SYNTHESIS AND APPLICATIONS OF BULKY RHODIUM(III) INTERCALATORS FOR THE RECOGNITION OF DNA MISMATCHES

Thesis by

Jonathan Ross Hart

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2006

(Defended May 24, 2006)

Acknowledgements:

I would especially like to thank my research advisor, Professor Jacqueline Barton. Your enthusiasm for chemistry has always been an inspiration. I am endebted to you for your ideas and perspectives which have turned me into a better chemist. I appreciate the projects you have suggested, for allowing me to pursue some ideas of my own and for providing me and the rest of the group with resources in order to make it happen. I also appreciate your ability to bring together a group of talented and intelligent people within the group and am glad to have been a part of it.

I would like to thank the many members of the Barton group have helped me in various ways and I especially thank the following people. Henrik Junicke started me out in lab, taught me gel electrophoresis and inorganic synthesis. Kim Copeland taught me PCR and associated techniques. Alex Schnyder and Jae Yoo taught me plenty of synthetic tricks and were always ready to celebrate over a pint. Tashica Williams who taught me gel electrophoresis again and made a great corner mate. Christopher Treadway helped me to understand ruthenium synthesis, and always had something entertaining on the computer. You pointed me in the direction of the finest eating establishments in the LA area. Melanie O'Neill always encouraged me and got me interested in many subjects I probably would never have known existed. Eva Rüba is a suberb synthetic chemist from who I learned a great deal and who helped keep my spirits up through trying times. Sarah Delaney, who taught me cell culture and set up the entire cell culture lab, I always enjoyed our conversations on science or whatever and am still disappointed I could never figure out the PCR problem. I'd like to thank Irvin Lau for being a good friend, being helpful in lab, trying new places to eat and turning me onto plenty of computer generated distractions both in lab and at home. Brian Zeglis, Russ Ernst and Paul Lee for technical assistance, many helpful discussions and for making the corner a fun place to work.

Finally, I would like to thank Mo Renta who has helped me with so many things over the years it is hard to keep track. You were an invaluable resource.

I would like to thank Neha Das and Yan Qi for their fruitful summer undergraduate projects.

I would like to thank those who I have collaborated with outside of Caltech. Marty Johnson from Applied Biosystems provided plasmids, equipment, chemicals and advice on the SNP project. Oleg Glebov, who knows more about human cell culture than anyone I have ever met, for teaching me how to culture cells the right way, and for many helpful discussions and suggestions. Ilan Kirsch for enlightening me about mismatch repair deficient cancers.

Finally, I would like to thank my friends and family. I would like to thank Mom, Dad and Grandma for their encouragement and support over my years here at Caltech. I appreciate the many things you have done over the years that have let me get to this point. To Jennifer, thanks for the support you have given me and for always having a place for me to visit at get away from everything. Tony thanks for always being a good friend over these many years. I appreciate the times out on the town, the Sunday dinners and all the times you've been there for me. David, Spooner and Odyssey I'd like to thank you for all the fun times we've had together over the years. My roommates at the Mentor House, Greg, James, Ted, Ali, Jonathan thanks for putting up with me and making Pasadena a fun place to live. I'd also like to thank Ted for everything and agreeing to put up with me for another year.

Thank you all.

Abstract:

The recognition of DNA base mismatches is of considerable interest for both the diagnosis and treatment of mismatch repair-deficient cancers. Two new mismatch recognition complexes have been synthesized. The first, $[Rh(bpy)_2(phzi)]^{3+}$ (phzi=benzo[a]phenazine-5,6-quinone diimine), recognizes DNA mismatches with high specificity and affinity, 1×10^7 M^{-1}, two orders of magnitude stronger than $[Rh(bpy)_2(chrysi)]^{3+}$ (chrysi=chrysene-5,6-quinone diimine), the parent complex that binds single thermodynamically-destabilized base-mismatch sites in duplex DNA. The second, $[Rh(bqdi)_2(chrysi)]^{3+}$, is able to recognize more stable mismatches such as the G-G mismatch.

These complexes have been applied in a variety of ways. A method has been developed for the discovery of new single nucleotide polymorphisms, SNPs, within a sequence of interest amplified from pooled genomic DNA. SNPs are readily detected using these mismatch selective molecules without false positives; allele frequencies as low as 0.05 can be detected.

Upon photoexcitation, the rhodium(III) diimine complexes cleave DNA by hydrogen atom abstraction from the sugar to yield 3'-phosphate terminated DNA that is inactive for enzymatic modification. This 3'-phosphate can be removed using T4-polynucleotide kinase opening up the possibility of enzymatic modification at the site of rhodium cleavage. The cleavage site can be fluorescently labeled. Terminal transferase can also be used to attach a homopolymer tail tagging the damage site, allowing the amplification of the DNA up to the damaged site.

This assay can also be employed towards the development of early cancer diagnostics. Some cancers are deficient in the repair of DNA base mismatches. As a consequence, these cells have an increased number of mismatches within their genome. These mismatches in extracted genomic DNA were cleaved using mismatch-specific rhodium complexes. The cleavage sites were labeled with radioactivity, allowing the number of mismatch sites to be quantitated. A significant number of sites were cleaved in the mismatch repair deficient DU145 cell line, 1 base/3000 bp, while no sites were cleaved in the mismatch repair proficient cell line SW620. This method may present a new method for the detection of mismatch repair deficiency.

These mismatch-specific complexes also are shown to have an antiproliferative effect on mismatch repair deficient cell lines. Mismatch repair deficiency is a contributing factor in both hereditary and sporadic human cancers. Both $[Rh(bpy)_2(chrysi)]Cl_3$ and $[Rh(bpy)_2(phzi)]Cl_3$ show a stronger antiproliferative effect against MMR deficient cells than proficient cells. Effects of stereoisomers, incubation time, and UV irradiation are also demonstrated.

Table of Contents

List of Figures and Tables

Chapter 1: Introduction

1.1: Introduction to deoxyribonucleic acid

Deoxyribonucleic acid, DNA, is the genetic storage material of cells. Very early genetic studies by Griffith showed that heat inactivated, cell free extracts from pathogenic bacteria could be used to transform nonpathogenic bacteria.[1] The agent responsible for the transformation of these bacteria was purified and discovered to consist exclusively of deoxyribonucleic acid.[2] This experiment, as well as others with eukaryotes, firmly established DNA as the primary carrier of genetic information.

DNA is a long polymer of individual deoxyribonucleotides, each of which are composed of a base, a sugar and a phosphate, shown in **figure 1.1**. The bases encode the genetic information while the sugar and phosphates participate in a structural role. The bases present in DNA are adenine (A), thymine (T), guanine (G) and cytosine (C). Early x-ray fiber diffraction studies were used by Watson and Crick to elucidate the double helical structure of DNA we know today.[3] A DNA duplex is composed of two antiparallel DNA strands whose bases interact through hydrogen bonding and aromatic stacking. Because of the chemical functionalities presented, adenine pairs well with thymine and guanine with cytosine. When so arranged, the base-paired units are virtually identical in width, 10.8 Å. This initial model was later refined by single crystal x-ray diffraction studies.

The structure of DNA also solves the issue of genetic information inheritance. One DNA duplex can be used as a template for two identical duplexes.[4] If the DNA duplex were opened, exposing the functionalities of the bases, complementary bases could be assembled to form a new complementary DNA. By this process each DNA strand encodes for its own complement. *In vivo*, DNA polymerases catalyze the reaction of deoxyribonucleotide triphosphates to produce complementary DNA to a parent

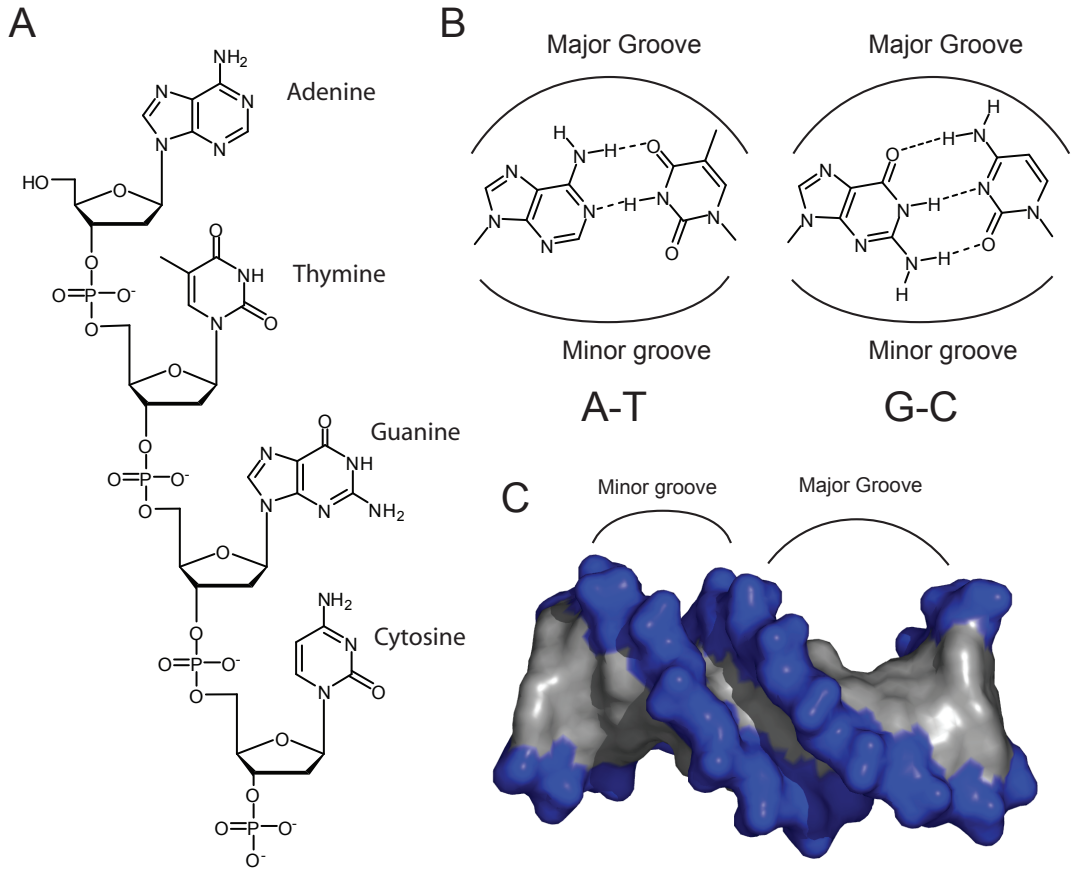

Figure 1.1: Deoxyribonucleic acid. A. Structures of the 4 natural DNA bases attached to a sugar phosphate backbone. B. The two Watson-Crick base pairs A-T and G-C and major and minor groove are indicated. C. When two complementary strands are annealed the DNA forms a B-form double helix. Major and minor grooves are indicated. Bases are colored grey with the sugar phosphate backbone in blue.

template strand.[5] In this way the original DNA duplex is copied, a necessary step in cell division.

The copying of genetic material is a particularly important step in the reproduction of cells. The primary focus of DNA polymerases is to replicate the parent DNA with as high a degree of fidelity as possible.[6] If just the catalytic domain of a DNA polymerase were used to synthesize DNA, approximately 1 base is inserted incorrectly in 10^4 base pairs.[7] This error rate is unacceptable for even the simplest of organisms which have genomes of ~10^6 base pairs, not to mention eukaryotes such as humans which have ~10^9 base pairs.[8] In addition to the catalytic domain of DNA polymerase, several other mechanisms are employed to ensure high fidelity DNA replication. First, the DNA polymerase enzyme also includes a proofreading domain which checks the base pairs immediately after DNA synthesis, excising those which are incorrectly incorporated. This boosts the fidelity of the DNA polymerase to 1 incorrect base in ~10^7–10^8 base pairs. Finally, postreplication repair extends the fidelity even further to <1 base pair in 10^9.[7] These incorrectly inserted bases, if not repaired, will form permanent mutations upon further DNA replication.

These incorrectly inserted bases arise from two distinct sources. First, bases within a DNA strand[9] as well as nucleotide triphosphate pools[10] can undergo various chemical reactions within the cell such as oxidation, reduction, and alkylation. These chemically modified bases present different functionalities and may not be recognized or inserted correctly by DNA polymerase.[11] These errors are detected by primarily by base excision repair enzymes,[12] and phosphatases to eliminate unnatural nucleotide triphosphates.[10] The second source is from the incorrect insertion of a natural DNA base into an incorrect DNA base pair. These mismatched base pairs are repaired by the MutS/L “long patch” repair pathway.[13]

1.2: Consequences of DNA mismatches

1.2.1: Physical consequences of DNA mismatches

DNA bases which are not paired with Watson-Crick partner are termed "base mismatches." These base mismatches are of considerable interest for their role in the formation of natural mutations. The proposed hydrogen bonding structures are shown in **figure 1.2**. Base mismatches in DNA have been studied by a variety of methods including x-ray crystallography, NMR, and melting temperatures. Through these physical studies it is hoped that the mechanism by which mismatches are distinguished *in vivo* can be determined.

Several structures of DNA mismatches have been obtained by single crystal x-ray crystallography. Structures of G-A,[14] G-G,[15] C-A,[16] and G-T[17] mismatches have been obtained, shown in **figure 1.3**. All of the structures obtained have two internal mismatches and are modifications of the Drew-Dickerson dodecamer.[18] These structures all show the DNA to be in essentially B-form conformation without kinks or extrahelical bases. The individual mispaired bases do adopt unusual conformations in order to maximize their hydrogen bonding and aromatic stacking. Some arguments have been made as to the validity of these crystal structures. The only crystal structures obtained have been for more thermodynamically stable mismatches in a particularly stable duplex. Perhaps the reason these crystallize may be they are B-form while others are not.

In addition to x-ray crystallography, NMR studies have been undertaken to address concerns that the x-ray structures may have been exceptions. NMR can be used to determine dynamics as well as structure of DNA molecules under more physiologically relevant conditions.[19] Solution structures of various mismatched duplexes have confirmed that the overall structure of the DNA is essentially B-form in solution. Varying of the sequence context of some mismatches such as the G-T mismatch do result in

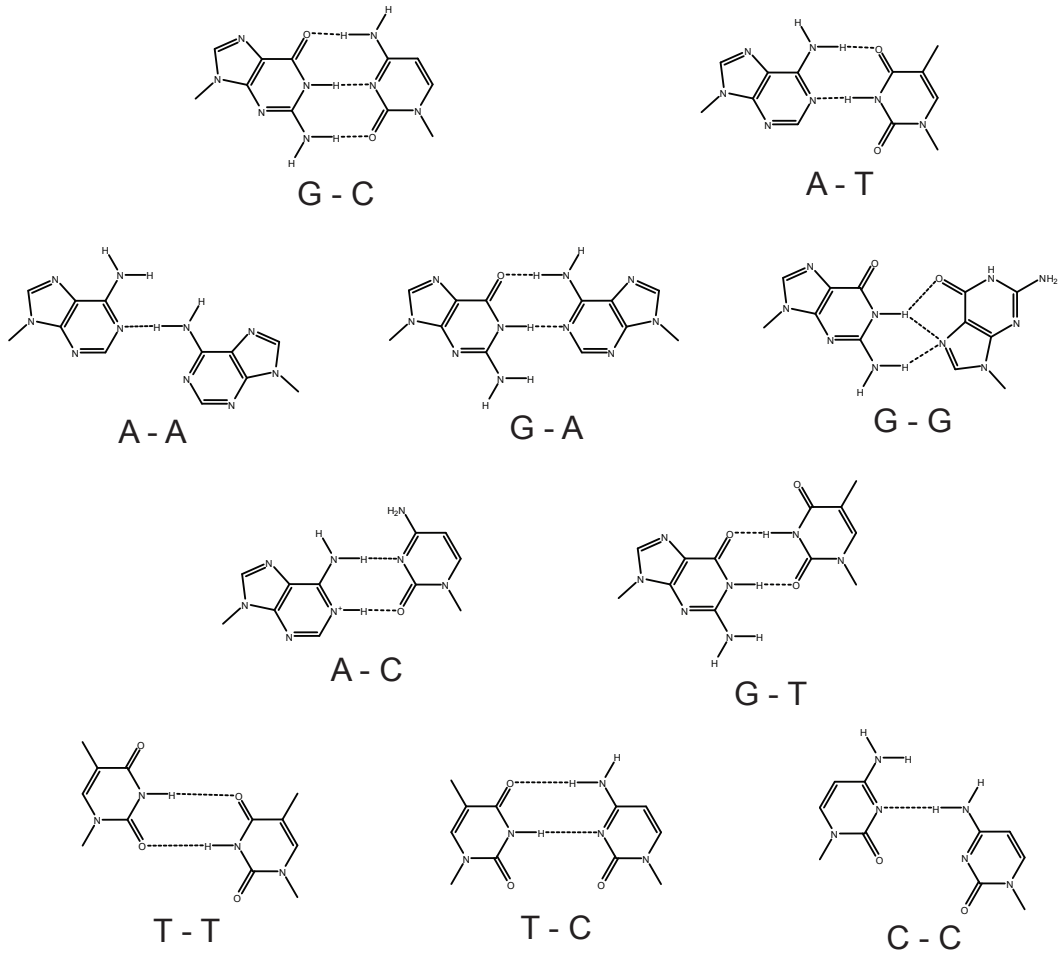

Figure 1.2: Structures of the possible base pairs. At the top we see the Watson-Crick base pairs, G-C and A-T, for comparison. Below these are the various mismatches.

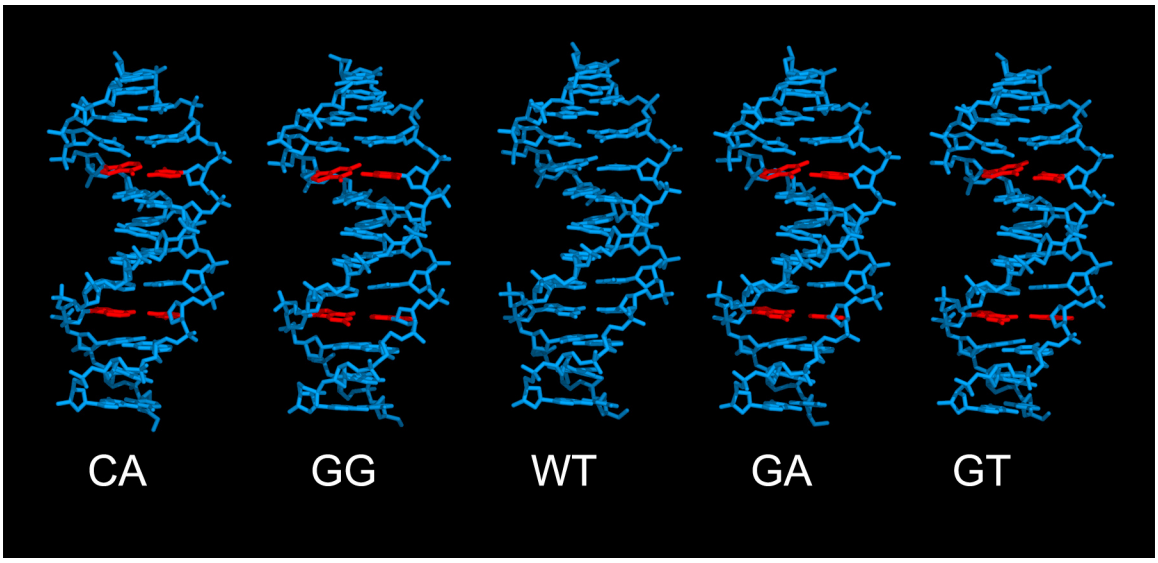

Figure 1.3: Crystal structures of DNA containing mismatched bases. Four duplexes are labeled with the mismatch they contain, while the central duplex without a mismatch is shown for comparison. The mismatched bases are highlighted in red. The overall structure is not significantly perturbed, but those bases which are mismatched do vary slightly.

conformational changes of the mismatched bases, a switch occurs between two possible hydrogen bonding schemes.[20; 21] Initial ^{1}H NMR studies have shown increased rates of solvent exchange of protons involved in hydrogen bonding.[19] It is expected that in order for this solvent exchange to occur, mismatched bases must undergo a conformational change to expose the bases to bulk solvent. This implies thermodynamic destabilization of the mismatched bases as compared to the natural DNA bases, a fact which can be quantitatively measured using variable temperature NMR. Further NMR studies have confirmed that the mismatched bases undergo a conformational switch which occurs faster than the NMR time scale.[21] However this occurs without disruption of the overall canonical B-form structure of the mismatched DNA. Other local perturbations in dynamics are observed in addition to the mismatched bases.[22] ^{13}C NMR relaxation studies have shown that base pairs adjacent to a mismatch experience a decrease in dynamical motion as compared to other Watson-Crick base pairs.[21]

These structural studies have suggested that DNA containing a mismatch has an equilibrium structure which is essentially the same as standard B-form DNA. However, in solution the mismatched bases are seen to be more fluctional. This suggests that perhaps when stress is applied to mismatched DNA, the DNA will form a kink at the mismatched site, as shown in **figure 1.4**.[23] However no such kink has been experimentally observed.

In addition to these structure and dynamics studies, mismatches have been studied energetically as well.[24] Two sources of thermodynamic data are available, melting temperatures by UV spectroscopy, which gives data about the DNA duplex as a whole, and variable temperature NMR which can provide thermodynamic data about each individual base pair.

Melting temperature data were the first data to show that mismatches are destabilizing to DNA duplexes as a whole. Using melting temperatures it is possible to

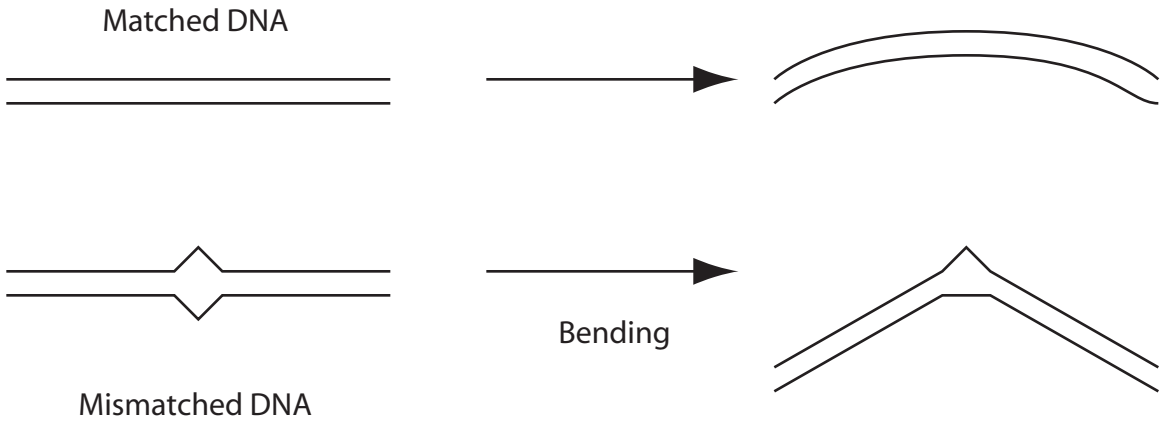

Figure 1.4: Schematic representation of the mismatched DNA bending theory. Two duplexes are shown on the left, matched and mismatched. When these duplexes are subjected to bending, different results are proposed. Matched DNA is thought to form a smooth bend under bending stress, while mismatched DNA, because of conformational flexibility, is thought to form a kink at the mismatch site. Some believe that this helps mismatch repair proteins to locate mismatched base pairs.

compare the stability of different mismatches quickly and easily. This led to an initial ranking of the stability of different base pairs, G-C > A-T > G-G ~ G-T ~ G-A > T-T ~ A-A > C-T ~ C-A > C-C.[20] These rankings do vary slightly depending on sequence context, but the trends remain similar.[24] Guanine containing mismatches tend to be the most stable of mismatches, but are destabilizing as compared to the corresponding Watson-Crick base pair. Their stability is attributed to the formation of strong hydrogen bonds and a large aromatic surface area. The C-C mismatch is often the most destabilizing mismatch with weak hydrogen bonding low aromatic surface area. Some have suggested the C-C mismatch can become protonated to form a wobble base pair, but this would involve the burying of a charge within the hydrophobic base stack of DNA. Using this method, standard thermodynamic parameters $\Delta H°$, $\Delta S°$ and $\Delta G°$ of the melting transition can be determined. NMR methods can provide even more data. Using ^{1}H NMR of the DNA imino region at different temperatures it is possible to determine melting temperatures of the individual bases in duplexes.[19]

The combination of melting data with NMR data has allowed the creation of standard tables of $\Delta G°$ parameters of all base pairs in every sequence context, shown in **table 1.1**. If we look at a particular duplex and ask how much can a single mismatch inserted into a duplex destabilize the DNA, the most destabilizing mismatch is a C-C mismatch in a 5'-ACT-3' sequence context, where the C forms a C-C mismatch. This particular mismatch destabilizes the DNA duplex by 2.7 kcal/mol. We can also ask the inverse question, how much can a mismatch stabilize a duplex? The G-G mismatch in a 5'-GGC-3' sequence context stabilizes the duplex by -2.5 kcal/mol. While this is stabilizing overall, this still falls short of a Watson-Crick G-C and A-T base pair which impart a -4.08 kcal/mol or -2.74 kcal/mol stabilization respectively. In all cases, if a mismatch were replaced by a matched base pair it would have a stabilizing effect on the duplex. This is illustrated by $\Delta G°$ parameters for all base pairs shown in sequence

GX/CY	A	C	G	T
A	0.17	0.81	-0.25	-1.3
C	0.47	0.79	-2.24	0.62
G	-0.52	-1.84	-1.11	0.08
T	-1.44	0.98	-0.59	0.45

CX/GY	A	C	G	T
A	0.43	0.75	0.03	-1.5
C	0.79	0.7	-1.84	0.62
G	0.11	-2.17	-0.11	-0.5
T	-1.28	0.4	-0.32	-0.1

AX/TY	A	C	G	T
A	0.61	0.88	0.14	-1
C	0.77	1.33	-1.44	0.64
G	0.02	-1.28	-0.13	0.71
T	-0.88	0.73	0.07	0.69

TX/AY	A	C	G	T
A	0.69	0.92	0.42	-0.6
C	1.33	1.05	-1.3	0.97
G	0.74	-1.45	0.44	0.43
T	-1	0.75	0.34	0.68

Table 1.1: Thermodynamics of different mismatches. ΔG^{o} values(kcal/mol) for different internal mismatches with a neighboring matched base. In all cases the Watson-Crick base pairs, highlighted in yellow, are more stable than any of the possible mismatches. To obtain a value for a mismatch, CC for example, in a sequence context, say 5'-GCG-3' with 3'-CCC-5' you would add 0.79 + 0.70 to obtain a net 1.49 kcal/mol destabilization. X refers to the row while Y refers to the column.

context in **table 1.1**.

1.2.2: Biological sources and consequences of DNA mismatches

1.2.2.1: Natural sources of DNA mismatches in vivo

Mismatches are formed *in vivo* by a variety of processes. DNA synthesis, genetic recombination, cytosine deamination and somatic hypermutation all produce mismatches to greater or lesser extents. As previously mentioned, DNA synthesis is an extremely high fidelity process under normal circumstances, misincorporating a single base in 10^9 bases.[8] More recent work has shown that under nonideal circumstances the misincorporation rate can be much higher.[25] For example, humans have 14 known DNA polymerases.[26] The majority of DNA synthesis occurs during the replication of the human genome by Polδ and/or Polε, very high fidelity polymerases. However, these polymerases are incapable of replication of chemically damaged bases. When these polymerases encounter a damaged base, they dissociate from the DNA and are replaced by a lower fidelity polymerase in a process known as translesion synthesis.[27] Translesion synthesis is performed by Polη in the case of pyrimidine dimers or Polι or Polζ in other cases.[11] These polymerases are more permissive and therefore capable of incorporating a base across from a damaged base, but often an incorrect one. These polymerases also lack 3'→5' proofreading domains which in combination with their permissiveness can misincorporate a base in $10–10^3$ base pairs.[26] Translesion synthesis often only occurs for 100–1000 base pairs after which the high fidelity polymerases continue synthesis.

Genetic recombination is an essential process in meiosis which can also form DNA mismatches. During recombination, homologous chromatids can form a crossover complex.[28] This complex eventually forms a four way junction called a Holliday intermediate. These junctions can "slide" exchanging a single strands between two

homologous chromatids; often several thousand base pairs are exchanged. If the DNA is not exactly identical, mismatches are produced. Eventually an enzyme called a resolvase cuts the four way junction, which may or may not result in recombination, regardless, mismatches which were formed by four-way junction sliding remain. This process accounts for some interesting non-Mendelian genetics, but it is not thought to be a major source of mismatches *in vivo* except in the case of germ cell production.

Cytosine deamination is a natural chemical reaction whereby the minor imine tautomeric form of cytosine is hydrolysed to produce a uracil.[29] This produces a G-U mispair which if left unpaired will result in an A-U and eventually an A-T base pair upon two rounds of DNA replication. The G-U mispair is recognized and the uracil is removed by a base excision repair enzyme uracil DNA glycosylase, UDG.[30] This generates an abasic site in the DNA which is repaired by an AP endonuclease and DNA polymerase.

Intentional somatic mutation only occurs within a single eukaryotic cell type, B lymphocytes. In the production of immunoglobulin several processes are used to increase diversity, one of which is somatic hypermutation, SHM. During SHM variable regions of the immunoglobulin gene undergo mutations at a vastly higher rate of 1 base per 1000 base pairs. While the exact mechanisms leading to SHM are not fully understood it appears that DNA mismatches are a necessary component. It is currently believed that SHM occurs by the action of a activation-induced cytosine deaminase, AID.[29] This enzyme catalyzes the conversion of cytosine to uracil in single stranded DNA. Because of intentional incomplete repair this process can lead to G-U mismatches and abasic sites during DNA replication, which produce all possible mutations upon replication.[31]

1.2.2.2: Fate and consequences of mismatches

In general, DNA mismatches *in vivo* are recognized and repaired by MutS/L/H DNA repair pathway or homologs in higher organisms.[32] This pathway, also known

simply as mismatch repair, MMR, detects a mismatch and replaces the ~1 kb of DNA in the process. This type of DNA repair is present in all organisms from *E. coli* to humans.

In this pathway, shown in **figure 1.5**, a MutS dimer recognizes a mispaired base and binds the DNA. MutS or its homologs are able to recognize most DNA mismatches although some mismatches such as C-C are not detected. A MutL dimer can then bind to the MutS dimer signaling and coordinating a mismatch repair event. This MutS/L complex discriminates parent strand from daughter strand by a poorly understood process and then recruits MutH, or an as yet unknown endonuclease in eukaryotes, to nick the mispaired daughter strand. The nicked strand is digested by exonucleases in either 5'→ 3' or 3'→5' directions towards the mismatched base. Finally, a new DNA strand is synthesized by a DNA polymerase followed by a ligation to seal the patch. In addition to the role of MMR in DNA repair, the associated proteins also can trigger cell cycle arrest or apoptosis.[33]

If mismatches are not repaired before another round of DNA replication, the mismatch will become a permanent mutation. Under normal circumstances this would be a rare event, but if mismatch repair were somehow deactivated this process is significantly enhanced. This MMR deficient phenotype has been observed in numerous types of cancer.[34]

1.3: Recognition of DNA mismatches, enzymes

Recognition of mismatched natural DNA bases occurs primarily by MutS dimer or heterodimer of homologs in higher organisms.[32] However, there is a second specialized protein which recognizes only G-T mismatches formed from cytosine deamination. This protein is the very short patch repair protein Vsr.[35] These two proteins have been studied by a variety of biochemical and structural methods, the results of which show a remarkable contrast in two strategies for mismatch recognition.

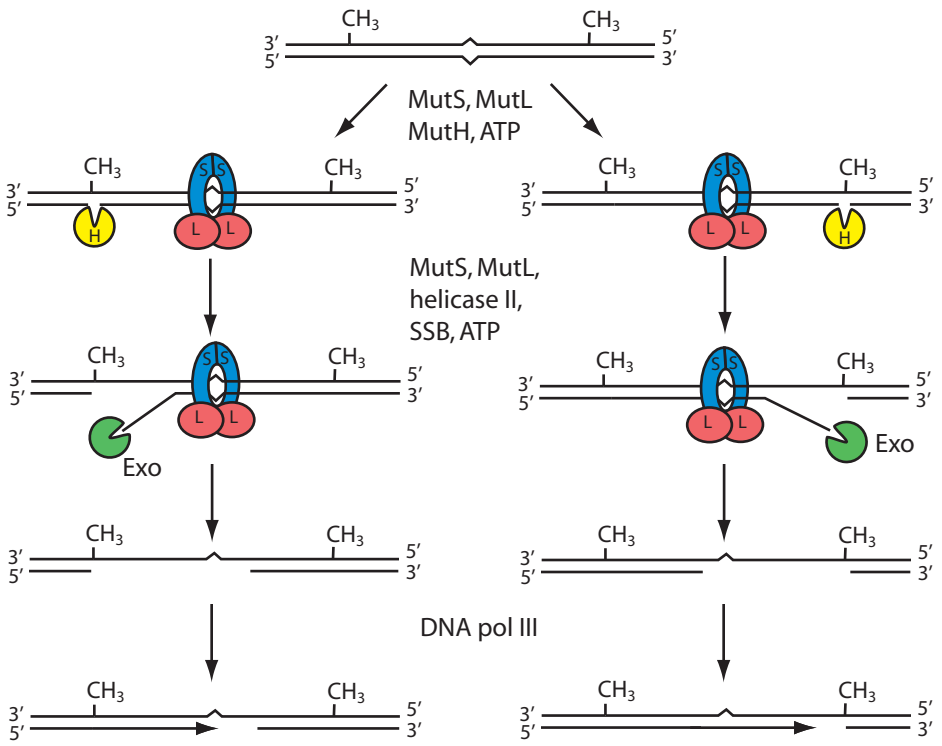

Figure 1.5: Mismatch repair in prokaryotes. A general scheme outlining the major steps in mismatch repair. A similar process is believed to occur in eukaryotes. MutS finds a mismatch base and is bound by MutL. MutL activates MutH to nick the mismatched daughter strand on either the 5' or 3' side of the mismatch. Exonucleases digest the DNA to the mismatched site. DNA polymerase fills in the gap followed by ligation.

MutS is responsible for the initial binding event which is then recognized by other members of the mismatch repair pathway to signal a lesion to repair. DNA-protein interactions have been studied by a variety of biochemical and structural methods. While the method by which mismatches are recognized is not clear, significant progress has been made in this field.

