On the pairing rules for recognition in the minor groove of DNA by pyrrole–imidazole polyamides

Sarah White, Eldon E Baird and Peter B Dervan

Background: Cell-permeable small molecules that target predetermined DNA sequences with high affinity and specificity have the potential to control gene expression. A binary code has been developed to correlate DNA sequence with side-by-side pairings between *N*-methylpyrrole (Py) and *N*-methylimidazole (Im) carboxamides in the DNA minor groove. We set out to determine the relative energetics of pairings of Im/Py, Py/Im, Im/Im, and Py/Py for targeting G·C and A·T base pairs. A key specificity issue, which has not been previously addressed, is whether an Im/Im pair is energetically equivalent to an Im/Py pair for targeting G·C base pairs.

Results: Equilibrium association constants were determined at two five-base-pair sites for a series of four six-ring hairpin polyamides, in order to test the relative energetics of the four aromatic amino-acid pairings opposite G·C and A·T base pairs in the central position. We observed that a G·C base pair was effectively targeted with Im/Py but not Py/Im, Py/Py, or Im/Im. The A·T base pair was effectively targeted with Py/Py but not Im/Py, Py/Im, or Im/Im.

Conclusions: An Im/Im pairing is energetically disfavored for the recognition of both A·T and G·C. This specificity will create important limitations on undesirable slipped motifs that are available for unlinked dimers in the minor groove. Baseline energetic parameters will thus be created which, using the predictability of the current pairing rules for specific molecular recognition of double-helical DNA, will guide further second-generation polyamide design for DNA recognition.

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

Correspondence: Peter B Dervan
E-mail: dervan@cco.caltech.edu

Key words: binary code, gene regulation, hydrogen bond, molecular recognition

Received: **6 June 1997**
Accepted: **1 July 1997**

Chemistry & Biology August 1997, 4:569–578
http://biomednet.com/elecref/1074552100400569

Introduction

Crescent-shaped polyamides containing *N*-methylpyrrole (Py) and *N*-methylimidazole (Im) amino acids bind cooperatively as antiparallel dimers in the minor groove of the DNA helix [1]. Their sequence-specificity depends on the side-by-side pairings of the *N*-methylpyrrole and *N*-methylimidazole amino acids. A pairing of Im opposite Py (Im/Py) targets a G·C base pair, while Py/Im targets C·G (see Figure 1) [1–5]. A Py/Py pairing is degenerate and targets both T·A and A·T base pairs [1–9]. Pyrrole–imidazole polyamides have been shown to be cell permeable and to inhibit the transcription of genes in cell culture [10]. This provides impetus to explore the scope and limitations of their use for DNA recognition, particularly the energetics and structural details of their remarkably simple binary code.

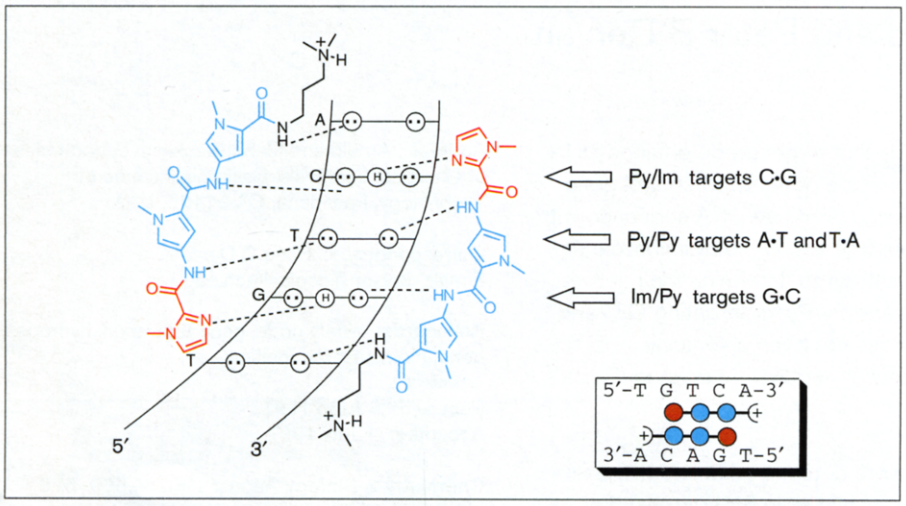
Recognition of a G·C base pair in the DNA minor groove

Recognition of a G·C base pair by the Im/Py pairing requires precise positioning for the key hydrogen bond between the Im N3 and the exocyclic amine of guanine [11,12]. Given the central location of the guanine exocyclic amine group in the DNA minor groove [13–15], the question arises of whether an Im/Im pairing might also target a G·C base pair [16]. Remarkably, even in the first report on the binding specificity of the three-ring polyamide homodimer $(Im–Py–Py–Dp)_2$ (Dp: dimethylaminopropylamide), there were qualitative data to suggest that there was indeed a binding preference for placing the Im/Py pair opposite the G·C base pair [1] (Figure 1). It would be useful to determine the generality of the aromatic amino acid pairing preferences and to compare the relative energetics of the four possible pairings of Im and Py for the recognition of G·C and T·A base pairs. Here, we describe an experimental design that uses an oriented six-ring hairpin polyamide to allow the relative energetic preferences of four different binary combinations (Im/Py, Py/Im, Im/Im, and Py/Py) to be tested opposite a G·C or an A·T base pair (Figure 2).

The hairpin polyamide motif

Three-ring polyamide subunits covalently coupled by a γ-aminobutyric acid linker (γ) form hairpin structures with five-base-pair target DNA sequences. [17–20]. For example, according to the pairing rules, polyamides with the sequences Im–*Im*–Py-γ-Py–*Py*–Py-β (β: β-alanine) and Im–*Py*–Py-γ-Py–*Py*–Py-β (central amino acid pairing italicized) would be expected to bind to 5′-TG*G*TA-3′

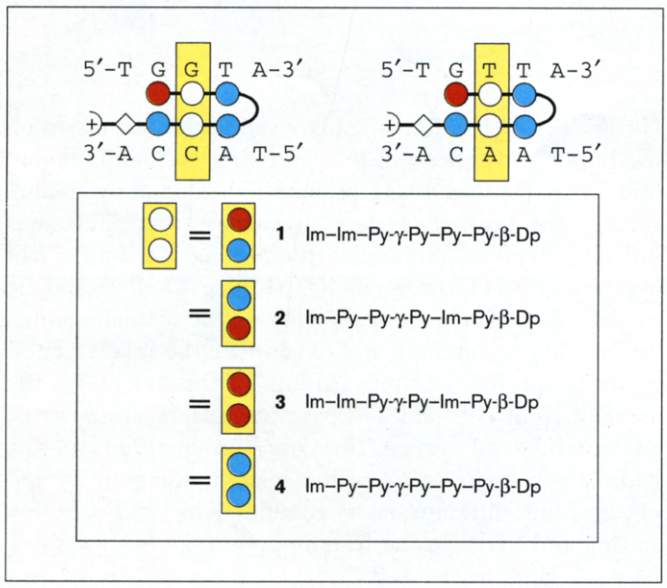
Figure 1



Binding models for antiparallel three-ring polyamide subunits, Im–Py–Py-Dp (Py, *N*-methylpyrrole, blue; Im, *N*-methylimidazole, red; Dp, dimethylaminopropylamide), in complex with 5′-TGTCA-3′. Circles with two dots represent the lone pairs of the N3 atoms of purines and the O2 atoms of pyrimidines. Circles containing an H represent the N2 hydrogens of guanines. Putative hydrogen bonds are illustrated by dotted lines. In the schematic binding model (inset), the Py and Im rings are represented as blue and red circles, respectively.

and 5′-TG*T*TA-3′ (central base italicized) sequences, respectively. Selective substitution of the central amino acid of each three-ring polyamide subunit allows four possible ring pairings at a unique central location within the hairpin structure that can be placed opposite a G·C or T·A base pair (Figure 2).

