rpm. 20 μL of each starter culture was used to inoculate a 10 mL LB culture with/without the addition of methyl viologen (sterile) or 1 mM rhamnose. Cells were placed back in the shaking incubator for 1–1.5 hours. Samples were then centrifuged and resuspended in 10 mL NCE+glucose culture which was then grown to a density of 10^9 cells/mL at 37°C (~ 5 hours), 250 rpm. Cell density was determined by serial dilution plating a 10 μL aliquot of the NCE+glucose culture onto NCE+glucose solid medium followed by incubation at 37°C for 24 hours. 5 mL of this culture was centrifuged in a clinical tabletop centrifuge at 4°C and plated on NCE+lactose solid medium and then incubated at 37°C for 36 hours.

Materials for the microtiter screen Strains were purchased from the KEIO knockout collection from CGSC Coli Center at Yale (*19*). ΔsoxR; ΔuvrA; ΔmutM; ΔmutY; ΔiscU; Δnth-; ΔdinG=kanamycin resistant and Δlacz WT: BW25113 Δ*(araD-araB)567,* Δ*lacZ4787*(::rrnB-3*), λ⁻, rph-1,* Δ*(rhaD-rhaB)568, hsdR514.* Strain stocks were prepared from the glycerol filter disk by placing the disk on LB solid medium, adding ~ 100–200 μL LB media, and streaking the cells. A

single colony was added to 1 mL LB with 1 μL kan (34 μg /mL) and cultured in a shaking incubator overnight at 37 °C. The culture was combined with 500 μL 50% glycerol stock solution and stored at -80 °C. The MutM$^-$ strain was generously donated by Professor Sheila David at UC Davis. Reporter and control plasmids used in the assay were prepared by Helen Segal from pBBR1MCS (gentamicin/tetracycline resistant) or pUC19 (ampicillin resistant). The CPRG kit is purchased from Gelatins. CPRG=chlorophenol red β-galactopyranoside which cleaves β-galactosidase to galactose and chlorophenol red. BCA total protein assay is performed at concurrently with β-galactosidase standard curves, following the protocols in the thermo scientific. Standards are made with 2000-0 μg/mL BSA to generate curves. Only a small volume of cell lysate (6 μL) is used in the experiment as lysis buffer is not compatible with the BCA assay.

Helper screen for Lac$^+$ reversion (no plasmids; 96-well plastic plate) Cells were streaked onto LB/antibiotic plates and incubated at 37 °C 12 hours. Each colony was added to 200 μL LB in a 96-well microtiter plate and grown in a shaking incubator at 500 rpm at 37 °C ~ 3 hours. The OD of each well was measured at 600 nm using a microtiter plate reader. 1 μL starter culture was used to inoculate 200 μL fresh LB. The plate was left to shake for 1 hour and centrifuged for 5 minutes at 250 g. Supernatant was extracted, and the pellet was resuspended in 1x lysis buffer (50 μL). Cell lysate was incubated 10–15 minutes at room temperature and gently swirled. Standard dilution buffer (50 μL) was added to new wells. A blank sample contained 50 μL lysis buffer to which 50 μL dilution buffer was added. 50 μL of each sample was then added to wells that had been filled with standard dilution buffer. 100 μL of 1xCPRG substrate solution was added to each well and incubated at room temp. Absorbance at 570-595 nm was measured using the plate reader.

Helper screen for Lac$^+$ reversion (***polypropylene plate: 96-well***) The edges of 2.2 mL deep well polypropylene plates were filled with 1 mL LB. Each colony from streaked plates, described above, was added to 1 mL LB and grown for 12 hours shaking at 500 rpm. The plate was placed on ice. 1 mL LB was inoculated with 5 μL starter culture and allowed to shake 6.5 hours. 50 μL of each sample was transferred to a fresh microtiter plate and centrifuged for 5 min at 250 g. The protocol then continued as described above.

Helper screen for Lac$^+$ reversion (***polypropylene plates 96-well plate no plasmids/restreak strains***) Strains were streaked on LB/antibiotic plates and incubated at 37 °C 12 hours. Strains were then restreaked onto LB/antibiotic plates for normalization and incubated at 37 °C 12 hours. Starter cultures were made by selecting individual colonies from the restreaked plate and placing them in 1 mL LB. Cultures were grown overnight (12 hours) at 500 rpm. 1 mL LB was then inoculated with 5 μL starter culture and grown for 7 hours in a shaking incubator. 50 μL-200 μL was then transferred to a new 96-well microtiter plate and spun down 12 min at 1390 rpm. The plate was flipped to gently remove supernatant. The protocol returns to the original description above.

Helper screen for Lac$^+$ reversion (transformation with plasmids) Cells were streaked onto LB/antibiotic plates and incubated 37 °C 12 hours. 1 mL LB starter cultures (1 colony each) were grown overnight at 500 rpm. 1 mL LB was inoculated with 10 μL starter culture and grown to an OD at 600 nm of 0.3–0.4 (approx. 2–3 hours at 500 rpm). The plate was placed on ice for 30 min. Cells were centrifuged at 2500 rpm for 10 min. Supernatant is decanted by flipping the plate. Cells were resuspended in 1 mL ice cold $CaCl_2$ solution and centrifuged 5 min at 2863 rpm (60 mM CaCl2, 15% glycerol, 10 mM PIPES (pH 7): PIPES was first dissolved in water, the pH was adjusted, and the $CaCl_2$ was added). Supernatant was decanted, 1 mL $CaCl_2$ solution was added,

and the cells were placed on ice for 30 min. Cells were then centrifuged 5 min at 2750 rpm. Pellets were resuspended in 300 μL $CaCl_2$ and aliquoted into 75 μL fractions for transformation in a pre-chilled microtiter plate. The plate was frozen with liquid nitrogen and thawed on ice. 0.5 μL-1 μL plasmid was added to cells and swirled. This was left on ice for 10 min. The plate was held in a 42 °C hot water bath for 2 min and put on ice for 2 min.

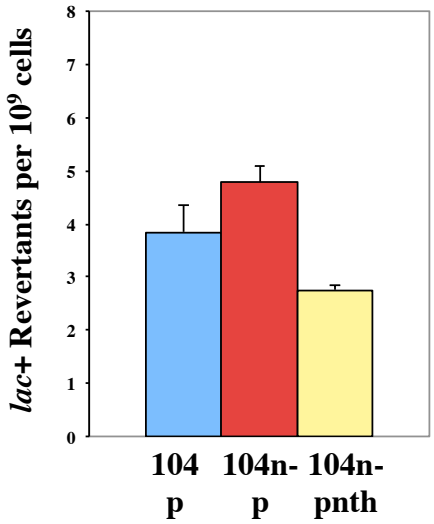
Cells that had been transformed with plasmid were recovered in 1 mL LB in a deep well polypropylene plate (incubate 1 hour 37 °C) followed by plating or culture with the appropriate antibiotic. Another method was performed as follows: 200 μL LB was added to the microtiter plate containing the cells and incubated 1 hour at 37 °C. After this, 200 μL was plated onto LB/antibiotic plate and incubated at 37 °C 12 hours to observe colony growth. A colony from the plate could then be added to 1 mL LB with the appropriate antibiotic and allowed to shake at 37 °C for 12 hours at 500 rpm. 1 mL LB with the appropriate antibiotic was then inoculated with 10 μL of the overnight culture. The cells were then grown to an OD at 600 nm of 0.3–0.4. The protocol then returned to the original CPRG assay described above. Alternatively, rather than growing colonies on plates, the 200 μL of recovered transformed cells could be transferred to a polypropylene plate with 800 μL LB with antibiotic. As an example, if the plasmid is resistant to gentamicin, 0.4 μL of 50 μg/mL gent.=20 μg/mL total concentration would be added. Cultures were grown at 37 °C for 12 hours at 500 rpm. 1 mL LB with the appropriate antibiotic was then inoculated with 10 μL of the overnight culture. The cells were then grown to an OD at 600 nm of 0.3-0.4/~ 1 hour. 50 μL of the culture was extracted, pelleted down, and resuspend. The protocol then returned to the original CPRG assay described above.

Results and Discussion

Cooperation Between MutY and EndoIII Inside the Cell We have used genetic methods to test the CT scanning model *in vivo*. We can assay for the cooperation among repair proteins facilitated by DNA-mediated signaling. If these proteins are able to help each other in their search for damage using DNA CT, variants of EndoIII with decreased or increased capability to carry out CT will alter MutY activity *in vivo*. Assays for MutY (CC104) and EndoIII activity inside *E. coli* cells have already been developed (*12*). Prior to adding oxidative stress to the experiments, we first wanted to reproduce previous results (*12*). Utilizing the CC104 MutY activity reporter strain (Figure 4.5), 3.8 ± 0.5 *lac*+ revertants are observed per 10^9 cells. When the gene encoding EndoIII (*nth*) is inactivated in CC104 (CC104 *nth*-), the *lac*+ reversion frequency observed is 4.8 ± 0.3, more a factor of ~ 1.3-fold over CC104. Thus, loss of EndoIII does have a small but significant effect on MutY activity *in vivo*. This loss in activity is consistent with a loss in helper function by EndoIII, as predicted; the lower activity of MutY without EndoIII could reflect the lack of cooperative searching via DNA CT. Significantly, when a plasmid encoding EndoIII (pnth) is transformed back into the cells, full MutY activity is restored, and the revertant count decreases to 2.8 ± 0.1. These findings are consistent with values that have been reported previously by our lab; however, the number of revertants counted in the experiment is low (<5 colonies). We aimed to increase the overall revertant count and, subsequently, decrease the error in these measurements. To generate more lesions for MutY repair (A:8-oxo-G), oxidative stress was introduced into the assay.

When the complex is irradiated with 365 nm light, $Rh(phi)_2(bpy)^{3+}$ is an effective photooxidant and consequently, promotes guanine oxidation primarily at the 5′-G of 5′-GG-3′ doublets found in the DNA sequence (*24-26*). Thus, we have employed this complex to generate base mutations, such as 8-oxo-guanine, identical to naturally occurring lesions in the cell that MutY is responsible for repairing.

Figure 4.5. Bar graph showing *lac*+ revertants for CC104/p, CC104 *nth*-/p, and CC104 *nth*-/pnth where p denotes inclusion of an empty vector. *Lac*+ revertants are reported as the average number *lac*+ colonies that arise per 10^9 cells plated on minimal lactose media containing ampicillin. Data for the CC104 strains are shown based an independent experiment, containing 10 replicates per strain.



$Rh(phi)_2(bpy)^{3+}$ is added to cells while they are growing LB. The cells are then irradiated for 1 hour while they are in minimal media, spun down, and plated onto NCE+lactose to observe revertants. Irradiating the cells in the presence of $Rh(phi)_2(bpy)^{3+}$ increases the overall number of revertants in the assay (9.3 ± 1.8 *lac*+ revertants are observed per 10^9 cells); however, when the gene encoding EndoIII (*nth*) is inactivated in CC104 (CC104 *nth*-), the *lac*+ reversion frequency remains a factor of 2–3-fold over CC104 (17 ± 2.4 *lac*+ revertants) (Figure 4.6).

We then considered whether a redox-cycling agent, methyl viologen, that generates global oxidative stress, might further enhance the number of lesions, thereby increasing the workload for MutY. In this case, when EndoIII is knocked out of CC104 strains we would expect a large decrease in MutY activity. Methyl viologen (50 μM or 100 μM) is introduced into the minimal media during the last step of the assay. Notably, the cells treated with methyl viologen do not grow as well as the control samples. The ODs of the cells at 600 nm prior to plating are 0.6 (-methyl viologen) and 0.1 (+methyl viologen). Figure 4.7 shows that revertant counts of the cells treated with methyl viologen are much lower than those of wild-type samples. Further, growth curves (Figure 4.8) reveal that in LB, the cells treated with methyl viologen grow at similar rates as wild-type strains, but $Rh(phi)_2bpy^{3+}$ significantly slows the growth rate. From this, we have learned that rather than adding methyl viologen to cells in minimal media, when they are already starved for nutrients, we should introduce the oxidative stress while the cells are growing in LB.

Methods to Increase Lac$^+$ Revertants in the Helper Function Assay Additionally, we have purchased a new CC104 strain from the CGSC Yale facility as a comparison for the CC104 strain we have been using for the *in vivo* assays. We explore whether we can increase the *lac*+ revertant counts simply by employing a fresh bacterial strain.

Figure 4.6. Bar graph showing *lac*+ revertants for CC104 and CC104 *nth*- where Rh indicates that the samples were treated with 50 μM $Rh(phi)_2bpy^{3+}$ and irradiated in minimal media for 1 hour with a solar simulator (340 nm internal cutoff). There is a threefold difference from CC104 to CC104 *nth*- and a twofold difference from CC104+Rh to CC104 *nth*-+Rh. There is about a two- to threefold increase from –Rh to +Rh overall.

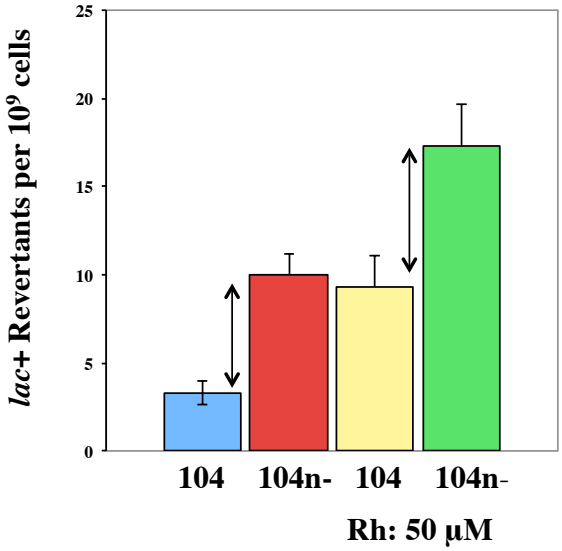


Figure 4.7. Bar graph showing *lac*+ revertants for CC104 and CC104 *nth*- where MV indicates that the cells, in minimal media, were treated with 100 μM or 50 μM methyl viologen (MV) to induce oxidative stress. *Lac*+ revertants are reported as the average number *lac*+ colonies that arise per 10^9 cells plated on minimal lactose media.

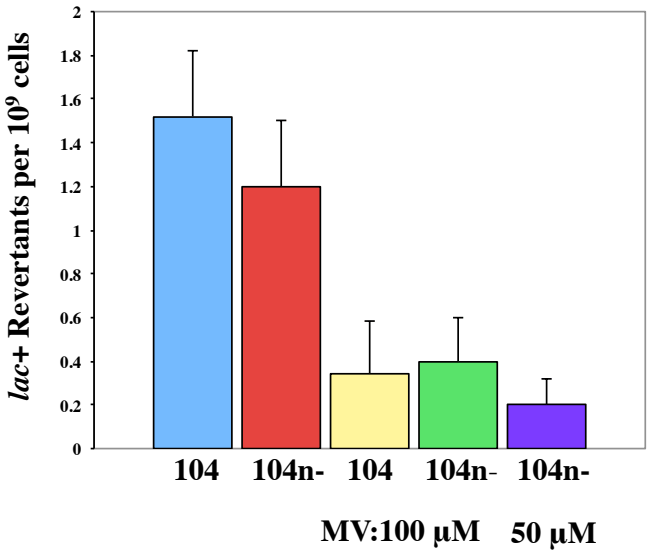
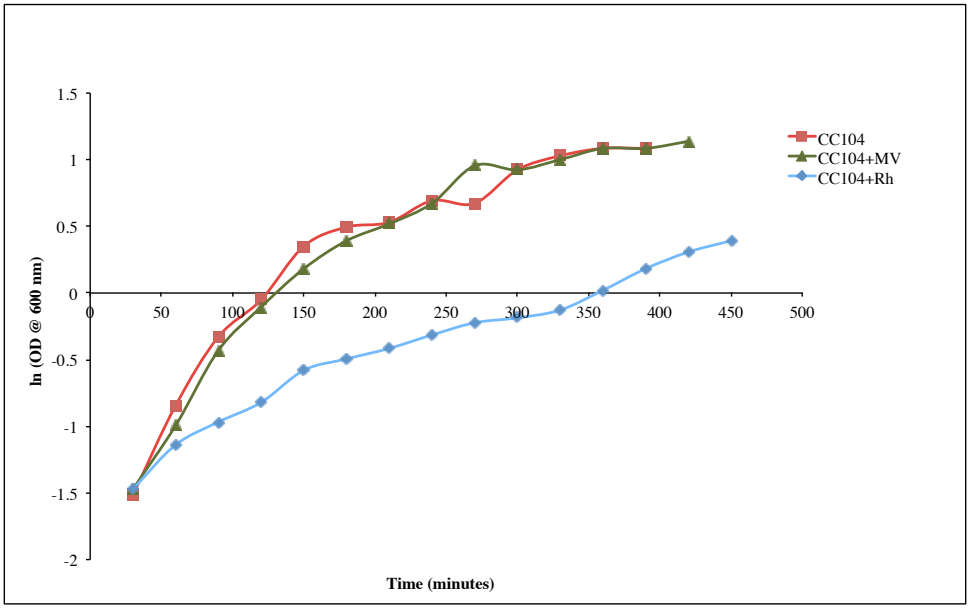


Figure 4.8. Growth curves comparing the OD at 600 nm of untreated CC104 strain cells in LB (red), + 50 μM $Rh(phi)_2bpy^{3+}$ (blue), and + 50 μM methyl viologen (green). Cells treated with $Rh(phi)_2bpy^{3+}$ grow at half the rate of the other strains.



Here, the CGSC CC104 MutY activity reporter strain yields only 0.7 ± 0.2 *lac*+ revertants per 10^9 cells. Yet, surprisingly, the original CC104 *nth*- strain produces a *lac*+ reversion frequency of 5 ± 0.6, more than a factor of > 6 over CC104 (Table 4.1). When the experiment is repeated in a separate trial, the number of *lac*+ revertants in CC104 *nth*- strain is twofold more than that of CC104, consistent with the previous results. Regardless of the OD of the cells at 600 nm prior to plating on NCE+lactose solid media, we consistently observe a twofold increase in revertants from wild-type to the EndoIII knockout strain (Table 4.1). However, the new strain does not improve the *lac*+ revertant count or the standard error in the assay.

In an effort to increase the number of lesions in the cells, we have designed and employed three strategies: i) purge cultures with argon while they are growing, and once they are starving in minimal media, expose them to oxygen (Figure 4.2) ii) grow the LB cultures in an anaerobic chamber (10% H_2, 10% CO_2, and 80% N_2 atmosphere), expose the cells to oxygen while they are in minimal media (and add the redox-cycling agent, methyl viologen) (Figure 4.2), and iii) grow the LB cultures over time in the presence of methyl viologen at concentrations of 50-500 μM, pellet the cells, and resuspend them in minimal media (Figure 4.3) (*22*). The first two methods do not increase total *lac*+ revertants (Table 4.2); however, the third method produces counts slightly higher than those we have observed in prior experiments (Table 4.2). When the cells are transformed with plasmids and treated with methyl viologen, the revertant counts are higher than those reported earlier. Adding methyl viologen to the LB cultures has generated more oxidative stress in the cells. At high concentrations of methyl viologen (100 μM) CC104 *nth*-/p *lac*+ revertants decrease from 5.7 ± 0.4 to 5.1 ± 0.5. This trend is consistent in samples treated with 500 μM methyl viologen: CC014/p reversion frequency is 3.2 ± 0.4, CC104 *nth*-/p increases to 16 ± 4, CC104 *nth*-/pnth decreases to 13 ± 2, and CC104 *nth*-/Y82W decreases further to 11 ± 1.9 (Figure 4.9).

Table 4.1. *Lac*$^+$ revertants are reported as the average number of *lac*$^+$ colonies that arise per 10^9 cells plated on minimal lactose media.

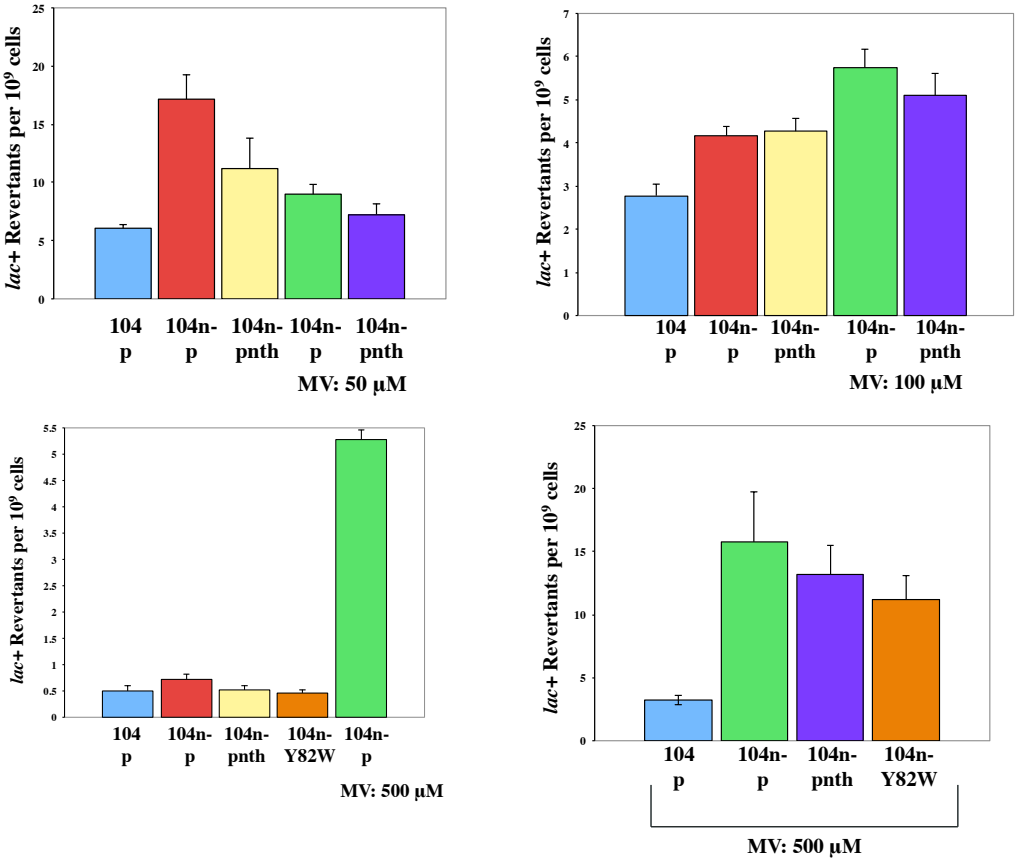
OD @ 600 nm	CC104	CC104 *nth*-	Increase
0.5-0.6	0.73 ± 0.2	5 ± 0.6	6.8
0.5-0.6	0.96 ± 0.2	1.8 ± 0.2	1.9
0.5[a]	1.0 ± 0.2	1.9 ± 0.1	2
0.7[a]	1.9 ± 0.2	3.2 ± 0.3	1.7
0.5-0.7[a]	1.4 ± 0.2	2.4 ± 0.2	1.7

a) These data represent a single set of experiments with 5 replicates at each OD measurement.

Table 4.2. *Lac*$^+$ revertants are reported as the average number of *lac*$^+$ colonies that arise per 10^9 cells from CC104 and CC104 *nth*- strains plated on minimal lactose media as a function of oxidative stress. These data represent at least five replicates in each trial with high and low values removed. Error reported is standard error based on the number of replicates reported.

Strain	*Lac*+ Revertants	Increase
No Oxidative Stress		
CC104	0.9 ± 0.1	
CC104 *nth*-	1.4 ± 0.1	1.5
Argon Bubbled		
CC104	1.7 ± 0.3	
CC104 *nth*-	2.5 ± 0.6	1.5
Argon + (100 µM MV)		
CC104	1.3 ± 0.2	
CC104 *nth*-	1.4 ± 1.0 (**2.1 ± 1.3**)	1 (**1.6**)
Anaerobic Chamber		
CC104	1.2 ± 0.2	
CC104 *nth*-	1.7 ± 0.3	1.4
Trial 1: Resuspend Pellet (50 µM MV)		
CC104	1.6 ± 0.1	
CC104 *nth*-	2.3 ± 0.3 (**3.2 ± 0.2**)	1.4 (**2**)
Trial 2: Resuspend Pellet (50 µM MV)		
CC104	1.5 ± 0.1	
CC104 *nth*-	2.7 ± 0.4 (**3.7 ± 0.3**)	1.8 (**2.5**)

Figure 4.9. Bar graph showing *lac*+ revertants for CC104/p, CC104 *nth*-/p, and CC104 *nth*-/pnth where p denotes inclusion of an empty vector. *Lac*+ revertants are reported as the average number *lac*+ colonies that arise per 10^9 cells plated on minimal lactose media containing ampicillin. Methyl viologen (MV) is added at the concentrations indicated below (50, 100, and 500 μM). In the presence of 500 μM, there is an increase in the overall number of revertants.



Y82W is an EndoIII mutant that is able to perform DNA CT, and this data supports that the mutant can cooperate with MutY to help it repair damage (*27*).

Concurrently, we have examined a CC104 strain that, rather than incorporating an antibiotic cassette in place of EndoIII, (*nth*-), it lacks the gene entirely (Δ*nth)*. In parallel, we have compared CC104 Δ*nth* strains that either wild-type EndoIII reinserted (WTnth) or that have Y82A mutant EndoIII (Y82A). Although the growth rates of CC104, CC104 *nth*-, and CC104 Δ*nth,* and CC104 Δ*nth*WTnth (not shown) are identical (Figure 4.10), the CC104 Δ*nth* strain behaves differently in the *lac*+ reversion assay. With methyl viologen (50 μM), CC104 yields 2.9 ± 0.2 revertants, and CC104 Δ*nth* increases slightly less than twofold to 4.9 ± 1. Upon the addition of 1 mM rhamnose, to induce the expression of WTnth and Y82A in the bacterial cultures, the trend is absent (Table 4.3). When we compare CC104, CC104 Δ*nth,* and CC104 *nth*- in the assay (using 50 μM methyl viologen), CC104 *nth*- consistently shows an increase in revertant counts, but the CC104 Δ*nth* strain does not.

High-Throughput (96-Well-Plate) Screen Following the protocols described in the materials and methods section, we have begun developing a high-throughput screen (Figure 4.11) to identify proteins that may cooperate with MutY in a DNA-mediated manner to find lesions in the genome. Each strain is transfected with a plasmid containing *lacZ*. Mutant *lacZ* has been PCR-amplified from the CC104 strain alongside a positive control, wild-type *lacZ*, from the wild-type MG1655 strain. These genes have been inserted into both pBBR1MCS and pUC19 plasmids (*28*), which contain tetracycline, gentamicin, or ampicillin resistance cassettes. The antibiotic resistance ensures that the strains contain the reporter plasmid after transfection.

Figure 4.10. Growth curves comparing the OD at 600 nm of untreated CC104 strain cells in LB (blue), CC104 Δ*nth* (red), and CC104 *nth*- (green). All of the strains show identical growth curves.

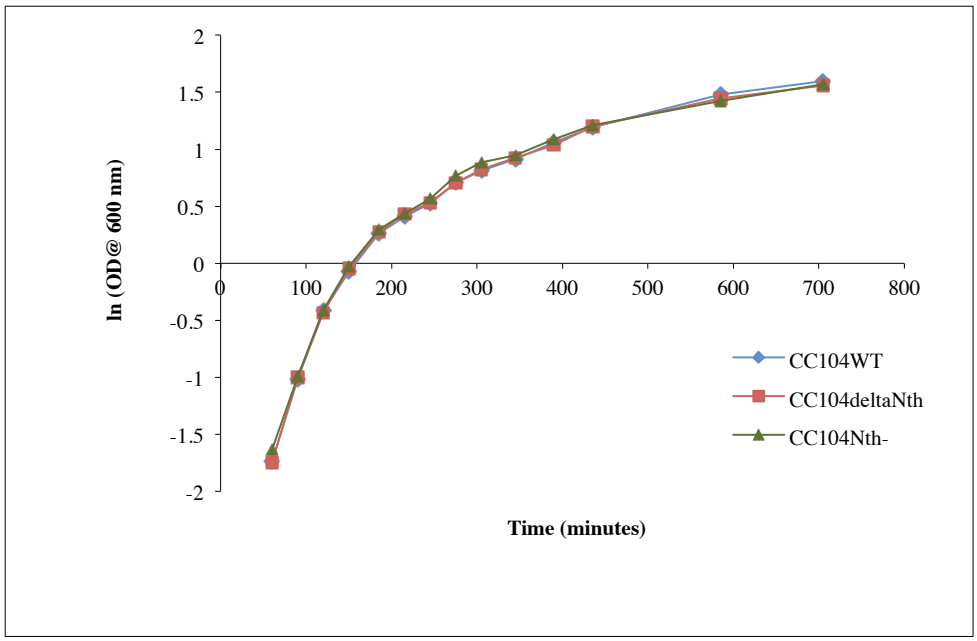
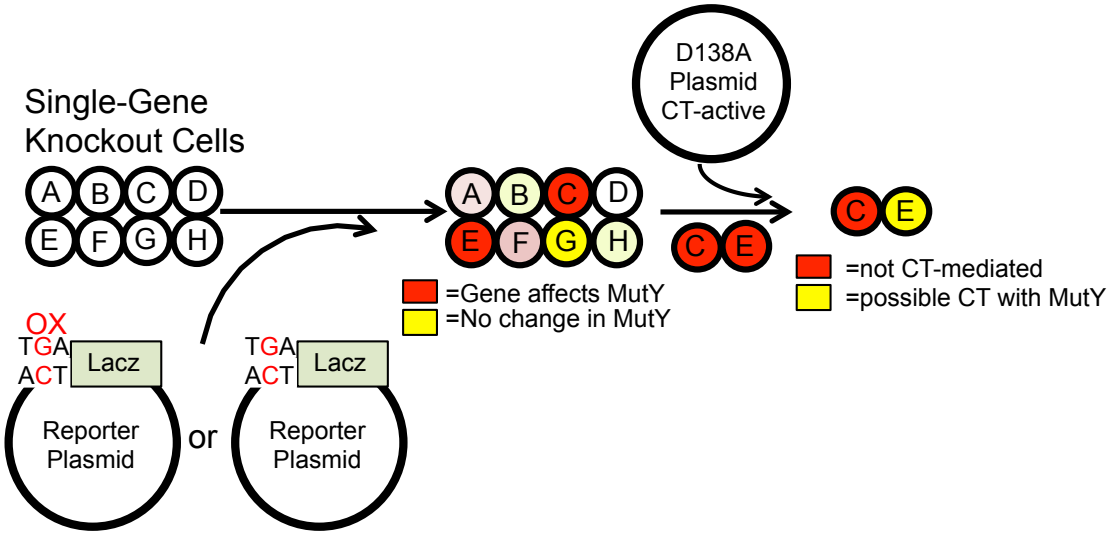


Table 4.3. *Lac*$^+$ revertants are reported as the average number of *lac*$^+$ colonies that arise per 10^9 cells from CC104, CC104 Δ*nth*-, and CC104 *nth*- strains plated on minimal lactose media as a function of oxidative stress and rhamnose induction. Error reported is standard error based on the number of replicates reported. There is no increase in revertant counts from wild-type CC104 to CC104 Δ*nth*.

Strain	*Lac*+ Revertants
CC104 (**+ 50 µM MV**)	1.3 ± 0.2 (**2.9 ± 0.2**)
CC104 *Δnth* (**+ 50 µM MV**)	2.2 ± 0.4 (**4.9 ± 0.9**)
1 mM Rhamnose, 50 µM MV	
CC104	1.8 ± 0.2
CC104 *Δnth*	1.8 ± 0.2
CC104 *Δnt*h/nth reinserted	3.2 ± 1
CC104 *Δnth*/Y82A	1.5 ± 0.3
50 µM MV (**+ 1 mM Rhamnose**)	
CC104	1.7 ± 0.3 (**1.36 ± 0.2**)
CC104 *Δnth*	1.3 ± 0.2 (**1.52 ± 0.2**)
No MV (**+ 1 mM Rhamnose**)	
CC104	1.7 ± 0.2 (**1.9 ± 0.3**)
CC104 *Δnth*	1.6 ± 0.2 (**1.7 ± 0.4**)
50 µM MV	
CC104	1.6 ± 0.1
CC104 *Δnth*	1.3 ± 0.1
50 µM MV	
CC104	1.3 ± 0.1
CC104 *Δnth*	1.4 ± 0.3
CC104 *nth*-	4.9 ± 1.4

Figure 4.11. A flowchart of the helper screen is shown. Single-gene knockout cells are transformed with plasmids that report on MutY activity. Yellow circles indicate the lack of β-galactosidase activity, and red circles represent β-galactosidase production. If the knocked out enzyme helps MutY detect sites of DNA damage, there will be an increase in the number of revertants, and thus an increase in β-galactosidase activity. An increase in in β-galactosidase activity will be visualized as an increase in absorbance at 572 nm. If the protein that has been knocked out helps MutY in a DNA-mediated process, transforming the cells with D138A, a mutant that is CT-proficient but glycolytically inactive, will restore MutY activity reverting the well to a yellow color.

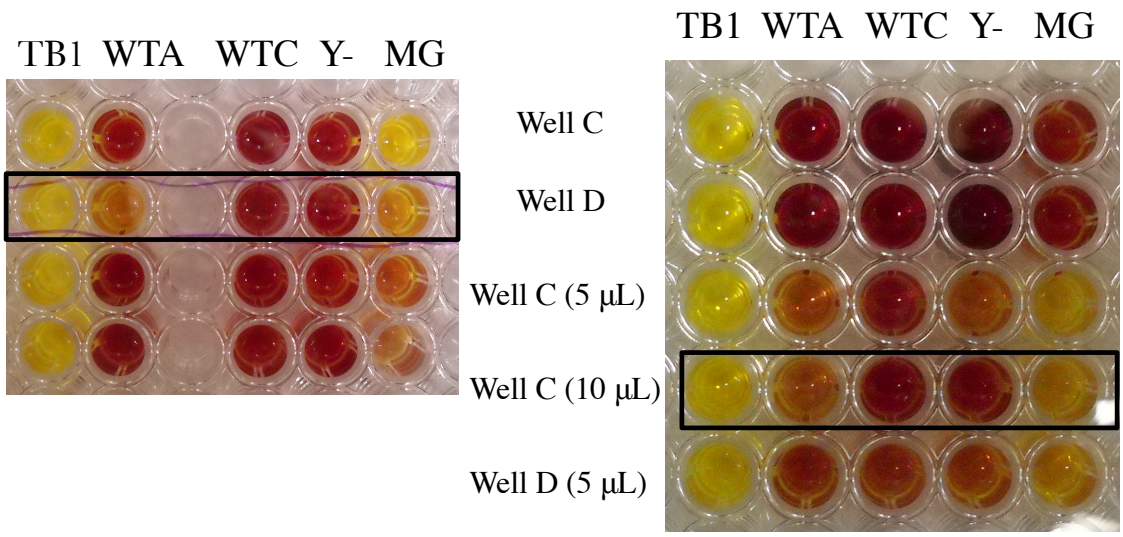


Prior to utilizing the vectors, we ran a variety of controls to reduce the technical error in the microtiter-based assay. Control strains, TB1 and MG1655, were chosen as TB1 ΔlacZ strain and MG1655 is wild-type-these were streaked onto LB plates. First, TB1, a control strain that does not produce β-galactosidase, and MG1655 are streaked onto LB solid media and incubated overnight at 37 °C. Starter cultures are grown for 4 hours in a shaking incubator at 500 rpm. The OD measurements at 600 nm for the strains were: TB1: 2.0 and MG1655: 2.0. 1 μL of the starter culture is used to inoculate 200 μL LB, incubated, and grown for 3 hours at 500 rpm. The ODs of the cells are TB: 0.5 and MG1655: 0.8. Upon the addition of CPRG substrate, the wells containing MG1655 bacterial strains turn red, as expected, and the wells with TB1 remain yellow.

