Chapter 1

Introduction to DNA Recognition By Minor Groove-Binding Polyamides

Background and Significance

The complete set of biochemical instructions required for the sustenance of life is contained within the genomes of all known organisms.[1] The genomes of complex organisms are arranged into discrete linear sequences of nucleic acids (*genes*) which constitute blueprints for the organism's proteins. The faithful replication and expression of this genetic material direct the growth, assembly, and function of the cellular machinery. The central role played by DNA in biological systems has made it a long-standing target for the diagnosis and treatment of human illness. The completion of the massive international initiative to sequence the human genome, generating some 30,000-40,000 putative genes, highlights the potential utility of DNA recognition in the molecular life sciences as well as in pharmaceutical discovery.[2,3] Understanding the functional significance of genome sequence variation and organization, as well as the complex molecular interplay between various genes and gene products, remain central questions for human medicine. Designed small molecules capable of sequence specific recognition of predetermined DNA sequences are powerful tools that can be directed toward interpreting the human genome, and these molecules may ultimately prove valuable as therapeutic agents.[4]

Structural Features of DNA

Double helical DNA consists of two complementary, antiparallel polydeoxyribonucleotide strands associated by specific hydrogen bonding interactions between nucleotide bases.[5] The sugar phosphate backbone of paired

strands defines the helical grooves, within which the edges of the heterocyclic bases are exposed (Figure 1.1A). The biologically relevant B-form structure of the DNA

