Toward an Understanding of the Chemical Etiology for DNA Minor-Groove Recognition by Polyamides

by **Michael A. Marques**, **Raymond M. Doss**, **Adam R. Urbach**, and **Peter B. Dervan***

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

Dedicated to Professor *Dieter Seebach* on the occasion of his 65th birthday

Crescent-shaped polyamides composed of aromatic amino acids, *i.e.*, 1-methyl-1*H*-imidazole **Im**, 1-methyl-1*H*-pyrrole **Py**, and 3-hydroxy-1*H*-pyrrole **Hp**, bind in the minor groove of DNA as 2:1 and 1:1 ligand/DNA complexes. DNA-Sequence specificity can be attributed to shape-selective recognition and the unique corners or pairs of corners presented by each heterocycle(s) to the edges of the base pairs on the floor of the minor groove. Here we examine the relationship between heterocycle structure and DNA-sequence specificity for a family of five-membered aromatic amino acids. By means of quantitative DNase-I footprinting, the recognition behavior of polyamides containing eight different aromatic amino acids, *i.e.*, 1-methyl-1*H*-pyrazole **Pz**, 1*H*-pyrrole **Nh**, 5-methylthiazole **Nt**, 4-methylthiazole **Th**, 3-methylthiophene **Tn**, thiophene **Tp**, 3-hydroxythiophene **Ht**, and furan **Fr**, were compared with the polyamides containing the parent-ring amino acids **Py**, **Im**, and **Hp** for their ability to discriminate between the four *Watson–Crick* base pairs in the DNA minor groove. Analysis of the data and molecular modeling showed that the geometry inherent to each heterocycle plays a significant role in the ability of polyamides to differentiate between DNA sequences. Binding appears sensitive to changes in curvature complementarity between the polyamide and DNA. The **Tn**/**Py** pair affords a modest 3-fold discrimination of T·A *vs.* A·T and suggests that an S-atom in the thiophene ring prefers to lie opposite T not A.

1. Introduction. – Many diseases are related to aberrant gene expression, and the ability to reprogram transcription in a cell by small molecules could be important in biology and human medicine. Minor-groove-binding polyamides, which bind predetermined DNA sequences offer a chemical approach to artificial gene regulation. These molecules are based on analogues of the 1-methyl-1*H*-pyrrole ring of amino acid residue **Py** of the natural products netropsin and distamycin A, which have been shown to bind in the minor groove of DNA in 1:1 and 2:1 ligand/DNA stoichiometries [1–4] (*Fig. 1*). **Py** is specific for A·T and T·A base pairs due to steric exclusion of $H_2N-C(2)$ of guanine (G-NH2) [1–3]. Base-pair specificity can be altered by changing the functional group(s) presented to the floor of the DNA minor groove. Stabilizing and destabilizing interactions with the different edges of the four *Watson–Crick* base pairs are modulated by specific H-bonds and, importantly, steric fit or shape complementarity. For example, the 1-methyl-1*H*-imidazole residue **Im** presents the DNA with the N-atom and its lone pair sp^2 orbital, which can accept a H-bond from G-NH2 [1–6]. Additionally, 3-hydroxy-1*H*-pyrrole residue **Hp** presents an OH group that is sterically accommodated opposite T not A and, in addition, can donate H-bonds to the O=C(2) of thymine [7] [8]. For discrimination of each of the *Watson–Crick* base pairs, the 2:1 stoichiometry involving unsymmetrical antiparallel cofacial pairs appears to be the best solution such that **Im**/**Py** is specific for G·C, **Py**/**Im** for C·G, **Hp**/**Py** for T·A, and **Py**/**Hp** for A·T.

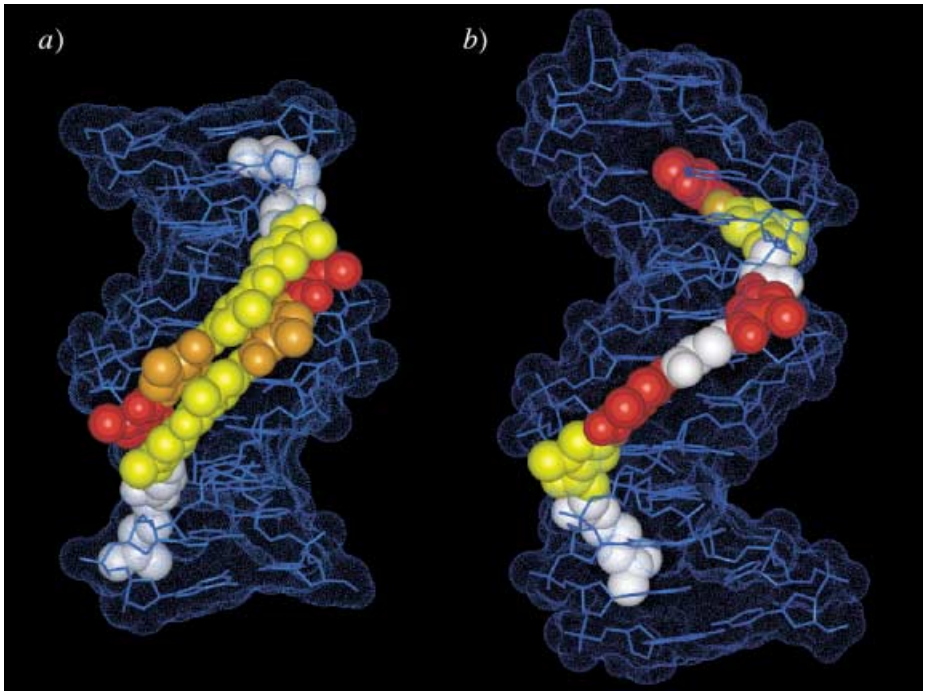


Fig. 1. *Structures of polyamides bound to DNA:* a) *2:1 motif determined by X-ray crystallography* [8]; b) *1:1 motif determined by NMR* [19]. DNA is shown as a stick model in blue. Polyamides are shown as space-filling models, with imidazole residues in red, hydroxypyrrole in orange, pyrrole in yellow, and aliphatic residues in white.

The pairing rules have proven useful for the recognition of hundreds of DNA sequences by polyamides. However, sequence-dependent DNA structural variation (such as minor-groove width) makes binding affinity and specificity at many DNA sequences unpredictable, which leads us to continue our search for new aromatic amino acid residues of slightly different shape (curvature and twist) for minor-groove recognition. Importantly, we find that the **Hp** residue can degrade over time in the presence of acid or free radicals, and a robust replacement for the **Hp**/**Py** pair suitable for use in biological studies is desirable. Several five-membered heterocyclic residues other than **Py**, **Im**, and **Hp** have been investigated previously, including furan **Fr** [9], thiazole **Nt** [10][11], and 1*H*-pyrazole **Pz** [11][12], with no new specificity uncovered. This raises the issue whether there is something 'special' about the 1-methyl-1*H*-pyrrole analogs **Py**, **Im**, and **Hp** for minor-groove recognition. We attempt here to broaden the repertoire of aromatic five-membered heterocycles for DNA recognition by synthesizing and characterizing a family of five-membered heterocyclic carboxamides grouped by the type of functionality directed toward the floor of the DNA minor groove (*Fig. 2*). Analogs of **Py** would be 1-methyl-1*H*-pyrazole **Pz** and 1*H*-pyrrole **Nh**, which project a H-atom toward the floor of the minor groove. Analogs of **Im** would be 5-methylthiazole **Nt** and furan **Fr**, which project lone pairs of electrons from the N- or O-atom. An analog of **Hp** is the 3-hydroxythiophene **Ht**, which projects an OH group. The 4-methylthiazole **Th**, 3-methylthiophene **Tn**, and thiophene **Tp** project a large

heteroatom, the S-atom to the floor of DNA and likely represent a new class for shape-selective recognition. We chose to maintain a *five-membered heterocyclic framework* in our library to retain overall the crescent shape of the polyamide ligand and to observe the effects of small structural changes resulting from single atom substitution on DNA base-pair specificity. One anticipates that substitution of atoms in the five-membered ring projecting away from the DNA minor groove (*i.e.*, the non-reading frame) will have effects on bond lengths and bond angles of each heterocycle as well and allows us to ask how tolerant is DNA to subtle alterations in curvature and twist of the minor-groove-binding ligand.

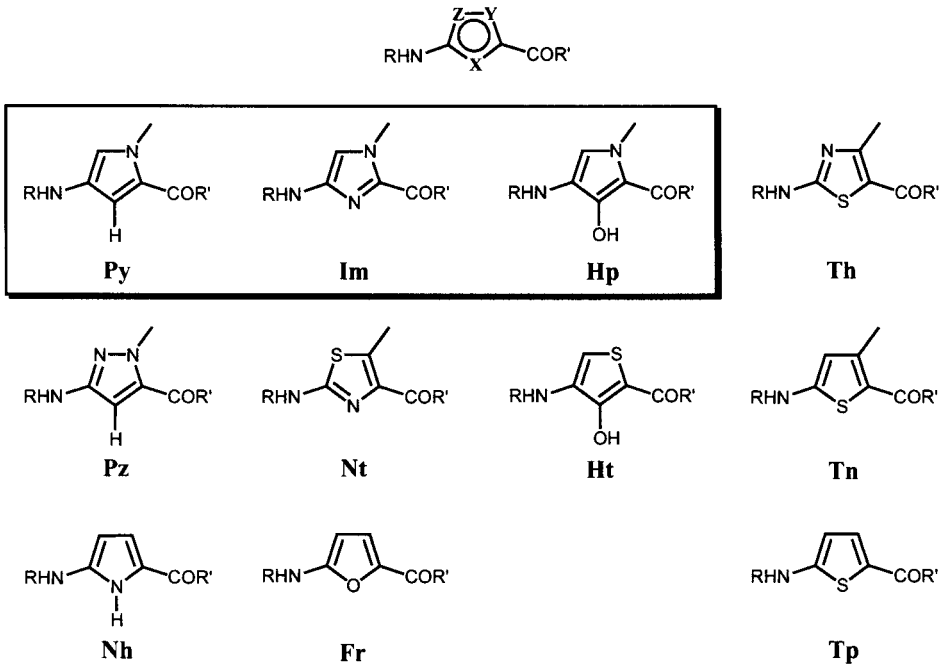


Fig. 2. *Family of five-membered heterocyclic amino acids studied here.* Centered above is a formula showing the five-membered heterocyclic framework with the variable positions labeled X, Y, and Z. The parent **Im**, **Py**, and **Hp** residues are boxed. All residues are shown with the functionality that faces the DNA minor groove pointed down (X).

The covalent head-to-tail linkage of polyamide subunits in a 2 : 1 complex results in a hairpin oligomer with increased DNA affinity and sequence specificity [13][14]. The hairpin motif avoids the ambiguity of slipping between the side-to-side stacked subunits [15], and 'locks' individual ring pairings in a predictable cofacial manner. In a formal sense, the hairpin oligomer provides a predictable foldamer for studying the DNA recognition characteristics of new ring pairings. We have reported previously the sequence specificities of **Py**/**Py**, **Hp**/**Py**, **Pz**/**Py**, and **Th**/**Py** pairings at a single position within the hairpin-polyamide-sequence context Im-Im-**X**-Py-γ-Im-**Py**-Py-Py-β-Dp (**X** = **Py**, **Hp**, **Pz**, and **Th**; $\gamma = \gamma$-aminobutanoic acid; $\beta = \beta$-alanine; Dp = *N*,*N*-dimethylpropane-1,3-diamine) opposite the four *Watson–Crick* base pairs within the

sequence context 5′-ATGG**X**CA-3′ (**X** = A, T, G, and C) [11] [17]. We found that **Pz**/**Py** can mimic **Py**/**Py** and, as a surprising result, **Th**/**Py** bound all four base pairs with low affinity and no specificity. Undaunted, we broaden our data set here to include **Nh**, **Fr**, **Ht**, **Tp**, and **Tn** residues each paired with **Py** in the same hairpin context for comparison (*Fig. 3*). For the sake of completeness and to provide a comparative analysis of heterocycle behavior opposite the four *Watson–Crick* base pairs in *both stoichiometries* (*Fig. 3*), we analyze polyamides also in a 1:1 polyamide/DNA complex (1:1 motif) of type Im-β-ImPy-β-**X**-β-Im-Py-β-Dp (**X** = **Py**, **Hp**, **Nh**, **Ht**, **Fr**, **Nt**, **Tn**, and **Th**), which

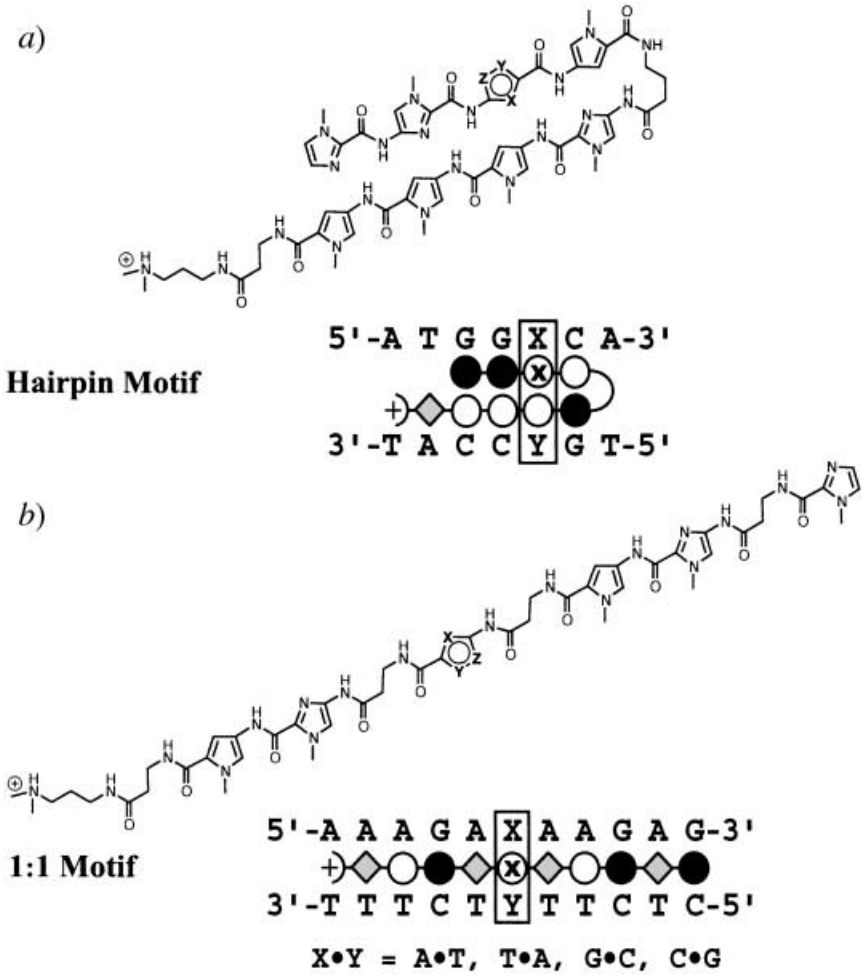


Fig. 3. *Schematic illustrating the examination of sequence selectivity against the four* Watson–Crick *base pairs within* a) *hairpin motifs and* b) *1:1 motifs.* Each chemical structure has a variable residue containing X, Y, and Z-labeled positions, which are designated in *Fig. 2*. The dot models shown below each chemical structure illustrate the binding mode with the polyamide shown inside its target DNA sequence: ● = imidazole residue; ○ = pyrrole residue; ◆ = β-alanine residue; semicircle ⊃ = γ-aminobutanoic acid turn residue connecting the two subunits; ⊗ = novel heterocycle residue.

bind DNA sequences 5′-AAAGA**X**AGAAG-3′ (**X** = A, T, G, and C) in a single orientation [16–19].

Quantitative DNase-I footprinting was used to determine the equilibrium association constant for each complex. *Ab initio* computational modeling of the heterocyclic amino acids was implemented to derive their inherent geometric and electronic parameters to guide interpretation of the experimental outcome. The combination of experiment and modeling provides insight to the origin of DNA-sequence discrimination by polyamides.

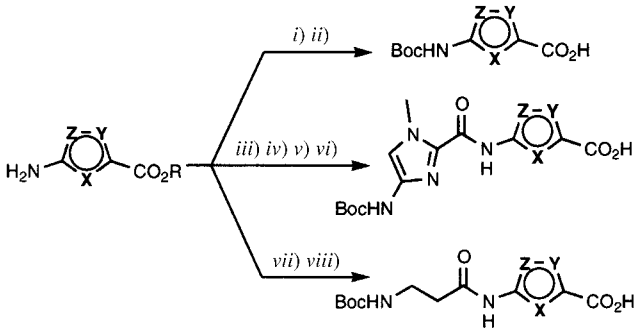
2. Monomer, 'Dimer', and Polyamide Synthesis. – Polyamides were synthesized manually on solid support by the stepwise addition of monomeric and 'dimeric' Boc-protected amino acids, **1**–**17** (*Schemes 1–7*). Hydroxythiophene(**Ht**)-containing polyamides were prepared by deprotecting the 3-methoxy analogue **Mt**. Boc-protected amino acids for **Im**, **Hp**, **Py**, **Pz**, and **Th** monomers were synthesized according to previously reported procedures [11][17][20]. Syntheses of core amino-ring-carboxylic acid alkyl ester (NH_2-X-OR) structures **1**–**4** for X = **Fr**, **Nh**, **Tn**, and **Mt** are shown in *Schemes 2–4* and *6*. Boc-protected monomeric amino acids **5**–**12** suitable for olid-phase synthesis were prepared in two steps from their NH_2-X-OR analogues (*Schemes 1–7*). However, aminofuran and aminothiophene precursors were unreactive to coupling on solid support, and, therefore, 'dimers' **13**–**17** were preformed in solution under strong acylation conditions prior to solid-phase coupling. Synthetic schemes for all Boc-protected amino acids used in this study follows (*Schemes 2–7*).

Furan Ring (**Fr**) *Derivatives.* The nitro-furan ester **18** was prepared from methyl furan-2-carboxylate [21]. The amino ester **1** was then synthesized from **18** by treatment with H_2 (500 psi) and Pd/C (*Scheme 2*) and isolated as the free base from a AcOEt solution by precipitation with hexanes. The free base is a stable crystalline solid at room temperature. The (nitro-imidazolyl)-furan ester **19** (NO_2-Im-Fr-OMe) was prepared by condensing the 1-methyl-4-nitro-1*H*-imidazol-2-yl trichloromethyl ketone (NO_2-Im-$COCCl_3$) with **1** in AcOEt at 35°. The 'dimer' product **19** is rather insoluble in AcOEt and began to precipitate upon formation. Reduction of **19** with H_2 (500 psi) and Pd/C, followed by addition of 2M HCl in Et_2O, gave the hydrochloride salt **20** ($HCl \cdot H_2N$-Im-Fr-OMe). Salt **20** was Boc-protected (Boc = (*tert*-butoxy)carbonyl) with $(Boc)_2O$, *N*,*N*-diisopropylethylamine (DIEA), and DMF at 60° for 12–18 h to give **21** (Boc-Im-Fr-OMe; see **15**, with CO_2Me instead of CO_2H). Elevated temperatures and extended reaction times were necessary, presumably due to the poor nucleophilicity of the NH_2 group at the imidazole ring. Saponification of **21** in 1N aqueous NaOH and MeOH at room temperature provided the target 'dimer' **15** (Boc-Im-Fr-OH).

