CHAPTER ONE

Introduction to DNA Binding and Recognition

The text of this chapter was taken in part from a book chapter co-authored with Benjamin S. Edelson, Professor Joel M. Gottesfeld (Scripps), and Professor Peter B. Dervan (Caltech).

Dervan PB, Fechter EJ, Edelson BS, Gottesfeld JM; Regulation of Gene Expression with Pyrrole-Imidazole Polyamides, *Pseudo-peptides in Drug Discovery*, Nielsen, PE, Ed.; Wiley-VCH **2004**, 121-152.

Background

DNA has been widely recognized for more than fifty years as the storehouse of genetic information for all living organisms on earth.[1] Over time, evolutionary pressure created the linear combination with only four base pairs that direct more than 30,000 genes to fabricate the molecules of human life.[2,3] Normal gene expression producing these molecules is delicately regulated through the interplay of activators and repressors but can deviate into disease states through mutations or intervention of pathogens.[4] Small molecules that influence the tools driving abnormal gene expression would have powerful therapeutic applications.[5] One promising approach is to disrupt the regulatory protein-DNA interface with designed ligands that recognize a gene's signature DNA sequence.[6]

Structural Features of DNA

The structure of DNA contains four natural heterocyclic bases, adenine (A), thymidine (T), guanine (G), and cytosine (C), linked through deoxyribose-phosphate and hydrogen-bonded in an antiparallel fashion with a complementary strand.[7] The linear combination of pi-stacked bases creates a helical twist with a wide and shallow major groove, and a narrow and deep minor groove due to the asymmetric projection of the sugar-base linkage between complementary strands.[7] The floor of each groove consists of exposed base-pair edges to offer subtle electronic and formal differentiation of DNA sequences through the projection of hydrogen bond donors and accepters (Figure 1.1).

Many natural proteins utilize these features, in addition to van der Waals and electrostatic interactions, to bind specific DNA sequences in the major groove. For example, the zinc finger[8] and leucine zipper[9] motifs each target specific DNA sequences through noncovalent interactions with the protein side chains. For many years studies were

aimed at understanding natural DNA recognition of the major groove with proteins due to its relatively rich network of hydrogen bonding recognition elements. However, due to the complex nature of protein structure, the rational design of proteins to recognize any sequence of interest has remained a challenge.

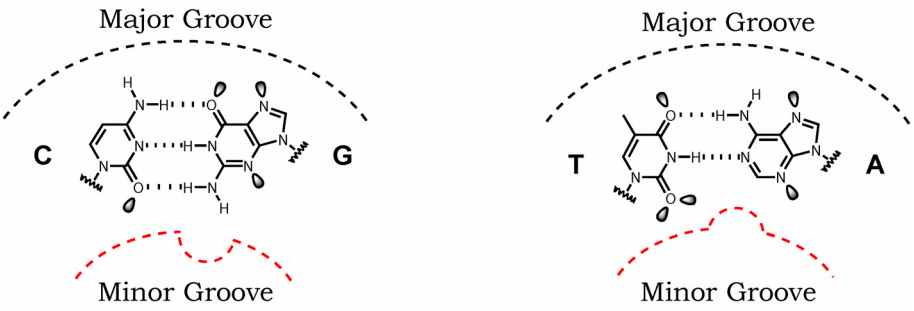


Figure 1.1 Functionality of DNA base pairs. All revealed H's are hydrogen bond donors and drawn orbitals are hydrogen bond acceptors. The top black outline indicates the major groove and the bottom red outline indicates the minor groove. The minor groove shows the protruding exocyclic NH_2 group of guanidinium and the asymmetric bump and the missing functionality of adenine as the asymmetric cleft.

Intercalation

Small molecules that intercalate DNA form an important class of widely studied compounds used for cancer chemotherapy.[10] Intercalators bind DNA by mimicking stacked base pairs and inserting their chromophores between two DNA base pairs (Figure 1.2).[11] During this process the deoxyribose-phosphate backbone of the DNA helix is unwound by 10

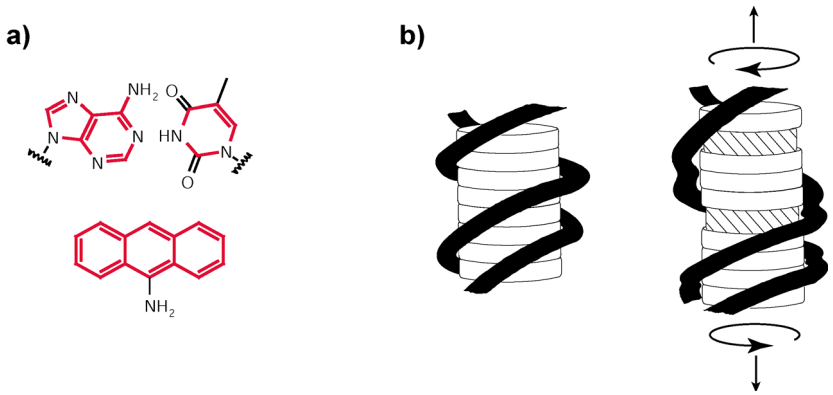


Figure 1.2 Lerman intercalation model. a) Structural similarities between hydrogen bonded base pairs and the flat aromatic fused ring system of an intercalator (9-aminoacridine shown). b) Cartoon model of standard B-form helix (left) and intercalated helix (right). Base-pairs and intercalators are represented as white and striped disks, respectively.

– 26 degrees, extended 3.4Å, and greatly stiffened to create a localized structural distortion that is usually confined to a few base pairs of the intercalation site. In addition to stacking interactions, several structurally complex intercalators (such as daunomycin and actinomycin D, Figure 1.3) exhibit additional stability from hydrogen bonding between appended sugars and the DNA base-pair edges, adding modest specificity to an otherwise non-specific class of DNA binding agents.

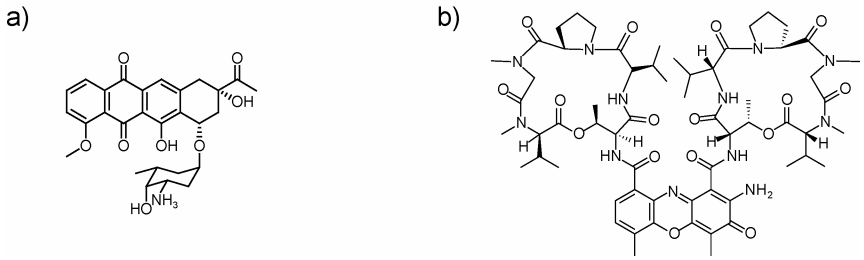


Figure 1.3 Intercalating natural products. a) Daunomycin and b) Actinomycin D.

Minor Groove-Binding Ligands

Several small molecule natural products, including calicheamicin,[12] Hoeschst 33258,[13] and distamycin,[14] have been shown to bind DNA with high affinity and modest sequence specificity within the minor groove. However, the potential to use small molecules for sequence-selective recognition was greatly enhanced when Wemmer and coworkers revealed by x-ray and NMR structural studies that the crescent-shaped distamycin molecule could bind A,T tracts as an antiparallel dimer (Figure 1.4).[15-17] This finding made sense of early results toward designing G-C selective binders[18] and raised the possibility for the design of minor groove binders that are selective for sequences containing mixed G-C and A-T base pairs.

After a 20-year search, the Dervan lab at Caltech demonstrated that synthetic analogs of the *N*-methylpyrrole (Py) carboxamide ring afford a set of heterocycles that can be

combined as unsymmetrical ring pairs in a modular fashion to recognize specifically a large repertoire of DNA sequences with affinities and specificities comparable to DNA-binding proteins.[19] The remainder of this chapter describes advances in the field of DNA-binding polyamides, cellular and nuclear uptake properties, and recent biological applications.

