EXPLORING THE DIRECT AND INDIRECT READOUT OF DNA WITH PHENANTHRENEQUINONE DIIMINE COMPLEXES OF RHODIUM(III)

Dissertation by

Robert Terbrueggen

Advisor Jacqueline K. Barton

In Partial Fulfillment of the

Requirements for the Degree of Doctor of Philosophy

in Chemistry

California Institute of Technology

Pasadena, California

1996

(Submitted October 12, 1995)

Acknowledgments

I would like to thank my family for all of the support they have given me over the years. Throughout these seemingly endless years that I have been a student, my family has always been extremely encouraging. My outlook on life is a direct result of the time that I have spent with them.

I would like to thank all of my friends for their support throughout my graduate career. Out of all of the memories I have to look back on over the last 4 years, it is the times I have spent playing with friends which standout in my mind. Thoughts of the ski trips and the scuba diving have kept me going during those times when my lab work was not proceeding as desired. I would especially like to acknowledge Steve O'Connor and Foley Weems. I thank Steve for being my roommate for 3 years, and for always being willing to go out and play. I thank Foley for agreeing to move down to Manhattan Beach with me. Living at the beach has made the past 6 months of my graduate career truly memorable. I shall really miss the beach volleyball and the body surfing.

I thank my advisor Jackie Barton for the encouragment and direction she has given me over the past 4 years. The importance of Jackie's guidance and support througout my graduate career is immeasurable. I would especially like to thank her for being understanding and allowing me to work at my own pace.

I would like to thank the past and present members of the Barton group. My lab mates are a wonderful bunch of people to work with. The friendliness of this group has made coming into work enjoyable, and the knowledge and intelligence of this group of people is truly amazing. I would especially like to thank Achim Krotz for his help as I was starting out as a first year graduate student. I would like to thank Dan Hall for his synthetic consultation, Tim Johann for the

numerous discussions we have had, and Tom Shields for being a great person to share a lab bay with. I would also like to thank the members of the Carreira and Myers groups for allowing me to steal dry solvents and supplies when I ran out at 11 o'clock at night. I would especially like to acknowledge Justin DuBois of the Carreira group for his synthetic advice. I thank Mo Renta for her immense help over the last 4 years. Mo's assistance has eased the burden of being a graduate student. I would especially like to thank her for the help she has given me while I was writing my thesis.

Attending Cal Tech has been a truly wonderful experience. Cal Tech is not only a great place to do science, but it is the social aspects of the school that have made it special for me. I thank the members of the Cal Tech football team for the good times I had throughout my 3 years on the team. I would also like to thank the many people that I have played basketball, flag football and softball with. It is these diversions which have kept me going throughout my graduate career. All work and no play makes Bob a dull boy, and the Cal Tech community gave me many opportunities to play.

Abstract

Phenanthrenequinone diimine (phi) complexes of rhodium(III) have been designed and characterized in order to investigate the principles of direct and indirect readout of double helical DNA. The metallointercalator 1-$Rh(MGP)_2phi^{5+}$ (MGP = 4-guanidylmethyl-1,10-phenanthroline) binds via intercalation in the major groove of DNA and upon irradiation promotes DNA strand scission. The Λ-enantiomer, Λ-1-$Rh(MGP)_2phi^{5+}$, binds at subnanomolar concentrations to the 6 base pair sequence, 5'-CATATG-3', with enantiospecificity. An essential feature of this recognition is the sequence-specific unwinding of the DNA helix which permits direct contacts between guanidinium functionalities on the metal complex and guanine residues. Deazaguanine substitutions were used to establish direct contacts between the N7 nitrogen atoms of guanine and the guanidinium moiety on the metal complex. Through an assay developed to test for sequence-specific DNA unwinding, a 70 ± 10 degrees unwinding of the sequence 5'-CATATG-3' upon binding by Λ-1-$Rh(MGP)_2phi^{5+}$ was established. Thus, the sequence-dependent twistability of DNA plays an important role in determining the sequence specificity of the complex. The Δ-enantiomer, Δ-1-$Rh(MGP)_2phi^{5+}$, binds preferentially to the 6 base pair sequence, 5'-CATCTG-3'. The hierarchy of recognition sites determined in photocleavage studies on oligonucleotides suggests that DNA recognition by this complex also involves sequence specific contacts by the guanidinium functionalities. Photocleavage studies indicate additional similarities in the recognition of Δ and Λ-1-$Rh(MGP)_2phi^{5+}$. Both enantiomers of 1-$Rh(MGP)_2phi^{5+}$ display increased binding specificity relative to the parent complex, $Rh(phen)_2phi^{3+}$. The exchange rates of both enantiomers are also decreased at least a 1000-fold relative to $Rh(phen)_2phi^{3+}$. Studies in which the

length of the linker arm between the core of the metal complex and the guanidinium moiety was varied demonstrate that proper orientation of the guanidinium moiety is an essential feature of complex specificity. As the length of the linker arm increases, the binding specificity of the complex decreases. DNA recognition studies with $Rh(APB)_2phi^{5+}$ (APB = 4-(3-aminopropyl)-4'-2,2'-bipyridine) have demonstrated that the amino moiety can also be used to alter the sequence specificity of phi complexes of rhodium(III), although the sequence specificity of this complex is reduced greatly as compared with Δ- and Λ-1-$Rh(MGP)_2phi^{5+}$. This work therefore demonstrates that the guanidinium moiety may be used to enhance both the binding affinity and specificity of phi complexes of rhodium(III). In mimicking DNA binding proteins, molecules which recognize their binding sites through direct and indirect readout of the DNA can be designed. Importantly, this study highlights a new structural element of DNA recognition, the sequence-dependent twistability of the DNA helix. This sequence-dependent twistability may be an essential feature of the recognition of sequences by DNA-binding proteins and may be powerfully exploited in future design.

Table of Contents

LIST OF FIGURES

Chapter 1:

LIST OF SCHEMES

Chapter 3:

Chapter 5:

LIST OF TABLES

Chapter 1: Introduction

1.1. Principles of DNA Recognition for Small Molecules and Proteins

Deoxyribonucleic acid (DNA) is the storehouse of all genetic information. Accessing the information encoded in an organism's genome is a complicated process involving the controlled interactions of thousands of different molecules. This process is governed by the principles of molecular recognition. The ability to understand the types of interactions that occur between proteins and DNA, and the energetics that is associated with these interactions, is an exciting challenge which offers the reward of having a better understanding of the mysteries of life. As our understanding of biological processes and molecular recognition improves, it is becoming a viable option to manipulate proteins and DNA, so as to prevent disease. One potential scenario for the prevention of disease is to synthesize molecules which selectively block access to the DNA, such that the genes which code for undesirable processes cannot be read. The design of molecules capable of functioning as sequence specific transcriptional inhibitors truly tests our understanding of many key principles of molecular recognition.

1.2. DNA Structural Variations

Our understanding of DNA structure has changed since the original structure of DNA was proposed by Watson and Crick over 40 years ago. Watson and Crick proposed that DNA is a double-helical molecule in which the heterocyclic base pairs (bp), CG and AT, occupy the center of the helix and the sugar-phosphate backbone wraps around the outside in a right-handed fashion. An important part of this model is that two polynucleotide chains running in opposite orientation come together through hydrogen bonding to form the double helix. The inner base pairs are perpendicular to the helix axis and the AT base pairs are held together by two hydrogen bonds and the GC base pairs are held together by 3. Importantly, two distinct grooves through which the edge of the bases

could be accessed by proteins and other molecules would exist (major and minor groove) (Figure 1.1).

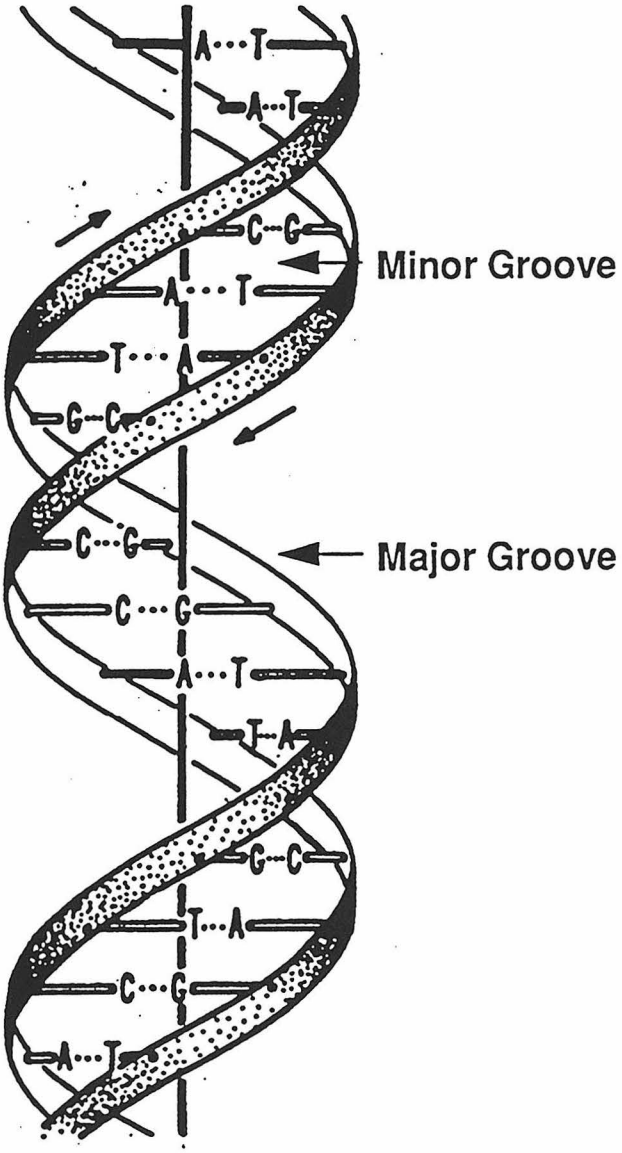
Subsequent studies have verified the general structure of the Watson-Crick model. However, x-ray analysis has demonstrated that the DNA helix is not nearly as uniform in structure as was originally thought (109). In fact, large variations for the normal helical parameters are often observed at individual base steps. In an effort to quantify these deviations, it was necessary to create a set of parameters by which to catalog DNA structure (108). Some of the more frequently sited helical parameters are propeller twist, helical twist, base rise, and base pair roll (Figure 1.2).

In addition to local variations, large global differences in DNA structure also occur. Three distinct DNA shapes have been identified, A, B and Z-form DNA (Figure 1.3). Both A and B-DNA are right-handed helices and vary in the number of base pairs per turn of the helix and in the depth and width of the major and minor grooves. Z-DNA has an overall left-handed helicity and has primarily been observed for sequences possessing stretches of 5'-GC-3' dinucleotide repeats. A summary of some basic helical parameters for these different DNA forms are shown in Table 1.1. Other structural polymorphisms of the DNA helix have been observed which could be important for DNA recognition. Sequences which form cruciforms, bends, tetraplexes and triple helices have been identified (Figure 1.4).

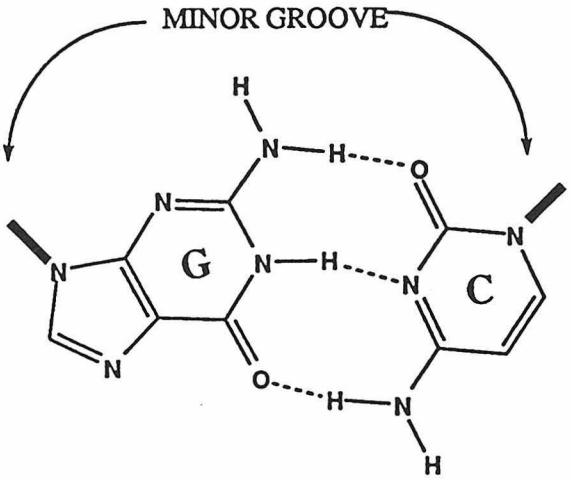
A major obstacle in the development of transcriptional inhibitors is our incomplete understanding of how DNA structure varies with sequence. Molecular modeling has been incapable of solving this problem, because biomolecules like DNA and proteins can assume thousands of different possible conformations, and the energy differences between alternate DNA conformations might be so slight (1-2 kCal out of 200 kCal total stabilization energy) that a 10 percent error in energy minimization calculations would yield the wrong structure. Furthermore, the thermal energy available to molecules at room temperature is sufficient to allow interconversion between different conformational states.

Figure 1.1. Schematic illustration of DNA (A.) Ribbon diagram of B-DNA. The sugar phosphate backbone is represented by the vertical ribbons and the base pairs are represented by horizontal lines. The base pairs are perpendicular to the helical axis and two distinct grooves (as determined by the ribbon backbone) exist (major and minor). (B) Detailed drawing of the DNA bases pairs. The glycosidic bond is represented by the heavy line and acts as the dividing line between the major and minor grooves of the DNA. Guanine (G) pairs with cytosine (C) and Adenine (A) pairs with Thymine (T). Figure adapted from ref. 1.

A.


Minor Groove
Major Groove

B.


MINOR GROOVE
G
C
MAJOR GROOVE

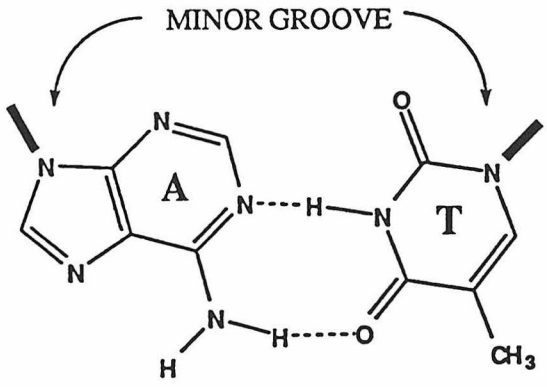
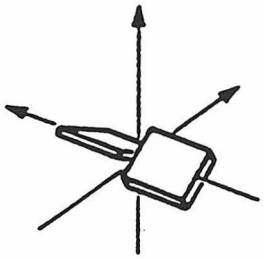
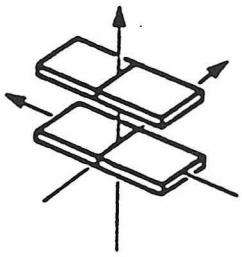

MINOR GROOVE
A
T
MAJOR GROOVE

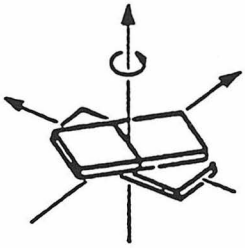
Figure 1.2. Schematic depiction of the four most commonly cited helical parameters. All drawings show 5' to 3' orientation as going from the bottom to the top of the page. (Top Left) Propeller twist- Refers to the planarity of the hydrogen bonding interactions between the two bases of a base pair. Positive propeller twist is towards the base to the 3' side when viewed from the major groove. (Top Right) Base rise- Distance between consecutive base pairs. (Bottom left) Helical twist- Rotation about the helical axis of one base pair with respect to the next. Looking down the helical axis (5' to 3'), in right-handed DNA, the bottom base is rotated in a clockwise (positive) direction with respect to the top base. (Bottom right) Base roll- Rotation of a base pair about the base pair axis (perpendicular to helical axis). When viewed from the major groove, positive roll is in the 3' direction. Adapted from ref. 108.



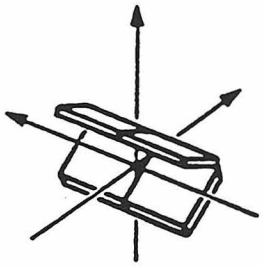
Propeller twist (ω)



Rise (D_z)



Twist (Ω)



Roll (ρ)

Figure 1.3. A-, B-, and Z-form DNA (left to right). Side view and top view. Adapted from ref. 1.

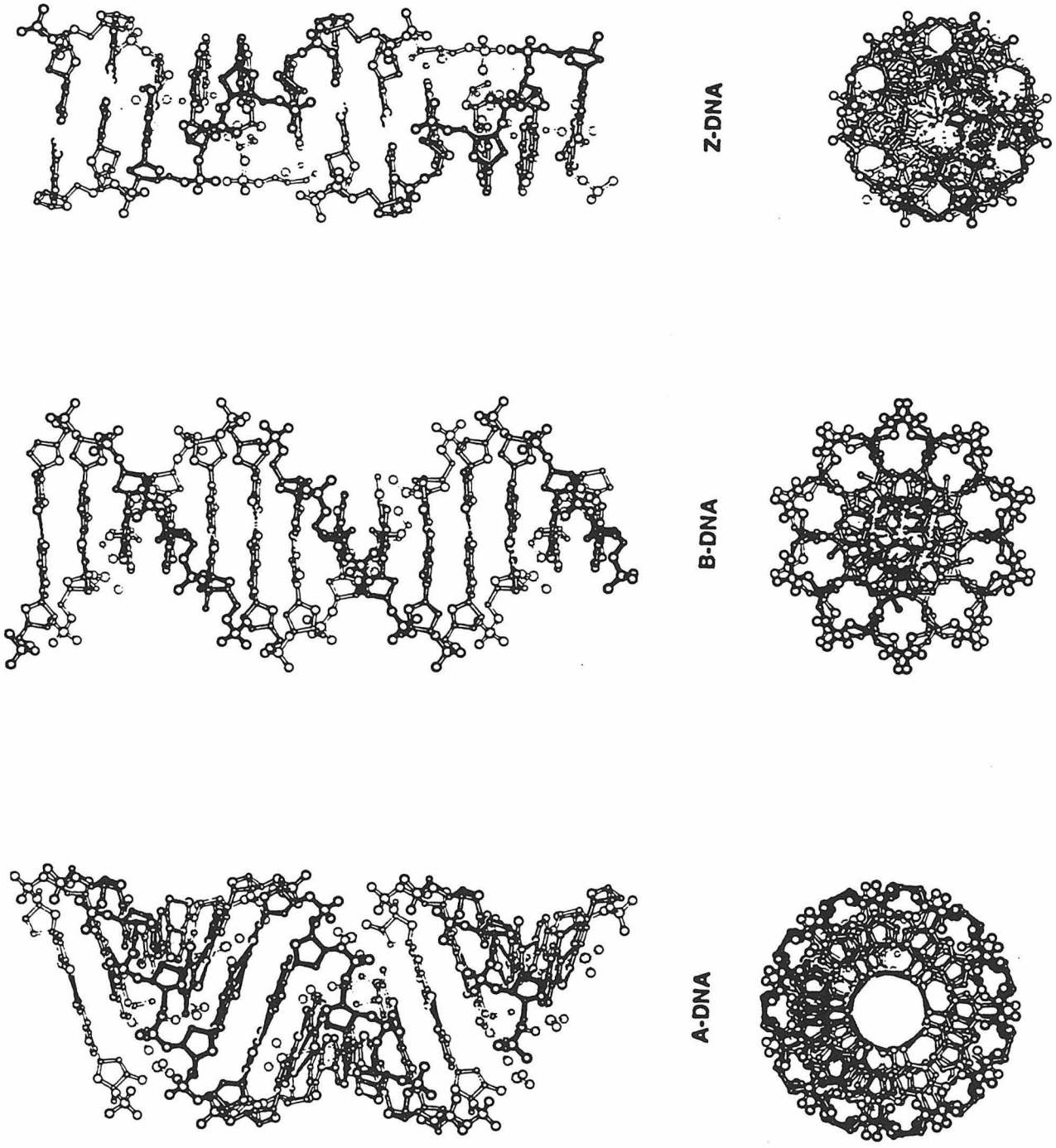
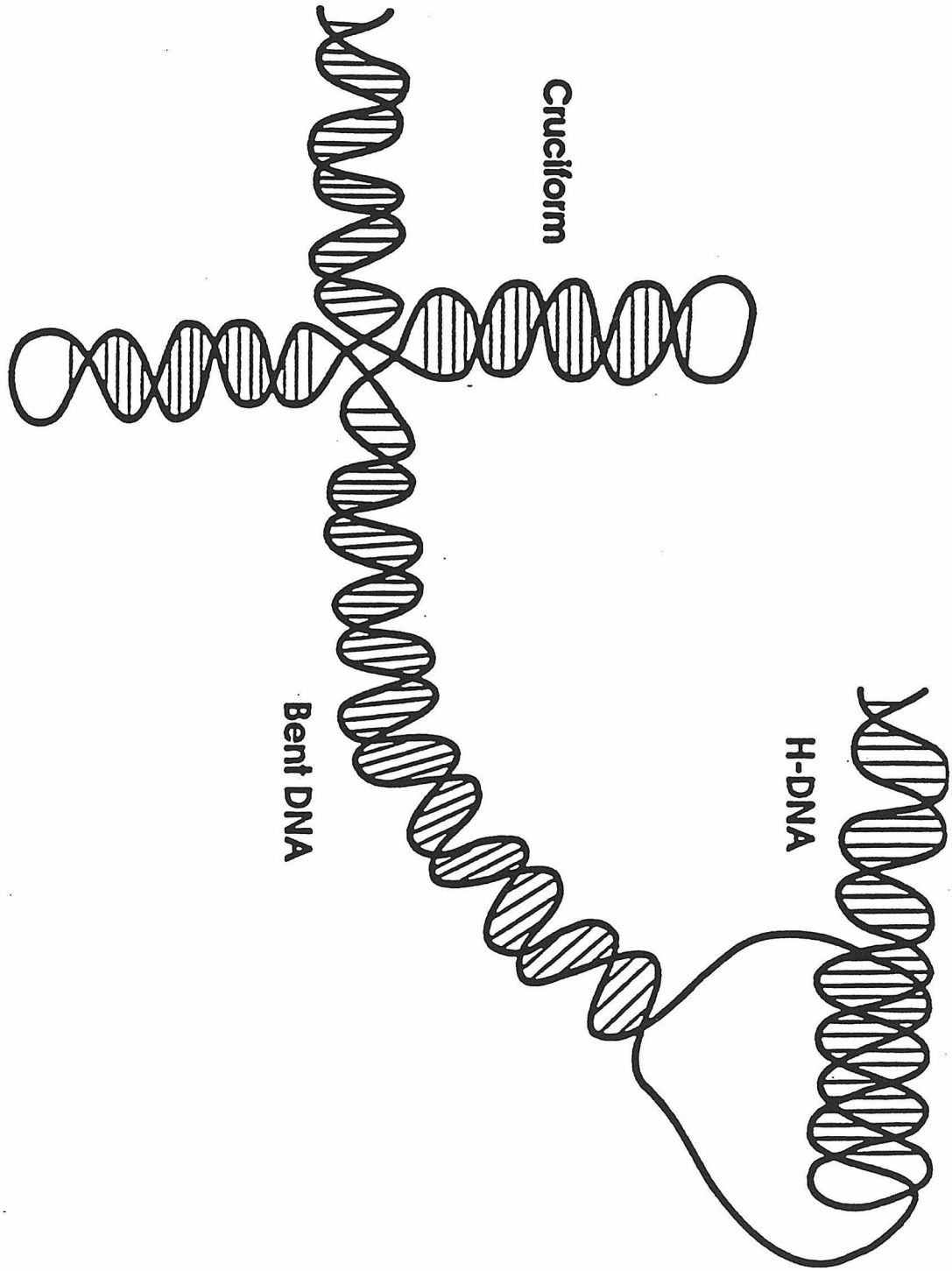

Z-DNA
B-DNA
A-DNA

Table 1.1 Structural Characterisitics of DNA. Adapted from ref. 1.

Table 1.1. Structural Characteristics of DNA According to Type[a]

Feature	A[b]	B[b]	Z
helix diameter	23 Å	19.3 Å	18.4 Å
pitch	28.2 Å	33.8 Å	45 Å
base pairs per turn	11 bp	10 bp	12 bp
groove width[c] -major[d]	2.7 Å	11.7 Å	-
minor	11.0 Å	5.7 Å	2.7 Å
groove depth[c] -major[d]	13.5 Å	8.5 Å	-
minor	2.8 Å	7.5 Å	9.0 Å
rise	2.56 Å	3.38 Å	3.7 Å
inclination	10° to 20°	-5.9° to -16.4°	-7°
base pair displacement	4.4 to 4.9 Å	-0.2 to -1.8 Å	positive[e]

Figure 1.4. Line drawing illustration of structural polymorphisms in DNA. Shown are a DNA cruciform, a bent segment and H-DNA or hinged-DNA (which contains a triple-helix region).


Cruciform
Bent DNA
H-DNA

Experiments which provide information about the sequence dependent variations in DNA structure will greatly aid the effort to design molecules with predictable DNA sequence specificities.

Techniques such as X-ray crystallography are helping to decipher how DNA structure varies with sequence. Numerous crystal structures of DNA oligonucleotides have been solved and the structural diversity of helix has been verified time and again. However, it should be noted that the structure of the DNA in the crystal may not necessarily reflect the structure of the DNA in solution. As mentioned earlier, the barrier to interconversion between different conformations of a biological molecule is often small, and it is possible that crystal packing forces may be sufficient to alter the DNA structure. Furthermore, it is possible that a given DNA fragment might exist in a variety of different conformations in solution and phenomenoms like DNA superhelicity can induce unusual structures in the DNA. Thus, X-ray crystallography is not as well suited for the study of DNA structure as one might hope. As a result, it is necessary to explore other techniques which can give information about DNA structure.

DNA footprinting and electrophoretic mobility studies are often employed to test for DNA bending and unwinding. Chemical probes and enzymatic probes are frequently used to investigate DNA structure. Early studies used the DNA degradation enzyme DNase I to probe DNA (1, 3), and more recently, small molecules like methidiumpropyl-EDTA-iron(II) (MPE-Fe(II)) (4,5) and N-methyl-N-nitrosourea (6) are being used to probe the structure of DNA both in the presence and the absence of DNA binding molecules. It is of interest that many of these small molecular probes were developed during earlier DNA studies. One noteworthy example is the metal complex $Rh(phen)_2phi^{3+}$ which has recently been shown to be an effective probe of propeller twisting in B-DNA (7), a parameter which has been found to be very difficult to study.

Electrophoretic mobility studies are becoming increasingly important for the study of DNA structure. These studies rely upon the analysis of anomalies in the electrophoretic

mobility of DNA fragments to assess, qualitatively and quantitatively, variations in the roll and helical twist of the DNA. Large variations in the base roll of the DNA can result in a macroscopic bend in the DNA and variations in the helical twist can result in an altered DNA helical repeat. Recent studies have demonstrated that DNA A-tracts possess an inherent bend (9-12) and, more importantly, this inherent bend is being taken advantage of to design assays which give data about other DNA parameters like helical twist (13,14).

1.3. DNA Binding Proteins

In the last decade, X-ray diffraction and NMR have been used to examine DNA recognition by proteins and small molecules. These data, in conjunction with mutagenesis and footprinting experiments, has greatly increased our knowledge of how proteins recognize their DNA operator sites. Studies have shown that proteins use two primary mechanisms to recognize their DNA operator sites, direct and indirect readout of the DNA sequence. Direct readout refers to amino acid side chains of a protein making specific contacts with the DNA bases pairs. Most frequently, these contacts take the form of hydrogen bonds; however, van der Waals contacts play an important role. Direct contacts often take place between amino acid side chains and the charged phosphate backbone. However, in many cases, these contacts result in an increase in non-specific DNA binding affinity of the protein rather than a site specific increase. The exception is when these phosphate contacts require an unusual DNA structure. Indirect readout of the DNA sequence involves a protein or small molecule recognizing the shape of the DNA helix. Although numerous proteins have been shown to bind to nearly canonical B-form DNA, an increasing number of proteins are being discovered which bind to an unwound and/or bent DNA site. The interplay between direct and indirect readout is very important, and many direct contacts between the protein and the DNA would not be possible unless the DNA was deformed. Most crystallographic studies tend to concentrate on the direct

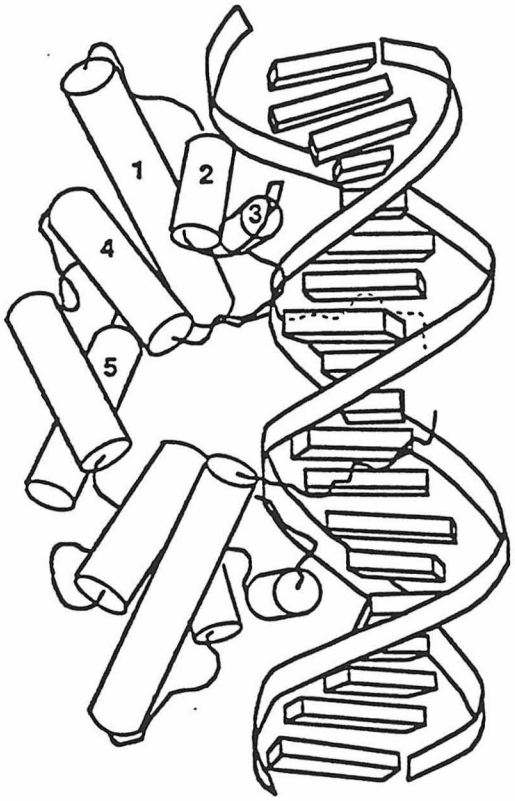
readout aspect of protein recognition, but in most cases, the indirect readout plays an equally important role.

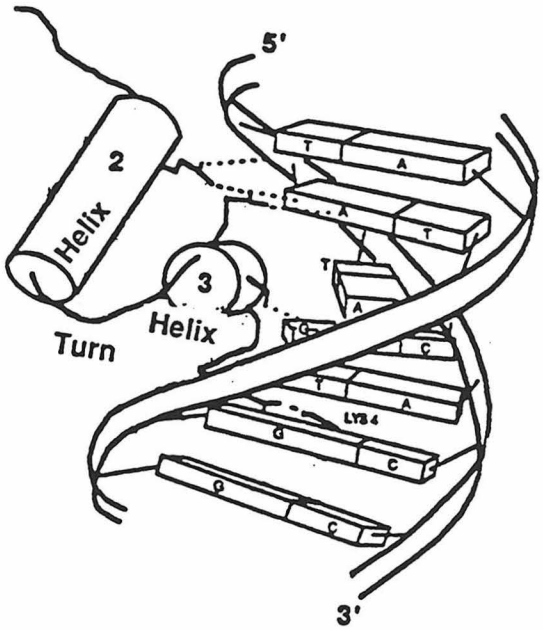
The first structural motif for a DNA binding protein to be identified was the helix-turn-helix (HTH) motif. This motif was originally discovered by structural comparisons between the DNA binding proteins CAP (catabolite activator protein) and cro (21), and for a while it was thought to be the primary motif used by most proteins to recognize DNA . However, in recent years, a number of other widely used motifs have been identified; the zinc-finger, the homeodomain, the leucine zipper, the β-sheet, and the steroid receptor motifs are some of the best understood. Many recent reviews do an excellent job of summarizing past work on these various motifs (22-25). These DNA binding proteins primarily recognize DNA from the major groove, and recognition usually involves a combination of direct and indirect readout of the DNA sequence. The fact that there are so many different families of DNA-binding proteins demonstrates that there is no set code by which to read the DNA helix. All DNA binding proteins use the same general principles to interact with the DNA; it is only the way in which the recognition element (α-helix or β-sheet) of the protein is presented to the DNA that varies.

1.3.1. Helix-Turn-Helix (HTH) Motif

The HTH motif has been shown to be used by the Λ repressor protein (26), the P_{22} repressor (27), the 434 repressor (28), and CAP (29). The HTH recognition motif consists of an α-helix followed by a turn region and then a second α-helix. The length of the helices can vary and the most highly conserved residues are a glycine in the turn region and some hydrophobic residues which are important for the packing of the α-helices against each other (23). The first α-helix of the HTH domain does not directly interact with the DNA base pairs. Instead it acts as a positioning arm and assures proper positioning of the second α-helix or recognition helix. The recognition helix packs against the first helix and lies in the major groove of the DNA (24) (Figure 1.5). The amino acid side chains of the

Figure 1.5. (Left) Sketch of the N-terminal domain of the Λ-repressor dimer, a HTH protein, binding to its DNA operator site. Cylinders are used to represent regions of α-helix. (Right) Expanded view of the helix-turn-helix region of the Λ-repressor interacting with its DNA operator site. The recognition helix (helix 3) is bound in the major groove and is oriented parallel to the DNA base pairs. Helix 2 helps orient helix 3 with respect to the DNA. Adapted from ref. 24.


1
2
3
4
5

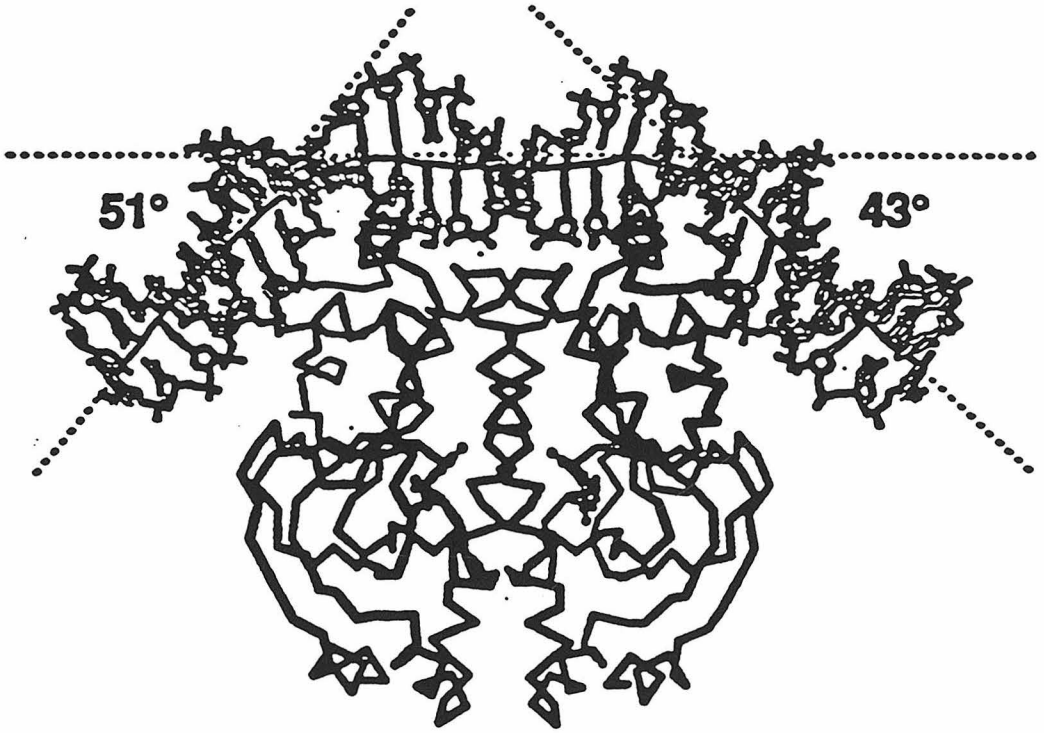

5'
2
Helix
3
Helix
Turn
T
A
A
T
T
A
G
C
T
A
G
C
G
C
LYS 4
3'

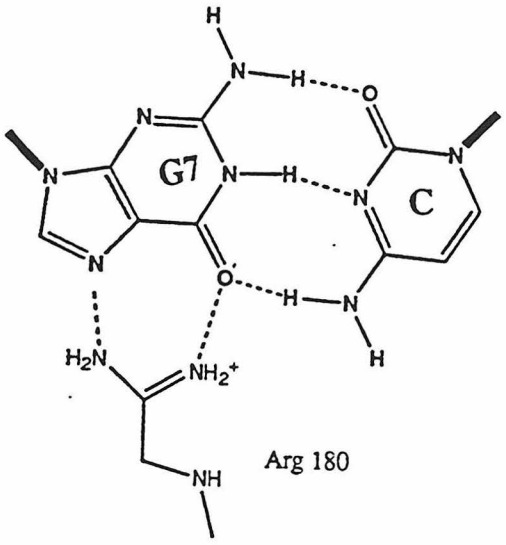
recognition helix contact the DNA base pairs and this unit can be thought of as a flexible reading head. The helical positioning of the amino acids which make base contacts shows some degree of conservation and it has been postulated that by changing the identity of these conserved amino acids, it might be possible to design DNA binding proteins with new sequence specificities (30, 31). Base specific contacts between amino acid side chains and the DNA reveal that proteins recognize their DNA sites by forming hydrogen bonds and van der Waals contacts with the DNA base pairs. While there is no invariant code that proteins use to target the DNA, there are some general trends that are followed. The amino acids which possess charged functional groups or functional group which can form hydrogen bonds tend to be important for recognition. The most frequently used amino acids are Arg, Lys, His, Asn and Gln. To a much lesser extent, Asp, Glu, Ser and Thr are used, as well as the hydrophobic chains of the nonpolar amino acids.

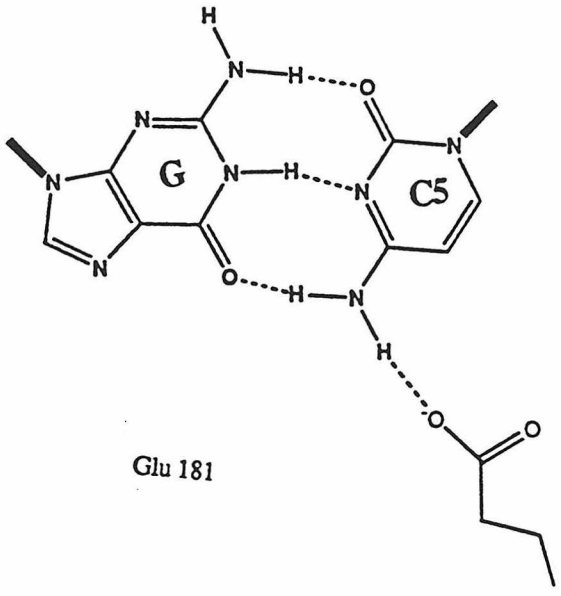
The crystal structure of the E. coli catabolite activator protein (CAP) [or cAMP (adenosine 3'-5'-monophosphate) receptor protein] is illustrative of how the recognition helix of HTH proteins interact with their DNA operator sites (29). The structure of CAP bound to a 30-base pair palindromic (5'-C_{15}GAAAAGTGTGACAT$_1$ATGTCACACTTT CG-3') DNA duplex was determined to 3 angstroms resolution. CAP is a transcription activator which requires the presence of cAMP for effect. The 209 amino acid protein contains a large amino-terminal domain which binds cAMP and a C-terminal domain which binds DNA. CAP-cAMP complex binds to the DNA duplex as a dimer (like most HTH proteins) and interacts with 27 of 30 base pairs in the crystallized DNA duplex. Importantly, the protein induces a large 90 degree bend in the DNA (8) which is important to the recognition process.

The recognition helix of HTH proteins is inserted into the major groove of the DNA and is oriented such that its helical axis is parallel to the plane defined by the DNA base pairs (Figure 1.6). Three side amino acid side chains make base specific contacts with the DNA. The guanidinium group of Arg 180 forms a bidentate contact with the N7 and

Figure 1.6. (Top) Illustration of the CAP dimer bound to its operator site. The DNA is bent by approximately 90 degrees. Adapted from ref. 29. (Bottom) Schematic illustration of hydrogen bonding interactions between Arg 180 and the N7 and O6 of guanine 7, and Glu 181 and the N4 of cytosine 5 of the operator site.


51°
43°


G7
C
Arg 180


G
C5
Glu 181

O6 of G7, the carboxylate of Glu 181 interacts with the N4 of C5, and Arg 185 interacts with either the O6 or N7 of G5 and the O4 of T6. The contacts by Arg 180 and Glu 181 were predicted by mutagenesis studies but not the Arg 185 contact (Figure 1.6). The sharp 90 degree bend that is induced in the DNA upon CAP binding is caused by a 40 degree kink between base pairs 5 and 6 (TG step) and an asymmetric 10 degree bend between base pairs 10 and 11. The 40 degree kink is caused by a substantial increase in the roll angle of base pair 6 which results in an unstacking of the base step. A large preference for a TG base step at this position has been demonstrated suggesting that this base step might be more easily unstacked than other base steps. The energy that is lost by this destacking is regained through phosphate contacts and through an increase in hydrophobic interactions between the protein and the DNA. In all, there are three base specific contacts between amino acids and the DNA bases and 11 phosphate contacts (per half site). Thus the direct readout of the DNA sequence by the recognition helix would appear to play a minor role in DNA binidng affinity relative to the indirect readout of the DNA.

It should be noted that CAP is an extreme example of the importance of indirect readout; however, indirect readout of the DNA sequence by other HTH proteins is also evident. Mutagenesis experiments on other DNA operator sequences have revealed that non-contacted base pairs can play an important role in the recognition process. For example, the central T·A base step of the P22 operator site is not directly contacted by the protein. However, mutations to a different sequence results in a 10-fold decrease in protein affinity (32). A similar result has been observed for the central bases in the 434 repressor operator site (33). The Trp repressor (34) also shows a large reliance on the sequence of non-contacted DNA bases. Thus, HTH recognize their DNA sites by both direct and indirect read-out of the DNA sequence.

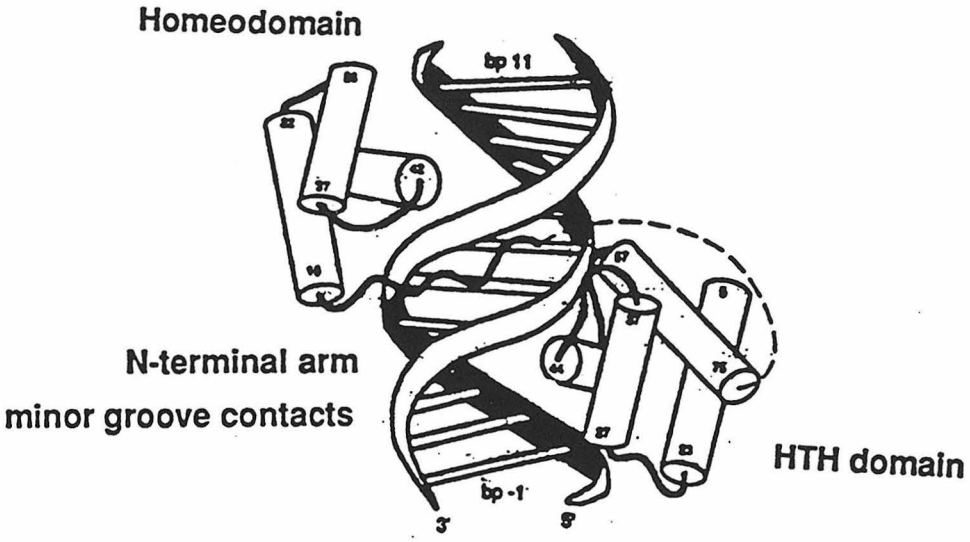
1.3.2. Homeodomain

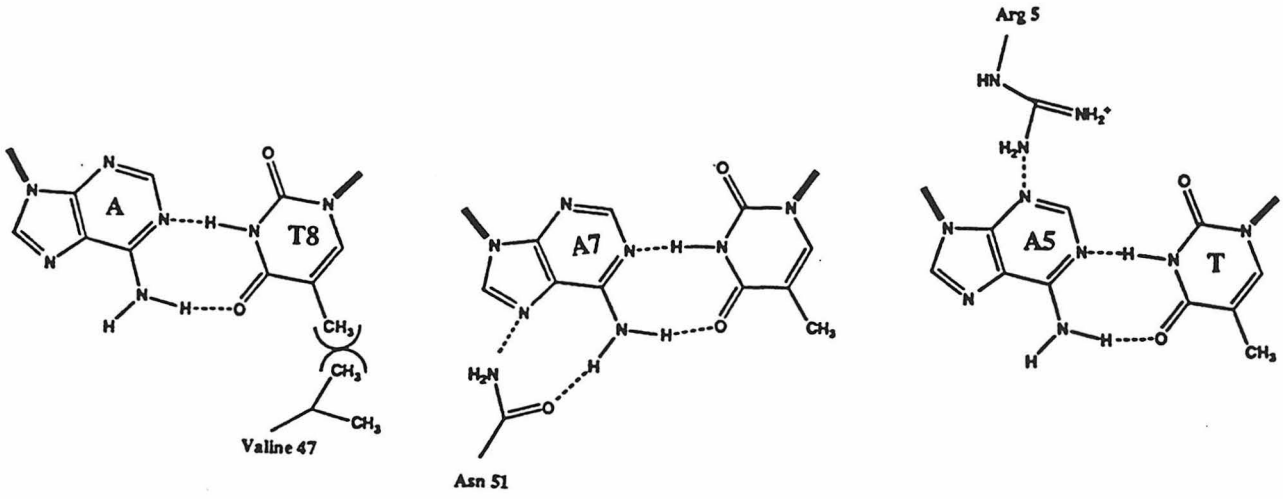
The homeodomain motif is very similar to that of HTH proteins. Both families insert a single α-helix into the major groove of the DNA; however, in HTH proteins, the recognition helix packs against a single α-helix at the protein interface, while in the homeodomain proteins, the recognition helix packs against a pair of antiparallel helices and this pair of helices serves to orient the recognition helix with respect to the DNA (24) (Figure 1.7). Another important difference is that homeodomain folds into a stable element when removed from the protein scaffold while HTH proteins do not. The X-ray structure of a number of homeodomain proteins have been determined, and they all appear to recognize their DNA sites in a similar manor.

One interesting feature of homeodomain proteins is that in addition to recognizing the DNA by insertion of an α–helix into the DNA major groove, the protein uses an amino-terminal arm region to reach around and make DNA contacts in the minor groove of the DNA (35-37). Many aspects of DNA recognition by homeodomain proteins are exemplified in the structure Oct-1 (38) bound to its eight base pair operator site, 5'-ATGCAAAT-3'. Oct-1 is an interesting protein in that it contains a 75 amino acid HTH recognition unit followed by a variable linker of 15-30 amino acids and a 60 amino acid homeodomain recognition unit. The HTH unit and the homeodomain units act in tandem to recognize the eight base pair operator site. The 5'-ATGC-3' unit is recognized by the HTH unit and the 5'-AAAT-3' is targeted by the homeodomain. The use of two recognition domains in tandem is in many ways analogous to the tendency of HTH to bind as dimers, but the strategy employed by Oct-1 (and zinc finger proteins) allows for the recognition of non-palindromic DNA sequences. The HTH unit of Oct-1 recognizes the sequence 5'-ATGC-3' in a manor which is completely analogous to other HTH proteins like the Λ and 434 repressor proteins and will not be described here.

The homeodomain region of Oct-1 resembles that of many HTH proteins. The DNA recognition unit consists of a three helix bundle in which the first two helices are

Figure 1.7. (Top) Illustration of the homeodomain protein Oct-1 bound to its DNA operator site. The HTH domain and the homeodomain regions of the protein act bind in tandem to bind consecutive major grooves of the DNA. The bridging linker arm makes contacts in the minor groove of the DNA. The DNA is nearly canonical B-form. Adapted from ref. 38. (Bottom) Schematic illustration of (left) van der Waals contacts between Val 47 and the methyl group of thymine 8 (center) hydrogen bonding interactions between the amido moiety of Asn 51 and adenine 7 (right) minor groove contact between the guanidinium moiety of Arg 5 and the N3 of adenine 5.


Homeodomain
bp 11
N-terminal arm
minor groove contacts
HTH domain
bp -1
3'
5'


A
T8
Valine 47
A7
Asn 51
Arg 5
A5
T

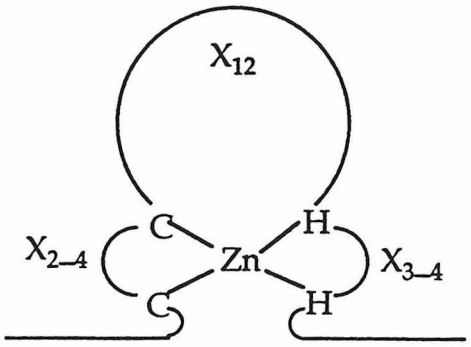
above the DNA site and the third helix packs against them and is inserted into the major groove of the DNA. Three amino acid side chains of the recognition helix, Val-47, Asn-51, and Cys-50, make well-defined contacts with the DNA base stack. Val-47 forms a van der Waals contact with the methyl group of T8. Asn-51 hydrogen bonds to the N6 (amino group) and N7 of A7 and Cys-50 appears to make a van der Waals contact with the methyl groups of T9 and T10 (two residues outside of the highly conserved octamer site). It should be noted that it is very rare for a cysteine residue to be involved in the direct read-out of the DNA sequence. The N-terminal arm of the homeodomain reaches around to the minor groove of the recognition site and Arg-5 appears to donate a hydrogen bond to the N3 of A5 (Figure 1.7). In addition to these base specific contacts, a number of phosphate contacts are also evident in the structure. However, unlike the CAP protein, the DNA is nearly canonical B-form. Thus, recognition by Oct-1 seems to be attributed mainly to direct readout of the DNA sequence.

1.3.3. Zinc-Finger Proteins

One of the most promising DNA recognition motifs for the development of proteins with new sequence selectivities is the zinc finger motif. The zinc finger motif is defined by the consensus sequence Cys-X_{2or4}-Cys-X_{12}-His-X_{3-5}-His and this invariant Cys-Cys-His-His (CCHH) arrangement coordinates a single zinc atom which is important for proper folding of the recognition domain (24). Importantly, a single zinc finger has been shown to fold into the proper secondary structure in the absence of the large protein scaffold (39). The zinc finger domain consists of an antiparallel β-sheet region followed by a region of α-helix (Figure 1.8) (40-42). Similar to Oct -1, zinc finger proteins use multiple domains in tandem to recognize large sequences. In fact, the Xenopus Xfin protein is thought to contain 27 distinct zinc finger domains (43).

It is the α-helical region of the zinc finger which primarily contacts the DNA, and x-ray analysis has shown that this portion of the finger resides in the major groove of the

Figure 1.8. (Top) Schematic illustration of a single folded zinc finger. (Bottom) Sequence of the three zinc finger domains of Zif268. Sequence is aligned to show the similarities between the three fingers.



1 5 10 15 20 25 30
MERPYA CPVES CDRRFS RSDELT RHIRIHTGQK

35 40 45 50 55 60
PFQ CRI__ CMRNFS RSD HLTT HIRTHTGEK

65 70 75 80 85 90
PFA CDI__ CGRKFA RSDERK RHTKIHLRQKD

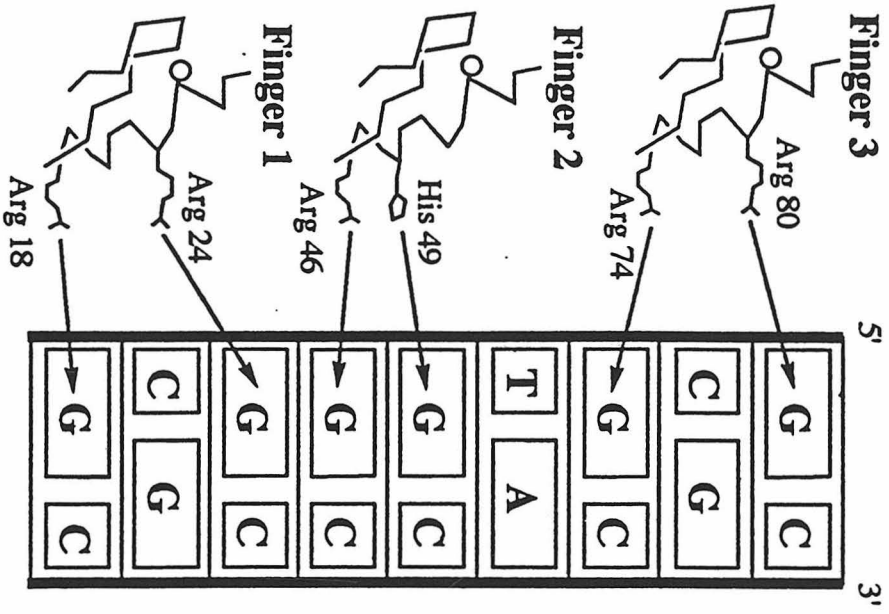
ANTI-PARALLEL BETA SHEET

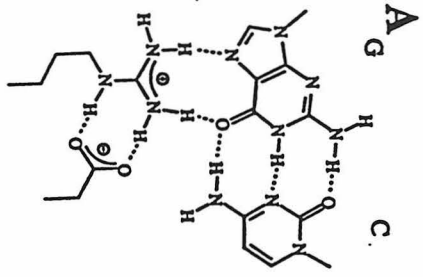
ALPHA HELIX

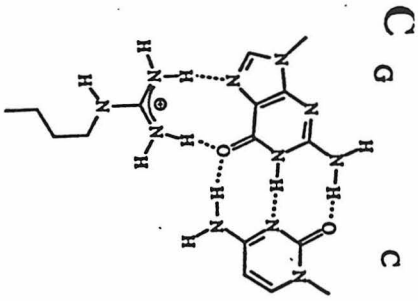
DNA (40,44). A number of different zinc fingers have been characterized in recent years (45) and our understanding of how these proteins recognize their DNA operator sites is more complete than any other family of DNA binding proteins, despite their relatively recent discovery. The mouse protein Zif268 is an excellent example of how zinc fingers recognize their DNA binding sites (40). This protein contains three zinc finger domains which act in tandem to recognize the guanine rich nine base pair site 5'-GCGTGGGCG-3'. Each domain recognizes a three base pair site and the protein is in antiparallel orientation relative to the DNA operator site such that finger 1 recognizes the terminal 5'-GCG-3' triplet (Figure 1.9). Zif268 recognizes its operator site by direct readout of the DNA sequence. Fingers 1 and 3 recognize the triplet 5'-GCG-3' and do so in similar ways. The arginine residue immediately prior to the α-helical region (-1) and an Arg in the sixth position of the α-helix recognize the third and first guanine residues of the DNA triplet by forming a bidentate contact with the N7 and O6 of this purine base (Figure 1.9). The Arg in the -1 position is also contacted by the Asp in the second position of the α-helix. This contact is believed to help orient the arginine with respect to the DNA. The second zinc finger domain which recognizes the triplet 5'-TGG-3' has an Arg residue in position -1 like fingers 1 and 3 and recognizes the 3' guanine of its triplet site; however, it lacks an Arg in the sixth position. Instead it has a His in the +3 position of the α-helix and uses this residue to target the central guanine of its recognition site.

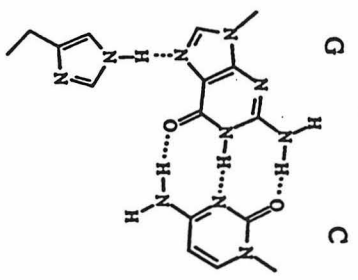
The mode by which Zif268 recognizes its DNA sequence exemplifies many of the key aspects of direct readout of the DNA sequence. The DNA operator site of Zif268 is nearly canonical B-form DNA in the crystal structure and the protein appears to code for guanine base pairs with arginine residues in the -1 and +6 positions of the α-helical region of the finger. The central DNA triplet of operator site differs from the outer triplets and this variation is reflected in the lack of a arginine residue in the +6 position and the presence of a histidine residue in the +3 position. Furthermore, the fact that a His residue can also be used to code for a guanine exemplifies the degeneracy of the code by which proteins

Figure 1.9. (Left) Illustration of the three fingers of Zif268 recognizing the sequence 5'-GCGTGGGCG-3'. (Right) Schematic illustration of interactions between the amino acids side chains of Zif268 and the guanine residues of the operator site. (A) The hydrogen bonding contact between the guanidinium moiety of the arginine in the -1 position of the α-helix and the N7 and O6 of guanine in the 3' position of the triplet site is illustrated. The Asp residue is used to help orient the guanidinium moiety. (B) The hydrogen bonding contact between the His in the +3 position of the a-helix and the N7 of guanine is illustrated for finger 2. (C) The hydrogen bonding contact between the guanidinium moiety of the arginine in the +6 position of the α-helix and the N7 and O6 of guanine in the 5' position of the triplet site is illustrated.


Finger 3
Arg 80
Arg 74
Finger 2
His 49
Arg 46
Finger 1
Arg 24
Arg 18
5'
3'
G C
C G
G C
T A
G C
G C
C G
G C


A
G
C


C
G
C


B
G
C

recognize DNA sites. The x-ray structure of another zinc finger protein, Gli, (44) bound to its DNA site has also been published and while the way in which the individual fingers bind to their DNA site it not identical, the overall mode of binding is very similar. One significant difference in the structures is that the DNA in the Gli structure in not B-form DNA. It shows a region of A-like DNA where the DNA is underwound and has a deep major groove. It also contains a region where the DNA is overwound relative to B-DNA and displays a shallow major groove.

Many research groups have been successful in altering the specificity of zinc finger proteins. Berg and co-workers have shown how it is possible to change the specificity of the zinc finger Sp1 by altering the identity of the residues at positions -1, +3 and +6 of the α-helical portion of the zinc finger domain (46). Pabo and co-workers have shown that a fusion protein containing the first two fingers of Zif268 and the homeodomain of the Oct-1 is specific for a sequence containing the homeodomain preferred binding site (5'-TAATTA) adjacent to the zinc finger site (5'-NGGGNG-3') (47). Most recently, both Pabo and Klug have demonstrated how phage display libraries can be used to select for Zif268 mutants which bind to almost any DNA sequence, and in most cases it is only necessary to modify positions -1, +2, +3, and +6 of the α-helical region of the zinc finger domain (48,49).