Plastic microtiter plates accommodate only very small volumes (< 500 μL) and are susceptible to contamination, thus we have switched to using deep well (2.2 mL) polypropylene plates. In each experiment, the edges of the plate are filled with 1 mL LB. We have examined the controls for TB1, CC104, CGSC CC104, CC104 Y-, and MG1655 strains. OD values, just before the assay, are: TB1: 0.03, CC104 (WT Amie): 0.36, Blank: 0, CC104 (CGSC): 0.3, MG1655: 0.5 and CC104Y-: 0.4. 50 μL is removed from these samples and transferred to a fresh plastic 96 well microtiter plate. The protocol then continues as described above. Figure 4.12 reveals that half of the MutY- cultures turn red before WT (Amie); however, WT (CGSC) turns red at the same time as MutY-. One row (colony) follows the trend that we expect: the wild-type CC104 strain stays yellow longer than MutY-. In this assay, just as in the *lac*+ reversion assay of individual strains, when MutY is knocked out of cells, lesions persist.

The experiment has been repeated, and the cells grow faster: OD measurements of TB1: 0.9, CC104 (WT Amie): 0.3, CC104 (CGSC): 0.2, MG1655: 0.6 and CC104Y-: 0.6. Colonies are diluted to observe whether the concentration of cells affects our ability to observe β-galactosidase production (Figure 4.12).

Figure 4.12. Microtiter 96-well plates with TB1, wild-type CC104, ΔmutY, and MG1655 strains. Red wells indicate β-galactosidase activity. TB1 does not produce β-galactosidase and remains yellow throughout the assay. Rows with black boxes yielded absorbance patterns that would be expected for each strain. (Right) Dilutions of cell lysates do not alter the results.



CPRG is added in opposite directions to each row to ensure uniformity and reduce technical error. Here, we confirm that the protocol we have used for lysis/dilution yields reproducible results (Figure 4.12).

Transfecting Cells with Plasmid (96-Well Plate) In order to develop the screen, it is necessary to transform the cells with plasmids that report on MutY activity. To this end, we have investigated whether we can reproducibly transfect cells in a 96-well microtiter plate. Thus far, using heat-shock methods, we have successfully transfected WT, Nth-, and MutM- strains with reporter and control plasmids (*29*). OD values for the strains post-transformation are 0.2-0.3, within error of each other. We have expanded our studies to strains purchased from the KEIO knockout collection: ΔsoxR; ΔuvrA; ΔmutM; ΔmutY; ΔiscU; Δnth-; ΔdinG. Since all of the strains are kanamycin resistant, we examine whether it is required for cellular growth. We find that all of the strains, with the exception of ΔiscU, grow at similar rates (Table 4.4). It is not surprising that the ΔiscU strain does not reach the same OD as the other strains as iscU is a scaffolding enzyme that is required to build iron-sulfur clusters and deliver them to a variety of other proteins (*30, 31*). When this gene is absent, proteins that require metal cofactors may not fold or function properly. Absorption measurements at 570 nm are recorded over time using a microtiter plate and normalized to control wells (Figure 4.13). MG1655 (wild-type) turns red after a short time; whereas, all strains from the knockout collection, which are ΔlacZ, remain yellow.

Using reporter plasmids that are resistant to tetracycline/gentamicin, the assay has been repeated with WT, Nth-, MutY-, and MutM- strains. Results are reported as Miller Units versus time, where Miller Units=Absorbance 570 nm/Absorbance 630 nm (time)(volume)*1000 (*32-33*). MutM- produces more β-galactosidase than other strains. Curiously, MutY- remains yellow (Figure 4.14).

Table 4.4. OD values (at 600 nm) of strains from the KEIO knockout collection. Δiscu does not grow in the microtiter plates as well as the other strains.

MG	nth	mutY	mutM	uvR	iscu	dinG	soxR	TB1	Kan?
0.5	0.54	0.5	0.5	0.45	0.2	0.58	0.47	0.4	No
0.56	0.5	0.5	0.6	0.5	0.26	0.68	0.47	0.43	No
0.53	0.46	0.43	0.48	0.5	0.3	0.55	0.42	-	Yes
0.56	0.55	0.57	0.56	0.48	0.34	0.61	0.44	0.49	Yes

Figure 4.13. Absorbance at 570 nm vs. time (minutes) after the addition CPRG substrate. MG1655 produces β-galactosidase and turns red over time.

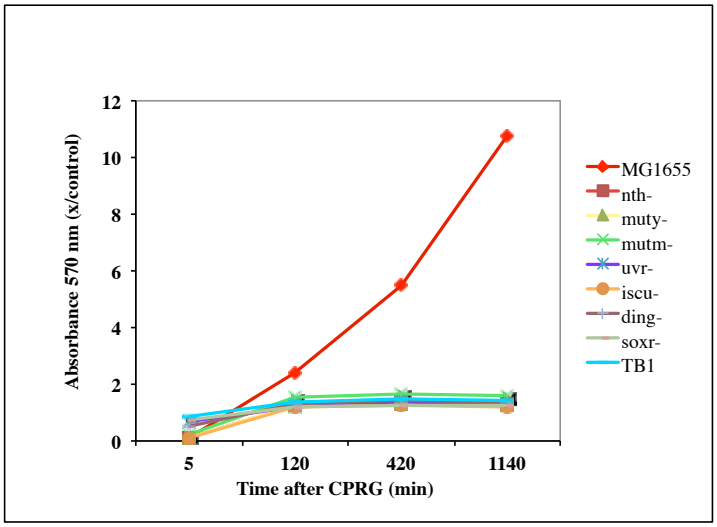
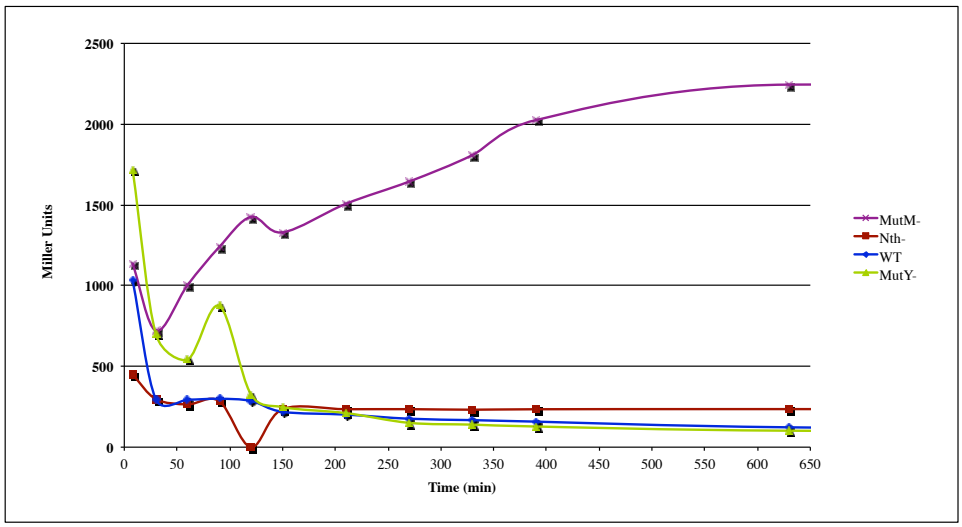


Figure 4.14. Miller Units vs. time (post the addition of CPRG substrate). ΔmutM, ΔmutY, Δnth, and WT (BW25113) strains have been transformed with reporter plasmids. ΔmutM produces more β-galactosidase than the other knockout strains.



Just as we enhanced *lac*+ revertants in the helper function, we introduced methyl viologen in the screen to increase the number of lesions that MutY must repair. After the cells are grown to an OD at 600 nm of 0.4-0.6, 50 μM methyl viologen is added to the culture.

For each trial, standards are made using 100-0.39 milliunits of β-galactosidase. In addition, a BCA protein assay is performed with 200-0 μg/mL BSA standards to normalize for the total protein concentration in the cell lysate (Figure 4.15). In the presence of methyl viologen, WT, MutM-, MutY-, and Nth- strains that have been transfected with reporter plasmids produce more β-galactosidase than those with control vectors, which is significant. On the other hand, the difference between WT and MutY- is not large enough to identify proteins that affect MutY activity (Figure 4.16). The WT strain (with reporter plasmids) turns red quickly, which convolutes comparisons between the strains. Starving the cells in minimal media (NCE) does not enhance our ability to differentiate WT and MutY- strains, which are the two controls for the experiment (Figure 4.17).

Lastly, we have randomized the position of strains within the microtiter plate to reduce experimental error. When we examine WT and MutY- strains that have been transfected plasmids, we observe an increase in absorption at 570 nm over 7 hours only in the samples that carry reporter vectors (Figure 4.18). Once again, however, the difference between WT and MutY- strains is very small (0.4 vs. 1), which provides a challenge in further developing this screen.

Summary

The helper function assay is a valuable tool we can utilize to investigate metalloproteins that may cooperate in a DNA-mediated search process, thus we have employed multiple strategies to both enhance the signal and reduce the error of our measurements. The results described here support the model that EndoIII helps MutY search DNA for damage, yet there are still ways in which the assay may be improved.

Figure 4.15. (Top) Representative β-galactosidase standard curve (Bottom) BCA standard curve to normalize for total protein concentration

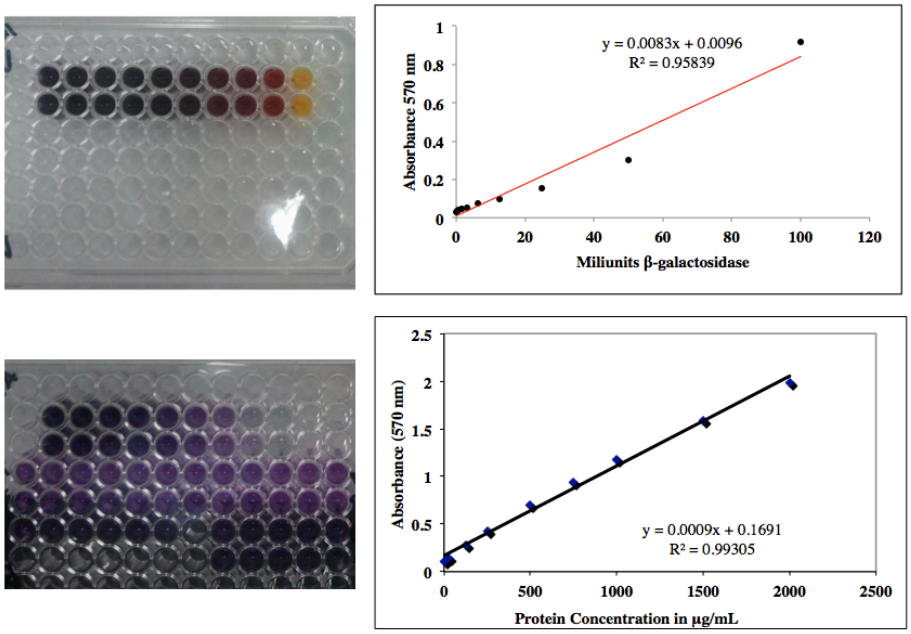


Figure 4.16. Miller Units over time (minutes) for ΔmutY, Δnth, ΔmutM, and WT (BW25113) strains transformed with reporter plasmid and treated with 50 μM methyl viologen. Wild-type strains have also been transformed with control plasmids and remain yellow over time.

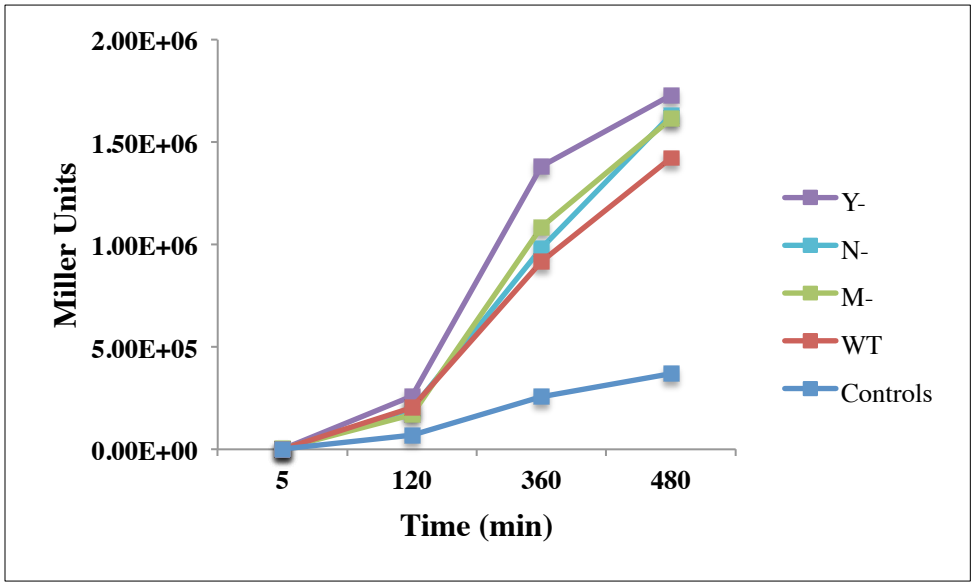


Figure 4.17. Miller Units over time (minutes) for ΔmutY, Δnth, ΔmutM, and Δiscu, and WT (BW25113) strains transformed with reporter plasmid and incubated in minimal media (NCE). Wild-type strains have also been transformed with control plasmids and remain yellow over time.

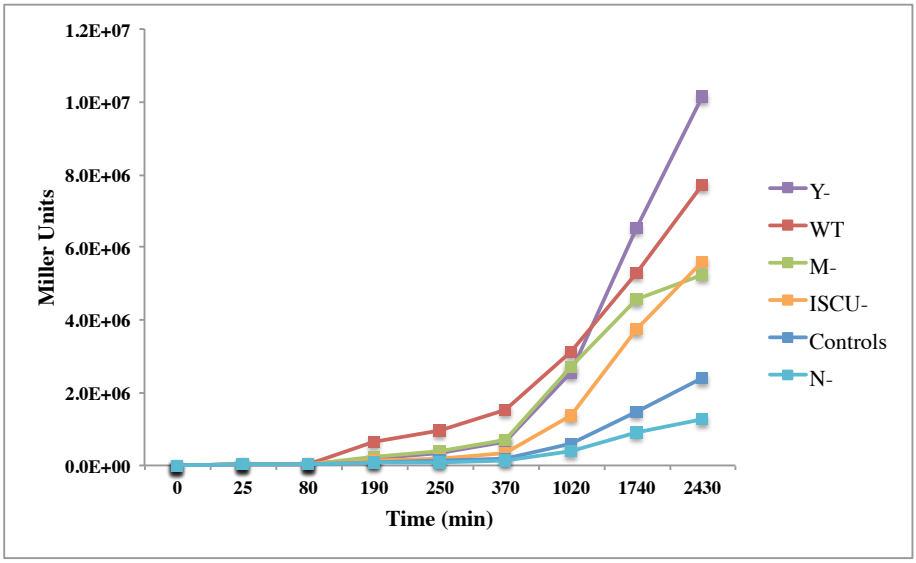
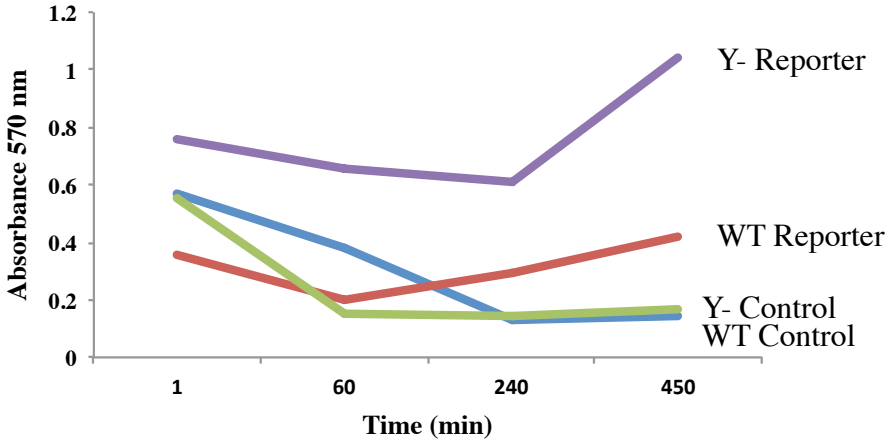


Figure 4.18. Absorbance at 570 nm vs. time (minutes). ΔmutY and WT (BW25113) strands have been transformed with reporter plasmid (purple and red, respectively) or control plasmid (green and blue). When ΔmutY is transformed with reporter plasmid, the cells produce more β-galactosidase. The controls remain yellow over time, as MutY is active and no β-galactosidase is generated.



Although our initial attempts to increase reversion events using the oxidants: $Rh(phi)_2bpy^{3+}$ and methyl viologen enhance the overall signal, they significantly inhibit cellular growth. Then, rather than using chemical reagents, we exposed the cells to environments that were free of oxygen. After the cells were grown under these conditions, they were shocked with air or methyl viologen to generate oxidative lesions. This technique was not successful, but it did inspire the development of a new method: treatment of LB cultures with methyl viologen. One of the main reasons that the initial studies with methyl viologen were futile was that the cells died during the experiment. Thus, instead of subjecting cells to the reagent while they are in minimal media, we have incorporated a modified step in the procedure. The cells are first treated with methyl viologen while they are growing in LB media. Following this step, the cultures are spun down, resuspended in minimal media, and grown according to the original protocol. From these studies, we find that 50 μM methyl viologen increases the number of *lac*+ revertants. Concurrently, a new EndoIII strain (CC104 Δ*nth*) has been compared to the original EndoIII knockout strain (CC104 *nth*-). Since EndoIII is completely deleted from the Δ*nth* strain, we expected a large decrease in MutY activity; however, we do not observe this trend in the assay.

Proteins from a variety of repair pathways and organisms that contain a redox-active [4Fe4S] cluster are being discovered. We would like to expedite the hunt for these specialized metalloproteins so that we can characterize them and determine whether they are able to participate in a DNA-mediated search within the cell. To this end, we are working toward the development of a high-throughput screen to find genes that alter MutY activity. Though we have methodically improved the microtiter-based technique, performed 96-well plasmid transformations, and successfully constructed the independent components, the sensitivity of the assay is not optimal. Applying the protocols that we have used to improve the helper function assay, methyl viologen and minimal media have been incorporated in the screen to enhance the signal, but there is no

differential between control strains. These studies have established that generating and quantitating oxidative damage are key requirements to assay for MutY activity. Only when we can do this, will we be able to utilize the screen to identify other proteins containing a redox-active [4Fe4S] cluster that might signal MutY in the search for damage.

REFERENCES

1. Boon, E. M., Livingston, A. L., Chmiel, N. H., David, S. S., and Barton, J. K. (2003) *Proc. Natl. Acad. Sci. USA 100*, 12543.

2. Boal, A. K., Yavin, E., Lukianova, O. A., O'Shea, V. L., David, S. S., and Barton, J. K. (2005) *Biochemistry 44*, 8397.

3. Genereux, J. C., Boal, A. K., and Barton, J. K. (2010) *J. Am. Chem. Soc. 132*, 891.

4. Genereux, J. C., and Barton, J. K. (2010) *Chem. Rev. 110*, 1642.

5. Kelley, S. O., Boon, E. M., Barton, J. K., Jackson, N. M., and Hill, M. G. (1999) *Nucleic Acids Res. 27*,4830.

6. Kelley, S. O., Jackson, N. M., Hill, M. G., and Barton, J. K. (1999) *Angew. Chem. Int. Ed. 38*, 941.

7. Boon, E. M., Salas, J.E., and Barton, J.K. (2002) *Nat. Biotechnol. 20*, 282.

8. Kelley, S. O., and Barton, J. K. (1999) *Science 283*, 375.

9. Kelley, S. O., Holmlin, R. E., Stemp, E. D. A., and Barton, J. K. (1997) *J. Am. Chem. Soc. 119*, 9861.

10. Gorodetsky, A. A., Boal, A. K., and Barton, J. K. (2006) *J. Am. Chem. Soc. 128*, 12082.

11. Yavin, E., Boal, A. K., Stemp, E. D. A., Boon, E. M., Livingston, A. L., O'Shea, V. L., David, S. S., and Barton, J. K (2005) *Proc. Natl. Acad. Sci. 102*, 3546.

12. Boal, A. K., Genereux, J., Sontz, P., Gralnick, J. A., Newman, D. K., and Barton, J. K. (2009) *Proc. Natl. Acad. Sci. USA 106*, 15237.

13. Kuo, C. F., McRee, D. E., Fisher, C. L., O'Handley, S. F., Cunningham, R. P., and Tainer, J. A. (1992) *Science 258*, 434.

14. Thayer, M. M., Ahern, H., Xing, D., Cunningham, R. P., and Tainer, J. A. (1995) *EMBO J. 14*, 4108.

15. Cupples, C. G., and Miller, J. H. (1989) *Proc. Natl. Acad. Sci. USA 86*, 5345.

16. Nghiem, Y., Cabrera, M., Cupples, C. G., and Miller, J. H. (1988) *Proc. Natl. Acad. Sci. USA 85*, 2709.

17. Kreutzer, D. A., and Essigmann, J. M. (1998) *Proc. Natl. Acad. Sci. USA 95*, 3578.

18. Tano, K., Iwamatsu, Y., Yasuhira, S., Utsumi, H., and Takimoto, K. (2001) *J. Radiat. Res. 42*, 409.

19. Baba, T., Ara, T., Hasegawa, M., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B.L., and Mori, H. (2006) *Mol. Syst. Biol. 2*, 1.

20. Davis, R. W., Botstein, D., and Roth, J. R. (1980) *Advanced Bacterial Genetics: A Manual for Genetic Engineering* (Cold Spring Harbor Laboratory Press).

21. Miller, J. H. (1992) *A Short Course in Bacterial Genetics: A Laboratory Manual and Handbook for Escherichia coli and Related Bacteria.* (Cold Spring Harbor Laboratory Press).

22. Ono, T., Nunoshiba, T., Kuraoka, I., Hiratsu, K., Yamada, K., and Nishioka, H. (1991) *The Science and Engineering Review of Doshisha University 32*, 35.

23. Romano, C. A. (2011) DNA-mediated charge transfer between [4Fe-4S] cluster glycosylases, Thesis in Chemistry, California Institute of Technology, Pasadena, CA.

24. Barton, J. K., Olmon, E. D., and Sontz, P. A. (2011) *Coord. Chem. Rev. 255*, 619.

25. Merino, E. J., Boal, A. K., and Barton, J. K. (2008) *Curr. Opin. Chem. Biol. 12*, 229.

26. Lee, P. E., Demple, B., and Barton, J. K. (2009) *Proc. Natl. Acad. Sci. 106*, 13164.

27. Romano, C. A., Sontz, P. A., and Barton, J. K. (2011) *Biochemistry 50*, 6133.

28. Kovach, M. E., Elzer, P. H., Hill, D. S., Robertson, G. T., Farris, M. A., Roop, R. M. II, and Peterson, K. M. (1995) *Gene 166*, 175.

29. Michaels, M. L., Christina, C., Grollman, A. P., and Miller, J. H. (1992). *Proc. Natl. Acad. Sci., 31*, 10964.

30. Kim, J. H., Füzéry, A. K., Tonelli, M., Ta, D. T., Westler, W. M., Vickery, L. E., and Markley, J. L. (2009) *Biochemistry 48*, 6062.

31. Mochel, F., Knight, M. A., Tong, W. H., Hernandez, D., Ayyad, K., Taivassalo, T., Andersen, P. M., Singleton, A., Touault, T. A., Fischbeck, K. H., and Haller, R. G. (2008) *Am. J. Hum. Genet. 82*, 652.

32. Miller, J. H. (1972) *Experiments in molecular genetics.* (Cold Spring Harbor Laboratory Press).

33. Zhang, X., and Bremer, H. (1995) *J. Biol. Chem. 270*, 11181.

CHAPTER 5

DNA Charge Transport as a First Step in Coordinating the Detection of Lesions by Repair Proteins

Adapted from Sontz, P. A., Mui, T. P., Fuss, J. O., Tainer, J. A., and Barton, J. K. (2012) *Proc. Natl. Acad. Sci. USA* **109**, 1856

P.A.S. performed all AFM experiments. T.P.M. characterized XPD and L325V electrochemically. J.O.F and J.A.T. provided XPD and L325V protein. We thank Alison Parisian for technical assistance and Eric Olmon for preparation and purification of WT and Y82A EndoIII protein.

INTRODUCTION

Proteins in cellular repair pathways are charged with finding and correcting genomic lesions arising from a variety of sources including oxidative stress, UV radiation, and environmental mutagens (*1*). Specific repair proteins are allocated to target different types of DNA damage for a concerted attack. For example, the nucleotide excision repair (NER) pathway involves proteins that repair lesions that largely distort the helical structure of DNA in a variety of ways; dipyrimidine adducts and cis-platin cross-links are two structurally distinct examples. In contrast, different base excision repair (BER) glycosylases remove specific base lesions; in *Escherichia coli*, for example, Endonuclease III (EndoIII) targets oxidized pyrimidines while MutY repairs oxo-guanine-adenine mismatches (*1-4*).

Given that DNA facilitates charge transport (CT) over long molecular distances (*5, 6*) and this charge transport chemistry is sensitive to the wide variety of lesions that perturb DNA base stacking (*7, 8*), our laboratory has proposed that repair proteins exploit this unique property of DNA to search for damage (*9-11*). Several DNA repair proteins contain redox-active [4Fe4S] clusters that are not required for folding (*12, 13*). Examples of these proteins from the BER pathway, EndoIII and MutY, are activated toward oxidation as they bind DNA (*14-16*). While EndoIII effectively removes oxidized pyrimidines, the repair protein is found in very low copy number in *E. coli* (~ 500 copies per cell) (*17, 18*). Surprisingly, MutY, which removes adenine from 8-oxo-G:A mismatches, is found in even lower copy number with only ~ 30 copies per cell (*19*). These low copy numbers, along with the low selectivity of these proteins for their substrates relative to unmodified DNA, begs the question of how they can so effectively find and repair their target lesions in the genome. Importantly, mutations in the human MutY homolog (MUTYH) directly correlate to the development of colorectal cancer (*20, 21*).

We have recently examined whether DNA CT may play some role in how these BER proteins find their site (*22, 23*). Using DNA electrochemistry and AFM experiments, we found that the

ability of EndoIII mutants to localize in the vicinity of a base mismatch correlates with their ability to carry out DNA CT. Moreover, using a genetics assay, we found that EndoIII cooperates with MutY *in vivo* in finding MutY lesions, but that a CT-deficient mutant of EndoIII cannot similarly aid MutY (*22*). Interestingly, more proteins involved in genome maintenance that contain [4Fe4S] clusters are emerging, many with no clear structural or enzymatic role for their clusters. One such protein is XPD, which is part of the transcription factor IIH (TFIIH) machinery and is involved in nucleotide excision repair (NER) (*24, 25*). Here we consider if XPD, which is not involved in BER, similarly utilizes DNA CT to relocalize in the vicinity of lesions, and whether signaling between different DNA-binding proteins can occur so as to coordinate the search for damage (*26, 27*).

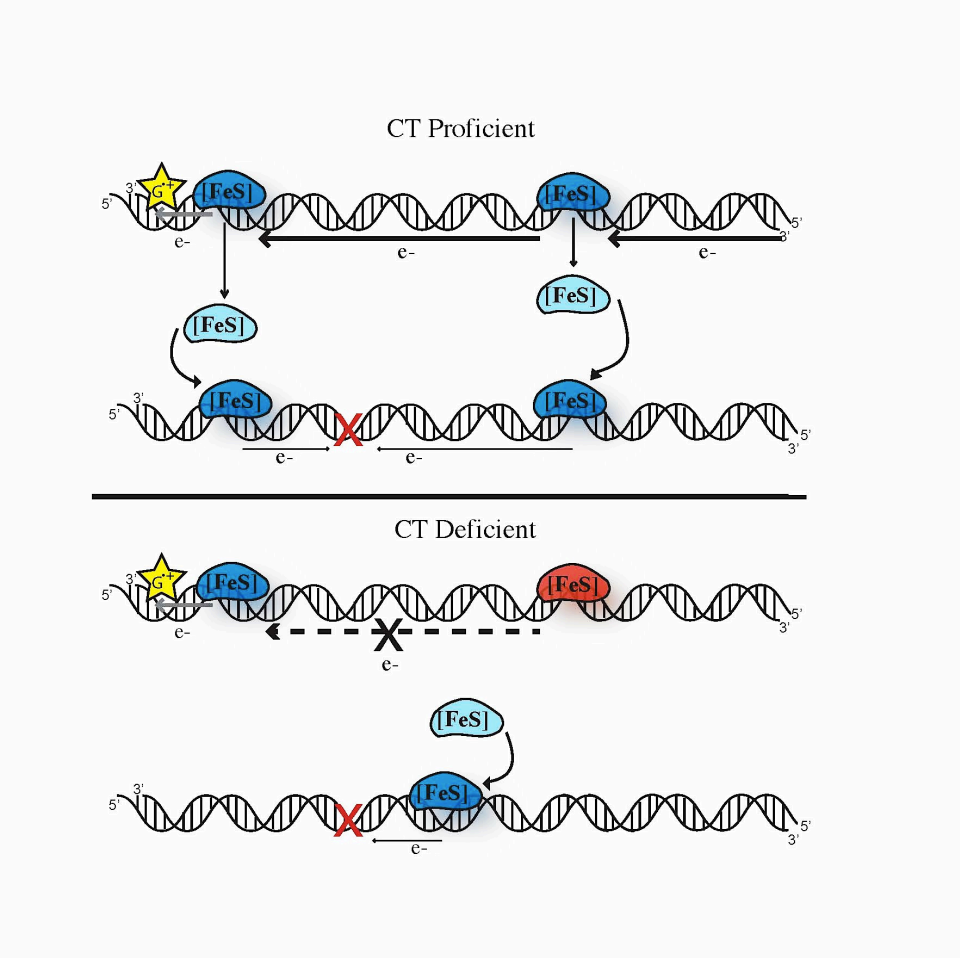
XPD, an ATP-dependent 5′-3′ helicase from the NER pathway, is a major component of TFIIH, the transcriptional and repair machinery that unwinds damaged DNA for lesion repair in NER (*28, 29*). Mutations in the human *XPD* helicase gene (ERCC2) lead to the genetic disorders trichothiodystrophy (TTD), Cockayne syndrome (CS), and xeroderma pigmentosum (XP) (*28-31*). There is now evidence that mutations specifically in the iron-sulfur cluster domain of XPD result in TFIIH instability, thus leading to TTD (*28, 30*). Similarly, mutations in the analogous region on the related *FancJ* gene (which also encodes a helicase with an iron-sulfur cluster) causes predisposition to early-onset breast cancer (*25, 32, 33*). We have demonstrated that the [4Fe4S] cluster in *Sulfolobus acidocaldarius* (Sa)XPD is redox-active, with a redox potential equivalent to that of BER proteins EndoIII and MutY (~ 100 mV vs. NHE) (*34*).

Here we explore, using AFM, whether XPD can participate in DNA-mediated signaling to localize in the vicinity of a DNA mismatch. Similar to base lesions, a single base mismatch inhibits DNA CT but importantly, the base mismatch is not a substrate for repair by XPD (*35*). In our model for facilitated search by DNA CT (Figure 5.1), the search for DNA damage is initiated, under

oxidative stress, by the one-electron oxidation of a DNA-bound protein by a nearby guanine radical; the oxidation takes the cluster from a 2+ oxidation state to a 3+ oxidation state. But it is DNA CT between two repair proteins that leads to the redistribution in the vicinity of a lesion. With a DNA-bound potential shifted negative versus free protein (*14*), the proteins are expected to have a lower affinity for DNA in their reduced form. Thus, in the model, DNA-mediated reduction of one protein by the other leads to the dissociation of the reduced protein from DNA, effectively giving overall dissociation of repair proteins away from a region of the genome without lesions. Our model relies on the sensitivity of CT to proper π-stacking of the bases between the donor protein and the distant acceptor protein (*22, 23*); if instead there is an intervening lesion, DNA-mediated CT does not occur, and the repair protein remains bound and can processively move to the lesion and carry out enzymatic repair. In this process, the repair proteins eventually relocalize in the vicinity of lesions or any modification that inhibits DNA CT. As a result the overall search regime for the repair proteins is made smaller than the full genome: they preferentially localize where needed for repair.

Despite various studies of EndoIII and MutY independently, it still remains to be established whether these proteins from different repair pathways (NER and BER) cooperate with one another in a coordinated search. Since a C:A mismatch, for example, does not distort the DNA helix, it would be expected to evade the biological NER pathway. Alternatively, if enzymes rely on DNA/protein CT as a first step in finding damage, then any protein that can carry out DNA CT can participate in the search, integrating repair pathways that were previously thought to be separate. To examine this coordinated search, AFM is used to visualize mixtures containing two types of proteins: i) those that are able to carry out DNA CT and ii) those that are defective at DNA CT. AFM, a technique that allows for visualization of single-molecule protein/DNA interactions, can be used to observe proteins bound to long strands (3.8 kbp) of DNA containing a site-specific mismatch (C:A) (*23, 36-39*).

Figure 5.1. Model for damage detection and redistribution by repair proteins. (Top) Given two repair proteins, containing [4Fe4S] clusters, that are able to carry out DNA CT, such as XPD or EndoIII (blue), CT can proceed through well-matched unmodified DNA. Driven by protein oxidation by guanine radicals (yellow) formed under oxidative stress, DNA CT occurs between the DNA-bound proteins, promoting the dissociation of reduced protein (light blue). DNA CT does not occur, however, in the presence of an intervening mismatch or lesion (red X); with the intervening mismatch or lesion, repair proteins are not reduced from a distance, and therefore do not dissociate but instead remain bound in the vicinity of the lesion. (Bottom) If repair proteins are mutants such as XPD L325V or EndoIII Y82A (red) and are unable to perform DNA CT (dashed line), they cannot send or receive signals to locate damage. In the presence of a lesion, these proteins then do not preferentially redistribute in the vicinity of the mismatch.



Binding density ratios calculated from counting strands and proteins in AFM images can then give us clues to the first step in lesion detection. Here, not only do we examine whether proteins from the NER pathway redistribute in the vicinity of lesions as a function of their ability to perform DNA CT but also whether they can cooperate with other repair proteins in the DNA-mediated search for damage.

Materials and Methods All chemicals were purchased from Sigma Aldrich. All enzymes were purchased from New England Biolabs unless otherwise specified. Mica surfaces were purchased from SPI supplies. Silicon AFM probes were purchased from Budget Sensors. Oligonucleotides were purchased from IDT or synthesized on a 3400 DNA synthesizer (Applied Biosystems).