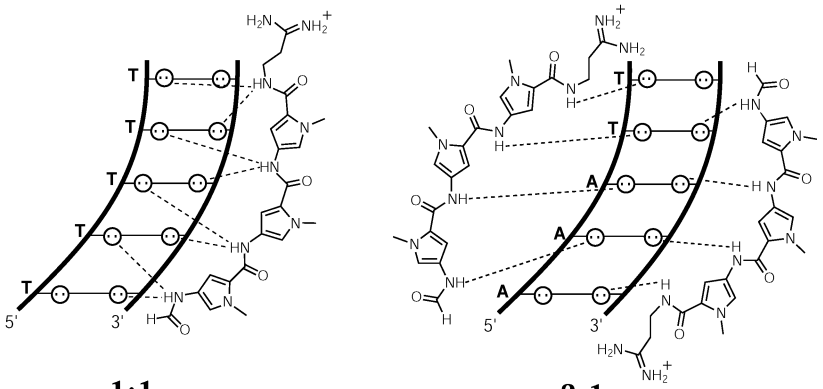


Figure 1.4 Schematic representation of the two modes of distamycin:DNA complexes with putative hydrogen bonds shown as dashed lines. Circles with dots represent lone pairs of N(3) of purines and O(2) of pyrimidines.

Pairing Rules

As mentioned, the four Watson-Crick base pairs can be differentiated on the minor groove floor by the specific positions of hydrogen bond donors and acceptors, as well as by differences in molecular shape (Figure 1.1).[19] The exocyclic NH_2 of guanine presents a bump on the edge of a G·C base pair, whereas a T·A base pair presents a cleft. A key study in the early 1990's demonstrated that the N-methylimidazole-(Im)-containing polyamide ImPyPy bound to the five bp sequence 5'-WGWCW-3' (where W = A or T).[18] This result was rationalized in terms of the formation of a 2:1 polyamide-DNA complex,[18] subsequently verified by NMR,[20] in which an antiparallel ring pairing of Im stacked against Py could specifically distinguish a G·C from a C·G base pair.

The Im/Py pair has been explored by extensive studies, including analyses of binding

in hundreds of different sequence contexts. Crystal structures confirmed the existence of a hydrogen bond between the Im nitrogen and the exocyclic NH_2 of guanine when the Im/Py pair binds opposite the G·C base pair.[21] The preference for a linear hydrogen bond, coupled with the unfavorable angle to an Im over the cytosine side of the base pair, provides a basis for the ability of an Im/Py pair to discriminate specifically G·C from C·G. These crystal structures also revealed other key ligand-DNA interactions, such as a series of hydrogen bonds between the amide groups of the polyamides and the edges of the bases on the adjacent DNA strand. Thermodynamic investigations dissected binding free energies into enthalpic and entropic contributions, revealing that the sequence selectivity of the Im/Py pair is driven by a favorable enthalpic contribution.[22]

Within the context of Watson-Crick base pair recognition by unsymmetrical heterocyclic ring pairs and informed by high resolution crystallographic data from a polyamide-DNA complex, the N-methyl-3-hydroxypyrrole (Hp) monomer was designed as a thymine-selective recognition element when paired across from Py (Figures 1.5 and 1.6).[23] It was anticipated for steric reasons that a substituent such as hydroxyl would not "fit" opposite A but would be accommodated at T. Crystal structures of two different Hp-containing polyamides, as their 2:1 complexes with DNA, have been determined at high resolution.[24, 25]

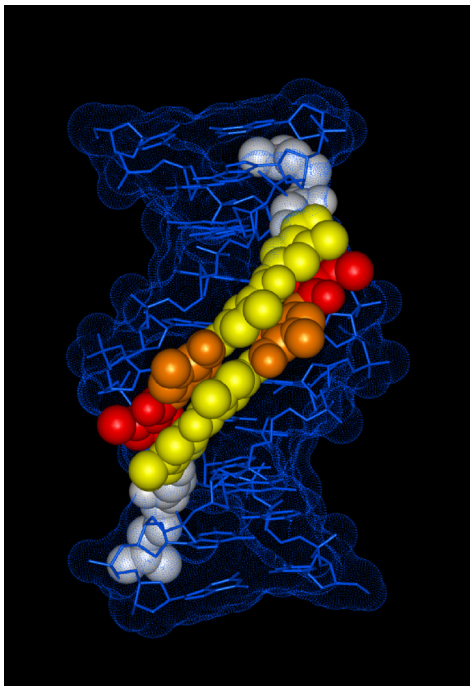


Figure 1.5 X-ray crystal structure of ImHpPyPy-*β*-Dp (Dp = dimethyl-amino propylamine) bound in a 2:1 complex with its target DNA site, 5'-AGTACT-3'.

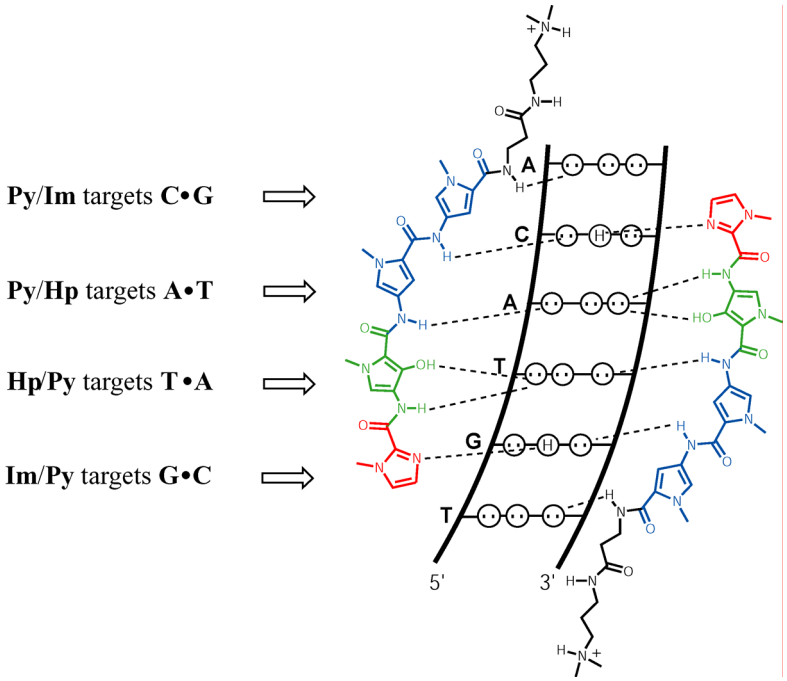


Figure 1.6 Pairing rules for polyamide recognition of all four Watson-Crick base pairs of DNA. Putative hydrogen bonds are shown as dashed lines. Circles with dots represent lone pairs of N(3) of purines and O(2) of pyrimidines, and circles containing an H represent the 2-amino group of guanine.

An Hp/Py pair was shown to distinguish T·A from A·T, G·C, and C·G base pairs using a combination of specific hydrogen bonds between the hydroxyl and the thymine-O2, along with shape selective recognition of an asymmetric cleft between the thymine-O2 and adenine-C2 (Figures 1.1 and 1.6). Together, three rings-Py, Im, and Hp-can be combined as unsymmetrical pairs to recognize specifically each of the four Watson-Crick base pairs; Im/Py is specific for G·C and Hp/Py for T·A. These interactions can be conveniently described as 'pairing rules' (Figures 1.6), which should be considered as guidelines for polyamide designs. Antiparallel polyamide dimers bind B-form DNA, and there are limitations regarding sequences targeted due to the sequence dependent microstructure of DNA. In certain cases of low-affinity polyamides, replacing the aromatic Py residue with an aliphatic β-Ala (β) residue enhances the affinity. The β/Im pair is specific for C·G, whereas the β/Py is specific for A,T over G,C base pairs.[26] Nonetheless, particular sequences (such as

polypurine tracts) have been difficult to recognize using polyamides.

