Sequence-Specific DNA Binding by a Rhodium Complex: Recognition Based on Sequence-Dependent Twistability†

Robert H. Terbrueggen and Jacqueline K. Barton*

Division of Chemistry, California Institute of Technology, Pasadena, California 91125

Received March 22, 1995®

ABSTRACT: The chemical construction of small molecules targeted to DNA depends upon the sequence-dependent structure of the double helix. Here we describe a new structural element to be considered in the sequence-specific recognition of DNA, sequence-dependent DNA twistability. The importance of sequence-dependent DNA twistability is demonstrated in the DNA recognition properties of a novel synthetic rhodium intercalator, Λ-1-Rh(MGP)$_2$phi^{5+}. This metallointercalator, containing pendant guanidinium groups, binds in the major groove of DNA 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. Through an assay developed to test for sequence-specific DNA unwinding, a 70 ± 10° unwinding of the sequence 5′-CATATG-3′ is established with specific binding by the metal complex. This sequence-dependent twistability may be an essential feature of the recognition of sequences by DNA-binding proteins and may be exploited in future design.

The rational design of sequence-specific DNA-binding molecules requires an understanding of the principles governing DNA recognition. Crystal structures of DNA−protein complexes have pointed to the importance of both direct and indirect readout mechanisms (Pabo & Sauer, 1992; Steitz, 1990; Harrison & Aggarwal, 1990). One illustration of the direct readout of a DNA sequence is apparent in the crystal structure of the transcription factor Zif268 bound to its 9 base pair (bp) recognition sequence (Pavletich & Pabo, 1991). In this structure, three distinct zinc finger domains act in unison in the major groove of the helix to recognize 5′-GCGTGGGCG-3′ primarily through direct contacts between side-chain functionalities and the six guanines on one strand; five of the six contacts involve a bidentate hydrogen bond between the N7 and O6 of guanine and the guanidinium moiety of an arginine. A remarkable illustration of indirect readout can be seen in the crystal structure of the TATA-box binding protein (TBP) with its consensus sequence (Kim, J. L., et al., 1993; Kim, Y. C., et al., 1993). The protein is bound from the minor groove side to a severely distorted DNA sequence, in which the DNA helix is unwound by nearly 120° over 8 bp and is bent toward the major groove by nearly 100°. Only six hydrogen bonds to 4 TA base pairs are evident, and direct contacts between side-chain functionalities of TBP and the DNA base pairs are insufficient to account for the DNA sequence specificity.

Here we describe the sequence-specific DNA binding of a new metallointercalator, Λ-1-Rh(MGP)$_2$phi^{5+} [MGP = 4-(guanidylmethyl)-1,10-phenanthroline; phi = phenanthrenequinone diimine] (Figure 1). This small, synthetic-complex mimics DNA-binding proteins such as Zif268 and TBP not only in terms of affinity and specificity for its DNA target but also in terms of the principles and strategies utilized in establishing this recognition. Λ-1-Rh(MGP)$_2$phi^{5+}, through intercalation in the major groove, binds at nanomolar concentrations preferentially to a 6 bp sequence, 5′-CATATG-3′. Direct contacts between guanidinium functionalities on the metal complex and guanine residues are found to contribute to the sequence recognition. Importantly, however, we find that recognition through these contacts requires the unwinding of the 6 bp sequence. The overall sequence specificity of the complex therefore relies directly upon a sequence-dependent twistability of DNA.

EXPERIMENTAL PROCEDURES

Materials. The synthesis and purification of Rh(MGP)$_2$phi^{5+} will be described elsewhere. A 330 base pair 5′-γ-^{32}P- or 3′-α-^{32}P-labeled *Acc*I/*Drd*I fragment of pBR322 (New England Biolabs) was prepared by standard methods. 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.

Photocleavage Reactions on 5′- and 3′-^{32}P Restriction Fragments. The ^{32}P end-labeled 330 base pair restriction fragment and rhodium complexes were incubated together in a 1.7 mL siliconized Eppendorf tube at 23 °C for 5 min prior to irradiation with 313 nm light for 8 min. The DNA/Rh ratio was 50:1 with 20 μL irradiation reactions (10 mM

† This work has been supported by the NIH (GM33309), which we gratefully acknowledge.

* To whom correspondence should be addressed.

® Abstract published in *Advance ACS Abstracts,* June 1, 1995.