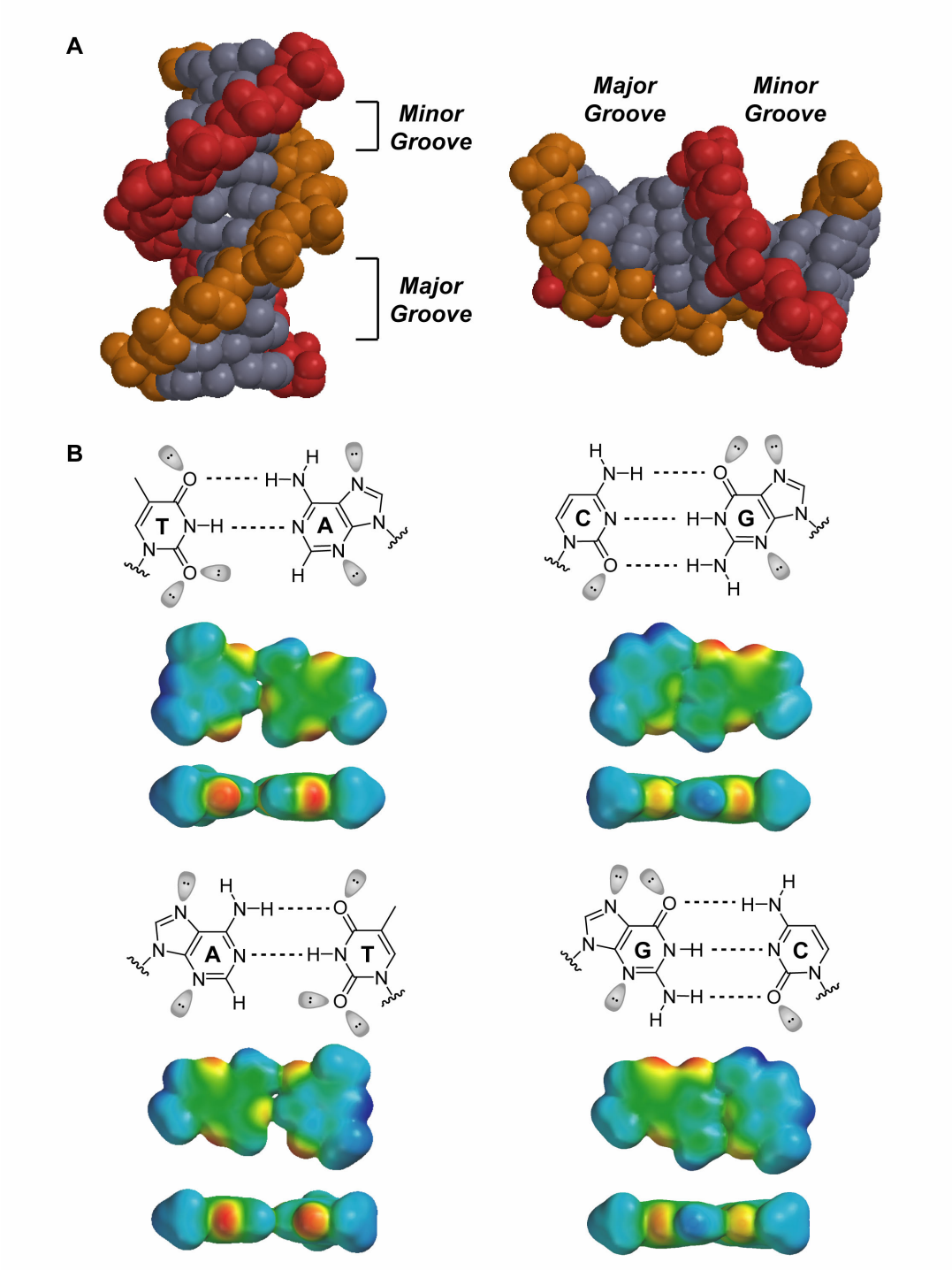

Figure 1.1 Structural features of the DNA double helix. **(A)** Molecular rendering of B-form DNA double helix. Sugar phosphate backbones of complementary nucleotide strands are shown in red and orange. **(B)** Chemical features and electropotential surfaces of Watson-Crick base pairings.

double helix is characterized by a shallow, wide major groove and a deep, narrow minor groove.[6] The chemical features of the molecular surfaces presented by a given DNA sequence in either groove are distinct, forming the basis for molecular recognition by small molecules and proteins (Figure 1.1B).

Nature's Solutions to DNA Recognition

Nature employs a diverse array of structural motifs for DNA recognition by proteins, using combinations of electrostatic interactions with the sugar phosphate backbone and van der Waals contacts with the nucleobases within the helical grooves to facilitate specificity (Figure 1.2A).[7-9] The structures of DNA-binding proteins, while complex in their own right, are often altered by association with their target DNA sequence. Complex formation often involves more than one protein component and can result in gross distortions of the canonical B-form double helix.[9] Attempts to engineer proteins with novel DNA binding properties constitute an active area of research in biochemistry; however, to date a general recognition code correlating amino acid composition with target DNA sequence has not been reported.[10] While a majority of DNA-binding proteins rely on major groove contacts for specificity, nature has also evolved a set of small molecules that target the minor groove.[11-13] Among these natural products, netropsin and distamycin are particularly attractive lead compounds owing to their modular nature, reduced size, and diminished conformational freedom with respect to native proteins. The amenability of these *polyamides* to synthetic manipulation has allowed the physical principles governing their interaction with the minor groove to be investigated in great detail.[14]

