Molecular recognition of DNA base pairs by the formamido/pyrrole and formamido/imidazole pairings in stacked polyamides

Karen L. Buchmueller[1], Andrew M. Staples[1], Peter B. Uthe[1], Cameron M. Howard[1,2], Kimberly A. O. Pacheco[3], Kari K. Cox[1], James A. Henry[1], Suzanna L. Bailey[1], Sarah M. Horick[1], Binh Nguyen[2], W. David Wilson[2] and Moses Lee[1,*]

[1]Department of Chemistry, Furman University, Greenville, SC 29613, USA, [2]Department of Chemistry, Georgia State University, Atlanta, GA 30303, USA and [3]Department of Chemistry and Biochemistry, University of Northern Colorado, Greeley, CO 80639, USA

ABSTRACT

Polyamides containing an N-terminal formamido (f) group bind to the minor groove of DNA as staggered, antiparallel dimers in a sequence-specific manner. The formamido group increases the affinity and binding site size, and it promotes the molecules to stack in a staggered fashion thereby pairing itself with either a pyrrole (Py) or an imidazole (Im). There has not been a systematic study on the DNA recognition properties of the f/Py and f/Im terminal pairings. These pairings were analyzed here in the context of f-ImPyPy, f-ImPyIm, f-PyPyPy and f-PyPyIm, which contain the central pairing modes, –ImPy– and –PyPy–. The specificity of these triamides towards symmetrical recognition sites allowed for the f/Py and f/Im terminal pairings to be directly compared by SPR, CD and ΔT_M experiments. The f/Py pairing, when placed next to the –ImPy– or –PyPy– central pairings, prefers A/T and T/A base pairs to G/C base pairs, suggesting that f/Py has similar DNA recognition specificity to Py/Py. With –ImPy– central pairings, f/Im prefers C/G base pairs (>10 times) to the other Watson–Crick base pairs; therefore, f/Im behaves like the Py/Im pair. However, the f/Im pairing is not selective for the C/G base pair when placed next to the –PyPy– central pairings.

INTRODUCTION

The development of a diverse group of polyamides that bind specific DNA sequences (1–3) is a promising arena for the design of new therapeutics (4,5), and it has also provided substantial information on the structure and function of DNA (6–8). Consequently, a thorough understanding of the interactions and dynamics between polyamides and DNA can have a major impact on drug design and DNA molecular recognition, possibly beyond the realm of polyamides.

Distamycin A is a naturally occurring polyamide with antibacterial properties and is the basis for the closely related synthetic triheterocyclic polyamides (triamides) (9–11). Triamides are a good model system for investigating the structure–function relationship between polyamide components and DNA sequence-specific recognition. Triamides bind as antiparallel dimers in the minor groove of DNA, such that the positively charged C-termini are distal from one another (12–14). The ability for two triamide molecules to stack in the minor groove of DNA, rather than just one molecule binding, is vital for the recognition of both DNA strands and, therefore, to reduce the degeneracy of sequence selectivity.

Recently, our group has demonstrated that the arrangement of the imidazole (Im) and pyrrole (Py) moieties, and the inclusion of an N-terminal formamido (f) group are crucial in the design of useful polyamides (15,16). Distamycin A consists of three pyrrole rings and selectively binds AT-rich DNA (9–11). Stacked pyrroles (Py/Py) from the two separate molecules of the homodimer are unable to distinguish between A/T and T/A base pairs; however, pyrrole still provides the strongest binding affinity for adenine and thymine bases over other heterocyclic moieties that can distinguish between these base pairs (17–19). Incorporation of imidazole rings into polyamides considerably advanced DNA sequence recognition. Im/Py stacked pairs are very selective for G/C base pairs and Py/Im pairs recognize C/G base pairs (13,20). These findings allowed for the design of polyamides to target specific, also known as cognate,

DNA sequences (21–23). Distamycin A has a formamido group (f) at the N-terminus, but for synthetic reasons this group is often omitted when novel polyamides are designed (16,24). Recent work has shown that the formamido group is a very important component of polyamide design (16). First, the formamido group allows for the triamide to bind as a staggered dimer, such that the N-terminal formamido group of one molecule is stacked opposite the C-terminal heterocycle of the second molecule (Figure 1A). The staggered binding mode, rather than the overlapped mode observed for non-formylated polyamides (Figure 1B), allows for six base pairs to be recognized by the polyamide dimer. More importantly, it has been shown that the formamido group improves DNA binding affinity by one to two orders of magnitude. For example, f-ImPyPy binds its cognate DNA, TGCA, with a $K_{eq} = 1.2 \times 10^7 \text{ M}^{-1}$, but the ImPyPy counterpart binds to its cognate, GTC, with a $K_{eq} = 1.4 \times 10^5 \text{ M}^{-1}$, an 85-fold reduction in binding affinity for the non-formylated triamide (16).

In addition, the context of imidazole and pyrrole moieties within the triamide must be considered when designing polyamide-based DNA binding agents (15). The triamide dimers can be dissected into two morphologically distinct units: the central and terminal pairing modes (Figure 1A). For example, the central pairings are underlined for the f-ImPyIm/ImPyIm-f homodimer and denoted as –ImPy–. The terminal pairing group consists of the remaining parts of the dimer and are designated the f/Im terminal pairs. There is a distinct trend that relates the content of the central pairing mode to the strength of binding affinity. The strongest central pairing motif is –ImPy–, and in decreasing affinity for their respective Watson–Crick cognate sequences are –PyPy– then –PyIm– and –ImIm–. Interestingly, f-ImPyIm exhibits one-order of magnitude better affinity for its cognate DNA than does distamycin A for AATT (15). These findings are significant because the language of DNA sequence recognition by polyamides has been expanded to include 'words' of two base pairs, instead of the existing paradigm of recognizing one 'letter' or base pair at a time.
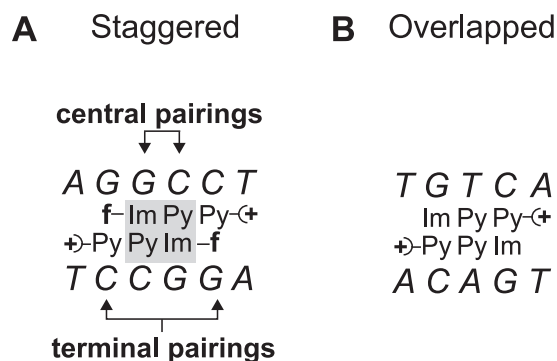
Even though the recognition rules have been elucidated for the stacked heterocyclic pairs in the central recognition motif, the sequence preference for formamido/pyrrole (f/Py) and formamido/imidazole (f/Im) terminal pairings has not been systematically studied. This study is important because the formamido group increases both the binding affinity and DNA recognition site size (16); in addition to being an essential component of the natural products distamycin A. For distamycin A and other formylated oligopyrroles (number of pyrroles = 1, 3–5), the preference of f/Py for A/T and T/A base pairs over C/G or G/C base pairs has been well established (9–11,25,26). Thus, it seems that f/Py pairings, at least behave in a similar fashion to Py/Py pairings. f/Im has been shown to bind to C/G base pairs by the f-ImImIm homodimer (27) or to both C/G and T/A base pairs by the f-ImIm homodimer (28). However, f-ImIm exhibits a low binding affinity for DNA, which relates to low specificity among a variety of sequences (15) and, therefore, the f/Im may exhibit base pair selectivity in other environments. Molecular recognition properties of the f/Py and f/Im to sequences that contain A/T-rich, mixed AT/GC- and GC-rich target sites are described herein.

