Extended hairpin polyamide motif for sequence-specific recognition in the minor groove of DNA

John W Trauger, Eldon E Baird and Peter B Dervan

Background: Three-ring polyamides containing *N*-methylimidazole and *N*-methylpyrrole amino acids bind sequence-specifically to double-helical DNA by forming side-by-side complexes in the minor groove. Simple pairing rules relate the amino-acid sequence of a pyrrole–imidazole polyamide to its expected DNA target site, and polyamides that target a wide variety of DNA sequences have been synthesized. We have shown previously that two three-ring subunits could be linked together by an aliphatic amino acid, increasing the binding affinity of the polyamide and, in some cases, increasing the length of the target sequence. We set out to determine whether different types of linkers could be used in a single molecule to generate a nine-ring polyamide molecule that would bind to specific DNA sequences.

Results: A nine-ring pyrrole–imidazole polyamide, containing two different amino acid linkers, β-alanine and γ-aminobutyric acid, has been synthesized and shown to specifically bind a designated nine-base-pair target site at subnanomolar concentration in a novel extended hairpin conformation.

Conclusions: The amino acids γ-aminobutyric acid and β-alanine optimally link three-ring pyrrole–imidazole subunits in 'hairpin' and 'extended' conformations, respectively. Both aliphatic amino acids can be combined to generate a nine-ring polyamide that specifically recognizes a nine-base-pair target site with very high affinity.

Address: Arnold and Mabel Beckman Laboratories of Chemical Synthesis, California Institute of Technology, Pasadena, CA 91125, USA.

Correspondence: Peter B Dervan
e-mail: dervan@starbase1.caltech.edu

Key words: binding affinity, linker amino acid, pyrrole-imidazole, solid-phase synthesis

Received: **28 Mar 1996**
Revisions requested: **23 April 1996**
Revisions received: **26 April 1996**
Accepted: **26 April 1996**

Chemistry & Biology May 1996, **3**:369–377

 ISSN 1074-5521

Introduction

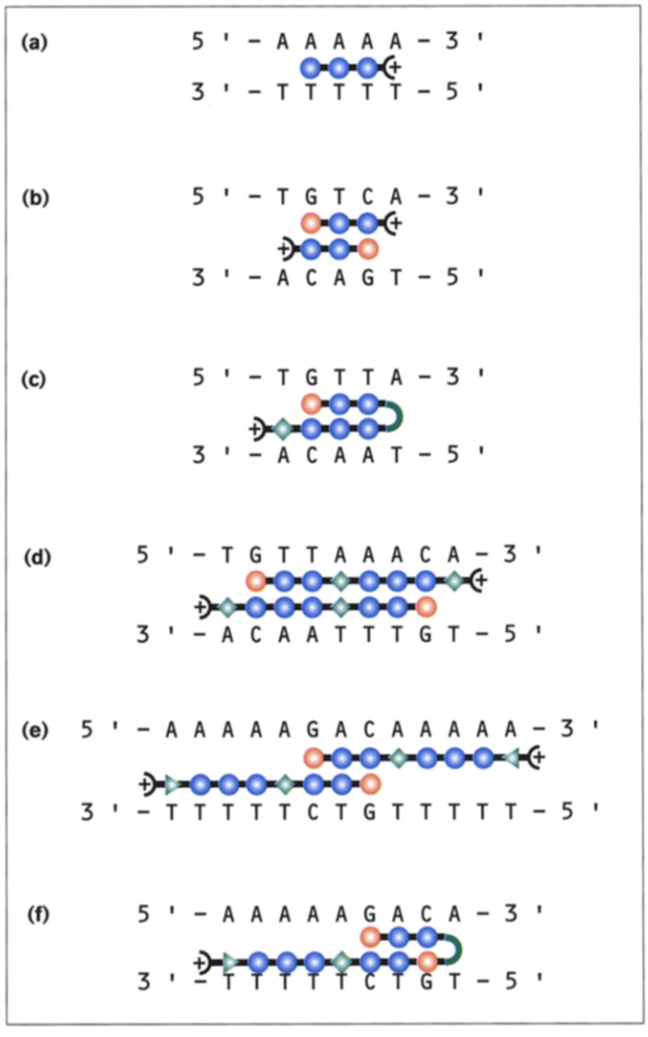
The three-ring polyamide ImPyPy-Dp (Im = *N*-methylimidazole, Py = *N*-methylpyrrole, Dp = dimethylaminopropylamide) was found by footprinting and affinity cleavage studies to specifically bind to the sequence 5'-TGTCA-3' [1]. In the $(ImPyPy\text{-}Dp)_2$•5'-(A,T)G(A,T)C(A,T)-3' complex, each polyamide makes specific contacts with one strand on the floor of the minor groove such that the sequence specificity depends on the sequence of side-by-side amino acid pairings between the two polyamide molecules [1–3]. A pairing of imidazole opposite pyrrolecarboxamide targets a G•C base pair, and a pairing of pyrrolecarboxamide opposite imidazole targets a C•G base pair [1–3]. A pyrrole–pyrrole combination is partially degenerate and targets both T•A and A•T base pairs [1–6]. Specificity for G,C base pairs results from the formation of a hydrogen bond between the imidazole N3 atom and the exocyclic amino group of guanine [1,2]. The generality of these pairing rules has been demonstrated by targeting a wide variety of sequences [7–11].

Hairpin and extended binding motifs

In addition to elucidating the scope and limitations of the sequence-specificity rules, we have made efforts to increase the DNA-binding affinity of polyamides by covalently linking three-ring polyamide subunits [12–15]. Some of the polyamide molecules that have been studied and their interactions with DNA are depicted schematically in Figure 1. The polyamide ImPyPy-γ-PyPyPy-Dp, containing a 'turn' amino acid, γ-aminobutyric acid (γ), specifically binds to the designated target site 5'-TGTTA-3' in a 'hairpin' conformation (Fig. 1c) with an equilibrium association constant of $K_a = 8 \times 10^7\ M^{-1}$. This is an increase of ~800-fold relative to unlinked three-ring polyamides [15]. Addition of a carboxy-terminal β-alanine residue enhances both the binding affinity and sequence specificity of hairpin polyamides [16] (Fig. 1b).

The six-ring polyamide ImPyPy-β-PyPyPy-Dp containing an internal β-alanine (β) residue specifically binds as a dimer in an extended conformation to two designated target sites, 5'-TGTTAAACA-3' (9 base pairs (bp)) and 5'-AAAAAGACAAAAA-3' (13 bp), with equilibrium association constants of $K_a = 8 \times 10^8\ M^{-1}$ and $K_a = 5 \times 10^9\ M^{-1}$, respectively. This represents increases of 10-fold for the 9-bp recognition sequence and 100-fold for the 13-bp sequence over the formally *N*-methylpyrrole-linked polyamide ImPyPy-Py-PyPyPy-Dp [17]. Addition of a carboxy-terminal amino acid regulates the specificity between the 13-bp and 9-bp binding modes. Two ImPyPy-β-PyPyPy-β-Dp polyamides, each containing a carboxy-terminal β-alanine residue, can bind directly opposite one

Figure 1



Schematic models of specific polyamide–DNA complexes. **(a)** 1:1 complex of distamycin, **(b)** 2:1 complex of ImPyPy-Dp, **(c)** 'hairpin' complex of ImPyPy-γ-PyPyPy-β-Dp, **(d)** 9 bp and **(e)** 13 bp 'extended' complexes formed by ImPyPy-β-PyPyPy-β-Dp and ImPyPy-β-PyPyPy-G-Dp, respectively, and **(f)** 'extended hairpin' polyamide ImPyPy-γ-ImPyPy-β-PyPyPy-G-Dp bound to a nine base pair target site. Red and blue circles represent imidazole and pyrrole rings, respectively. Green triangles and diamonds represent glycine and β-alanine, respectively. γ-aminobutyric acid is represented as a curved green line.

another (Fig. 1d), whereas two ImPyPy-β-PyPyPy-G-Dp polyamides, each containing a carboxy-terminal glycine residue, strongly favor binding in the 13-bp binding mode (Fig. 1e). The 13-bp binding mode integrates the 2:1 and 1:1 polyamide–DNA binding motifs at a single site [18]. The ImPyPy moieties of two ImPyPy-β-PyPyPy-G-Dp polyamides bind the central 5'-AGACA-3' sequence in a 2:1 manner, as does the ImPyPy homodimer, and the PyPyPy moieties of the polyamides bind the all-A,T flanking sequences as in the 1:1 distamycin–DNA complexes (Fig. 1a).

