

# **Chapter 1**

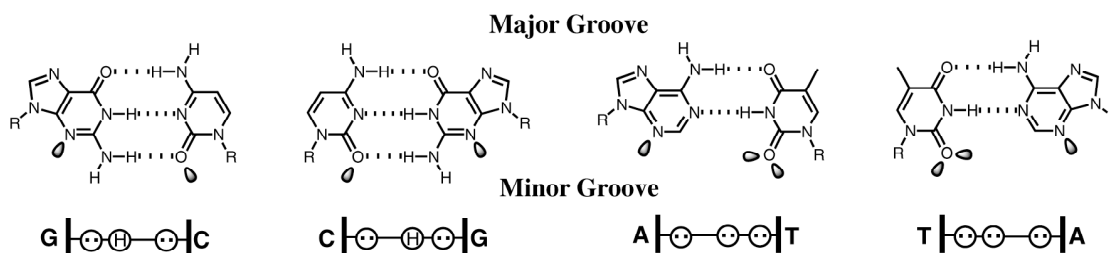
## **Introduction**

## Background

DNA is the basis for life.<sup>1</sup> Composed of deoxyribose-phosphate polymers containing any of the four heterocyclic bases, adenine (A), thymidine (T), guanine (G), and cytosine (C), DNA is a self-organized chemical code. Through inter-strand hydrogen bonds, antiparallel strands of polymer are organized to form a higher helical structure.<sup>2</sup> The linear arrangement of the bases provides chemical instructions in the form of genes. There are approximately 30,000-40,000 genes in man.<sup>3,4</sup> Understanding the complex relationship between DNA (gene expression) and well-being is of singular importance to molecular biology and human medicine. Aberrant gene expression, either in the form of deleterious products or abnormal regulation, is responsible for a variety of disease states.<sup>5</sup> Thus there is interest, particularly in the area of small molecules, to design compounds to regulate and study the process of gene expression and organization.<sup>6</sup>

## DNA recognition

Canonical B-form of DNA is characterized by a wide (12 Å) and shallow major groove and a narrow (4-6 Å) and deep minor groove.<sup>7</sup> Sequence-dependent structural variations, conformational properties, solvent, and counterion organization all contribute to structural variation from the canonical form.<sup>7</sup> The chemical distinction between the



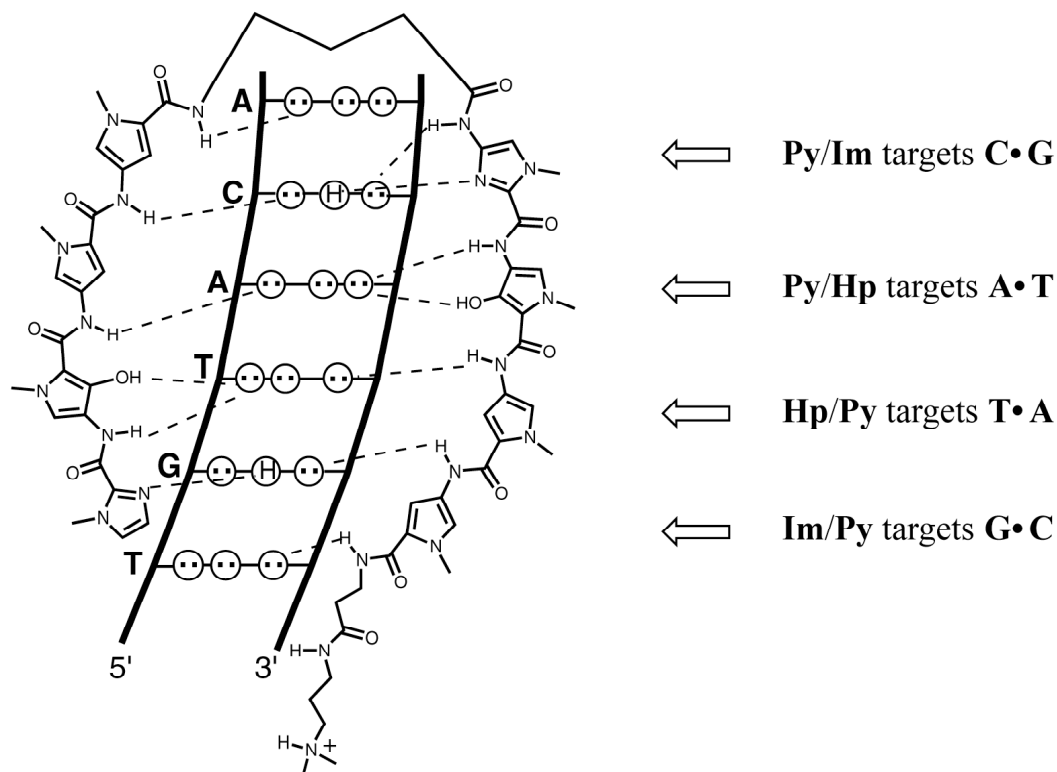
**Figure 1.1** The edges of the base pairs present different hydrogen bond donor and acceptor combinations to the major and minor grooves in the DNA double helix. The top edge of the bases, as drawn, projects into the major groove, and the bottom projects into the minor groove. Below the base pairs, the minor groove is shown in schematic representation, where circles with dots represent lone pairs on N(3) of purines (A,G) and O(2) of pyrimidines (C,T), and circles containing an H represent the 2-amino group of G.