Figure 2



Models of the expected hairpin complexes of Im–Im–Py-γ-Py–Py–Py-β-Dp (β: β-alanine; **1**), Im–Py–Py-γ-Py–Im–Py-β-Dp (**2**), Im–Im–Py-γ-Py–Im–Py-β-Dp (**3**), and Im–Py–Py-γ-Py–Py–Py-β-Dp (**4**) complexed with 5′-TGGTA-3′ and 5′-TG*T*TA-3′. Unfilled white circles may be either Py or Im; and blue and red circles, are Py and Im, respectively. Diamonds represent β-alanine and the curved line joining the two polyamide subunits represents the γ-aminobutyric acid linker. The central pairings of Im/Py, Py/Im, Im/Im and Py/Py with G·C and T·A are highlighted with yellow boxes.

Four six-ring polyamides (Figure 3), Im–Im–Py-γ-Py–Py–Py-β-Dp (**1**), Im–Py–Py-γ-Py–Im–Py-β-Dp (**2**), Im–Im–Py-γ-Py–Im–Py-β-Dp (**3**), and Im–Py–Py-γ-Py–Py–Py-β-Dp (**4**), containing central amino pairings of Im/Py, Py/Im, Im/Im, and Py/Py in a hairpin structure were synthesized by solid phase methods [21]. The corresponding EDTA analogs (Figure 3), Im–Im–Py-γ-Py–Py–Py-β-Dp-EDTA (**1-E**), Im–Py–Py-γ-Py–Im–Py-β-Dp-EDTA (**2-E**), Im–Im–Py-γ-Py–Im–Py-β-Dp-EDTA (**3-E**), and Im–Py–Py-γ-Py–Py–Py-β-Dp-EDTA (**4-E**), were also constructed, in order to confirm the single binding orientation of each hairpin–DNA complex.

Here, we report the DNA-binding affinities, orientations, and sequence-selectivity of the four polyamides for two five-base-pair binding sites, 5′-TG*T*TA-3′ and 5′-TG*G*TA-3′, which vary at one unique third position (italicized). Three separate techniques were used to characterize the DNA-binding properties of the polyamides: affinity cleavage [22,23], MPE·Fe(II) footprinting (MPE: methidiumpropyl-EDTA) [24,25], and DNase I footprinting [26–28]. Affinity cleavage studies were used to determine the specific binding orientation and stoichiometry of each hairpin–DNA complex. The binding location and site size were determined accurately by MPE·Fe(II) footprinting, and quantitative DNase I footprint titration was used for measuring the equilibrium association constants (K_a) for the binding of the polyamide to the designated sequences.

Results and discussion

Binding-site size by MPE·Fe(II) footprinting

MPE·Fe(II) footprinting on 3′ and 5′ ^{32}P end-labeled 302-base-pair restriction fragments (see the Materials and methods section) revealed that the polyamides bind and discriminate the two five-base-pair sites, 5′-TG*T*TA-3′ and 5′-TG*G*TA-3′ (Figure 4). At micromolar concentrations,

polyamides **1** and **3** (Figure 3) bound to 5′-TG*G*TA-3′ > 5′-TG*T*TA-3′, whereas polyamides **2** and **4** (Figure 3) bound to 5′-TG*T*TA-3′ > 5′-TG*G*TA-3′.

Binding orientation by affinity cleavage

Affinity cleavage experiments using hairpin polyamides modified with EDTA·Fe(II) at the carboxyl terminus were used to determine the orientation and stoichiometry of polyamide binding. Experiments were performed on the same 3′ and 5′ ^{32}P end-labeled 302-base-pair restriction fragments (Figure 5; see the Materials and methods section). In all cases, the observed cleavage patterns were 3′ shifted, consistent with minor-groove occupancy [22]. A single cleavage locus proximal to the 5′ side of both the 5′-TG*T*TA-3′ and 5′-TG*G*TA-3′ binding sites confirmed that the four polyamides bound each discreet site with a single orientation. The observation of a single cleavage locus is consistent only with an oriented 1:1 complex in the minor groove of DNA and rules out dimeric overlapped or extended binding motifs [29]. A 1:1 oriented but extended motif would require at least an 8-base-pair binding site, which is inconsistent with the high-resolution MPE footprinting data on both target sites. The hairpin complex Im–Py–Py-γ-Py–Py–Py-Dp·5′-TGTTA-3′ has recently been characterized by direct nuclear magnetic resonance methods [30].

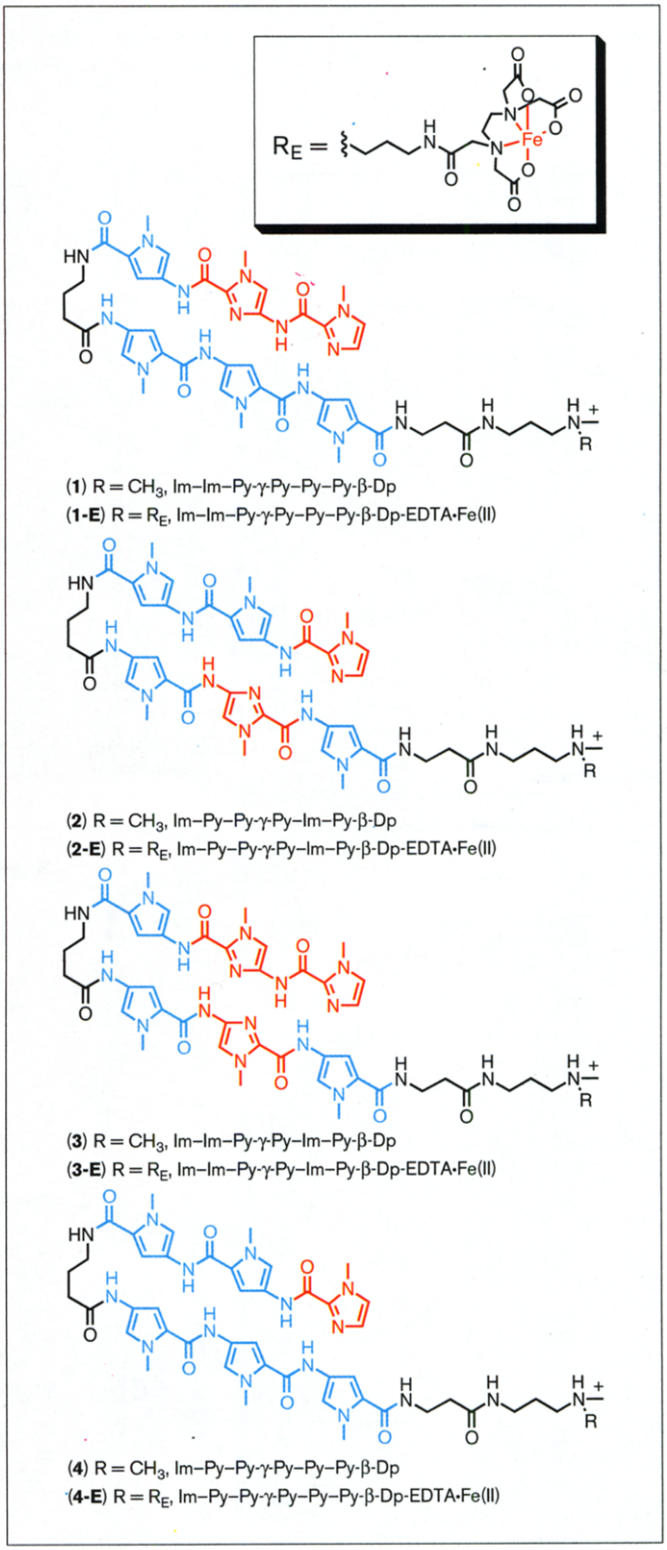
Energetics by quantitative DNase I footprint titrations

