

Chapter 1

Introduction

1.1. Deoxyribonucleic acid (DNA)

DNA is the biological macromolecule that encodes the genetic information necessary for life. The molecule consists of four different bases on a phosphodiester-linked deoxyribose sugar backbone. The intertwined strands of two DNA molecules form a double helix with hydrogen bonds between the Watson-Crick base pairs such that thymine (T) pairs with adenine (A), and cytosine (C) pairs with guanine (G). The structure of B-form DNA is a right-handed helix containing ten base pairs per turn, and the plane of each hydrogen-bonded base pair lies perpendicular to the helix axis.^{1,2} A wide major groove and a narrow minor groove line the helix and are available for binding (Figure 1.1).² The human genome contains roughly three billion base pairs that encode 20,000

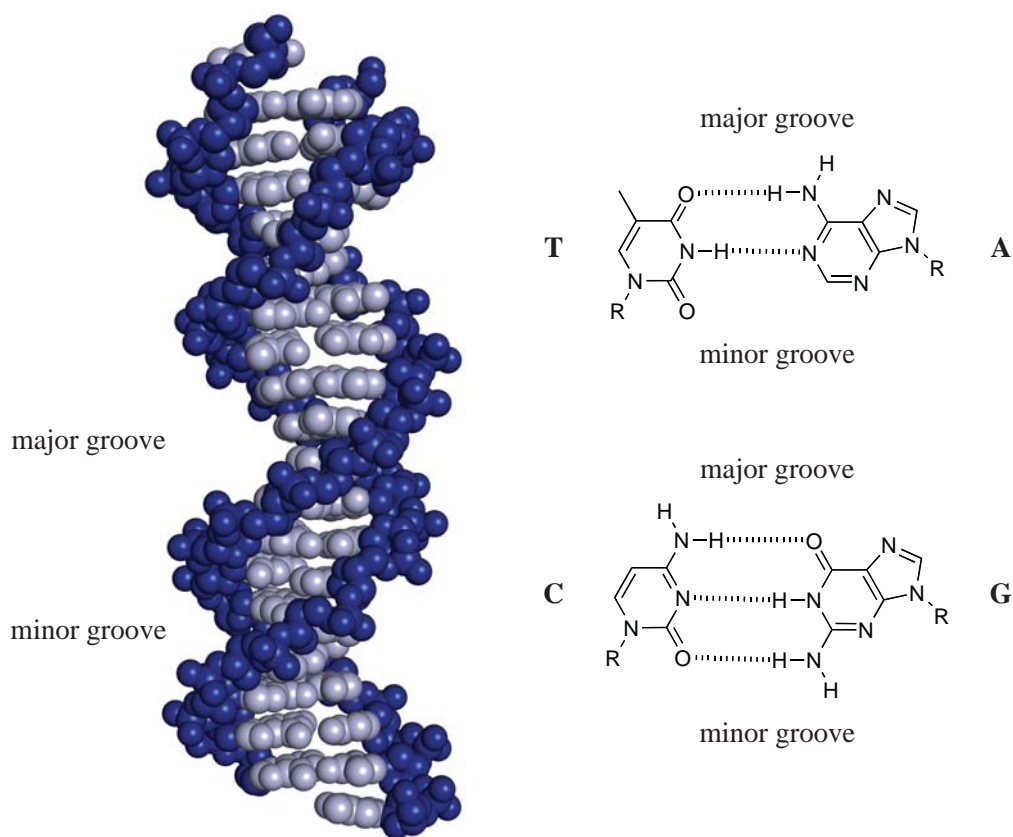


Figure 1.1. The structure of DNA. (left) The phosphodiester-linked deoxyribose backbone is shown in blue, and the Watson-Crick base pairs are shown in gray (PDB accession code: 1YSA⁵). (right) The chemical structures of the DNA bases thymine (T), adenine (A), cytosine (C), and guanine (G) are shown as hydrogen-bonded base pairs.⁴⁰

to 25,000 protein-coding genes.³ In addition to DNA sequences that code for proteins, genes are associated with regulatory regions that control transcription. In the context of the living cell, DNA-binding proteins and the interactions between proteins modulate gene expression.^{4,5}

1.2. DNA-binding small molecules

Nature has evolved a number of small molecules, including actinomycin, echinomycin, and calicheamicin, that bind to specific sequences of DNA. Distamycin A and netropsin are A,T-binding oligopeptidic antibiotic drugs of bacterial origin whose structures contain three and two *N*-methylpyrrole (Py) carboxamide units, respectively.⁶⁻⁹ The X-ray crystal structure of the 1:1 complex between netropsin and 5'-CGCGAATTCGCG-3' shows that the crescent-shaped natural product binds in the minor groove of B-form DNA by displacing water molecules on the spine of hydration.^{10,11} The NMR structure of the 2:1 complex between distamycin and 5'-CGCAAATTGGC-3' indicates that the two distamycin molecules bind the central 5'-AAATT-3' region in an antiparallel orientation with expansion of the minor groove relative to the 1:1 ligand-DNA complex.¹² The structures of the 1:1 and 2:1 distamycin-DNA complexes are depicted in Figure 1.2.^{13,14}