It has been demonstrated that γ-aminobutyric acid does not optimally link polyamide subunits in extended conformations, and β-alanine does not optimally link polyamide subunits in hairpin conformations. The polyamide ImPyPy-β-PyPyPy-Dp binds the hairpin site 5'-TGTTAgacc-3' with an association constant of $K_a \leq 2 \times 10^6\ M^{-1}$, a decrease of ≥ 40-fold relative to ImPyPy-γ-PyPyPy-Dp, and displays a cooperative binding isotherm (eq. 2, n = 2; see Materials and methods) in quantitative footprinting experiments at this site, consistent with mismatched binding as an intermolecular dimer [15]. The polyamide ImPyPy-γ-PyPyPy-Dp binds the 13-bp site 5'-AAAAAGACA-AAAA-3' with an association constant of $K_a = 6 \times 10^6\ M^{-1}$, a decrease of 700-fold relative to ImPyPy-β-PyPyPy-Dp, and displays a Langmuir binding isotherm (eq. 2, n = 1) at this site, consistent with mismatched binding as a hairpin. The site 5'-TGTTAAACA-3' is a match site for both hairpin (5'-TGTTA-3' and 5'-AAACA-3') and extended binding by ImPyPy-X-PyPyPy-Dp polyamides. ImPyPy-γ-PyPyPy-Dp and ImPyPy-β-PyPyPy-Dp both bind to this sequence with high affinity ($K_a = 1 \times 10^8\ M^{-1}$ and $K_a = 8 \times 10^8\ M^{-1}$, respectively), but display, respectively, a Langmuir isotherm consistent with hairpin binding, and a cooperative isotherm consistent with extended binding. Remarkably, given their simplicity and similarity, γ-aminobutyric acid and β-alanine selectively link polyamide subunits in hairpin and extended conformations, respectively.

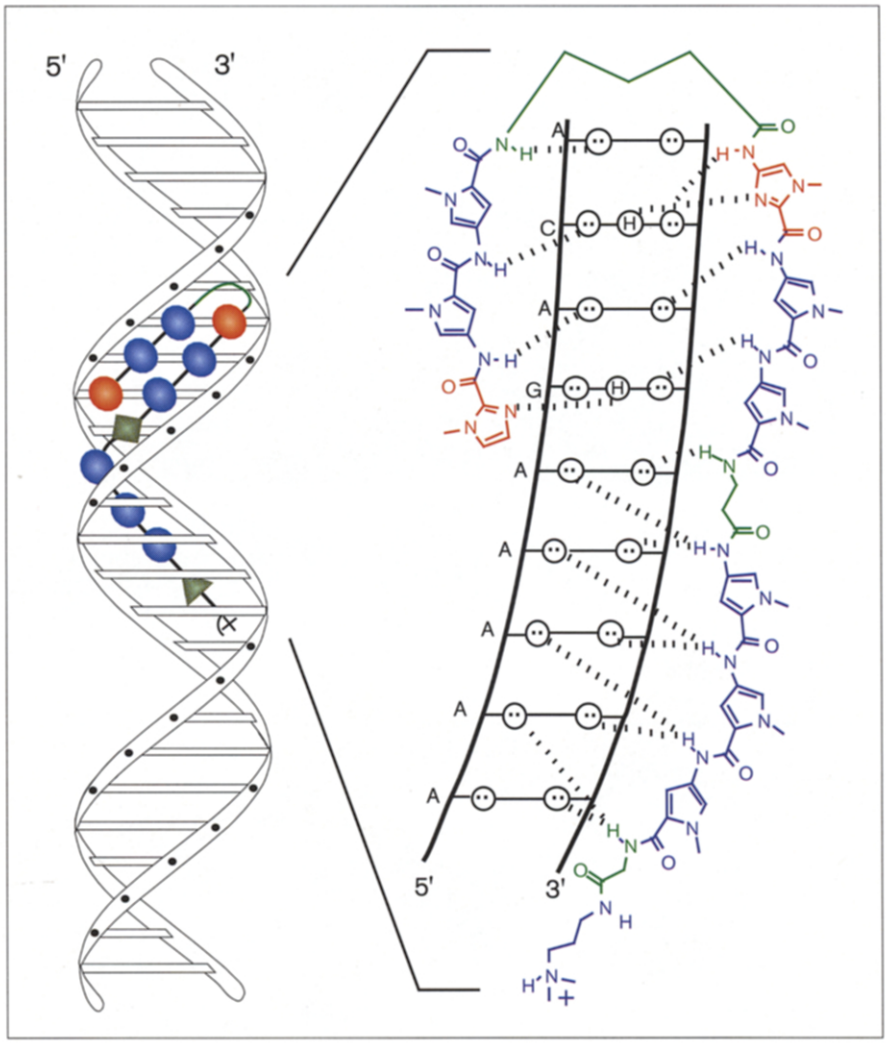
Extended hairpin motif

The results described above suggest that γ-aminobutyric acid and β-alanine could be combined within a single polyamide with predictable results. We report here the synthesis of the nine-ring polyamide ImPyPy-γ-ImPyPy-β-PyPyPy-G-Dp and its association constant for the designated 9-bp target site 5'-AAAAAGACA-3'. The expected ImPyPy-γ-ImPyPy-β-PyPyPy-G-Dp•5'-AAA-AAGACA-3' 'extended hairpin' complex integrates the 1:1 and 2:1 polyamide–DNA motifs at a single site (Fig. 1f, Fig. 2). Each of the three linkers within the polyamide is expected to fulfill a specific function: γ links subunits in a hairpin conformation, β links subunits in an extended conformation, and G confers specificity for the 1:1 binding mode at the carboxyl terminus of the polyamide.

In principle, the 'extended hairpin' motif could provide a general motif for targeting hairpin binding sites having an $(A,T)_4$ flanking sequence. Since six-ring hairpin polyamides already have high DNA-binding affinity ($K_a \approx 10^8\ M^{-1}$), adding an optimally linked three-ring subunit should provide a polyamide with very high DNA-binding affinity. To determine the increase in binding affinity provided by the carboxy-terminal PyPyPy subunit, the six-ring hairpin polyamide ImPyPy-γ-ImPyPy-β-Dp

Figure 2

Models of the expected 'extended hairpin' complex of ImPyPy-γ-ImPyPy-β-PyPyPy-G-Dp with 5'-AAAAAGACA-3'. (Left) Red and blue circles represent imidazole and pyrrole rings, respectively. Green triangles and diamonds represent glycine and β-alanine, respectively, and γ-aminobutyric acid is represented as a curved green line. (Right) Circles with dots represent lone electron pairs on N3 of purines and O2 of pyrimidines, and circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dashed lines.



was synthesized and its equilibrium association constant for the site 5'-AGACA-3' determined.

