Chapter 1

Introduction to DNA-Binding Polyamides

1.1 DNA-Binding Molecules

The genetic code for all organisms is encoded in the double stranded molecule DNA. While DNA itself contains all of an organism's genes, transcription is regulated by an intricate network of interactions. The binding of protein transcription factors to specific sequences of DNA is an important part of gene regulation. With the human genome estimated to contain over 25,000 genes,[1, 2] DNA-binding biological or chemical molecules involved in transcriptional regulation must be able to bind to specific sequences of DNA with high affinity and specificity.

The structure of DNA consists of two antiparallel polydeoxyribonucleotide strands intertwined as a double helix.[3] Hydrogen bonds between specific base pairs hold the two strands together. The pairing between DNA bases is highly specific: adenine pairs with thymine, and guanine pairs with cytosine. The DNA helix possesses two grooves, a broad shallow major groove and a narrow and deep minor groove. Differences in the surfaces of these grooves provide the basis for selective DNA recognition.

Distamycin A, composed of three aromatic N-methylpyrrole rings, is a relatively simple natural product capable of DNA recognition. This small organic molecule has been shown to bind preferentially in the minor groove of A,T tracts in 1:1 and 2:1 complexes with DNA.[4, 5] The amide backbone of distamycin A makes a series of hydrogen bonds to the N3 of adenine and O2 of thymine in the DNA sequence. In addition, the crescent shape of distamycin allows it to follow the curvature of the minor groove and make van der Waals contacts with its walls.[4, 5]

The Dervan group has shown that polyamides, small organic analogues of distamycin A, are able to bind DNA in a sequence specific fashion.[6, 7] *N*-methylpyrrole (Py) heterocycle carboxamides when bound to DNA in a 2:1 complex were found to have high affinity for A and T base pairs. It was then shown that the use of *N*-methyl-imidazole (Im) paired with Py is specific for the G·C basepair.[8, 9] The exocyclic amine of guanine sterically clashes with pyrrole but is tolerated by imidazole which it is also capable of making a hydrogen bond with, thus imparting specificity for the G·C base pair.