Recently, two structures of the MutS protein bound to mismatched substrate have been obtained, one from *E. coli*[36] and the other from *T. aquaticus*[37]. The structures shows a clamp structure which grabs the DNA and kinks the DNA duplex at a rather severe 60° angle. This kinking of duplex causes the DNA to adopt a structure which shows the typical features of A form DNA. The primary interaction with the DNA is through electrostatic interactions of the DNA with the high positive charge of the interior of the MutS protein. These interactions are supplemented by a phenylalanine which is intercalated in the DNA at the mismatch.[38]

This has led to the theory that MutS finds mismatched bases by forming a sliding clamp which bends the DNA as it slides looking for sites which are especially easy to bend.[23] This theory is supported by NMR data suggesting increased flexibility of the DNA at mismatched sites.[21] This theory however is not supported by the repair activity of MutS.[39] MutS has been found to have high repair activity toward G-T mismatches and almost none against C-C mismatches. The G-T mismatch is among the most stable and would be predicted by this theory to be inefficiently repaired. The C-C mismatch is very destabilized and would be predicted to be repaired very well.

AFM studies were carried out using MutS with matched and mismatched DNA.[40] While these samples were dried onto a mica substrate and therefore not under physiological conditions, an intriguing result was obtained. When matched DNA binds to MutS, it does so in the kinked form. However, when mismatched DNA is bound to MutS, it is in equilibrium between two states, kinked and unkinked. This bent/unbent

equilibrium delicately balances the stabilization from the insertion of the phenylalanine versus the destabilized unkinked protein conformation in the linear DNA form. This gives at least a tenuous explanation of why pyrimidine-pyrimidine mismatches are not recognized. Since there is little or no stabilization from the insertion of the phenylalanine, the equilibrium is placed toward the bent conformation. This unkinked form is proposed to be the ultimate recognition complex to which MutL binds, but this has not yet been confirmed.

These studies also suggest that overall there is little difference in binding affinity of MutS for matched or mismatched DNA. Instead the protein remains bound to the DNA with or without mismatch present. The recognition event is likely not protein binding, but rather the change in protein conformation upon mismatch binding that begins DNA mismatch repair.

The very short patch repair takes a completely different approach to finding mismatches. The Vsr protein only repairs G-T mismatches which are produced from 5-methylcytosine deamination in *E. coli*.[41] 5-methylcytosine is created in some bacteria by the Dcm methylase as part of a phage resistance mechanism involving methylation dependent restriction endonucleases.[42] 5-methylcytosine has the unfortunate consequence that when deamination occurs it becomes a natural DNA base, unlike unmethylated cytosine which will become uracil. Upon recognition of a G-T mismatch, Vsr nicks the DNA on the 5' side of the mismatch. DNA polymerase I binds this nick and performs a nick translation reaction repairing the mismatch. The single strand break is then sealed by a DNA ligase. The cytosine is then methylated by the dcm methylase.

Recently, a crystal structure of the Vsr enzyme bound to DNA was obtained.[41] This shows a structure which is remarkably similar to a type II restriction enzyme. However, the enzyme binds to the mismatched DNA as a monomer. Numerous electrostatic and hydrogen bond contacts are made with the DNA. The mismatch is

discriminated by the insertion of an aromatic wedge composed of two tryptophans and a phenylalanine. This insertion causes a large distortion in the DNA, a gap of 6 Å and a 44° kink forcing the site to be probed into the restriction enzyme catalytic pocket. Here the G-T mismatch is discriminated from the G-C base pair by conformation. If the G-T wobble base pair is bound, its distorted structure will cause the phosphodiester linkage into the endonuclease pocket cleaving the substrate. Remarkably this enzyme is able to perform this discrimination without the use of ATP, inferring that the recognition process is a thermodynamically favorable one, although it is possible that the cleaved enzyme-DNA complex cannot dissociate without assistance. This protein is also an example of damage recognition without extrahelical bases which are present in MutS, uracil DNA glycosylase and MutY.[43] This lack of base flipping may be the reason that this protein is able to recognize a mismatch without external energy.

1.4: Recognition of DNA mismatches, small molecules

Several small molecule methods of mismatch recognition have also been developed. The ways in which these small molecules recognize mismatched DNA vary. Many compounds which bind matched DNA duplexes also will bind mismatched substrates as well. For example the classic intercalator ethidium bromide[44] and the minor groove binding DAPI[45] are capable of binding mismatched as well as matched DNA with similar affinity. For the purposes of this section, only compounds which are capable of recognition of mismatches in preference to matched DNA will be discussed.

1.4.1: Mismatch recognition using polyamides

Polyamides are capable of the recognition of mismatched bases through minor groove interactions.[46] Work within the Dervan group have generated methods to target polyamides to virtually any sequence within DNA. Since the polyamides form

interactions with bases of both strands, one must only insert the proper building blocks to recognize a mismatched base pair. This method depends on multiple interactions however and so can only recognize a particular mismatch within a given sequence.

This method has been successfully applied to the recognition of G-T mismatches within DNA by the polyamide f-ImImIm.[47] This compound binds within the minor groove of DNA as a dimer. Binding of the compound in the minor groove can be observed by DNAse I footprinting[48] and by surface plasmon resonance. Surface plasmon resonance measurements can be used to calculate binding constants of 6×10^6 M^{-1} for the G-T mismatch. This complex has a significantly decreased affinity towards matched DNA (CCGG), 2.1×10^5 M^{-1} or non-cognate sequences (AATT) 6×10^3 M^{-1}.

1.4.2: Mismatch recognition using naphthyridine

Work within the Saito and Nakatani groups have focused on the use of the naphthyridine moiety. Originally, 2-amino-7-methylnaphthyridine was shown to stabilize single guanine bulges.[49] The mechanism proposed for this stabilization was by the intercalation of the naphthyridine moiety across from the bulged base and forming strong hydrogen bonds with the bulged guanine base. The nitrogens of the naphthyridine hydrogen bond with the guanine stabilizing its interaction with the DNA, as shown in **figure 1.6A**. The DNA is stabilized through aromatic stacking of the naphthyridine.

In addition to bulges, dimers of naphthyridine, formed by attaching a alkyl linker between two exocyclic amine groups, are able to recognize a variety of different mismatches. G-G mismatches are readily recognized by an amide linked naphthyridine dimer.[50] The interaction of this dimer with the G-G mismatch is very similar to the single guanine bulge mentioned above. Specific interaction with the bases of the G-G mismatch is observed by DNAse I footprinting assays.[51] The two naphthyridine moieties insert within the DNA base stack and hydrogen bond with the mismatched guanine bases. This

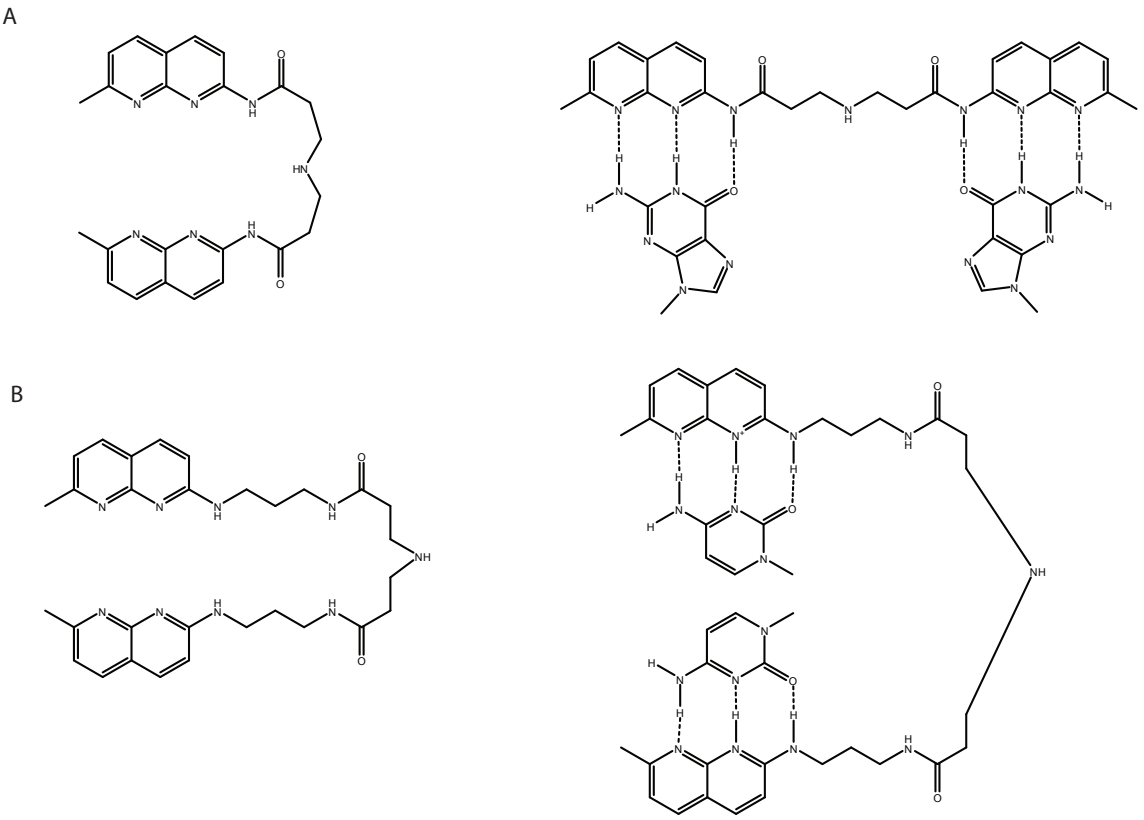

Figure 1.6: Naphthyridine dimer structures and interactions with mismatches. Two different naphthyridine dimers have been constructed which bind mismatches. In A an amide linked structure will bind to G-G mismatches. In B an alkyl linked strcuture will bind to C-C mistmatches because of the shift in pKa associated with the different linker.

interaction can be observed readily using NOESY.[50] For quantitative measurements of binding surface plasmon resonance was employed.[52] Naphthyridine dimer was functionalized to a SPR substrate and mismatched DNA was applied in solution. This method allows for the measurement of both binding rates and dissociation rates from which K_D can be calculated. These complexes bind the G-G mismatch tightly with affinities of 0.1–4 × 10^6 M^{-1}, depending on sequence context.

A modification of the original dimer, linked through an alkyl chain rather than an amide bond was also synthesized, shown in **figure 1.6b**.[53] This compound is able to recognize C-C mismatches in addition to a few others: C-T, C-A, and T-T. This is surprising considering that the recognition element remains the same, but only the linker is changed. This has been explained by conducting experiments while varying pH. It is believed that the recognition of the C-C mismatch is accomplished by protonation of a heterocyclic amine. Affinities for this molecule are also quite high, 1–8 × 10^6 M^{-1}, again varying with sequence context. It is not completely specific for the C-C mismatch; C-T, C-A and T-T mismatches are all bound with decreased affinity. Additionally fully matched duplexes are also recognized with about an 80-fold lower affinity.

1.4.3: Mismatch recognition using metallointercalators

1.4.3.1: Metallointercalator background

Previous work within the Barton group has focused on the use of coordinatively saturated d^6 octahedral metal complexes and their interactions with DNA.[54] These compounds can interact with DNA by a variety of different routes. Compounds such as $[Ru(bpy)_3]Cl_2$ interact only weakly with DNA via electrostatic interactions. Using ligands with a larger aromatic surface area such as $[Ru(phen)_3]Cl_2$ can result in minor groove binding where some aromatic surface area becomes buried within the

hydrophobic DNA minor groove. Finally some complexes such as $[Ru(bpy)_2(dppz)]Cl_2$ or $[Rh(bpy)_2phi]Cl_3$ where a large aromatic ligand extends further from the metal center can interact with DNA via intercalation from the major groove.

Since these complexes are low-spin d^6, they are coordinatively saturated and inert to substitution. This inertness and definite geometry allows for the construction of complexes with a rational positioning of functional groups. In this way we can divide the complexes into several pieces: the intercalating ligand, ancillary ligands and the metal center.

The intercalating ligand, typically a large planar heterocycle, forms the basis for the majority of the affinity the complex has for DNA. The source of this affinity is several-fold. First interactions of the hydrophobic ligand with the hydrophobic interior of the DNA base stack through van der Waals contacts contribute to the stabilization of the interaction. Second, removing the large aromatic surface area from the water causes the release of ordered waters whose increased entropy likely more than offsets the decrease in entropy from binding. Lastly, the interaction with the DNA is strengthened by strong dipole-dipole interactions, through fixed dipoles of the DNA and metal complex.

Ancillary ligands can serve two functions. First, they donate several electron pairs which allow us to form a coordinatively saturated complex. Beyond this, they can simply be innocent bystanders in the interaction with DNA, or can provide functionality which modifies the interaction of the complex with DNA. If we choose a ancillary ligand such as 2,2'-bipyridine, bpy, this ligand has little functionality which can interact with the DNA. On the other hand if we choose a ligand such as ethylene diamine, this ligand provides a hydrogen bond donor to the complex.[55] This ligand produces a favorable hydrogen bond with guanines increasing its affinity toward certain sequences in the DNA. Unfavorable interactions are also possible, in the case of 1,10-phenanthroline, phen, an unfavorable steric clash between the 2 and 9 hydrogens of the phenanthroline ring with

functionalities on the DNA causes this complex to bind to 5'-pyrimidine-pyrimidine-purine-3' sequences.[56]

Finally, the center of the octahedral complex, the metal ion, serves multiple purposes. In general, rhodium (III) is the metal ion of choice, although others have been used such as ruthenium (II), iridium (III), osmium (III) and nickel (II). Each metal ion has its own unique abilities when used to make a metallointercalator. Rhodium (III) complexes with a diimine intercalating ligand, such as phi, can cleave DNA at their intercalated site by irradiation with UV light.[57] The UV excitation causes a ligand to metal charge transfer forming a ligand radical capable of hydrogen atom abstraction at the intercalated site.[58] Ruthenium(II) complexes are typically used for their unique photophysical properties. Tris diimine complexes, such as the light switch compound $[Ru(bpy)_2(dppz)]Cl_3$, when excited by visible light can emit luminescence.[59] These compounds can also be used as photooxidants when an appropriate external quencher is added.[60] Nickel(II) complexes have been used because they are paramagnetic which makes them useful for NMR studies.[61]

An example of designed interactions with DNA is the complex Δ-α-[Rh(R,R-2,9-dimethyl-trien)(phi)]$^{3+}$, shown in **figure 1.7**.[62] The metal of choice for this complex was rhodium(III) for both for photocleavage and substitutional inertness. This complex uses the phi intercalating ligand which provides a basis for the interaction of the complex with the DNA. The ancillary ligand adds many interactions to the complex. The triethylenetetraamine (trien) ligand provides axial amines similar to ethylene diamine, and similarly hydrogen bonds to 5'-GC-3' base steps. To this ligand two methyl groups are added. These methyl groups are designed to interact favorably with the methyl groups of thymine. This makes a compound which is designed to bind to the overall sequence 5'-TGCA-3'.

This compound does bind the designed cognate sequence. Upon UV irradiation

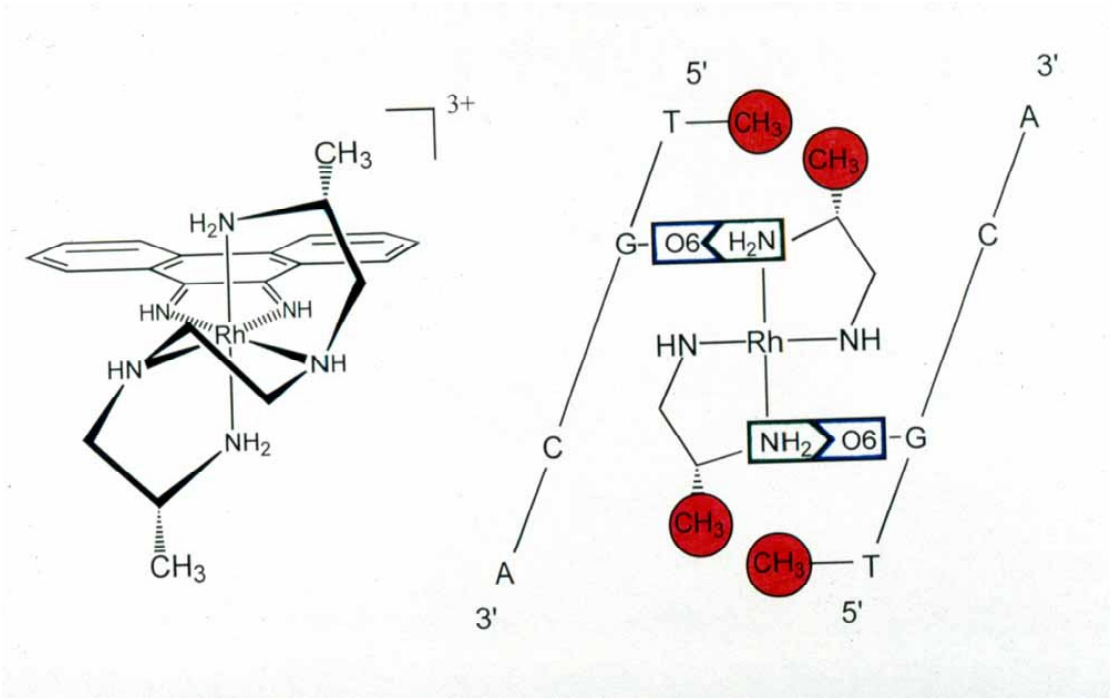

Figure 1.7: Δ-α-[Rh(R,R-dimethyl-trien)(phi)]$^{3+}$. Shown on the left is the complex with the stereochemistry shown. On the right is the complex schematically interacting with the cognate 5'-TGCA-3' sequence. Axial amines hydrogen bond with the O6 of guanine. The methyl groups are shown interacting with the methyl group of thymine.

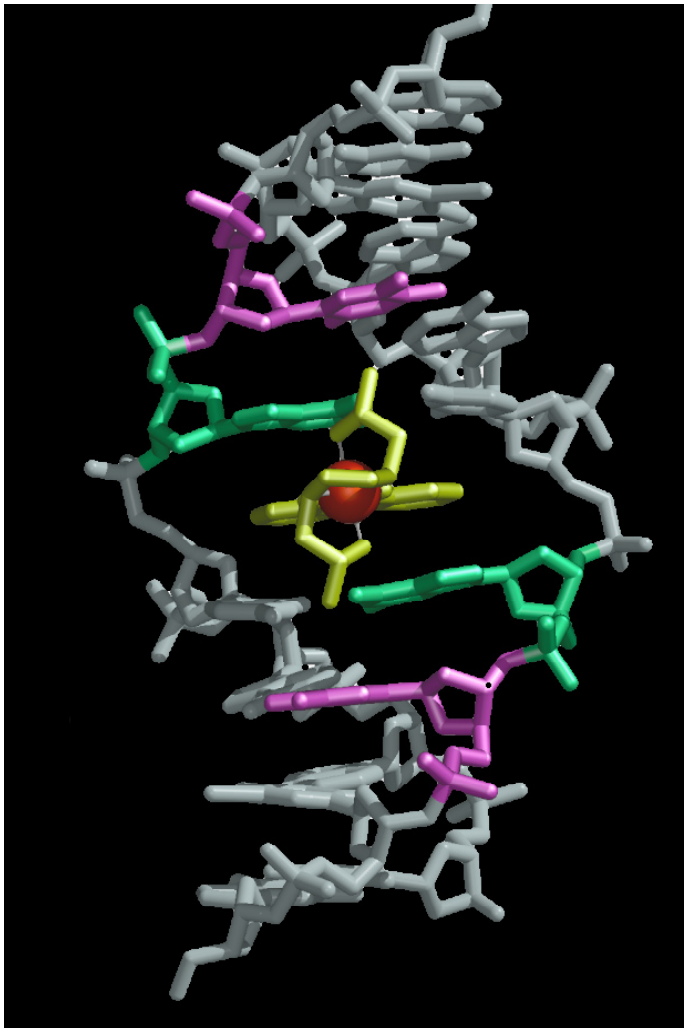

Figure 1.8: Crystal structure of Δ-α-[Rh(R,R-dimethyl-trien)(phi)]$^{3+}$ bound to DNA.

this compound is cleaves at the 5'-TGCA-3' site with moderate photoefficiency. Using a titration it is possible to determine the binding constant of 9×10^7 M^{-1} for this complex to the desired sequence.[63]

This high affinity and specificity were key factors in obtaining both a crystal structure[64] and NMR structure[65] of this complex with its cognate DNA, shown in **figure 1.8**. Using the crystal structure we are able to verify the various designed interactions. Additionally this structure conclusively demonstrates that the phi ligand intercalates from the major groove with an untwisting of the DNA. The phi ligand is buried with almost no free space between the ligand and the sugar rings.

1.4.3.2: Mismatch recognition by metallointercalators

Using the principles outlined above metallointercalators were synthesized which are able to bind DNA mismatches. Similar to the phenanthroline complex described above, mismatches were targeted by creating a ligand which would have an unfavorable interaction at any intercalated site other than the destabilized mismatch site.

This was achieved not by modification of the ancillary ligands, but by modification of the intercalating ligand. As seen from the crystal structure, the phi ligand forms a tight complex with matched DNA. As seen in the space filling model, it is the perfect size for intercalation, filling all available space within the intercalation site. If we were to expand the width of the intercalating ligand, the intercalating ligand would sterically clash with sugar rings. This would be energetically prohibitive for the complex to bind to matched sites. At destabilized sites such as mismatches, intercalation may be possible. As the ligand intercalates steric clash with the sugar rings will be inevitable. If the mismatched bases are significantly destabilized, the loss of hydrogen bonding between the bases may be more than offset by increased aromatic stacking, dipole interactions and removing the ligand from water.

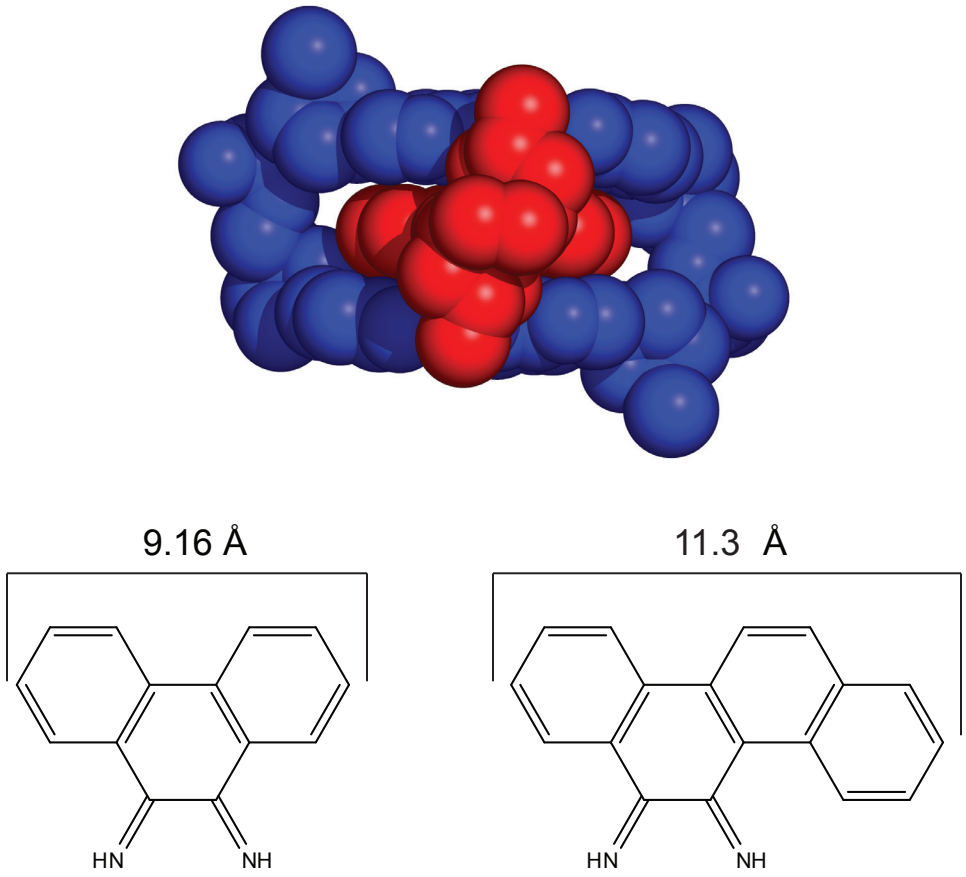

Figure 1.9: Intercalated phi ligand and comparison to chrysi. At the top a view of the intercalated phi ligand from crystal structure data. There is little room for expansion as the phi ligand already makes contact with the oxygen of the sugar ring. At the bottom is a comparison of the widths of the phi ligand and chrysi ligand. The addition of a fourth ring will make the intercalator too large to intercalate at stable sites in the DNA

This design was implemented by modifying the phi intercalating ligand, fusing another ring to form chrysene quinone diimine, chrysi, shown in **figure 1.9**.[66] This ligand was incorporated into a rhodium(III) complex, $[Rh(bpy)_2(chrysi)]Cl_3$, Rhchrysi. This compound was found to bind mismatches with moderate affinity 1×10^5 M^{-1} and high specificity; mismatches are bound ~1000 times tighter than matched base pairs.

A series of experiments were conducted to determine which mismatches are recognized by this complex. In addition to the mismatch itself, base pairs surrounding the mismatch site, the sequence context, of the mismatch were also varied. The sequence context has a profound effect on mismatch stability as shown previously. The conclusion of these studies was that Rhchrysi bound 80% of all possible mismatches. When the cleavage intensity is plotted against mismatch stability, we can see a clear effect of mismatch stability on the binding and cleavage of $[Rh(bpy)_2(chrysi)]Cl_3$.[67]

Additional studies were undertaken to determine just how far this complex could be pressed as far as specificity. For these experiments, two plasmids were created which differed at one site in 2725 base pairs. These plasmids were linearized, denatured and annealed to form a mismatch at the desired site. Using an alkaline agarose gel, cleavage can be seen, at a single site, approximately corresponding to the mismatch site. If this duplex is further restricted and run on a denaturing polyacrylamide gel, we can see that the mismatch site, and only the mismatch site, is cleaved within the whole plasmid, shown in **figure 1.10**.[68]

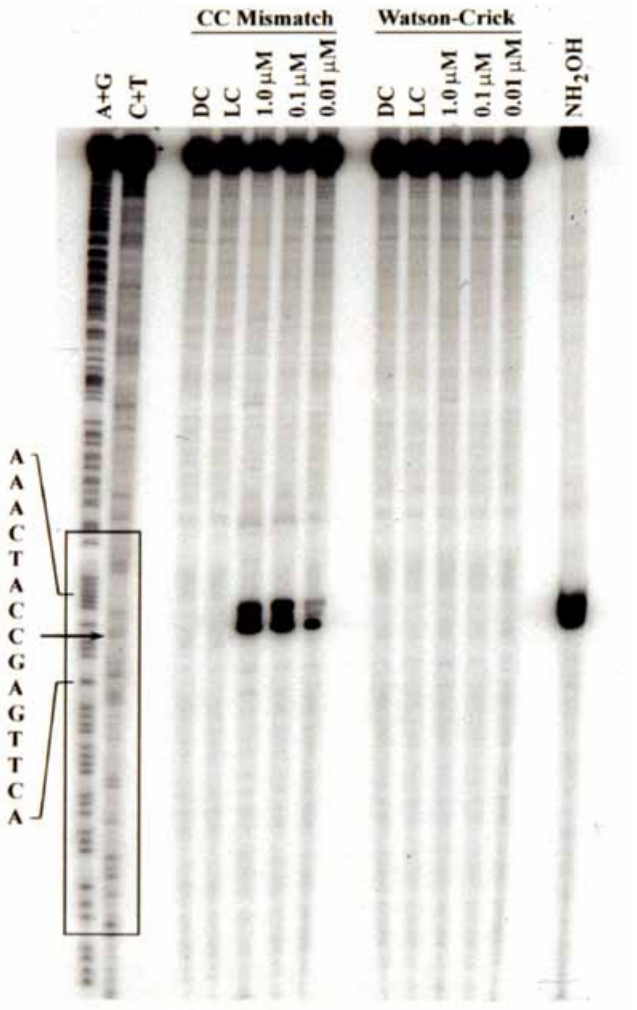

Figure 1.10: Specific cleavage of a single mismatch in a 2725 base pair plasmid. Two plasmids were created which differed by a single G → C base substitution. These were linearized, denatured and annealed to form a C-C mismatch. The mismatch was cleaved and the resulting DNA was separated by denaturing polyacrylamide gel.

1.5: References

1. Griffith, F. (1928). The significance of pneumococcal types. *J. Hyg.* **27**, 113-159.

2. Avery, O. T., MacLeod, C. M., McCarty, M. (1944). Chemical nature of the substance inducing transformation of pneumococcal types. Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus Type III. *J. Exp. Med.* **79**, 137-158.

3. Watson, J. D., Crick, F. H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737-738.

4. Watson, J. D., Crick, F. H. (1953). Genetical implications of the structure of deoxyribonucleic acid. *Nature* **171**, 964-967.

5. Friedberg, E. C. (2006). The eureka enzyme: The discovery of DNA polymerase. *Nat. Rev. Mol. Cell. Biol.* **7**, 143-7.

6. Kunkel, T. A. (2004). DNA replication fidelity. *J. Biol. Chem.* **279**, 16895-16898.

7. Schaaper, R. M. (1993). Base selection, proofreading, and mismatch repair during DNA replication in *Escherichia coli*. *J. Biol. Chem.* **268**, 23762-23765.

8. Drake, J. W., Charlesworth, B., Charlesworth, D., Crow, J. F. (1998). Rates of spontaneous mutation. *Genetics* **148**, 1667-1686.

9. Burrows, C. J., Muller, J. G. (1998). Oxidative nucleobase modifications leading to strand ccission. *Chem. Rev.* **98**, 1109-1152.

10. Russo, M. T., Blasi, M. F., Chiera, F., Fortini, P., Degan, P., Macpherson, P., Furuichi, M., Nakabeppu, Y., Karran, P., Aquilina, G., Bignami, M. (2004). The oxidized deoxynucleoside triphosphate pool is a significant contributor to genetic instability in mismatch repair-deficient cells. *Mol. Cell. Biol.* **24**, 465-474.

11. Prakash, S., Johnson, R. E., Prakash, L. (2005). Eukaryotic translesion synthesis DNA polymerases: Specificity of structure and function. *Annu. Rev. Biochem.* **74**, 317-353.

12. Berti, P. J., McCann, J. A. (2006). Toward a detailed understanding of base excision repair enzymes: transition state and mechanistic analyses of N-glycoside hydrolysis and N-glycoside transfer. *Chem. Rev.* **106**, 506-555.

13. Iyer, R. R., Pluciennik, A., Burdett, V., Modrich, P. L. (2006). DNA mismatch repair: functions and mechanisms. *Chem. Rev.* **106**, 302-323.

14. Brown, T., Hunter, W. N., Kneale, G., Kennard, O. (1986). Molecular structure of the G.A base pair in DNA and its implications for the mechanism of transversion mutations. *Proc. Natl. Acad. Sci. USA* **83**, 2402-2406.

15. Skelly, J. V., Edwards, K. J., Jenkins, T. C., Neidle, S. (1993). Crystal structure of an oligonucleotide duplex containing G.G base pairs: influence of mispairing on DNA backbone conformation. *Proc. Natl. Acad. Sci. USA* **90**, 804-808.

16. Hunter, W. N., Brown, T., Kennard, O. (1987). Structural features and hydration of a dodecamer duplex containing two C.A mispairs. *Nucleic Acids Res.* **15**, 6589-6606.

17. Hunter, W. N., Brown, T., Kneale, G., Anand, N. N., Rabinovich, D., Kennard, O. (1987). The structure of guanosine-thymidine mismatches in B-DNA at 2.5Å resolution. *J. Biol. Chem.* **262**, 9962-9970.

18. Drew, H. R., Wing, R. M., Takano, T., Broka, C., Tanaka, S., Itakura, K., Dickerson, R. E. (1981). Structure of a B-DNA dodecamer: Conformation and dynamics. *Proc. Natl. Acad. Sci. USA* **78**, 2179-2183.

19. Arnold, F. H., Wolk, S., Cruz, P., Tinoco, I., Jr. (1987). Structure, dynamics, and thermodynamics of mismatched DNA oligonucleotide duplexes $d(CCCAGGG)_2$ and $d(CCCTGGG)_2$. *Biochemistry* **26**, 4068-4075.

20. Peyret, N., Seneviratne, P. A., Allawi, H. T., SantaLucia, J., Jr. (1999). Nearest-neighbor thermodynamics and NMR of DNA sequences with internal A.A, C.C, G.G, and T.T mismatches. *Biochemistry* **38**, 3468-3477.

21. Isaacs, R. J., Rayens, W. S. & Spielmann, H. P. (2002). Structural differences in the NOE-derived structure of G-T mismatched DNA relative to normal DNA are correlated with differences in ^{13}C relaxation-based internal dynamics. *J. Mol. Biol.* **319**, 191-207.

22. Isaacs, R. J., Spielmann, H. P. (2004). Insight into G-T mismatch recognition using molecular dynamics with time-averaged restraints derived from NMR spectroscopy. *J. Am. Chem. Soc.* **126**, 583-590.

23. Isaacs, R. J., Spielmann, H. P. (2004). A model for initial DNA lesion recognition by NER and MMR based on local conformational flexibility. *DNA Repair (Amst.)* **3**, 455-464.

24. SantaLucia, J., Jr., Hicks, D. (2004). The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.* **33**, 415-440.

25. Wang, G., Alamuri, P., Humayun, M. Z., Taylor, D. E., Maier, R. J. (2005). The *Helicobacter pylori* MutS protein confers protection from oxidative DNA damage. *Mol. Microbiol.* **58**, 166-176.

26. Kunkel, T. A. (2003). Considering the cancer consequences of altered DNA polymerase function. *Cancer Cell* **3**, 105-110.

27. Lehmann, A. R. (2005). Replication of damaged DNA by translesion synthesis in human cells. *FEBS Lett.* **579**, 873-876.

28. Gerton, J. L., Hawley, R. S. (2005). Homologous chromosome interactions in meiosis: diversity amidst conservation. *Nat. Rev. Genet.* **6**, 477-487.

29. Samaranayake, M., Bujnicki, J. M., Carpenter, M., Bhagwat, A. S. (2006). Evaluation of molecular models for the affinity maturation of antibodies: Roles of cytosine deamination by AID and DNA repair. *Chem. Rev.* **106**, 700-719.

30. Wibley, J. E., Waters, T. R., Haushalter, K., Verdine, G. L., Pearl, L. H. (2003). Structure and specificity of the vertebrate anti-mutator uracil-DNA glycosylase

SMUG1. *Mol. Cell.* **11**, 1647-1659.