Hairpin and Cycle Motifs

Covalently linking the two antiparallel polyamide strands results in molecules with increased affinity and specificity (Figure 1.7). Currently, the "standard" motif is the eight-ring hairpin, in which a γ-aminobutyric acid linker (γ-turn) connects the carboxylic terminus of one polyamide to the amino terminus of another. Compared to the unlinked homodimers, hairpin polyamides display ~100-fold higher affinity, with the γ-turn demonstrating selectivity for A,T over G,C base pairs, presumably due to a steric clash between the aliphatic turn unit and the exocyclic amine of guanine.[27] Eight-ring hairpins, which bind 6 bp, were shown to have affinity and sequence specificity similar to DNA-binding proteins (i.e. K_d < 1 nM).[28] NMR studies confirmed that the γ-turn locks the register of the ring pairings, preventing the ambiguity of slipped dimers.[29] Hairpin polyamides retain the orientation preferences of unlinked antiparallel polyamides, aligning N→C with respect to the 5'→3' direction of the adjacent DNA strand.[30]

For some hairpins, however, "reversed binding" (a C→N alignment of the polyamide with respect to the 5'→3' direction of the adjacent DNA strand) has been observed as the preferred orientation.[30] By introducing an amino substituent at the α position of the γ-turn, reversed binding is disfavored because of a steric clash between the amino substituent and the floor of the minor groove.[31] Furthermore, the chiral turn maintains the specificity of hairpins and increases the affinity. Covalently linking the C- and N-termini of a hairpin polyamide affords cyclic polyamides, which have slightly lower specificities but higher affinities for their target DNA sequences compared to analogous hairpin molecules bearing

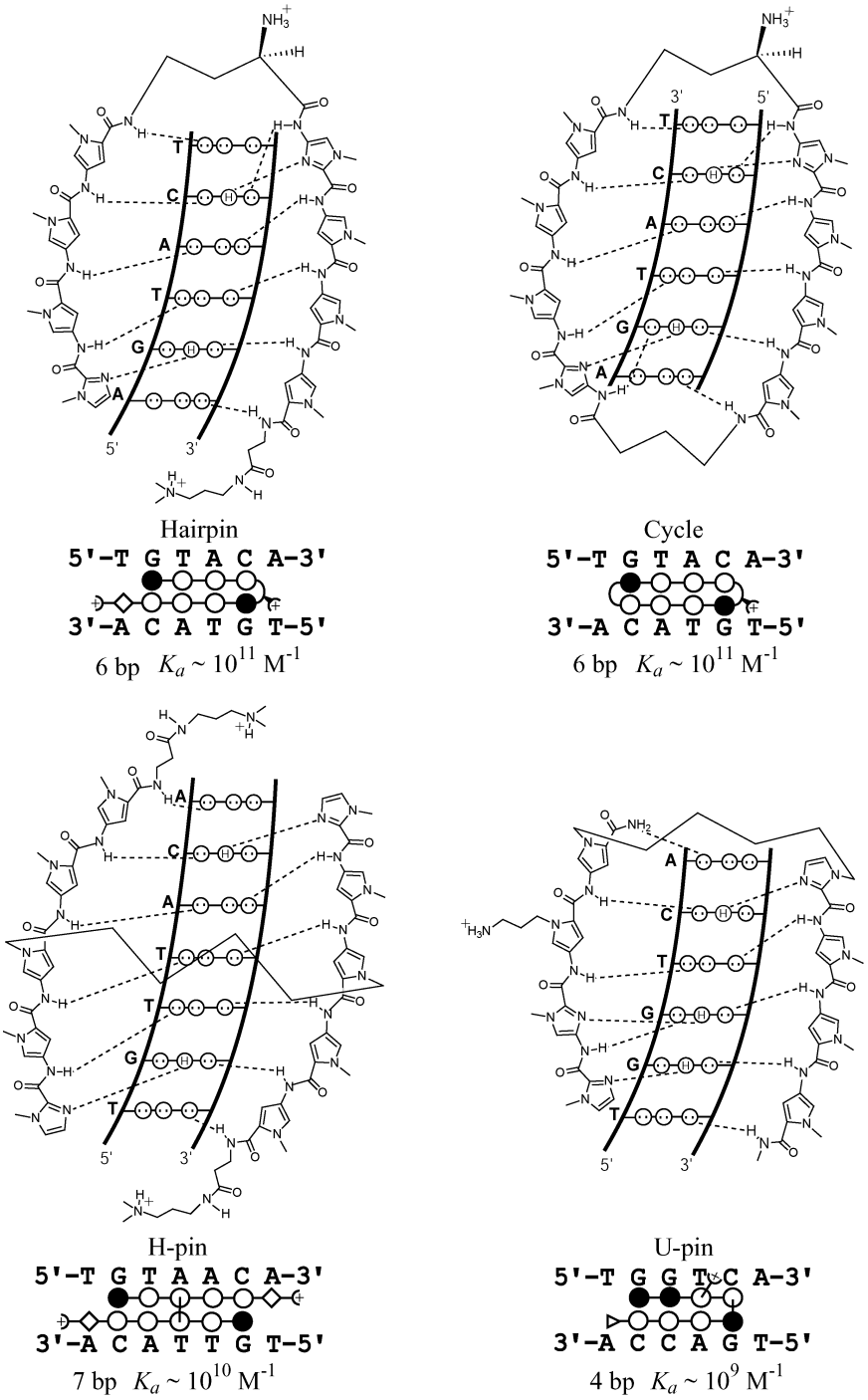


Figure 1.7 Polyamide-DNA binding motifs with equilibrium association constants (K_a). Hairpin: amino-substitution at the α-position of the γ-turn residue leads to enhanced binding affinity (10-fold) without loss of specificity, higher orientational selectivity, and offers potential for further substitution. Cycle: Cyclic polyamides show higher affinity than analogous hairpin molecules with the same number of cationic groups and eliminate all possibility of extended 1:1 binding. H-pin and U-pin: compared to their non-linked analogs, H-pins and U-pins exhibit higher binding affinity. The black and open circles represent Im and Py rings, respectively; diamonds represent *β*-alanine residues; and plus signs next to diamonds represent Dp residues. A curved line connecting the sides of two circles represents the γ-aminobutyric acid turn, and a curved line with a wedge and a plus sign represents the chiral $(R)^{H2N}$γ-turn. For the H-pin and U-pin, curved lines connecting the centers of two circles represent alkyl linkers attached to the N-methyl positions of the aromatic rings, a straight line with a plus sign projecting from the center of a ring represents a propylamine -$(CH_2)_3NH_3^+$ group, and a triangle represents a methyl amide. Other symbols are defined in Figure 1.6.

the same number of cationic groups (Figure 1.5).[32, 33]

H-Pin and U-Pin Motifs

Polyamides also can be linked, via the ring nitrogens, with an alkyl spacer that projects away from the minor groove. When placed in the center of a polyamide, the resultant branched molecule has been termed an H-pin; when placed at the end, a U-pin (Figure 1.7). H-pin polyamides bind with high affinity and good specificity, shown by ~50-fold lower affinity for single bp mismatch sites—that is, for sites constructed with one target Watson-Crick bp replaced by a disfavored bp.[34] Recent efforts to improve the synthetic methods for H-pins using alkene metathesis on a solid support have enabled a detailed study of the optimal alkyl linker length, demonstrating that four and six methylene units provide the highest affinities.[35] U-pin polyamides behave similarly.[36] The affinity of an eight-ring U-pin is most comparable to a six-ring hairpin polyamide, likely due to a loss of two hydrogen bond donors upon removal of the γ-turn element. Thus, the dimeric Py-Im U-turn element may be thought of as a C·G specific replacement for the γ-turn (Figure 1.7). In combination with removal of the β-Ala tail (selective for W), H-pin and U-pin polyamides could potentially bind purely G,C sites, a sequence type that has been difficult to target with other polyamide motifs.