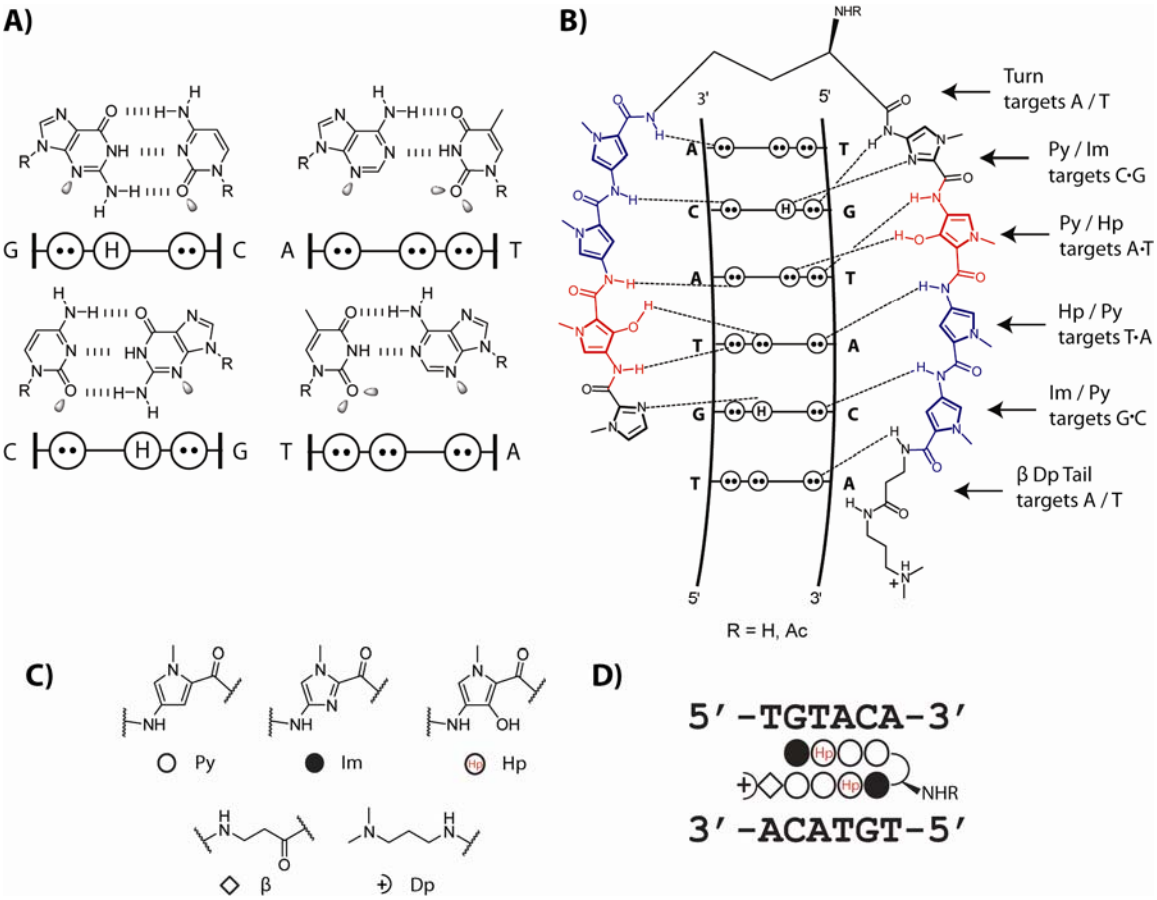

Figure 1.1 Molecular recognition of the minor groove of DNA. (A) Minor groove hydrogen-bonding patterns of Watson-Crick base pairs. Circles with dots represent lone pairs of N(3) of purines and O(2) of pyrimidines, and circles containing an H represent the 2-amino group of guanine. The R group represents the sugar-phosphate backbone of DNA. Electron lone pairs projecting into the minor groove are represented as shaded orbitals (B) Binding model for the complex formed between ImHpPyPy-γ-ImHpPyPy-β-Dp and a 5'-TGTACA-3' sequence. Putative hydrogen bonds are shown as dashed lines. (C) Chemical structures of commonly used monomers in polyamides along with their ball and stick representations. D) Ball-and-stick model of the hairpin polyamide shown in B.

Similarly, the use of *N*-methyl-3-hydroxypyrrole (Hp) which can form a combination of hydrogen bonds between the hydroxyl and the thymine O2 can specifically recognize the T·A base pair when it is paired across from Py.[10] Thus a set of pairing rules have been established whereby Im/Py, Py/Im, Hp/Py, Py/Hp pairings are specific for the G·C, C·G, T·A, and A·T base pairs, respectively. The Py/Py pairing is degenerate for A·T or T·A. In the most commonly used motif, two antiparallel strands of polyamides are linked N-terminus to C-terminus using γ-aminobutyric acid to form a hairpin polyamide.[11, 12] In general polyamides are usually oriented N to C in the 5' to 3' direction of the DNA