1.3. Recognition of the DNA minor groove by pyrrole-imidazole polyamides

A modular code for molecular recognition of the four Watson-Crick base pairs has been empirically determined using pyrrole-imidazole polyamides.^{15,16} These synthetic ligands recognize the chemical features presented in the minor groove of DNA as a form of digital readout (Figure 1.3). The *N*-methylimidazole (Im) analog of distamycin binds the five base-pair sequence 5'-WGWCW-3' (where W = A or T) by formation of a 2:1 polyamide-DNA complex.^{17,18} The two ImPyPy molecules are stacked together as an antiparallel dimer in the DNA minor groove with Im paired unsymmetrically across from Py to recognize the G•C base pair. Hydrogen bonding between the Im nitrogen and the

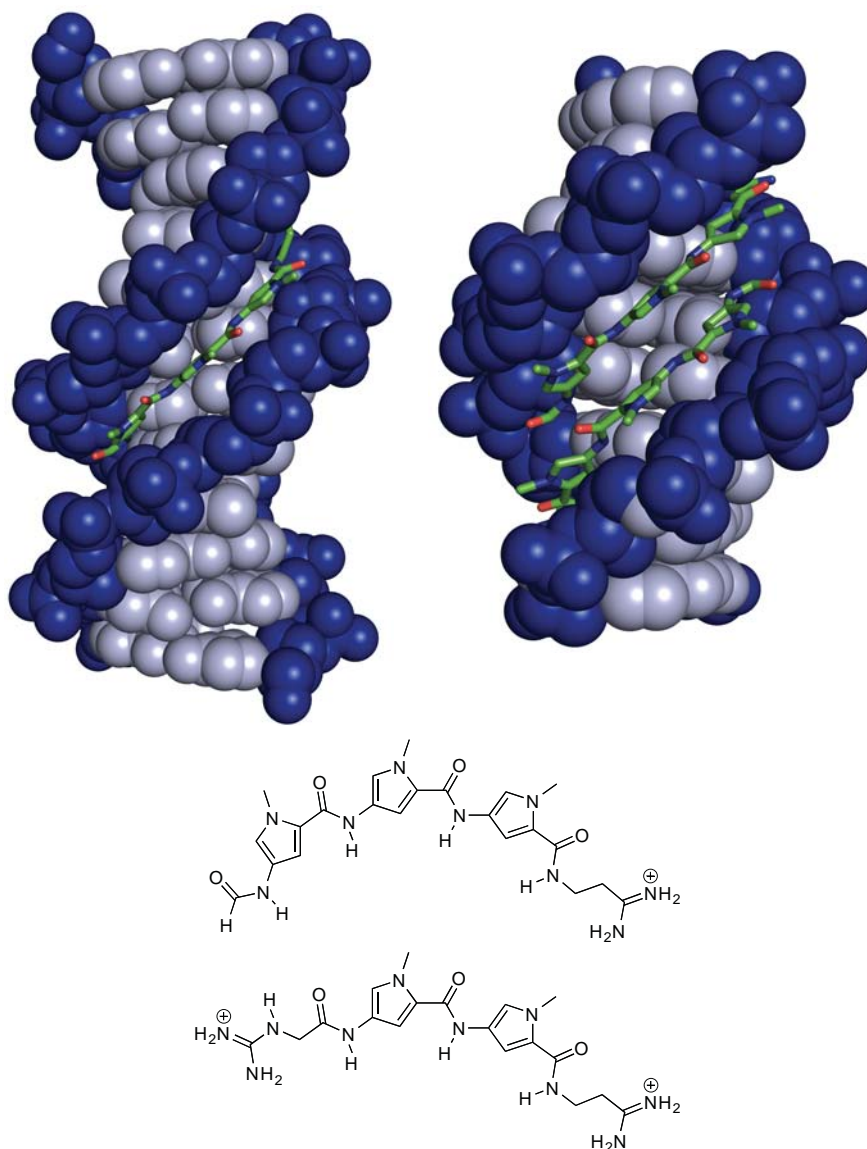


Figure 1.2. Structures of distamycin A bound to DNA as 1:1 and 2:1 complexes (PDB accession codes: 2DND¹³ and 378D¹⁴). The DNA backbone is shown in blue, the base pairs are gray, and distamycin is shown in stick representation. The chemical structures of distamycin A (top) and netropsin (bottom) are shown below the crystal structures.⁴⁰

exocyclic amine of guanine helps to confer specificity for G•C over C•G, A•T and T•A.¹⁹ In this way, the Im/Py pair recognizes G•C and the Py/Im pair codes for C•G, while the degenerate Py/Py pair does not discriminate between A•T and T•A. To complete the task of recognizing all four Watson-Crick base pairs, *N*-methylhydroxypyrrole (Hp) was paired

