Structural Basis of DNA Recognition by Synthetic Ligands

Thesis by:

Clara Louise Kielkopf

Advisor:

Professor Douglas C. Rees

In Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy

California Institute of Technology

Pasadena, California

2000

(Submitted August 17, 1999)

Acknowledgments

First I would like to thank my graduate adviser, Prof. Douglas C. Rees. I am constantly amazed by Prof. Rees's excellent crystallographic knowledge, and am very thankful for his help determining and analyzing these structures, as well as for his patient explanations.

The crystallographic studies of polyamide-DNA complexes were carried out in collaboration with Prof. Peter B. Dervan and the members of his group, Eldon Baird, Sarah White, James Turner, Jason Szewczyk, Ryan Bremer and David Herman. I am very appreciative of Prof. Dervan's incredible support and enthusiasm. Eldon has been the best of collaborators who never forgets the big picture; he and Sarah are amazing scientists. I am very grateful to them for the exciting projects and flashes of humor ☺.

Likewise, the crystallographic studies of metallointercalator-DNA interactions were determined in collaboration with Prof. Jacqueline K. Barton and members of her group, Kitty Erkkila and Brian Hudson. I am grateful for having experienced Prof. Barton's infectious enthusiasm for science, and regard her as model of excellence. I thank Kitty for her cheerfulness and help. I especially thank Sheila David and Leemor Joshua-Tor for their efforts to determine the structure of a different intercalator-DNA complex, and hope that one day this crystallographic mystery may be solved.

I thank past and present members of the Rees group for their assistance with structure solutions, especially Hermann Schindelin, Caroline Kisker, Cathy Drennan, Tina Iverson, Andy Yeh, Rob Spencer, Art Chirino, Michael Williamson and Phoebe Ray. I would also like to thank Joseph Wedekind and Prof. David B. McKay for their help. I am especially grateful to Joe for his excellent and caring crystallographic advice, and look forward to a time when we can be together.

I would like to thank the members of my thesis committee, Prof. Richard W. Roberts, Prof. Jacqueline K. Barton, Prof. Peter B. Dervan and Prof. Douglas C. Rees for

their time and guidance. I would also like to thank Prof. Pamela J. Bjorkman for her support and advice. In addition, I need to thank Prof. John W. Brown, for introducing me to biochemistry at the University of Louisville.

Finally, I would like to thank my family- my mother for fun phone conversations, my grandma for her sweetness (fudge), and especially my father, who has always been there for me. Thank you!

Abstract

The DNA double helix presents functional groups in the major and minor grooves that can be used by ligands for readout of the base pair sequence. The overall flexibility and shape also distinguishes various sequences. Although the information displayed in the major groove is more diverse, small DNA-binding polyamides predictably distinguish all four base pairs in the minor groove. In this work, the structural basis of this recognition has been studied using x-ray crystallographic techniques, and is described for G•C (Chapter 2) and T•A base pairs (Chapter 3-4). The polyamides directly read the DNA sequence using a combination of specific hydrogen bonds and shape selection. A second class of ligands called intercalators bind to DNA via the major groove by slipping a planar aromatic ligand between the bases. Although this shape-selective DNA binding is relatively nonspecific, the substitution of the ancillary ligands on octahedral metallointercalators allows specific sequences to be recognized in the major groove. The high resolution crystal structure of a designed metal complex bound to its target site reveals the nature of these specific contacts, as well as the precise stacking of the ligand between the bases. A detailed understanding of specific DNA recognition by small molecules is important for their further development as tools for molecular biology and medicine. These structures reveal how synthetic ligands can distinguish all four base pairs in both the major and the minor grooves of DNA.

Table of Contents

List of Figures

Chapter 1:

Chapter 2:

Chapter 3:

List of Tables

Chapter 1

Introduction

DNA (deoxyribonucleic acid) is considered by many to be a structurally uninteresting polymer composed of repetitive nucleotides. Nevertheless, DNA is the genetic material of life on earth, and as such, demands our attention. Nature's protein machinery recognizes DNA sequences to perform the functions of replication and transcription. Small synthetic molecules that would permeate living cells and specifically manipulate DNA-binding proteins or nucleic acid would be useful tools for molecular biology and medicine. To achieve this goal, the structural features of DNA and its complexes with ligands must be understood.

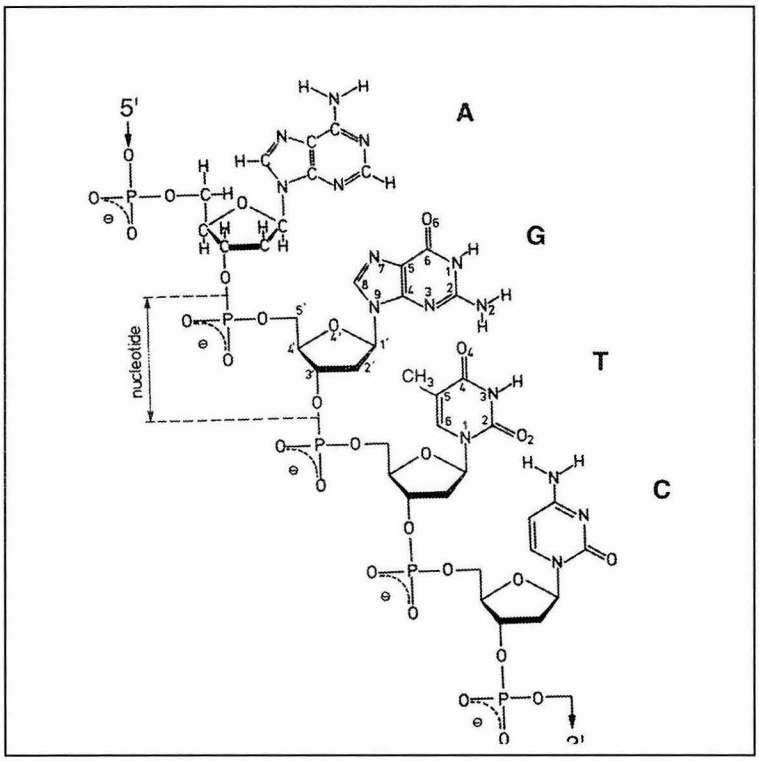
1.1 DNA Structure

The chemical structure of DNA is a linear sequence of nucleotides, monomeric units containing deoxyribose, a base and a phosphate, with a directionality defined by the 5' and 3' sides of the sugar moiety (Figure 1.1a). Four different types of nucleotides are used in biological DNA, distinguished by base, adenine (A), guanine (G), cytosine (C) or thymine (T). In the DNA strand, these aromatic bases prefer to stack upon one another, thereby excluding water and participating in favorable π-interactions, and form hydrogen bonds with other DNA molecules via the functional groups on the base edges (Saenger, 1984). When DNA is packaged in the cell, the nucleotides A and T, or G and C, usually pair to form a double helix (Figure 1.1b) with the bases hung like rungs on a ladder between the phosphodiester backbones. However, other base pairs and higher order interactions important for biological function also occur, and the resulting structures are far from boring. These include the biologically common B-form described above, the A-form that occurs in dehydrated conditions and in RNA-DNA hybrids, the inverse Z-form of some GC-rich sequences, and multi-stranded structures (Figure 1.2) (Rich, 1993; Schwartz *et al.*, 1999).

The flexible torsion angles of the sugar-phosphate backbone (Figure 1.3a) and the variety of hydrogen bond donors and acceptors on the edges of the bases (Figure 1.1) allows these many different structures to form, depending on the sequence, salt, pH and other conditions (Saenger, 1984; Rich, 1993; Hartmann & Lavery, 1996). In particular, the DNA conformation is sometimes indicated by the torsion angles of the deoxyribose sugars, which prefer to be nonplanar, or puckered, with one or two of the atoms displaced from the plane formed by the other three or four. Sugar puckering modes are named after the atom that is displaced the most, followed by either *endo* if on the 5' side of the furanose, or *exo* if on the 3' side (Figure 1.3b). For example, the B-form of a DNA

duplex favors the C2'-*endo* sugar conformation, whereas the A-form is commonly C3'-*endo* (Altona & Sundaralingam, 1972).

Figure 1.1. **(a)** Chemical structure of DNA. The four different nucleotide monomers of DNA, adenosine (A), guanosine (G), cytidine (C) and thymidine (T) are linked by phosphodiester bonds. The convention for directionality of the chain is from the 5' oxygen of the sugar to the 3', as indicated by the arrows (Saenger, 1984).



(b) Watson Crick base pairs commonly found in double-stranded DNA at neutral pH. Top, A•T, bottom, G•C.

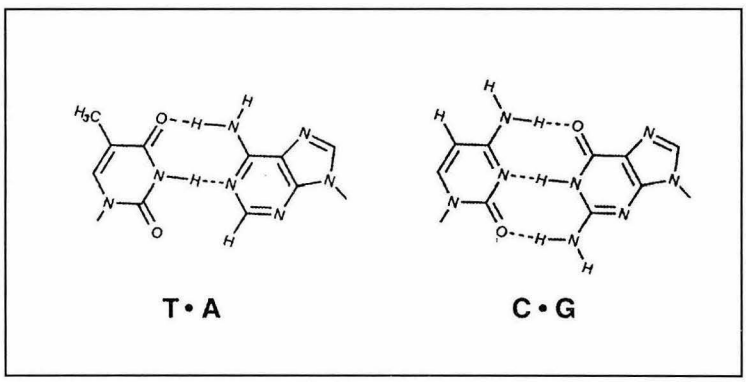


Figure 1.2. Examples of the wide range of possible DNA structures. From left to right, idealized B-form, A-form, Z-form and triple helical DNA oligonucleotides. Coordinates were generated using the program *Insight II* (BIOSYM / Molecular Simulations).

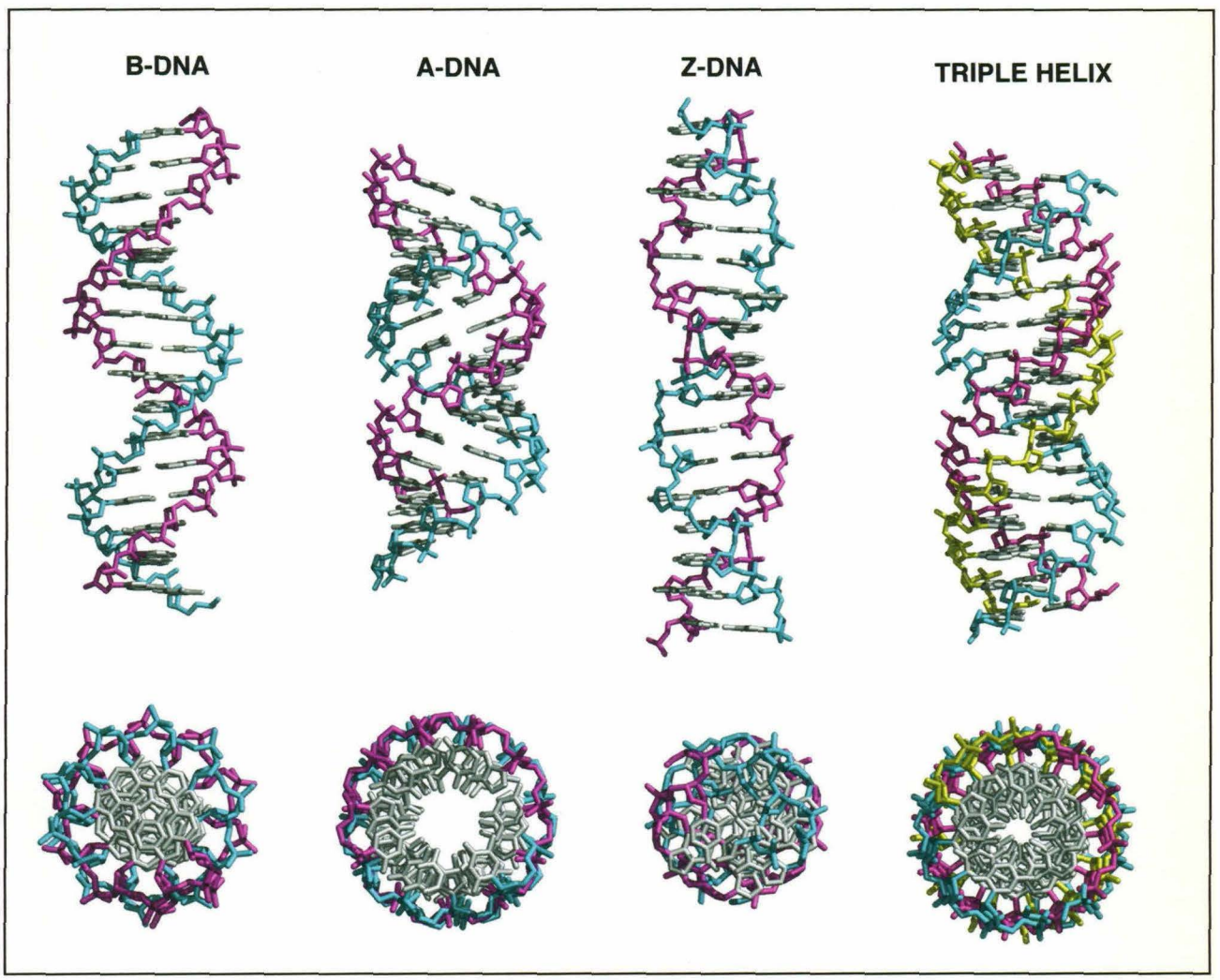
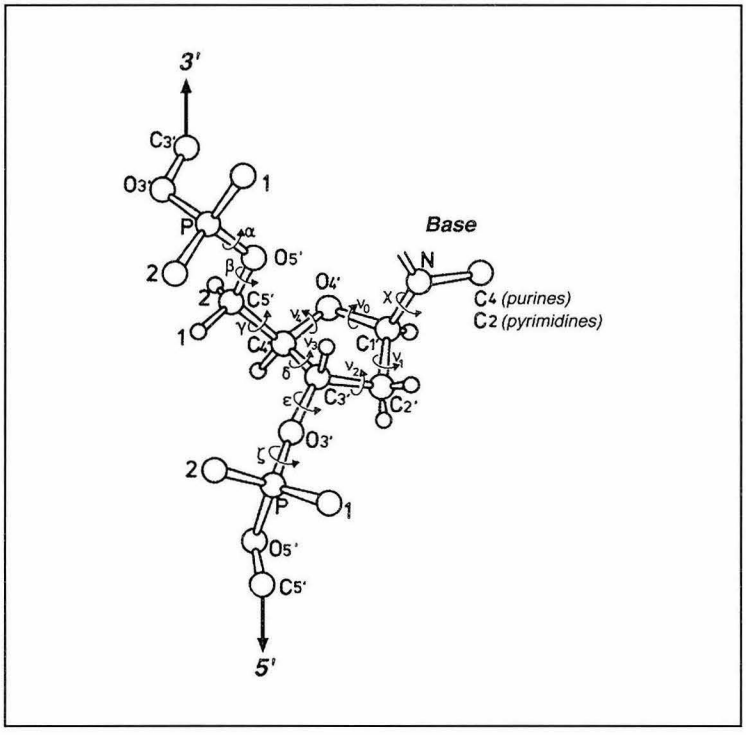


Figure 1.3. **(a)** Definition of the torsion angles of a polynucleotide (Saenger, 1984, Hartmann & Lavery, 1996). The torsion angle refers to the orientation of flanking atoms about the central bond identified by the arrow and the Greek symbol.



(b) Schematic diagram of sugar puckering modes (Saenger, 1984; Altona & Sundaralingam, 1972). Far left, hypothetical planar diagram with C1'-O4'-C4' plane hatched. Middle, C3'-*endo*. Far right, C2'-*endo*.

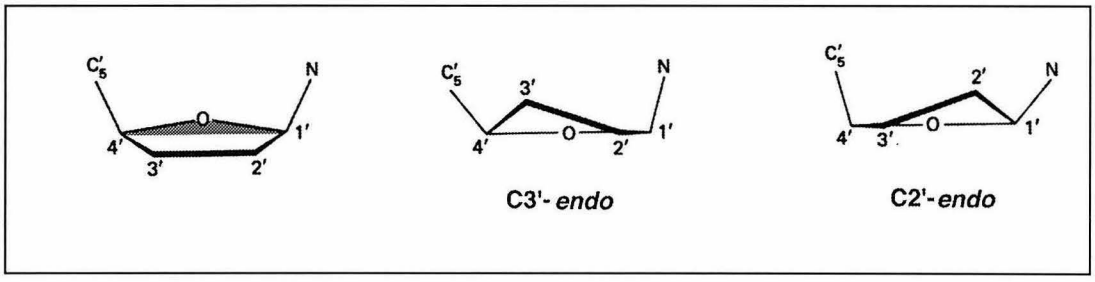
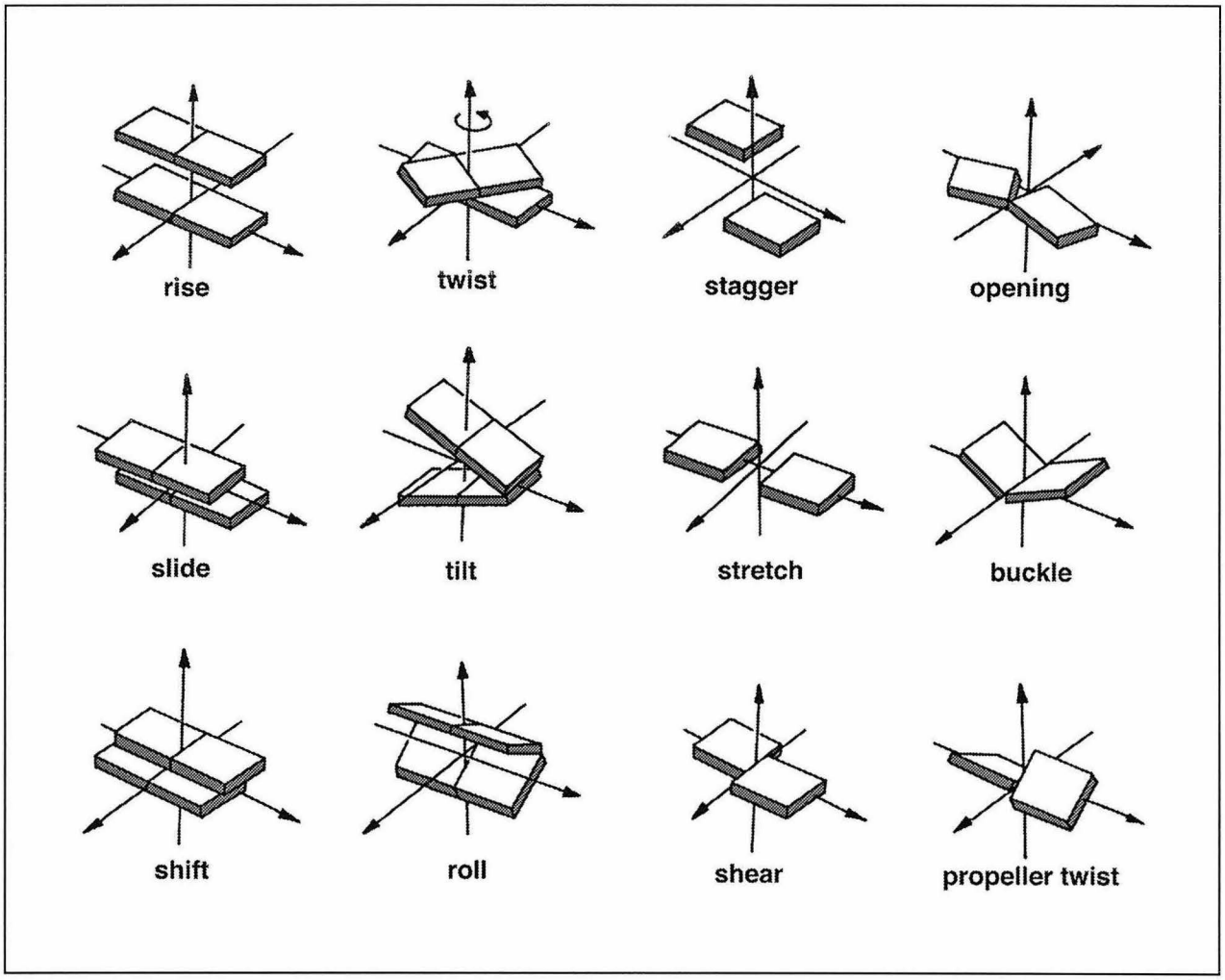


Figure 1.4. Schematic diagram of some of the most commonly used helical parameters for nucleic acids. The arrows indicate the coordinate system of the helix. The right-to-left diagonal points into the minor groove, whereas the left-to-right vector is between the phosphodiester backbones of the duplex, and the vertical arrow is the helix axis. The movement portrayed in each diagram corresponds to the positive sense of the parameter shown, as implemented by the program Curves (Lavery & Sklenar, 1988).



Other parameters have been devised to measure the detailed conformation of double helical DNA (Lavery & Sklenar, 1988) (Figure 1.4). Translational and rotational properties considering each base pair to be a single entity can be measured between consecutive pairs. The most straightforward of these are the helical twist and rise per nucleotide unit along the helical axis, respectively. These are charactistic of the DNA conformation (Saenger, 1984; Hartmann & Lavery, 1996), as illustrated in Table 1.1. More subtle parameters of this type include the following: *slide* towards the phosphodiester backbone, *shift* into the DNA grooves, *tilt* angle between the pairs towards the backbone, and *roll* angle between the base pairs toward the DNA grooves. Parameters between the individual bases in a pair are also measured. Translational properties include *stagger* between the bases along the helical axis, *stretch* toward the backbone, and *shear* into the grooves, whereas rotational ones consist of *opening* of the base pair toward the DNA grooves around the helix axis, *buckle* cupping of the bases and *propeller twist* along the long axis between the bases. The width of the DNA grooves also varies with the overall conformation, with an exposed, shallow minor groove and narrow, deep major groove for A-form nucleic acids, as opposed to B-DNA, in which the major groove is wider and more accessible that the minor groove (Table 1).

Table 1.1. Some Typical Helix Parameters for A- and B-type DNA.

			Major Groove (Å)		Minor Groove (Å)			
	Twist (°)	Rise (Å)	Width	Depth	Width	Depth	Sugar Pucker	Torsion χ (°)
A-DNA	33	2.8	3.0	13.5	11.0	3.0	C3'-*endo*	-160
B-DNA	36	3.4	12.0	8.5	5.5	7.5	C2'-*endo*	-100

1.2 Recognition of DNA

These characteristic shapes of the grooves of double helical DNA are recognized in cells using complementary protein surfaces. In addition to governing overall shape and flexibility, the base pairs display potential hydrogen bond donors and acceptors in the helical grooves. In double helical DNA at neutral pH, these duplexes are usually formed by the Watson-Crick base pairs G•C and T•A (Figure 1.5 & 1.6). For direct sequence readout, proteins match side chains with these functional groups of the target site, whereas shape selectivity depends upon the overall fold or flexibility of the macromolecules resulting in complementary surfaces. Both of these are fundamental aspects of DNA recognition, used for example by transcription factors (Rhodes *et al.*, 1996) and for high fidelity DNA replication (Matray & Kool, 1999).

The presence of only four common base pair combinations (G•C, C•G, A•T or T•A) in double helical DNA has created the misleading impression that a sequence-dependent structural code would be easier to define for DNA than proteins. Although the Watson-Crick base pairing of two DNA strands can often be predicted, after more than a decade of oligonucleotide structures, the flexible phosphodiester backbone and sensitivity to the solution environment has largely aggravated such attempts to predict detailed aspects of DNA structure (Grzeskowiak, 1996). DNA structure is far from static and often adapts to recognize the protein or other ligand. For example, the minor groove of duplex DNA splays open and bends 100° to bind TATA-binding protein (Burley, 1996), and small DNA aptamers have been selected *in-vitro* to bind amino acids and other small molecules (Lin & Patel, 1998).

Before the first structure of a DNA-ligand complex had been determined, it was predicted that most sequence-specific recognition would occur in the major groove of B-form DNA rather than the minor (Seeman *et al.*, 1976). This belief was based upon both shape and hydrogen bonding considerations; the width of the major groove is

approximately the same as the diameter of a protein α-helix, and the variety of functional groups is greater in the major groove. In fact, to a first approximation the base pairs have twofold symmetry in the minor groove, so that theoretically only a binary recognition code would be possible (Figure 1.5 & 1.6). Consequently, it was predicted that A•T/T•A could be distinguished from G•C/C•G by the guanine exocyclic amine, but discrimination of all four base pairs in the minor groove would be unlikely.

With many protein-DNA cocrystal structures now determined (Steitz, 1990; Nelson, 1995) (Figure 1.7), it has been confirmed that usually sequence-specific contacts are formed with the major groove of the double helix. Therefore, the major groove has become the primary target of strategies to decipher an amino acid-base pair code for the design of peptides with predetermined sequence preferences and prediction of the highest affinity DNA-binding site from primary amino acid sequence. Although there has been limited success in the case of *in vitro* selection using zinc fingers (Choo & Klug, 1997; Elrod-Erikson *et al.*, 1998), there is still no simple code for targeted recognition of DNA using proteins.

Apart from proteins, some types of small molecules also bind to DNA (Figure 1.8). Comparatively speaking, many of these small ligands interact primarily with the minor groove (Geierstanger & Wemmer, 1995), perhaps because of a better fit than that of protein secondary structure and little evolutionary pressure for sequence specificity. Often in combination with groove binding, some types of DNA-binding molecules insert an aromatic ligand between the bases (Berman & Young, 1981), a strategy that is employed by proteins as well (for examples, see Burley, 1996; Swaminathan, K. *et al.*, 1997; Ohndorf *et al.*, 1999). For example, actinomycin-D intercalates preferentially at 5'-CG-3' sequences with peptide groups positioned in the minor groove (Takusagawa *et al.*, 1982), whereas arugomycin threads the DNA duplex via intercalation, so that its oligosaccharides bind in both the major and minor grooves (Searle *et al.*, 1992). The molecules distamycin-A and netropsin bind entirely in the minor groove of AT-tracts

without intercalation (Zimmer & Wahnert, 1986). In addition to binding as a monomer, the uncharged distamycin-A displays a second, dimeric binding mode (Pelton & Wemmer, 1989). The compound mithramycin also binds as a dimer in the minor groove, but prefers GC-rich sequences and distorts the DNA structure (Sastry & Patel, 1993). The simple distamycin-A composition of repeating aromatic amino acids made it an attractive starting point for design of a sequence-specific small molecule that could be used to manipulate DNA-based chemistry by larger proteins within the cell.

Figure 1.5. Watson-Crick base pairs of duplex DNA. A schematic model for the minor groove is indicated below the base pairs, and the major groove is above. The minor groove is approximately twofold symmetric due to the hydrogen bond acceptors on each of the bases, and the location of the guanine exocyclic amine close to the center of the groove, whereas all four base pairs can easily be distinguished from the major groove.

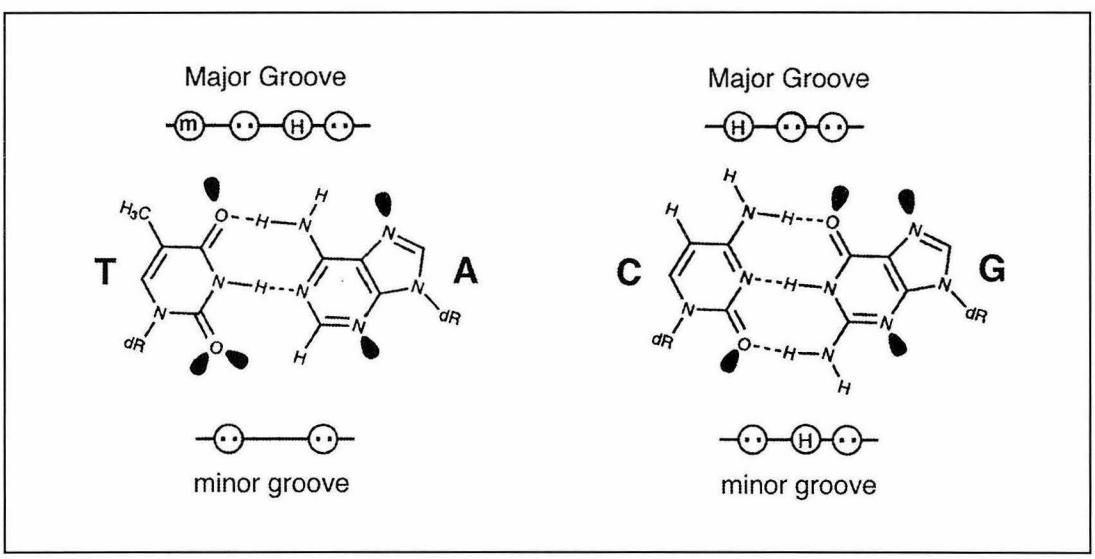


Figure 1.6. Protein's view of the grooves displayed by the base pairs in B-form DNA. The nitrogen atoms are shown in blue, oxygens red and hydrogens are colored white. (**a**) Top left, minor groove of AT base pairs. The two hydrogen bond acceptors are Thy-O2 and Ade-N3. Top right, major groove of AT base pairs. In addition to the hydrogen bond acceptors Thy-O4 and Ade-N7, the thymine methyl group and the hydrogen bond donor Ade-N6 break the twofold symmetry. The methyl group of thymine is shown in yellow. (**b**) Bottom left, minor groove of GC base pairs. The hydrogen bond acceptors are similar to those of the AT minor groove, Cyt-O2 and Gua-N3, but are differentiated by the hydrogen bond donor Gua-N2 located close to the center of the groove. Bottom right, the GC major groove has an asymmetric distribution of hydrogen bond donors (Cyt-N4) and acceptors (Gua-N7 and Gua-O2).

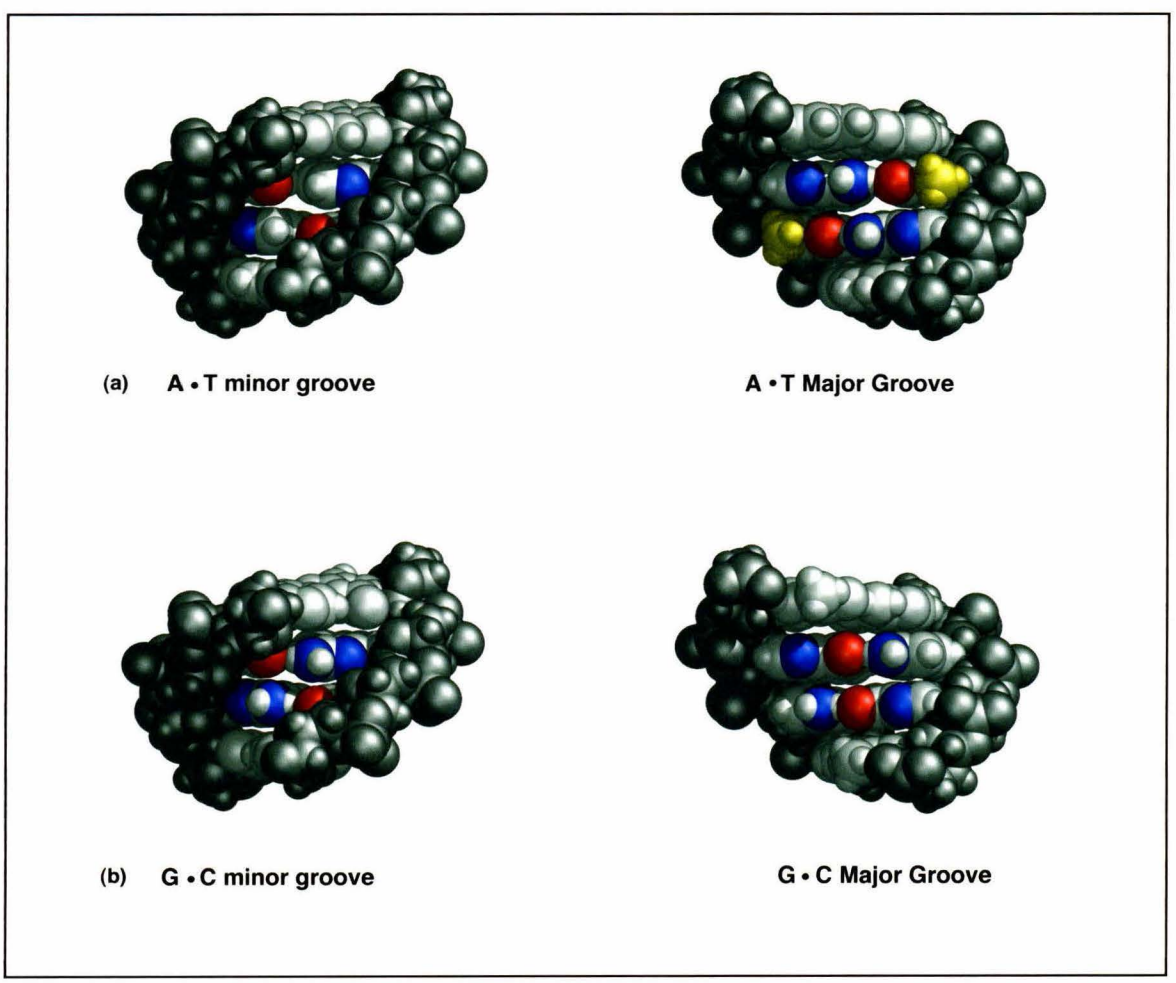


Figure 1.7. Some examples of protein-DNA cocrystal structures. Although the folds of the DNA binding domains differ, the sequence specific contacts are usually with the major groove.

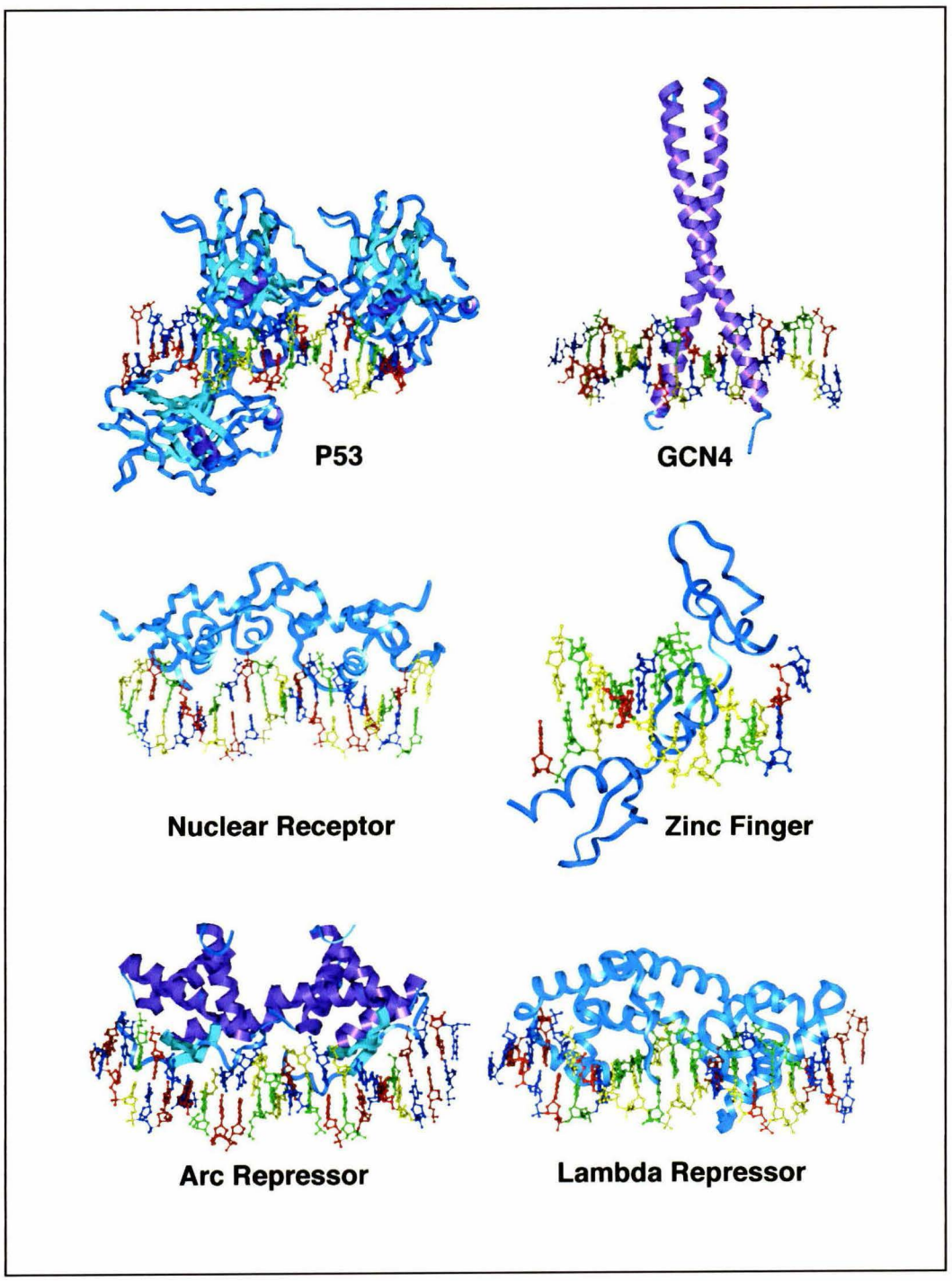
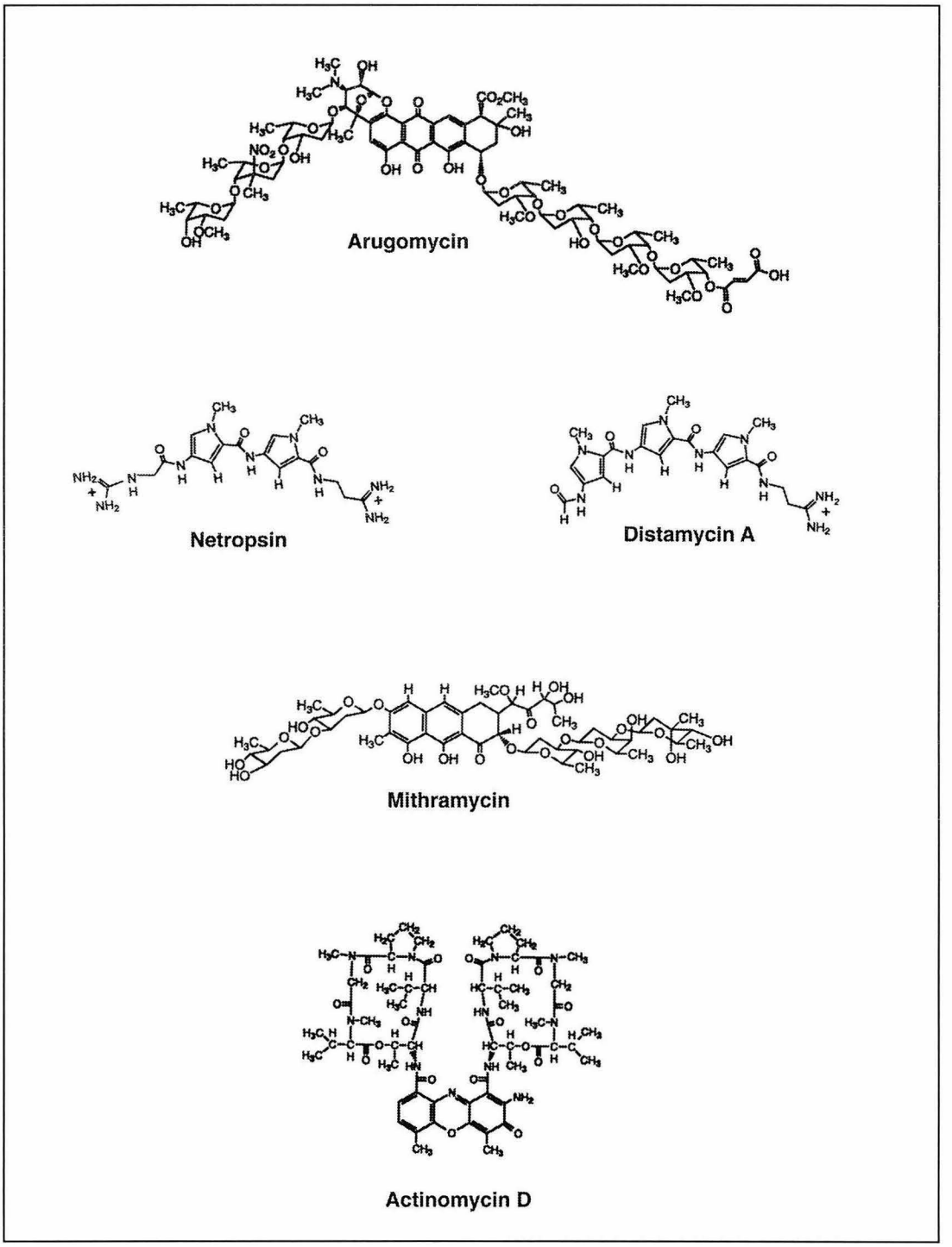


Figure 1.8. Selected examples of naturally-occurring, small DNA ligands.