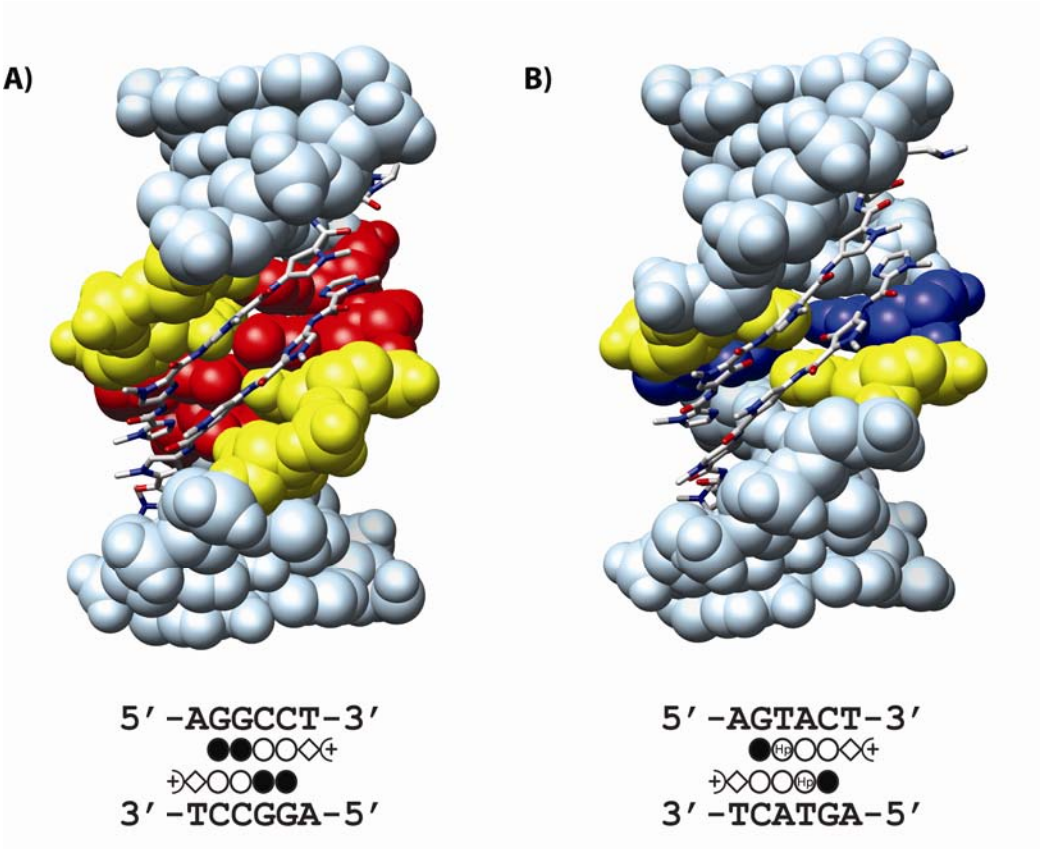

Figure 1.2 Crystal structures showing polyamides bound in a 2:1 complex with DNA. A) Structure of a polyamide homodimer bound to a 5'-GGCC-3' core sequence. Guanine and cytosine are colored red and yellow respectively. A Ball-and-stick model of the Im-Im-Py-Py-β-Dp homodimer is shown at bottom. B) Structure of the polyamide homodimer Im-Hp-Py-Py-β-Dp bound to a 5'-GTAC-3' core sequence. Adenine and thymine are colored blue and yellow respectively. The Ball-and-stick model of the homodimer is shown at bottom.

strands, although exceptions have been observed.[12]

Given the large size of the human genome, the ability to target increasingly long sequences of DNA is important for gene regulation, as one often wants to limit the number of sites affected by any given molecule to prevent unwanted side effects. A standard eight-ring hairpin polyamide has specificity for 6 bp of DNA. It has been shown that polyamide curvature no longer matches that of DNA after five contiguous rings,[13] limiting the length of sequences that can be targeted with this motif. As a result, a variety of extended motifs have been developed to target longer sequences than what is possible with a single hairpin.[7] Two six-ring polyamides linked head-to-head or turn-to-tail have been shown to be able to recognize sequences as long as 10 bp with high affinity.[14-16]

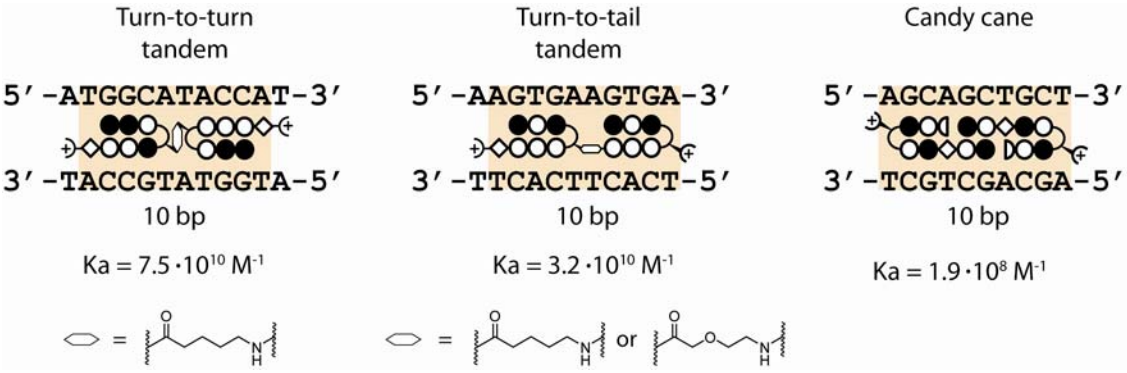

Figure 1.3 Polyamide motifs for binding extended sequences. Ball-and-stick models as well as the corresponding binding site size and affinity are shown.

One of the most active areas of research using polyamides is gene regulation. Polyamides targeted to the promoter region of a gene can be used to disrupt transcription factor binding and inhibit gene transcription.[7, 17] Recent successes in inhibiting vascular endothelial growth factor (VEGF) and androgen receptor (AR) have demonstrated the power of this approach.[18-20] By interfering with the binding of hypoxia-inducible factor (HIF-1) to the hypoxia response element (HRE), polyamides inhibited the hypoxia-

induced gene expression of VEGF and other HIF-1 related genes in cultured cells. Similarly, polyamides to the androgen response element resulted in inhibition of androgen-induced expression of prostate-specific antigen (PSA) and other AR-regulated genes in cultured prostate cancer cells. In a complimentary field of research, there has been much progress in the use of polyamides for the creation artificial transcription factors.[21-24] The goal of these ambitious studies is to combine polyamides which function as a DNA-binding domain with minimal activation domains that can recruit the transcriptional machinery. In this manner, polyamides can be used as programmable molecules for gene regulation.

1.2 Addressing higher-order DNA structures

While the ability to address B-form DNA with polyamides has been well-characterized, it was not known whether polyamides could successfully bind to higher-order DNA structures. Nature uses higher-order DNA structures to compact the DNA in cells. Chromosomes consist of compacted chromatin, which is itself comprised of nucleosomes. Nucleosomes, comprised of DNA wound around a core of histone proteins, represent the most basic unit used by eukaryotes for condensing their DNA. The super-helical structure of the DNA wrapped around the histones in nucleosomes, as well as the partial blockage of DNA sites that face the histone core, present unique challenges for molecular recognition of DNA in these architectures.

