Results and Discussion

The Importance of β-Alanine for Recognition of the Minor Groove of DNA¹

Purpose. The recent finding by Laemmli and coworkers that the β-linked polyamide Im-β-ImPy-β-Im-β-Im-β-Dp binds GAGAA tracks in a 1:1 stoichiometry and in a single orientation raises some interesting questions regarding the nature of this binding mode (Janssen et al., 2000). The favorable Im-β-Im composition (Turner et al., 1998) is present in their polyamide, leading us to ponder its importance and generality. On one hand, the 1:1 complex is an important observation by Laemmli, which could expand the sequence repertoire for DNA targeting. On the other hand, the fact that β-linked Py/Im polyamides can bind *both* 1:1 and 2:1 in the minor groove raises the issue that a single polyamide molecule may bind very different sequences depending on stoichiometry (Figure 10).

Approach. To explore the effects of β linkage at different polyamide positions, we synthesized three polyamides: Im**Py**ImPy- β -Im**Py**ImPy- β -Dp (**1**) and Im- β -ImPy- β -Im- β -ImPy- β -Dp (**2**), which vary in their Im-**X**-Im composition (X= Py or β); and Im- β -ImPy**Py**Im- β -ImPy- β -Dp (**3**), wherein an internal β residue is changed to Py (Figure 11). To ask if a single ligand can target different sequence contexts depending on stoichiometry, footprinting and affinity cleavage experiments were performed for **1**-**3** at designed 2:1 and 1:1 sites, 5'-AGCGCAGCGCT-3' and 5'-AAGAGAAGAG-3', respectively.

¹ The text of this section is taken from Dervan and Urbach, 2001.



Figure 10 Stoichiometry-dependent sequence targeting. Polyamide **2**, shown here, can bind in 2:1 (A) and 1:1 (B) ligand:DNA complexes, specifying the very different sequences, 5'-AGCGCAGCGCT-3 and 5'-AAGAGAAGAG-3', respectively.



Figure 11 (A) Selective $Py \rightarrow \beta$ substitutions in polyamides 1, 2, and 3 examined in this study. (B) Chemical structures of these compounds and their EDTA conjugates, 1E, 2E, and 3E. Polyamide sequences are indicated beneath each structure.

Synthesis. Polyamides 1-3 were synthesized in 12 steps from Boc- β -Pamresin (1.25 g resin, 0.26 mmol/g substitution) using previously described solidphase methods (Baird and Dervan, 1996). Nonterminal imidazole residues were introduced as dimers, Boc-Py-Im-COOH and Boc- β -Im-COOH. The polyamide was cleaved from the solid support by aminolysis of the resin ester linkage using dimethylaminopropylamine (Dp) or 3,3'-diamino-N-methyldipropylamine (Dp-NH₂). Products were purified by reversed-phase preparatory HPLC to provide ImPyImPy- β -ImPyImPy- β -Dp (1), ImPyImPy- β -ImPyImPy- β -Dp-NH₂ (1-NH₂), Im- β -ImPy- β -Im- β -ImPy- β -Dp (2), Im- β -ImPy- β -ImPy- β -Dp-NH₂ (2-NH₂), Im- β -ImPyPyIm- β -ImPy- β -Dp (3), and Im- β -ImPyPyIm- β -Dp-NH₂ (3-NH₂). The purified polyamides with primary amines at the C terminus were treated with an excess of EDTA dianhydride (DMSO/NMP, DIEA, 55 °C, 15 min) and the remaining anhydride was hydrolyzed (0.1 N NaOH, 55 °C, 10 min). The EDTA conjugates were then purified by reversed-phase preparatory HPLC to yield **1E**, **2E**, and **3E**.

MPE•Fe(II) Footprinting and Affinity Cleaving. The plasmid pAU9 was constructed, which contains two binding sites, 5'-AGCGCAGCGCT-3' and 5'-AAGAGAAGAG-3', in order to examine 2:1 and 1:1 polyamide:DNA binding, respectively. MPE•Fe(II) footprinting (Van Dyke and Dervan, 1983) on the 3'and 5'-³²P end-labelled 253 base pair *EcoRI/Pvull* restriction fragment (3') or PCR product (5') from plasmid pAU9 (28.6 mM HEPES, 285.7 mM NaCl, pH 7.0, 22°C) reveals that polyamides **2** and **3**, each at 100 nM concentration, bind the designed sites 5'-AGCGCAGCGCT-3' and 5'-AAGAGAAGAG-3' (Figure 12). No footprinting is observed for polyamide **1** at concentrations <1 μ M.

Affinity cleaving experiments (Schultz and Dervan, 1984) on the same DNA fragments (28.6 mM HEPES, 285.7 mM NaCl, pH 7.0, 22 °C) with the EDTA•Fe(II) analogs **1E**, **2E**, and **3E** reveals 3'-shifted cleavage patterns indicating minor groove binding for polyamides **2E** (at 10 nM) and **3E** (at 100 nM) (Figure 12-A). Polyamides **2E** and **3E** cleave both sides of site,

Figure 12 (A) MPE footprinting (polyamides 2 and 3) and affinity cleaving experiments (2E and 3E) on the 5'-end-labelled 253 bp PCR product from pAU9: lane 1, intact DNA; lane 2, G reaction; lane 3, A reaction; lane 4, MPE standard; lanes 5-7, 10 nM, 30 nM, and 100 nM 2, respectively; lanes 8-10, 10 nM, 30 nM, and 100 nM 3, respectively; lanes 11-13, 1 nM, 3 nM, and 10 nM 2E, respectively; lanes 14-16, 10 nM, 30 nM, and 100 nM 3E, respectively. Relative cleavage intensities and MPE footprints for each binding site are illustrated at right by arrows and boxes, respectively. (B, top) Quantitative DNase I footprint titration experiments on the 5'-end-labelled 253 bp PCR product from pAU9 (left) with Im- β -ImPy- β -Im- β -ImPy- β -Dp, **2**: lane 1, intact DNA; lane 2, G reaction; lane 3, A reaction; lane 4, DNase I standard; lanes 5-15, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, and 100 nM 2, respectively; and (right) with Im-\beta-ImPvPvIm- β -ImPv- β -Dp, **3**: lane 1, intact DNA; lane 2, G reaction; lane 3, A reaction; lane 4, DNase I standard; lanes 5-15, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, 100 nM, 300 nM, and 1 µM 3, respectively. 2:1 and 1:1 binding sites are shown boxed at the right of each gel. (bottom) Binding isotherms for each DNase I footprint titration experiment directly above. The data points for the 5'-AGCGCAGCGCT-3' (2:1) site are indicated by filled circles and the 5'-GAGAAGAGAA-3' (1:1) site by open circles. The solid lines are best fit Langmuir binding isotherms obtained by a nonlinear least-squares algorithm, as described in the Experimental.



5'-AGCGCAGCGCT-3', consistent with binding as an antiparallel dimer. However, **2E** and **3E** reveal cleavage only at the 5'-end of the site, 5'-AAGAGAAGAG-3', suggesting one orientation and a 1:1 polyamide:DNA binding stoichiometry, similar to the finding of Laemmli (Janssen et al., 2000). No cleavage is observed for polyamide **1E** at concentrations $\leq 1 \mu$ M.

Quantitative DNase I Footprint Titrations. Quantitative DNase I footprint titration experiments (Brenowitz et al., 1986; Trauger and Dervan, 2001) were performed to determine the apparent equilibrium dissociation constants (K_D = concentrations of polyamide bound at half-saturation) of **1-3** at each of the target sites (Figure 12-B). The 5'-AGCGCAGCGCT-3' site was bound by **2** and **3** with similar affinities, $K_D = 0.63$ nM and 1.9 nM, respectively, displaying cooperative binding isotherms (n=2) for both molecules, consistent with 2:1 binding. However, the 5'-AAGAGAAGAG-3' site was bound 16-fold more tightly by polyamide **2** ($K_D = 0.02$ nM) than polyamide **3** ($K_D = 0.32$ nM), displaying noncooperative binding isotherms (n=1) for both molecules, consistent with 1:1 stoichiometry. Polyamide **1** bound with poor affinity and at concentrations ≥100 nM in a nonspecific manner (Table 1).

		_		
Polyamide 1	Polyamide 2	Polyamide 3		
2:1				
5'-A G C G C A G C G C T-3'	5'-A G C G C A G C G C T-3'	5'-A G C G C A G C G C T-3'		
+>♦○●○●◆◇●○●	$+) \diamond \bigcirc \bigcirc \diamond \odot \diamond \bigcirc \diamond \bigcirc \bigcirc \diamond \bigcirc \bigcirc \bigcirc \bigcirc \bigcirc \bigcirc \bigcirc \bigcirc \bigcirc$	+)♦○●♦●○○●♦●		
3'-TCGCGTCGCGA-5'	3'-TCGCGTCGCGA-5'	3'-TCGCGTCGCGA-5'		
≥100 nM	0.63 nM (±0.08)	1.9 nM (±0.1)		
1:1				
5'-A A G A G A A G A G-3'	5'-A A G A G A A G A G-3'	5'-AAGAGAAGAG-3'		
+)♦○●○●◆●○●○●	+)♦○●♦●♦○●♦●	+)♦○●♦●○○●♦●		
3'-TTCTCTTCTC-5'	3'-TTCTCTTCTC-5'	3'-TTCTCTTCTC-5'		
≥100 nM	0.020 nM (±0.002)	0.32 nM (±0.04)		

Table 1 Apparent Equilibration Dissociation Constants (K_D)

Discussion. Polyamides **2** and **3**, which preserve the Im-β-Im unit, bind DNA sites of similar size (10 and 11-base pairs), but remarkably different sequence compositions, which is related to the stoichiometry of complexation in the minor groove. In the original "lexitropsin" model, based on 1:1 binding of polyamides to DNA, Im was proposed to specify $G \cdot C/C \cdot G > A \cdot T/T \cdot A$ (Kopka et al., 1985). We now know this not to be the case. In the study detailed in the next section, Im, in the sequence context of polyamide **2**, is shown to bind all four base pairs (within a factor of 2) (Urbach and Dervan, 2001). From the crystal structure of the 1:1 netropsin:DNA complex, we understand the molecular mechanism by which Py specifies $A \cdot T/T \cdot A > G \cdot C/C \cdot G$ (Kopka et al., 1985). However, this 1:1 recognition code of Py selecting $A \cdot T/T \cdot A > G \cdot C/C \cdot G$ must now be modified to include the judicious placement of β for $A \cdot T/T \cdot A$ recognition to reset the curvature in the 1:1 motif. This is substantiated by our observation that Im-β-Im binds 5′-GAG -3′ with higher affinity than Im-Py-Im (Table 1).

There are several implications for the addition of a 1:1 motif to current 2:1 targeting of mixed G,C/A,T sequences of DNA in the minor groove. Assuming cell permeability is related to size, one could argue that a 1:1 binding molecule will occupy larger binding site sizes for the same molecular weight as a hairpin polyamide. In fact, the difference of one methylene unit (β vs. γ) can direct whether a polyamide binds extended as 1:1 complex or folds as a hairpin (Mrksich et al., 1994; Trauger et al., 1996a). We face the added complexity of the same molecule binding very different DNA sites depending on stoichiometry. If specificity is the goal, this raises the challenge for chemists to design next-generation polyamides, which enforce 2:1 vs. 1:1 binding in the minor groove.

Toward Rules for 1:1 Polyamide:DNA Recognition²

Purpose. At the forefront of the endeavor to control gene expression by small molecules is the elucidation of chemical principles for direct read-out of predetermined sequences of double-stranded DNA. Although there exists a significant body of literature on the 1:1 mode of binding (Lown et al., 1986), much of this was carried out before quantitative footprinting methods were introduced to the field. In an effort to characterize more rigorously the 1:1 motif, we address several questions quantitatively: 1) Can a 1:1 recognition code be established, which uses individual ligand residues (e.g. Py, Im, Hp or β) to specify individual Watson•Crick base pairs? 2) How is orientation related to overall DNA sequence type? and 3) What is the effect of ligand size on binding affinity in the 1:1 motif?