1.3 Minor Groove Ligands: Polyamides

The natural product distamycin A was rationally modified by one-atom substitution of a nitrogen for the C3 carbon of one of the N-methyl-pyrrole (Py) residues, with the intention of changing the sequence-specificity of that residue from an A•T to G•C base pair. Surprisingly, the resulting synthetic polyamide ImPyPyDp (Im, N-methyl-imidazole-carboxamide; Dp, dimethylpropylamide) displayed a strong preference to bind a five base pair sequence 5'-WGWCW-3' (where W represents either A or T) containing two G•C base pairs (Wade *et al.*, 1992). This sequence preference combined with affinity cleaving data suggested that the compounds were binding in an antiparallel dimeric mode previously characterized for distamycin, with Im/Py pairs of polyamide residues targeting the G•C base pairs and Py/Py bound opposite A•T or T•A. Subsequent structural studies confirmed this 2:1 binding model, showing that the Im was opposite G, whereas Py bound on the C side of the base pair (Mrksick *et al.*, 1992).

To test the generality of this sequence preference, the heterodimer ImPyPyDp/PyPyPyDp was determined to bind to the predicted sequence 5'-WGWWW-3' (Mrksich & Dervan, 1993). Furthermore, the compound ImPyImPyDp was successfully designed to bind the sequence 5'-WGCGCW-3', composed entirely of GC base pairs (Geierstanger *et al.*, 1994; Mrksich & Dervan, 1995). The preference of the Dp tail for AT-rich sequences explains the requirement for flanking AT base pairs on the designated target sites. In spite of the presumed similarity among the base pairs in the minor groove, a two-letter amino acid code to predictably recognize G•C, C•G and A•T/T•A base pairs had been successfully deciphered using simple compounds in the minor groove.

These pairing rules have been expanded and improved using aliphatic linkers to synthesize a variety of polyamide topologies and motifs (Figure 1.9). A β-alanine residue (β) preceding the C-terminal Dp enhances DNA-binding affinity and specificity (Parks *et al.*, 1996) and displays a strong preference for A•T/T•A flanking sequences

(Swalley *et al.*, 1999). Primarily, the C-terminal β residue allows use of Boc-β-Pam resin for solid phase synthesis, producing large quantities of polyamides for chemical and biological studies (Baird & Dervan, 1996). Inserting a β/β pairing allows the binding site size to be extended beyond the natural limits of aromatic polyamides (Trauger *et al.*, 1996). In addition, β/Py and β/Im pairings specify A•T/T•A and C•G, respectively (Turner *et al.*, 1998). Linking the polyamides in a hairpin conformation using γ-aminobutyric acid (γ) increases affinity and specificity by several orders of magnitude (Mrksich *et al.*, 1994). Additional restriction of the topology into a cyclic motif further increases affinity, but has the drawback of reduced specificity (Cho *et al.*, 1995). However, addition of a chiral positive charge to the linker by use of diaminobutyric acid (daba) improves both specificity and affinity of hairpin (Hermann *et al.*, 1998) and cycle polyamides (Hermann *et al.*, 1999a). Furthermore, conjugation of two hairpins together, similar to nature's strategy of threading several DNA binding domains together in a single polypeptide chain, is valid strategy to increase the binding site size (Hermann *et al.*, 1999b). Thus, continued evolution of polyamide design has resulted in a variety of topological arrangements that can target predictable sequences with affinities and specificities that rival those of DNA binding proteins.

Until recently, the specificity of Im-Py polyamides was limited to discriminating G•C, C•G and A•T/T•A. To break this partial degeneracy, a new residue, 3-hydroxypyrrole, was designed and found to preferentially bind T•A over A•T, G•C and C•G base pairs as an Hp/Py pairing (White *et al.*, 1998; 1999). This completed polyamide recognition of all four bases pairs; an Im/Py targets G•C whereas Py/Im prefers C•G (White *et al.*, 1997), and Py/Py pairs are partially degenerate for A•T and T•A base pairs (White *et al.*, 1996). An Hp/Py specifies T•A and Py/Hp binds opposite A•T (Figure 1.10) (White *et al.*, 1998; 1999). Nevertheless, in light of the functional similarity of the four base pairs in the minor groove, the structural basis of the pairing rules was unknown. Since Im-Py polyamides are cell permeable and capable of

inhibiting transcription of specific genes in living cells (Gottesfeld *et al.*, 1997; Dickinson *et al.*, 1998), it was important to provide a structural model for DNA binding that could be used as a reference for thermodynamic studies.

Figure 1.9. Schematic representation of representative polyamide-DNA motifs. The equilibrium association constant (M^{-1}) is indicated below the compound.

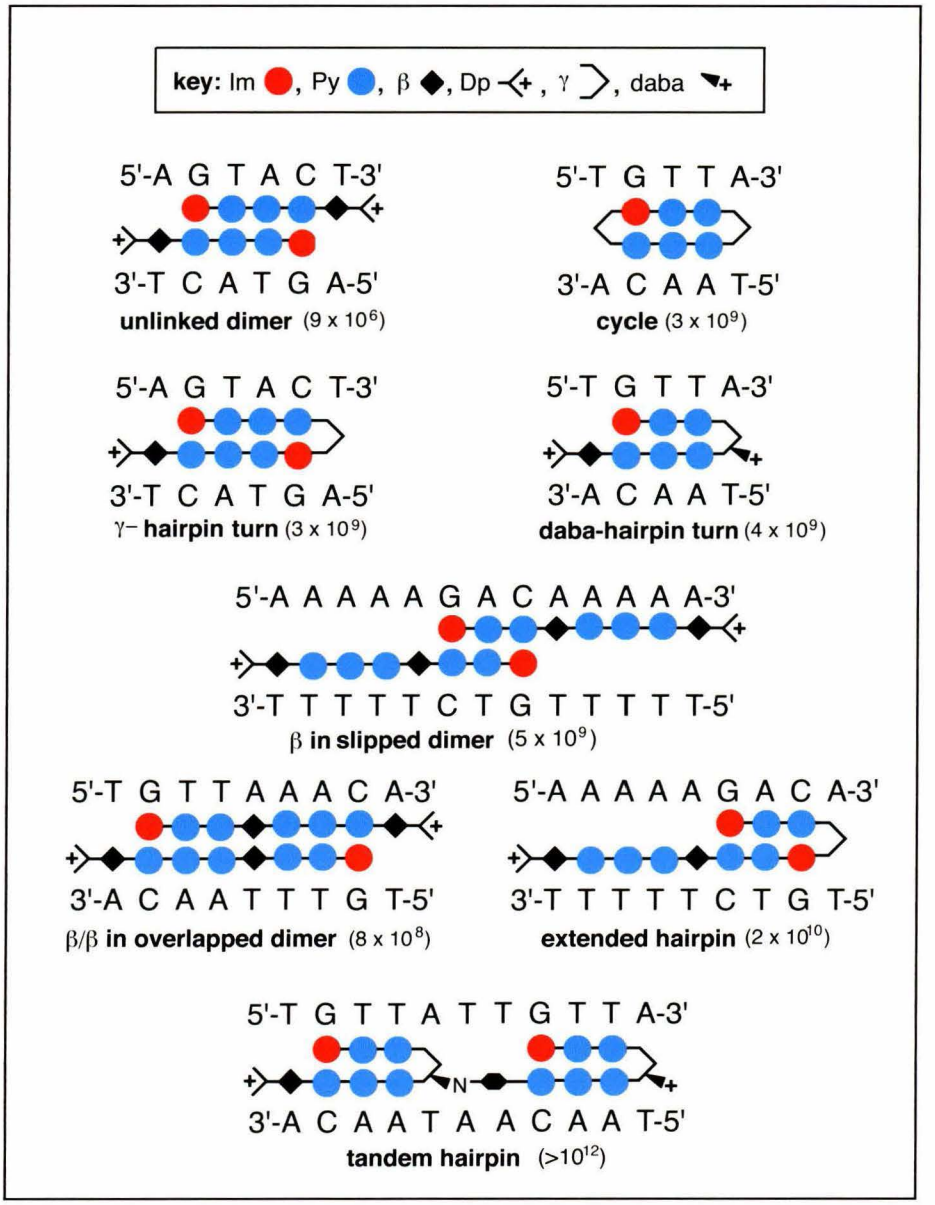
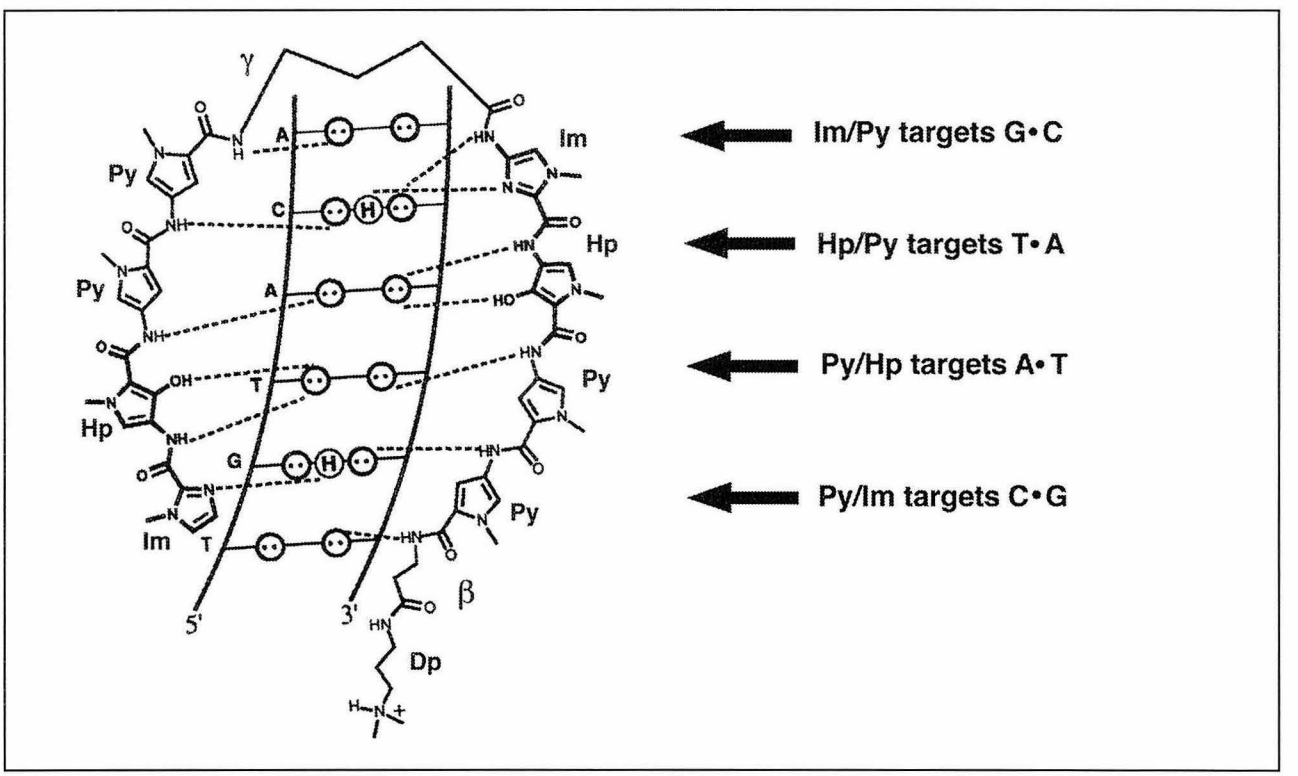


Figure 1.10. Pairing rules for minor groove recognition.



1.4 Intercalation via the Major Groove: Metallointercalators

Unlike most small DNA-binding molecules, octahedral complexes of rhodium(III) containing the phi ligand (where phi is 9,10-phenanthrenequinone diimine) intercalate B-form DNA while orienting additional functional groups in the major groove, as demonstrated by NMR (David, S. S. & Barton, 1993; Collins *et al.*, 1994; Shields & Barton, 1995b; Hudson *et al.*, 1995; 1998) and chemical studies (Sitlani *et al.*, 1992; Krotz *et al.,* 1993). Despite the inherent nonspecificity of interactions between the intercalating phi and the aromatic surfaces of the bases, site-selectivity can be achieved by modification of the ancillary ligands to recognize the diverse functional groups in the major groove. In general, classes with large aromatic substitutions recognize the overall shape of the binding site, whereas those with aliphatic axial amines also interact with the detailed groups displayed on the edges of the bases (Johann & Barton, 1996).

Most phi complexes of rhodium(III) are photoreactive, with high energy irradiation (313 nm) resulting in direct DNA strand scission, whereas lower energy wavelengths (365 nm) promote long range photooxidation (Holmlin *et al.*, 1997). Cleavage of the bound site provides a convenient assay for specificity and affinity of rhodium intercalators. Furthermore, long range electron transfer is a sensitive function of the stacking between the intervening base pairs (Hall & Barton, 1997; Kelley & Barton, 1999). These qualities of photoreactivity and site-selectivity have allowed metal complexes to be employed in a variety of ways, including as probes of nucleic acid tertiary structure (Chow & Barton, 1992; Terbrueggen & Barton., 1995; Campisi *et al.*, 1994), for mismatch detection (Jackson *et al.*, 1999), to repair DNA damage (Dandliker *et al.*, 1997) and for the study of protein-nucleic acid interactions (Odom *et al.*, 1999; Rajski *et al.*, 1999).

On the most basic level, the handedness of the octahedral coordination geometry affects the affinity of the $Rh[phi]^{+3}$ complexes for nucleic acids, due to steric clashes with

the walls of the grooves. For most compounds, including $[Rh(phen)_2phi]^{+3}$ and $[Rh(DPB)_2phi]^{+3}$ (phen, 9,10-phenanthroline; DPB, 4,4'-diphenyl-2,2'-bipyridyl) shown in Figure 1.12 (Pyle *et al.*, 1990; Sitlani & Barton, 1994), the right-handed isomer (Δ) binds with higher affinity for the right-handed DNA double helix than does the left-handed configuration (Λ). Due to the malleability of DNA, there are exceptions; for example the compound Λ-$[Rh(MGP)_2phi]^{+3}$ (MGP, 4-guanidylmethyl-1,10-phenanthroline) significantly distorts the DNA structure, allowing it to specifically target a six-base pair sequence with an affinity of $\sim 10^8$ M^{-1} (Terbrueggen & Barton, 1995; Franklin & Barton, 1998; Terbrueggen *et al.*, 1998).

The complementary fit of the Δ-enantiomer of $[Rh(phen)_2phi]^{+3}$ in the major groove of B-form DNA allows its use to investigate overall nucleic acid conformation (Chow & Barton, 1992). In addition to this straightforward global discrimination, $[Rh(phen)_2phi]^{+3}$ can detect finer variations of DNA structure, preferentially binding highly propellar twisted sites with an open major groove that minimize the steric clash between the ancillary ligands and the DNA (Pyle *et al.*, 1990; Campisi *et al.*, 1994). Striking home the potential specificity of shape recognition, a second compound with no hydrogen bond donors or acceptors, $[Rh(DPB)_2phi]^{+3}$, binds with affinity and selectivity comparable to DNA binding proteins (Sitlani *et al.*, 1993).

Direct hydrogen bonds and van der Waals interaction via the ancillary ligands in the major groove can also be used for DNA recognition by rhodium intercalators. The simple complex $[Rh(NH_3)_2phi]^{+3}$ intercalates 5'-GC-3' steps, with the ammonia groups positioned to form hydrogen bonds with the guanine-O6 (Krotz *et al.*, 1993a; Collins *et al.*, 1994). Complexes containing aliphatic axial amines, such as Δ-$[Rh(en)_2phi]^{+3}$ (en, ethylene diamine) or Δ-$[Rh([12]aneN_4)phi]^{+3}$ (Figure 1.11), display a similar capability to target 5'-GC-3' sites (Krotz *et al.*, 1993a; Shields & Barton, 1995). By comparison, Δ-$[Rh([12]aneS_4)phi]^{+3}$ lacks hydrogen bond donors and correlates with GC-specificity.

In addition to hydrogen bonding interactions, it was observed that Λ-$[Rh(en)_2phi]^{+3}$ binds 5'-TX-3' steps as well as 5'-GC-3' via hydrophobic contact between the methylene carbons of the phi complex and thymine-methyl group. The second generation compound Δ-α-$[Rh(Me_2trien)phi]^{+3}$ (Me_2trien, (2*R*, 9*R*)-diamino-4,7-diazadecane) (Krotz *et al.*, 1993b) incorporates both of these established functionalities by addition of methyl groups to the ethylene diamine skeleton (Figure 1.12). The intent was that the methyl groups would interact with the thymines, while the axial amines would target guanines, resulting in specificity for the major groove of 5'-TGCA-3' sequences. Following synthesis and painstaking separation of 12 possible isomers (Krotz & Barton, 1994), photocleavage data and H^1-NMR analysis indicated that this design was a success (Krotz *et al.*, 1993b; Hudson *et al.*, 1995; 1998).

Rhodium intercalators are a class of compounds with intriguing photoreactive capabilities. Ironically, these major groove ligands lack the predictive DNA recognition of polyamides in the minor groove. Nevertheless, the octahedral coordination geometry presents a scaffold of known orientation that can be substituted with synthetic groups for sequence specific recognition by shape selection or direct readout. Adding to the thorough global structural characterizations $Rh[phi]^{+3}$ complexes by H^1-NMR, an atomic resolution view of an intercalated DNA helix would identify details of interactions that facilitate DNA photooxidation and site-selectivity.

Figure 1.11. Chemical structures of selected phi complexes of rhodium(III).

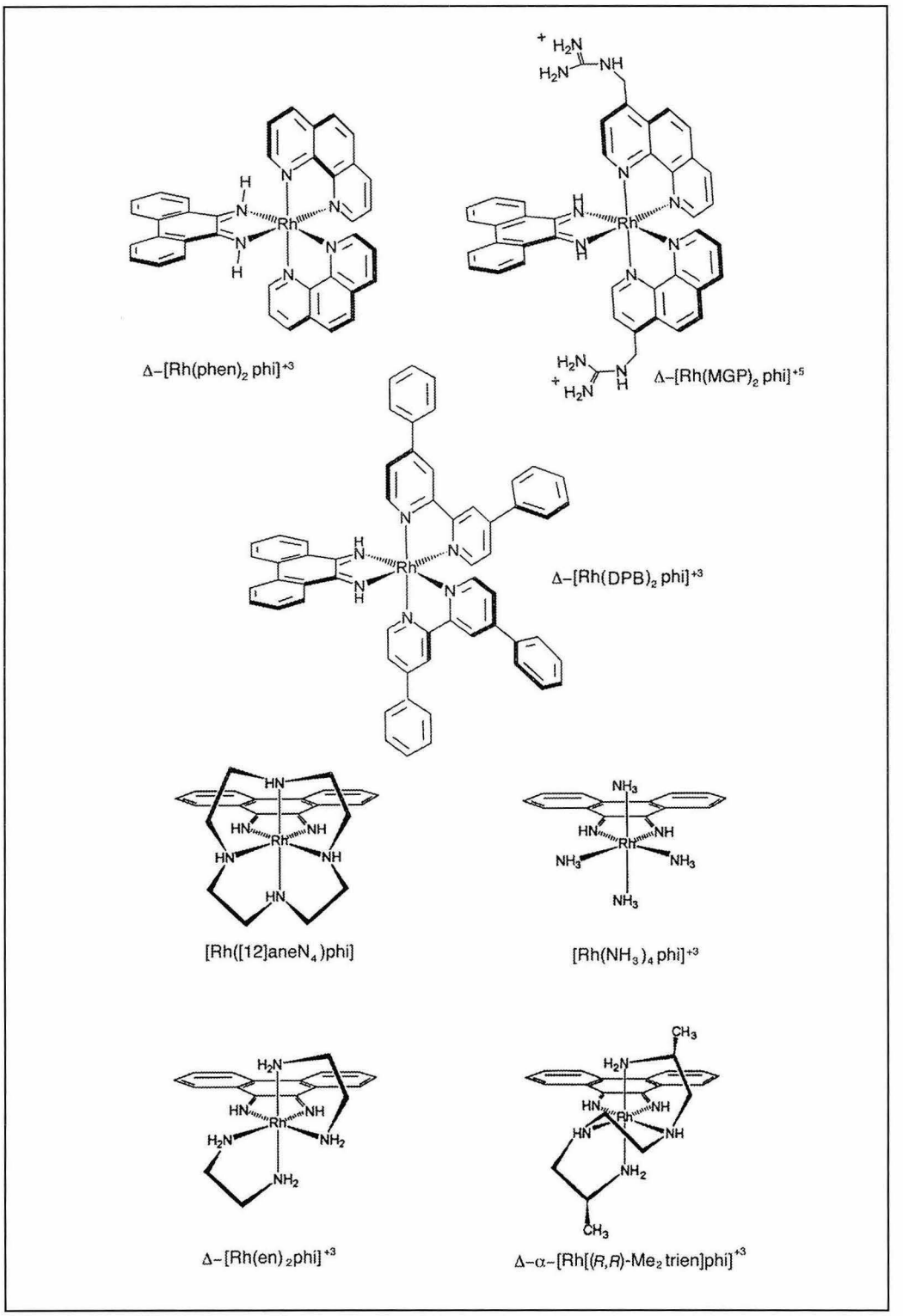
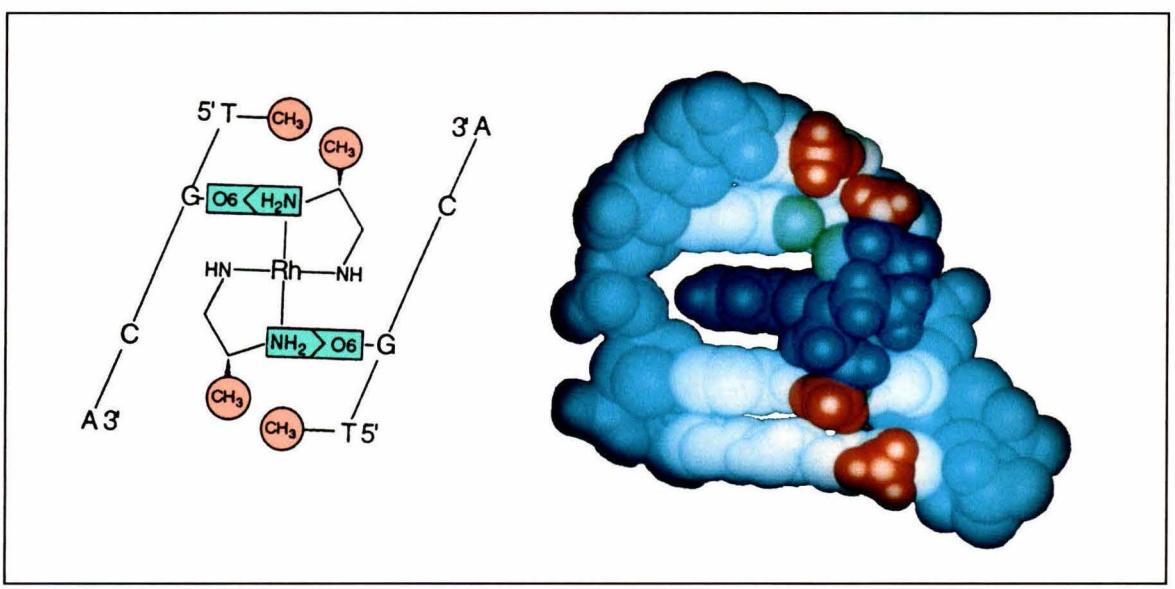


Figure 1.12. Design of a site-selective rhodium intercalator Δ-α-[Rh(Me_2trien)phi]$^{+3}$. A schematic DNA binding model of the interactions of the 5'-TGCA-3' site with the ancillary ligands is shown to the left. To the right, a CPK representation of a molecular model of predicted interactions generated using the program *Insight II* (BIOSYM / Molecular Simulations). Sequence-specific hydrophic interactions are colored orange, hydrogen bonds are shown in green. Adapted from Krotz *et al.*, 1993b.



1.5 References

Altona, C. & Sundaralingam, M. (1972). Conformational analysis of the sugar ring in nucleosides and nucleotides. A new description using the concept of pseudorotation. *J. Amer. Chem. Soc.* **94**, 8205-8212.

Baird, E. E. & Dervan, P. B. (1996). Solid phase synthesis of polyamides containing imidazole and pyrrole amino acids. *J. Am. Chem. Soc.* **118**, 6141-6146.

Berman, H. M. & Young, P. R. (1981). The interaction of intercalating drugs with nucleic acids. **10**, 87-114.

Burley, S. K. (1996). The TATA box binding protein. *Curr. Opin. Struct. Biol.* **6**, 69-75.

Campisi, D., Morii, T. & Barton, J. K. (1994). Correlations of crystal structures of DNA oligonucleotides with enantioselective recognition by $[Rh(phen)_2phi]^{+3}$ - probes of DNA propeller twisting in solution. *Biochem.* **33**, 4130-4139.

Cho, J., Parks, M. E. & Dervan, P. B. (1995).Cyclic polyamides for recognition in the minor groove of DNA. *Proc. Natl. Acad. Sci.* **92**, 10389-10392.

Choo, Y. & Klug, A. (1997). Physical basis of a protein-DNA recognition code. *Curr. Opin. Struct. Biol.* **7**, 117-125.

Chow, C. S. & Barton, J. K. (1992). Transition metal complexes as probes of nucleic acids. *Method. Enzymol.* **212**, 219-242.

Collins, J. G., Shields, T. P. & Barton, J. K. (1994). H^1-NMR of $[Rh(NH_3)_4phi]^{+3}$ bound to $d(TGGCA)_2$; classical intercalation by a nonclassical octahedral metallointercalator. *J. Am. Chem. Soc.* **116**, 9840-9846.

Dandliker, P. J., Holmlin, R. E. & Barton, J. K. (1997). Oxidative thymine dimer repair in the DNA helix. *Science.* **275**, 1465-1468.

David, S. S. & Barton, J. K. (1993). NMR evidence for specific intercalation of Δ-$[Rh(phen)_2phi]^{+3}$ in $[d(GTCGAC)_2$. *J. Am. Chem. Soc.* **115**, 2984-2985.

Dickinson, L. A., Gulizia, R. J., Trauger, J. W., Baird, E. E., Mosier, D. E., Gottesfeld, J. M. & Dervan, P. B. (1998). Inhibition of RNA polymerase II transcription in human cells by synthetic DNA-binding ligands. *Proc. Natl. Acad. Sci. USA.* **95**, 12890-12895.

Elrod-Erikson, M., Benson, T.E. & Pabo, C.O. (1998). High-resolution structures of variant Zif268-DNA complexes: implications for understanding zinc finger-DNA recognition. *Structure* **6,** 451-464.

Franklin, S. J. & Barton, J. K. (1998). Differential DNA recognition by enantiomers of $Rh(MGP)_2phi$: A combination of shape selection and direct readout. *Biochem.* **37**, 16093-16105.

Geierstanger, B. H., Mrksich, M., Dervan, P. B. & Wemmer, D. E. (1994). Design of a G•C specific DNA minor groove binding peptide. *Science.* **266**, 646-650.

Geierstanger, B. H. & Wemmer, D. E. (1995). Complexes of the minor groove of DNA. *Annu. Rev. Biophys. Biomol. Struct.* **24**, 463-493.

Gottesfeld, J. M., Neely, L., Trauger, J. W., Baird, E. E. & Dervan, P. B. (1997). Regulation of gene expression by small molecules. *Nature*. **387**, 202-205.

Grzeskowiak, K. (1996). Sequence dependent structural variation in B-DNA. *Chem. Biol.* **3**, 785-790.

Hall, D. B. & Barton, J. K. (1997). Sensitivity of DNA-mediated electron transfer to the intervening π-stack: A probe for the integrity of the DNA base stack. *J. Am. Chem. Soc.* **119**, 5045-5046.

Hartmann, B. & Lavery, R. (1996). DNA structural forms. *Quart. Rev. Biophys.* **29**, 309-368.

Hermann, D. M., Baird, E. E. & Dervan, P. B. (1998). Stereochemical control of the DNA binding affinity, sequence specificity and orientation preference of chiral hairpin polyamides in the minor groove. *J. Am. Chem. Soc.* **120**, 1382-1391.

Hermann, D. M., Turner, J. M., Baird, E. E. & Dervan, P. B. (1999a) Cycle polyamide motif for recognition of the minor groove of DNA. *J. Am. Chem. Soc.* **121**, 1121-1129.

Hermann, D. M., Baird, E. E. & Dervan, P. B. (1999b) Tandem hairpin motif for recognition in the minor groove of DNA by pyrrole-imidazole polyamides. *Chem. Eur. J.* **5**, 975-983.

Holmlin, R. E., Dandliker, P. J. & Barton, J. K. (1997). Charge transfer through the DNA base stack. *Angew. Chem. Int. Ed. Engl.* **36**, 2714-2730.

Hudson, B. P., Dupureur, C. M. & Barton, J. K. (1995). H^1-NMR structural evidence for the sequence specific design of an intercalator Δ-α-[Rh(Me_2trien)phi]$^{+3}$ bound to d(GAGTGCACTG)$_2$. *J. Am. Chem. Soc.* **117**, 9379-9380.

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

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

Johann, T. W. & Barton, J. K. (1996). Recognition of DNA by octahedral coordination complexes. *Phil. Trans. R. Soc. Lond. A*. **354**, 299-324.

Kelley, S. O. & Barton, J. K. (1999). Electron transfer between bases in double helical DNA. *Science*. **283**, 375-381.

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

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

Krotz, A. H. & Barton, J. K. (1994). Rational construction of metallointercalators: Syntheses, structures and characterizations of isomers of ((2*S*, 9*S*)-diamino-4,7-diazadecane)(9,10-phenanthrene diimine-rhodium(III).

Lavery, R. & Sklenar, H. (1988). The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biolmolec. Struct. Dynam.* **6**, 63-91.

Lin, C. H. & Patel, D. J. (1998). Encapsulating an amino acid in a DNA fold. *Nat. Struct. Biol.* **3**, 1046-1050.

Matray, T. J. & Kool, E. T. (1999). A specific partner for abasic damage in DNA. *Nature.* **399**, 704-708.

Mrksich, M., Wade, W. S., Dwyer, T. J., Geierstanger, B. H., Wemmer, D. E. & Dervan, P. B. (1992). Antiparallel side-by-side dimeric motif for sequence-specific recognition in the minor groove of DNA by the designed peptide 1-methylimidazole-2-carboxamide netropsin. *Proc. Natl. Acad. Sci. USA*. **89**, 7586-7590.

Mrksich, M. & Dervan, P. B. (1993). Antiparallel side-by-side heterodimer for sequence-specific recognition in the minor groove of DNA by a distamycin 1-methylimidazole-2-carboxamide pair. *J. Am. Chem. Soc*. **115**, 2572-2576.

Mrksich, M., Parks, M. E. & Dervan, P. B. (1994). Hairpin peptide motif. A new class of oligopeptides for sequence-specific recognition in the minor groove of double helical DNA. *J. Am. Chem. Soc*. **116**, 7983-7988.

Mrksich, M. & Dervan, P. B. (1995). Recognition in the minor groove of DNA at 5'-(A,T)GCGC(A,T)-3' by a four ring tripeptide dimer. Reversal of the specificity of natural product distamycin. *J. Am. Chem. Soc*. **117**, 3325-3332.

Nelson, H. C. M. (1995). Structure and function of DNA binding proteins. *Curr. Opin. Gen. Dev.* **5**, 180-189.

Odom, D. T., Parker C. S. & Barton J. K. (1999). Site-specific inhibition of transcription factor binding to DNA by a metallointercalator. *Biochem.* **38**, 5155-5163.

Ohndorf, U-.M., Rould, M. A., He, Q., Pabo, C. O. & Lippard, S. J. (1999). Basis for recognition of cisplatin-modified DNA by high-mobility-group proteins. *Nature.* **399**, 708-712.

Parks, M. E., Baird, E. E. & Dervan, P. B. (1996). Optimization of the hairpin polyamide design for recognition of the minor groove of DNA. *J. Am. Chem. Soc.* **118**, 6147-6152.

Pelton, J. G. & Wemmer, D. E. (1989). Structural characterization of a 2:1 distamycin A-d(CGCAAATTGGC) complex by two-dimensional NMR. *Proc. Natl. Acad. Sci. USA.* **86**, 5723-5727.

Pyle, A. M., Morii, T. & Barton, J .K. (1990). Probing microstructures in double helical DNA with chiral metal complexes: Recognition of changes in base-pair propellar twisting in solution. *J. Am. Chem. Soc.* **112**, 9432-9434.

Rajski S. R., Kumar S., Roberts R. J. & Barton J. K. (1999). Protein-modulated DNA electron transfer. *J. Am. Chem. Soc.* **121**, 5615-5616.

Rhodes, D., Schwabe, J. W. R., Chapman, L. & Fairall, L. (1996). Towards an understanding of protein-DNA recognition. *Phil. Trans. R. Soc. Lond. B.* **351**, 501-509.

Rich, A. (1993). DNA comes in many forms. *Gene.* **135**, 99-109.

Saenger, W. (1984). *Principles of Nucleic Acid Structure.* (Springer-Verlag, New York, NY).

Sastry, M. & Patel, D. J. (1993). Solution structure of the mithramycin-dimer DNA complex. *Biochemistry.* **32**, 6588-6604.

Schwartz T., Rould M. A., Lowenhaupt K., Herbert A. & Rich A. (1999). Crystal structure of the Z alpha domain of the human editing enzyme ADAR1 bound to left-handed Z-DNA. *Science.* **284**, 1841-1845.

Searle, M. S., Bicknell, W., Wakelin, L. P. & Denny, W. A. (1991). Anthracycline antibiotic arugomycin binds in both grooves of the DNA helix simultaneously: and NMR and molecular modeling study. *Nucleic Acids Res.* **19**, 2897-2906.

Seeman, N. C., Rosenburg, J. M. & Rich, A. (1976). Sequence specific recognition of double helical nucleic acids by proteins. *Proc. Natl. Acad. Sci. USA.* **73**, 804-808.

Shields, T. P. & Barton, J. K. (1995a). Sequence selective DNA recognition and photocleavage: A comparison of enantiomers of $[Rh(en)_2phi]^{+3}$. *Biochemistry.* **34**, 15037-15048.

Shields, T. P. & Barton, J. K. (1995b). Structural examination of enantioselective intercalation; H^1-NMR of $[Rh(en)_2phi]^{+3}$ isomers bound to $d(GTGCAC)_2$. *Biochemistry.* **34**, 15049-15056.

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

Sitlani, A., Dupureur, C. M. & Barton, J. K. (1993). Enantiospecific palindromic recognition of 5'-d(CTCTAGAG)-3' by a novel rhodium intercalator: Analogies to a DNA binding protein. *J. Am. Chem. Soc.*, **115**, 12589-12590.

Sitlani, A., Long, E. C., Pyle, A. M. & Barton, J. K. (1994). Sequence specific recognition of DNA by phenanthrenequinone diimine complexes of rhodium(III); importance of steric and van der Waals interactions. *Biochemistry.* **33**, 12100-12108.

Steitz, T. A. (1990). Structural studies of protein-nucleic acid interaction: Sources of sequence-specific binding. *Quart. Rev. Biophys.* **23**, 205-280.

Swalley, S. E., Baird, E. E. & Dervan, P. B. (1999). Effects of γ-turn and β-tail amino acids on sequence specific recognition of DNA by hairpin polyamides. *J. Am. Chem. Soc.* **119**, 6953-6961.

Swaminathan, K., Flynn, P., Reece, R. J. & Marmorstein, R. (1997). Crystal structure of a PUT3-DNA complex reveals a novel mechanism for DNA recognition by a protein containing a Zn_2Cys_6 binuclear cluster. *Nat. Struct. Biol.* **4**, 751-759.

Takusegawa, F., Dabrow, M., Neidle, S. & Bermann, H. M. (1982) The structure of a pseudo-intercalated complex between actinomycin and the DNA binding sequence d(GpC). *Nature*. **296**, 466-469.

Terbrueggen, R. H. & Barton, J. K. (1995). Sequence specific DNA binding by a rhodium complex; recognition based on sequence specific twistability. *Biochem.* **34**, 8227-8234.

Terbrueggen, R. H., Johann, T. W. & Barton, J. K. (1998). Functionalized rhodium intercalators for DNA Recognition. *Inorg. Chem.* **37**, 6874-6883.

Trauger, J. W., Baird, E. E. & Dervan, P. B. (1996). Extension of sequence-specific recognition in the minor groove of DNA by pyrrole-imidazole polyamides to 9-13 base pairs. *J. Am. Chem. Soc.* **118**, 6160-6166.

Turner, J. M., Swalley, S. E., Baird, E. E. & Dervan, P. B. (1998). Aliphatic/aromatic amino acid pairings for polyamide recognition in the minor groove of DNA. *J. Am. Chem. Soc.* **120**, 6219-6226.