MPE·Fe(II) footprinting combined with affinity cleavage experiments indicated that each polyamide bound the designated five-base-pair target site as a 1:1 hairpin complex in the minor groove. This single and consistent mode of binding allowed a valid thermodynamic comparison for the central-ring amino acid pairing of each polyamide that recognized the central base pair of each designated target site. Quantitative DNase I footprint titration experiments [24–26] (see the Materials and methods section) were performed to determine the equilibrium association constants for the bound sites (Figure 6). The 5′-TG*G*TA-3′ site was bound by the polyamides with decreasing affinity: Im–*Im*–Py-γ-Py–*Py*–Py-β-Dp (**1**) >> Im–*Py*–Py-γ-Py–*Im*–Py-β-Dp (**2**) ≈ Im–*Im*–Py-γ-Py–*Im*–Py-β-Dp (**3**) ≈ Im–*Py*–Py-γ-Py–*Py*–Py-β-Dp (**4**). The 5′-TG*T*TA-3′ site was bound with decreasing affinity: Im–*Py*–Py-γ-Py–*Py*–Py-β-Dp (**4**) > Im–*Py*–Py-γ-Py–*Im*–Py-β-Dp (**2**) ≈ Im–*Im*–Py-γ-Py–*Py*–Py-β-Dp (**1**) > Im–*Im*–Py-γ-Py–*Im*–Py-β-Dp (**3**). Remarkably, the association constant for each site varied 100-fold between the four polyamides, indicating a sensitivity to a single atomic substitution within the central-ring amino acids (Figure 7).

The Im/Py pair

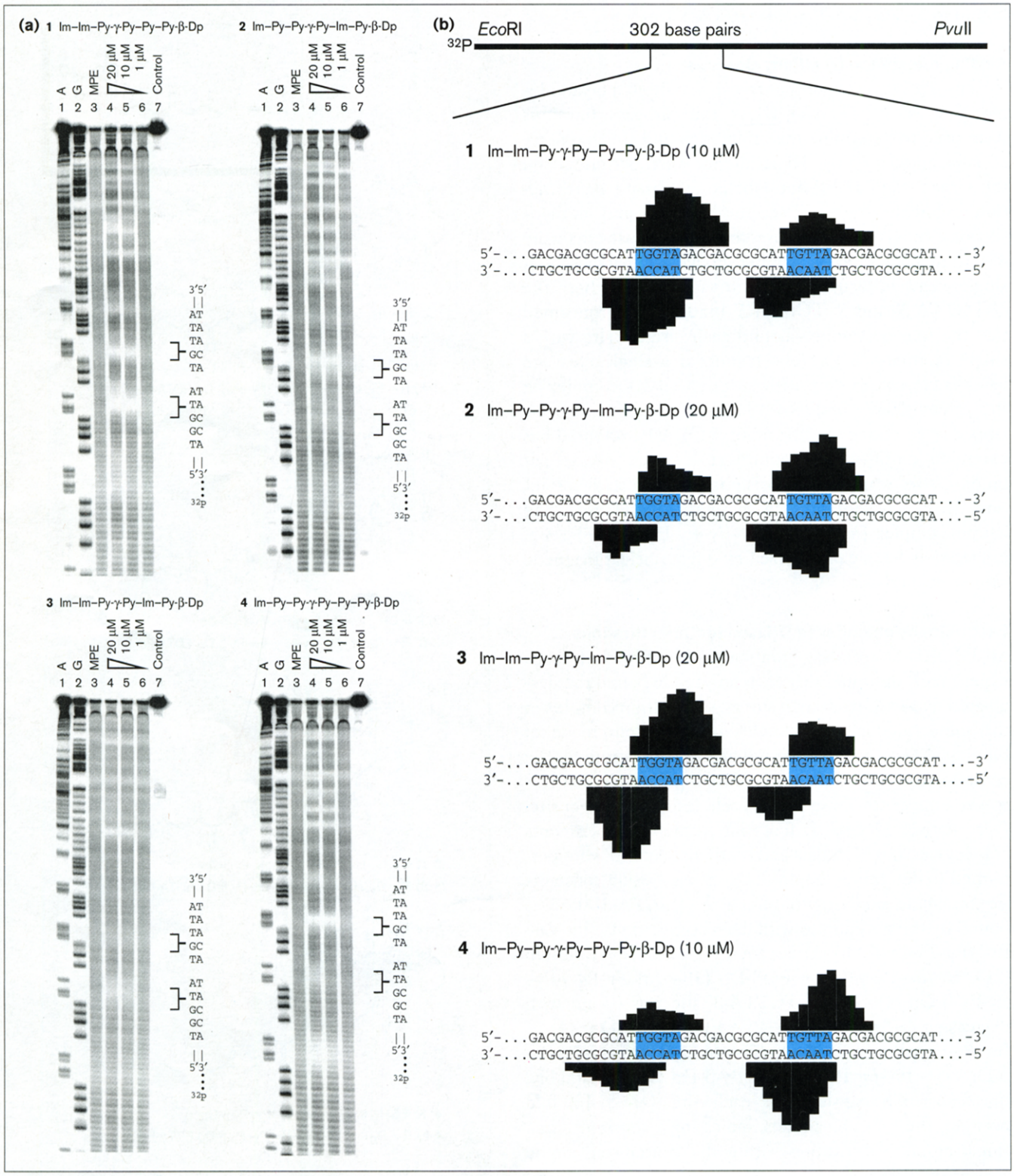
Among the four ligands, Im–*Im*–Py-γ-Py–*Py*–Py-β-Dp (**1**; central Im/Py pairing italicized) bound to the 5′-TG*G* TA-3′ site, which contained a central G·C base pair, with the highest affinity ($K_a = 9.0 \times 10^7\ M^{-1}$). This selectivity