Specificity of Py, Im, Hp and β

Approach. The polyamide Im- β -ImPy- β -Im- β -ImPy- β -Dp (2) was chosen as the template to examine the specificity at Im, Py and β residues in an oriented 1:1 complex with DNA. Because Hp/Py pairs were shown to discriminate T•A from A•T in the 2:1 motif, a second ligand Im- β -ImHp- β -Im- β -ImPy- β -Dp (4) was prepared to explore any possible specificity the Hp residue may have in a 1:1 complex. Specificity at a single and unique carboxamide position was determined by varying a single base pair within the parent sequence context, 5'-AAAGAGAAGAG-3', to all four Watson•Crick base pairs and comparing the relative affinities for the four possible complexes. To meet this end, three

² The text of this section is taken from Urbach and Dervan, 2001.



Figure 13 Examination of sequence selectivity at a single imidazole (Im), beta alanine (β), pyrrole (Py), or hydroxypyrrole (Hp) position within the parent context, 5'-AAAGAGAAGAG-3'. Imidazole and pyrrole rings are represented as shaded and nonshaded circles, respectively; β -alanines are shown as gray diamonds; and hydroxypyrrole is indicated by a circle containing the letter H.

plasmids were cloned, each containing four binding sites: pAU8 (for Im), 5'-AAAGAXAAGAG-3'; pAU15 (for β), 5'-AAAGXGAAGAG-3'; and pAU12 (for Py and Hp) 5'-AAAGAGAGAXGAG-3', where **X** = A, T, G, and C (Figure 13). The Hp-containing polyamide (**4**) was synthesized by solid phase methods, as described previously (Urbach et al., 1999).

DNA Binding Affinity and Sequence Specificity. Quantitative DNase I footprinting was carried out for polyamides **2** and **4** on PCR products of pAU8, pAU15, and pAU12 (Figure 14). The variable base pair position was chosen opposite the amino acid residue in question. Specificity of Im: polyamide **2** binds the four DNA sites 5'-AAAGAXAAGAG-3' (**X** =A, T, G, C), with similar high affinities, $K_a = 2.6 - 1.1 \times 10^{10} \text{ M}^{-1}$, revealing that Im tolerates all four base pairs.

Figure 14 Quantitative DNase I footprint titration experiments for polyamide **2** on the 298 bp, 5'-endlabelled PCR product of plasmids pAU8 (A), pAU15 (B), and pAU12 (C), as well as polyamide **4** on the PCR product of pAU12 (D): (A) and (C), lane 1, intact DNA; lane 2, G reaction; lane 3, A reaction; lane 4, DNase I standard; lanes 5 - 14, 300 fM, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM **2**, respectively; (B) lane 1, intact DNA; lane 2, G reaction; lane 3, A reaction; lane 4, DNase I standard; lanes 5 - 15, 100 fM, 300 fM, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM 1, respectively; (D) lane 1, intact DNA; lane 2, G reaction; lane 3, A reaction; lane 4, DNase I standard; lanes 5 - 14, 10 pM, 300 fM, 1 pM, 3 pM, 10 pM, 30 pM, 100 nM, 300 nM, 100 nM, 300 nM **4**, respectively. Each footprinting gel is accompanied by the following: (above) binding schematic with the mutated position boxed; (left, top) chemical structure of the monomer of interest; and (left, bottom) Langmuir binding isotherms for the four designed sites. θ_{norm} values were obtained using a nonlinear, leastsquares algorithm.



27

Residue	A•T	T∙A	G•C	C•G
Im (2) ^c	$2.5(\pm 0.2) \ge 10^{10}$	$1.1 (\pm 0.1) \ge 10^{10}$	$2.6(\pm 0.4) \ge 10^{10}$	$1.3 (\pm 0.3) \ge 10^{10}$
β (2)	$2.4(\pm 0.1) \ge 10^{10}$	$1.3(\pm 0.1) \ge 10^{10}$	$4.3 (\pm 1.4) \ge 10^8$	$7.8(\pm 1.9) \ge 10^8$
Py (2)	$3.4(\pm 0.3) \ge 10^{10}$	$1.5(\pm 0.2) \ge 10^{10}$	$1.8(\pm 0.2) \ge 10^9$	$8.6(\pm 1.5) \ge 10^8$
Hp (4)	$1.6(\pm 0.2) \ge 10^9$	$1.3 (\pm 0.2) \ge 10^8$	$4.9(\pm 0.8) \ge 10^7$	$1.0(\pm 0.3) \ge 10^7$

Table 2 Equilibrium Association Constants, K_a (M⁻¹)^{a,b}

^{**a**} Values reported are the mean values from at least three DNase I footprint titration experiments, with the standard deviation given in parentheses. ^{**b**} Assays were performed at 22 $^{\circ}$ C in a buffer of 10 mM Tris•HCl, 10 mM KCl, 10 mM, MgCl₂, and 5 mM CaCl₂ at pH 7.0. ^{**c**} The number in parentheses indicates the compound containing the unique residue.

Specificity of Py and β: polyamide **2** binds the target sites, 5'-AAAGAGA**X**GAG-3' and 5'-AAAGXGAAGAG-3', respectively (X = A, T, G, C), with high affinity and in both cases displays a preference for $A \bullet T$ and $T \bullet A > G \bullet C$ and $C \bullet G$ by at least a factor of 10. Remarkably, substituting one Py residue with Hp afforded the most specific polyamide (4), which binds the sequences 5'-AAAGAGAXGAG-3' with a modest loss in affinity, characteristic of the Hp residue (White et al., 1998; Kielkopf et al., 1998b), and with a tenfold single site preference for X = A > T > TG/C (Table 2). Langmuir binding isotherms for each complex fit well to an n = 1Hill equation (see Experimental for equation), which is consistent with a 1:1 ligand:DNA stoichiometry (Figure 14). To establish that the specificity of Im and β is not position dependent, controls were performed on polyamide 2 at different Im (pAU16, 5'-AAAGAGAAXAG-3') and β positions (pAU13, 5'-AAAGAGXAGAG-3'). The observed complex affinities and sequence specificities were similar to that described above for pAU8 and pAU15, respectively. The more significant effect of position on the sequence specificity of Hp is presented in the next section.

Discussion. This quantitative study helps to elucidate the current state-ofthe-art for 1:1 polyamide:DNA complexes and creates a baseline for future specificity studies. An Im residue binds each of the four Watson•Crick base pairs with high affinity, whereas a β or Py residue prefers A,T over G,C base pairs but does not discriminate A from T. Steric inhibition between the exocyclic amino group of guanine and the Py and β residues may explain their A,T preference, as previously suggested by Dickerson from x-ray structural analysis of 1:1 complexes. Based on the study of netropsin bound in a 1:1 complex with DNA, the promiscuous nature of the Im residue accepting G,C as well as A,T was anticipated (Kopka et al., 1985). The unanticipated result of this study is the observation that an Hp residue in this polyamide sequence context can distinguish one of the four Watson•Crick base pairs. Whether the hydroxyl moiety lies asymmetrically in the cleft between A and T and makes a specific hydrogen bond to the O2 of T, as observed for Hp/Py recognition of T•A (Kielkopf et al., 1998b), is unclear. However, the fact that a single aromatic carboxamide residue can select one of the four Watson•Crick base pairs within the 1:1 motif is an encouraging step toward a set of rules for DNA recognition similar to the 2:1 motif. Whether new aromatic residues can be invented to complete a 1:1 recognition code is addressed in the next section.

Specificity of Novel Heterocyclic Amino Acids³

Approach. In an effort to improve upon the 1:1 recognition code established in the previous section, we address the issues of whether new heterocycles can be developed to discriminate between the four Watson-Crick base pairs in the 1:1 motif, and whether we can understand the relationship between overall heterocycle structure and DNA sequence specificity. This specificity has been attributed largely to the unique functionality presented by each heterocycle to the floor of the minor groove. However, little has been done to assess the ramifications of functional groups pointing away from the DNA. Figure 15 shows a family of five-membered aromatic, heterocyclic residues grouped into columns by the type of functionality directed toward the DNA minor groove. Py and 1H-pyrrole (Nh) project a hydrogen with positive



Figure 15 Family of five-membered aromatic heterocycles.

³ The text of this section is taken from Marques et al., in preparation.

potential toward the DNA; Im, 5-methylthiazole (Nt), and furan (Fr) project an sp² lone pair from nitrogen or oxygen; Hp and 3-hydroxythiophene (Ht) project a hydroxyl group; and 4-methylthiazole (Th) and 4-methylthiophene (Tn) project an sp² lone pair from sulfur. Comparative analysis of new residues within this five-membered heterocyclic framework should enable us to retain overall ligand morphology and to observe the effects of small structural changes, such as single atom substitution, on DNA base pair specificity. Specificity at the unique carboxamide position was determined by varying a single base pair (**X**) within the sequence context 5'-AAAGAXAAGAG-3' to all four Watson•Crick base pairs and comparing the relative affinities for the four possible complexes. *Ab initio* computational modeling of the heterocyclic amino acids was implemented to derive their inherent geometric and electronic parameters. The combination of these techniques has provided an interesting perspective on the origin of DNA sequence discrimination by polyamides.

Synthesis. Synthesis of Boc-protected Nh, Fu, Ht, Nt, and Tn amino acids and the corresponding polyamides Im-β-ImPy-β-X-β-ImPy-β-Dp (**X** = unique heterocycle) required new solution and solid-phase synthetic methodologies, which were developed by Michael Marques and Raymond Doss and will be reported elsewhere (Marques et al., in preparation). The structures of polyamides Im-β-Im-Py-β-**Py**-β-Im-Py-β-Dp (**5**), Im-β-Im-Py-β-**Hp**-β-Im-Py-β-Dp (**6**), Im-β-Im-Py-β-**Nh**-β-Im-Py-β-Dp (**7**), Im-β-Im-Py-β-**Ht**-β-Im-Py-β-Dp (**8**), Im-β-Im-Py-β-**Fr**-β-Im-Py-β-Dp (**9**), Im-β-Im-Py-β-**Nt**-β-Im-Py-β-Dp (**10**), Im-β-Im-Py-β-**Tn**-β-Im-Py-β-Dp (**11**), and Im-β-Im-Py-β-**Th**-β-Im-Py-β-Dp (**12**), with the parent Im polyamide (**2**) are shown in Figure 16.



Figure 16 Chemical structures for 1:1 polyamides containing novel heterocyclic residues, with the variable positions in the central ring indicated by the letters X, Y, and Z. A binding model is shown at top with the variable polyamide position indicated by a circle containing the letter X proximal to the variable base pair, $X \cdot Y$.

DNA Binding Affinity and Sequence Specificity. Quantitative DNase I footprinting was carried out for polyamides **5** – **12** on the 298 bp PCR product of pAU8 (Figures 17 and 18). The variable base pair position was installed opposite the amino acid residue in question. Equilibrium association constants (K_a) for 1:1 polyamides containing Im, Py, and Hp residues tested against the four Watson-Crick base pairs have been discussed. However, in that study only the Im specificity experiment was performed at the more flexible central residue, as with the new polyamides reported here. Therefore, new polyamides containing Py and Hp residues at the central position have been included in this study for a

Figure 17 (A-D) Quantitative DNase I footprint titration experiments for polyamides **5-8**, respectively, on the 298 bp, 5'-end-labelled PCR product of plasmid pAU8: (A and B) lane 1, intact DNA; lane 2, G reaction; lane 3, A reaction; lane 4, DNase I standard; lanes 5-15, 100 fM, 300 fM, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM polyamide, respectively. (C) lane 1, intact DNA; lane 2, G reaction; lane 3 A reaction; lane 4, DNase I standard; lanes 5-15, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM polyamide, respectively. (D) lane 1, intact DNA; lane 2, G reaction; lane 3 A reaction; lane 4, DNase I standard; lanes 5-15, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, 100 nM polyamide, respectively. (D) lane 1, intact DNA; lane 2, G reaction; lane 3 A reaction; lane 4, DNase I standard; lanes 5-15, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, 100 nM polyamide respectively. Each footprinting gel is accompanied by the following: (left, top) chemical structure of the residue of interest; and (left bottom) Langmuir binding isotherm for the four designed sites. θ_{norm} values were obtained using a nonlinear least-squares fit.