0006-2960/95/0434-8227$09.00/0 © 1995 American Chemical Society

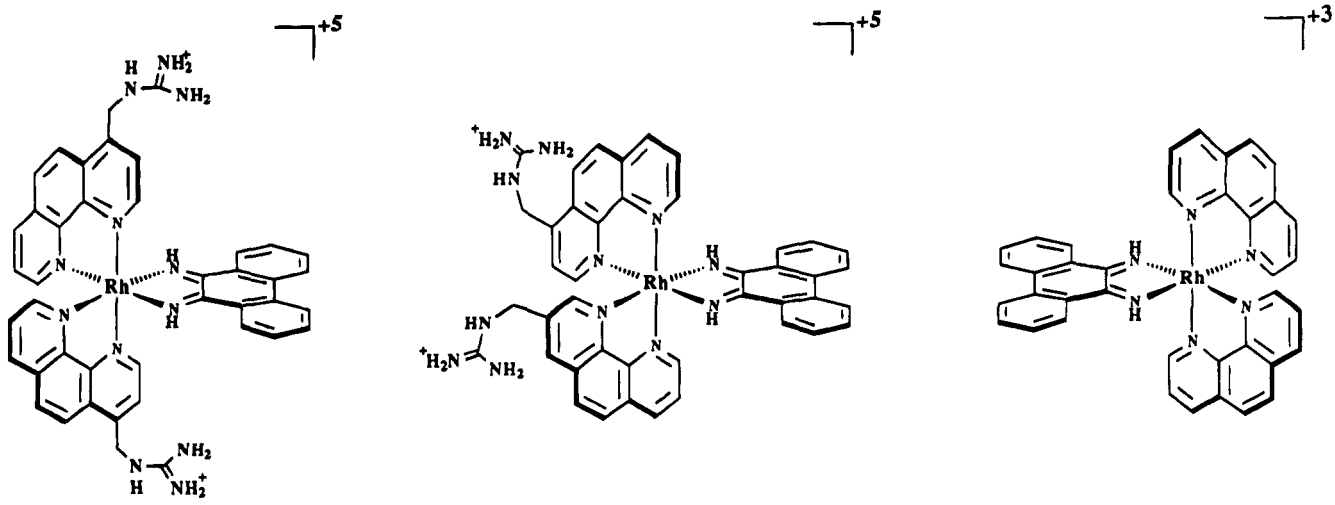


FIGURE 1: Schematic illustrations of Λ-1-Rh(MGP)$_2$phi^{5+} (left), Λ-2-Rh(MGP)$_2$phi^{5+} (middle), and Δ-Rh(phen)$_2$phi^{3+} (right).

sodium cacodylate, 40 mM NaCl, pH 7.0). Samples were ethanol precipitated immediately after irradiation and resuspended in denaturing gel loading dye (10 000 cpm/μL). Following separation on an 8% denaturing polyacrylamide gel, the samples were analyzed using a phosphorimager (Molecular Dynamics).

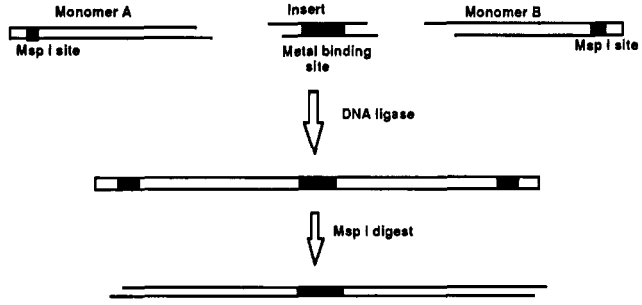
Photocleavage Reactions on the 5′-^{32}P End-Labeled 76 Base Pair Oligonucleotide. A 76 base pair long oligonucleotide, 5′-ACTGACTGCCTAGGTATAAACTTAAGCATAAACATATGCGTACGTATATACGTAGCCACGTGGGTACCACTGACTG-3′, and its complement were synthesized and purified by gel electrophoresis (8% denaturing polyacrylamide gel). The oligonucleotide (both strands) was 5′-^{32}P labeled by standard protocols. Each radiolabeled strand was annealed with its cold complement by incubation at 90 °C for 4 min 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 min prior to irradiation with 313 nm light for 20 min. The DNA/Rh ratio was 50:1 with 20 μL irradiation reactions (10 mM sodium 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). Following separation on an 8% denaturing polyacrylamide gel, the samples were analyzed using a phosphorimager (Molecular Dynamics). Relative cleavage was determined as the sum of intensities at the central T and A sites for Λ-1-Rh(MGP)$_2$phi^{5+} minus the intensity of cleavage by Λ-2-Rh(MGP)$_2$phi^{5+} at the same two sites. This value yields the cleavage enhancement at each site.

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′-TCAGCATATACCCATATGAGCATATGTCCTGA-3′ (A), 5′-TCAGGACATATGCTCATATGGGTATATGCTGA-3′ (B), 5′-TCAGCATATACCCATATGAGCATATGTCCTGA-3′ (C), and 5′-TCAGGACATATGCTCATATGGGTATATGCTGA-3′ (D), where G represents the unnatural oligonucleotide 7-deazaguanine. Photocleavage experiments on double-stranded oligonucleotides AB, AD, BC, and CD were carried out as described above for the 76 base pair long oligonucleotide. 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.

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° bend is encoded by five in-phase 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)$_2$phi^{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 a AT step.