1.3.4. Steroid Receptor Proteins

At first glance, the steroid receptor proteins appear to be a subset of the zinc finger proteins which uses a 4 Cys arrangement to coordinate the zinc atom instead of a His-His-Cys-Cys arrangement (24, 51). However, the steroid receptor motif is definitely an example of a unique structural motif. The best way to demonstrate the uniqueness of the steroid receptor motif is by contrasting it with the zinc finger motif. The best characterized member of this family is the glucocorticoid receptor (50). An 86 amino acid fragment of the rat glucocorticoid receptor (residues 440-525) bound to two different DNA fragments

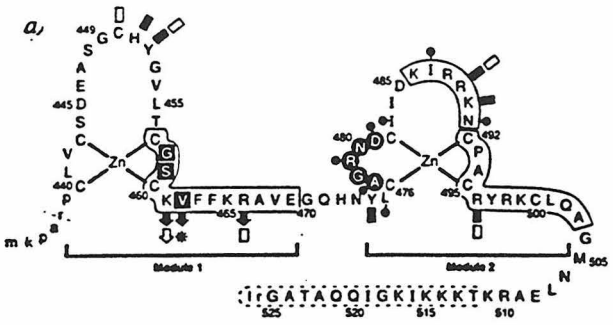
which differ in the spacing between consecutive receptor binding sites was recently reported (50). The amino acid sequence of the 86 amino acid DNA binding domain fragment is shown in Figure 1.10. As illustrated, the DNA binding element consists of two zinc binding modules which contain regions of β-sheet and α-helix. Importantly, members of the steroid receptor family all possess only two of these modules while there appears to be no limit to the number of modules that a zinc finger protein possesses.

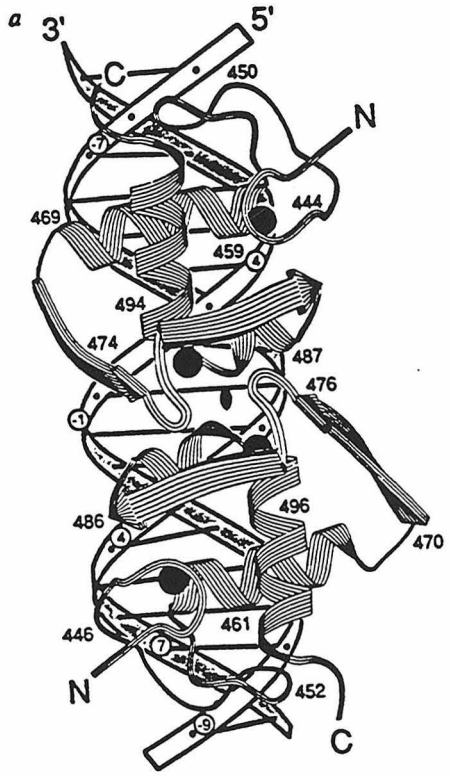
The glucocorticoid receptor protein binds to the GRE DNA as a dimer (4 modules) with each 2 module monomer recognizing a six base pair half site, 5'-AGAACA-3'. The first module of each monomer binds in the major groove of the DNA and makes base specific contacts, while the second module packs against the first module and only makes phosphate contacts. Thus while the fingers of the zinc finger proteins appear to act independently, the modules of the glucocorticoid receptor protein fold into a single globular unit. Base specific contacts between Arg 466 and the N7 and O6 of G4, Val 462 and the methyl group of T5 and a water mediated hydrogen bond between Lys 461 and G-7 are evident (Figure 1.10). DNA sequence discrimination not only relies on base specific contacts between the protein and the DNA, but it also requires the correct spacer length between consecutive DNA half sites. The protein requires a three base pair spacer between half sites for specific recognition. Thus, the glucocorticoid receptor protein recognizes its DNA operator site in a manner which differs significantly from zinc finger proteins.

1.3.5. Leucine Zipper Proteins

Leucine zipper proteins are a family of DNA binding proteins which were first discovered based upon amino acid sequence analysis (52). Analysis showed the presence of a heptad repeat of leucine residues followed by a region possessing a large number of basic residues, i.e., Lys and Arg (Figure 1.11). Subsequent studies revealed that the leucine region is highly helical in solution and is used as a dimerization domain in which the hydrophobic leucines of the two monomers pack against each other. The basic region

Figure 1.10. (Top) Sequence of the glucocorticoid receptor used for crystallographic analysis. Zinc coordination and α-helical regions (boxed) are shown. Residues which make dimer interface contacts are indicated by solid dots. Residues making phosphate contacts are indicated by solid rectangles. (Middle) Schematic representation of the glucocorticoid receptor interacting with its DNA operator site. Ribbon diagram indicates regions of α-helix and β-sheet. Zinc atoms are indicated by solid circles. (Bottom) Sketch indicating contacts between the recognition helix of the glucocorticoid receptor and the DNA. Illustrated are the direct contacts between K (Lys) 461, V (Val) 462, and R (Arg) 466. Figure is adapted from ref. 50.


a
Module 1
Module 2


a
3'
5'
C
N
450
444
469
459
494
474
487
476
496
470
486
461
446
452
N
C

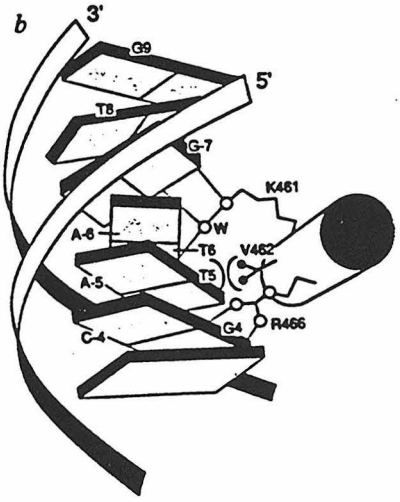
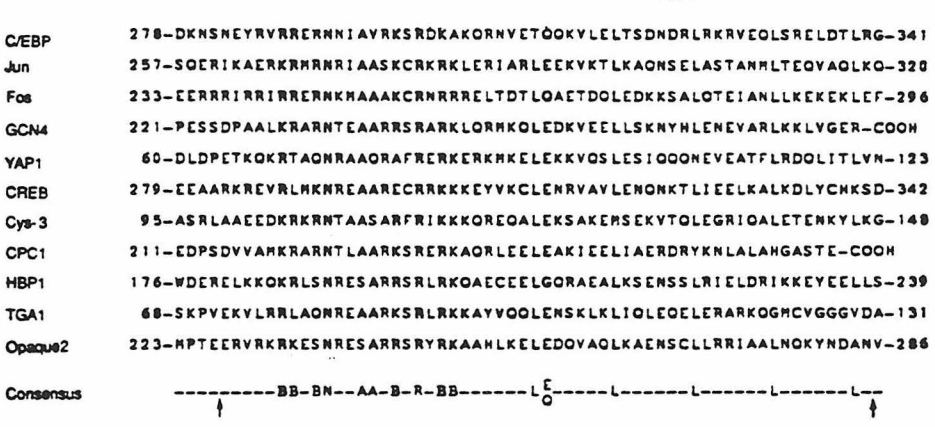
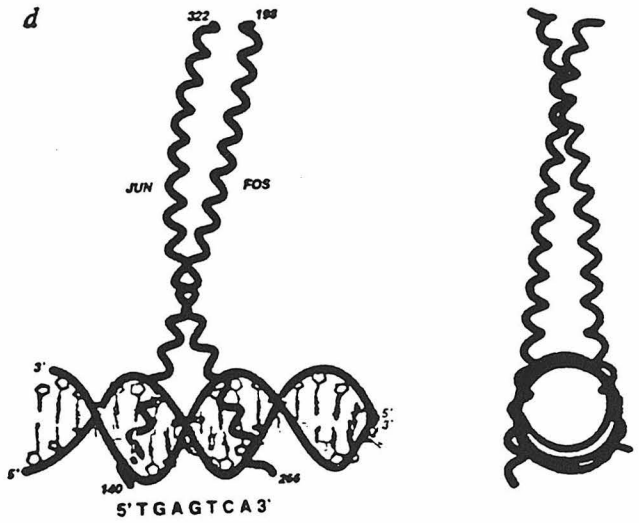
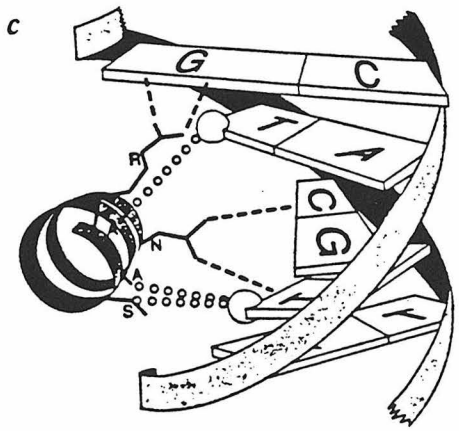

b
3'
G9
5'
T8
G-7
K461
W
A-6
T6
V462
T5
A-5
G4
R466
C-4

Figure 1.11 (Top) Amino acid sequence comparison for a number of different leucine zipper proteins. Lines mark conserved amino acids in the basic and leucine zipper regions. Adapted from ref. 52. (Middle) Overall structure of Fos/Jun dimer bound to DNA. Shown is two different views. Adapted from ref. 55. (Bottom) Sketch of the conserved contacts between the operator site and the α-helical basic region of Fos, Jun, and GCN4. Adapted from ref. 55.


Protein
BR–A
BR–B
Leucine zipper
C/EBP 278–DKNSNEYRVRRERNNIAVRKSRDKAKQRNVETQQKVLELTSDNDRLRKRVEQLSRELDTLRG–341
Jun 257–SQERIKAERKRMRNRIAASKCRKRKLERIARLEEKVKTLKAQNSELASTANMLREQVAQLKQ–320
Fos 233–EERRRIRRIRRERNKMAAAKCRNRRRELTDTLQAETDQLEDEKSALQTEIANLLKEKEKLEF–296
GCN4 221–PESSDPAALKRARNTEAARRSRARKLQRMKQLEDKVEELLSKNYHLENEVARLKKLVGER–COOH
YAP1 60–DLDPETKQKRTAQNRAAQRAFRERKERMKELEKKVQSLESIQQQNEVEATFLRDQLITLVN–123
CREB 279–EEAARKREVRLMKNREAARECRRKKKEYVKCLENRVAVLENQNKTLIEELKALKDLYCHKSD–342
Cys-3 95–ASRLAAEEDKRKRNTAASARFRIKKKQREQALEKSAKEMSEKVTQLEGRIQALETENKYLKG–148
CPC1 211–EDPSDVVAMKRARNTLAARKSRERKAQRLEELEAKIEELIAERDRYKNLALAHGASTE–COOH
HBP1 176–WDERELKKQKRLSNRESARRSRLRKQAECEELGQRAEALKSENSSLRIELDRIKKEYEELLS–239
TGA1 68–SKPVEKVLRRLAQNREAARKSRLRKKAYVQQLENSKLKLIQLEQELERARKQGMCVGGGVDA–131
Opaque2 223–MPTEERVRKRKESNRESARRSRYRKAAHLKELEDQVAQLKAENSCLLRRIAALNQKYNDANV–286
Consensus ----------BB–BN--AA–B–R–BB------L[E/Q]-----L------L------L------L--


d
322
198
JUN
FOS
3'
5'
5'
3'
140
264
5'TGAGTCA3'


c
G
C
T
A
C
G
R
N
S

is also α-helical and makes direct contacts with the DNA. Basic region swap experiments have verified that this region is primarily responsible for the sequence specificities of the leucine zipper proteins (24, 56).

Structural studies have shown that the dimerized protein is shaped like a thick, short stemmed Y in which the zipper region is the stem of the Y and extends out from the DNA while the forked region of the Y contacts the major groove on opposite sides of the DNA helix in a scissors grip type orientation (53). Importantly, leucine zipper proteins are able to form heterodimers which increases the potential number of sequences that a given set of zipper proteins can recognize. The best characterized members of the leucine zipper family are GCN4 (54,57) and the Fos/Jun heterodimer (55). Both dimers show similar packing of the zipper region as well as a similar scissors grip hold on the DNA operator site.

The Fos/Jun heterodimer recognizes the asymmetric AP-1 site, 5'-TGAGTCA-3', and is capable of binding the site in either of two orientations which differ by a 180 degree rotation about the dyad axis of the complex (55). In both orientations, the central guanine residue is targeted with the guanidinium group of the amino acid arginine. Other contacts include hydrophobic interactions between an alanine residue and the methyl group of thymine and a cross strand bidentate contact between an asparagine residue and the amino group of cytosine and the O4 of thymine (Figure 1.11). Surprisingly, basic regions of Fos, Jun, and GCN4 all interact with the AP-1 site in the same way (55, 57). It should also be noted that the DNA operator site shows a slight 10 degree bend towards the coiled-coil (zipper region), and another leucine zipper protein, GCN4, has been shown to distinguish between DNA target sites based on an intrinsic bend in the DNA (58,59).

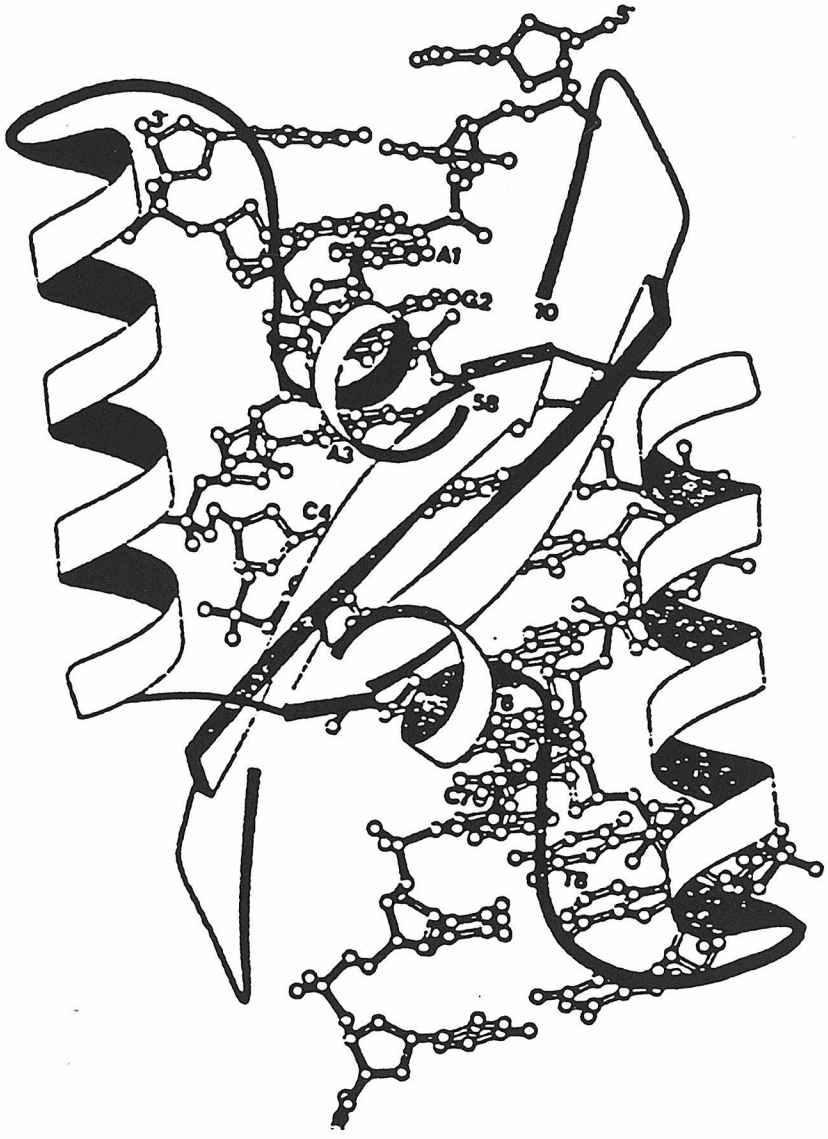
A number of research groups have investigated whether it is possible to replace the dimerization domain of the leucine zipper proteins with a different kind of dimerization element. If the leucine zipper domain is replaced with a disulfide bond, this basic region only complex can still bind its target site (60). Schepartz and co-workers have recently demonstrated that replacement of the dimerization domain of GCN4 with a metal

dimerization domain results in viable metal protein complexes (61). Importantly, a series of peptide/metal complexes were synthesized in which the relative orientation of the basic regions was varied and it was discovered that both the affinity and specificity varied depending on the orientation of the basic regions.

1.3.6. β-Sheet Motif

All of the above mentioned DNA binding proteins recognize their operator sites by inserting an α-helix into the major groove of the DNA. All that really varies in each system is that the protein uses a different structural motif for the presentation of the α-helix to the DNA. There exists a small family of DNA repressor proteins that use a β-sheet motif to recognize their DNA site (62). The structures of the Arc repressor (63, 64), the related Mnt repressor (65) and the Met repressor (66) have been determined and all three of these proteins insert a double-stranded antiparallel β-sheet into the major groove of the DNA. The E. coli met repressor protein controls the expression of many of the proteins in the methionine biosynthetic pathway (67). This 104-amino acid protein recognizes the palindromic consensus sequence 5'-AGACGTCT-3' (met box). The protein binds cooperatively as a dimer and the crystal structure of the met repressor protein dimer bound to the 19-base self-complementary oligonucleotide, 5'-$TTA_1GACGTCT_8A_9GACGTCT$ A_{17}-3' (containing two met boxes), was solved to 2.8 angstroms resolution. Repressor binding requires the presence of S-adenosylmethionine (SAM). Each subunit of the dimer is made up of three α-helices and a single β-sheet. In the bound dimer, the antiparallel double-stranded β-sheet is inserted into the major groove centered on met box (Figure 1.12). The side chains of Lys 23 and Thr 25 make hydrogen bonds to the edges of the DNA base pairs. Lys 23 hydrogen bonds with the O6 and N7 of G2 (and G10') and Thr 25 hydrogen bonds with the N7 of A3 (and A11'). The DNA in this crystal structure is mainly B-form. Surprisingly, the T_8A_9 base step between the met boxes is not contacted by the protein, but if it is mutated from a TA to a GC or AT step, a 76-fold decrease in

Figure 1.12. Sketch of the Met repressor dimer interacting with the met box. Ribbon diagram is used to indicate regions of α-helix and β-sheet. Adapted from ref. 66.


A1
C2
T10
S8
A3
C4
G
C7
T6

binding affinity is seen. This step is overwound in the crystal structure, and this extra winding must be necessary for proper orientation of the two met boxes with respect to each other. The central ten base pairs of the DNA duplex are B-form DNA. However, both ends show a 25 degree bend towards the major groove. Thus a large element of indirect readout of the DNA sequence is evident in this crystal structure. It should be noted that despite the difference in protein structural elements being used, the method by which the met repressor recognizes its DNA site very closely resembles that of other DNA binding proteins.

1.4. General Principles of DNA Recognition by Proteins

At first glance it appears that proteins use a diverse array of structural motifs to recognize their DNA operator sites. However, careful inspection reveals that these different recognition motifs are all very closely related. The HTH, zinc finger, leucine zipper, homeodomain, and steroid receptor families all insert an α-helix into the major groove of DNA, and side chains of the amino acids in the α-helix make contacts with the DNA base stack. To make matters simpler, out of the 20 possible amino acid side chains, the only ones used with any great frequency are arginine, lysine, histidine, asparagine and glutamine. The β-sheet motif used by the Met, Arc and Mnt repressor proteins represents a unique mode for DNA recognition in that it not only uses an anti-parallel double stranded β-sheet to recognize its DNA site, but each strand of this double stranded β-sheet comes from a different monomer of the dimer. However, the protein still binds in the major groove of the DNA and the same amino acid side chains employed in the α-helical recognition modes are used. The primary difference between all of the different recognition motifs is the way in which the protein presents the recognition elements to the DNA. In HTH proteins and homeodomain proteins, the protein uses one (or two) α-helices to assure proper orientation of the recognition helix with respect to the rest of the protein. In the zinc finger and steroid receptor proteins, a coordinated zinc atom is used to

insure proper folding and hence orientation of the DNA recognition element. In the leucine zipper proteins, it appears that the dimerization domain is used to orient the protein.

In addition to making base specific contacts with the DNA, every DNA binding protein appears to make a number of nonspecific contacts with the sugar phosphate backbone. These contacts increase the non-specific DNA binding affinity of the protein, unless formation of the interaction requires a unique DNA structure. Most DNA binding proteins have a nonspecific DNA binding affinity of 10^5-10^6 M^{-1}. By having a moderate nonspecific affinity for DNA, the proteins increase the chances of the protein being associated with the DNA and finding the specific site. Similarly, the DNA binding domains tend to be basic in nature, which results in a net positive charge on the DNA binding domain and increases the chances of the protein binding to the negatively charged DNA polymer. Proteins like CAP gain sequence specificity through phosphate contacts because many of these contacts can only be made to the distorted DNA binding site. In contrast, phosphate contacts in proteins like Zif268 should not help sequence specificity because most DNA sequences should be in position to make the contacts.

Most DNA binding proteins use some aspect of dimerization to increase their specificity. HTH and β-sheet motif proteins tend to form symmetric dimers with themselves. Leucine zipper proteins can form either symmetric or asymmetric dimers and zinc finger proteins form "covalent" dimers. Since many DNA binding proteins bind as homo-dimers, many operator sites are palindromic. Non-covalent dimer formation has many advantages over covalent dimer formation, and for this reason most proteins form noncovalent dimers.

A number of different DNA polymorphisms have been shown to be important for DNA recognition. DNA bending appears to be most frequently used; however, helical twist variation is also important. Conceptually, it is easy to understand how shape selection works. What isn't understood is why only some proteins possess a large element of shape selection in the recognition of their DNA sites. As mention above, the DNA binding

protein CAP shows a large bend in its operator site which is important for recognition, and the met repressor shows both a 50 degree bend and an overwinding of the central base step. Other operator sites such as the recognition site of Zif268, and Oct-1 are nearly canonical B-form DNA. Why do some proteins recognize severely distorted DNA sites and not others? It is possible that some recognition motifs, like zinc fingers are better suited for direct readout of the DNA sequence and tend to recognize sequences which prefer to be B form. As more DNA/protein structures become available, maybe a clearer trend will emerge; however, at the present time the answer is not known.

Our challenge as scientists remains to take advantage of what we have learned about DNA recognition for proteins and apply it in the development of small molecular probes for DNA. It is likely that many of the principles of recognition for proteins cannot be applied directly to small molecule recognition. However, concepts such as shape selection and recognition by hydrogen bonding and van der Waals contacts are important for small molecules and many of these principles have already been demonstrated in small molecule recognition.

1.5. DNA Recognition by Small Molecules

The development of small molecules for the sequence specific recognition of DNA is a challenging goal. In general, small molecules, like cAMP or a steroid like estrogen, usually owe their function to interactions with proteins. The vast majority of pharmaceuticals that are currently available for the treatment of diseases rely on some protein mediated interaction for their effect. However, some small molecules, like Fe(II)-bleomycin (68) and cis-platin (111), owe their effectiveness as drugs to direct interactions with DNA. Footprinting agents like MPE-Fe(II) (4) are used extensively to investigate the DNA binding specificities of proteins and small molecules. Anthracycline based antibiotics such as the pluramycins (74) bind by intercalation into the minor groove of DNA . Classical minor groove binders such as distamycin target stretches of DNA with

four to five contiguous A·T base pairs (75). Purely synthetic molecules like $Rh(phen)_2phi^{3+}$ have been shown to recognize DNA sites from the major groove and have a sequence specificity which can be attributed to shape selection (76-78). By studying these molecules, the rules that govern the recognition of DNA by small molecules can be learned. With this information, it might be possible to design next generation small molecules capable of exquisite sequence specificities.

The majority of small molecules currently being studied which bind DNA noncovalently either bind by intercalation or bind non-intercalatively in the minor groove of DNA. It is believed that small molecules prefer the smaller minor groove of B-DNA because it allows for better van der Waals contacts and increased charge neutralization (79). Since small molecules primarily bind in the minor groove and proteins primarily bind in the major groove, it seems likely that some of the rules for DNA recognition might be different.

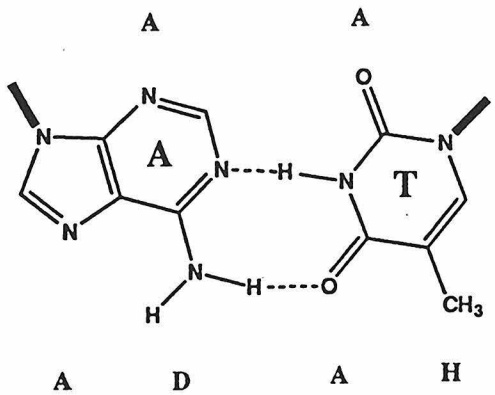
The preference of small molecules for the minor groove of DNA can make the design of sequence specific DNA binding agents problematic. The minor groove of DNA is a virtual desert of hydrogen bonding and van der Waals contacts compared to the major groove. Furthermore, inspection of the arrangement of hydrogen bond donors, acceptors, and hydrophobic groups in the minor groove reveals that it is very difficult to distinguish AT from TA base pairs and GC from CG base pairs (Figure 1.13). In contrast, each base pair has very distinct features when viewed from the major groove, which allows for more precise recognition. Thus, it is very difficult to design minor groove binding agents with sequence selectivities comparable to what has been seen for proteins.

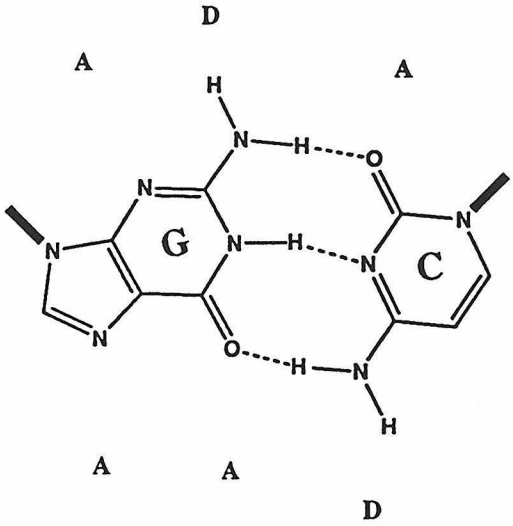
1.6. Nonintercalating Groove Binding Agents

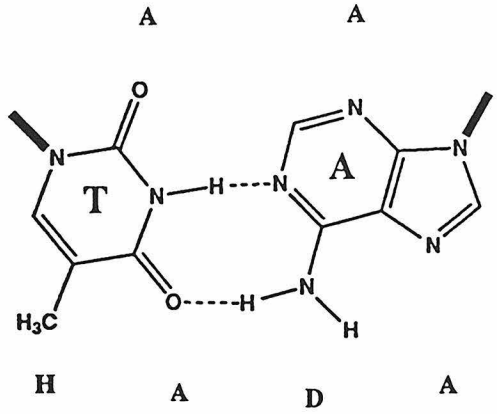
1.6.1. Distamycin and Related Molecules

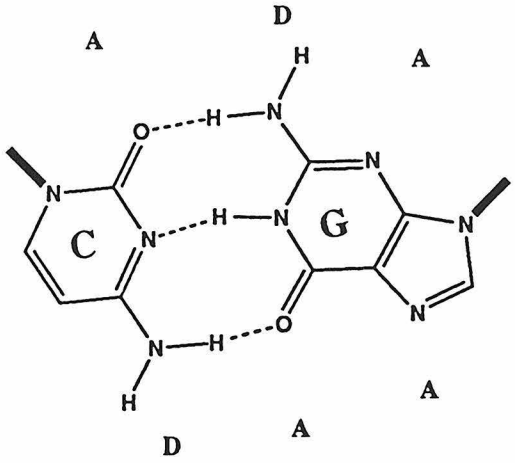
Distamycin and netropsin are naturally occurring antitumor agents which bind in the minor groove of DNA at sites of four or more continuous A·T base pairs. These

Figure 1.13 Comparison of the pattern of hydrogen bond donors (D), acceptors (A) and hydrophobic groups (H) for the major and minor grooves of the four possible base pairs arrangements. It is difficult to distinguish AT from TA base pairs, and GC from CG base pairs from the minor groove.


A
A
A
T
CH3
A
D
A
H


D
A
A
G
C
A
A
D

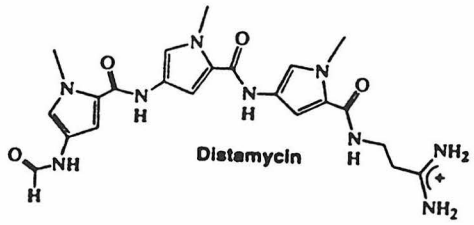

A
A
T
A
H3C
H
A
D
A

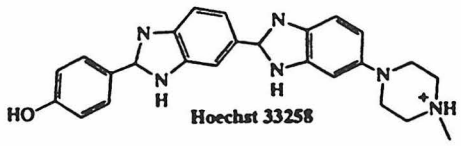

D
A
A
C
G
A
D
A

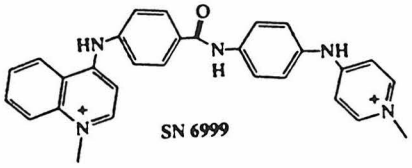
crescent shaped molecules bind in a side on fashion in which hydrogen bonds, van der Waals contacts, and electrostatics all contribute to the binding affinity and specificity (Figure 1.14). It is believed that part of sequence specificity exhibited by netropsin and distamycin can be attributed to the fact that long tracts of adenines exhibit a narrower minor groove than mixed DNA sequences, thus promoting better van der Waals contacts with the walls of the minor groove (75). Additionally, the bulky amino group of GC base pairs sterically interferes with tight groove binding between the ligand and the floor of the minor groove. An interesting feature of the drug distamycin is that at sites of 5 or more continuous A·T base pairs, distamycin can bind to DNA as either a 1:1 complex or as a 2:1 complex. To accommodate the second ligand in the 2:1 complex, it is estimated that the groove must widen by approximately 3.5 angstroms (75) (Figure 1.15). Thus, whether distamycin binds as a 2:1 complex or as a 1:1 complex can be used to assess minor groove width and flexibility. It was discovered that sequences such as 5'-AAAAA-3' favor 1:1 binding while more flexible sequences like 5'-ATATA-3' only bind in the 2:1 mode.

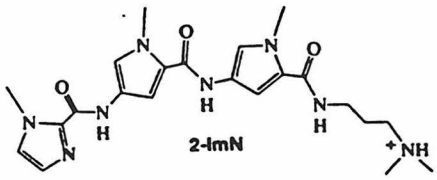
Attempts at expanding the recognition of netropsin and distamycin to include G·C base pairs has focused on the incorporation of strategically placed hydrogen bond acceptors. Derivatives like 1-methylimidazole-2-carboxamide-netropsin (2-ImN) have been synthesized to make use of the hydrogen-bonding capabilities of the imidazole group to select for sequences with G·C base pairs (Figure 1.15). 2-ImN binds to the sequence 5'-(A,T)G(A,T)C(A,T)-3' as a side-by-side anti-parallel dimer in which the imidazole nitrogen interacts with the guanine C-2 amino group (80-84) (Figure 1.15). Another molecule, ImPImP (imidazole-pyrrole-imidazole-pyrrole), has been designed to recognize an all G·C site (85,86) (Figure 1.15). This complex also binds as a 2:1 dimer and targets the sequence 5'-(A,T)GCGC(A,T)-3'. Similar to distamycin and netropsin, the sequence specificity can be attributed to van der Waals contacts and hydrogen bonding.

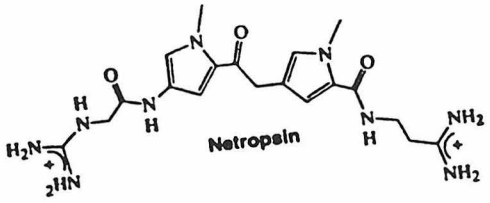
Figure 1.14. Schematic illustration of some minor groove binding drugs.


Distamycin


Hoechst 33258


SN 6999


2-ImN


Netropsin

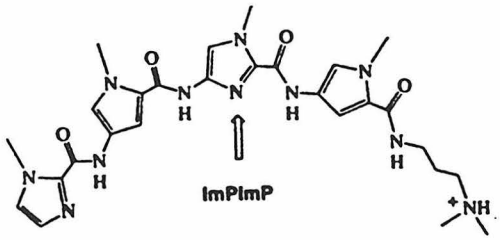
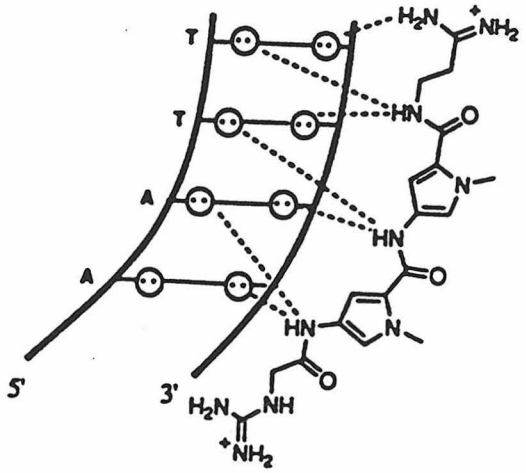
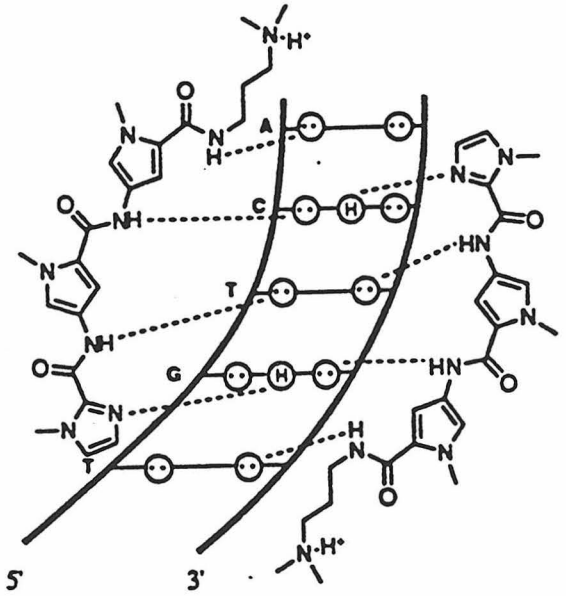

ImPImP

Figure 1.15. (Top) Illustration of the 1:1 complex between netropsin and 5'-TTAA-3'. Dotted lines indicate hydrogen bonds between netropsin and N2 of adenine and the O2 of thymine. (Bottom) Homodimeric binding of 2-ImN and 5'-TGTCA. Dotted lines indicate hydrogen bonds between the complex and the O2 of thymine and the N3 of adenine (dotted open-circles), as well as hydrogen bonds to the amino group of guanine (circled H). Adapted from ref. 83.



B



1.6.2. Hoechst 33258 and Other Dyes

Hoechst 33258 (Figure 1.14) is a fluorescent dye which is frequently used in the imaging of DNA. Similar to netropsin and distamycin, Hoechst 33258 binds at sites of four or more contiguous A·T base pairs with a side on geometry and has a binding affinity of approximately 5 x 10^8 M^{-1} at 5 °C(75, 87). This positively charged crescent shaped molecule recognizes DNA by forming hydrogen bonds to the N3 and O2 of adenine and thymine and through van der Waals contacts with the walls of the minor groove.

The synthetic drug SN-6999 has a higher overall charge than Hoechst 33258 and a similar number of aromatic rings for interactions with the walls of the minor groove, yet SN-6999 has been shown to have a binding affinity of only 2 x 10^6 M^{-1}. This 100-fold decrease in binding affinity is most likely due to the shape difference between the two molecules (75,88). Hoechst 33258, like distamycin and netropsin, is crescent shaped and mirrors the surface of the minor groove, while SN-6999 is more extended and is not able to fully match the inherent cylindrical nature of the minor groove. Hence, shape complementarity is important for determining the binding affinity of small molecules.

1.7. Minor Groove Intercalators

A number of potent antitumor agents and antibiotics are avid DNA intercalators. These molecules all contain a planar, aromatic, multi-ring unit for insertion in the DNA helix. The exact nature of the ring unit and the substitution pattern around the ring varies from system to system. The most famous DNA intercalator is the fluorescent dye ethidium bromide. However, many important therapeutic molecules owe their biological activity to intercalation. The process of intercalation requires a slight unwinding of the DNA helix (usually about 20 degrees) and an increase in the helical rise from 3.4 to 6.8 angstroms at the site of intercalation (89). Studies have shown that the favorable energetics of protecting the intercalator from solvent more than compensates for the energy required to deform the DNA. The majority of molecules appear to intercalate from the minor

groove; however, numerous metallointercalators which bind from the major groove have been reported.

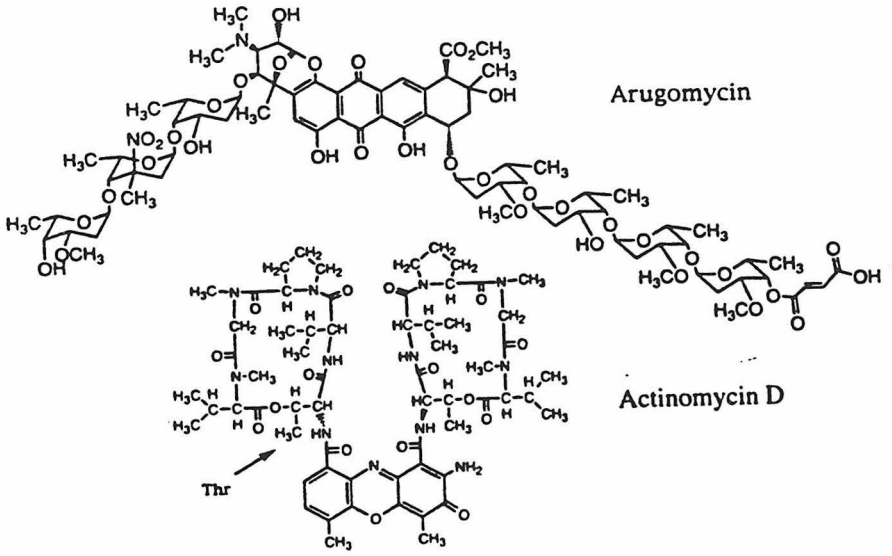
The anthracycline-based antibiotics nogalamycin(90), arugomycin (91), and pluramycin (92) all possess ring systems with two appended oligosaccharide units and all bind to DNA via intercalation (Figure 1.16). An interesting feature of these molecules is that upon binding, the anthracycline unit intercalates into the DNA base stack, and a sugar moiety projects into *both* the major and the minor groove of the DNA (Figure 1.17). For this to occur, one of the sugar moieties must be threaded through the DNA helix. This threading process requires a major distortion of the DNA helix and illustrates the flexibility of DNA. The sequence specificity of these antibiotics is surprisingly limited and has been demonstrated to rely more on the different chemical reactivities of the base pairs than on noncovalent interactions (92).

Another intercalator, actinomycin D, preferentially intercalates in GC base steps from the minor groove side (Figure1.16). This peptide based antibiotic binds to DNA such that both peptide rings are located in the minor groove (Figure 1.17). This molecule's sequence specificity is believed to be due to hydrogen bonds between the backbone carbonyls of threonine residues in the peptide rings and the amino group of guanine (93), thus demonstrating an element of direct readout of the DNA sequence. Additionally, binding of Actinomycin D requires a substantial widening of the minor groove of the DNA, and the wider minor groove of GC sequences should help contribute to binding specificity.

1.8. Metal Complexes as DNA Recognition Agents

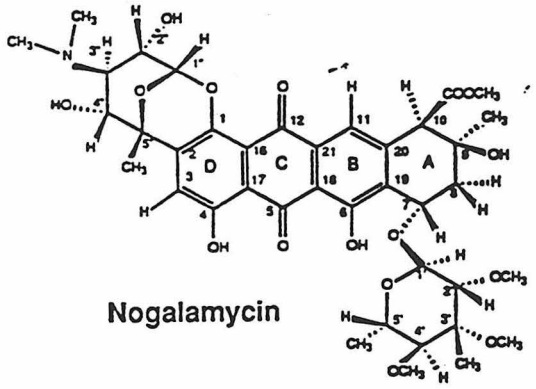
The diverse chemistry and geometries available to transition metal complexes makes them ideal probes for DNA structure. Simple molecules like Fe(II)·EDTA rely on the metal center for their DNA cleavage properties. The extensively studied Fe(II)·bleomycin (Fe-bleo) makes use of the metal center for both chemical reactivity and

Figure 1.16. Schematic illustration of some antibiotic intercalators. Adapted from ref. 75 and 92.

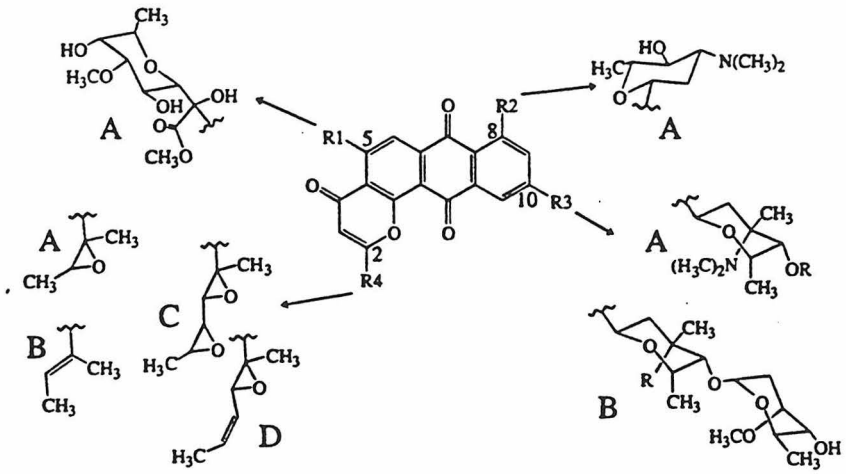


Arugomycin

Actinomycin D

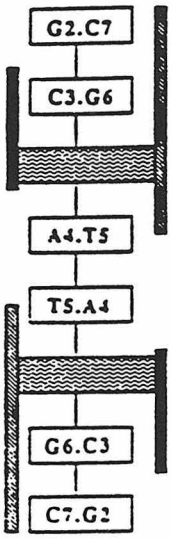


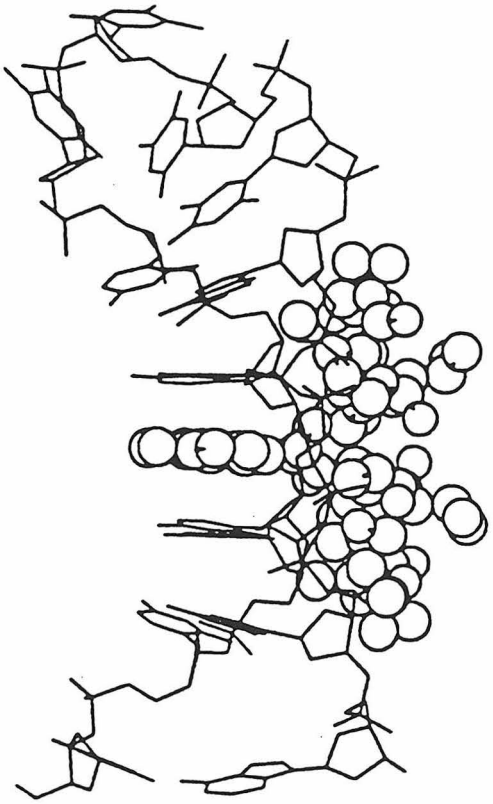
Nogalamycin



Compound	R1	R2	R3	R4	R'
altromycin B	A	H	B	A	$N(CH_3)_2$
altromycin H	OH	H	B	A	$N(CH_3)_2$
altromycin I	OH	H	B	A	$NHCH_3$
akinomycin	CH_3	H	A	C	H
pluramycin	CH_3	A	A	D	acetyl
neopluramycin	CH_3	A	A	B	acetyl
hedamycin	CH_3	A	A	C	H
rubiflavin A	CH_3				

Figure 1.17. (Top) Schematic illustration of nogalamycin intercalated in DNA. Notice the protrusion of moieties into both grooves of the DNA. Adapted from ref. 90. (Bottom) Illustration of Actinomycin D bound to DNA. The aromatic portion of the molecule intercalates into DNA and the peptide rings reside in the minor groove of the DNA. Adapted from ref. 93.


G2.C7
C3.G6
A4.T5
T5.A4
G6.C3
C7.G2



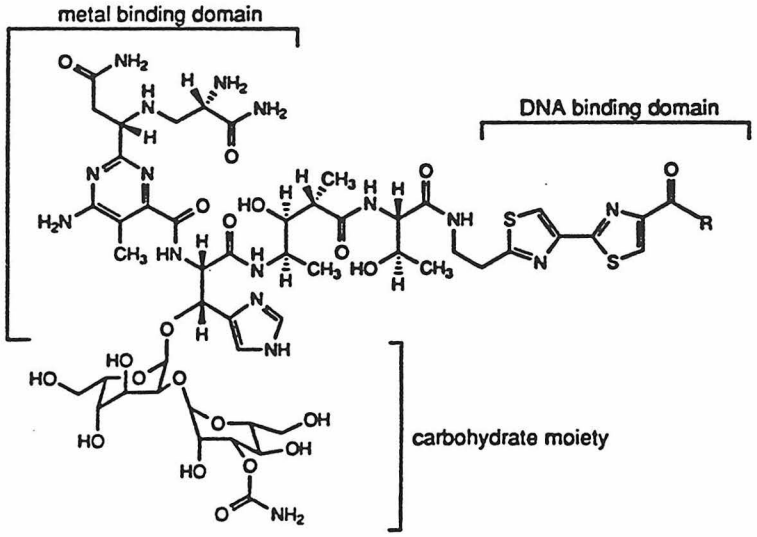
structural stability (68). Fe-bleo represents a highly evolved DNA cleaving agent which has served as a blueprint for many synthetic transition metal based DNA binding agents and is worthy of detailed examination.

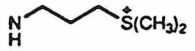
1.8.1. Iron(II)-Bleomycin

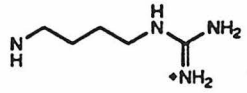
Fe-bleo has three distinct domains (Figure 1.18): a DNA binding domain, a metal binding domain, and a carbohydrate moiety. The carbohydrate region of bleomycin is thought to be important for membrane permeability and tumor recognition, and in some instances, the carbamoyl group of the mannose ring has been shown to participate in metal coordination (69). The DNA binding domain is a variability substituted bithiazole moiety and different members of the bleomycin family are distinguished by the substitution of the bithiazole moiety. It was originally thought that the bithiazole moiety of Fe-bleo was the primary determinant of the sequence selectivity of Fe-bleo. However, recent studies have shown that while the bithiazole moiety provides substantial DNA binding affinity via partial intercalation and electrostatic interactions in the minor groove of B-DNA, it does so in a sequence neutral fashion (70).

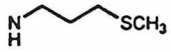
The heart of Fe-bleo is its metal binding domain. The metal binding domain of bleomycin coordinates a variety of transition metals (68) in a square pyramidal arrangement (71) (Figure 1.18) using the primary and secondary amines of the β-aminoalaninamide moiety, the N-1 of the pyrimidine ring, and the N-3 of the imidazole, and surprisingly, the deprotonated amide of the histidine moiety also coordinates the metal center. The metal domain not only is responsible for the chemical reactivity of Fe-bleo, but it also appears to be responsible for the predominantly 5'-GT-3' and 5'-GC-3' sequence specificity of the complex. The NMR solution structure of a zinc substituted bleomycin bound to a small DNA duplex has verified that the molecule is bound in the minor groove of DNA (72). The NMR data suggests that the minor groove of the DNA is distorted

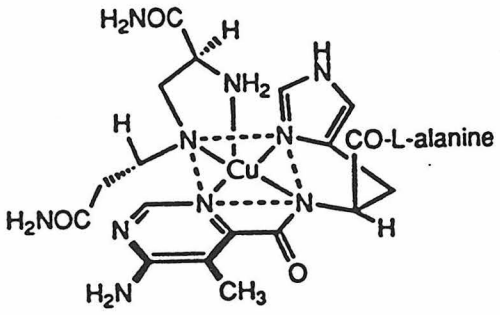
Figure 1.18. (Top) Illustration of bleomycin. The molecule possesses three distinct domains: the metal binding domain, the DNA binding domain and the carbohydrate moiety. Adapted from ref. 68. (Bottom) The metal binding domain of bleomycin has been shown to coordinate a metal ion in a square pyramidal arrangement. Adapted from ref. 112.



bleomycin A_2 R = 

bleomycin B_2 R = 

bleomycin demethyl A_2 R = 



upon complex binding; however, the exact nature of interactions that the complex uses to determine its sequence specificity is not known.

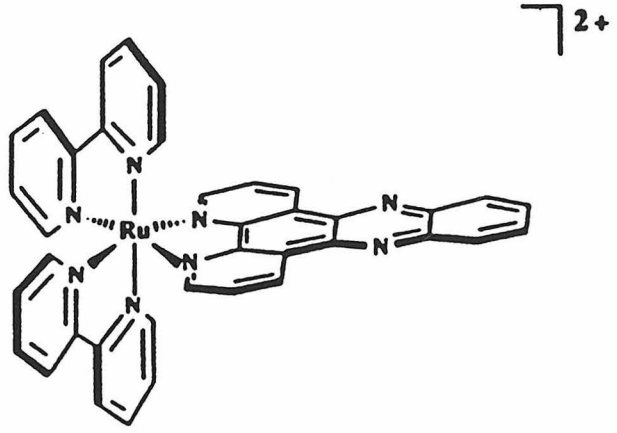
The metal domain is responsible for the DNA damage which gives Fe-bleo its therapeutic usefulness. An iron-oxo species is thought to be responsible for the cleavage process, and much evidence suggests that the active species is an unusual Fe(V)-oxo species. However, other reactive species cannot be strictly ruled out (68). Strand scission is caused by abstraction of the C4'-hydrogen of the ribose sugar followed by combination with O_2 or H_2O. Subsequent rearrangements lead to strand scission (73,68). Importantly, the C4'-hydrogen is known to reside in the minor groove of B-DNA and further supports the notion of the bleomycin complex binding in the minor groove.

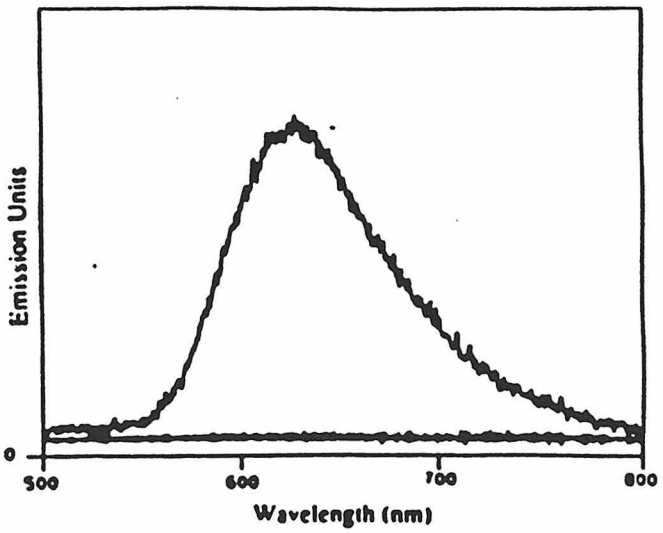
1.8.2 Rhodium and Ruthenium Based DNA Probes

Unlike most other small molecules which bind DNA, the rigid geometries of transition metal complexes allowed us to build well defined probes of DNA structure, and by varying the central metal core, it has been possible to tune the spectroscopic properties and chemical reactivities of the metal complexes. Ru(II) polypyridyl complexes have been shown to display dramatic increases in their luminescence properties upon binding to DNA. The complex bis(2,2'-bipyridine) dipyrido[3,2-a:2',3'-c]phenazine ruthenium(II), $[Ru(bpy)_2dppz]^{2+}$, shows no luminescence in aqueous solution due to quenching by water. However, upon intercalation into the DNA, the phenazine nitrogens are protected from solvent and the complex shows intense luminescence (94,95) (Figure 1.19). Importantly, this molecular "light switch" varies in luminescence intensity depending on the global conformation of the DNA, demonstrating that the molecule is capable of sensing the local geometry of the DNA site. Research is currently focused on developing biosensors based on the photoluminescence properties of $Ru(bpy)_2dppz^{2+}$ and similar molecules (96).

The nature of the binding interactions between tris(phenanthroline) Ru(II) $[Ru(phen)_3^{3+}]$ and B-DNA are governed by shape-selection (97). The Δ-enantiomer is

Figure 1.19. (Top) Schematic illustration of the luminescent probe $Ru(bpy)_2dppz^{2+}$. (Bottom) Study-state emission spectrum of $Ru(bpy)_2dppz^{2+}$ in the absence (flat line) and the presence of B-form poly d(GC). In aqueous solution, no luminescence of $Ru(bpy)_2dppz^{2+}$ is observed. However, upon addition of DNA, a large emission signal is observed. Adapted from ref. 94.


2+
N
N
Ru
N
N
N
N
N
N


Emission Units
0
500
600
700
800
Wavelength (nm)

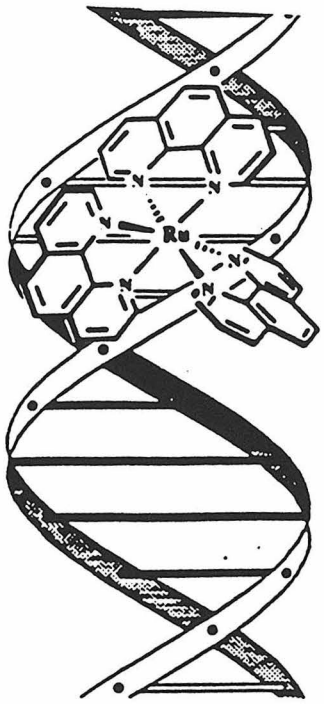
favored for binding by intercalation in the major groove of B-DNA over the Λ-enantiomer due to unfavorable steric clashes between the non-intercalating ancillary ligands and the sugar-phosphate backbone (Figure 1.20). In contrast, the Λ-enantiomer is favored for surface binding in the minor groove. The complex tris(3,4,7,8-tetramethylphenanthroline) Ru(II) or $[Ru(TMP)_3]^{2+}$ is sterically excluded from binding by both intercalation and groove binding in B-DNA due to the bulky substitutions on all three ligands (2). However, the Λ-enantiomer of the complex is able to bind to the wide, shallow minor groove of A-DNA (Figure 1.21).

The metal complex tris(4,7-diphenyl-1,10-phenanthroline) Rh(III) or $Rh(DIP)_3^{3+}$ has been shown capable of probing folded DNA structures (15-17). This bulky, hydrophobic molecule which lacks hydrogen bonding functionalities binds to unusual DNA structures like cruciforms and holiday junctions (15,16). This unique specificity is thought to be attributed to the ability of these DNA protrusions to offer a large hydrophobic surface which protects the molecule from water. $Rh(DIP)_3^{3+}$ has also been shown to recognize what might be potential "sign posts" which mark the beginning and end of genes in SV40 (Figure 1.22) (110). This complex also cleaves adjacent to the branch points of introns in Adenovirus E1A genes and the Simian Virus 40 T-antigen (17,98) (Figure 1.23). The presence of structural markers next to intron branch points suggests a mechanism for the splicing of genes. Work aimed at deciphering the nature of these structural markers has only succeeded in partially characterizing these unusual structures (17), and more work needs to be done.

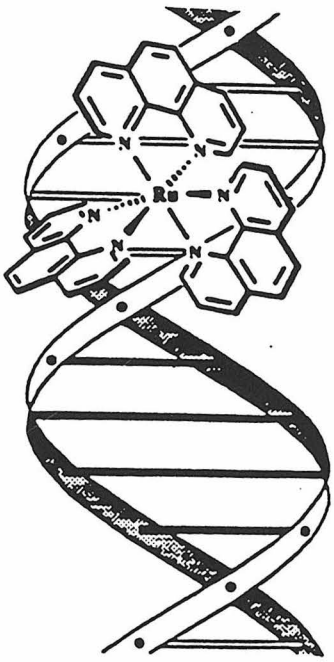
1.8.3. Phenanthrenequinone Diimine Complexes of Rhodium(III)

Phenanthrenequinone diimine (phi) complexes of rhodium(III) occupy a unique niche in the development of sequence specific DNA binding molecules. Phi complexes of rhodium(III) bind to B-DNA via intercalation of the phi ligand in the major groove of DNA and upon photoactivation promote DNA strand scission (76-78, 99). Intercalative

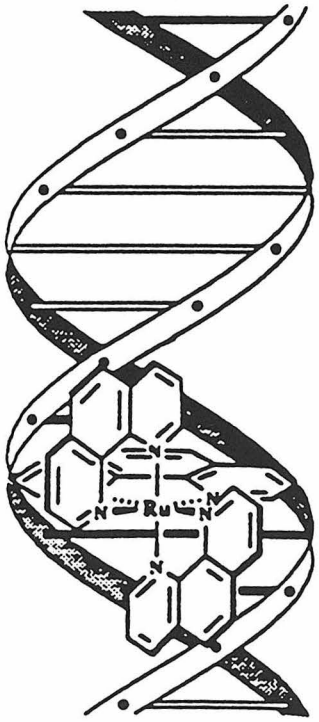
Figure 1.20 Schematic illustration explaining the enantiospecificity of $Rh(phen)_3^{2+}$. (Top) Groove binding favors the Λ-enantiomer, while (Bottom) intercalation favors the Δ-enantiomer. Enantioselectivity is controlled by steric clashes between the phenanthroline ligands and the sugar-phosphate backbone. In the surface bound form, the Λ-enantiomer mirrors the shape of the DNA. In the intercalated form, the Δ-enantiomer mirrors the shape of the groove.