Figure 18 (A-D) Quantitative DNase I footprinting experiments for polyamides **9-12**, respectively, on the 298 bp, 5'-end-labelled PCR product of plasmid pAU8: lane 1, intact DNA; lane 2, G reaction; lane 3, A reaction; lane 4, DNase I standard; lanes 5-15, 100 fM, 300 fM, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, respectively. Each footprinting gel is accompanied by the following: (left, top) Chemical structure of the residue of interest; and (left bottom) Langmuir binding isotherm for the four designed sites. θ_{norm} values obtained using a nonlinear least-squares fit. Isotherms for C and D were generated from gels run out to a final concentration of 1 uM (not shown).



more controlled comparison. Polyamide **5** (Py) binds with very high affinity (K_a ~ 6 x 10¹⁰ M⁻¹) at the **X** = A, T sites (5'-AAAGAXAAGAG-3') with a 5- to 10-fold preference over **X** = G, C (Table 3). Polyamide **6** (Hp) binds with lower affinity (K_a ~ 3 x 10⁹ M⁻¹) but with similar specificity to **5**, preferring **X** = A, T > G, C by 5- to 10-fold. The Nh-containing polyamide (7) bound with very high affinity to the **X** = A, T sites (K_a = 7.5 x 10¹⁰ M⁻¹) but with a mere 3- to 5-fold selectivity over the high-affinity **X** = G, C sites. Compound **8** (Ht) bound with subnanomolar affinities to the **X** = A, T sites, similar to **6** but with \geq 40-fold specificity for **X** = A, T > G, C. Polyamide **9** (Fr) showed high affinity for the **X** = A, T sites (K_a ~ 10¹⁰ M⁻¹) with a small 2- to 4-fold preference over **X** = **G**, **C**. The 5-methylthiazole-containing polyamide (**10**, Nt), which places the thiazole ring *nitrogen* into the floor of the minor groove, bound all four sites with similar high affinities (K_a ~ 5 x 10⁹ M⁻¹). Thiophene-containing polyamide (**11**, Tn) showed modest single-site specificity, binding the **X** = **A** site at K_a = 3.0 x 10¹⁰ M⁻¹ with 5-fold preference over

				-
Ring	A•T	T∙A	G•C	C•G
Im (2) ^c	$2.5(\pm 0.2) \ge 10^{10}$	$1.1 (\pm 0.1) \ge 10^{10}$	$2.6(\pm 0.4) \ge 10^{10}$	$1.3 (\pm 0.3) \ge 10^{10}$
Py (5)	$7.2(\pm 0.3) \ge 10^{10}$	$5.3(\pm 0.1) \ge 10^{10}$	$3.2(\pm 0.4) \ge 10^9$	9.4 (±0.2) x 10^9
Нр (б)	$3.9(\pm 0.1) \ge 10^9$	$2.5(\pm 0.3) \ge 10^9$	$5.3 (\pm 0.5) \ge 10^8$	$1.9(\pm 0.5) \ge 10^8$
Nh (7)	7.5 (±0.2) x 10^{10}	$7.4(\pm 0.1) \ge 10^{10}$	$1.6 (\pm 0.2) \ge 10^{10}$	$2.3(\pm 0.1) \ge 10^{10}$
Ht (8)	$2.8(\pm 0.5) \ge 10^9$	$1.6(\pm 0.6) \ge 10^9$	$3.8(\pm 1.3) \ge 10^7$	$3.7 (\pm 0.7) \ge 10^7$
Fr (9)	$2.2(\pm 0.5) \ge 10^{10}$	$1.0(\pm 1.3) \ge 10^{10}$	$4.4(\pm 0.5) \ge 10^9$	$5.0(\pm 0.5) \ge 10^9$
Nt (10)	$5.4(\pm 0.9) \ge 10^9$	$2.9(\pm 0.6) \ge 10^9$	$8.0(\pm 1.3) \ge 10^9$	$4.2(\pm 0.6) \ge 10^9$
Tn (11)	$3.0(\pm 0.2) \ge 10^{10}$	$5.7 (\pm 0.4) \ge 10^9$	8.1 (±0.4) x 10^7	$8.3 (\pm 0.2) \times 10^7$
Th (12)	$1.5(\pm 0.2) \ge 10^{10}$	$3.0(\pm 0.7) \ge 10^9$	$6.5 (\pm 0.5) \ge 10^6$	$7.4(\pm 0.5) \ge 10^6$

Table 3 Equilibrium Association Constants, K_a (M⁻¹)^{a,b} for Polyamides Containing Novel Heterocycles (X) within Im-β-ImPy-β-X-β-ImPy-β-Dp

^a Values reported are the mean values from at least three DNase I footprint titration experiments, with the standard deviation given in parentheses. ^b Assays were performed at 22 $^{\circ}$ C in a buffer of 10 mM Tris•HCl, 10 mM KCl, 10 mM, MgCl₂, and 5 mM CaCl₂ at pH 7.0. ^c The number in parentheses indicates the compound containing the unique residue.

X = **T** and ~ 70-fold preference over **X** = **G**, **C**. 3-methylthiophene-containing polyamide (12, Th), which places the thiazole ring *sulfur* into the floor of the minor groove, bound with similar **X** = A, T affinity as **11** (Tn) but with > 400-fold preference over **X** = G, C. In all cases, binding isotherms fit well to an n = 1 Hill equation, which is consistent with a 1:1 polyamide:DNA stoichiometry (Figures 17 and 18).

Calculations. Molecular modeling calculations were preformed by Michael Marques using *Spartan Essential* software package (Wavefunction Inc.). Each ring was first minimized using an AM1 model, followed by *Ab initio* calculations using the Hartree-Fock model and a 6-31G* polarization basis set. Each heterocycle exhibited a unique geometric and electronic profile (Figure 19). Bonding geometry for imidazole, pyrrole, and 3-hydroxypyrrole were in excellent agreement with coordinates derived from x-ray structures of polyamides containing these heterocycles (Kielkopf et al., 1998a; 1998b). The overall curvature of each monomer was calculated to be the sweep angle (θ)



Ring	Х	Y	Z	θ (degrees)	Charge on X (e)
Fr	0	C-H	C-H	126	-0.31
Nt	Ν	C-Me	S	127	-0.60
Ht	O-H	S	C-H	133	+0.40
Nh	N-H	C-H	C-H	136	+0.34
Im	Ν	N-Me	C-H	137	-0.71
Py	C-H	N-Me	C-H	146	+0.21
Нр	O-H	N-Me	C-H	148	+0.50
Tn	S	C-Me	C-H	149	-0.21
Th	S	C-Me	Ν	153	-0.25

Figure 19 Geometric and electrostatic profiles for nine heterocyclic amino acids, derived from *ab initio* molecular modeling calculations using *Spartan Essential* software (Wavefunction, Inc.). (Top) Schematic illustrating the amide-ring-amide angle of curvature, θ . X, Y, and Z denote variable functionality at the different ring positions for each heterocycle. (Bottom) Table listing the functional groups at X, Y, and Z, along with the angle θ , and the electrostatic partial charge on X. For Ht, Nh, Py, and Hp, the positive charge on X is listed for the H atom.

created by the theoretical intersection of the two ring-to-amide bonds in each ring. The structures were ranked by increasing θ as follows: Fr > Nt > Ht > Nh > Im > Py > Hp > Tn > Th. The ring atom in closest proximity to the floor of the DNA minor groove was examined for partial charge. The structures were ranked by decreasing partial charge on this atom as follows: Hp > Ht > Nh > Py > Tn > Th > Fr > Im.

Discussion. The single-subunit•DNA complexes of the 1:1 motif provide a relatively flexible system for the exploration of novel recognition elements. Due to the conformational freedom imparted by the β residues, changes in heterocycle geometry do not have as much of an impact on DNA sequence recognition as in the hairpin motif (Marques et al., in preparation). In fact, all 1:1 compounds in this study bind with high affinity to the **X** = A, T sites but with varying degrees of **X** = A, T > G, C specificity. The high-resolution solution structure presented later in this Thesis reveals an important register of amide NH groups with the purine N3 and pyrimidine O2 groups on the floor of the DNA minor groove (Urbach et al., 2002). Given this alignment as a driving force for DNA recognition in the 1:1 motif, one may view the subtle differences in heterocycle curvature as merely placing the central ring atom (**X** in Figure 19) closer to or farther from the DNA. In this view, increasing the ring curvature decreases the polyamide-DNA intimacy, thereby diminishing DNA specificity. The results presented here fit well within this ideology.

Polyamides **5** (Py) and **7** (Nh) present a hydrogen with a positive potential to the minor groove floor. Both compounds exhibit a modest 3- to 5-fold selectivity for $\mathbf{X} = \mathbf{A}$, T > G, C, but **7** binds with higher affinity to all sites. The selectivity is probably due to the negative steric X-H to G-NH₂ interaction (X =

C3 for Py and N1 for Nh), which was predicted by Dickerson and coworkers for netropsin and supported by NMR studies discussed later in this Thesis (Kopka et al., 1985; Urbach et al., 2002). The higher affinity for 7 may be attributed to a combination of greater positive charge on N1-H and higher ring curvature, both of which should reduce specificity.

Polyamides 2 (Im), 9 (Fr), and 10 (Nt) present a small atom with an sp² lone pair directed toward the minor groove floor. Polyamide 2 was discussed in the previous section, binding all sites with high affinity and displaying virtually no discrimination between sites. 9 and 10 behave quite similarly. It is likely that the small atom (N for Im and Nt or O for Fr) presented to the DNA provides no steric clash with G-NH₂, and therefore all sites are bound with similarly high affinities.

Polyamides **6** (Hp) and **8** (Ht) present a hydroxyl group to the DNA minor groove. Previously in both hairpin and 1:1 systems, hydroxypyrrole successfully discriminated between A•T and T•A base pairs (White et al., 1998; Urbach and Dervan, 2001). Yet when flanked on both sides by β -alanine residues, as with the Hp compound presented here, single base-pair specificity is lost. This loss may be attributed to a larger degree of conformational freedom afforded to the Hp ring by the *two* flanking aliphatic linkers, (Urbach et al., 2002). Nonetheless, both **6** and **8** exhibit significant **X** = A, T > G, C specificity, as expected from a negative 3C-OH to G-NH₂ steric clash.

Polyamides **11** (Tn) and **12** (Th) present a sulfur atom with an sp² lone pair to the DNA minor groove. These compounds exhibit substantial $\mathbf{X} = \mathbf{A}$, T > G, C specificity ranging from \geq 70 to \geq 2300-fold. This remarkable selectivity may be attributed to the decreased curvature of thiazole and thiophene rings, which forces a more intimate interaction of the large sulfur atom and the minor groove floor. In the case of X = G and C, this interaction is very negative, resulting in complete loss in measurable binding affinity.

Sequence Dependence of Polyamide Orientation

Approach. Because Im binds all four base pairs at a single position, it would be interesting to ask what happens if one varies simultaneously the base pairs proximal to all four Im residues. To meet this end, the plasmid pAU18 was prepared, which contains the four binding sites 5'-AAAXAXAAXAAAA-3' (X = G, C, A, and T). Equilibrium association constants were determined for the complexes, and the orientation at each site was determined by affinity cleaving experiments.

DNA Binding Affinity and Sequence Specificity. Quantitative DNase I footprint titrations were performed to determine the equilibrium association constants for polyamide **2** at the designed sites on pAU18 (Figure 20). A slight preference for the **X** = G site is revealed, although all measured sites are bound with high affinity. Affinity cleavage analysis with polyamide **2** confirms a single orientation when **X** = G (Figure 20), consistent with 1:1 binding, which reverses when **X** = C, similar to observations made by Laemmli and coworkers (Janssen et al., 2000a). Although the **X** = A site lacks a DNase I footprint because of a characteristic lack of cleavage and is oriented to the 5' side of this binding site, similar to **X** = G. The cleavage pattern is broader than that observed at the other sites, likely because of an ensemble of slipped complexes. The **X** = T site reveals cleavage, however, on both sides of the binding site and to different extents.