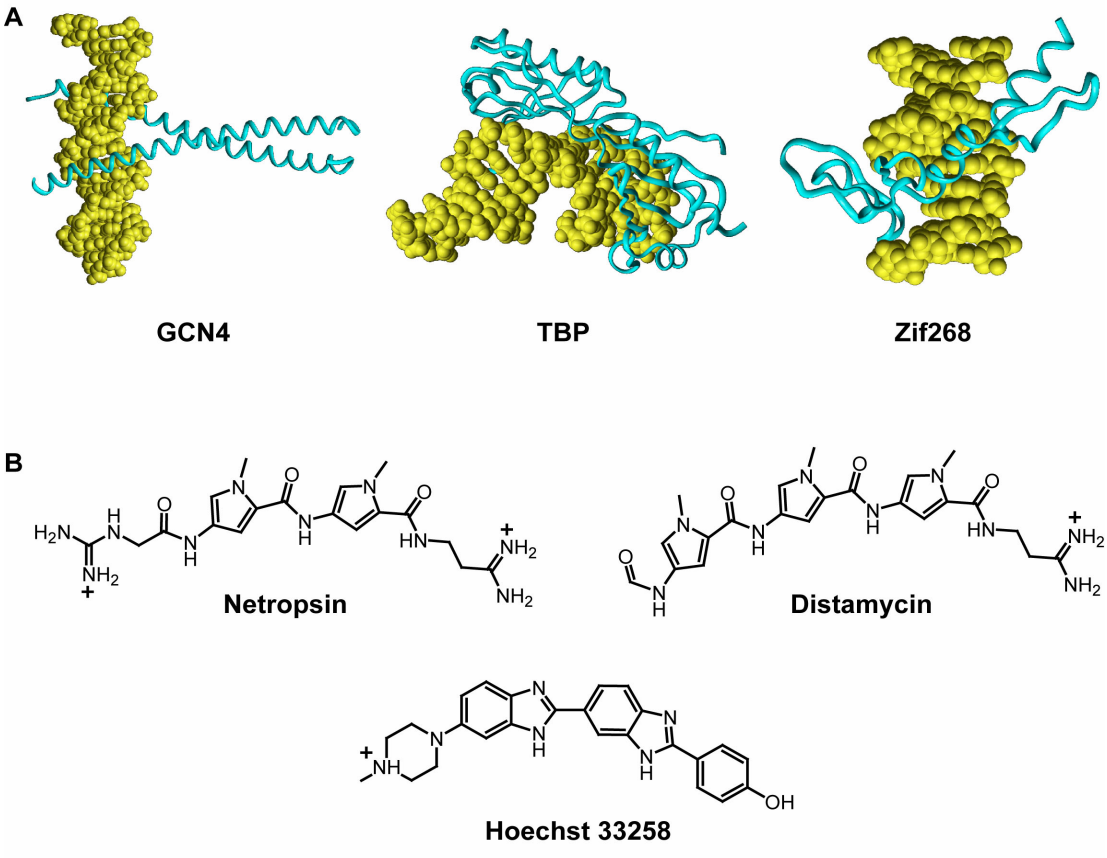

Figure 1.2 Naturally occurring ligands for DNA. **(A)** X-ray crystal structures of protein-DNA complexes. DNA-binding proteins use major (GCN4, Zif268) and minor (TBP) groove contacts to recognize their target sequences**. (B)** Crescent-shaped natural products known to bind the minor groove of DNA**.**

Designed Minor Groove Recognition

More than two decades of research in the laboratories of Prof. Peter B. Dervan have engendered a new paradigm for sequence specific recognition of the DNA minor groove.[4,15-17] DNA-binding polyamides composed of N-methylimidazole (Im), N-methylpyrrole (Py), and N-methyl-3-hydroxypyrrole (Hp) are crescent-shaped molecules that bind the minor groove as antiparallel dimers. DNA association is driven by a combination of van der Waals and hydrogen bonding interactions. Side-by-side pairings of aromatic residues stack five-membered heterocycles against

each other and the walls of the minor groove, positioning the polyamide backbone and aromatic 3-substituents for intimate contacts with the edges of nucleotide bases on the adjacent DNA strand. The cofacial arrangement of side-by-side pairings exploits the specific pattern of hydrogen bond donors and acceptors, as well as subtle variations in the molecular shape of the minor groove floor to distinguish the Watson-Crick base pairs (Figure 1.3A).[18]

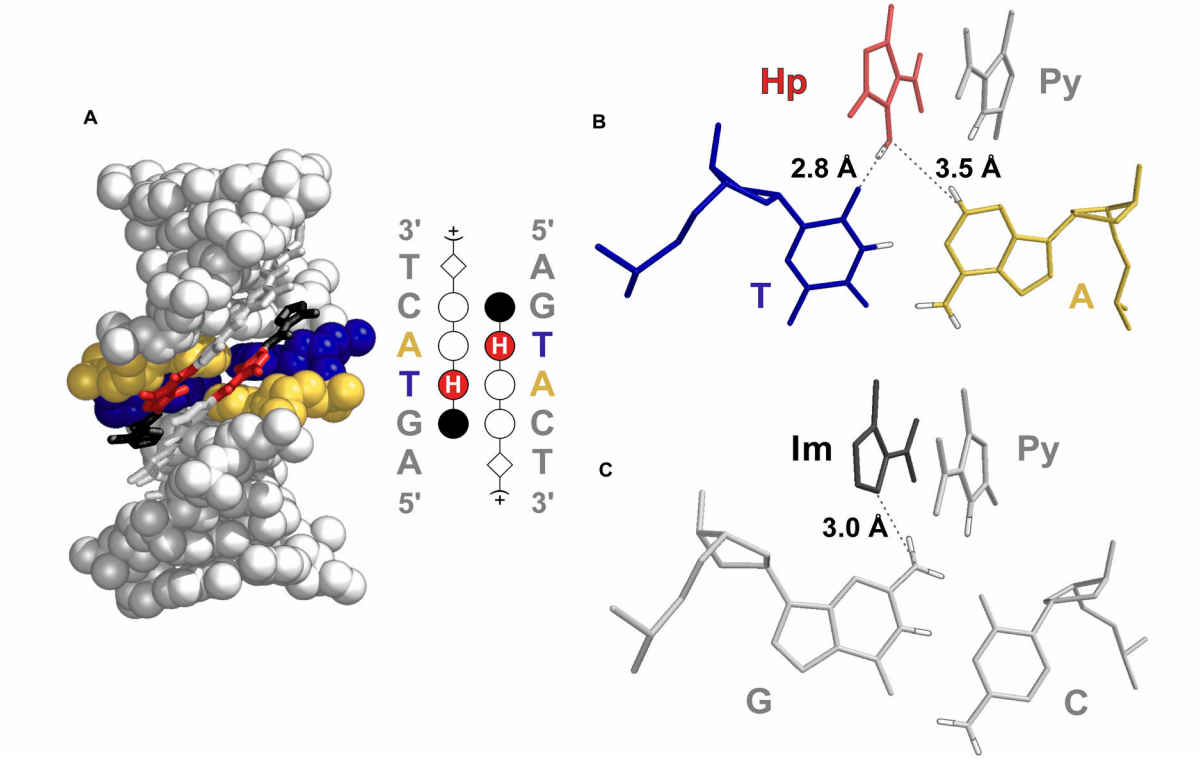

Figure 1.3 The molecular details of minor groove recognition by polyamides. **(A)** Spacefilling and ball-and-stick representation of polyamide homodimer bound in the minor groove. **(B)** Physical basis for selectivity of Hp/Py pairings. **(C)** Physical basis for selectivity in Im/Py pairings.