31. Seki, M., Gearhart, P. J., Wood, R. D. (2005). DNA polymerases and somatic hypermutation of immunoglobulin genes. *EMBO Rep.* **6**, 1143-1148.

32. Kunkel, T. A., Erie, D. A. (2005). DNA mismatch repair. *Annu. Rev. Biochem.* **74**, 681-710.

33. O'Brien, V., Brown, R. (2006). Signalling cell cycle arrest and cell death through the MMR System. *Carcinogenesis* **27**, 682-692.

34. Arzimanoglou, II, Gilbert, F., Barber, H. R. (1998). Microsatellite instability in human solid tumors. *Cancer* **82**, 1808-1820.

35. Bhagwat, A. S., Lieb, M. (2002). Cooperation and competition in mismatch repair: Very short-patch repair and methyl-directed mismatch repair in *Escherichia coli*. *Mol. Microbiol.* **44**, 1421-1428.

36. Lamers, M. H., Perrakis, A., Enzlin, J. H., Winterwerp, H. H., de Wind, N., Sixma, T. K. (2000). The crystal structure of DNA mismatch repair protein MutS binding to a G-T mismatch. *Nature* **407**, 711-717.

37. Obmolova, G., Ban, C., Hsieh, P., Yang, W. (2000). Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. *Nature* **407**, 703-710.

38. Yamamoto, A., Schofield, M. J., Biswas, I., Hsieh, P. (2000). Requirement for Phe36 for DNA binding and mismatch repair by Escherichia coli MutS protein. *Nucleic Acids Res.* **28**, 3564-3569.

39. Brown, J., Brown, T., Fox, K. R. (2001). Affinity of mismatch-binding protein MutS for heteroduplexes containing different mismatches. *Biochem. J.* **354**, 627-633.

40. Wang, H., Yang, Y., Schofield, M. J., Du, C., Fridman, Y., Lee, S. D., Larson, E. D., Drummond, J. T., Alani, E., Hsieh, P., Erie, D. A. (2003). DNA bending and unbending by MutS govern mismatch recognition and specificity. *Proc. Natl.*

Acad. Sci. USA **100**, 14822-14827.

41. Tsutakawa, S. E., Jingami, H., Morikawa, K. (1999). Recognition of a TG mismatch: The crystal structure of very short patch repair endonuclease in complex with a DNA duplex. *Cell* **99**, 615-623.

42. Lieb, M., Bhagwat, A. S. (1996). Very short patch repair: reducing the cost of cytosine methylation. *Mol. Microbiol.* **20**, 467-473.

43. Morikawa, K., Shirakawa, M. (2000). Three-dimensional structural views of damaged-DNA recognition: T4 endonuclease V, *E. coli* Vsr protein, and human nucleotide excision repair factor XPA. *Mutat. Res.* **460**, 257-275.

44. Hernandez, L. I., Zhong, M., Courtney, S. H., Marky, L. A., Kallenbach, N. R. (1994). Equilibrium analysis of ethidium binding to DNA containing base mismatches and branches. *Biochemistry* **33**, 13140-13146.

45. Trotta, E., Paci, M. (1998). Solution structure of DAPI selectively bound in the minor groove of a DNA T.T mismatch-containing site: NMR and molecular dynamics studies. *Nucleic Acids Res,* **26**, 4706-4713.

46. Dervan, P. B., Edelson, B. S. (2003). Recognition of the DNA minor groove by pyrrole-imidazole polyamides. *Curr. Opin. Struct. Biol.* **13**, 284-299.

47. Lacy, E. R., Cox, K. K., Wilson, W. D., Lee, M. (2002). Recognition of T*G mismatched base pairs in DNA by stacked imidazole-containing polyamides: surface plasmon resonance and circular dichroism studies. *Nucleic Acids Res.* **30**, 1834-1841.

48. Lacy, E. R., Nguyen, B., Le, M., Cox, K. K., C, O. H., Hartley, J. A., Lee, M., Wilson, W. D. (2004). Energetic basis for selective recognition of T*G mismatched base pairs in DNA by imidazole-rich polyamides. *Nucleic Acids Res.* **32**, 2000-2007.

49. Nakatani, K., Sando, S., Saito, I. (2000). Recognition of a single guanine bulge by

2-acylamino-1,8-naphthyridine. *J. Am. Chem. Soc.* **122**, 2172-2177.

50. Nakatani, K., Sando, S., Kumasawa, H., Kikuchi, J. & Saito, I. (2001). Recognition of guanine-guanine mismatches by the dimeric form of 2-amino-1,8-naphthyridine. *J. Am. Chem. Soc.* **123**, 12650-12657.

51. Nakatani, K., Sando, S., Saito, I. (2001). Improved selectivity for the binding of naphthyridine dimer to guanine-guanine mismatch. *Bioorg. Med. Chem.* **9**, 2381-2385.

52. Nakatani, K., Sando, S., Saito, I. (2001). Scanning of guanine-guanine mismatches in DNA by synthetic ligands using surface plasmon resonance. *Nat. Biotechnol.* **19**, 51-55.

53. Kobori, A., Nakatani, K., Saito, I. (2003). Detection of a C-C mismatched base pair by SPR assay using sensors immobilized with a cytosine-specific ligand. *Biochemistry* **42**, 8610.

54. Erkkila, K. E., Odom, D. T., Barton, J. K. (1999). Recognition and reaction of metallointercalators with DNA. *Chem. Rev.* **99**, 2777-2795.

55. Krotz, A. H., Kuo, L. Y., Shields, T. P., Barton, J. K. (1993). DNA recognition by rhodium(III) polyamine intercalators: Considerations of hydrogen-bonding and van der Waals interactions. *J. Am. Chem. Soc.* **115**, 3877-3882.

56. Sitlani, A., Long, E. C., Pyle, A. M., Barton, J. K. (1992). DNA photocleavage by phenanthrenequinone diimine complexes of rhodium(III): Shape-selective recognition and reaction. *J. Am. Chem. Soc.* **114**, 2303-2312.

57. Pyle, A. M., Long, E. C., Barton, J. K. (1989). Shape-selective targeting of DNA by (phenanthrenequinone diimine)rhodium(III) photocleaving agents. *J. Am. Chem. Soc.* **111**, 4520-4522.

58. Turro, C., Hall, D. B., Chen, W., Zuilhof, H., Barton, J. K., Turro, N. J. (1998). Solution photoreactivity of phenanthrenequinone diimine complexes of rhodium

and correlations with DNA photocleavage and photooxidation. *J. Phys. Chem. A* **102**, 5708-5715.

59. Hartshorn, R. M., Barton, J. K. (1992). Novel dipyridophenazine complexes of ruthenium(II): Exploring luminescent reporters of DNA. *J. Am. Chem. Soc.* **114**, 5919-5925.

60. Stemp, E. D. A., Arkin, M. R., Barton, J. K. (1997). Oxidation of guanine in DNA by $[Ru(phen)_2(dppz)]^{3+}$ using the flash-quench technique. *J. Am. Chem. Soc.* **119**, 2921-2925.

61. Rehmann, J. P., Barton, J. K. (1990). ^{1}H NMR studies of tris(phenanthroline) metal-complexes bound to oligonucleotides: Structural characterizations via selective paramagnetic relaxation. *Biochemistry* **29**, 1710-1717.

62. Krotz, A. H., Barton, J. K. (1994). Rational construction of metallointercalators: Syntheses, structures, and characterizations of isomers of ((2s,9s)-diamino-4,7-diazadecane)(9,10-phenanthrenequinone diimine)rhodium(III). *Inorg. Chem.* **33**, 1940-1947.

63. Krotz, A. H., Hudson, B. P., Barton, J. K. (1993). Assembly of DNA recognition elements on an octahedral rhodium intercalator: Predictive recognition of 5'-TGCA-3' by Δ-$[Rh[(R,R)\text{-}Me_2trien]Phi]^{3+}$. *J. Am. Chem. Soc.* **115**, 12577-12578.

64. Kielkopf, C. L., Erkkila, K. E., Hudson, B. P., Barton, J. K., Rees, D. C. (2000). Structure of a photoactive rhodium complex intercalated into DNA. *Nat. Struct. Biol.* **7**, 117-121.

65. Hudson, B. P., Barton, J. K. (1998). Solution structure of a metallointercalator bound site specifically to DNA. *J. Am. Chem. Soc.* **120**, 6877-6888.

66. Jackson, B. A., Barton, J. K. (1997). Recognition of DNA base mismatches by a rhodium intercalator. *J. Am. Chem. Soc.* **119**, 12986-12987.

67. Jackson, B. A., Barton, J. K. (2000). Recognition of base mismatches in DNA by

5,6-chrysenequinone diimine complexes of rhodium(III): A proposed mechanism for preferential binding in destabilized regions of the double helix. *Biochemistry* **39**, 6176-6182.

68. Jackson, B. A., Alekseyev, V. Y., Barton, J. K. (1999). A versatile mismatch recognition agent: Specific cleavage of a plasmid DNA at a single base mispair. *Biochemistry* **38**, 4655-4662.

Chapter 2: Metallointercalators

2.1: Introduction to mismatch binding with metallointercalators

2.1.1: General notes on synthesis

2.1.1.1: Rhodium coordination complexes

The synthesis of rhodium(III) coordination complexes is relatively straightforward as much work has been done in this field previously. In general, the desired complexes fall into one of two categories: plain intercalators and conjugates. The intercalators by themselves typically employ an intercalating ligand and ancillary ligands to suit the purpose. The ancillary ligands may or may not supply additional functionality for DNA binding, but typically they are chosen for their stability and synthetic accessibility. Typical ancillary ligands would include 2,2'-bipyridine, 1,10-phenanthroline, or ammonia.

The synthesis of rhodium(III) complexes typically begins with the $RhCl_3 \cdot xH_2O$ salt. This salt is can be readily converted to a variety of metallointercalators containing one intercalating ligand, and ancillary ligands. Ancillary ligands are installed first with a hydrazine promoted ligand exchange. This leads to a cis-dichloro complex which can be reacted with triflic acid followed by ammonia to form a cis-diammine complex. This cis-diammine can be used in a quinone condensation reaction to form the desired diimine intercalating ligand.

Many ligand exchange reactions of rhodium(III) are promoted by the inclusion of catalytic amounts of hydrazine within the reaction.[1] While the reaction with hydrazine is not fully understood, it is believed that hydrazine, being a two electron reductant, reduces a small amount of rhodium(III) to rhodium(I) which is substitutionally more labile. The rhodium(I) complex undergoes a rapid ligand exchange followed by electron transfer

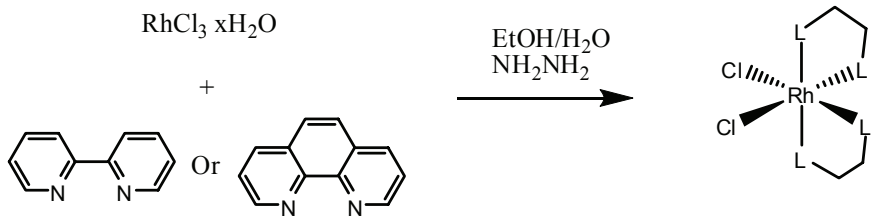

Figure 2.1: Synthesis of rhodium(III) complexes by hydrazine promoted ligand exchange. The addition of hydrazine to the reaction mixture promotes a wide variety of ligand exchange reactions of rhodium(III). This occurs by formation of a catalytic amount of rhodium(I) which can exchange ligands more readily followed by electron transfer to another rhodium(III) center. These additions proceed in a stepwise fashion from mono to bis to tris.

to another metal center. Single electron reductants have been shown to not catalyze ligand exchange even those with similar reducing potentials to hydrazine.[1] This suggests that rhodium(II) is not involved, but it is still possible that it is involved. Additional information has shown that these reactions ligand exchanges proceed in a stepwise reaction.[2] This is important as it allows for the creation of heteroleptic species in a rational manner.

Exchange of chloride ligands is achieved using trifluoromethanesulfonic acid, triflic acid.[3; 4] Rhodium complexes can simply be dissolved in neat triflic acid followed by precipitation with ether and filtration to recover the complex. This method is quick, ~1 hour and efficient, yields are often quantitative, for the removal of chloride ligands and replacing them with the weakly coordinating triflate ligand. This triflate can be replaced by virtually any ligand by simply dissolving the complex and ligand in solution. The major drawback of this method is that triflic acid is extremely acidic and ligands which are labile to acid are obviously incompatible. One alternative to this procedure includes simply boiling the complex in water which will form the labile diaquo complex although the reaction times can be exceedingly long, and yields are often not as good as the triflate reaction.[5]

The triflate complexes are then replaced with ammonia by introduction of concentrated ammonium hydroxide or condensation of anhydrous ammonia gas, using a dry ice bath.[3; 4] Either of these methods can be employed to form amine complexes from the corresponding triflates. These diamine complexes can then be condensed with a quinone to form the desired complexes.

Three ligand complexes will typically need an additional ligand introduced after condensation which is achieved using the aforementioned hydrazine promoted ligand exchange reaction.

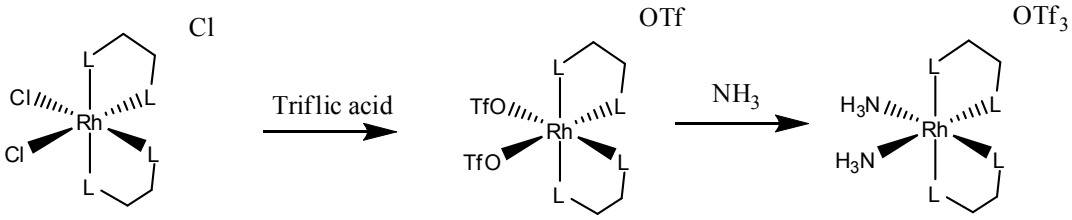

Figure 2.2: Triflic acid for ligand exchange. Triflic acid can be used to remove chloride ligands from rhodium by simply dissolving the complex in neat triflic acid. These triflate ligands can be easily exchanged for a variety of ligands including the cis-diamine desired for o-quinone condensation.

2.1.1.2: Quinone condensation

In rhodium(III) metallointercalators, diimine intercalating ligands are normally used. These intercalating ligands are synthesized by the o-quinone condensation method.[6] This reaction is based upon the extensive work of Sargeson involving intermolecular[7] and intramolecular[8] condensation reactions. Schiff base condensations have been shown to be possible with coordinated ligands with retention of configuration at the metal center.[8; 9] These methods can be used to create a variety of bimetallic species.[10]

Before this method was introduced, these hydrogen imines were typically very difficult to access synthetically. Typically these ligands were synthesized by the coordination of a diamine followed by air oxidation to form a diimine.[11] This method has been employed extensively in the synthesis of phi complexes. The other alternative method is the *in situ* deprotection of a trimethysilylimine.[12] This produces the diimine in solution which is subject to rapid decomposition unless it is trapped by coordination to the metal center. This method was employed in the synthesis of $[Rh(phi)_3]Cl_3$ and $[Ru(phi)_3]Cl_2$. Both of these methods are capable of producing diimine complexes, albeit at low yields and under inconvenient anaerobic conditions.

In the quinone condensation method, a metal bound cis-diamine is deprotonated by a base and attacks the quinone to form a hemiaminal intermediate.[8] This intermediate can then dehydrate to form the desired metal bound imine. In its original, form this reaction was performed in a mixture of acetonitrile and water with sodium hydroxide as the base.[6] Reactions proceed rapidly with 9,10-phenanthrene quinone to form the phi ligand. Additionally the sterically hindered chrysi ligand can also be synthesized by this method although reaction times are increased.

While this system works well for the synthesis of the chrysi ligand, it is not desirable for the synthesis of phzi complexes.[13] Phzi complexes can be synthesized this

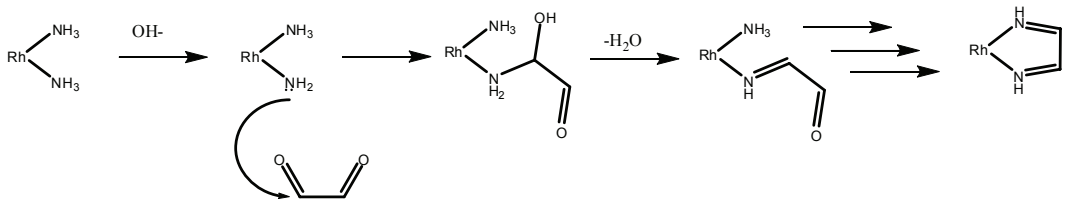

Figure 2.3: O-quinone condensation. A cis diammine complex is deprotonated by base. This coordinated amide can then attack the quinone carbon to form a hemiaminal intermediate. This can then dehydrate to form an imine. This is followed by an intramolecular condensation reaction to form the desired diimine ligand.

way, but are susceptible to hydrolysis in the presence of water. Non-aqueous conditions are desired. Phzi complexes can be synthesized using acetonitrile with a small amount of water added to dissolve the sodium hydroxide. Using this reduced amount of water does significantly reduce the decomposition of the quinone and boost the yield of the diimine complex.

Essentially any metal-bound ammine can be used in this reaction with a few caveats. Not all metal-bound ammines have the same pKa.[14] Rhodium(III) amines have a pKa of ~10 whereas ruthenium(II) amines are much more resistant to deprotonation. The pKa of a bound ammine is also strongly affected by the trans ligand.[15] In the case of $[Rh(ppy)_2(NH_3)_2]^+$ the ammines require much more basic conditions to be deprotonated because of the strongly donating trans carbon donor. A strong donor trans to an ammine also makes the ammine more labile. This can make a condensation reaction unrealistic because the added base will simply replace the bound ammine in some cases. If a strong sterically hindered base is employed, it may be possible to perform quinone condensations in these cases.

It is interesting to note that using a completely water free solvent system is possible. Using acetonitrile as a solvent with triethylamine as a base causes rapid production of chrysi. This is a bit surprising considering that the typical imine formation reaction requires the participation of solvent protons. It is possible that an intramolecular proton transfer occurs which can effect the same transition.

2.1.2: Assessing the DNA binding properties of metallointercalators

2.1.2.1: Direct photocleavage

The simplest method for assessing binding of a metal complex to a particular site is UV excited photocleavage.[16] Most rhodium diimine complexes are capable

of hydrogen atom abstraction from their excited states.[17] This direct hydrogen atom abstraction occurs without diffusion, marking the site where the complex is bound. This cleavage of the DNA can be analyzed by denaturing polyacrylamide gel electrophoresis, separating the DNA based upon molecular weight and charge. The sites can be determined by comparison to the Maxam-Gilbert sequencing reactions. Cleavage intensities can be determined by photography or phosphorimagery.

Nonspecific complexes, such as $[Rh(bpy)(phi)_2]Cl_3$, show cleavage at most bases.[16] A series of cleavage reactions is performed varying the concentration of the metallointercalator with a constant concentration of DNA. These complexes can be characterized by a non-specific DNA binding constant. This can be determined by the quantitation of the parent uncleaved DNA versus the intergrated intensity of the entire lane. This fraction cleaved is plotted versus complex concentration and fit to a sigmoidal curve. The inflection point of the graph can then be used to determine the binding constant of the complex

Specific complexes such a $[Rh(bpy)_2(chrysi)]Cl_3$ have one preferred binding site and several undesirable sites.[18] Experiments can be performed like those for nonspecific complexes although nonspecific binding may complicate analysis of binding. An alternative method is to hold the ratio of complex to DNA constant and increase both simultaneously. This type of experiment allows the specific binding constant to be determined without the complication of nonspecific binding. Fraction cleaved is determined and plotted versus increasing concentration. The data are plotted and fit to a sigmoidal curve. The inflection point is again used to determine the binding constant via a slightly different formula. Non-specific binding constants are determined in a method identical to that for a nonspecific complex.

2.1.2.2: Photofootprinting

If the complex is not proficient at direct photocleavage, binding of the complex to DNA can still be assessed by footprinting. Typically weak binding footprinting agents are used for these reactions such as $[Fe(EDTA)]^{2-}$. This allows for easy determination of binding constants by footprinting.

Footprinting using strong binding agents is also possible, although it introduces another equilibrium into the reaction. Footprinting using strong binding nonspecific complexes can be used for qualitative analysis of binding, but cannot provide a quantitative measure of binding unless the specific binding constants for the complex at every site in a duplex are determined. This may be possible for a very short duplex, but in general the analysis required would be prohibitive.

The binding of several complexes has been assessed through these footprinting competition experiments. A regime is chosen where the footprinting complex is tightly bound to its DNA. Then varying concentrations of the competing complex are added. This will cause the footprinting complex to become dissociated from the site and the cleavage of the DNA is reduced. This reduction in cleavage is measured with phosphorimagery, graphed and apparent binding constant measured. This value is then used to compute the correct binding constant. This method cannot lead to an accurate determination of the binding constant when the unknown complex is a tighter binder than footprinting complex.

2.2: $[Rh(bpy)_2(phzi)]Cl_3$, a high affinity binder of DNA mismatches

2.2.1: Introduction

The targeting of single base mismatches in DNA represents a challenging problem in molecular recognition. The mismatched site may be of variable sequence and within a variable sequence context. What distinguishes such a site is instead the local destabilization in opposing bases because of the absence of Watson-Crick hydrogen bonding.[19] While challenging, however, the development of small molecules targeted to mismatches offers many applications. Site-specific mismatch probes could be used in discovery efforts to identify single nucleotide polymorphisms. Moreover, such molecules could provide the basis for the development of novel chemotherapeutics. Many cancers are associated with a deficiency in mismatch-repair.[20; 21] Hence, by directing small molecules to the accumulated mismatches, a cancer-specific targeting strategy could be envisioned.

In our laboratory, we have exploited the local helix destabilization associated with mispairing in the design of a metal complex targeted to mismatches.[18; 22] Shown in **figure 2.5** is a rhodium complex containing the benzo[a]phenazine quinone diimine (phzi) ligand as well as chrysene quinone diimine (chrysi) and phenanthrene quinone diimine (phi) ligands. Octahedral rhodium(III) complexes containing the phi ligand bind avidly to double helical DNA by intercalation[23]; upon photoactivation, direct DNA strand cleavage is also promoted at the bound site.[16] By tuning the ancillary ligands, complexes can be prepared that are either site specific or sequence neutral. $[Rh(bpy)phi_2]^{3+}$, for example, binds to B-DNA with little site-selectivity[16], while

---

Adapted from Junicke, H., Hart, J. R., Kisko, J., Glebov, O., Kirsch, I. R. & Barton, J. K. (2003). A rhodium(III) complex for high-affinity DNA base-pair mismatch recognition. *Proc Natl Acad Sci U S A* **100**, 3737-3742.

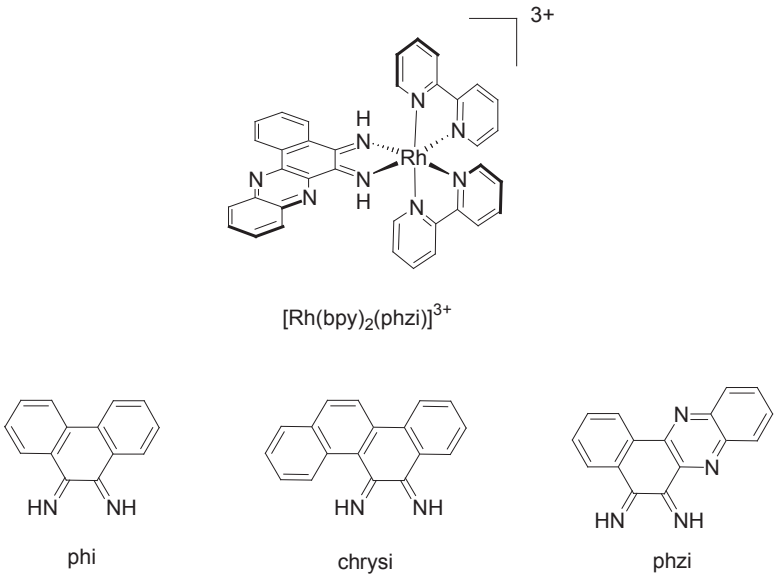

Figure 2.4: Bulky metallointercalators. Shown above is $[Rh(bpy)_2(phzi)]^{3+}$, below, the phi, chrysi, and phzi intercalating ligands. The phzi ligand is of a similar width to the chrysi ligand, but the inclusion of heterocyclic nitrogens increases affinity to mismatched DNA.

[Rh(S,S-dimethyltrien)(phi)]$^{3+}$ specifically targets the sequence 5'-TGCA-3'.[24] The high resolution crystal structure of [Rh(S,S-dimethyltrien)(phi)]$^{3+}$ bound to a DNA octamer shows site-specific intercalation by the complex from the major groove side of the helix, with site-discrimination determined through an ensemble of hydrogen bonding contacts and methyl-methyl interactions.[25] The intercalated phi ligand itself resembles another base pair, stacked at a distance of 3.4 Å between neighboring base pairs, and with an expanse across the ligand of dimensions matching that of the base pairs above and below.

The chrysi complex represents our first generation complex targeted to mismatches.[18] The site specificity of the complex is derived from the fact that it is a more bulky intercalator with an expanse exceeding that of the well-matched base pair. Owing to shape selection, the complex is unable to bind to well-matched B-form DNA. However, at sites where a base pair is destabilized, intercalation of the chrysi complex can occur. [Rh(bpy)$_2$chrysi]$^{3+}$ has been shown to be both a general and remarkably specific mismatch recognition agent.[26; 27] Specific DNA cleavage is observed at over 80% of mismatch sites in all the possible single base pair sequence contexts around the mispaired bases. Moreover, the complex was found to recognize and photocleave at a single base mismatch in a 2725 base pair linearized plasmid heteroduplex.

Here we describe the synthesis and application of a second generation mismatch recognition agent, [Rh(bpy)$_2$phzi]$^{3+}$. This complex shares with the chrysi complex a sterically demanding intercalating ligand. However, the phzi ligand, as an aromatic heterocycle, offers the possibility of greater stabilization through stacking within the mismatched site. Indeed, [Rh(bpy)$_2$phzi]$^{3+}$ binds mismatched sites with a binding affinity increased by two orders of magnitude without a loss in site-selectivity. This high affinity allows the application of the complex in the differential targeting of DNA in a mismatch-repair deficient cell line.

2.2.2: Synthesis and experimental

Materials. Commercially obtained chemicals were used as received. Bipyridine, triflic acid, 2,3-dichloro-1,4-napthoquinone and phenylenediamine were purchased from Aldrich. $RhCl_3 \bullet 2H_2O$ was obtained from Pressure Chemical.

Instrumentation. Electronic spectra were recorded on a Beckman DU 7400 UV-visible spectrophotometer. Mass spectra (ESI) were measured on an LCQ mass spectrometer. High-performance liquid chromatography was performed on a Waters 996 system using a Vydac C_{18} column (4.0 mL/min liquid phase, linear gradient over 50 min from 100% 50 mM NH_4OAc, pH 7.0 to 100% acetonitrile). NMR spectra were recorded on a Varian Mercury 300 MHz instrument. DNA synthesis was performed on an ABI 392 DNA/RNA synthesizer using reagents from Glen Research. DNA was purified using Glen Research Poly-Pak II cartridges followed by reverse phase HPLC. Photocleavage reactions were carried out using a 1000 W Oriel Hg/Xe arc lamp with a monochromator fitted with a 300 nm cutoff filter and an IR filter.

Synthesis of Benzo[a]phenazinequinone. This compound has previously been synthesized.[28] The synthetic procedure used is elaborated here. The quinone is synthesized through an electophilic aromatic substitution followed by an imine condensation. This intermediate is then oxidized by nitric acid to the desired quinone. 2,3-dichloro-1,4-naphthaquinone (9 g, 40 mmol) and 1,2-phenylenediamine (4 g, 37 mmol) were dissolved in 200 mL pyridine. The solution was heated with a boiling water bath for 1 hour. The dark red solution was filtered and the solid was collected. The solid was dissolved in 100 mL of 100 ºC pyridine and filtered again. The solution was cooled to -20 ºC and the zwitterionic intermediate, 6-pyridinium-benzo[a]phenazin-5-olate, was isolated by filtration. The collected solid was freed from excess pyridine by placing under high vacuum for several days, although the oxidation will proceed without

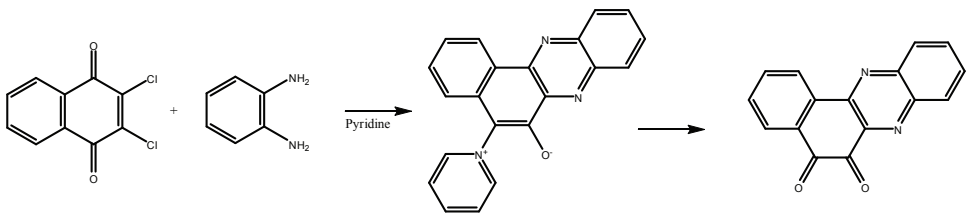

Figure 2.5: Synthesis of phzi quinone. 2,3-dichloro-naphtho-1,4-quinone is reacted with o-phenylene diamine in pyridine to produce a zwitterionic species. This species can then be oxidized by nitric acid under acidic conditions to form the desired phzi quinone.

interference if some pyridine remains.

The intermediate was oxidized to the desired benzo[a]phenazine-5,6-quinone by oxidation with nitric acid. 1 g of the 6-pyridinium-benzo[a]phenazin-5-olate was dissolved in 35 mL glacial acetic acid, 4 mL nitric acid, and 2.3 mL of water. The red brown solution was heated to 80 ºC for 1 hour. The product was precipitated by pouring of the solution into 100 mL of cold water. The precipitate was filtered from solution and washed 3 times with ethanol and 3 times with ether to yield a yellow powder with a slight green tinge. ^{1}H NMR ($CDCl_3$, 300MHz) 8.78ppm (d, 1H, 7.9Hz), 8.24ppm (d, 1H, 8.6Hz), 8.21ppm (d, 1H, 10Hz), 8.15ppm (d, 1H, 7.4Hz), 8.01ppm (t, 1H, 6.8Hz), 7.94ppm (t, 1H, 6.4Hz), 7.90ppm (t, 1H, 7.8Hz), 7.73ppm (t, 1H, 7.6Hz)

Synthesis and Characterization. *Rac*-[Rh(bpy)$_2$phzi]Cl$_3$, (bis(2,2'-bipyridine)benzo[a]phenazine diimine rhodium(III) chloride). *Rac*-[Rh(bpy)$_2$(phzi)]Cl$_3$ was synthesized using the o-quinone condensation method previously employed in the synthesis of [Rh(bpy)$_2$(chrysi)]Cl$_3$.[6] [Rh(bpy)$_2$(NH$_3$)$_2$]OTf$_3$ (100 mg, 0.1 mmol) and benzo[a]phenazinequinone (35 mg, 0.125 mmol), were prepared according to literature procedure.[5; 28] These were dissolved in dry acetonitrile (50 mL). Aqueous sodium hydroxide (2 mL, 0.2 M) was added and the solution was rapidly stirred for 30 minutes at ambient temperature. The reaction was quenched with the addition of 35 mL water and 0.2 N HCl to neutralize. The acetonitrile was removed *in vacuo*. The remaining solution was filtered from the brown solid and loaded on a CM-C25 Sephadex cation exchange column (resin equilibrated with 0.05 M $MgCl_2$). A yellow organic ligand passes right through. After rinsing with water, 0.05 M $MgCl_2$, a green fraction elutes which was collected and concentrated on a Seppac-C18 column and eluted with a mixture of acetonitrile/water/TFA (1:1:0.001) and lyophilized to dryness. Yield: 23mg (30% yield). Microanalysis found (calc): C 54.8 (55.4), H 3.6 (3.3), N 15.3 (14.8). ^{1}H NMR (d6-DMSO, 300 MHz): δ 14.88 (s), 14.70 (s), 9.03 (m, 4H), 8.90 (d, 2H, 7.2 Hz), 8.72 (d, 1H,

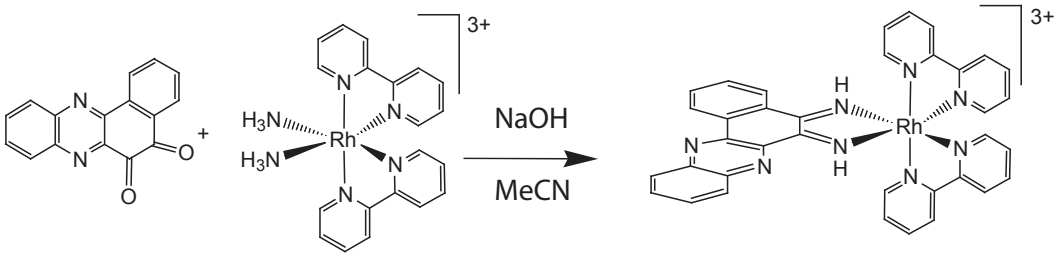

Figure 2.6: Synthetic scheme for *rac*-$[Rh(bpy)_2(phzi)]^{3+}$ using the condensation strategy. Phzi quinone and $[Rh(bpy)_2(NH_3)_2]^{3+}$ are dissolved in dry acetonitrile. The reaction proceeds with the addition of several equivalents of sodium hydroxide to produce the desired rhodium complex. The complex is purified by ion exchange chromatography.

6.1 Hz), 8.60 (t, 2H, 8.5 Hz), 8.54 (d, 1H, 6.1 Hz), 8.47 (t, 2H, 8.5 Hz), 8.32 (d, 1H, 8.5 Hz), 8.20 (d, 1H, 8.5 Hz), 8.11 (t, 1H, 7.3 Hz), 8.03 (m, 3H), 7.94 (t, 1H, 6.1Hz), 7.84 (t, 1H, 7.3Hz), 7.75 (m, 3H), 7.69 (d, 1H, 6.1 Hz) ppm. UV/vis (H_2O, pH 5): 245 nm, (89,800 $M^{-1}cm^{-1}$); 304 nm (65,800 $M^{-1}cm^{-1}$), 314 nm (67,300 $M^{-1}cm^{-1}$), 343 nm (39,300 $M^{-1}cm^{-1}$). ESI-MS(cation) 671($M-2H^+$) observed, 671 calc.