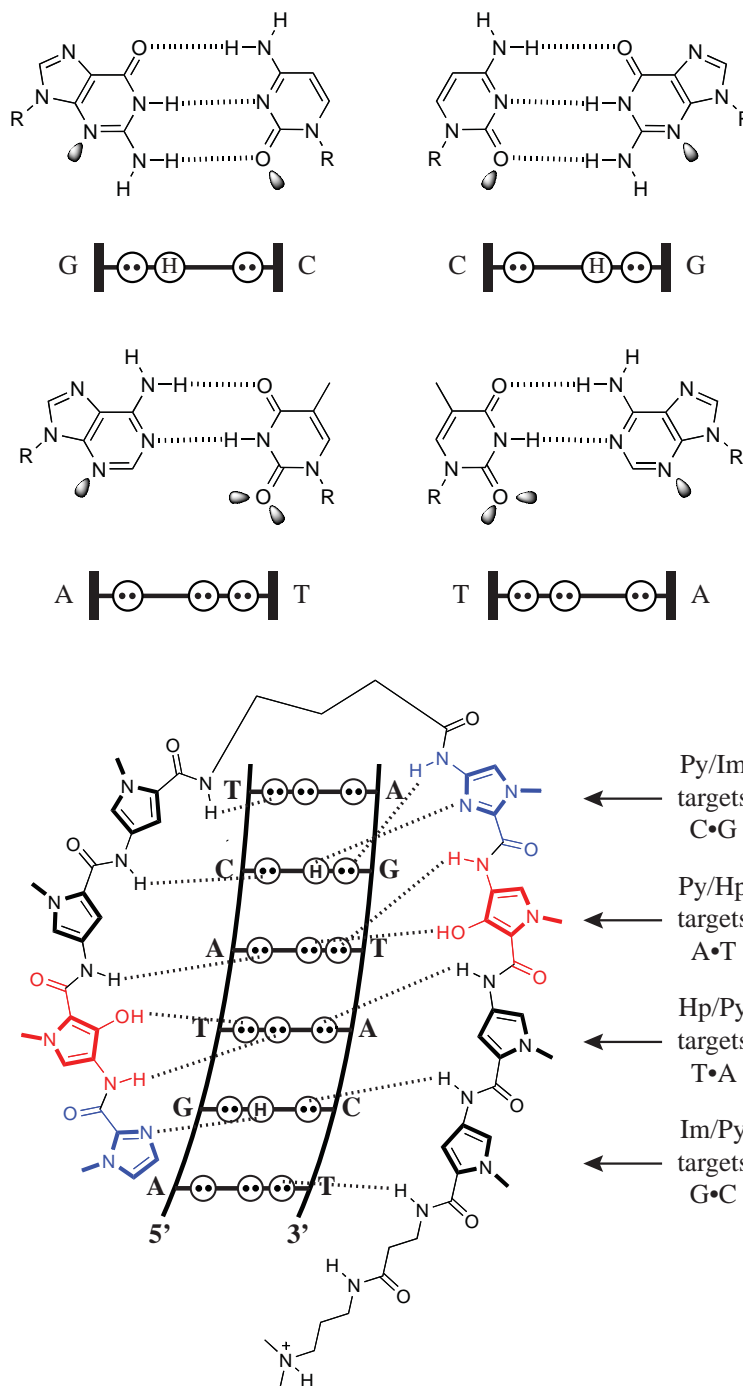


Figure 1.3. Molecular recognition of the minor groove of DNA. (top) Minor groove hydrogen bonding patterns of Watson-Crick base pairs. Circles with dots represent lone pairs of N(3) of purines and O(2) of pyrimidines, and circles containing an H represent the 2-amino group of guanine. The R group represents the sugar-phosphate backbone of DNA. Electron pairs projecting into the minor groove are represented as shaded orbitals. (bottom) Binding model for the complex formed between ImHpPyPy- γ -ImHpPyPy- β -Dp and a 5'-AGTACT-3' sequence. Putative hydrogen bonds are shown as dashed lines.¹⁶

across from Py to recognize T•A over A•T.²⁰⁻²² The X-ray crystal structure of the 2:1 complex between ImHpPyPy and 5'-CCAGTACTGG-3' provides structural verification of the pairing rules (Figure 1.4).