An outline of the procedure for generating full-length substrates is shown in Scheme 1. The outer invariant segments were ligated to opposite ends of the 5′-^{32}P end-labeled spacer inserts. The full-length oligonucleotides (142−148 bp) were then cut with *Msp*I to generate truncated fragments (116−122 bp) with complementary sticky ends. The assay substrates were synthesized as follows using a modified version of the procedure described by Crothers (Koo et al., 1990). Oligonucleotide strands of monomers A

Scheme 1



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 to 23 °C over a 1 h 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 of polynucleotide kinase (Boehringer Mannheim, 10 units/μL) in a 50 μL reaction mixture (70 mM Tris·HCl, pH 7.6, 10 mM $MgCl_2$, and 5 mM dithiothreitol) at 37 °C for 2 h. 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 h. After being labeled with [γ-^{32}P]ATP, 2 μL of 10 mM cold ATP and an additional 10 units of polynucleotide kinase were added. The reaction mixture 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 h, 50 μL of 10 mM EDTA was then added, and the reaction mixture was heated at 65 °C for 30 min 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 h at 37 °C, 10 μL of 10× nondenaturing gel loading dye was added, and the reaction was loaded onto a 6% nondenaturing 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.

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, and 50 μg/mL BSA); Rh(MGP)$_2$phi^{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 of T4 DNA ligase (Boehringer Mannheim; 1 unit/μL) was added. Time points were taken at 0, 1, 5, 10, and 30 min. The aliquots (8 μL) were mixed with 8 μL of 50 mM EDTA solution to quench the reaction. Cyclization reactions were run in both 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.

RESULTS AND DISCUSSION

Λ-1-Rh(MGP)$_2$phi^{5+} was first prepared as a derivative of the parent complex Rh(phen)$_2$phi^{3+} (Pyle et al., 1989; Sitlani et al., 1992) (Figure 1), which contains ancillary guanidinium functionalities for direct interaction with guanines on the DNA helix. Δ-Rh(phen)$_2$phi^{3+} preferentially targets 5′-py-py-pu-3′ sequences through shape selection (Campisi et al., 1994), and Rh(MGP)$_2$phi^{5+} was constructed to target the subset of Rh(phen)$_2$phi^{3+} sites which contained flanking guanines. A variety of phi complexes of rhodium have been prepared previously which bind DNA by intercalation in the major groove (David & Barton, 1993; Collins et al., 1994; Dupureur & Barton, 1994) and which recognize different DNA sites depending upon the symmetry and placement of functional groups on the ancillary, nonintercalating ligands (Sitlani & Barton, 1994; Sardesai et al., 1994; Sitlani et al., 1993; Krotz et al., 1993a). (Δ,α)-(*R*,*R*)-[Rh(Me$_2$trien)phi]$^{3+}$ (Me$_2$trien = 2,9-diamino-4,7-diazadecane), for example, in mimicking the direct readout of DNA sequences by proteins, has been shown to intercalate into the predicted sequence 5′-TGCA-3′ through a mixture of hydrogen bonds and van der Waals contacts (Krotz et al., 1993a,b). Thus far, the phi complexes of rhodium(III) all promote DNA strand cleavage upon photoactivation. For these complexes, as for Rh(MGP)$_2$phi^{5+}, the DNA cleavage products have been consistent with direct abstraction of the C3′ hydrogen atom by the photoexcited phi (Sitlani et al., 1992). This nondiffusible photocleavage chemistry is useful in marking sites of intercalation.

DNA Photocleavage by Rh(MGP)$_2$phi^{5+} Isomers. The sequence selectivity in photocleavage by Rh(MGP)$_2$phi^{5+} isomers, shown in Figure 2, may be understood in part on the basis of the positioning of the guanidinium groups in the complex relative to the intercalative phi ligand. Since 4-(guanidylmethyl)-1,10-phenanthroline is an asymmetric ligand, synthesis of Rh(MGP)$_2$phi^{5+} generates three different positional isomers in a statistical distribution of 1:2:1. The recognition pattern of Rh(MGP)$_2$phi^{5+} complexes is seen to be both isomer-specific and enantiospecific. Cleavage with racemic 2-Rh(MGP)$_2$phi^{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)$_2$phi^{3+}; this result is expected given the positioning of the guanidinium moieties and the increase in charge. In contrast, Λ-1-Rh(MGP)$_2$phi^{5+},