Unsymmetrical pairings of Im and Hp with Py underlie the sequence specificity of minor groove-binding polyamides.[19] A pairing of Im opposite Py (Im/Py) targets a G•C base pair, while Py/Im targets C•G.[20] The physical basis for the selectivity of these pairings is a linear hydrogen bond formed between the N3 of Im

and the exocyclic amine of guanine (Figure 1.3C).[21] A Py/Py pairing is degenerate for A•T and T•A, owing to the similarity of these bases in the minor groove. An Hp/Py pairing targets T•A, while Py/Hp targets A•T. The selectivity of Hp is derived from a combination of hydrogen bonding with the thymine O2 and shape-selective discrimination of the asymmetric cleft created by the bulkier adenine ring of a T•A base pair (Figure 1.3B).[18,22,23] The molecular details of DNA recognition by minor groove binding polyamides, collectively referred to as the *pairing rules*, have been validated by exhaustive physical studies using a range of techniques, including DNAse I footprinting, X-ray crystallography, and multidimensional NMR spectroscopy.

The Hairpin Motif

Covalent linkage of two antiparallel polyamide strands reduces the entropic costs of DNA association, giving rise to designed ligands with affinities and specificities rivaling naturally occurring proteins.[24,25] The hairpin motif connects the N-terminus of one polyamide strand with the C-terminus of a second strand through an aliphatic γ-aminobutyric acid residue (γ) (Figure 1.4). Hairpin polyamides usually exhibit the same orientational preference as unlinked dimers, aligning N$\rightarrow$C with respect to the 5'$\rightarrow$3' direction of the adjacent DNA strand.[26,27] Some hairpins have been observed to bind DNA in a reverse orientation and others have demonstrated the propensity to unfold, binding DNA as a single extended strand; however, both of these issues can be resolved by introduction of a chiral amino group at the α carbon of the γ *turn* residue.[28] Both chiral and achiral turn residues are marked by

pronounced selectivities for A•T and T•A base pairs, presumably owing to potential steric clashes with the exocyclic amine of guanine.[25,28]

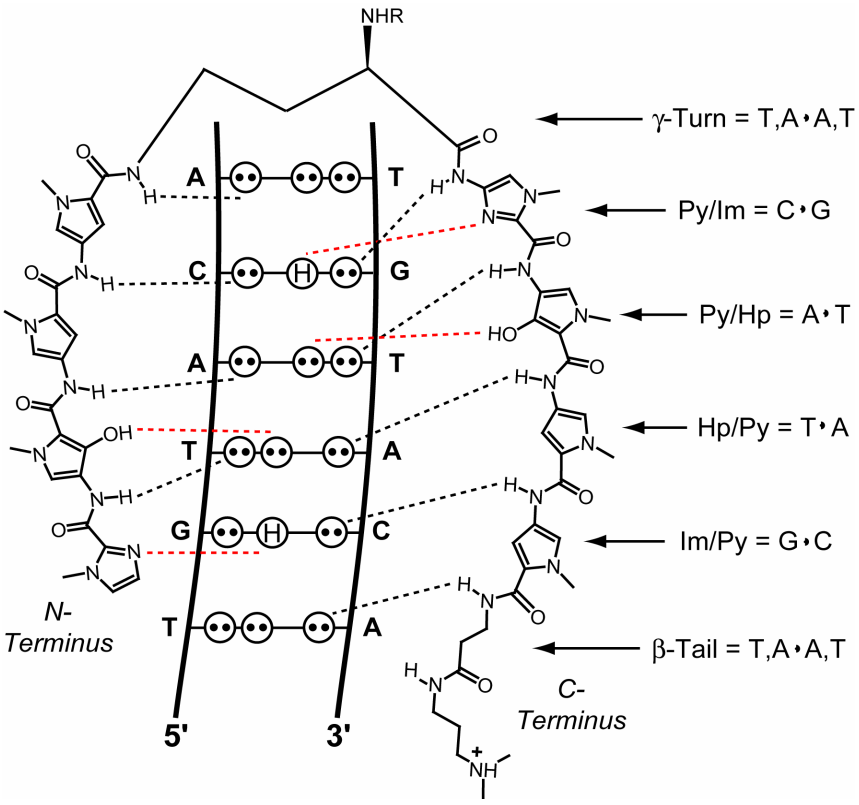

Figure 1.4 Hydrogen-bonding model of the hairpin motif illustrating the pairing rules.