In addition to the higher-order structures found in nature, the use of complex DNA architectures has found profound use in structural DNA nanotechnology.[25, 26] Beginning with pioneering work by Ned Seeman, it was originally demonstrated how

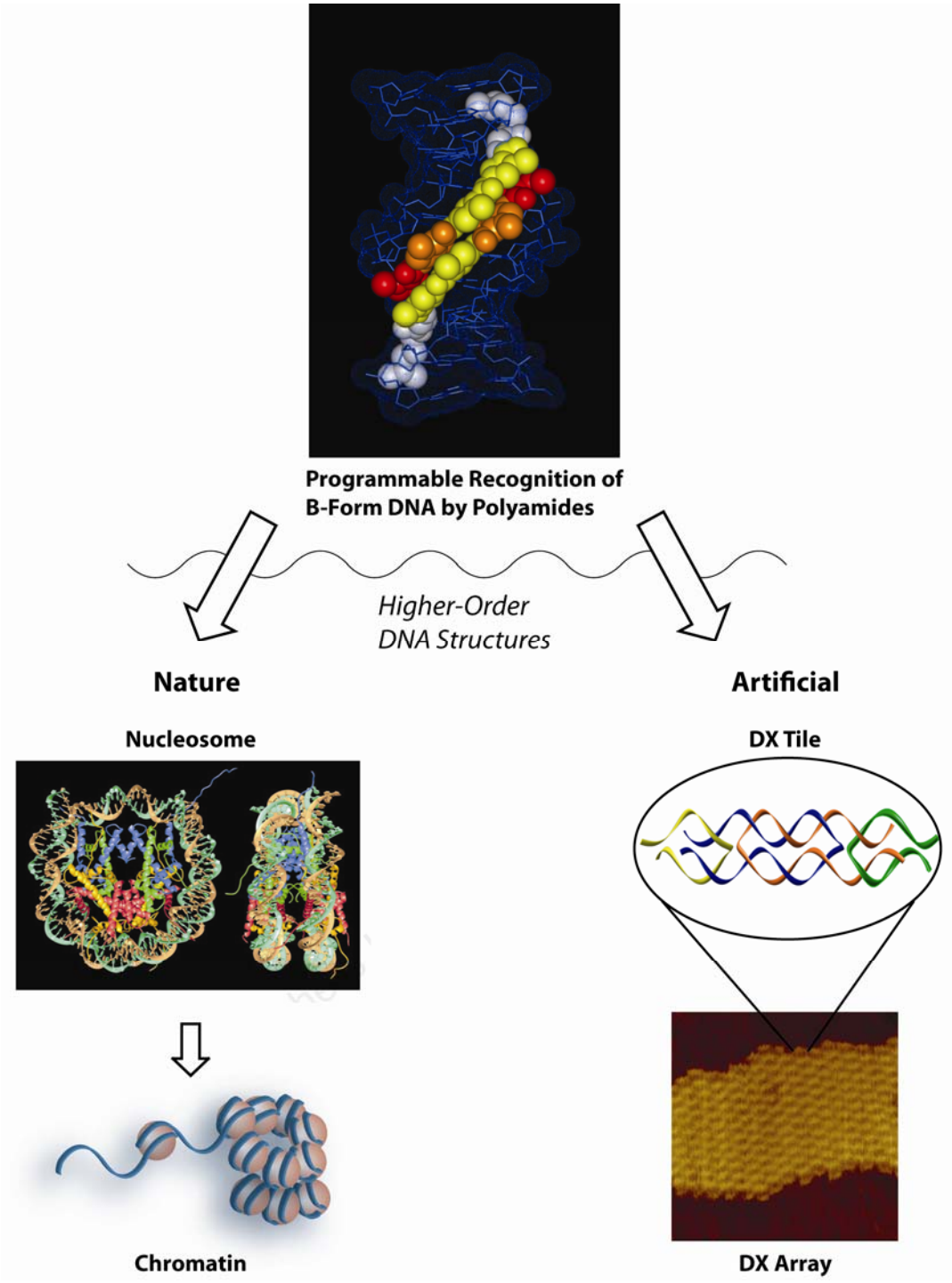

Figure 1.4 Overview of higher-order DNA structures. At top, recognition of B-form DNA by ImHpPyPy-β-Dp polyamide. Higher-order DNA structures are shown in increasing complexity. Nature uses the nucleosome to condense DNA,[27] which is in turn compacted into chromatin. At right, artificial DNA nanostructures can be used to create complex 2-dimensional DNA arrays.

DNA could be made to self-assemble into rigid 2-dimensional structures.[28] These Holliday junction mimics consist of two helices of DNA linked by stable crossover junctions. These stable DNA structures, known as DX tiles, possess four sets of sticky ends, and can be made to assemble into 2-dimensional arrays.[29] Since then, a large number of DNA architectures have been described, based upon or extending the basic DX motif.

While the origins and functions of nucleosomes and DX-arrays are quite distinct, they both raised a common question. Can polyamides that target B-form DNA be used to address these higher-order DNA structures? The primary goal of the investigations contained in this thesis has been to answer that question.