Figure 3



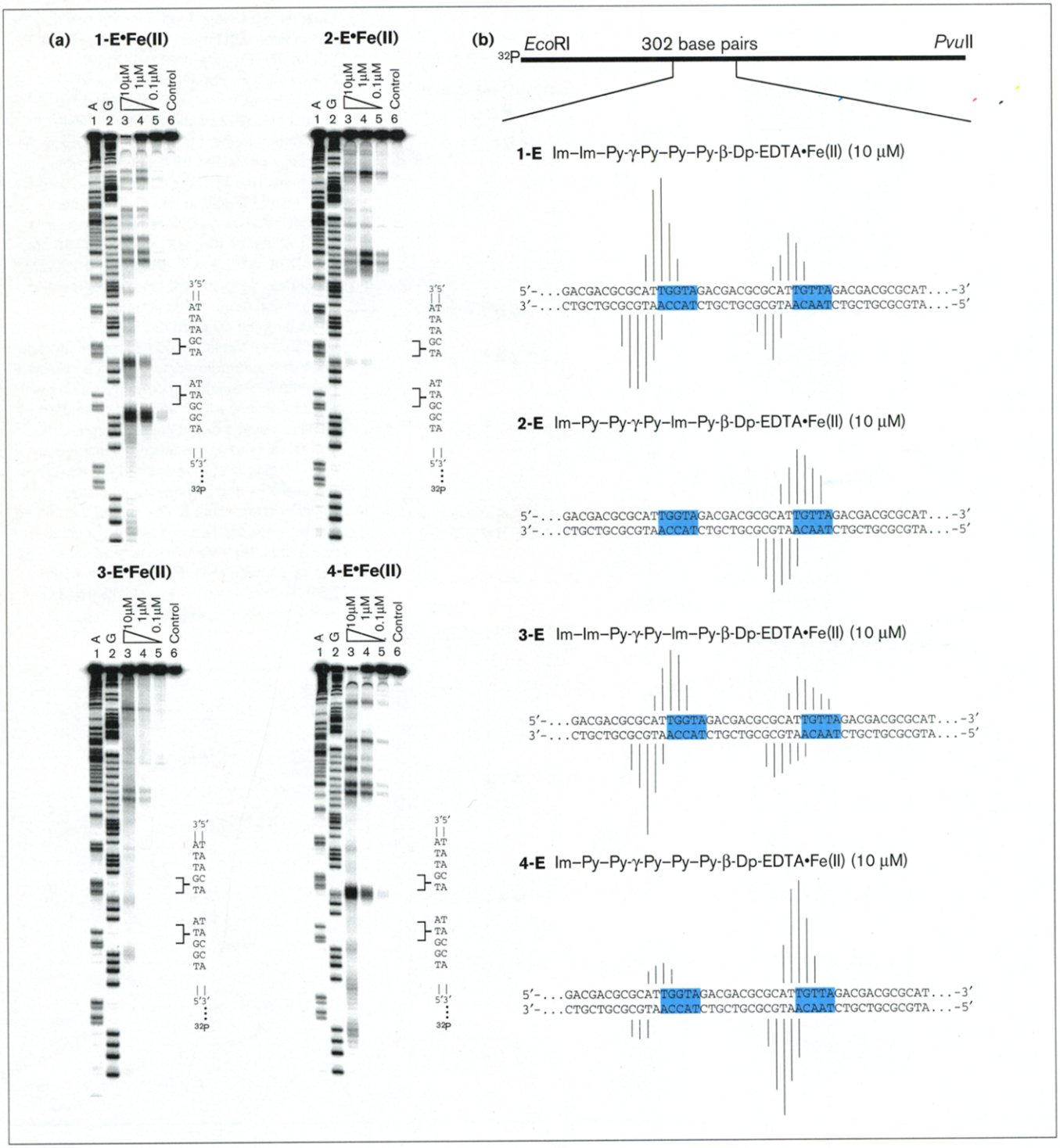
Structures of the four polyamides and their EDTA analogs. Im–Im–Py-γ-Py–Py–Py-β-Dp, **1**; Im–Py–Py-γ-Py–Im–Py-β-Dp, **2**; Im–Im–Py-γ-Py–Im–Py-β-Dp, **3**; and Im–Py–Py-γ-Py–Py–Py-β-Dp, **4**; Im–Im–Py-γ-Py–Py–Py-β-Dp-EDTA·Fe(II), **1-E**; Im–Py–Py-γ-Py–Im–Py-β-Dp-EDTA·Fe(II), **2-E**; Im–Im–Py-γ-Py–Im–Py-β-Dp-EDTA·Fe(II), **3-E**; and Im–Py–Py-γ-Py–Py–Py-β-Dp-EDTA·Fe(II), **4-E**.

Figure 4



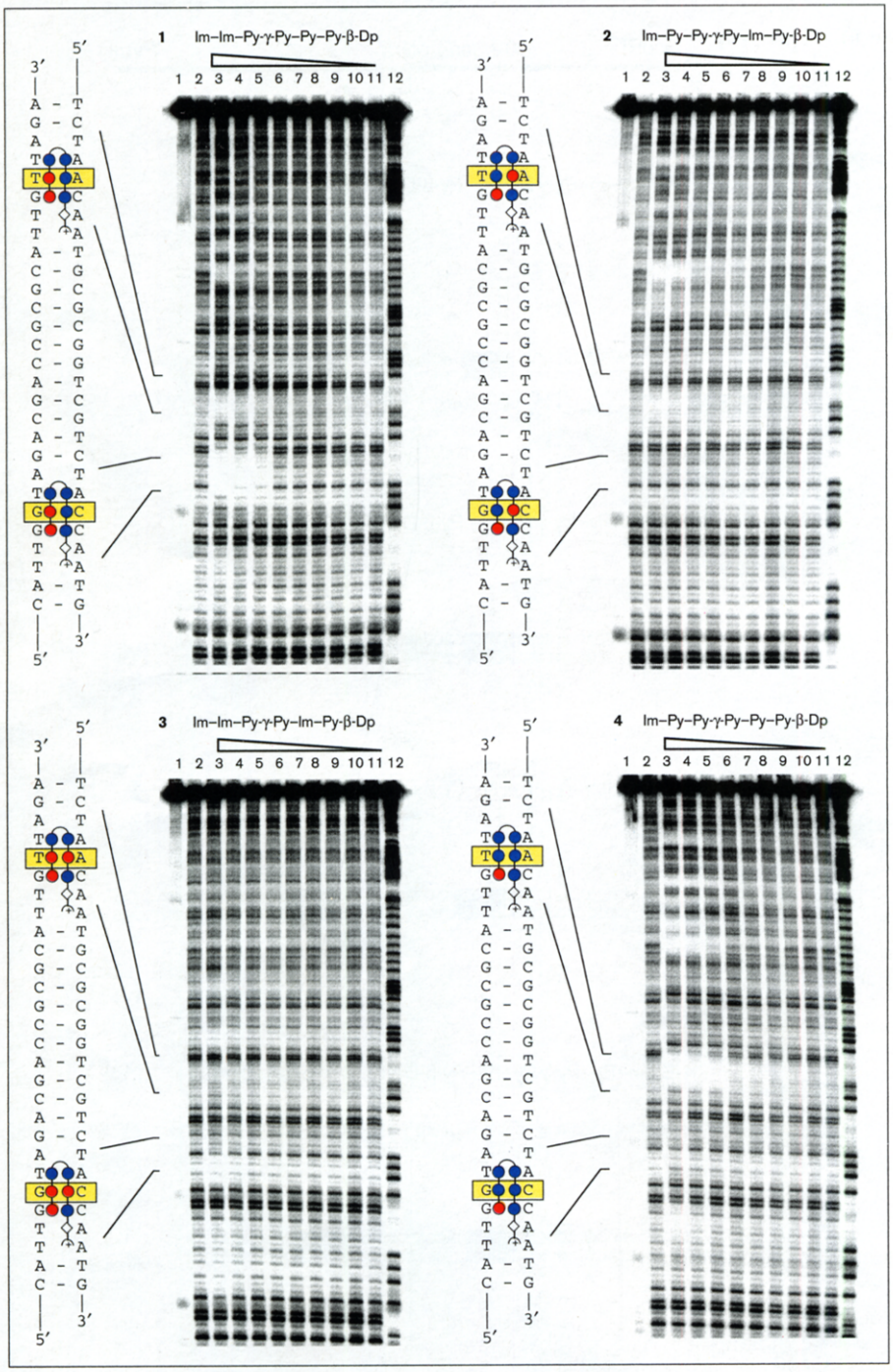
MPE·Fe(II) footprinting experiments on the 3′ ^{32}P end-labeled 302-base-pair *EcoRI*/*PvuII* restriction fragment derived from the plasmid pDEH4. **(a)** The 5′-TGGTA-3′, and 5′-TG*T*TA-3 sites are shown on the right-hand side of each autoradiogram. Lane 1, adenosine reaction; lane 2, guanine reaction; lane 3, MPE·Fe(II) standard; lanes 4–6: 20 μM, 10 μM, and 1 μM polyamide, respectively; lane 7, intact DNA. All lanes contained 30 kcpm 3′-radiolabeled DNA, 25 mM Tris-acetate buffer (pH 7.0), 10 mM NaCl, and 100 μM/base pair calf thymus DNA. **(b)** An illustration of the 302-base-pair restriction fragment with the position of the sequence indicated and MPE·Fe(II) protection patterns of each polyamide. Bar heights are proportional to the relative protection from cleavage at each band. The binding sites 5′-TG*G*TA-3′ and 5′-TG*T*TA-3′ are highlighted in blue.