Binding to the 9-bp target site 5'-ATATAGACA-3' was also investigated. Based on previous results with the distamycin analog Ac-PyPyPy-Dp (Ac = acetyl), a preference of ImPyPy-γ-ImPyPy-β-PyPyPy-G-Dp for 5'-AAAAAGACA-3' over 5'-ATATAGACA-3' is expected.

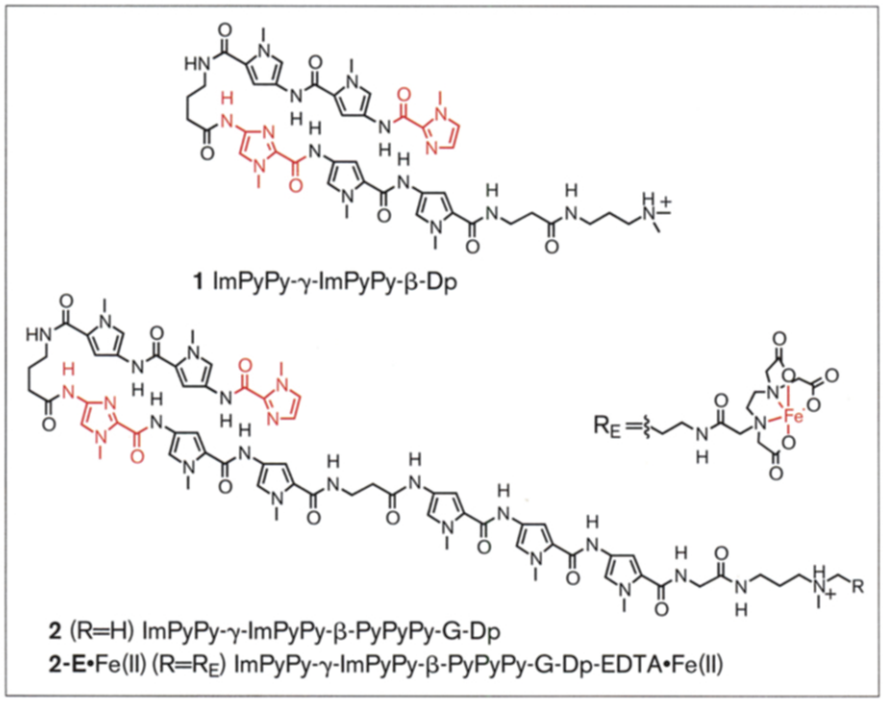
Equilibrium association constants for the polyamides ImPyPy-γ-ImPyPy-β-Dp (compound **1**) and ImPyPy-γ-ImPyPy-β-PyPyPy-G-Dp (compound **2**) for two 9-bp target sites 5'-AAAAAGACA-3' and 5'-ATATAGACA-3' located on a 247-bp restriction fragment, and to additional binding sites identified on the fragment, were determined by quantitative DNase I footprint titration experiments. We also carried out affinity cleavage experiments using the EDTA•Fe(II) polyamide ImPyPy-γ-ImPyPy-β-PyPyPy-G-Dp-EDTA•Fe(II) (**2-E**•Fe(II)) to determine the DNA-binding orientation and qualitative DNA-binding affinity and specificity of this molecule (Fig. 3).

Results and discussion

Synthesis

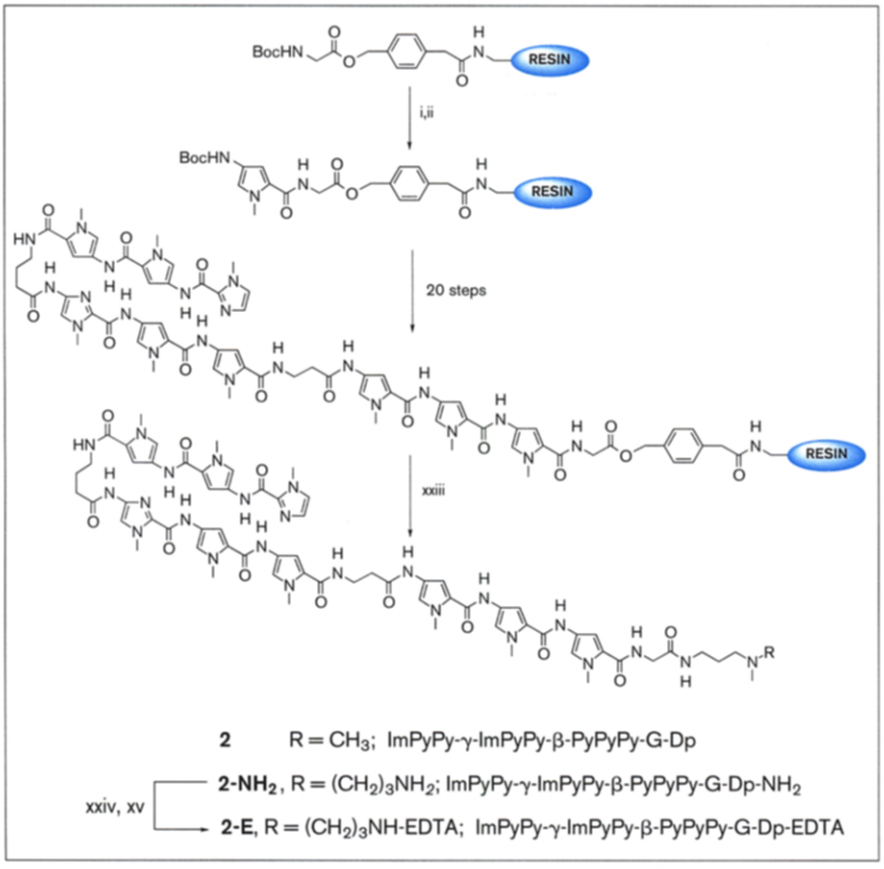
All polyamides were prepared in high purity using solid-phase synthetic methodology [19]. Polyamides **1** and **2** (Fig. 3) were assembled in a stepwise manner on Boc-β-alanine-Pam resin and Boc-glycine-Pam-resin, respectively. Polyamides **1**, **2** and **2-NH$_2$** were cleaved from the support with an appropriate primary amine and purified by reversed-phase high-pressure liquid chromatography (HPLC) (Fig. 4) to provide 10–30 mg of polyamide. Polyamide **2-NH$_2$**, which contains a primary amine group suitable for modification, was treated with an excess of the dianhydride of EDTA, unreacted anhydride was hydrolyzed, and the EDTA-modified polyamide **2-E** was isolated by reversed-phase HPLC.

Figure 3



Structures of polyamides ImPyPy-γ-ImPyPy-β-Dp (compound **1**), ImPyPy-γ-ImPyPy-β-PyPyPy-G-Dp (compound **2**), and ImPyPy-γ-ImPyPy-β-PyPyPy-G-Dp-EDTA•Fe(II) (compound **2-E•Fe(II)**).

Figure 4



Synthetic scheme for preparation of polyamides **2**, **2-NH$_2$** and **2-E**. Cycling protocols consist of trifluoroacetic acid (TFA) deprotection followed by coupling with HOBt-activated Boc-pyrrole or Boc-imidazole ester (steps i–xxii). The resin is then cleaved by treatment with a primary amine (step xxiii) and purified by reversed-phase HPLC. Polyamide **2-NH$_2$** is subsequently modified with EDTA dianhydride (steps xxiv and xxv), providing the affinity-cleaving analog **2-E**.