DNA bases is the pattern of hydrogen bond donors and acceptors (heteroatoms) present at the edges of the base pairs (Figure 1.1). Proteins that recognize DNA take advantage of specific hydrogen bonding, van der Waals, and electrostatic interactions to sequence specifically bind to biologically important DNA sequences. Thus the arrangement of the bases relates to the phrasing of chemical features in the DNA major and minor grooves that DNA binding proteins “read.”<sup>8</sup> The zinc finger<sup>9</sup> and leucine zipper<sup>10</sup> motifs represent some of the better known structural motifs for DNA recognition. No single motif exists that represents a general amino acid•base pair code for all DNA sequences.<sup>11</sup> Recognition by some proteins, such as the ubiquitous DNA-bending transcription factor TBP,<sup>12</sup> involves large distortions of the target DNA from its common B-form structure.

A survey of the field of DNA recognition by small molecules provides many natural products that recognize DNA, with varying degrees of sequence specificity, by diverse binding modes that include groove binding and intercalation.<sup>13-15</sup> The minor groove is a particular rich environment for recognition by small molecules (Figure 1.1).<sup>16,17</sup> Many of these DNA recognizing compounds, such as the oligosaccharide calicheamicin,<sup>18,19</sup> CC-1065,<sup>20,21</sup> and Hoechst 33258,<sup>22,23</sup> have allowed chemists to modularly synthesize numerous derivatives in order to understand and modify their sequence specificity and binding affinity. Distamycin,<sup>24</sup> which binds a five base pair A,T sequence in the minor groove, has proven particularly well suited for modifications leading to rationally controlled sequence specificity.<sup>22,25-28</sup>

## The polyamide scaffold for sequence specific molecular recognition in the minor groove

Research in the laboratories of Professor Peter B. Dervan has led to the development of minor groove binding distamycin derivatives that can recognize a large number of predetermined DNA sequences with affinities that rival natural transcription factors.<sup>6,29,30</sup> DNA binding polyamides contain *N*-methylpyrrole (Py) carboxamides, the structural repeat of distamycin, as well as *N*-methylimidazole (Im) and *N*-methyl-3-hydroxypyrrole (Hp) carboxamides. DNA recognition depends on side-by-side amino acid pairings in the minor groove that stack the aromatic rings against each other and the walls of the groove allowing backbone amide hydrogens and the substituents at the 3-



**Figure 1.2** The pairing rules for polyamide recognition in the minor groove.

position of the Py, Im, and Hp residues to make specific contacts with the edges of the DNA bases. Each pair of polyamide residues is selective for a specific DNA base pair based on steric factors and the matching of hydrogen bond donor and acceptor functionalities (Figure 1.2). A pairing of Im opposite Py (Im/Py) targets a G•C base pair,

while Py/Im targets C•G. A Py/Py pairing is degenerate, targeting both A•T and T•A base pairs. An Hp opposite a Py (Hp/Py) discriminates T•A from A•T, while Py/Hp targets A•T in preference to T•A and both of these from G•C and C•G. Footprinting, NMR, and X-ray structure studies validate these pairing rules for DNA minor groove recognition (Figure 1.2).<sup>29-31</sup>