A similar explanation can be evoked for the T, A selectivity of aliphatic β-alanine (β) and N,N-dimethylaminopropylamine (Dp) residues frequently found as the C-terminal *tail* of hairpin polyamides. These tail residues are not critical to DNA recognition by hairpins, and can be replaced with other residues or removed altogether.[29] In fact, alternate tails are often more beneficial with regard to cellular uptake and binding affinities. β has proven more useful in the context of *internal* pairings opposite Im, Py, and other β residues.[30,31] While the rise per residue of

polyamides correlates well with the pitch of the B-form DNA helix, they are inherently more curved than the minor groove, and beyond five consecutive aromatic rings, the shape of the polyamide is no longer complementary to DNA.[32] Internal β residues relax the curvature of polyamide strands, allowing longer DNA sequences to be targeted.

The affinity of minor groove-binding polyamides for predetermined DNA sequences often allows them to interfere with the association of transcription factors with their binding sites in promoters, leading to inhibition of transcription. A range of transcription factors and promoters have been successfully targeted by hairpin polyamides, including the TFIIIA zinc finger[33,34] and the HIV-1[35,36] promoter (Figure 1.5A). Despite the successes of the hairpin motif, the array of sequences it can be used to target is limited by the T, A selectivity of the aliphatic turn moiety. Other polyamide motifs, employing different strategies for covalent linkage of polyamide strands, have been developed, including cycles[37], H-pins[38], and U-pins (Figure 1.5B).

Limitations of Classic Minor Groove Recognition

The ability of minor groove-binding polyamides to target predetermined DNA sequences is hindered by the context dependence of the thymine-selective Hp residue. The context of an aromatic residue within a given polyamide can have pronounced effects on the specificity of its interaction with DNA. In particular, the absence of a second groove-anchoring carboxamide moiety in *N-terminal* residues affords them more conformational freedom than their *internal* counterparts. The

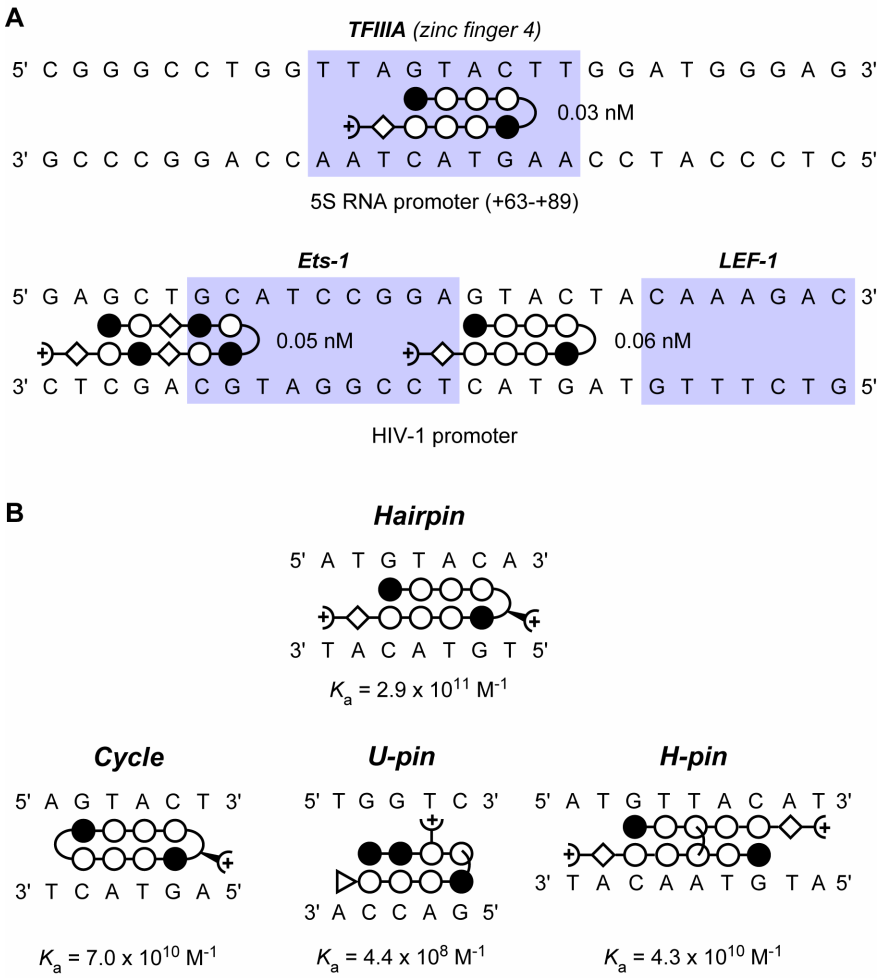

Figure 1.5 Biological applications and implications for polyamide motifs. **(A)** Representative examples of hairpin polyamide applications in the control of gene expression. Transcription factor binding sites are shown in shaded boxes. **(B)** Ball-and-stick schematic for covalently linked polyamide motifs.