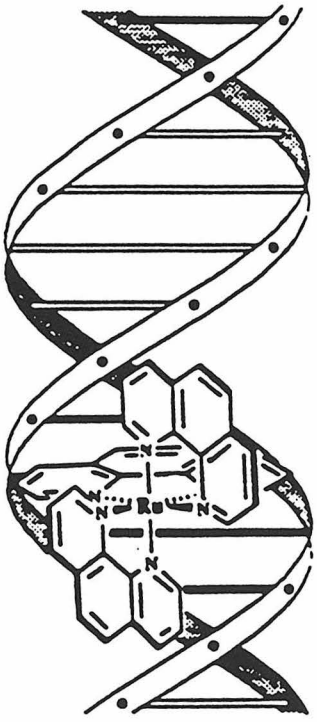
Λ



Δ

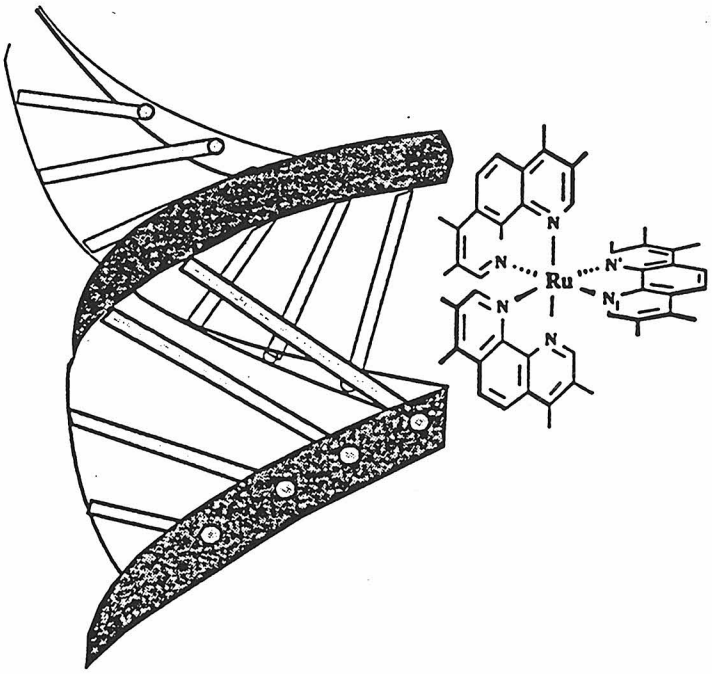


Λ



Δ

Figure 1.21 Illustration of the binding of $Rh(TMP)_3^{2+}$ to the minor groove of A-form DNA. $Rh(TMP)_3^{2+}$ prefers the wide, shallow minor groove of A-DNA (Top) to the deep, narrow minor groove of B-DNA (Bottom).


N
N
Ru
N'
N
N
N

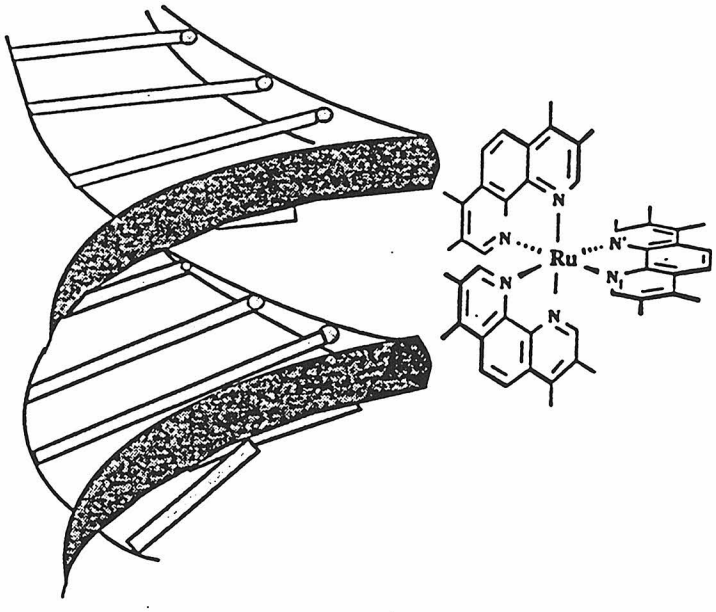
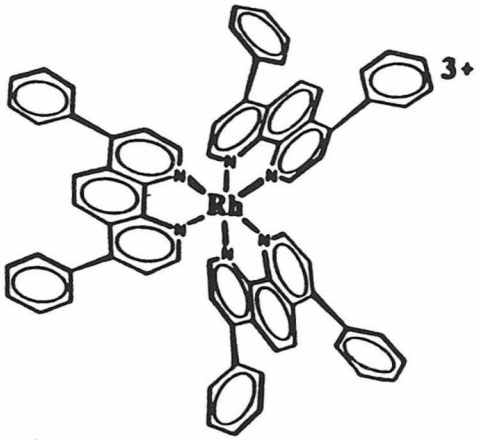

N
N
Ru
N'
N
N
N

Figure 1.22 (Top) Schematic illustration of the metal complex $Rh(DIP)_3^{3+}$. (Bottom) The metal complex $Rh(DIP)_3^{3+}$ has been shown to target sites at the beginning and end of SV40. Sites of cleavage are indicated by the symbol Λ. Adapted from ref. 110.


3+
Rh
N

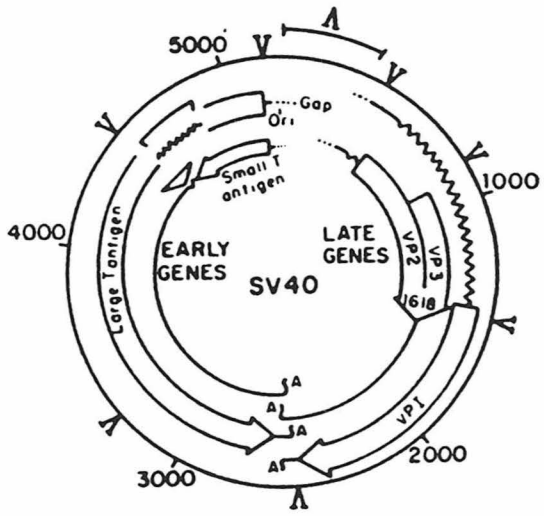
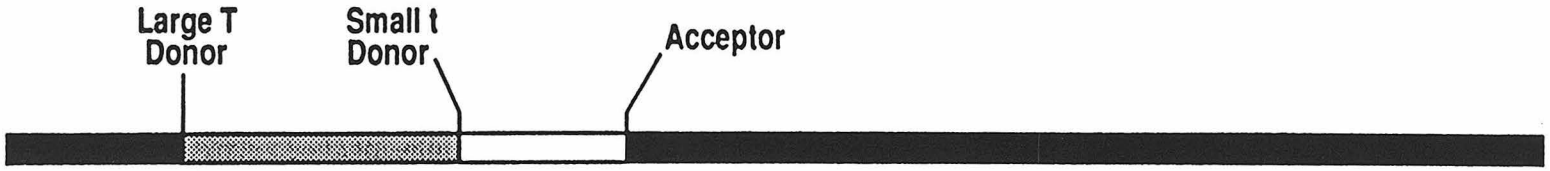
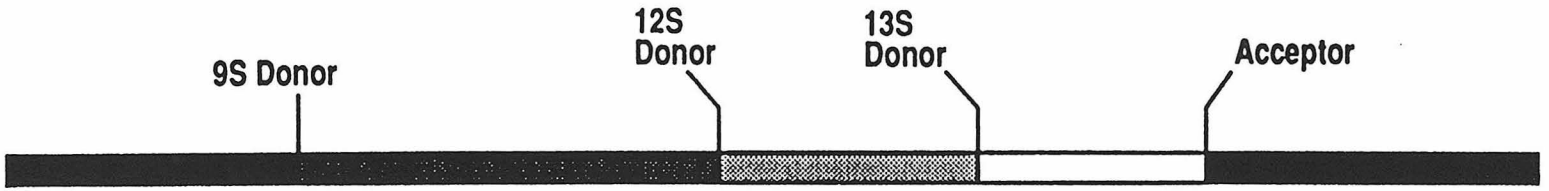

5000
1000
2000
3000
4000
Gap
Ori
Small T antigen
Large T antigen
EARLY GENES
LATE GENES
SV40
VP2
VP3
VP1
1618

Figure 1.23. Schematic illustration of cleavage by $Rh(DIP)_3^{3+}$ on Adenovirus E1A gene and Simian Virus 40 T Antigen gene. Arrows indicate points of cleavage. The two genes are represented by bars. Solid boxes represent the exons and unshaded regions represent the introns. On the SV 40 T antigen gene (Top), cleavage is observed at the acceptor (strong) and donor sites as well as the branch point. (Bottom) On the 2EIA gene, strong cleavage is only observed at the branch point. Adapted from ref. 17.

Simian Virus 40 T Antigen Transcription Unit



Adenovirus 2 E1A Transcription Unit

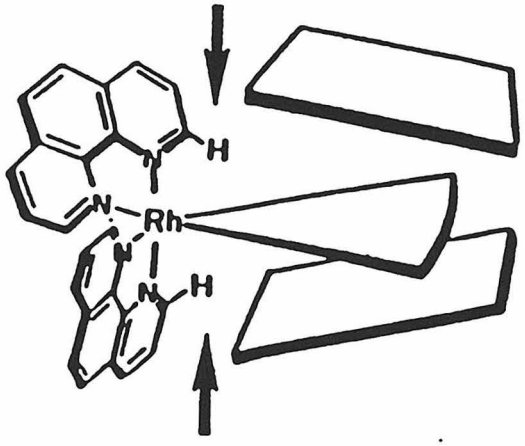


binding gives these molecules a large DNA binding affinity ($>10^6\ M^{-1}$). Importantly, phi ligand intercalation into double stranded DNA appears to be a relatively non-specific reaction, and it is the non-intercalating ancillary ligands of the complex which determine the sequence specificity of the complex (13, 77, 100, 101). Phi complexes of rhodium(III) can serve as a molecular scaffold offering access to the major groove of DNA. The intimate association that intercalation provides allows for the construction of molecules which demonstrate the principles of direct and indirect readout of the DNA helix.

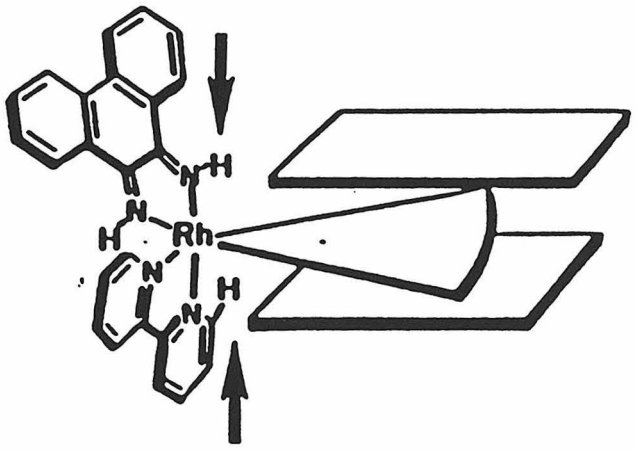
The molecules $Rh(phen)_2phi^{3+}$ and $Rh(phi)_2bpy^{3+}$ recognize DNA primarily by shape selection (77). Similar to the results with $Rh(phen)_3^{3+}$, the Δ-enantiomer of both complexes bind preferentially to the right-handed DNA helix. To a first approximation, the two complexes appear to be relatively similar and might be expected to show the same DNA sequence specificities. However, $Rh(phi)_2bpy^{3+}$ binds DNA in a relatively sequence neutral fashion and $Rh(phen)_2phi^{3+}$ demonstrates a remarkable level of sequence specificity. Closer examination of the geometry of the two molecules reveals that the ancillary ligands in $Rh(phi)_2bpy^{3+}$ are pulled back from the DNA while the 2,9-hydrogens of the phenanthroline ligands project into the major groove of the DNA (Figure 1.24). The steric constraints that are imposed by the phenanthroline ligands leads to selective targeting of DNA sequences with a open major groove, where the 2,9-hydrogens will not clash with the DNA base stack.

Shape selection in DNA recognition is truly demonstrated in the DNA recognition characteristics of $Rh(4,4'\text{-diphenylbpy})_2phi^{3+}$(102, 103). Owing to the steric constraints imposed by the phenyl substitution, this complex displays a remarkable degree of sequence specificity. An unexpected feature of complex binding is that in addition to being remarkably specific, it also models some other features which are commonly attributed to DNA binding proteins. This complex has been shown to be specific for the palindromic site 5'-CTCTAGAG-3', and thermodynamic experiments have suggested that two molecules of $Rh(4,4'\text{-diphenylbpy})_2phi^{3+}$ might bind simultaneously and cooperatively to

Figure 1.24. Illustration explaining the difference in sequence specificity between $Rh(phen)_2phi^{3+}$ and $Rh(phi)_2bpy^{3+}$. (Left) The 2,9-hydrogens of $Rh(phen)_2phi^{3+}$ project into the major groove of the DNA and can sterically clash with the base pairs. Propeller twisting can "open" the major groove to avoid steric clashes. (Right) The metal complex $Rh(phi)_2bpy^{3+}$lacks these steric clashes because the non-intercalated phi ligand is pulled back from the DNA helix. This allows $Rh(phi)_2bpy^{3+}$ to bind in a sequence neutral fashion. Adapted from ref. 77.



$[Rh(phen)_2(phi)]^{3+}$



$[Rh(phi)_2(bpy)]^{3+}$

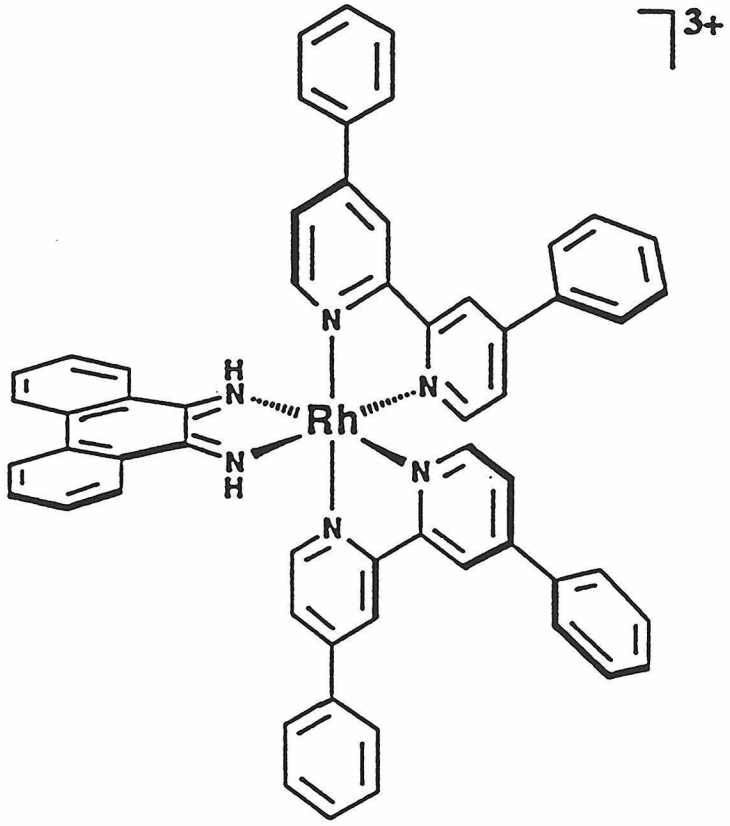
this site through non-covalent dimerization of the two complexes (Figure 1.25) (102). Thus, similar to DNA binding proteins, non-covalent interactions between two metal complexes can be used to increase the specificity of the metal complexes.

Phi complexes of rhodium(III) possessing saturated polyamine ancillary ligands have been used to demonstrate the principles of direct readout of the DNA helix. The complexes $Rh(en)_2phi^{3+}$ (en = ethylenediamine), $Rh(NH_3)_4phi^{3+}$ and $Rh([12]aneN_4)phi^{3+}$ (Figure 1.26) all have been shown to possess a 5'-GC-3' specificity which most likely is due to hydrogen bonds between the axial amines and the O6 of guanine(101) (Figure 1.27). Importantly, the molecule $Rh([12]aneS_4)phi^{3+}$ which lacks axial hydrogen bonding functionalities shows no preference for 5'-GC-3' steps. In addition to targeting 5'-GC-3' steps, the Λ-enantiomer of $Rh(en)_2phi^{3+}$ shows a preference for 5'-TX-3' steps. This additional site preference is thought to occur due to van der Waals contacts between the methylene groups of the ancillary ethylenediamine ligands and the methyl group of thymine (Figure 1.27). Importantly, replacement of thymine with uracil (which lacks the methyl group) abolishes this specificity.

Building upon the DNA recognition principles demonstrated with $Rh(en)_2phi^{3+}$, the metal complex (Δ,α)-(R,R)-$[Rh(Me_2trien)phi]^{3+}$ (Me_2trien=2,9-diamino-4,7-diazadecane) was designed to recognize the sequence 5'-TGCA-3' through a mixture of hydrogen bonding and van der Waals contacts (100). As illustrated in Figure 1.28, the methyl group of the complex should be positioned to make a methyl-methyl interaction with the thymines in the outside positions and the axial amines should be capable of hydrogen bonding with the O6 of the guanine residues. NMR studies and DNA photocleavage studies have subsequently demonstrated this model to be correct (100, 104).

Systems which combine the sequence neutrality of $Rh(phi)_2phen^{3+}$ complexes with the DNA sequence specificity of the DNA recognition helix of the P_{22} repressor protein have been explored (Figure 1.29) (105). Peptide/transition metal complex conjugates offer a unique method by which to combine the predictable sequence specificity

Figure 1.25. Schematic illustration of the binding of $Rh(DPB)_2phi^{3+}$ binding to the eight base pair sequence 5'-CTCTAGAG-3'. Cleavage is seen at the C on opposite strands (arrow). Sequence specificity is thought to be partially due to non-covalent dimerization of two metal complexes. Adapted from ref. 5.


3+
H
N
N
H
Rh
N
N
N
N

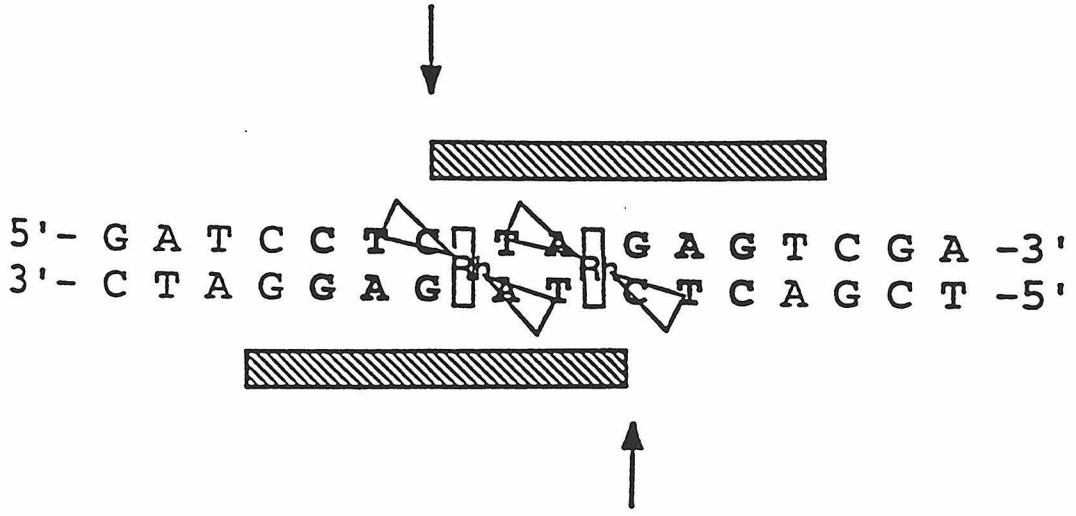
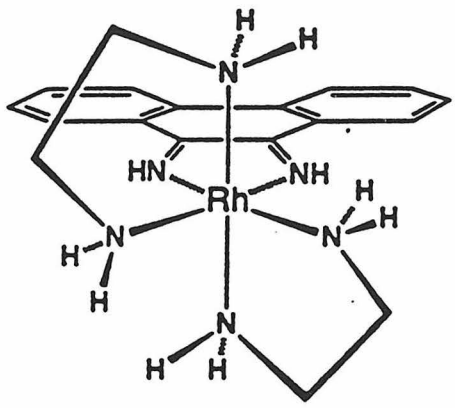
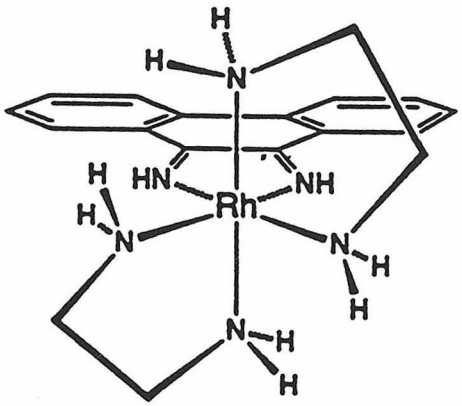
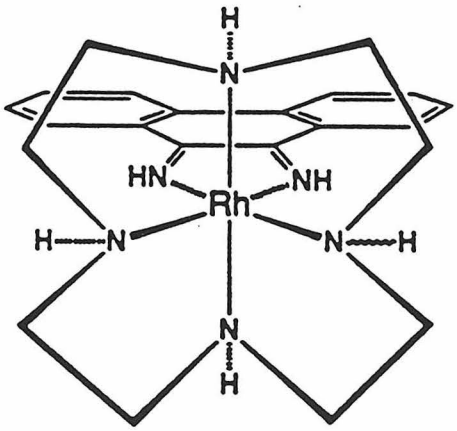

5'- G A T C C T C T A G A G T C G A -3'
3'- C T A G G A G A T C T C A G C T -5'
Rh
Rh

Figure 1.26. Schematic illustration of phi complexes of rhodium(III) which contain polyamine and polythio ancillary ligands. Complexes shown are Λ-$Rh(en)_2phi^{3+}$ (top left), Δ-$Rh(en)_2phi^{3+}$ (top right), $Rh([12]aneN_4)phi^{3+}$ (bottom left) and $Rh([12]aneS_4)phi^{3+}$ (bottom right). Note that all of the complexes except $Rh([12]aneS_4)phi^{3+}$ possess hydrogen bond donors in their axial positions. Adapted from ref. 101.


H
H
N
HN
NH
Rh
H
H
N
N
H
H
H
N
H
H


H
H
N
H
H
HN
NH
Rh
N
N
H
H
N
H
H


H
N
HN
NH
Rh
H
N
N
H
N
H

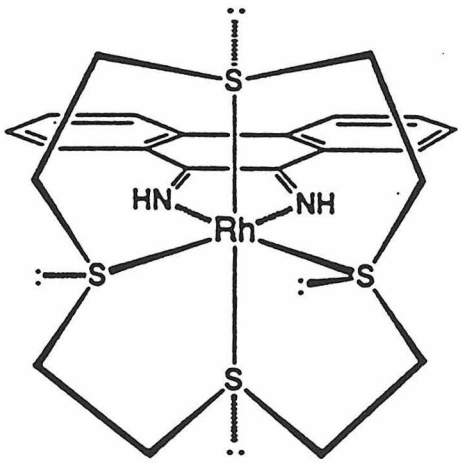
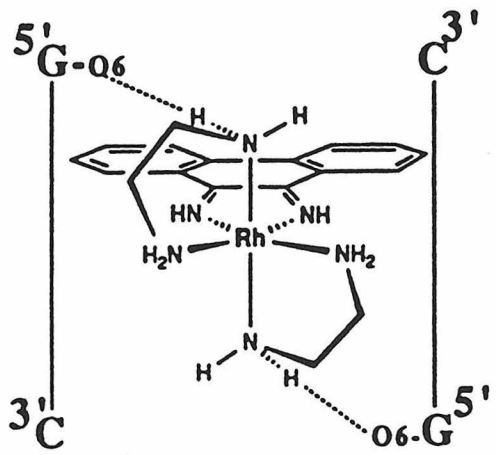
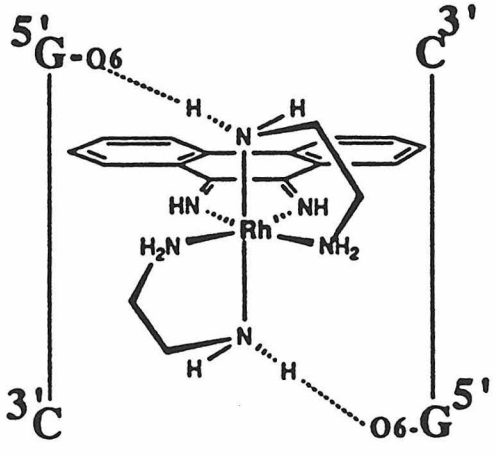

S
HN
NH
Rh
S
S
S

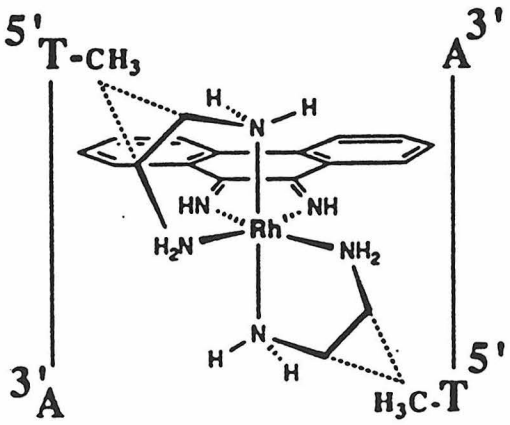
Figure 1.27. Schematic illustration explaining the sequence specificity of Δ and Λ-$Rh(en)_2phi^{3+}$. (Top) Both enantiomers show a preference for 5'-GC-3' sites due to hydrogen bonds between the axial amines and the O6 of guanine. However, Λ-$Rh(en)_2phi^{3+}$ also 5'-TX-3' sites. This additional specificity is thought to be due to van der Waals contacts between the ethylene diamine aliphatic backbone and the methyl group of thymine. The Λ-enantiomer should be able to make this van der Waals contact while the Δ-enantiomer cannot. Adapted from ref. 101.

Λ-[Rh(en)$_2$ phi]$^{3+}$



Δ-[Rh(en)$_2$ phi]$^{3+}$





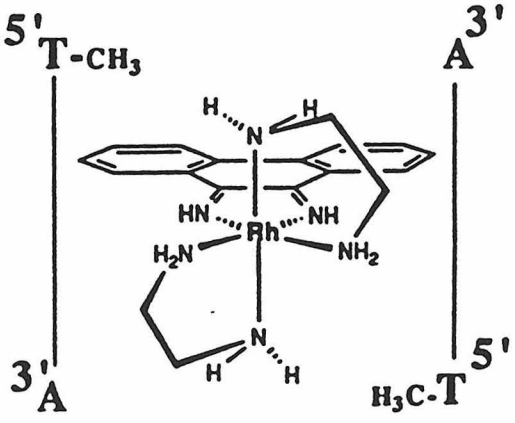


Figure 1.28. Schematic illustration showing the (left) structure of (Δ,α)-(R,R)-[Rh(Me_2trien)phi]$^{3+}$. This molecule was predicted to recognize the four base pairs sequence 5'-TGCA-3' through a combination of hydrogen bonding and van der Waals contacts (right).

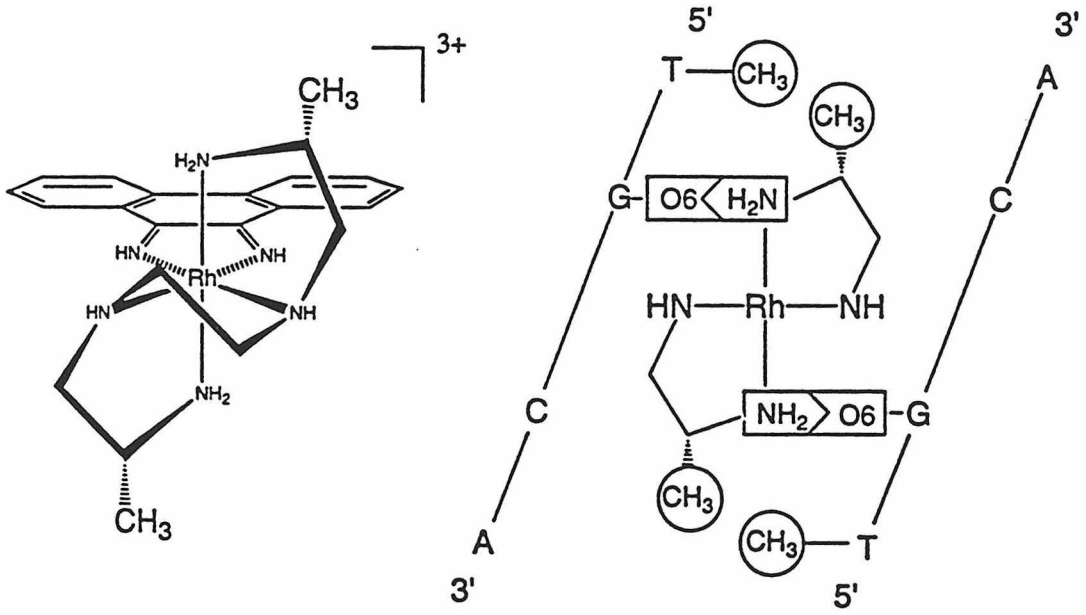
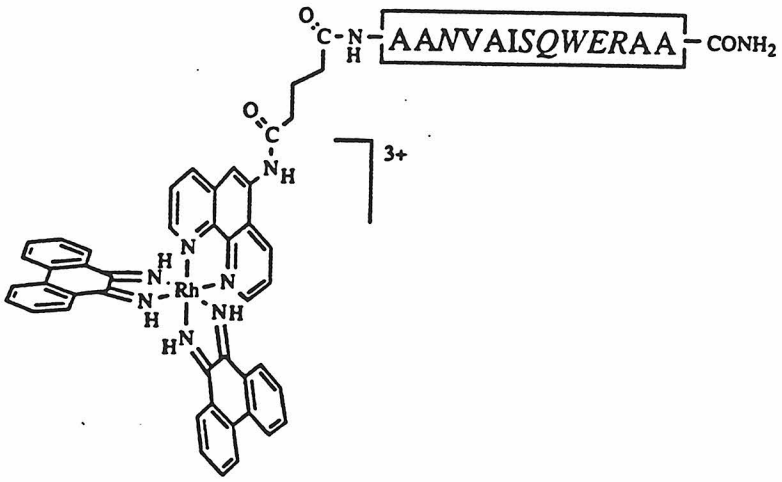
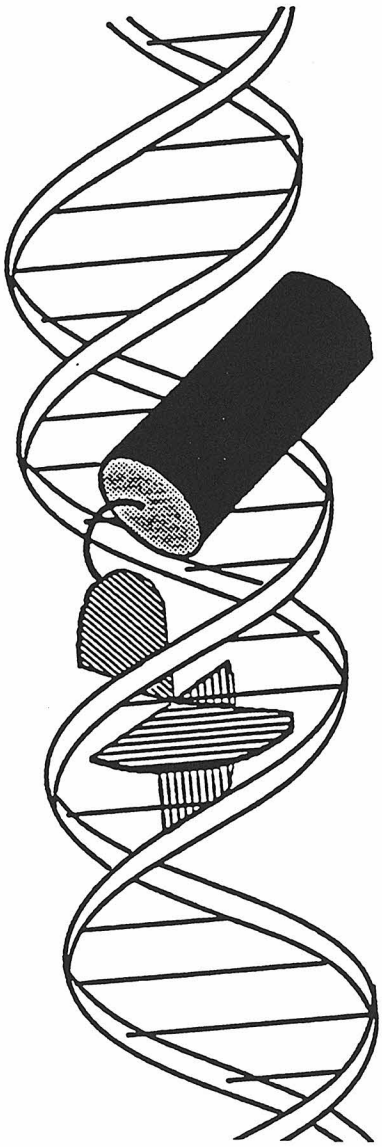

3+
CH3
H2N
HN
NH
Rh
HN
NH
NH2
CH3
5'
3'
T
CH3
CH3
A
G
O6
H2N
C
HN
Rh
NH
C
NH2
O6
G
CH3
CH3
T
A
3'
5'

Figure 1.29. (Top) Schematic illustration of the metal-peptide chimera. (Bottom) A proposed model of the metal-chimera interacting with the DNA. The metal complex intercalates into the DNA and delivers the peptide (cylinder) to the major groove. Adapted from ref. 5.


AANVAISQWERAA
CONH2
3+
Rh



of the DNA recognition elements of DNA binding proteins with the DNA binding affinity of phi complexes of rhodium(III). A lone α-helix or zinc finger module has been shown to have insufficient DNA binding affinity in the absence of the large protein. By attaching these elements onto the $Rh(phi)_2phen^{3+}$ skeleton, it might be possible to synthesize small metal-peptide chimeras capable of sequence specific binding DNA. Initial studies have shown that $Rh(phi)_2phen^{3+}$ complexes are effective at delivering peptide elements to the major groove of DNA, and the attached peptide element can be used to control the sequence specificity of these otherwise sequence neutral metal complexes. The putative operator sequence of the P_{22} repressor protein is 5'-ANTNAAG-3' and weak cleavage is observed at this sequence for the chimera complex. However, the strongest site targeted is 5'-CCA-3'. It is not known why the 5'-CCA-3' site is preferentially targeted over the operator site of P_{22}; however, it is likely that the recognition helix might not be presented to the DNA in the same orientation as it is in the native protein. Targeting of 5'-CCA-3' by the metal-peptide chimers was shown to be dependent on proper folding of the peptide a-helix, and subsequent studies demonstrated that a glutamate residue at position 10 of the peptide is an essential element of this recognition. Other peptide/metal conjugate systems which make use of single zinc fingers are currently being investigated (106).

1.9. Implications

Studies have shown that hydrophobic interactions play a major role in the binding of small molecules to DNA. This is not surprising since the majority of studies take place in aqueous solutions (conditions which should minimize the importance of hydrogen bonding) and unlike proteins, small molecules are unable to "create" their own hydrophobic environment. The majority of small molecules possess some sort of aromatic ring system (imidazole, anthracycline, phenanthrene) which either intercalates into the DNA base stack or nestles into the minor groove and forms extensive van der Waals

contacts with the sugar phosphate backbone. Similar to proteins, many small molecules bind as dimers to DNA.

Hydrogen bonding interactions usually serve to alter the specificity of small molecules but do not substantially affect the binding energetics of the complex as a whole. Owing to the limited number of potential hydrogen bond donors and acceptors which exist in the minor groove of DNA, minor groove binders tend to possess either A·T or G·C specificity, as opposed to being able to distinguish accurately between all four possible base pairs arrangements. Thus, the degree of sequence specificity which can be achieved with minor groove binders is severely limited.

Phi complexes of rhodium(III) represent a unique class of small molecules in that they bind via intercalation in the **major** groove of DNA. Similar to other small molecules, the driving force for complex binding is primarily hydrophobic in nature. This major groove preference is not understood, but other metallointercalators have also been shown to share this characteristic. The importance of shape-selection in determining the sequence specificity of phi complexes of rhodium(III) cannot be overstated, and numerous complexes which recognize DNA solely by shape selection have been reported. Additionally, a number of phi complexes of rhodium(III) with polyamine ancillary ligands, whose sequence specificity is determined by direct readout of the DNA helix, have been reported. The sequence specificity of these complexes is determined by hydrogen bonding and van der Waals contacts. While these complexes show enhanced sequence specificity, little enhancement in binding affinity has been observed. Importantly, the principles of direct read-out of the DNA sequence has made it possible to design molecules with predictive sequence specificities.

Small molecules which demonstrate the principles of direct or indirect readout of the DNA have therefore been reported. However, no clear example of a small molecule which simultaneously combines both elements of DNA recognition has been described. The sequence specificity of most DNA binding proteins relies upon the combination of

both mechanisms of DNA recognition. In this thesis we explore using the guanidinium moiety to enhance the DNA binding specificity and affinity of $Rh(phen)_2phi^{3+}$. As seen with DNA binding proteins, the guanidinium moiety is frequently used to recognize guanines residues from the major groove. By designing and synthesizing a metal complex capable of presenting the guanidinium moiety to the major groove of DNA in an orientation such that it can read the DNA base stack, we have been able to build a metal complex with unprecedented DNA binding affinity and specificity. Importantly, a major facet of this DNA recognition is the ability to couple the direct readout of the DNA sequence with a sequence dependent conformational change in the DNA.

Through the application of a new assay which we developed to test for sequence dependent unwinding of the DNA helix, we were able to demonstrate the importance of sequence dependent twistability in the DNA recognition. At the same time, it was discovered that 5'-TA-3' base steps have a tendency towards unwinding, and may be unwound transiently in solution. This structural deformation could play an important role in determining the sequence specificity of many DNA-binding proteins and might offer a possible mechanism for the initiation of strand separation by DNA polymerases. Herein we describe the synthesis and characterization of this complex and its DNA binding interactions, as well as many others which were used to verify the uniqueness of this DNA recognition.

1.10. References and Notes

1. Saenger, W., *Principles of Nucleic Acid Structure*, Springer-Verlag New York, Inc. (1984)
2. Mei, H.Y. and Barton, J.K., *J. Am. Chem. Soc.*, **108**, 7414 (1986)
3. Travers, A.A. and Klug, A., *In DNA Topology and Its Biological Effect* (eds. Cozzarelli, N.R. and Wang, J.C.), Cold Spring Harbor Laboratory Press, New York, 57-106 (1990)
4. Hertzberg, R.P. and Dervan, P.B., *Biochemistry*, **23**, 3934-3945 (1984)
5. Sardesai, N.Y., *Ph.D. Thesis*, California Institute of Technology, Pasadena, California (1995)
6. Lawley, P.D. and Shah, S.A., *Biochem. J.*, **128**, 117-132 (1972)
7. Campisi, D., Morii, T. and Barton, J.K., *Biochemistry*, **33**, 4130-4139 (1994)
8. Wu, H.M. and Crothers, D.M., *Nature*, **308**, 509-513 (1984)
9. Crothers, D. M. and Drak, J., *Meth. Enzym..*, **212,** 46-71(1992)
10. Crothers, D. M., Drak, J., Kahn, J. D., and Levine, S. D., *Meth. Enzym.*, **213,** 3-29 (1992)
11. Kahn, J. D., Yun, E. and Crothers, D. M., *Nature*, **368**, 163-166 (1994)
12. Koo, H. S., Drak, J.; Rice, J. A. and Crothers, D. M., *Biochem.*, **29**, 4227-4234 (1990)
13. Terbrueggen, R.H. and Barton, J.K., *Biochemistry*, **34**, 8227-8234 (1995)
14. Niederweis, M. and Hillen, W., *Electrophoresis* , **14**, 693-698 (1993)
15. Kirshenbaum, M.R., Tribolet, R. and Barton, J.K., *Nucl. Acids. Res.*, **16**, 7943 (1988)
16. Waldron, K., Voulgaris, T. and Barton, J.K., unpublished results
17. Lee, I. and Barton, J.K., *Biochemistry*, **32**, 6121-6127 (1993)
18. Chow, C.S. and Barton, J.K., *J. Am. Chem. Soc.*, **112**, 2839 (1990)

19. Chow, C.S., Behlen, L.S., Uhlenbeck, O. C. and Barton, J.K., *Biochemistry*, **31**, 972 (1992)

20. Lim, A.C. and Barton, J.K., unpublished results

21. Steitz, T.A., Ohlendorf, D.H., McKay, D.B., Anderson, W.F. and Matthews, B., *Proc. Nat. Acad. Sci. USA*, **79**, 3097-3100 (1982)

22. Steitz, T.A., *Q. Rev. Biophysics*, **23**, 205-280 (1990)

23. Harrison, S.C. and Aggarwal, A.K., *Ann. Rev. Biochem.*, **59**, 933-969 (1990)

24. Pabo, C.O. and Sauer, R.T., *Annu. Rev. Biochem.*, **61**, 1053-1095 (1990)

25. Johnson, P.F. and McKnight, S.L., *Ann. Rev. Biochem.*, **58**, 799 (1989)

26. Jordan, S.R. and Pabo, C.O., *Science*, **242**, 893-899 (1988)

28. Aggarwal, A.K., Rodgers, D.W., Drottar, M., Ptashne, M. and Harrison, S.C., *Science*, **242**, 893-899 (1988)

29. Schultz S.C., Shields, G.C. and Steitz, T.A., *Science*, **253**, 1001-1007 (1991)

30. Wharton, R.P., Brown, E.L. and Ptashne, M., *Cell*, **38**, 361-369 (1984)

31. Wharton, R.P. and Ptashne, M., *Nature*, **316**, 601-605 (1985)

32. Wu, L. and Koudelka, G.B., *J. Biol. Chem.*, **268**, 18975-18981 (1993)

33. Koudelka, G.B., Harrison, S.C. and Ptashne, M., *Nature*, **326**, 886-888 (1987)

34. Otwinowski, Z., Sxhevitz, R.W., Zhang, R.-G., Lawson, C.L., Joachimiak, A., Marmorstein, R.Q., Luisi, B.F. and Sigler, P.B., *Nature*, **335**, 321-329 (1988)

35. Kissinger, C.R., Liu, B., Martin-Blanco, E., Kornberg, T.B. and Pabo, C.O., *Cell*, **63**, 579-590 (1990)

36. Otting, G. et al., *EMBO J.*, **9**, 3085-3092 (1990)

37. Wolberger, C., Vershon, A. K., Liu, B., Johnson, A.D. and Pabo, C.O., *Cell*, **67**, 517-528 (1991)

38. Klemm, J.D., Rould, M.A., Aurora, R., Herr, W. and Pabo, C.O., *Cell*, **77**, 21-32 (1994)

39. Frankel, A.D., Berg, J.M. and Pabo, C.O., *Proc. Natl. Acad. Sci. USA*, **84**, 4841-4845 (1987)

40. Pavletich, N. L. and Pabo, C. O., *Science,* **252,** 809-817 (1991)

41. Parraga, G. et al., *Science*, **241**, 1489-1492 (1988)

42. Lee, M.S., Gippert, G.P., Soman, K.V., Case, D.A. and Wright, P.E., *Science*, **245**, 635-637 (1989)

43. Altaba, R.A., Perry-O'Keefe, H. and Melton, D.A., *EMBO J.*, **6**, 3065 (1987)

44. Pavletich, N.P. and Pabo, C.O., *Science*, **261**, 1701-1707 (1993)

45. Berg, J.M., *Metal-Binding Domains in Nucleic Acid-Binding and Gene-Regulatory Proteins*, 143-185

46. Dejarlais, J.R. and Berg, J.M., *Proc. Natl. Acad. Sci. USA*, **90**, 2256-2260 (1993)

47. Pomerantz, J.L., Sharp, P.A. and Pabo, C.O., *Science*, **267**, 93-96 (1995)

48. Rebar, E.J. and Pabo, C.O., *Science*, **263**, 671-673 (1994)

49. Choo, Y. and Klug, A., *PNAS*, **91**, 11163-11167 (1994)

50. Luisi, B.F., Xu, W.X., Otwinowski, Z., Freedman, L.P., Yamamoto, K.R., and Sigler, P.B., *Nature*, **352**, 497-505 (1991)

51. Evans, R.M., *Science*, **240**, 889-895 (1988)

52. Landschulz, W.H., Johnson, P.F. and McKnight, S.L., *Science*, **240**, 1759-1764 (1988)

53. Vinson, C.R., Sigler, P.B. and McKnight, S.L., *Science*, **246**, 911-916 (1989)

54. O'Shea, E.K., Klemm, J.D., Kim, P.S. and Alber, T., *Science*, **254**, 539-544 (1991)

55. Glover, J.N.M. and Harrison, S.C., *Nature*, **373**, 257-261 (1995)

56. Nakabeppu, Y. and Nathans, D., *EMBO J.,* **8**, 3833-3841 (1989)

57. Ellenberger, T.E., Brandl, C.J., Struhl, K. and Harrison, S.C., *Cell*, **71**, 1223-1237 (1992)

58. Paolella, D.N., Palmer, C.R. and Schepartz, A., *Science*, **264**, 1130-1133 (1994)

59. Konig, P. and Richmond, T.J., *J. Mol. Biol.*, **233**, 139-154 (1993)

60. Talanian, R.V., Mc Knight, C.J., and Kim, P.S., *Science*, **249**, 769-771 (1990)

61. Cuenoud, B. and Schepartz, A., *Science*, **259**, 510-513 (1993)

62. Suzuki, M., *Protein Engineering*, **8**, 1-4 (1995)

63. Breg, J.N., van Opheusden, J.H.J., Burgering, M.J.M., Boelens, R. and Kaptein, R., *Nature*, **346**, 586-589 (1990)

64. Raumann, B.E., Rould, M.A., Pabo, C.O. and Sauer, R.T., *Nature*, **367**, 754-757 (1994)

65. Burgering, M.J.M., Boelens, R., Gilbert, D.E., Berg, J.N., Knight, K.L., Sauer, R.T. and Kaptein, R., *Biochemistry*, **33**, 15036-15045 (1993)

66. Somers, W.S. and Phillips, S.E.V., *Nature*, **359**, 387-393 (1992)

67. Old, I.G. et al., *Prog. Biophys. Mol. Biol.*, **56**, 145-184 (1991)

68. Kane, S.A. and Hecht, S.M., *Nucl. Acid. Res. and Mol. Biol.*, **49**, 313-352 (1994)

69. Oppenheimer, N.J., Rodriguez, L.O. and Hecht, S.M., *Proc. Natl. Acad. Sci. USA*, **76**, 5616 (1979)

70. Kane, S.A., Natrajan, A. and Hecht, S.M., *J. Biol. Chem.*, **269**, 10899-10904 (1994)

71. Y. Iitaka, Nakamura, H., Nakantani, T., Fujii, A., Takita, T. and Umezawa, H., *J. Antibiot.*, **31**, 1070 (1978)

72. Manderville, R.A., Ellena, J.F. and Hecht, S.M., *J. Am. Chem. Soc.*, **116**, 10851-10852 (1994)

73. Stubbe, J. and Kozarich, J.W., *Chem. Rev.*, **87**, 1107-1136 (1987)

74. Sun, D., Hansen, M. and Hurley, L., *J. Am. Chem. Soc.*, **117**, 2430-2440 (1995)

75. Geierstanger, B.H. and Wemmer, D.E., *Annu. Rev. Biophys. Biomol. Struct.*, **24**, 463-493 (1995)

76. Pyle, A.M., Long, E.C. and Barton, J.K., *J. Am. Chem. Soc.*, **111**, 4520-4522 (1989)

77. Sitlani, A., Long, E.C., Pyle, A.M., and Barton, J.K., *J. Am. Chem. Soc.*, **114**, 2303-2312 (1992)

78. David, S.D.and Barton, J.K., *J. Am. Chem. Soc.*, **115**, 2984-2985 (1993)

79. Zimmer, C. and Wahnert, U., *Prog. Biophys. Mol. Biol.*, **47**, 31-112 (1986)

80. Geierstanger, B.H., Jacobsen, J.P., Mrksich, M., Dervan, P.B. and Wemmer, D.E., *Biochemistry*, **33**, 3055-3062 (1994)

81. Wade, W.S., Mrksich, M. and Dervan, P.B., *Biochemistry*, **32**, 11385-11389 (1993)

82. Mrksich, M., Wade, W.S., Dwyer, T.J., Geierstanger, B.H., Wemmer, D.E. and Dervan, P.B., *Proc. Natl. Acad. Sci. USA*, **89**, 7586-7590 (1992)

83. Mrksich, M. and Dervan, P.B., *J. Am . Chem. Soc.*, **115**, 2572-2576 (1993)

84. Wade, W.S., Mrksich, M. and Dervan, P.B., *J. Am. Chem. Soc.*, **114**, 8783-8794 (1992)

85. Mrksich, M. and Dervan, P.B., *J. Am. Chem. Soc.*, **117**, 3325-3332 (1995)

86. Geierstanger, B.H., Mrksich, M., Dervan, P.B. and Wemmer, D.E., *Science*, **266**, 646-650 (1994)

87. Loontiens, F.G., McLaughlin, L.W., Diekmann, S. and Clegg, R.M., *Biochemistry*, **30**, 182-189 (1991)

88. Gao, Y.-G., Sriram, M., Denny, W.A. and Wang, A.H.-J., *Biochemistry*, **32**, 9639-9648 (1993)

89. Wang, A.H.J., Nathans, J., van der Marel, G., van Boom, J.H. and Rich, A., *Nature*, **276**, 471-474 (1978)

90. van Houte, L.P.A., van Gargeren, C.J. and Patel, D.J., *Biochemistry*, **32**, 1667-1674 (1993)

91. Searle, M.S., Bicknell, W., Wakelin, L.P. and Denny, W.A., *Nucl. Acids Res.*, **19**, 2897-2906 (1991)

92. Sun, D., Hansen, M. and Hurley, L., *J. Am. Chem. Soc.*, **117**, 2430-2440 (1995)

93. Lin, X., Chen, H. and Patel, D.J., *J. Biomol. NMR*, **1**, 323-347 (1991)

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

95. Jenkins, Y., Friedman, A.E., Turro, N.J. and Barton, J.K., *Biochemistry*, **31**, 10809 (1992)

96. Jenkins, Y. and Barton, J.K., unpublished results

97. Barton, J.K., Science, **233**, 727-734 (1986)

98. Lee, I., Ph.D. Thesis, California Institute of Technology, Pasadena, California (1993)

99. Pyle, A.M., Long, E.C. and Barton, J.K., *J. Am. Chem. Soc.*, **111**, 4520 (1989)

100. Krotz, A. H., Hudson, B. P. and Barton, J. K. (1993) *J. Am. Chem. Soc.*, **115**, 12577-12578 (1993)

101. Krotz, A. H., Kuo, L. Y., Shields, T. P. & Barton, J. K., *J. Am. Chem. Soc.* **115**, 3877-3882 (1993)

102. Sitlani, A., Dupureur, C. M. and Barton, J. K., *J. Am. Chem. Soc.*, **115**, 12589-12590 (1993)

103. Sitlani, A. and Barton, J.K., *Biochemistry*, **33**, 12100-12108 (1994)

104. Hudson, B.H., Dupureur, C.M. and Barton, J.K., in press

105. Sardesai, N. Y., Zimmerman, K. and Barton, J. K., *J. Am Chem. Soc.*, **116**, 7502-7508 (1994)

106. Lin, S.C. and Barton, J.K., unpublished results

107 Terbrueggen, R.H. and Barton, J.K., manuscript in preparation

108. EMBO Workshop on DNA Curvature and Bending, *EMBO J.*, **8**, 1-4 (1989)

109. Dickerson, R.E. and Drew, H.R., *J. Mol. Biol.*, **149**, 761-786 (1981)

110. Muller, A.L., Raphael and Barton, J.K., *Proc. Natl. Acad. Sci. USA*, **84**, 1764-1768 (1987)

111. Bellon, S.F. and Lippard, S.J., *Biophys. Chem.*, **35**, 179-188 (1990)

Chapter 2: DNA Recognition by Rhodium Intercalators Functionalized with the Guanidinium Moiety: Δ and Λ-1-Rh(MGP)$_2$phi^{5+}

2.1. Introduction

As illustrated in chapter 1, DNA binding proteins rely on the combination of direct and indirect readout of the DNA for sequence specificity and affinity. In contrast, no clear example of a small molecule which is capable of utilizing both mechanisms simultaneously has been described. We were interested in designing molecules capable of displaying both mechanisms of DNA recognition. It might be possible to accomplish this goal by functionalizing molecules which have been demonstrated to recognize DNA by indirect readout, such as Rh(phen)$_2$phi^{3+}(1). As demonstrated by the zinc finger protein Zif268(2), the guanidinium moiety of the amino acid arginine is frequently used to target guanine bases in the major groove of DNA. The guanine specificity of arginine is attributed to the formation of a bidentate hydrogen bond between the guanidinium moiety and the N7 and O6 of guanine (Figure 2.1). The metal complex Rh(MGP)$_2$phi^{5+} (MGP = 4-guanidylmethyl-1,10-phenanthroline) was therefore designed to combine the shape selectivity of Rh(phen)$_2$phi^{3+} with the guanine-coding properties of the guanidinium moiety. [Rh(phen)$_2$phi^{3+}] possesses a DNA binding affinity of about 10^{-7} M^{-1} and targets DNA sites with an open major groove (2). NMR experiments have verified that the binding occurs by intercalation of the phi ligand into the major groove of DNA (3). It was postulated that Rh(MGP)$_2$phi^{5+} (Figure 2.2) would target the subset of Rh(phen)$_2$phi^{3+} sites which contained flanking guanines.

Figure 2.1. Illustration of the bidentate hydrogen bond formed between the guanidinium moiety and the N7 and O6 of guanine.

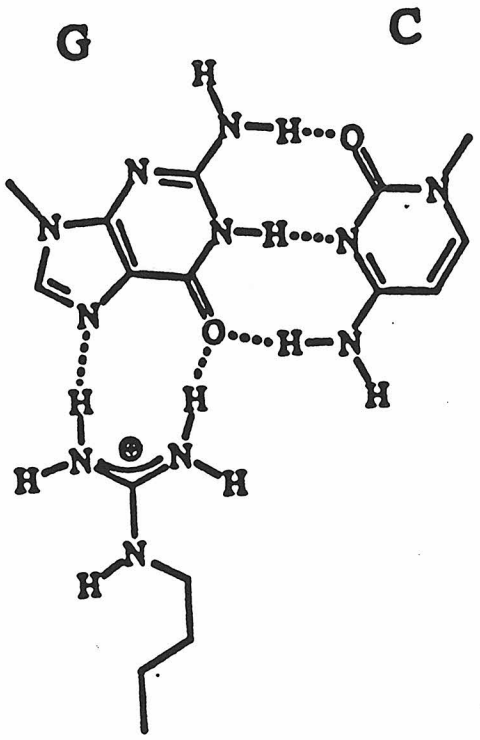
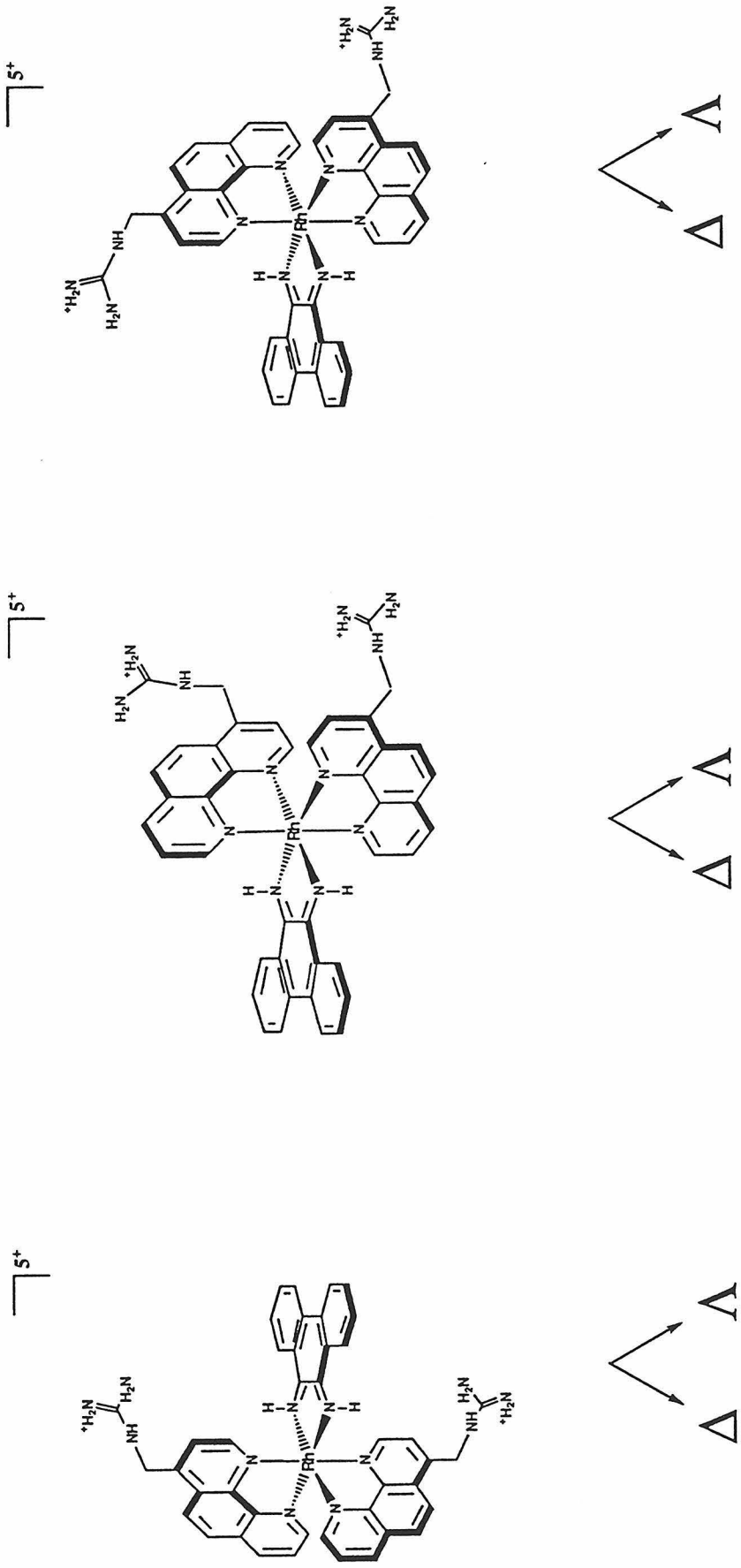

G
C

Figure 2.2. Schematic illustration of the different isomers of $Rh(MGP)_2phi^{5+}$. Isomers shown are (Left) Λ-1-$Rh(MGP)_2phi^{5+}$ (Center) Δ-2-$Rh(MGP)_2phi^{5+}$ (Right) Δ-3-$Rh(MGP)_2phi^{5+}$.


