

Nuclear Localization of Polyamide-Fluorescein Conjugates in Cell Culture

Thesis by

Benjamin S. Edelson

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2005

© 2005

Benjamin S. Edelson

All Rights Reserved

Table of Contents

Table of Contents.....	iii
List of Figures and Tables.....	iv
Chapter 1 Introduction: Recognition of the DNA Minor Groove by Pyrrole- Imidazole Polyamides.....	1
Chapter 2 Nuclear Localization of Pyrrole-Imidazole Polyamide-Fluorescein Conjugates in Cell Culture.....	35
Chapter 3 Influence of Structural Variation on Nuclear Localization of DNA- Binding Polyamide-Fluorophore Conjugates.....	54

List of Figures and Tables

Chapter 1		Page
Figure 1	Pairing rules for recognition of DNA by polyamides	3
Figure 2	Schematic of polyamide:DNA interactions	4
Figure 3	X-Ray crystal structure of a polyamide:DNA complex	5
Figure 4	Polyamide motifs	7
Figure 5	Extended polyamide motifs	11
Figure 6	Novel 5-membered heterocyclic rings	13
Figure 7	Schematic of polyamide and DNA curvatures	14
Figure 8	Benzimidazole derivatives in polyamides	15
Figure 9	Solid phase methods for polyamide synthesis	16
Figure 10	Structures of DNA:protein complexes that have been inhibited by polyamides	19
Figure 11	Examples of DNA-binding proteins that have been inhibited by polyamides	19
Figure 12	Two approaches to gene regulation by polyamides	20
Figure 13	X-ray crystal structure of a polyamides bound to the NCP	25
 Chapter 2		
Figure 1	Structures of conjugates 1-22 and LysoTracker Red 23	42
Figure 2	Cellular localization images for 2 and 3	43
Figure 3	Colocalization of polyamide 3 and Hoechst in live cells	44
Figure 4	Cell Uptake Data for compounds 1-22	48

Figure 5	DNase I footprint titration with compound 3	49
Figure 6	Energy dependence of nuclear localization of compound 3	50

Chapter 3

Figure 1	Structures of conjugates 1-20	60
Figure 2	Uptake profiles for compounds 1-20	61
Figure 3	Structures of conjugates 21-36	65
Figure 4	Uptake profiles for compounds 21-36	66
Figure 5	Structures of conjugates 37-46	68
Figure 6	Uptake profiles for compounds 37-46	69
Figure 7	Structures of conjugates 47-62	70
Figure 8	Uptake profiles for compounds 47-62	72
Figure 9	Structures of conjugates 63-77	73
Figure 10	Uptake profiles for compounds 63-77	74
Figure 11	Structures of conjugates 78-89	76
Figure 12	Uptake profiles for compounds 78-89	77
Figure 13	Structures of conjugates 90-100	78
Figure 14	Uptake profiles for compounds 90-100	79

Chapter 1

Introduction: Recognition of the DNA Minor Groove by Pyrrole-Imidazole Polyamides

The text of this chapter was taken in part from a manuscript coauthored with Professor Peter B. Dervan (Caltech)

(Dervan, P. B.; Edelson, B. S. "Recognition of the DNA minor groove by pyrrole-imidazole polyamides" *Cur. Opin. Struct. Bio.*, **2003**, 13, 284

Introduction

Distamycin A is a crescent-shaped natural product that preferentially binds to (A,T) sequences in the minor groove of DNA as 1:1 and 2:1 ligand-DNA complexes [1,2]. Analogs of the *N*-methylpyrrole (Py) rings of these polyamides afford a set of five-membered heterocycles that can be combined—as unsymmetrical ring pairs—in a modular fashion to recognize predetermined DNA sequences with affinity and specificity comparable to DNA-binding proteins (Figures 1 and 2) [3,4]. We describe here recent advances in the field of DNA-binding polyamides, including structural verification of binding models, new heterocycles for recognition, cellular and nuclear uptake properties, and recent biological applications.

Pairing rules

In a formal sense, the four Watson-Crick base pairs (bps) can be differentiated on the minor groove floor by the specific positions of hydrogen bond donors and acceptors, as well as subtle differences in molecular shape [5]. A key study in the early 1990's demonstrated that an imidazole- (Im)-containing polyamide ImPyPy bound to the five bp sequence 5'-WGWCW-3' (where W = A or T) instead of the 5'-WGWWW-3' sequence, which would be expected for a 1:1 polyamide-DNA complex [6]. This surprising result was rationalized in terms of the formation of a 2:1 polyamide-DNA complex, subsequently verified by NMR [7], in which an antiparallel ring pairing of Im stacked against Py could specifically distinguish a G·C from a C·G base pair. This discovery pointed toward a new paradigm of unsymmetrical ring pairs for specific recognition in the DNA minor groove (Figure 1) [6].

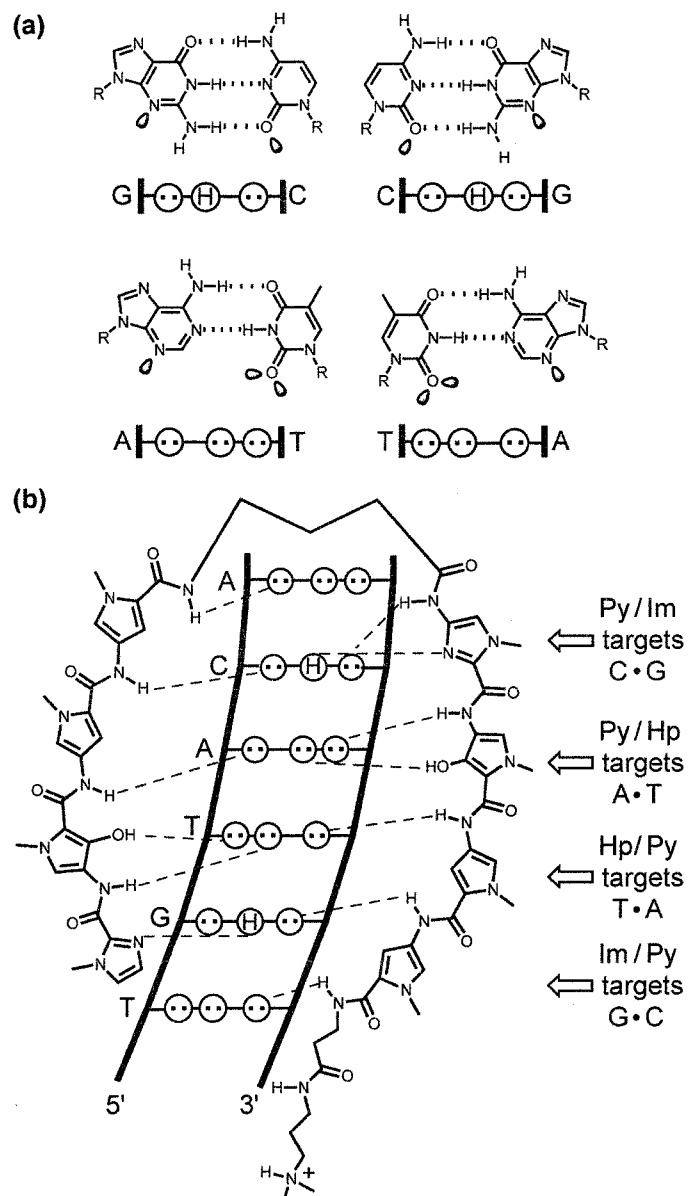


Figure 1. Molecular recognition of the minor groove of DNA. **(a)** Minor groove hydrogen bonding patterns of Watson-Crick bps. Circles with dots represent lone pairs of N(3) of purines and O(2) of pyrimidines, and circles containing an H represent the 2-amino group of guanine. The R group represents 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.

The Im/Py pair has been explored by extensive studies, including analyses of binding in hundreds of different sequence contexts, and crystal structures confirming the existence

of a hydrogen bond between the Im nitrogen and the exocyclic amine of guanine [8]. Moreover, the energetic 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. This crystal structure 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 (Figure 2). 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 [9].

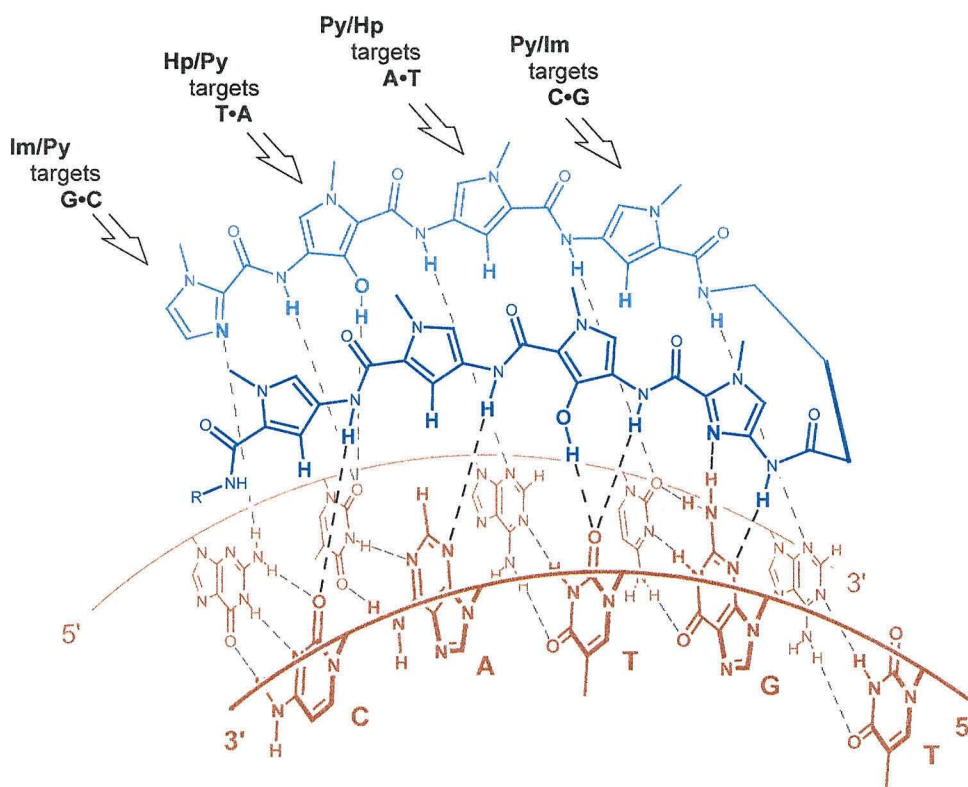


Figure 2. Schematic of a pyrrole-imidazole polyamide-DNA complex. Perspective view into the minor groove with the curvature lessened and the twist between the base pairs removed, illustrating hydrogen bonding patterns and shape complementarity. The R group represents β -Dp. This figure presents an alternative view of the complex shown in Figure 1 (b). The phosphate backbone for each DNA strand is indicated by a solid curved line with the 5' and 3' ends noted.

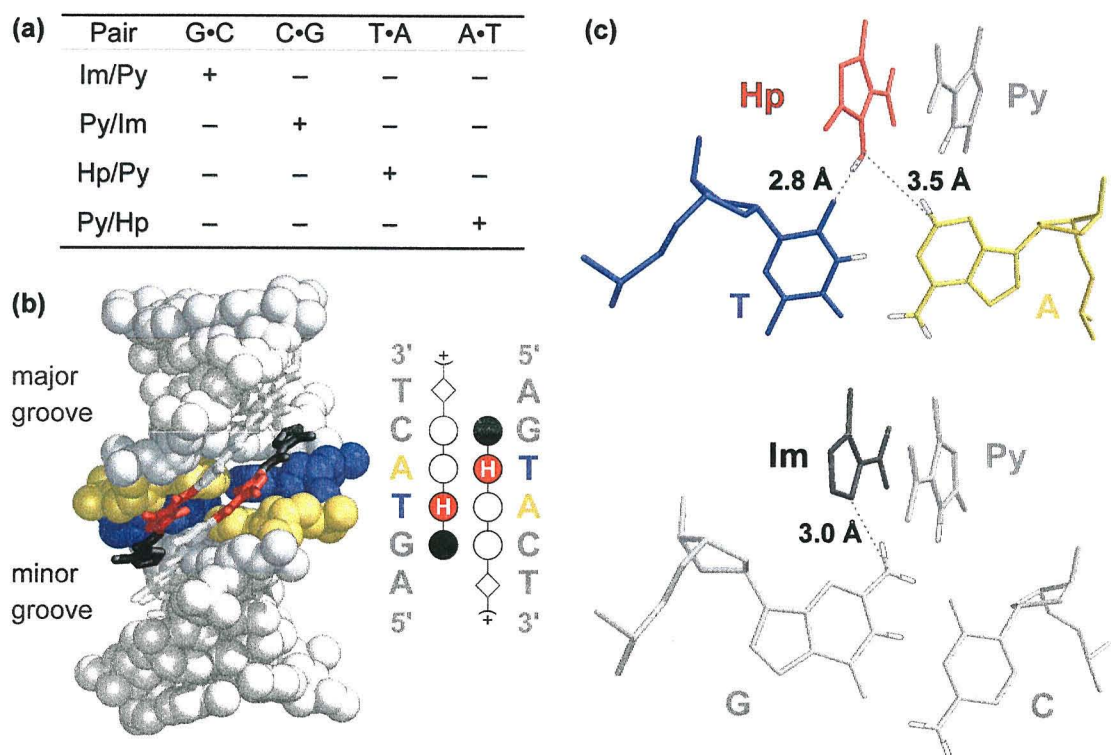


Figure 3. Pairing rules for minor groove recognition. **(a)** Table indicating pairing code for Py, Im, and Hp. Plus and minus signs indicate favored and disfavored interactions, respectively. **(b)** Space-filling model of (ImHpPyPy-β-Dp)₂·5'-CCAGTACTGG-3' (PDB code 407D). In the central binding region, adenosines are yellow and thymidines blue. Hp is red, Py is gray, and Im is black. A schematic is shown to the right: the black and open circles represent Im and Py rings, respectively; red circles with an H represent Hp rings; diamonds represent β-alanine residues; and plus signs next to diamonds represent Dp residues. **(c)** Detail views of the complex shown in (b), indicating structures of the Hp/Py pair interacting with the T•A base pair (top) and the Im/Py pair interacting with the G•C base pair (bottom). Dashed lines indicate interatomic distances between O, C, and N atoms. The Hp oxygen may form a favorable hydrogen bond with the adenine-C2–H. As in this case, C–H hydrogen bonds are strongest between aromatic carbons adjacent to nitrogen atoms with oxygen hydrogen bond acceptors [10]. Hydrogens involved in recognition were added using PC Spartan (Wavefunction, Inc.) and are colored white with gray outlines.

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 hydroxypyrrole (Hp) monomer was designed as a T-selective recognition element when paired across from Py (Figure 3) [5]. Crystal structures of two different Hp-containing polyamides, as their 2:1 complexes with DNA, have been determined at high resolution [10,11]. 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 (Figure 3). Hp polyamides bind with lower affinity than their Py counterparts [5], yet the structural basis for this difference has remained elusive. Partial melting of the T·A bp recognized by an Hp/Py ring pair was thought to be a possible explanation [10], but the distortion was observed in only one of two crystal structures. A consistent lengthening of the amide-DNA hydrogen bond on the C-terminal side of the Hp residue is observed in both structures, although this may be inadequate to account for the energetic differences [11]. A recent computational study by Liedl and coworkers argues that desolvation of the hydroxyl group upon insertion into the minor groove accounts for the energetic penalty [12]. Together, the rings Py, Im, and Hp can be combined 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 1-3).

Affinity and Specificity / Linked Motifs

Covalently linking the two antiparallel polyamide strands results in molecules with increased affinity and specificity. 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 (Figures 1 and 2). 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 with the exocyclic amine of guanine [13]. 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) [14]. As well, NMR studies confirmed that the γ -turn locks the register of the ring pairings, preventing the ambiguity of slipped dimers, in which the polyamide chains bind to DNA partially in the 2:1 mode and partially in the 1:1 mode [15]. Hairpin compounds retain the orientation preferences of extended polyamides, aligning N \rightarrow C with respect to the 5' \rightarrow 3' direction of the adjacent DNA strand [16].

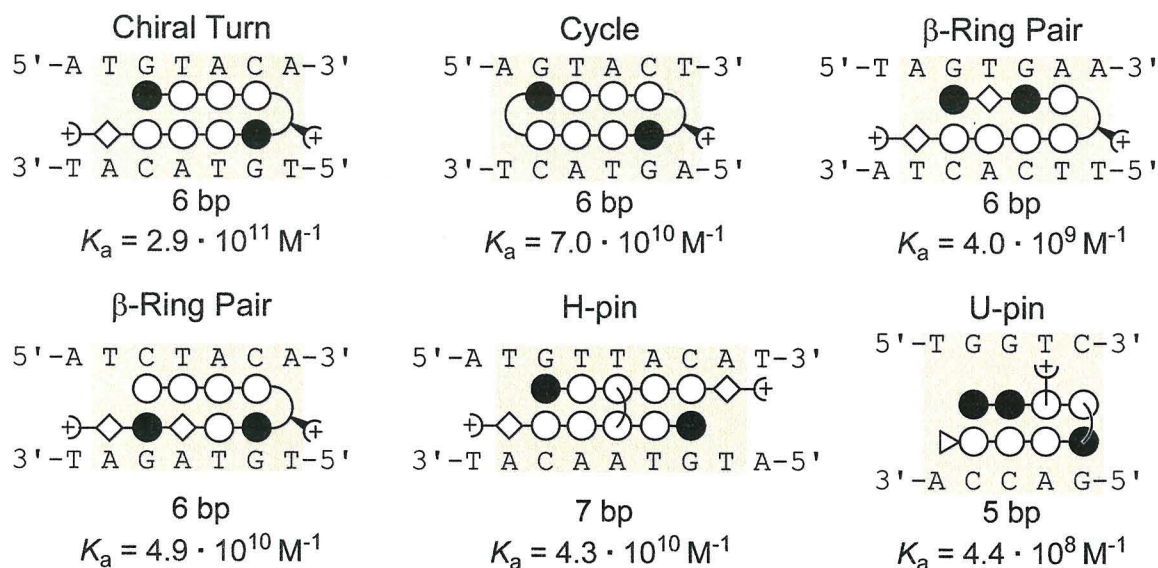


Figure 4. Polyamide-DNA binding motifs with equilibrium association constants (K_a). Chiral turn: amino-substitution at the α -position of the γ -turn residue leads to enhanced binding affinity (10-fold) without loss of specificity, higher orientational selectivity, and offers potential for further substitution [27]. Cycle: covalently linking the C and N termini of a hairpin polyamide eliminates all possibility of extended 1:1 binding. Cyclic polyamides show higher affinity than analogous hairpin molecules with the same number of cationic groups [21]. β /Ring Pair: the β -Ala residue relaxes the ligand curvature. In some cases, polyamides with a β /ring pair show improved affinity and specificity compared to analogs with a ring/ring pair [27,60]. H-pin and U-pin: compared to their non-linked analogs, H-pins and U-pins exhibit higher binding affinity ([23,24] and unpublished observations). 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 center 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 $-(\text{CH}_2)_3\text{NH}_3^+$ group, and a triangle represents a methyl amide.

For some hairpins, however, reversed binding (where there occurs a C \rightarrow N alignment of the polyamide with respect to the 5' \rightarrow 3' direction of the adjacent DNA strand) has been observed as the dominant orientation [16]. In other cases, the folded hairpin mode is

favored only ~5-fold over the linear 1:1 binding mode, in which the polyamide does not fold back on itself [17]. A solution to both problems comes from introducing an amino substituent at the α position of the γ -turn (Figure 4) [18]. Reversed binding is disfavored because of a steric clash between the amino substituent and the floor of the minor groove [18]; extended 1:1 binding is similarly destabilized by the chiral (*R*)^{H₂N} γ -turn, an effect that is amplified by acetylating the amine group [17]. Furthermore, the chiral turn maintains the specificity of hairpins and increases the affinity, most likely due to electrostatic interactions between the cationic amine group and the anionic DNA backbone. Substitution on a β -alanine-linker (β) has also been shown to influence the tendency of a polyamide to bind in an extended versus a hairpin conformation [19]. Covalently linking the C and N termini of a hairpin polyamide eliminates all possibility of extended binding, and such cyclic polyamides have slightly lower specificities but higher affinities for their target DNA sequences compared to analogous hairpin molecules bearing the same number of cationic groups (Figure 4) [20,21]. These differences can be attributed primarily to changes in the rates of dissociation from DNA, as the association rates for hairpin and cyclic polyamides are essentially diffusion-limited [22].

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 4). 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 [23]. Recent efforts to improve the synthetic methods for H-pins have enabled a detailed study of the optimal alkyl linker length, demonstrating that four and six methylene units provide the highest affinity [24]. U-pin polyamides behave similarly (unpublished observations). The affinity of an eight-ring U-pin is more comparable to a hairpin polyamide with six rather than eight rings, 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 (Figure 4) may be thought of as a C·G specific replacement for the γ -turn. In combination with removal of the β -Ala tail (see below), H-pin and U-pin polyamides could potentially bind purely G,C sites, a sequence type that it has not been possible to target with other polyamide motifs.

Certain DNA sequences (including G,C tracts) have been challenging sites for high affinity recognition with hairpin polyamides. For example, sites containing the sequence 5'-GNG-3' are often bound relatively poorly. Structural data has provided insight in certain cases. High-resolution crystal structures of different polyamide dimer-DNA complexes consistently display a large negative propeller twist in all targeted base pairs, an orientation that generally favors formation of three-center hydrogen bonds in the major groove [8,10,11]. Such intraduplex bonds cannot be formed in the sequences 5'-GCG-3' and 5'-GAT-3'. Accordingly, it has been problematic to target these sequences with high affinity.

In the case of 5'-GNG-3' sequences, replacing the aromatic Py residue with a flexible β -Ala (β) residue enhances the affinity (Figure 4). The β -Ala unit may allow the flanking

Im rings to orient better, while relieving the requirement for propeller twisted base pairs [11]. The β /Im pair is specific for C·G, whereas the β /Py and β / β pairs are specific for A,T over G,C base pairs [25,26]. For polyamides targeted to sites containing multiple 5'-GNG-3' sequences, binding enhancement is particularly dramatic upon incorporation of β /ring pairs: Im β ImPy- γ -Im β ImPy- β -Dp (Dp = dimethylaminopropylamine) binds the site 5'-TGCGCA-3' with 100-fold higher affinity than the purely ring-ring paired analog ImPyImPy- γ -ImPyImPy- β -Dp [26]. Incorporation of a β -Ala also enhanced the DNA-binding properties of a polyamide with an N-terminal pyrrole [27]. A Py/Im pair in this position generally displays poor selectivity for its target C·G base pair, but flanking the Im residue with a β -Ala improves the selectivity of the polyamide. For example, the molecule PyPyPyPy- γ -ImPy β Im- β -Dp binds its target site 5'-TCTACA-3' with subnanomolar affinity and 5- to 25-fold weaker binding to single bp mismatch sites (Figure 4).