Alternatively, 'dimer' **22** (Boc-*β*-Fr-OMe; see **13**, with CO_2Me instead of CO_2H) was synthesized from **1** by coupling to the symmetrical anhydride of Boc-*β*-alanine in DMF, DIEA, and *N*,*N*-dimethylpyridin-4-amine (DMAP) (*Scheme 2*). The anhydride was pre-formed in minutes with dicyclohexylcarbodiimide (DCC) in CH_2Cl_2 at room temperature. The amino-furan ester **1** was then added as a solution in DMF and DIEA, followed by the transacylation catalyst DMAP. The 5-amino group at the furan ring is significantly unreactive [22], and attempts to couple it on solid support with reagents such as DCC and HOBt (1-hydroxy-1*H*-benzotriazol), HBTU (2-(1*H*-benzotriazol-1-

Scheme 1. a) *Formation of Monomer and 'Dimer' Boc-protected Amino Acids.* b) *Monomeric and 'Dimeric' Units*

a)



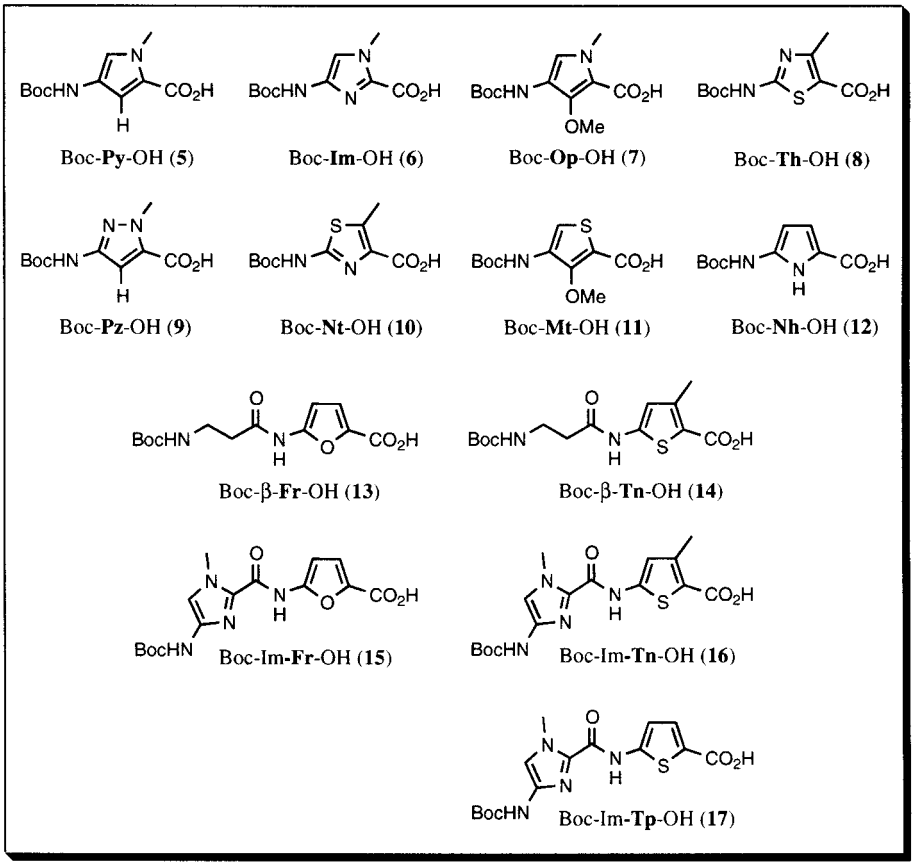
i) For Py (X = C–H, Y = N–Me, Z = C–H, R = Me), Im (X = N, Y = N–Me, Z = C–H, R = Et), Op (X = C–OMe, Y = N–Me, Z = C – H, R = Et), Th (X = S, Y = C–Me, Z = N, R = Et), Pz (X = C–H, Y = N–Me, Z = C–H, R = Et), Nt (X = N, Y = C–Me, Z = S, R = H), and Nh (X = N–H, Y = Z = C–H, R = Et), $(Boc)_2O$, *N,N*-diisopropylethylamine (DIEA), DMF, 60°, 12–18 h; for Mt (X = C–OMe, Y = S, Z = C–H, R = Me), $(Boc)_2O$, DIEA, Et_3N, CH_2Cl_2, 60°, 12–18 h. *ii*) For Py (X = C–H, Y = N–Me, Z = C–H, R = Me), Im (X = N, Y = N–Me, Z = C–H, R = Et), Op (X = C–OMe, Y = N–Me, Z = C–H, R = Et), Th (X = S, Y = C–Me, Z = N, R = OEt), Pz (X = C–H, Y = N–Me, Z = C–H, R = Et), and Nh (X = N–H, Y = Z = C–H, R = Et), 1N NaOH, MeOH, r.t., 3–4 h; for Nt (X = N, Y = C–Me, Z = S, R = H), 1N NaOH, MeOH, r.t., 1 h; for Mt (X = C–OMe, Y = S, Z = C–H, R = Me), KOH, MeOH, 60°, 4–6 h. *iii*) For Fr (X = O, Y = Z = C–H, R = Me), Tn (X = S, Y = C–Me, Z = C–H, R = Me), and Tp (X = S, Y = Z = C–H, R = Me), O_2N-Im-$COCCl_3$, AcOEt, DIEA, 35°, 10–12 h. *iv*) For Fr (X = O, Y = Z = C–H, R = Me), Tn (X = S, Y = C–Me, Z = C–H, R = Me), and Tp (X = S, Y = Z = C–H, R = Me), H_2, 10% Pd/C, AcOEt, r.t., 1.5 h. *v*) For Fr (X = O, Y = Z = C–H, R = Me), Tn (X = S, Y = C–Me, Z = C–H, R = Me), and Tp (X = S, Y = Z = C–H, R = Me), $(Boc)_2O$, DIEA, DMF, 60°, 12–18 h. *vi*) For Fr (X = O, Y = Z = C–H, R = Me), 1N NaOH, MeOH, r.t., 3 h; for Tn (X = S, Y = C–Me, Z = C–H, R = Me) and Tp (X = S, Y = Z = C–H, R = Me), 1N NaOH, MeOH, 60°, 4–6 h. *vii*) For Fr (X = O, Y = Z = C–H, R = Me) and Tn (X = S, Y = C–Me, Z = C–H, R = Me), $(Boc\text{-}\beta\text{-}Ala)_2O$, DMF, DIEA, 40°, 12–18 h. *viii*) For Fr (X = O, Y = Z = C–H, R = Me), 1N NaOH, MeOH, r.t., 3 h; for Tn (X = S, Y = C–Me, Z = C–H, R = Me), 1N NaOH, MeOH, 60°, 4–6 h.

yl)-1,1,3,3-tetramethyluronium hexafluorophosphate), HATU, PyBrOP, PyBOP, and TFFH, were unsuccessful. Formation of the Boc-β-Fr-OMe 'dimer' **22** occurred slowly at elevated temperature. Heating above 40° did not affect the rate of reaction. Several different solvent systems were tried, but DMF was found to be optimal. Saponification of **22** was accomplished with a mixture of MeOH and 1N NaOH at room temperature to provide the target **13** (Boc-β-Fr-OH), which is suitable for standard solid-phase protocols [20].

1-H-*Pyrrole Ring* (**Nh**) *Derivatives.* Pyrrolyl trichloromethyl ketone **23** was prepared by adding 1*H*-pyrrole to a mixture of trichloroacetyl chloride in Et_2O at 0°, then warming to room temperature and stirring overnight, followed by precipitation from hexanes. Nitration of **23** was accomplished with Ac_2O and nitric acid at −40° to provide the 5-nitro regioisomer **24** as the major product (*Scheme 3*). Regiocontrol of the nitration appears to depend significantly on the reaction temperature. Higher temperatures provided unfavorable mixtures of 4-nitro and 5-nitro regioisomers. Treatment of trichloromethyl ketone **24** with EtONa/EtOH at room temperature gave the nitro-pyrrole ester **25** (NO_2-Nh-OEt) in good yield. Reduction of **25** with H_2 (500 psi) and Pd/C, followed by the addition of 2M HCl in Et_2O, provided the

Scheme 1 (cont.)

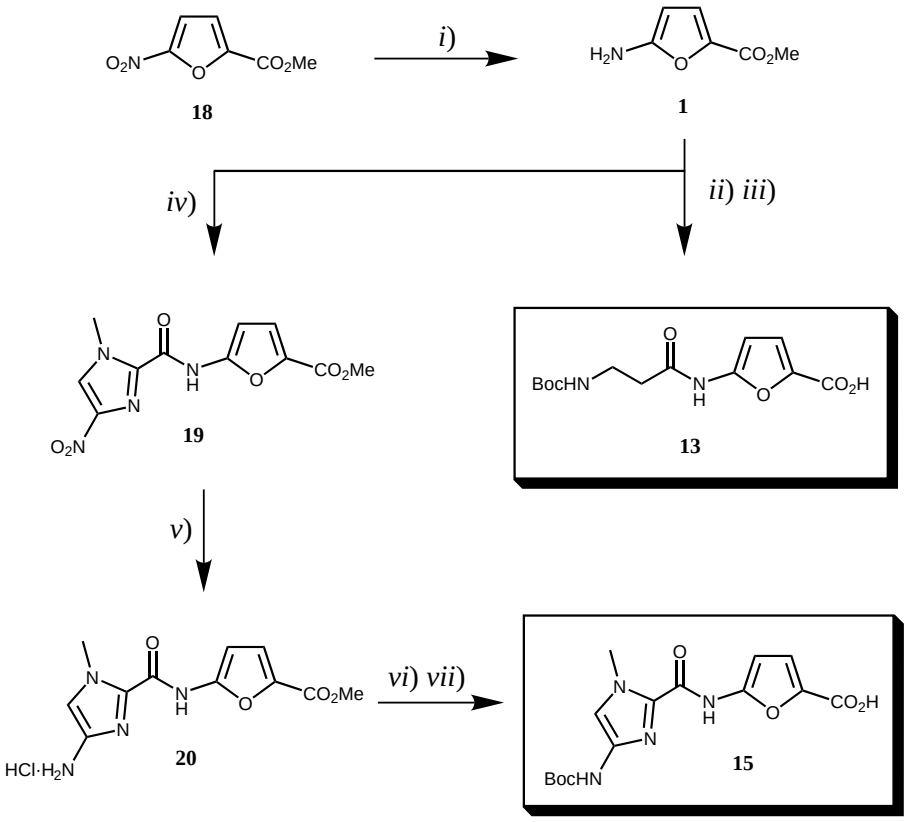
b)



hydrochloride salt **2** ($HCl \cdot H_2N$-Nh-OEt), which was then Boc-protected ($(Boc)_2O$, DIEA, and DMF) to yield **26** (Boc-Nh-OEt) which was saponified (1N NaOH, MeOH, room temperature) to the final monomer unit **12** (Boc-Nh-OH).

3-Methylthiophene Ring (**Tn**) *Derivatives.* The acyclic precursor to the thiophene-ring system was prepared by a *Knoevenagel* reaction involving acetoacetate and cyanoacetic acid, providing **27** as a mixture of (*E*)- and (*Z*)-regioisomers in moderate yield after vacuum distillation. Treatment of **27** with sulfur flakes and Et_2NH in EtOH yielded the cyclized aminothiophene, and addition of conc. HCl solution precipitated the hydrochloride salt **3** ($HCl \cdot H_2N$-Tn-OMe) (*Scheme 4*). Formation of the (nitro-imidazolyl)-thiophene ester **28** (NO_2-Im-Tn-OMe), followed by reduction and Boc-protection to provide **29** ($HCl \cdot H_2N$-Im-Tn-OMe) and **30** (Boc-Im-Tn-OMe; see **16**, with CO_2Me instead of CO_2H), respectively, was accomplished by the procedures described above for the furan compounds **19**, **20**, and **21**. The Boc-β-Tn-OMe 'dimer' **31** (see **14**; with CO_2Me instead of CO_2H) was prepared as described above for **22**, by coupling **3** with the symmetric anhydride of Boc-β-alanine. The NH_2 group at the thiophene ring displays low reactivity comparable to furan. Elevated temperature was

Scheme 2

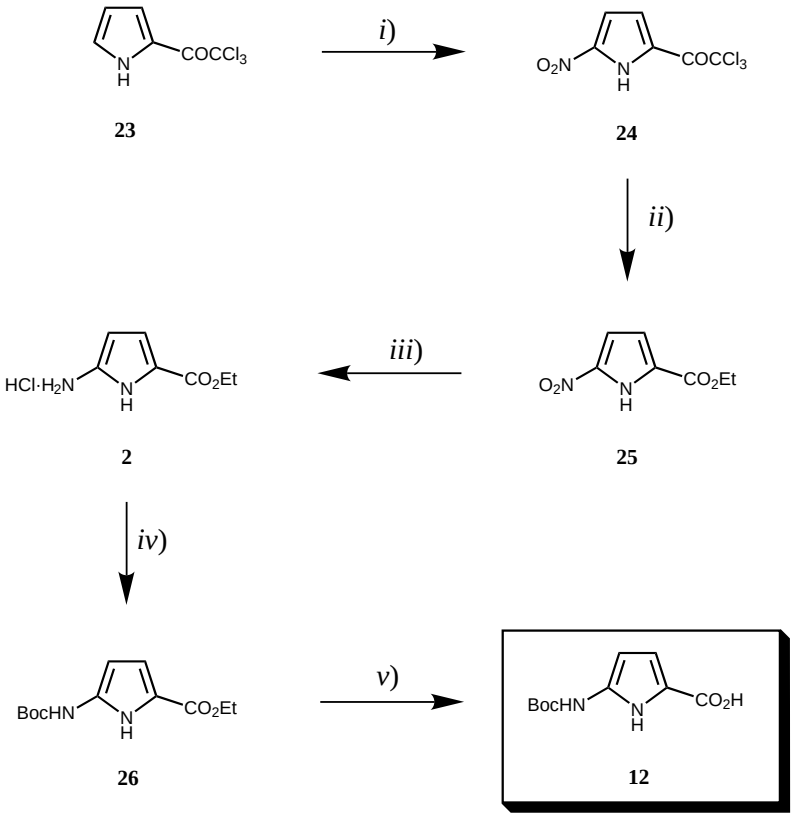


i) H_2, Pd/C, AcOEt, 500 psi, r.t. *ii*) (Boc-β-Ala)$_2$O, DMF, DIEA, DMAP, 40°. *iii*) 1N NaOH, r.t. *iv*) NO_2-Im-$COCCl_3$, AcOEt, 35°. *v*) H_2, Pd/C, AcOEt, 500 psi, r.t. *vi*) $(Boc)_2O$, DIEA, DMF, 60°. *vii*) 1N NaOH, MeOH, r.t.

necessary to completely saponify methyl esters **30** and **31**. Saponification was carried out in 1N NaOH and MeOH at 60° for 4–6 h to give the 'dimers' **16** (Boc-Im-Tn-OH) and **14** (Boc-β-Tn-OH), respectively.

Thiophene Ring (**Tp**) *Derivatives.* Treatment of commercially available 5-nitro-thiophene-2-carboxaldehyde in acetone with a mixture of NaOCl and Na_2HPO_4 in H_2O, gave the nitro-thiophene acid **32** (NO_2-Tp-OH). Esterification of **32** by refluxing for 48 h in a mixture of H_2SO_4 and MeOH provided the nitro ester **33** (NO_2-Tp-OMe) (*Scheme 5*). Reduction of **33** with a mixture of tin(II) chloride dihydrate and HCl in EtOH gave the hydrochloride salt **34** ($HCl \cdot H_2N$-Tp-OMe). Formation of the 'dimer' **35** (NO_2-Im-Tp-OMe), followed by reduction to **36** ($HCl \cdot H_2N$-Im-Tp-OMe), Boc-protection to **37** (Boc-Im-Tp-OMe; see **17**, with CO_2Me instead of CO_2H, and saponification to the target 'dimer' **17** (Boc-Im-Tp-OH) proceeded as described above for furan compounds **19**, **20**, **21**, and **15**.

Scheme 3

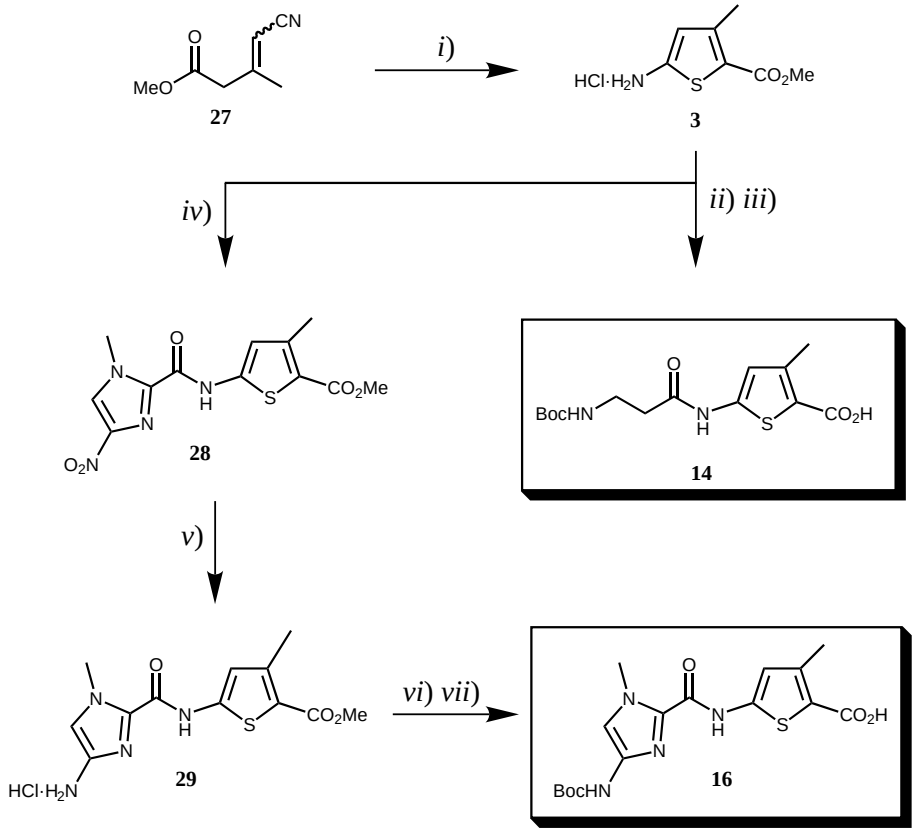


i) HNO_3, Ac_2O, $-40°$. *ii*) EtONa, EtOH, reflux. *iii*) H_2, Pd/C, DMF, HCl, Et_2O, r.t. *iv*) $(Boc)_2O$, DIEA, DMAP, DMF, r.t. *v*) 1N NaOH, MeOH, r.t.

3-Methoxythiophene Ring (**Mt**) *Derivatives.* The hydroxythiophene methyl ester **38** (Ht-OMe) was synthesized by a cyclization reaction between methyl thioglycolate and methyl-2-chloroacrylate in MeONa/MeOH [23]. Ester **38** was nitrated with a mixture of conc. H_2SO_4 and HNO_3 at $-10°$ to give 4-nitro-3-hydroxythiophene ester **39** (NO_2-Ht-OMe) as the major regioisomer after column chromatography (*Scheme 6*). Treatment of **39** with CH_2N_2 in Et_2O afforded the methyl ether **40** (NO_2-Mt-OMe; see **39**, with OMe instead of OH) in near quantitative yield (Ht = 3-hydroxythiophene residue, Mt = 3-methoxythiophene residue). Reduction of **40** with a mixture of tin(II) chloride dihydrate, HCl, and EtOH gave the hydrochloride salt **4** ($HCl \cdot H_2N$-Mt-OMe). Boc-protection was accomplished by heating a mixture of **4**, $(Boc)_2O$, Et_3N, and CH_2Cl_2 at 60° for 12 h $\rightarrow$ **41** (Boc-Mt-OMe; see **11**, with CO_2Me instead of CO_2H). Saponification of **41** was achieved with methanolic KOH and heating at 50° for 6 h to give the final monomer **11** (Boc-Mt-OH).

5-Methylthiazole Ring (**Nt**) *Derivative.* The Boc-protected 2-amino-5-methylthiazole acid **10** (Boc-Nt-OH) was synthesized on a multi-gram scale by brominating 2-

Scheme 4

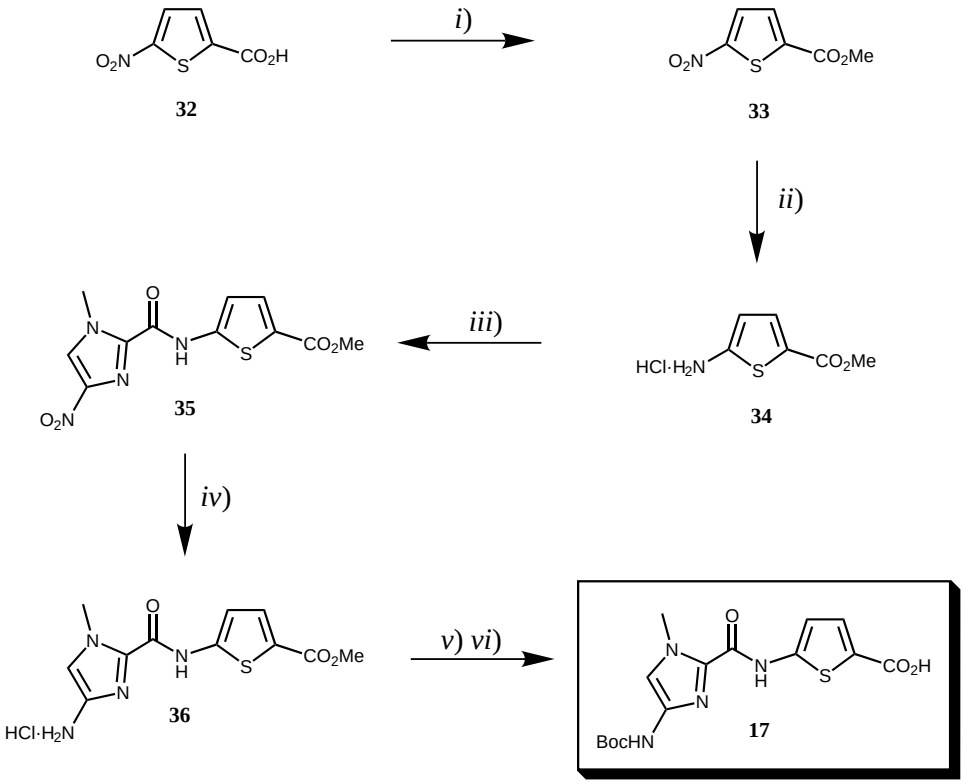


i) S, Et_2NH, EtOH, r.t. *ii*) (Boc-β-Ala)$_2$O, DMF, DIEA, DMAP, 40°, *iii*) 1N NaOH, MeOH, 60°. *iv*) NO_2-Im-$COCCl_3$, AcOEt, 35°. *v*) H_2, Pd/C, AcOEt, 500 psi, r.t. *vi*) $(Boc)_2O$, DIEA, DMF, 60°. *vii*) 1N NaOH, MeOH, 60°.

oxobutanoic acid, followed by condensation with thiourea and Boc-protection of the amine (*Scheme 7*). For best results, the Br_2 should be added dropwise over at least 2 h, as the reaction is autocatalytic and highly exothermic. Also, the thiourea should be added in small portions with vigorous stirring. Boc-protection was accomplished by dissolving the crude material in DMF and DIEA, followed by the addition of $(Boc)_2O$ and stirring at 60° for 12 h. The material was then stirred in a solution of MeOH and 1N NaOH for ester saponification to provide the target compound **10** (Boc-Nt-OH).