Wade, W. S., Mrksich, M. & Dervan, P. B. (1992). Design of peptides that bind in the minor groove of DNA at 5'-$(A,T)_5$G(A,T)C(A,T)-3' sequences by a dimeric side-by-side motif. *J. Am. Chem. Soc.* **114**, 8783-8794.

White, S., Baird, E. E. & Dervan, P. B. (1996). Effects of the A•T/T•A degeneracy of pyrrole-imidazole polyamide recognition in the minor groove of DNA. *Biochemistry*. **35**, 12532-12537.

White, S., Baird, E. E. & Dervan, P. B. (1997). On the pairing rules for recognition in the minor groove of DNA by pyrrole-imidazole polyamides. *Chem. Biol.* **4**, 569-577.

White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E. & Dervan, P. B. (1998). Recognition of the four Watson-Crick base pairs in the DNA minor groove by synthetic ligands. *Nature.* **391**, 468-471.

White, S., Turner, J. M., Szewczyk, J. W., Baird, E. E. & Dervan, P. B. (1999). Affinity and specificity of multiple hydroxypyrrole/pyrrole ring pairings for coded recognition of DNA. *J. Am. Chem. Soc.* **121**, 260-261.

Zimmer, C. & Wahnert, U. (1986). Nonintercalating DNA-binding ligands: specificity of the interaction and their use as tools in biophysical, biochemical and biological investigations of the genetic material. *Prog. Biophys. Molec. Biol.* **47**, 31-112.

Chapter 2

Structural Basis for G•C Recognition in the DNA Minor Groove

Abstract: Small molecules which target specific DNA sequences offer a potentially general approach for the regulation of gene expression. Pyrrole-imidazole polyamides represent the only class of synthetic small molecules that can bind predetermined DNA sequences with affinities and specificities comparable to DNA binding proteins. Antiparallel side-by-side pairing of two aromatic amino acids, imidazole (Im) and pyrrole (Py), distinguish G•C from C•G, and both from A•T/T•A base pairs. A high resolution x-ray crystal structure of a four-ring pyrrole-imidazole polyamide specifically bound as a dimer to a six-base pair predetermined DNA site reveals a structural framework of hydrogen bonds and interactions with the walls of the minor groove that underlies the pairing rules for DNA recognition.

Publication: Kielkopf, Baird, Dervan & Rees (1998). *Nature Struct. Biol.* **5**, 104-109.

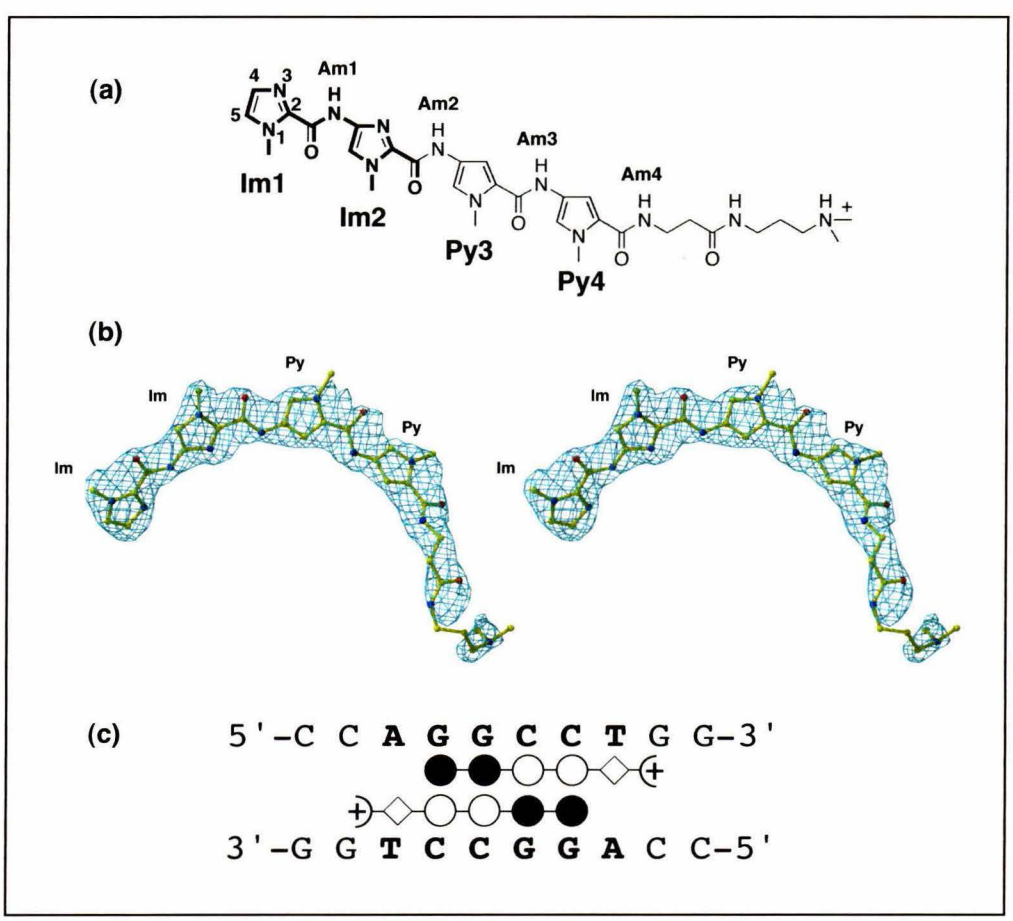
2.1 Introduction

Pairing rules have been developed to rationally determine the sequence-specificity of minor-groove-binding polyamides containing *N*-methylpyrrole (Py) and *N*-methylimidazole (Im) amino acids (Trauger *et al.*, 1996; White *et al.*, 1997a). An Im/Py pair distinguishes G•C from C•G and both of these from A•T/T•A base pairs (Wade *et al.*, 1992; Mrksich *et al.*, 1992; Trauger *et al.*, 1996; White *et al.*, 1997). A Py/Py pair specifies A,T from G,C but does not distinguish A•T from T•A (Pelton & Wemmer, 1989; White *et al.*, 1996). The antiparallel dimer in the DNA minor groove formed by a polyamide containing both imidazole and pyrrole was confirmed by NMR structure studies (Mrksich *et al.*, 1992; Geierstanger *et al.*, 1994). Subsequent high resolution x-ray studies of distamycin, $(PyPyPy)_2$• DNA, have provided the structure of the bound Py/Py pair (Chen *et al.*, 1994; 1997). The Im/Py and Py/Im pairings, which allowed the design of small molecules which recognize any predetermined DNA sequence at subnanomolar concentration, have not yet been characterized by use of x-ray crystallographic methods (Trauger *et al.*, 1996; Swalley *et al.*, 1997a; b). Because these ligands have been shown to be cell permeable and to inhibit the transcription of specific genes in cell culture (Gottesfeld *et al.*, 1997), it is important to understand the molecular basis for this non-biological recognition.

We report here the atomic resolution structure of a polyamide of sequence composition ImImPyPy bound to an oligonucleotide duplex containing its cognate 5'-GGCCW-3' site (where W represents either A or T) as predicted by the pairing rules. In addition to providing a molecular basis for predictable read-out of the DNA minor groove, the structure addresses key issues for the design of enhanced second generation polyamides: the register of the polyamide dimer with respect to the DNA double helix critical for expanding binding site size (Kelley *et al.*, 1996), the position of the paired polyamide rings and the base pairs for the design of novel pairings to break the A•T/T•A

degeneracy (White *et al.*, 1996), and the determinants of high affinity binding in order to increase affinity for sequences which have been difficult to target (Swalley *et al.*, 1997b).

Figure 2.1. (a) Diagram of polyamide ImImPyPy-β-Dp (Im = *N*-methylimidazole carboxamide, Py = *N*-methylpyrrole carboxamide, β = β-alanine, Dp = dimethylaminopropylamide, Am = amide). Rings are numbered starting from the N-terminal imidazole (Im1, Am1) to the C-terminal pyrrole (Py4, Am4). **(b)** Omit $|F_o|$ - $|F_c|$ electron density map for one of the ImImPyPy-β-Dp molecules, shown with 2σ contour. **(c)** Schematic DNA-binding model, with the sequence of the co-crystallized oligonucleotide. Filled circles represent imidazoles; unfilled, pyrroles; and diamonds β-alanine.



2.2 Overall Structure of the Complex

The pyrrole-imidazole polyamide ImImPyPy-β-Dp was co-crystallized with the oligonucleotide 5'-CC**AGGCCT**GG-3' (Figure 2.1). The polyamide selectively binds the six base pair target sequence, 5'-WGGCCW-3', with an equilibrium dissociation constant, K_d = 42 nM (Swalley *et al.*, 1997b). The structure was solved by MIR with multiple-crystal form and noncrystallographic symmetry-averaging to improve the initial phase information (Methods). One of the brominated derivatives (BrCyt6) diffracts to 2.0 Å resolution with synchrotron radiation, and was used for the refinement of the structure. Two independent polyamide molecules and one duplex oligonucleotide are present in the asymmetric unit of the crystal, although the conformations of the polyamides are very similar (0.5 Å r.m.s. difference). Unlike the crystal structure of the distamycin dimer (Chen *et al.*, 1994; 1997), the crystal contacts are not mediated by the polyamide molecules.

Consistent with structural studies of other pyrrole-imidazole polyamides, the polyamide ImImPyPy-β-Dp is bound as an antiparallel head-to-tail dimer (Figure 2.2), burying 39% of the total available surface area deep in the minor groove of the B-form oligonucleotide (Figure 2.3a). The N- to C-terminal orientation of each polyamide recognizes the adjacent 5' to 3' nucleic acid strand, a distinct orientation preference that is crucial for a given pyrrole-imidazole polyamide to predictably target a single site (White *et al.*, 1997b). Minor groove recognition is accomplished by two aromatic residues, each from independent molecules in the dimer, that complement each target base pair of the duplex DNA. The four Im/Py, Im/Py, Py/Im, and Py/Im pairs center over and read the core of the oligonucleotide, G•C, G•C, C•G, C•G, with a series of hydrogen bonds and van der Waals interactions. Flanking the central GGCC sequence, the smooth minor grooves of the A•T/T•A base pairs accommodate the hydrophobic β-alanines. The ImImPyPy-β-Dp molecules are fully overlapped: the last residue of one polyamide is

paired with the first residue of the other in the plane of the bases. The binding mode with the C-terminal pyrroles unpaired, referred to as slipped, is prevented because the resulting Im/Im pair is unfavorable for DNA binding (White *et al.*, 1997a). The fully-overlapped conformation of the designed polyamide contrasts with the staggered register of distamycin (Chen *et al.*, 1994; 1997) and ImIm homodimers (Kopka *et al.*, 1997) observed in crystallographic structures, although it is consistent with the NMR characterization of the designed polyamide ImPyImPy-β-Dp (Geierstanger *et al.*, 1994). In practice, slipped binding modes are eliminated by linking antiparallel molecules with a γ-aminobutyric acid hairpin turn (Trauger *et al.*, 1996; Swalley *et al.*, 1997b).

The rings of the polyamide dimer are approximately adjacent in the planes of the base pairs. However, when viewed perpendicular to the long axis of the polyamide, the residues are offset such that the aromatic rings of one molecule stack over the carboxamides of the other. Because the molecules track the minor groove, the normals of the imidazoles and pyrrole planes are inclined 50° with respect to the base pairs (Methods). By simple geometric constraints, the centers of the imidazoles or pyrroles of one molecule are expected to be closest to the bond between the carbonyl carbon and the aromatic ring of the other. Within error, this prediction is observed, although the carbonyl carbons are actually closer than the bonds to the centers of aromatic rings (Figure 2.3b). In the NMR (Pelton & Wemmer, 1989) and crystallographic (Chen *et al.* 1994; 1997) models of distamycin dimers, the aromatic rings of one molecule also stack over the carbonyl carbons of the other. Despite differences in global slipped versus maximum overlap, the C-terminal three residues of the ImImPyPy-β-Dp dimer superimpose with 0.3 Å difference on those of the distamycin dimer. The stacking interaction between pyrrole-imidazole polyamides is a consequence of, and may contribute to, side-by-side pairing and cooperative binding to B-form DNA.

Figure 2.2. Stereo diagram of the (ImImPyPy-β-Dp)$_2$ / DNA complex, with the polyamides in yellow, bases pairs in cyan, and phosphate backbone schematized as red ribbons.

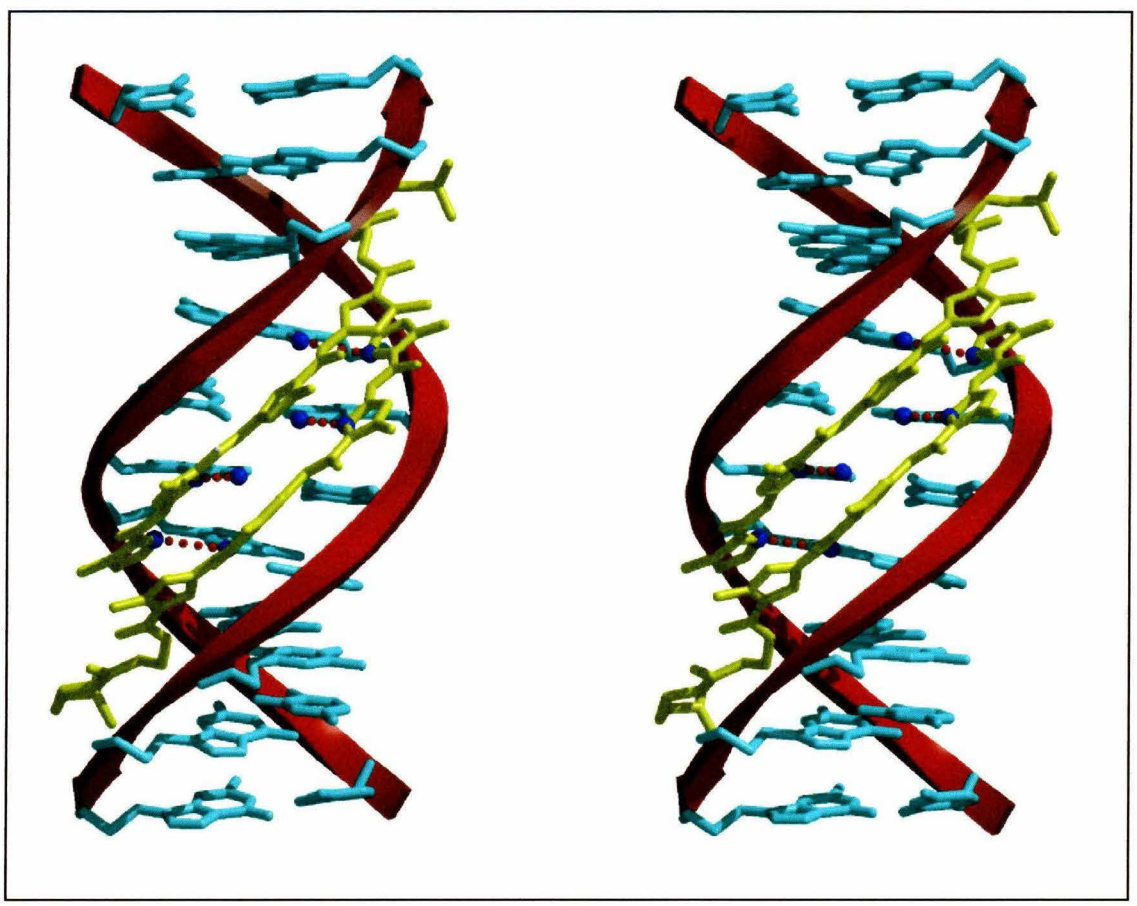
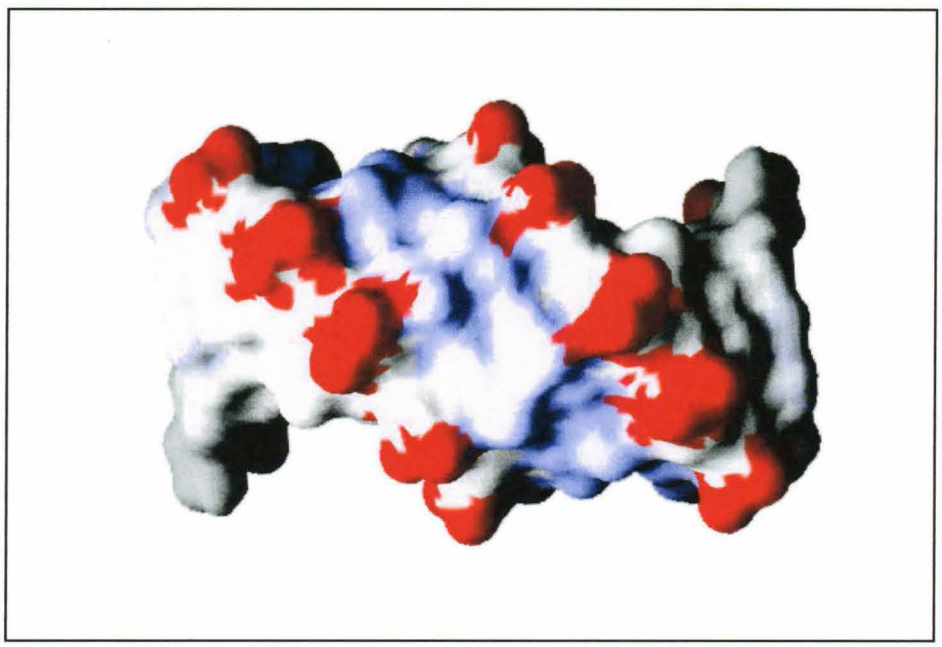
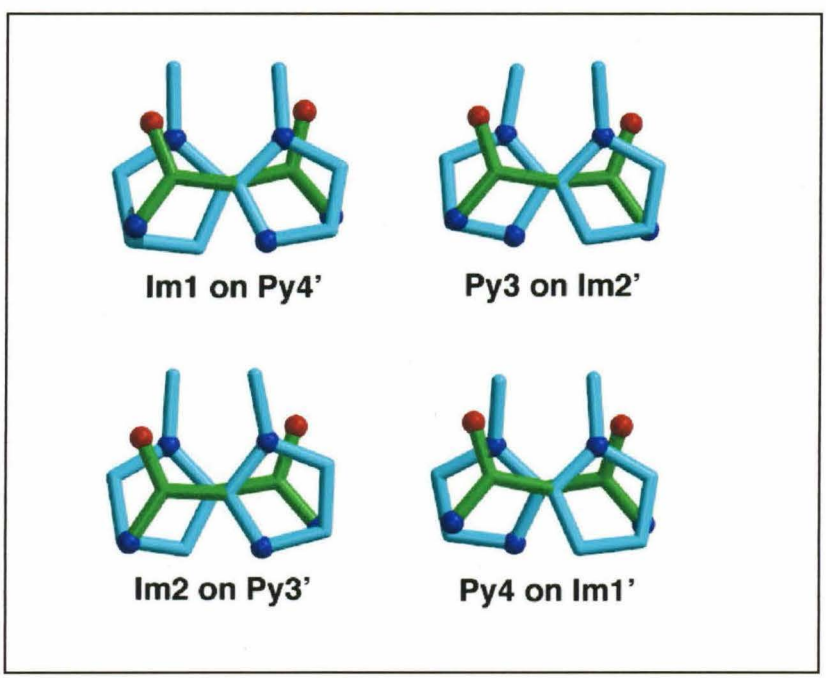


Figure 2.3. **(a)** Grasp surface showing that the polyamide dimer completely fills the minor groove. Negative, red; positive, blue.



(b) The four independent amide-aromatic stackings. The aromatic rings center approximately over the carbonyl carbon.



2.3 Hydrogen-bond Interactions

The imidazoles and the amide groups of the polyamides form a series of hydrogen bonds with the edges of the bases on the adjacent DNA strand (Figure 2.4a). Contributing to the affinity, rather than specificity, of the polyamide for DNA, the amides linking the aromatic rings contribute hydrogen bonds to the purine N3 or pyrimidine O2. As opposed to the three-center hydrogen bonds donated by the amides of the netropsin monomers to successive A•T pairs of the binding site (Kopka *et al.*, 1985), each amide is within hydrogen bond distance of a single base (3.0 Å average, Table 2.1), and beyond contact with the preceding or following N3 or O2 (5.1 Å average). Although the β-alanines of one molecule in the dimer and the dimethylaminopropylamide tails are not well-ordered, the β-alanine of the other polyamide molecule crosses over to contribute a hydrogen bond from the C-terminal amide to the adenine N3 of the opposite DNA strand.

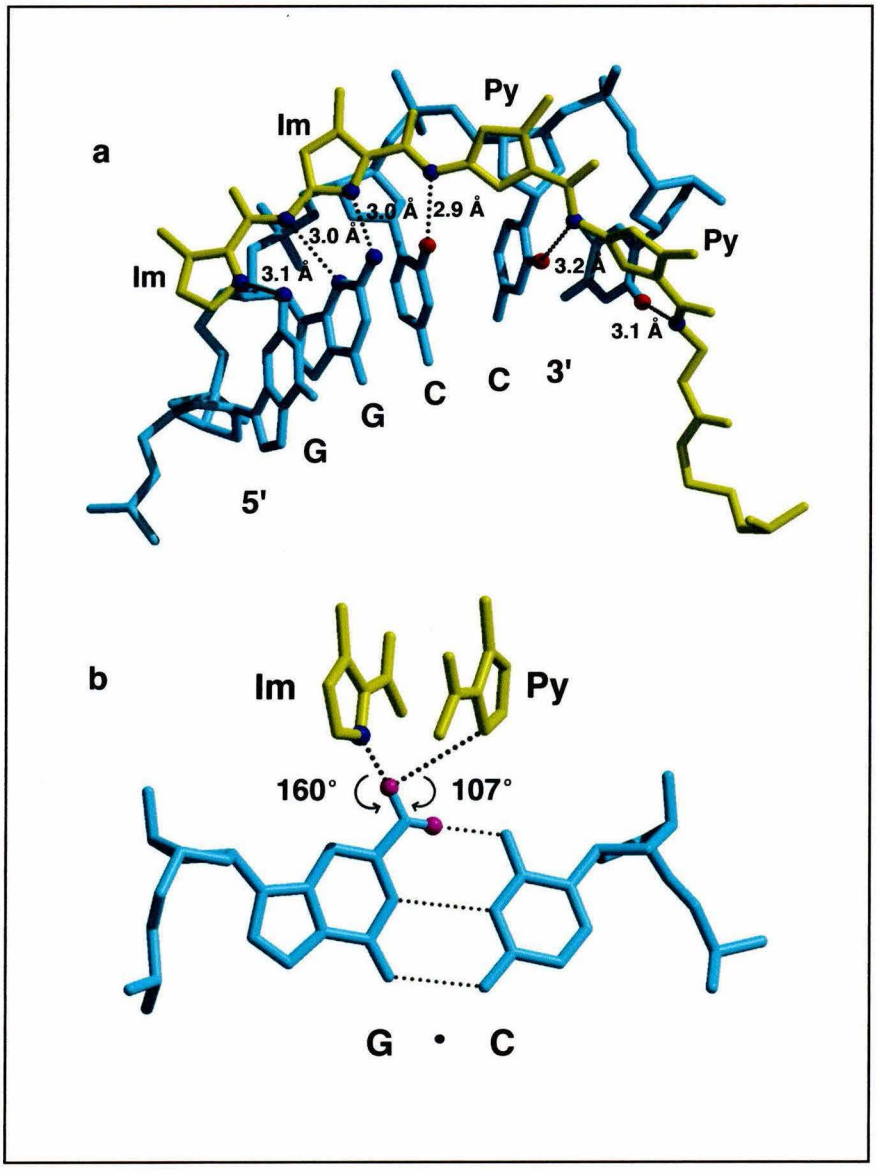
The Im/Py pairing targets the G•C base pair by packing the imidazole in the minor groove with favorable geometry to form a hydrogen bond with the free hydrogen of the guanine exocyclic amine (N2). Although the guanine N2 is approximately in the center of the minor groove (Seeman *et al.*, 1976) (0.3 Å closer to the Gua C1' atom, on average, when measured in the central 5'-GGCC-3' sequence of the liganded and free (Heinemann & Alings, 1989; Goodsell *et al.*, 1993) structures of the oligonucleotide duplex) an Im/Py pair has a 100-fold greater affinity for a G•C base pair than does a Py/Im pair. Individual hydrogen atoms cannot be distinguished in the structure; however, hydrogens consistent with small molecule structures can be assigned using the program X-PLOR (Brünger, 1992). The Gua N2-H ··· Im N3 angle (~160°, Figure 2.4b) falls within most probable range of hydrogen bond angles, whereas the possible hydrogen bond angle to an imidazole over the cytosine of the base pair, extrapolated from the angle to the opposite pyrrole (~107°), is less than the most acute angle (132°) observed in a survey of two-center hydrogen bonds (Jeffrey, 1997). Furthermore, with an imidazole,

instead of the pyrrole, on the cytosine side of the G•C base pair, the ring would be expected to move deeper into the minor groove to form a hydrogen bond, further compressing the angle. The possible hydrogen bond angle (~93°) to the other hydrogen of the guanine amine is even less favorable, and in addition the hydrogen is already engaged in the Watson-Crick base pair. The energetic preference for a linear hydrogen bond, coupled with the unfavorable angle to an imidazole over the cytosine side of the base pair, provides a structural basis for the ability of an Im/Py pair to specifically discriminate G•C from C•G.

Table 2.1. Hydrogen-bond distances between polyamide and DNA.

Polyamide Atom:	ImImPyPyβDp1 DNA strand 1	Distance (Å)	ImImPyPyβDp2 DNA strand 2	Distance (Å)
Im1-N3	G4-N2	3.0	G14-N2	3.1
Im2-N3	G5-N2	3.0	G15-N2	2.9
Am1-N	G5-N3	3.1	G15-N3	2.8
Am2-N	C6-O2	2.9	C16-O2	2.8
Am3-N	C7-O2	3.0	C17-O2	3.4
Am4-N	T8-O2	3.1	T18-O2	3.1
β-ala, Am-N	A13-N3	3.0	A3-N3	-

Figure 2.4. **(a)** Isolated view of one of the ImImPyPy-β-Dp molecules and hydrogen bonds with the adjacent DNA strand. The hydrogen bond lengths given are the average of the ncs-related strands. **(b)** Geometry of the Im/Py polyamide pair interacting with the G5•C16 base pair. An imidazole on the cytosine side of the base pair would be unable to form an optimum hydrogen bond with the hydrogen of the guanine exocyclic amine. This allows the Im/Py pair to discriminate the G•C base pair from C•G.



2.4 Helical Parameters

To evaluate the structural changes associated with polyamide binding, the DNA conformation can be compared with the structure of the same sequence in the absence of ligand (Heinemann & Alings, 1989), and to an oligonucleotide with a central 5'-GGCC-3' sequence and an isomorphous crystal lattice (Goodsell *et al.*, 1993) (Table 2.2a). The twist, pitch, bend and sugar pucker of the DNA helix is B-form and equivalent between the structures. By contrast, the central two G•C base pairs are pried apart by ~20° in the minor groove by the ImImPyPy-β-Dp dimer. The Gua-N2H to Cyt-O2 Watson-Crick hydrogen bonds are elongated by 0.3 Å on average, with respect to the unliganded oligonucleotide structures, by reorientation of these base pairs to optimize the intermolecular hydrogen bonds with the imidazoles as well as to the opposing nucleotide strand. Perhaps due to the slightly different conformation of the leading imidazole ring, described below, the G•C base pairs at the borders of the central 5'-GGCC-3' binding site do not open, and the hydrogen bonds with the first imidazole ring are slightly longer than to the second (Table 2.1).

The minor groove of G•C rich sequences (~6.4 Å) is inherently wider than that of ideal fiber B-form DNA (~5.4 Å) (Heinemann & Alings, 1989; Goodsell *et al.*, 1992). With respect to the unliganded oligonucleotides, the ImImPyPy-β-Dp dimer expands the minor groove by ~1 Å to 7.4 Å, slightly less than the 7.8 Å minor groove widths observed in the distamycin dimer crystal structures (Chen *et al.*, 1994; 1997). The polyamides are enclosed tightly within the nucleic acid backbone, engaged by interactions between the deoxyribose O4's and the first imidazole or the third pyrrole residues. The electron density between the sugar oxygens and these aromatic rings is continuous (Figure 2.5), and the distance between the nucleic acid oxygens and the rings is in some cases less than the sum of the atomic van der Waals radii (Table 2.3). The distance from the sugar O4' is shortest to the center rather than to the atoms of the

leading imidazole, although the methylated nitrogen and adjacent carbon of one of the polyamides are also close. The sugar O4' is closer to the third pyrrole C3 atom than to any of the other atoms or the center of the ring. However, the predicted C-H ··· O4' angle is too acute (70.8°) to be a C-H hydrogen bond (Jeffrey, 1997), and is probably an interaction similar to that between the sugar O4' and the leading imidazole. Close interactions between the DNA O4' and the drug aromatic rings were also observed in the structures of the distamycin dimer (Chen *et al.*, 1994; 1997) and ImIm bound to DNA (Kopka *et al.*, 1997). Intimate deoxyribose O4' contacts with aromatic systems are well-documented and have been suggested to play important roles in the stability of many diverse nucleic acid structures (Bugg *et al.*, 1971; Egli & Gessner, 1995).

Calculations based on 1:1 PyPyPy•DNA crystal structures led to the notion that the rise-per-pyrrole was greater than the pitch of the DNA helix (Goodsell & Dickerson, 1986). The four consecutive pyrrole-imidazole polyamide residues allow calculation of the best-fit helical parameters of the polyamide dimer for comparison with the DNA helix (Table 2.2b). Strikingly, the rise-per-aromatic amino acid residue is very similar to the rise-per-base pair. However, the twist of the polyamide helix is 14° overwound with respect to the twist of the binding site. When the polyamide is extended by application of the helical transformations to the best fit coordinates of the monomer, a steric clash with the B-form oligonucleotide results. The fifth residue is apparently acceptable, but the sixth has bad contacts, and by the seventh residue, the polyamide is penetrating the base pairs (Figure 2.6). Quantitative DNase I footprint titrations demonstrate that pyrrole-imidazole polyamide binding affinity increases with length up to five residues, plateaus between five and seven residues, then plummets as the length is increased further (Kelley *et al.*, 1996). The insertion of a β-alanine linker within a series of imidazole or pyrrole-carboxamide subunits restores the increase in binding affinity with length (Swalley *et al.*, 1997a). The helical parameters of the polyamide provide a structural basis for these observations: rather than correcting a translational discrepancy, the flexible linker resets

the rotation of the relatively rigid pyrrole or imidazole carboxamides to fit the curvature of B-DNA.

When the rotation of the residues of ImImPyPy-β-Dp are compared with the corresponding monomers of the idealized polyamide helix, the obvious outliers are the leading imidazoles. The first imidazole residue is rotated by 20° with respect to the ideal coordinates in one polyamide, and 14° in the other. The singularity of the N-terminal imidazoles can be explained in terms of their unique interactions with the nucleic acid backbone. These include the sugar O4'-aromatic contact discussed above, and in addition, an interaction between the edge of the first aromatic ring and the preceding deoxyribose O4' (Table 2.3). The latter interaction could be described as a interaction with the quadrupole moment of the aromatic ring (Mecozzi *et al.*, 1996), or as a C-H hydrogen bond (Jeffrey, 1997). The C-H ··· O4' angle is nearly linear, and the carbon to oxygen distance is less than the sum of the atomic van der Waals radii (Table 2.3). For the idealized monomer, the distances between the centers and the edges of the leading imidazoles to the sugar O4's are much longer. The imidazole with the largest deviation from the idealized coordinates is also the one with the closest interactions. Since the leading imidazole twists in order to achieve close contacts with the DNA backbone, it is probable that the contacts are energetically favorable.

Figure 2.5. $2|F_o| - |F_c|$ map, contoured at 1σ, illustrating the close interaction between the nucleic acid O4' and the leading imidazole ring of one of the polyamides.

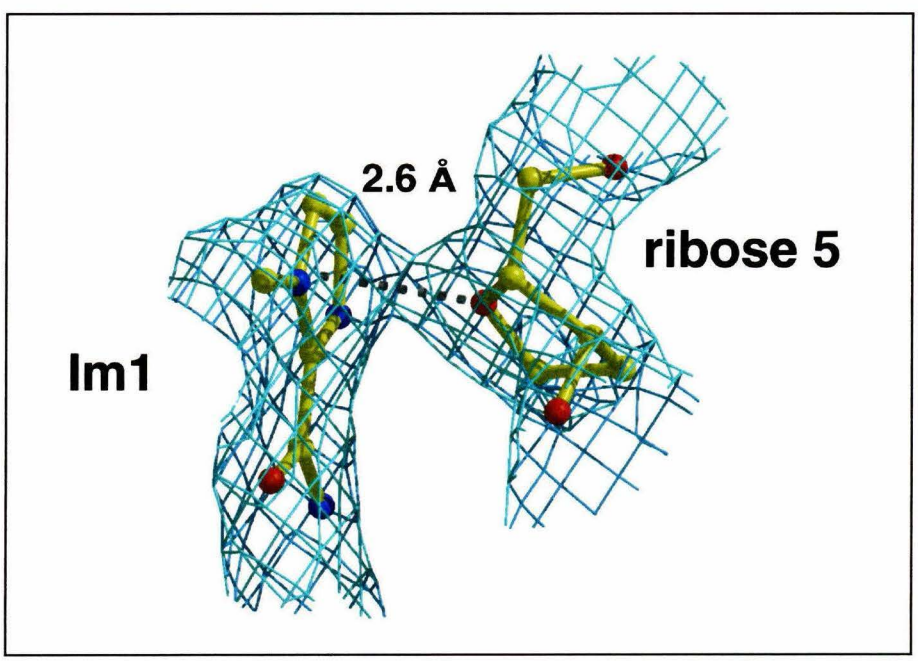


Table 2.2. Helical parameters.

(a)

Helical Parameters[1]	CCA**GGCC**TGG ImImPyPy-β-Dp	CAT**GGCC**TAG isomorphous crystal[2]	CCA**GGCC**TGG no ligand[3]
Minor groove width (Å)[4]	7.4	6.5	6.4
Bend (°)	9.9	5.0	8.1
Twist (°)	35.4	35.4	35.8
Rise (Å)	3.4	3.4	3.4
Propellar twist (°)	-12.9	-12.9	-11.4
Phase angle (°)	160 (C2' endo)	150	147
Overall opening (°)	-0.5	6.6	2.3
Individual central base-pair openings: G4,G5,C6,C7 (°)	0.8, -12.4, -11.5, 0.3	4.8, 16.2, 10.9, -1.3	-0.6, 4.0, 5.2, 0.0

(b)

	Polyamide helix	DNA helix
Twist	49°	35°
Rise	3.5 Å	3.4 Å

[1] (Lavery & Sklenar, 1988)
[2] (Goodsell *et al.*, 1993)
[3] (Heinemann & Alings, 1989)
[4] Calculated for central GGCC sequence

Figure 2.6. An extended polyamide helix. Least squares minimized, mean coordinates for a polyamide monomer were translated and rotated by the parameters of the polyamide helix. After a period of approximately seven residues, the 14° greater twist of the polyamide with respect to B-form DNA manifests as clash with the walls of the minor groove. The number of contacts closer than 3.4 Å between the DNA nucleotides and the polyamides is indicated by a color gradient, that ranges from blue, for 1 to 4 contacts, via red, for 94 to 109 contacts, to yellow for completely passing through the DNA helix.

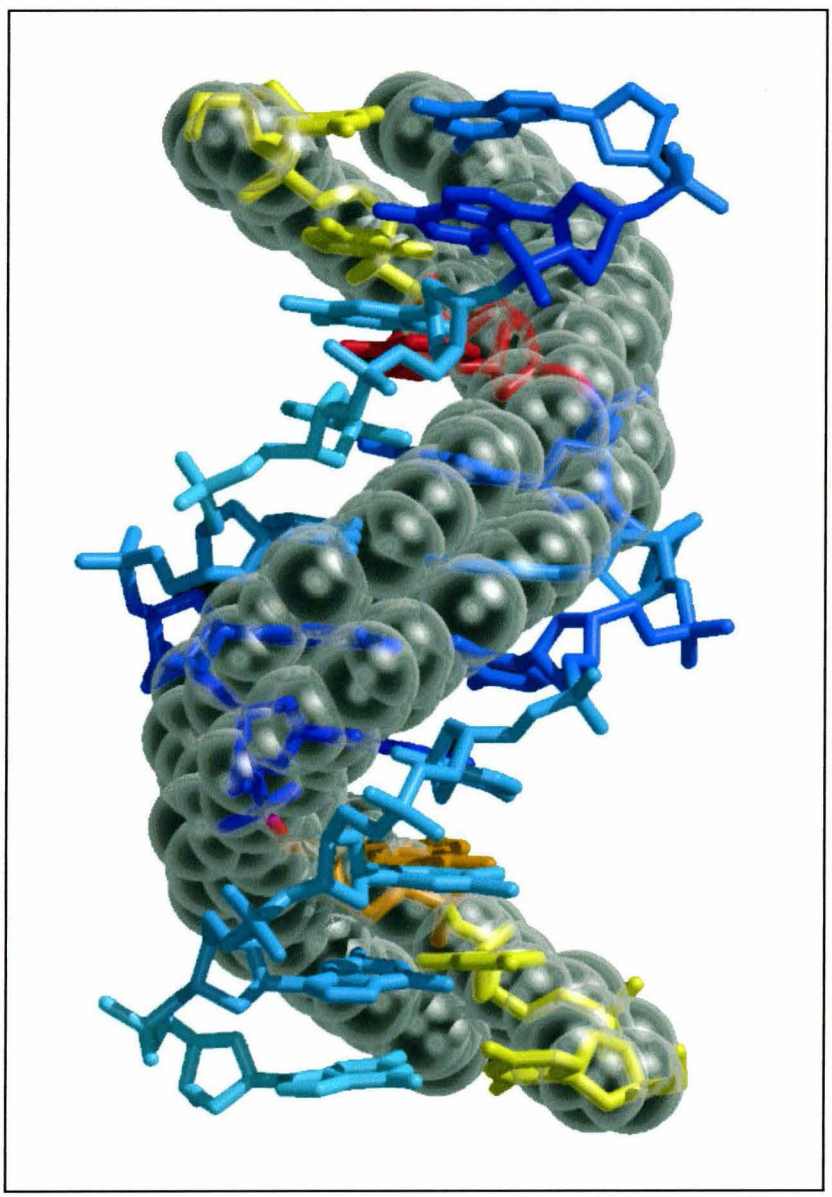


Table 2.3. Distances for aromatic ring-O4' interactions.