Though many different recognition motifs based on pyrrole-imidazole polyamides are possible, the design known as the hairpin (Figure 1.2) has been extensively studied.<sup>32</sup> The peptide strands of Py/Im aromatic amino acids are connected by an alkyl amino acid—either  $\beta$ -aminobutyric acid ( $\beta$ ), or the chiral, amine-functionalized derivative (*R*)-2,4,-diaminobutyric acid ((*R*)<sup>H<sub>2</sub>N</sup> $\beta$ )-in an antiparallel orientation relative to each other and, generally, a 5'  $\rightarrow$  3', N  $\rightarrow$  C orientation relative to the phosphate backbone of DNA. The linker between Py/Im strands is considered the “turn” of the hairpin, and alkyl units  $\beta$ -alanine ( $\beta$ ) and *N,N*-dimethylaminopropylamine (Dp) are typically used to form a C-terminal “tail.” All of these alkyl units are specific for both A,T base pairs.  $\beta$ -alanine ( $\beta$ ) can also be used internally to form  $\beta/\beta$ ,  $\beta$ /Py, and  $\beta$ /Im pairs, which has extended the binding site size amenable to recognition by the hairpin polyamides and in certain cases results in increased affinity and specificity. Tandem hairpin dimers have also been produced which bind with high affinity and specificity.<sup>34,35</sup> The hairpin motif is also a versatile scaffold for the production of conjugates with diverse functions including sequence-specific DNA-alkylation,<sup>36,37</sup> recruitment of Topoisomerase I,<sup>38</sup> and delivery of fluorescent labels.<sup>39-41</sup>

### Scope of this work

Conceptually the division of this thesis is into three parts, one biological, one nucleic acids recognition chemistry, and one spectroscopic. In the first part, Chapter 2, we collaborate with Professor David Margolis at the University of Texas, Southwestern Medical Center, to explore polyamide interactions with the repressor complex sequence

(RCS) of the long-terminal repeat (LTR) of the integrated HIV-1 gene. We demonstrate, in Chapter 2, that polyamides affect HIV-1 gene expression levels by displacing endogenous repressor transcription factors that downregulate LTR expression and lead to latent HIV-1 infection. This evidence in Chapter 2 supports the role of the host transcription factors LSF and YY1 in regulating HIV-1 post-infection latency. In Chapter 3, we explore binding orientation within the 2- $\beta$ -2 polyamide scaffold, often observing a “reversed” orientation for polyamide binding (C $\rightarrow$  N, 5' $\rightarrow$  3'), and we report the first use of (S)-2,4-diamino butyric acid to successfully influence polyamide specificity and orientation when bound in the minor groove. Chapters 4 and 5 are devoted to the discussion of polyamides that act as “light-up” probes for the fluorescent detection of DNA in homogenous solution. By quenching the fluorescence from the xanthene fluorophore to which it is attached, and only allowing fluorescence when the polyamide binds in the minor groove, we report a new method for sequence specific fluorescent detection of native DNA. Chapter 5, the fruits of a collaboration with Alexander Dunn of Professor Harry Gray’s group, is devoted to spectroscopic studies to understand the chemistry responsible for polyamide quenching of fluorescence and sequence specific increase in quantum yield upon polyamide binding in the minor groove. We postulate that quenching arises from a non-specific intramolecular hydrophobic collapse between the polyamide and the fluorophore, and we propose an excited-state charge-transfer reduction of the fluorophore singlet excited state as the likely mechanism for electronic quenching.

## References

1. Lewin, B. *Genes V*; Oxford University Press: New York, 1994.
2. Dickerson, R. E.; Drew, H. R.; Conner, B. N.; Wing, M.; Fratini, A. V.; Kopka, M. L. *Science* **1982**, *216*, 475.
3. Consortium, I. H. G. S. *Nature* **2001**, *409*, 860.
4. Venter, J. C. et al. *Science* **2001**, *291*, 1304.
5. Tjian, R. *Sci. Am.* **1995**, *2*, 54.
6. Gottesfeld, J. M.; Neely, L.; Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1997**, *387*, 202-205.
7. Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.
8. Pabo, C. O.; Sauer, R. T. *Annu. Rev. Biochem.* **1992**, *61*, 1053-1095.
9. Pavletich, N. P.; Pabo, C. O. *Science* **1991**, *252*, 809.
10. Ellenberger, T. E.; Brandl, C. J.; Struhl, K.; Harrison, S. C. *Cell* **1992**, *71*, 1223.
11. Kissinger, C. R.; Liu, B.; Martin-Blanco, E.; Kornberg, T. B.; Pabo, C. O. *Cell* **1990**, *63*, 579.
12. Kim, Y.; Geiger, J. H.; Hahn, S.; Sigler, P. B. *Nature* **1993**, *365*, 512-520.
13. Gao, X. L.; Mirau, P.; Patel, D. J. *J. Mol. Biol.* **1992**, *223*, 259-279.
14. Kamitori, S.; Takusagawa, F. *J. Mol. Biol.* **1992**, *225*, 445-456.
15. Guan, Y.; Sakai, R.; Rinehart, K. L.; Wang, A. H.-L. *J. Biomol. Struct. Dyn.* **1993**, *10*, 793-818.
16. Wemmer, D. E.; Dervan, P. B. *Current Opinion in Structural Biology* **1997**, *7*, 355-361.
17. Geierstanger, B. H.; Wemmer, D. E. *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 463-493.