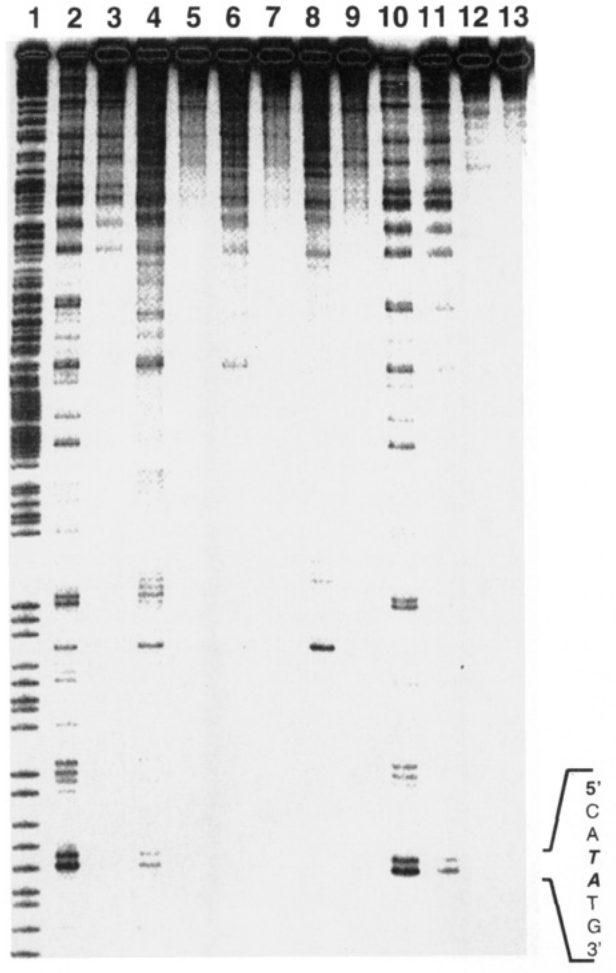
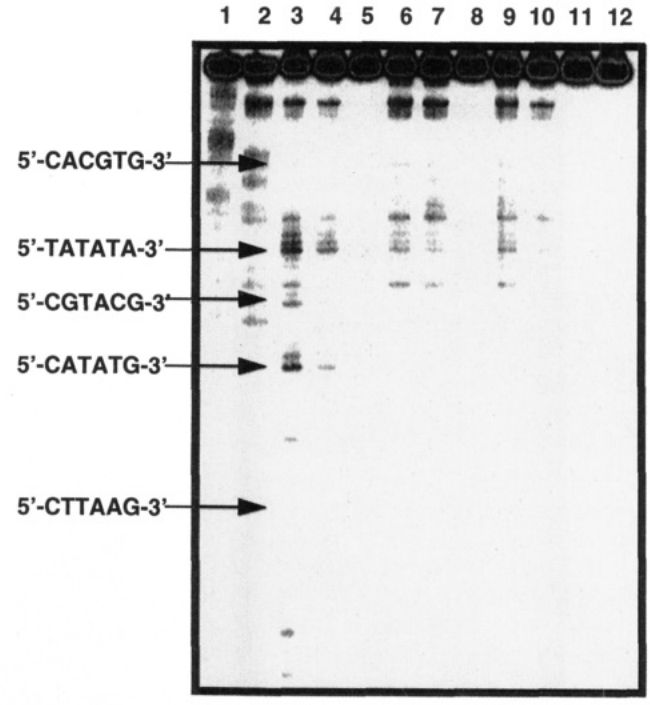


FIGURE 2: Comparison of photocleavage patterns for Rh(phen)$_2$phi^{3+}, 1-Rh(MGP)$_2$phi^{5+}, and 2-Rh(MGP)$_2$phi^{5+} on a 3′-^{32}P end-labeled 325-mer restriction fragment (pBR322 cut by *Acc*I/*Drd*I). Lanes: 1, Maxam−Gilbert A+G reaction; 2 and 3, *rac*-1-Rh(MGP)$_2$phi^{5+} (50 and 5 nM); 4 and 5, *rac*-2-Rh(MGP)$_2$phi^{5+} (50 and 5 nM); 6 and 7, *rac*-Rh(phen)$_2$phi^{3+} (50 and 5 nM); 8 and 9, Δ-1-Rh(MGP)$_2$phi^{5+} (50 and 5 nM); 10 and 11, Λ-1-Rh(MGP)$_2$phi^{5+} (50 and 5 nM); 12, Λ-1-Rh(MGP)$_2$phi^{5+} (50 nM) without irradiation; 13, DNA without metal and with irradiation. The photocleavage pattern is found to vary depending upon the disposition of the guanidinium groups in the complex. The sites targeted are both isomer specific and enantiospecific. Cleavage at low concentrations is seen for Λ-1-Rh(MGP)$_2$phi^{5+} at 5′-CATATG-3′ but not for Δ-1-Rh(MGP)$_2$phi^{5+}, *rac*-2-Rh(MGP)$_2$phi^{5+}, or *rac*-Rh(phen)$_2$phi^{3+}.

the isomer in which both guanidinium groups are pointed over the phi ligand, or when intercalated, toward the DNA, shows a specific pattern of photocleavage which is strong at lower concentrations. On the restriction fragment shown, Δ-1-Rh(MGP)$_2$phi^{5+} specifically targets the sequence 5′-ATCTG-3′ (cleavage at the C); this site represents one of the family of sites targeted by Δ-Rh(phen)$_2$phi^{3+}. The Λ-isomer, however, is responsible for the bulk of sites targeted. All of the sites cleaved by Λ-1-Rh(MGP)$_2$phi^{5+} contain a central 5′-TA-3′ base step, and sequence discrimination is apparent over 6 base pairs. In contrast to both Rh(phen)$_2$phi^{3+} and Δ-1-Rh(MGP)$_2$phi^{5+}, where a single cleavage site is found to the 5′-side of the intercalation site, Λ-1-Rh(MGP)$_2$phi^{5+} shows cleavage at both nucleotides of the intercalation site (Sitlani et al., 1992). Cleavage titrations reveal that Λ-1-Rh(MGP)$_2$phi^{5+} targets 5′-CATATG-3′, the strongest site seen in Figure 2, at concentrations as low as