Polyamide Atom:	ImImPyPy-βDp1 DNA strand1[1] (Å)	ImImPyPy-βDp2 DNA strand2 (Å)	Sum of van der Waals radii (Å)
Im1 to G5 or G15-O4':			
Im1-N1	3.0	2.8	2.9
Im1-C2	3.0	2.8	3.0
Center	2.8 (3.1)	2.6 (3.0)	
Py3 to C7 or C17-O4':			
Py3-C3	2.8	2.8	3.0
Center	3.1	3.3	

Polyamide/DNA atoms:	C ··· O4' distance[5](Å)	C-H ··· O4' angle (°)	Sum of van der Waals radii (Å)
Im1-C4 to G4-O4'	3.8 (4.1)	161.3	4
Im1-C4 to G14-O4'	3.6 (4.5)	164.0	4

[5]Numbers in parentheses are corresponding distance measured to best fit polyamide helix.

2.5. Conclusion

The hydrogen bond array necessary for specific recognition of the DNA helix by pyrrole-imidazole polyamides has been predicted through a combination of thermodynamic and NMR studies (Wade *et al.*, 1992; Mrksich *et al.*, 1992; Geierstanger *et al.*, 1994; Swalley *et al.*, 1997b; Pilch *et al.*, 1996). The crystal structure reported here provides direct confirmation that all potential hydrogen bonds are formed between the imidazole N3 and the exocyclic amine of guanine as well as hydrogen bonds between the imidazole- and pyrrolecarboxamide NH and the respective N3 of Gua and the O2 of Cyt. The perfect match of the rise-per-aromatic amino acid residue with the pitch of the DNA duplex puts to rest a long standing question regarding curvature and register of the ligand•DNA complex (Goodsell & Dickerson, 1986). The overwound curvature of more than four consecutive rings explains the observation that flexible β-alanine residues reset an optimum fit of the polyamide dimer with the DNA helix at long binding sites (Swalley *et al.*, 1997a). Additionally, these results suggest that selectively placed 'β-spring' amino acids may allow for improved positioning of Im residues in order to recognize certain problematic G,C-rich sequences (Swalley *et al.*, 1997b). A wealth of biochemical and structural information has identified general principles of protein-DNA recognition, with a versatile recognition code emerging from studies of certain zinc finger-DNA complexes (Choo & Klug, 1997). A similar database of high resolution small molecule•DNA structures is notably absent. The knowledge of the position of each atom within a DNA-bound synthetic G•C-specific ligand provides a new baseline model of molecular design for minor groove recognition.

2.6 Methods

Crystal preparation. The four ring polyamide ImImPyPy-β-Dp was synthesized by solid-phase methods, and purified by reverse-phase high-performance liquid chromatography (HPLC) (Baird & Dervan, 1996; Swalley *et al.*, 1997a). Synthetic deoxyoligonucleotides were purified twice by reverse-phase HPLC, annealed to form double-stranded DNA, and then incubated with a 2:1 ratio of polyamide to DNA duplex prior to crystallization. Crystals were obtained from a solution of 0.6 mM duplex DNA, 1.3 mM polyamide, 12% 2-methyl-2,4-pentanediol, 35 mM Ca acetate, 10 mM Tris pH 7.5 equilibrated in sitting drops against a reservoir of 25% MPD at 4° C.

X-ray structure determination. Native crystals grew in space group C2 with unit cell dimensions **a**=63.6 Å, **b**= 26.9 Å, **c**=43.5 Å, β=123.6°, with one duplex molecule per asymmetric unit (a.s.u.). Datasets for the C2 crystals and the BrCyt6 data set 1 were collected on an R-axis IIc imaging plate detector with CuKα radiation. All remaining data were collected at Stanford Synchrotron Radiation Laboratory (SSRL) beamline 7-1 with a MAR Research imaging plate detector at wavelength 1.08 Å (Table 2.4). Data was processed with DENZO and scaled with SCALEPACK (Otwinowski, 1993). Further manipulations were performed with the CCP4 program suite (Bailey, 1994), unless otherwise indicated. Phases from a molecular replacement solution (Navaza, 1994) in space group C2 using B-DNA as the starting model resulted in uninterpretable electron density for the polyamide and a crystallographic R-factor of 36%. Consequently, 5-Bromo-uridine (BrUra) and 5-Bromo-cytosine (BrCyt) heavy atom derivatives were prepared. Although the BrUra crystal form was isomorphous with the native, the BrCyt substituted oligonucleotides crystallized in a different space group ($P2_12_12_1$, **a**=34.4 Å, **b**=39.0 Å, **c**=44.4 Å, with one duplex molecule per asymmetric unit). The bromine positions were obtained from difference Fourier maps calculated with molecular

replacement phases, and refined using PHASES (Furey & Swaminathan, 1997). A polyamide dimer was clearly visible in the experimental map calculated with multiple isomorphous replacement (MIR) phases. The phases were improved with non-crystallographic averaging, averaging between the crystal forms, and solvent flattening with the program Dmmulti (Cowtan, 1994). Although of higher resolution, the BrCyt6 data set collected at the synchrotron suffered from the diffuse scattering of the pseudo-continuous DNA helices of the crystal lattice. To improve the quality of the data, BrCyt6 synchrotron and R-axis data sets were scaled and merged for use in the refinement (Otwinowski, 1993).

Refinement. The coordinates were fit into the improved electron density map using the program O (Jones *et al.*, 1991), and refined with the program X-PLOR (Brünger, 1992) using the DNA parameters of Parkinson *et al.* (1991) and polyamide parameter and topology files prepared with use of XPLO2D (Kleygwegt & Jones, 1994). Restraints on the non-crystallographic symmetry (ncs) and on the planarity of the bases, peptide bonds, pyrroles, and imidazoles were included until the latter stages of the refinement, as guided by the behavior of the free R-factor. The final r.m.s. difference is 0.5 Å between ncs-related strands. After inclusion of 70 water molecules above 3σ level in difference maps, the crystallographic R-factor is 22.3%, free R-factor 25.9% (for the free R-factor, 5% of the data were set aside from the refinement), for all data, with bulk solvent and overall anisotropic B-factor corrections, between 20.0 and 2.0 Å resolution. The r.m.s. deviations from ideal bond lengths and angles are 0.028 Å and 2.2°, respectively. Apart from the C-terminal Dp tails of the polyamides, which are partially occupied, the polyamide molecules are well defined in an omit electron density map (Figure 2.1). Refined coordinates have been deposited in the NDB, accession code GDJ057.

Structure analysis. The angle between the normals to the planes of the polyamide rings and the walls of the minor groove was calculated by minimizing $\Sigma(\Psi_o - \Psi_c)^2$, where $\Psi(n,\omega,\theta,\alpha) = \mathrm{Cos}^{-1}[\{\mathrm{Cos}(n\omega)\}(\cos^2\alpha \sin^2\theta + \sin^2\alpha) + \cos^2\alpha \cos^2\theta]$ (Ψ, angle between plane i and plane (i + n) normals; ω, twist of the helix; α, tilt of plane normal from helix axis out of minor groove; θ, tilt of plane normal from helix axis along minor groove).

DNA helical parameters (Table 2.2a) were calculated using the program Curves (Lavery & Sklenar, 1988). Parameters for the polyamide helix (Table 2.2b) were obtained by least squares minimization of a function describing the helix using the program Mathematica, given the observed coordinates, and assuming that the polyamide helix is coaxial with the DNA helix. Residual = $\Sigma_I[y-\{(y_o- b)\mathrm{Cos}(n\Phi)-(z_o-c)\mathrm{Sin}(n\Phi)+b\}]^2$ + $[z-\{(z_o-c)\mathrm{Cos}(n\Phi)-(z_o-c)\mathrm{Sin}(n\Phi)+c\}]^2+\{x-(x_o+nP)\}^2$, i.e., the square difference between the observed and calculated equation for a helix over all atom types *i* for each monomer n, given that the helical axis is parallel to the **x** axis (verified from DNA helix axis coordinates output by Curves). Φ, twist of helix per residue; P, translation along helix axis per residue; (b,c) coordinates of helix axis in yz plane; (x,y,z) experimental coordinates of atom type *i* in monomer n; (x_o,y_o,z_o) best fit coordinates for atom type *i*. Parameters x_o, y_o, z_o, b, c, Φ, and P are minimized. Helical parameters calculated by this method for the DNA, using the sugar C1' and purine N2 or pyrimidine N3 coordinates of the central 5'-GGCC-3' are consistent with those from Curves (Table 2.2), and provide a reference for the polyamide helical parameters. Polyamide helical parameters calculated by three alternative methods give equivalent results (not shown). For measurements, hydrogens with standard geometry were assigned by use of X-PLOR (Brünger, 1992) and the atomic van der Waals radii C, 1.6; N, 1.5; O, 1.4, H; 1.0 Å (Weast, 1978) were used. Surface area was calculated with the program GRASP (Nicholls, 1993), and figures were prepared by use of MOLSCRIPT (Kraulis, 1991), BOBSCRIPT (Esnouf, 1997) and RASTER3D (Bacon & Anderson, 1988; Merritt & Murphy, 1994).

Acknowledgments. We are grateful to the NIH for research support, the NSF for a predoctoral fellowship to C.L.K., and the HHMI for a predoctoral fellowship to E.E.B. We thank J.E. Wedekind, H. Schindelin, C. Kisker, L. Joshua-Tor, P.M. Takahara and the members of the Rees group for guidance during the structure determination. The rotation camera facility at SSRL is supported by the Department of Energy and NIH.

Table 2.4. Crystallographic statistics.

Parameter	Native	BrUra8	BrCyt2	BrCyt7	BrCyt6 Data 1	BrCyt6 Data 2	BrCyt6 Data 1 & 2, merged
Spacegroup	C2	C2	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
d (Å)	2.5	2.7	2.2	2.5	2.4	2.0	2.0
Completeness (Last Shell) %	94.5 (99.6)	99.6 (98.8)	94.9 (95.8)	93.1 (99.6)	90.0 (92.9)	88.2 (88.5)	93.9 (88.5)
Reflections Measured (Unique)	5,837 (2,065)	5,114 (1,697)	13,521 (3,169)	10,427 (2,144)	10,256 (2,199)	15,075 (3,867)	25,331 (4,124)
Average I/σ	14.2 (3.6)	19.2 (7.9)	28.2 (11.1)	18.2 (6.6)	56.9 (13.8)	57.4 (9.3)	82.2 (9.5)
R_{sym}[1] (Last Shell) %	7.3 (24.9)	4.9 (16.2)	3.1 (21.9)	6.8 (23.2)	2.2 (11.1)	2.2 (17.2)	2.7 (14.5)
R_{cullis} (15-3 Å)[2]	-	0.59	0.71	0.85	-	-	-
Phasing Power[3]	-	2.25	1.86	1.14	-	-	-

[1] $R_{sym}=\sum_{hkl}\sum_{I}|I_i - \langle I\rangle|/\sum_{hkl}\sum_{I}\langle I\rangle$ where I_i is the ith measurement and $\langle I\rangle$ is the weighted mean of all measurements of I.

[2] $R_{cullis}=\sum_{hkl} ||F_{ph}| - |F_p||/\sum_{hkl}|F_{ph} - F_p|$

[3] Phasing power is the mean value of the heavy atom derivative structure factor amplitude divided by the residual lack of closure error.

2.7 References

Bacon, D. J. & Anderson, W. F. (1988). A fast algorithm for rendering space filling molecule pictures. *J. Mol. Graph.* **6**, 219-220.

Bailey, S. (1994). The CCP4 suite- Programs for protein crystallography. *Acta Crystallogr. D* **50**, 760-763.

Baird, E. E. & Dervan, P. B. (1996). Solid phase synthesis of polyamides containing imidazole and pyrrole amino acids. *J. Am. Chem. Soc.* **118**, 6141-6146.

Brunger, A. T. *X-PLOR Version 3.1. A system for X-ray Crystallography and NMR,* (University Press, New Haven, 1992).

Bugg, C. E., Thomas, J. M., Sundaralingam, M. & Rao, S. T. (1971). Stereochemistry of nucleic acids and their constituents. X. Solid state base-stacking patterns in nucleic acid constituents and polynucleotides. *Biopolymers.* **10**, 175-219.

Chen, X., Ramakrishnan, B., Rao, S. T. & Sundaralingam, M. (1994). Binding of two distamycin A molecules in the minor groove of an alternating B-DNA duplex. *Nature Struct. Biol.* **1**, 169-175.

Chen, X., Ramakrishnan, B. & Sundaralingam, M. (1997). Crystal structures of side-by-side binding of distamycin to AT-containing DNA octamers d(ICITACIC) and d(ICATATIC). *J. Mol. Biol.* **267**, 1157-1170.

Choo, Y. & Klug, A. (1997). Physical basis of a protein-DNA recognition code. *Curr. Opin. Struct. Biol.* **7**, 117-125.

Cowtan, K. *Joint CCP4 amd ESF-EACBM Newsletter on Protein Crystallography* **31**, 34-38 (1994).

Egli, M. & Gessner, R. V. (1995). Stereoelectronic effects of deoxyribose O4' on DNA conformation. *Proc. Natl. Acad. Sci. USA* **92**, 180-184 (1995).

Esnouf, R. M. (1997). An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph.* **15**, 132-134.

Furey, W. & Swaminathan, S. in *Meth. Enzymology* (eds. Carter, C. & Sweet, R.) (Academic Press, Orlando, FL, 1995).

Geierstanger, B. H., Mrksich, M., Dervan, P. B. & Wemmer, D. E. (1994b). Design of a G•C specific DNA minor groove binding peptide. *Science*. **266**, 646-650.

Goodsell, D. & Dickerson, R. E. (1986). Isohelical analysis of DNA groove-binding drugs. *J. Med. Chem.* **29**, 727-733.

Goodsell, D. S., Kopka, M. L., Cascio, D. & Dickerson, R. E. (1993). Crystal structure of CATGGCCATG and its implications for A-tract bending models. *Proc. Natl. Acad. Sci. USA* **90**, 2930-2934.

Gottesfeld, J. M., Neely, L., Trauger, J. W., Baird, E. E. & Dervan, P. B. (1997). Regulation of gene expression by small molecules. *Nature*. **387**, 202-205.

Heinemann, U. & Alings, C. (1989). Crystallographic study of one turn of GC-rich B-DNA. *J. Mol. Biol.* **223**, 369-381 (1989).

Jeffrey, G. A. *An Introduction to Hydrogen bonding* (Oxford University Press, New York, 1997).

Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog.* **A47**, 110-119.

Kelly, J. J., Baird, E. B. & Dervan, P. B. (1996). Binding site size limit of the 2:1 pyrrole-imidazole polyamide-DNA motif. *Proc. Natl. Acad. Sci. USA* **93**, 6981-6985.

Kleygwegt, G. J. & Jones, T. A. In *From first map to final model* (eds Hubbard, R. & Waller, D. A.) 59-66 (SERC Daresbury Laboratory, UK; 1994).

Kopka, M. L., Goodsell, D. S., Han, G. W., Chiu, T. K., Lown, J. W. & Dickerson, R. E. (1997). Defining GC-specificity in the minor groove: side-by-side binding of the di-imidazole lexitropsin to C-A-T-G-G-C-C-A-T-G. *Structure* **5**, 1033-1046.

Kopka, M. L., Yoon, C., Goodsell, D., Psura, P. & Dickerson, R. E. (1985). The molecular origin of DNA-drug specificity in netropsin and distamycin. *Proc. Natl. Acad. Sci. USA* **82**, 1376-1380.

Kraulis, P. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946-950.

Lavery, R. & Sklenar, H. (1988). The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biolmolec. Struct. Dynam.* **6**, 63-91.

Mecozzi, S., West, A. P. & Dougherty, D. A. (1996). Cation-pi interactions in aromatics of biological and medicinal interest: Electrostatic potential surfaces as a useful qualitative guide. *Proc. Natl. Acad. Sci. USA* **93**, 10566-10571 (1996).

Merritt, E. A. & Murphy, M. E. P. (1994). Raster3D version 2.0- a program for photorealistic molecular graphics. *Acta Crystallog.* **D50**, 869-873.

Mrksich, M., Wade, W. S., Dwyer, T. J., Geierstanger, B. H., Wemmer, D. E. & Dervan, P. B. (1992). Antiparallel side-by-side dimeric motif for sequence-specific recognition in the minor groove of DNA by the designed peptide 1-methylimidazole-2-carboxamide netropsin. *Proc. Natl. Acad. Sci. USA* **89**, 7586-7590.

Navazza, J. (1994). AmoRe, an automated package for molecular replacement. *Acta Crystallog.* **A50**, 157-163.

Nicholls, K. A., Baradwaj, R. & Honig, B. (1993). GRASP- Graphical representation and analysis of surface properties. *Biophys. J.* **64**, A166.

Otwinowski, Z. in *Data Collection and Processing* (eds Sawyer, L., Isaacs, N. & Bailey, S.) 56-62 (SERC Daresbury Laboratory, UK; 1993).

Parkinson, G., Vojtechovsky, J., Clowney, L., Brünger, A. T. & Berman, H. M. (1991). New parameters for the refinement of nucleic acid-containing structures. *Acta Crystallog.* **A47**, 110-119.

Pelton, J. G. & Wemmer, D. E. (1989). Structural characterization of a 2-1 distamycin-A •d(CGCAAATTTGCG)$_2$ complex by two-dimensional NMR. *Proc. Natl. Acad. Sci. USA* **86**, 5723-5727.

Pilch, D. S., Poklar, N., Gelfand, C. A., Law, S. M., Breslauer, K. J., Baird, E. E. & Dervan, P. B. (1996). Binding of a hairpin polyamide in the minor groove of DNA: Sequence-specific enthalpic discrimination. *Proc. Natl. Acad. Sci. USA* **93**, 8306-8311.

Seeman, N. C., Rosenburg, J. M. & Rich, A. (1976). Sequence specific recognition of double helical nucleic acids by proteins. *Proc. Natl. Acad. Sci. USA* **73**, 804-808.

Swalley, S. E., Baird, E. E. & Dervan, P. B. (1997). A pyrrole-imidazole polyamide motif for recognition of eleven base pair sequences in the minor groove of DNA. *Chem. Eur. J.* **3**, 1608-1616.

Swalley, S. E., Baird, E. E. & Dervan, P. B. (1997). A pyrrole-imidazole polyamide motif for recognition of eleven base pair sequences in the minor groove of DNA. *J. Am. Chem. Soc.* **119**, 6953-6961.

Trauger, J. W., Baird, E. E. & Dervan, P. B. (1996a). Recognition of DNA by designed ligands at subnanomolar concentrations. *Nature*, **382**, 559-561.

Wade, W. S., Mrksich, M. & Dervan, P. B. (1992). Design of peptides that bind in the minor groove of DNA at 5'-(A,T)$_5$G(A,T)C(A,T)-3' sequences by a dimeric side-by-side motif. *J. Am. Chem. Soc.* **114**, 8783-8794.

Weast, R. C.(ed.) in *Handbook of Chemistry and Physics.* (CRC Press, Inc. Florida, 1996).

White, S., Baird, E. E. & Dervan, P. B. (1996). Effects of the A•T/T•A degeneracy of pyrrole-imidazole polyamide recognition in the minor groove of DNA. *Biochemistry.* **35**, 12532-12537.

White, S., Baird, E. E. & Dervan, P. B. (1997a). On the pairing rules for recognition in the minor groove of DNA by pyrrole-imidazole polyamides. *Chem. Biol.* **4**, 569-577.

White, S., Baird, E. E. & Dervan, P. B. (1997b). Orientation preferences of pyrrole-imidazole polyamides in the minor groove of DNA.*J. Am. Chem. Soc.* **119**, 8756-8765.

Chapter 3

Structural Basis for T•A Recognition in the DNA Minor Groove

Abstract: Polyamide dimers containing three types of aromatic rings, pyrrole (Py), imidazole (Im), and hydroxypyrrole (Hp), afford a small molecule recognition code that discriminates among all four Watson-Crick base pairs in the minor groove. The crystal structure of a specific polyamide dimer-DNA complex establishes the structural basis for distinguishing T•A from A•T base pairs. Specificity for the T•A base pair is achieved via distinct hydrogen bonds between pairs of substituted pyrroles on the ligand and the O2 of T and N3 of A. In addition, shape-selective recognition of an asymmetric cleft between the thymine-O2 and the adenine-C2 is observed. Although similarities among the base pairs in the minor groove have been emphasized in the past, the structure illustrates differences that allow specific minor groove recognition.

Publication: Kielkopf, White, Szewczyk, Turner, Baird, Dervan & Rees (1998). *Science.* **282**, 111-115.

3.1 Introduction

Before the first structure of a molecule bound to DNA had been determined, specific recognition of double helical B-form DNA was predicted to occur primarily in the major, rather than the minor, groove (Seeman *et al.*, 1976). This proposal was based upon the observation that for A,T base pairs, the hydrogen-bond acceptors at N3 of adenine and O2 of thymine are similarly placed and lack any prominent distinguishing feature (Figure 3.1). Subsequent structures of DNA-binding domains co-crystallized with DNA supported this idea, since most of the specific contacts were made with the major groove (Steitz, 1990). The principle that 'the major groove is a better candidate for sequence-specific recognition than the minor groove' (Branden & Tooze, 1991) continues to provide the basis for strategies to decipher rules for protein-DNA recognition. Although there has been remarkable progress in the design of zinc fingers to recognize the major groove (Choo & Klug, 1997), no protein structure motif has been identified which provides an α-amino acid-base pair code for the minor groove. Eight ring hairpin polyamides have affinities and specificities that rival those of major groove binding proteins (Trauger *et al.*, 1996; White *et al.*, 1997), and have been shown to permeate living cells and inhibit specific gene expression (Gottesfeld *et al.*, 1997). The side-by-side pairing of the residues in the polyamide dimer determines the DNA sequence recognized. An Im/Py pair distinguishes G•C from C•G and both of these from A•T/T•A base pairs (Wade *et al.*, 1992; Trauger *et al.*, 1996; White *et al.*, 1997), and the structural basis of this discrimination is now understood (Mrksich *et al.*, 1992; Geierstanger *et al.*, 1994; Kielkopf *et al.*, 1998). However, a structural understanding for how an Hp/Py pair distinguishes T•A from A•T, and both from G•C/C•G (White *et al.*, 1998), has yet to be established. To address this, we have determined the co-crystal structure of a polyamide of sequence ImHpPyPy-β-Dp (Figure 3.2), bound as a dimer to a self-complementary ten base pair oligonucleotide containing all four Watson-Crick base pairs, 5'-CC**AGTACT**GG-3' (binding site in bold, β, β-alanine, Dp, dimethylamino-propylamide) (Figure 3.3). The structure of the polyamide ImPyPyPy-β-Dp, containing a Py/Py pair that does not distinguish A•T and T•A (White *et al.*, 1996, Pelton & Wemmer, 1989), bound to the same

duplex was solved for comparison. In both the ImHpPyPy and ImPyPyPy structures, the polyamides bind as antiparallel dimers centered over the target GTAC sequence in the minor groove of a B-form DNA duplex (Figure 3.3). The N to C-terminal orientation of each fully overlapped polyamide is parallel to the adjacent 5' to 3' strand of DNA, consistent with previous chemical (White *et al.*, 1997) and structural studies of polyamide dimers (Pelton & Wemmer, 1989; Geierstanger *et al.*, 1994; Chen *et al.*, 1994; 1997; Kopka *et al.*; 1997).

Figure 3.1. Anatomy of the T•A base pair. Arrows indicate potential sites for discrimination of A•T from T•A in the major and minor grooves. Lone pair electrons in the minor groove are shown as ovals and Watson-Crick hydrogen bonds of the base pair as dotted lines. Arrows for previously described sites (Seeman *et al.*, 1976) are colored black, sites identified in this work are in red. The type of potential recognition is labeled: a, hydrogen bond acceptor, d, hydrogen bond donor, vdW, van der Waals.

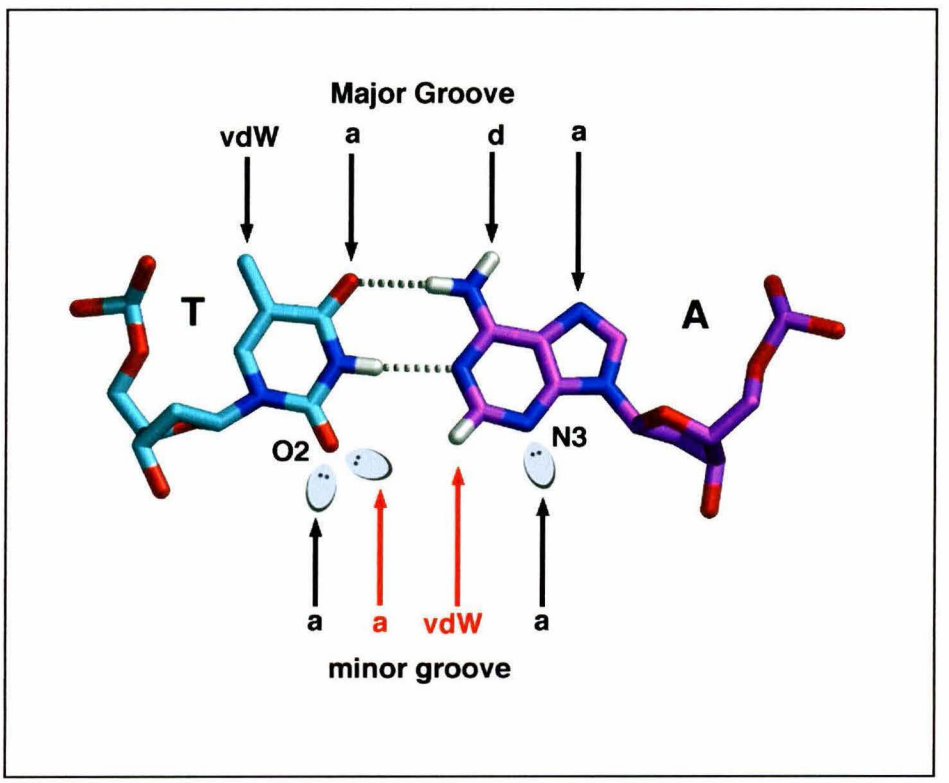


Figure 3.2. An omit $|F_o|$ - $|F_c|$ electron density map for one of the ImHpPyPy polyamide molecules, contoured at 1.5 σ, clearly shows the position of the 3-hydroxyl group. The numbering of the atoms used in the text is indicated below on the chemical structure. The Hp is red; the Py that would be paired with it is yellow. The Im, Py, β and Dp are colored silver.

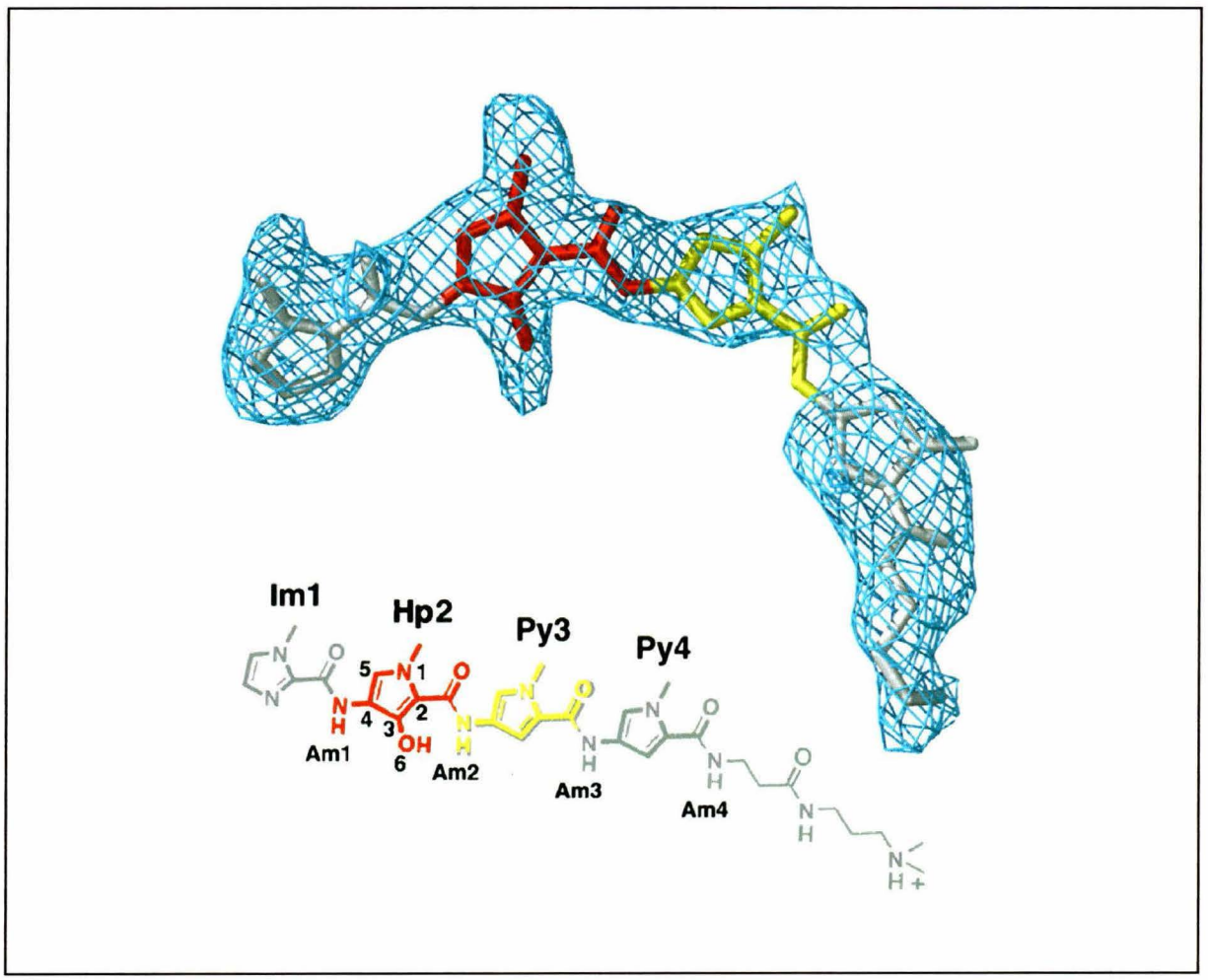
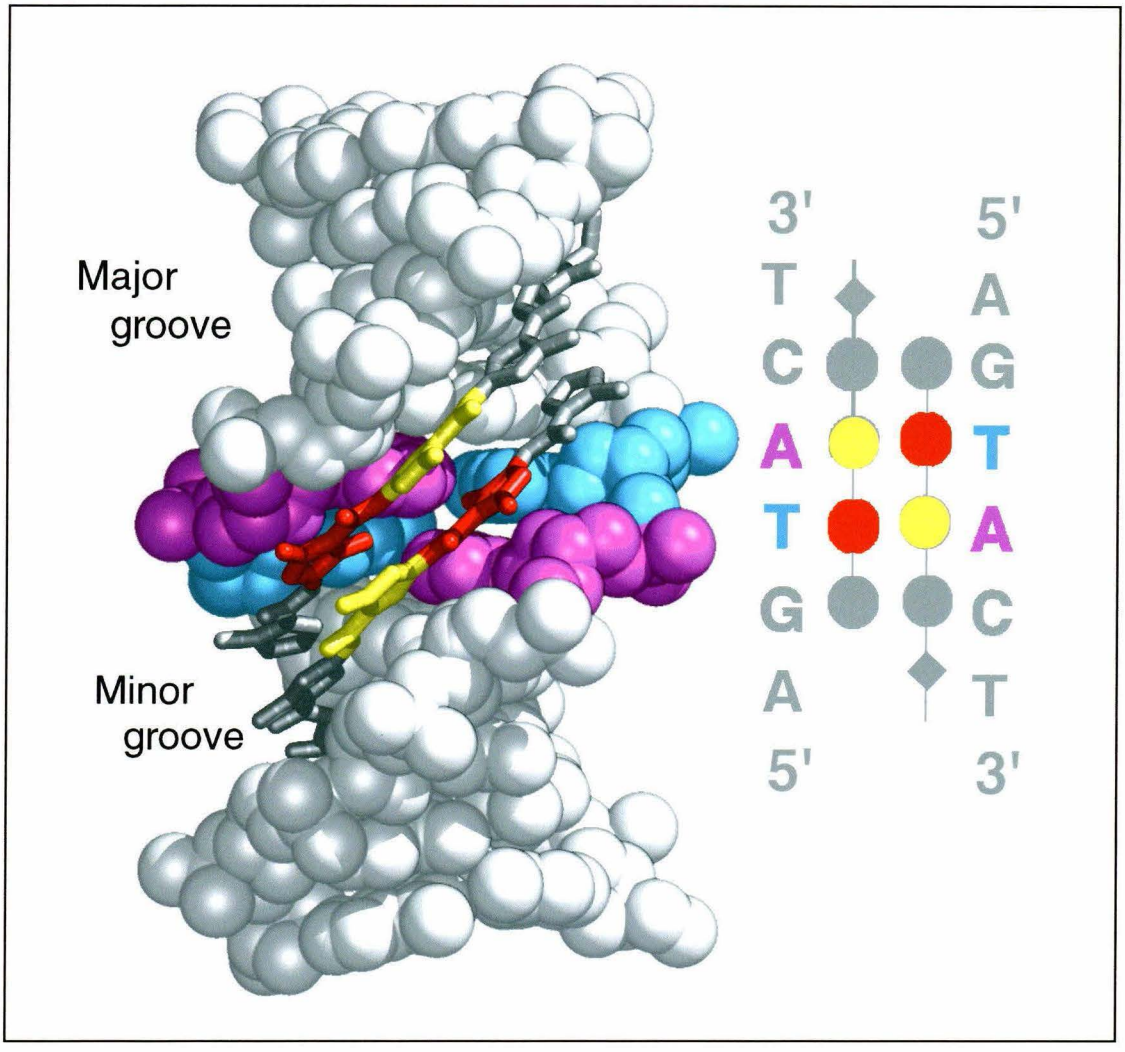


Figure 3.3. Structure of $(ImHpPyPy)_2$•5'-CC**AGTACT**GG-3'. Adenosine is purple and thymidine cyan; polyamide is colored as above. A schematic, space-filling model is shown to the side, with the aromatic residues of the polyamide indicated by filled circles, β by the diamond, and Dp by stick. The overall structure of $(ImPyPyPy)_2$•5'-CC**AGTACT**GG-3' is similar.



3.2 T•A Specific Interactions

Although the functional groups of adenine and thymine are very similar in the minor groove, the number of lone pairs on the hydrogen bond acceptors is different: a thymine-O2 has two free lone pairs, whereas an adenine-N3 has only one (Figure 3.1). The amide nitrogens of the ligand form hydrogen bonds with the purine-N3 (A or G) or pyrimidine-O2 (T or C). As a result, the hydrogen bond potential of adenine-N3 is filled when an imidazole-pyrrole polyamide is bound, but the thymine-O2 has the capacity to accept an additional hydrogen bond. We find that both the hydroxyl group of the Hp, and the amide-NH of the preceding residue, form hydrogen bonds with the target thymine-O2 of the adjacent DNA strand (Figure 3.4a). A similar interaction between the Hp and the adenine-N3 would be impossible without loss of the hydrogen bond from the preceding amide-NH.

Despite a hydrogen bond of favorable length (Table 3.1) between Hp and thymine-O2, it was possible that the position of the hydroxyl and the amide out of the plane of the thymine-O2 sp^2-hybridized lone pairs would weaken the hydrogen bonds. The thymine-C2=Ô2···O6-Hp and thymine-C2=Ô2···N-Am angles and their out-of-plane and in-plane components were calculated to be 17 or 35° out of the plane, for the Hp and Am respectively, and 25° in the plane (Table 3.2). The observed values for the components of the hydroxyl and amide hydrogen bond angles with thymine were found to be comparable to hydrogen bond angles between carbonyls and waters in protein structures, which range from ~0-60° for both the in- and out-of-plane angles (Baker & Hubbard, 1984). In addition, the out-of-plane thymine-C2=Ô2···N-Am components in the ImPyPyPy structure are approximately the same as those of ImHpPyPy, indicating that formation of an additional hydrogen bond with the hydroxyl does not substantially perturb the hydrogen bond geometry between the amide and the thymine-O2.

In addition to the difference in number of lone pairs of the adenine-N3 versus thymine-O2, adenine is also distinguished from thymine by a bulkier aromatic ring. Although the adenine-C2-H does not protrude into the minor groove like the guanine exocyclic amine, the additional carbon results in an asymmetric cleft in the minor groove of a T•A base pair (White *et al.*, 1998;

Wong & Bateman, 1994) (Figure 3.1). The adenine-C2 of the ImHpPyPy structure contacts the Hp hydroxyl (Figure 3.4b). Modeling the target thymine as an adenine reveals that the C2 carbon of a mismatch 'adenine' opposite an Hp residue would sterically overlap the hydroxyl by 1-2 Å (depending upon the hydrogen positions). Furthermore, the orientation of the Hp hydroxyl observed in the ImHpPyPy structure, 3.5 Å from the adenine-C2, with an average adenine-C2-H···O6-Hp angle of 165° (depending upon the hydrogen positions) (Table 3.1), indicates that the Hp-O6 forms a favorable C-H hydrogen bond with the adenine-C2-H. As in this case, C-H hydrogen bonds are strongest between aromatic carbons adjacent to nitrogen atoms with oxygen hydrogen bond acceptors (Wahl & Sundaralingam, 1997). Shape selective recognition of the asymmetric cleft is the second feature that allows the Hp/Py pair to discriminate T•A from A•T.

