Chapter 2

Turn-to-Tail Dimerizations of Hairpin Polyamides on Duplex DNA Templates

(Poulin-Kerstien, A. T.; Dervan, P. B. "DNA-Templated Dimerizations of Hairpin Polyamides." *J. Am. Chem. Soc.* **2003**, *125*, 15811–15821)

Abstract

Double-helical DNA accelerates the rate of ligation of two six-ring hairpin polyamides, which bind adjacent sites in the minor groove via a 1,3-dipolar cycloaddition to form a tandem dimer. The rate of the templated reaction is dependent on DNA sequence as well as on the distance between the hairpin binding sites. The tandem-dimer product of the DNA-templated reaction has improved binding properties with respect to the smaller hairpin fragments. Since cell and nuclear uptake of DNA-binding polyamides will likely be dependent on size, this is a minimum first step toward the design of self-assembling small gene-regulating fragments to produce molecules of increasing complexity with more specific genomic targeting capabilities.

Introduction.

Small molecules that bind a large repertoire of DNA sequences and modulate transcription could be useful in biology and medicine.[1, 2] Polyamides comprised of three aromatic amino acids, N-methylpyrrole (Py), N-methylimidazole (Im), and N-methyl-3-hydroxypyrrole (Hp), distinguish the four Watson-Crick base pairs by a set of pairing rules.[3] Connecting the two antiparallel strands of aromatic amino acids with a γ-aminobutyric acid (γ) creates a hairpin motif capable of binding to match DNA with increased affinity and sequence specificity.[4-6] For applications in gene regulation within biological systems, binding-site size may be critical because longer sequences should occur less frequently in a gigabase-sized genome. For this reason, the design of ligands capable of targeting >10 base pairs (bp) of DNA remains an important goal in the area of polyamide design.[3, 7-9] An ideal DNA-binding polyamide used for gene regulation must possess several properties: high affinity to DNA, such that it can compete with cellular DNA-binding proteins; specificity, to distinguish its targeted match site from mismatch sites; and favorable cell and nuclear uptake properties in order to reach its targeted DNA on nuclear chromatin.

DNA-binding eight-ring hairpin polyamides possess excellent affinity and sequence specificity, but they target only 6 bp.[3, 7-9] Various polyamide motifs have been designed to target longer sequences. High resolution X ray studies reveal that polyamides containing more than five contiguous aromatic ring pairings are over curved with respect to the DNA helix, which results in a loss of the hydrogen bonds and van der Waals interactions responsible for binding affinity.[10, 11] Replacement of one or more internal Py carboxamides with more flexible β-alanine (β) residues relaxes the polyamide

curvature and allows longer hairpin polyamides to bind DNA with restored affinity.[12-14] These types of flexible motifs have been used to target up to 16 base pairs of DNA.[12-14] However, polyamides containing internal β-alanines are also able to bind in 1:1 ligand-DNA stoichiometries, thereby decreasing their specificity.[15, 16] Another approach to increase polyamide binding-site size has been to covalently link two hairpin modules to form hairpin dimers. Dimers linked both "turn-to-tail" and "turn-to-turn" have excellent affinity and specificity to DNA sequences up to 10 bp in length.[17-19] Though likely satisfying the DNA-binding criteria to target unique sequences within large genomic DNA, hairpin dimers do not possess the favorable cell and nuclear uptake properties of smaller hairpins, presumably due to size and shape.[20, 21]

Many DNA-binding transcription factors and complexes rely on dimerization or multimerization of DNA recognition elements that each occupy 4–6 base pairs and target unique contiguous sites in genomic DNA.[22, 23] This cooperative, combinatorial association of gene-regulatory proteins may be a useful strategy to overcome the kinetic problems associated with finding long contiguous sequences of DNA. As polyamides are designed to target longer sequences of DNA, a similar kinetic barrier may be encountered.

Single-strand DNA has long been known to template chemical reactions by bringing reactive functionalities in close proximity.[24-45] Annealing of two adjacent complementary single strands of DNA brings the reactive groups together. Remarkably, Liu and coworkers report that single-stranded DNA can template various chemical reactions in systems in which 1–30 nucleobases of single-stranded DNA separate the template site from the reactive site, with little observable change in reactivity.[26, 30-33] The

single-stranded template DNA serves to coordinate chemical reactivity and is not a desired part of the final product.[26, 30-33]

In this report we explore *double-stranded DNA-templated reactions* combining hairpin modules in the minor groove with the overall goal of producing larger polyamides capable of targeting longer sequences.[46] Two different six-ring hairpin polyamides were designed such that when their match sites are adjacent on the DNA a thermal reaction at 37 °C (pH 7.0) would afford a covalent bond between the hairpin modules, forming a tandem-dimer structure *in situ* (Figure 2.1). The tandem-hairpin dimer product should have improved DNA-binding properties over the smaller hairpin subunits. This type of scheme differs from hybridization-based DNA-templated chemistry in that we are using the double-helical structure of DNA to template covalent bond formation. In this method
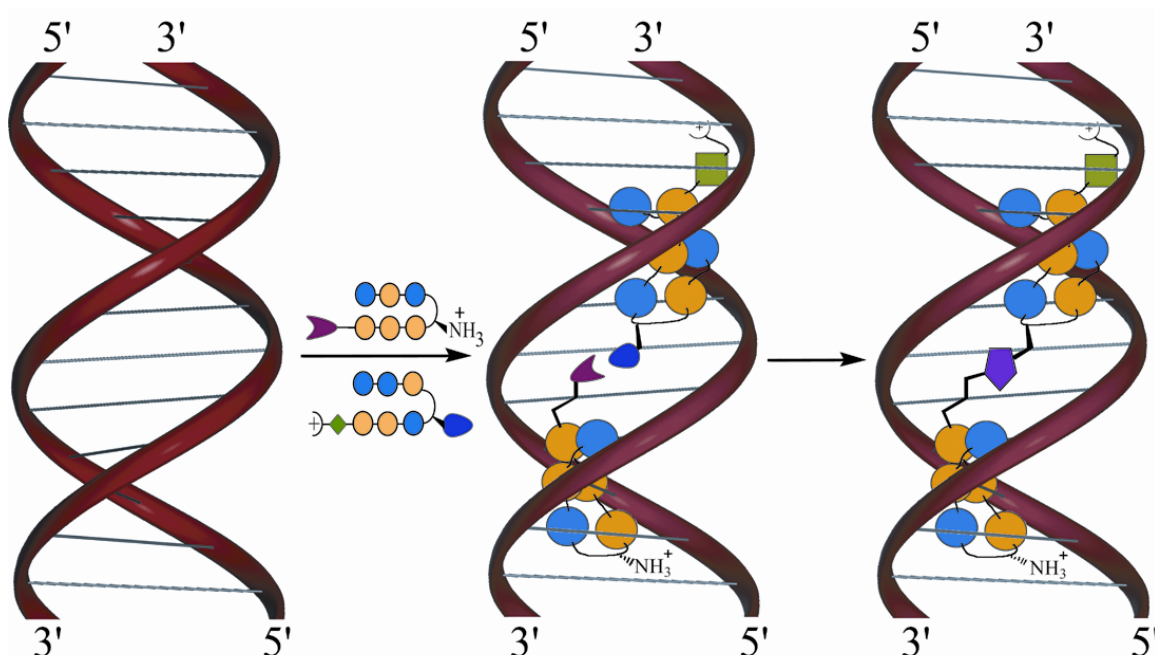

Figure 2.1. Schematic model of DNA-templated tandem hairpin formation. Polyamides functionalized with complementary reactive groups (red and blue shapes) bind to contiguous match sites on DNA. The reactive groups are placed in close proximity, causing them to form a covalent bond, linking the hairpins (purple pentagon).

pyrrole-imidazole polyamides read by "pairing rules" the unique hydrogen bonding pattern presented by the edges of the Watson-Crick base pairs in the minor groove of DNA. The DNA binding hairpin polyamide modules become part of a ligand-DNA supramolecular complex containing four oligomers: two DNA strands and two juxtaposed polyamides. *The sequence information encoded in the DNA base pairs becomes encoded in the polyamide product molecule.*[46]

Results.

Chemical Ligation Reaction. The reaction chosen to ligate the DNA-binding hairpins together is the 1,3-dipolar cycloaddition reaction of an acetylene and azide.[47-52] The active site of acetylcholine esterase has been shown to template the formation of a high-affinity inhibitor from azide- and alkyne-functionalized building blocks utilizing the Hüisgen 1,3-cycloaddition.[49] The thermal cycloaddition reaction requires no cofactors and tolerates reaction conditions similar to those inside a cell.[47-52] The simultaneous binding of two ligands on adjacent addressable sites should accelerate the reaction that connects them. To create unique points of reactivity in the minor groove, an acetylene on the γ-turn of one hairpin would be placed proximal to the azide on the C-terminus of an adjacent hairpin in tandem orientations affording a 1,2,3-triazole in the linker of a tandem hairpin dimer product (Fig. 2.2).
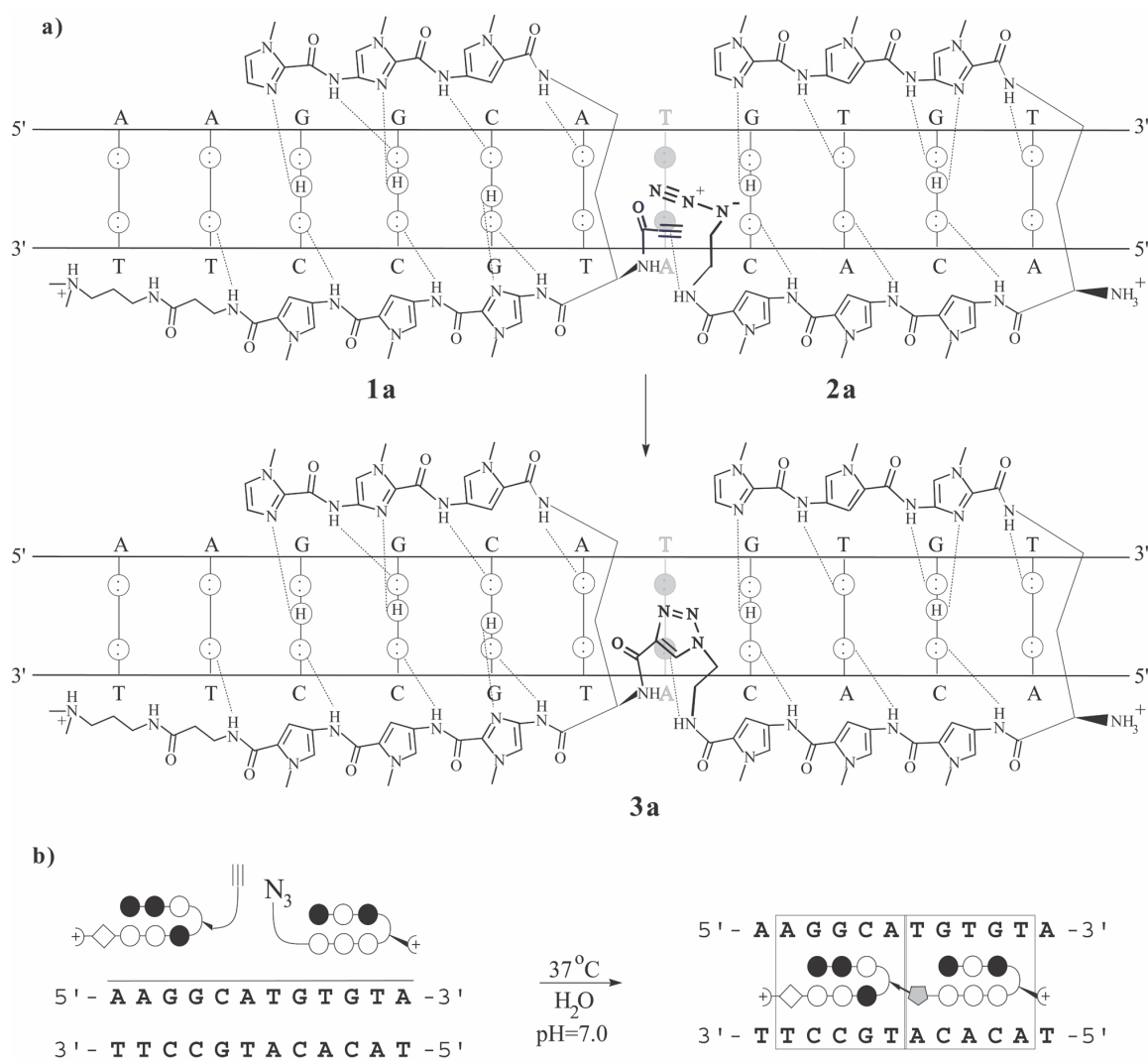
Figure 2.2. a) Hydrogen bonding model of the hairpin-DNA complex, **1a**, **2a**, at the 10-bp match (zero intervening bp) site 5'-AGGCATGTGT-3' (top), and reaction product tandem **3a** complexed with DNA (bottom). Circles with two dots represent the lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dashed lines. b) Schematic model of alkyne **1a** and azide **2a** binding to their match sites and forming **3a**, which recognizes the entire 10-bp binding site. Im and Py residues are represented by filled and open circles, respectively. The β-residue is represented by a diamond. The triazole linker is represented as a pentagon. Individual binding sites for the starting hairpin polyamides (**1a** and **2a**) are boxed.