Figure 20 (Left) Quantitative DNase I footprint titration experiment for polyamide **2** on the 298 bp, 5'-end-labelled PCR product from plasmid pAU18: lane 1, intact DNA; lane 2, G reaction; lane 3, A reaction; lane 4, DNase I standard; lanes 5 - 14, 300 fM, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM **2**, respectively. (Middle) Affinity cleavage experiment with polyamide **2E** on the same PCR product of plasmid pAU18: lanes 1-3, 1 nM, 3 nM, and 10 nM **2E**, respectively; lane 4, A reaction; lane 5, G reaction; lane 6, intact DNA. (Right) Schematic illustrating the observed affinity cleavage patterns with arrows representing relative cleavage intensities. Polyamides are drawn as oriented, 1:1 complexes, as observed by affinity cleavage. Equilibrium association constants, K_a , are listed below each binding site.

Because of the similarity in DNA sequence at this site when read $5' \rightarrow 3'$ or $3' \rightarrow 5'$, it is possible that polyamide **2** is binding 1:1 at this site but with a slight preference for one orientation. However, due to the plasticity of 5'-TA-3' steps (Dickerson, 2001), this site likely has the capacity to accommodate two ligands, and therefore the binding mode is unclear.

Discussion. It is interesting that although a single Im residue displays no significant base preference, the multiple G to C base mutation experiment described here demonstrates that the four Im residues taken together profoundly influence the orientation preference of the polyamide with respect to G•C vs. C•G base pairs. The similarity of orientation for X = G and A suggests that the purine rich strand dominates the orientation preference, i.e., polyamide N – C aligns in the minor groove 3' – 5' with the purine-rich strand. This phenomenon is addressed more extensively in the discussion of the high-resolution NMR structure. A hypothesis is offered there to explain the G/C-dependence of polyamide orientation based on the inherent bonding geometry of an Im residue combined with the large negative propeller twisting of base pairs in the 1:1 complex.

Ligand Size Limitations in the 1:1 Motif

Approach. Determination of macromolecular structure using NMR methods requires extensive deconvolution of two-dimensional spectra. To simplify this process, the size of the molecule or complex can be reduced. However, spectral quality, and therefore the quality of the final structure, depends on a high-affinity complex. For this reason, we were interested in elucidating the effects of ligand size on binding affinity in the 1:1 motif. Using the high-affinity, eleven-residue polyamide Im- β -ImPy- β -Im- β -ImPy- β -Dp (**2**) as the parent for this study, a series of truncated polyamides, designed as analogues of **2** containing 9, 8, and 6 residues and retaining the overall Im- β -ImPy- β -type sequence (Figure 21), were prepared and examined for binding affinity in complex with the parent sequence 5'-AAAGAGAAGAG-3'.



Figure 21 The effect of ligand size on binding affinity in the 1:1 motif. The 11 residue parent polyamide (2) is shown at the top. At left, 2 is shown with gray boxes drawn to indicate the residues that were removed in the design of polyamides 13 - 16. The number of residues and equilibrium binding constants are given at right.

DNA Binding Affinity and Sequence Specificity. Polyamides ImPy-β-Im-β-ImPy-β-Dp (13), Im-β-ImPy-β-Dp (14), ImPy-β-Im-β-Dp (15), and Imβ-ImPy-β-Dp (16) were prepared by standard solid phase protocols (Baird and Dervan, 1996). The equilibrium association constants for polyamides 13 - 16against the sequence 5'-AAAGAGAAGAG-3' were determined by quantitative DNase I footprinting on the 298 base pair PCR product of pAU18 (Figure 21). It was found that the 9-residue polyamide (13) binds with nanomolar affinity (K_a = 10^9 M⁻¹). However, the 8-residue polyamide (14) binds with 100-fold reduced affinity (K_a = 10^7 M⁻¹). Binding affinities for the 6-residue polyamides (15 and 16) were too low to be determined by this method. **Discussion.** These results clearly demonstrate the precipitous decrease in binding affinity as a function of ligand size. The 100-fold decrease in binding affinity from 9 to 8 residues demonstrates an important limitation of the 1:1 motif. However, since longer binding sites are required for greater specificity in a genomic context, this limitation is not so unfavorable. For the purposes of determining a high-resolution structure by NMR, the higher affinity polyamides 2 and 13 are desirable because their conformational mobility on the NMR time scale is likely to be considerably reduced as a function of their affinity. Because 13 is two residues smaller than 2, while still retaining high affinity, 13 is the preferred target for NMR studies. The question of how far the DNA binding site can be reduced while retaining a high-affinity complex will be addressed in the next section.

NMR Structure of a 1:1 Polyamide-DNA Complex⁴

Purpose. High-resolution structural studies of small molecule-DNA interactions improve our understanding of molecular recognition processes and aid in the design of sequence-specific DNA-binding ligands. In the course of these thesis studies, several observations were made that would require structural data for explanation. In particular, we were interested in elucidating the role of β -alanine, as well as the sequence-dependent orientation of the ligand (Janssen et al., 2000a; Dervan and Urbach, 2001; Urbach and Dervan, 2001).

Approach. In order to understand the unique mode of 1:1 recognition, we set out to solve the high-resolution solution structure of an Im- and β-rich polyamide bound to its purine-rich cognate match site in a 1:1 complex. The solution structure of the polyamide ImPy-β-Im-β-ImPy-β-Dp bound to a full turn of the 13-mer DNA duplex 5'-CC**AAAGAGAAG**CG-3'•5'-CG**CTTCTCTTT**GG-3' (match site bolded) is presented here (Figure 22). Footprinting and affinity



Figure 22 Nomenclature for the 1:1 complex studied by NMR. The chemical structure of polyamide **13** is shown in black with relevant protons labeled in red. Residue names are shown below the structure, linked by line segments. (inset) Polyamide binding model: DNA residue numbers are indicated in red, along with a dot model of the bound ligand.

⁴ The text of this section is taken from Urbach et al., 2002.

cleavage experiments were employed to determine the minimal complex size that retains the established properties of 1:1 polyamide-DNA recognition (Janssen et al., 2000a; Urbach and Dervan, 2001). The DNA and polyamide were synthesized and purified in large quantity, and a sample of the 1:1 complex was prepared at 3.67 mM concentration (10 mM sodium phosphate, pH 7.0). Twodimensional TOCSY, NOESY, and DQF-COSY experiments were performed, and the spectra were fully assigned. Distance constraints were derived from the NOESY spectra using hybrid matrix relaxation calculations in the MARDIGRAS program. The structural ensemble was generated and refined using AMBER 6.0 software.

DNA Binding Affinity and Ligand Orientation. To simplify the NMR spectral assignments, several truncated polyamides were screened in order to identify the minimal high-affinity ligand size. The ligand size study detailed in the previous section was the first round of this iterative process of design, synthesis, and characterization, wherein it was found that the polyamide ImPy-β-Im-β-ImPy-β-Dp (13) is the smallest ligand of its class that retains high affinity. In the second iteration, we were interested in determining the minimal DNA binding site for polyamide **13** that retains high affinity. To this end, three DNA binding sites of sequence type 5'-ccAAAGAGAAGA_ncg-3' (flanking G•C clamps in lower case) were examined as candidates for NMR studies, such that the only variable among the sites was the number of A•T base pairs (**A**_n, **n** = 0, 1, and 2) beyond the N-terminus of the polyamide binding site. To determine binding affinity, site size, and orientation, the DNA sequences were cloned into plasmid pAU20 and characterized in complex with polyamide **13** by DNase I and methidiumpropyl-EDTA (MPE) footprinting. Affinity cleavage analysis was



Figure 23 Biophysical characterization. (lanes 1-15) Quantitative DNase I footprint titration experiment for polyamide 13 on the 289-bp 5'-end-labeled PCR product from pAU20: lane 1, intact DNA; lane 2, G reaction; lane 3, A reaction; lane 4, DNase I standard; lanes 5-15, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, 100 nM 13, respectively. (lanes 16-25) MPE and affinity cleavage experiments with 13 and 13E, respectively, on the 289-bp 5'-end-labelled PCR product from pAU20: lane 16, intact DNA; lane 17, G reaction; lane 18, A reaction; lane 19, MPE standard; lanes 20-22, 10 nM, 30 nM, 100 nM 13, respectively; lanes 23-25, 10 nM, 30 nM, 100 nM 13E, respectively. (lanes 26-35) MPE and affinity cleavage experiments with 13 and 13E, respectively, on the 289-bp 3'-end-labelled restriction fragment of pAU20: lane 26, intact DNA; lane 27, G reaction, lane 28, A reaction, lane 29, MPE standard; lanes 30-32, 10 nM, 30 nM, 100 nM 13, respectively; lanes 33-35, 10 nM, 30 nM, 100 nM **13E.** respectively. (right) Analysis of the polyamide•d(CCAAAGAGAAGCG)•d(CGCTTCTCTTTGG) binding site at the bottom of the gels: (right top) Langmuir binding isotherm for the DNAse titration with the binding constant (K_a) , as determined from a nonlinear least squares fit, shown in bold type; (right bottom) schematics illustrating observed protection (middle) and cleavage (bottom) patterns derived from the MPE and affinity cleavage experiments, respectively. The relative heights of the bars and arrows indicate relative intensities of protection and cleavage, respectively. Polyamides are drawn as oriented, 1:1 complexes, as observed in these experiments.

performed on the C-terminal EDTA conjugate (**13E**) (Figure 23). Quantitative DNase I footprinting of compound **13** on pAU20 revealed essentially equivalent binding affinities for the three sites at $K_a \sim 2 \times 10^9$ M⁻¹, indicating that the additional A•T base pairs flanking the binding site (**A**_n) were not necessary to stabilize the complex. MPE and affinity cleavage experiments of compounds **13**

and **13E**, respectively, show the polyamide bound to its match sequence, 5'-AAAGAGAAG-3', oriented N-C with the 3'-5' direction of the purine-rich strand. Both experiments show a 3'-shift between upper and lower strands, which is characteristic of minor groove binding (Taylor et al., 1984). The smallest sequence, 5'-ccAAAGAGAAGcg-3', was chosen for NMR studies because it retains 1:1 polyamide-DNA binding properties and high affinity.

Titration to 1:1 Polyamide:DNA Stoichiometry. The titration of polyamide **13** to the NMR sample of d(CCAAAGAAGAAGCG) • d(CGCTTCTCTTTGG) is shown in Figure 24. Chemical shift perturbation in the uncrowded DNA imino region upon polyamide addition was used to monitor the degree of complex formation. The number of imino peaks doubles upon addition of sub-stoichiometric amounts of polyamide. A single set of imino peaks is restored at a 1:1 polyamide-DNA stoichiometry. Only one set of



Figure 24 Far downfield region (9.0 - 14.5 ppm) of the ¹H NMR (at 600 MHz, 25 °C) spectra in 9:1 D₂O:H₂O of d(C C A A A G A G A G A G C G) • d(C G C T T C T C T T T G G) with (A) no ligand added; (B) 1:4 ligand:DNA; (C) 1:2 ligand:DNA; (D) 3:4 ligand:DNA; and (E) 1:1 ligand:DNA stoichiometry. Molar ratios are indicated at right. Ligand aromatic amide NH protons are indicated by asterisks. Peaks at left are from thymine and guanine imino protons.

polyamide signals is observed, as indicated with asterisks in Figure 24, which further confirms a 1:1 ligand:DNA stoichiometry. Terminal base pair imino resonances were not observed.