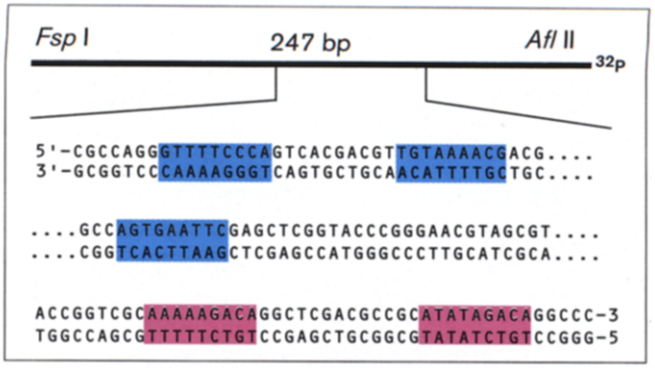
DNA-binding orientation

We carried out affinity cleavage experiments [1,20] using ImPyPy-γ-ImPyPy-β-PyPyPy-G-Dp-EDTA•Fe(II) (**2-E**•Fe(II)) and the 5'- or 3'-^{32}P end-labeled 247-bp pJT4 *Afl*II/*Fsp*I restriction fragment (Fig. 5). This polyamide selectively binds to the 5'-AAAAAGACA-3' and 5'-ATATAGACA-3' target sequences at subnanomolar concentration. A single 3'-shifted cleavage pattern is observed at each 9-bp site, indicating that the polyamide is bound in one orientation with the carboxyl terminus at the 5' end of the 5'-AAAAAGACA-3' or 5'-ATATAGACA-3' sequences (Fig. 6).

DNA-binding affinity and specificity

The exact locations and sizes of all binding sites were determined first by preliminary footprinting experiments using the cleavage reagent methidiumpropyl EDTA•Fe(II) [21] (data not shown). Quantitative DNase I footprint titration experiments [22–24] on the 3'-^{32}P-labeled 247-bp restriction fragment (in 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, pH 7.0, 22 °C) revealed that ImPyPy-γ-ImPyPy-β-PyPyPy-G-Dp specifically binds the 5'-AAAAAGACA-3' and 5'-ATATAGACA-3' target sequences with equilibrium association constants of $K_a \approx 2 \times 10^{10}$ M^{-1} and $K_a = 8 \times 10^9$ M^{-1}, respectively (Figs 7,8 and Table 1). The polyamide binds to additional sites on the restriction fragment with lower affinity. For comparison, the

Figure 5



Partial sequence of the restriction fragment used for footprinting and affinity cleaving experiments. The two match sites (pink) and three mismatch sites (blue) are highlighted.

Figure 6

Compound **2-E**•Fe(II) cleaves the 247-bp pJT4 *Afl*II/*Fsp*I restriction fragment at specific sites. **(a)** Storage phosphor autoradiogram of an 8 % denaturing polyacrylamide gel used to separate the fragments generated by cleavage reactions using polyamide **2-E**•Fe(II) on the 3'-^{32}P-end-labeled restriction fragment. Lanes 1 and 5: A-track sequencing lanes; lanes 2–4: cleavage products obtained in the presence of 0.1 nM, 1 nM, or 10 nM concentration of polyamide **2**•Fe(II), respectively; lane 6: intact DNA. **(b)** Schematic representation of the bound polyamide and cleavage sites in the affinity-cleavage experiment. **(c)** Observed cleavage intensities at the designated match sites at 100 pM and 0.1 nM concentrations of polyamide **2-E**•Fe(II). Arrow lengths are proportional to the amount of cleavage at the indicated base.

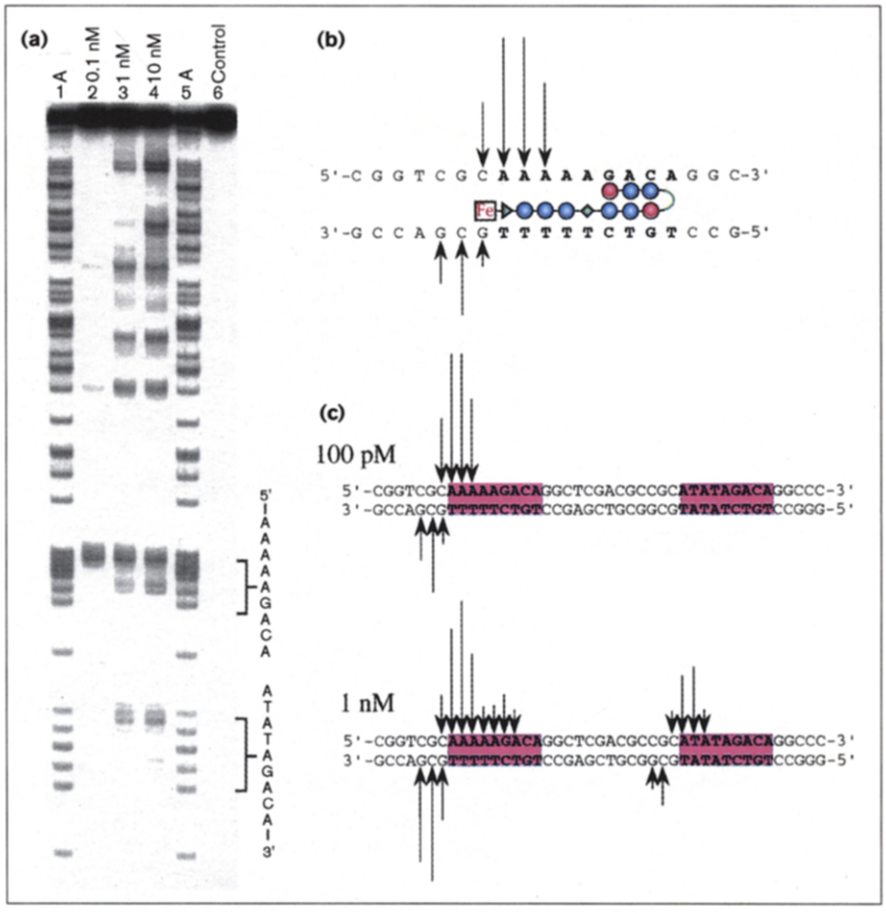
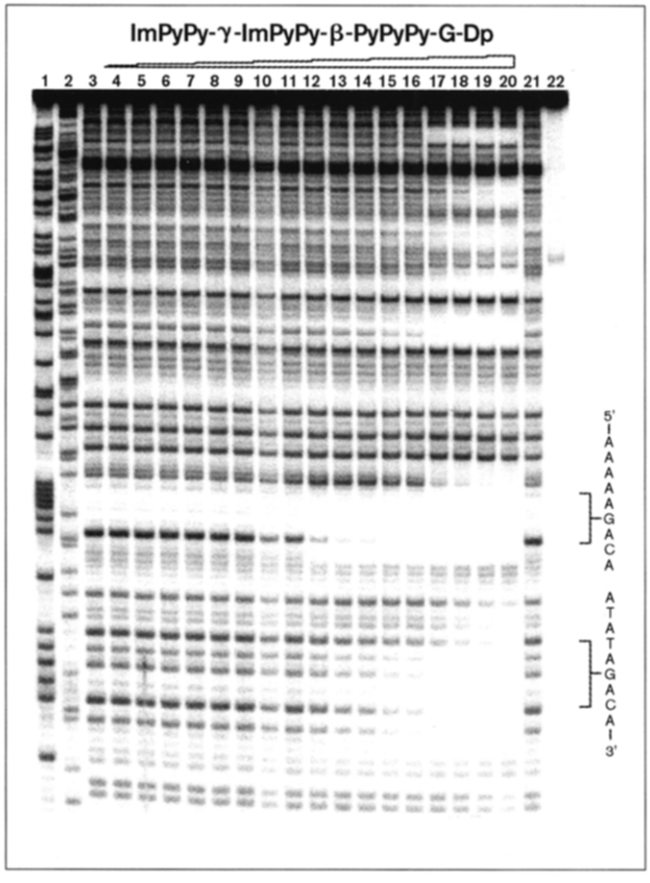


