Pergamon

SCIENCE DIRECT®

Bioorganic & Medicinal Chemistry 11 (2003) 4333–4340

BIOORGANIC & MEDICINAL CHEMISTRY

Shape Selective Recognition of T·A Base Pairs by Hairpin Polyamides Containing N-Terminal 3-Methoxy (and 3-Chloro) Thiophene Residues

Shane Foister, Michael A. Marques, Raymond M. Doss and Peter B. Dervan*

The Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125, USA

Received 14 April 2003; accepted 23 July 2003

Abstract—Hairpin polyamides selectively recognize predetermined DNA sequences with affinities comparable to naturally occurring proteins. Internal side-by-side pairs of unsymmetrical aromatic rings within the minor groove of DNA distinguish each of the four Watson–Crick base pairs. In contrast, N-terminal ring pairs exhibit less specificity, with the exception of Im/Py targeting G·C base pairs. In an effort to explore the sequence specificity of new ring pairs, a series of hairpin polyamides containing 3-substituted-thiophene-2-carboxamide residues at the N-terminus was synthesized. An N-terminal 3-methoxy (or 3-chloro) thiophene residue paired opposite Py displayed 6- (and 3-) fold selectivity for T·A relative to A·T base pair, while disfavoring G,C base pairs by >200-fold. Our data suggests shape selective recognition with projection of the 3-thiophene substituent (methoxy or chloro) to the floor of the minor groove.

Introduction

Polyamides composed of *N*-methylpyrrole (Py), *N*-methylimidazole (Im), and *N*-methylhydroxypyrrole (Hp) amino acids are crescent-shaped ligands that bind sequence specifically in the minor groove of DNA and have the potential to modulate gene expression. The specificity of DNA recognition arises from interactions between the edges of the Watson–Crick base pairs and antiparallel aromatic amino acid ring pairs oriented N→C with respect to the 5′→3′ direction of the DNA helix.[1–3] Covalent head-to-tail linkage of two polyamide strands by γ-aminobutyric acid constitutes the hairpin motif, in which opposing residues from each strand are locked into cofacial pairs.[4,5] Im/Py distinguishes G·C from C·G and both of these from T·A/A·T base pairs while a Py/Py pair binds both T·A and A·T in preference to G·C/C·G. The exocyclic amino group of guanine imparts G·C specificity to Im/Py pairs through formation of a specific hydrogen bond with N3 of Im. Binding of Py/Py is disfavored at G,C base pairs by destabilizing steric interactions between the C3-H of Py and the guanine amino group.[6,7] The replacement of C3-H of one Py with hydroxyl creates the Hp/Py pair which exploits the steric fit and hydrogen bond acceptor potential of thymine-O2 as well as the destabilizing steric interaction with the bulkier adenine ring to gain specificity for T·A.[8,9]

The above pairing rules have been used to design hundreds of synthetic ligands that bind predetermined DNA sequences. However, many sequences remain difficult to target, likely due to sequence dependent microstructure variations in minor groove width or curvature. Furthermore, the specificity of cofacial aromatic amino acid pairings depend on their context (position) within a given hairpin polyamide. For example, Im/Py pairings show comparable specificity for G·C at both *terminal* and *internal* positions.[1] Conversely, Hp/Py pairings do not specify T·A at the N-terminus of hairpin polyamides.[10] The context dependence of Hp is presumably a result of the conformational freedom inherent to an N-terminal aromatic residue. The absence of a second 'groove-anchoring' carboxamide allows terminal rings to bind DNA in either of two conformations. For a terminal Hp residue, a rotamer with the hydroxyl recognition element oriented away from the floor of the minor groove could be stabilized by intramolecular hydrogen bonding between the C3-OH and the carbonyl oxygen of the 2-carboxamide. For terminal 2-hydroxybenzamide residues, some measure of T·A selectivity

*Corresponding author. Tel.: +1-626-395-6002; fax: +1-626-683-8753; e-mail: dervan@caltech.edu

0968-0896/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/S0968-0896(03)00502-9

was recovered by creating steric bulk at the 6-position to force the hydroxyl recognition element into the groove.[10] However, N-terminal pairings capable of binding T·A, with affinity and specificity comparable to those of Im/Py for G·C, remain to be devised.

The fidelity of minor groove recognition by N-terminal Im/Py pairings in hairpin polyamides can be rationalized by a combination of both stabilizing and destabilizing forces which favors the rotamer with N3 in the groove and *N*-methyl out. Rotation of a terminal Im residue in the opposite conformer, orienting N3 away from the minor groove, would create unfavorable lone pair interactions with the proximal carboxamide oxygen, disrupt a favorable hydrogen bond with the exocyclic amine of G, and project an *N*-methyl group to the DNA floor which is presumably sterically unfavorable. We address in this paper whether T recognition element could be designed using the asymmetric cleft of a T·A base pair as the basis for *shape selective discrimination*. Recent work from our group has indicated that the polarizable sulfur atom of thiophene heterocycles might serve this purpose.[11]

Our experimental design anticipated that substitution of the 3-position of a thiophene-2-carboxamide scaffold could be used to favor an anti ('sulfur down') conformation at the N-terminus by disfavoring contact of the 3-substituent with the floor of the minor groove (Fig. 1). It was envisioned that the electronic properties of the 3-substituent might be used to tune the polarization of the sulfur atom, allowing a more complementary fit with thymine in the minor groove. We attempt here to expand the repertoire of DNA sequences that can be targeted using hairpin polyamides by investigating the DNA recognition properties of a series of N-terminal residues consisting of 3-substituted-thiophene-2-carboxamide heterocycles. Quantitative DNAse I footprinting was used to determine the affinity of eight novel N-terminal 3-substituted thiophene rings residues, paired opposite Py, for each of the four Watson–Crick base pairs (Fig. 2). Ab initio computational modeling was used to guide interpretation of the experimental results.

Results

Monomer synthesis (Fig. 3)

Methyl 3-aminothiophene-2-carboxylate was Boc-protected and the resulting ester was saponified to yield 3-[(*tert* - butoxy)carbonylamino] - 2 - thiophenecarboxylic acid (**11**). Methyl 3-hydroxythiophene-2-carboxylate (**12**) was prepared by cyclization of methylthioglycolate and methyl-2-chloroacrylate in methanolic sodium methoxide.[12] Alkylation of (**12**) with iodomethane and subsequent hydrolysis of the methyl ester gave 3-methoxy-2-thiophenecarboxylic acid (**13**). 3-Fluorothiophene-2-carboxylic acid (**14**) was synthesized as described previously.[13] The remaining 3-substituted-thiophene-2-carboxylic acids were obtained from commercial sources.