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 [28]. Crystal structures of polyamide-DNA complexes have consistently 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 [8,10,11]. 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 [8].

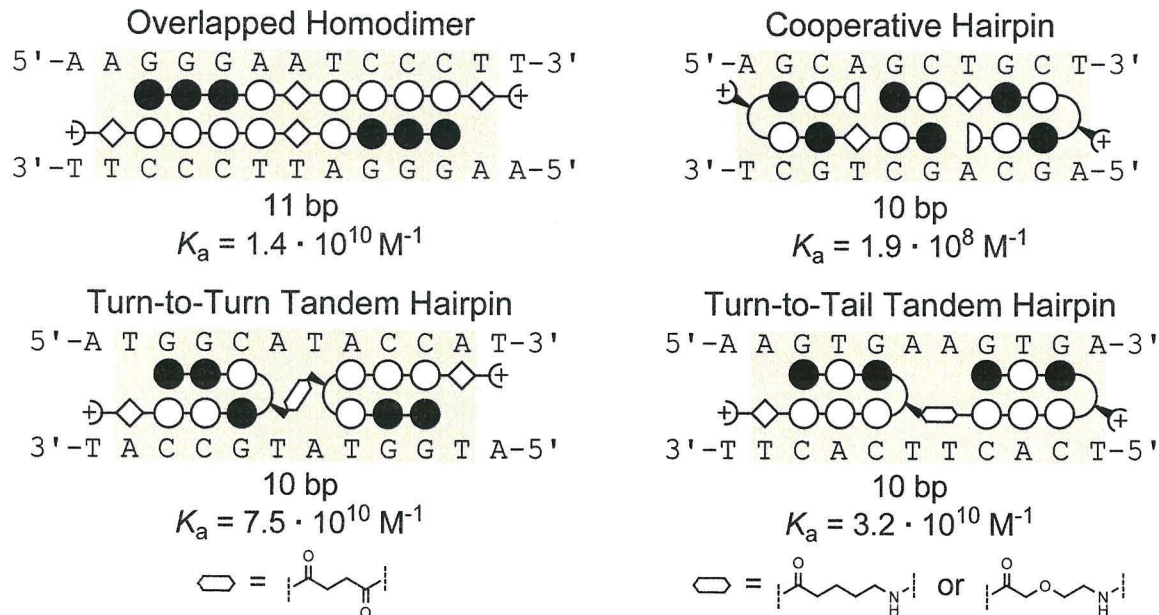


Figure 5. Schematic representation of extended polyamides at their binding sites. Overlapped homodimer: fully overlapping extended homodimer recognizing 11 bp [25]. Cooperative hairpin: a cooperatively binding hairpin polyamide can extend the binding site size to 10 bp without an increase in the molecular weight of the ligand [29]. Tandem-hairpins: linked either turn-to-turn or turn-to-tail, tandem polyamides recognize large DNA sequences with good specificity and excellent binding affinity [30,32]. For the turn-to-tail tandem hairpin, the 5-aminovaleric acid-linked polyamide and the 3-oxo-5-aminovaleric acid-linked polyamide display the same K_a . Half circles represent ethanolamine $\text{-NH(CH}_2\text{)}_2\text{OH}$ groups, and extended hexagons represent the linkers depicted at the bottom. Other symbols are defined in Figure 4.

The flexibility of β -Ala can be used to relax the curvature of polyamides, and molecules designed to bind as overlapped homodimers can recognize 11 bp of DNA with subnanomolar affinities (Figure 5) [25]. Another motif utilizing dimerization to increase binding site size is the cooperative hairpin [29]. Both of these motifs require a palindromic target site, and have the potential to bind in non-cooperative modes, albeit with lower affinities. Tandem hairpin polyamides, linked either turn-to-turn or turn-to-tail, resolve both issues (Figure 5) [30,31]. For linking two six-ring hairpin polyamides turn-to-tail, 5-aminovaleric acid is the optimal aliphatic linker, furnishing polyamides that target 10 bp sites [32]. Although a single early example showed preferential binding to

an 11 bp sequence [31], 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 bp mismatch site [32]. In addition, a tandem polyamide 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 5). A Gly-Im linker was only slightly less effective than a valeric acid linker and may allow targeting of sites with a G-C base pair in the linker region. The turn-to-turn tandem dimer also binds longer sequences with subnanomolar affinity, with a four-carbon linker proving optimal [30]. For both types of tandems, however, specificity (expressed in terms of affinity for match over single bp 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 dioxo-PEG linker to stain insect or vertebrate telomeres with remarkable selectivity in fixed cells and chromosome spreads [33].

Exploration of New Ring Systems for Minor Groove Recognition

To explore more broadly the structural landscape for minor groove recognition, a panel of five-membered aromatic heterocycles was synthesized and incorporated into DNA-binding polyamides (Figure 6) [34,35]. Because we have found that the Hp residue can degrade over time in the presence of acid or free radicals, a more robust T-selective element will be needed for biological applications [35]. Remarkably, none of the heterocycles tested (in the context of an eight-ring hairpin) revealed major advantages over conventional pyrrole-imidazole polyamides. Although ring pairings of Py with

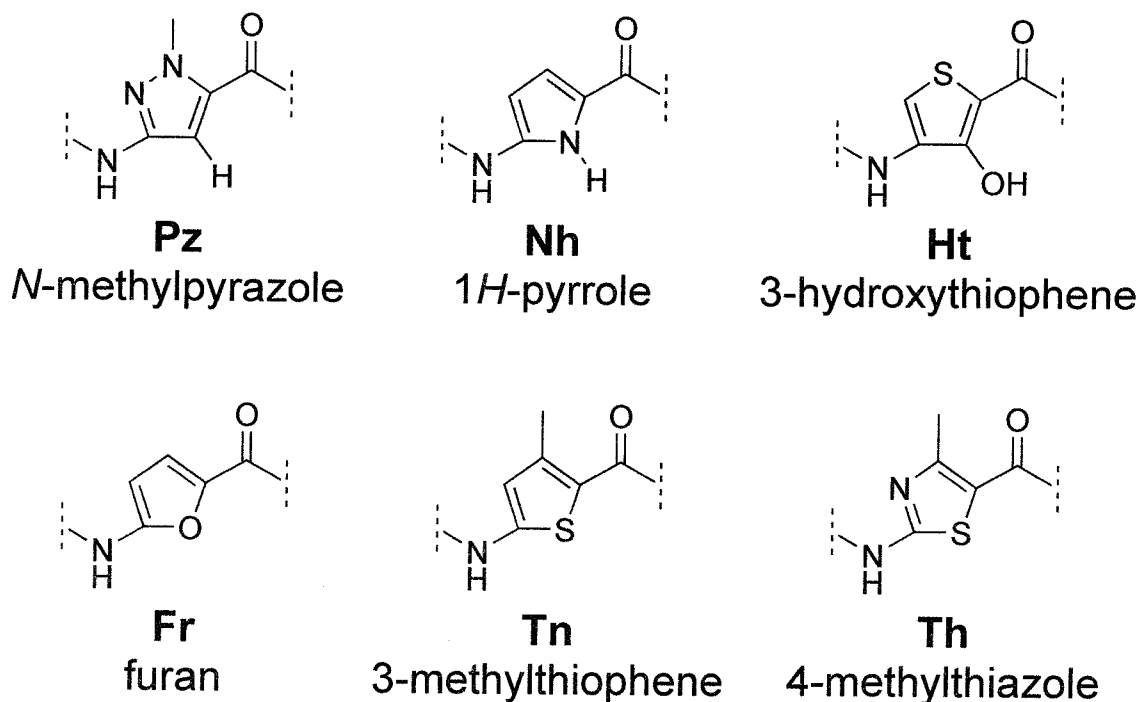


Figure 6. Novel five-membered heterocyclic amino acids that have been incorporated into hairpin polyamides. All residues are shown with the functionality that faces the DNA minor groove towards the bottom-right.

either *N*-methylpyrazole (Pz) or 3-methylthiophene (Tn) are exceptionally selective for A/T bases, displaying no binding to G/C bases in the measured range, incorporation of 4-methylthiazole (Th), furan (Fr), or 3-hydroxythiophene (Ht) abolishes binding entirely. Molecular modeling indicates that relatively small changes to a single heterocycle present in a contiguous four-ring system can induce large alterations in the curvature and electronic properties of the ligand (Figure 7) [35]. There appears to be a very narrow window of five-membered ring architectures that are effective for minor groove binding.

The benzimidazole ring system represents a different structural framework, which is amenable to functionalization on the six-membered ring and appears to impart a curvature that is complementary to DNA. Indeed, the classic minor groove-binding Hoechst dyes are composed of benzimidazole units, and a number of derivatives of these

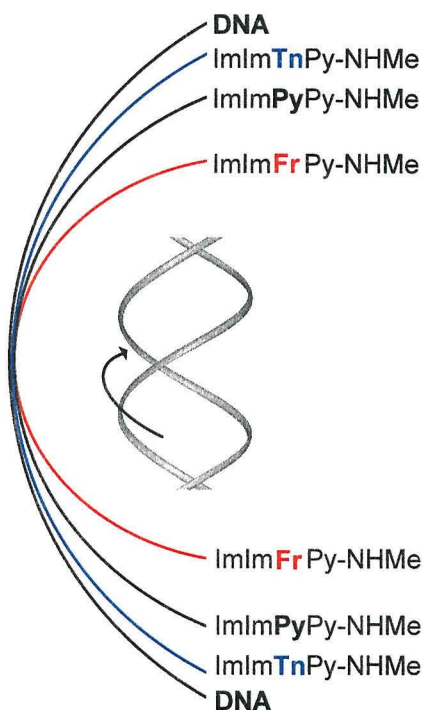


Figure 7. Schematic illustrating the curvatures of the DNA helix and of several four-ring polyamide subunits containing Tn, Py, and Fr heterocycles. The curves represent the overall curve of each ligand as determined by *ab initio* molecular modeling [35]. The Fr ring induces substantial curvature, and an eight-ring hairpin polyamide containing this residue does not bind to its target DNA sequence at concentrations up to 1 μ M.

molecules have been prepared [36,37]. We have recently incorporated benzimidazole derivatives into the backbone of hairpin polyamides in a manner that preserves critical hydrogen bonding contacts and overall molecular shape (Figure 8) [38]. The hydroxybenzimidazole (Hz) and imidazopyridine (Ip) rings are introduced into polyamides as dimeric subunits PyHz and PyIp, respectively, in which the Py ring is directly connected to the benzimidazole derivative without an intervening amide bond. DNase I footprinting indicates that the Hz/Py and Ip/Py pairs are functionally identical to the analogous five-membered ring pairs Hp/Py and Im/Py, and pairing rules for these rings are given in Table 1 [38]. Importantly, Hz-containing polyamides are chemically

robust, making the Hz/Py pair a strong candidate for replacing Hp/Py in biological studies.

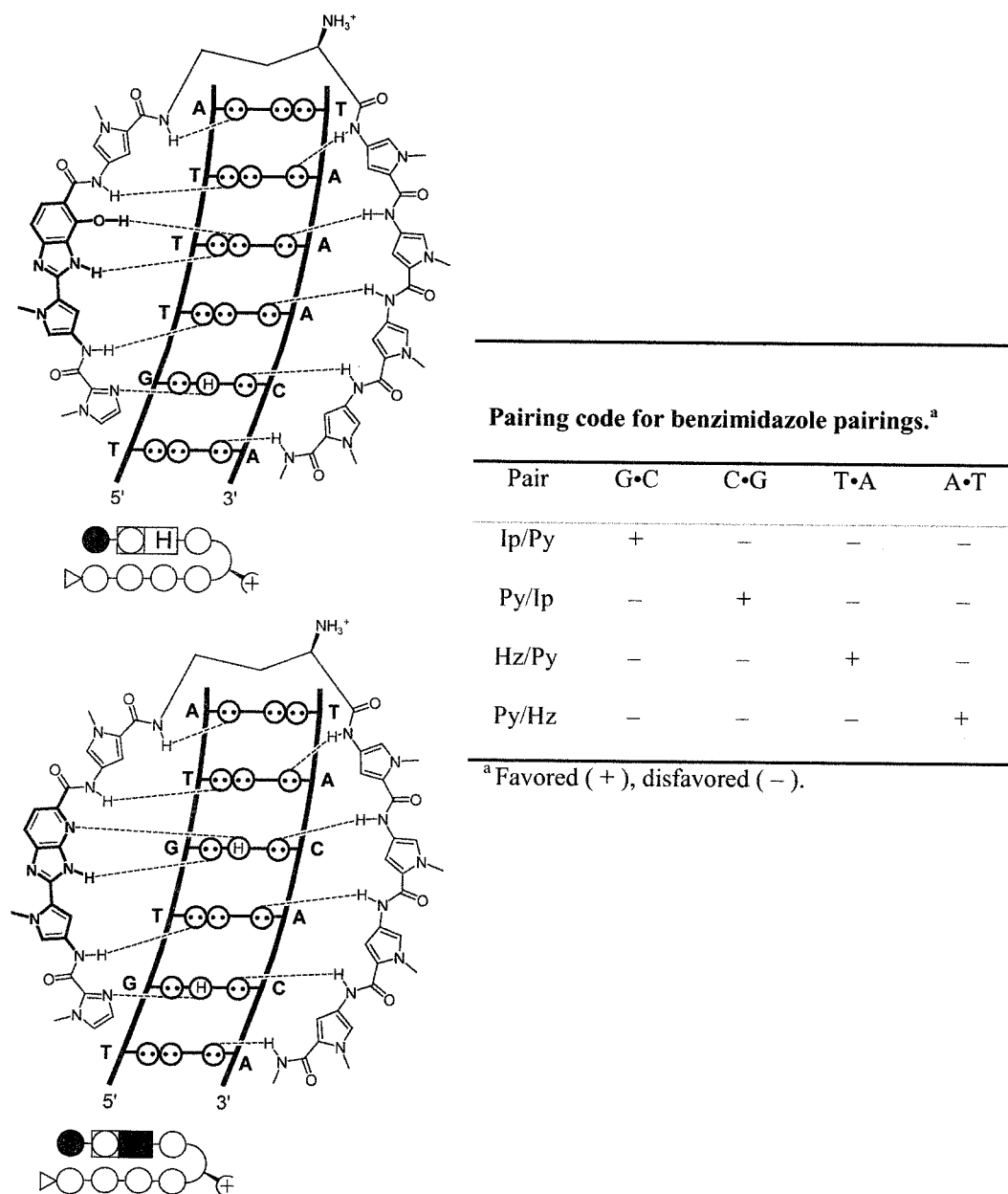


Figure 8. Recognition of the DNA minor groove with benzimidazole-derivatives. Circles with dots represent lone pairs of N(3) of purines and O(2) of pyrimidines, and circles containing an H represent the 2-amino group of guanine. Putative hydrogen bonds are shown as dashed lines. The dimeric units Py-hydroxybenzimidazole (Hz) and Py-imidazopyridine (Ip), at left and at right, respectively, are shown in bold. Schematics of the hairpin polyamides are given below the binding scheme: the rectangle containing an open circle and the letter H represents the dimer Py-Hz, and the rectangle containing an open circle and a shaded box represents the Py-Ip dimer. Other symbols are defined in Figure 4. At right is a table indicating the recognition code for 6-5 fused systems. + indicates favored; - indicates disfavored.

Synthetic Methods

The investigation of minor groove-binding polyamides was greatly accelerated by the implementation of solid-phase synthesis [39]. 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 [40]. Recently, Pessi and coworkers used a sulfonamide-based safety-catch resin to prepare derivatives of hairpin polyamides [41]. Upon activation of the linker, resin-bound polyamides were readily cleaved with stoichiometric quantities of nucleophile to provide thioesters or peptide conjugates.

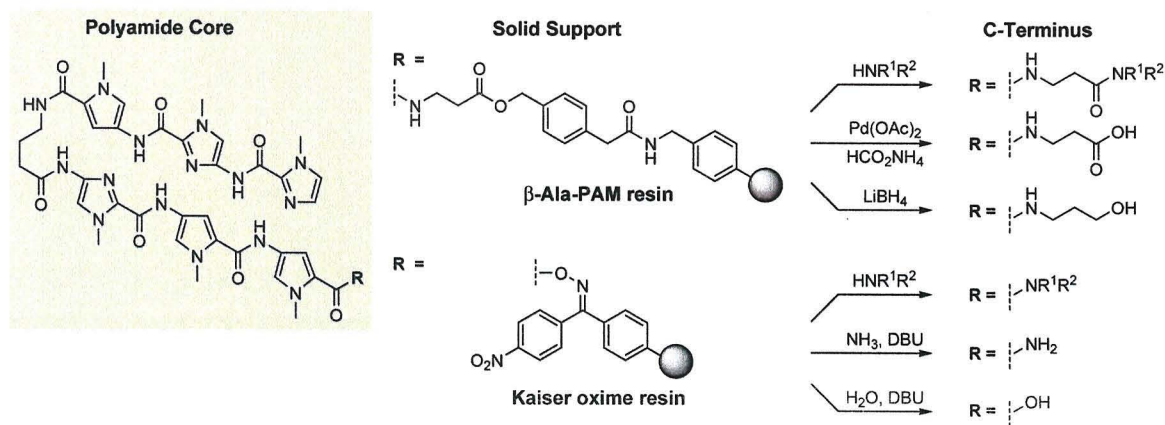


Figure 9. Schematic of potential C-termini available from specific solid supports. Polyamides synthesized on the Kaiser oxime resin can have shorter C-terminal groups than molecules prepared on β -Ala-PAM resin. The amine HNR^1R^2 may be a primary or secondary alkyl amine.

While allowing rapid preparation of a range of polyamides, all of these resins install a T,A selective β -Ala residue at the C-terminus, which places limits on the DNA sites that can be targeted [42]. The shortest tail available from these resins is a propanolamide, obtained by reductive cleavage [21]. Polyamides prepared on Boc-Gly-PAM resin can be reductively cleaved to obtain ethanolamide tails [29], but it was expected that further

truncation of the C-terminus would be necessary for tolerance of G,C at the tail position [43]. The Kaiser oxime resin was therefore adapted to polyamide synthesis, allowing the preparation of polyamides with incrementally shorter C-termini (Figure 9). These molecules display the desired tolerance for G,C bases while maintaining high affinities [43]. Moreover, removing the β -Ala residue may prove critical for specific applications in the future.

Inhibition of Gene Expression

Polyamides can bind with high affinity to a wide range of DNA sites, and can often competitively displace proteins from DNA (Figures 10 and 11). One approach to modifying gene expression involves inhibition of key transcription factor (TF)-DNA complexes in a designated promoter, thus interfering with recruitment of RNA polymerases (Figure 12). Significantly, because there are considerably fewer oncogenic TFs than potentially oncogenic signaling proteins, TF inhibition represents a uniquely promising approach to cancer treatment [44]. The transcription factor TFIIA was chosen as a first target because it regulates a relatively small number of genes and because contacts between the nine zinc-finger protein and the minor groove had been established. A polyamide bound in the recognition site of TFIIA suppressed transcription of 5S RNA genes by RNA polymerase III *in vitro* and in cultured *Xenopus* kidney cells [45]. Further studies used polyamides in combination with recombinant derivatives of TFIIA subunits to elucidate essential minor groove contacts for the binding of this TF [46].

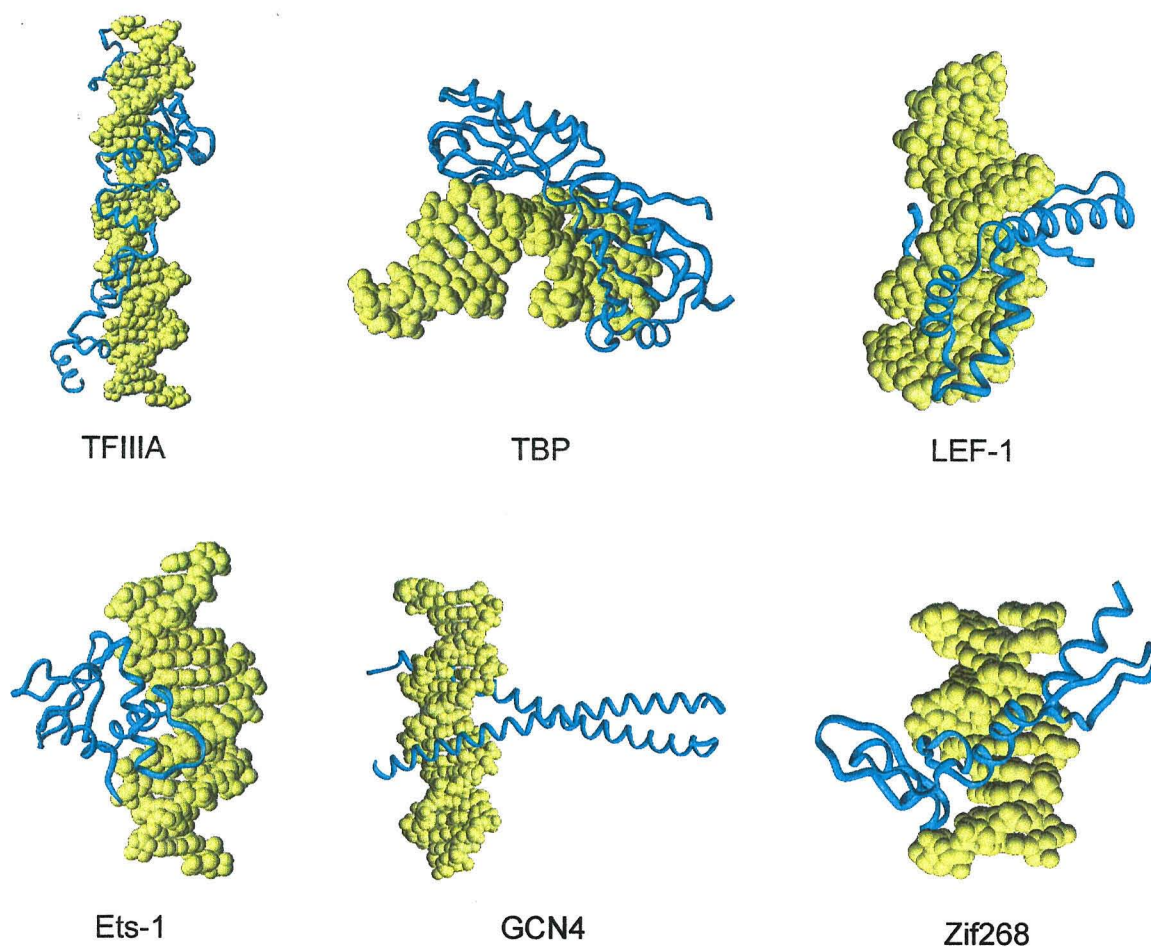


Figure 10. Structures of several different protein-DNA complexes that have been inhibited by polyamides. TFIIIA is a zinc finger, TBP is a minor groove-binding protein, LEF-1 is an HMG box, Ets-1 is a winged-helix-turn-helix, GCN4 is a bZIP protein, and Zif 268 is a zinc finger.