5+
Rh
Λ
Δ
5+
Rh
Λ
Δ
5+
Rh
Λ
Δ

2.2. Experimental

2.2.1. Materials. The synthesis and characterization of $Rh(MGP)_2phi^{5+}$ is described in chapter 5 of this thesis. [α-^{32}P]dATP and [γ-^{32}P]-ATP were obtained from NEN-Dupont. A 330 base pair 3'-α-^{32}P-labeled AccI/DrdI fragment of pBR322, a 140 base pair and 180 base pair 5'-γ-^{32}P- or 3'-α-^{32}P-labeled EcoRI/PvuII fragment of pUC-18, and a 636 base pair and 190 base pair 5'-γ-^{32}P- or 3'-α-^{32}P-labeled EcoRI/PvuII/EcoRV fragment of pBR322 where prepared by standard methods (4). Oligonucleotides were synthesized via the phosphoramidite method (5) on an ABI 391 DNA-RNA synthesizer and purified by reverse-phase HPLC. Phosphoramidites, solid supports, and reagents were purchased from ABI and Glen Research. All restriction enzymes were purchased from New England Biolabs and Boehringer Mannheim. The concentrations of all rhodium solutions were determined by UV-VIS spectroscopy using ε = 19,400 $M^{-1}cm^{-1}$ at 360 nm. DNA concentrations were determined using ε = 6600 $M^{-1}cm^{-1}$ at 260 nm.

2.2.2. Instrumentation. All photocleavage experiments were performed using an Oriel Model 6140 1000 W Hg/Xe lamp fitted with a monochromator and a 300 nm cut-off filter. Quantitation of the metal complexes was performed on a Cary 219 (Varian) spectrophotometer. High performance liquid chromatography (HPLC) was carried out on a Waters 600E system equipped with a Waters 484 tunable detector. A Vydac reverse phase protein and peptide C18 column was used for HPLC separations. Gel electrophoresis experiments were quantified using a Molecular Dynamics phosphorimager and ImageQuant software.

2.2.3. Photocleavage Reactions on 5' and 3'-^{32}P Restriction Fragments. The ^{32}P end-labeled restriction fragment and rhodium complexes were incubated together in a 1.7 mL siliconized Eppendorf tube at 23 °C for 5 minutes prior to

irradiation with 313 nm light for 8 minutes. The DNA/Rh ratio was maintained at 50:1 by addition of unlabeled calf thymus DNA (Pharmacia) in 20 μL irradiation reactions (10 mM sodium cacodylate, 40 mM NaCl, pH 7.0). Samples were frozen on dry ice immediately after irradiation, and lyophilized to dryness. The pellet was resuspended in 15 μL on denaturing gel loading dye and 3 μL was loaded onto an 8% denaturing polyacrylamide gel and electrophoresed at 90 watts (@2000 V) for one and a half hours. The gel was transferred to paper and dried prior to being analyzed using a phosphorimager (Molecular Dynamics).

2.2.4. Photocleavage Reactions on 5'-^{32}P-Endlabeled Oligonucleotides. The oligonucleotides, 5'-ACTGACTGCCTAGGTATAAACTTAAGCATAAACA TATGCGTACGTATATACGTAGCCACGTGGGTACCACTGACTG-3' , and its complement were synthesized for use in determining the consensus sequence for Λ-1-Rh(MGP)$_2$phi^{5+}. A series of oligonucleotides, (**A1/A2**) 5'-GTAGGTGAACCTCGGACGATCTCAACATCTGTTAATCTTGGTTTCTA AGGTTCTTCCGCTTCAGAAGTTCACAGGTGA-3', (**B1/B2**) 5'-CTGTAGGT GGTCGACAATCGATGCCCGGGACACGTGACTTCTGACTTCAGACTGC AGACACGTGACATCTGAGGTCAC-3', (**C1/C2**) 5'-TAGTTCGGCGTCGAC GCTCGAGACTTCAGAGTCTACTCTCTAGTGTGCAGAATCTATGTTCTA ACCGGAATC-3', (**D1/D2**) 5'-ACTGGCCTCATCTGCTCTTCAGCAGCTGCT GCAGGTCTAGGGCAT-3', and their complements were synthesized for use in determining the sequence specificity of Δ-1-Rh(MGP)$_2$phi^{5+}. The oligonucleotides (both strands) were 5'-^{32}P labeled by standard protocols (4). Each oligonucleotide strand was annealed with a slight excess of its cold complement by incubation at 90 °C for 4 minutes followed by cooling to 23 °C over a 1 h period. The double stranded oligonucleotides and the different rhodium complexes were incubated together at 23 °C for 5 minutes prior to irradiation with 313 nm light for 10

minutes. The DNA/Rh ratio was 100:1 with 20 μL irradiation reactions (10 mM sodium cacodylate, 40 mM NaCl, pH 7). Samples were frozen on dry ice immediately after irradiation, and lyophilized to dryness. The pellet was resuspended in 15 μL on denaturing gel loading dye and 3 μL was loaded onto an 8% denaturing polyacrylamide gel and electrophoresed at 90 watts (@2000 V) for one and a half hours. The gel was transferred to paper and dried prior to being analyzed using a phosphorimager (Molecular Dynamics).

2.2.5. Substitution of 7-Deazaguanine. Photocleavage experiments on 5'-^{32}P-end labeled 32 base pair long oligonucleotides were carried out. These oligonucleotides were designed to contain a central complex binding site in which the effect of incorporating 0, 1, or 2 deazaguanines could be studied. Control sites were placed both 5' and 3' to the modified site. Four different oligonucleotides were synthesized; 5'-TCAGCATATACCCATATG AGCATATGTCCTGA-3' (**A**), 5'-TCAGGACATATGCTCATATGGGTATATGC TGA-3' (**B**), 5'-TCAGCATAT ACCCATAT*G*AGCATATGTCCTGA-3' (**C**), and 5'-TCAGGACATATGCTCA TAT*G*GGTATATGCTGA-3' (**D**) where *G* represents the unnatural oligonucleotide 7-deazaguanine. The oligonucleotides was 5'-^{32}P-labeled by standard protocols. Each radio-labeled strand was annealed with its cold complement by incubating at 90°C for 4 minutes followed by cooling to 23°C over a 1 hour period. Photocleavage experiments on double stranded oligonucleotides **AB**, **AD**, **BC**, and **CD** were carried out. The double stranded oligonucleotides and the different rhodium complexes were incubated together at 23°C for 5 minutes prior to irradiation with 313 nm light for 8 minutes. DNA/Rh ratio 50:1, 20 μL irradiation reactions (10 mM Na cacodylate, 40 mM NaCl, pH 7.0). Samples were frozen on dry ice immediately following irradiation, dried, and resuspended in denaturing gel loading dye (10,000 cpm/μL). After separation on a 20%-

denaturing polyacrylamide gel, the samples were analyzed using a phosphorimager. Photocleavage intensity was taken to be the total cleavage seen at both the central T and A of the modified site and the control sites. Overall cleavage intensity was adjusted to take into account unequal lane loading. Cleavage on both strands of the DNA duplex was quantitated and the cleavage at the control sites varied by less than 10% on the different oligonucleotides.

2.2.6. Qualitative Determination of Exchange Rates for $Rh(MGP)_2phi^{5+}$. A series of solutions containing the 3'-^{32}P labeled 330 base pair restriction fragment from pBR322 and 5 mM calf thymus DNA were prepared (10 mM sodium cacodylate, 40 mM NaCl, pH 7). Rhodium complex was added to give a final concentration of 0.1 μM (20 μL sample). The samples were incubated at 23 °C for 5 minutes. A zero point sample was irradiated at 313 nm for 4 minutes. To all of the other solutions was added calf thymus DNA to give a final DNA concentration of 100 uM (DNA/Rh ratio 1000:1). Samples were incubated at 23 °C for 5 minutes, 30 minutes, 1 hour, 3 hours, and 24 hours prior to irradiation. The samples were irradiated for 4 minutes at 313 nm and frozen on dry ice immediately after irradiation. The pellet was resuspended in 15 μL on denaturing gel loading dye and 3 μL was loaded onto an 8% denaturing polyacrylamide gel and electrophoresed at 90 watts (@2000 V) for one and a half hours. The gel was transferred to paper and dried prior to being analyzed using a phosphorimager (Molecular Dynamics).

2.3. Results and Discussion

2.3.1 DNA Photocleavage by $Rh(MGP)_2phi^{5+}$ Isomers. Since 4-guanidylmethyl-1,10-phenanthroline is an asymmetric ligand, synthesis of $Rh(MGP)_2phi^{5+}$ generates three different positional isomers in a statistical

distribution of 1:2:1. Each isomer is in turn an equal mixture of 2 enantiomers (Figure 2.2). Photocleavage by the isomers of $Rh(MGP)_2phi^{5+}$ is shown in Figure 2.3. All isomers promote DNA strand scission upon irradiation.

The recognition pattern of $Rh(MGP)_2phi^{5+}$ complexes is seen to be both isomer-specific and enantiospecific. Cleavage with racemic 2-$Rh(MGP)_2phi^{5+}$ (lane 4 and 5), the isomer in which the guanidinium groups are directed away from the phi ligand, reveals a pattern which is identical to but more intense compared to the cleavage pattern of *rac*-$Rh(phen)_2phi^{3+}$ (lane 6 and 7); this result is expected given the positioning of the guanidinium moieties and the increase in charge. In contrast, racemic 1-$Rh(MGP)_2phi^{5+}$ (lane 2 and 3), the isomer in which both guanidinium groups are pointed over the phi ligand, or when intercalated, towards the DNA, shows a specific pattern of photocleavage which is strong at lower concentrations. Photocleavage with the two enantiomers of 1-$Rh(MGP)_2phi^{5+}$, Λ-1-$Rh(MGP)_2phi^{5+}$ and Δ-1-$Rh(MGP)_2phi^{5+}$, reveals that the majority of these enhanced sites can be attributed to Λ-1-$Rh(MGP)_2phi^{5+}$ (lane 10 and 11). However, there is one site of enhanced cleavage which can be attributed to recognition by Δ-1-$Rh(MGP)_2phi^{5+}$ (lane 8 and 9).

2.3.2. DNA Cleavage Characteristics of Λ-1-$Rh(MGP)_2phi^{5+}$. All of the sites cleaved by Λ-1-$Rh(MGP)_2phi^{5+}$ (Figure 2.3) contain a central 5'-TA-3' base step, and sequence discrimination is apparent over 6 base pairs. In contrast to both $Rh(phen)_2phi^{3+}$ and Δ-1-$Rh(MGP)_2phi^{5+}$, where a single cleavage site is found at the site of the intercalation, Λ-1-$Rh(MGP)_2phi^{5+}$ shows cleavage at both nucleotides of the intercalation site (6). Titration experiments reveal that Λ-1-$Rh(MGP)_2phi^{5+}$ targets 5'-CATATG-3', the strongest site seen in Figure 2.3, at concentrations as low as 250 picomolar, more than two orders of magnitude lower in concentration than observed with $Rh(phen)_2phi^{3+}$ (Figure 2.4).

Figure 2.3. Comparison of photocleavage patterns for $Rh(phen)_2phi^{3+}$, 1-$Rh(MGP)_2phi^{5+}$, and 2-$Rh(MGP)_2phi^{5+}$ on a P^{32}-3'-endlabeled 330 base pair restriction fragment (pBR322 cut by Acc I/Drd I). Lane 1: Maxam-Gilbert A+G reaction; lanes 2 and 3: *rac*-1-$Rh(MGP)_2phi^{5+}$ (50 nM and 5 nM); lanes 4 and 5: *rac*-2-$Rh(MGP)_2phi^{5+}$ (50 nM and 5 nM), lanes 6 and 7: *rac*-$Rh(phen)_2phi^{3+}$ (50 nM and 5 nM); lanes 8 and 9: Δ-1-$Rh(MGP)_2phi^{5+}$ (50 nM and 5 nM); lanes 10 and 11: Λ-1-$Rh(MGP)_2phi^{5+}$ (50 nM and 5 nM); lane 12: Λ-1-$Rh(MGP)_2phi^{5+}$ (50 nM) without irradiation; lane 13: DNA without metal and without irradiation.

Sites indicated are the strongest sites of photocleavage for Δ-1-$Rh(MGP)_2phi^{5+}$ (5'-CATCTG-3') and Λ-1-$Rh(MGP)_2phi^{5+}$ (5'-CATATG-3'). Note the double cleavage band for Λ-1-$Rh(MGP)_2phi^{5+}$.

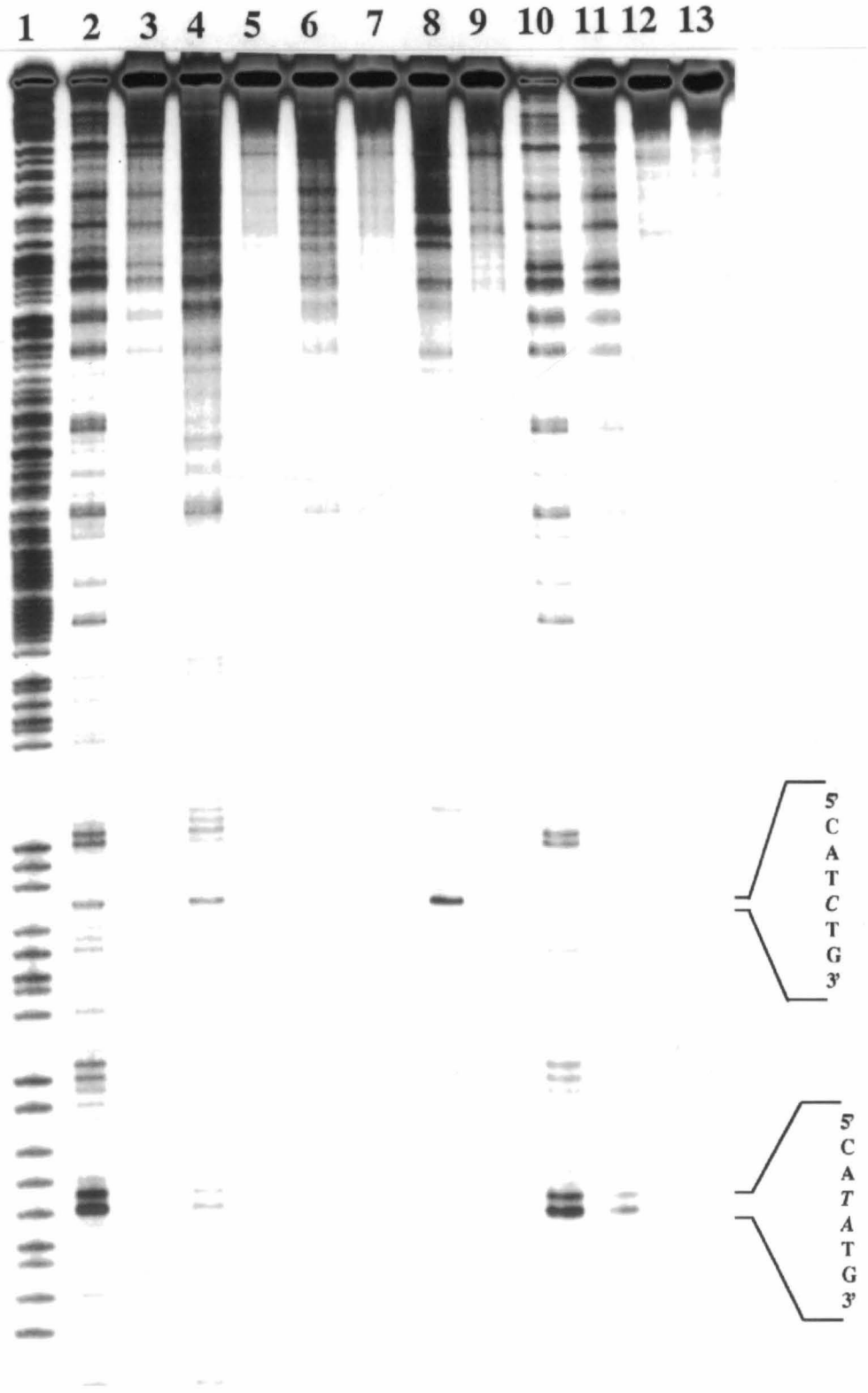
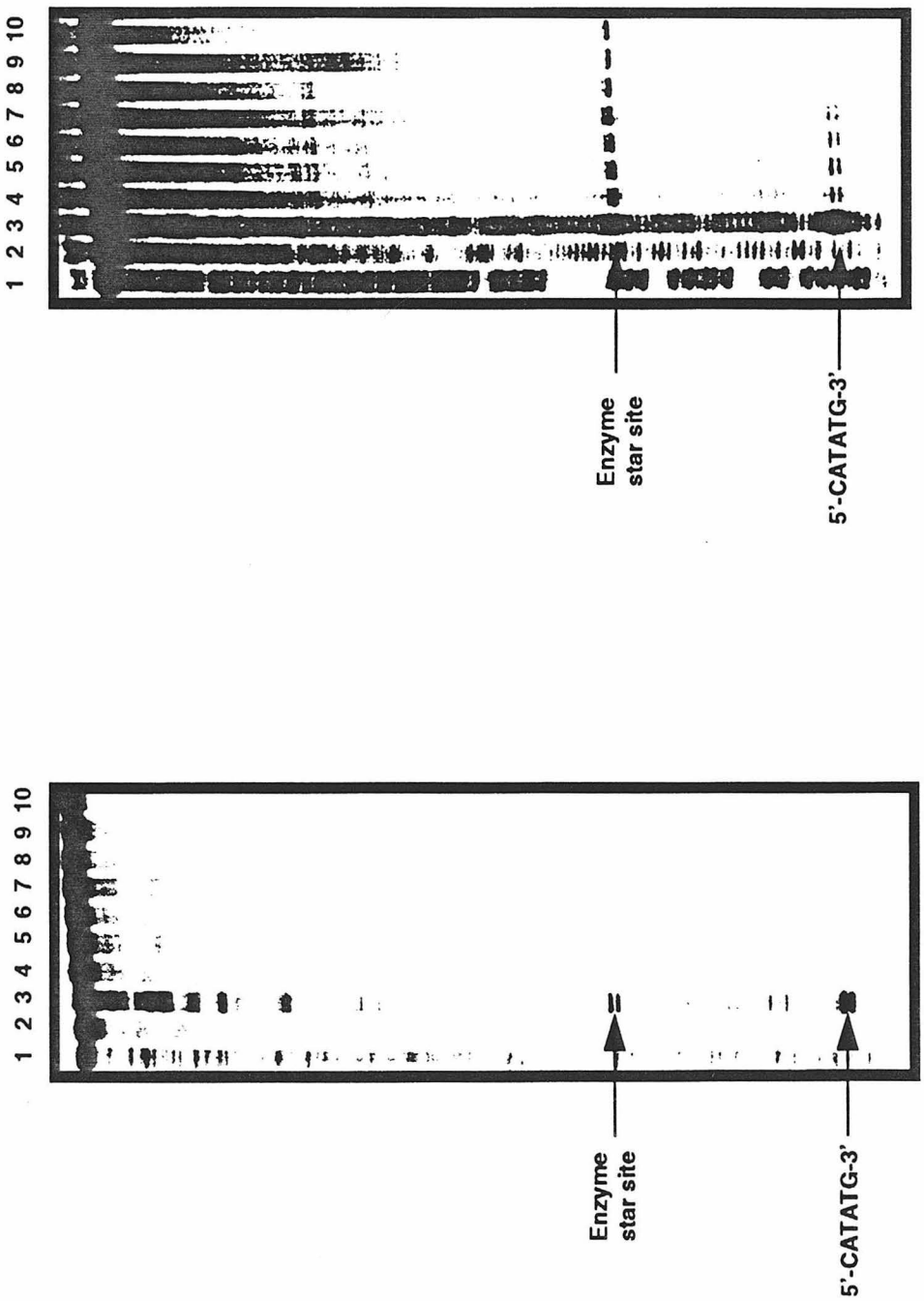

1 2 3 4 5 6 7 8 9 10 11 12 13
5' C A T C T G 3'
5' C A T A T G 3'

Figure 2.4. Photocleavage titration of Λ-1-Rh(MGP)$_2$phi^{5+} on a P^{32}-3'-endlabeled 330 base pair restriction fragment (pBR322 cut by Acc I/Drd I). Photos on the left and the right side of the page are identical, except for the exposure level. Lane 1: Maxam-Gilbert A+G reaction; Lane 2: Maxam-Gilbert C+T reaction; Lanes 3 through 8: Photocleavage with decreasing amounts of Λ-1-Rh(MGP)$_2$phi^{5+}. No carrier DNA (DNA concentration of about 200 nM). Lane 3: 10 nM Λ-1-Rh(MGP)$_2$phi^{5+}; Lane 4: 1 nM Λ-1-Rh(MGP)$_2$phi^{5+}; Lane 5: 750 pM Λ-1-Rh(MGP)$_2$phi^{5+}; Lane 6: 500 pM Λ-1-Rh(MGP)$_2$phi^{5+}; Lane 7: 250 pM Λ-1-Rh(MGP)$_2$phi^{5+}; Lane 8: 100 pM Λ-1-Rh(MGP)$_2$phi^{5+}; Lane 9: 10 nM Λ-1-Rh(MGP)$_2$phi^{5+}-no irradiation; Lane 10: DNA only w/ irradiation.

Only two cleavage sites are apparent on the restriction fragment. The lower site is 5'-CATATG-3' and the upper site (star site) is a secondary site for the restriction enzyme. Cleavage by Λ-1-Rh(MGP)$_2$phi^{5+} is apparent down to 250 pM.


1 2 3 4 5 6 7 8 9 10
Enzyme star site
5'-CATATG-3'
1 2 3 4 5 6 7 8 9 10
Enzyme star site
5'-CATATG-3'

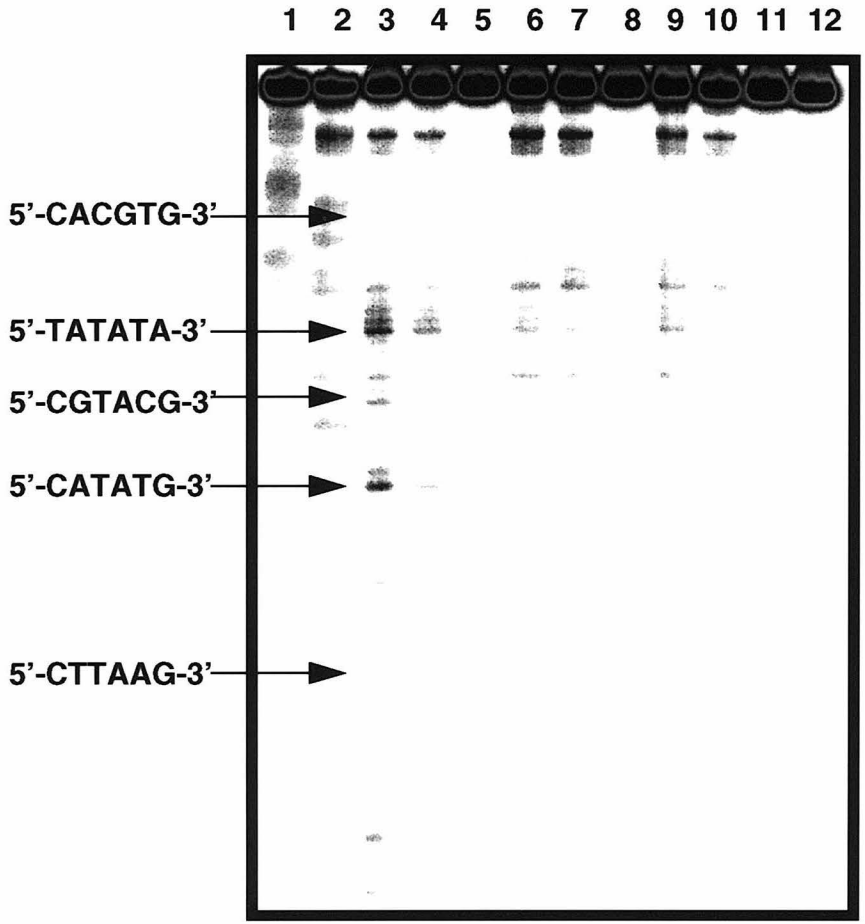
2.3.3 Determination of the Consensus Sequence for Λ-1-Rh(MGP)$_2$phi^{5+}. The consensus sequence for cleavage by Λ-1-Rh(MGP)$_2$phi^{5+} was determined using a 76 base pair DNA duplex which was constructed to contain 14 different sequence variations of 6 bp sites which contain a central 5'-TA-3' base step (Figure 2.5). As on the larger restriction fragment, cleavage by Rh(MGP)$_2$phi^{5+} is enantiospecific and isomer-specific. Closer examination reveals the hierarchy for sites recognized by Λ-1-Rh(MGP)$_2$phi^{5+}. As evident in Figure 2.5, there is a stringent requirement for a central 5'-TA-3' step; changing the central step to a 5'-CG-3' abolishes cleavage. Sites containing a C at position 1 and G at position 6 are found to be preferred but not essential. For example, 5'-TATATA-3' is cleaved with 62% of the intensity of 5'-CATATG-3'. An alternation of pyrimidines and purines is clearly preferred; cleavage at 5'-CTTAAG-3' shows 16% of the cleavage intensity compared to 5'-CATATG-3'. Overall, at nanomolar and higher concentrations, Λ-1-Rh(MGP)$_2$phi^{5+} cleaves preferentially on the consensus site 5'-CATATG-3', with cleavage at both the internal T and A.

2.3.4. Effects of Deazaguanine Substitution. The differences in recognition between Λ-1-Rh(MGP)$_2$phi^{5+} and Λ-2-Rh(MGP)$_2$phi^{5+} indicate a participation of the guanidinium groups in the recognition, but to establish where they contact on the DNA helix, we directly compared cleavage of 32 bp oligonucleotides containing the native recognition site and a site with deazaguanine substitutions. In mimicking Zif268 (2), one would expect that the guanidinium moiety would hydrogen bond with the N7 and O6 atoms of guanine; incorporation of 7-deazaguanine would eliminate the possibility of hydrogen bonding to the N7 position (7).

Figure 2.6 summarizes the results of making single and double base substitutions. Incorporation of a single deazaguanine in the palindromic sequence 5'-CATATG-3' results in either 22% or 39% reduction in cleavage, depending upon

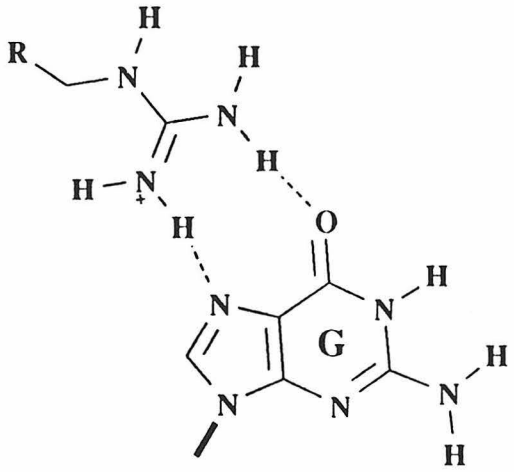
Figure 2.5. (Left) Determination of the consensus sequence targeted by Λ-1-$Rh(MGP)_2phi^{5+}$ on a 76-mer oligonucleotide containing 14 different 6 bp sequence variations with a central 5'-TA-3' base step. Lanes 1 and 2: A+G and C+T sequencing reactions (Maxam-Gilbert); lanes 3, 4, and 5: Λ-1-$Rh(MGP)_2phi^{5+}$(1.0, 0.1, 0.01 μM); lanes 6, 7, and 8: Λ-2-$Rh(MGP)_2phi^{5+}$ (1.0, 0.1, and 0.01 μM); lanes 9, 10, and 11: *rac*-$Rh(phen)_2phi^{3+}$(1.0, 0.1, and 0.01 μM); lane 12: DNA without irradiation and without rhodium.

The Table provides a comparison for several representative sites. In particular the rigorous requirement for a central 5'-TA-3' is evident. The importance of alternation in pyrimidines and purines as well as the preference for a cytosine in position 1 and a guanine in position 6 are illustrated.

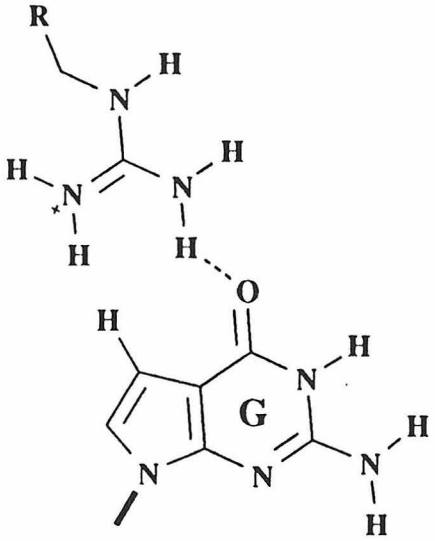


Site	Relative Cleavage %
5'-CATATG-3'	100
5'-CACGTG-3'	0
5'-CGTACG-3'	73
5'-TATATA-3'	62
5'-CTTAAG-3'	16

Figure 2.6. Probing hydrogen bonding interactions using 7-deazaguanine substitutions. Above are illustrated the hydrogen bonding interactions between guanine and the guanidinium moiety. The guanidinium functionality of Λ-1-$Rh(MGP)_2phi^{5+}$ may form a single hydrogen bond with the O6 oxygen atom of 7-deazaguanine while it can form a bidentate hydrogen bond with both the N7 nitrogen and O6 oxygen atoms on guanine. (Below) Illustrated schematically are the oligonucleotides that were synthesized to probe for hydrogen bonding interactions and a summary of the results. The effect of both single and double deazaguanine (G) substitutions were investigated. Note that 5'-CATATA-3' is cleaved with an intensity close to 5'-CATATG-3' and greater than 5'-CATATG-3'. Thus one guanidinium may maintain an N7 and O6 contact while the other may contact primarily the N7.



Guanine



Deaza-Guanine
G*

	Relative cleavage intensity
5'-TCGA **CATATA** CC **CATATG*** AG **CATATG** TCCTGA-3' 3'-AGCT **GTATAT** GG***GTATAC** TC **GTATAC** AGGACT-5'	40.1
5'-TCGA **CATATA** CC **CATATG*** AG **CATATG** TCCTGA-3' 3'-AGCT **GTATAT** GG **GTATAC** TC **GTATAC** AGGACT-5'	67.9
5'-TCGA **CATATA** CC **CATATG** AG **CATATG** TCCTGA-3' 3'-AGCT **GTATAT** GG **GTATAC** TC **GTATAC** AGGACT-5'	100

which guanine is substituted. Despite the palindromic nature of the binding site for Λ-1-Rh(MGP)$_2$phi^{5+}, there is significant variation between the amount of reduction in cleavage for the two different single deazaguanine substitutions. The sequences flanking the binding site must therefore play some role. When deazaguanines are incorporated on both strands, the overall cleavage intensity is reduced by 60%. This experiment indicates that contact with the N7 position of guanine directly stabilizes the bound complex. Furthermore, these results verify that Λ-1-Rh(MGP)$_2$phi^{5+} binds in the DNA major groove.

2.3.5. DNA Binding Characteristics of Δ-1-Rh(MGP)$_2$phi^{5+}. Δ-1-Rh(MGP)$_2$phi^{5+} shows enhanced cleavage relative to Rh(phen)$_2$phi^{3+} at 5'-CATCTG-3' with cleavage at the italicized C (Figure 2.3). Photocleavage results on other 5' and 3'-labeled restriction fragments reveal a much more complicated pattern for recognition by Δ-1-Rh(MGP)$_2$phi^{5+} than merely Rh(phen)$_2$phi^{3+} sites with flanking G·C base pairs. Δ-1-Rh(MGP)$_2$phi^{5+} intercalates into a central 5'-T*C*-3' step and cleaves at the 3'C. Surprisingly, photocleavage results on the opposite strand show light cleavage at the G of the 5'-A*G*-3' sequence, or no cleavage at all. Previous results have shown that photocleavage asymmetries are indicative of the orientation of the metal complex in the intercalation site (7,8). Based upon the cleavage asymmetry seen here, it would appear that Δ-1-Rh(MGP)$_2$phi^{5+} is canted in the intercalation site. Photocleavage titration experiments with Δ-1-Rh(MGP)$_2$phi^{5+} reveal a loss of cleavage between 1 and 5 nanomolar, which represents about a 10-fold increase in binding affinity as compared to Rh(phen)$_2$phi^{3+}.

2.3.6. Determination of the Consensus Sequence for Δ-1-Rh(MGP)$_2$phi^{5+}. A series of oligonucleotides were designed and synthesized to evaluate how DNA

sequence variations affect the cleavage intensity of Δ-1-Rh(MGP)$_2$phi^{5+}. Initial oligonucleotide sites were designed with a central 5'-TC-3' step and flanking base pairs were varied symmetrically due to the C_2 symmetry of Δ-1-Rh(MGP)$_2$phi^{5+}. Photocleavage results show the consensus site to be 5'-CATCTG-3' (Figure 2.7).

As illustrated in Table 2.1, changes in the consensus sequence of both Λ-1-Rh(MGP)$_2$phi^{5+} and Δ-1-Rh(MGP)$_2$phi^{5+} have very similar effects. For example, changing the outside C·G base pairs to A·T results in approximately a 25% decrease in photocleavage for both enantiomers. Importantly, Λ-1-Rh(MGP)$_2$phi^{5+} requires that the internal 4 base pairs of the recognition sequence be alternating 5'-purine- pyrimidine-purine-pyrimidine-3' for photocleavage to be observed. Comparatively, however, Δ-1-Rh(MGP)$_2$phi^{5+} does not as rigorously control the internal portion of the recognition sequence. This is evident by comparing the photocleavage intensity observed for the two complexes at the sites 5'-CTXXAG-3'. Λ-1-Rh(MGP)$_2$phi^{5+} shows an 84% reduction in photocleavage intensity, while Δ-1-Rh(MGP)$_2$phi^{5+} only shows a 52% reduction in photocleavage. This is further demonstrated by the effect seen when the central two base pairs of the intercalation site are changed. Changing the central base pairs of the consensus site for Λ-1-Rh(MGP)$_2$phi^{5+} completely abolishes photocleavage and only results in a 90% decrease for Δ-1-Rh(MGP)$_2$phi^{5+}. It should be noted that completely abolishing cleavage is a much stronger affect than a 90% reduction in cleavage intensity.

It is interesting that the external sequences of the consensus sites for Δ and Λ-Rh(MGP)$_2$phi^{5+} are found to be identical and it is the internal sequences that vary. Comparatively, the external portions of the two enantiomers are identical and it is the internal cores of the molecules that vary. This suggests that the guanidinium moieties of the Δ-1-Rh(MGP)$_2$phi^{5+} and Λ-1-Rh(MGP)$_2$phi^{5+} might be interacting with their DNA sites in a similar manner. This result furthermore

Figure 2.7. Determination of the consensus sequence for Δ-1-Rh(MGP)$_2$phi^{5+} on a 77 bp DNA duplex containing symmetrically varied 6 bp sites with a central 5'-TC-3' base step. Lanes 1 and 2: A+G and C+T sequencing reactions (Maxam-Gilbert); Lane 3; 0.1 μM Δ-1-Rh(GEB)$_2$phi^{5+} ; Lane 4: 0.1 mM Δ-1-Rh(MGP)$_2$phi^{5+}. The indicated site of cleavage is 5'-CAT*C*TG-3' with cleavage at the italicized C and no corresponding cleavage on the opposite strand.

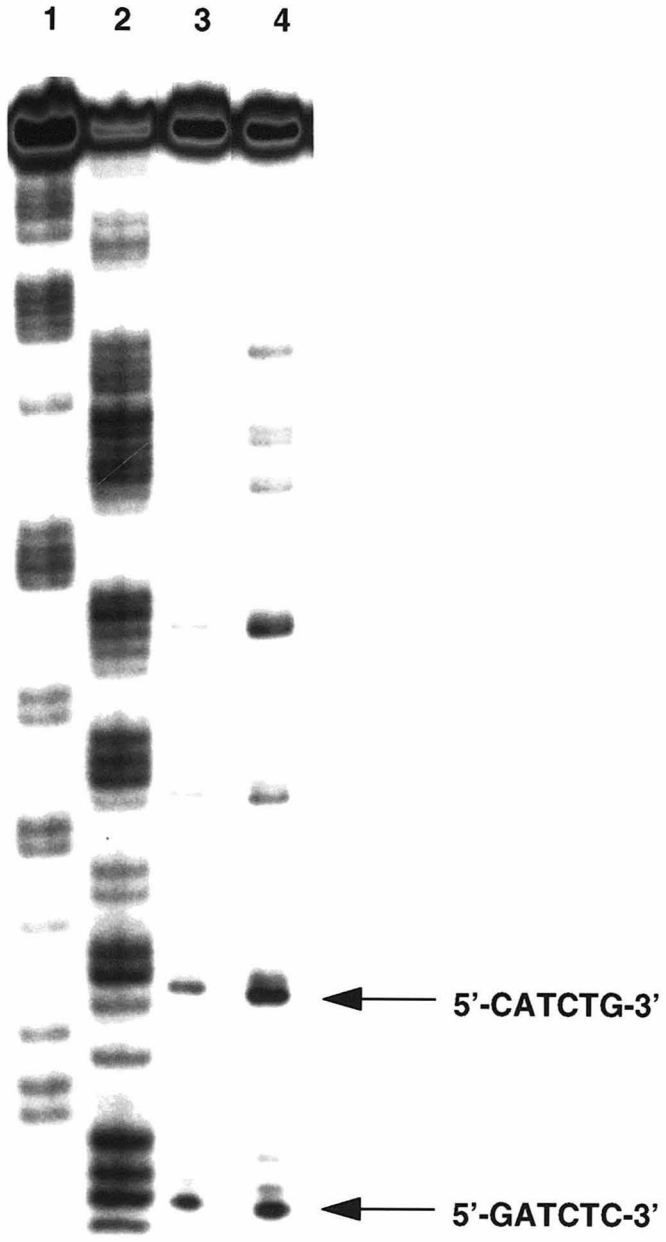

1
2
3
4
5'-CATCTG-3'
5'-GATCTC-3'

Table 2.1. Comparison of the the relative cleavage intensity of Λ-1-$Rh(MGP)_2phi^{5+}$ and Δ-1-$Rh(MGP)_2phi^{5+}$ as a function of variation in consenus sequence. The strongest cleavage site for each complex is set at 100%.

Effect of Variations in the Consensus Sites.

Site	Λ-1-Rh(MGP)$_2$phi^{5+}	Site	Δ-1-Rh(MGP)$_2$phi^{5+}
5'-CATATG-3'	100	5'-CATCTG-3'	100
5'-CACGTG-3'	0	5'-CAGCTG-3'	10
5'-TATATA-3'	73	5'-TATCTA-3'	75
5'-GATATC-3'	67	5'-GATCTC-3'	68
5'-CTTAAG-3'	16	5'-CTTCAG-3'	48

implies a modular nature to the manner in which these metal complexes recognize their DNA sites.

2.3.7. Qualitative Determination of Exchange Rates. The exchange rate of a molecule can potentially affect its ability to inhibit a reaction and function as a chemotherapeutic. We were interested in determining whether Δ and Λ-1-$Rh(MGP)_2phi^{5+}$ show a decrease in exchange rates with DNA as compared to $Rh(phen)_2phi^{3+}$. NMR and photocleavage experiments have previously shown that the parent molecule $Rh(phen)_2phi^{3+}$ has an exchange rate on oligonucleotides of many times per second (3, 9). A photocleavage assay was therefore used to investigate the rate at which Δ- and Λ-1-$Rh(MGP)_2phi^{5+}$ dissociate from a radiolabeled DNA fragment in the presence of a large excess of unlabeled carrier DNA. If Δ- and Λ-1-$Rh(MGP)_2phi^{5+}$ possess an exchange rate which is comparable to the exchange of $Rh(phen)_2phi^{3+}$, then the metal complexes should quickly equilibrate between the hot and cold DNA, and photocleavage intensity should decrease rapidly. Conversely, if the guanidinium substitution decreases the exchange rate substantially, then the cleavage intensity should decrease at a much slower rate. As shown in Figure 2.8, strong cleavage for both complexes is still observed after 3 hours, indicating that the exchange rates of Δ- and Λ-1-$Rh(MGP)_2phi^{5+}$ are much slower than that of $Rh(phen)_2phi^{3+}$. Rough estimates suggest that a 100 to 1000-fold decrease in DNA exchange rates are observed compared to the parent $Rh(phen)_2phi^{3+}$.

2.4. Implications for Design. Λ-1-$Rh(MGP)_2phi^{5+}$ is the first example of a small molecule which uses both direct and indirect readout of the DNA sequence for sequence selectivity. Deazaguanine substitutions have demonstrated the existence of direct contacts between the metal complex and the N7 nitrogens of

Figure 2.8. Determination of exchange rates for Δ-1-Rh(MGP)$_2$phi^{5+} and Λ-1-Rh(MGP)$_2$phi^{5+}. Photocleavage on a P^{32}-3'-endlabeled 330 base pair restriction fragment (pBR322 cut by Acc I/Drd I). Lanes 1 and 12: DNA only with irradiation.; Lane 2: 0.1 μM Λ-1-Rh(MGP)$_2$phi^{5+}, DNA/Rh 10:1, zero time point; Lane 3-6: 0.1 μM Λ-1-Rh(MGP)$_2$phi^{5+}, DNA/Rh 1000:1; Lane 3: 5 minutes incubation; Lane 4: 30 minutes incubation; Lane 5: 1 hour incubation; Lane 6: 3 hour incubation; Lane 7: 0.1 μM Δ-1-Rh(MGP)$_2$phi^{5+}, DNA/Rh 10:1, zero time point; Lane 8-11: 0.1 μM Δ-1-Rh(MGP)$_2$phi^{5+}, DNA/Rh 1000:1; Lane 3: 8 minutes incubation; Lane 9: 30 minutes incubation; Lane 10: 1 hour incubation; Lane 11: 3 hour incubation.

Note: Photocleavage is still strong after 3 hours for both Λ and Δ-1-Rh(MGP)$_2$phi^{5+}.


1 2 3 4 5 6 7 8 9 10 11 12
5'-CATCTG-3'
5'-CATATG-3'

guanine, and the rigorous requirements for the internal sequence of the complex's binding sites suggest a strong element of indirect readout. Importantly, the similarity between the consensus sites for Δ and Λ-1-Rh(MGP)$_2$phi^{5+} (5-CATCTG-3' versus 5'-CATATG-3'), as well as the similar effect that consensus site sequence variations have on the cleavage intensity of both complexes, suggests that Δ-1-Rh(MGP)$_2$phi^{5+} also recognizes its sites through a combination of direct and indirect readout.

The ability to design and synthesize molecules which are capable of recognizing DNA by both direct and indirect readout of the DNA sequence offers exciting possibilities for potential pharmaceuticals. Direct readout of the DNA sequence allows for an element of sequence predictability in design, while indirect readout gives an extremely high level of sequence specificity. Furthermore, indirect readout offers the potential of recognizing large DNA sites with a little molecule. This is possible since the molecule doesn't need to span the whole DNA structure, such as a cruciform; it only needs to recognize a structural element of the DNA which is only present when the DNA is fully "folded." Thus, it is possible for a complex to recognize large sequences even though it many directly contact only three or four base pairs. As we learn more about DNA structure, it should become possible to design small molecules which are capable of predictably utilizing both elements of DNA recognition. In the future, it might be possible to achieve a high enough level of specificity that small molecules that are capable of functioning as sequence specific transcriptional inhibitors can be synthesized. Functionalities like the guanidinium moiety can play an important role in achieving this goal.

2.5. References and Notes

1. Sitlani, A., Long, E.C., Pyle, A.M. and Barton, J.K., *J. Am. Chem. Soc.*, **114**, 2303 (1992)
2. Pavletich, N.L. and Pabo, C.O., *Science*, **252**, 809 (1991)
3. David, S.D. and Barton, J.K., *J. Am. Chem. Soc.*, **115**, 2984 (1993)
4. Maniatis, T., Fritsch, E.F. and Sambrook, J., *Molecular Cloning*, Cold Spring Harbor Laboratory (1982)
5. Caruthers, M.H., Barone, A.D., Beaucage, S.L., Dodds, D.R., Fisher, E.F., McBride, L.J., Matteucci, M., Stabinsky, Z. and Tang, J-Y., *Methods Enz.*, **154**, 287 (1987)
6. Pyle, A.M., Long, E.C., and Barton, J.K., *J. Am. Chem. Soc.*, **111**, 4520 (1989)
7. Jain, S.K., Inman, R.B. and Cox, M.M., *J. Biol. Chem.*, **267**, 4215 (1992)
8. Sardesai, N.Y., Zimmerman, K. and Barton, J.K., *J. Am Chem. Soc.*, **116**, 7502 (1994)
9. NMR experiments have shown that the complex is in intermediate exchange on the NMR time-scale. This suggests that the complex is coming on and off the helix many times every second.

Chapter 3: A New Facet of Indirect Readout: Recognition Based on the Sequence Dependent Twistability of the DNA Helix

3.1. Introduction

Recognition by Λ-1-Rh(MGP)$_2$phi^{5+} is difficult to reconcile based simply upon the functionalization of the parent Rh(phen)$_2$phi^{3+}. Λ-1-Rh(MGP)$_2$phi^{5+}, unlike Rh(phen)$_2$phi^{3+}, binds preferentially at sites containing a central 5'-TA-3' base step. Despite the periphery of the complex being modified, it is the internal part of the sequence which seems to be most rigorously controlled. Furthermore, Δ-enantiomers of a wide variety of octahedral polypyridyl complexes, including Rh(phen)$_2$phi^{3+}, are favored for intercalation into the right-handed DNA helix (1), yet it is Λ-1-Rh(MGP)$_2$phi^{5+} which shows a higher DNA binding affinity than its corresponding Δ-isomer. Moreover, both Rh(phen)$_2$phi^{3+} and Λ-1-Rh(MGP)$_2$phi^{5+} cleave DNA with similar quantum efficiencies, yet both the central T and A sites are cleaved by Λ-1-Rh(MGP)$_2$phi^{5+} while Rh(phen)$_2$phi^{3+} cleaves singly. In an effort to reconcile these inconsistencies, experiments aimed at elucidating the exact nature of the interaction of Λ-1-Rh(MGP)$_2$phi^{5+} with its consensus sequence were performed.

3.2. Experimental

3.2.1. Materials. Oligonucleotides were synthesized via the phosphoramidite method on an ABI 391 DNA-RNA synthesizer and purified by reverse phase HPLC unless stated otherwise. Phosphoramidites, solid supports and reagents were purchased from ABI and Glen Research. T4 DNA ligase and all restriction enzymes were purchased from New England Biolabs and Boehringer Mannheim. DNA concentrations were determined by absorption spectroscopy.

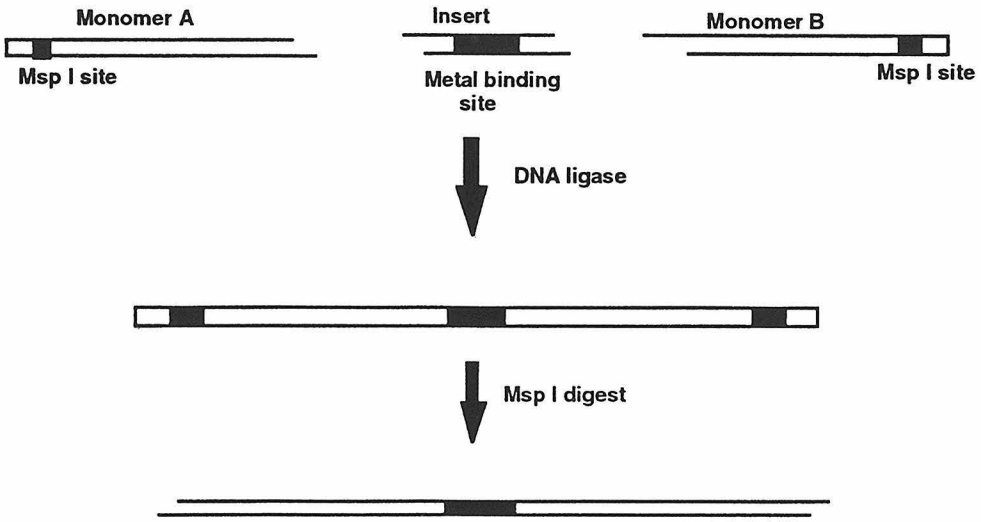
3.2.2. Instrumentation. Absorption spectra were recorded on a Cary 219 spectrophotometer. High performance liquid chromatography (HPLC) was carried out on a Waters 600E system equipped with a Waters 484 tunable detector. A Vydac reverse phase protein and peptide C18 column was used for HPLC separations. Gel electrophoresis experiments were quantified using a Molecular Dynamics phosphorimager and ImageQuant software. Computer modelling was performed on a Silicon Graphics Indigo work station using InsightII (Biosym.).

3.2.3 DNA Unwinding Assay: Substrate Preparation. The different oligonucleotide substrates used in the unwinding assay were synthesized in three sections. Two different invariant outer segments, monomer A and monomer B (see figure), each containing a 64 base pair double stranded region and a 3 base pair 5' overhang (monomer A and B have different overhangs) were synthesized. The sequence of the long bent segments is such that a 100 degree bend is encoded by 5 inphase A_6 tracts. Furthermore, a terminal Msp I site was incorporated to allow for the generation of complementary 2 base pair overhangs in the final full length oligonucleotides following restriction enzyme digestion. The central insert oligonucleotides varied from 10 to 16 base pairs in length and each possessed two different 3 base pair 3'-overhangs. The central inserts (+0 to +6) all contained a binding site for $Rh(MGP)_2phi^{5+}$ as well as spacer DNA to establish the proper phasing between the invariant outer segments. The inserts +0AT and +3AT were synthesized as controls in which the central TA step of the complex binding site, 5'-CATATG-3', is inverted to an AT step.

A schematic of the procedure for generating full length substrates is shown in Scheme 1. The outer invariant segments were ligated to opposite ends of the ^{32}P-5'-end labeled spacer inserts. The full length oligonucleotides (142-148 bp)

were then cut with Msp I to generate truncated fragments (116-122) with complementary sticky ends.

Scheme 3.1.



The assay substrates were synthesized as follows using a modified version of the procedure described by Crothers (2). Oligonucleotide strands of monomers A and B were made on a DNA synthesizer as were the oligonucleotide strands for the nine different inserts indicated in Figure 5. The strands of monomers A and B were purified by reverse phase HPLC followed by purification by gel electrophoresis (10% denaturing polyacrylamide gel). The spacer inserts were purified by reverse phase HPLC. For both the monomers and the spacer inserts, the complementary strands were mixed and annealed by cooling from 95°C to 23°C over a 1 hour period. The duplex monomers A and B (10 μg each) were phosphorylated (in the same reaction tube) with cold ATP (1.0 mM) using 15 units polynucleotide kinase (Boehringer Mannheim, 10 units/μL) in a 50 μL reaction mixture (70 mM Tris•HCl, pH 7.6, 10 mM $MgCl_2$, 5 mM dithiothreitol) at 37°C for 2 hours. The annealed inserts (2 μg each) were separately phosphorylated with 10 μCi of [γ-^{32}P]ATP by 8 units of polynucleotide kinase in a 20 μL reaction mixture at 37°C for 1 hour. After labeling with [γ-^{32}P]ATP, 2 μL of 10 mM cold ATP and an additional 10 units of polynucleotide kinase were added. The reaction

was heated at 37°C for an additional hour. Following phosphorylation of both the inserts and the monomers, the inserts were combined with the monomers (final volume 70 μL for each insert), and 10 μL of 50 mM EDTA was added to quench the reaction. The reaction was phenol/chloroform extracted and ethanol precipitated. The pellets were resuspended in 47 μL of buffer (70 mM Tris•HCl, pH 7.6, 10 mM $MgCl_2$, 1 mM ATP and 5 mM dithiothreitol) and 3 μL of T4 DNA ligase (New England Biolabs, 400 units/μL) was added. The reaction was allowed to proceed at 23°C for 8 hours; 50 μL of 10 mM EDTA was then added and the reaction was heated at 65°C for 30 minutes longer. The ligation reactions were then phenol/chloroform extracted and ethanol precipitated. The DNA pellets were resuspended in 95 μL of buffer (50 mM NaCl, 10 mM Tris-HCL, pH 7.9, 10 mM $MgCl_2$ and 1 mM dithiothreitol) and 5 μL of Msp I restriction enzyme (New England Biolabs, 20 units/μL) was added. Following digestion for 4 hours at 37°C, 10 μL of 10 X non-denaturing gel loading dye was added, and the reaction was loaded onto a 6% non-denaturing preparatory gel. Following electrophoresis, the gel was autoradiographed, and the bands corresponding to the full length sticky-end fragments were isolated (116-122 bp). Following electroelution, the DNA was ethanol precipitated and stored at -70°C. Prior to usage in cyclization experiments, the DNA was resuspended in water and quantitated by UV-visible spectroscopy. (See appendix 1 for additional explanation.)

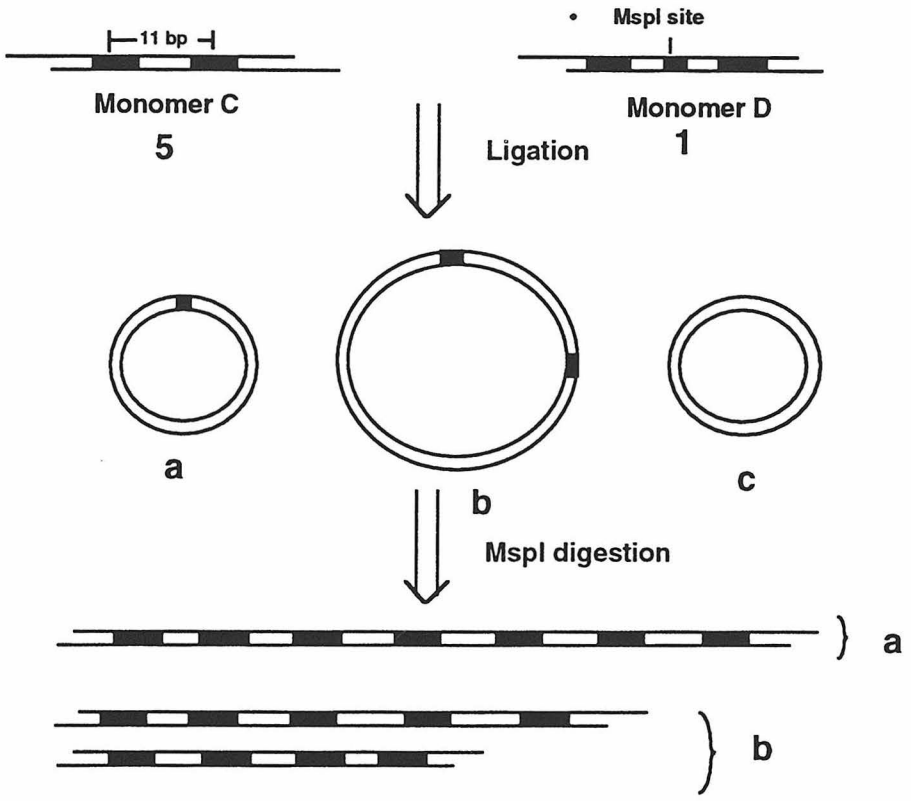
3.2.4. Cyclization Experiments. Ligation reactions were carried out as follows: DNA ends = 2.0×10^{-7} M; DNA ligase = 8.0×10^{-10} M in DNA ligase buffer (50 mM Tris-HCl, pH 7.8, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP, 50 μg/mL BSA); $Rh(MGP)_2phi^{5+}$ = 1.0×10^{-6} M. The original reaction volume was 120 μL and all reactions were carried out at ambient temperature. An 8 μL aliquot was withdrawn prior to ligase addition for a zero-time point and then 8.0 μL T4 DNA

ligase (Boehringer Mannheim;1u/μL) was added. Time points were taken at 0, 1, 5, 10, and 30 minutes. The aliquots (8 μL) were mixed with 8 μL 50 mM EDTA solution to quench the reaction. Cyclization reactions were run both in the presence and absence of Λ-1-Rh(MGP)$_2$phi^{5+}. Following separation on a 6% native polyacrylamide gel, the samples were analyzed using a phosphorimager (Molecular Dynamics). The fraction of cyclized product formed (compared to total counts in the lane) was quantitated. See appendix 1 for additional information.