Figure 5



Affinity cleavage experiments on the 3′ ^{32}P end-labeled 302-base-pair *Eco*RI/*Pvu*II restriction fragment derived from the plasmid pDEH4. **(a)** The targeted binding sites are indicated on the right-hand side of the autoradiograms. The storage phosphor autoradiograms were obtained from the 8% denaturing polyacrylamide gels used to separate the fragments generated by affinity cleavage experiments: lanes 1 and 2, A and G sequencing lanes; lanes 3–6, digestion products obtained in the presence of **1-E**, **2-E**, **3-E**, or **4-E**: 10 μM, 1 μM, and 0.1 μM polyamide, respectively; lane 7, intact DNA. All reactions contained 20 kcpm 3′ ^{32}P restriction fragment, 25 mM Tris-acetate (pH 7.0), 20 mM NaCl, 100 μM/base pair calf thymus DNA. **(b)** An illustration of the 302-base-pair restriction fragment with the position of the sequence indicated. Results from affinity cleavage with 10 μM Im–Im–Py-γ-Py–Py–Py-β-Dp-EDTA·Fe(II), **1-E·Fe(II)**; Im–Py–Py-γ-Py–Im–Py-β-Dp-EDTA·Fe(II), **2-E·Fe(II)**; Im–Im–Py-γ-Py–Im–Py-β-Dp-EDTA·Fe(II), **3-E·Fe(II)**; and Im–Py–Py-γ-Py–Py–Py-β-Dp-EDTA·Fe(II), **4-E·Fe(II)**. Line heights are proportional to the relative cleavage intensities at each base pair. The binding sites 5′-TG*G*TA-3′ and 5′-TG*T*TA-3′ are highlighted in blue.

Figure 6



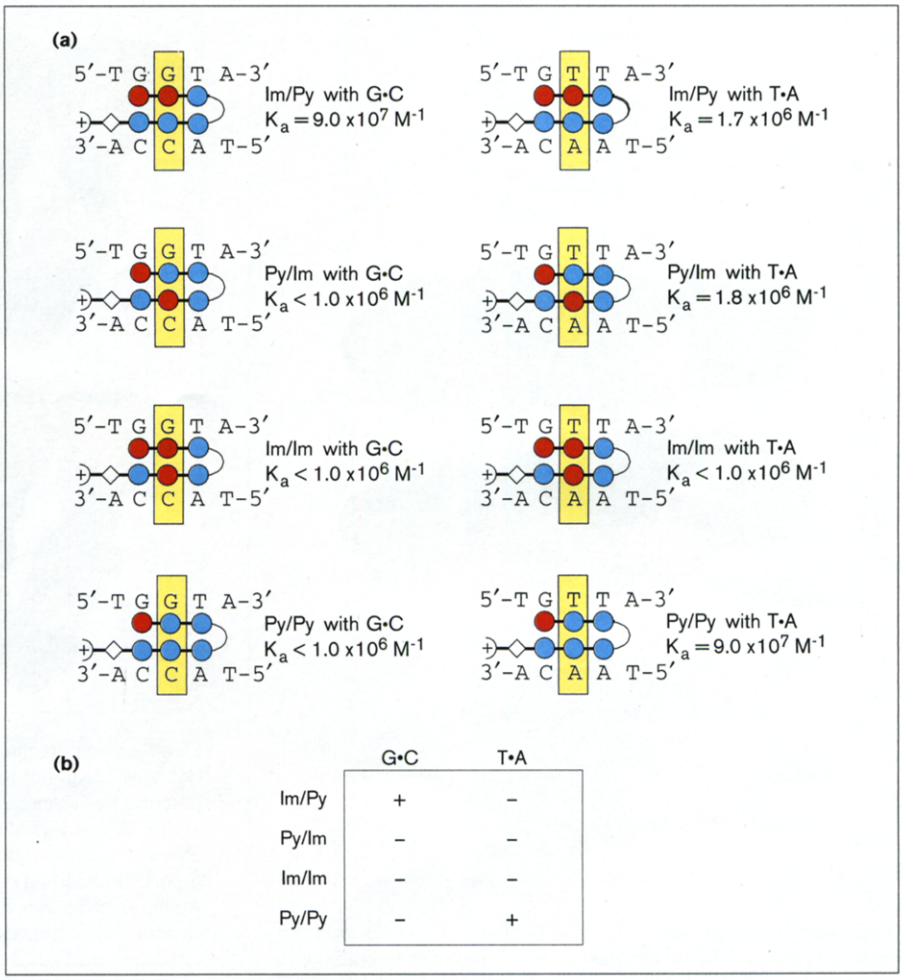
Quantitative DNase I footprint titration experiment with Im–Im–Py-γ-Py–Py–Py-β-Dp (**1**), Im–Py–Py-γ-Py–Im–Py-β-Dp (**2**), Im–Im–Py-γ-Py–Im–Py-β-Dp (**3**), and Im–Py–Py-γ-Py–Py–Py-β-Dp (**4**), on the 3′ ^{32}P end-labeled 302-base-pair *Eco*RI/*Pvu*II restriction fragment from plasmid pDEH4. All reactions contained 20 kcpm restriction fragment, 10 mM Tris-HCl (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$. Lane 1, intact DNA; lane 2, DNase I standard; lanes 3–11 contain 1 μM, 100 nM, 65 nM, 10 nM, 6.5 nM, 4 nM, 2.5 nM, 500 pM, and 100 pM polyamide, respectively; lane 12, G reaction. The portion of the restriction fragment containing the targeted binding sites is indicated on the left-hand side of the storage phosphor autoradiograms that were obtained from the 8% denaturing polyacrylamide gels used to separate the fragments generated by affinity cleavage experiments. Hairpin polyamide binding models are indicated at each binding site; blue and red circles represent Py and Im rings, respectively. Diamonds represent β-alanine and the curved line between the two polyamide subunits represents the γ-aminobutyric acid linker. The central pairings of Im/Py, Py/Im, Im/Im and Py/Py with G·C and T·A are highlighted with yellow boxes.

indicates that Im/Py is the optimal ring pairing for recognition of a G·C base pair. The sequence specificity of the Im/Py pairing was underscored by the 50-fold reduced affinity ($K_a = 1.7 \times 10^6\ M^{-1}$) when the Im/Py pair was placed opposite a T·A base pair at 5′-TG*T*TA-3′.