Polyamide synthesis (Fig. 4)

Polyamide resin (**R1**) was prepared using manual solid-phase synthetic techniques described previously.[14] Treatment of this resin with trifluoroacetic acid (80% TFA in CH_2Cl_2) yielded a support-bound amine that was subsequently acylated with the appropriate, HBTU-activated, thiophene-2-carboxylic acids. Acylation of (**R1**) by (**11**) and removal of the Boc protecting group with TFA yielded resin (**R3**) which was cleaved with dimethylaminopropylamine (Dp) to give polyamide (**3**). Treatment of (**R3**) with acetic anhydride prior to

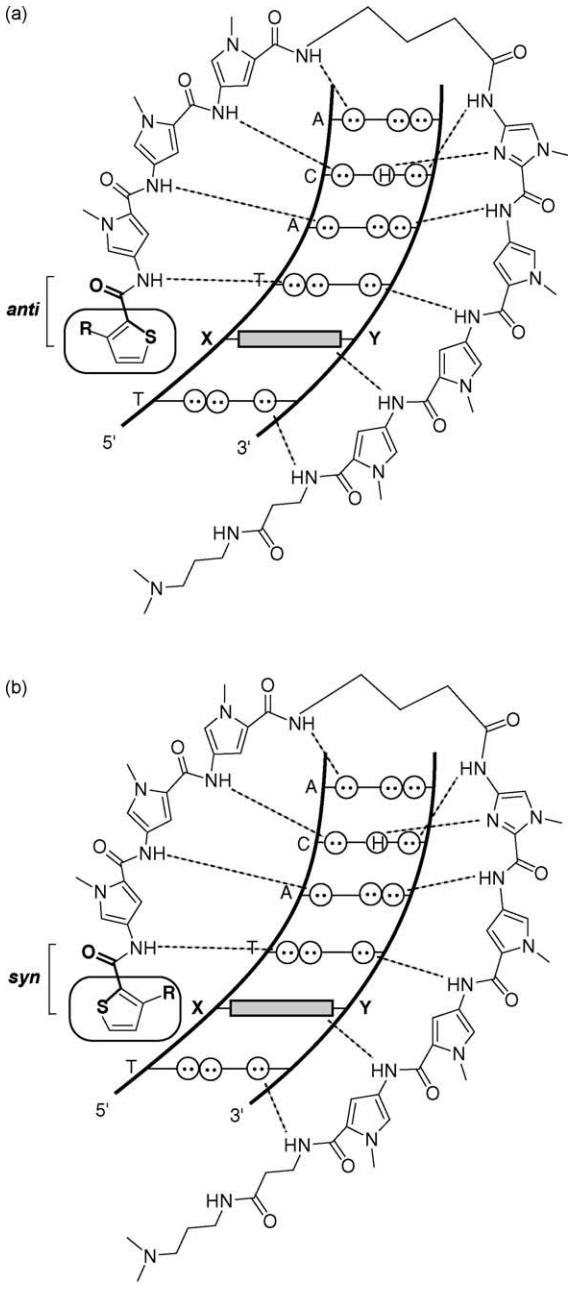

Figure 1. Proposed binding models for hairpin polyamides with 5′-TXTACA-3′ site. A circle enclosing two dots represents lone pairs of N3 of purines and O2 of pyrimidines. A circle containing an H represents the exocyclic amine of guanine. Putative hydrogen bonds are indicated by dashed lines; (a) N-terminal residue drawn in 'sulfur down' syn conformation; (b) N-terminal residue drawn in 'sulfur up' anti conformation.

cleavage with Dp gave polyamide (**4**). The remaining polyamides (**1**, **2**, **6**–**8**) were cleaved from resin with Dp immediately following acylation of the carboxylic acid. Treatment of (**6**) with sodium thiophenoxide in DMF gave (**5**). Crude products were purified by reversed-phase HPLC and characterized by MALDI-TOF mass spectrometry.

DNA binding energetics

Quantitative DNase I footprinting titration experiments (10 mM Tris–HCl, 10 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, pH 7.0, 22 °C) were performed on 5′-^{32}P end-labeled, 285 bp PCR product from plasmid pCW15.[10] This plasmid contains four binding sites that vary at a single N-terminal position, 5′-A T **N** T A C A-3′, where N = T, A, G, C. The DNA sequence specificity of novel thiophene-2-carboxamides was evaluated by comparing their affinities for each Watson–Crick base pair to those of *N*-methylimidazole (Im) and *N*-methylpyrrole (Py) (Fig. 5 and Table 1). The divergent behavior of control polyamides **9** and **10** illustrate the need for development of new N-terminal residues. A terminal Im/Py pairing **9** binds its match sequence, 5′-A T **G** T A C A-3′, with high affinity ($K_a = 7\times10^{10}$ M^{-1}) while showing >15-fold preference for G·C relative to T·A, A·T, and C·G base pairs. Terminal Py/Py pairings **10**, on the other hand, are characterized by little sequence specificity, binding T·A, A·T, and G·C with comparable affinity.

Within the thiophene-2-carboxamide series, *an unsubstituted thiophene ring* Tp **1** paired with Py shows little

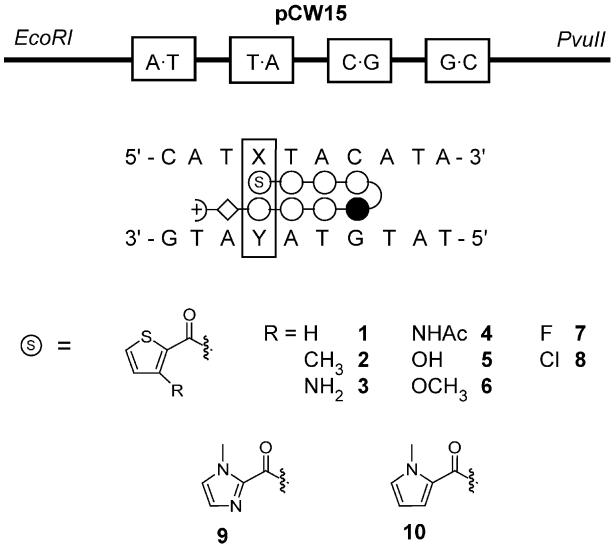

Figure 2. (top) pCW15 plasmid design; (bottom) ball and stick model of hairpin polyamides varying the N-terminal residue. Shaded and non-shaded circles represent imidazole and pyrrole residues, respectively. A circle containing an S denotes an N-terminal thiophene (R = 1–8) residue.