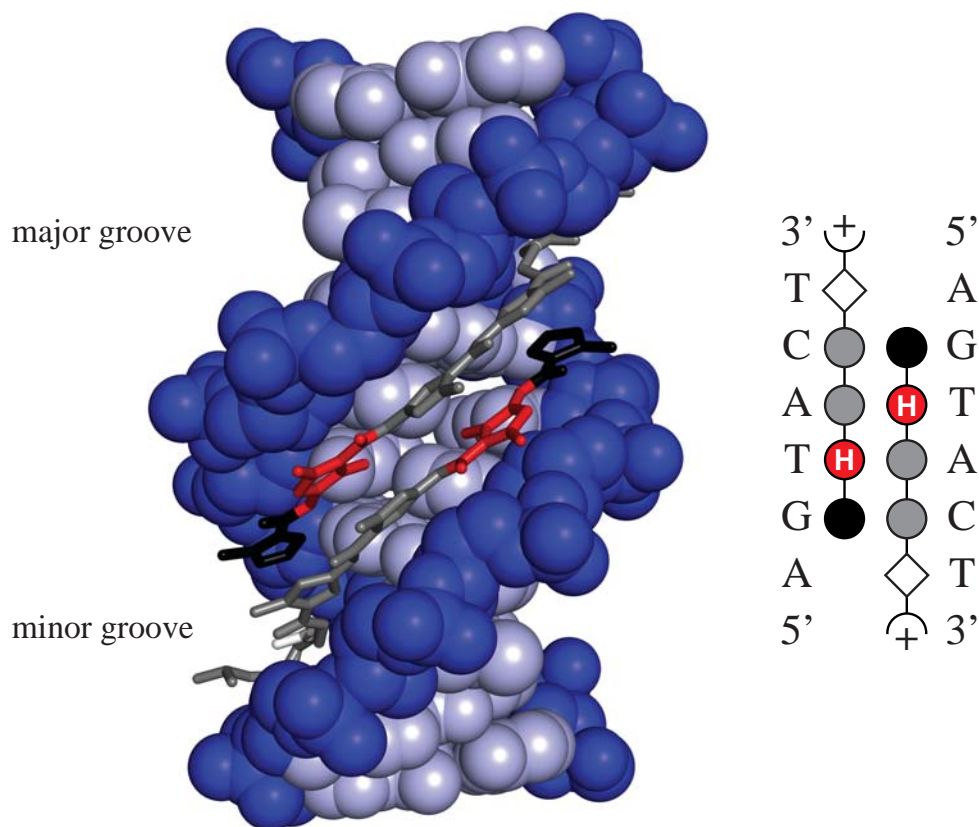


Figure 1.4. Crystal structure of the 2:1 complex formed between ImHpPyPy-β-Dp and the 5'-CCAGTACTGG-3' sequence (PDB accession code: 407D²¹). Polyamides are shown at right in a ball-and-stick model, where a black circle represents Im, a red circle with an H represents Hp, a gray circle represents Py, a white diamond represents β-alanine, and a half circle with a plus represents Dp.

1.4. Hairpin polyamide motif

The pre-organization of two polyamide subunits by a covalent linkage improves the binding affinity and sequence specificity of these ligands.²³⁻²⁶ While several different binding motifs have been explored, the hairpin motif provides excellent affinity and specificity for DNA and can be accessed through a facile synthetic route.²³ In the hairpin motif, a γ -aminobutyric acid (γ) covalently links the carboxylic terminus of one polyamide with the amino terminus of the other. Eight-ring hairpin polyamides bind their six base-pair DNA match sequences with subnanomolar affinities.^{27,28}

Structural features of the hairpin motif contribute to sequence-specific DNA recognition (Figure 1.5). The γ residue that forms the turn of the hairpin recognizes A•T and T•A base pairs. A chiral 2,4-diaminobutyric acid improves affinity while retaining specificity for A•T and T•A.²⁹ The increased binding affinity can be attributed to electrostatic interactions between the negatively charged DNA backbone and the positively charged free amine on the turn. Furthermore, the chiral turn allows stereochemical control, enforcing the N→C polyamide alignment with respect to the 5'→3' orientation of the adjacent DNA strand.^{29,30}

The microstructure of DNA is sequence-dependent, and polyamides composed of Py and Im rings are overly curved compared to the DNA minor groove. Aliphatic β -alanine residues have been incorporated at internal positions to reset the register between contiguous aromatic ring pairings.³¹ The β /Py and Py/ β pairs, as well as the β / β degenerate pair, recognize A•T and T•A base pairs.^{31,32} The β -alanine residue on the tail, as well as the 3-(dimethylamino)propylamine (Dp) tail itself, both code for A•T and T•A.³³ Hairpin polyamide design typically includes an Im/Py pairing at the hairpin N-terminus in order to impart G•C specificity at this position. However, when 3-chlorothiophene (Ct) is paired across from Py at the N-terminus, the Ct/Py pair exhibits specificity for T•A.³⁴ In this way, the eight-ring hairpin polyamide CtPyPyIm-(R)^{RHN} γ -PyImPyPy- β -Dp recognizes the seven base-pair sequence 5'-WWTWCGW-3' in the DNA minor groove.