Spectral Assignments. Methods for assigning protons in polyamide•DNA complexes have been well established (Pelton and Wemmer, 1988; Pelton and Wemmer, 1989; Mrksich et al., 1992; Dwyer et al., 1992; Dwyer et al., 1993; Geierstanger et al., 1996). DNA chemical shifts for the 1:1 complex are listed in Table 4. The intramolecular DNA crosspeaks were assigned as previously described (Wütrich, 1986). 16 pairs of H5'/H5" protons were assigned nonstereospecifically. All other nonexchangeable protons were

	u6/u8	CU5 / λ 2U	י 1 נו	<u>ц</u> 2 і	U 2"	יכם	ц / і	CU1 / TU3	CH2	CH/1	CU/2
Strand 1	10/10	CHJ/AZH	пт	ΠZ	112	пэ	114	GH1/1H5	GIIZ	CH41	CII4Z
C1	7 72	5 95	5 98	2 03	2 / 8	1 66	1 09			6 93	7 01
C2	7 47	5 74	5 28	1 83	2 13	4.00	4 01			6 81	8 73
A3	8 25	7 41	5 56	2 76	2 76	5 01	4 32				
A4	8 19	7 31	5 48	2.76	2 71	5 05	4 35				
Δ5	8 08	7.66	5 9/	2.70	2 71	1 98	1 15				
G6	7 13	7.00	5 11	1 93	2.54	4.72	3 28	12 54	7 76		
A7	7 62	8 0/	5 56	2 06	2.54	1.71	3 04	12.54			
68	7 13		5 12	2.00	2.05	1 90	1 03	12 23	6 58		
20	7 93	7 61	5 40	2.10	2 70	1 91	3 82	12.25	0.50		
A10	7 01	7.01	6 04	2.15	2.70	5 02	1 27				
G11	7 20	7.94	5 30	2.00	2.75	1 63	3 52	12 57	8 12		
C12	6 96	/ 91	1 85	1 /1	1 83	4.62	3 28	12.57	0.12	6 11	8 16
C12	7 97	4.91	4.0J	2 36	2 64	4.02	1 20	not ob	not ob	0.11	0.10
GI2	/.0/		0.10	2.30	2.04	4.09	4.20	not ob	. HOL OD	•	
Strand 2											
C14	7.58	5.89	5.76	1.70	2.33	4.67	4.04			7.02	8.35
G15	7.92		5.96	2.68	2.79	4.99	4.38	13.28	7.23		
C16	7.44	5.39	6.04	2.06	2.61	4.67	4.35			6.42	7.96
T 17	7.26	1.52	5.85	1.78	2.50	4.65	3.45	14.53			
Т18	7.17	1.57	5.76	1.91	2.49	4.67	3.40	14.47			
C19	7.53	5.53	6.17	2.15	2.43	4.88	3.99			6.69	8.25
т20	7.13	1.52	5.44	1.91	2.37	4.70	3.41	14.41			
C21	7.52	5.44	6.10	2.15	2.45	4.71	4.06			6.56	7.86
т22	7.16	1.48	5.68	1.93	2.44	4.68	3.04	14.19			
т23	7.26	1.70	5.56	1.78	2.29	4.62	3.08	13.93			
т24	7.20	1.69	5.63	1.88	2.13	4.86	3.94	14.03			
G25	7.90		5.55	2.70	2.70	4.97	4.33	12.96	6.81		
G26	7 85		6 1 9	2 30	2 5/	4 65	1 20	not ob	not ob		

Table 4. DNA Chemical Shift Assignments for the 1:1 Polyamide-DNA Complex.*

* Nonlabile proton chemical shifts are from the NOESY spectrum in D20 (75 ms mixing). Labile proton chemical shifts are from the NOESY spectrum in H20 (100 ms mixing).

Residue:	Im-1	Py-2	Beta-3	Im-4	Beta-5	Im-6	Py-7	Beta-8		Dp-9
H4-1 H3 H5 NCH-3	7.21 7.33 4.12	6.77 7.77 4.03		 7.62 4.12		 7.73 4.14	6.46 7.55 3.94			
methylene protons		C12-H1 C12-H2 C13-H1 C13-H2	2.91 2.94 3.70 4.01	C20-H1 C20-H2 C21-H1 C21-H2	2.56 3.04 3.61 4.13		C34-H1 C34-H2 C35-H1 C35-H2	2.42 2.65 3.43 3.53	C37-H1 C37-H2 C38-H1 C38-H2 C39-H1 C39-H2	3.12 3.38 1.94 1.94 3.12 3.18
Amides	NH-1 9.65	NH-2 8.24	NH-3 9.87	NH-4 8.35	NH-5 11.14	NH-6 9.26	NH-7 8.08	NH-8 8.66	Dimethyl	2.90

Table 5. Polyamide Chemical Shift Assignments.*

* Nonlabile proton chemical shifts are from the NOESY spectrum in D20 (75 ms mixing). Labile proton chemical shifts are from the NOESY spectrum in H20 (100 ms mixing).

assigned, and all exchangeable protons were assigned except for those on the terminal C1•G26 and C13•G14 base pairs. Sequential assignments for guanine and thymine imino protons were determined from the NOESY spectrum in protiated solvent. These facilitated the assignment of A-H2 and C-amino protons. The C-amino protons correlate strongly to each other and to the vicinal C-H5 proton, allowing identification of C-H5 – C-H6 crosspeaks. It is noteworthy to point out that the process of assigning imino \rightarrow amino \rightarrow A-2H/CH5/CH6 was straightforward, and proved essential for the unambiguous assignment of the aromatic-H1' region. NOEs were observed between each A-H2 proton and the H1' protons of its 3'-neighbor and its base-paired thymidine's 3'-neighbor. This pattern is indicative of a narrow minor groove (Nadeau and Crothers, 1989).



Figure 25 Expansion of the aromatic and amide region of the NOESY spectrum of the 1:1 polyamide • $d(CCAAAGAGAAGCG) \cdot d(GCGTTCTCTTTGG)$ complex (9:1 H₂O:D₂O, 10 mM sodium phosphate, pH = 7.0; 25 °C; 75 ms mixing). Sequential aromatic to H1' connectivities for the purine-rich strand are shown as dotted lines; those for the pyrimidine-rich strand as solid lines. Crosspeaks are labeled according to their chemical shifts along ω_1 (vertical axis, label beside the peak) and along ω_2 (horizontal axis, label above or below the peak). Labeling conventions for the DNA are residue name, residue number, proton name (e.g. A7H2 = Adenine 7, H2 proton); ligand is named as proton name, residue number (e.g. NH-7 = amide NH of pyrrole 7). Red asterisks above or below a peak indicate a cross-strand A-2H to H1' NOE (type *a* distance, Nadeau and Crothers, 1989).

Nomenclature for polyamide protons is shown in Figure 22. All polyamide protons were assigned, and their chemical shifts are listed in Table 5. Amide NH, Py-H3, and Im-H4 protons were assigned based on intra- and intermolecular connectivities, as observed in the NOESY spectrum in H_2O . Methylene protons for each β -alanine residue were identified by a unique set of strong intermolecular contacts with each A-H2 proton. Geminal pairs were assigned nonstereospecifically based on the patterns observed in the DQF-COSY



Figure 26 Model of the 1:1 complex showing observed intermolecular NOEs from the polyamide to the floor and walls of the minor groove as solid and dashed gray lines, respectively. The DNA is shown as a ladder with residue numbers at each rung. Lines to a residue number show a contact to H1'; lines ending between residue numbers indicate H4' and/or H5'/H5" protons.

and NOESY experiments. Ring H5 protons were identified by contacts to H4' ribose protons as well as strong contacts to the respective proximal N-methyl. The **-C**₁₆ intermolecular connectivity pattern observed for these protons defines the orientation of each Im and Py residue such that the N-methyl group points out of the groove. Polyamide chemical shifts are similar to those previously reported by Wemmer and coworkers for 1:1 and 2:1 complexes (Mrksich et al., 1992; Geierstanger et al., 1996; de Clairac et al., 1999).

> The NOESY spectrum of the 1:1 complex in H_2O is shown in Figure 25. Sequential ligand assignments were based on a clear pattern of intermolecular NOEs from ligand protons to assigned DNA protons. Strong intermolecular NOEs between Py-H3 and A-H2 protons, as well as contacts between amide NH and H1' protons in the binding site, orient the polyamide N-C with respect to the 3' – 5'

direction of the purine-rich strand. Each polyamide residue contacts H1', H4', and/or base protons on *both strands* of DNA. This is further evidence for 1:1 complex formation in the minor groove with a single binding orientation. Selected intermolecular contacts are illustrated in Figure 26, with solid and dashed lines indicating contacts to the floor and walls of the minor groove, respectively. A homogeneous distribution of contacts is observed, which unambiguously defines the position of the ligand within the minor groove. The NOESY and spectra are provided in Appendix A, with each region expanded and annotated.

Distance Constraints. Polyamide binding induces a high dispersion of NOESY crosspeaks. This enabled the identification of a large number of discrete NOEs, which allowed the use of 508 distance constraints for structure calculations. The distribution of experimental constraints is relatively homogeneous throughout the binding region with fewer constraints for the terminal base pairs. The majority of constraints were derived from the NOESY spectrum in D_2O , with an additional 115 distance constraints from the NOESY spectrum in H_2O . Methods for converting NOE intensities to upper bound distance constraints are detailed in the Experimental section. Forty Watson-Crick hydrogen bond constraints were applied on the basis of observed cytosine amino, and guanosine and thymidine imino chemical shifts in the spectral region indicative of cross strand hydrogen bonding. The final list of 548 experimental constraints is provided in Appendix B and is available from the Brookhaven Data Bank under accession code 1LEJ.

Structure Calculations. Following the approach of Chazin and coworkers, a significant effort was made during the structure calculations to

sample and represent conformational space consistent with the input data (Eis et al., 1997; Schnell et al., 1999). The starting ensemble of 40 structures, which differ considerably in helical geometry (RMSD = 3.46 Å), were generated using the Nucleic Acids Builder program (NAB; Macke and Case, http://www.scripps.edu/case). These structures were positioned with the polyamide and docked using a restrained molecular dynamics (rMD) simulated annealing (SA) protocol in the AMBER 6.0 software package (Kollman et al., 2000). The docked structures converged to an RMSD of 1.37 Å, and subsequent rMD SA did not improve the total energy or RMSD of the ensemble. The 40 docked structures were sorted by increasing residual constraint violation energy,

Table 6. Statistics for the Final StructuralEnsemble of the Polyamide•DNA Complex					
E _{total}	-1407.4 (±17.0) ^a				
E _{Leonard Jones}	-583.7 (±14.7) ^a				
Eviolation	$0.9 \ (\pm 0.2)^{a}$				
NOE Violations ≥ 0.13 Å	0 ^b				
NOE Violations ≥ 0.10 Å	0.42 ^b				

Progression of the NMR Refinement (I	RMSD, Å) ^c
40 Starting DNA Structures	3.46
40 Docked Complexes	1.37
12 Lowest Constraint Energy Structures	1.12 (0.76)
Core Eight Base Pair Binding Site ^d	0.80 (0.54)

^a Average AMBER energies in kcal/mol. ^b Average number of violations per structure. ^c Mean pairwise rmsd between the structures. ^d The core binding site includes all heavy atoms in DNA residues A4 – G11, C16 – T23, and polyamide residues Im1 – Py7. The rmsd from the mean is given in parentheses.

and the 12 structures of lowest violation energy were chosen as the final ensemble (RMSD = 1.12 Å). A summary of relevant statistics for the structure calculations is given in Table 6. These statistics demonstrate excellent agreement with the input data (low violation energy) as well as nonbonding stabilization of the structure due to ligand binding (large negative Leonard Jones energy).

Discussion

Confirmation of Oriented 1:1 Binding. Several independent measurements confirm formation of the 1:1 ligand:DNA complex presented here, in which the polyamide is oriented N–C with respect to the 3'–5' direction of the purine-rich strand. The binding isotherm derived from quantitative DNase I footprinting (Figure 23) fits an n=1 Hill equation. The 1D NMR titration (Figure 24) shows the appearance of a single set of polyamide resonances with concomitant shifting of the DNA peaks at 1:1 stoichiometry. By contrast, polyamides that form 2:1 complexes with high cooperativity are known to do so even at low (0.25:1) ligand:DNA ratios, resulting in the existence of two sets of DNA resonances at 1:1 stoichiometry (Pelton and Wemmer, 1989; Mrksich et al., 1992; de Clairac et al., 1999). Additionally, the NOESY spectrum reveals a large number of intermolecular contacts from ligand protons to both strands of DNA with similar intensity. This is in sharp contrast to reported 2:1 complexes where ligands contact only the proximal strand (Pelton and Wemmer, 1989; Mrksich et al., 1992; de Clairac et al., 1999).