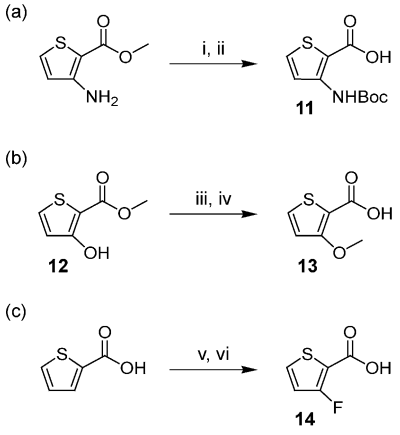

Figure 3. (a) Synthesis of 3-[(*tert*-butoxy)carbonylamino]-2-thiophenecarboxylic acid (**11**): (i) Et_3N, Boc_2O, DMAP, acetone; (ii) 50% NaOH, MeOH. (b) Synthesis of 3-methoxy-2-thiophenecarboxylic acid (**13**): (iii) K_2CO_3, CH_3I, acetone, acetonitrile, reflux; (iv) 50% NaOH, MeOH. (c) Synthesis of 3-Fluoro-2-thiophenecarboxylic acid (**14**): (v) *n*BuLi (2.2 equiv), THF, −78 °C, 0.5 h; (vi) $(PhSO_2)_2NF$, THF, −78 °C→rt.

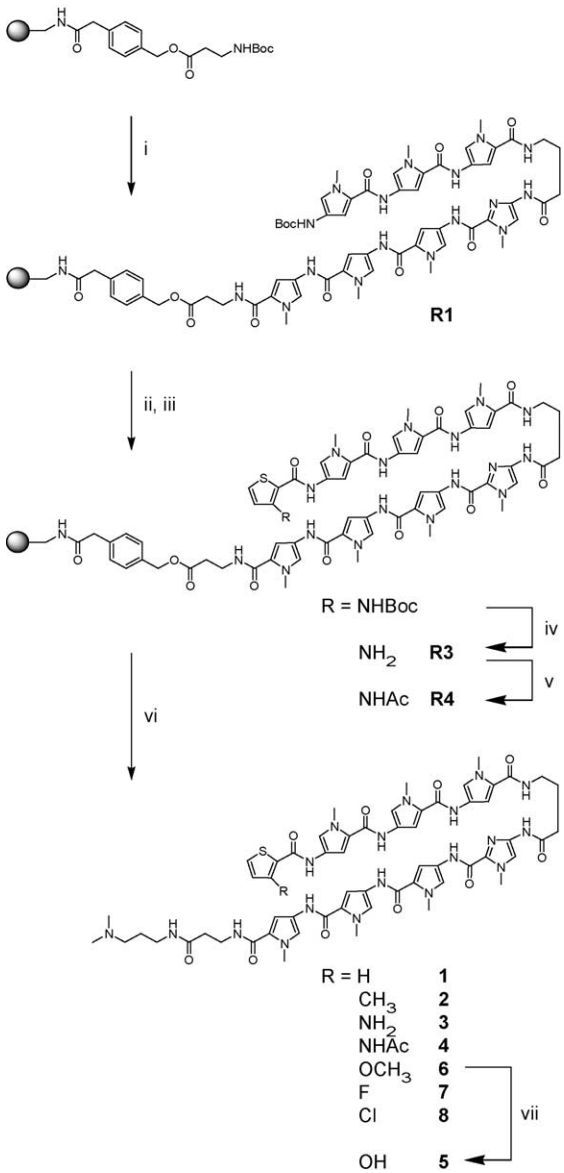

Figure 4. Synthesis of hairpin polyamides: (i) Synthesis of polyamide resin by standard solid-phase techniques;[14] (ii) TFA, CH_2Cl_2; (iii) 3-*R*-thiophene-2-CO_2H, HBTU, DMF, DIEA; (iv) TFA, CH_2Cl_2; (v) Ac_2O, DMF, DIEA; (vi) Dp, 40 °C; (vii) PhSH, NaH, DMF, 100 °C.