Table 3.1. Hydrogen Bond Distance (Å) and Donor-H···Acceptor Angle (°).[1]

Polyamide-DNA Atoms	ImHpPyPy1 to DNA Strand 1		ImHpPyPy2 to DNA Strand 2		ImPyPyPy1 to DNA Strand 1		ImPyPyPy2 to DNA Strand 2	
	length (Å)	angle (°)	length (Å)	angle (°)	length (Å)	angle (°)	length (Å)	angle (°)
Im1N3···GuaN2	3.0	158	3.0	153	3.2	164	3.2	157
Am1N···ThyO2	2.5	134	3.0	139	2.8	136	2.8	137
Am2N···AdeN3	3.2	144	3.8	148	3.0	156	2.8	159
Am3N···CytO2	3.5	147	3.4	135	3.1	157	2.9	166
Am4N···ThyO2	3.1	134	3.1	132	3.0	132	2.7	147
Hp2O6···AdeC2	3.5	160	3.4	179	-	-	-	-
Hp2O6···ThyO2	2.8	167	2.8	173	-	-	-	-
	ImHpPyPy1, Intramolecular		ImHpPyPy2, Intramolecular					
Hp2O6··· Am2N	2.9	131	3.0	130	-	-	-	-

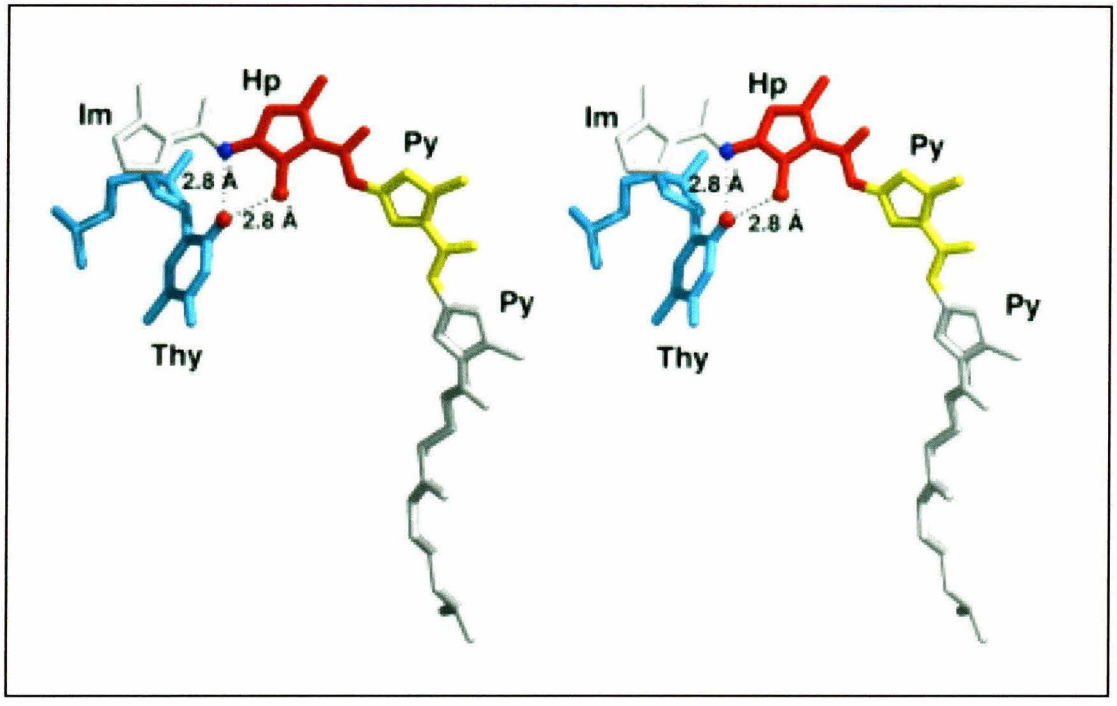
[1] Distances were measured between donor and acceptor atoms. Angles are those formed by donor-H···acceptor atoms. Hydrogens were assigned with standard geometry to sp2-hybridized groups by use of X-PLOR (Brünger, 1992). The hydroxyl hydrogen was assigned with standard bond length and angle in the orientation closest to the Thy-O2 yet having favorable staggered conformation with respect to the hydroxypyrrole ring.

Table 3.2. Analysis of Hydrogen Bond Acceptor Angles (°) at the Target Thymines.[2]

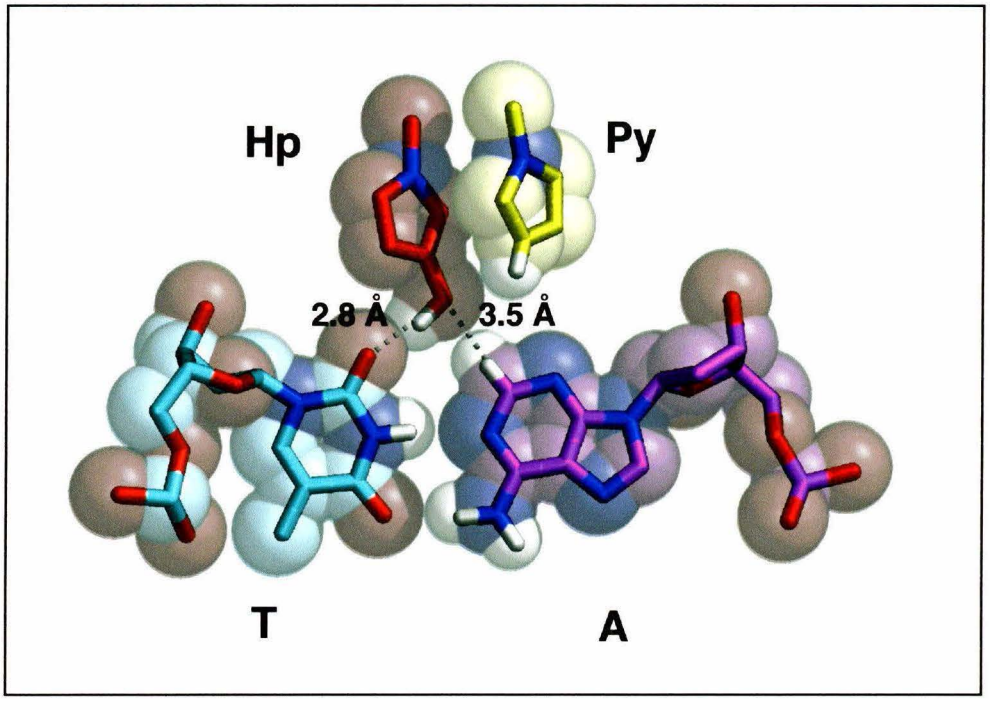
Polyamide-DNA Atoms	ImHpPyPy1 to DNA Strand 1	ImHpPyPy2 to DNA Strand 2	ImPyPyPy1 to DNA Strand 1	ImPyPyPy2 to DNA Strand 2
ThyC=Ô···HpO6	150 (145)	154 (155)	-	-
In-plane	21 (18)	28 (33)	-	-
Out-of-plane	16 (14)	18 (17)	-	-
ThyC=Ô···Am1N	142 (138)	144 (141)	147 (145)	132 (125)
In-plane	28 (21)	31 (21)	20 (6)	36 (33)
Out-of-plane	34 (43)	32 (40)	29 (36)	35 (48)

[2] Angles in parentheses measured to 'hydrogen' of donor atom.

Figure 3.4. **(a)** The hydroxypyrrole donates two hydrogen bonds to the Thy-O2.



(b) Space-filling model of the Hp/Py pair interacting with the T•A base pair shows the Hp-OH tightly fits the cleft formed by the adenine-C2H.



3.3 Conformational Changes

The sugar-phosphate backbones in the ImHpPyPy and ImPyPyPy structures superimpose with 0.75 Å rms difference. In both structures, the oligonucleotides have the standard B-DNA features of 35° twist, 3.4 Å rise per residue and C2'-endo sugar pucker, but are distinguished from ideal B-form by a strong propeller twist and opening of the target T•A base pairs. However, the Hp/Py pairs induce a change in the T•A base pairs from no shear (-0.2 average displacement between the bases in the base pair, perpendicular to the helix axis) to a large positive shear (1.2 Å, average) (Table 3.3). The movement of the bases past one another may result from the Hp-O6 contact with the adenine-C2 pressing the adenine of the target base pair back into the major groove. The increased displacement between the bases stretches the Watson-Crick hydrogen bonds between them by 0.5 Å, on average (Table 3.4). Although the specificity of hydroxypyrrole-containing polyamides is greatly increased for T•A over A•T, the affinities are slightly reduced relative to the pyrrole-counterparts. For example, ImHpPyPy-β-Dp and ImPyPyPy-β-Dp bind a 5'-AGTACT-3' site with equilibrium dissociation constants of 344 and 48 nM, respectively (Methods). The energetic penalty due to the partial 'melting' of the target T•A base pairs could account for the 1.2 kcal/mol reduction in binding affinity (Brameld *et al.*, 1997).

The change in the shear in the presence of the Hp/Py versus the Py/Py pair is more dramatic for one of the two crystallographically independent T•A base pairs than for the other (2.2 Å compared to 0.4 Å). A buffer molecule from the crystallization solution, tris-(hydroxymethyl)-aminomethane molecule (Tris), is bound in the major groove of this A•T base pair of the ImHpPyPy structure (Figure 3.5). No evidence for a corresponding buffer molecule was found in the major groove of the ImPyPyPy structure. The Tris molecule, bound in the major groove selectively in the presence of an Hp/Py pair in the minor groove, suggests that hydroxypyrrole-containing polyamides may be used as an indirect lever to manipulate interactions of proteins with the major groove.

The hydrogen bonds between the amides of each ImPyPyPy polyamide and the purine-N3 or pyrimidine-O2 of the adjacent DNA strand are maintained for the ImHpPyPy polyamide. However, the hydrogen bonds between the DNA and the ImHpPyPy amides are longer for the residues that follow the Hp than those observed for the ImPyPyPy complex (Table 3.1). The hydroxyl forms an intramolecular hydrogen bond with the following amide, causing the hydrogen bond of that amide with the adenine-N3 to become bifurcated and therefore weaker (Figure 3.6). This may be an additional source of the slightly decreased affinity of the Hp-containing polyamides relative to the pyrrole counterparts.

Figure 3.4. In the structure ImHpPyPy-DNA complex, a Tris molecule is bound in the major groove of one of the T•A base pairs. The $2|F_o|-|F_c|$ electron density is shown at 1σ contour level.

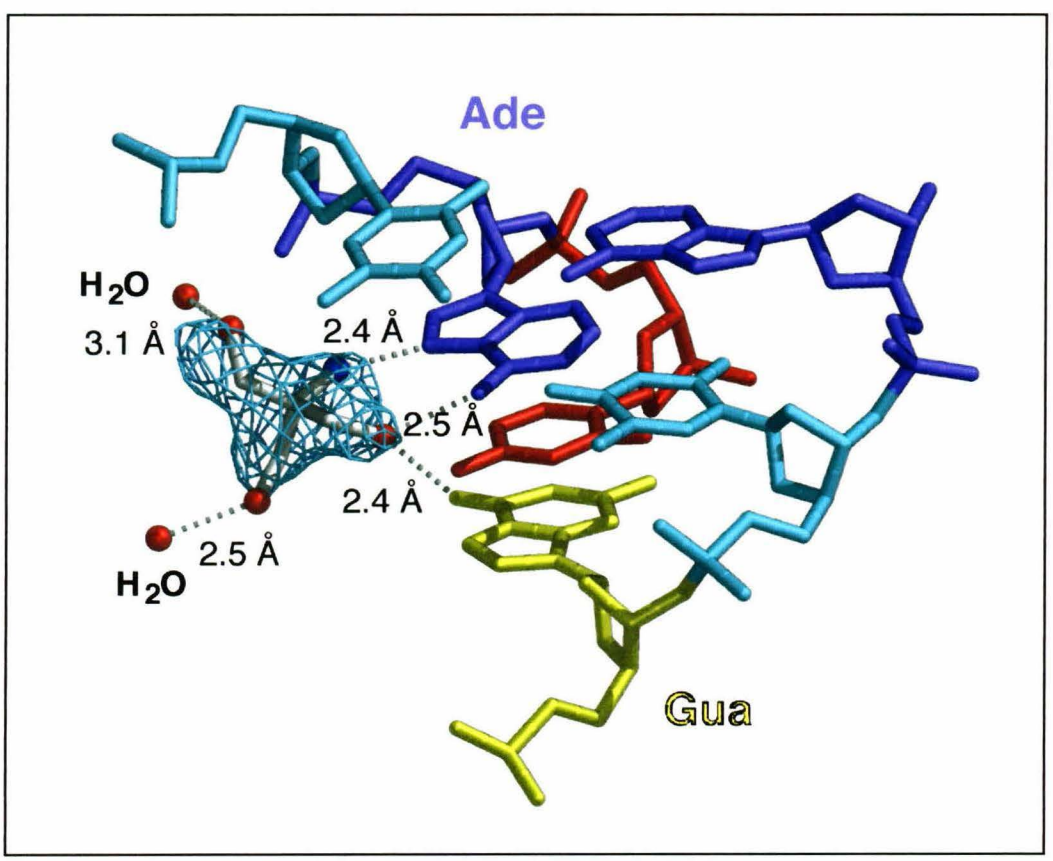


Figure 3.5. The hydrogen bonds between ImHpPyPy and one strand of DNA, indicated by dashed lines. An intramolecular hydrogen bond is formed between the hydroxyl group and the following amide.

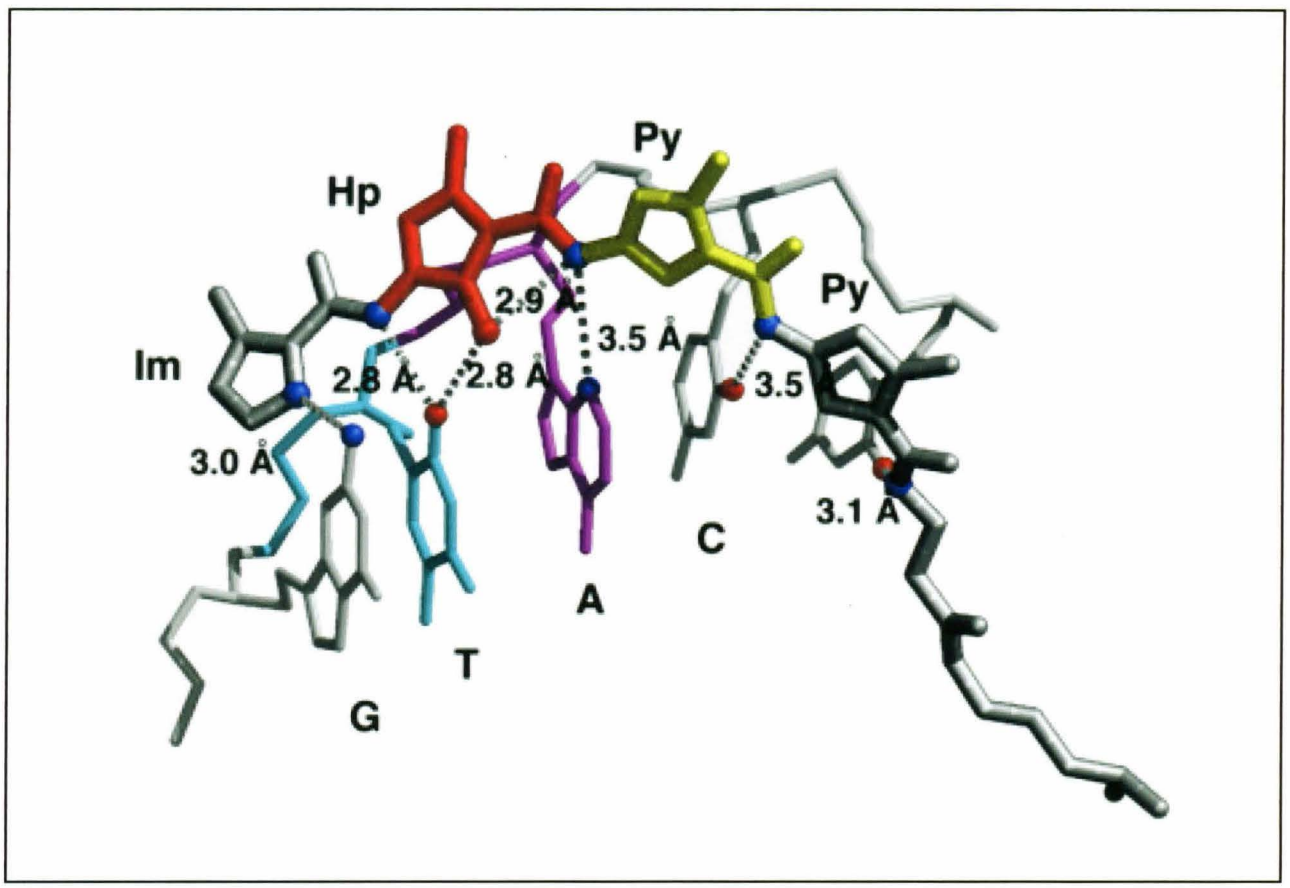


Table 3.3. DNA Helical Parameters.[3]

	Shear[4] (Å)		Propeller Twist[5] (°)		Opening[6] (°)	
	ImHpPyPy	ImPyPyPy	ImHpPyPy	ImPyPyPy	ImHpPyPy	ImPyPyPy
5' 3'						
G•C	-0.7	-0.7	-13.8	-8.8	3.6	-3.1
T•A	0.1	-0.3	-19.5	-24.5	-8.8	-14.3
A•T[7]	2.2	0.0	-19.1	-27.5	-24.9	-27.5
C•G	-0.1	-0.1	-13.6	-5.8	-5.3	-5.8
3' 5'						
B-DNA[8]	0.0		4.1		-4.0	

	ImHpPyPy	ImPyPyPy	B-DNA
Minor Groove Width (Å)	7.9	7.9	6.0
Sugar Pucker (°)	C2'-endo	C2'-endo	C2'-endo
Helical Twist (°)	35	35	36
Helical Pitch (Å)	3.4	3.3	3.4

[3] Calculated using the program Curves4 (Lavery & Sklenar, 1989).
[4] Shear is the displacement between base pairs, in plane of base pairs.
[5] Propeller twist, angle between the planes of the bases in the base pair.
[6] Opening, angle between the base pairs, in the plane of the bases, due to flexing about helix axis.
[7] The base pair of the ImHpPyPy structure with bound Tris.
[8] Ideal B-DNA generated using the program Insight.

Table 3.4. Watson-Crick Hydrogen Bond Lengths for Central GTAC.

Base Pair	ImHpPyPy (Å)	ImPyPyPy (Å)	Difference (Å)
Gua4-N2···Cyt17-O2	2.7	2.7	0.0
Gua4-N1···Cyt17-N3	2.8	2.9	0.1
Gua4-O6···Cyt17-N4	2.7	3.0	0.3
Thy5-N3···Ade16-N1	**3.5**	**3.2**	**0.3**
Thy5-O4···Ade16-N6	**3.6**	**3.0**	**0.6**
Ade6-N1···Thy15-N3	**3.3**	**2.7**	**0.6**
Ade6-N6···Thy15-O4	**3.2**	**2.6**	**0.6**
Cyt7-O2···Gua14-N2	2.6	2.7	0.1
Cyt7-N3···Gua14-N1	2.9	2.7	0.2
Cyt7-N4···Gua14-O6	3.0	2.7	0.3

3.4 Conclusion

These studies have established how a designed ligand can predictably discriminate A•T from T•A in the minor groove, utilizing the double hydrogen bond acceptor potential of the thymine-O2 and the asymmetry of the adenine-C2 cleft. The structure eliminates the possibilities that a bulky substitution at the pyrrole 3-position might 1) sterically clash with the thymine-O2 (Kopka *et al.*, 1997), or 2) cause a gross distortion of the DNA duplex (Kim *et al.*, 1995a, Kim *et al.*, 1995b). In addition, the structural basis of minor groove recognition by a synthetic molecule raises the interesting question of whether naturally occurring DNA-binding proteins may use similar principles to distinguish between the base pairs in the minor groove. [9]

---

[9] For example, certain HMG-domain proteins recognize a TT step with a conserved tyrosine that has been suggested to hydrogen bond with the thymine-O2 and pack against the adenine-C2 (Travers, 1995) and the N-terminal arm of homeodomain proteins has weak T•A specificity (Ades & Sauer, 1995) which may or may not be related to an arginine side chain that satisfies two hydrogen bonds with the thymine-O2 (Wilson *et al.*, 1995; Hirsch & Aggarwal, 1995; Tucker-Kellogg *et al.*, 1997).

3.5 Methods

The polyamides ImHpPyPy-β-Dp and ImPyPyPy-β-Dp were synthesized by solid-phase methods (Baird & Dervan, 1996) using Boc-protected 3-methoxypyrrole, imidazole and pyrrole amino acids. Identity and purity of the polyamides was confirmed by ^{1}H NMR and matrix-assisted laser-desorption ionization time-of-flight mass spectrometry. Equilibrium dissocation constants were determined by quantitative Dnase I footprint titration experiments on the DNA restriction fragment pJT8 (Trauger *et al.*, 1996) as described (White *et al.*, 1997). Synthetic deoxyoligonucleotides were synthesized with the 5'-trityl on, and purified with two rounds of reverse phase on a C8 column (Pharmacia). The ImHpPyPy and ImPyPyPy•5'-CC**AGTACT**GG-3' crystals grew under similar conditions to those of ImImPyPy•5'-CC**AGGCCT**GG-3'.

The ImHpPyPy and ImPyPyPy structures crystallized in an isomorphous lattice (Table 3.5), and were solved by molecular replacement using a B-DNA model (Navaza, 1994). The ImHpPyPy data set was collected on beamline 9-1 at the Stanford Synchrotron Radiation Laboratory (SSRL), with a MAR Research image plate detector at wavelength 0.98 Å. The ImPyPyPy data set was collected on an R-Axis IIC image plate using CuKα radiation produced by a Rigaku RU200 rotating anode generator with double focusing mirrors and a Ni filter. Both data sets were collected on flash-cooled crystals. The data was processed with DENZO/SCALEPACK (Otwinowski, 1993). Free-R sets comprising 5% of the data were chosen to contain the same reflections in resolution shells that overlapped between the data sets. All data was used, with bulk solvent correction and anisotropic B-scale applied using the program X-PLOR (Brünger, 1992), and no sigma cutoff. The polyamide β-Dp tails of both structures were modeled in alternate conformations. The planarity of the bases, aromatic rings, and peptide bonds were restrained throughout the refinement. Topology and parameter files for polyamides and Tris were generated by use of XPLO2D (Kleygwegt & Jones, 1994), and nucleic acid parameters were those of Parkinson *et al.* (1991). Figures were prepared by use of

Molscript, Bobscript and Raster3D (Bacon & Anderson, 1988; Kraulis, 1991; Merrit & Murphy, 1994; Esnouf, 1997).

Acknowledgments. We are grateful to the NIH for research support, to the NSF for a predoctoral fellowship to C.L.K., to J. Edward Richter for an undergraduate fellowship to J.M.T., and to the HHMI for a predoctoral fellowship to E.E.B. We thank J.E. Wedekind, C. L. Drennan, T. M. Iverson and M. Williamson for assistance with data collection and S. Horvath for oligonucleotide synthesis. The rotation camera facility at SSRL is supported by the DOE and NIH. Coordinates have been deposited in the Nucleic Acid Database with accession numbers BDD002 (ImHpPyPy) and BDD003 (ImPyPyPy).

Table 3.5. Data collection and refinement statistics.

Data Collection[10]		
	ImHpPyPy	ImPyPyPy
Space Group	C2	C2
Unit Cell (Å)	**a**=60.4,**b**=30.5,**c**=42.8 β=120.5°	**a**=60.7,**b**=30.2,**c**=43.5 β=123.6°
Resolution (Å)	2.2-17.0	2.1-16.0
Measured Reflections	9893	12267
Unique Reflections	3070	3417
Completeness (%)	91.5(91.0)	91.6(82.9)
I/σ(I)	28.8(3.2)	31.8(4.1)
R_{sym}[11] (%)	3.0(25.7)	3.8(17.9)
Refinement		
R_{crys}[12] (%)	21.0	21.9
R_{free} (%)	23.7	24.2
Rms deviation of bond lengths (Å)	0.016	0.012
Rms deviation of bond angles (°)	2.17	1.29
Number of Nonhydrogen Atoms		
DNA	404	404
Polyamide	96	96
Tris	8	-
Water Molecules	65	87

[10] Values in parentheses are for the highest resolution shell: 2.28-2.20 Å for ImHpPyPy, and 2.18-2.10 Å for the ImPyPyPy data sets.

[11] $R_{sym} = \sum_{hkl}\sum_{i}|I_i - \langle I\rangle|/\sum_{hkl}\sum_{i}\langle I\rangle$ where I_i is an intensity I for the ith measurement of a reflection with indices hkl and $\langle I\rangle$ is the weighted mean of all measurements of I.

[12] $R_{cryst} = \sum_{hkl}||F_{obs}(hkl)| - k|F_{calc}(hkl)||) / \sum_{hkl}|F_{obs}(hkl)|$ for the working set of reflections, R_{free} is R_{cryst} for 5% of the reflections excluded from the refinement.

3.6 References

Ades, S. E. & Sauer, R. T. (1995). Specificity of minor-groove and major-groove interactions in a homeodomain-DNA complex. *Biochemistry* **34**, 14601-14608.

Bacon, D. J. & Anderson, W. F. (1988). A fast algorithm for rendering space filling molecule pictures. *J. Mol. Graph.* **6**, 219-220.

Baker E. N. & Hubbard, R. E. (1984). Hydrogen-bonding in globular proteins. *Prog. Biophys. Molec. Biol.* **44**, 97-179.

Bailey, S. (1994). The CCP4 suite- Programs for protein crystallography. *Acta Crystallogr. D* **50**, 760-763.

Baird, E. E. & Dervan, P. B. (1996). Solid phase synthesis of polyamides containing imidazole and pyrrole amino acids. *J. Am. Chem. Soc.* **118**, 6141-6146.

Brameld, K., Dasgupta, S. & Goddard, W. A. (1997). Distance dependent hydrogen bond potentials for nucleic acid base pairs from ab initio quantum mechanical calculations (LMP2/cc-pVTZ) *J. Amer. Phys. Chem. B.* **101**, 4851-4859.

Branden, C. & Tooze, J. *Introduction to Protein Structure* (Garland Publishing, New York, 1991).

Brunger, A. T. *X-PLOR Version 3.1. A system for X-ray Crystallography and NMR* (University Press, New Haven, 1992).

Chen, X., Ramakrishnan, B., Rao, S. T. & Sundaralingam, M. (1994). Binding of two distamycin A molecules in the minor groove of an alternating B-DNA duplex. *Nature Struct. Biol.* **1**, 169-175.

Chen, X., Ramakrishnan, B. & Sundaralingam, M. (1997). Crystal structures of side-by-side binding of distamycin to AT-containing DNA octamers d(ICITACIC) and d(ICATATIC). *J. Mol. Biol.* **267**, 1157-1170.

Choo, Y. & Klug, A. (1997). Physical basis of a protein-DNA recognition code. *Curr. Opin. Struct. Biol.* **7**, 117-125.

Esnouf, R. M. (1997). An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph.* **15**, 132-134.

Geierstanger, B. H., Mrksich, M., Dervan, P. B. & Wemmer, D. E. (1994b). Design of a G•C specific DNA minor groove binding peptide. *Science.* **266**, 646-650.

Gottesfeld, J. M., Neely, L., Trauger, J. W., Baird, E. E. & Dervan, P. B. (1997). Regulation of gene expression by small molecules. *Nature.* **387**, 202-205.

Hirsch, J. A. & Aggarwal, A. K. (1995). Structure of the even-skipped homeodomain complexed to AT-rich DNA: New perspectives on homeodomain specificity. *EMBO J.* **14**, 6280-6291.

Kielkopf, C. L., Baird, E. E., Dervan, P. B. & Rees, D. C. (1998). Structural basis for G•C recognition in the DNA minor groove. *Nature Struct. Biol.* **5**, 104.

Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. (1993). Crystal structure of a yeast TBP-TATA box complex. *Nature.* **365**, 512-520.

Kim, J. L., Nikolov, D. B. & Burley, S. K. (1993). Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature.* **365**, 520-527.

Kleygwegt, G. J. & Jones, T. A. In *From first map to final model* (eds Hubbard, R. & Waller, D. A.) 59-66 (SERC Daresbury Laboratory, UK; 1994).

Kopka, M. L., Goodsell, D. S., Han, G. W., Chiu, T. K., Lown, J. W. & Dickerson, R. E. (1997). Defining GC-specificity in the minor groove: side-by-side binding of the di-imidazole lexitropsin to C-A-T-G-G-C-C-A-T-G. *Structure* **5**, 1033-1046.

Kraulis, P. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946-950.

Lavery, R. & Sklenar, H. (1988). The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biolmolec. Struct. Dynam.* **6**, 63-91.

Merritt, E. A. & Murphy, M. E. P. (1994). Raster3D version 2.0- a program for photorealistic molecular graphics. *Acta Crystallog.* **D50**, 869-873.

Mrksich, M., Wade, W. S., Dwyer, T. J., Geierstanger, B. H., Wemmer, D. E. & Dervan, P. B. (1992). Antiparallel side-by-side dimeric motif for sequence-specific recognition in the minor groove of DNA by the designed peptide 1-methylimidazole-2-carboxamide netropsin. *Proc. Natl. Acad. Sci. USA* **89**, 7586-7590.

Navazza, J. (1994). AmoRe, an automated package for molecular replacement. *Acta Crystallog.* **A50**, 157-163.

Otwinowski, Z. in *Data Collection and Processing* (eds Sawyer, L., Isaacs, N. & Bailey, S.) 56-62 (SERC Daresbury Laboratory, UK; 1993).

Parkinson, G., Vojtechovsky, J., Clowney, L., Brünger, A. T. & Berman, H. M. (1991). New parameters for the refinement of nucleic acid-containing structures. *Acta Crystallog.* **A47**, 110-119.

Pelton, J. G. & Wemmer, D. E. (1989). Structural characterization of a 2-1 distamycin-A •d(CGCAAATTTGCG)$_2$ complex by two-dimensional NMR. *Proc. Natl. Acad. Sci. USA* **86**, 5723-5727.

Seeman, N. C., Rosenburg, J. M. & Rich, A. (1976). Sequence specific recognition of double helical nucleic acids by proteins. *Proc. Natl. Acad. Sci. USA* **73**, 804-808.

Steitz, T. A. (1990). Structural studies of protein nucleic acid interaction: the source of sequence specific binding. *Quart. Rev. Biophys.* **23**, 203-250.

Trauger, J. W., Baird, E. E. & Dervan, P. B. (1996a). Recognition of DNA by designed ligands at subnanomolar concentrations. *Nature*, **382**, 559-561.

Travers, A. A. (1995). Reading the minor groove. *Nature Struct. Biol.* **2**, 615-618.

Tucker-Kellogg, L., Rould, M. A., Chambers, K. A., Ades, S. E., Sauer, R. T. & Pabo, C. O. (1997). Engrailed (Gln50->Lys) homeodomain-DNA complex at 1.9 angstrom resolution: structural basis for enhanced affinity and altered specificity. *Structure* **5**, 1047-1054.

Wade, W. S., Mrksich, M. & Dervan, P. B. (1992). Design of peptides that bind in the minor groove of DNA at 5'-(A,T)$_5$G(A,T)C(A,T)-3' sequences by a dimeric side-by-side motif. *J. Am. Chem. Soc.* **114**, 8783-8794.

Wahl, M. C. & Sundaralingam, M. S. (1997). C-H---O hydrogen bonding in biology*TIBS* **22**, 97-102.

White, S., Baird, E. E. & Dervan, P. B. (1996). Effects of the A•T/T•A degeneracy of pyrrole-imidazole polyamide recognition in the minor groove of DNA. *Biochemistry*. **35**, 12532-12537.

White, S., Baird, E. E. & Dervan, P. B. (1997a). On the pairing rules for recognition in the minor groove of DNA by pyrrole-imidazole polyamides. *Chem. Biol.* **4**, 569-577.

White, S., Baird, E. E. & Dervan, P. B. (1997b). Orientation preferences of pyrrole-imidazole polyamides in the minor groove of DNA. *J. Am. Chem. Soc.* **119**, 8756-8765.

White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E. & Dervan, P. B. (1998). Recognition of the four Watson-Crick base pairs in the DNA minor groove by synthetic ligands. *Nature*. **391**, 468-471.

Wilson, D. S., Guether, B., Desplan, C. & Kuriyan, J. (1995). High resolution crystal structure of a paired (pax) class cooperative homeodomain dimer on DNA. *Cell* **82**, 709-719.

Wong, J. M. & Bateman, E. (1994). TBP-DNA interactions in the minor groove distinguish between A-T and T-A base pairs. *Nucl. Acids Res.* **22**, 1890-1896.

Chapter 4

Structural Effects of Sequence on T•A Recognition in the DNA Minor Groove

Abstract: Synthetic polyamides composed of three types of aromatic amino acids, N-methylimidazole (Im), N-methylpyrrole (Py) and N-methyl-3-hydroxypyrrole (Hp) bind specific DNA sequences as antiparallel dimers in the minor groove. The side-by-side pairings of the residues in the dimer afford a general recognition code that allows all four base pairs to be predictably recognized. To examine the structural consequences of changing the sequence composition on T•A recognition in the minor groove, crystal structures of polyamides ImPyHpPy and the pyrrole counterpart ImPyPyPy bound to the six base pair target site 5'-AGATCT-3' in a ten base pair oligonucleotide have been determined to a resolution of 2.27 and 2.15 Å, respectively. The structures demonstrate that the principles of Hp/Py recognition of T•A are consistent between different sequence contexts. However, a general structural explanation for the slight, non-additive reduction in binding affinity due to introduction of the hydroxyl group is less clear. Comparison with other polyamide-DNA cocrystal structures reveals structural themes and differences that may relate to sequence preference.

Manuscript in preparation: Kielkopf, Bremer, White, Szewczyk, Turner, Baird, Dervan & Rees

4.1 Introduction

Polyamides are synthetic ligands composed of N-methylimidazole (Im), N-methylpyrrole (Py) and N-methyl-3-hydroxypyrrole (Hp) amino acids, that target specific DNA sequences with affinities and specificities that can rival those of DNA binding proteins (Trauger *et al.*, 1996a, 1998a; Swalley *et al.*, 1997; Turner *et al.*, 1997; Herman *et al.*, 1999). Im-Py polyamides have been shown to be cell-permeable, and to inhibit transcription of designated genes in living cells (Gottesfeld *et al.*, 1997; Dickinson *et al.*, 1998). Surprisingly, polyamides predictably distinguish all four base pairs in the minor groove, despite the symmetric distribution of functional groups displayed on the base pair edges (Seeman *et al.*, 1976). These molecules bind as antiparallel dimers and read the DNA sequence using the side-by-side pairings of the aromatic residues. An Im/Py pair discriminates a G•C from C•G, and both from A•T/T•A base pairs (Wade *et al.*, 1992; Wade *et al.*, 1993; Mrksich & Dervan, 1993; White *et al.*, 1997b). A Py/Py pair is partially degenerate for T•A and A•T base pairs (Pelton & Wemmer, 1989, 1990; White *et al.*, 1996), whereas an Hp/Py pair distinguishes T•A from A•T and G•C/C•G base pairs (White *et al.*, 1998, 1999). In practice, the molecules in the dimer may be covalently linked in a hairpin turn using the aliphatic residue γ-amino-butyric acid (γ-turn), thereby improving specificity and affinity (Mrksich *et al.*, 1994). In addition, solid phase synthesis of polyamides (Baird & Dervan, 1996) is facilitated by a C-terminal β-alanine residue (β) preceding a dimethylpropylamine (Dp). This β-Dp tail enhances both DNA-binding affinity and specificity (Parks *et al.*, 1996) and displays a strong preference for flanking A•T/T•A base pairs (Swalley *et al.*, 1999). The structural basis of these pairing rules has been established by NMR (Pelton & Wemmer, 1989, 1990; Mrksich *et al.*, 1992; Geierstanger *et al.*, 1994a, b; de Clairac *et al.*, 1997, 1999) and crystallographic analysis (Chen *et al.*, 1994, 1997; Kielkopf *et al.*, 1998a, b; Mitra *et al.*, 1999). An Im/Py targets G•C via a directional hydrogen bond from the Im-N3 to the free guanine-N2H (Geierstanger *et al.*, 1994b; Kielkopf *et al.*, 1998a), and an Hp/Py pair recognizes T•A using a combination of specific hydrogen bonds between the hydroxyl with the thymine-O2 and shape selective recognition of the bulkier adenine ring (Kielkopf *et al.*, 1998b).

The generality of the pairing rules has been thoroughly established by targeting a wide variety of DNA sequences. Detailed characterization of the polyamide-DNA sequence preferences has identified certain classes of Im-Py-Hp/DNA base pair combinations that are more difficult to distinguish, including longer sequences, 5'-GN*G*-3' and 5'-G*A*-3' steps (problematic nucleotide italicized). The upper limit in binding site size for Im-Py polyamides before an energetic penalty occurs has been determined to be seven base pairs, or five paired aromatic residues (Kelly *et al.*, 1996). This binding site size limit for polyamides composed of aromatic amino acids has been explained subsequently as resulting from a mismatch between the curvature of the polyamide and the twist of the DNA (Kielkopf *et al.*, 1998a). Flexible β-alanine residues introduced internally, as A•T/T•A specific β/β and Py/β pairings or Im/β pairings that target G•C, reset the register of the polyamide to that of the DNA (Trauger *et al.*, 1996b; Turner *et al.*, 1998; de Clairac *et al.*, 1999). Using a combination of these motifs, sequences at least sixteen base pairs in length can be targeted with high affinity (Trauger *et al.*, 1998b). Furthermore, inserting a β-alanine preceding an Im residue allows 5'-GN*G*-3' sequences to be targeted with greatly increased affinity and specificity relative to the Py counterpart (Turner *et al.*, 1998). Therefore, even in difficult cases polyamide motifs can be fine-tuned to provide minor structural adjustments that restore predictable specificity.

Despite the similarities between the functional groups displayed by adenine and thymine in the minor groove, an Hp/Py pair was successfully designed to distinguish T•A from A•T base pairs (White *et al.*, 1998a, b; Kielkopf *et al.*, 1998b). Although relatively degenerate for T•A and A•T base pairs, Py/Py pairs also display an unexplained 2-12-fold preference to bind 5'-G*T*-3' steps over 5'-G*A*-3' (White *et al.*, 1996). This inherent sequence preference is exhibited by the Hp as well, so that equivalent Im-Py-Hp polyamides that vary only in the placement of the Hp residue are consistently several fold more specific when the desired target sequence is 5'-G*T*-3' rather than 5'-G*A*-3' (White *et al.*, 1998a, b). In addition to this fine point of sequence preference, another secondary characteristic of the Hp-residue is that the hydroxyl group energetically destabilizes DNA-binding affinity by a few-fold relative to the Py/Py counterpart (White *et al.*, 1998a, 1999). Fortunately, this effect is not additive, so that multiple Hp residues can be

introduced without significant reduction in the equilibrium association constant for the target site (White *et al.*, 1999). X-ray structural analysis of an Hp-containing polyamide suggests that elongation of the Watson-Crick hydrogen bonds of the target T•A base pairs or bifurcation of the amide-DNA hydrogen bond following the Hp residue by the hydroxyl group may contribute to the reduction in binding affinity (Kielkopf *et al.*, 1998b). However, the generality of these structural observations for different TA sequences and their relative contribution to the overall binding energetics have not been investigated.