Polyamides were then used to target viral genes transcribed by RNA polymerase II. 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 [47]. In human blood lymphocytes, treatment with the two polyamides in combination inhibited viral replication by more than 99%, with no significant decrease in cell viability. Inhibition of viral replication is indirect evidence for specific transcription inhibition by

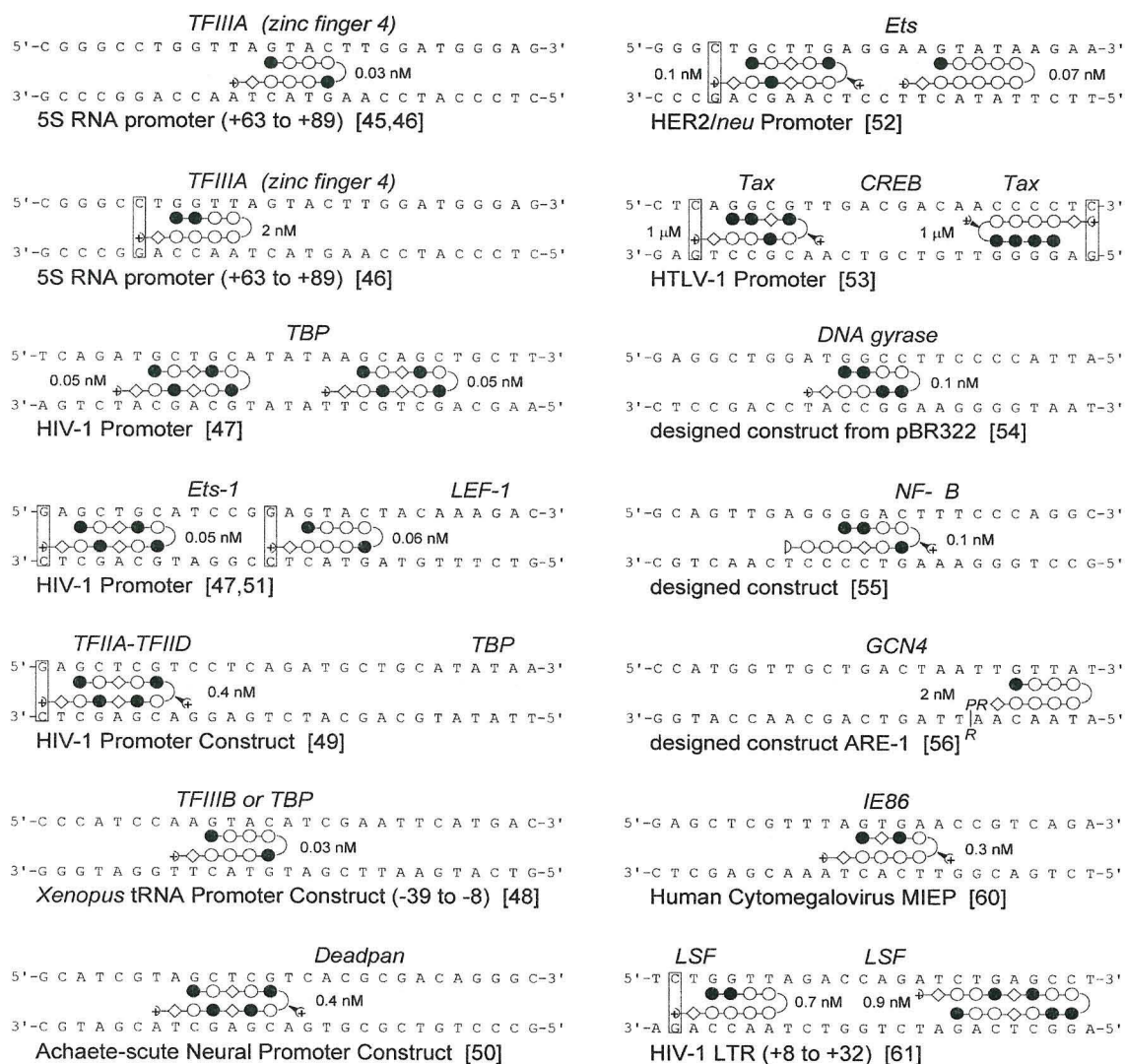


Figure 11. Examples of DNA-binding proteins that have been inhibited by polyamides. Approximate K_d values are given next to each polyamide. Where available, the values are from experiments conducted on the depicted DNA sequence. The name of the protein is italicized above a shaded box indicating its DNA binding site. Open boxes indicate mismatches between the Dp tail and G,C base pairs [42]. The promoter or construct is identified below the DNA sequence and is followed by the reference number in brackets. The HIV-1 Promoter Construct, *Xenopus* tRNA Promoter Construct, and Achaete-scute Neural Promoter Construct are sample DNA sequences from studies that employed promoter scanning. Precise binding sites for the TFIIA-TFIID complex and for TFIIIB are not identifiable from promoter scanning. RPR represents an Arg-Pro-Arg C-terminus, and the half circle represents a propanolamine $-NH(CH_2)_3OH$ group. All other symbols are defined in Figure 4.

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 [47].

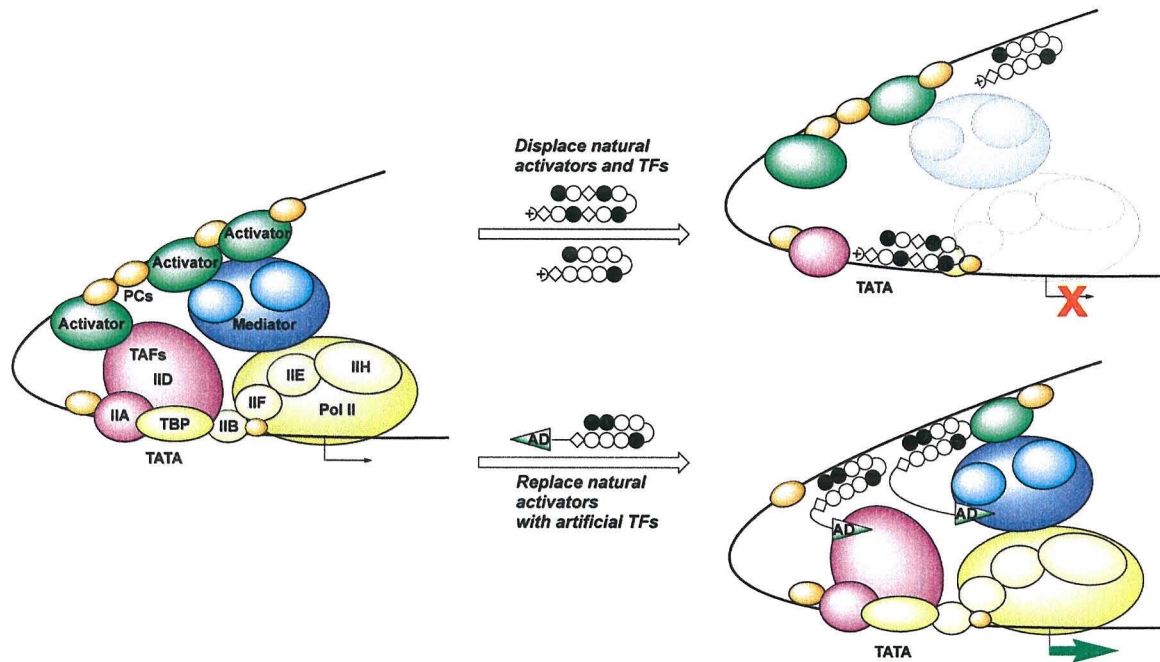


Figure 12. Two approaches to modifying gene expression with polyamides. At left is a model for the transcription machinery adapted from [78]. RNA polymerase II (Pol II) and general transcription factors are yellow, distally located activators are green, the mediator complex is blue, TBP-associated factors (TAFs) are purple, and architectural factors such as PC4 are orange. Polyamides are depicted as in Figure 4, and a small activation domain (AD) is indicated by the green triangle. The location of the TATA box is noted, and the transcription start site is indicated by a bent arrow. The large open arrows indicate analogous systems rather than direct effects of polyamide addition. Polyamides can downregulate gene expression by interfering with the binding of general TFs or activators in the promoter region of a gene (top). Polyamides conjugated to an AD function as artificial transcription factors and can upregulate gene expression by recruiting or stabilizing components of the transcription machinery (bottom). *In vitro* examples have used an engineered promoter that incorporates multiple polyamide binding sites [63-65].

This early biological result spurred a variety of biochemical studies of the interactions of various polyamides with the basal transcription machinery and TF-DNA complexes.

Two studies have used promoter scanning to identify sites where polyamide binding inhibits transcription [48,49]. The method uses a series of DNA constructs with designed polyamide binding sites at varying distances from the transcription start site. Essential minor groove contacts were identified for a subunit of TFIIB (possibly TBP) in a *Xenopus* tRNA promoter [48], as well as for TFIID-TFIIA and TBP in the HIV-1 core promoter [49]. The binding of the homodimeric basic-helix-loop-helix TF Deadpan was investigated using a variant of promoter scanning [50]. A series of duplex

oligonucleotides based on a *Drosophila* neural promoter were designed, incorporating polyamide binding sites on different sides of the Deadpan recognition sequence, and in different orientations. The TF-DNA complex was inhibited only by a polyamide binding upstream of the homodimer, establishing an asymmetric binding mode for this TF.

The binding of Ets-1 to the HIV-1 enhancer was examined in greater detail, and polyamides were shown to inhibit the formation of a ternary Ets-1–NF- κ B–DNA complex [51]. Ets-1 is a winged-helix-turn-helix TF, and its key phosphate contacts on either side of the major groove can be disrupted by a polyamide in the adjacent minor groove. The report provided the first evidence for cooperative DNA binding by Ets-1 and NF- κ B to the HIV-1 enhancer sequence [51]. A different Ets binding site in the HER2/*neu* promoter was targeted with hairpin polyamides that successfully blocked Ets-DNA complex formation and transcription of the HER2/*neu* oncogene in a cell-free system [52].

Several other protein-DNA interactions have been inhibited with polyamides. In the human T-cell leukemia virus type 1 (HTLV-1) promoter, polyamides targeted to G,C rich regions flanking the viral CRE sites inhibit binding of the Tax protein and Tax transactivation *in vitro* [53]. Bacterial gyrase recognizes a short 5'-GGCC-3' site, and a polyamide targeted to this sequence inhibited gyrase-catalyzed strand cleavage at nanomolar concentrations [54]. NF- κ B is a TF crucial for development, viral expression, inflammation, and anti-apoptotic responses. The most common form is a p50-p65 heterodimer, which binds DNA in the major groove, making several phosphate contacts

throughout the binding site. Polyamides targeted to the minor groove opposite p50, but not p65, inhibit DNA binding by NF- κ B [55].

Some purely major-groove binding TFs, such as the basic-region leucine zipper (bZIP) protein GCN4, can co-occupy the DNA helix in the presence of polyamides [56].

Modified polyamides with an attached Arg-Pro-Arg tripeptide can interfere with major-groove binding proteins by disrupting key phosphate contacts, distorting the DNA by charge neutralization, or sterically invading the major groove. An Arg-Pro-Arg-polyamide conjugate successfully inhibited the binding of GCN4 to DNA [56], and further optimization yielded a polyamide derivative with an alkyl diamine substituent that was 10-fold more potent [57]. Polyamide conjugates that distort DNA more severely, as by delivery of an intercalator, are promising candidates for site selective inhibition of any DNA-binding protein [58].

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 [59]. Polyamides induced specific gain- and loss-of-function phenotypes when fed to developing *Drosophila* embryos.

Gene Activation

Polyamides can upregulate transcription in two main ways: derepression or 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 [60]. 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 [61]. As with other systems, only polyamides matched to the correct protein binding site induced significant effects. Several existing drug 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 [61]. This therapeutic 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 12) [62]. Such artificial transcription factors have been synthesized and evaluated in cell-free transcription assays [63-65]. 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 [63]. 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 analogue [64]. Since the linker length had been shown to influence activation efficiency, a set of molecules with rigid oligoproline linkers between the polyamide and the activation domain was synthesized [65]. The oligoproline linkers act as rigid “molecule rulers,” and optimal activation was observed with a Pro₁₂ linker, about 36Å in length.

Nucleosomes

In eukaryotic cells, DNA is tightly packaged by compaction into chromatin, and changes in chromatin structure can alter accessibility of specific sequences and actively 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 [66]. 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 amino-terminal tails of histone H3 and H4 [66].

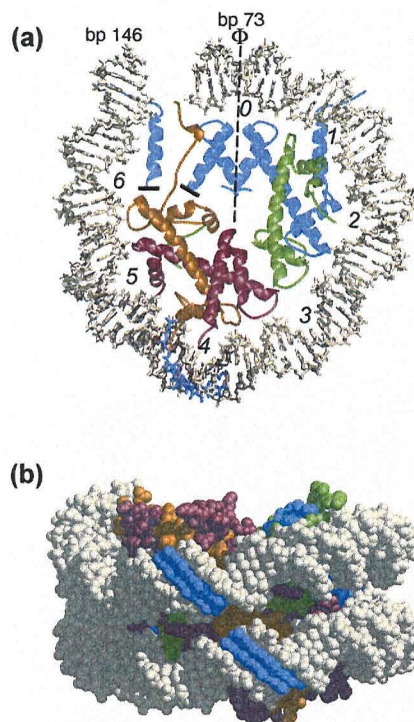


Figure 13. Overview of a nucleosome core particle (NCP)-polyamide cocrystal structure (PDB code 1M18) [67]. **(a)** Partial structure, viewed down the superhelical axis. Base-pairs 68–146 (shown in white) and associated proteins (H3, blue; H4, green; H2A, yellow; H2B, red) are shown. The equivalent region of DNA from the NCP with no additional ligands is superimposed, based on an alignment of the histone octamer atoms, and is shown in brown. Superhelix locations 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. **(b)** Side-view of the structures in an orientation that best displays the bound polyamides is shown in a CPK representation. DNA, histones, and polyamide are colored as in (a).

Subsequently, the structures of three of these polyamide-nucleosome core particle complexes were determined by X-ray crystallography (Figure 13) [67]. 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 ($K_d \sim 1$ nM) only the relatively non-specific polyamide ImPyPyPy- γ -PyPyPyPy- β -Dp inhibits temperature-induced nucleosome translocation [67]. 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.

Although polyamides can block transcription by targeting promoter elements, they do not affect transcription when bound in the RNA coding regions of DNA [47]. Presumably, the strand melting required for RNA polymerase progression disrupts the minor groove and displaces polyamides. To investigate potential effects on transcription through a nucleosome, hairpin polyamides were targeted to sites on the nucleosome positioning sequence of the sea urchin 5S gene [68]. The two molecules that blocked heat-induced nucleosomal translocation also blocked transcription by T7 RNA polymerase. Each of these polyamides binds with high affinity ($K_d \sim 1$ nM) to a single site in the nucleosome construct, potentially implying that placement is critical. Nonetheless, the positions of these sites are distinct from those occupied by the compound ImPyPyPy- γ -PyPyPyPy- β -Dp in the crystal structure [67]. Although the precise mechanisms involved in nucleosome repositioning remain in question, it appears that in some cases DNA can “roll” over the histones, and certain polyamides can act as chocks to prevent the DNA from moving [67,68].

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 [69]. In cell types that had shown robust responses to polyamides, such as primary human T-cells [47], polyamide-dye conjugates were observed to enter the nuclei of live cells [69]. However, in the majority of cell lines, polyamides were excluded from the nucleus. Co-staining with organelle-specific dyes indicates that polyamides are often trapped in lysosomes and other cytoplasmic vesicles [70], such that cells treated with polyamides can give a false nuclear signal upon fixing, even if they are washed extensively [69]. We have recently found specific combinations of linkers and fluorophores that enable the nuclear localization of polyamides in a broad range of mammalian cell lines, which will be reported in due course [71].

Double Strand DNA Detection

We are hopeful that solutions to several key problems connected with molecular recognition and biological trafficking are in sight. Although the scope and limitations of these advances must be examined, we are now keenly interested in using polyamides as multi-purpose tools for DNA detection. Fluorescent derivatives may find broad application. Laemmli and coworkers developed the telomere-specific dye conjugates mentioned previously [33], and Trask and coworkers have painted human chromosomes with centromere-specific hairpin polyamide-dye conjugates [72]. Polyamide-based

molecular beacons [73], which display fluorescence enhancement upon binding to a target sequence, are distinguished from DNA-based probes because the double strand DNA target does not have to be denatured. Analysis of native duplex DNA allows a broader range of detection possibilities, such as identification of mismatched (non-Watson-Crick) DNA base pairs [73], and potentially, tracking of specific DNA sequences in live cells. The combination of polyamides, multiple fluorophores capable of FRET, and microfluidic technologies could provide the basis for unique DNA detection devices.

Outlook and Future Directions

Our long-term aspiration has always been the control of gene expression in living systems. This goal can now be pursued with renewed vigor. Polyamides and the pairing rules allow for the “digital read-out” of predetermined DNA sequences, and now that it appears that polyamides can be modified with improved cell uptake and nuclear localization properties in mammalian cells, the question of specificity on a genomic scale will be critical. What will be the minimum requirements for selectively altering transcription of a single gene? It may be that multiple “small polyamides” (each targeting 6 bp) acting in concert will be more effective than one single large molecule (targeting 10-12 bp). Put rather differently, what are the genomic effects of treating cells with polyamides? Such genome-wide analyses have already begun [74], and in our lab, gene chip experiments in both yeast and human systems indicate that an eight-ring polyamide—recognizing only six bp of DNA—affects the transcription of only a handful of genes. Importantly, we observe that polyamides targeted to different DNA sequences generate distinct expression profiles.

Although gene suppression techniques have made extraordinary advances, most notably with the advent of RNAi technology [75,76], DNA-binding polyamides are the only cell-permeable small molecules capable of recognizing specific, predetermined sequences of DNA. Furthermore, polyamide-peptide conjugates may be used as artificial transcription factors to up-regulate transcription [63-65], whereas it is not clear how RNAi could be used as a general tool for transcriptional up-regulation. A major challenge, which we have only recently begun to address, will be cell and nuclear uptake trafficking of such functionalized conjugates, including polyamides bearing moieties for activation [64] or protein recruitment [77].

References

1. Arcamone F, Nicoletti V, Penco S, Orezzi P, Pirelli A. *Nature* 1964, **203**:1064.
2. Pelton JG, Wemmer DE. *Proc Natl Acad Sci USA* 1989, **86**:5723-5727.
3. Gottesfeld JM, Turner JM, Dervan PB. *Gene Expr* 2000, **9**:77-91.
4. Dervan PB. *Bioorg Med Chem* 2001, **9**:2215-2235.
5. White S, Szewczyk JW, Turner JM, Baird EE, Dervan PB. *Nature* 1998, **391**:468-471.
6. Wade WS, Mrksich M, Dervan PB. *J Am Chem Soc* 1992, **114**:8783-8794.
7. Mrksich M, Wade WS, Dwyer TJ, Geierstanger BH, Wemmer DE, Dervan PB. *Proc Natl Acad Sci USA* 1992, **89**:7586-7590.
8. Kielkopf CL, Baird EE, Dervan PB, Rees DC. *Nat Struct Biol* 1998, **5**:104-109.
9. Pilch DS, Poklar N, Baird EE, Dervan PB, Breslauer KJ. *Biochemistry* 1999, **38**:2143-2151.
10. Kielkopf CL, White S, Szewczyk JW, Turner JM, Baird EE, Dervan PB, Rees DC. *Science* 1998, **282**:111-115.
11. Kielkopf CL, Bremer RE, White S, Szewczyk JW, Turner JM, Baird EE, Dervan PB, Rees DC. *J Mol Biol* 2000, **295**:557-567.
12. Wellenzohn B, Loferer MJ, Trieb M, Rauch C, Winger RH, Mayer E, Liedl KR. *J Am Chem Soc* 2003, **125**:1088-1095.
13. Mrksich M, Parks ME, Dervan PB. *J Am Chem Soc* 1994, **116**:7983-7988.
14. Trauger JW, Baird EE, Dervan PB. *Nature* 1996, **382**:559-561.
15. deClairac RPL, Geierstanger BH, Mrksich M, Dervan PB, Wemmer DE. *J Am Chem Soc* 1997, **119**:7909-7916.
16. White S, Baird EE, Dervan PB. *J Am Chem Soc* 1997, **119**:8756-8765.
17. Urbach AR. Pasadena: California Institute of Technology: 2002.
18. Herman DM, Baird EE, Dervan PB. *J Am Chem Soc* 1998, **120**:1382-1391.