1.3 Targeting the Nucleosome Core Particle

In eukaryotes, DNA is found as chromatin, a highly compact complex of DNA and protein. The primary unit of chromatin is the nucleosome core particle (NCP) consisting of a 147 bp sequence of DNA wrapped approximately twice around an octamer of histone proteins. A short linker strand between 20 and 80 bp of DNA links successive NCPs.[30]

Chromatin exists in numerous states which can be transcriptionally active in the form of euchromatin, or transcriptionally inactive as in heterochromatin. The transition between these two states is thought to be regulated at the most basic level by changes in the nucleosome structure and the surface of the histone octamer.[30] When DNA is packaged in nucleosomes, only 70% of the DNA surface is solvent accessible as the portion that faces the histone core is sterically blocked. Along the outer surface, the two

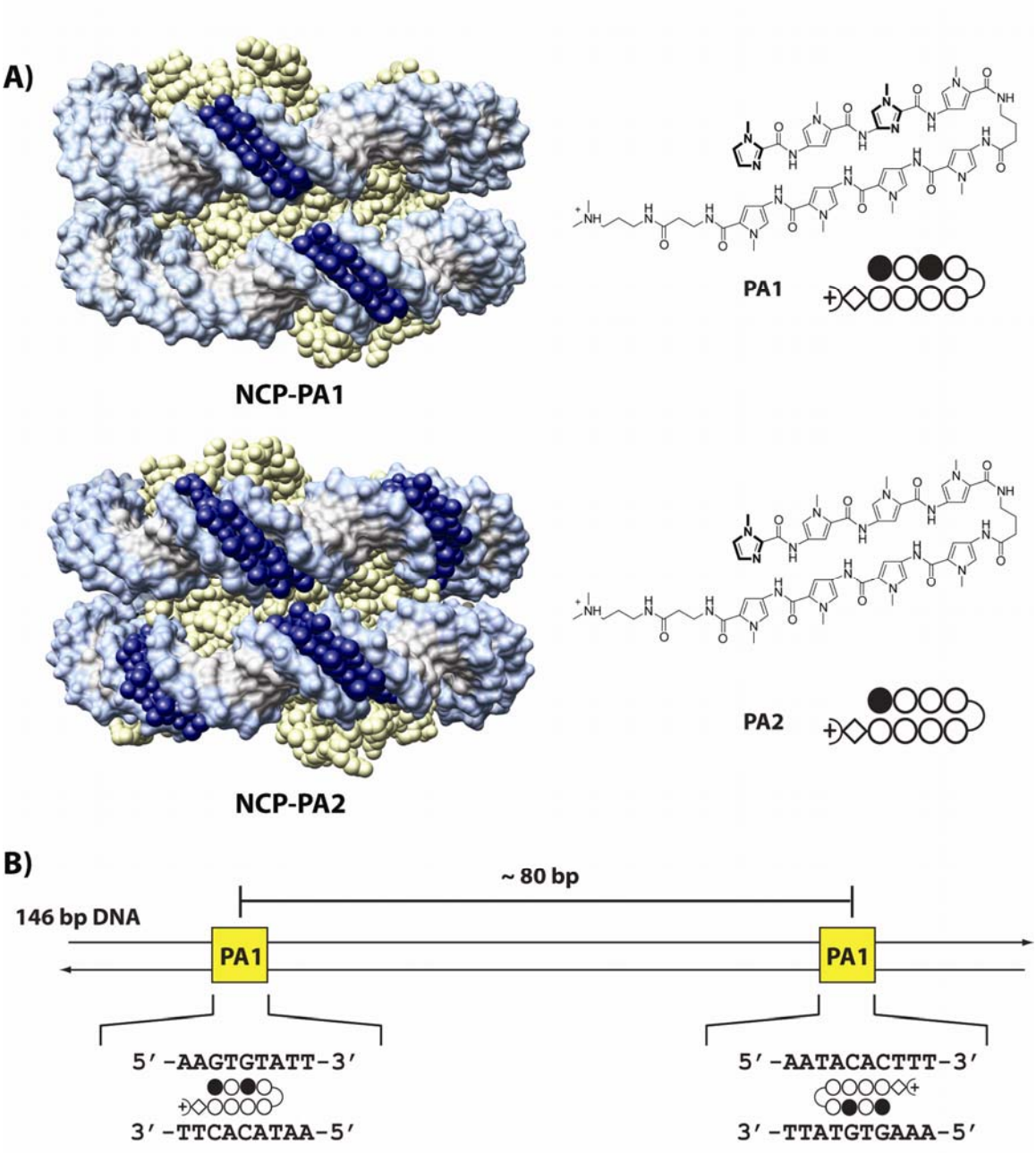

Figure 1.5 Crystal structures of hairpin polyamides bound to the NCP. A) Chemical structures and ball-and-stick models are shown for each polyamide next to the corresponding crystal structure. **PA1** and **PA2** are colored navy blue in structures. (PDB code 1m18 and 1m19) B) Schematic of the palindromic 146 bp DNA strand of the NCP with the binding sites for **PA1** shown in yellow. The sites are separated by ~80 bp.

gyres of DNA stack, forming "supergrooves" of DNA in which the major and minor grooves are aligned to provide a 14 – 16 bp platform for molecular recognition. The

ability to bind to and modulate the stability of nucleosomes would have obvious importance for gene regulation in eukaryotes.