Hairpin polyamide design. For the design of the hairpin polyamide reactants the flexibility and distance of the linker that connects the reactive alkyne and azide functionalities to the DNA-binding polyamides were chosen based on model building. With regard to reaction times at 37 °C, we anticipated that the inherent reactivity of the

starting materials could be tuned to afford templated intramolecular reaction half-lives on the order of hours, rather than seconds or weeks. For example, alkynes possessing an electron-withdrawing group, such as an amide or carboxylic ester, are more reactive in dipolar cycloadditions than their alkyl counterparts (up to 10^5 increase in reaction rates).[47, 51, 52] Alkyne-functionalized polyamides ImImPy-(R)-γ^{NH}[COC≡CH]-ImPyPy-β-Dp (**1a**), ImImPy-(R)-γ^{NH}[CO(CH$_2$)$_2$C≡CH]-ImPyPy-β-Dp (**1b**) were synthesized (Dp = 3-(Dimethylamino)propylamine, γ = (R)-2,4-diaminobutyric acid) with two different linker lengths; **1a** possessing the rigid alkynyl amidate functionality and **1b** possessing the more flexible alkyl alkyne. Our criteria for screening linkers for the DNA-templated reaction was to select reactive partners that do not react in solution at μM concentration (37 °C) but react in a reasonable time (hours) on DNA. Azide-functionalized polyamides ImPyIm-(R)-γ^{NH_2}-PyPyPy-(CH$_2$)$_2$-N$_3$ (**2a**) and ImPyIm-(R)-γ^{NH_2}-PyPyPy-(CH$_2$)$_3$-N$_3$ (**2b**) were synthesized which contain either an ethyl (for **2a**) or propyl (for **2b**) linker connecting the azide moiety to the C-terminus. These molecules were chosen because the 1,2,3-triazole linker (Tr) formed in the pairing of either **2a–b** with **1a** best approximated the length of the 5-aminovaleric acid linker found in our previous studies to be optimal for "turn-to-tail" tandem affinity and specificity to a 10 bp binding site.[18]

Precursor **4** was synthesized on resin using Boc chemistry and was liberated by aminolysis with neat Dp (Figure 2.3).[53] Six-ring hairpins **1a** and **1b** were completed by the installation of the alkynyl functionality on the α-amine of the γ-turn residue using either propiolic acid (for **1a**) or 1-pentynoic acid (for **1b**) and DCC coupling conditions according to modified literature procedures.[54] Azide-functionalized polyamide

precursors **6a–b** were synthesized on Kaiser oxime resin[55] and were liberated from solid

support by aminolysis with either 2- or 3-aminoethanol (**6a** and **6b**, respectively).
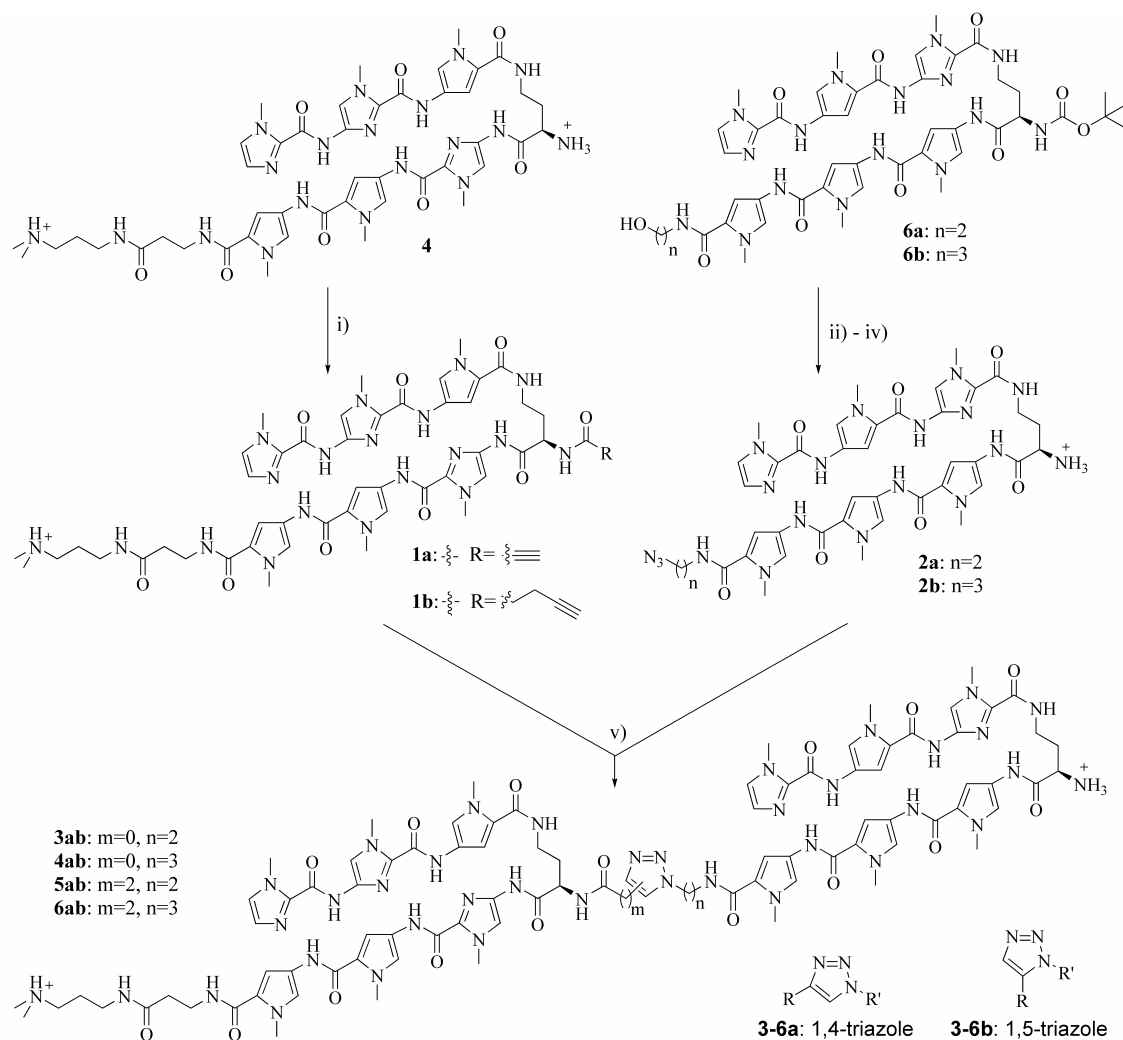

Figure 2.3. Synthesis of hairpins **1a–b** and **2a–b** and tandem hairpin dimers **3ab–6ab**: i) propiolic acid (for **1a**) or 5-pentynoic acid (for **1b**), DCC, DMF/CH₃CN 1:1, 0 ᵒC, 6 h; ii) DIEA, CH₂Cl₂, 0 ᵒC, 15 min, then MsCl, rt, 2 h; iii) NaN₃, DMF, 70 ᵒC, 12 h; (iv) CH₂Cl₂-TFA 1:1, 15 min; (v) neat, 60 ᵒC, 5–14 days.

Mesylation of the resulting alcohols, followed by displacement with sodium azide

resulted in the appropriately functionalized C-termini.[56] Compounds **2a** and **2b** were

obtained following Boc-deprotection of the α–amino group of the γ-turn residue with

50% TFA-CH₂Cl₂.

Authentic samples of the expected tandem products ImImPy-(R)-[ImPyIm-(R)$^{H_2N}\gamma$–PyPyPy-(CH$_2$)$_2$-Tr-(OC)]$^{HN}\gamma$-ImPyPy-β-Dp (**3ab**), ImImPy-(R)-[ImPyIm-(R)$^{H_2N}\gamma$–PyPyPy-(CH$_2$)$_3$-Tr-(OC)]$^{HN}\gamma$-ImPyPy-β-Dp (**4ab**), ImImPy-(R)-[ImPyIm-(R)$^{H_2N}\gamma$–PyPyPy-(CH$_2$)$_2$-Tr-(CH$_2$)$_2$(OC)]$^{HN}\gamma$-ImPyPy-β-Dp (**5ab**), and ImImPy-(R)-[ImPyIm-(R)$^{H_2N}\gamma$–PyPyPy-(CH$_2$)$_3$-Tr-(CH$_2$)$_2$(OC)]$^{HN}\gamma$-ImPyPy-β-Dp (**6ab**) were synthesized by heating dry, powdered mixtures of **1a-b** and **2a-b** at 60 °C for 5 days.

1,3-dipolar cycloadditions between azides and alkynes are known to produce both 1,4- and 1,5-substituted triazole ring products. In the case of alkyl substituted reactants, the stereoelectronics of the reaction pathways leading to each regioisomer are approximately equal, leading to equal ratios of the two regioisomers. When the alkynyl reactant is substituted with an electron withdrawing group, the pathway leading to the 1,4-regioisomer becomes favored, producing this isomer as the major product.[47, 51, 52] The thermal reaction between activated alkyne **1a** and azide **2a** favors the 1,4-regioisomeric product **3a** over the 1,5-regioisomeric product **3b** by a ratio of 20:1.[57] Regioisomers of substituted triazole rings are known to have distinct chemical shifts for the lone aromatic proton, and this distinction is the basis for our assignment of regioisomers.[58] Similarly, the reaction between the activated alkyne **1a** and the longer azide **2b** to form **4ab** gives a product ratio of 20:1, presumably **4a**:**4b**. When the unactivated alkyl alkyne **1b** is paired with either azide **2a** or **2b** to form **5ab** and **6ab**, respectively, each thermal coupling yields the two expected regioisomers in a ratio of 1:1. We anticipate the DNA-templated cycloaddition in the minor groove might produce different regioisomeric ratios due to steric constraints within the polyamide/DNA complex.

DNA-templated tandem formation.

Reactivity and hairpin binding site separation preference (Table 2.1). A key issue in evaluating the DNA-templated reaction was to determine if the cycloaddition reaction was accelerated in the presence of match DNA and in reasonable yields. In addition, we anticipated that reactions would be sensitive to the distance separating the hairpin-binding modules.

Table 2.1. Relative pseudozero-order rate constants (s^{-1}) for DNA-templated tandem formation when the reaction is performed with the hairpin binding sites separated by zero, one, or two intervening base pairs.[a-c]

	5'-AGGCATGTGT-3' A	5'-AGGCAATGTGT-3' B	5'-AGGCAAATGTGT-3' C	No DNA
1a + 2a	16290 (±54)	≤ 1	≤ 1	≤ 1
1a + 2b	13040 (±42)	≤ 1	≤ 1	≤ 1
1b + 2a	200 (±6)	≤ 1	≤ 1	≤ 1
1b + 2b	132 (±1)	21 (±1)	≤ 1	≤ 1

[a] The reported rates are normalized with respect to the nontemplated reaction between **1a** and **2a** at 1 μM, and are the average values obtained from three kinetics experiments, with the error for each data set indicated in parentheses.[b] The assays were carried out at 37 °C at pH 7.0 in the presence of 3 mM Tris-HCl, 3 mM KCl, 3 mM MgCl$_2$, and 1.7 mM CaCl$_2$, 1 μM each polyamide, 1μM DNA.[c] Rate data were taken from the linear phase of product formation (four time points over 5.25 hours), except for experiments with **1b**, which were taken from three time points over two weeks.

Tandem-hairpin polyamides have previously been shown to bind to both 10- and 11-bp sites, with zero and one base pairs separating the hairpin binding sites, respectively. To assess the appropriate polyamide separation distance for the DNA-templated tandem-forming cycloaddition, the duplex DNA templates 5'-GGGGT**AGGCATGTGT**AGGGG-3' (**A**), 5'-GGGGT**AGGCAA**T**GTGT**AGGGG-3' (**B**) and 5'-GGGGT**AGGCA**AA**TGTGT**AGGGG-3' (**C**) were synthesized (Figure 2.4a).

Each duplex contains five bp match sites for the two hairpin polyamides **1(a–b)** and **2(a–b)** separated by zero, one, or two base pairs, respectively.

Reactions were performed with equal concentrations of each hairpin polyamide and DNA (2 mM Tris-HCl, 2 mM KCl, 2 mM $MgCl_2$, 1 mM $CaCl_2$, pH 7.0, 37 °C). Analytical reverse-phase HPLC was used to monitor the cycloaddition reactions. MALDI-TOF MS was used to verify HPLC assignments. Experiments were carried out at 1 μM concentrations of DNA and hairpin polyamides.

When any pair of hairpin polyamides (**1a** + **2a**, **1a** + **2b**, **1b** + **2a**, **1b** + **2b**) is combined in solution at 1 μM concentration of each polyamide in the absence of DNA, no tandem product is observed after two weeks at 37 °C (pH = 7.0). The detection limit is approximately 0.1% of the total starting material. Thus, all tandem-forming cycloadditions proceed in less than 0.1% yield after 2 weeks in the absence of DNA.

When alkyne **1a** and ethyl azide **2a** are incubated with duplex **A** (zero intervening bases) at 1 μM concentrations, tandem product **3ab** is formed in detectable amounts after 45 minutes (37 °C, pH = 7.0). Quantitation of HPLC traces shows that **3ab** is formed in approximately 35% yield in 5 hours. We observe that the reaction does not progress significantly after this, exhibiting 43% product formation after 24 hours. When **1a** and **2a** are incubated with either of the longer duplexes **B** or **C** (1 and 2 intervening base pairs, respectively), tandem product **3ab** is not detected after 2 weeks. When **1a** and propyl azide **2b** are incubated with duplex **A** (zero intervening bases) at 1 μM, tandem product **4ab** is formed 20% more slowly than **3ab** from **1a** and **2a** on the same duplex. **1a** and **2b** produce no observable product formation on either duplex **B** or **C**. When less reactive alkyne **1b** and ethyl azide **2a** are incubated with duplex **A** (zero intervening

bases), **5ab** is formed in 20% yield after 2 weeks. When either one or two intervening

bases is present (duplex **B** and **C**, respectively), **5ab** is not detected. Product **6ab** is

formed on duplex **A** (zero intervening bases) from the long, flexible alkyne **1b** and propyl

azide **2b** in 13% yield in 2 weeks. These long linkers are able to span an additional bp,

with product **6ab** forming from this pairing on duplex **B** in 6% yield in 2 weeks.