DNA Preparation and Photocleavage Experiments. The oligonucleotides, 5'-GAT GTC GGT C*C*C ACG ATG GTG ACG GAT TAC C-3' and 5'-GAG TTG GTA ATC CGT CAC CAT CGT G*C*G ACC GAC ATC ATG CG-3', where *C* denotes the postion of the mismatch, were synthesized on an ABI 392 DNA/RNA synthesizer, using standard phosphoramidite solid-phase synthesis, and initially purified on Poly-Pak II cartridges and further purified by HPLC (98% 100 mM NH_4OAc/2% acetonitrile to 70% 100 mM NH_4OAc/30% acetonitrile over 30 minutes). The single strands were then 5' labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase. The labeled strands were further purified by gel electrophoresis (20% denaturing polyacrylamide gel), eluted from the gel via soaking in TEAA, ethanol precipitated, and annealed in the presence of unlabeled DNA. To the labeled 31/41-mer duplex (2 μM in 50 mM NaCl, 10 mM TrisHCl, pH 8.5) was added either $[Rh(bpy)_2phzi]Cl_3$ or $[Rh(bpy)_2chrysi]Cl_3$, and the sample was irradiated. After irradiation, all samples were lyophilized, denaturing formamide loading dye was added, and the samples were electrophoresed on a 20% polyacrylamide denaturing gel. The photocleavage results were quantitated via phosphorimagery (Molecular Dynamics Phosphorimager).

2.2.3: DNA photocleavage of a single base mismatch

As is evident in **figure 2.8**, photoactivation of $[Rh(bpy)_2phzi]^{3+}$ bound to a mismatched site leads to direct strand cleavage. Upon photoactivation of

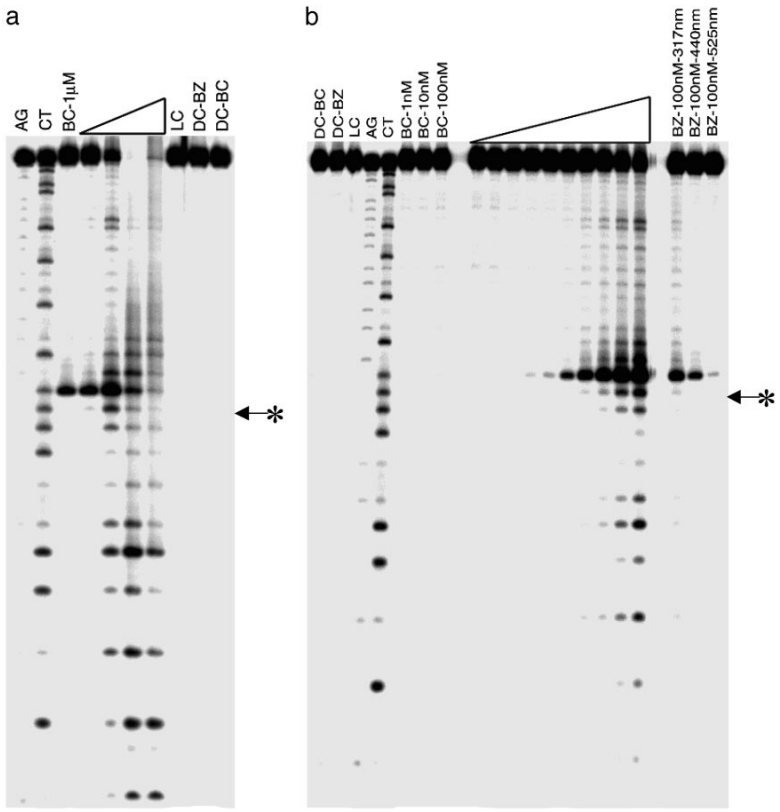

Figure 2.7: Photocleavage of mismatched DNA with Rhphzi. Phosphoimagery of 20% polyacrylamide gels shows mismatch-specific cleavage by the rhodium(III) complex (a) after DNA photocleavage by 1 μM $[Rh(bpy)_2chrysi]Cl_3$ (BC) and 0.1 to 100 μM $[Rh(bpy)_2phzi]Cl_3$ (BZ) on [5'-32P]-labeled CC mismatch-containing, indicated by *, 31/41-mer DNA duplex (1 μM). All samples were prepared in 10 mM Tris pH 8.0, 50 mM NaCl. All irradiations were at 313 nm for 15 min. (b) after DNA photocleavage by varying concentration of $[Rh(bpy)_2chrysi]Cl_3$ (BC) and $[Rh(bpy)_2phzi]Cl_3$ (BZ) as well as at different wavelengths for $[Rh(bpy)_2phzi]Cl_3$. It is notable, that BC does not cleave the DNA at concentrations up to 100 nM whereas BZ already shows a cleavage at 10 nM at the mismatch site. Lanes labeled AG and CT refer to Maxam-Gilbert sequencing reactions. DC refers to the dark control DNA sample without irradiation, and LC refers to the light control sample irradiated in the absence of metal complex.

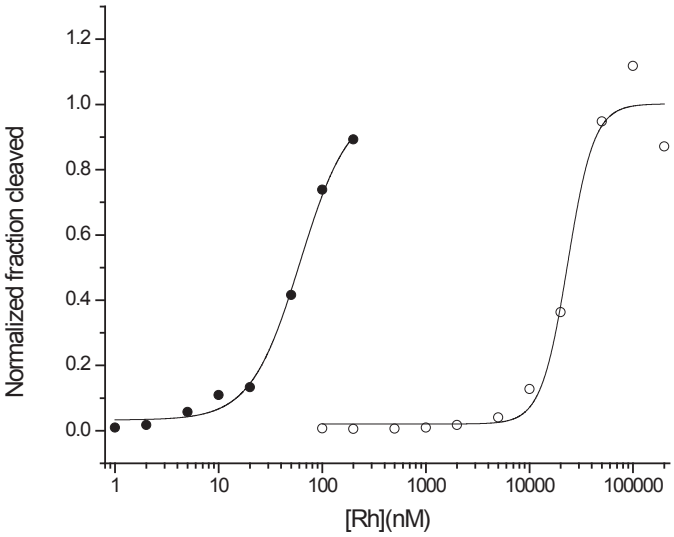

Figure 2.8: Binding of Rhphzi to mismatched and matched DNA. Binding isotherms for $[Rh(bpy)_2phzi]^{3+}$ targeted to a C-C mismatch-containing oligonucleotide (closed circles) and the highest affinity well-matched (open circle) site in the same duplex. Photocleavage reactions were performed in 50 mM NaCl, 10 mM Tris·Cl, pH 8.5 buffer with irradiation 313 nm radiation for 15 min. The concentration of the oligomer varied from 1×10^{-10} to 4×10^{-4} M with the rhodium complex at 10-fold lower concentration. Samples were eluted through 20% denaturing polyacrylamide gels and the data analyzed using a Molecular Dynamics Phosphorimager and ImageQuant software. Cleavage is observed at the base 3' to the C-C mismatch on the shorter strand. For comparison, the specific binding affinity of the corresponding chrysi complex to this site is 3.1×10^{-5} M^{-1}.

$[Rh(bpy)_2phzi]^{3+}$ in the presence of a mismatch-containing DNA duplex, site-specific photocleavage is observed at the base 3' to the mismatch site. Both the position of strand cleavage and the level of specificity are in accordance with that seen using the chrysi complex. Significantly, however, specific cleavage of the mismatch site by $[Rh(bpy)_2phzi]^{3+}$ can be seen at metal concentrations as low as 10 nM.

Titrations as a function of concentration, where the ratio of metal/DNA is kept constant, can be used to establish the specific binding constant to an individual site. **Figure 2.9** shows the quantitation for such a photocleavage titration. The specific binding constant for $[Rh(bpy)_2phzi]^{3+}$ at the C-C mismatch in this oligonucleotide is 1×10^7 M^{-1}. This compares to a value of 3×10^5 M^{-1} for $[Rh(bpy)_2chrysi]^{3+}$ under comparable conditions.

Not only does $[Rh(bpy)_2phzi]^{3+}$ bind more tightly to mismatched sites, photoactivation experiments reveal also that the complex has approximately a five-fold higher photoefficiency than the chrysi analogue in cleaving the C-C mismatch. The last three lanes in **figure 2.8** also show the photoefficiency for the complex at different wavelengths. Remarkably, the cleavage can be seen even at visible wavelengths of 525 nm.

Site-specific binding by $[Rh(bpy)_2phzi]^{3+}$ can also be compared quantitatively to non-specific binding to well-matched sites on the duplex (**figure 2.9**). For this oligonucleotide, the highest cleavage intensity other than that for the C-C site is evident within the duplex region 5'-GT**C**GG-3'. The determination of the binding constant for this site yielded a value of 4×10^5 M^{-1}, almost two orders of magnitude weaker than that for the mismatch. This represents the highest well-matched site affinity, so that the average non-specific affinity is weaker. This level of specificity is comparable to that seen with the chrsyi complex.[18]

$[Rh(bpy)_2phzi]^{3+}$ targets not only the C-C mismatch but other mismatches as

well. The full family of mismatches was examined through photocleavage experiments on a 31 base-pair oligomer by varying the base sequence at the central mismatch site. The pyrimidine mismatches in particular, including C-T, T-T, and T-C, are also readily cleaved. By quantitative photocleavage titration, the binding affinity for the C-A mismatch is found to be 3×10^6 M^{-1}, while that for C-T is 6×10^7 M^{-1}. Overall, targeting appears not to be base specific, but mismatch specific. As with the parent chrysi complex, then, site specificity depends upon the local helix destabilization associated with mispairing.

2.2.4: Photocleavage efficiency

Also of interest, while $[Rh(bpy)_2phzi]^{3+}$ has the ability to target a variety of different mismatches with high affinity, it is not able to cleave all mismatches with equal photoefficiency. This may determined through quantitative photocleavage titration; at maximum binding, different intensities of photocleavage are observed. The different photoefficiencies can be understood in terms of variations in bound geometries within the different mismatch site, since photocleavage is expected to depend upon access of the photoexcited ligand radical to hydrogen atoms on the sugar.[16] Several mismatches, such as the C-C and T-T, are cleaved with much better efficiency than $[Rh(bpy)_2(chrysi)]^{3+}$; others such as C-A are bound tightly but do not exhibit efficient cleavage.

2.3: $[Rh(bqdi)_2(chrysi)]^{3+}$, a binder of G-G DNA mismatches

2.3.1: Introduction

While $[Rh(bpy)_2(phzi)]Cl_3$ is a higher affinity complex, it retains similar mismatch recognition properties to the $[Rh(bpy)_2(chrysi)]Cl_3$ complex. It was desired to synthesize a complex which would be able to recognize mismatches which were not recognized by the parent complexes. In general, these mismatches are the guanine containing

mismatches G-G, G-A, and G-T.

The strategy for the creation of this new complex was simple. Guanine containing mismatches are not stabilizing in comparison to Watson-Crick base pairs, but are simply more stable than other mismatches.[29] If the affinity of a complex could be boosted slightly perhaps it would be possible to target these guanine containing mismatches.

Previous work within our group has demonstrated the ability of coordinated amines in the recognition of guanine bases. Several complexes have been synthesized which demonstrate this increased affinity for guanine rich sites.[30] In these complexes, amines coordinated in the axial position of a metallointercalator hydrogen bond with the O6 of the guanine base. This hydrogen bonding is readily seen in the crystal structure of $[Rh(dimethyl\text{-}trien)(phi)]Cl_3$.[25]

The goal then is to add such functionality to one of our mismatch intercalators to see if the increase in binding affinity will be sufficient to bind guanine containing mismatches. While this may seem simple enough there are two complicating factors. First, the binding of $[Rh(bpy)_2(chrysi)]Cl_3$ to DNA has never been adequately described. While intercalation from the major groove would seem to be the obvious binding mode, there is little evidence for or against this. Additionally the location of the mismatched bases is also very important. If these bases are in an extrahelical conformation upon binding, they will not be available for hydrogen bonding.

The other complication is the photoefficiency of rhodium amine complexes. Upon UV excitation of a rhodium intercalator, a transient species capable of hydrogen atom extraction is formed.[17] It has been previously shown that when ancillary ligands are changed from bipyridine or phenanthroline to amines the yield of photocleavage is significantly reduced. For this reason, the 1,2-benzoquinone diimine, bqdi, ligand was also employed.[31] This ligand has been shown to not be an intercalator of DNA and provides an axial imine moiety which, similar to an amine, should be capable of

hydrogen bonding with the O6 of guanine. Additionally it was believed that these bqdi complexes would show a higher photoefficiency for photocleavage as compared to ammine complexes. However these complexes do not show efficient photocleavage, but binding to G-G mismatches can still be observed.

2.3.2: Experimental

Synthesis of $[Rh(NH_3)_4(chrysi)]Cl_3$. $[Rh(NH_3)_6]OTf_3$ was prepared by the method of Sargeson as described previously.[3] $[Rh(NH_3)_6]OTf_3$, 100 mg, was reacted with 39 mg chrysene quinone in 50 mL acetonitrile, 50 mL water with 0.2 mL 1 M sodium hydroxide as a catalyst. Acetonitrile is removed under vacuum followed by filtration, which removes the unreacted chrysi quinone. The desired complex is separated from unreacted $[Rh(NH_3)_6]OTf_3$ by Seppac C_{18} cartridge and eluted with 1:1:0.001 acetonitrile: water:TFA. This forms the desired $[Rh(NH_3)_4(chrysi)]OTf_3$ complex in 65% yield, 85 mg. The NMR of this species is complex at room temperature due to slowly interconverting hydrogen bonded structures. At higher temperatures these interconvert more rapidly, for this reason NMR data of this complex is taken in DMSO at 50 ºC. ^1H NMR (d6-DMSO, 300 MHz): δ 13.30 (s), 12.32 (s), 8.876 (t, 1H, 7.7Hz), 8.787 (d, 1H, 7.9Hz), 8.57-8.51 (m,2H), 8.358 (dd, 1H, 8.9Hz, 4.6Hz), 8.145 (d, 1H, 7.7Hz), 7.85-7.70 (m, 4H), 4.73-4.54 (broad m, 6H), 3.862 (s, 3H), 3.636 (s, 3H) ppm. UV/vis (H_2O, pH 5): 263 nm, (60,900 $M^{-1}cm^{-1}$); 283 nm (38,100 $M^{-1}cm^{-1}$), 326 nm (12,600 $M^{-1}cm^{-1}$), 413 nm (12,000 $M^{-1}cm^{-1}$). MALDI-MS(cation) 425m/z ($M-2H^+$) observed, 427m/z calc.

Synthesis of $[Rh(bqdi)_2(chrysi)]Cl_3$. $[Rh(NH_3)_4(chrysi)]OTf_3$, 25 mg, with 100 mg 1,2-phenylene diamine in a simple displacement reaction in 100 mL water with a crystal of hydrazine hydrochloride as a catalyst. This solution was allowed to stir under air atmosphere for 3 days during which time a reaction occurs to produce

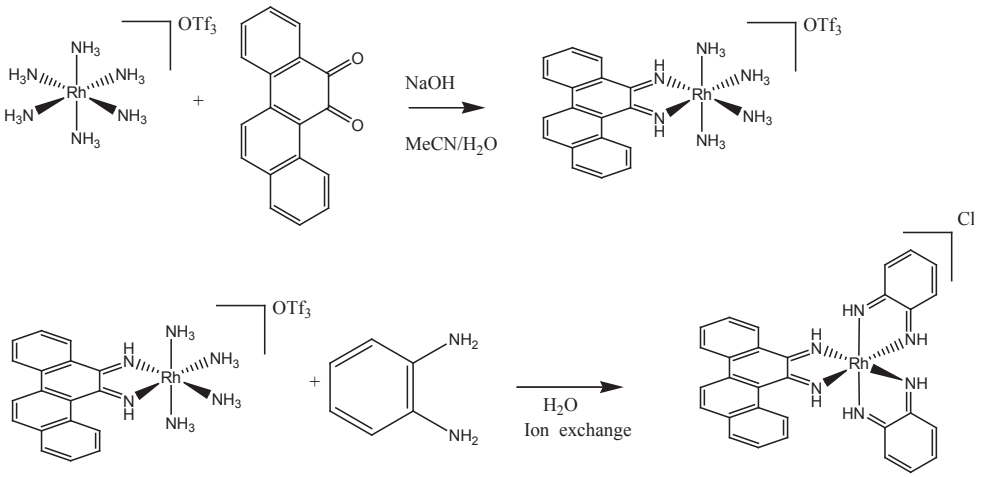

Figure 2.9: Synthesis of $[Rh(NH_3)_4(chrysi)]^{3+}$ and $[Rh(bqdi)_2(chrysi)]^{3+}$. $[Rh(NH_3)_6]OTf_3$ and chrysene quinone are dissolved in a 3:1 acetonitrile:water mixture with excess of sodium hydroxide to form $[Rh(NH_3)_4(chrysi)]^{3+}$. This is purified by reverse phase chromatography. The pure $[Rh(NH3)_4(chrysi)]^{3+}$ is stirred with an excess of o-phenylene diamine in water for 3 days under oxygen atomosphere. The compound is purified from the reaction mixture by ion exchange chromatography

$[Rh(pda)_2(chrysi)]OTf_3$ which is not isolated, but upon standing oxidizes with air to form the desired $[Rh(bqdi)_2(chrysi)]OTf_3$. The solution is extracted 5 times with 20 mL dichloromethane to remove excess ligand. The resulting solution of $[Rh(bqdi)_2(chrysi)]OTf_3$ was purified by ion exchange chromatography on Sephadex CM-25 resin using 0.1 M $MgCl_2$ as an eluant. This also exchanges the triflate ions for chloride. The eluant is collected onto a SepPac C18 reverse phase cartridge and eluted with 1:1:0.001 acetonitrile:water:TFA. The solution is then dried under vacuum to isolate the desired complex. Yield 2.5 mg, 13% yield. UV/vis (H_2O, pH 5): 263 nm, (56,400 $M^{-1}cm^{-1}$), 323 nm (14,200 $M^{-1}cm^{-1}$), 399 nm (12,600 $M^{-1}cm^{-1}$). MALDI-MS observed 573 m/z ($M-4H^+$) calculated 573 m/z.

DNA Preparation and Photocleavage Experiments $[Rh(bqdi)_2(chrysi)]^{3+}$. The oligonucleotides, 5'-GAT GTC GGT C*C*C ACG ATG GTG ACG GAT TAC C-3' and 5'-GGT AAT CCG TCA CCA TCG TG*C* GAC CGA CAT C -3', where *C* denotes the postion of the mismatch, were synthesized on an ABI 392 DNA/RNA synthesizer, using standard phosphoramidite solid-phase synthesis, and initially purified on Poly-Pak II cartridges and further purified by HPLC (98% 100 mM NH_4OAc/2% acetonitrile to 70% 100 mM NH_4OAc/30% acetonitrile over 30 minutes). The single strands were then 5' labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase. The labeled strands were further purified by gel electrophoresis (20% denaturing polyacrylamide gel), eluted from the gel via soaking in TEAA, ethanol precipitated, and annealed in the presence of unlabeled DNA. To the labeled 31-mer duplex (2 μM in 50 mM NaCl, 10 mM TrisHCl, pH 7.0) was added either $[Rh(bqdi)_2chrysi]Cl_3$ or $[Rh(bpy)_2chrysi]Cl_3$, and the sample was irradiated for 15 minutes using a solar simulator. After irradiation, all samples were lyophilized, denaturing formamide loading dye was added, and the samples were electrophoresed on a 20% polyacrylamide denaturing gel. The photocleavage results were quantitated via phosphorimagery (Molecular Dynamics Phosphorimager).

Competition between $[Rh(bqdi)_2(chrysi)]Cl_3$ and $[Rh(bpy)_2(chrysi)]Cl_3$. In this gel, 5 μL of 5 μM $[Rh(bpy)_2(chrysi)]Cl_3$ and 10 μL of 2.5 μM DNA, sequence is the same used in the previous experiment, was added to 5 μL of $[Rh(bqdi)_2(chrysi)]Cl_3$, whose concentration varied from 1 μM to 500 μM. This would allow the $[Rh(bqdi)_2(chrysi)]Cl_3$, which does not cleave, to compete with the $[Rh(bpy)_2(chrysi)]Cl_3$, which does cleave. These samples were irradiated by a solar simulator for 15 minutes. After this the samples were separated by gel electrophoresis and imaged by phosphorimagery. The data was analyzed as the fraction of DNA cleaved and plotted against concentration. The data was then fitted to a sigmoidal curve with an binding constant of 1×10^6 M^{-1}.

Footprinting the binding of $[Rh(bqdi)_2(chrysi)]Cl_3$ to a G-G mismatched DNA duplex with $[Rh(bpy)_2(phi)]Cl_3$. The experiment was conducted by mixing 1 μM DNA duplex, 5'-$^{32}P_i$-GAT GTC GGT C*G*C ACG ATG GTG ACG GAT TAC C-3' and 5'-GGT AAT CCG TCA CCA TCG TG*G* GAC CGA CAT C -3', 8 μM $[Rh(bpy)_2(phi)]Cl_3$ and a gradient of $[Rh(bqdi)_2(chrysi)]Cl_3$ from 125 μM to 25 nM and in the presence of 50 mM NaCl, 10 mM Tris pH 7.0. These samples were irradiated for 3 minutes. The cleaved DNA was analyzed by gel electrophoresis. The gel was imaged by phosphorimagery. The DNA spots were quantitated by Imagequant and plotted as the fraction of spot intensity to the lane intensity (fraction cleaved). The data were fit to a sigmoidal curve with an apparent binding constant of 4×10^5 M^{-1}. Nonspecific sites are also noted, but they are at least an order of magnitude less than the G-G mismatch.

2.3.3: Mismatch binding by bqdi complexes

After the initial synthesis of $[Rh(bqdi)_2(chrysi)]Cl_3$, its properties with DNA where investigated. A series of experiments were conducted with $[Rh(bqdi)_2(chrysi)]Cl_3$

and C-C mismatched DNA. This mismatch being the most destabilizing provides the best DNA binding site for the newly synthesized complex to determine photocleavage properties. A broad spectrum source was used for irradiation and upon irradiation for up to 3 hours, no cleavage of DNA was seen, an example of the lack of cleavage is shown in **figure 2.10.** $[Rh(bpy)_2(chrysi)]Cl_3$ is shown as example of photocleavage.

This result can be taken one of two ways. Either the complex does not bind mismatches and therefore does not cleave, or it binds mismatches and has a low photoefficiency. To distinguish between these two possibilities, a competition experiment between $[Rh(bpy)_2(chrysi)]Cl_3$ and $[Rh(bqdi)_2(chrysi)]Cl_3$ was conducted. To a solution containing 5 μM duplex containing a C-C mismatch, 5 μM $[Rh(bpy)_2(chrysi)]Cl_3$, 10 mM Tris pH 7.0, 50 mM NaCl, varying amounts of $[Rh(bqdi)_2(chrysi)]Cl_3$ were added. This experiment, shown in **figure 2.11**, demonstrates that the bqdi complex is competent in binding a C-C mismatch. From this competition data it is possible to determine the binding constant of the added intercalator, in this case $[Rh(bqdi)_2(chrysi)]Cl_3$. The binding constant of $[Rh(bpy)_2(chrysi)]Cl_3$ for a particular mismatch in the absence of $[Rh(bqdi)_2(chrysi)]Cl_3$ is easily determined by a simple titration. When an additional intercalator is added to the system, equilibrium is established between the two intercalators. Using a titration of $[Rh(bqdi)_2(chrysi)]Cl_3$ against a constant amount of $[Rh(bpy)_2(chrysi)]Cl_3$, it is possible to determine the affinity of the new intercalator towards a particular site, in this case a C-C mismatch. This relative measurement along with the absolute binding affinity of $[Rh(bpy)_2(chrysi)]Cl_3$ for the particular site can be used to calculate the binding affinity of the intercalator. From this competition data, a binding constant to the C-C mismatch of $7 \times 10^5\ M^{-1}$ can be calculated. From these results, we must infer that the bqdi complex can bind to mismatches, but has a low photoefficiency for DNA photocleavage.

Finally, an experiment was conducted with this complex to determine if

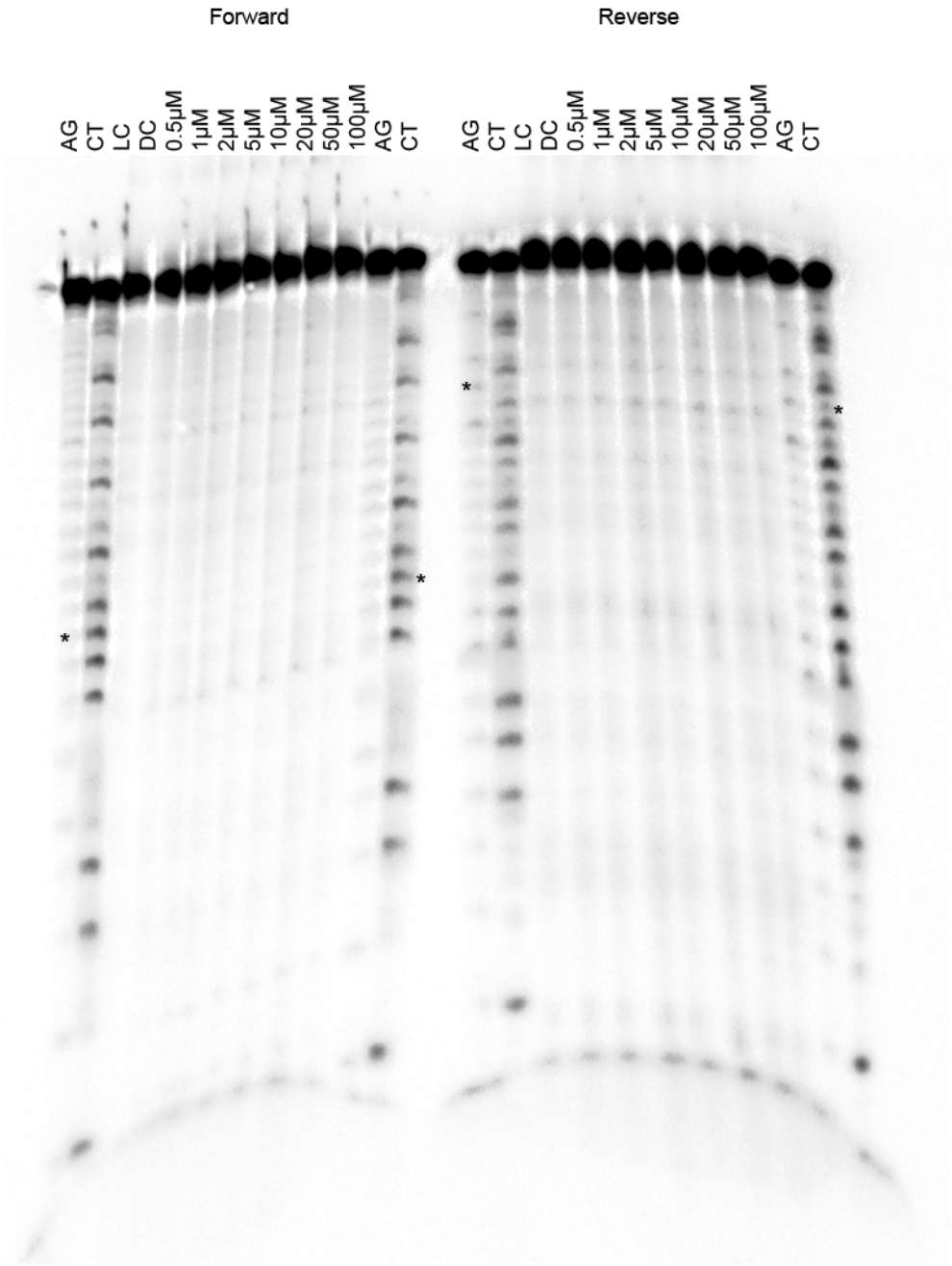

Figure 2.10: Photocleavage with $[Rh(bqdi)_2(chrysi)]^{3+}$. Photocleavage experiment with a C-C mismatched DNA duplex in the presence of 50 mM NaCl and 10 mM Tris pH 7.0. The mismatch site is indicated by the star. No photocleavage is noted, variation of irradiation time does not produce cleavage.

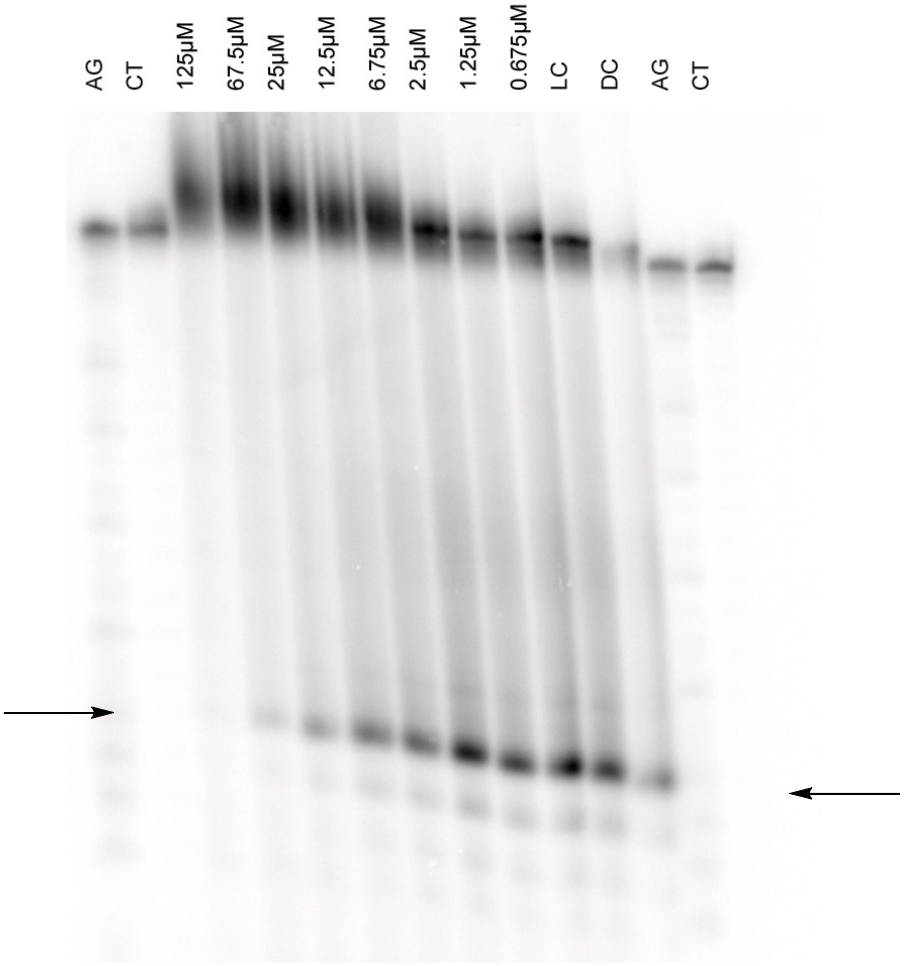

Figure 2.11: Competition between $[Rh(bqdi)_2(chrysi)]^{3+}$ and $[Rh(bpy)_2(chrysi)]^{3+}$. Increasing amounts of the bqdi complex are added to 5μM DNA containing a C-C mismatch, 5'-$^{32}P_i$-GAT GTC GGT C***C***C ACG ATG GTG ACG GAT TAC C-3' and the unlabeled complement with indicated C mismatched with a C, at the indicated site with 5 μM $[Rh(bpy)_2(chrysi)]^{3+}$ in 50 mM NaCl, 10 mM Tris pH 7.0. The bqdi complex effectively competes for the C-C mismatch site to which the bpy complex is bound. A binding constant of 1 x 10^6 M^{-1} can be calculated from the fit of this data.

indeed it could bind a G-G mismatch. Because $[Rh(bpy)_2(chrysi)]Cl_3$ does not bind G-G mismatches it cannot be used. Binding sites of non-photoreactive major groove intercalators can be determined by footprinting using the non-specific $[Rh(bpy)_2(phi)]Cl_3$. This complex cleaves DNA by abstracting a hydrogen atom from the DNA backbone upon UV excitation. Upon the addition of another intercalator, $[Rh(bpy)_2(phi)]Cl_3$ can be displaced from DNA duplex reducing the damage at a particular site on the DNA. This reduction in damage is called a footprint and it indicates the presence of a competing species at that site. This complex is capable of binding both matched and mismatched base pairs within DNA, with little sequence specificity.

Using $[Rh(bpy)_2(phi)]Cl_3$ as a footprinting agent, The binding of the bqdi complex to a G-G mismatch can be assessed. Solutions containing 1 μM G-G mismatched DNA duplex, 8 μM $[Rh(bpy)_2(phi)]Cl_3$, 10 mM Tris pH 7.0, and 50 mM NaCl. To this mixture various amounts of $[Rh(bqdi)_2(chrysi)]Cl_3$ were introduced before irradiation. These experiments clearly show a footprint at the G-G mismatch site. At higher concentrations the complex effectively blocks all sites on the DNA. While quantitative measurements of binding affinity cannot be obtained from this data, the binding constant of this complex to a G-G mismatch is at least $4 \times 10^5\ M^{-1}$. Additionally, the binding constant at the G-G mismatch is at least 10 times higher than that for the matched base pairs.

This complex represents the first attempt at the recognition of highly stabilizing mismatch. This mismatch G-G in this sequence context, 5'-C***G***C-3' is stabilizing to the DNA duplex by -1.22 kcal/mol, as calculated using the nearest neighbor method.[29] It seems likely that this complex is capable of recognizing other highly stable mismatches. Additional studies will be required to determine which mismatches and in what sequence contexts this complex will bind.