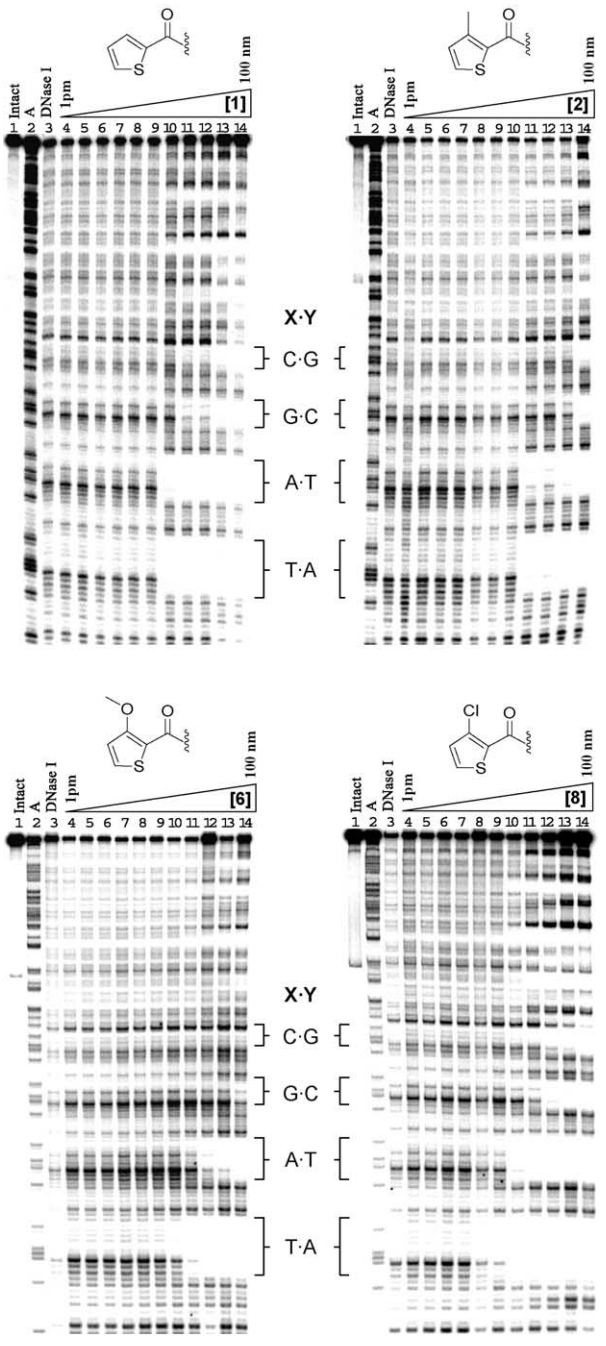

Figure 5. Quantitative DNase I footprint titration experiments for polyamides **1**, **2**, **6**, and **8** on pCW15 PCR product. Lane 1, intact DNA; lane 2, A reaction; lane 3, DNase I standard; lanes 4–14, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, 100 nM polyamide, respectively. The chemical structure of each N-terminal residue is included at the top of the gel and the four binding sites are labeled.

sequence specificity. Addition of a methyl group at the 3-position exerts a dramatic effect on sequence specificity; A,T favored over G,C. Polyamide **2** binds both T·A and A·T with a 140-fold preference for T,A relative to G,C. Amino **3**, acetamido **4**, or hydroxyl **5** substituents at the 3-position of thiophene all distinguish T,A from G,C but again do not distinguish T·A from A·T. Remarkably, a 3-methoxythiophene **6** paired with Py shows good affinity for T·A ($K_a = 2\times10^9$ M^{-1}) with 6-fold selectivity for T·A relative to A·T and >200-fold

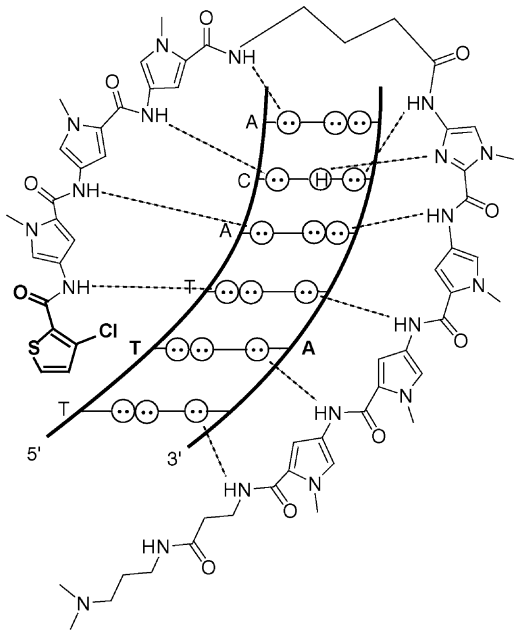

Figure 6. Hypothetical binding model to explain selectivity for T·A over A·T. The thiophene is in the *anti* conformation (sulfur away from the groove) and the 3-chloro substituent points to the minor groove floor.

specificity relative to G,C. Fluoro **7** and chloro **8** substituted thiophene paired with Py afford higher binding affinities for T·A but a lower selectivity (3-fold) for T·A over A·T (Fig. 6).

Molecular modeling (Table 2)

Molecular modeling was performed using the *Spartan Essential* software package.[15] N-Terminal residues were first minimized as methyl-2-carboxamides using an AM1 model. The resulting geometry was then subjected to ab initio calculation using the Hartree–Fock model with a 6-31G* polarization basis set. The partial electrostatic charge of the sulfur atom, δ_S, and the partial charge of the peripheral atom of the 3-substituent, δ_R, were examined for each novel thiophene residue. The electronic influences of 3-substituents on the polarization of the sulfur atom follow expected trends, with partial electronic charge, δ_S, decreasing as follows: **4** > **7** > **1** > **6** > **5** > **8** > **2** > **3**. The electronic surfaces presented by the 3-substituents, δ_R, were also calculated and found to decrease as follows: **5** > **3** > **4** > **1** > **6** > **2** > **7** > **8**.

The relative energy differences between minimized syn and anti conformations was also examined for each new thiophene ring. Hairpins **1**–**5** show a preference for the anti, or 'sulfur down,' conformation which may be attributed to lone pair repulsions between the sulfur atom and the carbonyl oxygen of the 2-carboxamide moiety. This bias can be reinforced by favorable hydrogen bonding interactions between 3-substituents and the carboxamide as in polyamides **3**–**5**. By contrast, polyamides **6**–**8** display a bias for the syn, or 'sulfur up,' conformation, possibly owing to more severe electronic clashes between the electron rich 3-substituents and the

Table 1. Equilibrium association constants (M^{-1})[a]

Polyamide	Ring pairing	R	T·**A**	A·T	G·C	C·G
1	Tp(1)/Py	H	$6.0\ (0.7)\times10^{9}$	$4.7\ (0.7)\times10^{9}$	$4.3\ (0.4)\times10^{8}$	$2.2\ (0.3)\times10^{9}$
2	Tp(2)/Py	CH_3	$2.3\ (0.4)\times10^{9}$	$1.4\ (0.2)\times10^{9}$	$1.0\ (0.4)\times10^{7}$	$1.0\ (0.3)\times10^{7}$
3	Tp(3)/Py	NH_2	$6.3\ (1.0)\times10^{9}$	$4.6\ (0.6)\times10^{9}$	$7.8\ (0.9)\times10^{8}$	$2.2\ (0.3)\times10^{8}$
4	Tp(4)/Py	NHAc	$5.9\ (0.3)\times10^{9}$	$2.9\ (0.1)\times10^{9}$	$6.6\ (0.4)\times10^{8}$	$6.0\ (0.2)\times10^{8}$
5	Tp(5)/Py	OH	$6.2\ (0.6)\times10^{9}$	$4.5\ (0.6)\times10^{9}$	$2.1\ (0.3)\times10^{8}$	$8.4\ (0.1)\times10^{7}$
6	Tp(6)/Py	OCH_3	$2.0\ (0.4)\times10^{9}$	$3.2\ (0.6)\times10^{8}$	$\leq1.0\times10^{7}$	$\leq1.0\times10^{7}$
7	Tp(7)/Py	F	$1.2\ (0.2)\times10^{10}$	$3.9\ (0.3)\times10^{9}$	$3.7\ (0.4)\times10^{8}$	$2.9\ (0.3)\times10^{8}$
8	Tp(8)/Py	Cl	$1.3\ (0.2)\times10^{10}$	$3.7\ (0.2)\times10^{9}$	$3.1\ (0.6)\times10^{8}$	$2.1\ (1.1)\times10^{8}$
9	Im/Py		$3.8\ (0.3)\times10^{9}$	$2.8\ (0.2)\times10^{9}$	$7.0\ (0.9)\times10^{10}$	$3.2\ (0.4)\times10^{9}$
10	Py/Py		$5.1\ (0.6)\times10^{9}$	$3.1\ (0.3)\times10^{9}$	$1.1\ (0.1)\times10^{9}$	$2.6\ (0.3)\times10^{8}$