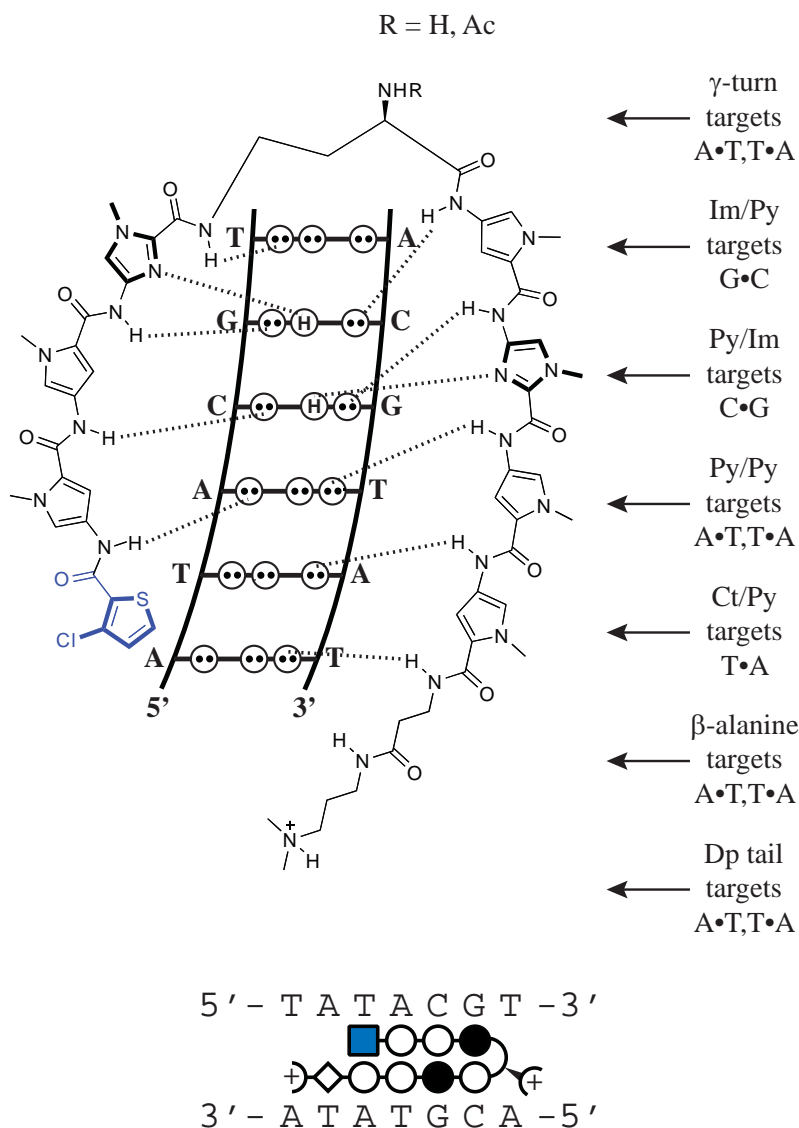


Figure 1.5. Binding model for the complex formed between CtPyPyIm-(R)^{RHN}γ-PyImPyPy-β-Dp and a 5'-TATACGT-3' sequence. A ball-and-stick model is shown at bottom, where a blue square represents Ct, a black circle represents Im, a white circle represents Py, a white diamond represents β-alanine, a half circle with a plus represents Dp, and the chiral 2,4-diaminobutyric acid turn residue is shown as a semicircle connecting the two subunits with a solid wedge linking a half-circle with a plus.

Methods for solid-phase synthesis of hairpin polyamides have been developed using standard Boc coupling chemistry.^{35,36} Aminolysis furnishes the polyamide, which can be further derivatized on the turn or tail residues. The selection of either β-Ala-PAM or Kaiser oxime resin yields different C-terminal groups following cleavage with a primary

amine (Figure 1.6). The use of β -Ala-PAM resin leaves a residual A,T-specific β -alanine moiety at the C-terminus, while oxime resin produces a shorter polyamide. Access to a diverse range of polyamide binding motifs and linkers has proven valuable for experiments in cell culture.