19. Woods CR, Ishii T, Wu B, Bair KW, Boger DL. *J Am Chem Soc* 2002, **124**:2148-2152.
20. Herman DM, Turner JM, Baird EE, Dervan PB. *J Am Chem Soc* 1999, **121**:1121-1129.
21. Melander C, Herman DM, Dervan PB. *Chem-Eur J* 2000, **6**:4487-4497.
22. Baliga R, Baird EE, Herman DM, Melander C, Dervan PB, Crothers DM. *Biochemistry* 2001, **40**:3-8.
23. Greenberg WA, Baird EE, Dervan PB. *Chem-Eur J* 1998, **4**:796-805.
24. Olenyuk B, Jitianu C, Dervan PB. *Journal of the American Chemical Society* in press.
25. Swalley SE, Baird EE, Dervan PB. *Chem-Eur J* 1997, **3**:1600-1607.
26. Turner JM, Swalley SE, Baird EE, Dervan PB. *J Am Chem Soc* 1998, **120**:6219-6226.
27. Wang CCC, Ellervik U, Dervan PB. *Bioorg Med Chem* 2001, **9**:653-657.
28. Kelly JJ, Baird EE, Dervan PB. *Proc Natl Acad Sci USA* 1996, **93**:6981-6985.
29. Trauger JW, Baird EE, Dervan PB. *Angew Chem-Int Edit* 1998, **37**:1421-1423.
30. Weyermann P, Dervan PB. *J Am Chem Soc* 2002, **124**:6872-6878.
31. Herman DM, Baird EE, Dervan PB. *Chem-Eur J* 1999, **5**:975-983.
32. Kers I, Dervan PB. *Bioorg Med Chem* 2002, **10**:3339-3349.
33. Maeshima K, Janssen S, Laemmli UK. *Embo J* 2001, **20**:3218-3228.
34. Nguyen DH, Szewczyk JW, Baird EE, Dervan PB. *Bioorg Med Chem* 2001, **9**:7-17.
35. Marques MA, Doss RM, Urbach AR, Dervan PB. *Helv Chim Acta* 2002, **85**:4485-4517.
36. Minehan TG, Gottwald K, Dervan PB. *Helv Chim Acta* 2000, **83**:2197-2213.
37. Lown JW: In *Advances in DNA sequence specific agents*. Edited by Jones GB, Chapman BJ: Elsevier Science; 2003. vol 4.]
38. Renneberg D, Dervan PB. *J Am Chem Soc* 2003

39. Baird EE, Dervan PB. *J Am Chem Soc* 1996, **118**:6141-6146.
40. Wurtz NR, Turner JM, Baird EE, Dervan PB. *Org Lett* 2001, **3**:1201-1203.
41. Fattori D, Kinzel O, Ingallinella P, Bianchi E, Pessi A. *Bioorg Med Chem Lett* 2002, **12**:1143-1147.
42. Swalley SE, Baird EE, Dervan PB. *J Am Chem Soc* 1999, **121**:1113-1120.
43. Belitsky JM, Nguyen DH, Wurtz NR, Dervan PB. *Bioorg Med Chem* 2002, **10**:2767-2774.
44. Darnell JE. *Nat Rev Cancer* 2002, **2**:740-749.
45. Gottesfeld JM, Neely L, Trauger JW, Baird EE, Dervan PB. *Nature* 1997, **387**:202-205.
46. Neely L, Trauger JW, Baird EE, Dervan PB, Gottesfeld JM. *J Mol Biol* 1997, **274**:439-445.
47. Dickinson LA, Gulizia RJ, Trauger JW, Baird EE, Mosier DE, Gottesfeld JM, Dervan PB. *Proc Natl Acad Sci USA* 1998, **95**:12890-12895.
48. McBryant SJ, Baird EE, Trauger JW, Dervan PB, Gottesfeld JM. *J Mol Biol* 1999, **286**:973-981.
49. Ehley JA, Melander C, Herman D, Baird EE, Ferguson HA, Goodrich JA, Dervan PB, Gottesfeld JM. *Mol Cell Biol* 2002, **22**:1723-1733.
50. Winston RL, Ehley JA, Baird EE, Dervan PB, Gottesfeld JM. *Biochemistry* 2000, **39**:9092-9098.
51. Dickinson LA, Trauger JW, Baird EE, Dervan PB, Graves BJ, Gottesfeld JM. *J Biol Chem* 1999, **274**:12765-12773.
52. Chiang SY, Burli RW, Benz CC, Gawron L, Scott GK, Dervan PB, Beerman TA. *J Biol Chem* 2000, **275**:24246-24254.
53. Lenzmeier BA, Baird EE, Dervan PB, Nyborg JK. *J Mol Biol* 1999, **291**:731-744.
54. Simon H, Kittler L, Baird E, Dervan P, Zimmer C. *FEBS Lett* 2000, **471**:173-176.
55. Wurtz NR, Pomerantz JL, Baltimore D, Dervan PB. *Biochemistry* 2002, **41**:7604-7609.
56. Bremer RE, Baird EE, Dervan PB. *Chem Biol* 1998, **5**:119-133.

57. Bremer RE, Wurtz NR, Szewczyk JW, Dervan PB. *Bioorg Med Chem* 2001, **9**:2093-2103.
58. Fechter EJ, Dervan PB. Submitted.
59. Janssen S, Cuvier O, Muller M, Laemmli UK. *Mol Cell* 2000, **6**:1013-1024.
60. Dickinson LA, Trauger JW, Baird EE, Ghazal P, Dervan PB, Gottesfeld JM. *Biochemistry* 1999, **38**:10801-10807.
61. Coull JJ, He GC, Melander C, Rucker VC, Dervan PB, Margolis DM. *J Virol* 2002, **76**:12349-12354.
62. Ansari AZ, Mapp AK. *Curr Opin Chem Biol* 2002, **6**:765-772.
63. Mapp AK, Ansari AZ, Ptashne M, Dervan PB. *Proc Natl Acad Sci USA* 2000, **97**:3930-3935.
64. Ansari AZ, Mapp AK, Nguyen DH, Dervan PB, Ptashne M. *Chem Biol* 2001, **8**:583-592.
65. Arora PS, Ansari AZ, Best TP, Ptashne M, Dervan PB. *J Am Chem Soc* 2002, **124**:13067-13071.
66. Gottesfeld JM, Melander C, Suto RK, Raviol H, Luger K, Dervan PB. *J Mol Biol* 2001, **309**:615-629.
67. Suto RK, Edayathumangalam RS, White CL, Melander C, Gottesfeld JM, Dervan PB, Luger K. *J Mol Biol* 2003, **326**:371-380.
68. Gottesfeld JM, Belitsky JM, Melander C, Dervan PB, Luger K. *J Mol Biol* 2002, **321**:249-263.
69. Belitsky JM, Leslie SJ, Arora PS, Beerman TA, Dervan PB. *Bioorg Med Chem* 2002, **10**:3313-3318.
70. Crowley KS, Phillion DP, Woodard SS, Schweitzer BA, Singh M, Shabany H, Burnette B, Hippenmeyer P, Heitmeier M, Bashkin JK. *Bioorg Med Chem Lett* in press.
71. Edelson BS, Best TP, Nickols, NG, Dervan PB. *Proc. Natl. Acad. Sci. U.S.A.* 2003, **100**, 12063-12075.
72. Gygi MP, Ferguson MD, Mefford HC, Lund KP, O'Day C, Zhou P, Friedman C, van den Engh G, Stolowitz ML, Trask BJ. *Nucleic Acids Res* 2002, **30**:2790-2799.

73. Rucker VC, Foister S, Melander C, Dervan PB. *J Am Chem Soc* 2003, **125**:1195-1202.
74. Supekova L, Pezacki JP, Su AI, Loweth CJ, Riedl R, Geierstanger B, Schultz PG, Wemmer DE. *Chem Biol* 2002, **9**:821-827.
75. Agami R. *Curr Opin Chem Biol* 2002, **6**:829-834.
76. Kamath RS, Fraser AG, Dong Y, Poulin G, Durbin R, Gotta M, Kanapin A, Le Bot N, Moreno S, Sohrmann M, et al. *Nature* 2003, **421**:231-237.
77. Wang CCC, Dervan PB. *J Am Chem Soc* 2001, **123**:8657-8661.
78. Malik S, Roeder RG. *Trends Biochem Sci* 2000, **25**:277-283.

Chapter 2

Nuclear Localization of Pyrrole-Imidazole Polyamide-Fluorescein Conjugates in Cell Culture

The text of this chapter was taken in part from a manuscript coauthored with Professor Timothy P. Best, Nicholas G. Nickols, and Peter B. Dervan (Caltech)

(Best, T. P.; Edelson, B. S.; Nickols, N. G.; Dervan, P. B.; “Nuclear Localization of Pyrrole-Imidazole Polyamide-Fluorescein Conjugates in Cell Culture” *Proc. Natl. Acad. Sci. U.S.A.*, **2003**, *100*, 12063)

Abstract

A series of hairpin pyrrole-imidazole polyamide-fluorescein conjugates were synthesized and assayed for cellular localization. Thirteen cell lines, representing eleven human cancers, one human transformed kidney cell line, and one murine leukemia cell line, were treated with 5 μ M polyamide-fluorescein conjugates for 10-14 h, then imaged by confocal laser scanning microscopy. A conjugate containing a β -alanine residue at the C-terminus of the polyamide moiety showed no nuclear localization, while an analogous compound lacking the β -alanine residue was strongly localized in the nuclei of all cell lines tested. The localization profiles of several other conjugates suggest that pyrrole-imidazole sequence and content, dye choice and position, linker composition, and molecular weight are determinants of nuclear localization. The attachment of fluorescein to the C-terminus of a hairpin polyamide results in an approximate 10-fold reduction in DNA-binding affinity, with no loss of binding specificity with reference to mismatch binding sites.

Introduction

Small molecules that preferentially bind to predetermined DNA sequences inside living cells would be useful tools in molecular biology, and perhaps human medicine. The effectiveness of these small molecules requires not only that they bind to chromosomal DNA site-specifically, but also that they be permeable to the outer membrane and gain access to the nucleus of living cells. Polyamides containing the aromatic amino acids N-methylpyrrole (Py), N-methylimidazole (Im) and N-methyl-3-hydroxypyrrole (Hp) bind DNA with affinities and specificities comparable to naturally occurring DNA-binding proteins (1,2). A set of pairing rules describes the interactions between pairs of these heterocyclic rings and Watson-Crick base pairs within the minor groove: Im/Py is specific for G•C, Hp/Py is specific for T•A and Py/Py binds both A•T and T•A. Exploitation of polyamides to target sequences of biological interest has yielded results in a number of cell-free systems (3-8).

Extension of these *in vitro* biological results to cellular systems has proven to be cell-type dependent. Polyamides exhibited biological effects in primary human lymphocytes (9), human cultured cell lines (10), and when fed to *Drosophila* embryos (11,12). However, attempts to inhibit the transcription of endogenous genes in cell lines other than insect or T-lymphocytes have met with little success. For example, polyamides that down-regulate transcription of the HER2/*neu* gene in cell-free experiments display no activity in HER2-overexpressing SK-BR-3 cells (13).

To determine if these results were due to poor cellular uptake or nuclear localization, a series of polyamides incorporating the fluorophore Bodipy FL was synthesized. Intracellular distribution of these molecules in several cell lines was then

determined by confocal laser scanning microscopy (14). Cells which demonstrated robust responses to polyamides, such as T-lymphocyte derivatives, showed staining throughout the cells, including the nucleus. In other cell lines studied, however, treatment with polyamide-Bodipy conjugates produced a punctate staining pattern in the cytoplasm, with no observable signal in the nuclei. Bashkin and coworkers recently reported that an eight-ring polyamide-Bodipy conjugate colocalized with LysoTracker Red DND 99 (a lysosome and *trans*-Golgi stain, **23**) in several cell lines, indicating that the punctate staining pattern was presumably due to trapping of the polyamide in acidic vesicles (15). In contrast, an eight-ring polyamide-fluorescein conjugate, **1**, was shown to accumulate in the nuclei of HCT-116 human colon cancer cells.

We have found that a similar eight-ring polyamide-fluorescein conjugate, **2**, with a single Py to Im change, is excluded from the nuclei of thirteen different mammalian cell lines, whereas removal of the β -Ala residue in the linker, affording **3**, enables nuclear localization in all of these cell lines, with no obvious toxicity. This raises the issue: what are the molecular determinants—fluorophore, position of attachment, linker composition, polyamide sequence and size—of uptake and nuclear localization in cultured cells? Understanding nuclear accessibility in a wide variety of living cells is a minimum first step toward chemical regulation of gene expression with this class of molecules. A further issue will be whether polyamides modified for optimal cellular and nuclear uptake retain favorable DNA-binding affinity and sequence specificity. We have synthesized twenty-two polyamide-fluorophore conjugates with incremental changes in structure and examined their intracellular distribution in thirteen cell lines.

Materials and Methods

For more details, see *Supporting Materials and Methods*, which is published as supporting information on the PNAS web site, www.pnas.org.

Chemicals

Polyamides **1** and **2** were prepared by solid phase methods on Boc- β -alanine-PAM-resin (Peptides International; Louisville, KY) (16). All other polyamides were synthesized by solid phase methods on the Kaiser oxime resin (Nova Biochem; Laufelfingen, Switzerland) (17). After cleavage with the appropriate amine and reverse-phase HPLC purification, polyamides were allowed to react at room temperature for ~3 h at ~0.01 M in *N,N*-dimethylformamide with fluorescein isothiocyanate (FITC; compounds **1-3** and **5-22**) or the N-hydroxysuccinimidyl ester of BODIPY-FL (**4**), as well as 20 eq of Hünig's base, to yield polyamide-dye conjugates. The purity and identity of the dye conjugates were verified by analytical HPLC, UV-vis spectroscopy and MALDI-TOF mass spectrometry. All fluorescent dye reagents were from Molecular Probes. Chemicals not otherwise specified were from Aldrich.

Confocal Microscopy

Adherent cell lines were trypsinized for 5-10 min at 37°C, centrifuged for 5 min at 5°C at 2000 rpm, and resuspended in fresh medium to a concentration of 1.25×10^6 cells/mL. Suspended cell lines were centrifuged and resuspended in fresh medium to the same concentration. Incubations were performed by adding 150 μ L cells into culture dishes equipped with glass bottoms for direct imaging (MatTek Corporation). Adherent cells

were grown in the glass-bottom culture dishes for 24 h. The medium was then removed and replaced with 142.5 μ L of fresh medium. Then 7.5 μ L of the 100 μ M polyamide solution was added and the cells were incubated in a 5% CO₂ atmosphere at 37°C for 10-14 h. Suspended cell line samples were prepared in a similar fashion, omitting trypsinization. These samples were then incubated as above for 10-14 h. Imaging was performed with a 40X oil-immersion objective lens on a Zeiss LSM 510 META NLO laser scanning microscope with a Coherent Chameleon 2-photon laser, or on a Zeiss LSM 5 Pascal inverted laser scanning microscope.

Energy Dependence Experiments

Inhibitory medium was prepared by supplementing glucose- and sodium pyruvate-free DMEM (Gibco #1196025) with 2-deoxyglucose (6 mM) and sodium azide (10 mM) (18). Cells were grown, trypsinized, resuspended, plated and incubated as above. After 24 h of growth at 37°C in 5% CO₂, the medium was removed and replaced with either 142.5 μ L fresh normal DMEM medium or 142.5 μ L inhibitory DMEM medium. The cells were incubated for 30 min, then treated with 7.5 μ L of 100 μ M compound **3**. The cells were then incubated for 1 h, followed by confocal imaging as above. Samples in inhibitory medium were then treated by removal of the medium, washing and removal of 200 μ L of normal medium, replacement with 142.5 μ L of normal medium and addition of 7.5 μ L of 100 μ M compound **3**. These samples were incubated for 1 h and then imaged once more.

DNase I Footprinting Titration Experiments

A 3'-[³²P]-labeled restriction fragment from the plasmid pDEH9 was generated in accordance with standard protocols and isolated by nondenaturing gel electrophoresis (19,20).

Results

Structures for all of the compounds synthesized are listed in Figure 1. Sample images for compounds **2** and **3** in two cell lines are shown in Figure 2. To show that compound **3** localizes to the nuclei of live cells, samples were treated with **3**, the nuclear stain Hoechst 33342, and the dead-cell stain Sytox Orange (Figure 3). The uptake characteristics of compounds **1-22** were examined in thirteen cell lines by confocal microscopy, and each sample was rated qualitatively for the extent of nuclear localization (Figure 4).

The molecules showing the highest degree of nuclear uptake in most cell lines stained nuclei very brightly with reference to the background fluorescence caused by fluorescent agent in the medium. Agents **1, 3, 5, 6, 11-14**, and **22** showed such high levels of uptake in many cell lines. Many compounds exhibited a very scattered uptake profile in the series of cells studied, and compounds **2, 4, 15**, and **21**, showed no significant nuclear staining in any cell line. Of the entries in Figure 4 indicating poor nuclear uptake properties, most reflect lysosomal staining, as indicated by costaining with LysoTracker Red DND-99, as well as fluorescence in the intercellular medium. Only compound **2** was found exclusively in the medium, showing no uptake in cells.

The DNA-binding properties of **3** were assessed by DNase I footprinting titrations on the plasmid pDEH9, which bears the 5'-TGGTCA-3' match site and discrete single

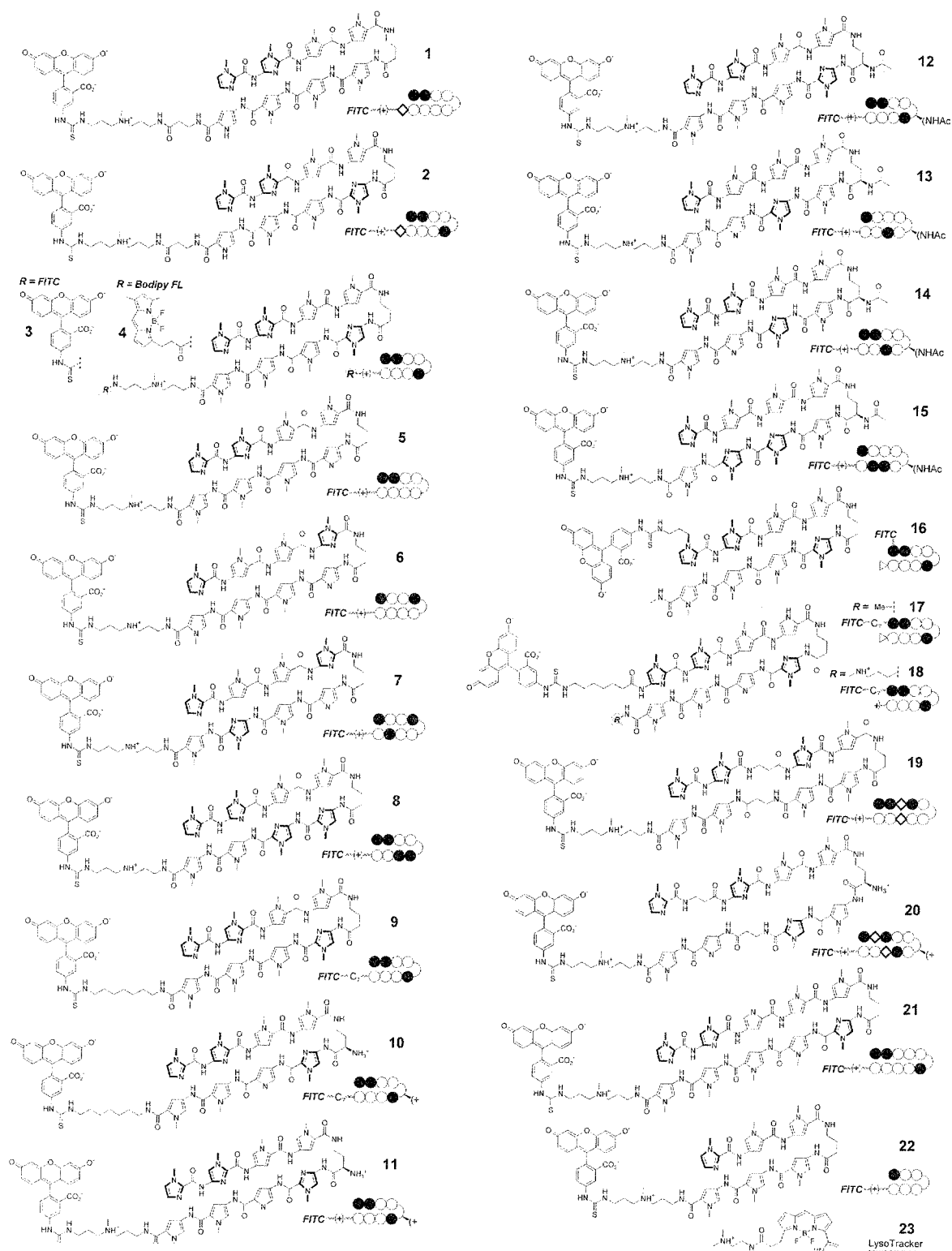


Figure 1. Structures of compounds 1-22, synthesized for this work, and 23, purchased from Molecular Probes.

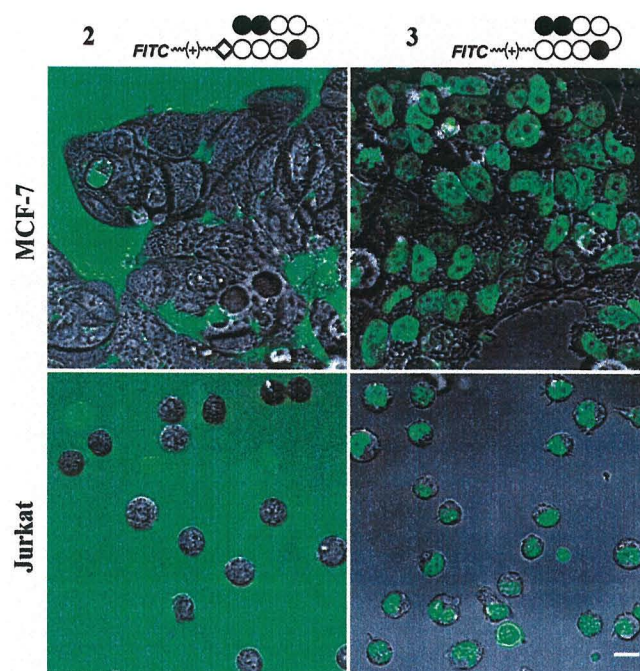


Figure 2. Cellular localization of polyamide-fluorescein conjugates. (*Top*) Adherent MCF-7 cells were treated with compound **2** (*Upper left*) or **3** (*Upper right*) for 10-14 h at 5 μ M. Compound **2** was excluded from the cells entirely, while compound **3** localized to the nucleus. (*Bottom*) Suspended Jurkat cells were similarly treated, and show similar results. (Scale bar = 10 μ m.)

base-pair mismatch sites. Conjugate **3** bound to the match site with a K_a of $1.6 (\pm 0.3) \times 10^9 \text{ M}^{-1}$, and showed specificity over mismatch sites by >100-fold (Figure 5).

To investigate the energy dependence of the cellular uptake mechanism of compound **3**, HeLa cells growing under normal conditions were incubated for 30 min in either normal DMEM medium or inhibitory DMEM medium, then treated with **3** for 1 h prior to confocal imaging. The cells growing in normal medium showed clear nuclear staining, whereas the cells growing in inhibitory medium displayed very little to no discernable staining. Subsequent washing and replacement of the inhibitory medium with normal medium (supplemented with 5 μ M **3**), resulted in nuclear staining after 1 h, comparable to that seen in the sample grown continuously in the normal medium (Figure 6).