[a]Values reported are mean results determined by at least three DNase I footprint titrations, with standard deviation given in parentheses. Assays were performed at 22 °C in a buffer containing 10 mM Tris–HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$ at pH 7.0.

Table 2. Physical properties determined by molecular modeling[a]

Polyamide	R	d_S[b]	d_R[b] (R)[c]	E_{syn}–E_{anti}[d]	A_R/A_H[e]
1	H	−0.065	0.124, (**CH**)	0.262	1.00
2	CH_3	−0.093	0.036, (**CH_3**)	1.739	1.11
3	NH_2	−0.117	0.426, (**NH_2**)	10.289	1.07
4	NHAc	−0.057	0.320, (**NH**)	10.308	1.31
5	OH	−0.069	0.512, (**OH**)	7.142	1.04
6	OCH_3	−0.068	0.063, (**OCH_3**)	−7.298	1.17
7	F	−0.061	−0.227, (**F**)	−5.043	1.03
8	Cl	−0.076	−0.106, (**Cl**)	−13.293	1.09

[a]Ab initio calculations were performed with Spartan Essential software package using Hartree–Fock model with 6-31G* polarization basis set.
[b]Partial electrostatic charges are given in arbitrary units.
[c]Partial charges given for atoms in bold.
[d]Energy differences are reported in kcal/mol.
[e]Ratio of surface area, A, of 3-substituent to hydrogen.

carboxamide relative to those of the sulfur atom. Finally, the solvent exposed surface area of each 3-substituted thiophene was compared to the unsubstituted thiophene ring to assess the steric contribution of the 3-substituent and surface area was found to increase in the following order: **1** < **7** < **5** < **3** < **8** < **2** < **6** < **4**.

Discussion

The observed equilibrium association constants for polyamides **1**–**5** support an anti conformation for the N-terminal thiophene residue. The binding preference of these compounds for T·A/A·T relative to G·C/C·G might be a result of unfavorable steric clashes between the sulfur atom and the exocyclic amino group of guanine. The binding properties of N-terminal, 3-methylthiophene-2-carboxamide residues also correlate well with values derived from internal contexts, where the sulfur down conformation is stringently enforced.[11]

It was envisioned that polyamides **6**–**8** would assume an anti conformation by sterically disfavoring contact between the bulky 3-substituents and the floor of the minor groove. Quantitative DNase I footprinting revealed modest selectivity for T·A relative to A·T and excellent specificity for both of these over G·C/C·G. However, binding properties of 3-methoxy- and 3-chlorothiophene residues determined at the N-terminus do not correlate with those derived from internal positions. Furthermore, molecular modeling indicated that unfavorable lone pair interactions favor the sulfur up conformation. Taken together, these results could suggest that the *electron rich methoxy and halogen groups are projected toward the minor groove*. The greater size of the methoxy group relative to the halogens might account for the lower affinity of polyamide **6** relative to **7** and **8**. The greater T·A selectivity of **6** could stem from the more complementary positive electronic surface presented to the thymine carbonyl by the methoxy protons relative to the negatively polarized halogen atoms.

N-Terminal 3-methoxy (or 3-chloro) thiophene-2-carboxamide residues when paired with Py demonstrate selectivity for T·A versus A·T. This represents an important step toward expanding the array of DNA sequences that can be targeted by minor groove-binding polyamides. Confirmation of the novel binding model with chloro (or methoxy) responsible for minor groove shape selective recognition, not sulfur, must await structure studies in solution by NMR.

Experimental

Materials

Methyl-2-chloroacrylate was obtained from Acros. Benzenethiol, di-*tert*-butyl dicarbonate (Boc_2O), fluorotrichloromethane (NMR grade), iodomethane, methyl thioglycolate, 3-methyl-2-thiophenecarboxylic acid, *N*,*N*-diisopropylethylamine (DIEA), *N*,*N*-dimethylaminopropylamine (Dp), *N*,*N*-dimethylaminopyridine (DMAP), *N*,*N*-dimethylformamide (DMF), *N*-fluorobenzenesulfonimide, potassium carbonate, sodium metal, tetrahydrofuran (THF), 2-thiophenecarboxylic acid, and triethylamine (TEA) were purchased from Aldrich. Methyl 3-amino-2-thiophenecarboxylate and 3-chloro-2-thiophenecarboxylic acid were obtained from Alfa Aesar. Boc-β-alanine-(4-carbonylaminomethyl)-benzyl-ester-copoly(styrene-divinylbenzene)resin (Boc-β-Pam-resin) and 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) were

purchased from NOVA Biochem. Trifluoroacetic acid (TFA) was purchased from Halocarbon. *N*-Butyllithium was obtained as a solution in hexanes from Strem. All other solvents were reagent grade from EM.

NMR spectra were recorded on a Varian spectrometer at 300 MHz in DMSO-d_6 or $CDCl_3$ with chemical shifts reported in parts per million relative to residual solvent. Fluorotrichloromethane was used as an internal standard for ^{19}F NMR. UV spectra were measured on a Hewlett-Packard Model 8452A diode array spectrophotometer. High-resolution EI mass spectra were recorded at the Mass Spectrometry Laboratory at the California Institute of Technology. Matrix-assisted, laser desorption/ionization time of flight mass spectrometry (MALDI-TOF-MS) was conducted at the Protein and Peptide Micronanalytical Facility at the California Institute of Technology.