The affinity cleavage and MPE footprinting experiments (Figure 23) reveal a characteristic 3'-shift of cleavage and protection intensities, respectively, between the upper and lower DNA strands, which is indicative of minor groove binding (Taylor et al., 1984). The existence of cleavage on only one side of the binding site, even when the [polyamide] >> [DNA], is evidence for a single binding orientation. The pattern of intermolecular contacts observed in the NOESY spectra (Figures 25 and 26) confirms the DNA binding site and orientation, in addition to the ligand–DNA recognition code of Im to G•C, and Py and β to A/T base pairs.



Figure 27 Stereo view of the final ensemble of 12 structures superimposed over the binding site. DNA is shown in blue, ligand in gold. The binding site is shown as a lighter color in the center of the complex.

Characterization of the Complex. The final ensemble of 12 structures is shown in Figure 27 with the core binding site highlighted in the center. The binding site is well defined, and it consists of the five aromatic residues linked by two flexible β residues (ImPy- β -Im- β -ImPy). The RMSD for the ensemble is 1.12 Å, converging to 0.80 Å for all heavy atoms in the binding site (Table 6). The ends of the DNA as well as the C-terminal β -Dp tail of the polyamide sample a larger region of conformational space, and consequently are more poorly defined. The DNA has average B-form values for rise per residue (3.3 Å) and inter-base-pair twist (38°). Sugar pucker values span the range from C2' endo to C1' exo, as previously observed for A-tract structures (Celda et al., 1989), with an average phase angle value of 148°. Helical parameters are listed in Appendix C.

Minor Groove Width and Propeller Twist. Figure 28 shows plots of minor groove width and propeller twist for the complex, which were calculated using the CURVES program, as described by Lavery and coworkers (Lavery and Sklenar, 1988; Stofer and Lavery, 1994). Overall, the complex displays a narrow minor groove and a large negative propeller twist, which are features typically associated with A-tract structures (Crothers and Shakked, 2001). The narrow minor groove is confirmed by the observation of medium intensity cross-strand NOEs (type *a* distance, Nadeau and Crothers, 1989) between each adenine H2 and the H1' of its base-paired thymidine's 3'-neighbor (red stars in Figure 25). The minor groove in the structural ensemble is much narrower than observed for ligand-free duplex DNA containing similar sequences – 5'-AAAGAA-3' by NMR (MacDonald et al., 2001) as well as 5'-AAAGAAA-3' and short A-tracts by x-ray crystallography (Han et al., 1997; Shatzky-Schwartz et al., 1997). However, it has



Figure 28 Plots of average propeller twist (top) and minor groove width (bottom) at each DNA residue, determined using the CURVES program as described by Lavery and coworkers (Lavery and Sklenar, 1988; Stofer and Lavery, 1994). Average values are connected by a solid line, and the y-axis error bars indicate one standard deviation from the average over the ensemble of 12 structures. Schematics illustrating the helical parameters are given in the upper right of each plot.

been proposed that ligand binding in a 1:1 mode can induce the walls of the minor groove to close down in order to maximize van der Waals contacts (Dickerson, 2001; Bostock-Smith et al., 2001). Large negative propeller twist is commonly associated with a narrow minor groove. Lu and coworkers observe a significant decrease in propeller twist upon interruption of an A-tract with guanine, and they attribute this to a disruption in the spine of hydration by the G-NH₂ group (MacDonald et al., 2001). We observe a consistently large degree of propeller twist throughout the complex, even though there are multiple guanines interrupting the short A-tracts. Therefore, this effect is likely to be stabilized by the ligand, as discussed below. Helical parameters are provided in Appendix C.



Figure 29 Top view of the polyamide•DNA mean structure with (left) and without (right) the ligand. The DNA is shown as a stick model in black, and the ligand is rendered with a molecular surface in gold.

Ligand Structure. Figure 29 displays an axial view of the DNA helix with and without the polyamide shown. It is clear from this perspective that the DNA has a characteristic B-form geometry. The helical axis is relatively straight, and the polyamide wraps completely around the DNA, binding to a full turn of 9.5 base pairs. The ligand is bound in the minor groove to its cognate nine base pair match site, as determined by footprinting studies, with Im across from G•C, and Py and β across from A•T base pairs. The orientation of the polyamide is N – C with the 3' – 5' direction of the purine-rich strand.

The core binding site contains β residues in positions 3 and 5 (ImPy- β 3-ImPy – see Figure 22). The structural ensemble reveals a single orientation for β 5. However, equal populations of two binding modes are observed for β 3, which adopts a straight conformation and a bent conformation with virtually identical AMBER energies. We were chiefly interested in assessing the impact of

these two conformations on DNA recognition. Therefore, we averaged the coordinates for the 5 bent and 7 straight conformers within the final ensemble, minimized the mean structures, and then superimposed the flanking Py2 and Im4 rings of the two mean structures. The superposition reveals an interesting result: the two β -alanine conformations do not affect the positioning or relative orientation of the flanking aromatic rings. Specifically, the dihedral between the aromatic ring planes of Py2 and Im4 are virtually identical (30° and 33°) in spite of the two β -alanine conformations. This result could be attributed to a higher definition of position for β 5 from intermolecular NOEs, or it could be that the role of β -alanine is more to provide the flexibility needed for polyamide residues to properly align with DNA base pairs and less for specific base recognition.

Amide–DNA Interactions. In the absence of a ligand, minor groove hydration is thought to stabilize the propeller twisted base pairs via hydrogen bonding to the N3 and O2 groups on the proximal purine and pyrimidine residues, respectively (Dickerson and Drew, 1981). Netropsin (Kopka et al., 1985) and distamycin (Coll et al., 1987) have been observed to displace the spine of hydration by forming bifurcated bonds from each amide NH to the same N3 and O2 atoms in the 1:1 motif. Indeed, we observe this phenomenon. Figure 30 illustrates the interactions observed between ligand NH and purine-N3/pyrimidine-O2 groups, showing the core binding site of the NMR structure at left with the average (± standard deviation) distances given in the schematic at right. Amide NH groups to the C-terminal side of each Im residue have longer NH to N3 distances than would be ideal for hydrogen bonding. This small deviation could be due to competition from the Im-N3 to properly hydrogen



Figure 30 Schematic of polyamide NH to purine N3 and pyrimidine O2 contacts for DNA residues A4-G11•C16-T23 and polyamide residues Im1-Py7. The figure has three parts, which are all in vertical register. (left) Mean structure of the polyamide•DNA complex showing the base pairs as a stack down the center with purines in dark gray, pyrimidines in light gray, and ligand residues in yellow. Hydrogen bond donor and acceptor pairs are in blue and red, respectively. Amide NH to purine N3 and pyrimidine O2 interactions are indicated by dashed lines; Imidazole-N3 to guanine NH2 hydrogen bonds are shown as thick dotted lines. (center) Polyamide chemical structure. (right) Diagram of amide NH to purine N3 and pyrimidine O2 interactions. The DNA is shown as a ladder with each rung containing the residue number as well as purine N3 and pyrimidine O2 atoms indicated by bold "N" and "O," respectively. The ligand is illustrated as a long bar with imbedded circles containing the amide NH number, as defined in Figure 22. Dashed lines connect NH to N3 and O2 atoms. Average distances over the final 12 structures for these interactions are given in red next to the dashed lines, with standard deviations in parentheses.

bond to G-NH₂. The distance values are roughly equal from the left to right side of the minor groove, which shows the relatively central location of the ligand between the two DNA strands on the floor of the minor groove. The distances are also similar along the length of the DNA (top to bottom in Figure 30), attesting to an optimal ligand–DNA register along the length of the minor groove between polyamide and DNA residues for a complete turn of the helix. This is evidence for excellent shape complementarity between polyamide and DNA, and it suggests that part of the driving force for polyamide–DNA register is the proper alignment of amide NH groups with respect to the DNA base pairs.

The Lexitropsin Model. Dickerson and Lown proposed that substitution of one or more Py units in netropsin with Im should confer G/C specificity in a 1:1 complex (Kopka et al., 1985; Lown et al., 1986). Attempts have been made to verify this specific interaction in x-ray cocrystal structure analysis of imidazolecontaining netropsin analogs with DNA sequences containing G/C base pairs (Goodsell et al., 1995; Kopka et al., 1997). However, Im does not bind across from G in these structures, and therefore the key Im-N3 to G-NH₂ hydrogen bond in 1:1 binding was not verified by a high-resolution structure. Considerable x-ray and NMR structural data exists to support the formation of Im-N3 to G-NH₂ hydrogen bonds for the Im/Py pair in the 2:1 motif (Kielkopf et al., 1998a; Mrksich et al., 1992; Dwyer et al., 1992; Geierstanger et al., 1994). In the 1:1 complex presented here, we find a specific oriented hydrogen bond from Im-N3 to G-NH₂. In fact, the ImPy sections of the polyamide presented here are structurally similar to the compounds originally synthesized by Lown and coworkers (Lown et al., 1986; Goodsell et al., 1995). Figure 31 shows a portion of the binding site with polyamide residues Im6- β 5-Im4 binding across from DNA purine residues G6-A7-G8, respectively. The Im-N3 to G-NH₂ hydrogen bonds (Im6-G6 lower left, Im4-G8 upper right) are shown by dotted lines.

Hydrogen bonding frequency was tested using the CARNAL module of the AMBER 6.0 software package (Kollman et al., 2000). The final mean structure was subjected to 15 ps of restrained molecular dynamics at 25 °C, and snapshots of the trajectory were analyzed for hydrogen bond formation within 3.3 Å N-N



Figure 31 View into the minor groove of the mean structure, with polyamide residues $Im6-\beta5-Im4$ proximal to DNA residues G6-A7-G8, respectively. The DNA bases are in black; the backbone is in blue; the polyamide is in white. Hydrogen bonds between imidazole N3 and guanine NH2 are shown by dotted lines connecting the donor (blue) and acceptor (red) pairs.

distance and 30° N-H•••N angle. It was found for 30 snapshots that Im1 and Im6 residues, which are locked into two contiguous ring subunits (i.e., Im1-Py2 and Im6-Py7), form hydrogen bonds with greater frequency (93 %) than the central Im4 residue (80 %). This result suggests that the central Im4, which is flanked by two β residues, is sampling more conformational space and is therefore more conformationally flexible at room temperature than Im1 and Im6.

The Importance of \beta-Alanine. We have previously reported a study in which the Im- β -Im and Im- β -Py subunits of a 1:1 binding polyamide were replaced with Im–Py–Im and Im–Py–Py, respectively (Dervan and Urbach, 2001). We found that the β to Py substitution was tolerated for Im- β -Py with a small energetic penalty. However, the $Im-\beta-Im$ to Im-Py-Im mutation completely eliminated specific binding. In accordance with β /Py mutation studies in the 2:1 motif, it is likely that the β is needed to reset the register for the following (Cterminal) residue (Turner et al., 1998). It appears that β is necessary for the two flanking Im residues in $Im-\beta-Im$ to orient properly in order to form hydrogen The Im- β -Py subunit forms only one Im•G hydrogen bond, and bonds. therefore the flexibility is not as critical. This is further supported by the ringring dihedrals between the plane of Im4 and the planes of Py2 (Im $-\beta$ -Py) and Im6 (Im $-\beta$ –Im). Figure 32 shows axial views of these dihedrals with average values (± standard deviation) for the 12 structures in the final ensemble. The Im/Py dihedral for $Im-\beta-Py$ of 33° is similar to the Im-Py-Py dihedrals (measured for bolded residues) observed in 2:1 crystal structures (Kielkopf et al., 1998a; 1998b). The Im/Im dihedral for Im- β -Im subunit is a significant 17°



Figure 32 Polyamide ring– β –ring dihedrals. The polyamide from the mean structure is shown as a stick model with a transparent molecular surface. Imidazoles are in red, with β -alanines in white and pyrroles in blue. (A) View down the dihedral axis between Py2 and Im4 (Py– β –Im). (B) View down the dihedral axis between Im4 and Im6 (Im– β –Im). Arrows point along the plane of each ring, perpendicular to the view. θ and θ' prime denote the average ring to ring dihedral angles (± standard deviation) over the 12 final structures.

larger. Although a high-resolution structure has yet to be determined for an Im–Py–Im-containing compound, we suggest that the increase in dihedral for Im– β –Im is due to the need for proper orientation in order to maximize Im-N3 to G-NH₂ hydrogen bonding. It is interesting to point out that only one β -alanine conformation is observed in Im4– β 5–Im6 vs. two degenerate conformers for Py2– β 3–Im4. If indeed the two β 3 conformers are real, it is conceivable that β 5 is conformationally constrained due to the propensity of the flanking Im residues to orient properly for hydrogen bond formation and thus force a larger ring-ring dihedral.