It was shown that polyamides are able to bind to reconstituted nucleosome core particles with high affinity and specificity for their target DNA sites.[31] Subsequently, high resolution X-ray crystal structures were obtained showing the NCP in complex with hairpin polyamides.[32] Surprisingly, polyamides were able to access many sites facing away, and even partially facing towards the histone proteins. In addition polyamide binding had no adverse effect upon nucleosome stability or reconstitution.

The presence of a supergroove on the NCP containing two aligned minor groove binding sites for **PA1**, led to the hypothesis that two polyamides joined together with a linker of sufficient length should be able to bind a single supergroove as a linked dimer. Three turn-to-turn dimers were synthesized containing the same polyamide joined by linkers containing 2, 3, and 4 ethylene glycol units. All three clamps were shown to bind

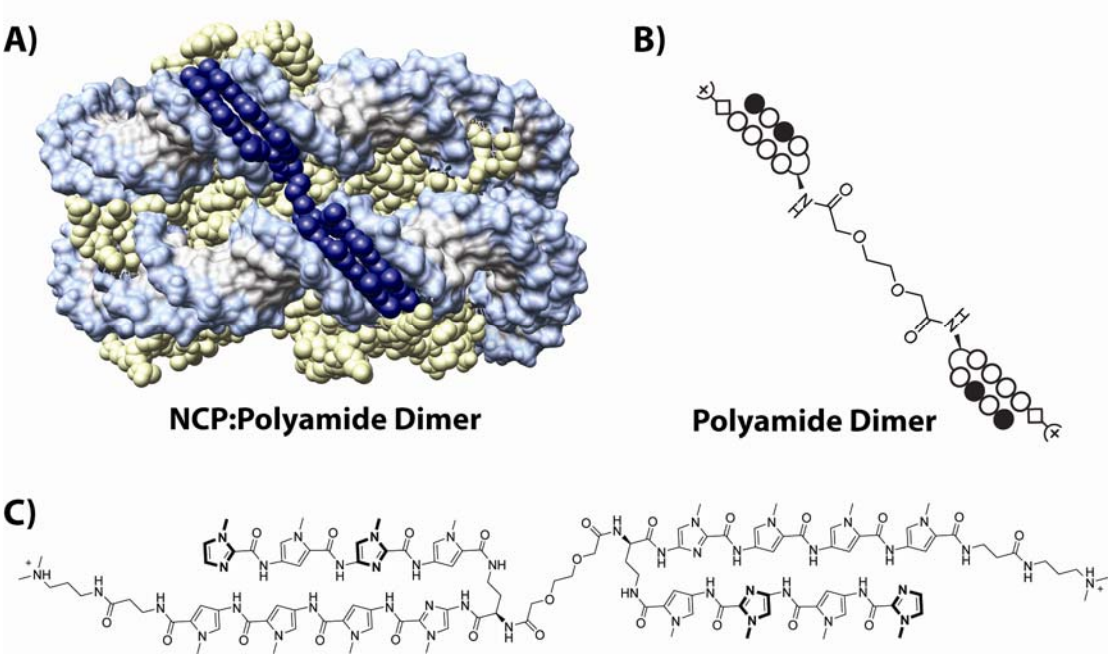

Figure 1.6 Crystal structure of the NCP bound by a turn-to-turn tandem polyamide. A) Structure of the NCP bound by a polyamide dimer. The polyamide is shown in navy blue. (PBD code 1S32) B) Ball-and-stick model of the turn-to-turn polyamide dimer, C) Chemical Structure of the polyamide dimer.

to the NCP with similar affinities in the low nanomolar range.[33] The crystal structure shows that the polyamide dimer binds in the predicted super groove with each polyamide aligned against its target DNA sequence and the PEG linker spanning the ~12 Å gap between the two gyres of DNA.[33] High resolution crystal structures were obtained showing the dimers bound in the supergroove of the NCP as predicted. The polyamide dimer was referred to as a "clamp" because of its ability to clamp the two gyres of DNA