How is this complex able to recognize this stable mismatch? The $[Rh(bpy)_2(chrysi)]^{3+}$ has been shown to bind mismatches if they are destabilized to a

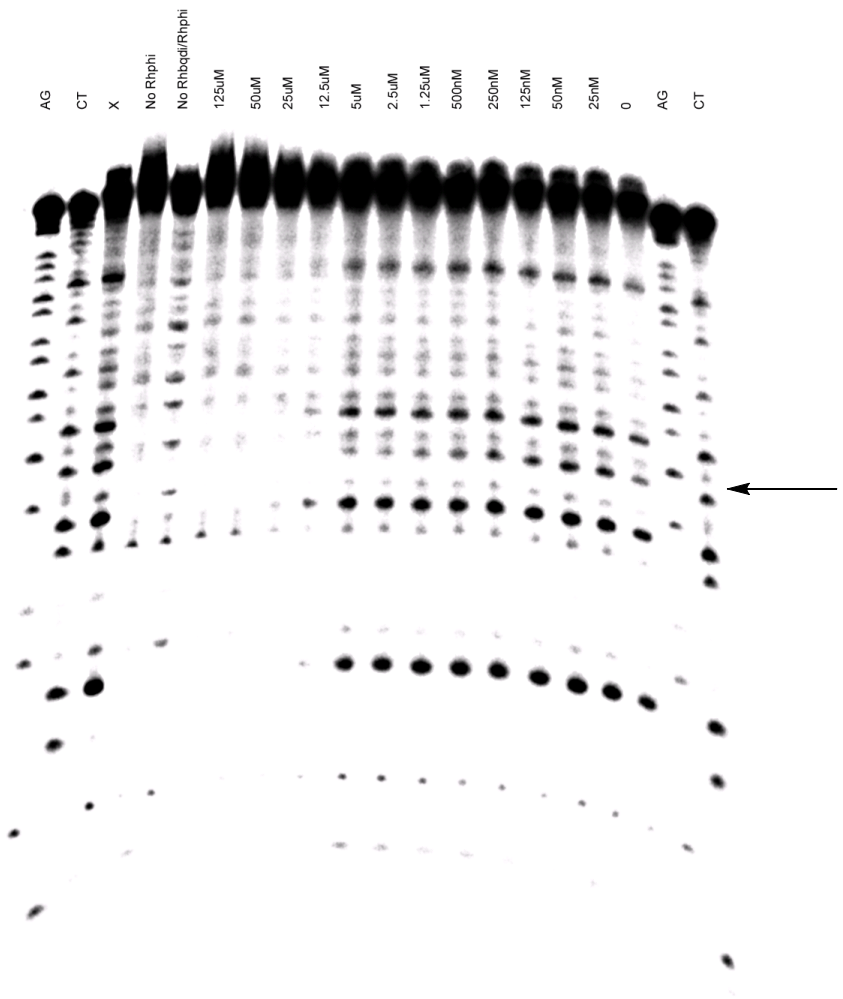

Figure 2.12: Photofootprinting of $[Rh(bqdi)_2(chrysi)]^{3+}$ binding to a G-G mismatch. Since $[Rh(bqdi)_2(chrysi)]^{3+}$ is not capable of photocleavage, photofootprinting was used as an alternative. 1 μM DNA containing a G-G mismatch, 5'-$^{32}P_i$-GAT GTC GGT C***G***C ACG ATG GTG ACG GAT TAC C-3' and the unlabeled complement with indicated G mismatched with a G, was mixed with 8 uM $[Rh(bpy)_2(phi)]^{3+}$ and photocleaved in the presence of varying concentrations of $[Rh(bqdi)_2(chrysi)]^{3+}$, as indicated on the gel. The reaction was performed with 50 mM NaCl and 10 mM Tris·Cl as buffer. The location of the G-G mismatch is indicated by the arrow. Both specific and nonspecific behaviors are noted.

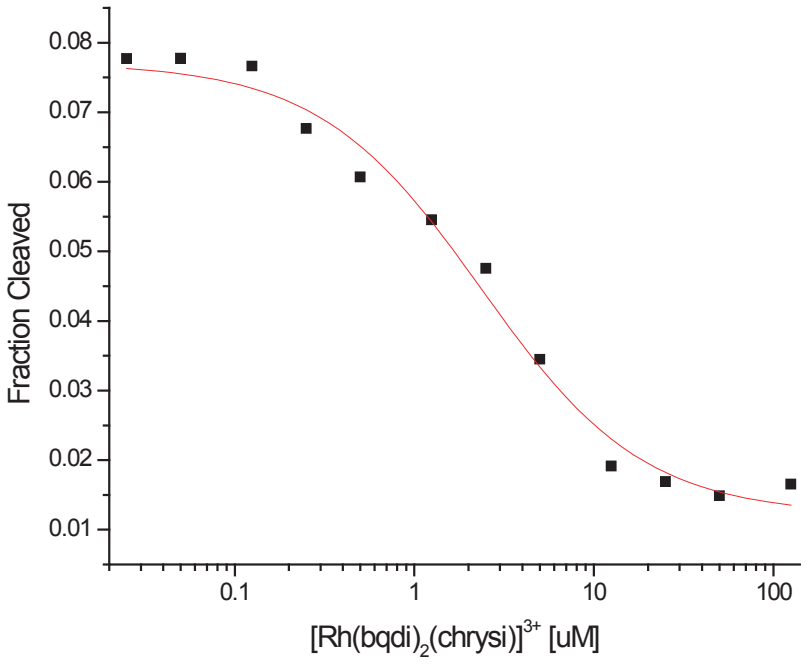

Figure 2.13: Binding of $[Rh(bqdi)_2(chrysi)]^{3+}$ to a G-G mismatch. The quantified intensity of the G-G mismatched site, from **figure 2.12**, expressed as fraction cleaved, is plotted against $[Rh(bqdi)_2(chrysi)]^{3+}$ concentration. The apparant binding constant is 4×10^5 M^{-1}. The actual binding constant will be this value or higher. The nonspecific binding constant is at least an order of magnitude less. 1μM DNA containing a G-G mismatch, 5'-$^{32}P_i$-GAT GTC GGT C***G***C ACG ATG GTG ACG GAT TAC C-3' and the unlabeled complement with indicated G mismatched with a G, was mixed with 8 μM $[Rh(bpy)_2(phi)]^{3+}$ and photocleaved in the presence of varying concentrations of $[Rh(bqdi)_2(chrysi)]^{3+}$. The reaction was performed with 50 mM NaCl and 10 mM Tris·Cl as buffer.

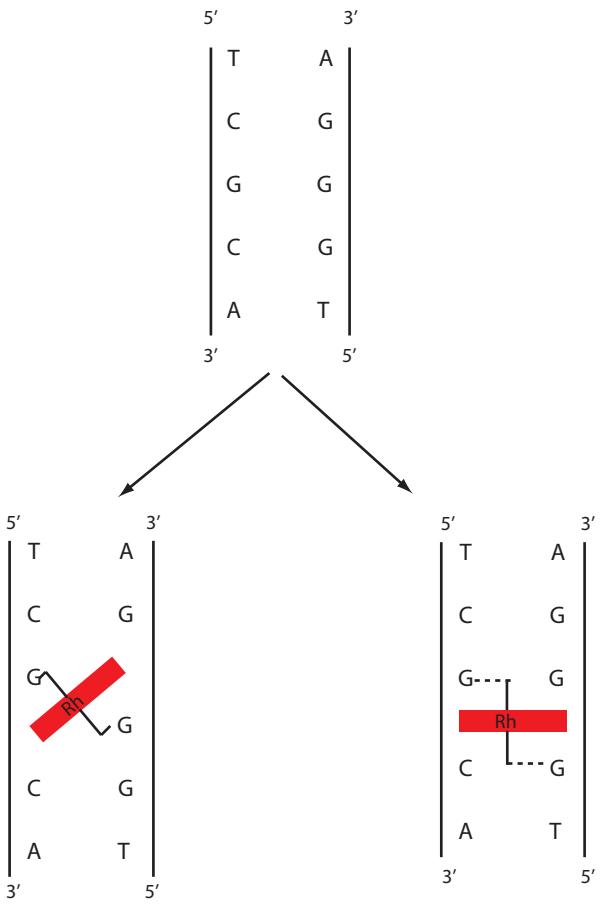

Figure 2.14: Two schematic representations of $[Rh(bqdi)_2(chrysi)]^{\#+}$ binding. It seems likely that the chrysi ligand intercalates into the DNA base stack. This will allow the imines of the bqdi ligand to hydrogen bond with the guanine bases present. Two modes of intercalation can be proposed. On the left, the complex inserts in a canted mode, stacking with both guanines, and hydrogen bonding with both. This will distrupt the hydrogen bonding of the G-G mismatch and would likely severly kink the DNA. On the right, the complex inserts in a normal intercalative mode, stacking with the G-G mismatch and hydrogen bonding with it and nearby G-C base pair.

certain amount.[27] It would seem then that this complex has increased affinity to this particular site. This small increase in affinity may be sufficient to push the complex over the energetic barrier to a tighter bound form.

The structure of how the chrysi ligand binds to the mismatch is also of considerable interest. One possibility is that the axial imines hydrogen bond with the O6 of guanine similar to the parent ammine complexes.[24; 25; 30] In this case, the chrysi ligand intercalates and stacks with both the G-G mismatch and a C-G base pair, hydrogen bonding with the guanines of both the mismatch and the matched base pair. The other possibilities are that the chrysi ligand intercalates in a canted fashion with the two guanines of the mismatch on opposite sides of the intercalated chrysi ligand. In this situation it would seem that this binding mode would place a severe kink within the DNA and hydrogen bonding between the guanines would be lost, but a hydrogen bond from the coordinated imine to both of the guanines would be possible. It seems unlikely that extrahelical bases could be involved as the both hydrogen bonding and aromatic stacking of the guanine mismatch would both be disrupted and could hardly be recovered by the intercalation of the complex. Minor groove binding also seems unlikely as the complex, though smaller, would still seem to be far too large to associate with any depth in the minor groove. Perhaps future studies will elucidate the structure of the intercalated chrysi ligand.

2.3: References

1. Figgis, B. N., Gillard, R. D., Nyholm, R. S., Wilkinson, G. (1964). Action of reducing agents on pyridine complexes of rhodium(III). *J. Chem. Soc.*, 5189.

2. Mckenzie, E. D., Plowman, R. A. (1970). Rhodium(III) compounds with 1,10,-

phenanthroline and 2,2-bipyridyl. *J. Inorg. Nucl. Chem.* **32**, 199.

3. Dixon, N. E., Lawrance, G. A., Lay, P. A., Sargeson, A. M. (1983). (Trifluoromethanesulfonato-O)pentaammine complexes: Versatile synthetic intermediates. *Inorg. Chem.* **22**, 846-847.

4. Dixon, N. E., Lawrance, G. A., Lay, P. A., Sargeson, A. M. (1984). Synthetically versatile (trifluoromethanesulfonato)metal amine complexes. *Inorg. Chem.* **23**, 2940-2947.

5. Gidney, P. M., Gillard, R. D., Heaton, B. T. (1972). 1,10-Phenanthroline and 2,2'-bipyridyl complexes of rhodium(III). *J. Chem. Soc. Dalt. Trans.* 2621-2628.

6. Murner, H., Jackson, B. A., Barton, J. K. (1998). A versatile synthetic approach to rhodium(III) diimine metallointercalators: Condensation of o-quinones with coordinated cis-ammines. *Inorg. Chem.* **37**, 3007-3012.

7. Evans, I. P., Everett, G. W., Sargeson, A. M. (1975). Intermolecular imine formation: Condensation of organic-molecules with hexammine Ru(III) and Pt(IV) ions. *J. Chem. Soc. Chem. Comm.*, 139-140.

8. Harrowfield, J. M., Sargeson, A. M. (1974). Reactions of coordinated nucleophiles: Intramolecular imine formation. *J. Am. Chem. Soc.* **96**, 2634-2635.

9. Gainsford, A. R., Sargeson, A. M. (1978). Stereoselective and regioselective condensation: Reactions of coordinated aminoacetone. *Aust. J. Chem.* **31**, 1679-1688.

10. MacDonnell, F. M., Bodige, S. (1996). Efficient stereospecific syntheses of chiral ruthenium dimers. *Inorg. Chem.* **35**, 5758.

11. Pyle, A. M., Chiang, M. Y., Barton, J. K. (1990). Synthesis and characterization of physical, electronic, and photochemical aspects of 9,10-phenanthrenequinone diimine complexes of ruthenium(II) and rhodium(III). *Inorg. Chem.* **29**, 4487-4495.

12. Pyle, A. M., Barton, J. K. (1987). Synthesis and spectroscopic characterization of the purple tris(phenanthrenequinone diimine)ruthenium(II) ion. *Inorg. Chem.* **26**, 3820-3823.

13. Junicke, H., Hart, J. R., Kisko, J., Glebov, O., Kirsch, I. R., Barton, J. K. (2003). A rhodium(III) complex for high-affinity DNA base-pair mismatch recognition. *Proc. Natl. Acad. Sci. USA* **100**, 3737-3742.

14. Palmer, J. W., Basolo, F. (1960). Effect of transition metal ion on rates of hydrogen exchange in metal ammines. *J. Inorg. Nucl. Chem.* **15**, 279-286.

15. Palmer, J. W., Basolo, F. (1960). Effect of ligands on rates of hydrogen exchange of substituted metal ammines. *J. Phys. Chem.* **64**, 778-780.

16. Sitlani, A., Long, E. C., Pyle, A. M., Barton, J. K. (1992). DNA photocleavage by phenanthrenequinone diimine complexes of Rhodium(III): Shape-selective recognition and reaction. *J. Am. Chem. Soc.* **114**, 2303-2312.

17. Turro, C., Hall, D. B., Chen, W., Zuilhof, H., Barton, J. K., Turro, N. J. (1998). Solution photoreactivity of phenanthrenequinone diimine complexes of rhodium and correlations with DNA photocleavage and photooxidation. *J. Phys. Chem. A* **102**, 5708-5715.

18. Jackson, B. A., Barton, J. K. (1997). Recognition of DNA base mismatches by a rhodium intercalator. *J. Am. Chem. Soc.* **119**, 12986-12987.

19. Peyret, N., Seneviratne, P. A., Allawi, H. T., SantaLucia, J. (1999). Nearest-neighbor thermodynamics and NMR of DNA sequences with internal A-A, C-C, G-G, and T-T mismatches. *Biochemistry* **38**, 3468-3477.

20. Duval, A., Hamelin, R. (2002). Mutations at coding repeat sequences in mismatch repair-deficient human cancers: Toward a new concept of target genes for instability. *Cancer. Res.* **62**, 2447-2454.

21. Jacob, S., Praz, F. (2002). DNA mismatch repair defects: role in colorectal

carcinogenesis. *Biochimie* **84**, 27-47.

22. Boon, E. M., Kisko, J. L., Barton, J. K. (2002). Detection of DNA base mismatches using DNA intercalators. *Redox Cell Biology and Genetics, Pt B* **353**, 506-522.

23. Erkkila, K. E., Odom, D. T., Barton, J. K. (1999). Recognition and reaction of metallointercalators with DNA. *Chem. Rev.* **99**, 2777-2795.

24. Krotz, A. H., Hudson, B. P., Barton, J. K. (1993). Assembly of DNA recognition elements on an octahedral rhodium intercalator: Predictive recognition of 5'-TGCA-3' by Δ-[Rh[(R,R)-Me_2trien]Phi]$^{3+}$. *J. Am. Chem. Soc.* **115**, 12577-12578.

25. Kielkopf, C. L., Erkkila, K. E., Hudson, B. P., Barton, J. K., Rees, D. C. (2000). Structure of a photoactive rhodium complex intercalated into DNA. *Nat. Struct. Biol.* **7**, 117-121.

26. Jackson, B. A., Alekseyev, V. Y., Barton, J. K. (1999). A versatile mismatch recognition agent: Specific cleavage of a plasmid DNA at a single base mispair. *Biochemistry* **38**, 4655-4662.

27. Jackson, B. A., Barton, J. K. (2000). Recognition of base mismatches in DNA by 5,6-chrysenequinone diimine complexes of rhodium(III): A proposed mechanism for preferential binding in destabilized regions of the double helix. *Biochemistry* **39**, 6176-6182.

28. Vanallan, J. A., Reynolds, G. A. (1963). Polynuclear heterocycles VI: Reaction of 2,3-dichloro-1,4-naphthoquinone with aromatic amines. *J. Org. Chem.* **28**, 1019.

29. SantaLucia, J., Hicks, D. (2004). The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.* **33**, 415-440.

30. Krotz, A. H., Kuo, L. Y., Shields, T. P., Barton, J. K. (1993). DNA recognition by rhodium(III) polyamine intercalators: Considerations of hydrogen-bonding and van der Waals interactions. *J. Am. Chem. Soc.* **115**, 3877-3882.

31. Fu, P. K. L., Bradley, P. M., Turro, C. (2003). Stabilization of duplex DNA structure and suppression of transcription in vitro by bis(quinone diimine) complexes of rhodium(III) and ruthenium(II). *Inorg. Chem.* **42**, 878-884.

Chapter 3: Single Nucleotide Polymorphism Discovery by Targeted DNA Photocleavage

3.1: Introduction

Single nucleotide polymorphisms (SNPs) have been the focus of research for both their roles in disease[1] and use as markers in genetics[2]. Occuring as frequently as 1 SNP per 1000 base pairs, they are by far the most common mutation present in the human genome[3,4]. It is estimated that there are 10 million SNPs with an allele frequency greater than 1%[5] and approximately 5 million so called common SNPs with frequencies exceeding 10%[6]. Most SNPs have no effect on gene expression or gene products and therefore appear to be phenotypically silent.

It has been suggested that developing a dense map of SNP locations can significantly aid in the identification of genetic factors associated with disease[7]. Owing to the presence of recombination hot spots[8,9], the genome is organized into haplotype blocks that infrequently undergo recombination separated by areas of frequent recombination. On which regions might SNP discovery efforts be focused? Because of the great cost associated with full scale mapping of SNPs throughout the genome, current efforts have been directed toward SNP discovery within genes likely associated with disease or disease predisposition[10]. This strategy then necessarily excludes discovery within regions of the genome of unknown function.

Although many have touted the benefits of SNPs, it is because discovering the location of low frequency SNPs remains a costly and time consuming challenge that

---

Adapted from Hart, J. R., Johnson, M. D., Barton, J. K. (2004). Single-nucleotide polymorphism discovery by targeted DNA photocleavage. *Proc. Natl. Acad. Sci. USA* **101**, 14040-14044.

SNP discovery efforts have been restricted. Currently, there are few alternatives to resequencing for the initial discovery of these SNPs. Resequencing is expensive in terms of materials, labor, and information processing.[11,12] A relatively large population must be sequenced throughout the genome[7], or alternatively every haplotype block and intervening hotspots[3], in order to create a high density SNP map of the genome. Sequence data from each individual, or pool, must be compared to all other individuals, or pools, to detect the presence of a SNP within that sequence. Current techniques require the sequencing of a particular region of the genome several times in order to locate a rare SNP[11,12]. Additionally these methods suffer from a high false positive rate of 10%-15%. Once the SNP has been localized, many existing technologies can be applied to determine alleles and allele frequency of the SNP for a broader population[13]. For this reason the development of a new assay for the discovery and localization of SNPs would be particularly attractive.

As an alternative to resequencing as a method of SNP discovery, we have adapted our mismatch-specific rhodium(III) metallointercalators[14,15] to the task of detecting low frequency sequence variations. While DNA mismatches may be of varied sequence and within varied sequence contexts, the factor that distinguishes mispaired DNA bases from Watson-Crick base pairs is the local instability arising from decreased hydrogen bonding and aromatic stacking of the mismatched bases.[16] This destabilization is exploited in site recognition by $[Rh(bpy)_2(phzi)]Cl_3$ (phzi = 3,4-benzo[a]phenazine quinone diimine) and $[Rh(bpy)_2(chrysi)]Cl_3$ (chrysi = 9,10-chrysene quinone diimine). The phi (phenanthrenequinone) ligand is an optimal size for intercalation in well matched DNA interbase pair sites, based upon crystallography.[16a] However, the analogous but more bulky phzi and chrysi ligands are wider than a well matched base pair. This shape selective approach allows intercalation of the bulky chrysi and phzi complexes in the destabilized region of the DNA mismatch, but not within stable well paired

regions. $Rh(bpy)_2(chrysi)^{3+}$ (Rhchrysi) and $Rh(bpy)_2(phzi)^{3+}$ (Rhphzi) have been shown to be remarkably specific and high affinity binders of DNA mismatches. Upon irradiation, these complexes promote direct strand cleavage of the DNA backbone selectively neighboring the mismatched site. It has been documented that Rhchrysi targets 80% of all possible base mismatches, with full variation of the sequence context as well as identity of the mismatch, and the sites targeted correlate closely with their thermodynamic instability[17]. Importantly, owing to the high specificity of these complexes, the targeting of a single base mismatch within a 2725 bp DNA duplex by photocleavage has also been demonstrated[18].

We have exploited our mismatch-specific rhodium complexes to develop a new methodology for the detection of SNPs within genomic DNA, **figure 3.1**. The area of the genome to be probed for SNPs is amplified by PCR from pooled genomic DNA. SNPs in the PCR product are turned into mismatches by denaturing and annealing the DNA. The site-specific targeting of DNA mismatches by the metal complexes provides a method for cleaving the DNA and therefore marking the SNP site. The SNPs can then be detected using capillary gel electrophoresis.

Resequencing does however have some advantages over our method, specifically resequencing gives a clear picture of the alleles present and their frequencies, albeit approximately, in addition to the SNPs location. Regardless of the method of initial SNP discovery, larger populations could be more economically genotyped after the initial localization of the SNP to give a clearer picture of SNP frequency and distribution within a larger and more diverse population.

In contrast to the resequencing method, our new assay allows for low frequency SNPs to be detected directly and unambiguously from a pooled sample, greatly reducing the costs for the localization of new SNPs. Specifically fewer PCR reactions and no cycle sequencing is required. Additionally, information processing is greatly reduced as

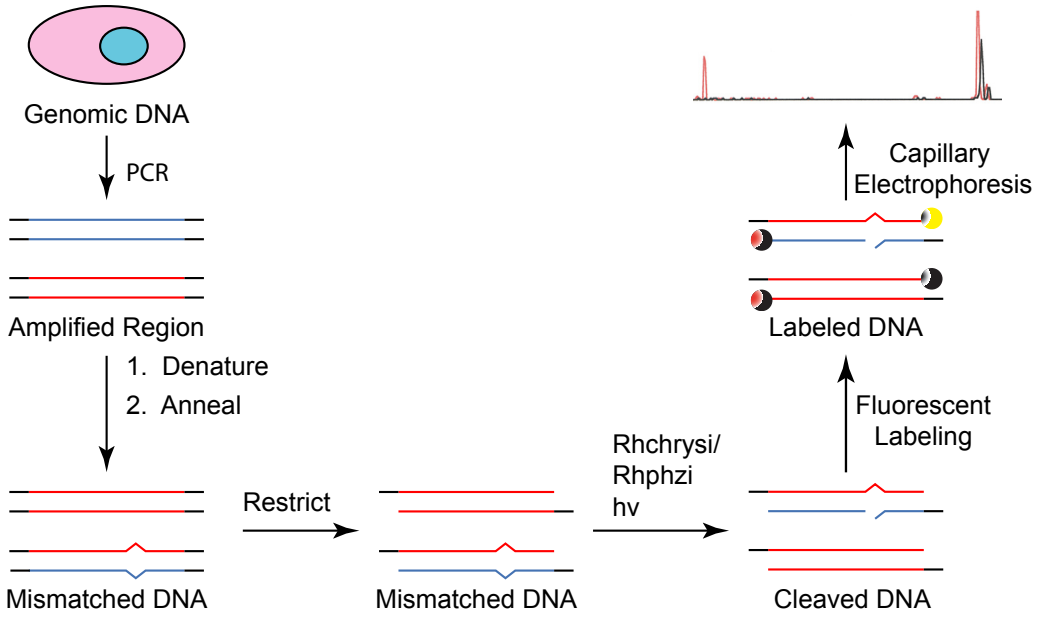

Figure 3.1: Strategy for detecting an SNP from genomic DNA in pooled samples. The portion of genomic DNA to be analyzed is selected and amplified by PCR. These PCR products are denatured and annealed to produce mismatches at the locations of heterozygous SNPs. The DNA is restricted to create a 3' underhang to be labeled. The mismatched DNA is then treated with either Rhchrysi or Rhphzi and irradiated to cleave the DNA at the locations of the mismatches, corresponding to the SNP sites. The cleaved products are fluorescently labeled, followed by separation and detection by capillary electrophoresis.

individual samples need not be compared to determine the presence of an SNP.

3.2: Materials and experimental

Materials. $[Rh(bpy)_2(phzi)]Cl_3$ and $[Rh(bpy)_2(chrysi)]Cl_3$ were synthesized according to published procedures.[14,15] Restriction enzymes, EcoRI, XhoI, and ClaI, calf intestinal alkaline phosphatase(CIAP), and Taq DNA polymerase were purchased from Roche Molecular Biochemicals. The SNaPshot labeling kit was provided by Applied Biosystems. Primers used were synthesized using an ABI 394 DNA synthesizer by standard phophoramidite chemistry and purified using Glen Research PolyPakII columns. Plasmids used as templates in this assay were provided by Applied Biosystems. Pooled human genomic DNA was purchased from Roche Molecular Biochemicals.

Preparation of mismatched templates. Both the plasmid templates and pooled human genomic DNA were amplified using PCR with the primers, F 5'-TTT TTT ATC GAT CGC GTT GGC CGA TTC ATT AAT G-3' and R 5'-TTT TTT CTC GAG GCT GCG CAA CTG TTG GGA AG-3', or F 5'-AAA ATC GAT AGA ACA AAA GGA TAA GGG CTC AG-3' and R 5'-AAA CTC GAG GTG TGG CCA TAT CTT CTT AAA CG-3' respectively. These products were sequenced by standard fluorescent dye terminator cycle sequencing, shown in **figure 3.2** and **figure 3.3**. After the amplification of the DNA, Primers and unincorporated dNTPs were degraded using a combination of 6 U ExoI and 20 U CIAP. The DNA product was purified using a QIAquick PCR purification column eluting into 10 mM Tris HCl buffer, pH 8.0. DNA was thermally denatured under these low ionic strength conditions by heating at 99 ºC for 25 minutes. DNA was annealed in the presence of 60 mM Tris HCl pH 7.5, 10 mM $MgCl_2$, 100 mM NaCl, 1 mM dithioerythritol, by heating to 95 ºC for 10 minutes and linearly ramping temperature to 4 ºC over a period of 150 minutes. This generates mismatches at heterozygous SNP sites. Restriction sites present in the primers were cut using 5 U ClaI

```
0         1         2         3         4         5         6
12345678901234567890123456789012345678901234567890123456789012345678901234567890
TTTTTTCTCGAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATT
ACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTT
TTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTGTTATCCCAAAACCACAAAGA
NAAATCATAAAATGAACAGGGGGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTGGC
GTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAA
CATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCAC
ATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCA
TTAATGAATCGGCCAACGCGATCGATAAAAAA
```

Figure 3.2: Sequence of PCR product generated from plasmids. The full product is 459 bases in length, restricted product is 440 bases in length. The polymorphic site denoted N in the sequence can be A, T, G, or C, that designation corresponding to the base on the forward strand.

```
0         1         2         3         4         5         6
12345678901234567890123456789012345678901234567890123456789012345678901234567890
AAAATCGATAGAACAAAAGGATAAGGGCTCAGAGAGCTTCAGGGATATGTGATGGACTCA
CCAGGTGAGGCCGCCAGACTGCTGCAGGGGAAGCAAAGGAGAAGCTGAGAAGATGAAGGA
AAAGTCAGGGTCTGGAGGGGCGGGGGTCAGGGAGCTCCTGGGAGATATGGCCACATGTAG
CGGCTCTGAGGAATGGGTTACAGGAGACCTCTGGGGAGATGTGACCACAGCAATGGGTAG
GAGAATGTCCAGGGCTATGGAAGTCGAGTATGGGGACCCCCCCTTAACGAAGACAGGGCC
ATGTAGAGGGCCCCAGGGAGTGAAAGAGCCTCCAGGACCTCCAGGTATGGAATACAGGGG
ACGTTTAAGAAGATATGGCCACACACCTCGAGTTT
```

Figure 3.3: Sequence of PCR product from TNF gene. Full product is 395 base pairs, the -862 SNP is located at position 281. After restriction the duplex is 382 base pairs in length.

and 5 U XhoI incubated at 37 ºC for 60 minutes.

Detection of mismatches generated at the polymorphic sites. The annealed, mismatched DNA was cleaved using 500 nM Rhchrysi or 200 nM Rhphzi, in the presence of 30 mM Tris HCl pH 7.5, 5 mM $MgCl_2$, 50 mM NaCl, 0.5 mM dithioerythritol, by irradiating samples with 442 nm or 340 nm light, respectively, with a 1000 W Oriel Hg/Xe arc lamp fitted with a monochromator, a 295 nm UV cutoff filter, and an IR filter (Thermo-Oriel). Cleavage can also be induced using a standard 302 nm transilluminator, inverted at a distance of 4 cm from the top of the sample or 365 nm "blacklight" also a distance of 4 cm. Irradiation times using a transilluminator or blacklight are approximately 60 times longer than those required for the arc lamp. Following cleavage samples were dried and fluorescent tags were introduced by single base extension using the SNaPshot kit. The fluorescently labeled products were analyzed by capillary electrophoresis performed on an ABI 310 Prism instrument using a 47 cm x 50 μm capillary, POP4 polymer, dye set E5 with LIZ size standards.

3.3: Preparation of mismatched DNA templates for SNP detection

Typically, rhodium-mediated cleavage of DNA has been performed on relatively short oligonucleotide duplexes and the cleavage products have been monitored using denaturing polyacrylamide gel electrophoresis[19]. Additionally, duplexes containing mismatches have been generated by the high temperature annealing of synthetic single stranded DNA[14]. These conditions provide superb control and sensitivity over reaction conditions allowing the determination of binding constants, etc. In contrast, for the detection of SNPs within longer biologically derived DNA duplexes, these methods are not applicable.

As an alternative, DNA duplexes are generated by PCR using a polymorphic

template. These duplexes are then fully denatured at high temperatures and low ionic strength to generate single stranded DNA[20]. These single strands are then annealed at a higher ionic strength by slow cooling to ambient temperature. In this way, the forward and reverse strands anneal randomly so as to obtain a mixture of products, in which DNA mismatches are generated at the sites of SNPs. The degree of strand exchange can be calculated from the maximum cleavage intensities observed for equimolar polymorphic templates. In this system, strand exchange exceeds 45%, and approaches the statistical limit. We have chosen to use this low ionic strength method for denaturation over the previously demonstrated sodium hydroxide denaturing method[18] to eliminate the time consuming desalting steps.

In addition to the creation of mismatched templates for SNP cleavage, a method of detecting the DNA and its cleaved products was needed. Radionucleotides and phosphorimagery were used previously.[14,15,18] Fluorescent tagging is preferable because of the abundance of instruments available for the rapid and high throughput analysis of fluorescently labeled polynucleotides, primarily for sequencing. To incorporate fluorescent tags site specifically, PCR primers used in amplification were tagged with a ClaI restriction site on the forward primer and an XhoI restriction site on the reverse primer. After cleavage with the restriction enzymes, these restriction sites may be labeled by polymerase and standard dideoxyribonucleotide dye-terminators used in sequencing and primer extension genotyping.[20] This results in unique fluorescent tagging of both 3' ends of the DNA simultaneously.

Templates for these studies were derived from two sources. First for basic characterization of the system, a set of four plasmid templates containing a single base substitution in their insert was used. Secondly, a pooled human DNA source from 80 individuals was used as a template in the detection of a known SNP in the tumor necrosis factor (TNF) gene.

Table 3.1: Homozygous and heterozygous templates and their photocleavage

SNP Alleles present[a]	Mismatches generated[b]	Phzi Photocleavage[c]	Chrysi Photocleavage[d]
AT	none		
TA	none		
GC	none		
CG	none		
AT + TA	AA + TT	x	
AT + GC	AC + GT	x	x
AT + CG	AG + CT		
TA + GC	TC + GA	x	x
TA + CG	TG + CA	x	x
GC + CG	GG + CC	x	x

[a]The base pair(s) at the polymorphic site present in the template plasmid. The first base indicates the template plasmid used. [b]Upon denaturing and annealing of the PCR amplified template, mismatches can be generated as indicated if heterozygousity exists. DNA duplexes are denatured in the presence of 10 mM Tris pH 8.5, and annealed with the addition of buffer to make a final concentration of 60 mM Tris HCl pH 7.5, 10 mM MgCl2, 100 mM NaCl, 1 mM dithioerythritol. This buffer is diluted to half with the addition of metal complex to 500 nM Rhchrysi or 200 nM Rhphzi. Both [c]Rhphzi and [d]Rhchrysi generate site specific cleavage products, upon irradiation at 340 nm for 1 hr or 442 nm for 1 hr respectively, as denoted by X.

3.4: Discovery of SNPs by photocleavage

Photocleavage and analysis using the various plasmid templates were performed and representative results are shown in **figure 3.2**. For plasmid samples not containing a heterozygous locus, no cleaved products are detected, **figure 3.3**. Irrespective of rhodium addition or time of irradiation, a low background is detected in all samples, typically present with an intensity of <2% of the parent peak. In contrast, samples containing a heterozygous locus, formed by the mixing of dissimilar templates before annealing, clearly show the formation of cleaved products after irradiation. Within the sequence context of our plasmid, the forward strand is preferentially cleaved in all cases. The size of the cleaved product aligns perfectly with the location of the polymorphic site. Upon extensive irradiation, small amounts of cleaved reverse strand are also noted. This is expected as rhodium complexes of this type cleave the DNA duplex asymmetrically[18].

A summary of the possible polymorphisms and their detected cleavage with both Rhchrysi and Rhphzi is shown in **table 3.1**. As is evident, a variety of SNPs are detected using both complexes. Also noteworthy, however, is the lack of reaction with the generated AAA site, consistent with previous studies of Rhchrysi targeting as a function of sequence context. Within this sequence context, because of base slippage, the adenine-containing mismatches are exceptionally stable, with hydrogen bonding of the mismatched adenine base with the thymines present on the opposite strand; thus the stabilized site is not targeted. These data emphasize the fact that specific cleavage is a clear indicator of the SNP, but the lack of reaction does not preclude the presence of an SNP creating a thermodynamically stable mismatch.