Mismatched (C:A) strand synthesis Four primers with the following sequences were synthesized using standard phosphoramidite chemistry:

1) 5′-GTACAGAGTTCAGTCGGCATCCGCTTACAGACAAGC-3′ (forward),

2) 5′-CCGGTAACTATCGTCTTGAGTCC-3′ (reverse),

3) 5′-GACTGAACTCTGTACCTGGCACGACAGGTTTCCCG-3′ (forward),

4) 5′-GACTGAACTCTATACCTGGCACGACAGGTTTCCCG-3′ (forward)

The underlined bases highlight the location of a 2′-*O*-methyl residue (Glen Research). Primers were phosphorylated using 5 U PNK, 10% PNK buffer, 0.5 mM ATP for 5 hrs at 37 °C. Primers were purified using phenol-chloroform extraction followed by ethanol precipitation (*40*). After being dried in vacuo, primers were redissolved in 20 μL water and used in separate PCR reactions (*41*) using pUC19 as a template to generate two duplexes 1610 bp and 2157 bp (matched), each containing one 14-nucleotide single-strand overhang. Each 100 μL PCR reaction contained 50 pmol of each of two primers, 1X *Taq* buffer (100 mM Tris-HCl, 15 mM $MgCl_2$, 500 mM KCl, pH 8.3), 0.2 mM each dNTP, 1 ng plasmid template pUC19 and 3 U *Taq* polymerase (Roche). A

typical step program for PCR was as follows: After incubation at 94°C for 10 minutes, 34 cycles were performed as follows: 94°C for 1 minute, (52°C for primers 1+2, 54°C for primers 2+3 or 2+4), for 1 min, and 72°C for 3 minutes. The PCR product was then suspended in 50 mM NaCl/ 5 mM phosphate buffer and quantitated. Separate duplexes were annealed at 65 °C for 8 min in 10 mM Tris buffer, then cooled to 20 °C during 2 h. A total of 15 units of T4 DNA ligase and 10% T4 ligase buffer were added (total reaction volume ~ 20 μL) and incubated overnight at 16 °C, followed by deactivation for 10 min at 65 °C to yield the 3767 bp (mismatched) long strand. We did not bring the ligation reaction to completion, so as to obtain a mixture of DNA samples that were equivalent other than the presence of the mismatch at the ligation site. The DNA duplexes (ligated and unligated) were then analyzed by 0.6% agarose gel electrophoresis. Single DNA strands complementary to short duplex overhangs were ordered from IDT: 5′-GACTGAACTCTGTAC-3′ Tm=41.6 ° C (1.6 kbp duplex overhang), 5′-GTACAGAGTTCAGTC-3′ Tm=41.6 ° C (2.2 kbp matched duplex overhang), and 5′-GTATAGAGTTCAGTC-3′ Tm=37.5 ° C (2.2 kbp mismatched duplex overhang). Single-strand DNA was purified using reversed-phased HPLC and verified with MALDI-TOF mass spectrometry.

Protein purification and expression EndoIII and Y82A EndoIII were expressed from the pNTH10 expression vector and purified as described previously (*22*). EndoIII and Y82A were stored in 20 mM sodium phosphate pH 7.5, 100 mM NaCl, 20% glycerol, and 0.5 mM EDTA buffer. Protein concentrations were determined using the UV-visible absorbance of the [4Fe4S] cluster (410 nm, ε = 17,000 $M^{-1}cm^{-1}$) (*42*). XPD was purified as previously described (*34*).

AFM experiments AFM experiments were performed using the protocol similar to that reported previously (*23*). Stock DNA solution contained 50–200 ng of total DNA (~ 6 μM) composed of

the mixture of ligated 3.8 kbp duplexes and the two unligated duplexes (1.6 and 2.2 kb) in 6 mM $MgCl_2$/Tris-EDTA buffer. The 1.6 kb overhang complement (60 μM) was added to the DNA solution to block the 14 bp single-strand overhangs generated by PCR. This sample was then incubated overnight at 4 °C. XPD protein was dialyzed against the protein buffer (20 mM phosphate, 100 mM NaCl, 1 mM EDTA, 5% glycerol, pH 7.5 and filtered prior to use) to remove residual DTT. The concentration of individual proteins were determined by UV/visible spectrophotometry (Beckman DU 7400) using ε=17,000 $M^{-1}cm^{-1}$ at 410 nM for the [4Fe4S] cluster. After addition of excess 2.2 kb duplex complement (60 μM), XPD (0.6 μM) was added to the stock DNA solution to block single strand overhangs. This protein/DNA solution was incubated at 4 °C overnight. Sample was then deposited (5–10 μL) onto a freshly cleaved mica surface for 1–2 min, rinsed with 2 mL of water, and dried under argon. Mutant XPD protein (L325V) was added to a stock solution of 50 ng DNA for a final protein concentration of 0.6 μM. Deposition conditions were identical to that for WT XPD- DNA samples after incubation at 4 °C overnight.

For mixed protein experiments (WT^{XPD}/$WT^{EndoIII}$, WT^{XPD}/$Y82A^{EndoIII}$, $WT^{EndoIII}$/$L325V^{XPD}$, and WT^{XPD}/$L325V^{XPD}$), XPD, EndoIII, L325V, or Y82A were added to the prepared DNA solution described previously at equimolar (1:1) concentrations (0.3 μM each) prior to incubation at 4 °C overnight. WT^{XPD}/$Y82A^{EndoIII}$ (3:1) mixtures were prepared by adding 0.45 μM WT XPD and 0.15 μM Y82A to the DNA solutions. Protein/DNA complexes were formed with DNA solution containing ~ 200 ng of the mixture of PCR products and overhang complements (~ 6 μM DNA) in 6 mM $MgCl_2$ /Tris-EDTA buffer at 4°C overnight. The reaction mixture was then deposited (5–10 μL) on the mica surface for 1–2 minutes, rinsed with 2 mL water and dried under argon.

AFM instrumentation Silicon AFM Probes purchased from Budget Sensors, with a spring constant of 3 N/m and a resonance frequency of 75 kHz, were used in a Digital Instruments Multimode SPM. Images were captured in air with scan areas of 2 x 2 μm^2 or 3 x 3 μm^2 in tapping mode, at an amplitude of 0.54–2.00 V and at a scan rate of 3.05 Hz. Scan rates of 3.05 Hz were used in order to obtain images of higher quality.

Binding density ratio calculations WSxM software was used to measure general DNA contour lengths and height profiles of the proteins as described previously (*23, 43*). For each data set, images from at least three independent samples were analyzed, compared, and pooled (>200 long or short strands). Distinguishable strands and protein positions were counted by hand. The binding density ratio, r, is defined as the ratio of the proteins bound on long strands divided by proteins bound on short strands. The ratio is normalized for length by dividing by 1.9 kbp, which is the average length of the short strands. Binding affinities were found by determining the number of proteins bound per kilobasepair strand. The uncertainty was determined through the total number of proteins observed.

Protein electrochemistry Protein electrochemistry was performed as previously described (*34*). Briefly, individual proteins samples were dialyzed to remove residual DTT and quantified based on 410 nm absorbance. Protein was then added to a DNA modified electrode containing a 9 nucleotide 5′ single-strand overhang. Cyclic voltammograms were then obtained using Ag/AgCl reference electrodes, Pt auxiliary electrode at 50 mV/s scan rate on a CH Instruments 620C electrochemical analyzer.

RESULTS

Experimental strategy using AFM to probe protein distribution We have previously demonstrated with BER proteins that the ability of a protein to perform DNA CT directly correlates to its redistribution in the vicinity of base lesions or mismatches that inhibit CT (*23*). We tested XPD redistribution promoted through DNA CT by preparing DNA strands containing a single C:A mismatch, a modification that we know to inhibit CT (*44*), alongside DNA strands containing no mismatches. To distinguish the strands in microscopy experiments, we made mismatched strands > 1200 base pairs longer than the matched strands. AFM was utilized to gather images of DNA and bound protein that could be further analyzed to examine the propensity of XPD to redistribute. ATP was not incorporated in AFM experiments, as XPD helicase activity should not be required for a DNA-mediated redistribution.

Long (3.8 kilobase) DNA duplexes containing a site-specific C:A mismatch and short (2.2 and 1.6 kilobases) well-matched duplexes of the same total sequence were constructed. To prepare these DNA samples, pUC19 plasmid was amplified with primers incorporating a 2′-*O*-methyl residue to yield two short duplexes containing 14-nucleotide single-strand overhangs (*23, 41*). Upon ligation of these two duplexes, 3.8 kbp strands were formed that contained only one C:A mismatch in the middle of the strand. The ligation reaction was not taken to completion so as to have a mixture of well-matched short and mismatched long strands for protein distribution studies. Although a C:A mismatch effectively inhibits DNA CT (*44*), it is not a lesion that is preferentially bound by XPD. However, the 14-nucleotide overhangs generated with PCR are specific substrates for XPD helicase (K_d~1 μM) (*25, 45*). We directly observed this preference in initial AFM experiments. Protein assignments were verified through analysis of their 3–4 nm heights in the images; without protein, features of this dimension were not observed and larger heights (> 7 nm) indicated salt precipitates or protein aggregation. Only clearly identifiable long or short strands and bound proteins were

counted. XPD protein position was determined based on the distance of the protein from the end of the strand.

We examined distribution with and without blocking the ends of short strands, given that XPD has some preferential binding affinity for single-strand overhangs. Comparing raw protein position (middle versus ends) on long and short strands, XPD exhibited a large preference for the ends (~ 300 bp range) of short duplexes containing overhangs (> 50% of bound proteins) (Figure 5.2). To block XPD from binding to the 14 bp segment, overhang complements were added to DNA/protein solutions in excess. We found, when we blocked the overhangs, no preferential binding to the short strands was observed (*vide infra*).

Detection of XPD Complexes AFM images of DNA-bound protein complexes are represented in Figure 5.3. XPD protein bound to matched DNA (Figure 5.3 top) can be easily distinguished based on strand height profile of tapping mode images. Zoomed-in images display clearly identifiable long and short strands with protein bound (Figure 5.3 bottom). Images of DNA and protein complexes were acquired with a scan size of 2 x 2 μm^2 or 3 x 3 μm^2 at a scan rate of 3.05 Hz. Since AFM images vary with surface cleavage, sample wetness, deposition time and volume, > 200 long or short strands were counted for at least three independent samples. The uncertainty was based on the total number of proteins counted. Importantly, we measured the relative binding affinity of mutants and WT proteins, and in all of the samples described, the number of proteins bound per base pair remained constant. Thus, any changes we see in distribution are not due to differences in binding affinities of proteins.

In order to probe the DNA CT properties of a redox-active DNA helicase, XPD and DNA complexes were examined with AFM. XPD shows redistribution onto long mismatched strands with a ratio of protein binding densities, r (long/short) of 1.54 ± 0.08. We observed 0.23 proteins per kilobase of long strand and 0.15 proteins per kilobase of short strand.

Figure 5.2. Distribution of XPD on DNA containing short strands with overhangs. Protein position (middle versus ends) on long and short strands is shown in blue and red respectively. The middle region of the short strands is ~ 1300 bp. The middle of the long strands is ~ 3200 bp. Ends of the strands are ~ 300 bp.

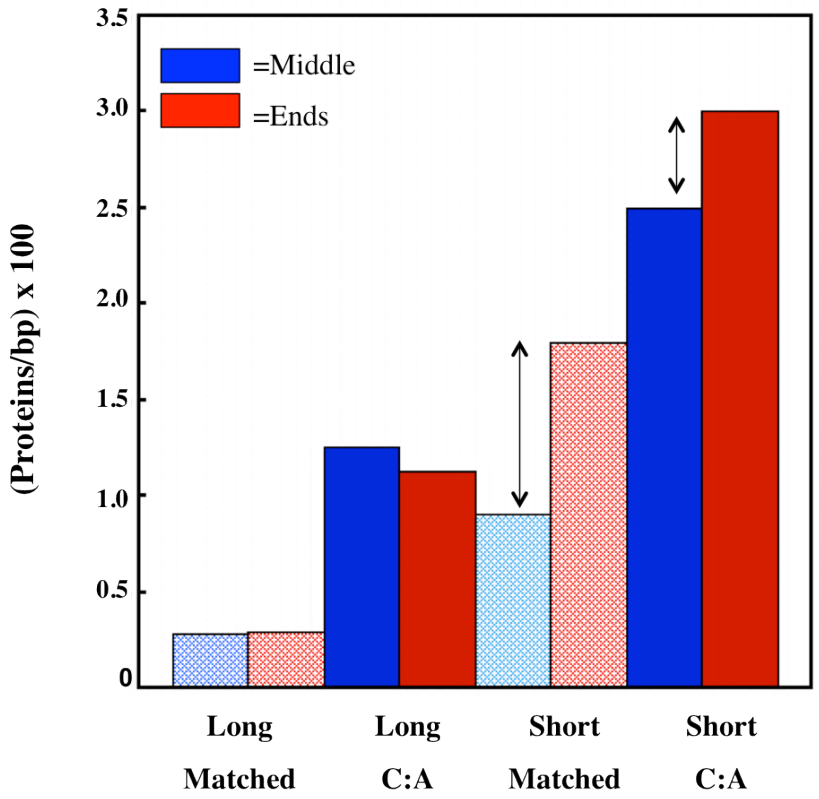
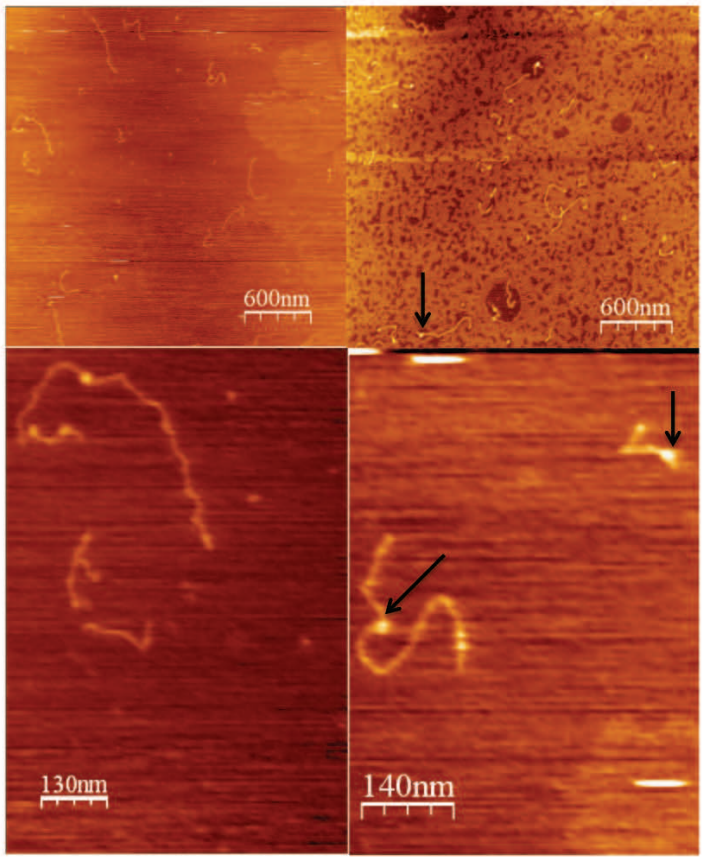


Figure 5.3. Atomic force microscopy to visualize DNA-bound proteins. Tapping mode AFM images of DNA and XPD protein on mica imaged in air. The matched strands and protein (dots) are visible on the surface (top). Since the single-strand overhangs on the short duplexes have been blocked by annealing short oligomers, XPD is bound in random positions on the DNA. Zoomed-in images of long (3.8 kbp) and short (1.6 and 2.2 kbp) DNA strands with bound proteins (black arrows) are shown below. Images were acquired with a scan size of 2 x 2 μm^2 or 3 x 3 μm^2, at a rate of 3.05 Hz with a data scale of 10 nm/ 0.5 V.



If, instead, we examine XPD distribution when long and short strands are fully matched, we see a binding density ratio of 0.94 ± 0.05. Previously, for EndoIII we found r, for (long/short) was 1.6 ± 0.09 (*22*). We expect a ratio of 1 if there were an equal distribution of proteins on matched and mismatched strands (*22*). Thus, similar to EndoIII distribution, XPD, an NER protein with DNA-bound redox-activity, redistributes to localize in the vicinity of a C:A mismatch.

We also tested the possible redistribution of a CT-deficient mutant of SaXPD, (L461 in human XPD). L325V, aligns with mutated residues in human XPD and *Schizosaccharomyces pombe* Rad 15 that are associated with TTD, XP, and XP/CS (*25, 31, 33*). It should be noted, however, that the structural scaffold required to form the TFIIH machinery is not disrupted in the L325V mutant. As evident in Figure 5.4, the L325V XPD mutant exhibits an electrochemical signal that is less than half that of WT XPD, indicating it is deficient at performing DNA CT. Interestingly, L325V is 30 Å away from the [4Fe4S] cluster (Figure 5.5), yet affects protein/DNA CT. Significantly, when we examine whether this electrochemically deficient mutant redistributes onto the mismatched strand in our AFM assay, we find that there is no preference for the mismatched strands; L325V exhibits a protein binding density ratio, r (long/short) of 1.14 ± 0.06 in mismatched samples, which is within error of the fully matched binding density ratio r (long/short) of 1.05 ± 0.07 (Figure 5.6). Thus, as with the CT-deficient EndoIII mutant Y82A (*22*), we find a correlation between the inability of L325V to perform CT and its lack of redistribution onto mismatched strands. Even though the [4Fe4S] cluster is intact, this mutant protein cannot cooperate with other proteins using DNA CT.

Atomic force microscopy of protein mixtures We have established that DNA-mediated CT occurs on a much faster time scale than protein diffusion (*46, 47*), and we have characterized several mutants of EndoIII, from the BER repair pathway that exhibit similar midpoint potentials of ~ 80 (±30) mV vs. NHE yet differ in their ability to carry out DNA CT (*14, 23, 26*).

Figure 5.4. DNA CT with the L325V mutant. Cyclic voltammogram (CV) of WT SaXPD (8 μM) (green) and L325V SaXPD (8 μM) (red) on DNA modified electrodes after 90 minutes

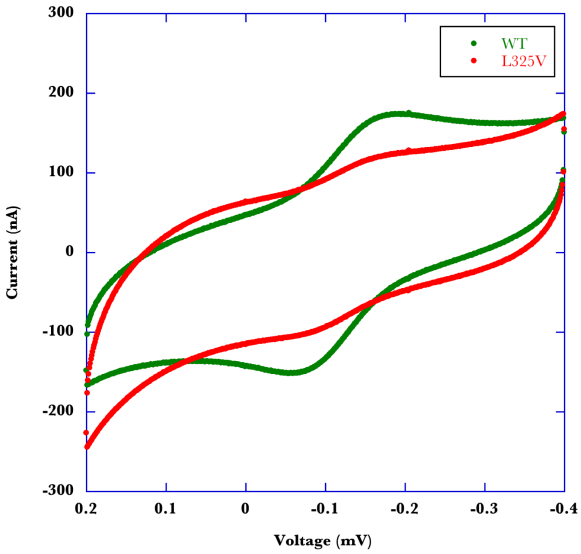


Figure 5.5. Crystal structure of SaXPD (pdb ID: 3crv). The structure shown was adapted from reference (*25*). L325 is highlighted in red, and aromatic residues are shown in orange.

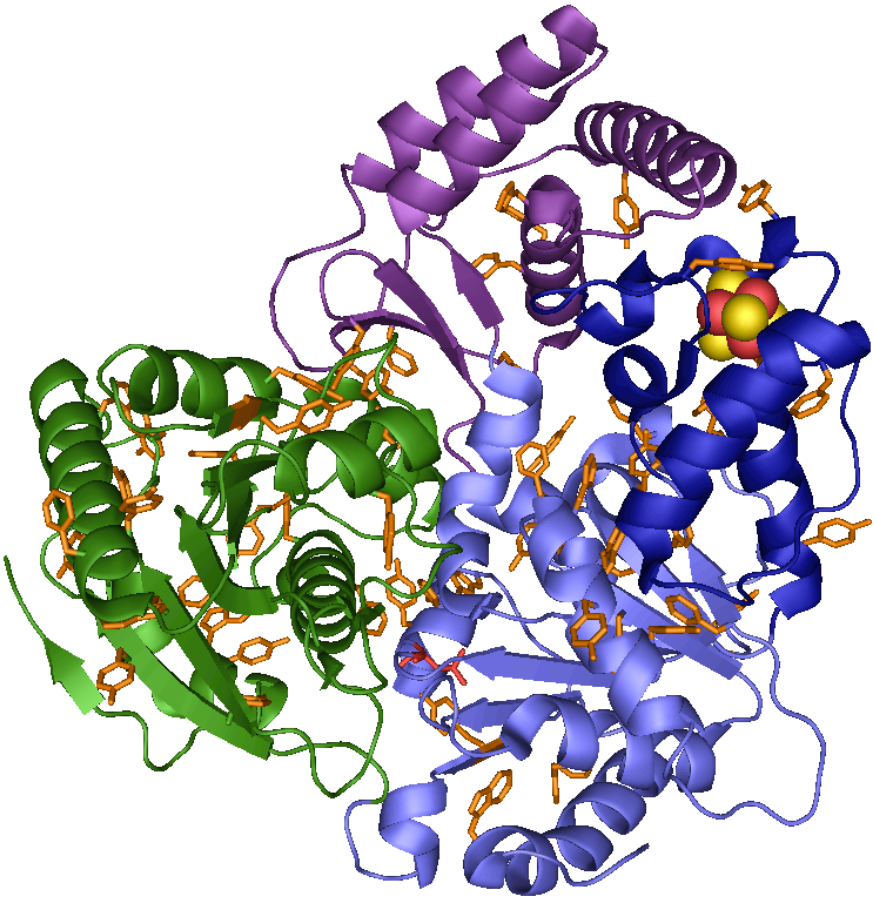
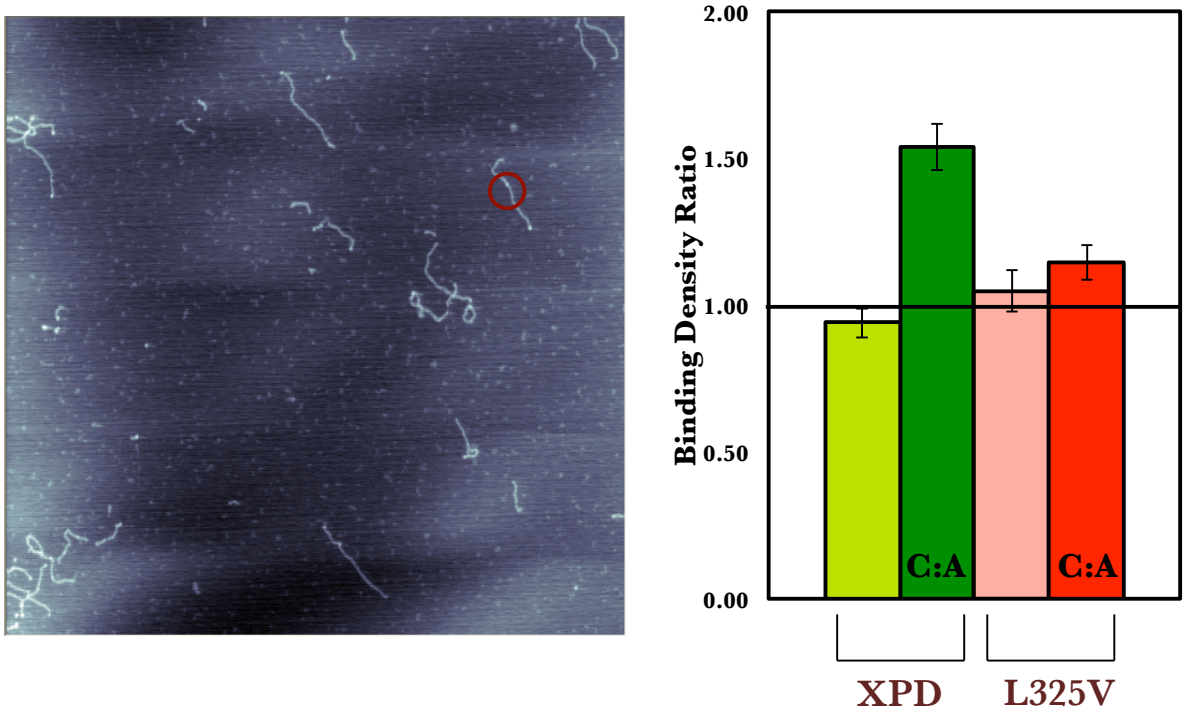


Figure 5.6. (Left) Representative AFM image of L325V XPD bound to mismatched DNA strands (red circle). The image was acquired at a rate of 3.05 Hz with a data scale of 10 nm. (Right) Quantitation of XPD and mutant L325V protein density ratios (< 10% uncertainty) where C:A indicates a mismatch is contained in long strands. The unmarked bars show the control measurements for fully matched long and short strands. XPD redistributes onto mismatched strands. L325V, CT-deficient, does not show redistribution.



We have not, however, yet provided evidence that different proteins can signal one another through DNA-mediated CT. AFM provides an opportunity to examine this issue by assaying protein mixtures.

To test this, WT XPD and EndoIII, both proficient at DNA CT, were mixed in a 1:1 equimolar ratio with DNA. Short duplex overhangs were blocked with excess complementary single strand. Figure 5.7 shows representative images of fully matched DNA strands or mismatched DNA strands incubated overnight with XPD/EndoIII 1:1 protein ($WT^{XPD}/WT^{EndoIII}$). XPD is twice as large as EndoIII (65 and 32 kDa, respectively) (pdb IDs: 3crv, 1p59). The average height of the protein in AFM studies can be estimated using the equation: $R=0.717(M)^{1/3}$ where R is the radius of the protein globule in Å and M is the mass of the protein in Da (*48*). Thus, XPD is expected to have a height of 5.6 nm, which is within error of EndoIII (4.6 nm). While in general we cannot distinguish them, select images (Figure 5.7 inset) are suggestive of both EndoIII and XPD binding. XPD (0.3 μM)/EndoIII (0.3 μM) 1:1 protein mixture shows that XPD and EndoIII redistribute onto long mismatched strands, with a binding density ratio r (long/short) of 1.75 ± 0.13. Fully matched controls yield a binding ratio r (long/short) of 1.02 ± 0.07. Importantly, controls were also performed with only XPD (0.3 μM) under the same conditions as mixed protein experiments. Consistent with XPD at a higher concentration (0.6 μM), XPD at half the concentration (0.3 μM) redistributes onto long mismatched strands (r = 1.38 ± 0.07) with no redistribution in fully matched samples (r = 1.08 ± 0.07) (Figure 5.8). Significantly, mixing XPD and EndoIII protein results in a binding density ratio of 1.75, which is higher than that of each protein separately (Figure 5.9). This suggests that these two proteins signal one another to localize in the vicinity of the lesion.

In order to explore cooperativity between XPD and EndoIII, we then replaced WT EndoIII with mutant Y82A EndoIII, a protein that binds to DNA but cannot perform DNA-mediated CT efficiently (*22*).

Figure 5.7. Representative tapping mode AFM images. Fully matched long and short DNA strands (left) or mismatched long DNA and matched short DNA strands (right) are incubated overnight with XPD/EndoIII 1:1 protein mixture. Inset shows zoomed-in view of a long DNA strand with both large (~ 6 nm) and small (~ 3.5 nm) proteins bound.

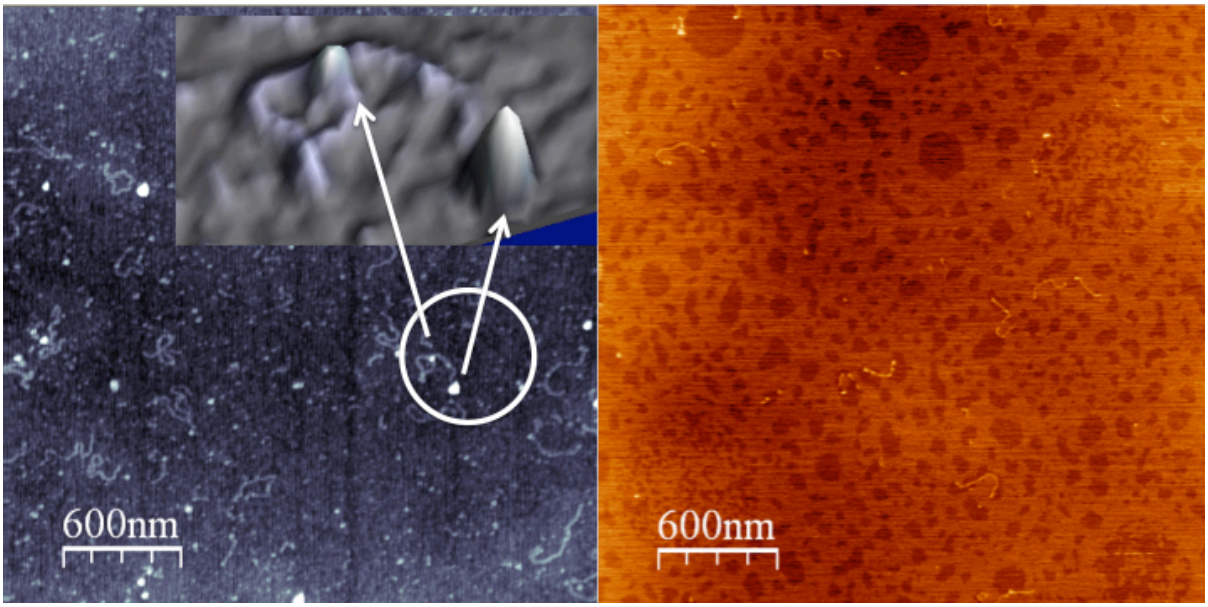
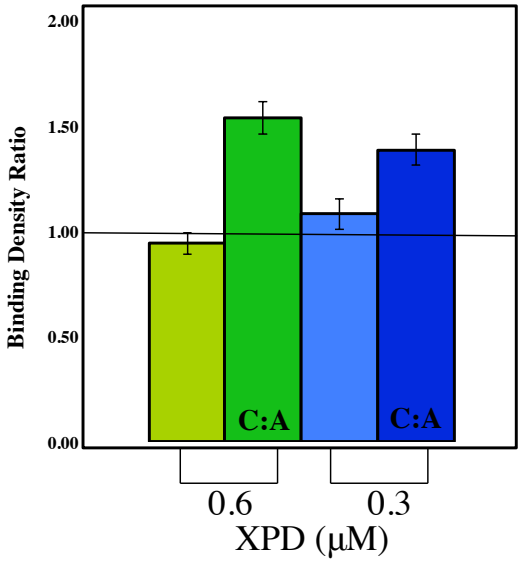


Figure 5.8. Quantitation of XPD (0.6 μM) and XPD (0.3 μM) protein density ratios (< 10% uncertainty) where C:A indicates a mismatch is contained in long strands. XPD redistributes at both concentrations.



In previous studies, Y82A did not redistribute in mismatched samples (r = 0.9 ± 0.1) (*22*). When XPD is mixed with Y82A 1:1 (WT^{XPD}/$Y82A^{EndoIII}$), there is no redistribution onto mismatched strands (r = 0.98 ± 0.05). WT^{XPD}/$Y82A^{EndoIII}$ matched controls are within error of mismatched results with a binding density ratio of 1.11 ± 0.07 (Figure 5.9). The number of proteins bound per kilobasepair remains the same between WT^{XPD}/$Y82A^{EndoIII}$ and WT^{XPD}/$WT^{EndoIII}$ mixtures (0.13 proteins/kbp) (Table 5.1). When WT^{XPD} is titrated into $Y82A^{EndoIII}$ at a ratio of 3:1, increasing the probability that two proteins bound on the same strand are both electrochemically active, the proteins once again redistribute onto mismatched strands (Figure 5.10).

Complementing WT^{XPD}/$WT^{EndoIII}$ and WT^{XPD}/$Y82A^{EndoIII}$ mixture studies, we also investigated signaling between EndoIII and XPD mutant L325V. Similar to EndoIII Y82A, XPD L325V cannot redistribute onto mismatched strands. WT^{XPD}/$L325V^{XPD}$ 1:1 mixtures were examined initially to determine whether L325V, deficient in DNA CT, could signal XPD. WT^{XPD}/$L325V^{XPD}$ mixtures showed no redistribution (r = 0.88 ± 0.04). This result was within error of matched controls, with a binding density ratio revealing even a slight preference for short strands (r = 0.90 ± 0.05). L325V inhibits XPD from relocalizing in the vicinity of lesions, but does L325V have any influence on the search for damage by EndoIII BER protein? $WT^{EndoIII}$/$L325V^{XPD}$ mixtures were examined and exhibited no difference in mismatched versus matched samples, where r = 1.02 ± 0.05 and 0.94 ± 0.04, respectively. As with the WT^{XPD}/$Y82A^{EndoIII}$ mixture, here we see that L325V binding alters EndoIII redistribution. Figure 5.9 shows binding density ratio comparisons for WT^{XPD}/$WT^{EndoIII}$, WT^{XPD}/$Y82A^{EndoIII}$, $WT^{EndoIII}$/$L325V^{XPD}$, and WT^{XPD}/$L325V^{XPD}$ 1:1 protein mixtures.

In the case of the mixtures of active proteins and CT-deficient mutants, we might have expected binding density ratios to be equal to that of pure samples of active repair protein. However, we are testing these redistributions at protein loadings on the DNA strands of ~ 2 per strand. Thus if one of the proteins is CT-defective, there is no protein to which the CT-proficient protein may signal.

Figure 5.9. Summary of binding density ratios for XPD (red text)/EndoIII (blue text) and mutant mixtures. Quantitation of protein density ratios (< 10% uncertainty) where C:A indicates a mismatch is contained in long strands. $WT^{XPD}/WT^{EndoIII}$ mixtures 1:1 (purple), both proteins CT proficient, redistribute onto mismatched strands. $WT^{XPD}/Y82A^{EndoIII}$ and $WT^{EndoIII}/L325V^{XPD}$ 1:1 mixtures (green and orange, respectively), Y82A and L325V both mutants deficient in CT, do not show redistribution. $WT^{XPD}/L325V^{XPD}$ 1:1 mixtures (blue) do not show redistribution, as XPD cannot perform DNA-mediated signaling if the L325V mutant is present.

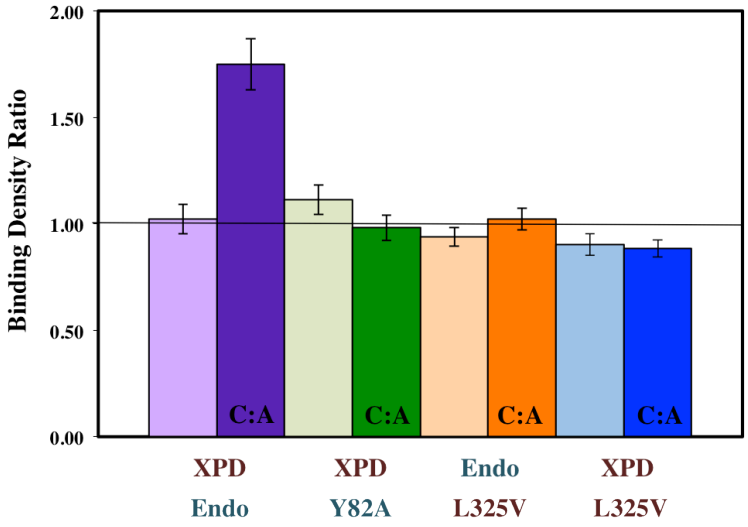


Figure 5.10. Titration with WT XPD. Quantitation of protein density ratios (< 10% uncertainty) where C:A indicates a mismatch is contained in long strands. WT^{XPD}/$Y82A^{EndoIII}$ mixtures 1:1 (purple), both proteins CT proficient, redistribute onto mismatched strands. WT^{XPD}/$Y82A^{EndoIII}$ 1:1 mixtures (green), Y82A mutant deficient in CT, do not show redistribution. WT^{XPD}/ $Y82A^{EndoIII}$ 3:1 mixtures (teal) show that redistribution redistribution is restored, as WT XPD is present at higher concentrations.