Polyamide and DNA sequence design

Four triheterocyclic polyamides (triamides) were used to study the binding preference for formamido/heterocycle stacked pairings: f-PyPyPy, f-PyPyIm, f-ImPyPy and f-ImPyIm (Figure 2A). These triamides contain one of two different central pairing motifs: –ImPy– and –PyPy–, which are the best pyrrole and imidazole containing central pairings in terms of binding affinity to Watson–Crick sequences. The other four combinations of pyrrole and imidazole-containing triamides were not investigated because they exhibited low DNA binding affinity, even for their cognate sequences. The DNA sequences designed for the studies contained symmetric recognition sites, so that the two formamido/heterocycle pairings within a single dimer were in the same local environment. The core recognition site, –GC–, for f-ImPyPy and f-ImPyIm remained constant, while the terminal recognition sequence was varied to all four base pairs (Figure 1B). Control sequences, CCGG and TCGA, were added to alter the central recognition DNA site. The DNA sequences for f-PyPyPy and f-PyPyIm include the same –AT– DNA core recognition site for the central pairing mode, and the terminal pairing recognition site was tested using AATT and CATG (A/T and C/G terminal recognition sites, respectively).

METHODS

General

The surface plasmon resonance (SPR), ΔT_M, and circular dichroism (CD) experiments were performed in MES20 (10 mM 2-(*N*-morpholino)ethanesulfonic acid, 200 mM NaCl, 1 mM EDTA, pH 6.2) or PO$_4$20 buffer (10 mM sodium phosphate, 200 mM Na$^+$, 1 mM EDTA, pH 6.2) at room temperature (24–25°C). Polyamides behave the same in both buffering systems (control experiments not shown). DNA sequences were chemically synthesized and desalted by Qiagen or Midland Certified Reagent Company, with purity of >98% after HPLC purification. Oligonucleotides needed
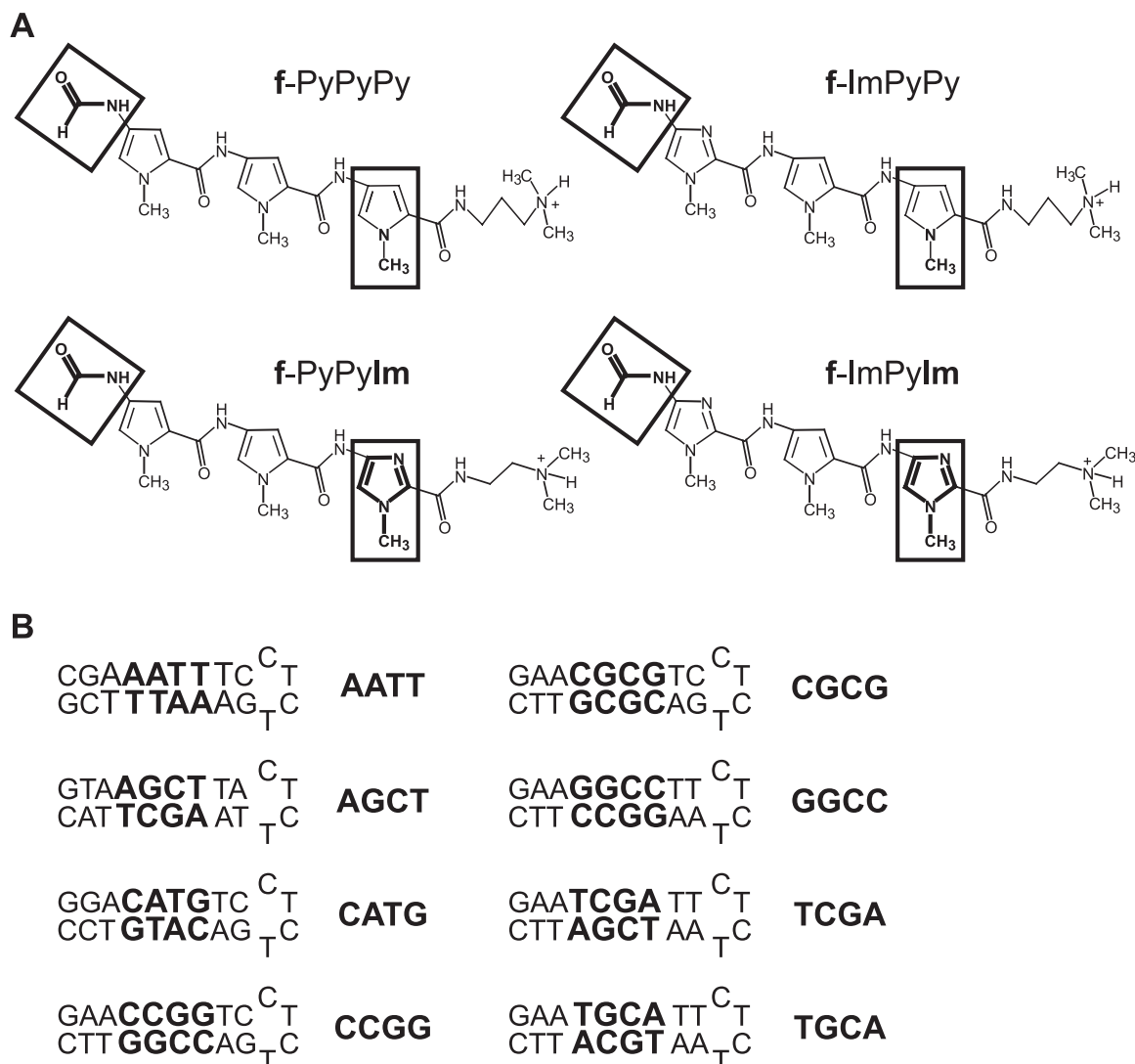
A Staggered **B** Overlapped

central pairings

A G G C C T   T G T C A
 f– Im Py Py–↻   Im Py Py–↻
 ↺–Py Py Im –f   ↺–Py Py Im
T C C G G A   A C A G T

terminal pairings

Figure 1. Staggered and overlapped orientations of triamide dimers that form within the DNA minor groove. (**A**) Staggered dimers are preferred by formylated triamides. In the staggered orientation, heterocycles stack (pair) to form the 'central pairing' (gray box). The central pairing is composed of two adjacent sets of stacked heterocycles. The C-terminal heterocyclic group is not included in the central pairing and is stacked on the formamido moiety, to form the 'terminal pairing'. (**B**) Overlapped dimers are preferred by non-formylated triamides. In the overlapped orientation, all three heterocycles are engaged in heterocycle–heterocycle stacking.