Monomer synthesis

3-[(*tert*-Butoxy)carbonylamino]-2-thiophenecarboxylic acid (11). A mixture of methyl 3-amino-2-thiophenecarboxylate (2.53 g, 15.9 mmol), Boc_2O (7.64 g, 35 mmol), and DMAP (2.04 g, 16.7 mmol) was dissolved in acetone (15 mL) and TEA (5 mL). The reaction mixture was stirred vigorously for 4 h and diluted to a volume of 75 mL with dichloromethane. The resulting solution was washed with cold 1 N HCl (3×50 mL), 1 N NaOH (3×50 mL), and brine (50 mL). The dichloromethane solution was then dried over $MgSO_4$, filtered, and concentrated in vacuo to yield a yellow oil. The crude product was loaded onto a short plug of silica and eluted with 9:1 hexanes/ethyl acetate to yield a pale yellow solid (1.2 g) that was used without further purification. The solid was dissolved in methanol (76 mL) and 50% NaOH (4 mL) and the mixture was stirred for 4 h. The reaction was diluted to a volume of 160 mL with water and concentrated briefly in vacuo. The remaining aqueous solution was washed with diethyl ether (2×80 mL), cooled in an ice bath, and cautiously acidified to pH 2 with sulfuric acid. The suspension was washed with ethyl acetate (3×50 mL) and the combined organic washes were dried over $MgSO_4$, filtered, and concentrated in vacuo to yield (**11**) as a white solid (0.79 g) in 69% yield over two steps. ^1H NMR (DMSO-d_6) δ 9.43 (s, 1H), 7.80 (d, $J=5.4$ Hz, 1H), 7.72 (d, $J=5.4$ Hz, 1H), 1.46 (s, 9H); ^{13}C NMR (75 MHz, DMSO-d_6) δ 165.8, 151.8, 144.9, 133.1, 121.2, 109.9, 81.5, 28.6; EI-MS *m*/*e* 243.0563 (M^+ calculated 243.0565 for $C_{10}H_{13}NO_4S$).

Methyl 3-hydroxy-2-thiophenecarboxylate (12). To dry methanol (81 mL), under nitrogen, was added sodium metal (3.68 g, 304 mmol). After H_2 evolution has ceased, the solution was cooled to 0 °C and methyl thioglycolate (10 g, 179 mmol) was added dropwise. A solution of methyl-2-chloroacrylate (10.88 g, 179 mmol) in methanol (21 mL) was then added slowly, resulting in the formation of yellow precipitate. The solution was allowed to warm to ambient temperature and stirred for 2 h. The solvent was removed in vacuo to give a dark yellow solid that was acidified to pH 2 with 4 N HCl. The resulting aqueous solution was extracted with dichloromethane (3×150 mL) and the combined organic solutions were washed with water (3×150 mL), dried over $MgSO_4$, filtered, and concentrated to give a dark oil. The oil was subjected to column chromatography on silica gel (20:1 hexanes/ethyl acetate) to give (**12**) (18.4 g) as a crystalline solid in 64% yield. TLC (20:1 hexanes/ethyl acetate) R_f 0.47; ^1H NMR ($CDCl_3$) δ 9.58 (s, 1H), 7.59 (d, $J=5.7$ Hz, 1H), 6.75 (d, $J=4.8$ Hz, 1H), 3.90 (s, 3H); ^{13}C NMR (75 MHz, $CDCl_3$) δ 166.8, 164.7, 131.7, 119.4, 52.2; EI-MS *m*/*e* 158.0039 (M+ calculated 158.0038 for $C_6H_6O_3S$).

3-Methoxy-2-thiophenecarboxylic acid (13). A mixture of (**12**) (2.3 g, 14.5 mmol), K_2CO_3 (5.02 g, 36.3 mmol), and iodomethane (10.4 g, 73 mmol) was suspended in acetone (25 mL) and acetonitrile (5 mL). The resulting mixture was stirred vigorously at reflux for 3 h. The reaction was filtered and the resulting solid was washed with acetone and dichloromethane. The reaction and washes were combined and concentrated in vacuo to yield a yellow solid (1.9 g) that was used without further purification. The yellow solid was dissolved in methanol (17 mL) and 50% NaOH (3 mL) and was stirred for 3 h. The reaction was diluted to 40 mL with water and concentrated briefly in vacuo to yield a suspension. The aqueous suspension was washed with diethyl ether (2×25 mL), cooled to 0 °C, and acidified to pH 2 with 10% sulfuric acid. The aqueous mixture was then washed with dichloromethane (3×50 mL) and the combined organic washes were dried over sodium sulfate, filtered, and concentrated in vacuo to give a yellow oil. The oil was suspended in 3:1 petroleum ether/dichloromethane at −20 °C overnight. Filtration gave (**13**) as a finely divided white solid (0.736 g) in 33% yield over two steps. TLC (4:1 ethyl acetate/hexanes) R_f 0.5; ^1H NMR (DMSO-d_6) δ 12.4 (s, 1H), 7.74 (d, $J=5.7$ Hz, 1H), 7.06 (d, $J=5.4$ Hz, 1H), 3.85 (s, 3H); ^{13}C NMR (75 MHz, DMSO-d_6) δ 163.0, 161.9, 131.9, 118.0, 109.9, 59.4; EI-MS *m*/*e* 158.0034 (M+ calcd 158.0038 for $C_6H_6O_3S$).

3-Fluoro-2-thiophenecarboxylic acid (14). 2-Thiophenecarboxylic acid (1.7 g, 13.3 mmol) was dissolved in anhydrous THF (30 mL) and the solution was cooled to −78 °C under Ar, with stirring. *n*-Butyllithium (18.3 mL, 29.3 mmol) in hexanes was added to the above solution and the mixture was stirred for 30 min. A solution of *N*-fluorobenzenesulfonimide (5 g, 15.9 mmol) in THF (30 mL) was then added and the resulting solution was stirred at −78 °C for 4 h and allowed to warm to ambient temperature over a period of 6 h. The reaction was diluted with diethyl ether (100 mL), cooled to 0 °C, and 1 N HCl (15 mL) was added to give a biphasic mixture. The aqueous layer was isolated and washed with diethyl ether (3×50 mL). The combined ethereal layers were dried over $MgSO_4$, filtered, and concentrated in vacuo to yield an orange oil. The oil was subjected to column chromatography on silica gel using 1:1 hexanes/ethyl acetate as the eluent. (**14**) was obtained as a slightly brown solid (0.777 g) in 40% yield. TLC (1:1 ethyl acetate/hexanes) R_f 0.17; ^1H NMR ($CDCl_3$) δ 10.7 (s, 1H), 7.53 (dd, $J=5.4$, 3.6 Hz, 1H),

6.89 (d, $J=5.4$ Hz, 1H); ^{13}C NMR (75 MHz, $CDCl_3$) δ 166.2 (d, $J=3.5$ Hz), 161.5 (d, $J=278$ Hz), 132.0 (d, $J=10$ Hz), 118.9 (d, $J=24.7$ Hz), 113.6; ^{19}F NMR (282 MHz, $CDCl_3$, $CFCl_3$) δ −65.2 (d, $J=6$ Hz); EI-MS m/e 145.9838 (M+ calcd 145.9838 for $C_5H_3FO_2S$).