Template mismatch tolerance: 10bp binding site (Table 2.2)**.** To assess if the

cycloaddition is sequence-specific with respect to the template, the duplexes

Table 2.2. Relative pseudozero-order rate constants (s^{-1}) for DNA-templated tandem formation of **3ab** from **1a** and **2a**. This data was collected on the duplex template (zero intervening bp) to probe tandem formation when there are mismatches under one or both hairpin sites.[a-d]

Concentration	5'-AGGCATGTCT-3' **A**	5'-AGGCATGT**C**T-3' **D**	5'-AGG**G**ATGT**C**T-3' **E**	No DNA
1.0 μM	16290 (±51) [42.7% (±0.7)]	10441 (±210) [28.2% (±0.2)]	1540 (±30) [16.3% (±0.2)]	1 [≤0.1%]
750 nM	17130 (±92) [47.3% (±0.7)]	11033 (±141) [30.4% (±0.4)]	793 (±3) [7% (±1)]	1 [≤0.1%]
500 nM	15420 (±77) [41% (±2)]	8927 (±327) [26% (±1)]	1050 (±22) [7.7% (±0.1)]	1 [≤0.1%]

[a] The reported rate constants are normalized with respect to that of **1a** and **2a** with no DNA template and are the average values obtained from three kinetics experiments, with the error in each data set indicated in parentheses. [b] The assays were carried out at 37 °C at pH 7.0 in the presence of 3 mM Tris-HCl, 3 mM KCl, 3 mM $MgCl_2$, and 1.7 mM $CaCl_2$, and the listed concentration of each polyamide and DNA. [c] Rate data was taken from the linear phase of product formation (4 time points over 5.25 hours). [d] In brackets is the amount of tandem **3ab** formed after 24-hour incubation (expressed as a percentage with 100% being complete conversion to product).

corresponding   to   5'-GGGGT**AGGCATGTCT**AGGGG-3'   (**D**)   and   5'-

GGGGT**AGGGATGTCT**AGGGG-3' (**E**) were synthesized (Figure 2.4a).   These

duplexes contain a single bp mismatch under one hairpin binding site (**D**) and a single bp

mismatch under each of the two hairpin polyamide binding sites (**E**) for the designed site

with zero intervening bp. This site was chosen because hairpin site-separation preference results indicated the 10 bp site to be optimal for tandem formation.

Because **1a** and **2a** showed the best DNA-templated reaction rate and yield, this pair was chosen for mismatch tolerance studies. Tandem product **3ab** is formed on double-stranded DNA template **D** 0.65 times as fast as on duplex **A** (zero intervening base pairs, match site). Product formation goes to only 28% completion on duplex **D** after 24 hours. When an additional mismatch is introduced (duplex **E**), **1a** and **2a** react to form **3ab** 0.029 times as fast as at the match site (duplex **A**). Product formation reaches only 16% after 24 hours (Figure 2.4b).

Reaction order and product verification. While non-templated couplings of **1b** with either **2a** or **2b** yield **5ab** and **6ab**, respectively, in equal regioisomeric ratios, when the reaction between **1b** and **2a** is templated on duplex **A** (zero intervening bp), a single regioisomer is formed. Likewise, when duplex **B** (one intervening bp) is used to template the formation of **6ab** from **1b** and **2b**, the reaction produces only a single regioisomer. When the reaction between **1b** and **2b** is templated by duplex **A** (zero intervening bp), two regioisomers are produced in a ratio of 3:1. When activated alkyne **1a** is paired with either **2a** or **2b** on template **A**, a single product isomer is produced.

When **1a** and **2a** are assayed on duplex DNA templates **A**, **D**, and **E** at either 750 nM or 500 nM concentrations, the relative rates of tandem formation are similar to those observed at 1 μM concentrations (Figure 2.4c). After 24 hours, the tandem-forming reaction is 47% complete at 750 nM and 41% complete at 500 nM. The reaction mixture of **1a** and **2a** at 1 μM concentration on DNA template **A** was taken at 8.25 hours and an
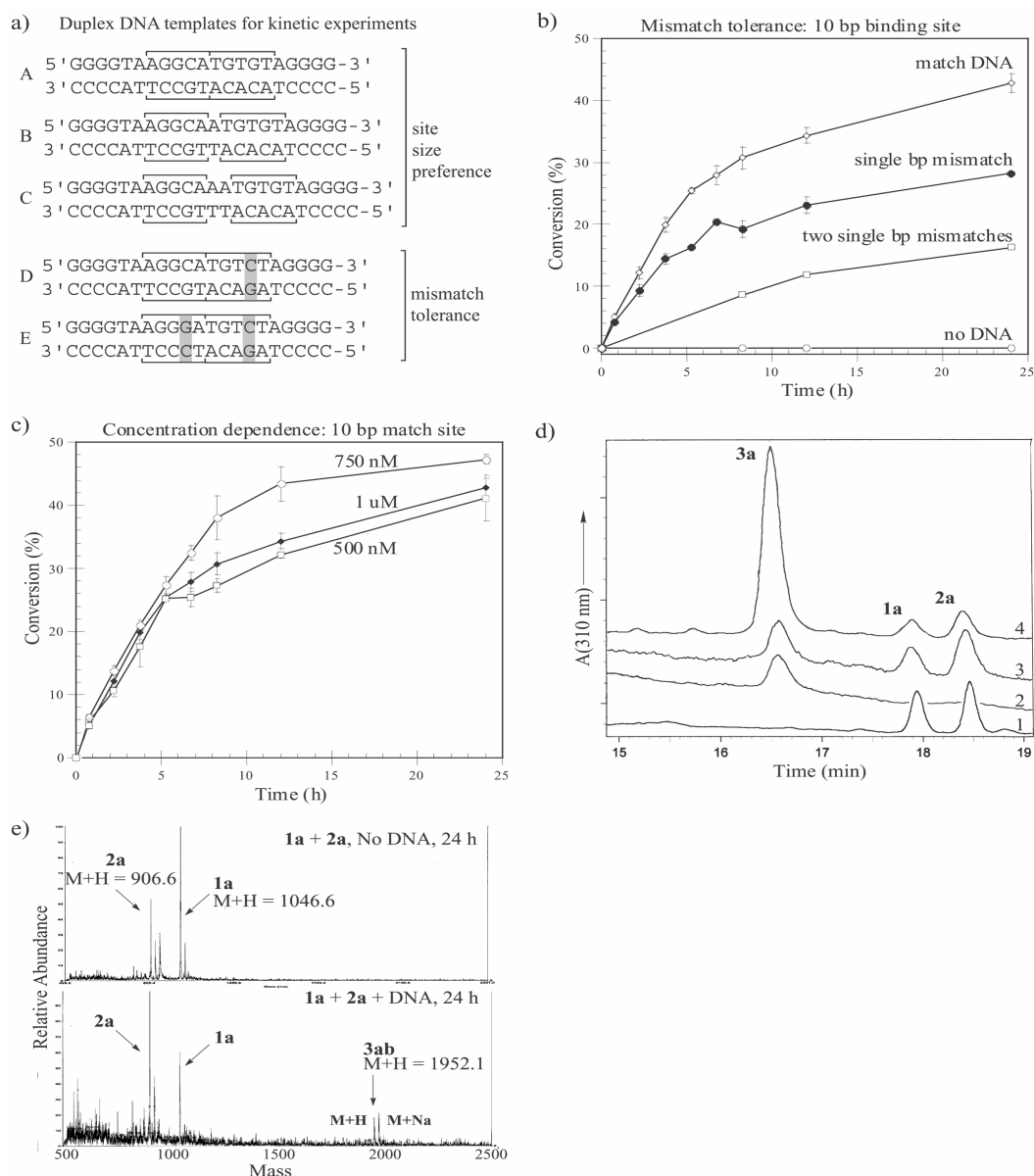
a) Duplex DNA templates for kinetic experiments

A 5'GGGGTAAGGCATGTGTAGGGG-3'
  3'CCCCATTCCGTACACATCCCC-5'

B 5'GGGGTAAGGCAATGTGTAGGGG-3'
  3'CCCCATTCCGTTACACATCCCC-5'

C 5'GGGGTAAGGCAAATGTGTAGGGG-3'
  3'CCCCATTCCGTTTACACATCCCC-5'

} site size preference

D 5'GGGGTAAGGCATGTCTAGGGG-3'
  3'CCCCATTCCGTACAGATCCCC-5'

E 5'GGGGTAAGGGATGTCTAGGGG-3'
  3'CCCCATTCCCTACAGATCCCC-5'

} mismatch tolerance

b) Mismatch tolerance: 10 bp binding site

match DNA
single bp mismatch
two single bp mismatches
no DNA

Conversion (%) vs Time (h)

c) Concentration dependence: 10 bp match site

750 nM
1 uM
500 nM

Conversion (%) vs Time (h)

d)

A(310 nm) vs Time (min)

3a, 1a, 2a

1, 2, 3, 4

e)

1a + 2a, No DNA, 24 h

2a M+H = 906.6
1a M+H = 1046.6

1a + 2a + DNA, 24 h

2a
1a
3ab M+H = 1952.1
M+H | M+Na

Relative Abundance vs Mass

Figure 2.4. a) Sequences of the short DNA duplexes used in examining the kinetics of DNA-templated tandem polyamide formation. Duplexes **A–C**: site size preference. Duplexes **D–E**: mismatch tolerance at the optimal 10 bp template site length. b) Mismatch tolerance rate data for the formation of tandem **3ab** from **1a** and **2a** at 1 μM concentrations at the 10 bp template site. Open diamond, duplex **A** 5'-AGGCATGTGT-3'; closed circle, duplex **D** 5'-AGGCATGTCT-3'; open square, duplex **E** 5'-AGGGATGTCT-3'; open circle, no DNA template. Each data point is an average of three kinetics experiments. c) Rate data for formation of tandem **3ab** from **1a** and **2a** on duplex **A** 5'-AGGCATGTGT-3'. Closed diamond, 1 μM; open circle, 750 nM; open square, 500 nM. Each data point is an average of three kinetics experiments. d) HPLC product verification from a single kinetics experiment between **1a** and **2a** on duplex **A** 5'-AGGCATGTGT-3' at 1 μM: Trace 1) **1a** + **2a** on duplex **A**, 0 hours; Trace 2) 1,4-regioisomeric **3a** with duplex **A**; Trace 3) **1a** + **2a** on duplex **A**, 8 hours, 37 °C; Trace 4) **1a** + **2a** on duplex **A**, 8 hours, + authentic sample of 1,4-regioisomeric **3a**. e) MALDI-TOF mass spectrometry product verification: No DNA) Mass spectra of **1a** + **2a** at 1 μM with no DNA, 24 h, 37 °C; + DNA) Mass spectra of **1a** + **2a** at 1 μM with duplex **A** 5'-AGGCATGTGT-3', 24 h, 37°C, arrow highlights mass corresponding to **3ab**.

authentic sample of **3a** (verified as the 1,4-regioisomer) was added. When analyzed by HPLC, no new peaks were observed while the putative product peak grew in absolute magnitude (Figure 2.4d). Purification of reaction mixtures by ZipTip C_{18}-charged pipette tips and subsequent analysis of the 50% CH_3CN eluent shows a peak corresponding to the mass of **3ab** only when the DNA template **A** is present (Figure 2.4e).

Quantitative DNase I footprinting (Tables 2.3 and 2.4). Once it was established that **1a** and **2a** were optimal reaction partners for the templated cycloaddition, the equilibrium association constants and sequence specificity of these hairpin polyamides and the 1,4-regioisomeric tandem product **3a** were analyzed by quantitative DNase I footprinting.[59] The polyamides were characterized on a DNA fragment from pATK1 which contains the sites 5'-AGGCATGTGT-3', 5'-AGGCAATGTGT-3', and 5'-AGGCAAATGTGT-3'. This allowed the comparison of binding affinities for DNA sequences including zero, one, or two intervening base pairs between the hairpin polyamide binding sites. In order to assess the sequence specificity of mismatches on overall tandem binding to the entire 10-bp site, 5'-AGGCATGTGT-3', the polyamides were assayed against a restriction fragment of DNA from pATK2 containing the match site, two single-bp mismatch sites targeted to each half-site (5'-AGGGATGTGT-3' and 5'-AGGCATGTCT-3') and the double-bp mismatch site 5'-AGGGTAGTCT-3' (Figure 2.5).

Six-ring hairpin module **1a** binds to its designed match sites 5'-AGGCA-3' on the restriction fragments from both pATK1 and pATK2 with modest affinity ($K_a = 3.5 \times 10^8$ M^{-1}). On the restriction fragment from pATK2, which contains the mismatch site 5-AGGGA-3', polyamide **1a** shows sequence specificity, favoring the match site over the

mismatch site by >340-fold (Figure 2.6). The affinity compares favorably with that found previously for the parent polyamide ImImPy-γ-ImPyPy-β-Dp ($K_a = 6.2 \times 10^7$ M^{-1}) footprinted at the site 5'-TGGCT-3' on pSES-TL1.[60]
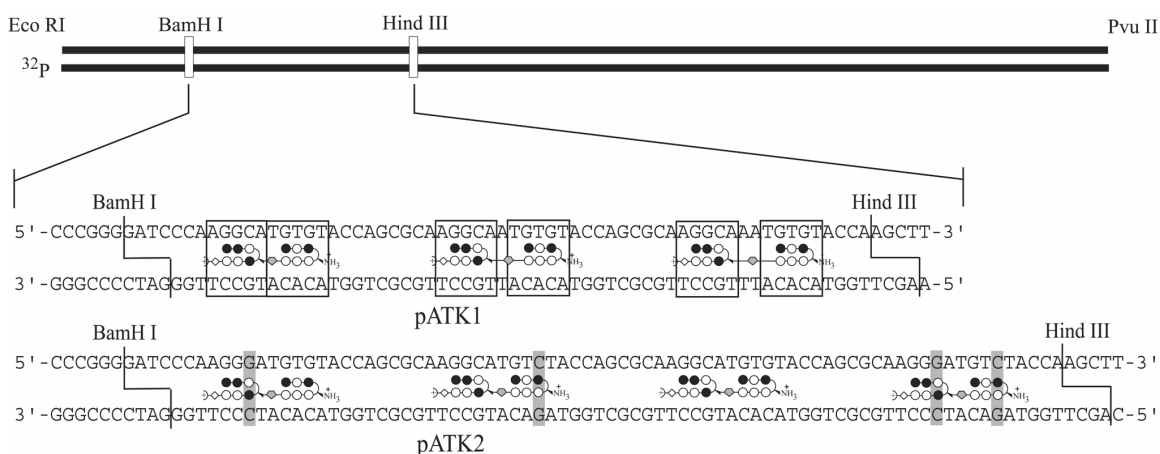

Figure 2.5. Sequences of the *Bam*HI/*Hind*III inserts from the *Eco*RI/*Pvu*II restriction fragments from pATK1 and pATK2 representing the designed sites for quantitative DNase footprinting assays. (See Supplemental Fig. 1 for footprinting gels.) pATK1- sites for the 10-bp (zero intervening bp), 11-bp (one intervening bp), and 12-bp (two intervening bp) match sites. Designed hairpin binding sites are boxed. pATK2- mismatch tolerance at the 10-bp (zero intervening bp) site. Sites for the two single bp mismatches, the match, and the double bp mismatch. Designed mismatch bases are boxed.