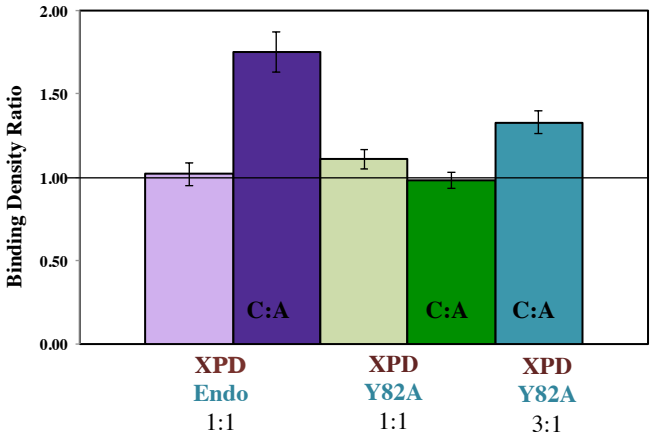


Table 5.1. Binding Density Ratios for Atomic Force Microscopy Samples

Sample	BDR[a,b]	Long Strands	Proteins	Short Strands	Proteins	Error (%)[c]	Prot./Kbp
0.6 uM XPD	0.94	307	216	360	142	5.3	0.19
0.6 uM XPD (CA)	1.54	354	304	323	95	5	0.20
0.3 uM XPD	1.08	304	123	502	99	6.7	0.11
0.3 uM XPD (CA)	1.38	376	271	354	97	5.2	0.18
L325V 0.6 uM	1.05	221	149	251	85	6.5	0.18
L325V 0.6 uM (CA)	1.14	311	236	415	145	5	0.19
XPD/EndoIII 1:1	1.02	221	127	257	76	7	0.15
XPD/EndoIII 1:1 (CA)	1.75	253	150	370	66	7.6	0.13
XPD/Y82A 1:1	1.11	468	183	665	123	6	0.10
XPD/Y82A 1:1 (CA)	0.98	496	258	538	150	5	0.14
EndoIII/L325V 1:1	0.94	327	263	412	186	4.7	0.22
EndoIII/L325V 1:1 (CA)	1.02	361	217	553	171	5.1	0.16
XPD/L325V 1:1	0.90	309	174	408	134	5.7	0.16
XPD/L325V 1:1 (CA)	0.88	564	329	653	227	4.2	0.16

a) Binding Density Ratio (BDR) is reported as (number of proteins bound on long strands (3.6 kbp)) /number of long strands) divided by the (number of proteins on short strands (1.9 kbp))/number of short strands) normalized for length. (CA) represents samples with long mismatched strands.

b) These data represent three individual samples with 10–20 images per sample. L325V is a mutant of wild-type XPD, and Y82A is a mutant of wild-type EndoIII.

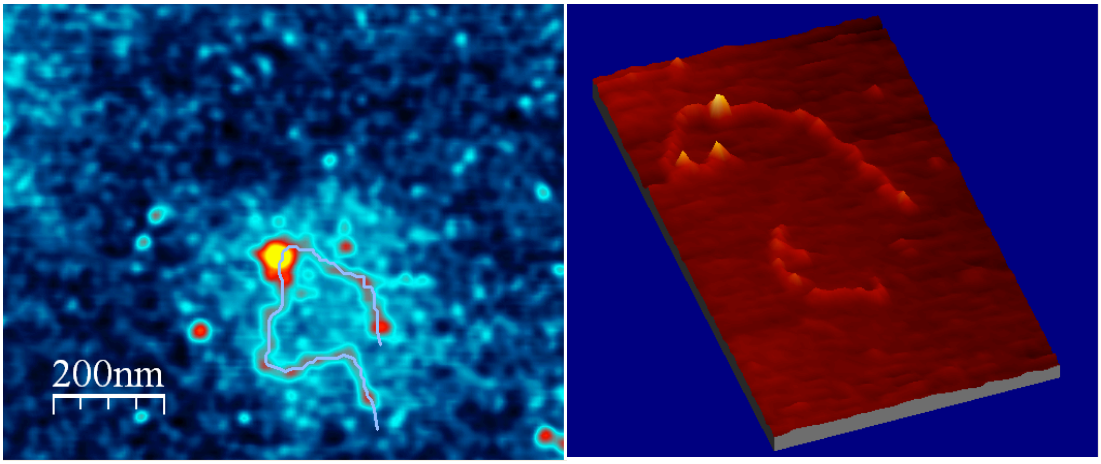
c) Error is reported as the (total number of proteins divided by the square root of the total number of proteins counted) times 100 for each type of sample.

Certainly the striking differences seen between CT-active protein mixtures and those containing CT-deficient mutants support the idea that the proteins can carry out DNA CT to one another. Having established that this ability to redistribute in the vicinity of mismatches depends on the ability of the protein to carry out DNA CT (*23*), these results thus provide evidence for DNA CT as a means of inter-protein signaling.

Discussion

The data presented here indicate that XPD, an archaeal protein from the NER pathway, may cooperate with other proteins that are proficient at DNA CT to localize in the vicinity of damage. XPD, a superfamily 2 DNA helicase with 5′-3′ polarity, is a component of TFIIH that is essential for repair of bulky lesions generated by exogenous sources such as UV light and chemical carcinogens (*28, 49, 50*). XPD contains a conserved [4Fe4S] cluster suggested to be conformationally controlled by ATP binding and hydrolysis (*25*). Mutations in the iron-sulfur domain of XPD can lead to diseases including TTD and XP, yet the function of the [4Fe4S] cluster appears to be unknown (*23, 28, 29*). Electrochemical studies have shown that when BER proteins MutY and EndoIII bind to DNA, their [4Fe4S] clusters are activated toward one electron oxidation (*14, 26*). XPD exhibits a DNA-bound midpoint potential similar to that of EndoIII and MutY when bound to DNA (~ 100 mV vs. NHE), indicative of a possible role for the [4Fe4S] cluster in DNA-mediated CT (*34*). For EndoIII we have also already determined a direct correlation between the ability of proteins to redistribute in the vicinity of mismatches as measured by AFM, and the CT proficiency of the proteins measured electrochemically (*23*). Thus, we may utilize single-molecule AFM as a tool to probe the redistribution of proteins in the vicinity of base lesions (Figure 5.11) and in so doing, the proficiency of the protein to carry out DNA CT.

Figure 5.11. Representative tapping mode AFM images of XPD. (Left) Zoomed-in view of a DNA strand with XPD protein bound with height scale shown in color (yellow is ~ 4–6 nm). (Right) AFM to visualize multiple XPD proteins bound to a single matched strand of DNA.



Here we show that, like the BER protein EndoIII, XPD, involved both in transcription and NER, redistributes in the vicinity of a lesion. Importantly, this ability to relocalize is associated with the ability of XPD to carry out DNA CT. The mutant L325V is defective in its ability to carry out DNA CT and this XPD mutant also does not redistribute effectively onto the mismatched strand.

Moreover, these data provide evidence that two different repair proteins, each containing a [4Fe4S] cluster at similar DNA-bound potential, can communicate with one another through DNA-mediated CT. This result becomes more interesting still given that in the experiments conducted here, the proteins are from completely different organisms. Nonetheless, what is critical is that the protein clusters have similar DNA-bound potentials, facilitate many electron exchanges, and have the ability to carry out DNA-mediated CT. Furthermore, no signaling to effect the redistribution in the vicinity of lesions occurs when one partner is CT-deficient.

Lesion detection by repair proteins, based on our model, depends on the maximum distance over which DNA-mediated CT can occur and the percentage of proteins oxidized (*6, 22, 51*). While we have documented that DNA CT occurs over at least 100 bp, we do not yet know the maximum distance for this process (*5, 52*). If CT could proceed only over short distances (<500 bp), more than 6 proteins would be required for signaling across strands. Instead, only 1–3 proteins are bound to the long matched or mismatched strands in AFM experiments, suggesting that DNA CT is occurring over distances much greater than those we have measured electrochemically.

How might this signaling be utilized inside the cell? Surely, this ability to redistribute in the vicinity of lesions reduces the search process required to find lesions across the genome. The higher the concentration of total proteins involved in signaling that are at similar potentials, the more efficient the search process becomes (*22*). Indeed, XPD may utilize DNA-mediated CT to signal its presence and perhaps to “call off” other proteins from other repair pathways. Various DNA-binding proteins, such as those involved in repair and other DNA transactions, have been found to

contain iron-sulfur domains and other redox cofactors (*53-56*). DNA-mediated signaling among DNA-binding proteins that are involved in maintaining the integrity of the genome allows a coordination of repair, transcription and replication processes.

REFERENCES

1. Kuraoka, I., Bender, C., Romieu, A., Cadet, J., Wood, R. D., and Lindahl, T. (2000) *Proc. Natl. Acad. Sci. USA* 97, 3832.

2. Sancar, A., Lindsey-Boltz, A., Ünsal-Kaçmaz, K., and Linn, S. (2004) *Annu. Rev. Bioch. 73*, 39.

3. David, S. S., and Williams, S. D. (1998) *Chem. Rev. 98*, 1221.

4. David, S. S., O'Shea, V. L., and Kundu, S. (2007). *Nature 447*, 941.

5. Slinker J. D., Muren, N. B., Renfrew, S. E., and Barton, J. K. (2011) *Nature Chem. 3*, 230.

6. Núñez, M. E., Hall, D. B., and Barton, J. K. (1999) *Chem Biol. 6*, 85.

7. Kelley, S. O., Boon, E. M., Barton, J. K., Jackson, N. M., and Hill, M. G. (1999) *Nucleic Acids Res. 27*, 4830.

8. Boal, A. K., and Barton, J. K. (2005) *Bioconj. Chem. 16*, 312.

9. Genereux, J. C., Boal, A. K., and Barton J. K. (2010) *J. Am. Chem. Soc. 132*, 891.

10. Boon, E. M., Livingston, A. L., Chmiel, N. H., David, S. S., and Barton, J. K. (2003) *Proc. Natl. Acad. Sci. USA 100*, 12543.

11. Barton, J. K., Olmon, E. D., and Sontz, P. A. (2011) *Coord. Chem. Rev. 255*, 619.

12. Kuo, C. F., McRee, D. E., Fisher, C. L., O'Handley, S. F., Cunningham, R. P., and Tainer, J. A. (1992) *Science 258*, 434.

13. Thayer, M. M., Ahern, H., Xing, D., Cunningham, R. P., and Tainer, J. A. (1995) *EMBO J. 14*, 4108.

14. Gorodetsky, A. A., Boal, A. K., and Barton, J. K. (2006) *J. Am. Chem. Soc. 128*, 12082.

15. Porello, S. L., Cannon, M. J., and David, S. S. (1998) *Biochemistry 37*, 6465.

16. Lukianova, O. A., and David, S. S. (2005) *Curr. Opin. Chem. Biol. 9*, 145.

17. Fromme, J. C., and Verdine, G. L. (2003) *EMBO J. 22*, 3461.

18. Genereux, J. C., Boal, A. K., and Barton, J. K. (2010) *J. Am. Chem. Soc. 132*, 891.

19. Fromme, J. C., Banerjee, A., Huang, S. J., and Verdine, G. L. (2004) *Nature 427*, 652.

20. Cheadle J. P., and Sampson, J. R. (2007) *DNA Repair 6*, 274.

21. Fodde, R. (2002) *Eur. J. Cancer 38*, 867.

22. Boal, A. K., Genereux, J. C., Sontz, P. A., Gralnick, J. A., Newman, D.K., and Barton, J. K. (2009) *Proc. Natl. Acad. Sci. 106,* 15237.

23. Romano, C. A., Sontz, P. A., and Barton, J. K. (2011) *Biochemistry 50*, 6133.

24. Gorbalenya, A. E., and Koonin, E. V. (1993) *Curr. Opin. Struct. Biol. 3*, 419.

25. Fan, L., Fuss, J. O., Cheng, Q. J., Arvai, A. A., Hammel, M., Roberts, V. A., Cooper, P. K., and Tainer, J. A. (2008) *Cell 133*, 789.

26. Boal, A. K., Yavin, E., Lukianova, O. A., O'Shea, V. L., David, S. S., and Barton, J. K. (2005) *Biochemistry 44*, 8397.

27. Yavin, E., Boal, A. K., Stemp, E. D. A., Boon, E. M., Livingston, A. L., O'Shea, V. L., David, S. S., and Barton, J. K (2005) *Proc. Natl. Acad. Sci. 102*, 3546.

28. Lehmann, A. R. (2008) *DNA Repair 7*, 1912.

29. Fuss, J. O., and Tainer, J. A. (2011) *DNA Repair 10*, 697.

30. Takayama, K., Salazar, E. P., Broughton, B. C., Lehmann, A. R., Sarasin, A., Thompson, L. H., and Weber, C. A. (1996) *Am. J. Hum. Genet. 58*, 263.

31. Taylor E. M., Broughton, B. C., Botta, E., Stefanini, M., Sarasin, A., Jaspers, N. G. J., Fawcett, H., Harcourt, S. A., Arlett, C. F., and Lehmann, A. R. (1997) *Proc. Natl. Acad. Sci. USA 94*, 8658.

32. Rudolf, J., Makrantoni, V., Ingledew, W. J., Stark, M. J. R., and White, M. F. (2006) *Molecular Cell 23*, 801.

33. Wolski S. C., Kuper, J., Hänzelmann, P., Truglio, J. J., Croteau, D. L., Van Houten, V., and Kisker, C. (2008) *PLos Biol. 6*, 1332.

34. Mui, T. P., Fuss, J. O., Ishida, J. P., Tainer, J. A., and Barton, J. K. (2011) *J. Am. Chem. Soc. 133*, 16378.

35. Rudolf, J., Rouillon, C., Schwarz-Linek, U., and White, M. F. (2010) *Nucleic Acids Res. 38*, 931.

36. Sun, H. B., and Yokota, H. (2000) *Anal. Chem. 72*, 3138.

37. Sorel, I., Piétrement, O., Hamon, L., Baconnais, S., Le Cam, E., and Pastré, D. (2006) *Biochemistry 45*, 14675.

38. Pastré, D., Hamon, L., Sorel, I., Le Cam, E., Curmi, P. A., and Piétrement, O. (2010) *Langmuir 26*, 2618.

39. Jiang, Y., and Marszalek, P. E. (2011) *EMBO J. 30*, 2881.

40. Davis, R. W., Botstein, D., and Roth, J. R. (1980) *Advanced Bacterial Genetics: A Manual for Genetic Engineering* (Cold Spring Harbor Laboratory Press).

41. Donahue, W. F., Turczyk, B. M., and Jarrell, K. A. (2002) *Nucleic Acids Res. 30*, e95.

42. Cunningham, R. P., Asahara, H., Bank, J. F., Scholes, C. P., Salerno, J. C., Surerus, K., Munck, E., McCracken, J., Peisach, J., and Emptage, M. H. (1989) *Biochemistry 28,* 4450.

43. Horcas I., Fernández, R., Gómez-Rodríguez, J. M., Colchero, J., Gómez-Herrero, J., and Baro, A. M. (2007) *Rev. Sci. Instrum. 78*, 013705.

44. Boon, E. M., Ceres, D. M., Drummond, T. G., Hill, M. G., and Barton, J. K. (2000) *Nat. Biotech. 18*, 1096.

45. Rudolf, J. (2006) Thesis submitted for the degree of Doctor of Philosophy, University of St. Andrews, Scotland, UK.

46. *Charge Transfer in DNA: From Mechanism to Application* (2005), ed. Wagenknecht H. A. (Wiley- VCH, Weinheim).

47. Berg, O. G., Winter, R. B., and von Hippel, P. H. (1981) *Biochemistry 20*, 6929.

48. *Atomic Force Microscopy Biomedical Methods and Applications* (2004), ed. Braga P. C., Ricci D. (Humana Press; Totowa, NJ).

49. Liu H., Rudolf, J., Johnson, K. A., McMahon, S. A., Oke, M., Carter, L., McRobbie, A. M., Brown, S. E., Naismith, J. H., and White, M. F. (2008) *Cell*, *133*, 801.

50. Broughton, B. C., Berneburg, M., Fawcett, H., Taylor, E. M., Arlett, C.F., Nardo, T., Stefanini, M., Menefee, E., Price, V. H., Queille, S., Sarasin, A., Bohnert, E., Krutmann, J., Davidson, R., Kraemer, K. H., and Lehmann, A. R. (2001) *Hum. Mol. Gen. 19*, 2539.

51. Yavin, E., Stemp, E. D. A., O'Shea, V. L., David, S. S., and Barton, J. K. (2006) *Proc. Natl. Acad. Sci. USA 103*, 3610.

52. Núñez, M. E., Noyes, K. T., and Barton, J. K. (2002) *Chem. Biol. 9*, 403.

53. Lee, P. E., Demple, B., and Barton, J. K. (2009) *Proc. Natl. Acad. Sci. 106*, 13164.

54. Augustyn, K. E., Merino, E. J., and Barton, J. K. (2007) *Proc. Natl. Acad. Sci. USA 104*, 18907.

55. Merino, E. J., and Barton, J. K. (2008) *Biochemistry 47*, 1511.

56. Klinge, S., Hirst, J., Maman, J. D., Krude, T., and Pellegrini, L. (2007) *Nat. Struct. Mol. Biol. 14*, 875.

CHAPTER 6

Visualizing XPD Helicase Activity

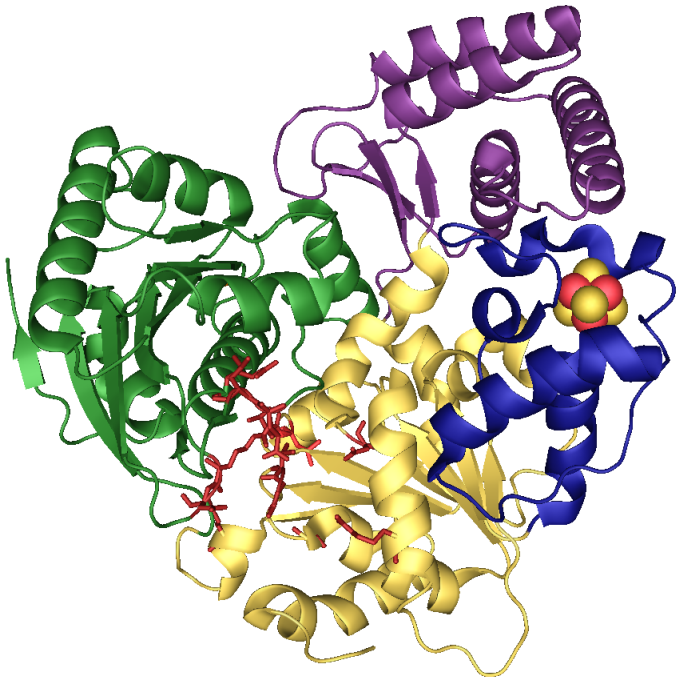
P.A.S. designed and performed the AFM experiments with summer undergraduate student, Alison Parisian. J.O.F and J.A.T. provided XPD.

INTRODUCTION

In the cell, DNA is under incessant attack from sources including reactive oxygen species, UV-light, and metabolic byproducts (*1*). Subsequently, lesions that arise threaten the integrity of the genome. Cells, however, possess a variety of repair pathways to efficiently detect and fix sites of damage. One of these, nucleotide excision repair (NER), is thought to be the most versatile of these pathways (*2*). Specifically, NER repairs lesions that largely distort the helical structure of DNA; guanine-cisplatin adducts and benzo[a]pyrene-guanine adducts are distinct examples and notably, in humans, NER machinery is the only defense against photoproducts (*2*). In particular, XPD and XPB, 5′-3′ and 3′-5′ helicases, respectively, are components of the ten-subunit transcription factor (TFIIH) responsible for binding near the damaged site and unwinding the DNA that surrounds the lesion, thereby exposing it for repair (*2*). Mutations in XPD and XPB have been associated with severe diseases such as xeroderma pigmenostum, Cockayne syndrome, and trichothiodystrophy (*2-4*). The catalytic core of XPD is comprised of four domains: helicase domain 1 (HD1), helicase domain 2 (HD2), Arch domain, and of interest to us, a [4Fe4S] domain (Figure 6.1) (*3, 4*). Over 50% of the known point mutations that are associated with diseases are found in these regions (*3*). Additionally, XPD contains a loop motif in the [4Fe4S] cluster domain similar to the FeS cluster loop motif (FCL) found in MutY, a BER protein from *E. coli* that exhibits a DNA-bound redox potential of ~ 100 mV vs. NHE (*2, 5, 6*). Two arginines, R88 in XPD and R153 in MutY, are conserved, similarly oriented, and shield the [4Fe4S] cluster (*2*). We then ask whether the resemblance between MutY and XPD extends beyond the presence of an [4Fe4S] cluster to a relationship, mediated by DNA charge transport (CT), between proteins to cooperate in the search for damage within the genome.

Helicase activity of XPD is required for repair but not for transcription (*3*). Because of its proximity to the ATP-binding site (HD1/HD2 interface), the [4Fe4S] cluster may be coupled to

Figure 6.1. SaXPD crystal structure with domains HD1 (yellow) and HD2 (green) forming the ATP-binding interface (red). The [4Fe4S] domain holds the cluster (blue: residues 81–149), and the Arch domain is in purple. Structure constructed based on reference (*3*) using pdb ID: 3crv.



ATP binding and hydrolysis (*3*). Wild-type XPD has an ATP-hydrolysis rate of 0.55 mol ATP/mol/s and a rate of unwinding ssDNA containing a 5′ overhang of 2.22 bp/min (human XPD: K_d=46 nM) (*3*). Loss of the [4Fe4S] cluster abolishes the helicase activity, however, this is not necessarily coupled to the ability of the protein to bind DNA nor to its structure and folding properties (*3*). For example, a conserved mutation found in disease phenotypes, G34R lacks ATPase activity and helicase activity yet it binds ssDNA with an affinity 155% that of wt XPD (*3*). K84H, a mutation at a residue near the [4Fe4S] cluster decreases helicase activity but does not affect ATPase rates. K84H XPD binds ssDNA with an affinity (K_d=80 nM) ~ 58% that of wt (*3*). Currently, there is no crystal structure of XPD bound to DNA, but recently, we have employed a variety of techniques to investigate whether XPD is indeed another protein containing a [4Fe4S] cluster that exhibits a DNA-bound redox potential conducive to relaying charge to and from other proteins (*6-10*).

Using electrochemical methods, we have demonstrated that XPD has a DNA-bound potential similar to that of BER enzymes we have examined the past (~ 80 mV vs. NHE) (*9*). Moreover, we directly observe that the efficiency of DNA CT is enhanced in an ATP-dependent manner (*9*). As ATP is added to XPD bound to DNA on an electrode, the current increases. When we substitute ATP with a non-hydrolyzable analogue, XPD is not able to perform DNA CT as efficiently. Also, consistent with biochemical studies by Tainer et al. (*3*), the SaXPD G34R mutant described above does not mediate DNA CT as well as wt XPD (*9)*. Thus, we have used this assay not only to examine the DNA-bound redox potential of XPD but also to directly measure helicase activity. Does the ability of XPD to find lesions correlate with the addition of ATP? Does the addition of ATP help XPD “lock in” to the DNA strand to cooperate with other proteins in a DNA-mediated search?

To test this, we have employed single-molecule atomic force microscopy (AFM) to investigate

the ability of XPD to find lesions in strands of DNA (*10*). In an effort to link the efficiency of XPD to perform DNA CT with its biochemical function, here, we induce the helicase activity of XPD by incorporating ATP in the experiments, at concentrations comparable to those applied in electrochemical studies. Other AFM studies with helicases have established that it is indeed possible to track proteins as they unwind DNA (*11-16*). Notably, as these enzymes separate the DNA duplex, strands will appear as "forks" in high-resolution images (*11*). To this end, we have substituted ATP with a caged-ATP analogue, which, importantly, we can trigger with UV-light, thereby releasing energy to activate XPD helicase. Upon irradiation, we observe forked DNA structures, indicative of XPD helicase activity, yet we are still working to increase the speeds with which we gather images. With improved resolution, we hope to observe XPD redistribution and helicase activity in "real-time".

Materials and Methods All chemicals were purchased from Sigma Aldrich. All enzymes were purchased from New England Biolabs unless otherwise specified. Mica surfaces were purchased from SPI supplies. Silicon AFM probes were purchased from Budget Sensors. Oligonucleotides were purchased from IDT or synthesized on a 3400 DNA synthesizer (Applied Biosystems).

Mismatched (C:A) Strand Synthesis Mismatched and matched strands were prepared according to the protocol described in Chapter 5.

Protein Purification and Expression Protein concentrations were determined using the UV-visible absorbance of the [4Fe4S] cluster (410 nm, ε = 17,000 $M^{-1}cm^{-1}$) (*17*). XPD was purified as previously described (*9*).

AFM Experiments AFM experiments were performed using the protocol similar to that reported previously (*8, 10*). Stock DNA solution contained 50–200 ng of total DNA (~ 6 μM) composed of the mixture of ligated 3.8 kbp duplexes and the two unligated duplexes (1.6 and 2.2 kb) in 6 mM $MgCl_2$/Tris-EDTA buffer. The 1.6 kb overhang complement (60 μM) was added to the DNA solution to block the 14 bp single-strand overhangs generated by PCR. This was then incubated overnight at 4 °C. XPD protein was dialyzed against the protein buffer (20 mM phosphate, 100 mM NaCl, 1 mM EDTA, 5% glycerol, pH 7.5 and filtered prior to use) to remove residual DTT. The concentration of individual proteins were determined by UV/visible spectrophotometry (Beckman DU 7400) using ε=17,000 $M^{-1}cm^{-1}$ at 410 nM for the [4Fe-4S] cluster. After addition of excess 2.2 kb duplex complement (60 μM), XPD (0.6 μM) was added to the stock DNA solution. This protein/DNA solution was incubated at 4 °C overnight**.** Sample was then deposited (5–10 μL) onto a freshly cleaved mica surface for 1–2 min, rinsed with 2 mL of water, and dried under argon.

AFM Studies of XPD + 2 mM ATP Samples containing 34.25 μL Tris EDTA buffer, 14 μL 20 mM MgCl2 (to a concentration of 6 mM), 0.75 μL of the appropriate DNA ligation products (6 μM), and 0.66 μL (60 μM) complementary strand (primer 2 in excess of DNA) were prepared and incubated for 24 hours at 4 ºC. The samples were removed from the refrigerator and complementary strand (primer 3 for matched primer 4 for mismatched) was added in excess. The sample was vortexed and quickly centrifuged to mix the contents, then split into two 25 μL aliquots. 1.5 μL of 10 μM XPD protein was added to each aliquot (to a concentration of 0.6 μM) and samples were incubated for another 24 hours at 4 ºC. For 'incubated' samples, 1 μL of 50 mM ATP stock was added to each sample (2 mM working concentration) directly after protein addition and incubated with the sample for 24 hours. For 'deposition samples' ATP was added to

the sample (2 mM) immediately before surface deposition.

AFM Studies of XPD + 5 mM ATP Samples containing 34.25 µL Tris EDTA buffer, 12.5 µL 20 mM $MgCl_2$ (to a concentration of 5 mM), and 0.75 µL of the appropriate DNA ligation products (6 µM) were prepared and incubated for 24 hours at 4 ºC. The samples were removed from the refrigerator and split into two 25 µL aliquots. 1.5 µL of 10 µM XPD protein was added to each aliquot (to a concentration of 0.6 µM) and the samples were incubated for another 24 hours at 4 ºC. For 'incubated' samples, 2.5 µL of 50mM ATP stock (5 mM working concentration) was added to each sample directly after protein addition and incubated with the sample for 20 minutes. For 'deposition samples' ATP was added to the sample (5 mM) immediately before surface deposition.

Caged ATP studies An initial sample of 34.25 µL Tris EDTA buffer, 14 µL 20 mM $MgCl_2$ (to a concentration of 6 mM), and 0.75 µL the appropriate ligation reaction of DNA (6 µM) was prepared and incubated for 24 hours at 4 ºC. The sample was then removed from the refrigerator and split into two 25 µL aliquots. 1.5 µL of 10 µM XPD protein was added to each aliquot (to a concentration of 0.6 µM) and samples were incubated for another 24 hours at 4 ºC. In incubated samples, 1 µL of 50 mM NPE-caged ATP (Invitrogen) was added (2 mM) directly following the addition of protein and incubated with the sample for 24 hours. As an alternative method, ATP was added (2 mM) 20 minutes before deposition onto the surface. The sample was not dried to completion. The sample was irradiated either after it was deposited onto the mica surface, or from directly above the eppendorf test tube. Prior to activation of caged ATP (355 nm) with a small handheld light, the overhead lights were kept off. The imaging procedure is still being optimized.

AFM to Probe XPD Helicase Activity 6 mM $MgCl_2$ and 6 μM DNA in Tris-EDTA buffer is incubated overnight at 4 °C. 0.6 μM XPD was added, and the sample was incubated overnight once again with 5 mM ATP alongside a control that has water added instead of ATP. An aliquot of the sample was be imaged at this point. XPD will preferentially bind to the short strand overhangs (14 bp).

AFM Instrumentation Silicon AFM Probes purchased from Budget Sensors, with a spring constant of 3 N/m and a resonance frequency of 75 kHz, were used in a Digital Instruments Multimode SPM. Images were captured in air with scan areas of 2 x 2 μm^2 or 3 x 3 μm^2 in tapping mode, at an amplitude of 0.54–2.00 V and at a scan rate of 3.05 Hz. Scan rates of 3.05 Hz were used in order to obtain images of higher quality.

Binding Density Ratio Calculations WSxM software was used to measure general DNA contour lengths and height profiles of the proteins as described previously (*8, 10*). For each data set, images from at least three independent samples were analyzed, compared, and pooled (> 200 long or short strands). Distinguishable strands and protein positions were counted by hand. The binding density ratio, r, is defined as the ratio of the proteins bound on long strands divided by proteins bound on short strands. The ratio is normalized for length by dividing by 1.9 kbp, which is the average length of the short strands. Binding affinities were found by determining the number of proteins bound per kilobasepair strand. The uncertainty was determined through the total number of proteins observed.

Results and Discussion

Atomic Force Microscopy of XPD + ATP We have previously demonstrated with BER proteins that the ability of a protein to perform DNA CT directly correlates to its redistribution in the vicinity of base lesions or mismatches that inhibit CT (*8, 10*). We have tested XPD, an NER protein, with respect to redistribution promoted through DNA CT by preparing DNA strands containing a single C:A mismatch, a modification that we know to inhibit CT (*18*), alongside DNA strands containing no mismatches. Although a C:A mismatch effectively inhibits DNA CT (*18*), it is not a lesion that is preferentially bound by XPD. However, the 14-nucleotide overhangs generated with PCR are specific substrates for XPD helicase (K_d~1 μM) (*19, 20*). To block XPD from binding to the 14 bp segment, overhang complements were added to DNA/protein solutions in excess. We found, when we blocked the overhangs, no preferential binding to the short strands was observed. Protein assignments were verified through analysis of their 3–4 nm heights in the images; without protein, features of this dimension were not observed and larger heights (> 7 nm) indicated salt precipitates or protein aggregation. Only clearly identifiable long or short strands and bound proteins were counted. XPD showed redistribution onto long mismatched strands with a ratio of protein binding densities, r (long/short) of 1.54 ± 0.08. If, instead, we examined XPD distribution when long and short strands were fully matched, we observed a binding density ratio of 0.94 ± 0.05 (*10*).

Using electrochemical techniques, we have previously observed an increase in the efficiency of XPD to perform DNA CT in an ATP-dependent manner. In order to understand whether ATP also enhances the propensity of the protein to find lesions in the genome, we have performed AFM studies to measure XPD redistribution in the presence of ATP. Here, wild-type (wt) XPD was mixed with 2 mM ATP, the energy source for the enzyme, and DNA. Short duplex overhangs were blocked with excess complementary single strand. Figure 6.2 shows a representative image of DNA strands incubated for 24 hours with XPD and 2 mM ATP. In parallel, XPD and DNA were treated

with 2 mM ATP and immediately deposited for imaging. Consistent with XPD redistribution in the absence of ATP, as shown in Figure 6.3, we find that XPD with 2 mM ATP redistributes onto mismatched strands when the samples are immediately deposited ($r = 1.35 \pm 0.06$), we expect a ratio of 1 if there were an equal distribution of proteins on matched and mismatched strands (*7*). However, when the samples are incubated overnight with 2 mM ATP, there is a preference for the short strands both when strands are matched and mismatched ($r = 0.78 \pm 0.05$ and $r = 0.83 \pm 0.04$, respectively) (Table 6.1), similar to what we would expect if we did not block the overhangs.

Importantly, we measured the relative binding affinity of XPD, and in all of the samples described, the number of proteins bound per base pair remained constant at ~ 0.20 proteins/kbp (Table 6.1). Thus, any changes we see in distribution are not due to differences in binding affinities. This result indicates that over the 24-hour incubation, even though we added complementary strands to block the overhangs, we have induced XPD helicase activity. Notably, the preference we observe in the incubated samples directly reflects the enzyme binding to DNA and beginning to unwind at the junction.

In order to minimize the bias of XPD toward short strands and increase our chances of observing an enhancement in protein redistribution, we decreased the sample incubation time to 20 minutes and changed the concentration of ATP to 5 mM. As Figure 6.4 shows, the baseline binding density ratios for all of the samples is greater than 1, though in each case we observe a slight increase in redistribution onto the mismatched strands.