Figure 2. Polyamide DNA molecules used to study the sequence preference of formamido (f) terminal pairings. (**A**) The four triheterocyclic polyamides (triamides) with pyrrole (Py) and imidazole (Im) groups are f-PyPyPy, f-ImPyPy, f-PyPyIm and f-ImPyIm. The C-terminal heterocyclic moieties, which are involved in terminal pairings, are boxed. The formamido and C-terminal imidazole moieties are bold. (**B**) Nine different DNA hairpin molecules are named for their target sequences.

for the SPR experiments were biotinylated at the 5′-terminus. For CD and DNA melting studies, the oligonucleotides were not biotinylated. The oligonucleotides were suspended in TE buffer (10 mM Tris, 1 mM EDTA, pH 7.5) and stored at −20°C. Polyamides were synthesized as previously reported (15–16), and were homogeneous by 500 MHz ^{1}H-NMR analyses. Polyamides were resuspended in water containing one mole equivalent of HCl to stock concentrations of ∼300 μM and stored at 4°C.

Surface plasmon resonance

SPR experiments were performed using either a BIACORE 2000 or 3000 instrument (Biacore AB) as described previously with the DNA hairpins shown in Figure 2B (15,16,27,29). Steady state and kinetic data were analyzed as described previously (16). Experimental error is ±10% for k_a and most k_d and K_{eq} values. The error is larger, ±20%, for k_d values faster than 0.1 s^{-1} and K_{eq} below 5×10^5 M^{-1} and above 5×10^5 M^{-1}.

k_a and k_d cannot be determined when they are faster than 1 s^{-1}. Fitting errors are less than ±5% for K_{eq}, and are ∼25% for the individual K_1 and K_2 values due to the correlation of variables. Errors estimated from reproducibility are ±10% when K_{eq} or k_d values are between 5×10^5 and 5×10^7 M^{-1} or 0.1 and 0.001 s^{-1}, respectively. Errors increase to ±20% for K_{eq} values between 5×10^7 and 5×10^8 M^{-1} and for k_d between 0.1 and 1 s^{-1}. k_d values are difficult to determine when greater than 1 s^{-1} by biosensor-SPR methods. k_a values have ±10% errors.

Circular dichroism and DNA thermal melts

Experiments were performed as described previously (15–16). These experiments utilized 11 bp hairpin oligonucleotides that are simply extended by 2 bp (5′-CG . . . . . CG-3′) from the open end of the hairpin, and are otherwise identical to those shown in Figure 2 with the following exception: AATT (5′-GGC GAA ATT TC CTCT GA AAT TTC GCC), with the addition to the 9 bp DNA underlined. The circular dichroism data was

normalized (CD mdeg/positive peak height at λ_{max}) such that the peak height of the positive DNA band, in the absence of triamide, is comparable for all experiments.

RESULTS

Formamido/imidazole terminal pairing with the –ImPy– central pairing motif

Binding of f-ImPyIm to various DNA sequences was monitored by SPR. Sensorgram examples are shown in Figure 3A. Binding isotherms were fit to the steady-state data assuming a 2:1 triamide:DNA complex formation. Formation of a 2:1 f-ImPyIm to DNA complex is substantiated by a 2-fold higher response at saturation (RU_{sat}) than the calculated maximum response for 1:1 complex formation (RU_{max}) (16). Therefore, two binding constants were determined for each complex formed (K_1 and K_2), and the equilibrium constant (K_{eq}) is reported as the (K_1K_2)$^{1/2}$ in Table 1.

The K_{eq} for f-ImPyIm binding to TGCA, AGCT and GGCC are 1.2×10^7, 5.5×10^6 and 8.3×10^6 M^{-1}, respectively. The second molecule of f-ImPyIm binds TGCA and GGCC more tightly to the DNA than does the first ($K_2/K_1 \geqslant 10^4$); therefore, formation of each of these complexes exhibit very strong positive cooperativity. K_1 could not be delineated from K_2 for f-ImPyIm with AGCT and are, therefore, not reported. Binding of f-ImPyIm to DNA sequences that lack the –GC– recognition site necessary for the central –ImPy– pairing motif (CCGG and TCGA) were also studied. These DNA sequences exhibit considerably weaker binding affinities that are each

between two and three orders of magnitude lower than what was previously reported for f-ImPyIm binding to CGCG ($K_{eq} = 1.9 \times 10^8$ M^{-1}). f-ImPyIm also exhibits positively cooperative binding to CCGG and TCGA, with approximately one order of magnitude lower K_1 than K_2 for each DNA. The general finding that f-ImPyIm binds to these sequences with positive cooperativity is not surprising, according to previous work. DNA molecules with high GC content often have a wide (0.5–0.6 nm) minor groove (11), in which two polyamide molecules can readily stack without significant widening of the minor groove (16,27,29–31).

It is notable that f-ImPyIm has a lower binding affinity for DNA molecules that lack the –GC– central recognition site. For example, the TGCA and TCGA DNA sequences are each bound by f-ImPyIm, but TGCA is clearly saturated at a much lower polyamide concentration than is TCGA (Figure 4A).

Binding constants were also derived from the association (k_{a1} and k_{a2}) and dissociation (k_{d1} and k_{d2}) rate constants [$K_{eq} = ((k_{a1}/k_{d1})*(k_{a2}/k_{d2}))^{1/2}$]. These binding constants are in good agreement with the equilibrium constants from steady-state analysis (Table 2). K_{eq} of f-ImPyIm with TGCA and GGCC were both determined to be 1.7×10^7 M^{-1}, which are comparable to the K_{eq} observed by steady state, respectively. K_{eq} for f-ImPyIm with AGCT, CCGG and TCGA could not be determined because the association and dissociation rates were too fast, which is consistent with previous observations that faster kinetics correlate with lower binding affinities (15,23).

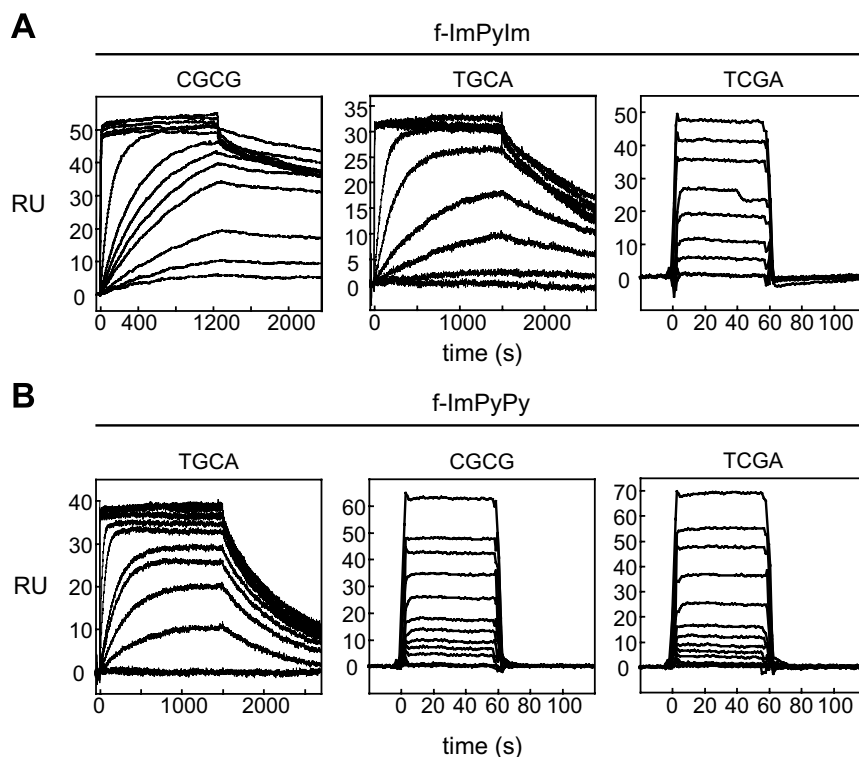f-ImPyIm binds to CGCG with 16-, 37- and 24-fold stronger affinity than it binds to TGCA, AGCT and GGCC,

Figure 3. Surface plasmon resonance (SPR) sensorgrams with f-ImPyIm and f-ImPyPy. CGCG, TGCA and TCGA DNA hairpins were titrated with up to 40 µM of f-ImPyIm (**A**) and f-ImPyPy (**B**).