Six-ring hairpin module **2a** binds the designed match sites 5'-TGTGT-3' with $K_a = 9.2 \times 10^7$ M^{-1}. On the restriction fragment from pATK2, which contains the mismatch site 5'-TGTCT-3', polyamide **2a** favors its match site by >10-fold (Figure 2.7). Previous studies showed that the parent polyamide, with an additional positive charge, ImPyIm-(R)^{H2N}γ-PyPyPy-β-Dp binds match site 5'-AGTGA-3' on pIK2 with $K_a = 1.2 \times 10^9$ M^{-1}.

The 1,4-triazole regioisomer of the tandem-hairpin **3a** was characterized on the restriction fragment from pATK1 containing the zero intervening bp 5'-AGGCATGTGT-3', one intervening bp 5'-AGGCAATGTGT-3', and two intervening bp 5'-AGGCAAATGTGT-3' match binding sites. Polyamide **3a** binds to both the 10 bp ($K_a = 6.2 \times 10^9$ M^{-1}) and 11

Figure 2.6 Left: Quantitative DNase I footprinting experiments with hairpin **1a** (shown at top) on the 3'-^{32}P-labeled DNA fragment derived from plasmid pATK1. From left to right: lane 1, A-specific reaction; lane 2, G-specific reaction; lanes 3–13 1 μM, 300 nM, 100 nM, 30 nM, 10 nM, 3 nM, 1 nM, 300 pM, 100 pM, 30 pM, 10 pM **1a**; lane 14, DNase I standard; lane 15, intact DNA. Right: Quantitative DNase I footprinting experiments with hairpin **1a** on the 3'-^{32}P-labeled DNA fragment derived from plasmid pATK2. From left to right: lane 1, A-specific reaction; lane 2, G-specific reaction; lanes 3–13 100 nM, 30 nM, 10 nM, 3 nM, 1 nM, 300 pM, 100 pM, 30 pM, 10 pM, 3 pM **1a**; lane 13, DNase I standard; lane 14, intact DNA. All reactions contained 15 Kcpm labeled DNA and were carried out at 22 °C at pH 7.0 in the presence of 10 mM Tris-HCl, 10 mM KCl, 10 mM MgCl2, and 5 mM CaCl2 with an equilibration time of 36 hours. Designed binding sites where equilibrium association constants were obtained are shown to the right side of the gel. Equilibrium association constants are listed next to each binding site.
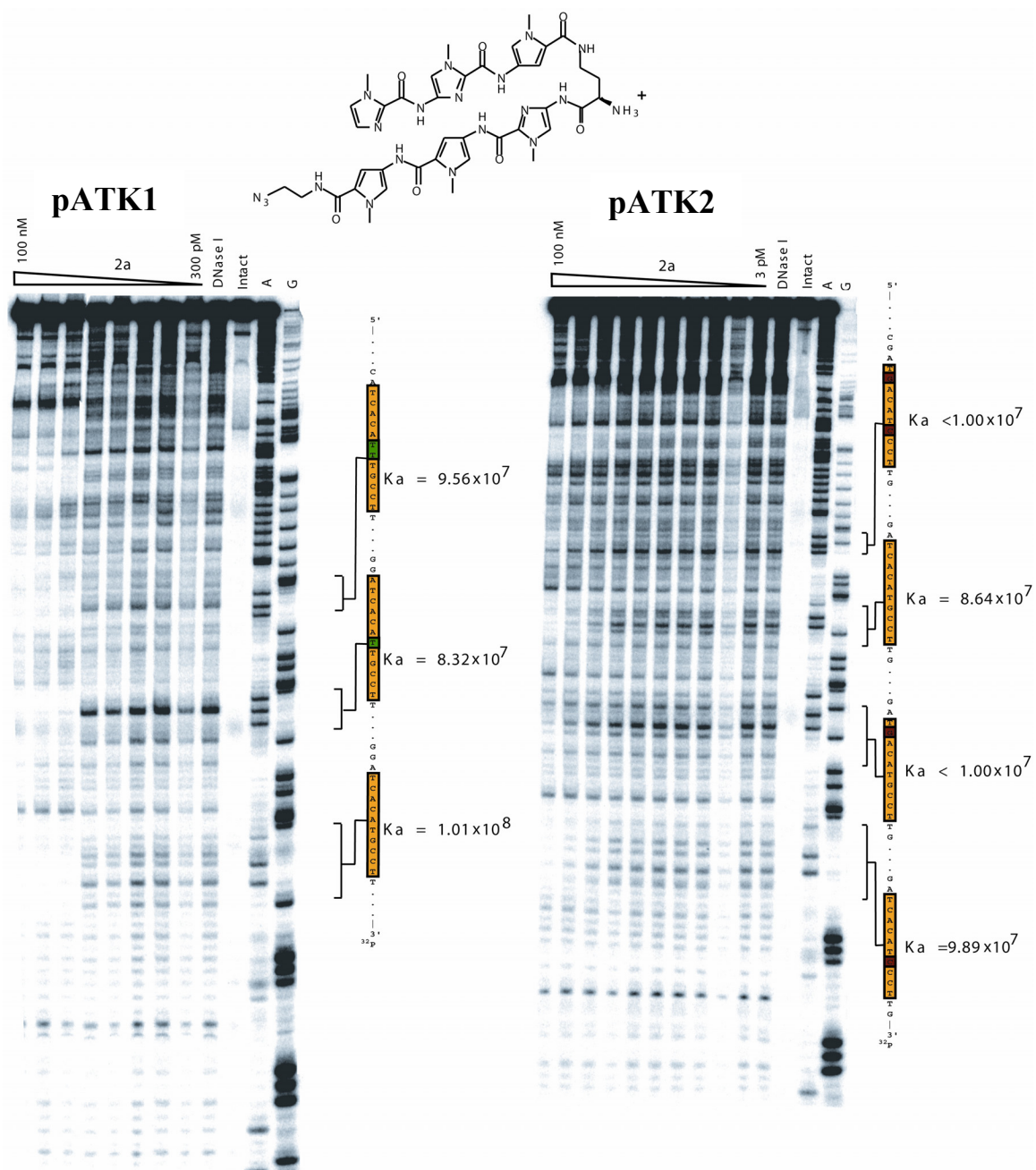
Figure 2.7 Left: Quantitative DNase I footprinting experiments with hairpin **2a** (shown at top) on the 3'-^{32}P-labeled DNA fragment derived from plasmid pATK1. From left to right: lanes 1–8 100 nM, 30 nM, 10 nM, 3 nM, 1 nM, 300 pM **2a**; lane 9, DNase I standard; lane 10, intact DNA; lane 11, A-specific reaction; lane 12, G-specific reaction. Right: Quantitative DNase I footprinting experiments with hairpin **2a** on the 3'-^{32}P-labeled DNA fragment derived from plasmid pATK2. From left to right: lanes 1–10 100 nM, 30 nM, 10 nM, 3 nM, 1 nM, 300 pM, 100 pM, 30 pM **2a**; lane 11, DNase I standard; lane 12, intact DNA; lane 13, A-specific reaction; lane 14, G-specific reaction. All reactions contained 15 Kcpm labeled DNA and were carried out at 22 °C at pH 7.0 in the presence of 10 mM Tris-HCl, 10 mM KCl, 10 mM MgCl2, and 5 mM CaCl2 with an equilibration time of 36 h. Designed binding sites are shown to the right side of the gel. Equilibrium association constants are listed next to each binding site.
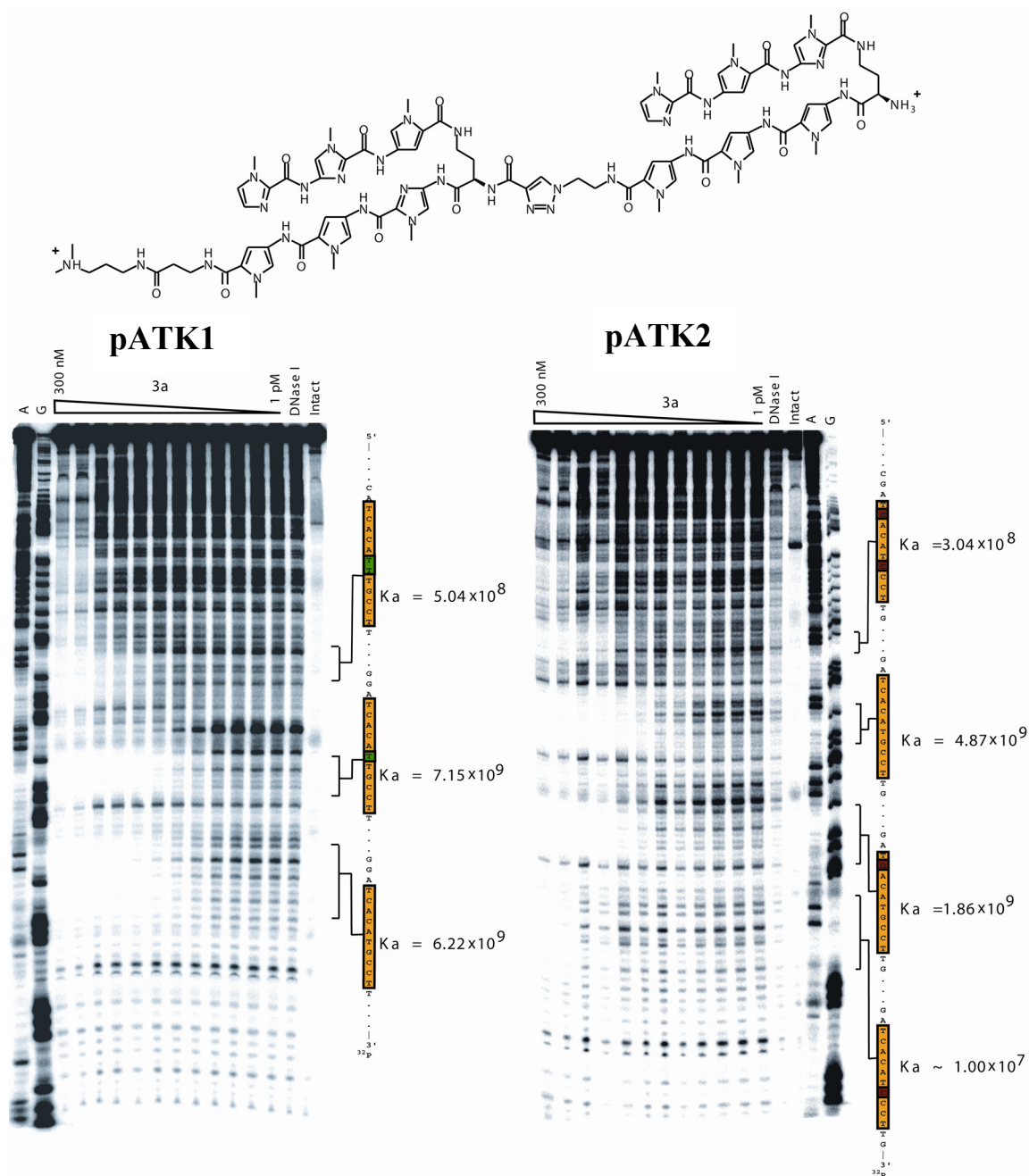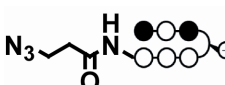
Figure 2.8 Left: Quantitative DNase I footprinting experiments with hairpin dimer **3a** (shown above) on the 3'-^{32}P-labeled DNA fragment derived from plasmid pATK1. Lane 1, A-specific reaction; lane 2, G-specific reaction; lanes 3–14 300 nM, 100 nM, 30 nM, 10 nM, 3 nM, 1 nM, 300 pM, 100 pM, 30 pM, 10 pM, 3 pM, 1 pM **3a**; lane 15, DNase I standard; lane 16, intact DNA. Right: Quantitative DNase I footprinting experiments with hairpin dimer **3a** on the 3'-32P-labeled DNA fragment derived from plasmid pATK2. Lanes 1–12 300 nM, 100 nM, 30 nM, 10 nM, 3 nM, 1 nM, 300 pM, 100 pM, 30 pM, 10 pM, 3 pM, 1 pM **3a**; lane 13, Intact DNA; lane 14, DNase I standard; lane 15, A-specific reaction; lane 16, G-specific reaction. All reactions contained 15 Kcpm labeled DNA and were carried out at 22 °C at pH 7.0 in the presence of 10 mM Tris-HCl, 10 mM KCl, 10 mM MgCl2, and 5 mM CaCl2 with an equilibration time of 36 h. Designed binding sites where equilibrium association constants were obtained are shown to the right side of the gel. Equilibrium association constants are listed next to each site.

bp site (K_a = 7.1x10^9 M^{-1}) with high affinity. The size of the footprint indicates that the polyamide is protecting the entire binding site from cleavage by DNase I. At the 12 bp site, the polyamide shows lower affinity (K_a = 5.4x10^8 M^{-1}), while still protecting the entire 12 bp binding site (Figure 2.8).