It had been shown previously that polyamide-Bodipy conjugates stain the nuclei of T-lymphocytes, but no other cell type tested, and most commonly produced a punctate cytoplasmic staining pattern (14). Our studies indicated that a polyamide-fluorescein conjugate, **2**, uniformly proved refractory to nuclear uptake in several human cancer cell lines. Elimination of the β -alanine residue at the carboxy-terminal end of the polyamide

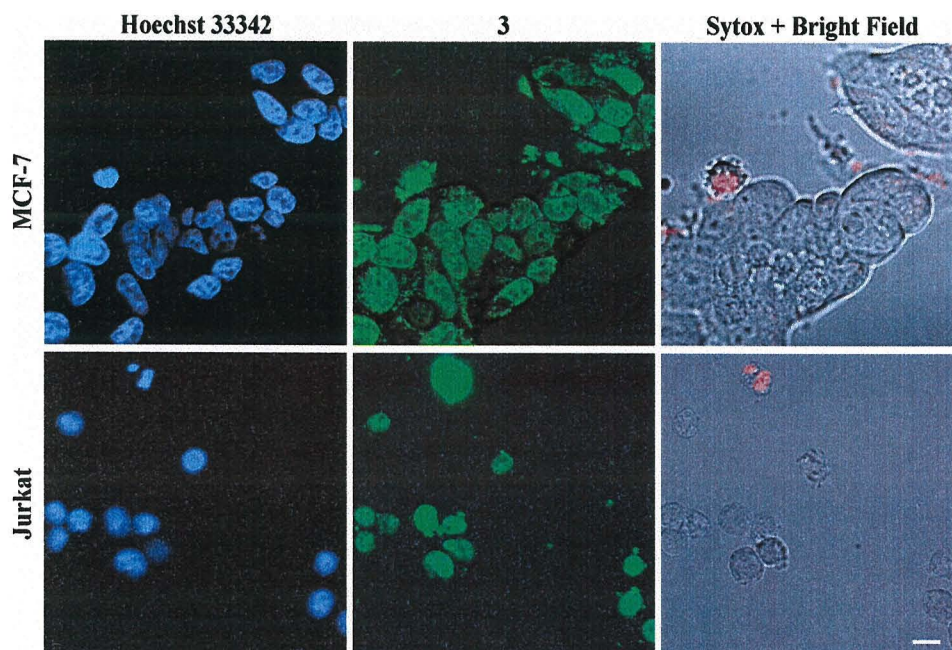


Figure 3. Colocalization of polyamide **3** and Hoechst in live cells, imaged using sequential single- and two-photon excitation. (*Top*) MCF-7 cells were treated with the nuclear stain Hoechst 33342 (15 μ M), compound **3** (5 μ M), and the dead cell stain Sytox Orange (0.5 μ M). Fluorescence signals from Hoechst (*Upper left*, blue) and compound **3** (*Upper center*, green) colocalize in cell nuclei. (*Upper right*) Overlay of the visible light image (grayscale) and the Sytox Orange fluorescence image (red), indicating that the majority of cells are alive. (*Bottom*) Jurkat cells were treated similarly, and show similar results. (Scale bar = 10 μ m.)

afforded compound **3**, which, surprisingly, showed excellent nuclear staining properties in the cell lines examined. Attachment of the Bodipy-FL fluorophore to the polyamide precursor of compound **3** provided compound **4**. This molecule showed no nuclear staining, sequestering itself in cytoplasmic vesicles, indicating that some characteristic of polyamide-Bodipy conjugates differing from that of polyamide-FITC conjugates, and not the C-terminal β -alanine, prevents their trafficking into the nucleus. It is interesting to

note that the structure of LysoTracker Red DND-99 (**23**) includes both a tertiary alkyl amine, similar to that often used in hairpin polyamide tails, and a Bodipy moiety.

It became our intent to explore the structure-space of polyamide-fluorophore conjugates to overcome cellular exclusion and lysosomal trapping, allowing the polyamides to travel to the nucleus. To explore the criteria that permit uptake of **3**, and that prevent that of **4**, several variations on the structure of the compound were made and their effects determined by confocal microscopy.

Wishing to explore the effect of Py/Im sequence and content of hairpin polyamides on cellular trafficking, we synthesized compounds **5-8**. Compounds **5** and **6**, both containing two imidazole residues, showed a high degree of nuclear staining in all cell lines studied. Three-imidazole compound **7** showed intermediate levels of nuclear staining in the cell lines studied. The nuclear localization of four-imidazole compound **8** was quite poor in all cell lines tested. The difference in nuclear staining levels exhibited by compounds **3** and **7** shows that Py/Im sequence alone is an important determinant of nuclear uptake, though overall content (in terms of the number of Py and Im residues) may also be a factor.

To explore the effect of the positively-charged linker on nuclear uptake, compound **9** was synthesized. This agent showed nearly global abrogation of nuclear uptake efficiency versus the analog containing a tertiary amine. Further, substitution of the γ -aminobutyric acid turn (γ -turn) of **9** with the [(R)- α -amino]- γ -diaminobutyric acid turn ($^{H2N}\gamma$ -turn) provided **10**, which restored most of the nuclear uptake properties of **3**. This suggests that the overall charge of the molecule, and perhaps the placement of that charge, is important variables in nuclear uptake of polyamide-dye conjugates.

The $\text{H}^{2\text{N}}\gamma$ -turn is a structural element commonly included in polyamide design. Replacement of the γ -turn of **3** with the $\text{H}^{2\text{N}}\gamma$ -turn provided **11**. This molecule showed nuclear staining in all cell lines tested, save NB4, though often to a lesser degree than **3**. It is unclear whether this reduction in uptake efficiency is due to an increase in the overall positive charge of the molecule, a more branched structure than the linear γ -linked hairpin, or the positioning of a positive charge medial in the molecule.

Selective acetylation of the polyamide precursor to **11**, followed by FITC conjugation, provided **12**. This molecule exhibited excellent nuclear uptake, as good or better than both **11** and **3**. The excellent uptake of **12** argues against branching as a negative determinant of nuclear staining. This result also prompted us to synthesize **13-15** to probe the generality and flexibility of turn acetylation across several polyamide sequences. The uptake of **13** and **14** was mostly nuclear. Compound **15**, on the other hand, was a very poor nuclear stain. This result reaffirms the importance of polyamide sequence on nuclear uptake.

We next synthesized several conjugates to explore the uptake effects of different points of attachment of the fluorophore-linker moiety to the polyamide. An *N*-propylamine linkage from the terminal imidazole and a methylamide tail were incorporated into **16**, which showed poor nuclear uptake in nearly all cell lines tested. We thought it possible that the poor uptake properties of conjugate **16** was due to some non-linearity introduced into the overall structure of the molecule by the linkage at the 1-position of the *N*-terminal ring of the polyamide. Consequently, conjugates **17** and **18** were synthesized, appending the linker-fluorophore moiety to the *N*-terminal amine. Although compound **17** has an overall structure and charge distribution very similar to

that of **9**, it possesses better nuclear uptake properties, particularly in suspended cell lines. Subsequent addition of a positive charge at the C-terminal end of the polyamide through employment of an *N,N*-dimethylaminopropylamine tail (compound **18**) boosts uptake in all cell lines studied, save NB4. Once again, this suggests that the addition or deletion of a single charge may (but not necessarily will) have a strong effect on nuclear uptake. In fact, a comparison of compounds **3**, **10**, and **18** (differing in the placement of a positive charge) with **9** and **17** (which lack a positive charge) seems to indicate that the positioning of this charge is less influential than its presence.

Another common structural feature utilized in DNA-binding polyamides is the addition of one or more β -alanine residues. This amino acid is often used to relax overcurvature of a polyamide backbone with reference to the minor groove of DNA and permits the targeting of longer sequences than that of eight-ring hairpins. Compound **19**, which includes this structural element, and **20**, which includes both β -alanines and the $\text{H}_2\text{N}\gamma$ -turn, showed fair to poor nuclear uptake in the cell line series. The reduced uptake levels could be due to the increased molecular flexibility imparted by the β -alanine residues, increased molecular weight, decreased recognition by a cellular import protein, or some combination of these criteria.

In order to determine whether or not the added size imparted to **19** and **20** by the β -alanine residues was likely a major contributor to their poor uptake, the ten-ring compound **21** was synthesized. Its ubiquitously poor nuclear uptake suggests that molecular weight may, indeed, play a prominent role in uptake character. The excellent uptake character of the six-ring compound **22** furthers this hypothesis, though Py/Im sequence in the 6-ring and 10-ring contexts is also likely to be important.

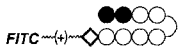
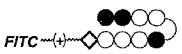
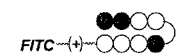
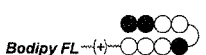

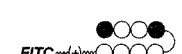
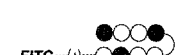
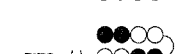
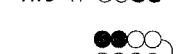
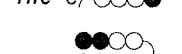
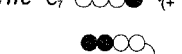







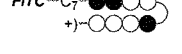
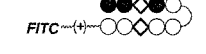

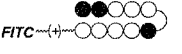
		DLD-1	HeLa	MCF-7	SK-BR-3	786-O	293	LN-CaP	PC3	MEL	NB4	Jurkat	CCRF-CEM	MEG-01
	1	+	++	++	++	+	+	++	++	+	++	++	++	++
	2	-	--	--	--	--	--	--	--	--	--	--	--	--
	3	+	++	++	++	++	++	++	++	++	++	++	++	++
	4	-	-	-	-	--	--	--	-	-	-	-	--	-
	5	++	++	++	++	++	++	++	++	++	++	++	++	++
	6	++	++	++	++	++	++	++	++	++	++	++	++	++
	7	+	++	+	+	+	+	+	+	--	-	++	++	+
	8	-	+	+	-	--	-	-	-	--	--	--	-	--
	9	-	--	+	+	--	--	-	-	--	--	-	-	--
	10	+	++	+	+	+	--	++	+	+	+	++	+	+
	11	+	++	+	+	+	++	++	++	+	--	++	+	++
	12	++	++	++	++	++	++	++	++	++	++	++	++	++
	13	++	++	++	++	++	++	++	++	++	+	++	++	++
	14	++	++	++	++	+	++	++	++	+	-	++	++	++
	15	--	--	--	--	--	-	--	--	--	--	--	--	--
	16	+	--	+	+	--	-	--	-	--	-	+	-	--
	17	+	-	++	+	--	--	--	+	+	+	+	+	+
	18	+	++	++	+	-	+	-	+	++	-	++	+	+
	19	-	+	-	+	-	--	+	+	+	++	-	--	+
	20	-	++	+	-	+	-	+	+	--	--	+	+	+
	21	-	-	--	-	--	--	-	--	--	--	--	--	--
	22	++	++	++	++	++	++	++	++	++	++	++	++	++

Figure 4. Uptake profile of compounds 1-22 in 13 cell lines: “+ +” Indicates nuclear staining exceeds that of the medium; “+” indicates nuclear staining \leq that of the medium, but still prominent; “-” indicates very little nuclear staining, with the most fluorescence seen in the cytoplasm and/or medium; “--” indicates no nuclear staining.

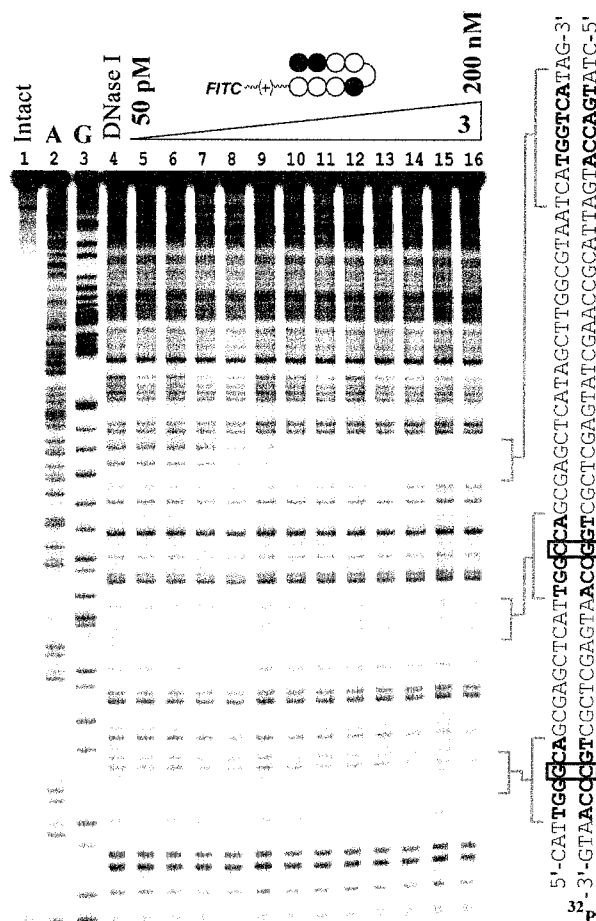


Figure 5. Quantitative DNase I footprinting titration. Compound **3** binds the 3'-TGGTCA-5' site with an affinity $K_a = 1.6 (\pm 0.3) \times 10^9 \text{ M}^{-1}$. Compound **3** does not bind the single base-pair mismatch sites shown at concentrations $\leq 200 \text{ nM}$.

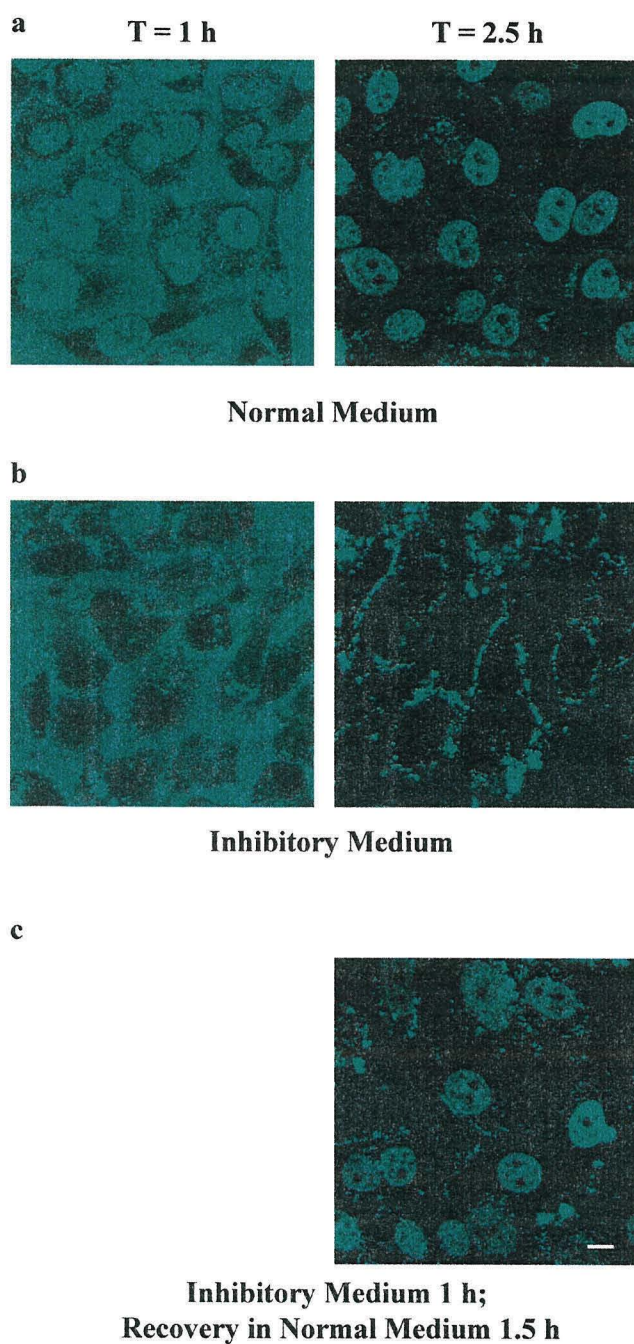


Figure 6. ATP Dependence of nuclear localization of compound 3. HeLa cells were grown and plated as usual. (a) HeLa cells were washed, incubated in normal DMEM medium for 30 min, supplemented with 5 mM 3, incubated for 1 h, and imaged, showing roughly equivalent amounts of compound localized both in the nucleus and in the medium (left). The cells were incubated for a further 1 h 30 min, and then imaged once more (right), showing exclusively nuclear localization. (b) HeLa cells were washed, incubated in inhibitory DMEM medium for 30 min, supplemented with 5 mM 3, incubated for 1 h, and imaged, showing localization at the cellular membranes, as well as in the medium (left). The cells were incubated for a further 1 h 30 min, and then imaged once more (right), showing exclusively membranous localization. (c) Recovery experiment: HeLa cells treated as in b were washed 2x with normal DMEM medium, incubated for 1 h 30 min in normal DMEM medium supplemented with 5 mM 3, then imaged, showing recovery of nuclear localization upon reinitiation of ATP synthesis.

Conclusions

This study demonstrates the cellular localization profile of a host of polyamide-dye conjugates in a wide variety of cell lines. In general, compounds exhibiting good nuclear uptake properties have several common elements: an eight-ring polyamide DNA-binding domain, one or more positive charges incorporated within either the linker or the turn residue, and a conjugated fluorescein fluorophore. This study also demonstrates that each cell line possesses a unique uptake profile for the panel of compounds presented to it. These profiles will be important in choosing a cell line and compound architecture appropriate to a given experiment.

Conjugation of the fluorophore to the polyamide DNA-recognition domain results in an approximately 10-fold reduction in DNA-binding affinity as compared with the parent polyamide, with retention of binding specificity over mismatch sites. This quality, along with the nuclear uptake results, suggests that fluorophore-conjugated polyamides may be employed directly in experiments designed to take place in living mammalian cells.

Clearly, there are many criteria at work, each one having its share of influence upon nuclear uptake of polyamide-dye conjugates. The extension of the structure-space of compounds known to both bind chromosomal DNA specifically and with high affinity, as well as to traffic to the nucleus of living cells is the object of current investigations. The illumination of these possibilities will permit the further study of transcriptional regulation in living systems.

References

1. Dervan, P.B. & Edelson, B.S. (2003) *Curr. Opin. Struct. Biol.* **13**, 284–299.
2. Dervan, P.B. (2001) *Bioorg. Med. Chem.* **9** 2215-2235.
3. Dickinson, L.A., Trauger, J.W., Baird, E.E., Ghazal, P., Dervan, P.B., & Gottesfeld, J.M. (1999) *Biochemistry* **38**, 10801-10807.
4. Gottesfeld, J.M., Turner, J.M., & Dervan, P.B. (2000) *Gene Expression* **9**, 77-91.
5. Wang, C.C.C. and Dervan, P.B. (2001) *J. Am. Chem. Soc.* **123**, 8657-8661.
6. Ansari, A.Z., Mapp, A.K., Nguyen, D.H., Dervan, P.B., & Ptashne, M. (2001) *Chem. Biol.* **8**, 583-592.
7. Gottesfeld, J.M., Melander, C., Suto, R.K., Raviol, H., Luger, K., & Dervan, P.B. (2001) *J. Mol. Biol.* **309**, 615-629.
8. Ehley, J.A., Melander, C., Herman, D., Baird, E.E., Ferguson, H.A., Goodrich, J.A., Dervan, P.B., & Gottesfeld, J.M. (2002) *Mol. Cell. Biol.* **22**, 1723-1733.
9. Dickinson, L.A., Gulizia, R.J., Trauger, J.W., Baird, E.E., Mosier, D.E., Gottesfeld, J.M., & Dervan, P.B. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 12890-12895.
10. Coull, J.J., He, G.C., Melander, C., Rucker, V.C., Dervan, P.B., & Margolis, D.M. (2002) *J. Virol.* **76**, 12349-12354.
11. Janssen, S., Durussel, T., & Laemmli, U.K. (2000) *Mol. Cell* **6**, 999-1011.
12. Janssen, S., Cuvier, O., Muller, M., & Laemmli, U.K. (2000) *Mol. Cell* **6**, 1013-1024.
13. Chiang, S.Y., Bürli, R.W., Benz, C.C., Gawron, L., Scott, G.K., Dervan, P.B., & Beerman, T.A. (2000) *J. Biol. Chem.* **275**, 24246-24254.
14. Belitsky, J.M., Leslie, S.J., Arora, P.S., Beerman, T.A., & Dervan, P.B. (2002) *Bioorg. Med. Chem.* **10**, 3313-3318.

15. 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. (2003) *Bioorg. Med. Chem. Lett.*, **13**, 1565-1570.
16. Baird, E.E. & Dervan, P.B. (1996) *J. Am. Chem. Soc.* **118**, 6141-6146.
17. Belitsky, J.M., Nguyen, D.H., Wurtz, N.R., & Dervan, P.B. (2002) *Bioorg. Med. Chem.* **10**, 2767-2774.
18. Richardson, W.D., Mills, A.D., Dilworth, S.M., Laskey, R.A., & Dingwall, C. (1988) *Cell* **52**, 655-664.
19. Heckel, A. & Dervan, P.B. (2003) *Chem. Eur. J.* **9**, 3353-3366.
20. Trauger, J.W. & Dervan, P.B. (2001) *Methods Enzymol.* **340**, 450-466.

Chapter 3

Influence of Structural Variation on Nuclear Localization of DNA-Binding Polyamide-Fluorophore Conjugates

The text of this chapter was taken in part from a manuscript coauthored with Timothy P. Best, Bogdan Olenyuk, Nicholas G. Nickols, Raymond M. Doss, Shane Foister, Alexander Hekel, and Professor Peter B. Dervan (Caltech)

(Edelson, B. S.; Best, T. P.; Olenyuk, B.; Nickols, N. G.; Doss, R. M.; Foister, S.; Hekel, A.; Dervan, P. B.; "Influence of Structural Variation on Nuclear Localization of DNA Binding Polyamide-Fluorophore Conjugates" *Nuc. Acids Research.*, **2002**, 32, 2802)

Abstract

A pivotal step forward in chemical approaches to controlling gene expression is the development of sequence-specific DNA-binding molecules that can enter live cells and traffic to nuclei unaided. DNA-binding polyamides are a class of programmable, sequence-specific small molecules that have been shown to influence a wide variety of protein-DNA interactions. We have synthesized over 100 polyamide-fluorophore conjugates and assayed their nuclear uptake profiles in thirteen mammalian cell lines. The compiled dataset, comprising 1300 entries, establishes a benchmark for the nuclear localization of polyamide-dye conjugates. Compounds in this series were chosen to provide systematic variation in several structural variables, including dye composition and placement, molecular weight, charge, ordering of the aromatic and aliphatic amino-acid building blocks, and overall shape. Nuclear uptake does not appear to be correlated with polyamide molecular weight or with the number of imidazole residues, although the positions of imidazole residues affect nuclear access properties significantly. Generally negative determinants for nuclear access include the presence of a β -Ala-tail residue and the lack of a cationic alkyl amine moiety, whereas the presence of an acetylated 2,4-diaminobutyric acid-turn is a positive factor for nuclear localization. We discuss implications of this data on the design of polyamide-dye conjugates for use in biological systems.