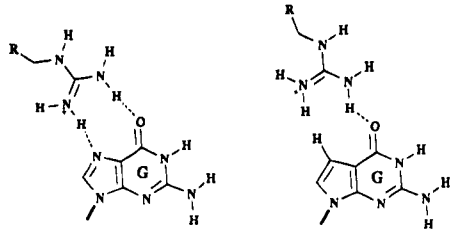


Site	Relative Cleavage %
5'-CATATG-3'	100
5'-CACGTG-3'	0
5'-CGTACG-3'	73
5'-TATATA-3'	62
5'-CTTAAG-3'	16

FIGURE 3: Determination of the consensus sequence targeted by Λ-1-Rh(MGP)$_2$phi^{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); 3−5, Λ-1-Rh(MGP)$_2$phi^{5+} (1.0, 0.1, and 0.01 μM); 6−8, Λ-2-Rh(MGP)$_2$phi^{5+} (1.0, 0.1, and 0.01 μM); 9−11, *rac*-Rh(phen)$_2$phi^{3+} (1.0, 0.1, and 0.01 μM); 12, DNA with 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 is illustrated.

250 pM, more than 2 orders of magnitude lower in concentration compared to Rh(phen)$_2$phi^{3+}.

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 (Figure 3) on a 76-mer DNA duplex which was constructed to contain 14 different sequence variations of 6 bp sites which contain a central 5′-TA-3′ base step. As on the larger restriction fragment, cleavage by Rh(MGP)$_2$phi^{5+} is enantiospecific and isomer specific. Closer examination reveals the heirarchy for sites recognized by Λ-1-Rh(MGP)$_2$phi^{5+}. As evident in Figure 3, 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% 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



	Relative cleavage intensity
5'-TCGA **CATATA** CC **CATATG** AG **CATATG** TCCTGA-3' 3'-AGCT **GTATAT** GG **GTATAC** TC **GTATAC** AGGACT-5'	40
5'-TCGA **CATATA** CC **CATATG** AG **CATATG** TCCTGA-3' 3'-AGCT **GTATAT** GG **GTATAC** TC **GTATAC** AGGACT-5'	61
5'-TCGA **CATATA** CC **CATATG** AG **CATATG** TCCTGA-3' 3'-AGCT **GTATAT** GG **GTATAC** TC **GTATAC** AGGACT-5'	78
5'-TCGA **CATATA** CC **CATATG** AG **CATATG** TCCTGA-3' 3'-AGCT **GTATAT** GG **GTATAC** TC **GTATAC** AGGACT-5'	100

FIGURE 4: 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)$_2$phi^{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. Also illustrated schematically are the oligonucleotides that were synthesized to probe for hydrogen-bonding interactions and a summary of the results. The effects of both single and double deazaguanine (G) substitutions were investigated. Note that 5'-CATATA-3' is cleaved with an intensity close to that of 5'-CATATG-3' and greater than that of 5'-CATATG-3'. Thus one guanidinium may maintain an N7 and O6 contact while the other may contact primarily the N7.

on the consensus site 5'-C(A/G)TA(C/T)G-3', with cleavage at both the internal T and A.

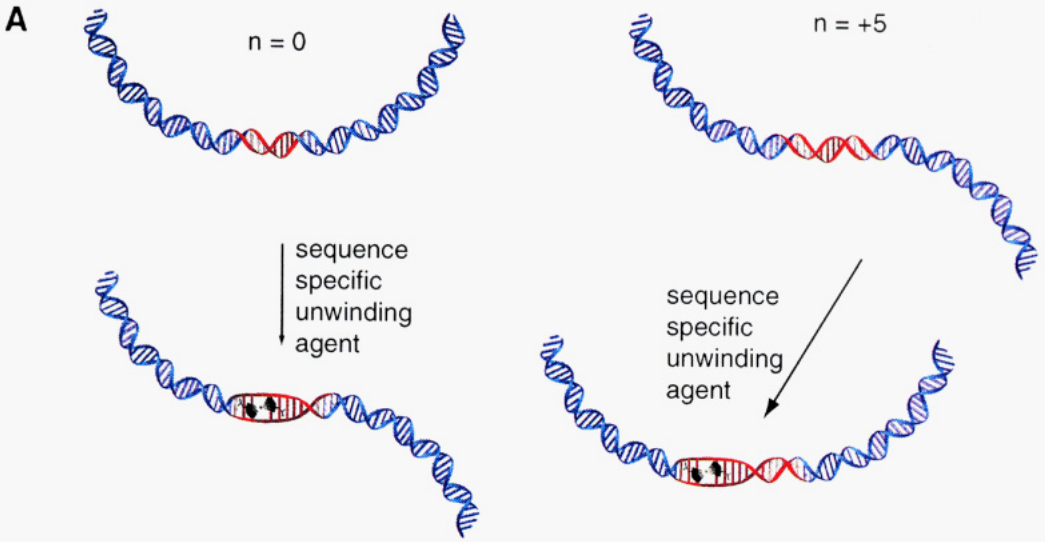
Effects of Deazaguanine Substitution. The differences in recognition between Λ-1-Rh(MGP)$_2$phi^{5+} and Λ-2-Rh(MGP)$_2$-phi^{5+} indicate a participation of the guanidinium groups in the recognition, but to establish where they contact on the DNA helix, we directly compared cleavage of 32 bp oligonucleotides containing the native recognition site and a site with deazaguanine substitutions. In mimicking Zif268, 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 (Jain et al., 1992). Figure 4 summarizes the results of making single- and double-base substitutions. Incorporation of a single deazaguanine (G) in the palindromic sequence 5'-CATAT G-3' results in either 22% or 39% reduction in cleavage,depending upon which guanine is substituted. When deazaguanines are incorporated into 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.