Polyamides. Hairpin and 1:1 motif polyamides were synthesized manually from Boc-β-Pam resin in a stepwise fashion with Boc-protected monomeric and 'dimeric' amino acids (*Scheme 8*) according to solid-phase protocols [20]. Polyamides containing 3-methoxythiophene (Mt) were deprotected by treatment with sodium benzenethiolate in DMF (100°, 2 h) to provide the **Ht** analogues after HPLC purification.

Scheme 5

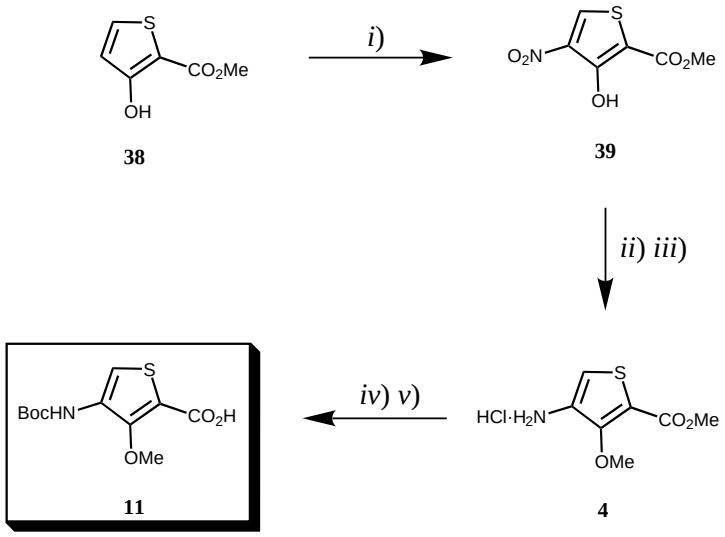


i) H_2SO_4, MeOH, reflux. *ii*) $SnCl_2 \cdot 2\,H_2O$, 95% EtOH, 37% HCl soln., 35°. *iii*) NO_2-Im-$COCCl_3$, AcOEt, 35°. *iv*) H_2, Pd/C, AcOEt, 500 psi, r.t. *v*) $(Boc)_2O$, DIEA, DMF, 60°. *vi*) 1N NaOH, MeOH, 60°.

3. DNA Affinity and Sequence Specificity in the Hairpin Motif. – Quantitative DNase-I-footprinting titrations [24] were carried out for the following polyamides on the 278-base-pair PCR product of plasmid pDHN1 [11]: Im-Im-**Im**-Py-γ-Im-**Py**-Py-Py-β-Dp (**Im**/**Py** pair; **42**), Im-Im-**Nh**-Py-γ-Im-**Py**-Py-Py-β-Dp (**Nh**/**Py** pair; **43**), Im-Im-**Tn**-Py-γ-Im-**Py**-Py-Py-β-Dp (**Tn**/**Py** pair; **44**), Im-Im-**Tp**-Py-γ-Im-**Py**-Py-Py-β-Dp (**Tp**/**Py** pair; **45**), Im-Im-**Ht**-Py-γ-Im-**Py**-Py-Py-β-Dp (**Ht**/**Py** pair; **46**), and Im-Im-**Fr**-Py-γ-Im-**Py**-Py-Py-β-Dp (**Fr**/**Py** pair; **47**). The DNA-sequence specificity of each polyamide at a single ring-pairing position (bolded in the sequences listed above) was determined by varying a single DNA base pair within the parent-sequence context, 5′-TGG**X**CA-3′, to all four *Watson–Crick* base pairs (**X** = A, T, G, C) and comparing the relative affinities of the resulting complexes (*Fig. 4*). The variable base-pair position was installed opposite the novel heterocycle/pyrrole pairing in question, according to previously reported specificity studies on eight-ring hairpin polyamides [7][11].

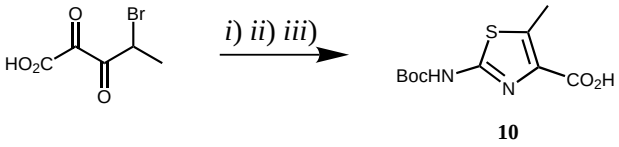
Equilibrium association constants (K_a) for eight-ring polyamides containing **Hp**/**Py**, **Py**/**Py**, **Pz**/**Py**, and **Th**/**Py** pairings against the four DNA sites used in this study have been reported [11][17] and are included in *Table 1* for comparison with values

Scheme 6



i) HNO_3, H_2SO_4, 0°. *ii*) CH_2N_2, r.t. *iii*) $SnCl_2 \cdot 2\,H_2O$, HCl, EtOH, 40°. *iv*) $(Boc)_2O$, CF_3COOH, CH_2Cl_2, 60°. *v*) 1N NaOH, MeOH, 50°.

Scheme 7



i) CSN_2H_4, neat, r.t. *ii*) $(Boc)_2O$, DIEA, DMF, DMAP, 60°. *iii*) 1N NaOH, MeOH, 35°.

presented here. As expected, polyamide **42** (**Im**/**Py** pair) exhibited single-site specificity (5′-TGG**X**CA-3′, **X** = G) at modest affinity ($K_a = 4.5 \cdot 10^8$ M^{-1}) with fourteen-fold preference over **X** = C and at least 400-fold preference over **X** = A, T. Polyamide **43** (**Nh**/**Py** pair) bound with high affinity to the **X** = A, T sites ($K_a \approx 10^{10}$ M^{-1}) in preference to **X** = G, C by about ten-fold. Hairpin **44** (**Tn**/**Py** pair) bound with high affinity to **X** = T, A ($K_a \approx 10^9$ M^{-1}), with a 3-fold preference for T · A > A · T, and 800-fold preference over the **X** = G, C sites. The thiophene analogue, **Tp**, lacking the methyl group was examined to probe possible effects on DNA binding caused by the 3-methyl group. The **Tp**/**Py** pair (**45**) was found to display similar recognition properties as the **Tn**/**Py** pair (**44**). Remarkably, hairpins containing **Ht**/**Py** and **Fr**/**Py** pairs (**46** and **47**, resp.) demonstrated no binding to the designed sites at concentrations up to 1 μM.

4. DNA Affinity and Sequence Specificity in the 1:1 Motif. – Quantitative DNase-I-footprinting titrations were carried out for the following polyamides on the 298-base-

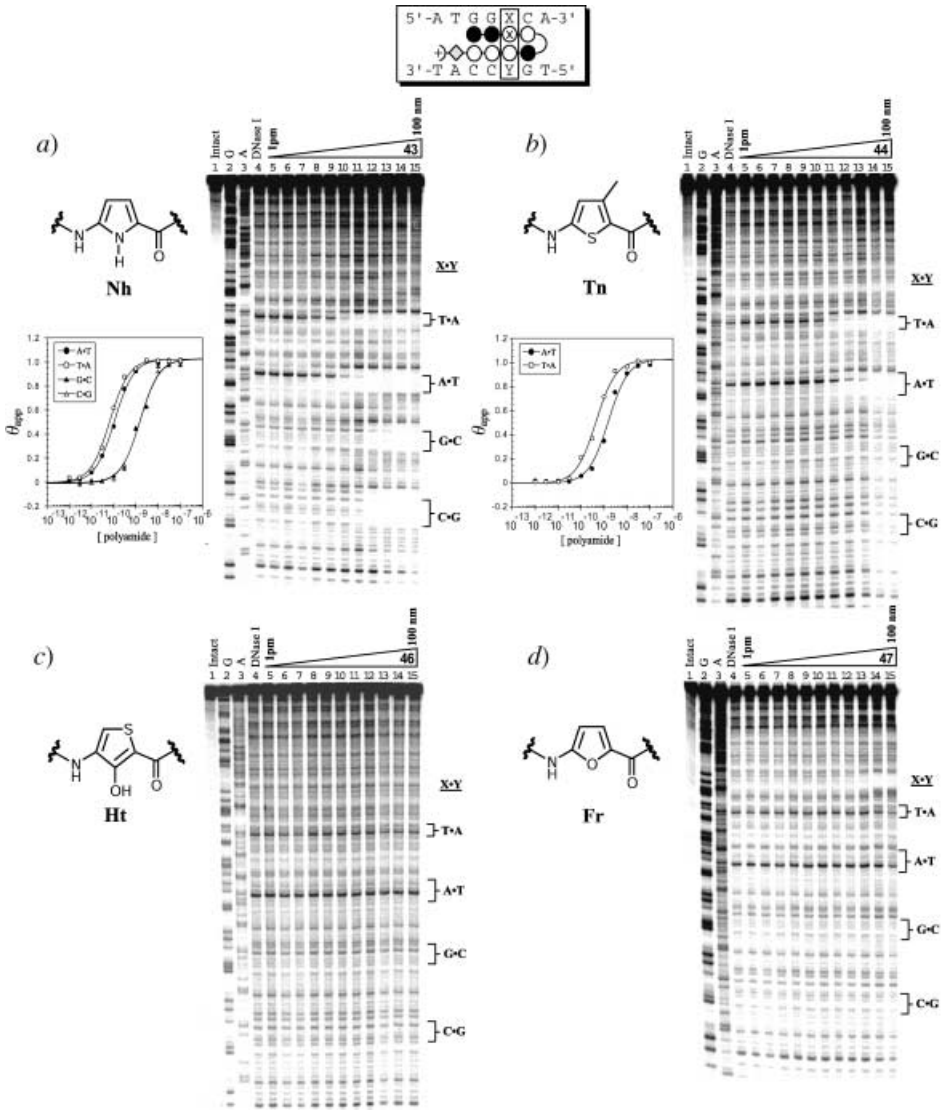


Fig. 4. *Quantitative DNase-I-footprinting experiments in the hairpin motif for* a*) polyamides* **43**, b*) polyamide* **44**, *c) polyamide* **46**, *and* d*) polyamide* **47** *on the 278-base-pair, 5′-end-labelled PCR product of plasmid DHN1. Lane 1*, intact DNA; *Lane 2*, G reaction; *Lane 3*, A 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 polyamide, resp. Each footprinting gel is accompanied by the following: chemical structure of the residue of interest (left, top) and, for *a*) and *b*), binding isotherms for the four designed sites (left, bottom). θ_{norm} values were obtained according to published methods [24][26]. A binding model for the haipin motif is shown centered at the top as a dot model with the polyamide bound to its target DNA sequence: ● = imidazole residue; ○ = pyrrole residue; ◆ = β-alanine residue; ⊃ = semicircle γ-aminobutanoic acid turn residue connecting the two subunits; ⊗ = novel heterocycle residue.

Scheme 8. *Solid-Phase Synthesis of Im-Im-***X***-Py-γ-Py-Py-Py-β-Dp* (arrows up from center) *and Im-β-ImPy-β-***X***-β-ImPy-β-Dp* (arrows down from center) *starting from commercially available Boc-β-Pam resin*

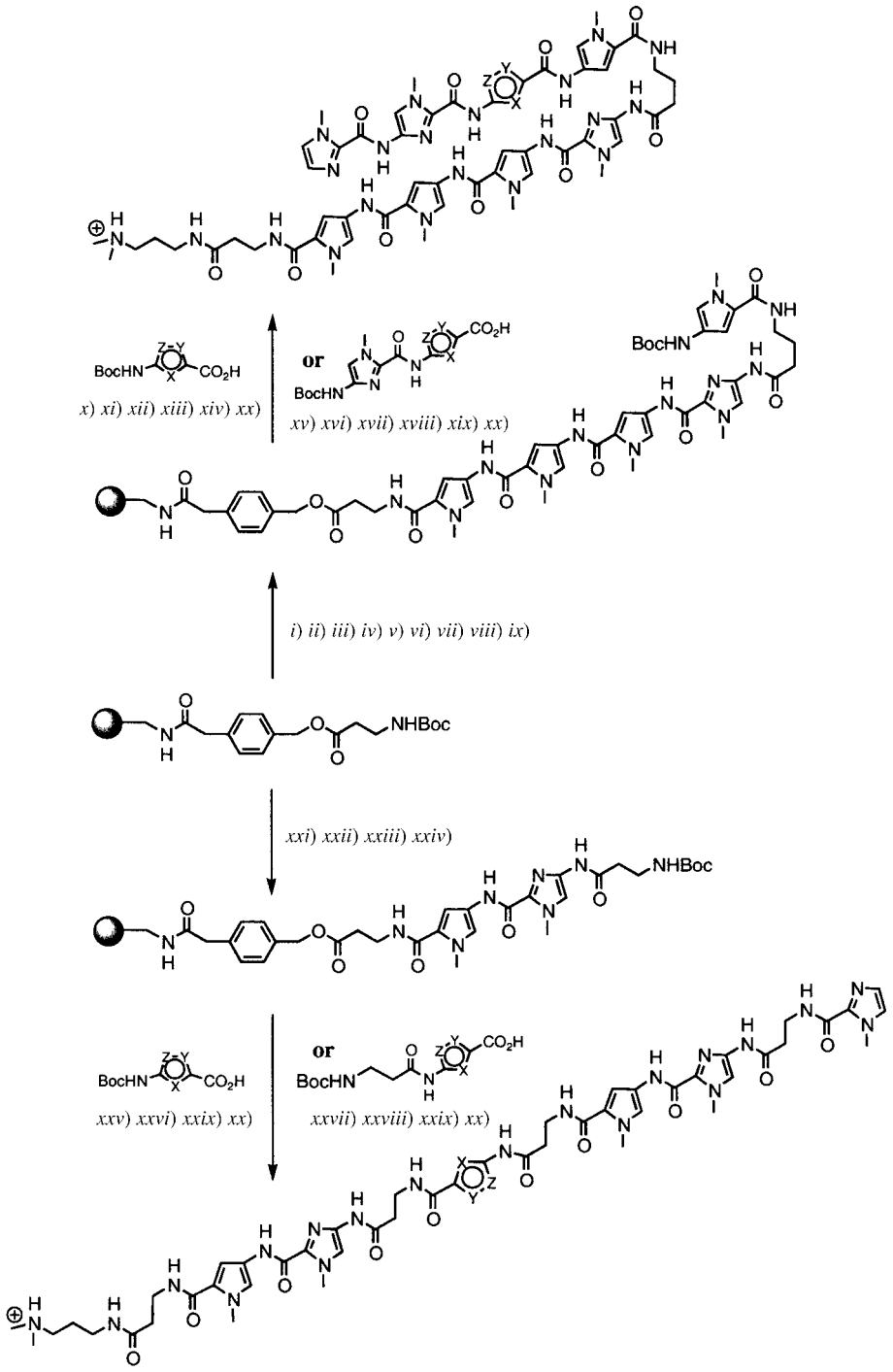


Table 1. *Hairpin Motif:* K_a [M^{-1}][a])[b])

Pair[c])	A·T	T·A	G·C	C·G
Im/Py (**42**)	$\leq 10^6$	$\leq 10^6$	$4.5 (\pm 0.7) \cdot 10^8$	$3.2 (\pm 0.5) \cdot 10^7$
Hp/Py	$8.1 (\pm 1.9) \cdot 10^7$	$1.6 (\pm 0.3) \cdot 10^9$	$5.5 (\pm 1.5) \cdot 10^7$	$7.9 (\pm 2.1) \cdot 10^7$
Py/Py	$3.1 (\pm 0.7) \cdot 10^9$	$4.7 (\pm 0.4) \cdot 10^9$	$2.2 (\pm 0.6) \cdot 10^8$	$2.5 (\pm 0.9) \cdot 10^8$
Nh/Py (**43**)	$8.5 (\pm 0.3) \cdot 10^9$	$1.1 (\pm 0.1) \cdot 10^{10}$	$9.2 (\pm 0.1) \cdot 10^8$	$8.2 (\pm 0.4) \cdot 10^8$
Pz/Py	$1.0 (\pm 0.5) \cdot 10^9$	$2.0 (\pm 0.3) \cdot 10^9$	$\leq 2 \cdot 10^7$	$\leq 2 \cdot 10^7$
Th/Py	$\leq 2 \cdot 10^7$	$\leq 2 \cdot 10^7$	$\leq 2 \cdot 10^7$	$\leq 2 \cdot 10^7$
Tn/Py (**44**)	$8.0 (\pm 0.4) \cdot 10^8$	$2.7 (\pm 0.2) \cdot 10^9$	$\leq 10^6$	$\leq 10^6$
Tp/Py (**45**)	$3.8 (\pm 0.5) \cdot 10^8$	$1.0 (\pm 0.3) \cdot 10^9$	$\leq 10^6$	$\leq 10^6$
Ht/Py (**46**)	$\leq 10^6$	$\leq 10^6$	$\leq 10^6$	$\leq 10^6$
Fr/Py (**47**)	$\leq 10^6$	$\leq 10^6$	$\leq 10^6$	$\leq 10^6$

[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° in a buffer of 10 mM *Tris*·HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$ at pH 7.0. [c]) The number in parentheses indicates the compound containing the unique pairing.

pair PCR product of pAU8 [30]: Im-β-Im-Py-β-**Py**-β-Im-Py-β-Dp (**48**), Im-β-Im-Py-β-**Hp**-β-Im-Py-β-Dp (**49**), Im-β-Im-Py-β-**Nh**-β-Im-Py-β-Dp (**50**), Im-β-Im-Py-β-**Ht**-β-Im-Py-β-Dp (**51**), Im-β-Im-Py-β-**Fr**-β-Im-Py-β-Dp (**52**), Im-β-Im-Py-β-**Nt**-β-Im-Py-β-Dp (**53**), Im-β-Im-Py-β-**Tn**-β-Im-Py-β-Dp (**54**), and Im-β-Im-Py-β-**Th**-β-Im-Py-β-Dp (**55**). The sequence specificity of each polyamide at a single carboxamide position (bolded in the sequences listed above) was determined by varying a single base pair within the parent DNA sequence context, 5′-AAAGA**X**AAGAG-3′, to all four *Watson–Crick* base pairs (**X** = A, T, G, C) and comparing the relative affinities of the resulting complexes (*Figs. 5* and *6*). The variable base-pair position was installed opposite the novel heterocycle in question, according to previously described specificity studies on 1:1 polyamide/DNA complexes [17].

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 reported. However, in that study, only the **Im** specificity experiment was performed at the 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 the study for a more controlled comparison. Polyamide **48** (**Py**) demonstrated very high affinity ($K_a \approx 6 \cdot 10^{10}$ M^{-1}) at the **X** = A, T sites (5′-AAAGA**X**AAGAG-3′) with a 5- to 10-fold

← *i*) 80% CF_3COOH/CH_2Cl_2, 0.4M PhSH. *ii*) Boc-Py-OBt, DIEA, DMF. *iii*) Repeat steps *i*) and *ii*) twice. *iv*) 80% CF_3COOH/CH_2Cl_2, 0.4M PhSH. *v*) Boc-Im-OH, HBTU, DIEA, DMF. *vi*) 80% CF_3COOH/CH_2Cl_2, 0.4M PhSH. *vii*) Boc-γ-OH, HBTU, DIEA, DMF. *viii*) 80% CF_3COOH/CH_2Cl_2, 0.4M PhSH. *ix*) Boc-Py-OBt, DIEA, DMF. *x*) 80% CF_3COOH/CH_2Cl_2, 0.4M PhSH. *xi*) Boc-**X**-OH, HBTU, DIEA, DMF. *xii*) 80% CF_3COOH/CH_2Cl_2, 0.4M PhSH. *xiii*) Boc-Im-OH, HBTU, DIEA, DMF. *xiv*) Im-$COCCl_3$, DIEA, DMF. *xv*) 80% CF_3COOH/CH_2Cl_2, 0.4M PhSH. *xvi*) Boc-**X**-OH, HBTU, DIEA, DMF. *xvii*) 80% CF_3COOH/CH_2Cl_2, 0.4M PhSH. *xviii*) Im-$COCCl_3$, DIEA, DMF. *xix*) Im-$COCCl_3$, DIEA, DMF. *xx*) *N,N*-Dimethylpropane-1,3-diamine, 85°. *xxi*) 80% CF_3COOH/CH_2Cl_2, 0.4M PhSH. *xxii*) Boc-Py-OBt, DIEA, DMF. *xxiii*) 80% CF_3COOH/CH_2Cl_2, 0.4M PhSH. *xxiv*) Boc-β-Im-OH [19], HBTU, DIEA, DMF. *xxv*) 80% CF_3COOH/CH_2Cl_2, 0.4M PhSH. *xxvi*) Boc-**X**-OH, HBTU, DIEA, DMF. *xxvii*) 80% CF_3COOH/CH_2Cl_2, 0.4M PhSH. *xxviii*) Boc-**X**-OH, HBTU, DIEA, DMF. *xxix*) Elongation and termination according to standard procedures [19].