Footprinting data provided earlier in this thesis has established a partial recognition code for the 1:1 motif, whereby Im, Py, and β will bind across from A/T base pairs, but only Im is able to tolerate G/C. The structure presented here reveals the specific interactions underpinning the Im–G specificity and provides the first high-resolution structural model of a β residue in a 1:1 complex with DNA. Strong NOEs are observed from methylene protons in each β residue to the proximal A-H2 proton on the floor of the minor groove, as observed in the 2:1 motif (de Clairac et al., 1999). The structural ensemble shows these protons to be directly in contact, with internuclear distances in the range of 2.3 to 2.5 Å. This would suggest a strong steric violation when β is placed across from a G•C base pair, which helps to explain the 20- to 50-fold loss in affinity for this mismatch.

The Sequence-Dependence of Ligand Orientation. Laemmli and coworkers made the remarkable observation that the β -linked polyamides in a 1:1 complex prefer a single orientation, N–C with respect to the 3' – 5' direction of



Figure 33 G/C-Dependence of polyamide orientation. (A) Schematic illustrating the optimal alignment of polyamide NH groups with the DNA bases and showing that the inherent geometry of the Im residue coupled with the propeller twist of the G/C base pair allows overlap between Im-N3 and G-NH₂ groups. Propeller twisted base pairs are shown as gray boxes (pyrimidines) crossing over black boxes (purines), with purine labels at right. The Im-N3 sp² orbital is shown as its lone pair in red. Guanine's NH₂ group is drawn as HN- in blue. The bases are rotated away from the center in order to project the natural 38° twist onto two dimensions. (B) Analogous section of the mean NMR structure showing Im4 next to DNA residues A7 – G8 – A9 • T20 – C21 – T22. Purine bases are shown in black; pyrimidines in gray. Hydrogen bond donors and acceptors are shown in blue and red, respectively. (C) Upon mutation of G to C, the model suggests that the sp² orbital of Im-N3 would not overlap with G-NH₂. Therefore, the ligand binds in the opposite orientation (D) in order to restore this interaction.

5' – AAGAGAAGAG – 3', but in the opposite C – N oritation with respect to this strand when the G residues are mutated to C. In the 1:1 complex the polyamide does not distinguish G•C and C•G base pairs (Janssen et al., 2000; Urbach and Dervan, 2001). Based on the structure presented here, we believe that polyamide orientation is governed by a combination of the inherent geometry in the amide-

Im-amide unit in combination with the negative propeller twisting of base pairs. We assume that the uniform alignment of amide NH groups observed here is the driving force for the register between polyamide and DNA. The sp² lone-pair orbital on the Im-N3 atom has an inherently different orientation with respect to the flanking NH groups based on covalent bonding geometry. Due to the propeller twisting of base pairs, the G-NH₂ group is oriented more favorably for interaction with the sp² orbital of Im-N3 when the polyamide is oriented N – C with respect to the 3' – 5' direction of the G-containing strand. Figure 33 details this model, showing a section of the NMR structure (part B) with a schematic of this orientation (part A). When the G residues are mutated to C, the lone pair of each Im-N3 will not overlap with G-NH₂ (part C) unless the ligand orients in the opposite direction (part D). This model could apply to the orientation preference at poly A (Urbach and Dervan, 2001) by considering the partial positive potential presented by A-H2, similar to G-NH₂ but weaker, which could interact favorably with Im-N3.

Summary. Solution structure determination by NMR is an essential tool for studying molecular recognition phenomena. DNA structure determined by NMR is often poorly defined in the absence of a bound ligand, and in the present study it was indeed the chemical shift dispersion induced by the ligand that allowed for the virtually complete assignment of all NMR spectra, which resulted in the use of a large number of NOE-derived distance constraints to enforce the geometry of the complex. This enabled the determination at high resolution of the solution structure of a 1:1 polyamide:DNA complex comprising one full turn of the DNA helix. The structure presented here offers a close look at a polyamide containing imidazole and β -alanine residues bound to its DNA match site in a 1:1 motif. The complex reveals B-form DNA with a narrow minor groove and a large degree of negative propeller twist, which has been demonstrated to be stabilized by bifurcated hydrogen bonds donated from each polyamide NH group to proximal purine N3 and pyrimidine O2 atoms. Stabilization of the negative propeller twist by this interaction, in addition to the inherently rigid and narrow minor groove, is thought to be the reason that polyamides would bind 1:1 in A-tract-like sequences, but would have difficulty binding as 2:1.

The observed homogeneous register of amide NH groups with respect to the DNA is though to be the driving force for optimal ligand-DNA alignment. If this is so, the previously established G/C-dependent orientation preference of the polyamide could be explained by an inherent asymmetry in the projected angle of the Im-N3 lone pair sp² orbital with respect to the amide NH groups. Therefore, overlap of this orbital with the propeller-twisted guanine's NH₂ group is optimal when the polyamide is oriented N – C with respect to the 3' – 5' direction of the guanine-containing strand.

The final structural ensemble reveals specific hydrogen bonds between Im-N3 and G-NH₂. The Im – β – Im subunit requires a large dihedral so that both rings orient properly to form hydrogen bonds. Additionally, we are now able to understand from a structural perspective the observed A/T specificity of β -alanine within the 1:1 motif, based on its close contacts with the floor of the minor groove. These results set the stage for a more critical design of polyamides that can discriminate between 1:1 and 2:1 binding modes.

Linker-Dependent Conformational Control of Polyamide-DNA Binding Modes⁵

Purpose. It has been demonstrated that polyamides of type Im-β-ImPy can bind in 1:1 and 2:1 modes depending on rules for recognition inherent to each motif (Dervan and Urbach, 2001). This ambiguity of sequence targeting depending on stoichiometry has posed a serious design problem for the DNA recognition field. If a polyamide is to be effectively sequence selective in a genomic context, it will be necessary to control its binding mode. For example, the polyamide of sequence $Im-\beta-ImPy-\beta-Im-\gamma-\beta-Dp$ (2) (Dp = dimethylaminopropylamide) targets 5'-AAAGAGAAGAG-3' as a 1:1 complex and 5'-TAGCGCAGCGCTA-3' as a 2:1 complex. The γ -linked analogue, Im- β -ImPy- γ -Im- β -ImPy- β -Dp (17), binds as a hairpin with high affinity to its target sequence, 5'-TAGCGCT-3' (Turner et al., 1998). Given the structural similarity between 2 and 17, i.e., one methylene unit difference, it was a simple prediction that **17** would bind in a 1:1 mode. We examined this possibility and found that, indeed, **17** binds in both hairpin and 1:1 conformations (Figure 34). Therefore, an accurate prediction of the target DNA sequence would require the ability to control the ligand conformation. Here we ask the question whether polyamide conformation, and therefore the polyamide•DNA binding mode, can be controlled by tailoring the linkage between subunits.

Approach. The concept of linker-controlled binding modes is a familiar one. Trauger et al. demonstrated that by increasing the aliphatic linker length from β to γ , the hairpin binding mode is favored over the extended 2:1 motif

⁵ The text of this section is taken from Urbach et al., in preparation.



Figure 34 Equilibrium between hairpin (left) and extended (right) conformational binding modes for compound 17. The cartoons at top represent DNA with polyamide bound in the respective conformation. DNA is shown as a ladder. Polyamides are shown as dot models, with shaded and nonshaded circles representing imidazole and pyrrole, respectively, and gray diamonds indicating beta-alanine. The y-turn residue is shown both as a semicircle connecting the two subunits and as the symbol, γ .

(Trauger et al., 1996b). Wemmer and coworkers showed by NMR that β-linked polyamides can adopt a hairpin conformation, although the hairpin conformation is more accessible by the γ linker (de Clairac et at., 1997). During the course of this study, Boger and coworkers reported that the turn conformation of a β linker can be reinforced by substituting the α-(R)-proton of β with -OCH₃ (Woods et al., 2002). Herman et al. demonstrated the use of α-(R)-amino-substituted γ (the Herman turn, ^{H₂N} γ) to increase the binding affinity of hairpin polyamides (Herman et al., 1998). We hypothesized that the Herman turn could be used to disfavor the extended 1:1 binding mode, while favoring the



Figure 35 Chemical structures of polyamides 2, 17, 18, 19, 20, and 21.

hairpin mode. Accordingly, the polyamide Im-β-ImPy-^{H₂N}γ-Im-β-ImPy-β-Dp (**18**) and its acetamide analogue, Im-β-ImPy-^{Ac}γ-Im-β-ImPy-β-Dp (**19**), were prepared (Figure 35). In addition, we were interested in studying the conformational effects of linkage on eight-ring polyamides, such as ImPyPyPy-β-ImPyPyPy-β-Dp (**20**) and ImPyPyPy-γ-ImPyPyPy-β-Dp (**21**). As described above, the β-rich series (**2**, **17**, **18**, and **19**) targets hairpin and 1:1 binding sites, 5'-TAGCGCT-3' and 5'-AAAGAGAAGAG-3', respectively. Polyamide **21** is known to bind as a hairpin with subnanomolar affinity to the sequence 5'-TAGTACT-3' (Trauger et al., 1996b). The sequence 5'-AAAAGAAAAG-3' was designed for extended 1:1 binding of polyamides **20** and **21** based on rules for 1:1 recognition (Urbach and Dervan, 2001). For simplicity, a plasmid containing all four binding sites was



Figure 36 The designed insert cloned into plasmid pAU27. The targeted recognition sites are shown in bold type. Two sets of polyamides are shown as dot models. The two at left represent compounds **20** and **21** bound in the predicted 1:1 and hairpin conformations. The two at right represent compounds **2**, **17**, **18**, and **18** bound in the predicted 1:1 and hairpin conformations. The variable linker position is shown as a square containing the letter X.

constructed (Figure 36). Equilibrium association constants and binding site sizes for these compounds at their target sites were determined in order to derive the relative preference of each compound for hairpin versus 1:1 binding modes.