Previously, three structures of unlinked, four-ring polyamide dimers bound to their designated target sites in ten base pair oligonucleotides have been determined by crystallographic methods. These include the GC-specific polyamide ImImPyPyβDp (Dp, dimethyl-propylamine) bound to 5'-CC**AGGCCT**GG-3' (predicted target site in bold) (Kielkopf *et al.*, 1998a), the TA-specific polyamide ImHpPyPyβDp and its parent compound for comparison, ImPyPyPyβDp, both bound to the oligonucleotide 5'-CC**AG*TA*CT**GG-3' (Kielkopf *et al.*, 1998b). To examine the structural consequences of reversing the TA sequences on polyamide-DNA interactions, the crystal structures of ImPyHpPyβDp•5'-CC**AG*AT*CT**GG-3' and its pyrrole-counterpart 5'-CC**AG*AT*CT**GG-3' are presented here. In particular, we wish to verify the generality of the structural basis of Hp/Py recognition of TA base pairs, especially given the relative difficulty of targeting 5'-G*A*-3' steps. Furthermore, by comparison of the Im-Py-Hp polyamide-DNA structures, it may be possible to find a structural explanation for the non-linear energetic destabilization of DNA binding due to introduction of the hydroxyl group and the preference for 5'-G*T*-3' steps over 5'-G*A*-3'. A detailed understanding of polyamide-DNA interactions will contribute to further development of these useful small molecules as medicinal and chemical tools, for which their ability to target specific genes is essential.

4.2 Overall Structure

The overall conformation of ImPyHpPyβDp and ImPyPyPyβDp•5'-CC**AGATCT**GG-3' are similar to those of other four-ring polyamide dimers bound to DNA (Figure 4.1). In both structures, the polyamides bind as an antiparallel dimer in the minor groove of B-form DNA, with the N- to C-terminal orientation of the polyamide parallel to the 5'- to 3'- orientation of the adjacent DNA strand. This is consistent with affinity cleaving experiments (White *et al.*, 1997), and previous structures of polyamide dimer-DNA complexes (Pelton & Wemmer, 1989, 1990; Mrksich *et al.*, 1992; Geierstanger *et al.*, 1994a, b; Chen *et al.*, 1994, 1997; Kielkopf *et al.*, 1998a, b; Mitra *et al.*, 1999; de Clairac *et al.*, 1999). In the ImImPyPyβDp•5'-CC**AGGCCT**GG-3' complex, it was noticed that some of the aromatic rings, in particular the center of the leading imidazole and 3-position of the third pyrrole, were very close to the O4' of the deoxyribose (Kielkopf *et al.*, 1998a). This effect is also found at the third residue position in the ImPyHpPyβDp structure (~2.6 Å O4'---C3 distance), where it may be enhanced by the electron withdrawing effect of the hydroxyl group. In all of the structures, the leading Im interacts most closely with a deoxyribose-O4' and sets the register for the residues that follow.

The polyamides in the dimer are fully overlapped, although a binding mode with the terminal residues unpaired would be possible given that the polyamides are not linked by a γ-turn. Although such slipped binding motifs are found in structures of distamycin and Im-Im homodimers bound to DNA, previously determined structures of unlinked Im-Py and Im-Py-Hp polyamides overlap to recognize the target site according to the pairing rules (Geierstanger *et al.*, 1994b ; Kielkopf *et al.*, 1998a, b; de Clairac *et al.*, 1999). Accordingly, the polyamides in these structures fully overlap to target the 5'-AGATCT-3' sequence, forming the predicted Im/Py, Hp/Py, Py/Hp and Py/Im pairs in the ImPyHpPy structure, and Im/Py, Py/Py, Py/Py and Py/Im pairs in the ImPyPyPy structure.

Although the residues match to complement the base pairs, the aromatic ring of one residue does not stack against the ring of its partner. To analyze the microscopic register between the polyamides in the dimer, the intersection between the center of one aromatic ring and

the normal vector to the least squares minimized plane of its paired aromatic residue was calculated (Figure 4.2). Among the different structures, these points cluster along the bond between the aromatic ring and the carboxamide carbon, and vary from ImPyPyPyβDp•5'-CC**AGTACT**GG-3' where the aromatic rings are slightly offset but close to one another, to ImImPyPyβDp•5'-CC**AGGCCT**GG-3', in which the aromatic rings stack opposite the carboxamide carbons. The stacking between the polyamides of both the ImPyHpPyβDp and ImPyPyPyβDp•5'-CC**AG*A*TCT**GG-3' structures is closer to ring-on-carboxamide, whereas the ImHpPyPyβDp•5'-CC**AG*T*ACT**GG-3' structure is closer to ring-on-ring. Thus, the precise register between the molecules in the dimer may be a function of both the polyamide and the DNA sequence. Polyamides containing mostly Py/Py pairs are closer to ring-on-ring, those with Im/Py pairs ring-on-carboxamide, Hp/Py in-between; those targeting alternating purine-pyrimidine DNA sequences tend toward ring-on-ring stacking versus purine-purine steps ring-on-carboxamide. The combination of these preferences would give rise to the observed distribution of registers. Nevertheless, these guidelines, drawn from a few experimental points, are clearly violated by the structures of distamycin dimers bound to DNA (Chen *et al.*, 1994, 1997; Mitra *et al.*, 1999). Although composed of Py/Py pairs bound to an alternating purine-pyrimidine tract, the interaction between the distamycin molecules is clearly ring-on-carboxamide, and may be modulated by the presence of a leading formyl group.

Pairwise comparison of the DNA conformation between the different structures reveals the same trends in the r.m.s. distances among the DNA atoms as observed for the variations in register; the structures of the polyamides bound to 5'-AGATC-3' are more similar to those bound to 5'-AGGCCT-3' and to one another than to the 5'-AGTACT-3' complexes. The ImPyHpPyβDp and ImPyPyPyβDp•5'-CC**AGATCT**GG-3' DNA atoms superimpose with an r.m.s. distance of 0.5 Å, and are as similar to the coordinates for ImImPyPyβDp•5'-CC**AGGCCT**GG-3' (Kielkopf *et al.*, 1998a) as to one another (ImPyHpPy-ImImPyPy r.m.s. 0.5 Å; ImPyPyPy•GATC-ImImPyPy r.m.s. 0.3 Å). Surprisingly, superposition with the AT-rich counterparts ImHpPyPyβDp and ImPyPyPyβDp•5'-CC**AGTACT**GG-3' (Kielkopf *et al.*, 1998b) is a poorer match (ImPyHpPy-ImHpPyPy r.m.s. 0.9 Å; ImPyHpPy-ImPyPyPy•G*T*AC

r.m.s. 1.0 Å; ImPyPyPy•G*A*TC-ImPyPyPy•G*T*AC r.m.s. 1.0 Å; ImPyPyPy•G*A*TC-ImHpPyPy r.m.s. 0.9 Å). In comparison, the 5'-AGATCT-3' binding site of the of ImPyHpPyβDp and ImPyPyPyβDp structures superimposes nearly as well with this sequence in the structure of the DNA dodecamer 5'-CGT**AGATCT**ACG-3' (Leonard & Hunter, 1993), despite the absence of the polyamide, different flanking sequences and oligonucleotide length (ImPyHpPy-AGATCT r.m.s. 1.0 Å; ImPyPyPy-AGATCT r.m.s. 1.2 Å).

Figure 4.1. **(a)** Structure of $(ImPyHpPy)_2$•5'-CC**AGATCT**GG-3'. Adenosine is purple and thymidine cyan; polyamide is colored as above. **(b)** Overall structure of $(ImPyPyPy)_2$•5'-CC**AGATCT**GG-3'. Schematic models of the bound sites are shown to one side.

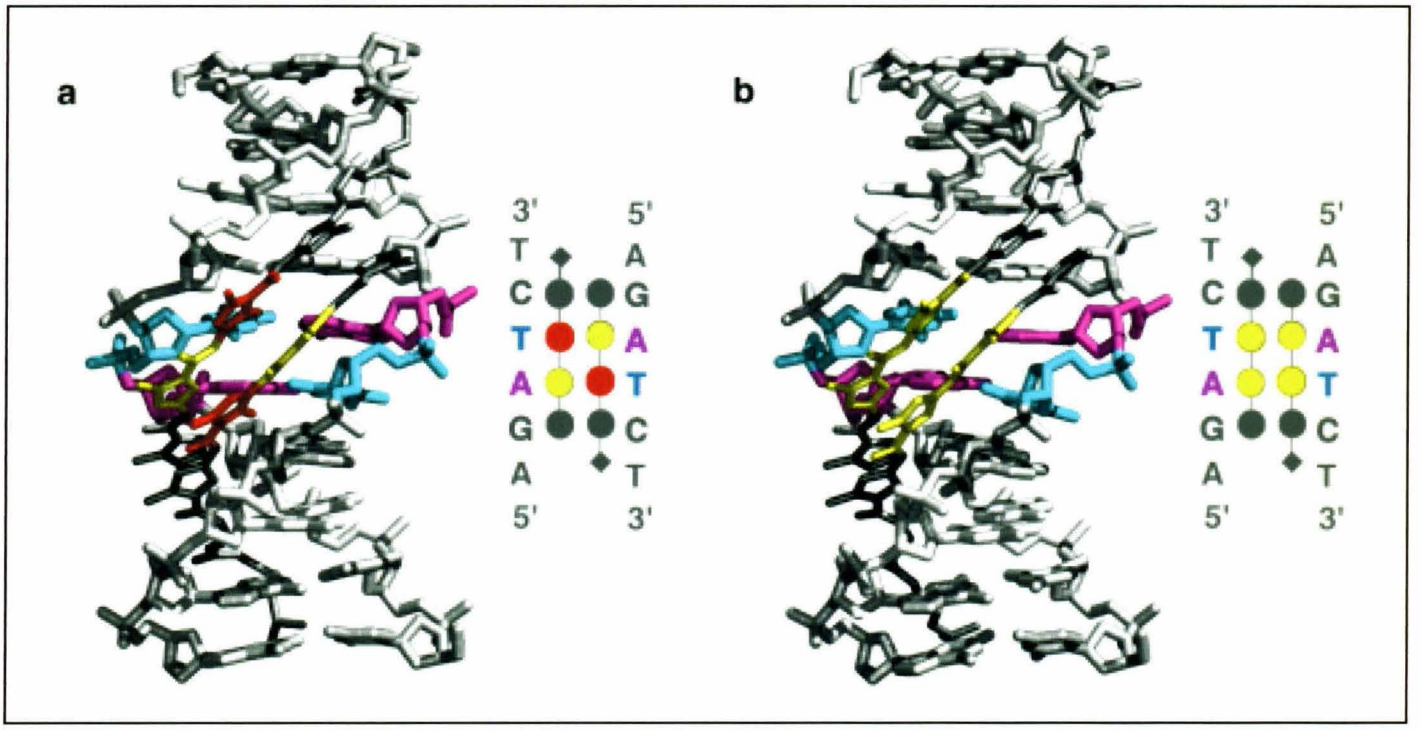
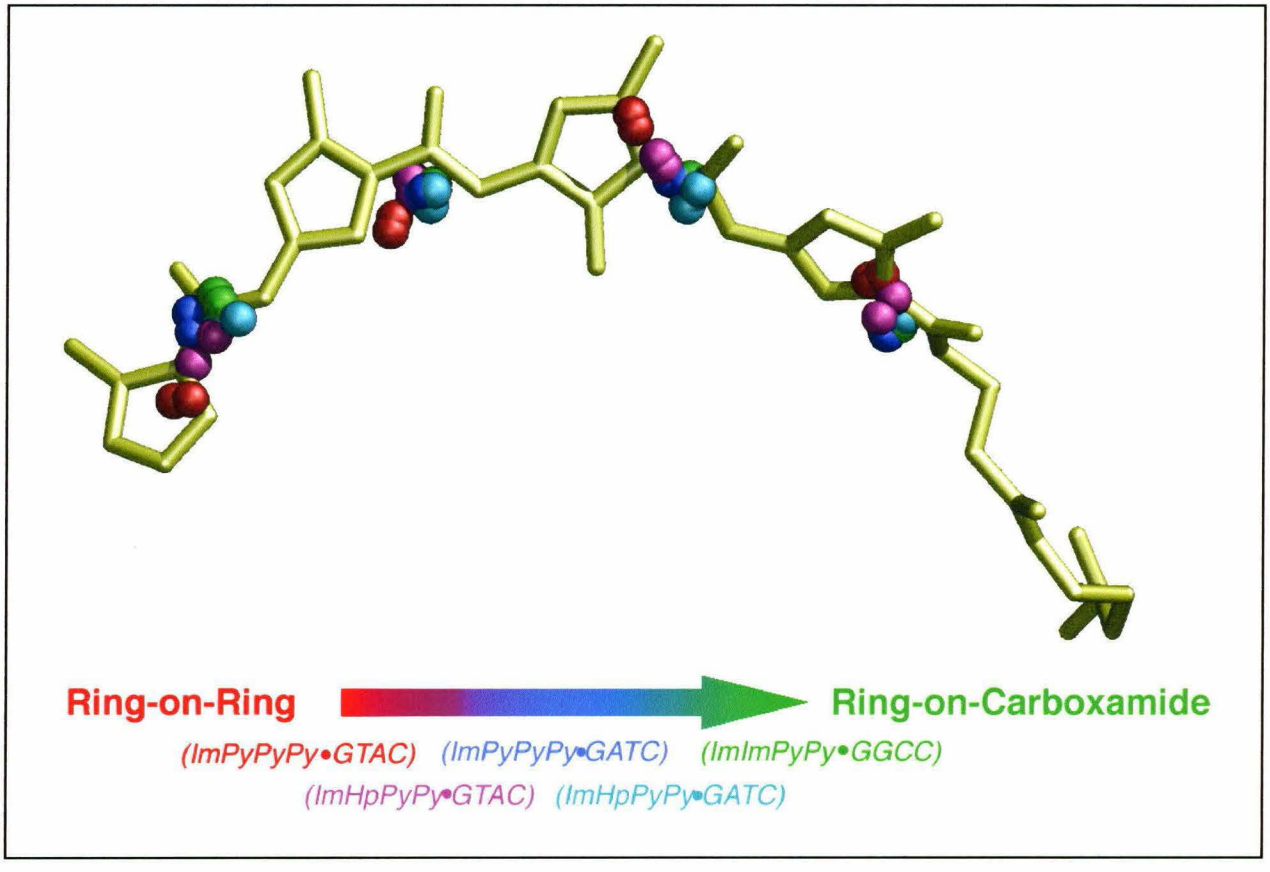


Figure 4.2. Sequence-dependent register differences between the different polyamide dimer-DNA structures. Each polyamide from the five different polyamide dimer-DNA cocrystal structures were superimposed upon one of the ImPyHpPy polyamides, which for simplicity is shown. The spheres indicate the intersection of the normal from the center of the paired aromatic ring to the least squares minimized plane of the residue shown, and represent the stacking between the paired residues. These are colored according to the structure they are derived from, and shaded from red for ring-on-ring stacking, via blue to green for ring-on-amide: Red, ImPyPyPy•GTAC; purple, ImHpPyPy•GTAC; blue, ImPyPyPy•GATC; cyan, ImPyHpPy•GTAC and green, the ImImPyPy•GGCC dimer.



4.3 Comparison of DNA Helical Parameters

For the ImPyHpPyβDp structure and its pyrrole counterpart, the twist, rise per residue and sugar pucker of the bound DNA remains B-form (Table 4.1). The ImPyPyPyβDp and ImImPyPyβDp polyamides reverse the buckle of the unliganded DNA, whereas by comparison the Hp of ImPyHpPyβDp and ImPyPyPyβDp simply flattens the base pairs. In addition, the Hp residue consistently induces a positive displacement between the target TA bases rather than the small negative shear observed both in the presence and absence of other polyamide residues. All of the polyamide dimers consistently increase the opening of the target base pairs toward the major groove, and induce a large negative propeller twist, regardless of the sequence recognized. The presence of the polyamide dimer widens the minor groove to ≥7.0 Å from the ~4 Å minor groove width of AT rich sequences. This difference (~3.0 Å) is much more significant than the difference when compared with the inherently wider minor groove of GC rich sequences (1.0 Å). The high propeller twisting of the base pairs in light of the widened minor groove is surprising, since propeller twist often accompanies a narrow minor groove in B-DNA structures (Yuan *et al.*, 1992; Leonard & Hunter, 1993). This discrepancy supports the idea that many factors interact to determine DNA conformation, so that strict relationships between helical parameters are not often observed between different sequences.

Likewise, from analysis of several B-DNA structures, it was concluded that the minor groove width of G-A steps tends to be narrower than that of G-T steps (Yanagi *et al.*, 1991). Since this sequence-dependent variation in groove width is a possible explanation for the binding preferences of polyamide dimers (White *et al.*, 1996), the minor groove widths of the polyamide binding sites were compared (Table 4.2). Indeed, there appears to be a difference between the groove widths of the GATC and the GTAC sequences in the presence of both the Hp-containing polyamides and the Py-parent compounds, with the bound GTAC site ~0.5 Å wider on average. A difference is also observed between the structures of the unliganded DNA sites (Leonard & Hunter, 1993; Ding *et al.*, in preparation). Since this effect is observed between all of the polyamide complexes, it is unlikely to result from variations in crystal packing. However, given

the interaction between the O4' of the deoxyribose and the polyamide aromatic rings, it is uncertain whether the small differences contribute unfavorably to the overall binding affinity for GATC versus GTAC sequences.

Table 4.1. DNA helical parameters.[1]

	GATC, no ligand[2]	ImPyHpPy	ImPyPyPy
Shear (Å) A•T	-0.2	0.5	-0.2
T•A	0.0	0.1	0.0
Buckle (°)A•T	11.1	1.0	-10.8
T•A	4.4	0.2	1.6
Propeller (°)A•T	-21.3	-31.1	-29.8
T•A	-24.8	-10.7	-31.4
Opening (°)A•T	-4.3	-7.0	-12.6
T•A	-3.2	-14.0	-9.4
Sugar pucker	C2' endo	C2' endo	C2' endo
Twist (°)	36.2	38.9	36.9
Rise (Å)	3.5	3.4	3.2

[1] Calculated using the program Curves4 (Lavery & Sklenar).

[2] Leonard & Hunter, 1993.

Table 4.2. Minor Groove Width[3] (Å).

Base Pair:	A•T	G•C	T•A	A•T	C•G	T•A
ImPyPyPy•GTAC	7.2	7.9	7.7	7.8	8.4	7.9
ImHpPyPy•GTAC	7.7	7.7	7.4	7.4	8.7	8.5
Base Pair:	A•T	G•C	A•T	T•A	C•G	T•A
ImPyPyPy•GATC	8.2	7.6	6.8	7.2	8.0	8.2
ImPyHpPy•GATC	7.5	7.2	6.7	7.2	7.9	8.3
Average Difference, polyamide•GTAC-polyamide•GATC	-0.4	0.4	0.7	0.4	0.6	-0.1
Base Pair:	A•T	G•C	T•A	A•T	C•G	T•A
GTAC, DNA[4]	11.5	8.0	5.5	3.8	6.4	9.5
Base Pair:	A•T	G•C	A•T	T•A	C•G	T•A
GATC, DNA[5]	8.5	7.5	4.8	3.2	5.2	8.3
Average Difference, GTAC DNA-GATC DNA	3.0	0.5	0.7	0.6	1.2	1.2

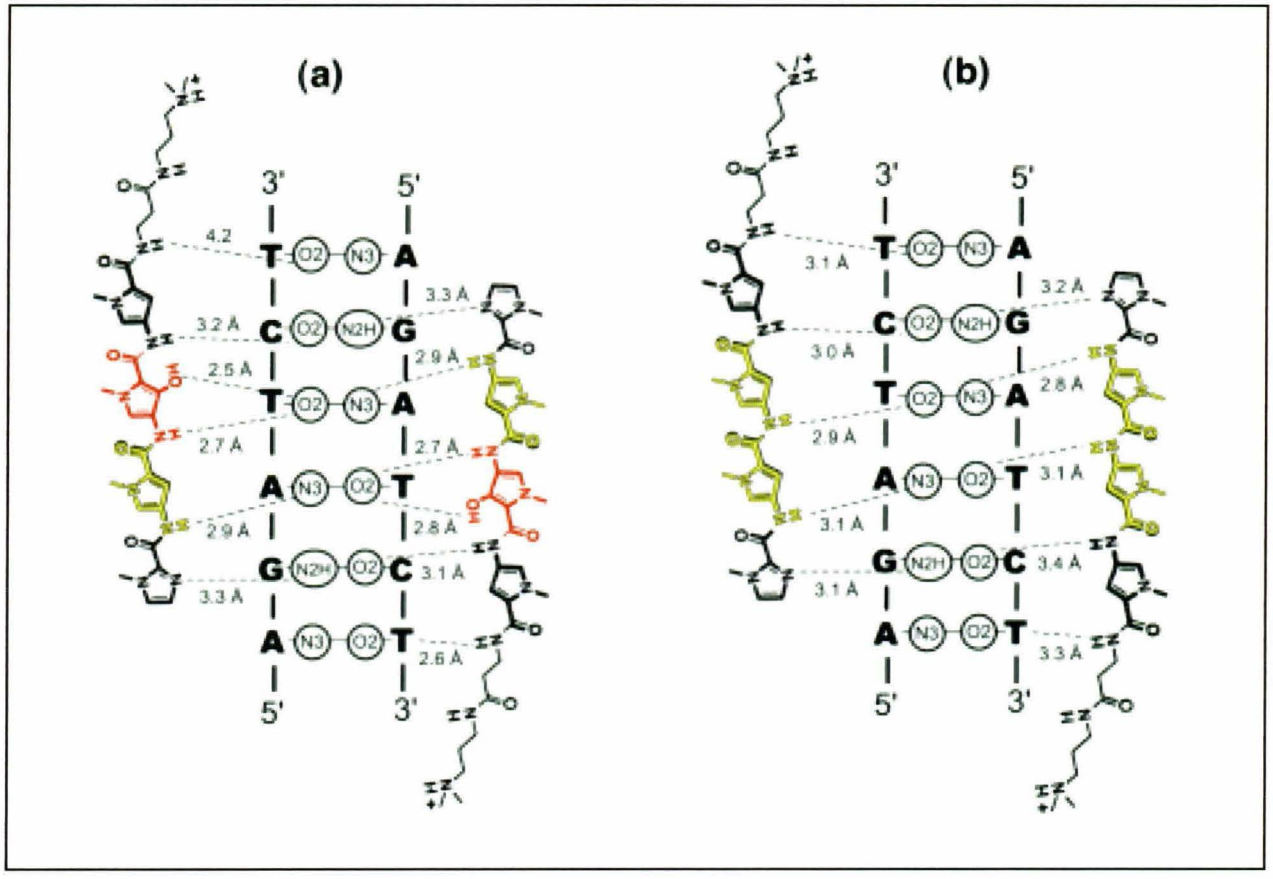
[3] Average of minor groove width for indicated base step, calculated using the program Curves4 (Lavery & Sklenar).
[4] Ding *et al.* in preparation.
[5] Leonard & Hunter, 1993.

4.4 Polyamide-DNA Hydrogen Bonds

The N3 and O2 atoms of the bases form hydrogen bonds with the amide nitrogens of the ImPyHpPyβDp and ImPyPyPyβDp polyamides (Figure 4.3). Reversing the TA sequence of the binding site does not affect these hydrogen bonds, nor the additional specific hydrogen bond between the hydroxyl and the thymine. The ImPyHpPyβDp polyamide recognizes the thymine-O2 in an identical manner to ImHpPyPyβDp. Between the amide nitrogen and the hydroxyl group of the Hp, both the lone pairs of the thymine-O2 accept hydrogen bonds from the polyamide, whereas the single lone pair of adenine-N3 can accept only one. Secondly, the hydroxyl oxygen contacts the C2 carbon of the bulky adenine ring, and would overlap the carbon by ~2 Å if the target thymine is modeled as a mismatch adenine. Therefore, the slight preference for G-T over G-A steps cannot be due to direct perturbation of the interactions between the Hp residue and target T•A pairs.

As was observed in the ImHpPyPyβDp structure, the hydroxyl is positioned to form an intramolecular hydrogen bond with the amide of the following residue (2.9 Å distance). This forces the amide to donate two hydrogen bonds, one with the DNA and the other to the hydroxyl oxygen, thereby weakening the interaction with the DNA. Accordingly, in both Hp-containing structures, this amide-DNA hydrogen bond is slightly elongated (3.2-3.8 Å).

Figure 4.3. Chemical structure of the polyamides and schematic diagram of hydrogen bonding interactions with the 5'-AGATCT-3' binding site, for **(a)** ImHpPyPy and **(b)** ImPyPyPy. Hydrogen bonds are shown by dashed lines, and numbers indicate distances between non-hydrogen atoms.



4.4 Intra-duplex Hydrogen Bonds

In the structure of the Hp-containing polyamide ImHpPyPyβDp bound to 5'-AGTACT-3', the Watson Crick hydrogen bonds of the target TA base pairs were lengthened by ~0.5 Å in the presence of the Hp/Py pairs. However, ImPyHpPyβDp recognizes 5'-GATC-3' without perturbing these base pair hydrogen bonds relative to the pyrrole-counterpart (Table 4.3). Instead of resulting directly from the presence of the Hp residue, the lengthened hydrogen bonds may have been a secondary effect of a buffer molecule, tris-(hydroxymethyl)-aminomethane (Tris), from the crystallization conditions bound in the major groove of the ImHpPyPyβDp structure (Figure 4.4). Although such a buffer molecule has not been observed in any of the other polyamide-DNA complex structures, the structure of the oligonucleotide 5'-CCAGTACTGG-3', determined in the absence of polyamide, shows a hydrated calcium molecule interacting via bound water molecules to the same site (Ding *et al.*, in preparation). Therefore, the functional groups displayed on the edges of the bases may serve to bind small molecules in the major groove of the 5'-GT-3' step.

Cross-strand hydrogen bonds between a member of one base pair and another in the following step are conceptually possible whenever a donor amine is opposed diagonally with an acceptor oxygen across the DNA groove (Figure 4.5). In the major groove, these intrastrand hydrogen bonds can occur for 5'-C-C-3', 5'-C-A-3', 5'-A-A-3' and 5'-A-C-3' sequences (and complementary strands 5'-G-G-3', 5'-G-T-3', 5'-T-T-3' and 5'-T-G-3'). Although not required for cross-strand hydrogen bond formation, base pairs with a large propeller twist, such as observed in the polyamide-DNA structures, are usually engaged in these three-center hydrogen bonds (Heinemann & Alings, 1989, Yanagi *et al.*, 1991).

The binding sites of the polyamide dimer-DNA structures were analyzed for these three-center hydrogen bonds using the criteria of Heinemann and Alings (1989); the three angles about the hydrogen position should be ~90° or greater, and the sum of the angles ~360° to indicate coplanarity. The ImImPyPyβDp•GGCC complex, at the 5'-G-G-3' and 5'-C-C-3' steps, and ImHpPyPyβDp or ImPyPyPyβDp•GTAC at 5'-G-T-3' and 5'-A-C-3', are theoretically capable

of forming cross-strand hydrogen bonds. Indeed, these anticipated interactions are observed, although borderline in some cases where the N-H---O angle is 80-87° (Table 4.5). Interestingly, a similar close approach between equivalent atoms is observed even in the absence of an appropriate hydrogen bond donor, possibly induced by the propeller twisting of the base pairs in the presence of the polyamides. Thus, for GATC sequences, the Gua-O6 and Thy-O4 are forced into a close contact that would presumably be uncomfortable due to the electronegative quality of both atoms. Accordingly, it is these sequences that are unable to form the intraduplex hydrogen bonds, 5'-GCG-3' and 5'-GAT-3', that are more problematic to target with high affinity. For example, the compound ImPyImPy-γ-ImPyImPyβDp has 100-fold lower affinity for 5'-GCGC-3' than does the analogous polyamide ImImPyPy-γ-ImImPyPyβDp for its respective match site 5'-WGGCCW-3' (Swalley *et al.*, 1997). Likewise, ImPyPy-γ-ImHpPyβDp displays 100-fold lower affinity for its match site 5'-WGATW-3' than does ImHpPy-γ-ImPyPyβDp for 5'-WGTAW-3' (White, *et al.*, 1999). Therefore, a correlation exists between the ability of some sequence steps to form three-center hydrogen bonds in the presence of highly propeller twisted base pairs, and the DNA binding affinity of the corresponding polyamide.

Table 4.3. Watson-Crick Hydrogen Bond Lengths for Central AT Base Pairs (Å).

Base Pair	ImPyHpPy	ImPyPyPy	Difference
Ade5-N1···Thy16-N3	3.0	3.2	-0.2
Ade5-N6···Thy16-O4	2.9	2.9	0.0
Thy6-N3···Ade15-N1	3.1	3.3	-0.2
Thy6-O4···Ade15-N6	2.9	3.1	-0.2

Figure 4.4. Structure of representative TA base pairs from the ImPyHpPy (left) and ImHpPyPy (right) structures. There is a Tris molecule bound in the major groove of the ImHpPyPy structure. Watson-Crick hydrogen bond lengths for the base pair shown are indicated.

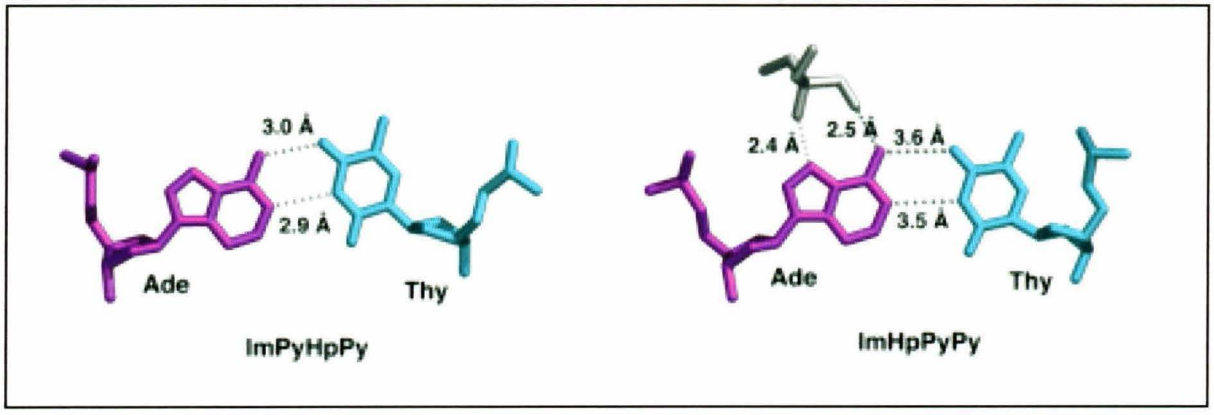


Figure 4.5. Schematic diagram of a three-center hydrogen bond in the DNA major groove. '*N3*' represents either Ade-N6 or Cyt-N4, '*O2*' and '*O1*' either Thy-O4 or Gua-O6. Numbering is consistent with Yanagi *et al.*

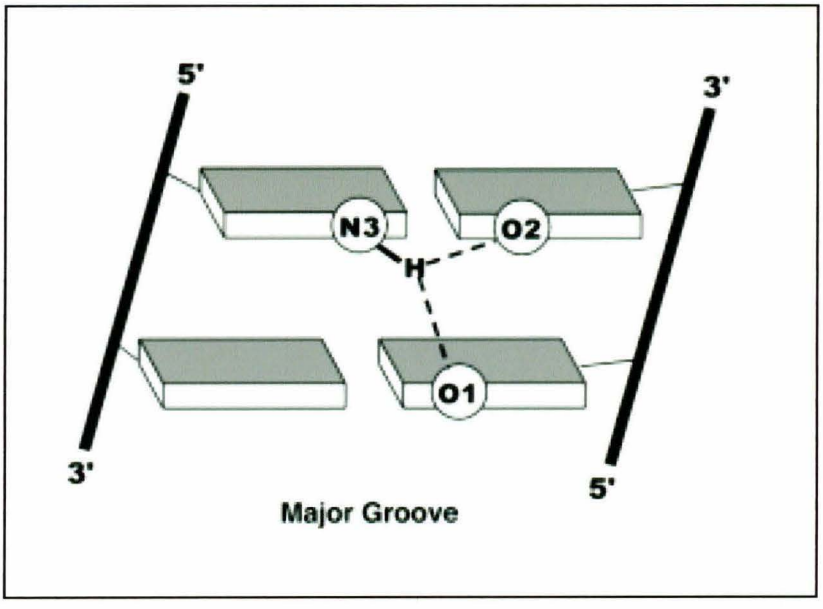


Table 4.4. Comparison of three-center interactions in the major groove. Distances R_{xy} are the separation between the atoms x and y, as defined in Figure 5, where atoms 1 and 2 are either Thy-O4 or Gua-O6, and atom 3 is either Ade-N6 or Cyt-N4. The angle A_{xy} is measured between x, hydrogen H, and y. Hydrogens were assigned using CNS (Brünger *et al.*, 1998). For structures containing more than one DNA strand in the crystallographic asymmetric unit, values are listed for both strands of the palindromic DNA site, with the second italicized and indicated by a prime.

		Distance (Å)		Angle (°)			
Structure	Base Step	R_{13}	R_{23}	A_{12}	A_{13}	A_{23}	Sum
ImPyPyPy•GATC	Gua4-Ade5	2.9	2.9	-	-	-	-
ImPyPyPy•GATC'	*Gua14-Ade15'*	2.7	2.7	-	-	-	-
ImPyHpPy•GATC	Gua4-Ade5	2.6	3.0	-	-	-	-
ImPyHpPy•GATC'	*Gua14-Ade15'*	2.9	2.9	-	-	-	-
GATC	Gua4-Ade5	3.4	2.6	-	-	-	-
GATC'	*Gua14-Ade15'*	3.1	2.7	-	-	-	-
ImPyPyPy•GTAC	Gua4-Thy5	3.1	2.6	128	91	157	376
ImPyPyPy•GTAC'	*Gua14-Thy15'*	3.3	3.0	114	95	167	376
ImHpPyPy•GTAC	Gua4-Thy5	2.8	3.6	143	87	150	370
ImHpPyPy•GTAC'	*Gua14-Thy15'*	3.2	3.2	112	80	166	358
GTAC	Gua4-Thy5	3.1	2.9	117	84	158	359
ImImPyPy•GGCC	Gua4-Gua5	2.7	2.8	133	85	157	375
ImImPyPy•GGCC'	*Gua14-Gua15'*	2.8	2.9	126	91	165	382
GGCC	Gua4-Gua5	3.1	2.9	105	87	168	360
GCGC	Gua4-Cyt5	3.1	2.9	-	-	-	-

4.6 Conclusions

The structures of ImPyHpPyβDp and ImPyPyPyβDp bound to 5'-CC**AG*AT*CT**GG-3' were undertaken to achieve a better understanding of TA recognition and the detailed affinities of polyamides for different TA sequences. From these structures, T•A recognition by Hp/Py follows two principles that are general in different sequence contexts; formation of specific hydrogen bonds with the thymine-O2 and shape selective recognition of the adenine-C2. The structural basis of the non-additive, ~2-10 fold reduction in binding affinity upon addition of the hydroxyl group is more difficult to address. Originally, it was calculated that the partial melting of the target TA hydrogen bonds in the structure of ImHpPyPyβDp bound to DNA would produce sufficient energetic penalty to account for the difference in affinity. However, the elongation of the TA hydrogen bonds is not observed in the ImPyHpPyβDp•5'-AGATCT-3' structure, so this cannot be a general mechanism.

Although perhaps inadequate to account for the energetic difference between Hp-polyamides and their Py- counterparts, lengthening of the amide-DNA hydrogen bond following the hydroxyl group is observed in both Hp-containing structures. One attractive feature of this as an explanation for the reduction in affinity is that it is not necessarily additive. Since optimization of the polyamide-DNA hydrogen bonds becomes a compromise as the number of aromatic residues is increased beyond five (Kelley *et al.*, 1996; de Clairac *et al.*, 1999), interfering with the amide-DNA hydrogen bond following the Hp residue would have less of a consequence for longer polyamides. This fits the observation that addition of multiple Hp residues in a ten-ring hairpin polyamide does not further reduce the binding constant (White *et al.*, 1999). Nevertheless, the structural basis of such a small energetic difference is difficult to unambiguously pinpoint without higher resolution structures coupled with further thermodynamic studies.

The basis of the higher affinity and specificity of Im-Py-Hp polyamides for 5'-G*T*-3' versus 5'-G*A*-3' and 5'-G*GC*-3' over 5'-G*CG*-3' sequences is also difficult to ascribe to a single type of interaction. Comparison of the different polyamide dimer-DNA structures shows that

favorable cross-strand hydrogen bonds can form in the major groove of certain sequences in the presence of the propeller-twisted base pairs induced by the polyamide, whereas for others, an unfavorable electrostatic interaction results instead. This may contribute to the difficulty of targeting 5'-G*A*-3' and 5'-G*CG*-3' sequences. A trend for the 5'-GATC-3' sequences to display a narrower minor groove than 5'-GTAC-3' may also relate to differences in affinity. Since 5'-G*CG*-3' sequences can be efficiently targeted by introducing a β-alanine prior to the specific imidazole, disparity of the helical parameters of the polyamide and the DNA probably contribute to difficulty targeting this sequence. Additionally, the β-alanine may relieve the requirement for propeller twisted base pairs so that unfavorable interactions are not incurred for these sequences. Further assessment of this speculation awaits a high resolution structure of a polyamide-DNA complex containing β-alanine residues.

Although many potentially important interactions may be observed structurally, it is highly difficult to deduce which provide important contributions to the overall binding affinity (Clackson & Wells, 1995). This is clearly demonstrated by the absence of partial melting of the AT base pairs in the second structure of an Hp-containing polyamide, a feature that was previously believed to be the basis of the slightly reduced affinity for these compounds relative to the Py- counterparts. Therefore, the coupling of multiple structures with thermodynamic analysis may provide a more reliable picture of the underlying interactions than would a single structure alone. The increasing efficiency of high resolution structure determination makes the routine determination of multiple structures for this purpose an increasingly realistic goal.

4.7 Methods

Crystallization: The polyamides ImPyHpPyβDp and ImPyHpPyβDp were synthesized by solid phase methods (Baird & Dervan, 1996) using Boc-protected 3-methoxypyrrole, imidazole and pyrrole amino acids and purified by reverse phase chromatography. Identity and purity of the polyamides was confirmed by ^{1}H nuclear magnetic resonance and matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Synthetic deoxyribonucleotides were synthesized without removal of the 5'-dimethoxy-trityl group, and purified by two rounds of reverse phase on a C8 column (Pharmacia). The ammonium salt of the purified oligonucleotide was incubated with excess polyamide (tri-fluoro-acetate salt) prior to crystallization. Crystallization conditions are similar to those used for ImImPyPyβDp•5'-CCAGGCCTGG-3' (Kielkopf *et al.*, 1998a).