Polyamide synthesis

Hairpin polyamides were synthesized from intermediate resin (**R1**) that was prepared according to published protocols using Boc-β-alanine-Pam resin (50 mg, 0.59 mmol/g).[14] Products were purified by reversed-phase HPLC and characterized by MALDI-TOF mass spectrometry.

(1). Resin (**R1**) was treated with 80% TFA in dichloromethane and washed thoroughly. A solution of 2-thiophenecarboxylic acid (19 mg, 0.148 mmol) and HBTU (28 mg, 0.079 mmol) in DMF (0.45 mL) and DIEA (0.5 mL) was mixed at 40 °C for 25 min and poured onto the deprotected resin. The resin slurry was shaken for 4 h at room temperature and filtered. After washing with DMF, the resin was cleaved with Dp (1 mL) at 40 °C for 4 h. The crude product was purified by reversed-phase HPLC to afford (**1**) as a white solid upon lyophilization (3.3 mg, 9% recovery). MALDI-TOF-MS m/z 1224.23 (1224.53 calcd for M + H).

(2). Resin (**R1**) was treated with 80% TFA in dichloromethane and washed thoroughly. A solution of 3-methyl-2-thiophenecarboxylic acid (21 mg, 0.148 mmol) and HBTU (28 mg, 0.079 mmol) in DMF (0.45 mL) and DIEA (0.5 mL) was mixed at 40 °C for 25 min and poured onto the deprotected resin. The resin slurry was shaken for 4 h at room temperature and filtered. After washing with DMF, the resin was cleaved with Dp (1 mL) at 40 °C for 4 h. The crude product was purified by reversed-phase HPLC to afford (**1**) as a white solid upon lyophilization (3.0 mg, 8.2% recovery). MALDI-TOF-MS m/z 1238.35 (1238.54 calcd for M + H).

(2). Resin (**R1**) was treated with 80% TFA in dichloromethane and washed thoroughly. A solution of 3-methyl-2-thiophenecarboxylic acid (21 mg, 0.148 mmol) and HBTU (28 mg, 0.079 mmol) in DMF (0.45 mL) and DIEA (0.5 mL) was mixed at 40 °C for 25 min and poured onto the deprotected resin. The resin slurry was shaken for 4 h at room temperature and filtered. After washing with DMF, the resin was cleaved with Dp (1 mL) at 40 °C for 4 h. The crude product was purified by reversed-phase HPLC to afford (**2**) as a white solid upon lyophilization (3.0 mg, 8.2% recovery). MALDI-TOF-MS m/z 1238.35 (1238.54 calcd for M + H).

(3). Resin (**R1**) was treated with 80% TFA in dichloromethane and washed thoroughly. A solution of (**11**) (36 mg, 0.148 mmol) and HBTU (28 mg, 0.079 mmol) in DMF (0.45 mL) and DIEA (0.5 mL) was mixed at 40 °C for 25 min and poured onto the deprotected resin. The resin slurry was shaken for 4 h at room temperature and filtered. After washing with DMF and dichloromethane, the resin was treated with 80% TFA in dichloromethane. The resin was filtered and washed before cleavage with Dp (1 mL) at 40 °C for 4 h. The crude product was purified by reversed-phase HPLC to afford (**3**) as a slightly yellow solid upon lyophilization (3.4 mg, 9.4% recovery). MALDI-TOF-MS m/z 1239.46 (1239.54 calcd for M + H).

(4). Resin (**R1**) was treated with 80% TFA in dichloromethane and washed thoroughly. A solution of (**11**) (36 mg, 0.148 mmol) and HBTU (28 mg, 0.079 mmol) in DMF (0.45 mL) and DIEA (0.5 mL) was mixed at 40 °C for 25 min and poured onto the deprotected resin. The resin slurry was shaken for 4 h at room temperature and filtered. After washing with DMF and dichloromethane, the resin was treated with 80% TFA in dichloromethane. The resin was filtered, neutralized and shaken in a solution of acetic anhydride (0.2 mL), DIEA (0.2 mL) and DMF (1.6 mL) for 30 min. The resin was then filtered and washed with DMF before cleavage with Dp (1 mL) at 40 °C for 4 h. The crude product was purified by reversed-phase HPLC to afford (**4**) as a pale yellow solid upon lyophilization (4.2 mg, 11.2% recovery). MALDI-TOF-MS m/z 1281.62 (1281.55 calcd for M + H).

(5). A solution of sodium hydride (40 mg, 60% oil dispersion) and thiophenol (0.1 mL) in DMF (0.15 mL) was heated to 100 °C and a solution of (**6**) (1.3 mg, 1 μmol) in DMF (0.25 mL) was added. After 2 h, the reaction mixture was cooled to 0 °C and 20% TFA in water (7 mL) was added. The aqueous solution was washed three times with diethyl ether (8 mL) and was subjected to preparative, reversed-phase HPLC to afford (**5**) as a white solid upon lyophilization (0.6 mg, 50% recovery). MALDI-TOF-MS m/z 1241.09 (1240.52 calcd for M + H).

(6). Resin (**R1**) was treated with 80% TFA in dichloromethane and washed thoroughly. A solution of (**13**) (23 mg, 0.148 mmol) and HBTU (28 mg, 0.079 mmol) in DMF (0.45 mL) and DIEA (0.5 mL) was mixed at 40 °C for 25 min and poured onto the deprotected resin. The resin slurry was shaken for 4 h at room temperature and filtered. After washing with DMF and dichloromethane, the resin was cleaved with Dp (1 mL) at 40 °C for 4 h. The crude product was purified by reversed-phase HPLC to afford (**6**) as a white solid upon lyophilization (3.3 mg, 8.9% recovery). MALDI-TOF-MS m/z 1255.96 (1255.39 calcd for M + H).