3.5: Photolysis with common light sources

A mercury-xenon arc lamp and monochromator are typically employed in our

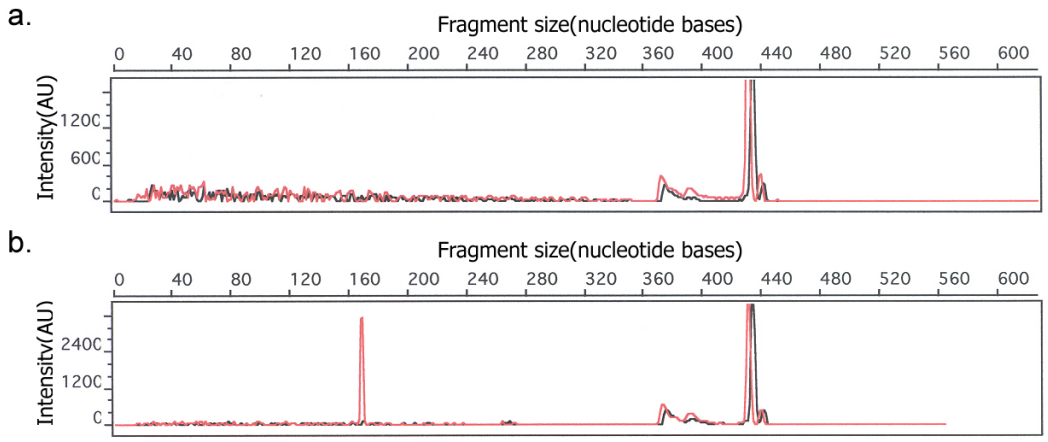

Figure 3.4: Representative capillary electrophoresis traces demonstrating the detection of an SNP within a plasmid. The red trace represents the forward strand while the black corresponds to the reverse strand. (a) When a homozygous template, in this case the homozygous plasmid A, is tested by the assay procedure, no cleaved products are detected. The full length PCR product is present, 436 bases in length. (b) When two templates, here G and C plasmids, are mixed, denatured, and reannealed to generate a heterozygous site, mismatches are generated. After photocleavage for 30 min at 442 nm with 500 nM Rhchrysi, and fluorescently tagging, the cleaved product, 170 bases in length, is evident (see experimental section). The horizontal scale shows the fragment size in units of nucleotides. The vertical scale is fluorescent signal intensity in arbitrary units. We ascribe the small peak ~380 bases in length to overloading of the sample.

photocleavage reactions. While this lamp is convenient because of its high intensity light, other more common laboratory light sources may be utilized in the photocleavage reaction. Both a 302 nm transilluminator used in the visualization of agarose gels, and a 365 nm "blacklight" have been used to activate the metal complex for DNA cleavage, although at a somewhat diminished rate owing to the decreased intensity of these sources. For example, while the arc lamp is capable of driving the cleavage reaction to completion in 20 minutes, the 365 nm blacklight can only cleave 28% in 4 hours and the 302 nm transilluminator can cleave 15% in 4 hours. Before any light source is used with this assay, controls are conducted in the absence of rhodium complex to determine if the light alone can damage the DNA. None of these lamps are capable of detectably damaging DNA. Typical white fluorescent light and incandescent light do not produce sufficient ultraviolet light to produce significant strand scission.

3.6: Detection of an SNP from a biological source

Ultimately this method is not designed for the detection of an SNP present in a synthetic plasmid, but rather it is designed for the discovery of previously unknown SNPs from biological sources. We therefore sought to examine a target SNP known to be present in the general population and well characterized as to its allele frequency as a test of the methodology. Here then, the known SNP, -307, in the TNF promoter[22] was targeted for detection.

To determine the presence and frequency of this SNP within our pooled genomic DNA sample, this SNP was detected by the generally accepted resequencing method, **figure 5a**. This shows demonstrates the presence of the -307 SNP within our genomic sample. The frequency of this SNP within the pooled sample is 11% A with the remaining 89% C.

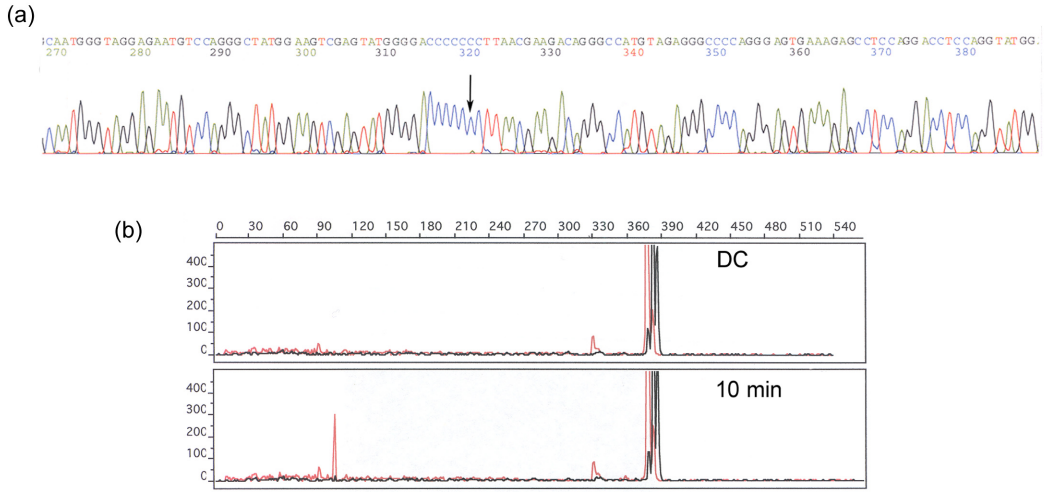

Figure 3.5: The discovery of an SNP in the TNF promoter region by two methods, (a) resequencing and (b) photocleavage with Rhchrysi. In (a) The TNF promoter was sequenced using a pooled genomic DNA sample. A portion of this sequencing data is shown highlighting a polymorphic site, demonstrating C allele frequency of 89% with a minor A allele frequency of 11%. In (b) using our targeting assay, this SNP is readily detected by capillary electrophoresis. After photocleavage with 500 nM Rhchrysi at 442 nm for 10 min, the parent band, 377 bases in length, is cleaved and a new peak appears at 105 bases. This fragment length corresponds exactly with the location of the SNP detected by sequencing in (a).

A pooled DNA sample, taken from a mix of genomic DNA from 80 individuals, was amplified using PCR primers designed to target the region of the SNP in the TNF promoter. After amplification, the DNA was treated under the same conditions as the plasmid derived DNA (*vide supra*). Following denaturation and annealing, the 375 length PCR fragments, incubation with 500 nM Rhchrysi, photolysis, and fluorescent tagging, the samples were analyzed by capillary electrophoresis. As is evident in **figure 5b**, a peak appears with a length corresponding precisely to the location of the SNP. The contrast in ease of discovery of this SNP using sequencing versus our SNP photocleavage methodology is noteworthy.

3.7: Sensitivity and generality

The effect of varying the allele frequency was also examined. Rather than mixing approximately equal amounts of two alleles, various fractions of the two alleles were mixed and assayed to simulate allele frequencies from 0.01 to 0.99. In this way the sensitivity of this method is measured. Frequencies as low as 5% are detectable above the background present, **figure 3.4**.

The sensitivity of this method is ultimately determined by the presence of background peaks. These peaks are present regardless of rhodium complexes or UV irradiation. Therefore this background has been tentatively assigned to PCR pause sites. If this background can be reduced, the sensitivity of the method can be enhanced still further.

Utilizing a general mismatch recognition complex should in theory allow for any SNP to be detected. In practice, however, neither Rhchrysi nor Rhphzi are able to detect all mismatches. Both complexes recognize mismatches because of their local destabilization within the DNA duplex. Thus in our assay, SNPs that lead to a stable

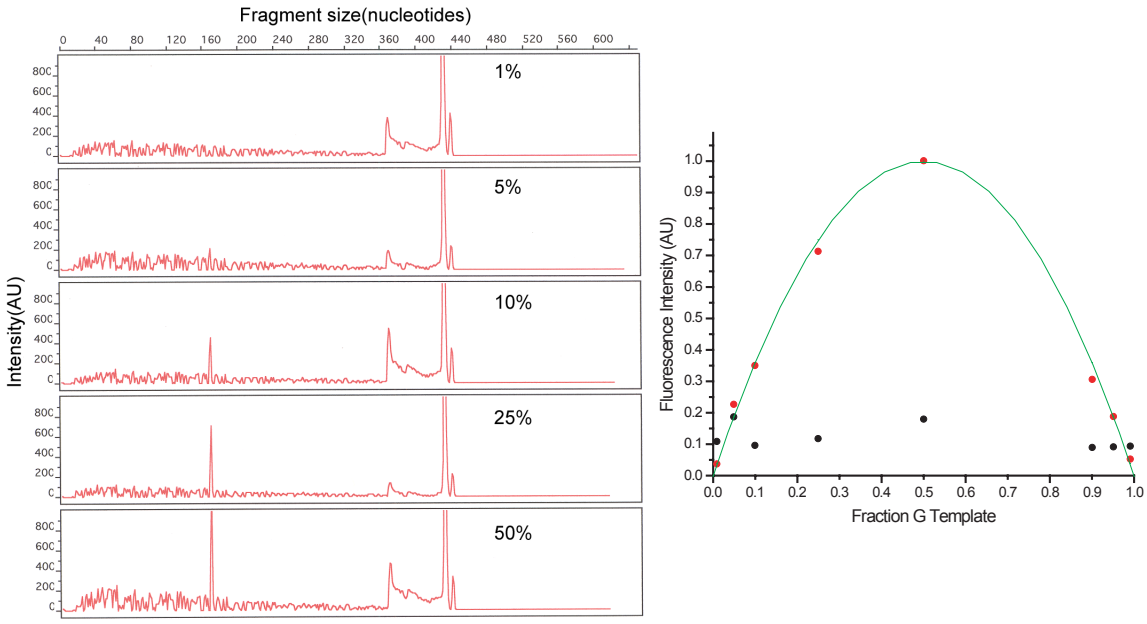

Figure 3.6: A plot showing the effect of varying the allele frequency on the fluorescent intensity of the cleaved products. The fraction of G template added to C template was varied over a range of 0.01 to 0.99 to imitate the effect of varying the allele frequencies on this assay. This fraction of G, which may be thought of as an allele frequency, is plotted on the horizontal axis. The normalized fluorescence intensity of the cleaved peak is plotted in red. The normalized background peak intensity is plotted in black. All fluorescence measurements are normalized to the uncleaved parent peak intensity. For fractions of 0.05 or greater, the cleaved peak exceeds background and can therefore be detected as an SNP. A line consistent with a fit to the statistical mixture is included to guide the eye.

mismatch are not detected. In general, this is not a significant problem, since in a statistical mixture, two different mismatches are formed from any SNP. One tends to be a particularly destabilized mismatch while the other is much less so. For example, G→C SNP will produce both a stable G-G mispair and a destabilized C-C mispair. While our complexes are not able to detect a G-G mispair, the C-C mispair is easily recognized and cleaved. Nonetheless, illustrated here is an alternate example where photocleavage is not noted. In the case of the A/C heterozygote neither mismatch generated is cleaved. The sequence surrounding the mismatch is 5'-AXA-3'/3'-TYT-5'. Here photocleavage is not observed owing to slippage and site stabilization. Specific photocleavage by the Rh intercalator therefore does indeed provide an unambiguous determination of the presence of a SNP. However, the lack of reaction does not establish the absence of an SNP. This assay is then expected to yield the rapid discovery of SNPs across the genome without the complication of false positives.

3.8: References

1. Erichsen, H. C., Chanock, S. J. (2004). SNPs in cancer research and treatment. *Br. J. Cancer* **90**, 747-751.

2. Vignal, A., Milan, D., SanCristobal, M., Eggen, A. (2002). A review on SNP and other types of molecular markers and their use in animal genetics. *Genet. Sel. Evol.* **34**, 275-305.

3. Carlson, C. S., Eberle, M. A., Rieder, M. J., Smith, J. D., Kruglyak, L., Nickerson, D. A. (2003). Additional SNPs and linkage-disequilibrium analyses are necessary for whole-genome association studies in humans. *Nat. Genet.* **33**, 518-521.

4. Cooper, D. N., Smith, B. A., Cooke, H. J., Niemann, S., Schmidtke, J. (1985). An

estimate of unique DNA sequence heterozygosity in the human genome. *Hum. Genet.* **69**, 201-205.

5. Botstein, D., Risch, N. (2003). Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat. Genet.* **33 Suppl**, 228-237.

6. Kruglyak, L., Nickerson, D. A. (2001). Variation is the spice of life. *Nat. Genet.* **27**, 234-236.

7. Kruglyak, L. (1999). Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat. Genet.* **22**, 139-144.

8. Patil, N., Berno, A. J., Hinds, D. A., Barrett, W. A., Doshi, J. M., Hacker, C. R., Kautzer, C. R., Lee, D. H., Marjoribanks, C., McDonough, D. P., Nguyen, B. T., Norris, M. C., Sheehan, J. B., Shen, N., Stern, D., Stokowski, R. P., Thomas, D. J., Trulson, M. O., Vyas, K. R., Frazer, K. A., Fodor, S. P., Cox, D. R. (2001). Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science* **294**, 1719-1723.

9. Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., Liu-Cordero, S. N., Rotimi, C., Adeyemo, A., Cooper, R., Ward, R., Lander, E. S., Daly, M. J., Altshuler, D. (2002). The structure of haplotype blocks in the human genome. *Science* **296**, 2225-2229.

10. Martin, A. M., Athanasiadis, G., Greshock, J. D., Fisher, J., Lux, M. P., Calzone, K., Rebbeck, T. R., Weber, B. L. (2003). Population frequencies of single nucleotide polymorphisms (SNPs) in immuno-modulatory genes. *Hum. Hered.* **55**, 171-178.

11. Kwok, P. Y., Deng, Q., Zakeri, H., Taylor, S. L., Nickerson, D. A. (1996). Increasing the information content of STS-based genome maps: identifying

polymorphisms in mapped STSs. *Genomics* **31**, 123-126.

12. Taillon-Miller, P., Piernot, E. E., Kwok, P. Y. (1999). Efficient approach to unique single-nucleotide polymorphism discovery. *Genome. Res.* **9**, 499-505.

13. Kwok, P. Y., Carlson, C., Yager, T. D., Ankener, W., Nickerson, D. A. (1994). Comparative analysis of human DNA variations by fluorescence-based sequencing of PCR products. *Genomics* **23**, 138-144.

14. Jackson, B. A., Barton, J. K. (1997). Recognition of DNA base mismatches by a rhodium intercalator. *J. Am. Chem. Soc.* **119**, 12986-12987.

15. Junicke, H., Hart, J. R., Kisko, J., Glebov, O., Kirsch, I. R., Barton, J. K. (2003). A rhodium(III) complex for high-affinity DNA base-pair mismatch recognition. *Proc. Natl. Acad. Sci. USA* **100**, 3737-3742.

16. Peyret, N., Seneviratne, P. A., Allawi, H. T., SantaLucia, J. (1999). Nearest-neighbor thermodynamics and NMR of DNA sequences with internal A-A, C-C, G-G, and T-T mismatches. *Biochemistry* **38**, 3468-3477.

17. Kielkopf, C. L., Erkkila, K. E., Hudson, B. P., Barton, J. K., Rees, D. C. (2000). Structure of a photoactive rhodium complex intercalated into DNA. *Nat. Struct. Biol.* **7**, 117-121.

18. Jackson, B. A., Barton, J. K. (2000). Recognition of base mismatches in DNA by 5,6-chrysenequinone diimine complexes of rhodium(III): a proposed mechanism for preferential binding in destabilized regions of the double helix. *Biochemistry* **39**, 6176-6182.

19. Jackson, B. A., Alekseyev, V. Y., Barton, J. K. (1999). A versatile mismatch recognition agent: specific cleavage of a plasmid DNA at a single base mispair. *Biochemistry* **38**, 4655-4662.

20. Pyle, A. M., Long, E. C., Barton, J. K. (1989). Shape-selective targeting of DNA by (phenanthrenequinone diimine)rhodium(III) photocleaving agents. *J. Am.*

Chem. Soc. **111**, 4520-4522.

21. Gotoh, O., Wada, A., Yabuki, S. (1979). Salt-concentration dependence of melting profiles of lambda phage DNAs: Evidence for long-range interactions and pronounced end effects. *Biopolymers* **18**, 805-824.

22. Baena, A., Leung, J. Y., Sullivan, A. D., Landires, I., Vasquez-Luna, N., Quinones-Berrocal, J., Fraser, P. A., Uko, G. P., Delgado, J. C., Clavijo, O. P., Thim, S., Meshnick, S. R., Nyirenda, T., Yunis, E. J., Goldfeld, A. E. (2002). TNF-alpha promoter single nucleotide polymorphisms are markers of human ancestry. *Genes Immun.* **3**, 482-487.

Chapter 4: Phosphatase Assisted DNA Modification

4.1: Introduction

In order to observe reactions of rhodium intercalators with DNA, a label must be incorporated into the DNA. This label can be either radioactive or fluorescent. Typically these are incorporated into DNA by an enzymatic method, although duplex DNA can be detected using fluorescent intercalators or groove binders without the need for modifying the DNA directly. The 5' terminus can be labeled with a radioactive phosphate by the kinase activity of T4 polynucleotide kinase, PNK.[1] The 3' end of DNA can be labeled either fluorescently or radioactively using terminal transferase.[2; 3] As an alternative to end labeling of DNA, labeling the DNA at the site cleaved by rhodium intercalators was explored.

Rhodium intercalators abstract a hydrogen atom from a sugar ring of the DNA backbone at the intercalation site.[4] Upon UV excitation, an electron is transferred from the intercalated ligand to the metal center, forming a ligand radical.[5] This radical is capable of abstracting a hydrogen atom from the sugar ring, forming a radical which decomposes to yield a strand break. Depending on which hydrogen is abstracted from the sugar ring different decomposition paths are possible. In the case of $[Rh(phen)_2(phi)]Cl_3$, primarily the 3' hydrogen is thought to be extracted. Irrespective of the hydrogen atom abstracted, sugar radicals primarily decompose to a 3'-phosphate terminated DNA and a 5'-phosphate terminated DNA with a base and sugar fragment missing. Other complexes are believed to react via a similar hydrogen atom abstraction mechanism, although the hydrogen removed may vary.

Rhodium complexes are not unique in their production of 3'-phosphate terminated DNA. 3'-phosphate lesions are formed from a variety of other damage events. DNA damage stemming from hydrogen atom abstraction at any sugar position results, at

least partially, in 3'-phosphate termini.[6] Typical sources of hydrogen atom abstraction are reactive oxygen species such as hydroxyl radical or metal bound reactive oxygens such as copper or manganese oxo complexes, x-rays and gamma rays, ene-diyne drugs and photoactivated rhodium intercalator complexes. Additionally, 3'-phosphate lesions are produced in the course of hydrolysis of DNA bases damaged by oxidation.[7] If DNA bases become oxidized and are treated with a strong base such as piperidine, this results in hydrolysis and β elimination which will produce a strand break with both a 3'-phosphate and 5'-phosphate termini. *In vivo,* enzymatic repair also results in 3'-phosphate termini.[8]

These 3'-phosphate lesions are problematic because they are not substrates for the wide variety of DNA modifying enzymes. In order to modify this terminus we must first remove the 3'-phosphate blocking lesion. This repair is suspected to occur *in vivo* by either the apurinic endonuclease, APE1,[9] or polynucleotide kinase 3'-phosphatase, PNKP.[8; 10] PNKP is thought to be involved with the repair of single and double strand breaks formed by hydrogen atom abstraction, primarily high energy radiation such as x-rays or gamma rays.[11] Since these enzymes are not commonly available, the more common T4-PNK enzyme has been used in these experiments.[12] T4-PNK is a bacteriophage enzyme which catalyzes the same reactions as PNKP of higher organisms. Bacteria which T4 phage infect attempt to sabotage the T4 replication by expressing an endonuclease capable creating of 3'-phosphate termini.

4.2: Experimental

Materials. Primers and oligonucleotides were synthesized on an ABI 394 DNA synthesizer using standard phosphoramidite chemistry. The DNA was purified using PolyPakII cartridges. Human genomic DNA, FastStart polyermase, calf intestinal phosphatase, shrimp alkaline phosphatase, were purchased from Roche. Rhodium

complexes were synthesized as previously described. ExoI, terminal transferase, and T4-PNK were purchased from New England Biolabs. Remaining chemicals were purchased from Sigma.

Demonstration of T4-PNK repair. 1 μM of *rac*-$[Rh(bpy)_2(chrysi)]^{3+}$ was added to 1 μM of ^{32}P labeled mismatched DNA duplex template, 5'-$^{32}P_i$-TTA GGA TCA T**C**C ATA TA-3' and 5'-TAT ATG **C**AT GAT CCT AA-3' dissolved in 50 mM NaCl with 10 mM sodium phosphate buffer pH 7.1. This mixture was irradiated for 15 minutes using a solar simulator. T4-PNK repair was performed by adding 3 μL of 10x PNK buffer (700 mM Tris-HCl, 100 mM $MgCl_2$, 5 mM DTT pH 7.6), 6 μL water, and 10 U T4-PNK. This mixture was incubated for 1 hour and the reaction stopped by heating to 90ºC for 15 minutes. The DNA samples were run on a denaturing polyacrylamide gel and imaged using phosphorimagery.

Fluorescent labeling of the cleaved site in a duplex. A pooled human genomic DNA sample was amplified using primers specific for the TNF promoter region, F:AGA GAT AGA ACA AAA GGA TAA GGG CTC AG and R:GTG TGG CCA TAT CTT CTT AAA CG using Roche FastStart High Fidelity polymerase according to standard proceedure. After polymerization, 20 U calf intestinal alkaline phosphatase and 6 U exonuclease I were added. The DNA was further purified by using a QIAgen PCR cleanup column and eluting in 10 mM Tris pH 8.0. This DNA was denatured, by heating to 99ºC for 20 minutes and annealed, by the addition of buffer and slow cooling to room temperature, to generate a final concentration of 20 mM Tris pH 7.0 and 100 mM NaCl; this denaturation and annealing generates mismatches. These mismatches were cleaved by 1 μM $[Rh(bpy)_2(chrysi)]^{3+}$ upon irradiation at 440 nm for 25 minutes. After cleavage, 80 U T4 polynucleotide kinase and 4 uL T4 ligase buffer were added to remove terminal

3'-phosphates. This mixture was then dried under reduced pressure and labeled using Applied Biosystems's SNaPshot kit following its procedures. The fluorescently labeled products were separated and detected using an ABI 310 prism capillary electrophoresis instrument.

Protocol used in testing phosphatase assisted trasferase tagging PCR(PATT-PCR). DNA was PCR amplified from two plasmids which were polymorphic at a single site, one containing the G allele, the other contained the C allele with F: CGC GTT GGC CGA TTA ATT AAT G and R: GCT GCG CAA CTG TTG GGA AG using Taq polymerase from Roche Biochemicals under standard conditions. After polymerization, 20 U calf intestinal alkaline phosphatase and 6 U exonuclease I were added. The DNA was further purified by using a QIAgen PCR cleanup column and eluting in 10 mM Tris pH 8.0. This DNA was denatured, by heating to 99ºC for 20 minutes and annealed, by the addition of buffer and slow cooling to room temperature, to generate a final concentration of 20 mM Tris pH 7.0 and 100 mM NaCl; this denaturation and annealing generates mismatches. These mismatches were cleaved by 1 μM $[Rh(bpy)_2(chrysi)]^{3+}$ upon irradiation at 440 nm for 25 minutes. After cleavage 80 U T4 polynucleotide kinase and 4 uL T4 ligase buffer were added to remove terminal 3'-phosphates. The DNA was ethanol precipitated and dried under reduced pressure to remove the T4 ligase buffer. The nicks in the DNA were tagged using terminal transferase under the following conditions: 400 U recombinant terminal transferase, 200 mM potassium cacodylate, 25 mM Tris HCl, 5 μg BSA, 0.75 mM $CoCl_2$, 5 μM dGTP in a total volume of 20 uL pH 6.6. The reaction mixture was incubated at 37C for 1 hour and stopped by denaturing at 70ºC for 20 minutes. The tagged products were then amplified by Taq polymerase in two reactions under standard conditions, except with an annealing temperature of 45ºC using the primers F, as described above, F and C_{15} or R and C_{15}. The products of this reaction were

then analyzed by agarose gel electrophoresis.

4.3: Repair of 3'-phosphate termini using T4-PNK

While the mechanism of cleavage has only been studied for a few rhodium complexes, the previous work with other intercalators has shown that 3'-phosphate termini occur in most hydrogen atom abstraction reactions.[4] Because the chrysi and phzi intercalating ligands have not been structurally characterized with DNA, we cannot predict which hydrogen will be abstracted. It could be possible that a different hydrogen atom would be extracted and hence different products could form. Despite this, most abstractions will result in 3'-phosphate termini to some extent.[6] Initially, validation of the 3'-phosphate terminus and repair using T4-PNK was sought.

A photocleavage reaction was performed with 1 μM $[Rh(bpy)_2(chrysi)]Cl_3$ with 1 μM C-C mismatched DNA, shown in **figure 4.1**. Products from photocleavage reactions were compared by denaturing gel electrophoresis and determined to have identical electrophoretic mobility as the Maxam-Gilbert reactions. Maxam-Gilbert reactions produce only 3'-phosphate termini upon reaction with piperidine. This suggests that 3'-phosphates are the major product in these reactions.

Subsequently, mass spectra of the products from the cleavage reactions were obtained by Jens Brunner confirming these observations as being bona fide 3'-phosphates. Shortly after photocleavage some metastable intermediates are formed, but after many hours these decompose to 3'-phosphate termini. It seems likely from these data that the hydrogen atom abstractions by chrysi and phzi complexes are the same, while the reaction using the phi ligand is distinct.

Following the confirmation of 3'-phosphate terminated DNA, enzymatic hydrolysis of the phosphate was attempted using T4-PNK. This reaction was found to proceed rapidly producing a slower migrating band which would be expected from the

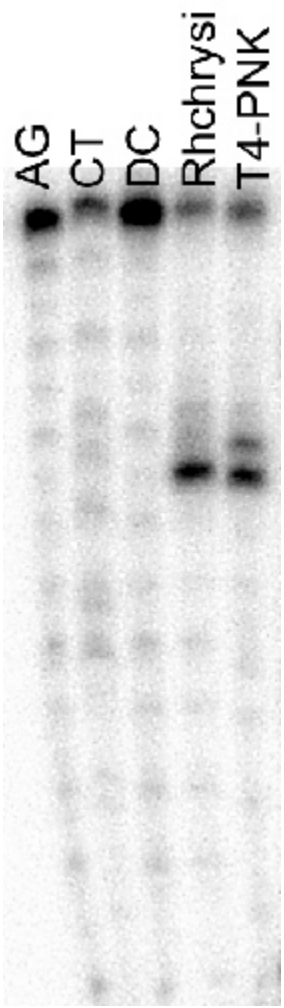

Figure 4.1: Demonstration of T4-PNK repair. An oligonucleotide duplex containing a C-C mismatch, DC, is cleaved with Rhchrysi. The cleaved band comigrates with the band in the AG Maxam-Gilbert reaction, indicating the presence of a 3'-phosphate. This can be removed to form a slower migrating band with T4-PNK.

decreased charge of the fragment. Using this reaction, it should be possible to repair the DNA at the cleavage site to make it a substrate for further enzymatic reaction

4.4: Utilizing 3'-phosphate repair for a new method of labeling DNA fragments

Using T4-PNK we are able to repair 3'-phosphate lesions to produce a 3'-hydroxyl end to the DNA. Once this repair is complete, we can use the many other enzymes, such as DNA polymerases, DNA ligases, and terminal transferase, to functionalize the DNA. In essence, we can now mark the site of damage within the DNA.

The initial application of this new method was the introduction of a fluorescent label at the site of damage. This allows a significant simplification of the procedure for the discovery of SNPs. The scheme outlined in **figure 4.2** shows how this method would function in practice. The DNA template would be created in a similar manner, but utilizing standard primers rather than restriction tagged primers. The DNA would then be denatured and annealed to produce mismatches at the sites of SNPs. Mismatch specific rhodium complexes would be added and irradiated. The 3'-phosphate terminated DNA strand is repaired by T4-PNK and finally labeled using the SNaPshot kit.

In practice, this method does work although only modestly. This method was used to detect a single low frequency SNP within the TNF gene. Under these conditions a new peak was detected 275 bases in length corresponding to the known SNP site in this sequence, as shown in **figure 4.3**. A standard restriction site labeled analysis is shown for comparison. The opposite side of the fragment is analyzed under these conditions so the peak sizes are necessarily different. Without the irradition of rhodium complex or without addition of PNK, no cleaved product is observed.

However, this method is not without its problems. Rhodium complexes cause the base 3' to the mismatched base to be eliminated, as illustrated in **figure 4.4**. After

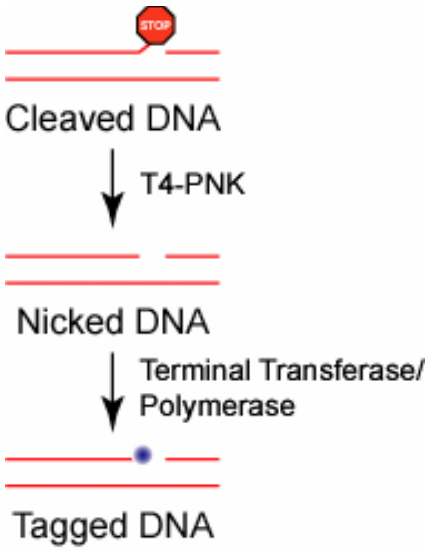

Figure 4.2: The incorporation of a fluorescent tag at the cleavage site. After oxidative cleavage of a site within the DNA, a 3'-phosphate remains. This phosphate can be removed with T4-PNK. After this the DNA becomes a substrate for labeling by polymerase or terminal transferase

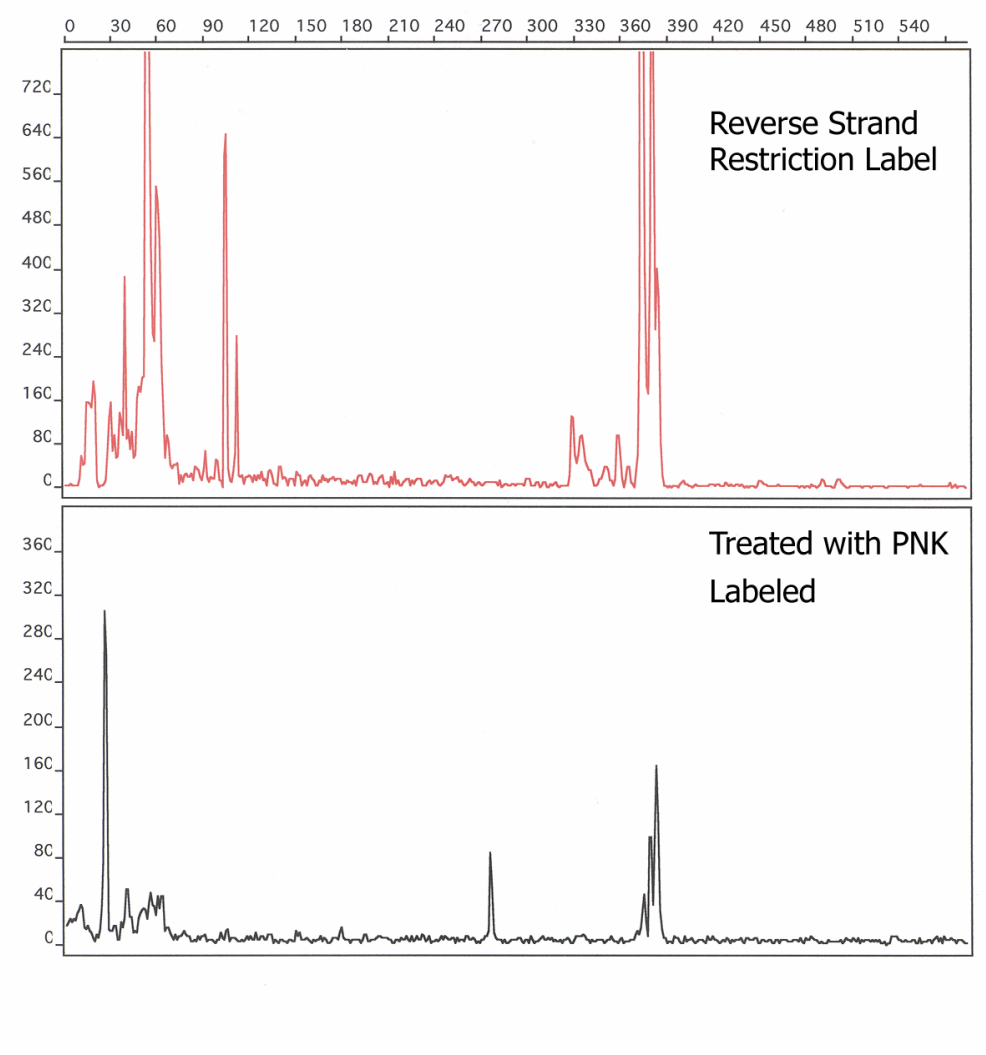

Figure 4.3: Fluorescent labeling at a $[Rh(bpy)_2(chrysi)]^{3+}$ cleavage site. A. A cleaved mismatch, 101 bp, is visualized by labeling at the 3' end of the DNA, shown for comparison. B. A label was introduced into the DNA at the cleavage site, 281 bp, after 3'-phosphate removal. No label incorporation is seen without photocleavage or phosphate removal.

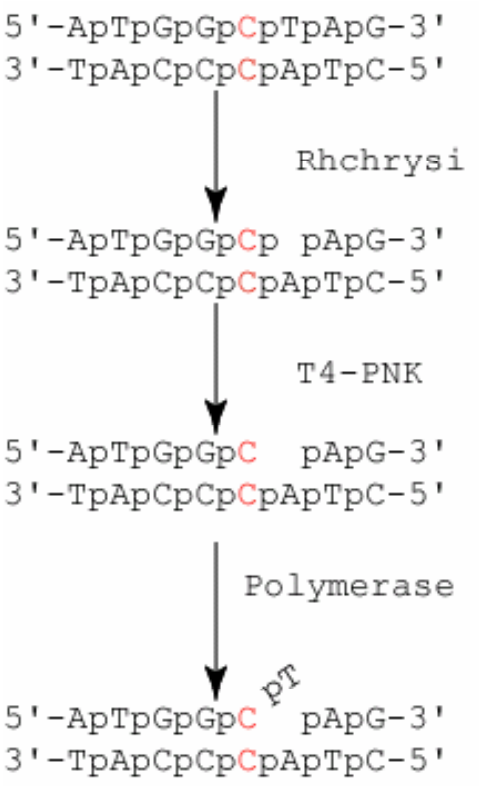

Figure 4.4: Poor labeling efficiency when labeling at the Rhchrysi cleavage site. When Rhchrysi cleaves a mismatch site, the base on the 3' side of the mismatch is actually removed. This leaves the mismatched base within the DNA. When polymerase is used to incorporate a fluorescent base, it must extend a terminally mismatched primer.

repair, this will leave the mismatched base as the 3'-hydroxyl terminus of the DNA. If we consider this situation was a primer and template in normal PCR, we have produced a mismatched primer and the biggest problem is that the mismatch corresponds to the last base of the primer. It is interesting that we see any extension of the mismatched base, but this non-ideal situation results in a 60-fold reduction in labeling efficiency. It is possible that if the T4-PNK were replaced by another enzyme, this problem could be bypassed. For example, APE1 catalyses the removal of 3'-phosphate termini by the hydrolysis of the 3' nucleotide and the phosphate. In this method, the mismatched base would be removed alleviating the problem.