3.2.5. Geometrical Analysis of Cyclization Data. Geometrical analysis of the spaced A-tracts was performed on a silicon graphics workstation using InsightII. The DNA fragments were modeled by assuming a smooth 100 degree bend for each set of 5 A_6-tracts, and a straight segment of DNA connecting them. A helical rise of 3.4 Å and a helical twist of 35 degrees was used. Each A-tract fragment is 53 bp longs or 180 Å, and the length of the center piece varies from 34 Å (n= 0) to 54.4 angstroms (n = 6). The bent fragment was further approximated by the vector of the curve (158 Å) which projects from the ends of the center fragment at a 130 degree angle. When the two A-tracts are completely inphase, the end to end distance is 237 Å. Different angles of unwinding where modeled for each fragment and the end to end distances were calculated.

3.2.6. DNA Bending Assay: Substrate Preparation. The method for the preparation of the linear multimers is outlined in Scheme 2. Identical procedures were used for preparing linear multimers with 10.5 and 13.0 bp phasing repeats, and all that was changed was the sequence of the starting monomers.

Scheme 3.2.



The procedure was adapted from ref. 2. The oligonucleotide strands were synthesized on a DNA synthesizer and purified by reverse-phase HPLC. For each monomer, the complementary strands were mixed and annealed by heating to 90 °C for 4 minutes followed by cooling to 23 °C over a 1 hour period. The duplex monomer C (8 μg) was phosphorylated with cold ATP (0.6 mM) and 20 units of Polynucleotide Kinase in a 50 μL solution (70 mM Tris·HCl (pH 7.6), 10 mM $MgCl_2$, and 5 mM dithiothreitol) at 37 °C for 2 hours. The reaction was quenched by addition of 2.5 μL of 500 mM EDTA. Double stranded monomer D (1.6 μg) was phosphorylated with 350 μCi of [γ-^{32}P]ATP and 10 units of polynucleotide kinase at 37 °C for 1 hour (same buffer as above). 3 μL of 1 mM cold ATP and 10 units of polynucleotide kinase was then added and the reaction was allowed to proceed for 1 hour longer at 37 •C before quenching with 2.5 μL of 500 mM

EDTA. Monomer A and B phosphorylation reactions were combined. Following phenol/chloroform extraction and ethanol precipitation, the pellet was resuspended in 43 μL of ligation buffer (same as above except also 1 mM ATP) and 800 units (400u/μL) of DNA ligase was added. The reaction was allowed to proceed at 37 °C for 4 hours, then 3.0 μg (5 μL) of additional phosphorylated (cold) monomer D was added. After 2 hours more at 23 °C, the reaction was quenched with 2.5 μL 500 mM EDTA. Following extraction and ethanol precipitation, the pellet was resuspended in 95 μL buffer (50 mM NaCl, 10 mM Tris·HCl (pH 7.9), 10 mM $MgCl_2$ and 1 mM DTT) and 100 units Msp I (20 u/μL) was added. The reaction was incubated at 37 °C for 10 hours and the reaction mixtures were loaded directly onto a 6% nondenaturing preparatory polyacrylamide gel. The gel was run at 500 V for 5 hours at 4 °C. The bands were viewed by autoradiography and excised from the gel. Following electroelution, the fragments were used in cyclization assays.

3.2.7. Cyclization Reactions. Cyclization reactions were run at 23 •C with a multimer concentration of 3.5×10^{-9} M ends and a rhodium concentration (when present) of 1.0×10^{-7} M. Cyclization reactions were run in 70 mM Tris·HCl (pH 7.6), 1 mM ATP, 10 mM $MgCl_2$, and 5 mM dithiothreitol. The original reaction volume was 120 μL and all reactions were carried out at ambient temperature. An 8 μL aliquot was withdrawn prior to ligase addition for a zero-time point and then 8.0 μL T4 DNA ligase (Boehringer Mannheim;1u/μL) was added. Time points were taken at 0, 1, 5, 10, and 30 minutes. The aliquots (8 μL) were mixed with 8 μL 50 mM EDTA solution to quench the reaction. Cyclization reactions were run both in the presence and absence of Λ-1-$Rh(MGP)_2phi^{5+}$ and 2-$Rh(MGP)_2phi^{5+}$. Following separation on a 6% native polyacrylamide gel, the samples were analyzed using a phosphorimager (Molecular Dynamics).

3.3. Results and Discussion

3.3.1. Molecular Modeling of Λ-1-Rh(MGP)$_2$phi^{5+} Bound to 5'-CATATG-3'.

Molecular modeling was carried out using a Silicon Graphics Iris Indigo workstation with Insight II (Biosym). Λ-1-Rh(MGP)$_2$phi^{5+} was modeled bound to its consensus site 5'-CATATG-3' in which the central 5'-TA-3' step was unwound by 20 degrees (a standard intercalation site) and separated by 6.8 Å. The phi ligand of Λ-1-Rh(MGP)$_2$phi^{5+} was inserted into the TA step in a symmetrical fashion. The rest of the sequence was canonical B-form DNA (Figure 3.1). Molecular modeling shows that the guanidinium moieties in Λ-1-Rh(MGP)$_2$phi^{5+} should be unable to form a hydrogen bond with the N7 of the guanine bases in the first and sixth positions of the recognition site if the DNA (minus the intercalation site) is standard B-form, and we know from the deazaguanine experiment that such a contact does take place. Subsequent modeling in which the helical twist between consecutive base pairs was varied suggests that it is necessary to unwind the six base pair sequence by approximately 150 degrees to be able to make strong (non-water mediated) hydrogen bonds (Figure 3.2). Such a distortion would be unprecedented in the binding of small molecules to DNA.

Such a large unwinding might help to explain the binding characteristics of Λ-1-Rh(MGP)$_2$phi^{5+}. If the DNA sequence 5'-CATATG-3' is significantly unwound over the six base pair recognition site, then the right-handed helicity of the site would be diminished as would the preference for right-handed Δ-isomers as intercalators. In addition, when the DNA site is unwound, the deoxyribose sugars 5'- and 3'- to the intercalation site would become more symmetrically disposed with respect to the intercalated phi ligand, so that the photoexcited complex would have comparable access in abstracting a hydrogen from either the 3'- or 5'- sugars, yielding a double cut in photocleavage experiments. Lastly and perhaps most importantly, the untwisting of the DNA will place the guanines in the first and sixth

Figure 3.1. Molecular modeling of Λ-1-Rh(MGP)$_2$phi^{5+} bound to the canonical B-form sequence, 5'-CATATG-3'. Λ-1-Rh(MGP)$_2$phi^{5+} (yellow) is intercalated into the 5'-TA-3' step which is unwound by 20 degrees. The guanidinium moiety of the complex (yellow and blue) is unable to align with the N7 and O6 (blue) of the guanine 3 bases away.

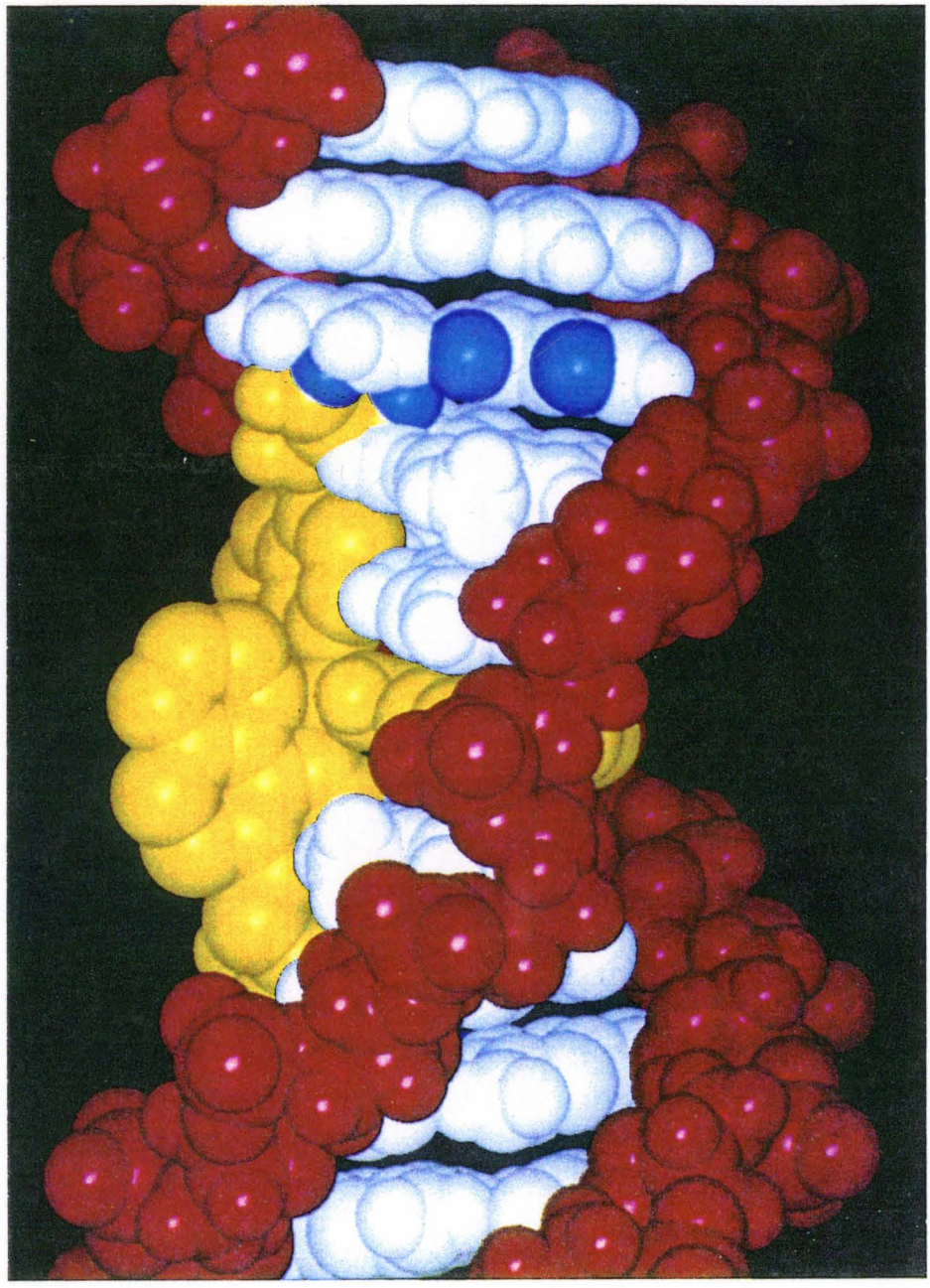
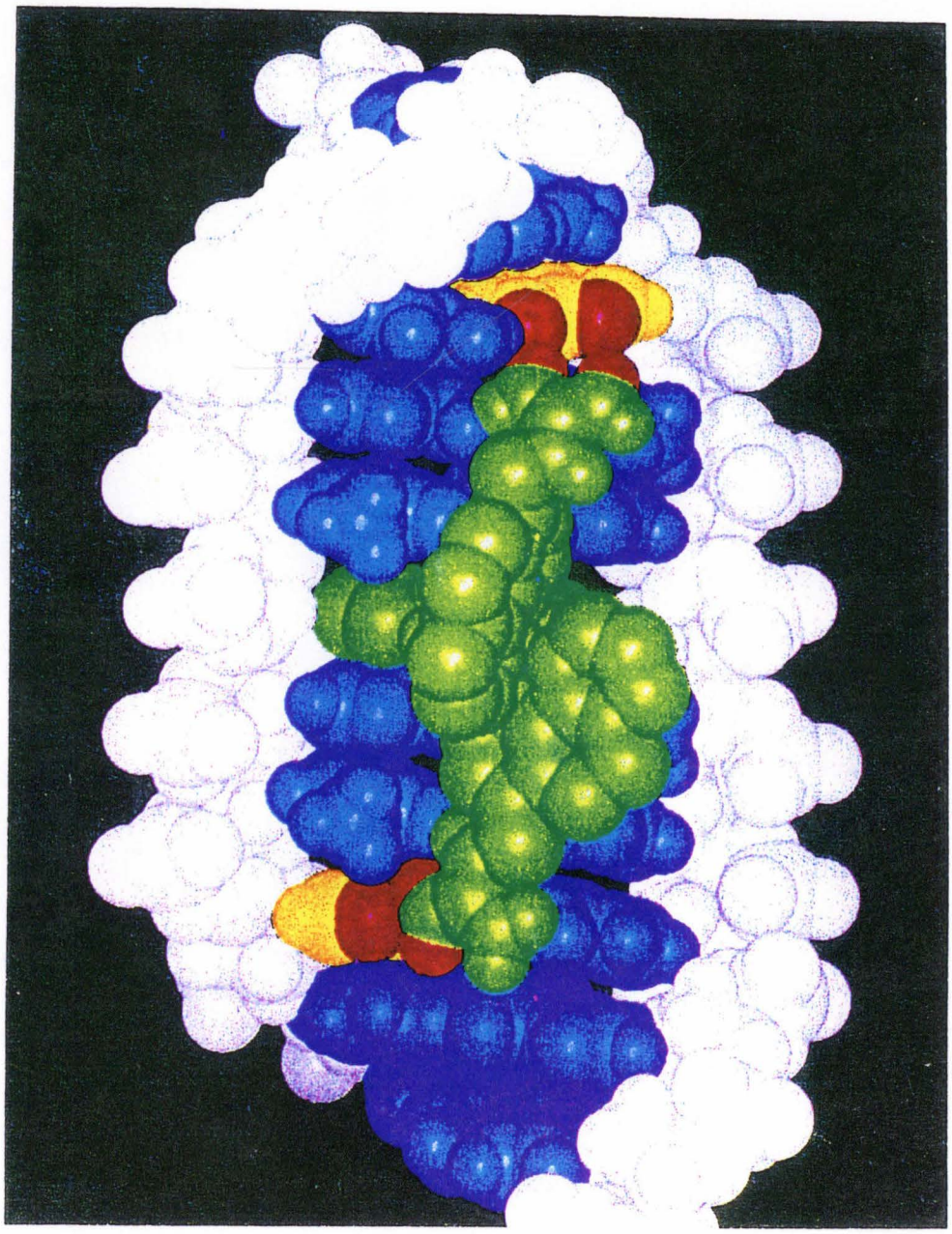


Figure 3.2. Molecular modeling of Λ-1-Rh(MGP)$_2$phi^{5+} bound to a highly unwound 5'-CATATG-3' sequence. The DNA sequence is unwound by nearly 150 degrees with respect to normal B-form DNA. The central 5'-TA-3' step is unwound by 34 degrees and the other base steps are unwound by 28 degrees. As a result of the unwinding, the guanines in the first and sixth positions of the recognition site are on the same face of the helix and can be contacted simultaneously by the metal complex. The guanidinium moiety at the top of the page (green and red) can make a bidentate contact with the N7 and O6 (red and yellow) of the guanine in position 1. The other guanidinium moiety is able to make a single hydrogen bond with the N7 of the guanine in the sixth position.



position on the same face of the DNA polymer so that the complex could make both necessary N7 contacts.

While DNA unwinding of this magnitude is unprecedented in small molecules, numerous protein operator sites are severely distorted upon protein binding. The best known examples are the TATA-box Binding Protein (TBP) (5) and the Catabolite Activator Protein (CAP) (6). The DNA binding properties of CAP have already been discussed in chapter 1. In the crystal structure, TBP is bound to the eight base pair site 5'-TATATAAA-3' which is unwound by nearly 120 degrees and bent towards the major groove by nearly 100 degrees (Figure 3.3). The six base pair site, 5'-TATATA-3', which closely resembles the consensus site for Λ-1-Rh(MGP)$_2$phi^{5+}, is unwound by nearly 85 degrees.

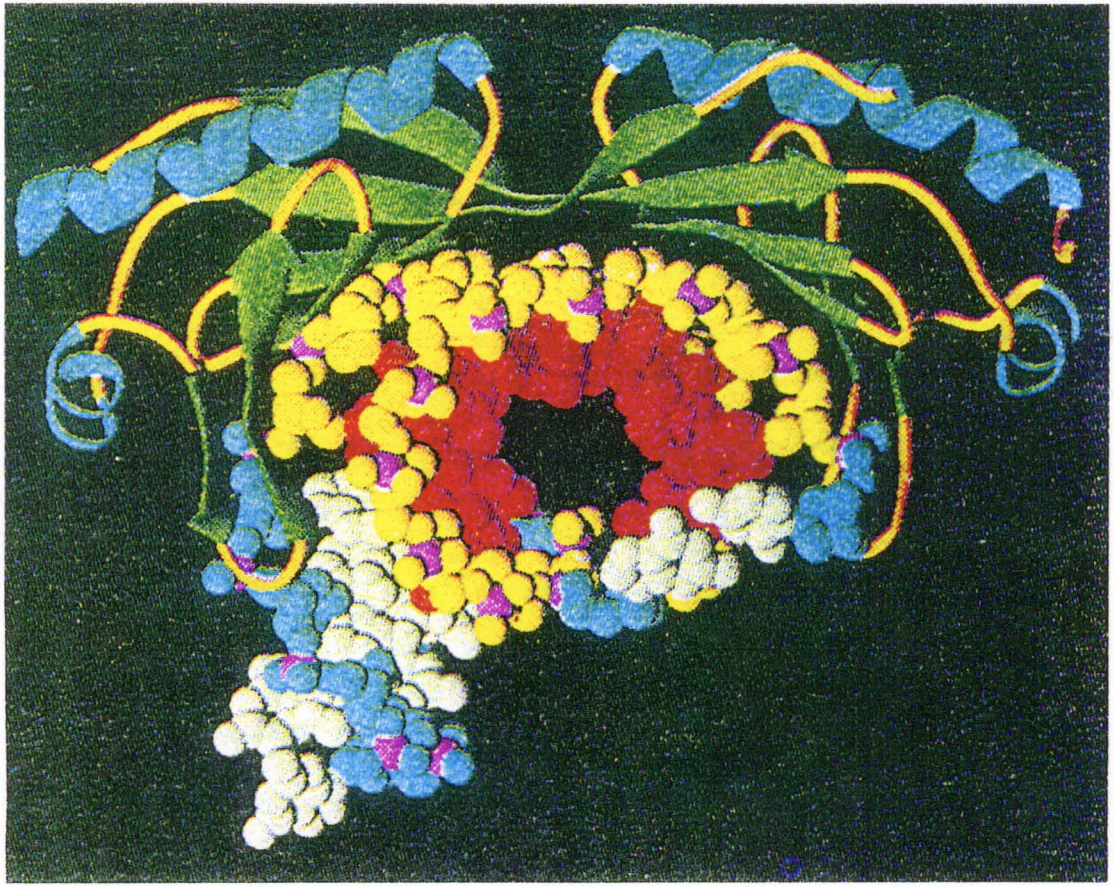
3.3.2. DNA Unwinding Assay. Figure 3.4 illustrates an assay we designed to test for sequence-specific DNA unwinding in solution. This assay was derived based upon the elegant assays developed by Crothers and co-workers to test for DNA bending (2,7,8,10-17). Our assay also resembles a procedure adopted by Hillen to examine DNA unwinding by the Tet repressor (18). We synthesized a series of oligonucleotides in which two DNA segments, each with a bend of approximately 100 degrees encoded by five in-phase A_6 tracts, were separated by a spacer sequence containing the 5'-CATATG-3' binding site (Figure 3.5). The spacer length was varied from 10 bp to 16 bp (1 to 1.5 turns of the helix), so that when the two A tracts are in phase, we may define n, the spacer insert length, equal to 0, and when the two tracts are out of phase by half a helical turn, n equals 5 (19). At the ends of the oligonucleotide were placed two base complementary overhangs to allow for the facile intramolecular cyclization of an appropriately bent oligomer (10-17). The rate of cyclization of these different oligonucleotides by T4 DNA ligase

was then measured as a function of spacer length in the presence and absence of rhodium complex (Figure 3.6).

The results are given in Figure 3.7. In the absence of rhodium complex, the rate of cyclization is seen to be high when both A-tract segments are essentially in phase (n = 0, 1). For the in phase oligonucleotide with n = 1, for example, cyclization is seen to be inhibited by Λ-1-Rh(MGP)$_2$phi^{5+}; this observation is consistent with the idea that the metal complex unwinds the DNA and brings the bent segments out of phase. The oligonucleotide with n = 3, in contrast, displays an extent of cyclization which is consistent with having bent segments being out of phase in the absence of metal complex. The presence of the metal complex serves also to unwind this DNA, which brings the bent segments into phase and *increases* the rate of cyclization. Whether the metal complex acts as an inhibitor or an activator of intramolecular cyclization therefore depends upon the relative phasing of the bent segments (20). To establish that the metal-promoted unwinding was localized to the binding site, we also compared ligation for two spacers, 0AT and 3AT, containing the sequence 5'-CAATTG-3', which is not bound preferentially by the metal complex. In the presence of rhodium, no activation of cyclization for the 3AT fragment is observed; unwinding in the presence of the metal complex therefore requires site-specific binding. Subsequent photocleavage studies on the cyclization fragments showed cleavage only at the central TA base step of the binding site and no corresponding cleavage in the AT oligonucleotides.

3.3.3. Analysis of Cyclization Data. The amount of cyclized product obtained should vary as a function of the distance between the two ends of the cyclization fragment. Geometrical analysis was performed on model fragments in which two bent pieces of DNA (53 bp) are separated by a 11 to 16 base pair spacer as a function of unwinding angle. A 100 degree smooth bend was used to estimate the

Figure 3.3. (Left) Computer modeling representation of TBP (ribbon diagram (top)) bound to its DNA operator site (CPK model). TBP binds in the minor groove of the DNA and severely distorts the DNA. Adapted from ref. 5. (Right) Illustration of how the nearly 100 degrees of unwinding in the TBP operator site is distributed over the 8 base pair site. The top numbers represent the observed helical twist for each base step, and the bottom numbers represent the amount of unwinding relative to B-form DNA. The total unwinding over the 5'-TATATA-3' sequence is 84 degrees.



21	18	23	3	21	21	23
-13	-16	-11	-31	-13	-13	-11

5'-T|A|T|A|T|A A A-3'
3'-A T A T A T T T-5'

-13
-16
-11
-31
-13

-84

Figure 3.4. Assay for sequence specific DNA unwinding with Λ-1-$Rh(MGP)_2phi^{5+}$. A schematic illustration of the unwinding assay. An oligonucleotide is constructed which contains two bent segments, each with a bend of approximately 100 degrees encoded by 5 phased A_6 tracts, and a central spacer with an unwinding agent binding site. When the spacer is 1 turn of the helix (n=0), the two bent segments are in phase and may be easily cyclized with T4 DNA ligase. In the presence of the unwinding agent, the two bent segments twist out of phase into a structure which is not easily cyclized. When the two segments are instead separated by 1.5 turns of the helix (n=5) , addition of the unwinding agent will generate a structure which is easily cyclized.

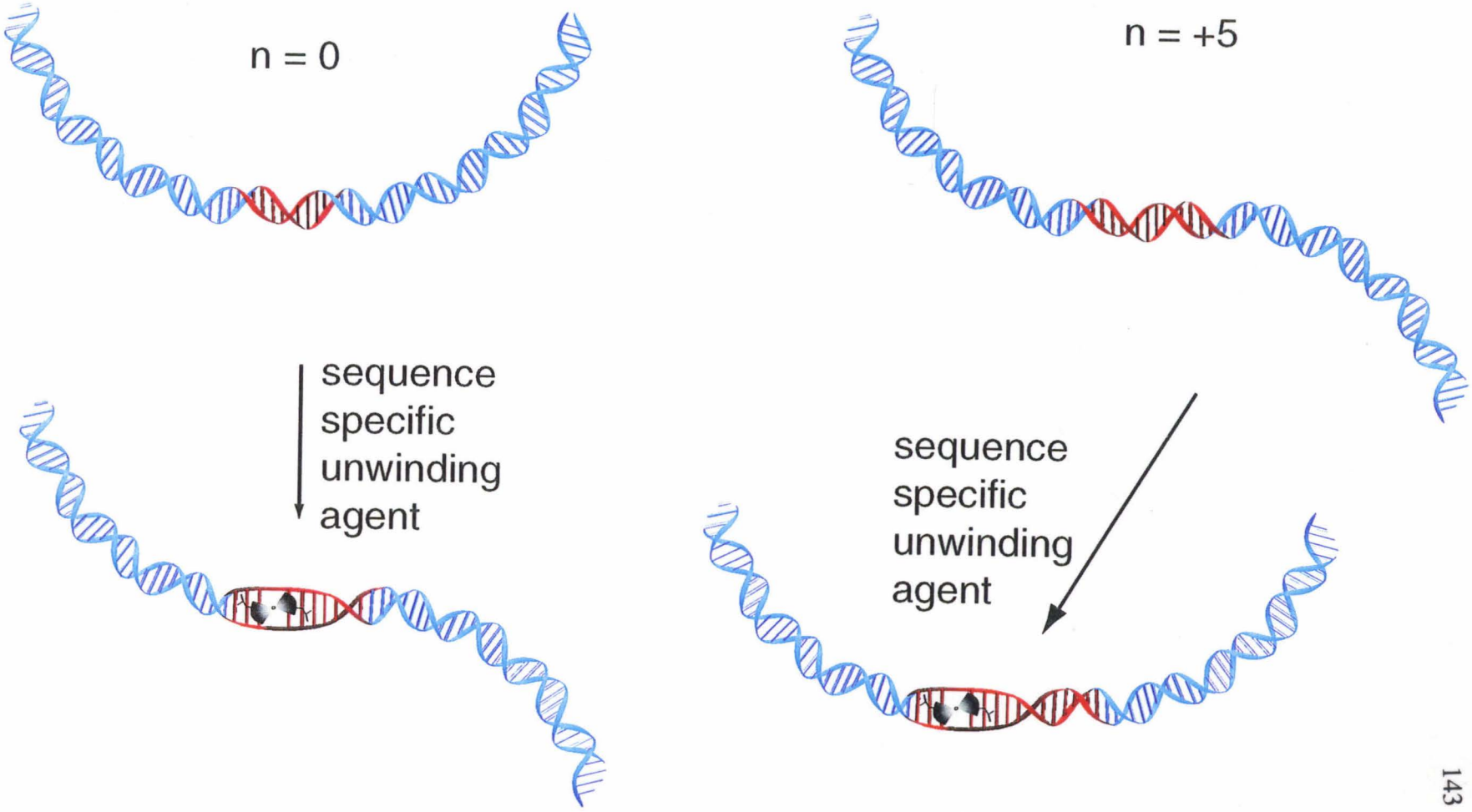

n = 0
sequence
specific
unwinding
agent
n = +5
sequence
specific
unwinding
agent

Figure 3.5. Illustration of the different oligonucleotides synthesized to probe for DNA unwinding. The spacer between the bent segments contains the recognition site (blue) as well as an insert to establish the phasing. When the spacer is 10 bp long, the spacer insert length, n, equals 0, and the two segments are in phase. Segments 0AT and 3AT represent control fragments without the complex binding site.

n =

+0 5'-$A_6A_6GA_6A_6GA_6$ — CGCATATGTC — $A_6A_6GA_6A_6GA_6$-3'

+0AT 5'-$A_6A_6GA_6A_6GA_6$ — CGCAATTGTC — $A_6A_6GA_6A_6GA_6$-3'

+1 5'-$A_6A_6GA_6A_6GA_6$ — CGCATATGTTC — $A_6A_6GA_6A_6GA_6$-3'

+2 5'-$A_6A_6GA_6A_6GA_6$ — CGCATATGGTGC — $A_6A_6GA_6A_6GA_6$-3'

+3 5'-$A_6A_6GA_6A_6GA_6$ — CGCATATGTTCGG — $A_6A_6GA_6A_6GA_6$-3'

+3AT 5'-$A_6A_6GA_6A_6GA_6$ — CGCAATTGTTCGG — $A_6A_6GA_6A_6GA_6$-3'

+4 5'-$A_6A_6GA_6A_6GA_6$ — CGCATATGTTCAGG — $A_6A_6GA_6A_6GA_6$-3'

+5 5'-$A_6A_6GA_6A_6GA_6$ — CGCATATGTTCAGTC — $A_6A_6GA_6A_6GA_6$-3'

+6 5'-$A_6A_6GA_6A_6GA_6$ — CGCATATGTTCAGGTC — $A_6A_6GA_6A_6GA_6$-3'

5'- CGAAAAAACG-3'
3'- GCTTTTTTGC-5' = A_6

Figure 3.6. Illustration of cyclization reactions. (Top Left) In the absence of metal complex, the inphase A-tract fragment (n = 0) quickly cyclizes in the presence of DNA ligase. (Bottom Left) Upon addition of Λ-1-Rh(MGP)$_2$phi^{5+}, the amount of cyclized product, for the n = 0 fragment, decreases significantly. (Top Right) In the absence of metal complex, the out-of-phase A-tract fragment (n = 4) does not cyclize; instead it undergoes a bimolecular reaction to form multimers. (Bottom Right) In the presence of Λ-1-Rh(MGP)$_2$phi^{5+}, the n = 4 fragment forms an appreciable amount of cyclized fragment.

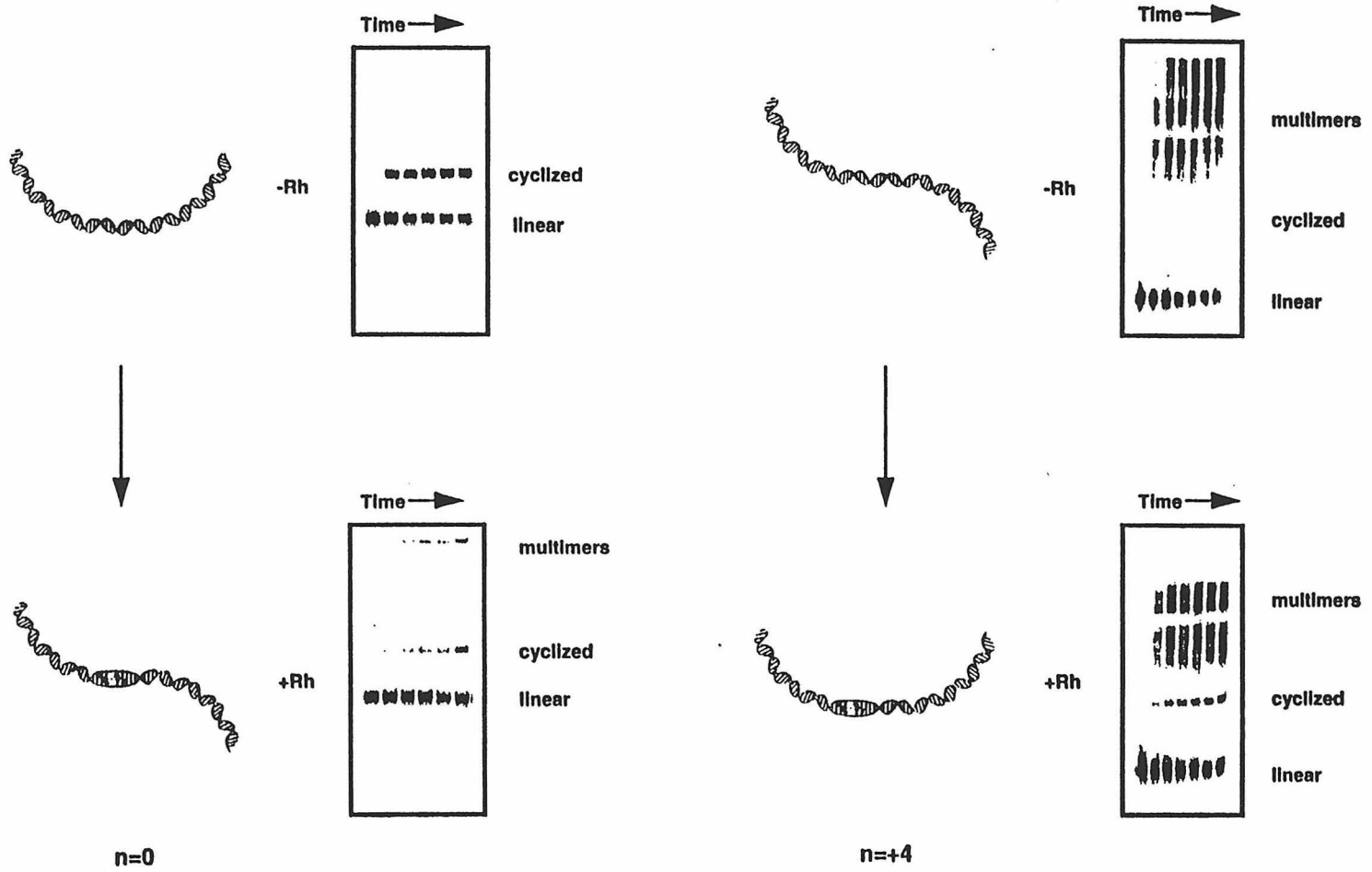
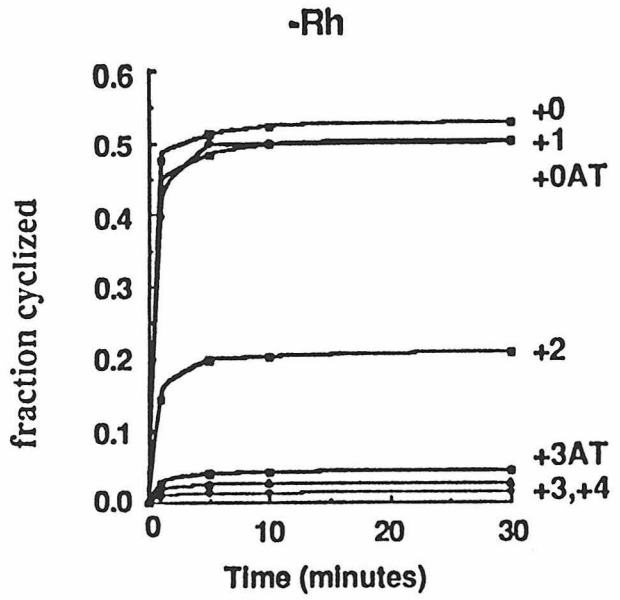
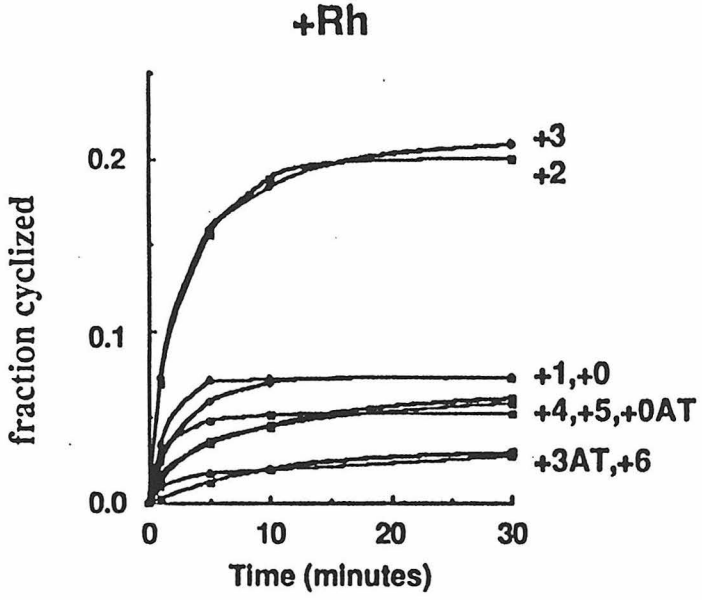
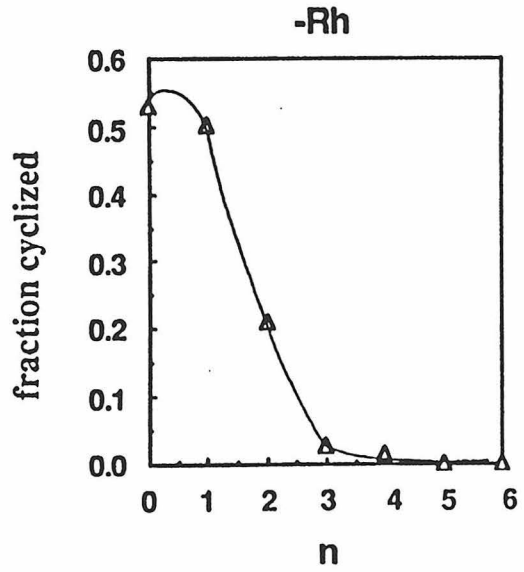

Time
-Rh
cyclized
linear
Time
-Rh
multimers
cyclized
linear
Time
+Rh
multimers
cyclized
linear
Time
+Rh
multimers
cyclized
linear
n=0
n=+4

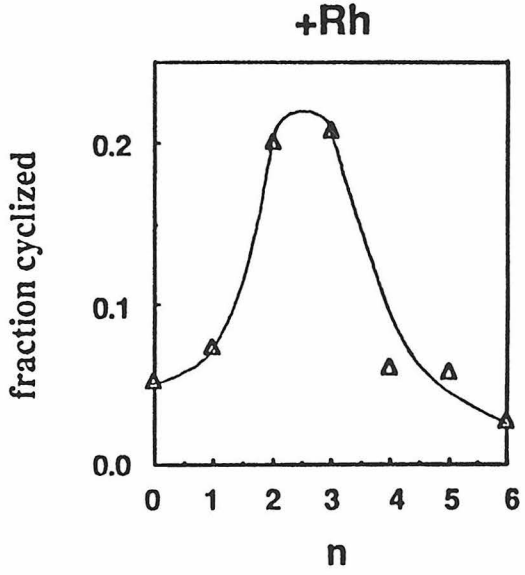
Figure 3.7. The fraction of intramolecular oligonucleotide cyclization by T4 DNA ligase as a function of time is plotted for bent fragments of different spacer insert lengths in the absence (left) and presence (right) of Λ-1-Rh(MGP)$_2$phi^{5+}(1μM). Below are shown the maximum amounts of intramolecular cyclization product in the absence (left) and presence (right) of Λ-1-Rh(MGP)$_2$phi^{5+} as a function of n.

In the absence of rhodium, the peak in amount of cyclized product formed is seen between n=0 and n=1, consistent with a helical repeat of 10.5 bp (n=0.5). In the presence of rhodium, the maximum is shifted to between n=2 and n=3. Note that the absolute amount of intramolecular cyclization is reduced in the presence of rhodium. The data was compared to a plot of end to end distance versus n for different degrees of unwinding. Based on this analysis, in the presence of Λ-1-Rh(MGP)$_2$phi^{5+}, the sequence 5'-CATATG-3' is found to be unwound by 70 ± 10 degrees.


-Rh
fraction cyclized
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
10
20
30
Time (minutes)
+0
+1
+0AT
+2
+3AT
+3,+4


+Rh
fraction cyclized
0.2
0.1
0.0
0
10
20
30
Time (minutes)
+3
+2
+1,+0
+4,+5,+0AT
+3AT,+6


-Rh
fraction cyclized
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
1
2
3
4
5
6
n

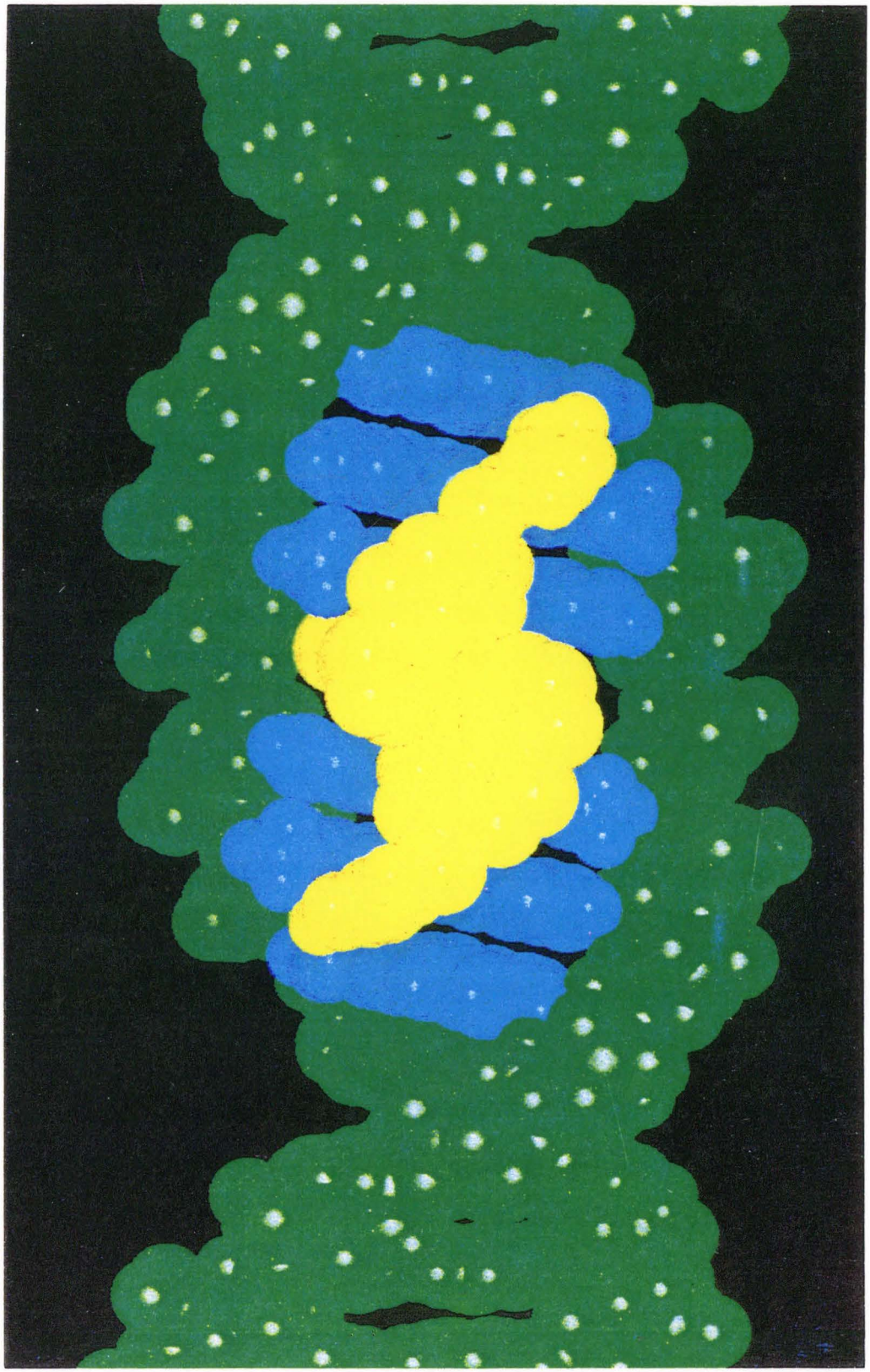

+Rh
fraction cyclized
0.2
0.1
0.0
0
1
2
3
4
5
6
n

A-tract shape, and a helical rise of 3.4 angstroms and helical twist of 35 degrees per base pair were used. These data can be used to estimate the relative amounts of cyclized products for the different cyclization fragments as a function of unwinding angle. The data obtained for the amount of cyclized fragment in the absence and presence of Λ-1-Rh(MGP)$_2$phi^{5+} as a function of spacer length indicate that Λ-1-Rh(MGP)$_2$phi^{5+}, in binding to 5'-CATATG-3', unwinds the oligonucleotide by 70 ± 10 degrees. Figure 3.8 illustrates a structural model for Λ-1-Rh(MGP)$_2$phi^{5+} bound to its DNA binding site based on these data.

3.3.4. DNA Bending upon Binding of Λ-1-Rh(MGP)$_2$phi^{5+} to 5'-CATATG-3'. The crystal structure of TBP bound to its operator site reveals a substantial bend in the DNA fragment. Computer modeling suggests that no such bend should be present in the binding of Λ-1-Rh(MGP)$_2$phi^{5+} to DNA. However, bending experiments were still performed to verify this result. In an assay similar to the one originally used by Crothers to assess the degree of bending in A-tracts (2), a series of DNA fragments in which multiple binding sites (8-20) for Λ-1-Rh(MGP)$_2$phi^{5+} were placed in phase were synthesized (See Scheme 1 in experimental). Two sets of oligonucleotides were synthesized: one set used a helical repeat of 10.5 base pairs between sites, and the other used a helical repeat of 13 base pairs to accommodate the 70 degrees of unwinding of the DNA helix which occurs upon the binding of Λ-1-Rh(MGP)$_2$phi^{5+}. To allow for cyclization of these fragments, 5 bp complementary overhangs were placed on each end. The amount of fragment cyclization by T4 DNA ligase in the presence and absence of Λ-1-Rh(MGP)$_2$phi^{5+} was then measured.

If the DNA site is bent appreciably upon complex binding, then the amount of cyclized fragment should increase in the presence of Λ-1-Rh(MGP)$_2$phi^{5+} due to a decrease in the end to end distance of the fragment. As the number of sites is

Figure 3.8. Molecular modeling of Λ-1-Rh(MGP)$_2$phi^{5+} bound to its recognition site, 5'-CATATG-3'. Λ-1-Rh(MGP)$_2$phi^{5+} (yellow) is shown intercalated in the major groove into the central 5'-TA-3' base step of the DNA duplex 5'-GGCG*CATATG*CGGG-3'. The 6 bp complex binding site, 5'-CATATG-3', is shown in blue and the DNA flanking sequences, which are in the canonical B-form, are in green. In this model the binding site, 5'-CATATG-3', is symmetrically unwound by 70 degrees; the central TA step is unwound by 34 degrees and the flanking base steps are unwound by 9 degrees each. It is noteworthy that the complex can span all six base pairs of the binding site in this model and that the guanidinium arms are well positioned for contacts with the guanines in position 1 and position 6. In the absence of this unwinding, Λ-1-Rh(MGP)$_2$phi^{5+} is unable to fully span the recognition site. Molecular modeling was carried out using a Silicon Graphics Iris Indigo Workstation with Insight II (Biosym).



increased, the amount of cyclized product obtained should increase up to a maximum and then decrease as the bend goes past 360 degrees. By dividing 360 degrees by the number of sites at the maximum in cyclized product, the bending angle in degrees is obtained. No increase in cyclization was obtained in the presence of metal complex for both sets of fragments (10.5 bp and 13 bp phasing) indicating that the complex does not bend the DNA substantially upon binding. We can place a lower limit on bend detection at about 10 degrees.

Importantly, it was observed that when a complex binding site is immediately adjacent to the sticky end overhang (ligation point), the ligation reaction is inhibited by the presence of Λ-1-Rh(MGP)$_2$phi^{5+} but not by 2-Rh(MGP)$_2$phi^{5+}. If a spacer sequence is inserted between the complex binding site and the ligation point, then no inhibition is observed. Thus it appears that sequence specific binding by Λ-1-Rh(MGP)$_2$phi^{5+} is capable of inhibiting the action of DNA T4 ligase.

3.3.5. Comparison of Λ-1-Rh(MGP)$_2$phi^{5+} and TBP. It is interesting to compare features of binding to DNA by Λ-1-Rh(MGP)$_2$phi^{5+} and TBP. In the crystal structure, TBP is bound to a similar sequence, 5'-TATATAAA-3', and the 6 bp 5'-TATATA-3' are unwound similarly by 85 degrees (5). Λ-1-Rh(MGP)$_2$phi^{5+}, however, unlike TBP, does not promote DNA bending. This lack of bending is not surprising given that TBP binds in the minor groove and bends the DNA towards the major groove, while Λ-1-Rh(MGP)$_2$phi^{5+} is bound in the major groove; any tendency of the sequence to bend towards the major groove should be blocked by the rhodium complex.

One may also consider whether both Λ-1-Rh(MGP)$_2$phi^{5+} and TBP promote the unwinding of the DNA sequence or whether both molecules recognize a DNA site which is transiently unwound in the absence of any binding agent.

Transient unwinding could lead to the trapping of the unwound structure by either TBP or the metal complex without penalty energetically. Our experiments do not allow us to distinguish between these possibilities. We also cannot determine how the 70 degree unwinding is distributed over the 6 bp site. It is noteworthy that in a variety of crystal structures, 5'-TA-3' steps tend to be significantly overwound or underwound, consistent with the notion that the TA-segments are inherently twistable. The DNA operator sites of the Trp repressor (21), the P_{22} repressor (22), the 434 repressor (23), the Met repressor (24) and Ner Repressor (25) all possess 5'-TA-3' base steps which are not contacted by the DNA bindng protein, yet still have a substantial effect on the proteins binding affinity. In all cases, helical twist is believed to play an important role. Work by Drew and Travers on enzyme hypersensitivity of TA-rich regions also supports this idea (26). Sequence-dependent DNA twistability may also account for the ability of TA-rich segments to accommodate either one or two distamycins in the minor groove (27). The biological implications of underwinding at TA rich regions are enormous. TBP is required by all three eukaryotic RNA polymerases (28-30). An inherent part of polymerase reaction is DNA strand separation, and the first step of strand separation is unwinding of the DNA helix. Thus, unwinding at TA steps could play a major role in the transcription process.

3.4. Implications for Design

The notion of sequence-dependent DNA twistability leads to important implications in the rational design of novel DNA-binding molecules. If one considers the DNA duplex as a static entity, then sequence-specificity may be built upon an ensemble of weak interactions assembled between functional groups on the binding molecule and the DNA site. As with some DNA-binding proteins, this strategy of direct readout will yield a hierarchy of sequence preferences, with

perhaps 1-2 kilocalorie discrimination between the target sequence and single base mismatches. But the notion of sequence-dependent twistability or more generally of sequence-dependent conformational switches, once coupled to functional group interactions and incorporated into design, offers a significantly higher level of sequence discrimination to be achieved. Indeed, indirect readout mechanisms based upon sequence-dependent flexibilities might be exploited by DNA-binding proteins in particular under circumstances where a high level of discrimination is required, as for example in the case of binding to the TATA box to activate transcription. Our challenge remains to delineate how different DNA sequences vary both structurally and dynamically so as to incorporate more rationally this notion of sequence-dependent twistability into design.

3.5. References and Notes

1. J. K. Barton, *Science*, **233**, 727 (1986)
2. H. S. Koo, J. Drak, J. A. Rice and D. M. Crothers, *Biochem.*, **29**, 4227 (1990)
3. N. L. Pavletich and C. O. Pabo, *Science*, **252**, 809 (1991)
4. S. K. Jain, R. B. Inman and M. M. Cox, *J. Biol. Chem.*, **267**, 4215 (1992)
5. Y. C. Kim, J. H. Geiger, S. Hahn and P. B. Sigler, *Nature*, **365**, 512 (1993)
6. Schultz S.C., Shields, G.C. and Steitz, T.A., *Science*, **253**, 1001-1007 (1991)
7. D. M. Crothers, J. Drak, J. D. Kahn, S. D. Levine, *Meth. Enzym.*, **213**, 3 (1992)
8. J. D. Kahn and D. M. Crothers, *Proc. Natl. Acad. Sci. USA*, **89**, 6343 (1992)
9. D. M. Crothers, M. R. Gartenberg and T. E. Shrader, *Meth. Enzym.*, **208**, 118 (1991)
10. S. S. Zinkel and D. M. Crothers, *Nature*, **328**, 178 (1987)
11. R. S. Tang and D. E. Draper, *Nucl. Acid Res.*, **22**, 835 (1994)
12. R. E. Harrington, *Electrophoresis*, **14**, 732 (1994)
13. J. D. Kahn, E. Yun and D. M. Crothers, *Nature*, **368**, 163 (1994)
14. J. Drak and D. M. Crothers, *Meth. Enzym.*, **212**, 46 (1992)
15. J. A. Rice, D. M. Crothers, A. L. Pinto, and S. J. Lippard, *Proc. Natl. Acad. Sci. USA*, **85**, 4158 (1988)
16. D. M. Crothers, T. E. Haran and J. G. Nadeau, *J. Biol. Chem.*, **265**, 7093 (1990)
17. H. S. Koo and D. M. Crothers, *Proc. Natl. Acad. Sci. USA*, **85**, 1763 (1988)
18. M. Niederweis and W. Hillen, *Electrophoresis*, **14**, 693 (1993)

19. Since the helical repeat for A_6 tracts has been shown to be 10.5 bp (16), we synthesized A_6 tracts which alternated between 10 and 11 bp in length.

20. Since the helical repeat for A6 tracts has been shown to be 10.5 bp (16), we synthesized A6 tracts which alternated between 10 and 11 bp in length.

21. Z. Otwinosksi et. al., *Nature*, **335**, 321 (1988)

22. Wu, L. and Koudelka, G.B., *J. Biol. Chem.*, **268**, 18975-18981 (1993)

23. Koudelka, G.B. and Carlson, P., *Nature*, 355, 89-91 (1992)

24. W. S. Somers and S. E. V. Phillips, *Nature*, **359**, 387 (1992)

25. Benevides, J.M. et al., *Biochemistry*, **33**, 10701-10710 (1994)

26. H. R. Drew, J. R. Weeks and A. A. Travers, *EMBO*, **4**, 1025 (1985).

27. B. H. Geierstanger, T. J. Dwyer, Y. Bathini, J. W. Lown and D. E. Wemmer, *J. Am. Chem. Soc.*, **115**, 4474 (1993).

28. Sharp, P.A., *Cell*, **68**, 819-821 (1992)

29. White, R.J. and Jackson, S.P., *Trends Genet.*, **8**, 284-288 (1992)

30. Rigby, P.W.J., *Cell*, **72**, 7-10 (1993)

3.6. Appendix 1: Cyclization Assay

Synthesis of Spaced A-tract fragments- There are some improvements which can be made in the procedure described in the text for the synthesis of the A-tracts. The text procedure involves ligating two "conserved" bent pieces to opposite ends of a center spacer which is 11- 16 bp in length. Following ligation, the fragments are restricted with Msp I to generate complementary two bp overhangs. The following changes should result in improved yields of fragment:

1.) Longer center pieces should be used to insure that the oligonucleotides are completely double-stranded during the ligation process. This can be accomplished by shortening the outer fragments and including two of the A-tract segments (one on each side) in the center piece.

2.) Instead of using 3 bp overhangs for connecting the fragments, 4 or 5 bp overhangs can be used.

3.) Longer lengths of DNA after the Msp I sites on the outer spacers should be used to insure high cleavage efficiency. At least 1 turn of the helix should be used.

4.) It is important to insure complete phosphorylation of the ends of the pieces prior to ligation. One potential way to accomplish this is to make the pieces longer than necessary, but to include an enzyme restriction site at the proper location to truncate the fragment down to the correct size. Following digestion, the ends should be completely phosphorylated.

Cyclization of A-tracts. Cyclization reactions are carried out as described in the text. An important aspect of these reactions is knowing the concentration of DNA fragment which is present in the reaction. One procedure for determining the concentration of DNA is given in reference 1. An alternative method is to measure the absorption at 260 nm for each radioactive fragment prior to use. It is important

to use the same absolute concentration of DNA and rhodium in each reaction to insure reproducibility.

A rhodium to DNA sites ratio of 5:1 was used in all of the cyclization reactions. Attempts to use a 1:1 ratio gave irreproducible results. Since a higher than 1:1 ratio is used, it is necessary to perform control experiments to verify the results. Photocleavage experiments reveal single site binding. However, experiments in which the center site is changed from 5'-CATATG-3' to either 5'-CAATTG-3' or some other site are the best controls. It is very important that these controls are carried out, because without them another explanation for the observed results is that four different complexes intercalate throughout the fragment and act in unison to give the large unwinding.

Analysis of data. Following separation by electrophoresis, the data was analyzed using a phosphorimager. The amount of cyclized product was quantitated and reported as a percent of the total radioactivity in the lane. This automatically corrects for unequal lane loading.

It is also important to measure the amount of uncyclized fragment in the lane. In all reactions, regardless of spacer length or whether metal complex is absent or present, the amount of unreacted fragment should be about the same. This is true because the slow step of the reaction is the dissociation of the enzyme from the DNA, and not the formation of the cyclized or multimer product. Only reactions which show similar amounts of uncyclized fragment should be compared.

Similarly, because the slow step of the reaction is the dissociation of the enzyme from the DNA, reactions originally show a quick burst of cyclized product and then level off. After 10 minutes the reactions are essentially finished, and by a half an hour, no further change is observed. For comparison between samples, the 30 minute time points are used. Importantly, if the amount of cyclized product for

each sample is plotted versus time, the sample which shows the most cyclized product at 1 minute also shows the most product at 30 minutes.

It is interesting to note that the amount of cyclized product formed in the presence of rhodium seems to be less than what is formed in the absence of rhodium. For the samples +0 and +1, about 50% of the fragment cyclized is formed in the absence of rhodium. However, in the presence of rhodium, the +2 and +3 fragments only form about 20% of the cyclized product. If the complex unwinds the DNA by about 70 degrees, then in the presence of rhodium, the +2 and +3 fragments should be essentially in phased. If this is true, then they would be expected to show about 50% of the cyclized product. Furthermore, the amount of unreacted fragment is essentially the same in the presence and absence of rhodium, showing that the rhodium complex is not inhibiting the enzyme. Why is there this drop in cyclized product? The most likely explanation is that the excess metal complex (Rh:DNA sites 5:1) is interacting with the A-tracts, either by groove binding or intercalating, and is acting to stiffen and straighten them. This should result in a decrease in the bend of the A-tracts and reduce the amount of cyclized product that is formed. Thus, to avoid confusion, the with rhodium and without rhodium data are treated separately. For determining the amount of unwinding, it is the peak in the amount of cyclized product formed which matters, not the magnitude of this number.