Table 2.3. Hairpin half-site separation preference. Equilibrium association constants K_a [M^{-1}] for polyamides **1a**, **2a,** and **3a**.[a-c]
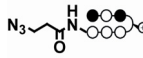
Polyamide		5'-<u>AGGC</u>A<u>TGTGT</u>-3'	5'-<u>AGGC</u>AA<u>TGTGT</u>-3'	5'-<u>AGGC</u>AAA<u>TGTGT</u>-3'
	1a	2.8 × 10^8 (±0.4)[d]	3.4 × 10^8 (±0.4)[d]	3.0 × 10^8 (±0.2)[d]
	2a	1.0 × 10^8 (±0.3)[e]	8.3 × 10^7 (±0.2)[e]	9.6 × 10^7 (±0.3)[e]
	3a	6.2 × 10^9 (±0.9)	7.1 × 10^9 (±0.4)	5.4 × 10^8 (±0.3)

[a] The reported association constants K_a are the average values obtained from three DNase I footprint titration experiments, with the standard deviation for each data set indicated in parentheses. [b] The assays were carried out at 22 °C at pH 7.0 in the presence of 10 mM Tris-HCl, 10 mM KCl, 10 mM MgCl$_2$, and 5 mM CaCl$_2$ with an equilibration time of 12 h. [c] Specificities are given in brackets under the K_a values and are calculated as K_a(match)/K_a(mismatch). [d] Affinity constants for **1a** measured on the 5'-aGGCa-3' and 5'-aGGGa-3' sites. [e] Affinity constants for **2a** measured on the 5'-tGTGt-3' and 5'-tGTCt-3' sites.

Mismatch tolerance for the binding of polyamide **3a** was then examined at the 10 bp binding site using the restriction fragment from pATK2 containing the 10 bp (zero intervening bp) match site (5'-AGGCATGTGT-3'), the 10 bp site with a mismatch under **2a** (5'-AGGCATGT<u>C</u>T-3'), the 10 bp site with a mismatch under module **1a** (5'-AGG<u>G</u>ATGTGT-3') and the 10 bp site with a single-bp mismatch under each of the two hairpins (5'-AGG<u>G</u>ATGT<u>C</u>T-3'). Polyamide **3a** again binds its match site with high affinity (K_a = 4.9x10^9 M^{-1}). When a single base pair mismatch is present under the **2a** half site, polyamide **3a** binds with reduced affinity (K_a = 1.9x10^9 M^{-1}). When a single base pair mismatch is present under the **1a** half-site, the binding affinity of **3a** is greatly

reduced ($K_a \leq 1.0\times10^7$ M^{-1}). When a single base pair mismatch is present at both hairpin

sites, the equilibrium association constant of **3a** is similarly reduced ($K_a \leq 1.0\times10^7$ M^{-1}).

In each case where we measure a binding constant, the entire length of the 10 bp binding

site is protected from cleavage by DNase I.

Table 2.4. Mismatch tolerance at the 10 bp binding site. Equilibrium association constants K_a [M^{-1}] for polyamides **1a**, **2a**, and **3a**.[a-c]

Polyamide	5'-AGGCATGTGT-3'	5'-AGG**G**ATGTGT-3'	5'-AGGCATGT**C**T-3'	5'-AGG**G**ATGT**C**T-3'
1a	3.7×10^8 (± 0.4)^d	$\leq 1 \times 10^7$ (± 0.4)^d [≥ 37]	3.4×10^8 (± 0.2)^d	$\leq 1 \times 10^7$ (± 1.5)^d [≥ 37]
2a	8.6×10^7 (± 0.3)^e	9.9×10^7 (± 0.2)^e	$\leq 1 \times 10^7$ (± 0.3)^e [≥ 9]	$\leq 1 \times 10^7$ (± 0.4)^e [≥ 9]
	4.9×10^9 (± 0.9)	$\leq 1 \times 10^7$ (± 0.4) [≥ 490]	1.6×10^9 (± 0.3) [3]	$\leq 1 \times 10^7$ (± 0.4) [≥ 490]

[a] The reported association constants K_a are the average values obtained from three DNase I footprint titration experiments, with the standard deviation for each data set indicated in parentheses. [b] The assays were carried out at 22 °C at pH 7.0 in the presence of 10 mM Tris-HCl, 10 mM KCl, 10 mM MgCl$_2$, and 5 mM CaCl$_2$ with an equilibration time of 12 h. [c] Specificities are given in brackets under the K_a values and are calculated as K_a(match)/K_a(mismatch). [d] Affinity constants for **1a** measured on the 5'-aGGCa-3' and 5'-aGGGa-3' sites. [e] Affinity constants for **2a** measured on the 5'-tGTGt-3' and 5'-tGTCt-3' sites.

Cell Uptake of Fluorescein-Conjugated Polyamides.

Fluorescein-conjugated analogs of polyamides **1a** and **2a** (**9-17**) were synthesized

according to standard procedures (Figure 2.9).[20, 21] Fluorescein was incorporated either

as the diacetylated 5'-amide, or from fluorescein isothiocyanate (FITC) to yield the

thiourea-linked dye.

Conjugates were screened against a panel of human adherent and nonadherent cell

lines according to literature procedures.[20, 21] The data is summarized in Table 2.5.

Alkyne-functionalized conjugates **9** and **10** were able to localize to the nuclei of most
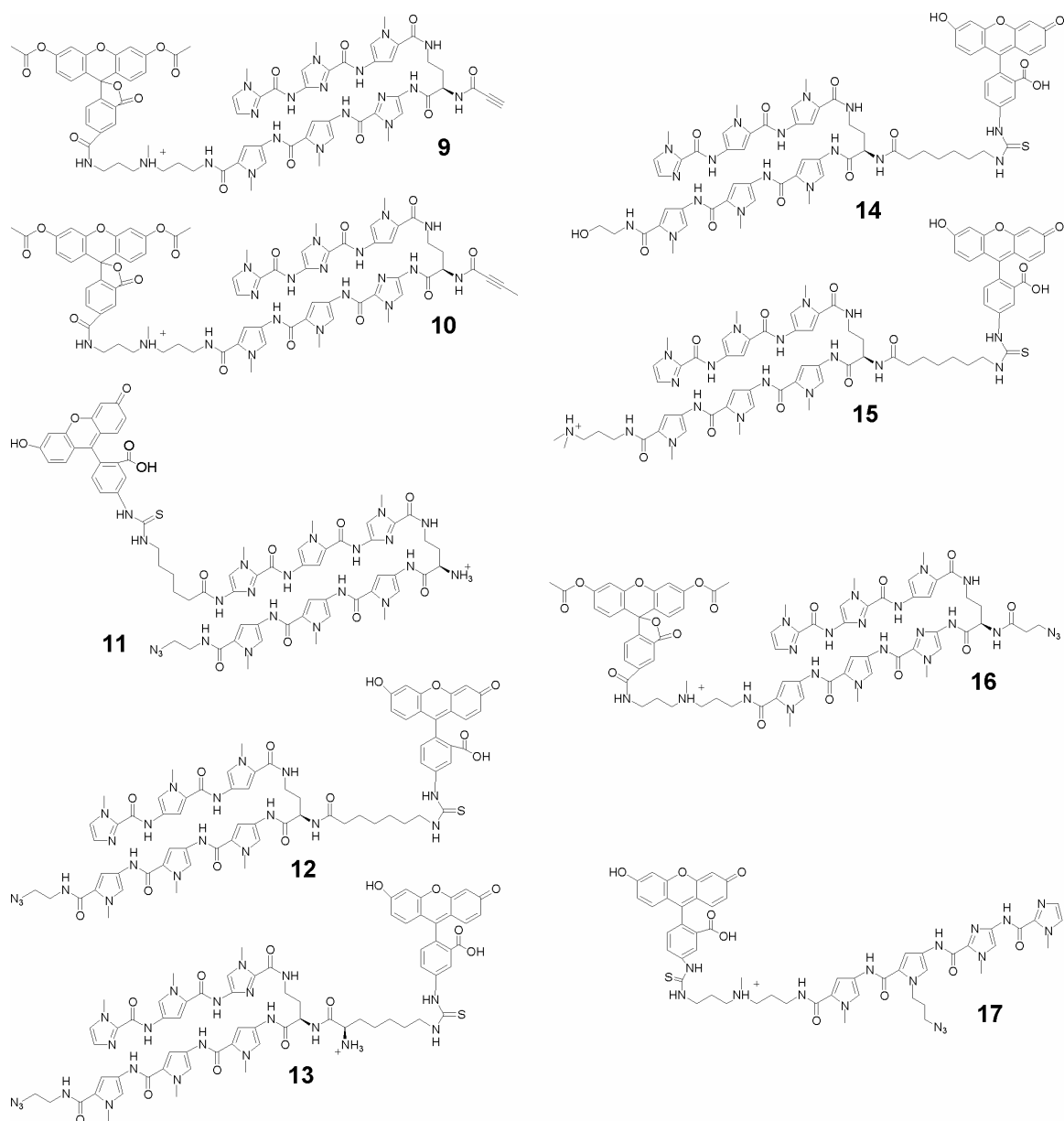
Figure 2.9 Molecular structures of fluorescein dye conjugates synthesized for cell uptake studies.

living cells tested. Conjugate **9**, while showing good nuclear localization, also exhibits

an interesting staining pattern on the surface of the cells. Conjugate **10**, which

is a methyl-protected version of the terminal alkyne **9**, also shows good nuclear

localization without the association with the cell matrix proteins observed with **9**.

		MCF-7	HeLa	PC3	LN-CaP	SK-BR-3	DLD-1	786-O	293	Jurkat	CEM	MEG-01	MEL	NB4
9	Ac₂FAM⁓(+)⁓⬤⬤◯◯◯–propargyl	+	+	+	+	−	+	− −	+	+ +	+ +	−	+	+ +
10	Ac₂FAM⁓(+)⁓◯◯⬤◯⬤–propargyl	+ +	+	+	+	+	+ +	+	+	+ +	+	+	+	+
11	FITC⁓C₇⁓⬤◯⬤ N₃⁓◯◯◯⁓(+	+	− −	− −	− −	−	− −	− −	− −	− −	− −	− −	− −	− −
12	N₃⁓◯◯◯◯⁓C₇⁓FITC	+	− −	− −	− −	− −	− −	− −	− −	− −	− −	− −	− −	− −
13	N₃⁓◯◯◯⁓(+)Lys⁓FITC	+	− −	− −	− −	− −	− −	− −	− −	−	−	− −	− −	− −
14	HO⁓◯◯◯⁓C₇⁓FITC	+	− −*	− −	− −	− −	− −	− −	− −	− −	− −	− −	− −	− −
15	+⁓◯◯◯⁓C₇⁓FITC	−	+*	+	+ +	+	+	−*	− −	+ +	+ +	+	+ +	+
16	Ac₂FAM⁓(+)⁓◯◯◯⁓N₃	+	+	+	− −	+	+ᵇ	−	− −	+	−	− −	− −	− −
17	FITC⁓(+)⁓◯◯⬤⬤ (N₃)	+ +	−	−	+	+	− −	− −	− −	−	+ᵇ	− −	− −	− −

Table 2.5. Cell uptake data for alkyne- and azide-functionalized fluorescein-polyamide conjugates. At left is the schematic representation of each compound. Black and white circles represent imidazole and pyrrole carboxamides, respectively. Data table is filled out with subjective determinations of nuclear localization. Double plus represents a strong nuclear localization; plus represents nuclear accessibility; minus represents weak nuclear localization; double minus represents exclusion from nucleus.

Azide-functionalized conjugates **11**, **12,** and **13** were entirely excluded from the nuclei of most living cells tested, the exception being MCF-7 (kidney cancer) cells, which allow modest uptake (Figure 2.10).

In order to determine whether the azide moiety is responsible for the lack of nuclear localization, controls **14** and **15**, which differ from **12** only in the tail functionality, were assayed against the panel of cells. Terminal alcohol conjugate **14** exhibits a pattern of localization similar to azide **12**, while Dp-functionalized **15** shows strong nuclear localization in many of the cell lines tested.

In order to verify that molecular geometry was not the cause of conjugates **11**–**13** being excluded from the nuclei, conjugate **16** was synthesized. This polyamide possesses

9 - jurkat cells   **11 - jurkat cells**

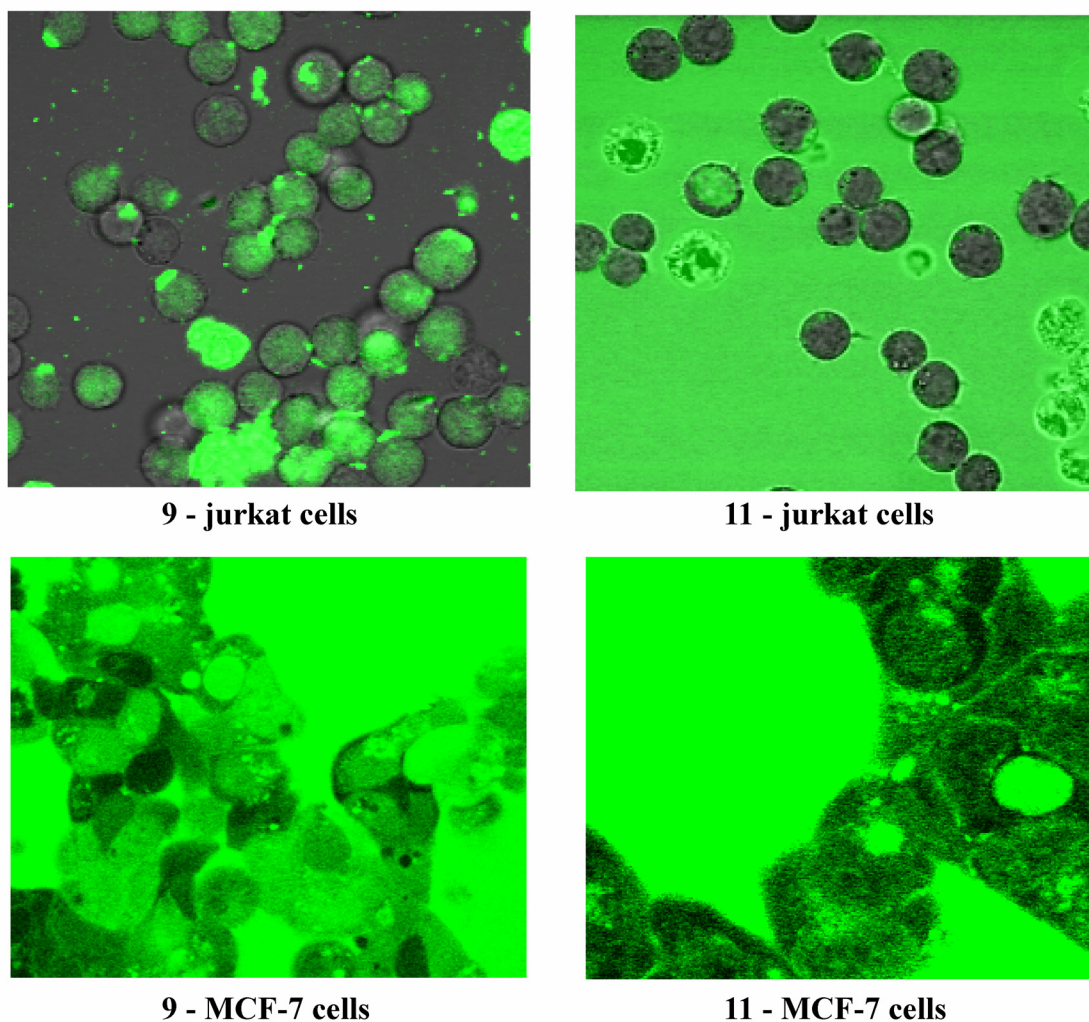**9 - MCF-7 cells**   **11 - MCF-7 cells**

Figure 2.10. Fluorescent images taken with a confocal microscope. At top are shown images with compound **9** and **11** in jurkat cells. Compound **9** shows clear nuclear localization and is rated a "++." Compound **11** is completely excluded from the nuclei, and is rated a "- -." At bottom is shown images with compounds **9** and **11** in mcf-7 cells. Both compounds show some nuclear localization while the majority of the fluorescence is outside the cells and are rated "+." Cells with homogeneously bright areas are dead.

identical geometry to conjugates **9** and **10**, with the dye conjugated to the C-terminal tail and the azide conjugated to the turn amine. This polyamide was able to access the nuclei in several cell lines.