Table 1. Binding constants (M^{-1}) and thermal stability of the complexes (ΔT_M)

Triamide	Technique	CGCG	TGCA	AGCT	GGCC	CCGG	TCGA
f-lmPylm	SPR (M^{-1})	$1.9 \times 10^{8a,b}$	1.2×10^7	5.5×10^6	8.3×10^6	2.2×10^{5b}	1.0×10^5
	ΔT_M (°C)	7.8^b	8.5	5.6	7.0	1.1^b	0.8
f-lmPyPy	SPR (M^{-1})	8.9×10^{4b}	7.4×10^{6c}	8.1×10^6	2.9×10^6	7.0×10^4	9.4×10^{4b}
	ΔT_M (°C)	2.0^b	11.0^b	8.5	2.6		2.5
Triamide	Technique	CATG	AATT	TCGA	CACG		
f-PyPylm	SPR (M^{-1})	4.4×10^{5b}	4.2×10^{5b}		9.4×10^4		
	ΔT_M (°C)	1.0^b	5.8^b	1.3	1.3		
f-PyPyPy	SPR (M^{-1})	$<5 \times 10^4$	3.2×10^{6c}				
	ΔT_M (°C)	0.1	9.3^b	1.0	-1.0		

[a]Determined from kinetic analysis.
[b,c]These values were taken from Refs (15,16), respectively.

Table 2. Kinetic rate constants derived from SPR

Compound	DNA sequence	$k_{a1}{}^a$ (M^{-1} s^{-1})	k_{d1} (s^{-1})	k_{a2} (M^{-1} s^{-1})	k_{d2} (s^{-1})	$K_{eq}{}^b$ (M^{-1}) kinetic	$K_{eq}{}^c$ (M^{-1}) steady state
f-ImPyIm	CGCG	5.9×10^4	0.017	1.1×10^6	1.1×10^{-4}	1.9×10^8	ND
f-ImPyIm	TGCA	1.2×10^5	96	1.1×10^8	5.1×10^{-4}	1.7×10^7	1.2×10^7
f-ImPyIm	GGCC	3.4×10^4	12	8.8×10^7	8.8×10^{-4}	1.7×10^7	8.3×10^6
f-ImPyIm	AGCT	$>10^6$	>1	$>10^6$	>1	ND	5.5×10^6
f-ImPyIm	CCGG	$>10^6$	>1	$>10^6$	>1	ND	2.2×10^5
f-ImPyIm	TCGA	$>10^6$	>1	$>10^6$	>1	ND	1.0×10^5
f-ImPyPy	TGCA	4.3×10^4	1.4	1.2×10^7	1.3×10^{-3}	1.7×10^7	7.4×10^6
f-ImPyPy	AGCT	4.0×10^5	61	4.3×10^7	1.5×10^{-3}	1.4×10^7	8.1×10^6
f-ImPyPy	GGCC	$>10^6$	>1	$>10^6$	>1	ND	2.9×10^6
f-ImPyPy	CGCG	$>10^6$	>1	$>10^6$	>1	ND	8.9×10^4
f-ImPyPy	CCGG	$>10^6$	>1	$>10^6$	>1	ND	7.0×10^4
f-ImPyPy	TCGA	$>10^6$	>1	$>10^6$	>1	ND	9.4×10^4
f-PyPyIm	CTAG	6.6×10^4	0.41	4.6×10^6	2.2×10^{-2}	5.9×10^5	1.2×10^6

ND, not determined.
[a]Data were fit as previously described in Ref. (21).
[b]K_{eq} is calculated directly from the kinetic analysis. $K_{eq} = (K_1 K_2)^{1/2} = [(k_{a1}/k_{d1}) \times (k_{a2}/k_{d2})]^{1/2}$.
[c]K_{eq} is calculated from steady-state measurements, see Table 1.
[d,e]These values were taken from Refs (15,16), respectively.
[f]Association and dissociation rates were too fast for the detection limits of BIACORE.
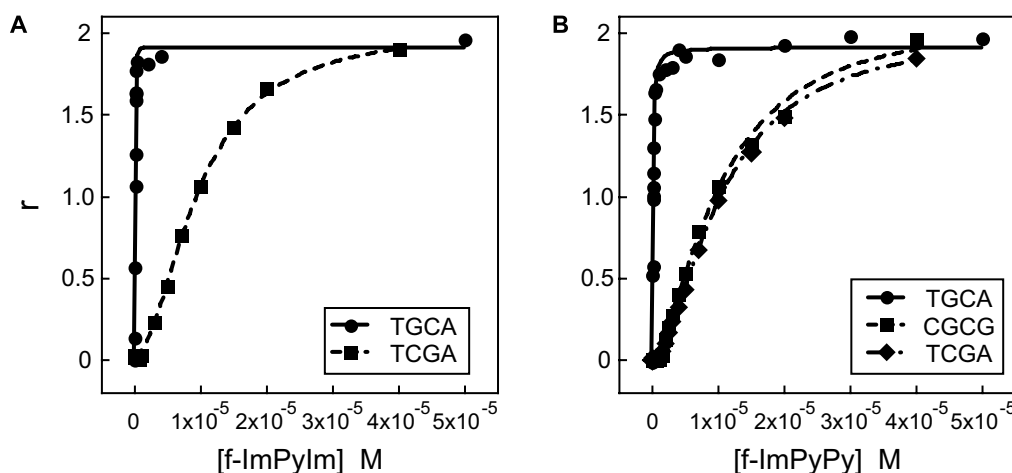
Figure 4. Steady-state analysis of f-ImPyIm (**A**) and f-ImPyPy (**B**). The SPR responses are normalized such that $r = RU_{sat}/RU_{max}$. Data were fit by $((K_1[\text{triamide}]) + (2K_1K_2[\text{triamide}]^2))/(1 + (K_1[\text{triamide}]) + (K_1K_2[\text{triamide}]^2))$; where triamide concentrations are reported in molarity and represent the free (unbound) concentration.

respectively. The sensorgrams in Figure 3A empirically show the significant slowing of the dissociation rate for the (f-ImPyIm)$_2$–CGCG complex compared to (f-ImPyIm)$_2$–TGCA. This slow dissociation of f-ImPyIm results in a higher

binding affinity for CGCG. These sequence-dependent variations in K_{eq} show that f-ImPyIm binds with significant affinity for CGCG over DNA sequences that also contain the central recognition motif (–GC–). Thus, the f/Im terminal pairing,

when adjacent to the –ImPy– central pairing, has the following DNA base pair preference: C/G > T/A > G/C ∼ A/T.