The Py/Im pair

The polyamide Im–*Py*–Py-γ-Py–*Im*–Py-β-Dp (**2**; central Py/Im pairing italicized) bound to the 5′-TG*G*TA-3′ site (central G·C base pair) with 100-fold reduced affinity relative to polyamide **1** (central Im/Py pairing). Remarkably, given the central location of the exocyclic 2-amino group of guanine in the minor groove, the Py/Im pairing is disfavored relative to the Im/Py pair for the recognition of a G·C base pair. Although the nitrogen of the 2-amino group of guanine was displayed in a similar location for both G·C and C·G, the proton available for hydrogen bond recognition in the minor groove had a strand-specific directionality

Figure 7



Models of the expected hairpin complex of Im–Im–Py-γ-Py–Py–Py-β-Dp (**1**), Im–Py–Py-γ-Py–Im–Py-β-Dp (**2**), Im–Im–Py-γ-Py–Im–Py-β-Dp (**3**), and Im–Py–Py-γ-Py–Py–Py-β-Dp (**4**) complexed with 5′-TGGTA-3′ and 5′-TG*T*-TA-3′. Blue and red circles represent Py and Im rings, respectively. Diamonds represent β-alanine and the curved line between the two polyamide subunits represents the γ-amino-butyric acid linker. The central pairings of Im/Py, Py/Im, Im/Im and Py/Py with G·C and T·A are highlighted with yellow boxes. The binding affinity for each complex as determined from quantitative DNase I footprinting is shown to the right of each complex. **(b)** A summary of the binding affinities: +, favorable; −, no binding.

(Figure 8). The directionality requirements for effective hydrogen bond formation thus allow discrimination of G·C by the Im/Py and Py/Im pairs. Validation of this interpretation awaits the completion of high resolution X-ray crystallographic structure studies which are in progress (C.L. Kielkopf, E.E.B., P.B.D. and D.C. Rees, unpublished observations). Placing the Py/Im pair of **2** opposite T·A at the 5′-TG*T*TA-3′ target site resulted in similar affinity (K_a = $1.8 \times 10^6\ M^{-1}$) to placing the Im/Py pair **1** opposite T·A. It should be noted that the Py/Im hairpin polyamide **2** recognized a 5′-AG*C*TT-3′ match site present on the restriction fragment (central C·G base pair) with high affinity.

The Py/Py pair

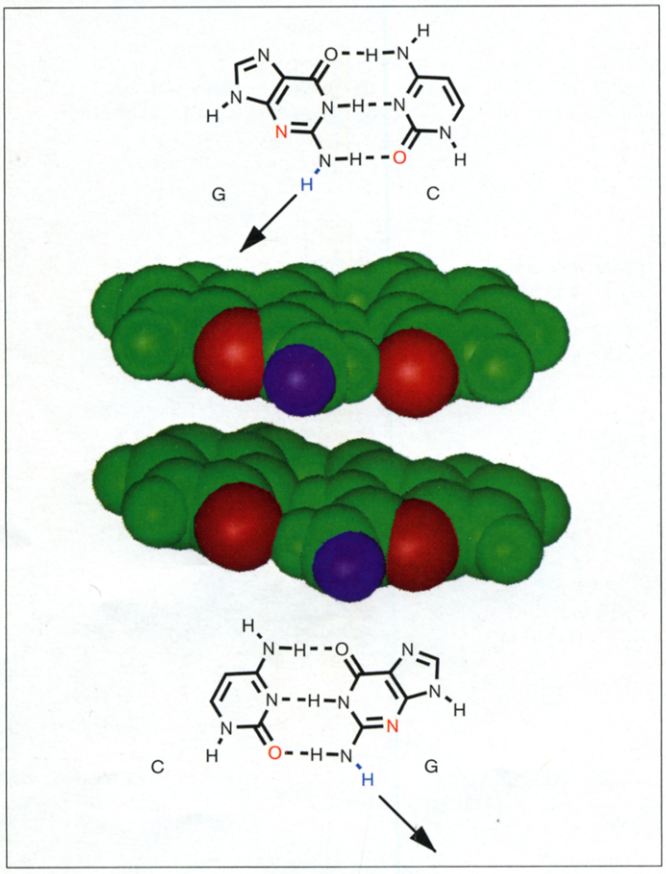
The polyamide Im–*Py*–Py-γ-Py–*Py*–Py-β-Dp (**4**; central Py/Py pairing) bound to the 5′-TG*T*TA-3′ site (central T·A base pair) with 100-fold enhanced affinity relative to the 5′-TG*G*TA-3′ site (central G·C base pair). The discrimination of A·T/T·A base pairs from G·C/C·G base pairs by the Py/Py pairing is likely to be due to the exocyclic amine groups of guanine which present a steric hindrance to deep polyamide binding in the minor groove. Im-rich hairpin polyamides, however, recognize G·C sequences with affinities and specificities similar those of Py-rich polyamides that recognize A·T rich sequences [31], indicating (as will become evident below) that additional energetic parameters may be important for high-affinity recognition.

The Im/Im pair

The hairpin Im–*Im*–Py-γ-Py–*Im*–Py-β-Dp (central Im/Im pairing italicized) bound to both designated sites, 5′-TG*T*-TA-3′ and 5′-TG*G*TA-3′ with > 100-fold reduced affinity relative to polyamides **1** and **4** (central Py/Py and Im/Py pairs, respectively). The reduced binding energetics of the Im/Im pair may result from desolvation of the N3 of Im upon binding to DNA [32]. Such an unfavorable desolvation could be compensated for when a hydrogen bond is formed between Im and the exocyclic 2-amino group of guanine, but not compensated for upon placement of Im opposite A, T, or C bases.

Alternatively, specificity may result from electrostatic interactions between the aromatic ring carboxamides and the floor of the DNA minor groove [33]. The DNA minor

Figure 8



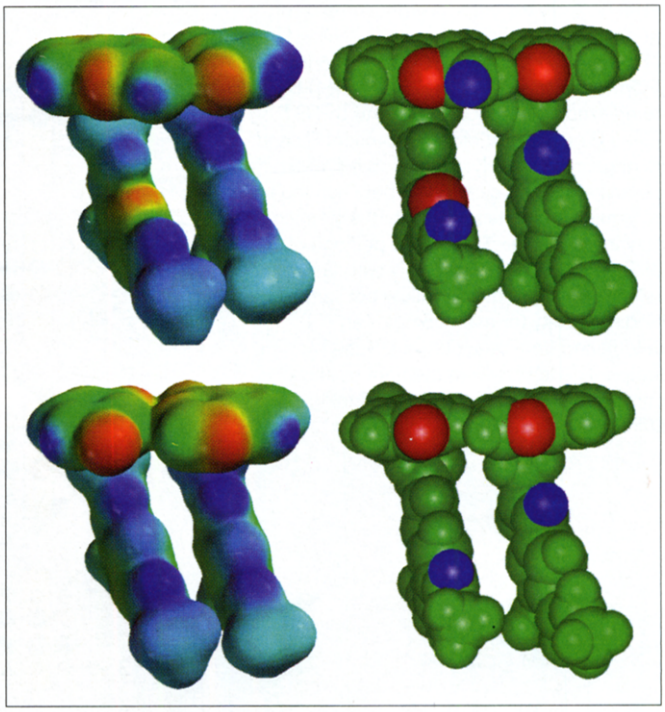
Space-filling models of the G·C and C·G base pairs as viewed from the minor groove of DNA, generated using B-form DNA coordinates provided in InsightII. The O2 and N3 atoms are highlighted in red. The asymmetrically directed amino proton is highlighted in blue.

groove displays a negative electrostatic potential at A·T and T·A base pairs [34]. The only positive potential located on the minor groove floor is at the exocyclic 2-amino group of guanine [34]. The Py ring displays a positive potential across the interface with the floor of the minor groove, providing complementary electrostatic interactions at A, T, and C bases (Figure 9). The Im ring displays a negative potential which can interact favorably with the 2-amino group of G, but may interact unfavorably with the other three bases.