As a final control, the short, 4-ring polyamide-azide-dye conjugate **17** was synthesized. Interestingly, this smaller molecule was still only able to access the nuclei of a few cells.

Discussion

This work represents an exploratory effort towards using the minor groove of double-helical DNA to template a chemical reaction wherein the polyamide product encodes the base sequence content of the DNA template. The DNA-templated tandem polyamide formation occurs through the Hüisgen 1,3-dipolar cycloaddition reaction in aqueous media at 37 °C (pH = 7.0). The minor groove of DNA appears to impart some additional steric constraint upon the 1,3-dipolar cycloaddition. In each pairing, the product ratio of regioisomers is increased in the templated versus the non-templated reaction. While non-templated pairings of the activated alkyne **1a** with either **2a** or **2b** resulted in a 20:1 ratio of regioisomers, these reactions, when templated by DNA, produce the major thermal product exclusively. Non-templated couplings of unactivated alkyne **1b** resulted in formation of both regioisomers in equal amounts. However, when templated on DNA, the ratio was increased to the point where only a single isomer was produced (**1b** with **2a**, duplex **A**; and **1b** with **2b**, duplex **B**), or to a ratio of 3:1 (**1b** with **2b**, duplex **A**). These results can be rationalized by the fact that the 1,4-regioisomer is formed by an antiparallel approach of the two reactants while the 1,5-regioisomer is formed from a parallel approach of the two reactants. When the reactive partners are sequestered in the minor groove of DNA, the linkers are forced to span the space between the two hairpin binding sites and approach each other in an antiparallel fashion, thereby

favoring the pathway leading to the 1,4-regioisomer. The fact that any of the minor isomer is produced by the longest linkers on the shortest template suggests that the space between the hairpin binding sites is short enough for the long linkers to approach each other in a parallel fashion while still being close enough to react.

Both the activated alkyne (**1a**) and the alkyl alkyne (**1b**) were unreactive toward either azide (**2a–b**) in the absence of DNA template up to 1 μM concentrations. Hairpins **1a** and **2a** at similar concentrations form tandem **3a** on a template **A** with a rate increase of greater than 10,000-fold over the non-templated version. Hairpin **1a** also forms tandem- dimer **4ab** on the 10 bp duplex **A** (zero intervening base pairs) when paired with the longer, more flexible azide **2b**. However, the rate of tandem formation was slower than the rate of formation from **1a** and **2a**. This decreased rate is perhaps due to the additional flexibility in the linker, which allows the reactants more freedom to adopt non-productive conformations. Hairpin **1b**, which exhibits a longer alkyne linkage (2 additional methylene units), was also tested with the azido-functionalized hairpins **2a** and **2b**. Both pairings resulted in tandem dimer formation on the 10 bp site. The pairing of **1b** and **2b** also showed product formation on the longer 11 bp site, with the 10 bp site preferred by a ratio of 6:1. However, the rate of product formation by these pairings was 800 times slower than the rate of formation between **1a** and **2a** on the 10 bp binding site. The decreased rate may be due to the inherent differences in reactivity between **1a** and **1b**. Thus, while DNA template **A** increases the rate of cycloadditions between **1b** and either azide more than two orders of magnitude, the reaction still takes weeks to proceed to moderate yields. The rate of DNA-templated cycloaddition with the activated alkynyl

amidate **1a** is increased more than 10,000-fold by template **A**, which causes the reaction to proceed in hours, a timescale that is relevant to biological applications.

The double-helical DNA-templated cycloaddition also shows a dependence on the flexibility of the linkers between the chemical reactants and the DNA-binding domains. The longer, more flexible linkers (**1b** + **2b**) allow the cycloaddition to proceed at both the 10 bp and 11 bp template sites. However, the more restricted linker (**1a** + **1b**) limits the intervening distance over which cycloaddition may occur to a single bp. Both reactivity and orientation of the azide and alkyne moieties appears critical to optimal reactivity.

The cycloaddition reaction is also shown to be dependent not only upon the spacing between the hairpins, but also upon the sequence composition of the DNA under the template recognition elements of the **1a** and **2a** optimal pair. When a single base-pair mismatch is present under the azide-functionalized hairpin polyamide **2a** at the 10 bp binding site, the rate of the tandem-forming cycloaddition is nearly halved. Additionally, when a single base-pair mismatch is introduced under each of the hairpin modules at the 10 bp binding site, the cycloaddition reaction is reduced 30-fold.

When the concentration of reacting species is varied from 1 μM to 500 nM, the initial rates of cycloadduct **3a** formation do not significantly change, which implies that pseudozero-order kinetics apply to this reaction. While the rate equations that describe this process fully are likely complex (and not the focus of this paper), polyamide association to DNA is known to be near diffusion limited.[61] Once the polyamides occupy both half-sites on DNA, the cycloaddition is an *intramolecular* reaction that competes with complex dissociation. We believe that at concentration regimes near the IC_{50} values for **1a** and **2a** (~10 nM), the kinetics would reflect the DNA bound intramolecular

cycloaddition reaction. It is clear from the DNase I footprint titration experiments that at concentrations between 1 μM and 500 nM, hairpins **1a** and **2a** cannot fully distinguish their match and mismatch sites, likely occupying both sites. Thus, tandem **3a** formation observed on duplex **D** and **E** (10 bp site, 1 or 2 mismatch bases, respectively) is most likely a consequence of complex formation despite the presence of mismatch DNA. Furthermore, the fact that the reactions proceed to less than 50% yield is most likely a consequence of the ability of polyamides **1a** and **2a** to occupy mismatch DNA sites and thus bind in non-product-forming orientations. The kinetic results on the templated formation of **3a** on duplex **E** (10 bp site, two single bp mismatches) further support this hypothesis. At concentrations of 1 μM, **3a** is formed in 16% yield, while at 750 nM and 500 nM concentrations **3a** is formed in only 8% yield. This suggests a threshold exists between 1 μM and 750 nM that defines the ability of these polyamides to recognize their match site in preference to the double base pair mismatch site. It follows that at some lower concentration, there is another threshold that allows the templated cycloaddition to discriminate its match site from a single-bp mismatch. Thus, at concentrations at which these polyamides can fully distinguish their match site, the ratio of product formation on match versus mismatch DNA as well as the overall tandem yield should increase.

Once establishing that the pair of hairpins **1a** and **2a** showed the most favorable template-directed cycloaddition with respect to rate and specificity, the binding properties of the hairpin starting materials and tandem product **3a** were analyzed by quantitative DNase I footprinting. The tandem dimer formed between **1a** and **2a** shows a >12-fold increase in binding affinity over either of the two hairpin starting materials. At the 10 bp site, tandem **3a** exhibits only a modest 3-fold specificity for its match site over a single

base pair mismatch under the azide-functionalized **2a**, but shows good specificity over a single base pair mismatch under the alkyne-functionalized **1a** and over a double base pair mismatch (>400 fold). Dimer **3a** targets both the 10 and 11 bp sequences of DNA with high affinity and specificity. However, because the product is formed solely at the 10 bp site, and subsequently exhibits high affinity for that site, its rate of dissociation from that match site should be very slow, rendering it specific for the 10 bp site.

Finally, because our eventual goal is to form these tandem structures inside living cells, nuclear localization of alkyne and azide-functionalized polyamide-fluorescein conjugates was assayed. The alkyne-functionalized polyamides **9** and **10** were able to strongly localize to the nuclei of most cells tested. Interestingly, the terminal alkyne showed some staining along the cellular membrane. Perhaps the alkyne moiety is undergoing some Michael-type reaction with cell matrix proteins, thus causing the unusual staining pattern. Protection of the terminal alkyne with a methyl group (compound **10**) led to a molecule with good nuclear localization and decreased membrane staining.

Compounds functionalized with an azide at the tail were unable to access the nuclei of most cells tested. That control compound **15**, with the Dp tail in place of the ethyl azide, shows strong nuclear localization in many cell lines indicates that the azide functionality of compounds **11–13** may be a negative determinant for polyamide uptake. That compound **16** localized to the nuclei of several cell lines indicates that the placement of the azide group is important for cell uptake, and that polyamides with the azide moiety conjugated to the γ-aminobutyric acid turn residue possess better uptake properties than their C-terminal azide counterparts.

Conclusion.

Double-helical DNA is capable of templating the site-specific formation of tandem-hairpin dimers. At μM concentrations, the reaction between hairpin polyamides **1a** and **2a** in the DNA minor groove exhibit a 10,000-fold rate enhancement. The templated 1,3-dipolar cycloaddition reaction is sensitive to separation distance between adjacent DNA binding sites. The Watson-Crick sequence information in the DNA helix is encoded by the sequence composition in the pyrrole-imidazole polyamide product. In this supramolecular system, the input molecule is DNA and the output molecule is an organic product with "improved function" with regard to DNA recognition properties (i.e., increased binding-site size and higher affinity). By extension, regarding the design of chemical systems that ligate nonenzymatically in the minor groove of DNA in a sequence dependent fashion,[46] other reactive functional pairs can be considered— such as the Diels-Alder reaction, Staudinger ligation, and S_N2-type reactions— that fit the criteria of thermal reactivity at 37 °C in water with orthogonality towards chemical moieties found inside living systems.

Regarding self-assembly in live cells, we anticipate at least two technical hurdles. Cell uptake remains a large hurdle.[20, 21] Because the azide- functionalized polyamides do not exhibit favorable uptake properties, other ligation reactions, such as those listed above, that utilize polyamides with different functionality, may have to be used. Alternatively, because the turn-linked azide compound did show improved cell uptake, perhaps azide and alkyne pairs can be used for biological templating reactions if

molecules are designed to incorporate the azide only at this position. Secondly, when the starting materials are presented with a gigabase of DNA, will the polyamides be able to avoid the large number of single- and double-bp mismatch sites in order to equilibrate to their unique contiguous match sites? Perhaps we may need to start by targeting repeat regions of DNA such as the centromeres or telomeres that are known to aggregate polyamides.[62] In may be necessary to move to longer, more specific polyamides as starting materials. Finally, the DNA target is condensed one-millionfold on chromatin, and ligation may be better attempted across adjacent minor grooves between two aligned superhelical gyres ("supergrooves" as reaction platforms).[63, 64]

Experimental

Materials

Boc-β-alanine-(4-carboxamidomethyl)-benzyl-ester-copoly(styrene-divinylbenzene) resin (Boc-β-ala-PAM resin) was purchased from Peptides International. Oxime resin was purchased from Nova Biochem. Oligonucleotides for kinetics and footprinting were purchased from the Biopolymer Synthesis and Analysis Facility at the California Institute of Technology. Propiolic acid, sodium azide, and methanesulfonyl chloride were purchased from Aldrich. All other synthetic and footprinting reagents were as previously described.

^{1}H NMR spectra were recorded on a Varian Mercury 300 instrument. UV spectra were recorded on a Beckman Coulter DU 7400 diode array spectrophotometer. Autoradiography was performed with a Molecular Dynamics Typhoon PhosphorImager. MALDI mass spectra were obtained on a Voyager De PRO time-of-flight mass

spectrometer (Applied BioSystems) operated at an accelerating voltage of +20 dV. The samples were dissolved in 50% CH_3CN: 0.1% $TFA-H_2O$ and applied to the target in a α-cyanohydroxycinnamic matrix. The mass spectrometer was calibrated with a calibration mixture provided by the instrument manufacturer. DNA sequencing was performed at the Sequence/Structure Analysis Facility (SAF) at the California Institute of Technology. HPLC analysis was performed on a Beckman Gold system using a RAINEN C_{18}, Microsorb MV, 5 μm, 300 x 4.6 mm reversed-phase column in 0.1% (w/v) $TFA-H_2O$ with acetonitrile as eluent and a flow rate of 1.0 mL/min, gradient elution 1.25% CH_3CN/min. Preparatory HPLC was carried out on a Beckman HPLC using a Waters DeltaPak 25 x 100 mm, 100 μm C_{18} column, 0.1% (w/v) TFA- H_2O, 0.4% CH_3CN/min. Water was obtained from a Millipore MilliQ water purification system and all buffers were 0.2 μm filtered. All reagent-grade chemicals were used without further purification unless otherwise stated.

ImImPy-(R)^{H_2N}γ-ImPyPy-β-Dp (4) Synthesized on solid support according to literature procedures.[53]   UV (H_2O) λ_{max} 310 nm (51540).   MALDI-TOF-MS calcd. for $C_{45}H_{60}N_{19}O_8$ (M + H): 994.5. Found 994.6.