Binding Site Size

For biological applications, binding site size may be critical because longer sequences would be expected to occur less frequently in the genome. Yet, beyond five contiguous rings, the binding affinity of polyamides decreases.[37] Crystal structures of polyamide-DNA complexes have shown that the polyamide rise per residue matches the pitch of the B-DNA

helix—that is, the spacing of the polyamide rings matches the spacing of the DNA base pairs.[21, 24] However, polyamides, which are inherently crescent-shaped, are slightly more curved than the minor groove of DNA, such that beyond five consecutive rings the shape of a polyamide is no longer complementary to DNA.[21]

The flexibility of β-Ala can be used to relax the curvature of polyamides, and molecules designed to bind as overlapped homodimers containing a β/β pair can recognize 11 bp of DNA with subnanomolar affinities (Figure 1.8).[38] Another motif utilizing dimerization to increase binding site size is the cooperative hairpin dimer (Figure 1.8).[39] Both of these motifs require a palindromic target site and have the potential to bind in non-

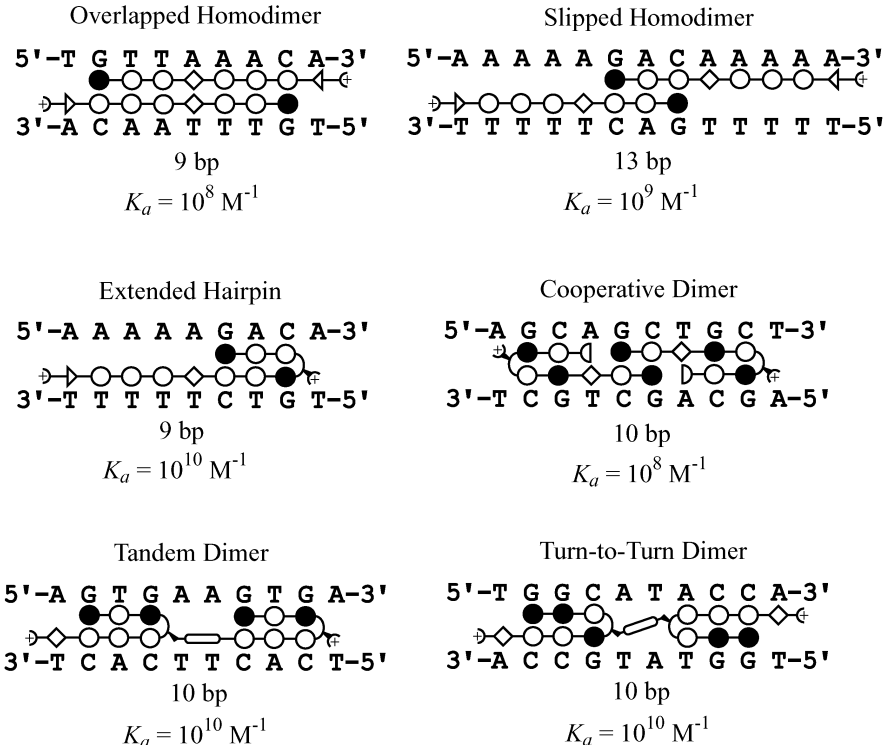


Figure 1.8 Polyamide-DNA binding motifs targeting longer DNA sequences. Overlapped and slipped homodimers: depending on the sequence context, six-ring polyamides with central β-Ala residues can bind to DNA as fully overlapped homodimers, recognizing 11 bp, or as slipped homodimers, recognizing 13 bp. Extended hairpin: extended conformation increases binding site size (to 9 bp) and enhances binding affinity. Cooperative dimer: a cooperatively binding hairpin polyamide can extend the binding site size to 10 bp without an increase in the molecular weight of the ligand. Tandem dimer: linked turn-to-tail, tandem polyamides recognize large DNA sequences with good specificity and excellent binding affinity. Turn-to-turn dimer: similar to tandem dimers, turn-to-turn dimers increase binding site size and affinity. Half circles represent ethanolamine -$NH(CH_2)_2OH$ groups, and a triangle represents glycine. Other symbols are defined in Figure 1.6.

cooperative modes, albeit with lower affinities. Hairpin dimers covalently linked turn-to-tail or turn-to-turn resolve both issues (Figure 1.8).[40-42] To link two six-ring hairpin polyamides turn-to-tail, 5-aminovaleric acid is the optimal aliphatic linker, furnishing tandem dimers that target 10 bp sites. Although an early example showed preferential binding to an 11 bp sequence, a broader study with additional examples showed that the valeric acid linker effectively balances affinity, selectivity for a 10 bp match site, and specificity over a double base pair mismatch site.[41] In addition, a tandem hairpin dimer connected by the ether linker 3-oxo-5-aminovaleric acid displayed a DNA-binding affinity equal to the valeric acid-linked compound but was more selective for the 10 bp binding site (Figure 1.8). The turn-to-turn dimer also binds longer sequences with subnanomolar affinity, with a four-carbon linker proving optimal.[42] For both types of hairpin dimers, however, specificity (expressed in terms of affinity for match over single base pair mismatch sites) is often poor, and remains a challenge. Nonetheless, an impressive application of tandem polyamide dimers was demonstrated by Laemmli and co-workers, who employed tandem hairpins with a dioxa-PEG linker to stain insect or vertebrate telomeres, $(TTAGG)_n$ or $(TTAGGG)_n$ repeats, respectively, with remarkable selectivity in fixed cells and chromosome spreads.[43]

Solid Phase Methods

The investigation of minor groove-binding polyamides was greatly accelerated by the implementation of solid-phase synthesis.[44] Originally demonstrated on Boc-β-Ala-PAM resin with Boc-protected monomers, it was also shown that Fmoc chemistry could be employed with suitably protected monomers and Fmoc-β-Ala-Wang resin (Figure 1.9).[45]

While allowing rapid preparation of a range of polyamides, these resins install a T,A selective β-Ala residue at the C-terminus, which places limits on the DNA sites that can be

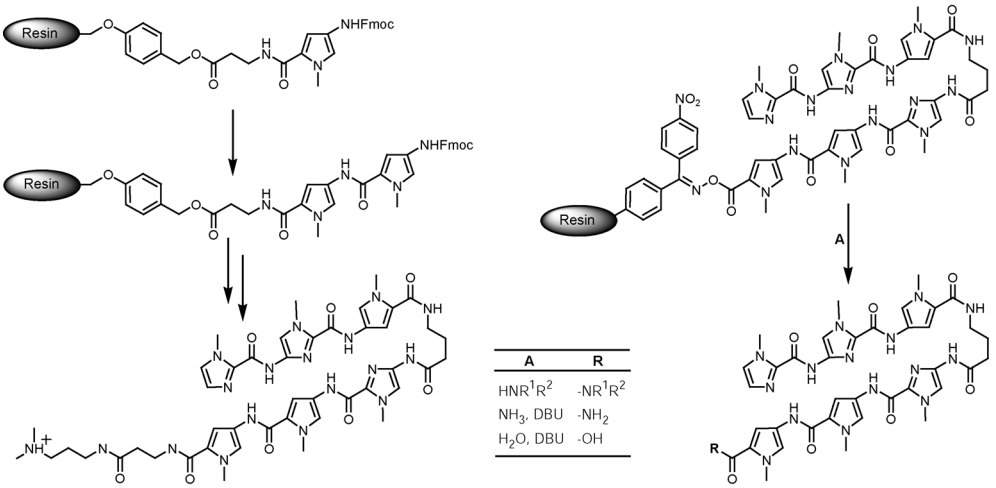


Figure 1.9 Variations to solid phase synthesis of polyamides. Use of Fmoc monomers on β-Ala-Wang resin (left) provides polyamides containing a β-alanine residue near the C-termini. Polyamides synthesized on the Kaiser oxime resin (right) can have shorter C-terminal groups than molecules prepared on β-Ala-PAM resin. The amine HNR^1R^2 may be a primary or secondary alkyl amine.

targeted. The shortest tail available from these resins is a propanolamide, obtained by reductive cleavage. Polyamides prepared on Boc-Gly-PAM resin can be reductively cleaved to obtain ethanolamide tails, but it was expected that further truncation of the C-terminus would be necessary for tolerance of G,C at the tail position.[46] The Kaiser oxime resin was therefore adapted to polyamide synthesis, allowing the preparation of polyamides with shorter C-termini (Figure 1.9). These molecules display the desired tolerance for G,C bases while maintaining high affinities.[47]

Sequence-Specific Alkylation of DNA

The availability of sequence-specific DNA binding molecules led to the development of bifunctional polyamides that covalently react with the minor groove of DNA. Two classes of alkylating agents were conjugated to the hairpin 'turn' unit and bound proximal to their alkylation sites by a DNA binding polyamide (Figure 1.10).[48, 49] Polyamide-*seco*-CBI conjugates showed alkylation at adenines proximal to the binding site within 12 h at nanomolar concentrations.[49] Likewise, polyamide-chlorambucil conjugates specifically

alkylate predetermined sites in the minor groove, but with slightly lower alkylation at mismatch sites than the CBI counterparts.[48] The slower rate of alkylation for the chlorambucil moiety in the minor groove likely allows for an increased specificity of alkylation to polyamide match sites. This class of sequence-specific DNA binding alkylators also inhibits polymerase elongation during transcription.