18. Liu, C.; Smith, B. M.; Ajito, K.; Komatsu, H.; GomezPaloma, L.; Li, T. H.; Theodorakis, E. A.; Nicolaou, K. C.; Vogt, P. K. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 940-944.
19. Kalben, A.; Pal, S.; Andreotti, A. H.; Walker, S.; Gange, D.; Biswas, K.; Kahne, D. *J. Am. Chem. Soc.* **2000**, *122*, 8403-8412.
20. Boger, D. L.; Schmitt, H. W.; B.E., F.; Hendrick, M. P. *J. Org. Chem.* **2001**, *66*, 6654-6661.
21. Boger, D. L.; Stauffer, F.; Hendrick, M. P. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 2021-2024.
22. Satz, A. L.; Bruice, T. C. *Acc. Chem. Res.* **2002**, *35*, 86-95.
23. Minehan, T. G.; Gottwald, K.; Dervan, P. B. *Helv. Chim. Acta.* **2000**, *83*, 2197-2213.
24. Arcamone, F.; Penco, S.; Prezzi, P. G.; Nicolella, V.; Pirelli, A. *Nature* **1964**, *203*, 1064.
25. Bailly, C.; Chaires, J. B. *Bioconjugate Chem.* **1998**, *9*, 513-538.
26. Boger, D. L.; Dechantsreiter, M. A.; Ishii, T.; Fink, B. E.; Hendrick, M. P. *Bioorg. Med. Chem.* **2000**, *8*, 2049-2057.
27. O'Hare, C. C.; Mack, D.; Tandon, M.; Sharma, S. K.; Lown, J. W.; Kopka, M. L.; Dickerson, R. E.; Hartley, J. A. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 72-77.
28. Dyatkina, N. B.; Roberts, C. D.; Keischer, J. D.; Dai, Y.; Nadherny, J. P.; Zhang, W.; Schmitz, U.; Kongpachith, A.; Fung, K.; Novikov, A. A.; Lou, L.; Velligan, M.; Khorlin, A. A.; Chen, M. S. *J. Med. Chem.* **2002**, *45*, 805-817.
29. Dervan, P. B. *Bioorg. Med. Chem.* **2001**, *9*, 2215-2235.
30. Dervan, P. B.; Bürli, R. W. *Curr. Opin. Chem. Biol.* **1999**, *3*, 688-693.
31. Kielkopf, C. L.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Science* **1998**, *282*, 111-115.
32. Gottesfeld, J. M.; Turner, J. M.; Dervan, P. B. *Gene Expression* **2000**, *9*, 77-91.

33. Turner, J. M.; Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 6219-6226.
34. Herman, D. M.; Baird, E. E.; Dervan, P. B. *Chem.-Eur. J.* **1999**, *5*, 975-983.
35. Maeshima, K.; Janssen, S.; Laemmli, U. K. *EMBO J.* **2001**, *20*, 3218-3228.
36. Chang, A. Y.; Dervan, P. B. *J. Am. Chem. Soc.* **2000**, *122*, 4856-4864.
37. Wurtz, N. R.; Dervan, P. B. *Chem. Biol.* **2000**, *7*, 153-161.
38. Wang, C. C. C.; Dervan, P. B. *J. Am. Chem. Soc.* **2001**, *123*, 8657-8661.
39. Janssen, S.; Durussel, T.; Laemmli, U. K. *Mol. Cell* **2000**, *6*(5), 999-1011.
40. Maeshima, K.; Janssen, S.; Laemmli, U. K. *EMBO J.* **2001**, *20*, 3218-3228.
41. Gygi, M. P.; Ferguson, M. D.; Mefford, H. C.; Kevin, P. L.; O'Day, C.; Zhou, P.; Friedman, C.; van den Engh, G.; Stobwitz, M. L.; Trask, B. J. *Nucleic Acids Res.* **2002**, *30*, 2790-2799.