Introduction

Cell-permeable small molecules that preferentially bind to predetermined DNA sequences inside living cells would be useful tools in molecular biology, and perhaps human medicine. Minor groove-binding polyamides containing the aromatic amino acids *N*-methylpyrrole (Py), *N*-methylimidazole (Im), *N*-methyl-3-hydroxypyrrole (Hp), and other related aromatic heterocycles bind DNA with affinities and specificities comparable to naturally occurring DNA-binding proteins (1-4). DNA recognition by polyamides is described by a code of side-by-side amino acid pairings that are oriented N→C with respect to the 5'→3'-direction of the DNA helix in the minor groove: Im/Py is specific for G•C, Hp/Py is specific for T•A, and Py/Py binds both A•T and T•A (1).

Polyamides have been shown to influence a wide variety of protein-DNA interactions in solution (5-15), yet similar experiments in cell culture have proven to be dependent on cell-type (16-18). Studies with fluorescent bodipy-labeled polyamides indicate that these conjugates are excluded from the nuclei of most cells, with the notable exceptions of lymphoid and myeloid cell types (19,20). One report suggested that the fluorophore itself may play a role in cellular uptake: an eight-ring polyamide-fluorescein conjugate was shown to accumulate in the nuclei of HCT-116 colon cancer cells (21). Expanding upon this lead, we recently described a set of polyamide-fluorescein conjugates that localized to the nuclei of thirteen live mammalian cell lines (22). Relatively small structural alterations, such as differences in Py/Im sequence, caused dramatic changes in nuclear localization. The current study seeks to elucidate more fully the structural requirements for nuclear localization of polyamide-fluorophore conjugates. This is a minimum first step in exploring the use of these DNA-binding ligands for gene

modulation in cell culture experiments.

Materials and Methods

Chemicals

Polyamides were synthesized by solid phase methods on Boc- β -ala-PAM resin (Peptides International, Louisville, KY) (23) or on Kaiser oxime resin (Nova Biochem, Laufelfingen, Switzerland) (24). Synthetic protocols for second-generation building blocks (incorporated into polyamides shown in Figure 3 and polyamides **90** and **91**, Figure 13) are essentially as described (2-4,25). U-pin **93** and H-pins **94-99** were synthesized according to methods described in (26) and (27), respectively. After cleavage with the appropriate amine and reverse-phase HPLC purification, polyamides were dissolved in DMF and treated with diisopropylethylamine (DIEA) (20 eq) followed by the fluorophore in the form of an isothiocyanate, an *N*-hydroxysuccinimidyl ester, or a free acid activated *in situ* with HBTU. Fluorophores were delivered as solutions in DMF or DMSO. After reacting at rt for ~3 h, the resulting dye conjugates were purified by HPLC. Peracetylated polyamides **17**, **34**, **95**, **98**, and **99** were obtained by treating the precursor polyamides with acetic anhydride and DIEA in DMF for ~30 min. For compounds with a free (*R*)-2,4-diaminobutyric acid ($^{\text{H}_2\text{N}}\gamma$ -turn) moiety (**29-31**, **76**, **78**, **79**, and **100**), the protected $^{\text{FmocHN}}\gamma$ -turn amine was deprotected (20% piperidine-DMF) and reprotected as the Boc derivative (Boc₂O, DIEA, DMF) immediately prior to cleavage of the polyamide from resin. After conjugation of the dye moiety, these compounds were treated with TFA and purified by HPLC. The identity and purity of each compound was verified by analytical HPLC, UV-visible spectroscopy, and matrix-assisted laser

desorption ionization/time-of-flight mass spectrometry (MALDI-TOF). All fluorescent dye reagents were from Molecular Probes. Chemicals not otherwise specified were from Sigma-Aldrich.

Cell Cultures

The human cancer cell lines MCF-7, PC3, LNCaP, DLD-1, 786-O, Jurkat, CCRF-CEM (CEM), MEG-01, and NB4 were cultured in a 5% CO₂ atmosphere at 37°C in supplemented RPMI medium 1640. The human cancer cell line HeLa, the murine leukemia cell line MEL, and the transformed human kidney cell line 293 were grown as above in supplemented DMEM. The human cancer cell line SK-BR-3 was cultured as above in supplemented McCoy's medium. All media were supplemented with 10% fetal bovine serum (Irvine Scientific) and 1% penicillin/streptomycin solution (Sigma).

Confocal Microscopy

Adherent cell lines (MCF-7, HeLa, PC3, LNCaP, SK-BR-3, DLD-1, 786-O, and 293) were trypsinized for 5–10 min at 37°C, centrifuged for 5 min at 5°C at 900 g, and resuspended in fresh medium to a concentration of 1.25×10^6 cells per ml. Suspended cell lines (Jurkat, CEM, MEG-01, MEL, and NB4) were diluted in fresh medium to the same concentration. Incubations were performed by adding 150 µl of cells into culture dishes equipped with glass bottoms for direct imaging (MatTek, Loveland, OH).

Adherent cells were grown in the glass-bottom culture dishes for 24 h. The medium was then removed and replaced with 147 µl (or 142.5 µl) of fresh medium. Then 3 µl (or 7.5 µl) of the 100 µM polyamide solution was added for final polyamide concentration of 2

μM (or 5 μM), as noted in the data tables. Cells were incubated in a 5% CO_2 atmosphere at 37°C for 10–14 h. Suspended cell line samples were prepared in a similar fashion, omitting trypsinization. These samples were then incubated as above for 10–14 h. Imaging was performed with a x40 oil-immersion objective lens on a Zeiss LSM 5 Pascal inverted laser scanning microscope. The optical slice was set to 2.2 μm . Images were line-averaged 2, 4, 8, or 16 times and were obtained at a 0.8 $\mu\text{s}/\text{pixel}$ scanning rate. Polyamide-dye conjugate fluorescence and visible light images were obtained using standard filter sets appropriate for fluorescein, rhodamine, or Texas Red.

Results

We have synthesized 100 polyamide-fluorophore conjugates, whose chemical structures are shown along with schematic ball-and-stick representations (e.g., Figure 1). We examined the nuclear localization properties of these compounds in thirteen mammalian cell lines, representing eleven human cancers, one human transformed kidney cell line, and one murine leukemia cell line. Cells were allowed to incubate with polyamide-fluorophore conjugate in the surrounding medium (2 or 5 μM , as indicated, 10–14 h) and the extent of nuclear uptake was analyzed by confocal microscopy. Polyamide-dye conjugates display a range of uptake efficiencies, varying from strong nuclear concentration, producing brightly stained nuclei with little or no signal in the medium or cytoplasm, to an absence of nuclear accumulation, in which case the molecules may be trapped in vesicles or excluded from cells entirely. Within a single sample, staining is usually remarkably homogenous, displaying minimal cell-to-cell variation. The extent of nuclear localization for each sample was assigned one of four qualitative ratings, and

these data are organized into tables showing the ball-and-stick structure of each compound and the level of uptake in each cell line (e.g., Figure 2). A thin vertical line separates data collected in adherent cell lines (at left, MCF-7 through 293) from data

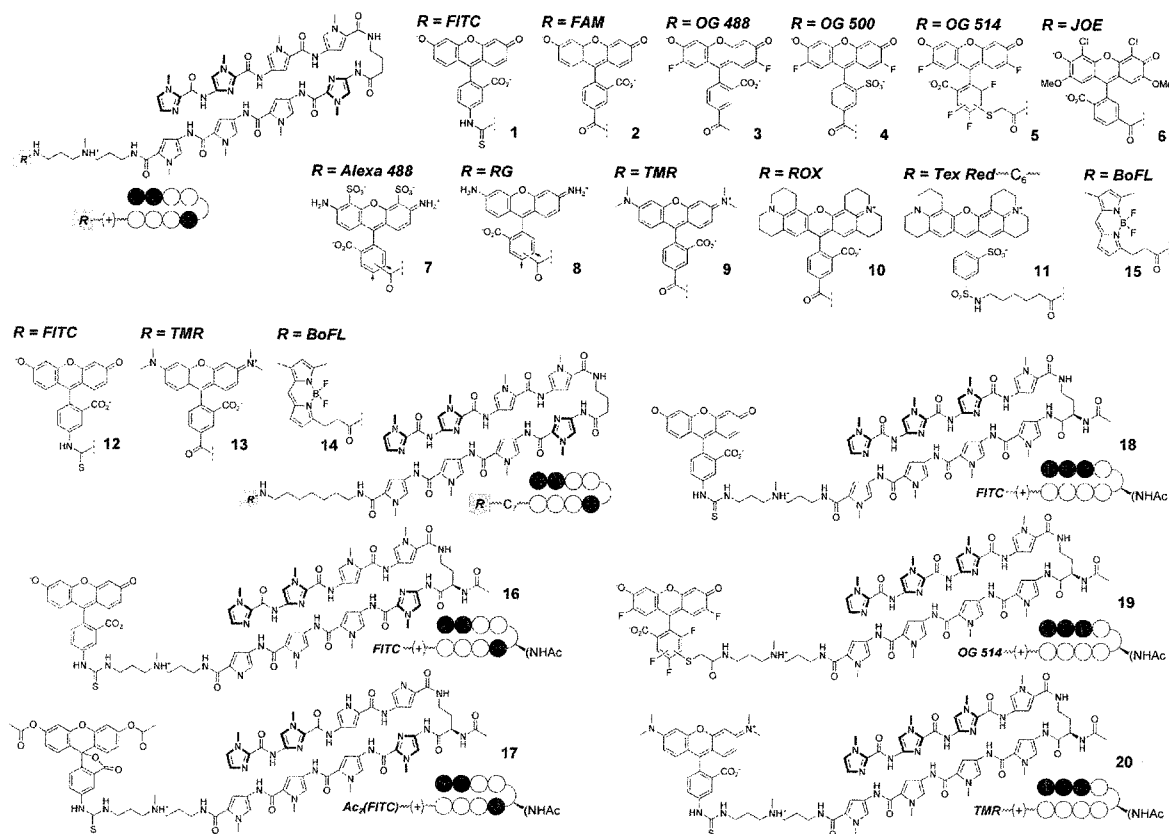


Figure 1. Dye composition: chemical and ball-and-stick structures for polyamides 1-20.

collected in suspended cell lines (right, Jurkat through NB4). Within these two sets, more permissive cell lines – those that generally display stronger nuclear staining with polyamide conjugates – are towards the left. Thus, MCF-7 and Jurkat are the most permissive cell types of the adherent and suspended cells, respectively. Shaded groupings of compounds indicate structurally related conjugates, as discussed in the text.

Figure 2. Dye composition: uptake profile of polyamides 1-20 in thirteen cell lines. Alternating highlighting indicates groups of chemically similar polyamides, as described in the text. Data for adherent and suspended cells are towards the left and right, respectively, separated by a light gray vertical line. Lighter shades of blue represent stronger nuclear localization. ++, 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. Except where noted, polyamide concentration was 2 μM . ^aPolyamide described previously (22), assayed at 5 μM . ^bAssayed at 5 μM .

staining in the cell lines tested. There appears to be flexibility in the specific type of chemical linkage, as thiourea (**1**, resulting from FITC conjugation) and amide (**2-6**) linkers are effective, including the extended thioether amide-linkage of Oregon Green[®] (OG) 514 in compound **5**. Fluorination of the aromatic rings is well tolerated (**3-5**), and replacing the carboxylic acid with a sulfonic acid (as in OG 500 conjugate **4**) appears to be somewhat beneficial in several cells lines (compare compounds **3** and **4**). The greater photostability of the fluorinated OG derivatives, as compared to fluorescein, may be advantageous for fluorescence imaging applications (28,29). The JOE fluorophore, a fluorescein derivative with chloro and methoxy substituents, has red-shifted absorption and emission spectra compared to fluorescein, such that their spectra can be distinguished. JOE conjugate **6** shows moderate to strong uptake in all cell lines, providing a second color for multicolor fluorescence applications.

Compounds **7-11**, conjugated to rhodamine derivatives, display poorer nuclear localization profiles than analogous fluorescein conjugates. The sulfonated rhodamine-dye Alexa 488 is a highly photostable alternative to fluorescein; however, conjugate **7** was completely excluded from the nuclei of all cells tested. Rhodamine green derivative **8**, lacking sulfonates, showed slightly stronger nuclear localization, particularly in adherent cell lines. The additional alkyl groups in tetramethyl rhodamine (TMR) conjugate **9** impart enhanced uptake properties in most cell lines, compared to **8**. Depending upon the cell type (as well as the ring sequence and composition of the polyamide, *vide infra*), polyamide-TMR conjugates may be useful because of the photostable, red/orange fluorescence of the TMR dye. Compound **10** is conjugated to the related ROX fluorophore, and conjugate **11** incorporates a Texas Red fluorophore with a

hexanoic acid spacer, a dye/linker combination that has been used to label polyamides for telomere staining experiments (30). Both polyamides, **10** and **11**, are consistently excluded from live cells.

In compounds **12-14**, a seven-carbon alkyl chain links the dye – FITC, TMR, and bodipy FL (BoFL), respectively – to the polyamide. For comparison, BoFL derivative **15**, incorporating the standard cationic triamine linker, is also included. As described previously, the BoFL conjugate in this motif displays only minimal evidence of nuclear access in live cells (22). For each alkyl-linked conjugate, the loss of the tertiary amine in the linker worsened the uptake profile (compare **1** and **12**, **9** and **13**, and **15** and **14**). It was shown previously that alkyl-linked conjugate **12** performed much more poorly than triamine-linked conjugate **1**, and that addition of an amino group at the turn residue, using the chiral $\text{H}_2\text{N}\gamma$ -turn, restored most of the uptake properties (see compound **78**, below) (22). In contrast adding amino groups to the dye, as in alkyl-linked TMR conjugate **13**, abolishes nuclear access.

Acetylating fluorescein produces a non-fluorescent, uncharged moiety; cleavage of the acetates by esterases subsequently unmask the fluorophore. Polyamide **16**, as described previously (22), demonstrates that the $\text{AcHN}\gamma$ -turn does not impede uptake, and the analogous diacetyl-FITC conjugate **17** stains nuclei with similar effectiveness. It should be noted that no special effort was taken to remove esterases from the growth medium, such that some amount of **16** is likely to be in solution when cells are treated with **17**.

Polyamides **18-20** have the same number of Im and Py residues as compounds **1-17** but in a different sequence. With the exception of the less permissive suspended cells

MEL and NB4, FITC conjugate **18** (an isomer of **16**) and OG 514 conjugate **19** display modest to strong nuclear staining. In contrast, TMR conjugate **20** is a very poor nuclear stain. Comparing analogous compounds, conjugates **18** and **19** display only slightly poorer uptake profiles than those of conjugates **1** and **5**, respectively, whereas TMR conjugate **20** is a much worse nuclear stain than polyamide **9**. Such comparisons indicate that modifications such as ring sequence and fluorophore structure interact to affect nuclear uptake in ways that are not yet predictable.

Second-Generation Rings

We have recently described several second-generation aromatic ring systems that improve the DNA-binding properties of polyamides (3,4). When paired against Py, a chlorothiophene (Ct) residue at the N-terminal cap position targets a T•A base pair (4). The hydroxybenzimidazole (Hz) moiety, incorporated as part of a dimeric subunit, appears to be a chemically stable replacement for the Hp residue (3), and a CtHz-cap dimer has recently been found to target the sequence 5'-TT-3' (R.M. Doss, M.A. Marques, S. Foister, and P.B. Dervan, unpublished results). Figure 3 shows the chemical structures of polyamide-fluorophore conjugates incorporating second-generation rings, and Figure 4 displays their uptake properties.

Ct-cap compounds **21-24** have the same ring sequence but different dyes. Polyamides **21-23**, conjugated to fluorescein derivatives, display strong nuclear localization, whereas TMR conjugate **24** displays a moderate uptake profile that is generally similar to **9** and better than **20**. Conjugates **25-28** demonstrate the effects of different ring sequences on nuclear uptake of thiophene-cap conjugates. Whereas **25**

displays good to excellent nuclear staining, its isomer **26** is excluded from all cells tested.

Other conjugates with three contiguous Im residues – methylthiophene-cap polyamide

27, and ten-ring hairpin **28** – also display very poor uptake properties. The only exception

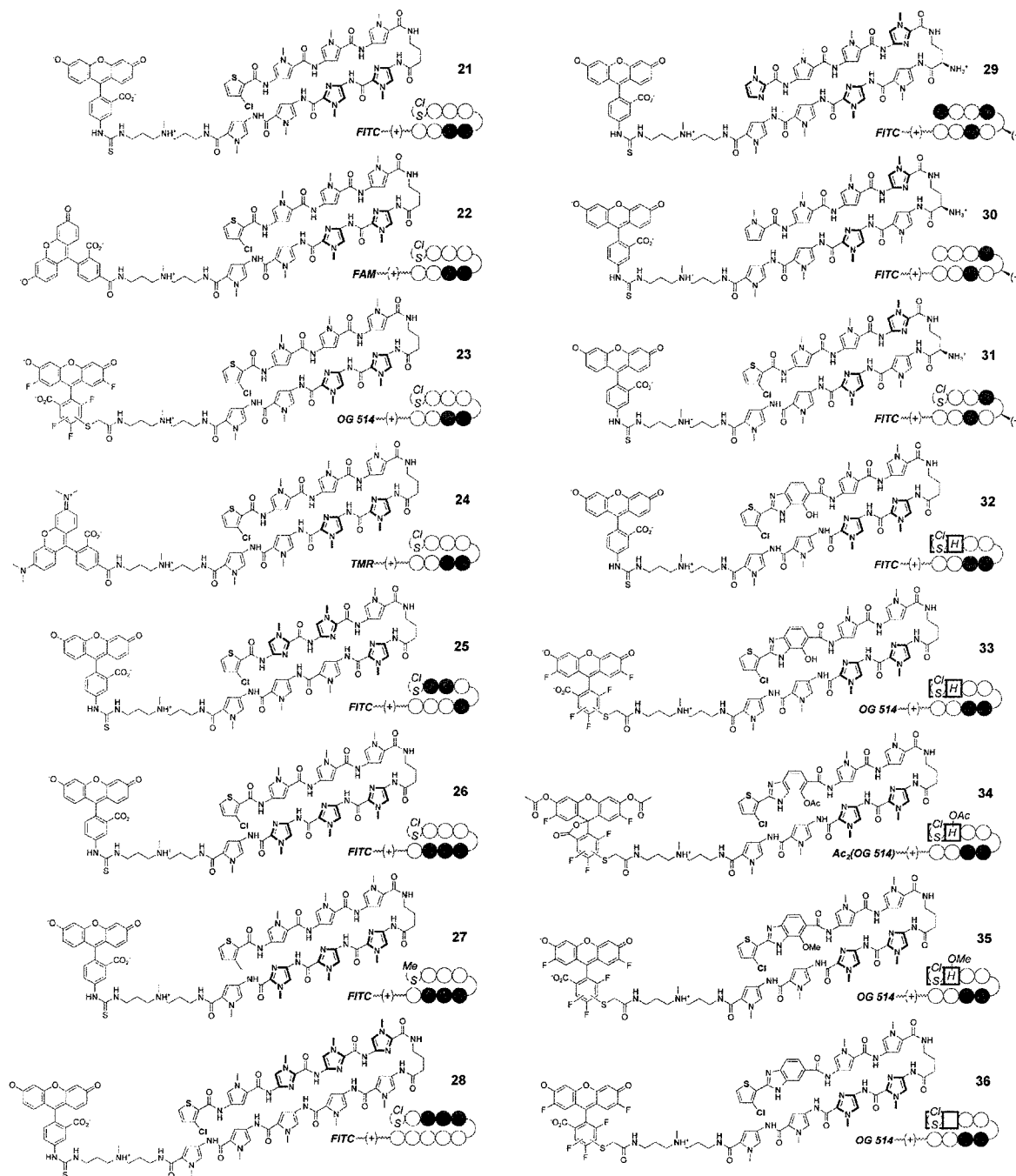


Figure 3. Second-generation rings: chemical and ball-and-stick structures for polyamides 21-36.

is relatively modest and heterogeneous uptake of **27** in MEG-01 cells. Contrasting these

data with the generally good to excellent nuclear staining displayed by **18** and **19**, which also contain three contiguous imidazole residues, reemphasizes the complexity of the interactions between polyamide-fluorophore conjugates and cells.

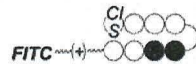
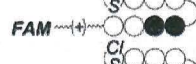

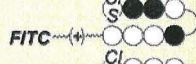

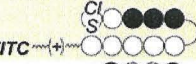
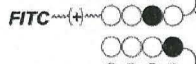
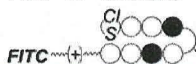
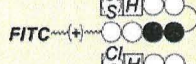

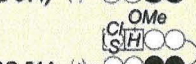
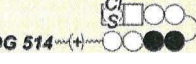
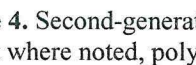
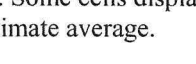
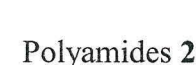

		MCF-7	HeLa	PC3	LN-CaP	SK-BR-3	DLD-1	786-O	293	Jurkat	CEM	MEG-01	MEL	NB4
	21	++	++	++	++	++	++	++	++	++	++	++	++	+
	22	++	++	++	+	++	++	+	+	++	++	++	+	+
	23	++	++	++	+	++	++	+	++	++	++	++	+	+
	24	+	-	+	+	+	+	--	-	++	+	+	++	+
	25	+	++	++	++	+	+	+	+	+	+	+	+	+
	26	--	--	--	--	--	--	--	--	--	--	--	--	--
	27	--	--	--	--	--	-	--	--	--	--	+ ^b	--	--
	28	--	--	--	--	--	--	--	--	--	--	--	--	--
	29 ^a	++	++	++	++	++	++	++	++	++	++	++	++	++
	30 ^a	+	+	+	+	-	+	+	+	+	-	-	-	--
	31 ^a	++	++	++	++	+	++	++	+	++	++	++	++	+
	32	+	-	+	--	--	+	--	--	--	--	--	--	--
	33	+	--	--	+ ^b	--	-	--	--	--	--	--	--	--
	34	--	--	--	--	--	--	--	--	--	--	--	--	--
	35	--	-	--	--	--	--	--	--	--	--	--	--	--
	36	-	--	+	+	-	+	--	--	+ ^b	+ ^b	--	+ ^b	+ ^b

Figure 4. Second-generation rings: uptake profile of polyamides **21-36**. Symbols are defined in Figure 2. Except where noted, polyamide concentration was 2 μ M. ^aAssayed at 5 μ M. ^bHighly heterogeneous uptake profile. Some cells display clear nuclear localization, while others show none. The value shown is an approximate average.