Structure determination: Crystals were flash cooled to -160° C using 35% 2-methyl-2,4-pentane-diol as a cryoprotectant prior to data collection. X-ray diffraction data were collected on an R-Axis IV image plate using CuKα radiation produced by a Rigaku RU-H3RHB rotating anode generator with double focusing mirrors and a Ni filter, and processed with DENZO/SCALEPACK (Otwinowski & Minor, 1997). The structures were solved by molecular replacement using the stand-alone version of the program AmoRe (Navazza, 1994) and the structure of ImPyPyPyβDp•5'-CCAGTACTGG-3' mutated to the appropriate sequence as starting models. The structures were refined against all data with bulk solvent correction and anisotropic B-scale using CNS (Brünger *et al.*, 1998), and restrained to the DNA parameters of Parkinson *et al.* (1991) and dictionaries for the polyamides constructed with the help of XPLO2D (Kleygwegt & Jones, 1994). Manual model building was performed using the program O (Jones *et al.*, 1991). All additional data manipulation was performed by use of programs in the CCP4 package (CCP4, 1994). An unbiased free R-factor comprising 6% of the data was excluded from the refinement (Brünger, 1993). Figures were drawn by use of Bobscript (Esnouf, 1997). Molscript (Kraulis, 1991), and Raster3D (Bacon & Anderson, 1988; Merritt & Murphy, 1994).

Quality of the model: The final model of ImPyHpPyβDp bound to 5'-CC**AGATCT**GG-3' contains two polyamides, one DNA duplex and 30 water molecules in the crystallographic asymmetric unit; the ImPyPyPyβDp counterpart consists of two polyamides, one DNA duplex and 52 water molecules (Table 4.5). The crystallographic R-factor for the ImPyHpPyβDp structure is 23.7%, free R-factor 23.8% for all data from 17.0 to 2.27 Å resolution. Similarly, the crystallographic R-factor for ImPyPyPyβDp•5'-CC**AGATCT**GG-3' is 22.9%, free R 22.9% from 17.0 to 2.15 Å resolution. The structures have reasonable geometry, both with r.m.s. deviations in bond lengths and angles of 0.02 Å and 2.1°, respectively. In all of the structures, the β-Dp tails have poorly defined electron density, and the β-Dp tail of one of the ImPyHpPyβDp molecules was modeled with an alternate conformation. Otherwise, the omit electron density for the polyamides is well-defined, and reveals the position of the Hp residue (Figure 4.6). The r.m.s. deviation between the DNA atoms of the two strands related by non-crystallographic symmetry is 0.5 Å for both structures. The ImPyPyPyβDp•5'-CC**AGATCT**GG-3' structure was determined in space group $P2_12_12_1$, and shares the crystal packing of ImImPyPyβDp•5'-CC**AGGCCT**GG-3', whereas ImPyHpPyβDp complex crystallized in space group $P3_1$. In contrast, the previously solved ImHpPyPyβDp and ImPyPyPyβDp•5'-CC**AGATCT**GG-3' structures were both determined in space group C2. Although there are no lattice contacts that would be anticipated to affect the structural analysis, it should be kept in mind that these complexes do not share identical crystal packing environments.

Acknowledgments: We are grateful to the NIH for research support, to the NSF and NIH for predoctoral fellowships to C.L.K. and R. E. B., to J. Edward Richter for an undergraduate fellowship to J.M.T., and to the HHMI for a predoctoral fellowship to E.E.B. We thank A. Chirino and M. Williamson for assistance with computational and x-ray equipment respectively, and S. Horvath for oligonucleotide synthesis.

Figure 4.6. **(a)** Omit $|F_o|$ - $|F_c|$ electron density map for one of the ImPyHpPy polyamides molecules, contoured at 1.5 σ. **(b)** Corresponding electron density for one of the ImPyPyPy polyamides. The polyamide residues are colored consistently throughout; Hp is red, and Py residues paired with Hp or Py are yellow. The residues involved in Im/Py pairings are shown in gray.

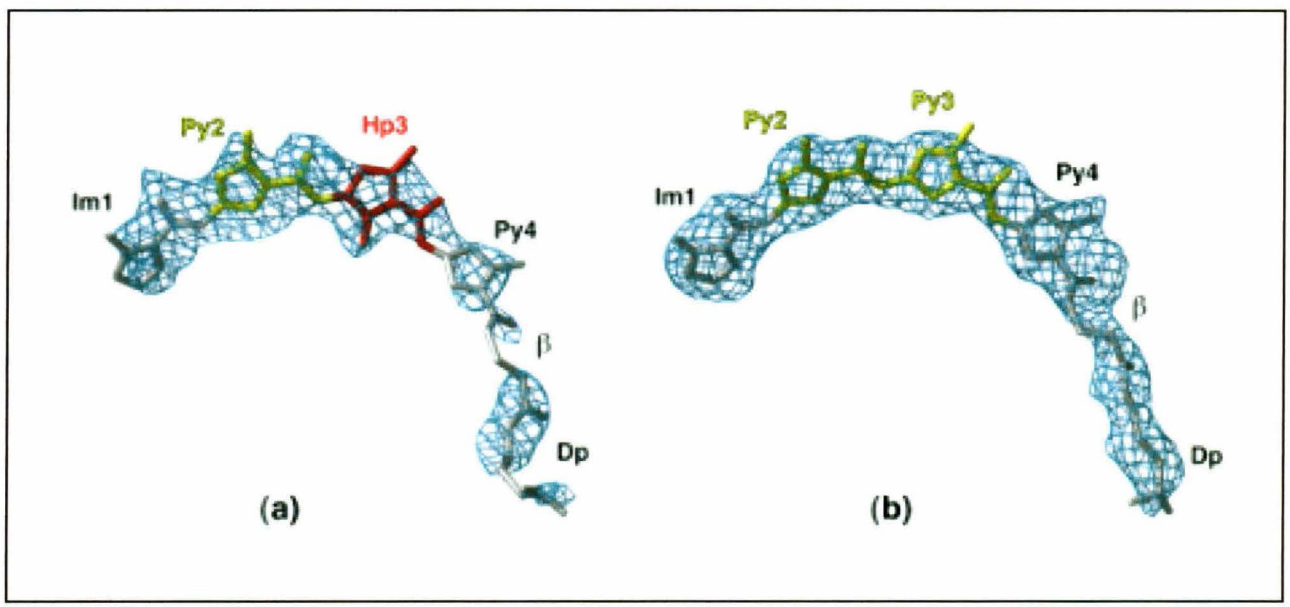


Table 4.5. Crystallographic data and refinement statistics.

Data Collection[6]		
	ImPyHpPy	ImPyPyPy
Space Group	$P3_1$	$P2_12_12_1$
Unit Cell (Å)	**a**=**b**=31.2,**c**=46.1	**a**=34.6,**b**=38.7,**c**=47.2
Resolution (Å)	2.27-17.0	2.15-17.0
Measured Reflections	7261	21424
Unique Reflections	2680	3618
Completeness (%)	96.1(87.0)	97.2(94.2)
I/σ(I)	42.3(3.1)	33.8(7.4)
R_{sym}[7] (%)	2.3(20.4)	4.1(11.9)
Refinement		
R_{crys}[8] (%)	23.7	22.9
R_{free} (%)	23.8	22.9
Rms deviation of bond lengths (Å)	0.02	0.02
Rms deviation of bond angles (°)	2.1	2.1
Number of Nonhydrogen Atoms		
DNA	404	404
Polyamide	96	94
Water Molecules	30	52

[6] Values in parentheses are for the highest resolution shell: 2.35-2.27 Å for ImPyHpPy, and 2.25-2.15 Å for the ImPyPyPy data sets.

[7] $R_{sym} = \sum_{hkl}\sum_{I}|I_i - \langle I\rangle|/\sum_{hkl}\sum_{I}I_i$ where I_i is an intensity I for the ith measurement of a reflection with indices hkl and $\langle I\rangle$ is the weighted mean of all measurements of I.

[8] $^3R_{cryst} = \sum_{hkl}||F_{obs}(hkl)| - k|F_{calc}(hkl)||) / \sum_{hkl}|F_{obs}(hkl)|$ for the working set of reflections, R_{free} is R_{cryst} for 5% of the reflections excluded from the refinement.

4.8 References

Bacon, D. J. & Anderson, W. F. (1988). A fast algorithm for rendering space filling molecule pictures. *J. Mol. Graph.* **6**, 219-220.

Baird, E. E. & Dervan, P. B. (1996). Solid phase synthesis of polyamides containing imidazole and pyrrole amino acids. *J. Am. Chem. Soc.* **118**, 6141-6146.

Brünger, A. T. (1993). Assessment of phase accuracy by cross-validation: the free R value, methods and applications. *Acta Crystallog.* **D49**, 24-36.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallog.* **D54**, 905-921.

Chen, X., Ramakrishnan, B., Rao, S. T. & Sundaralingam, M. (1994). Binding of two distamycin A molecules in the minor groove of an alternating B-DNA duplex. *Nature Struct. Biol.* **1**, 169-175.

Chen, X., Ramakrishnan, B. & Sundaralingam, M. (1997). Crystal structures of side-by-side binding of distamycin to AT-containing DNA octamers d(ICITACIC) and d(ICATATIC). *J. Mol. Biol.* **267**, 1157-1170.

Clackson, T. & Wells, J. A. (1995). A hot spot of binding energy in a hormone-receptor interface. *Science.* **267**, 383-386.

Collaborative Computational Project No. 4. (1994). The CCP4 Suite: programs for protein crystallography. *Acta Crystallog.* **D50**, 760-763.

de Clairac, R. P. L., Geierstanger, B. H., Mrksich, M., Dervan, P. B. & Wemmer, D. E. (1997). NMR characterization of hairpin polyamide complexes with the minor groove of DNA. *J. Am. Chem. Soc.* **119**, 7909-7916.

de Clairac, R. P. L., Seel, C. J., Geierstanger, B. H., Mrksich, M., Baird, E. E., Dervan, P. B. & Wemmer, D. E. (1999). NMR characterization of the aliphatic β/β pairing for recognition of base pairs in the minor groove of DNA. *J. Am. Chem. Soc.* **121**, 2956-2964.

Dickinson, L. A., Gulizia, R. J., Trauger, J. W., Baird, E. E., Mosier, D. E., Gottesfeld, J. M. & Dervan, P. B. (1998). Inhibition of RNA polymerase II transcription in human cells by synthetic DNA-binding ligands. *Proc. Natl. Acad. Sci. USA* **95**, 12890-12895.

Ding, S., Kielkopf, C. L., Kuhn, P. & Rees, D. C. Crystal structure of a B-form oligonucleotide at 0.74 Å resolution. *In preparation.*

Esnouf, R. M. (1997). An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph.* **15**, 132-134.

Geierstanger, B. H., Jacobson, J. P., Mrksich, M., Dervan, P. B. & Wemmer, D. E. (1994a). Structural and dynamic characterization of the heterodimeric and homodimeric complexes of distamycin and 1-methylimidazole-2-carboxamide-netropsin bound to the minor groove of DNA. *Biochemistry*. **33**, 3055-3062.

Geierstanger, B. H., Mrksich, M., Dervan, P. B. & Wemmer, D. E. (1994b). Design of a G•C specific DNA minor groove binding peptide. *Science*. **266**, 646-650.

Gottesfeld, J. M., Neely, L., Trauger, J. W., Baird, E. E. & Dervan, P. B. (1997). Regulation of gene expression by small molecules. *Nature*. **387**, 202-205.

Herman, D. M., Baird, E. E. & Dervan, P. B. (1999). Tandem hairpin motif for recognition in the minor groove of DNA by pyrrole-imidazole polyamides. *Chem. Eur. J.* **5**, 975-983.

Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog.* **A47**, 110-119.

Kelly, J. J., Baird, E. B. & Dervan, P. B. (1996). Binding site size limit of the 2:1 pyrrole-imidazole polyamide-DNA motif. *Proc. Natl. Acad. Sci. USA* **93**, 6981-6985.

Kielkopf, C. L., Baird, E. E. & Dervan, P. B. (1998a). Structural basis for G•C recognition in the DNA minor groove. *Nature Struct. Biol.* **5**, 104-109.

Kielkopf, C. L., White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E. & Dervan, P. B. (1998b). A structural basis for recognition of A•T and T•A base pairs in the minor groove of B-DNA. *Science*. **282**, 111-115.

Kleygwegt, G. J. & Jones, T. A. In *From first map to final model* (eds Hubbard, R. & Waller, D. A.) 59-66 (SERC Daresbury Laboratory, UK; 1994).

Kraulis, P. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946-950.

Lavery, R. & Sklenar, H. (1988). The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biolmolec. Struct. Dynam.* **6**, 63-91.

Leonard, G. A. & Hunter, W. N. (1993). Crystal and molecular structure of d(CGTAGATCTACG) at 2.25 Å resolution. *J. Mol. Biol.* **234**, 198-208.

Merritt, E. A. & Murphy, M. E. P. (1994). Raster3D version 2.0- a program for photorealistic molecular graphics. *Acta Crystallog.* **D50**, 869-873.

Mitra, S. N., Wahl, M. C. & Sundaralingam, M. (1999). Structure of the side-by-side binding of distamycin to d(GTATATAC)$_2$. *Acta Crystallog.* **D55**, 602-609.

Mrksich, M. & Dervan, P. B. (1993). Antiparallel side-by-side heterodimer for sequence-specific recognition in the minor groove of DNA by a distamycin 1-methylimidazole-2-carboxamide pair. *J. Am. Chem. Soc.* **115**, 2572-2576.

Mrksich, M., Parks, M. E. & Dervan, P. B. (1994). Hairpin peptide motif. A new class of oligopeptides for sequence-specific recognition in the minor groove of double helical DNA. *J. Am. Chem. Soc.* **116**, 7983-7988.

Mrksich, M., Wade, W. S., Dwyer, T. J., Geierstanger, B. H., Wemmer, D. E. & Dervan, P. B. (1992). Antiparallel side-by-side dimeric motif for sequence-specific recognition in the minor groove of DNA by the designed peptide 1-methylimidazole-2-carboxamide netropsin. *Proc. Natl. Acad. Sci. USA* **89**, 7586-7590.

Navazza, J. (1994). AmoRe, an automated package for molecular replacement. *Acta Crystallog.* **A50**, 157-163.

Otwinowski, Z. & Minor, W. (1997) Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326.

Parkinson, G., Vojtechovsky, J., Clowney, L., Brünger, A. T. & Berman, H. M. (1991). New parameters for the refinement of nucleic acid-containing structures. *Acta Crystallog.* **A47**, 110-119.

Parks, M. E., Baird, E. E. & Dervan, P. B. (1996). Optimization of the hairpin polyamide design for recognition of the minor groove of DNA. *J. Am. Chem. Soc.* **118**, 6147-6152.

Pelton, J. G. & Wemmer, D. E. (1989). Structural characterization of a 2-1 distamycin-A •d(CGCAAATTTGCG)$_2$ complex by two-dimensional NMR. *Proc. Natl. Acad. Sci. USA* **86**, 5723-5727.

Pelton, J. G. & Wemmer, D. E. (1990). Binding modes of distamycin-A with d(CGCAAATTTGCG)$_2$ determined by two-dimensional NMR. *J. Am. Chem. Soc.* **112**, 1393-1399.

Seeman, N. C., Rosenburg, J. M. & Rich, A. (1976). Sequence specific recognition of double helical nucleic acids by proteins. *Proc. Natl. Acad. Sci. USA* **73**, 804-808.

Swalley, S. E., Baird, E. E. & Dervan, P. B. (1997). Discrimination of 5'-GGGG-3', 5'-GCGC-3' and 5'-GGCC-3' sequences in the minor groove of DNA by eight-ring hairpin polyamides. *J. Am. Chem. Soc.* **119**, 6953-6961.

Swalley, S. E., Baird, E. E. & Dervan, P. B. (1999). Effects of γ-turn and β-tail amino acids on sequence specific recognition of DNA by hairpin polyamides. *J. Am. Chem. Soc.* **119**, 6953-6961.

Trauger, J. W., Baird, E. E. & Dervan, P. B. (1996a). Recognition of DNA by designed ligands at subnanomolar concentrations. *Nature*, **382**, 559-561.

Trauger, J. W., Baird, E. E. & Dervan, P. B. (1996b). Extension of sequence-specific recognition in the minor groove of DNA by pyrrole-imidazole polyamides to 9-13 base pairs. *J. Am. Chem. Soc.* **118**, 6160-6166.

Trauger, J. W., Baird, E. E. & Dervan, P. B. (1998a). Cooperative hairpin dimers for recognition of DNA by pyrrole-imidazole polyamides. *Angew. Chem., Int. Ed. Engl.* **37**, 1421-1423.

Trauger, J. W., Baird, E. E. & Dervan, P. B. (1998b). Recognition of 16 base pairs in the minor groove of DNA by a pyrrole-imidazole polyamide dimer. *J. Am. Chem.* **120**, 3534-3535.

Turner, J. M., Baird, E. E. & Dervan, P. B. (1997). Recognition of seven base pair sequences in the minor groove of DNA by ten-ring pyrrole-imidazole polyamide hairpins. *J. Am. Chem. Soc.* **119**, 7636-7644.

Turner, J. M., Swalley, S. E., Baird, E. E. & Dervan, P. B. (1998). Aliphatic/aromatic amino acid pairings for polyamide recognition in the minor groove of DNA. *J. Am. Chem. Soc.* **120**, 6219-6226.

Wade, W. S., Mrksich, M. & Dervan, P. B. (1992). Design of peptides that bind in the minor groove of DNA at 5'-$(A,T)_5$G(A,T)C(A,T)-3' sequences by a dimeric side-by-side motif. *J. Am. Chem. Soc.* **114**, 8783-8794.

Wade, W. S., Mrksich, M. & Dervan, P. B. (1993). Binding affinities of synthetic peptides, pyridine-2-carboxamidenetropsin and 1-methylimidazole-2-carboxamidenetropsin, that

form 2/1 complexes in the minor groove of double helical DNA. *Biochemistry*. **32**, 11385-11389.

White, S., Baird, E. E. & Dervan, P. B. (1996). Effects of the A•T/T•A degeneracy of pyrrole-imidazole polyamide recognition in the minor groove of DNA. *Biochemistry*. **35**, 12532-12537.

White, S., Baird, E. E. & Dervan, P. B. (1997a). Orientation preferences of pyrrole-imidazole polyamides in the minor groove of DNA. *J. Am. Chem. Soc.* **119**, 8756-8765.

White, S., Baird, E. E. & Dervan, P. B. (1997b). On the pairing rules for recognition in the minor groove of DNA by pyrrole-imidazole polyamides. *Chem. Biol.* **4**, 569-577.

White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E. & Dervan, P. B. (1998a). Recognition of the four Watson-Crick base pairs in the DNA minor groove by synthetic ligands. *Nature*. **391**, 468-471.

White, S. (1998b). Ph.D. thesis. Department of Chemistry and Chemical Engineering. California Institute of Technology.

White, S., Turner, J. M., Szewczyk, J. W., Baird, E. E. & Dervan, P. B. (1999). Affinity and specificity of multiple hydroxypyrrole/pyrrole ring pairings for coded recognition of DNA. *J. Am. Chem. Soc.* **121**, 260-261.

Yuan, H., Quintana, J. & Dickerson, R. E. (1992). Alternative structures for alternating poly(dA-dT) tracts: the structure of the B-DNA decamer C-G-A-T-A-T-A-T-C-G. *Biochemistry*. **31**, 8009-8021.

missing

Chapter 5

Structural Basis of Intercalation via the DNA Major Groove

Abstract: One of the common ways that molecules interact with nucleic acids is by inserting a planar ligand between the base pairs. Intercalating complexes of rhodium(III) bind DNA via the major groove, and are strong photo-oxidants that can promote long range electron transfer through the DNA double helix. The 1.2 Å resolution crystal structure of a designed, sequence-specific rhodium intercalator details the conformational changes that occur upon intercalation of a small molecule within a DNA helix. The aromatic ligand is deeply stacked between the target base pairs of five independent duplexes present in the crystallographic asymmetric unit. Based upon the structure, an intercalator may be considered as an additional base pair with specific functional groups positioned within the DNA groove. This observation has implications for expanding the current repertoire of sequence-specific intercalators and for understanding long range electron transfer in DNA.

Manuscript in preparation: Kielkopf, Erkkila, Hudson, Barton & Rees.

5.1 Introduction

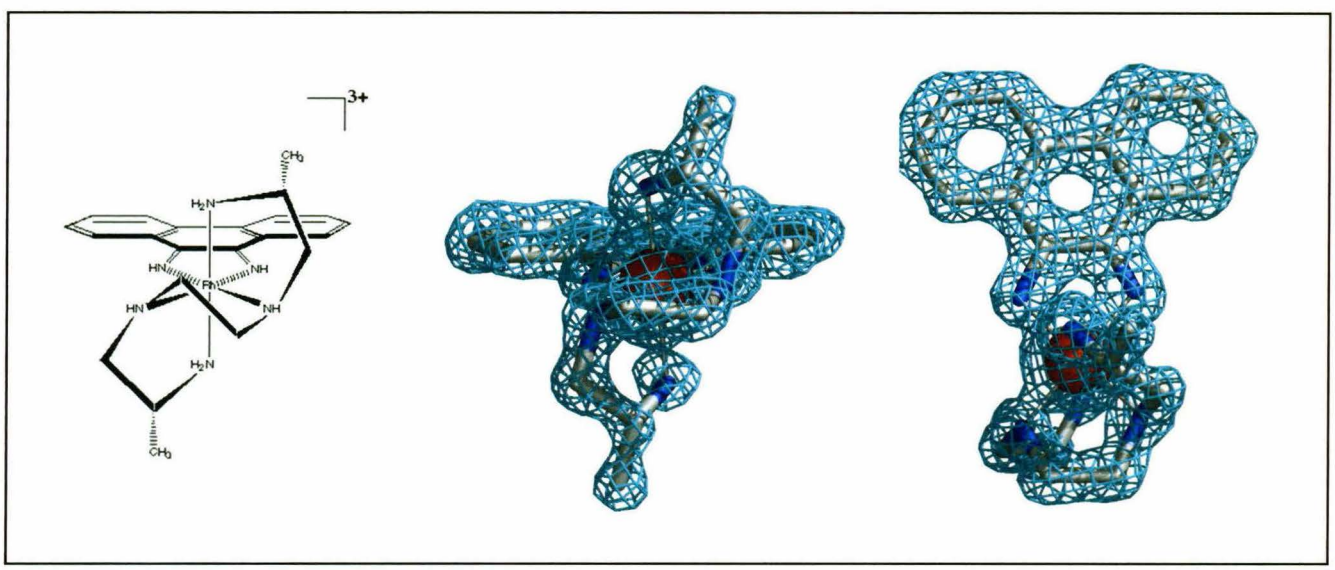
Before the elucidation of the Watson-Crick model for the DNA double helix, it was realized that many planar, aromatic compounds had mutagenic activity due to their interaction with DNA. Subsequent biochemical and fiber diffraction studies by Lerman in the early 1960's resulted in the intercalation model of binding (Lerman, 1961; Wilson & Jones, 1981; Gago, 1998); the aromatic small molecules bound to essentially B-form DNA by insertion of a planar ligand between the base pairs. Although this binding mode resulted in high affinities of the ligand for double stranded DNA, the majority of the interactions were with the aromatic surfaces of the bases rather than with the distinctive functional groups in the DNA grooves. Consequently, most intercalators, such as ethidiums and acridines are relatively nonspecific DNA binding agents.

In contrast with most naturally occurring intercalators, octahedral complexes of Rh(III) containing the phi ligand (where phi is 9,10-phenanthrenequinone diimine) constitute a class of intercalators with the ability to specifically bind DNA (Johann & Barton, 1996). The octahedral geometry of these complexes provides a rigid framework that predictably orients the ancillary, non-intercalating ligands in the major groove, and these ancillary ligands can be synthetically modified to alter the sequence specificity and reactivity of the molecule. As such, rhodium intercalators can compete directly with DNA binding proteins (Odom *et al.*, 1999). Furthermore, these metallointercalators are photoactive, and can carry out both short and long range photoinduced oxidation of DNA through reactions that are a sensitive function of their stacking within the DNA helix (DNA (Sitlani *et al.*, 1992; Hall *et al.*, 1996; Johann & Barton, 1996; Dandliker *et al.*, 1997). Metallointercalators have been exploited extensively for site-specific targeting of DNA (Johann & Barton, 1996), for the photochemical repair of DNA lesions (Dandliker *et al.*, 1997), as photoluminescent reporters of DNA (Murphy & Barton, 1993), in mismatch detection (Jackson *et al.*, 1997; 1999) and as probes of DNA structure (Chow

& Barton, 1992; Terbrueggen & Barton, 1995). An atomic resolution structure of the ligand-bound DNA site would contribute to understanding the specificity and reactivity of metallointercalators as well as provide a rational basis for subsequent design.

The compound Δ-α-[Rh[(R,R)-Me_2trien]phi]$^{3+}$ (where (R,R)-Me_2trien is 2*R*,9*R*-diamino-4,7-diazadecane) is an intercalator that was successfully designed to recognize the major groove of the four base pair sequence 5'-TG|CA-3' (where | indicates phi insertion) (Krotz *et al.* 1993). Subsequent NMR analysis revealed that the complex specifically interacts with the predicted 5'-TG|CA-3' binding site (Hudson & Barton, 1998). To determine the sugar pucker, backbone conformation, sequence specific interactions and π-stacking of a DNA helix bound to an intercalator, Δ-α-[Rh[(R,R)-Me_2trien]phi]$^{3+}$ was cocrystallized with the eight base pair oligonucleotide 5'-G-dIU-TGCAAC-3' (dIU, 5-iodo-deoxy-uridine). Although the native sequence 5'-GTTGCAAC-3' also cocrystallized with the intercalator, the iodinated crystals diffracted to 1.2 Å resolution and were used for the refinement and subsequent analysis of the structure (Methods).

Figure 5.1. Representative omit $|F_o|$ - $|F_c|$ electron density map for one of the intercalators, contoured at 4.0 σ, viewed from two perpendicular orientations with the chemical structure indicated to one side.



5.2 Overall Structure and π-Stacking Interactions

Five end-to-end complexes in the asymmetric unit of the crystal provide five independent views of an intercalated duplex (Figure 5.2). Intercalation occurs in the center of each of the five double-stranded oligonucleotides, via the major groove of the predicted 5'-TGCA-3' binding site (Figure 5.3). Although the 3' end of one DNA strand is closest to the 5' end of the next, instead of the +36° twist that would be expected between consecutive nucleotides of B-form DNA, the twist relating the ends of the duplexes is –26°. The sum of the twists between the consecutive base pairs in all five duplexes, including the –26° intermolecular twists, is 3 x 360°; the molecules form three complete helical turns of DNA. Thus, a pseudo-continuous DNA helix that traverses an end-to-end distance of over 150 Å is observed in the crystallographic asymmetric unit.

The intercalation site is unwound ~10° by insertion of the ligand, and the rise is doubled (Table 5.1). However, different types of intercalators and flanking DNA sequences display a wide range of twists (Lerman, 1961; Sobell *et al.*, 1977; Wang *et al.*, 1978; Saenger, 1984; Wang *et al.*, 1986; 1987; Kamitori & Takusagawa, 1992; Gasper *et al.*, 1998; Hudson & Barton, 1998), indicating that the exact twist of an intercalation site is probably flexible. Unlike intercalated dinucleotides, the sugar pucker at the binding sites for the rhodium complexes is consistently characteristic of B-form DNA. This difference may result from the context of a longer DNA helix, which avoids the greater flexibility and crystal packing forces at the ends of shorter oligonucleotides. In addition, the torsion angles display similar values to those of B-form DNA. Therefore, from this structure, a gross distortion of the sugar-phosphate backbone conformation is not required to separate the base pairs and admit an intercalating ligand.

Likewise, introduction of the rhodium intercalator does not require disruption of the base stacking of the double helix. The length of the Watson-Crick hydrogen bonds between the base pairs are unperturbed by intercalation by the phi ligand. This contrasts

with the crystal structure of a copper porphyrin, hemi-intercalated one base pair from the end of an oligonucleotide, in which the terminal base is flipped out of the double helix (Lipscomb *et al.*, 1996). Likewise, the structure of triostin A bound at the ends of an oligonucleotide reveals that this type of intercalator induces Hoogsteen base pairs in the DNA (Wang *et al.*, 1986). Based on the structures, it would appear that these molecules would be poor electron transfer agents because they interfere with optimal stacking of the base pairs.

The nature of the π interactions between the phi ligand of the rhodium complex and the surrounding bases is similar to that of consecutive base pairs in a DNA duplex (Bugg *et al.*, 1971). The aromatic π clouds are positioned over atoms at the edges of the bases below, to interact favorably with the π-orbitals of the bases. The phi ligand stacks entirely upon both of the guanine bases with little contact to the cytosines. More than 60% of the hydrophobic surface area of the ligand is buried upon formation of the complex, which may provide an energetic basis for the tight binding ($K_d \sim 10^{-8}$ M). The phi ligand is deeply inserted into the DNA base pair stack, with the rhodium atoms only 4.6 Å from the helix axis, as compared to the 10 Å radius of the DNA cross-section. Since disturbing the DNA π-stack profoundly affects long range electron transfer in the DNA helix (Kelley & Barton, 1999), the tightly stacked base pairs associated with the $Rh[phi]^{+3}$ complex provides a structural basis for the efficient long range electron transfer associated with metallointercalators.

Exposure of the intercalated rhodium complex to higher energy light can result in direct DNA strand scission via hydrogen abstraction from the 5' side of the intercalation site (Krotz *et al.*, 1993; Shields & Barton, 1995). The electronegative deoxyribose O4' atoms at the 3' side of the binding site are in close proximity (≥3.0 Å) to the edge of the phi ligand. Similar interactions between the sugar O4' of the nucleic acid backbone and aromatics often stabilize nucleic acid structures (Bugg *et al.*, 1971; Egli & Gessner, 1995). It is an intriguing possibility that the sugar O4' may stabilize the phi-cation

radical generated by photo-excitation of the rhodium complex below the 5'-deoxyribose moiety, and thereby facilitate DNA cleavage.

Figure 5.2. The five complexes in the asymmetric unit of the crystal.

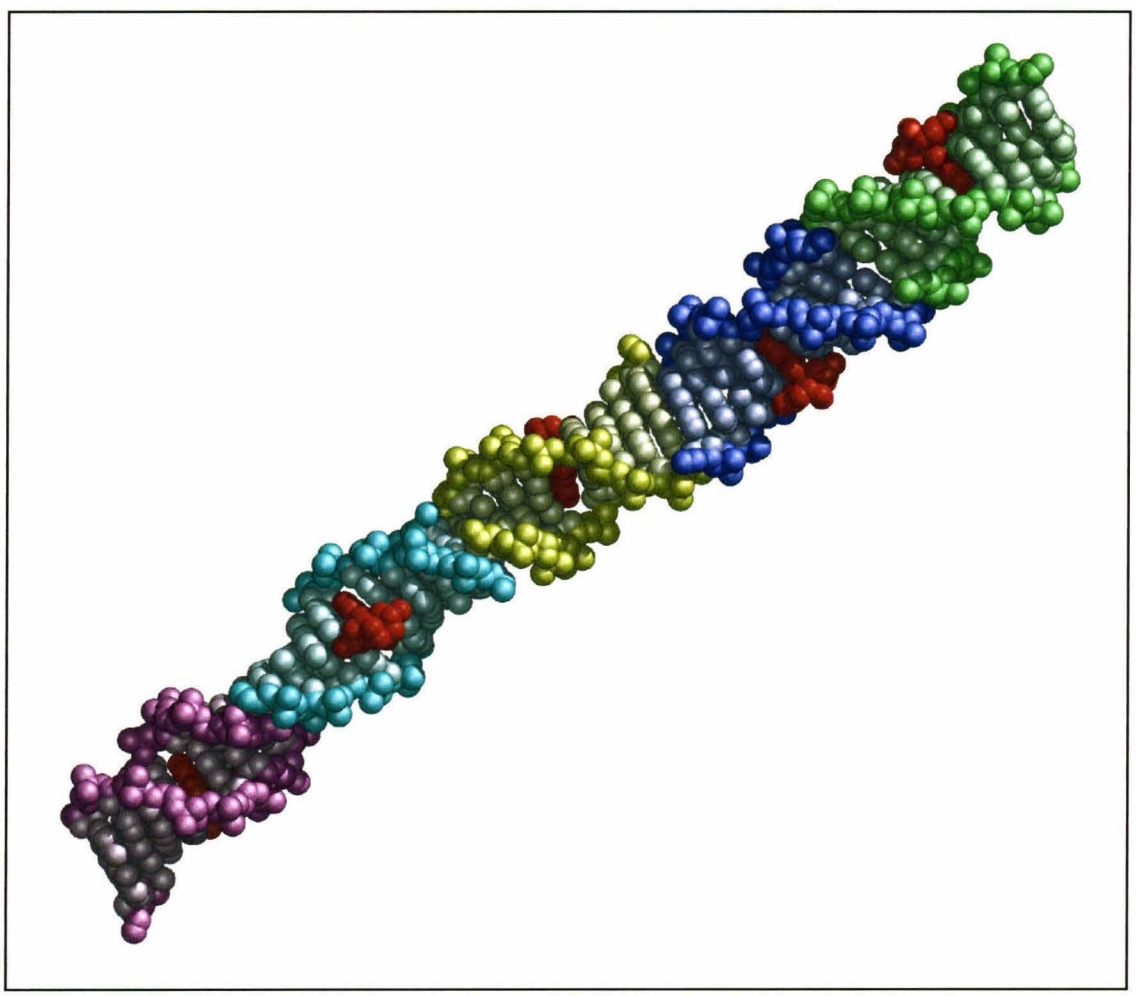


Figure 5.3. Two views of one of the Rh- DNA complexes, one into the major groove, the other rotated about the helix axis. Thymine is colored purple and guanine, green. A schematic DNA binding model of the predicted contacts is shown to the left.

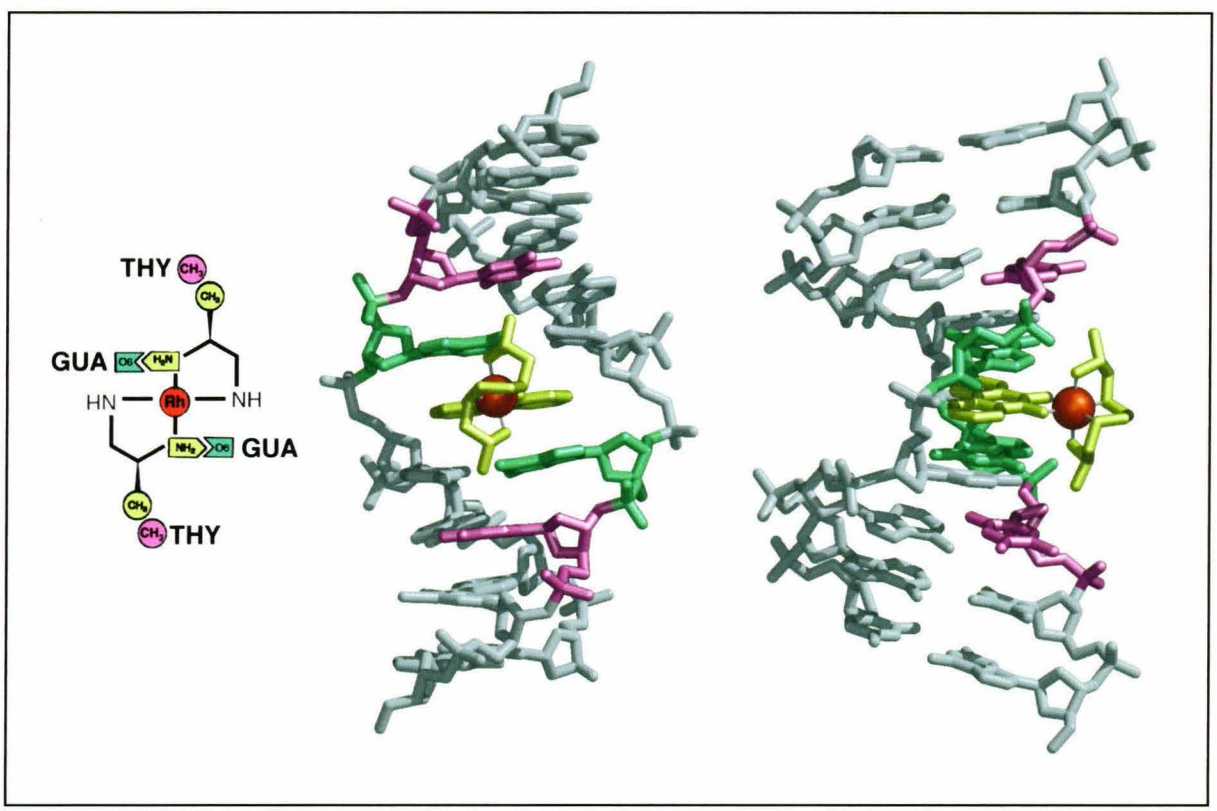


Table 5.1. DNA conformation.[1]

Base Step	G/dIU	dIU/T	T/G	G/C	C/A	A/A	A/C	B-DNA[¶2]
Twist (°)	34	31	46	23	45	33	30	36
Rise (Å)	3.4	3.2	2.8	7.3	3.2	3.2	3.2	3.3
Shift (Å)	-0.3	0.4	0.2	0.6	0.1	0.1	0.9	-0.1
Slide (Å)	-1.0	-0.2	2.3	1.2	1.8	-0.1	-1.0	-0.8
Roll (°)	1.1	5.1	-6.8	-0.8	-14.3	4.3	-1.4	-3.6
β(O5'-C5', °)	180	180	146	180	165	158	161	-146
P-P distance(Å)	-	6.6	6.4	7.0	6.6	7.0	6.3	6.5
Bend (°)	5	12	9	13	17	4	4	4

Nucleotide	G	DIU	T	G	C	A	A	C	B-DNA[¶]
Buckle[‖](°)	0.6	-0.4	5.7	14.2	-13.5	-4.7	0	1.1	0
χ(C1'-N, °)	-108	-111	-90	-80	-85	-91	-106	-140	-98
Sugar Pucker	C2'-endo	C2'-endo	C2'-endo	C2'-endo	C2'-endo	C2'-endo	C2'-endo	C2'-endo	C2'-Endo

	Major Groove Width	Major Groove Depth	Minor Groove Width	Minor Groove Depth
Average	12.1	-1.4	7.1	6.9
B-DNA	11.2	4.0	5.9	4.7

[1] Calculated using the program Curves (*Lavery & Sklenar, 1988*). Average values for the five duplexes are given.
[2] Ideal B-form DNA.

5.2 Sequence Specific Interactions

In addition to the π-interactions between the DNA and the aromatic ligand, the intercalator forms sequence specific interactions with the DNA major groove. The rhodium complex was originally designed to target the 5'-TGCA-3' site by introducing functionalities onto the non-intercalated framework of the complex to interact with base functionalities within the major groove (Krotz *et al.*, 1993). As intended, the axial methyl groups of the ligand are a comfortable 3.5-4.1 Å distance from the thymine methyl carbons (Figure 5.4). Another feature of the molecular design is evident in the hydrogen bonds donated by the axial amines of the metal complex to the guanine-O6 of the DNA (Figure 5.5). These atoms are an average 3 Å apart, and if hydrogens are assigned to the axial amines, the N-H···O6 angle is a favorable ~165° (Baker & Hubbard, 1984). The specificity for guanines arises because adenosine and cytosine have an sp^2-hybridized exocyclic amine, which is a poor hydrogen bond acceptor, in place of the guanine-O6. Although thymines also present an oxygen to the major groove, both the orientation and the distance of the thymine-O2 places it in an unfavorable position to accept a hydrogen bond from the axial amines.