(7). Resin (**R1**) was treated with 80% TFA in dichloromethane and washed thoroughly. A solution of (**14**) (22 mg, 0.148 mmol) and HBTU (28 mg, 0.079 mmol) in DMF (0.45 mL) and DIEA (0.5 mL) was mixed at 40 °C for 25 min and poured onto the deprotected resin. The resin slurry was shaken for 4 h at room temperature and filtered. After washing with DMF and dichloromethane, the resin was cleaved with Dp (1 mL) at 40 °C for 4 h. The crude product was purified by reversed-phase HPLC to afford (**7**) as a white solid upon lyophilization (2.6 mg, 7.0% recovery). MALDI-TOF-MS m/z 1242.20 (1242.52 calcd for M + H).

(8). Resin (**R1**) was treated with 80% TFA in dichloromethane and washed thoroughly. A solution of 3-chloro-2-thiophenecarboxylic acid (24 mg, 0.148 mmol)

and HBTU (28 mg, 0.079 mmol) in DMF (0.45 mL) and DIEA (0.5 mL) was mixed at 40 °C for 25 min and poured onto the deprotected resin. The resin slurry was shaken for 4 h at room temperature and filtered. After washing with DMF and dichloromethane, the resin was cleaved with Dp (1 mL) at 40 °C for 4 h. The crude product was purified by reversed-phase HPLC to afford (**8**) as a white solid upon lyophilization (3.8 mg, 10.1% recovery). MALDI-TOF-MS *m*/*z* 1258.86 (1258.49 calcd for M + H).

(9). Resin (**R1**) was treated with 80% TFA in dichloromethane and washed thoroughly. A solution of 2-trichloroacetyl-1-methylimidazole (34 mg, 0.148 mmol) in DMF (0.45 mL) and DIEA (0.5 mL) was poured onto the deprotected resin. The resin slurry was shaken for 4 h at 40 °C and filtered. After washing with DMF and dichloromethane, the resin was cleaved with Dp (1 mL) at 40 °C for 4 h. The crude product was purified by reversed-phase HPLC to afford (**9**) as a yellow solid upon lyophilization (2.5 mg, 6.9% recovery). MALDI-TOF-MS *m*/*z* 1222.03 (1222.58 calcd for M + H).

(10). Resin (**R1**) was treated with 80% TFA in dichloromethane and washed thoroughly. A solution of *N*-methylpyrrole-2-carboxylic acid (19 mg, 0.148 mmol) and HBTU (28 mg, 0.079 mmol) in DMF (0.45 mL) and DIEA (0.5 mL) was mixed at 40 °C for 25 min and poured onto the deprotected resin. The resin slurry was shaken for 4 h at room temperature and filtered. After washing with DMF and dichloromethane, the resin was cleaved with Dp (1 mL) at 40 °C for 4 h. The crude product was purified by reversed-phase HPLC to afford (**10**) as a white solid upon lyophilization (2.7 mg, 7.5% recovery). MALDI-TOF-MS *m*/*z* 1222.12 (1221.58 calcd for M + H).

DNA reagents and materials

Oligonucleotide primers SF1 (5′-AATTC-GAGCTCGGTACCGGGG-3′) and SF2 (5′-CTGGCACGACAGGTTTCCCGA-3′) were synthesized by the Biopolymer Synthesis Center at the California Institute of Technology. Products from PCR amplification of the pCW15 using 5′-[γ-^{32}P]-labeled SF1 and SF2 were purified on a 7% non-denaturing polyacrylamide gel. Glycogen (20 mg/mL), dNTPs (PCR nucleotide mix), and all enzymes, unless otherwise stated, were purchased from Boehringer-Mannheim. Deoxyadenosine [γ-^{32}P]triphosphate was obtained from ICN. Calf thymus DNA (sonicated, deproteinized) and DNase I (7500 units/mL, FPLC pure) were from Amersham Pharmacia. AmpliTaq DNA polymerase was obtained from Perkin-Elmer and was used with provided buffers. Tris·HCl, DTT, RNase-free water, and 0.5 M EDTA were from United States Biochemical. Calcium chloride, potassium chloride, and magnesium chloride were purchased from Fluka. Tris-borate-EDTA was from GIBCO and bromophenol blue was from Acros. All reagents were used without further purification.

DNase I footprinting experiments were performed according to standard protocols.[16]

Acknowledgements

We are grateful to the National Institutes of Health for research support, the National Science Foundation for a predoctoral fellowship to S.F., the Parson's Foundation for a fellowship to M.A.M., and Caltech for a James Irvine Fellowship to R.M.D.

References and Notes

1. Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783.
2. Dervan, P. B.; Burli, R. W. *Curr. Opin. Chem. Biol.* **1999**, *3*, 688.
3. Dervan, P. B. *Bioorg. Med. Chem.* **2001**, *9*, 2215.
4. Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 3663.
5. Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1996**, *382*, 559.
6. White, S.; Baird, E. E.; Dervan, P. B. *Chem. Biol.* **1997**, *4*, 569.
7. Kielkopf, C. L.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Nat. Struct. Biol.* **1998**, *5*, 104.
8. White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Nature* **1998**, *391*, 468.
9. Kielkopf, C. L.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Science* **1998**, *282*, 111.
10. Ellervik, U.; Wang, C. C. C.; Dervan, P. B. *J. Am. Chem. Soc.* **2000**, *122*, 9354.
11. Marques, M. A.; Doss, R. M.; Urbach, A. R.; Dervan, P. B. *Helv. Chim. Acta* **2002**, *85*, 4485.
12. Huddleston, P. R.; Barker, J. M. *Synth. Commun.* **1979**, *9*, 731.
13. Taylor, E. C.; Ping, Z. *Org. Prep. Proc. Int.* **1997**, *29*, 221.
14. Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6141.
15. SPARTAN ESSENTIAL Copyright® 1991–2001 by Wavefunction Inc.
16. Trauger, J. W.; Dervan, P. B. *Methods Enzymol.* **2001**, *340*, 450.