f/Im terminal pairing with the –PyPy– central pairing motif

The binding preference of f/Im was studied in the context of triamide f-PyPyIm with the CATG and AATT DNA hairpins. Steady-state analysis of the SPR data shows that there is identical affinity of f-PyPyIm to CATG and AATT ($K_{eq} = 4 \times 10^5$ M^{-1}). Binding of f-PyPyIm to CATG and AATT DNA hairpins exhibit similarly fast association and dissociation rates. Thus, in the context of the –PyPy– central pairing, the f/Im terminal pairing has little to no preference for C/G base pairs over A/T base pairs.

f/Py terminal pairing with the –ImPy– central pairing motif

The steady-state response of f-ImPyPy binding to the four GC-containing DNA hairpins (CGCG, TGCA, AGCT and GGCC) and two control DNA sequences (CCGG and TCGA) was monitored by SPR. The resulting sensorgrams were fit with a 2:1 triamide:DNA binding isotherm, and the RU$_{sat}$ responses for each complex were twice the values calculated for RU$_{max}$, indicating that f-ImPyPy binds as a dimer to the DNA sequences tested (sensorgram and steady-state fit examples shown in Figures 3B and 4B).

Binding affinities of f-ImPyPy to GGCC and CGCG are approximately 4- and 100-fold lower than binding to AGCT and TGCA, which exhibit nearly identical affinities ($K_{eq} \sim 8 \times 10^6$ M^{-1}) (Table 1). f-ImPyPy binds to the –CG– central recognition site containing CCGG and TCGA with K_{eq} lower than 10^5 M^{-1}. Binding constants derived from the kinetic rates for the association and dissociation of f-ImPyPy with TGCA or AGCT are consistent with those determined by steady state (Table 2). The kinetics were too fast to accurately establish the association and dissociation rates for f-ImPyPy binding to GGCC, CGCG and TCGA. The specificity of f-ImPyPy for TGCA over CGCG and TCGA can be empirically determined by the visual comparison of the slow association and even slower dissociation rates for binding to TGCA and the fast kinetics observed for CGCG and TCGA in Figure 3B. Together, the kinetic and steady-state analysis of the SPR data show that TGCA is a much better target than CGCG and TCGA.

The SPR analysis of f-ImPyPy with the four –GC– variants and the control DNA hairpins, CCGG and TCGA, show that f-ImPyPy is specific only for DNA that contain the –GC– central pairing sequence. The f/Py terminal pairing has the highest affinity for T/A and A/T base pairs and a 2- to 3-fold lower affinity for the G/C base pair. Interestingly, f/Py has a considerably reduced affinity for the C/G from the T/A or A/T base pairs (80- to 90-fold).

f/Py terminal pairing with the –PyPy– central pairing motif

f-PyPyPy has been previously shown to bind with good affinity as a dimer to AATT by SPR ($K_{eq} = 3.2 \times 10^6$ M^{-1}) (16). No SPR response is measured until titration of 30 µM f-PyPyPy to CATG and, therefore, the binding affinity must be

below 5×10^4 M^{-1}. Thus, f-PyPyPy has a considerably greater affinity for AATT over CATG, which is likely to be due to the higher affinity of the f/Py terminal pairings for the A/T base pair over the C/G base pair.

DNA thermal melting of the triamide–DNA complexes

DNA thermal melting experiments were performed to monitor the ability of the triamides to stabilize the temperature-dependent denaturation of double-stranded DNA (Table 1). The base-pair specificity of the f/Im terminal pairing was probed in the context of the –ImPy– central pairing. f-ImPyIm stabilized the CGCG, TGCA and GGCC DNA hairpins ($\Delta T_M = 7.8$, 8.5 and 7.0°C). (f-ImPyIm)$_2$–AGCT exhibits a slightly lower ΔT_M of 5.6°C and CCGG and TCGA complexed with f-ImPyIm each have negligible ΔT_M. Interestingly, thermal melting analysis of the (f-ImPyIm)$_2$–CGCG complex does not exhibit a higher ΔT_M than the (f-ImPyIm)$_2$–TGCA and (f-ImPyIm)$_2$–GGCC complexes, as would be expected from the SPR data.

The (f-ImPyPy)$_2$–TGCA and (f-ImPyPy)$_2$–AGCT complexes each have high ΔT_M (11.0 and 8.5°C, respectively). The (f-ImPyPy)$_2$–GGCC, (f-ImPyPy)$_2$–CGCG and (f-ImPyPy)$_2$–TCGA complexes exhibit significantly lower ΔT_M (2.6, 2.0 and 2.5°C). These ΔT_M values correlate well with the SPR analysis. f-ImPyPy has low affinities for CGCG and TCGA, which contain the C/G terminal base pair and the –CG– central pairing recognition sites, respectively. Interestingly, by SPR f-ImPyPy binds GGCC with only a 2- to 3-fold reduction in affinity compared to TGCA and AGCT; however, the thermal melting experiments show no improved stability for the (f-ImPyPy)$_2$–GGCC complex over the (f-ImPyPy)$_2$–CGCG and (f-ImPyPy)$_2$–TCGA complexes.

–PyPy– containing triamides, f-PyPyPy and f-PyPyIm, increase the stability of the AATT DNA ($\Delta T_M = 5.8$ and 9.3°C, respectively), but do not increase the stability of CATG or TCGA (Table 1). With the exception of the (f-PyPyIm)$_2$–CATG complex, this is in good agreement with the SPR data. The slightly different trends observed between thermal melts and SPR for some of these triamides is not contradictory because binding affinities can exhibit significant temperature dependence and the thermal melting experiments inherently probe the complexes at a higher temperature (above 50°C) than do SPR and CD experiments (24–25°C). Therefore, slight differences in the relative affinity of the polyamide for the DNA are probably due to temperature.