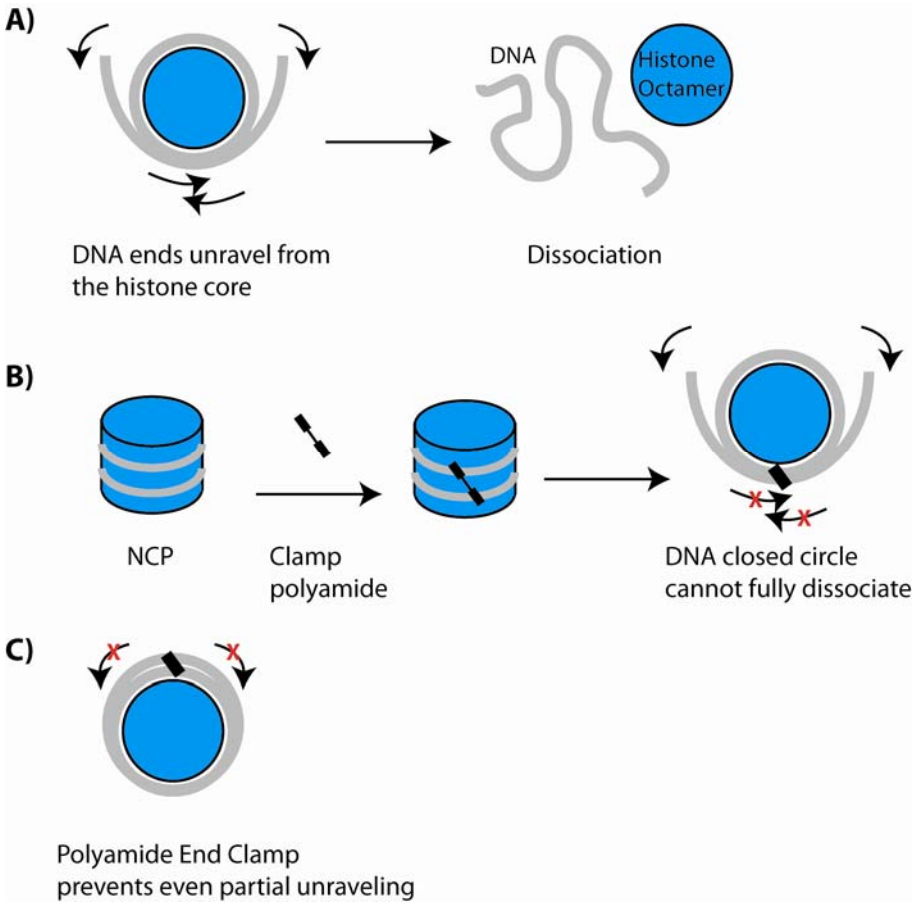

Figure 1.7 Diagram illustrating models for NCP dissociation. A) Model for the dissociation of the NCP into DNA and the histone proteins. B) The binding of the polyamide clamp to the NCP creates a full circle of DNA which is thought to prevent full unraveling and to explain the observed increase in NCP stability. C) A polyamide clamp targeted to the ends of the DNA could be able to prevent even partial unraveling.

on the NCP together, resulting in an increase in NCP stability. This is presumably due to the formation of a closed circle of DNA around the histone core which would be unable to fully unravel as illustrated in Figure 1.6B.

One unexpected result of this study was that the clamp polyamide in complex with the NCP resulted in the growth of extremely large, well-ordered crystals. This is presumed to be a direct effect of the stabilization effect the clamp has on the NCP. The future use of polyamide clamps to stabilize the NCP, could prove useful for generating high quality crystals needed for future crystallography studies of nucleosomes and chromatin.

1.4 Scope of Work

As outlined in the previous sections, polyamides have been previously used to target both B-form DNA, and DNA present in the nucleosome core particle, a biological DNA architecture. The uniting theme of the work described in this thesis has been the expansion of the scope of DNA architectures that can be targeted using sequence-specific DNA-binding polyamides. Primary investigations into using polyamides to target a DX array, a two-dimensional DNA nanostructure, are described in Chapter 2. These studies were the first attempts at using polyamides to target a higher-order DNA structure that can be used for 2-dimensional array assembly. Affinity cleavage experiments first demonstrated the ability of polyamides to target a variety of binding sites on a single DX tile. Additional AFM studies demonstrated the use of polyamides in recruiting protein to 2-dimensional DX arrays. In addition, several polyamide-biotin conjugates were used to

assemble proteins on a four component DX array in distinct patterns. In Chapter 3, we examine another DNA nanostructure, DNA origami, and the ability of polyamide conjugates to target these structures. Finally, Chapter 4 describes the work done in targeting nucleosomes, a biological DNA architecture, with polyamide dimers. The use of novel polyamide dimers to further stabilize the NCP, as well as the use of the NCP as a unique template for chemical reactions is examined. It is hoped that the work described herein will expand the use of polyamides from chemical agents for gene regulation, to tools for molecular biology, and reagents for the creation of complex molecular assemblies.

1.5 References

1. Venter, J. C., et al., The sequence of the human genome *Science* **2001,** 291, 1304-1351.

2. Lander, E. S., et al., Initial sequencing and analysis of the human genome *Nature* **2001,** 409, 860-921.

3. Dickerson, R. E., et al., The anatomy of A-DNA, B-DNA, And Z-DNA *Science* **1982,** 216, 475-485.

4. Pelton, J. G.; Wemmer, D. E., Binding modes of distamycin A with d(CGCAAATTTGCG)2 determined by 2-dimensional NMR *J. Am. Chem. Soc.* **1990,** 112, 1393-1399.