Recognition Based upon Sequence-Dependent Unwinding. Recognition by Λ-1-Rh(MGP)$_2$phi^{5+} is difficult to reconcile simply on the basis of the functionalization of the parent Rh(phen)$_2$phi^{3+}. Λ-1-Rh(MGP)$_2$phi^{5+}, unlike the parent Rh(phen)$_2$phi^{3+}, binds preferentially at sites containing a central 5'-TA-3' base step. Despite the 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+} (David & Barton, 1993), are favored for intercalation into the right-handed DNA helix (Barton, 1986), 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 with 5'-asymmetry. Importantly, molecular modeling of Λ-1-Rh(MGP)$_2$phi^{5+} intercalated into 5'-CATATG-3', constructed with a central intercalation site unwound by 20° (Sitlani et al., 1992; Sitlani & Barton, 1994), reveals that the guanidinium groups on the complex are unable to reach the N7 positions of guanines in the 1 and 6 positions if the DNA base pairs flanking the intercalation step are in the canonical B-form.

These results are understandable in the context of a sequence-dependent twistability of DNA. If the DNA sequence 5'-CATATG-3' is significantly unwound over the 6 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 would place the guanines in the first and sixth positions on the same face of the DNA polymer so that the complex could make both necessary N7 contacts.

Figure 5 illustrates an assay we designed to test for sequence-specific DNA unwinding in solution. This assay was derived on the basis of the elegant assays developed by Crothers and co-workers to test for DNA bending (Koo et al., 1990; Crothers et al., 1992; Kahn et al., 1994; Crothers & Drak, 1992). Our assay also resembles a procedure adopted by Hillen to examine DNA unwinding by the Tet repressor (Niederweis & Hillen, 1993). We synthesized a series of oligonucleotides in which two DNA segments, each with a bend of approximately 100° encoded by five in-phase A_6 tracts, were separated by a spacer sequence containing the 5'-CATATG-3' binding site. The spacer length was varied from 10 to 16 bp (1−1.5 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 areout of phase by half a helical turn, n equals 5. At the ends of the oligonucleotide were placed two base complementary overhangs to allow for the facile intramolecular cyclization of an appropriately bent oligomer. 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.

The results are also given in Figure 5. In the absence of the 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