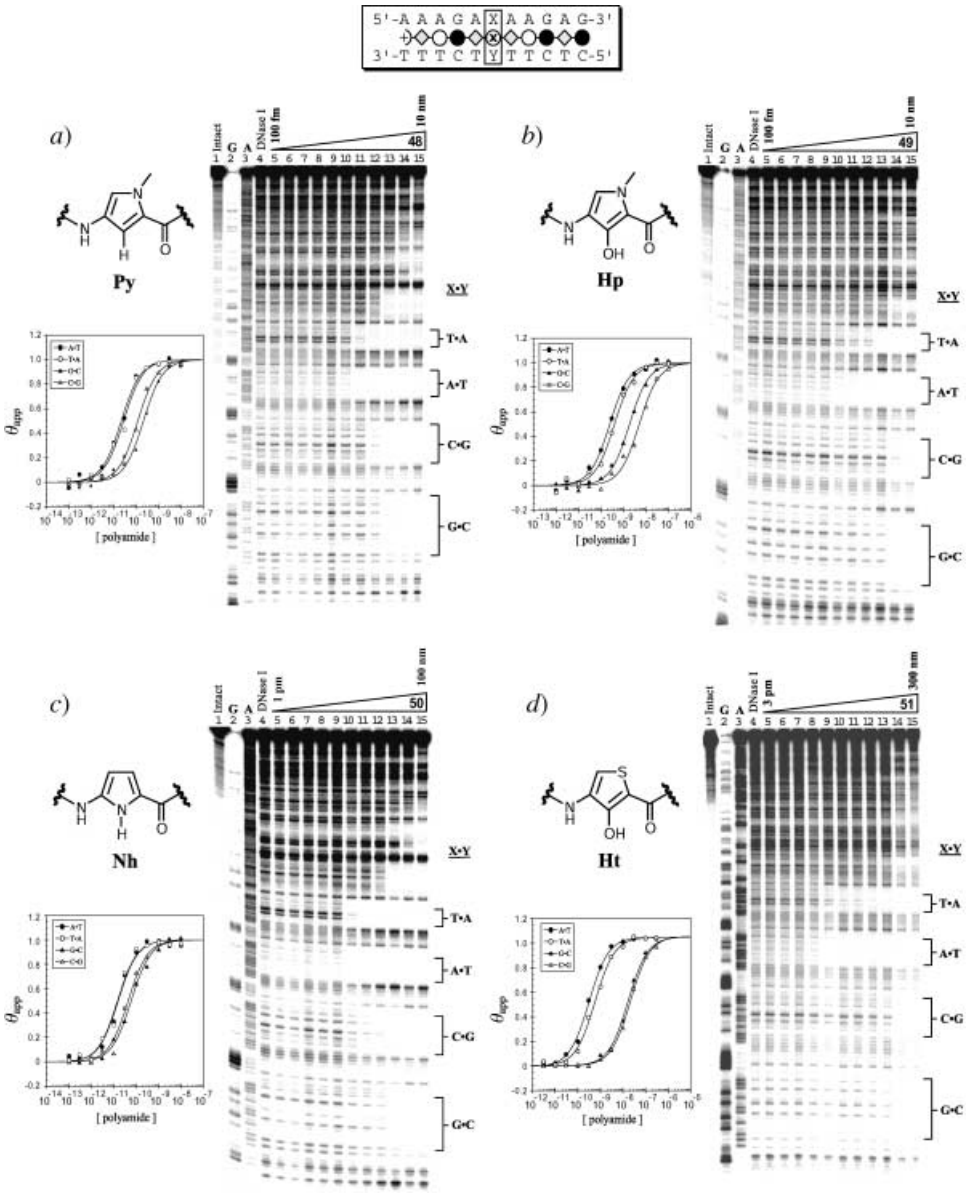


Fig. 5. *Quantitative DNase-I-footprinting experiments for* a) *polyamide* **48**, b) *polyamide* **49**, c) *polyamide* **50**, *and* d) *polyamide* **51** *on the 298-base-pair, 5′-end-labelled PCR product of plasmid pAU8. a*) and *b*): *Lane 1*, intact DNA; *Lane 2*, G reaction; *Lane 3*, A reaction; *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, and 10 nM polyamide, resp. *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, 30 nM, and 100 nM polyamide, resp. *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, and 300 nM polyamide, resp. Each footprinting gel is accompanied by the following: chemical structure of the residue of interest (left, top) and binding isotherm for the four designed sites (left, bottom). θ_{norm} values were obtained according to published methods [24]. A binding model for the 1:1 motif is shown centered at the top as a dot model with the polyamide bound to its target DNA sequence: ● = imidazole residue; ○ = pyrrole residue; ◆ = β-alanine residue; ⊗ = novel heterocycle residue.

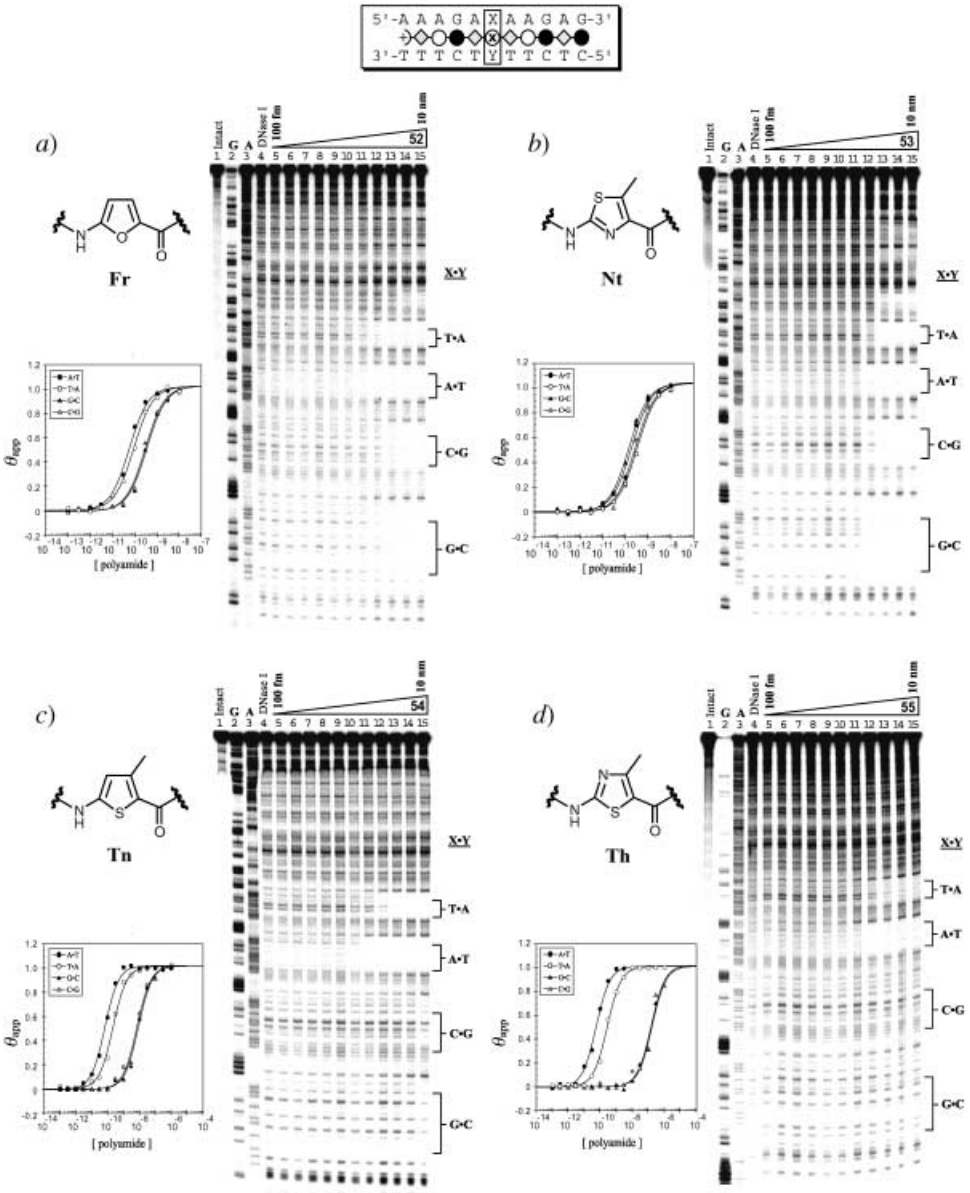


Fig. 6. *Quantitative DNase-I-footprinting experiments for* a) *polyamide* **52**, b) *polyamide* **53**, c) *polyamide* **54**, *and* d) *polyamide* **55**, *on the 298-base-pair, 5′-end-labelled PCR product of plasmid pAU8. Lane 1*, intact DNA; *Lane 2*, G reaction; *Lane 3*, A reaction; *Lane 4*, 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, and 10 nM polyamide resp. Each footprinting gel is accompanied by the following: chemical structure of the residue of interest (left, top) and binding isotherms for the four designed sites (left, bottom). Isotherms for *c*) and *d*) were generated from gels run out to a final concentration of 1 μM (not shown). θ_{norm} values were obtained according to published methods [24]. A binding model for the 1:1 motif is shown centered at the top as a dot model with the polyamide bound to its target DNA sequence: ● = imidazole residue; ○ = pyrrole residue; ◆ = β-alanine residue; ⊗ = novel heterocycle residue.

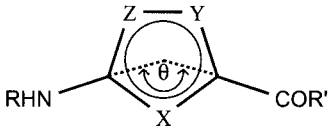
preference over **X** = G, C (*Table 2*). Polyamide **49** (**Hp**) bound with lower affinity ($K_a \approx 3 \cdot 10^9$ M^{-1}) but with similar specificity to **48**, preferring **X** = A, T > G, C by 5- to 10-fold. The **Nh**-containing polyamide **50** bound with very high affinity to the **X** = A, T sites ($K_a = 7.5 \cdot 10^{10}$ M^{-1}) but with a mere 3- to 5-fold selectivity over the high-affinity **X** = G, C sites. Polyamide **51** (**Ht**) bound with subnanomolar affinities to the **X** = A, T sites, similar to **49** but with ≥40-fold specificity for **X** = A, T > G, C. Polyamide **52** (**Fr**) showed high affinity for the **X** = A, T sites ($K_a \approx 10^{10}$ M^{-1}) with a small 2- to 4-fold preference over **X** = G, C. The 5-methylthiazole-containing polyamide **53** (**Nt**), which places the thiazole-ring N-atom into the floor of the minor groove, bound all four sites with similar high affinities ($K_a \approx 5 \cdot 10^9$ M^{-1}). Thiophene-containing polyamide **54** (**Tn**) showed specificity for A, T *vs.* G, C and a preference for a single A·T site. The 4-methylthiazole-containing polyamide **55** (**Th**), which places the thiazole-ring *S-atom* into the floor of the minor groove, bound with similar **X** = A, T affinity as **54** (**Tn**) but with >400-fold preference over **X** = G, C. In all cases, binding isotherms fit well to an $n = 1$ *Hill* equation, supporting a 1:1 polyamide/DNA stoichiometry (*Figs. 4–6*).

Table 2. *1:1 Motif:* K_a [M^{-1}][a])[b])

Ring[c])	A·T	T·A	G·C	C·G
Im	2.5 (± 0.2) · 10^{10}	1.1 (± 0.1) · 10^{10}	2.6 (± 0.4) · 10^{10}	1.3 (± 0.3) · 10^{10}
Py (**48**)	7.2 (± 0.3) · 10^{10}	5.3 (± 0.1) · 10^{10}	3.2 (± 0.4) · 10^9	9.4 (± 0.2) · 10^9
Hp (**49**)	3.9 (± 0.1) · 10^9	2.5 (± 0.3) · 10^9	5.3 (± 0.5) · 10^8	1.9 (± 0.5) · 10^8
Nh (**50**)	7.5 (± 0.2) · 10^{10}	7.4 (± 0.1) · 10^{10}	1.6 (± 0.2) · 10^{10}	2.3 (± 0.1) · 10^{10}
Ht (**51**)	2.8 (± 0.5) · 10^9	1.6 (± 0.6) · 10^9	3.8 (± 1.3) · 10^7	3.7 (± 0.7) · 10^7
Fr (**52**)	2.2 (± 0.5) · 10^{10}	1.0 (± 1.3) · 10^{10}	4.4 (± 0.5) · 10^9	5.0 (± 0.5) · 10^9
Nt (**53**)	5.4 (± 0.9) · 10^9	2.9 (± 0.6) · 10^9	8.0 (± 1.3) · 10^9	4.2 (± 0.6) · 10^9
Tn (**54**)	3.0 (± 0.2) · 10^{10}	5.7 (± 0.4) · 10^9	8.1 (± 0.4) · 10^7	8.3 (± 0.2) · 10^7
Th (**55**)	1.5 (± 0.2) · 10^{10}	3.0 (± 0.7) · 10^9	6.5 (± 0.5) · 10^6	7.4 (± 0.5) · 10^6

[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° in a buffer of 10 mM *Tris*·HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$ at pH 7.0. [c]) The number in parentheses indicates the compound containing the unique pairing.

5. Molecular-Modeling Calculations. – Modeling calculations were preformed with the 'Spartan Essential' software package [25]. Each ring was first minimized by means of an AM1 model, followed by *ab initio* calculations by means of the *Hartree–Fock* model and a 6-31G* polarization basis set. Each heterocycle exhibited a unique geometric and electronic profile (*Fig. 7*). Bonding geometries for the imidazole, pyrrole, and 3-hydroxypyrrole residues were in excellent agreement with coordinates derived from X-ray structures of polyamides containing these heterocycles [7][9]. The overall curvature of each monomer was calculated to be the sweep angle (θ) 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** > **Pz** > **Tp** > **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** > **Pz** > **Py** > **Tn** = **Tp** > **Th** > **Fr** >



Ring	X	Y	Z	θ[°]	Charge on X [e]
Fr	O	C-H	C-H	126	–0.31
Nt	N	C-Me	S	127	–0.60
Ht	C-OH	S	C-H	133	+0.40
Nh	N-H	C-H	C-H	136	+0.34
Im	N	N-Me	C-H	137	–0.71
Py	C-H	N-Me	C-H	146	+0.21
Hp	C-OH	N-Me	C-H	148	+0.50
Tn	S	C-Me	C-H	149	–0.21
Pz	C-H	N-Me	N	151	+0.23
Tp	S	C-H	C-H	152	–0.21
Th	S	C-Me	N	153	–0.25

Fig. 7. *Geoetric and electrostatic profiles for eleven heterocyclic amino acids, derived from* ab initio *molecular-modeling calculations with 'Spartan Essential'* software [25]. Left: 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 (see *Fig. 2*). Right: functional groups at X, Y, and Z, along with the angle θ and the electrostatic partial charge at X, for **Ht**, **Nh**, **Py**, **Hp**, and **Pz**, the positive charge at X is listed for the H-atom.

Im. Four-ring subunits containing the sequence Im-Im-**X**-Py (**X** = **Py**, **Pz**, **Nh**, **Im**, **Fr**, **Hp**, **Ht**, **Th**, **Tn**, and **Tp**) were constructed and subjected to AM1 and *ab initio* calculations as described above to examine overall subunit curvature and planarity (*Figs. 8* and *9*).

6. Discussion. – Here we explore the effects of varying single atom positions in five-membered aromatic heterocycles on the ability of polyamides to discriminate the four *Watson–Crick* base pairs in the minor groove of DNA. In this experimental design, the incremental scheme of DNA and polyamide sequence allows for the comparison of binding affinities for a {4} × {10} array of complexes containing unique combinations of {*Watson–Crick* base pair} × {five-membered aromatic heterocycle} at a single position (*Tables 1* and *2*). This quantitative analysis combined with computational modeling of the different heterocycles has led to insight into the etiology of DNA-sequence discrimination by polyamides.

The success of **Im**/**Py** and **Hp**/**Py** pairs at discriminating between the four *Watson–Crick* base pairs has been attributed to the shape of the functional groups directed toward the floor of the minor groove. Therefore, the heterocycles discussed here will be divided into groups based on the types of 'bumps and holes' presented to the complementary bumps and holes created by each *Watson–Crick* base pair on the floor of the DNA minor groove. For the sake of clarity and brevity, hairpin polymaides will be referred to in bold as their unique amino acid pair, *e.g.*, **Tn**/**Py** for polyamide Im-Im-**Tn**-Py-γ-Im-**Py**-Py-Py-β-Dp (**44**). The 1:1 polyamides will be referred to in bold as their unique heterocycle, *e.g.*, **Tn** for polyamide Im-β-Im-Py-β-**Tn**-β-Im-Py-β-Dp (**54**). DNA sequences will be identified as the variable base position within each motif, *e.g.*, **X** = G in the hairpin motif for 5′-TGG**X**CA-3′.

The Pairing Rules. Discrimination of the four *Watson–Crick* base pairs by means of *unsymmetrical* cofacial pairs of aromatic amino acids has proven to be a key insight for

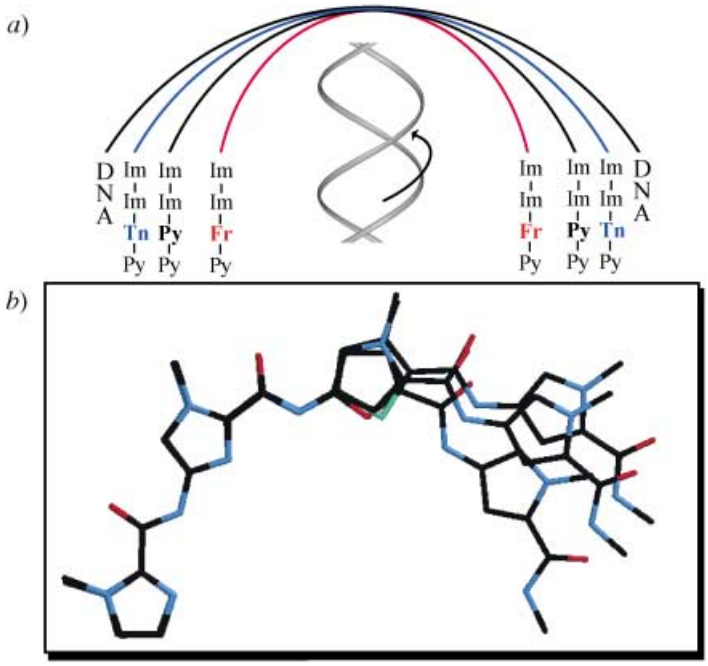


Fig. 8. a) *Schematic illustrating the curvatures of four-ring polyamide subunits containing* **Tn**, **Py**, *and* **Fr** *heterocycles with respect to one another and the DNA helix*; b) ab initio *models of polyamide subunits* (Im-Im-**X**-Py, **X** = **Tn**, **Py**, and **Fr**) *superimposed to demonstrate the significant difference in curvature resulting from atomic substitution* (H-atoms are not shown)

minor-groove DNA recognition. The four-ring contiguous subunits described here (*e.g.* Im-Py-Py-Py) are π-conjugated, which limits their conformational flexibility. Consequently, small changes in individual ring curvature can cause greater effects on overall oligomer curvature.