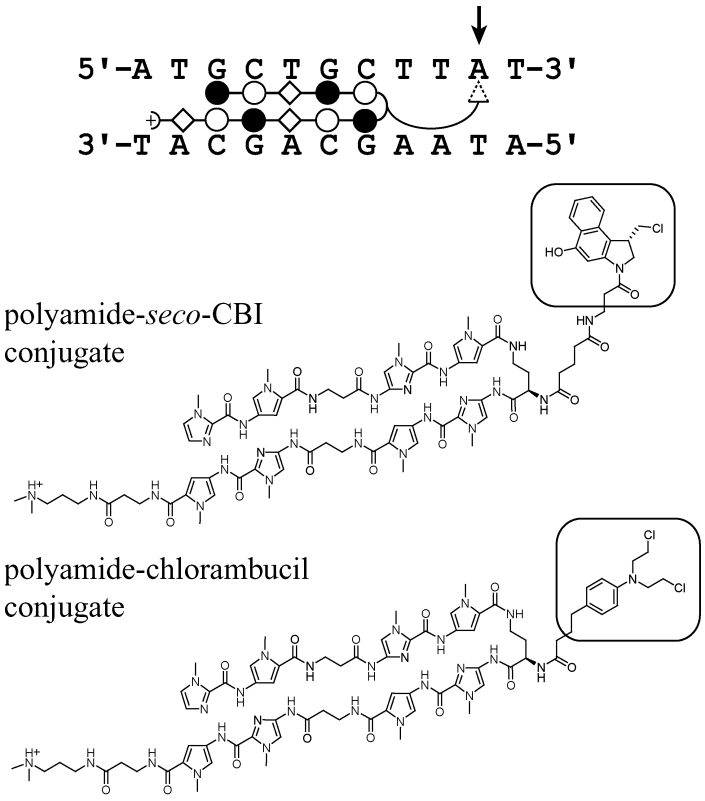
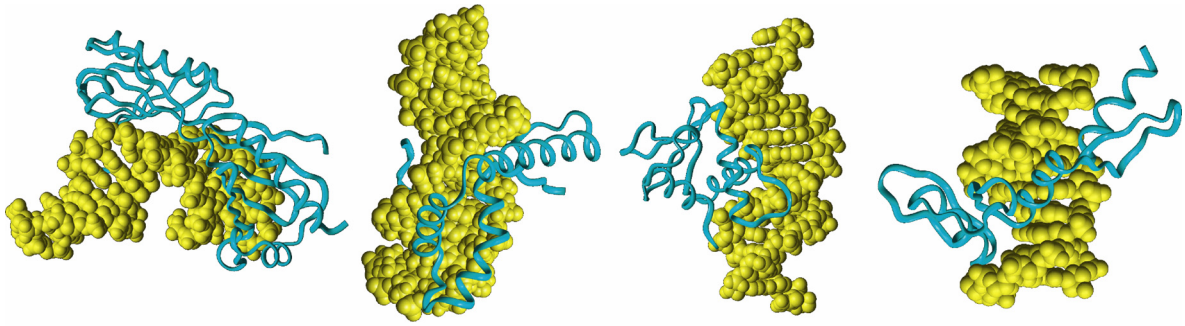


Figure 1.10 Alkylating polyamide binding site model and structures of polyamide-alkylator conjugates. The hatched triangle represents the alkylating agent. All other symbols are defined in Figure 1.6. The CBI and chlorambucil alkylator domains are boxed.

Inhibition of Gene Expression

Polyamides bind with high affinity to a wide range of DNA sites and often competitively displace proteins from DNA. One approach to modifying gene expression involves inhibition of key transcription factor (TF)-DNA complexes (Figures 1.11) in a designated promoter, thus interfering with recruitment of RNA polymerases.[50] Significantly, because there are considerably fewer oncogenic TFs than potentially oncogenic signaling proteins, TF inhibition represents a uniquely promising approach to cancer treatment.[51] The transcription factor TFIIIA was chosen as a first target because it regulates a relatively small number of genes and because the contacts between the nine zinc-finger protein and the minor groove had been established. A polyamide bound in the recognition site of TFIIIA suppressed transcription of 5S RNA genes by RNA polymerase III *in vitro* and



TBP LEF-1 Ets-1 Zif268

Figure 1.11 Examples of four different protein-DNA complexes that have been inhibited by polyamides. TBP is a minor groove-binding protein, LEF-1 is an HMG box, Ets-1 is a winged-helix-turn-helix, and Zif268 is a zinc finger.

in cultured *Xenopus* kidney cells.[50] Further studies used polyamides in combination with recombinant derivatives of TFIIIA subunits to elucidate essential minor groove contacts for the binding of this TF.[52]

Polyamides were then used to target viral genes transcribed by RNA polymerase II (Figure 1.12). The HIV-1 enhancer/promoter contains binding sites for multiple transcription factors, including TBP, Ets-1, and LEF-1. Two hairpin polyamides designed to bind DNA sequences immediately adjacent to the binding sites for these TFs specifically inhibited binding of each transcription factor and HIV-1 transcription in a cell-free assay.[53] In human blood lymphocytes, treatment with the two polyamides in combination inhibited viral replication by 99%, with no significant decrease in cell viability. Inhibition of viral replication is *indirect* evidence for specific transcription inhibition by polyamides because other modes of action could be involved such as modulation of T-cell activation pathways. However, RNase protection assays indicated that the two polyamides did not alter the RNA transcript levels of several cytokine and growth factor genes, suggesting that polyamides do affect transcription directly.

In a unique example of altering gene expression by modifying chromatin structure in a complex organism, Laemmli and coworkers targeted satellite regions of *Drosophila*

chromosomes with polyamides.[54] Remarkably, polyamides induced specific gain- and loss-of-function phenotypes when fed to developing *Drosophila* embryos.

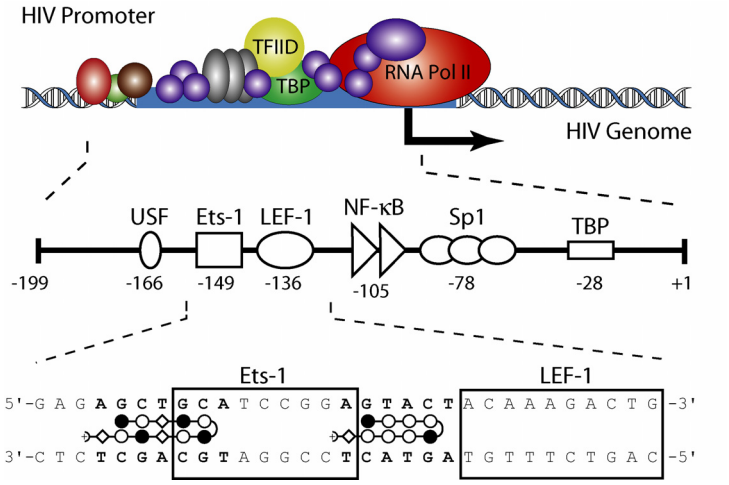


Figure 1.12 Polyamide inhibition of HIV-1 replication. The HIV promoter is expanded to show the binding sites of crucial transcription factors. The schematic is further expanded to show the binding sites for two polyamides that target the flanking regions of the Ets-1 transcription factor binding region (boxed). All other symbols are defined in Figure 1.6.