Figure 6.2. Representative AFM image of XPD + 2 mM ATP. Proteins that bind to DNA strands are tallied for binding density ratio measurements. Scan size: 3 x 3 μm^2, scan rate: 3.05 Hz

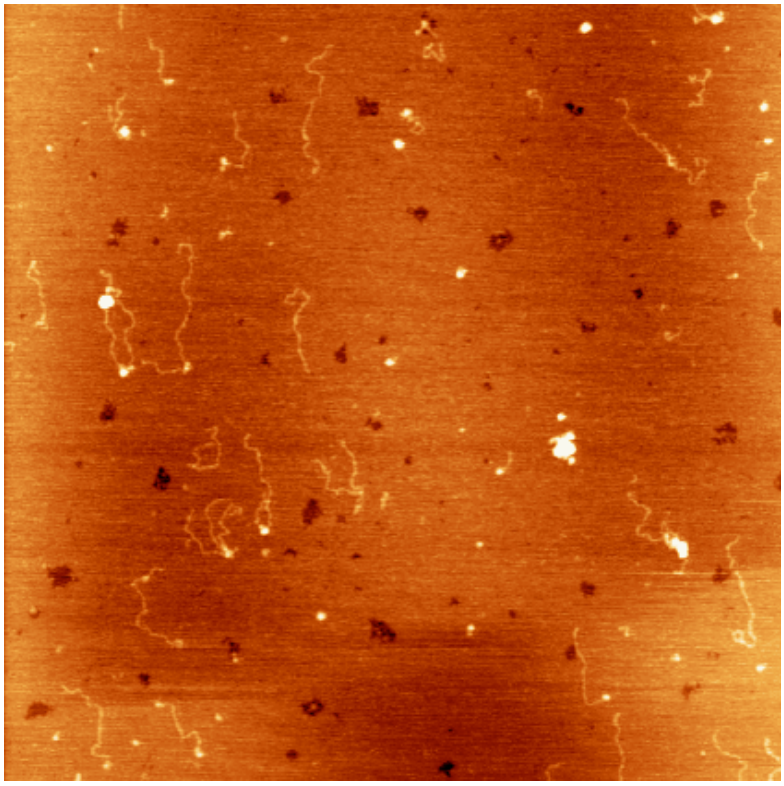


Figure 6.3. Quantitation of XPD density ratios (< 10% uncertainty) where C:A indicates a mismatch is contained in long strands. Samples were incubated with 2 mM ATP for 24 hours (left) or immediately deposited (right). Redistribution only occurs when ATP is added and the samples are immediately deposited and imaged by AFM.

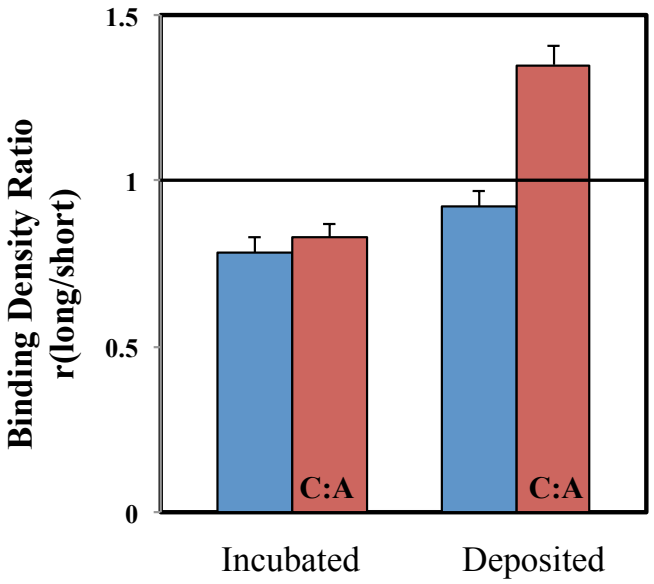


Table 6.1. Binding Density Ratios for Atomic Force Microscopy Samples XPD + 2 mM ATP

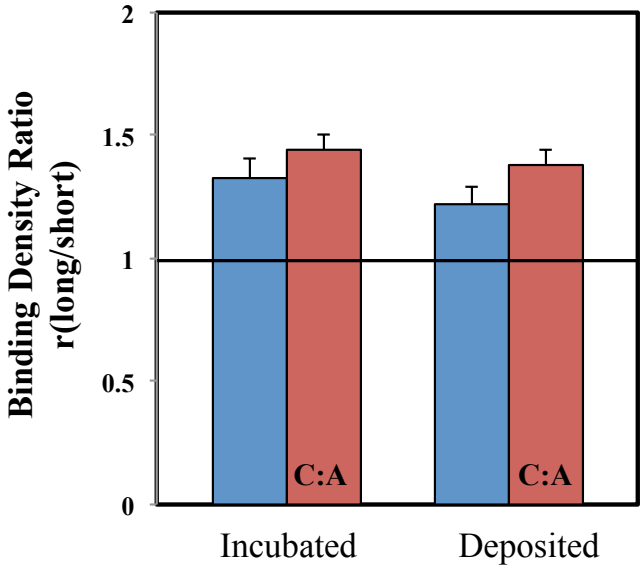
Sample (2 mM ATP)	BDR[a,b]	Long Strands	Proteins	Short Strands	Proteins	Error (%)[c]	Prot./Kbp
Incubated	0.78	258	171	246	110	6	0.19
Incubated (CA)	0.83	393	260	433	181	4.8	0.19
Deposited	0.92	410	291	315	128	4.9	0.19
Deposited (CA)	1.35	335	295	435	149	4.8	0.21

a) Binding Density Ratio (BDR) is reported as (number of proteins bound on long strands (3.6 kbp)) /number of long strands) divided by the (number of proteins on short strands (1.9 kbp))/number of short strands) normalized for length. (CA) represents samples with long mismatched strands.

b) These data represent three individual samples with 10–20 images per sample.

c) Error is reported as the (total number of proteins divided by the square root of the total number of proteins counted) times 100 for each type of sample.

Figure 6.4. Quantitation of XPD density ratios (< 10% uncertainty) where C:A indicates a mismatch is contained in long strands. Samples were incubated with 5 mM ATP for 20 minutes (left) or immediately deposited (right). XPD has a preference for the long strands in all of the samples.



For C:A mismatched samples that were incubated for 20 minutes, r = 1.44 ± 0.07 whereas in the fully matched samples r = 1.33 ± 0.08. The difference may be somewhat more pronounced in the values when the samples are immediately deposited: r = 1.38 ± 0.06 and r = 1.22 ± 0.07 mismatched and matched, respectively. Importantly, the number of proteins bound per base pair remained constant at ~ 0.20 proteins/kbp (Table 6.2). Notably, there were more features of larger heights (> 4 nm) in the samples that had 5 mM ATP. Thus, the high concentration of ATP may have inhibited sample deposition, disrupted the quality of the images, or increased DNA/protein aggregation, resulting in a biased redistribution ratio.

Caged-ATP Studies AFM studies provide us with a "snapshot" of protein redistribution, giving us insight on the first step of lesion detection. Similar to BER proteins that we have examined in the past, XPD is a repair protein from the NER pathway that exhibits a DNA-bound redox potential ~ 80 mV vs. NHE (*9*). Notably, however, XPD is also a 5′-3′ helicase, a property that we can explore in 'real-time.' Thus, we have employed a strategy to trigger the helicase activity of XPD in the presence of ATP. Following the protocol we developed to observe XPD with 2 mM ATP, here we replaced ATP with NPE-caged ATP (Figure 6.5), an analogue in which a blocking group is attached to the gamma phosphate. Upon irradiation with 355 nm light (UV), the analogue is deprotected and released in as native ATP.

Based on other AFM studies of helicases (*11-16*), DNA structures that are forked indicate that the protein has begun unwinding the strand. In our studies, we did not add complementary single strands, leaving the short strand overhangs free for XPD to bind. The DNA/protein samples were incubated for a few minutes and deposited onto a freshly cleaved mica surface. Images of DNA and protein complexes were acquired in the dark with a scan size of 2 x 2 μm^2 or 3 x 3 μm^2 at a scan rate of 1–3.05 Hz.

Table 6.2. Binding Density Ratios for Atomic Force Microscopy Samples XPD + 5 mM ATP

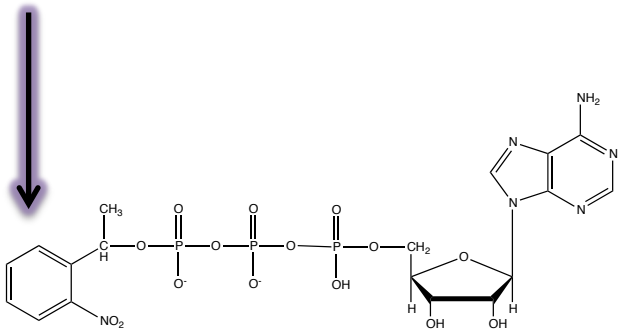
Sample (5 mM ATP)	BDR[a,b]	Long Strands	Proteins	Short Strands	Proteins	Error (%)[c]	Prot./Kbp
Incubated	1.33	210	168	256	81	6.3	0.19
Incubated (CA)	1.44	367	334	417	139	4.6	0.22
Deposited	1.22	316	228	289	90	5.6	0.18
Deposited (CA)	1.38	401	348	539	179	4.4	0.21

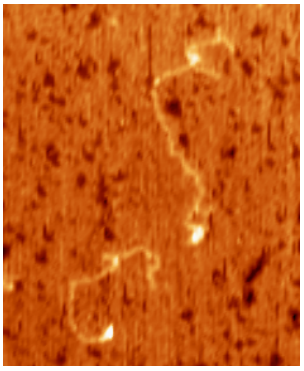
a) Binding Density Ratio (BDR) is reported as (number of proteins bound on long strands (3.6 kbp)) /number of long strands) divided by the (number of proteins on short strands (1.9 kbp))/number of short strands) normalized for length. (CA) represents samples with long mismatched strands.

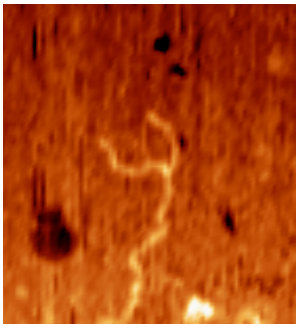
b) These data represent three individual samples with 10–20 images per sample.

c) Error is reported as the (total number of proteins divided by the square root of the total number of proteins counted) times 100 for each type of sample.

Figure 6.5. AFM of XPD/DNA treated with caged-ATP to visualize XPD helicase activity. (Top) Picture of AFM instrument with small UV-light to irradiated caged-ATP at 355 nm. Irradiation deprotects the compound, releasing ATP that the protein can utilize as an energy source. (Bottom) Schematic of XPD bound to duplex with short overhang. AFM images display DNA strands that are forked. XPD is bound at the junction.







Once we gathered an image of a DNA strand with bound protein, we irradiated the sample with a handheld UV-light while quickly scanning the region of interest. With this method, we could not detect any changes in DNA structure. In another method, the sample was irradiated in the test tube for 10 minutes prior to deposition onto the mica surface. Interestingly, in this example, we observed a large percentage of forked DNA structures. Significantly, many of the forks exhibited XPD bound at the junction (Figure 6.5). We are now optimizing this strategy so that we can observe changes in the DNA structure as a function of helicase activation (sample irradiation).

Visualizing XPD Helicase Activity Complementing AFM studies of XPD with caged-ATP to visualize helicase activity, we have imaged XPD in the absence and presence of 5 mM ATP. We can compare the percentage of forked DNA structures, which are indicative of DNA unwinding. We expect that in the presence of ATP, when XPD is activated, more strands will be forked.

Concurrently, we would like to observe XPD redistribution, thus we have increased the scan rate in an attempt to visualize the protein association with or dissociating from the DNA strand.

Here, the short strand overhangs remain unblocked and are accessible to XPD. We find several differences between samples that contain ATP and those that do not (Figures 6.6, 6.7, 6.8). Although there are not many forked DNA structures overall, we observe curled structures, where the DNA turns to meet at junctions in the presence of 5 mM ATP. We have considered including other proteins that perform DNA CT to probe whether we observe a shift in redistribution onto mismatched strands as we initiate XPD helicase activity.

Figure 6.6. AFM images of 6 μM DNA and 0.6 μM XPD with (left) or without (right) 5 mM ATP. Scan size: 3 x 3 μm^2, scan rate: 3.05 Hz. Arrows highlight proteins bound to DNA strands. Short DNA strands contain 14 base single-strand overhangs. Although there are more DNA strands that can be visualized in the images on the right, the concentration of the samples is not different. The images show that with ATP, more of the strands are forked or curled, structures that we do not see as much if ATP is not present.

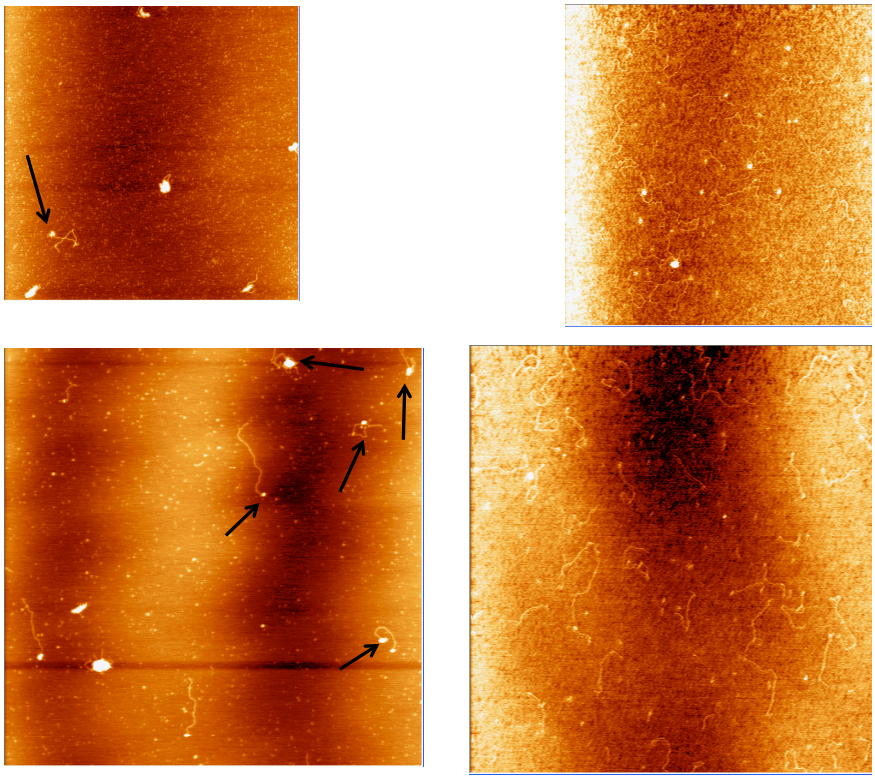


Figure 6.7. AFM images of 6 μM DNA and 0.6 μM XPD with (left) or without (right) 5 mM ATP. Scan size: 3 x 3 μm^2, scan rate: 3.05 Hz. Arrows highlight proteins bound to DNA strands. Short DNA strands contain 14 base single-strand overhangs. Although there are more DNA strands that can be visualized in the images on the right, the concentration of the samples is not different. The images show that with ATP, more of the strands are forked or curled, structures that we do not see as much if ATP is not present.

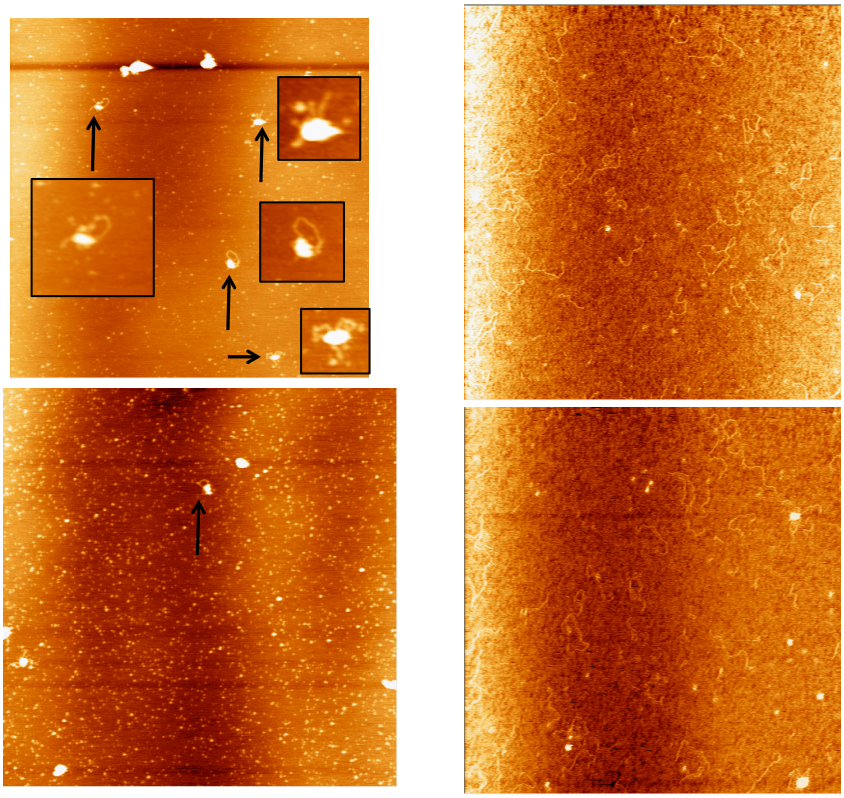
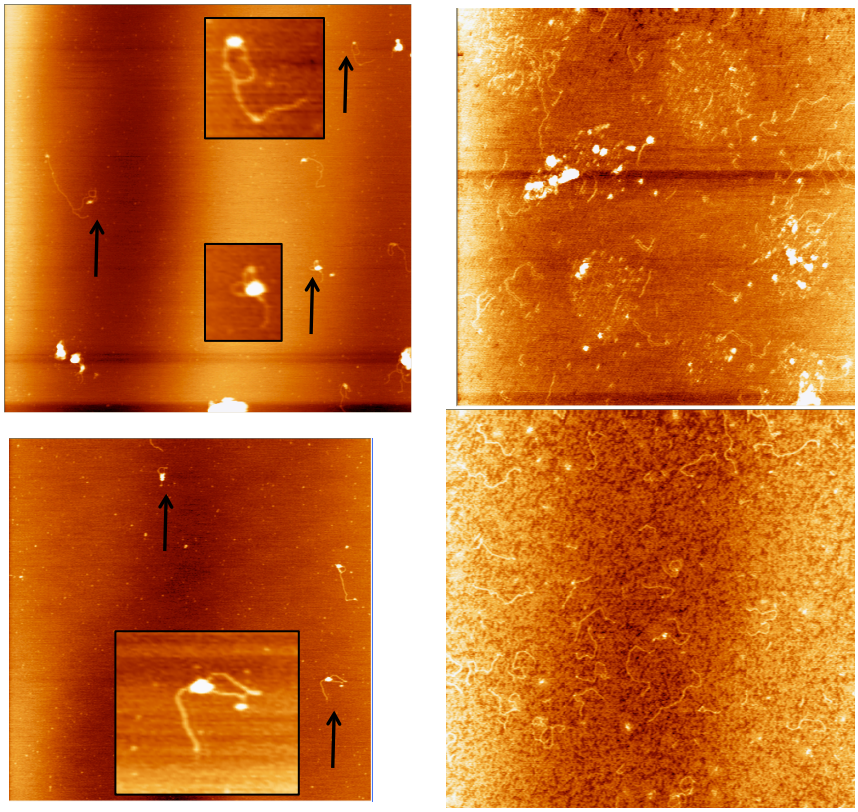


Figure 6.8. AFM images of 6 μM DNA and 0.6 μM XPD with (left) or without (right) 5 mM ATP. Scan size: 3 x 3 μm^2, scan rate: 3.05 Hz. Arrows highlight DNA forks that have protein bound. Short DNA strands contain 14 base single-strand overhangs. Although there are more DNA strands that can be visualized in the images on the right, the concentration of the samples is not different. The images show that with ATP, more of the strands are forked or curled, structures that we do not see as much if ATP is not present.



Summary

In these studies we demonstrate that XPD, in the presence of 2 mM ATP, redistributes in the vicinity of mismatches similar to wild-type protein. Since XPD exhibits helicase and ATPase activity, it is not surprising that incubating the samples for 24 hours prior to imaging results in a preference of XPD for short strands that contain single-strand overhangs. Upon increasing the concentration of ATP to 5 mM, there is a redistribution onto long strands irrespective of a mismatch. Thus, our efforts toward enhancing the efficiency of DNA CT with ATP to improve the ability of XPD to find lesions were not successful. However, significantly, this early work reveals that AFM may be employed to "visualize" XPD helicase activity. When protein and DNA samples are treated with caged-ATP and activated with UV-light, forked structures appear in AFM images, indicative of DNA unwinding. We observe proteins bound at the junction of the forks, exactly where we would expect XPD to bind as it initiates helicase activity. Applying what we have learned, in the most recent experiments, we have left short-strand overhangs accessible to XPD. When 5 mM ATP is added to these samples, we observe curled DNA strands, forks, and other structures with XPD, once again, bound at the junctions. Future studies using AFM to visualize XPD redistribution and helicase activity with/without ATP will help us understand the link between the ability of a protein to perform DNA CT and its propensity to cooperate with other DNA-bound proteins to find and correct lesions. We could then extend our studies to examine XPD mutants relevant to disease.

REFERENCES

1. Kuraoka, I., Bender, C., Romieu, A., Cadet, J., Wood, R. D., and Lindahl, T. (2000) *Proc. Natl. Acad. Sci. USA* 97, 3832.

2. Wolski S. C., Kuper, J., Hänzelmann, P., Truglio, J. J., Croteau, D. L., Van Houten, V., and Kisker, C. (2008) *PLos Biol. 6*, 1332.

3. Fan, L., Fuss, J. O., Cheng, Q. J., Arvai, A. A., Hammel, M., Roberts, V. A., Cooper, P. K., and Tainer, J. A. (2008) *Cell 133*, 789.

4. Liu H., Rudolf, J., Johnson, K. A., McMahon, S. A., Oke, M., Carter, L., McRobbie, A. M., Brown, S. E., Naismith, J. H., and White, M. F. (2008) *Cell 133*, 801.

5. Lukianova, O. A., and David, S. S. (2005) *Curr. Opin. Chem. Biol. 9*, 145.

6. Gorodetsky, A. A., Boal, A. K., and Barton, J. K. (2006) *J. Am. Chem. Soc. 128*, 12082.

7. Boal, A. K., Genereux, J. C., Sontz, P. A., Gralnick, J. A., Newman, D.K., and Barton, J. K. (2009) *Proc. Natl. Acad. Sci. 106,* 15237.

8. Romano, C. A., Sontz, P. A., and Barton, J. K. (2011) *Biochemistry 50*, 6133.

9. Mui, T. P., Fuss, J. O., Ishida, J. P., Tainer, J. A., and Barton, J. K. (2011) *J. Am. Chem. Soc. 133*, 16378.

10. Sontz, P. A., Mui, T. P., Fuss, J. O., Tainer, J. A., and Barton, J. K. (2012) *Proc. Natl. Acad. Sci. USA 109*, 1856.

11. Henn, A., Medalia, O., Shi, S. P., Steinberg, M., Franceschi, F., and Sagi, I. (2001) *Proc. Natl. Acad. Sci. USA 98*, 5007.

12. Marsden, S., Nardelli, M., Linder, P., and McCarthy, J. E. G. (2006) *J. Mol. Biol. 361*, 327.

13. Zhang, W., Dillingham, M. S., Thomas, C. D., Allen, S., Roberts, C. J., and Soultanas, P. (2007) *J. Mol. Biol. 371*, 336.

14. Liu, Y. Y., Wang, P. Y., Dou, S. X., Wang, W. C., Xie, P., Yin, H. W., Lin, Y., Jian, Y. X., Fang, X. H., and Xi, X. G. (2005) *Sci. Technol. Adv. Mat. 6*, 842.

15. Lia, G., Indrieri, M., Owen-Hughes, T., Finzi, L., Podesta, A., Milani, P., and Dunlap, D. (2008) *J. Biophoton. 1*, 280.

16. Yeeles, J. T. P., van Aelst, K., Dillingham, M. S., and Moreno-Herrero, F. (2011) *Mol. Cell. 42*, 806.

17. Cunningham, R. P., Asahara, H., Bank, J. F., Scholes, C. P., Salerno, J. C., Surerus, K., Munck, E., McCracken, J., Peisach, J., and Emptage, M. H. (1989) *Biochemistry 28,* 4450.

18. *Charge Transfer in DNA: From Mechanism to Application* (2005), ed. Wagenknecht H. A. (Wiley- VCH, Weinheim).

19. Fan, L., Fuss, J. O., Cheng, Q. J., Arvai, A. A., Hammel, M., Roberts, V. A., Cooper, P. K., and Tainer, J. A. (2008) *Cell 133*, 789.

20. Rudolf, J. (2006) Thesis submitted for the degree of Doctor of Philosophy, University of St. Andrews, Scotland, UK.

CHAPTER 7

Charge Photoinjection in Intercalated and Covalently Bound $[Re(CO)_3(dppz)(py)]^+$-DNA Constructs Monitored by Time-Resolved Visible and Infrared Spectroscopy

Adapted from Olmon, E. D., Sontz, P. A., Blanco-Rodríguez, A. M., Towrie, M., Clark, I. P., Vlček, A. Jr., and Barton, J. K. (2011) *J. Am. Chem. Soc.* **133**, 13718

E.D.O. and P.A.S. prepared oligonucleotide conjugates and constructs. P.A.S. performed oxidative DNA cleavage experiments. E.D.O. performed spectroelectrochemistry and transient absorption studies, and A.M.B.R., M.T., I.P.C., and A.V. performed TRIR studies.

INTRODUCTION

The ability of DNA to mediate charge transport (CT) has been established using a variety of redox-active probes and in a great diversity of experimental systems (*1-3*). The efficiency of DNA-mediated CT is affected by several factors, including the extent of electronic coupling between the probe and the DNA base stack, coupling within the base stack itself, the driving force of the CT reaction, and the base sequence. DNA CT has been observed over long molecular distances with little attenuation (*4-6*), suggesting its utility in molecular-scale devices (*7-9*) and in biological systems (*2, 10-13)*. Many of the properties of DNA CT have been elucidated in experiments involving the slow accumulation of oxidative damage at low potential guanine sites. While such methods remain useful in the investigation of DNA CT, a general probe for direct, time-resolved monitoring of these processes remains elusive.

Time-resolved infrared (TRIR) spectroscopy offers several advantages over other time-resolved methods for the study of CT events (*14*). With the proper choice of IR-active probe and solvent medium, changes in the absorption pattern of well-resolved, transient IR bands provide kinetic information on specific photophysical, chemical, and biochemical processes, together with structural characterization of the excited states and reaction intermediates involved. One common family of probes are coordination complexes of the type $[Re(CO)_3(N,N)(L)]^n$, where N,N stands for an α-diimine ligand such as 2,2′-bipyridine (bpy), phenanthroline (phen), or dppz (dipyrido[3,2-*a*:2′,3′-*c*]phenazine) and L represents an axial ligand, often Cl (n = 0) or functionalized pyridine (n = 1+) (*15-23*). Photophysical or photochemical reactions involving these Re complexes are manifested in TRIR spectra as changes in the intensities and positions (energies) of absorption bands due to CO stretching vibrations of the $Re(CO)_3$ group, ν(C≡O). Variation of the N,N and L ligands affords fine control over the excited-state characters and energetics (*16, 18-20, 22-27*). These complexes have also proven useful as biochemical probes for fluorescence imaging (*28*), for

monitoring the dynamics of structural fluctuations (*29, 30*), and especially, for triggering photoinduced electron transfer (ET) (*31*). Information on ET kinetics and intermediates provided by TRIR is more direct than that obtained using UV/visible time-resolved spectroscopic methods due to their low specificity. Recently, the presence of tryptophan along the ET pathway in $Re(CO)_3$(4,7-dimethyl-1,10-phenanthroline)-modified azurin was shown to increase the rate of ET (*32-34*). Although other coordination complexes, such as dicarbonyl Ru species, $W(CO)_5$(4-cyanopyridine), and $[Ru(bpy)(CN)_4]^{2-}$ have been employed as TRIR probes, tricarbonyl Re complexes have been studied much more extensively (*14, 16, 35*). TRIR can also be used to monitor changes in the vibrational frequencies and IR band intensities of organic functionalities in ET assemblies (*36*). Of particular interest, TRIR spectra were recorded following the 267 nm excitation of the four canonical nucleotides and of poly(dG-dC)·poly(dG-dC) and poly(dA-dT)·poly(dA-dT) (*37*). In that work, the lifetimes of the transient states of the free nucleotides ranged from 2.2 to 4.7 ps, while those of the polymers were an order of magnitude longer. Upon 200 nm photoionization of 5′-dGMP and poly(dG-dC)·poly(dG-dC), evidence for the formation of the guanine radical was observed by TRIR as the growth of a transient band at 1702 cm^{-1} (*38*). In other experiments, TRIR was used to observe the triplet state of thymine and of 2′-dT (*39*), as well as to unravel the pH-dependent photophysics of 5′-G, 5′-GMP, and poly(G) (*40*). Importantly, these studies indicate that TRIR can be used to monitor photoinduced changes of DNA and of $[Re(CO)_3(N,N)(L)]^n$ simultaneously, making it possible to investigate both the donor and the acceptor sites of Re-DNA CT assemblies. Although interactions between Re complexes and DNA have been studied by UV/visible spectroscopy (*41, 42*), these interactions had not been investigated by vibrational methods until very recently (*43, 44*).

Here, TRIR spectroscopy is used in conjunction with other methods to observe the DNA-mediated oxidation of guanine in DNA by photoexcited $[Re(CO)_3(dppz)(py'\text{-}OR)]^+$, where py′-OR represents pyridine functionalized at the 4 position (Figure 7.1). The influence of guanine on the

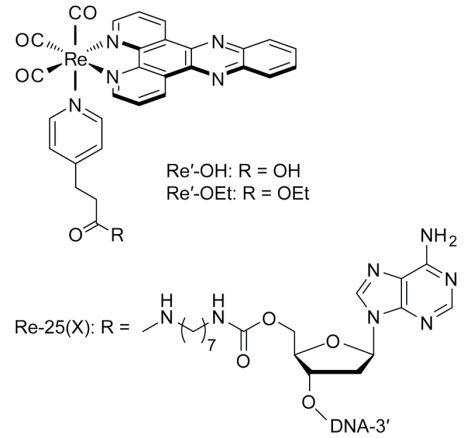
photochemical behavior of the Re complex bound to DNA is investigated by comparing results obtained in four different DNA contexts, including two in which the complex is covalently tethered to specific locations on the duplex. The data presented show that the photoexcited Re complex can oxidize guanine at a distance of several bases away by DNA-mediated CT and that this process can be monitored on the ps to μs time scale by TRIR. The results of this study, in which TRIR is used for the first time to observe DNA-mediated CT between photooxidants and guanine in well-defined covalent constructs, shows that the DNA sequence surrounding the metal complex binding site has a large influence on the photophysics and photochemistry of the system.

Materials and Methods Most reagents for metal complex synthesis and coupling were purchased from Sigma-Aldrich unless otherwise indicated. 3-(pyridin-4-yl)propanoic acid (py′-OH) was purchased from Chess GmbH (Mannheim, Germany). Reagents for DNA synthesis were purchased from Glen Research (Sterling, VA). All reagents were used as received.

Complex and Conjugate Synthesis Preparation of $[Re(CO)_3(dppz)(py'\text{-}OH)](PF_6)$ was adapted from previously described methods (*41*). Following the synthesis, the PF_6^- counter anion was exchanged (QAE Sephadex A-25 resin, GE Healthcare) for chloride ion in order to increase the solubility of the complex in aqueous media. Because facile proton loss from the carboxylic acid-modified pyridine ligand results in an overall neutral zwitterionic species, altering the extent of electrostatic repulsion between complex molecules and of electrostatic attraction to DNA, the protected ethyl ester version of the complex, $[Re(CO)_3(dppz)(py'\text{-}OEt)]^+$ (py′-OEt = ethyl 3-(pyridin-4-yl)propanoate), was used for some experiments.

Oligonucleotide Synthesis and Modification Oligonucleotides were synthesized using standard solid-phase phosphoramidite chemistry on an Applied Biosystems 3400 DNA synthesizer. Tethers were appended to the 5′-OH termini of resin-bound oligonucleotides: Holmlin et al. (*45*).

Figure 7.1. Schematic illustration of $[Re(CO)_3(dppz)(py'\text{-}OR)]^+$, the covalent linker, and the DNA sequences used for studies of guanine oxidation. Experiments involving AT-30 and GC-30 were conducted in the presence of the free complex Re′-OH. In the covalent assemblies Re-25(G) and Re-25(I), the Re photooxidant is tethered to the 5′ end of one strand via a peptide linkage.



AT-30

5′-TTTATATTATTAAATAAATTTTATATATTT-3′
3′-AAATATAATAATTTATTTAAAATATATAAA-5′

GC-30

5′-CCCGCGCCGCCGGGGCGGGCCCCGCGCGCCC-3′
3′-GGGCGCGGCGGCCCGCCCGGGGCGCGCGGG-5′

Re-25(X), X = G, I

The alkyl tether was added to the DNA strand by successive treatment with carbonyldiimidazole and diaminononane. Agitation of the resin-bound, amine modified DNA strands in the presence of excess (5 mg) $[Re(CO)_3(dppz)(py'\text{-}OH)]Cl$, *O*-(benzotriazol-1-yl)-*N,N,N',N'*-tetramethyluronium hexafluorophosphate (HBTU), 1-hydroxybenzotriazole hydrate (HOBT), and diisopropylethylamine (DIEA) in anhydrous DMF for 24 hrs resulted in covalent attachment of the metal complex to the DNA. Cleavage from the resin was effected by incubation in NH_4OH at 60 °C for 6 h. Oligonucleotides were purified by reversed-phase HPLC and characterized by MALDI-TOF mass spectrometry. Oligonucleotide concentrations were determined by UV/visible spectrophotometry (Beckman DU 7400). Annealing was accomplished by incubating solutions containing equimolar amounts of complementary strands in buffer (10 mM NaP_i, 50 mM NaCl buffer; pH 7.0) at 90 °C for 5 min followed by slow cooling over 90 min to ambient temperature.

Assay for Oxidative DNA Damage Oxidative DNA cleavage experiments were performed using a protocol adapted from Zeglis (*46*) with the following adjustments. Oligonucleotides were labeled at the 3′-end by incubating a mixture of 2 μL single-stranded DNA (100 μM), 5 μL [α-^{32}P]-dTTP (Perkin Elmer), 2 μL terminal transferase (TdT; New England Biolabs), 5 μL $CoCl_2$ solution (included with TdT), and 5 μL terminal transferase reaction buffer (included with TdT) for 2 h at 37 °C. Before gel purification, strands were incubated at 90 °C for 20 min in 100 μL 10% aqueous piperidine to induce cleavage of damaged strands. Following purification and annealing, samples (10 μL, 2 μM) were irradiated in parallel for 2 hrs using a solar simulator (Oriel Instruments) fitted with a 340 nm internal long pass filter. Samples were then treated with 0.2 units calf thymus DNA and 10% piperidine (v/v), heated for 30 min at 90 °C, and dried *in vacuo*. After gel electrophoresis, oxidative damage was quantified by phosphorimagery (ImageQuant). Sample counts are reported as % of total counts per lane and were corrected by subtracting the dark control.

Spectroelectrochemistry UV/visible spectroelectrochemistry was carried out using a custom-built, optically transparent, thin-layer electrode (OTTLE) cell (path length = 0.1 mm) consisting of vapor-deposited platinum working and pseudoreference electrodes and a Pt-wire auxiliary electrode (*47*). The potential of the cell was controlled by a potentiostat (CH Instruments Model 650A electrochemical workstation). Samples consisted of saturated solutions of metal complexes in dry acetonitrile with 0.1 M Bu_4NPF_6 electrolyte. Samples were degassed by bubbling argon and introduced into the optical cell using a gas-tight syringe prior to measurement. The cell was held at a reducing potential, and spectra were acquired on a Hewlett Packard 8452A diode array spectrophotometer every 4 s until the sample was fully reduced.