Figure 7



Polyamide **2** binds to specific DNA sequences with subnanomolar affinity. The storage phosphor autoradiogram is shown of an 8 % denaturing polyacrylamide gel used to separate the products of DNase I digestion in quantitative footprint titration experiments with polyamide **2** on the 247-bp pJT4 *Afl*II/*Fsp*I restriction fragment. Lanes 1–2: A- and G-track sequencing lanes; lanes 3 and 21: DNase I digestion products obtained in the absence of polyamide; lanes 4–20: DNase I digestion products obtained in the presence of 1 pM, 2 pM, 5 pM, 10 pM, 15 pM, 25 pM, 40 pM, 65 pM, 0.1 nM, 0.15 nM, 0.25 nM, 0.4 nM, 0.65 nM, 1 nM, 2 nM, 5 nM, and 10 nM polyamide; lane 22: intact DNA. All reactions were carried out in the presence of 15 000 cpm of restriction fragment, 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$ and 5 mM $CaCl_2$ at pH 7.0.

six-ring hairpin polyamide ImPyPy-γ-ImPyPy-β-Dp binds to the sequences 5'-aaaaAGACA-3' and 5'-atatAGACA-3' (lower-case letters are upstream of the recognition sequence) with association constants of $K_a = 5 \times 10^7\ M^{-1}$ and $K_a = 9 \times 10^7\ M^{-1}$, respectively.

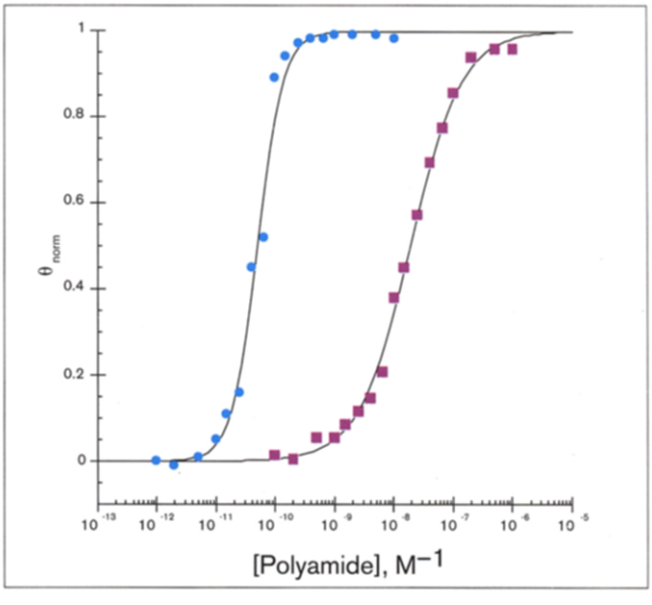
Relative to the six-ring polyamide ImPyPy-γ-ImPyPy-β-Dp, the nine-ring polyamide ImPyPy-γ-ImPyPy-β-PyPyPy-G-Dp binds to the sequences 5'-AAAAAGACA-3' and 5'-ATATAGACA-3' with ~400-fold and ~100-fold higher affinity, respectively. Similar binding enhancements have recently been reported in a separate system [25]. Addition of a carboxy-terminal PyPyPy subunit using a β-alanine linker is thus an effective strategy for increasing the DNA-binding affinity of hairpin polyamides that bind adjacent to an $(A,T)_4$ sequence.

Polyamide ImPyPy-γ-ImPyPy-β-PyPyPy-G-Dp binds to several mismatched sites present on the 247-bp restriction fragment with high affinity (Table 1). The two highest affinity mismatch sites, 5'-GAATTCACT-3' ($K_a = 4.5 \times 10^9\ M^{-1}$) and 5'-GTTTTCCCA-3' ($K_a = 2.5 \times 10^9\ M^{-1}$), are bound with at least five-fold reduced affinity relative to the optimal match site 5'-AAAAAGACA-3' (formally mismatched base pairs are underlined). This value may, however, be a lower limit due to the uncertainty in the very high equilibrium association constant for the optimal match site (see Materials and methods). In contrast, the six-ring polyamide ImPyPy-γ-ImPyPy-β-Dp binds more strongly to the match site 5'-AGACA-3' than to the single base-pair mismatch sites 5'-ATTCA-3' and 5'-TTACA-3' by a factor of 10. Although it remains a challenge to optimize the specificity of the extended hairpin polyamide motif, it is significant that a wholly designed, synthetic polyamide specifically binds to a designated nine base-pair site at subnanomolar concentrations.

Significance

Small molecules that specifically bind to any sequence in the human genome would be useful tools in molecular biology and potentially in human medicine. Pyrrole–imidazole polyamides bind sequence-specifically

Figure 8



The 'extended hairpin' polyamide **2** binds to the recognition sequence 5'-AAAAAGACA-3' with ~400-fold higher affinity than the 'hairpin' polyamide **1**. Data from quantitative DNase I footprinting experiments using the polyamides **1** (purple squares) and **2** (blue circles) are shown. Each data point shows the average value obtained from three footprinting experiments.

Table 1

Equilibrium association constants (M^{-1})

Binding site	Polyamide: ImPyPy-γ-ImPyPy-β-Dp	Polyamide: ImPyPy-γ-ImPyPy-β-PyPyPy-G-Dp
Match sites		
5′-AAAA**AGACA**-3′	5.2 x 10^7 (0.5)	2.0 x 10^{10} (0.4)
5′-ATAT**AGACA**-3′	9.1 x 10^7 (1.1)	8.1 x 10^9 (0.1)
Mismatch sites		
5′-G**AATTCA**CT-3′	≤8 x 10^6	4.5 x 10^9 (1.0)
5′-CGTT**TTACA**-3′	9.2 x 10^6 (1.6)	1.6 x 10^9 (0.3)
5′-GTTTTCCCA-3′	<10^6	2.5 x 10^9 (0.7)
5′-GGCGATTAAGTTG-3′	<10^6	8.9 x 10^8 (0.6)
5′-TCGCTATTACGCCA-3′	<10^6	1.5 x 10^9 (0.2)

Values reported are the mean values obtained from four DNase I footprint titration experiments. The standard deviation for each value is indicated in parentheses. 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 $CaCl_2$. The 5 base pair ImPyPy-γ-ImPyPy-β-Dp binding sites are in bold type. The portion of each site occupied by the carboxy-terminal PyPyPy subunit of ImPyPy-γ-ImPyPy-β-PyPyPy-G-Dp, as evidenced by affinity cleavage experiments with ImPyPy-γ-ImPyPy-β-PyPyPy-G-Dp-EDTA•Fe(II), is underlined.

to DNA by forming side-by-side complexes in the minor groove. A simple code has been developed to rationally alter the sequence-specificity of these minor-groove-binding molecules, and has guided the design of polyamides that specifically target a wide variety of DNA sequences.

The results described here demonstrate that two simple amino acids, γ-aminobutyric acid and β-alanine, optimally connect polyamide subunits in different and predictable conformations. A designed three-subunit polyamide incorporating both linkers specifically binds to a designated nine-base-pair target site at subnanomolar concentrations. The high binding affinity and sequence-specificity of pyrrole–imidazole polyamides, coupled with a broad targetable sequence repertoire and an efficient solid-phase synthesis methodology, provide a powerful class of non-natural molecules for sequence-specific recognition of double helical DNA.