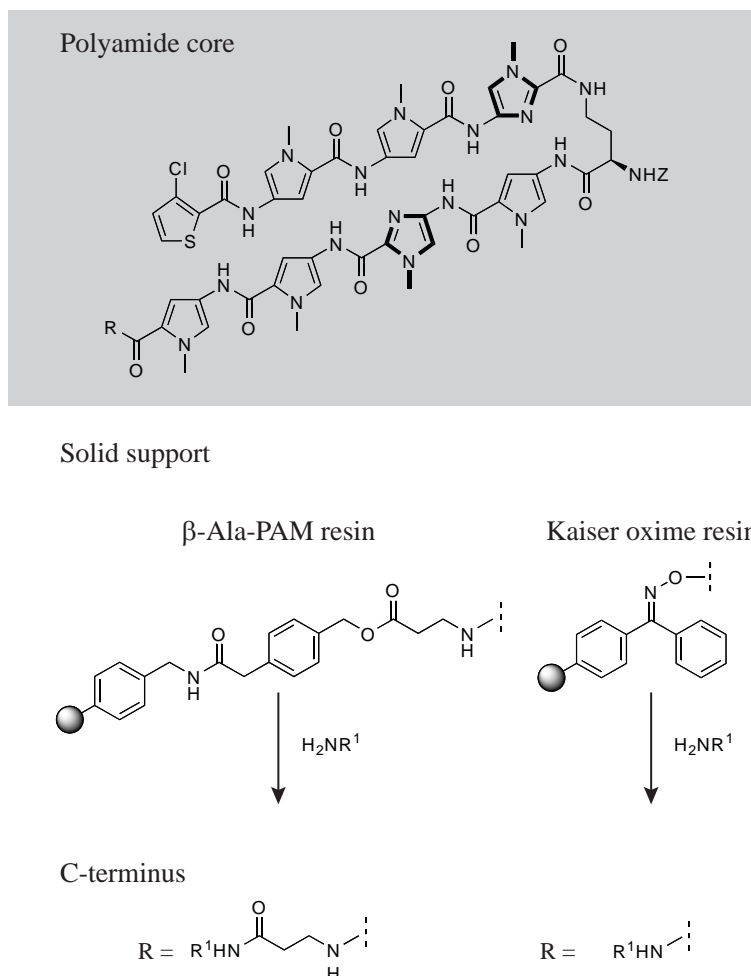


Figure 1.6. Synthesis of polyamides on solid support. β -Ala-PAM resin and Kaiser oxime resin yield different C-terminal groups following cleavage with a primary amine. Z = H, Ac, or Boc, if further derivatization is required.¹⁶

1.5. Nuclear localization of hairpin polyamide conjugates in cell culture studies

Confocal microscopy has been used to determine the uptake profiles of polyamide-fluorophore conjugates in a range of cell lines.³⁷⁻⁴⁰ A brief excerpt is presented in Figure 1.7. Polyamide **1** localizes to the nucleus, while polyamide **2** is excluded from the nucleus, which reveals that uptake is core-dependent. Removal of the β -alanine residue to produce

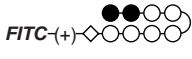
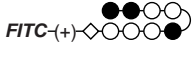
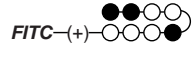
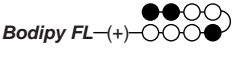
		DLD-1	HeLa	MCF-7	SK-BR-3	786-O	293	LN-CaP	PC3	MEL	NB4	Jurkat	CCRF-CEM	MEG-01
	1	+	++	++	++	+	+	++	++	+	++	++	++	++
	2	-	--	--	--	--	--	--	--	--	--	--	--	--
	3	+	++	++	++	++	++	++	++	++	++	++	++	++
	4	-	-	-	-	--	--	--	-	-	-	-	--	-

Figure 1.7. Uptake profile of polyamide-fluorophore conjugates **1-4** in 13 cell lines.³⁸ ++ indicates that nuclear staining exceeds that of the medium; +, nuclear staining less than or equal to that of the medium, but still prominent; -, very little nuclear staining, with the most fluorescence seen in the cytoplasm and/or medium; --, no nuclear staining.^{38,39} In the ball-and-stick model, a black circle represents Im, a white circle represents Py, a white diamond represents β -alanine, “(+)” represents the 3,3'-diamino-*N*-methyldipropylamine linker, and a half-circle represents the γ -aminobutyric acid turn residue.

polyamide **3** rescues uptake, suggesting that the incorporation of β -alanine may be lethal for uptake. The poor nuclear localization of polyamide **4** indicates that fluorophore selection is crucial.⁴¹ Based on these results, fluorescein has been the fluorescent dye of choice in recent experiments.