Circular dichroism analysis of the triamide–DNA complexes

The –GC– central recognition site containing DNA hairpins were investigated by circular dichroism (CD) as a function of titrated f-ImPyIm and f-ImPyPy (Figure 5A). The triamides are not chiral and, therefore, CD of these compounds do not result in any peaks. However, a peak was induced at ∼320 nm upon titration of the triamide. This peak is indicative of the compounds binding in the minor groove of the DNA (32,33). The compound was titrated until no change was seen in the spectra, which indicates that the DNA binding sites were saturated. Visual comparison of each of the spectra in Figure 5A demonstrates that the peak induced at saturating concentrations of f-ImPyIm to CGCG is greater than any of the other
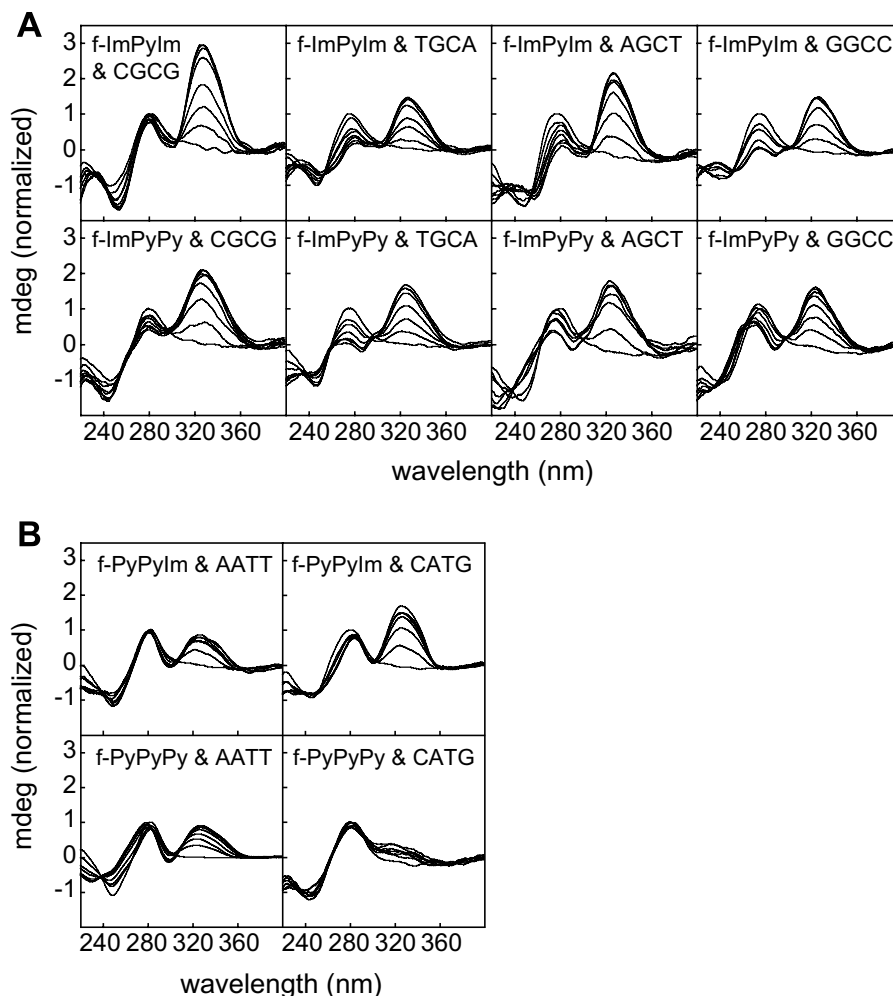
Figure 5. Comparative circular dichroism experiments of the triamides with different terminal pairing iterations. (**A**) f-ImPyIm and f-ImPyPy binding to four –GC– containing DNA sequences. (**B**) f-PyPyIm and f-PyPyPy binding to AATT and CATG, which contain the –AT– recognition site. Data were normalized (CD mdeg/positive peak height at λ_{max}).

induced peaks. This exceptional response may be correlated with the high binding affinity observed by SPR for f-ImPyIm with CGCG ($K_{eq} = 1.9 \times 10^8$ M^{-1}). The other three –GC– DNA sequences are also bound with good affinity by f-ImPyIm ($K_{eq} = 10^6$ to 10^7 M^{-1}), which may explain the good response upon addition of f-ImPyIm in the CD experiments. The four –GC– sequences show approximately the same response at saturating f-ImPyPy (lower panels, Figure 5A). One possible explanation for this CD data is that the –GC– recognition site and the –ImPy– central pairing motif are structurally well aligned, resulting in strong CD signals, even when binding is weak, as in the case of f-ImPyPy and CGCG. Ultimately, the DNA region and the large induced peak height may point to remarkable structural features of f-ImPyIm, CGCG and their complex.

Titration of f-PyPyIm and f-PyPyPy to AATT and CATG also show that these triamides bind in the DNA minor groove and f-PyPyPy shows negligible response upon addition to CATG. These CD spectra correlate with the SPR and ΔT_M data.

DISCUSSION

Herein, we have described a systematic study of the DNA sequence specificity of f/Im and f/Py pairings in polyamides. Together, results from surface plasmon resonance, circular dichroism and thermal melting experiments provided new insight into using formamido groups in polyamide design. This study was necessitated by the discovery that formamido groups enhance the binding affinity by one to two orders of magnitude in addition to extending the binding site size over otherwise identical non-formylated polyamides (16). Previous work showed that the –ImPy– and –PyPy– central pairing elements have higher affinity and specificity for their Watson–Crick cognate sequences than do –PyIm– and –ImIm– (15); therefore, –PyIm– and –ImIm– were not used in this study. New rules, pertaining to formamido groups, are now added to those already necessary in polyamide design (5,15).

The f/Py terminal pairing was studied using the f-ImPyPy and f-PyPyPy triamides and DNA hairpins with symmetric recognition sites (Figure 2). The central pairing recognition sites, –GC– and –AT–, were held constant for their respective

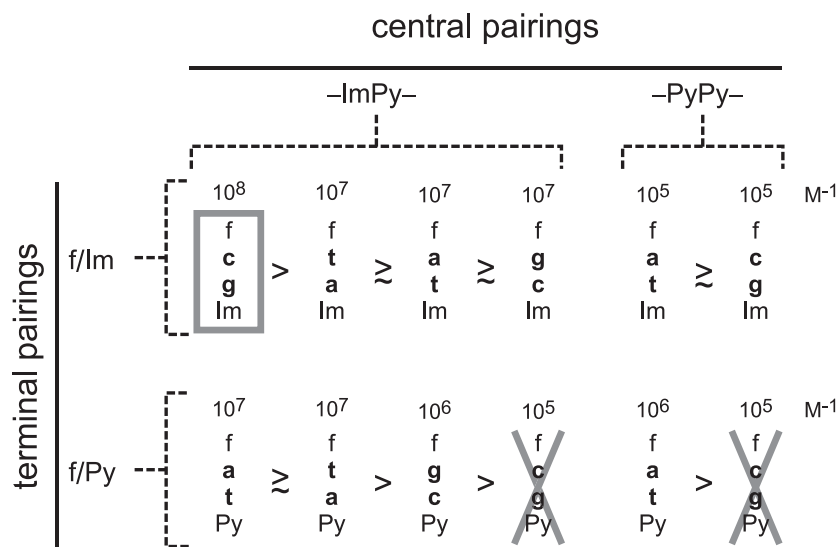**Figure 6.** Recognition scheme of the f/Im and f/Py pairings for Watson–Crick base pairs. Also depicted in this figure is the range of binding constants (10^5–10^8 M^{-1}) for the f/Im and f/Py terminal pairings when placed adjacent to –ImPy– and –PyPy– central pairings. The gray box indicates selective binding by f/Im to C/G and Xs denote the poor match of f/Py and the C/G base pair.