ImImPy-(R)[HC≡COC]^{HN}γ-ImPyPy-β-Dp (1a)   Synthesized from **4** according to modified literature procedures.[53] **4** (30 μmol, 30 mg) was dissolved in 1 mL DMF and 1.5 mL acetonitrile and cooled to 0 °C. Propiolic acid (30 μmol, 2.1 mg) was added, followed by DCC (30 μmol, 6.2 mg), and the solution stirred for 15 min at 0 °C then 1 h at rt. Pure product was obtained after purification by reversed-phase HPLC, followed by

lyophilization of appropriate fractions as a white powder (7.8 mg, 7.5 μmol, 25% recovery); UV (H_2O) λ_{max} 310 nm (51540). ^{1}H NMR (DMSO-d_6) δ 10.35 (s, 1H), 10.31 (s, 1H), 10.13 (s, 1H), 9.90 (s, 1H), 9.71 (s, 1H), 9.22 (bs, 1H), 9.13 (d, 1H, J=4.8 Hz), 8.03 (t, 2H, J=5.7 Hz), 8.02 (m, 2H), 7.55 (s, 1H), 7.45 (s, 1H), 7.44 (s, 1H), 7.43 (s, 1H), 7.24 (d, 1H, J=1.8 Hz), 7.21 (d, 1H, J=1.8 Hz), 7.15 (d, 1H, J=1.5 Hz), 7.13 (d, 1H, J=2.1 Hz), 7.06 (s, 1H), 6.99 (d, 1H, J=1.5 Hz), 6.86 (d, 1H, J=2.1 Hz), 4.51 (q, 1H, J=6.3 Hz), 4.23 (s, 1H), 3.99 (s, 6H), 3.94 (s, 3H), 3.82 (s, 3H), 3.78 (s, 6H), 3.36 (q, 2H, J=5.7 Hz), 3.22 (m, 2H), 3.09 (q, 2H, J=6.0 Hz), 2.98 (m, 2H), 2.72 (d, 6H, J=4.8 Hz), 2.33 (t, 2H, J=7.2 Hz), 1.95 (m, 2H), 1.72 (m, 2H), MALDI-TOF-MS calcd. for $C_{48}H_{60}N_{19}O_9$ (M + H): 1046.5. Found 1046.6.

ImImPy-(R)[HC≡C(CH$_2$)$_2$OC]^{HN}γ-ImPyPy-β-Dp (1b)   Synthesized from **4** (3 μmol, 3.1 mg) using a similar procedure to the synthesis of **1a**, substituting 1-pentynoic acid (10 μmol, 1.0 mg) for propiolic acid. Product was obtained as a white powder (2.2 mg, 2.1 μmol, 70% recovery); UV (H_2O) λ_{max} 310 nm (51540). MALDI-TOF-MS calcd. for $C_{50}H_{64}N_{19}O_9$ (M + H): 1074.5. Found 1074.4.

ImPyIm-(R)^{BocHN}γ-PyPyPy-(CH$_2$)$_2$-OH (5a) Product obtained as a white powder upon aminolysis from oxime resin using ethanolamine (neat) (35 mg, 40 μmol, 8% recovery); UV (H_2O) λ_{max} 310 nm (51540). MALDI-TOF-MS calcd. for $C_{45}H_{57}N_{16}O_{10}$ (M + H): 981.4. Found 981.2

ImPyIm-(R)^{BocHN}γ-PyPyPy-(CH$_2$)$_2$-OMs (6a)   Synthesized from **5a** according to modified literature procedures. **5a** (40 μmol, 35 mg) was dissolved in 750 μL CH$_2$Cl$_2$ and cooled to 0 °C. DIEA (0.3 mmol, 38 mg) was added and the solution stirred for 15 min. Methanesulfonyl chloride (90 μmol, 15 mg) was added and the reaction stirred for 10 min at 0 °C then 1 h at room temperature. The solvent was removed by evaporation and the compound purified by reversed-phase HPLC. Product was obtained as a white powder upon lyophilization of the appropriate fractions (30.7 mg, 32 μmol, 80%); UV (H$_2$O) λ_{max} 310 nm (51540).   MALDI-TOF-MS calcd. for C$_{46}$H$_{59}$N$_{16}$O$_{12}$S (M + H): 1059.4. Found 1059.6.

ImPyIm-(R)^{H_2N}γ-PyPyPy-(CH$_2$)$_2$-N$_3$ (2a)   Compound **6a** (32 μmol, 30.7 mg) was dissolved in 750 μL DMF. Sodium azide (1 mmol, 65 mg) was added and the solution stirred at 65 °C for 12 h.[56]   The solvent was evaporated and the residue taken up in 1 mL 50% TFA-CH$_2$Cl$_2$ and allowed to stand for 10 min. Pure product was obtained by reversed-phase HPLC purification followed by lyophilization of the appropriate fractions (9.06 mg, 10 μmol, 31% over 2 steps); UV (H$_2$O) λ_{max} 310 nm (51540). ^{1}H NMR (DMSO-d_6) δ 10.50 (s, 1H), 10.41 (s, 1H), 10.11 (s, 1H), 9.97 (s, 1H), 9.92 (s, 1H), 8.35 (t, 2H, J=5.7 Hz), 8.25 (m, 4H), 7.50 (s, 1H), 7.39 (s, 2H), 7.24 (s, 1H), 7.22 (s, 1H), 7.19 (s, 1H), 7.14 (s, 1H), 7.03 (s, 2H), 6.92 (s, 1H), 6.88 (s, 1H), 3.97 (s, 3H), 3.92 (s, 3H), 3.82 (s, 9H), 3.78 (s, 3H), 3.39 (m, 3H), 2.99 (m, 2H), 1.99 (m, 2H), 1.21 (t, 2H, J=6.0 Hz). MALDI-TOF-MS calcd. for C$_{40}$H$_{48}$N$_{19}$O$_7$ (M + H): 906.4. Found 906.6.

ImPyIm-(R)^{H_2N}γ-PyPyPy-(CH$_2$)$_3$-N$_3$ (2b) Compound **2b** was prepared according to the procedure for **2a**, substituting 3-aminoethanol for ethanolamine for the nucleophilic resin cleavage. Product was obtained in 2.3% overall yield from resin bound starting material; UV (H$_2$O) λ_{max} 310 nm (51540). MALDI-TOF-MS calcd. for C$_{41}$H$_{50}$N$_{19}$O$_7$ (M + H): 920.2. Found 920.4.

ImImPy-(R)-[ImPyIm-(R)^{H_2N}γ–PyPyPy-(CH$_2$)$_2$-Tr-(OC)]^{HN}γ-ImPyPy-β-Dp　　　　(3a)

Compounds **1a** (5 μmol, 5.23 mg) and **2a** (5 μmol, 4.5 mg) were dissolved in 1 mL 20% (v/v) acetonitrile in water. The solution was lyophilized and the powder formed into a tight pellet. The pellet was heated under argon for 6 days at 55 °C. Product was obtained as a white powder after reversed-phase HPLC purification and lyophilization of the appropriate fractions (800 nmol, 1.56 mg, 16%). UV (H$_2$O) λ_{max} 310 nm (103080). ^{1}H NMR (DMSO-d_6) δ 10.51 (s, 1H), 10.41 (s, 1H), 10.38 (s, 1H), 10.29 (s, 1H), 10.13 (s, 1H), 10.11 (s, 1H), 9.97 (s, 1H), 9.91 (s, 2H), 9.73 (s, 1H), 9.33 (bs, 1H), 8.63 (s, 1H), 8.56 (d, 1H, J=7.5 Hz), 8.35 (t, 2H, J=5.5 Hz), 8.33 (d, 2H, J=4.5 Hz), 8.29 (m, 2H), 8.19 (t, 1H, J=5.5 Hz), 8.06 (t, 2H, J=6.0 Hz), 8.03 (t, 2H, J=5.5 Hz), 7.56 (s, 1H), 7.51 (s, 1H), 7.47 (s, 1H), 7.46 (s, 1H), 7.40 (s, 1H), 7.39 (s, 1H), 7.24 (s, 1H), 7.22 (s, 2H), 7.17 (s, 1H), 7.15 (s, 2H), 7.14 (s, 1H), 7.07 (s, 1H), 7.04 (s, 2H), 6.98 (s, 1H), 6.29 (s, 1H), 6.88 (s, 1H), 6.87 (s, 1H), 4.73 (m, 4H), 4.59 (m, 5H), 4.00 (s, 3H), 3.99 (s, 3H), 3.98 (s, 3H), 3.95 (s, 3H), 3.93 (s, 3H), 3.85 (s, 6H), 3.83 (s, 3H), 3.80 (s, 3H), 3.78 (s, 3H), 3.77 (s, 3H), 3.65 (m, 2H), 3.37 (m, 2H), 3.23 (m, 2H), 3.11 (q, 2H, J=6.0 Hz), 3.01 (m, 2H), 2.75 (s, 3H), 2.74 (s, 3H), 2.35 (t, 2H, J=6.5 Hz), 2.09 (m, 2H), 2.00 (m, 6H),

1.72 (qu, 2H, J=8.0 Hz), 1.23 (m, 6H). MALDI-TOF-MS calcd. for $C_{88}H_{107}N_{38}O_{16}$ (M + H): 1952.8. Found 1952.1.

ImImPy-(R)-[ImPyIm-(R)^{H_2N}γ−PyPyPy-(CH$_2$)$_3$-Tr-(OC)]^{HN}γ-ImPyPy-β-Dp (4ab)

Prepared from **1a** and **2b** according to the procedure for **3a** (15% yield); UV (H$_2$O) λ_{max} 310 nm (103080). MALDI-TOF-MS calcd. for $C_{89}H_{109}N_{38}O_{16}$ (M + H): 1966.9. Found 1966.7.

ImImPy-(R)-[ImPyIm-(R)^{H_2N}γ−PyPyPy-(CH$_2$)$_2$-Tr-(CH$_2$)$_2$(OC)]^{HN}γ-ImPyPy-β-Dp (5ab)

Prepared from **1b** and **2a** according to the procedure for **3a** (29% yield); UV (H$_2$O) λ_{max} 310 nm (103080). MALDI-TOF-MS calcd. for $C_{90}H_{111}N_{38}O_{16}$ (M + H): 1980.9. Found 1980.5.

ImImPy-(R)-[ImPyIm-(R)^{H_2N}γ−PyPyPy-(CH$_2$)$_3$-Tr-(CH$_2$)$_2$(OC)]^{HN}γ-ImPyPy-β-Dp (6ab)

Prepared from **1b** and **2b** according to the procedure for **3a** (25% yield); UV (H$_2$O) λ_{max} 310 nm (103080). MALDI-TOF-MS calcd. for $C_{91}H_{113}N_{38}O_{16}$ (M + H): 1994.9. Found 1994.8.

Polyamides 9 and 10.

Resin-bound, Fmoc-protected precursor polyamide was treated with 20% piperidine in DMF for 20 min. The resin was washed and then treated with 5 mole

equivalents of t-butoxycarbonyl (BOC) anhydride in 4:1 DMF:DIEA for 2 h. BOC-protected polyamide was liberated from resin with neat 3,3'-N-methyaminodipropylamine. Cleaved product was purified by reversed phase HPLC

(25 % yield). Product was then reacted with 5'fluorescein carboxylic acid diacetate (1.2 mole equivalents), PyBOP (1.5 equiv.), in 4:1 DMF:DIEA. When reaction was complete (by analytical HPLC), the solvent was removed *in vacuo*, and the residue taken up in 50% TFA:DCM for 20 min. The reaction was dried, and the purified by reversed-phase HPLC (70% yield). Final alkyne functionalization was performed as for compounds **1a** and **b** (5% overall yield). UV (H_2O) λ_{max} 310 nm (51540). MALDI-TOF-MS calcd. (M + H): **9** 1461.56, **10** 1475.50. Found **9** 1461.8, **10** 1475.9.

Polyamides 11 and 12.

Resin-Bound, Fmoc-protected polyamide precursors were treated with 20% piperidine in DMF for 20 min. 6-amino hexanoic acid (**11**) or lysine (**12**) were then conjugated to the liberated amine using standard amide bond–forming conditions. The polamides were then liberated from resin in neat ethanolamine, and purified by reversed-phase HPLC (20% yield). The terminal hydroxyl groups were transformed into azides according to the procedure for **2a**. The compounds were then coupled to FITC for an overall yield of 2%. UV (H_2O) λ_{max} 310 nm (51540). MALDI-TOF-MS calcd. (M + H): **11** 1421.56, **12** 1437.54. Found **9** 1421.7, **10** 1437.7.

Polyamide 13.

Boc-Im-OH was coupled in place of the terminal Im-OH cap during standard solid phase synthesis of the precursor polyamide. 6-amino hexanoic acid was then coupled. Fluorescein diacetate was then coupled onto the hexanoic acid. The polyamide-dye conjugate was then liberated from resin using ethanolamine, and the azide installed according to the procedure for **2a**. UV (H_2O) λ_{max} 310 nm (51540). MALDI-TOF-MS calcd. (M + H): 1424.49. Found 1424.6.

Polyamides 14 and 15.

Made according to the procedure for **2a**, using ethanolamine (**14**) or Dp (**15**) to cleave the polyamide from resin. FITC was used to conjugate the dye to the turn (12% overall yield). UV (H_2O) λ_{max} 310 nm (51540). MALDI-TOF-MS calcd. (M + H): **14** 1396.49, **15** 1438.59. Found **14** 1396.8, **15** 1438.9.

Polyamide 16.

On resin, 3-azido propionic acid was conjugated to the turn amine. Polyamide was then liberated from resin using 3,3'-N-methylaminodipropylamine. FITC was then used to conjugate the dye to the tail (18% overall yield). UV (H_2O) λ_{max} 310 nm (51540). MALDI-TOF-MS calcd. (M + H): 1506.52. Found 1506.7.

Polyamide 17.

Core polyamide was built using N-propanol, Boc-Py-OH. Polyamide was cleaved from resin using mono-BOC-protected 3,3'-N-methylaminodipropylamine. The

hydroxyl was transformed into an azide according to the procedure for **2a**. The BOC group was removed in TFA:DCM, and FITC used to couple the dye onto the C-terminal amine (10% overall yield). UV (H_2O) λ_{max} 310 nm (34360). MALDI-TOF-MS calcd. (M + H): 1080.16. Found 1080.9.