specificity of Im residues is comparable in either context; however, the specificity of Hp residues is greatly reduced in N-terminal pairings. Additionally, Hp is prone to acid- and radical-catalyzed decomposition in solution and Hp-containing polyamides often exhibit diminished binding affinities.

Novel Heterocycles for Recognition of DNA

The shortcomings of Hp, in both terminal and internal contexts, have inspired the search for new thymine-selective heterocyclic residues. The reduced specificity of N-terminal Hp/Py pairings has been attributed to rotation of the terminal aromatic ring, orienting the hydroxyl recognition element away from the minor groove floor.[39] This undesirable conformation is presumably stabilized by intramolecular hydrogen bonding with the proximal 2-carboxamide oxygen (Figure 1.6). Efforts to constrain the torsional freedom of terminal Hp residues by installing acetyl or formyl groups at the 4-position of the aromatic ring have proven unsuccessful, suggesting that 3-

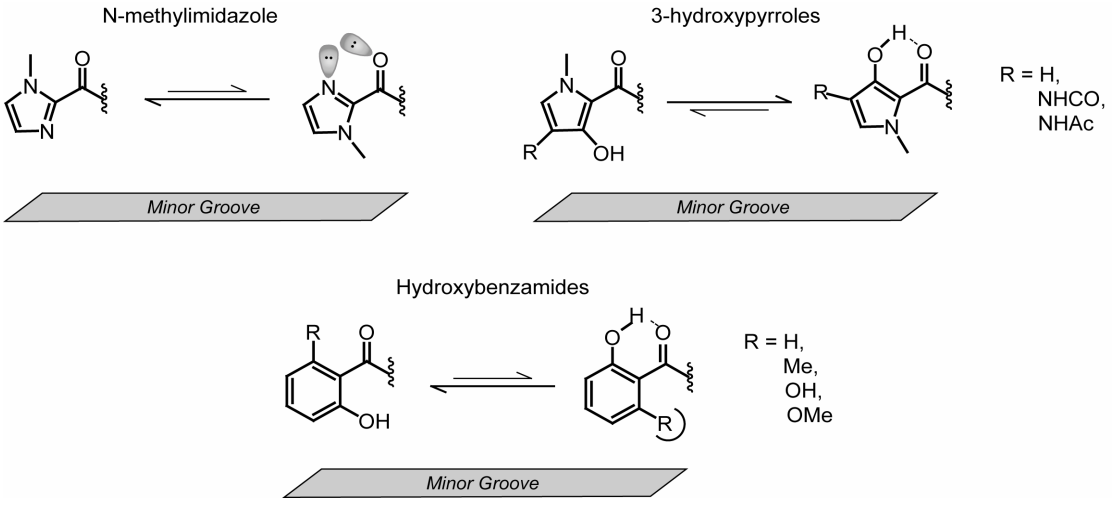

Figure 1.6 N-terminal aromatic residues for specific DNA recognition.

hydroxy-substituted five-member heterocycles are not optimal for recognition in terminal pairings.

Early work in the development of minor groove-binding polyamides demonstrated that N-terminal pairings of pyridine-2-carboxamide opposite Py could target G•C base pairs.[40] The capacity of both five-member and six-member

aromatic scaffolds to serve as guanine-selective residues suggested that hydroxybenzamides might serve as terminal Hp replacements. Towards this end, the specificity of a series of 6-substituted-2-hydroxybenzamide residues was investigated in the context of a terminal pairing opposite Py in a hairpin polyamide. Potential steric clashes between the 6-substituent and the floor of the minor groove were envisioned as a means of forcing the hydroxyl entity into the minor groove. While these new rings did not exhibit the degree of thymine-specificity shown by internal Hp residues, a 2-hydroxy-6-methoxybenzamide residue was found to have a modest preference for T•A base pairs at subnanomolar concentrations (Figure 1.6).[40] Collectively, the body of work on N-terminal pairings seems to indicate that hydroxyl recognition elements are not optimal for terminal discrimination of thymine.