4.5: Phosphatase assisted transferase tagging

The study of damaged sites in genomic DNA is a difficult subject. While it is possible to look for damaged DNA bases by HPLC methods, there are relatively few methods for localizing damaged sites within genomic DNA. Two methods currently exist which are capable of analyzing DNA damage on single nucleotide level. The first method is ligation mediated PCR, LM-PCR.[13] In LM-PCR, DNA is cleaved by some agent or DNA lesions are converted to strand breaks by piperidine or enzymatic treatment.[14; 15] A single primer is added which is specific for a single site in the genome of interest. This primer is bound and extended by polymerase to form a blunt end to the DNA. This is followed by a ligation of a known sequence to the blunt end of the DNA molecule and amplification of the DNA by PCR. This sounds simple and elegant, but in practice this method is difficult. The problem is typically in the blunt end ligation, which has a poor efficiency.

A more recent method is the terminal transferase dependent PCR.[16] In this method, primer extension is performed several times to produce some single stranded DNA. This extension depends on the ability of the polymerase chosen to occasionally

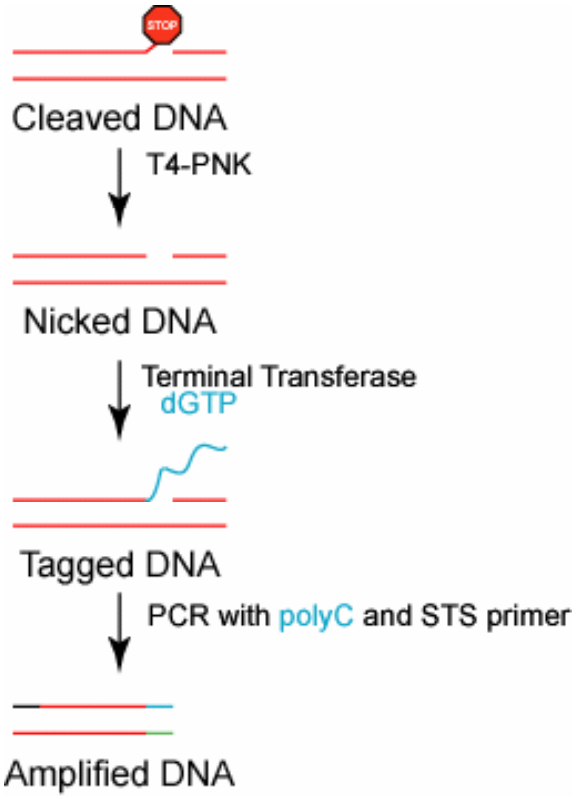

Figure 4.6: Phosphatase assisted transferase tagging PCR (PATT-PCR). Oxidatively cleaved DNA contains a 3'-phosphate terminus. This can be cleaved by T4-PNK to give an enzymatically active 3'-hydroxyl. A homopolymer tail can then be incorporated at the cleavage site, marking it. A primer complementary to the homopolymer tail and another primer can be used to observe the cleavage site by PCR.

stop or occasionally bypass any damaged sites in the DNA, without the addition of repair enzymes. The singled stranded DNA is then tailed with terminal transferase using ribonucleotides. These ribonucleotides are poor substrates for terminal transferase which can typically only incorporate a few, typically 3, bases. These 3 bases are annealed to a known DNA duplex with a complementary C tailed overhang. This then forms the primed complex for amplification. This method while an improvement over the LM-PCR method still suffers from some problems. The main issue is the polymerase pausing at damaged sites. Because polymerases vary in pausing at damaged sites, different results may be seen. Regardless of the type of DNA damage, oxidation, reduction, adduct or protein crosslink, these pauses may occur resulting in ambiguous data.

Using T4-PNK repair provides another possible route to localizing DNA damage, as shown in **figure 4.6**. DNA can be damaged *in vivo* through a variety of different established methods such as dimethylsulfate[13] or oxidative stress.[15] The DNA can be extracted from the cells using by lysis and purified from the other cellular components through established methods. The genomic DNA can then be digested with repair enzymes or treated simply with piperidine. Using terminal transferase, a polynucleotide tail comprised of a single DNA base can be incorporated at the newly repaired 3'-hydroxyl terminated DNA break, effectively labels the site. Then by performing PCR the product length will correspond to the damaged site. This method has several advantages over the previously described methods in that it does not involve an unpredictable primer extension step or a blunt end ligation. Additionally, this method allows for the observation of the 5' side of the break, while the other methods both analyze the 3' side.

In practice, this method has a complication. Terminal transferase will append a homopolymer tail without difficulty; however, the length of the tail will be heterogeneous. This problem is solved by the PCR amplification. Since we choose

a specific primer length, say 15 base pairs, the primer can bind anywhere within the homopolymer tail. For every amplification step, some of the homopolymer tail is lost until the substrate and primer lengths become equal. The lengthening of the homopolymer tail is not expected, as we can select a length and therefore annealing temperature at which primer binding will be permissible. Lengthening could only occur under conditions where fewer bases would be annealed and therefore the process would be less favorable than the desired annealing.

This method has been demonstrated in a model system. A PCR product of the TNF promoter region was prepared. This promoter region contains an SNP at a particular site which can be converted into a mismatch through denaturing and annealing. This mismatch can be cleaved by $[Rh(bpy)_2(chrysi)]Cl_3$ introducing a single strand oxidative nick at a known location within the DNA duplex. This oxidative nick contains a 3'-phosphate terminus which can be repaired by T4-PNK. Terminal transferase can then be used to attach a homopolymer tail at this site. Because a short DNA duplex was used rather than genomic DNA, there are two additional 3' termini which will also be tailed at the end of the duplex. PCR can then be used to detect the products from either the forward primer site or the reverse site.

An example of this procedure is shown in **figure 4.7**. In this experiment the DNA duplex was tagged with terminal transferase using different nucleotides. There is a definite new product ~280 bp in length when the forward primer is used. This seems to correspond with the position of the SNP plus the additional 15 bases from the homopolymer tail. Using the reverse primer, little cleavage is observed as would be expected. Using higher resolution capillary electrophoresis analysis, the new peak is 274 bases in length which corresponds exactly with the SNP site, shown in **figure 4.8**. The full length product can also be seen at 440 bases.

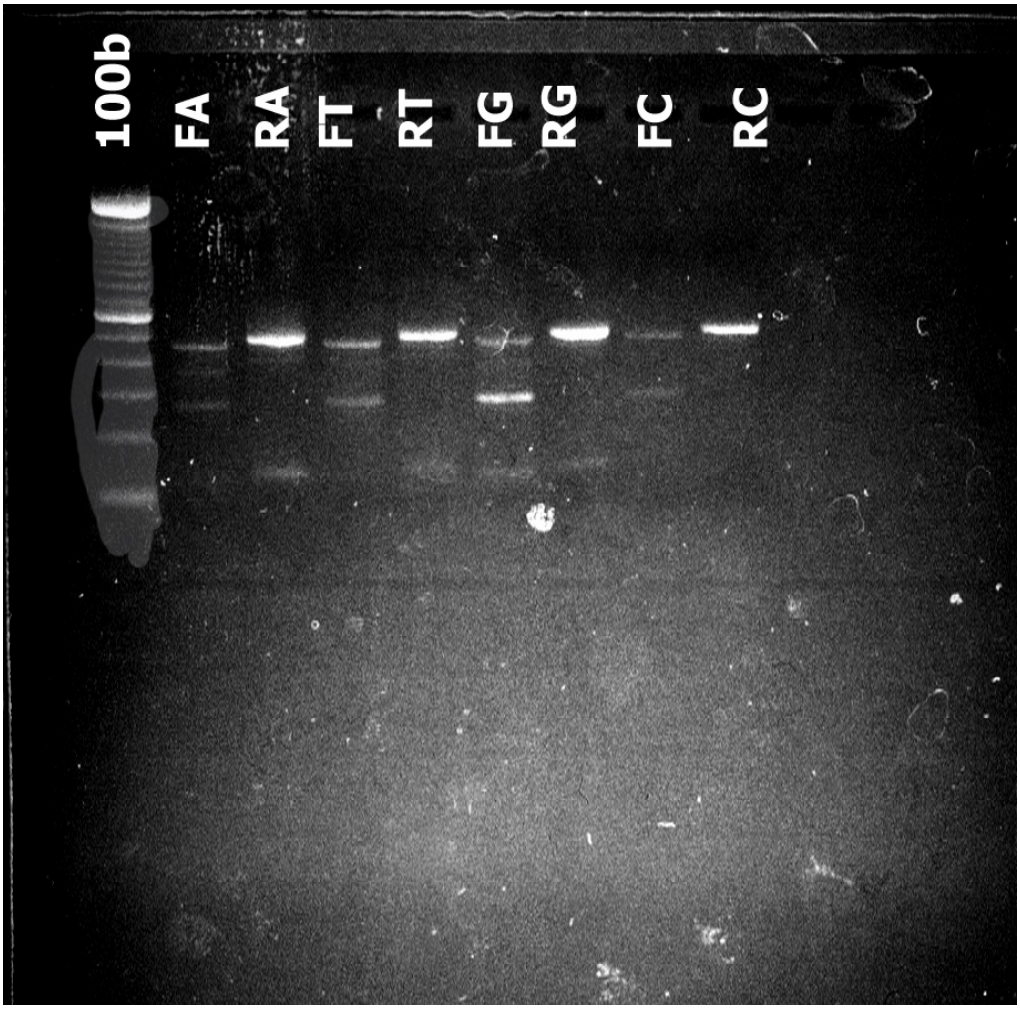

Figure 4.7: PATT-PCR by agarose gel. A SNP within the TNF promoter was cleaved by a mismatch specific complex as described within the text. PATT-PCR was performed using all possible homopolymer tails, poly G tailing being the obvious best. There a new band is clearly visible at ~300 base pairs in the forward primer reactions.

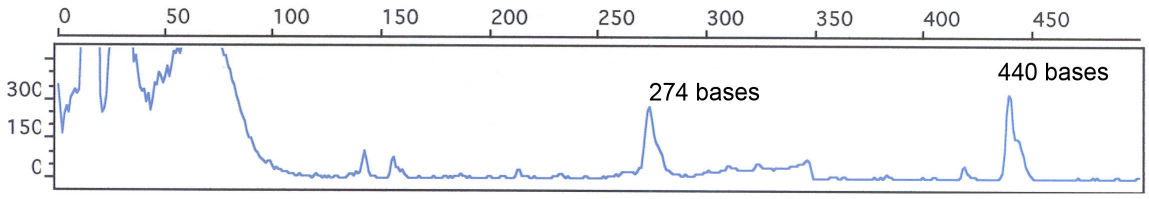

Figure 4.8: Analysis of a PATT-PCR reaction by capillary electrophoresis. An identical reaction to that performed in **figure 4.7** is shown here using capillary electrophoresis. The new band is 274 bases in length which exactly corresponds to the SNP site in the TNF gene. The peak at 440 bases corresponds to the PCR template used in the reaction. Peaks below 100 bases are from incomplete extension.

4.6: References

1. Richardson, C. C. (1965). Phosphorylation of nucleic acid by an enzyme from T4 bacteriophage-infected E*scherichia coli*. *Proc. Natl. Acad. Sci. U S A* **54**, 158.

2. Hayes, F. N., Mitchell, V. E., Ratliff, R. L., Schwartz, A. W., Williams, D. L. (1966). Incorporation efficiency of small oligo-5'-nucleotide initiators in terminal deoxyribonucleotide transferase reaction. *Biochemistry* **5**, 3625.

3. Trainor, G. L., Jensen, M. A. (1988). A procedure for the preparation of fluorescence-labeled DNA with terminal deoxynucleotidyl transferase. *Nucleic Acids Res.* **16**, 11846-11846.

4. Sitlani, A., Long, E. C., Pyle, A. M., Barton, J. K. (1992). DNA photocleavage by phenanthrenequinone diimine complexes of rhodium(III): Shape-selective recognition and reaction. *J. Am. Chem. Soc.* **114**, 2303-2312.

5. Turro, C., Hall, D. B., Chen, W., Zuilhof, H., Barton, J. K., Turro, N. J. (1998). Solution photoreactivity of phenanthrenequinone diimine complexes of rhodium and correlations with DNA photocleavage and photooxidation. *J. Phys. Chem. A* **102**, 5708-5715.

6. Pogozelski, W. K., Tullius, T. D. (1998). Oxidative strand scission of nucleic acids: Routes initiated by hydrogen abstraction from the sugar moiety. *Chem. Rev.* **98**, 1089-1107.

7. Burrows, C. J., Muller, J. G. (1998). Oxidative nucleobase modifications leading to strand scission. *Chem. Rev.* **98**, 1109-1151.

8. Wiederhold, L., Leppard, J. B., Kedar, P., Karimi-Busheri, F., Rasouli-Nia, A., Weinfeld, M., Tomkinson, A. E., Izumi, T., Prasad, R., Wilson, S. H., Mitra, S., Hazra, T. K. (2004). AP endonuclease-independent DNA base excision repair in human cells. *Mol. Cell.* **15**, 209-220.

9. Parsons, J. L., Dianova, I. I., Dianov, G. L. (2004). APE1 is the major 3 '-

phosphoglycolate activity in human cell extracts. *Nucleic Acids Res.* **32**, 3531-3536.

10. Jilani, A., Ramotar, D., Slack, C., Ong, C., Yang, X. M., Scherer, S. W., Lasko, D. D. (1999). Molecular cloning of the human gene, PNKP, encoding a polynucleotide kinase 3'-phosphatase and evidence for its role in repair of DNA strand breaks caused by oxidative damage. *J. Biol. Chem.* **274**, 24176-24186.

11. Vance, J. R., Wilson, T. E. (2001). Repair of DNA strand breaks by the overlapping functions of lesion-specific and non-lesion-specific DNA 3'-phosphatases. *Mol. Cell. Biol.* **21**, 7191-7198.

12. Wang, L. K., Lima, C. D., Shuman, S. (2002). Structure and mechanism of T4 polynucleotide kinase: An RNA repair enzyme. *EMBO J.* **21**, 3873-3880.

13. Mueller, P. R., Wold, B. (1989). *In vivo* footprinting of a muscle specific enhancer by ligation mediated PCR. *Science* **246**, 780-786.

14. Denissenko, M. F., Pao, A., Tang, M. S., Pfeifer, G. P. (1996). Preferential formation of benzo[a]pyrene adducts at lung cancer mutational hotspots in P53. *Science* **274**, 430-432.

15. Rodriguez, H., Akman, S. A., Holmquist, G. P., Wilson, G. L., Driggers, W. J., LeDoux, S. P. (2000). Mapping oxidative DNA damage using ligation-mediated polymerase chain reaction technology. *Methods* **22**, 148-156.

16. Komura, J., Riggs, A. D. (1998). Terminal transferase-dependent PCR: A versatile and sensitive method for in vivo footprinting and detection of DNA adducts. *Nucleic Acids Res.* **26**, 1807-1811.

Chapter 5: Mismatches in Genomic DNA

5.1: Introduction to mismatch repair and genomic mismatches

Cells are equipped with DNA repair machinery that systematically repairs various lesions within DNA. Mismatch repair, MMR, is responsible for the detection and repair of base pair mismatches within the genome.[1] These proteins are essential for the maintenance of genomic DNA and if the MMR system is disrupted through mutation[2; 3] or epigenetic silencing[4] the mutation rate within the cell increases dramatically.

The human mismatch repair machinery is composed of several key proteins.[1] First, DNA mismatches are detected by a MutS homolog, either MutSα or MutSβ, composed of MSH2 and MSH6 or MSH2 and MSH3, respectively. The two forms of MutS have somewhat different recognition properties, MutSα is believed to target mainly mismatches while MutSβ is believed to target primarily insertion deletion loops. After recognition of a mismatch by MutS, a MutL homolog binds to MutS are coordinates the repair. At least two forms of MutL are present in humans, MutLα and MutLβ, formed from MLH1 and either PMS2 or PMS1. MLH1 has been shown to form heterodimers with other proteins as well. The function of these various forms of MutL is not well understood.

Because MSH2 or MLH1 are involved in all mismatch repair, their loss is detrimental to the entire process of mismatch repair.[5] *In vivo*, mutations in these genes can cause both an increase in mutation rate, the mutator phenotype,[2] and instability of short repeats of DNA, microsatellite instability.[6] These two types of genetic instability within humans are responsible for hereditary non-polyposis colon cancer or HNPCC.[7] Additionally, 18% of all solid tumors[8] and a significant number of leukemias[9] also are mismatch repair deficient. The loss of the other mismatch repair proteins can also cause increases in certain types of genetic instability which may also lead to cancer.[10]

In addition to being a predisposing factor for cancer, loss of mismatch repair also confers resistance to many common chemotherapeutics.[11] Mismatch repair deficient tumors have been reported to be resistant to alkylating agents[12], platinum drugs[13] and metabolic analogs[14]. It is believed that the mismatch repair proteins are involved to some extent with the recognition of these lesions. The recognition of these lesions then would normally cause the induction of apoptosis.[15]

Typically, the way in which mismatch repair deficiency is assessed is by investigating several microsatellite regions in the genome.[16] These are amplified by PCR and length of the microsatellite is investigated by gel electrophoresis. Additionally, immunological methods for the detection of MSH2 or MLH1 are sometimes also employed. Depending of the microsatellites selected for assay, different outcomes can be obtained.[17]

Within our group we have developed bulky rhodium metallointercalators which can bind mismatches in DNA, shown in **figure 5.1**. These compounds bind mismatches by intercalation and upon UV photoexcitation can cleave the DNA at the mismatch site.[18; 19] Our first generation complex $[Rh(bpy)_2(chrysi)]Cl_3$, Rhchrysi, can bind mismatches with moderate affinity and high specificity.[18] The high specificity of Rhchrysi is beautifully demonstrated in its ability to cleave a single mismatched site in a 2700 base pair duplex.[20] The second generation complex $[Rh(bpy)_2(phzi)]Cl_3$, Rhphzi, utilized heterocyclic nitrogens to boost the affinity of the complex.[19] Rhphzi is capable of binding mismatched sites with 100 nM affinity and similar specificity to Rhchrysi. Using these complexes, a method for the detection of mismatched DNA bases in genomic DNA has been developed.

Previously, a method for the detection of mismatches within genomic DNA was developed using these mismatch specific rhodium complexes.[19] In this method, genomic DNA, extracted from cell cultures, was cleaved by a restriction enzyme into smaller

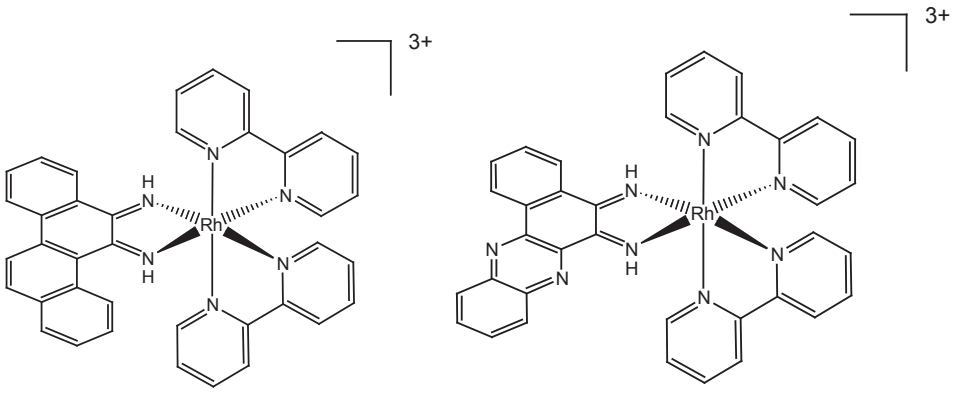

Figure 5.1: Structures of Rhchrysi and Rhphzi. Rhchrysi is our first generation mismatch binding molecule. It is capable of binding to 80% of the possible mismatches with moderate affinity. Rhphzi is a higher affinity complex which has similar recognition properties to Rhchrysi.

pieces with an average length of 4 kb. These fragments were then labeled with ^{32}P-phosphate using T4-PNK. This genomic DNA was then cleaved by either Rhchrysi or Rhphzi and light. The fragments were then separated using an alkaline agarose gel and visualized with phosphorimagery. The change in the fragment distribution was then used to quantify the degree to which mismatches were contained within the genomic DNA.

This method while effective is difficult in both performance and analysis. A simpler scheme was envisioned wherein a radioactive phosphate or fluorescent tag would be installed at the cleavage site. This would provide a method for a new straightforward analysis of mismatches within genomic DNA. This method may prove to be an invaluable tool in the early detection of HNPCC and other MMR deficient cancers as mismatched bases occur with higher frequency than insertion/deletion loops.

5.2: Experimental

Cell lines, Media and Culture. The DU145 and SW620 cell lines were obtained from the ATCC. Cells were grown in Leibovitz L-15 for SW620 or Minimum Essential Media (MEM) alpha modification for DU145, supplemented with 10% fetal bovine serum, 100 U/mL penicillin, 100 ug/mL streptomycin. Cells were grown in Corning Costar tissue culture flasks and dishes at 37°C under 5% CO_2 atmosphere.

Isolation of DNA. DNA was isolated from cultured cells using the QIAgen DNeasy kit and protocol. The DNA was quantitated by UV spectroscopy and diluted to 100 μM DNA bases.

Labeling at mismatched sites. Isolated cellular DNA is placed into a solution with 500 nM Rhchrysi, 50 mM NaCl, and 10 mM Tris pH 7.0 in a final reaction volume

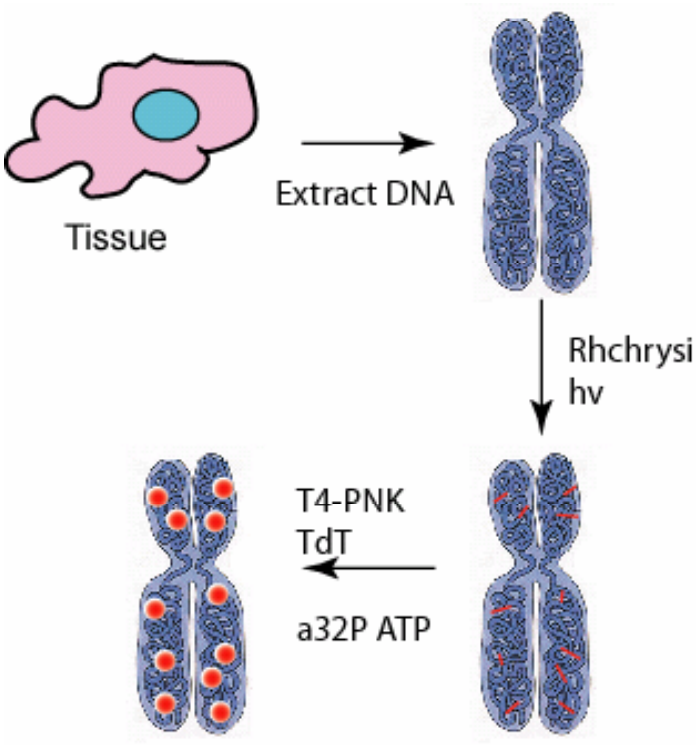

Figure 5.1: Detection of mismatches in genomic DNA. Genomic DNA is extracted from cells grown in cell culture. The genomic DNA is mixed with the mismatch specific metal complex Rhchrysi which cleaves DNA at mismatched sites with UV irradiation. Radioactivity is incorporated at the cleaved sites using terminal transferse. The radioactivity is then quantitated by scintillation counting.

of 20 uL. These are irradiated with a solar simulator, UVB/UVC filtered, for 15 minutes. To the 20 uL reaction, 5 uL of 5x of a special PNK buffer is added which contains 50 mM MgOAc, 100 mM Tris OAc pH 7.9, 5 mM DTT, and 5 mM ATP. To this 10 U, in 1 uL, of PNK is added and allowed to react 1 hour at 37 ºC. After this point 10 U, in 1 uL, of SAP are added to destroy the ATP in the PNK buffer. This is allowed to react for 1 hour followed by inactivation of PNK and SAP at 75 ºC for 15 minutes. After this the reaction mixture is adjusted for the terminal transferase reaction by the addition of 3 uL of a 10x terminal transferase additive buffer which contains 500 mM KOAc and 2.5 mM $CoCl_2$. Along with this 20 U, in 1 uL, of terminal transferase and 1 uL of α^{32}P-ddATP are added. Following this reaction, 10 U of SAP are again added to convert the ^{32}P into inorganic phosphate which is more easily removed.

Quantification of ^{32}P incorporated into DNA. Radioactive phosphate incorporation was analyzed by the acid precipitation of DNA onto filter paper.[21] Approximately 1 cm^2 pieces of filter paper are cut and spotted with 10 uL of the reaction mixtures. The paper are numbered with indelible marks, either graphite pencil or laser printed numbers, and placed into a 5% trichloroacetic acid, TCA, solution to precipitate onto the filter paper in a batch precipitation process in an Erlenmeyer flask. Squares without applied radioactivity are used as a control the acid precipitation process. The unreacted ATP which was converted to inorganic phosphate dissolves easily in the acid solution. This is rinsed 3x with the TCA solution then dehydrated with absolute ethanol. As the solution drains from the flask radioactivity is monitored by Geiger counter. If significant radioactivity remains in solution after the third rinse with TCA, more rinses may be performed. The filter paper is then allowed to dry and precipitated radioactivity is quantitated by dry scintillation by Cherenkov quenching.

5.3: Quantification of mismatches in genomic DNA using terminal transferase

As an alternative to the alkaline agarose gel method, a new method was developed. Instead of looking at the shifts of labeled pieces of DNA, DNA is labeled with radioactivity at the sites were cleavage occurs. This provides a move straightforward method for the detection of mismatched sites within the DNA.

Genomic DNA is isolated from cells grown in culture. This genomic DNA is then cleaved using a mismatch specific complex such as Rhchrysi. Because of the mechanism of cleavage, the cleaved site will have a 3'-phosphate which is enzymatically inactive. The 3'-phosphate is removed using T4-PNK. After removal, the nick can be radioactively labeled using terminal transferase and α^{32}P-ddATP. The DNA is precipitated onto a small piece of filter paper and quantitated by scintillation counting.

Two different cell lines were grown to produce genomic DNA. The SW620 cell line is an epithelial cell line from colorectral adenocarcinoma.[22] This cell line is mismatch repair-proficient and has been shown to have normal to high amounts of the key MSH2 and MLH1 proteins.[23] The other cell line is DU145 which is an epithelial cell line from a prostate carcinoma.[24] This cell line is mismatch repair-deficient and has no MLH1 and low MSH2.[25] These two cell lines were used for testing of this method of detecting mismatches in genomic DNA.

This method was first tested using the nonspecific DNA intercalator $[Rh(bpy)_2(phi)]^{3+}$. This complex should bind equally well to genomic DNA with or without DNA mismatches and should consequently shown similar amounts of ^{32}P incorporation. Experiments using this complex were conducted at a range of concentrations which correspond to 1/10000 base pairs, 1/1000 base pairs, 1/100 base pairs or 1/10 base pairs. As can be seen from **figure 5.2**, both DU145 and SW620 show similar levels of radioactive incorporation as should be expected. When large amounts of rhodium complex are used, the amount of radioactivity incorporated decreases. This is

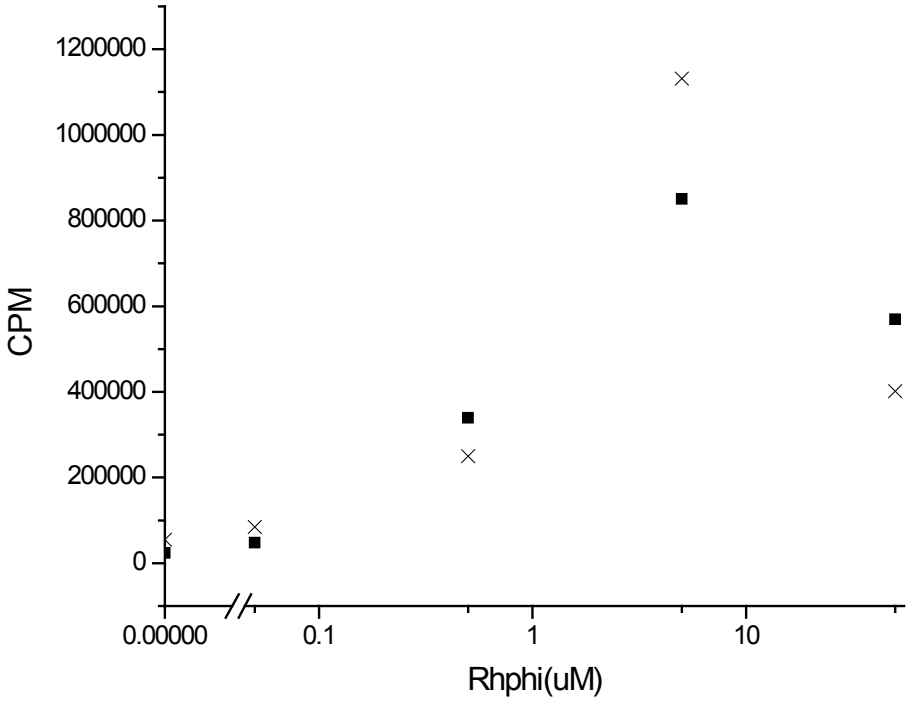

Figure 5.2: Incorporation of radioactivity after photocleavage using the nonspecific Rhphi complex. If a nonspecific DNA cleavage agent is used similar amounts of radioactivity are incorporated into the genomic DNA of both DU145 and SW620. Genomic DNA was extracted from the cell lines. The DNA was dissolved in 50mM NaCl, 10mM Tris pH 7.0. DNA concentrations were adjusted to 300mM bases.

likely because the genomic DNA is extremely damaged and cleaved to small fragments. This has been shown to decrease the efficiency of nucleotide incorporation by terminal transferase.[26]

When genomic DNA was cleaved by the mismatch-specific complex $[Rh(bpy)_2(chrysi)]^{3+}$, radioactivity is only incorporated into the genomic DNA of the DU145 mismatch repair deficient cell line, as shown in **figure 5.3**. The amount of radioactivity increases with increasing amounts of $[Rh(bpy)_2(chrysi)]^{3+}$. At the same time the genomic DNA of SW620 does not show any increase in radioactivity incorporation with $[Rh(bpy)_2(chrysi)]^{3+}$ of any concentration. This is surprising as at high concentrations of rhodium complex, some nonspecific binding would be expected based purely upon the nonspecific binding constant. This should cause some cleavage and labeling of the SW620 cell line. However, binding and photocleavage efficiency are not one in the same. Upon photoexcitation of rhodium complexes a short lived ligand radical forms which is capable of hydrogen atom abstraction.[27] If this radical is not positioned correctly, as is likely the case with nonspecific binding, hydrogen atom abstraction does not occur as readily as in the case of mismatch binding.

If the amounts of radioactivity are compared between $[Rh(bpy)_2(chrysi)]^{3+}$ and $[Rh(bpy)_2(phi)]^{3+}$ cleavage experiments, the number of $[Rh(bpy)_2(chrysi)]^{3+}$ binding sites per base pair of genomic DNA can be estimated. In this case, the data suggest that 1 base in 3000 base pairs forms a site for $[Rh(bpy)_2(chrysi)]^{3+}$ binding.

The amount of radioactivity incorporated by this method will be proportional to the number of breaks within the DNA from all sources. The genomic DNA is sheared to some extent by its isolation. As no precautions are taken to reduce shearing, typically the DNA is fragmented into 50–100 kb pieces. If cells are experiencing either necrosis or apoptosis, the genomic DNA may be extremely fragmented which will skew the data substantially.

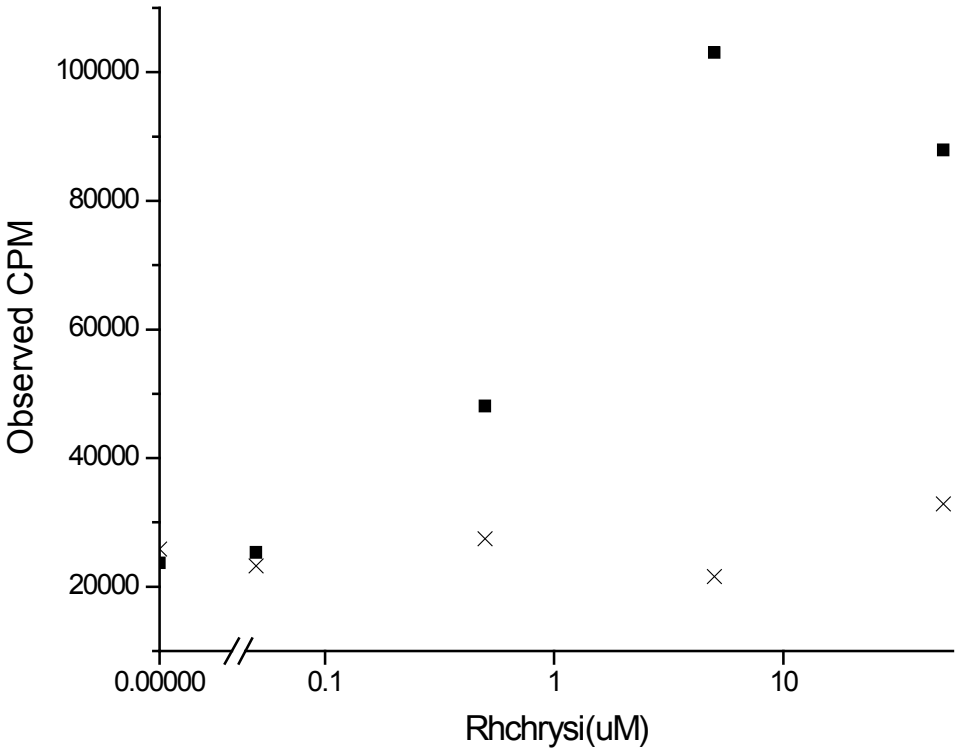

Figure 5.3: Radioactivity incorporated using the mismatch specific complex Rhchrysi. Increasing amounts of radioactivity are into the genomic DNA from the mismatch repair deficient DU145 cell line, with increasing amounts of Rhchrysi. Under the same conditions essentially no radioactivity above background is incorporated into the genomic DNA of the mismatch repair proficient SW620 cells. The DNA was dissolved in 50 mM NaCl, 10 mM Tris pH 7.0. The concentration of DNA was adjusted to 300 mM bases. The concentration of Rhchrysi is as indicated. Samples were irradiated for 15 minutes and labeled according to the experimental proceedure.

These data is in direct conflict with previously measured mutation rates in mismatch repair deficient cell lines which place mutation rates at ~1 in 10^6/base/generation.[2; 28] These experiments used loss of function methodology where *hprt*, necessary for 6-thioguanine to be toxic, must become mutated in order for the cell to survive. Since hprt is located on the X chromosome there is only one copy of the gene which must become mutated. These experiments are generally neutral to the type of mutation, transition, transversion or frameshift, but only mutations at particular sites can produce a loss of function mutant. Because only a small fracation of the bases within the gene are functionally assessed, one mutation does correspond to a single mismatch, but one mismatch may or may not correspond to a loss of function mutation.