B

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

C

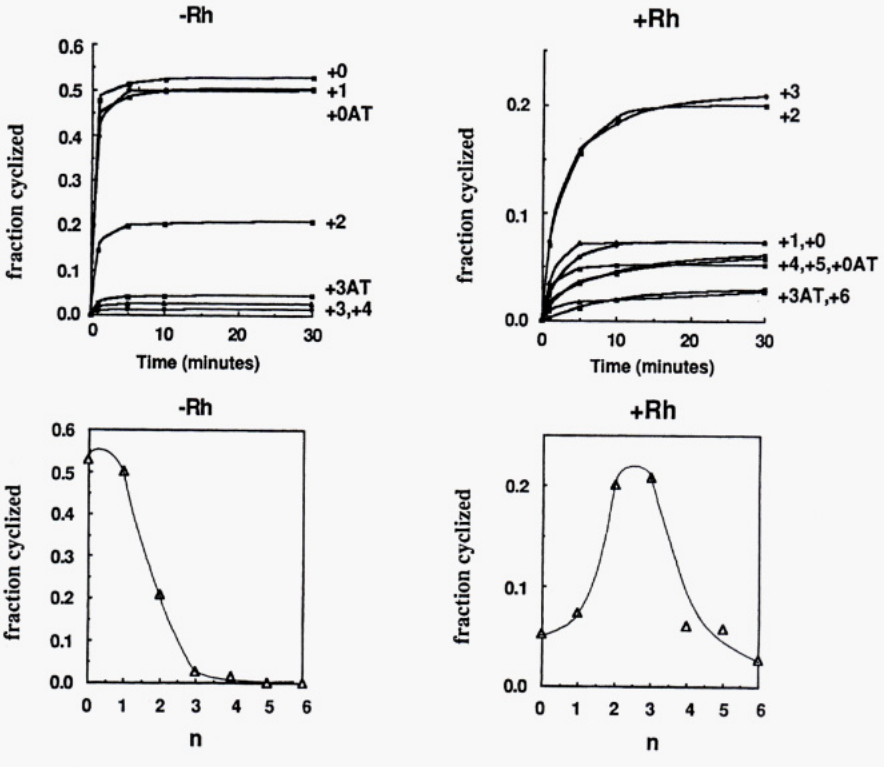


FIGURE 5: Assay for sequence-specific DNA unwinding with Λ-1-Rh(MGP)$_2$phi^{5+}. (A) Schematic illustration of the unwinding assay. An oligonucleotide is constructed which contains two bent segments, each with a bend of approximately 100° encoded by five phased A_6 tracts and a central spacer with an 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. (B) 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. (C) The fraction of intramolecular oligonucleotide cyclization by T4 DNA ligase as a function of time plotted for bent fragments of different spacer insert lengths in the absence (left) and presence (right) of Λ-1-Rh(MGP)$_2$phi^{5+} (1 μM). 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 were compared to a plot of end to end distance versus n for different degrees of unwinding. On the basis of 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°.

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 the 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 or intramolecular cyclization therefore depends upon the relative phasing of the bent segments.

In this assay it is necessary to establish that a single metal complex is bound and is localized to its recognition site. Under the assay conditions, photocleavage using Λ-1-Rh(MGP)$_2$phi^{5+} was carried out on two oligonucleotides, 3TA, which contains the binding site, and 3AT, which differs as a result of a two-base inversion to the sequence 5′-CAATTG-3′. The experiment revealed sequence-specific cleavage only on 3TA within the 5′-CATATG-3′ binding site, with cleavage as expected at the central T and A. To establish further that the unwinding was only associated with the binding site, we compared ligation of the two oligonucleotides. 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. Analyses of the amounts of cyclization for the series of fragments containing the recognition site 5′-CATATG-3′ 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 its site, unwinds the oligonucleotide by 70 ± 10°. Figure 6 illustrates a structural model for Λ-1-Rh(MGP)$_2$phi^{5+} bound to its DNA-binding site based on these data.

Comparisons in Recognition to DNA-Binding Proteins. 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′ is unwound similarly by 85° (Kim, J. L., et al., 1993; Kim, Y. C., et al., 1993). Λ-1-Rh(MGP)$_2$phi^{5+}, however, unlike TBP, does not promote DNA bending. Using an assay similar to that established by Crothers (Koo et al., 1990), in which we synthesized a series of DNA fragments with different numbers of phased 5′-CATATG-3′, we were unable to detect a bend in the DNA as a result of complex binding. This lack of bending is not surprising given that TBP binds in the minor groove and bends the DNA toward the major groove, while Λ-1-Rh(MGP)$_2$phi^{5+} is bound in the major groove; any tendency of the sequence to bend toward 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° 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 (Somers & Phillips, 1992; Otwinosksi, 1988; Kim, J. L., et al., 1993; Kim, Y. C., et al., 1993). Work on enzyme hypersensitivity of TA-rich regions also supports this idea (Drew et al., 1985). 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 (Geierstanger et al., 1993).

Implications for the Rational Design of Sequence-Specific DNA-Binding Molecules. 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 kcal 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