While the utility of the Hp/Py pairing is limited, the physical principles governing its T•A selectivity have provided valuable insight for the design of novel, internal thymine-specific residues. The feasibility of purely shape-selective recognition of the T•A asymmetric cleft has been explored with a family of five-membered aromatic residues designed to present different molecular surfaces toward the minor groove floor (Figure 1.7).[41,42] None of these heterocycles rivaled

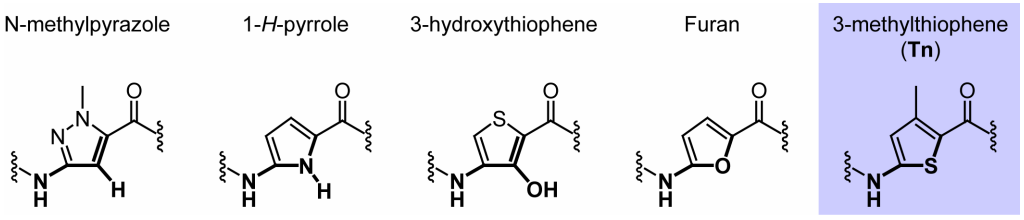

Figure 1.7 Novel heterocycles for thymine-selective recognition of the minor groove. The molecular surface presented to the minor groove floor is shown in bold.

the specificity of Hp/Py pairings, however, 3-methylthiophene (Tn) paired opposite Py was somewhat selective for T•A relative to A•T and showed no binding to C, G base pairs. The selectivity of Tn has been attributed to the preference of the polarizable sulfur atom of thiophene to lie over the smaller thymine ring of a T•A base pair.

A more radical approach to internal Hp replacements used a benzimidazole ring system, reminiscent of Hoechst subunits, as a scaffold for thymine selectivity (Figure 1.8).[43,44] These novel residues preserve the hydrogen bonding capacity and overall molecular shape of classic polyamides. The absence of an intervening amide bond between a benzimidazole residue and the adjacent heterocycle accounts for subtle changes in the curvature of a polyamide strand. This variation seems to be tolerated in the context of hairpins, though, as hydroxybenzimidazole (Hz) and imidazopyridine (Ip) residues paired opposite Py have specificities comparable to Hp/Py and Im/Py, respectively. Importantly, the phenol recognition element of Hz is chemically robust, making it a strong candidate for replacement of Hp.

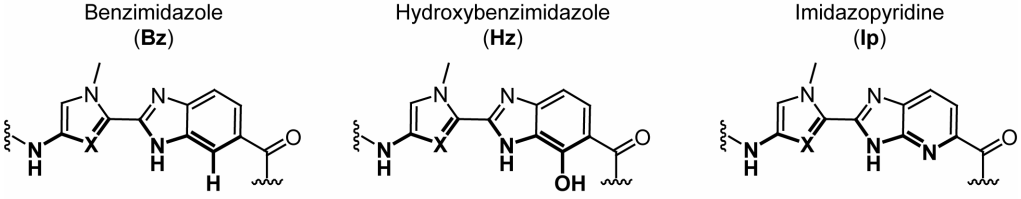

Figure 1.8 Bicyclic aromatic heterocycles for minor groove discrimination. Molecular surfaces proposed to interact with the minor groove floor are shown in bold.

Scope of this Work

This dissertation describes work focused on the application of polyamide conjugates, as well as efforts to expand the repertoire of predetermined DNA sequences amenable to recognition of by minor groove-binding polyamides. Chapter 2 discusses the design and synthesis of fluorophore-polyamide conjugates and the building blocks required for their construction. Chapter 3 presents self-quenched fluorophore-polyamide conjugates as sequence specific probes for DNA. Chapter 3A summarizes the spectroscopic and DNA recognition properties of this class of polyamides, while Chapter 3B suggests potential applications of these molecules in the context of larger DNA fragments. Chapter 4 presents another class of hairpin conjugates, containing cyanine fluorophores. These conjugates exhibit fluorescence resonance energy transfer (FRET) when bound to proximal sites in the minor groove.

Chapters 5, 6, and 7 focus on more fundamental studies of molecular recognition by polyamides in the minor groove. Chapter 5 discusses efforts to expand the sequence specificity of N-terminal pairings within the hairpin motif. Chapter 5A introduces N-terminal, 3-substituted-thiophene-2-carboxamide residues capable of shape-selective recognition of T•A base pairs. Chapter 5B describes the DNA-binding properties of other N-terminal, heterocyclic residues. The cellular localization of polyamides containing N-terminal thiophene residues, as well as examples of their use in biological settings are also presented in Chapter 5B. Chapter 6 compares the abilities of 3-methylthiophene (Tn) and hydroxybenzimidazole (Hz) residues to target adjacent T•A / A•T base pairs within

the same hairpin polyamide. Chapter 7 describes novel 3-chlorothiophene-hydroxybenzimidazole dimers capable of targeting adjacent thymine bases at the N-terminus of hairpin polyamides.