Materials and methods

NMR spectra were recorded on a GE 300 instrument operating at 300 MHz (^1H) or 75 MHz (^{13}C). Spectra were recorded in DMSO-d_6 with chemical shifts reported in parts per million relative to residual DMSO-d_5. UV spectra were measured on a Hewlett-Packard Model 8452A diode array spectrophotometer. Matrix-assisted, laser desorption/ionization time of flight mass spectrometry (MALDI-TOF MS) was carried out at the Protein and Peptide Microanalytical Facility at the California Institute of Technology. HPLC analysis was performed either on a HP 1090 M analytical HPLC or a Beckman Gold system using a Rainen C18, Microsorb MV, 5 μm, 300 x 4.6 mm reversed-phase column in 0.1 % (w/v) trifluoroacetic acid (TFA) with acetonitrile as eluent and a flow rate of 1.0 ml min^{-1}, gradient elution 1.25 % acetonitrile min^{-1}. Preparative HPLC was carried out on a Beckman instrument using a Waters DeltaPak 25 x 100 mm 100 μm C_{18} column in 0.1 % (w/v) TFA, gradient elution 0.25 % CH_3CN min^{-1}. Water was obtained from a Millipore Milli-Q water purification system. Reagent-grade chemicals were used unless otherwise stated. *Escherichia coli* XL-1 Blue competent cells were obtained from Stratagene. Restriction endonucleases were purchased from either New England Biolabs or Boehringer-Mannheim. Sequenase (version 2.0) was obtained from United States Biochemical, and DNase I (FPLC pure) was obtained from Pharmacia. Stocks of [α-^{32}P]-thymidine-5′-triphosphate (≥ 3000 Ci $mmol^{-1}$), [α-^{32}P]-deoxyadenosine-5′-triphosphate (≥ 6000 Ci $mmol^{-1}$), and [γ-^{32}P]-adenosine-5′-triphosphate were purchased from Du Pont/NEN.

*ImPyPy-γ-ImPyPy-β-Dp (compound **1**)*

Polyamide **1** was prepared by machine-assisted solid phase methods as a white powder (17 mg, 56 % recovery). HPLC room temperature (RT): 26.1 min; UV λ_{max} (ε): 234 nm (39 300), 312 nm (53 200); ^1H NMR (DMSO-d_6): δ 10.53 (s, 1 H), 10.27 (s, 1 H), 10.04 (s, 1 H), 9.96 (s, 1 H), 9.94 (s, 1 H), 9.2 (br s, 1 H), 8.08 (m, 3 H), 7.49 (s, 2 H), 7.44 (s, 1 H), 7.31 (d, 1 H, J = 1.0 Hz), 7.23 (d, 1 H, J = 1.1 Hz), 7.19 (m, 3 H), 7.10 (s, 1 H), 6.92 (d, 1 H, J = 1.1 Hz), 6.90 (d, 1 H, J = 1.1 Hz). 4.01 (s, 3 H), 3.97 (s, 3 H), 3.86 (m, 6 H), 3.82 (m, 6 H), 3.41 (q, 2 H, J = 6.0 Hz), 3.22 (q, 2 H, J = 5.9 Hz), 3.13 (q, 2 H, J = 5.9 Hz), 3.0 (q, 2 H, J = 5.6 Hz), 2.76 (d, 6 H, J = 4.8 Hz), 2.37 (m, 4 H), 1.78 (m, 4 H); MALDI-TOF MS: 979.3 (979.1 calc. for M+H).

*ImPyPy-γ-ImPyPy-β-PyPyPy-G-Dp (compound **2**)*

Polyamide **2** was prepared by machine-assisted solid phase methods as a white powder (12 mg, 19 % recovery). HPLC RT: 29.5 min; UV λ_{max}(ε): 238 nm (53 900), 312 nm (71 100); ^1H NMR (DMSO-d_6): δ 10.46 (s, 1 H), 10.24 (s, 1 H), 9.96 (s, 1 H), 9.90 (m, 5 H), 9.2 (br s, 1 H), 8.25 (m, 1 H), 8.00 (m, 3 H), 7.44 (s, 1 H), 7.39 (s, 1 H), 7.26 (d, 1 H, J = 1.3 Hz), 7.24 (d, 1 H, J = 1.5 Hz), 7.20 (m, 2 H), 7.16 (m, 2 H), 7.13 (m, 2 H), 7.11 (d, 1 H, J = 1.4 Hz), 7.05 (d, 1 H, J = 1.4 Hz), 7.03 (d, 1 H, J = 1.5 Hz), 6.93 (d, 1 H, J = 1.3 Hz), 6.87 (m, 2 H), 6.84 (d, 1 H, J = 1.5 Hz), 3.97 (s, 3 H), 3.92 (s, 3 H), 3.82 (m, 9 H), 3.79 (m, 6 H), 3.76 (m, 6 H), 3.73 (m, 2 H), 3.44 (q, 2 H, J = 5.0 Hz), 3.17 (m, 4 H), 3.03 (m, 2 H), 2.74 (d, 6 H, J = 4.8 Hz), 2.50 (m, 2 H) 2.33 (t, 2 H, J = 6.7 Hz), 1.77 (m, 4 H); MALDI-TOF MS: 1402.2 (1402.5 calc. for M+H).

*ImPyPy-γ-ImPyPy-β-PyPyPy-G-Dp-NH_2 (compound **2-NH_2**)*

Polyamide **2-NH_2** was prepared by machine-assisted solid phase methods as a white powder (29 mg, 59 % recovery). HPLC RT: 21.5 min; ^1H NMR (DMSO-d_6): δ 10.50 (s, 1 H), 10.27 (s, 1 H), 9.96 (s, 1 H), 9.93 (m, 5 H), 9.2 (br s, 1 H), 8.27 (t, 1 H, J = 5.1 Hz), 8.03 (m, 3 H), 7.90 (s, 3 H), 7.45 (s, 1 H), 7.40 (s, 1 H), 7.27 (d, 1 H, J = 1.3 Hz), 7.25 (d, 1 H, J = 1.4 Hz), 7.22 (m, 2 H), 7.18 (m, 2 H), 7.17 (d, 1 H, J = 1.4 Hz), 7.14 (d, 1 H, J = 1.3 Hz), 7.11 (m, 2 H), 7.06 (d, 1 H, J = 1.5 Hz), 6.94 (d, 1 H, J = 1.3 Hz), 6.88 (m, 2 H), 6.84 (d, 1 H, J = 1.4 Hz), 3.97 (s, 3 H), 3.93 (s, 3 H), 3.83 (m, 9 H), 3.80 (m, 6 H), 3.76 (m, 6 H), 3.72 (d, 2 H, J = 5.2 Hz), 3.43 (q, 2 H, J = 5.0 Hz), 3.17 (m, 6 H), 3.11 (q, 2 H, J = 5.3 Hz), 2.85 (q, 2 H, J = 5.2 Hz), 2.73 (d, 3 H, J = 3.9 Hz), 2.51 (t, 2 H, J = 6.5 Hz), 2.35 (t, 2 H, J = 6.7 Hz), 1.92 (quintet, 2 H, J = 6.8 Hz), 1.78 (m, 4 H); MALDI-TOF MS: 1445.6 (1445.6 calc. for M+H).