central pairing polyamide motifs. In the context of the –ImPy– central pairing, the f/Py pairing preferred A/T and T/A base pairs by 2- to 3-fold higher affinity over the G/C base pair (Table 1). Interestingly, f/Py had very low affinity for the C/G base pair (∼100 fold lower than A/T) regardless of the central pairing (Figure 6). Therefore, f/Py pairings behave much like Py/Py pairings in their preference for A/T and T/A base pairs. However, the f/Py pairings do have some tolerance for the G/C base pairs. The poor binding of f/Py to C/G base pairs is not surprising, because placing the pyrrole of the f/Py pair towards the N2 position of the guanine results in steric hindrance with the exocyclic amino group, much like that expected for Py/Py and C/G base pairs (13,14,20). The terminal f/Py pairings behave much like the terminal Py/Py of non-formylated triamides. It was previously shown that the non-formamido PyPyPy compound interacted with a 9 bp, AT-rich DNA hairpin, CGAAATTTC, with ∼30-fold greater affinity over the GAACTGGTC DNA hairpin. Binding to the second sequence must place at least one of the Py/Py terminal pairings opposite a G/C or C/G base pair, which is not an ideal binding sequence for nf-PPP (16). The findings provided herein are also in good agreement with previous studies on distamycin A and its formylated oligopyrrole analogs, which proved that these compounds strongly prefer AT-over GC-rich DNA sequences and, consequently, the f/Py pairings must prefer the A/T and T/A base pairs (11,25,26). Therefore, our results are consistent with prior work and systematically show that for polyamide design purposes, f/Py should be used to recognize A/T or T/A base pairs.

f-ImPyIm and f-PyPyIm were used to study the sequence specificity of the f/Im terminal pairing motif. As stated above, the central pairing recognition sites were held constant for the appropriate central pairing motif. Kinetic analysis of the SPR experiments demonstrated that the f/Im pairing in the f-ImPyIm triamide was specific by at least one-order of magnitude for the C/G base pair over the T/A, A/T and G/C base pairs (Figure 6). Interestingly, within this series, the

formamido group has a preference for pyrimidines over purines. When f/Im and –ImPy– are part of the same triamide, then f/Im has strong selectivity for C/G over the other three base pairs. In this setting, the imidazole of f/Im is most likely placed such that a favorable hydrogen bond is formed with the N2 of guanine (13,20). Prior work showed that the Im/Py pairings of the non-formylated ImPyPy recognize the G/C base pairs of a GTC containing DNA hairpin with ∼25-fold better affinity than DNA sequences that did not contain either GTC or GAC (15). With these sequences, the Im/Py pairings would be placed opposite C/G, A/T or T/A base pairs, and these results are similar to our results with f-ImPyIm suggesting that the f/Im behaves like a terminal Py/Im pairing (15).

In contrast, the f/Im pairing of the f-PyPyIm triamide had no selectivity for the C/G base pair over the A/T base pair. Thus, base pair specificity of the f/Im terminal pairing appears to be more complex than is the f/Py pairing. Therefore, inclusion of f/Im in polyamide design should be used with care, but with the right target DNA and polyamide content it can demonstrate significant selectivity for the C/G base pair.

This research has further demonstrated the impressive binding affinity and sequence selectivity of f-ImPyIm for its cognate sequence, CGCG. f-ImPyIm has the same shape and molar mass as Distamycin A (differs only by 3 g/mol). The heterocyclic content and, therefore, their cognate sequences are the main differences between these two triamides; however, f-ImPyIm recognizes CGCG with approximately one order of magnitude better affinity than mother nature recognizes AATT with distamycin A (15). In addition, f-ImPyIm has good selectivity for CGCG over similar sequences that also contain the –GC– recognition site (TGCA, AGCT and GGCC), or are GC rich (GGCC and CCGG). The –ImPy– motif and its –GC– recognition site cannot be the only reason f-ImPyIm binds so tightly to CGCG because f-ImPyPy binds its cognate, AGCT, with approximately one-order of magnitude lower affinity (15). f/Im, when partnered with

the –ImPy– central pairing, both increases binding affinity to DNA and improves specificity for a single DNA sequence.

Several factors are probably influential in the strong affinity and high specificity of f-ImPyIm for CGCG, these may include DNA structure, polyamide structure, DNA dynamics and complex conformation. Base pair sequence influences DNA structure; for example, 5′-CpG-3′ and 5′-GpC-3′ steps can result in significantly altered DNA (34,35) and their conformations are highly dependent on the neighboring DNA base pairs (36). Interestingly, the conformation of CGCG and CCGG containing DNA are essentially identical (37,38). If DNA is also a component of the uniqueness of f-ImPyIm recognizing CGCG, then differences must arise from subtleties, such as flexibility or hydration, that are not immediately apparent. Structural and energetic studies on f-ImPyIm, CGCG and their complex are underway to provide a better understanding of this impressive complex.