References

1. H. S. Koo, J. Drak, J. A. Rice and D. M. Crothers, *Biochem.*, **29**, 4227

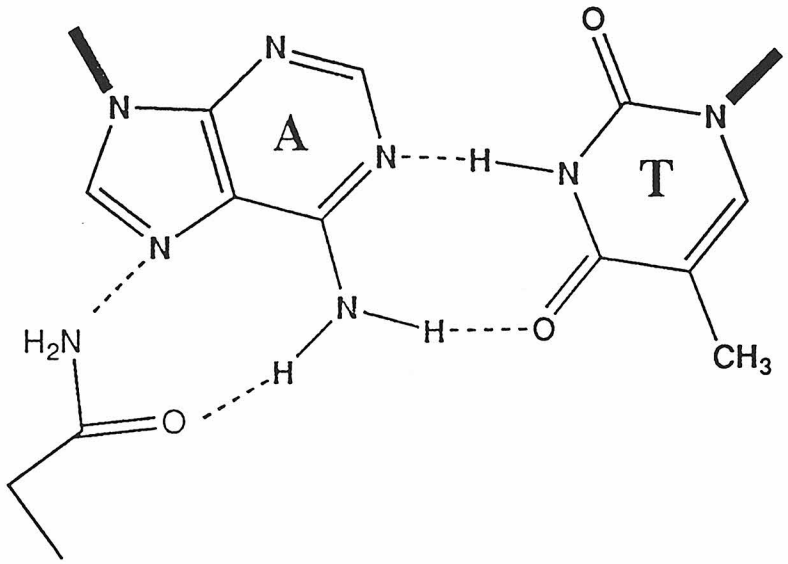
Chapter 4: Investigation of Other Functional Group Interactions: Variations on $Rh(MGP)_2phi^{5+}$

4.1. Introduction

In chapters 2 and 3, we showed how functionalization of the ancillary ligands of $Rh(phen)_2phi^{3+}$ may be used to alter the sequence specificity of this shape-selective DNA intercalator. By tethering a methylguanidinium moiety from the C4 position of phenanthroline ring, metallointercalators, which are capable of recognizing DNA sites by both shape selection and hydrogen bonding interactions, were synthesized. In this chapter, we investigate the application of other functional groups in an effort to explore the scope and limitations of this strategy. We hope to answer the following questions. Is the guanidinium moiety the only functional group which is capable of altering the sequence specificity of metallointercalators or will others functionalities work? How important is the charge on the functional group? How important is the spatial positioning of the functional group with respect to the metal complex core? To answer these questions, the DNA photocleavage properties of phi complexes of rhodium(III) which possess tethered amine and amide functionalities, as well as guanidylethyl and guanidylpropyl moieties, were investigated.

DNA mutagenesis experiments, X-ray structures and NMR solution structures of protein/DNA complexes (see chapter 1) (1-3) have demonstrated that like the guanidinium moiety, the amino group and the amido group play an important role in determining the sequence specificity of many proteins. The amido moiety has the ability to code for adenine base pairs by forming a bidentate hydrogen bond with N7 and C6 amino group of adenine (Figure 4.1). The amino

Figure 4.1. Illustration of the bidentate hydrogen bond formed between the amido moiety and the N7 and C6 amino group of adenine.

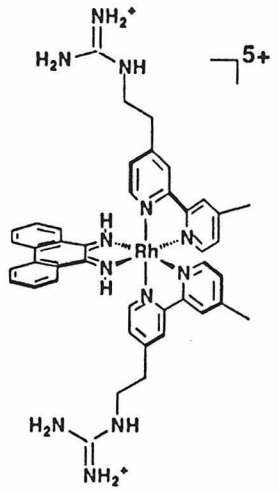

A
T
N
H
O
H_2N
CH_3

group can act as a general proton donor for hydrogen bonding interactions, yet has been seen to preferentially interact with the DNA base guanine.

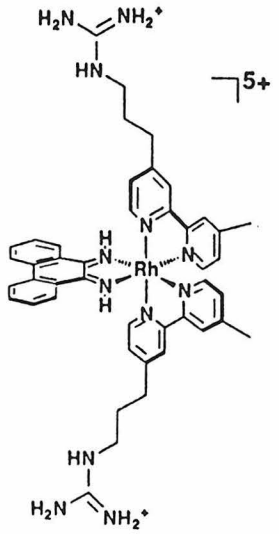
The metal complex $Rh(AEB)_2phi^{3+}$ (Figure 4.2) (AEB = 4-(2-amidoethyl)-4'-methylbipyridine) was designed to target the subset of $Rh(bpy)_2phi^{3+}$ sites with flanking A·T base pairs. Like $Rh(phen)_2phi^{3+}$, the sequence specificity of $Rh(bpy)_2phi^{3+}$ has been shown to be governed by shape-selection (4). However, the bipyridyl ligand is much more easily derivatized than the phenanthroline ligand (5). The distance from the bipyridyl ring system to the terminal nitrogen of amide moiety is roughly equal to the distance from the phenanthroline ring to the terminal nitrogen of the guanidinium moiety in $Rh(MGP)_2phi^{5+}$. Thus, $Rh(AEB)_2phi^{3+}$ can be thought of as an analog of $Rh(MGP)_2phi^{5+}$ which uses the amido moiety instead of the guanidinium moiety to recognize DNA. Importantly, the amido moiety is uncharged in solution and should be able to address the issue of functional group charge in DNA recognition.

The metal complex $Rh(APB)_2phi^{5+}$ (Figure 4.2) (APB = 4-(3-aminopropyl)-4'-methylbipyridine) was designed to target the subset of $Rh(bpy)_2phi^{3+}$ sites which possess flanking guanine bases. Like $Rh(AEB)_2phi^{3+}$, the distance from the bipyridyl ring system to the terminal nitrogen of the amine moiety is roughly equal to the distance from the phenanthroline ring to the terminal nitrogen of the guanidinium moiety in $Rh(MGP)_2phi^{5+}$. This complex has the same overall charge as $Rh(MGP)_2phi^{5+}$, but it uses the amino moiety to recognize DNA. This complex can be used to determine if the DNA recognition seen for $Rh(MGP)_2phi^{5+}$ is unique to the guanidinium moiety or if it can be duplicated with a charged amino group. Importantly, the aminopropyl moiety is much more rotationally unrestrained than the guanidylmethyl moiety and $Rh(APB)_2phi^{5+}$ can also be used to study how linker arm flexibility affects complex specificity.

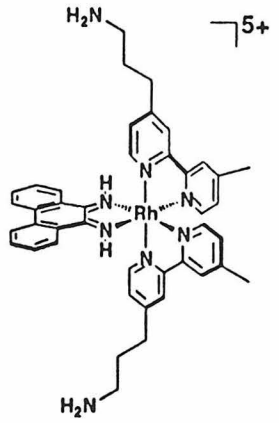
Figure 4.2. Schematic illustration of (Top Left) Δ-1-Rh(GEB)$_2$phi^{5+} (Top Right) Δ-1-Rh(GPB)$_2$phi^{5+} (Bottom Left) Δ-1-Rh(APB)$_2$phi^{5+} (Bottom Right) Δ-1-Rh(AEB)$_2$phi^{3+}.



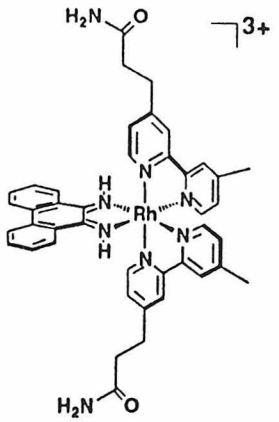
Δ-1-Rh(GEB)$_2$phi^{5+}



Δ-1-Rh(GPB)$_2$phi^{5+}



Δ-1-Rh(APB)$_2$phi^{5+}



Δ-1-Rh(AEB)$_2$phi^{5+}

The DNA recognition properties of these complexes will help develop our understanding of how the identity of the tethered functional group affects the DNA recognition properties of phi complexes of rhodium (III). Furthermore, these studies could lead to the development of an "arsenal" of functional groups which can be used to design and build molecules with predictable sequence selectivities.

4.2. Experimental

4.2.1. Materials. The synthesis and characterization of all metal complexes are described in chapter 5 of this thesis. DNA restriction fragments and oligonucleotides were prepared as described in chapter 2. The concentrations of all rhodium solutions were determined by UV-VIS spectroscopy using ε = 19,400 $M^{-1}cm^{-1}$ at 360 nm. DNA concentrations were determined using ε = 6600 $M^{-1}cm^{-1}$ at 260 nm.

4.2.2. Instrumentation. As described in chapter 2.

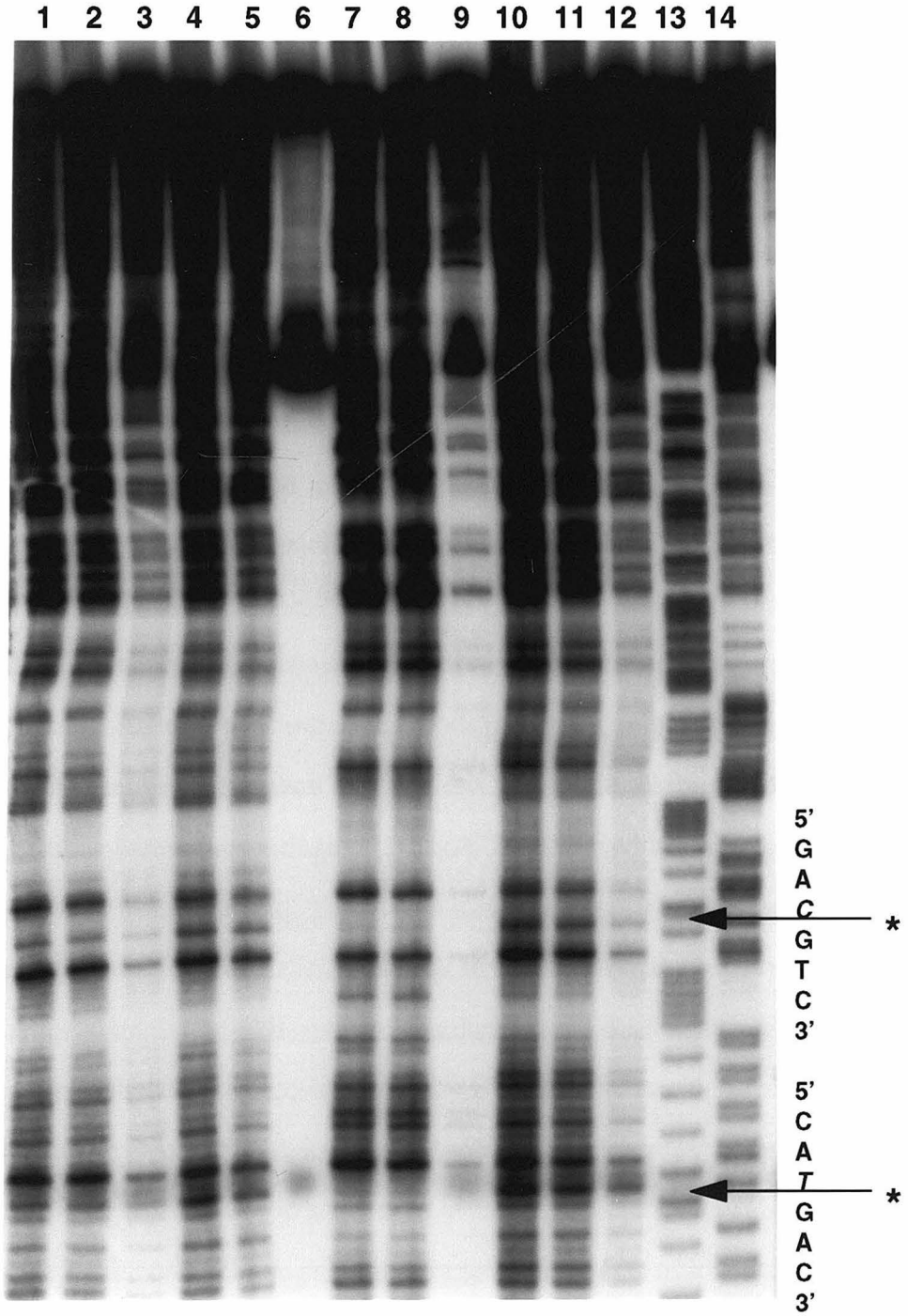
4.2.3. Methods. DNA photocleavage experiments on restriction fragments and oligonucleotides were performed as described in chapter 2.

4.3. Results

4.3.1. DNA Photocleavage by $Rh(AEB)_2phi^{3+}$ and $Rh(APB)_2phi^{5+}$. The DNA photocleavage of $Rh(AEB)_2phi^{3+}$ and $Rh(APB)_2phi^{3+}$ on a 3'-^{32}P-endlabeled 636 bp restriction fragment is shown in figure 4.3. Both complexes promote DNA strand scission upon irradiation with UV light and appears to do so with a quantum efficiency which is similar to that of the parent molecule, $Rh(bpy)_2phi^{3+}$. Since the bipyridine ligands that were used to make the complexes are asymmetric, the complexes were obtained as a 1:2:1 mixture of three different stereoisomers. The effect that the functional moieties have on the DNA recognition

Figure 4.3. Photocleavage with the different isomers of $Rh(APB)_2phi^{5+}$ and $Rh(AEB)_2phi^{3+}$ on a 3'-^{32}P-end-labeled 636 bp restriction fragment. Lane 1: 1.0 μM *rac*-1-$Rh(AEB)_2phi^{5+}$; Lane 2: 0.1 μM *rac*-1-$Rh(AEB)_2phi^{5+}$; Lane 3: 0.01 μM *rac*-1-$Rh(AEB)_2phi^{5+}$; Lane 4: 1.0 μM *rac*-2-$Rh(AEB)_2phi^{5+}$; Lane 5: 0.1 μM *rac*-2-$Rh(AEB)_2phi^{5+}$; Lane 6: 0.01 μM *rac*-2-$Rh(AEB)_2phi^{5+}$; Lane 7: 1.0 μM *rac*-1-$Rh(APB)_2phi^{5+}$; Lane 8: 0.1 μM *rac*-1-$Rh(APB)_2phi^{5+}$; Lane 9: 0.01 μM *rac*-1-$Rh(APB)_2phi^{5+}$; Lane 10: 1.0 μM *rac*-2-$Rh(APB)_2phi^{5+}$; Lane 11: 0.1 μM *rac*-2-$Rh(APB)_2phi^{5+}$; Lane 12: 0.01 μM *rac*-2-$Rh(APB)_2phi^{5+}$; Lane 13: Maxam-Gilbert A+G reaction; Lane 14: Maxam-Gilbert C+T reaction.

Astericks make sites of decreased cleavage for 1-$Rh(APB)_2phi^{5+}$.


1 2 3 4 5 6 7 8 9 10 11 12 13 14
5' G A C G T C 3'
*
5' C A T G A C 3'
*

properties of the complexes can be determined by comparing the photocleavage pattern of the different isomers. (The same designation system that was used for $Rh(MGP)_2phi^{5+}$ is employed here; see chapter 2).

Comparison of the photocleavage pattern of 1-$Rh(AEB)_2phi^{3+}$ (lanes 1-3) with that of 2-$Rh(AEB)_2phi^{3+}$ (lanes 4-6) reveals little difference in the sequence specificity of the two molecules. The overall pattern of cleavage is essentially the same, as well as the intensity of cleavage at high concentrations. One noticeable difference is that 1-$Rh(AEB)_2phi^{3+}$ shows an increase in photocleavage intensity at low concentrations compared to 2-$Rh(AEB)_2phi^{3+}$ and $Rh(phen)_2phi^{3+}$ (lane 3 versus lane 6). However, this enhancement is nonspecific in nature.

The DNA photocleavage by the amine complex, $Rh(APB)_2phi^{5+}$, is also shown in figure 4.3. Comparison of the photocleavage of 1-$Rh(APB)_2phi^{5+}$ (lanes 7-9) with that of 2-$Rh(APB)_2phi^{5+}$ (lanes 10-12) can be used to determine the effect that the aminopropyl substitution has on the DNA photocleavage properties of the complex. At high concentrations, both isomers display a photocleavage pattern which is more intense than that of either the parent $Rh(bpy)_2phi^{3+}$ or the amido complex. Direct comparison of 1-$Rh(APB)_2phi^{5+}$ with 2-$Rh(APB)_2phi^{5+}$ shows some obvious differences. The overall cleavage intensity of the isomer with both arms directed *away* from the DNA, 2-$Rh(APB)_2phi^{5+}$, shows more intense photocleavage than 1-$Rh(APB)_2phi^{5+}$. Furthermore, 1-$Rh(APB)_2phi^{5+}$ shows some sites in which the photocleavage intensity is reduced substantially over that of the parent $Rh(bpy)_2phi^{3+}$ and 2-$Rh(APB)_2phi^{5+}$. These sites are marked by asterisks in figure 4.3. The 6 bp sequences of these sites are 5'-CA*T*GAC-3' and 5'-GA*C*GTC-3' with photocleavage at the italicized bases. It is of interest that there are strong photocleavage sites for the complex two base pairs away from both sites of reduced intensity. It is poosible that binding at the site two bp away blocks binding at these sites of reduced cleavage.

In Figure 4.4, the intensity of photocleavage by 1-Rh(APB)$_2$phi^{5+} and 1-Rh(AEB)$_2$phi^{3+} is directly compared with that of 1-Rh(MGP)$_2$phi^{5+} and Rh(phen)$_2$phi^{3+} on a 3'-^{32}P-endlabeled 330 bp restriction fragment. The strongest site for 1-Rh(APB)$_2$phi^{5+} (lane 7) and 1-Rh(AEB)$_2$phi^{3+} (lane 6) is 5'-CATCTG-3', the consensus site for Δ-1-Rh(MGP)$_2$phi^{5+} (lane 9). Importantly, neither complex cleaves strongly at 5'-CATATG-3', the consensus site for Λ-1-Rh(MGP)$_2$phi^{5+}. The overall intensity of photocleavage by the amino complex, 1-Rh(APB)$_2$phi^{5+}, is greater than that of 1-Rh(MGP)$_2$phi^{5+} (lane 3); however, 1-Rh(APB)$_2$phi^{5+} lacks the specificity of 1-Rh(MGP)$_2$phi^{5+}. The amido complex, 1-Rh(AEB)$_2$phi^{5+}, shows decreased photocleavage relative to 1-Rh(MGP)$_2$phi^{5+} and more closely resembles Rh(phen)$_2$phi^{3+}in both affinity and specificity.

4.3.2. Effect of Variation in Linker Arm Length: Photocleavage with Rh(GEB)$_2$phi^{5+} and Rh(GPB)$_2$phi^{5+}. To test the effect of variation in the length of the linker arm between the metal complex and the guanidinium moiety, the DNA photocleavage properties of two new metal complexes, Rh(GEB)$_2$phi^{5+} (4-(2-guanidylethyl)-4'-methyl-2.2'-bipyridine) and Rh(GPB)$_2$phi^{5+} (4-(3-guanidylpropyl)-4'-methyl-2,2'-bipyridine) (Figure 4.2), were investigated. Both of these complexes have the same general shape as Rh(MGP)$_2$phi^{5+}, but the linker arm between the bipyridyl ligand and the guanidinium is varied from methyl to ethyl to propyl. Both complexes promote DNA strand scission upon irradiation, and both complexes display photocleavage quantum efficiencies which closely resemble that of Rh(phen)$_2$phi^{3+}. Photocleavage by rac-1-Rh(GEB)$_2$phi^{5+} and rac-1-Rh(GPB)$_2$phi^{5+} on a 3'- labeled 330 bp restriction fragment is shown in Figure 4.5.

The DNA photocleavage pattern of 1-Rh(GEB)$_2$phi^{5+} varies depending on the spatial positioning of the ethylguanidinium moiety (Figure 4.5). This is evident

Figure 4.4. Photocleavage with $Rh(phen)_2phi^{3+}$, $Rh(MGP)_2phi^{5+}$, $Rh(GEB)_2phi^{5+}$, $Rh(GPB)_2phi^{5+}$, $Rh(AEB)_2phi^{3+}$, and $Rh(APB)_2phi^{5+}$ on a 3'-^{32}P-labeled 330 bp restriction fragment. Lane 1: Maxam-Gilbert A+G reaction; Lane 2: 0.1 μM *rac*-1-$Rh(phen)_2phi^{5+}$; Lane 3: 0.1 μM *rac*-1-$Rh(MGP)_2phi^{5+}$; Lane 4: 0.1 μM *rac*-1-$Rh(GEB)_2phi^{5+}$; Lane 5: 0.1 μM *rac*-1-$Rh(GPB)_2phi^{5+}$; Lane 6: 0.1 μM *rac*-1-$Rh(AEB)_2phi^{3+}$; Lane 7: 0.1 μM *rac*-1-$Rh(APB)_2phi^{5+}$; Lane 8: 0.1 μM Δ-1-$Rh(MGP)_2phi^{5+}$; Lane 9: 0.1 μM Δ-1-$Rh(GEB)_2phi^{5+}$; Lane 10: 0.1 μM Λ-1-$Rh(MGP)_2phi^{5+}$; Lane 11: 0.1 μM Λ-1-$Rh(GEB)_2phi^{5+}$; Lane 12: DNA only.

Note: Only Λ-1-$Rh(MGP)_2phi^{5+}$ photocleaves at 5'-CATATG-3', and Λ-1-$Rh(MGP)_2phi^{5+}$ is the only metal complex which gives double cleavage bands.

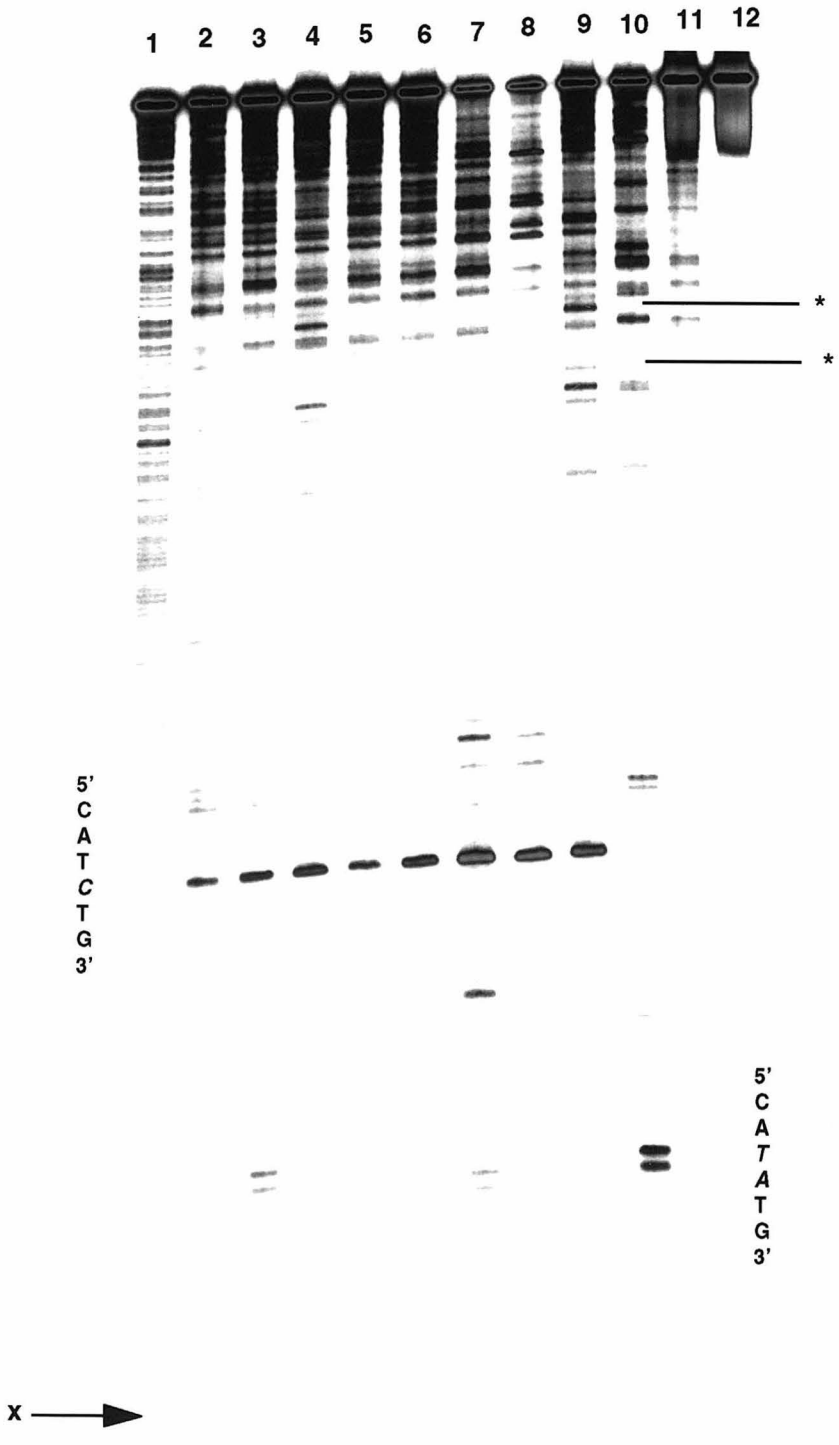
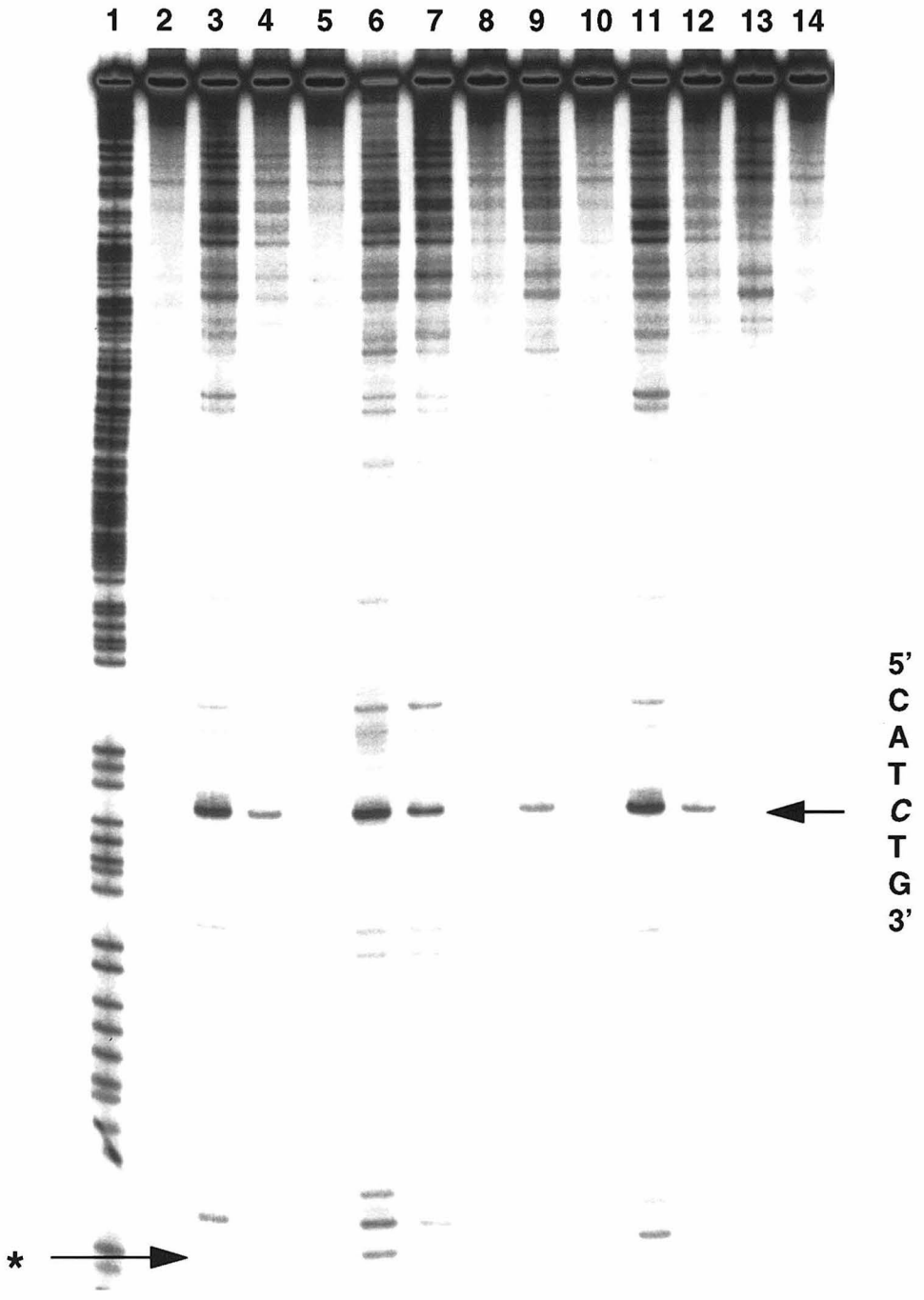

1 2 3 4 5 6 7 8 9 10 11 12
*
*
5' C A T C T G 3'
5' C A T A T G 3'
X

Figure 4.5. Photocleavage with $Rh(GEB)_2phi^{5+}$ and $Rh(GPB)_2phi^{5+}$ on a 3'-^{32}P-labeled 330 bp restriction fragment. Lane 1: Maxam-Gilbert A+G reaction; Lane 2: DNA only; Lane 3: 0.1 μM *rac*-1-$Rh(GEB)_2phi^{5+}$; Lane 4: 0.01 μM *rac*-1-$Rh(GEB)_2phi^{5+}$; Lane 5: 0.01 μM *rac-2*-$Rh(GEB)_2phi^{5+}$; Lane 6: 0.1 μM *rac-2*-$Rh(GEB)_2phi^{5+}$; Lane 7: 0.1 μM *rac*-1-$Rh(GPB)_2phi^{5+}$; Lane 8: 0.01 μM *rac-1*-$Rh(GPB)_2phi^{5+}$; Lane 9: 0.1 μM *rac*-2-$Rh(GPB)_2phi^{5+}$; Lane 10: 0.01 μM *rac*-2-$Rh(GPB)_2phi^{5+}$; Lane 11: 0.1 μM Δ-1-$Rh(GEB)_2phi^{5+}$; Lane 12: 0.01 μM Δ-1-$Rh(GEB)_2phi^{5+}$; Lane 13: 0.1 μM Λ-1-$Rh(GEB)_2phi^{5+}$; Lane 14: 0.01 μM Λ-1-$Rh(GEB)_2phi^{5+}$

Cleavage is isomer specific and enantiospecific for $Rh(GEB)_2phi^{5+}$ and $Rh(GPB)_2phi^{5+}$.


1 2 3 4 5 6 7 8 9 10 11 12 13 14
5'
C
A
T
C
T
G
3'
*

by comparing the photocleavage of 1-Rh(GEB)$_2$phi^{5+} (lane 3 and 4) with 2-Rh(GEB)$_2$phi^{5+} (lane 5 and 6). The best site for both isomers is 5'-CATCTG-3'; however, cleavage by 1-Rh(GEB)$_2$phi^{5+} is stronger at low concentrations (lane 4 versus lane 5). The guanidylethyl complex also displays a significant decrease in cleavage intensity at 5'-TCTCAG-3' (arrow) and, similar to the amine complex, this site is 2 bp away from a strong site of cleavage.

Comparison of the amount of photocleavage by the Δ (lane 11 and 12) and Λ (lane 13 and 14) enantiomers of 1-Rh(GEB)$_2$phi^{5+} show that only the Δ enantiomer shows any appreciable cleavage. This is in contrast to what was seen with 1-Rh(MGP)$_2$phi^{5+}. Importantly, similar to Rh(MGP)$_2$phi^{5+}, DNA photocleavage is isomer specific and enantiomer specific.

Photocleavage with the guanidylpropyl complex (Figure 4.5), 1-Rh(GPB)$_2$phi^{5+}, shows that the photocleavage intensity for this complex varies depending on which isomer is used. Photocleavage by 1-Rh(GPB)$_2$phi^{5+} (lane 7) is significantly more intense than the photocleavage seen by 2-Rh(GPB)$_2$phi^{5+} (lane 9). However, no cleavage is observed for either isomer at low concentrations. Both isomers show less intense photocleavage than what is seen for either isomer of Rh(GEB)$_2$phi^{5+}.

In Figure 4.4, the intensity of photocleavage by 1-Rh(GEB)$_2$phi^{5+} and 1-Rh(GPB)$_2$phi^{5+} is directly compared with that of 1-Rh(MGP)$_2$phi^{5+} and Rh(phen)$_2$phi^{3+} on a 3'-^{32}P-endlabeled 330 bp restriction fragment. The strongest site for the guanidylethyl and the guanidylpropyl complex is 5'-CATCTG-3', the same as it is for the guanidylmethyl complex. Importantly, as the length of the linker arm increases from methyl to ethyl to propyl, the enhancement in photocleavage that is observed relative to Rh(phen)$_2$phi^{3+} decreases. In fact, the intensity of photocleavage by the guanidylpropyl complex at 5'-CATCTG-3' shows a slight decrease with respect to Rh(phen)phi^{3+}.

While the overall photocleavage pattern seen for 1-Rh(GEB)$_2$phi^{5+} is very similar to that of 1-Rh(MGP)$_2$phi^{5+}, there are some differences. As indicated by the asterisks at the top of the gel, 1-Rh(GEB)$_2$phi^{5+} photocleaves at two sites which are not targeted by 1-Rh(MGP)$_2$phi^{5+} . However, the overall cleavage intensity at these sites is greatly reduced compared to photocleavage at 5'-CATCTG-3'.

Direct comparison between the enantiomers of 1-Rh(MGP)$_2$phi^{5+} and 1-Rh(GEB)$_2$phi^{5+} is also shown in figure 4.4. Comparison of the Δ-enantiomers (lanes 8 and 9) shows that the overall pattern of photocleavage is very similar for both complexes. The sequence specificity of Rh(MGP)$_2$phi^{5+} is slightly higher. Importantly, the strongest site of cleavage for both Δ-1-Rh(MGP)$_2$phi^{5+} and Δ-1-Rh(GEB)$_2$phi^{5+} is 5'-CATCTG-3'. Quantitation of the photocleavage using a phosphorimager reveals that cleavage by Δ-1-Rh(MGP)$_2$phi^{5+} is about 50% more intense.

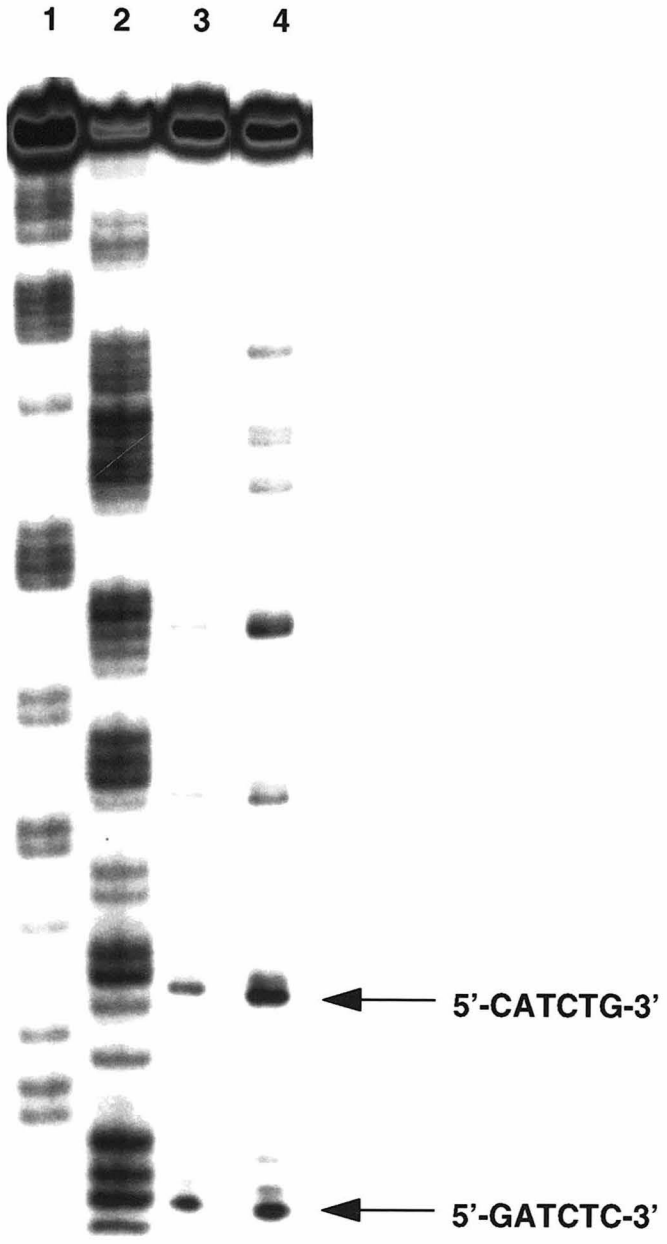
Comparison of Λ-1-Rh(MGP)$_2$phi^{5+} and Λ-1-Rh(GEB)$_2$phi^{5+} shows that only Λ-1-Rh(MGP)$_2$phi^{5+} photocleaves to any appreciable extent. The differences in the photocleavage of the two complexes is dramatic. The consensus site of Λ-1-Rh(MGP)$_2$phi^{5+} is shown at the bottom of the gel, and this site is clearly not targeted by Λ-1-Rh(GEB)$_2$phi^{5+}.

4.3.3. Determination of the Consensus Recognition Site for Δ-1-Rh(GEB)$_2$phi^{5+}.

The consensus recognition site for Δ-1-Rh(GEB)$_2$phi^{5+} was determined using the same oligonucleotides as were used to determine the consensus recognition site for Δ-1-Rh(MGP)$_2$phi^{5+} in chapter 2. Figure 4.6 shows the photocleavage results for both Δ-1-Rh(MGP)$_2$phi^{5+} and Δ-1-Rh(GEB)$_2$phi^{5+}. On this oligonucleotide, Δ-1-Rh(MGP)$_2$phi^{5+} shows much more intense photocleavage than does Δ-1-Rh(GEB)$_2$phi^{5+}.

Figure 4.6. Determination of the consensus sequence for Δ-1-Rh(GEB)$_2$phi^{5+} on a 77 bp DNA duplex containing symmetrically varied 6 bp sites with a central 5'-TC-3' base step. Lanes 1 and 2: A+G and C+T sequencing reactions (Maxam-Gilbert); Lane 3: 0.1 μM Δ-1-Rh(GEB)$_2$phi^{5+} ; Lane 4: 0.1 mM Δ-1-Rh(MGP)$_2$phi^{5+}. The indicated sites of cleavage are 5'-CAT*C*TG-3' and 5'-GATCTC-3' with cleavage at the italicized C and no corresponding cleavage on the opposite strand.

Note the differences in photocleavage between Δ-1-Rh(GEB)$_2$phi^{5+} and Δ-1-Rh(MGP)$_2$phi^{5+} .


1
2
3
4
5'-CATCTG-3'
5'-GATCTC-3'

The photocleavage data were quantitated using a phosphorimager, and the results are given in Table 4.1. The consensus binding site for Δ-1-Rh(GEB)$_2$phi^{5+} is 5'-GATCTC-3'. However, photocleavage at 5'-CATCTG-3' is only slightly less intense. Similar to what was previously seen with Δ-1-Rh(MGP)$_2$phi^{5+} , photocleavage by Δ-1-Rh(GEB)$_2$phi^{5+} was evident on only one strand. This suggests that Δ-1-Rh(GEB)$_2$phi^{5+} may be canted in the binding site (6,7).

4.4. Discussion

The synthetic versatility of phi complexes of rhodium(III) allows us to explore many aspects of DNA recognition. Amino, amido and guanidyl derivatives of Rh(bpy)$_2$phi^{3+} retain the DNA binding and photocleavage properties of the parent complex. Complex binding appears to occur through intercalation of the phi ligand from the major groove of the DNA, and photocleavage products are consistent with direct abstraction of the C3' hydrogen.

As demonstrated in chapter 2 with Rh(MGP)$_2$phi^{5+}, by comparing the different stereoisomers of asymmetrically substituted complexes, it is possible to assess the effect that various substitutions have on the DNA recognition properties of phi complexes of rhodium(III). These data give information about how structural modifications affect the sequence specificity of phi complexes of rhodium(III).

4.4.1. Effect of Charge on the DNA Recognition. Many of the substitutents which we have explored are positively charged under the conditions of the DNA photocleavage experiments. Complexes which possess the guanidinium or amino moiety have a higher overall charge than their parent complexes (+5 versus +3). This increase in charge might be expected to increase the non-specific binding

Table 4.1 Comparison of the relative cleavage as a function of variation in consensus sequence for Δ-1-Rh(GEB)$_2$phi^{5+} and Δ-1-Rh(MGP)$_2$phi^{5+} .

Table 4.1 Effect of sequence variations on the recognition of Δ-1-Rh(MGP)$_2$phi^{5+} and Δ-1-Rh(GEB)$_2$phi^{5+}.

Site	Δ-1-Rh(MGP)$_2$phi^{5+}	Δ-1-Rh(GEB)$_2$phi^{5+}
5'-CATCTG-3'	100	44
5'-CAGCTG-3'	10	5
5'-TATCTA-3'	75	27
5'-GATCTC-3'	68	49
5'-CTTCAG-3'	48	17

affinity of the metal complex and/or increase the likelihood of the recognition moiety being associated with the negatively charged DNA.

The results presented in this chapter show that increased charge alone is insufficient to alter substantially the photocleavage properties of phi complexes of rhodium(III). This is most clearly demonstrated by the guanidylpropyl complex, $Rh(GPB)_2phi^{5+}$. This complex has a higher overall charge than $Rh(phen)_2phi^{3+}$, yet $Rh(phen)_2phi^{3+}$ shows slightly more intense photocleavage (Figure 4.4). The negligible effect that charge plays is not surprising since the photocleavage experiments are performed in 10 mM sodium cacodylate/40 mM NaCl buffer. Under these conditions, the DNA should be well solvated and the overall negative charge of the polymer will be screened by the sodium cations.

While charge by itself is insufficient to yield increased photocleavage, properly oriented charge can have an effect. This is evident in comparing the photocleavage intensity of the amido complex with that of the amino complex. Both molecules possess NH functionalities which can act as hydrogen bond donors, yet the amino complex shows substantially increased photocleavage relative to the amido and the guanidinium complexes (Figure 4.3 and 4.4) Thus, charge and functional group orientation must be considered together and not separately for assessing the effect that a recognition moiety has on the DNA photocleavage of a complex.

4.4.2. The Importance of the Spatial Orientation. The importance of proper spatial positioning of the recognition moiety is demonstrated by the differences in photocleavage which are seen for the different stereoisomers of the $Rh(bpy)_2phi^{3+}$ derivatives. In every case, the photocleavage pattern of isomer 1 compared with isomer 2 shows some differences which can be attributed to the position of the recognition orientation. These photocleavage results reveal that large changes in the

spatial orietantion of the recognition moieties can affect binding, but what about small changes?

The series of guanidinium complexes in which the linker arm is varied from methyl to ethyl to propyl demonstrates the effect of small changes in spatial positioning. The effect of increasing the length of the linker arm depends on the site being recognized. DNA recognition by the Λ-1-Rh(MGP)$_2$phi^{5+} is very sensitive to the orientation of the linker arm. Only Λ-1-Rh(MGP)$_2$phi^{5+} shows appreciable photocleavage at the site 5'-CATATG-3. In contrast, the recognition of 5'-CATCTG-3' by Δ-1-Rh(MGP)$_2$phi^{5+} is not as sensitive to the length of the linker arm. Δ-1-Rh(GEB)$_2$phi^{5+} still shows enhanced cleavage at 5'-CATCTG-3' relative to Rh(phen)$_2$phi^{3+}; however, the overall photocleavage intensity is about half of what is observed for Λ-1-Rh(MGP)$_2$phi^{5+} (See table 4.1). Increasing the linker arm to the propyl results in complete loss of enhancement. Thus, the spatial positioning of the recognition moiety is sensitive to small changes in the orientation.

4.4.3. Effect of Linker Arm Flexibility. The flexibility of the linker arm connecting the recognition moiety to the complex core plays an important role in determining the sequence selectivity of a metal complex. As the flexibility of the linker arm is increased, it becomes more difficult to predict the interactions that are taking place between the recognition moiety and the DNA. To complicate matters, all of the recognition moieties explored in this study are capable of interacting with both the DNA bases or the phosphate backbone. This is best exemplified in the photocleavage pattern observed for the amine complex. The amine complex shows a strong enhancement in DNA photocleavage relative to Rh(phen)$_2$phi^{3+} and 1-Rh(MGP)$_2$phi^{5+}. However, this photocleavage enhancement is very nonspecific in nature. This lack of specificity is most likely a result of a combination of things.

The long propyl arm is very flexible and the amino group could potentially form a hydrogen bond with every DNA base and the phosphate backbone as well.

Linker arm flexibility is most likely responsible for some of the differences observed between Δ-1-Rh(MGP)$_2$phi^{5+} and Δ-1-Rh(GEB)$_2$phi^{5+}. As seen in Figure 4.4, Δ-1-Rh(GEB)$_2$phi^{5+} targets some sites which are not recognized by Rh(phen)$_2$phi^{3+} or Δ-1-Rh(MGP)$_2$phi^{5+}. These sites are indicated by an arrow with an asterisk. The sequence of lower site is 5'-GAG*C*TG-3'. This sequence very closely resembles the best sites found for Δ-1-Rh(GEB)$_2$phi^{5+} in the oligonucleotide experiments (5'-GATCTC-3' and 5'-CATCTG-3'). The recognition of this site is most likely a result of the more flexible linker arm being able to compensate for the lack of the internal 5'-TC-3', which the more rigid Δ-1-Rh(MGP)$_2$phi^{5+} is unable to. The ability of Δ-1-Rh(GEB)$_2$phi^{5+} to recognize both 5'-GATCTC-3' and 5'-CATCTG-3' with almost equal affinity is also an indication of the flexibility of the ethyl linker. Δ-1-Rh(MGP)$_2$phi^{5+} shows a significant decrease in photocleavage upon changing the recognition sequence from 5'-CATCTG-3' to 5'-GATCTC-3' (see table 4.1).

As the linker arm flexibility increases, it should also be less favorable entropically for the recognition moiety to interact with the DNA. Thus complexes with flexible linker arms would be expected to have lower DNA binding affinities with all other things being equal. While none of the molecules used in this study are capable of directly testing this entropic effect, it is possible that one of the reasons that Rh(GPB)$_2$phi^{5+} shows little photocleavage enhancement is that it is entropically unfavorable for the recognition moiety to be constrained in such a manner that it can interact with the DNA. One possible solution to this problem is to use more rigid linker arms. This might allow for the recognition of larger DNA sites without the unwanted entropy penalties or the unpredictably of flexible linkers.

4.4.4. Functional Group Comparison. The intensity and specificity of DNA photocleavage for the series of metal complexes explored in this chapter is an indication of the ability of these different functional groups to alter the sequence specificity and binding affinity of phi complexes of rhodium(III). As demonstrated in chapter 2, the guanidylmethyl group is able to enhance significantly the binding specificity of $Rh(phen)_2phi^{3+}$ at low concentrations. The results in this chapter suggest that the amino group is able to enhance the binding affinity of metal complexes, but it lacks the sequence specificity of the guanidylmethyl moiety. As mentioned previously, this lack of specificity is partially the result of the flexibility of the linker arm. Future complexes should investigate the sequence specificity of amino complexes which have shorter linker arms and more rigid linker arms.

The amido complex, $1\text{-}Rh(AEB)_2phi^{3+}$, shows a slight enhancement in photocleavage intensity at low concentrations; however, this enhancement appears to be nonspecific in nature. Previous work in our laboratory has shown that the molecule $Rh(4,4'\text{-diamidobpy})_2phi^{3+}$ shows a significant increase in photocleavage relative to the parent complex (4). However, this increase was also nonspecific in nature. The lengthening of the linker arm in $1\text{-}Rh(AEB)_2phi^{3+}$ appears merely to reduce the effect of the amido moiety. It seems that the amido moiety is less well suited for enhancing the sequence specificity of phi complexes of rhodium(III) than either the amino or amido moiety.

4.5. Implications for Future Design. Similar to what has been seen with DNA binding proteins, the nature of the recognition moiety and its orientation with respect to the DNA is very important in determining the sequence specificity of phi complexes of rhodium(III). Designing small molecules capable of high DNA binding specificity and affinity requires proper placement of functional groups.

Molecules which use of flexible linker systems to compensate for "improper" placement of functional moieties are likely to suffer from low sequence specificity.

As illustrated in chapter 3, molecules which couple direct readout of the DNA sequence with a conformational change in the DNA offer huge potential for the design and synthesis of highly specific molecules. This finding is reinforced by our inability to mimic the recognition seen by Λ-1-Rh(MGP)$_2$phi^{5+} with complexes possessing the amido, amino, guanidylethyl and guanidylpropyl moiety. Future design should concentrate on incorporating the amino and guanidinium moiety in rigid molecules which require some DNA structural perturbation for proper alignment of functional groups.

4.5. References and Notes

1. Pabo, C.O. and Sauer, R.T., *Ann. Rev. Biochem.*, **61**, 1053 (1992)
2. Steitz, T.A., *Q. Rev. Biophys.*, **23**, 205 (1990)
3. Harrison, S.C. and Aggarwal, A.K., *Q. Rev. Biophys.*, **59**, 933 (1990)
4. Sitlani, A. and Barton, J.K., *Biochem.*, **33**, 12100 (1994)
5. Delleciana, L., Hamachi, I. and Meyer, T.J., *J. Am. Chem. Soc.*, **54**, 1731-1735 (1989)
6. Sardesai, N. Y., Zimmerman, K. and Barton, J. K., *J. Am Chem. Soc.*, **116**, 7502-7508 (1994)
7. Sitlani, A., Dupureur, C. M. and Barton, J. K., *J. Am. Chem. Soc.*, **115**, 12589-12590 (1993)

Chapter 5: Synthesis and Characterization of Functionalized Phi Complexes of Rhodium(III)

5.1. Introduction

The rigid geometries of transition metal complexes allow us to build well defined probes of DNA structure. Phenanthrenequinone diimine (phi) complexes of rhodium(III) occupy a unique niche in the rapidly expanding family of transition metal based DNA binding molecules. Phi complexes of rhodium(III) bind in the major groove of DNA via intercalation of the phi ligand into the DNA base stack. This intercalative mode of binding has served as a molecular platform in many elegant studies and these studies have helped develop our understanding of the principles of DNA recognition by small molecules (see chapter 1).

As the complexity of the phi complexes of rhodium(III) has increased, it has been necessary to devise new synthetic procedures for their synthesis. The guanidinium moiety of $[Rh(MGP)_2phi]Cl_5$ is incapable of surviving the harsh conditions of previously reported synthetic protocols (3,4). As a result, a new one-pot synthesis for phi complexes of rhodium(III) has been developed. This procedure uses mild reaction conditions which allows complexes possessing unprotected amino, amido and guanidyl functionalities to be synthesized. This procedure requires fewer steps and gives higher yields than previously reported methods.

The higher overall charge of the guanidinium and amino complexes, and the asymmetric nature of the ligands has also made it necessary to develop new methods for the purification and separation of the different stereoisomers. New and improved procedures for the separation of the enantiomers have also been devised. As a result, it is now possible to make, purify and resolve the enantiomers of many molecules which would have been unobtainable a few years ago.

In anticipation of the next generation of phi complexes of rhodium(III), procedures for the synthesis of complexes possessing three different ligands have also been devised. Previous methods for the synthesis of phi complexes of rhodium(III) which possess two different ancillary ligands gave a mixture of many different compounds (8). This procedure suffered from scrambling between the ancillary ligands and gave a mixture of $[RhX_2phi]Cl_3$, $[RhXYphi]Cl_3$ and $[RhY_2phi]Cl_3$ complexes (where X and Y are bidentate chelates). When the X and Y ligands are asymmetric in nature, an inseparable mixture of ten different compounds is obtained. A new synthetic protocol in which each ligand is added in a controlled fashion has been devised. This procedure has recently succeeded in the preparation the metal complex (4-methyl-1,10-phenanthroline)(4,4'-dimethyl-2,2'-bipyridine)(phenanthrenequinone diimine)rhodium(III)trichloride as a mixture of only two isomers. This new procedure should allow for the synthesis of new phi complexes of rhodium(III) which build upon the DNA recognition principles which have been illustrated in chapters 3, 4 and 5 of this thesis.

Described herein are the synthetic protocols for the metal complexes and ligands described in chapters 2, 3 and 4. Emphasized are the improvements made in synthesis and purification. Also described is the synthesis of the three ligand complex (4-methyl-1,10-phenanthroline)(4,4'-dimethyl-2,2'-bipyridine)(phenanthrenequinone diimine)rhodium(III)trichloride.

5.2. Experimental

5.2.1. Materials. All chemicals used were obtained from commercial sources. No further purification was performed before usage. 40 μm silica gel was used in column chromatography. All solvents were reagent grade or higher.

5.2.2. Instrumentation. 1H-NMR spectra were recorded on a 300 MHz GE QE Plus spectrometer. Ultraviolet-visible (UV-vis) spectra were recorded on a Cary 219 spectrophotometer (Varian). Circular dichroism studies were performed on a Jasco J-500a or J-600 spectrometer. High performance liquid chromatography (HPLC) was carried out on a Waters 600E system equipped with a Waters 484 tunable detector. A Vydac reverse phase protein and peptide C18 column was used for HPLC separations. Enantiomers were separated on a Chiralcel OD-R reverse phase HPLC column (Chiral technologies).

5.2.3. Synthesis of Ligands and Metal Complexes.

4-carboxaldehyde-1,10-phenanthroline (1). 4-methyl-1,10-phenanthroline (8.2 g, 2.2 mmole) was dissolved in 450.0 mL of dioxane/ 4% H_2O. Selenium oxide (18.7 g, 168.5 mmole) was added and a nitrogen atmosphere was maintained over the reaction mixture. After refluxing for 6 hours, the reaction was filtered hot through celite. The celite was washed with 100 mL of absolute ethanol and the ethanol wash was combined with the filtrate. The solvent was removed by rotary evaporation and the reddish oil was purified by chromatography (silica, 95/5 $CHCl_3$/Ethanol) to yield a yellowish solid. Yield = 74%.

^1H NMR ($CDCl_3$, ppm) 7.80 (1H, dd), 8.10 (2H, m), 8.40 (1H, d), 9.1 (1H, d), 9.3 (1H, d), 9.55 (1H, d), 10.65 (1H, s)

4-carboaldoxime-1,10-phenanthroline (2) The aldehyde (**1**) (6.0 g, 28.8 mmole) was dissolved in 100 mL of ethanol. Hydroxylamine hydrochloride (6.0 g, 86.3 mmole) and pyridine (6 mL) were added. The reaction was heated to reflux under an inert atmosphere for 6 hours. The resulting grayish precipitate was filtered off. The precipitate was dissolved in 100 mL of H_2O, filtered, and the solvent was removed under vacuum to yield a white solid. Yield = 65%.

^{1}H NMR (DMSO, ppm) 8.20 (1H, dd), 8.40 (2H, m), 8.85 (1H, d), 9.05 (1H, d), 9.15 (1H, s), 9.20 (1H, d), 9.30 (1H, d), 12.75 (1H, s); ^{13}C NMR (D_2O, ppm) 122.7, 124.0, 125.2, 125.8, 126.7, 128.6, 135.6, 137.0, 139.7, 143.6, 145.6, 145.9, 147.8

4-(aminomethyl)-1,10-phenanthroline (3) 4-carboaldoxime-1,10-phenanthroline(**2**) (3.0 g, 13.5 mmole) was dissolved in water (100 mL) containing 2% concentrated HCl. The reaction mixture was hydrogenated using an excess of hydrogen gas (1 atm.) over 500 mg of 10% palladium on charcoal. The progress of the reaction was monitored via an inline bubbler. Upon completion, the reaction was filtered through celite and the filtrate was concentrated to yield the amine as an off-white solid. Yield = 71%.

^{1}H NMR: (H_2O(pH 3), ppm) 4.75 (2H, s), 8.8 (1H, d), 7.95 (1H, d), 8.1 (2H, m), 8.8 (1H, d), 9.0 (1H, d), 9.1 (1H, d); ^{13}C NMR (D_2O, ppm): 40.0, 124, 124.5, 125, 126, 126.5, 129, 135, 137, 142, 144, 147, 151

4-guanidylmethyl-1,10-phenanthroline (4). 4-aminomethyl-1,10-phenanthroline(**3**) (0.5 g, 2.1 mmole) was dissolved in 100 mL of water. $ZnCl_2$ (285 mg, 2.1 mmol) was added and the pH of the reaction mixture was raised to pH 6.0 by addition of Na_2CO_3 (1.0 M). 2-methyl-2-thiopseudourea hydrogen sulfate (1.15 g, 4.14 mmole) was added and the reaction mixture was heated to reflux for 5 days under nitrogen. The reaction was purified by column chromatography (silica; 88% $CHCl_3$/10% ethanol/2% conc. HCl). Yield = 63%.