Py/**Py**, **Pz**, **Py**, and **Nh**/**Py** present an H-atom with positive potential to the minor-groove floor. As reported previously, **Py**/**Py** exhibits preference for the **X** = A, T binding sites with > 10-fold selectivity for **X** = A, T > G, C. **Pz**/**Py** behaves with similar affinity as **Py**/**Py**, but greater discrimination for A, T *vs.* G, C. **Nh**/**Py** is similar to **Py**/**Py** and binds all sites with higher affinity than **Py**/**Py** and **Pz**/**Py**. The A · T/T · A preference for the pairs **Py**/**Py**, **Pz**/**Py**, and **Nh**/**Py** originate from a steric interaction with G-NH2, yet there are subtle differences among these compounds, which undoubtedly derive from their slightly different shape. **Py** is known to be over-curved with respect to the DNA helix [6][8][27–29]. The calculations described here provide an amide-ring-amide intersection angle (θ in *Fig. 7*) of 146° and a partial positive charge at the H–C(3) of +0.21. By contrast, **Pz** is somewhat less curved than **Py** ($\theta = 151°$) with similar charge, and **Nh** is considerably more curved ($\theta = 136°$) with a greater charge of +0.34. The reduced curvature of **Pz**/**Py** should make it more complementary to the DNA curvature, and therefore, the steric H–C(4) to G-NH2 clash would be exacerbated, resulting in lower affinity for the **X** = G, C sites. **Nh**/**Py** is more curved

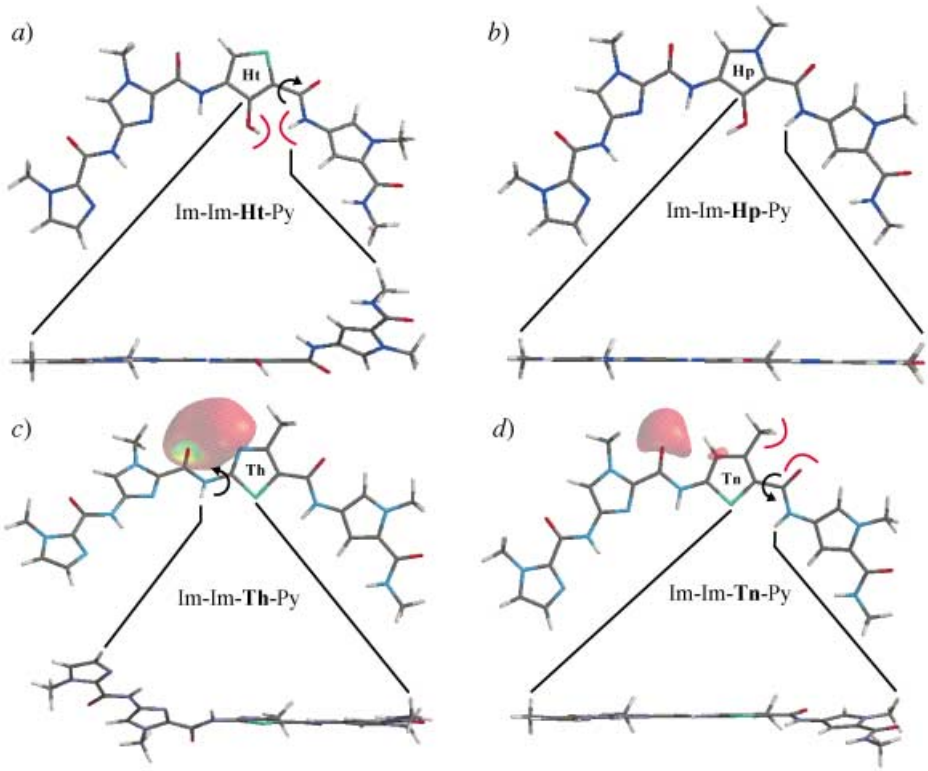


Fig. 9. Ab initio *models of four-ring polyamide subunits* (Im-Im-**X**-Py, **X** = **Ht**, **Hp**, **Th**, and **Tn**): a) *Im-Im-***Ht***-Py subunit demonstrating a dihedral* (shown as a curved arrow) *created at the ring–carboxamide juncture due to destabilizing eclipsing interaction* (shown as red arcs) *between HO–C(3) and the proximal carboxamide proton*; b) *Im-Im-***Hp***-Py subunit demonstrating coplanarity of the contiguous ring system;* c) *Im-Im-***Th***-Py subunit showing the negative isopotential surface* (lone-pair density) *in red and the dihedral* (shown as a curved arrow) *resulting from lone-pair repulsion interaction between the thiazole N-atom and the proximal carbonyl group*; d) *Im-Im-***Tn***-Py subunit showing the negative isopotential surface* (lone-pair density) *in red and the small dihedral* (shown as a curved arrow) *resulting from a destabilizing eclipsing interaction* (shown as red arcs) *between the thiophene Me group and the proximal carboxamide.* Atomic substitution of N to C–H for **Th** to **Tn** removes lone-pair repulsion interaction between the ring N-atom and the proximal carbonyl group.

and more charged, thereby reducing negative steric effects and increasing binding affinity.

Im/**Py** and **Fr**/**Py** present an N- or O-atom with sp^2 lone-pair electrons directed toward the minor-groove floor. Their DNA-recognition behavior is strikingly different. **Im**/**Py** preferentially targets **X** = G. On the other hand, **Fr**/**Py** shows a complete loss of DNA-binding affinity. Based on the established principles for G·C recognition by an **Im**/**Py** pair, it was not unreasonable to expect that the **Fr**/**Py** pair could be a positive recognition element for G·C as well, with the **Fr** O-atom acting as a H-bond acceptor to G-NH2. However, calculations for **Fr** reveal tight over-curvature, with the amide-ring-amide angle decreased by more than ten degrees with respect to **Im** and more than twenty degrees with respect to **Py**. This property causes a pronounced effect on the

entire Im-Im-**Fr**-Py subunit (*Fig. 8*), such that complex formation is no longer energetically favorable. This view is further reinforced by results in the more flexible 1:1 motif, wherein the **Fr**-containing polyamide binds all designed sites with high affinity.

Hp/**Py** and **Ht**/**Py** present an OH group (HO–C(3)) to the minor-groove floor. **Hp**/**Py** displays a preference for **X** = T that breaks the A·T/T·A degeneracy of **Py**/**Py** pairs. The **Hp**/**Py** pair presents OH opposite T not A. Structural studies reveal the origin of this specificity to reside in the H-bond formed between HO–C(3) of **Hp** and O=C(2) of thymine (T-O2), and to shape recognition of the asymmetric cleft in the T·A base pair [8][29]. However, a loss in affinity is typically observed for polyamides containing **Hp**/**Py** pairs compared to the **Py**/**Py** pair revealing that there is an energetic penalty for the gain in selectivity [30]. This loss in affinity 'trade-off' may be attributed to unfavorable steric interactions between the HO–C(3) and the minor-groove floor or perhaps differential solvation between **Hp**/**Py** and **Py**/**Py** in H_2O. The **Ht**/**Py** pair was designed with the hope to improve upon **Hp**/**Py** by increasing the ring curvature, thus reducing the steric interaction while maintaining the HO–C(3) to T-O2 H-bond. Remarkably, DNase-I footprinting for **Ht**/**Py** *reveals a complete loss in binding affinity.* As seen with **Fr**/**Py**, increased ligand curvature may be responsible for disrupting hairpin binding. *Ab initio* calculations on the Im-Im-**Ht**-Py subunit reveal an unfavorable eclipsing interaction between HO–C(3) and the proximal carboxamide proton. As shown in *Fig. 9*, this steric clash may force rotation about the ring–amide bond, which would twist the subunit out of plane. Given the snug fit of stacked hairpin subunits within the DNA minor groove, a large distortion in ligand planarity may not be tolerated. **Ht** in the flexible 1:1 motif binds its target sites with high affinity, which further underscores the pronounced effects of ring geometry within the conformationally constrained hairpin motif.

Th/**Py**, **Tn**/**Py**, and **Tp**/**Py** pairs present a large S-atom with an sp^2 lone pair to the minor-groove floor. The thiazole analog **Th**/**Py** has been shown previously to afford poor affinity overall and no discrimination of the *Watson–Crick* base pairs [11]. It was thought that the thiazole S-atom was too large to be easily accommodated within the closely packed cofacial rings in the hairpin·DNA complex. However, *ab initio* calculations on thiazolecarboxamide reveal an unfavorable interaction between the lone pair of the thiazole N(3) and the proximal carbonyl O-atom (*Fig. 9*). Although this interaction exists for the pyrazole N(2), the effect on thiazole is much greater due to the large S-atom forcing the thiazole N-atom into closer proximity with the carboxamide. Consequently, the polyamide subunit may twist out of plane about the amide–ring bond to alleviate electronic strain, as shown in *Fig. 9*. As with **Ht**/**Py**, the diminished DNA binding affinity of **Th**/**Py** may be due to its non-planar conformation. These effects are not observed for **Th** in the 1:1 motif.

The negative electronic interaction and hence non-planar conformation should be alleviated if the N-atom on the back corner of thiazole were replaced with C–H, as in the case of thiophene (**Tn** and **Tp**). Remarkably, **Tn**/**Py** binds to the **X** = A and T sites with high affinity, but with no observable binding to **X** = G, C. There is a modest 3-fold preference for T > A. Therefore, the S-atom is *not* too large to be accommodated within the tightly packed hairpin·DNA complex. It appears that the large S-atom of thiophene prefers to sit opposite T not A for the **Tn**/**Py** pair in this sequence context. This experimental finding is *opposite* the result predicted by *Lown* and *Dickerson* from

model building [31]. The reduced curvature of thiophene should exacerbate steric effects between the S-atom and the minor-groove floor, sterically permissive for A, T recognition but resulting in a >800-fold loss in binding affinity at the **X** = G, C sites. An *ab inito* calculation of the Im-Im-**Tn**-Py subunit reveals a likely steric interaction between the thiophene Me group and the proximal carboxamide O-atom, which may force the subunit to twist slightly out of plane. Although structural data on polyamide · DNA complexes reveals a tolerance to small amide–ring dihedrals, a thiophene without the Me substituent (**Tp**) and the **Tp**/**Py** pair were tested as a control. **Tp**/**Py** exhibited virtually identical DNA recognition behavior as **Tn**/**Py**.

In retrospect, in our search for a **Hp**/**Py** replacement, our group had great hopes for the **Ht**/**Py** to distinguish T · A from A · T. Yet this new **Ht** ring system was a complete disappointment. On the other hand, the earlier negative result with the thiazole **Th**/**Py** pair suggested that the S-atom on the corner of the ring was not a viable recognition element and that the thiophene **Tn**/**Py** or **Tp**/**Py** pairs were not likely to lead to T · A *vs.* A · T discrimination. The fact that **Tn** or **Tp** prefers to bind opposite T not A is the most significant unanticipated lead to emerge from the study.

The 1 : 1 Motif. The 1 : 1 motif has emerged recently as a way to target certain purine-rich DNA sequences (*e.g.*, GAGAA) with high affinity [16][17]. Although the cofacial pairing of rings in the hairpin motif offers a greater chance to differentiate between *Watson–Crick* base pairs, the single-subunit · DNA complexes of the 1 : 1 motif provide a relatively flexible system for the exploration of novel shape-selective recognition elements. Due to the conformational freedom imparted by the β-alanine residues, changes in heterocycle geometry do not have such a pronounced impact on the rest of the molecule. Therefore, specificity may be more difficult to achieve in this motif. In fact, all 1 : 1 polyamides described here bind with high affinity to the **X** = A, T sites but with varying degrees of **X** = A, T > G, C specificity. Structural studies reveal an important register of amide NH groups with the purine N(3) and pyrimidine O=C(2) groups on the floor of the DNA minor groove [19]. 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 *Fig. 7*) 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 model.

Py and **Nh** present a H-atom with a positive potential to the minor-groove floor. Both compounds exhibit a modest 3- to 5-fold selectivity for **X** = A, T > G, C, but **Nh** binds with higher affinity to all sites. The selectivity is probably due to the unfavorable steric X to G-NH2 interaction (X = H–C(3) for **Py** and H–N(1) for **Nh**) postulated for netropsin and supported by recent NMR studies [1][19]. The higher affinity for **Nh** may be attributed to a combination of greater positive charge at H–N(1) and higher ring curvature, both of which should reduce specificity.

Im, **Fr**, and **Nt** present a small atom with an sp^2 lone pair directed toward the minor-groove floor. **Im** has been reported previously [16][17], binding all sites with high affinity and displaying virtually no discrimination between sites. **Fr** and **Nt** 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 affinity.

Hp and **Ht** present an OH group to the DNA minor groove. In a different polyamide context, **Hp** discriminated between A·T and T·A base pairs in the 1:1 motif [30]. In this case, **Hp** is flanked on both sides by β-alanine residues, and specificity is lost. This loss may be attributed to a larger degree of conformational freedom afforded to the **Hp** ring by the aliphatic linkers [17]. Nonetheless, both **Hp** and **Ht** exhibit significant **X** = A, T > G, C specificity, as expected from a negative HO–C(3) to G-NH2 steric clash. **Ht** is slightly more specific, which may result from the non-planarity of this ligand as discussed above for **Ht**/**Py** in the hairpin motif.

Tn and **Th** present an S-atom with an sp^2 lone pair to the DNA minor groove. These compounds exhibit substantial **X** = 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 S-atom and the minor-groove floor. In the case of **X** = G, C, this interaction is very negative, resulting in a dramatic loss in binding affinity. **Th** is more specific than **Tn**, which is probably due to the curvature-induced non-planarity of Im-Im-**Tn**-Py, as discussed above for the hairpin motif (*Fig. 7*).

7. Conclusions. – Our understanding of the origin of DNA-sequence discrimination by polyamides has been improved by combining the tools of quantitative DNase-I footprinting and computational molecular modeling to establish a correlation between polyamide structure and DNA-sequence specificity (*Tables 3* and *4*). We believe that the footprinting results are best explained by differences in the overall heterocycle structure. Each heterocyclic amino acid has an inherently unique shape, which results in varying degrees of curvature complementarity between the polyamide and the DNA minor groove. Given that pyrrole is over-curved with respect to the DNA helix, reducing heterocycle curvature should increase the polyamide-DNA fit. Consequently, the polyamide would have greater sensitivity to changes in DNA structure and, therefore, greater DNA-sequence selectivity. On the other hand, increasing heterocycle curvature should decrease sensitivity to changes in DNA sequence. In addition, over-curvature can induce ligand non-planarity deriving from destabilizing eclipsing interactions. These results suggest that merely considering the functional group facing

Table 3. *Pairing Specificity of the Hairpin Motive*

Specificity for	Pair	A·T	T·A	G·C	C·G
Lone pair to floor	**Im/Py**	–	–	+	–
	Fr/Py	–	–	–	–
C–H or N–H to floor	**Py/Py**	+	+	–	–
	Nh/Py	+	+	–	–
	Pz/Py	+	+	–	–
C–OH to floor	**Hp/Py**	–	+	–	–
	Ht/Py	–	–	–	–
S to floor	**Th/Py**	–	–	–	–
	Tn/Py	+	+	–	–
	Tp/Py	+	+	–	–

Table 4. *Specificity in the 1:1 Motif*

Pair	A·T	T·A	G·C	C·G
Im	+	+	+	+
Py	+	+	−	−
Hp	+	+	−	−
Nh	+	+	+	+
Ht	+	+	−	−
Fr	+	+	+	+
Nt	+	+	+	+
Tn	+	+	−	−
Th	+	+	−	−

the minor-groove floor is insufficient for an accurate prediction of DNA-recognition behavior.

Curvature effects are amplified in contiguous-ring polyamides, where continuous π-conjugation limits conformational flexibility. Furthermore, the packing of cofacial polyamide subunits in the minor groove, as with the hairpin motif, provides additional preorganization. The dense functional array offered to the DNA by ring pairs affords greater promise for sequence discrimination. By contrast, the 1:1 motif offers a high degree of flexibility but a less dense functional array and, therefore, a lower capacity for DNA-sequence selectivity.

In retrospect, it is remarkable that a library of eight five-membered heterocycles (and hence, new heterocycle pairs) reveals very little new leads for sequence discrimination. This implies that the solution to DNA recognition by **Im**/**Py**, **Py**/**Im**, **Hp**/**Py**, and **Py**/**Hp** could be a narrow structural window. Perhaps the most novel and useful lead is the possibility that the **Tn**/**Py** pair will distinguish T·A from A·T. The next step is to test multiple **Tn**/**Py** pairs in different A,T-rich contexts to validate whether this is a potential breakthrough or not. This study is underway and results will be reported in due course.

Experimental Part

1. *General. N,N*-Dimethylformamide (DMF), *N,N*-diisopropylethylamine (DIEA), thiophenol, Et_2NH, *N,N*-dimethylpropane-1,3-diamine (Dp), Et_3N, methyl furane-2-carboxylate, oxobutanoic acid, methyl acetoacetate, cyanoacetic acid, trichloroacetyl chloride, 1*H*-pyrrole, Na metal, methyl thioglycolate, methyl 2-chloroacrylate, tin(II) chloride dihydrate, and thiourea were purchased from *Aldrich*. [(*tert*-Butoxy)carbonyl]-β-alanine-(4-carbonylaminomethyl)-benzyl-ester-copoly(styrene-divinylbenzene)resin (Boc-β-Pam resin), dicyclohexylcarbodiimide (DCC), 1-hydroxy-1*H*-benzotriazole (HOBt), 2-(1*H*-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), *N,N*-dimethylpyridin-4-amine (DMAP), and Boc-β-alanine were purchased from *NOVA Biochem*. CF_3COOH was purchased from *Halocarbon*. All other solvents were reagent grade from *EM*. Oligonucleotide inserts were synthesized by the Biopolymer Synthesis Center at the California Institute of Technology. Glycogen (20 mg/ml), dNTPs (PCR nucleotide mix), and all enzymes, unless otherwise stated, were purchased from *Boehringer-Mannheim*. pUC19 was purchased from *New England Biolabs*, and deoxyadenosine [γ-^{32}P]triphosphate was provided by *ICN*. Calf thymus DNA (sonicated, deproteinized) and DNase I (7500 units/mL, FPLC pure) were from *Amersham Pharmacia*. AmpliTaq DNA polymerase was from *Perkin-Elmer* and used with the provided buffers. *Tris*·HCl, DTT, RNase-free H_2O, and 0.5M EDTA were from *United States Biochemical*. Calcium chloride, potassium chloride, and magnesium chloride were purchased from *Fluka*. *Tris* borate EDTA buffer was from *GIBCO*, and bromophenol blue was from *Acros*. All reagents were used without further purification. Column chromatography (CC): silica gel *60*,

EM. TLC: *J. T. Baker*. UV Spectra: *Hewlett-Packard 8452A* diode array spectrophotometer. IR Spectra: in cm^{-1}. NMR Spectra: *Varian* spectrometer at 300 MHz, (D_6)DMSO or $CDCl_3$ solns.; δ in ppm rel. to residual solvent, *J* in Hz. High-resolution FAB- and EI-MS: recorded at the Mass Spectroscopy Laboratory at the University of California, Los Angeles; in *m/z*. Matrix-assisted, laser desorption/ionization time of flight (MALDI-TOF)-MS: recorded at the Protein and Peptide Microanalytical Facility at the California Institute of Technology.

2. *Monomer and 'Dimer' Synthesis. 5-Aminofuran-2-carboxylic Acid Methyl Ester* (NH_2-Fr-OMe; **1**). Methyl 5-nitrofuran-2-carboxylate (**18**) was prepared by published methods [21] in 84% yield. TLC (hexanes/AcOEt 5:2). R_f 0.7. ^{1}H-NMR ((D_6)DMSO): 7.78 (*d*, *J* = 3.9, 1 H); 7.59 (*d*, *J* = 3.9, 1 H); 3.38 (*s*, 3 H). ^{13}C-NMR ((D_6)DMSO): 157.7; 152.8; 144.5; 120.4; 113.6; 53.5. EI-MS: 171.117 (M^+, $C_6H_5NO_5^+$; calc. 171.117).

A mixture of **18** (3 g, 17.5 mmol) and 10% Pd/C (0.3 g) in AcOEt (25 ml) was hydrogenated in a *Parr* apparatus at 500 psi and r.t. for 1.5 h (TLC monitoring). The mixture was filtered over a 2.5-cm pad of *Celite* to remove Pd/C. The filtrate was cooled to $-20°$, and hexanes were added until a white precipitate was formed. The precipitate (2.1 g) was collected by vacuum filtration and washed with Et_2O: **1** (85%). TLC (hexanes/AcOEt 5:2): R_f 0.25. IR (film): 3395, 3322, 1684, 1627, 1528, 1441, 1338, 1297, 1199, 1153. ^{1}H-NMR ((D_6)DMSO): 7.13 (*d*, *J* = 3.6, 1 H); 6.60 (*s*, 2 H); 5.09 (*d*, *J* = 3.6, 1 H); 3.66 (*s*, 3 H). ^{13}C-NMR ($CDCl_3$): 123.8; 123.0; 121.9; 88.3; 86.3. EI-MS: 141.042 (M^+, $C_6H_7NO_3^+$; calc. 141.042).