*ImPyPy-γ-ImPyPy-β-PyPyPy-G-Dp-EDTA (compound **2-E**)*

EDTA-dianhydride (50 mg) was dissolved in 1 ml DMSO/NMP solution and 1 ml DIEA by heating at 55 °C for 5 min. The dianhydride solution was added to ImPyPy-γ-ImPyPy-β-PyPyPy-G-Dp-NH_2 (9.0 mg, 5 μmol) dissolved in 750 μl DMSO. The mixture was heated at 55 °C for 25 min, treated with 3 ml 0.1 M NaOH, and heated at 55 °C for 10 min. 0.1 % TFA was added to adjust the total volume to 8 ml and the solution

purified directly by reversed-phase HPLC to provide ImPyPy-γ-ImPyPy-β-PyPyPy-G-Dp-EDTA as a white powder (3 mg, 30 % recovery after HPLC purification). MALDI-TOF MS: 1720.1 (1719.8 calc. for M+H).

Preparation of ^{32}P-labeled DNA

Plasmid pJT4 was prepared by hybridizing two sets of 5'-phosphorylated, complementary oligonucleotides, 5'-CCGGGAACGTAGCGTACCGGTCGCAAAAAGACAGGCTCGA-3' and 5'-GGCGTCGAGCCTGTCTTTTTGCGACCGGTACGCTACGTTC-3', and 5'-CGCCGCATATAGACAGGCCCAGCTGCGTCCTAGCTAGCGTCGTAGCGTCTTAAGAG-3' and 5'-TCGACTCTTAAGACGCTACGACGCTAGCTAGGACGCAGCTGGGCCTGTCTATATGC-3', and ligating the resulting duplexes to the large pUC19 *AvaI*/*SalI* restriction fragment. The 3'-^{32}P end-labeled *AflII*/ *FspI* fragment was prepared by digesting the plasmid with *AflII* and simultaneously filling in using Sequenase enzyme, [α-^{32}P]-deoxyadenosine-5'-triphosphate, and [α-^{32}P]-thymidine-5'-triphosphate, digesting with *FspI*, and isolating the 247-bp fragment by nondenaturing gel electrophoresis. The 5'-^{32}P-end-labeled *AflII*/*FspI* fragment was prepared using standard methods. A and G sequencing were carried out as described [26,27]. Standard methods were used for all DNA manipulations [28].

Affinity-cleavage reactions

All reactions were executed in a total volume of 400 μl. A stock solution of compound **2-E** or H_2O was added to a solution containing labeled restriction fragment (15 000 cpm), affording final solution conditions of 20 mM HEPES, 200 mM NaCl, 50 μg ml^{-1} glycogen, and pH 7.3. Subsequently, 20 μl of freshly prepared 20 mM $Fe(NH_4)_2(SO_4)_2$ was added and the solution allowed to equilibrate for 20 min. Cleavage reactions were initiated by the addition of 40 μl of 50 mM dithiothreitol, allowed to proceed for 12 min at 22 °C, then stopped by the addition of 1 ml of ethanol. Reactions were precipitated and the cleavage products separated using standard methods [28]. Next, 10 μl of a solution containing calf thymus DNA (140 μM base-pair) (Pharmacia) and glycogen (2.8 mg ml^{-1}) was added, and the DNA precipitated. The reactions were resuspended in 1 x TBE/80 % formamide loading buffer, denatured by heating at 85 °C for 10 min, and placed on ice. The reaction products were separated by electrophoresis on an 8 % polyacrylamide gel (5 % cross-link, 7 M urea) in 1 x TBE at 2000 V. Gels were dried and exposed to a storage phosphor screen (Molecular Dynamics). Relative cleavage intensities were determined by volume integration of individual cleavage bands using ImageQuant software.

Quantitative DNase I footprint titration experiments

All reactions were executed in a total volume of 400 μl. A polyamide stock solution or H_2O (for reference lanes) was added to an assay buffer containing radiolabeled restriction fragment (15 000 cpm), affording final solution conditions of 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, pH 7.0, and either 1 pM–10 nM polyamide or no polyamide (for reference lanes). The solutions were allowed to equilibrate at 22 °C for 12 h for polyamide **1** or 36 h for polyamide **2**. Footprinting reactions were initiated by the addition of 10 μl of a DNase I stock solution (at the appropriate concentration to give ~55% intact DNA) containing 1 mM dithiothreitol and allowed to proceed for 7 min at 22 °C. The reactions were stopped by the addition of 50 μl of a solution containing 2.25 M NaCl, 150 mM EDTA, 0.6 mg ml^{-1} glycogen, and 30 μM base pair calf thymus DNA, and ethanol precipitated. Reactions were resuspended in 1 x TBE/80 % formamide loading buffer, denatured by heating at 85 °C for 10 min, and placed on ice. The reaction products were separated by electrophoresis on an 8 % polyacrylamide gel (5 % cross-link, 7 M urea) in 1 x TBE at 2000 V. Gels were dried and exposed to a storage phosphor screen (Molecular Dynamics).

Quantitation and data analysis

Data from the footprint titration gels were obtained using a Molecular Dynamics 400S PhosphorImager followed by quantitation using ImageQuant software (Molecular Dynamics). Background-corrected volume integration of rectangles encompassing the footprint sites and a reference site at which DNase I reactivity was invariant across the titration generated values for the site intensities (I_{site}) and the reference intensity (I_{ref}). The apparent fractional occupancy (θ_{app}) of the sites were calculated using the equation:

$$\theta_{app} = 1 - \frac{I_{site} / I_{ref}}{I^{o}_{site} / I^{o}_{ref}} \quad (1)$$

where I^{o}_{site} and I^{o}_{ref} are the site and reference intensities, respectively, from a control lane to which no polyamide was added. The ($[L]_{tot}$, θ_{app}) data points were fit to a general Hill equation (eq. 2) by minimizing the difference between θ_{app} and θ_{fit}:

$$\theta_{app} = \theta_{min} + (\theta_{max} - \theta_{min}) \frac{K_a^n [L]^n_{tot}}{1 + K_a^n [L]^n_{tot}} \quad (2)$$

where $[L]_{tot}$ is the total polyamide concentration, K_a is the equilibrium association constant, and θ_{min} and θ_{max} are the experimentally determined site saturation values when the site is unoccupied or saturated, respectively. The data were fit using a nonlinear least-squares fitting procedure with K_a, θ_{max}, and θ_{min} as the adjustable parameters. For polyamide ImPyPy-γ-ImPyPy-β-Dp, binding isotherms for the 5'-AGACA-3' target sites were adequately fit by Langmuir isotherms (eq. 2, n = 1), consistent with formation of 1:1 polyamide-DNA complexes. For ImPyPy-γ-ImPyPy-β-PyPyPy-G-Dp, steeper binding isotherms (eq. 2, n = 1.8–2.2) were observed at the target sites 5'-AAAAAGACA-3' and 5'-ATATAGACA-3'. We believe that the steepness of these isotherms is due to the very high equilibrium association constants at these sites, and emphasize that treatment of the data in this manner does not represent an attempt to model a binding mechanism. Rather, we have chosen to compare values of the apparent equilibrium association constant, a value that represents the concentration of ligand at which a site is half-saturated. The binding isotherms were normalized using the following equation:

$$\theta_{norm} = \frac{\theta_{app} - \theta_{min}}{\theta_{max} - \theta_{min}} \quad (3)$$

Four sets of data were used in determining each association constant. The method for determining association constants used here involves the assumption that $[L]_{tot} \approx [L]_{free}$, where $[L]_{free}$ is the concentration of polyamide free in solution (unbound). For very high association constants this assumption becomes invalid, resulting in underestimated association constants. In the experiments described here, the DNA concentration is estimated to be ~5 pM. As a consequence, apparent association constants greater than ~10^{10} M^{-1} should be regarded as lower limits.