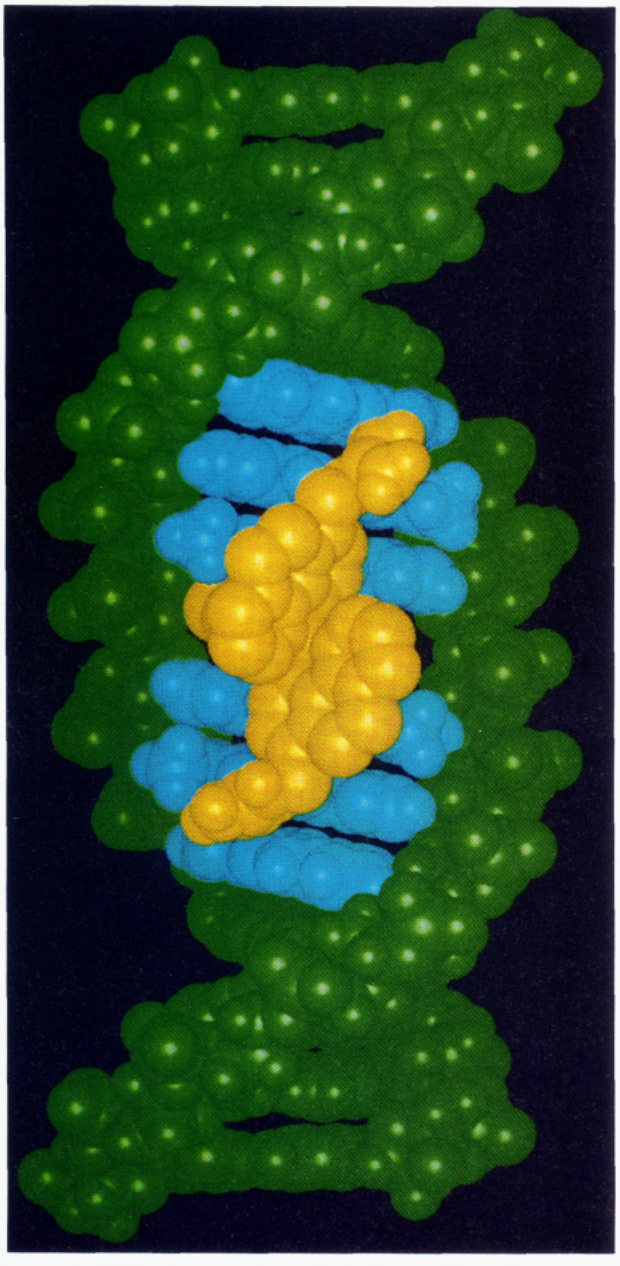


FIGURE 6: Molecular modeling of Λ-1-Rh(MGP)$_2$phi^{5+} bound to its recognition site, 5′-CATATG-3′. Λ-1-Rh(MGP)$_2$phi^{5+} (yellow) is shown intercalated in the major groove into the central 5′-TA-3′ base step of the DNA duplex 5′-GGCG*CATATG*CGGG-3′. The 6 bp complex binding site, 5′-CATATG-3′, is 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°, the central TA step is unwound by 34°, and the flanking base steps are unwound by 9° each. It is noteworthy that the complex can span all 6 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).

vary both structurally and dynamically so as to incorporate more rationally this notion of sequence-dependent twistability into design.

REFERENCES

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

Campisi, D., Morii, T., & Barton, J. K. (1994) *Biochemistry 33*, 4130−4139.

Collins, J. G., Shields, T. P., & Barton, J. K. (1994) *J. Am. Chem. Soc. 116*, 9840−9846.

Crothers, D. M., & Drak, J. (1992) *Methods Enzymol. 212*, 46−71.

Crothers, D. M., Drak, J., Kahn, J. D., & Levine, S. D. (1992) *Methods Enzymol. 213*, 3−29.

David, S. D., & Barton, J. K. (1993) *J. Am. Chem. Soc. 115*, 2984−2985.

Drew, H. R., Weeks, J. R., & Travers, A. A. (1985) *EMBO J. 4*, 1025−1032.

Dupureur, C. M., & Barton, J. K. (1994) *J. Am. Chem. Soc. 116*, 10286−10287.

Geierstanger, B. H., Dwyer, T. J., Bathini, Y., Lown, J. W., & Wemmer, D. E. (1993) *J. Am. Chem. Soc. 115*, 4474−4482.

Harrison, S. C., & Aggarwal, A. W. (1990) *Q. Rev. Biophys. 59*, 933−969.

Jain, S. K., Inman, R. B., & Cox, M. M. (1992) *J. Biol. Chem. 267*, 4215−4222.

Kahn, J. D., Yun, E., & Crothers, D. M. (1994) *Nature 368*, 163−166.

Kim, J. L., Nikolov, D. B., & Burley, S. K. (1993) *Nature 365*, 520−528.

Kim, Y. C., Geiger, J. H., Hahn, S., & Sigler, P. B. (1993) *Nature 365*, 512−520.

Koo, H. S., Drak, J., Rice, J. A., & Crothers, D. M. (1990) *Biochemistry 29*, 4227−4234.

Krotz, A. H., Hudson, B. P., & Barton, J. K. (1993a) *J. Am. Chem. Soc. 115*, 12577−12578.

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

Niederweis, M., & Hillen, W. (1993) *Electrophoresis 14*, 693−698.

Otwinowksi, Z., Schevitz, R. W., Zhang, R. G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F., & Sigler, P. B. (1988) *Nature 335*, 321−329.

Pabo, C. O., & Sauer, R. T. (1992) *Annu. Rev. Biochem. 61*, 1053−1095.

Pavletich, N. L., & Pabo, C. O. (1991) *Science 252*, 809−817.

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

Sardesai, N. Y., Zimmerman, K., & Barton, J. K. (1994) *J. Am. Chem. Soc. 116*, 7502−7508.

Sitlani, A., & Barton, J. K. (1994) *Biochemistry 33*, 12100−12108.

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

Sitlani, A., Dupureur, C. M., & Barton, J. K. (1993) *J. Am. Chem. Soc. 115*, 12589−12598.

Somers, W. S., & Phillips, S. E. V. (1992) *Nature 359*, 387−393.

Steitz, T. A. (1990) *Q. Rev. Biophys. 23*, 205−280.

BI950651L