Significance

A binary code has been developed to correlate DNA sequence with polyamide sequence composition. The results described here demonstrate the sequence specificity of the four possible individual pyrrole and imidazole ring amino acid pairings: the Im/Py pairing recognizes G·C, but is disfavored when placed opposite C·G, A·T and T·A; the Py/Im pairing targets C·G, but is disfavored opposite G·C, A·T and T·A base pairs; the Py/Py pairing recognizes T·A/A·T base pairs but is disfavored opposite G·C/C·G; and an Im/Im pairing is disfavored with any of the four base pairs, breaking a potential degeneracy for recognition by preventing unlinked polyamide dimers from binding in certain slipped motifs. These results will guide further second-generation polyamide design for DNA recognition.

Figure 9



Electrostatic potential maps (left) and space-filling models (right) for Im/Py with G·C (top) and Py/Py with T·A (bottom). Electrostatic potentials were calculated using MacSpartan version 1.0 as described in [39]. Areas of positive potential are represented as blue surface and areas of negative potential as red surface; neutral regions appear green. Space-filling models were generated using InsightII software. Atoms involved in hydrogen bond formations are highlighted, with donors and acceptors represented as blue and red, respectively.

Materials and methods

The solid phase synthesis of six-ring hairpin polyamides has been described [9,17]. Epicurian coli® XL-1 Blue supercompetent cells were obtained from Stratagene (La Jolla, CA, USA). All enzymes were purchased from Boeringher-Mannheim (Indianaoplis, IN, USA) and were used according to the supplier's recommended protocol in the activity buffer provided. Sequenase (version 2.0) was obtained from United States Biochemical (Cleveland, OH, USA). [α-^{32}P]-Thymidine-5′-triphosphate (≥ 3000 Ci/mmol), [α-^{32}P]-deoxyadenosine-5′-triphosphate (≥ 6000 Ci/mmol), and [γ-^{32}P]-adenosine-5′-triphosphate were purchased from Du Pont/NEN (Boston, MA, USA).

Preparation of ^{32}P-labeled DNA

Plasmid pDEH4 was prepared by hybridizing a complementary set of synthetic oligonucleotides, 5′-CTAGACGACGCGCATTATTAGACGA-CGCGCATTGGTAGACGACGCGCATTGTTAGACGACGCGCATTG-TCAGACGACGCGCATTGCA-3′ and 5′-ATGCGCGTCGTCTGACA-ATGCGCGTCGTCTAACAATGCGCGTCGTCTACCAATGCGCGTC-GTCTAATAATGCGCGTCGT-3′ and ligating the resulting duplex to the large pUC19 *XbaI*/*PstI* restriction fragment. The 3′ ^{32}P end-labeled

*Eco*RI/*Pvu*II fragment was prepared by digesting the plasmid with *Eco*RI and simultaneously filling in using Sequenase, [α-^{32}P]-deoxyadenosine-5′-triphosphate, and [α-^{32}P]-thymidine-5′-triphosphate, digesting with *Pvu*II, and isolating the 302-base-pair fragment by nondenaturing gel electrophoresis. The 5′ ^{32}P end-labeled *Eco*RI/*Pvu*II fragment was prepared using standard methods. A and G sequencing were carried out as described [35,36]. Standard methods were used for all DNA manipulations [37].

MPE·Fe(II) footprint titrations

All reactions were performed in a total volume of 40 µl. A polyamide stock solution (**1**, **2**, **3**, or **4**) or H_2O (for reference lanes) was added to an assay buffer containing labeled restriction fragment (30,000 cpm), affording final solution conditions of 25 mM Tris-acetate (pH 7.0), 10 mM NaCl, 100 µM/base pair calf thymus DNA, 5mM dithiothreitol (DTT). Solutions were incubated at 22°C for 24 h. A fresh 50 µM MPE·Fe(II) solution was made from 100 µl of a 100 µM methidiumpropyl-EDTA (MPE) solution and 100 µM ferrous ammonium sulfate ($Fe(NH_4)_2(SO_4)_2{\cdot}6H_2O$) solution. Then, 4 µl of the 50 µM MPE·Fe(II) solution was added, and the solution was allowed to equilibrate for 10 min at 22°C. Cleavage was initiated by the addition of 4 µl of a 50 mM dithiothreitol solution and allowed to proceed for 15 min at 22°C. Reactions were stopped by ethanol precipitation, and were resuspended in 1 × TBE/80% formamide loading buffer, denatured by heating at 85°C for 15 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 × TBE at 2000 V for 1.5 h. Gels were dried and exposed to a storage phosphor screen [38]. Relative cleavage intensities were determined by volume integration of individual cleavage bands using ImageQuant software (Sunnyvale, CA, USA).

Affinity cleavage reactions

All reactions were executed in a total volume of 40 µl. A stock solution of polyamide (**1-E**, **2-E**, **3-E**, or **4-E**) or H_2O was added to a solution containing labeled restriction fragment (20,000 cpm), affording final solution conditions of 25 mM Tris-acetate (pH 7.0), 20 mM NaCl, 100 µM/base pair calf thymus DNA. Solutions were incubated for a minimum of 4 h at 22°C. Subsequently, 4 µl of freshly prepared 100 µM $Fe(NH_4)_2(SO_4)_2$ was added and the solution allowed to equilibrate for 20 min. Cleavage reactions were initiated by the addition of 4 µl of 100 mM dithiothreitol, allowed to proceed for 30 min at 22°C, then stopped by the addition of 10 µl of a solution containing 1.5 M NaOAc (pH 5.5), 0.28 mg/ml glycogen, and 14 µM/base pair calf thymus DNA, and ethanol precipitated. The reactions were resuspended in 1 × TBE/80% formamide loading buffer, denatured by heating at 85°C for 15 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 × TBE at 2000 V for 1.5 h. Gels were dried and exposed to a storage phosphor screen. 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 40 µl. A polyamide stock solution (**1**, **2**, **3**, or **4**) or H_2O (for reference lanes) was added to an assay buffer containing radiolabeled restriction fragment (20,000 cpm), affording final solution conditions of 10 mM Tris-HCl (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, and either 0.1 nM–1 µM polyamide or no polyamide (for reference lanes). The solutions were allowed to equilibrate at 22°C for 24 h. Footprinting reactions were initiated by the addition of 4 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 10 µl of a solution containing 1.25 M NaCl, 100 mM EDTA, 0.2 mg/ml glycogen, and 28 µM/base pair calf thymus DNA, and ethanol precipitated. Reactions were resuspended in 1 × TBE/80% formamide loading buffer, denatured by heating at 85°C for 15 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 × TBE at 2000 V for 1.5 h. Gels were dried and exposed to a storage phosphor screen (Molecular Dynamics, Sunnyvale, CA, USA).