Recent data have suggested that much higher mutation rates may occur under non-ideal growth conditions. One such non-ideal condition is oxidative stress. Mismatch repair deficient cell lines have been shown to have greatly enhanced mutation rates under oxidative conditions, presumably because mismatch repair does repair some oxidized lesions.[29; 30] These mutation rates only tell us about the number of unrepaired mutagenic lesions within the genomic DNA after a whole cell cycle worth of DNA repair. It seems reasonable based upon the efficiency of base excision repair of oxidized lesions within DNA that the number of lesions may be greater by an order of magnitude or possibly more than the mutation rate.

The DU145 cell line is considered incompetent for mismatch repair because of the complete lack of the MLH1 protein. Despite this cellular extracts from DU145 are able to repair a variety of mismatches in an *in vitro* repair assay.[25] Other mismatch repair deficient cell lines have also shown this depressed mismatch repair activity.[2; 6] These observations suggest that perhaps they are not completely mismatch repair deficient but rather mismatch repair compromised. Perhaps higher levels of mismatches occur than the mutation rate alone indicates. No direct measurements of mismatches in genomic DNA

have been made previously so perhaps there are more than we have previously expected.

5.4: References

1. Kunkel, T. A. & Erie, D. A. (2005). DNA mismatch repair. *Annu. Rev. Biochem.* **74**, 681-710.

2. Parsons, R., Li, G. M., Longley, M. J., Fang, W. H., Papadopoulos, N., Jen, J., Delachapelle, A., Kinzler, K. W., Vogelstein, B. & Modrich, P. (1993). Hypermutability and mismatch repair deficiency in RER+ tumor-cells. *Cell* **75**, 1227-1236.

3. Parker, A. R., Leonard, C. P., Hua, L., Francis, R. O., Dhara, S., Maitra, A. & Eshleman, J. R. (2004). A subgroup of microsatellite stable colorectal cancers has elevated mutation rates and different responses to alkylating and oxidising agents. *Br. J. Cancer* **90**, 1666-1671.

4. Herman, J. G., Umar, A., Polyak, K., Graff, J. R., Ahuja, N., Issa, J. P. J., Markowitz, S., Willson, J. K. V., Hamilton, S. R., Kinzler, K. W., Kane, M. F., Kolodner, R. D., Vogelstein, B., Kunkel, T. A. & Baylin, S. B. (1998). Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. *Proc. Natl. Acad. Sci. U S A* **95**, 6870-6875.

5. Peltomaki, P. (2001). Deficient DNA mismatch repair: a common etiologic factor for colon cancer. *Hum. Mol. Genet.* **10**, 735-740.

6. Umar, A., Boyer, J. C., Thomas, D. C., Nguyen, D. C., Risinger, J. I., Boyd, J., Ionov, Y., Perucho, M. & Kunkel, T. A. (1994). Defective mismatch repair in extracts of colorectal and endometrial cancer cell-lines exhibiting microsatellite instability. *J. Biol. Chem.* **269**, 14367-14370.

7. Papadopoulos, N. & Lindblom, A. (1997). Molecular basis of HNPCC: Mutations

of MMR genes. *Hum. Mutat.* **10**, 89-99.

8. Arzimanoglou, I. I., Gilbert, F. & Barber, H. R. K. (1998). Microsatellite instability in human solid tumors. *Cancer* **82**, 1808-1820.

9. Karran, P., Offman, J. & Bignami, M. (2003). Human mismatch repair, drug-induced DNA damage, and secondary cancer. *Biochimie* **85**, 1149-1160.

10. Kolodner, R. D., Tytell, J. D., Schmeits, J. L., Kane, M. F., Das Gupta, R., Weger, J., Wahlberg, S., Fox, E. A., Peel, D., Ziogas, A., Garber, J. E., Syngal, S., Anton-Culver, H. & Li, F. P. (1999). Germ-line MSH6 mutations in colorectal cancer families. *Cancer Res.* **59**, 5068-5074.

11. Fink, D., Aebi, S. & Howell, S. B. (1998). The role of DNA mismatch repair in drug resistance. *Clin. Cancer Res.* **4**, 1-6.

12. Karran, P. (2001). Mechanisms of tolerance to DNA damaging therapeutic drugs. *Carcinogenesis* **22**, 1931-1937.

13. Aebi, S., KurdiHaidar, B., Gordon, R., Cenni, B., Zheng, H., Fink, D., Christen, R. D., Boland, C. R., Koi, M., Fishel, R. & Howell, S. B. (1996). Loss of DNA mismatch repair in acquired resistance to cisplatin. *Cancer Res.* **56**, 3087-3090.

14. Stojic, L., Brun, R. & Jiricny, J. (2004). Mismatch repair and DNA damage signalling. *DNA Repair* **3**, 1091-1101.

15. O'Brien, V. & Brown, R. (2006). Signalling cell cycle arrest and cell death through the MMR System. *Carcinogenesis* **27**, 682-692.

16. Rodriguez-Bigas, M. A., Boland, C. R., Hamilton, S. R., Henson, D. E., Jass, J. R., Khan, P. M., Lynch, H., Perucho, M., Smyrk, T., Sobin, L. & Srivastava, S. (1997). A National Cancer Institute workshop on hereditary nonpolyposis colorectal cancer syndrome: Meeting highlights and Bethesda guidelines. *J. Natl. Cancer Inst.* **89**, 1758-1762.

17. Hatch, S. B., Lightfoot, H. M., Garwacki, C. P., Moore, D. T., Calvo, B. F.,

Woosley, J. T., Sciarrotta, J., Funkhouser, W. K. & Farber, R. A. (2005). Microsatellite instability testing in colorectal carcinoma: Choice of markers affects sensitivity of detection of mismatch repair-deficient tumors. *Clin. Cancer Res.* **11**, 2180-2187.

18. Jackson, B. A. & Barton, J. K. (1997). Recognition of DNA base mismatches by a rhodium intercalator. *J. Am. Chem. Soc.* **119**, 12986-12987.

19. Junicke, H., Hart, J. R., Kisko, J., Glebov, O., Kirsch, I. R. & Barton, J. K. (2003). A rhodium(III) complex for high-affinity DNA base-pair mismatch recognition. *Proc. Natl. Acad. Sci. U S A* **100**, 3737-3742.

20. Jackson, B. A., Alekseyev, V. Y. & Barton, J. K. (1999). A versatile mismatch recognition agent: Specific cleavage of a plasmid DNA at a single base mispair. *Biochemistry* **38**, 4655-4662.

21. Bollum, F. J. (1959). Thermal Conversion of Nonpriming Deoxyribonucleic Acid to Primer. *J. Biol. Chem.* **234**, 2733-2734.

22. Leibovitz, A., Stinson, J. C., Mccombs, W. B., Mccoy, C. E., Mazur, K. C. & Mabry, N. D. (1976). Classification of human colorectal adenocarcinoma cell lines. *Cancer Res.* **36**, 4562-4569.

23. Taverna, P., Liu, L., Hanson, A. J., Monks, A. & Gerson, S. L. (2000). Characterization of MLH1 and MSH2 DNA mismatch repair proteins in cell lines of the NCI anticancer drug screen. *Cancer Chemother. Pharmacol.* **46**, 507-516.

24. Stone, K. R., Mickey, D. D., Wunderli, H., Mickey, G. H. & Paulson, D. F. (1978). Isolation of a Human Prostate Carcinoma Cell Line (Du 145). *Int. J. Cancer* **21**, 274-281.

25. Yeh, C. C., Lee, C. & Dahiya, R. (2001). DNA mismatch repair enzyme activity and gene expression in prostate cancer. *Biochem. Biophys. Res. Commun.* **285**, 409-413.

26. Hayes, F. N., Mitchell, V. E., Ratliff, R. L., Schwartz, A. W. & Williams, D. L. (1966). Incorporation efficiency of small oligo-5'-nucleotide initiators in terminal deoxyribonucleotide transferase reaction. *Biochemistry* **5**, 3625.

27. Sitlani, A., Long, E. C., Pyle, A. M. & Barton, J. K. (1992). DNA photocleavage by phenanthrenequinone diimine complexes of rhodium(III): shape-selective recognition and reaction. *J. Am. Chem. Soc.* **114**, 2303-2312.

28. Russo, M. T., Blasi, M. F., Chiera, F., Fortini, P., Degan, P., Macpherson, P., Furuichi, M., Nakabeppu, Y., Karran, P., Aquilina, G. & Bignami, M. (2004). The oxidized deoxynucleoside triphosphate pool is a significant contributor to genetic instability in mismatch repair-deficient cells. *Mol. Cell. Biol.* **24**, 465-474.

29. Mazurek, A., Berardini, M. & Fishel, R. (2002). Activation of human MutS homologs by 8-oxo-guanine DNA damage. *J. Biol. Chem.* **277**, 8260-8266.

30. Macpherson, P., Barone, F., Maga, G., Mazzei, F., Karran, P. & Bignami, M. (2005). 8-Oxoguanine incorporation into DNA repeats *in vitro* and mismatch recognition by MutSα. *Nucleic Acids Res.* **33**, 5094-5105.

Chapter 6: Hypersensitivity of Mismatch Repair Deficient Cells to Bulky Rhodium(III) Intercalators

6.1: Introduction

DNA repair is essential for maintenance of genomic integrity. Mismatch repair, MMR, is responsible for the recognition and repair of mispaired DNA bases, small insertion or deletion loops and some chemically damaged bases.[1; 2] With a deficiency in MMR, the mutation rate in cells is enhanced[3] and repetitious DNA sequences can become unstable.[4]

MMR deficiency is an integral part of the formation of some cancers. Loss of functional MLH1, MSH2, MSH3, MSH6, PMS1 or PMS2 cause MMR deficiency.[5] Those who have inherited a defective copy of any of these genes are especially at risk of developing certain endothelial cancers, primarily colon cancer.[6] This syndrome is termed hereditary non-polyposis colon cancer, HNPCC. In addition to this syndrome, approximately 18% solid tumors which are tested for mismatch repair are found to be deficient.[7] While mismatch repair is not essential to cancer, it can greatly accelerate the rate at which other essential mutations are introduced.[8; 9]

In addition to being a predisposing factor for the formation of tumors, mismatch repair deficiency is also associated with a tolerance of many common chemotherapeutics including alkylating agents, platinum compounds, topoisomerase inhibitors, and metabolic analogs.[10] Mismatch repair deficient cell lines have been shown to be hypersensitive to certain drugs including topoisomerase inhibitors captothecin and etoposide[11], DNA double strand break inducing bleomycin[12], and methylating agent temozolomide in combination with methoxyamine.[13] It has also been suggested that

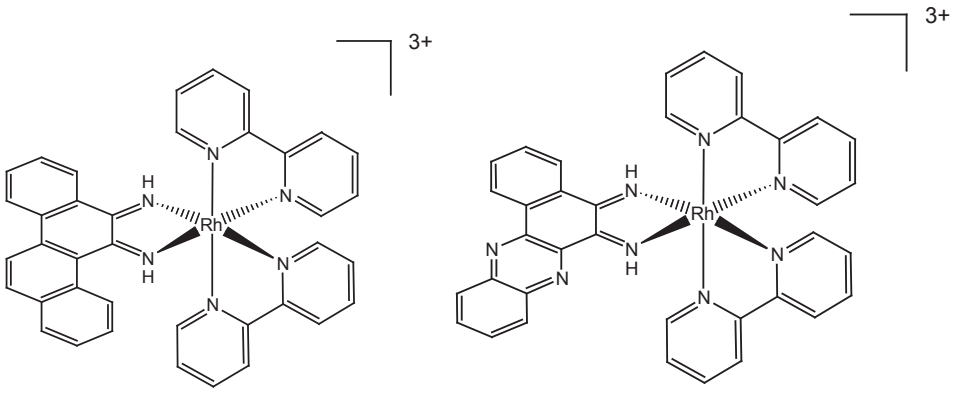

Figure 6.1: Structures of Rhchrysi and Rhphzi. Rhchrysi is our first generation mismatch binding molecule. It is capable of binding to 80% of the possible mismatches with moderate affinity. Rhphzi is a higher affinity complex which has similar recognition properties to Rhchrysi.

recurrent tumors which are drug resistant obtain this trait through MMR deficiency.[14] Because there are few compounds which are active against MMR deficient cells, new compounds are being sought which will be active against mismatch repair deficient tumors.

Within our group we have developed bulky rhodium metallointercalators which can bind mismatches in DNA, shown in **figure 6.1**.[15; 16] These compounds bind mismatches by intercalation and upon UV photoexcitation can cleave the DNA at the mismatch site. Our first generation complex $[Rh(bpy)_2(chrysi)]Cl_3$, Rhchrysi, can bind mismatches with moderate affinity and high specificity. The high specificity of Rhchrysi is beautifully demonstrated in its ability to cleave a single mismatched site in a 2700 base pair duplex.[17] The second generation complex $[Rh(bpy)_2(phzi)]Cl_3$, Rhphzi, utilized heterocyclic nitrogens to boost the affinity of the complex.[16] Rhphzi is capable of binding mismatched sites with 100 nM affinity and similar specificity to Rhchrysi. Here we report our initial *in vivo* experiments with Rhchrysi and Rhphzi in model mismatch repair deficient and proficient cell lines.

6.2: Experimental

Materials. *Rac*-$[Rh(bpy)_2(chrysi)]Cl_3$ and *rac*-$[Rh(bpy)_2(phzi)]Cl_3$ were synthesized using established procedures.[15; 16] *Rac*-$[Rh(bpy)_2(chrysi)]Cl_3$ was resolved into Δ and Λ isomers using potassium antimonyl tartrate as previously described.[18] Salts were exchanged to chloride before use using Sephadex QAE ion exchange resin. Media and supplements were purchased from Invitrogen. 5-bromodeoxyuridine, antibodies, buffers and peroxidase substrate were purchased in kit format from Roche Molecular Biochemicals. Irradiations were performed using a Oriel 1000 W Hg/Xe solar simulator using a UVB/C blocking filter.

Cell lines, Media and Culture. The HCT116N and HCT116O cell lines were

obtained from the National Cancer Institute (Bethesda, MD).[19] Cells were grown in RPMI-1640 media supplemented with 10% fetal bovine serum, 2 mM L-glutamine, 0.1 mM non-essential amino acids, 1 mM sodium pyruvate, 100 U/mL penicillin, 100 ug/mL streptomycin, 400 ug/mL geneticin (G418). Cells were grown in Corning Costar tissue culture flasks and dishes at 37 °C under 5% CO_2 atmosphere.

Cell proliferation assays. Cell proliferation was measured using the 5-bromouracil incorporation assay.[20] Cells were plated in 96 well plates at 2000 cells/well and grown for 24 hours. At this point, various concentrations of rhodium complexes were added. The cultures were cell cultures were allowed to grow for 48 hours or 6–96 hours as indicated. Cells were either irradiated or not followed by the addition of 5-bromodeoxyuridine, BrdU. Irradiation of cells were performed by a solar simulator adapted for irradiation from the bottom of the well plates. The cells were grown for an additional 24 hours. The BrdU incorporation was assayed by antibody assay using established procedures.

Rhodium isomers photocleavage of mismatched DNA. 1 μM of either Δ or Λ $[Rh(bpy)_2(chrysi)]^{3+}$ were added to 1 μM of ^{32}P labeled mismatched DNA duplex template, 5'-$^{32}P_i$-TTA GGA TCA T**C**C ATA TA-3' and 5'-TAT ATG **C**AT GAT CCT AA-3' dissolved in 50 mM NaCl with 10 mM sodium phosphate buffer pH 7.1. This mixture was irradiated for 10 minutes using a solar simulator. The irradiated DNA was run on a denaturing polyacrylamide gel and imaged using phosphorimagery.

6.3: Antiproliferative effects of rhodium compounds

The cytotoxicity of our compounds was assessed in the HCT116 derived cell lines. These cell lines have had an extra copy of either chromosome 2, HCT116O, or 3, HCT116N, inserted in the cell. Placing a copy of chromosome 3 within this cell line corrects the mismatch repair deficiency by replacing the MLH1 gene with a functional

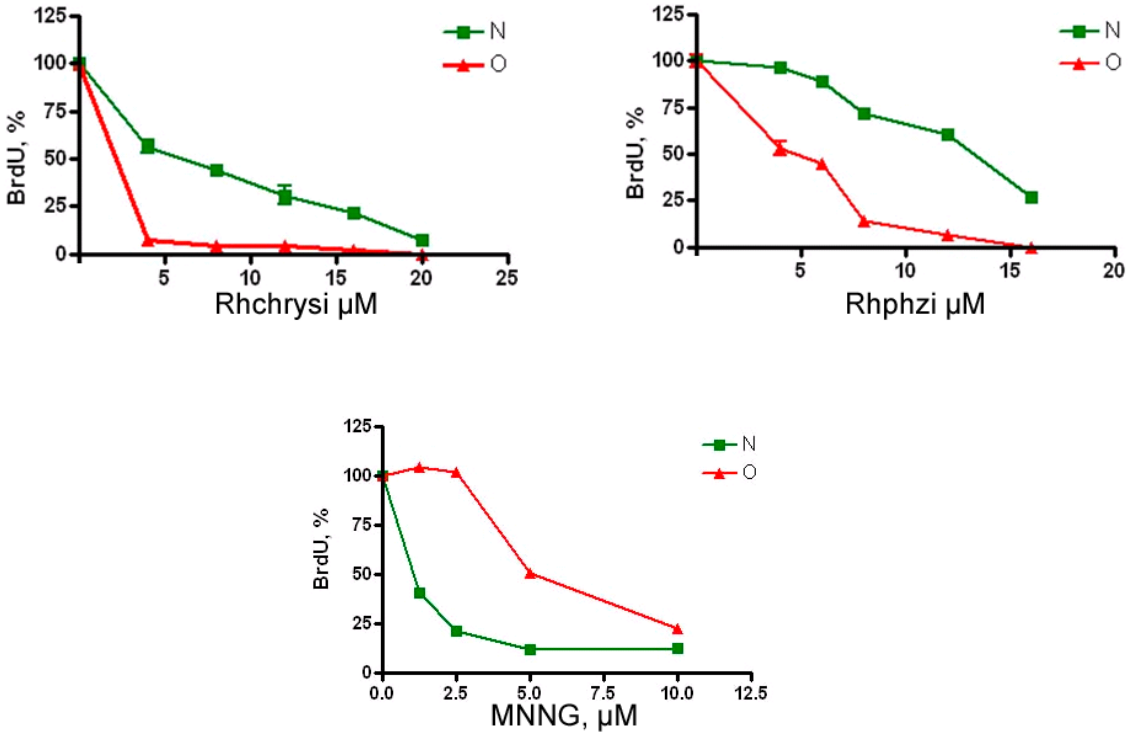

Figure 6.2: Differential antiproliferative effect of MNNG, Rhchrysi and Rhphzi on mismatch repair deficient and proficient cell lines. Mismatch repair deficient, O, and proficient, N, cell lines derived from the HCT116 cell line were treated with the methylating agent MNNG or Rhchrysi or Rhphzi. As is typical the mismatch repair deficient population is resistant to the action of MNNG[15], but is sensitive to Rhchrysi and Rhphzi.

copy normal promoter. These cell lines give us a convenient method for testing whether or not a particular compound will differentially target mismatch repair deficient cells, since these cells are essentially genetically identical except for mismatch repair proficiency.

These HCT116 derived cell lines were treated with either Rhchrysi or Rhphzi for 48 hours at varying concentrations. The results are shown in **figure 6.2**. For both of the tested compounds we see a significant inhibition of proliferation of the mismatch repair deficient strain. This contrasts with the action of MNNG, a DNA methylating agent, which is more toxic to the mismatch repair proficient cell line.[19]

The effect of incubation time was also investigated, the results are shown in **figure 6.3**. Cells were exposed to various concentrations of Rhchrysi for varying amounts of time from 6 to 96 hours. In this series it is clear that maximum effect is obtained after 48 hours of incubation.

Experiments were also conducted using enantiomerically pure Rhchrysi. Shown in **figure 6.5** is an example of the enatiospecific interaction of Rhchrysi with a C-C mismatch within an oligonucleotide. Upon excitation of the complex with UV radiation, the complex cleaves the DNA at the site of intercalation. For this particular sequence only the Δ isomer binds to the DNA. In general, the Δ isomer binds to mismatched DNA more tightly than the Λ isomer.

The antiproliferative effects of these two isomers were also assessed. Results of this experiment are shown in **figure 6.4**. Clearly, the Δ isomer is the isomer responsible for the antiproliferative effect. Since the opposite enantiomer is inactive, we suggest that the active compound is the intact rhodium complex, and no decomposition of the metal complex occurs intracellularly to form the active species. These complexes are capable of cleaving DNA upon UV irradiation.[15; 16] Some increase in toxicity is noted with 15 minutes of irradiation and only for the Δ isomer.

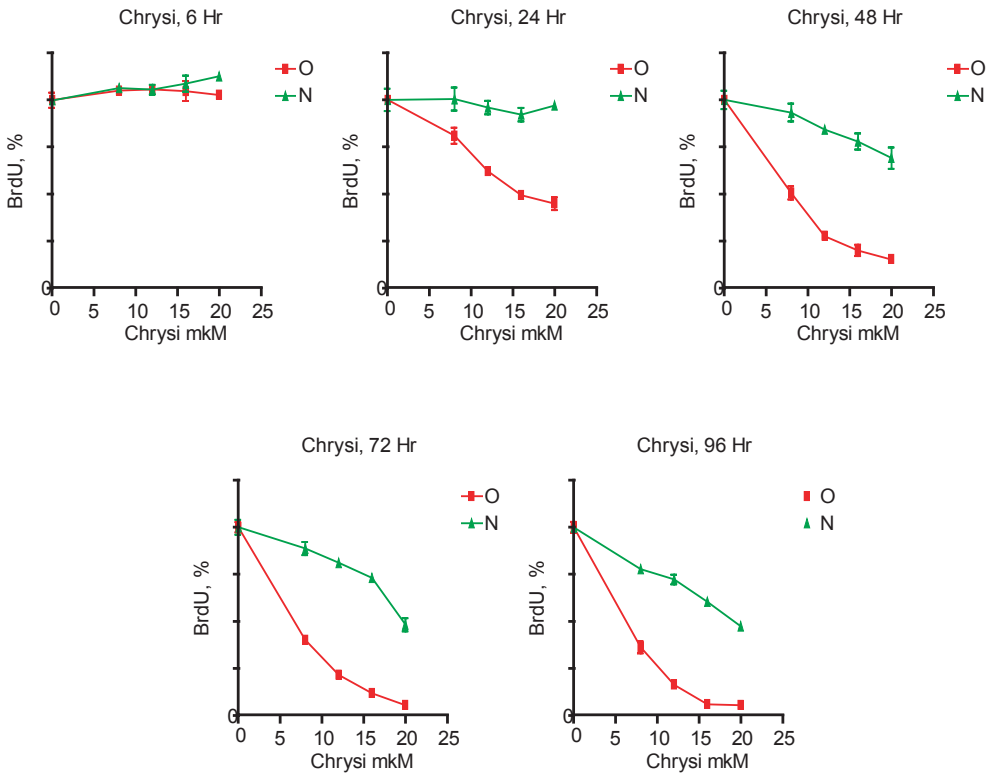

Figure 6.3: Effect of varying drug incubation time on cell proliferation. Cells were exposed to varying concentrations of Rhchysi for different incubation times as shown. Antiproliferative effects increased with longer incubation times upto 48 hours. Increases were seen at all concentrations of Rhchrysi.

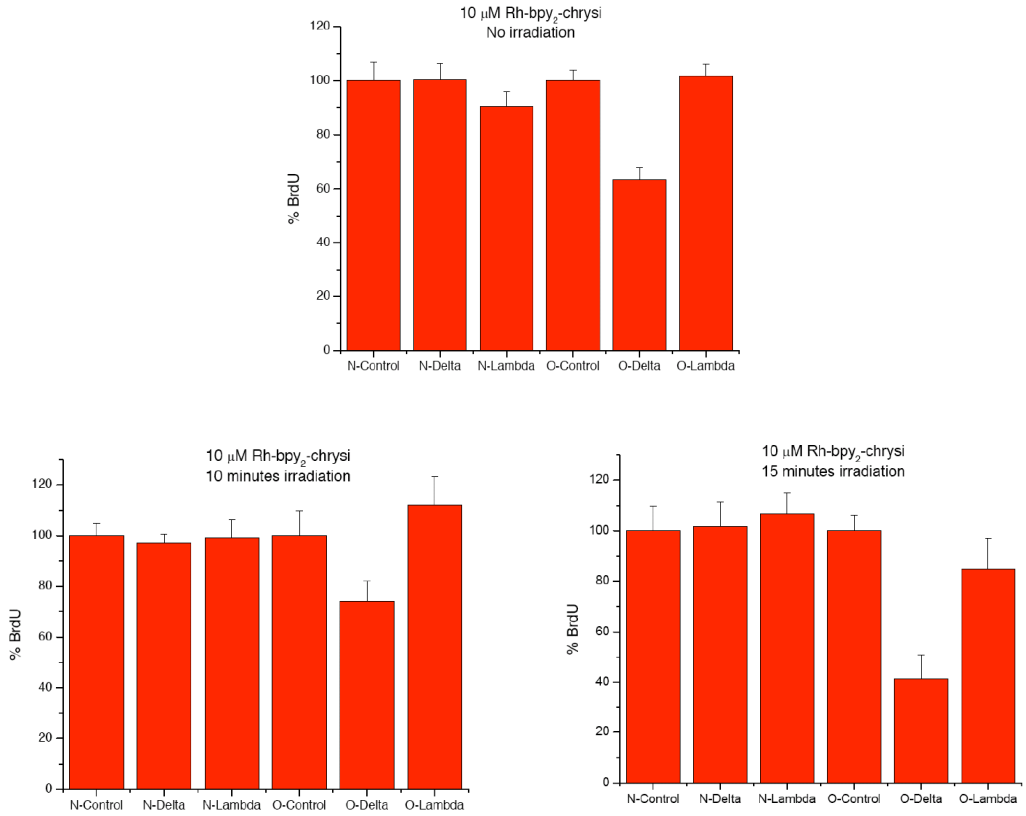

Figure 6.4: Antiproliferative effect of different stereoisomers and irradiation of Rhchrysi. The octahedral metal center of Rhchrysi and Rhphzi is chiral producing two different isomers. These isomers can be separated and tested for antiproliferative effects individually. As shown above the Δ isomer appears to be the only active compound. This is significant as it shows that the compound is not dissociating within the cell. These complexes are photoreactive with DNA *in vitro*, so the effect of irradiation was also explored. As shown a small enhancement of antiproliferative effect is noted upon irradiation

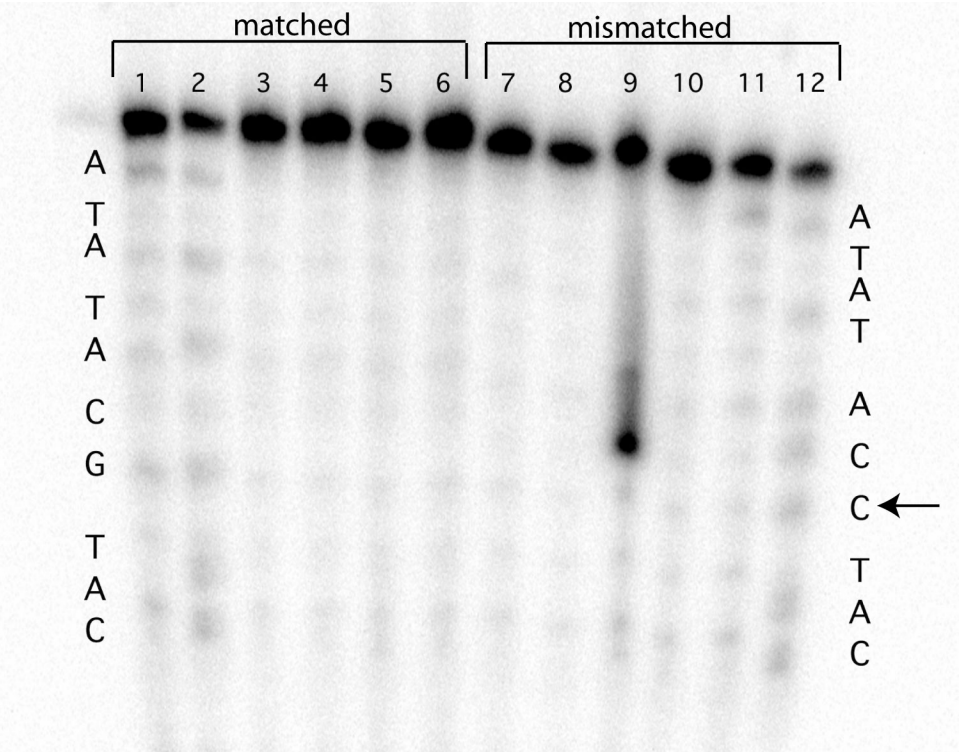

Figure 6.5: Effect of stereoisomers on photocleavage. A C-C mismatch is incorporated at the site indicated by the arrow on the right side. Cleavage is only noted with Δ-Rhchrysi in the presence of a mismatch. 1, 11 are AG and 2, 12 are CT Maxam-Gilbert reactions. 3 and 7 are light controls, without Rhchrysi. 4 and 8 are dark controls with *rac*-Rhchrysi. 5 and 9 are reactions using Δ-Rhchrysi. 6 and 10 are reactions with Λ-Rhchrysi.

MNNG is typical of most chemotherapeutics which target genomic DNA. MMR deficient cell lines are resistant to the majority of alkylating drugs, platinum compounds and metabolic analogs.[10] The cytotoxicity and antiproliferative effects these compounds have on cancerous cells are due in part to recognition of the drug induced genomic DNA damage by the mismatch repair system.[21] This recognition causes cell cycle arrest and in some cases apoptosis.

While these results are by no means conclusive, they are not inconsistent with genomic DNA being the target of these compounds. These compounds do intercalate into DNA *in vitro* at mismatched sites with high affinity and specificity. While the mechanism of action is not clearly understood for these complexes, the ability of these complexes to inhibit growth of mismatch repair deficient cells is intriguing.

6.4: References

1. Kunkel, T. A., Erie, D. A. (2005). DNA mismatch repair. *Annu. Rev. Biochem.* **74**, 681-710.

2. Kolodner, R. D., Marsischky, G. T. (1999). Eukaryotic DNA mismatch repair. *Curr. Opin. Genet. Dev.* **9**, 89-96.

3. Bhattacharyya, N. P., Skandalis, A., Ganesh, A., Groden, J., Meuth, M. (1994). Mutator phenotypes in human colorectal-carcinoma cell-lines. *Proc. Natl. Acad. Sci. USA* **91**, 6319-6323.

4. Strauss, B. S. (1999). Frameshift mutation, microsatellites and mismatch repair. *Mut. Res. Rev. Mut. Res.* **437**, 195-203.

5. Papadopoulos, N., Lindblom, A. (1997). Molecular basis of HNPCC: Mutations of MMR genes. *Hum. Mutat.* **10**, 89-99.

6. Peltomaki, P. (2001). Deficient DNA mismatch repair: A common etiologic factor for colon cancer. *Hum. Mol. Genet.* **10**, 735-740.

7. Arzimanoglou, I. I., Gilbert, F., Barber, H. R. K. (1998). Microsatellite instability in human solid tumors. *Cancer* **82**, 1808-1820.

8. Loeb, L. A. (2001). A mutator phenotype in cancer. *Cancer Res.* **61**, 3230-3239.

9. Loeb, L. A., Loeb, K. R., Anderson, J. P. (2003). Multiple mutations and cancer. *Proc. Natl. Acad. Sci. USA* **100**, 776-781.

10. Fink, D., Aebi, S., Howell, S. B. (1998). The role of DNA mismatch repair in drug resistance. *Clin. Cancer Res.* **4**, 1-6.

11. Jacob, S., Aguado, M., Fallik, D., Praz, F. (2001). The role of the DNA mismatch repair system in the cytotoxicity of the topoisomerase inhibitors camptothecin and etoposide to human colorectal cancer cells. *Cancer Res.* **61**, 6555-6562.

12. Li, H. R., Shagisultanova, E. I., Yamashita, K., Piao, Z., Perucho, M., Malkhosyan, S. R. (2004). Hypersensitivity of tumor cell lines with microsatellite instability to DNA double strand break producing chemotherapeutic agent bleomycin. *Cancer Res.* **64**, 4760-4767.

13. Liu, L. L., Nakatsuru, Y., Gerson, S. L. (2002). Base excision repair as a therapeutic target in colon cancer. *Clin. Cancer Res.* **8**, 2985-2991.

14. Karran, P., Offman, J., Bignami, M. (2003). Human mismatch repair, drug-induced DNA damage, and secondary cancer. *Biochimie* **85**, 1149-1160.

15. Jackson, B. A., Barton, J. K. (1997). Recognition of DNA base mismatches by a rhodium intercalator. *J. Am. Chem. Soc.* **119**, 12986-12987.

16. Junicke, H., Hart, J. R., Kisko, J., Glebov, O., Kirsch, I. R., Barton, J. K. (2003). A rhodium(III) complex for high-affinity DNA base-pair mismatch recognition. *Proc. Natl. Acad. Sci. USA* **100**, 3737-3742.

17. Jackson, B. A., Alekseyev, V. Y., Barton, J. K. (1999). A versatile mismatch recognition agent: Specific cleavage of a plasmid DNA at a single base mispair. *Biochemistry* **38**, 4655-4662.

18. Murner, H., Jackson, B. A., Barton, J. K. (1998). A versatile synthetic approach to rhodium(III) diimine metallointercalators: Condensation of o-quinones with coordinated cis-ammines. *Inorg. Chem.* **37**, 3007-3012.

19. Koi, M., Umar, A., Chauhan, D. P., Cherian, S. P., Carethers, J. M., Kunkel, T. A., Boland, C. R. (1994). Human-chromosome-3 corrects mismatch repair deficiency and microsatellite instability and reduces n-methyl-n'-nitro-n-nitrosoguanidine tolerance in colon-tumor cells with homozygous hMLH1 mutation. *Cancer Res.* **54**, 4308-4312.

20. Gratzner, H. G. (1982). Monoclonal-antibody to 5-bromodeoxyuridine and 5-iododeoxyuridine: A new reagent for detection of DNA-replication. *Science* **218**, 474-475.

21. O'Brien, V., Brown, R. (2006). Signalling cell cycle arrest and cell death through the MMR System. *Carcinogenesis* **27**, 682-692.