Gene Activation

Polyamides can upregulate transcription by inhibition of a repressor protein (derepression) or by recruitment of transcriptional machinery. For example, a hairpin polyamide was shown to block binding of the repressor IE86 to DNA, thereby upregulating transcription of the human cytomegalovirus MIEP.[55] A more complex case involves derepression of the integrated HIV-1 long terminal repeat (LTR). The human protein LSF binds in the promoter region at the LTR and recruits YY1, which then recruits histone deacetylases (HDACs). HDACs subsequently maintain LTR quiescence, which has been implicated in HIV latency by maintaining a silent stock of pathogen. Three different live cell models demonstrated that polyamides can inhibit LSF binding and increase expression of integrated HIV-1 promoter (Figure 1.13).[56] As with other systems, only polyamides matched to the correct protein binding site induced significant effects. Several existing drug

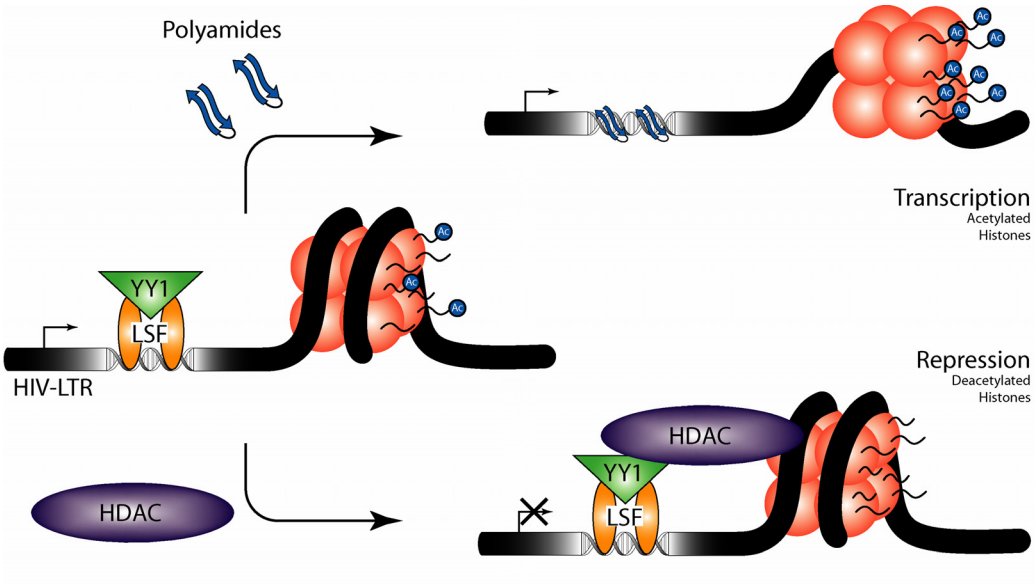


Figure 1.13 Dynamic model for polyamide intervention in LSF_2/YY1 induction of latency. The host factor LSF (orange oval) is shown bound to the HIV long terminal repeat (LTR) and recruits YY1 (green triangle) followed by HDAC (blue oval), mediating LTR repression by deacetylating histones (red circles). LSF is blocked upon binding of polyamides (connected blue arrows), resulting in increased LTR expression.

treatments can reduce HIV-1 levels in the blood to below detectable amounts, yet the virus inevitably returns in infected patients. Derepression by inhibition of LSF-DNA binding may eventually allow HIV to be fully eradicated by drug treatments. This approach is particularly promising because LSF is a human protein, which could make the target less susceptible to resistance by HIV-1 mutations.

Recruitment of transcriptional machinery is a fundamentally different approach to gene activation. Polyamides can be thought of as artificial DNA binding domains that can be linked to an activation domain (Figure 1.14). Such artificial transcription factors have been synthesized and evaluated in cell-free transcription assays.[57, 58] A hairpin polyamide tethered by a 36-atom straight chain linker to the short (20-residue) peptide activation domain AH gives robust activation of transcription with a size of only 4.2 kDa. Replacing the AH peptide with the shorter yet more potent activator VP2 derived from the activator domain of the viral activator VP16 and reducing the linker from 36 to 8 atoms provided a "minimal"

polyamide-peptide conjugate, 3.2 kDa in size, which activated transcription slightly more effectively than the larger analog.[58]

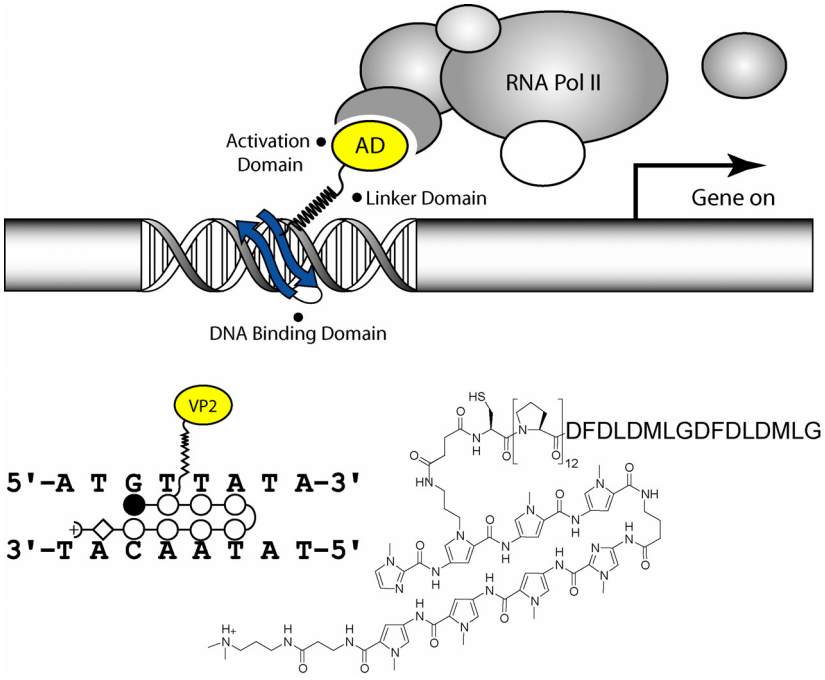


Figure 1.14 Activation of gene transcription by artificial transcription factors. (Top) The artificial activator is composed of three separate functional domains. The DNA binding domain consists of the pyrrole/imidazole polyamides (shown as connected blue arrows). A tethered linker domain (shown as a coil) connects the DNA binding domain to the peptide activation domain (AD, shown as a yellow oval). Upon binding, the artificial transcription factor recruits the necessary transcriptional machinery for gene activation. (Bottom-left) Ball-and-stick model for a polyamide conjugated to the VP2 activation domain. Symbols are as in Figure 1.6. (Bottom-right) Structure of the polyamide-VP2 conjugate with the polyproline linker domain in brackets.