ACKNOWLEDGEMENTS

REFERENCES

1. Wemmer,D.E. (2000) Designed sequence-specific minor groove ligands. *Annu. Rev. Biophys. Biomol. Struct.*, **29**, 439–461.
2. Bailly,C. and Chaires,J. (1998) Sequence-specific DNA minor groove binders. Design and synthesis of netropsin and distamycin analogues. *Bioconjug. Chem.*, **9**, 513–538.
3. Dervan,P.B. (2001) Molecular recognition of DNA by small molecules. *Bioorg. Med. Chem.*, **9**, 2215–2235.
4. Uil,T.G., Haisma,H.J. and Rots,M.G. (2003) Therapeutic modulation of endogenous gene function by agents with designed DNA-sequence specificities. *Nucleic Acids Res.*, **31**, 6064–6078.
5. Dervan,P.B. and Edelson,B.S. (2003) Recognition of the DNA minor groove by pyrrole-imidazole polyamides. *Curr. Opin. Struct. Biol.*, **13**, 284–299.
6. Yang,X.-L. and Wang,A.H.-J. (1999) Structural studies of atom-specific anticancer drugs acting on DNA. *Pharmacol. Ther.*, **83**, 181–215.
7. Yang,X.-L., Hubbard,R.B., Lee,M., Tao,Z.-F., Sugiyama,H. and Wang,A.H.-J. (1999) Imidazole–imidazole pair as a minor groove recognition motif for T:G mismatched base pairs. *Nucleic Acids Res.*, **27**, 4183–4190.
8. Edayathumangalam,R.S., Weyermann,P., Gottesfeld,J.M., Dervan,P.B. and Luger,K. (2004) Molecular recognition of the nucleosomal 'supergroove'. *Proc. Natl Acad. Sci. USA*, **101**, 6864–6869.
9. Pelton,J.G. and Wemmer,D.E. (1989) Structural characterization of a 2:1 distamycin A. d(CGCAAATTGGC) complex by two-dimensional NMR. *Proc. Natl Acad. Sci. USA*, **86**, 5723–5727.
10. Dwyer,T.J., Geierstanger,B.H., Bathini,Y., Lown,J.W. and Wemmer,D.E. (1992) Design and binding of a distamycin A analog to d(CGCAAGTTGGC)·d(GCCAACTTGCG): synthesis, NMR studies, and implications for the design of sequence-specific minor groove binding oligopeptides. *J. Am. Chem. Soc.*, **114**, 5911–5919.
11. Pelton,J.G. and 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.
12. Kopka,M.L., Yoon,C., Goodsell,D., Pjura,P. and Dickerson,P.E. (1985) The molecular origin of DNA-drug specificity in netropsin and distamycin. *Proc. Natl Acad. Sci. USA*, **82**, 1376–1380.
13. Lown,J.W., Krowicki,K., Bhat,U.G., Skorobogaty,A., Ward,B. and Dabrowiak,J.C. (1996) Molecular recognition between oligopeptides and nucleic acids: novel imidazole-containing oligopeptides related to netropsin that exhibit altered DNA sequence specificity. *Biochemistry*, **25**, 7408–7416.
14. Lee,M., Rhodes,A.L., Wyatt,M.D., Forrow,S. and Hartley,J.A. (1993) GC base sequence recognition by oligoimidazolecarboxamide and C-terminus modified analogues of distamycin deduced from CD, ¹H-NMR and MPE footprinting studies. *Biochemistry*, **32**, 4237–4245.
15. Buchmueller,K.L., Staples,A.M., Howard,C.M., Horick,S.M., Uthe,P.B., Le,N.M., Cox,K.K., Nguyen,B., Pacheco,K.A.O., Wilson,W.D. and Lee,M. (2005) Extending the language of DNA molecular recognition by polyamides: unexpected influence of imidazole and pyrrole arrangement on binding affinity and specificity. *J. Am. Chem. Soc.*, **127**, 742–750.
16. Lacy,E.R., Le,N.M., Price,C.A., Lee,M. and Wilson,W.D. (2002) Influence of a terminal formamido group on the sequence recognition of DNA by polyamides. *J. Am. Chem. Soc.*, **124**, 2153–2163.
17. Marques,M.A., Doss,R.M., Urbach,A.R. and Dervan,P.B. (2002) Toward an understanding of the chemical etiology for DNA minor-groove recognition by polyamides. *Helv. Chim. Acta*, **85**, 4485–4517.
18. Renneberg,D. and Dervan,P.B. (2003) Imidazopyridine/pyrrole and hydroxybenzimidazole/pyrrole pairs for DNA minor groove recognition. *J. Am. Chem. Soc.*, **125**, 5707–5716.
19. Wellenzohn,B., Loferer,M.J., Trieb,M., Rauch,C., Winger,R.H., Mayer,E. and Liedl,K.R. (2003) Hydration of hydroxypyrrole influences binding of ImHpPyPy-beta-Dp polyamide to DNA. *J. Am. Chem. Soc.*, **125**, 1088–1095.
20. Kissinger,K., Krowicki,K., Dabrowiak,J.C. and Lown,J.W. (1987) Molecular recognition between oligopeptides and nucleic acids: monocationic imidazole lexitropsins that display enhanced GC sequence dependent DNA binding. *Biochemistry*, **26**, 5590–5595.
21. Dervan,P.B. and Burli,R.W. (1999) Sequence-specific DNA recognition by polyamides. *Curr. Opin. Chem. Biol.*, **3**, 688–693.
22. Wemmer,D.E. and Dervan,P.B. (1997) Targeting the minor groove of DNA. *Curr. Opin. Struct. Biol.*, **7**, 355–361.
23. White,S., Baird,E.E. and Dervan,P.B. (1997) On the pairing rules for recognition in the minor groove of DNA by pyrrole–imidazole polyamides. *Chem. Biol.*, **4**, 569–578.
24. Hawkins,C.A., Peláez de Clairac,R., Dominey,R.N., Baird,E.E., White,S., Dervan,P.B. and Wemmer,D.E. (2000) Controlling binding orientation in hairpin polyamide DNA complexes. *J. Am. Chem. Soc.*, **122**, 5235–5243.
25. Marky,L.A., Younquist,R.S., Dervan,P.B., Arcamone,F. and Breslauer,K.J. (1987) Length dependent thermodynamics of minor groove binders. *Proc. of 5th Conversation in Biomolec. Stereodynamics.*, State University of New York at Albany, Academic Press, NY, pp. 198–199.
26. Luck,G., Zimmer,C., Reinert,K.-E. and Arcamone,F. (1977) Specific interactions of Distamycin A and its analogs with (A·T) rich and (G·C) rich duplex regions of DNA and deoxypolynucleotide. *Nucleic Acids Res.*, **4**, 2655–2670.
27. Lacy,E.R., Cox,K.K., Wilson,W.D. and Lee,M. (2002) Recognition of T∗G mismatched base pairs in DNA by stacked imidazole-containing polyamides: surface plasmon resonance and circular dichroism studies. *Nucleic Acids Res.*, **30**, 1834–1841.
28. Kopka,M.L., Goodsell,D.S., Han,G.W., Chiu,T.K., Lown,J.W. and Dickerson,R.E. (1997) Defining GC-specificity in the minor groove: side-by-side binding of the di-imidazole lexitropsin to C-A-T-G-G-C-C-A-T-G. *Structure*, **5**, 1033–1046.
29. Lacy,E.R., Nguyen,B., Le,M., Cox,K.K., O'Hare,C., Hartley,J.A., Lee,M. and Wilson,W.D. (2004) Energetic basis for selective recognition of T∗G mismatched base pairs in DNA by imidazole-rich polyamides. *Nucleic Acids. Res.*, **32**, 2000–2007.
30. Yang,X.-L., Kaenzig,C., Lee,M. and Wang,A. (1999) Binding of AR-1-144, a tri-imidazole DNA minor groove binder, to CCGG sequence analyzed by NMR spectroscopy. *Eur. J. Biochem.*, **263**, 646–655.
31. Lee,M., Krowicki,K., Hartley,J.A., Pon,R.T. and Lown,J.W. (1988) Molecular recognition between oligopeptides and nucleic acids: influence of van der Waals contacts in determining the 3′-terminus of DNA sequences read by monocationic lexitropsins. *J. Am. Chem. Soc.*, **110**, 3641–3649.
32. Lyng,R., Rodger,A. and Norden,B. (1991) The CD of ligand–DNA systems. 1. Poly (dG-dC) B-DNA. *Biopolymers*, **31**, 1709–1720.
33. Lyng,R., Rodger,A. and Norden,B. (1992) The CD of ligand–DNA systems. 2. Poly (dA-dT) B-DNA. *Biopolymers*, **32**, 1201–1214.
34. Mauffret,O., Monnot,M., Lanson,M., Armier,J. and Fermandjian,S. (1989) Conformational variations in d(TGACGTCA) and its reverse sequence d(ACTGCAGT): a joint circular dichroism and nuclear

magnetic resonance study. *Biochem. Biophys. Res. Commun.*, **165**, 602–614.

35. Lefebvre,A., Fermandjian,S. and Hartmann,B. (1997) Sensitivity of NMR internucleotide distances to B-DNA conformation: underlying mechanics. *Nucleic Acids Res.*, **25**, 3855–3862.

36. Packer,M.J., Dauncey,M.P. and Hunter,C.A. (2000) Sequence-dependent DNA structure: tetranucleotide conformational maps. *J. Mol. Biol.*, **295**, 85–103.

37. Isaacs,R.J., Rayens,W.S. and Spielmann,H.P. (2002) Structural differences in the NOE-derived structure of G–T mismatched DNA relative to normal DNA are correlated with differences in (13)C relaxation-based internal dynamics. *J. Mol. Biol.*, **319**, 191–207.

38. Timsit,Y., Vilbois,E. and Moras,D. (1991) Base-pairing shift in the major groove of (CA)n tracts by B-DNA crystal structures. *Nature*, **354**, 167–170.