UV/Visible Emission and Transient Absorption Spectroscopy Steady-state emission spectra were recorded on a Fluorolog-3 spectrofluorometer (Jobin Yvon) using 2 mm slits. Scattered excitation light was rejected from the detector by appropriate filters. Reported spectra are averages of at least five consecutive measurements.

All time-resolved UV/visible spectroscopic measurements were carried out at the Beckman Institute Laser Resource Center. Nanosecond luminescence decay measurements and transient absorption (TA) measurements were performed using the third harmonic (355 nm) of a 10 Hz, Q-switched Nd:YAG laser (Spectra-Physics Quanta-Ray PRO-Series) as the excitation source (8 ns pulse width, 5 mJ/pulse). Probe light was provided by a synchronized, pulsed 75 W Hg-Xe arc lamp (PTI model A 1010), and detection was accomplished using a photomultiplier tube (Hamamatsu R928) following wavelength selection by a double monochromator (Instruments SA DH-10). Scattered light was rejected using suitable filters. The samples were held in 1 cm path length quartz cuvettes (Starna) equipped with stir bars. TA measurements were made with and without excitation, and were corrected for background light, scattering, and fluorescence.

Picosecond emission decay measurements (*48*-51) were performed using the third harmonic of a regeneratively amplified mode-locked Nd:YAG laser (355 nm, 1 ps pulse width after amplification) as the excitation source and a picosecond streak camera (Hamamatsu C5680, photon-counting mode) as the detector. Emission was observed under magic angle conditions using a 550 nm long-pass cutoff filter.

TRIR Spectroscopy The ULTRA instrument at the STFC Rutherford Appleton Laboratory has been used. It is described in detail elsewhere (*52*). In brief, a titanium sapphire laser-based regenerative amplifier (Thales) produces 800 nm, ~ 50 fs pulses at a 10 kHz repetition rate. The laser output is split in two parts, one of which is either frequency doubled or used to drive an OPA (Light Conversion, TOPAS) equipped with SHG and SFG units to produce a pump beam at 400 or 355 nm, respectively. The second pumps a TOPAS OPA, yielding signal and idler beams that are difference frequency mixed to generate ~ 400 cm^{-1} broad mid IR probe pulses. An optical delay line is used to introduce a delay between the pump and probe beams, and the mid IR probe spectrum is recorded at a given time delay using two 128 element HgCdTe detectors (Infrared Associates). For ns-μs measurements, the sample was pumped with 355 nm, 0.7 ns FWHM pulses (AOT, AOT-YVO-20QSP/MOPO), and probed with electronically synchronized 50 fs IR pulses (*53*). The sample solutions were placed in a round dip *ca*. 0.75 mm deep, drilled into a CaF_2 plate and tightly covered with a polished CaF_2 window. The cell was scanned-rastered across the area of the dip in two dimensions to prevent laser heating and decomposition of the sample. FTIR spectra measured before and after the experiment demonstrated sample stability.

Fitting Methods TRIR data were simulated at each time delay as a series of Gaussian terms in order to extract kinetic data from overlapping transient bands. The area of each Gaussian was

calculated, and kinetic decays were constructed as the change in area with delay time. Nanosecond time-resolved emission, TRIR, and TA data were fit by nonlinear least-squares analysis using IGOR Pro software (Wavemetrics). Model functions consisted of a linear series of exponential terms of the form $y(t) = \Sigma\, a_i \exp[-t/\tau_i]$, where a_i and τ_i are the pre-exponential factor and lifetime, respectively, of the *i*th term. Up to three exponential terms were included until reasonable fits were obtained. For time-resolved emission data, the percent relative contribution reported in Table 7.1 represents the number of photons emitted at the probe wavelength by each emissive population, and is calculated as $a_n\tau_n/\Sigma a_i\tau_i$ (the area under the decay for the *n*th exponential term normalized to the total area under the decay curve). For TRIR and TA data, the percent relative contribution represents the change in absorbance of species *n* extrapolated to time $t = 0$, and is calculated as $a_n/\Sigma a_i$.

Picosecond emission data were collected at 1 ns, 5 ns, and 50 ns time ranges and spliced together before fitting. Data were compressed logarithmically in time prior to fitting in order to decrease the bias of long time data on the fit. These data could not be fit well to a series of exponential terms and were instead analyzed by the maximum entropy method using a MATLAB (MathWorks) routine written at Caltech (*48-51*).

Results

Research Strategy and Design of Re-DNA CT Assemblies With the aim to establish DNA oxidation by electronically excited rhenium tricarbonyl-diimine complexes, we have employed a newly developed Re sensitizer, $[Re(CO)_3(dppz)(py'\text{-}OR)]^+$ (R = H, Re′-OH; or R = CH_2CH_3, Re′-OEt), which can be covalently linked to DNA (Figure 7.1). Three design elements make this a promising probe for the study of DNA-mediated CT.

Table 7.1. Decay Lifetimes from Least-Squares Fits (λ_{ex} = 355 nm)

Technique	Detail	Sample	Lifetimes, seconds (% Relative Contribution[a])					
			10^{-9}	10^{-8}	10^{-7}	10^{-6}	10^{-5}	$>10^{-5}$ [b]
TRIR[c]	$Re(CO)_3$ Bleach Recovery	AT-30 + Re'-OH		1.4 (28)	3.0 (31)		2.5 (41)	
		GC-30 + Re'-OH		6.3 (43)	5.5 (49)		1.1 (8)	
		Re-25(I)		2.6 (16)	4.8 (30)		2.8 (54)	
		Re-25(G)	4.8[d]		2.9 (65)	9.1 (35)		
	MLCT (2071 cm^{-1})	AT-30 + Re'-OH	8.8 (38)		5.6 (38)			Long (23)
		Re-25(I)		3.2 (36)	8.9 (44)			Long (20)
	IL (2030 cm^{-1})	AT-30 + Re'-OH			1.5 (31)		2.4 (69)	
		GC-30 + Re'-OH			1.8 (51)	1.0 (42)	2.2 (7)	
		Re-25(I)				3.0 (34)		Long (66)
		Re-25(G)	4.0[d]		3.0 (58)	9.2 (42)		
	$G^{\bullet+}/G^{\bullet}$ (1702 cm^{-1})	GC-30 + Re'-OH			2.1[d]	5.8		
	DNA Bleach Recovery	AT-30 + Re'-OH			1.4 (48)	5.0 (18)		Long (35)
		GC-30 + Re'-OH		5.9 (31)	9.6 (69)			
		Re-25(I)			3.1 (88)			Long (12)
		Re-25(G)		2.9[d]	8.4 (53)	1.0 (47)		
NS Visible TA[e]	λ_{probe} = 475 nm	AT-30 + Re'-OH			4.9 (17)		2.7 (83)	
		GC-30 + Re'-OH			2.7 (42)	2.0 (58)		
		Re-25(I)			9.6 (42)		2.0 (58)	
		Re-25(G)			4.4 (37)		1.4 (63)	
NS Emission[e,f]	λ_{probe} = 570 nm	AT-30 + Re'-OH	2.9 (34)	2.4 (23)	5.7 (43)			
		GC-30 + Re'-OH	3.2 (39)	2.7 (35)	2.4 (26)			
		Re-25(I)	5.3 (17)	2.6 (21)	5.4 (62)			
		Re-25(G)	3.9 (42)	2.1 (30)	4.5 (28)			

[a] Determined by different methods for absorption and emission; see Experimental Section. [b] "Long" indicates incomplete decay. [c] Uncertainty estimated at 20%. [d] These values reflect an increase in intensity. [e] Uncertainty estimated at 10%. [f] Processes faster than 8 ns are convoluted with instrument response.

The first is the incorporation of TRIR-active carbonyl ligands. Re carbonyl-diimine complexes are useful probes in TRIR spectroscopic experiments due to the intense and well-resolved bands corresponding to carbonyl stretching modes. These modes are extremely sensitive to changes in electron density distribution, molecular structure and environment (*17, 24, 30, 32, 54, 55)*. The second design element is the inclusion of the planar dppz ligand. By incorporating dppz on the metal center, we ensure effective electronic coupling with the DNA base stack. Indeed, the binding constants for intercalating dppz complexes such as $[Ru(bpy)_2(dppz)]^{2+}$ and $[Ru(phen)_2(dppz)]^{2+}$ are greater than 10^6 M^{-1} (*56*). While the binding of complexes like $[Re(CO)_3(dppz)(py'\text{-}OR)]^+$ is weaker (10^5 M^{-1}) (*41, 42,* 57) due to its lower electrostatic charge, the decrease of the molar absorptivity of its near-UV absorption band (i.e., hypochromicity) upon incubation with DNA, as well as an increase in the melting temperature of the bound DNA duplex by approximately 5 °C (depending on the sequence), indicate that this Re complex indeed binds by intercalation. The third design element is the ability to covalently attach the complex to DNA via carboxyalkyl-modified pyridine incorporated at the axial coordination site. The covalent link between the complex and the DNA strand, while flexible, restricts diffusion of the unbound complex, ensuring a higher percentage bound than if the complex were allowed to diffuse freely. In addition, the covalent link enables us to define the DNA sequence at the binding region, eliminating sequence effects as a variable. Physical models suggest that in the equilibrium geometry (disregarding frayed end effects) tethering restricts binding to the region within three base pairs from the end of the duplex.

The DNA duplexes used were designed to test for the effect of the DNA sequence on the efficiency of DNA oxidation. For systems in which guanine, an effective hole trap, is placed near the expected binding site of the Re complex, charge injection may be followed by facile back electron transfer (BET). Such nonproductive reactions are competitive with permanent charge trapping at guanine sites (*58-60*). The frequency of nonproductive events can be reduced by replacing guanine at the Re binding site with inosine (I), a base analog that has a higher oxidation

potential than guanine [$E°(I^{\bullet+}/I)$ ≈1.5 V vs. NHE; $E°(G^{\bullet+}/G)$ = 1.29 V vs. NHE] (*60-64*). With these considerations in mind, four DNA sequences were designed (Figure 7.1). Two of them contain only adenine and thymine (AT-30) or guanine and cytosine (GC-30) and are expected to reveal the effect of the absence or presence, respectively, of strong guanine thermodynamic hole traps on DNA oxidation by noncovalently bound Re′-OH. Two DNA sequences were also designed to test for the effect of neighboring guanine on the efficiency of long range DNA oxidation by covalently-bound Re. These are Re-25(G), which contains guanine next to the Re binding site, and Re-25(I), in which guanine is replaced by inosine.

Sensitizer Characterization The photophysics of $[Re(CO)_3(dppz)(py'\text{-}OH)]^+$ and $[Re(CO)_3(dppz)(py'\text{-}OEt)]^+$ are very similar, suggesting that modification at the py′ carbonyl has little effect on the energetics of the complex. For example, each complex exhibits absorption maxima at 364 and 382 nm ($\varepsilon \approx 11{,}000\ M^{-1}\ cm^{-1}$) (*41, 65*), with a tail that extends into the visible region (*66*). The emission spectra of both complexes show maxima at 554 and 595 nm. At 570 nm, Re′-OH and Re′-OEt each show a biexponential emission decay in acetonitrile, with lifetimes on the order of 200 ns (~ 10%) and 10 μs (~ 90%), tentatively attributed to emission from different types of ^{3}IL states (*66*). Tethering the Re species to DNA, therefore, is expected to have negligible influence on the energetics of the complex.

The reduction potential of the emissive ^{3}IL state(s), $E°(*Re^+/Re^0)$, of the Re label can be estimated as the sum of the ground-state reduction potential, $E°(Re^+/Re^0)$, and the zero-zero excited-state energy, ($E_{0,0}$) (*67*).The exact value of $E_{0,0}$ is unknown, but it is estimated to lie between the energy at which the excitation and emission spectra coincide (480 nm, 2.58 eV) and the energy of the emission maximum in aqueous solution (570 nm, 2.18 eV). For Re′-OEt in acetonitrile, $E°(Re^+/Re^0)$ was reported as –850 mV vs. NHE (*66*), predicting the excited-state reduction potential to lie

between 1.33 and 1.73 eV. As an oxidant, electronically excited Re′-OEt is clearly strong enough to oxidize guanine, and it may be strong enough to oxidize adenine [$E°(A^{\bullet+}/A)$ = 1.42 V vs. NHE] (*68*). The latter reaction, however, is expected to be slower due to the lower driving force.

Hole injection into the DNA base stack must coincide with reduction of the metal complex. In order to characterize this reduced state independently, IR spectroelectrochemical reduction of saturated Re′-OEt in acetonitrile was carried out (Figure 7.2). Before reduction, the spectrum exhibits a band at 2036 cm^{-1} assigned to the totally symmetric in-phase ν(C≡O) vibration A′(1), and a band at 1932 cm^{-1} due to quasidegenerate totally symmetric out-of-phase A′(2) and equatorial antisymmetric A″ ν(C≡O) vibrations (*17, 54, 69)*. Reduction results in a bathochromic shift of these bands to 2029 cm^{-1} and 1922 cm^{-1}, respectively. This shift is similar to that observed previously (*23*) upon reduction of [$ReCl(CO)_3(dppz)$] and its small magnitude is consistent with occupation of the phenazine π* orbital of the dppz ligand in the [$Re^{I}(CO)_3(dppz^{\bullet-})$(py′-OEt)] reduction product (*24*). Subsequent regeneration of the initial species via reoxidation was 95% complete, suggesting partial irreversible decomposition of the electrogenerated product; however, these decomposition products are not expected to interfere in time-resolved spectroscopic experiments employing fast photocycles. An attempt was made to duplicate the experiment in D_2O buffer (10 mM NaP_i, 50 mM NaCl, pD 7.0) in order to generate spectra that would be more directly comparable to TRIR measurements conducted in D_2O buffer. Although the low solubility of the complex and the strong background absorbance of the solvent in this energy region prevented precise analysis, band positions, widths, and relative intensities were similar to those observed in acetonitrile solutions.

Oxidative Damage Pattern of Re-25(G) and Re-25(I) Observed by PAGE Figures 7.3 and 7.4 show DNA-mediated oxidative damage in 2 μM solutions of Re-25(G) and Re-25(I) observed after 2 hours of broadband (λ > 340 nm) irradiation and 20% PAGE analysis.

Figure 7.2. Steady-state FTIR spectra (bottom) of saturated Re′-OEt in acetonitrile recorded during bulk reduction using an OTTLE cell. Arrows indicate spectral changes that occur upon reduction. The difference in absorbance between the fully reduced species and the initial species is also shown (top).

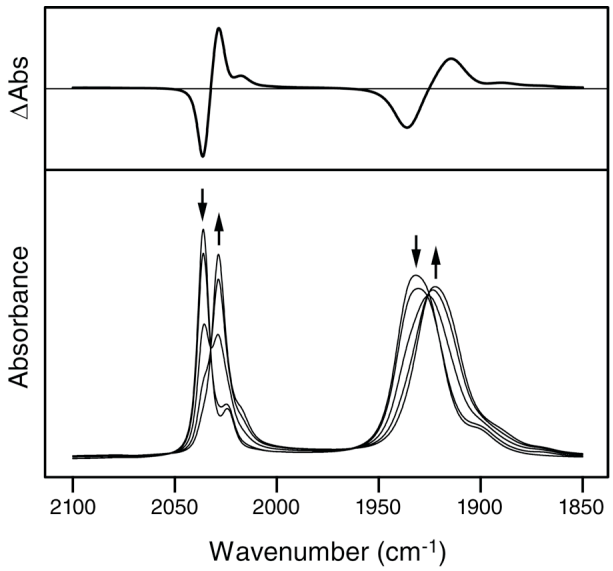


Figure 7.3. (Left) Oxidative damage with $Re(CO)_3(dppz)(py)^+$ photooxidant in Re-25(I) assemblies. 20% PAGE after irradiation of functionalized 3′-dTTP-^{32}P-labeled DNA assemblies (2 μM). A+G and C+T are Maxam-Gilbert sequencing lanes. Following 2 hour irradiation with a solar simulator, samples were treated with 10% piperidine (v/v), heated for 30 min at 90 °C, and dried. Lane 3 shows background damage in the absence of photooxidant and irradiation. LC indicates the light control, which is nonfunctionalized assembly with irradiation. DC indicates the dark control, which is the functionalized assembly without irradiation but with subsequent piperidine treatment. Re25IA refers to the assembly with inosine near the complex binding site. DNA-mediated oxidative damage at the double guanine site and other guanine positions is indicated with an arrow. (Right) Oxidative damage with $Re(CO)_3(dppz)(py)^+$ photooxidant 20% PAGE gel analysis

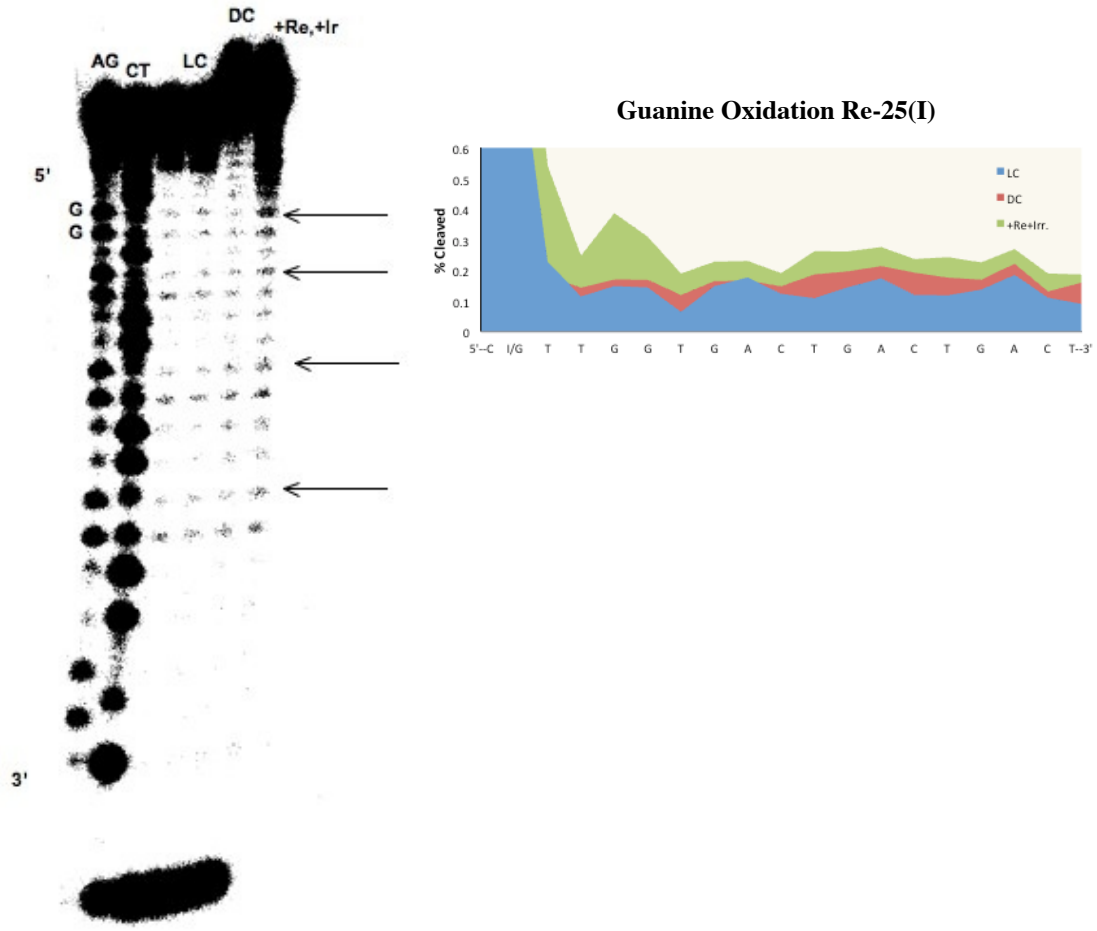
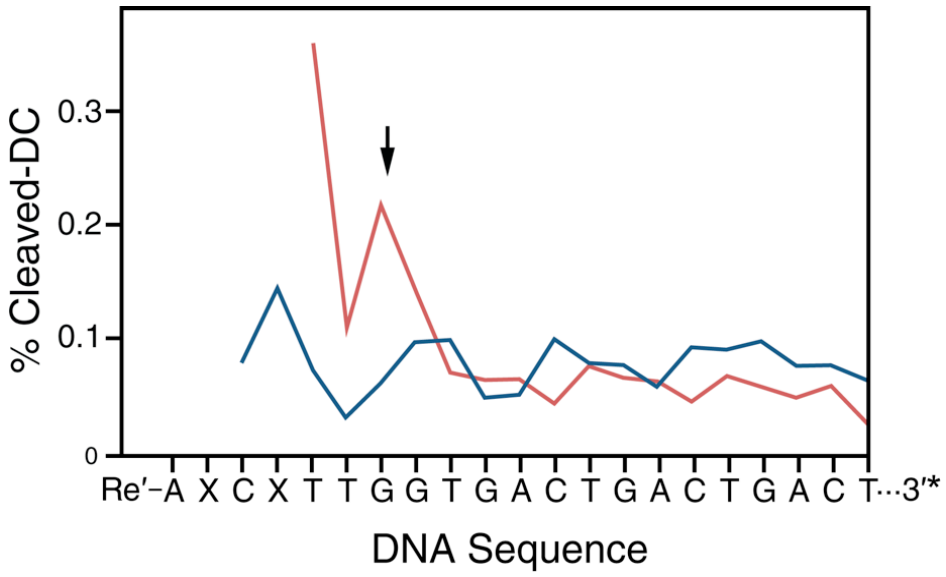


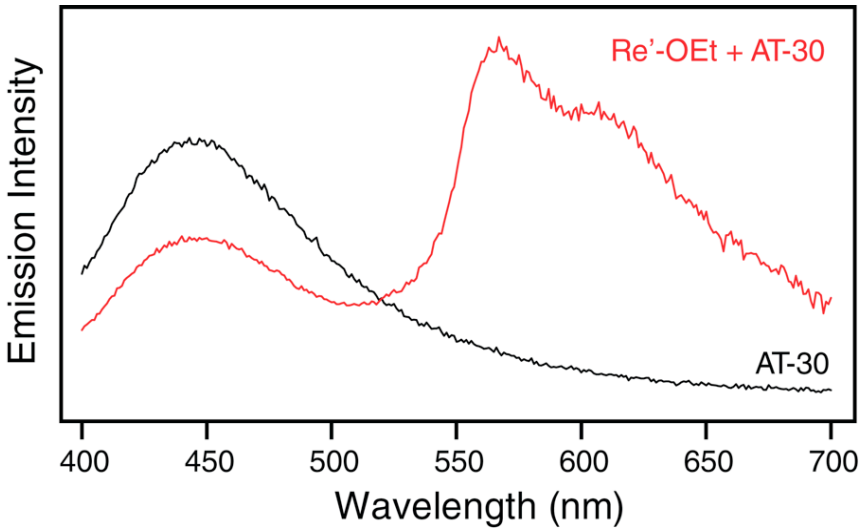
Figure 7.4. Quantification of oxidative damage observed for Re-25(I) (X = I; red) or Re-25(G) (X = G; blue) by PAGE analysis. Aqueous samples containing 3′-[α-^{32}P]-radiolabelled (indicated by *) Re-DNA constructs (2 μM) were irradiated for 2 h and treated with piperidine to induce cleavage at damaged bases. Cleavage products were separated by 20% PAGE and imaged by phosphorimagery. Quantitation was accomplished by normalizing counts at each site to total counts per lane. Traces were corrected for false positives by subtracting the dark control (DC). The arrow indicates the 5′-guanine of a 5′-GG-3′ doublet. Re is expected to bind 2–3 bases in from the 5′-end of the duplex.



Damage occurs as base radicals, formed following hole injection by the excited Re complex, react with solution species such as H_2O or O_2 to form irreversible products (*70, 71*). Subsequent treatment of the 3′-[^{32}P]-labeled DNA with piperidine induces cleavage at damage sites. For both Re-25(G) and Re-25(I), damage is observed primarily at the 5′-G site of the 5′-GG-3′ doublet, several bases distant from the Re complex binding site predicted from physical models. Importantly, the low concentrations used in these experiments preclude interstrand damage (i.e., it is unlikely that the Re moiety of one construct will intercalate into the base stack of another). The observation of damage at the 5′-GG-3′ site indicates that long-range photoinduced hole injection from the Re label to DNA indeed occurs, consistent with results obtained for a similar Re-DNA conjugate (*66*). However, the extent of damage is consistently greater in the case of Re-25(I) than Re-25(G).

Emission Measurements Many Re tricarbonyl complexes of dppz behave as DNA light switches (*18-23, 41, 42*), much like their Ru counterparts (*72*), and the complexes studied here are no exception. In the absence of DNA, negligible emission is observed from an aqueous solution of Re′-OH or Re′-OEt; however, in the presence of AT-30 and in the Re-25(I) sample, a prominent emission band is observed that exhibits a maximum at 570 nm and a shoulder near 600 nm, resembling the emission spectrum seen for similar Re complexes in organic solvents (*19, 21, 22, 41, 42, 57, 65*). By comparison with the emission of $[Ru(bpy)_3]^{2+}$ in deaerated acetonitrile, the emission quantum yield of Re′-OEt in the presence of AT-30 in buffer is 0.008 (as compared to 0.062 for $[Ru(bpy)_3]^{2+}$ (*73*). In the presence of GC-30 and in the Re-25(G) sample, the emission is much less intense, the maxima are shifted to 585 nm, and no shoulder is observed. Steady-state emission spectra of AT-30 alone and in the presence of Re′-OEt are shown in Figure 7.5. Interestingly, the DNA oligomers used in this study are themselves emissive under 355 nm excitation, giving rise to a broad band near 450 nm that tails into the visible region.

Figure 7.5. Steady-state emission spectra of 20 μM Re′-OEt in the presence of 0.5 mM (base pairs) AT-30 (red), and of 0.5 mM (base pairs) AT-30 (black) alone following excitation at 355 nm. Samples were prepared in D_2O buffer (10 mM NaP_i, 50 mM NaCl; pD 7.0).



All efforts were made to ensure that this is not an effect of the instrument, solvent, scattering, or impurities. Such emission, ascribed to excitons or charge transfer excited states, has previously been observed in DNA oligomers but not in calf thymus DNA (*74-76*). The Re-loaded AT-30 sample shows overlapping DNA and Re′-OEt emission. By scaling and subtracting the emission band due to DNA alone, it is possible to isolate emission from only the intercalated complexes. Significantly, emission from Re′-OEt becomes strongly quenched on going from AT-30 to GC-30 (Figure 7.6). A similar decrease is observed for Re-25(G) compared to Re-25(I). The concentrations of DNA and of the Re complex are the same in all of the samples, but the intensity of emission decreases as AT-30 ≈ Re-25(I) > Re-25(G) ≈ GC-30.

Differences in emission intensity are also observed in time-resolved measurements carried out on the nanosecond timescale with a PMT detector (response time 8 ns) and on the picosecond timescale using a streak camera (response time 55 ps). On the nanosecond timescale, the time-integrated emission intensity of Re-25(G) at 570 nm is 14% that of Re-25(I), and the intensity of GC-30 is 12% that of AT-30, following the trend observed in stationary spectra. Even on the picosecond timescale, the instantaneous emission intensity extrapolated to t=0 is lower in the GC-30 and Re-25(G) samples than in the AT-30 and Re-25(I) samples, respectively. In addition, on this time scale, the time-integrated emission intensity of Re-25(G) is 79% that of Re-25(I), and the intensity of GC-30 is 69% that of AT-30. These observations clearly indicate reaction(s) between electronically excited Re complex and DNA occurring on the picosecond to nanosecond timescale. On the basis of results of the PAGE experiment, hole transfer from *Re to G is most likely a prominent contributing reaction pathway.

The emission decay of the four DNA samples is highly multiexponential, with lifetimes varying over four orders of magnitude, from ~ 100 ps to ~ 500 ns. The present data do not allow us to attribute individual emission decay components to particular species present in the solution, although steady-state results suggest that DNA excimer emission contributes significantly (~ 20 %)

to the total decay. After accounting for DNA excimer emission, which decays with a lifetime of only a few ns (*75*), about half of the Re emission decays within 50 ns, and the remainder persists for hundreds of ns. Maximum entropy fitting of the emission decays yields several distributions of rate constants (Figure 7.7) that vary only slightly between samples, and in every sample, the majority component has a lifetime of less than 1 ns. Notably, while most of the lifetimes are shortened slightly on going from AT-30 to GC-30 and from Re-25(I) to Re-25(G), no decay component is observed that corresponds to quenching of the excited Re sensitizer by guanine. Considering the significant quenching in steady-state measurements of the GC-30 and Re-25(G) samples, it seems that quenching at the reactive binding site(s) is ultrafast, probably tens of picoseconds or faster, but involves only a fraction of the excited population.

Time-Resolved Infrared (TRIR) Spectra Whereas emission spectra provide evidence for ultrafast hole injection from electronically excited Re into the GC-30 and Re-25(G) samples, TRIR has the potential to characterize the reacting state(s) of the Re complex and to detect products and intermediates. To this effect, TRIR spectra were investigated in the picosecond (1–100 ps) and nanosecond-to-microsecond time domains in the regions of the $Re(CO)_3$ ν(C≡O) and DNA organic carbonyl vibrations.

Typical picosecond TRIR spectra obtained in the ν(C≡O) region after 355 nm excitation are shown in Figure 7.8 for AT-30 and GC-30. The spectra measured 1 ps after excitation show negative bands due to bleaching of the ground-state absorption (2036 and 1939 cm^{-1}) and broad transient bands at 2026 and 1908 cm^{-1}. Over the course of time, both features decay in intensity while a sharp band grows in at 2031 cm^{-1} (overlapping with the 2036 cm^{-1} bleach) together with a broad band between 1915 and 1935 cm^{-1}.

Figure 7.6. Steady-state emission spectra of 20 μM Re′-OEt and 0.5 mM DNA (base pairs) in D_2O buffer (10 mM NaP_i, 50 mM NaCl; pD 7.0) solution following excitation at 355 nm. Emission spectra of Re′-OEt with AT-30 (red) or GC-30 (black) have been corrected for emission from DNA alone.

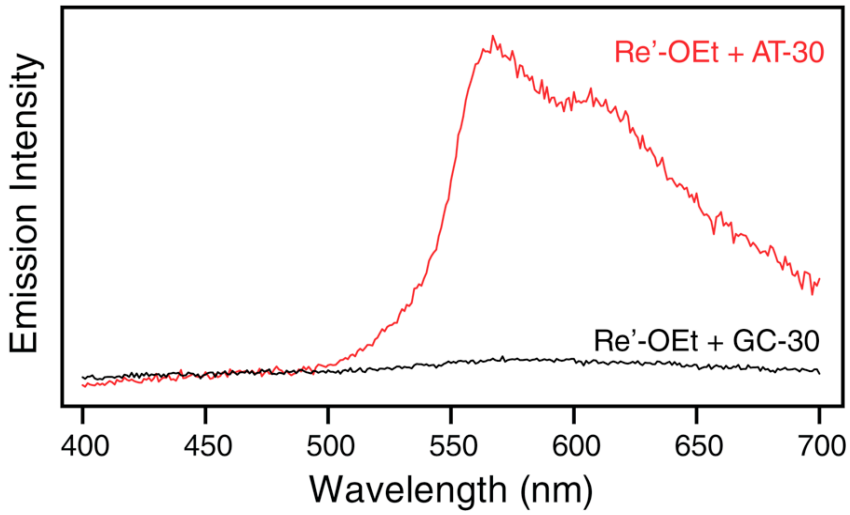


Figure 7.7. Lifetime distributions from maximum entropy analysis of emission from (64 μM) Re′-OH in the presence of 1.6 mM (base pairs) DNA and of 64 μM Re-25(I) or Re-25(G) measured on the picosecond timescale (λ_{ex} = 355 nm, 1 ps pulse width). Samples were prepared in D_2O buffer (10 mM NaP_i, 50 mM NaCl; pD 7.0) and were irradiated at 355 nm. Probability P is plotted as a function of rate k. Large distributions at k = 10^{11}-10^{12} s^{-1} are caused by convolution of the measurement signal with instrumental noise. The emission decay from Re′-OH in buffer is expected to be monoexponential; the complex distribution of rates observed here may be due to the formation of aggregates (solubility is quite low) or it may simply be an effect of the low emission intensity observed for this sample.

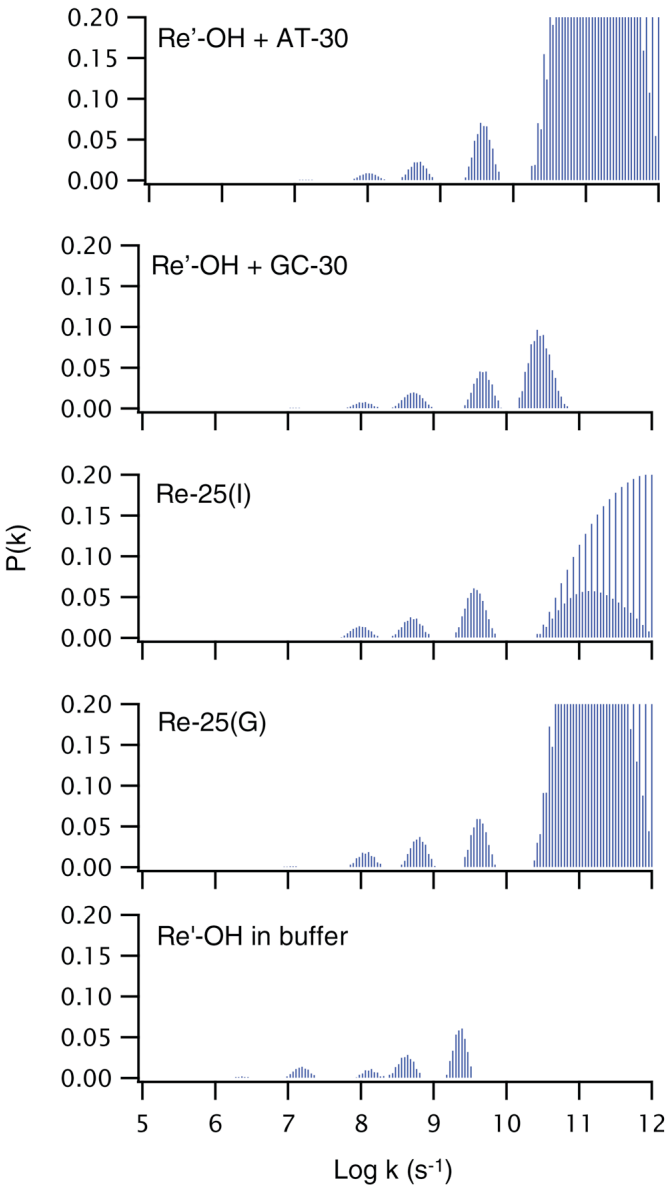
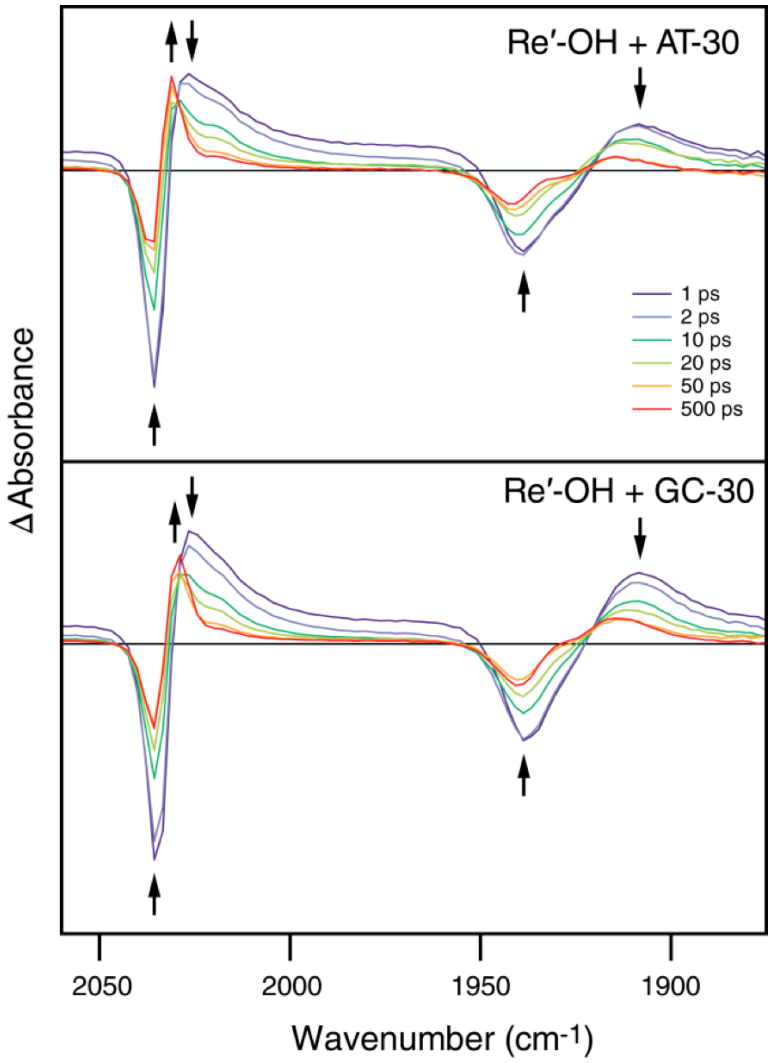


Figure 7.8. Picosecond-timescale TRIR difference spectra of 4.8 mM (base pairs) AT-30 (top) or GC-30 (bottom) with 0.5 mM Re′-OH measured at specified time delays after 355 nm, 50 fs excitation. Each probe data point is separated by ca. 2.1 cm^{-1}. Arrows indicate changes in the spectra with time. Delay times displayed are a subset of the data collected. (Similar TRIR spectra of Re-25(I) and Re-25(G) are shown in Figure 7.10.)