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:

$$\theta_{app} = 1 - \frac{I_{site}/I_{ref}}{I^0_{site}/I^0_{ref}} \quad (1)$$

where I^0_{site} and I^0_{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 Langmuir binding isotherm (Equation 2, n = 1) by minimizing the difference between θ_{app} and θ_{fit}, using the modified Hill equation:

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

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 fitted using a nonlinear least-squares fitting procedure with K_a, θ_{max}, and θ_{min} as the adjustable parameters. All acceptable fits had a correlation coefficient of $R > 0.97$. At least three sets of data were used in determining each association constant. All lanes from each gel were used unless visual inspection revealed a data point to obviously flawed relative to neighboring points. The binding isotherms were normalized using the following equation:

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

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 S.W., 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)$_5$G(A,T)C(A,T) sequences by a dimeric side-by-side motif. *J. Am. Chem. Soc.* **114**, 8783-8794.
2. Mrksich, M., Wade, W.S., Dwyer, T.J., Geirstanger, 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. 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.
5. Trauger, J.W., Baird, E.E. & Dervan, P.B. (1996). Recognition of DNA by designed ligands at subnanomolar concentrations. *Nature* **382**, 559-561.
6. Pelton, J.G. & Wemmer, D.E. (1989). Structural characterization of a 2-1 distamycin A·d(CGCAAATTTGGC)$_2$ complex by two-dimensional NMR. *Proc. Natl Acad. Sci. USA* **86**, 5723-5727.
7. Pelton, J.G. & Wemmer, D.E. (1990). Binding modes of distamycin-A with d(CGCAAATTTGCG)$_2$ determined by two-dimensional NMR *J. Am. Chem. Soc.* **112**, 1393-1399.

8. Chen, X., Ramakrishnan, B., Rao, S.T. & Sundaralingham, M. (1994). Binding of 2 distamycin-A molecules in the minor groove of an alternating B-DNA duplex. *Nat. Struct. Biol.* **1**, 169-175.
9. White, S, Baird, E.E. & Dervan, P.B. (1996). Effects of the A·T/T·A degeneracy of pyrrole-imidazole polyamide recognition in the minor groove of DNA. *Biochemistry* **35**, 12532-12537.
10. Gottesfeld, J.M., Nealy, L., Trauger, J.W., Baird, E.E. & Dervan, P.B. (1997). Regulation of gene expression by small molecules. *Nature* **387**, 202-205.
11. Pilch, D.S., *et al.*, & Dervan, P.B. (1996). Binding of a hairpin polyamide in the minor groove of DNA: sequence-specific enthalpic discrimination. *Proc. Natl Acad. Sci. USA* **93**, 8306-8311.
12. 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.
13. Wing, R., *et al.*, & Dickerson, R.E. (1980). Crystal structure analysis of a complete turn of DNA. *Nature* **287**, 755-758.
14. Seeman, N.C., Rosenberg, J.M. & Rich, A. (1976). Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl Acad. Sci. USA* **73**, 804-808.
15. Steitz, T.A. (1990). Structural studies of protein nucleic-acid interaction-the sources of sequence-specific binding. *Quart. Rev. Biophys.* **23**, 205-280.
16. Dwyer, T.J., Geierstanger, B.H., Bathini, Y., Lown, J.W. & Wemmer, D.E. (1992). Design and binding of a Distamycin A analog to d(CGCAAGTTGGC) - synthesis, NMR-studies, and implications for the design of sequence-specific minor groove binding oligopeptides. *J. Am. Chem. Soc.* **114**, 5911-5919.
17. 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.
18. Trauger, J.W., Baird, E.E. & Dervan, P.B. (1996). Extended hairpin polyamide motif for sequence-specific recognition in the minor groove of DNA. *Chem. Biol.* **3**, 369-377.
19. Parks, M.E., Baird, E.E. & Dervan, P.B. (1996). Optimization of the hairpin polyamide design for recognition of the minor groove of DNA. *J. Am. Chem. Soc.* **118**, 6147-6152.
20. Parks, M.E., Baird, E.E. & Dervan, P.B. (1996). Recognition of 5′-(A,T)GG(A,T)$_2$-3′ by hairpin polyamides in the minor groove of DNA. *J. Am. Chem. Soc.* **118**, 6153-6159.
21. Baird, E.E. & Dervan, P.B. (1996). Solid phase synthesis of polyamides containing imidazole and pyrrole amino acids. *J. Am. Chem. Soc.* **118**, 6141-6146.
22. Taylor, J.S., Schultz, P.G. & Dervan, P.B. (1984). DNA affinity cleavage. Sequence specific cleavage of DNA by distamycin-EDTA·Fe(II) and EDTA-distamycin·Fe(II). *Tetrahedron* **40**, 457.
23. Dervan, P.B. (1986) Design of sequence-specific DNA-binding molecules. *Science* **232**, 464-471.
24. Van Dyke, M.W., Hertzberg, R.P. & Dervan, P.B. (1982). Map of distamycin, netropsin, and actinomycin binding sites on heterogeneous DNA. DNA cleavage inhibition patterns with methidiumpropyl-EDTA·Fe(II). *Proc. Natl Acad. Sci. USA*, **79**, 5470-5474.
25. Van Dyke, M.W. & Dervan, P.B. (1984). Echinomycin binding sites on DNA. *Science* **225**, 1122-1127.
26. 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.
27. 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.
28. 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.
29. Trauger, J.W., Baird, E.E., Mrksich, M. & Dervan, P.B. (1996). Extension of sequence-specific recognition in the minor groove of DNA to 9-13 base-pairs. *J. Am. Chem. Soc.* **118**, 6160-6166.
30. De Clairac, R.P.L., Geierstanger, B.H., Mrksich, M., Dervan, P.B. & Wemmer, D.E. (1997). NMR characterization of hairpin polyamide complexes with the minor groove of DNA. *J. Am. Chem. Soc.*, in press.
31. Swalley, S.E., Baird, E.E. & Dervan, P.B. (1997). Discrimination of 5′-GGGG-3′, 5′-GCGC-3′, and 5′-GGCC-3′ sequences in the minor groove of DNA by eight-ring hairpin polyamides. *J. Am. Chem. Soc.* **119**, 6953-6961.
32. Singh, S.B., Wemmer, D.E. & Kollman, P.A. (1994). Relative binding affinities of distamycin and its analog to d(CGCAAGTTGGC)·d(GCCAACTTGCG) comparison of simulation results with experiment. *Proc. Natl Acad. Sci. USA* **91**, 7673-7677.
33. Zimmer, C. & Wahnert, U. (1986). Nonintercalating DNA binding ligands: specificity of the interaction and their use as tools in biophysical, biochemical, and biological investigations of the genetic material. *Prog. Biophys. Mol. Biol.* **47**, 31-112.
34. Pullman, B. (1990). Molecular mechanisms of specificity in DNA-antitumour drug interactions. *Adv. Drug Res.* **18**, 1-113.
35. Maxam, A.M. & Gilbert, W.S. (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* **65**, 499-560.
36. Iverson, B.L. & Dervan, P.B. (1987). Adenine-specific DNA chemical sequencing reaction. *Methods Enzymol.* **15**, 7823-7830.
37. Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989). *Molecular Cloning.* (2nd edn). Cold Spring Harbor Laboratory Press, Cold Spring Harbor New York.
38. Johnston, R.F., Picket, S.C. & Barker, D.L. (1990). Autoradiography using storage phosphor technology. *Electrophoresis* **11**, 355-360.
39. Mecozzi, S., West, A.P. & Dougherty, D.A. (1996). Cation-pi interactions in aromatics of biological and medicinal interest-electrostatic potential surfaces as a useful qualitative guide. *Proc. Natl Acad. Sci. USA* **93**, 10566-10571.