Acknowledgements

We are grateful to the National Institutes of Health (Grant GM-27681) for research support, to the National Science Foundation for a predoctoral fellowship to J.W.T., and to the Howard Hughes Medical Institute for a predoctoral fellowship to E.E.B.

References

1. Wade, W.S., Mrksich, M. & Dervan, P.B. (1992). Design of peptides that bind in the minor groove of DNA at 5'-(A,T)G(A,T)C(A,T)-3' sequences by a dimeric side-by-side motif. *J. Am. Chem. Soc.* **114**, 8783–8794.
2. Mrksich, M., Wade, W.S., Dwyer, T.J., Geierstanger, B.H., Wemmer, D.E. & Dervan, P.B (1992). Antiparallel side-by-side motif for sequence-specific recognition in the minor groove of DNA by the designed peptide 1-methylimidazole-2-carboxamide netropsin. *Proc. Natl. Acad. Sci. USA* **89**, 7586–7590.
3. Wade, W.S., Mrksich, M. & Dervan, P.B. (1993). Binding affinities of

synthetic peptides, pyridine-2-carboxamidenetropsin and 1-methylimidazole-2-carboxamidenetropsin, that form 2:1 complexes in the minor groove of double-helical DNA. *Biochemistry* **32**, 11385–11389.
4. Pelton, J.G. & Wemmer, D.E. (1989). Structural characterization of a 2-1 distamycin A•d(CGCAAATTTGGC) complex by two-dimensional NMR. *Proc. Natl. Acad. Sci. USA* **86**, 5723–5727.
5. Pelton, J.G. & Wemmer, D.E. (1990). Binding modes of distamycin-A with $d(CGCAAATTTGCG)_2$ determined by 2-dimensional NMR. *J. Am. Chem. Soc.* **112**, 1393–1399.
6. Chen, X., Ramakrishnan, B., Rao, S.T. & Sundaralingam, M. N. (1994). Binding of two distamycin-A molecules in the minor groove of an alternating B-DNA duplex. *Nat. Struct. Biol.* **1**, 169–175.
7. Mrksich, M. & Dervan, P.B. (1993). Antiparallel side-by-side heterodimer for sequence-specific recognition in the minor groove of DNA by a distamycin 1-methylimidazole-2-carboxamide pair. *J. Am. Chem. Soc.* **115**, 2572–2576.
8. Geierstanger, B.H., Jacobsen, J-P., Mrksich, M., Dervan, P.B. & Wemmer, D.E. (1994). Structural and dynamic characterization of the heterodimeric and homodimeric complexes of distamycin and 1-methylimidazole-2-carboxamide-netropsin bound to the minor-groove of DNA. *Biochemistry* **33**, 3055–3062.
9. Geierstanger, B.H., Dwyer, T.J., Bathini, Y., Lown, J.W. & Wemmer, D.E. (1993). NMR characterization of a heterocomplex formed by distamycin and its analog 2-ImD with d(CGCAAGTTGGC)–d(GCC-AACTTGCG) – preference for the 1/1/1 2-ImD–Dst–DNA complex over the 2/1 2-ImD–DNA and the 2/1 Dst–DNA complexes. *J. Am. Chem. Soc.* **115**, 4474–4482.
10. Geierstanger, B.H., Mrksich, M., Dervan, P.B. & Wemmer, D.E. (1994). Design of a G•C-specific minor groove-binding peptide. *Science* **266**, 646–650.
11. Mrksich, M. & Dervan, P.B. (1995). Recognition in the minor-groove of DNA at 5'-(A,T)GCGC(A,T)-3' by a four-ring tripeptide dimer – reversal of the specificity of the natural product distamycin. *J. Am. Chem. Soc.* **117**, 3325–3332.
12. Mrksich, M. & Dervan, P.B. (1993). Enhanced sequence-specific recognition in the minor-groove of DNA by covalent peptide dimers – bis(pyridine-2-carboxamidonetropsin)$(CH_2)_{3-6}$. *J. Am. Chem. Soc.* **115**, 9892–9899.
13. Dwyer, T.J., Geirstanger, B.H., Mrksich, M., Dervan, P.B. & Wemmer, D.E. (1993). Structural analysis of covalent peptide dimers, bis(pyridine-2-carboxamidonetropsin)$(CH_2)_{3-6}$, in complex with 5'-TGACT-3' sites by two-dimensional NMR. *J. Am. Chem. Soc.* **115**, 9900–9906.
14. Mrksich, M. & Dervan, P.B. (1994). Design of a covalent peptide heterodimer for sequence-specific recognition in the minor-groove of double-helical DNA. *J. Am. Chem. Soc.* **116**, 3663–3664.
15. Mrksich, M., Parks, M.E. & Dervan, P.B. (1994). Hairpin peptide motif – a new class of oligopeptides for sequence-specific recognition in the minor-groove of double-helical DNA. *J. Am. Chem. Soc.* **116**, 7983–7988.
16. Parks, M.E., Baird, E.E. & Dervan, P.B. Optimization of the hairpin polyamide design for recognition of the minor groove of DNA. *J. Am. Chem. Soc.*, in press.
17. Trauger, J.W., Baird, E.E. & Dervan, P.B. Extension of sequence-specific recognition in the minor groove of DNA to 9-13 base-pairs. *J. Am. Chem. Soc.*, in press.
18. Geierstanger, B.H., Mrksich, M., Dervan, P.B. & Wemmer, D.E. (1996). Extending the recognition site of designed minor groove binding molecules. *Nat. Struct. Biol.* **3**, 321–324.
19. Baird, E.E. & Dervan, P.B. Solid-phase synthesis of polyamides containing imidazole and pyrrole amino acids. *J. Am. Chem. Soc.*, in press.
20. Schultz, P.G., Taylor, J.S. & Dervan, P.B. (1982). Design and synthesis of a sequence-specific DNA cleaving molecule – (distamycin-EDTA)iron(II). *J. Am. Chem. Soc.* **104**, 6861–6863.
21. Hertzberg, R.P. & Dervan, P.B. (1982). Cleavage of double-helical DNA by (methidiumpropyl-EDTA)iron(II). *J. Am. Chem. Soc.* **104**, 313–315.
22. Fox, K.R. & Waring, M.J. (1984). DNA structural variations produced by actinomycin and distamycin as revealed by DNase-I footprinting. *Nucleic Acids Res.* **12**, 9271–9285.
23. Brenowitz, M., Senear, D.F., Shea, M.A. & Ackers, G.K. (1986). Quantitative DNase footprint titration – a method for studying protein–DNA interactions. *Methods Enzymol.* **130**, 132–181.
24. Brenowitz, M., Senear, D.F., Shea, M.A. & Ackers, G.K. (1986). Footprint titrations yield valid thermodynamic isotherms. *Proc. Natl. Acad. Sci. USA* **83**, 8462–8466.
25. Nicolaou, K.C., Smith, B.M., Ajito, K., Komatsu, H., Gomez-Paloma, L. & Tor, Y. (1996). DNA–carbohydrate interactions. Specific binding of head-to-head and head-to-tail dimers of the calicheamicin oligosaccharide to duplex DNA. *J. Am. Chem. Soc.* **118**, 2303–2304.
26. Maxam, A.M. & Gilbert, W.S. (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* **65**, 499–560.
27. Iverson, B.L. & Dervan, P.B. (1987). Adenine-specific DNA chemical sequencing reaction. *Nucleic Acids Res.* **15**, 7823–7830.
28. Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989). Molecular Cloning. (2nd edn). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.