DNA Binding Affinity and Sequence Specificity. The synthesis and characterization of **2**, **17**, and **21** have been reported previously (Trauger et al., 1996b; Turner et al., 1998; Urbach and Dervan, 2001). Polyamides **18**, **19**, and **20** were prepared according to standard solid phase protocols (Baird and Dervan, 1996; Herman et al., 1998). Quantitative DNase I footprint titrations (Trauger and Dervan, 2001) were performed for polyamides **2** and **17-21** on the 288 bp PCR product of pAU27 in order to compare the equilibrium association constants for the resulting complexes. Polyamides **2**, **17**, **18**, and **19** were designed to bind in hairpin and 1:1 modes at DNA sites 5'-TAGCGCT-3' and 5'-AAAGAGAAGAG-3', respectively (Figure 37). Compound **2** (β linker) binds to the 1:1 site with very high affinity ($K_a = 1.5 \times 10^{10} \text{ M}^{-1}$), displaying more than 150-fold specificity versus the hairpin site ($K_a = 9.7 \times 10^7 \text{ M}^{-1}$) (Table 7). Polyamide **17** (γ linker) binds with similar affinities to the 1:1 ($K_a = 1.3 \times 10^8 \text{ M}^{-1}$) and 2:1 ($K_a = 7.6 \times 10^8 \text{ M}^{-1}$) sites, showing a modest 5-fold preference for the hairpin site. The ^{H₂N}γ-linked



Figure 37 (A - D) Quantitative DNase I footprinting experiments for polyamides **2**, **17**, **18**, and **19**, respectively, on the 288 bp, 5'-end-labeled PCR product of plasmid pAU27: lane 1, intact DNA; lane 2, G reaction; lane 3, A reaction; lane 4, DNase I standard; lanes 5-15, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, 100 nM polyamide, respectively. Each footprinting gel is accompanied by the following: (right top) Schematic illustrating the observed protection pattern derived from the MPE footprinting experiment, with the polyamides shown in the observed conformation; (right bottom) Langmuir binding isotherms derived from the DNase I footprinting experiment for the two designed sites, 5'-AAAGAGAAGAG-3' and 5'-TAGCGCT-3', as determined from a nonlinear least squares fit.



Figure 38 (A and B) Quantitative DNase I footprinting experiments for polyamides **20** and **21**, respectively, on the 288 bp, 5'-end-labeled PCR product of plasmid pAU27: lane 1, intact DNA; lane 2, G reaction; lane 3, A reaction; lane 4, DNase I standard; lanes 5-15, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, 100 nM polyamide, respectively. Each footprinting gel is accompanied by the following: (right top) Langmuir binding isotherms derived from the DNase I footprinting experiment for the two designed sites, 5'-AAAAAGAAAAG-3' and 5'-TAGTACT-3', as determined from a nonlinear least squares fit; (right bottom) schematic illustrating the observed protection pattern derived from the MPE footprinting experiment, with the polyamides shown in the observed conformation.

polyamide **18** binds with very high affinity to the hairpin site ($K_a = 1.2 \times 10^{10} \text{ M}^{-1}$), exhibiting more than 150-fold specificity versus the 1:1 site ($K_a = 7.6 \times 10^7 \text{ M}^{-1}$), which is an almost exact reversal of specificity compared to compound **2**. Polyamide **19** ($^{Ac}\gamma$ -linker) binds at high affinity to the hairpin site ($K_a = 8.5 \times 10^9 \text{ M}^{-1}$), with greater than 500-fold specificity versus the 1:1 site ($K_a = 1.6 \times 10^7 \text{ M}^{-1}$).

Polyamides **20** and **21** were designed to bind in hairpin and 1:1 modes at DNA sites 5'-TAGTACT-3' and 5'-AAAAAGAAAG-3', respectively (Figure 38). Compound **20** (β linker) binds with similar affinity (K_a \approx 3 x 10⁹ M⁻¹) to both

ibrium Association Constants, K _a (M ⁻¹)*								
Linker	Hairpin	1:1	Specificity (Hairpin/1-1)					
	5'-TAGCGCT-3'	5'-AAAGAGAAGAG-3'						
β	$9.7 (\pm 0.8) \ge 10^7$	$1.5(\pm 0.5) \ge 10^{10}$	0.0065					
γ	$7.6(\pm 0.6) \ge 10^8$	$1.3 (\pm 0.5) \ge 10^8$	5.8					
$\mathbf{H_2N}_{\gamma}$	$1.2(\pm 0.2) \ge 10^{10}$	$7.6(\pm 0.2) \ge 10^7$	160					
Acy	$8.5 (\pm 0.5) \ge 10^9$	$1.6(\pm 0.4) \ge 10^7$	530					
		1						

 $2.7(\pm 0.7) \ge 10^9$

 $1.8(\pm 0.2) \times 10^9 b$

Table 7. Equili

Polyamide Im-β-ImPy-(2) (17)(18)(19)

ImPyPyPy-

(20)

(21)

β

γ

* Values reported are the mean values from at least three DNase I footprint titration experiments, with the standard deviation given in parentheses. Assays were performed at 22 °C in a buffer of 10 mM Tris+HCl, 10 mM KCl, 10 mM MgCl₂, and 5 mM CaCl₂ at pH 7.0 ** Observed binding sites are indicated as boxes a and b.

5'-TAGTACT-3'

 $3.9(\pm 0.7) \ge 10^9$

 $1.0(\pm 0.2) \ge 10^{10}$

designed sites (Table 7). Polyamide **21** (γ -linker) binds with very high affinity to the hairpin site ($K_a = 1.0 \times 10^{10} \text{ M}^{-1}$), exhibiting a modest 5-fold selectivity over its single base-pair mismatch hairpin site, 5'-AAGAAAA-3' (mismatch base bolded), within the 1:1 binding site ($K_a = 1.8 \times 10^9 \text{ M}^{-1}$). All binding isotherms fit well to an n = 1 Hill equation, which is consistent with 1:1 polyamide:DNA stoichiometry (Figures 37 and 38).

Binding Site Size. Evidence for hairpin and 1:1 binding modes is provided by methidium propyl EDTA (MPE) footprinting, wherein the binding site size can be determined at high resolution (Trauger and Dervan, 2001). Binding modes were deduced from MPE footprint sizes on the basis of previously characterized hairpin and 1:1 complexes for polyamides 2, 17, and 21 at identical sites (Trauger et al., 1996b; Turner et al., 1998; Urbach and Dervan, 2001). For example, the eight-ring hairpin, **21**, occupies a seven base-pair binding site, whereas the prototypical 1:1 binder, 2, occupies an eleven base-pair binding site. Figures 37 and 38 display occupation histograms derived from the

1.4

5.6

MPE gels (not shown), illustrated to the right of the respective DNase gel. Polyamides **2**, **17**, **18**, and **19** were designed to bind in hairpin and 1:1 modes at DNA sites 5'-TAGCGCT-3' and 5'-AAAGAGAAGAG-3', respectively (Figure 37); and polyamides **20** and **21** were designed to bind in hairpin and 1:1 modes at DNA sites 5'-TAGTACT-3' and 5'-AAAAAGAAAAG-3', respectively (Figure 38). In all cases, polyamides occupy their target hairpin sites as hairpins. Polyamides **2** and **17-20** bind to their target 1:1 sites in an extended 1:1 binding mode. However, polyamide **21** occupies a smaller site within its designed 1:1 site. Close inspection of the occupied site reveals the binding mode of **21** as a hairpin bound to a single base-pair mismatch site (Figure 38 at left). Inspection of the relative footprint sizes in the DNase gels further supports the MPE results.

Discussion. Linker-dependent conformational control of binding modes is established here for two distinct classes of hairpin polyamides–those containing contiguous rings, and those with beta alanine-endowed conformational flexibility. In the flexible series, **2**, **17**, **18** and **19**, compounds of type Im- β -ImPy-**X**-Im- β -ImPy- β -R (**X** = β , γ , ^{H₂N} γ , and ^{Ac} γ ; R = Me or Dp) bind in hairpin and extended 1:1 modes to the designed sites, 5'-TAGCGCT-3' and 5'-AAAGAGAAGAG-3', respectively (Figure 37). Previous studies in this thesis on the β -linked polyamide **2** demonstrated its high affinity for the 1:1 site. The >150-fold specificity of **2** for 1:1 versus hairpin binding sites, presented here, may be attributed to the steric destabilization involved in the β residue adopting a hairpin conformation (de Clairac et al., 1997). Moreover, structural studies in 1:1 and 2:1 modes support the notion that β confers optimal polyamide-DNA alignment in an extended conformation (de Clairac et al., 1999; Urbach et al., 2002). Inspection of the DNase gel reveals that **2** bound with very high affinity to the 5'-AAAAAGAAAAG-3' site. Although this site was not intended for **2**, the rules for 1:1 recognition would suggest this as a match site (Urbach and Dervan, 2001). The γ -linked polyamide **17** contains one additional methylene unit in the linker, which provides the flexibility needed to fold into a hairpin, as established previously (de Clairac et al., 1997; Turner et al., 1998). Interestingly, **17** discriminates least between the hairpin and 1:1 sites. This result poses an important problem because hairpin polyamides, which most commonly employ the γ linker, are the most sequence-specific class of DNA-binding polyamides. Therefore, a linker that favors hairpin formation while disfavoring 1:1 and other alternative binding modes would be of great value.

It was predicted from molecular modeling of the 1:1 NMR complex that adding functionality to an aliphatic linker in the 1:1 motif would cause steric destabilization of this binding mode. Furthermore, we postulated that incorporation of the Herman turn ($^{H_2N}\gamma$) would favor hairpin binding while disfavoring the 1:1 mode. Indeed, this result was observed for the $^{H_2N}\gamma$ -linked polyamide **18**, which shows >150-fold preference for the hairpin binding site versus the 1:1 site. Therefore, polyamides **2** and **18** exhibit a reversal of preference for the hairpin and 1:1 binding sites. It should be noted that **18** contains a truncated tail in order to reduce bias in this series of compounds by maintaining a single positive charge. We have also tested the doubly charged Dp-tail analogue of **18**, finding that its recognition properties are virtually identical to **18**. Inspection of the data shown in Figure 37 reveals that **18** is relatively nonspecific, binding with high affinity to many other sites on the plasmid. Acetylating the amino group of the Herman turn and restoring the charged Dp tail markedly reduces all mismatch affinities with a negligible decrease in hairpin site affinity, as observed for compound **19**. Furthermore, the $^{Ac}\gamma$ -linked polyamide **19** exhibits >500-fold preference for hairpin versus 1:1 binding modes, a reversal of specificity effectively 82,000-fold compared to **2**. The exceptional specificity of **19** may be attributed to the limited mobility of the bulky acetamide group at the turn position, which may in turn limit alternative binding modes. These results demonstrate that by changing the linker residue we can control the ligand conformation and hence the DNA sequence preference.

Polyamides **20** and **21** contain contiguous four-ring subunits, which are inherently limited in conformational flexibility due in part to the continuous π conjugation in each subunit. Therefore, it is expected that these compounds should display different recognition properties than polyamides **2**, **17**, **18**, and **19**. The β-linked polyamide **20** binds in an extended mode to the designed 1:1 sequence, 5'-AAAAAGAAAAG-3', and as a hairpin to the designed site, 5'-TAGTACT-3'. However, unlike the β-linked polyamide **2** described above, **20** does not show a preference for hairpin or 1:1 modes, binding both sites with similar subnanomolar affinities. On the other hand, the γ-linked compound **21** binds as a hairpin to both designed sites, tolerating a single base-pair mismatch within the 1:1 binding site, 5'-AAGAAAA-3', rather than adopting an extended conformation. This result provides further evidence that contiguous ring polyamides linked by γ have inherently unique DNA recognition properties, and as such are not inclined to adopt an extended binding conformation (Trauger et al., 1996a). Implications for the Design of Minor Groove Binding Polyamides. The results presented here indicate that hairpin and extended modes of polyamide-DNA binding, which are dependent on the ligand conformation, can be controlled by the choice of linkage between subunits (Figure 39). For β -rich polyamides, the hairpin binding mode is favored by incorporating a substituted γ residue, such as H_2N_{γ} or Ac_{γ} , to link the flexible polyamide subunits. In addition to disfavoring the 1:1 mode, the Ac_{γ} linker substantially improves hairpin binding affinity and specificity. Alternatively, the 1:1 binding mode can be favored by using a β residue as the linker. We are particularly interested in favoring the hairpin binding conformation because of its high capacity for programmable DNA sequence specificity. Therefore, it is reassuring that for polyamides with multiple contiguous heterocycles, such as eight-ring polyamides, extended 1:1 binding modes do not compete well with the favored hairpin motif.



Figure 39 Control of polyamide-DNA binding modes. The γ -linked polyamide shown in the center binds well in both hairpin and 1:1 modes. Replacement of γ with β favors the 1:1 mode by 150-fold (left), whereas substitution of γ at the α -(R) position with acetamide (^{Ac} γ) favors the hairpin mode by 560-fold, thereby enforcing an 82,000-fold reversal of specificity.