*Methyl 5-{[(1-Methyl-4-nitro-1*H*-imidazol-2-yl)carbonyl]amino}furan-2-carboxylate* (NO_2-Im-Fr-OMe; **19**). A mixture of **1** (1 g, 7.08 mmol), NO_2-Im-$COCCl_3$ (2.31 g, 8.5 mmol), DIEA (1.1 g, 1.48 ml, 8.5 mmol), and AcOEt (14 ml) was stirred at 35° for 12 h. The mixture was cooled to r.t., and sufficient hexanes were added to completely precipitate a pale yellow solid. The precipitate was collected by vacuum filtration and washed with cold MeOH and Et_2O: **19** (1.18 g, in 57%). TLC (hexanes/AcOEt 5:2): R_f 0.20. IR (film): 3195, 3133, 1731, 1676, 1553, 1541, 1525, 1442, 1387, 1313, 1146. ^{1}H-NMR ((D_6)DMSO): 12.17 (br., 1 H); 8.65 (*s*, 1 H); 7.34 (*d*, *J* = 3.6, 1 H); 6.52 (*d*, *J* = 3.6, 1 H); 4.01 (*s*, 3 H); 3.77 (*s*, 3 H). ^{13}C-NMR ((D_6)DMSO): 158.6; 155.8; 150.0; 145.0; 137.3; 137.0; 127.8; 121.4; 99.0; 52.3; 37.3. EI-MS: 294.060 (M^+, $C_{11}H_{10}N_4O_6^+$; calc. 294.060).

*Methyl 5-{[(4-Amino-1-methyl-1*H*-imidazol-2-yl)carbonyl]amino}furan-2-carboxylate Hydrochloride* ($HCl \cdot H_2N$-Im-Fr-OMe; **20**). A mixture of **19** (1 g, 3.4 mmol) and 10% Pd/C (0.2 g) in AcOEt (7 ml) was hydrogenated in a *Parr* apparatus at 500 psi and r.t. for 1.5 h. The mixture was filtered over a 2.5-cm pad of *Celite* to remove the Pd/C. The AcOEt was evaporated and 2M HCl in Et_2O was added to give **20** (0.62 g, 61%). TLC (hexanes/AcOEt 5:2): R_f 0.15 (amine), 0.0 (salt). IR (film): 2985, 3008, 1708, 1689, 1537, 1318, 1193, 1142, 1119, 1018, 756, 668. ^{1}H-NMR ((D_6)DMSO): 11.73 (br., 1 H); 7.54 (*s*, 1 H); 7.33 (*d*, *J* = 3.6, 1 H); 6.51 (*s*, *J* = 3.6, 1 H); 3.96 (*s*, 3 H); 3.76 (*s*, 3 H). ^{13}C-NMR ((D_6)DMSO): 157.9; 155.0; 149.5; 136.3; 134.8; 129.5; 120.9; 118.7; 97.9; 51.6; 35.8. EI-MS: 264.086 (M^+, $C_{11}H_{12}N_4O_4^+$; calc. 264.086).

*Methyl 5-{{[4-{[(*tert*-Butoxy)carbonyl]amino}-1-methyl-1*H*-imidazol-2-yl}carbonyl}amino}furan-2-carboxylate* (Boc-Im-Fr-OMe; **21**). A mixture of **20** (0.5 g, 1.6 mmol), $(Boc)_2O$ (545 mg, 2.4 mmol), and DIEA (258 mg, 348 µl, 2 mmol) in DMF (5 ml) was stirred at 60° for 18 h. The mixture was added to ice-water (0.5 l) and the precipitate extracted with AcOEt (50 ml). The org. layer was dried (Na_2SO_4) and evaporated, and the resulting residue was subjected to CC (hexanes/AcOEt 1:1): **21** (376 mg, 62%). White solid. TLC (hexanes/AcOEt 1:1): R_f 0.65. IR (film): 3243, 2978, 1722, 1577, 1533, 1436, 1368, 1311, 1163, 1138, 755. ^{1}H-NMR ((D_6)DMSO): 11.11 (*s*, 1 H); 9.52 (*s*, 1 H); 7.33 (*d*, *J* = 3.3, 1 H); 6.47 (*d*, *J* = 3.3, 1 H); 3.90 (*s*, 3 H); 3.76 (*s*, 3 H); 1.44 (*s*, 9 H). ^{13}C-NMR ((D_6)DMSO): 158.6; 156.1; 153.6; 150.4; 137.7; 136.8; 133.1; 121.7; 115.3; 97.8; 79.8; 52.3; 35.9; 28.9. EI-MS: 364.138 (M^+, $C_{16}H_{20}N_4O_6^+$; calc. for 364.138).

*5-{{[4-{[(*tert*-Butoxy)carbonyl]amino}-1-methyl-1*H*-imidazol-2-yl}carbonyl}amino}furan-2-carboxylic Acid* (Boc-Im-Fr-OH; **15**). A mixture of **21** (0.3 g, 0.82 mmol), 1N NaOH (5 ml), and MeOH (1 ml) was stirred at r.t. for 3 h (TLC monitoring). The MeOH was evaporated and the aq. layer carefully adjusted to pH 2 with 1N HCl. The milky white precipitate was extracted with AcOEt, and the extract was dried (Na_2SO_4) and evaporated: **15** (270 mg, 94%). Fine white powder. TLC (hexanes/AcOEt 1:1, 10% AcOH): R_f 0.5. IR (film): 3231, 2917, 2856, 1688, 1542, 1311, 1259, 1163, 1119. ^{1}H-NMR ((D_6)DMSO): 10.94 (*s*, 1 H); 9.52 (*s*, 1 H); 7.30 (*s*, 1 H); 7.22 (*d*, *J* = 3.6, 1 H); 6.44 (*d*, *J* = 3.6, 1 H); 3.91 (*s*, 3 H); 1.44 (*s*, 9 H). ^{13}C-NMR ((D_6)DMSO): 160.1; 159.3; 149.9; 136.5; 135.5; 133.2; 123.6; 115.4; 101.4; 97.0; 82.3; 35.9; 28.9. EI-MS: 350.123 (M^+, $C_{15}H_{18}N_4O_6^+$; calc. 350.123).

*Methyl 5-{[3-{[(*tert*-Butoxy)carbonyl]amino}propanoyl}amino}furan-2-carboxylate* (Boc-β-Fr-OMe; **22**). A mixture of Boc-β-alanine (3.22 g, 17 mmol) and DCC (1.75 g, 8.5 mmol) in CH_2Cl_2 (25 ml) was stirred at r.t. for 30 min. To the above mixture was added **1** (0.6 g, 4.25 mmol) as a soln. in DMF (5 ml) and DIEA (0.741 ml, 0.55 g, 4.25 mmol), followed by the addition of DMAP (0.155 g, 1.27 mmol). The mixture was heated to 40°, stirred overnight, and then filtered to remove the dicyclohexylurea. The filtrate was poured into ice-water (0.5 l)

upon which time a crude white precipitate formed. The crude precipitate was extracted with AcOEt and subjected to CC (hexanes/AcOEt 5:2): **22** (1.1 g, 85%). Flaky white solid. TLC (hexanes/AcOEt 5:2) R_f 0.2. IR (film): 3372, 3234, 3036, 1957, 1728. ^{1}H-NMR ((D$_6$)DMSO): 11.58 (*s*, 2 H); 7.29 (*d*, *J* = 3.8, 1 H); 6.86 (*s*, 2 H); 6.36 (*d*, *J* = 3.8, 1 H); 3.76 (*s*, 3 H); 3.32 (*q*, *J* = 6, 2 H); 2.69 (*t*, *J* = 6, 3 H); 1.35 (*s*, 9 H). ^{13}C-NMR ((D$_6$)DMSO): 167.9; 159.4; 155.8; 150.1; 137.0; 120.8; 95.2; 78.3; 51.6; 36.1; 35.8; 29.0. EI-MS: 312.132 (M^+, $C_{14}H_{20}N_2O_6^+$; calc. 312.132).

*5-{[3-{[(*tert*-Butoxy)carbonyl]amino}propanoyl]amino}furan-2-carboxylic Acid* (Boc-*β*-Fr-OH; **13**). As described for **15**, with **22** (1.1 g, 3.52 mmol), 1N NaOH (15 ml), and MeOH (5 ml) for 4 h: **13** (0.97 g, 92%). Off-white solid. TLC (hexanes/AcOEt 5:2, 10% AcOH): R_f 0.6. IR (film): 3321, 3270, 3979, 1684, 1522. ^{1}H-NMR ((D$_6$)DMSO): 11.28 (*s*, 2 H); 6.98 (*d*, *J* = 3.6, 1 H); 6.82 (*s*, 2 H); 6.23 (*d*, *J* = 3.6, 1 H); 3.19 (*q*, *J* = 6, 2 H); 2.45 (*t*, *J* = 6, 2 H); 1.35 (*s*, 9 H). ^{13}C-NMR ((D$_6$)DMSO): 168.7; 159.7; 156.0; 150.6; 137.1; 120.8; 95.8; 78.3; 36.9; 36.6; 29.0. EI-MS: 298.117 (M^+, $C_{13}H_{18}N_2O_6^+$; calc. 298.116).

*2,2,2-Trichloro-1-(1*H*-pyrrol-2-yl)ethan-1-one* (**23**). A soln. of 1*H*-pyrrole (20.6 ml, 20 g, 298 mmol) and Et_2O (86 ml) was added dropwise to trichloroacetyl chloride (71.9 ml, 117 g, 644 mmol) with stirring at 0°. The mixture was allowed to warm to r.t. and stirred overnight. The solvent was evaporated, and the residue was reprecipitated from hexanes: **23** (24 g, 38%). White solid. TLC (hexanes/AcOEt 5:2): R_f 0.75. IR (film): 3322, 1656, 1388, 1136, 1035, 953, 842, 808, 754, 733, 688. ^{1}H-NMR ((D$_6$)DMSO): 12.4 (*s*, 1 H); 7.32 (*m*, *J* = 2.1, 1 H); 7.29 (*m*, *J* = 2.1, 1 H); 6.34 (*m*, *J* = 2.1, 1 H). ^{13}C-NMR ((D$_6$)DMSO): 172.5; 130.0; 122.3; 121.9; 112.0; 95.9. EI-MS: 210.936 (M^+, $C_6H_4Cl_3NO^+$; calc. for 210.936).

*2,2,2-Trichloro-1-(5-nitro-1*H*-pyrrol-2-yl)ethan-1-one* (**24**). A soln. of **23** (20 g, 95 mmol) and Ac_2O (111 ml) was cooled to −40° and treated dropwise with 70% nitric acid (8.24 ml) over 2 h. After completion of addition, the mixture was warmed to r.t. over 2 h and then cooled back down to −40°. Sufficient ice-water was added to pecipitate **24** (16.5 g, 68%). White solid. TLC (hexanes/AcOEt 5:2): R_f 0.6. IR (film): 3316, 1676, 1551, 1518, 1405, 1379, 1317. ^{1}H-NMR ((D$_6$)DMSO): 13.62 (*s*, 1 H); 8.33 (*d*, *J* = 3, 1 H); 7.66 (*d*, *J* = 3, 1 H). ^{13}C-NMR ((D$_6$)DMSO): 173.4; 137.6; 128.4; 122.0; 115.0; 94.4. EI-MS: 255.921 (M^+, $C_6H_3Cl_3N_2O_3^+$; calc. 255.921).

*Ethyl 5-Nitro-1*H*-pyrrole-2-carboxylate* (NO_2-Nh-OEt; **25**). To a mixture of **24** (10 g, 39 mmol) in EtOH (35 ml) at r.t. was added EtONa (4 g, 59 mmol). The mixture was stirred for 2 h, then quenched with H_2SO_4, and cooled to 0°. Ice-water was added (0.5 l) to precipitate **25** (7 g, 97%). Tan solid. TLC (hexanes/AcOEt 5:2): R_f 0.5. IR (film): 3263, 3152, 2979, 1687, 1565, 1508, 1365, 1323, 1207, 1017, 752. ^{1}H-NMR ((D$_6$)DMSO): 8.04 (*d*, *J* = 1.5, 1 H); 7.23 (*d*, *J* = 1.5, 1 H); 4.29 (*q*, *J* = 7.2, 2 H); 1.29 (*t*, *J* = 7.2, 3 H). ^{13}C-NMR ((D$_6$)DMSO): 160.0; 137.2; 124.9; 123.6; 110.1; 61.4; 14.9. EI-MS: 184.048 (M^+, $C_7H_8N_2O_4^+$; calc. 184.048).

*Ethyl 5-Amino-1*H*-pyrrole-2-carboxylate Hydrochloride* ($HCl \cdot H_2N$-Nh-OEt; **2**). A mixture of **25** (3 g, 16 mmol) and 10% Pd/C (0.3 g) in AcOEt (25 ml) was hydrogenated in a *Parr* apparatus at 500 psi and r.t. for 1.5 h. The mixture was filtered over a 2.5-cm pad of *Celite* to remove Pd/C. The filtrate was cooled to 20°, and HCl in Et_2O was added. Upon addition, the hydrochloride salt precipitated and was collected by vacuum filtration: **2** (2.4 g, 78%). White solid. TLC (hexanes/AcOEt 1:1): R_f 0.15 (amine), 0.0 (hydrochloride). IR (film): 2914, 1694, 1495, 1429, 1376, 1345, 1284, 1224, 1106, 1020, 965. ^{1}H-NMR ((D$_6$)DMSO): 12.21 (*s*, 1 H); 10.13 (*s*, 1 H); 7.10 (*d*, *J* = 1.8, 1 H); 6.74 (*d*, *J* = 1.8, 1 H); 4.24 (*q*, *J* = 7.2, 2 H); 1.26 (*t*, *J* = 7.2, 3 H). ^{13}C-NMR ((D$_6$)DMSO): 159.6; 121.3; 117.7; 115.6; 109.2; 60.0; 14.3. EI-MS: 154.074 (M^+, $C_7H_{10}N_2O_2^+$; calc. 154.074).

*Ethyl 5-{[(*tert*-Butoxy)carbonyl]amino}-1*H*-pyrrole-2-carboxylate* (Boc-Nh-OEt; **26**). As described for **21**, with **2** (2 g, 11 mmol), $(Boc)_2O$ (3.6 g, 16.5 mmol), DIEA (2.1 ml, 1.56 g, 12.1 mmol), and DMF (15 ml) for 12 h. Workup with ice-water (1 l) and AcOEt (150 ml). CC (hexanes/AcOEt 5:2) afforded **26** (2 g, 72%). Flaky white solid. TLC (hexanes/AcOEt 5:2): R_f 0.65. IR (film): 3296, 1683, 1570, 1384, 1315, 1264, 1249. ^{1}H-NMR ((D$_6$)DMSO): 11.48 (*s*, 1 H); 9.06 (*s*, 1 H); 6.93 (*d*, *J* = 1.8, 1 H); 6.58 (*d*, *J* = 1.8, 1 H); 4.21 (*q*, *J* = 7.2, 2 H); 1.41 (*s*, 9 H); 1.24 (*t*, *J* = 7.2, 3 H). ^{13}C-NMR ((D$_6$)DMSO): 160.1; 152.5; 119.1; 112.4; 105.2; 78.3; 59.4; 28.2; 14.4. EI-MS: 254.127 (M^+, $C_{12}H_{18}N_2O_4^+$; calc. 254.127).

*5-{[(*tert*-Butoxy)carbonyl]amino}-1*H*-pyrrole-2-carboxylic Acid* (Boc-Nh-OH; **12**). As described for **15** with **26** (2 g, 7.9 mmol), 1N NaOH (15 ml), and MeOH (5 ml) for 3 h: **12** (1.6 g, 92%). Off-white solid. TLC (hexanes/AcOEt 5:2, 10% AcOH): R_f 0.5. IR (film): 3329, 3153, 2969, 1691, 1586, 1549, 1434, 1374, 1250, 1167, 1117, 1057, 961, 762. ^{1}H-NMR ((D$_6$)DMSO): 11.32 (*s*, 1 H); 9.01 (*s*, 1 H); 6.88 (*s*, 1 H); 6.52 (*s*, 1 H); 1.41 (*s*, 9 H). ^{13}C-NMR ((D$_6$)DMSO): 162.3; 153.3; 125.4; 120.7; 112.6; 106.0; 79.0; 28.9. EI-MS: 226.095 (M^+, $C_{10}H_{14}N_2O_4^+$; calc. 226.095).

Methyl 4-Cyano-3-methylbut-3-enoate (**27**). A mixture of acetoacetate (30 g, 258 mmol), cyanoacetic acid (24 g, 284 mmol), NH_4OAc (3.98 g, 51.6 mmol), AcOH (6.65 ml, 6.98 g, 116 mmol), and benzene (75 ml) was stirred for 12 h at 145° (round-bottom flask, *Dean–Stark* apparatus, condenser). The mixture was allowed to

cool to r.t., washed with brine (0.3 l), sat. $NaHCO_3$ soln. (0.3 l), dried ($MgSO_4$), and evaporated. The crude product was distilled at 60°/0.1 Torr: **27** (23 g, 65%), (*E*)/(*Z*) mixture. Clear liquid. IR (film): 2957, 2221, 1741, 1437. ^{1}H-NMR ((D_6)DMSO): 5.69 (*q*, *J* = 0.6, 1 H); 5.62 (*q*, *J* = 0.6, 1 H); 3.61 (*s*, 3 H); 3.60 (*s*, 3 H); 3.42 (*s*, 2 H); 3.35 (*d*, *J* = 1.2, 2 H); 2.01 (*d*, *J* = 1.2, 3 H); 1.93 (*d*, *J* = 1.2, 3 H). ^{13}C-NMR ((D_6)DMSO): 170.1; 169.5; 158.4; 158.1; 117.4; 117.3; 99.6; 99.4; 52.8; 52.7; 42.8; 41.3; 23.6; 21.7. EI-MS: 139.063 (M^+, $C_7H_9NO_2^+$; calc. 139.063).

Methyl 5-Amino-3-methylthiophene-2-carboxylate Hydrochloride ($HCl \cdot H_2N$-Tn-OMe; **3**). Et_2NH (18.7 ml, 13.2 g, 181 mmol) was added dropwise to a mixture of **27** (23 g, 165 mmol) and S-flakes (5.28 g, 165 mmol) in EtOH (130 ml) and stirred at r.t. for 3 h. The mixture was concentrated to a minimal volume and placed in an ice bath. Conc. HCl was slowly added to the mixture to give a light orange solid. The precipitate was collected by vacuum filtration and washed repeatedly with Et_2O: **3** (19 g, 68%). TLC (hexanes/AcOEt 5:2): R_f 0.55 (amine), 0.0 (hydrochloride). IR (film): 3422, 3339, 3204, 2849, 1713, 1677, 1546, 1462, 1269, 1187, 1092. ^{1}H-NMR ((D_6)DMSO): 6.91 (*s*, 2 H); 5.76 (*s*, 1 H); 3.61 (*s*, 3 H); 2.62 (*s*, 3 H). ^{13}C-NMR ((D_6)DMSO): 163.5; 146.7; 145.5; 114.9; 114.7; 52.0; 16.6. EI-MS: 171.035 (M^+, $C_7H_9NO_2S^+$; calc. 171.035).

*Methyl 3-Methyl-5-{[(1-methyl-4-nitro-1*H*-imidazol-2-yl)carbonyl]amino}thiophene-2-carboxylate* (NO_2-Im-Tn-OMe; **28**). As described for **19**, from **3** (1 g, 4.8 mmol): **28** (0.87 g, 56%). Yellow solid. TLC (hexanes/AcOEt 5:2): R_f 0.60. IR (film): 3125, 1649, 1543, 1506, 1382, 1312, 1267. ^{1}H-NMR ((D_6)DMSO): 12.43 (*s*, 1 H); 8.66 (*s*, 1 H); 6.95 (*s*, 1 H); 4.03 (*s*, 3 H); 3.79 (*s*, 3 H); 2.41 (*s*, 3 H). ^{13}C-NMR ((D_6)DMSO): 163.5; 155.7; 145.1; 144.7; 143.6; 136.9; 128.0; 118.8; 117.2; 52.2; 37.5; 16.6. EI-MS: 324.053 (M^+, $C_{12}H_{12}N_4O_5S^+$; calc. 324.053).

*Methyl 5-{[(4-Amino-1-methyl-1*H*-imidazol-2-yl)carbonyl]amino}-3-methylthiophene-2-carboxylate hydrochloride* ($HCl \cdot H_2N$-Im-Tn-OMe; **29**). As described for **20**, from **28** (0.5 g, 1.5 mmol): **29** (321 mg, 63%). Pale yellow solid. TLC (hexanes/AcOEt 5:2): R_f 0.25 (amine), 0.0 (hydrochloride). IR (film): 3294, 1735, 1674, 1562, 1520, 1440, 1407, 1267, 1185, 1091. ^{1}H-NMR ((D_6)DMSO): 12.14 (*s*, 1 H); 7.47 (*s*, 1 H); 6.95 (*s*, 1 H); 3.99 (*s*, 1 H); 3.73 (*s*, 1 H); 2.39 (*s*, 1 H). ^{13}C-NMR ((D_6)DMSO): 177.9; 63.4; 157.6; 155.1; 144.1; 128.0; 118.6; 116.1; 52.2; 35.4; 16.4. EI-MS: 294.079 (M^+, $C_{12}H_{14}N_4O_3^+$; calc. for 294.079).