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

6. Dervan, P. B., Molecular recognition of DNA by small molecules *Bioorg. Med. Chem.* **2001,** 9, 2215-2235.

7. Dervan, P. B.; Edelson, B. S., Recognition of the DNA minor groove by pyrrole-imidazole polyamides *Curr. Opin. Struct. Biol.* **2003,** 13, 284-299.

8. Kielkopf, C. L.; Baird, E. E.; Dervan, P. D.; Rees, D. C., Structural basis for G center dot C recognition in the DNA minor groove *Nat. Struct. Biol.* **1998,** 5, 104-109.

9. Pilch, D. S., et al., The thermodynamics of polyamide-DNA recognition: Hairpin polyamide binding in the minor groove of duplex DNA *Biochemistry* **1999,** 38, 2143-2151.

10. Kielkopf, C. L., et al., A structural basis for recognition of A center dot T and T center dot A base pairs in the minor groove of B-DNA *Science* **1998,** 282, 111-115.

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

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

13. Kelly, J. J.; Baird, E. E.; Dervan, P. B., Binding site size limit of the 2:1 pyrrole-imidazole polyamide-DNA motif *Proc. Natl. Acad. Sci. U. S. A.* **1996,** 93, 6981-6985.

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

15. Kers, I.; Dervan, P. B., Search for the optimal linker in tandem hairpin polyamides *Bioorg. Med. Chem.* **2002,** 10, 3339-3349.

16. Weyermann, P.; Dervan, P. B., Recognition of ten base pairs of DNA by head-to-head hairpin dimers *J. Am. Chem. Soc.* **2002,** 124, 6872-6878.

17. Gottesfeld, J. M., et al., Regulation of gene expression by small molecules *Nature* **1997,** 387, 202-205.

18. Olenyuk, B. Z., et al., Inhibition of vascular endothelial growth factor with a sequence-specific hypoxia response element antagonist *Proc. Natl. Acad. Sci. U. S. A.* **2004,** 101, 16768-16773.

19. Nickols, N. G.; Dervan, P. B., Suppression of androgen receptor-mediated gene expression by a sequence-specific DNA-binding polyamide *Proc. Natl. Acad. Sci. U. S. A.* **2007,** 104, 10418-10423.

20. Nickols, N. G.; Jacobs, C. S.; Farkas, M. E.; Dervan, P. B., Modulating hypoxia-inducible transcription by disrupting the HIF-1-DNA interface *ACS Chem. Biol.* **2007,** 2, 561-571.

21. Mapp, A. K.; Ansari, A. Z.; Ptashne, M.; Dervan, P. B., Activation of gene expression by small molecule transcription factors *Proc. Natl. Acad. Sci. U. S. A.* **2000,** 97, 3930-3935.

22. Stafford, R. L., et al., Minimization of a protein-DNA dimerizer *J. Am. Chem. Soc.* **2007,** 129, 2660-2668.

23. Stafford, R. L.; Dervan, P. B., The reach of linear protein-DNA dimerizers *J. Am. Chem. Soc.* **2007,** 129, 14026-14033.

24. Arndt, H. D., et al., Toward artificial developmental regulators *J. Am. Chem. Soc.* **2003,** 125, 13322-13323.

25. Aldaye, F. A.; Palmer, A. L.; Sleiman, H. F., Assembling materials with DNA as the guide *Science* **2008,** 321, 1795-1799.

26. Seeman, N. C., DNA in a material world *Nature* **2003,** 421, 427-431.

27. Luger, K., et al., Crystal structure of the nucleosome core particle at 2.8 angstrom resolution *Nature* **1997,** 389, 251-260.

28. Fu, T. J.; Seeman, N. C., Dna Double-Crossover Molecules *Biochemistry* **1993,** 32, 3211-3220.

29. Winfree, E.; Liu, F. R.; Wenzler, L. A.; Seeman, N. C., Design and self-assembly of two-dimensional DNA crystals *Nature* **1998,** 394, 539-544.

30. Luger, K., Structure and dynamic behavior of nucleosomes *Curr. Opin. Genet. Dev.* **2003,** 13, 127-135.

31. Gottesfeld, J. M., et al., Sequence-specific recognition of DNA in the nucleosome by pyrrole-imidazole polyamides *J. Mol. Biol.* **2001,** 309, 615-629.

32. Suto, R. K., et al., Crystal structures of nucleosome core particles in complex with minor groove DNA-binding ligands *J. Mol. Biol.* **2003,** 326, 371-380.

33. Edayathumangalam, R. S., et al., Molecular recognition of the nucleosomal "supergroove" *Proc. Natl. Acad. Sci. U. S. A.* **2004,** 101, 6864-6869.