Polyamides **29-31**, each with a chiral H_2N γ -turn, differ only in cap-residue structure. Compounds **29** and **31**, with Im and Ct caps, respectively, are excellent nuclear stains, whereas Py-cap polyamide **30** is significantly less effective. It is remarkable that

an Im→Py conversion (**29** to **30**), formally an N→C-H shift, reduces nuclear uptake, but the more dramatic Im→Ct conversion affects uptake very little.

Based on the ring sequence of **21-24**, compounds **32-36** incorporate variants of the benzimidazole ring system. Although polyamides incorporating the Hz unit display favorable DNA-binding properties, FITC conjugate **32** stains the nuclei of only a few cell lines, and OG 514 conjugate **33** stains only MCF-7 and LNCaP cells. Modifications to the hydroxyl group, including acetylation (**34**) and conversion to a methoxy group (**35**), did not improve the uptake profile. Complete removal of the hydroxyl group (**36**) allows nuclear localization in many cell lines, though the uptake profile does not correlate with the general permissiveness of the cell types. Furthermore, removing the hydroxyl eliminates the ability to distinguish between A•T and T•A base pairs at that position (3).

Extended Hairpin Motif

Figure 5 shows the chemical structures of polyamide-dye conjugates **37-46** with C-terminal unpaired rings, and Figure 6 shows their uptake profiles. The molecules are grouped into pairs that differ with respect to the conjugated dye. Although the DNA-binding properties of the extended hairpin motif have not been studied as extensively as those of fully ring-paired hairpin polyamides (31,32), extended hairpin polyamides have been shown to interfere with NF-κB—DNA interactions (13) and to label specific heterochromatic regions of human chromosomes (29). Furthermore, tandem polyamides with extended hairpin subunits have been used to stain telomeres (30) and to displace a viral transcription factor from DNA (33). Overall, conjugates **37-46** display good to excellent nuclear localization properties.

Seven-ring polyamides **37** and **38**, conjugated to FITC and OG 514, respectively, are moderately effective nuclear stains. The larger nine-ring FITC conjugate **39** displays a significantly improved uptake profile, whereas OG 514 conjugate **40** is similar to the

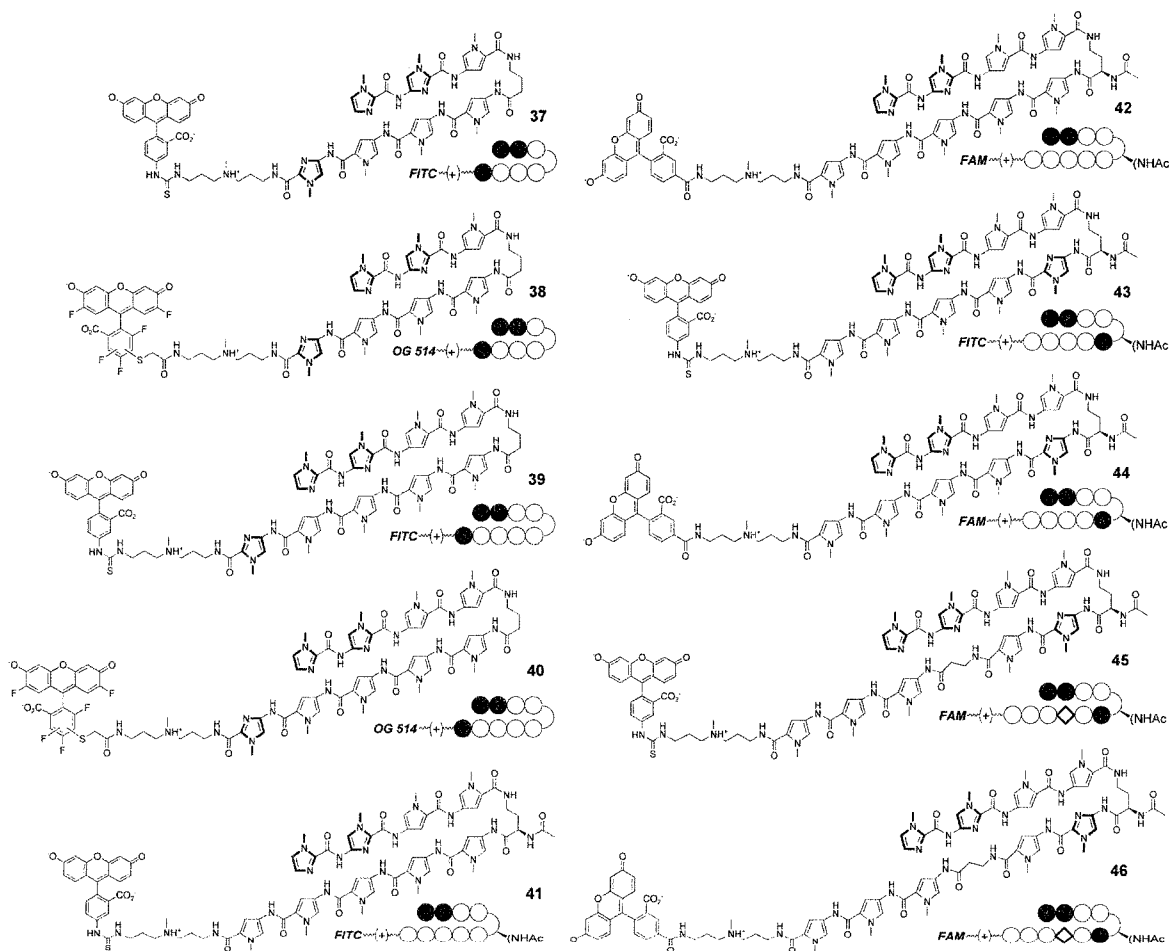


Figure 5. Extended hairpins: chemical and ball-and-stick structures for polyamides **37-46**.

seven-ring hairpins. Polyamides **41** and **42** have a similar nine-ring core, but with a single unpaired Py residue and a chiral $^{AcHN}\gamma$ -turn. Both the FITC and FAM conjugate (**41** and **42**, respectively) are excellent nuclear stains. From these compounds, one Py \rightarrow Im substitution gives **43** and **44**, respectively. In this case, FAM conjugate **44** performs better than FITC conjugate **43**. Based on a motif that was used in studies with NF- κ B (13), compounds **45** and **46** have two unpaired Py residues. Again, the FAM conjugate

displays the better uptake profile, strongly staining the nuclei of eight of the thirteen cell lines.

		MCF-7	HeLa	PC3	LN-CaP	SK-BR-3	DLD-1	786-O	293	Jurkat	CEM	MEG-01	MEL	NB4
	37	++	+	+	+	++	+	-	-	++	++	-	--	--
	38	+	++	+	+	+	+	+	+	+	+	--	--	-
	39	++	++	++	++	+	+	++	+	++	++	-	+	+
	40	+	+	+	-	+	+	+	+	++	+	--	-	+
	41	++	++	++	+	++	++	++	+	++	++	++	+	++
	42	++	++	++	++	++	++	++	++	++	++	++	++	++
	43	+	++	++	+	+	+	+	+	++	+	++	-	--
	44	++	++	++	++	++	+	+	--	++	+	++	++	+
	45	+	+	+	+	+	--	--	+	+	--	-	-	+
	46	++	++	++	+	--	+	+	+	++	++	++	++	++

Figure 6. Extended hairpins: uptake profile of polyamides 37-46. Symbols are defined in Figure 2.

Larger Polyamides

Somewhat larger than the extended hairpin polyamides **37-46** are fully ring-paired hairpins **47-54** and **56-62**. Figure 7 shows the structures of dye conjugates **47-62**, and Figure 8 displays their uptake profiles. Many of these compounds incorporate the flexible β -Ala residue, which is often substituted for Py to allow polyamides to adapt to sequence-dependent DNA microstructure and flexibility (34,35). In general, conjugates of this motif display poor to moderate nuclear staining, although there are scattered exceptional cases.

Conjugate **47**, described previously, displayed a particularly poor uptake profile. From **47**, an Im \rightarrow Py substitution produces **48**, which displays some nuclear staining in

eight cell lines. Chiral $^{AcHN}\gamma$ -turn analogs of **48**, conjugates **49** and **50**, display improved uptake properties, accessing the nuclei of most of the cell types, and showing excellent staining in some of the more permissive adherent cells. Polyamides **51** and **52** (related to

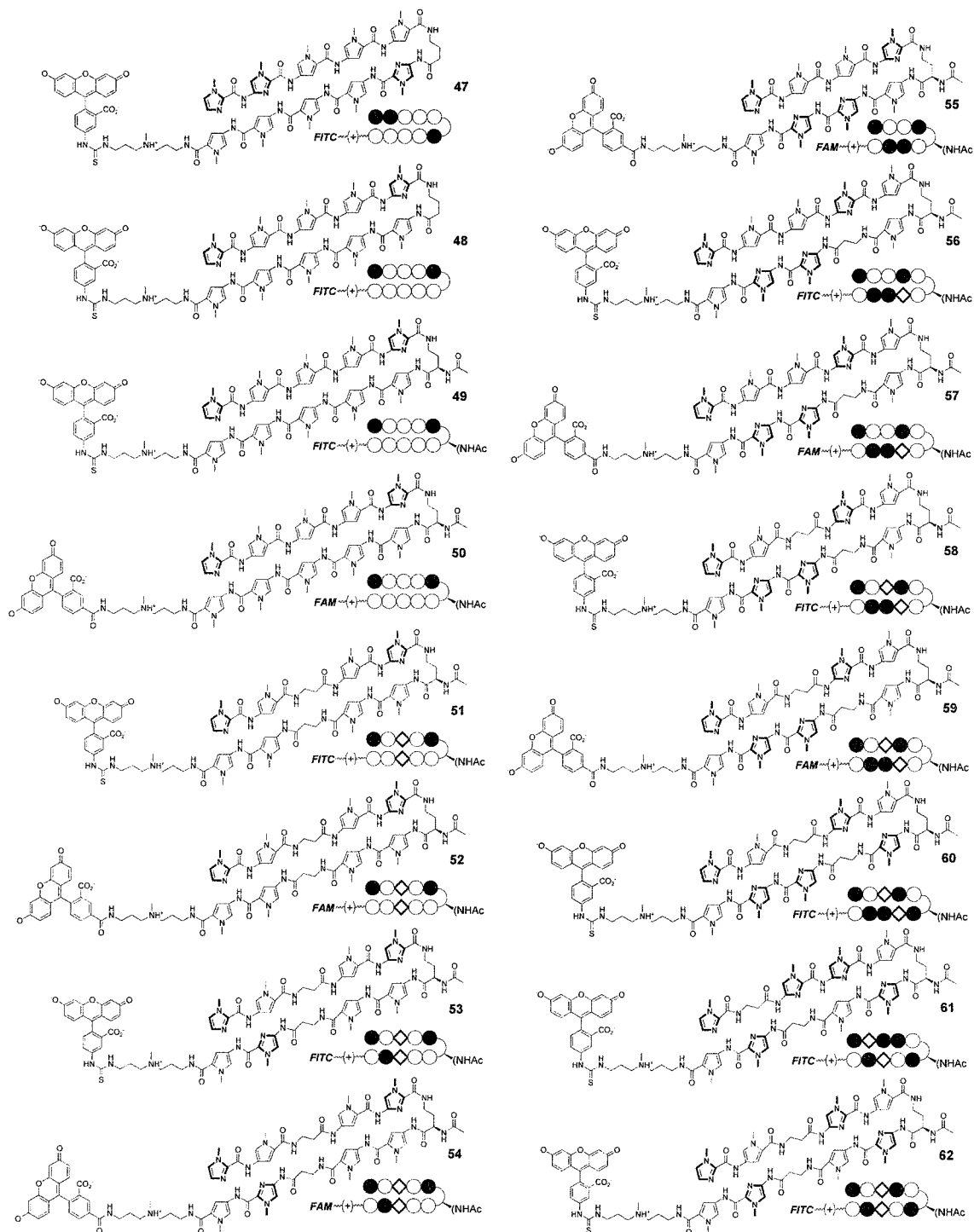


Figure 7. Larger polyamides: chemical and ball-and-stick structures for polyamides 47-62.



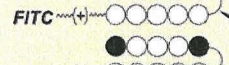










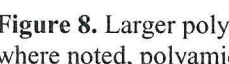

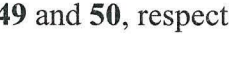
		MCF-7	HeLa	PC3	LN-CaP	SK-BR-3	DLD-1	786-O	293	Jurkat	CEM	MEG-01	MEL	NB4
	47^a	--	-	--	-	-	-	--	--	--	--	--	--	--
	48	+	+	+	--	--	+	-	+	+	+	-	-	-
	49	+	++	++	++	+	+	+	+	+	+	+	+	--
	50	+	++	++	+	++	+	+	-	+	-	+	+	-
	51	+	+	+	+	-	+	+	-	+	--	+	+	+
	52	+	-	+	+	--	--	--	--	-	--	--	+	-
	53	+	++	+	+	+	-	--	--	--	--	--	-	+
	54	+	--	-	-	+	--	--	+	-	++	--	-	+
	55	--	--	--	--	--	--	--	--	--	--	--	--	--
	56	+	-	+	--	--	--	--	--	--	--	--	--	--
	57	+	+	+	+	--	-	--	--	+	-	-	--	--
	58	+	+	+	+	-	-	+	-	--	-	--	-	--
	59	+	+	+	+	-	+	+	-	+	+	-	+	+
	60	-	+	--	+	-	-	-	--	--	--	--	--	-
	61	-	--	--	--	--	--	--	--	--	--	--	--	--
	62	-	-	--	-	-	-	-	-	--	--	--	--	--

Figure 8. Larger polyamides: uptake profile of polyamides **47-62**. Symbols are defined in Figure 2. Except where noted, polyamide concentration was 2 μ M. ^aPolyamide described previously (22), assayed at 5 μ M.

49 and **50**, respectively, by Py \rightarrow β -Ala substitutions of the central residues), are less efficient nuclear stains, with FAM conjugate **52** performing significantly more poorly. Compounds **53** and **54** result from an additional Py \rightarrow Im substitution. FITC conjugate **53** accesses the nuclei of the more permissive adherent cells, whereas FAM conjugate **54** displays a scattered uptake profile.

Polyamides **55-59** are designed to target essentially the same 6-bp DNA sequence. Eight-ring hairpin **55**, included as a reference, is excluded from all cell nuclei, whereas the larger polyamides access nuclei somewhat more successfully. In this set, internal β -

Ala substitution and FAM-conjugation are both positive determinants for nuclear uptake; accordingly, conjugate **59** is the most effective nuclear stain. Polyamide **60** is related to **58** by a Py→Im substitution, and **60** displays a much poorer uptake profile. Inverting the strands of **60** produces **61** – the result of a formal 180° rotation of the ring core – which is excluded from the nuclei of all cells tested. Conjugate **62** is based on a motif that has been employed in a variety of studies (18,20,36); however this FITC conjugate is excluded from cell nuclei.

C-terminal β -Ala Residues

In a previous study, we noted that adding β -Ala residue to the C-terminal (“tail”) position of a polyamide-dye conjugate could have dramatic effects on nuclear localization (22).

We therefore synthesized the molecules shown in Figure 9 to investigate the effects of β -Ala-tail residues. To allow for the possibility that the added β -Ala would enhance uptake, precursor ring systems were chosen to have a range of uptake efficiencies (Figure 10).

The polyamides are grouped by their core ring sequence. In almost every case, adding a β -Ala-tail residue reduces nuclear uptake efficiency. Indeed, even in cases in which both samples have the same qualitative rating, the conjugate with a β -Ala-tail shows slightly less intense nuclear staining.

Conjugate **63** is an excellent nuclear stain, and addition of a β -Ala-tail residue (compound **64**) affects uptake only minimally. In contrast, polyamide **65** displays a moderate to good uptake profile, and its β -Ala-tail analog **66** shows very poor nuclear staining. Although FITC conjugate **68** (an isomer of polyamide **1**) accesses nuclei slightly more effectively than FAM conjugate **67**, it is nonetheless a rather weak stain, and

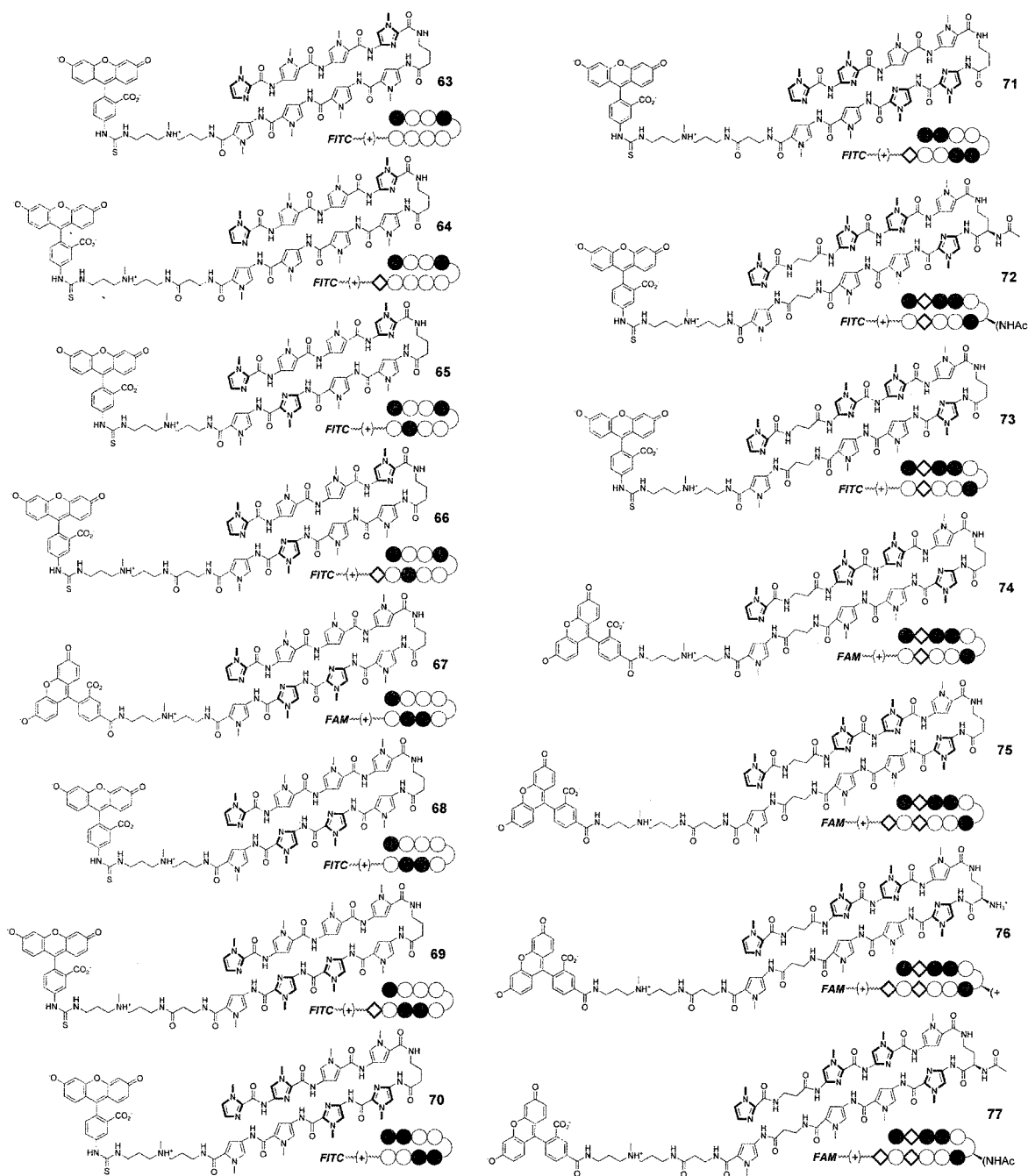


Figure 9. β -Ala tail residues: chemical and ball-and-stick structures for polyamides 63-77.

addition of a β -Ala-tail residue (conjugate **69**) completely inhibits nuclear access.

Conjugate **70**, described previously (22), accesses the nuclei of only two cell lines, and addition of a β -Ala-tail residue (compound **71**) does not change the uptake profile

Turn-linked conjugates

To investigate other dye-attachment points, conjugates were synthesized linked to the chiral turn element (**80-88**, Figure 11). For comparison, this set includes tail-conjugated polyamides **78**, **79**, and **89**. Figure 12 presents the uptake profiles of **78-89**.

Polyamides **78** and **79** are conjugated to FITC through seven- and six-methylene spacers, respectively. Neither compound stains nuclei as effectively as the analogous conjugate **1**, although **79** is quite close. Compared to **78**, polyamide **80** exchanges the positions of the cationic amine group and the dye linker. These turn-linked FITC conjugates are excellent nuclear stains, and the presence of an additional amine (in **81**) or a β -Ala-tail residue (in **82**) does not impede nuclear uptake. The sole exception is compound **81** in NB4, a cell line which was previously observed to exclude polyamide-dye conjugates with added positive charges (22).

Hairpin polyamides **83-88**, linked to BoFL through the turn, differ with respect to Py/Im composition, and are analogous to the tail-linked FITC conjugates **1**, **89**, **65**, **68**, **70**, and **48**, respectively. Although each eight-ring BoFL turn-conjugate is a less effective stain than the corresponding FITC tail-conjugate, relative uptake efficiency within each motif appears to be influenced primarily by ring composition. Indeed, ranked by average effectiveness, $83 \approx 84 > 85 > 86 \approx 87$, and correspondingly, $1 \approx 89 > 65 > 68 \approx 70$. Ten-ring BoFL-turn conjugate performs similarly to its FITC-tail analog **48**, although the $\text{AcHN}\gamma$ -turn derivative **49** is a significantly better nuclear stain than either conjugate. The favorable uptake properties of some of these BoFL-turn conjugates – particularly in more permissive cell lines – is somewhat surprising in comparison to the poor nuclear

localization profile of tail-linked BoFL conjugate **15** (Figure 2).