There was a possibility in the molecular design that the intercalator axial amines would hydrogen bond with the guanine-N7 as well as the O6. In nine of the ten interactions with guanines, the N7 atoms are at least 3.6 Å from the axial amines of the rhodium complex, too far for a strong hydrogen bond. Instead, the interaction is mediated by an ordered water molecule, present in all five duplexes, that accepts a hydrogen bond from the axial amine while donating one to the guanine-N7. One of the molecules shows an exception, an alternate conformation of the guanine that places the N7 within hydrogen bond distance of the axial amine (3.0 Å). As a consequence, the O6 are unfavorably close (2.5 Å) to the amine, and the stacking of the phi on the guanine changes. The second conformation of that guanine forms hydrogen bonds between the

O6 atoms and the axial amines of the intercalator, as observed for the other molecules. A different type of intercalator-bound water molecule, found in all five complexes, accepts hydrogen bonds from the phi-imine and equatorial amine. The pseudo-twofold symmetry of the binding site, applied to the two types of water molecules, generates four water molecules specifically bound to the nine of the ten DNA strands (Figure 5.6). Therefore, as for many DNA-binding proteins, ordered water molecules form an integral part of the DNA-ligand interface (Schwabe, 1997).

Figure 5.4. Space-filling model of the binding site. The C5A, C5, C6, and C2' atoms of the thymines are shown in purple, whereas the other DNA atoms are cyan. The methyl carbons of the intercalator are yellow, other intercalator atoms are silver.

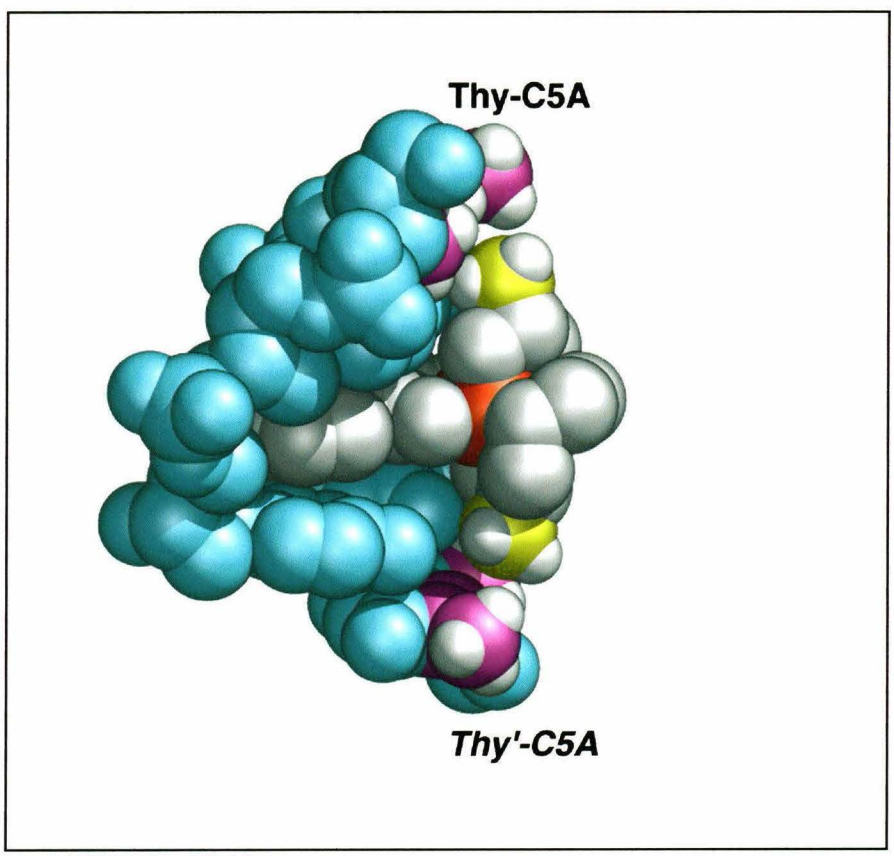


Figure 5.5. Specific hydrogen bonds between the intercalator and the guanines of the intercalation site.

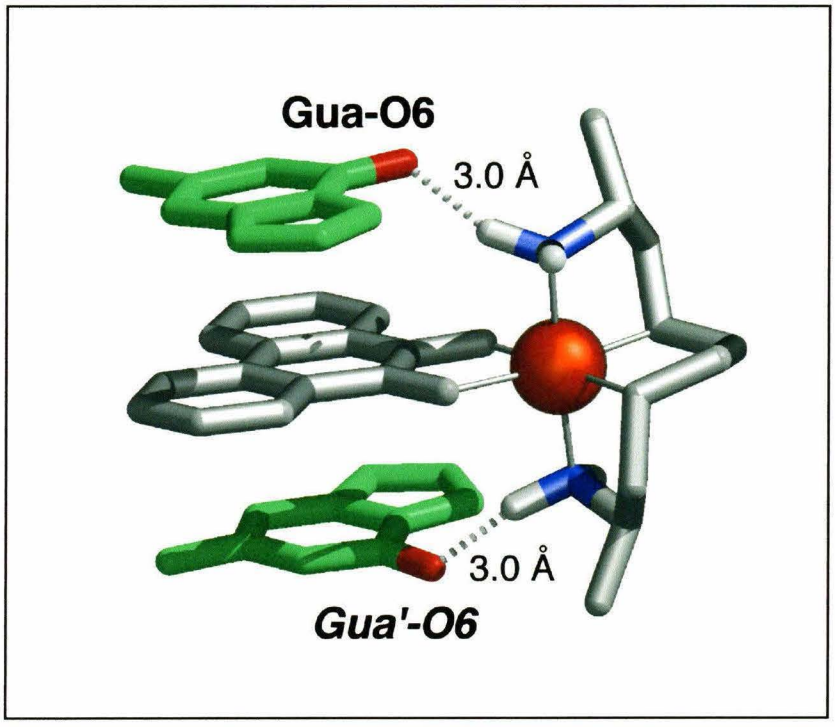
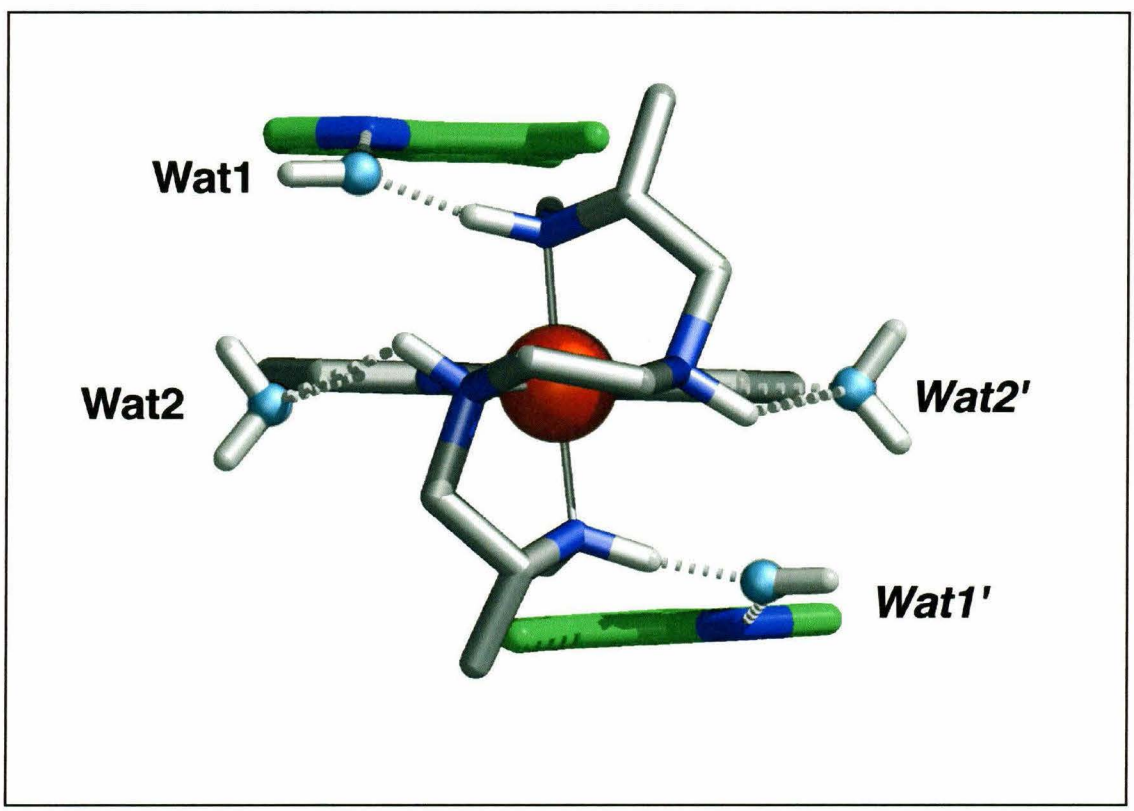


Figure 5.6. Hydrogen bonds between ordered water molecules and the intercalator.



5.4 Conclusion

The ability to target specific DNA sequences will be crucial for the development of intercalators as medical and chemical tools. Importantly, the observation of the designed contacts in the structure indicates that substitution of the ancillary ligands of octahedral Rh[phi]$^{3+}$ complexes is a viable strategy to target specific DNA sequences in the major groove (Krotz & Barton, 1994). This structure validates the original intercalation model proposed by Lerman almost four decades ago (Lerman, 1961); an intercalator resembles a base pair. The sugar pucker and rise per residue maintained, although variations in helical twist may occur with variations in sequence. There is no disruption of the helical π-stack by the metallointercalators in the three complete turns of intercalated DNA in the asymmetric unit of the crystal. Since electron transfer is exquisitely sensitive to changes in the stacking between the electron donors and acceptors, this is a key observation for understanding the long range electron transfer through the DNA helix initiated by metallointercalators.

5.6 Methods

The rhodium complex Δ-α-[Rh[(R,R)-Me_2trien]phi]$^{3+}$ was synthesized and isolated in isomerically pure form as described previously (Krotz & Barton, 1994; Hudson & Barton, 1998). Bright red crystals were obtained from a solution of 1 mM double-stranded oligonucleotide, 1.7 mM Rh(Me_2trien)phi^{3+}, 50 mM calcium acetate, 5 mM sodium cacodylate pH 7.0, and 10% 2-methyl-2,4-pentanediol (MPD) equilibrated in sitting drops versus a reservoir of 20% MPD at 4° C.

Crystals grew in space group C2 with unit cell dimensions of a=94.8 Å, b=24.0 Å, c=96.7 Å, β=93.4°. Data sets from flash-cooled crystals were collected on an R-Axis IV image plate using CuKα radiation produced by a Rigaku RU-H3RHB rotating anode generator with double focusing mirrors and a Ni filter, and processed with DENZO/SCALEPACK (Otwinowski, 1993). Because of the poor quality of the native crystals, the structure was solved by single isomorphous replacement with anomalous scattering (f''=6.8 electrons for I at λ=1.54 Å) using brominated crystals as the native (5'-G-dBrU-TGCAAC-3') and the corresponding iodinated sequence (5'-G-dIU-TGCAAC-3') for the derivative (Table 5.2). Heavy atom sites were located and refined using the program SOLVE (Terwilliger, 1999). The structure was refined using the program SHELXH (Sheldrick & Schneider, 1997) against data from an iodinated crystal collected on beamline 9-1 at the Stanford Synchrotron Radiation Laboratory (SSRL) and merged with data collected on the R-Axis IV to include the low resolution.

Molecules 1, 3 and 4 are well-defined, with average B-factors of 25, 19 and 25 Å^2, whereas molecules 2 and 5 are more mobile, with average B-factors of 40 and 42 Å^2. Pairwise r.m.s. differences between molecules 1-2, 1-3, 1-4, 1-5, 2-3, 2-4, 2-5, 3-4, 3-5 and 4-5 are 0.7, 0.3, 0.6, 0.6, 0.8, 0.9, 0.8, 0.5, 0.6 and 0.5 Å, respectively. Figures were drawn by use of Bobscript and Raster 3D (Kraulis, 1991; Merrit & Bacon, 1997; Esnouf, 1999).

Acknowledgments: We are grateful to the NIH for research support and a fellowship to K.E.E, and to the NSF and NIH for predoctoral fellowships to C.L.K. and K.E.E. We thank J. E. Wedekind, A. J. Chirino and R. Bau for assistance with the structure solution and S. Horvath for oligonucleotide synthesis. The rotation camera at SSRL is supported by the U.S. Department of Energy and NIH. Coordinates have been deposited in the Nucleic Acid Database with accession number UD0005.

Table 5.2. Crystallographic Statistics.

Data collection[3]

Data set	Resolution (Å)	Observed reflections	Unique Reflections	Completeness	<I/σ>	R_{sym}[4] (%)
Bromine (native)	2.8	21447	5367	97.2 (96.7)	40.0 (24.0)	3.1 (5.7)
Iodine	1.2	360925	122830	92.5 (88.2)	25.0 (3.6)	6.8 (27.8)

Phasing statistics

Derivative	Resolution range	Sites	PhP[5] Acentric/centric	R_{cullis}[6]
Iodine	20.0-3.0 Å	10	1.68/1.54	0.59

Quality of model

DNA atoms	Intercalator atoms	Water molecules	R_{crys} (R_{free})[7]	r.m.s.d. bonds	r.m.s.d. angles
1610	145	374	17.0 (21.0)%	0.018 Å	1.2°

[3] Values in parentheses are for the highest resolution shell.

[4] $R_{sym} = \Sigma_{hkl}\Sigma_i|I_i - \langle I\rangle|/\Sigma_{hkl}\Sigma_i I_i$ where I_i is an intensity I for the ith measurement of a reflection with indices hkl and <I> is the weighted mean of all measurements of I.

[5] Phasing power is the mean value of the heavy atom derivative structure factor amplitude divided by the residual lack of closure error.

[6] $R_{cullis}=\Sigma_{hkl}\,||F_{ph}| - |F_p| - F_{h,calc}|/\Sigma_{hkl}|F_{ph} - F_p|$ for centric reflections only.

[7] $R_{cryst} = \Sigma_{hkl}||F_{obs}(hkl)| - k|F_{calc}(hkl)||) / \Sigma_{hkl}|F_{obs}(hkl)|$ for the working set of reflections, R_{free} is R_{cryst} for 5% of the reflections excluded from the refinement.

5.7 References

Baker E. N. & Hubbard, R. E. (1984). Hydrogen-bonding in globular proteins. *Prog. Biophys. Molec. Biol.* **44**, 97-179.

Bugg, C. E., Thomas, J. M., Sundaralingam, M. & Rao, S. T. (1971). Stereochemistry of nucleic acids and their constituents. X. Solid state base-stacking patterns in nucleic acid constituents and polynucleotides. *Biopolymers.* **10**, 175-219.

Chow, C. S. & Barton, J. K. (1992). Transition metal complexes as probes of nucleic acids. *Method. Enzymol.* **212**, 219-242.

Dandliker, P. J., Holmlin, R. E. & Barton, J. K. (1997). Oxidative thymine dimer repair in the DNA helix. *Science.* **275**, 1465-1468.

Egli, M. & Gessner, R. V. (1995). Stereoelectronic effects of deoxyribose O4' on DNA conformation. *Proc. Natl. Acad. Sci. USA* **92**, 180-184 (1995).

Esnouf, R. M. (1997). An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph.* **15**, 132-134.

Gago, F. (1998). Stacking interactions and intercalative DNA binding. *Methods.* **14**, 277-292.

Gasper, S. M., Armitage, B., Shui, X. Q., Hu, G. G., Yu, C. J., Schuster, G. B. & Williams, L. D. (1998). Three-dimensional structure and reactivity of a

photochemical cleavage agent bound to DNA. J. *Am. Chem. Soc.* **120**, 12402-12409.

Hall, D. B. & Barton, J. K. (1997). Sensitivity of DNA-mediated electron transfer to the intervening π-stack: A probe for the integrity of the DNA base stack. *J. Am. Chem. Soc.* **119**, 5045-5046.

Holmlin, R. E., Dandliker, P. J. & Barton, J. K. (1997). Charge transfer through the DNA base stack. *Angew. Chem. Int. Ed. Engl.* **36**, 2714-2730.

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

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

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

Johann, T. W. & Barton, J. K. (1996). Recognition of DNA by octahedral coordination complexes. *Phil. Trans. R. Soc. Lond. A*. **354**, 299-324.

Kamitori, S. & Takusagawa, F. (1992). Crystal structure of the 2:1 complex between d(GAAGCTTC) and the anticancer drug actinomycin D. *J. Mol. Biol.* **225**, 445-456.

Kelley, S. O. & Barton, J. K. (1999). Electron transfer between bases in double helical DNA. *Science*. **283**, 375-381.

Kraulis, P. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946-950.

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

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

Krotz, A. H. & Barton, J. K. (1994). Rational construction of metallointercalators: Syntheses, structures and characterizations of isomers of ((2*S*, 9*S*)-diamino-4,7-diazadecane)(9,10-phenanthrene diimine-rhodium(III).

Lavery, R. & Sklenar, H. (1988). The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biolmolec. Struct. Dynam.* **6**, 63-91.

Lerman, L. S. (1961). Structural considerations on the interaction of DNA and acridines. *J. Mol. Biol.* **3**, 18-30.

Lipscomb, L. A., Zhou F. X., Presnell S. R., Woo R. J., Peek M. E., Plaskon R. R. & Williams L. D. (1996). Structure of a DNA porphyrin complex. *Biochemistry.* **35**, 2818-2823.

Merritt, E. A. & Murphy, M. E. P. (1994). Raster3D version 2.0- a program for photorealistic molecular graphics. *Acta Crystallog.* **D50**, 869-873.

Murphy, C. J. & Barton, J. K. (1993). Ruthenium complexes as luminescent reporters of DNA. *Methods Enzymol.* **226**, 576-594.

Odom, D. T., Parker C. S. & Barton J. K. (1999). Site-specific inhibition of transcription factor binding to DNA by a metallointercalator. *Biochem.* **38**, 5155-5163.

Otwinowski, Z. in *Data Collection and Processing* N. I. L. Sawyer, S. Bailey, Ed. (SERC Daresbury Laboratory, UK, 1993) pp. 56-61.

Saenger, W. *Principles of Nucleic Acid Structure* (Springer-Verlag, New York, NY; 1984).

Schwabe, J. W. R. (1997). The role of water in protein DNA interactions. *Curr. Opin. Struct. Biol.* **7**, 126-134.

Sheldrick, G. M. & Schneider, T. R. (1997). SHELXL: High-resolution refinement. *Methods Enzymol.* **277**, 319-343.

Shields, T. P. & Barton, J. K. (1995a). Sequence selective DNA recognition and photocleavage: A comparison of enantiomers of $[Rh(en)_2phi]^{+3}$. *Biochemistry.* **34**, 15037-15048.

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

Sobell, H. M., Tsai, C. C., Jain, S. C. & Gilbert, S. G. (1977). Visualization of drug-nucleic acid interactions at atomic resolution. III. Unifying structural concepts in understanding protein-DNA interactions and their broader implications in understanding protein-DNA interactions. *J. Mol. Biol.* **114**, 333-365.

Terbrueggen, R. H. & Barton, J. K. (1995). Sequence specific DNA binding by a rhodium complex; recognition based on sequence specific twistability. *Biochem.* **34**, 8227-8234.

Terwilliger, T. C. (1999). SOLVE. http://www.solve.lanl.gov.

Wang, A. H.-J., Nathans, J., van der Marel, G., van Boom, J. H. & Rich, A. (1978). Molecular structure of a double helical DNA fragment intercalator complex between deoxy CpG and a terpyridine platinum compound. *Nature* **276,** 471-474.

Wang A. H. -J., Ughetto, G., Quigley, G. J. & Rich, A. (1986). Interactions of quinoxaline antibiotic and DNA: the molecular structure of a triostin A-d(GCGTACGC) complex. *J. Biomolec. Struct. Dynamics* **4**, 319-342.

Wang, A. H. -J., Ughetto, G., Quigley, G. J. & Rich, A. (1987). Interactions between an anthracycline antibiotic and DNA: Molecular structure of daunomycin complexed to d(CGTACG) at 1.2 Å resolution. *Biochemistry* **26**, 1152-1163.

Appendix

Crystallization of Small Ligand-DNA Complexes

The purpose of this appendix is to record some strategies I used for co-crystallization of DNA oligonucleotides with small synthetic ligands. In addition, there are a couple of helpful references on the subject (Wang & Gao, 1990; Dock-Bregeon & Moras, 1992). Crystallization of small molecule-DNA complexes can be reduced to a few simple points: 1) oligonucleotide length, 2) palindromic sequences, 3) two possible crystallization conditions and 4) bromination of the oligonucleotide. Of course crystallization is mostly trial and error, but well-designed experiments apparently help the chance of success.

Only a few different oligonucleotide lengths and resulting crystal forms are present for each DNA conformation in the Nucleic Acid Database (NDB) (Berman *et al.*, 1998). Consequently, the NDB serves as a useful resource when designing sequences to crystallize. For example, ten and twelve base pair lengths are common for B-form oligonucleotides. As a result of the 10_1 B-DNA screw axis (Saenger, 1984), the 10mers consistently crystallize as pseudo-continuous helices, whereas the 12mers overlap by two base pairs, with the major groove of one molecule tucked into the minor groove of the next. Accordingly, in retrospect a twelve base pair oligonucleotide was a poor choice to co-crystallize with a polyamide with an RPR-tail (Bremer *et al.*, 1998) (Table A.1). The motif of end-end helix packing is common, not only for B-form 10mers, but also for intercalator-DNA complexes. There are many examples of the latter, but with few exceptions they are all variations of the hexamer 5'-C|GTAC|G-3' (where | indicates the intercalation site) with two intercalators, one base pair from each end (for example, see Wang *et al.*, 1980). The exceptions are the structures of nogalamycin•5'-CCC|GGG-3'

(Cruse *et al.*, 1996) and actinomycin-D•5'-GAAG|CTTC-3' (Kamitori & Takusagawa, 1992). Bearing this in mind, six and eight base pair lengths were exploited to obtain co-crystals of Rh[en_2]phi•5'-CTG|CAG-3' and Rh[Me_2trien]phi•5'-GTTG|CAAC-3'. (Table A.2).

The symmetry of the sequence also affects crystallization properties. Most DNA oligonucleotides that have been successfully crystallized are self-complementary, with pseudo-twofold symmetry. Asymmetric sequences are often associated with disorder in the crystals (Nelson *et al.*, 1987; DiGabrielle *et al.*, 1989, Han *et al.*, 1998). Although I have tried to crystallize asymmetric sequences with hairpin polyamides, the rare crystals that resulted did not diffract (Table A.1). One alternative when crystallizing an asymmetric DNA binding molecule is to design the oligonucleotide to contain two binding sites related by the twofold symmetry. This was used to cocrystallize Rh[bpy_2]chysi•5'-GCTGCAAC-3' (intercalation sites either one or two base pairs from the ends of the sequence) (Table A.2). This strategy was also applied to the six-ring hairpin polyamide and 5'-UTACTAGTAA-3' (Table A.1). Unfortunately, this experiment failed, perhaps since the binding sites are somewhat too close to one another. Better results might be obtained with a polyamide with the shorter ethanol rather than Dp tail, to preserve use of a ten base pair oligonucleotide.

In addition to symmetry, there is an apparent trend to the type of base pairs at the ends of the oligonucleotide. The majority of ten base pair oligonucleotides that crystallize as A-form begin with either G or A, whereas B-form structures usually have a C or T at the 5' end. This may relate to the stacking preferences of the bases in the steps between oligonucleotides in the crystal. Furthermore, usually the 5' and 3' steps are GC base pairs.

Apart from DNA sequence, the crystallization conditions themselves are rather monotonous. Most DNA crystal grow from 2-methyl-2,4-methyl-pentanediol (MPD) and a divalent cation. For a given sequence, the conditions that I routinely screen are 20-30%

MPD in 3% increments, 10 mM buffer (tris-(hydroxymethyl)-aminomethane pH 8.0 in the presence of polyamides, sodium cacodylate with rhodium intercalators), and either 70-120 mM calcium acetate or 20 mM magnesium chloride / 1-4 mM spermine at 4° C. The concentration of the double-stranded oligonucleotide is usually 1.35 mM for polyamides with ten-mers, and 1.6-2.0 mM for intercalators with eight base pair oligonucleotides. Polyamide-DNA complexes have not crystallized in my hands in the presence of magnesium, whereas metallointercalators will crystallize with DNA under either condition. Athough the quantity and type of divalent ion is important, the main variable in the screen is the ligand:DNA ratio, which can be quite excessive in some cases. From experience, if the ligand has a binding constant less than $\sim 10^5$, it is impossible to obtain cocrystals (Table A.1).

The DNA was synthesized with the 5'-protecting group on, and after removal of base-labile protecting groups by incubation in NH_4OH at 55°C for overnight (a few days at room temperature for halogenated sequences) purified by two rounds of reverse phase on a C8 column (Pharmacia) with 40% acetonitrile / 100 mM ammonium acetate pH 7.0 as the eluant. Between chromatographic rounds, the 5'-trityl group was removed by incubating in glacial acetic acid for half an hour. Prior to crystallization, oligonucleotides were desalted using a Sep-Pac C18 cartridge (Waters), and the resulting white, fluffy material is the NH_4^+ salt. To prevent ammonia from interfering with crystallization of the rhodium intercalator, sequences intended for this use were washed with several volumes of 100 mM NaCl followed by water once bound to the C18 column. The salts of the ligands varied: polyamides were consistently the tri-fluoro-acetate (TFA) salt, whereas intercalators were either the TFA or the Cl^- salt, without effect on crystallization.

Another detail of the crystallization of metallointercalator-DNA complexes is that the crystals often take quite long to form. The setups are performed in the dark, then the trays kept in a closed cooler in the cold room for several months before 1-2 single, high quality crystals grow from crystalline clusters or oil. These crystals are highly colored by

the presence of the intercalator. Polyamide-DNA crystals, on the other hand, usually grow in 3-14 days and are colorless or pale yellow (Figure A.1).

Once crystals have been grown, use of brominated or halogenated DNA often changes the habit or diffraction qualities of the crystal. Although this is infuriating when searching for heavy atom derivatives, it can be an asset when collecting high resolution data. In retrospect, rather than multiple isomorphous replacement, the structure of ImImPyPy•5'-CCAGGCCTGG-3' in space group P2(1)2(1)2(1) (Chapter 2) could have been solved based upon the differences between a brominated sequence and an iodinated one, as was used to solve the structure of Rh[Me2trien]phi•5'-GdIUTGCAAC-3' (Chapter 5). Instead, oligonucleotides with bromination at every position were screened for crystallization to solve the ImImPyPy co-structure. Only those containing BrC or BrU crystallized, and the BrC derivatives were in several space groups, all different from that of the native crystals. The preference of BrG and BrA to have the *syn* rather than *anti-* conformation may relate to their inability to crystallize (Saenger, 1984). However, once grown, the brominated crystals did diffract to higher resolution than their parent, an observation that also held true for Rh[Me2trien]phi•5'-GdIUTGCAAC-3'. In the latter case, it was difficult to collect data on the native crystals at all due to their high mosaicity and multiple, overlapping lattices, and in fact the structure could not be solved using the native data. Thus bromination of the DNA is a strategy for improving crystal quality.

Table A.1 Other Polyamide-DNA Crystallization Trials.[1]

Sequence	Results
ImPyβImPyβImPyβDp 5'-CAGCAGCTGCTG-3'	*Standard Screen* *Results:* No crystals. Gel.
AcImPyPyPyβDp 5'-CCAGTACTGG-3	*Standard Screen* *Results*: No crystals. Precipitate.
DpβPyPyImIm 5'-CCACCGGTGG-3	*Standard Screen.* *Results:* No crystals. Oil.
DpβPyPyImIm 5'-CCAC-BrC-GGTGG-3	*Standard Screen.* *Results:* No crystals. Oil.
DpβImImPyPyAc 5'-CCAGGCCTGG-3	*CaOAc / MPD conditions.* *Results:* No crystals. Precipitate. Also tried soaking ImImPyPy-GGCC crystals in excess reverse polyamide for one week, with three changes of buffer. Still bound as a mixture of two orientations (or two polyamides).
ImPyImPyβDp 5'-CCAGCGCTGG-3	*CaOAc / MPD conditions.* *Results:* Crystals did not contain polyamide.
ImPyImPyβDp 5'-CATG-BrC-GCATG-3	*CaOAc / MPD conditions.* *Results:* Precipitate. No crystals.
ImβImPyβDp 5'-CATGCGCATG-3	*CaOAc / MPD conditions.* *Results:* Crystals did not contain polyamide.
ImHpPyPyβDp 5'-CCAGATCTGG-3 (polyamide-DNA mismatch)	*CaOAc / MPD conditions.* *Results:* Precipitate. No crystals.
ImHpPyPyβDp 5'-CATGATCATG-3 (polyamide-DNA mismatch)	*CaOAc / MPD conditions.* *Results:* Precipitate. No crystals.
ImHpPyPyβDp 5'-CCTG-BrT-ACTGG-3	*CaOAc / MPD conditions.* *Results:* Crystals did not contain polyamide.
ImPyPyPyβRPR-NH_2 5'-CGTAGATCTACG-3	*Standard Screen*, also with hexanediol and Na Cac. *Results:* Crystals (from hexanediol) did not contain polyamide. Solved by MIR (whoops!).
ImPyPyPyβDp 5'- CGTAGATCTACG-3 (test)	*CaOAc / hexanediol.* *Results:* Crystals did not contain polyamide.
ImPyPyPyβRPR-NH_2 5'-CCAGTACTGG-3	*CaOAc / MPD conditions.* *Results:* Crystals did not contain polyamide.

ImPyPyPyPyβRPR-NH_2 5'- CATGATCATG-3	*CaOAc / MPD conditions.* *Results:* No crystals. Precipitate.
ImPyPy-γ-PyPyPyPyβDp 5'-CGGTGTGTTAGG-3' 3'- CCACACAATCCG-5' (10mer, 1 base overhang)	*Standard Screen*, with more extensive screens of divalents and polyamines. *Results:* Oil. No crystals.
ImPyPy-γ-PyPyPyPyβDp 5'-CGGTGTGTTAGGC-3'	*Standard Screen*, with more extensive screens of divalents and polyamines. *Results:* Oil. No crystals.
ImPyPy-γ-PyPyPyPyβDp 5'-CGCGTGTTAGCG-3'	*Standard Screen* as in text, with more extensive screens of divalents and polyamines. *Results:* 'Crystals' from 20% MPD, 2 mM spermine, MgOAc. Fiber diffraction.
ImPyPyPy-$(R)^{H2N}$γ- PyPyPyPyPyβDp 5'-CAAAGAAAAG-3	*Standard Screen* as in text. *Results:* Precipitate. No crystals.
ImPyPyPy-γ-PyPyPyPyPyβDp 5'-CAAAGAAAAG-3	*Standard Screen* as in text. *Results:* Mechanically twinned crystals, did not diffract, 0.6 mM ds DNA, 0.66 mM polyamide; 70 mM CaOAc, 15 mM Mes pH 6.5, 27% MPD.
ImPyPyPy-γ-PyPyPyPyPyβDp 5'-CGAAGAAACG-3	*Standard Screen.* *Results:* No crystals. Precipitate.
ImPyPyPy-γ- PyPyPyPyPyβRPR-NH_2 5'-CGAAGAAACG-3	*Standard Screen.* *Results:* Precipitate, single small crystals after several months. Did not diffract.
ImPyPyPy-γ-ImPyPyPyPyβDp 5'-CCAGTACAGG-3	*CaOAc / MPD conditions.* *Results:* One crystal from precipitate. Did not diffract or was too small.
ImPyPyPy-γ- PyPyPyPyPyβRPR-NH_2 5'-CAAAGAAAAG-3'	*Standard Screen* as in text, except MgOAc in addition to $MgCl_2$, and spermidine in addition to spermine. *Also screened*: NaCac pH 6.0, Mes pH 6.5, hexanediol, Peg 1K, Peg 2K, Peg mme2K, BOG, 200 mM NH_4OAc (DNA sequence was relatively insoluble), 0.35-0.9 mM dsDNA. *Results:* Tiny crystals, 200 mM NH_4OAc / 0.08% BOG premixed with complex to promote solubility; 0.34 mM ds DNA, 0.37 mM polyamide; reservoir 80 mM $MgOAc_2$ / 30% Peg 2K / 50 mM NaCac pH 6.5

ImPyPy-γ-PyPyPyβDp 5'-UTACTAGTAA-3' (2:1)	*Standard screen.* *Results:* Tiny crystals, 20 mM CaOAc / 12% MPD premixed with complex to promote solubility; 0.7 mM dsDNA, 1.6 mM polyamide; reservoir 70 mM CaOAc / 20-23% MPD, not layered with reservoir.
ImPyPyPy-$(R)^{H2N}$γ- ImPyPyPy 5'-CCAGTACTGG-3'	*CaOAc / MPD conditions.* *Results:* Small crystals from precipitate, ~25 x 25 micron, after 2 weeks. Very mosaic, weak diffraction to 2.9 Å. Probably space group P2(1), **a**=34.4 **b**=42.3 **c**=37.2 β=93.7, 2 duplex / asu. At >25% MPD, obtained DNA only crystals as well.

[1]Crystallization trials not described in other chapters. *Standard Screen* is described in text. Abbreviations: OAc, acetate; ds, double stranded; BOG β-octyl-glucoside; Cac, cacodylate; Peg, polyethylene glycol; Mes, morpholino-ethane-sulfonic acid; $(R)^{H2N}$γ, (*R*)-2,4-diamino-butyric-acid.

Table A.2 Other Rhodium Intercalator-DNA Crystallizations.[1]

Sequence	Results
Rh[en_2]phi 5'-CTGCAG-3'	*Standard Screen* *Results:* Red crystals, P4(2)2(1)2, **a**= 52.3 **c**=57.1. Data collected at SSRL to 2.8 Å resolution. Two duplex per asymmetric unit.
Rh[en_2]phi 5'-BrC-TGCAC-3'	*Standard Screen* *Results:* Red crystals, P3(1)12 or P3(2)12, **a**=54.5 **c**=80.9. Four duplex per asymmetric unit. Data collected on rotating anode to 4.0 Å resolution.
Rh[en_2]phi 5'-GAAGCTTC-3'	*Standard Screen* *Results:* Precipitate.
Rh[en_2]phi 5'-CTTGCAAG-3'	*Standard Screen* *Results:* Pink crystals, don't diffract.
Rh[Me_2trien]phi 5'-CCTGCAGG-3'	*Standard Screen* *Results:* Precipitate.
Rh[Me_2trien]phi 5'-GGTGCACC-3'	*Standard Screen* *Results:* Precipitate.
Rh[Me_2trien]phi 5'-CATGCATG-3'	*Standard Screen* *Results:* Red crystalline spherulites.
Rh[Me_2trien]phi 5'-CT-dU-GCATG-3'	*Standard Screen* *Results:* Pink crystals, do not diffract.
Rh[Me_2trien]phi 5'-CTGCAG-3'	*Standard Screen* *Results:* Shower thin red plates.
Rh(phen)$_2$phi 5'-GTCTAGAC-3'	Purified DNA given to K. Erkkila
Rh(phen)$_2$phi 5'-CCUAGG-3'	Purified DNA given to K. Erkkila
Rh[bpy_2]phi 5'-CGGCAATTCCCG-3' (CC-mismatch)	*Standard Screen* *Results:* Red Oil
Rh[bpy_2]phi 5'-CGATCGACCG-3' (AC-mismatch)	*Standard Screen* *Results:* Red Oil
Rh[bpy_2]phi 5'-GCTGCAAC-3' (AC-mismatch)	*Standard Screen* *Results:* Red Oil after 2 weeks, single red crystals after 6 weeks, I-centered orthorhombic, **a**=54.5 **b**= 41.1 **c**=60.3. One strand per asymmetric unit, data collected to 2.7 Å on rotating anode.

References

Berman H. M., Zardecki C. & Westbrook J. (1998). The nucleic acid database: A resource for nucleic acid science. *Acta Crystallogr.* **D54**, 1095-1104.

Bremer, R. E., Baird, E. E. & Dervan, P. B. (1998). Inhibition of major-groove-binding proteins by pyrrole-imidazole polyamides with an Arg-Pro-Arg positive patch. *Chem. Biol.*, **5**, 119-133.

Cruse, W. *et al.* A Continous transition from A-DNA to B-DNA in the 1:1 Complex between Nogalamycin and the Hexamer dCCCGGG. *J. Biol. Chem.* **271**, 15558-15567 (1996).

DiGabriele, A.D., Sanderson, M.R. & Steitz, T.A. (1989). Crystal Lattice Packing Is Important in Determining the Bend of a DNA Dodecamer Containing an Adenine Tract. *Proc. Nat. Acad. Sci. USA*, **86,** 1816-1820.

Dock-Bregeon, A.-C. & Moras, D. (1992). In Crystallization of Nucleic Acids and Proteins. Ducruix, A. & Giege, R., Ed.; Oxford University, New York; pp. 152-160.

Han, G.W., Kopka, M.L., Cascio, D., Grzeskowiak, K. & Dickerson R.E. (1997). Structure of a DNA Analog of the Primer for HIV-1 Second Strand Synthesis. *J. Mol. Biol.* **269**, 811-826.

Kamitori, S. & Takusagawa, F. (1992). Crystal structure of the 2:1 complex between d(GAAGCTTC) and the anticancer drug actinomycin D. *J. Mol. Biol.* **225**, 445-456.

Nelson, H.C.M., Finch, J.T., Luisi, B.F. & Klug, A. (1987). The Structure of an Oligo(dA).Oligo(dT) Tract and Its Biological Implications. *Nature.* **330**, 221-226.

Saenger, W. (1984) *Principles of Nucleic Acid Structure.* (Springer-Verlag, New York, NY).

Wang, A.H.-J., Ughetto, G., Quigley, G.J. & Rich, A. (1987). Interactions between an Anthracycline Antibiotic and DNA Molecular Stucture of Daunomycin Complexed to d(CpGpTpApCpG) at 1.2-Angstroms Resolution. *Biochemistry*, **26,** 1152-1163.

Wang, A. H. -J. & Gao, Y.-G. (1990). Crystallization of oligonucleotides and their complexes with antitumor drugs. *Methods.* **1**, 91-99.

Figure A.1 Examples crystals of small ligands with DNA.

(a) ImImPyPyβDp•5'-CCAGGCCTGG-3'

(b) Rh[Me_2trien]phi•5'-GTTGCAAC-3'.