References

1) Brown, T. A. *Genomes*; Wiley-Liss: New York, 1999.

2) Consortium, I. H. G. S. *Nature* **2001**, *409*, 860.

3) Venter, J. C. et. al. *Science* **2001**, *291*, 1304.

4) Gottesfeld, J. M.; Neely, L.; Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1997**, *387*, 202.

5) Dickerson, R. E.; Drew, H. R.; Conner, B. N.; Wing, M.; Fratini, A. V.; Kopka, M. L. *Science*, **1982**, *216*, 475.

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

7) Pabo, C. O.; Sauer, R. T. *Annu. Rev. Biochem.* **1992**, *61*, 1053.

8) Pavletich, N. P.; Pabo, C. O. *Science* **1991**, *252*, 809.

9) Kim, Y.; Geiger, J. H.; Hahn, S.; Sigler, P. B. *Nature* **1993**, *365*, 512.

10) Kissinger, C. R.; Liu, B.; Martin-Blancco, E.; Kornberg, T. B.; Pabo, C. O. *Cell* **1990**, *63*, 579.

11) Gao, X. L.; Mirau, P.; Patel, D. J. *J. Mol. Biol.* **1992**, *223*, 259.

12) Kamitori, S.; Takusagawa, F. *J. Mol. Biol.* **1992**, *225*, 445.

13) Guan, Y.; Sakai, R.; Rinehart, K. L.; Wang, A. H.-L. *J. Biomol. Struct. Dyn.* **1993**, *10*, 793.

14) Bailly, C.; Chaires, J. B. *Bioconj. Chem.* **1998**, *9*, 513.

15) Dervan, P. B. *Bioorg. Med. Chem.* **2001**, *9*, 2215.

16) Dervan, P. B.; Bürli, R. W. *Curr. Opin. Chem. Biol.* **1999**, *3*, 688.

17) Dervan, P. B.; Edelson, B. S. *Curr. Opin. Struct. Biol.* **2003**, *13*, 284.

18) White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Deran, P. B. *Nature* **1998**, *391*, 468.

19) Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783.

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

21) Kielkopf, C. L.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Nat. Struct. Biol.* **1998**, *5*, 104.

22) Kielkopf, C. L.; White, S.; Szewczyk, J. W.; Turner, J. W., Baird, E. E.; Dervan, P. B. *Science* **1998**, *282*, 111.

23) Kielkopf, C. L.; Bremer, R. E.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *J. Mol. Biol.* **2000**, *295*, 557.

24) Mrksich, M.; Parks, M. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 7983.

25) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1996**, *382*, 559.

26) deClairac, R. P. L.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1997**, *119*, 7909.

27) White, S.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 8756.

28) Herman, D. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 1382.

29) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 1113.

30) Turner, J. M.; Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 6219.

31) Wang, C. C. C.; Ellervik, U.; Dervan, P. B. *Bioorg. Med. Chem.* **2001**, *9*, 653.

32) Kelly, J. J.; Baird, E. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 6981.

33) Gottesfeld, J. M.; Neely, L.; Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1997**, *387*, 202.

34) Neely, L.; Trauger, J. W.; Baird, E. E.; Dervan, P. B.; Gottesfeld, J. M. *J. Mol. Biol.* **1997**, *274*, 439.

35) Dickinson, L. A.; Gulizia, R. J.; Trauger, J. W.; Baird, E. E.; Mosier, D. E.; Gottesfeld, J. M.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 12890.

36) Dickinson, L. A.; Trauger, J. W.; Baird, E. E.; Dervan, P. B.; Graves, B. J.; Gottesfeld, J. M. *J. Biol. Chem.* **1999**, *274*, 12765.

37) Herman, D. M.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 1121.

38) Olenyuk, B.; Jitianu, C.; Dervan, P. B. *J. Am. Chem. Soc.* **2003**, *125*, 4741.

39) Ellervik, U.; Wang, C. C. C.; Dervan, P. B. *J. Am. Chem. Soc.* **2000**, *122*, 9354.

40) Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783.

41) Nguyen, D. H.; Szewczyk, J. W.; Baird, E. E.; Dervan, P. B. *Bioorg. Med. Chem.* **2001**, *9*, 7.

42) Marques, M. A.; Doss, R. M.; Urbach, A. R.; Dervan, P. B. *Helv. Chim. Acta.* **2002**, *85*, 4485.

43) Minehan, T. G.; Gottwald, K.; Dervan, P. B. *Helv. Chim. Acta.* **2000**, *83*, 2197.

44) Renneberg, D.; Dervan, P. B. *J. Am. Chem. Soc.* **2003**, *125*, 5707.