The gained insights into cellular uptake properties of polyamides have proven instrumental in the successful design of gene regulation studies. A polyamide-fluorescein conjugate has been shown to decrease hypoxia-inducible transcription of vascular endothelial growth factor (VEGF) in cultured HeLa cells.^{42,43} The binding of this polyamide to the hypoxia response element (HRE) disrupts the binding of hypoxia-inducible factor to the HRE. Inhibition of VEGF expression has been used to assay the nuclear localization of HRE-targeted polyamides. These studies have shown that the fluorescein dye can be minimized to a isophthalic acid moiety, yielding higher-affinity conjugates that exhibit comparable biological activity.⁴⁴ A hairpin polyamide-isophthalic acid conjugate has been shown to downregulate prostate-specific antigen (PSA) expression in prostate cancer cells by binding the androgen response element in the PSA promoter.⁴⁵ This prevents the binding of androgen receptor (AR) and inhibits androgen-induced PSA expression.

1.6. Scope of this work

In Chapter 2, we complete a library of 27 hairpin Py-Im polyamides that bind 7-base-pair sequences of the general form 5'-WWGNNNW-3' (where W = A or T, N = W, G, or C). A table of binding affinities and sequence contexts for this completed 27-member library has been assembled for the benefit of the chemical biology community interested in molecular control of transcription. In Chapter 3, quantitative fluorescence-based methods have been developed to determine the nuclear concentration of polyamide-fluorescein conjugates in cell culture. Confocal laser scanning microscopy and flow cytometry techniques are utilized to plot calibration curves, from which the nuclear concentration can be interpolated. Although confocal microscopy and flow cytometry generate disparate values, taken together these experiments suggest that the polyamide concentration inside the cell nucleus is lower than it is outside the cell. In Chapter 4, to further our understanding of C-terminal tail linkage effects on sequence specificity, the equilibrium association constants of hairpin polyamide conjugates were measured by quantitative DNase I footprint titration experiments. These results indicate that linkers and functional R groups on the tails of hairpin polyamide conjugates have recognition properties that should be considered in the design of these molecules to target DNA binding sites. Furthermore, these β -alanine- C_3 -linked polyamide conjugates are shown to decrease hypoxia-inducible transcription of vascular endothelial growth factor (VEGF) in cultured HeLa cells. In Chapter 5, polyamide conjugates designed to target the Oct4 octamer DNA element modulate the expression levels of Oct4-driven genes in P19 mouse embryonal carcinoma cells and R1 mouse embryonic stem (ES) cells.

References

1. Wing, R.; Drew, H.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E., *Nature* **1980**, 287, 755-758.
2. Dickerson, R. E.; Drew, H. R.; Conner, B. N.; Wing, R. M.; Fratini, A. V.; Kopka, M. L., *Science* **1982**, 216, 475-485.
3. Collins, F. S.; Lander, E. S.; Rogers, J.; Waterston, R. H., *Nature* **2004**, 431, 931-945.
4. Pabo, C. O.; Sauer, R. T., *Annu. Rev. Biochem.* **1984**, 53, 293-321.
5. Ellenberger, T. E.; Brandl, C. J.; Struhl, K.; Harrison, S. C., *Cell* **1992**, 71, 1223-1237.
6. Arcamone, F.; Nicoletti, V.; Penco, S.; Orezzi, P.; Pirelli, A., *Nature* **1964**, 203, 1064-&.
7. Finlay, A. C.; Hochstein, F. A.; Sobin, B. A.; Murphy, F. X., *J. Am. Chem. Soc.* **1951**, 73, 341-343.
8. Vantamelen, E. E.; White, D. M.; Kogon, I. C.; Powell, A. D. G., *J. Am. Chem. Soc.* **1956**, 78, 2157-2159.
9. Waller, C. W.; Wolf, C. F.; Stein, W. J.; Hutchings, B. L., *J. Am. Chem. Soc.* **1957**, 79, 1265-1266.
10. Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E., *Proc. Natl. Acad. Sci. U. S. A.* **1985**, 82, 1376-1380.
11. Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E., *J. Mol. Biol.* **1985**, 183, 553-563.
12. Pelton, J. G.; Wemmer, D. E., *Proc. Natl. Acad. Sci. U. S. A.* **1989**, 86, 5723-5727.
13. Coll, M.; Frederick, C. A.; Wang, A. H. J.; Rich, A., *Proc. Natl. Acad. Sci. U. S. A.* **1987**, 84, 8385-8389.
14. Mitra, S. N.; Wahl, M. C.; Sundaralingam, M., *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **1999**, 55, 602-609.
15. Dervan, P. B., *Bioorg. Med. Chem.* **2001**, 9, 2215-2235.