*Methyl 5-{{[4-{[(*tert*-Butoxy)carbonyl]amino}-1-methyl-1*H*-imidazol-2-yl}carbonyl}amino}-3-methylthiophene-2-carboxylate* (Boc-Im-Tn-OMe; **30**). As described for **21**, from **29** (300 mg, 0.91 mmol): **30** (239 mg, 67%). Pale yellow solid. TLC (hexanes/AcOEt 1:1): R_f 0.7. IR (film): 3424, 3219, 2995, 1750, 1704, 1677, 1571, 1251, 1141. ^{1}H-NMR ((D_6)DMSO): 11.78 (*s*, 1 H); 9.33 (*s*, 1 H); 7.32 (*s*, 1 H); 6.90 (*s*, 1 H); 3.91 (*s*, 1 H); 3.73 (*s*, 1 H); 2.39 (*s*, 1 H); 1.43 (*s*, 9 H). ^{13}C-NMR ((D_6)DMSO): 162.8; 155.7; 152.8; 144.0; 143.4; 136.7; 132.4; 117.0; 115.6; 115.3; 79.4; 51.3; 35.2; 26.9; 15.9. EI-MS: 394.131 (M^+, $C_{17}H_{22}N_4O_5S^+$; calc. 394.131).

*5-{{[4-{[(*tert*-Butoxy)carbonyl]amino}-1-methyl-1*H*-imidazol-2-yl}carbonyl}amino}-3-methylthiophene-2-carboxylic Acid* (Boc-Im-Tn-OH; **16**). As described for **15**, with **30** (200 mg, 0.507 mmol), MeOH (1 ml), and 1N NaOH (5 ml) at 60° for 6 h: **16** (158 mg, 82%). Pale tan solid. TLC (hexanes/AcOEt 1:1, 10% AcOH): R_f 0.8. IR (film): 3400, 2976, 3231, 2961, 1722, 1678, 1589, 1253, 1179, 1091. ^{1}H-NMR ((D_6)DMSO): 11.68 (*s*, 1 H); 9.35 (*s*, 1 H); 7.32 (*s*, 1 H); 6.87 (*s*, 1 H); 3.93 (*s*, 3 H); 2.38 (*s*, 3 H); 1.44 (*s*, 9 H). ^{13}C-NMR ((D_6)DMSO): 164.0; 158.3; 155.6; 142.8; 136.7; 117.3; 117.1; 115.2; 99.4; 81.4; 35.2; 28.1; 15.8. EI-MS: 380.115 (M^+, $C_{16}H_{20}N_4O_5S^+$; calc. 380.115).

*Methyl 5-{{3-{[(*tert*-Butoxy)carbonyl]amino}propanoyl}amino}-3-methylthiophene-2-carboxylate* (Boc-*β*-Tn-OMe; **31**). A mixture of Boc-*β*-alanine (1 g, 5.28 mmol) and DCC (545 mg, 2.64 mmol) in CH_2Cl_2 (10 ml) was stirred at r.t. for 30 min. The mixture was then filtered into a round-bottom flask containing **3** (382 mg, 1.8 mmol), DIEA (322 µl, 239 mg, 1.8 mmol), DMAP (100 mg, 0.8 mmol), and DMF (8 ml). The mixture was heated at 45° (TLC monitoring). Additional symmetrical anhydride (1.4 equiv.) was added every 8 h, as needed, until completion of the reaction. The mixture was then added to brine (0.2 l) and extracted twice with AcOEt (50 ml). The org. layer was washed with sat. $NaHCO_3$ soln. (0.1 l) and 10 mM HCl (0.1 l), dried (Na_2SO_4), and evaporated, and the crude residue subjected to CC (hexanes/AcOEt 5:2): **31** (392 mg, 62%). Flaky white powder. TLC (hexanes/AcOEt 5:2): R_f 0.32. IR (film): 3348, 3450, 2981, 1684, 1568, 1522, 1445, 1252. ^{1}H-NMR ($CDCl_3$): 10.11 (*s*, 1 H); 6.49 (*s*, 1 H); 3.79 (*s*, 3 H); 3.49 (*q*, *J* = 6, 2 H); 2.64 (*t*, *J* = 6, 2 H); 1.40 (*s*, 9 H). ^{13}C-NMR ($CDCl_3$): 168.7; 164.1; 157.0; 144.9; 143.4; 128.4; 117.3; 116.1; 80.4; 51.7; 37.0; 34.8; 16.4. EI-MS: 342.124 (M^+, $C_{15}H_{22}N_2O_5S^+$; calc. 342.124).

*5-{{3-{[(*tert*-Butoxy)carbonyl]amino}propanoyl}amino}-3-methylthiophene-2-carboxylic Acid* (Boc-*β*-Tn-OH; **14**). As described for **15**, with **31** (200 mg, 0.58 mmol), MeOH (4 ml), and 1N NaOH (15 ml) at 60° for 6 h: **14** (180 mg, 94%). Light yellow solid. TLC (hexanes/AcOEt 5:2, 10% AcOH): R_f 0.5. IR (film): 3255, 2976, 2954, 1674, 1569, 1522, 1445, 1253. ^{1}H-NMR ((D_6)DMSO): 12.35 (*s*, 1 H); 11.42 (*s*, 1 H); 6.89 (*s*, 1 H); 6.48 (*s*, 1 H); 3.21 (*q*, *J* = 6, 2 H); 2.50 (*t*, *J* = 6, 2 H); 2.37 (*s*, 3 H); 1.35 (*s*, 9 H). ^{13}C-NMR ((D_6)DMSO): 168.2;

164.1; 155.4; 143.3; 143.1; 116.8; 115.0; 77.7; 36.3; 35.6; 28.2; 15.7. EI-MS: 328.109 (M^+, $C_{14}H_{20}N_2O_5S^+$; calc. 328.109).

5-Nitrothiophene-2-carboxylic Acid (NO_2-Tp-OH; **32**). A mixture of NaOCl (26.35 g, 291 mmol) and $Na_2HPO_4 \cdot H_2O$ (30.3 g, 219 mmol) in H_2O (250 ml) was added dropwise to a soln. of commercially available 5-nitrothiophene-2-carboxaldehyde (5 g, 31.8 mmol) in acetone (0.6 l) at r.t. Upon completion of addition, TLC showed total consumption of the starting aldehyde. The mixture was washed with hexanes (0.1 l) and acidified to pH 2 with 1N HCl. The mixture was extracted with Et_2O (3×0.1 l) and the extract dried (Na_2SO_4) and evaporated: **32** (3.74 g, 68%). White solid. TLC (hexanes/AcOEt 5:2, 10% AcOH): R_f 0.55. IR (film): 3118, 3109, 2876, 1688, 1680, 1512, 1350, 1336, 1274. ^{1}H-NMR ((D$_6$)DMSO): 8.12 (*d*, $J = 4.2$, 1 H); 7.73 (*d*, $J = 4.2$, 1 H). ^{13}C-NMR ((D$_6$)DMSO): 162.2; 154.5; 141.0; 132.6; 130.5. EI-MS: 172.978 (M^+, $C_5H_3NO_4S^+$; calc. 172.978).

Methyl 5-Nitrothiophene-2-carboxylate (NO_2-Tp-OMe; **33**). A mixture of **32** (3.5 g, 20.2 mmol), conc. H_2SO_4 (0.2 g, 110 µl, 2.0 mmol), and MeOH (50 ml) was refluxed for 48 h. The MeOH was evaporated and the residue neutralized with 1N NaOH. The mixture was extracted with AcOEt (2×0.1 l) and the extract dried (Na_2SO_4) and evaporated: **33** (3.4 g, 91%). Crystalline white solid. TLC (hexanes/AcOEt 1:1): R_f 0.8. IR (film): 3476, 3115, 1730, 1705, 1535, 1508, 1423, 1360, 1282, 1250, 1191, 997, 856, 748, 732. ^{1}H-NMR ((D$_6$)DMSO): 8.21 (*d*, $J = 3.9$, 1 H); 7.80 (*d*, $J = 3.9$, 1 H); 3.87 (*s*, 3 H). ^{13}C-NMR ((D$_6$)DMSO): 161.2; 155.0; 138.4; 133.3; 130.4; 54.0. EI-MS: 186.994 (M^+, $C_6H_5NO_4S^+$; calc. 186.994).

Methyl 5-Aminothiophene-2-carboxylate Hydrochloride ($HCl \cdot H_2N$-Tp-OMe; **34**). Conc. HCl soln. (5.8 ml) was added dropwise to a mixture of **33** (0.3 g, 1.6 mmol) and $SnCl_2 \cdot 2\,H_2O$ (2.43 g, 12.8 mmol) in 95% EtOH (5.8 ml) at r.t. Sufficient cooling was necessary to keep the reaction temp. under 35°. The mixture was stirred at 35° for 2 h. The EtOH was evaporated and the aq. layer washed with hexanes (2×50 ml). The aq. layer was neutralized with 1N NaOH to pH 9 ($\rightarrow$ milky white emulsion), the mixture was extracted several times with AcOEt (50 ml), and the org. phase was dried (Na_2SO_4) and evaporated to give a thin yellow film. Addition of 2M HCl in Et_2O gave **34** (220 mg, 71%). White solid. TLC (hexanes/AcOEt 1:1): R_f 0.55 (amine), 0.0 (hydrochloride). IR (film): 3219, 1731, 1706, 1471, 1272, 1088, 739. ^{1}H-NMR ((D$_6$)DMSO): 7.33 (*d*, $J = 4.5$, 1 H); 6.70 (*s*, 2 H); 5.87 (*d*, $J = 4.5$, 1 H); 3.65 (*s*, 3 H). ^{13}C-NMR ((D$_6$)DMSO): 163.5; 162.8; 136.2; 105.1; 51.8. EI-MS: 157.020 (M^+, $C_6H_7NO_2S$; calc. 157.020).

*Methyl 5-{[(1-Methyl-4-nitro-1*H*-imidazol-2-yl)carbonyl]amino}thiophene-2-carboxylate* (NO_2-Im-Tp-OMe; **35**). As described for **19**, from **34** (200 mg, 1.0 mmol): **35** (164 mg, 51%). Yellow solid. TLC (hexanes/AcOEt 5:2): R_f 0.50. IR (film): 3133, 1698, 1672, 1560, 1543, 1521, 1455, 1379, 1313, 1265, 1098, 746.9. ^{1}H-NMR ((D$_6$)DMSO): 12.55 (*s*, 1 H); 8.68 (*s*, 1 H); 7.64 (*d*, $J = 3.9$, 1 H); 7.16 (*d*, $J = 3.9$, 1 H); 4.04 (*s*, 3 H); 3.77 (*s*, 3 H). ^{13}C-NMR ((D$_6$)DMSO): 163.1; 155.6; 145.8; 136.8; 132.6; 128.0; 123.5; 115.3; 52.6; 37.5. EI-MS: 310.037 (M^+, $C_{11}H_{10}N_4O_5S^+$; calc. 310.037).

*Methyl 5-{[(4-Amino-1-methyl-1*H*-imidazol-2-yl)carbonyl]amino}thiophene-2-carboxylate Hydrochloride* ($HCl \cdot H_2N$-Im-Tp-OMe; **36**). As described for **20**, from **35** (150 mg, 0.48 mmol): **36** (107 mg, 70%). Off-white solid. TLC (AcOEt): R_f 0.45 (amine), 0.0 (hydrochloride). IR (film): 3344, 3204, 2954, 1691, 1673, 1561, 1511, 1458, 1438, 1343, 1275, 1100. ^{1}H-NMR ((D$_6$)DMSO): 12.20 (*s*, 1 H); 7.63 (*d*, $J = 4.2$, 1 H); 7.40 (*s*, 1 H); 7.11 (*d*, $J = 4.2$, 1 H); 3.99 (*s*, 3 H); 3.76 (*s*, 3 H). ^{13}C-NMR ((D$_6$)DMSO): 177.4; 162.3; 159.4; 155.0; 145.4; 131.8; 128.5; 113.8; 51.7; 35.6. EI-MS: 280.063 (M^+, $C_{11}H_{12}N_4O_3S^+$; calc. 280.063).

*Methyl 5-{{{4-{[(*tert*-Butoxy)carbonyl]amino}-1-methyl-1*H*-imidazol-2-yl}carbonyl}amino}thiophene-2-carboxylate* (Boc-Im-Tp-OMe; **37**). As described for **21**, from **36** (100 mg, 0.35 mmol): **37** (90 mg, 66%). White solid. TLC (hexanes/AcOEt 1:1): R_f 0.75. IR (film): 3282, 2964, 1707, 1692, 1673, 1573, 1550, 1368, 1341, 1273, 1159, 1096. ^{1}H-NMR ((D$_6$)DMSO): 11.91 (*s*, 1 H); 9.32 (*s*, 3 H); 7.62 (*d*, $J = 4.2$, 1 H); 7.33 (*s*, 1 H); 7.07 (*d*, $J = 4.2$, 1 H); 3.94 (*s*, 3 H); 3.76 (*s*, 3 H); 1.44 (*s*, 9 H). ^{13}C-NMR ((D$_6$)DMSO): 162.3; 155.5; 152.8; 145.7; 136.7; 132.4; 131.8; 122.0; 115.4; 113.5; 51.7; 35.2; 28.1. EI-MS: 380.115 (M^+, $C_{16}H_{20}N_4O_5S^+$; calc. 380.115).

*5-{{{4-{[(*tert*-Butoxy)carbonyl]amino}-1-methyl-1*H*-imidazol-2-yl}carbonyl}amino}thiophene-2-carboxylic acid* (Boc-Im-Tp-OH; **17**). As described for **15**, from **37** (90 mg, 0.23 mmol): **17** (80 mg, 91%). White solid. TLC (hexanes/AcOEt 5:2, 10% AcOH): R_f 0.6. IR (film): 3246, 2965, 1674, 1556, 1509, 1456, 1314, 1271, 1237, 1162, 1102, 1023, 750. ^{1}H-NMR ((D$_6$)DMSO): 11.79 (*s*, 1 H); 9.31 (*s*, 1 H); 7.52 (*d*, $J = 3.9$, 1 H); 7.32 (*s*, 1 H); 7.04 (*d*, $J = 3.9$, 1 H); 3.93 (*s*, 3 H); 1.43 (*s*, 9 H). ^{13}C-NMR ((D$_6$)DMSO): 163.4; 155.4; 152.8; 145.2; 136.7; 132.5; 131.3; 123.8; 115.2; 113.4; 79.1; 35.2; 28.1. EI-MS: 366.100 (M^+, $C_{15}H_{18}N_4O_5S^+$; calc. 366.100).

Methyl 3-Hydroxythiophene-2-carboxylate (Ht-OMe; **38**). To dry MeOH (81 ml) under N_2 was added Na metal (3.68 g, 304 mmol). After H_2 evolution had stopped, the soln. was cooled to 0°, and methyl thioglycolate (10 g, 179 mmol) was added dropwise. Methyl 2-chloroacrylate (10.88 g, 179 mmol) in MeOH (21 ml) was then added dropwise ($\rightarrow$ cloudy yellow precipitate). The soln. was allowed to warm to r.t. and stirred for 2 h,

whereupon the precipitate turned dark brown. Evaporation gave a dark yellow solid which was treated with 4N HCl (→ pH 2). The aq. layer was extracted with CH_2Cl_2 (3 × 150 ml), the org. layer was washed with H_2O (3 × 150 ml), dried ($MgSO_4$), and evaporated, and the residual dark oil was subjected to CC (hexanes/AcOEt 20:1): **38** (18.4 g, 64.7%). Clear crystalline solid. TLC (hexanes/AcOEt 20:1): R_f 0.47. IR (film): 3334, 3112, 2955, 1716, 1664, 1552, 1444, 1415, 1350, 1296, 1208, 1104, 1032, 781. ^{1}H-NMR ($CDCl_3$): 9.58 (*s*, 1 H); 7.59 (*d*, *J* = 5.7, 1 H); 6.75 (*d*, *J* = 4.8, 1 H); 3.90 (*s*, 3 H). ^{13}C-NMR ($CDCl_3$): 131.7; 119.4; 52.2. EI-MS: 158.004 (M^+, $C_6H_6O_3S^+$; calc. 158.004).

Methyl 3-Hydroxy-4-nitrothiophene-2-carboxylate (NO_2-Ht-OMe; **39**). Ester **38** (15 g, 98 mmol) was added to conc. H_2SO_4 (48 ml) and stirred until homogeneous. The soln. was then cooled to −10 to 0°, and HNO_3 (4.3 ml) in conc. H_2SO_4 (24 ml) was added dropwise with sufficient cooling to keep the temp. below 0°. After the addition, the soln. was stirred at 0° for 3 h. The resulting black soln. was added to ice and extracted with CH_2Cl_2 (3 × 150 ml), the extract was dried ($MgSO_4$) and evaporated, and the resulting residue was chromatographed (silica gel; hexanes/AcOEt 5:1): **39** (8.2 g, 37.6%). Yellow solid. TLC (hexanes/AcOEt 5:1): R_f 0.12. IR (film): 3107, 1674, 1561, 1520, 1446, 1368, 1267, 1212, 1127, 974, 900, 842, 773. ^{1}H-NMR ($CDCl_3$): 10.15 (*s*, 1 H); 8.43 (*s*, 1 H); 3.97 (*s*, 3 H). ^{13}C-NMR ($CDCl_3$): 164.6; 155.6; 132.6; 53.0. EI-MS: 202.989 (M^+, $C_5H_5NO_5S^+$; calc. 202.989).

Methyl 3-Methoxy-4-nitrothiophene-2-carboxylate (NO_2-Mt-OMe; **40**). A mixture of **39** (1.5 g, 7.4 mmol) and THF (29.5 ml) was cooled to 0°. CH_2N_2 (341 mg, 27 ml, 8.12 mmol) in Et_2O was slowly added by means of a plastic funnel. After several seconds, N_2 evolution ceased, and the soln. was allowed to warm to r.t. A few drops of glacial AcOH were added to ensure the complete consumption of CH_2N_2. The solvent was evaporated: **40** (1.49 g, 93%). Yellow solid. TLC (hexanes/AcOEt 1:1): R_f 0.72. IR (film): 3115, 2960, 1725, 1555, 1506, 1453, 1435, 1387, 1353, 1281, 1199, 1110, 1059, 954, 772. ^{1}H-NMR ($CDCl_3$): 8.38 (*s*, 1 H); 4.10 (*s*, 1 H); 3.94 (*s*, 3 H). ^{13}C-NMR ($CDCl_3$): 156.5; 131.3; 63.9; 53.0. EI-MS: 217.004 (M^+, $C_7H_7NO_5S^+$; calc. 217.004).

Methyl 4-Amino-3-methoxythiophene-2-carboxylate Hydrochloride ($HCl \cdot H_2N$-Mt-OMe; **4**). A mixture of **40** (800 mg, 3.69 mmol) and $SnCl_2 \cdot 2\,H_2O$ (6.66 g, 29.5 mmol) in 95% EtOH (29.5 ml) was stirred vigorously at r.t. Conc. HCl soln. (29.5 ml) was added dropwise, and the soln. was heated at 35° for 6 h. The mixture was removed from heat and adjusted to pH 9 with 4N NaOH (15 ml). The resulting white emulsion was extracted with AcOEt (3 × 100 ml), dried ($MgSO_4$), and evaporated. After addition of a small amount of fresh AcOEt, 2M HCl in Et_2O was added to precipitate crude **4**. The salt was filtered and taken directly on to the next step.

*Methyl 4-{[(*tert*-Butoxy)carbonyl]amino}-3-methoxythiophene-2-carboxylate* (Boc-Mt-OMe; **41**). A mixture of **4** (1.0 g, 4.47 mmol), Et_3N (498 mg, 0.68 ml, 4.92 mmol), and $(Boc)_2O$ (1.0 g, 4.92 mmol) in CH_2Cl_2 (9 ml) was stirred at 60° for 12 h. The soln. was washed with sat. NH_4Cl soln. (3×) dried ($MgSO_4$), and evaporated, and the resulting red solid was subjected to CC (hexanes/AcOEt 5:1): **41** (528 mg, 41.3%). White solid. TLC (hexanes/AcOEt 10:1): R_f 0.24. IR (film): 3437, 3329, 2979, 1772, 1716, 1530, 1440, 1376, 1230, 1165, 1081, 1055, 993, 861, 778. ^{1}H-NMR ($CDCl_3$): 7.58 (*s*, 1 H); 6.82 (*s*, 1 H); 4.06 (*s*, 3 H); 3.85 (*s*, 3 H); 1.52 (*s*, 9 H). ^{13}C-NMR ($CDCl_3$): 161.5; 152.6; 151.5; 130.2; 112.0; 111.2; 85.5; 81.3; 62.9; 52.3; 28.6; 28.0. EI-MS: 287.083 (M^+, $C_{12}H_{17}NO_5S^+$; calc. 287.083).