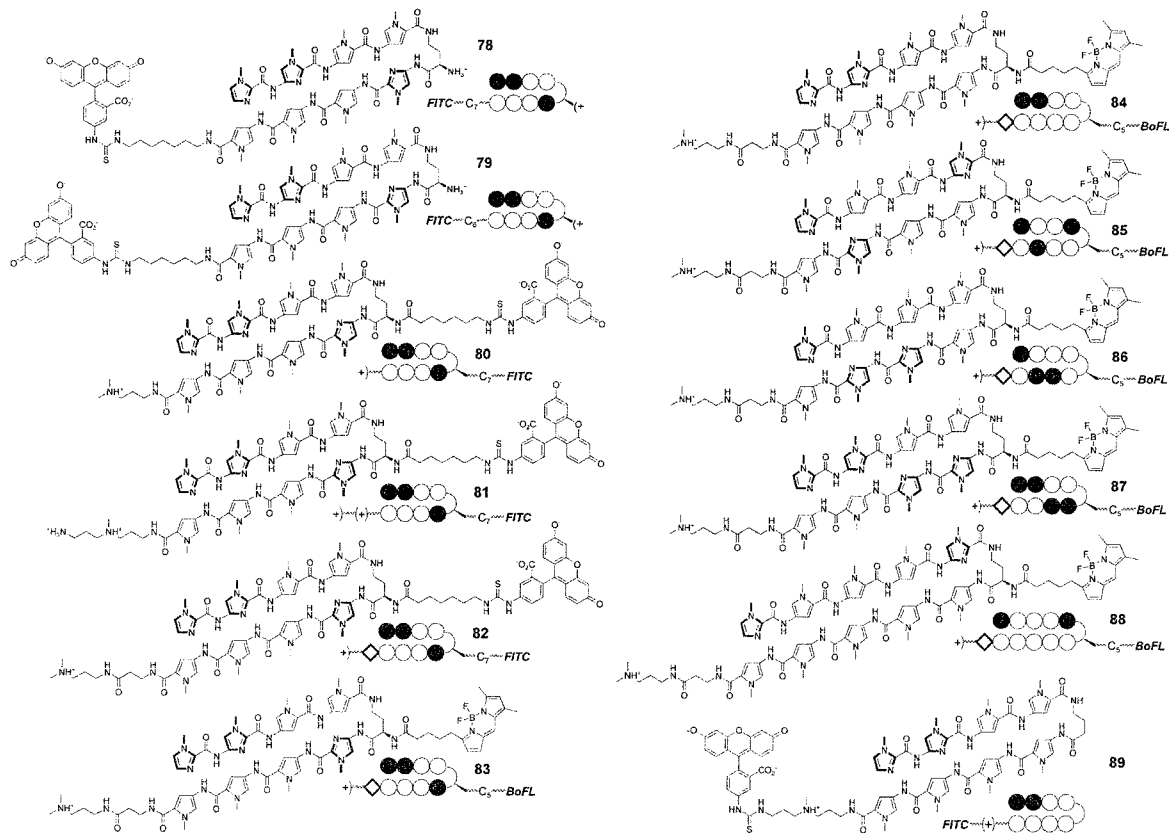


Figure 11. $H_2N\gamma$ -Turn-linked conjugates: chemical and ball-and-stick structures for polyamides **78**–**89**.


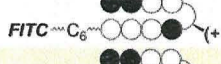
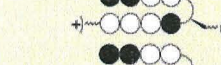







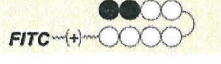
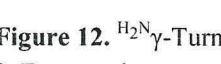
		MCF-7	HeLa	PC3	LN-CaP	SK-BR-3	DLD-1	786-O	293	Jurkat	CEM	MEG-01	MEL	NB4
	78 ^a	+	++	+	++	+	+	+	+	+	++	+	+	+
	79	++	++	++	++	++	++	+	++	++	++	++	++	++
	80 ^b	++	++	++	++	++	++	++	++	++	++	++	+	+
	81 ^b	++	++	++	++	++	++	++	++	++	++	++	++	--
	82 ^b	++	++	++	++	++	++	++	++	++	++	++	+	+
	83	+	++	++	++	+	--	--	--	+	+	--	-	+
	84	+	++	++	++	+	-	-	--	+	+	--	-	+
	85	-	+	+	++	-	--	-	--	+	+	--	--	-
	86	--	--	--	--	--	--	--	--	--	--	--	--	--
	87	--	--	--	--	--	--	--	--	+	--	--	--	--
	88	+	+	++	++	--	-	-	--	+	+	--	--	-
	89 ^a	++	++	++	++	++	++	++	++	++	++	++	++	++

Figure 12. $\text{H}_2\text{N}\gamma$ -Turn-linked conjugates: uptake profile of polyamides **78-89**. Symbols are defined in Figure 2. Except where noted, polyamide concentration was 2 μM . ^aPolyamide described previously (22), assayed at 5 μM . ^b Assayed at 5 μM .

Shapes

Beyond hairpin polyamides derivatized at the turn, a variety of unique polyamide shapes have been described, including *N*-methyl substituted hairpins (13-15,37,38), cycles (39,40), U-pins (26), H-pins (27,41), and tandem hairpins (30,33,42). Figure 13 shows the chemical structures of dye conjugates of various shapes, and Figure 14 presents their uptake profiles.

FITC conjugate **90**, linked through an *N*-propylamine Im-cap residue, is a poor nuclear stain, as described previously (22). After evaluating other compounds, and noting that effective compounds incorporated a cationic amine group, we synthesized compound **91** (an isomer of conjugate **1**), which demonstrated good to excellent uptake properties. Correspondingly, cyclic polyamide **92** displayed a very poor uptake profile. In contrast,

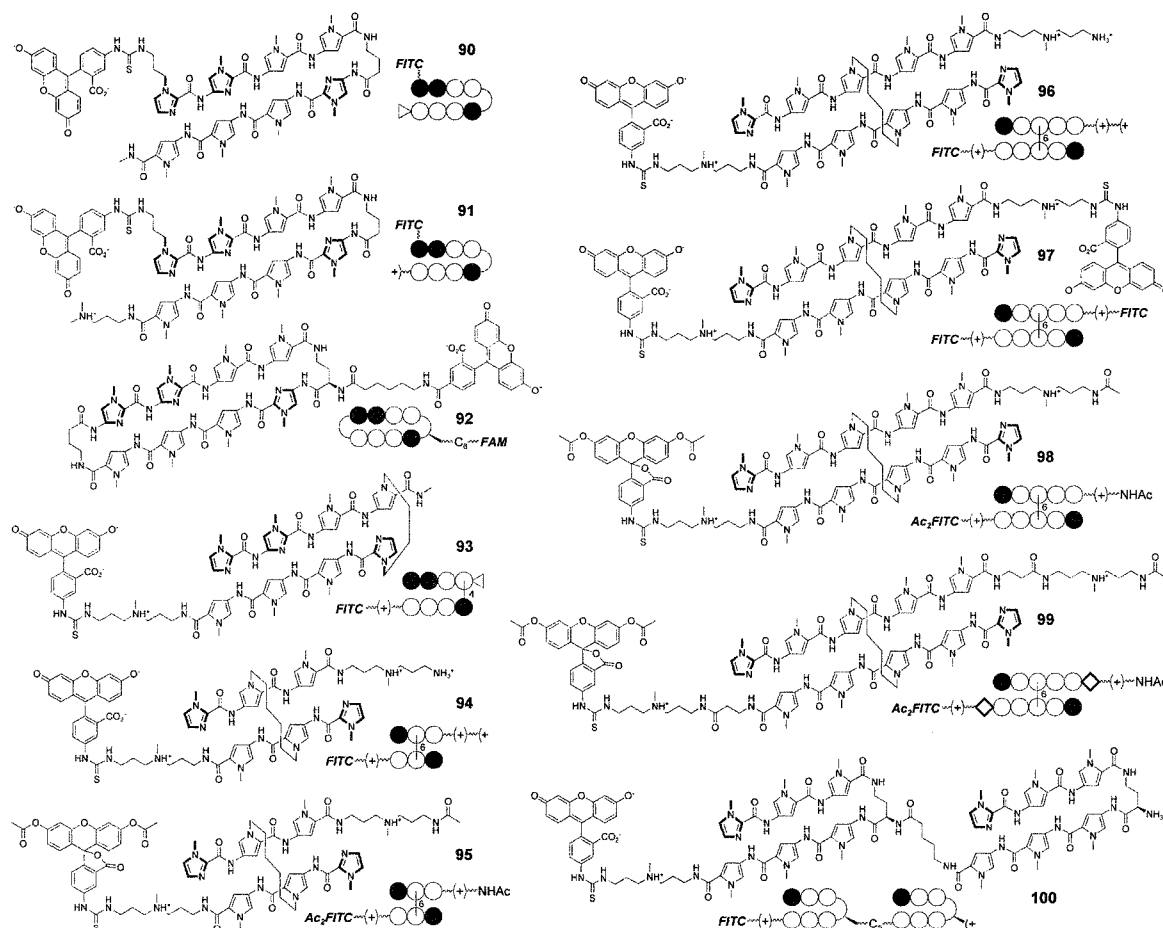


Figure 13. Shape variations: chemical and ball-and-stick structures for polyamides **90-100**.

U-pin polyamide **93** is a moderately good nuclear stain, especially in adherent cells.

Uptake data for H-pin polyamides indicates that a motif cannot be evaluated by a handful of molecules alone. Of the six-ring H-pins, **95** (the peracetylated derivative of **94**) showed highly heterogeneous staining in a small number of cell lines. Within a single microscope frame, some cells were brightly stained, while others were completely dark – where indicated, the data in Figure 14 represent a rough average across the cells. Ten-ring H-pins **96** and **97** were excluded from all cells tested, whereas **98** (the ten-ring analog of **95**) showed good to excellent staining in all but the least permissive cell lines. In line with trends observed earlier, adding C-terminal β -Ala residues (compound **99**) diminishes

molecular weight, may intensify or diminish nuclear staining (Figure 8). The acetylated ^{AcHN}γ-turn, which adds molecular weight compared to an unsubstituted γ-turn, is one of the few consistently positive factors for nuclear localization (for example, compare **1** and **16**, **48** and **49**, or **72** and **73**). Generally negative determinants for nuclear access include the presence of a β-Ala-tail residue and the lack of a cationic alkyl amine moiety.

Although there is no general correlation between the number of Im residues and uptake efficiency, the positions of Im residues clearly affect nuclear access. For example, structural isomers often display very different uptake profiles (compare eight-ring polyamides **1**, **65**, and **68**, or larger conjugates **58**, **62**, and **72**). Furthermore, a Py→Im exchange is nearly always a negative determinant for nuclear access (except at the cap position, compare **29** and **30**), and the impact is highly dependent on the position of the residue and the overall polyamide motif. For example, in eight-ring polyamide-dye conjugates the effect of a single Py→Im exchange may be very small (**89** versus **1**), modest (**63** versus **65**), or severe (**1** versus **70**).

The role of charge in nuclear uptake is particularly intriguing, though uptake does not appear to correlate with net charge. At neutral pH, fluorescein conjugates that efficiently access cell nuclei have a net charge of -1 (the majority of conjugates, exemplified by **1**) or a net charge of 0 (such as **29**, **31**, and **81**). However, polyamide-fluorescein conjugates have been observed to accumulate in acidic vesicles (21 and data not shown), in which the pH is between ~4.5 and 6 (43). Because the pK_a's of the phenolic protons of fluorescein and OG derivatives are ~6.4 and ~4.7, respectively (44), the phenol moiety will be fully or partially protonated in acidic vesicles. Thus, in such subcellular structures, polyamide-fluorescein conjugates such as **1** would have a net

charge of 0, whereas conjugates such as **29** would have a net charge of +1.

It may be that the ability to change protonation state enhances the nuclear localization properties of fluorescein conjugates relative to rhodamine and bodipy derivatives, which are pH insensitive. TMR and BoFL conjugates having a net charge of 0 (compounds **13** and **14**) or +1 (compounds **8-11**, **15** and **83-88**) all perform significantly more poorly than analogous fluorescein conjugates. It is remarkable that moving the positive charge from the linker to the dye is not tolerated whatsoever (compare triamine-linked FAM conjugate **2** and alkyl-linked TMR conjugate **13**, Figure 2), whereas relatively well-tolerated alterations include moving the positive charge to the turn (compare FITC conjugates **1** and **78**) and moving the dye to the cap residue, effectively increasing the distance between the anionic and cationic moieties (compare FITC conjugates **1** and **91**).

Conjugation to fluorescein appears to facilitate uptake for this class of molecules, though it is not essential: TMR conjugate **9** is an excellent nuclear stain in several cell lines, and turn-linked BoFL conjugates **83**, **84**, and **88** display improved uptake profiles compared to the tail-linked BoFL derivative **15**. Indeed, polyamides without any attached fluorophore have been shown to induce biological effects in living cells, albeit in a limited number of cell types. Whereas **62** was excluded from cells, an analog of **62** without a dye-label altered gene expression in live lymphoid cells, and a non-dye-labeled chorambucil derivative of **62** was shown to alkylate genomic DNA in live cells (20). This example highlights the distinct properties of dye-conjugated versus unlabeled polyamides. One cannot be considered as a “proxy” for the other, regarding cellular localization characteristics or other properties. Fluorescent polyamide conjugates are

unique because their ability to access nuclei can be assessed directly, and it is our intent to use such labeled polyamides in biological studies involving live cell systems.

Although the uptake data presented here do not allow for prediction of nuclear uptake properties *a priori*, rough design guidelines are apparent. Synthesizing and analyzing a small, focused library of polyamide-fluorophore conjugates appears to be the optimal approach. Key points of variation are ring sequence (if possible on a given target), dye composition (FITC, FAM, OG 514, etc.) and position of conjugation, turn substitution (γ -turn or $^{\text{AcHN}}\gamma$ -turn), and β -Ala incorporation. Examples of such focused libraries are compounds **55-59**, **72-77**, and **96-99**.

For use in live-cell studies, polyamides must now be optimized along three axes: DNA-binding affinity and specificity, *in vitro* biochemical activity, and nuclear localization. For the first and second axes, assays such as DNase footprinting, gel shift, and *in vitro* transcription are well established, and studies based on these techniques continue to expand the scope and utility of DNA-binding polyamides (3,12,15,45). The current study employs an assay based on confocal microscopy to establish a benchmark dataset, comprising 1300 entries, for nuclear localization of DNA-binding polyamide-dye conjugates. Live-cell studies with these compounds are currently in progress, as are efforts to elucidate the energy-dependent mechanism of uptake (22). Understanding nuclear accessibility in a wide variety of living cells is a minimum first step toward chemical regulation of gene expression with this class of molecules.

References

1. Dervan,P.B. and Edelson,B.S. (2003) *Curr. Opin. Struct. Biol.*, **13**, 284-299.
2. Briehn,C.A., Weyermann,P. and Dervan,P.B. (2003) *Chem.-Eur. J.*, **9**, 2110-2122.
3. Renneberg,D. and Dervan,P.B. (2003) *J. Am. Chem. Soc.*, **125**, 5707-5716.
4. Foister,S., Marques,M.A., Doss,R.M. and Dervan,P.B. (2003) *Bioorg. Med. Chem.*, **11**, 4333-4340.
5. Dickinson,L.A., Trauger,J.W., Baird,E.E., Ghazal,P., Dervan,P.B. and Gottesfeld,J.M. (1999) *Biochemistry*, **38**, 10801-10807.
6. Dickinson,L.A., Trauger,J.W., Baird,E.E., Dervan,P.B., Graves,B.J. and Gottesfeld,J.M. (1999) *J. Biol. Chem.*, **274**, 12765-12773.
7. McBryant,S.J., Baird,E.E., Trauger,J.W., Dervan,P.B. and Gottesfeld,J.M. (1999) *J. Mol. Biol.*, **286**, 973-981.
8. Chiang,S.Y., Burli,R.W., Benz,C.C., Gawron,L., Scott,G.K., Dervan,P.B. and Beerman,T.A. (2000) *J. Biol. Chem.*, **275**, 24246-24254.
9. Wang,C.C.C. and Dervan,P.B. (2001) *J. Am. Chem. Soc.*, **123**, 8657-8661.
10. Ehley,J.A., Melander,C., Herman,D., Baird,E.E., Ferguson,H.A., Goodrich,J.A., Dervan,P.B. and Gottesfeld,J.M. (2002) *Mol. Cell. Biol.*, **22**, 1723-1733.
11. Yang,F., Belitsky,J.M., Villanueva,R.A., Dervan,P.B. and Roth,M.J. (2003) *Biochemistry*, **42**, 6249-6258.
12. Fechter,E.J. and Dervan,P.B. (2003) *J. Am. Chem. Soc.*, **125**, 8476-8485.
13. Wurtz,N.R., Pomerantz,J.L., Baltimore,D. and Dervan,P.B. (2002) *Biochemistry*, **41**, 7604-7609.

14. Arora,P.S., Ansari,A.Z., Best,T.P., Ptashne,M. and Dervan,P.B. (2002) *J. Am. Chem. Soc.*, **124**, 13067-13071.
15. Arndt,H.D., Hauschild,K.E., Sullivan,D.P., Lake,K., Dervan,P.B. and Ansari,A.Z. (2003) *J. Am. Chem. Soc.*, **125**, 13322-13323.
16. Gottesfeld,J.M., Neely,L., Trauger,J.W., Baird,E.E. and Dervan,P.B. (1997) *Nature*, **387**, 202-205.
17. Janssen,S., Cuvier,O., Muller,M. and Laemmli,U.K. (2000) *Mol. Cell.*, **6**, 1013-1024.
18. Wang,Y.D., Dziegielewski,J., Wurtz,N.R., Dziegielewska,B., Dervan,P.B. and Beerman,T.A. (2003) *Nucleic Acids Res.*, **31**, 1208-1215.
19. Belitsky,J.M., Leslie,S.J., Arora,P.S., Beerman,T.A. and Dervan,P.B. (2002) *Bioorg. Med. Chem.*, **10**, 3313-3318.
20. Dudouet,B., Burnett,R., Dickinson,L.A., Wood,M.R., Melander,C., Belitsky,J.M., Edelson,B., Wurtz,N., Briehn,C., Dervan,P.B. *et al.* (2003) *Chem. Biol.*, **10**, 859-867.
21. Crowley,K.S., Phillion,D.P., Woodard,S.S., Schweitzer,B.A., Singh,M., Shabany,H., Burnette,B., Hippenmeyer,P., Heitmeier,M. and Bashkin,J.K. (2003) *Bioorg. Med. Chem. Lett.*, **13**, 1565-1570.
22. Best,T.P., Edelson,B.S., Nickols,N.G. and Dervan,P.B. (2003) *Proc. Natl. Acad. Sci. U. S. A.*, **100**, 12063-12068.
23. Baird,E.E. and Dervan,P.B. (1996) *J. Am. Chem. Soc.*, **118**, 6141-6146.
24. Belitsky,J.M., Nguyen,D.H., Wurtz,N.R. and Dervan,P.B. (2002) *Bioorg. Med. Chem.*, **10**, 2767-2774.
25. Wurtz,N.R. (2002) Pasadena, CA, USA, p 99.
26. Heckel,A. and Dervan,P.B. (2003) *Chem.-Eur. J.*, **9**, 3353-3366.

27. Olenyuk,B., Jitianu,C. and Dervan,P.B. (2003) *J. Am. Chem. Soc.*, **125**, 4741-4751.
28. Delmotte,C. and Delmas,A. (1999) *Bioorg. Med. Chem. Lett.*, **9**, 2989-2994.
29. Gygi,M.P., Ferguson,M.D., Mefford,H.C., Lund,K.P., O'Day,C., Zhou,P., Friedman,C., van den Engh,G., Stolowitz,M.L. and Trask,B.J. (2002) *Nucleic Acids Res.*, **30**, 2790-2799.
30. Maeshima,K., Janssen,S. and Laemmli,U.K. (2001) *Embo J.*, **20**, 3218-3228.
31. Trauger,J.W., Baird,E.E. and Dervan,P.B. (1996) *Chem. Biol.*, **3**, 369-377.
32. Trauger,J.W., Baird,E.E. and Dervan,P.B. (1998) *Angew. Chem.-Int. Edit.*, **37**, 1421-1423.
33. Schaal,T.D., Mallet,W.G., McMinn,D.L., Nguyen,N.V., Sopko,M.M., John,S. and Parekh,B.S. (2003) *Nucleic Acids Res.*, **31**, 1282-1291.
34. Swalley,S.E., Baird,E.E. and Dervan,P.B. (1997) *Chem.-Eur. J.*, **3**, 1600-1607.
35. Turner,J.M., Swalley,S.E., Baird,E.E. and Dervan,P.B. (1998) *J. Am. Chem. Soc.*, **120**, 6219-6226.
36. Gottesfeld,J.M., Belitsky,J.M., Melander,C., Dervan,P.B. and Luger,K. (2002) *J. Mol. Biol.*, **321**, 249-263.
37. Rucker,V.C., Foister,S., Melander,C. and Dervan,P.B. (2003) *J. Am. Chem. Soc.*, **125**, 1195-1202.
38. Bremer,R.E., Wurtz,N.R., Szewczyk,J.W. and Dervan,P.B. (2001) *Bioorg. Med. Chem.*, **9**, 2093-2103.
39. Herman,D.M., Turner,J.M., Baird,E.E. and Dervan,P.B. (1999) *J. Am. Chem. Soc.*, **121**, 1121-1129.

40. Baliga,R., Baird,E.E., Herman,D.M., Melander,C., Dervan,P.B. and Crothers,D.M.
(2001) *Biochemistry*, **40**, 3-8.
41. Greenberg,W.A., Baird,E.E. and Dervan,P.B. (1998) *Chem.-Eur. J.*, **4**, 796-805.
42. Kers,I. and Dervan,P.B. (2002) *Bioorg. Med. Chem.*, **10**, 3339-3349.
43. Lodish,H., Berk,A., Zipursky,L.S., Matsudaira,P., Baltimore,D. and Darnell,J. (2000)
Molecular Cell Biology. 4th ed. W. H. Freeman and Co., New York, NY.
44. Lin,H.J., Szmecinski,H. and Lakowicz,J.R. (1999) *Anal. Biochem.*, **269**, 162-167.
45. Marques,M.A., Doss,R.M., Urbach,A.R. and Dervan,P.B. (2002) *Helv. Chim. Acta*,
85, 4485-4517.