Nucleosomes

In eukaryotic cells, DNA is tightly packaged by compaction into chromatin, and changes in chromatin structure can alter the accessibility of specific sequences and affect components of the molecular machinery in the nucleus. The fundamental repeating unit of chromatin is the nucleosome, comprising a 20-80 bp DNA linker region and the nucleosome core particle—roughly two tight superhelical turns of DNA (147 bp in length) wrapped around a disk of eight histone proteins. The ability of DNA-binding proteins to recognize their cognate sites in chromatin is restricted by the structure and dynamics of nucleosomal

DNA and by the translational and rotational positioning of the histone octamer. Using six different hairpin polyamides, it was shown that sites on nucleosomal DNA facing away from the histone octamer, or even partially facing the octamers, are fully accessible.[59] Remarkably, one section of 14 consecutive base pairs—more than a full turn of the DNA helix—was accessible for high affinity polyamide binding. The only positions very poorly bound by polyamides were sites near the amino-terminal tails of histone H3 or histone H4. Removal of either tail allowed polyamides to bind, suggesting that the structure of the DNA and perhaps its rotational position are strongly influenced by the N-terminal tails of histone H3 and H4.[59]

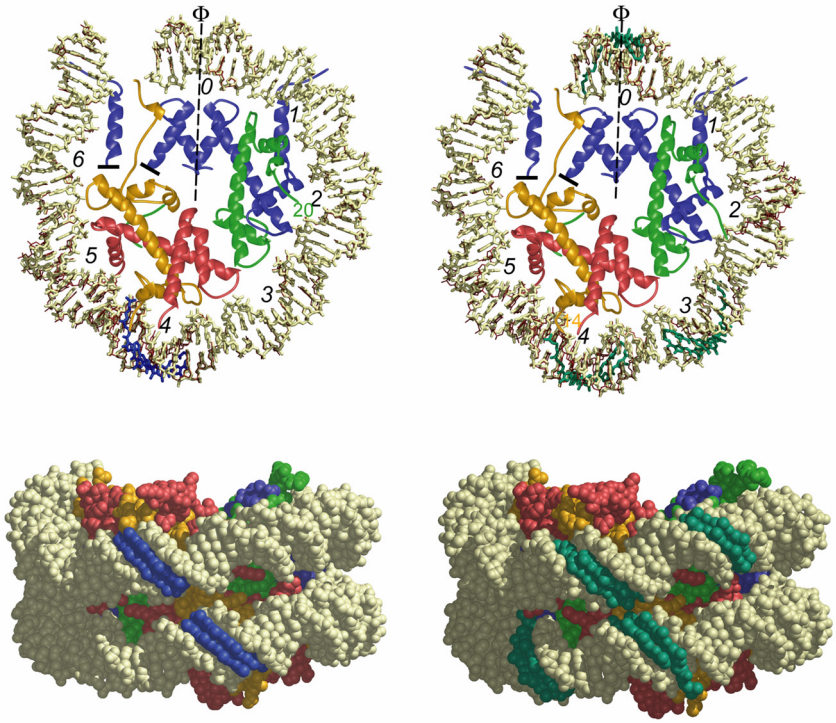


Figure 1.15 Nucleosome core particle (NCP)-polyamide cocrystal structures (PDB codes 1M18 and 1M19). (Top) Partial structure, viewed down the superhelical axis. Base-pairs 68–146 (shown in white) and associated proteins (H3, blue; H4, green; H2A, yellow; H2B, red) are shown for each complex. Superhelix locations (SHLs) are labeled as each major groove faces inward. The location of the dyad axis is indicated by a broken line, and the central base-pair (base-pair 73) is indicated by Φ. Bound polyamides are shown in blue (left) and green (right). (Bottom) Side-view of the structures in an orientation that best displays the bound polyamides is shown in a CPK representation.

Subsequently, the structures of three of these polyamide-nucleosome core particle complexes were determined by X-ray crystallography (Figure 1.15).[60] The histone octamer is unaffected by polyamide binding, but the nucleosomal DNA undergoes significant structural changes at the ligand binding sites and the adjacent regions. Significantly, distortions in DNA twist can propagate over long distances without disrupting histone-DNA contacts, giving a potential mechanistic rationale for the role of twist diffusion in nucleosome translocation. Although the three polyamides display very similar affinities for their binding sites in the α-satellite nucleosome particle, only the relatively non-specific polyamide ImPyPyPy-γ-PyPyPyPy-β-Dp inhibits temperature-induced nucleosome translocation.[60] This may indicate that ligand positioning is critical, such that a single properly placed polyamide would effectively block translocation; or that the small effects of a single bound ligand can be amplified, such that a combination of several different polyamides would block translocation.

Nuclear Uptake

DNA-binding polyamides can inhibit and influence a wide variety of protein-DNA interactions in solution, yet effectiveness in cell culture has proved to be dependent on cell type. A series of fluorescently labeled polyamides was prepared to analyze the intracellular distribution of these molecules in a panel of cell lines.[61] In cell types that had shown robust responses to polyamides, such as primary human T-cells, fluorescent polyamide-bodipy conjugates were observed to enter the nuclei of live cells.[61] However, in the majority of cell lines, polyamide-bodipy conjugates were excluded from the nucleus. Co-staining with organelle-specific dyes indicates that polyamide-bodipy conjugates are often trapped in lysosomes and other cytoplasmic vesicles,[62] such that cells treated with polyamides can give

a false nuclear signal upon fixing even if they are washed extensively. Bashkin and coworkers have demonstrated that a polyamide-bodipy conjugate will traffic to the nucleus a human cell line in the presence of verapamil, a p-glycoprotein inhibitor.[62] Importantly, it appears that specific combinations of linkers and fluorescein that enable the nuclear localization of polyamides in a broad range of mammalian cell lines in the absence of verapamil.[63] The uptake of polyamide-fluorescein conjugates appears to be dependent on both the polyamide structure and the cell line studied.[64]

Scope of this work

This thesis describes the design and synthesis of hybrid molecules that recognize and intercalate specific sequences of DNA. Chapter two illustrates some limitations of polyamides to inhibit transcription factor complexes and describes the initial design of polyamide-intercalator conjugates to overcome them. A series of polyamides of varying linker length were coupled to acridine and the DNA binding characteristics, including affinity, specificity, binding site size, and unwinding were studied. The series was then examined for its ability to inhibit the binding of two transcription factors: GCN4 and Sp1/Sp3, and its potential use as topoisomerase II inhibitors and downregulators of transcription *in vitro*. Alternative attachment of the linker and acridine was also investigated. Finally, compounds were synthesized to probe the cellular uptake properties of intercalator linked polyamides. Chapter three describes the design and synthesis of molecules that bis-intercalate and bind to DNA in a sequence-specific manor. The DNA binding characteristics including affinity, specificity, and unwinding were again investigated. Attempts were then made to understand the potency of polyamide bisintercalators compared with polyamide monointercalators. Chapter four details an additional polyamide-intercalator motif, modeled

after the antibiotic actinomycin D, which contains an intercalating chromophore flanked by two hairpin polyamides. Chapter five investigates applications of polyamide-intercalator conjugates to detect specific DNA sequences through fluorescent enhancement. The conjugates utilize fluorescence qualities of thiazole orange, which fluoresces only when bound to DNA. Finally, the appendix contains a somewhat distinct study utilizing polyamides to target the HTLV-1 promoter and inhibit viral replication.

References

1. Lewin, B., Genes V; Oxford University Press: New York, 1994.
2. Consortium, I. H. G. D., *Nature* **2001**, 409, 860.
3. Venter, J. C. et al., *Science* **2001**, 291, 1304.
4. Tjian, R., *Sci. Am.* **1995**, 2, 54.
5. Denison, C. and T. Kodadek, *Chem. Biol.* **1998**, 5, R129-R145.
6. Gottesfeld, J. M., L. Neely, J. W. Trauger, E. E. Baird, and P. B. Dervan, *Nature* **1997**, 387, 202-205.
7. Saenger, W., *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.
8. Pavletich, N. P. and C. O. Pabo, *Science* **1991**, 252, 809.
9. Ellenberger, T. E., C. J. Brandl, K. Struhl, and S. C. Harrison, *Cell* **1992**, 71, 1223.
10. Denny, W. A., *Anticancer Drug. Des.* **1989**, 4, 241-263.
11. Lerman, L. S., *J. Mol. Biol.* **1961**, 3, 18-30.
12. Kalben, A., S. Pal, A. H. Andreotti, S. Walker, D. Gange, K. Biswas, and D. Kahne, *J. Am. Chem. Soc.* **2000**, 122, 8403-8412.
13. Minehan, T. G., K. Gottwald, and P. B. Dervan, *Helv. Chim. Acta.* **2000**, 83, 2197-2213.
14. Arcamone, F., S. Penco, P. G. Prezzi, and V. Nicolella, A. Pirelli, *Nature* **1964**, 203, 1064.
15. Kopka, M. L, C. Yoon, D. Goodsell, and P. Pjura, and R. E. Dickerson, *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 1376-1380.
16. Coll, M., C. A. Frederick, A. H. J. Wang, and A. Rich, *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 8385-8389.
17. Pelton, J. G. and D. E. Wemmer, *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 5723-5727.