*4-{[(*tert*-Butoxy)carbonyl]amino}-3-methoxythiophene-2-carboxylic Acid* (Boc-Mt-OH; **11**). A mixture of **41** (250 mg, 0.87 mmol) and KOH (48.8 mg, 0.87 mmol) in dry MeOH (1 ml), was heated to 50° for 6 h. The soln. was added to CH_2Cl_2 (5 ml) and H_2O (5 ml). The aq. layer was washed with CH_2Cl_2 (3 × 10 ml) and acidified to pH 3 with 1N HCl (3.5 ml). The aq. soln. was then extracted with CH_2Cl_2 (3 × 10 ml), and the extract was dried ($MgSO_4$) and evaporated; **11** (193 mg, 81%). Off-white solid. TLC (hexanes/AcOEt 10:1): R_f 0.24. IR (film): 3400, 1699, 1526, 1438, 1369, 1230, 1154, 1053. ^{1}H-NMR ($CDCl_3$): 7.68 (*s*, 1 H); 6.84 (*s*, 1 H); 4.08 (*s*, 3 H); 1.53 (*s*, 9 H). ^{13}C-NMR ($CDCl_3$): 165.8; 152.7; 130.4; 112.0; 111.2; 81.4; 63.3; 62.3; 28.7. EI-MS: 273.067 (M^+, $C_{11}H_{15}NO_5S^+$; calc. 273.067).

*2-{[(*tert*-Butoxy)carbonyl]amino}-5-methylthiazole-4-carboxylic Acid* (Boc-Nt-OH; **10**). The 2-oxabutanoic acid (10 g, 98 mmol) was treated dropwise with Br_2 (8 ml, 25 g, 157 mmol) while stirring. Upon completion of addition, the mixture was stirred until the red color of Br_2 had dissipated. Thiourea (14.8 g, 196 mmol) was then added in portions, and stirring was continued overnight. The mixture was acidified with conc. HCl soln., and the precipitated hydrochloride was filtered and washed with cold EtOH. The crude solid was taken into DMF (50 ml), then DIEA (10 ml) and $(Boc)_2O$ (21.4 g, 98 mmol) were added. The mixture was stirred at 60° for 12 h, then diluted with AcOEt, and washed with brine (3×). The combined org. phase was dried (Na_2SO_4) and evaporated. The resulting crude oil was dissolved in MeOH (0.1 l) and 1N NaOH (0.1 l) and stirred at r.t. for 1 h. The MeOH was then evaporated, and the aq. layer was washed with Et_2O (2 × 0.1 l). The aq. phase was acidified to pH 2 with 1N HCl and extracted with AcOEt (3 × 0.1 l). The combined org. phase was dried (Na_2SO_4) and evaporated: **10** (11 g, 44%). White flaky solid. TLC (hexanes/AcOEt 5:2, 10% AcOH): R_f 0.4.

IR (film): 3191, 2978, 1714, 1669, 1578, 1562, 1317, 1165. ^{1}H-NMR ((D$_6$)DMSO): 11.54 (*s*, 2 H); 2.54 (*s*, 3 H); 1.44 (*s*, 9 H). ^{13}C-NMR ((D$_6$)DMSO): 164.1; 155.4; 137.1; 136.9; 81.9; 28.6; 12.94; 11.61. EI-MS: 258.067 (M^+, $C_{10}H_{14}N_2O_4S^+$; calc. 258.067).

3. *Hairpin-Polyamide Synthesis.* Polyamides were synthesized from Boc-β-alanine-Pam resin (50 mg, 0.59 mmol/g) and purified by prep. HPLC according to published manual solid-phase protocols [20].

Im-Im-Nh-Py-γ-Im-Py-Py-Py-β-Dp (**43**). Boc-Nh-OH (**12**; 33 mg, 0.147 mmol) was incorporated by activation with HBTU (53 mg, 0.140 mmol), DIEA (50 µl), and DMF (300 µl). The mixture was allowed to stand for 15 min at r.t. and then added to the NH_2-Py-γ-Im-Py-Py-Py-β-Pam resin. Coupling was allowed to proceed for 1.5 h at r.t. After Boc deprotection, Boc-Im-OH (35 mg, 0.147 mmol) was activated with HBTU (53 mg, 0.140 mmol), DIEA (50 µl), and DMF (300 µl). The mixture was allowed to stand for 15 min at r.t. and then added to the NH_2-Nh-Py-γ-Im-Py-Py-Py-β-Pam resin. Coupling was allowed to proceed for 1.5 h at r.t., and determined to be complete by anal. HPLC. After Boc deprotection, the terminal imidazole residue was added by means of Im-$COCCl_3$: Im-$COCCl_3$ (67 mg, 0.295 mmol), DIEA (50 µl), and DMF (600 µl) were added to NH_2-Im-Nh-Py-γ-Im-Py-Py-Py-β-Pam resin. Coupling was allowed to proceed for 2 h at 37°, and determined to be complete by anal. HPLC. To Im-Im-Nh-Py-γ-Im-Py-Py-Py-β-Pam resin (50 mg) in a 20-ml scintillation vial, Dp (1 ml) was added. The mixture was allowed to stand for 2 h at 85° with occasional agitation. The resin was then filtered, and the soln. was diluted to 8 ml with 0.1% CF_3COOH soln. The sample was purified by reversed-phase HPLC to provide **43** (2 mg, 5.6% recovery). Fine white powder after lyophilization. MALDI-TOF-MS (monoisotopic): 1209.59 ($[M+H]^+$, $C_{56}H_{69}N_{22}O_{10}^+$; calc. 1209.56).

Im-Im-Im-Py-γ-Im-Py-Py-Py-β-Dp (**42**). Boc-Im-OH (**6**) was incorporated according to previously described procedures [20]. The terminal imidazole residue was incorporated and the compound purified as described for **43** to provide **42** (2.6 mg, 6.0% recovery). Fine white powder after lyophilization. MALDI-TOF-MS (monoisotopic): 1210.56 ($[M+H]^+$, $C_{56}H_{70}N_{23}O_{10}^+$; calc. 1224.56).

Im-Im-Tn-Py-γ-Im-Py-Py-Py-β-Dp (**44**). Boc-Im-Tn-OH (**16**; 56 mg, 0.147 mmol) was incorporated by activation with HBTU (53 mg, 0.140 mmol), DIEA (50 µl), and DMF (300 µl). The mixture was allowed to stand for 15 min at r.t. and then added to the NH_2-Py-γ-Im-Py-Py-Py-β-Pam resin. Coupling was allowed to proceed for 24 h at 37°. After Boc deprotection, the terminal imidazole residue was incorporated as described for **43**. The compound was cleaved from the resin and purified as described for **43**: **44** (2.1 mg, 5.7% recovery). Fine white powder after lyophilization. MALDI-TOF-MS (monoisotopic): 1240.53 ($[M+H]^+$, $C_{57}H_{70}N_{21}O_{11}S^+$; calc. 1240.53).

Im-Im-Tp-Py-γ-Im-Py-Py-Py-β-Dp (**45**). Boc-Im-Tp-OH (**17**) was incorporated as described for **44**: **45** (1.8 mg, 4.9% recovery). Fine white powder after lyophilization. MALDI-TOF-MS (monoisotopic): 1226.53 ($[M+H]^+$, $C_{57}H_{70}N_{21}O_{11}S^+$; calc. 1226.52).

Im-Im-Ht-Py-γ-Im-Py-Py-Py-β-Dp (**46**). Boc-Mt-OH (**11**; 42 mg, 0.147 mmol) was incorporated by activation with HBTU (53 mg, 0.140 mmol), DIEA (50 µl), and DMF (300 µl). The mixture was allowed to stand for 15 min at r.t. and then added to the NH_2-Py-γ-Im-Py-Py-Py-β-Pam resin. Coupling was allowed to proceed for 20 h at 37°. After Boc deprotection, Boc-Im-OH (35 mg, 0.147 mmol) was activated with HBTU (53 mg, 0.140 mmol), DIEA (50 µl), and DMF (300 µl). The mixture was allowed to stand for 15 min at r.t. and then added to the NH_2-Mt-Py-γ-Im-Py-Py-Py-β-Pam resin. Coupling was allowed to proceed for 40 h at 37°, and determined to be complete by anal. HPLC. After Boc deprotection, the terminal imidazole residue was incorporated as described for **43**. The compound was cleaved from the resin and purified as described for **43** to provide the methoxy-protected Mt-containing polyamide Im-Im-Mt-Py-γ-Im-Py-Py-Py-β-Dp (2.0 mg, 5.4% recovery): White powder after lyophilization. MALDI-TOF-MS (monoisotopic): 1256.54 ($[M+H]^+$, $C_{57}H_{70}N_{21}O_{11}S^+$; calc. 1256.53).

The Mt-containing polyamide was then dissolved in DMF (200 µl) and added to a suspension of NaH (40 mg; 60% oil dispersion) and PhSH in DMF (400 µl) that was preheated for 5 min at 100°. The mixture was heated for 2 h at 100° and then cooled to 0°, and 20% CF_3COOH soln. (7.0 ml) was added. The aq. layer was washed with Et_2O (3 × 8 ml) and then diluted to a total vol. of 9.5 ml with 0.1% CF_3COOH soln. The mixture was then purified by reversed-phase HPLC to give the deprotected Ht-containing polyamide **46** (0.83 mg, 41% recovery). Fine white powder after lyophilization. MALDI-TOF-MS (monoisotopic): 1242.51 ($[M+H]^+$, $C_{56}H_{68}N_{21}O_{11}S^+$; calc. 1242.51).

Im-Im-Fr-Py-γ-Im-Py-Py-Py-β-Dp (**47**). Boc-Im-Fr-OH (**15**; 51 mg, 0.147 mmol) was incorporated by activation with HBTU (53 mg, 0.140 mmol), DIEA (50 µl), and DMF (300 µl). The mixture was allowed to stand for 15 min at r.t. and then added to the NH_2-Py-γ-Im-Py-Py-Py-β-Pam resin. Coupling was allowed to proceed for 1.5 h at r.t. After Boc deprotection, the terminal imidazole residue was incorporated as described for **43**. The compound was cleaved from the resin and purified as described for **43**: **47** (1.5 mg, 4.2% recovery). Fine

white powder after lyophilization. MALDI-TOF-MS (monoisotopic): 1210.54 ($[M+H]^+$, $C_{56}H_{68}N_{21}O_{11}^+$; calc. 1210.54).

4. *1:1 Motif Polyamide Synthesis.* – Polyamides were synthesized from Boc-β-alanine-Pam resin (50 mg, 0.59 mmol/g) and purified by prep. HPLC according to published manual solid-phase protocols [20].

Im-β-Im-Py-β-Py-β-Im-Py-β-Dp (**48**). To Im-β-Im-Py-β-Py-β-Im-Py-β-Pam resin (120 mg) in a 20-ml scintillation vial, Dp (2 ml) was added. The mixture was allowed to stand for 2 h at 85° with occasional agitation. The resin was then filtered, and the soln. was diluted to 8 ml with 0.1% CF_3COOH soln. The sample was purified by reversed-phase HPLC: **48** (12 mg, 15.3% recovery). Fine white powder after lyophilization. MALDI-TOF-MS (monoisotopic): 1107.70 ($[M+H]^+$, $C_{50}H_{66}N_{20}O_{10}^+$; calc. 1107.53).

Im-β-Im-Py-β-Hp-β-Im-Py-β-Dp (**49**). The polyamide was synthesized, deprotected, and purified according to the previously published protocol [17]: **49** (5.6 mg, 7.0% recovery). White powder after lyophilization. MALDI-TOF-MS (monoisotopic): 1124.20 ($[M+H]^+$, $C_{50}H_{67}N_{20}O_{11}^+$; calc. 1124.19).

Im-β-Im-Py-β-Nh-β-Im-Py-β-Dp (**50**). Boc-Nh-OH (**12**; 271 mg, 1.2 mmol) was incorporated by activation with DCC (247 mg, 1.2 mmol) and HOBt (141 mg, 1.2 mmol) in DMF (2 ml). The mixture was shaken at 37° for 30 min and filtered into the reaction vessel containing NH_2-β-Im-Py-β-Pam resin. DIEA (400 µl) was added, and coupling was allowed to proceed for 1.5 h at r.t. After Boc deprotection, Boc-β-OH (227 mg, 1.2 mmol) was activated with HBTU (432 mg, 1.14 mmol), DIEA (400 µl), and DMF (2 ml). The mixture was allowed to stand for 15 min at r.t. and then added to the NH_2-Nh-β-Im-Py-β-Pam resin. Coupling was allowed to proceed for 2 h at 37°, and determined to be complete by anal. HPLC. The compound was cleaved from the resin and purified as described for **48**: **50** (9 mg, 11.4% recovery). Fine white powder after lyophilization. MALDI-TOF-MS (monoisotopic): 1107.70 ($[M+H]^+$, $C_{49}H_{64}N_{20}O_{10}^+$; calc. 1107.53).

Im-β-Im-Py-β-Ht-β-Im-Py-β-Dp (**51**). The polyamide was synthesized from Boc-Mt-OH (**11**), deprotected, and purified as described for **46**: **51** (1.1 mg, 2.8% recovery). White powder after lyophilization. MALDI-TOF-MS (monoisotopic). 1126.43 ($[M+H]^+$, $C_{49}H_{64}N_{19}O_{11}S^+$; calc. 1126.47).

Im-β-Im-Py-β-Fr-β-Im-Py-β-Dp (**52**). Boc-β-Fr-OH (**13**; 369 mg, 1.2 mmol) was incorporated by activation with DCC (247 mg, 1.2 mmol) and HOBt (141 mg, 1.2 mmol) in DMF (2 ml). The mixture was shaken at 37° for 30 min and filtered into the vessel containing the NH_2-β-Im-Py-β-Pam resin. DIEA (400 µl) was added, and coupling was allowed to proceed for 1.5 h at r.t. The compound was cleaved from the resin and purified as described for **48**: **52** (6 mg, 7.7% recovery). White powder after lyophilization. MALDI-TOF-MS (monoisotopic): 1094.50 ($[M+H]^+$, $C_{49}H_{63}N_{19}O_{11}^+$; calc. 1094.60).

Im-β-Im-Py-β-Nt-β-Im-Py-β-Dp (**53**). Boc-Nt-OH (**10**; 309 mg, 1.2 mmol) was incorporated by activation with HBTU (432 mg, 1.14 mmol), DIEA (400 µl), and DMF (2 ml). The mixture was allowed to stand for 15 min at r.t. and then added to the NH_2-β-Im-Py-β-Pam resin. Coupling was allowed to proceed for 20 h at 37°. After Boc deprotection, Boc-β-OH (227 mg, 1.2 mmol) was activated with HBTU (432 mg, 1.14 mmol), DIEA (400 µl), and DMF (2 ml). The mixture was allowed to stand for 15 min at r.t. and then added to the NH_2-Nt-β-Im-Py-β-Pam resin. Coupling was allowed to proceed for 48 h at 37°, and determined to be complete by anal. HPLC. The compound was cleaved from the resin and purified as described for **48**: **53** (4.2 mg, 5.2% recovery). Fine white powder after lyophilization. MALDI-TOF-MS (monoisotopic): 1125.50 ($[M+H]^+$, $C_{49}H_{64}N_{20}O_{10}S^+$; calc. 1125.49).

Im-β-Im-Py-β-Tn-β-Im-Py-β-Dp (**54**). Boc-β-Tn-OH (**14**; 393 mg, 1.2 mmol) was incorporated by activation with HBTU (432 mg, 1.14 mmol), DIEA (400 µl), and DMF (2 ml). The mixture was allowed to stand for 15 min at r.t. and then added to the NH_2-β-Im-Py-β-Pam resin. Coupling was allowed to proceed for 20 h at 37°. The compound was cleaved from the resin and purified as described for **48**: **54** (5.8 mg, 7.2% recovery). Fine white powder after lyophilization. MALDI-TOF-MS (monoisotopic): 1124.50 ($[M+H]$, $C_{50}H_{65}N_{19}O_{10}S^+$; calc. 1124.49).

Im-β-Im-Py-β-Th-β-Im-Py-β-Dp (**55**). Compound **55** was prepared by the protocol used for **53**, but from Boc-Th-OH (**8**): **55** (6.0 mg, 7.5% recovery). Fine white powder after lyophilization. MALDI-TOF-MS (monoisotopic): 1125.60 ($[M+H]^+$, $C_{49}H_{64}N_{20}O_{10}S^+$; calc. 1125.49).

5. *Footprinting Experiments.* Plasmids pDHN1 and pAU8 were constructed and 5′-radiolabeled as previously described [11][30]. DNase-I-footprint titrations were performed according to standard protocols [24].

We are grateful for financial support from the *National Institutes of Health* and Caltech for a *James Irvine Fellowship* to *R. M. D.*

REFERENCES

[1] M. L. Kopka, C. Yoon, D. Goodsell, P. Pjura, R. E. Dickerson, *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 1376.
[2] J. W. Lown, K. Krowicki, U. G. Bhat, A. Skorobogaty, B. Ward, J. C. Dabrroiak, *Biochemistry* **1986**, *25*, 7408.
[3] J. G. Pelton, D. E. Wemmer, *J. Am. Chem. Soc.* **1989**, *112*, 1393.
[4] P. B. Dervan, *Bioorg. Med. Chem*. **2001**, *9*, 2215.
[5] W. S. Wade, M. Mrksich, P. B. Dervan, *J. Am. Chem. Soc*. **1992**, *114*, 8783.
[6] C. L. Kielkopf, E. E. Baird, P. B. Dervan, D. C. Rees, *Nat. Struct. Biol*. **1998**, *5*, 104.
[7] S. White, J. W. Szewczyk, J. M. Turner, E. E. Baird, P. B. Dervan, *Nature (London)* **1998**, *391*, 468.
[8] C. L. Kielkopf, S. White, J. W. Szewczyk, J. M. Turner, E. E. Baird, P. B. Dervan, D. C. Rees, *Science (Washington, D.C.)* **1998**, *282*, 111.
[9] M. Lee, K. Krowicki, R. G. Shea, J. W. Lown, R. T. Pon, *J. Mol. Recogn.* **1989**, *2*, 84.
[10] S. K. Sharma, M. Tandon, J. W. Lown, *J. Org. Chem*. **2000**, *65*, 1102.
[11] D. H. Nguyen, J. W. Szewczyk, E. E. Baird, P. B. Dervan, *Bioorg. Med. Chem*. **2001**, *9*, 7.
[12] J. Z. Zheng-Yun, P. B. Dervan, *Bioorg. Med. Chem*. **2000**, *8*, 2467.
[13] M. Mrksich, P. B. Dervan, *J. Am. Chem. Soc*. **1994**, *116*, 3663.
[14] J. W. Trauger, E. E. Baird, P. B. Dervan, *Nature (London)* **1996**, *382*, 559.
[15] J. W. Trauger, E. E. Baird, M. Mrksich, P. B. Dervan, *J. Am. Chem. Soc*. **1996**, *118*, 6160.
[16] S. Janssen, T. Durussel, U. K. Laemmli, *Mol. Cell* **2000**, *6*, 999.
[17] A. R. Urbach, J. W. Szewczyk, S. White, J. M. Turner, E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc*. **1999**, *121*, 11621.
[18] P. B. Dervan, A. R. Urbach, in 'Essays in Contemporary Chemistry, from Molecular Structure Toward Biology', Eds. G. Quinkert and M. V. Kisakürek, Verlag Helvetica Chimica Acta, Zürich, 2000.
[19] A. R. Urbach, J. J. Love, S. A. Ross, P. B. Dervan, *J. Mol. Biol*. **2002**, *320*, 55.
[20] E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc*. **1996**, *118*, 6141.
[21] B. D. Dorsey, J. M. Hoffman Jr., S. A. Joseph, S. L. McDaniel, *J. Heterocycl. Chem*. **1995**, *32*, 1283.
[22] K. E. Rao, R. G. Shea, B. Yadagiri, J. W. Lown, *Anti-Cancer Drug Des.* **1990**, *5*, 3.
[23] P. R. Huddleston, J. M. Barker, *Synth. Commun.* **1979**, *9*, 731.
[24] J. W. Trauger, P. B. Dervan, *Methods Enzymol*. **2001**, *340*, 450.
[25] 'Spartan Essential', Wavefunction Inc., 1991 – 2001.
[26] D. M. Herman, J. M. Turner, E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc*. **1999**, *121*, 1121.
[27] J. J. Kelly, E. E. Baird, P. B. Dervan, *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 6981.
[28] R. P. L. de Clairac, C. L. Seel, B. H. Geierstanger, M. Mrksich, E. E. Baird, P. B. Dervan, D. E. Wemmer, *J. Am. Chem. Soc*. **1999**, *121*, 2956.
[29] C. L. Kielkopf, R. E. Bremer, S. White, J. W. Szewczyk, J. M. Turner, E. E. Baird, P. B. Dervan, D. C. Rees, *J. Mol. Biol*. **2000**, *295*, 557.
[30] A. R. Urbach, P. B. Dervan, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 4343.
[31] M. L. Kopka, D. S. Goodsell, G. W. Han, T. K. Chiu, J. W. Lown, R. E. Dickerson, *Structure* **1997**, *5*, 1033.

Received August 5, 2002