Preparation of duplex DNA for kinetic experiments. Duplex DNA was prepared by incubating equal amounts of complementary sets of synthetic oligonucleotides at 90 °C for 10 min, then slowly allowing them to cool to rt. Resulting duplex DNA was quantified by UV by the relationship 1 OD_{260} unit = 50 μg/mL duplex DNA. Duplex DNA was stored at -20 °C in water.

Cycloaddition reactions. All kinetic reactions were performed in 1.7 mL presiliconized microcenterfuge tubes obtained from VWR International. Total reaction volumes were 1.2 mL aqueous solutions of equimolar concentrations of each hairpin polyamide and DNA (2 mM Tris-HCl, 2 mM KCl, 2 mM $MgCl_2$, 1 mM $CaCl_2$, pH 7.0, 37 °C). Reactions were monitored by HPLC by direct injection of reaction samples onto a RAINEN C_{18}, Microsorb MV, 5 μm, 300 x 4.6 mm reversed-phase column in 0.1% (w/v) TFA-H_2O with acetonitrile as eluent and a flow rate of 1.0 mL/min, gradient elution 0.5% CH_3CN/min. Peaks were quantified using the Beckman Coulter GOLD software package. Verification of product was determined by MALDI TOF-MS of ~40 μL samples of each reaction concentrated on a ZipTip 2 mg C_{18} pipette tip eluted with 75% (v/v) acetonitrile in 0.1% (w/v) TFA.

Construction of plasmid DNA. Plasmids pATK1 and pATK2 were prepared by hybridization of complementary sets of synthetic oligonucleotides. The hybridized inserts were individually ligated into *Bam*HI/*Hind*III linearized pUC19 using T4 DNA ligase. *E. coli* JM109 high efficiency competent cells were then transformed with the ligated plasmid. Plasmid DNA from ampicilin-resistant white colonies was isolated using a Qiagen Wizard MidiPrep kit. The presence of the desired insert was determined by dideoxy sequencing. Concentration of prepared plasmid was determined by UV by the relationship 1 OD_{260} unit = 50 µg/mL duplex DNA.

Preparation of ^{32}P-end-labeled restriction fragments. Plasmids pATK1 and pATK2 were linearized with *Eco*RI and *Pvu*II restriction enzymes. The linearized plasmids were then treated with Klenow enzyme, deoxyadenosine 5'-[α-^{32}P]triphosphate and thymidine 5'-[α-^{32}P]triphosphate for 3' labeling. The reactions were loaded onto a 7% nondenaturing polyacrylamide gel. The desired bands were visualized by autoradiography and isolated. Chemical sequencing reactions were done according to published methods.

Quantitative DNase I footprinting.[59] DNase I footprinting reactions were carried out as previously described. Photostimulable storage phosphorimaging plates (Storage Phosphor Screen from Molecular Dynamics) were pressed flat against gel samples and exposed for 12–16 hours. Imaging of Storage Phosphor screens was accomplished on a Molecular Dynamics 425E PhosphorImager and the data analyzed using ImageQuant v. 3.2 software.

Binding energetics. Quantitative DNase I footprint titration experiments[59] (10 mM Tris-HCl, 10 mM KCl, 10 mM MgCl$_2$, 5 mM CaCl$_2$, pH 7.0, 22 °C) were performed on the 3'-^{32}P end labeled 270 bp *Eco*RI/*Pvu*II restriction fragment from pATK1 and the 3'-^{32}P end labeled 261 bp *Eco*RI/*Pvu*II restriction fragment from pATK2. Equilibrium association constants for polyamides **1a**, **2a**, and **3a** on the designed binding sites were determined by calculating a fractional saturation value at the site, for each polyamide concentration, and fitting the data to a modified Hill equation.

Cell uptake studies. Done with the help of Tim Best and Ben Edelson. Polyamide/cell incubations were performed by adding 150 μL cells into culture dishes equipped with glass bottoms for direct imaging. Incubations were done in medium containing 1 μM polyamide at 37 ºC for 10–14 hours. Imaging was then performed with a Zeiss LSM 5 Pascal inverted laser scanning microscope using 488 nm laser excitation with a standard fluorescein filter set.

References

1. Darnell, J. E., *Nature Reviews Cancer*, **2002,** *2*, 740–749.

2. Pandolfi, P. P., *Oncogene*, **2001,** *20*, 3116–3127.

3. Dervan, P. B.; Edelson, B. S., *Current Opinion in Structural Biology*, **2003,** *13*, 284–299.

4. deClairac, R. P. L.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E., *Journal of the American Chemical Society*, **1997,** *119*, 7909–7916.

5. Mrksich, M.; Parks, M. E.; Dervan, P. B., *Journal of the American Chemical Society*, **1994,** *116*, 7983–7988.

6. Trauger, J. W.; Baird, E. E.; Dervan, P. B., *Nature*, **1996,** *382*, 559–561.

7. Bailly, C.; Chaires, J. B., *Bioconjugate Chemistry*, **1998,** *9*, 513–538.

8. Faria, M.; Giovannangeli, C., *Journal Of Gene Medicine*, **2001,** *3*, 299–310.

9. Reddy, B. S. P.; Sharma, S. K.; Lown, J. W., *Current Medicinal Chemistry*, **2001,** *8*, 475–508.

10. Kelly, J. J.; Baird, E. E.; Dervan, P. B., *Proceedings of the National Academy of Sciences of the United States of America*, **1996,** *93*, 6981–6985.

11. Kielkopf, C. L.; Baird, E. E.; Dervan, P. D.; Rees, D. C., *Nature Structural Biology*, **1998,** *5*, 104–109.

12. Swalley, S. E.; Baird, E. E.; Dervan, P. B., *Chemistry-a European Journal*, **1997,** *3*, 1600–1607.

13. Trauger, J. W.; Baird, E. E.; Dervan, P. B., *Journal of the American Chemical Society*, **1998,** *120*, 3534–3535.

14. Trauger, J. W.; Baird, E. E.; Mrksich, M.; Dervan, P. B., *Journal of the American Chemical Society*, **1996,** *118*, 6160–6166.

15. Urbach, A. R.; Dervan, P. B., *Proceedings of the National Academy of Sciences of the United States of America*, **2001,** *98*, 4343–4348.

16. Urbach, A. R.; Love, J. J.; Ross, S. A.; Dervan, P. B., *Journal of Molecular Biology*, **2002,** *320*, 55–71.

17. Herman, D. M.; Baird, E. E.; Dervan, P. B., *Chemistry-a European Journal*, **1999,** *5*, 975–983.

18. Kers, I.; Dervan, P. B., *Bioorganic & Medicinal Chemistry*, **2002,** *10*, 3339–3349.

19. Weyermann, P.; Dervan, P. B., *Journal of the American Chemical Society*, **2002,** *124*, 6872–6878.

20. Best, T. P.; Edelson, B. S.; Nickols, N. G.; Dervan, P. B., *Proceedings of the National Academy of Sciences of the United States of America*, **2003,** *100*, 12063–12068.

21. Edelson, B. S.; Best, T. P.; Olenyuk, B.; Nickols, N. G.; Doss, R. M.; Foister, S.; Heckel, A.; Dervan, P. B., *Nucleic Acids Research*, **2004,** *32*, 2802–2818.

22. Goodsell, D. S.; Olson, A. J., *Annual Review of Biophysics and Biomolecular Structure*, **2000,** *29*, 105–153.

23. Mikhailov, V. S., *Molecular Biology*, **1999,** *33*, 498–510.

24. Bridson, P. K.; Orgel, L. E., *Journal of Molecular Biology*, **1980,** *144*, 567–577.

25. Bruick, R. K.; Dawson, P. E.; Kent, S. B.; Usman, N.; Joyce, G. F., *Chemistry & Biology*, **1996,** *3*, 49–56.

26. Calderone, C. T.; Puckett, J. W.; Gartner, Z. J.; Liu, D. R., *Angewandte Chemie-International Edition*, **2002,** *41*, 4104–4108.

27. Dolinnaya, N. G.; Sokolova, N. I.; Gryaznova, O. I.; Shabarova, Z. A., *Nucleic Acids Research*, **1988,** *16*, 3721–3738.

28. Ferris, J. P.; Huang, C. H.; Hagan, W. J., *Nucleosides & Nucleotides*, **1989,** *8*, 407–414.

29. Fujimoto, K.; Matsuda, S.; Takahashi, N.; Saito, I., *Journal of the American Chemical Society*, **2000,** *122*, 5646–5647.

30. Gartner, Z. J.; Grubina, R.; Calderone, C. T.; Liu, D. R., *Angewandte Chemie-International Edition*, **2003,** *42*, 1370–1375.

31. Gartner, Z. J.; Kanan, M. W.; Liu, D. R., *Journal of the American Chemical Society*, **2002,** *124*, 10304–10306.

32. Gartner, Z. J.; Kanan, M. W.; Liu, D. R., *Angewandte Chemie-International Edition*, **2002,** *41*, 1796–+.

33. Gartner, Z. J.; Liu, D. R., *Journal of the American Chemical Society*, **2001,** *123*, 6961–6963.

34. Kanaya, E.; Yanagawa, H., *Biochemistry*, **1986,** *25*, 7423–7430.

35. Lohrmann, R.; Bridson, P. K.; Bridson, P. K.; Orgel, L. E., *Science*, **1980,** *208*, 1464–1465.

36. Lohrmann, R.; Orgel, L. E., *Journal of Molecular Biology*, **1980,** *142*, 555–567.

37. Luo, P. Z.; Leitzel, J. C.; Zhan, Z. Y. J.; Lynn, D. G., *Journal of the American Chemical Society*, **1998,** *120*, 3019–3031.

38. Naylor, R.; Gilham, P. T., *Biochemistry*, **1966,** *5*, 2722–2789.

39. Niemeyer, C. M., *Current Opinion in Chemical Biology*, **2000,** *4*, 609–618.

40. Orgel, L. E.; Lohrmann, R., *Accounts of Chemical Research*, **1974,** *7*, 368–377.

41. Sokolova, N. I.; Ashirbekova, D. T.; Dolinnaya, N. G.; Shabarova, Z. A., *Febs Letters*, **1988,** *232*, 153–155.

42. Summerer, D.; Marx, A., *Angewandte Chemie-International Edition*, **2002,** *41*, 89–90.

43. Xu, Y. Z.; Karalkar, N. B.; Kool, E. T., *Nature Biotechnology*, **2001,** *19*, 148–152.

44. Xu, Y. Z.; Kool, E. T., *Journal of the American Chemical Society*, **2000,** *122*, 9040–9041.

45. Ye, J. D.; Gat, Y.; Lynn, D. G., *Angewandte Chemie-International Edition*, **2000,** *39*, 3641–3643.

46. Luebke, K. J.; Dervan, P. B., *Journal of the American Chemical Society*, **1989,** *111*, 8733–8735.

47. Huisgen, R., *Profiles, Pathways, and Dreams*. American Chemical Society: Washington, DC, 1994.

48. Kolb, H. C.; Finn, M. G.; Sharpless, K. B., *Angewandte Chemie-International Edition*, **2001,** *40*, 2004–2006.

49. Lewis, W. G.; Green, L. G.; Grynszpan, F.; Radic, Z.; Carlier, P. R.; Taylor, P.; Finn, M. G.; Sharpless, K. B., *Angewandte Chemie-International Edition*, **2002,** *41*, 1053–1356.

50. Wang, Q.; Chan, T. R.; Hilgraf, R.; Fokin, V. V.; Sharpless, K. B.; Finn, M. G., *Journal of the American Chemical Society*, **2003,** *125*, 3192–3193.

51. Warrener, R. N.; Butler, D. N.; Margetic, D.; Pfeffer, F. M.; Russell, R. A., *Tetrahedron Letters*, **2000,** *41*, 4671–4675.

52. Huisgen, R., *1,3-Dipolar Cycloaddition Chemistry*. Wiley-Interscience: New York, 1984; Chapter 1.

53. Baird, E. E.; Dervan, P. B., *Journal of the American Chemical Society*, **1996,** *118*, 6141–6146.

54. Brunton, S. A.; Jones, K., *Journal of the Chemical Society-Perkin Transactions 1*, **2000**, 763–768.

55. Belitsky, J. M.; Nguyen, D. H.; Wurtz, N. R.; Dervan, P. B., *Bioorganic & Medicinal Chemistry*, **2002,** *10*, 2767–2774.

56. Saxon, E.; Bertozzi, C. R., *Science*, **2000,** *287*, 2007–2010.

57. Fazio, F.; Bryan, M. C.; Blixt, O.; Paulson, J. C.; Wong, C. H., *Journal of the American Chemical Society*, **2002,** *124*, 14397–14402.

58. Hlasta, D. J.; Ackerman, J. H., *Journal of Organic Chemistry*, **1994,** *59*, 6184–6189.

59. Trauger, J. W.; Dervan, P. B., Footprinting methods for analysis of pyrrole-imidazole polyamide/DNA complexes. In *Drug-Nucleic Acid Interactions*, 2001; Vol. 340, pp 450–466.

60. Swalley, S. E.; Baird, E. E.; Dervan, P. B., *Journal of the American Chemical Society*, **1999,** *121*, 1113–1120.

61. Baliga, R.; Baird, E. E.; Herman, D. M.; Melander, C.; Dervan, P. B.; Crothers, D. M., *Biochemistry*, **2001,** *40*, 3–8.

62. 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.; Trask, B. J., *Nucleic Acids Research*, **2002,** *30*, 2790–2799.

63. Edayathumangalam, R. S.; Weyermann, P.; Gottesfeld, J. M.; Dervan, P. B.; Luger, K., *Proceedings of the National Academy of Sciences of the United States of America*, **2004,** *101*, 6864–6869.

64. Suto, R. K.; Edayathumangalam, R. S.; White, C. L.; Melander, C.; Gottesfeld, J. M.; Dervan, P. B.; Luger, K., *Journal of Molecular Biology*, **2003,** *326*, 371–380.