^{1}H NMR:(D_20, ppm) 4.5 (2H, s), 8.5 (2H, m), 8.0 (1H, d), 7.65 (1H, d), 7.5 (1H, m), 7.35 (1H, d), 7.15 (1H, d); ^{13}C NMR (D_2O, ppm) 41.0, 121.6, 122.1, 125.8, 126.3, 127.7, 128.4, 135.2, 136.2, 140.4, 145.4, 148.2, 154.1, 161.4

Lithiation of 4,4'-dimethyl-2,2'-bipyridine (11). All glassware was dried for several hours at 110 ^{o}C , assembled hot, and flushed thoroughly with argon. The bipyridine was dried for several hours under vacuum. A 500 mL three-necked

flask was fitted with two rubber septa and a 250 mL dropping funnel also sealed with a rubber septum. Dry THF (Aldrich, 10 mL) and diisopropylamine (7 mL, 50 mmole) were added by syringe. The flask was cooled to -78 ^{o}C with an acetone-dry ice slurry. Butyllithium (2.5 M in hexanes, 20 mL, 50 mmole) was added by syringe and the mixture was stirred for 15 minutes. 4,4'-dimethyl-2,2'-bipyridine (10.0 g, 54.2 mmole) dissolved in 240 mL of dry THF was quickly added to the dropping funnel against a strong flow of argon and added dropwise over 10 minutes. The dark brown (almost bluish) mixture was stirred for two hours.

4-(2-Hydroxyethyl)-4'-methyl-2,2'-bipyridine (5). Paraformaldehyde (3.0 g , 99.0 mmole) was dried under vacuum for 3 hours and added to the lithiated bipyridine (**11**) against a strong flow of argon. After one hour, the low-temperature bath was removed and the reaction was stirred overnight under a slow flow of argon. The reaction turned bright yellow. Brine (120 mL) was added and the reaction was extracted with CH_2Cl_2 (2 x150 mL). The CH_2Cl_2 was dried over anhydrous sodium sulfate and the solvent was removed to yield an oily mixture. The reaction was purified by column chromatography (silica; 19:1 CH_2Cl_2:Methanol). The first spot (Rf=0.70) is 4,4'-dimethyl-2,2'-bipyridine which can be recrystallized from ethyl acetate. The monoalcohol elutes at Rf=0.45 and the dialcohol at Rf=0.26. 4-(2-Hydroxyethyl)-4'-methyl-2,2'-bipyridine (5) was recovered as a yellow oil. Yield = 42.5%.

^{1}H NMR ($CDCl_3$, ppm) 2.45 (1H, s), 2.96 (2H, t), 3.98 (2H, t), 7.13 (1H, d), 7.19 (1H, d), 8.20 (1H, s), 8.23 (1H, s), 8.49 (1H, d), 8.56 (1H, d); ^{13}C NMR ($CDCl_3$, ppm) 21.60 , 39.02 , 62.55, 122.18 , 122.65, 124.98, 125.15, 148.80 , 149.13, 149.45, 149.70, 156.20, 156.40.

4-(2-Bromoethyl)-4'-methyl-2,2'-bipyridine · 2HBr (6). 4-(2-Hydroxyethyl)-4'-methyl-2,2'-bipyridine(**5**) (4.5 g, 21.3 mmole) was refluxed in a

48% solution of HBr (60 mL) for 8 hours. The solvent was removed under vacuum and the resulting dark brown powder was dissolved in 200 mL of ethanol with warming. The product was precipitated by addition of ether (500 mL). The brown sticky solid was filtered and washed with ether. Yield = 69.5 %.

^{1}H NMR (D_2O, ppm) 2.4 (3H, s), 3.2 (2H, t), 3.65 (2H, t), 7.45 (1H, d), 7.6 (1H, d), 8.0 (1H, s), 8.15 (1H, s), 8.45 (1H, d), 8.55 (1H, d); ^{13}C NMR (D_2O, ppm) 22, 32, 37, 123, 125, 127, 127.2, 143, 145, 145.1, 147, 154, 159

4-(2-Phthalimidoethyl)-4'-methyl-2,2'-bipyridine (7). 4-(2-Bromoethyl)-4'-methyl-2,2'-bipyridine· 2 HBr (**6**) (4.5 g, 10.4 mmole) was dissolved in 0.2 M NaOH (350 mL). The solution was extracted with $CHCl_3$ (2 x 350 mL). The extracts were washed with water (350 mL) and dried over anhydrous Na_2SO_4. Removal of the solvent under vacuum was stopped before completion to avoid decomposition. DMF (110 mL) and potassium phthalimide (3.2 g, 17.2 mmole) were added. The mixture was heated at 50 ^{o}C for 12 hours. The reaction was cooled and 400 mL of water was added. The solution was extracted with $CHCl_3$ (2 x 400 mL), washed with 0.2 M NaOH (400 mL) and water (400 mL), and dried over anhydrous Na_2SO_4. The mixture was purified by chromatography (silica). The first product which was eluted off with 9:1 toluene: ethyl acetate (Rf = 0.47) was an elimination product. The solvent was changed to 2:1 toluene: ethyl acetate and the desired product was collected as a pure white powder. Yield = 11.0 %.

^{1}H NMR ($CDCl_3$) 2.45 (3H, s), 3.1 (2H, t), 4.05 (2H, t), 7.2 (1H, d), 7.3 (1H, d), 7.8 (2H, m), 7.9 (2H, m), 8.30 (1H, s), 8.4 (1H, s), 8.6 (1H, d), 8.65 (1H, d)

This major product is the elimination product, 4-vinyl-4'-methyl-2,2'-bipyridine, and is produced in 40.5% yield.

^{1}H NMR ($CDCl_3$, ppm) 2.5 (3H, s), 5.55 (1H, d), 6.05 (1H, d), 6.8 (1H, dd), 7.15 (1H, d), 7.3 (1H, d), 8.25 (1H, s), 8.4 (1H, s), 8.55 (1H, d), 8.65 (1H, d) ^{13}C

NMR ($CDCl_3$, ppm) 21.6, 34.5, 38.5, 122, 122.4, 123.7, 124.4, 125.1, 132.3, 134.4, 148.3, 148.5, 149.3, 149.7, 156.0, 168.4

4-(2-aminoethyl)-4'-methyl-2,2'-bipyridine (8). 4-(2-Phthalimidoethyl)-4'-methyl-2,2'-bipyridine(**7**) (1.0 g, 2.86 mmole) was dissolved in 40 mL of ethanol with heating. Hydrazine hydrate (0.210 mL, 4.29 mmole) was added and the mixture was refluxed for 6 hours. The reaction was poured into brine (100 mL) and the pH was raised to 13 by addition of 50% NaOH. The solution was extracted with $CHCl_3$ (2 x 150 mL) and dried over anhydrous Na_2SO_4. The solvent was removed to yield a clear oil.

^1H NMR ($CDCl_3$, ppm) 2.44 (3H, s), 2.80 (2H, t), 3.10 (2H, t), 7.2 (2H, m), 8.23 (1H, s), 8.24 (1H, s), 8.56 (1H, d), 8.6 (1H, d); ^{13}C NMR ($CDCL_3$, ppm): 21.6, 38.8, 42.5, 122, 122.5, 124.7, 125.2, 149.3, 149.6, 149.8, 152, 156, 156.2

4-(2-guanidylethyl)-4'-methyl-2,2'-bipyridine (9). 4-(2-aminoethyl)-4'-methyl-2,2'-bipyridine(**8**) (0.6 g, 1.70 mmole) was dissolved in 50 mL of water and zinc (II) acetate (0.37 g, 1.70 mmole) was added. 2-methyl-2-thiopseudourea hydrogen sulfate (4.1 g, 14.3 mmole) was added and the pH of the solution was adjusted to 7.5 by the addition of Na_2CO_3 (1.0 M). The reaction was heated to reflux under a nitrogen atmosphere for 5 days. The reaction was purified by column chromatography (silica). The unreacted 2-methyl-2-thiopseudourea was eluted with 78% chloroform/18% ethanol/ 4% conc. HCl and the solvent was switched to 78% chloroform/18% methanol/ 4% conc. HCl to elute the guanidylethylbipyridine ligand. This ligand is unstable when stored dry. Therefore, the ligand was immediately dissolved in water (pH 2).

^1H NMR (D_2O, ppm) 2.45 (3H, s), 2.95 (2H, t), 3.35 (2H, t), 7.55 (1H, d), 7.65 (1H, d), 8.05 (1H, s), 8.15 (1H, s), 8.45 (1H, d), 8.55 (1H, d); ^{13}C NMR (D_2O, ppm) 22.0, 36.0, 40.1, 123, 123.5, 128, 129, 144, 147, 149, 154, 157, 159, 163

$C_{14}H_{18}N_5$ calc. 256.33 Found m/z 256 (M+)

4-[2-(1,3-dioxolan-2-yl)ethyl]-4'-methyl-2,2'-bipyridine (10). 2-Bromomethyl-1,3-dioxolane (8.01 mL, 12.93 g., 77.4 mmole) was added to the lithiated bipyridine (**11**) with a syringe. After one hour the low-temperature bath was removed and the reaction was stirred overnight still under a slow flow of argon. After the reaction had turned bright yellow, brine (120 mL) was added and the mixture extracted with CH_2Cl_2. The combined extracts were dried over anhydrous Na_2SO_4 and concentrated under vacuum to yield yellow oil.

4-(2-Formylethyl)-4'-methyl-2,2'-bipyridine (12). The crude acetal mixture (**10**) was dissolved in 50 mL of 1 M HCl and heated for 4 hours at 50 °C. After cooling, the solution was neutralized with saturated aqueous $NaHCO_3$ and extracted with $CHCl_3$ (3 x 150 mL). The extracts were dried over anhydrous Na_2SO_4 and concentrated under reduced pressure.

4-(2-carboxylethyl)-4'-methyl-2,2'-bipyridine (13). The crude 4-(2-formylethyl)-4'-methyl-2,2'-bipyridine (**12**) was dissolved in 100 mL of acetone. An excess of $KMnO_4$ (7.0 g, 44.3 mmole) was added in 3 portions. The bright purple solution was stirred at 23 °C for 12 hours. The acetone was removed under vacuum and the residue was dissolved in 100 mL H_2O and heated at 95 °C for 2 hours. The reaction was cooled to 4 °C overnight and the MnO_2 was removed by filtering twice through celite. The filtrate was acidified to pH 4.8 using 1M HCl and the solvent was removed under vacuum. The product was purified by column chromatography (silica: 1:1 toluene: ethyl acetate).

4-(2-amidoethyl)-4'-methyl-2,2'-bipyridine (14). 4-(2-carboxyethyl)-4'-methyl-2,2'-bipyridine (**13**) (542 mg, 2.27 mmole) was dissolved in 50 mL of methanol and 2 mL of concentrated H_2SO_4 and heated to reflux for 5 hours. The reaction was cooled to room temperature and neutralized with saturated Na_2CO_3 solution. The solvent was removed under vacuum and resuspended in 100 mL 50/50 CH_2Cl_2/H_2O. The mixture was extracted with CH_2Cl_2 (2 x 50 mL) and

dried over anhydrous Na_2SO_4. The solvent was removed under vacuum to yield the methyl ester. The solid was redissolved in 25 mL of methanol saturated with ammonia and stirred in a sealed flask at 55 °C for 12 hours. The product was purified on silica (20:1 CH_2Cl_2/methanol). Yield = 71%.

^1H NMR ($CDCl_3$, ppm) 2.4 (3H, s), 2.6 (2H, t), 3.0 (2H, t), 5.5 (2H, m), 7.1 (1H, d), 7.15 (1H, d), 8.2 (1H, s), 8.22 (1H, s), 8.5 (1H, d), 8.55 (1H, d)

4-(3-Hydroxypropyl)-4'-methyl-2,2'-bipyridine (15). The crude 4-(2-formylethyl)-4'-methyl-2,2'-bipyridine (**12**) was dissolved in 50 mL of ethanol. Sodium borohydride (0.79 g., 20.8 mmole) was added and the reaction was stirred for 30 minutes. Brine was added (120 mL) and the mixture was extracted with CH_2Cl_2. The extracts were dried over anhydrous Na_2SO_4 and the CH_2Cl_2 was removed under vacuum. The crude reaction mixture was purified by column chromatography (silica; ethyl acetate). The first spot (Rf=0.53) is 4,4'-dimethyl-2,2'-bipyridine which can be recrystallized from ethyl acetate. The monoalcohol eltues at Rf=0.25 and the dialcohol at Rf=0.05. 4-(3-Hydroxypropyl)-4'-methyl-2,2'-bipyridine (**15**) was recovered as a clear oil. Yield = 35%.

^1H NMR ($CDCl_3$, ppm) 1.95 (2H, m), 2.43 (3H, s), 2.80 (2H, t), 3.67 (2H, t), 7.1 (2H, m), 8.2 (2H, m), 8.51 (1H, d), 8.55 (1H, d); ^{13}C NMR ($CDCl_3$, ppm): 21.62, 31.98, 33.43, 61.96, 121.84, 122.57, 124.41, 125.13, 148.71, 149.24, 149.49, 152.45, 156.35, 156.49.

4-(3-Bromopropyl)-4'-methyl-2,2'-bipyridine · 2HBr (16). 4-(3-Hydroxypropyl)-4'-methyl-2,2'-bipyridine (**15**) (4.25 g, 18.6 mmole) was refluxed in a 48% solution of HBr (60 mL) for 8 hours. The solvent was removed and the resulting light brown powder was dissolved in 200 mL of ethanol with warming. The product was precipitated by addition of ether (500 mL). The white powder was filtered and washed with ether. Yield = 88.5%.

^{1}H NMR (D_2O, ppm) 2.2 (1H, m), 2.56 (3H, s), 2.94 (1H, t), 3.42 (2H, t), 7.64 (1H, d), 7.69 (1H, d), 8.12 (1H, s), 8.17 (1H, s), 8.55 (1H, d), 8.58 (1H, d); ^{13}C NMR (D_2O, ppm): 21.87, 32.40, 33.61, 33.83, 124.35, 125.24, 127.78, 128.30, 144.16, 146.59, 146.69, 146.73, 158.56, 158.95.

4-(3-Phthalimidopropyl)-4'-methyl-2,2'-bipyridine (17). 4-(3-Bromopropyl)-4'-methyl-2,2'-bipyridine· 2HBr (**16**) (6.87 g, 15.2 mmole) was dissolved in 0.2 M NaOH (350 mL). This solution was extracted with $CHCl_3$ (2 x 350 mL). The $CHCl_3$ extracts were washed with water (350 mL) and dried over anhydrous Na_2SO_4. The removal of the solvent was stopped before completion, in order to prevent decomposition. DMF (110 mL) and potassium phthalimide (3.2 g, 17.2 mmole) were added. The mixture was heated at 50 ^oC for 12 hours. The reaction was cooled and 400 mL of water was added. The solution was extracted with $CHCl_3$ (2 x 400 mL), washed with 0.2 M NaOH (400 mL) and water (400 mL), and dried over anhydrous Na_2SO_4. The mixture was purified by chromatography (silica, 2:1 toluene: ethyl acetate). Yield = 70.5%.
^{1}H NMR ($CDCl_3$, ppm) 2.09 (2H, m), 2.44 (3H, s), 2.80 (2H, t), 3.79 (2H, t), 7.12 (1H, d), 7.17 (1H, d), 7.7 (2H, m), 7.85 (2H, m), 8.18 (1H, s), 8.22 (1H, s), 8.53 (2H, m); ^{13}C NMR ($CDCl_3$, ppm): 21.59, 29.37, 33.15, 38.00, 121.52, 122.39, 123.63, 124.12, 125.07, 132.39, 134.35, 148, 149.33, 149.54, 152, 157, 169.

4-(3-aminopropyl)-4'-methyl-2,2'-bipyridine (18). 4-(3-Phthalimidopropyl)-4'-methyl-2,2'-bipyridine (**17**) (1.0 g, 2.80 mmole) was dissolved in 40 mL of ethanol with heating. Hydrazine hydrate (0.210 mL, 4.29 mmole) was added and the mixture was refluxed for 6 hours. The reaction was poured into brine (100 mL) and the pH was raised to 13 by addition of 50% NaOH. The solution was extracted with $CHCl_3$ (2 x 150 mL) and dried over anhydrous Na_2SO_4. The solvent was removed yielding a clear oil.

^{1}H NMR ($CDCl_3$, ppm) 1.85 (2H, m), 2.44 (3H, s), 2.75 (4H, m), 7.15 (2H, m), 8.23 (2H, s), 8.56 (2H, m); ^{13}C NMR ($CDCl_3$, ppm): 21.55, 33.14, 34.48, 41.98, 121.58, 122.38, 124.23, 125.04, 148.50, 149.27, 149.42, 152.51, 156.30, 156.51.

4-[3-guanidylpropyl]-4'-methyl-2,2'-bipyridine (19). 4-(3-aminopropyl)-4'-methyl-2,2'-bipyridine (**18**) (630 mg, 2.80 mmole) was dissolved in 50 mL 80/20 H_2O/DMSO and 2-methyl-2-thiopseudourea hydrogen sulfate (3.2 g, 11.2 mmole) and zinc acetate (630 mg, 2.80 mmole) were added. The pH of the solution was adjusted to 7.5 by addition of Na_2CO_3 (1.0 M) and the reaction was heated at reflux under a nitrogen atmosphere for 5 days. The reaction was purified by flash column chromatography. The unreacted 2-methyl-2-thiopseudourea was eluted with 78% chloroform/18% ethanol/ 4% conc. HCl and then the solvent was switched to 78% chloroform/18% methanol/ 4% conc. HCl to elute the guanidylpropyl ligand. This ligand is unstable when stored dry. Therefore, it was immediately dissolved in water (pH 2). Yield = 28%.

^{1}H NMR (D_2O (pH 3), ppm) 1.75 (2H, t), 2.45 (3H, s), 2.75 (2H, t), 2.95 (2H, t), 7.8 (2H, m), 8.1 (2H, m), 8.55 (1H, d), 8.57 (1H, d) $C_{15}H_{20}N_5$ calc. 270.33 m/z 270 (M+)

Bis(4-guanidylmethyl-1,10-phenanthroline)(phenanthrenequinone diimine)rhodium(III) pentachloride, [Rh(MGP)$_2$phi]Cl$_5$ (20). $Rh(NO_3)_3.xH_2O$ (30 % Rh, 100 mg, 0.294 mmole) and hydrazine hydrochloride (5 mg) were dissolved in 20 mL 50/50 DMSO/H2O. The reaction was degassed and placed under an N_2 atmosphere. 9,10-diaminophenanthrene (65 mg, 0.312 mmole) was added against a strong flow of nitrogen, and the reaction was heated to reflux for 2 hours under nitrogen. 4-guanidylmethyl-1,10-phenanthroline (**4**) (168 mg, 0.588 mmole) dissolved in 10 mL 50/50 DMSO/H_2O (pH 7) was added via syringe. The reaction was stirred at reflux for 12 hours, and then the pH was adjusted to pH =2 by addition of HCl (1.0 M). H_2O (50 mL) was added and the

reaction was stirred open to the air for 24 hours. The reaction was filtered and the dark orange supernatant was purified by cation exchange chromatography (Sephadex SP C-25, 0.3 M $MgCl_2$) to yield $[Rh(MGP)_2phi]Cl_5$ as a mixture of 3 isomers. Yield = 13.5%. The three stereoisomers were present a 1:2:1 ratio. The three isomers were separated by reverse-phase HPLC (isocratic 86/14 H_2O w/1% trifluoroacetic acid/acetonitrile). The first isomer to elute is the symmetric isomer with both arms over the phi ligand, 1-$[Rh(MGP)_2phi]Cl_5$. The second isomer to elute is the asymmetrical isomer, 3-$[Rh(MGP)_2phi]Cl_5$. The third isomer to elute is the symmetric isomer with both arms away from the phi ligand, 2-$[Rh(MGP)_2phi]Cl_5$. The enantiomers of each stereoisomer were separated by cation-exchange chromatography (Sephadex SP C-25/potassium antimonyl tartrate (0.15M)). The isomers were loaded individually onto a resin (24 inches high) preequilibrated with 0.15M potassium antimonyl tartrate. Over a 12 hour period, enantiomer separation was obtained. The Λ-enantiomer of each isomer eluted first.

Alternatively, the stereo-isomers and enantiomers can be separated simultaneously on a large cation exchange column (Sephadex SP C-25/potassium antimonyl tartrate (0.15M)). About 30 mg of filtered complex (determined by UV-vis) is loaded onto a 36 inch high and 2 inch in diameter column of preequilibrated resin. Over a 12 hour period, the different isomers are allowed to separate on the resin. Immediately prior to elution of the first orange band, the column is stopped and the resin is removed from the column. The orange bands are removed (4-6 bands depending on separation) and the enantiomers are eluted from the resin slices with 0.5 M $MgCl_2$. Very good separation of the Δ- and Λ-enantiomers of each isomer are obtained. It is sometimes necessary to further separate the stereo-isomers by HPLC due to overlap. The enantiomers elute in the following order: 1-Λ, 2-Λ, 3-Λ, 1-Δ, 2-Δ, 3-Δ.

1-[Rh(MGP)$_2$phi]Cl$_5$.- ^{1}H NMR(D$_2$O (pH 3), ppm) 5.05 (4H, dd), 7.4 (2H, t), 7.65 (4H, m), 7.85 (2H, d), 7.97 (2H, d), 8.03 (2H, d), 8.15 (2H, d), 8.35 (4H, dd), 8.75 (2H, d), 8.85 (2H, d) RhC$_{42}$H$_{38}$N$_{12}$ calc. 813.74; Found FABMS m/z 813 (M^+)

UV-visible spectrum pH 5, H$_2$O-Isobestic point at 366. Maximum at 380 and 272 (See appendix).

CD (pH 5)- (Δ)peaks at 279 (-) $\Delta\varepsilon_{281} = -2.7 \times 10^2\ M^{-1}\ cm^{-1}$, 259 (+) and 210 (-); (Λ)-peaks at 279 (+) $\Delta\varepsilon_{281} = +2.7 \times 10^2\ M^{-1}\ cm^{-1}$, 259 (-), 210 (+) (See appendix).

2-[Rh(MGP)$_2$phi]Cl$_5$- ^{1}H NMR (D$_2$O (pH 3), ppm) 4.95 (4H, dd), 7.4 (2H, t), 7.6 (2H, d), 7.7 (2H, t), 7.85 (2H, d), 8.05 (4H, m), 8.15 (2H, d), 8.35 (4H, dd), 8.95 (2H, d), 9.0 (2H, d) RhC$_{42}$H$_{38}$N$_{12}$ calc. 813.74; Found FABMS m/z 813 (M^+)

UV-visible spectrum pH 5, H$_2$O-Isobestic point at 366. Maximum at 380 and 272 (See appendix).

CD (pH 5)- (Δ)peaks at 279 (-), 259 (+) and 210 (-); (Λ)-peaks at 279 (+), 259 (-), 210 (+) (See appendix).

3-[Rh(MGP)$_2$phi]Cl$_5$- ^{1}H NMR (D$_2$O (pH 3), ppm) 4.95 (2H, dd), 5.05 (2H, dd), 7.4 (2H, d), 7.55 (1H, d), 7.65 (3H, m), 7.75 (1H, d), 7.82 (1H, d), 7.95 (1H, d), 8.0 (3H, m), 8.1 (2H, d), 8.3 (4H, m), 8.75 (1H, d), 8.82 (1H, d), 8.9 (1H, d), 9.0 (1H, d) Rh C42H$_{38}$N$_{12}$ calc. 813.74; Found FABMS m/z 813 (M^+)

UV-visible spectrum pH 5, H$_2$O-Isobestic point at 366. Maximum at 380 and 272 (See appendix).

CD (pH 5)- (Δ)peaks at 279 (-), 259 (+) and 210 (-); (Λ)-peaks at 279 (+), 259 (-), 210 (+) (See appendix).

Bis(4-(2-amidoethyl)-4'-methyl-2,2'-bipyridine)(phenanthrenequinone diimine)rhodium(III) trichloride, [Rh(AEB)$_2$phi]Cl$_3$ (21). RhCl$_3$.xH$_2$O

(25mg, 0.199 mmole) was dissolved in 5 mL of H_2O and placed under an N_2 atmosphere. 9,10-diaminophenanthrene (25 mg, 0.12 mmole) was added against a strong N_2 flow and the reaction was heated to reflux for 2 hours. 4-(2-amidoethyl)-4'-methyl-2,2'-bipyridine (**14**) (59.0 mg, 0.245 mmole) was dissolved in 15 mL of DMF and added via syringe to the rhodium reaction. The reaction was stirred at reflux for 4 hours. The pH of the reaction was adjusted to 1 with addition of HCl (1.0 M) and the reaction was added to 50 mL of H_20. The reaction was air oxidized by stirring open to the air for 24 hours. The reaction was filtered and purified by HPLC. Yield = 40%. The stereo-isomers were separated by reverse-phase HPLC (isocratic 84/16 H_2O w/1% trifluoroacetic acid/acetonitrile). The isomers eluted in the same order as for $[Rh(MGP)_2phi]Cl_5$(**20**). $RhC_{42}H_{40}N_8O_2$ calc. 791.7; Found ESI m/z 789.5 (minus 2 H)

1-$[Rh(AEB)_2phi]Cl_3$- ^1H NMR (D_2O (pH 3), ppm) 2.45 (6H, s), 2.65 (4H, t), 3.1 (4H, t), 7.3 (4H, s), 7.4 (2H, t), 7.55 (2H, d), 7.7 (2H, t), 8.1 (4H, t), 8.15 (2H, d), 8.4 (4H, d)

UV-visible spectrum pH 5, H_2O- peaks at 370, 311, 300 (see appendix).

2-$[Rh(AEB)_2phi]Cl_3$- ^1H NMR (D_2O (pH 3), ppm) 2.5 (6H, s), 2.6 (4H, t), 3.05 (4H, t), 7.4 (6H, m), 7.5 (2H, d), 7.7 (2H, t), 8.1 (4H, t), 8.15 (2H, d), 8.4 (4H, d)

UV-visible spectrum pH 5, H_2O- peaks at 370, 311, 300 (see appendix).

3-$[Rh(AEB)_2phi]Cl_3$-^1H NMR (D_2O (pH 3), ppm) 2.4 (3H, s), 2.5 (3H, s), 2.6 (4H, m), 3.0 (2H, t), 3.1 (2H, t), 7.3 (2H, s), 7.4 (4H, t), 7.5 (1H, d), 7.55 (1H, s), 7.7 (2H, t), 8.1 (2H, s), 8.13 (2H, s), 8.17 (1H, d), 8.2 (1H, d), 8.35 (2H, s), 8.45 (2H, s)

UV-visible spectrum pH 5, H_2O- peaks at 370, 311, 300 (see appendix).

Bis(4-(3-aminopropyl)-4'-methyl-2,2'-bipyridine)(phenanthrenequinone diimine)rhodium(III) pentachloride, $[Rh(APB)_2phi]Cl_5$ (22). The same synthethic procedure as for $Rh(AEB)_2phi^{3+}$ (**21**) was used but 4-(3-aminopropyl)-

4'-methyl-2,2'-bipyridine (**18**) was substituted as the bipyridine ligand. Yield = 25%. The stereo-isomers were separated by reverse-phase HPLC (isocratic 85/15 H_2O w/1% trifluoroacetic acid/acetonitrile). The isomers eluted in the same order as for $[Rh(MGP)_2phi]Cl_5$(**20**). $RhC_{42}H_{44}N_8$ calc. 763.7 ; Found ESI m/z 761.5 (minus 2 H)

1-$Rh(APB)_2phi^{5+}$-^1H NMR (D_2O (pH 3), ppm) 2.0 (4H, m), 2.45 (6H, s), 2.95 (8H, m), 7.3 (4H, m), 7.4 (2H, t), 7.6 (2H, d), 7.7 (2H, t), 8.1 (4H, t), 8.2 (2H, d), 8.4 (2H, s), 8.5 (2H, s)

UV-visible spectrum pH 5, H_2O- peaks at 370, 311, 300 (see appendix).

2-$Rh(APB)_2phi^{5+}$-^1H NMR (D_2O (pH 3), ppm) 1.95 (4H, m), 2.5 (6H, s), 2.8 (4H, t), 2.9 (4H, t), 7.3 (4H, m), 7.4 (2H, t), 7.5 (2H, d), 7.65 (2H, t), 8.1 (4H, m), 8.2 (2H, d), 8.4 (2H, s), 8.45 (2H, s)

UV-visible spectrum pH 5, H_2O- peaks at 370, 311, 300 (see appendix).

3-$Rh(APB)_2phi^{5+}$-^1H NMR (D_2O (pH 3), ppm) 1.95 (4H, m), 2.45 (3H, s), 2.5 (3H, s), 2.9-3.0 (6H, m), 7.25-7.45 (6H, m), 7.5 (1H, d), 7.6 (1H, d), 7.7 (2H, t), 8.1 (2H, m), 8.2 (2H, m), 8.25 (2H, d), 8.4 (2H, s), 8.45 (2H, s)

UV-visible spectrum pH 5, H_2O- peaks at 370, 311, 300 (see appendix).

Bis(4-(2-guanidylethyl)-4'-methyl-2,2'-bipyridine)(phenanthrenequinone diimine)rhodium(III) pentachloride, $[Rh(GEB)_2phi]Cl_5$ (23)-The same synthethic procedure as for $Rh(AEB)_2phi^{3+}$ (**21**) was used but 4-(2-guanidylethyl)-4'-methyl-2,2'-bipyridine (**9**) was substituted as the bipyridine ligand and DMSO was used as a solvent instead of DMF. Yield = 20%. The stereoisomers and the enantiomers were separated simultaneously using cation exchange chromatography (Sephadex SP C-25/potassium antimonyl tartrate (0.15M)) The separation was performed using the same procedure as was used for $[Rh(MGP)_2phi]Cl_5$(**20**). The enantiomers eluted in the same order. Each enantiomer was further purified by reverse-phase HPLC

(isocratic 84/16 H_2O w/1% trifluoroacetic acid/acetonitrile). The isomers eluted in the same order as for [Rh(MGP)$_2$phi]Cl$_5$(**20**). $RhC_{42}H_{44}N_{12}$ calc. 819.7; Found ESI m/z 819.5

1-Rh(GEB)$_2$phi^{5+}- ^{1}H NMR (D_2O (pH 3), ppm) 2.4 (3H, s), 2.9 (4H, t), 3.35 (4H, t), 7.3 (4H, s), 7.4 (2H, t), 7.55 (2H, d), 7.7 (2H, t), 8.1 (4H, t), 8.2 (2H, d), 8.35 (2H, s), 8.4 (2H, s)

UV-visible spectrum pH 5, H_2O- peaks at 370, 311, 300 (see appendix).

CD (pH 5)- (Δ)peaks at 317 (-) $\Delta\varepsilon_{281}$ = -55 M^{-1} cm^{-1}, 306 (-), 253 (+), 229 (-); (Λ)-peaks at 317 (+) $\Delta\varepsilon_{281}$ = +51 M^{-1} cm^{-1}, 306 (+), 253 (-), 229 (+) (See appendix).

2-Rh(GEB)$_2$phi^{5+}- ^{1}H NMR (D_2O (pH 3), ppm) 2.5 (6H, s), 2.85 (4H, t), 3.3 (4H, t), 7.4 (6H, m), 7.55 (2H, d), 7.7 (3H, t), 8.1 (4H, t), 8.15 (2H, d), 8.4 (4H, d)

UV-visible spectrum pH 5, H_2O- peaks at 370, 311, 300 (see appendix).

3-Rh(GEB)$_2$phi^{5+}- ^{1}H NMR (D_2O (pH 3), ppm) 2.4 (3H, s), 2.5 (3H, s), 2.85 (2H, t), 2.95 (2H, t), 3.35 (4H, m), 7.3 (4H, m), 7.4 (2H, t), 7.5 (1H, d), 7.55 (1H, d), 7.7 (2H, t), 8.1 (4H, m), 8.17 (1H, d), 8.22 (1H, d), 8.35 (2H, d), 8.45 (2H, d)

UV-visible spectrum pH 5, H_2O- peaks at 370, 311, 300 (see appendix).

Bis(4-(3-guanidylpropyl)-4'-methyl-2,2'-bipyridine)(phenanthrenequinone diimine)rhodium(III) pentachloride, [Rh(GPB)$_2$phi]Cl$_5$ (24)- The same synthethic procedure as for Rh(AEB)$_2$phi^{3+} (**21**) was used but 4-(3-guanidylpropyl)-4'-methyl-2,2'-bipyridine (**19**) was substituted as the bipyridine ligand and DMSO was used as a solvent instead of DMF. Yield = 18%. The three stereoisomers were separated by reverse-phase HPLC (isocratic 84/16 H_2O w/1% trifluoroacetic acid/acetonitrile). The isomers eluted in the same order as for [Rh(MGP)$_2$phi]Cl$_5$(**20**). $RhC_{44}H_{48}N_{12}$ calc. 847.8; Found ESI m/z 847.6

1-$Rh(GPB)_2phi^{5+}$-^1H NMR (D_2O (pH 3), ppm) 1.8 (4H, m), 2.4 (6H, s), 2.8 (4H, t), 3.0 (4H, t), 7.3 (4H, s), 7.4 (2H, t), 7.55 (2H, d), 7.7 (2H, t), 8.1 (4H, t), 8.2 (2H, d), 8.35 (2H, s), 8.4 (2H, s)

UV-visible spectrum pH 5, H_2O- peaks at 370, 311, 300 (see appendix).

2-$Rh(GPB)_2phi^{5+}$-^1H NMR (D_2O (pH 3), ppm) 1.75 (4H, m), 2.5 (6H, s), 2.75 (4H, t), 2.95 (4H, t), 7.3-7.4 (6H, m), 7.5 (2H, d), 7.7 (2H, t), 8.1-8.2 (6H, m), 8.35 (2H, s), 8.4 (2H, s)

UV-visible spectrum pH 5, H_2O- peaks at 370, 311, 300 (see appendix).

Hydroxonium tetrachloro- (4-methyl-1,10-phenanthroline)rhodate(III)-H[Rh(mephen)Cl_4] (25). $RhCl_3 \cdot 3\ H_2O$ (670 mg, 3.2 mmole) and 4-methyl-1,10-phenanthroline (1.86 g, 9.6 mmole) are dissolved in 30 mL of 2.0 M HCl. The reaction was heated to reflux for 1 hour and the yellowish 4-methyl-1,10-phenanthrolinium[Rh(4-methyl-1,10-phenanthroline)Cl_4] was filtered off. This solid was washed with 20 mL of $CHCl_3$ and dried. All of the solid were dissolved in 100 mL of 12 M HCl and heated to reflux for 1 hour. The reaction was filtered hot and the filtrate was saved. The filtered off precipitate was redissolved in 100 mL concentrated HCl and heated to reflux for 1 hour. The reaction was filtered and the filtrates were combined. The solvent was removed under vacuum. The resulting solid was dissolved in 50/50 ethanol/H_2O and loaded onto an anion exchange column. The product was eluted with 50/50 0.4 M HCl/ethanol. Yield = 41%.

^1H NMR (DMSO, ppm) 3.0 (3H, s). 8.1 (1H,d), 8.2 (1H, m), 8.4 (2H, dd), 8.9 (1H, d), 9.9 (1H, d), 10.0 (1H, d)

(4-methyl-1,10-phenanthroline)(4,4'-dimethyl-2,2'-bipyridine)(phenanthrenequinone diimine)rhodium(III) trichloride-[Rh(mephen)(DMB)(phi)]Cl_3. H[Rh(4-methyl-1,10-phenanthroline)Cl_4] (50 mg, 0.114mmol) and 4,4'-dimethyl-2,2'-bipyridine (21 mg, 0.125 mmole) were

dissolved in 25 mL of 50/50 DMSO/H_2O. The pH of the reaction was adjusted to @7 with the addition of 1.0 M Na_2CO_3 solution. The reaction was degassed and placed under a nitrogen atmosphere. The reaction was heated to reflux for 3 hours. 9,10-diaminophenanthrene (26.2 mg, 0.125 mmole) was added against a strong flow of nitrogen, and the reaction was heated under nitrogen to 110 °C for 4 hours. 1.0 M HCl (1 mL)was added and the reaction was air oxidized for 24 hours. The reaction was purified by cation exchange (Sephadex SP C-25, 50/50 acetonitrile/0.3 M HCl) and HPLC verified that there were only two isomers present.

^1H NMR (mixture of two isomers) See appendix 1.

UV-visible spectrum, pH 5, H_2O; peaks at 381, 312, 301 and 271 (see appendix).

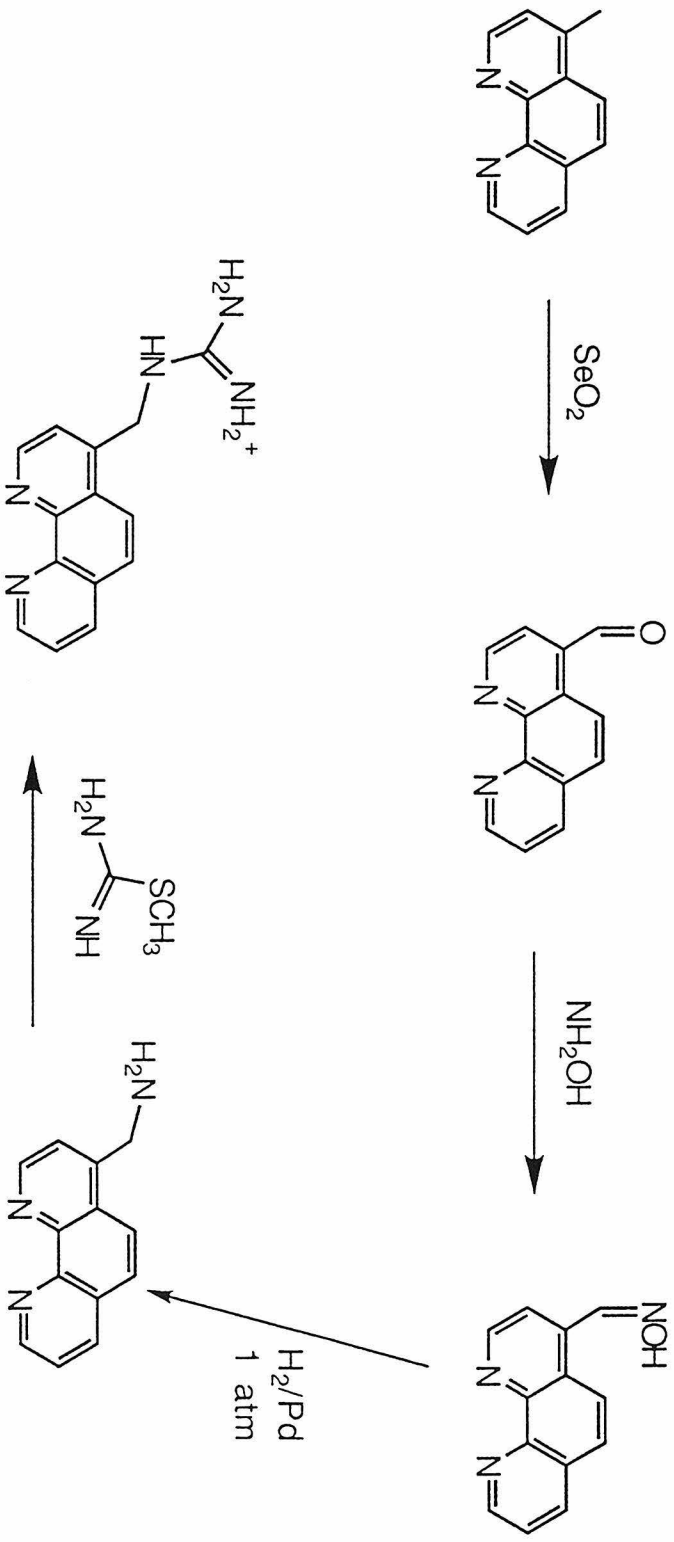
5.3. Results and Discussion

5.3.1 Ligand Synthesis

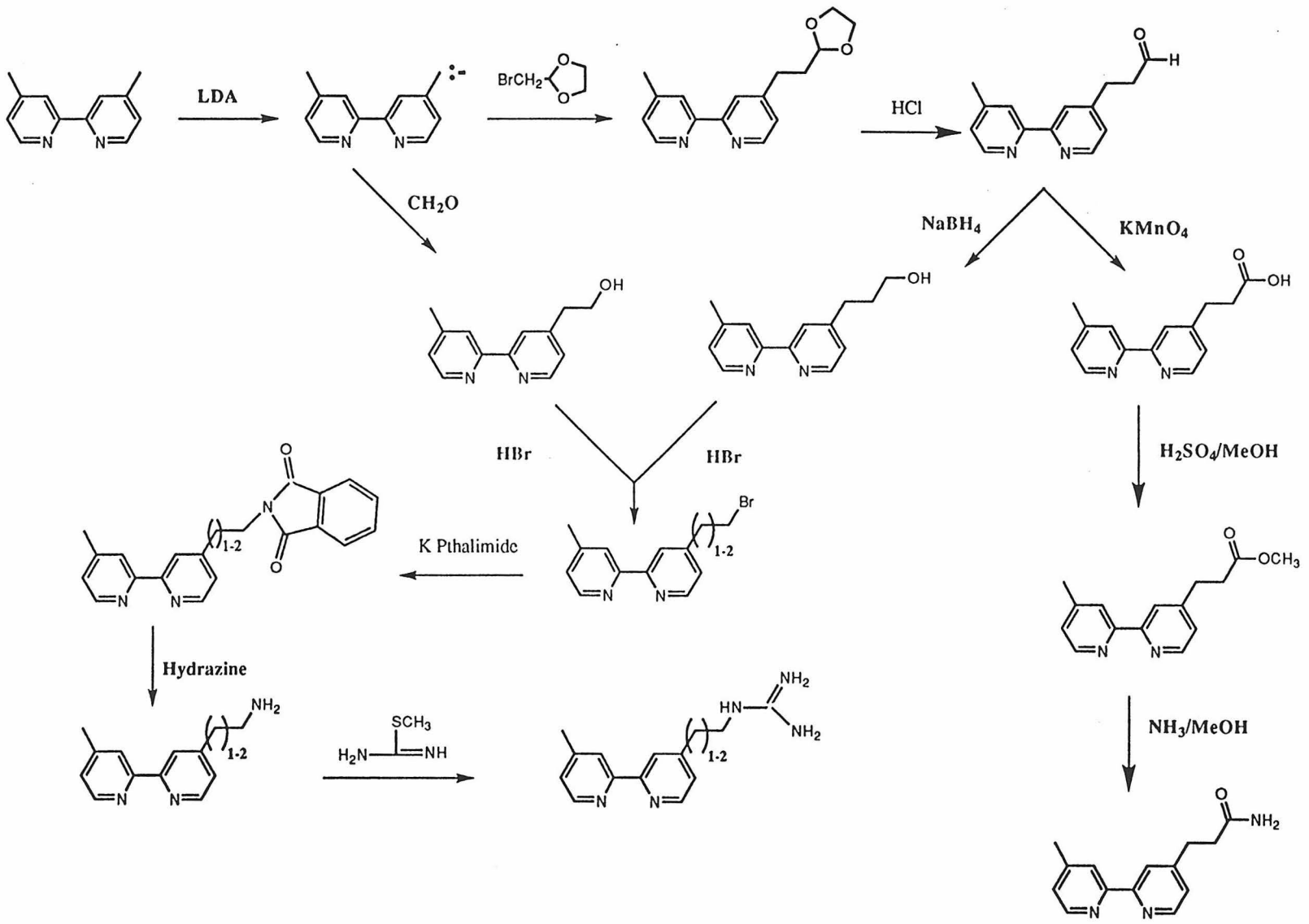
The synthetic protocols for making the different ligands are outlined in schemes 5.1 and 5.2. The synthetic protocols for 4-methylguanidinium-1,10-phenanthroline were adapted from ref. 1, and the protocols for the bipyridine ligands were adapted from ref. 2.

The synthesis of the ligands proceeded as expected with one notable exception. Conversion of 4-aminomethyl-1,10-phenanthroline to 4-guanidylmethyl-1,10-phenanthroline (Scheme 5.1) proved to be a very problematic reaction. Initial attempts to make the guanidinium ligand resulted in the decomposition of the starting material. It was subsequently discovered that it is necessary to add one equivalent of zinc salt to the reaction mixture to stabilize the reaction intermediate. Removal of the zinc ion during purification gives a stable ligand.

Scheme 5.1. Synthetic route for the synthesis of 4-guanidylmethyl-1,10-phenanthroline.


SeO2
NH2OH
H2/Pd
1 atm
H2N
SCH3
NH
NOH
O
H2N
HN
NH2+

Scheme 5.2. Synthetic route for the synthesis of the various bipyridine derivatives.


LDA
BrCH2
HCl
CH2O
NaBH4
KMnO4
OH
OH
OH
HBr
HBr
H2SO4/MeOH
Br
1-2
K Pthalimide
1-2
OCH3
Hydrazine
NH2
1-2
SCH3
H2N
NH
HN
NH2
NH2
NH3/MeOH
NH2

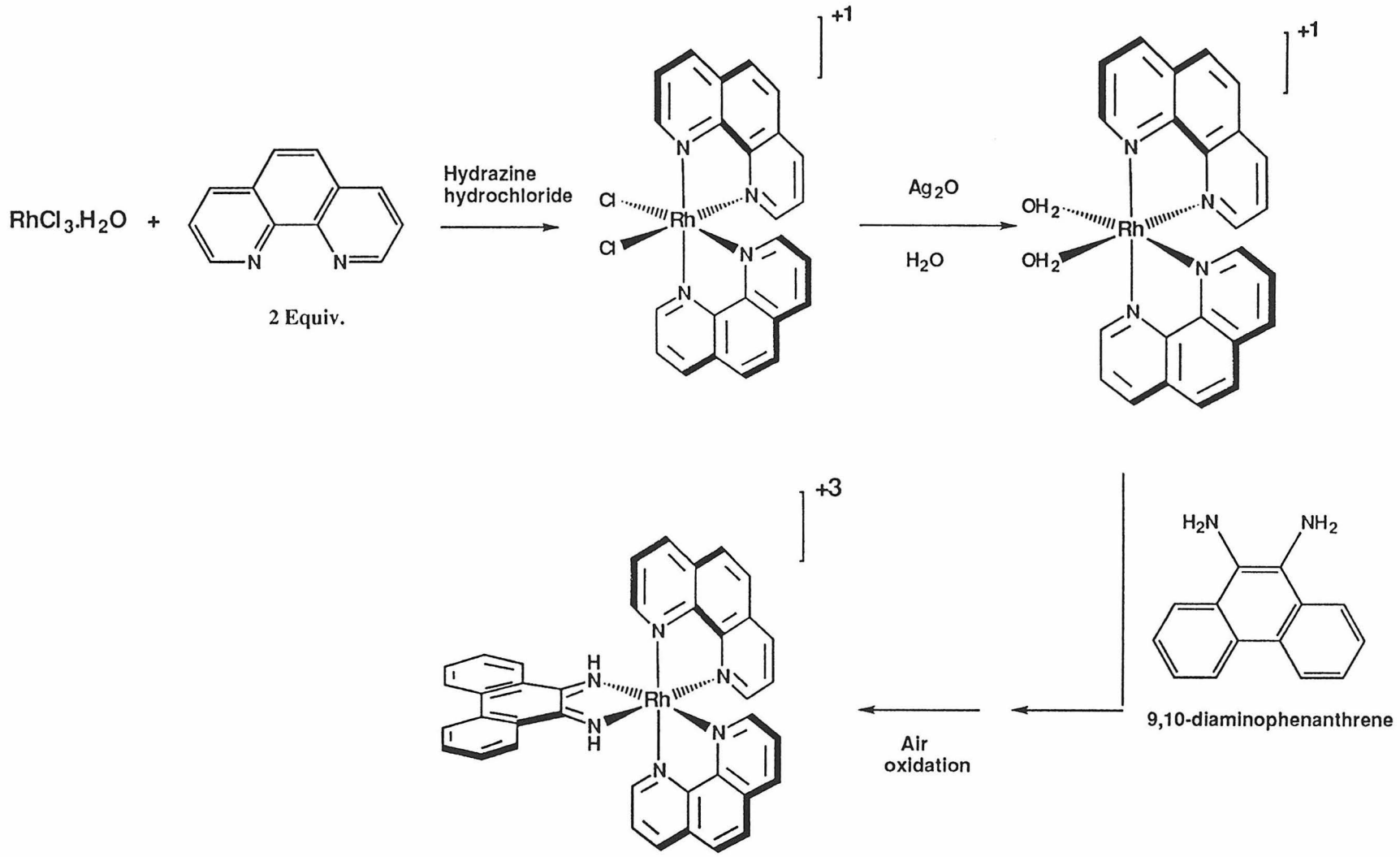
5.3.2. Synthesis of Metal Complexes: One-Pot Synthesis

Previously reported procedures (3,4) for the synthesis of $[Rh(phen)_2phi]Cl_3$ and $[Rh(bpy)_2phi]Cl_3$ complexes has involved reacting $RhCl_3$ with 2 equivalents of phen or bpy (X) in the presence of a catalytic amount of hydrazine hydrochloride to generate $[RhX_2Cl_2]Cl$. Following purification, the strongly coordinating chlorides are exchanged for weakly coordinating triflate (OTf) or aquo ligands. This step requires the addition of silver oxide or silver triflate, which subsequently precipitate from the reaction as AgCl. $[RhX_2(H_2O)_2](OTf)$ is then reacted with 1 equivalent of 9,10-diaminophenanthrene (DAP), and following overnight air oxidation, the desired compound, $[RhX_2phi](OTf)_3$, is obtained (Scheme 5.3). A major drawback of this procedure is that it is very time consuming and gives low yields.

Initial attempts to synthesize $Rh(MGP)_2phi^{5+}$ and $Rh(APB)_2phi^{5+}$ using the procedure outlined in Scheme 5.3 resulted in phi complexes of rhodium(III) in which the guanidinium moiety and amino moiety were no longer intact. Degradation of these functionalities was subsequently traced to the silver salts that were used to convert $[RhX_2Cl_2]Cl$ to $[Rh(X_2)(H_2O)_2]Cl$. Alternative methods for the removal of the chlorides either gave no reaction (thallium salts) or resulted in a mixture of many products. As a result, a new method by which to synthesize $Rh(MGP)_2phi^{5+}$ and other phi complexes of rhodium (III) was developed.

This new procedure is a one-pot synthesis and utilizes either $Rh(NO_3)_3$ as a starting material if a derivative of $[Rh(phen)_2phi]Cl_3$ is desired or $RhCl_3$ if a derivative of $[Rh(bpy)_2phi]Cl_3$ is the target compound. The ability to synthesize phi complexes of rhodium(III) with aliphatic ancillary ligands in a one-pot synthesis has been reported previously (5) and the procedure by which to synthesize $[Rh(phen)_2phi]Cl_3$ and $[Rh(bpy)_2phi]Cl_3$ derivatives was adapted from this.

Scheme 5.3. Previously reported synthetic route for $Rh(phen)_2phi^{3+}$. See ref. 3 and ref. 4.


RhCl3.H2O +
2 Equiv.
Hydrazine hydrochloride
Cl
Cl
Rh
N
+1
Ag2O
H2O
OH2
OH2
H2N
NH2
9,10-diaminophenanthrene
Air oxidation
H
+3

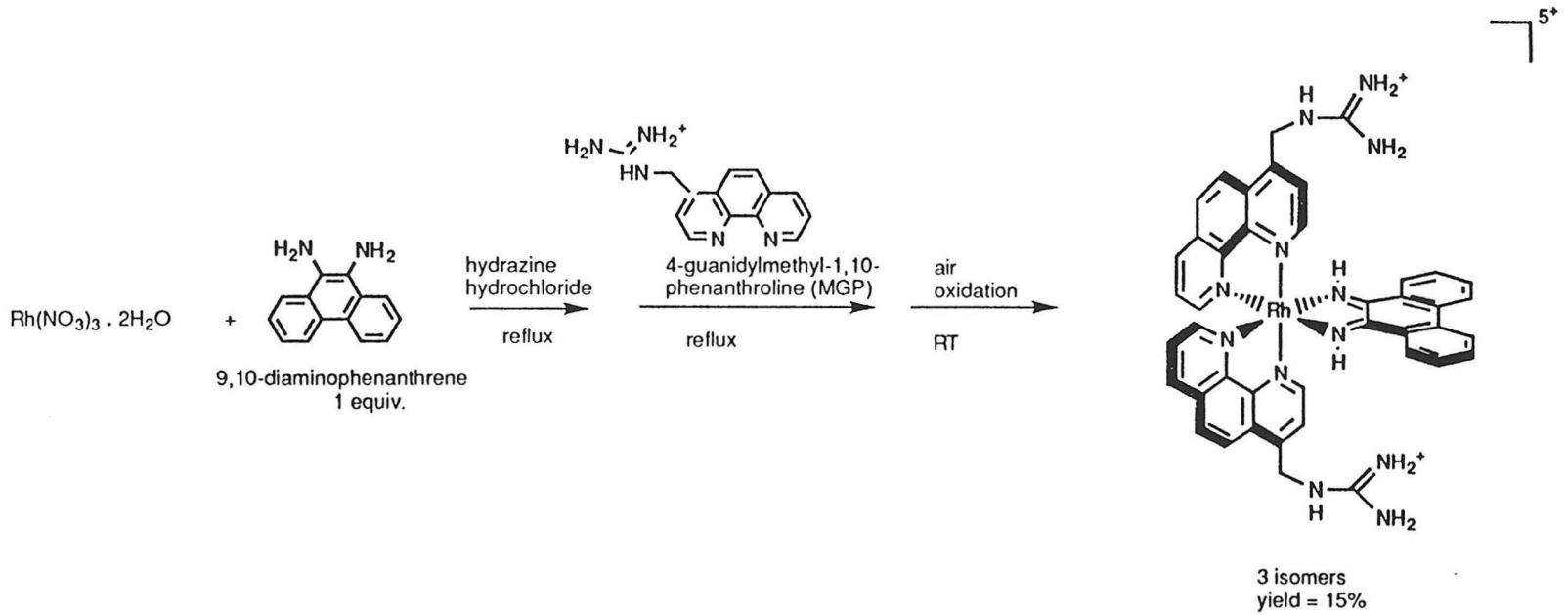
The one-pot synthesis is outlined in Scheme 5.4. A slight excess of 9,10-diaminophenanthrene is heated to reflux with $Rh(NO_3)_3 \cdot xH_2O$ under a nitrogen atmosphere for 2 hours. A solution of 4-guanidylmethyl-1,10-phenanthroline (or other ligand) is added via cannula to the rhodium solution, and the reaction is stirred at reflux for 12 hours. The pH of the solution is adjusted to about 1 by addition of dilute HCl, and 50 mL of distilled water is added. The reaction is stirred open to the air for 24 hours to allow for complete oxidation of the metal complex to generate $[Rh(MGP)_2phi]Cl_5$. The resulting solution is filtered to remove insoluble materials. The overall yield of complex is about 15%. Derivatives of $[Rh(bpy)_2phi]Cl_3$ are synthesized using a similar protocol except it is possible to use $RhCl_3$ and shorter reaction times.

An important aspect of this one-pot reaction is the difference in reactivity of the phenanthroline ligand as compared to the bipyridine ligand. The phenanthroline ligand coordinates rhodium much more slowly than the bipyridine ligand. This reactivity difference is most likely related to the difference in flexibility of the ring systems of the two ligands. As a result, it is possible to use either $RhCl_3$ or $Rh(NO_3)_3$ in the bipyridine reaction, while only the more easily substituted $Rh(NO_3)_3$ will work for phenanthroline. It is preferable to use $RhCl_3$ whenever possible because it gives a cleaner reaction than $Rh(NO_3)_3$, and the overall yields are slightly better.

5.3.3. Purification and Separation of Isomers

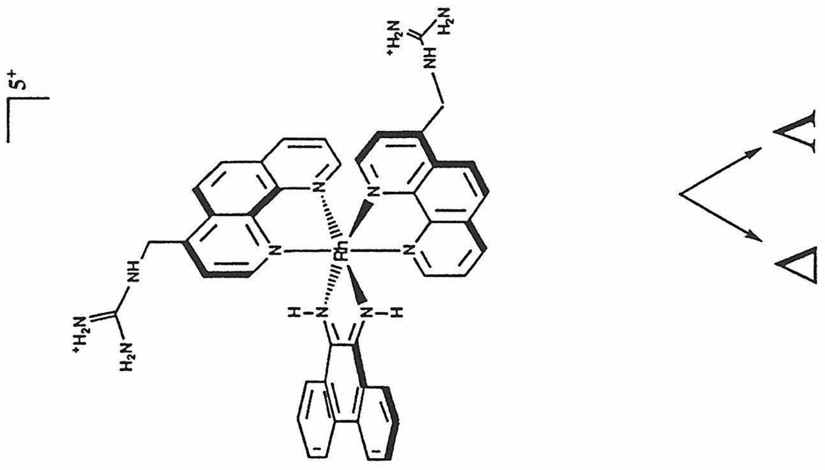
Due to the asymmetric nature of the bipyridine and phenanthroline ligands used, the derivatives of $[Rh(phen)_2phi]Cl_3$ and $[Rh(bpy)_2phi]Cl_3$ were generated as a mixture of three isomers which occur in a statistical distribution of 1:2:1. Each stereoisomer also exists as an equal mixture of 2 enantiomers (Figure 5.1).