These new transients partially overlap with the parent bleaches at 2036 and 1939 cm^{-1}; hence, the growth of the transients is accompanied by a decrease in the intensities of both bleaches and a distortion of the band shape of the 1939 cm^{-1} bleach. The down-shift in the energies of the transient bands from the ground-state positions is typical of ππ* ^{3}IL(dppz) excited states (*19-32, 26, 27, 54*). Tentatively, we attribute the initially formed 2026 and 1908 cm^{-1} transient bands to a hot ^{3}IL state localized at the phen part of the dppz ligand, ^{3}IL(phen). Subsequent electron density reorganization and cooling produce another ^{3}IL state localized predominantly at the phenazine part, ^{3}IL(phz), manifested as the sharp 2036 cm^{-1} band. The ^{3}IL(phz) IR spectrum is more similar to that of the ground states than the ^{3}IL(phen) spectrum since the electronic changes in ^{3}IL(phz) occur further away from the Re center. The excited-state conversion is largely completed in the first 100 ps. The spectra measured at 100 and 500 ps also show a shoulder at ~ 2020 cm^{-1} that probably corresponds to a residual population of the IL(phen) state. The ps spectra do not show any bands attributable to Re→dppz MLCT states, which are expected to occur at higher energies.

The GC-30 sample shows very similar behavior (Figure 7.8, bottom); however, there is one important difference: the ~ 2031 cm^{-1} ^{3}IL(phz) feature at longer time delays (> 50 ps) is much weaker relative to the initially formed transient than in the case of AT-30. In accordance with the ultrafast GC-30 emission intensity quenching, we attribute this deficiency to a partial picosecond quenching of the ^{3}IL state(s) by CT with guanine to produce $[Re^{I}(CO)_3(dppz^{\bullet -})(py'\text{-}OH)]$ and $G^{\bullet +}$. The lack of IR features in the TRIR spectra due to the reduced Re complex is likely caused by two factors. The first is very close similarity with the spectrum of the ^{3}IL(phz) state (compare with Figure 7.2); the second is very fast BET that regenerates the ground state and keeps the concentration of the reduced state low. The persistence of the 2031 cm^{-1} band of GC-30 into the nanosecond-to-microsecond domain (see below) demonstrates that the relaxed ^{3}IL(phz) state of Re′-OH shows little reactivity, if any. This spectral feature could also correspond to a population of Re

complexes that are protected from solvent quenching by DNA binding but are not well coupled to the base stack.

The picosecond TRIR spectrum of AT-30 in the DNA region is very similar to that measured in the nanosecond time domain (Figure 7.9). The spectra show instantaneous formation of bleach bands at ca. 1618 (weak), 1635, 1660, and 1690 (weak) cm^{-1} that are not accompanied by the formation of transients. These bleaches originate from a decrease in the intensity of the nucleobase carbonyl IR bands upon excitation, rather than band shifts, and they compare well with bleaches observed upon direct 267 nm photoexcitation of nucleic acid polymers (*37*). The GC-30 sample shows strong bleaches at about 1577, 1619 (weak), 1648 and 1679 cm^{-1}, again without the formation of transients. Notably, on the picosecond timescale there is no evidence of a transient due to oxidized $G^{\bullet+}$ or $G^{\bullet}$, which would be expected at ~ 1700 cm^{-1} (*38, 43, 44*). The absence of such a transient is consistent with the ultrafast BET proposed above.

Picosecond TRIR spectra (Figure 7.10) of the Re-25(I) and Re-25(G) samples in both the $Re(CO)_3$ ν(C≡O) and the DNA carbonyl regions closely resemble those of the AT-30 and GC-30 samples, respectively. Importantly, the ^{3}IL(phz) band intensity at 100−500 ps is much lower for Re-25(G) than Re-25(I) relative to the initial transient, again indicating ultrafast *Re→G CT. Absence of any $[Re^{I}(CO)_3(dppz^{\bullet-})(py'\text{-}OH)]$ or $G^{\bullet+}/G^{\bullet}$ IR features suggests ultrafast BET, as in the case of GC-30.

TRIR spectra recorded between 1 ns and 10 μs after photoexcitation are shown in Figure 7.9. The spectral patterns are very similar to those obtained in picosecond experiments at 100 ps and longer: The IL(phz) bands, as well as the bleaches in the DNA region, appear prominently in all four samples. Despite these similarities, closer examination reveals several important spectral differences. The AT-30 and Re-25(I) samples both show a weak isolated positive band at 2070 cm^{-1} and a broad, positive absorbance near 1980 cm^{-1}.

Figure 7.9. Nanosecond-timescale TRIR difference spectra showing changes in the IR absorbance of systems containing 0.5 mM $[Re(CO)_3(dppz)(py'\text{-}OR)]^+$ and 4.8 mM DNA (base pairs) following 355 nm excitation. Both the $Re(CO)_3$ ν(C≡O) (1860–2150 cm^{-1}) and the DNA C=O stretching (1550–1850 cm^{-1}) regions are shown. Arrows indicate changes in the spectra with time. Delay times displayed are a subset of the data collected. The growth of the signal at ~ 1700 cm^{-1} in the GC-30 sample is shown in the inset.

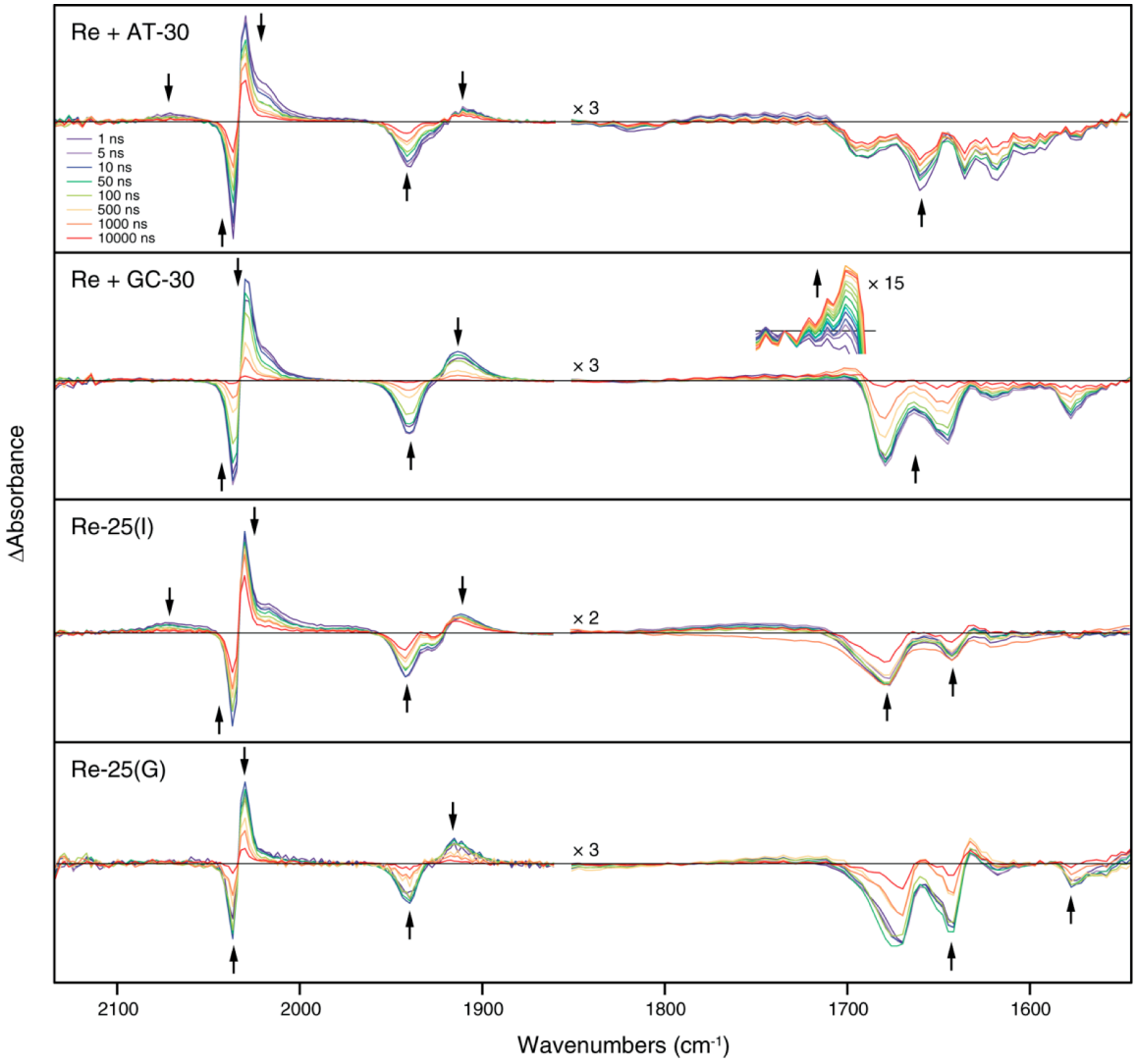
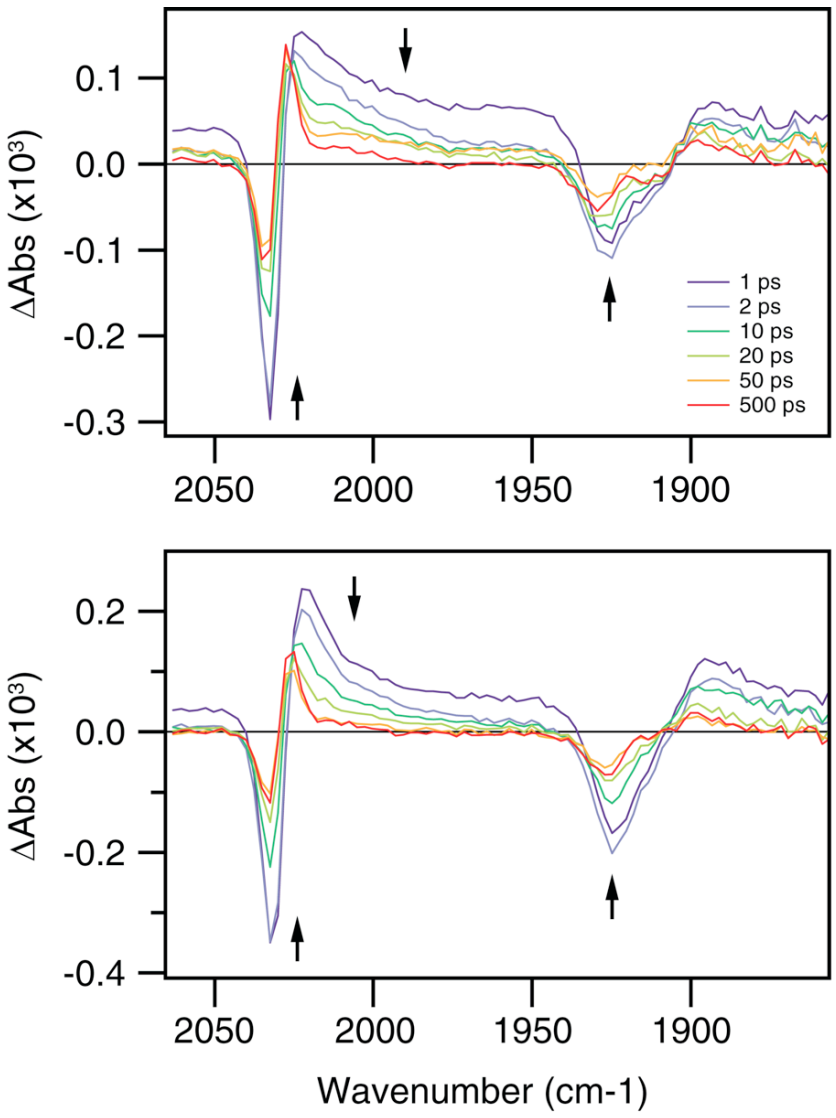


Figure 7.10. Picosecond-timescale TRIR difference spectra of 100 μM Re-25(I) (top) and Re-25(G) (bottom) measured at the indicated delay times after 355 nm, 50 fs excitation. Each probe data point is separated by ca. 2.1 cm^{-1}. Arrows indicate changes in the spectra with time. Delay times displayed are a subset of the data collected.



The 2070 cm^{-1} band can be assigned definitively to the MLCT excited state based on analyses of related complexes (*19-23, 35-37, 54, 69, 77*). This assignment predicts two additional low-intensity absorption bands near 2015 cm^{-1} and 1960 cm^{-1} due to hypsochromic shift of the A′(2) and A″ modes upon excitation of the complex into the MLCT state. These features are probably encompassed by the broad unresolved absorption between 1960 cm^{-1} and 1990 cm^{-1} and eclipsed by the much stronger absorption of IL states at higher energies. The MLCT features are absent in the GC-30 and Re-25(G) spectra. The AT-30 and Re-25(I) samples also exhibit a pronounced shoulder near 2020 cm^{-1} that is weaker for GC-30 and nearly absent in the Re-25(G) sample. This shoulder grows in intensity with increasing sample irradiation during the experiment, so it is in part related to transient absorption of a side photoproduct. However, its greater intensity in the AT-30 and Re-25(I) samples may be due to the presence of an underlying MLCT band or residual population of the ^{3}IL(phen) state, as observed in the picosecond experiments (see above). Importantly, on the nanosecond timescale, TRIR spectra of GC-30 in the DNA region show a growing band at ~ 1700 cm^{-1} attributable to the oxidized guanine radical, $G^{\bullet+}$ or $G^{\bullet}$. This transient is very similar to that observed at 1702 cm^{-1} in both 5′-dGMP and poly(dG-dC)·poly(dG-dC) upon 200 nm photoionization, which was assigned to oxidized guanine (although the particular ionic state of this radical was not determined) (*38, 43, 44*).

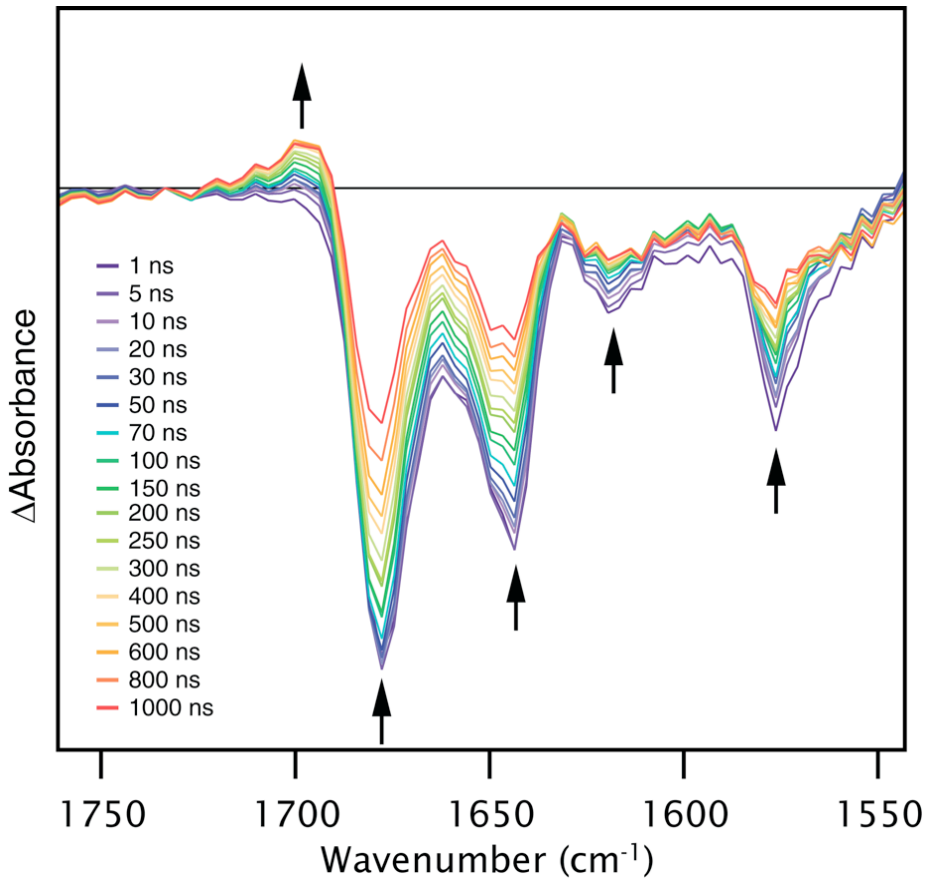
The nanosecond kinetic behavior of the four samples differs substantially in several ways (Table 7.1): (i) The bleach recoveries and ^{3}IL(phz) decays of the AT-30 and Re-25(I) are largely composed of long-lived components (≥ 20 μs) with smaller contributions on the time scale of tens to hundreds of nanoseconds. The occurrence of such slow microsecond processes, which have no counterparts in emission decays, indicates the presence of long-lived, non-emissive ^{3}IL excited states or transient species. (ii) The AT-30 and Re-25(I) MLCT band at ~ 2070 cm^{-1} is fully formed in the 1 ns spectra and decays monotonically over time (9 and 32 ns, respectively). In general, the lifetimes of the MLCT bands are significantly different than those of the IL bands, showing that the ^{3}IL and ^{3}MLCT

states are not equilibrated. Importantly, the 2070 cm^{-1} MLCT band is completely absent in the spectra of the GC-30 and Re-25(G) samples, probably due to very fast quenching of the ^{3}MLCT excited state by guanine. (iii) As compared to AT-30 and Re-25(I), both GC-30 and Re-25(G) show faster ^{3}IL decay and bleach recovery. (iv) Direct IR evidence for $G^{\bullet+}/G^{\bullet}$ formation was obtained for the GC-30 sample, where a band appears with a lifetime of 210 ns in the DNA spectral region at ~ 1700 cm^{-1} and then decays over ~ 20 μs with a lifetime estimated roughly as 6 μs, Figure 7.11.

As in the picosecond TRIR spectra, we do not see any distinct signals attributable to the reduced Re sensitizer [$Re^{I}(CO)_3(dppz^{\bullet-})(py'\text{-}OH)$]. This again is because its IR spectrum is nearly identical with that of the ^{3}IL(phz) state; moreover, the yield of reduced Re species is low due to efficient BET.

Visible TA Transient absorption decay in the visible spectral range at 475 nm was investigated in order to compare the TRIR kinetics specific to the $Re(CO)_3$ moiety with those of the dppz part of the chromophore. A single exponential term was a poor model for the transient decay, indicating that more than one transient species exists during the course of the measurement. Biexponential fit parameters for the transient decays are shown in Table 7.1. It should be noted that the TA experiments were performed with a time resolution of about 10 ns, so they only provide information on the slower kinetics and longer-lived intermediates. Still, the TA decay lifetimes for each sample are comparable to the decay of the TRIR band near 2030 cm^{-1}, including the lifetime shortening upon guanine incorporation near the Re binding site. It follows that the same states and processes are monitored by both methods. In a similar system, the TA spectrum of the reduced state following 355 nm excitation of a Re-DNA conjugate could not be distinguished from the spectrum of the excited state, presumably due to the greater concentration of the excited state and the strong similarity between the two spectra (*66*).

Figure 7.11. Nanosecond-timescale TRIR difference spectra showing changes in the IR absorbance of 4.8 mM (base pairs) GC-30 in the presence of 0.5 mM Re′-OH following excitation at 355 nm. Arrows indicate changes in the spectra with time. The increase in absorbance at ~ 1700 cm^{-1} is clearly displayed.



However, a change in the lifetime of the transient upon DNA binding suggested that DNA-mediated quenching by guanine was taking place. A similar effect is expected for the conjugates studied here.

Discussion

Interactions Between [Re(CO)$_3$(dppz)(py'-OR)]$^+$ and DNA Strong interactions between intercalating metal complexes and DNA are well known. As observed with several other dppz-bearing cationic metal complexes, incubation with DNA results in hypochromicity of the electronic spectrum and increased luminescence of the complex (*78-80*). Certainly, the light switch effect is a strong indicator of intercalative binding. Biexponential emission decays observed for other light switch complexes bound to DNA, such as dppz complexes of Ru, have been attributed to the existence of two different intercalative binding modes: a perpendicular mode, in which the metal-phenazine axis of the dppz ligand lies along the DNA dyad axis, and a side-on mode, in which the metal-phenazine axis lies along the long axis of the base pairs (*81*). In a similar way, the multiexponential emission decays observed for the Re complexes are probably due in part to the existence of several binding modes. Emission decay lifetimes of intercalated complexes are also affected by the DNA sequence to which they are bound (*82-84*). Although the range of DNA binding sites in the tethered complexes is limited, the tether is flexible enough to allow for binding at any of several locations, each of which may have a different effect on the luminescence lifetime. Similarly, for non-tethered samples, slight variations in the sequence at the binding site may contribute differently to the overall decay. DNA sequence effects, therefore, also contribute to the multiexponential emission decay kinetics of bound complexes.

The bleaches observed in the organic carbonyl stretching region (1600 cm^{-1} to 1700 cm^{-1}) of the TRIR spectra could be another indication of the strong interaction between the complexes and DNA. It is possible that such bleach signals arise from direct photoexcitation of DNA, but excited

states thus generated are expected to persist for only a few nanoseconds (*76*). On the contrary, the microsecond DNA bleach recovery lifetimes, commensurate with Re excited-state lifetimes observed herein, indicate that the bleached signals originate from perturbation of the bases upon photoexcitation of electronically coupled Re chromophore. A similar effect was observed previously upon 400 nm photoexcitation of $[Ru(dppz)(tap)_2]^{2+}$ intercalated nonspecifically into poly(dG-dC)·poly(dG-dC) (*85*). In that work, a series of overlapping bleach and transient signals in the organic carbonyl stretching region at short times (2 ps to 2 ns) was attributed to guanine oxidation by excited Ru via a proton-coupled electron transfer (PCET) mechanism. Such a mechanism seems unlikely in our system because of the absence of TRIR transients that could be assigned to changes in cytosine carbonyl stretching frequency.

Guanine Oxidation by *[Re(CO)$_3$(dppz)(py′-OR)]$^+$ Previous work has shown that extended irradiation of mixtures of $[Re(CO)_3(dppz)(py)]^+$ and supercoiled plasmid DNA at $\lambda > 350$ nm results in nicks in the DNA backbone (*42*). In that work, the yield of cleavage did not depend on the concentration of singlet oxygen, suggesting that cleavage is the result of direct oxidation of guanine by the excited complex. The experimental results described here provide further evidence for the oxidation of guanine in DNA duplexes by photoexcited $[Re(CO)_3(dppz)(py′\text{-}OR)]^+$. In PAGE experiments, oxidation was observed preferentially at the 5′-guanine of the 5′-GG-3′ doublet. Importantly, the observation of oxidation at this site, at least three base pairs removed from the Re binding site, indicates that long-range DNA-mediated CT has occurred. The preferential oxidation of the 5′-guanine of the doublet is typical for long-range DNA-mediated CT processes (*86, 87)*. This pattern is due to localization of the injected hole at guanine, the site of lowest oxidation potential (*68*). Once localized on guanine, proton transfer with base-paired cytosine results in the formation of the neutral guanine radical ($k \approx 10^7$ s^{-1}) (*38*). In this state, the radical is quite stable, and can persist for > 1 ms (*88*). In the present study, a greater yield of guanine damage was observed by

PAGE at the guanine doublet in Re-25(I) than in Re-25(G). This result can be attributed to the effect of the flanking guanines in Re-25(G). For each photon absorbed, CT may occur to any low potential guanine site that is well-coupled to the probe. Statistically, transfer to and trapping at the guanine doublet is more probable in Re-25(I) than in Re-25(G) since CT to inosine is expected to be thermodynamically less favorable. The long-range DNA-mediated oxidation of guanine observed in the gel experiment is not surprising, given the favorable driving force and strong electronic coupling between the complex and DNA.

The spectroscopic data are also consistent with guanine oxidation. By both steady-state and time-resolved emission, the luminescence intensity of each AT-30 and Re-25(I) is greater than that for GC-30 and Re-25(G), respectively. In early work, a similar disparity in the emission intensity of $[Re(CO)_3(dppz)(py)]^+$, a known DNA light-switch complex, bound to poly(dA)·poly(dT) versus poly(dG)·poly(dC) was ascribed to steric inhibition of binding to the latter duplex (*42*). Such an interpretation falls short on several accounts. First, it cannot explain the difference in emission intensity observed between Re-25(I) and Re-25(G); exchanging guanine for inosine at the Re binding site is expected to present a negligible change in steric interactions between the complex and the duplex. Second, it is not consistent with the equal degree of hypochromicity observed in the electronic spectrum of a similar Re complex when bound to either poly(dG-dC)·poly(dG-dC) or poly(dA-dT)·poly(dA-dT) (*57*). Finally, it contradicts the strong luminescence observed from the bulkier light switch complex $[Ru(bpy)_2(dppz)]^{2+}$ bound to poly(dG-dC)·poly(dG-dC) (*72, 89*). A more consistent explanation involves facile quenching of the Re excited state by guanine (*43, 44, 57*). CT from excited Re to guanine accounts well for our observation that the Re-25(G) and GC-30 samples, in which guanine neighbors the intercalation site, show less emission than the Re-25(I) and AT-30 samples, in which direct interaction between the complex and guanine is prevented.

TRIR spectra reported above provide further information on the rate and mechanism of guanine oxidation in GC-30 and Re-25(G). The reduced yield of the IL(phz) state relative to AT-30 and Re-

25(I) suggests that *Re→G CT involves the IL(phen) state and occurs on a comparable time scale as the IL(phen)→IL(phz) conversion, namely a few tens of picoseconds. In addition, the absence of MLCT features in spectra observed on the nanosecond timescale shows that parallel CT involving the MLCT state occurs with a sub-nanosecond lifetime. Under some circumstances, IL(phz) could be reactive as well, but we have no direct evidence for a process involving this state. The rate of *Re→G CT cannot be determined exactly by TRIR because the spectral patterns of the IL-excited and reduced states cannot be distinguished. Nevertheless, the picosecond-timescale CT rates are further corroborated by comparison of the instantaneous (t = 0) emission intensity between samples. On the nanosecond timescale, the four samples give similar emission decay rates, although the integrated emission intensity is much less for Re-25(G) and GC-30 than for Re-25(I) and AT-30.

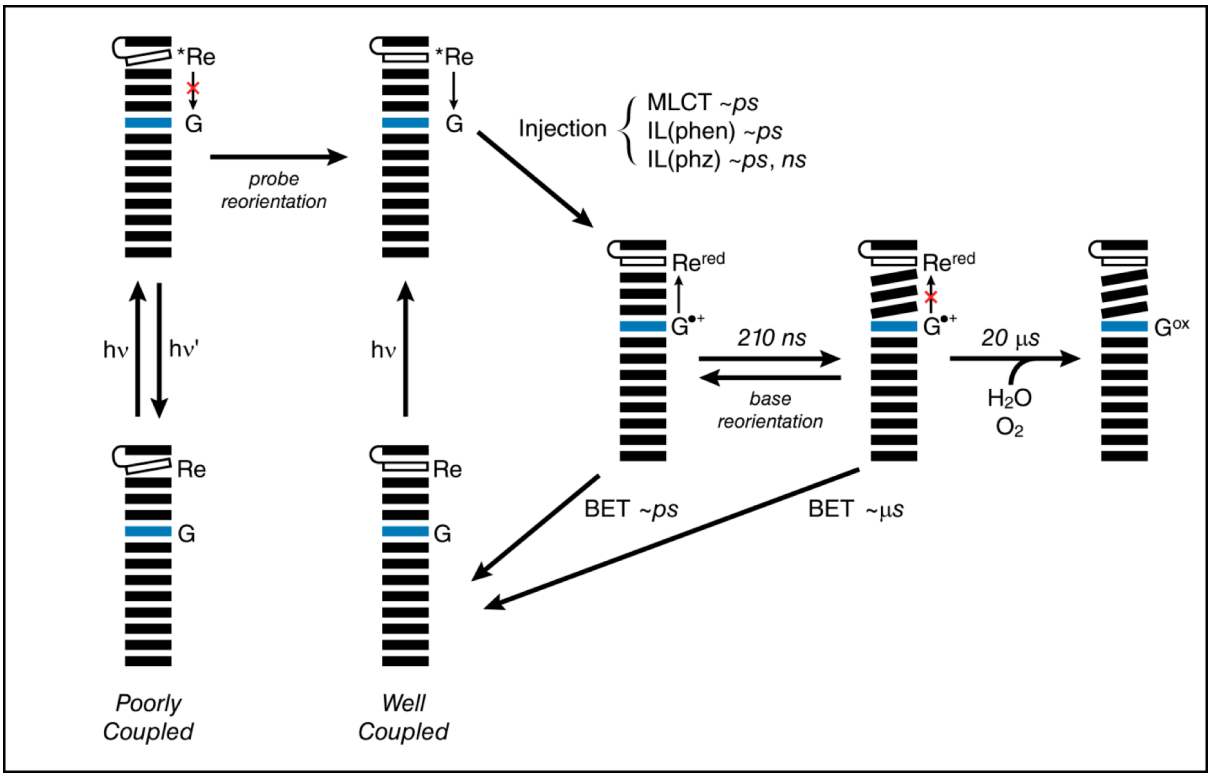
The reason for the absence of a guanine oxidation signal in TRIR spectra of Re-25(G) and Re-25(I) is unclear, but it may be an effect of the mixed base sequence used in these constructs. In previous studies of guanine oxidation by $[Ru^{III}(phen)_2(dppz)]^{3+}$, a strong transient was observed in the visible region that was attributed to the neutral guanine radical when the complex was intercalated in poly(dG)·poly(dG) or poly(dG-dA)·poly(dC-dT), but no signal was seen when the complex was intercalated in poly(dG-dT)·poly(dC-dA) (*88*). This difference was attributed to sequence dependent variations in the redox potential of guanine or structural variations, which would alter the coupling in the system.

Long-Lived Transient States In addition to the reactive Re states, TRIR and TA measurements indicate that one or more non-emissive transient states persists long after the emissive species has been depleted. We have established that the long-lived transients are composed primarily of mixtures of Re in the ^{3}IL(phz) excited state and in the reduced state, $[Re(CO)_3(dppz^{\bullet-})(py'\text{-}OR)]$. The long-lifetime decay processes observed by these absorption methods therefore contain

contributions from the decay of these two states. From the ^{3}IL(phz) state, the decay is likely due to internal conversion to the ground state. From the reduced state, the decay is caused by charge recombination, i.e., BET. The observation of long-lived transients in the AT-30 sample and the possibility for the oxidation of by excited Re indicate that some amount of charge injection may be possible in the absence of guanine. However, the lack of evidence for the formation of $A^{\bullet+}$ and the relatively strong emission observed in the AT-30 system suggest that if CT with adenine occurs, it is slow, minimally competitive with emission, and followed by fast BET.

Suggested Mechanism of DNA-Mediated Guanine Oxidation Based on spectroscopic evidence, a model can be generated for the oxidation of guanine by excited $[Re(CO)_3(dppz)(py'\text{-}OR)]^+$ (Figure 7.12). Photoexcitation of the complex populates a mixture of close lying IL(phen), IL(phz) and MLCT excited states, presumably spin-triplets, that are clearly observed by TRIR. (Such mixture of states also has been observed experimentally in analogous rhenium tricarbonyl complexes, and has been verified in computational models (*16, 17, 19, 21-23, 26, 27, 41, 90*). On the basis of our TRIR results, it appears that different excited states are more or less likely to participate in DNA-mediated CT. The MLCT state in particular, which is not observed in samples where the excited complex is in direct contact with guanine, seems to be more easily quenched than the IL states. The CT reactivity appears to decrease in the order MLCT > IL(phen) > IL(phz). It is also possible that conversion between excited states affects the apparent rates and yields observed for charge injection or emission. The reaction pathways from the excited state are also governed by the extent of electronic coupling in the system, which itself is determined by the dynamics of the probe and of the bases themselves (*66*). At the instant of excitation, two major populations exist. The first involves complexes which are poorly bound or which are bound to DNA in orientations that are not conducive to electron transfer. In this population, the mechanism of relaxation involves either quenching by water, observed for Ru-dppz complexes in polar, protic solvents (*56, 91*), or emission.

Figure 7.12. The proposed model for the oxidation of guanine by photoexcited $[Re(CO)_3(dppz)(py'\text{-}OR)]^+$. Photoexcitation in the poorly coupled system results in emission ($h\nu'$) or nonradiative decay to the ground state. Photoexcitation in the well-coupled system results in charge injection over an arbitrary distance to form reduced $[Re(CO)_3(dppz)(py'\text{-}OR)]^0$ (Re^{red}) and the guanine radical cation ($G^{\bullet+}$). During the excited state lifetime of the complex, the poorly coupled system may undergo reorientation, allowing charge injection. From the charge-separated state, facile back electron transfer (BET) competes with charge migration and trapping, resulting either in no reaction or the formation of permanent oxidation products. Base motions may result in isolation of the injected charge, favoring the trapping pathway.