18. Wade, W. S., M. Mrksich, and P. B. Dervan, *J. Am. Chem. Soc.* **1992**, *114*, 8783-8794.

19. Dervan, P. B., *Bioorg. Med. Chem.* **2001**, *9*, 2215-2235.

20. Mrksich, M., W. S. Wade, T. J. Dwyer, B. H. Geierstanger, D. E. Wemmer, and P. B. Dervan, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 7586-7590.

21. Kielkopf, C. L., E. E. Baird, P. B. Dervan, and D. C. Rees, *Nat. Struct. Biol.* **1998**, *5*, 104-109.

22. Pilch, D. S., N. Poklar, E. E. Baird, P. B. Dervan, and K. J. Breslauer, *Biochemistry* **1999**, *38*, 2143-2151.

23. White, S., J. W. Szewczyk, J. M. Turner, E. E. Baird, and P. B. Dervan, *Nature* **1998**, *391*, 468-471.

24. Kielkopf, C. L., S. White, J. W. Szewczyk, J. M. Turner, E. E. Baird, P. B. Dervan, and D. C. Rees, *Science* **1998**, *282*, 111-115.

25. Kielkopf, C. L., R. E. Bremer, S. White, J. W. Szewczyk, J. M. Turner, E. E. Baird, P. B. Dervan, and D. C. Rees, *J. Mol. Biol.* **2000**, *295*, 557-567.

26. Turner, J. M., S. E. Swalley, E. E. Baird, and P. B. Dervan, *J. Am. Chem. Soc.* **1998**, *120*, 6219-6226.

27. Mrksich, M., M. E. Parks, and P. B. Dervan, *J. Am. Chem. Soc.* **1994**, *116*, 7983-7988.

28. Trauger , J. W., E. E. Baird, and P. B. Dervan, *Nature* **1996**, *382*, 559-561.

29. deClairac, R. P. L., B. H. Geierstanger, M. Mrksich, P. B. Dervan, and D. E. Wemmer, *J. Am. Chem. Soc.* **1997**, *119*, 7909-7916.

30. White, S., E. E. Baird, and P. B. Dervan, *J. Am. Chem. Soc.* **1997**, *119*, 8756-8765.

31. Herman, D. M., E. E. Baird, and P. B. Dervan, *J. Am. Chem. Soc.* **1998**, *120*, 1382-1391.

32. Herman, D. M., J. M. Turner, E. E. Baird, and P. B. Dervan, *J. Am. Chem. Soc.* **1999**,

121, 1121-1129.

33. Melander, C., D. M. Herman, and P. B. Dervan, *Chem.-Eur. J.* **2000**, *6*, 4487-4497.
34. Greenberg, W. A., E. E. Baird, and P. B. Dervan, *Chem.-Eur. J.* **1998**, *4*, 796-805.
35. Olenyuk, B., C. Jitianu, and P. B. Dervan, *J. Am. Chem. Soc.* **2003**, 125, 4741.
36. Heckel, A. and P. B. Dervan, *Chem. Eur. J.* **2003**, *9*, 1-14.
37. Kelly, J. J., E. E. Baird, and P. B. Dervan, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 6981-6985.
38. Swalley, S. E., E. E. Baird, and P. B. Dervan, *Chem.-Eur. J.* **1997**, *3*, 1600-1607.
39. Trauger, J. W., E. E. Baird, and P. B. Dervan, *Angew. Chem.-Int. Edit.* **1998**, *37*, 1421-1423.
40. Herman, D. M., E. E. Baird, and P. B. Dervan, *Chem.-Eur. J.* **1999**, *5*, 975-983.
41. Kers, I. and P. B. Dervan, *Bioorg. Med. Chem.* **2002**, *10*, 3339-3349.
42. Weyermann, P. and P. B. Dervan, *J. Am. Chem. Soc.* **2002**, *124*, 6872-6878.
43. Maeshima, K., S. Janssen, and U. K. Laemmli, *Embo. J.* **2001**, *20*, 3218-3228.
44. Baird, E. E. and P. B. Dervan, *J. Am. Chem. Soc.* **1996**, *118*, 6141-6146.
45. Wurtz, N. R., J. M. Turner, E. E. Baird, and P. B. Dervan, *Org. Lett.* **2001**, *3*, 1201-1203.
46. Swalley, S. E., E. E. Baird, and P. B. Dervan, *J. Am. Chem. Soc.* **1999**, *121*, 1113-1120.
47. Belitsky, J. M., D. H. Nguyen, N. R. Wurtz, and P. B. Dervan, *Bioorg. Med. Chem.* **2002**, *10*, 2767-2774.
48. Wurtz, N. R. and P. B. Dervan, *Chem. & Biol.* **2000**, *7*, 153-161.
49. Chang, A. Y. and P. B. Dervan, *J. Am. Chem. Soc.* **2000**, *122*, 4856-4864.
50. Gottesfeld, J. M., L. Neely, J. W. Trauger, E. E. Baird, and P. B. Dervan, *Nature* **1997**, *387*, 202-205.

51. Darnell, J. E., *Nat. Rev. Cancer* **2002**, *2*, 740-749.

52. Neely, L., J. W. Trauger, E. E. Baird, P. B. Dervan, and J. M. Gottesfeld, *J. Mol. Biol.* **1997**, *274*, 439-445.

53. Dickinson, L. A., R. J. Gulizia, J. W. Trauger, E. E. Baird, D. E. Mosier, J. M. Gottesfeld, and P. B. Dervan, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 12890-12895.

54. Janssen, S., O. Cuvier, M. Muller, and U. K. Laemmli, *Mol. Cell* **2000**, *6*, 1013-1024.

55. Dickinson, L. A., J. W. Trauger, E. E. Baird, P. Ghazal, P. B. Dervan, and J. M. Gottesfeld, *Biochemistry* **1999**, *38*, 10801-10807.

56. Coull, J. J., G. C. He, C. Melander, V. C. Rucker, P. B. Dervan, and D. M. Margolis, *J. Virol.* **2002**, *76*, 12349-12354.

57. Mapp, A. K., A. Z. Ansari , M. Ptashne, and P. B. Dervan, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 3930-3935.

58. Ansari, A. Z., A. K. Mapp, D.H. Nguyen, P. B. Dervan, and M. Ptashne, *Chem. & Biol.* **2001**, *8*, 583-592.

59. Gottesfeld, J. M., C. Melander, R. K. Suto, H. Raviol, K. Luger, and P. B. Dervan, *J. Mol. Biol.* **2001**, *309*, 615-629.

60. Suto, R. K., R. S. Edayathumangalam, C. L. White, C. Melander, J. M. Gottesfeld, P. B. Dervan, and K. Luger, *J. Mol. Biol.* **2003**, *326*, 371-380.

61. Belitsky, J. M., S. J. Leslie, P. S. Arora, T. A. Beerman, and P. B. Dervan, *Bioorg. Med. Chem.* **2002**, *10*, 3313-3318.

62. Crowley, K. S. et al., *Bioorg. Med. Chem. Lett*. **2003**, *13*, 1565-1570.

63. Best, T. P., B. S. Edelson, N. G. Nickols, and P. B. Dervan, *Proc. Natl. Acad. Sci. USA*, **2003**, 100, 12063.

64. Edelson, B. S., T. P. Best, B. Olenyuk, N. G. Nickols, R. Doss, S. Foister, A. Heckel, and P. B. Dervan, *Nucl. Acids Res.* **2004**, 32, 2802.