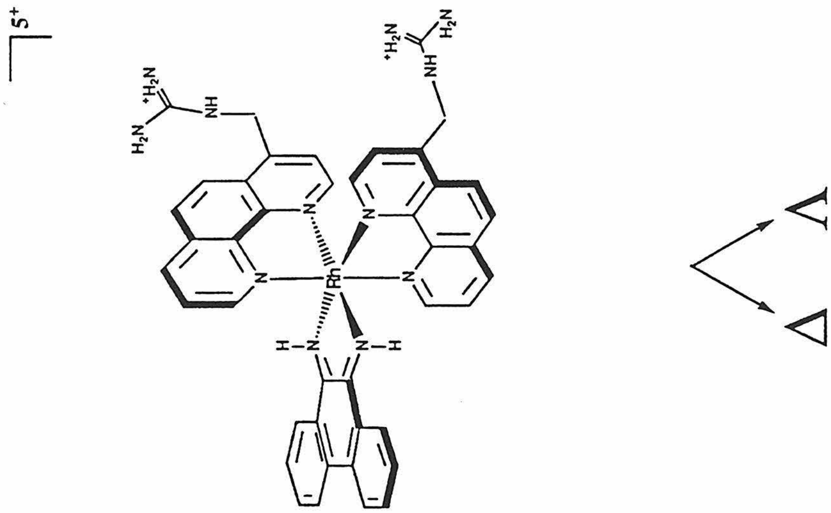
Scheme 5.4. Schematic of the one-pot synthesis used to make functionalized $Rh(phen)_2phi^{3+}$ and $Rh(bpy)_2phi^{3+}$ complexes.

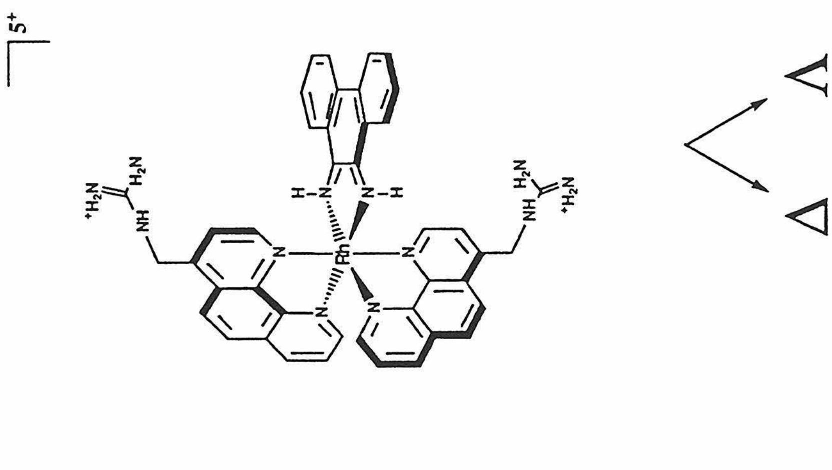


3 isomers
yield = 15%

Figure 5.1. Illustration of the 6 different isomers of $Rh(MGP)_2phi^{5+}$. (Left) the C_2 symmetric isomer 1-$Rh(MGP)_2phi^{5+}$, (Center) the C_2 symmetric isomer 2-$Rh(MGP)_2phi^{5+}$ and (right) the asymmetrical isomer 3-$Rh(MGP)_2phi^{5+}$. Each isomer is also an equal mixture of 2 enantiomers.


5+
Rh
Δ
Λ


5+
Rh
Δ
Λ


5+
Rh
Δ
Λ

The reaction mixtures were purified by cation exchange chromatography. Importantly, the guanidinium and amino complexes could not be eluted with HCl or NaCl. It was instead necessary to use a divalent cation like $MgCl_2$. This behavior is most likely the result of the increased charge on the complex (+5 versus +3). Following cation exchange, the different stereoisomers of the complex were separated by reverse phase HPLC. It was possible to separate the positional isomers of every metal complex described in this thesis by HPLC. The best separation is obtained under isocratic elution conditions.

The Δ and Λ-enantiomers of $[Rh(MGP)_2phi]Cl_5$ and $[Rh(GEB)_2phi]Cl_5$ have been completely resolved. Initial attempts at resolving the enantiomers of $[Rh(MGP)_2phi]Cl_5$ by cation exchange chromatography using the previously described tris(L-cysteinylsulfinato)cobaltate (III) ($[Co(L\text{-}cysu)_3]^{3-}$) as a chiral eluant were unsuccessful. As a result an alternative method for the separation of the enantiomers was developed.

It was discovered that the enantiomers of $[Rh(MGP)_2phi]Cl_5$ could be separated on a chiral HPLC column. It was possible to obtain baseline resolution between the Δ and Λ enantiomers of $[Rh(MGP)_2phi]Cl_5$ on a reverse-phase cellulose tris(3,5-dimethylphenyl carbamate) support (6). The impetus for choosing this column was that the support has a slight propeller nature to it, which could potentially match the chirality of one of the enantiomers but not the other. This column has been shown capable of resolving many of the compounds which are currently being explored in the Barton group.

It has subsequently been discovered that the enantiomers of $[Rh(MGP)_2phi]Cl_5$ and $[Rh(GEB)_2phi]Cl_5$ can be completely resolved by cation exchange using potassium antimonyl tartrate as a chiral eluant. Potassium antimonyl tartrate ($K_2[Sb_2(C_4H_2O_6)_2]\cdot 3H_2O$) offers many advantages over $[Co(L\text{-}cysu)_3]^{3-}$. Potassium antimonyl tartrate is inexpensive and can be readily obtained

from Aldrich, while it is necessary to synthesize $[Co(L\text{-}cysu)_3]^{3-}$. More importantly, the cobalt complex is light sensitive and orange/brown in color, making it extremely difficult to view the separation of the orange colored rhodium enantiomers. In contrast, potassium antimonyl tartrate is stable and colorless in solution. A last advantage is that potassium antimonyl tartrate is more water soluble, making it possible to make more concentrated solutions when necessary.

Recently, we have been able to separate all six enantiomers of $[Rh(MGP)_2phi]Cl_5$ and $[Rh(GEB)_2phi]Cl_5$ on a single cation exchange column using potassium antimonyl tartrate as a chiral eluant. This improvement represents a significant reduction in time for the separation of all six enantiomers. This procedure has yet to be tried with the amino and amido compounds; however, a similar separation should be attempted with all future complexes.

5.3.4. Characterization of the Stereoisomers

The stereoisomers are generated in the expected 1:2:1 ratio with the most prevalent isomer being the asymmetrical molecule and the two lesser isomers being symmetrical molecules. Structural assignments of the two symmetric isomers were made by NMR comparison with $[Rh(phen)_2phi]Cl_3$. The NMR assignments for $[Rh(phen)_2phi]Cl_3$ have previously been published (7), and these data show that the C3 proton of the ancillary phenanthroline ring is shifted downfield by 0.4 ppm when it is pointed towards the phi ligand as opposed to being pointed towards the other phenanthroline ligand. This change in chemical shift is most likely a reflection of the structural differences between the phi and the phenanthroline ligand.

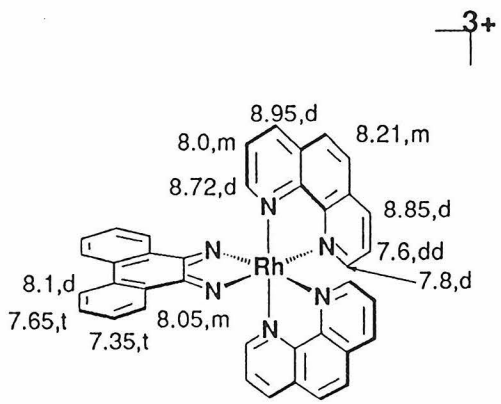
The C3 hydrogen of phenanthroline ligand is situated above the aromatic ring systems of the other two ligands in the $[Rh(phen)_2phi]Cl_3$ complex. In this location, the aromatic ligand ring current increases the amount of shielding at this

hydrogen. Comparison of the structural differences, between the phi ligand and the phen ligand, reveal that the phi ligand is extended away from the metal center relative to the phen ligand. Thus, the C3 hydrogen projects further into the ring currents of the phenanthroline ligand and is shielded to a greater extent. Hence, the C3 proton is shifted upfield when it is towards the phenanthroline ligand. This trend holds true in the derivatized $[Rh(phen)_2phi]Cl_3$ complexes and makes it possible to assign the isomers of $[Rh(MGP)_2phi]Cl_5$ and other complexes (Figure 5.2). Similar reasoning has been used to assign the isomers of the bipyridyl derivatives.

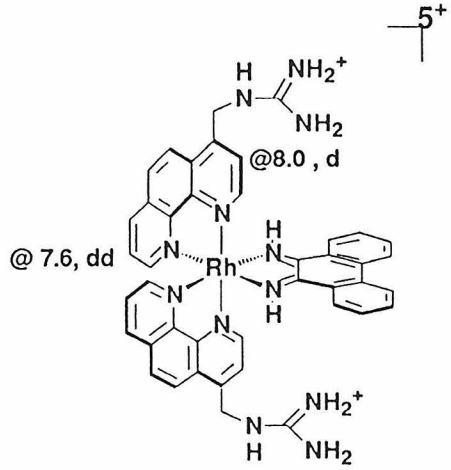
5.3.5. Characterization of the Δ and Λ-Enantiomers.

The assignments of the Δ- and Λ-configurations of the different enantiomers were made by comparison of the circular dichroism (CD) spectra of the new complexes with those previously reported for the different enantiomers of $[Rh(phen)_2phi]Cl_3$ (7). The CD spectra for $[Rh(MGP)_2phi]Cl_5$ and $[Rh(GEB)_2phi]Cl_5$ are shown in the appendix of this chapter. The CD spectra of the enantiomers are equal and opposites as required. The absorption band at 280 nm in the $[Rh(phen)_2phi]Cl_3$ and the absorption band at 310 nm in the $[Rh(bpy)_2phi]Cl_3$ are very distinctive, and can be used to assign the configuration of the complexes. In the Λ-enantiomer, this absorption is positive and in the Δ-enantiomer, it is negative. Comparison of relative mobility of the complexes on cation exchange columns with the same chiral eluant and comparison of the photocleavage patterns of the different enantiomers on DNA fragments have further verified the assignments.

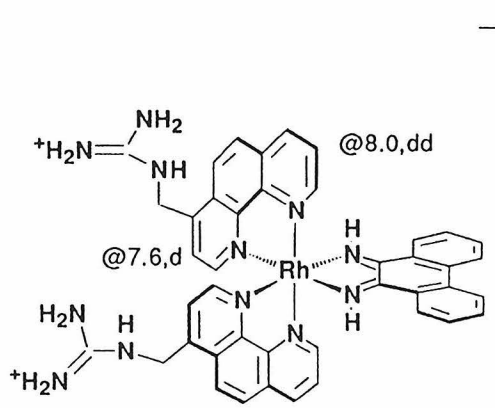
Figure 5.2. Assigning the isomers of $Rh(MGP)_2phi^{5+}$ by NMR. The C3 hydrogen of $Rh(MGP)_2phi^{5+}$ is shifted by 0.4 ppm depending of its location with respect to the phi ligand.



Δ-Rh(phen)$_2$phi^{3+}



Λ-1-Rh(MGP)$_2$phi^{5+}



Λ-2-Rh(MGP)$_2$phi^{5+}

5.3.6. Synthesis of Phi Complexes of Rhodium(III) Possessing Three Different Ligands.

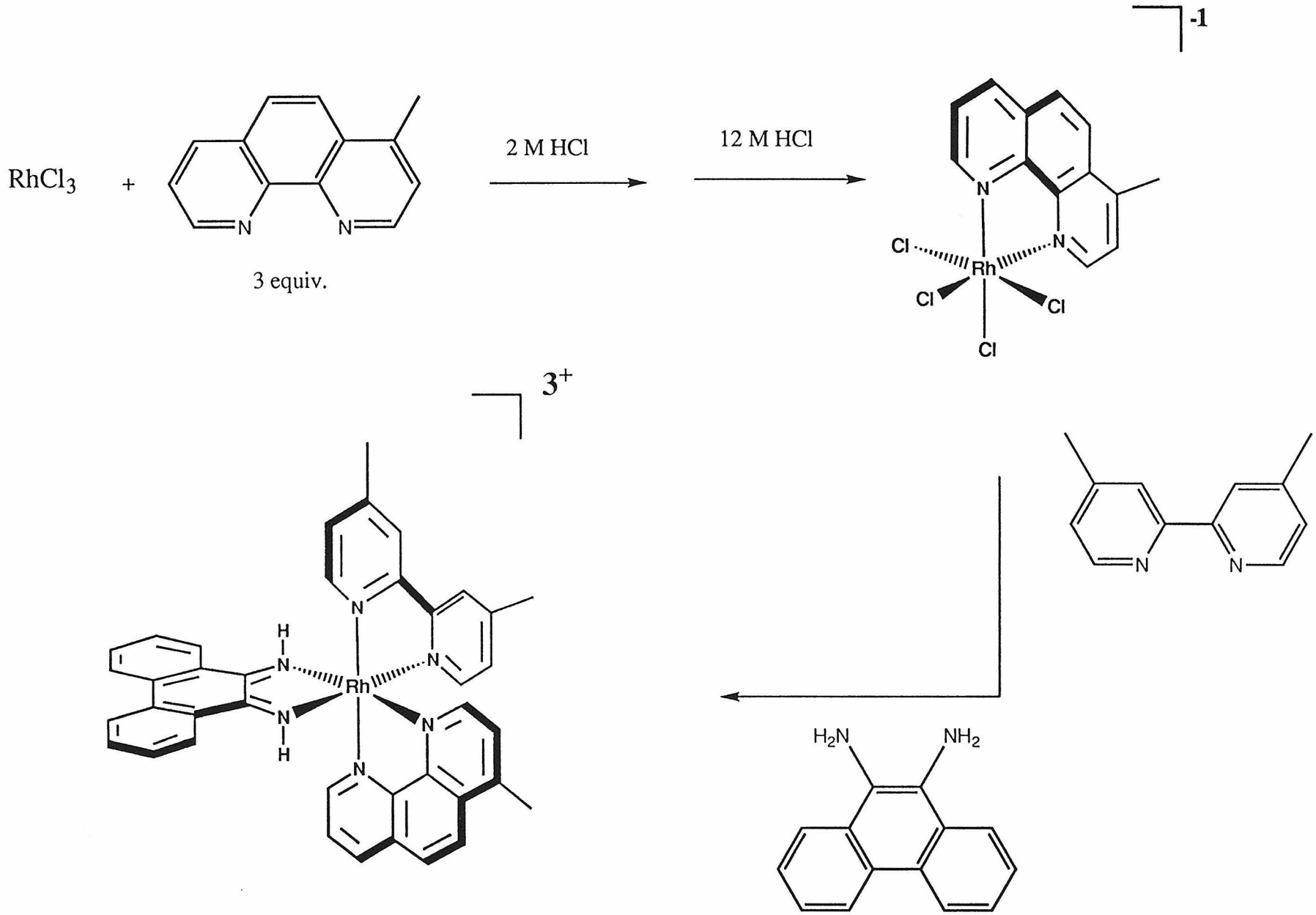
The next generation of transition metal complexes for DNA recognition will most likely possess three different ligands. One can envision making complexes which have one ligand which can intercalate into the DNA, a second ligand, like 4-guanidylmethyl-1,10-phenanthroline, which targets guanines, and a third ligand which either targets adenines or possibly induces dimerization between complexes. Previous work has shown that it is possible to synthesize phi complexes of rhodium(III) which possess three different ligands (8). However, these methods gave monophi complexes in which the two ancillary ligands added in a statistical fashion such that a mixture of bisphenanthroline, bisbipyridine and the desired monophenanthroline/monobipyridine complexes was obtained. Importantly, a driving force for the addition of only one phi ligand was observed.

In an effort to avoid scrambling of the ligands, a new synthetic protocol was devised. As outlined in scheme 5.5, hydroxonium tetrachloro-(4-methyl-1,10-phenanthroline)rhodate(III) was synthesized as previously described by McKenzie (9). This molecule was then reacted with one equivalent of 9,10-diaminophenanthrene, followed by 1 equivalent of 4,4-dimethylbipyridine. After air oxidation, the desired complex (4-methyl-1,10-phenanthroline)(4,4'-dimethyl-2,2'-bipyridine)(phi)rhodium(III)trichloride was obtained as a mixture of two isomers. Importantly, no scrambling between the phenanthroline and bipyridine ligands was observed, and this procedure should be adaptable to almost any system.

5.4. Conclusion

The desire to synthesize phi complexes of rhodim(III) which possess novel functionalities such as the guanidinium moiety has led to many improvements in

Scheme 5.5. Synthetic route for the synthesis of (4-methyl-1,10-phenanthroline)(4,4'-dimethyl-2,2'-bipyridine)(phi)rhodium(III).


RhCl3
+
3 equiv.
2 M HCl
12 M HCl
-1
Cl
Rh
Cl
Cl
Cl
N
N
3+
H
N
N
H
Rh
N
N
N
N
N
H2N
NH2
2 isomers

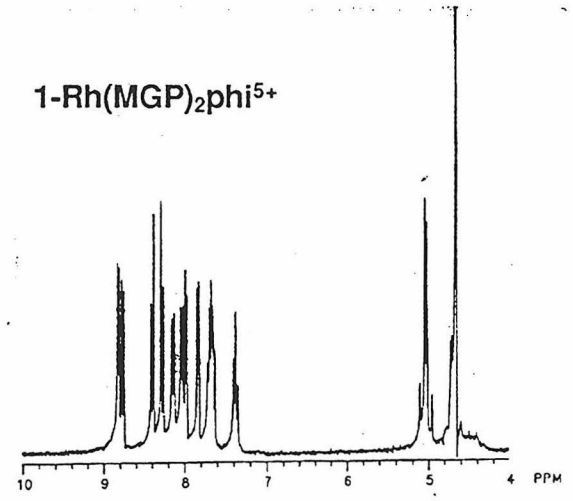
improvement over older methods (8) in that scrambling of the ancillary ligands is avoided. This protocol is also general in nature and should be amenable to a wide variety of functionalities.

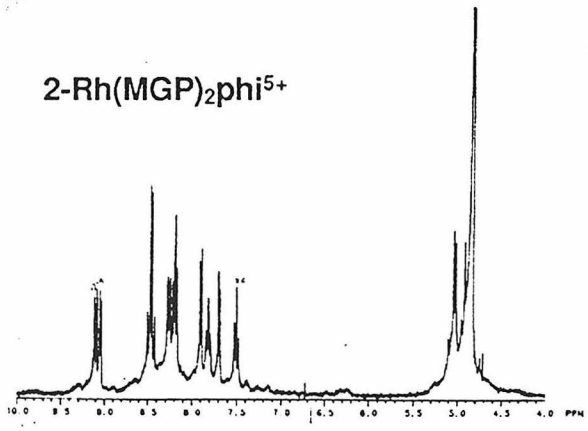
These advances in synthetic techniques set the stage for the development of new phi complexes of rhodium(III) with even more impressive DNA recognition capabilities than $[Rh(MGP)_2phi]Cl_5$. With these synthetic tools in hand, it is now necessary to concentrate on the design of complexes which make use of the DNA recognition principles that have been demonstrated by the collection of molecules described here.

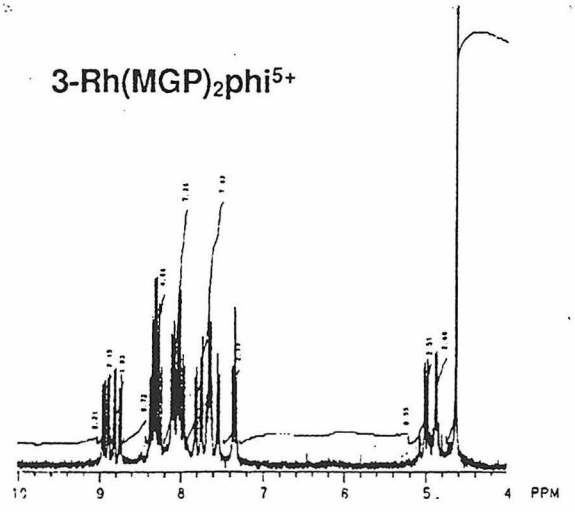
5.6. References

1. Chandler, Deady, and Reiss, *J. Heterocyclic Chem.*, **18**, 599 (1981)
2. Delleciana, L., Hamachi, I. and Meyer, T.J., *J. Am. Chem. Soc.,* **54**, 1731- 1735 (1989)
3. Sitlani, A. and Barton, J.K., *Biochem.*, **33**, 12100 (1994)
4. Pyle, A.M., Long, E.C. and Barton, J.K., *J. Am. Chem. Soc.,* **111,** 4520 (1989)
5. Krotz, A. H., Hudson, B. P. and Barton, J. K. (1993) *J. Am. Chem. Soc.,**115***, 12577-12578 (1993)
6. Enantiomers were separated on a Chiralcel OD-R reverse phase HPLC column (chiral technologies). The eluant was acetonitrile/aqueous phase (pH 2, 0.5 M $HClO_4/NaClO_4$). The best separation was obtained under isocratic conditions at 78/22 acetonitrile/aqueous phase.
7. David, S.D. and Barton, J.K., *J. Am. Chem. Soc.,* **115**, 2984 (1993)
8. Kingsbury, K. and Barton, J.K., unpublished results
9. McKenzie, E.D. and Plowman, R.A., *J. inorg. Nucl. Chem.*, **32**, 199-212 (1970).

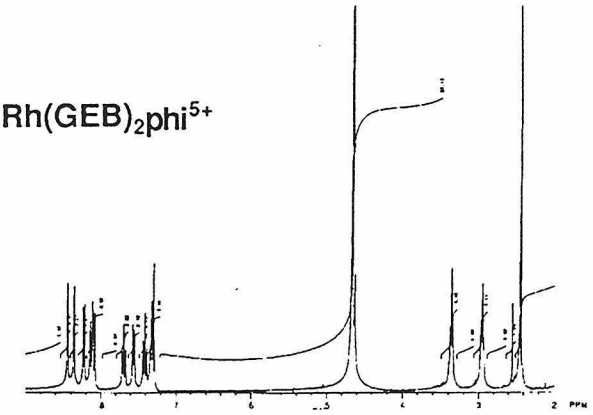
5.7 Appendix 1: Characterization Data



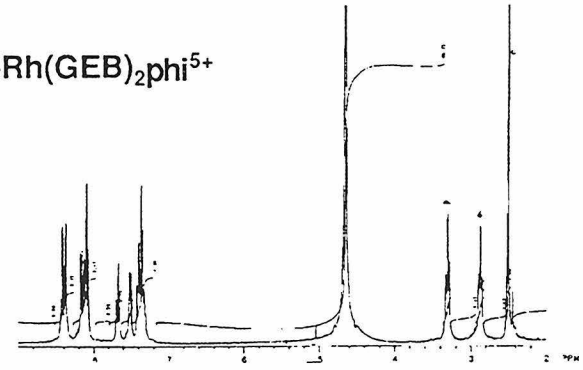




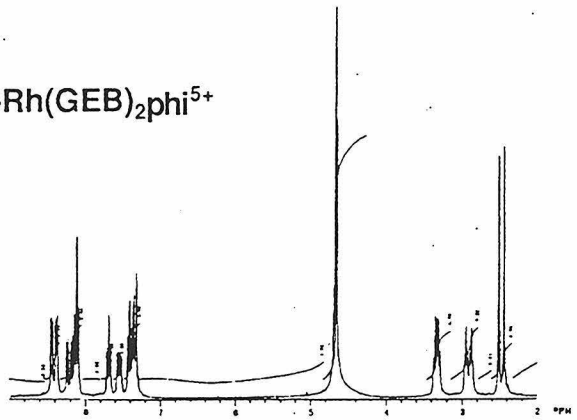
1-Rh(GEB)$_2$phi^{5+}

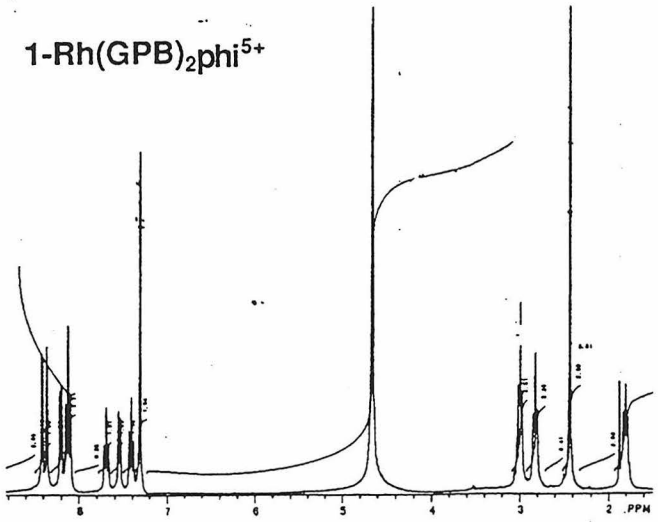


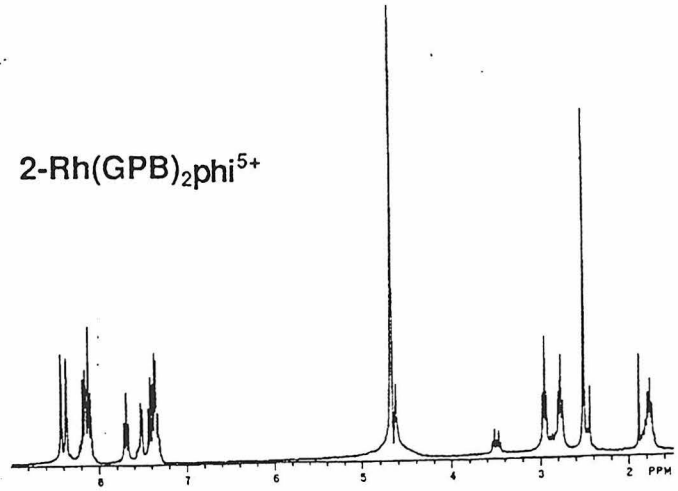
2-Rh(GEB)$_2$phi^{5+}

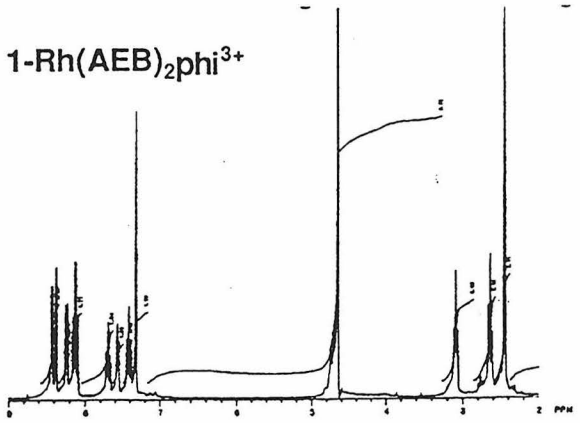


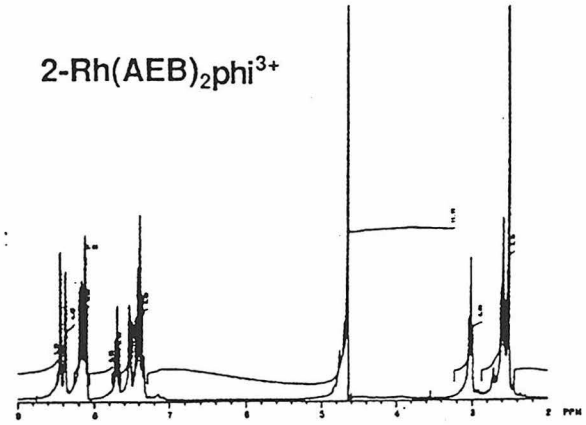
3-Rh(GEB)$_2$phi^{5+}

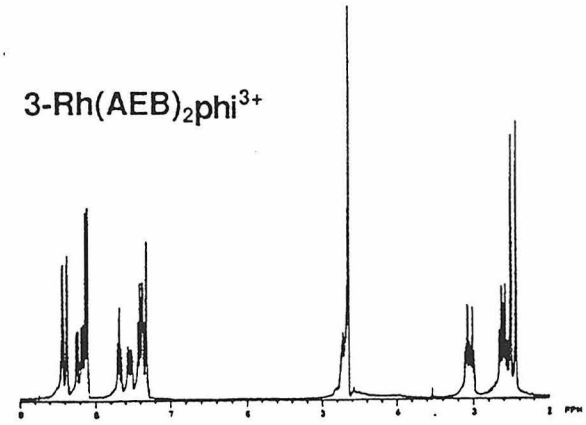



1-Rh(GPB)2phi5+
8
7
6
5
4
3
2 PPM

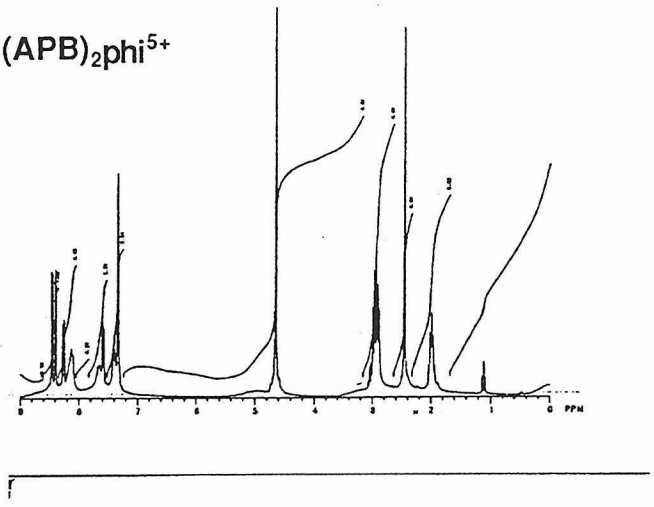

2-Rh(GPB)2phi5+
8
7
6
5
4
3
2 PPM


1-Rh(AEB)2phi3+

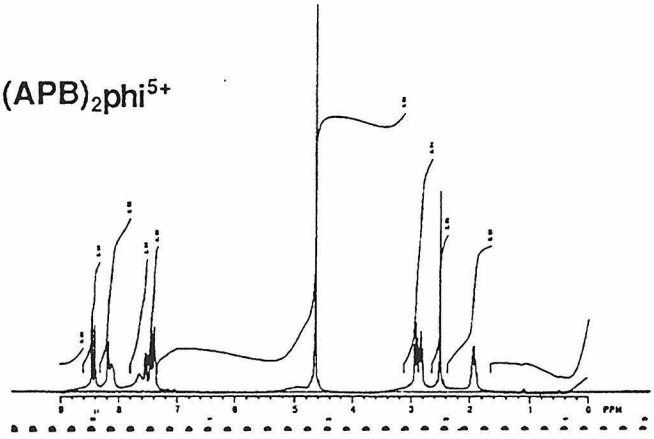

2-Rh(AEB)2phi3+


3-Rh(AEB)2phi3+

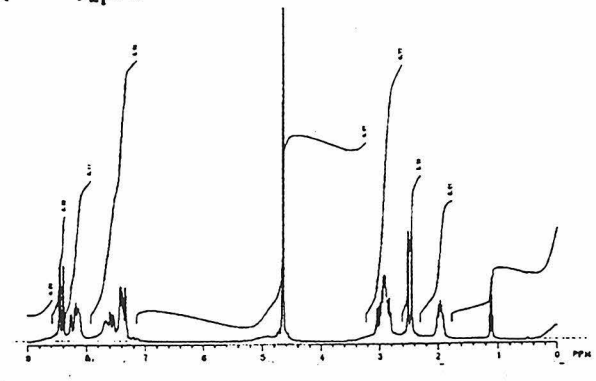
1-Rh(APB)$_2$phi^{5+}

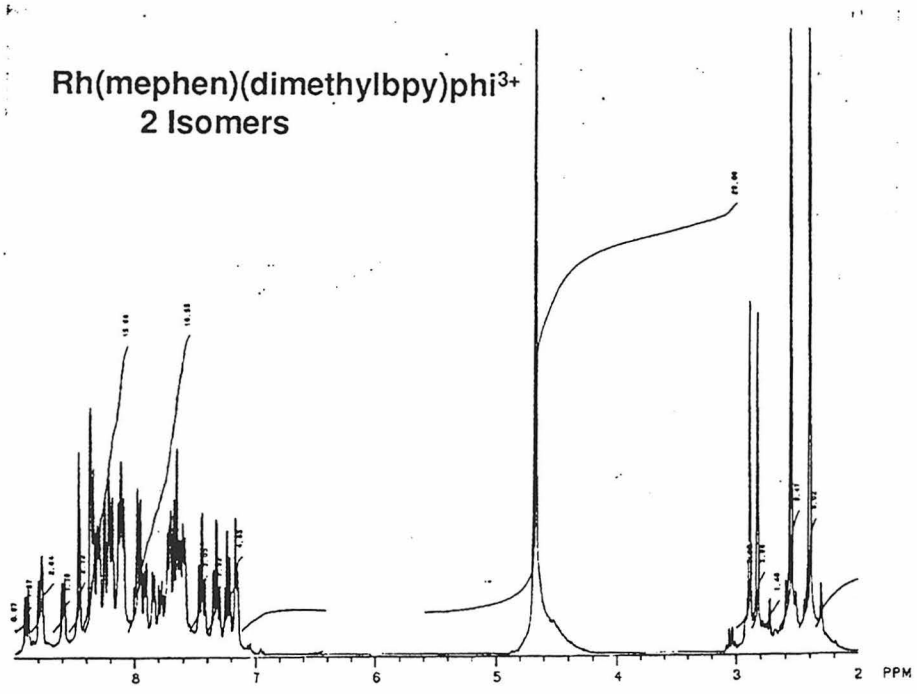


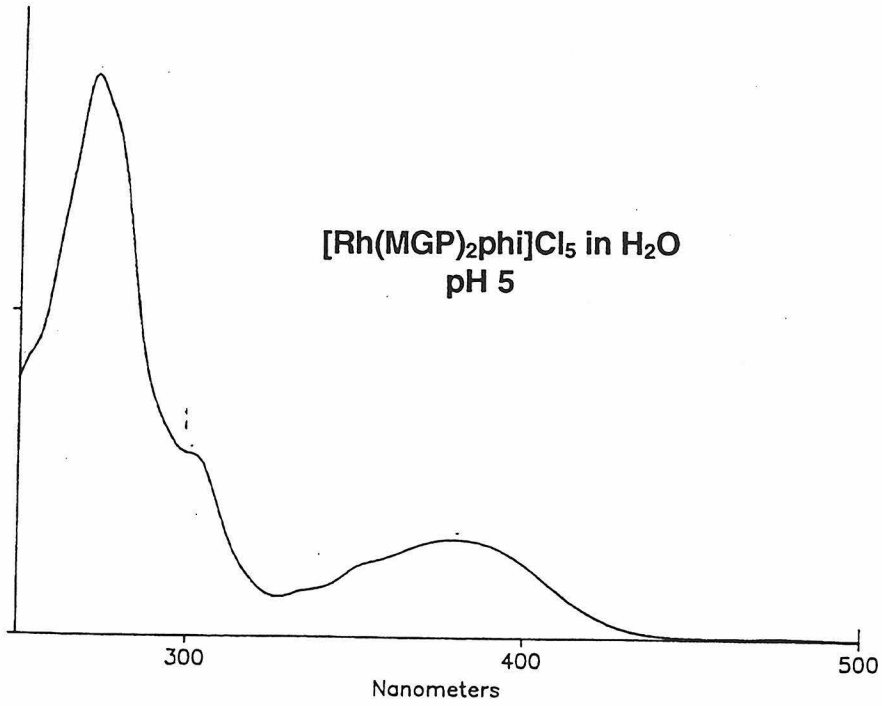
2-Rh(APB)$_2$phi^{5+}

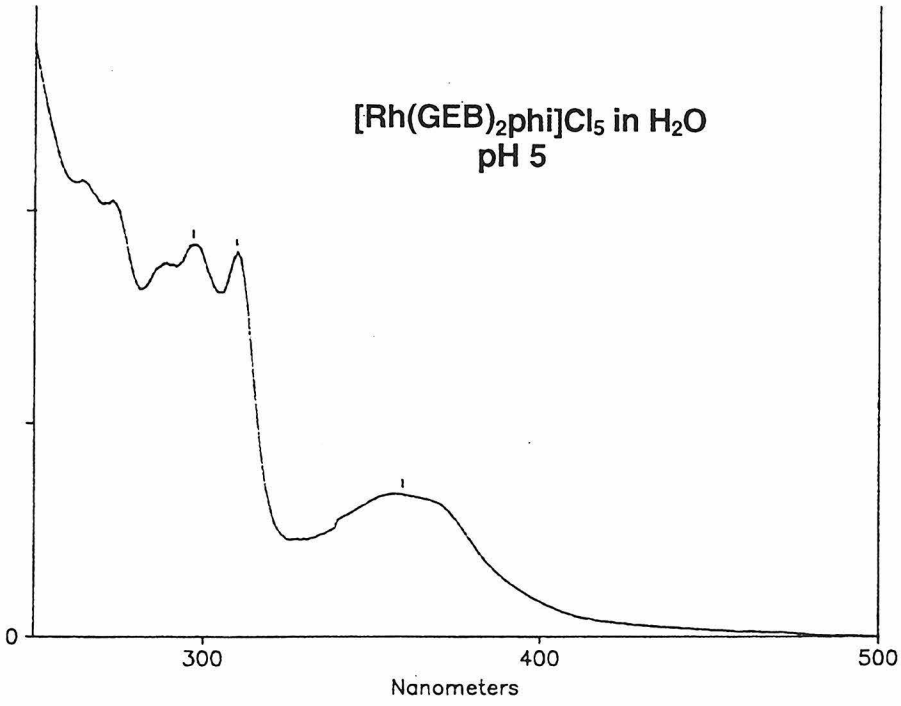


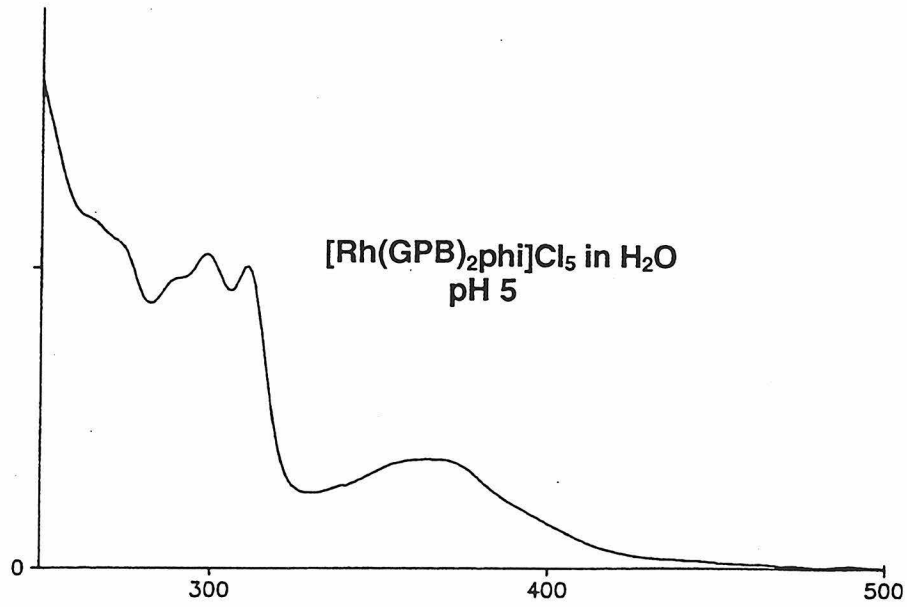
3-Rh(APB)$_2$phi^{5+}

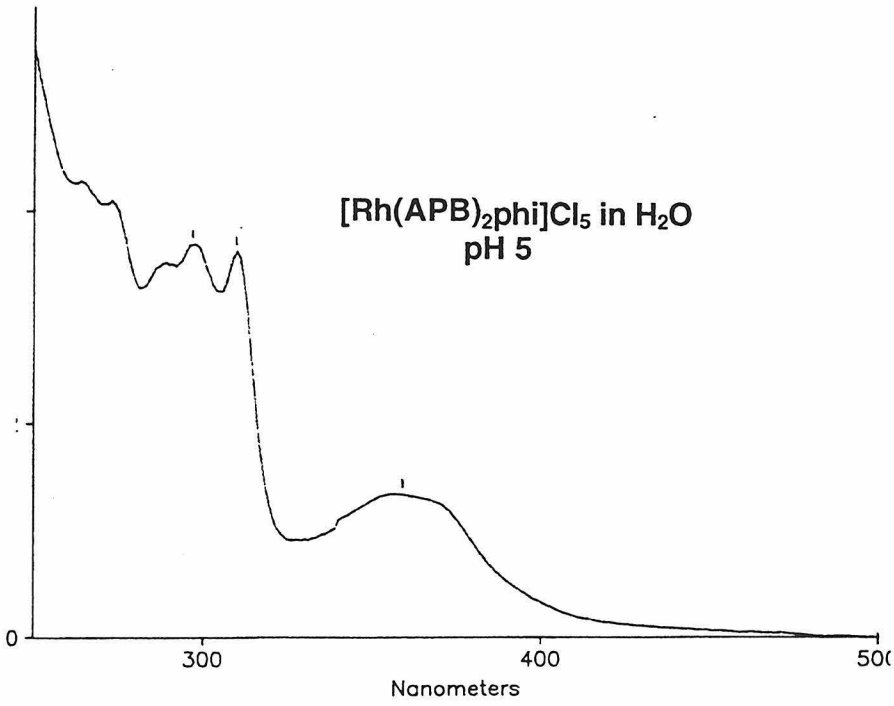


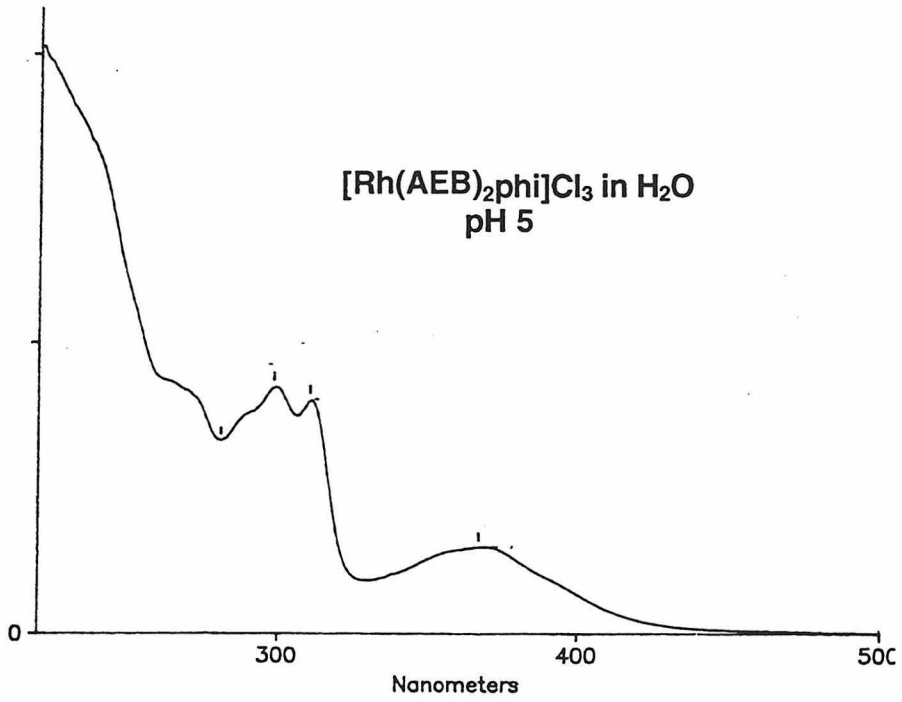

Rh(mephen)(dimethylbpy)phi3+
2 Isomers
8
7
6
5
4
3
2
PPM

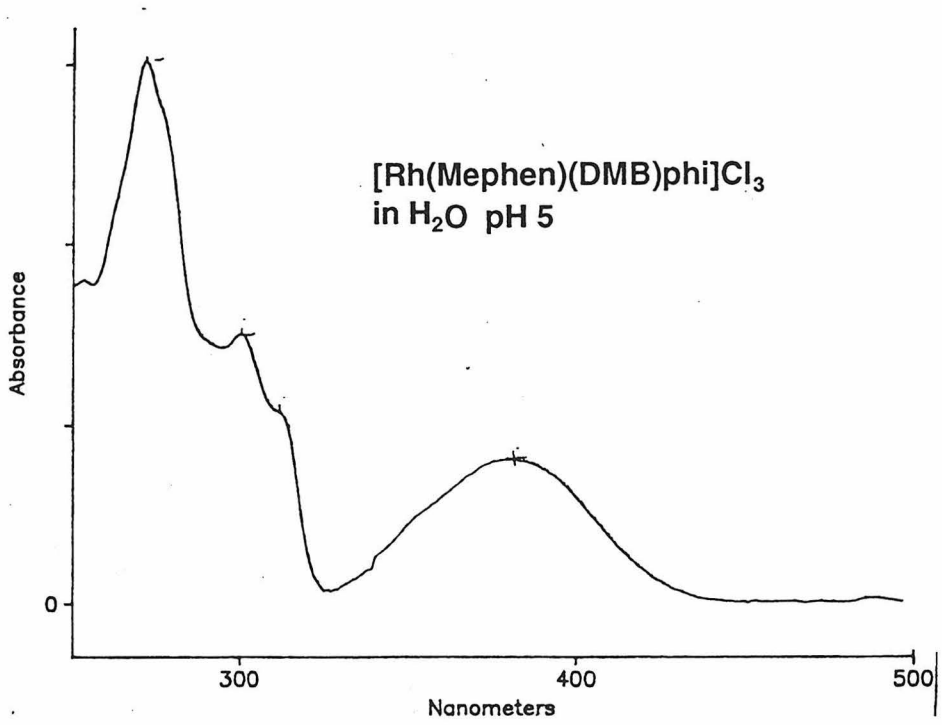

[Rh(MGP)2phi]Cl5 in H2O
pH 5
300
400
500
Nanometers

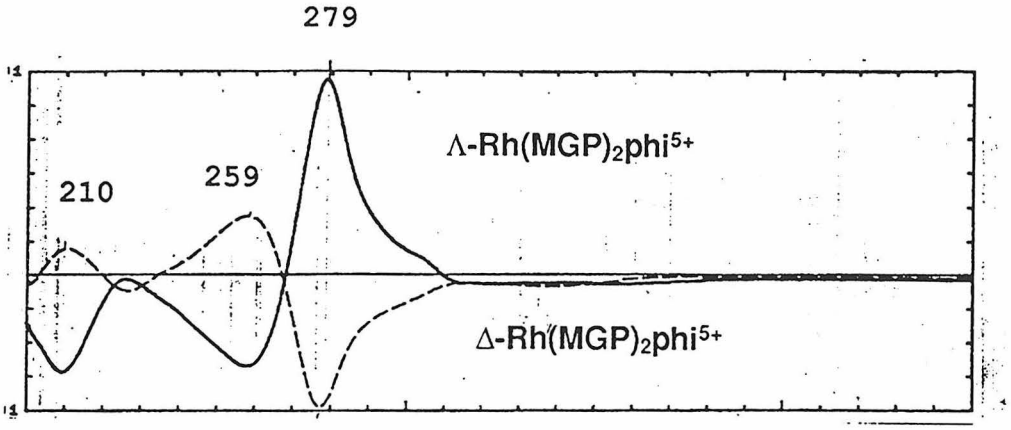

[Rh(GEB)2phi]Cl5 in H2O
pH 5
0
300
400
500
Nanometers

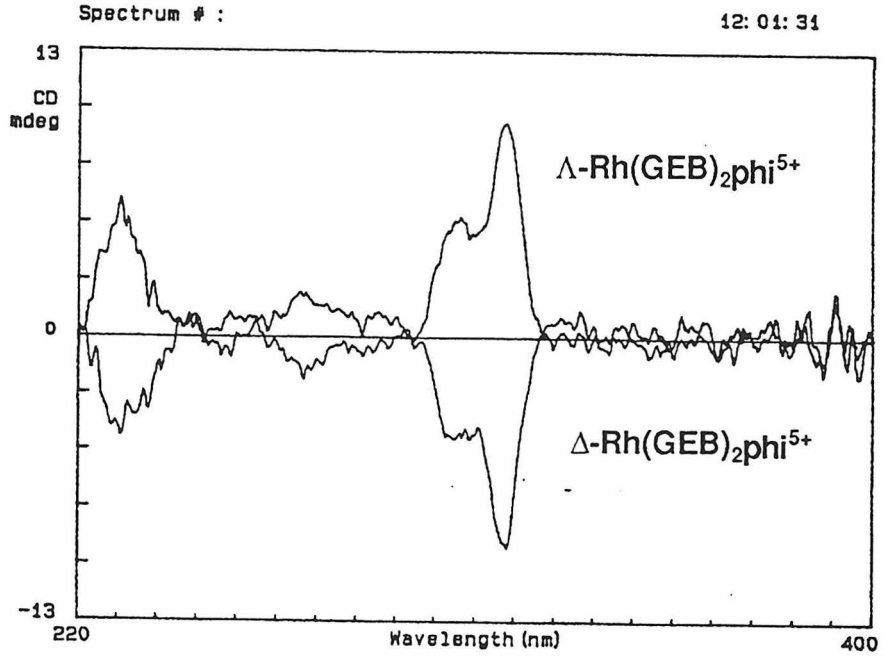

[Rh(GPB)2phi]Cl5 in H2O
pH 5
0
300
400
500


[Rh(APB)2phi]Cl5 in H2O
pH 5
0
300
400
Nanometers


[Rh(AEB)2phi]Cl3 in H2O
pH 5
0
300
400
500
Nanometers


[Rh(Mephen)(DMB)phi]Cl3
in H2O pH 5
Absorbance
0
300
400
500
Nanometers


279
210
259
Λ-Rh(MGP)2phi5+
Δ-Rh(MGP)2phi5+


Spectrum # :
12: 01: 31
13
CD
mdeg
0
-13
220
Wavelength (nm)
400
Λ-Rh(GEB)2phi5+
Δ-Rh(GEB)2phi5+

Chapter 6: Conclusions and Perspectives.

Both enantiomers of 1-Rh(MGP)$_2$phi^{5+} have been shown to recognize DNA through direct and indirect readout of the DNA helix. By simultaneously utilizing both elements of DNA recognition, Λ-1-Rh(MGP)$_2$phi^{5+} displays exquisite sequence specificity and unprecedented DNA binding affinity (binding at subnanomolar concentrations). An essential feature of this recognition is the sequence-specific unwinding of the DNA helix. By taking advantage of the sequence-dependent twistability of the DNA helix, Λ-1-Rh(MGP)$_2$phi^{5+} is able to achieve a level of specificity difficult to achieve through a small ensemble of non-covalent contacts.

Subsequent studies with Δ-1-Rh(MGP)$_2$phi^{5+} have demonstrated many similarities in the way in which this complex recognizes its DNA binding site. The consensus binding sites for Δ-1-Rh(MGP)$_2$phi^{5+} and Λ-1-Rh(MGP)$_2$phi^{5+} only differ by one base in the central portion of the recognition sequence: 5'-CATCTG-3' versus 5'-CATATG-3'. Similarly, it is the inner portions (Δ versus Λ enantiomer) of the complexes which vary and the outer portions are identical. This suggests that the guanidinium moieties of Λ-1-Rh(MGP)$_2$phi^{5+} and Δ-1-Rh(MGP)$_2$phi^{5+} might be interacting with their DNA sites in a similar manner. Furthermore, this result paints a modular nature to the manner in which these metal complexes recognize their DNA sites, since the inner core of the complex seems to recognize one piece of the recognition sequence (central base step) and the guanidinium arms appear to be responsible for the outer portion of the recognition sequence.

The modular fashion by which Δ and Λ-1-Rh(MGP)$_2$phi^{5+} recognize their DNA sites suggests a possible mechanism for the design of metal complexes with predictable sequence specificities. It might be possible to combine small molecular fragments with known sequence specifities to construct complexes which recognize

larger DNA sequences. The usefulness of this approach has been demonstrated in the predictable sequence specificity of (Δ,α)-(R,R)-[Rh(Me$_2$trien)phi]$^{3+}$(1). Designs which make use of the guanidinium moiety and/or 1-Rh(MGP)$_2$phi^{5+} have the advantage of possibly targeting sequence-dependent distortions in the DNA.

If both Δ and Λ-1-Rh(MGP)$_2$phi^{5+} recognize their sequences in a similar manner, how is it that Λ-1-Rh(MGP)$_2$phi^{5+} displays approximately a 100-fold increase in binding affinity relative to the Δ-enantiomer? Studies with the metal complexes Rh(GEB)$_2$phi^{5+} and Rh(APB)$_2$phi^{5+} have helped to possibly explain this difference. Both of these complexes are capable of mimicking the recognition seen for Δ-1-Rh(MGP)$_2$phi^{5+}, but are unable to target the sites recognized by Λ-1-Rh(MGP)$_2$phi^{5+}. Rh(GEB)$_2$phi^{5+} was synthesized to investigate the importance of linker arm length between the metal complex and the guanidinium moiety, and Rh(APB)$_2$phi^{5+} was designed to test whether an amino group is capable of carrying out the same recognition as the guanidinium moiety. Both of these molecules have demonstrated that the sequence specificity observed for Λ-1-Rh(MGP)$_2$phi^{5+} is very dependent on the nature of the recognition moiety and its proximity with respect to the hydrophobic core of the complex. Thus suggesting that both the "fit" of Λ-1-Rh(MGP)$_2$phi^{5+} into its recognition site as well as the nature of the recognition moiety is very important. These results suggest that Λ-1-Rh(MGP)$_2$phi^{5+} shows a high degree of shape complementarity with its recognition site, which should result in additional stabilization energy in the binding of this complex. Perhaps only Λ-1-Rh(MGP)$_2$phi^{5+} is properly positioned for this additional stabilization from unwinding.

We have demonstrated how coupling a sequence dependent conformational switch to functional group interactions can result in a high level of sequence specificity. However, our current knowledge of DNA structure is insufficient to be

able to apply this principle with any degree of predictability. Thus, a major obstacle in the design of complexes with high levels of DNA specificity is our inability to predict the structure of DNA. One potential application for the DNA twistability assay, which we developed to test for sequence specific unwinding by Λ-1-$Rh(MGP)_2phi^{5+}$, is that it can be applied to the study of DNA structure in the absence of metal complexes. This assay might be capable of assessing differences in helical twist for individual dinucleotide steps. As our knowledge of DNA structure grows, it should be possible to design molecules which rationally incorporate indirect as well as readout of the DNA sequence.

Current efforts in our laboratory are aimed at developing applications for Δ and Λ-1-$Rh(MGP)_2phi^{5+}$, as well as other metal complexes. The high sequence selectivity, and the slow exchange rates of Δ and Λ-1-$Rh(MGP)_2phi^{5+}$, might make them capable of acting as transcriptional inhibitors. Recent work in our laboratory has shown that Λ-1-$Rh(MGP)_2phi^{5+}$ is capable of sequence specifically inhibiting the Klenow fragment of DNA polymerase I, DNA Sequenase (a modified version of T7 DNA polymerase) and AMV reverse transcriptase (2). Additionally, Δ-2-$Rh(MGP)_2phi^{5+}$, the isomer with both arms pointed away from the DNA when intercalated, has been shown to selectively target tertiary elements of the transcription activation response element (TAR) (3). This RNA segment has been implicated in the amplification process for HIV. An interesting feature of its function is its ability to undergo a conformation change to form a triple helix site in the presence of argininamide (4-8). Δ-2-$Rh(MGP)_2phi^{5+}$ seems capable of inducing this conformational change and subsequently photocleaves at the triple site.

Small molecules which bind in the major groove of DNA appear to offer the best opportunities for designing complexes with high sequence selectivity, and the knowledge obtained from DNA binding proteins seems to be most easily

extrapolated over to major groove binders. Studies with small molecules suggests that a modular approach to recognition can be used to design sequence specific binding agents. The next generation of small molecules should be aimed at testing this notion, and other principles of recognition such as targeting conformational changes in the DNA structure offer untapped potential for the design of highly selective metal complexes.

6.1. References and Notes

1. Krotz, A. H., Hudson, B. P. and Barton, J. K. (1993) *J. Am. Chem. Soc.*, **115**, 12577-12578 (1993)
2. Johann, T.W. and Barton, J.K., unpublished results
3. Lim, A.C. and Barton, J.K., unpublished results
4. Puglisi, J.D., Chen, L., Frankel, A.D. and Williamson, J.R., *Proc. Natl. Acad. Sci. USA*, **90**, 3680-3684 (1993)
5. Frankel, A.D., *Protein Sci.*, **1**, 1539-1542 (1992)
6. Tan, R.Y. and Frankel, A.D., *Biochemistry*, **31**, 10288-10294 (1992)
7. Puglisi, J.D., Tan, R.Y., Calman, B.J., Frankel, A.D. and Williamson, J.R., *Science*, **257**, 76-80 (1993)
8. Tao, J.S. and Frankel, A.D., *Proc. Natl. Acad. Sci.*, **89**, 2723-2726 (1992)