Emission is expected to occur primarily from the ^{3}IL state, as reported for $[Re(CO)_3(dppz)(py)]^+$ in acetonitrile (*19*). In the second population, the excited complex is well coupled to the DNA. Here, excited state quenching via positive charge (i.e., hole) injection into the DNA duplex is the preferred reaction pathway. Indeed, primarily coherent CT at a distance of ten base pairs was observed in systems utilizing 2-aminopurine as a hole donor (*60*). Such processes are rapid. In systems involving DNA-bound ethidium, DNA-mediated CT over distances of several bases was observed to occur in 5 ps (*92*). Further, emission quenching is not limited to the population that exists in a CT-active configuration at the moment of excitation; reorientation of the bound oxidant to generate such a configuration may occur within the lifetime of the excited state. The rate of reorientation for DNA-bound ethidium is 75 ps (*92*), although for a larger molecule such as $[Re(CO)_3(dppz)(py'\text{-}OR)]^+$, this rate may be slower. Following charge separation, charge recombination (BET) may occur. After all, the ground-state oxidation of $[Re^I(CO)_3(dppz^{\bullet-})(py'\text{-}OH)]^0$ [$E°(Re^+/Re^0)$ = -0.85 V vs. NHE] by $G^{\bullet+}$ [$E°(G^{\bullet+}/G)$ = 1.29 V vs. NHE] (*68*) is thermodynamically favorable, and immediately after charge separation, the system exists in a CT-active state. Back reaction along this pathway is consistent with the absence of a guanine signal at short times in TRIR experiments. While this non-productive reaction pathway can be invoked to explain some of the experimental observations, additional pathways must be operative; quantitative deactivation of the charge separated state via short-range BET would prevent the eventual formation of permanent oxidative damage. A third population, then, involves molecules that are well coupled during charge injection, but that lose coupling before BET can take place due to reorientational motion of either the probe or the bases. The holes thus isolated within the base stack are quite stable and can migrate away from the site of injection, further reducing the probability for BET to occur and increasing the yield of permanent oxidative damage (*60*). Charge migration is limited in rate by stacking and destacking motions of the duplex, which form transient delocalized

electronic domains (*93, 94*). The 210 ns rate of formation for the guanine radical signal observed at ~ 1700 cm^{-1} by TRIR in the GC-30 sample may therefore reflect the rate of this conformational gating.

Concluding Remarks

Complexes that contain IR-active moieties show promise as probes for the study of DNA CT. In this work, we have used PAGE and time-resolved spectroscopy to observe the oxidation of guanine in DNA by photoexcited $[Re(CO)_3(dppz)(py'\text{-}OR)]^+$. Although no direct evidence for this reaction is afforded by UV/visible methods, fast excited-state quenching by guanine provides indirect evidence that oxidation is taking place. Direct evidence for the formation of guanine oxidation products is observed biochemically by PAGE analysis and spectroscopically by TRIR following photoexcitation of Re′-OH in the presence of GC-30. Similarities between the spectral features and kinetics of this system with those of other DNA sequences containing guanine allow us to conclude that the photochemical processes observed in the GC-30 sample are general. In these systems, the rate of guanine oxidation (herein 210 ns) is dictated largely by motions of the bases, which allow for long-range charge separation and prevent BET, rather than by the intrinsic photophysics of the photosensitizer complex. In this respect, the role of Re′-OH is similar to that of other photooxidants that have been used in DNA CT studies (*60, 93, 94*).

Unlike the well-known $[Ru(phen)_2(dppz)]^{2+}$ DNA "light-switch", rhenium(I) tricarbonyl-dppz complexes are strong enough photooxidants to inject positive charge into DNA directly from their electronically excited state(s), i.e., without the use of an external quencher and involvement of diffusion-controlled steps. This allows for ultrafast charge injection, with possible applications in mechanistic studies of DNA-mediated CT and in development of DNA-based photonic devices. However, the present study indicates that charge injection by $[Re(CO)_3(dppz)(py'\text{-}OR)]^+$ preferentially involves the initially populated IL(phen) and the minor MLCT states, with the long-

lived ^{3}IL(phz) state showing little reactivity, if any. This, together with fast BET, limits the reaction yield. From the experimental point of view, Re tricarbonyl-diimines have the advantage of being both ET phototriggers and probes by virtue of their sensitive IR spectral responses to changes in the electron density distribution (*32, 34*). However, in the particular case of dppz complexes, the TRIR spectral analysis is complicated by the close resemblance of ^{3}IL(phz) and reduced-state spectral patterns that renders the two species indistinguishable. It is suggested that optimization of the Re-photooxidant structure will improve both the charge injection efficiency and the IR spectral response.

A complete picture of DNA CT requires the observation of processes on very different time scales. At the instant of photoexcitation, the extent of coupling between the probe and the base stack, and between the bases themselves, defines two populations of DNA: one that is CT-active and one that is CT-inactive. The outcomes of fast processes, such as fluorescence and charge injection, are determined based on the relative sizes of these populations. At longer times, base motions change the energetic landscape, offering alternative reaction pathways, such as charge migration and trapping, that were not available immediately after excitation. TRIR allows for the observation of processes at all of these time scales, making it a valuable addition to the methods employed for the study of DNA-mediated CT.

REFERENCES

1. Genereux, J. C., and Barton, J. K. (2010) *Chem. Rev. 110*, 1642.

2. Barton, J. K., Olmon, E. D., and Sontz, P. A. (2010) *Coord. Chem. Rev. 255*, 619.

3. Schuster, G. B., (2004) Ed. *Topics in Current Chemistry: Long-Range Charge Transfer in DNA I*, Springer-Verlag.

4. Núñez, M. E., Hall, D. B., and Barton, J. K. (1999) *Chem. Biol. 6*, 85.

5. Slinker, J. D., Muren, N. B., Renfrew, S. E., and Barton, J. K. (2011) *Nature Chem. 3*, 228.

6. Augustyn, K. E., Genereux, J. C., and Barton, J. K. (2007) *Angew. Chem. Int. Ed. 46*, 5731.

7. Gorodetsky, A. A., Buzzeo, M. C., and Barton, J. K. (2008) *Bioconj. Chem. 19*, 2285.

8. Guo, X., Gorodetsky, A. A., Hone, J., Barton, J. K., and Nuckolls, C. (2008) *Nature Nanotech. 3*, 163.

9. Slinker, J. D., Muren, N. B., Gorodetsky, A. A., and Barton, J. K. (2010) *J. Am. Chem. Soc. 132*, 2769.

10. Genereux, J. C., Boal, A. K., and Barton, J. K. (2010) *J. Am. Chem. Soc. 132*, 891.

11. Boal, A. K., Genereux, J. C., Sontz, P. A., Gralnick, J. A., Newman, D. K., and Barton, J. K. (2009) *Proc. Natl. Acad. Sci. USA 106*, 15237.

12. Lee, P. E., Demple, B., and Barton, J. K. (2009) *Proc. Natl. Acad. Sci. USA 106*, 13164.

13. Augustyn, K. E., Merino, E. J., and Barton, J. K. (2007) *Proc. Natl. Acad. Sci. USA 104*, 18907.

14. Turner, J. J., George, M. W., Johnson, F. P. A., and Westwell, J. R. (1993) *Coord. Chem. Rev. 125*, 101.

15. Kumar, A., Sun, S. S., and Lees, A. J. (2010) *Top. Organomet. Chem. 29*, 1.

16. Stufkens, D. J., and Vlček, A. (1998) *Coord. Chem. Rev. 177*, 127.

17. Vlček, A. (2010) *Top. Organomet. Chem. 29*, 73.

18. Dyer, J., Grills, D. C., Matousek, P., Parker, A. W., Towrie, M., Weinstein, J. A., and George, M. W. (2002) *Chem. Commun. 8,* 872.

19. Dyer, J., Blau, W. J., Coates, C. G., Creely, C. M., Gavey, J. D., George, M. W., Grills, D. C., Hudson, S., Kelly, J. M., Matousek, P., McGarvey, J. J., McMaster, J., Parker, A. W., Towrie, M., and Weinstein, J. A. (2003) *Photochem. Photobiol. Sci. 2*, 542.

20. Kuimova, M. K., Alsindi, W. Z., Dyer, J., Grills, D. C., Jina, O. S., Matousek, P., Parker, A. W., Porius, P., Sun, X. Z., Towrie, M., Wilson, C., Yang, J., and George, M. W. (2003) *Dalton Trans. 21*, 3996.

21. Kuimova, M. K., Sun, Z., Matousek, P., Grills, D. C., Parker, A. W., Towrie, M., and George, M. W. (2007) *Photochem. Photobiol. Sci. 6*, 1158.

22. Dyer, J., Creely, C. M., Penedo, J. C., Grills, D. C., Hudson, S., Matousek, P., Parker, A. W., Towrie, M., Kelly, J. M., and George, M. W. (2007) *Photochem. Photobiol. Sci. 6*, 741.

23. Kuimova, M. K., Alsindi, W. Z., Blake, A. J., Davies, E. S., Lampus, D. J., Matousek, P., McMaster, J., Parker, A. W., Towrie, M., Sun, X.-Z., Wilson, C., and George, M. W. (2008) *Inorg. Chem. 47*, 9857.

24. Vlček, A., and Záliš, S. (2007) *Coord. Chem. Rev. 251*, 258.

25. George, M. W., and Turner, J. J. (1998) *Coord. Chem. Rev. 177*, 201.

26. Vlček, A., and Busby, M. (2006) *Coord. Chem. Rev. 250*, 1755.

27. Busby, M., Matousek, P., Towrie, M., and Vlček, A. (2007) *Inorg. Chim. Acta 360*, 885.

28. Lo, K. K.W., Louie, M.W., and Zhang, K. Y. (2010) *Coord. Chem. Rev. 254*, 2603.

29. Blanco-Rodríguez, A. M., Busby, M., Grădinaru, C., Crane, B. R., Di Bilio, A. J., Matousek, P., Towrie, M., Leigh, B. S., Richards, J. H., Vlček, A., and Gray, H. B. (2006) *J. Am. Chem. Soc. 3*, 3552.

30. Blanco-Rodríguez, A. M., Busby, M., Ronayne, K., Towrie, M., Grădinaru, C., Sudhamsu, J., Sýkora, J., Hof, M., Záliš, S., Di Bilio, A. J., Crane, B. R., Gray, H. B., and Vlček, A. (2009) *J. Am. Chem. Soc. 3*, 11788.

31. Gray, H. B., and Winkler, J. R. (2009) *Chem. Phys. Lett. 483*, 1.

32. Blanco-Rodríguez, A. M., Di Bilio, A. J., Shih, C., Museth, A. K., Clark, I. P., Towrie, M., Cannizzo, A., Sudhamsu, J., Crane, B. R., Sýkora, J., Winkler, J. R., Gray, H. B., Záliš, S., and Vlček, A. (2011) *Chem. Eur. J. 17*, 5350.

33. Connick, W. B., Di Bilio, A. J., Hill, M. G., Winkler, J. R., and Gray, H. B. (1995) *Inorg. Chim. Acta 240*, 169.

34. Shih, C., Museth, A. K., Abrahamsson, M., Blanco-Rodríguez, A. M., Di Bilio, A. J., Sudhamsu, J., Crane, B. R., Ronayne, K. L., Towrie, M., Vlček, A., Richards, J. H., Winkler, J. R., and Gray, H. B. (2008) *Science 320*, 1760.

35. Encinas, S., Morales, A. F., Barigelletti, F., Barthram, A. M., White, C. M., Couchman, S. M., Jeffery, J. C., Ward, M. D., Grills, D. C., and George, M. W. (2001) *J. Chem. Soc., Dalton Trans. 22*, 3312.

36. Towrie, M., Doorley, G. W., George, M. W., Parker, A. W., Quinn, S. J., and Kelly, J. M. (2009) *Analyst 134*, 1265.

37. Kuimova, M. K., Dyer, J., George, M. W., Grills, D. C., Kelly, J. M., Matousek, P., Parker, A. W., Sun, X. Z., Towrie, M., and Whelan, A. M. (2005) *Chem. Commun.* 1182.

38. Kuimova, M. K., Cowan, A. J., Matousek, P., Parker, A. W., Sun, X. Z., Towrie, M., and George, M. W. (2006) *Proc. Natl. Acad. Sci. USA 103*, 2150.

39. Hare, P. M., Middleton, C. T., Mertel, K. I., Herbert, J. M., and Kohler, B. (2008) *Chem. Phys. 347*, 383.

40. McGovern, D. A., Doorley, G. W., Whelan, A. M., Parker, A. W., Towrie, M., Kelly, J. M., and Quinn, S. J. (2009) *Photochem. Photobiol. Sci. 8*, 542.

41. Stoeffler, H. D., Thornton, N. B., Temkin, S. L., and Schanze, K. S. (1995) *J. Am. Chem. Soc. 117*, 7119.

42. Yam, V. W.W., Lo, K. K. W., Cheung, K. K., and Kong, R. Y. C. (1997) *J. Chem. Soc., Dalton Trans. 14*, 2067.

43. Smith, J. A., George, M. W., and Kelly, J. M. (2011) *Coord. Chem. Rev. 253,* 2021.

44. Cao, Q., Creeley, C. M., Dyer, J., Easun, T., Grills, D. C., Hudson, S., McGovern, D. A., McMaster, J., Pitchford, J., Smith, J. A., Sun, X. Z., Kelly, J. M., and George, M. W. (2011) *Photochem. Photobiol. Sci. 10,* 1355.

45. Holmlin, R. E., Dandliker, P. J., and Barton, J. K. (1999) *Bioconj. Chem. 10*, 1122.

46. Zeglis, B. M., and Barton, J. K. (2007) *Nat. Protoc. 2*, 357.

47. Boyle, P. D., Boyd, D. C., Mueting, A. M., and Pignolet, L. H. (1988) *Inorg. Chem. 27*, 4424.

48. Pletneva, E. V., Gray, H. B., and Winkler, J. R. (2005) *J. Molec. Biol. 345*, 855.

49. Pletneva, E. V., Gray, H. B., and Winkler, J. R. (2005) *Proc. Natl. Acad. Sci. USA 102*, 18397.

50. Kimura, T., Lee, J. C., Gray, H. B., and Winkler, J. R. (2007) *Proc. Natl. Acad. Sci. USA 104*, 117.

51. Kimura, T., Lee, J. C., Gray, H. B., and Winkler, J. R. (2009) *Proc. Natl. Acad. Sci. USA 106*, 7834.

52. Greetham, G. M., Burgos, P., Cao, Q., Clark, I. P., Codd, P. S., Farrow, R. C., George, M. W., Kogimtzis, M., Matousek, P., Parker, A. W., Pollard, M. R., Robinson, D. A., Xin, Z. J., and Towrie, M. (2010) *Appl. Spectrosc. 64*, 1311.

53. Towrie, M., Gabrielsson, A., Matousek, P., Parker, A. W., Blanco-Rodríguez, A. M., and Vlček, A. (2005) *Appl. Spectrosc. 59*, 467.

54. Dattelbaum, D. M., Omberg, K. M., Hay, P. J., Gebhart, N. L., Martin, R. L., Schoonover, J. R., and Meyer, T. J. (2004) *J. Phys. Chem. A 108*, 3527.

55. Gamelin, D. R., George, M. W., Glyn, P., Grevels, F. W., Johnson, F. P. A., Klotzbücher, W., Morrison, S. L., Russell, G., Schaffner, K., and Turner, J. J. (1994) *Inorg. Chem. 33*, 3246.

56. Turro, C., Bossmann, S. H., Jenkins, Y., Barton, J. K., and Turro, N. J. (1995) *J. Am. Chem. Soc. 117*, 9026.

57. Ruiz, G. T., Juliarena, M. P., Lezna, R. O., Wolcan, E., Feliz, M. R., and Ferraudi, G. (2007) *Dalton Trans. 3*, 2020.

58. Dohno, C., Stemp, E. D. A., and Barton, J. K. (2003) *J. Am. Chem. Soc. 125*, 9586.

59. Williams, T. T., Dohno, C., Stemp, E. D. A., and Barton, J. K. (2004) *J. Am. Chem. Soc. 126*, 8148.

60. Genereux, J. C., Wuerth, S. M., and Barton, J. K. (2011) *J. Am. Chem. Soc. 133*, 3863.

61. Yoo, J., Delaney, S., Stemp, E. D. A., and Barton, J. K. (2003) *J. Am. Chem. Soc. 125*, 6640.

62. O'Neill, M. A., Dohno, C., and Barton, J. K. (2004) *J. Am. Chem. Soc. 126*, 1316.

63. Genereux, J. C., Augustyn, K. E., Davis, M. L., Shao, F., and Barton, J. K. (2008) *J. Am. Chem. Soc. 130*, 15150.

64. The redox potentials of the canonical bases described in this work were determined by pulse radiolysis of the free nucleosides and are therefore estimates of the potentials of the bases in the DNA polymer environment. For a summary of experimentally-determined guanine redox potentials in different contexts, see Ref. 1.

65. Yam, V. W. W., Lo, K. K. W., Cheung, K. K., and Kong, R. Y. C. (1995) *J. Chem. Soc., Chem. Commun. 3*, 1191.

66. Olmon, E. D., Hill, M. G., and Barton, J. K. (2011) *Inorg. Chem. 50*, 12034.

67. Juris, A., Balzani, V., Barigelletti, F., Campagna, S., Belser, P., and Von Zelewsky, A. (1988) *Coord. Chem. Rev. 84*, 85.

68. Steenken, S., and Jovanovic, S. V. (1997) *J. Am. Chem. Soc. 119*, 617.

69. Dattelbaum, D. M., Omberg, K. M., Schoonover, J. R., Martin, R. L., and Meyer, T. J. (2002) *Inorg. Chem. 41*, 6071.

70. Burrows, C. J., and Muller, J. G. (1998) *Chem. Rev. 98*, 1109.

71. Williams, T. T., Odom, D. T., and Barton, J. K. (2000) *J. Am. Chem. Soc.122*, 9048.

72. Friedman, A. E., Chambron, J. C., Sauvage, J. P., Turro, N. J., and Barton, J. K. (1990) *J. Am. Chem. Soc. 112*, 4960.

73. Caspar, J. V., and Meyer, T. J. (1983) *J. Am. Chem. Soc*. *105*, 5583.

74. Kwok, W. M., Ma, C., and Phillips, D. L. (2009) *J. Phys. Chem. B 113*, 11527.

75. Markovitsi, D., Gustavsson, T., and Vayá, I. (2010) *J. Phys. Chem. Lett. 1*, 3271.

76. Banyasz, A., Vayá, I., Changenet-Barret, P., Gustavsson, T., Douki, T., and Markovitsi, D. (2011) *J. Am. Chem. Soc. 133*, 5163.

77. Kuimova, M. K., Gordon, K. C., Howell, S. L., Matousek, P., Parker, A. W., Sun, X. Z., Towrie, M., and George, M. W. (2006) *Photochem. Photobiol. Sci. 5*, 82.

78. Hiort, C., Lincoln, P., and Nordén, B. (1993) *J. Am. Chem. Soc. 115*, 3448.

79. Holmlin, R. E., and Barton, J. K. (1995) *Inorg. Chem. 34*, 7.

80. Shao, F., Elias, B., Lu, W., and Barton, J. K. (2007) *Inorg. Chem. 46*, 10187.

81. Hartshorn, R. M., and Barton, J. K. (1992) *J. Am. Chem. Soc.114*, 5919.

82. Jenkins, Y., Friedman, A. E., Turro, N. J., and Barton, J. K. (1992) *Biochemistry 31*, 10809.

83. Holmlin, R. E., Stemp, E. D. A., and Barton, J. K. (1998) *Inorg. Chem. 37*, 29.

84. Stemp, E. D. A., Holmlin, R. E., and Barton, J. K. (2000) *Inorganica Chimica Acta 297*, 88.

85. Elias, B., Creely, C., Doorley, G. W., Feeney, M. M., Moucheron, C., Kirsch-DeMesmaeker, A., Dyer, J., Grills, D. C., George, M. W., Matousek, P., Parker, A. W., Towrie, M., and Kelly, J. M. (2008) *Chem. Eur. J. 14*, 369.

86. Stemp, E. D. A., Arkin, M. R., and Barton, J. K. (1997) *J. Am. Chem. Soc. 119*, 2921.

87. Saito, I., Takayama, M., Sugiyama, H., Nakatani, K., Tsuchida, A., and Yamamoto, M. (1995) *J. Am. Chem. Soc. 117*, 6406.

88. Hall, D. B., Holmlin, R. E., and Barton, J. K. (1996) *Nature 382*, 731.

89. Murphy, C. J., Arkin, M. R., Ghatlia, N. D., Bossmann, S., Turro, N. J., and Barton, J. K. (1994) *Proc. Natl. Acad. Sci. USA 91*, 5315.

90. MacQueen, D. B., and Schanze, K. S. (1991) *J. Am. Chem. Soc. 113*, 7470.

91. Olson, E. J. C., Hu, D., Hörmann, A., Jonkman, A. M., Arkin, M. R., Stemp, E. D. A., Barton, J. K., and Barbara, P. F. (1997) *J. Am. Chem. Soc. 7863*, 11458.

92. Wan, C., Fiebig, T., Kelley, S. O., Treadway, C. R., Barton, J. K., and Zewail, A. H. (1999) *Proc. Natl. Acad. Sci. USA 96*, 6014.

93. O'Neill, M. A., Becker, H. C., Wan, C., Barton, J. K., and Zewail, A. H. (2003) *Angew. Chem. Int. Ed. 42*, 5896.

94. Shao, F., O'Neill, M. A., and Barton, J. K. (2004) *Proc. Natl. Acad. Sci. USA 101*, 17914.

CHAPTER 8

Summary and Perspective

DNA is an exquisite biomolecule that stores genetic information to encode proteins and regulate pathways, but even more remarkable, its unique structure imparts unprecedented function that allows it to mediate charge transport over long molecular distances, and potentially relay signals within the cell (*1-4*). DNA is composed of a sugar-phosphate backbone that encloses nucleotide base pairs layered on top of one other. The π-stacked bases of the DNA duplex make it identical to other conductive molecular wires, such as graphite (*1-4*). Using a variety of spectroscopic, electrochemical, biochemical techniques, we have investigated the fundamental properties of DNA CT (*1-11*). Importantly, the redox probes that we employ to monitor CT must be well-coupled to the DNA base pair stack, thus, for our studies, we are limited to molecules that we can design and build in the laboratory. Yet, Nature has already evolved incredible redox probes of its own: metalloproteins. Significantly, we are discovering that proteins from a variety of organisms and pathways incorporate redox-active cofactors, which are specialized "engines" that allow them to "tap-into" DNA CT. Therefore, just as a gold electrode can be used to send charge through DNA to a distally bound redox acceptor (*4*), it is possible that proteins take on both of these roles, simultaneously, within the cell.

One major repair pathway of the cell, base excision repair (BER), involves many proteins, some of which are in very low copy number (~ 30 copies), to scan the entire genome for damage (*12-15*). A subset of BER proteins contain a [4Fe4S] cluster that is not required for structure or folding; we then asked, why is the [4Fe4S] cluster present in these proteins? Our initial electrochemical experiments revealed that MutY and EndoIII, BER proteins, exhibit DNA-bound redox potentials of ~ 100 mV vs. NHE (*15-18*), which is relevant to the potentials found inside a cell. From these studies, we found that as MutY or EndoIII binds DNA, the [4Fe4S] cluster is activated toward oxidation 3+, and thus necessarily, has a > 1000 higher affinity (K_d) for DNA (*15*). Notably, BER proteins must search through thousands of bases to find damage, and assuming facilitated diffusion

and instant interrogation, the copy number of these proteins is not sufficient to locate lesions before the cell divides (*15*). We have proposed a model, however, in which bound proteins utilize DNA-mediated CT to efficiently scan regions of the genome, allowing them to reduce the population of sites to search by a localization of protein in the vicinity of lesions. In this model, there are four fundamental requirements: i) DNA CT occurs over distances of 100−500 bp; ii) lesions attenuate DNA CT; iii) repair proteins exhibit DNA-bound redox potentials of ~ 80−100 mV vs. NHE and are able to efficiently perform CT; iv) there is a balance of reduced and oxidized protein (*15*). We have previously, and are currently, performing experiments to investigate each requirement we propose for the DNA-mediated search. In a complementary way, the studies described here are directed toward monitoring the entire coordination process that occurs among repair proteins for lesion detection.

In the second chapter, we detail an assay we have developed to determine whether repair proteins could cooperate in a DNA-mediated search for damage. Atomic force microscopy (AFM) is employed to obtain a "snapshot" of the search for lesions, which is reported as a redistribution ratio. We have incorporated a C:A lesion that inhibits DNA CT but is not a specific substrate for the repair protein in the assay into a long DNA strand (3.8 kbp). In this assay, strands and proteins are visualized and tallied from multiple samples and images based on height (nm). When EndoIII protein is incubated with long mismatched strands and short matched duplexes, we observe a higher density of proteins bound to the mismatched strands. Consistent with the model, this observation correlates with the ability of EndoIII to perform DNA CT, which has been characterized electrochemically. Alternatively, in the fully matched controls, there is no lesion to "home-in on", and the distribution ratio is ~ 1, indicating there is an equal number of proteins bound to long and short strands. Importantly, the model depends on the percentage of protein in the oxidized/reduced states. To this end, we add small concentrations of peroxide to the protein/DNA samples and

examine them with AFM. Only in the mismatched samples, we find the binding density ratio increases proportional to the peroxide treatment, indicating that indeed, when more protein is oxidized, it is bound in the vicinity of the lesion. From these experiments, AFM emerges as a tool, complementary to other methods we have developed in the lab, that we can apply to investigate whether proteins coordinate to find damage (*15*).

We then wondered: which residues are involved in mediating charge between the DNA and the [4Fe4S] cluster? Chapter three discusses a number of EndoIII mutants that we have characterized based on their ability to perform DNA CT and their propensity to redistribute in AFM studies (*19*). We find that the ability of a protein to localize in the vicinity of a lesion correlates with its ability to send and receive charge. Importantly, we note that the Y82A EndoIII mutant, which is deficient in CT, has been associated with colorectal cancer in humans (*15*), thus we must consider the implications of a DNA-mediated search process *in vivo*.

Applying genetic techniques, we have begun exploring this model in *E. coli* cells (*15*). The assay has been designed to measure the ability of MutY, a BER protein to repair oxidative damage *in vivo*. Our search model benefits from cooperation among metalloproteins that exhibit similar DNA-bound potentials. Thus we have asked whether eliminating one of these "helper" proteins disrupts the search partner left behind. We find that indeed this is the case, as we observe a decrease in the ability of MutY to fix damage when EndoIII is knocked out genetically. In chapter four, we describe methods designed to improve the signal output of this assay and ultimately reduce error in the measurements. Additionally, we present efforts toward developing a high-throughput screen, based on the helper function assay, to identify other proteins that may cooperate with MutY in a DNA-mediated search for damage.

Could the DNA-mediated search also occur among metalloproteins from different repair pathways? Excitingly, XPD, an archaeal protein from the nucleotide excision repair (NER) pathway, incorporates a [4Fe4S] cluster (*20, 21*). Similar to the [4Fe4S] cluster found within BER

proteins, the cluster is not required for XPD to fold and, additionally, is labile. XPD is a 5′-3′ helicase, one component of the transcription factor IIH machinery involved in binding near bulky lesions and exposing them for repair (*20, 21*). Significantly, XPD has a similar DNA-bound redox potential to EndoIII (~ 80 mV vs. NHE) (*22*). Notably, the efficiency of DNA CT increases upon the addition of ATP, as its helicase activity is activated (*22*). Might XPD also participate in the search process? AFM results, presented in chapter five, establish that XPD redistributes onto long mismatched strands. In these studies, the short strand overhangs are blocked with complementary single strand, so that XPD does not have a bias toward the short strands. Expanding upon these observations, we then utilized AFM to investigate whether proteins from different pathways (BER vs. NER) could cooperate with one another. To this end, we mixed XPD and EndoIII, proteins from distinct organisms as well as pathways, and discovered that they localize in the vicinity of lesions (*23*)! If, instead, we introduced a mutant either of EndoIII or XPD, defective at CT, we did not observe this redistribution (*23*). This result is exciting, considering these two proteins are from two different organisms. DNA signaling appears to depend only upon the [4Fe4S] clusters in the DNA binding proteins and this ability to couple through CT.

Probing the model further, we use AFM in an attempt to observe XPD helicase activity in conjunction with redistribution. In chapter 6, we find no difference in redistribution when we add ATP to XPD/DNA samples. In an effort to activate helicase activity, and observe unwinding in 'real-time', we have substituted ATP with a caged-analogue that is deprotected with UV-light. Post-irradiation, we observe forked DNA structures, indicative of XPD helicase activity. We are now in the process of expanding these studies using different concentrations of ATP and scanning speeds.

We have also probed DNA CT using time-resolved IR (TRIR) spectroscopy. In the first step of the search model, a reduced metalloprotein binds DNA, and consequently, gets oxidized. Since the guanine base is the most susceptible to oxidation, we have considered that guanine oxidation may

trigger the metalloprotein to bind DNA, initiating the search process. We have designed a Re-based metal complex, a photooxidant, to initiate this process. Significantly, we observe guanine oxidation via spectroscopic and biochemical techniques (*24*).

As we identify the metalloprotein players in the DNA-mediated search damage, it is clear that many pieces of this puzzle are interchangeable. In order to participate, proteins must bind DNA and simply have the means (i.e., redox-active moiety) to harness DNA-mediated CT. We have the methods and tools to characterize the DNA-bound redox potentials of proteins from a variety of pathways. Described here, AFM and *in vivo* assays will help us establish whether there is signaling between the proteins we identify in the future. Metalloproteins containing [4Fe4S] clusters are emerging rapidly: polymerases, primases, and transcription factors (*25*). We can even consider proteins from other organisms.

Pathogenic actinobacteria such as *Mycobacterium tuberculosis*, *Cornebacterium diphtheria*, and *Streptomyces coelicolor*, possess specific mechanisms that help them survive and thrive under extreme cellular conditions including oxidative stress (*26*). Recently, WhiB-like (Wbl) proteins have been identified as major components of redox control in actinobacteria (*27-29*). What is the biochemical role of these proteins? Importantly, these proteins bind DNA, act as transcription factors, and incorporate a [FeS] cluster cofactor that is stable under anaerobic conditions (*30*). What is the role of these [FeS] clusters in sensing redox conditions? Might Wbl proteins also be involved in DNA-mediated signaling?

We now move toward identifying and understanding the players in this search model. Just as nature takes advantage of the resources it has, so will we employ the tools we have designed to probe DNA-mediated CT in a biological context.

REFERENCES

1. Kelley, S. O., Boon, E. M., Barton, J.K., Jackson, N. M., and Hill, M. G. (1999) *Nucleic Acids Res. 27*, 4830.

2. Boon, E. M., Ceres, D. M., Drummond, T. G., Hill, M. G., and Barton, J. K. (2000) *Nat. Biotechnol. 18*, 1096.

3. Boal, A. K., and Barton, J. K. (2005) *Bioconj. Chem*. *16*, 312.

4. Slinker, J. D., Muren, N. B., Renfrew, S. E., and Barton, J. K. (2011) *Nat. Chem. 3*, 228.

5. Boon, E. M., Livingston, A. L., Chmiel, N. H., David, S. S., and Barton, J. K. Barton (2003) *Proc. Natl. Acad. Sci. 100*, 12543.

6. Núñez, M. E., Holmquist, G. P., and Barton, J. K. (2001) *Biochemistry 40*, 12465.

7. Boon, E. M., Salas, J. E., and Barton, J. K. (2002) *Nat. Biotechnol. 20,* 282.

8. Rajski, S. R., and Barton, J. K. (2001) *Biochemistry 40*, 5556.

9. Gorodetsky, A. A., Ebrahim, A., and Barton, J. K. (2008) *J. Am. Chem. Soc. 130*, 2924.

10. Boon, E. M., and Barton, J. K. (2002) *Curr. Opin. Struct. Biol. 12*, 320.

11. Núñez, M. E., Noyes, K. T., and Barton, J. K. (2002) *Chem. Biol. 9*, 403.

12. Demple, B., and Harrison, L. (1994) *Annu. Rev. Biochem. 63*, 915.

13. David S. S., O'Shea V. L., and Kundu, S. (2007) *Nature 447*, 941.

14. Francis, A. W., Helquist, S. A., Kool, E. T., and David, S. S. (2003) *J. Am. Chem. Soc. 125*, 16235.

15. Boal, A. K., Yavin, E., Lukianova, O. A., O'Shea, V. L., David, S. S., and Barton, J. K. (2005) *Biochemistry 44*, 8397.

16. Cowan, J. A., and Lui, S. M. (1998) *Adv. Inorg. Chem. 45*, 313.

17. Gorodetsky, A. A., Boal, A. K., and Barton, J. K. (2006) *J. Am. Chem. Soc. 128*, 12082.

18. Boal, A. K., Genereux, J., Sontz, P., Gralnick, J. A., Newman, D. K., and Barton, J. K. (2009) *Proc. Natl. Acad. Sci. USA 106*, 15237.

19. Romano, C. A., Sontz, P. A., and Barton, J. K. (2011) *Biochemistry 50*, 6133.

20. Fan, L., Fuss, J. O., Cheng, Q. J., Arvai, A. A., Hammel, M., Roberts, V. A., Cooper, P. K., and Tainer, J. A. (2008) *Cell 133*, 789.

21. Wolski S. C., Kuper, J., Hänzelmann, P., Truglio, J. J., Croteau, D. L., Van Houten, V., and Kisker, C. (2008) *PLos Biol. 6*, 1332.

22. Mui, T. P., Fuss, J. O., Ishida, J. P., Tainer, J. A., and Barton, J. K. (2011) *J. Am. Chem. Soc. 133*, 16378.

23. Sontz, P. A., Mui, T. P., Fuss, J. O., Tainer, J. A., and Barton, J. K. (2012) *Proc. Natl. Acad. Sci. USA 109*, 1856.

24. Olmon, E. D., Sontz, P. A., Blanco-Rodríguez, A. M., Towrie, M., Clark, I. P., Vlček, A., Jr., and Barton, J. K. (2011) *J. Am. Chem. Soc. 133*, 13718.

25. Netz, D. J. A., Stith, C. M., Stümpfig, M., Köpf, G., Vogel, D., Genau, H. M., Stodola, J. L., Lill, R., Burgers, P. M. J., and Pierik, A. J. (2012) *Nature Chem. Bio. 8*, 125.

26. Saini, V., Farhana, A., and Steyn, A. J. C. (2012) *Antiox. Redox Sign*. *16*, 687.

27. Hutter, B., and Dick, T. (1999). *Res. Microbiol. 150*, 295.

28. den Hengst, C. D., and Buttner, M. J. (2008) *BBA—Gen. Subj*. *1780*, 1201.

29. Fowler-Goldsworthy, K., Gust, B., Mouz, S., Chandra, G., Findlay, K. C., and Chater, K. F. (2011) *Microbiology 157*, 1312.

30. Singh, A., Guidry, L., Narasimhulu, K. V., Mai, D., Trombley, J., Redding, K. E., Giles, G. I., Lancaster Jr., J. R., and Steyn, A. J. C. (2007) *Proc. Natl. Acad. Sci. USA 104*, 11562.