16. Dervan, P. B.; Edelson, B. S., *Curr. Opin. Struct. Biol.* **2003**, 13, 284-299.
17. Wade, W. S.; Mrksich, M.; Dervan, P. B., *J. Am. Chem. Soc.* **1992**, 114, 8783-8794.
18. Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B., *Proc. Natl. Acad. Sci. U. S. A.* **1992**, 89, 7586-7590.
19. Kielkopf, C. L.; Baird, E. E.; Dervan, P. D.; Rees, D. C., *Nat. Struct. Biol.* **1998**, 5, 104-109.
20. White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B., *Nature* **1998**, 391, 468-471.
21. 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.
22. 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, 557-567.
23. Mrksich, M.; Parks, M. E.; Dervan, P. B., *J. Am. Chem. Soc.* **1994**, 116, 7983-7988.
24. Greenberg, W. A.; Baird, E. E.; Dervan, P. B., *Chem.-Eur. J.* **1998**, 4, 796-805.
25. Melander, C.; Herman, D. M.; Dervan, P. B., *Chem.-Eur. J.* **2000**, 6, 4487-4497.
26. Heckel, A.; Dervan, P. B., *Chem.-Eur. J.* **2003**, 9, 3353-3366.
27. Trauger, J. W.; Baird, E. E.; Dervan, P. B., *Nature* **1996**, 382, 559-561.
28. Trauger, J. W.; Dervan, P. B., *Methods in Enzymology* **2001**, 340, 450-466.
29. Herman, D. M.; Baird, E. E.; Dervan, P. B., *J. Am. Chem. Soc.* **1998**, 120, 1382-1391.
30. White, S.; Baird, E. E.; Dervan, P. B., *J. Am. Chem. Soc.* **1997**, 119, 8756-8765.
31. Turner, J. M.; Swalley, S. E.; Baird, E. E.; Dervan, P. B., *J. Am. Chem. Soc.* **1998**, 120, 6219-6226.
32. Wang, C. C. C.; Ellervik, U.; Dervan, P. B., *Bioorg. Med. Chem.* **2001**, 9, 653-657.

33. Swalley, S. E.; Baird, E. E.; Dervan, P. B., *J. Am. Chem. Soc.* **1999**, 121, 1113-1120.
34. Foister, S.; Marques, M. A.; Doss, R. M.; Dervan, P. B., *Bioorg. Med. Chem.* **2003**, 11, 4333-4340.
35. Baird, E. E.; Dervan, P. B., *J. Am. Chem. Soc.* **1996**, 118, 6141-6146.
36. Belitsky, J. M.; Nguyen, D. H.; Wurtz, N. R.; Dervan, P. B., *Bioorg. Med. Chem.* **2002**, 10, 2767-2774.
37. Belitsky, J. M.; Leslie, S. J.; Arora, P. S.; Beerman, T. A.; Dervan, P. B., *Bioorg. Med. Chem.* **2002**, 10, 3313-3318.
38. Best, T. P.; Edelson, B. S.; Nickols, N. G.; Dervan, P. B., *Proc. Natl. Acad. Sci. U. S. A.* **2003**, 100, 12063-12068.
39. Edelson, B. S.; Best, T. P.; Olenyuk, B.; Nickols, N. G.; Doss, R. M.; Foister, S.; Heckel, A.; Dervan, P. B., *Nucleic Acids Res.* **2004**, 32, 2802-2818.
40. Stafford, R. L. Ph.D. Thesis, California Institute of Technology, Pasadena, CA, 2008.
41. Crowley, K. S.; Phillion, D. P.; Woodard, S. S.; Schweitzer, B. A.; Singh, M.; Shabany, H.; Burnette, B.; Hippenmeyer, P.; Heitmeier, M.; Bashkin, J. K., *Bioorg. Med. Chem. Lett.* **2003**, 13, 1565-1570.
42. Olenyuk, B. Z.; Zhang, G. J.; Klco, J. M.; Nickols, N. G.; Kaelin, W. G.; Dervan, P. B., *Proc. Natl. Acad. Sci. U. S. A.* **2004**, 101, 16768-16773.
43. Nickols, N. G.; Jacobs, C. S.; Farkas, M. E.; Dervan, P. B., *ACS Chem. Biol.* **2007**, 2, 561-571.
44. Nickols, N. G.; Jacobs, C. S.; Farkas, M. E.; Dervan, P. B., *Nucleic Acids Res.* **2007**, 35, 363-370.
45. Nickols, N. G.; Dervan, P. B., *Proc. Natl. Acad. Sci. U. S. A.* **2007**, 104, 10418-10423.